McGRAW-HILL PUBLICATIONS IN THE ZOOLOGICAL SCIENCES

A. FRANKLIN SHULL, Consulting Editor

QUANTITATIVE ZOOLOGY

SELECTED TITLES FROM

McGRAW-HILL PUBLICATIONS IN THE ZOOLOGICAL SCIENCES

A. FRANKLIN SHULL, *Consulting Editor*

Chapman · ANIMAL ECOLOGY

Goldschmidt · PHYSIOLOGICAL GENETICS

Graham · FOREST ENTOMOLOGY

Haupt · FUNDAMENTALS OF BIOLOGY

Metcalf and Flint · INSECT LIFE

Mitchell · GENERAL PHYSIOLOGY

Mitchell and Taylor · LABORATORY MANUAL OF GENERAL PHYSIOLOGY

Pearse · ANIMAL ECOLOGY

Reed and Young · LABORATORY STUDIES IN ZOOLOGY

Riley and Johannsen · MEDICAL ENTOMOLOGY

Rogers · TEXTBOOK OF COMPARATIVE PHYSIOLOGY

LABORATORY OUTLINES IN COMPARATIVE PHYSIOLOGY

Senning · LABORATORY STUDIES IN COMPARATIVE ANATOMY

Shull · EVOLUTION

HEREDITY

Shull, LaRue, and Ruthven · PRINCIPLES OF ANIMAL BIOLOGY

Simpson and Roe · QUANTITATIVE ZOOLOGY

Snodgrass · PRINCIPLES OF INSECT MORPHOLOGY

Van Cleave · INVERTEBRATE ZOÖLOGY

Welch · LIMNOLOGY

Wieman · GENERAL ZOOLOGY

AN INTRODUCTION TO VERTEBRATE EMBRYOLOGY

Wolcott · ANIMAL BIOLOGY

There are also the related series of McGraw-Hill Publications in the Botanical Sciences, of which Edmund W. Sinnott is Consulting Editor, and in the Agricultural Sciences, of which Leon J. Cole is Consulting Editor.

QUANTITATIVE ZOOLOGY

Numerical Concepts and Methods in the Study of Recent and Fossil Animals

BY

GEORGE GAYLORD SIMPSON, PH.D.

Associate Curator of Vertebrate Palaeontology, The American Museum of Natural History

AND

ANNE ROE, PH.D.

FIRST EDITION

McGRAW-HILL BOOK COMPANY, INC.

NEW YORK AND LONDON

1939

THE MAPLE PRESS COMPANY, YORK, PA.

To

R. A. FISHER

RAYMOND PEARL

and

ALEXANDER G. RUTHVEN

Leaders in devising, explaining, and applying
sound quantitative methods in the life sciences

PREFACE

The exclusion of zoology from the roster of the exact sciences has usually been a subject of self-congratulation for zoologists and of reproach for their more mathematically inclined associates. The abstract mathematician is inclined to believe that the reduction to numbers of all descriptions of the phenomena of nature is not only theoretically possible but also practically desirable. Some zoologists would grant the theoretical possibility; but few conceive of it as at present practicable, and perhaps none admit its general desirability. The zoologist is not, and surely should not be, interested in reducing his observations or theories to a purely numerical basis simply because he likes numbers. His interest is not at all in formulas or digits, but in animals. He is concerned with the anatomy, behavior, and relationships of these animals; and he quite properly refuses to fit his studies into any a priori framework, such as that of formal mathematics or statistics.

In zoology, numbers and formulas are of no interest or value for their own sake but only to the extent that they may be the best means of describing and of interpreting what animals are and do. In presenting to zoologists and paleontologists a book especially devoted to numerical concepts and containing many mathematical symbols and formulas, it is essential to state this fact at the outset and to maintain and stress this viewpoint throughout, as we have done. The symbols are merely shorthand expressions for concepts that necessarily enter into most work in zoology. The formulas are only the most convenient and usable way of summarizing operations that have logical, common-sense meanings.

If in the course of using such mathematical methods these zoological, nonmathematical implications are lost sight of the zoologist will also lose sight of the whole purpose of his work and will fall into futility or even absurdity, although his arithmetic is perfectly correct. While urging and facilitating the use of numerical methods, the authors have tried at every point to

guard against these grave dangers and to insist that the methods be used zoologically, not by rote and not as mathematical abstractions.

Whether from inertia, from ignorance, or from natural mistrust engendered by some ill-advised efforts, or from a combination of these, most zoologists and paleontologists have distrusted the overt use of any but the very simplest and most obvious numerical methods. They have, in fact, been dealing with some very subtle and difficult numerical concepts, but too often they have failed to recognize them as such, and frequently this has resulted in their serious misuse. There is little need for the use of more numerical concepts and methods in zoology, but there is great need for recognizing them as such and for according them more correct and careful treatment. To the still rather slight extent that this improvement has been made and the great extent to which it can be made, the data of zoology are susceptible of marked betterment, its conclusions can be made more logical and reliable, and it can reach out to interpretive phases that are of basic importance and that have now been barely touched.

If zoology and paleontology have lagged behind most other sciences in their numerical methods, a major reason has been the extreme difficulty of learning the methods that are known in other fields and of adapting them to this one. In order to obtain the mathematical and statistical information pertinent to his own problems, the zoologist has had to wade through great masses of difficult material, most of it not directly useful to him and none specifically arranged for his purposes.

This was the experience of the senior author of the present text some ten years ago when he set out to work toward a conscious, rational numerical methodology, the absence of which was increasingly apparent to him in his own work and in that of almost all his colleagues. In order to accomplish anything in this line without abandoning his regular work, it was necessary to seek the aid and collaboration of someone who, without being a professional mathematician, was thoroughly familiar with the desirable mathematical and statistical concepts and was accustomed to using them in practical research in some life science. The junior author of this volume possessed these qualifications, and for several years she has devoted much of her time to the joint research in methodology of which this book is one result.

It was first anticipated that the senior author would write the more zoological and the junior author the more mathematical parts of the book, but in the end each author contributed so largely to both and the two aspects of the subject became so completely fused that individual responsibility cannot be assigned for any passage. For the sake of uniformity of style and treatment, the actual words of the final manuscript were written by the senior author, but the work involved has been equally shared and the collaboration is complete.

This text presupposes no knowledge whatever of statistics, and of mathematics no more than elementary algebra. If the reader is willing to take for granted a few derivations and transformations of formulas, almost all the procedures recommended can be carried out with no more elaborate preparation than a knowledge of arithmetic. Anyone studying college zoology or engaged in research in this field can understand and use these methods, however slight his mathematical interests or training. Many of the passages, essential as they are for our subject, deal with quantitative concepts without involving any actual mathematical operations.

The discussion of concepts is based on long experience in the use of numerical data in research in several different life sciences and on wide reading of publications presenting such data. Every procedure recommended or explained has been tested in connection with some real zoological problem. A great mass of material was rejected because it appeared to be of relatively little use to the average zoologist, and many procedures were modified to adapt them to this purpose. Much of the material in the first few chapters is well known to working zoologists, but it is important to present it to students more consistently and fully than has been customary; even advanced workers may profit by an explicit statement of some of the fundamentals underlying their practices. Some current procedures, even of the simplest sort, are open to criticism and require defense or abandonment.

Most of the more advanced procedures, especially those of a statistical sort, are widely employed in other sciences but have not generally been adopted by zoologists or modified for their use. From their special point of view, the authors find that some statistical procedures in common use are inadequate or erroneous, and these have been modified or criticized, without intending

any general survey of statistical theory. Some new concepts and methods have been proposed, for instance, those discussed in connection with single specimens.

Although statistics necessarily figures in much of this work, this is a book on the methodology of zoology, and not on statistics or even (as a whole) on the use of statistics in zoology. Advanced statisticians, if any refer to this work, may detect in it what will appear to them two principal shortcomings; these should be briefly mentioned here. In the first place, the authors have not followed Fisher and some of the other most specialized recent statisticians in using a double set of symbols and terms, one for the parameters of populations and one for the calculated estimates of these parameters. They have explained and have maintained the important distinction here involved, probably to a greater extent than in any other introduction to the use of statistics in a particular field of research, but they believe that the use of two systems of notation, instead of clarifying this distinction, would prove confusing, repellent, and unnecessarily difficult for their special audience. In fact, within this field the dual system would not be much more logical than the one that has been used, for the group concepts involved are not two but three—samples, actual populations, and theoretical or abstract populations—or even four, since the theoretical population envisioned by hypothetical deduction from the sample is not the same as the abstraction derivable from the real population.

These distinctions, which may appear unduly subtle to a zoological reader if this is read before the body of the text, but which should later become sufficiently clear, are related to the problem of inverse probability. This problem is essential in the philosophy of numerical method, and it underlies the whole structure of statistics as a logical treatment of data. The authors have been aware of the problem, and they have endeavored to avoid errors in this respect (errors commonly made even by professional statisticians), but they have not included any explicit and detailed treatment of inverse probability. A condensed discussion would be almost incomprehensible to the average zoological reader; an adequate discussion would carry the book far beyond reasonable length and belongs in an advanced treatment of statistics rather than in an introductory work on zoological method.

As the clearest method of explanation, examples have been freely introduced, generally based on real and typical zoological data. These data have been drawn from many sources, as credited in each case. With few exceptions any calculations based on these data have been made by the present authors, rather than by the original publisher of the raw data. The calculation has been done twice: first by hand or with simple aids available to all zoologists and second on modern electric calculating machines. The purpose of the duplication was not only to eliminate errors so far as possible but also to test the practicability and the accuracy of carrying out such work by the simplest means, within the reach of any student in this field.

Calculating machines are a great convenience, but most zoologists do not now have them; they are not necessary for correct numerical treatment of zoological data, and to have assumed that they would be used would greatly have decreased the usefulness and acceptance of this work. Formulas for their use are therefore relegated to an appendix, and in the main text it is assumed that calculation will be done by hand or with an inexpensive pocket abacus and a slide rule, plus a few numerical tables—equipment readily available to and easily used by any student of zoology.

Without making themselves responsible for our judgment or use of the data, several colleagues, including specialists in almost all the various divisions of zoology and paleontology, have suggested sources of numerical data and discussions and exemplifications of numerical methods. Among those to whom the authors are indebted for this or similar assistance are Dr. E. H. Colbert, Dr. J. E. Hill, Mr. J. T. Nichols, Dr. G. K. Noble, and Mr. J. T. Zimmer. Mr. Morton Jellinek criticized certain statistical parts of the manuscript and pointed out some pitfalls in this field. The authors are also indebted to Miss Evelyn Horton and to the statistical laboratory of Teachers College, Columbia University, for the use of calculating machines.

GEORGE GAYLORD SIMPSON,
ANNE ROE.

NEW YORK,
May, 1939.

CONTENTS

PREFACE vii

LIST OF TABLES xvii

CHAPTER I

TYPES AND PROPERTIES OF NUMERICAL DATA 1

The Meaning of Numbers in Zoology—Data from Direct Observation—Ratios and Indices.

CHAPTER II

MENSURATION 16

Requirements of Good Measurement—Systems of Mensuration—Accuracy and Refinement—Bias and Consistency.

CHAPTER III

FREQUENCY DISTRIBUTIONS AND GROUPING 34

Frequency Distributions—Grouping—Frequency Distributions and Grouping of Discontinuous Variates—Numerical Qualitative Grouping—Criteria for Secondary Grouping.

CHAPTER IV

PATTERNS OF FREQUENCY DISTRIBUTIONS 56

Graphic Representation—The Meaning of Distribution Patterns—General Types of Distribution Patterns—The Normal Curve—The Binomial Distribution—The Poisson Series—Cumulative Distributions: the Ogive.

CHAPTER V

MEASURES OF CENTRAL TENDENCY 85

Arithmetic Mean—Median—Mode—Other Measures of Central Tendency—The Meaning of Average, Typical, and Normal.

CHAPTER VI

MEASURES OF DISPERSION AND VARIABILITY 106

Observed Range—Mean Deviation—Standard Deviation—Semi-interquartile Range—Relative Dispersion and Variability.

CHAPTER VII

THE NORMAL CURVE 129

The Normal Curve and Reality—Parameters of the Normal Curve—Areas and Ordinates—Curve Fitting—Skewness—Kurtosis.

CHAPTER VIII

PROBABILITY AND RELIABILITY 149

Standard Errors—Calculation of Standard Errors—Consistency, Efficiency, and Sufficiency—Significant Figures.

CHAPTER IX

SAMPLING 166

Characteristics of Good Samples—Faunal Sampling—Paleontological Sampling—Sampling Limits.

CHAPTER X

COMPARISONS OF SAMPLES 186

Comparison of a Single Specimen with a Larger Sample—Comparison of the Means of Two Samples—Heterogeneity of Samples.

CHAPTER XI

SMALL SAMPLES AND SINGLE SPECIMENS 203

Deviations from the Mean in Small Samples—Significance of Means of Single Samples—Comparison of the Means of Two Small Samples—Significance of Single Observations.

CHAPTER XII

CORRELATION 222

The Correlation Coefficient—Significance of r—Transformation of r to z—Difference between Two Correlation Coefficients—Cause and Effect and Spurious Correlation—Partial Correlation—Rank Correlation.

CHAPTER XIII

REGRESSION 255

Scatter Diagrams and Trend—Concepts and Rough Approximation of Regression—Regression Coefficients—Significance of Regression—Curvilinear Regression—The Correlation Ratio.

CHAPTER XIV

ASSOCIATION 285

Contingency Classifications—Tests of Significance—$2 \times c$ Contingency Tables—Small Samples—Multiple and Spurious Association—Hypotheses of Dependence.

CHAPTER XV

GRAPHIC METHODS 305

Types of Diagrams—Special Types of Graphic Frequency Distributions—Comparison of Ranges and Means—Proportions and Percentages—Time Series and Periodic Phenomena—Pictorial Diagrams.

CHAPTER XVI

GROWTH 336

Age Groups—Sampling for Growth—Absolute Growth and Growth Rates—Proportions and Size—Heterogony.

APPENDIX

CALCULATION 375
SYMBOLS 380
FORMULAS 382
GLOSSARY 391

BIBLIOGRAPHY

MATHEMATICAL AND STATISTICAL REFERENCES 399
ZOOLOGICAL REFERENCES 401

INDEX 407

LIST OF TABLES

TABLE		PAGE
I.	Areas of the Normal Curve	137
II.	Ordinates of the Normal Curve	139
III.	Sampling Limits	184
IV.	Relationship of P and d/σ	191
V.	Relationship of d/P.E. and P	191
VI.	Table of t	206
VII.	Table of V, D, and Antilog D	217
VIII.	Percentage Difference between Two Observations and Corresponding Minimum Value of V in a Population Including Both	219
IX.	Significance of the Correlation Coefficient r for Different Sizes of Samples (Using the Distribution of t as a Measure of Probability)	236
X.	Corresponding Values of r and z	239
XI.	Standard Errors of z and of the Difference between Two Values of z from Samples of Equal Size	241
XII.	Values of χ^2 Corresponding to Given Values of P and Given Degrees of Freedom	292
XIII.	Criteria of the Significance of the Difference between Two Means Automatically Assumed by the Dice-Leraas Method of Graphic Comparison, with Given Ratios of Sizes of Samples and of Their Standard Errors of Means	320

QUANTITATIVE ZOOLOGY

CHAPTER I

TYPES AND PROPERTIES OF NUMERICAL DATA

THE MEANING OF NUMBERS IN ZOOLOGY

When a zoologist sets out to describe or discuss any animals, he almost inevitably finds that he is using some numbers. Usually, measurements of the dimensions of individual animals are given; the proportions of the different parts of the animal are considered; different animals are compared as to size and proportions; abundance or scarcity of a species may be mentioned; the number of teeth, scales, fin rays, vertebrae, and the like are recorded; and in many other ways essentially numerical facts and deductions enter into the work. Commonly these observations are given by actual numbers; but not infrequently they may be expressed in words and without the use of figures. When it is said that one species is larger than another, that a given animal is abundant in a certain area, or that a mammal lacks canine teeth, for instance, this is only a verbal expression of a numerical idea. If such an observation can be reduced to concrete figures, the expression will usually be made more accurate and more succinct. Even if it cannot well be expressed except in words, the essentially numerical nature of the concept demands recognition and requires knowledge of the properties of numbers and of the ways in which they should be used and understood.

Numbers are of several different sorts, not always clearly distinguished, and each sort has its own meaning and properties. When it is said that a bird lays clutches of 4 eggs each, that its eggs are 4 cm. long, and that the bird was observed to leave its nest 4 times in one day, the number 4 is being used in three quite different and not interchangeable ways. These illustrate the three fundamental types of numerical data in zoology. In the first instance, the figure 4 is a count of discrete objects; it means that there were 4 such objects, no more and no less, in the unit of

observation (the clutch). Fractions and indeterminate numbers obviously do not exist in such observations. In saying that any given object is 4 cm. long, on the contrary, a measurement and not a count is given. It does not mean that the length is exactly 4 cm. or that the only adjacent possibilities are that it might be 3 or 5 cm. in length, but only that the length is nearer to 4 than to 3 or to 5, that it is greater than 3.5 and less than 4.5, without specifying just where in that range the absolute measurement lies. Such numbers may be fractional, in practice almost invariably are so if the observation is closely accurate, and often they are practically or really indeterminate. There is a continuous, infinite series of possibilities, and the figure given merely limits the observation to a particular part of this series. In the final instance the number 4 is again a count and again means exactly 4 with the only adjacent possibilities 3 or 5; yet it is essentially different from the first use of the figure. It is a count not of concrete, discrete objects but of the number of times that something occurred, a frequency.

In correct logical and mathematical procedure these three kinds of numbers are quite different sorts of data, and procedures and deductions proper with one may be entirely wrong in application to another. They are all means of giving particular values for things that vary in nature, and this is in the broadest sense the only and the whole reason for making any zoological observation. Zoology is entirely concerned with the study of things, of whatever sort, that vary in nature and that are in any way related to animal morphology and behavior. Such variables as cannot be reduced to numerical expression are attributes. Variables that can be expressed as counts or measurements, and hence in numbers, are given the special name "variate," a term that, in numerical zoology and in the pages of this book, has this special meaning, not being synonymous with "variable" but designating a special kind of variable. Since the number of times that a thing occurs is a numerical observation, whether the thing itself is an attribute or a variate, any sort of zoological variable can give rise to numerical data; and these observations are called "frequencies," also here used in this special sense rather than in its nearly synonymous but broader colloquial meaning.

As shown in the example of the bird and its eggs, variates include two different types of observations recorded by two

distinct kinds of numbers. Variates, like the number of eggs in a clutch, that can only take certain definite values, almost always integral, are called discontinuous variates because no matter how numerous the observations they never form a continuous series. An increase or decrease in value cannot be indefinitely or infinitely small but proceeds by a definite jump to a next higher or lower value. Variates, such as the length of an egg, that theoretically can take any of an infinite series of different values are called continuous variates because their numerical values express approximate position on a continuous scale. The difference between two observations may be indefinitely or infinitely small.

The principal sorts of primary numerical data available in zoological research are thus as follows:

1. Measurements of continuous variates, among others the following:
   - *a.* Linear dimensions.
   - *b.* Areas.
   - *c.* Volumes.
   - *d.* Weights.
   - *e.* Angles.
   - *f.* Temperatures.
   - *g.* Periods of time.
2. Counts of discontinuous variates, especially of elements serially arranged in animal structures or any separate but related or homologous structures, etc.
3. Counts of frequencies:
   - *a.* Of continuous variates.
   - *b.* Of discontinuous variates.
   - *c.* Of attributes, actions, or related phenomena.

DATA FROM DIRECT OBSERVATION

The raw data for the numerical analysis and synthesis of zoological materials must be derived from direct observation. In starting work, for instance, on an unstudied group of specimens these observations are in most cases lists of the specimens with simple measurements suspected of being significant and verbal notes of qualitative differences. As study progresses, some of these first observations will, in all probability, prove to be unimportant for the object in view and will be discarded, while new observations of the same sort but of different variates or attributes may prove to be desirable. When the work has progressed to the point of recognizing particular groupings, whether qualitative or

quantitative, it becomes possible to compile numerical values of a different category: frequencies, that is, counts of the numbers of observations belonging, in a given respect, to one of the categories recognized. This operation often derives numerical data from observations that are not numerical in their own character. Thus the presence or absence of a keel on a given tooth cusp would not primarily be expressed by a number; but if it appears that this has some significance for the work being done, it becomes subject to numerical analysis and statistical study when the number of specimens with the keel and the number without it are counted. Or, in taxonomic work, after all the specimens have been identified, the number of individuals in the collection belonging to each species gives numerical data involving biological conclusions not themselves of a numerical character. From the point of view of basing inferences of a higher order on the data and particularly of using statistics as a basis for such inferences, all of these direct numerical observations are primary observations or raw data, even though, as in the last example given, the possibility of making them comes only after the making of many secondary observations necessary in order to recognize the groups involved.

There is an almost unlimited variety of different sorts of primary numerical data possible under each of the broad categories of continuous variates, discontinuous variates, and frequencies, about as many sorts as there are different zoological problems to be solved. In the field of animal morphology and taxonomy the greater number of useful continuous variates are linear dimensions. Areas have some significance, for instance, the area of grinding teeth in mammals, important in considering food habits, or of the caudal fin in fishes, essential in studying their locomotion. Areas have, however, the serious disadvantage that they cannot be directly measured but must be calculated from linear dimensions or otherwise. This calculation is often difficult, may introduce errors or inaccuracies, and involves certain obscure peculiarities analogous to those of ratios, discussed on a later page. For these reasons it is usually preferable and possible to treat the same problems by the use of the more directly measurable dimensions from which the area would be calculated, its linear components. Volume is to even greater degree open to objection on the same grounds and if it must be calculated from linear dimensions should generally be used only if

the problem cannot well be attacked in any other way. It can, however, also be measured directly, as by displacement of liquids or filling cavities with a measured volume of fine shot or similar substances, and in such cases may be reliable and useful. Among mammals, cranial capacity is an important character properly recorded in this way.

In most cases, weight, rather than volume, is used in recording and comparing the bulk of animals and of their various organs. It is directly measurable to any desirable degree of accuracy and is usually at least as directly related to the zoological problem under attack as is volume. In problems of growth, it is the most important single factor; and it is also conveniently used in considering the relative strength of muscles, development of glands, etc. A few special problems of laboratory technique are involved, for instance, in recording weights of animals that ingest large quantities of heavy food or in weighing parts, like bones, in which the weight is greatly altered by methods of preservation and preparation. These can all be met by the simple but sometimes neglected principle of assuring that measurements to be compared were obtained under exactly comparable conditions. In paleontology, weight has, however, few valid uses, for measurable weights of fossils have no constant relationship to those of the living animal and are not reliably comparable with each other. Occasionally brain casts are compared by weight, but this is really an indirect comparison of volume; and the same is true of relative weights of plaster casts of bones, which do have an approximately constant relationship to the volume (but not to the weight) of the original bone.

Angles measure an important category of animal characters not measurable in any other way; and the numerical results, aside from the inconvenience of not being simple decimal numbers, are continuous variates subject to much the same sort of comparison and analysis as are linear dimensions. They record such biologically and taxonomically important characters as craniofacial flexion, limb angulation, or axial rotation of skeletal processes. The exact measurement of angles in zoological material is difficult but can usually be adequately achieved by methods of graphic projection.

Temperature typifies a class of physiological characters, to which basal metabolism, blood pressure, pulse rate, and numerous

others also belong, that are essentially continuous variates and may be treated as such mathematically.[1] They clearly can be related to taxonomy, although it is generally impractical to use them in that way, but they are principally involved in biological problems, where they are of the greatest importance. The measurement of periods of time delimited by some animal activity is also important in physiological, biological, and ecological studies, and these are also continuous variates. Pulse or respiration rate may also be expressed as the period between pulsations or respirations; periods of incubation and gestation are time-period variates; so are length of life, time of hibernation, or length of oestrous cycle; and there are numerous other essential time measurements involved in zoological research.

Discontinuous variates, not always recognized as such, are almost as abundant as continuous in zoological data. They are of major importance in taxonomy because they often have more limited individual and specific range and variability than do continuous variates and hence may characterize genera or higher groups. Their character and significance are more often obvious on inspection and without analysis, although this is not always true. Dental, vertebral, and phalangeal formulas often characterize super-specific categories and usually are of obvious significance. Cuspule or striation counts on mammal teeth, fin-ray counts on fishes, feather or egg counts for birds, blood-corpuscle counts for any vertebrate, and many others are discontinuous variates commonly highly variable and demanding some formal analysis for their successful interpretation. Any serial or repetitive structures are discontinuous variates whatever the scope of the taxonomic or other category within which they vary; and all may, if desirable, be treated as such statistically by methods discussed on later pages.

Frequencies, which are fundamental in the most useful statistical procedures, are expressed or implied in any collection of numerical data. They are simply counts of individuals belonging to any selected category. The categories may be based on any measurements or counts of variates, either as observed or as

[1] Pulse rate would appear at first sight to be a discontinuous variate, since discrete pulsations are counted; but in fact it is a continuous variate. The number of pulsations per minute accurately involves a fraction and can theoretically take any value in a continuous series.

gathered secondarily into groups. The categories may, furthermore, be based on any logical consideration, even one wholly non-numerical or fundamentally subjective. Thus frequencies may be based on simple attributes, such as the presence or absence of a vestigial tooth (which may, however, also be considered a discontinuous variate) or differences in geological or geographical origin. They may be counts of the individuals of each species in a certain collection, counts of the number of known species in each of several genera, counts of the species of a given fauna grouped by their probable habits of life, etc., each of these and the innumerable other possibilities having a definite bearing on some type of zoological research. All observations involve frequencies, even if the frequency be 1 (or 0, the characteristic sought not being found in any case); and in many cases these frequencies are at least as essential to consideration of the problem in hand as are other types of data.

Since a continuous variate may theoretically take any of an infinite series of values, it follows that absolutely accurate measurements of any two values of such a variate would never be the same and consequently that the frequency of any one value would always be one and the concept of frequency useless. In fact it has been pointed out that such absolutely accurate measurements are not possible (or desirable) and that the measured and recorded value of the continuous variate is in practice only a conventional means of defining a greater or smaller span on the continuous scale within which the real or absolute value is known to lie. Thus the record 3 mm. means that the true value is known to be greater than 2.5 and less than 3.5 mm., 3.1 mm. means that the true dimension is greater than 3.05 and less than 3.15 mm., etc.[1] Thus the record 3.1 mm. can include various different exact values of a continuous variate between 3.05 and 3.15 mm., and recorded values of continuous variates can and do have frequencies greater than 1 in practice. The groups of values thus brought together can be made larger or smaller at will, and a similar sort of grouping may be applied to discontinuous variates, so that the frequencies can be manipulated into the form most advantageous for the problem in hand, a subject discussed in detail in Chap. III.

[1] In practice, in approximating measurements by dropping the last digit, it is customary to count 3.05 to 3.14 as 3.1, 3.15 to 3.24 as 3.2, etc.

RATIOS AND INDICES

Ratios, products, indices, and other numbers obtained by the combination in various ways of two or more numbers are themselves raw numerical data from a statistical point of view; but they are secondary, not derived from direct observation, and they have properties unlike those of numbers obtained by direct observation. Of these, the most important are ratios, which give by a single number or expression the relative sizes of two other numbers. A valid ratio is the quotient of two numbers which must express observations of the same sort, for instance, linear dimensions, and must be in the same units, for instance, millimeters. The resulting ratio is independent of the absolute size of the original figures; for instance, 5 : 10 is the same ratio as 500 : 1,000 and is also independent of the original units of measurement; for instance, 5:10 mm. is the same ratio as 5:10 years. The result, ordinarily expressed as .5 for all these examples, is a pure number divorced from any particular system of mensuration.

Ratios of two continuous variates are in proper and widespread use in zoology, and they express characters that are of fundamental importance. They have, however, certain peculiar and generally ignored properties that must be kept in mind and that may in some cases make conclusions based on them inaccurate or even invalid. They are themselves continuous variates, and the numbers in which they are written are of the indefinite kind that express approximate position in a continuous series; but the accuracy and limits implied are not the same as for the direct measurements on which the ratios are based.

Ratios frequently vary more than do the dimensions on which they are based. Thus if the lengths of a given sample of homologous teeth vary from, say 0.9 to 1.1 mm. and the widths also vary from 0.9 to 1.1. mm., the possible length: width ratios vary from 0.8 to 1.2, a markedly greater range. The relative variabilities of ratios and of their constituent dimensions are tied up in an intricate way with the correlation between the latter (see Chap. XII).

The most confusing characteristic of ratios is that they are grouped in a peculiar way not determinable by simple inspection of the figures and that this may be a source of error in basing deductions on them. A length recorded as 1.0 mm. is known to be

somewhere between .95 and 1.05 on the continuous scale, a simple and obvious relationship, but this is not true of a ratio recorded as 1.0. For instance, a length:width ratio of 1.1:1.1 mm. would be recorded as 1.0, but its real value may be anywhere between .92 and 1.09, or in round figures from .9 to 1.1.[1] Furthermore this peculiarity may result in writing two really different ratios as the same or two really identical ratios as different. It has been shown that the ratio 1.1:1.1 may really be anywhere from approximately .9 to 1.1. Similarly the ratio 1.0:1.1 may really lie anywhere from .8 to 1.0, a range widely overlapping that of the other and apparently different ratio. These possible limits of the true values of ratios are never obvious from inspection of the ratio, and ratios recorded as identical may in fact have quite different limits. Thus the real value of the ratio 9:9 is somewhere in the range .90–1.11 and that of 1.9:1.9 somewhere in the range .95–1.05—a considerable difference in accuracy, but written as a single figure, according to usual practice, these ratios are given as identical.

These difficulties are far outweighed by the usefulness of ratios; but they must be understood, and it should not be supposed that a figure representing a ratio is necessarily as accurate as those on which it was based or that it is on the same standing. If minute differences are important and the status of ratios doubtful, it may occasionally be advisable to abandon these and deal with the problem directly from the original measurements.

Ratios may also be usefully based on discontinuous variates and on frequencies. The ratio of dorsal to lumbar vertebral counts, for instance, may express an important character in the clearest way, or, as another example, the ratio of number of individuals (frequency) with skulls longer than a selected standard to that of those with skulls shorter than the standard may be a valuable means of characterizing the group as a whole. Ratios based on such data are themselves discontinuous variates. They do not have the disadvantages of ratios that are based on continuous variates, but they have an extraordinary peculiarity of their own: although discontinuous they are usually fractional and sometimes indeterminate. An example will make this clear. Suppose that

[1] Because each of the component dimensions, as far as shown by the record of them, may be anywhere from 1.05 to 1.14. If the real length is 1.05 and the real width 1.14, the real ratio is .92; and in the converse case the real ratio is 1.09.

each of two discontinuous variates can take the value 1, 2, 3, or 4. Ratios between these two can take the values shown in Example 1.

EXAMPLE 1.—RATIOS BETWEEN TWO DISCONTINUOUS VARIATES, EACH WITH VALUES OF 1 TO 4

(Hypothetical data)

1:4	= 0.25
1:3	= 0.333 . . .
1:2, 2:4	= 0.5
2:3	= 0.666 . . .
3:4	= 0.75
1:1, 2:2, 3:3, 4:4	= 1
4:3	= 1.333 . . .
3:2	= 1.5
2:1, 4:2	= 2
3:1	= 3
4:1	= 4

This series of 11 possible values is irregular and follows no obvious system; 7 of the values are fractional, and 3 are infinite repeating decimals. Nevertheless they are the possible values of a discontinuous variate. Each value is definite and exact, not an approximation or group symbol as for a continuous variate; and under the postulated conditions they are the only values that the variate can take, intermediates between them being impossible. It is also noteworthy that more combinations of the original dimensions result in a ratio of 1 than any other figure, a peculiarity that also may strongly affect conclusions based on such ratios.

The word "index" is used in somewhat different technical senses in statistics and in morphological zoology. In the latter, unlike statistics, an index is obtained directly from individual measurements and not from statistical or group data. The uses of the words "ratio" and "index" in comparative morphology are not strictly standardized. "Index" is generally used for a figure obtained by dividing a given dimension by some particular larger dimension of the same anatomical element and multiplying by 100 (or expressing as a percentage). Unless the dimensions are otherwise specified, it is generally understood that they are the minimum and maximum dimensions of the anatomical unit. "Ratio" is a more general word, which may include indices, but usually refers to proportions between dimensions of different anatomical elements.[1]

[1] Hue (1907) defines "index" as always based on these dimensions, but this seems an unnecessary limitation. Troxell (1915) limits "ratios" to comparisons of homologous dimensions of different individuals and calls all

The useful broad categories of ratios based on dimensions appear to be as follows:

1. The ratio (or index) of two different dimensions of one anatomical element of one individual.
2. The ratio of two analogous dimensions, one for each of two anatomical elements of one individual.
3. The ratio of homologous dimensions of homologous anatomical elements of two individuals.

Theoretically there are four other possibilities:

4. The ratio of nonanalogous dimensions (*e.g.*, length and width) of different elements of one individual.
5. The ratio of nonhomologous measurements of homologous elements of two individuals.
6. The ratio of analogous dimensions of nonhomologous elements of two individuals.
7. The ratio of nonanalogous dimensions of nonhomologous elements of two individuals.

But these last four ratios are of little or no practical value. They make comparisons of things that usually are not related to each other in some simple, orderly, and significant way, or, differently expressed, they obscure significant dependence by introducing variables too numerous or essentially independent.

The three useful categories of ratios of dimensions express different sorts of characters or concepts, and the inferences based on them are of different kinds. Indices or ratios under (1) are essentially unit characters not markedly unlike linear dimensions in the concept involved. The index (breadth $\times$ 100)/length of a given tooth is a simple character for that tooth, as are its breadth and its length individually. Such indices are sometimes designated by their supposed or actual correlation with some other function or character; *e.g.*, the index (length $\times$ 100)/breadth of a limb bone has been called the "speed index" because it is advanced as a hypothesis or supported as a theory that the larger the value of this index the more rapid, in general, the locomotion of the animal. Even aside from the fact that this is not a constant relationship (and even that the exact opposite can be demonstrated to be true in some cases), this naming of a ratio by the inference that is expected to be drawn from it is unsound.

proportions of one individual "indices." Although there are many such variant usages, the exact meaning is usually clear from the context.

The conclusions that may be drawn from numerical data should not be confused with the data themselves.

Ratios under (2) of the above list express a different sort of character, for they are descriptive of a larger anatomical unit than that measured by either of the primary figures from which the ratio is derived. Thus if teeth are used as examples again, the ratio length of trigonid/length of talonid belongs to this category, and it expresses numerically a character of the tooth as a whole, whereas neither of the direct measurements applies to the whole tooth. Similarly length of humerus/length of radius is a character of the forelimb, and length of humerus/length of femur of the locomotive apparatus as a whole. By "analogous dimensions" is meant length against length, etc. There may well be some relationship between the length of one element and, say, the width of another; but this is a somewhat confusing concept and one of little practical use.

Ratios listed under (3) are far the most common in zoological work and in some form or other are almost universally employed. Thus the statement that one species is larger than another is merely a crude expression of a ratio of this sort. A statement, such as is frequently made, that one species is 20 per cent larger than another is, however, a gross misstatement of the actual facts. What it usually means is that some given dimension of one specimen of one species is 20 per cent larger than the same dimension of one specimen of another species. That all the dimensions of all the individuals of one species should be 20 per cent larger than the corresponding dimensions of all the individuals of another species is impossible; and it is preferable to say what is really meant. This exemplifies the usefulness of defining species, whenever possible, by the statistical constants of their several variates, rather than by individual values of these variates, and of always specifying the particular variate involved.[1]

[1] Troxell (1915) confines the word "ratio" to ratios of our class (3) and makes the astonishing statement (p. 615) that "apparently nowhere in the literature has such an application been made of 'ratios,' comparing one type to another." It is fairly obvious that such ratios have always been used, in one form or another, in almost every description of a species. However, it appears from the context that what is particularly meant is more specific: the use of ratios between the dimensions of various fossils and those of some one more complete specimen (particularly a recent skeleton) and the comparison of different individual fossils by means of their ratios to such a

There is a higher, derived category of ratios the formal recognition of which is infrequent but which are often implied and which may be useful, that is, the ratio of two ratios. Thus the ratio of the cephalic index of one specimen to that of another is a ratio of two ratios which can be written in this way:

$$\frac{\text{Breadth of skull } A \times 100}{\text{Length of skull } A} : \frac{\text{breadth of skull } B \times 100}{\text{length of skull } B}$$

This is a means of comparison as logical as the ratio of linear dimensions, for instance,

$$\frac{\text{Breadth of skull } A}{\text{Breadth of skull } B}$$

but the ratio of two ratios suffers in an exaggerated degree from the peculiar and disadvantageous properties of ratios in general and should be used with the greatest caution.

Ratios and indices may be expressed numerically in several different ways:

1. As the unreduced ratio of the actual measurements, *e.g.*, 5:10 mm., or 5:10.
2. As a fraction, *e.g.*, $\frac{5}{10}$ or $\frac{1}{2}$.
3. As a quotient, *e.g.*, 0.5.
4. As a percentage, *e.g.*, 50 per cent.
5. As a quotient multiplied by a constant, *e.g.* (using 100 as the constant, the usual form), 50.

For purposes of inference or of analysis, ratios are still raw data. Their only essential difference from the numbers on which they are based is that they express a different sort of character.

---

complete specimen. The application of this principle in certain specific circumstances is useful and already generally understood. Thus if it is important to compare two incomplete specimens that have no parts in common, this obviously can only be done through their comparison with a third individual, as nearly similar to them as possible, in which parts also present in each of them are known. The results are frequently unreliable and in such cases the method is used only *faute de mieux*. As a general principle of research it is undesirable to compare two things not directly with each other but each with a third thing extraneous to the particular problem being studied. It is of course quite sound when the third thing, or standard of comparison, is not actually extraneous to the problem, and it is necessary when a comparison is essential but cannot be made more directly; but it is never so reliable as direct comparison.

In general the further study of ratios follows the same lines as for any other raw numerical data. As with any other data, however, their morphological meaning and arithmetical derivation must be kept clearly in mind. For instance, in considering variation and variability, the fact that linear dimensions may or do vary within related groups relative to the mean of each group does not warrant the assumption, a priori, that ratios based on these dimensions will vary in the same or a similar way, since the ratios are not dimensions but are pure numbers derived from but independent of the mean dimensions (see also Coefficient of Variation, Chap. VI).

In general, it seems probable that biological characters and relationships of the sort involved in ratios are adequately and most clearly expressed by the mean value of the ratio and some form of correlation of the measurements involved in the ratio (rather than in terms of the distribution of the ratio itself; see Chap. XII).[1]

There are several other types of calculated but essentially raw data that are analogous to ratios in expressing by one number some relationship between two measurements but that involve distinct concepts and operations. Few of these are in use in zoology, and we know of none likely to be of really general value; but some may be useful in certain special problems. Such a figure is, for instance, (length + width)/2 (sometimes called a "module"). This may be a useful concept in cases where there is no marked functional difference between the two dimensions and they tend to vary about the same or approximated means. In cases where one dimension tends to increase as the other decreases,[2] this module will generally vary less than does either

[1] The mean of a ratio is most easily calculated by dividing the mean of one set of measurements by the mean of the other, but note that this is not proper unless all the measurements are paired. That is, if, for instance, a mean length:width ratio is sought, it may be obtained by dividing the mean length by the mean width if each length is accompanied by the corresponding width of the same specimen and each width by the corresponding length. This is not always true of paleontological data; and if unpaired measurements have been included in the means of the two primary measurements, the ratio of each individual must be taken and the mean of the ratios then calculated. The difference is seldom significant, but this procedure is technically the correct one.

[2] Negative correlation or inverse covariation.

dimension (whereas the ratio will vary more than either) and may be useful on that account. Measures of area, length × width, are in a sense analogous and have the same property of tending to be less variable than the length or width if these two have an inverse relation to each other. There are numerous cases in nature (*e.g.*, the surfaces of grinding teeth) where the functional character is the module or area rather than the linear dimensions; and it would be interesting to examine a number of these and to see whether they are not better described in terms of a module or similar figure than in terms of the original linear measurements or of area.

Various limb modules, such as

$$\frac{\text{Length humerus} + \text{length radius}}{2}$$

or

$$\frac{\text{Length tarsus} + \text{length metatarsus}}{2},$$

are also logical concepts that may serve to bring out relationships not immediately visible from the original measurements; and many similar formulas will suggest themselves in the course of special investigations.

From the standpoint of any particular problem, the purpose is to find a figure, all the elements of which are related to the problem, which has some property not possessed by the primary elements, such as being more constant than are they or varying in some definite and ascertainable way with respect to a different variate, to function, etc. Other general possibilities to bear in mind aside from ratios (quotients), modules (arithmetic means), areas (and analogous cases, products), and deviations (remainders of subtraction) are geometric means (roots of products), as well as many secondary or tertiary figures such as powers of ratios or of deviations, quotients of modules. Such figures should only appear in the final work, however, if they did really prove to express characters or have useful properties other than those of the original measurements.

CHAPTER II

MENSURATION

REQUIREMENTS OF GOOD MEASUREMENT

An infinite number of numerical observations may be made on any one zoological specimen, and each may be made in many different ways. The first approach to a problem in this field is decision as to what is to be measured and how. The most important criteria of good numerical observations are that they should be:

1. Of a logical unit.
2. Related to a definite problem.
3. Significant.
4. Adequate.
5. Well delimited.
6. Standardized or well specified.
7. Accurate.
8. As refined as desirable.
9. Unbiased and consistent.

Paleontologists seem to use illogical and nonunit measurements more often than do neozoologists, for instance, such a measurement as length from the second premolar through the first molar in a mammal. This has no natural unity, measures no biologically important single character, and is poor for comparisons (the only purpose of taking it) because the measurement is not likely to be available in the literature for other specimens and on many specimens otherwise comparable it may be impossible to make. Measurements of each single tooth should be given, and measurements of groups of teeth should be of natural groups—of the whole cheek series, of all the premolars, or of all the molars.

A general principle of measurement, involved in several of the criteria listed above and violated in the example of illogical measurement just given, is that those measurements are usually best that permit the greatest number of valid comparisons. In paleontology such violations of the principle as that of the

example are generally caused by incompleteness of the material. It is possible to measure only what is preserved, but this is hardly worth while at all unless natural units can be measured. Such an odd and relatively useless dimension as length P^2-M^1 is probably given, instead of measurements of individual teeth, on the premise that the percentage of error will be less for a large measurement than for a small one. This argument, however, merely shows that the technique used should be one producing accuracy at the desirable degree of refinement whatever the size of the measurement. In fact, in paleontology, the premise here is often fallacious, for a longer measurement is more likely to be affected by distortion than a shorter, so that its accuracy as an estimate of what the dimension was in the living animal may be as low as for a smaller dimension or even considerably lower. This is particularly true in dealing with teeth or similar series in which the individual elements are usually little distorted but the series as a whole is frequently seriously distorted.

As regards relationship to a problem in hand, this requirement is so obvious and so rarely transgressed as to require no emphasis except to point out that the relationship should be as direct and as simple as possible and that the problems of other workers should also be kept in mind to some extent. Brain growth, for instance, can be studied from the skull dimensions or endocranial capacity, and in some cases must be because other data are unobtainable; but neither is related directly and simply to the question studied, and the best measurement is weight or volume of the brain itself. It is, however, pertinent to give measurements that will be useful to others working on related problems even though they may not be necessary for the purposes of the immediate enquiry. In taxonomy many standardized dimensions may be quite unnecessary to define a species or subspecies and yet should be included as a regular practice to facilitate future work.

It is always better in assembling raw data to take too many measurements than too few. At this stage in research it is commonly inadvisable to adhere too rigidly to a criterion of direct relationship and preferable to measure most variates with any conceivable bearing on a problem, for in this way important and unsuspected relationships are often discovered. Such data require in any case careful analysis. Certain of them will probably turn out not to be significant or necessary to demonstrate the

point at issue. In this case (except for standard dimensions that will surely be useful to others or in the future), they should be discarded, no matter how much work has been involved in obtaining and evaluating them. Zoological literature is replete with long tables of measurements that prove nothing and the publication of which was unnecessary, expensive, and really a discourtesy to other students. In this respect the methods, largely statistical, discussed in succeeding chapters are invaluable. They provide definite tests as to whether measurements really are significant, facilitating the selection of essential and rejection of nonessential data; and they also assist in reducing raw data to the most compact and most useful form.

Equally common and perhaps still more open to criticism is the gathering and publication of inadequate numerical data. In discussing a species from a taxonomic point of view, it is usually unnecessary to give all the pertinent dimensions for each of a large series of specimens; but at least this does make the data available and is preferable to the practice of using only the dimensions of the type or of giving only the mean dimensions of the whole series. Many studies that purport to deal with variation give only the maximum and minimum dimensions observed, or sometimes these and the mean. Occasionally the number of specimens involved is also given, but the frequency of omission of this absolutely essential datum is remarkable. For a real study of variation and indeed for most purposes of valid comparison, such data, even if the means and the number of specimens are recorded, are a little better than nothing, but not much better. Far more important are data on the way in which the observations were distributed about the mean, on the probable relationship of the observed extremes and mean to those of the whole population from which the sample was drawn, on the probable significance of differences in ranges and means, and similar questions. Measurements and other observations are inadequate if they do not permit the calculation of such data, and the publication of results is inadequate if such data are not obtained and recorded.

Some measurements are useless or nearly so because they have no well-defined limits and hence cannot approach an adequate standard of accuracy and refinement. For instance, an attempt has been made to use the distance of the narrowest point on a slender limb bone from the proximal end of that bone as a

numerical character of animals. The sides of the limb bone being nearly parallel, its narrowest point is so vaguely defined that any reasonable degree of accuracy is impossible and the character, although real, is generally useless because it is not well delimited.

To be comparable, measurements must be taken in the same way; and to be comprehensible, they must be exactly specified. The first of these requirements is largely mechanical and depends on adequacy of equipment, practice, and experimentation to produce sufficiently consistent results. Absolute consistency is impossible, but assurance is necessary that it is approached closely enough not to affect the results derived from data. Specification demands mention not only of exactly what measurement was taken but also of exactly how it was taken, unless both are certainly obvious or understood by the readers addressed. In taxonomic work on fishes, which is largely based on proportions (*i.e.*, ratios) because few ichthyologists have learned how to use linear dimensions properly with their sort of materials, proportions are usually obtained by using dividers which are set at the smaller dimension, the integers of the proportional value of the larger dimension stepped out with the dividers, and the fractional excess estimated by eye. Such grossly inaccurate methods are not to be condemned on that score alone if the low accuracy is really adequate for the purposes intended;[1] but clearly they are not comparable with more refined methods, and their use should be specified. Similarly mammalogists usually measure the longer dimensions with a ruler and the shorter with calipers; but some use ruler and some calipers for both, some use simple dividers read against a ruler, and some use other methods, such as measuring with the short end of proportional dividers and reading the long end against a ruler. The refinement of each method is different and may need specification, even though in this case all the methods mentioned are sufficiently refined for the usual purposes.

The condition of the material and the way in which it is held for measurement also affect accuracy and comparability and may require specification. Living animals, dead unprepared animals, animals in preservatives, and skins all differ to some extent in dimensions; and the different preservatives and methods of

[1] But some of the results suggest that they are frequently not sufficiently accurate as used.

preparation may also have effects so different as to render measurements incomparable. Sumner (1927) has shown, for instance, that the mean total length of 10 mice (*Peromyscus maniculatus gambelii*) was 166.65 mm. at time of death and 164.10 two hours later.[1] Differences between freshly killed animals and skins as customarily preserved in collections are usually still greater. Measurements of one specimen held free, one lying flat, and one stretched out may also differ considerably for some types of material, especially live or freshly killed.

Specification of the thing measured is equally important, and current practices are still more varied and confusing. Checking over some recent literature, the dimension given simply as "length" for mammal teeth was found to have been applied in at least six different ways:

1. Greatest distance between planes tangential to the margin of the crown and at right angles to the longitudinal skull axis.
2. Distance between planes tangential to the crown margin, parallel to each other, and approximately parallel to the anterior and posterior edges of the crown.
3. Greatest horizontal distance along the outer or inner face of a tooth (*e.g.*, along the ectoloph).
4. Distance from anterior to posterior borders along the midline of a tooth.
5. Greatest diameter of the tooth crown (sometimes longitudinal, sometimes transverse, sometimes vertical, and generally somewhat oblique).
6. Distance from tip of crown to tip of root.

Probably other usages are also current. Obviously "length of tooth" is a meaningless designation unless some further specification is made or distinctly understood.

It is, however, most usual for a dimension taken to be the maximum distance between parallel planes tangential to the designated anatomical element. For length, the planes are usually considered to be oriented vertically to the axis of the body through the axial anatomical divisions and their parts—teeth, skull, vertebrae, etc.—and vertically to the proximodistal axis for nonaxial elements—ribs, limbs, etc. Width is the dimension at right angles to the length and most nearly in a horizontal plane, and depth or height the dimension at right angles to these two and nearly in a vertical plane. These definitions apparently conform to a consensus at present and, although not recognized

[1] The difference is probably real and significant although Sumner does not give the data by which this can be evaluated.

rules, might well be made so. In some groups, specialists understand other conventional designations without specification; but in general any departure from these general definitions should be specified.

SYSTEMS OF MENSURATION

Experience in ascertaining the most useful measurements, the irksomeness of fully specifying a dimension each time it is used, and the need to make the work of different observers as comparable as possible have led to some standard systems of mensuration more or less generally used within the various zoological groups and for various types of zoological problems. There is not and cannot be a single standardized system for zoology in general. Even the vertebrates differ too much in structure, their dimensions differ too much in significance, and the variety of problems that arise is too great for such an end to be practicable or desirable. Systems already in use are so numerous that they cannot be usefully summarized but only a few examples briefly mentioned in a general work like this; and the student must become familiar with those employed in his own field through its special literature and then adopt them or replace them by one specifically suited to his own problems.

Measurement of dimensions of animals is most suited for reduction to a standard system, supplemented in most instances by some counts of discontinuous variates. In most cases the principal purpose of such a system is taxonomic, and it usually concentrates on external characters.

For fishes, few standardized linear dimensions seem to be generally recognized, and the customary numerical data are proportions, usually the total length of the fish divided by a lesser dimension such as length of head or depth of body, with counts of scales (usually along, above, and below the lateral line), of spines, and of fin rays. Such a system is briefly explained by Nichols (1918), and variations are exemplified in all systematic monographs on fishes. Relatively little work seems to have been done on the careful mensuration of fish skulls or other internal structures, but an excellent system has been proposed by Gregory (1933).

Among lizards and snakes few linear dimensions except total length and tail length are commonly used in taxonomic work,

which is mainly based on discontinuous variates such as tooth counts, scale counts on rather elaborate systems, and counts of elements in repetitive color patterns. An excellent exemplification of such a system is given by Blanchard (1921). For turtles the simple linear dimensions of carapace, plastron, tail, etc., are the usual numerical data. Kälin (1933) has given a complicated system for numerical study of the crocodile skull involving numerous linear dimensions and 12 indices.

The measurement of birds for taxonomic purposes is more nearly standardized than for most lower groups, Ridgway's system usually being employed. Because it is so widely accepted and as an example of such a system, its standard measurements, all linear dimensions, are listed (from Ridgway 1901):

Length.—From tip of bill to tip of tail. (This may differ greatly in recently killed birds and prepared skins and may also be difficult to measure accurately.)

Wing.—From the anterior side of the carpal bend to the tip of the longest primary (feather).

Tail.—From between the shafts of the middle pair of rectrices at the base, pressed as far forward as possible without splitting the skin, to the extremity of the longest rectrix.

Culmen.—From the tip of the bill to the edge of the feathers on the dorsal side. (This is sometimes called "bill" if it extends to the true base of the bill and "exposed culmen" if the base is partly covered by feathers.)

Depth of Bill at Base.—From the lower edge of the mandibular rami to the highest portion of the culmen.

Width of Bill at Base.—Across the chin between the outside of the gnathidea at their base.

Tarsus.—From the tibiotarsal joint on the outer side to distal end of the tarsus.

Middle Toe.—From the distal end of the tarsus to the base of the claw, not including the claw unless so stated.

Graduation of Tail.—From the end of the outermost rectrix to that of the middle or longest, the tail being closed.

As proposed, the length was to be taken with tape or ruler, the other measurements with dividers (then read against a ruler). As with all such systems, the whole series of measurements is not invariably made (frequently only wing, tail, and culmen); and in some groups other measurements may be needed. Except for the total length, these dimensions are nearly the same on skins as on the living birds.

Among mammals the standard external measurements (see Anthony 1925 or Sumner 1927) are:

Total Length.—Tip of nose to tip of tail, laid out as nearly as possible in a straight line. In this and all other measurements the limits are on the skin, with hair excluded.

Tail Length.—From tip of tail to point where tail breaks from body when held at right angles to the latter. This is usually designated "length of tail vertebrae"; but the designation is false, for it often excludes the most proximal vertebrae.

Body Length.—This is not measured but is the total length minus the tail length.

Foot Length.—From the posterior edge of the heel (*i.e.*, the skin over the extremity of the calcaneum) to the end of the longest toe, including the claw, stretched out as straight as possible. Sumner proposes always using the left foot unless it is injured.

Ear Length.—From the tip of the ear either on the medial side to the junction with the crown of the head, when it is specified as "from crown," or on the lateral side to the bottom of the notch, specified as "from notch." Sumner also proposes using the left ear as far as possible.

Weight.—This is hardly a standard measurement, for it is still usually neglected in making systematic collections; but it is recognized as important, and efforts are being made to procure its more frequent record.

It is customary to take the longer dimensions with a rule and the shorter with calipers or dividers. Except for foot length, all these measurements may be significantly different in the living animal and in prepared specimens, so that they are generally taken on the freshly killed animal and any deviation from this practice must be specified.

Taxonomists tend for practical reasons to concentrate on external characters like those given above for birds and for mammals, especially when they are interested in smaller groups, such as species and subspecies. These characters are superficial, both literally and figuratively, and so are not very reliable for the taxonomy of higher groups. They are usually not available in fossils, which, with unimportant exceptions, can be studied only by osteology and dentition. Their comparison with living animals also requires study of the hard parts in the latter. Numerical and other characters derived from the teeth and skeleton are of great value and are widely used in mammalogy, both recent and fossil. As regards the skeleton, they are of equal value among the lower vertebrates but have as yet been less used for recent animals.

Paleontological mensuration differs little from that of the hard parts of recent animals. Fossil material is almost invariably less complete, so that a standardized system of a few measure-

ments is less practical and requirements must in each case be adjusted to possibilities. Fossil bones are also commonly distorted so that their measurements are generally less reliable than are those of recent animals; this may make some measurements, especially those of proportions, unusable. Some groups of extinct animals are so unlike any living forms that they present a different problem in mensuration. All of these factors also militate against systematization of paleontological data; but they do not make it impossible, and there is room for great improvement in this respect. In much paleontological publication, aside from a few obvious measurements, the numerical data either are inadequate or seem to have been selected at random and without rational criteria.

Perhaps the most detailed system of osteological mensuration for mammals is that of Duerst (1926), who also gives references to and synonymy with the practices of other workers. Osborn's elaborate studies on the osteometry and craniometry of perissodactyls (especially 1912 and 1929), although based on a single group of mammals, also repay close study by anyone engaged in gathering numerical zoological data. Within the limitless field of special problems, only two strikingly different and suggestive examples will be mentioned. Zeuner (1934) has used a system of cranial angles as a basis for biological inferences regarding rhinoceroses, and Soergel (1925) has employed numerical and mathematical procedures in studying footprints and inferring from them the sort of animal that made them.

Aside from dimensions and counts like those mentioned above, color is a very important character in the study of recent animals. Usually this is roughly described in the vernacular; or an attempt, much better but still inexact, is made to match the color against a standard chart, of which Ridgway's (1912) is the most widely used. The most precise method would be by photometric spectroscopic analysis, but this is such an elaborate and exacting process that it is impractical in most zoological work. Numerical data on color can, however, be obtained more simply with a color top or a tint photometer. In using a color top (see Collins 1923), segments of white, black, and a set of standard colors, usually complementary and primary, are adjusted to match the color to be measured when they are spun so as to blend into a single color. Adjustment of the segments by trial and error is a long process,

and the matching is subjective and does not give very consistent results. In a tint photometer (see Sumner 1927), reflected light from a white surface and from the colored object to be measured are viewed simultaneously through a color filter, and the light from the white surface is cut down by a diaphragm until it matches in intensity that from the object, giving a relative measure of the amount of light, of the wave lengths passed by the screen, that is reflected by the object. The percentage of closure of the diaphragm is read from a scale and recorded numerically. If several screens are used and a reading taken for each, a good numerical measure of color can thus be obtained. The procedure is reasonably rapid and simple, and the estimate of relative intensity of light is easier and involves less subjective inconsistency than the matching of colors. This method also has drawbacks, especially its requirement of a complex apparatus and the fact that it does not measure the whole color but only certain components in it (the color bands passed by the filters). Without the use of an impracticably large number of filters, the color could not be reproduced exactly from data gathered in this way. This is, however, the most practical valid method of reducing color to exact numerical terms that has yet been devised.

ACCURACY AND REFINEMENT

Since a continuous variate may take any of an infinite series of values, its exact measurement would require an infinite number of decimal places[1] and is impossible. Such a measurement is therefore never an exact value but always a conventional representation of a range within which the exact value is believed to occur. If the exact value does in fact occur within the indicated range, the measurement is accurate, regardless of how many or how few places are recorded. The smaller the indicated range (the more places accurately recorded), the more refined the measurement. Thus if an exact value of a given variate is 2.30749 . . . (plus an infinite series of smaller decimals), measurements and records, as 2, 2.3, 2.31, 2.307, or 2.3075, are all accurate since each designates a range within which the exact value does in fact lie, the limits implied being respectively 1.5–2.5, 2.25–2.35, 2.305–2.315, 2.3065–2.3075, and 2.30745–2.30755. The implied ranges are

[1] This would be true even if the exact value were an integer, for this could be represented only by an infinite series of zeros after the decimal point.

thus respectively 1, .1, .01, .001, and .0001; and the measurements are progressively more refined because the ranges within which the exact value is shown to lie are progressively smaller.

In practice a point of refinement is eventually reached where accuracy is no longer possible. Thus in the preceding example the material and methods might well be such that the third decimal would not always be measured as .007 but might be measured within a range of error of .005–.009, the whole measurement recorded as 2.305, 2.306, 2.307, 2.308, or 2.309, of which only the middle value is accurate since the true value does not, in fact, lie within the ranges implied by the others. This range of error has nothing to do with the range implied by the recorded figures, which are accurate only so far as they can be carried out without any range of error appearing.

With some exceptions, which are discussed later, refinement should not be carried beyond the limit of accuracy. It is not, however, true that refinement should always be carried to that limit. It frequently happens, for instance, that accurate measurements can be made to two decimal places but that the results obtained from the actual use of these measurements would have been the same if they had been carried only to one place. In such a case it is a waste of labor and space to make or publish the more refined measurements. Two factors, accuracy of measurement and refinement desirable for the purpose in mind, are thus involved in the ideal of making measurements neither more nor less refined than necessary and of making them accurate to that optimum degree of refinement.

Accuracy of measurement depends on the nature of the material, equipment used in measuring, and personal factors such as bias and consistency (reliability) discussed in the following section. The degree of accuracy obtained under given circumstances can be determined experimentally. As an example, one of the present authors made a measurement of tooth length such as is routine work in vertebrate paleontology, repeating the measurement independently on six different days, using a low-power binocular microscope and a caliper calibrated to .1 mm. The results were as follows:

13.0	13.1
13.3	13.0
13.2	12.9

Expressed in integral millimeters, these measurements are all 13, while in tenths of a millimeter they range from 12.9 to 13.3, averaging 13.1. From the distribution of these measurements and other criteria extraneous here, it is certain that the exact value was somewhere in the range 13.0–13.2. The measurements are thus all accurate to two figures (13); for that implied range (12.5–13.5) certainly includes the true value. They are not accurate to three figures (one decimal place); for no single one of these more refined figures certainly includes the true value, and two of them certainly do not (12.9, 13.3). This is nevertheless a case in which records to three figures, one inaccurate, are preferable to the accurate two-figure measurements.[1] All the three-place figures, even the most divergent, are closer to the exact value than are the limits 12.5–13.5, implied by the two-place figure 13. As a general criterion, inaccurate figures are useful if their range of error is less than the implied range of the accurate figures available, in other words, if the range of error for any one place is less than 10 units, as in the example the range in the first decimal place is .4, less than 1.0.

The smallest of the six measurements made in this experiment is certainly 98 per cent or more of the exact value and the largest 102 per cent or less. It is thus certain that any one measurement was within 2 per cent of the real value of the dimension measured. Supposing, as other experiments show to be highly probable, that this represents the degree of accuracy generally obtainable with such equipment and with little or no personal bias,[2] it is possible to work out a schedule for measuring that will assure gathering all the useful data and none that is so inaccurate as to be useless. With a probable range of error of 4 per cent, as in this experiment, such a schedule would be:

[1] It might, of course, be preferable to have the third figure also accurate if it is to be recorded; but the example is chosen to illustrate the fact that one inaccurate figure may sometimes be legitimately recorded and that valid results may be based on such figures. The reasons for this phenomenon and the circumstances under which advantage may be taken of it are discussed in Chap. VIII. With much paleontological material, as in this example, it is a practical impossibility to achieve complete three-figure accuracy; yet three-figure records may be better than two.

[2] In this case the same experiment was conducted independently with two other observers equally accustomed to using the same equipment on similar material and gave substantially the same results.

Between .2 and 2, use two decimal places (.20–1.99).
Between 2 and 20, use one decimal place (2.0–19.9).
Between 20 and 200, use units (20–199).
Etc.

Another expression of the same rule is: Under the given or similar conditions of material and technique, record three digits if the first (to the left) is 1, and otherwise record only two. In practice this means that a record of a tooth as 15 mm. in length is, for practical purposes, absolutely accurate and a record as 15.8 is a better approximation for most purposes although not absolutely accurate, but a record as 15.82 is in no respect better than as 15.8. If the measurements are large, it is advisable to change the unit so that no number larger than 199 need be used. Thus, under these conditions, 390 mm. should be recorded as 39 cm.; for the figure 390 implies a range 389.5–390.5, whereas the range really intended is 385–395 mm.. expressed by 39 cm., *i.e.*, 38.5–39.5 cm.

Such a rule naturally is valid only for the given conditions, but there is no difficulty in applying similar methods to any sort of measurement. If the degree of accuracy obtained proves to be insufficient for the purposes in mind, a refinement of technique and increase of accuracy are usually possible. Considerable inaccuracy is inseparable from the nature of some material, and in such cases, refinement of technique is useless, and only problems soluble by the relatively inaccurate data can be usefully attacked. In most paleontological work the degree of accuracy shown by the preceding example is quite adequate for the purposes involved, and in many cases a markedly higher degree of accuracy is impossible. In some other fields such measurements would be grossly inadequate, and accurate four- or five-digit measurements may be both possible and desirable.

In the absence of any other criterion, it is proper to record as many digits as are accurate or are found to be useful approximations by tests like that just described. When refinement may be increased indefinitely by changes in technique, there nevertheless comes a point beyond which it is useless to go; and for the determination of this point statistical methods provide the best criterion. If a series of specimens is to be measured, the most useful rule is to measure the largest and smallest specimens and then to adopt a minimum unit of measurement such that it is

contained at least 16 and up to 24 times in this range. If an adequate series is not available, a much rougher but still useful rule applicable to most linear dimensions is simply to record three digits.[1]

In practice this means that measurements of a variate ranging, say, from 10 to 12 mm. should be taken to .1 mm. This would give 20 steps within the range, which sufficiently meets the first rule. If the range were 75–95 mm., no decimal places need be recorded, for there are 20 integral steps in the range. The first example conforms also to the second rule. The second does not; but the rougher rule would result only in making measurements somewhat more refined than necessary, which does no harm. Except in a few special cases it is useless greatly to exceed the requirements of either rule, and unnecessary work can thus be avoided. For instance, with a range 10–12 mm., measurements to .01 mm., giving 200 steps within the range, even though entirely accurate would generally serve no useful purpose; and the refinements of technique and added labor involved in making such minute measurements would simply be wasted.[2]

In the great majority of cases these rules ensure data that will provide a maximum of useful information, sufficing for efficient statistics and for any other usual zoological purpose. It is not

[1] These may here be advanced as rules of thumb, but their real basis and validity will appear when statistical methods have been discussed. For reference, their derivation is as follows. Efficient statistics, those that give more accurate information for a given sample than could have been derived from any smaller sample, generally require grouping with a class interval not exceeding one-fourth of the standard deviation of the distribution. The smallest class interval available is the smallest unit of mensuration used for the raw data. The standard deviation cannot be exactly determined before the raw data are all gathered; but the range can readily be obtained, and the standard deviation is almost never less than one-sixth or more than one-fourth of this. Hence the efficient class interval and the largest suitable unit of measurement should be contained not less than 16 and preferably up to 24 times in the range. The more general but rougher rule depends on the fact that the range is seldom less than 15 per cent of any single value of the variate, generally much more. If it were at this low figure, a value 100 would usually imply a range of 15 or more, 200 of 30 or more, 300 of 45 or more, etc.; hence a record of three digits will almost always make available at least 16 steps within the range, which complies with the preceding rule.

[2] This applies only to the raw data. Statistical constants are often significant to more places than are these original observations.

true that the rules must be met in order to provide useful data. Measurements of the optimum refinement are not always practicable, and yet such substandard data are not on this account any less accurate and may be highly useful. They are merely less efficient.

BIAS AND CONSISTENCY

One of the most troublesome difficulties in using numerical data is bias, a tendency to favor some hypothesis or to lean toward a numerical result not purely objective. In this sense, bias is assumed to be unconscious and to have no flavor of disingenuousness. It usually arises either in sampling, discussed in the chapter devoted to that subject, or in measurement.

Bias in measurement is subjective and personal. It usually takes such forms as tendency to overrun or not quite to reach the accurate figure for the measurement in question, tendency toward or away from integral or some other certain values, or tendency to favor or oppose a given hypothesis. The existence of a tendency to overrun or underrun measurements can usually be detected by having two workers independently make a large series of measurements of the same objects in the same way. If the average result obtained by one is significantly smaller than that of the other, the existence of bias may be assumed and similar further tests made to determine whose the bias is, what its direction, and what its amount. The same sort of bias may often be both detected and corrected by taking measurements in duplicate in two different directions, for instance by opening calipers to the dimension sought, then closing them to it, and taking the mean if the two differ.

There is also a tendency when taking a series of homologous or numerically closely similar measurements to make them more nearly similar than is correct. This tendency, almost universal if attention is not paid to it, may be largely eliminated (1) by deliberately ignoring preceding readings and (2), when using calipers, by throwing them far off the last measurement before bringing them to the next. This precaution is an essential feature of good measuring technique.

If not forewarned, many students have a bias toward integral values; and if detected, this may be overcompensated by bias away from them. Such bias with respect to particular numbers

can usually be detected by checking over a large series of measurements of many different sorts and determining whether any one final digit occurs oftener than would be likely by chance. Care must be taken that the data are not such as would really tend to be concentrated about any one number in the last place. Tendency to favor a hypothesis is perhaps the most obscure bias of all and the most difficult to detect or to avoid. If there is any real possibility of such bias, measurements may be made by a worker not acquainted with the hypothesis in question.

In addition to the forms of bias mentioned, there are also biases of procedure, of instruments, and of materials. Some systems of dealing with specimens consistently make them appear longer or shorter than others. Biased instruments, such as one that does not return precisely to zero when closed or an inaccurately calibrated ruler, naturally produce biased results. Measurements of shrunken or swollen skins and other specimens are biased with respect to fresh materials. Inexact or incorrect specification of the dimension measured also produces an effect analogous to bias. The correctives for all these are fairly obvious.

The possibility of bias can generally be reduced to insignificance by duplication of measurement (perhaps varying the direction), by maintenance of an objective attitude, by carefully standardized procedure, by the use of highly refined instruments, by recording exactly what the measuring instrument says, by ignoring the purpose of the measurements as far as possible while they are being made, and by recording the results in smaller units than are to be used in ensuing calculation or publication. For trained observers some of these precautions are automatic, and others are unnecessary; but the complete elimination of bias is very difficult.

The distinctive feature of bias is some degree of consistency, a tendency to deviate from the ideal more often in some particular direction than in others. Since the usual purpose of measurements is to make comparisons, such deviations may have little or no effect on the conclusions drawn. Thus, such a form of bias as the almost unavoidable shrinkage of dead materials may be of no importance if it is sufficiently consistent; and the deviations from live measurements are hardly to be considered as bias as long as comparisons are made only between specimens comparably preserved. Similarly a worker may have a marked bias, and yet

it may not affect his comparisons so long as he is highly consistent and uses only measurements made by himself. It is a well-recognized fact in zoology that measurements made by one observer are always better comparable than those made by two or more different observers. Here there is not only the element of bias as it has hitherto been defined but also the related element of personal idiosyncrasies in the exact definition and orientation of measurements, which even the most rigidly standardized systems of mensuration do not wholly eliminate.

The factor of consistency is thus at least as important as that of bias, strictly speaking. Both factors are visible in such examples as that given by Sumner (1927) in recording the means obtained by each of three different observers measuring the same sample of 10 specimens on two successive days. The figures, for tail length in a sample of *Peromyscus maniculatus gambelii*, are given in Example 2.

EXAMPLE 2.—MEAN MEASUREMENTS TAKEN BY THREE OBSERVERS ON TWO DAYS
(Data from Sumner)

	First day	Second day
Sumner	74.9	74.4
Second observer	70.9	72.2
Third observer	70.2	71.1

The second and third observers were clearly biased with respect to Sumner, or he with respect to them; for his mean is on both days considerably larger than theirs.[1] The consistency involved is of two sorts: that for the figures of a single observer and that between those given by different observers. Each observer is reasonably consistent with himself, Sumner more so than the other two. The figures of the second and third observers are fairly consistent, but those of Sumner are not consistent with theirs. In fact the figures strongly suggest that the second and third observers used the same technique in nearly the same way and that Sumner used a different technique—on the face of the

[1] He does not give the data on which the significance of the difference could be evaluated, but it is so marked that it may reasonably be assumed to be significant.

figures it does not necessarily follow that Sumner's technique was more accurate or more refined, although this also is hinted. Such is, in fact, the case. Sumner measured the specimens on a special measuring frame with calipers calibrated to .1 mm., and the other two measured the loose specimens with a ruler. Incidentally the figures clearly show that measurement to .1 mm. was here unduly refined, even for Sumner's more precise methods, and that the last digit is not in any case either accurate or useful.[1]

Although the method is somewhat laborious, both of these sorts of consistency can be measured. Any rather long series of measurements is made by the same observer at two different times or by two observers, and the two sets of measurements are correlated (as explained in Chap. XII), the result being a coefficient that measures the agreement between the two sets and hence the consistency of the two series of measurements of the same objects. This is commonly called a "coefficient of reliability," but the term is misleading. Reliability more logically relates to the approximation of recorded to true values. This coefficient actually measures only consistency. An observer may have a strong bias, hence be unreliable (unless the bias is known, measured, and corrections applied); yet if this is consistent, the coefficient in question will give him a high rating. In fact since bias always involves some sort of consistency this method is not adapted to detecting or measuring this common defect in numerical data.

[1] Much the more so since these are means and not raw measurements. A large range of error of measurement in the first decimal place is implied even with Sumner's unusually fine technique.

CHAPTER III

FREQUENCY DISTRIBUTIONS AND GROUPING

FREQUENCY DISTRIBUTIONS

The first step in reducing original observations to more compact form and in preparing to draw any sort of conclusions from them is to tabulate them in the form of a frequency distribution. A frequency is the number of observations that fall into any one defined category, and a frequency distribution is a list of these categories with the frequency of each. Such distributions are the basis for almost all important numerical operations in zoology, and the use of numerical data depends on the definition of the categories or groups in which the data are to be placed.

Qualitative Distributions.—The grouping need not be and often is not in itself numerical. A common zoological grouping is that of the taxonomic system, the group being a subspecies, species, genus, or larger category in the hierarchy. Frequencies are employed when it becomes necessary to count the number of individuals within a given taxonomic unit observed under certain conditions: the number observed in traversing a defined area, the number caught by fishing operations, etc. Example 3 is a frequency distribution of this sort.

EXAMPLE 3.—SPECIMENS OF FOSSIL MAMMALS FROM THE SCARRITT QUARRY, MONTANA
(Original data)

Species	No. of specimens
Ectypodus hunteri	57
Leptacodon cf. *tener*	6
Bessoecetor thomsoni	30
Palaeosinopa senior	3
Unuchinia asaphes	1
Plesiadapis anceps	10
Carpodaptes hazelae	11
Phenacolemur frugivorus	3
Litolestes notissimus	61
Condylarths or Creodonts, undetermined	9
Pantolambdid undetermined	2

In other studies the groups may be defined ecologically, and the frequencies may be either of individuals or of species or genera observed within certain limits. Thus for the Bridger (Middle Eocene of Wyoming) mammalian fauna as known to Matthew (1909), a frequency distribution can be compiled as in Example 4.

EXAMPLE 4.—DISTRIBUTION OF BRIDGER MAMMALIAN FAUNA BY HABITAT TYPE
(Data from Matthew 1909)

Habitat	No. of known genera	No. of specimens in American Museum collection
Aerial	0	0
Arboreal:		
Surely	13	184
Probably	11	485
Terrestrial	17	314
Fossorial	3	8
Amphibious or aquatic	1	12

The groups may be geographic or based on habits and activities or on nonnumerical anatomical characters. Examples 5 and 6 will suggest the wide range of possibilities of this sort.

EXAMPLE 5.—PHYSIOLOGICAL CONDITION OF SPECIMENS OF THE EASTERN CHIPMUNK, *Tamias striatus*, TAKEN IN JULY
(Data from Schooley 1934)

Condition	Frequency
With embryos	11
Has ovulated recently	14
Not gravid, no recent ovulation	22

EXAMPLE 6.—BILL COLOR IN THE EUROPEAN STARLING, *Sturnus vulgaris vulgaris*, IN FEBRUARY
(Data from Hicks 1934)

Color	Frequency
Yellow (bills more than 85 % yellow)	33
Partial (20 to 85 % yellow)	20
Dark (less than 20 % yellow)	29

Quantitative Distributions.—Qualitative groupings give frequency distributions of nonnumerical variables, attributes. The analysis and use of such data are treated in Chap. XIV. This

and the several succeeding chapters are devoted more particularly to distributions of variates, those in which the categories are defined by numerical characters.

It was shown in the preceding chapters that the values of a continuous variate are already grouped as they are originally observed and recorded. Thus 9.2 mm. is not an absolute measurement but is the designation of the midpoint of a group of measurements, the absolute values of which may be anything greater than 9.15 and less than 9.25. Such a figure as 9.2 in this usage is thus simply a conventional way of representing a group with the limits 9.15 and 9.25.

When continuous variates are designated by a single figure, this figure is almost always the midpoint of the implied group. There are, however, some exceptions, and these may lead to serious numerical errors if not detected and adjusted to the usual convention. Thus in representing ages, this is usually done by the lower limit of the group. A child 2 years old is not between the ages of 1 year 6 months (1.5 years) and 2 years 6 months (2.5) but between the ages of 2 and 3. In all statistical operations, including taking the mean, commonest of all such operations, this has a strong influence as the following hypothetical distribution shows:

Recorded age	Frequency
2	6
3	20
4	5

Calculated on these data in the ordinary way (more fully expounded in Chap. V), the mean or average age of this group of infants would appear to be 3.0 years. The calculation is, however, invalid unless the records are adjusted to represent group midpoints, thus:

Midpoint of age group	Frequency
2.5	6
3.5	20
4.5	5

The mean age is now correctly found to be 3.5 years, a decided difference.

Some other age records are even more confusing. For instance, horse breeders advance the nominal age of all horses, whenever they were really foaled, on January 1, so that a "1-year-old

horse" may in reality be anything from just over 0 to just under 2 years in age, a "2-year-old" between 1 and 3 years, etc.[1]

Almost all numerical procedures are based on the convention that the figure recorded is the midpoint of the group; and if this is not true of a given set of data an adjustment must be made.

GROUPING

In making measurements the record is to the nearest unit, which may be at any point on the decimal scale, and the implied grouping is of the sort just discussed, with the record understood as the midpoint of a group extending one-half unit below and above this point. In compiling frequency distributions, it is often advisable to expand the group limits (called here secondary grouping), thus giving fewer groups and higher group frequencies. Such groups usually are represented not by a single figure,[2] but by two figures, called the "group limits" or "class limits," joined by a hyphen.

The relationship between the original measurements, the so-called group limits, the real limits of the groups so designated, and the midpoints of the groups is somewhat confusing, so much so that most textbooks on statistics and biometry give an incorrect usage and one that has led to errors, sometimes significant, in a great deal of the work published in this field even by professional statisticians. If original measurements are taken to .1 mm., then the classes of their distribution are designated by a series of single figures each .1 mm. larger than the last, as in Example 7.

EXAMPLE 7.—DISTRIBUTION OF MEASUREMENTS AS USUALLY GIVEN
(Hypothetical data)

Measurement	Frequency
9.1	1
9.2	5
9.3	10
9.4	7
9.5	3
9.6	2

[1] Apparently a special sort of grouping with the midpoint designated; but the situation is more complex than this, for the relationship of designation to group depends on the time of the year when the records were made. This particular case is of little interest to zoologists, but similar anomalies in data should be noted and taken into account.

[2] Except in the special case of "rounding" numbers; see p. 44.

If this were translated into the real group limits of the measurements, a form of record unnecessarily complex and never employed, although many errors might have been avoided in this way, it would read as in Example 8.

EXAMPLE 8.—DISTRIBUTIONS OF MEASUREMENTS BY REAL GROUP LIMITS

(Same data as in Example 7)

Limits of measurements	Frequency
9.05–9.15	1
9.15–9.25	5
9.25–9.35	10
9.35–9.45	7
9.45–9.55	3
9.55–9.65	2

If, now, it is decided to gather these measurements into larger groups, these new groups are usually designated by the smallest and the largest original measurements placed in them:

Group	Frequency
9.1–9.3	16
9.4–9.6	12

These figures 9.1–9.3 and 9.4–9.6 are what are called the group or class limits, but obviously they are not real limits. If they were, there would be no place in this grouping for measurements that lie between 9.3 and 9.4. Moreover, since all measurements of 9.1, 9.2, and 9.3 were placed in the 9.1–9.3 group, and since these measurements included, respectively, those between 9.05 and 9.15, 9.15 and 9.25, and 9.25 and 9.35, all measurements between 9.05 and 9.35 are included in the group labeled 9.1–9.3, and these are the real limits of that group. Similarly the real limits of the group labeled 9.4–9.6 are 9.35–9.65. This peculiarity arises from the fact that the figures 9.1, 9.3, 9.4, and 9.6 were themselves the midpoints, and not the limits, of groups.

The midpoints of the new groups thus formed are respectively 9.2 and 9.5. In calculations based on grouped data, these midpoints are taken to approximate the average value of all the measurements placed in the corresponding groups; and calculations are based on the midpoints, the fixing of the true value of which is thus important. For purposes of calculation the midpoint is often added in tabulating a frequency distribution of this sort:

Group	Midpoint	Frequency
9.1–9.3	9.2	16
9.4–9.6	9.5	12

This implies that although the 16 measurements placed in the group called 9.1–9.3 are really scattered from 9.05 to 9.35, their average value is about 9.2, which is exactly in the middle of their range.

Most workers have hitherto assumed that the lower figure in the group designation is in fact its lower limit, so that the group 9.1–9.3 does not include any measurements below 9.1 (not even 9.0999 . . .) and does include all measurements between 9.1 (exactly) and the lower limit of the next group, 9.4, or that the group limits are 9.10000 . . . (to as many decimal places as desired) and 9.39999 . . . (to as many places as desired). This assumption is very seldom true. With some types of data it could perhaps be made true, although this is almost never done; but measurements of the sort here under discussion cannot be placed in groups of which this assumption is true, although almost all biometricians have supposed that this could be and was done. Since everything from 9.05 to 9.15 was measured and recorded as 9.1, it is obviously impossible to place those of them below 9.10 in one group and those above in another. All go into the 9.1–9.3 group. Similarly measurements above 9.35 are not, as these workers imply, placed in the 9.1–9.3 group but in the 9.4–9.6 group.

If this assumption were correct and the lower so-called group limits were the real lower limits, then the midpoint of the group 9.1–9.3 would be 9.25, not 9.2, and they would arrange the distribution given correctly above in the following incorrect way:

Group	Midpoint	Frequency
9.1–9.3	9.25	16
9.4–9.6	9.55	12

Abundant examples of this erroneous identification of the midpoint may be found in almost any textbook of statistics or

report on research in which grouped data were used. The same error is also common, with even less reason, in dealing with grouped distributions of discontinuous variates.

The relationships between recorded measurements, conventionally stated class limits, real limits and midpoints, and the false limits and midpoints so often used are clearly shown in the accompanying diagram (Fig. 1). The magnitude of the groups formed is designated by the class interval, which is the distance from any point within a group, such as the lower stated limit or the midpoint, to the corresponding point in the next higher or lower group. The class interval is .3 in the examples just dis-

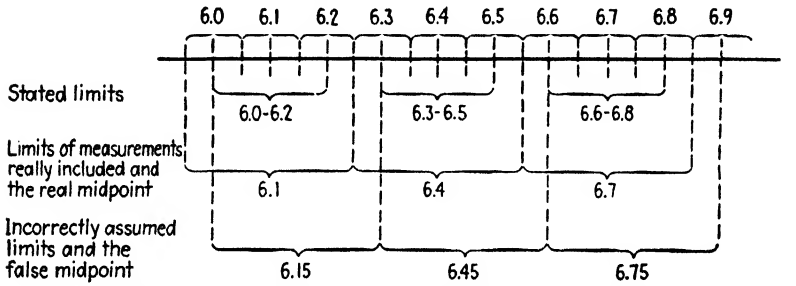


FIG. 1.—Midpoints and limits in primary and secondary grouping. The horizontal line represents the scale of all possible measurements of a continuous variate. The numbers above this line are original measurements, to .1 mm., which are in fact the midpoints of primary groups the ranges of which are shown by the brackets beneath the recorded measurements. Below the line is indicated secondary grouping with interval .3 mm. The stated limits are in terms of the original measurements and are shown not to correspond with the real limits. The lowest set of brackets shows an assumption commonly made as to limits and midpoints demonstrated by the diagram to be incorrect.

cussed. Although it is usual and preferable for most purposes to designate secondary groups by their conventional limits, a distribution may also be given by midpoints alone, even though the grouping is larger than that of the original measurements. If the classes are designated by one number and the difference between successive designations is not a single unit, it may be understood that the numbers are midpoints of enlarged groups and not measurements.

In some cases, classes are designated with a closer approach to the real limits by giving the second number one or two added places, generally filled by 9's, for instance, 9.1–9.399. Such a group does nearly approach the limits often erroneously assumed for a group 9.1–9.3, that is, it is meant to include everything

greater than 9.100 and less than 9.400. Of course, measurements to a single decimal place cannot be grouped in this way. The exception to the practice and usages already explained is only

EXAMPLE 9.—LENGTH OF P_4 IN A SAMPLE OF THE EXTINCT MAMMAL *Ptilodus montanus* FROM THE GIDLEY QUARRY
(Original data)

A

Original measurement, mm. (class interval .1 mm.)	Frequency
7.1	1
7.2	1
7.3	0
7.4	0
7.5	6
7.6	2
7.7	5
7.8	8
7.9	4
8.0	4
8.1	8
8.2	8
8.3	6
8.4	8
8.5	7
8.6	6
8.7	0
8.8	0
8.9	0
9.0	0
9.1	1

B

Class limits (interval .3 mm.)	Midpoints	Frequency
7.1–7.3	7.2	2
7.4–7.6	7.5	8
7.7–7.9	7.8	17
8.0–8.2	8.1	20
8.3–8.5	8.4	21
8.6–8.8	8.7	6
8.9–9.1	9.0	1

C

Class limits (interval .3 mm.)	Midpoints	Frequency
6.9–7.1	7.0	1
7.2–7.4	7.3	1
7.5–7.7	7.6	13
7.8–8.0	7.9	16
8.1–8.3	8.2	22
8.4–8.6	8.5	21
8.7–8.9	8.8	0
9.0–9.2	9.1	1

D

Class limits (interval .5 mm.)	Midpoints	Frequency
6.8–7.2	7.0	2
7.3–7.7	7.5	13
7.8–8.2	8.0	32
8.3–8.7	8.5	27
8.8–9.2	9.0	1

apparent, if this sort of designation is correctly used, for it really implies data to as many places as are given by the second figure. The group 9.1–9.399 in this type of designation is, or should be, the same as 9.100–9.399 in the type of designation previously

explained. The real midpoint of this group is 9.2495. It is generally assumed to be 9.25, which is not exactly right but is almost always an adequate approximation.

Example 9 shows a frequency distribution in terms of the original measurement to .1 mm. and with three different secondary groupings, two with class interval .3 mm. but with the limits at different points on the scale and one with class interval .5 mm.

To compile such frequency distributions, it is first necessary to make the measurements to as fine a point as will be required for any desirable grouping. These records will be irregularly scattered, for it is not practical to make them in the order of their magnitudes. The next procedures are to write down all the steps from the smallest to largest in the unit of measurement (to .1 mm. in the example), to tally against this from the original measurements, and then to reduce the tally marks to numbers. This results in the first form of distribution given above.[1] If a larger unit of secondary grouping is to be employed, the interval to be used and the point at which to start (or positions of the midpoints, determined by this) are decided and the frequencies taken from the distribution of the measurements.[2] This may be facilitated as in Example 10, using data from sections of the first two distributions, in Example 9.

This also facilitates the selection of the best secondary grouping, discussed in the next section.

It is customary to speak of the distribution in terms of the original measurements (*i.e.*, with the class interval equal to the smallest unit of measurement) as "ungrouped" and of a distribution with larger class interval as "grouped," but this is here avoided. Especially in conjunction with the record of measurements by midpoints rather than by limits, it obscures the fact that the measurements are really grouped (if of a continuous variate), a fact that should always be kept in mind.

Nondecimal Units.—Some workers take measurements in units that are not decimal and yet write them in the ordinary way, *e.g.*, measure only to half millimeters but record these as decimals.

[1] In some cases it is unnecessary to take this step if only larger groups are to be used—tallying may be directly against the class limits employed. It is, however, usually convenient also to have the original measurements in the form of a frequency distribution.

[2] Or tallied from the original observations.

This practice is confusing and indefensible in the face of the universal convention as to limits in decimal measurements. Such an author will record 2.3 mm. as 2.5, because it is nearer to that than to 2.0. By 2.5 he means a group 2.25–2.75, but his reader can only infer that 2.5 stands for the group 2.45–2.55 according to convention, a group to which the measurement does not really belong. It would be preferable to write the measurement as 2½ mm., thus showing that the unit of measurement was ½ mm.

EXAMPLE 10.—SECONDARY GROUPING, OR DECREASING THE NUMBER OF GROUPS IN A FREQUENCY DISTRIBUTION
(Part of data of Example 9)

Original measurements	Frequency	Frequency (interval .3 mm.)	Limits (interval .3 mm.)	Midpoints
7.7 7.8 7.9	5 8 4	17	7.7–7.9	7.8
8.0 8.1 8.2	4 8 8	20	8.0–8.2	8.1
8.3 8.4 8.5	6 8 7	21	8.3–8.5	8.4

and that the group implication is that the dimension is nearer 2½ than any other multiple of ½, *i.e.*, that the class limits are 2¼–2¾. Such a record has, however, the serious drawback that an integer, as 2, does not indicate the unit of measurement. This difficulty could be overcome only by writing all measurements as fractions, multiples of the unit used, *i.e.*, writing 2 as 4/2 and 2½ as 5/2 if the unit was ½ mm. Even this is clumsy and makes subsequent calculation based on the measurements difficult. Still worse are cases in which nondecimal fractional measurements are used but the fractional unit is not the same for the different measurements to be compared, for instance, one measurement may be recorded as 3⅓, another to be compared with this as 3⅛, another as 3⅕, etc.[1] It is practically impossible

[1] This has been a common practice in recording fish proportions.

to base valid frequency distributions and make accurate comparisons and calculations on such data. The general solution of these difficulties is to make measurements in decimal units whenever possible and, when this is not possible or for some special reason is undesirable, to make records by class limits, not by class midpoints. Thus 2½ mm. should be recorded decimally as 2.3–2.7 (conventional limits) or 2.25–2.75 (real limits), not as 2.5.[1]

Rounding Figures.—The practice of rounding figures, that is, of reducing the number of places or digits recorded, is a special case of grouping in which the class intervals are increased in powers of 10. Thus a measurement 2.132 mm. has a class interval of .001 mm. If rounded to 2.13, the interval is multiplied by 10, becoming .01; or if rounded to 2, the interval is multiplied by 1,000, becoming 1.

Rounding introduces a special difficulty not present in enlarging groups by multiplying intervals by any number not a power of 10. If the multiplier is not decimal, the real class limits are not integral in the last place recorded in the original measurement; but with a decimal multiplier they are. Thus if records are to .1 mm., interval .1, and the interval is increased to, say .4 mm., the new class limits are not integral tenths but in multiples of .05. A measurement of 2.5 mm. (real limits 2.45–2.55) will then fall into such a group as 2.4–2.7 (real limits 2.35–2.75). There is no ambiguity in placing any multiple of .1 mm. in such a group. But if the class interval is multiplied by 10, making it 1 mm. and rounding the figures to integral millimeters, real class limits become 1.5–2.5, 2.5–3.5, etc. (not 1.45–2.55 or any similar multiples of .05). Where, now, is a measurement of 2.5 to be placed? It could either become 2, of which it is the upper limit, or 3, of which it is the lower limit.

A possible solution for a series of such measurements would be to round half of them into the lower and half into the upper group, but this is seldom practical. Moreover, the operation of rounding is applied to single figures as often as to series. Another and better solution is to go back and carry out the original measurement, or the calculation from which a figure to be rounded was derived, to more places, until it is no longer on a limit of the rounded group. Thus, if a number 2.35 is to be rounded to

[1] The class midpoint 2.5 would, however, be used in some subsequent operations, such as finding the mean.

multiples of .1 mm., it could be placed with equal reason either as 2.3 or as 2.4. If, however, the original measurement or calculation be carried out to another place, the figure might prove to be 2.347, showing that it should be rounded to 2.3, or 2.351, to be rounded to 2.4, etc. If the third decimal place proves to be zero, that is if the figure is 2.350, a fourth place will have to be obtained, when, for instance, 2.3498 will be rounded to 2.3 or 2.3502 to 2.4.

That is the only accurate, unbiased way of rounding numbers that happen to fall at the class limits. In dealing with measurements, however, it is seldom practicable. The original specimens may not be available, or more refined measurements may not be accurately possible. Even with a number derived from calculation, its carrying to more places may be unnecessarily laborious and may take it beyond any reasonable limits of accuracy and significance. It is therefore customary in rounding numbers to take arbitrary limits at .5 below and .4 above the midpoint of the rounded group: 2.35 is rounded to 2.4; 2.34, to 2.3; 1.5 or 2.4, to 2; but 1.4 to 1, and 2.5 to 3.

This convention introduces bias into figures. The real class limits of 2.3 are 2.25 (or 2.250)–2.35 (2.350). The rounding convention makes the limits used actually 2.245–2.345, or .005 lower than the real limits. Thus, rounding by this rule tends to reduce the recorded values of numbers by an amount equal to one-half the smallest unit of the original number. This means that if numbers to two decimal places are rounded to one decimal place, they will, on an average, be reduced in value by .005; or if numbers to one decimal place are rounded to integers, they will, on an average, be reduced in value by .05. This is a small difference; it usually has no significance, and the slight loss of accuracy is compensated by the simplicity of rounding in this way. At the same time the existence of this bias should be known; and if it has any probable bearing on the results obtained, it should be eliminated either by not rounding or by rounding in the accurate way explained above.

FREQUENCY DISTRIBUTIONS AND GROUPING OF DISCONTINUOUS VARIATES

The quantitative distributions and groupings hitherto considered have all been of continuous variates. Discontinuous

variates also yield quantitative frequency distributions that, with a few exceptions, have the same properties and are the basis for the same sorts of calculations and statistical operations as the distributions of continuous variates. Typical but varied examples of discontinuous variates are given in Example 11 A–D.

Such data may be grouped in the same way as those of continuous variates. In such grouping, the recorded counts are to be considered as midpoints and not as limits, and the midpoints of the groups are halfway between the original counts.[1] Among numerous other ways, the last example given could be grouped as shown in Example 12.

EXAMPLE 11.—DISTRIBUTIONS OF DISCONTINUOUS VARIATES

A. A variable physiological function. Number of breaths taken in a single breathing period by a young Florida manatee (data from Parker 1922)

Breaths taken (the variate)	No. of times observed (frequency)
1	16
2	13
3	2
4	2

B. Variable reproduction. Number of young in nests of tree swallows, *Iridoprocne bicolor* (data from Low 1933)

No. of young (variate)	No. of nests (frequency)
1	1
2	4
3	7
4	31
5	56
6	17
7	4

C. Discontinuously variable anatomical character. Number of serrations on the last lower premolars of specimens of the extinct mammal, *Ptilodus montanus*

No. of serrations (variate)	No. of specimens (frequency)
13	8
14	19
15	2

[1] Most statisticians use the counts as if the first were a lower limit and, for instance, give the midpoint of a group 40–42 of a discontinuous variate as 41.5; but this is certainly incorrect. The midpoint of 40–42 is 41.

EXAMPLE 11.—DISTRIBUTIONS OF DISCONTINUOUS VARIATES.—(*Continued*)

D. Discontinuously variable anatomical character. Number of caudal scutes in the king-snake, *Lampropeltis getulus getulus* (data from Blanchard 1921)

No. of caudals (variate)	No. of specimens (frequency)
40	3
41	2
42	4
43	4
44	4
45	7
46	6
47	3
48	9
49	2
50	5
51	2
52	1
53	2
54	3
55	0
56	1
57	0
58	1

For a discontinuous variate, the raw data, the individual counts, are not grouped but are absolute values. The 6 specimens of *Lampropeltis g. getulus* with recorded caudal counts of 46 do not have from 45.5 to 46.5 caudals. Each of the 6 has exactly 46 caudals.[1]

In the grouping with interval 4 (or any other even number), the midpoints are fractional and are not values that the variate can actually take. This introduces a series of imaginary values if the midpoints are used in calculation, and the conception is difficult and (some believe) logically objectionable. For this reason it may be preferred to use only groupings with intervals of odd numbers, so that the midpoints will be integers and will be real values of the variate. In fact, however, the use of imaginary values as midpoints is not mathematically invalid, the results

[1] Occasionally there is some doubt in the observations; it may be, for instance, that a count could equally well be called 5 or 6. In such cases, if decision is really impossible by any objective means, it may be valid to record the count as 5½, and this may even produce greater accuracy in the derived statistical data. If this method is employed, attention should be called to the exact usage, since there is no general understanding of such a convention.

based on them are just as accurate as those based on real values as midpoints, and there is no reason why they should not be used if more convenient in other respects.

For purposes of calculation the ungrouped data are usually used for discontinuous variates. This is more accurate, and the number of steps is seldom so great as to make the calculation unduly laborious. The form of the distribution is, however,

EXAMPLE 12.—TWO FORMS OF SECONDARY GROUPINGS OF DATA FROM EXAMPLE 11D

A. Interval 3			B. Interval 4		
Group	Midpoint	Frequency	Group	Midpoint	Frequency
40–42	41	9	40–43	41.5	13
43–45	44	15	44–47	45.5	20
46–48	47	18	48–51	49.5	18
49–51	50	9	52–55	53.5	6
52–54	53	6	56–59	57.5	2
55–57	56	1			
58–60	59	1			

usually clearer if some grouping is employed, especially with the small samples usual in zoology.

NUMERICAL QUALITATIVE GROUPING

In the distributions of variates discussed so far, the categories in which the observations are grouped and for which frequencies are recorded are themselves quantitative concepts. It is also possible and often is highly useful to employ categories that are defined by numerical data but that are conceptually qualitative and the consideration and analysis of which should be with the viewpoint and methods of the study of attributes rather than of variates. Because the categories are defined numerically, such groupings are easily confused with truly quantitative distributions, and it is important to recognize the distinction.

One of the commonest of such arrangements of data, especially useful in studying association (see Chap. XIV), is the division of a frequency distribution into two groups, one in which the values of the variate exceed and one in which they are less than

a given value. The value selected may be the midpoint of the distribution or may be at a break in the distribution or at any other point suggested by the problem in hand. In any case, the resulting twofold grouping, although literally quantitative, is effectively qualitative. It is a division not into a series of equal, quantitative steps, but into larger and smaller qualitative groups. Such a division, from data for a continuous variate, using a break in the distribution as the division point, is given in Example 13.

EXAMPLE 13.—TOTAL LENGTH OF FEMALES OF THE KING-SNAKE, *Lampropeltis elapsoides elapsoides*
(Data from Blanchard 1921)

A. Quantitative grouping (interval 25 mm.)		B. Qualitative grouping	
Length	Frequency	Length	Frequency
150–174	1	Less than 300	9
175–199	2	Greater than 300	25
200–224	3		
225–249	2		
250–274	0		
275–299	1		
300–324	0		
325–349	2		
350–374	2		
375–399	5		
400–424	4		
425–449	4		
450–474	3		
475–499	2		
500–524	3		

Such a grouping might be made, for instance, to see whether larger size and smaller size, as attributes, can be associated with greater age and lesser age, or with occurrence in two different regions, or with any other factors.[1]

Sometimes a multiple grouping will be suggested as in Example 14.

[1] The association is almost certainly with age in this example, although the data to prove this are not available.

EXAMPLE 14.—TOTAL LENGTH OF THE RECENT FISH *Caranx melampygus*
(Data from Nichols 1935)

Qualitative length groups	Frequencies
104–115	4
120–125	7
128–136	5
147–169	2
195–208	2

Such a grouping suggests a quantitative frequency distribution and may be mistaken for one or may even be supposed by its author to be one. If it were so intended, it would be a very careless and unsound grouping, and conclusions based on it would be inaccurate and probably invalid. As a qualitative grouping, however, it is valid and useful. In the example these are qualitative age groups necessarily defined indirectly in terms of size and shown to be associated with various growth phenomena.

CRITERIA FOR SECONDARY GROUPING

In designating numerical groups, it must be clear whether the numerical designation is the midpoint, lower limit, upper limit, or both limits, and whether the limit is absolute or is in terms of the original measurements. It is assumed that a single number designates a midpoint unless the contrary is explicitly stated. If only the lower limit or only the upper limit is given, this usage must be specified. If two figures separated by a dash are given, these are the two limits. It may usually be assumed that these are given in terms of the original measurements, hence that they are midpoints of the smaller groups of observation from which the larger groups have been derived. If the figures are intended as absolute limits, they are generally and should always be distinguished either in words or by added decimal points on the second figure. Thus 20–22, designating a group for a continuous variate, will be assumed to be in terms of original measurement and hence to have the true limits 19.5–22.5 and midpoint 21; but 20–21.99 is assumed to represent absolute limits, not 19.5–22.5 but 20–22, the midpoint still being 21. It has been noted that two-place data cannot be distributed within groups of such absolute limits. An observation 20 would be

indeterminate as to group, for its value can be anything from 19.5 to 20.5 and if it is below 20.0 it does not belong in the 20–21.99 group. It does belong, however, in a 20–22 group (in terms of original measurements).

For discontinuous variates the group limits are also assumed to be in terms of the original counts.

The first requirement of valid grouping, and one of the most frequently ignored, is that the group designations be accurately defined and that all the data enter the groups as designated.

In quantitative grouping the next requirements are that the groups should be of equal size and mutually exclusive. Within a single distribution, groups such as, for instance, 10.5–11.4 and 11.0–11.9 or 10.5–11.4 and 11.5–11.9 are incompatible and should never be used. The data must be arranged in groups such as 10.5–10.9, 11.0–11.4, 11.5–11.9, or any other series of equal and exclusive steps. In qualitative grouping, whether on numerical or other definitions, the criterion of equal size does not apply; but the principles of unity and exclusion are equally important and sometimes are more obscure so that failures in this respect are common in the literature. It is frequently stated that a given character is present in a certain number of cases (has a certain frequency) and absent or indeterminate in so many others. This grouping is invalid because the second group may or does include among the indeterminate cases some that had the character and hence belong in the first group. The twofold grouping is not mutually exclusive, and there are really three groups: present, absent, and indeterminate. But presumably it is the presence or absence of the character that is being studied, so that the indeterminate specimens really have nothing to contribute to the problem and should not be included in the data.[1]

Decision as to what secondary grouping is to be made depends on the uses to which the groups are to be put. These uses are

[1] This simple logic is so often contravened that it apparently is not obvious and requires statement. Commonly the form of the error is to say that 50 per cent of the specimens have the character and that in the other 50 per cent it is absent or indeterminate; or, slightly better but still wrong in most cases, that 50 per cent have the character, 30 per cent do not, and 20 per cent are indeterminate. The correct expression of these facts is that of the determinable specimens 62.5 per cent have the character and 37.5 per cent do not.

discussed in detail in the following chapters, and the purposes and procedures of grouping will be clearer when these chapters

EXAMPLE 15.—TWO ARRANGEMENTS OF SECONDARY GROUPING, WITH THE SAME INTERVAL

Original measurement	Frequency	A. Grouped frequency (interval .3)	B. Same in different position
8.7	0		
8.8	0	1	
8.9	1		1
9.0	0		
9.1	8	18	
9.2	10		25
9.3	7		
9.4	5	14	
9.5	2		14
9.6	7		
9.7	0	7	
9.8	0		

Distribution A			Distribution B		
Class limits	Midpoints	Average of individual measurements included	Class limits	Midpoints	Average of individual measurements included
8.7–8.9	8.8	8.90	8.8–9.0	8.9	8.90
9.0–9.2	9.1	9.15	9.1–9.3	9.2	9.20
9.3–9.5	9.4	9.36	9.4–9.6	9.5	9.51
9.6–9.8	9.7	9.60			

In the first grouping A, the frequencies are irregularly distributed within each of the classes shown. As a result, the class midpoints differ by .10, .05, .04, and .10 from the means of the individual measurements which, in calculation from such grouped data, they are taken to represent. In the second grouping B, the frequencies are almost symmetrically distributed in each class, and the class midpoints agree almost exactly with the means of the individual measurements.

have been read. In general the purposes of secondary grouping are to simplify calculation and to bring out formal characteristics

of the distribution. Frequently, especially with small samples, the same grouping will not serve well for both purposes.

Grouping is defined by the class interval and by the position of any one limit or midpoint. The class interval together with the total range of the observations to be grouped determines how many classes or steps there will be in the grouped distribution. This, in turn, determines the concentration or dispersion of frequencies. Since the total frequency is fixed, if there are fewer classes each will tend to have a higher frequency; and if there are more each will tend to have a lower frequency. In grouping for calculation the number of classes should generally be between 15 and 25. If the original data cover only 25 or fewer steps, calculation should be based on these data and not on further grouping.[1] It is rarely advisable to group discontinuous variates for calculation.

In calculating from secondary grouped data, the class midpoints are taken to represent all the observations included in the class. It therefore follows that in grouping for this purpose that arrangement is best that produces groups in which the midpoint of each class most nearly corresponds with the mean of the individual values included in the class, or in other words in which the individual values in each class are most symmetrically distributed around the midpoint of the class. If the secondary grouping is done from a frequency distribution of the individual values (original measurements), the degree to which this ideal is approached and the position of the classes that best corresponds with it may easily be determined by inspection and trial. Thus in Example 15, an extract from a (hypothetical) distribution, arrangement B is clearly better than arrangement A although the interval is the same.

Even if the secondary grouped distribution is not to be used in calculation, it is well to follow this criterion as much as possible. The more nearly these conditions can be fulfilled, the more proper it is to reduce the number of classes or to increase the class interval.

[1] Some of the examples in this book do not follow this rule. They are given to show the method involved and not for the sake of their concrete results. The method is sometimes easier to comprehend if fewer groups are used than should be employed in actual research, and space is also saved with this gain in clarity. In practice, also, fewer than the ideal number of groups may sometimes be used to obtain a quick result in preliminary work.

If the frequencies are low, the whole sample small, even though the number of steps (classes) in the original data is considerably greater than 25, secondary grouping for calculation may be unjustified. In addition to the criterion just explained, in such cases this grouping, without reducing the number of classes below 15, should produce a marked concentration of frequencies in some classes and should tend to eliminate frequencies of 0 or 1 except toward the ends of the distribution.

Grouping that is satisfactory for calculation is also satisfactory for the further purpose of reducing the bulk of the data for publication. All that is required for publication is that satisfactory results be derivable from the data published, so that a compact table on which accurate calculations can be based is just as good as a much longer and more diffuse table of the raw measurements.

The way in which secondary grouping can be used to bring out the form of the distribution is well exemplified by the figures for caudal scutes of *Lampropeltis getulus getulus* on pages 47–48. The frequency distribution of the original data is long and irregular, and it is difficult to detect any pattern in it. When these are grouped with interval 3, giving 7 classes, a very definite pattern emerges. When they are grouped with interval 4, 5 classes, a similar pattern is evident but it is now so compressed as to be less clear. Evidently for these data a secondary grouping with interval 3 is better than the raw data or than any other grouping to show the distribution pattern. That secondary grouping is best for this purpose which most clearly and smoothly brings out such a pattern, a criterion that will be more easily applied when the sorts of patterns involved have been considered in the next and subsequent chapters.

Secondary grouping for this purpose generally requires fewer classes than are advisable for calculation, and this is particularly true with small samples such as are usual in zoology. The number of classes should usually be less than 16 and more than 4, while for calculation they should if possible be more than 15. In grouping for pattern, it is often an advantage to have an odd number of classes; for this will give a middle class, an important point in most zoological distributions. It should tend to smooth out any small random fluctuations in the frequencies so that they tend to rise or to fall steadily through several successive classes. In Example 12, with interval 3, they rise through the first three

and fall through the last five classes in an orderly way, and in the raw data (Example 11) they reverse direction eleven times. The grouping should tend to eliminate frequencies of 0 within the distribution and also any very low frequencies except toward the ends. In the example mentioned, the raw data have two internal zeros and several low frequencies of 1 to 3 far from the ends, and the grouped data have no internal zeros and have relatively low frequencies only in the last two classes, where they may be expected to occur in any case.

CHAPTER IV

PATTERNS OF FREQUENCY DISTRIBUTIONS

GRAPHIC REPRESENTATION

A frequency distribution has characteristics of its own, not seen in the isolated observations, and these are properties of the data as a whole on which the most important deductions and comparisons can be based. The essential characteristic of a distribution is a pattern formed by the rise and fall of the values of the frequencies as the values of the variate increase. This pattern is shown by the distribution in numerical form, but it almost always stands out more clearly if it is made into a diagram or picture, and such graphic distributions may convey much of the information in the most simple, rapid, and concise way.

In all graphs of frequency distributions, the values of the variate are laid down on a horizontal line, the x-axis, as abscissa, starting at the lower left-hand corner of the diagram, and frequencies are measured from the same point upward along the vertical y-axis as ordinate. For purposes of such plotting, the axes may be called the X- and f-axes, X being a conventional symbol for the value of a variate and f for its frequency, which in these cases takes the place of the conventional y in mathematical curve plotting.

Aside from a few exceptional cases, the initial value of the f-axis should be 0. It would be preferable also to begin the X-axis scale at 0; but if the lowest X of the distribution is a large number, as it often is, this means that a large blank space will occur to the left of the diagram. In such cases it is usually advisable to begin the X-scale at an arbitrary number shortly below the lowest observed value of X.

The simplest way to construct such a diagram is to place dots at points defined by the pairs of corresponding X and f values. These are not very satisfactory because the scattered dots do not

readily suggest a pattern[1] and the magnitudes involved are not readily grasped.

A dot diagram of this sort is changed into a frequency polygon by drawing a line from each dot to the next. The line is preferably joined to the edge of the diagram by including on each side a value of the variate for which the frequency is 0.[2] In a frequency polygon (if both ends have 0 frequency), the whole area is proportional to the total frequency, and the distances from the points (usually angles of the polygon) to the X-axis are proportional to the class frequencies. This type of diagram has the disadvantages that the verticals to the X-axis are proportional to frequencies only at these points, where the frequency is supposed to be concentrated, and that the areas above the X-axis for the given classes, magnitudes generally clearer to the eye than are the linear distances, are not proportional to the frequencies. The principal advantage of the frequency polygon is that it nearly resembles a curve, the theoretical form to which the angular pattern is to be related. This advantage is generally not great for anyone accustomed to the use and characters of distributions, and frequency polygons are not very commonly used. They should particularly be avoided if there are abrupt changes of slope, which tend to make the polygon misleading.

The commonest and for most purposes the best graphic representation of a frequency distribution is by a histogram. To make a histogram, a vertical line is erected at each class limit, and these are connected across their tops by horizontal lines at a height equal (on the f-scale) to the frequency of the class. If raw measurements or counts are used, it should be remembered that these are midpoints, and the class limits should be drawn halfway between them. If secondarily grouped measurements are used, the true limits and not those of any convention are to be used (see Chap. III). Sometimes in the final drawing vertical lines within the diagram are omitted and only the external boundaries of the histogram drawn, but such a diagram is usually harder to read.

[1] They are also liable to confusion with a scatter diagram, which is quite different from a frequency distribution (see page 255).

[2] Occasionally there is no such value, when the diagram is simply framed arbitrarily, or the abscissa $X = 0$ will have a frequency above 0, when the line ends at the f-axis.

The f-scale is marked to the left of the diagram either in units or in convenient multiples, as by fives or tens. The unit of the X-scale should be the class interval, and designations should be either at (true) limits or at midpoints. The latter is usually preferable, and in either case the numbers should be so placed as to leave no doubt as to their positions in the classes. In frequency polygons and in graphs of discontinuous variates the designations on the X-scale must represent midpoints.

EXAMPLE 16.—FREQUENCY DISTRIBUTIONS

A. Widths of last upper molars of the extinct mammal *Acropithecus rigidus* (original data)

Measurement	Frequency	Measurement	Frequency
5.4	1	6.1	8
5.5	0	6.2	6
5.6	0	6.3	2
5.7	5	6.4	0
5.8	2	6.5	2
5.9	5	6.6	0
6.0	4	6.7	1

B. Size of sets of song-sparrows, *Melospiza melodia beata* (data from Nice 1933)

No. eggs (X)	No. nests (f)	No. eggs (X)	No. nests (f)
1	1	4	25
2	2	5	14
3	19		

In a histogram each class is represented by a rectangle. The horizontal widths of these are all the same, and their heights are proportional to the class frequencies. The areas are therefore also proportional to the class frequencies, the great advantage of this sort of diagram.[1]

The same distribution may be represented by histograms of markedly different superficial aspect depending on where the classes are placed and on their magnitude. If the class interval is increased, the frequencies of many or all classes will also be increased and those with frequencies already high will be increased the most. The histogram with larger class intervals will therefore rise higher and will have greater relief.

[1] Sometimes the rectangles are not of equal width, different class intervals being employed in different parts of the distribution. The result is exceedingly confusing, and this method should be employed only in very exceptional cases.

This is a characteristic only of the secondary grouping and not of the distribution, so that it is necessary to recognize essentially the same types of curves with different groupings and also to employ, as far as possible, the same class intervals for two distributions that are to be compared. In placing the classes on the scale, the position that gives the most symmetrical result is usually preferable.

Figures 2, 3, and 4 show various forms of graphic representation of frequency distributions using the two sets of data given in Example 16, one of a continuous and one of a discontinuous variate.

THE MEANING OF DISTRIBUTION PATTERNS

If a frequency polygon were based on a series of observations that can be multiplied indefinitely and at the same time made more refined at will, it would be possible to decrease the class interval and at the same time to increase the number of observations so that the class frequencies remained reasonably large. Continuing this process, a condition would be reached when the dots, the angles of the polygon, were so close together that they became indistinguishable; for the horizontal distance between any two successive dots is equal to the class interval, and this is made indefinitely small. The polygon would then cease to have visible corners and angles and would become a smooth curve. The same procedure applied to a histogram would produce the same result, since the horizontal lines forming the tops of the rectangles would become shorter and shorter with decrease of the class interval, to which they are equal, until eventually they would appear only as points which would coalesce and form a curve.

This curve that is approached as a limit when the class intervals are decreased and the total frequency increased is the ideal pattern of the corresponding frequency distribution. In practice the curve cannot be obtained in this way; for no method of measurement is sufficiently refined for the indefinite reduction of the class interval, nor can the number of observations ever be really increased indefinitely. The true ideal curve would, indeed, only be reached when the class interval reached zero and the total frequency infinity, an obvious impossibility in practice. Yet the approach of the distribution to this purely theoretical

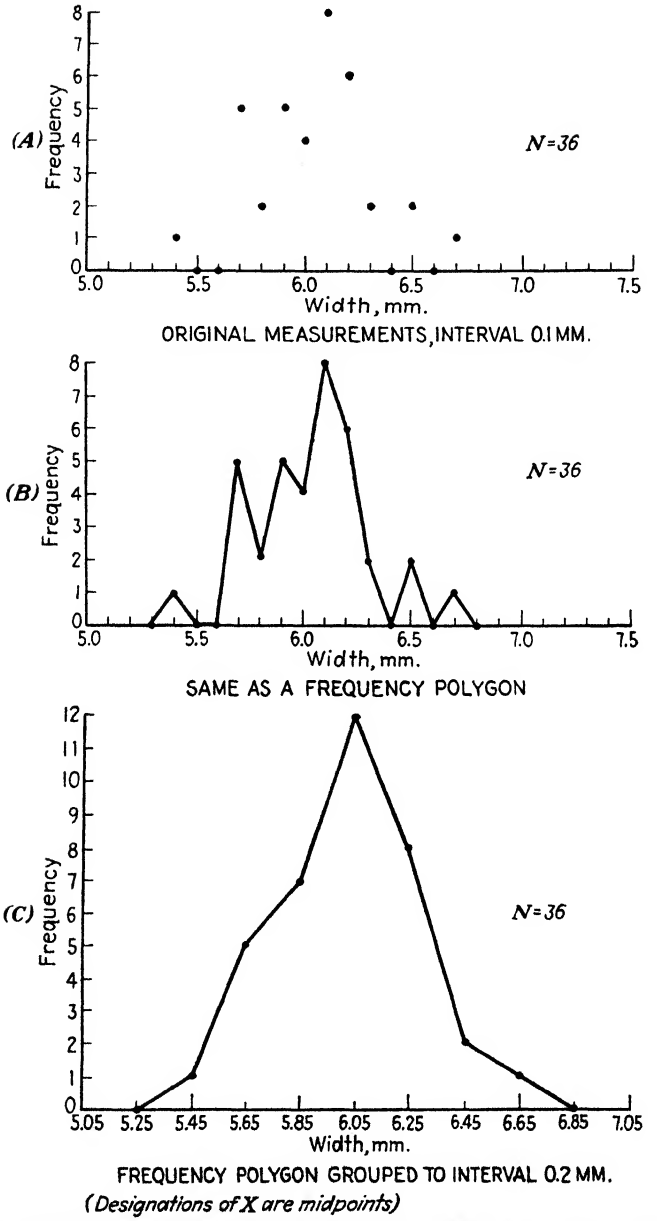
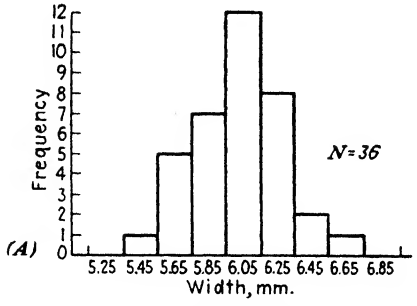
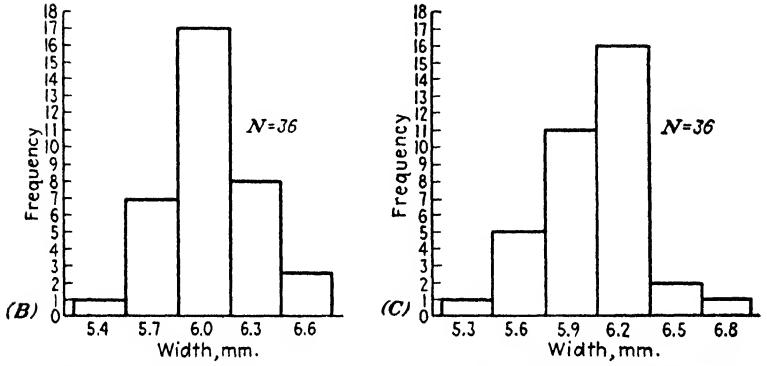


FIG. 2.—Graphic representations of a continuous frequency distribution. Width of the last upper molar in the fossil mammal *Acropithecus rigidus* (data of Example 16*A*). *A*, the raw data plotted by dots. *B*, the raw data as a frequency polygon. *C*, frequency polygon of the data regrouped to interval .2 mm., more clearly bringing out the real form of the distribution.



HISTOGRAM, GROUPED WITH INTERVAL 0.2 MM.
(Designations of X are midpoints)



HISTOGRAMS, GROUPED WITH INTERVAL 0.3 MM., WITH LIMITS IN DIFFERENT POSITIONS
(Designations of X are midpoints)

FIG. 3.—Histograms of a continuous frequency distribution (same data as in Fig. 2). *A*, regrouped to interval .2 mm., corresponding to the polygon of Fig. 2*C*. *B*, regrouped to interval .3 mm., showing change of form by broadening of class intervals; the grouping of Fig. 3*A* is preferable as it is more refined and yet gives an equally clear pattern. *C*, regrouped to interval .3 mm. with the midpoints taken at different values; the position shown in Fig. 3*B* is preferable as it more correctly shows the essential symmetry of the distribution.

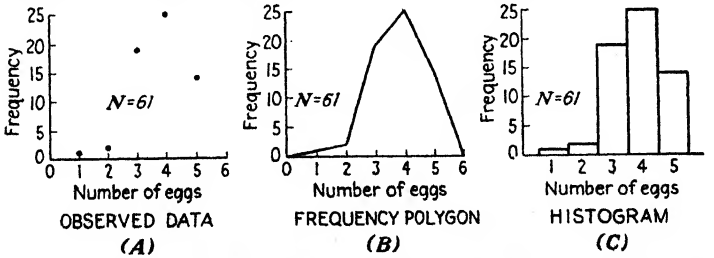


FIG. 4.—Graphic representations of a discontinuous frequency distribution. Number of eggs in nests of the bird *Melospiza melodia beata* (data of Example 16B). *A*, the raw data plotted by dots. *B*, the same plotted as a frequency polygon. *C*, the same plotted as a histogram.

limit is a real phenomenon, and the theoretical curve is the best possible representation of the distribution as a whole. The study of a frequency distribution thus commonly involves setting up a hypothesis as to the curve represented by the data of the actual observations and estimating the mathematical constants by which the curve can best be defined.

The importance to the zoologist of these operations and of the numerical values assigned to constants of frequency distribution curves is very great and can hardly be overemphasized. They give him a way of describing not single individual specimens or observations but the whole series of specimens and observations available to him. They permit estimates of the characters of the whole population from which his series is drawn and measurements of the probability that two sets of observations are drawn from the same or from different populations. The importance of such data for taxonomy is obvious and their importance is equally great for any consideration of variates. In fact, single observations on variates are practically meaningless in themselves. They are only a means to an end, and the end is estimating characters of frequency distributions. The essential character of a variate is its frequency distribution pattern, and no work in this field is really valid and useful unless it is done from this point of view.

GENERAL TYPES OF DISTRIBUTION PATTERNS

In the great majority of cases the characters, anatomical, physiological, psychological, or other, with which a zoologist deals are distributed in such a way that certain classes of these variates are more frequently observed than others and that the frequency becomes progressively less as the classes are farther in either direction from these most common values. This fact, so often seen by anyone dealing with zoological data that it becomes a basic assumption of the science, is often called Quetelet's law or, better, principle.[1] As with most of the so-called laws of biology

[1] L. A. J. Quetelet (1796–1874) was a Belgian astronomer, meteorologist, and statistician. He observed that the individual characters of man are distributed about the value for the hypothetical "average man" in the same way that single observations of mathematical probabilities are arranged around a mean result. This opened the way for the application of the theory of probabilities to the study of numerical observations on biological and zoological materials.

and zoology, there are some exceptions; but these are rare and usually belong to certain distinctive classes of data, so that zoological variates may generally be assumed to fall into a pattern approximately specified by Quetelet's law.

A large number of specific types of curves have been observed in frequency distributions of zoological variates. The distinction and specification of many of these require such extensive data and such intricate mathematical procedures that they are of little or no use to the zoologist, and, even if not entirely beyond

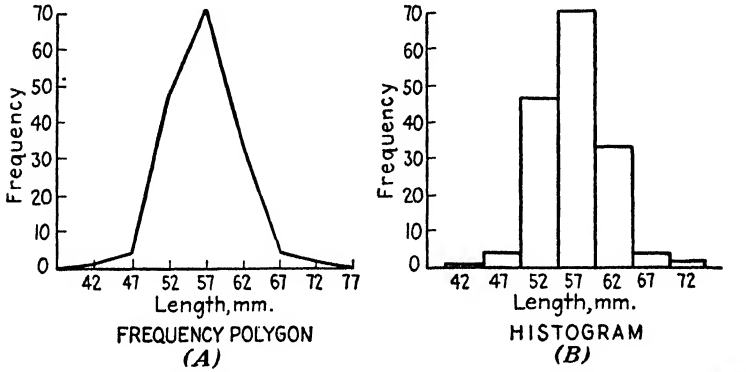


FIG. 5.—An essentially symmetrical frequency distribution following Quetelet's law. Lengths of the fish *Pomolobus aestivalis* (data of Example 17). *A*, as a frequency polygon. *B*, as a histogram. The degree of asymmetry observed is usual in samples of essential symmetrical populations.

his powers, such work would be a waste of time and effort. Moreover, many of these curves, most of those commonly involved in zoological work, approach a few standard types so closely that they are most usefully studied as approximations of these standard curves and specified in terms of the latter with, if necessary, estimates of deviation from them.

All such curves can be classed in four general groups:

1. Those high at a midpoint and sloping away nearly symmetrically on each side of this.
2. Those with a high point not at the midpoint of the distribution and sloping away from this with moderate asymmetry.
3. Those with the high point near or at one end of the distribution and strong asymmetry.
4. Those with a low point within the distribution and rising at both ends.

These are not absolutely clear-cut categories: 1 grades into 2, 2 into 3, and 3 into 4; but a given distribution can usually be referred to one of these general types.

Absolute symmetry almost never occurs in a limited set of observations, indeed so rarely that its appearance may be viewed with suspicion. Distributions nearly enough symmetrical to be considered as essentially so are, however, common. This is the ideal form of most animal characters that follow Quetelet's law. Numerous examples appear in the pages of this work; and Example 17, given in graphic form in Fig. 5, serves to illustrate the type here:

EXAMPLE 17.—FREQUENCY DISTRIBUTION. LENGTHS OF SPECIMENS OF THE GLUT-HERRING *Pomolobus aestivalis*, CAUGHT IN CHESAPEAKE BAY, OCTOBER 16–31

(Data from Hildebrand and Schroeder 1927)

Length, mm.	Frequency
40–44	1
45–49	4
50–54	47
55–59	71
60–64	33
65–69	4
70–74	2

This is slightly asymmetrical since classes at corresponding distances on each side of the middle do not have the same frequencies and the frequency below the middle class, 52, is greater than that above, 39.[1] It is, however, an approximation to symmetry about as close as is commonly to be expected unless the total frequency is very large. The highest frequency is in the middle class, and the distribution does taper off steadily for equal distances on both sides of this.

Moderately asymmetrical curves are spoken of as being moderately skewed and may be loosely defined as those in which the highest frequency is definitely not near the middle or near the ends of the distribution. Skewed curves in which the right-hand limb tapers off more gradually than the left-hand, hence in which the class with highest frequency is below the middle of the distribution, are said to be positively skewed, or skewed to the right. Similarly those with the left-hand limb longer and the class with highest frequency above the middle are nega-

[1] Curves may, however, be asymmetrical without the total frequencies above and below the middle being unequal.

EXAMPLE 18.—FREQUENCY DISTRIBUTIONS. STANDARD LENGTHS OF SAMPLES OF THE FLYING-FISH, *Parexocoetus brachypterus hillianus*, COLLECTED IN THE ATLANTIC DURING TWO DIFFERENT MONTHS
(Data from Bruun 1935)

Standard lengths, cm.	Frequencies	
	A. November	B. April
1.5– 1.9	1	0
2.0– 2.4	1	0
2.5– 2.9	0	0
3.0– 3.4	0	15
3.5– 3.9	0	35
4.0– 4.4	0	13
4.5– 4.9	0	13
5.0– 5.4	0	5
5.5– 5.9	0	3
6.0– 6.4	0	0
6.5– 6.9	0	2
7.0– 7.4	1	1
7.5– 7.9	1	0
8.0– 8.4	0	1
8.5– 8.9	1	1
9.0– 9.4	2	0
9.5– 9.9	2	0
10.0–10.4	21	0
10.5–10.9	46	1
11.0–11.4	19	0
11.5–11.9	4	0
12.0–12.4	0	0
12.5–12.9	0	1

The distribution for November is skewed to the left, or negatively, since the class with highest frequency is well above the middle class; it is the nineteenth of 23 classes. The distribution for April is skewed to the right, or positively, because the class with highest frequency is far below the middle class; it is the second of 20 classes.[1] The skewing in this instance is so well marked that it might, especially for the April sample, be considered an example of extreme rather than of moderate skew.

[1] The biological significance of the skewing and its reversal at different seasons in this example are clearly related to the existence of a restricted spawning season and to changing growth rates. If it were possible to gather a sample of these fishes all of the same age, the curve would almost surely be symmetrical. As in many cases of marked asymmetry this is probably thus due to heterogeneity of the sample. It is not a characteristic of length distribution in specimens essentially similar in everything but length.

tively or left skewed. Interesting examples (Fig. 6) of the two types are furnished by the data in Example 18 on samples of the same subspecies of fish caught at different times in the year.

Every gradation from no skewing to extreme skewing is encountered. Indeed, as will be shown in Chap. VII, a slight degree of skew, usually to the right, is to be expected with many zoological variates and may usually be ignored. A large skew, however, demands recognition and explanation either as a character of the variate or a result of peculiarities in the sample.

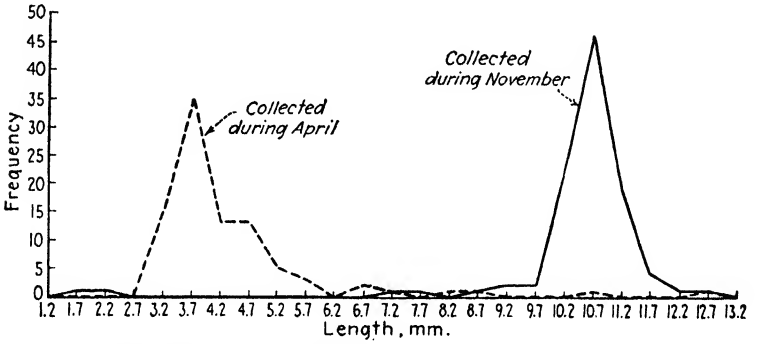


FIG. 6.—Moderately but significantly asymmetrical frequency distributions. Lengths of the fish *Parexocoetus brachypterus hillianus* (data of Example 18). The polygon in continuous outline represents the November catch and is skewed to the left. The polygon in broken outline represents the April catch and is skewed to the right.

It is possible to measure the degree of skewing and to estimate its significance (see Chap. VII).

The most extreme form of skew is one in which a terminal class (in practice usually the lowest) has the highest frequency. Such a curve with its high point at one end and hence with the frequencies dropping at first rapidly and later increasingly slowly to zero is usually called a J-shaped distribution.[1] It does not follow Quetelet's law; and with most characters of animals, especially continuous variates, it will be found that an apparently J-shaped distribution is simply a humped very asymmetrical distribution with the class intervals too large. Thus if the April distribution of Example 18 were grouped with interval 2.0 instead of .5 it would appear as in Example 19.

[1] Although it is almost always reversed, with the high point to the left, and does not rise again at the other end, as would a reversed letter J.

EXAMPLE 19.—REGROUPING OF EXAMPLE 18

Length	Frequency
3.0– 4.9	76
5.0– 6.9	10
7.0– 8.9	3
9.0–10.9	1
11.0–12.9	1

This appears to be a J-shaped distribution but is not really so, since splitting up the first class into much smaller classes would show the frequencies falling steadily to 0 at this end of the series as they do more slowly at the other end.

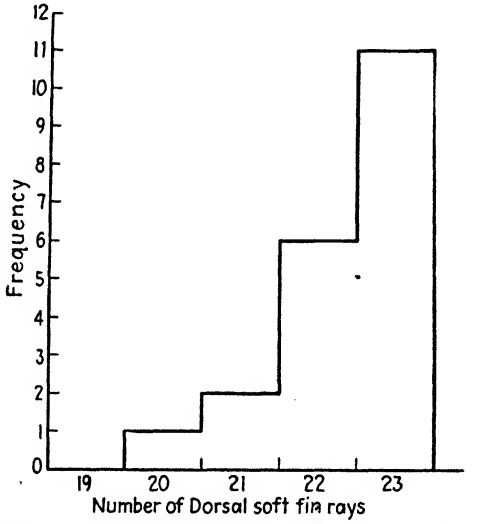


FIG. 7.—An extremely asymmetrical or J-shaped frequency distribution. Number of dorsal soft fin rays in the fish *Caranx melampygus* (data of Example 20B). Such a left-skewed J-shaped distribution is less common than one skewed to the right (*e.g.*, Fig. 11).

In zoology, true J-shaped distributions are usually of discontinuous variates, anatomical or otherwise, although this is not the commonest form even for discontinuous variates. Simply from the logical and biological aspects of the problem it is usually obvious when a terminal class may be expected to have the highest frequency and hence when a J-shaped curve is the probable pattern of the frequency distribution. Such variates are either structures, events, etc., not normally occurring but occasionally observed in varying number, or more rarely, variates in which the usual value is never exceeded or is always reached

EXAMPLE 20.—J-SHAPED DISTRIBUTIONS

A. Number of times individual female snowshoe-hares were live-trapped (data from Aldous 1937)

No of times trapped	No. of hares
1	365
2	163
3	103
4	58
5	33
6	14
7	6
8	4
9	1
10	0
11	0
12	0
13	1

B. Number of dorsal soft fin rays in the fish *Caranx melampygus* (data from Nichols 1935)

No. of fin rays	No. of fishes
20	1
21	2
22	6
23	11

Obviously more individuals will be trapped once only under ordinary conditions than will be trapped two or more times, so that a J-shaped distribution such as actually occurs in A is to be expected. This cannot be made into a moderately skewed distribution by splitting the classes, since an animal cannot be trapped a fractional number of times.[1]

As it stands, B is a J-shaped distribution skewed to the left. The species usually has 23 such rays, and as far as these data show it never has more but may have less. It is probable, in this and in most analogous cases, that the J-shape is illusory, however, and is only a chance result in a small sample. It is highly probable that further search would result in finding some individuals with more than 23 rays; for most distributions of this sort are only moderately skewed, and there is no obvious reason why this should be extremely skewed. Most J-shaped distributions in which the class with highest frequency is not 0 (or 1 in cases like example A), would probably lose the J-shape if a very large total frequency were available; and this pattern in such a case is distinguished from the sort of asymmetrical distributions previously discussed only by being still more skewed.

[1] That is, the variate is discontinuous with integral values. Nor can a slope downward to 0 be added, for as it stands the frequency of the class 0 comprises the number of animals not trapped at all (unknown but presumably larger than those trapped once). Also since the subject of enquiry is retrapping, an equally or more valid form would be to use number of times retrapped as the variate, when class 0 would have frequency 365, class 1 frequency 163, etc.

or exceeded. Example 20 gives two such distributions (see also Figs. 7 and 11).

U-shaped distributions, rising to a peak at both ends, are of rare occurrence with any type of data and are almost nonexistent for variates used in zoology. Some of the distributions of the special sort mentioned at the end of this section may approximate the U-shape, however. In other instances an apparently U-shaped distribution may result from incorrect sampling or similar errors in technique resulting in the fusion of two curves strongly skewed in opposite directions. Thus the combined data for the whole sample of *Parexocoetus brachypterus hillianus* from which the skewed distributions on page 65 were drawn (from Bruun 1935) could be presented in this form:

Standard length	Frequency
1.5– 4.4	66
4.5– 7.4	28
7.5–10.4	39
10.5–13.4	83

As it stands, this is U-shaped; but this is because most of the specimens were collected in two months, November and April, when their distributions were strongly and opposingly skewed. If the specimens had been collected in about equal numbers at all times of the year or if only those collected at one time were counted, the distribution would not be U-shaped.[1] Moreover if the class intervals were made smaller, as they should be, it would be obvious that this is not a U-shaped curve but two moderately skewed curves. An apparently U-shaped distribution of zoological variates is usually an indication of faulty procedure or of heterogeneity of the materials included.

There are a few zoological variates that tend to fall into a curve more complex than those already mentioned. Generally the presence of two high points on a curve is a sign that the sample is heterogeneous and that the curve is really composed of two or more curves that should, if possible, be separated. An exception to this rule is the possibility that the variate may naturally take only low or high values, a rarity with zoological materials. Thus the Patagonian rhea frequently lays one or a few eggs in isolated spots but otherwise tends to concentrate a

[1] Bruun presents his data correctly, and we have recast it in incorrect form only as an example.

large number of eggs in one spot, a crude nest. Figures on this do not seem to be available; but the observed habit suggests a hypothetical distribution of this general form, as in Example 21.

EXAMPLE 21.—HYPOTHETICAL DATA ON SETS OF RHEA EGGS

No. of eggs	No. of sets
1– 5	20
6–10	5
11–15	10
16–20	15
21–25	20
26–30	10
31–35	5

The curve begins high, falls, then rises to a second apex, then falls again. Even such cases, however, may properly and most conveniently be considered as composed of two separate curves, in the example, a J-shaped curve of sets not in nests and an approximately symmetrical curve of sets in nests.

THE NORMAL CURVE

The complete specification of a distribution curve would involve giving a mathematical formula with the values of the variate X and the frequencies f as the variables and with such numerical constants as were needed. To do this for every distribution would, however, involve an enormous amount of difficult and essentially useless labor. The simplest and most useful method is to relate the various distributions as far as possible to some mathematical curve that they approximate, that has a logical bearing on distributions, and that has useful and readily apprehended properties.

By far the most important of such standard curves is that called the normal curve, Gaussian curve,[1] or Laplace's normal curve.[2] This curve was discovered in considering the theory of probabilities, and the way in which it was developed can most simply be explained by the classic example of coin tossing.

If a single coin is tossed, supposing the coin itself and the method of tossing to be without any bias, there are two ways in

[1] K. F. Gauss (1777–1855) was a German mathematician and geodesist who published on numerical series, including that from which the normal curve is derived.

[2] P. S. Laplace (1749–1827) was a French astronomer who studied the theory of probabilities and laid the foundation on which many statistical procedures are based.

which it can fall, head or tail, of which one is as probable as the other. The probability of each can be expressed by a fraction with the number of possibilities (2 in this case) as the denominator and the number agreeing with a set condition (that it be a head or that it be a tail, 1 in each case) as numerator. Thus the probabilities with a single coin are

No. of heads = value of variate (X)	Probability, or theoretical relative frequency
0	$\frac{1}{2}$
1	$\frac{1}{2}$

If two coins are tossed, there are four possibilities: the first can be a head and the second likewise a head; first head, and second tail; first tail, and second head; first tail, and second tail. The theoretical distribution is now:

No. of heads	Relative frequency
0	$\frac{1}{4}$
1	$\frac{2}{4}$
2	$\frac{1}{4}$

Since these are relative frequencies, they may be multiplied by the denominator (or only the numerators read) to give a theoretical integral distribution, making the values just given 1, 2, and 1. In this way one can build up a whole series of distributions, the first few of which are as in Example 22. They are presented in graphic form in Fig. 8.

EXAMPLE 22.—THEORETICAL INTEGRAL FREQUENCIES. COIN TOSSING

No. of heads	With 1 coin	With 2 coins	With 3 coins	With 4 coins	With 5 coins	With 6 coins	With 7 coins
0	1	1	1	1	1	1	1
1	1	2	3	4	5	6	7
2	..	1	3	6	10	15	21
3	..	..	1	4	10	20	35
4	..	..	..	1	5	15	35
5	..	..	..	..	1	6	21
6	..	..	..	..	..	1	7
7	..	..	..	..	..	..	1

Of course if 1 coin is tossed only 2 times, we will not always get 1 head and 1 tail; and if 2 coins are tossed only 4 times, the result

frequently will not be to get 2 heads once, 2 tails once, and 1 head and 1 tail twice. These are, however, the proportions in which these results tend to be distributed; and if the experi-

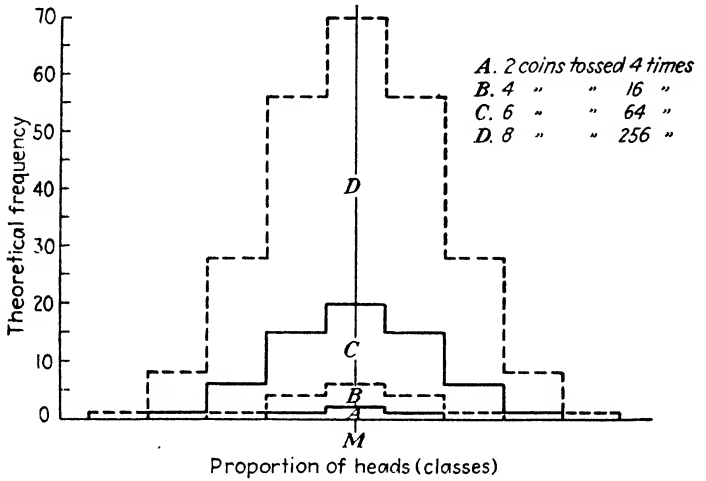


FIG. 8.—Changes in symmetrical binomial probability histograms as the number of possible different outcomes (classes) and the number of possibilities of occurrence (frequencies) are increased so as to keep the chances of occurrence integral. The diagrams are centered so that the middle classes, those representing the most probable outcome, are in the same central position in each case. The chances represented are those of obtaining specified numbers of heads by tossing different numbers of coins, the middle class in each case representing the chances of obtaining half heads (and half tails). *A*, 2 coins tossed 4 times. *B*, 4 coins tossed 16 times. *C*, 6 coins tossed 64 times. *D*, 8 coins tossed 256 times. (See Example 22.)

ment is made a great many times, these relative values will be approximated, as shown by experiment in tossing 4 coins 80 times, the results of which are given in Example 23.

EXAMPLE 23.—OBSERVED AND CALCULATED FREQUENCIES IN TOSSING 4 COINS 80 TIMES

No. of heads	Observed frequency	Calculated frequency
0	4	5 (1 × 5)
1	23	20 (4 × 5)
2	32	30 (6 × 5)
3	15	20 (4 × 5)
4	6	5 (1 × 5)

Although the observed frequency, a real thing, almost never exactly corresponds with the calculated frequency, a theoretical thing, the latter and the general type of distribution on which it is based are the best general description of such observations. This is the ideal toward which the observations tend, and they approach it closer and closer as more and more observations are made. The calculation has the advantage of giving an expression valid for an infinitely large number of observations, and for general descriptive purposes this is more enlightening than the more irregular data of a few observations.

All the theoretical frequency distributions given above have about the same form. They have a high midpoint,[1] and they fall away on each side of this with perfect symmetry. The variate in this example is discontinuous, and the frequencies hence always go up by definite steps. A graph of them would never be a real curve, but the distribution approaches a curve as a limit. If the number of coins used is increased indefinitely, the number of classes, in this case always one more than the number of coins for any one toss, also increases. If now the graphic representation of the class interval is made smaller and smaller, a point will be reached where the eye cannot distinguish the interval, and the boundary of the graphic distribution will appear to be a curve. This curve is the limit of distributions based on the laws of probability when the chances of an event's occurring or not occurring (as of a head turning up or not turning up when a coin is tossed) are equal, and it is called the normal curve. A distribution that approaches it as a limit as the frequency is increased is a normal distribution (see Fig. 9).

It was Quetelet's discovery that this curve, originally derived mathematically to represent theoretical probabilities, also represents the way in which many zoological characters are distributed in nature. It was found empirically by observation and comparison that numerous zoological frequency distributions do in fact closely approach normal curves if the frequencies are made large enough. It is also theoretically likely that they should do so, although the factors governing such fluctuations are so complex that theoretical demonstration of exact equivalence to a

[1] When the number of classes is even, the midpoint falls between the two middle classes, which are therefore equal; but this does not change the essential character of the distribution.

normal curve is not very satisfactory. It does not matter to the zoologist whether the data really exactly represent a normal curve or something so like one that the difference has no effect on the zoological conclusions to be drawn.

The normal curve is of fundamental zoological value in two ways: as an approximation of zoological frequency distribution patterns and as a representation of probabilities under certain conditions. Its importance in both these respects is so great

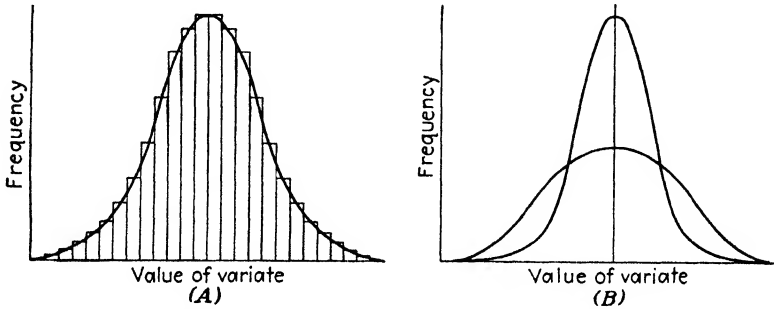


FIG. 9.—The normal curve. *A*, the normal curve as approximation or limit of the symmetrical binomial probability distribution. The histogram represents a symmetrical binomial with 24 classes. The curve is normal and approximates this histogram in area and form. If the classes of the binomial and the included frequencies were increased indefinitely, the outline of the histogram would approach that of the curve as a limit. *B*, two curves, both normal and with equal means, but one (the higher) with less and the other (the lower) with greater dispersion. (Note that the higher curve is not leptokurtic and the lower curve not platykurtic: lepto- and platykurtic curves are more or less peaked than the most nearly corresponding normal curves, whereas both these curves are normal curves.)

that a special chapter (Chap. VII) is later devoted to the properties of the normal curve and to how these are used. Mathematical consideration of the curve is therefore deferred to that chapter.

In considering a normal curve as an approximation of an actual distribution, it is necessary first to specify the particular normal curve that most nearly approximates the distribution. This is done by three numbers calculated from the data: the average value of the variate, which shows where the center of the curve will lie on the scale; the frequency, which shows how high the curve will be; and a more complex figure, the standard deviation (Chap. VI), which shows how rapidly the curve falls off on each side of the midpoint.

A given distribution will usually differ in some degree from one truly normal, even though the resemblance is close enough for most practical purposes. The most important of such deviations are that a distribution may be sharper or flatter than the normal, a property called kurtosis, or that it may be asymmetrical, a property called skewness. Measurements of kurtosis and skewness are discussed in Chap. VII. For the normal curve, both kurtosis and skewness are zero; and if these are slight, it is usually not worth while to go through the complex process of measuring them, and the nearest normal curve may be taken as an adequate approximation.

THE BINOMIAL DISTRIBUTION

If the probability that an event will occur is represented by p, the probability that it will not occur by q, and the number of trials made, or opportunities for occurrence existing, by n, then the probability of occurrence of the event 0, 1, 2, 3, etc., up to n times is given by the algebraic multiplying out or expansion of the expression

$$(p + q)^n$$

This expression is a binomial, it is expanded by the binomial theorem, and the series resulting from the expansion is a binomial distribution. The coin-tossing series given in the last section are in fact binomial distributions in which p and q are equal; both $= \frac{1}{2}$. The normal curve is the limit approached by such a series when n is increased indefinitely, and the normal distribution is in this sense a special case of the binomial distribution.[1] When p and q are equal, the limit of the binomial series is a

[1] The binomial distribution was in fact known first and the normal distribution developed from it. The theorist and mathematician insist on keeping binomial and normal distributions in separate categories. The former deals only with discrete steps, hence is applicable only to discontinuous variates; and the latter is a continuous curve and hence can exactly correspond only to a continuous distribution. Although correct in theory, this distinction has little bearing on zoological practice. A real sample of a continuous variate is never continuous in fact, and its approach to the theoretical normal curve is no closer than that of a discontinuous variate. In the actual operations useful to a zoologist, who is employing statistics only as a means to an end, it is usually as valid to apply the statistics of the normal curve to a discontinuous as to a continuous distribution; and when this can be done, the results are more readily useful than would be the statistics of a binomial distribution.

normal curve. When this condition is nearly or quite fulfilled, it is usually proper as a practical convention to consider the distribution as normal, rather than as binomial.

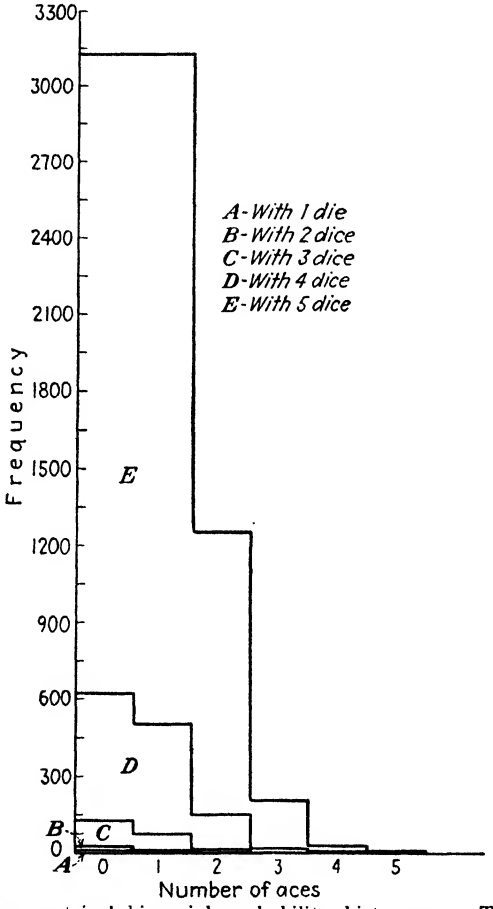


FIG. 10.—Asymmetrical binomial probability histograms. The histograms represent the chances of throwing a given number of aces with different numbers of dice and frequencies increased so as to make the chances of occurrence integral. *A*, with 1 die. *B*, with 2 dice. *C*, with 3 dice. *D*, with 4 dice. *E*, with 5 dice. (See Example 24, and compare Fig. 8.)

The practical use of the binomial distribution is in studying discontinuous variates when p and q are not equal and the difference between them is a significant element in the problem. Inequality of p and q introduces skew into the distribution, and

the stronger the inequality the greater the skew. As an analogy to the coin-tossing series, the number of times that a 1 will appear in throwing dice may be considered. Here there are six possibilities and only one of them fulfills the condition, so that the probability of occurrence p is $\frac{1}{6}$ and the probability of nonoccurrence q is $\frac{5}{6}$. Distributions for small numbers of dice are given in Example 24, and represented graphically in Fig. 10.

EXAMPLE 24.—BINOMIAL DISTRIBUTION RESULTING FROM DICE THROWING

Number of 1's appearing (X)	Theoretical integral frequencies				
	With 1 die	With 2 dice	With 3 dice	With 4 dice	With 5 dice
0	5	25	125	625	3125
1	1	10	75	500	3125
2	..	1	15	150	1250
3	..	..	1	20	250
4	..	..	...	1	25
5	..	..	...	...	1

These are strongly skewed because p and q are very unequal. Every gradation from perfect symmetry to still stronger skewing may appear in binomial distributions. The value of any one class frequency in such a distribution is given by the expression

$$N\left(\frac{n(n-1)(n-2)\cdots(n-X+1)}{X(X-1)(X-2)\cdots 1}\right)p^X q^{n-X}$$

where N = the total frequency.

n = the number of opportunities of occurrence (the number of dice in the example).

X = any possible value of the variate (the steps in the discontinuous distribution or, if the observations are grouped, the class midpoints).

p = the probability of occurrence.

q = the probability of nonoccurrence.

It happens only rather rarely with zoological data that it is more advantageous to use the binomial distribution than to use the normal distribution as an adequate approximation. Example 25, from data on human families, is one in which the binomial distribution is properly used and brings out facts less clear in the normal distribution:

EXAMPLE 25.—BINOMIAL DISTRIBUTION. NUMBER OF BOYS IN GERMAN FAMILIES OF 8 CHILDREN
(Data from Geissler; binomial distribution from Fisher 1936)

No. of boys in family	No. of families observed	Theoretical binomial distribution	Theoretical normal distribution
0	215	165.22	209.69
1	1,485	1,401.69	1,676.90
2	5,331	5,202.65	5,871.25
3	10,649	11,034.65	11,742.50
4	14,959	14,627.60	14,678.13
5	11,929	12,409.87	11,742.50
6	6,678	6,580.24	5,871.25
7	2,092	1,993.78	1,676.90
8	342	264.30	209.69

The normal distribution agrees fairly well with the observed facts, but it is noticeable that above the middle class the observed frequencies are considerably higher and below this most of them are lower than the normal frequencies. The divergence is so marked and so regular that it demands explanation. Now the normal distribution assumes that a child born to a German family is as likely to be a boy as a girl, in other words that the probability p of a child's being a boy is equal to the probability q of its not being a boy. The divergence of the observed frequencies suggests that this is false and that in fact p is greater than q. This conclusion could safely be drawn from the two distributions, and it is the only conclusion that could be. A study of actual figures on the sexes of all children born in Germany shows that boys are in fact more numerous than girls, hence that p is greater than q. If now the values of p and q thus determined are used, the theoretical binomial distribution shown above is obtained. This agrees better with the observed facts than does the normal distribution, but now that allowance has been made for the sex ratio in the whole country an additional conclusion can be drawn from these particular figures. It is noted that the more extreme classes are in excess, their frequencies larger than the calculated frequencies in the binomial distribution; and it can be shown by methods not pertinent in this chapter that this result is significant, is not due to chance. The new biological conclusion, not discernible without the use of the binomial series, is that large German families tend to have either more girls or more boys than does the general population, or in other words that large families tend to run to one sex instead of having the two sexes represented about equally.

THE POISSON SERIES

It has been shown that the normal curve is the limit of the binomial $(p + q)^n$ when p and q are equal and n is made increas-

ingly larger. When p and q become increasingly unequal, or in other words when one becomes very small, and n is made increasingly larger so that both np and nq remain finite quantities even though p or q becomes very small, then this binomial approaches another limit that also is of use in some zoological work. This limit is called the Poisson series,[1] and it is the theoretical distribution of a discontinuous variate the probability of the occurrence of which is very small but which does occur if enough observations are made.

The general expression for any one class frequency in a Poisson distribution is

$$Ne^{-M}\left[\frac{M^X}{X(X-1)(X-2)\cdots 1}\right]$$

where N = the total frequency.
e = the number 2.71828 (base of Napierian logarithms).
M = the mean value of the observations.
X = any possible value of the variate.

The series thus begins as follows:

Value of variate (X)	Class frequency (f)
0	Ne^{-M}
1	$Ne^{-M}(M)$
2	$Ne^{-M}(M^2/2)$
3	$Ne^{-M}(M^3/6)$
4	$Ne^{-M}(M^4/24)$

Most of the pertinent cases are distributions of discontinuous variates involving sampling operations. The most important use of the series is in showing how nearly the results of sampling agree with those theoretically to be expected, giving a check on the correctness and efficacy of the sampling. The biological importance of the series was first recognized in dealing with hemacytometer counts, and it finds considerable use in analogous biological problems. In more strictly zoological work, to which this book is devoted, the Poisson series is of relatively little use.

[1] Named for S. D. Poisson (1781–1840), a French mathematician who published on the series in 1837. It has since been independently discovered by several other statisticians.

The present brief section is devoted to it principally for reference on the few occasions when it will be helpful to zoologists, and the reader may prefer to omit close study of this section until such an occasion arises.

The use of the series and a relatively easy way to carry out the calculations involved are best shown by a simple example, which is provided by a sampling operation in paleontology closely analogous to hemacytometer counting (Example 26).

EXAMPLE 26.—A POISSON SERIES

Distribution of the number of specimens of the extinct mammal *Litolestes notissimus* found in each of 30 squares of horizontal quarry surface about one meter on a side (original data)

No. of specimens per square (X)	No. of squares (f)
0	16
1	9
2	3
3	1
4	1
5 and over	0

N = total number of squares = 30

M = mean number of specimens per square = .73

$$\log e = .43429$$
$$\log e^{-M} = .43429 \times (-.73) = -.31703 = 9.68297 - 10$$
$$e^{-M} = \text{antilog } 9.68297 - 10 = .482$$

From these data, the relative frequencies are calculated as follows:

$X = 0 \quad e^{-M} = .48$

$X = 1 \quad e^{-M}M = .482 \times .73 = .35$

$X = 2 \quad e^{-M}\left(\frac{M^2}{2}\right) = .482 \times .269 = .13$

$X = 3 \quad e^{-M}\left(\frac{M^3}{6}\right) = .482 \times .066 = .03$

$X = 4 \quad e^{-M}\left(\frac{M^4}{24}\right) = .482 \times .014 = .01$

$X > 5 \quad 1 - (\text{sum of preceding terms}) = 1 - 1.00 = .00^*$

The theoretical frequencies, those that the sample would take if it had this mean and total frequency and were perfectly distributed in a Poisson series, are calculated by multiplying each relative frequency by the total frequency ($N = 30$, in the example). The sum of the relative frequencies is always 1, which

*The value of the remainder of the series is always greater than zero, but the record .00 indicates that it is less than .005 and for the purposes of the present calculations is negligible.

EXAMPLE 27.—POISSON SERIES
Comparison of observed and theoretical frequencies

No. of specimens per square	Observed frequencies	Theoretical frequencies from Poisson series
0	16	14.4
1	9	10.5
2	3	3.9
3	1	.9
4	1	.3
5 and over	0	.0

In view of the small size of the sample, the agreement is remarkably close.

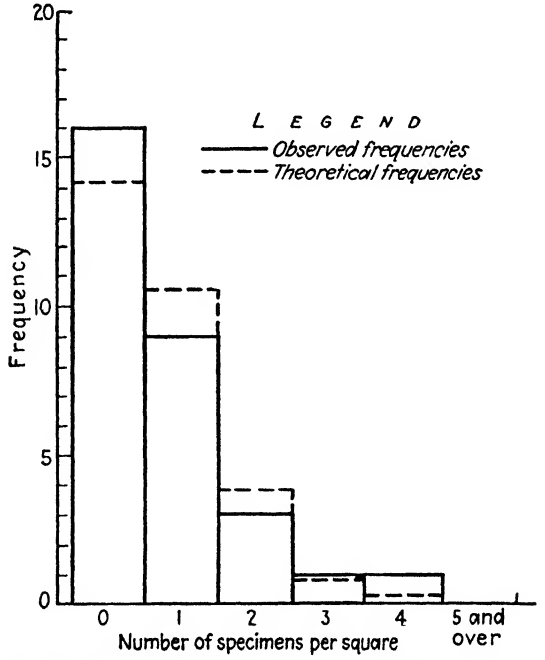


FIG. 11.—Histograms of a Poisson series and of an observed distribution approximating it in form. The solid lines represent numbers of specimens of the extinct mammal *Litolestes notissimus* found in each of 30 sq. m. of quarry surface (data of Example 26). The broken lines show an approximately equivalent Poisson distribution (see Example 27).

may be used, as above, to get a relative frequency for all higher classes at any point.[1]

The observed frequencies and the theoretical frequencies so obtained compare as in Example 27 and Fig. 11.

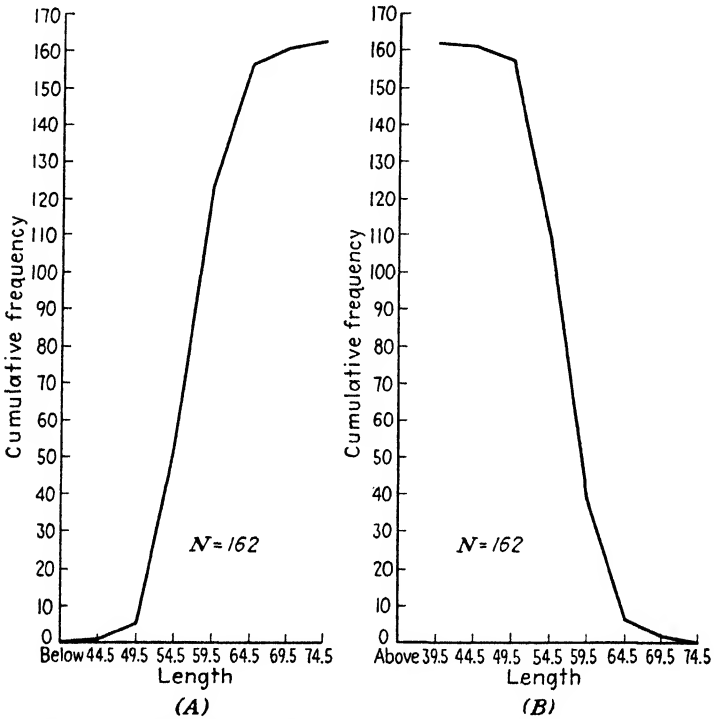


FIG. 12.—Graphs of cumulative distributions. Lengths of the fish *Pomolobus aestivalis* (data of Example 17 as rearranged in Example 28). *A*, frequencies cumulative from below. *B*, frequencies cumulative from above. These graphs approximate right- and left-handed ogive curves.

It will be noted that a Poisson series, in terms of relative frequencies, is fully determined by the mean so that a distribution

[1] For instance the low frequencies for $X = 3$ and $X = 4$ might have been considered unimportant in the example and total relative frequency for all values greater than 2 obtained by subtracting the sum of the first three frequencies from unity: $1 - (.48 + .35 + .13) = 1 - .96 = .04$. The total relative frequency above $X = 2$ is thus approximately .04, giving with $N = 30$ an absolute theoretical frequency of 1.2. Carrying out to two subsequent steps, as in the text, shows that this is distributed more precisely as .9 for $X = 3$ and .3 for $X = 4$.

that falls into such a series is adequately defined by this single number.[1]

The Poisson series just calculated is strongly asymmetrical, and this is true of any in which M, the value of the mean, is

EXAMPLE 28.—ORDINARY AND CUMULATIVE DISTRIBUTIONS
(Data from Example 17)

Ordinary		Cumulative			
		A		B	
Length, mm.	Frequency	Length	f	Length	f
40–44	1	Below 44.5	1	Above 39.5	162
45–49	4	Below 49.5	5	Above 44.5	161
50–54	47	Below 54.5	52	Above 49.5	157
55–59	71	Below 59.5	123	Above 54.5	110
60–64	33	Below 64.5	156	Above 59.5	39
65–69	4	Below 69.5	160	Above 64.5	6
70–74	2	Below 74.5	162	Above 69.5	2

The two cumulative distributions are approximately mirror images of each other and would be exactly so were the frequency distribution perfectly symmetrical. The figure given for the terminal class (the last class in cumulative distributions of type A and the first in those of type B) is always the total frequency. Note that classes are designated by limits and that care should be taken to use the real limits and not, as is usually done, the conventional limits. Cumulative distribution A in the example would usually have the classes listed as "below 45," "below 50," etc.; but this is wrong.

low. If M is 1 or less, the resulting distribution is J-shaped. With M low but above 1, the distribution is humped but strongly skewed. With high values of M, it becomes more nearly symmetrical and may approximate a normal distribution. In such a case, as for relatively symmetrical binomial distributions, it is often useful and simpler to compare the distribution with the normal rather than with the theoretically more correct Poisson series.

[1] It is also a peculiarity of this series that the variance (as defined in Chap. VI) is equal to the mean. Approximation of variance and mean thus suggests that a distribution may be in a Poisson series, although it does not prove this to be the case.

CUMULATIVE DISTRIBUTIONS: THE OGIVE

In the distributions previously discussed in this chapter, the frequency within each class is given. It is sometimes more convenient and may be more directly related to a problem in hand to give the total frequency above or below each class, and distributions given in this way are called cumulative. The distribution given on page 64 is used as an example, repeated in the usual form and given in the two types of cumulative distributions in Example 28.

The curve approached by normal distributions when placed in cumulative form is called an ogive, and its approximate shape can be judged from the accompanying graph (Fig. 12). While cumulative distributions are often useful as such and as graphic diagrams, their numerical characters are more easily and usefully calculated from the ordinary frequency distributions, and it is unnecessary to consider the mathematical properties of the ogive.

CHAPTER V

MEASURES OF CENTRAL TENDENCY

Most zoological variates are so distributed as to be more frequent near some one value and to become less and less frequent in departing from this value in either direction, a fact of experience summed up in Quetelet's principle. The normal distribution is one that complies fully with this principle; and so, in essentials, do many distributions that differ from the normal in various ways such as kurtosis or skewness. Clearly the most important things to observe and to measure in such a distribution are (1) the point around which the observations tend to cluster and (2) the extent to which they are concentrated around this point.

To say simply that observations tend to cluster around some particular values or, in more technical language, that they have a central tendency is an expression too vague to lead to an accurate measurement or estimate of the tendency without further qualification. The important feature may be a mean value of all the observations (and "mean" is also a loose word requiring specification), or the point at which the highest frequencies are observed, or the point where the middle observation lies, or even the middle of the observed range of values. In a given distribution these various values may all be different, and the different ways of looking at central tendency and its meaning involve several different constants used to designate or measure it. This general group of constants are all called averages. The most important of these are the arithmetic mean, the median, and the mode, discussed in that order below; and at least four other, relatively unimportant measures of central tendency are also in occasional use.

ARITHMETIC MEAN

The arithmetic mean is used so much more than any other that it is usually simply called the mean; other types of means are

specified in the rare instances when they are used. This is by far the most widely employed statistical constant, and everyone who uses numerical data at all has at some time calculated a

EXAMPLE 29.—CALCULATION OF THE ARITHMETIC MEAN OF THE LENGTH OF THE THIRD UPPER PREMOLAR OF THE EXTINCT MAMMAL *Ptilodus montanus*

(Original data)

Measurements (X)
3.0
2.8
3.4
3.2
3.0
2.9
2.6
3.3
3.1
2.9
2.9
3.0
2.8
2.9
2.7
2.9
3.1
2.8
3.0
3.1
3.0

$\Sigma(X) = 62.4$ (sum of measurements taken)

$N = 21$ (number of measurements taken)

$$M = \frac{\Sigma(X)}{N} = \frac{62.6}{21} = 2.97$$

mean.[1] The mean is an average obtained by adding together all the observed values and dividing by the number of observations. Its general formula is

$$M = \frac{\Sigma(X)}{N}$$

[1] Even zoologists who scorn statistics commonly use arithmetic means (which is what they usually imply by the loose term "average"), without being aware that these are the basic statistical constants and without bothering to think just what they indicate or what their properties are.

where M = the mean (arithmetic only).
Σ (Greek capital sigma) = the result of adding together all the data indicated by the symbol following it in parentheses.
X = any given value of the variate.
N = the number of observations made, the total frequency.

These symbols and a few others explained as they appear are used consistently throughout the present book, and learning this shorthand notation greatly simplifies not only the explanation of these processes but also their use.[1]

EXAMPLE 30.—SAME DATA AS IN EXAMPLE 29, RECAST AS A FREQUENCY DISTRIBUTION AND MEAN BASED ON THIS

Measurements (class midpoints) (X)	Frequencies (f)	Frequencies times class midpoints (fX)
2.6	1	2.6
2.7	1	2.7
2.8	3	8.4
2.9	5	14.5
3.0	5	15.0
3.1	3	9.3
3.2	1	3.2
3.3	1	3.3
3.4	1	3.4
	$\Sigma(f) = N = 21$	$\Sigma(fX) = 62.4$

$$M = \frac{\Sigma(fX)}{N} = \frac{62.6}{21} = 2.97$$

An instance of the simplest possible calculation of a mean is given in Example 29.

Even with such a short series of measurements, the duplications and irregularity make this a clumsy and time-wasting

[1] Certain of these symbols, like Σ, are used in the same way by almost all authors. Regarding others there is wide divergence of usage. The mean, for instance, is often designated by $\bar{X}$. Any system is adequate if consistent in the book used, but unfortunately this lack of uniformity requires caution when more than one text or reference work is employed. We have tried to select symbols simple typographically, as nearly self-explanatory as possible, and in as wide use as any others.

method. If the number of observations were large, say several hundred, it would be practically impossible, without the use of an adding machine. It is much better to place the observations in a frequency distribution and then to calculate the mean with a modified but arithmetically equivalent formula

$$M = \frac{\Sigma(fX)}{N}$$

in which f designates a class frequency. The same calculation made in this way is shown in Example 30.

When the number of classes is small as in this example (only 9 classes), it is better to work from the original measurements. If the number of classes is very large (generally not unless it exceeds 20 or 25), it may, however, be advisable to increase the class interval and group the data more broadly. In such a case the operation is carried out by the same formula, remembering that X is the true class midpoint. This is done in Example 31 with the same data as that of the last two examples for the sake of comparison although in ordinary practice the secondary grouping would not be justified in this case.

EXAMPLE 31.—SAME DATA AS IN LAST TWO EXAMPLES, GROUPED WITH INTERVAL .2 MM. FOR CALCULATION OF THE MEAN

Class	Class midpoint (X)	Class frequency (f)	Frequency times class midpoint (fX)
2.6–2.7	2.65	2	5.30
2.8–2.9	2.85	8	22.80
3.0–3.1	3.05	8	24.40
3.2–3.3	3.25	2	6.50
3.4–3.5	3.45	1	3.45
		$\Sigma(f) = N = 21$	$\Sigma(fX) = 62.45$

$$M = \frac{\Sigma(fX)}{N} = \frac{62.45}{21} = 2.97$$

The mean here obtained differs very slightly from that based on the original data because of inaccuracies introduced by the grouping.[1]

[1] As pointed out in considering grouping, many statisticians would call these class midpoints 2.7, 2.9, etc. This would make the mean 3.02, far from being as close an approximation of the more accurate figure based on the original measurements as is obtained by using the real class midpoints.

Calculation of the mean from grouped data depends on the assumption that each class midpoint does not differ significantly from the mean value of the observations that fall in the class. This assumption is more likely to be true with small class intervals than with large, because the midpoint cannot differ from the mean for the class by more than half the class interval. It is also more likely to be true with high class frequencies than with low, because if many observations enter into a single class they are likely to be well scattered in it and hence to have a mean value near its midpoint. Grouping always involves some inaccuracy, even when it is only the grouping of the original measurements of a continuous variate; but if these sources of inaccuracy are kept in mind, its extent seldom is great enough to affect the final result significantly. If there is any question about this, it is invariably true that the mean as calculated from grouped data is within one-half class interval of the true mean, and usually it will be within one-tenth class interval or even less. Despite the unduly large interval in the example just given, the calculated mean is probably not more than .01 from the true mean, and the class interval is .2.

When only the mean is to be calculated, the use of the formula $\Sigma(fX)/N$ is often sufficiently easy and may be the most satisfactory. In calculating certain measures of dispersion, however, it is necessary to obtain deviations; and in such cases another method of calculating the mean may be more convenient.[1] This is based on assuming the mean and then calculating the difference between the assumed mean and true mean. The formula is

$$M = A + \frac{\Sigma(fd_A)}{N} = A + c$$

in which A is the assumed mean and d_A is the difference between the assumed mean and the class midpoint. Any number may be used as the assumed mean, and it does not matter how far it later proves to be from the real mean. The expression $\Sigma(fd_A)/N$

[1] The method about to be explained is usually called the "short method" and that based on $\Sigma(fX)/N$ the "long method." Since the so-called short method is apparently less direct and more complex, these terms may not seem well applicable. Aside from convenience, when a whole set of statistical constants is to be calculated, however, the arithmetic involved in the short method is often easier, especially with large samples.

is the correction for the assumed mean and is often designated by c (for "correction"). Example 32 shows the use of this formula.

EXAMPLE 32.—CALCULATION OF THE MEAN BY THE DEVIATION METHOD, LENGTHS OF *Pomolobus aestivalis*
Data as in Example 17

Classes	Class midpoints	Frequencies (f)	d_A	fd_A
40–44	42	1	−15	− 15
45–49	47	4	−10	− 40
50–54	52	47	− 5	−235
				−290
55–59	57	71	0	+235
60–64	62	33	+ 5	+165
65–69	67	4	+10	+ 40
70–74	72	2	+15	+ 30
		$N = 162$		$\Sigma(fd_A) = -55$

A (arbitrarily chosen) = 57

$$c = \frac{\Sigma(fd_A)}{N} = \frac{-55}{162} = -.34$$

$$M = A + c = 57 - .34 = 56.66$$

For convenience in tabulation, lines are drawn above and below the value of A, and in the fd_A column, instead of that value (which is, of course, zero) the sums of negative and positive values of fd_A are noted here. The column value $\Sigma(fd_A)$ is the algebraic sum of these.

To demonstrate that the value taken as A has no influence on the outcome of the calculations, the same operation is shown in Example 33 with A taken at a different point.

The logic of this superficially esoteric method is quite simple. From the definition of the arithmetic mean, it follows that the sum of all the individual deviations from the mean is always 0. Hence if A were guessed at the real mean, $\Sigma(fd_A)$ would be 0. The farther A is from M, the larger this sum $\Sigma(fd_A)$ will be, either positively or negatively depending on its direction from M; and this increase is likewise proportional to N, since $N = \Sigma(f)$. Thus $\Sigma(fd_A)/N$ is always the distance between A and M.

This method may be made still easier arithmetically by omitting the use of class midpoints and taking the deviations in

EXAMPLE 33.—SAME AS THE LAST EXAMPLE, WITH A DIFFERENT ASSUMED MEAN

Classes	Class midpoints	f	d_A	fd_A
40–44	42	1	−25	− 25
45–49	47	4	−20	− 80
50–54	52	47	−15	− 705
55–59	57	71	−10	− 710
60–64	62	33	− 5	− 165
				−1,685
65–69	67	4	0	+ 10
70–74	72	2	+ 5	+ 10
		N = 162		$\Sigma(fd_A)$ = −1,675

A = 67
c = −1,675/162 = −10.34
M = 67 − 10.34 = 56.66

This gives the same value for the mean; but it is clear that it is not so easy arithmetically since higher numbers have to be multiplied to get fd_A. For this reason it is better to assume A as the midpoint of the class with the highest frequency if this is reasonably near the center of the distribution. For the same reason it is better to assume A at a midpoint, although it may be taken as any number whatever.

terms of class intervals, later applying a correction for this. In this case, using the formula

$$M = A + c$$

as before, a new c is used, from the formula

$$c = \frac{\Sigma(fd_A)}{N} i$$

where d_A = deviation in terms of class intervals.
i = the class interval.

The same problem calculated in this way is given in Example 34.

This seems offhand to be a long way around to reach the desired result; but (aside from the relationship to calculations discussed in the next chapter) the arithmetic in the last example is really simpler than in any other way of obtaining the mean, and the chances of errors are fewer.

It is valid to base a mean on two or more other means; but this requires logical consideration of what is being done and in most cases involves a correction, or weighting, if the means are derived from distributions with different total frequencies. If one wishes to calculate the mean length of species within a genus, this can be done by the ordinary formula $M = \Sigma(fX)/N$ in which X is a

EXAMPLE 34.—SAME DATA AS PRECEDING PROBLEM, WITH MEAN CALCULATED BY DEVIATIONS IN TERMS OF CLASS INTERVALS

Classes	f	d_A (in class intervals)	fd_A
40–44	1	−3	− 3
45–49	4	−2	− 8
50–54	47	−1	−47
			−58
55–59	71	0	+47
60–64	33	+1	+33
65–69	4	+2	+ 8
70–74	2	+3	+ 6
	$N = 162$		$\Sigma(fd_A) = -11$

$i = 5$

A = midpoint of class 55–59 = 57

$$c = \frac{\Sigma(fd_A)}{N} i = \frac{-11}{162} \times 5 = -.34$$

$$M = A + c = 57 - .34 = 56.66$$

value of the mean for any one species and N is the total number of species. If, however, one wishes to calculate the average size of the (available) individuals within a genus or if one wishes to average the means for two or more different samples of a single species, the result will not be valid unless the means are weighted according to the frequency of the sample on which each was based. This can be done by multiplying each mean by its corresponding frequency (N), adding the figures thus obtained, and dividing by the sum of the frequencies, which can be expressed in formula by

$$M = \frac{N_1M_1 + N_2M_2 + N_3M_3 \cdots}{N_1 + N_2 + N_3 \cdots}$$

Since N for the whole combined sample is equal to

$$N_1 + N_2 + N_3,$$

etc., and since $N_1M_1 = \Sigma(f_1X_1)$, and $\Sigma(fX)$ for the whole sample is $\Sigma(f_1X_1) + \Sigma(f_2X_2) + \Sigma(f_3X_3)$, etc., it is evident that this formula is merely a convenient form for the special case exactly equivalent to the usual formula $M = \Sigma(fX)/N$.

The mean has many properties that make it the most widely used and most useful of all averages. Its wide use and the relative simplicity and ready comprehensibility of the concept are in themselves advantages. The computation is simple, and it may be calculated, by one method or another, from few data. It can be treated algebraically, for instance, a mean may be derived from a series of other means,[1] which is not true of some other averages in current use. It also has several advantages or properties the significance of which involves operations discussed in later chapters: *viz.*, that the sum of the deviations about it is zero, that the sum of the squares of these deviations is less than for any other point, and that its standard error is less than for the median. It also has the property, sometimes advantageous but more often a disadvantage, that it is strongly affected by extreme values among the observations.

Widespread use of the mean as if it were in itself a sufficient summary of the data and as if means were directly and simply comparable betrays inadequate comprehension of just what a mean is and how means are validly compared. For purposes of comparison, a mean drawn from a limited number of observations is merely an estimate of what the mean would be if all possible observations had been made. For instance, a mean length for 10 specimens of a given species is an estimate of the mean length for all individuals belonging to that species. This is necessarily the logical background, since explicitly or implicitly the purpose in such a case is to characterize and compare species, not simply chance groups of specimens.

A calculated mean is never exactly equal to the general mean of which it is an estimate. The larger the sample, the closer the estimate; so that N, total frequency, should be given for every mean and should enter into any comparisons of means.

[1] But generally when this is done, the means must be weighted according to the frequencies represented by each (see preceding paragraphs).

The accuracy of estimate also depends on the dispersion of the distribution. Obviously the estimate will be closer if most individual observations are near the mean than if the observations are widely scattered. The real significance of the mean as calculated thus also involves some measure of dispersion (see Chap. VI), and such a measure should also be given for each mean and must also enter into a valid comparison of means.

Although the mean is the most useful and the basic statistic, it is thus not adequate alone for the purposes for which it is commonly used.

MEDIAN

The median is the value of the middle observation in a frequency distribution. If it is to be recorded in units not smaller than the class interval, it can be read directly from the frequency distribution with little or no calculation. For instance, in Example 30 (page 87), N is 21; so the middle observation is the eleventh from either end, and this is at once seen to be in the class with midpoint 3.0. The class interval is .1; so this value of the median is accurate to the first decimal place. If the same data are grouped with interval .2 (Example 31, page 88), the median is seen to lie in the class 3.0–3.1; *i.e.*, it is somewhere between 2.95 and 3.15, but it cannot be read directly to the first decimal place.

If, as is usually desirable, the median is to be calculated to more places than the original data or in smaller units than the class interval, it is necessary to estimate its position within the class by interpolation. In Example 30, the middle value is the eleventh, which is the lowest of the 5 in the class with midpoint 3.0 and implied true limits 2.95–3.05. For purposes of calculation it is assumed that the observations within the group are evenly distributed. Thus with $f = 5$, as here, it may be assumed that the middle one of these 5 is at the class midpoint, that those on each side are at a distance equal to $\frac{1}{5}$ class interval above and below the midpoint, and that the last two are at distances of $\frac{2}{5}$ class interval above and below the midpoint. This leaves a distance of $\frac{1}{2} \times \frac{1}{5}$ or $\frac{1}{10}$ class interval between the most divergent observations and the true class limits. Such considerations lead to the following general formula for interpolation in finding a median:

$$\text{Median} = L_l + \frac{(\frac{1}{2} + n - 1)i}{f} = L_l + \frac{(n - \frac{1}{2})i}{f}$$

where L_l = the true lower limit of the class in which the median lies.

n = the serial number of the desired observation within the class.

i = the class interval.

f = the frequency of the median class.

In Example 30 the median class has the true limits 2.95–3.05. Hence $L_l = 2.95$. The median observation is the eleventh, and there are 10 below the median class; so the median observation is the first in that class, and $n = 1$. Also $i = .1$, and $f = 5$. The median is then

$$\text{Median} = 2.95 + \frac{(1 - \frac{1}{2}) \times 1}{5} = 2.95 + .01 = 2.96$$

Calculation from the same data grouped with interval .2 (Example 31) gives

$$L_l = 2.95 \qquad n = 1 \qquad i = .2 \qquad f = 8$$

$$\text{Median} = 2.95 + \frac{(1 - \frac{1}{2}) \times .2}{8} = 2.95 + .01 = 2.96$$

The preceding formula assumes that the median is found by counting up from the lower end of the distribution. There is no particular advantage in doing so, and the same result can be achieved in counting down from the upper end. In this case the formula is

$$\text{Median} = L_u - \frac{(n - \frac{1}{2})i}{f}$$

where L_u = the upper limit.

The data of Example 30 would then be

$L_u = 3.05$ $\quad n = 5$ (in counting down, there are 6 observations before the median class and hence one wants the fifth in that class)

$i = .1$ $\quad f = 5$

$$\text{Median} = 3.05 - \frac{(5 - \frac{1}{2}) \times 1}{5} = 3.05 - .09 = 2.96$$

The result is the same as before, checking the correctness of the procedure[1] (see Fig. 13).

In the example thus far treated, the total frequency 21 is odd, so that there is a middle observation. When not obvious on inspection, the middle observation in such cases can be determined by the expression $(N + 1)/2$, N being the total frequency.[2]

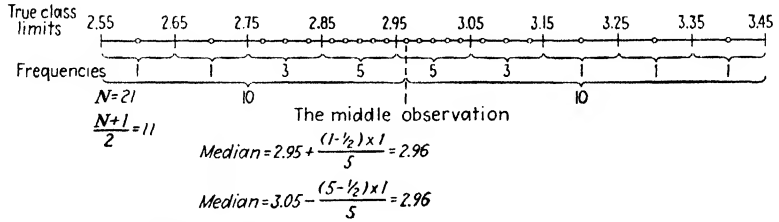


FIG. 13.—Diagram showing the calculation of the median. Length of the third upper premolar of the extinct mammal *Ptilodus montanus* (data of Example 29). Measurements were to .1 mm., implying the true primary class limits shown. The positions of the measurements within each primary class are not, in fact, known; but in calculating the median it is assumed that observations in each class are arranged symmetrically and evenly about the midpoint, as represented in the diagram. As there are 21 observations, the middle observation is the eleventh. The position of this observation under the assumptions made is the median and is given by the formulas, in one case by calculating from the lower and in the other from the upper end of the distribution.

If the total frequency is even, there is not in fact any one middle observation, but there are two, one above and one an equal distance below the middle of the series. Thus there cannot really be a median according to the definition of that average. The

[1] Here again we find ourselves in disagreement with many statisticians. The formulas for the median in most books on statistics are Median $= L_l + \frac{n}{f}i$ or $L_u - \frac{n}{f}i$. These formulas do not agree with the assumption that the observations are evenly distributed within the class. The first formula moves them all up toward the upper limit and gives too high a value for the median and the second moves them downward and gives too low a value. That this is true is easily seen in Fig. 14 and also from the fact that the two formulas do not give the same result, as do the correct formulas given in our text. For Example 30, the first of these erroneous formulas gives 2.97, and the second gives 2.95. The correct result is 2.96. As in other cases, many statisticians also use incorrect class limits.

[2] Some statisticians use $N/2$, but this is obviously wrong. The middle observation of 21, for instance, is $(21 + 1)/2 = 11$. There are 10 observations above the eleventh and 10 below, so that the eleventh is in fact the middle as a matter of sheer common sense. The formula $N/2$ gives $21/2 = 10.5$, leaving 10 observations on one side and 11 on the other.

distribution, however, can be thought of as essentially having a median regardless of the chance occurrence of an even or odd total frequency. It is a valid and necessary convention to modify the strict definition of the median as being the value of a single middle observation and to add that it may also be the midpoint

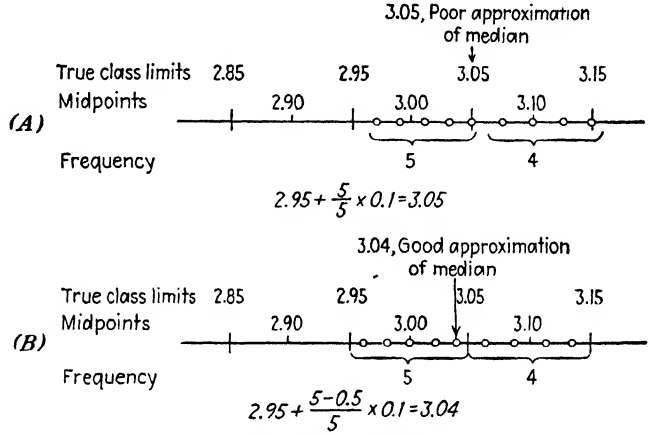


FIG. 14.—Diagram showing meaning of alternative methods of calculating the median. Both methods are meant to assume that the observations are evenly distributed within each class. *A*, calculation on the usual formula $L_l + \frac{N}{f}i$, showing that the calculation does not really suppose even distribution within the classes and hence is contrary to the assumption and gives incorrect results. *B*, calculation on the formula $L_l + \left(\frac{n - .5}{f}\right)i$, showing close approximation to the stated assumption and to the correct value.

between the values of two middle observations. The data of Example 32 (page 90) give these results in calculating the median:

$N = 162$

Middle observation (*i.e.*, in this case, midpoint between two middle observations) $= \frac{N + 1}{2} = \frac{163}{2} = 81.5$

There are 52 observations below the middle class; hence the one wanted in that class has the serial order $81.5 - 52 = 29.5 = n$.

The middle class has the real limits 54.5–59.5; so $L_l = 54.5$.

$f = 71$

$i = 5$

$$\text{Median} = L_l + \frac{(n - \tfrac{1}{2})i}{f} = 54.5 + \frac{(29.5 - .5)5}{71}$$
$$= 54.50 + 2.04 = 56.54$$

This may be checked by using the formula for the upper limit.

$$L_u = 59.5$$
$$n = 81.5 - 39 = 42.5$$
$$f = 71$$
$$i = 5$$
$$\text{Median} = L_u - \frac{(n - \tfrac{1}{2})i}{f} = 59.5 - \frac{(42.5 - .5)5}{71}$$
$$= 59.5 - 2.96 = 56.54$$

The result is the same, and the correctness of the formulas in application to distributions with N an even number is confirmed.

There is a simple but not particularly useful graphic method of finding the median. The distribution is put in cumulative form and plotted as an ogive. A horizontal line is now drawn at a position corresponding to $(N + 1)/2$ on the f-scale. A perpendicular from the point where this intersects the curve gives the value of the median on the X-scale.

The median is usually easier to calculate than the mean, and it has a few other advantageous properties. It is less distorted by extreme values than is the mean or is less sensitive to extreme values—often advantageous although it may also be a disadvantage. It can be calculated from some imperfect data, for instance, when the more divergent observations are grouped as so much "and over" or "and under," when the mean cannot. But the mean can also be calculated from data inadequate for a median, and either can be calculated from any properly collected and tabulated data. An observation selected at random is as likely to be above the median as below. The sum of the deviations (ignoring negative signs) is less about the median than about any other point.

These properties make the median an advantageous average in certain cases, but in ordinary practice they are outweighed by disadvantages. It is impossible to base a median on other medians. Medians cannot enter into many important algebraic calculations. They cannot be compared so simply and accurately as can means. Their standard errors are larger than for means.

In general the mean is a far more important and useful average than is the median. One essential use of the median (which also requires the use of the mean) is to approximate the value of the mode, as explained in the next section.

In a normal or in any other perfectly symmetrical distribution, the median is equal to the mean. Since actual distributions are rarely completely symmetrical, however, there is usually a small difference between these two averages.

MODE

For normal distributions and for those nearly like these and following Quetelet's principle, it is a fair generalization to say that an average is a value around which observations tend to cluster. The idea of the observations being crowded toward an average value applies well to the median and almost as well to the mean in such cases. In extremely skewed or J-shaped distributions, however, an average like the mean is not really a nucleus or a point of concentration of values. It is still true that the observations are arranged around the mean in such cases, and the calculation of the mean is still an essential part of their study, but the point around which they are really clustered may be well removed from the mean.

For studying distributions in which there is a significant degree of skewness, it is therefore necessary to have another sort of average, one that really designates in all cases a center of clustering or piling up of observations. Such a value is that at which the frequency is greatest, and this average is called the mode.

In a frequency distribution so grouped as to approach a fairly smooth curve, that is, with one class of outstanding frequency and with the frequencies of the other classes falling away evenly and definitely from this, the class in which the mode must occur, that with the highest frequency, is obvious on inspection. Thus in Example 32 (page 90), the mode is evidently in the class 55–59. If there is no single class with decidedly highest frequency, however, the position of the mode cannot be so closely approximated by simple inspection. Thus neither in Example 30 (page 87) nor in the same data more broadly grouped (Example 31, page 88) can one class be selected by inspection as containing the mode. If the size of the classes could be reduced and the total frequency increased indefinitely, it would be possible eventually to show a

definite modal class with interval as small as desired and so to determine the mode to any desired degree of accuracy, but this is impossible in practice. The total frequency cannot be increased at will, and if the class intervals are unduly decreased, the distribution becomes irregular and the mode cannot be placed at all.

The only accurate way to calculate the mode is to fit to a distribution the closest possible ideal mathematical frequency curve and then to calculate the point at which this curve has the highest ordinate. Such close curve fitting is an extremely complex process and requires more extensive data than are commonly available in zoology. The most accurate determination of the mode therefore has no practical value in zoology.

There are, however, several methods of approximating the mode that are useful in zoology and that give values accurate enough for practical purposes. The first and least refined of these is that already mentioned, grouping a distribution so that it is regular and has one class of outstanding frequency, then taking that class as an approximation of the mode.

A second method takes advantage of the fact that for moderately skewed curves the median lies at about one-third of the distance from the mean to the mode. This empirical rule, found to be closely followed by all but strongly skewed distributions, depends on the fact that the mode is not at all affected by extreme values, the median is somewhat affected, and the mean is most strongly affected, the effect on it being about one and one-half times as strong as on the median. This relationship gives an approximate formula for the mode:

$$\text{Mode} = \text{mean} - 3\,(\text{mean} - \text{median}) = 3\,\text{median} - 2M$$

This formula cannot be used for extremely skewed distributions, and in them approximation by inspection is the only easy and practical method.

There are several other methods giving still closer approximations; but they are also more complex mathematically, and the two mentioned suffice for any ordinary zoological work.

In the distribution of Example 32 (page 90) the mode is seen to lie in the class 55–59, *i.e.*, within the true limits 54.5–59.5. From calculations already given, the mean of this distribution is 56.66 and the median 56.54. Then since

$$\text{Mode} = \text{mean} - 3\ (\text{mean} - \text{median})$$
$$\text{Mode} = 56.66 - 3\ (56.66 - 56.54) = 56.66 - .36 = 56.30$$

This gives a reasonably accurate value which is within the class selected as modal by inspection.[1]

The importance of the mode is that since it is the value taken by the greatest number of observations it is in that sense the most typical. It can often be approximated roughly by simple inspection with no calculation, and it is independent of extreme values. It has the serious disadvantage that its exact calculation is practically impossible with limited data and is in any case extremely difficult and that, like the median, its usefulness for further calculations and for comparisons is far less than that of the mean. It may have the further disadvantage that for very small samples, such as are common in zoology, the mode may be quite indeterminate or may even be said, as far as a given concrete sample is concerned, not to exist.[2]

In practice the most important property of the mode and the only usual reason for its use is its being unaffected by extreme values. For instance, in a right skewed curve there is an excess of high values, to the right in a graph. These affect the mean, so that it lies well to the right of the mode and the difference between mean and mode thus provides a measure of skewness (see Chap. VII). Like the median, the mode is equal to the mean in a normal or other perfectly symmetrical distribution. In skewed distributions, the only ones for which its use is worth while, the mode may be a zoologically more important average than any other.

OTHER MEASURES OF CENTRAL TENDENCY

Several other measures of central tendency have been devised and are in occasional use, but they have relatively little practical value except in a few special problems, and only four of them will be mentioned here.

[1] It is not at the midpoint of the modal class, and such a midpoint is not a good approximation of the mode. The mode always lies on the side of the midpoint toward the greater frequencies in the classes beyond the modal class.

[2] Because the frequencies of all the measurements may be 1, or the differences in frequencies so slight as to have no discernible significance.

Range Midpoint.—This value is obtained by adding the lowest and highest observed values and dividing by 2. It is thus determined entirely by the extreme values and depends more on chance than on any real characteristic of the distribution. It is mentioned here only to observe that it has no practical use and should not be employed. It is generally avoided but occasionally appears in zoological work, sometimes with the wholly unwarranted assumption that it approximates or is equal to the arithmetic mean.

Geometric Mean.—The geometric mean is obtained by multiplying all the observed values and taking the Nth root of the product (N being total frequency as before). In mathematical notation

$$\text{Geometric mean} = \sqrt[N]{X_1 \cdot X_2 \cdot X_3 \cdot \ldots \cdot X_N}$$

X being the value of one observation of the variate.

The calculation can be greatly simplified by the fact that the logarithm of the geometric mean is the arithmetic mean of the logarithms of the individual observations.

The geometric mean has many of the advantages of the arithmetic mean; but it is relatively difficult to compute, the concept is not easily grasped or used, and it is indeterminate when negative values or zero occur among the observations. Its principal use is in the computation of index numbers as used especially in commercial statistics.

Harmonic Mean.—The harmonic mean is the reciprocal of the arithmetic mean of the reciprocals of the observed values.

$$\frac{1}{H} = \frac{1}{N}\sum\left(\frac{1}{X}\right)$$

H being the usual symbol for the harmonic mean.

The harmonic mean is always smaller than the geometric mean based on the same data, and the geometric mean is always smaller than the arithmetic mean. The harmonic mean is used in averaging rates.

Quadratic Mean.—The quadratic mean is the square root of the arithmetic mean of the squares of the observed values.

$$\text{Quadratic mean} = \sqrt{\frac{\Sigma(X^2)}{N}}$$

It is seldom used as such, but it is involved in some methods of calculating measures of dispersion (see Chap. VI).

The various measures of central tendency are illustrated graphically in Fig. 15.

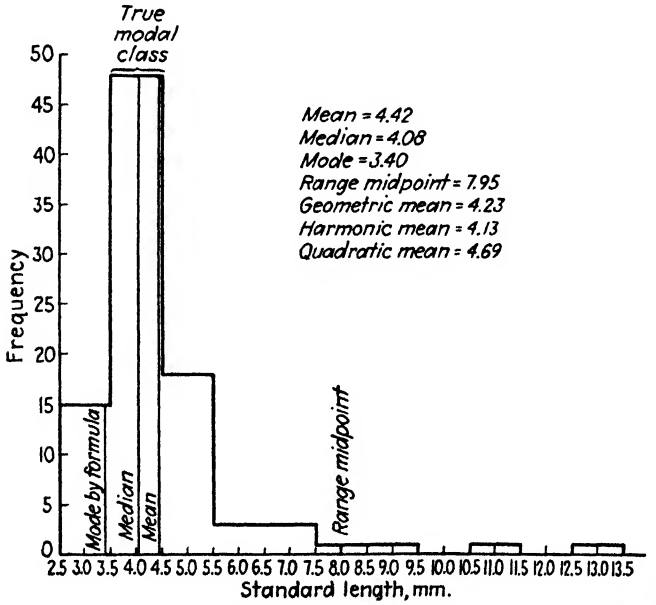


FIG. 15.—Diagram illustrating the relationship between various measures of central tendency in an asymmetrical distribution. Standard lengths of specimens of the fish *Parexocoetus brachypterus hillianus*, April collection. Note the following points: the range midpoint gives no conception of the actual distribution; the mean is on the right of the median in the diagram, that being the side of the distribution with the longer tail; the mode is on the other side of the median. Note also that, because of the extreme skew of the distribution, the mode calculated from the rough formula, mode = 3 median − 2 mean, is obviously incorrect.

THE MEANING OF AVERAGE, TYPICAL, AND NORMAL

An average, as defined and used in this chapter, is any constant measuring central tendency in a frequency distribution. It is generally given as a single figure, representing a point or a small class in the distribution around which the observations tend in some way to cluster. Since this clustering is a complex phenomenon, there are many sorts of averages, each with its own distinctive properties.

In common speech the word "average," if it is intended to be used exactly, is generally taken to signify only one of the many

averages of technical usage, the arithmetic mean. It is, however, seldom used so exactly in the vernacular and generally implies that the average is a large group including all but a few strongly aberrant observations. Hence occasional outbursts of indignation that a third, or some other large fraction, of our human population lives in conditions below the average, or has an income below average, or is below average intelligence. When one considers what averages really are, such statements are obviously ridiculous and tell nothing about the real distribution of living conditions, income, or intelligence. Obviously if the average in question is the median or (more approximately) if it is the mean, half the population must inevitably be below average in every respect. If the mean is high, a person may be far below average and yet be living luxuriously. This and analogous widespread fallacies in the use of words are both amusing and dangerous in the mouths of legislators. The reason for emphasizing them here is that many zoologists tend to carry over this looseness of thought into their work and to confuse vernacular and technical usages of such words as average.

The word typical is also subject to such confusion, so much so that it is hard to give it an exact sense. It implies the existence of a standard of comparison, a type. In the vernacular the type is sometimes supposed to be the average, or a member of a large group around the mean, and sometimes is a sort of ideal by no means really average. The usual description of the "typical American" has more to do with what the speaker or writer wishes were the mean in our population than with what the mean really is. In somewhat better defined usage, one as common in scientific as in popular language, the typical condition is taken to be that most frequent. Typical then signifies in more technical language belonging to a modal group. This usage, proper but requiring definition, is in turn often confused with the strictly technical use of types in zoology. The type of a taxonomic group is the basis and standard of comparison for that group, but it need not necessarily be and very frequently is not in a modal class in the frequency distribution for the taxonomic division. It may be far removed from any average and is quite likely to be, since it is usually the first specimen that came to hand by chance. The type of a species is thus not, or not necessarily, typical in any of the more usual senses of the word.

The word "normal" in the vernacular is subject to a curious dual usage, in which two mutually exclusive ideas are confused and confounded. It is supposed in the first place that the normal is a sort of average and in the second place that it means the absence of some particular sort of variation regardless of the fact that such variations do occur and hence do in some degree characterize the average. Physicians are the worst technical offenders in this sense, and medical literature is full of equivocations resulting from this double usage. It is assumed that the normal condition is the mean condition and also that the normal condition is one without any pathological factors. Normal cannot mean both these things at once. If anyone in a population is ill in any way, as of course is true of all populations of any size, then the mean condition of the whole population is one of partial illness. The typical condition, in the sense of the modal condition, may or may not include pathological factors, but the mean condition always does. In practice the modal condition usually does also. Perfect health is relatively rare. It is in an extreme, not a middle, position in the frequency distribution of health, and normal health in this sense is an unusual and not an average condition.

It is more reasonable in such a distribution to think of all the observations that really fit into the distribution as normal. It is as normal to be on the point of death as to be in perfect health. The smallest member of a species is as normal as the largest or as one of mean size. It is unfortunate that the word normal is used in a still more special and logically unrelated way in the name of the "normal distribution." This is a highly special and technical use of the word to mean not only conforming to a pattern but to one particular pattern of distribution. By no means all normal variations, in any but this one special sense, fall into a normal distribution.

CHAPTER VI

MEASURES OF DISPERSION AND VARIABILITY

The determination of any of the various averages gives a point or a small group around which observations are arranged in some way. In most cases the arrangement is that they tend to cluster around this value, to be crowded toward it, or to pile up on it, with frequencies falling away from it in both directions.[1] The

EXAMPLE 35.—HYPOTHETICAL DISTRIBUTIONS TO SHOW DIFFERENT DISPERSION WITH IDENTICAL RANGES AND MEANS

A		B	
Measurements	*f*	Measurements	*f*
4.9	2	4.9	1
5.0	3	5.0	1
5.1	2	5.1	2
5.2	1	5.2	5
5.3	4	5.3	11
5.4	3	5.4	3
5.5	2	5.5	2
5.6	1	5.6	0
5.7	2	5.7	1

In both of these the range is 4.9–5.7 and in both the mean is 5.28, yet they are decidedly different types of distribution. If, for instance, these were lengths of individuals of a given species, it would be justified with distribution B to consider the values 4.9 and 5.7 as exceptional; and if an individual were found with length 4.7 or 5.9, there would be some probability that it did not belong to this species. But with distribution A the values 4.9 and 5.7 are not exceptional, and there is no probability against reference of individuals measuring 4.7 or 5.9 to the species. In biological terms a character distributed like A is more and one distributed like B is less variable, even though in the samples at hand the range covered may be the same in each case.

[1] As pointed out in the last chapter, this is invariably true only of the mode, but it is an approximate statement of the usual relation of frequencies to other averages.

determination of such a point, essential as it is, does not tell enough about the real nature of the distribution. It is necessary to know also about how far the observations extend on each side of this point and about how fast the frequencies fall away from it, or, expressing the same thing from a different point of view, to what extent they are piled up around it. It makes a great difference in the conclusions to be drawn from a series of measurements whether they run, say, from 2.8–7.6 or from 4.9–5.7, although in both cases the mean may be the same.

It also is a very essential factor whether the frequencies are rather evenly scattered or are strongly concentrated at some point, even though the range and the mean of the observations may be the same in either case, as shown by Example 35.

The adequate measurement of these very important characteristics of frequency distributions is one of the greatest problems of zoology, and it is one toward which most zoological work so far published makes little progress. There are, however, good methods of making such measurements, called measures of dispersion, and this chapter is devoted to the most useful of these.

OBSERVED RANGE

The observed range is the difference between the highest and the lowest observed values of a variate, usually and most usefully expressed by giving these extreme observed values although, strictly speaking, the observed range is not these values but the difference between their limits. Thus in Example 36, giving data that will be used throughout this chapter so that the different measures and means of calculation can be easily compared, the observed range is best recorded as 52–68 mm. The difference between these, the actual value of the observed range, is 17 mm.[1]

The observed range, usually but with some danger of confusion simply called the range, is a useful datum and should be given whenever pertinent, but it has many drawbacks and it is not a good measure of dispersion. Because of its simplicity, obvious meaning, and requirement of no calculation it is frequently given in zoological publication, which is desirable; but unfortunately it

[1] Not 16 mm. as is usually stated in analogous cases. The range really lies between the implied limits of the extreme measurements, *i.e.*, between 51.5 and 68.5 mm. in the example, not between the recorded measurements, which are midpoints and not limits.

is often given without any way to assess its value and it is often assumed to be an adequate representation of a distribution and to be a significant measure of variability, which it is not. Indeed most zoologists who have discussed variability and variation, perhaps the most important single aspect of zoological work, have made this erroneous assumption, making their work relatively useless and revealing ignorance of a basic concept of zoology, what variability really is.

EXAMPLE 36.—TAIL LENGTH IN MILLIMETERS OF SPECIMENS OF THE DEER-MOUSE, *Peromyscus maniculatus bairdii*, BOTH SEXES, IN THE 1-YEAR AGE CLASS, TAKEN NEAR ALEXANDER, IOWA

(Data from Dice 1932)

Tail length (X)	No. of observations (f)
52	1
53	0
54	3
55	0
56	3
57	8
58	7
59	11
60	11
61	10
62	6
63	14
64	6
65	3
66	1
67	1
68	1

In the first place, it is clear that the observed range is dependent on the number of observations made. If only one is made, the observed range is zero. Certainly this does not mean that the species, or other category, measured does not vary at all in nature. If two observations are made, the observed range may be large or small but will probably be small. In general the probability is that the more observations are made the larger will be the observed range. Unless the total frequency is also given, an observed range is thus meaningless. Even if the total frequency is given, the meaning of the observed range is uncertain, for its increase with increased number of observations depends in large

measure on chance and its value with any given N is a matter of probability, usually with a large element of uncertainty, rather than of any simple and easily calculable relationship.

Any variate does have a real range. In any given species, for instance, there really does exist in nature one individual that is the largest and one that is the smallest. The difference between these, the real as opposed to the observed range, is an important significant character of the species or more generally of any variate; but it is never surely available. The chances of actually observing the largest and smallest of all existing values of any variate are obviously very small, and in most cases it would be impossible to know that they were the extreme values even if they were observed.

The observed range is significant only as an approximation of the real range. With small samples the approximation is so poor in most cases and so unreliable in all that the observed range is not very helpful. With large samples the observed range may be a relatively good approximation of the real range; but even in such cases a more reliable estimate of the real range can be obtained by calculation from other measures of dispersion, especially from the standard deviation (see the following paragraphs).

In any event, as the distributions on page 106 show, the range, especially observed but even real, does not give all the desired or necessary information about dispersion and variability. In terms of frequency curves, it shows at best only where the curve ends and tells nothing of the equally or more important shape of the curve between the ends. For all these reasons the observed range is the poorest of all the measures of dispersion.

MEAN DEVIATION

The range is dependent on only two values, the most extreme available. Clearly a better measure of dispersion can be obtained if all the values are taken into consideration and some figure calculated that depends on their spread from a given point, an average. Of such measures, the simplest is the mean deviation, which is the arithmetic mean of the distances of the individual observations from the arithmetic mean of the whole distribution. The fact that some observations are above the mean (have larger values) and some below is represented in usual calculation by

making the former distances from the mean positive and the latter negative. If this were done in calculating the mean deviation, it follows from the definition of the mean that the mean deviation about the mean would always be zero. In fact, the concern here is with the distances from the mean, that is, the deviations, not with their direction from the mean, so that all the deviations are taken to be positive or, as it is usually but less logically expressed, the signs are ignored. The calculation is by the following formula:

$$\text{M.D.} = \frac{\Sigma(fd)}{N}$$

in which M.D. is the mean deviation, d any one deviation from the arithmetic mean (in either direction), and the other symbols as

EXAMPLE 37.—CALCULATION OF MEAN DEVIATION FROM THE DATA OF EXAMPLE 36

X	f	fX	d	fd
52	1	52	8.43	8.43
53	0	0	7.43	0
54	3	162	6.43	19.29
55	0	0	5.43	0
56	3	168	4.43	13.29
57	8	456	3.43	27.44
58	7	406	2.43	17.01
59	11	649	1.43	15.73
60	11	660	.43	4.73
61	10	610	.57	5.70
62	6	372	1.57	9.42
63	14	882	2.57	35.98
64	6	384	3.57	21.42
65	3	195	4.57	13.71
66	1	66	5.57	5.57
67	1	67	6.57	6.57
68	1	68	7.57	7.57
	$N = \Sigma(f) = 86$	$\Sigma(fX) = 5,197$		$\Sigma(fd) = 211.86$

$$M = \frac{\Sigma(fX)}{N} = \frac{5,197}{86} = 60.43$$

d (for each class) $= X - M$ (sign ignored)

$$\text{M.D.} = \frac{\Sigma(fd)}{N} = \frac{211.86}{86} = 2.46$$

previously explained.[1] The calculation is exemplified in Example 37.

The mean deviation is a distance that is in the same units as the measurements involved, or, for a discontinuous variate, it is a count of the same units as the original data. The mean deviation calculated in Example 37 is 2.46 mm.

EXAMPLE 38.—CALCULATION OF MEAN DEVIATION ABOUT THE MEDIAN FROM THE DATA OF EXAMPLE 36

X	f	d	fd
52	1	8.41	8.41
53	0	7.41	0
54	3	6.41	19.23
55	0	5.41	0
56	3	4.41	13.23
57	8	3.41	27.28
58	7	2.41	16.87
59	11	1.41	15.51
60	11	.41	4.51
61	10	.59	5.90
62	6	1.59	9.54
63	14	2.59	36.26
64	6	3.59	21.54
65	3	4.59	13.77
66	1	5.59	5.59
67	1	6.59	6.59
68	1	7.59	7.59
	$N = 86$		$\Sigma(fd) = 211.82$

$$\frac{N+1}{2} = 43.5$$

Median class = 60 $\quad L_l = 59.5$

$n = 43.5 - 33 = 10.5$

$f = 11$

$i = 1$

Median $= 59.5 + {}^{10}\!/_{11} = 60.41$

$$\text{M.D.}_{\text{median}} = \frac{\Sigma(fd)}{N} = \frac{211.82}{86} = 2.46$$

The mean deviation can be calculated not only about the mean, the usual practice, but also about any other average. The only other one commonly used is the median. A mean deviation

[1] Explanation of the commonly used symbols is not repeated each time they are used. They are all also given in the Appendix for reference (without having to go back over the text to search for them).

about the median should be so specified, since mean deviation otherwise is taken to be about the mean. Example 38 shows the calculation about the median.

It is noteworthy that $\Sigma(fd)$ is less for the median, 211.82 in the example, than for the mean, 211.86, bearing out the statement made in the last chapter that this figure is less for the median than for any other average. In the example the mean and median are so nearly the same that the mean deviations about them are the same to two decimal places.[1]

There is a short method for calculating mean deviations analogous to that for calculating the mean, explained in the preceding chapter. Since it is seldom necessary to calculate a mean deviation, however, and since with most zoological data this can be done readily enough by the long method, space is not taken here to exemplify the short method.

It is easy to understand what the mean deviation signifies, and this measure of dispersion is relatively easy to calculate, although the difference from the standard deviation in ease of calculation will not be found great. If there are large erratic deviations beyond the bulk of the distribution, they usually disturb the mean deviation less than the standard deviation, and if they are not considered significant for the problem, the mean deviation may in such cases be preferable. In general, however, the mean deviation is not the best measure of dispersion. Its use in algebraic calculations is inconvenient, and its relationships to the normal curve and the theory of errors and its use in comparing means or other constants are also relatively inconvenient and not so well worked out as for the standard deviation. In almost every case the standard deviation, which will be discussed next, is preferable. The mean deviation has been introduced at this point and explained at this length not so much to recommend its use as because it provides a simple introduction and logical background for the problems of dispersion and the use of deviations in general.

STANDARD DEVIATION

If the direction from the mean is taken into account, or in other words if the signs are retained, the sum of the deviations about

[1] Further decimal places in these instances would not be significant for the problem. This question of significant places is discussed in Chap. VIII.

the mean is always zero—a statement obviously true because the process of finding the mean is simply a method of determining the point above and below which the deviations are equal in quantity although opposite in sign. The mean deviation would therefore always be zero except for the expedient of ignoring the directions of the deviations. The standard deviation is a measure obtained without recourse to this artifice, but taking advantage of the fact that any number, positive or negative, has a positive square. One has, then, only to square the deviations to be sure that their sum will always be positive and, in practice, will always be greater than zero.

The mean of the squares of the deviations is called the variance of the distribution and it is fundamentally the most important statistical constant measuring dispersion. The formula is simple:[1]

$$v_M = \frac{\Sigma(d^2)}{N}$$

in which v_M is the variance about the mean.

[1] R. A. Fisher (1936) gives the formula (in his own notation) as

$$v_M = \frac{\Sigma(d^2)}{N - 1}$$

and remarks, "In large samples the difference between these formulae is small, and that using N may claim some theoretical advantage if we wish an estimate to be used in conjunction with the estimate of the mean from the same sample, as in fitting a frequency curve to the data; in general it is best to use $(N - 1)$. In small samples the difference is still small compared to the probable error, but it becomes important if a variance is estimated by averaging estimates from a number of small samples." Many other biometricians do not as yet accept Fisher's formula. Without questioning the logic involved, it appears that the formula $\Sigma(d^2)/N$ is usually better for the uses that we contemplate. It is now in universal use, so that results are strictly comparable only if it is employed. Even for small samples, as Fisher observes, the value obtained is generally not significantly different from that of Fisher's formula. It is generally used, in zoological work, in conjunction with the mean and in ways related to that for which Fisher also grants it a theoretical advantage. It does not appear to be a question of right and wrong but only of different conventions or ways of treating the data. The usual procedures and tables presuppose that this formula rather than Fisher's will be used. For these reasons we retain it in general and suggest that Fisher's formula may be advantageous only in cases like those of averaging estimates derived from a number of small samples, a procedure relatively rare in zoology as opposed to experimental biology, and for special uses with very small samples, on which see Chap. XI.

In practical use, it is almost always more convenient to use the square root of the variance. This reduces the magnitude to one directly comparable to the deviations themselves and hence is particularly adapted to such uses as considering the significance of individual deviations and in general serves better the purposes for which a measure of dispersion is wanted. This constant, the square root of the variance, is the standard deviation

$$\sigma = \sqrt{\frac{\Sigma(d^2)}{N}}$$

in which σ (Greek lower-case sigma) is the almost universally used symbol for the standard deviation.

The calculation of σ by a method analogous to the long method of finding the mean is given in Example 39.

The variance v_M around the mean M has a simple relationship to the variance v_A around any other point A:

$$v_A = v_M + d_A^2$$

where d_A = the difference between the mean and the other point $(M - A)$.[1]

This relationship proves the statement previously made that the variance is less around the mean than around any other point, because d_A^2 is always positive and has the value 0 only when $A = M$. For this and other reasons the variance and the standard deviation are always taken around the mean; but the relationship to variances around other points is used to shorten the calculation of σ by using an assumed mean A, analogous to the

[1] If d is a deviation from the mean, a deviation from A is $d + (M - A)$, or $d + d_A$. If deviations from A are designated by δ,

$$\begin{aligned} \delta &= d + d_A \\ \delta^2 &= d^2 + 2dd_A + d_A^2 \\ \Sigma(\delta^2) &= \Sigma(d^2) + 2d_A\Sigma(d) + Nd_A^2 \end{aligned}$$

Since $\Sigma(d)$ is always 0, the second term vanishes:

$$\begin{aligned} \Sigma(\delta^2) &= \Sigma(d^2) + Nd_A^2 \\ \frac{\Sigma(\delta^2)}{N} &= \frac{\Sigma(d^2)}{N} + d_A^2 \end{aligned}$$

which is the same as

$$v_A = v_M + d_A^2$$

short method of calculating M given in the preceding chapter. The variance is calculated around an assumed mean in terms of class intervals, the midpoint of the modal class usually being taken as the assumed mean since this simplifies the arithmetic;

EXAMPLE 39.—CALCULATION OF STANDARD DEVIATION BY THE LONG METHOD FROM THE DATA OF EXAMPLE 36

Tail length (X)	No. observations (f)	Deviation from mean		
		d	d^2	fd^2
52	1	−8.43	71.0649	71.0649
53	0	−7.43	55.2049	0
54	3	−6.43	41.3449	124.0347
55	0	−5.43	29.4849	0
56	3	−4.43	19.6249	58.8747
57	8	−3.43	11.7649	94.1192
58	7	−2.43	5.9049	41.3343
59	11	−1.43	2.0449	22.4939
60	11	− .43	.1849	2.0339
61	10	.57	.3249	3.2490
62	6	1.57	2.4649	14.7894
63	14	2.57	6.6049	92.4686
64	6	3.57	12.7449	76.4694
65	3	4.57	20.8849	62.6547
66	1	5.57	31.0249	31.0249
67	1	6.57	43.1649	43.1649
68	1	7.57	57.3049	57.3049
	$N = 86$			$\Sigma(fd^2) = 795.0814$

$M = 6 \qquad 0.43$

$$\sigma = \sqrt{\frac{\Sigma(fd^2)}{N}} = \sqrt{\frac{795.0814}{86}} = \sqrt{9.2451} = 3.04^*$$

and from this is subtracted the square of the difference between the true mean and the assumed mean, the difference also being in class intervals. The square root of the remainder is the standard deviation in terms of class intervals; and if these are not 1, this is multiplied by the class interval to give the standard deviation in ordinary units. The general formula is

* It should be noted that the calculations in this and other examples in which the arithmetic is somewhat difficult need not be carried out arithmetically and are not in practice. Logarithms, tables of squares and roots, a slide rule, a calculating machine, or any combination of these are to be used (see Appendix).

$$\sigma = i\sqrt{\frac{\Sigma(fd_A^2)}{N} - c_1^2}$$

where d_A = the deviation from the assumed mean in terms of class intervals.

i = the class interval.

c_1 = the difference between mean and assumed mean in terms of class intervals.

This correction factor can be calculated otherwise but is most easily found from the relationship

$$c_1 = \frac{\Sigma(fd_A)}{N}$$

The correction factor c, for finding the mean, is equal to ic_1.

The calculation of M and σ by the short method is shown in Example 40.

The slight differences in M and σ obtained in this example from the more accurate values of Example 39 are caused by the grouping, not by any inaccuracy in the short method. The short method can be applied equally well to the original measurements, and the long method can be applied to the data grouped to class interval 2 (or any other, using class midpoints); and when the grouping is the same, the two methods give the same results. The short method has been exemplified here only with the larger grouping for the sake of brevity.[1]

[1] The use of grouped data to obtain the variance implies that the frequencies are concentrated at the class midpoints. In fact they may be scattered over the whole class interval, and this difference of fact from theory introduces an element of inaccuracy. If the frequency distribution is continuous and if it tapers off gradually to zero in both directions, the average effect is to increase the variance by the constant amount $\frac{1}{12} = .0833$, in terms of the units in which the deviation is measured. The subtraction of this figure from the variance (before the standard deviation is obtained from the variance) is Sheppard's correction. This is inapplicable to discontinuous distributions or to strongly skewed distributions and is in any case only an average effect: the actual effect in a given case cannot be measured and may be much less, especially with small samples. The application of the correction is, indeed, so doubtful in practice and the difference that it makes so slight that in practical work it is commonly omitted unless there is strong reason to believe it applicable and the data permit and the problem demands extreme accuracy. Although hardly called for in most zoological computations, it is mentioned because of these rarer instances when it should be

EXAMPLE 40.—CALCULATION OF STANDARD DEVIATION FROM DATA OF EXAMPLE 36 BY THE SHORT METHOD, AFTER SECONDARY GROUPING

Tail length	Number of observations	d_A	fd_A	fd_A^2
52–53	1	−4	− 4	16
54–55	3	−3	− 9	27
56–57	11	−2	−22	44
58–59	18	−1	−18	18
			−53	
60–61	21		48	
62–63	20	1	20	20
64–65	9	2	18	36
66–67	2	3	6	18
68–69	1	4	4	16
	$N = 86$		$\Sigma(fd_A) = -5$	$\Sigma(fd_A^2) = 195$

$i = 2$

A = midpoint of class 60–61 = 60.5

$$c_1 = \frac{\Sigma(fd_A)}{N} = \frac{-5}{86} = -.058 \qquad c_1^2 = .0034$$

$$c = ic_1 = 2 \times -.058 = -.12$$

$$M = A + c = 60.5 + (-.12) = 60.38$$

$$\sigma = i\sqrt{\frac{\Sigma(fd_A^2)}{N} - c_1^2} = 2\sqrt{\frac{195}{86} - .0034}$$

$$= 2\sqrt{2.2640} = 2 \times 1.505 = 3.01$$

The fd_A^2 column is obtained by multiplying fd_A by d_A, and it is not necessary to obtain d_A^2 separately.

With distributions reasonably near symmetry, the mean deviation is usually about four-fifths of the standard deviation. In

---

used and because it will occasionally be encountered in the literature. In Example 39 the application of Sheppard's correction would make $\sigma = 3.03$, as opposed to 3.04 without the correction, a difference not significant and not certainly a real increase in accuracy. In Example 40 it would make $\sigma = 2.95$ as opposed to 3.01, a difference that may well be a decrease, not increase, in accuracy.

Most authors apply the correction, if at all, only to data grouped with intervals larger than the original unit of measurement. But the original measurements are just as much grouped as are any secondary groupings based on them, and it would seem that the correction if applicable to either would be applicable to both.

Example 37 the mean deviation was found to be 2.46, and in Example 39, using the same data, the standard deviation is 3.04.

$$\frac{2.46}{3.04} = .81$$

which is very nearly four-fifths, as called for by this rule.[1]

The standard deviation, like the mean deviation, is an absolute figure in the same units as those of the original measurements or counts. Although usually written as in Example 39 where it is recorded as simply 3.04, it must be remembered that this is not an abstract number or a relative value but is itself a measurement, in this case 3.04 mm.

The standard deviation is usually the best available measure of dispersion. It is not much more difficult to calculate than is, for instance, the mean deviation; it can conveniently be used in algebraic calculation; and its relationship to the normal curve and to the theory of errors is relatively simple and direct. Its properties and uses are so extensive and important that they are given in the course of subsequent chapters rather than attempting to sum them all up briefly here, but one important point may be mentioned. If, as is so often the case, a normal distribution is an adequate approximation of an actual given distribution, then about 68 per cent of the observations will be distant not over one standard deviation above or below the mean, about 95½ per cent will be distant not over two standard deviations, and for practical purposes all the observations (theoretically about 99¾ per cent) will be within three standard deviations of the mean. From these relationships it follows that 6σ is a much better approximation of the real range and $(M - 3\sigma)$ and $(M + 3\sigma)$ of its limits than are the observed range and observed limits. In Example 39, based on an excellent sample of a nearly normal distribution, 6σ is about 18 and $(M - 3\sigma)$–$(M + 3\sigma)$ is about 51–69. The observed range is 17, 51.5–68.5. The agreement is excellent and confirms the relationship. Few distributions are as good as this; and the observed range is usually irregular and considerably smaller relative to 6σ or, it may safely be inferred, to the real range. The observed range is generally from 4σ to 6σ, a fact that may be used

[1] Based on the fact that in the normal curve, which is approximated by such distributions as are included in the rule, M.D. = $.7979\sigma$.

as a rough check in computing σ to see whether the value obtained is of the right magnitude.

SEMI-INTERQUARTILE RANGE

Quartiles measure the values of a variate below which lie one-fourth, two-fourths, or three-fourths of the observations and are designated respectively as the first, second, and third quartiles. Obviously the second quartile, with two-fourths or one-half of the observations below it, is the median, and it is usually called by that name, only the first and third quartiles being explicitly called quartiles.

The calculation of the first and third quartiles is the same as for the median except that a different value is given to n for each.

$$Q_1 = L_l + \frac{(n_1 - \tfrac{1}{2})i}{f}$$

$$Q_3 = L_l + \frac{(n_3 - \tfrac{1}{2})i}{f}$$

in which n_1 is found by subtracting the total frequency below the first quartile class from $(N + 1)/4$ and n_3 by subtracting that below the third quartile class from $3(N + 1)/4$.[1]

[1] As with the median, most statisticians use N where we use $N + 1$, and apparently none uses the correction of subtracting $\frac{1}{2}$ from n_1 or n_3; but the same logic applies here as for the median, as discussed in Chap. V.

For the data of Example 40 (page 117) N is 86, and i is 2.

$$\frac{N + 1}{4} = 21.75 \qquad \frac{3(N + 1)}{4} = 65.25$$

The first quartile must lie in the class 58–59, lower limit 57.5, frequency 18, and the frequency below this class is 15; therefore

$$n_1 = 21.75 - 15 = 6.75$$

Similarly the third quartile is in the class 62–63, lower limit 61.5, frequency 20, frequency below this 54; so

$$n_3 = 65.25 - 54 = 11.25$$

Placing these values in the formulas,

$$Q_1 = 57.5 + \frac{(6.75 - .5)2}{18} = 57.5 + .69 = 58.19$$

$$Q_3 = 61.5 + \frac{(11.25 - .5)2}{20} = 61.5 + 1.08 = 62.58$$

The median, which is Q_2, has already been found to be 60.41. In a sym-

The distance between Q_1 and Q_3 is a measure of dispersion. Half the observations are piled up within this distance, a quarter below it and a quarter above. The measure usually taken, however, is the semi-interquartile range, half of this distance, which is the average distance (regardless of direction) from the median to the first and third quartiles. It is also called the quartile deviation.

$$\text{Q.D.} = \frac{Q_3 - Q_1}{2}$$

in which Q.D. is the quartile deviation or semi-interquartile range. In the example used above

$$\text{Q.D.} = \frac{Q_3 - Q_1}{2} = \frac{62.58 - 58.19}{2} = \frac{4.39}{2} = 2.20$$

In distributions approximately normal, this is usually about two-thirds of the standard deviation.[1] In the example

$$\frac{\text{Q.D.}}{\sigma} = \frac{2.20}{3.04} = .72$$

a value somewhat higher than expected under the empirical rule, because the distribution is not, in fact, quite symmetrical, but still approximately two-thirds.

The quartiles and the semi-interquartile range have much the same advantages and disadvantages as the median. They are relatively simple and obvious characteristics of a distribution and are not infrequently employed on that account; but they do not lend themselves to calculation and comparisons so well as do the mean and the standard deviation, and their general use is not recommended.

In place of quartiles, quintiles, dividing the frequencies into 5 equal groups, deciles, dividing them into 10, and percentiles, dividing them into 100 groups, are sometimes useful in special cases, mostly outside the field of zoology. They are calculated like the median, getting an appropriate n for each by starting

---

metrical distribution, this lies halfway between first and third quartiles, which it nearly does in this case:

Median or $Q_2 - Q_1 = 60.41 - 58.19 = 2.22$

Q_3 – median or $Q_2 = 62.58 - 60.41 = 2.17$

[1] Because in the normal curve Q.D. = $.6745\sigma$.

with multiples of $(N + 1)/5$ for quintiles, of $(N + 1)/10$ for deciles, and of $(N + 1)/100$ for percentiles. If such figures are of any use, it is easier and usually sufficiently accurate to get them graphically, plotting the distribution as an ogive and drawing horizontal lines at heights corresponding to the desired fractions of the total frequency. The median can also be obtained in this way, as already mentioned, and so, of course, can the quartiles.

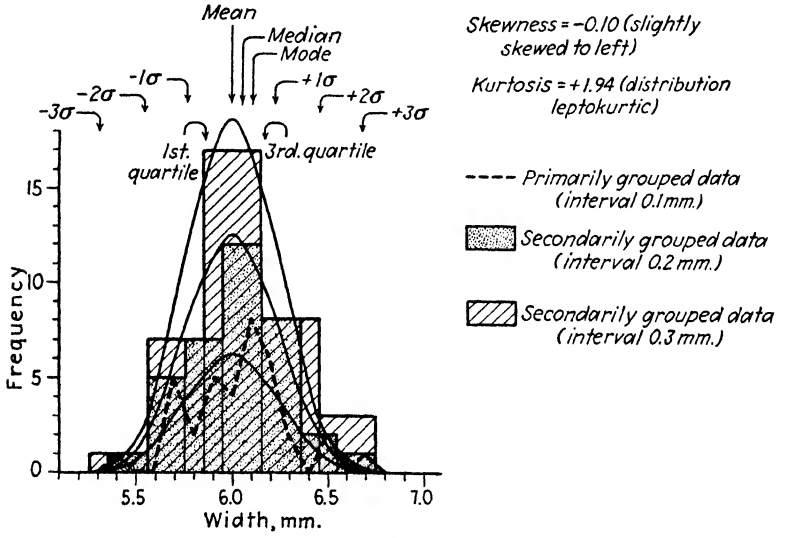


FIG. 16.—Various representations and group measures of an approximately normal frequency distribution.

Widths of last upper molars of the extinct mammal *Acropithecus rigidus*. Broken line, frequency polygon of the raw (primarily grouped) measurements (see Fig. 2). The histograms show the same data secondarily grouped with interval .2 mm. and .3 mm. (see Fig. 3). The three curves are normal curves roughly fitted to these three forms of the data. Their essential parameters are exactly the same; and they differ only because the broader groupings give fewer classes, hence higher class frequencies, and hence larger ordinates in the curves. The principal measures of central tendency and of dispersion are given. Note, in this satisfactory sample of moderate size, the close approximation of $M \pm 3\sigma$ to the observed range. The coefficients of skewness and of kurtosis measure the principal departures from the normal distribution, neither very marked.

RELATIVE DISPERSION AND VARIABILITY

All the measures of dispersion so far discussed are absolute. They are themselves measurements or counts of the variate, and their interest is that they are particular values of the variate lying at certain fixed and characteristic points in the given distribution. Valuable as such measurements are, they do not make possible a direct comparison of the dispersion and variability of variates of different absolute mean sizes, and just such comparisons are what

the zoologist usually has in mind when he talks of variability. In most fields a measure of relative variability is useful, and in zoology it is indispensable.

The fact that elephants, for instance, may have a standard deviation of 50 mm. for some linear dimension and shrews a standard deviation of .5 mm. for the same dimension does not necessarily mean that the elephants are more variable, in the essential zoological sense, than are the shrews. The elephants are a hundred times the size of the shrews in any case, and we should expect the absolute variation also to be about a hundred times as great without any essential difference in functional variability. The solution of this problem is very simple: it is necessary only to relate the measure of absolute variation to a measure of absolute size. The best measures to use for this purpose are the standard deviation and the mean, and since their quotient is always a very small number it is convenient to multiply it by 100. The resulting figure is a coefficient of variation, or of variability.[1] The formula is $V = 100\sigma/M$, V being the standard symbol for this coefficient of variation. The value of V, unlike that of σ, is usually spoken of as a pure number, divorced from any unit of measurement. More accurately, it measures a characteristic in units peculiar to itself, and its magnitude is not commensurate with the lengths, weights, or other units of standard mensuration.

The valid use of V depends on the assumption that variation as a biological function is relative to absolute size or, in terms of distributions, that absolute dispersion increases in direct proportion to the mean in variates of essentially the same variability from a biological point of view. It is logical in our understanding of what variability is that this should be so. It has also been empirically determined by many calculations on samples of all sorts of animals that this is at least approximately true of most of the variates used in zoology.

The comparison of values of V derived from different distributions is almost invariably valid and useful if the variates are homologous. If they are not, experience suggests that the com-

[1] Proposed by Karl Pearson (1857–1936), English leader in biometrics. As much as any one man, he was responsible for the extension and development of statistical and other advanced quantitative methods in the life sciences.

parison is still generally valid if the variates are analogous and belong to the same category, for instance, if they are all linear dimensions of anatomical elements, the usual case. It is also helpful that the units of measurement have no influence on the comparison as long as they are in one category: a V derived from measurements in millimeters is directly comparable to one from measurements in feet. As a rule, however, coefficients of variation for variates of essentially different categories cannot be usefully compared. V for a continuous and V for a discontinuous variate, V for a temperature and V for a mass, V for a linear dimension and V for an area, and the like, are not to be considered comparable unless this is shown to be warranted by logic and by experience. It is also necessary to bear in mind that the implication that absolute dispersion relative to the mean measures biological variability is only a broad rule, mainly empirical in foundation, and that it is always open to exception. V is in every case a good measure of relative dispersion, but relative dispersion is not always a good measure of variability. It is, however, usually so, and V is by far the most useful measure of this characteristic that has yet been proposed. Its use involves only the common-sense necessity of remembering what V is and being sure that this is really what one wishes to compare.

Discernment of the meaning of a value of V is largely a matter of experience. Its interpretation on functional zoological grounds depends on nonnumerical biological knowledge. We have compared dozens of V's for linear dimensions of anatomical elements of mammals. As a matter of observation, the great majority of them lie between 4 and 10, and 5 or 6 are good average values. Much lower values usually indicate that the sample was not adequate to show the variability. Much higher values usually indicate that the sample was not pure, for instance, that it included animals of decidedly different ages or of different minor taxonomic divisions. If the sample is adequate and reasonably unified, then different values of V generally represent in a clear and useful way inherent differences in variability.

For the data of Example 39 (page 115), V is calculated as follows:

$$V = \frac{100\sigma}{M} = \frac{100 \times 3.04}{60.43} = 5.03$$

EXAMPLE 41.—SKULL MEASUREMENTS OF MALES, ADULT UNLESS OTHERWISE NOTED, OF THE NORTHERN WHITE-TAILED DEER, *Odocoileus virginianus borealis*, FROM NEW ENGLAND, ALL BUT ONE OR TWO FROM MAINE
(Data from Phillips 1920)[1]

Variate	V
1. Palatal length	4.32
2. Audito-basal length	4.55
3. Length lower tooth row	4.97
4. Zygomatic width	5.11
5. Length upper tooth row	5.39
6. Orbital width	5.62
7. Mastoid width	7.10
8. Length of nasals	9.89
9. Antler length, adults	13.78
10. Antler length, 95 adults and 13 juveniles	18.66

The first six variates listed show about average variability, and such differences in V as exist show moderate increase in this characteristic in the order of listing. The seventh and eighth variates must be considered markedly variable for an adult sample of one sex of a single subspecies. In the case of length of nasals the variability is doubtless influenced by variability of suture form, always rather high. The figure for adult antler length is very high. Biologically there are doubtless two factors here: (1) variability proper, which is evidently high for this dimension; and (2) the fact that antler size in individual deer does not, like the other dimensions listed, reach a maximum and then remain nearly constant but declines markedly in adults past their prime. To get a measure strictly of variability for this dimension it would be necessary for all the animals observed to have been in the same year of their age and merely selecting adults is not adequate. The last figure given has no real usefulness; for to the two factors just mentioned is added marked heterogeneity by mixing in some juvenile specimens, an unwarranted procedure.[2] This is stressed to point out once more a fact that cannot be overemphasized: that the mathematical procedures in this book are not mystic formulas and have in and for themselves no usefulness or validity except as they help to record and interpret useful and valid zoological facts. They are not to be viewed as producing numbers important per se, but only in the light of their zoological (and logical) meaning.

[1] V's should generally not be published without also giving other constants of the distribution, and we do so only for the sake of illustrating this one point. Phillips also gives N, M, σ, observed range, and dimension of type, an unusually good publication of numerical data in its most usefully complete and yet compact form. It should be noted, however, that he quotes an incorrect definition of the standard deviation and that there are certain unnecessary and confusing irregularities and ambiguities in his data and conclusions.

[2] It might be a useful thing to calculate the mean length of antlers in all males of the subspecies, regardless of age. This could be done only by

This is a usual value for such a variate in a relatively unified sample. It shows this dimension to have about average variability.

The interesting series of values in Example 41, one of the few such series available in the published literature, also shows about how values of V commonly run.

The means for the dimensions of Example 41 run from 4.27 to 49.8. The second dimension listed has a mean (26.65) about three times that of the third (8.31), but their V's (4.55 and 4.97) and variability are almost the same. This tends to corrob-

EXAMPLE 42.—N AND V FOR WIDTH OF M^1 IN SAMPLES OF THE TREE-HYRAX, *Dendrohyrax dorsalis emini*
(Calculated from measurements by Hatt 1936)

Sample	N	V
1. Males only, 1 locality, Niapu	10	4.2
2. Combined males from 4 localities, Niapu, Akenge, Gamangui, and Ngayu	16	4.0
3. Females only, 2 localities, Niapu and Medje	8	3.1
4. All specimens, males and females, 5 localities	24	3.8

Sample 1 of this example is as homogeneous as a zoological sample could well be. The specimens are all of about the same age (adult), are of one sex, come from a single locality, and belong to a single subspecies. V is relatively low, as would be expected because it is strictly a measure of variability, other causes of dispersion being excluded by the purity of the sample. Even within the pure sample this is one of the least variable dimensions, as other measurements given by the original author show. Sample 2 is less homogeneous as to locality, but all the specimens were taken within the range of a single race and the resulting V is not significantly different from that of the entirely homogeneous sample. The females are, in this dimension, even less variable than the males, probably a real sex difference although proof of this would involve analysis not pertinent in this chapter. The combined sample of males and females is slightly (not significantly) less variable than a sample of males alone. The reason for this is that the females are in this case less variable than the males without differing significantly in absolute size. If the females did differ in mean size (in this dimension) from the males, the V of the combined sample would be higher than for either sex alone and would not be exclusively a measure of variability but of this plus sex divergence.

---

measuring adults and juveniles in the same proportion in which they usually occur in nature, which certainly is not 95:13 as in this sample. Even in such a case, however, V would not really measure variability in a biological sense and would have little value.

orate the belief that this measure is a valuable one even for widely unequal and for nonhomologous but broadly analogous dimensions.

Example 42 shows V's for relatively small samples, such as are more common in zoology.

As is discussed at greater length in Chap. IX, it is practically impossible to get a paleontological sample as homogeneous as a well-selected group of recent animals. The sexes usually cannot be separated, age groups may not be clear, and there will in many cases be unavoidable small differences in geological age and in race. It may therefore be expected that V's based on fossil material will run higher than those based on well-recorded and -selected recent material. They do so in general, but the differences are not great if the paleontological sample is also as well selected as the exigencies of collection permit. Example 43 is typical of many values of V based on fossils.

EXAMPLE 43.—COEFFICIENTS OF VARIABILITY OF LINEAR DIMENSIONS OF TEETH OF THE EXTINCT MAMMAL *Litolestes notissimus*, SEXES INDISTINGUISHABLE, DERIVED FROM A SINGLE QUARRY BUT NOT EXACTLY THE SAME STRATIGRAPHIC LEVEL

(Original data)

Variate	V
Length P_4	6.6
Width P_4	7.1
Length M_1	5.1
Width M_1	6.5
Length M_2	5.6
Width M_2	5.6
Length M_3	6.9
Width M_3	4.3

One of the essentials in good taxonomy is to select characters that are relatively little variable within a taxonomic group, for taxonomic comparisons are more easily and more reliably based on these than on highly variable characters. The coefficient of variation is very useful as a guide in the selection of such characters, too often merely guessed at or accepted with no real criterion. For instance, in dealing with fossil mammals of the family Ptilodontidae, order Multituberculata, there is a strong temptation to use the third upper premolar in taxonomic definition because it shows such strong and clear-cut differences. In a sample of *Ptilodus montanus*, a member of this group, the V's for other linear tooth dimensions are for the most part around 5 or 6, with

only one as high as 10; but for length of P^3 this figure is 18.5. Obviously the tooth is extremely variable and is not a good taxonomic guide, or expressed in other words, its variations reflect intraspecific variability and not reliable taxonomic differences.[1]

A measure of relative dispersion that is implied in some zoological procedures is the ratio of the extreme observed values. Such a statement as that the largest individual of a species is 15 per cent larger than the smallest is a use of this sort of measurement of relative dispersion. The same sort of measurement is implied in comparing species when, for instance, it is given as a specific character that the type of one species is in some dimension 20 per cent larger than the type of another species. Stated in this usual way with no other data, the implication is that individuals within one species do not differ so much in size as do the types of the two species. Usually no evidence for this implied belief is given; and it is, in fact, usually wrong.

The best way to record this measure of relative dispersion would be (100 × highest observed value)/lowest observed value, giving the highest as a percentage of the lowest value. Of all measures of relative dispersion, this is undoubtedly the worst; and it is discussed only to draw attention to the fact that it is so poor and yet is implied in so much zoological work. It is also important to note that the figures thus obtained even for good samples of relatively low variability are usually higher than most zoologists seem to know or suspect. It is commonly stated or implied that 115 per cent, or 115, is a high value for a single species (or other natural group). Actually it is an unusually low

EXAMPLE 44.—$\frac{100 \times \text{HIGHEST OBSERVATION}}{\text{LOWEST OBSERVATION}}$ FOR DATA OF EXAMPLE 41, WITH CORRESPONDING V'S

	V
1. 125	4.32
2. 127	4.55
3. 130	4.97
4. 129	5.11
5. 142	5.39
6. 129	5.62
7. 143	7.10
8. 179	9.89

[1] The biological explanation of this extraordinarily high V seems to be that P^3 in this family is not functional and is being lost. Organs in this condition are usually extremely variable.

value. Example 44 illustrates this, as well as the irregularity of such a measurement.

These samples are large for zoology (91 to 96 individuals), and so the observed limits are somewhere near the true limits. For the more usual small samples such a measure bears little rational relationship to the real variability.

If such a measure were really desirable, it would be much better to calculate it as $100(M + 3\sigma)/(M - 3\sigma)$, using the best available approximation of the true range limits. This measure is not in use, nor is there any justification for it except as an improvement on the one that is in use. If σ is calculated, it is easier and better to use V.

Several other measures of relative dispersion are in occasional use and may be mentioned. They have no advantages over V; and it is urged that the latter alone be employed, keeping all such data comparable. Formulas for some such measures in use are

(1) $\dfrac{100 \times \text{M.D.}}{\text{Mean}}$

(2) $\dfrac{100 \times \text{M.D.}_{\text{median}}}{\text{Median}}$

(3) $\dfrac{100 \times \text{Q.D.}}{\text{Median}}$

(4) $\dfrac{100(Q_3 - Q_1)}{Q_3 + Q_1}$ also sometimes modified as $\dfrac{100(Q_3 - Q_1)}{2(Q_3 + Q_1)}$

The values of these various coefficients for a single distribution are given for comparison in Example 45.

EXAMPLE 45.—VARIOUS MEASURES OF RELATIVE DISPERSION CALCULATED FOR THE DISTRIBUTION GIVEN IN EXAMPLE 36 (PAGE 108)

$$V = \frac{100\sigma}{M} = \frac{100 \times 3.04}{60.43} = 5.03$$

$$\frac{100 \times \text{highest observation}}{\text{Lowest observation}} = \frac{100 \times 68}{52} = 131$$

$$\frac{100(M + 3\sigma)}{M - 3\sigma} = \frac{100 \times 69.55}{51.31} = 136$$

$$\frac{100 \times \text{M.D.}}{\text{Mean}} = \frac{100 \times 2.46}{60.43} = 4.08$$

$$\frac{100 \times \text{M.D}_{\text{median}}}{\text{Median}} = \frac{100 \times 2.46}{60.41} = 4.07$$

$$\frac{100 \times \text{Q.D.}}{\text{Median}} = \frac{100 \times 2.20}{60.41} = 3.64$$

$$\frac{100(Q_3 - Q_1)}{Q_3 + Q_1} = \frac{100 \times 4.39}{120.77} = 3.64$$

CHAPTER VII

THE NORMAL CURVE

The normal curve and the normal distribution have been briefly characterized in previous pages, and it has been mentioned that their great importance for the zoologist is two-fold: they approximate many actual distributions that he encounters, and they provide the background for determinations of probability and for comparisons of various measures. It is now necessary to consider in greater detail the way in which the normal distribution is used as an approximation of real distributions, how normal curves are specified, and what the properties are that underlie the various operations in which it is involved.

THE NORMAL CURVE AND REALITY

It is convenient to call some real distributions "normal" and to speak of many others as if they were merely imperfect representations of what would be normal distributions if only our data were complete. It is, however, necessary to bear in mind that this is only a manner of speaking, adopted to save constant verbose qualification and circumlocution, and to remember that no real distribution, even if all existing data on it were at hand, can ever actually follow the normal curve or any other purely mathematical concept. To the mathematician and statistician these mathematical abstractions are the ideal and are perhaps in a philosophical sense the ultimate reality. To them, the actual concrete data of observation are only an imperfect approach toward such an abstraction. The zoologist has and should have a very different point of view. His observed distributions are the reality with which he is concerned, and the normal (or any other) curve is only an imperfect but helpful means of describing these and of otherwise drawing useful deductions from them.

A curve as a representation of a frequency distribution is, at its simplest, a limit that a real distribution would reach if the number of observations were infinite. Obviously the number of

observations is not infinite. Within limits it can often be increased at will, but it is rigidly restricted by the fact that there is always a finite, even though usually large, number of possible observations, and it is more loosely but also inevitably restricted by the fact that it is seldom really practical to make more than a small fraction of all possible observations. Thus in fact a distribution never could become literally normal.

This discrepancy would involve little or no difference between the zoological and the mathematical point of view if it were a fact that all possible observations, if they were made, would ever be distributed so as to differ from the normal only in being finite rather than infinite in number. If, in other words, a frequency polygon of all existing values of a variate differed from a normal curve only in being outlined by a succession of exceedingly short straight lines instead of by a literal curve, then it would be proper to think of the distribution as merely an imperfect approach toward the normal curve. This, however, is also false. No matter how large a series of observations are made, the normal curve is not approached in this simple way by any distribution based on real things.

The fact that the normal curve approximates real distributions is largely a matter of empirical observation. It has been repeatedly tried and has been found rather closely to approximate certain types of distributions more often than not. It has never been found to be exactly like a real distribution, even in the sense of the preceding paragraph. There is no reason why the normal curve should not be a good approximation, on theoretical and logical grounds there is some positive probability that it would be, and from observation it is a fact that it often is. It does not, then, matter in use that exact equivalence does not exist. The absence of any exact relationship is easy to prove. In the normal curve, the value of f is always greater than 0, whatever the value of X. For very high values of X, f may be very small; but it is always a definite positive figure. This is good mathematics, but it is utter nonsense as zoology. If animal size were distributed normally, this would mean that we had only to collect enough specimens to find a mouse with ears a mile long or a snake long enough to girdle the earth at the equator.

To relate the normal curve to reality it is, then, necessary to make an assumption that is bad mathematics but good zoology, that is, that small fractional values of f do not exist at all. When

the normal curve tails off above and below the mean and the indicated class frequencies become less than unity even for a large total frequency, it is correctly assumed in zoological practice that f becomes 0 and that the curve touches the X-axis and ends, although it never does so mathematically. The approximation of the curve to a real distribution is accepted as valid only within a reasonable distance (about 3σ) of the mean.

The approximation of such a curve to a distribution of observations actually made provides a good numerical method of describing all of these available observations as a unit, instead of giving each one as an isolated thing. It also has another and still more important use. The observations that have been made constitute a sample drawn from a larger unit, from the total of all possible homologous observations. In studying a sample, it is desired to draw from it information regarding this larger unit, the population from which the sample is drawn. For instance if 10 specimens of one subspecies of rodent are collected and measured, it is desired to infer from these measurements what the distribution of these dimensions is in all the rodents of that subspecies, wherever they may be. All existing members of the subspecies together form the population, which cannot be made available as a whole, and the 10 individuals actually caught and measured are a sample of the population.

In order to draw such inferences about something not available from a relatively small sample that is available, it is necessary to assume some definite relation between the two. The theory and practice of sampling, discussed in more detail in Chap. IX, assume that the distribution of the sample is similar to that of the population. Its differences from and irregularities with respect to the population depend on the chances of collecting the data, and the probability that the numerical characters of the population lie within a certain range of those of the sample can be calculated from the properties of the normal curve. Moreover, if a curve can be fitted reasonably well to the sample, there is a definite and usually high probability that the population fits in much the same way a similar curve with higher frequencies. A mathematical curve is thus a means of arguing from the particular to the general, of deducing the properties of the population from a given sample, and for this purpose the normal curve is by far the most useful. The procedure and logic involved in this operation are to calculate from the sample numerical characteristics, called

constants or parameters, of which M and σ are the best and most important, that determine the form and position of a normal curve. The curve so determined is assumed hypothetically to represent the distribution of the population.[1]

PARAMETERS OF THE NORMAL CURVE

The mathematical equation of the normal curve is[2]

$$y = \frac{N}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}}$$

[1] This discussion differs from usual statements of statistical procedure in two possibly subtle but important respects. In the first place, statisticians usually speak of the "infinite population." For the zoologist, at least, there is no such thing as an infinite population, and he has no interest in this useless abstraction. He is interested in the real population, which exists even though all of it is not available and which obviously is not infinite. He is not, as statisticians usually say they are, calculating the properties of an infinite population. In the second place, statisticians usually state that they calculate the probability that a sample is drawn from a population with certain mathematical characteristics. On the contrary, what is actually done is to use the sample to set up a hypothetical distribution and to estimate the probability that a real and finite population is correctly represented by it. Correct comprehension of these two points will help to avoid errors in logic and in procedure.

R. A. Fisher, some of whose important publications are cited in the bibliography at the end of this volume, is one of the few statisticians who consistently maintain a point of view valid and useful to zoologists; his work is of outstanding value for anyone working in this field, although it is at times hard to follow without more mathematical knowledge than a zoologist is likely to have or really to need. Fisher has recently adopted the convention of using Greek letters for the parameters of theoretical curves and italic letters for the calculated constants that are estimates of these. For instance he uses s for a calculated standard deviation and σ for the unobtainable standard deviation of which s is an approximation. For our purposes it seems sufficient to state the distinction and to assume that it is understood hereafter. It hardly seems necessary in zoological practice to have a whole set of symbols for theoretical values never actually obtained or used. It also seems less confusing to use symbols in the more usual way even though their significance is viewed in a somewhat different light from that most common. We therefore do not follow Fisher's convention or use his dual system of symbols; for instance, we use σ for any standard deviation, as is usually done. It is, however, desirable to direct attention to the more important differences between our usages and Fisher's, since his is the work to which the student is best referred for further information on various points.

[2] The equation is derived by taking a symmetrical binomial series ($p = q$)

in which x and y are used in the usual mathematical way, corresponding in a frequency distribution to d and f, respectively; N is the total frequency; σ is any number and is the standard deviation of the corresponding distribution; π is the familiar value 3.1416, and e is another natural constant (base of Napierian logarithms) with the value 2.7183.

The numbers 2, π, and e that occur in the equation are invariable in value. N and σ may take any value, and for each specified set of values for these there is a different normal curve. If we know these values, it is possible to determine a value of y for any value of x or one of x for any of y and to plot out the whole curve. In other words, the values of N and σ determine the whole shape of the particular normal curve with which one has to do. Such elements in an equation are called constants or parameters—we shall use the name parameter because "constant" has so many other vernacular and technical meanings that it is confusing.

Given an actual sample, N, the number of observations in the sample, is known, and σ can be calculated as shown in the preceding chapter. These figures for the actual sample are also parameters of a normal curve, and the particular normal curve that they specify is taken as an approximation of the distribution represented by the sample.

According to its equation, the curve is taken as centered on the y-axis (the f-axis of a frequency distribution) and the values of x are distances from this center (they are d, not X, of a frequency distribution). To use the curve it is necessary to know also at just what point on an absolute scale (a scale of measurements or counts, X, as observed) is the center of the curve. This point is the mean M, and M is therefore the third and last parameter that must be calculated to specify the exact form and position of a normal curve.

---

and finding the equation of the curve that it approaches as a limit, a mathematical operation unnecessary here but mentioned in order to affirm that the complicated equation for the curve does have a rational derivation. Yule (see Bibliography) gives the operations involved in this derivation. In practice it is not necessary to solve this difficult equation or to plot the curve from it. Statisticians have calculated tables of all the necessary values so that zoologists and others using the curve as a practical means toward an end are spared any operations with the equation and can simply refer to the tables.

In using the curve one must thus have a frequency parameter, the best of which is N, a dispersion parameter, the best of which is σ, and a central tendency parameter, the best of which is M. It is possible to use other parameters; but generally their relationship to the curve is less simple and useful than the relationship of N, σ, and M. Another possible frequency parameter is the frequency of the middle class; and this does have an important special use, as shown later in this chapter. If the curve is really normal, median and mode are both equal to the mean, so that median and mode as central tendency parameters are of no use for the normal curve itself but are useful for showing the difference (by skewing) of an observed distribution from the normal distribution that most nearly approximates it. Mean deviation and semi-interquartile range could be substituted for σ as dispersion parameters because in the normal curve they have a constant relationship to σ;[1] but for this same reason there is no particular reason for using them, and there are good reasons for not using them.[2]

There are certain natural sets or families of parameters, and if the distribution is approximately normal the pertinent numerical data for any series of observations are best given by such a set.

1. N, M, and σ.
2. N, M, and M.D.
3. N, median, and $\text{M.D.}_{\text{median}}$.
4. N, median, and Q.D.

Of these the first is far the best, and the others should be used only in special cases. If they are used, it is highly desirable not to mix up parameters of different families, as some zoologists have done. It is, for instance, confusing and bad practice to use N, M, and Q.D., or N, median, and σ, for the Q.D. is measured from quartiles, of which the median, not the mean, is one, and σ is calculated from deviations around the mean, not the median.

For each parameter except N its standard error[3] should be given—the calculation and significance of these are explained in

[1] M.D. = $.7979\sigma$ and Q.D. = $.6745\sigma$ in normal distributions.

[2] They could, indeed, be used in comparison with σ to measure skewness in an only approximately normal distribution, but this is so much more simply done by a comparison of mean and mode that this use, also, is impractical.

[3] Or probable error, but the standard error is much to be preferred.

the next chapter. In addition to the parameters, it is generally useful to give the observed range and the coefficient of variation. Although so poor as an approximation, the observed range does convey some rough information about the sample in the most obvious way; and V is a highly useful figure in dealing with linear dimensions and some other variates.[1] The best set of data on a given sample thus comprises N, R (for observed range), M and its standard error, σ and its standard error, and V and its standard error. For the data of Example 36 (page 108), this set is as follows:

Variate	N	R	M	σ	V
Tail length	86	52–68	60.43 ± .33	3.04 ± .23	5.2 ± .4

This conveys all the useful information contained in the complete list of 86 original measurements, it expresses it much more succinctly, and it gives it in the clearest and most useful possible form.

If N is so low that the values of σ and V are too unreliable to be useful, they may be omitted. Even in such cases, however, it is usually best to give some measure of dispersion for purposes of comparison with other small samples, and it is always well to give either σ, $\Sigma(fd^2)$, or the variance, *i.e.*, $\Sigma(fd^2)/N$, no matter how small the sample is.

AREAS AND ORDINATES

The most important property of the normal curve is that the part of the curve beyond a distance of a given multiple of σ from the mean always encloses the same fraction of the total area, whatever the parameters of the particular curve in question may be. If, on a normal curve, vertical lines are drawn at points corresponding on the X-scale to $(M + \sigma)$ and $(M - \sigma)$, the area between these lines is always 68 per cent of the total area, and the area beyond them, in the two tails of the curve, is 32 per cent of the total area. If the lines are drawn at $(M + 2\sigma)$ and $(M - 2\sigma)$, the area between them is 95.5 per cent of the total, that outside of them 4.5 per cent. If the lines are at $(M + 3\sigma)$ and $(M - 3\sigma)$,

[1] V is also in one sense a parameter of the normal curve, but it is not a simple one since it combines two of the primary parameters (M and σ) and its usefulness is not as a parameter.

the area between is 99.7 per cent and that outside is .3 per cent. In accordance with the nonmathematical assumption that small fractional frequencies are nonexistent, this means in practice that the whole area of the curve is between $(M + 3\sigma)$ and $(M - 3\sigma)$. This is the basis of the statement (page 118) that these are good approximations of the limits of the real range in a roughly normal distribution.

The area enclosed by any part of the curve and verticals to the base line (X-axis) is proportional to the frequency of the group enclosed by such verticals. In a large number of observations distributed normally, 68 per cent of the observations would be greater than $(M - \sigma)$ and less than $(M + \sigma)$. This is absolutely true of the normal distribution, and it is for practical purposes sufficiently close to the truth in any real distribution approximately normal in form. The conception of probability as applied to sampling and various other operations is based on this property of the curve. If a population is approximately normal, the probability that a specimen selected from it by chance will lie within a certain distance of the mean is proportionate to the area enclosed by verticals at this distance from the mean on a normal curve. The chances that a single specimen taken at random from such a population will be between $(M - \sigma)$ and $(M + \sigma)$ are 68 in 100; the chances that it will not are 32 in 100. Expressed in another way, if a sample of 100 is taken, about 68 will probably be within those limits and about 32 outside them; in a sample of 50 specimens about 34 will be within the limits and about 16 outside them; etc.

Still another expression and use of this relationship involve comparing a specimen with a population. The parameters of the population have been estimated from a sample drawn from it and a value of σ obtained. Another specimen is found, and it is desired to know whether it does or does not belong to the population, for instance, whether it is of a certain species. Its deviation d from the mean for the sample is determined and then is expressed in terms of σ, by the quotient d/σ. The probability that it belongs to the population can then be read from a table of areas for various values of d/σ. If this quotient is 1, the probability that it does belong to the population is proportionate to the area outside the limits $(M + \sigma)$ and $(M - \sigma)$, or is 32 in 100. This is a large probability, so that the specimen may very well belong

to the population in question. If d/σ is 4, the probability is .006 in 100, so small as to be nonexistent in practice, and it is certain that the specimen does not belong to the population.

Table I gives the areas between and outside of limits expressed in terms of σ, or by d/σ.

TABLE I.—AREAS OF THE NORMAL CURVE

d/σ	Percentage of area between limits	Percentage of area outside limits
.00	0	100
.10	8	92
.20	16	84
.30	24	76
.40	31	69
.50	38	62
.60	45	55
.70	52	48
.80	58	42
.90	63	37
1.00	68	32
1.10	73	27
1.20	77	23
1.30	81	19
1.40	84	16
1.50	87	13
1.60	89	11
1.70	91.1	8.9
1.80	92.8	7.2
1.90	94.3	5.7
2.00	95.5	4.5
2.10	96.4	3.6
2.20	97.2	2.8
2.30	97.9	2.1
2.40	98.4	1.6
2.50	98.8	1.2
2.60	99.1	0.9
2.70	99.3	0.7
2.80	99.5	0.5
2.90	99.6	0.4
3.00	99.7	0.3
3.25	99.88	0.12
3.50	99.96	0.04
3.75	99.98	0.02
4.00	99.994	0.006

This table combines the areas on both sides of the mean. For instance, for $d/\sigma = 1.00$, the 68 per cent of the area between the limits is composed of 34 per cent above the mean and 34 per cent below it. The 32 per cent outside the limits consists of 16 per cent in the tail of the curve to the left, below -1σ, and 16 per cent in that to the right, above $+1\sigma$. It is sometimes desirable to know the chances that a specimen will differ from the mean in a given direction, and not in either direction as is given by this table. The areas appropriate for this problem are on one side of the mean only, and the percentages are therefore just half those given in the table. Such tables are, in fact, frequently given in this form, but for zoological purposes it is less generally useful; and if such a figure is desired it is easy simply to divide an entry of Table I by two.

Note that these values do not hold good for some uses with very small samples. Procedure and values for these special conditions are considered in Chap. XI.

Another occasional problem for which tables are available is to obtain the percentage of the area above or below a single limit, for instance, above or below $(M + \sigma)$. The area outside $(M + \sigma)$ and $(M - \sigma)$ on both sides is seen by Table I for the value $d/\sigma = 1.00$ and is 32 per cent. Then the area above $(M + \sigma)$ will be the corresponding value for one side of the curve only and will be 32 per cent/2 = 16 per cent. All the rest of the area is below this; so the area below $(M + \sigma)$ is 84 per cent.[1] This, too, is a type of normal curve area seldom used and easily calculated from the data of Table I, and so no separate table is given for it.

Just as these areas are proportionate to frequencies, the height of any vertical line from the curve to the base line (X-axis) is proportionate to the frequency at that point. These vertical lines are parallel to the y-axis (f-axis of a frequency distribution) and are in fact ordinates of the curve. The longest such ordinate is at the point of highest frequency, the mode of a distribution, which in the normal curve is also the mean. If this maximum ordinate at the mean is taken as unity, the proportionate length of an ordinate at any given distance from the mean is the same for

[1] Consisting of the area between $(M - \sigma)$ and $(M + \sigma)$, seen by the table to be 68 per cent, and the area below $(M - \sigma)$, which is 16 per cent. 68 + 16 = 84. The result is the same as the operation in the text, which is 100 − 16 = 84.

any normal curve, whatever its parameters. These values are given in Table II.

TABLE II.—ORDINATES OF THE NORMAL CURVE

d/σ (abscissas in terms of σ, with the origin at M)	Proportion of the corresponding ordinate to the maximum ordinate y_0, at the mean
0	1.000
.1	.995
.2	.980
.3	.956
.4	.923
.5	.883
.6	.835
.7	.783
.8	.726
.9	.667
1.0	.607
1.1	.546
1.2	.487
1.3	.430
1.4	.375
1.5	.325
1.6	.278
1.7	.236
1.8	.198
1.9	.164
2.0	.135
2.1	.110
2.2	.089
2.3	.071
2.4	.056
2.5	.044
2.6	.034
2.7	.026
2.8	.020
2.9	.015
3.0	.011
3.5	.0009
4.0	.0003

Since the normal curve is symmetrical, the ordinate at $(M + 2\sigma)$, shown by the table to be .135 times the maximum ordinate, will be the same as that at $(M - 2\sigma)$. The maximum ordinate is conveniently designated as y_0, signifying the value of y in the normal equation that corresponds with $x = 0$, or a devia-

tion from the mean of 0 in a distribution. The conversion factors for other ordinates, their proportion to y_0, are symbolized by a.

CURVE FITTING

These proportions of the ordinates in a normal curve are used to calculate the class frequencies of a normal distribution nearly

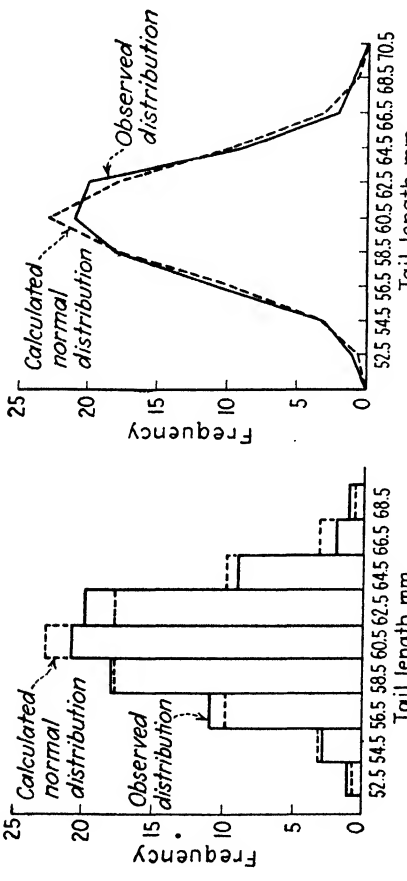


FIG. 17.—Approximation of a theoretical normal distribution to an observed distribution. Tail length in the mammal *Peromyscus maniculatus bairdii*. The continuous lines represent the observed facts, as given in Example 36, and the broken lines an approximately equivalent normal distribution, as calculated in Example 46. *A*, histogram. *B*, frequency polygon. The histogram permits clearer visual comparison of the observed and theoretical frequencies. The polygon more clearly shows the approximation to a normal curve.

corresponding with an actual distribution. This in turn makes it possible to draw up a histogram or frequency polygon for such a normal distribution or to draw its limiting normal curve without the effort of plotting the latter from its complex equation. By comparing the numerical values or graphs with those for the

observed distribution, the important resemblances and differences of the two can be seen.

In order to do this, it is evidently first necessary to get a value for y_0, the frequency in the class that includes the mean. This is done from the following relationship, always true of the normal curve.[1]

$$y_0 = \frac{N}{\sigma\sqrt{2\pi}}$$

The value of $\sqrt{2\pi}$ is 2.506628—the value 2.51 is usually sufficiently refined.

The frequency of this mean class in actual distributions is, however, also proportionate to the class interval, so that this value i must be inserted in the equation, giving

$$y_0 = \frac{Ni}{\sigma\sqrt{2\pi}}$$

The value of any ordinate other than y_0 will then be

$$y = ay_0$$

a being the proportionate value given in Table II and y being a value of the ordinate corresponding to a given d/σ. In practice this must be calculated for each class midpoint, obtaining a value of d/σ for each of these midpoints. The method of calculation is shown in Example 46, and the resulting approximation of a normal to a real distribution is also graphically shown in Fig. 17.

This agreement between the normal and the observed distributions is fairly good, about as close as it will be with most zoological data. A method for measuring the degree and significance of this agreement is given in Chap. XIV.

It is also possible to calculate a theoretical normal distribution not from the maximum class frequency (y_0) and the proportionate sizes of ordinates but from the total frequency and proportionate areas (as in Table I). This method gives slightly different results; but it has no particular advantage over the ordinate method and is somewhat more tedious, so that it is not necessary to give it

[1] The derivation of this equation may be taken for granted here. It will be noted that it is the multiplier in the equation for the normal curve (page 132).

EXAMPLE 46.—FITTING A THEORETICAL NORMAL DISTRIBUTION TO THE OBSERVED DISTRIBUTION OF EXAMPLE 40 (PAGE 117)

Tail length, mm.	Class midpoints	d	d/σ	Theoretical frequencies		Observed frequencies (f)
				a	ay_0	
52–53	52.5	8	2.7	.026	.6	1
54–55	54.5	6	2.0	.135	3.1	3
56–57	56.5	4	1.3	.430	9.8	11
58–59	58.5	2	.7	.783	17.9	18
60–61	60.5	0	.0	1.000	22.9	21
62–63	62.5	2	.7	.783	17.9	20
64–65	64.5	4	1.3	.430	9.8	9
66–67	66.5	6	2.0	.135	3.1	2
68–69	68.5	8	2.7	.026	.6	1

$N = 86 \qquad \sigma = 3.0 \qquad i = 2$

$M = 60.4$, but for purposes of calculating the theoretical normal distribution it is taken as at the midpoint of its class = 60.5.

$$y_0 = \frac{N \times i}{\sigma\sqrt{2\pi}} = \frac{86 \times 2}{3.0 \times 2.507} = \frac{172}{7.521} = 22.9$$

The values of a are read from Table II opposite the values here obtained for d/σ. Note that in these calculations the frequency of each class is assumed to be at its midpoint and the distribution is assumed to be symmetrical. It is, of course, unnecessary to calculate values for classes both above and below the middle class, since these are symmetrical.

here—it can be found in some textbooks of statistics (*e.g.*, Arkin and Colton) or can be worked out as an exercise.

These methods are only approximate. They do not inevitably lead to the theoretical normal distribution closest to the observed distribution but to one that is close enough for practical purposes. The relationship between the two is only approximate in any case, and great precision is usually a waste of effort. More exact curve fitting is a very difficult and complicated process even for a normal curve and exceedingly so for curves differing from the normal in any respect. Zoological data are seldom extensive enough to warrant such treatment, and even if they were the results would seldom justify the labor involved. Curve fitting more complicated than by the simple method just explained is one of the statistical procedures not necessary in zoology.

SKEWNESS

It has been mentioned that distributions approximately normal may yet differ from the normal in two significant ways, skewness and kurtosis. Skewness is a deviation from the normal by asymmetry of the distribution, which falls off more rapidly on one side of the high point than on the other. Its high point, the mode, no longer coincides with the mean, as in the normal distribution, for the mean is strongly affected by extreme values and the mode is not affected at all by them. Advantage is taken of these facts to devise a measure of skewness. The distance between the mean and the mode is such a measure, but it is not satisfactory. In the first place it depends not only on skewness but also on the degree of dispersion, and in the second place it is an absolute figure and hence is not readily used in comparisons of variates of different magnitudes. Both these objections are met by dividing the difference between these averages by the standard deviation, giving $S_k = (\text{mean} - \text{mode})/\sigma$, S_k being the usual symbol for this measure.

It was shown in Chap. V, however, that exact calculation of the mode is impractical and that it can most easily be approximated by the relationship

$$\text{Mode} = 3 \text{ median} - 2 \text{ mean}$$

Substituting this in the equation for skewness gives

$$S_k = \frac{\text{mean} - (3 \text{ median} - 2 \text{ mean})}{\sigma} = \frac{3(\text{mean} - \text{median})}{\sigma}$$

This is the best form for calculating this measure.[1] In a symmetrical distribution mean and median are equal; so S_k is 0. In a right-skewed distribution there are extreme values to the right of the averages; hence the mean is larger than the median, and the coefficient of skewness is positive. Similarly in a left-skewed distribution the coefficient is negative. There is no theoretical limit to the magnitude of the coefficient, but in fact it seldom is less than -1 or more than $+1$ in distributions even distantly approaching the normal.

[1] There is also a measure of skewness based on the quartiles and given by the expression $(Q_3 + Q_1 - 2 \text{ median})/\text{Q.D.}$ Its use is not recommended.

EXAMPLE 47.[1]—HEAD BREADTH OF CAMBRIDGE (ENGLAND) STUDENTS IN INCHES, EXEMPLIFYING A TYPICAL POSITIVELY SKEWED FREQUENCY DISTRIBUTION

(Data from Macdonell 1902)

X	f	d_A	fd_A	fd_A^2
5.5	3	−5	− 15	75
5.6	12	−4	− 48	192
5.7	43	−3	− 129	387
5.8	80	−2	− 160	320
5.9	131	−1	− 131	131
			− 483	
6.0	236	0	+1,098	
6.1	185	1	185	185
6.2	142	2	284	568
6.3	99	3	297	891
6.4	37	4	148	592
6.5	15	5	75	375
6.6	12	6	72	432
6.7	3	7	21	147
6.8	2	8	16	128
	$N = 1{,}000$		$\Sigma(fd_A) = 615$	$\Sigma(fd_A^2) = 4{,}423$

[1] Anthropological data are generally not used in the examples, but it is desired to show calculation for one example with a large N, here 1,000, and also to exemplify a type of moderately skewed curve very common in zoology by a distribution large enough to show the form clearly and with no question. In any case Cambridge undergraduates are zoological materials.

$$i = .1$$
$$c_1 = +615/1{,}000 = +.615$$
$$c = .615 \times .1 = .0615$$
$$M = 6.0 + .0615 = 6.062$$
$$c_1^2 = .3782$$
$$\sigma = .1\sqrt{\frac{4{,}423}{1{,}000} - .3782} = .1\sqrt{4.0448}$$
$$= .201$$
$$\text{Median} = 5.95 + \frac{23.1}{236} = 6.048$$
$$S_k = \frac{3(\text{mean} - \text{median})}{\sigma} = \frac{+.042}{.201} = +.21$$

This distribution is nearly normal, certainly near enough to use the parameters of the normal curve in dealing with it, but it does have a moderate positive skew. The approximate mode is 6.02, falling in the 6.0 class, and there are 8 classes above and only 5 below this, so that the skewness is visible on inspection.[1]

[1] Skewness may, however, be present and calculable without being obvious as it is here.

In the distribution of Example 46, the mean, as already calculated for the same data in Chap. VI, is 60.43, the median 60.41, and σ 3.04. Hence

$$S_k = \frac{3(60.43 - 60.41)}{3.04} = \frac{.06}{3.04} = +.02.$$

This is too small to be significant, and the distribution is essentially symmetrical. Example 47 shows the calculation of this coefficient for a distribution that is clearly skewed.

Moderately right-skewed distributions of this type are common in zoology, and, although an adequate census has not been made, they appear to be more common than are left-skewed distributions. In biological terms a right-skewed distribution indicates that large variants are more common than small variants within the sample representing the population. Although there are abundant exceptions, this seems to be a general tendency of morphological characters in zoology.

With the smaller samples usual in zoology, the values of S_k are inevitably somewhat erratic, and a general tendency can only be detected by comparing a considerable series of such values.

EXAMPLE 48.—COEFFICIENTS OF SKEWNESS FOR A SERIES OF SMALL SAMPLES OF TEETH OF THE EXTINCT MAMMAL *Litolestes notissimus*
(Original data)

Variate	N	S_k
Length P_4	10	+.19
Width P_4	11	−.52
Length M_1	19	−.21
Width M_1	19	+.52
Length M_2	28	−.17
Width M_2	29	0
Length M_3	24	+.45
Width M_3	24	+.19
Length P^4	7	0
Width P^4	7	−.51
Length M^1	10	+.29
Width M^1	10	+.83
Length M^2	13	−.18
Width M^2	13	−.54
Length M^3	8	+.60
Width M^3	9	+.43
Mean S_k: +.09		

Example 48 shows typical values for such a series with the usual small samples of zoology and paleontology.

It may be noted that the assumption on which the use of V and the study of variability in general are based involves a constant tendency for such characters to show a small positive skew. This assumption is that the dispersion is proportionate to the absolute value of the variate. If this is true as between different samples and different variates, it should be true also within a single distribution. The absolute dispersion should tend to be, or for an average of many different homologous distributions should be, greater for higher values of the variate than for lower and should increase steadily from the left-hand end of the graphic distribution through to the right-hand end. Since the values thus tend to be spread farther on the absolute scale above the mode than below it, a positive skew is involved. A given value below the mode will bear the same ratio to the mode that the mode will bear to a corresponding value (one of theoretically the same frequency) above the mode, instead of being simply equidistant from the mode as in the normal curve. From this relationship of ratios as opposed to simple linear distances, it follows that the geometric mean and the mode will tend to coincide, whereas in the normal distribution the mode and arithmetic mean coincide (the geometric mean being smaller than the arithmetic for any distribution).

It is probable that the slightly skewed form of curve thus determined is a better theoretical description of most zoological variates than is the normal curve. Basing calculations on the skewed curve would, however, be exceedingly difficult, and it would almost never really make the results significantly better in practice, however preferable in theory. The skewed curve thus determined differs very little from the normal, so little as to be wholly obscured by chance fluctuations in almost all zoological samples. Even with the large sample of Example 47 (page 144), skewness caused by this relationship is too small to have a really appreciable effect. The geometric mean for this distribution is 6.058, the arithmetic mean 6.062. Rounded to the number of places really accurate and significant these two are exactly the same, 6.1 or 6.06, and there is no demonstrable effect of this sort of skewness in the data. The observed significant skewness of this distribution is thus due not to this phenomenon but to some other biological factor.

KURTOSIS

Kurtosis is the property of being more pointed or flatter than a normal curve with the same parameters. It does not involve symmetry and usually cannot be detected by inspection unless it is very great. Unfortunately the measurement of this characteristic is somewhat laborious, although not unduly complicated. The best coefficient of kurtosis K is the value of the expression[1]

$$\frac{\Sigma(d^4)/N}{\sigma^4} - 3$$

It will be remembered that $\sigma = \sqrt{\frac{\Sigma(d^2)}{N}}$, so that $\sigma^4 = \left[\frac{\Sigma(d^2)}{N}\right]^2$, and this expression may sometimes be more advantageous in calculation.

When the coefficient of kurtosis is 0, the distribution is neither more peaked nor flatter than the normal curve and is sometimes spoken of as mesokurtic.

When this value is positive, the distribution is more peaked or sharper than the normal curve and is called leptokurtic.

When it is negative, the distribution is flatter than the normal curve and is called platykurtic.

Platykurtic distributions generally tend to have relatively large σ and V. The flattening usually reflects either high variability or some heterogeneity in the sample. Conversely, leptokurtic curves usually have lower σ and V, and the peakedness generally reflects low variability. The calculation of a coefficient of kurtosis is shown in Example 49.

[1] Sheppard has proposed a correction whereby $\Sigma(d^4)/N$ (known as the fourth moment around the mean) is replaced by $\frac{\Sigma(d^4)}{N} - \frac{\Sigma(d^2)}{N} + \frac{7}{240}$. The difference is usually slight, and it is generally not worth while to make this correction.

General statistical theory involves much consideration and use of a series of parameters called moments. The nth moment about the mean is calculated as $\Sigma(d^n)/N$, subject to various corrections and adjustments. Thus the first moment about the mean (always 0 for the normal or any other symmetrical distribution) is $\Sigma(d)/N$. The second is $\Sigma(d^2)/N$ and is the variance. The fourth moment is used in calculating the coefficient of kurtosis. Despite their theoretical importance and their use in fitting curves, moments other than those already discussed in the text have little practical use for such problems as are treated in this book.

EXAMPLE 49.—CALCULATION OF THE COEFFICIENT OF KURTOSIS FROM DATA OF EXAMPLE 40 (PAGE 117)

Tail length	f	d_A	d_A^4	fd_A^4
52–53	1	4	256	256
54–55	3	3	81	243
56–57	11	2	16	176
58–59	18	1	1	18
60–61	21			
62–63	20	1	1	20
64–65	9	2	16	144
66–67	2	3	81	162
68–69	1	4	256	256
	$N = 86$			$\Sigma(fd_A^4) = 1{,}275$

In terms of class intervals:

$$K = \frac{\frac{\Sigma(fd_A^4)}{N} - c_1^4}{\sigma_1^4} - 3$$

From Example 40:

$c_1 = .06$

$\sigma_1^2 = 2.26$

$$K = \frac{\frac{1{,}275}{86} - (.06)^4}{(2.26)^2} - 3 = \frac{14.8256 - .00001296}{5.1076} - 3 = 2.90 - 3.0 = -.10$$

Note that the formula is valid regardless of the units in which the deviations are taken so that the whole calculation can be carried out in class intervals without changing into the units of measurements. σ_1 is the standard deviation in terms of class intervals, hence in the example only half as large as σ, the class interval being 2. In using the short method, as in the example, it is, of course, necessary to correct both the sum of the fourth powers and σ_1, to refer them to the true mean.

The distribution is slightly platykurtic. By referring to Example 46 (page 142), it will be seen that the modal frequency is slightly lower and the adjacent frequencies slightly higher than those of a normal distribution. This makes the distribution less peaked and is the visible flattening that is measured by the coefficient of kurtosis.

CHAPTER VIII

PROBABILITY AND RELIABILITY

STANDARD ERRORS

When any measure of central tendency, of dispersion, or of variability has been calculated from a given sample, it is absolutely accurate for that sample within limits related only to the accuracy of the original observations. Viewed as devices for the description of actual samples, these values thus involve no inaccuracies or approximations beyond those readily visible and ascertainable in the technique of observation and in arithmetical calculation: they are themselves measures of real properties of real and available things and are not estimates from this point of view.

When, however, it is desired to infer the properties of a whole population from a given sample, the values calculated from the sample clearly are not accurate measures applicable to the whole population. It is impossible, within reason, for a sample that is only a fraction of the whole population to have exactly the same mean, standard deviation, coefficient of variation, or other parameters, as the population; for all these values depend on each particular individual observation in the sample, and these individual observations are selected by chance from the population. There is no conceivable way of selecting observations that will be distributed exactly like the whole population. If, however, the sample is properly taken, which means only that it is taken entirely at random, and if it is sufficiently large, its distribution will approximate that of the population, and the calculated parameters accurate for the sample will be more or less good approximations of those of the population. From this point of view the parameters calculated from the sample are estimates of the unknown and incalculable (but really existing) parameters of the population.

It is of prime importance in making any inferences about the population to know the probability that these estimates do repre-

sent the population parameters or (what more nearly expresses the intention of what is done) to know the probability that the population parameters lie within a certain distance of these sample parameters used as estimates. This is possible because of the fact that these sample parameters are themselves usually normally distributed. If, for instance, a number of separate samples of the same size were taken from a single population and the mean were calculated for each sample, each of the means thus obtained would be different, although usually only by a small fraction of their absolute values. This series of different but approximated means could then be placed in the form of a frequency distribution, and their distribution would be approximately normal,[1] with their mean in practice very close to the mean of the population and in theory exactly equal to it. By increasing the number of samples used, the population parameter could thus be estimated as closely as desired, and the value obtained would in effect become an actual measurement of the population parameter as accurate as the more direct measurement on a single sample (see Fig. 18).

It is not, in fact, practical to carry out this procedure. In dealing with any given problem, the usual procedure is to combine all the observations on a single sample and to obtain from it a single estimate of each desired parameter of the population. Advantage is then taken of the fact that this estimate is a chance value drawn from a normal distribution of values that would be obtained from a large number of samples of the same size as the one actually available. From the properties of the normal curve and corresponding normal distributions, discussed in the preceding chapter, it is then necessary only to know the standard deviation of this hypothetical distribution of sample parameters in order to know the probability that the particular estimate in hand lies within a given distance of the mean of that distribution, which signifies that it is within a given distance of the real value of the parameter for the whole population. Thus if the standard deviation of this sample parameter distribution is .31 and the calculated value of the parameter for the single sample is 5.20, it is a safe conclusion in almost all circumstances that 5.20 is

[1] This would be true of any parameter of a distribution itself approximately normal. For some parameters, notably the mean, it is true whether the population distribution is normal or not.

distant not over three times the standard deviation from the true value of the population parameter. In other words, it is demonstrated for all practical purposes that this parameter of the population is within a distance of .93 of the parameter of the sample, or within the range (5.20 − .93) to (5.20 + .93), *i.e.*, the range 4.27–6.13. It will probably be within a still smaller range,

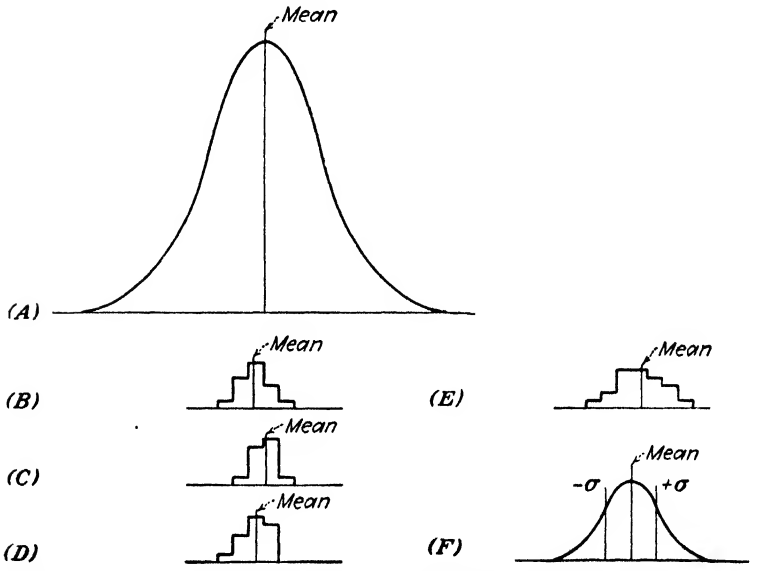


FIG. 18.—Diagrams to show the meaning of the standard error of the mean (σ_M). *A* is the normal frequency curve of a hypothetical population, the mean being indicated. *B*, *C*, *D*, and *E* represent histograms of distributions of small samples drawn from this population, the mean being shown in each case. This varies from sample to sample. *F* shows a hypothetical distribution of the mean values of a large series of such samples. The standard error of the mean of *A* is the standard deviation of this distribution (*F*).

the probabilities being given by the areas of sections of the normal curve as given in Table I (page 137).

In order to distinguish this standard deviation of a (hypothetical) distribution of sample parameters from the standard deviation of the distribution of values of the variate in question, it is called a standard error. Every calculated parameter when used as an estimate of a corresponding population parameter has its own standard error. The symbols for these are in each case σ

(because they are standard deviations although of a different sort from that hitherto considered) with a symbol of the parameter as a subscript. Thus σ_M is the standard error of the mean, σ_σ is the standard error of the standard deviation, and σ_V is the standard error of the coefficient of variation. The standard error is written with a plus and minus mark after the value of the sample parameter; thus $5.20 \pm .31$ means that the value of the parameter calculated from the sample is 5.20 and that as an estimate of the population parameter this has a standard error of .31.

It is most unfortunate that the word "error" is used in this connection, for its various vernacular uses lead to confusion. It seems to imply that there is something wrong with the figure to which a standard error is appended, that an attempt is being made to allow for some mistake made. Even some statisticians carelessly fall into this entirely wrong assumption. The fact is that the figure itself is perfectly accurate if correctly calculated and recorded and that if there were any errors in calculation, bias in observation, or other sources of inaccuracy in obtaining this figure, the standard error would give no information whatever about them and make no allowance for them. A correctly calculated and recorded figure is, however, accurate only for the data used in obtaining it, that is, for the sample. The standard error is not a correction of this result but is an estimate of the probability that the result applies also to the population from which the sample was drawn. The standard error is used not to give any information about the sample but to give information about the population.

A figure like $5.2 \pm .3$ is sometimes erroneously used to mean that the observed or calculated figure 5.2 is known to be inaccurate but that its true value is believed to be between 4.9 and 5.5. The correct way to record this is to give this range and make the record 4.9–5.5, or to say that the value is 5.2 to the nearest .3 or is 5.2 with class interval .6. The correct meaning of $5.2 \pm .3$ is that the observed or calculated value is accurately 5.2 and that its standard error is .3. Hence the corresponding value for the whole population is about 5.2; more exactly the odds are about 38 to 62 that it was in the range $5.2 - \frac{1}{2}(.3)$ to $5.2 + \frac{1}{2}(.3)$ or 5.05–5.35; about 68 to 32 that it was in the range 4.9–5.5; about 95 to 5 that it was in the range 4.6–5.8; etc., according

to the table of relative areas within the normal curve (Table I), up to the point where it is practically certain that this value was in the range 4.3–6.1.

When a sample is very small, and under some other special circumstances, it is not always safe to assume, as this procedure does, that sample parameters would fall into a normal distribution; these may be less good as estimates of population parameters than would appear from the usual use of standard errors. This is discussed in Chap. XI; the present chapter applies to the more common use and calculation of standard errors, valid in most cases.

Many authors use instead of the standard error a figure that they call the probable error, symbol *PE*. This is obtained by calculating the standard error, as explained below, and then multiplying by .6745. The probability that a population parameter will be within one standard error (plus or minus) of the sample parameter is 68 in 100, or .68. The probability that it will be within one probable error is 50 in 100, or ½. The simplicity of this fraction is the principal excuse for the common use of the probable error,[1] but it is not a valid reason. There is no special virtue or usefulness in this particular fraction. It enters into no criteria of significance or other special uses of the probable or standard error. Other and more important multiples of the probable error do not give simple fractions, for instance, for $\pm 2PE$ the probability is .82, and for $\pm 3PE$ it is .96. The probable error thus has no real advantage over the standard error, and it has decided disadvantages. It is theoretically less advantageous because it is not, like the standard error, a direct and simple parameter of a normal curve; and it is practically disadvantageous because to calculate it the standard error must first be obtained in any case and then subjected to another mathematical operation, involving some labor and chances for mistakes without gaining anything. In short the current tendency to use the probable error is supported only by the fact that it is current, and the sooner it is dropped the better.

Because of this difference of usage, an author should state whether he uses standard or probable errors. If he does not, it is necessary either to ignore his figures, since the two are too

[1] Perhaps also the apparently simple meaning of the name, but this is merely misleading, even more so than the name "standard error."

different to be in any degree interchangeable, or to calculate one from his data (if his publication provides any way of doing this) and see which he used.

CALCULATION OF STANDARD ERRORS

The larger a sample is, the more nearly its parameters are likely to approximate those of the population. The standard error therefore varies inversely with some quantity involving total frequency N, and the formula for calculating any standard error is therefore always a fraction with N appearing in some way in the denominator. If the dispersion is small, a sample of a given size will give a better estimate for the population than if the dispersion is large, since the chance of the sample observations being far from an average value is obviously less with small dispersion. Therefore the standard error varies directly with some quantity involving dispersion, and some measure of dispersion will always appear in the numerator of its formula.[1]

The following are the formulas for standard errors of the more important frequency-distribution measures so far discussed. When other measures are introduced in subsequent chapters, their standard error formulas will be given with them. The derivations of these formulas are somewhat complicated mathematical operations based on the theory of probabilities and will not be explained here. Most of them are given by Yule.

Standard error of the arithmetic mean:

$$\sigma_M = \frac{\sigma}{\sqrt{N}}$$

Standard error of the median:

$$\sigma_{\text{median}} = 1.2533\frac{\sigma}{\sqrt{N}} = 1.2533\sigma_M$$

[1] It should be noted in passing that some authors imply or even explicitly state that the size of the standard (or probable) error is also determined by the refinement of observation or accuracy of preceding calculations. Of course incorrect data produce incorrect standard errors; but the standard errors do not detect, allow for, or correct such mistakes. If the refinement of observation was adequate for the problem in hand, further refinement will have no appreciable influence on the standard error. Those who make opposite statements have evidently themselves been confused by the difference between the meaning of the word "error" in its vernacular and technical uses and have forgotten what a standard error really is.

Standard error of first or third quartiles:

$$\sigma_{Q_1} = \sigma_{Q_3} = 1.3626\frac{\sigma}{\sqrt{N}} = 1.3626\sigma_M$$

Standard error of standard deviation:

$$\sigma_\sigma = \frac{\sigma}{\sqrt{2N}}$$

Standard error of mean deviation:

$$\sigma_{\text{M.D.}} = .6028\frac{\sigma}{\sqrt{2N}} = .6028\sigma_\sigma$$

Standard error of the coefficient of variation:

$$\sigma_V = \frac{100\sigma}{M\sqrt{2N}} = \frac{V}{\sqrt{2N}} \quad *$$

Except that for σ_M, these formulas assume that the distributions are approximately normal.

The calculation of these for the data of Example 36, used throughout Chap. VI to illustrate sample parameters, is given in Example 50.

Generally adequate[1] records of these parameters would thus be:

$M = 60.4 \pm .3$	Median $= 60.4 \pm .4$	M.D. $= 2.46 \pm .14$
$\sigma = 3.04 \pm .23$	$Q_1 = 58.2 \pm .4$	
$V = 5.0 \pm .4$	$Q_3 = 62.6 \pm .4$	

In order to exemplify and clarify the relationships between samples, populations, their respective parameters, and standard

* This is an approximate formula only, the exact formula being

$$\sigma_V = \frac{V}{\sqrt{2N}}\left[\sqrt{1 + 2\left(\frac{V}{100}\right)^2}\right]$$

In practice the expression in brackets can almost always be ignored. For $V = 20$, as large a value as it commonly takes with zoological data, the expression in brackets is equal to 1.04, which would not affect the first decimal place of σ_V and would have no real significance. It need be used only with extraordinarily large values of V.

[1] The record is generally adequate when given to a smaller number of places than are used in calculation or for some special reasons. Criteria for this are given later in the present chapter.

EXAMPLE 50.—CALCULATION OF PRINCIPAL STANDARD ERRORS FOR DISTRIBUTION GIVEN IN EXAMPLE 36

$N = 86$
$M = 60.43$ Median $= 60.41$
$\sigma = 3.04$ $Q_1 = 58.19$ } from calculations in Chap. VI
M.D. $= 2.46$ $Q_3 = 62.58$
$V = 5.03$

$$\sqrt{N} = \sqrt{86} = 9.27$$
$$\sqrt{2N} = \sqrt{172} = 13.11$$
$$\sigma_M = \sigma/\sqrt{N} = 3.04/9.27 = \pm.33$$
$$\sigma_{\text{median}} = 1.25\sigma_M = 1.25 \times .33 = \pm.41$$
$$\sigma_{Q_1} = \sigma_{Q_3} = 1.36\sigma_M = 1.36 \times .33 = \pm.45$$
$$\sigma_\sigma = \sigma/\sqrt{2N} = 3.04/13.11 = \pm.23$$
$$\sigma_{\text{M.D.}} = .60\sigma_\sigma = .60 \times .23 = \pm.14$$
$$\sigma_V = V/\sqrt{2N} = 5.03/13.11 = \pm.38^*$$

errors, the experiment summed up in Example 51 was made. The length of the second lower molar was measured on the available specimens, 61 in all, of a species of fossil mammal and the standard data calculated. These figures (aside from N and R) may be considered as estimates of the parameters of a population consisting of all the members of this species that lived in the Bighorn Basin in the Lower Eocene. For experimental purposes it may be taken that there were only 61 members in this population, that we have them all, and that the parameters as calculated are those of the population—if this were really true the calculation of standard errors would, of course, be pointless. Now smaller samples can be drawn from the whole series and considered as samples drawn from a population.[1] The parameters of these smaller samples could then be directly compared with those of a known population.

To test the bearing of the experiment on zoological practice, a variate with rather high dispersion was deliberately chosen,

* By the more exact formula $\sigma_V = 5.03/13.11\sqrt{1 + 2(5.03/100)^2} = .38\sqrt{1.00506} = .38 \times 1.003 = \pm.38$. The correction makes no difference in the second decimal place and is insignificant.

[1] As a point of technique, the 61 observations of the large sample (the population of the experiment) were written on 61 exactly similar slips of paper, shuffled and spread out so that the numbers could not be seen, and a sample drawn. Then the slips were returned and reshuffled and a new sample drawn.

EXAMPLE 51.—COMPARISON OF PARAMETERS OF A POPULATION AND OF SAMPLES FROM IT

Length of M_2 in *Phenacodus primaevus* from the Gray Bull Formation, Bighorn Basin, Wyoming. All available specimens (61) and five smaller samples drawn from this (original data)

Sample	N	R	M	σ	V
A	61	10.6–13.8	12.12 ± .18	.96 ± .09	7.9 ± 0.7
B	10	10.7–13.5	12.26 ± .30	.95 ± .21	7.7 ± 1.7
C	10	10.8–13.6	12.07 ± .28	.88 ± .20	7.3 ± 1.6
D	10	10.7–12.8	11.93 ± .22	.71 ± .16	5.9 ± 1.3
E	5	10.9–13.8	12.18 ± .49	1.10 ± .35	9.1 ± 2.9
F	5	10.9–13.2	11.98 ± .40	.91 ± .29	7.6 ± 2.4

Weighted means for the five small samples

M	σ	V
12.09	.94	7.3

Differences of values for the small samples from that for the large sample divided by standard errors for small samples

Sample	M		σ		V	
	d	d/σ_M	d	d/σ_σ	d	d/σ_V
B	+.14	+.5	−.01	−.05	− .2	− .3
C	−.05	−.2	−.08	− .4	− .6	− .4
D	−.19	−.9	−.25	−1.6	−2.0	−1.5
E	+.06	+.1	+.14	+ .4	+1.2	+ .4
F	−.14	−.4	−.05	− .2	− .3	− .1

and the subsamples were made smaller than is considered valid by most statisticians. Despite these unfavorable circumstances, the experiment entirely confirms what has been said about standard errors, the only unexpected result being that in this experiment by chance the results obtained from the small samples are somewhat better estimates than would have been expected on the theory of standard errors. The parameters of the small samples are in every case closer than two times their standard errors to the "population parameters" (those of the large sample).

In only two cases (σ and V of sample D) are they over one standard error from the "population parameters."

It is noticeable that the standard errors are considerably larger for the smaller samples. For the samples with $N = 10$ they are around twice as large as for the large sample and for those with $N = 5$ about twice as large as for those with $N = 10$. The accuracy of the samples *as estimates* is thus seen to decrease rapidly with decrease in size of the sample, or in other words the small samples give only much broader ranges within which the population parameters can be inferred to lie and hence do not give as detailed or refined information about the population. Nevertheless they do give some information, and what they give is correct.

It may be emphasized again that the parameters of the small samples are inaccurate not in themselves but only as estimates of those of the large sample. Similarly the parameters of the large sample are accurate for that sample but are only estimates of the parameters of the real population. The samples with $N = 10$ give better information about the large sample than do those with $N = 5$. Thus sample D ($N = 10$) shows the mean of the large sample surely to be in the range 11.27–12.59 while sample E ($N = 5$) confines it surely only to the much larger range 10.71–13.65. Both implications are perfectly correct: the mean of the large sample, 12.12, is in both these ranges. The small samples also imply that the mean of the real population is within these ranges. The large sample surely fixes this parameter in the considerably smaller range 11.64–12.64. It is only in this sense that the large sample gives more or better information than do the small samples. It is a closer estimate and to that extent more reliable.

CONSISTENCY, EFFICIENCY, AND SUFFICIENCY

It has been shown that, in estimating a population parameter from a sample, the closeness of the estimate increases with an increase in the size of the sample; or in other words that the larger the sample the smaller the range within which the population parameter is shown to lie. The exact course of this relationship is different for different parameters. For the mean, for instance, the closeness of estimate is inversely proportionate to the square root of the number of observations in the sample so

that in samples of 5, 10, and 61 specimens, as in the experiment just given, the estimates are closer approximately in the ratio $\sqrt{5}:\sqrt{10}:\sqrt{61}$, not far from 1:2:4, which was seen to be roughly the inverse relation of the sizes of the actual calculated standard errors, which measure the closeness of estimate.

All estimated parameters tend thus to smaller standard errors as the sample size increases, with the result that they approach fixed values, accurate to any number of decimal places. If this exact value toward which they tend is also the exact value of the real parameter of the population, the measures based on samples are said to be consistent. The arithmetic mean is a consistent measure, and so are most of the others here recommended. Inconsistent measures are those that also tend to a fixed value with increased sample size but for which the fixed value is not exactly that of the population parameter. Generally such inconsistent measures are "outside the pale of decent usage";[1] but in some cases the value toward which they are tending is so close to the population parameter that the difference has no practical bearing on their use, and in such cases inconsistent measures are properly employed if they are considerably easier to calculate than the corresponding consistent measures. The standard deviation by the formula $\sqrt{\Sigma(d^2)/N}$ is an inconsistent measure that nevertheless is near enough to consistency for almost any practical purpose. The same is consequently true of V and all the standard errors into the calculation of which a standard deviation calculated in this way enters.

Efficiency as applied to numerical observations and operations means obtaining the maximum closeness of estimate from any given sample. For instance, if a sample of 100 observations gives about the same variance, and hence approximately the same standard deviation and standard errors, as a sample of 50 observations drawn from the same population, then the treatment of the large sample has been only 50 per cent efficient. Efficiency is influenced by the refinement and accuracy of measurement and by the magnitude and sort of secondary grouping. General rules for maintaining efficiency in these respects were given in Chaps. II and III. The best criteria are that the class midpoints, whether of original observation or of secondary grouping, should be about $\sigma/4$ distant from each other, that the records should be

[1] R. A. Fisher.

accurate to that point, and that the observations should be as evenly distributed as possible about the midpoint in each class.

As a matter of fact the influence on efficiency of unduly large grouping or (what amounts to the same thing) of unduly coarse measurement is often surprisingly slight, as Example 52 shows.

EXAMPLE 52.—INFLUENCE OF VARIOUS DEGREES OF GROUPING ON THE STANDARD DEVIATION AND ON THE STANDARD ERROR OF THE MEAN IN THE SAMPLES OF EXAMPLE 51 (PAGE 157)

Sample	N	i	No. of steps	σ	σ_M
A	61	.1	33	.96	± .18
A	61	.5	8	.98	± .18
A	61	1.0	4	.97	± .18
B	10	.1	29	.95	± .30
B	10	1.0	4	1.02	± .32
E	5	.1	30	1.10	± .49
E	5	1.0	4	1.17	± .50

For the large sample A, even the grouping to 1.0 mm. (or original measurement only to the nearest millimeter), giving the theoretically inexcusably low number of 4 steps in the distribution, hardly makes an appreciable difference in the efficiency of estimate. For the smaller samples the effect of this coarse grouping is appreciable but is not great. This small degree of difference cannot, however, be relied on and does not excuse using coarse grouping if fine is obtainable and theoretically better. It does tend to show that if fine grouping is not obtainable, the data are still useful even though less efficient. In any case, grouping finer than about $\sigma/4$ does not noticeably increase the efficiency and is useless. The estimates of the above samples with interval .25 are in every case as good as with interval .1, and those with interval .3 are not appreciably different.

The principal difference in closeness of estimates in such samples is caused by the total frequency. In practice, then, failure to use all of the available sample is the most important source of inefficiency of estimation—"inefficiency" is here used in a slightly different sense, but the effect is the same. The common zoological and to still greater degree paleontological custom of selecting from a sample a single specimen or a few specimens as types and describing only these thus results in very low numerical efficiency, not necessarily in relation to what could have been done with the specimens actually measured but in relation to what could have been done with the sample avail-

able. Unless the available sample is extraordinarily high, running into the hundreds or thousands, there is no excuse for not using it all.

Even if the sampling and grouping are adequate, there are some measures that are, inherently, mathematically less efficient than others. The efficiency of different formulas designed to measure the same thing can be calculated and compared, but the methods of doing this are rather laborious. In general the measures and formulas recommended in this book are all adequately efficient, efficiency 100 per cent or near enough to that for the purposes in mind.

There are a few measures that are not only consistent and efficient but also sufficient in the sense that such a measure gives in itself all of the information about a sample that is pertinent to a given problem. The arithmetic mean of a Poisson series is sufficient in this sense, and, under certain circumstances, the mean of a normal distribution may be so. For the general presentation of numerical zoological data there is no one sufficient measure, but for many particular problems there are sufficient measures. When these exist, they are the ones here recommended.

SIGNIFICANT FIGURES

In the record of any numerical value, significant figures are in a strict sense those digits that are accurate, that is, such that the last is correct within a half step and implies a range within which the exact value does lie. In a broader sense one digit beyond these may be considered significant if it is nearer the exact value than would be the limit of the range implied by the preceding digit. Thus if an exact value is 2.5834 and a measurement or calculation of it gives the result 2.5843, the first three digits 2.58 are significant in the strictest sense. They imply the range 2.575–2.585, and the exact value is in this range. The fourth digit is not significant in this sense because the exact value is not in its implied range 2.5835–2.5845. It is, however, significant in the broader sense, since it is nearer to the exact value than is the limit of the range implied by the last accurate figure, that is 2.584 is nearer to 2.583 than is 2.585. Thus this fourth digit, although inaccurate, is a better approximation than is the number to three digits. The fifth digit of the observed

or calculated result, however, gives no additional information and is not significant in any sense. It should be stressed that 0 is a digit and may have significance. The number 2 (as part of a continuous series) is not the same as the number 2.0, nor is the latter the same as 2.00. If 2.00 is accurate, it has three, not one, significant figures.

In general it is incorrect and misleading to record and publish figures that are not significant. As a regular practice they should be significant in the strict sense. In figures to be used in calculation the last recorded digit may be significant only in the broad sense if it appears that the result of the anticipated operation will thus be improved.

As applied to the making and recording of original observations, the identification of significant figures is relatively simple and has been discussed in Chap. II. The significance of figures resulting from calculation is equally important and is neither simple nor obvious so that it requires further consideration.

In the first place, it is necessary to remember that there are two kinds of numbers, continuous and discontinuous, and to know the kind involved in a given operation. Ungrouped discontinuous numbers are generally exact, and so is the result of any operation using only numbers of this class—the available number of significant digits is often infinite. Thus one-third of 10 eggs is 3.333333 . . . eggs. One could write a million 3's, and every one would be significant in the strict sense; but one-third of 10 in. is not 3.333333 . . . in.

The sum of 2, a discontinuous number, and of 3.12, a continuous number, is 5.12. The limits of 3.12 are 3.115 and 3.125; the 2 is exact; so the limits of the correct sum are 5.115 and 5.125, correctly represented by 5.12. In such a sum there are as many significant figures as in the continuous number. This is not, however, true if both numbers are continuous. Then the limits of 2 are 1.5 and 2.5, those of 3.12 are 3.115 and 3.125, and the limits of the sum are 4.615 and 5.625. There is no accurate answer to this problem: for the sum does not have even one digit significant in the strict sense, for 5 excludes possible exact values between 5.500 and 5.625 and the exact value is not necessarily in the range implied by the single digit. The single digit 5 is, however, significant in the broad sense, for it approximates the possible range of true values. Two digits are not significant in

any sense, since 5.1 is not necessarily closer to the exact value than either limit of the range implied by 5. The sum of 2 and 3.12, both being continuous numbers, is thus approximately 5. It is not 5.12 or 5.1. The range of the sum of continuous 2.0 and 3.12 is 5.065–5.175. Of the result 5.12, the first digit is strictly significant, the second broadly so, and the third is not significant.

The general rule for addition to be derived from these relationships is that in a sum reached by adding continuous numbers there are usually as many broadly significant places as in the number with the fewest strictly significant places and generally one fewer strictly significant places in the result. If continuous numbers are to be added, they should therefore all have the same number of significant places and this should be one greater than the number of strictly significant places desired in the result. Subtraction is analogous to addition, and the same rule applies.

Similar considerations apply to the results of multiplication and division. The product of discontinuous 2 and continuous 3.12 is 6.24, with all digits broadly and the first two strictly significant, the range of the product being 6.230–6.250. Unlike addition and subtraction, the multiplication of a discontinuous and a continuous number gives a result in which the last digit is generally broadly but not strictly significant. The number of broadly significant places in the result of such an operation is generally equal to the number of strictly significant places in the continuous number. Division, on the other hand, commonly increases the number of broadly significant places by one and retains the number of strictly significant places. The quotient 3.12/2 (2 being discontinuous) is in the range 1.5575–1.5625, so that in the recorded value 1.56 all the digits are strictly significant and an added 0, making 1.560, is broadly significant. Thus while 3.3333 . . . and so on indefinitely is correct as one-third of 10 eggs, one-third of 10 in. is 3 as far as strict significance goes and 3.3 if broad significance is allowed. In any case it is not 3.33.

Multiplication of two continuous numbers introduces remarkably complex relationships. The product of continuous 2 and 3.12 may readily be verified as having the possible range 4.6725–7.8125. Thus the arithmetic result 6.24 has no strictly significant digits, and only the 6 is broadly significant. The tendency here is toward the same rule as for addition: the result of the multipli-

cation of two continuous numbers has as many broadly significant places as the fewer strictly significant places of either number and has one fewer strictly significant places. The area of a square 99.9 mm. on one side and 77.7 on the other is accurately (but not exactly) 77 sq. cm. and is approximately 77.6 sq. cm. The result 7762.23 sq. mm. has three wholly insignificant and therefore misleading digits. The range within which the unknown exact result lies is 7753.3525–7770.5575 sq. mm.

The quotient of the continuous numbers 3.12/2 is in the range 1.246–2.083333 . . . . The arithmetic result 1.56 therefore has no strictly significant figures, and a broadly significant figure is obtained only by rounding to 2. The quotient 3.12/2.00 is in the range 1.55361 . . . –1.56641 . . . . The arithmetic result 1.56 has two strictly and one broadly significant digits. The tendency is the same as for multiplication.

The plain fact that with many of the numbers used by zoologists 2 and 2 is not 4, but any number from 3 to 5 makes it difficult to judge the significance of figures even in the results of simple operations like finding an arithmetic mean. For such operations as getting a standard error, it is practically hopeless to try to work out limits as has been done in the very simple examples given above. Fortunately there is another factor at work in these operations, and there is also an empirical but adequate way of deciding how many places are probably significant.

The saving factor is that of probability. Although 2 + 2 (both being continuous) may be anywhere from 3 to 5, it is often more likely to be about 4, or to be in the range 3.5–4.5, than to be in the partial ranges 3.0–3.5 and 4.5–5.0. Moreover if a series of such additions are made, there are likely to be about as many exact values in the low partial range 3.0–3.5 as in the high partial range 4.5–5.0; therefore these will tend to cancel out, and the mean result will in all probability really be in the range 3.5–4.5, accurately recorded as 4. The probability that this will be true is directly proportional to the number of operations or observations and inversely to their dispersion.

Thus the result of operations like finding the mean or the standard deviation tends to have more significant figures than the original observations rather than less. Like the standard error this tendency is related to total frequency and to dispersion,

but the relationship is inverse to that of the standard error. The larger the standard error, the fewer the significant figures.

On this probability is based the best rule for significant figures in calculated parameters: A calculated parameter may be considered as probably significant to the number of places indicated by the first digit (not 0) of one-third of the standard error. Thus if a standard error is .36, one-third of it is .12, and the corresponding parameter value may be taken as significant to the first decimal place. If the standard error were .24, it would be significant to the second decimal place.

From the considerations clear in the preceding discussion it is best to carry out all calculations to at least one more decimal place than will be significant in the result. In recording the result, however, it is seldom worth while to give more than the significant figures. It is best to give a measure and its standard error to the same number of places. If the value obtained is to be used in further calculations, it may sometimes be advisable to give one more place than is significant. In a list of comparable values it may also be neater or more useful to give all to the greatest number of places significant for any. In either case, record of the corresponding standard errors will give adequate warning that nonsignificant digits are recorded.

Application of this criterion to actual cases bears out the statement that parameters usually have more significant places than do the observations on which they are based. Thus from the data of Example 51 (page 157) it is seen that for sample A the mean has two significant decimal places, although the original measurements had only one; and indeed the mean still has two significant decimal places even when the original measurements are rounded to integers.[1]

[1] This increase in refinement or in number of significant places has also been verified experimentally by Pearl.

CHAPTER IX

SAMPLING

Any particular set of observations usually has little or no interest unless it reveals characteristics of broader scope and wider application than those actually observed. Even observations that are truly unique, such as those of abnormalities not repeated, have no value unless they cast light on more normal and widespread processes like heredity and embryology. Zoology is, or should be, a study of populations;[1] but a whole population cannot be brought into the laboratory or examined in the field, so that the only practical approach is by the method of samples. Undue preoccupation with what is actually observed and failure to relate it to broader issues and conclusions are a constant danger of the method. It is a failure to see the forest for the trees, the population being the forest and the sample a few trees from it. The logical transition from the particular to the general is the most difficult part of research, and it is the point where the student is most likely to go astray. Observation, in itself, is not science and has no value except as a basis for interpretation and some degree of generalization. In previous chapters, something has been said of the relationship between samples and populations. With this as a background, it is now possible to consider in more detail and with better understanding the actual process of sampling, which is the obtaining of the individual observations by means of which the population is to be studied.

[1] The word "population" in this sense is not only literal, applying to a natural assemblage of animals, but also figurative, applying to all existing phenomena of which a few are observed. Thus when specific characters are determined from a sample, the population is literal, the assemblage of all animals of the species. When an individual's behavior is studied, the population is figurative and twofold: it is (1) the whole of the individual's behavior in this respect, before, during, and after actual observation; and (2) the behavior of all animals in which that behavior follows recognizably similar patterns.

CHARACTERISTICS OF GOOD SAMPLES

The ideal representative of a population is a sample that is homogeneous, adequate, and unbiased, three requirements which mean that:

1. All individual observations in the sample belong to a single, defined population.
2. These observations include all the essential variations within the population.
3. These variations occur in the sample with about the same relative frequencies as in the population.

The more nearly a sample meets these three requirements, the better it is and the more reliable are conclusions based on it.

The requirement of homogeneity is in practice subject to at least one qualification and one exception. Often it is impossible to say that a given population is itself homogeneous, and the very purpose of sampling is to learn from the sample whether the population is pure or mixed with respect to the problem in hand. When, for instance, all the generally similar fossils from one horizon and locality are laid out, the first concern is to see whether they are a homogeneous taxonomic group or really represent two or more groups, and hence populations. Samples have in this respect one very important limitation. They can frequently prove beyond reasonable doubt that the population is heterogeneous, but they can never strictly prove that it is homogeneous. That is why assurance of homogeneity depends primarily on the specifications of the population and not on observations on the sample.

If a distribution of a sample is definitely bimodal and can be shown to combine two distributions with significantly different parameters, then the sample and the population are surely heterogeneous—some tests for this are discussed in the next chapter. If many different distributions can be made for different variates of a single sample and if none indicate heterogeneity, it becomes probable that the sample and population are homogeneous. This, however, is frequently impossible because only one variate is pertinent to the problem, or for various other reasons; and in any case the negative cannot be really proved, it cannot be established that the sample is not heterogeneous.

In taxonomic work, two closely related subspecies may be mingled in a sample and their combined variation may be such that it is impossible to separate them or even to establish the fact that two subspecies are present. This exemplifies the importance of population specification. If the sample were from populations known beforehand to be homogeneous in every respect, the subspecies would not be mingled and it would be possible from their separate samples to prove that they are in fact distinct and to show just what the distinctions are. Knowledge of homogeneity is derived from collection records, and in taxonomic studies it involves homogeneity:

1. Of place, derivation from a single locality or small area.
2. Of environment, ecological unity.
3. Of time, contemporaneity of the animals studied.
4. Of age, animals all in about the same stage of their lives.
5. Of sex.

Sometimes all of these specifications cannot be met or clearly determined, and this may introduce uncertainty into the results. Other specifications may also be necessary for the problem in hand, such as homogeneity of physiological condition (all animals gravid females, all undiseased, all with some specified disease, and the like). Good sampling starts with decision as to what homogeneity is for the problem to be attacked, in other words, with specification of the population. Good collecting involves as complete record of all specifications as possible so that samples of a population meeting any necessary requirements as to homogeneity can be drawn from the collection.

Another common purpose and result of sampling is to prove that difference in specifications did not produce evident heterogeneity. For instance, it may be found that samples from different localities (but otherwise homogeneous) are not significantly different. In taxonomy, the conclusion is that the same species (or smaller unit) does occur at both localities. If the populations compared are well specified, this may amount in practice to proof of homogeneity, although without good specification such proof is impossible.

Exceptions to the requirement of homogeneity are provided by problems in which heterogeneity of population is itself an element. Obviously this is not an exception to the requirement

for specification of population, for such problems cannot be attacked at all without such specifications. For instance, in the study of body temperature if the problem is individual variability, the population should be specified as homogeneous: in race, sex, age, perhaps weight or other growth characteristics; in physiological conditions except body temperature; and in environment, particularly environmental temperature. But if the purpose is to measure homothermy, the samples should deliberately be selected so as not to be homogeneous in environmental temperature but on the contrary to be as heterogeneous as possible in this respect, the environmental temperature being, however, exactly recorded for each body temperature observation.

Even in such problems, in which heterogeneity is deliberately and necessarily sought, homogeneity is also of prime importance. Every effort should be made to keep all the population specifications except the one under consideration as homogeneous as possible. For instance, in determining homothermy, heterogeneity in any respect except environmental temperature may obscure or falsify the result. If some animals are old and some young, some male and some female, some hibernating and some not, some resting and some exercising, some well and some ill, or some fat and some thin, the results will not be valid. The influence of each of these or of any other factors could be determined by keeping it as the variable and making the sample as homogeneous as possible in every other respect. The element of individual variability must also be eliminated or measured and allowed for. In fact this problem would be best attacked by multiple samples, each derived from a population comprising the body temperatures of a single animal over a short period and in one physiological state with varying environmental temperatures. Body temperature is taken here only as an example. The same sort of considerations applies to any sampling to determine relationships between two variable factors (see Chaps. XII and XIII).

It should hardly be necessary to add that observations for which an essential specification is lacking should never be included in a sample. If, for instance, there is any reason to believe that fossil animals from two successive geological horizons differ or if it is desirable to test whether they differ or not, specimens of unknown or inexactly known horizon must be omitted.

They must be omitted even if they greatly outnumber those of known horizon and even if the student thinks that he can make a subjective separation of them. If recorded population specifications are inadequate, the problem simply cannot be studied in this form and either the sample itself must prove heterogeneity (which it often will not do even though heterogeneity be a fact), or the problem must be abandoned as insoluble from the data available.

This is why such a large amount of material in collections is entirely unfit to provide samples that will really solve urgent and legitimate problems to which the collections are related. It is why poor collecting or collection by inadequately informed amateurs or by careless, venal, or dishonest persons is more likely to make problems insoluble than to help to solve them. It is why so many species and subspecies, especially those of a century or more ago but also a painfully large number up to the present day, are completely and permanently unrecognizable on the basis of the type specimens;[1] and hence why the rigidity of the International Rules of Nomenclature may make a valid nomenclature impossible.

The requirement that a sample should include all the essential variations of the population does not mean that it must include observations exactly at the range limits and at every possible point in between. This would, it is true, be ideal; but it is impossible either to obtain such observations in every case or to know them to be such if they are at hand. For quantitative observations, variates, it is enough if the observations are well distributed within the range so that they permit reasonable inference as to the population range. The extent to which they do this is well shown by the standard deviation and its standard error, as shown in the preceding chapter. Even from single specimens or inadequately small samples, some idea of the population range can be obtained (see Chap. XI), although of course the information is less exact and less reliable than for larger samples.

As regards qualitative characters, attributes, the basic sampling requirement is that each important variation should occur

[1] Many taxonomists refuse to admit the possibility, but it is a demonstrable fact that two specimens may not differ significantly in any respect and yet may belong to two valid, different, and recognizable subspecies or species.

at least once in the sample. The adequacy of the sample in this respect can be determined from the theory of sampling limits, which is discussed in a separate section later in the present chapter (page 182).

In zoology the sample size is usually fixed in practice by what can be obtained. Instances in which a sample can be made of any desired size are rare, and it is a good general principle to use all the available observations. The question of adequacy is not so much that of deciding how large a sample is desirable, but how adequate the sample actually in hand is and whether it suffices to solve a given problem. There is very seldom any excuse or reason for not using all the observations that can be obtained with proper specification. Availability means observations possible on all the collections accessible to the student or the observations on living wild or laboratory animals that can be made in the time available and before conditions change so as to involve a change in specifications. Even one observation can throw important light on a problem, and hundreds of observations are seldom too many to handle or so many that equally good results can be obtained from fewer. In the rare cases where the available sample is really too large or unnecessarily large, subsampling can be carried out on the same principles as sampling in general, considering the unduly large sample as if it were a population and sampling it to reach conclusions regarding its characteristics.

Bias may enter into qualitative sampling, but it is a still greater danger in quantitative sampling. If an essential variation of the population cannot be inferred from the sample or if the sample is such that inferences based on it as to the frequencies of variation in the population are incorrect, then the sample is biased, and conclusions based on it are unreliable or wrong. Sampling bias may be very difficult or even impossible to detect. If bias is suspected, it is sometimes possible to obtain a new sample from exactly the same population, preferably by a different sampling technique. The bias may then appear from significant differences between the samples, which should give essentially the same results if the sampling were unbiased in both cases.

The most practical way to avoid bias is to give careful consideration to the specifications and sampling technique so that

they cannot conceivably bias the result. A few examples of specifications and technique will suggest the innumerable ways in which bias can arise and the importance of considering all possibilities before selecting a sample. It was pointed out above that specimens of unrecorded horizon should be rejected in studying the relationships between fossils of successive strata. It is extremely unlikely that there is any relationship between a record or its lack and any morphological characters of the specimens. It is therefore safe to assume that this selection does not bias the result. But suppose, for instance, that specification of essential homogeneity of individual age were omitted and that, as often happens, the animals of one stratum were mostly juvenile and those of the other mostly mature. Many morphological characters are influenced by age, and the samples are therefore biased with respect to these characters by faulty specification. Or suppose that growth of a mammal limb bone is being studied and that it is specified that the limb bones be complete. A common reason for incompleteness is loss of epiphyses. This occurs only in young animals and so is directly related to the problem of growth, and the specification strongly biases the sample for the study of that problem. Again some detailed experiments were made on heritability of acquired skill in animals, and the sampling was done by opening a cage and taking the first animal that came to hand. Now the coming of the animal is evidently likely to be affected by its past experience, intelligence, or activity; and these are elements in its acquisition of skill. The sampling technique therefore very probably introduced bias into the results. Similarly field collecting sometimes tends to get an unduly large proportion of the more bulky or more active animals or to be otherwise biased. Often there is little that can be done about this, but it must be kept in mind in interpreting the results.

The theory of probability on which inference from sample to population depends assumes that the sample is taken at random. Conscious selection is involved only in specifying the population. The sample should be selected only by chance, and any element not chance and random may introduce bias. The sampling of laboratory animals, as in the learning-inheritance experiment, would be unbiased if each animal were numbered and the numbers written on identical balls or cards, thoroughly

mixed, and the required sample taken at random from these. In collecting wild animals, the method used should be as little selective as possible so that the collection will be random, or as nearly so as possible, within each specified population. Once the specimens are collected, it is necessary to assume that they are a random sample, or to allow for this if it is known that the collecting was biased. Then the whole collection is used as far as it meets population specifications or, in rare cases, subsampled by chance methods, such as drawing numbered cards or balls.

Sampling is random and meets the requirements for being unbiased when within a population any one individual observation is as likely to be drawn in a sample as any other. It follows that a population from which a sample has been drawn and not returned cannot give another absolutely unbiased sample, because there is no chance that the second sample will include any of the observations already withdrawn but nevertheless pertinent to the population. If the population is very large and if the first sample is unbiased, a subsequent sample will not in general be significantly biased. If, however, the first sample is biased or if the population is small, subsequent sampling will often be significantly biased. For instance, it is impossible to obtain an unbiased sample of male deer in this country, because the game laws assure the withdrawal from the population of a biased sample by hunters at each open season. It is likewise impossible to split up a large laboratory sample into a series of smaller unbiased samples—only the first sample can be unbiased, and the subsequent samples are more and more biased until the last is not selected by chance at all.[1]

It is, however, possible to devise suitable sampling techniques by which a sample can be split into subsamples biased with respect to each other or to the sample, but not significantly biased as representatives of the population for some particular problem. If this were not true, good parallel experiments on laboratory animals would not be possible. A sample split simultaneously into subsamples of equal size gives such essentially unbiased representatives of the population if the original sample was unbiased, if the splitting of it was at random, and if each subsample is of adequate size for the problem. It is also

[1] But repeated unbiased sampling can be done, as in Example 51 (p. 157), if each sample is returned to the population after being recorded.

possible for a sample to be biased in some respects, yet to be essentially unbiased in others. The bias is important only if it is related to the conclusions based on the sample.

FAUNAL SAMPLING

Most field collecting is a sampling operation in which the population is specified as all the animals of some zoological group (such as all insects, all fishes, all birds, all mammals, all carnivores, all cats) living in a given area. Data are recorded with the collections whereby smaller samples with more limiting specifications can be drawn from the collection. The requirements of different sorts of animals and of different types of problems are so various that no general schedule of data for all vertebrates is useful, but the following include the most commonly useful specifications:

Date of collection and, if possible, time of day.
Weather: temperature, light, humidity, and precipitation.
Place (geographic region).
Height or depth (from sea level and from local surface of land or water).
Station (local habitat).
Field identification (for help in sorting, not an adequate population specification).
Individual age (as closely as possible).
Sex.
Field measurements and counts, color notes, and other observations of variates best taken on recently dead material.
Method of taking (including such details for trapping, for instance, as type of trap, number of trap, interval since trap last visited, number of days for trap in same location, exact character of trap set, bait used).
Physiological condition (oestrous condition, sleeping, feeding, hibernating, shedding, etc.).

For fossils only the exact locality and stratigraphic horizon are pertinent field sample data.[1]

With good field data of this sort a collection can become a source of samples with almost any pertinent specifications. The collection is, however, a sample in itself, properly a faunal sample. Almost all older collections and most of those still made are intended only as qualitative faunal samples. The

[1] For convenience and museum record, not as sampling data, field identification, anatomical elements preserved, name of collector, and date of collection should also be recorded.

purpose is to include at least one representative of each taxonomic unit pertinent to the collecting, with no concern for relative frequencies. The collector simply moves into an area and employs as many different collecting methods in as many different places and ways as hold any promise of getting different species or smaller groups of the animals that interest him. The complexity of such qualitative sampling for large zoological divisions is suggested by methods used to get a qualitative sample of the fishes of Panama (Meek and Hildebrand 1923): drag net, set net, tide net, dip net, dynamite, two kinds of poison, hook and line, and dredge—and even this extensive list does not include all the methods of taking faunal samples of fishes. Each method took some species collected by no other.

Collection of this sort is skimming, the intention being, so to speak, to skim off a few representatives of each kind of animal in an area. If the samples could be of the same size for each kind, so much the better. The collection is not really considered as a sample itself but only as a series of samples specified by taxonomic considerations, to be sorted and studied separately in the laboratory.

In recent years increasing attention has been paid to faunal sampling as such, or to quantitative faunal sampling in which the purpose is not only to learn what kinds of animals live in a given area but also to observe their relative frequencies in that area and in the various local habitats or stations within it. This is not necessarily effected by killing and bringing back a faunal sample with the same relative composition as the fauna; in fact when possible it is preferable to collect a qualitative sample and to make the quantitative observations on the living animals. Good quantitative sampling, however, often does involve making a quantitative collection.

Quantitative sampling, whether on living animals or by collecting, involves counting individuals of each species[1] under such conditions that (1) the observed counts for the various species have the same relative sizes as do numbers of those species living in the area, or (2) the observed count for each species has an approximately known ratio to the whole number of individuals of that species in the area. Fulfilling either of these conditions

[1] Or subspecies or any other taxonomic group. "Species" is used in the text as an example and for brevity.

always involves considerable difficulty, and except with very localized and special groups neither one can ever be achieved with complete accuracy.

Collections made in the ordinary sort of qualitative sampling often give approximate quantitative data. For instance, with fishes that are all easily taken on hooks of the same size with the same bait, the catch will be a good quantitative sample, or a drag net with large enough mouth and small enough mesh to catch both larger and smaller fishes will generally give a fair quantitative sample of the water it traverses. But the drag-net sample will not be comparable with one taken by dredging or by spearing, and the hook-and-line sample will not be comparable with any of these.

The principal methods especially devised for quantitative sampling are the quadrat, station, and traverse methods and various adaptations and combinations of these. In the quadrat method the area to be sampled is subdivided into squares or quadrats, and an effort is made to count or collect all the animals living on each of a selected number of these quadrats. If the quadrats are well chosen and the observation or collection is made rapidly and well, the total number of animals in the area can be closely estimated by multiplying the quadrat observations by the ratio of the total area to the area of the selected quadrats.

The desirable size and number of such sampling quadrats necessarily vary greatly, depending on the animals sought, nature of the country, and other factors. The size of the quadrat must depend largely on the normal range of individuals of the species sought and should be large enough so that several individuals will occur on each quadrat. Collecting or observation should be as rapid and yet thorough as possible in order to include most or all of the animals in the quadrat when the work is begun but not to give time for others to wander in.[1] The quadrats should themselves be an adequate random sample of the area, including its various local habitats in the right proportions and chosen by chance, if possible by laying out on a map and taking every fifth, tenth, etc., quadrat mechanically. Not

[1] Unless, of course, the quite separate problem of wandering is also being investigated, when the initial work should also be rapid but should be continued in the same quadrats.

less than 10 quadrats is advisable, and a higher number is better if possible.

The distance between the chosen quadrats should also be great enough, relative to the size of the quadrats, so that the operations do not seriously disturb the animal population and, if collecting is the method, so that the local fauna will not be depleted or seriously unbalanced—this applies to any type of collecting, for the collector as such is not a proper agent of animal control. If sampling quadrats are well selected, even the complete extermination of the animals in them does no harm, for the relative numbers of animals of various species are not changed and the empty quadrats quickly fill up by normal increase around them. Skimming sampling is much more likely to upset a fauna.

It is not necessary or even desirable that each quadrat include a representative assemblage of all the species of the area. It is much better both as a sampling operation and as a conservation measure that the sampling quadrats be small and widely scattered, each with only a random and not complete representation of the fauna. If the quadrats are a good sample of the area, the animals of all combined will be a good sample of the fauna even though those of any one quadrat are not.

Perfect sampling is an impossible ideal. With relatively sedentary and easily captured or observed animals, the quadrat method seems to approach the ideal most closely. In other circumstances it may be poor or even wholly impractical.

The principle of the traverse method is the same as that of the quadrat method, but the observations are made along lines instead of in squares. The lines should be parallel, numerous enough to sample the area properly, and far enough apart for a single individual rarely to be recorded on different lines or, in collecting, for one line not to draw animals from the same area as any other. It is preferable to run the lines simultaneously if possible, and in any event as rapidly as may be. When the area has settled down, another set of traverses may be run as a check at right angles to the first. The system of moving trap lines covering a band is a combination of the quadrat and traverse methods, usually difficult to evaluate, but sometimes giving better absolute results.

The great drawbacks of simple traversing are that the size of the area of observation relative to the whole area of the region

sampled is very difficult or impossible to evaluate with any accuracy and that different species will be drawn from areas of different sizes. More wandering species will appear relatively more abundant (as they are likely to by any method, but especially by this); and larger, less timid, or more distinctive animals will be observed at greater distances. This cannot wholly be avoided by restricting observations to the path in which small and obscure species can be identified, for many animals tend to move out away from the line. Distance from the line is therefore an essential datum, and the interpretation of results may become very complex and uncertain. In favorable circumstances, however, traversing gives reasonably good data on relative abundance. It seldom gives a very adequate idea of absolute abundance although it may set a minimum.

Adapted forms of both quadrat and traverse methods are used in the station method.[1] The principle of this is to establish a number of stations of observation or collection and to record for each species the number of different stations at which it was taken (not, as in the quadrat and traverse methods, the number taken at each station). The stations may be quadrats, traps or trap lines, unit linear distances along traverse lines, scattered points or circles of observation, or time units. In using time stations, the period of observation or collection for any one region or place is divided into equal intervals, and the number of separate intervals in which each species was seen is the record.

This apparently simple method is more complex numerically than may appear at first sight. The relation of the data to either relative or absolute abundance is not direct and may be indeterminate, and it seems likely[2] that in most cases the data of ordinary qualitative sampling are more satisfactory, even quantitatively, than are those of the relatively elaborate station method. Surely well-conducted quadrat or traverse studies are preferable. Unlike these, the station method requires a high number of units, preferably 100 or more; and it logically requires

[1] Originally proposed for floral studies by Raunkiaer and by Gleason, extended to mammals by Kenoyer and in various other forms by Grinnell and Storer, by Linsdale, and by Dice (for references and further details see Dice 1931).

[2] As one of its proponents, Dice, also concludes.

units so small that the occurrence of any one species in them is relatively uncommon. If this condition is met, then most really rare species will be missed altogether; and if it is not met, then all the species except those really rare will be given high and about equal frequencies. Communal animals which are very abundant but only at limited spots will appear rare by the station method, and rare but solitary and widely dispersed species may appear abundant. In fact the station method really measures not relative frequency, as its users generally claim or suppose, but relative dispersion, which is quite a different thing in faunas as it is in variate distributions. As a measure of faunal dispersion it may prove to have considerable value. It is also fairly well adapted to studies of faunal fluctuation from hour to hour, season to season, or year to year, where frequency is a function of dispersion or where dispersion is the important factor.

PALEONTOLOGICAL SAMPLING

The problem of sampling in paleontology is very different from that of sampling among recent animals and constitutes a special problem, although the treatment of the samples, once obtained, is often the same in both fields. The most marked sampling difference is that, in paleontology, field collecting cannot meet any of the usual specifications except those of place and geological age. To these may be added a few very special data such as type of matrix and manner of occurrence, to some extent permitting specifications analogous to those of habitat or method of taking for recent animals. But many specifications often well filled from recent field data, like those for sex and individual age and those requiring that values of variates shall be those of living (or recently killed) animals, cannot be met at all by paleontological collecting; and heterogeneity in these respects can be determined, if at all, only from operations with the sample itself. Paleontological samples are thus always somewhat heterogeneous in fact as they come to the laboratory, and it is seldom possible to make them completely homogeneous by any amount of study and selection. They usually demand broader treatment, but within these broader limits work on them may be and should be just as accurate and just as useful as on recent materials.

Another peculiarity of paleontological samples is that they are always biased in some respects. They may not be biased for a particular problem, and it is often possible to determine the bias roughly and to allow for it; but its existence demands recognition.

The agencies by which paleontological faunal samples are biased are: (1) biotic areas or facies, by which the animals were to some extent sorted out while still alive; (2) agencies of burial and fossilization, by which the animals whose remains are actually preserved were selected; (3) agencies of exposure, by which some, but in no case all, of the preserved animals are made available to the collector; and (4) the collector, who finds and collects some but never all of the available fossils. This multiple sieve through which the animals are, so to speak, sifted before they become a sample for investigation exerts a pronounced influence on the nature of the sample. Nothing can be done about this (except, to some extent, as regards factor 4); but bearing these various factors in mind permits some judgment as to the degree and nature of the imperfection of a paleontological sample and may lead to considerable modification of the inferences based on it.

Factor 1, biotic areas or facies, means that every paleontological sample is biased at the start. Seldom or never does any geological formation yield all the types of animals living at any given time, even in the general area in question. It includes only animals that lived in some one type or a few types of surroundings, perhaps marshy, or desert, or (in practice extremely rare) mountainous. Study of the fauna itself, its entombing sediments, analogous recent faunas, etc., permit some judgment on this point. Another sort of facial bias exemplified in several collections is due to the fact that some species frequented the area in question only at certain periods of the year or of their lives.

Factor 2 is analogous to facies or tends to emphasize or modify facial differences, in that some types of vertebrates are much less likely to be fossilized than others. The usual relative scarcity of bird, bat, or higher primate remains in the paleontological record doubtless results principally from this cause. Other influences also may have a strong and fairly obvious selective influence; for instance, small animals necessarily pre-

dominate in deposits in narrow fissures, and large animals are likely to predominate in coarse gravels.

As far as it is natural, factor 3, exposure, is generally, but not invariably, an unbiased sampling agency. In a well-exposed formation, fossils are likely to be exposed in about the proportions in which they actually occur in the rock. Some incidental factor may, however, cause bias. For instance, a stratum laid down at some particular time may have been formed under conditions that gave it a distinctly different fossil content from adjacent strata and that also gave it a physical nature promoting better or worse exposure, thus definitely biasing the collection from the formation as a whole. Weathering of the bones after exposure may also be differential, those of some particular zoological or size group breaking down more rapidly under weathering and thus being abnormally few in the collection.

Collecting bias, factor 4, depends on ability and on attitude and purpose. It is well known that two collectors may consistently differ in their results, one finding, perhaps, a much higher percentage of small animals than the other. Also a collector who is instructed or who naturally tends to collect only for exhibition or for any specified and limiting aim will inevitably make a biased collection of the fauna as a whole and often even of any particular species in the fauna (for instance, may tend to collect chiefly large variants).

Two other factors may modify a paleontological sample for practical purposes, both affecting chiefly the apparent relative number of individuals of various groups in the collection. The first is that some animals actually have a much greater number of hard parts suitable for fossilization than have others. Thus an armadillo, thanks to its armor, has hundreds of bones more than has a rodent; therefore, a much greater number of pieces in a collection may not represent a correspondingly large number of individual animals in the population. The other factor might be called bias of identifiability. Although two species might be actually equally abundant in a collection, one would be recorded as more abundant if it were more readily identified from poorly preserved material. Or the apparent size of a species might be considerably greater than the real mean size if, for instance, its larger or adult specimens were more easily distinguished from some related form than its smaller or juvenile specimens.

SAMPLING LIMITS

For variates, the probable relationships between observations on a sample and characters of a population are generally well shown by the various parameters and their standard errors. No such means have been given for estimating the frequencies and distributions of attributes, nonnumerical characters, in the population from those of a sample. This can be done, although the results are not so definite or the implied limits usually so narrow as for variates.

The distribution of an attribute usually has only two classes: it is present or it is absent. Its frequency distribution consists only of the relative or absolute number of individuals that have and that do not have the character in question, and such a frequency distribution is not subject to the sort of analysis and synthesis explained in previous chapters. Nevertheless, there are many urgent zoological questions that can be answered only by such data. As regards sampling, such questions are:

Supposing that a character occurs in a certain proportion of the individuals of a species, how large a sample would be necessary to be sure, for all practical purposes, to include at least one individual with this character?

Given a sample of a certain size that does not include a given character, how large a proportion of the population can have had the character?

Given a sample of a certain size, what is the smallest proportion of the population in which a character must have occurred for it to be surely present in the sample?

Given a sample of a certain size in all the observations of which a character is present, what is the smallest proportion of the population that must have had the character?

The answers to these and many similar questions are given by the theory of sampling limits. The answers vitiate many conclusions that zoologists have based on small samples and show how necessary it is to pay especial attention to this important subject. To anticipate the results of the following enquiry, few zoologists would hesitate to say, for instance, that a character was absent in a species if 10 specimens did not show it; but all this really proves (assuming that the character is biologically possible in the species) is that not many more than half the individuals of the species had the character.

The sampling limits here involved are determined by a binomial of probability and by the assumed normal distribution of random samples. If a character is present in, say, 30 per cent of a population and absent in 70 per cent, then the probability of its occurrence in a sample is .3, and that of its nonoccurrence is .7. This means that if a great many samples of 10 specimens each were taken, on an average the character would occur in 3 specimens of a sample and be absent in 7. But this would be only an average result, and in fact such small samples are strongly affected by chance. It might happen that in a given sample the character would not occur at all or that it would occur up to 8 times instead of only 3. For a great many samples of this size a frequency distribution could be made, X being the number of times the character occurs and f the number of samples in each class. This distribution would be approximately normal; and therefore, if the standard deviation is obtained, the probability that X will be above or below any given value can be found from the table of areas of the normal curve. The standard deviation for such a distribution is

$$\sqrt{N \times p \times q}$$

where N = the total frequency of each sample (*i.e.*, 10 in the preceding discussion).

p = the probability that the character will appear, expressed as a fraction (.3).

q = the probability that it will not occur (.7).

In this case therefore

$$\sqrt{N \times p \times q} = \sqrt{10 \times .3 \times .7} = \sqrt{2.1} = 1.45$$

The mean value will be near the most probable value, which is 3. What we wish to know is the smallest number and greatest number of practically possible occurrences, in other words, the theoretical range limits; and study of the normal curve showed that these are practically always included between $(M - 3\sigma)$ and $(M + 3\sigma)$. In this case $(M - 3\sigma)$ is less than 0, which is the minimum possible value, and so 0 is the lower limit. $(M + 3\sigma)$ is $3 + 4.35 = 7.35$. But the values are always integral in fact, and it is well to play safe. Therefore, this can most safely be called 8; and it is established that in a sample of 10 specimens it is certain, for all practical purposes, that not more than 8 will have the character. These are the sampling limits for $N = 10$ and

$p = .3$.[1] Of course it remains true that the sample is more likely to have 3 than any other number of occurrences of the character, but this probability is not great enough to be a safe basis for conclusion with so small a sample.

On this basis, similar sampling limits can be calculated for any percentage of occurrence in the population and for a sample of any size, the lower limit being $(Np - 3\sqrt{Npq})$ and the upper $(Np + 3\sqrt{Npq})$, N being the size of the sample, p the fraction of occurrence in the population (expressed as a decimal), and q the fraction of nonoccurrence $(= 1 - p)$. If the calculated lower limit is negative, it is recorded as zero; and if the calculated upper limit exceeds the size of the sample, it is recorded as N, since 0 and N are actual absolute limits imposed by the conditions.

TABLE III.—SAMPLING LIMITS

Size of sample	Percentage of occurrence in population																	
N	10%		20%		30%		40%		50%		60%		70%		80%		90%	
5	0	3	0	4	0	5	0	5	0	5	0	5	0	5	1	5	2	5
10	0	4	0	6	0	8	0	9	0	10	1	10	2	10	4	10	6	10
15	0	5	0	8	0	10	0	12	1	14	3	15	5	15	7	15	10	15
20	0	6	0	9	0	12	1	15	3	17	5	19	8	20	10	20	14	20
25	0	7	0	11	0	15	3	18	5	20	7	23	10	25	14	25	18	25
30	0	8	0	13	1	18	4	20	6	23	10	26	13	29	17	30	22	30
40	0	10	0	16	3	21	6	25	10	30	14	34	19	37	24	40	30	40
50	0	12	1	19	5	25	9	31	14	36	19	41	25	45	31	49	38	50
75	0	16	4	26	10	35	17	43	24	51	32	58	40	65	49	71	60	75
100	1	19	8	32	16	44	25	55	35	65	45	75	56	84	68	92	81	99

[1] The most doubtful theoretical point is whether a very asymmetrical binomial, say for $p = .9$, $q = .1$, would not be too far from the normal curve for this reasoning to apply; but Pearl (1930), from whom this discussion is paraphrased with some changes, has shown that the difference is not significant even with this asymmetry.

Representative values, rounded to integers somewhat irregularly but so as to give safely high probabilities for each figure, are given in Table III. In each cell of the table the figure to the left is the lower limit and that to the right the upper limit. Thus for a sample of 20 and proportion of population occurrence 40 per cent, the limits are seen to be 1 and 15; in other words, in these conditions the sample would surely have the character represented once and would not have it represented more than fifteen times. Any conditions not covered by the table can easily be inserted in the formulas given above.

With this table or the formulas it is easy to answer questions like those posed at the beginning of this section. The following statements can be verified by the table:

If a character occurs in 20 per cent of the population, a sample of 50 specimens is necessary to be certain to include one with this character (in the 20 per cent column the number 1 first appears as lower limit in the row $N = 50$).

If a character does not occur in a sample of 15 specimens, it did not occur in more than 50 per cent of the population (in the $N = 15$ row, the 50 per cent column is the first in which the lower limit is not 0).

A character must occur in at least 40 per cent of the population to be surely represented in a sample of 20 specimens (in the $N = 20$ row, the 40 per cent column is the first with the lower limit above 0).

If a character is present in all individuals in a sample of 25 specimens, it must have occurred in at least 70 per cent of the population (in the $N = 25$ row, the 70 per cent column is the lowest with the upper limit equal to N).

It may be reiterated that these are limits of probability so great as to amount to certainty, the only ones on which conclusions may safely be based. Possibility and probability near an even chance, far from certainty, are different. Thus, if a character occurs in 20 per cent of the population, it is possible for it to occur in any sample of 1 or more individuals, and it is more likely than not to appear in any sample of 5 or more individuals, but it is only certain (for all practical purposes) to appear in a sample of 50 or more individuals. It is practically impossible for it to occur in all the individuals of any sample larger than 3 and improbable for it to occur in the whole of a sample of any size.

CHAPTER X

COMPARISONS OF SAMPLES

Most of the discussion up to this point has been devoted to the study of individual samples and of their relationships to populations. Probably the most frequent zoological operation with numbers, however, is in the comparison of two samples. Even in the study of a single sample, the usual aim is to obtain and present data that will permit subsequent comparisons. Such comparisons can be made intelligently and with reasonable objective probability only after the characteristics and relationships of samples and of distributions based on them are thoroughly understood. Once these ideas and the numerical operations based on them have been grasped, their use in comparisons is relatively simple; but without such data and concepts as means, standard deviations, and standard errors, comparisons are largely subjective, highly unreliable, and often erroneous or meaningless.

In comparing two samples, the primary operation is of course to see whether they are or are not different and what the degree of difference is. In practice two samples always are different, for the chances of two samples, even though drawn from exactly the same population, being identical in character are practically nil. The degree of difference may be slight or great, and the real purpose of the comparison is to judge from this difference whether the samples were or were not drawn from the same population. If the probability of their being from the same population is very small, so small that it is generally safe to consider it zero, then the difference between the samples is significant in a technical sense. If there is a definite possibility that the two samples are from one population and that their difference arose only from the chances of random sampling, then the difference is not significant—it is unreliable, and conclusions cannot safely be based on it.

This is the logic involved in any comparison, however poor and subjective may be the method of carrying it out. When a specimen is compared with a type and decision is made as to whether it belongs to the same species, the two are really being treated as

samples. The idea is to see whether the specimen to be identified was drawn from the same population, the species, as the type. This simple comparison as it is so often carried out is a thoroughly unscientific procedure. It involves no definite criterion of significance, nor idea of what the range of variation really is, no conception of the relationship of the type to the variation of the species as a whole, and no method of relating the specimen being identified to this specific distribution beyond a vague and subjective opinion that is shown by more reliable methods to be as often wrong as right. It is possible to obtain definite criteria on all these points, as will be shown in the present chapter.

No method of comparison is absolutely infallible. All methods rest on a basis of probability and so are conceivably open to exception. It is, however, possible to estimate closely what the probability is and to accept as significant only conclusions in which the probability of error is known to be so small that it can almost always be neglected. If, moreover, the conclusions are accompanied by the estimate of probability, everyone can judge what the chances of error are; and the published finding can hardly be considered erroneous even in the worst outcome, since it carries its own objective evidence and possible corrective.

Establishment of the significance of a difference between two samples is not in itself a zoological conclusion but only a datum that makes such a conclusion possible. The zoological conclusion is not numerical and cannot be reached mathematically. If two samples differ significantly, then it is certain for practical purposes that they represent different populations and that the degree of difference of the populations can be estimated; but it is not demonstrated what the populations are or why they differ, what the zoological meaning of the difference is. If two samples of zoological specimens are significantly different in morphological character, they may belong to different species or other taxonomic groups, they may represent different sexes, they may be of different age groups, they may have been affected by different food, one may have been affected by disease—these and many other possibilities remain to be considered and to have a choice made between them on zoological nonnumerical grounds. Correct numerical treatment does not assure a correct conclusion, but it makes such a conclusion possible. Incorrect numerical treatment makes a correct conclusion impossible except by blind luck.

In the present chapter three very common types of comparison are treated: comparison of a single specimen with a larger sample, comparisons of the means of two samples, and comparison of the parts of a single sample drawn from more than one population. Some almost equally important but more difficult comparisons are treated in subsequent chapters: for instance, the comparison of two very small samples or single specimens (Chap. XI), comparisons of related values of dependent variates (Chaps. XII and XIII), and comparisons of differences between two series of values of a variate (Chap. XIV). The general principles established by the cases treated in this chapter apply also to these more complex operations.

COMPARISON OF A SINGLE SPECIMEN WITH A LARGER SAMPLE

It was shown in Chap. VII that if the distribution of a variate is approximately normal in a population the chances that an observation drawn at random from the population will lie at or beyond a given distance from the mean are given by the table of areas of the normal curve lying outside of limits at that distance from the mean, the distance being measured in units equal to the standard deviation.

Taking the calculated values for a sample as estimates of the population parameters, it is thus simple to calculate from the mean and standard deviation the chances that any one observation belongs to the same population as the sample. It is necessary only to find the absolute distance of the value of the observation from the mean of the sample, a deviation that may be symbolized by d, to express in terms of the standard deviation σ by dividing by the latter, giving d/σ,* and to read the probability from the normal-curve area table (Table I, page 137). The "percentage of area between limits" is the probability (as parts of 100) that a deviation equal to or greater than the one in question would not be obtained by random drawing from the same population as the sample, and the "percentage of area outside limits" is the probability that such a deviation might be so obtained.

For instance, if an observation is distant 2σ (*i.e.*, $d/\sigma = 2$) from the sample mean, the probability that its deviation would not be equaled or exceeded by one drawn at random from a population

* Biological statisticians sometimes call d/σ the "coefficient of abmodality."

with the same parameters as the sample is 95½ per cent, or .955, and the probability that it might be is 4½ per cent, or .045. This means that if 1,000 single observations were drawn from the population the chances are that about 955 of them would be nearer the mean than the observation in question and only 45 as far away or farther; or in still other words that it would be necessary on an average to draw about 22 observations (1,000 ÷ 45) at random from the population to get one that deviated as far from the mean as does the observation at hand.

As stated, these probabilities do not logically apply to the single observation being studied. The measured probabilities are not those of drawing this particular observation from a given population but the probabilities of drawing any observation with a deviation greater or less than that of the observation in question. The reasoning from this to the sort of conclusion actually sought involves one more step, and again we insist that numerical data and formulas should never be used by rote but only with comprehension of the logic behind each such step. If the chances of drawing an observation from the given (or, more strictly, estimated) population with deviation equal to or greater than that of the particular observation at hand are very small, it is safe to assume that this observation was not derived from that population. If, on the other hand, the probability is large enough to make the chance one to be reckoned with, it does not follow that the particular observation did come from the population. This is not, as often stated, a measure of the probability that it belongs to the population. All that does follow in such a case as regards this particular observation is that it might have come from the population, that it cannot be safely concluded on these grounds that this was not its origin.

Example 53 shows the use of this method of comparison.

The deviation d is obtained by subtracting the mean of the sample with which comparison is made from the single observation being compared, and d/σ by dividing this result by the corresponding standard deviation of the sample. P is used, as it will be in all such comparisons, to represent the probability that the single deviation could have been equaled or exceeded by one drawn at random from the population represented by the sample. In this case, P is read from the normal-curve area table. However derived, it is best expressed as a decimal fraction instead of as

EXAMPLE 53.—COMPARISON OF DIMENSIONS OF TEETH IN A SINGLE SPECIMEN OF THE FOSSIL PIG *Chleuastochoerus*, FROM THE PLIOCENE OF CHINA, WITH THE DISTRIBUTION OF THESE DIMENSIONS IN A LARGER SERIES OF SPECIMENS OF THE SAME GENUS FROM A DIFFERENT LOCALITY

(Data from Pearson 1928)

Variate	15–16 specimens from locality 49		Single specimen from locality 30_2			
	M	σ	X (measurement)	d (difference from M for locality 49)	d/σ	P
Length M_1	13.6	.7	14.3	.7	1.0	.32
Width M_1	9.6	.5	9.4	− .2	− .4	.69
Length M_2	16.4	.8	16.7	.3	.4	.69
Width M_2	11.9	.7	10.9	−1.0	−1.4	.16

a percentage, giving the probability as a fraction of unity (1.00 = 100 per cent probability). It is customary to record the sign (+ or −) with the values of d/σ, as this usefully shows the direction of deviation; but sign is ignored in using the normal-curve area table. In these simpler comparisons it is not necessary or customary to publish the value of P (given here only for illustration), for the value of d/σ adequately carries the desired information to anyone familiar with such measures. In some more complex comparisons the value of P is the usual and best record.

This operation gives a good estimate of probability, but it remains to be decided what value is small enough to be ignored or to show that the deviation is significant. Decision on this point is necessarily somewhat arbitrary and must depend in some measure on the particular problem in hand. In principle it is necessary to consider what P means and what is its largest value that will give a sufficiently reliable criterion of significance for the problem in hand. In some cases, $P < .05$ may indicate a significant difference; and in other cases it may be necessary for P to be less than .01 to give sufficiently reliable conclusions. As a general statement applicable to most zoological problems,

$P < .01$ is almost always significant, $P < .02$ is usually sigficant, $P < .05$ is sometimes significant, and $P > .05$ is not significant. The deviations in Example 53 are not significant, since the smallest P is .16.[1]

Table IV, which is merely a different form of such data as are given in Table I, shows values of d/σ corresponding to certain values of P and hence serves to set up criteria of significance in terms of d/σ.

TABLE IV.—RELATIONSHIP OF P AND d/σ

P	d/σ	d/σ	P
.05	1.96	2.0	.045
.04	2.05	2.5	.012
.03	2.17	3.0	.003
.02	2.33	3.5	.0004
.01	2.58		
.005	2.81		
.001	3.29		

Roughly transferring the same general criteria of significance as those given for P into values of d/σ, it is reasonable to conclude that a value $d/\sigma > 3$ is almost always significant, $d/\sigma > 2.5$ is usually significant, $d/\sigma > 2$ is sometimes significant, and $d/\sigma < 2$ is not significant.

Although the use of probable errors is strongly deprecated, they are in fact used in many past publications, and the intelligent reading of these necessitates some criterion of significance in terms of the probable error P.E.; therefore, a summary table (Table V) of corresponding values of P and of deviations in terms of the probable error is also given.

TABLE V.—RELATIONSHIP OF d/P.E. AND P

d/P.E.	P
2.0	.18
2.5	.09
3.0	.04
3.5	.018
4.0	.007
4.5	.002

[1] Hence the single specimen is not shown by these variates to be from a different population from the sample from locality 49. It is not, however, proved by these figures necessarily to be from the same population.

In terms of d/P.E. it may be said that values >4 are almost always, >3.5 usually, >3 sometimes, and <3 not significant.

COMPARISON OF THE MEANS OF TWO SAMPLES

If two samples are drawn from a single population, the chances are enormously great that they will not have exactly the same mean. Their means will in practice be found always to differ, and if the operation were repeated indefinitely with many pairs of samples from the same population these differences would generally tend themselves to fall into an approximately normal distribution. This distribution would have 0 as a mean, since the true value of the difference of means in the original population is 0 (there is only one mean, or all are the same, in the population); and positive and negative differences would be distributed on each side in frequencies proportional to areas of the normal curve.

Given two samples, the purpose of comparing their means is to determine whether the two populations from which they are drawn can have had equal means. This is the practical approach to the problem of whether the two populations are really identical, since in that case their means must be the same, or to the related but distinct problem as to whether the two populations, although distinct, are or are not distinguishable by mean values of this particular variate. The method is to estimate a standard error of the difference between the means of two samples on the hypothesis that the means of the corresponding populations were equal. If the observed deviation in terms of this standard error, a form of d/σ, is greater than could reasonably arise under this assumption, for instance if it is greater than 3, it is established for any practical purpose that the hypothesis is incorrect, or in other words that the samples are from different populations and that those populations do have different means.

Such a standard error of the difference between two means can be estimated by using the calculated parameters of the two samples as follows:

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

This is numerically equivalent to the following two formulas (again using values calculated from the two samples):

$$\sigma_d = \sqrt{\frac{\Sigma(d_1^2) + \Sigma(d_2^2)}{N_1 N_2}}$$

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

The second of these three formulas, being fewer operations removed from the raw data, generally gives the best result; but the result from the others will not be substantially different, and they are sometimes more readily useful.

The formulas given in almost all texts on statistics and hitherto almost universally used are quite different.

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

or

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

These give results numerically equal to those of the preceding formulas only when the two samples are of the same size. If the samples are very unequal, these current formulas may lead to a decision as to significance different from those given by our formulas, although in the usual cases of large samples of more or less similar size the results are not importantly unlike.

The disagreement does not mean that one method is right and the other wrong, but that they involve different assumptions and hypotheses. The one to be used is that really based on the hypothesis which is to be tested. In most cases the use of the current formulas is incorrect because it does not really follow the user's intentions. The formula

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

tests whether the samples could come from populations with the same mean and variance, in other words, whether they really could come from one population. This is almost always what the zoologist intends to test, and the general use of the other formula is therefore commonly wrong.

This other formula

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

assumes that the variances of the two populations are different (if the sample variances are different), that they are in fact well estimated separately by the two samples, and tests whether the two samples could have come from two different populations with such distinct variances but with equal means. The first formula is the one to use, for instance, to test whether two samples might represent a single species. The second is correctly used to see whether two separate species differ significantly in the mean for some variate.[1]

[1] Since the first formula differs markedly from that in most current statistical use it is advisable to explain its derivation, although this is somewhat lengthy and may be skipped by readers not especially interested in the point.

If $\Sigma(d_1^2)/N_1$ is used, according to the formula hitherto employed, for the variance of one sample and $\Sigma(d_2^2)/N_2$ for that of the other, the best estimate of the variance of a single population from which both samples are supposed by hypothesis to come would be

$$\sigma^2 = \frac{\Sigma(d_1^2) + \Sigma(d_2^2)}{N_1 + N_2}$$

which is a method of pooling the variance from the two samples.

It has been shown that the variance of a mean is that of its distribution divided by N (*i.e.*, that $\sigma_M = \sigma/\sqrt{N}$; a standard error or deviation is the square root of the corresponding variance). Taking the value of σ^2 given above as the variance of the hypothetical population, the variance of the first sample mean M_1 would therefore be σ^2/N_1 and that of M_2 would be σ^2/N_2. The variance of the difference between the means would be the sum of their separate variances, because they can vary either toward or away from each other and the difference between them changes from an average value by an amount that is the sum of their individual departures from their own average values. The variance of this difference is therefore

$$\frac{\sigma^2}{N_1} + \frac{\sigma^2}{N_2} = \sigma^2\left(\frac{1}{N_1} + \frac{1}{N_2}\right) = \sigma^2\left(\frac{N_2 + N_1}{N_1 N_2}\right)$$

or, substituting the value of σ^2 given above,

$$\left[\frac{\Sigma(d_1^2) + \Sigma(d_2^2)}{N_1 + N_2}\right]\left(\frac{N_2 + N_1}{N_1 N_2}\right) = \frac{\Sigma(d_1^2) + \Sigma(d_2^2)}{N_1 N_2}$$

The standard error of the difference is the square root of this variance, that is,

$$\sigma_d = \sqrt{\frac{\Sigma(d_1^2) + \Sigma(d_2^2)}{N_1 N_2}}$$

Since $\sigma_{M_1} = \sigma_1/\sqrt{N_1}$, $\sigma_{M_1}^2 = \sigma_1^2/N_1$. $\sigma_1^2 = \Sigma(d_1^2)/N_1$; so $\sigma_{M_1}^2 = \Sigma(d_1^2)/N_1^2$,

The correct use of these two different sets of formulas is shown in Examples 54 and 55.

EXAMPLE 54.—COMPARISONS OF MEANS OF TWO SAMPLES. TEST OF THE HYPOTHESIS THAT TWO SAMPLES WERE DRAWN FROM ONE POPULATION OR FROM POPULATIONS WITH EQUAL MEANS AND VARIANCES

Mandible length in samples of *Peromyscus maniculatus bairdii:* A, from Alexander, Iowa; and B, from Grafton, N. D. (data from Dice 1932)

Sample	N	M	σ_M	σ_M^2
A	13	15.721	.175	.030625
B	43	15.492	.062	.003844

$$d = M_1 - M_2 = 15.721 - 15.492 = .229$$

$$\sigma_d = \sqrt{\frac{N_1}{N_2}\sigma_{M_1}^2 + \frac{N_2}{N_1}\sigma_{M_2}^2} = \sqrt{\frac{13}{43} \times .030625 + \frac{43}{13} \times .003844} = .148$$

$d/\sigma_d = .229/.148 = 1.55$. . . . The difference is not significant. The two samples could have been drawn from a homogeneous population.

In comparing two lots of specimens, for instance, in deciding whether they belong to one species or not, a sample is obtained from each lot for each of a number of different variates and each variate is compared. In these circumstances, judgment as to the significance of deviations is inevitably and properly colored by the relationships of all the variates instead of being decided separately in each case. For instance, if the deviations for most of the variates fall well short of significance but one is just over an arbitrary limit of significance, it is well to reexamine the problem and see whether this one may not in fact also be a chance deviation and not truly significant. Or if, on the other hand, many variates are near but just below the chosen limit of significance and some beyond it, the conditions of the problem may warrant

---

and $\Sigma(d_1^2) = N_1^2\sigma_{M_1}^2$, and similarly $\Sigma(d_2^2) = N_2^2\sigma_{M_2}^2$. Substituting these in the formula for σ_d gives

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

This derivation is based on a more succinct and rather difficult explanation by Fisher of the derivation for an analogous formula for small samples, given in our next chapter.

EXAMPLE 55.—TEST OF THE HYPOTHESIS THAT THE TWO DIFFERENT POPULATIONS REPRESENTED BY TWO SAMPLES HAD EQUAL MEANS
Ratio of distal width to length in the ulna of: A, a fossil turkey *Parapavo californicus*; B, an allied living species *Meleagris gallopavo* (data from Howard and Frost; see Howard 1927)

Sample	N	M	σ_M	σ_M^2
A	133	10.8	.033	.001089
B	29	10.1	.077	.005929

$$d = M_1 - M_2 = 10.8 - 10.1 = .7$$

$$\sigma_d = \sqrt{\sigma_{M_1}^2 + \sigma_{M_2}^2} = \sqrt{.001089 + .005929} = .084$$

$d/\sigma_d = .7/.084 = 8.3$. The result is strongly significant, and the two populations surely did have different means.

Neither of these methods is reliable when the total frequency of the two samples is less than about 25, and in such cases the special method of comparison given in the following chapter should be used.

lowering the criterion of significance—the many "nearly significant" deviations tend to reenforce each other and to be significant together, although not clearly so singly.

Such problems of the combined significance of sets of deviations are usually best decided on nonnumerical considerations, and in any case the zoological meaning of the figures should be the primary consideration. They are susceptible of further numerical treatment useful in some cases. One method suggested is to calculate d/σ_d for each variate and to square these values and add them. Supposing the criterion of significance to be $d/\sigma_d = 3$ and the number of variates so treated to be n, the combination of the deviations is taken to be significant if the sum is greater than $n \times 3^2 = n \times 9$. The relationship of this operation to probability is, however, complex and obscure; and its relationship to the zoological considerations involved is uncertain and probably spurious in many cases. For instance, if there were a series of nearly significant deviations, say with $d/\sigma_d = 2.9$, 2.8, and 2.7, and one above significance, say 3.2, the sum of their squares would be 33.78 and well below the combined criterion of significance $n \times 9 = 36$. In fact in most zoological problems it is logical to suppose that the combination of such values makes them more significant than they are singly, and not less so as this operation would suggest. This and other doubtful points make the use of

this and analogous methods in zoology open to serious doubt. They may violate the principle, open to no exceptions, that however correct mathematically a numerical operation may be, it is invalid in zoology unless clearly and logically related to the zoological facts being studied.

HETEROGENEITY OF SAMPLES

When a collection is brought in from the field, it is always highly heterogeneous. Except as a broad faunal sample it is not a sample of a single population for any study purposes but is made up of a mixture of samples of many different populations. The first step in study is to separate the collection as nearly as possible into samples, each drawn from a single population. This is done in the first instance by setting up population specifications that can be met from the field data. It will, for instance, usually be necessary to specify populations by localities or areas or by geological horizons and to separate the collection, according to the field data, into samples representative of these. Even for studies not primarily taxonomic, it almost always becomes necessary eventually to set up taxonomic specifications and to separate the samples into subspecies, species, genera, etc., and this is, of course, the principal aim of strictly taxonomic studies. Sometimes these specifications can be met from the field data, which means that the taxonomic groups are so readily recognizable that part of what is essentially laboratory study could in fact be done in the field. More often this is not true, and taxonomic separation of samples must be done from the collection as a secondary operation and not simply by filling specifications from field data.

The problem of splitting a heterogeneous sample into relatively homogeneous parts on the basis of its own characteristics thus arises from almost all field collecting. The same problem sometimes arises in other ways than taxonomic. It is, for instance, frequently desirable to learn from a sample whether it is essentially homogeneous or recognizably heterogeneous as to age groups or as to sex. Observations on habits or other nonmorphological characters may also prove to be heterogeneous. With laboratory observations and controlled experiments, such heterogeneity can and should usually be avoided beforehand by meeting careful specifications; but this cannot always be done in the laboratory and often cannot in the field.

EXAMPLE 56.—SPLITTING A HETEROGENEOUS SAMPLE

Distributions of length of last lower premolar in small members of the extinct mammalian group Ptilodontidae, a heterogeneous sample from approximately the same horizon and locality, Lower Fort Union of the Crazy Mountain Field, Mont. (original data)

X	Interval, .1 mm.	Interval, .3 mm.
2.3	0	
2.4	0	1
2.5	1	
2.6	2	
2.7	0	3
2.8	1	
2.9	2	
3.0	0	2
3.1	0	
3.2	2	
3.3	1	6
3.4	3	
3.5	0	
3.6	1	2
3.7	1	
3.8	0	
3.9	0	0
4.0	0	
4.1	0	
4.2	0	0
4.3	0	
4.4	0	
4.5	0	0
4.6	0	
4.7	0	
4.8	0	1
4.9	1	
5.0	1	
5.1	1	3
5.2	1	
5.3	1	
5.4	2	3
5.5	0	
5.6	0	
5.7	0	0
5.8	0	

The observations as made, interval .1 mm., are highly irregular. But by grouping secondarily with interval .3 mm., a considerable degree of regularity is attained, and it becomes clear and certain that at least two populations are represented, one by the specimens 2.5 to 3.7 mm. in length and the other by those 4.9 to 5.4 mm. in length. These are obviously separate distributions with a long span of zero frequencies between them. These two distributions and the two samples that can now be separated on their evidence may still be heterogeneous, but at least each is more nearly homogeneous than was the single sample before this division.

The easiest and the only conclusive method of splitting heterogeneous samples is by frequency distributions that are plainly multiple and that do not overlap. If the frequencies are low and are irregular, this cannot be observed or determined. If, however, the distribution can be so grouped as to give a fairly regular sequence and if it then has two definite modes and a definite break, with frequencies zero somewhere between these modes, it is obvious that two distributions are really present and they can be separated by inspection. A special example of this sort is given in Chap. XVI (Blanchard). Example 56 shows another such case.

It may happen, however, that there are two modes but that the ranges overlap extensively. Even though the sample is really heterogeneous, the modes may be so close together and the ranges so nearly coextensive that it is impossible in a given distribution to detect the heterogeneity or to be sure that an apparent bimodality reflects two populations and is not the chance result of random sampling of one population. In taxonomic studies and others in which the sample observations are derived from a series of specimens with several variates pertinent to the problem under consideration, the correct procedure in such cases is to make distributions for all such pertinent variates. If any of these distributions show a clear and definite separation into two or more, it is possible to separate the specimens into groups on this basis and then to separate the other observations according to the specimen groups, even though these observations do not themselves have surely bimodal distributions.

In practice, taxonomically heterogeneous collections of related animals that are sufficiently homogeneous as to locality and, for fossils, horizon usually have at least one variate the distribution of which falls decisively into two or more parts and reveals the taxonomic heterogeneity. With the specimens of Example 56, for instance, the first of the two groups there visible did not

clearly split into two in any other distribution, and therefore the suggestion of bimodality here seen (one apparent mode in group 2.6–2.8 and the other, more definite, in 3.2–3.4) was very probably the result only of chance. The second group, however, did plainly split into two distributions for several variates (*e.g.*, number of cusps on M_1) and therefore is really heterogeneous even though there is no clear hint of this in the distribution given in the example.

This usual result in zoology whereby heterogeneous samples are most reliably split in practice is made more probable by the zoological law of ecological incompatibility which is, in its simplest form, that closely related and closely similar animals do not usually live together in one environment at one time. This has nothing to do directly with mathematics, but like many zoological principles it is essential in interpreting the results of numerical analysis. It may be restated from the present point of view: Really separable taxonomic groups of animals such that all their variates have extensively overlapping distributions are very rarely found together in nature. Such completely intergrading groups, usually contiguous geographic races or successive geologic subspecies, do occur, but not normally in full association with each other. Distribution of variates from samples of such populations are shown in Example 57.

If such a heterogeneous sample, composed of two or more overlapping populations, cannot be split on the collecting data, it usually cannot be split at all, and the zoological problem is usually insoluble.[1] There are ways in which the heterogeneity can be detected, or at least suggested, but even in such cases there is usually nothing practical that can be done about it. It may, for instance, be possible to show that two overlapping populations are represented and even, in the best possible circumstances, to calculate what percentage of specimens in the region of overlap belongs to each; but still the samples could not be separated, and it could not be determined what actual observations were from each population.

[1] It is possible to think of any given heterogeneous distribution as caused by the addition of two or more normal distributions and to calculate what normal distributions would give this result. As a mathematical problem, this is soluble, although with great labor. As a zoological procedure, the solution is usually impractical or meaningless—it has no real place in zoology.

EXAMPLE 57

A. Distributions of the discontinuous variate serration number for P_4 of two geographically and geologically separate samples of fossil mammals of the genus *Ptilodus* (original data)

X	f		
	Sample from Torrejon, N. M.	Sample from Fort Union, Mont.	Both together
12	5	0	5
13	1	8	9
14	0	19	19
15	0	2	2

B. Distributions of tail length in geographically separate samples of the deer-mouse *Peromyscus maniculatus* (data from Dice 1932)

X	f		
	Sample from Alexander, Iowa	Sample from Grafton, N. D.	Both together
52–53	1	0	1
54–55	3	1	4
56–57	11	2	13
58–59	18	2	20
60–61	21	3	24
62–63	20	8	28
64–65	9	25	34
66–67	2	11	13
68–69	1	10	11
70–71	0	5	5
72–73	0	7	7
74–75	0	2	2
76–77	0	2	2

In each of the examples given, the combined distribution for the two samples is not visibly bimodal. It would be impossible to learn from such a sample alone that it is heterogeneous. But in both cases the samples from populations specified as to locality are from decisively different populations, as comparison of the means has shown. The samples cannot be split except by specification of population from the field data, again emphasizing the fact that populations may be quite different, taxonomically or otherwise, and yet be quite inseparable on the basis of a mixed sample derived from both.

There are other hints of heterogeneity that never amount to proof but may be suggestive. A distribution with an unduly large V, much larger than for the same variate in related samples, is likely to be heterogeneous, and so is one that is strongly platykurtic. For instance, the large V for the tenth variate in Example 41 (page 124) suggests that the sample was heterogeneous, which was the fact; and so does the evident platykurtosis of the combined distribution of Example 57B (page 201), which is also heterogeneous in fact. Some homogeneous distributions do, however, have high V's, and some are platykurtic, so that this is never conclusive evidence of heterogeneity. A significant deflection, a high V, and strong platykurtosis are no more than hints that something may be wrong with the sample; and the best thing to do in practice is to reexamine the sample and to try to get better data, permitting assurance of essential homogeneity by specification of population.

CHAPTER XI

SMALL SAMPLES AND SINGLE SPECIMENS

Statistical use of measures of central tendency, dispersion, and probability assumes for the most part that the samples involved are large and that the characters of a population may take any values and can be judged only from the dispersion of the sample and the calculated standard errors of its various parameters.

Measures and procedures that are only approximations or that are slightly inconsistent as estimates of population characteristics are sufficiently reliable and accurate in dealing with large samples, but some of them become unreliable when the samples are small. Many statisticians therefore do not use such methods on small samples, and some go so far as to deny that it is possible to obtain any useful information about a population from a small sample. There is no general agreement as to how big a "small sample" is, but it is often said that it is useless to try to do anything with fewer than 30 observations.

This attitude is all very well if samples of almost any desired size can be obtained. Certainly there is no excuse for using a small sample if a large sample is available. Yet if it were true that any sample with $N < 30$ is small and that small samples are useless in the estimation of population characteristics, then the sciences of zoology and paleontology would become practically futile. A homogeneous sample of 30 specimens is a rarity in paleontology, and a large part of zoology is also based on smaller samples than this. Some zoologists likewise feel that such samples do not warrant synthesis or calculation of group characters and agree with the statisticians that statistical methods are not applicable to their materials. They apparently do not perceive that this stultifies their whole work. Such a conclusion really means that it is impossible from a usual zoological sample to learn anything useful about a population; and if this be true, zoologists are not studying nature, species, or general principles but are only amassing meaningless and incoherent single observations.

The zoologists in question naturally do not draw this logical conclusion; their agreement with the more extreme statisticians is merely an excuse for their loose and sometimes absurdly incorrect use of numerical data. In fact if such data are to lead to any broad reliable useful conclusions, they can do so only if based on methods essentially statistical in background if not in exact procedure. If such methods are available and applicable, they must be used if really good results are sought. If they are not, the fact must be faced that zoology is a very minor science indeed and that most of its results are meaningless.

Fortunately statisticians more conversant with the problems of small samples and zoologists more rational and realistic in their attitude toward their own problems have shown beyond question that small samples and even single observations do, or can, give useful information about populations. In the first place many procedures are equally valid for small samples and for large. As descriptions of the sample itself, measures of averages or of dispersion are accurate whatever the size of the sample, whether it include one observation or a million. Estimates of population parameters and of the probable significance of deviations are closer and in this sense more reliable for large samples than for small. A sample of 10 observations does not give so close an estimate as one of 30, and for that matter one of 30 does not give so close an estimate as one of 100. Nevertheless for many measures the estimate from a sample of 10 is valid as far as it goes and leads to correct conclusions as long as its probable closeness to the population parameter is understood. The fact that a closer estimate, if it had been possible, might have been more useful does not mean that a broader estimate is useless. It is necessary to take the best available and to do as much with this as possible, not to adopt the attitude that anything less than perfection is hopelessly bad.

In the second place, even in instances where estimates of probability based on the theory of large samples are unduly unreliable in their application to small samples, there are usually special methods, modifications, or corrections that do serve reliably for small samples. Some of these, those dealing with the most important of all numerical procedures in zoology, are given in the present chapter.

Finally, the statistical assumption that estimates may take any values whatever and that limitations are placed on them only by the sample itself is true mathematically but not zoologically. This makes the rigidity of statistical theory merely academic for the practical zoologist, who is not interested in mathematics for its own sake, and enables him to reach conclusions that cannot be absolutely proved in theory but that are correct in practice in such a large proportion of cases that he is quite justified in accepting them. The zoologist can, for instance, often assume a priori with high zoological (even though with very low mathematical) probability that the relative dispersion of a population will fall within certain limits, say that V will be greater than 3 and less than 10 in a given case. This has two useful corollaries: It often makes methods sufficiently reliable for small samples although they would not be so without this assumption; and it provides special methods for rational and useful even though broad estimation from single specimens, from which practically nothing can be learned regarding populations on the basis of the purely mathematical theory of large samples. This does not mean that mathematical procedures can be dispensed with: on the contrary, it makes them more useful and more necessary.

DEVIATIONS FROM THE MEAN IN SMALL SAMPLES

It has been noted that the standard deviation calculated from the formula

$$\sigma = \sqrt{\frac{\Sigma(d^2)}{N}}$$

may lead to somewhat inaccurate estimates of probability for small samples and that in such cases the value

$$\sigma' = \sqrt{\frac{\Sigma(d^2)}{(N-1)}}$$

in which σ' is used for this special small sample standard deviation to distinguish it from σ as hitherto used, often gives a better result. If the purpose of calculating the standard deviation is to estimate the significance of deviations, as it usually is, it will be found in practice that σ' hardly ever leads to a value of d/σ' so different from d/σ as to affect the conclusion as long as N is 10 or

greater and that it seldom does so even for values of N less than 10 but greater than 5. In general, although contrary to strict mathematical theory, it is a safe rule in almost all zoological practice that it is not worth while to use σ' for samples of 15 or more observations and that under 15 it is usually worth while only when

TABLE VI.—TABLE OF t

N	$N_1 + N_2$	P			
		.10	.05	.02	.01
2	3	6.3	12.7	31.8	63.7
3	4	2.9	4.3	7.0	9.9
4	5	2.4	3.2	4.5	5.8
5	6	2.1	2.8	3.7	4.6
6	7	2.0	2.6	3.4	4.0
7	8	1.9	2.4	3.1	3.7
8	9	1.9	2.4	3.0	3.5
9	10	1.9	2.3	2.9	3.4
10	11	1.8	2.3	2.8	3.3
11	12	1.8	2.2	2.8	3.2
12	13	1.8	2.2	2.7	3.1
13	14	1.8	2.2	2.7	3.1
14	15	1.8	2.2	2.7	3.0
15	16	1.8	2.1	2.6	3.0
16	17	1.8	2.1	2.6	2.9
17	18	1.7	2.1	2.6	2.9
18	19	1.7	2.1	2.6	2.9
19	20	1.7	2.1	2.6	2.9
24	25	1.7	2.1	2.5	2.8
∞	∞	1.6	2.0	2.5	2.8

the result is very near the criterion of significance so that great exactness has some real bearing on the conclusions reached.

It was also noted, however, that the distribution of deviations or of differences between single specimens and means or between means of two samples can be referred to an approximately normal curve only for relatively large samples and that this assumption tends to be unreliable for small samples. The estimate of prob-

abilities by the areas in the normal curve therefore tends to break down with small samples and needs replacement by tables for other probability distributions, varying with the size of the sample. Such tables have been worked out,[1] and a summary extract from them, sufficient for any zoological purpose, is given in our Table VI.

The form of the table is different from that of the d/σ tables, which serve the same purpose for larger samples; for here N must be taken into account, and the relationships are so complex that it is advisable to record and publish the value of P, instead of assuming that a figure analogous to d/σ will adequately convey the information. In order to distinguish between the measure of relative deviation used in this table and that of a d/σ table, it is called t in the present case. In essentials, t is simply a value analogous to any d/σ adjusted to allow for the size of the sample. In the table, N is the size of the sample,[2] and P is the probability that a value would appear by chance in random sampling, expressed as a decimal fraction (1.00 is 100 per cent probability). The entries in the body of the table are values of t corresponding to the values of N to the left and those of P above them. Thus, for instance, if a sample of 13 specimens gives a value of t between 2.2 and 3.1, P is between .05 and .01; and if t is greater than 3.1, P is less than .01. The $N_1 + N_2$ column is for use in comparing two samples.

Supposing d/σ and t to be nearly equal, as they almost always are for $N > 15$ and often are even for smaller values of N, it will be seen that the general criteria of significance given above for d/σ apply reasonably well to t except for very small samples. Thus it was said that:

1. $d/\sigma > 3$ is always significant. For $t > 3$, in a sample of 15 observations P is less than .01; in one of 10 observations P is less than about .015, and for one of 5 observations P is less than about .04.

[1] By "Student," a pseudonymous English statistician. Their mathematical derivation is complex and need not be followed, so long as their logical basis and practical use are understood.

[2] If reference is made to the more extensive table of t in Fisher and some other texts, note should be made that Fisher's n (N^1, etc., of some other authors) is not the total frequency of the sample but one or two less. Fisher's n is $(N - 1)$ in our notation for a single sample and $(N_1 + N_2 - 2)$ for two samples.

EXAMPLE 58.—COMPARISON OF A SINGLE SPECIMEN WITH THE MEAN OF A SMALL SAMPLE, TAIL LENGTH OF THE EASTERN VESPER SPARROW, *Pooecetes gramineus gramineus*, ADULTS IN SEPTEMBER AT EASTHAM, MASS.

(Data for distribution from Broun 1933)

X	Sample			
	f	d_A	fd_A	fd_A^2
60	1	−3	−3	9
61	0	−2	0	0
62	1	−1	−1	1
			−4	
63	1	0	3	0
64	1	1	1	1
65	1	2	2	4
	$N = 5$		$\Sigma(fd_A) = -1$	$\Sigma(fd_A^2) = 15$

$i = 1$

$c_1 = -1/5 = -.2 \qquad c_1^2 = .04$

$M = 62.8$

$\sigma = i\sqrt{15/5 - .04} = \sqrt{2.96} = 1.7$

$$\sigma' = i\sqrt{\frac{\Sigma(fd_A^2)}{(N-1)} - c_1^2} = 1 \times \sqrt{\frac{15}{4} - .04} = \sqrt{3.71} = 1.9$$

Single specimen not known to belong to sample has $X = 68$.

$d = X - M = 5.2$

$t = d/\sigma' = 5.2/1.9 = 2.7$.

From table of t, P is greater than .05.

The deviation is not shown to be significant.

$d/\sigma = 5.2/1.7 = 3.1$.

The deviation would appear to be significant if this procedure were used.

The calculation is given in detail to show the short method of getting σ', which is the same as for σ except for one substitution of $(N - 1)$ for N. Ordinarily σ would not be used in such a case, but it is also given in the example in order to show that it can give misleading results for very small samples with deviations near the boundary of significance. As was emphasized for d/σ, this does not prove that the difference is not significant but only that the data do not show it to be so. A bird with tail length 68 cannot be concluded to be from a different population from the sample here given. With a larger sample it might or might not prove to be from a different population.

2. $d/\sigma > 2.5$ is usually significant. For $t > 2.5$, in a sample of 15 observations P is less than about .025; in one of 10 it is less than about .04, and in one of 5 it is less than about .08.

3. $d/\sigma > 2$ is sometimes significant. For $t > 2$, in a sample of 15 observations P is less than about .04; in one of 10 it is less than about .09, and in one of 5 it is less than about .1.

It is thus seen that these criteria hold well enough for any practical purpose if N is 15 or greater, that they hold reasonably well in most instances for N between 10 and 15, but that they are generally not good enough for use if N is less than 10. For small values of N, t also tends to be smaller than a corresponding d/σ, so that this adds to the unreliability of the d/σ criterion in samples smaller than 10. On the other hand in dealing with most morphological variates there is a strong zoological probability that the dispersion will be limited, often that V will be less than 10, and this nonmathematical consideration adds to the zoological reliability of d/σ, which in such circumstances is greater than purely statistical procedures allow. Such considerations warrant our use of slightly broader or less exact criteria of significance under these special circumstances than are allowed by most statisticians.

For estimating probability for the deviation of one observation from the mean of a small sample, t is calculated exactly like d/σ except that σ' is used:

$$t = \frac{d}{\sigma'} = \frac{X - M}{\sqrt{\Sigma(d^2)/(N-1)}}$$

The use of this method is shown in Example 58.

SIGNIFICANCE OF MEANS OF SINGLE SAMPLES

It has been shown that σ_M, the standard error of the mean, serves to estimate the probability that the population mean is approximated by the mean of a sample. This measure can thus be used to estimate the probability that the mean of the sample differs significantly from any value set up by hypothesis as expected. Supposing M_A to be a hypothetical mean, the calculation is

$$\frac{d}{\sigma} = \frac{M - M_A}{\sigma_M}$$

entering a table of d/σ with the value thus obtained. Many experiments are so set up as to give an expectation that the mean result will be 0 under defined conditions. M_A then being 0, the appropriate analogue of d/σ for judging whether a value of M greater or less than 0 is significant or whether it could have arisen by chance is

$$\frac{d}{\sigma} = \frac{M}{\sigma_M}$$

These are formulas for large samples. For small samples another form of σ_M, which may be represented by σ'_M, is calculated.

$$\sigma'_M = \frac{\sigma'}{\sqrt{N}} = \sqrt{\frac{\Sigma(d^2)}{N \cdot (N-1)}}$$

In such a case the value of t, to be used in place of d/σ, is

$$t = \frac{M - M_A}{\sigma'_M}$$

or, if M_A is 0,

$$t = \frac{M}{\sigma'_M}$$

entering the usual table of t with the value thus obtained.

COMPARISON OF THE MEANS OF TWO SMALL SAMPLES

The use of d/σ, in the form d/σ_d, for comparison of the means of two large samples was explained in the last chapter. The same thing can be done with small samples, making adjustments for the use of $(N-1)$ in place of N, thus obtaining a t which is a form of d/σ_d adapted to small samples.

$$t = \frac{(M_1 - M_2)\sqrt{N_1 N_2/(N_1 + N_2)}}{\sqrt{\frac{\Sigma(d_1^2) + \Sigma(d_2^2)}{N_1 + N_2 - 2}}}$$

in which M_1, N_1, and d_1 are respectively the mean, total frequency, and deviations from the mean for one sample and M_2, N_2, and d_2 are the same for the other sample.[1]

[1] This formula is equivalent to $d/\sigma_d = \dfrac{(M_1 - M_2)}{\sqrt{\frac{N_1}{N_2}\sigma^2_{M_1} + \frac{N_2}{N_1}\sigma^2_{M_2}}}$ except that $(N_1 - 1)$ and $(N_2 - 1)$ are used in place of N_1 and N_2 for esti-

If the data include the standard deviations of the two distributions or the standard errors of their means, one of the following numerically equivalent formulas may be used:

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

$$t = \frac{(M_1 - M_2)\sqrt{\dfrac{N_1 N_2}{N_1 + N_2}}}{\sqrt{\dfrac{N_1^2\sigma_{M_1}^2 + N_2^2\sigma_{M_2}^2}{N_1 + N_2 - 2}}}$$

These formulas assume that σ, not σ', is used.[1]

In consulting the table of t, the sample size is read in the $N_1 + N_2$ column. The calculation and use of this method are shown in Example 59.

It should be noted that t cannot be used to test the validity of the results of dissecting a distribution or splitting a heterogeneous sample. What t measures in such uses is the probability that a difference observed between two means could have been equaled or exceeded by chance sampling of one population. If a sample is divided into two on any rational basis, t will always show the means of the resulting two samples to be significantly different, in other words, not to have arisen by chance—but this is obvious, for

---

mating the variance (but not for other purposes) and that the formula is given in terms of variance instead of standard errors of the means. It has been developed into a form convenient for calculation, although this somewhat obscures its derivation. Since, with the exception noted, it is derived in the same way as the formula for large samples, discussed in the preceding chapter, the derivation is not given here.

[1] If σ' were used, the denominators would be

$$\sqrt{\frac{(N_1 - 1)\sigma_1^2 + (N_2 - 1)\sigma_2^2}{N_1 + N_2 - 2}} = \sqrt{\frac{N_1(N_1 - 1)\sigma_{M_1}^2 + N_2(N_2 - 1)\sigma_{M_2}^2}{N_1 + N_2 - 2}}$$

where

$$\sigma = \sigma' = \sqrt{\frac{\Sigma(d^2)}{N - 1}}$$

$$\sigma_M = \frac{\sigma'}{\sqrt{N}} = \sqrt{\frac{\Sigma(d^2)}{N(N - 1)}}.$$

the samples were separated deliberately and not by chance. This has no bearing on the probability that they are from different populations.

EXAMPLE 59.—COMPARISON OF THE MEANS FOR THE RATIO LENGTH M_1:LENGTH M_2 IN TWO SMALL SAMPLES OF THE FOSSIL MAMMALIAN GENUS *Notostylops* FROM THE TWO LOCALITIES IN PATAGONIA
(Original data)

Sample	N	M	$\Sigma(d^2)$
Cañadón Vaca	6	.82	.0072
Colhué-Huapí	3	.89	.0051

$$t = \frac{(M_1 - M_2)\sqrt{\frac{N_1 N_2}{N_1 + N_2}}}{\sqrt{\frac{\Sigma(d_1^2) + \Sigma(d_2^2)}{N_1 + N_2 - 2}}} = \frac{(.82 - .89)\sqrt{\frac{6 \times 3}{6 + 3}}}{\sqrt{\frac{.0072 + .0051}{6 + 3 - 2}}}$$

$$= \frac{.07\sqrt{2.00}}{\sqrt{.00176}} = \frac{.099}{.042} = 2.4$$

$N_1 + N_2 = 9$

From Table VI, P is seen to be .05. Significance is possible but dubious.[1]

SIGNIFICANCE OF SINGLE OBSERVATIONS

Perhaps the commonest of all procedures in zoology is the comparison of the linear dimensions of two specimens. In taxonomic work, for instance, especially in paleontology where large samples are exceptional, a species is often represented chiefly or solely by its type, and identification of other specimens or description of other species involves comparison with this one specimen. The methods hitherto discussed are obviously inapplicable; and much as they may be desired the comparison must be made without their aid. The procedure, however, is fundamentally the same as if groups of specimens and not single examples were available and is much clarified and placed on a much sounder basis if the relationship to such groups is recognized.

A single dimension of one specimen is part of a distribution even though it is the only known part. It is in every case

[1] It happens in this case that several other variates in the same samples lie near the border of significance, and some are decidedly significant. It is then a reasonably safe zoological conclusion that the difference shown in the example is also significant.

assumed, consciously or unconsciously, that in comparing these isolated observations they are really considered as members of groups. This involves certain assumptions:

1. That the observation in hand is in fact representative of a group of more or less similar observations possible in nature; in other words that it belongs to a population.
2. That it is more or less characteristic of that group.
3. That the group has a certain variability, the extent of which is at least vaguely in mind.

The second of these assumptions involves the belief that the observed value of the variate is within a limited range about tht unknown mean of the population. Now it has been seen thae the probability that the specimen is within a given distance of the mean can be closely estimated if the value of σ (the standard deviation) is known. In a normal curve, a little over 95 per cent of the area is included between $(M - 2\sigma)$ and $(M + 2\sigma)$; or in other words, the chances are somewhat more than 95 in 100 that any one random observation will be within a distance of 2σ of the mean. Study of actual distributions of variates of reasonably homogeneous samples will readily show that the observed range does in fact nearly coincide with these limits in most cases.

For present purposes, then, the theoretical or inferred range of a variate on which only one or a few observations are available may be taken to be from $(M - 2\sigma)$ to $(M + 2\sigma)$. This implies that the specimen at hand is within those limits or is not over 2σ from the mean, an assumption sufficiently probable to be acceptable as a tentative working basis, in default of better evidence.[1]

The value of σ is unknown; but if, as is necessary for the drawing of most sorts of inferences from single observations, a value of V, the coefficient of variation, is assumed, then by direct calculation from the equation $V = 100\sigma/M$, it is possible to calculate a hypothetical value of σ corresponding with any position that the unique specimen may be assumed to have in the distribution, that

[1] Hitherto, $(M \pm 3\sigma)$ has been noted as the best figure to use for theoretical range, and these limits are often nearly reached by very large samples. Here the problem is different. It is estimating where in the range a single observation would be likely to fall. Usually, $(M \pm 2\sigma)$ is adequate for this purpose. Moreover, as will be shown, the particular way in which these limits are used makes the conclusions usually still more reliable than appears from the use of 2σ.

is, with any value of its deviation from M and the consequent assumed position of M. By assuming a value for V and supposing the specimen to be first at $(M - 2\sigma)$ and second at $(M + 2\sigma)$, its extreme probable positions, the extreme probable positions of M and of σ for the otherwise unknown group of which the specimen is a member may be calculated.

In such biological inferences, the assumption of a value for V is readily guided by analogy with known calculated values from larger samples of other groups. It is, for instance, reasonable to assume that if V for a linear dimension of a tooth is known on fairly good data in one or several species and appears to be fairly constant, it will have about the same value in another sample of these species and will not differ greatly in another closely related species. In general, in dealing with linear dimensions of functional unreduced anatomical elements, V may be assumed probably not to exceed 10, and it is almost never so great as 15. Usually it is from 3 to 7, provided that the sample is reasonably homogeneous.

The general concept here involved may be most readily represented by graphic means. In graphs, the range, real or, as in this case, theoretical or inferred, is represented by a horizontal distance. From the meaning and derivation of V, it is clear that the greater the V, the greater the linear distance representing the range, provided that M remains the same. If, however, V is supposed to remain the same, the basis of the present procedure, but M can take different values, then the greater the mean, the greater the range. It is not necessary to the solution of the problem but would certainly greatly facilitate and clarify things if the representation of the range could be made the same for any given value of V or retain the same proportion to the value of V, whatever the value of M—in other words, if the ratio $\frac{(M + 2\sigma) - (M - 2\sigma)}{V}$ were constant. This will be true if instead of the direct values of the various figures involved their logarithms are used or if (what amounts to the same thing) they are plotted on logarithmic graph paper.[1]

An example, plotted by logarithms, is given in Fig. 19. In this figure, A represents known constants for the length of M_3 of the

[1] It seems unnecessary to prove this fact here. If in doubt, refer to any elementary discussion of the properties of logarithms.

fossil mammalian species *Plesiadapis gidleyi*. The circle is the mean, the short vertical lines plus and minus 1σ, and the longer vertical lines plus and minus 2σ. The horizontal line is the actually observed range. The dots in the series B represent the actual length of a single M_3 of the series *P. tricuspidens*, a close

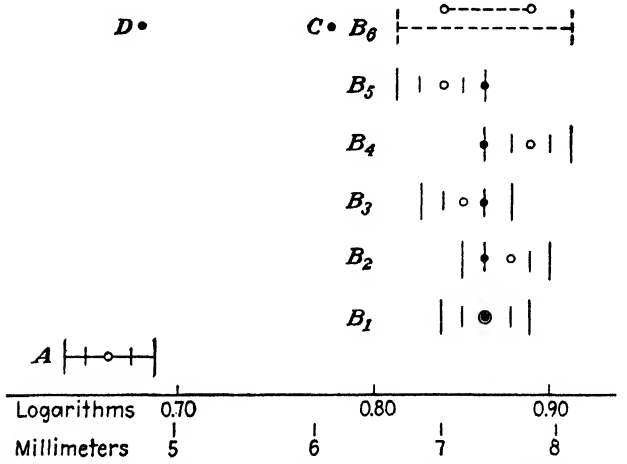


FIG. 19.—Graphic estimation of the probable variation of a species represented by a single specimen. The variate is length of last lower molar in specimens of the extinct mammalian genus *Plesiadapis*. The horizontal scale is logarithmic, and the vertical scale is not used. In *A* are shown the mean (circle), $(M \pm 1\sigma)$ (short verticals), and $(M \pm 2\sigma)$ (long verticals) for an adequate sample of *Plesiadapis gidleyi*. In *B*, all the black dots represent a single specimen of *Plesiadapis tricuspidens*. The short and long verticals are laid out as in *A* on the hypothesis that *V* (not σ) is the same for this species as for *P. gidleyi*. In B_1, the single specimen is assumed to be at the mean; in B_2, at $(M - \sigma)$; B_3, at $(M + \sigma)$; in B_4, at $(M - 2\sigma)$; in B_5, at $(M + 2\sigma)$. In B_6, the distance between verticals is the greatest theoretical range of this variate in *P. tricuspidens*, and the circles represent extreme theoretical positions of the true mean. This range is so far removed from that of *A* that it is certain that the two species are different in this variate. *C* and *D* represent single specimens that have been supposed to belong to *P. tricuspidens*. Their distance from the range shown in B_6 shows that they do not belong there.

ally of *P. gidleyi*. The assumption is made that *V* for this species is the same as for the latter. The corresponding σ values may then be shown by making the horizontal distances (plotted by logarithms) the same as in A. As in the latter, circles represent means, short vertical lines 1σ, and longer vertical lines 2σ.

In B_1 the specimen is assumed to be at the mean for its species, in B_2 at -1σ, in B_3 at $+1\sigma$, in B_4 at -2σ, and in B_5 at $+2\sigma$. The theoretical extreme probable range of the species is then the

extreme distance between -2σ of B_5 and $+2\sigma$ of B_4 and is shown above as a broken horizontal line. In practice the result may be reached by simply marking off on each side of the actual position of the specimen a distance equal to that between the longer vertical lines of A. Any specimen that does not fall within this range probably differs significantly from *P. tricuspidens* in the length of M_3. Any specimen within this range cannot be assumed with any probability to differ significantly, on these data, although of course it might prove to do so were fuller data available. The circles connected by a dotted line represent the extreme probable positions of the mean of the species to which B belongs.

The points C and D represent two isolated specimens that one student (Lemoine) believed to differ specifically from *P. tricuspidens* but that another student (Teilhard) referred to that species. Both are far outside the theoretical extreme range for the species, and the conclusion is that they almost certainly have significantly shorter last lower molars than in *P. tricuspidens* and therefore probably are not of that species.

The various values involved may also and somewhat more usefully be computed. This computation is much simplified by using the logarithms of the values, known and hypothetical, of the variate. The theoretical range corresponding to a given value of V will then have the same value regardless of the value of M, in other words, of the absolute size of the species. This theoretical range, on the distribution of logarithms, is then, by definition

$$D = \log (M + 2\sigma) - \log (M - 2\sigma)$$

Since $V = 100\sigma/M$,

$$\sigma = \frac{VM}{100}$$

Substituting this expression for σ in the formula for the theoretical range D,

$$D = \log \left(M + \frac{VM}{50}\right) - \log \left(M - \frac{VM}{50}\right)$$

And by a series of eliminations this becomes

$$D = \log (50 + V) - \log (50 - V)$$

Table VII gives values of D corresponding to integral values of V up to 20, and also antilogarithms of D, which represent the numerical value of $(M + 2\sigma)/(M - 2\sigma)$, the quotient of the upper theoretical limit divided by the lower, the use of which is explained below.

TABLE VII.—TABLE OF THE QUOTIENT OF UPPER LIMIT OF RANGE DIVIDED BY LOWER LIMIT (ANTILOG D) AND OF ITS LOGARITHM (D) CORRESPONDING TO A FIXED VALUE OF THE COEFFICIENT OF VARIATION (V)

V	D	antilog D
1	.017	1.04
2	.035	1.08
3	.052	1.13
4	.070	1.17
5	.087	1.22
6	.105	1.27
7	.122	1.33
8	.140	1.38
9	.158	1.44
10	.176	1.50
11	.194	1.56
12	.213	1.63
13	.231	1.70
14	.250	1.78
15	.269	1.86
16	.288	1.94
17	.308	2.03
18	.327	2.13
19	.347	2.23
20	.368	2.33

The general graphic solution of the problem may be made from these figures without the preliminary plotting of a known variate (A in the example above) and for any known value of V. On ordinary graph paper, place a dot at the point corresponding to the logarithm of the absolute value of the variate in question. Assume any desired value for V, and mark off on each side of the dot the distance designated by the corresponding D. The total distance thus indicated (twice D, with the observed value of the variate in the middle) is the extreme probable range of the hypothetical species, and a second specimen must fall outside this to be probably significantly different as regards this variate.

Another form of comparison between two specimens possibly of one species is to make the probable assumption that the smaller is at or above $(M - 2\sigma)$, the larger at or below $(M + 2\sigma)$, and to calculate the least value of V for a species which would include both. Subtract the logarithm of the smaller measurement from that of the larger, set this as equal to [log (50 + V) − log (50 − V)], and solve for V, or with this as D find the closest corresponding value of V in the table given above.

For instance, comparing B and D of the above example,

Length M_3 of B − 7.2 mm.....	log	.85733
Length M_3 of D − 4.8 mm.....	log	.68124
Difference in logs, D =		.17609
Minimum value of V =		10, from Table VII

Another exactly equivalent working of this type of problem is to divide the larger by the smaller number, when D will be the logarithm of the quotient (the subtraction of the logarithms, as above, being equivalent to a division and giving the logarithm of the quotient). The corresponding V may be found simply and with sufficient accuracy by considering this quotient as the antilog of D and finding the nearest V in the table above.

This relationship has an important application in a very common type of comparison different in form but basically the same as the examples just given. It is commonly said that a given dimension in one specimen is a certain percentage of, or percentage larger or smaller than, the same dimension in another specimen. This percentage is of course a quotient. If one value of a variate is 25 per cent greater than or 125 per cent of another, the quotient of the larger divided by the smaller is 1.25; and if the logarithm of this number is taken as D, the minimum probable V of a species including both can be read directly from the table. In this case, V is between 5 and 6. As before, the percentage, as a quotient, may be considered as antilog D and used to enter the table.

It is a commonly used criterion in some taxonomic work that a difference of 15 per cent in a linear dimension is in itself of probable specific value. This is traditional rule of thumb with no scientific basis. Clearly it implies some criterion as to variability, the most useful and exact measure of which (in dealing with analogous linear dimensions) is V. Now a difference of 15 per cent implies a quotient of 1.15, and the corresponding minimum probable value

of V is between 3 and 4. This is a small value even for a pure sample of a single local race, and the conclusion is that the "15 per cent rule" will inevitably separate some specimens of a single species.

The nearest integral values of V corresponding to a given percentage of difference are given in Table VIII. As before, V refers to the distribution of this variate in a hypothetical population to which both observations belong, on the assumption that one observation is at $(M - 2\sigma)$, and the other at $(M + 2\sigma)$.

TABLE VIII.—PERCENTAGE DIFFERENCE BETWEEN TWO OBSERVATIONS AND CORRESPONDING MINIMUM VALUE OF V IN A POPULATION INCLUDING BOTH

Percentage difference	V
5	1
10	2
15	3
20	5
25	6
30	7
40	8
50	10
60	12
70	13
80	14
90	16
100	17

In general, relying on no other criteria or characters, it is improper to assume that a difference in a linear dimension may characterize different taxonomic groups unless it implies a minimum V of at least 10, since V's as high as 10 or even somewhat higher are known to occur in such characters of pure races. Therefore if a percentage rule of thumb is to be used at all in comparing isolated specimens, a difference of 50 per cent is necessary to indicate sufficient probability of a real taxonomic distinction. This is, of course, true only in dealing with a single observation on a single variate and when there is no valid reason to assume a value for V lower than 10.

It should be emphasized that the purpose and result of this procedure are to prove the probability of a negative: to show that it is highly probable that two observations do *not* belong to one population. It cannot prove that they *do* belong to one population. It is sound procedure for the burden of proof to be on the

negative; and this is especially true in taxonomy, in which many inferences of this sort are involved. It must be assumed that any two comparable specimens are of one species unless it can be shown to be highly probable that they are not, and in general it is usually proper to assume that any two observations are from one population until the contrary is proved.

The degree of accuracy and probability achieved and the exact meaning of what is demonstrated in these procedures are made clear by the following considerations. Consider the distribution of the variate in question for the theoretically infinite population of which a single observation is available as a single normal curve of definite constants, and consider how much of this normal curve is actually included within the limits given by the above procedure (within the range $2D$ when plotted by logarithms). If the unique observation happened to be at the actual mean, the limits here used would include 99.99 per cent of the real curve, or for any practical purpose 100 per cent. Any observation on the population would certainly be included in the theoretical limits. Now the farther the observation really is from the real mean, the lower this figure will be. In approximately 955 cases out of 1,000, however, which is a sufficiently high probability to serve as a practical criterion, it will fall within the limits of $(M - 2\sigma)$ and $(M + 2\sigma)$. Now even if the specimen was actually at one of these limits on the real curve, the theoretical limits will include 97.73 per cent of the original curve, so that 955 times out of 1,000 the procedure here suggested will include within the theoretical limits at least 97.73 per cent (and generally much more, approaching 100 per cent) of all observations of the population. This high degree of probability is the basis for saying that if a second observation is not within these limits it is probably not from the same population as the first and may reasonably be separated. In order to reach this degree of probability for the negative conclusion, which is sought, the limits have necessarily been so placed that in every case they extend somewhat beyond the real (but unknown and indeterminable) limits of the population. For this reason, they may include observations that are in fact significantly different from those at hand but that cannot be shown to be so on these data. Therefore the positive proposition that all observations within the theoretical limits do belong to the same population is not true.

Finally the general validity of the procedure as establishing a reasonable degree of probability is shown by the fact that in less than 3 cases in 1,000 (0.27 per cent) will the unique specimen at hand deviate as much as 3σ from the unknown real mean in the distribution of the variate considered and that even in this extremely improbable case the theoretical limits here used would include 84 per cent of the real distribution.

CHAPTER XII

CORRELATION

The idea that two variates may be related so that one tends to vary in such a way as to maintain a fairly constant relationship to the values of the other is fundamental and widespread not only in zoology and in kindred sciences but also in the whole field of human thought. Tall parents tend to have tall children and short parents short children. Children tend to weigh more as they become older. Animals with larger heads tend also to have longer tails. The weather tends to be stormier when the sun has more spots. Increases in wages are generally accompanied by increases in living costs. Colored people tend to be darker the larger their percentage of African negro blood. A man is generally hungrier the longer he has gone without eating. These and thousands of other statements and ideas about every aspect of human life and of the universe are all examples of the correlation of two variates. All involve the idea that a change in one phenomenon or characteristic is usually accompanied by a change in some other.

It is not necessary to suppose that the relationship is constant and predictable in individual cases; indeed it is seldom so in practice. Tall parents do sometimes have short children, children do sometimes lose weight, numerous animals with large heads do have short tails, the weather is sometimes calm when the sun has many spots, etc., but on the whole these relationships hold good and it may confidently be stated from experience that they will always be true of averages if enough observations are made. It is obvious, however, that the correlation may differ in intensity or in the accuracy of its predictions.

Complete correlation or 100 per cent accuracy is almost confined to mechanical things and practically never occurs in biological data. That an automobile travels farther the more times its wheels turn around on the road is a positive correlation

that is 100 per cent true; but that larger individuals have larger offspring may be true only two-thirds of the time, and that larger animals dig deeper burrows may be true only one-third of the time. The intensity of a supposed correlation may also prove to be zero, or in other words the supposed relationship may not exist. For instance, the old idea that some plants grow better if planted in the dark of the moon is shown by recent experiments to be a correlation with value about zero. A negative correlation is one in which an increase in the value of one variate is usually accompanied by a decrease in the value of another, and those also are common in nature. For instance, molar-tooth crown height and age are negatively correlated in most mammals: the older a horse becomes, the less is the distance from grinding surface to roots on his cheek teeth.

Most data of this sort can be reduced to numerical terms, and once this is done it is possible to measure the exact degree of correlation shown by a given series of observations and from this as a sample to estimate the probable degree in a population. The importance of these procedures for zoology is so great and so obvious that it hardly needs emphasis, and mention of a few types of problems suffices to show how essential numerical correlation is in this field. Correlations between characters of parents and offspring show the effect and consistency of hereditary factors.[1] Correlation between environmental conditions (such as temperature) and physiological variates (such as respiration rate) or morphological characters (such as total length) show whether the latter are influenced by the former. Correlation of age with any physiological, psychological, or morphological characters reveals the phenomena of growth. Correlation of any two variates of an animal shows whether they are independent or linked together in some way, again throwing light on the phenomena of heredity and also on the functions of growth, on physiological mechanisms, and on many other problems. It is also interesting in this connection that different taxonomic groups are often characterized by different intensities of correlations of this type.

[1] Applied to human beings, this was the first field in which numerical correlation was used, and it sufficed to demonstrate and to measure heredity without experimentation and before anything was known of the mechanisms or laws of heredity.

THE CORRELATION COEFFICIENT

Several different means of measuring correlation numerically have been devised. They usually take as a basis a conception of perfect correlation as one in which an increase in the value of one variate is always accompanied by an exactly proportionate change (increase in positive correlation and decrease in negative) in the other. Thus in these series of paired observations

X	Y	X	Y
1	1	1	8
2	2	2	6
4	4	4	2

both are perfect correlations, the first positive, because every time that X increases the corresponding Y increases by an equal amount, and the second negative, because every time X increases Y decreases by an amount twice as great as the increase in X. Perfect correlation involves only the same relative change, not the same absolute change. The following series all have perfect positive correlation:

X	Y	X	Y	X	Y
1	.02	1	1.0	1	50
2	.04	2	1.5	2	100
4	.08	4	2.5	4	200

The measure of correlation is then a measure of how nearly the actual observations come to this relationship. The best measure would be analogous to the measure of dispersion for one variate by its variance or standard deviation but would involve a relationship between the variances of two variates as arranged in paired observations, their variance together or covariance. Such a measure, devised by Pearson, which has been found to have very useful properties, is the ratio of the mean product of corresponding deviations of the two variates[1] to the geometric mean of their

[1] It has been noted that a mean deviation is called a moment. This mean product of deviations is a product moment, and the correlation here discussed is generally called the product moment correlation.

separate variances. This measure is always symbolized by r, and the term "coefficient of correlation," if not qualified, generally refers to r, the formula for which is

$$r = \frac{\Sigma(d_X d_Y)}{\sqrt{\Sigma(d_X^2)\,\Sigma(d_Y^2)}}$$

in which d_X is any deviation of a value of X from the arithmetic mean of X and d_Y is the *corresponding* deviation of Y from the arithmetic mean of Y.[1] The formula is exactly equivalent to

$$r = \frac{\Sigma(d_X d_Y)}{N\sigma_X\sigma_Y}$$

in which N is the number of pairs of observations and σ_X and σ_Y are the standard deviations of X and Y respectively.[2]

This formula gives the value $+1$ for perfect positive and -1 for perfect negative correlation, 0 for absence of any correlation, and other values between these limits.[3] Two digits are usually significant, and it is customary to record two decimal places for r, although more may occasionally be useful.

[1] This is the same as the ratio of the mean product of the corresponding deviations to the geometric mean of the two variances and is derived as follows:

The ratio stated is exactly symbolized as

$$\Sigma(d_X d_Y)/N : \sqrt{\Sigma(d_x^2)\,\Sigma(d_y^2)/N_X N_Y},$$

N being the number of pairs and N_X and N_Y the number of observations of X and Y respectively. But under the conditions of correlation, values of X and Y are always paired; so N_X always equals N_Y, and either is equal to N. The expression thus becomes

$$\frac{\Sigma(d_X d_Y)}{N} \div \sqrt{\frac{\Sigma(d_X^2)\,\Sigma(d_Y^2)}{N^2}} = \frac{\Sigma(d_X d_Y)}{N} \times \frac{N}{\sqrt{\Sigma(d_X^2)\,\Sigma(d_Y^2)}} = \frac{\Sigma(d_X d_Y)}{\sqrt{\Sigma(d_X^2)\,\Sigma(d_Y^2)}}.$$

[2] This follows from the relationship of variance and standard deviation.

[3] It is, however, important that the values between 0 and ± 1 are not simply fractional parts of complete correlation in any simple or obvious sense and that the steps are not of equal value in relation to significance in most problems. Thus $r = .80$ does not mean 80 per cent of complete correlation, and the step from $r = .10$ to $r = .20$ is not so important a difference as that from $r = .80$ to $r = .90$. The relative complexity of these relations is among the arguments for transforming r into a measure with simpler properties (see the following pages).

EXAMPLE 60.—CALCULATION OF r

Correlation of number of tail scutes and tail length in females of the kingsnake, *Lampropeltis polyzona* (data from Blanchard 1921)

Caudal scutes X	Tail length Y	d_X	d_Y	$+d_Xd_Y$	$-d_Xd_Y$	d_X^2	d_Y^2
45	37	−5	−68	340	.....	25	4,624
46	130	−4	25	.....	100	16	625
46	137	−4	32	.....	128	16	1,024
47	156	−3	51	.....	153	9	2,601
48	49	−2	−56	112	.....	4	3,136
48	93	−2	−12	24	.....	4	144
49	68	−1	−37	37	.....	1	1,369
49	155	−1	50	.....	50	1	2,500
49	106	−1	1	.....	1	1	1
49	86	−1	−19	19	.....	1	361
51	142	1	37	37	.....	1	1,369
51	142	1	37	37	.....	1	1,369
52	53	2	−52	.....	104	4	2,704
52	51	2	−54	.....	108	4	2,916
52	50	2	−55	.....	110	4	3,025
53	149	3	44	132	.....	9	1,936
54	187	4	82	328	.....	16	6,724
55	146	5	41	205	.....	25	1,681
55	54	5	−51	.....	255	25	2,601
951	1,991			1,271	−1,009	167	40,710

$M_X = 50$ $M_Y = 105$ $\Sigma(d_Xd_Y) = +262$ $\sigma_X = 2.96$ $\sigma_Y = 46.3$

$$r = \frac{\Sigma(d_Xd_Y)}{\sqrt{\Sigma(d_X^2)\Sigma(d_Y^2)}} = \frac{262}{\sqrt{167 \times 40{,}710}} = \frac{262}{2{,}607} = +.10$$

or,

$$r = \frac{\Sigma(d_Xd_Y)}{N\sigma_X\sigma_Y} = \frac{262}{19 \times 3 \times 46} = \frac{262}{2{,}622} = +.10$$

(The slight difference in denominators is caused by rounding in obtaining σ_X and σ_Y; the values obtained for r are identical with the second decimal, which is sufficient for most purposes.)

The example has been chosen as one of a type common in zoological work, as being arithmetically relatively easy, and as testing a relationship that might well exist but the presence or absence of which is not obvious from the raw data. The value of r obtained is very small, +.10; in fact, as will be shown later in this chapter, it does not differ significantly from 0. As far as this sample shows, the number of caudal scutes is not correlated with the length of the tail.

The correlation coefficient may be calculated from a list of the corresponding measurements of the two variates, using either form of the formula, as shown in Example 60 (see Fig. 21, page 232).

Even with this simple example, this method of calculation is rather laborious; and with larger samples or where still more digits must be carried, it becomes almost prohibitive unless a calculating machine is available. There is, however, another method that facilitates reduction of the complexity of the data and of the arithmetic and also gives a good visual representation of the correlation. It involves making a correlation table and basing the calculation on this, using a procedure analogous to the short method of getting M and σ.

A correlation table is made on paper ruled into squares, with values of one variate scaled vertically at the left side and those of the other horizontally along the top of the table. The data can be grouped in any convenient way, and the values written on the scale are conventional class limits. Within the table there is thus a square, or cell, corresponding to any combination of values of the two variates, and the observed frequencies are entered in the cells.[1]

By totaling the frequency in each line, across the table from left to right, an ordinary frequency distribution for one variate is obtained and written to the right of the table. By totaling the columns, from top to bottom, such a distribution is obtained for the other variate and written below the table; M_X, M_Y, σ_X, and σ_Y can then be obtained from these distributions by the short method. The additional datum needed to get r is $\Sigma(d_X d_Y)/N$. A value $fd_X d_Y$ is calculated for each cell in which a frequency of 1 or more occurs. These may be written in the cells themselves, to one side, or, more conveniently, in a column to the right of the other tabulations, negative and positive values being put in different columns for ease in algebraic addition. In this connection it is helpful to note that the table, divided into quadrants by the mean row and mean column, has two diagonally opposite quadrants in which values of $d_X d_Y$ are positive and two in which they are negative. Which are positive and which negative

[1] In publication this is usually done in ordinary numerals. In tabulating from the original data before calculation, it is most easily done by tally checks.

EXAMPLE 61.—CORRELATION TABLES AND SHORT CALCULATIONS OF M_X, M_Y, σ_X, σ_Y, AND r_{XY}

A. Tail length and total length in females of the king-snake, *Lampropeltis polyzona* (same specimens as in Example 60; data from Blanchard, 1921).

X—tail length

Y—body length	30–49	50–69	70–89	90–109	110–129	130–149	150–169	170–189	f	d_A	fd_A	fd_A^2	$d_{A_X}d_{A_Y}$ +	$d_{A_X}d_{A_Y}$ −
1,350–1,449	..	....	...	...	....	...	...	1	1	+5	5	25	15	(none)
1,250–1,349	..	....	...	...	....	...	1	...	1	+4	4	16	8	
1,150–1,249	..	....	...	...	....	...	1	...	1	+3	3	9	6	
1,050–1,149	..	....	...	...	....	5	...	...	5	+2	10	20	10	
950–1,049	..	....	...	...	....	1	...	...	1	+1	1	1	1	
850–949	..	....	...	...	....	...	...	...	0	0	+23 −41			
750–849	..	....	...	1	....	...	...	...	1	−1	− 1	1	1	
650–749	..	....	...	1	....	...	...	...	1	−2	− 2	4	2	
550–649	..	....	1	...	....	...	...	...	1	−3	− 3	9	6	
450–549	..	1	...	...	....	...	...	...	1	−4	− 4	16	12	
350–449	1	4	...	...	....	...	...	...	5	−5	−25	125	80*	
250–349	1	....	...	...	....	...	...	...	1	−6	− 6	36	24	
f	2	5	1	2	0	6	2	1	19	—	−18	262	165	0
d_A	−4	− 3	−2	−1	0	+1	+2	+3	—	Totals for *Y*				
fd_A	−8	−15	−2	−2	+13 −27	6	4	3	−14	$\Sigma(d_{A_X}d_{A_Y}) = 165$				
fd_A^2	32	45	4	2	....	6	8	9	106					

Totals for *X*

* Note that deviations on the *Y*-scale are the same for all values in this row; therefore the products can be obtained by totaling deviations from A_X and multiplying by the deviation from A_Y. If the cells were each recorded separately, they would read $\{-5[1 \times (-4)]\} + \{-5[4 \times (-3)]\}$ or $(+20) + (+60) = +80$ for the row. Totaling deviations by rows, the procedure is $-5\{[1 \times (-4)] + [4 \times (-3)]\} = -5[(-16)] = +80$.

EXAMPLE 61. Continued on opposite page.

depends on the direction in which the scales are written. If the scale is as in the next example (Example 61), in the most usual form, the positive quadrants are the upper right and lower left, and the other two are negative. Taking d_X and d_Y from assumed means, as in the short method for M and σ, the result is properly $\Sigma(d_{A_X}d_{A_Y})/N$ and requires correction to give the desired value for deviations from the true mean. This correction factor is simply the product of the two correction factors, c_1 for X and Y, used for obtaining σ by the short method. Thus the true value of $\Sigma(d_Xd_Y)/N$ is

$$\frac{\Sigma(d_{A_X}d_{A_Y})}{N} - c_{1_X}c_{1_Y}$$

This gives the value in units equal to the class interval. Since in any case the class interval would appear in both numerator and denominator of the formula for r and so would cancel out, it is easier to use this and, of course, also to use σ_X and σ_Y in class interval units, or σ_X/i and σ_Y/i.

Although perhaps confusing when expressed in words, the operations involved are really simple and may best be learned by working examples. Two are given in Example 61 (see also Fig. 21, page 232).

The tabulation shows the following values, which in practice are not written down separately:

$\Sigma(fd_{A_X}) = -14$ $\quad \Sigma(fd_{A_Y}) = -18$

$\Sigma(fd^2_{A_X}) = 106$ $\quad \Sigma(fd^2_{A_Y}) = 262$

$c_{1_X} = -14/19 = -.737$ $\quad i_X = 20$ $\quad c_X = -14.7$

$c^2_{1_X} = .5432$ $\quad M_X = 120 - 14.7 = 105.3$

$\sigma_{1_X} = \sqrt{106/19 - .5432} = 2.244$

$c_{1_Y} = -18/19 = -.947$ $\quad i_Y = 100$ $\quad c_Y = -94.7$

$c^2_{1_Y} = .8968$ $\quad M_Y = 900 - 94.7 = 805.3$

$\sigma_{1_Y} = \sqrt{262/19 - .8968} = 3.59$

$c_{1_X}c_{1_Y} = (-.737)(-.947) = +.698$

$\frac{\Sigma(d_Xd_Y)}{N}$ in terms of class intervals $= \frac{165}{19} - .6979 = 7.9863$

$\sigma_X\sigma_Y$ in terms of class intervals $= 2.24 \times 3.59 = 8.0416$

$r_{XY} = \frac{\Sigma(d_Xd_Y)}{N} \times \frac{1}{\sigma_X\sigma_Y}$ either in units of measurement or, as here, in class intervals $= 7.9863/8.0416 = .9931 = .99$

The correlation is positive and is remarkably high, indeed almost perfect.

EXAMPLE 61.—CORRELATION TABLES AND SHORT CALCULATIONS OF M_X, M_Y, σ_X, σ_Y, AND r_{XY}.—(*Continued*)

B. Lengths of M_1 and M_2 in the fossil mammal *Phenacodus primaevus* from the Gray Bull formation of the Bighorn Basin (original data)

Y—length M_1	X—length M_2														
	10.0–10.4	10.5–10.9	11.0–11.4	11.5–11.9	12.0–12.4	12.5–12.9	13.0–13.4	13.5–13.9	f	d_A	fd_A	fd_A^2	$+d_{A_x}d_{A_y}$	$-d_{A_x}d_{A_y}$
13.5–13.9								1	1	+4	4	16	12	
13.0–13.4						2	1	1	4	+3	12	36	21	
12.5–12.9							1	2	3	+2	6	12	16	
12.0–12.4					2	1	2	1	6	+1	6	6	8	
11.5–11.9					1				1	0	+28 −17			
11.0–11.4		1	2		1	3			7	−1	− 7	7	4*	
10.5–10.9		3							3	−2	− 6	12	18	
10.0–10.4									0	−3	0			
9.5– 9.9	1								1	−4	− 4	16	6	
f	1	4	2	0	4	6	4	5	26	—	11	105	95	
d_A	−4	− 3	−2	−1	0	+1	+2	+3	—					
fd_A	−4	−12	−4	0	+29 −20	6	8	15	9					
fd_A^2	16	36	8	0		6	16	45	127					

* $-1\{[1 \times (-3)] + [2 \times (-2)] + [3 \times (+1)]\} = -1(-3 - 4 + 3) = -1 \times -4 = +4$

$c_{1_X} = \frac{9}{26} = .346$ $\quad i_X = .5$ $\quad c_X = .173$ $\quad c_{1_X}c_{1_Y} = .346 \times .423 = .146$

$c_{1_X}^2 = .1197$ $\quad M_X = 12.2 + .173 = 12.37$ $\quad \frac{\Sigma(d_X d_Y)}{N} = \frac{95}{26} - .146 = 3.51$

$\sigma_{1_X} = \sqrt{127\frac{1}{26} - .1197} = 2.18$ $\quad \sigma_X\sigma_Y = 2.18 \times 1.96 = 4.27$

$c_{1_Y} = 1\frac{1}{26} = .423$ $\quad i_Y = .5$ $\quad c_Y = .212$ $\quad r_{XY} = \frac{3.51}{4.27} = +.82$

$c_{1_Y}^2 = .1789$ $\quad M_Y = 11.7 + .212 = 11.91$

$\sigma_{1_Y} = \sqrt{105\frac{5}{26} - .1789} = 1.96.$

This is also a decisively significant positive correlation, but the value of r is lower and more usual for such data.

Aside from the relative ease of calculation,[1] placing the data in this tabular form also gives an immediate picture of the relationship. If the frequencies in the table tend to be arranged along an oblique line, then they show some correlation. If the

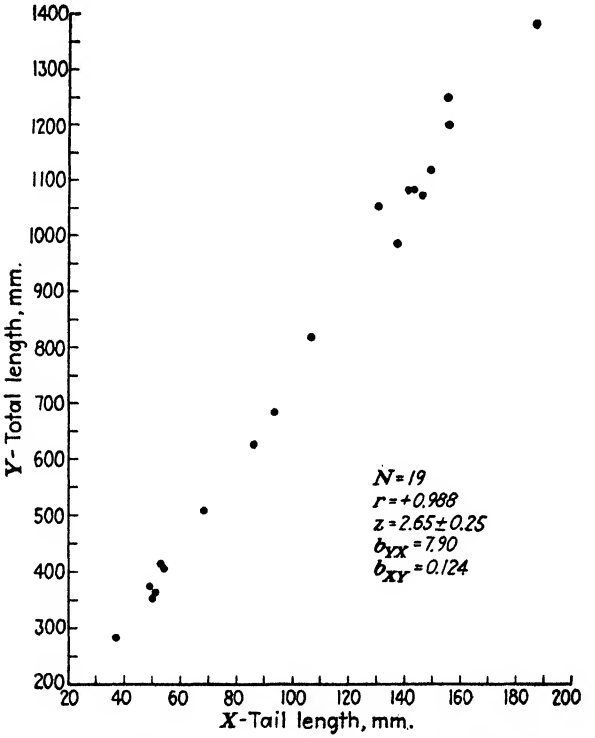


FIG. 20.—Scatter diagram of a nearly perfect positive correlation. Total length against tail length in females of the snake *Lampropeltis polyzona* (data of Example 61). The evident trend from lower left to upper right shows that the correlation is positive, and the very close clustering of the observations around a straight line shows that it is very strong (r = +.99).

line runs from lower left to upper right (as in Example 61), the correlation will be positive; if from upper left to lower right, negative. The straighter the line, the more closely the frequencies are concentrated on it, and the more nearly it bisects the

[1] If "relative ease" seems too strong an expression, note that these calculations give the useful and generally necessary values of M and σ for both X and Y, as well as r, and yet are not much more complicated and are no more difficult than getting M and σ for these two variates separately.

quadrants, the higher the correlation. In Example 61 the frequencies are more linear and concentrated in A than in B, and r is greater in A than in B. If, on the other hand, the frequencies are about equally scattered around the center of the diagram, are about equally divided between positive and negative quadrants, or are grouped roughly horizontally or vertically, the correlation will be low and probably not significant. Example 60 is such a case, as is clearly seen without calculation by putting it in tabular form as in Example 62.

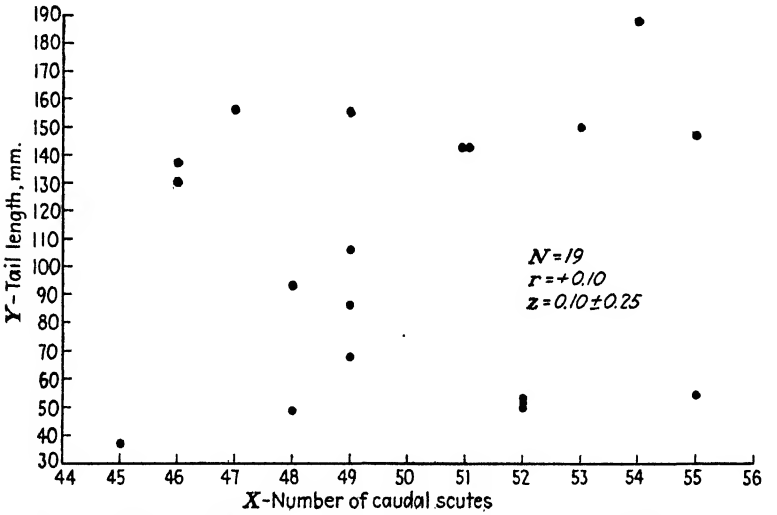


FIG. 21.—Scatter diagram of the relationship of two variables without significant correlation. Tail length against number of caudal scutes in females of the snake *Lampropeltis polyzona;* data of Example 60. The absence of evident trend and fairly even scattering of observations in all four quadrants show absence of correlation (r = +.10, which is far from significant).

The use of r is based on the assumption that a significant correlation will be not only linear but also rectilinear. There are, however, cases in which a strong correlation exists but is curvilinear. Such cases can usually be identified by sight in a correlation table. With a slight curve, r is usually still sufficiently valid and useful. With a strong curve, however, r is not a valid measure and should not be employed. The most common instances of this in zoology are in correlations of age and of variates involved in growth, which may be nearly linear for limited periods of growth but are almost always strongly curved

at some period in the whole life span. Some suggestions for handling this important and difficult type of relationship are given in Chap. XVI.

EXAMPLE 62.—RAW DATA OF EXAMPLE 60 (PAGE 226) IN TABULAR FORM

Y—tail length	X—caudal scutes											
		45	46	47	48	49	50	51	52	53	54	55
	180–194	..	..	..	..	..	..	..	..	..	1	
	165–179											
	150–164	..	..	1	..	1						
	135–149	..	1	..	..	..	..	2	..	1	..	1
	120–134	..	1									
	105–119	..	..	..	..	1						
	90–104	..	..	..	1							
	75– 89	..	..	..	..	1						
	60– 74	..	..	..	..	1						
	45– 59	..	..	..	1	..	..	..	3	..	..	1
	30– 44	1										

There is here no suggestion of linear arrangement, and the frequencies in positive and negative quadrants are about equal. As would be expected from this pattern, calculation showed the correlation to differ little from zero ($r = .10$). With a little practice, it is possible to see at a glance from such a table whether the correlation is likely to be significant or not, which may save the labor of calculation.[1]

SIGNIFICANCE OF r

Even if no real correlation, positive or negative, occurs in a population, it is very unlikely that a sample will show a correlation of 0.00. It will almost certainly show chance fluctuations one way or the other, and, as with estimates of the parameters of distributions, it is necessary to have some means of judging the probable value of r in the population as opposed to the more or less chance value in a sample. In the case of correlation, the most important point is to determine whether a deviation of r from zero could have arisen by chance. If it could, the observed

[1] But of course if the correlation might be significant, r should be calculated, and it should usually be calculated if the table shows it clearly to be significant. If it is certainly not significant, there is seldom any point in calculating r.

correlation cannot be taken as certainly real in the population. If it could not, a significant degree of correlation is demonstrated to exist in the population. The obvious method would be to calculate a standard error for r. Then, according to the theory of errors already explained, r would be certainly significantly different from zero if its value were three or more times its own standard error. This is, indeed, commonly used as a test of the significance of r, the formula for its standard error being

$$\sigma_r = \frac{1 - r^2}{\sqrt{N}}$$

or (for small samples) more consistently

$$\sigma_r = \frac{1 - r^2}{\sqrt{N - 1}}$$

Despite its wide use, however, this test is so uncertain and so often wholly erroneous that it should be abandoned. It is reasonably valid only for large samples and for very low values of r. In zoology, large samples for correlation are exceptional, and a measure valid only for them is of limited use. In any field it is not the low values of r that need testing, for they are seldom significant in any case, but those of medium or relatively high value, for instance, beyond $\pm .35$. For these, the standard error is usually so misleading that its use is worse than applying no test at all.

The reasons for this are: (1) that a small sample may give a very unreliable value for σ_r; and (2) and more important that even for large samples and relatively good estimates of σ_r of the population, the distribution of this is approximately normal only if the value of r is low. For increasingly high values of r, the distribution of σ_r rapidly becomes strongly asymmetrical; and if r is very high, the distribution of σ_r is practically J-shaped. The use of a standard error as a test of significance assumes that its distribution is approximately normal; and since this is far from being the case for σ_r when r is large, it follows that the test is not valid in such circumstances.

The problem, then, is to obtain from r, or some other measure of correlation, either a value that has a known and tabulated distribution for all values of N regardless of the absolute size of r or a value that is approximately normal in distribution regardless

of the absolute values of both N and r. Both these solutions are available, and either is incomparably preferable to the use of σ_r. The first solution is analogous to that of finding criteria for the usual distribution parameters as estimated from small samples and is solved in an analogous way, by obtaining from N and r a value that is distributed like t and that can be used as t in a table of t, N, and P. The following is such a value:

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

N is the number of pairs of observations. In using Table VI, it is read in the $N_1 + N_2$ (and not the ordinary N) column.

In application to Example 60, this gives

$$t = \frac{.10\sqrt{19-2}}{\sqrt{1-(.10)^2}} = \frac{.412}{.99} = .42$$

N is 19, and so it is seen in Table VI (page 206) that P is much greater than .10 (it is in fact between .6 and .7) and hence that r is not significant. The standard error in this case is

$$\sigma_r = \frac{1-r^2}{\sqrt{N}} = \frac{1-(.10)^2}{\sqrt{19}} = \frac{.99}{4.359} = .23$$

Thus, r/σ_r is .10/.23 = .43. Using this as d/σ shows that P is greater than .6 (it is in fact between .66 and .67). The results of the two tests are about the same, because the sample is not unduly small and the value of r is very small, the conditions under which the use of r/σ_r is valid.

For Example 61B, t is

$$\frac{.82\sqrt{26-2}}{\sqrt{1-(.82)^2}} = 7.02$$

and P is less than .01, showing that r is significant. For this same example, σ_r is .064 and r/σ_r = .82/.064 = 12.8. The values of r and N are so large that significance is certain by either test; but the numerical results are quite different, and the r/σ_r test greatly overestimates the significance, a most important point when the limit of significance is more closely approached.

By the use of this formula for t, it is clearly possible to calculate a value of r for any value of N such that any equal or greater value of r will give a value of P not greater than any chosen criterion of significance. The effort of calculating t in each case can thus be avoided by tabulating these values, which is done in sufficient detail for most zoological purposes in our Table IX.[1]

[1] This is taken from Fisher, with some modifications. Fisher's n is 2 less than our N in this table.

TABLE IX.—SIGNIFICANCE OF THE CORRELATION COEFFICIENT r FOR DIFFERENT SIZES OF SAMPLES
(Using the distribution of t as a measure of probability)

N	P			
	.10	.05	.02	.01
3	.9877	.9969	.9995	.9999
4	.900	.950	.980	.990
5	.81	.88	.93	.96
6	.73	.81	.88	.92
7	.67	.75	.83	.87
8	.62	.71	.79	.83
9	.58	.67	.75	.80
10	.55	.63	.72	.76
11	.52	.60	.69	.73
12	.50	.58	.66	.71
13	.48	.55	.63	.68
14	.46	.53	.61	.66
15	.44	.51	.59	.64
20	.38	.44	.52	.56
22	.36	.42	.49	.54
27	.32	.38	.45	.49
32	.30	.35	.41	.45
37	.27	.32	.38	.42
42	.26	.30	.36	.39
47	.24	.29	.34	.37
52	.23	.27	.32	.35
62	.21	.25	.29	.32
72	.20	.23	.27	.30
82	.18	.22	.26	.28
92	.17	.21	.24	.27
102	.16	.19	.23	.25

Using the number of pairs of observations as N, this table shows the value of r corresponding with each of the four stated values of P. A value of r greater than that entered means that P is less than that of the given column. Thus in Example 61B,

N is 26. For the next smaller N of the table, 22, an r of .54 corresponds with a P of .01. Hence in the example an r of .82 means a P considerably less than .01, and the table shows the r to be significant, without the trouble of calculating t. This is, nevertheless, a use of t, since it was used to calculate the table. The significance of r should always be judged from this table or by calculating t, and not from σ_r. It will be noted that very small samples, less than 7, will almost never give surely significant values of r, since calculated values of r greater than .90 are rarely encountered. In fact it is seldom worth while to calculate r for less than 10 pairs of observations unless a correlation table shows an unusually clear-cut linear arrangement.

TRANSFORMATION OF r TO z

Fisher, to whom are owed so many of the methods of valid numerical study of small samples, has worked out a method of transforming r into another measure, symbolized as z, which minimizes or does away with the disadvantages of r and also has valuable additional properties of its own. For almost all zoological purposes, z is a better measure than r, and its general use is urged. Since, however, r is at present more familiar, it is advisable to give both r and z in publication. This entails no extra work, since r must be calculated to obtain z. z is the limit approached by adding an infinite series consisting of odd powers of r each divided by its exponent, that is,

$$z = r + \frac{r^3}{3} + \frac{r^5}{5} + \frac{r^7}{7} + \cdots$$

Since only odd powers are used, z is negative when r is negative. It can also readily be seen that when $r = 0$, $z = 0$ and that when $r = \pm 1$, $z = \pm \infty$ (infinity). For very low values of r, z is nearly equal to r; but when r is large, z is relatively still larger, and small increments in r make large increments in z. The calculation of z is most easily carried out with Napierian (or natural) logarithms[1] as follows:

[1] Logarithms with the constant e (= 2.7183) as a base, symbolized by $\log_e$, instead of 10 as in common logarithms, $\log_{10}$ or simply log. Most trigonometric texts and mathematical handbooks give tables of $\log_e$ sufficiently detailed for present purposes. Our table of r and z (Table X) will make it seldom necessary to calculate z, but the formula given here and any table of $\log_e$ make it possible to do so should occasion arise.

$$z = \frac{\log_e (1 + r) - \log_e (1 - r)}{2}*$$

The method of calculation is shown in Example 63.

EXAMPLE 63.—TRANSFORMATION OF r TO z

A. Data of Example 60

$r = .10 \qquad 1 + r = 1.1 \qquad 1 - r = .9$

$\log_e 1.1 = 10.095 - 10$

$\log_e .9 = 9.895 - 10$

$\log_e (1 + r) - \log_e (1 - r) = .200$

$$z = \frac{\log_e (1 + r) - \log_e (1 - r)}{2} = \frac{.200}{2} = .10$$

For such a low value of r, z and r are practically identical.

B. Data of Example 61A

$r = .99 \qquad 1 + r = 1.99 \qquad 1 - r = .01$

$\log_e 1.99 = 10.688 - 10$

$\log_e .01 = 5.395 - 10$

$\log_e (1 + r) - \log_e (1 - r) = 5.293$

$z = 5.293/2 = 2.6465$ or 2.65

For such a high value of r, z and r are very unequal.

The most important properties of z are:

1. It is dependent on r but varies from $-\infty$ through 0 to $+\infty$, whereas r varies from -1 through 0 to $+1$.

2. When the value of r is low, below about .5, the value of z is nearly the same.

3. When the value of r is high, that of z is still higher; and the disproportion increases more rapidly the nearer r approaches its limits ± 1.

4. For the relatively unimportant lower levels of correlation, r and z thus have steps of about equal value; but for the more important higher levels, z has increasingly smaller steps relative to r and hence is an increasingly delicate measure as delicacy becomes more desirable.

5. Although z can theoretically take any values to $\pm\infty$, in practice it almost never exceeds ± 3.[1]

6. The standard error of z is independent of its own value and is the simplest of all commonly used standard errors to calculate. It is

$$\sigma_z = \frac{1}{\sqrt{N - 3}}$$

* This is equivalent to the other equation for z previously given, and it can easily be seen that it also gives values from 0 to $\pm\infty$ as r varies from 0 to ± 1. Thus when $r = 0$, this formula reduces to 0/2 which of course is 0; when $r = \pm 1$, $\log_e (1 - r)$ is $-\infty$, hence the formula gives $(\log_e 2 + \infty)/2 = +\infty$; and when $r = -1$, the formula becomes $(-\infty - \log_e 2)/2 = -\infty$.

[1] $z = \pm 3$ corresponds with $r = \pm .995$, a correlation so nearly perfect that it is almost never encountered.

7. When as often happens, several correlations are based on the same body of data, all the resulting z's have the same standard error, and all differences between them also have a single standard error. This is exceptionally convenient in itself and also makes it practical to table these standard errors. Neither is true of r.

TABLE X.—CORRESPONDING VALUES OF r AND z
(All values may be either + or −)

z	r	r	z
.1	.10	.05	.05
.2	.20	.10	.10
.3	.29	.15	.15
.4	.38	.20	.20
.5	.46	.25	.26
.6	.54	.30	.31
.7	.60	.35	.37
.8	.66	.40	.42
.9	.72	.45	.48
1.0	.76	.50	.55
1.1	.80	.55	.62
1.2	.83	.60	.69
1.3	.86	.65	.78
1.4	.89	.70	.87
1.5	.91	.75	.97
1.6	.92	.80	1.08
1.7	.94	.85	1.26
1.8	.947	.86	1.29
1.9	.956	.87	1.33
2.0	.964	.88	1.38
2.1	.970	.89	1.42
2.2	.976	.90	1.47
2.3	.980	.91	1.53
2.4	.984	.92	1.59
2.5	.987	.93	1.66
2.6	.989	.94	1.74
2.7	.991	.95	1.83
2.8	.993	.96	1.95
2.9	.994	.97	2.09
3.0	.995	.98	2.30
>3.0	1.00	.99	2.65
		1.00	3.0

The following points will be noted: r and z do not differ in the first decimal place below $z = .6$, $r = .5$; also, z goes from 1.08 to 2.65 while r is going from .80 to .99; and in this important part of the scale z has 158 steps (to two decimal places—it is, of course, a continuous variate) where r has only 20.

8. The significance of the difference between two values of z is very easy to estimate, that between two values of r very difficult.

9. Although the distribution of r changes enormously both in peakedness (or kurtosis) and in symmetry as the value of r changes, the shape of the distribution of z is almost constant whatever its value.

10. Although the distribution of r is never normal and is not even roughly so when the value of r is large, the distribution of z approaches the normal so closely that it may safely be called normal in practice whatever the value of z.

11. Although the value of r is too inaccurate for small samples to serve as a valid estimate of the correlation of the population, that of z is generally sufficiently accurate for this purpose even in small samples.

12. From properties just mentioned, it follows that σ_r is a very poor and often quite invalid means of testing the significance of r, while σ_z is almost always a valid and adequate means of testing the significance of z.

EXAMPLE 64.—CALCULATION OF σ_z AND z/σ_z FOR THE DATA OF EXAMPLE 63 (PAGE 238)

A.

$$N = 19 \qquad \sigma_z = \frac{1}{\sqrt{N-3}} = \frac{1}{\sqrt{16}} = \frac{1}{4} = .25$$

$$z = .10 \qquad \frac{z}{\sigma_z} = \frac{.10}{.25} = .40$$

The table of d/σ shows that P is greater than .6 (it is in fact between .68 and .69), and hence the value of z is not significant. Note that this result is very close to that of the accurate t test on the same data (page 235) which gave $t = .42$ and P between .6 and .7, while r/σ_r is .43. Even with this fair-sized sample and very low r, the r/σ_r value differs slightly from that of z/σ_z and is to that extent less reliable.

B.

$$N = 19 \qquad \sigma_z = \frac{1}{\sqrt{16}} = .25$$

$$z = 2.30 \qquad \frac{z}{\sigma_z} = \frac{2.30}{.25} = 9.2$$

This is decisively significant. Note that the value of σ_z is the same as in A, although the value of z is nearly 30 times greater. This peculiar property also increases the delicacy of the z/σ_z test for large values of r or z. For this same example, σ_r is .009, making r/σ_r about 110, a remarkably gross overestimate.

It thus appears that z is far superior to r in every respect except ease of calculation, and this difficulty is largely done away with by the use of tables. Fisher tables r and z in considerable detail. Our simpler and somewhat different Table X is adequate for most purposes.

Since z is approximately normal in distribution, z/σ_z is a good measure of significance, entering a table of d/σ with this value. This test and the calculation of σ_z are shown in Example 64.

Values of σ_z for samples of sizes usual in zoology are given in Table XI, which in most cases will make it unnecessary to calculate this value.

TABLE XI.—STANDARD ERRORS OF z AND OF THE DIFFERENCE BETWEEN TWO VALUES OF z FROM SAMPLES OF EQUAL SIZE

N	σ_z	σ_{d_z}
5	.71	1.00
6	.58	.82
7	.50	.71
8	.45	.63
9	.41	.58
10	.38	.53
11	.35	.50
12	.33	.47
13	.32	.45
14	.30	.43
15	.29	.41
16	.28	.39
17	.27	.38
18	.26	.36
19	.25	.35
20	.24	.34
21	.24	.33
22	.23	.32
23	.22	.32
24	.22	.31
25	.21	.30
26	.21	.29
27	.20	.29
28	.20	.28
29	.20	.28
30	.19	.27
35	.18	.25
40	.16	.23
50	.15	.21
75	.12	.17
100	.10	.14

N is the number of pairs for σ_z. For σ_{d_z} it is the number of pairs in each sample (not in both together). The table gives values of σ_{d_z} only when both

samples are of equal size. When they are not, the formula

$$\sigma_{d_z} = \sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}$$

must be used.

DIFFERENCE BETWEEN TWO CORRELATION COEFFICIENTS

As with means and other measures, it is often desirable or necessary to estimate whether the difference between the correlation coefficients of two samples is significant. This is practically impossible from r alone. It is possible to get a t value analogous to that for the difference between two means, but the work is complex and laborious. If, however, the r's are transformed to z's, the comparison can be made by a standard error

EXAMPLE 65.—COMPUTATION OF STANDARD ERRORS AND SIGNIFICANCE OF DIFFERENCES BETWEEN TWO CORRELATION COEFFICIENTS

A. Total length and tail length in males and females of *Lampropeltis polyzona* (raw data from Blanchard 1921)

	N	r	z
1. Males	24	.974	2.16 } Calculation
2. Females	19	.988	2.56 } not shown

$$\sigma_{d_z} = \sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}} = \sqrt{\frac{1}{21} + \frac{1}{16}} = \sqrt{.1101} = .33$$

$$d_z = z_2 - z_1 = 2.56 - 2.16 = .40$$

$$\frac{d_z}{\sigma_{d_z}} = \frac{.40}{.33} = 1.21.$$ The difference is not significant.

B. Total length and tail length in females of *Lampropeltis polyzona* and of *Lampropeltis elapsoides elapsoides* (raw data from Blanchard 1921)

	N	r	z
1. *L. polyzona*	19	.988	2.56
2. *L. e. elapsoides*	25	.899	1.47

$$\sigma_{d_z} = \sqrt{1/22 + 1/16} = \sqrt{.108} = .33$$

$$d_z = 2.56 - 1.47 = 1.09$$

$$\frac{d_z}{\sigma_{d_z}} = \frac{1.09}{.33} = 3.3.$$ The difference is significant.[1]

[1] For such comparisons it is often desirable to work out r and z more exactly than in Table X, and this has been done in the example. Using the table would, however, have given the same result but with slight overestimate of significance ($d_z/\sigma_{d_z} = 3.6$).

of the difference which is very easy to obtain. It is

$$\sigma_{d_z} = \sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}$$

in which N_1 and N_2 are the numbers of pairs of observations in the respective samples. The calculation and ensuing test of significance are shown in Example 65.

Tabling σ_{d_z} for all combinations of N_1 and N_2 would be involved and seems unnecessary. It often happens, however, that $N_1 = N_2$ (for instance in partial correlation, discussed later); and values of σ_{d_z} under this condition are given in Table XI.

CAUSE AND EFFECT AND SPURIOUS CORRELATION

Measures of correlation and tests of their significance demonstrate only that two variates are shown or are not shown by a given sample to vary in such a way as to tend to maintain a definite relation to each other. They tell nothing about the cause of the relationship but reveal only its existence or the lack of reliable evidence for it.[1] There is a danger in passing from these numerical results to biological conclusions that the relationship may be misunderstood or may, without due consideration, be assumed to represent cause and effect. The fallacy is that most common in vulgar thought, *post hoc, ergo propter hoc* (or *cum hoc propter hoc*), and is among the foundations of much of the untrue "natural wisdom" of the unlettered and of astrology, alchemy, and the other false sciences. On these discredited levels the fallacy is easy enough to detect, and numerical correlation itself disposes of many logically obscure cases by showing that the supposed or apparent correlation is not significant.

This fact, however, increases the danger of supposing that a causal relationship does exist when the correlation is found to be numerically significant, although even on this level common sense is an adequate safeguard against the more egregious blunders. It may be shown that rodents are more abundant in an area when crop prices are lower, in other words, that number of rodents and crop price show significant negative correlation; but no one would conclude that rodents tend to lower crop prices or that cheap crops tend to increase rodent families. It is obvious that the

[1] Of course they do not reveal nonexistence.

two phenomena are really quite independent except as both may depend on a third factor, crop abundance or, still more remotely, favorable weather. The relations may in other cases be so obscure that considerable analysis is necessary to differentiate cause and effect or to distinguish a real correlation from one that is spurious, being caused not by any true relationship between the variates considered but by their relationship to a third variate that has been omitted from the problem.

It is not true that any correlation between variates not truly related as cause and effect is to be considered spurious. Some of the most useful correlations are between two variates both of which are affected by the same unmeasured cause. In fact, such correlations may reveal the existence of an important variate or characteristic that cannot be directly observed and may serve to measure it indirectly by measuring its effects. Such correlations are common in zoology, and that between lengths of two adjacent teeth (Example 61B) is a good example. Obviously these are not cause and effect; but they are, or more properly their relationship is, the result of some cause that cannot otherwise be detected or measured. The data do not show what this cause is, whether genetic, environmental, or otherwise. Its existence having been detected in this way, its nature remains to be determined, if possible, by further observation, calculation, and experimentation.

There is no general rule for differentiating true and spurious correlation beyond the application of logic and testing of other possible correlations of the variates in question. It is, however, usually true that the correlation technique is properly applied to two variates that may be related as cause and effect or that may be analogous effects of a cause not directly observable.

Aside from purely spurious correlation, a true correlation may have its value falsified by the existence of another distinct correlation not excluded from the data and tending to either increase or decrease the correlation that is sought. For instance, correlation between thymus size and body size of a species of mammal may correctly be sought, and its result may throw important light on the biology of the species; but a good sample of the whole species will not give the correct value of the correlation. Thymus size will probably be found to be positively associated with body size in animals of the same age, but negatively correlated with body

size if age is not excluded because it shows a stronger negative correlation with age than positive correlation with body size.

Correlation either with age or with body size is real and significant, but correlation with either that does not exclude the effects of the other gives a false value. Such relationships may become very complex, but their disentanglement can often be accomplished in one of two ways. It may be possible to select a sample such that all but two of the variates involved, or possibly involved, in the correlation are made as nearly constant as possible, thus making possible a true observation of the correlation of those two variates. Sometimes, however, this will reduce the available sample so that it is too small to detect a significant correlation, and even with abundant data it may be impossible to keep all but two variates approximately constant. In such cases the only practical solution is, if possible, to measure all the variates involved in the correlation and to derive a corrected true correlation value for any two of them from the interrelationships of all.

PARTIAL CORRELATION

The process just mentioned is called partial correlation. When a measurable correlation involves more than two factors or variates, partial correlation is a technique for finding how much of the correlation value results from the relationship between any two of these variates. A variate the value of which depends on that of two or more other variates is said to show multiple correlation; and partial correlation, defined in another way, is the measurement of the relationship between the value of the given variate and that of any one of the other variates on which it depends. Partial correlation is used when it is not practical to keep all but two variates involved in correlation constant, and it gives a numerical result not significantly different from what would have been obtained had this been possible.

Partial correlation is possible only when corresponding values of the three or more variates involved can be obtained from the same body of data. When this is possible, the process is surprisingly simple in concept, although the calculation involved may be laborious, sometimes prohibitively so if there are more than three variates. With three variates the formula is

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

The subscripts 1, 2, and 3 refer to the three variates. The first step is to correlate these in pairs from the original data, correlating the first two and obtaining a coefficient r_{12}, the first and third, giving r_{13}, and the second and third, giving r_{23}. These are total correlations, because in obtaining them nothing was excluded—the value of r_{12} may still be influenced by that of r_{13} and of r_{23}. The symbol $r_{12.3}$ represents the partial correlation of variates 1 and 2 with the effect of the variation of 3 on either of them eliminated. Similarly a partial correlation $r_{13.2}$ (of 1 and 3 with 2 eliminated) may be obtained by transposing 2 and 3 throughout the formula and a partial correlation $r_{23.1}$ by transposing 1 and 3.

If there are four or more variates, it is necessary to eliminate them one at a time. Thus with four variates it is first necessary to obtain $r_{12.4}$, $r_{13.4}$, and $r_{23.4}$ by applying the formula three times and then to use these values, with the effect of 4 eliminated, in the formula to obtain $r_{12.34}$, the correlation of 1 and 2 with the effects of the variability of both 3 and 4 eliminated, as follows:

r_{12}, r_{13}, r_{14}, r_{23}, r_{24}, and r_{34} are first calculated, and then the following:

$$r_{12.4} = \frac{r_{12} - r_{14}r_{24}}{\sqrt{(1 - r_{14}^2)(1 - r_{24}^2)}}$$

$$r_{13.4} = \frac{r_{13} - r_{14}r_{34}}{\sqrt{(1 - r_{14}^2)(1 - r_{34}^2)}}$$

$$r_{23.4} = \frac{r_{23} - r_{24}r_{34}}{\sqrt{(1 - r_{24}^2)(1 - r_{34}^2)}}$$

$$r_{12.34} = \frac{r_{12.4} - r_{13.4}r_{23.4}}{\sqrt{(1 - r_{13.4}^2)(1 - r_{23.4}^2)}}$$

With four variates the rather laborious operation summarized in the formula thus has to be performed four times, and this labor increases more rapidly than does the number of variates: for five variates the formula is calculated ten times, for six, twenty times, and for ten, two hundred and twenty times.[1] In some sociological, medical, psychological, and kindred studies, partial correlation with large numbers of variates is used. For most

[1] The labor is shortened by using Miner's tables of $(1 - r^2)$ and $\sqrt{1 - r^2}$, cited in our bibliography, but remains considerable when the number of variates is more than three or four.

EXAMPLE 66.—COMPUTATIONS OF PARTIAL CORRELATIONS WITH THREE VARIATES

Maximum length, width, and height of crowns of last upper molars of the extinct mammal *Acropithecus rigidus* (original data)

Variates:

1. Length
2. Width
3. Height

$N = 28$

Total correlations:[1]

$r_{12} = +.355$

$r_{13} = +.795$

$r_{23} = -.046$

Partial correlations:

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} = \frac{.355 - .795 \times (-.046)}{\sqrt{(1 - .795^2)(1 - .046^2)}} = \frac{.355 + .037}{\sqrt{.368 \times .998}}$$

$$= \frac{.392}{.605} = +.648$$

$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{23}^2)}} = \frac{.811}{.933} = +.867$$

$$r_{23.1} = \frac{r_{23} - r_{13}r_{12}}{\sqrt{(1 - r_{13}^2)(1 - r_{12}^2)}} = \frac{-.328}{.562} = -.584$$

Tests of significance:

$$\sigma_z = \frac{1}{\sqrt{25}} = .20 \text{ (same for all } z\text{'s)}$$

$r_{12} = .355$	$z_{12} = .37$	$\frac{z}{\sigma_z} = \frac{.37}{.20} = 1.85$	Not significant.
$r_{13} = .795$	$z_{13} = 1.08$	$\frac{z}{\sigma_z} = \frac{1.08}{.20} = 5.40$	Surely significant.
$r_{23} = -.046$	$z_{23} = -.05$	$\frac{z}{\sigma_z} = \frac{-.05}{.20} = -.25$	Not significant.
$r_{12.3} = .648$	$z_{12.3} = .77$	$\frac{z}{\sigma_z} = \frac{.77}{.20} = 3.85$	All surely significant.
$r_{13.2} = .867$	$z_{13.2} = 1.32$	$\frac{z}{\sigma_z} = \frac{1.32}{.20} = 6.60$	
$r_{23.1} = -.584$	$z_{23.1} = -.67$	$\frac{z}{\sigma_z} = \frac{-.67}{.20} = -3.35$	

[1] These are calculated as in Example 61. The work of calculating the three total correlations is much less than three times the work of calculating one correlation because the correction factor and standard deviation (c_1 and σ_1, if in terms of class intervals) need to be calculated only once for each variate despite the fact that each is used twice in the course of obtaining the three total correlations. This part of the work, which is the most laborious, is therefore only one and one-half times as great for these three correlations as for one.

zoologists, the use of three variates or at most four is adequate. The process of obtaining $r_{12.3}$ is shown in Example 66.

The problem chosen, although relatively simple arithmetically as an example of partial correlation, has somewhat complex zoological connotations, and the partial correlations give information not visible or directly deducible from the original data and not so well shown by any other method of analysis. The unexpected nature of some of these results and their hardly obvious zoological meaning exemplify very clearly (1) the need for numerical analysis in such problems and (2) the fact that the results of such analysis are not zoological conclusions but must be interpreted logically and with care in the light of the real meaning of each operation, starting with the gathering of the original measurements. The zoological consideration of this problem and of its numerical conclusions will therefore be given in more detail than for most previous examples, since the reasoning involved is analogous to that for any problem of partial correlation and cannot be taken for granted.

The measurement of length was taken on the wearing surface of these teeth, and so it changes or may change with degree of wear. This is in this case almost invariably the maximum length of a tooth as preserved. It would be preferable to take a length measurement not affected by wear, but this was not done in this case for three reasons: because it had not usually been done in samples of allied animals, and hence the measurement as taken is more nearly comparable with data already available; because it is desirable to find out what influence wear does have on this length measurement; and because length on the wear surface can be accurately measured in all cases, and other lengths can be accurately measured only on isolated teeth out of the jaw. The maximum width is at the base of the tooth and had not been visibly affected by wear in any specimen measured. The maximum crown height would doubtless vary in unworn teeth, but all these teeth were in fact worn to various degrees. The differences in height caused by wear were obviously very much greater than any such differences could have been on the unworn teeth, so that this measurement is for all practical purposes a measure of degree of wear rather than of genetic differences in tooth heights. It is thus indirectly an approximate but useful measure of individual ages. Age cannot, of course, be measured directly for fossils; but it is certain that if it could it would show a very high positive correlation with molar wear and hence an equally high negative correlation with molar height. It can thus be reasonably assumed that molar height is a good approximate inverse indirect measure of age.

In order to understand and interpret this sample correctly, it is necessary to know and measure two things that are not apparent from the original data: the relationship between length and width and the relationship

between length and age, or wear, or molar height. We want, in other words, reliable measures of the correlation of 1 and 2 and of the correlation of 1 and 3. The total correlation z_{12} gives a result that is not significant, but evidently this may not be reliable; for if either 1 or 2 is also correlated with 3, which is known from the method of measurement to be probable for 1, then 3 also influences the value of z_{12}. Similarly it cannot be assumed that z_{13} is a reliable measure, for 2 may well be correlated with 1 and hence may influence this value. The only way to get the information sought is therefore somehow to eliminate any influence of 3 on the correlation of 1 and 2 and any influence of 2 on the correlation of 1 and 3.

One solution would be to select specimens in which height, 3, is the same for all and to obtain a value of z_{12} from them, then to select specimens of the same width, 2, and obtain a value of z_{13}. If this were done, z_{12} and z_{13} would be reliable measures of the relationships being investigated; but in this case it could not be done because the largest available samples with nearly constant height or width would (as the original measurements show) have included only three or four specimens, too small to give a useful value of z. Recourse is therefore had to partial correlation, which produces the same result but uses all the observations made. $z_{12.3}$ and $z_{13.2}$ are reliable measures of the two relationships sought.

z_{12} does not differ significantly from 0, but $z_{12.3}$ shows a significant positive correlation. The zoological conclusion is that for any given height, hence in any given wear stage, any given period of life, or for unworn teeth, greater length tends to be accompanied by greater width, and vice versa. It then follows, zoologically, that length and width are affected by some common influence, almost surely genetic, and further that they would tend to have an approximately constant ratio in teeth unaffected by wear. These are important conclusions essential to intelligent study of the specimens involved, and they could not be reached by any more direct consideration of the actual measurements available.

The value of z_{13} differs significantly from 0 and that of $z_{13.2}$ still more so, showing that a correlation existing between 1 and 2 or between 2 and 3 tended to reduce and to that extent to falsify the value of z_{13}. For constant width, there is a strong positive correlation between length and height. There are really two factors in this, both tending in the same direction so that correlation of length with only one of them would be less; but they cannot be separated on these data, and their separation is not really necessary. One factor is variation in original unworn height, and the other is variation in height caused by wear; but the former is certainly very slight compared with the latter (as known on zoological grounds, although not numerically demonstrable from these data), and it may safely be concluded that there is a significant negative correlation between length and wear and beyond that between length and age. We have, then, demonstrated beyond reasonable doubt that in these teeth the length of the grinding surface tends to become less as the animal grows older. Aside from the importance of this fact in itself, it also has an important practical bearing, for it shows that comparisons of lengths of these teeth with any others will not be valid unless height is also taken into account and its influence eliminated or discounted.

In this case the value $z_{23.1}$ does not measure any relationship of zoological importance or answer any question naturally suggested by the data. In the actual research there was no reason to calculate or discuss this correlation, but it is given in our example because the value will probably be found surprising and because it well illustrates the danger of hasty judgment or of faulty nonnumerical reasoning in dealing with numerical procedures. The value of z_{23} is very small and certainly does not differ significantly from 0, but the value of $z_{23.1}$ shows a fairly strong, surely significant negative correlation between width and height if the effect of length on these is discounted. On the face of it, this seems to mean that these teeth become wider (at the base, beyond the reach of the actual wear) as they are worn down, which might be considered either as a sensational discovery or as a manifest absurdity.

In fact this is not at all what the negative partial correlation means. It means only that in teeth of a given length those that are narrower will tend to be less worn. We know, from the value of $z_{13.2}$, that length and height are positively correlated. It can then logically be seen that a tooth in which the unworn length was greater will have to be worn down to some degree before its length becomes equal to that of a smaller unworn tooth and hence that when two teeth have the same length the tooth that was longer when unworn will tend at this stage to have the lesser height. But we also know from the value of $z_{12.3}$ that the tooth longer when unworn will, at any wear stage, tend to have the greater width. Hence it follows inevitably that in a miscellaneous sample with all stages of wear represented greater width will be associated with lesser height, in other words, that width and height will be negatively correlated. This is what the value of $z_{23.1}$ means, and it has no particular zoological significance, certainly not that suggested at first sight. It is simply a corollary of the high positive values of $z_{13.2}$ and $z_{12.3}$ and is seen to follow inevitably when the real meaning of these is kept in mind.

RANK CORRELATION

Aside from r, the only directly calculated measure of correlation in general use is Spearman's measure of rank correlation, usually designated by ρ (Greek lower-case rho). This method arranges the two series of values in the order of their magnitude and gives each a rank according to its position in the series. Thus the largest value of X is ranked as 1, the largest value of Y also as 1, the next value of each 2, etc. The differences in rank between corresponding values of X and Y are then recorded (none being regarded as negative) and the measure calculated as follows:

$$\rho = 1 - \frac{6\Sigma(d^2)}{N(N^2 - 1)}$$

where d = a difference in rank.

N = the number of pairs of observations.

EXAMPLE 67.—THE USE OF ρ
A. Calculation of ρ from data of Example 61B

Original measurements		Ranks		Difference in rank	
X	Y	X	Y	d	d^2
9.8	10.2	1	1	0	0
10.5	10.7	2.5	2.5	0	0
10.5	10.7	2.5	2.5	0	0
10.8	10.8	4	4.5	0.5	0.25
11.0	11.0	5.	6	1	1
11.1	11.4	6.5	7	0.5	0.25
11.1	12.1	6.5	9	2.5	6.25
11.3	12.6	8	14	6	36
11.4	12.8	10	17	7	49
11.4	10.8	10	4.5	5.5	30.25
11.4	12.6	10	14	4	16
11.9	12.3	12	10	2	4
12.2	12.4	13.5	11.5	2	4
12.2	12.0	13.5	8	5.5	30.25
12.3	13.7	15.5	25	9.5	90.25
12.3	13.0	15.5	18	2.5	6.25
12.4	13.2	17.5	20	2.5	6.25
12.4	12.4	17.5	11.5	6	36
12.5	13.8	19	26	7	49
12.7	13.5	20	22.5	2.5	6.25
12.8	13.3	21	21	0	0
13.0	12.7	22	16	6	36
13.1	13.1	23	19	4	16
13.2	13.6	24	24	0	0
13.4	12.6	25	14	11	121
13.5	13.5	26	22.5	3.5	12.25
					$\Sigma(d^2) = 556.50$

$N = 26$

$$\rho = 1 - \frac{6\Sigma(d^2)}{N(N^2-1)} = 1 - \frac{6 \times 556.50}{26(26^2-1)} = 1 - \frac{3{,}339}{17{,}550} = 1 - .19 = .81$$

r for these data is .82, a very close agreement.

If two or more observations have the same value and hence are tied for rank, they are best given the middle value of the ranks that they would occupy if different, and the next higher value is given its correct serial ranking. The value of ρ is generally about the same as that of r if the original distributions were approxi-

EXAMPLE 67.—THE USE OF ρ.—(*Continued*)

B. Demonstration of some relationships between ρ and r

1. The following hypothetical distribution has $r = +1$:

		Ranks		d	d^2
X	Y	X	Y		
1	1	1	1	0	0
2	2	2	2	0	0
4	4	3	3	0	0
					$\Sigma(d^2) = 0$

$$\rho = 1 - \frac{6 \times 0}{3(9 - 1)} = 1 - 0 = 1 \qquad \text{When } r = +1, \quad \rho = +1.$$

2. The following has $r = -1$:

		Ranks		d	d^2
X	Y	X	Y		
2	6	1	3	2	4
4	5	2	2	0	0
6	4	3	1	2	4
					$\Sigma(d^2) = 8$

$$\rho = 1 - \frac{6 \times 8}{3(9 - 1)} = 1 - 2 = -1 \qquad \text{When } r = -1, \quad \rho = -1.$$

mately normal; and ρ is somewhat easier to calculate than is r, but this is almost its only advantage. The use of ρ in zoology is not recommended except as a relatively quick way of finding whether r is likely to be significant, or when rank can be determined but accurate and equally spaced absolute values cannot, or when accuracy and exact tests of significance are not required.[1] The

[1] Some workers habitually calculate ρ, because this is easier, and then convert it to r by means of a formula that we prefer not to give. This procedure is thoroughly unsound and should never be followed. The conversion formula demands the assumption that the original observations on both variates were normally distributed. If this assumption is correct, the greatest possible difference between ρ and r is about .02, which is practically never a significant difference, and usually it is less. If the assumption is not correct, the difference between ρ and r may be greater than .02 and may be

EXAMPLE 67.—THE USE OF ρ.—(*Continued*)

3. The following has $r = 0$:

		Ranks		d	d^2
X	Y	X	Y		
1	1	1.5	1.5	0	0
1	3	1.5	3.5	2	4
3	1	3.5	1.5	2	4
3	3	3.5	3.5	0	0
					$\Sigma(d^2) = 8$

$\rho = 1 - \frac{6 \times 8}{4(16 - 1)} = 1 - .8 = .2$ When $r = 0$, ρ is small but is not necessarily 0. (It is 0 only when the distributions are normal.)

4. The following has $\rho = +1$:

X	Y	d_x	d_y	$d_x d_y$	d_x^2	d_y^2
1	1	−1.33	−1	1.33	1.7689	1
2	2	−0.33	0	0	0	0
4	3	1.66	1	1.66	2.7556	1
				$\Sigma(d_x d_y) = 2.99$	$\Sigma(d_x^2) = 4.5245$	$\Sigma(d_y^2) = 2$

$M_x = 2.33$

$M_y = 2.00$

$r = 2.99/\sqrt{4.5245 \times 2} = .99$ When $\rho = +1$, r is large but is not necessarily $+1$. (And when $\rho = -1$, r has a large negative value but is not necessarily -1.)

calculation is shown in Example 67, as well as a demonstration of some of the relationships between ρ and r.

From their respective formulas, it can be seen that ρ shows perfect correlation, $+1$, if all the corresponding ranks of X and Y are in the same order and hence if Y increases or decreases with every increase of X (or vice versa), regardless of

---

significant; but then the formula does not properly correct this and does not give a reliable value for r. In other words, when the formula works it is not needed, and when it is needed it does not work. ρ and r are not interconvertible under the ordinary conditions of research, and the attempt should not be made.

the amount of increase or decrease, while r does not show perfect correlation unless this condition is fulfilled and also the condition that the ratio between increments of X and Y is constant. For this reason, for certain types of curvilinear correlation ρ will give higher and perhaps more truly representative values than will r; but this relationship is usually unreliable in practical use, and ρ may overestimate or (less often) underestimate the true value of the correlation.

CHAPTER XIII

REGRESSION

SCATTER DIAGRAMS AND TREND

The present chapter is devoted to the graphic analysis and to certain special types of study of the relationship between two variates, involving data similar to that of correlation and methods supplementing those of correlation as discussed in Chap. XII. Correlation measures the intensity of such relationships but gives no additional information about its nature. A correlation table itself does give a picture of the nature of the relationships involved, but this can be more clearly and also more precisely shown on ordinary graph paper and can be measured or numerically indicated only by methods additional to those of correlation. Taking the values of one variate as X and those of the other as Y, scales of X and Y are laid out along the bottom and to the left, respectively, of the field of the proposed diagram on graph paper. A dot is then placed in the field at a point corresponding to each pair of observed values of X and Y. The result is called a scatter diagram (Figs. 20 and 21).

If there is a significant correlation between the two variates of a scatter diagram, the dots representing the observations will tend to be arranged along a line or in an elongate oval or elliptical figure, a sort of path across the diagram. If the correlation is positive, this line or the axis of this figure will trend from lower left to upper right of the field of the diagram, and there will be few or no dots in the upper left and lower right corners. Conversely, if the correlation is negative, the trend will be from upper left to lower right, and the upper right and lower left corners will be relatively or absolutely free of dots. Clearly, if the trend is from lower left to upper right, it means that values of X farther to the right, and hence higher, are usually associated with values of Y farther above the base line, and hence higher; and this is the relationship called positive correlation. It can similarly be seen that the opposite trend, downward to the right, is a graphic

representation logically related to what has been called negative correlation. If there is no apparent trend,[1] the correlation is slight or zero. The coefficient of correlation may, in fact, be considered as a relative measure of the closeness with which the observations are clustered about such an oblique axis or arranged along such an oblique line, a relationship that will become more evident as this chapter is studied.

There is involved in such a trend another important factor that correlation alone does not give. It is desirable not only to know the intensity of the trend, or closeness of clustering along it, but also to know on an average how much larger (or, in negative correlation, smaller) Y becomes with a given increase in X or how much larger (or smaller) X becomes with a given increase in Y, factors reflected in the diagram by the slope of lines of trend. These relationships are given the somewhat unfortunate name "regressions."[2] The average amount that X changes for a unit change of Y is the regression of X on Y, and the amount that Y changes for a unit change of X is the regression of Y on X. It will be found that these two regressions are seldom equal to each other or to the correlation coefficient. The simplest sort of regression is one in which one variate, best treated as X, consists of a series of single and unequal values and the other, Y, has one definite value for each of these. Such a relationship seldom arises

[1] Or if it appears to be horizontal or vertical, an effect caused by using a unit of measurement of X larger relative to its range than is the unit of measurement of Y to its range, producing an apparently vertical trend, or a unit of Y larger relative to its range than is the unit of X to its range, producing an apparently horizontal trend. If the units are so adjusted that the ranges of X and Y are represented on the diagram by approximately equal linear distances, a low or zero correlation will show no apparent trend, and any evident linear arrangement will almost certainly reflect a significant degree of correlation.

[2] Unfortunate because it implies the act of regressing or stepping back toward some fixed value, which is not the essential concept in regression as the term is now used technically. Like so many concepts and procedures useful in zoology, that of regression was first developed in the study of human populations (by Sir Francis Galton, 1822–1907, a great British pioneer in the field of statistics). It was first used to study the inheritance of stature in men and was regarded as a ratio expressing how much a son had "regressed" toward the mean for sons relative to his father's divergence from the mean for fathers. The modern concept of regression includes this but is much broader.

except in time series, when X represents elapsed time and Y is any variate observed (on the same individual or as a mean from the same sample) at each of several different times, or values of X. In such a case the regression and the trend sought are those of Y on X—it is wished to determine how much, on an average, Y changes with the lapse of a unit period of time. The opposite regression, X on Y, here usually has no logical significance—it is rarely desired to know how much average time elapses with a unit change in Y.

Since in such a case there is only one observation of X or Y for each value of X, it is possible to draw a single line between successive points on the scatter diagram; and this gives a rough indication of trend. Even if the trend is fairly well defined, however, chance fluctuations in the value of Y will usually make this line irregular; and it would indicate average trend better and have a closer correspondence with regression if it could be smoothed out in some way. There are several approximate methods of smoothing adapted to such examples, the simplest and, of course, least accurate or useful of which is to sketch in a line freehand that seems to strike a middle path between the fluctuations, the peaks and valleys, of the line drawn from the actual observations. Somewhat better is the semiaverage method, which is to divide the observations into two equal parts,[1] obtain a mean value for X and for Y for each of them, put the two points thus determined on the scatter diagram, and draw a straight line through them. If the trend is significantly curved or if there are many sharp deflections, this method is poor; but otherwise it gives a fairly good approximation of the trend.

A method less affected by these disadvantages but with some of its own is that of moving averages. An average value of Y is taken corresponding to the first three values of X, then to the second to fourth, third to fifth, etc.; and these successive values are plotted with the corresponding middle values of X and are connected by a line on the diagram. Any number of values may be averaged, instead of three as just stated; but the number should be small relative to the whole number of steps available, and the calculation is much simpler if it is an odd number, so that three or five is usual. Although occasionally helpful, these

[1] If there is an odd number of pairs of observations, it does not matter which part includes the extra observation and hence has one more.

methods borrowed from economic statistics are applicable in relatively few zoological problems, and it is not necessary here to consider their possibilities in further detail. They are generally useful only for time series, and even with these more exact methods are preferable if possible.

CONCEPTS AND ROUGH APPROXIMATION OF REGRESSION

Most regressions in zoology differ from such simple time series in that two or more observations of either X or Y may, and generally do, have a single value or fall within a single class of the distribution. In such a case it is not possible to connect the dots of the scatter diagram by a single line, and the simpler rough indications of trend are not applicable. Even in such cases, however, a fair approximation of the trend can be obtained with relatively little work by calculating an average value of Y for each class of the distribution of X and plotting these with the corresponding class midpoints of X. A line connecting these points will give an approximation of the trend of the regression of Y on X. A similar approximation of that of X on Y can be obtained and this line also plotted. The work is most conveniently carried out in a table similar to a correlation table but giving total frequency, $\Sigma(X)$ or $\Sigma(Y)$, and mean for each row and column. The calculation is shown in Example 68.

The example demonstrates, and it is easy to grasp logically, that the two regression lines do not ordinarily coincide. They do so only if the observations are all exactly along a single straight line, a condition practically never found in practice. Otherwise the regression line of X on Y will be distinctly different from that of Y on X. In the first instance, X is called the dependent, and Y the independent variable; and in the second, Y is the dependent variable, and X independent. The regression of X on Y means always that we are considering Y as independent and are estimating the dependence of X on it.

In some cases, dependence and independence are logical concepts in the problem, and it is not a matter of free choice whether X or Y is taken as the dependent variable. This is true in most time series, in regressions of effect on cause, and some others. Thus in the study of growth, total length or any other growth function is necessarily taken as the dependent variable, and time is necessarily the independent variable. The regression of time on

the growth function usually has little or no direct bearing on the problem. Or again in the study of such relationships as that of animal abundance or size to environmental conditions, it is certain that if any causal relationship is found it will be the environmental

EXAMPLE 68.—CALCULATION OF ROUGH REGRESSION COORDINATES Length of M_1 and of M_2 in the fossil mammal *Phenacodus primaevus* (original data)

		X—length of M_1										
		9.8–10.2 10.0	10.3–10.7 10.5	10.8–11.2 11.0	11.3–11.7 11.5	11.8–12.2 12.0	12.3–12.7 12.5	12.8–13.2 13.0	13.3–13.7 13.5	f	$\Sigma(fx)$	M_X
Y—length of M_2	13.8-14.2 14.0	....	....	....	....	....	1	....	....	1	12.5	12.5
	13.3–13.7 13.5	....	....	....	....	....	2	2	1	5	64.5	12.9
	12.8–13.2 13.0	....	....	....	1	....	2	1	....	4	49.5	12.4
	12.3–12.7 12.5	....	....	....	2	2	1	1	1	7	86.0	12.3
	11.8–12.2 12.0	....	....	1	....	1	....	....	....	2	23.0	11.5
	11.3–11.7 11.5	....	....	1	....	....	....	....	....	1	11.0	11.0
	10.8–11.2 11.0	....	....	2	1	....	....	....	....	3	33.5	11.2
	10.3–10.7 10.5	....	2	....	....	....	....	....	....	2	21.0	10.5
	9.8–10.2 10.0	1	....	....	....	....	....	....	....	1	10.0	10.0
	f	1	2	4	4	3	6	4	2			
	$\Sigma(fY)$	10.0	21.0	45.5	49.0	37.0	79.5	52.5	26.0			
	M_Y	10.0	10.5	11.4	12.3	12.3	13.3	13.1	13.0			

EXAMPLE 68.—CALCULATION OF ROUGH REGRESSION COORDINATES (*Continued*)

Regression of Y on X		Regression of X on Y	
Ordinates (class midpoints of X)	Abscissas (means of arrays of Y)	Ordinates (means of arrays of X)	Abscissas (class midpoints of Y)
10.0	10.0	10.0	10.0
10.5	10.5	10.5	10.5
11.0	11.4	11.2	11.0
11.5	12.3	11.0	11.5
12.0	12.3	11.5	12.0
12.5	13.3	12.3	12.5
13.0	13.1	12.4	13.0
13.5	13.0	12.9	13.5
		12.5	14.0

The resulting rough regression lines are shown in Fig. 22 (see also calculated regression lines in Fig. 23).

In the tabulation it is convenient to give both conventional class limits and class midpoints on the scales. The values of one variate that correspond with those of a single class of the other are called an array. Thus the array of X, in this example, corresponding to the class 12.3–12.7 of Y, is as follows:

Class	Frequency
11.3–11.7	2
11.8–12.2	2
12.3–12.7	1
12.8–13.2	1
13.3–13.7	1

The marginal values calculated are the frequencies f, summations of values $\Sigma(fX)$ or $\Sigma(fY)$, and means M_X or M_Y, for each array of each variate.

factor that is cause and the zoological variable that is result. It is necessary to take the regression of the latter on the former. The regression of environment on a zoological variable is practically meaningless. In such cases the usual and best practice is to plot the independent variable as X.

In many problems, however, perhaps the majority, the identification of one variable as logically independent and the other as logically dependent is impossible beforehand, or independence and dependence may be mere conventions with no logical meaning. In the study of regression, one must nevertheless necessarily be

held as dependent and one as independent. In problems of this class, it is therefore advisable to study and record both regressions, to consider each character successively as dependent and

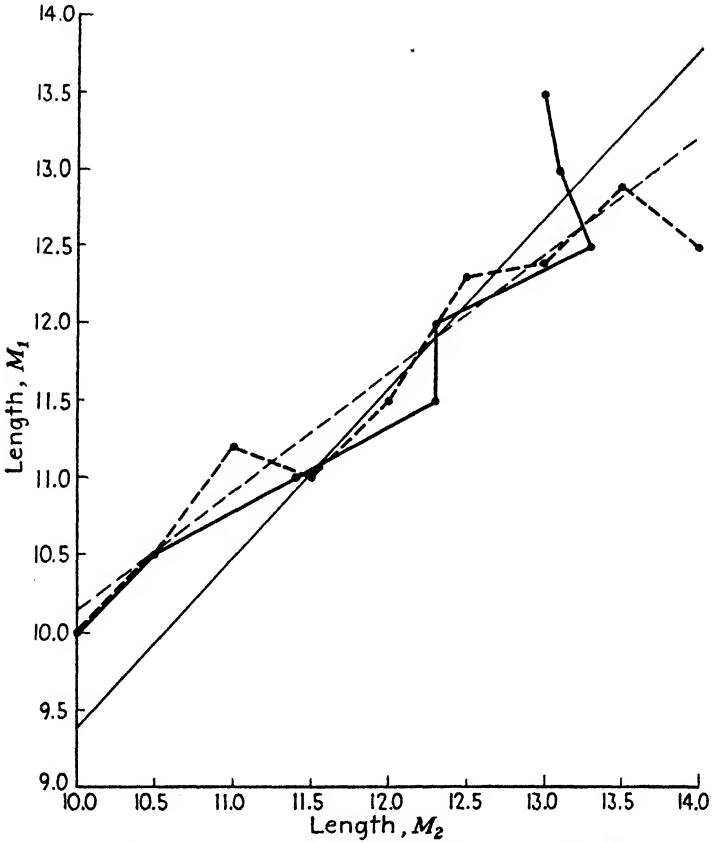


FIG. 22.—Rough approximation of regression lines. Lengths of the first and second lower molars of the fossil mammal *Phenacodus primaevus* (data of Example 68). The irregular lines are the rough approximations of the least-square fitted straight lines. Continuous lines represent regression of M_1 (or Y) on M_2 (or X) and the broken lines of M_2 on M_1. The approximation clearly is not very good, but better can hardly be expected from this easy but crude method in dealing with a rather small sample.

each as independent; it is the two regressions together and not simply one alone that are pertinent to the problem. The example just discussed, of the lengths of two adjacent teeth, is such a problem.

The functions and properties of the independent and of the dependent variables in regression are quite different. In any method of representing or of measuring regression, the independent variate is represented merely by the values it can take or by certain of these, such as class midpoints within its range, which were used in Example 68. The frequency distribution of this variate does not enter into the problem at all, and it may have any form without making any difference in the regression. The values of the independent variate may be arbitrarily selected, and the groups may be of any size or sort and need not even be equal to each other (although preferably so), so long as they are definitely determinable and errors in assigning observations to their proper group are few or none. The other factors of grouping will have little effect on the regression, but any errors may have a noteworthy effect; for the method by which they are normally neutralized, averaging, is not applied to this variate.

The frequency distribution of the dependent variate, on the contrary, has a very decided effect on the regression. The values of this variate are averaged, which not only makes the distribution a highly pertinent part of the problem but also tends to minimize the effects of errors, if they are truly random. Unbiased error in the dependent variate thus has little effect on the regression, although that in the independent variate generally has a strong effect. Arbitrary selection and grouping, unimportant in application to the independent variate, tend on the other hand to have an important and generally misleading effect on the regression if applied to the dependent variate. Of course if each variate is to be considered successively as independent and as dependent, the data must be in a form appropriate for either use.

REGRESSION COEFFICIENTS

Assigning an exact value to a regression involves determining a line of trend and then obtaining some numerical value that defines this line. The simplest case is when the trend is straight or nearly so, and the methods to be discussed first apply only to such regressions. Except in the extremely rare cases of perfect linearity of observations, several different lines may approximate the trend about equally well from different points of view or for different purposes, and the first point in seeking the most practical procedure is therefore to decide on a definition of the regression

line to be employed. The line that is generally adopted and that gives results most useful and most simply related to those of the best numerical procedures in general is a line such that the sum of the squares of the distances, or deviations, from it of the actual observations will be lower than for any other line. This is called the line of best fit, and the method is that of least squares.[1] What is needed is, then, an equation that can stand for any straight line and some method of calculating values to insert in this equation that will make the line that of best fit.

The following hypothetical series of values of X and Y has a single, straight regression line:

X	Y
1	2
2	4
3	6

A unit increase in X obviously means an increase of 2 in Y, and they also maintain a constant ratio $Y/X = 2$, so that a value of Y corresponding with any value of X can be calculated at once from the equation

$$Y = 2X$$

which is thus the equation for the straight line representing the regression of Y on X. Similarly

$$X = \frac{Y}{2}$$

is the equation of the regression of X on Y. General equations applicable to any case of this sort would be

$$Y = b_{YX}X$$
$$X = b_{XY}Y$$

in which b is any number, a constant or parameter to be calculated and inserted in the equation. b_{YX} is used to symbolize the value to be inserted in the equation for the regression of Y on X, and b_{XY} for that of X on Y. These quantities are called the regression coefficients.

In the simple regressions just given b_{YX} is 2 and b_{XY} is $\frac{1}{2}$. It will be noted that even if, as here, the two regression lines coincide, the two coefficients may be and commonly are different.

[1] Also used for curved regressions or for fitting frequency distributions.

Rectilinear regressions do not always take this form, however. A series like the following also has a single, straight regression line, yet the ratio $Y:X$ is not constant:

X	Y
1	3
2	5
3	7

When X increases by unity, Y increases by 2, as before, and so b_{YX} is again 2; yet the equation

$$Y = b_{YX}X$$

or

$$Y = 2X$$

is obviously incorrect. The equation must now become

$$Y = 1 + 2X$$

or

$$Y = a_Y + b_{YX}X$$

in which a is another constant, the value of Y when X is zero. In the example given or any just like it, the values of a_Y and of b_{YX} are easy to calculate, because all the values of Y lie exactly on its regression line and the lines for Y and for X coincide; but in practice this almost never occurs, and it is necessary to take the squares of their deviations from the line into account and to derive two equations from the basic equation so that both unknowns can be determined. The derivation of the formulas is not given here,[1] but it shows that the following two formulas, each in the general form of $Y = a_Y + b_{YX}X$, will give values of a_Y and of b_{YX} that satisfy the criterion of least squares.

$$\Sigma(Y) = Na_Y + b_{YX}\Sigma(X)$$
$$\Sigma(XY) = a_Y\Sigma(X) + b_{YX}\Sigma(X^2)$$

The values $\Sigma(X)$, $\Sigma(Y)$, $\Sigma(XY)$, $\Sigma(X^2)$, and N can all be calculated from the observed data; and the values of a_Y and b_{YX} can therefore be calculated from the two simultaneous equations. In practice the work can be considerably simplified because for purposes of obtaining b_{YX}, which simply measures slope of the lines, it does not matter what point on the X-scale is taken as 0

[1] It is given by Yule.

or what point on the Y-scale is taken as 0. This value can arbitrarily be assigned to the means of X and of Y and d_X and d_Y substituted for X and Y throughout. This will give the same value of b_{YX} as if X and Y were used. But it follows from the

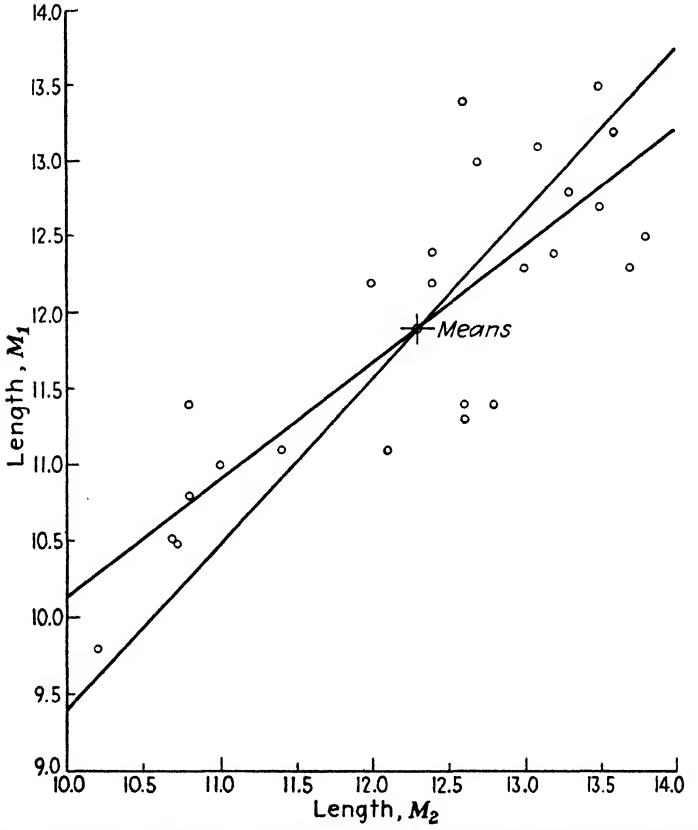


FIG. 23.—Scatter diagram and regression lines fitted by the method of least squares (same data as Fig. 22).

definition of the mean that $\Sigma(d_Y) = 0$ and $\Sigma(d_X) = 0$; so the first equation becomes

$$0 = Na_Y + 0$$

and since N is not 0, a_Y must be 0, and hence a does not enter into the regression equations when the means are taken as 0.[1] The

[1] In more mathematical terms, when the origin (in the graph) is taken at M_X, M_Y.

EXAMPLE 69.—CALCULATION OF COEFFICIENTS OF REGRESSION DIRECTLY FROM THE RAW OBSERVATIONAL DATA
(Same sample as in Example 61)

Tail length (X)	Total length (Y)	d_X	d_Y	$d_X d_Y$	d_X^2	d_Y^2
37	284	−68	−511	34,748	4,624	261,121
49	375	−56	−420	23,520	3,136	176,400
50	353	−55	−442	24,310	3,025	195,364
51	366	−54	−429	23,166	2,916	184,041
53	418	−52	−377	19,604	2,704	142,129
54	408	−51	−387	19,737	2,601	149,769
68	510	−37	−285	10,545	1,369	81,225
86	627	−19	−168	3,192	361	28,224
93	683	−12	−112	1,344	144	12,544
106	820	+ 1	+ 25	25	1	625
130	1,056	+25	+261	6,525	625	68,121
137	986	+32	+191	6,112	1,024	36,481
142	1,086	+37	+291	10,767	1,369	84,681
142	1,086	+37	+291	10,767	1,369	84,681
146	1,078	+41	+283	11,603	1,681	80,089
149	1,122	+44	+327	14,388	1,936	106,929
155	1,254	+50	+459	22,950	2,500	210,681
156	1,202	+51	+407	20,757	2,601	165,649
187	1,387	+82	+592	48,544	6,724	350,464
				312,604	40,710	2,419,218

$M_X = 105$

$M_Y = 795$ (calculated from the raw data—the different value in Example 61 is due to a nonsignificant error in calculating from coarser grouping.)

$$b_{YX} = \frac{\Sigma(d_X d_Y)}{\Sigma(d_X^2)} = \frac{312{,}604}{40{,}710} = 7.68$$

$$b_{XY} = \frac{\Sigma(d_X d_Y)}{\Sigma(d_Y^2)} = \frac{312{,}604}{2{,}419{,}218} = .129$$

The equations derived from the data of the example are

$$d_Y = 7.68 d_X$$
$$d_X = .129 d_Y$$

To put these in terms of X and Y, it is necessary to calculate a_Y and a_X, most simply done by these relationships:

$$a_Y = M_Y - b_{YX} M_X^*$$
$$a_X = M_X - b_{XY} M_Y$$

* Because a_Y is the value of Y when $X = 0$. When $X = 0$, $d_X = -M_X$,

EXAMPLE 69.—CALCULATION OF COEFFICIENTS OF REGRESSION DIRECTLY FROM THE RAW OBSERVATIONAL DATA.—(*Continued*)

With the data of the example these become

$$a_Y = 795 - 7.68 \times 105 = -11.4$$
$$a_X = 105 - .129 \times 795 = 2.4$$

Hence the regression equations are

$$Y = 7.68X - 11.4$$
$$X = .129Y + 2.4$$

From these the regression lines can readily be plotted. A straight line can be plotted from any two known different points on it. With the data in this form, the simplest way to plot the regression lines is to mark the point $X = M_X$, $Y = M_Y$, since both lines must pass through this, then to draw the regression line of Y through $X = 0$, $Y = a_Y$, and that of X through $X = a_X$, $Y = 0$. If a_Y and a_X are not calculated (and it is seldom necessary to give them in practice), the simplest method is to take an arbitrary integral difference of X from the mean, such as 10 or 100, and then to calculate the corresponding value of Y, as follows:

The coordinates sought are $(M_X + d_X)$ and $(M_Y + d_Y)$, M_X and M_Y being known and d_X arbitrarily selected. $d_Y = b_{YX}d_X$; so $M_Y + d_Y = M_Y + b_{YX}d_X$. For the data of this example, the following might be used:

$$d_X = 100 \qquad M_Y + d_Y = 795 + 7.68 \times 100 = 1{,}563$$
$$M_X + d_X = 205$$

So the regression line of Y on X passes through the point $X = 205$, $Y = 1{,}563$.

$$d_Y = 100 \qquad M_X + d_X = 105 + .129 \times 100 = 117.9$$
$$M_Y + d_Y = 895$$

So the regression line of X on Y passes through the point $X = 117.9$, $Y = 895$.

Both lines pass through $X = 105$, $Y = 795$; so both can now be plotted.

second equation becomes

$$\Sigma(d_Xd_Y) = 0 + b_{YX}\Sigma(d_X^2)$$

or

$$\Sigma(d_Xd_Y) = b_{YX}\Sigma(d_X^2)$$

and

$$b_{YX} = \frac{\Sigma(d_Xd_Y)}{\Sigma(d_X^2)}$$

---

so Y will be $M_Y - b_{YX}M_X$, which is therefore the value of a_Y sought. Similarly for a_X.

Similarly it can be shown that

$$b_{XY} = \frac{\Sigma(d_X d_Y)}{\Sigma(d_Y^2)}$$

$\Sigma(d_X d_Y)$, $\Sigma(d_X)$, and $\Sigma(d_Y^2)$ can be calculated directly from the observational data. Hence there is only one unknown in each of these equations, b_{YX} in one and b_{XY} in the other, and the values of these coefficients of regression can be readily and directly calculated from the data available. Such calculation is shown in Example 69. Figure 23 presents the data of Example 68 graphically.

The quantities $\Sigma(d_X d_Y)$, $\Sigma(d_X^2)$, and $\Sigma(d_Y^2)$ are also used in calculating the correlation coefficient and the last two also in calculating σ_X and σ_Y. With any given body of data it is obviously unnecessary to calculate these values separately in order to get the correlation coefficient, standard deviations, and regression coefficients. Instead of working out b_{YX} and b_{XY} separately, as was done in Example 69 in order to illustrate the process, it is thus practical to put the observations in a correlation table and to get all the desired constants or parameters from this one table, which can be done with very little more work than the calculation of any one of them. In fact, since b_{YX} and b_{XY} use only summations also entering into r, σ_X, and σ_Y, it is possible to obtain the latter first and then to calculate the regression coefficients from them, by these equations:

$$b_{YX} = r\frac{\sigma_Y}{\sigma_X}*$$

$$b_{XY} = r\frac{\sigma_X}{\sigma_Y}$$

* The equation is derived as follows:
It has been shown that

$$b_{YX} = \frac{\Sigma(d_X d_Y)}{\Sigma(d_X^2)}$$

whence, dividing both numerator and denominator by N,

$$b_{YX} = \frac{\Sigma(d_X d_Y)}{N\dfrac{\Sigma(d_X^2)}{N}}$$

(Footnote continued on opposite page.)

Using these formulas on the data of Example 61 (page 230), based on the same raw measurements as Example 69, gives results shown in Example 70.

EXAMPLE 70.—CALCULATION OF COEFFICIENTS OF REGRESSION FROM r, σ_X, AND σ_Y

(Same raw data as Examples 61 and 69)

From data of Example 61 we have

$\sigma_{1_X} = 2.244$

$\sigma_{1_Y} = 3.59$

$r = .993$

$i_X = 20$

$i_Y = 100$

$\sigma_X = i_X\sigma_{1_X} = 20 \times 2.244 = 44.9$

$\sigma_Y = i_Y\sigma_{1_Y} = 100 \times 3.59 = 359$

$$b_{YX} = r\frac{\sigma_Y}{\sigma_X} = .993 \times \frac{359}{44.9} = 7.94$$

$$b_{XY} = r\frac{\sigma_X}{\sigma_Y} = .993 \times \frac{44.9}{359} = .124$$

The results do not differ significantly from the more accurate values obtained in Example 69. The differences shown are principally caused by the coarser groupings of the observations and the greater rounding of various values in Example 61. Carried out with equal accuracy, the two methods give equal results. In this instance the less accurate but much more easily calculated values are accurate enough for practical purposes.

Since, in cases where it is worth while to calculate the regression coefficients, it is usually advisable also to calculate r, σ_Y, and σ_X, this is the usual method of obtaining these coefficients. It is to be emphasized, however, that this is only an arithmetical device

---

But

$$\frac{\Sigma(d_X^2)}{N} = \sigma_X^2$$

so

$$b_{YX} = \frac{\Sigma(d_X d_Y)}{N\sigma_X^2}$$

and multiplying numerator and denominator by σ_Y and then factoring

$$b_{YX} = \frac{\Sigma(d_X d_Y)\sigma_Y}{N\sigma_X^2\sigma_Y} = \frac{\Sigma(d_X d_Y)}{N\sigma_X\sigma_Y}\left(\frac{\sigma_Y}{\sigma_X}\right)$$

But it was shown in Chap. XII that $\Sigma(d_X d_Y)/N\sigma_X\sigma_Y = r$
so $b_{YX} = r(\sigma_Y/\sigma_X)$, and it can similarly be proved that $b_{XY} = r(\sigma_X/\sigma_Y)$.

adopted for convenience and made possible by the fact that some of the same values enter into the calculation of various of these coefficients and standard deviations. The formulas do not obviously suggest a clear and logical relationship of the various concepts involved and for this reason have been introduced only after the calculation of b_{YX} and b_{XY} from the original data was explained. The formulas may suggest that b_{YX} (and b_{XY}, understood to be analogous) is a sort of transformed r and so is, like z, in some way a measure of dispersion about the line of trend. This is not true. The arithmetical derivation of b_{YX} in this formula removes from its value any measure of dispersion. The dispersion (as involved in least squares) is used to fix the line taken as that of regression, but the coefficient of regression measures the slope of the line so determined and has nothing more directly to do with dispersion.

Conversely it might be supposed from this formula that r is a transformed regression coefficient and has something to do with some line and its slope, but this is equally untrue. It is a measure of dispersion only, with no direct relationship to slope.[1]

From these equations certain interesting and useful relationships between the two regression coefficients and that of correlation can be deduced:

One regression coefficient always has numerical value ± 1 or less (and a value of exactly ± 1 is of course very rare).[2]

[1] It seems easier to study correlation and regression in that order, and it is customary and usually easier to caculate r first and then get the regression coefficients with its aid; but logically and historically r is based indirectly on the regressions, not the latter on it. r measures the scatter or dispersion of the observations about the regression lines. This does not directly influence their slope, but it does influence their divergence from each other: when they coincide, $r = \pm 1$; and when they are at right angles, $r = 0$. This divergence can be expressed in terms of the two slopes. It is thus true that r is both arithmetically and logically a sort of average of the regression coefficients. It is, in fact, their geometric mean,

$$r = \sqrt{b_{YX} b_{XY}}.$$

[2] The maximum value of r is ± 1; so from their equations with r, σ_Y and σ_X, the maximum value of b_{YX} is $\pm(1/b_{XY})$. If b_{XY} is greater (arithmetically) than ± 1, then this expression and hence the maximum value of b_{YX} must be less than ± 1. If $b_{XY} = \pm 1$, then b_{YX} may also be ± 1 or less but cannot be greater.

The other regression coefficient may have any value from $+\infty$ to $-\infty$.[1]

Both regression coefficients and r have the same sign; all three are positive, or all three are negative.[2]

b_{YX}, b_{XY}, and r may all be equal, and in that case all are necessarily ± 1, a condition almost never obtaining with real data. If any two of these are equal, then the other is also equal, and all are ± 1.[3]

Unless all three are ± 1, one regression coefficient is always less than r and the other is always greater than r.[4]

If $r = \pm 1$, the two regression lines coincide, but the slope and hence the value of the corresponding regression coefficient of either of these may be anything and is independent of this value of r. When the lines coincide, r is always ± 1 (unless the regressions are horizontal or vertical), and the regression coefficients are always reciprocals.[5]

If $r = 0$, either the regression coefficients both are 0, or one is 0 and the other is infinity. The regression lines are thus either coincident and horizontal or vertical or at right angles, one horizontal and one vertical. If they are at right angles, they are necessarily horizontal and vertical (because otherwise they would not have the same sign), and r is necessarily 0. If they are coincident and both horizontal or both vertical, r may be 0 but may theoretically have any value between $+1$ and -1; this condition, however, practically never occurs in real data unless r is 0.[6]

[1] $b_{YX} = r(\sigma_Y/\sigma_X)$, and $b_{XY} = r(\sigma_X/\sigma_Y)$; so $b_{YX} = r^2/b_{XY}$ and $b_{XY} = r^2/b_{YX}$. As either regression coefficient approaches 0, the other approaches $\pm\infty$.

[2] Not subject to rigid mathematical proof. This is, in fact, a convention expressing the logic of the situation, not an inevitable numerical result. b_{YX} and b_{XY} necessarily have the same sign; but mathematically, however, r does not naturally have any sign and is in reality always $\pm$ (since, mathematically, our σ is $\pm$), but it is conventionally agreed to give it the sign of the regression coefficients.

[3] This follows in a simple and obvious manner from $b_{YX} = r^2/b_{XY}$ and $b_{YX} = r(\sigma_Y/\sigma_X)$, $b_{XY} = r(\sigma_X/\sigma_Y)$.

[4] $b_{YX} = r^2/b_{XY}$. If b_{XY} is greater than r, b_{YX} is less than r. If $b_{XY} = r$, both of these and b_{YX} are ± 1. If b_{XY} is less than r, b_{YX} is greater than r.

[5] When $r = \pm 1$ there is no scatter, all the observations are on one line, and this must represent both regression lines. Also since $b_{YX} = r^2/b_{XY}$ and $b_{XY} = r^2/b_{YX}$, if r is ± 1 it necessarily follows that $b_{YX} = 1/b_{XY}$ and $b_{XY} = 1/b_{YX}$; but either coefficient alone remains independent and may take any value whatever, subject only to the conventional (logical), not mathematical, requirement that all three coefficients have the same sign.

[6] $b_{YX} = r^2/b_{XY}$. If $r = 0$, the equation can be satisfied only by

$$\begin{array}{ll} b_{YX} = 0 & b_{XY} = 0 \\ b_{YX} = 0 & b_{XY} = \pm\infty \\ b_{YX} = \pm\infty & b_{XY} = 0 \end{array}$$

When $b_{YX} = 0$ and $b_{XY} = 0$, only $r = 0$ satisfies the equation. But when $b_{YX} = 0$ and $b_{XY} = \pm\infty$ or when $b_{YX} = \pm\infty$ and $b_{XY} = 0$, any

If r has any value between but not equal to 0 and ± 1, the two regression coefficients are not reciprocals, and both cannot be 0, they cannot be equal, and one may but both cannot be ± 1. The regression lines cannot coincide and cannot be at right angles but will be between these two conditions, and one of them can have any slope (the slope of the other being dependent on that of the one considered independent). The larger the value of r, the greater the divergence between the regression coefficients.

The ratio Y/X is constant (has the same value for any defined class of X) when $M_Y/M_X = b_{YX}$, and the ratio X/Y is constant when $M_X/M_Y = b_{XY}$. Both cannot be constant unless $r = \pm 1$, which practically never is true with real data.[1]

This last relationship is particularly important in zoological numerical theory. The extensive use of ratios in zoology is based on the assumption, sometimes expressed and almost always implicit, that ratios tend to be constant whatever the absolute value of the measurements entering into them. The approximate equivalence of M_Y/M_X to b_{YX} and of M_X/M_Y to b_{XY} will show whether this is really true or not. Often it is sufficiently so for practical purposes; but it may not be, and the assumption is not always justified. Ratios should be and will continue to be used because of their simplicity of calculation and obvious logical meaning; but the relationships that they seek to express

---

value of r satisfies the equation. The statements regarding the regression lines follow inevitably from these relationships of their slopes.

[1] $$Y = a_Y + b_{YX}X$$

so

$$\frac{Y}{X} = \frac{a_Y}{X} + b_{YX}$$

This value of Y/X contains the variable X and is itself necessarily variable unless the expression involving X is 0. This is always true and is only true when a_Y is 0, when $Y/X = b_{YX}$, which is a constant.
But

$$a_Y = M_Y - b_{YX}M_X$$

and Y/X is therefore constant when and only when

$$0 = M_Y - b_{YX}M_X$$

which is the same as

$$\frac{M_Y}{M_X} = b_{YX}$$

Similarly X/Y is constant when $M_X/M_Y = b_{XY}$.

If both Y/X and X/Y are constant, it follows from these equations that $b_{YX} = 1/b_{XY}$, and this is only true when r is ± 1.

are much more accurately expressed by correlation and regression, and the latter are much more reliable than ratios. For the data of Example 69,

$$\frac{M_Y}{M_X} = \frac{795}{105} = 7.57 \qquad b_{YX} = 7.68$$

$$\frac{M_X}{M_Y} = \frac{105}{795} = .132 \qquad b_{XY} = .129$$

Both ratios are very nearly constant. r has an extraordinarily high value in this example. If, as is much more common, r were lower, one or both of these ratios would be significantly inconstant.

The regression coefficients show the slope of the regression lines but do not show their absolute positions. These can be fixed only by giving a point through which each line passes. Mathematically the points $X = 0$, $Y = a_Y$, and $X = a_X$, $Y = 0$, are perhaps the most natural ones to choose, for a_Y and a_X are constants in the most general form of the regression equations; but these are seldom the best and are not the customary values to give in practice. In the first place they require the calculation of two numbers not needed for any other purpose; and in the second place they have no real and logical meaning when, as is true of so many zoological problems, X and Y do not, in fact, ever take the value 0. The values of M_X and M_Y are adequate and are the most natural and useful means of fixing the absolute position of the regressions. The single point determined by these is on the intersection of the lines and so fixes both of them; it thus has a special and important meaning; and M_X and M_Y are always needed in any case, and therefore their use involves no additional calculation.

Adequate study and representation of the (rectilinear) relationship between any two variables thus call for:

1. A measure of scatter about the lines of trend, or of the intensity of trend. r measures this and is useful in further calculation; and its transformed value z is the best measure.
2. Measures of the two directions of trend or slopes of the regression lines. b_{YX} and b_{XY} are the best measures of these slopes.
3. A measure of the absolute position of the regression lines. M_X and M_Y are the best measures for either line.

Thus in such problems the values of r, z, b_{YX} or b_{XY} or both, M_X, and M_Y should be calculated and given. Generally the

separate distributions of X and Y are also pertinent, and the whole family of values (each, except N, R, and r with its standard error) best given for the complete elucidation of such problems is:

For X	For Y
N_X	N_Y
R_X	R_Y
M_X	M_Y
σ_X	σ_Y
V_X	V_Y
b_{XY}	b_{YX}

For both

r_{XY} (and sometimes $r_{XY.z}$ or other partial correlation)

z

SIGNIFICANCE OF REGRESSION

One of the purposes of the regression coefficient is to predict what value of the dependent variate will correspond with a given value of the independent variate. Except in the usually only theoretical case of perfect correlation, such a prediction cannot be exact, for in practice the relationship is itself variable. That is, several somewhat different values of Y may be observed in combination with any one value of X. An array of values of Y, consisting of all those coupled with a single value or class of X, thus has its own distribution and its own standard deviation. Generally the mean of the array will be a point on the regression line or will, in other words, be the value of Y predicted by the regression equation. It would be possible to obtain a standard deviation for each array of Y, but this is not practical because it entails very laborious calculation and because the number of observations in a single array is often too small to give a useful standard deviation.

The usual practice is to obtain a combined standard deviation for all the deviations of Y from the regression line. The variance will be the sum of all the squares of all the deviations of Y from the corresponding values of Y, represented as Y_c, calculated from the regression equation, divided by N or, for small samples, by $(N - 2)$. The standard deviation sought is then the square root of this figure, or

$$S_Y = \sqrt{\frac{\Sigma[(Y - Y_c)^2]}{N - 2}}$$

using the formula in the form valid for both large and small samples. This special sort of standard deviation is called a standard error of estimate and symbolized as S_Y, Y being considered the dependent variate.

The calculation of the formula as written may be laborious because it calls for several series of calculated values, Y_c, $(Y - Y_c)$, $(Y - Y_c)^2$, and the summation of these, that are not required for any other purpose. It can be transformed into an arithmetically equivalent expression using only values already known and thus reducing the work to a relatively short arithmetical operation.

$$S_Y = \sigma_Y\sqrt{1 - r^2}\sqrt{\frac{N}{N-2}} \quad *$$

* $\Sigma[(Y - Y_c)^2] = \Sigma(Y^2) - NM_Y^2 - b_{YX}^2\Sigma(d_X^2)$.

Hence,

$$S_Y = \sqrt{\frac{\Sigma(Y^2) - NM_Y^2 - b_{YX}^2\Sigma(d_X^2)}{N-2}}$$

$$\Sigma(Y^2) = NM_Y^2 = \Sigma(d_Y^2)$$

$$b_{YX}^2 = \frac{[\Sigma(d_Xd_Y)]^2}{[\Sigma(d_X^2)]^2}$$

So,

$$S_Y = \sqrt{\frac{\Sigma(d_Y^2) - \dfrac{[\Sigma(d_Xd_Y)]^2}{\Sigma(d_X^2)}}{N-2}} = \sqrt{\frac{1 - \dfrac{[\Sigma(d_Xd_Y)]^2}{\Sigma(d_X^2)\Sigma(d_Y^2)}}{(N-2)\dfrac{1}{\Sigma(d_Y^2)}}}$$

$$r^2 = \frac{[\Sigma(d_Xd_Y)]^2}{\Sigma(d_X^2)\Sigma(d_Y^2)}$$

Hence,

$$S_Y = \sqrt{\left(\frac{(1 - r^2)}{N-2}\right)\left(\frac{\Sigma(d_Y^2)}{1}\right)}$$

$$\Sigma(d_Y^2) = N\sigma_Y^2$$

So

$$S_Y = \sigma_Y\sqrt{1 - r^2}\sqrt{\frac{N}{N-2}}$$

For large samples, N may be used in place of $(N - 2)$, so that the final expression is simply

$$S_Y = \sigma_Y\sqrt{1 - r^2}$$

On the other hand if, as is sometimes done, $(N - 1)$ is used for N in obtaining σ_Y, the formula becomes

$$S_Y = \sigma_Y\sqrt{1 - r^2}\sqrt{\frac{N-1}{N-2}}$$

Calculation and meaning of this value are shown in Example 71.

EXAMPLE 71.—CALCULATION OF STANDARD ERROR OF ESTIMATE FOR THE DATA OF EXAMPLE 70

$r = .993 \qquad N = 19$

$\sigma_Y = 359$

$$S_Y = \sigma_Y\sqrt{1 - r^2}\sqrt{\frac{N}{N-2}} = 359\sqrt{1 - .993^2}\sqrt{\frac{19}{17}} = 359 \times .117 \times 1.057 = 44.4$$

Considered as a standard deviation for any array of Y, this measures the probability that an observed value of Y will differ by a given amount from Y_c, or the probable range about Y_c (the calculated regression value) of Y. Thus for $X = 100$ the regression equation $Y_c = 7.68X - 11.4$ (see Example 60) gives $Y_c = 768 - 11.4 = 756.6$, which is the calculated value of Y, or the probable mean value of its array, when $X = 100$.

Taking $\pm 3\sigma$ as the probable range, the theoretical upper limit for Y is $Y_c + 3S_Y = 756.6 + 133.2 = 889.8$, and the theoretical lower limit is $Y_c - 3S_Y = 756.6 - 133.2 = 623.4$. This means that if X is 100, the data of this example prove, for all practical purposes, that Y will be between 623.4 and 889.8.

Prediction of the sort made possible by the standard error of estimate is sometimes highly important and useful. A principal use of the standard error of estimate is, however, to obtain an expression for the standard error of a regression coefficient. This coefficient has taken into account not merely the dispersion or deviation of the dependent variate but also that of the independent variate. Its variance is that of Y for any given value of X, divided by the sum of the deviations of X from its mean. The standard error $\sigma_{b_{YX}}$ is thus

$$\sigma_{b_{YX}} = \sqrt{\frac{S_Y^2}{\Sigma(d_X^2)}} = S_Y\sqrt{\frac{1}{\Sigma(d_X^2)}}$$

If S_Y is not needed, it is simpler to calculate this from previous data by one of the following equivalent equations:

$$\sigma_{b_{YX}} = \frac{\sigma_Y}{\sigma_X}\sqrt{\frac{1 - r^2}{N - 2}}$$

$$\sigma_{b_{YX}} = \sigma_Y\sqrt{\frac{1 - r^2}{\Sigma(d_X^2)}}\sqrt{\frac{N}{N - 2}}$$

$$\sigma_{b_{YX}} = \sqrt{\frac{S_Y^2}{\Sigma(d_X^2)}} \qquad S_Y^2 = \frac{\sigma_Y^2(1 - r^2)N}{N - 2}$$

(Footnote continued on opposite page.)

The calculation is shown in Example 72.

EXAMPLE 72.—CALCULATION OF STANDARD ERRORS OF REGRESSION COEFFICIENTS FOR THE DATA OF EXAMPLE 70

$N = 19 \qquad \sigma_X = 44.9 \qquad b_{YX} = 7.94$

$r = .993 \qquad \sigma_Y = 359 \qquad b_{XY} = .124$

$$\sigma_{b_{YX}} = \frac{\sigma_Y}{\sigma_X}\sqrt{\frac{1 - r^2}{N - 2}} = \frac{359}{44.9}\sqrt{\frac{1 - .993^2}{17}} = 8.00\sqrt{\frac{.0239}{17}} = 8 \times .0282 = .23$$

$$\sigma_{b_{XY}} = \frac{\sigma_X}{\sigma_Y}\sqrt{\frac{1 - r^2}{N - 2}} = \frac{44.9}{359} \times .0282 = .0035$$

Thence the two coefficients are:

$b_{YX} = 7.94 \pm .23$

$b_{XY} = .124 \pm .0035$

As with other standard errors, that of a regression coefficient can be used to test the significance of its deviation from 0 or any other given or hypothetical value. The deviation is divided by the standard error and the significance judged from a table of t (using the $N_1 + N_2$ column) or for large samples from a table of d/σ.

From these concepts and operations it is possible to proceed to an estimation of the significance of the difference between two regression coefficients by a method analogous to that for the difference between two means. The regression coefficients are

---

so,

$$\sigma_{b_{XY}} = \sqrt{\frac{\sigma_Y^2(1 - r^2)N}{\Sigma(d_X^2)(N - 2)}} \qquad \text{(the same as the second equation above)}$$

$$\Sigma(d_X^2) = N\sigma_X^2$$

so,

$$\sigma_{b_{YX}} = \sqrt{\frac{\sigma_Y^2}{\sigma_X^2}\left(\frac{1 - r^2}{N - 2}\right)} = \frac{\sigma_Y}{\sigma_X}\sqrt{\frac{1 - r^2}{N - 2}}$$

This last formula is valid for small or large samples and whether σ_X and σ_Y have been calculated as $\sqrt{\Sigma(d^2)/N}$ or as $\sqrt{\Sigma(d^2)/N - 1}$. For large samples, only, the approximation

$$\sigma_{b_{YX}} = \frac{\sigma_Y}{\sigma_X}\sqrt{\frac{1 - r^2}{N}}$$

is adequate. If σ_Y was taken as $\sqrt{\Sigma(d_Y^2)/(N - 1)}$, then the second form of the formula becomes

$$\sigma_{b_{YX}} = \sigma_Y\sqrt{\frac{1 - r^2}{\Sigma(d_X^2)}}\sqrt{\frac{N - 1}{N - 2}}$$

for small samples.

themselves complex measures; and this operation hence becomes very intricate, especially if the formula is made valid for small samples and for samples of unequal size, as it must be to have much utility. It is a more complicated procedure than almost any other given in this book; but it may be necessary in some research, and so it is given. The formula is

$$\sigma_{d_b} = \sqrt{\frac{N_1\sigma_{Y_1}^2(1 - r_1^2) + N_2\sigma_{Y_2}^2(1 - r_2^2)}{N_1 + N_2 - 4}\left(\frac{1}{N_1\sigma_{X_1}^2} + \frac{1}{N_2\sigma_{X_2}^2}\right)}$$

or,

$$\sigma_{d_b} = \sqrt{\frac{\Sigma(d_{Y_1}^2)(1 - r_1^2) + \Sigma(d_{Y_2}^2)(1 - r_2^2)}{N_1 + N_2 - 4}\left(\frac{1}{\Sigma(d_{X_1}^2)} + \frac{1}{\Sigma(d_{X_2}^2)}\right)}$$

The second form is somewhat more accurate in dealing with original data. The first uses only parameters usually published and so is helpful in comparing with previously published results. They are exactly equivalent, but rounding often makes the arithmetic results slightly different. The first assumes that σ is calculated as $\sqrt{\Sigma(d^2)/N}$.*

Significance is tested by taking the difference between the two regression coefficients, dividing by σ_{d_b}, and entering a table of t, using the value of $(N_1 + N_2 - 2)$ in the $(N_1 + N_2)$ column or of $(N_1 + N_2 - 3)$ in the N column.[1]

The calculation and use of this measure are shown in Example 73.

Cumbersome as this method is, it is the only one that permits reliable comparisons of regressions. Such comparisons are often essential, and the method therefore cannot be neglected in any competent review of numerical methods for zoology. Differences in regression may characterize not only sexes, as in the example given, but also age groups or taxonomic units. Differences in the regressions of time series and growth functions, as in relative rates of growth under different conditions or in different species,

* The meaning of the operation is brought out in more detail by Fisher. His procedure is very unlike ours but leads to essentially the same result. Ours brings together all the operations into one formula and expresses this in terms of values that may be assumed to be already available when this stage of study is reached.

[1] For large samples a table of d/σ may be used. For such samples also, $(N_1 + N_2)$ could be used in place of $(N_1 + N_2 - 4)$; but this is hardly any easier, and the formula as we give it applies to samples of all sizes.

EXAMPLE 73.—SIGNIFICANCE OF THE DIFFERENCE BETWEEN TWO REGRESSION COEFFICIENTS

Regression of tail length on total length in males and females of *Lampropeltis polyzona* (raw data from Blanchard 1921)

Sample 1 Males	Sample 2 Females
$N = 24$	$N = 19$
$\Sigma(d_Y^2) = 55{,}020$	$\Sigma(d_Y^2) = 40{,}710$
$\Sigma(d_X^2) = 2{,}275{,}088$	$\Sigma(d_X^2) = 2{,}419{,}218$
$\sigma_Y = 48$	$\sigma_Y = 46$
$\sigma_X = 308$	$\sigma_X = 357$
$r = .982$	$r = .993$
$b_{YX} = .153$	$b_{YX} = .124$

$$\sigma_{d_b} = \sqrt{\frac{N_1\sigma_{Y_1}^2(1 - r_1^2) + N_2\sigma_{Y_2}^2(1 - r_2^2)}{N_1 + N_2 - 4}\left(\frac{1}{N_1\sigma_{X_1}^2} + \frac{1}{N_2\sigma_{X_2}^2}\right)}$$

$$= \sqrt{\frac{24 \times 2{,}304 \times (1 - .982^2) + 19 \times 2{,}116 \times (1 - .993^2)}{24 + 19 - 4}} \times \sqrt{\left(\frac{1}{24 \times 94{,}864} + \frac{1}{19 \times 127{,}449}\right)}$$

$$= \sqrt{\frac{1{,}974.0672 + 562.8560}{39}\left(\frac{1}{2{,}276{,}736} + \frac{1}{2{,}421{,}531}\right)}$$

$$= \sqrt{65.0493 \times .00000085} = \sqrt{.000055291905} = .007436$$

It is, of course, unnecessary to work both formulas, but that will be done here to demonstrate that the results are substantially the same.

$$\sigma_{d_b} = \sqrt{\frac{\Sigma(d_{X_1}^2)(1 - r_1^2) + \Sigma(d_{Y_2}^2)(1 - r_2^2)}{N_1 + N_2 - 4}\left(\frac{1}{\Sigma(d_{X_1}^2)} + \frac{1}{\Sigma(d_{X_2}^2)}\right)}$$

$$= \sqrt{\frac{55{,}020 \times .0357 + 40{,}710 \times .0140}{39}\left(\frac{1}{2{,}275{,}088} + \frac{1}{2{,}419{,}218}\right)}$$

$$= \sqrt{64.9783 \times .00000085} = \sqrt{.000055231555}$$

$$= .007432$$

This is not significantly different from the result reached by the first formula. That it is different at all is caused by the different rounding in the two sets of data. The second formula goes back one step farther for its data and so is somewhat more likely to be reliable.

$$d_b = b_{YX_1} - b_{YX_2} = .153 - .124 = .029$$

$$t = \frac{d_b}{\sigma_{d_b}} = \frac{.029}{.007} = 4.14$$

$N_1 + N_2 - 2 = 41$. In the table of t, it is found that for any value of $(N_1 + N_2 - 2)$ greater than 5, a value of $t = 4.1$ gives a P smaller than .01. The difference is therefore significant. The regression of tail length on total length is significantly different in the males and females of this species.

are also highly pertinent in various zoological problems. These special uses of regressions are discussed more fully in Chap. XVI. The adequate measurement of regression is itself so complex that even long experience seldom permits reliable judgment of significance without a check by actual calculation.

CURVILINEAR REGRESSION

If a regression line is really curved, as is not infrequently true, the measures r and b may give misleading results. Since r is an inverse measure of relative dispersion about a straight line, it tends to give too low a value if the regression is not straight. The observations may be perfectly linear along a smooth curve and hence really be perfectly correlated, but they will obviously show deviations from any straight line that can be drawn and hence give a value of r less than 1. Similarly b shows only the slope of a straight line, and this may give a sort of average for the curve but will not be a good or reliable description of it. Moreover the curved line may have positive slope for part of its course and negative for part, so that the two parts may tend to neutralize each other when a straight line is used and give still more misleading values of r and b.

The equation for a straight line has only two constants, a and b as symbolized above, and is entirely determined by a single regression coefficient and a measure of absolute position, like a_Y or M_X and M_Y. A curved line, on the other hand, requires the determination of at least two coefficients for each regression and often requires many more. The calculation of these coefficients becomes very laborious as they increase in number, and the results also become less reliable. Even for curves that follow a relatively simple formula, the use of regression coefficients is seldom practical. For more complex or for irregular curves it may be impossible, and there is no general solution to the problem of fitting curved lines to regressions. The use of any regression coefficients for curvilinear data is so seldom practical in zoology that their calculation is not discussed here, but only some approximate methods of studying such regressions without the use of coefficients.[1] (One very important special type of curved regression is, however, discussed in detail in Chap. XVI.)

[1] The calculation of coefficients for a few types of curves has been worked out by some statisticians. Fisher discusses their calculation for curves of

If a regression is curvilinear, it may still be studied in one or more of the following ways:

1. The method of rough graphic approximation given on page 259 is as valid for curves as for straight lines and usually gives a fair visual conception of the relationship although it does not measure this.
2. If the curve is slight and its slope has the same sign throughout, the values of b_{YX} and b_{XY} are often sufficiently good approximations to serve all practical purposes.
3. With some data it is possible to divide the curve into segments, each of which is nearly rectilinear, or to select one nearly rectilinear segment that is in itself pertinent to the problem, and to apply rectilinear measures separately to these segments or to this one segment.
4. It may be possible to convert the data of the curve into some form still pertinent to the problem but rectilinear in regression or nearly so. The only relatively simple and important way to do this is in special cases where representing values of X, Y, or both by their logarithms makes the regression approximately rectilinear. The particular condition is that rate of change as a proportion remains constant although absolute change does not. For instance, the following is a (hypothetical) curvilinear regression that becomes rectilinear if the values of Y are replaced by their logarithms

X	Y
1	2
2	4
3	8
4	16
etc.	etc.

because Y is doubled each time that X increases by one unit.

THE CORRELATION RATIO

Pearson has devised a measure of correlation that is valid without regard to the form of the regression line and that thus presents a solution of the general problem of correlation for almost any relationship of two variables, no matter how irregular. This is called the correlation ratio and is usually symbolized by η (Greek lower-case eta). This quantity is the square root of one minus the ratio between the square of the mean standard deviation within the separate arrays and the square of the standard deviation of the whole distribution for one of the two variates.

---

the general form $Y = a + bX + cX^2 + dX^3$. . . . He also discusses regressions corresponding with multiple and partial correlations, which we likewise omit as being very complex and seldom of much practical use to zoologists. The curved regression $Y = bX^k$ is common in zoology, see Chap. XVI, where the calculations of b and k are explained.

The value is essentially the same whichever variate is used. Study of any correlation table will show that the mean standard deviation of the arrays will be about the same as that of the whole distribution if the correlation is low or about 0. Hence in such a case the ratio will be about 1, and 1 minus it will be about 0. On the other hand, the higher the correlation, the smaller will be the mean standard deviation of the arrays relative to that of the whole distribution, and hence the larger the value of 1 minus the ratio. If the correlation is perfect, the mean standard deviation and hence the ratio will be 0, and the value of η will be ± 1.

For rectilinear correlation the value of η is essentially the same as that of r. If the correlation (or regression) is curvilinear, however, r will underestimate the intensity of the correlation, but η will still measure it reliably. The difference between r and η is therefore also a measure of whether the regression is in fact significantly curvilinear.

η can be calculated much more simply than would appear from its derivation because it can be shown that the following simple formula gives the desired arithmetical result:

$$\eta = \frac{\sigma_{M_X}}{\sigma_X}$$

in which σ_{M_X} is the standard deviation of the means of the arrays of X.[1] This new datum is relatively easy to calculate, and on the whole η can be obtained as easily as r, if not more easily. As a first step it is well to calculate the rough regression as in Example

[1] By definition $\eta^2 = 1 - \sigma^2_{A_X}/\sigma^2_X$, if the mean standard deviation of the arrays of X be represented by σ_{A_X}.

From this

$$\eta^2 = \frac{\sigma^2_X - \sigma^2_{A_X}}{\sigma^2_X}$$

But it can be shown that $\sigma^2_{M_X} = \sigma^2_X - \sigma^2_{A_X}$.

Therefore,

$$\sigma^2 = \frac{\sigma^2_{M_X}}{\sigma^2_X}$$

and

$$\eta = \frac{\sigma_{M_X}}{\sigma_X}$$

68, for this gives at once the array means and frequencies needed to calculate σ_{M_X}. This further work is shown in Example 74.

EXAMPLE 74.—CALCULATION OF THE CORRELATION RATIO FOR THE DATA OF EXAMPLE 68

A. Using the distribution of X

Arrays (Y)	Mean of X for each array (M_{A_X})	$M_{A_X} - M_X$ (d)	d^2	f	fd^2
14.0	12.5	.5	.25	1	.25
13.5	12.9	.9	.81	5	4.05
13.0	12.4	.4	.16	4	.64
12.5	12.3	.3	.09	7	.63
12.0	11.5	− .5	.25	2	.50
11.5	11.0	−1.0	1.00	1	1.00
11.0	11.2	− .8	.64	3	1.92
10.5	10.5	−1.5	2.25	2	4.50
10.0	10.0	−2.0	4.00	1	4.00
				$N = 26$	$\Sigma(fd^2) = 17.49$

$\sigma_{M_X} = \sqrt{17.49/26} = .82$ $\quad \eta = \sigma_{M_X}/\sigma_X = .82/.95 = .863$

$\sigma_X = .95$

B. Using the distribution of Y

X	M_{A_Y}	d	d^2	f	fd^2
13.5	13.0	.7	.49	2	.98
13.0	13.1	.8	.64	4	2.56
12.5	13.3	1.0	1.00	6	6.00
12.0	12.3	0	0	3	
11.5	12.3	0	0	4	
11.0	11.4	− .9	.81	4	3.24
10.5	10.5	−1.8	3.24	2	6.48
10.0	10.0	−2.3	5.29	1	5.29
				26	24.55

$\sigma_{M_Y} = \sqrt{24.55/26} = .97$ $\quad \eta = .97/1.12 = .86$

$\sigma_Y = 1.12$

This result is practically the same as that based on X. r for these same data is .92.

Blakeman has suggested several tests of the significance of the difference between η and r. The following, modified after one of his formulas, is generally valid for samples of moderate or large size:

$$\frac{d}{\sigma} = \frac{\eta^2 - r^2}{2\sqrt{\frac{(\eta^2 - r^2)[(1 - \eta^2)^2 - (1 - r^2)^2 + 1]}{N}}}$$

Its use is shown in Example 75.

EXAMPLE 75.—COMPARISON OF r AND η, DATA OF EXAMPLE 74

$r = .82 \quad r^2 = .6724 \quad 1 - r^2 = .3276$

$\eta = .86 \quad \eta^2 = .7396 \quad 1 - \eta^2 = .2604$

$N = 26 \quad \eta^2 - r^2 = .0672$

$$\frac{d}{\sigma} = \frac{.0672}{2\sqrt{\frac{.0672(.2604^2 - .3276^2 + 1)}{26}}} = \frac{.0672}{2\sqrt{.00248}} = \frac{.0672}{.0996} = .67$$

This is not significant. Hence the data do not show this regression to be curvilinear.

Although η is relatively easy to calculate, perhaps easier than r, and applies to straight, curved, and many irregular regressions and not, like r, only to those rectilinear, it has so many disadvantages that it is not a very useful measure except as a makeshift for strongly curved regressions and as an approximate test of rectilinearity. One of the worst of these disadvantages is that it is not strictly consistent. Samples from a population that really has no correlation between two given variates will tend to give values of η definitely greater than 0.[1]

[1] The difference is sometimes significant, and the possibility makes an unreliable constant. Pearson has proposed a correction giving an approximate but somewhat more reliable value.

$$\eta' = \sqrt{\frac{\eta^2 - \frac{(k-3)}{N}}{1 - \frac{(k-3)}{N}}}$$

in which η' is the corrected correlation ratio and k is the number of arrays. In Example 75 this would be

$$\eta' = \sqrt{\frac{.7396 - \frac{5}{26}}{1 - \frac{5}{26}}} = \sqrt{.6776} = .82$$

This differs still less from r than does η, but the latter showed no significant curvilinearity and so was not misleading, nor is it generally unless near the limit of significance. The value 26 is a rather small N for the use of η in this way, although quite large enough to give good estimates of r and z.

CHAPTER XIV

ASSOCIATION

Correlation is possible only between variates with definitely ascertainable numerical values[1] and when each variate takes a considerable number of different values. The last two chapters have suggested how wide a variety of important problems may be treated by the methods of correlation and regression, but there remain many problems of a similar sort not subject to these methods.

Association is a relationship such that some category of observations tends to occur together with a category of some other given sort of observation more often than can be ascribed to chance alone. It reveals the existence of some kind of connection between two or more sorts of observations. Correlation is a special sort of association in which all the categories are numerical and each set of observations is divided into multiple categories. It is also necessary to have some method of detecting the presence of association when nonnumerical categories are included in the data or when a set of observations falls into only two categories or into a number too small to give reliable results by correlation methods. The general types of association not susceptible of correlation can be summarized and exemplified as follows:

1. Between a variate with multiple categories and a variate with few categories, *e.g.*, between depth of burrow and larger or smaller animals.
2. Between two variates with few categories, *e.g.*, between counts of dorsal and anal fin rays of fish, distribution of each only covering two or three classes.
3. Between a variate with multiple categories and an attribute, *e.g.*, between weight of fishes of a given species and geographic location.
4. Between a variate with few categories and an attribute, *e.g.*, between number of cuspules on a tooth and stratigraphic occurrence of a fossil mammal.
5. Between two attributes, *e.g.*, between sex and susceptibility to disease.

[1] Except rank correlation, which is possible without ascertaining any absolute values. Rank correlation is, however, only an approximation, and it is limited to variates the values of which, whether exactly known or not, do follow a definite and known serial order.

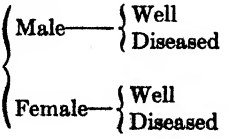
The same general methods can be applied to all these different problems and to any analogous to them. The variety of problems that can be treated by general methods of association is, indeed, much greater than of those that can be dealt with by correlation, and their importance is not less. It should be noted, also, that a variate for which only inadequate or inaccurate data are at hand can often be tested for association even though a correlation coefficient could not be based on it. It is necessary that the data suffice only for a reasonably good division into two or more categories. For instance, association may be tested by merely dividing a sample into smaller and larger observations by rough measurement or without actual measurement. Likewise a series of observations of a variate with multiple categories can be arbitrarily divided at any point into two parts and its association with some other variate or with an attribute tested, a procedure that may greatly simplify problems and reduce the work involved in studying them.

CONTINGENCY CLASSIFICATIONS

The simplest instances of association are those in which each set of observations has two categories. For the combination of the two sets, there are then four possible categories, and data arranged in this way are said to be placed in a fourfold or 2×2 classification. For instance, in studying the association of sex and susceptibility to disease, one set of observations has only the two categories, male and female, and the other only the two, well and diseased. The combination has the four categories:

Male and well.
Male and diseased.
Female and well.
Female and diseased.

This can also be arranged as a dichotomous classification:



In practice it is usually most convenient and comprehensible to arrange the data in what is called a contingency table, a set of

rectangular cells with the categories of one set of observations labeled at the top, those of the other at the left side, the corresponding frequencies entered in the cells, and the totals of rows and columns to the right and below the table. Such a table is shown in Example 75.

EXAMPLE 75.—CONTINGENCY TABLE OF GEOGRAPHIC LOCALITY AND NUMBER OF SERRATIONS ON LAST LOWER PREMOLAR IN CLOSELY SIMILAR MEMBERS OF THE FOSSIL MAMMALIAN GENUS *Ptilodus*

(Original data)

Number of serrations on P_4

Locality	Less than 14	More than 13	Totals
Montana	8	21	29
New Mexico	6	0	6
	14	21	35

General form of such 2×2 tables

First attribute or variate

Second attribute or variate	1st category	2d category	Totals
1st category	a	b	$a + b$
2d category	c	d	$c + d$
Totals	$a + c$	$b + d$	$a + b + c + d = N$

In order to test whether such data show any association, it is first necessary to establish what the frequencies would be if there were no association, *i.e.*, if the two sorts of things observed were completely independent. Obviously the numbers of observations in the two samples have nothing to do with association, nor have the total numbers of observations falling into any one category. The marginal totals, in other words, have no direct bearing on association, and in any specific problem they are to be taken as given and immutable. The next step is, then, to see what distribution of frequencies within the cells would give the marginal totals actually observed and would show complete theoretical agreement with the hypothesis that the two sets of observations do not influence each other.

These conditions would be fulfilled if the total $(a + b)$ were so divided that a and b had the same ratio as $(a + c)$ to $(b + d)$ and if all the other analogous ratios between cells were made equal to those between marginal totals. This would show complete independence; for then the cell frequencies would follow entirely from the totals, and the totals are not affected by any dependence between the two sets of observations. The conditions to be filled are, then, to keep the same totals and also to make

$$\frac{A}{B} = \frac{(a + c)}{(b + d)}$$
$$\frac{C}{D} = \frac{(a + c)}{(b + d)}$$
$$\frac{A}{C} = \frac{(a + b)}{(c + d)}$$
$$\frac{B}{D} = \frac{(a + b)}{(c + d)}$$

in which capital letters are used to indicate theoretical frequencies consistent with independence and lower-case letters to represent the observed totals. All the theoretical frequencies could be calculated by using these as simultaneous equations, but it is more convenient to use formulas by which each theoretical frequency can be calculated separately and directly from the marginal totals. The formulas best used are

$$A = \frac{(a + b)(a + c)}{N}$$
$$B = \frac{(a + b)(b + d)}{N}$$
$$C = \frac{(c + d)(a + c)}{N}$$
$$D = \frac{(c + d)(b + d)}{N}$$

The formulas can easily be remembered by the rule that the theoretical frequency of any cell is the total for the row in which it occurs multiplied by that for the column in which it occurs and divided by the total frequency.

A contingency table can be made with any number of cells, and the rule for finding the theoretical frequencies is the same whatever the size of the table. In practice there are seldom both

many rows and many columns in a table. 2 × 2 tables are the simplest and also the most common, 2 × 3, 2 × 4, and 2 × 5 tables are not uncommon, and 3 × 3 or 3 × 4 tables may also be useful occasionally. Larger tables are cumbersome and are seldom necessary. If a set of observations is on a variate and has many categories, it is usually better to lump these into two or, exceptionally, three. Attributes seldom have many categories and can often also be lumped if they are too finely subdivided for ease of handling.

The work of calculating theoretical frequencies in a simple 2 × 3 table is shown in Example 76.

EXAMPLE 76.—CONTINGENCY TABLE OF NUMBER OF SERRATIONS AND LENGTH OF LAST LOWER PREMOLARS OF THE FOSSIL MAMMAL *Ptilodus montanus* AND CALCULATION OF THEORETICAL FREQUENCIES ON THE HYPOTHESIS OF COMPLETE INDEPENDENCE

(Original data)

Length	Serrations			Totals
	13	14	15	
>7.9 mm.	4.1 / 0	9.8 / 14	1.0 / 1	15
<8.0 mm.	3.9 / 8	9.2 / 5	1.0 / 1	14
Totals	8	19	2	29

The theoretical frequencies are entered in the upper left corners of the cells. They are calculated as follows:

$$\frac{15 \times 8}{29} = 4.1 \qquad \frac{15 \times 19}{29} = 9.8$$

$$\frac{15 \times 2}{29} = 1.0 \qquad \frac{14 \times 8}{29} = 3.9$$

$$\frac{14 \times 19}{29} = 9.2 \qquad \frac{14 \times 2}{29} = 1.0$$

With a slide rule or machine these can be calculated easily and immediately from the table, and it is not necessary to put down any of the arithmetic.

TESTS OF SIGNIFICANCE

It is not to be expected that the frequencies actually observed in samples will correspond exactly with the theoretical frequencies or with their nearest integral values even if the variates and attributes studied are really completely independent in the

population. Chance necessarily plays a part, and the chance of complete agreement is always very small. What is needed, then, is to determine the probability that deviations from the theoretical frequencies equal to those observed could have arisen by chance in sampling a population in which the true proportions were those indicated by the theoretical frequencies. If such deviations could have arisen by chance, the data do not prove that the hypothesis of independence is inapplicable. If they could not have arisen by chance, then there is a significant disagreement with the hypothesis of independence, and it follows that there is significant association in the population.

The exact probability can be calculated,[1] but this process is too complicated to be satisfactory for practical use. It has, however, been shown that if for each cell the deviation of the observed from the theoretical frequency is squared and divided by the theoretical frequency, then the sum of these values for all the cells has a determinable distribution, depending also on the complexity and arrangement of the table, from which the desired probability can be determined.[2] This sum is generally called χ^2 (Greek chi). For a 2×2 table it is

$$\chi^2 = \frac{(A-a)^2}{A} + \frac{(B-b)^2}{B} + \frac{(C-c)^2}{C} + \frac{(D-d)^2}{D}$$

—and so on for larger tables, adding an analogous expression for each cell.

If the probability had to be calculated from χ^2 in each case, the work would be so difficult that the method would have no practical value, but this work has been done once for all, and corresponding values of χ^2 and of P have been tabled. The distribution of χ^2 depends on the number of degrees of freedom in the contingency table, so that this value must also be known in order to determine P from a table of χ^2. The number of degrees of freedom is the number of cells that could be filled in arbitrarily without disturbing the marginal totals. In a 2×2 table there is only one degree of freedom: if one cell is filled in, this determines the numbers that have to be placed in all the other cells in order

[1] Fisher shows the procedure.

[2] The proof of this statement and the calculation of the parameters of the distributions and of the probabilities are so complex and difficult mathematically that they are not given here.

to give the observed totals. In a 2×3 table there are two degrees of freedom. The degrees of freedom can be determined most simply by the rule

$$n = (r - 1)(c - 1)$$

where n = the number of degrees of freedom.
r = the number of rows in the contingency table.
c = the number of columns.
(Or r and c are the numbers of categories in the two sets.) Thus for a 3×5 table, n would be $(3 - 1)(5 - 1) = 8$. With this and the value of χ^2, Table XII shows the relationship of the value of P to the critical values concerned in estimating significance. If P is small, the data show significant association. The criteria of significance of P are the same as for tables of d/σ and P and of t and P, since P has the same logical meaning wherever it is used.[1]

If the theoretical frequencies are not desired for any reason except the calculation of χ^2, this can be simplified by following the formula

$$\chi^2 = \frac{N(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$$

for a 2×2 table. For larger tables it is generally simpler to calculate the separate values for each cell, as explained previously, except that in the special case of tables with two categories in one classification and two or more in the other there are special methods the most useful of which will be explained later.

The calculation of χ^2, both directly and by calculating the contribution to it of each cell, and its use in testing significance of association are shown in Example 77.

[1] Values for large numbers of degrees of freedom are seldom necessary for study of contingency tables; but χ^2 has other useful applications, as will be shown, and sometimes these involve many degrees of freedom. Several publications give more elaborate tables of χ^2, that in Fisher (on which our much simpler table is modeled) being perhaps the most convenient. For the higher values of n, generally above 30, the distribution of $\sqrt{2\chi^2}$ is approximately normal about $\sqrt{2n - 1}$ as a mean with 1 as its standard deviation. The probability can therefore be adequately estimated under these conditions by calculating

$$\sqrt{2\chi^2} - \sqrt{2n - 1}$$

and considering the result as a form of d/σ.

TABLE XII.—VALUES OF χ^2 CORRESPONDING TO GIVEN VALUES OF P AND GIVEN DEGREES OF FREEDOM

Degrees of freedom	P			
	.1	.05	.02	.01
1	2.7	3.8	5.4	6.6
2	4.6	6.0	7.8	9.2
3	6.3	7.8	9.8	11.3
4	7.8	9.5	11.7	13.3
5	9.2	11.1	13.4	15.1
6	10.6	12.6	15.0	16.8
7	12.0	14.1	16.6	18.5
8	13.4	15.5	18.2	20.1
9	14.7	16.9	19.7	21.7
10	16.0	18.3	21.2	23.2
12	18.5	21.0	24.1	26.2
14	21.1	23.7	26.9	29.1
16	23.5	26.3	29.6	32.0
18	26.0	28.9	32.3	34.8
20	28.4	31.4	35.0	37.6
25	34.4	37.7	41.6	44.3
30	40.3	43.8	48.0	50.9

There are two important things to remember about any such use of χ^2:

1. The method is valid only as applied to frequencies. If the data are expressed as proportions, percentages, rates, indices, or the like, the method cannot be used.

2. The result shows only the probability that deviations from the hypothesis of independence are caused by chance. If this probability is very small, the existence of dependence, that is, association, may be concluded. The method as here given and the value of χ^2 do not show: the degree of dependence; the nature of dependence; whether the dependence is direct, true, indirect, or spurious; what particular part of the table makes the result significant; or whether all the cells differ significantly from the hypothesis.

The test of association thus is a necessary preliminary to any reliable zoological conclusion, but it does not give a direct answer

EXAMPLE 77.—TEST OF ASSOCIATION BY χ^2. DORSAL AND ANAL RAYS OF THE FLYING-FISH, *Exocoetus obtusirostris*
(Data from Bruun 1935)

Anal rays	Dorsal rays		Totals
	12–13	14–15	
14	8.79 / 5	5.21 / 9	14
13	18.21 / 22	10.79 / 7	29
Totals	27	16	43

A. χ^2 from sums of contributions of cells:

$$\frac{(A-a)^2}{A} = \frac{3.79^2}{8.79} = 1.634$$
$$\frac{(B-b)^2}{B} = \frac{3.79^2}{5.21} = 2.757$$
$$\frac{(C-c)^2}{C} = \frac{3.79^2}{18.21} = .789$$
$$\frac{(D-d)^2}{D} = \frac{3.79^2}{10.79} = 1.331$$
$$\chi^2 = 1.634 + .789 + 2.757 + 1.331 = 6.511$$

B. χ^2 direct from raw data:

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} = \frac{43(5\times 7 - 9\times 22)^2}{14\times 29\times 27\times 16} = \frac{1{,}142{,}467}{175{,}392} = 6.514$$

The two results agree to the second decimal place, and the third is not significant. χ^2 is 6.51, and n is 1. From the table of χ^2, P is therefore very slightly greater than .01. Disagreement with the hypothesis of independence is almost surely significant, and association is established by the data beyond much doubt.

to many of the questions legitimately referred to the data. These may, however, usually be answered on a logical basis by reference to the contingency table, and for this essential purpose it is generally advisable or necessary to calculate the theoretical frequencies. Thus in Example 77 it is plain that the observa-

EXAMPLE 78.—RATIO METHOD OF CALCULATING χ^2 IN A $2 \times c$ TABLE. MORTALITY OF YOUNG AND OBSERVATION SUBSTATIONS OF THE TREE-SWALLOW, *Iridoprocne bicolor*
(Data from Low 1934)

Survival	Substations A	B	C	F	H	M	Totals
Hatched but not fledged (a')	27	11	11	3	5	10	$a = 67$
Fledged (b')	28	4	9	21	14	12	$b = 88$
Totals ($a' + b'$)	55	15	20	24	19	22	$a + b = N = 155$
Mortality ratio $\left(\frac{a'}{a' + b'}\right)$	.4909	.7333	.5500	.1250	.2632	.4545	$\frac{a}{a+b} = .4323$
$a'\left(\frac{a'}{a' + b'}\right)$	13.2543	8.0663	6.0500	.3750	1.3160	4.5450	$a\left(\frac{a}{a+b}\right) = 28.9641$

$$\sum\left[a'\left(\frac{a'}{a' + b'}\right)\right] = 33.6066 \qquad 1 - \left(\frac{a}{a+b}\right) = .5677$$

$$\chi^2 = \frac{\sum\left[a'\left(\frac{a'}{a'+b'}\right)\right] - a\left(\frac{a}{a+b}\right)}{\frac{a}{a+b}\left[1 - \left(\frac{a}{a+b}\right)\right]} = \frac{33.6066 - 28.9641}{.4323 \times .5677} = \frac{4.6425}{.24542} = 18.92$$

$n = (r - 1)(c - 1) = (2 - 1)(6 - 1) = 5$ $\quad P$ (from table) $< .01$

P is much less than .01, the deviation from independence is certainly significant, and mortality and substations are surely associated in some way. The data as here presented do not suggest why they are associated. They do, however, show that for some reason substations B and C had remarkably high mortality while F and H had remarkably low mortality, A and M being not far from average. The calculation in this example is easier by this method than would be calculating theoretical frequencies and contributions to χ^2 for each of 12 cells, and the quantities $a'/(a' + b')$ are ratios (mortality ratios) highly pertinent to the problem and desirable in themselves as well as for getting χ^2.

tions show excess frequencies in cells b and c and deficiency in cells a and d and hence that the nature of the association is that fewer dorsal rays are more often associated with fewer anal rays, more dorsal rays more often with more anal rays, fewer dorsal rays less often with more anal rays, and more dorsal rays less often with fewer anal rays than would be expected if the two variates were independent. In other words, the association plainly has the same nature as a positive correlation. The calculations also show that cell b contributes the most to χ^2, and hence departs most from the hypothesis of independence, and that cell c contributes the least and departs the least from the hypothesis.

$2 \times c$ CONTINGENCY TABLES

If one classification has two categories and the other has two or more, the theoretical frequencies and the cell contributions to χ^2 can be calculated as just explained and exemplified; but in some cases another method may be more convenient and equally or more enlightening. The data are set up in a contingency table with two rows and two or more columns. For each column the ratio of the number in the first row to the total is calculated, that is,

$$\frac{a'}{a' + b'}$$

if a' is taken as a value in the first row and b' as the corresponding value in the second row. The analogous ratio is calculated for the row totals, that is,

$$\frac{a}{a + b}$$

where a = the total for the first row.
b = the total for the second row.

The value of χ^2 is then

$$\chi^2 = \frac{\sum\left[a'\left(\frac{a'}{a' + b'}\right)\right] - a\left(\frac{a}{a + b}\right)}{\frac{a}{a + b}\left[1 - \left(\frac{a}{a + b}\right)\right]}$$

When there are many columns, this method may be quicker; and it is also advantageous when, as happens, the ratios $a'/(a' + b')$

and $a/(a + b)$ have a logical and pertinent connection with the subject of the investigation. The calculation is shown in Example 78.

SMALL SAMPLES

For the simplest calculation of χ^2 from a contingency table as hitherto considered, the samples should be relatively large. As an arbitrary but generally valid rule, the number of cells with small frequencies, say below 5, should be less than the number of degrees of freedom. Since a 2×2 table has only one degree of freedom, all four of its cells should have frequencies of 5 or greater. If a larger table has too many cells with low frequencies, it is often possible to eliminate these by combining categories and reducing the number of cells. Even if the smallest possible table, 2×2, still has cells with frequencies of 4 or less, it is possible to estimate P from χ^2 with considerable accuracy, but this requires an adjustment in the value of χ^2. The adjustment used is due to Yates and is also discussed in some detail by Fisher.

The reason for an adjustment is that the distribution of χ^2 is continuous, while that of the frequencies in a contingency table is necessarily discontinuous. The χ^2 distribution is approached as a limit by these discontinuous data, and if the frequencies are not unduly low the approach is sufficiently close to give a valid estimate of P from χ^2; but this is not reliable if the values of the table are determined largely by very low frequencies in it. Suppose, for example, that an observed frequency is 2 and that this is lower than the theoretical frequency. What the use of χ^2 measures, for that particular cell, is the probability that a frequency of 2, 1, or 0, that is, of 2 or less, would arise by chance if the variables were independent. The probability is measured in terms of the relative area of the curve for the distribution of χ^2 lying beyond the value of χ^2 resulting from a frequency of 2 in this cell. The next possible higher frequency is 3; but the curve for χ^2 is continuous, and there is an area between the points corresponding to frequencies 2 and 3. Moreover for the same reason there is no area corresponding exactly and only to the frequency 2 or 3; but each of these must be assumed to have a certain range, and the intervening area must be divided between them. The solution is to divide that area proportionately. This is done if the frequencies are treated as if they were midpoints of

groups of a continuous variate and the upper limit of 2 is taken as 2.5. The value of χ^2 wanted is that giving the probable occurrence not of 2 or less but of 2.5 or less. The exact probabilities have been calculated directly from the data for representative problems, and it has been shown that this adjustment does in fact lead to a significantly better estimate of P from the distribution of χ^2. It tends somewhat to underestimate significance but under the stated conditions is a closer estimate than is given by an unadjusted χ^2, and it is better to underestimate slightly than to overestimate markedly. With more evenly distributed and higher frequencies the unadjusted χ^2 does not markedly overestimate, and the adjusted χ^2 does markedly underestimate, significance.

EXAMPLE 79.—CALCULATION OF ADJUSTED χ^2 FROM A SMALL SAMPLE WEIGHTS AND DEPTHS OF BURROWS OF THE GROUND SQUIRREL, *Citellus columbianus columbianus*
(Data from Shaw 1926)
Raw data
Depth of burrow

Weight of animal	<29.5	>29.5	Totals
>452	1	10	11
<452	8	3	11
Totals	9	13	22

Same data, adjusted for the low frequencies

	<29.5	>29.5	Totals
>452	1.5	9.5	11
<452	7.5	3.5	11
Totals	9	13	22

$$\chi^2 = \frac{(1.5 \times 3.5 - 9.5 \times 7.5)^2 \times 22}{11 \times 11 \times 9 \times 13} = 6.77$$

P is slightly less than .01, and the association may be taken as significant. Larger ground squirrels dig deeper burrows. χ^2 calculated without adjustment is 9.6, leading to a great overestimate of significance.

This adjustment may be applied to any 2 × 2 table that needs it by increasing by .5 each of the observed frequencies that is lower than the corresponding theoretical frequency and decreasing by .5 each that is higher than the theoretical frequency, then calculating χ^2 in the ordinary way from these adjusted frequencies. This is shown in Example 79.

The adjusted value of χ^2 can also and more simply be calculated from the original (unadjusted) data by the usual formula.

$$\chi^2 = \frac{N(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$$

by adjusting the quantity $(ad - bc)$, the only one in this formula affected by the adjustment. If this quantity is positive, $N/2$ should be subtracted, and if it is negative, $N/2$ should be added to it.[1] The adjustment and this method of applying it are valid even if there are zero frequencies in the table. For the data of Example 79, the simplified calculation of adjusted χ^2 is

$$\chi^2 = \frac{22(1 \times 3 - 10 \times 8 + 22\tfrac{1}{2})^2}{11 \times 11 \times 9 \times 13} = 6.77$$

giving exactly the same value as before without the trouble of calculating theoretical frequencies and adjusted observed frequencies.

[1] If ad is greater than bc, making $(ad - bc)$ positive, the adjustment of the original data would make this expression

$$(a - .5)(d - .5) - (b + .5)(c + .5)$$

Multiplying these, it becomes

$$(ad - .5a - .5d + .25) - (bc + .5b + .5c + .25)$$

which may be combined as

$$ad - bc - .5(a + b + c + d)$$

and since $(a + b + c + d) = N$, this is

$$ad - bc - \frac{N}{2}$$

Similarly it can be shown that the adjustment gives

$$ad - bc + \frac{N}{2}$$

if bc is greater than ad.

EXAMPLE 80.—ASSOCIATION OF THE DATA OF EXAMPLE 79 WITH THE ADDED FACTOR OF SEX

A. Sex constant, association of weight and depth of burrow

All males:

Depth of burrow

Weight	<32.7	>32.7	Totals
>494	1	8	9
<494	3	4	7
Totals	4	12	16

Adjusted $\chi^2 = \frac{16(1 \times 4 - 8 \times 3 + 8)^2}{9 \times 7 \times 4 \times 12} = .76$ Not significant.

All females:

Depth of burrow

Weight	<24	>24	Totals
>325	2	2	4
<325	2	0	2
Totals	4	2	6

Adjusted $\chi^2 = \frac{6(2 \times 0 - 2 \times 2 + 3)^2}{4 \times 2 \times 2 \times 4} = .09$ Not significant.

B. Association of sex and depth of burrow

Depth of burrow

Sex	<21	>21	Totals
Male	2	14	16
Female	4	2	6
Totals	6	16	22

Adjusted $\chi^2 = \frac{22(2 \times 2 - 14 \times 4 + 11)^2}{16 \times 6 \times 6 \times 16} = 4.01$

$P < .05 > .02$ Possibly significant under these circumstances probably so.

EXAMPLE 80.—ASSOCIATION OF THE DATA OF EXAMPLE 79 WITH THE ADDED FACTOR OF SEX.—(*Continued*)

C. Association of sex and weight

Sex	Weight <375	Weight >375	Totals
Male	2	14	16
Female	5	1	6
Totals	7	15	22

Adjusted $\chi^2 = \frac{22(2 \times 1 - 14 \times 5 + 11)^2}{16 \times 6 \times 7 \times 15} = 7.091$ Surely significant.

Example 79 showed that animal weight and depth of burrow were associated, but the sample was heterogeneous and included both males and females. Example 80 shows that weight and depth of burrow are not, by these data, demonstrated to be associated for males only or for females only, that sex is surely associated with weight, and that sex is probably associated with depth of burrow. The general conclusion is that the males are heavier than the females and tend to dig deeper burrows. The data do not show whether they dig deeper burrows because they are heavier (or larger, or older) or for some other reason connected with their being males.[1]

MULTIPLE AND SPURIOUS ASSOCIATION

Association is closely analogous with correlation; and what has been said about the logic and pitfalls of correlation, especially as regards cause and effect, spurious correlation, and the like, is also true of association. Where three or more variables are or may be involved in a given problem of association, the best approach is to test the association of all these, two by two, and to base the zoological conclusion on the significance of these results. Two examples (80 and 81) and their consideration will make this clearer than a longer abstract discussion.

HYPOTHESES OF DEPENDENCE

It has been emphasized that the χ^2 technique does not directly test association but that it tests agreement with a hypothesis by showing the probability that observed deviations from that hypothesis could have arisen by chance. In testing association

[1] The author who published the data concluded that older animals dig deeper burrows. The data do not really prove or even suggest this conclusion.

the hypothesis set up is that of independence, and a significant deviation from the hypothesis is taken to prove dependence or association. It is equally valid and is often useful to set up a hypothesis of dependence on some particular law or set of values

EXAMPLE 81.—DATA OF EXAMPLE 78 RELATED TO TYPE OF NESTING BOXES PROVIDED AT THE VARIOUS SUBSTATIONS

	Type of box		
	Type 31 (substations B, C, and M)	Type 32 (substations A, F, H)	Totals
Hatched but not fledged.....	32	35	67
Fledged.....	25	63	88
Totals.....	57	98	155

$$\chi^2 = \frac{(32 \times 63 - 35 \times 25)^2 \times 155}{67 \times 88 \times 57 \times 98} = 6.02$$

$P < .02 > .01$ Probably significant.

Example 78 showed a sure association between substations and mortality or survival of the young birds, without giving any suggestion as to the cause of the association. The original data show that the substations differed not only in locality but also in the type of nesting boxes provided. It was therefore advisable to test the possibility that the mortality was associated with the type of nesting box, and Example 81 shows that it very probably was. Evidently the boxes of type 32 had a significantly better record of survival of young birds. It is not clear from these data whether there was also some other reason why some substations had higher mortality than others or whether this was entirely caused by the type of boxes used; but the boxes almost certainly had something to do with it.[1]

and to test by the use of χ^2 whether the observed values are in accord with the hypothesis. What can be shown is either that the hypothesis is wrong or that the data are harmonious with it; it is impossible to prove in any rigid and mathematical way that the hypothesis is right.

The principal use of this method is to test agreement between an observed frequency distribution and any theoretical distribu-

[1] The original author does not explicitly point this out but does mention that type 32 was a new (presumably improved) box used to replace type 31. The association technique incidentally provides a conclusive means of determining what boxes are best. Their acceptance by the birds should also be tested—the data given in this example do not permit this.

tion, such as a normal distribution, binomial series, or Poisson series. Each item contributes to χ^2 an amount determined by squaring the difference between observed and theoretical frequencies and dividing by the theoretical frequency. If the frequencies at either end are small and seem to make an undue contribution to χ^2, it is well to lump them in groups of so much "and over" or "and under," a procedure invalid in correlation but valid and often advisable in association and χ^2 technique. In such calculations the degrees of freedom are one fewer than the classes. The use of the test is shown in Example 82.

EXAMPLE 82.—χ^2 TEST OF OBSERVED DISTRIBUTION AND APPROXIMATELY FITTED NORMAL DISTRIBUTION
(Data of Example 46, page 142)

Class	Observed frequencies (f)	Theoretical frequencies (ay_0)	d ($f - ay_0$)	d^2	d^2/ay_0
55 and under.	4	3.8	.2	.04	.011
56–57.....	11	9.8	1.2	1.44	.147
58–59.....	18	17.9	.1	.01	.001
60–61.....	21	22.9	1.9	3.61	.158
62–63.....	20	17.9	2.1	4.41	.246
64–65.....	9	9.8	.8	.64	.065
66 and over...	3	3.8	.8	.64	.168
					$\chi^2 = \Sigma(d^2/ay_0) = .796$

Degrees of freedom: 6

P is very large, and the differences are not significant, that is, the observed distribution could very well have been derived from a normally distributed population, and the calculated normal distribution is an excellent fit to the data of observation. It does not, however, follow that a better theoretical fit is impossible.

Any hypothetical series of frequencies can be drawn up on the assumption that it represents the true proportions in a population, and its agreement or disagreement with an observed series of frequencies can be tested. This is equivalent to testing whether the hypothetical distribution could in fact represent the true proportions in the population from which the observed sample was drawn. An important use, for instance, is to test

whether variations occur with frequencies consistent with Mendelian ratios.

Another very important use of this technique is as a test of homogeneity. In this case χ^2 is calculated not from an observed and a hypothetical distribution but from two observed distributions. The hypothesis is that the two were drawn from the same population (the true proportions in which may be assumed to be represented by either sample), and the test shows whether this is possible or not. As a test for homogeneity it is often preferable to those hitherto given, since it takes all the items of the distributions into account and makes no assumption regarding the population distribution (for instance, whether or not it is normal, as other tests of homogeneity often assume) except that it is approximated by one or the other of the samples.

The method of Example 82 is applicable only when the two distributions have the same total frequency and hence cannot be used for two observed distributions with different totals. The methods of association, either by theoretical frequencies or by ratios in a $2 \times c$ table, are therefore to be used for this purpose. Two typical instances are given in Example 83.

EXAMPLE 83.—USE OF χ^2 TO TEST HOMOGENEITY OF SAMPLES

A. Two samples of length of P_4, both identified as the fossil mammal *Ptilodus montanus* and from approximately the same horizon and locality but collected at different times by different institutions (original data).

Calculation of χ^2 by ratio method

Classes	Sample 1	Sample 2	Totals	Ratios	Sample 1 × ratios
7.5 and below.....	5	3	8	.6250	3.1250
7.6–7.8.....	9	6	15	.6000	5.4000
7.9–8.1.....	7	9	16	.4375	3.0625
8.2–8.4.....	8	14	22	.3636	2.9088
8.5 and above.....	7	7	14	.5000	3.5000
Totals.....	36	39	75	.4800	17.9963

Total sample 1 × total of ratios = 17.2800

$$\chi^2 = \frac{17.9963 - 17.2800}{.48 \times .52} = 2.87$$

$n = 4$

The difference is not significant, and the samples are consistent with the hypothesis that both are from the same population.

EXAMPLE 83.—USE OF χ^2 TO TEST HOMOGENITY OF SAMPLES.—(*Continued*)

B. Three samples from different localities, hair counts of winter pelage of the mouse *Peromyscus maniculatus rubidus* (data modified from Heustis 1931). Calculation by contingency table. (The data have been artificially simplified for clearer exemplification of method.)

Hair types	Samples			Totals
	Coos Bay	Port Orford	Eugene	
Black overhairs	22.3 / 24	19.1 / 16	28.6 / 30	70
Large banded hairs	12.4 / 9	10.6 / 12	16.0 / 18	39
Fur hairs	103.3 / 105	88.3 / 90	132.4 / 129	324
Totals	138	118	177	433

Contributions to χ^2:

$$\frac{(24 - 22.3)^2}{22.3} = .130$$

$$\frac{(19.1 - 16)^2}{19.1} = .503$$

$$\frac{(30 - 28.6)^2}{28.6} = .069$$

$$\frac{(12.4 - 9)^2}{12.4} = .932$$

$$\frac{(12 - 10.6)^2}{10.6} = .185$$

$$\frac{(18 - 15.9)^2}{15.9} = .250$$

$$\frac{(105 - 103)^2}{103} = .028$$

$$\frac{(90 - 88.2)^2}{88.2} = .033$$

$$\frac{(132 - 129)^2}{132} = .087$$

Total.....2.217

$n = 4$

$\chi^2 = 2.22$

$P > .5$

The difference is not significant, and the samples are consistent with the hypothesis that all are from the same population.

CHAPTER XV

GRAPHIC METHODS

Almost any numerical data in zoology and many that are not numerical can be represented graphically. A good graph spreads before the eye in a unified and comprehensive way a picture of facts and of relationships that cannot be so clearly grasped, if at all, from any verbal or strictly numerical representation. Sometimes a graph may in itself permit an adequate solution of the problems arising from the data, but more often it does not supplant calculation and direct numerical treatment. Usually the two supplement each other, the graphic method giving an immediate and suggestive résumé of what the written methods reduce to exact values, prove, and interpret.

The most important graphic methods are those concerned with frequency distributions (see Chaps. III to VII) and with correlation and regression (Chaps. XII to XIII). These and a few other graphic methods (*e.g.*, for the comparison of single specimens, Chap. XI) have already been adequately explained and exemplified. They need no further discussion, but there are many other sorts of graphs. The possibilities are, indeed, almost unlimited; all cannot be considered, but only the more important, with examples of various sorts and some suggestions as to general principles. With so much basic knowledge and some ingenuity, special graphic methods can readily be devised for any particular problem.

TYPES OF DIAGRAMS

Most useful diagrams, although not all, belong to one of the following types:

1. *Point Diagrams.*—Here belong the point method of representing frequency distributions (page 56), scatter diagrams of correlation (page 231), and the like.

2. *Line Diagrams.*—These include frequency polygons (page 57), regression and other trend lines (page 261), theoretical curves like the normal curve (page 74), and any other diagram that relates the original discrete observations to some form of continuous line.

3. *Bar Diagrams.*—In these a line or a rectangle represents each category or variate, and its length is proportionate to the corresponding value. Histograms (page 57) are a special type of bar diagram. Others are mentioned below.

4. *Area Diagrams.*—A figure of standard shape is subdivided into areas proportionate to values to be represented. The most useful type, the pie diagram, uses a circle and subdivides it into sectors by drawing radii.

5. *Three-dimensional Diagrams,* such as correlation surfaces, discussed below.

6. *Pictorial Diagrams.*—This large and miscellaneous group includes maps, diagrammatic pedigrees and phylogenies, graphic representations of numerical properties by actual pictures of animals used in various ways, and many other methods. They include some concepts and methods not primarily numerical; but many are analogous to numerical methods, and some could be reduced to numbers if desired.

Most of these various types of diagrams involve a system of coordinates, a mesh, net, or field of some sort such that position, linear distance, angle, slope, or the like has a definite numerical value in the diagram. Most important are rectangular coordinates (Fig. 24A, B, C). All the diagrams given on preceding pages of this book have rectangular coordinates, and their general nature is already sufficiently clear. Arithmetic (Fig. 24A) coordinates, those usually employed, represent any two equal differences in values along one axis by equal linear distances. Usually the scales are the same for the X- and Y-axes if these represent analogous variates, and this is preferable when practical. Sometimes, however, the ranges of X and Y are so greatly unequal that an awkward or impossibly large diagram can only be avoided by giving the larger variate a smaller scale. When the two variables are not analogous, for instance, when one is a value of a variate and one a frequency as in histograms, there is no necessary relationship between the scales, and they are adjusted in each case to produce a convenient and enlightening result.

Rectangular coordinates may also be logarithmic or semilogarithmic (Fig. 24B and C). On a logarithmic scale, equal linear distances represent not equal absolute differences but equal ratios. Thus on arithmetic coordinates the distance between points scaled as 10 and 100 is ten times that between 1 and 10, but on logarithmic coordinates the distances are equal because $^{10}/_{100} = {}^{1}/_{10}$. Logarithmic coordinates are logarithmic on both X- and Y-axes, while semilogarithmic[1] coordinates are arithmetic

[1] Also called "arithlog."

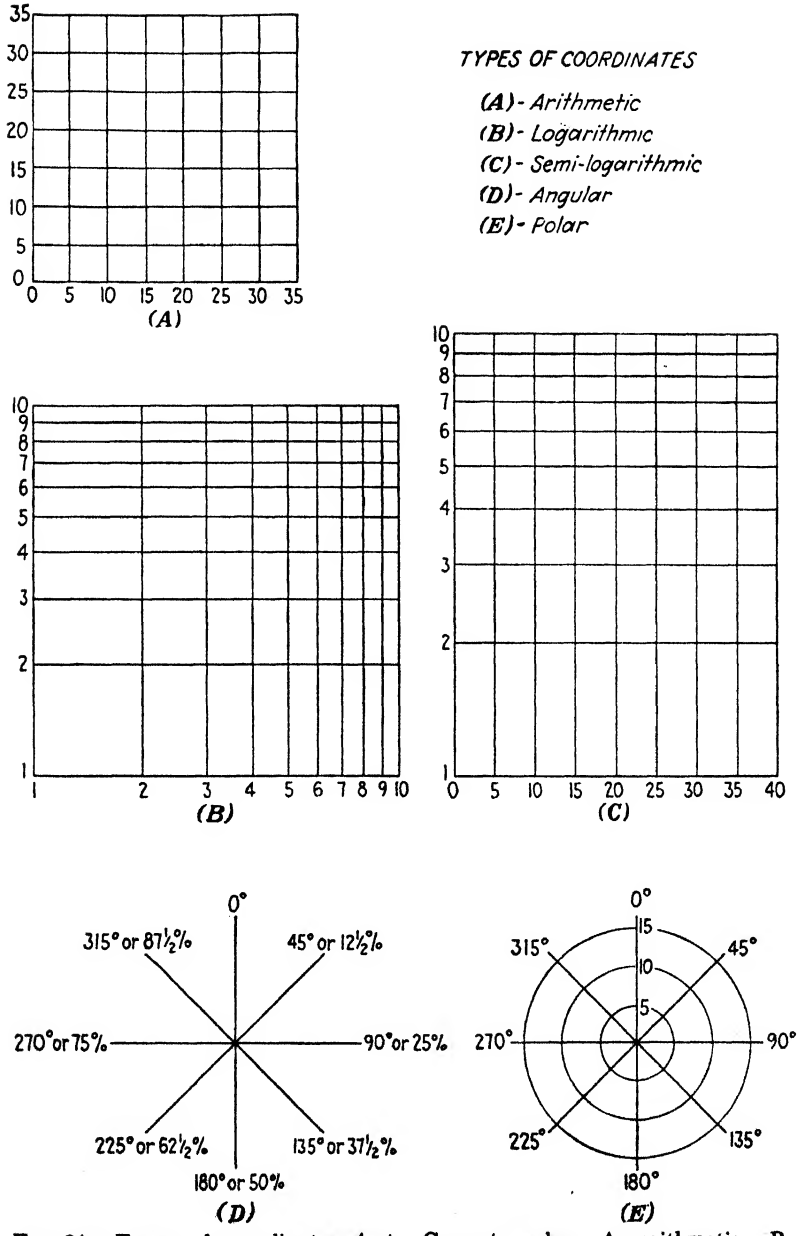


FIG. 24.—Types of coordinates A to C, rectangular. A, arithmetic. B, logarithmic. C, semilogarithmic or arithlog. D, angular. E, polar.

on the X- and logarithmic on the Y-axis. Such coordinates are often used for plotting rates, ratios, geometric progressions, and the like, because on them a geometric progression is plotted as a straight line, equal lines correspond with equal ratios, and equal slopes represent equal rates of change. Semilogarithmic coordinates are most commonly used for time series, plotting time arithmetically on the X-axis. They have the added advantage that if two comparable variates are being plotted in the same field and one is much larger than the other, the smaller is exaggerated, and the larger minimized; the comparison is thus clearer and more convenient than on arithmetic coordinates. Paper ruled logarithmically and semilogarithmically can be purchased. If such paper is not readily available, the same result can be obtained (but more laboriously) on arithmetic coordinates by plotting the logarithms of the values appearing in the data. It should be noted that there is no 0 on a logarithmic scale: its base line is 1, since the logarithm of 1 is 0. (The logarithm of 0 is $-\infty$ which of course cannot appear on the graph.)

Angular and polar coordinates (Figs. 24 D and E and 25) are also occasionally used, but are relatively unimportant. Angular coordinates represent a value by the angle between two lines diverging from a given point. There is thus only one scale, and values must almost necessarily be percentages or other fractions of a total, facts that make angular coordinates of very limited value except in the special form of pie diagrams. Polar coordinates are angular coordinates with another scale added—distance from the central point. They are of considerable value when one of the variables is in fact an angle or falls readily into circular form. For instance, they could be used to plot frequencies of cranial angles, and they are used to plot periodic annual or seasonal phenomena, dividing the angular scale into 30-degree segments each representing a month.[1]

The most common graphic representations of data are line diagrams on arithmetic rectangular coordinates. These are so widely used that a set of standards has been drawn up for them by a committee representing many fields of study. The essentials

[1] Also for directional phenomena such as winds, letting the angular scale represent the points of the compass. Such phenomena are rare in zoology but not altogether unknown. Birdbanding data, for instance, usually presented on maps, are well shown on polar coordinates (Fig. 25).

of these recommendations are as follows, with some modification and explanation pertinent to the special interests of this book:

1. The general arrangement should be from left to right, that is, with lower values (of X) to the left and higher to the right, and from bottom to top, with lower values (of Y) below and higher above.

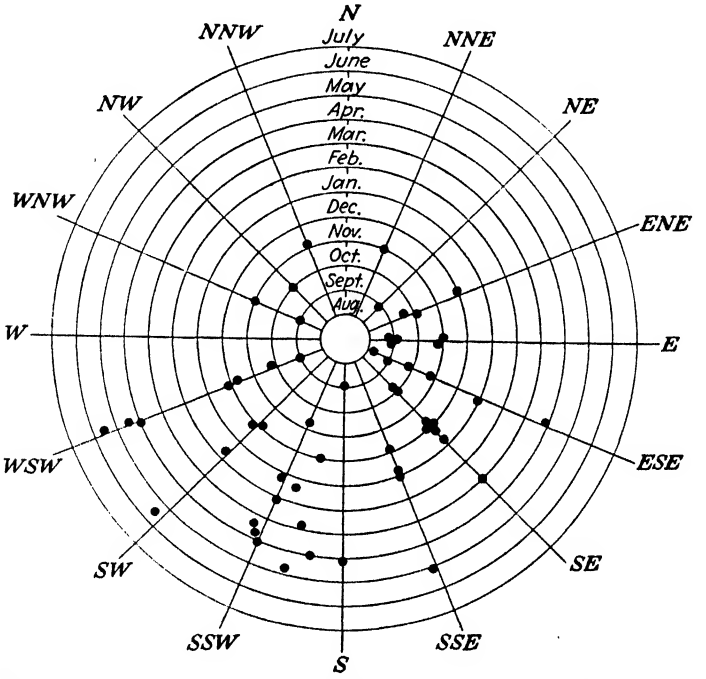


FIG. 25.—A graph on polar coordinates. Birdbanding data on herring gulls banded at Beaver Islands near St. James, Mich., and recovered during the first year (data from Eaton 1934). The angular distances, or directions of radii, indicate directions of the compass away from the banding place; and the concentric circles, or distances from the center, represent the dates of recovery, and hence elapsed time and age, in months.

2. Quantities should as far as possible be represented by or proportionate to linear magnitudes. In histograms and curves generally, areas are also important and necessary representations; but in histograms, specifically, these should be kept strictly proportionate to a linear magnitude (that of Y) by keeping the horizontal intervals equal.

3, 4. The zero lines should if possible be shown on the diagram, and if this leaves a large blank space it may be eliminated by a jagged break across the diagram. This recommendation is, however, unnecessary for much zoological work. The absence of the zero line is not misleading to anyone used to such diagrams if the scales are clearly marked.

5, 7. Coordinate lines that are natural limits, such as those for 0 or for 100 per cent, or that are otherwise exceptionally important should (or may) be emphasized; and others should not.

8. On logarithmic coordinates the limiting lines of the diagram should be powers of 10.

9. No more coordinate lines should be drawn than are necessary to guide the eye. It is often sufficiently clear and is generally neater simply to give scales at the left and bottom of a diagram and not to draw in any other coordinates.

10. The curve (or other noncoordinate diagram line) should be sharply distinguished from the coordinates, usually by being made heavier.

11. It is often advisable to emphasize individual observations, as distinct from a line based on them, as crosses or distinct dots on the diagram.

12. This recommendation has been combined with recommendation 1.

13, 14. Scales should be along the axes (seldom applicable to zoological diagrams) or to the left and at the bottom. Other pertinent data, formulas, etc., may, if desired, be arranged along the other two sides of the diagram or written within it.

15. The numerical data on which a diagram is based, if not clearly ascertainable from the diagram, should be given beside it or in the accompanying text.

16. Lettering should be clearly legible either as the diagram appears or after rotating it 90 deg. clockwise.

17. Diagrams should be clearly titled and should, as far as convenient, be self-explanatory without reference to an accompanying text.

SPECIAL TYPES OF GRAPHIC FREQUENCY DISTRIBUTIONS

The usual frequency polygons and histograms are limited to distributions of the absolute frequencies of a single variate with determinate numerical classes. Other types of graphs are necessary to represent such distributions for

1. Relative rather than absolute frequencies.
2. More than one variable; or
3. Attributes or variates in which the classes are not numerically determinate.

The representation of relative values, of frequencies or any other variables, is discussed on page 321.

The simplest method of representing the frequency distribution of more than one variable on a single diagram is simply to superimpose separate frequency polygons on the same field (see Fig. 6, page 66). They may be distinguished by the nature of the line used, solid, dashed, dotted, etc., or by shading the enclosed areas differently. If the magnitudes involved are about the same, the same scales may apply to both or all the distributions included,

but it may be necessary to give them separate scales. Such diagrams tend to become too involved to follow easily, and they should be avoided unless really simple, clear, and illustrative of an important relationship. Histograms can occasionally, but rarely, be combined in the same way without undue loss of clarity.

A second method particularly useful for histograms is to plot the combined distribution of two samples of the same

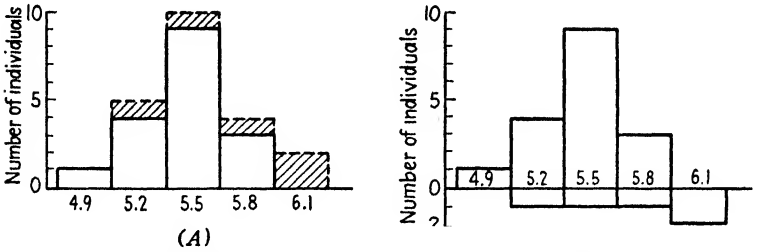


FIG. 26.—Combined histograms of two comparable distributions. Length of first lower molar in the extinct mammal *Anisonchus sectorius* (original data). A, samples added to each other; the clear areas represent a sample from New Mexico, and the crosshatched areas one from Montana. B, the same data with specimens from New Mexico above the horizontal axis and those from Montana below that axis.

variate, showing the contribution of the second by marking it off above the first and shading its area (Fig. 26A). Or, what amounts to the same thing, one sample can be plotted first and another then added above it. Three or more samples can be added together and plotted on the same chart in this way, and frequency polygons may be used instead of histograms.[1] For clarity it is important that the samples really be analogous and the variates homologous. It would, for instance, be valid and useful to plot in this way distributions of the same variates for males and females of one species collected together, or for two geographic samples of the same species; but it would usually be merely confusing to combine data on one variate for unrelated species or on two unlike variates for a single species.

A far less common solution of the same problem is to scale values on the Y-axis both above and below its zero point, at the intersection of the X-axis, and to plot one distribution in the

[1] The "band charts" of commercial statisticians are multiple frequency polygons of this sort.

ordinary way, above the X-axis, and one below it, as if reflected in a horizontal mirror (Fig. 26B).

Finally, it is possible to show the relationships of two frequency distributions very clearly by introducing a third dimension. The method is to prepare a correlation table and then to erect on each cell a column proportionate in height to the cell frequency. If the cells can be made very numerous and very small and still retain frequencies in most of them, the tops of the columns can be blended into a nearly smooth surface with hills and valleys corresponding to regions of greater and less frequencies. This is called the frequency surface or correlation surface, and it gives an almost ideal plastic representation of correlation. The construction of such three-dimensional models and their representation in a figure for publication are, however, too elaborate and too time-consuming for them to be used very extensively. Their reduction to two dimensions for a figure can be by perspective drawing or other oblique projection or by contour mapping like that of geographic maps.[1]

A scatter diagram of the correlation of three variables can be made by laying out the appropriate scales, two in a horizontal plane on a wooden or composition base and one vertical, and representing each triple observation by the head of a pin, its length determined on the vertical scale, inserted at the proper point on the horizontal base. Of several possible methods of representation on paper, perhaps the most practical if the observations are not too numerous is to represent each observation by a circle on the field of the horizontal scales with the third value given as a number in the circle.

A simple and almost always sufficient solution of the problem of graphic frequency distributions of attributes (and of numerically indeterminate variates) is to use a bar diagram. As in a histogram, each class or category is represented by a rectangle (or it may be in this case simply by a vertical line) with its height proportionate to the frequency represented. In bar diagrams, unlike histograms, a short space is generally left between successive categories, and each is separately labeled instead of being scaled continuously along the base of the diagram. The categories of attributes seldom have any necessary or logical order, and the usual practice is to arrange them

[1] Yule gives illustrations of both methods.

in the order of their frequencies, the highest to the left. It is a great advantage of this helpful and elegant method that almost any number of contiguous bars can be placed in one category, each representing a different sample, so that comparisons are

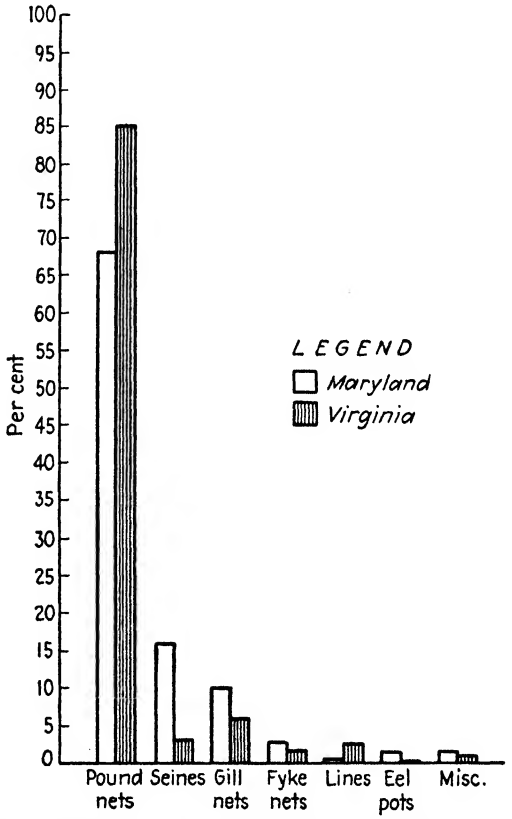


FIG. 27.—Bar diagram comparing categories of two different samples of an attribute. The attribute is method of collecting fishes in Chesapeake Bay during 1920, with the categories shown. Frequencies are given as percentages of total catch. The two samples are the Maryland catch (clear bars) and the Virginia catch (hatched bars). (*Data from Hildebrand and Schroeder* 1928.)

greatly facilitated. It is advisable in such cases to shade the bars differently for the different samples. A single diagram can thus show the distributions of an attribute in males and females, in young and adults, in different years, in samples from different localities, etc. A bar may be given in each category to show an average value for the samples represented. Samples may also

be added vertically or their component subsamples represented in the same way as for histograms of variates (see Figs. 27 and 28).

Pyramid diagrams may be used to represent distributions of attributes or, especially, variates. They are constructed by taking the rectangles of bar diagrams and histograms, turning them so that they are horizontal, and piling one on top of the other, centered on a vertical line, so that they look like an edge-

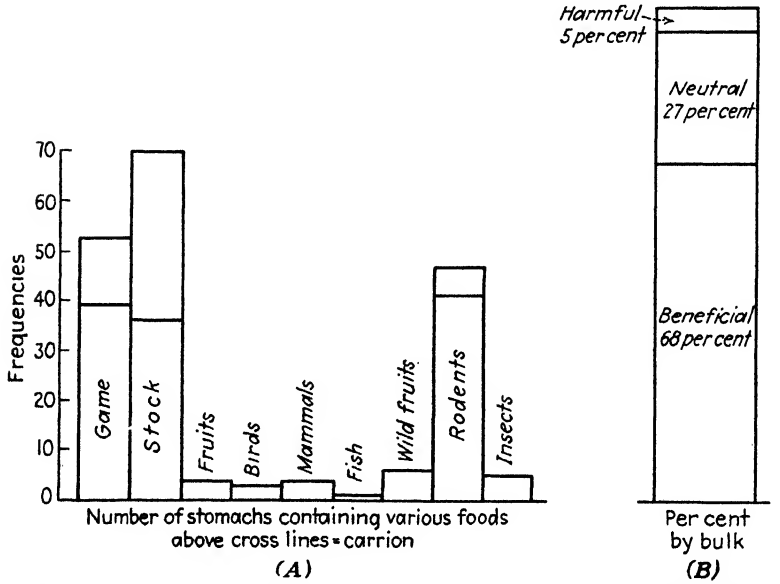


FIG. 28.—Bar diagrams of a frequency distribution of an attribute: food habits of the coyote. A, number of stomachs in a sample of 185 containing traces of each of 9 kinds of food; the distance above the crossline in 3 of these represents carrion. B, food contents from the same stomachs classified on an economic basis and represented by percentages of bulk. (*From Dixon* 1925.)

wise view of a stack of coins of different diameters but the same thickness. They have little advantage over ordinary bar diagrams and histograms and some disadvantages, and they are rarely used.

COMPARISON OF RANGES AND MEANS

The importance of comparisons of samples of homologous variates is great and has repeatedly been emphasized. The correct graphic representation of such comparisons may convey at a glance all that is most important in a long and complex series

of measurements and calculations, and this is one of the most useful of all graphic methods. The simplest form of the best of such methods was early employed by Ruthven[1] (1908). In it a vertical scale is used for the values of the variate, and each sample is represented by a vertical line, the ends of which correspond,

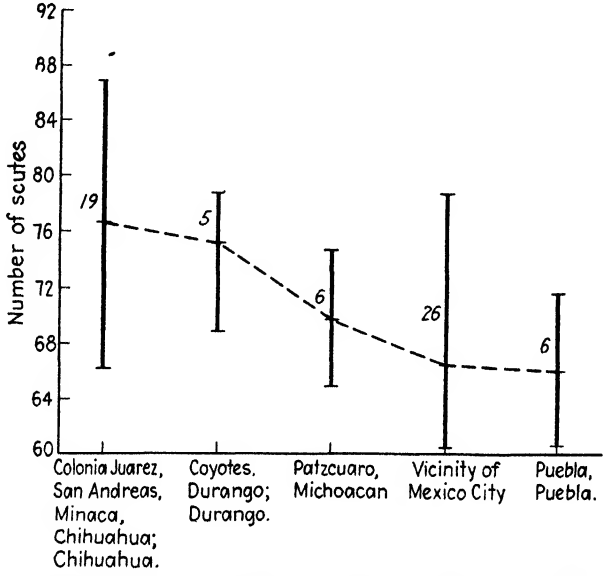


Fig. 29.—Simple graphic comparison of observed range and means. Number of subcaudal scutes in the snake *Thamnophis megalops* from five localities in Mexico. Values of the variate are scaled vertically, not horizontally as in most diagrams. The heavy vertical lines represent observed ranges, and the short crosslines are the means, connected by a broken line to facilitate comparison. The numbers to the left of the vertical bars are the sizes of the samples. (*From Ruthven* 1908, *somewhat modified.*)

on the scale, to the extreme observed values. A crossbar is placed on this line at the arithmetic mean.[2] If the samples have a natural sequence, for instance, geographically, a line, preferably broken to distinguish it from other lines in the diagram, connecting the successive means may be added, giving a graphic representation of a sort of trend.[3] It is also helpful to write

[1] Who states that it was suggested by Dr. Raymond Pearl.

[2] Ruthven used a cross (*X*), but a crossbar seems preferable. Ruthven also placed a circle at the middle of each bar (*i.e.*, at the mean of the two extreme values), but this does not seem to us to be useful.

[3] This was also done by Ruthven. Of course, such a line has little useful meaning if the samples have no natural sequence, although even in such a case it may be a helpful guide for the eye.

near the line for each sample a number indicating its total frequency (see Fig. 29).

If the differences between the samples can be expressed numerically, it might be practical and certainly would be useful to lay these values off on a horizontal scale and to make the distances between the vertical lines representing the samples proportionate to the numerical differences between them. This would, for instance, be possible in many cases of samples geographically separated (by miles, or latitude, or longitude) or samples taken at different times (at different hours, on different days, in different months, etc.) or in environments numerically different (in temperature, humidity, etc.).

Another improvement over current practice would be to place crossbars on or, when necessary, above or below the line for each sample to correspond with the values $(M + 3\sigma)$ and $(M - 3\sigma)$, for samples large enough to calculate σ with reasonable reliability. As has already been discussed, these points usually approximate the real population limits more closely than do the extremes of the actual observations.

A notable addition to such graphs was recently devised by Dice and Leraas (1936) and promises to be one of the most important of numerical methods in zoology. Crossbars are added at $(M + 2\sigma_M)$ (twice the standard error of the mean, not twice the standard deviation of the distribution) and $(M - 2\sigma_M)$, and these are connected by double vertical lines in order to define clearly the range so laid out. If these ranges overlap for two samples, it may usually (under conditions to be more exactly defined in the following paragraphs) be assumed that the differences between the means are not significant; and if they do not overlap, it may usually be assumed that these differences are significant. The comparison of two means by more exact methods is not difficult (see page 192), but it becomes laborious if many samples must be compared. Thus for five samples, 10 standard errors of differences between means would have to be calculated. The same information in less reliable but still generally suggestive form can be obtained by this graphic method with the calculation only of the means and their standard errors. Since the method is only approximate, it cannot entirely replace the exact calculation of the standard errors of differences between means; but it will usually show at a glance what differences

require testing in this way. It also presents the essential facts as a unified whole and in a way easily grasped (see Figs. 30 and 31).

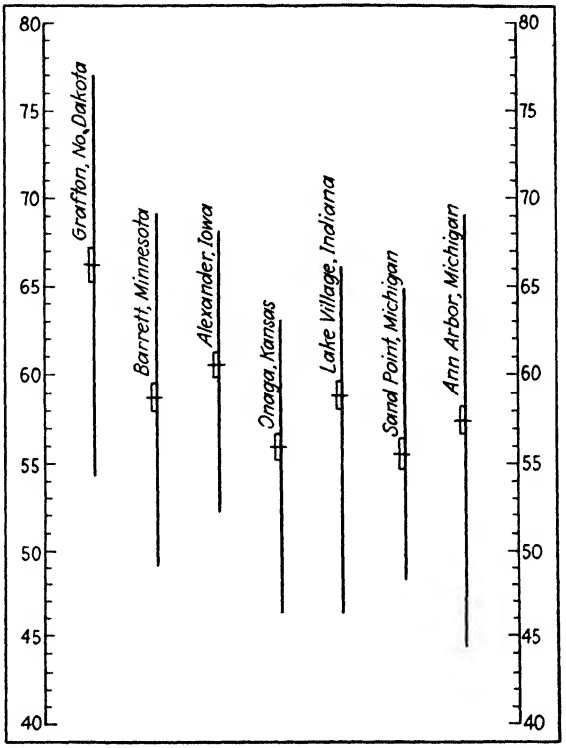


FIG. 30.—Graphic estimation of the significances of the difference between several means. Tail length in the mammal *Peromyscus maniculatus bairdii* from seven localities. Values of the variate are scaled vertically and given on both sides of the diagram to facilitate comparison by means of a straightedge. The vertical lines are observed ranges, and the small rectangles represent the range $(M + 2\sigma_M)$ to $(M - 2\sigma_M)$, with a crossline at M. If these rectangles do not overlap, it can be assumed under most conditions (as specified in the text) that the means differ significantly. (*From Dice* 1937.)

The relationship between the standard errors of two means and the standard error of their difference is given by the equation

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

* See p. 192. We here assume that the purpose is to test the probability that the samples are from one population, which will usually be the intention. If the purpose is to test differences between means of samples admittedly of

This relationship thus depends on the relative sizes of the two samples and the relative magnitudes of the standard errors of the

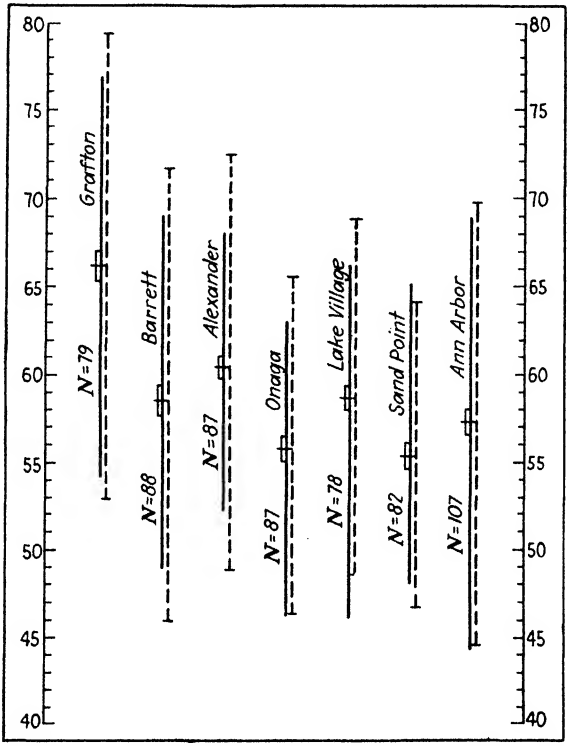


FIG. 31.—The data and method of Fig. 30 with the sizes of the samples and theoretical ranges ($M \pm 3\sigma$) added. The theoretical ranges are represented by broken vertical lines.

two means. For instance, if $N_1 = 2N_2$ and $\sigma_{M_2} = 2\sigma_{M_1}$, the value of σ_d can be thus expressed:

$$\sigma_d = \sqrt{\frac{(2N_2)^2\sigma_{M_1}^2 + N_2^2(2\sigma_{M_1})^2}{(2N_2)N_2}}$$

an equation that is easily reduced to

$$\sigma_d = \sqrt{\frac{8\sigma_{M_1}^2}{2}} = 2\sigma_{M_1} = \sigma_{M_2}$$

different populations, some of the limitations mentioned below are removed, and the method is still more generally valid.

Therefore under these conditions σ_d is equal to σ_{M_1} and is twice σ_{M_1}. Now the difference that is taken to be significant by the graphic method is $2\sigma_{M_1} + 2\sigma_{M_2}$; and when $\sigma_{M_2} = 2\sigma_{M_1}$, this is equal to $3\sigma_{M_2}$, or $6\sigma_{M_1}$. Therefore

$$\frac{d}{\sigma_d} = \frac{6\sigma_{M_1}}{2\sigma_{M_1}}$$

or

$$\frac{3\sigma_{M_2}}{\sigma_{M_2}} = 3.00$$

The graphic method takes this as the limit of significance under the stated conditions. It corresponds with $P = .003$, and though this is a somewhat low value as a criterion of significance its use is valid, and the graphic method is a good approximation under these circumstances. Similarly it is possible to calculate what value of d/σ_d is being taken as the criterion of significance with any other given relationships between N_1 and N_2 and between σ_{M_1} and σ_{M_2}. Because the method is so important and because it cannot be used intelligently without some conception of these values, we have calculated the values of d/σ_d and of P automatically taken as criteria of significance when the ratios N_1/N_2 and $\sigma_{M_1}/\sigma_{M_2}$ take certain representative values. The results are presented in Table XIII.[1]

When the value of P used as a criterion is too low, every difference shown as significant by this method is so, but some that are really significant will not appear as such on the diagram. Such an error is on the side of safety, but it invalidates the method (beyond the limits to which the table is carried) when

[1] The table was calculated by the following formulas, which can be used to obtain values for any conditions not approximated in the table.

$$\text{Let } a = \frac{N_1}{N_2} \quad \text{and} \quad b = \frac{\sigma_{M_1}}{\sigma_{M_2}}.$$

Then

$$\frac{\sigma_d}{\sigma_{M_2}} = \sqrt{\frac{a^2b^2 + 1}{a}}$$

and

$$\frac{d}{\sigma_{M_2}} = 2 + 2b$$

These formulas are of obvious derivation and simply sum up the procedure explained in the text and place it in the easiest form for calculation.

it means that significant differences pertinent to the problems studied will usually be missed. Too high a critical value of P, on the other hand, means that the method will show differences as significant when they are not. Circumstances that give rise to this condition, including those on the table remarked as "too high," make the method too misleading to be used at all.

Fortunately such circumstances will not often arise with zoological data. It will be noted that if neither N nor σ_M of one

TABLE XIII.—CRITERIA OF SIGNIFICANCE OF THE DIFFERENCE BETWEEN TWO MEANS AUTOMATICALLY ASSUMED BY THE DICE-LERAAS METHOD OF GRAPHIC COMPARISON WITH GIVEN RATIOS OF SIZES OF SAMPLES AND OF THEIR STANDARD ERRORS OF MEANS

N_1/N_2	$\sigma_{M_1}/\sigma_{M_2}$	σ_d/σ_{M_2}	d/σ_{M_2}	d/σ_d	P	Remarks on value of P as a criterion of significance
1	1	1.41	4	2.8	.005	Good
1	½	1.12	3	2.7	.007	Good
1	2	2.24	6	2.7	.007	Good
2	½	1.00	3	3.0	.003	Somewhat low
2	1	1.58	4	2.5	.012	Good
2	2	2.92	6	2.1	.04	Poor (high)
1	⅓	1.05	2⅔	2.5	.012	Good
1	3	3.16	8	2.5	.012	Good
2	⅓	.85	2⅔	3.1	.002	Somewhat low
2	3	4.30	8	1.9	.06	Too high
3	1	1.82	4	2.2	.03	Poor (high)
3	½	.87	3	3.4	.001	Poor (low)
3	2	3.51	6	1.7	.09	Too high
3	⅓	.82	2⅔	3.3	.001	Poor (low)
3	3	5.22	8	1.5	.13	Too high

sample is twice that of the other sample the criterion implied will be valid and will generally be good, with high reliability. The only exception is that some suspicion attaches to the results even in these cases when the larger sample also has the higher σ_M; when both ratios are 2 the criterion is $P = .04$, considered as usually valid by many workers but admitted not to be invariably reliable. If the samples are of about equal size, $N_1/N_2 = 1$, the results are good within the limits of our table and are in fact acceptable with values of $\sigma_{M_1}/\sigma_{M_2}$ up to about 10 (when $P = .05$; it becomes greater as this ratio increases). Such a

great difference between the standard errors of the means of homologous variates is very rare with zoological data, so that the method is almost always valid when the samples are of about equal size. When the ratio N_1/N_2 is greater than 2, however, the method is so generally poor or essentially unreliable that it should not be used, although, as the table shows, it still retains some validity with $N_1/N_2 = 3$ if $\sigma_{M_1}/\sigma_{M_2}$ is 1 or less. Noting the sample sizes on the diagram permits the reader to form some judgment as to reliability. It is a drawback of the method that the exact criterion of significance implied in it is variable and is difficult to judge, but within its limitations the method has advantages that outweigh this. In doubtful cases, σ_d should in any event be calculated and given in the text.[1]

PROPORTIONS AND PERCENTAGES

In some cases, for instance, in the most practical methods of colorimetry, observations are made and recorded as relative rather than absolute values. In other instances, it is helpful to reduce absolute data to relative values, usually to percentages, in order to facilitate comparisons. Samples of different sizes but with variations present in the same proportions may appear very different in a graph if absolute values are used. The plotting of percentages at once reveals their similarity. Such diagrams

[1] Dice and Leraas use probable errors, marking out $(M + 3\ \text{P.E.}_M)$ and $(M - 3\text{P.E.}_M)$ but, as they note, this differs little from $(M + 2\sigma_M)$ and $(M - 2\sigma_M)$. They imply that the method is more generally valid than now appears to be the case; and their calculations, which they do not give, evidently assumed that the samples were all of the same size or that comparisons were to be made only between populations known to be different. Yet their method is certainly valid in most cases and is of great value.

Rosahn (1935) has devised another graphic method for the same purpose. He makes a bar diagram with the vertical height of each bar scaled to the value of the mean for each sample. For the two samples he then calculates the standard error of the difference between their means and lays out on the diagram horizontal lines at intervals scaled to this value, so that the significance of the difference in height of the two bars can be observed visually. For multiple samples he takes a sort of average standard error of the differences and lays it on the diagram in the same way. He also assumes that the samples will be of about the same size and that the standard errors of their means will be nearly the same. The method is, under some special circumstances, more reliable than that of Dice and Leraas; but it is also much more laborious and has other disadvantages that seem to make it of little practical use for most zoological data, at least.

have, however, the one serious drawback that they do not take into account the varying reliabilities of the observed proportions or their probable significance. They should, therefore, be accom-

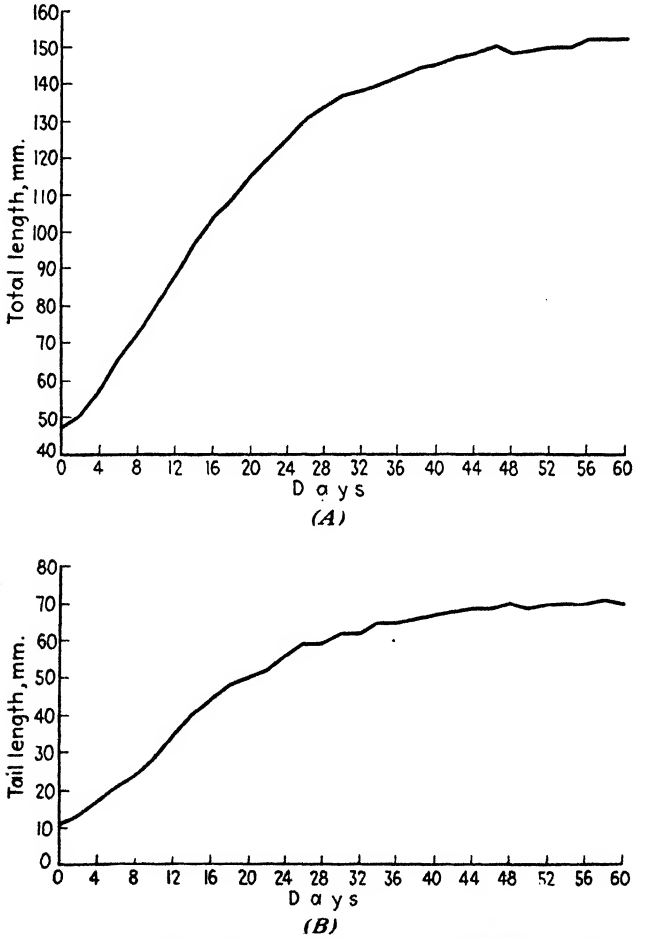


FIG. 32.—Growth curves of two variates recorded by absolute values. Mean measurements on identical samples of the mammal *Peromyscus maniculatus artemisiae*. A, growth curve of mean total length. B, growth curve of mean tail length. Because of the much smaller absolute values of B, the curves appear to be of different types. (*From Svihla* 1925.)

panied by numerical data, in the text, in the legend, or lettered on the diagram itself, by which these probabilities may be judged. It is usually advisable to state the absolute values represented

by the percentages or the total frequency included. Further analysis by various of the methods explained in previous chapters may also be necessary.

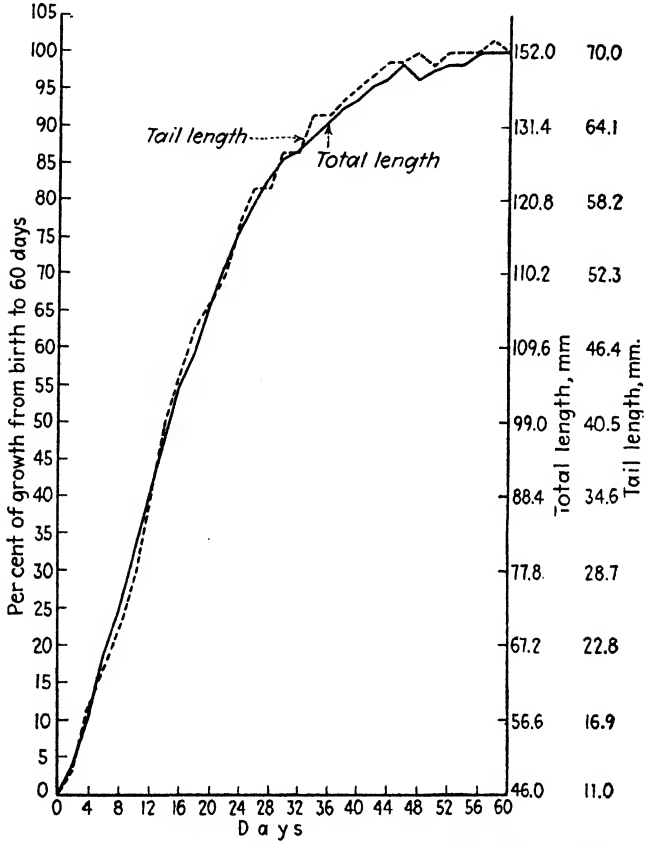


FIG. 33.—Growth curves of two variates recorded as percentages of total growth attained in the period of observation (same raw data as Fig. 32). The continuous line represents growth in total length, the broken line growth in tail length. Reduced to a proper basis for comparison, the two curves are seen to be almost identical in character.

Any graphic method representing absolute values may also be used to represent relative values. Frequency polygons, histograms, trend lines, and the like all lend themselves to this use. It is, however, often desired to compare relative values from two or more samples but also to retain the graphic record of absolute values. This problem can be neatly solved by plotting

the samples on a single diagram and equating critical values involved in the problem, for instance, making the graphic representation of total frequency the same for all samples or making the initial and terminal points of growth curves coincide. The values for the various samples are thus made graphically comparable and relative; the absolute values can be retained by calculating and showing a different scale for each sample. The

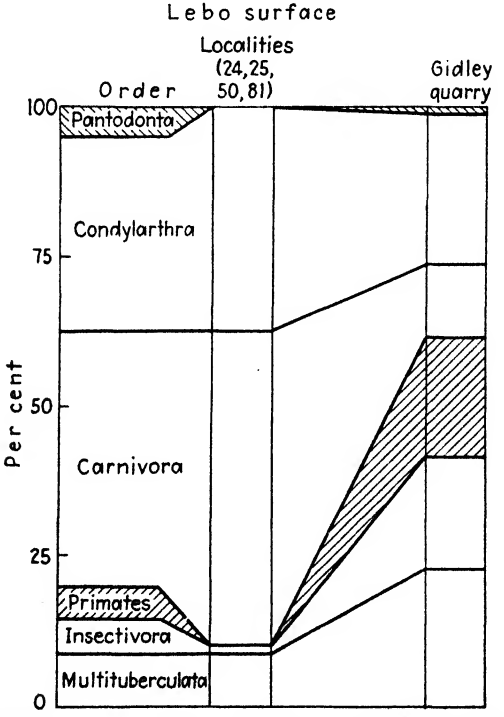


Fig. 34.—Graphic comparison of compositions of two samples by bar diagram. Surface and quarry collections of fossil mammals from the Paleocene of Montana (original data) subdivided by percentages of the various taxonomic groups (orders) represented.

method becomes too elaborate with many samples but is excellent for the comparison of a few. Figure 32 gives a typical example. As originally published, the growth curves were separate and were given by absolute values on the same scale for both. The two curves then appeared to be very different, and their author concluded that the growth of these two variates was quite distinctive. Figure 33 shows that the growth curves are really almost identical.

The apparent difference was not caused by kind or by relative rate of growth but simply by the different initial and terminal absolute magnitudes.

For representing relative frequencies of the different categories of an attribute (or more rarely the different classes of a variate), bar diagrams are useful. The samples are represented by rectangles all of the same height, and each is subdivided by horizontal lines into areas proportionate to the frequencies of the various categories. Connecting the homologous division lines by lines (often broken) from one rectangle to the next generally makes the visual comparison still simpler and more obvious (see Fig. 34).

A variation of this method is to represent each sample by a vertical line clear across the field of the diagram (for instance, from 0 to 100 per cent) and to record the proportions by drawing oblique or horizontal lines between the proper points on the various lines, these points being determined in the same way as the division lines of the rectangles when these are used. In drawing the diagram it is convenient to make the lines or rectangles representing the sample 10 cm. (100 mm.) in length or some integral multiple of this. Then the data in percentage can be directly transfered with no further calculation; for (using 100 mm.) 1 mm. on the line represents 1 per cent in the data. Following Sumner, this method is now generally used for graphic representations of results of tint photometry. The samples in this case are observations through different color screens, with one observation, a relative value as read, in each sample. The length of the vertical line for each sample represents the reading for pure white, and the point graphed is at the relative value of the reading for the object being observed (see Fig. 35).

Pie diagrams[1] are circles divided into sectors proportionate in area to the relative values of the categories represented. They do not lend themselves well to the representation of anything but proportions or percentages but are useful for this purpose. The procedure is to change the data into percentages and then to convert these into angles, measured in degrees, by multiplying by 3.6. The zero or first radius is drawn vertically upward from the center (to twelve o'clock on a clock dial). If there is no natural order of the categories represented, it is customary to

[1] Called sector diagrams by those who do not like pie in science.

place them in the order of diminishing size clockwise (see Fig. 36). If there is a natural order, the first sector lies immediately to the right of the zero radius, and the others follow clockwise. If there are many very small sectors, the radii may become too

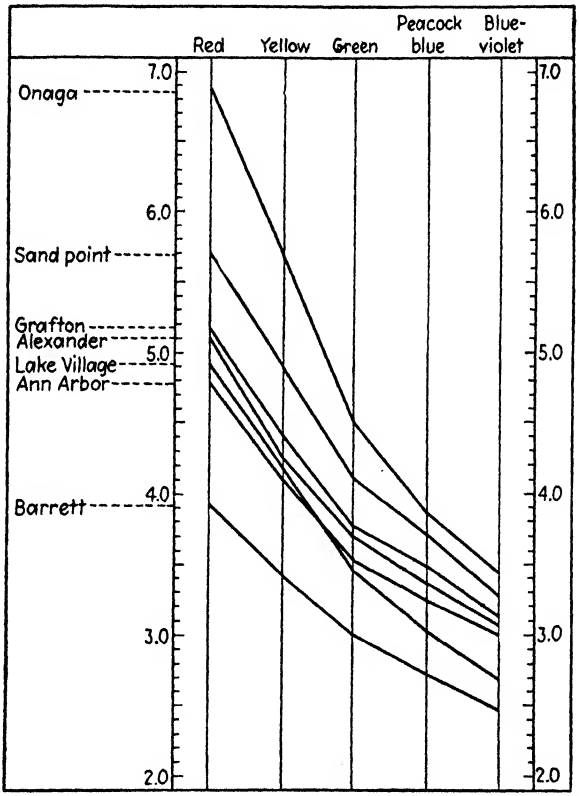


FIG. 35.—Graphic comparisons of tint photometry data. Skins of the mammal *Peromyscus maniculatus bairdii* from seven different localities. The scales on the two sides are tint photometer readings, and each of the five vertical lines between them represents the use of a filter of the designated standard color. Each of the oblique lines passes through the mean values for the sample from the given locality, and each line as a whole thus represents the essential color data for the corresponding sample. (*From Dice* 1937.)

crowded or tend to form a black blot near the center, and in this case it is advisable to draw small and large concentric circles and draw the radii only between these, not continuing them to the center. It is also possible to draw one or more still larger circles and to represent on them the proportions of the same data

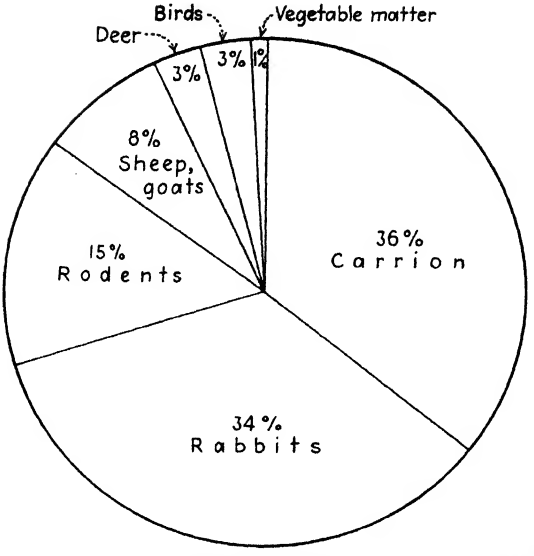


FIG. 36.—Pie diagram used for graphic representation of food habits. Winter stomach contents of coyotes in percentages. (*From Sperry* 1934.)

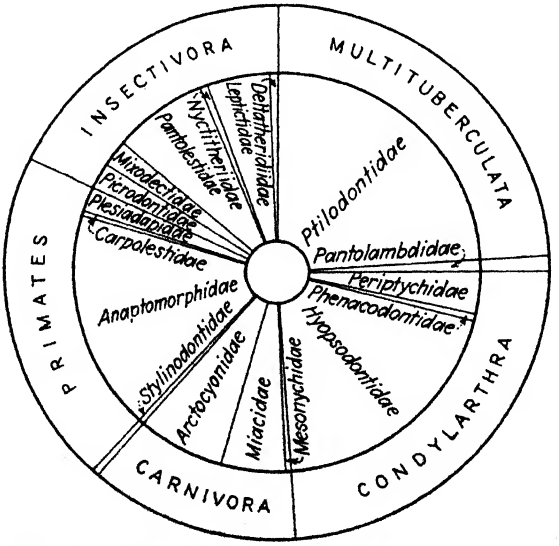


FIG. 37.—Modified pie diagram to show composition of a fauna. Percentages of specimens belonging to various taxonomic groups in a collection of Paleocene mammals from Montana (original data). The outer ring shows the orders represented and the inner ring their constituent families in this fauna. The center is left clear to avoid the confusion of many lines converging to a point.

arranged in different and broader categories; for instance, if the categories are taxonomic, an outer ring may show relative abundances by classes or orders and an inner ring or circle the same by suborders or families (see Fig. 37).

Pie diagrams convey certain simple ideas more directly to the eye than almost any other sort of graph, but they are not ideal for scientific comparisons. Their most extensive use has been to impress the layman and to popularize the numerical results of some types of investigation. For this purpose their appeal and their self-explanatory nature are often enhanced by putting in each segment, if it is large enough, an actual picture of the thing that it represents. For instance, if the categories are families of animals, a typical member of the family may be portrayed in each sector. In all cases the sectors should also be labeled in words; and usually it is advisable also to give in each a number indicating the percentage or, rarely, the absolute value that it represents.

TIME SERIES AND PERIODIC PHENOMENA

Most time series really involve at least three variables: time, frequencies, and values of a variate or categories of an attribute. Graphic representation usually eliminates or omits one of these, although all three can be shown if necessary. In series like growth curves, frequencies usually do not appear. The curve is based either on a single individual or on means for a group. In either case, frequencies need not be represented, for they are the same throughout. This is not invariably true of curves based on groups, but as far as possible these should not only have the same frequency throughout but also be based on precisely the same individuals. If this cannot be done, the deficiencies of the data should be very plainly stated, for the curve will then be relatively unreliable and may be very misleading.

Time series showing means may often be improved and may convey much additional and important information by plotting also a line, distinguished in some way such as being broken, through the points $(M + \sigma)$ and $(M - \sigma)$ or $(M + 3\sigma)$ and $(M - 3\sigma)$ for each time class. Changes in dispersion over periods of time are thus represented, and these may be even more enlightening than changes in means.

In other cases there are only one or a few categories of the variable being observed, and the time series is primarily concerned with frequencies, these then being plotted on the vertical axis against time on the horizontal axis. The frequencies may, of course, be represented either in relative or in absolute form.

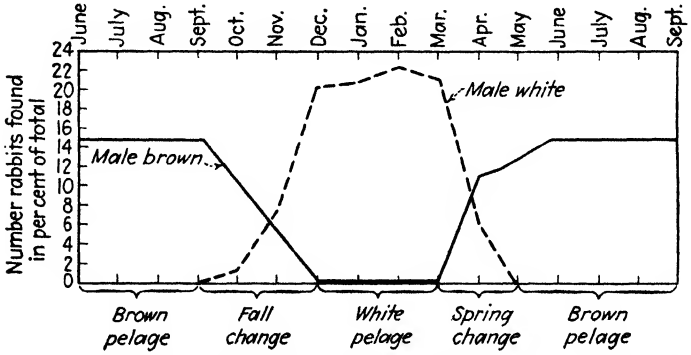


FIG. 38.—Graphic representation of periodic phenomena. Percentages of males in two color phases in total collections of snowshoe rabbits taken throughout the year, showing seasonal shift in dominant color phase. (*From Aldous* 1937.)

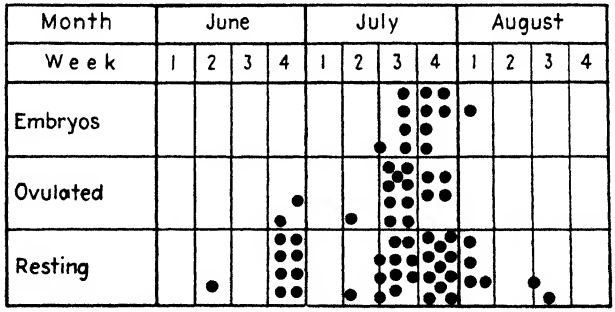


FIG. 39.—Point diagram of a periodic phenomenon. Samples of female Eastern American chipmunks collected during three summer months. Each dot represents an individual, classed according to three physiological (reproductive) conditions. The diagram shows a marked breeding season beginning late in June and culminating in July. (*From Schooley* 1934.)

If there is more than one category, lines for each, distinguished in some way, may be plotted on the same diagram (see Fig. 38). The same method may be used in a bar diagram, with one bar for each time class, its height proportionate to the frequency. Additional categories may be represented by subdividing the bars or by including in each time class a separate bar for each

category. Times, frequencies, and observed values may also all be shown at once in a diagram by various modifications of scatter or point diagrams (see Fig. 39).

Phenomena that are periodic or cyclical with a definite period can also be represented in any of these ways on rectangular coordinates, but often they are particularly adapted to the use of polar coordinates. In this case the angular scale represents time, 360 degrees being equated to the length of the cycle, and frequencies or values of a variate are represented by distance from the center.

PICTORIAL DIAGRAMS

The name pictorial diagrams is here used broadly to mean most sorts of diagrams not including a definite numerical scale or not drawn on a system of coordinates and not only those that literally involve pictures in the vernacular sense. Diagrams not essentially pictorial may be dressed up with or expressed in terms of pictures to make them more sprightly or to give them popular appeal. This is often done with pie diagrams, as has been mentioned. Another method is to represent classes (*e.g.*, time groups) by separate pictures of the thing being studied and to make the sizes of the pictures proportionate to frequencies or to values of a variate,[1] or each class may be given a number of repeated pictures of equal size in proportion to these frequencies or values.

Many sorts of pictorial diagrams useful in zoology are essentially nonnumerical and so fall outside the scope of this book, but a few may be mentioned. Graphic representations of animal phylogenies are of this type. Some of these are highly ingenious and manage to convey much information even in addition to the supposed lines of descent, for instance, about geographic distribution, environment, habits, or morphological changes

[1] These diagrams, widely used on posters and the like but not in scientific work, are always misleading. If, as is usually done, the height of the pictured person, animal, or thing is made proportionate to a frequency or other value, the differences are exaggerated in apparent value. If the presumed volume of the thing pictured is used for scaling, the differences are strongly minimized to the eye. What the eye actually judges in a printed picture is area, but the use of area for scaling in such diagrams is not so apparently logical as either height or volume and is almost never encountered.

(see Fig. 40). Usually, time is represented on a vertical scale in such charts, with the oldest forms at the bottom. Still more

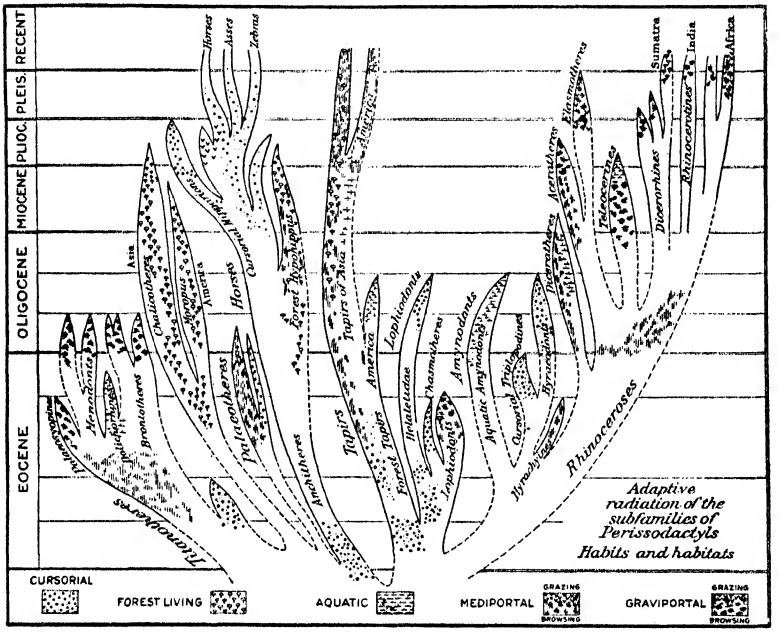


FIG. 40.—Pictorial diagram of phylogeny, habits, and habitats in a group of animals. Fossil and recent perissodactyls, as interpreted by Osborn. The vertical scale represents time, and the horizontal dimension (not scaled) is utilized to show the spread of the various phyla. Habits and habitats are shown by special patterns. (*From Osborn* 1929.)

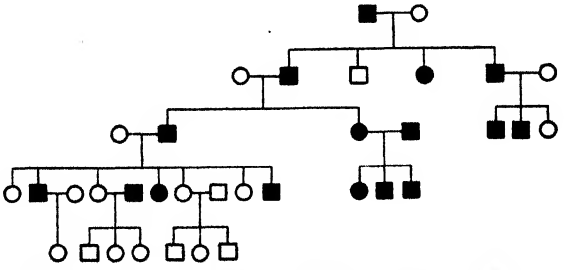


FIG. 41.—Chart representing genetic descent and inheritance. The chart represents a human family descending from the couple represented at the top. Squares represent males and circles females. Those in solid black represent bald individuals. (*From Snyder and Yingling* 1935.)

striking and convenient in many instances is a circular arrangement with phyla radiating from a central point. Charts of

genetic descent and relationship of individual animals also belong to this general group; and they, too, manage to express much more than descent alone by using symbols of different shape, shading, and subdivision for the various individuals (see

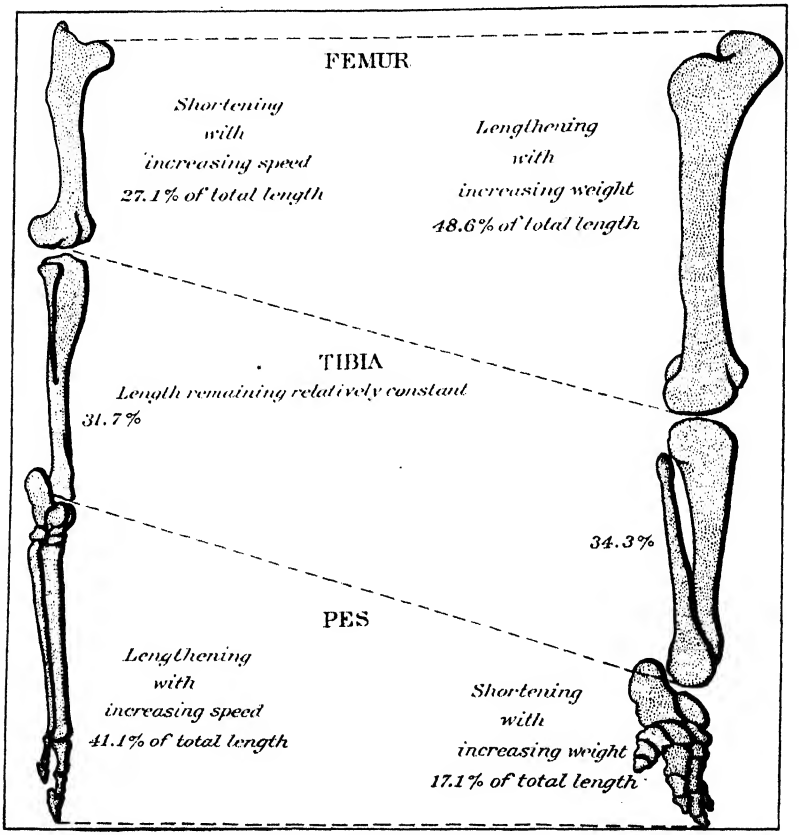


FIG. 42.—Pictorial diagram of contrasting anatomical proportions. Limbs of horse (left) and elephant (right), reduced to the same length, to show the marked difference in proportions of segments correlated with different habits and bulk. (*From Osborn* 1929.)

Fig. 41). Data dealing with geographic distribution of any sort are also frequently and well represented pictorially, on maps, bathymetric charts, cross sections of parts of the earth, pictures of landscapes, etc.

Among pictorial zoological diagrams, two of particular importance deal with essentially numerical ideas, even though usually

omitting actual numbers, and so merit special comment here. The first of these compares proportions or absolute dimensions of various sorts in ancestral and descendent or in otherwise related animals by presenting drawings of the anatomical elements concerned with homologous points connected by lines, usually broken (see Fig. 42). Numerical values of the proportions or absolute dimensions may also be inserted on the figure.

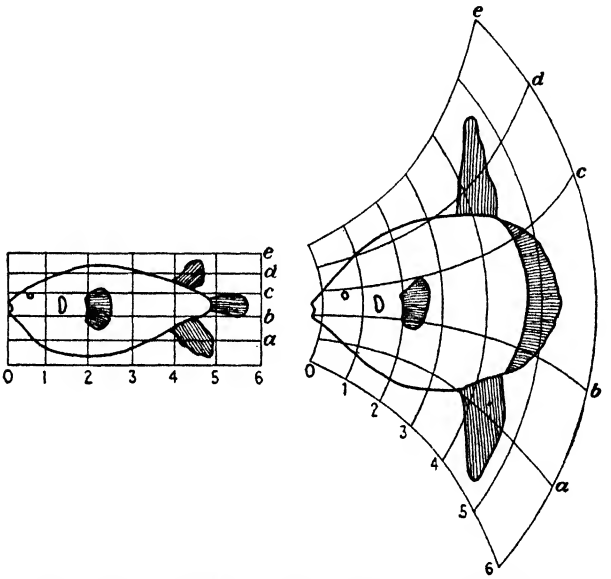


FIG. 43.—Pictorial comparison of proportions by rectangular and deformed coordinates. *Diodon*, left, has approximately normal fish proportions and is laid out on rectangular coordinates. A corresponding deformed coordinate system laid out on *Orthagoristus*, right, shows an almost perfectly regular expansion of the posterior part. (*From D'Arcy Thompson* 1917.)

A more complex, very interesting way of making similar comparisons in greater detail and taking into account areas and angles or directions as well as linear dimensions has been devised by D'Arcy Thompson (1917) and has since been used chiefly by paleontologists. On a drawing of all or any part of an animal a regular system of rectangular coordinates is superposed. The same parts of an allied animal are then drawn, and on it are superposed lines, some or all of which usually must be curved, such that each line passes as nearly as is practical through points homologous with those touched by a corresponding line on the

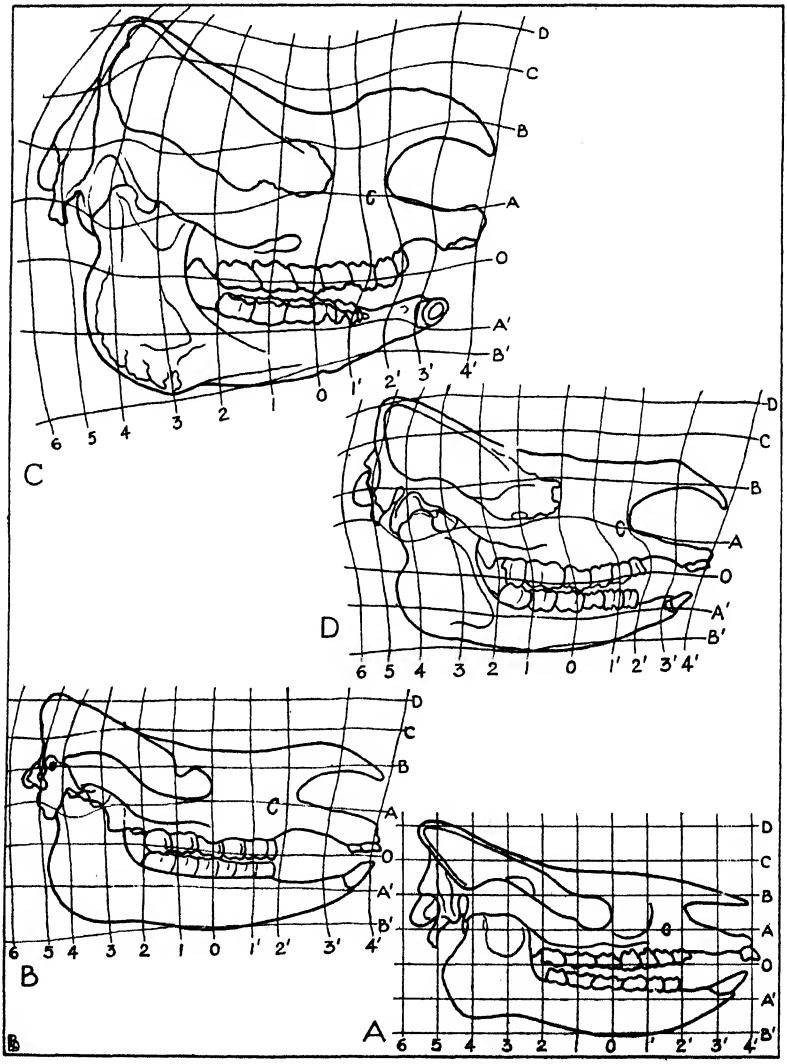


FIG. 44.—Changing proportions in phylogeny shown by deformed coordinates. C and D are skulls of modern rhinoceroses, D more primitive, to which B is approximately ancestral. A represents the primitive structure of the whole group and is laid out on rectangular coordinates, and the deformation of corresponding coordinates on the other skulls shows progressive irregular changes in proportions. A, *Subhyracodon occidentalis*. B, *Gaindatherium browni*. C, *Rhinoceros unicornis*. D, *Dicerorhinus sumatrensis*. (*After Colbert*, 1935.)

first drawing and each space between the lines covers approximately the same anatomical features as a corresponding square (or rectangle) on the first drawing. There is no way of making all the points and areas correspond absolutely if the animals differ markedly, but a close approximation is usually possible for forms that are visibly related. An effort is made to make the second system, the deformed coordinates, as simple as possible, with its lines smooth curves and not strongly irregular (see Fig. 43).

The method is most usefully applied by taking an ancestral or a presumably primitive animal as a standard of comparison, laying the rectangular coordinates on it, and then developing from it systems of deformed coordinates for more specialized descendants or relatives. The deformed coordinates then provide a vivid visualization of the differential expansion, contraction, and skewing that have taken place in the course of evolution (see Fig. 44). The method has also been used, not always with complete success but at least always with results more reliable than the usual guesswork, to effect harmonious restoration of missing parts in fossil animals and to reconstruct hypothetical intermediate stages between less and more specialized animals.[1]

[1] Osborn (1929) tried the experiment of drawing a short broad skull on a sheet of rubber, stretching the rubber until the drawing was long and narrow, and comparing with a naturally long and narrow related skull. What the experiment revealed was that the artificial harmonic distortion did not very closely approximate the naturally elongated skull, which differed from the short broad skull not simply in being relatively longer and narrower but also in the proportions of its various parts.

CHAPTER XVI

GROWTH

It would be impossible, in a work devoted to numerical methods in zoology generally, to consider every sort of data and of analysis involved in special zoological studies. There is, however, one group of problems that are of remarkable importance, that demand some numerical methods in addition to those already discussed, and that also illustrate in a clear, useful way the application and adaptation of general methods in approaching a specific problem. These problems, those arising from the numerical phenomena of growth, are therefore given special consideration in the present chapter. In addition to supplementing what has gone before, this chapter is in a sense a review and a large-scale example of the whole subject of numerical treatment of zoological data, for most of the principles and many of the methods already expounded here find application. It is also an introduction, elementary but adequate for any but extreme specialists, to what promises to be one of the most important and fruitful fields of strictly zoological and paleontological research and one that has only recently begun to be appreciated. It will be found also that many of the methods and results based on growth have far wider applications and implications.

AGE GROUPS

Since most animals have a more or less sharply defined breeding season, the young are normally born during a definite season or at least more commonly during some months of the year than during others. It thus often happens that in collections made over a short period of time the specimens will fall into distinct age groups rather than being evenly distributed as to age. This result is of great zoological significance. It is important in itself, as showing rates of growth, breeding seasons, and age of maturity. It also is important for taxonomy; for the different age groups produce an effect on distributions almost identical with that of

different species, and competent taxonomy must thus bear this possibility in mind. The study of variation has two aspects: variation within a single age group and variation correlated with age. The two must be separated for adequate analysis, and the methods of studying them are not the same.

The degree of distinctness of age groups depends on these factors:

1. Limitation of birth season. The shorter this is in relation to the life span of the animal, the sharper will be the age group. If there is no definite birth season (young are about as likely to appear at one time as another), there will be no age groups. Man well exemplifies a species in which age groups are practically undistinguishable for this reason.
2. Rapidity of growth, or of any changes correlated with age. Obviously if these changes are rapid, the discontinuity between age groups will be correspondingly great; and if they are slow, the discontinuity will become obscure or disappear. In almost all animals, growth is more rapid in early life than in maturity; and consequently the younger age groups are more clearly distinguished than the later. In mature animals, age groups usually are inseparable for this reason. The example given below, for *Hemidactylium scutatum*, clearly illustrates this.
3. Variability within any one age group. Highly variable groups will tend to overlap and less variable groups to be better distinguished.
4. Uniformity of growth within the species. This is a special case of the last factor. If some individuals grow much more rapidly than others, they may well overtake the slower members of the next older group; and hence the groups will soon tend to merge.
5. Period of sampling. Age groups can be clear-cut only if the whole sample is taken over a period of time that is short relative to the life span of the animal, or, more strictly, relative to the time between hatchings and seasons of parturition. If the sample is evenly distributed or taken at random over a longer period of time, the concentration of the sample at definite parts of the life cycle will not be likely to occur; and hence age groups will not be distinguishable. Sampling of recent animals can be controlled to any degree of accuracy in this respect. Even among fossil animals it will be shown in a later paragraph that seasonal sampling does occur.

Given a sample that is known or may reasonably be assumed to be homogeneous except in age, age groups may be recognized by biological or by statistical methods or by a combination of the two. The appropriate statistical method is obtaining frequency distributions for characters that change with age and then proceeding to observe or to test the normality, modality, and other characters of these distributions that show them to include different groups. The most obvious and generally the

most useful characters are linear dimensions involved in growth; and if age groups are recognizable at all, distributions of these dimensions will generally show distinct bi- or polymodality. The fact that growth is usually differential may also be used by obtaining frequency distributions of appropriate ratios. Thus in salamanders the tail is often smaller relative to the body in the young than in adults, so that the tail:body ratio may give clear age groups. Similarly in mammals the head generally grows more slowly than does the body, the legs may elongate more slowly than they become stouter, etc. If age groups exist, testing of all the possibly important characters of this sort will almost inevitably reveal and define them.

Biological age grouping depends on the recording for each specimen of a definite character that appears at approximately the same time in the life of any individual of the species. Such groups are not absolutely valid temporally, for there are no such characters that appear at exactly the same epoch in every individual; but many do distinguish years or even shorter periods. There is also the distinction that they do not depend, as do recognizable statistical age groups, on an actual discontinuity in the appearance of young individuals, the criterion not being such a discontinuity in group appearance but a milestone that every individual passes, at whatever season. When clear-cut, however, the biological groups do usually serve to distinguish groups that are also statistically valid and often permit the valid separation of the components of a polymodal frequency distribution.

Most obvious of biological age criteria, for recent animals, is the attainment of sexual maturity. More exact but less often available are periodic phenomena that leave records of their occurrence, such as fluctuations in rate of growth of fish scales or of elephant tusks. Frequently these are annual but cannot be assumed a priori to be so, each case demanding investigation of this point. With the period of fluctuation determined, the age of the animal can be closely determined by counting the alternations of slowly and rapidly deposited tissue, just as is done in determining the age of a tree by the rings in its wood. Among mammals the loss of deciduous and eruption of permanent teeth are also periodic phenomena that occur at approximately the same rate for all individuals of a species. The fusion of

epiphyses and sutures in mammalian skeletons also serves to distinguish young and mature individuals.

An excellent example of both statistical and biological methods and of their combination in determining age groups is given by Blanchard and Blanchard (1931) for the salamander, *Hemidactylium scutatum*. Their data for total length of males are reproduced in Fig. 45. In the histogram of all their observations, the distribution falls into two sharply distinguished groups, here lettered A and B.[1] A, including the smaller individuals, is well defined and seems evidently to comprise a homogeneous group. B, on the other hand, is clearly heterogeneous and has at least two modes. This group includes individuals both sexually mature and immature. On this biological criterion, a group B_1 can be dissected out that appears to be itself homogeneous. The remaining group B_2, including all the sexually mature males, gives a frequency distribution that is no longer clearly bimodal although it is skewed and tails off to the left in a way that suggests some heterogeneity.

The authors show that the characteristics of the whole distribution can be quite precisely related to biological facts. A, the numerous and sharply defined group of small sexually immature males, consists of individuals in their first year. The gap between this and B represents the growth made by the next preceding generation during the winter months before the appearance of the generation seen in A, a time when no salamanders were being born. Group B_1, large sexually immature males, includes individuals in their second year. In a quantitative sample, B_1 includes fewer individuals than A because of high mortality before reaching the second year and fewer than B_2 because the latter includes several generations and B_1 only one. It is well separated from A because of rapid growth in the first two years but overlaps B_2 because of slower growth thereafter, so that large variants in the second year are as long as small variants in the third year. B_2 includes the sexually mature males, in their third year and older. It is broad, skewed, and very obscurely polymodal because animals of more than one year are included; but growth is now so slow that yearly age groups can no longer be clearly separated.

The essential parameters of these groups are as follows:[2]

Age group	N	R	M	σ	V
A	81	25.8–38.3	31.93 ± .33	2.98 ± .23	9.3 ± .7
B	117	45.3–76.6	62.15 ± .75	8.10 ± .53	13.0 ± .9
B_1	24	45.3–56.8	49.34 ± .65	3.20 ± .46	6.5 ± .9
B_2	93	49.8–76.6	62.12 ± .64	6.12 ± .45	9.4 ± .7

[1] The Blanchards do not give their combined data in this form.

[2] Calculated from the grouped data given by the Blanchards as a histogram. The less extensive calculated data that they give do not agree precisely with these, presumably because they were based on raw measurements or on a different grouping.

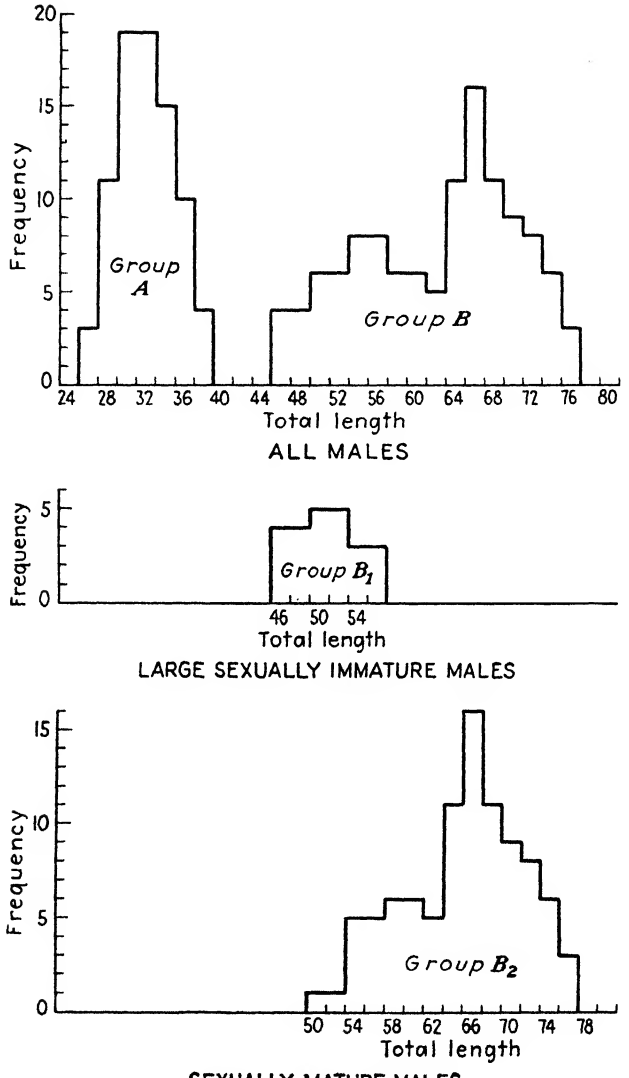


FIG. 45.—Age groups distinguished by discontinuous distribution of size and by physiology. Males of the salamander, *Hemidactylium scutatum*. The upper histogram shows the distribution of total length in the whole sample. Group A is manifestly distinct from Group B and was found to include individuals in their first year. B is clearly bi- or polymodal but cannot be clearly split mathematically. Physiological condition, recorded with the observations, permits the separation of B_1, individuals in their second year, and B_2, individuals older than two years. B_2 is heterogeneous but cannot be further subdivided on these data. (*Data from Blanchard and Blanchard* 1931.)

The variability of first-year individuals (A) is unexpectedly high, and the lower figure for the second year (B_1) suggests the advisability of investigating the significance and cause of this apparent change.[1] The variability of each of the separable components of B is notably lower than that for the whole group, and the greater variability of B_2 over B_1 agrees with the evidence that the former is still heterogeneous.

If sexual maturity is not determinable or is reached in the period from one breeding period to the next, such samples will generally be divisible only into two groups, a relatively well-defined immature group, born at the last hatching or bearing epoch, and a large heterogeneous but practically indissectible group of older individuals. This is, in fact, the usual condition in collections of many species of vertebrates.

In dealing with fossils, even though individual age can often be approximately determined, it is rare to find natural aggregation into age groups. The season of death (corresponding to date of collection of recent animals) cannot be controlled or determined. Many, perhaps most, fossil deposits include animals killed at various different seasons, so that age groups are not so sampled as to be kept distinct. Nevertheless it must commonly occur that in a given deposit animals were entombed more frequently or exclusively at one time of the year rather than any other, for instance, because of winter mortality, spring floods, summer droughts, or seasonal migrations to pastures and watering places. In such cases, definite age grouping may also be expected among fossils, and probably numerous examples will appear when paleontologists have become more inclined to look for and able to recognize such groupings.

One of the few examples yet recorded is that of the extinct horse *Merychippus primus* from the Sheep Creek beds of Nebraska. Matthew (1924) has shown that a large sample of this species can be separated on the basis of eruption and wear of teeth into four quite distinct groups of young and adolescent animals, with no intergradations between the groups, and an intergrading adult group perhaps separable into two to four additional age groups. Matthew's interpretation was that the younger, more clear-cut groups represent animals in their first, second, third, and fourth years and the more heterogeneous adult group, animals in their fifth year and later. He further inferred that parturition must have been seasonal in this species and that entombment of the animals must have occurred chiefly or wholly at one season, conclusions inescapable on the data presented. The reasons

[1] It will probably be found to be related to rapid growth, which usually produces high apparent variability.

for such seasonal fossilization are of course more obscure. Matthew suggested that entombment was in the dry season, either of one year or recurrently. In the famous deposit of complete skeletons of the little camel *Stenomylus*, also in Nebraska, similar age grouping has also been noted by Loomis, who considered the seasonal sampling the result of a flood, while Brown supposed it to have resulted from the rigors of winter. In any case occurrences show that age groups do occur in samples of fossil animals and therefore that seasonal sampling does exist in such deposits, whatever its cause.

SAMPLING FOR GROWTH

One of the most important and difficult elements in studies of growth is that of sampling. The technique of sampling for such studies can involve the following procedures, in their approximate order of accuracy and desirability:

1. Measuring parts affected by growth in each member of a group of live animals of known ages successively as each reaches certain fixed and preferably equidistant ages. The group should include the same individuals throughout, as far as possible, and the individual records should be kept distinct. This is the only method that gives adequate and complete information on growth. It gives a series of individual records and also a series of means from which population values can be estimated. It also shows the variation at various ages and possible changes in variability correlated with age and growth. As far as possible, factors other than age should be held constant, and all that might influence growth (*e.g.*, food, temperature, humidity) should be recorded and their bearing on the matter tested by partial correlation, association, or some analogous process. Generally this ideal technique can be applied only to laboratory animals, and then it has the disadvantage that it cannot be directly related to what actually happens in a wild population: it indicates only a sort of limit to what could happen under conditions more constant than those under which the animals really live. It is occasionally possible to sample wild animals in this way with groups that can be tagged or otherwise surely identified, that tolerate handling, and that remain available over considerable periods of time. It can, for instance, often be applied to very young or to relatively sedentary birds.

2. Measuring a single individual in the same way.

3. Measuring successive random samples from a population of known age. It is preferable to make live measurements and to return the animals to the population (laboratory or wild), for this leaves the succeeding samples unbiased. In this case each sample should include as large a percentage of the population as possible. Individual records have no value as such (unless as in method 1 the individuals are repeatedly observed and exactly identified); but the means and the dispersion are still useful, although less exact and less comparable than for method 1. It may, however, be impractical to take live samples of wild animals, and in any case the animals must

be killed if the growth of some element not measurable on live animals is being studied. In this case every sample after the first is inevitably biased. For instance, if the mean of the first sample happened to be above the population mean, the next sample will tend to be below the population mean, etc. Successive dead samples from one population cannot be truly random. Effects of this bias must be closely checked. For instance, if there were really no growth at all but if the mean of the first 5 samples of 10 happened to be below the mean for all 10 (and it will surely not be exactly the same and is as likely to be below as above), then the mean for the second 5 samples will certainly be above that for the first 5 and an incautious worker will conclude that growth has occurred.[1] This inevitable bias is minimized by making the samples as small as possible in relation to the population and yet large enough to reveal significant changes if they occur. The best balance between these two points depends on the peculiar conditions of each problem. It is a common laboratory technique to use up the whole population in successive samples, but this makes the later samples extremely biased. All the samples together should be only a small fraction of the available population, preferably not over one-tenth. Thus, if growth in laboratory animals is to be studied from 10 successive dead samples of 10 specimens each, about as small a series of observations as will give really reliable results, the samples should be drawn at random from a homogeneous laboratory stock of at least 1,000 animals. This counsel of perfection is almost never followed, but if it cannot be followed workers should be more hesitant in stating their results than they usually are. Sampling of wild populations should also be most carefully planned and scrutinized to be sure that the sampling itself is biasing the population as little as possible. Dead samples of single individuals for growth studies are relatively unreliable but are sometimes the best that can be done. The unreliable nature of the results must be fully appreciated. Finally, animals that belong to the population but that die while the sampling is in progress should be measured, if possible, and compared with the samples to determine whether their removal has significantly biased the population. If the mortality is high and the population small, this effect may be very appreciable, especially as the animals that die often average distinctly below the mean size of the population, so that if they are not taken into account growth will appear to be more rapid or more significant than it really is.

4. Successive samples of approximately known mean age but inexactly known individual age can be taken and studied as in method 3. The means obtained in this way for good samples will be about as reliable as those from method 3, but the indicated variation will have less significance, for it is here affected by growth as well as by variation, strictly speaking. This is usually the only or the best method available for wild animals, to which it can be applied when they have a fairly limited and known hatching or breeding period.

[1] This is undoubtedly one of several falsifying factors that have led certain observers to believe that the crowns of mammalian teeth grow after eruption.

5. Successive age groups in a large sample collected over a short period of time can be compared. With well-defined age groups the method is reasonably reliable. Its disadvantage is that the number of such groups, and hence the number of growth stages available, is almost always small, seldom more than two or three, which severely limits the information as to growth as a continuous process. Caution must also be used in comparing the younger groups with a combined adult group. The former, if recognizable at all, are usually well defined, often representing animals born in one year; and their mean size and mean age have real and reliable meanings. The adult group generally covers animals with much greater age differences, and the mean size is not a real and reliable measure of animals of any particular single mean age.

6. Age groups separated physiologically or anatomically from samples covering any period of time or, as with fossils, an indefinite period of time can be compared. This and, to much less extent, the preceding method are the only ones so far mentioned that are applicable to fossils. This method is inaccurate because the age criteria used are themselves variable as to age of appearance, because numerical values can seldom be given to the mean ages of the groups, because each group generally does not tend to have its values clustered around a mean, and because the successive groups usually do not cover equal or measurably related periods of time. Rate of growth cannot be studied from such data, and the results are generally somewhat indefinite, but some valuable information can be gathered.

7. It is occasionally possible to get some idea of growth trends from a sample that cannot be or is not sharply subdivided in any way correlated with age. Thus in a large sample of animals of unknown age and without sharply defined (discontinuous) physiological age criteria, it may yet be valid to assume that total length, or some analogous measurement, is closely correlated with age. If other measurements are then plotted against this, an idea, necessarily inaccurate but often suggestive, of their growth trend can be obtained. This method also does not give any measure of actual rates of growth, for that of the measurement used as a basis of comparison is not known. It may also give fallacious results because it starts out by assuming the truth of something that it seeks to prove, that is, that a measurement is correlated with age. As will be shown in detail, such data are invaluable in studies of relative growth, but they are of relatively little use in the study of absolute growth.

ABSOLUTE GROWTH AND GROWTH RATES

Without attempting an exhaustive discussion, some aspects of numerical work on growth may be considered, assuming that a good series of means or a good individual series of measurements at known ages is at hand. The principal elements in growth are:

1. Initial size.

2. Proportion of tissues with different growth characteristics. This may be so complicated that it is usual to assume that the whole of the material included in a given dimension grows in the same way. If this does not

give a good result, it may be assumed that a certain proportion of it does not grow at all and that the rest all grows in one way. It is probable, in some cases at least, that either of these assumptions is unreal and oversimplifies the real conditions; but they are proper if they approximate the observed results and reduce them to a basis that can be handled numerically without unreasonable labor.

3. Initial growth rate (of the whole or of the assumed growing proportion).
4. Changes in growth rate.

Growth can be completely described and fully studied if numerical values can be given to each of these four factors. Omission of any one of them makes the study incomplete and to that extent unreliable. The problem is essentially one of trend and can be discussed in graphic terms and represented by graphic methods. Factor 1 determines the starting point of the growth line and hence its absolute position on a graph. Factor 2 can seldom be measured directly and often must perforce be ignored; that is, it is usually assumed that growth is evenly distributed over the material between the limits of the dimension. If there are marked changes in growth rate that do not follow a defined pattern or formula, they can sometimes be brought into such a pattern by assuming the existence of a certain proportion of nongrowing material. This assumption and the consequent indirect measurement of factor 2 are largely numerical assumptions. They systematize and perhaps simplify study but are not necessarily very closely related to the real but obscure facts.

Factor 3 determines the initial slope of the growth line and hence in conjunction with the next factor determines the absolute slope throughout. Factor 4 determines the relative slopes of the growth line in its various parts. If there is no change in rate, the line is straight, and its initial slope is maintained throughout. This is almost never true for any growth line covering the whole life span of an animal or even for any large part of it. The principal and most difficult problem in the study of growth is that of evaluating this factor.

The concept of growth rate requires further specification. There are, in the first place, two quite distinct sorts of growth. In some cases, growth is a simple accretive or additive process. To the initial magnitude of the organ is added an increment which does not itself grow. The relationship is then given by the equation

$$Y_t = Y_0 + At$$

where Y_0 = the initial size of the organ.
A = the additive growth rate.
t = the time elapsed.
Y_t = the size of the organ at the end of this time.

If A is constant, then the growth line is straight, and A is simply its slope. It is then exactly analogous to a regression coefficient and can be calculated in the same way, fitting a straight line to the growth graph by the method of least squares (see Chap XIII). If A varies noticeably, which is usually the case if any considerable part of the individual's life is studied, then an estimate of the additive growth rate, a form of average, can be made for each interval between successive observations by the simple relationship

$$A = \frac{Y_t - Y_0}{t}$$

(Y_0 being taken as the value, or mean value, at the last time of observation and t as the duration of the interval between observations).

Among vertebrates, additive growth is mostly confined to organs like scales, feathers, hair, claws, teeth, and some types of horns. It is also found in concretionary structures such as otoliths or calculi. Among invertebrates, additive growth is more important because most of their hard parts, notably the shells of molluscs, tend to grow in this way. Most organs, however, do not have additive growth; nor does the body as a whole.

The other and more common type of growth is multiplicative, sometimes contrasted as compound-interest as opposed to simple-interest (additive) growth. In multiplicative growth the increment is itself alive and begins to grow as soon as produced. The increase is thus not by arithmetic addition, as in additive growth, but is by multiplication in a geometric series. The increment is not to be considered as a lump amount independent of the amount of material already in the organ but as dependent on the latter. The rate is not a fixed increment per unit of time but a percentage of preceding size per unit of time. This may be called the geometric as opposed to the arithmetic or additive growth rate. Thus if an organ at first weighs 50 grams and has increased to 60 grams 10 days later, its average absolute increment and its additive rate of growth are 1 gram per day, and its geometric increment is 20 per cent.

It is necessary here to distinguish between geometric increment and geometric rate of growth. The two are usually confused but are quite distinct. The organ that grows from 50 to 60 grams in 10 days has increased in weight by 20 per cent, its geometric increment, but it has not been growing at 20 per cent per 10 days nor at 2 per cent per day. Supposing, for the moment, that the increment is added instantaneously at the end of each day, the result of a geometric, compound-interest rate of 2 per cent per day is shown by the following series:

Days	Weight	Increment (2 % of weight at beginning of day)
0	50	1.00
1	51	1.02
2	52.02	1.0404
3	53.0604	1.061208
4	54.121608	1.08243216
5	55.20404016	1.1040808032
6	56.3081209632	1.126162419264
7	57.434283382464	1.14868566764928
8	58.58296905011328	1.1716593810022656
9	59.7546284311155456	1.19509256822310912
10	60.94972099737856512	

Discontinuous geometric growth at 2 per cent per day would thus cause 50 grams to increase to nearly 61, not to 60, in 10 days. This series is analogous to interest of 2 per cent per day compounded daily. In fact in such growth the compounding is not daily or at any other fixed discontinuous points but is continuous. The interest, so to speak, is continuously due, is paid instantaneously, and immediately becomes capital and starts paying interest itself. Such growth is represented by the equation

$$Y_t = Y_0 e^{Gt}$$

where Y_0 = initial magnitude.
Y_t = magnitude at time t, as before.
e is the base of Napierian logarithms.[1]

[1] Most succinctly so designated, but it may be noted here and is true in most other equations in which this constant appears that e enters into the equation not because it is the base of these logarithms but because it is the

G is the geometric or multiplicative growth rate, sometimes also called the "instantaneous growth rate" because it implies instantaneous compounding of the increments.

The practical determination of G demands converting the equation into logarithmic form. Using Napierian (natural) logarithms, it becomes

$$\log_e Y_t = \log_e Y_0 + Gt \log_e e$$

and since $\log_e e$ is, of course, 1, this is

$$\log_e Y_t = \log_e Y_0 + Gt$$

whence

$$G = \frac{\log_e Y_t - \log_e Y_0}{t}$$

If common logarithms (base 10) are used

$$\log Y_t = \log Y_0 + Gt \log e$$

and since $\log e$ (*i.e.*, $\log_{10} e$) is .434294, this is

$$\log Y_t = \log Y_0 + .434294\, Gt$$

whence

$$G = \frac{\log Y_t - \log Y_0}{.434294t}$$

or

$$G = 2.302585 \frac{\log Y_t - \log Y_0}{t}$$

There are four variables here; so any one can be found if the observational data supply the other three. Thus 50 grams growing at the geometric rate 2 per cent or .02 per day, compounded instantaneously, would grow to about 61.01 grams in 10 days, as shown by the following:

$Y_0 = 50$ $\quad \log_e Y_0 = 3.91202$
$G = .02$ per day $\quad t = 10$ days $\quad Gt = .2$
$\log_e Y_t = \log_e Y_0 + Gt = 3.91202 + .2 = 4.11202$
$Y_t = \text{antilog}_e Y_t = \text{antilog}_e 4.11202 = \text{about } 61.01$

---

sum (more strictly the limit of the sum) of the series

$$1 + \frac{1}{1} + \frac{1}{1 \times 2} + \frac{1}{1 \times 2 \times 3} + \frac{1}{1 \times 2 \times 3 \times 4} + \cdots$$

This series enters naturally into summation of a continuous geometric series, and there is nothing arbitrary or esoteric about the appearance of e in equations related to such series.

If G is constant, the growth line will be straight on semilogarithmic coordinates, plotting Y logarithmically and t arithmetically. G is the regression coefficient of log Y on t; and if G is constant, a least-squares line can be fitted and G calculated like b_{YX} (page 266), using log Y in place of the observed values throughout. G is approximately constant over short periods and often for longer periods in embryonic or earliest postnatal life, but it is seldom even roughly constant throughout life. It almost invariably drops as the organ in question nears its definitive size; and even if growth is continuous throughout life, the rate of growth becomes less. Constant multiplicative growth leads eventually to such enormous increments that the slowing down of growth is almost inevitable. For instance, if an animal weighed 10 pounds at birth and grew steadily at 1 per cent per day, in 10 years it would weigh more than 70,000,000,000,000,000 pounds! Yet most animals grow faster than 1 per cent per day in early stages of growth.

The value of G obtained by the procedure just explained is the value that would have effected the observed amount of growth in the given time if G had been constant. If the growth rate was approximately constant during that period, then G is a good estimate. If it was not constant, then G is to be considered as a useful form of average. Note also that G is in any case an actual value that the growth rate did have at some time during the period under consideration, even though the exact time cannot easily be specified if G is changing.

Change in growth rate can be most usefully represented by determining G for each of the successive time periods of observation. This gives not only the average rate for each of the successive periods, but also a series of actual rates, and so well represents the changes in rate that really occurred.

It is possible for the multiplicative growth rate to decrease as growth proceeds (as Y becomes larger) in such a way that $Y(e^{Gt} - 1)/t$, which is the mean increment for any given period, is constant. Then equal increments will be added in equal periods of time, and as far as any external manifestation is concerned growth will appear to be additive, not multiplicative. It is clear, however, that this is in reality only a special case of multiplicative growth; and it is preferable to treat it as such rather than as additive growth.

EXAMPLE 84.—TABULAR RECORDS OF GROWTH AND (Data from Jen-

Stage	Age, days from spawning	Elapsed time, days	No. of individuals	Mean length of head, mm.	Absolute increments, mm.	
					For interval	To date
I	49	0	200	36.1	——	——
II	63	14	209	53.4	17.3	17.3
III	77	28	198	68.1	14.7	32.0
IV	92	43	200	79.3	11.2	43.2
V	106	57	161	87.3	8.0	51.2

[1] These should not ordinarily be used, for they are confusing. They are often mistaken values of the latter.

It is thus possible to study growth from several different sorts of data, choice among which is to be determined by the kind of growth in question and by the purpose of the study. Most important among these are:

1. The data as observed—necessarily corresponding observations of magnitude Y and elapsed time t, with such other raw data as may be pertinent (number and nature of specimens, absolute ages, etc.).
2. Arithmetic or absolute increments—$Y_t - Y_0$.
   a. For each step between successive observations.
   b. Cumulative, giving the total increment since the first observation.
3. Arithmetic increments per unit of time, which are the average arithmetic growth rates—$(Y_t - Y_0)/t$.
   a. For each step.
   b. Cumulative from the first observation.
4. Geometric increments—$(Y_t - Y_0)/Y_0$.
   a. For each step.
   b. Cumulative.
5. Average geometric growth rates—$(\log_e Y_t - \log_e Y_0)/t$ [not $(Y_t - Y_0)/tY_0$, which gives average geometric increments per unit of time but not the geometric growth rates].
   a. For each step.
   b. Cumulative.

These are compared for a typical set of growth data in Example 84, and several are shown graphically in Fig. 46.

GROWTH RATES. LENGTH OF HEAD IN RAINBOW TROUT
kinson 1912)

Stage	Arithmetic growth rates, mm. per day		Geometric increments in %		Geometric growth rates in % per day		Mean geometric increments per day in %[1]	
	For interval	To date	For interval	To date	For interval	To date	For interval	To date
I	——	——	——	——	——	——	——	——
II	1.24	1.24	47.9	47.9	2.80	2.80	3.42	3.42
III	1.05	1.14	27.5	88.6	1.74	2.27	1.96	3.16
IV	.75	1.00	16.4	120	1.02	1.83	1.09	2.79
V	.57	.90	10.1	142	.69	1.55	.72	2.49

for the average geometric growth rates and are given here only for comparison with the true

It would seldom be necessary or advisable to give the data in all these different forms in any one study. In this case, growth is surely multiplicative, and so the geometric growth rates are most directly and truly related to the problem.[1]

As usually seen, a growth curve (arithmetic plot of magnitudes against time) is convex upward, rising very steeply at first and gradually leveling off until it becomes horizontal, or nearly so, as the adult size is reached. If, however, the whole curve is available, from the fertilized egg through the embryonic as well as the juvenile growth, the early part is seen to be concave upward, beginning nearly horizontal and curving upward until it becomes steep and runs into the postnatal curve. The point of inflection, where the concavity changes from upward (and to the left) to downward (and to the right) is that at which the increment per unit time, the arithmetic growth rate, is greatest. Birth usually occurs at about this time, so that the earlier part of the curve, with increments increasing more and more rapidly, is mainly or wholly embryonic, while the later part, with increments decreasing more and more slowly, is mainly or wholly

[1] All these data would be more enlightening if there were more steps, observations at shorter intervals; but the example shows the general nature of the different treatments, and a more elaborate series would be too bulky for our purpose.

postnatal. Most growth observations cover only the postnatal period, and especially its earlier part.

The arithmetic growth rate begins essentially at zero, rises to a maximum at about the period of birth, then falls off and eventually reaches zero when the animal stops growing. This curve is

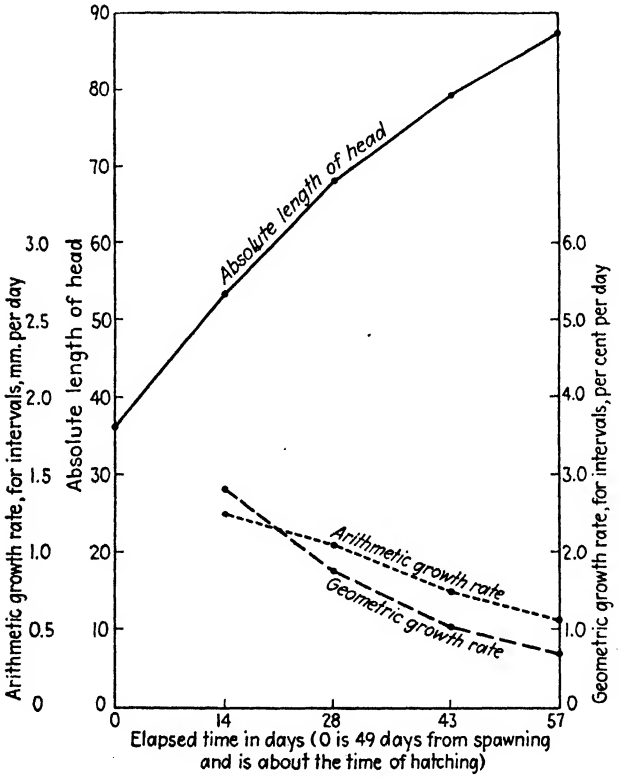


FIG. 46.—Growth curve and growth-rate curves. Length of head in rainbow trout (data of Example 84). (*Data from Jenkinson* 1912.)

usually strongly skewed because the period of accelerating arithmetic growth is generally much shorter than the period of decelerating growth, so that the peak of the curve is far to the left of its midpoint. The curve of the geometric growth rate usually follows a very different course. This rate is often very high in the earliest embryonic stages, then falls very rapidly and may level out so that the rate is relatively constant, or fluctuates about a nearly horizontal line of trend, through much of the

embryonic period. Near the time of maximum arithmetic growth, the geometric growth rate normally curves downward again and falls off, at first rapidly, then more and more slowly, until it finally reaches zero when maximum size is attained[1] (see Fig. 47).

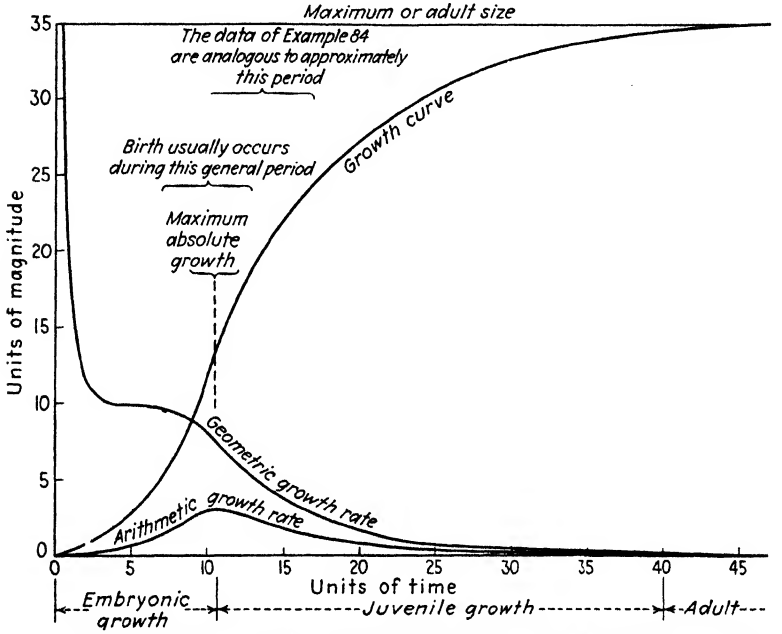


FIG. 47.—Idealized curves of growth and growth rates. Based on many sets of data but generalized and idealized. The relatively longer periods of maturity and senility are not included.

These relationships are subject to innumerable variations and modifications. Growth may proceed by cycles instead of in a single continuous curve. The rates are commonly highly irregular over short periods and show an even trend only over longer intervals. Some organs (like the thymus) may have a very different sort of growth pattern. These modifications cannot be considered here, where the concern is only with the general methods of study and not with its results. It is also often pos-

[1] The very earliest stages of embryonic growth are inadequately known, but there is some evidence that the geometric rate is then usually at its maximum. Embryonic growth after the earliest stages often approximates a logarithmic curve, indicating a relatively constant geometric growth rate.

sible to fit a mathematical curve to growth data over longer or shorter periods, but relatively exact methods of doing this may be very complex and laborious and belong in the domain of the specialist and the skilled mathematician.[1]

PROPORTIONS AND SIZE

Even more important than changes of absolute dimensions with growth are changes in proportions. It is a well-known fact that such changes do occur, for instance, that young mammals have larger heads relative to their bodies than do adults of the same species. It is less well known but nevertheless has been clearly shown that the same sorts of changes may appear not only with age but with animals of the same age but of different sizes, that is, that the smaller adults, for instance, may tend to have relatively larger heads than do larger individuals of the same age and species. It has even been demonstrated in numerous instances that the same sort of relationship may appear between larger and smaller species of a genus or in rarer cases even between larger and smaller genera of a family. The relationship of proportions and size thus has far wider implications than are involved in growth alone, although what appears to be the, or certainly a, basic law of these relationships was discovered

[1] The growth of a population usually follows what is called the logistic curve, generalized as

$$Y = \frac{K}{1 + Ce^{a_1 t + a_2 t^2 + \cdots a_n t^n}}$$

(see Pearl, 1930, and the work cited by him). Anatomical growth commonly follows an analogous pattern. There are various ways of applying to growth data empirical curves such as

$$Y = a \log X + b$$
$$Y = aX + b(\log X + c) + d$$
$$Y = (aX + b) + b(\log X + c) + d$$

(see Donaldson, 1924, where these and others are applied to relative growth in the rat). Such empirical curves are descriptive devices difficult to use and of limited value in zoology, strictly speaking, since for the most part they are not formalizations of the logical procedures involved in research but only artificial approximations of observed trends. More useful are methods that formalize certain hypotheses concerning growth and facilitate comparisons of these with actual observations (see, for example, Robb, 1929, and papers there cited by him, where the hypothesis that growth proceeds as does an autocatalytic chemical reaction is tested in this way).

from growth studies and has hitherto found its most important applications in that field.

Relative proportions of two parts of an animal do not depend directly on their respective absolute growth rates. If their geometric growth rates are the same at all times, that is, maintain the ratio 1, then the relative sizes of the absolute dimensions will remain the same, no matter what those relative sizes are. Nor does it then matter how much or in what way the growth rates change in absolute value, so long as both change in the same way and they maintain their ratio. The size of one structure Y can then always be calculated from that of another X by the simple relationship

$$Y = a + bX$$

In this equation, a is mathematically the value of Y when X is zero. Zoologically X cannot be zero, but a nevertheless has a biological significance: it is a basic size difference between X and Y, unaffected by growth. Probably this represents a certain proportion of nongrowing tissue, or possibly in some cases it is a conventionalization of a proportion of more slowly growing tissue. The constant b is mathematically the slope of the regression line, or more strictly (if the line is not straight) the slope of a tangent to it when $X = 1$. Zoologically it represents a basic ratio of the absolute sizes of the increments of Y and X: it is equal to the ratio of their arithmetic growth rates.

In practice it has been found that this relationship and others similar to it are all special cases of the more general equation

$$Y = a + bX^k$$

The value of k in this expression is the ratio of the geometric growth rates of Y and X.[1] This equation and all the others developed from it by giving fixed values to a, b, and k assume that these three values are constant in any given case. Especially as regards k, this is not necessarily true, and the equation

$$Y = a + bX^k$$

must itself be considered a special case of a still more general

[1] This is not obvious algebraically but can easily be proved by simple calculus (see Huxley, 1932) and can be empirically checked by application of the formula to actual cases of growth that follow this pattern.

relationship in which one or two additional variables occur, corresponding with variable amounts of tissue involved in growth and with a variable ratio of the growth rates of Y and X. Such forms, however, are too complex for most practical purposes. It has, moreover, been found that a, b, and k are in fact usually constant in zoological practice, if not over the whole of growth, at least over such considerable periods of it as to make forms derived from $Y = a + bX^k$ valid for most purposes. This discovery has the most important and in part unexpected bearings on zoology and paleontology in general, and this equation will help to solve many of the most basic and most puzzling problems of these sciences.

Before suggesting the sort of evidence and of conclusions involved in this relationship, it is well to examine in more detail the mathematical and zoological properties and relationships of the equation. According to whether the various constants take the values 0, 1, or neither, it has eight possible forms:[1]

(1) $Y = X$	(5) $Y = X^k$
(2) $Y = bX$	(6) $Y = bX^k$
(3) $Y = a + X$	(7) $Y = a + X^k$
(4) $Y = a + bX$	(8) $Y = a + bX^k$

Of these the first four have the growth ratio k, equal to 1. That is, Y and X grow at the same geometric rate. The regression of Y on X is then a straight line on arithmetic coordinates, and in practice the applicability of one of these formulas can be recognized by observation of this graphic relationship. The last four formulas, on the other hand, apply when Y and X do not grow at the same geometric rate (k is not equal to 1), but their growth rates do retain a constant ratio. These regressions are curvilinear on arithmetic coordinates but are rectilinear on logarithmic coordinates (or when the logarithms of X and Y are plotted on arithmetic coordinates). Their properties may be considered separately.

1. $Y = X$. This is the simplest possible relationship in relative growth and proportions and indicates, of course, that the

[1] Also theoretically possible are four other forms:

$Y = 0$ $\quad Y = a$ $\quad Y = b$ $\quad Y = a + b$

—but these are not possible in practice, for they make Y invariable, whereas it is defined as a variable for present purposes.

two anatomical elements remain of the same size (Fig. 48A). This is normally true, for instance, of two bilaterally symmetrical parts, such as right and left femora. The relationship is so simple and obvious as to require no analysis. It should, however, be noted that this and the subsequent more complex forms are relationships of biological growth or relative size factors and are only approximations to the detailed actual facts of given cases. That is, right and left femora, for instance, are never exactly and absolutely of the same size in nature. In a group, each shows some variability about its mean, and the correlation is not quite 1.00, although extremely high. Thus $Y = X$ is only a close, valid approximation of their actual relationships, but it is probably an exact statement of their relative growth and size tendencies as aside from incidental minor fluctuations caused by extraneous or nongenetic factors.

The graph of this equation on arithmetic coordinates is a straight line extending up and to the right from the origin ($Y = 0$, $X = 0$) at an angle of 45 deg. (or with a slope of 1).

2. $Y = bX$. The geometric growth rates are the same, the ratio Y/X is constant and is equal to b, but Y is not equal to X, being smaller if b is less than 1 and larger if b is larger than 1 (see Fig. 48B). This is approximately true of many variates and is the general expression of harmonious growth, without change of proportions. Confidence in this relationship as usually true underlies the general use of ratios, instead of absolute dimensions, in comparing different individuals and groups as in taxonomy. The confidence is, however, frequently misplaced. More thorough study tends to show that the relations $Y = a + bX$ and $Y = bX^k$ are both more widespread than is $Y = bX$, except as a gross approximation, and in neither of them does the ratio Y/X tend to be constant. It is generally recognized that this is often true of individual growth, that proportions often change according to the age of an animal; but it is less generally realized that proportions are often a variable function of absolute size regardless of differences in individual age.

The graph on arithmetic coordinates is a straight line of any slope (b is the slope) passing through the origin $Y = 0$, $X = 0$).

3. $Y = a + X$. The geometric growth rates are the same ($k = 1$), and every increase in one dimension is accompanied by an equal absolute increase in the other ($b = 1$); but there is a

basic difference in absolute size (a) (see Fig. 48C). The ratio of Y to X is variable: $\frac{Y}{X} = \frac{a+X}{X} = \frac{a}{X} + 1$. If a is positive, this ratio is greater than 1 and becomes smaller as X becomes

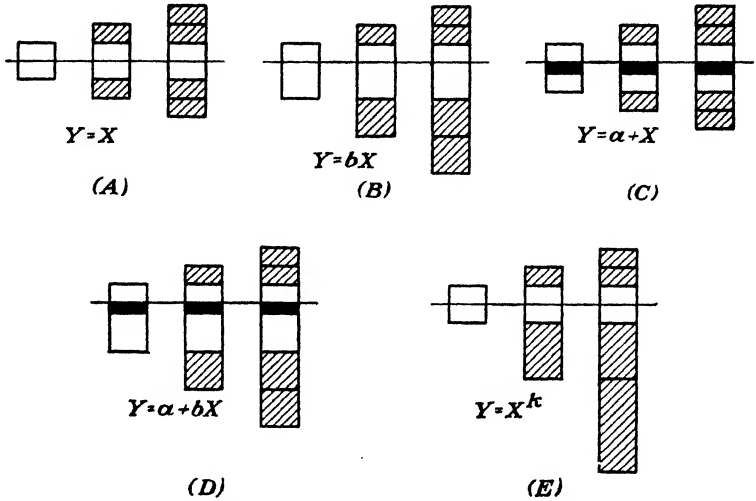


FIG. 48.—The four types of isogonic growth and the simplest case of heterogonic growth. In each diagram an initial size relation and two subsequent growth stages are shown, with the size of one variate (X) shown by the area above the base line and that of the other (Y) below. Increments of both are crosshatched. A, Y and X are initially equal and remain so, growing by equal increments; the arithmetic growth rates are the same, and the ratio of Y:X is constant and equal to 1. B, Y is initially twice as large as X and remains so, its increments being also twice as large ($b = 2$); the arithmetic growth rate is twice as large as that of X, and the ratio of Y:X is constant. C, Y is initially larger than X by a fixed amount a, represented in solid black, and grows by equal increments; the arithmetic growth rates are the same, but the ratio of Y:X is inconstant. D, Y is initially larger than X, and its increments are twice as large ($b = 2$); the arithmetic growth rate of Y maintains a constant ratio 2 to that of X, but the ratio of their absolute values is inconstant. A to D are the four basic types of isogonic relations in relative growth. E, Y is initially of the same size as X, and its geometric growth rate is constantly twice as great as that of X, making its increments progressively larger; the geometric growth rates maintain a constant ratio ($k = 2$ in the diagram), but the ratio of Y:X is inconstant. E is the simplest case of heterogonic relative growth. (*Based in part on a diagram by Robb* 1929.)

larger, approaching 1 as a limit. If a is negative, which can be and often is the case, the ratio Y/X is less than 1, becomes larger as X becomes larger, and also approaches 1 as a limit. This and the following equations illustrate the difference between zoological and mathematical possibilities. Because a mathe-

matical equation may usefully represent a zoological relationship, it does not follow that every correct mathematical solution of the equation is a zoological possibility. The inferences cannot be extended indefinitely beyond the field of observation, nor is any mathematical result to be accepted unless it is logical and meets the inherent nonmathematical postulates. In other words, the formulas apply as far, and only as far, as they are symbols of common-sense deductions.

In the use of these formulas in this section, it is in the first place understood that neither Y nor X can have a negative value. They represent measurements of anatomical variables; and these may be 0 in rare cases, that is, the element may disappear, but obviously cannot be less than 0. If the anatomical part is essential to life, it obviously cannot measure 0 but has a lower limit greater than 0, imposed physiologically, below which it is too small to function so that life ceases if it drops below this point. With less rigidity, it is also understood that Y is usually taken, by convention, as the smaller or less inclusive of the two anatomical variables. It is, for instance, to be used as tail length against X as total length or as weight of brain against X as total weight. In such cases the ratio Y/X obviously must be less than 1, and usually much less. There is a zoological upper limit of Y/X considerably lower than the mathematical limit. In some cases, however, it may be convenient to represent by Y a variable that can exceed X in size, for instance, tail length against X as body length. Then the ratio Y/X can be greater (or less) than 1, but its upper limit is still imposed physiologically: it is the point where the relative sizes are physiologically incompatible, for instance, where the tail becomes too large relative to the body to be nourished by the latter.

In the equation $Y = a + X$ if a is negative no value of X less than $(-a)$ is possible; for that would make Y negative, a zoological impossibility. Further, if Y represents a variate necessarily smaller than X, then a must be negative.

The graph on arithmetic coordinates is a straight line leaving the Y-axis at a, if a is positive, or the X-axis at $(-a)$ (*i.e.*, the positive arithmetic value of a), if a is negative, at 45°, slope 1, running to the right and upward.

This relationship is, in fact, rare and of little practical significance in zoology but does occur, at least approximately, and has

been treated at this length because it shows most clearly the meaning of a and the limitations of the zoological, as opposed to the mathematical, facts. The equation applies when the growth or increment rates of Y and X are constant and equal but their ratio is not. The calculation of a is obvious: it is simply $Y - X$ for corresponding values of those variates or, more accurately, for their mean values.

4. $Y = a + bX$. This relationship is true of relative size and growth when the geometric growth rates of Y and X are equal ($k = 1$) but the increments are unequal (b is not 1) and the ratio Y/X is not constant (see Fig. 48D). This is one of the two most important expressions of relative size and growth (the other being $Y = bX^k$) and is very commonly a good approximation of the facts. The equation is the same as that for rectilinear regression in general, and the calculation of a and b has been amply explained (Chap. XIII). The graph is as for form 2 of the equation except that the line may run in any direction with slope either greater or less than 1 (its angle greater or less than 45 deg.), as given by the value of b.

Several possible alternatives demand consideration, depending on possible ranges of a and b. In the first place, a and b may both be positive, or either one may be negative, but both cannot be negative (for then all mathematically possible values of Y would be negative, and this is zoologically impossible). The other important alternatives may be considered systematically:

A. $b > 0$. Y becomes larger as X becomes larger. In other words, regression and correlation are positive, and the line runs upward to the right.

a. $a > 0$. Y/X is necessarily greater than b and becomes smaller as X increases. Zoologically, Y cannot have any value less than a (because X would then be negative)

(1) $b > 1$. Y is always larger than X. Therefore this cannot be the case if, as by usual convention, Y is taken zoologically as an element necessarily smaller than X.

(2) $1 > b > 0$. Y is equal to X when $X = a/(1 - b)$, is greater than X when X is less than $a/(1 - b)$, and is less than X when X is greater than $a/(1 - b)$. $a/(1 - b)$ is, then, the lower zoological limit of X if Y must be smaller than X.

b. $a < 0$. Y/X is necessarily smaller than b and becomes larger as X increases. Zoologically X is always greater than $-a/b$; for if it were less, Y would be negative.

(1) $b > 1$. Y is smaller than X when X is smaller than $a/(1 - b)$, equal to X when $X = a/(1 - b)$, and greater than X when X is greater than $a/(1 - b)$. $a/(1 - b)$ is the upper zoological limit of X if Y must be smaller than X.

(2) $1 > b > 0$. Y is always smaller than X, and there is no obvious upper limit to the value of X. This is probably the most common rectilinear relationship in growth or between the magnitudes of two anatomical variates.

B. $b < 0$. Y becomes smaller as X becomes larger; regression and correlation are negative, and the line runs downward to the right. Such cases occur in zoology but are rare. It is very common for Y to become smaller relative to X as X becomes larger, *i.e.*, for Y/X to decrease with increase of X, but very uncommon for Y to become absolutely smaller as X increases.

a. $a > 0$. Y/X is greater than $(-b)$ and becomes smaller as X increases. X must be smaller than $a/-b$, and Y must be greater than a, for otherwise one or both would be negative. Y is greater than X when X is less than $a/(1 - b)$ and less than X when X is greater than $a/(1 - b)$. $a/(1 - b)$ is thus the lower zoological limit of X if Y must be less than X.

b. $b < 0$, $a < 0$, is a zoologically impossible combination.

5, 6. $Y = X^k$. $Y = bX^k$. These represent simple multiplicative growth such that the geometric rates of growth of Y and X maintain a constant ratio, k. The ratio Y/X is variable (Fig. 49) and is always b when $X = 1$. Y and X are equal ($Y/X = 1$) when $bX^{k-1} = 1$. $Y = X^k$ is a relatively unimportant special form when Y and X happen to be equal (both $= 1$) when $X = 1$. The value of b cannot be negative because that would make Y negative. There are the following special cases:

A. $k > 1$. Y increases as X increases and Y/X increases with them. That is, the two variables get larger together, but Y at a faster rate.

a. $b > 1$. Y is always larger than X. These conditions therefore cannot apply if Y is defined as the smaller variable. This sort of size relation is very rare in zoology.

b. $b < 1$. $Y = X$ when $X^{k-1} = 1/b$, which is thus the upper zoological limit of X if X must be larger than Y. This is a common and important relationship in zoology.

B. $1 > k > 0$. Y increases as X increases, but the ratio Y/X decreases as they increase. The two variables get larger together, but X at a faster rate. $Y = X$ when $X^{1-k} = b$, the zoological lower limit of X if Y must be smaller than X. This is also an important relative growth phenomenon in zoology.

C. $k < 0$. Y decreases as X increases and Y/X also decreases as X increases. If X^{1-k} is less than b, Y is larger than X; and this is therefore the lower zoological limit of X if Y must be smaller than X. Structures that become absolutely smaller as the body or any considerable part of it becomes absolutely larger are rare, so that this relationship is unusual in zoology, but it does occur.

For the graphs of these curves and the calculation of b and k, see below.

7, 8. $Y = a + X^k$. $Y = a + bX^k$. The theoretical importance of these equations is as great as for the preceding or is indeed greater, for $Y = a + bX^k$ is the generalization of all the other equations here given and often would give the best representation of the detailed facts. They have, however, little or no practical value at present; for no adequate tests or considerable use of them have been made, and the calculation of both a and k in the generalized equation is usually impractical. It is probable, however, that some of the instances of relative growth that do not follow either $Y = a + bX$ or $Y = bX^k$ and so have not yet been reduced to rule and analyzed would be found to follow $Y = a + bX^k$ were the values of all three constants determinable.

In practice, relative sizes, proportions, and relative growth can usually be approximated by either $Y = a + bX$ or $Y = bX^k$. The first is the equation to use if the regression is approximately rectilinear on arithmetic coordinates and the second if this is true only on logarithmic coordinates. Ordinary linear regression

has already been amply studied, but further notice of $Y = bX^k$ is necessary.

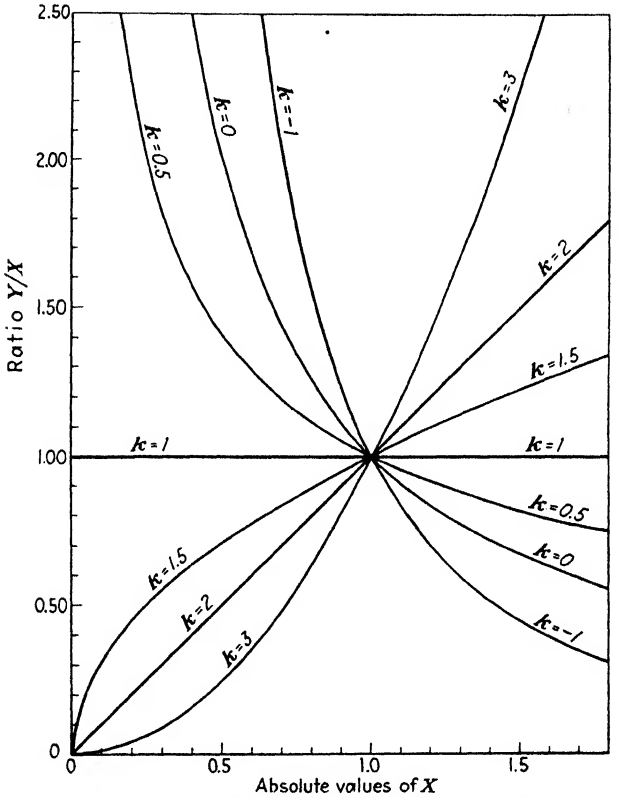


FIG. 49.—Dependence of ratios on absolute size and coefficients of heterogony. The ratio $Y:X$ is plotted against absolute values of X for different values of k in the equation $Y = bX^k$. When $k = 1$, the ratio $Y:X$ is constant. When k is less than 1, the ratio becomes at first rapidly and then more slowly less as X (the gross size of an animal, for instance) becomes greater. When k is greater than 1 and less than 2, the ratio becomes at first rapidly and then more slowly larger as X increases. When k is greater than 2, the ratio becomes at first slowly and then more rapidly larger as X increases. For simplicity, b has been taken as 1 in this diagram, but the forms of the curves would be the same for any values of b, only their positions on the coordinates shifting. By making such a diagram for any values of b and k, the nature of the shift in ratios with changes in size can be clearly shown for any given case.

Plotting this on logarithmic coordinates or plotting it on arithmetic coordinates by the logarithms of Y and X is equivalent to plotting the derived equation

$$\log Y = \log b + k \log X$$

This is an ordinary rectilinear regression equation and shows why the curve becomes a straight line in logarithmic plotting. That is, log Y is now being used instead of Y, log b is a constant and could equally well be symbolized as a, k is also a constant taking the place of b in the regression equation, and log X is

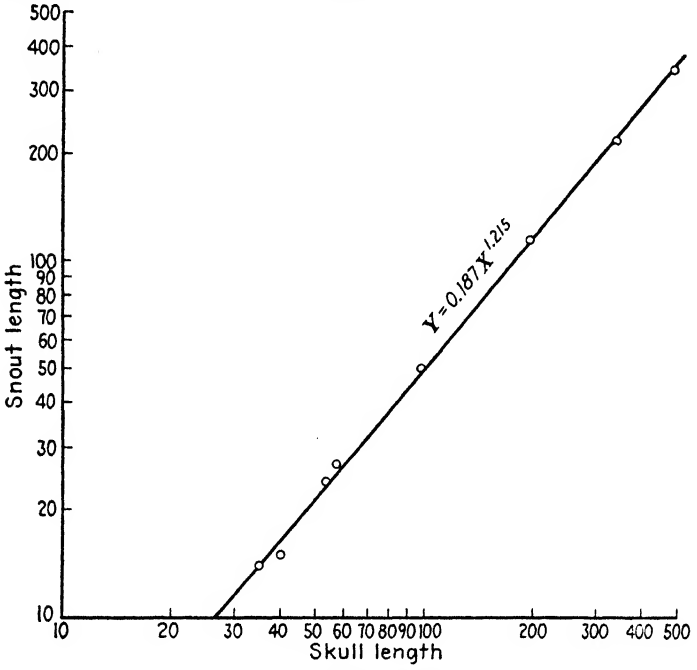


FIG. 50.—Logarithmic graph of observed and theoretical growth. Snout length against total skull length in *Alligator mississippiensis* (data of Example 85). The nearly rectilinear arrangement of the observations on logarithmic coordinates shows approximation to the form $Y = bX^k$, and the straight line representing the values $b = .187$, $k = 1.215$, very closely approximates the observed trend.

used for X; so this equation has the same form as $Y = a + bX$. It shows that k is the coefficient of regression of the logarithm of Y on the logarithm of X. If the original data, values of Y and X, be replaced by their logarithms, it is a relatively simple matter to fit a straight line to the logarithmic regression by the method of least squares, calculating the value symbolized in Chap. XIII as b_{YX}, which is then equal to the k of the present problem. log b can then be calculated as a was in Chap. XIII, and b taken as the antilog of this value.

EXAMPLE 85.—CALCULATION OF $Y = bX^k$ FOR REGRESSION OF SNOUT LENGTH ON SKULL LENGTH IN *Alligator mississippiensis*
(Data from Mook 1921)[1]

Skull length (X)	Snout length (Y)	Logarithms					
		X	Y	d_X	d_Y	$d_X d_Y$	d_X^2
35	14	1.544	1.146	−.471	−.574	.270354	.221841
40	15	1.602	1.176	−.413	−.544	.224672	.170569
53	24	1.724	1.380	−.291	−.340	.098940	.084681
57	27	1.756	1.431	−.259	−.289	.074851	.067081
97	50	1.987	1.699	−.028	−.021	.000588	.000784
194	115	2.288	2.051	+.273	+.341	.093093	.074259
339	217	2.530	2.336	+.515	+.616	.317240	.265225
488	340	2.688	2.531	+.673	+.811	.545803	.452929
		16.119	13.760			1.625541	1.337639

$M_{\log X} = 2.015 \qquad M_{\log Y} = 1.720$
$k = 1.625541/1.337639 = 1.215$
The line passes through the means of the logarithms; so

$$1.720 = \log b + 1.215 \times 2.015$$

whence $\log b = 1.720 - 2.448 = 9.272 - 10$
$b = .187$
The equation sought is thus $Y = .187X^{1.215}$.

If desired, theoretical values of Y corresponding to the observed values of X can then be calculated as follows. These define points on the smoothed logarithmic curve.

1.215 (log X)	With $\log b = 9.272 - 10$ added	Antilogs
1.876	1.148	14.1
1.946	1.218	16.5
2.095	1.367	23.3
2.134	1.406	25.5
2.414	1.686	48.5
2.780	2.052	112.7
3.074	2.346	221.8
3.266	2.538	345.1

[1] It is to be noted that these data are very inadequate, for they represent measurements of eight skulls only (of course, of different individuals). They are nevertheless used because they provide a numerically simple and brief example, our interest here being method and not results, and to show that even such poor data, analogous to many paleontological samples, give a surprisingly good result.

EXAMPLE 85.—CALCULATION OF $Y = bX^k$ FOR REGRESSION OF SNOUT LENGTH ON SKULL LENGTH IN *Alligator mississippiensis* (Data from Mook).—(*Continued*)

In this case, k is significantly greater than 1. If it were assumed to be 1, the rectilinear regression equation would be

$$Y = .711X - 15.9$$

Theoretical values of Y can also be calculated from this equation and contrasted as follows:

Observed values of Y	Theoretical values of Y			
	$Y = .711X - 15.9$		$Y = .187X^{1.215}$	
		d		d
14	9.0	+5.0	14.1	− .1
15	12.5	+2.5	16.5	−1.5
24	21.8	+2.2	23.3	+ .7
27	24.6	+2.4	25.5	+1.5
50	53.1	−3.1	48.5	+1.5
115	122	−7	113	+2
217	225	−8	222	−5
340	331	+9	345	−5

It is obvious that the values calculated from $Y = .187X^{1.215}$ agree much more closely with the observed facts; they are in every case closer to those actually observed than are the values from the other equation. The fact is that the length of the snout (preorbital part) of an alligator skull does grow faster than the length of the skull as a whole and that, within the period observed, its growth rate maintains an approximately constant ratio to that of the skull as a whole.

The work of calculating these constants and obtaining a relative size equation is shown in Example 85 and Fig. 50 (see also Fig. 51).

There are other, simpler ways to calculate k and b depending on rougher methods of fitting a straight line to the logarithmic data. As these methods of line fitting are much less reliable than that of least squares, the resulting values of k and b are to be viewed only as rough estimates and used only when greater reliability is not desirable or is not possible, and then only with a definite indication as to how the values were calculated. A straight line is determined by any two points, so that any two pairs of values of Y and X may be taken and assumed to lie on the regression line. Then

$$k = \frac{\log Y_t - \log Y_0}{\log X_t - \log X_0}$$

where X_t = a higher value.

X_0 = a lower value of the independent variate.

Y_t and Y_0 = corresponding values of the dependent variate.

The value of k may thus be judged, very roughly, by any two pairs of values of Y and X. Obviously, however, the more observations averaged into these values, the better the estimate. It is also true, with some reservations, that the estimate is better the farther apart are the values X_t and X_0. Averages of terminal groups may be used, if these groups have reasonably high frequencies, or the whole distribution may be split about equally and each of the two halves averaged. In any case the line used should be laid out on the log graph to see whether it does represent the apparent trend well.

HETEROGONY

Julian S. Huxley, whose interesting and invaluable book on the subject (Huxley 1932) should be read by all zoologists, has gone into the properties, meaning, and use of the equation $Y = bX^k$ and of the coefficient k, both of which he devised, in great detail. Some of what has already been said and most of what follows are based on his work, to which reference is to be made for many examples and a fuller discussion than is possible in this more general book.

An organ that grows at the same rate as the body as a whole is isogonic; the relationship isogony. In heterogonic growth, or heterogony, an organ grows at a different rate from that of the body as a whole or of another standard of comparison. In simple heterogony, the rate of the organ, although different, bears a constant ratio to that of the standard of comparison. In other words the relationship is that symbolized by $Y = bX^k$. The equation with numerical values of b and k inserted is a heterogony formula for any given case, and k (the ratio of the two rates of growth) is the constant differential growth ratio or, as we prefer to say, the coefficient of heterogony. The heterogony is said to be positive if it indicates an increase in the ratio Y/X with an increase in X, or in other words, if as X increases Y increases more rapidly, and negative if the ratio Y/X decreases as X increases, Y increasing less rapidly than X or decreasing as X

increases. Heterogony is positive if the coefficient is greater than 1 and negative if the coefficient is less than 1, as shown by the preceding discussion of the properties of the equation.[1]

It is customary and preferable to take a total weight or large dimension of the whole body or of a considerable part of it as the independent variable (X of the equation) and a weight or dimension of a part included in the independent variable or, in any case, of a lesser part as the dependent variable (Y of the equation). It is for this reason that we have noted the points where $X = Y$ as zoological limits of possible growth.

Huxley and others have shown that size relations do rather closely follow the formula for simple heterogony in a large number of cases among plants, invertebrates, and vertebrates. Here are a few examples selected at random from the dozens available:

Shoot weight against root weight in plants (from Pearsall, given by Huxley):

Daucus carota (carrot).—$k = .55$. Negative heterogony. The shoot becomes larger as the root grows, but the root grows faster than the shoot. The ratio of weights shoot: root becomes rapidly smaller as the root grows larger.

Pisum sativum (pea, etiolated, one of three experiments).—$k = 2.65$. Very high positive heterogony. The shoot grows faster than the root, and the ratio becomes rapidly larger as the root grows.

Nuclear diameter against cell diameter in oöcytes of *Hydractinia echinata* (from Teissier, given by Huxley).—$Y = 1.5X^{.69}$. Negative heterogony, the nucleus growing less rapidly than the cell as a whole. From the discussion on page 362, it follows that the zoological lower limit of C is reached when $X^{1-k} = b$. This is true when $X = 3.7$. Then, according to the formula, the nucleus would occupy the entire cell, and since this is impossible it follows either that cells as small as this or near it will not occur or that the heterogony formula is different for very small cells. In fact the smallest observed value of X was 6.8, suggesting that the relationship does effectively set a minimum cell size.

Weight of large chela against total weight in *Uca pugnax* (from Huxley).—$Y = .0073X^{1.62}$. (This example involves very abundant data and has been thoroughly analyzed; it is interesting as being a principal basis for Huxley's development of the method here discussed.)

Length of face against skull length in the sheep dog (on Becher's data, from Huxley): $Y = .28X^{1.5}$.

[1] Note that a negative coefficient always indicates negative heterogony, but that the heterogony may be negative although the coefficient is positive (but less than +1).

Length of face against skull length in the baboon *Papio porcarius* (Zuckerman's data, from Huxley): k is about 4.25, a remarkably high value. This and the last examples illustrate a phenomenon common among vertebrates, the fact that the preorbital part of the head grows more rapidly than the postorbital part. It is clear that the heterogony coefficients of this and other parts of primates are among their most important characters anatomically, functionally, and taxonomically. This very high k in the baboon gives it its "dog face" and is aberrant among primates. Man certainly has a very low corresponding value of k, and values for most other primates lie between these extremes.

Cranial capacity against skull length in *Mus bactrianus* (from Green): $Y = .0180X^{1.85}$. Green (1933) has shown that heterogony coefficients can be used to distinguish species and also that they reveal what differences in proportions are not of taxonomic value but are only functions of size. He has also shown the relationships of heterogony in hybrids to that of the parent species. This and the other coefficients given by him are also interesting because they are based on adult mice, not on juvenile growth. They show that the essential character, even in a comparison of adults, is not the ratio or proportion of two parts but their heterogony, a point also strongly emphasized by Huxley and others.

Facial length against skull length in the horse (from Robb): $Y = .30X^{1.23}$. Robb (1935) has examined this relationship in the approximately direct lineage of the horse, *Hyracotherium* to *Equus*, in recent adult horses of varying size, and in growth stages of recent horses, and he shows that the heterogony is approximately the same in all three series. The marked change in facial-cranial proportions in the evolution of the horse appears then simply as a function of size. It seems that there is no progressive evolution of face length as such but that on the contrary the genetic heterogony factor involved has not been affected by mutation. The adult size has become larger, and the heterogony coefficient already fixed in *Hyracotherium* has determined the relative facial length of all its descendants. This progressive change in proportions is not, therefore, orthogenetic as commonly stated but, in a sense, nongenetic. The persistence of an approximately fixed value of k also explains the apparent recapitulation as regards this proportion in the growth of the recent horse.

These few examples suggest something of the almost endless possibilities and the extraordinary value of Huxley's coefficient.

There are cases in which the two formulas

$$Y = a + bX$$
$$Y = bX^k$$

give almost equally good results. The regressions then are nearly linear both on arithmetic and on logarithmic coordinates, the value of k differs little from 1, and the value a differs (relative to absolute sizes of Y and X) little from 0. The first formula

ignores the difference of k from 1. It may be preferable when a is relatively large, that is, when it is large enough to have more influence on the result than does the difference between k and 1. The second formula ignores the difference of a from 0 and is preferable when the value of k has more influence on the result than does that of a. In practice the second formula is usually preferable.

The following three heterogony formulas are based on the same specimens of *Alligator mississippiensis* as Example 85 (data from Mook 1921):

X = median length of skull

1. Y = maximum width of skull. $Y = .56X^{.986}$
2. Y = postorbital length of skull. $Y = 1.43X^{.755}$
3. Y = preorbital length of skull. $Y = .187X^{1.215}$

Formulas 2 and 3 show significant heterogony, 2 negative and 3 positive.[1]

Huxley (1932) found that in the earliest embryonic growth periods, when the geometric growth rate is very high and when histological differentiation of the embryo is in progress, the conception of simple heterogony is seldom or never fulfilled and the laws of relative growth are quite different from those to be deduced from subsequent relationships. He proposes to call this phase histodifferentiation. After the completion of the basic histological differentiation in the embryo occurs the phase called

[1] These relate to two dimensions which together make up X. Therefore, if one has negative heterogony, the other necessarily must have about an equal degree of positive heterogony, which the formulas show to be the case. It would be misleading and logically incorrect to use the isogony formula $Y = a + bX$ for either of these, and in Example 85 it was shown for 3 that the equation in this form ($Y = .711X - 15.9$) does not fit the facts so well as the heterogony formula. In 1, however, the heterogony coefficient, .986 does not differ significantly from 1. The isogony formula therefore may fit these data about as well as the heterogony formula. It is

$$Y = .52X - .5$$

and by trial it is found that the two formulas do give about equally good approximations of the observations. The geometric rates of growth are about equal, and the ratio of width: length is about .52 and does not vary greatly, $a(-.5)$ being very small relative to the observed values of Y and X. Nevertheless it is better to consider this relationship by the heterogony and not the isogony formula. The growth involved is unquestionably geometric, not arithmetic, the factor a is small and no serious error results from ignoring it, and valid comparison with the distinctly heterogonic relationships of other parts of the skull can be made only by means of k.

by Huxley auxanodifferentiation. In this phase, relative growth often or usually follows the law of simple heterogony.

It has been found that instead of being about constant throughout (or during auxanodifferentiation) k may successively take two, three, or perhaps still more values. This is an unusual and

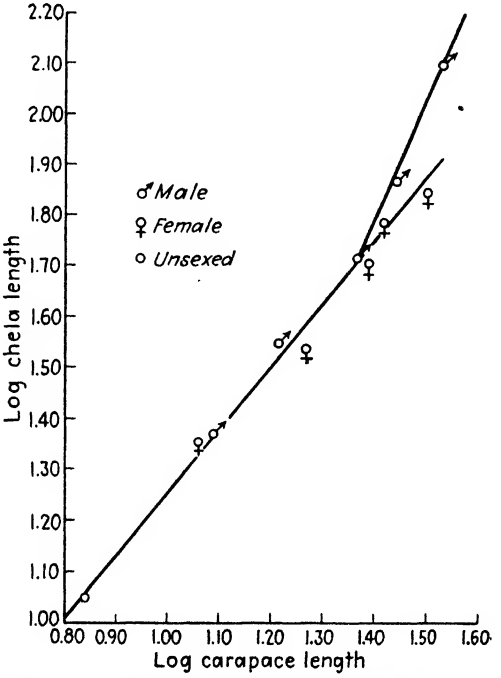


FIG. 51.—Change in heterogony coefficient in a single growth series. Chela length against carapace length in the crab *Palaemon bengalensis*, plotted by the logarithms of the measurements in millimeters. In the males the coefficient becomes abruptly larger in the older individuals, and there are two distinct growth phases. The females apparently have but one phase within the limits of these observations. [*From Huxley* 1932 (*based on data from Kemp*).]

unexpected phenomenon; for the examples studied do not show k simply as variable or as shifting gradually from one value to another,[1] but show it as practically constant for a period, then changing abruptly to a new value which in turn remains constant for some time. In some cases the line on the logarithmic plot

[1] There are, of course, cases in which k is variable or does show such shifts. Their study presents notable difficulties, and little has been done with them as yet except by purely empirical methods. The extraordinary thing is that k is so often constant or nearly so.

bends at a greater or smaller angle to its former course (see Fig. 51). In others it is suddenly offset or turns at an angle for a short distance and then resumes its former slope, continuing parallel to the earlier part of the line. The bending of the line represents a change in the value of both b and k, an offset a change in the value of b only. In some instances abrupt changes of this

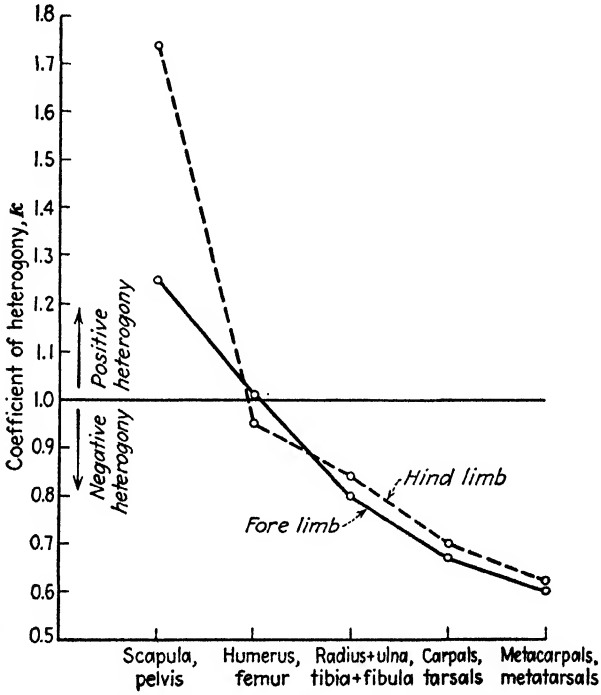


FIG. 52.—Growth gradients in limb segments. Weights of limb bones relative to vertebral column in young Suffolk sheep. Both fore and hind limbs show a strong regular growth gradient, with a distal center of negative heterogony. (*From Huxley* 1931.)

sort are known to be coincident with a well-defined physiological event, such as the onset of puberty. It may safely be assumed that this is true of all cases in which k (and b), otherwise constant, take two or more definite values in the course of growth.

A further important result of the study of heterogony has been the discovery and evaluation of growth gradients. It has been found that where any region of the body shows pronounced heterogony, positive or negative, there are usually also hetero-

gonic relationships within the smaller segments of this region and that the heterogony coefficients for these segments are not usually capricious and irregular in distribution but follow a definite gradient; for instance in dealing with limbs the heterogony is often most intense distally and becomes regularly less intense in more proximal parts. As an example of this, Huxley has given the following heterogony coefficients for designated limb segments in young sheep, weights of skeletal parts in the segment against the weight of the vertebral column:

Part	k
Scapula	1.25
Humerus	1.01
Radius + ulna	.80
Carpals	.67
Metacarpals	.60
Pelvis	1.74
Femur	.95
Tibia + fibula	.84
Tarsals	.70
Metatarsals	.62

In this phase, although the girdles show positive heterogony, the limbs as a whole are negatively heterogonic with respect to the vertebral column (or to the body as a whole). They show a steep gradient with a distal center of negative heterogony. These data can conveniently be represented graphically as in Fig. 52.

From such studies it appears that the body may be considered as covered by a general field of growth potential and that changes in proportions and the general phenomena of heterogony over the body are the results of the varying intensities of this potential, intensities distributed in orderly growth gradients. This, however, is a zoological conclusion beyond the results of numerical method. The greatest importance of this method in particular and of all numerical methods in general is that they provide a firm basis for the discovery and study of such fundamental zoological phenomena.

APPENDIX

CALCULATION

For calculation by any system, the orderly tabulation of data is essential to efficient work. As observations are made or the raw data first arranged for study, it is essential to record them as neatly and logically as possible, putting on a separate sheet each body of data that might possibly need to be considered as a separate sample. It is easy to combine results or data but is annoying and conducive to inaccuracy to work with confused or mixed records.

Many of the most difficult but most common calculations in numerical procedure are obviated by the use of tables. We have given summary tables adequate for almost all zoological purposes (see list, page xvii). Similar tables, in some cases in more detail, are given in most textbooks of statistics, and there are a few other sorts of statistical tables, more rarely necessary to the zoologist, to be found elsewhere (see, for instance, Miner and Pearson in the Bibliography). In addition to these tables, it is convenient or necessary to have a good table of common logarithms, preferably also one of Napierian logarithms, and a table of squares, square roots, and reciprocals. Several mathematical handbooks and texts, such as that edited by Hedrick (see Bibliography), give these in convenient form.

Calculation that cannot be reduced to consultation of tables seldom involves more than simple arithmetic and can all be done longhand if necessary. It may, however, become laborious when many places are to be carried or long lists of data added or otherwise combined. The use of logarithms greatly reduces work in multiplying, dividing, and taking powers and roots. Logarithms are helpful, for instance, in calculating standard deviations and standard errors.

Much time can be saved and greater accuracy can usually be maintained if some mechanical aids are available for calculation. Calculating machines are bulky, expensive, and not absolutely essential but are ideal if available. Lacking one of these, a small

pocket adding device and a good slide rule form a practical, portable, and inexpensive combination. In this text it has been assumed that nothing more elaborate than this is available. The examples have been laid out with this in mind, and most of them have been worked both by slide rule and by calculating machine, many of them also longhand, as a check and to assure the practicability and adequacy of simple means of calculation. The formulas have been given in forms well adapted to slide-rule or longhand calculation.

There are now on the market several practical adding devices of pocket size operating on the principle of the abacus. A slide rule for use in combination with a pocket abacus should be of polyphase type, readable to at least three significant figures, with two scales of squares A and B, an inverted scale CI, and the ordinary scales C and D. For abacus–slide-rule computation of the most used parameters, M, σ, and V, with their standard errors, this form is convenient:

X	f	fX	d	d^2	fd^2	N =
						$\sqrt{N}$ =
						$\sqrt{2N}$ =
						M =
						σ =
						V =
	$N = \Sigma f =$	$\Sigma(fX) =$			$\Sigma(fd^2) =$	

Under X, enter individual measurements and under f the frequency of each. Add column f by abacus, and enter the sum as N. Determine the square root of N and of $2N$ by slide rule or table, and enter. Multiply each X by the corresponding f, enter in the fX column, and add. Divide the result by N, and enter the quotient as M. Subtract each X from M, and enter in the d column. Square each value of d by slide rule or, preferably, tables, and enter in d^2 column. Multiply each d^2 by the corresponding f, enter in fd^2 column, and add. Further calculation is one continuous operation on the slide rule if the polyphase type mentioned above be used:

Set indicator to $\Sigma(fd^2)$ on scale A.
Slide N on scale B to indicator.
Read σ on scale D at index.

Set indicator to $\sqrt{N}$ on scale *CI*, read standard error of M at indicator on *D*.
Set indicator to $\sqrt{2N}$ on *CI*, read standard error of σ at indicator on *D*.
Set indicator to M on *CI*, read V at indicator on *D*.
Slide index to indicator.
Set indicator at $\sqrt{2N}$ on *CI*, read standard error of V at indicator on *D*.

Even such relatively complex calculations as that of t for comparing means of small samples can be done quickly and with sufficient accuracy on a slide rule. The following form can be used to calculate this value:

$M_1 - M_2 =$	Denominator of $t =$
$N_1N_2 =$	$t =$
$\Sigma(d_1^2) + \Sigma(d_2^2) =$	$P =$
$N - 2 =$	

The values in the left-hand column may be entered directly from the data for the two samples.

Having entered these values they may be placed in the formula for t and calculated or may be calculated on the slide rule directly as follows:

Set indicator to $[\Sigma(d_1^2) + \Sigma(d_2^2)]$ on *A*.
Slide $N - 2$ on *B* to indicator.
Read *D* at index and write down value. (This is the value of the denominator of the fractional formula for t).
Set indicator to (N_1N_2) on *A*.
Slide N on *B* to indicator.
Set indicator to $(M_1 - M_2)$ on *C*.
Slide value of denominator (as determined above) on *C* to indicator.
Read value of t at index on *D*.

Although the formula appears so complex, the whole calculation carried out in this manner need take no more than three or four minutes.

Although calculating machines are not necessary in zoological practice, they do have obvious advantages in speed and accuracy. Different makes of machines do not agree exactly in arrangement, so that precise instructions for their operation cannot be given here, but are provided with each type by the manufacturers. Simple adding machines have few advantages over an abacus for this purpose, and in speaking of calculating machines reference is to those that carry out most arithmetic operations

more or less automatically. These have two banks of windows in which the results appear.

In adding, the sum usually appears in the lower window, and the number of items added in the upper. Although a machine in proper repair is absolutely accurate, it is always possible to strike a wrong key inadvertently; all operations should be checked. With a machine that prints as it adds, addition can be checked by comparing the printed items with the original data. With machines that do not print, addition can only be checked by repetition.

In subtraction the minuend is first entered and appears in the lower window, the digit 1 appearing in the upper window. After subtraction this digit should disappear, leaving only a row of zeros in the upper window and the difference in the lower window. To check, add in the subtrahend, when the minuend should appear in the lower window.

In multiplication the machine is set for repetition, *i.e.*, so that the keyboard is not cleared automatically. The multiplicand is punched on the keys and then entered as many times as required, by the method specified in the instructions for any particular type of machine. In the end, the multiplier appears in the upper window, the product in the lower window, and the multiplicand still on the keys. If multiplier and multiplicand are correct as thus recorded, the product is correct.

In division the dividend is first entered and appears in the lower window. The upper window and keyboard are cleared, and the divisor is entered on the keyboard and subtracted repeatedly from the dividend according to instructions with the machine. In the end the quotient appears in the upper window and any remainder in the lower window. To check, multiply divisor and quotient, which should give the dividend.

Squaring is done by multiplication, and roots are taken by a somewhat complex method outlined in special instructions for each machine and essentially a series of subtractions. Squares can be checked by division and roots by multiplication.

Time will be saved in the end if each unit operation is checked before the next is begun. It is important to watch the position of a decimal point carefully throughout a series of operations.

Calculation longhand or by slide rule is made easier by secondary grouping, by multiplying in the frequencies, and especially

by calculating from deviations. The formulas given in this book have mostly been given in forms facilitating this simplification of arithmetic. If, however, a calculating machine is available, this arithmetic work is not a serious burden in any case, and the number of separate operations, rather than the numerical complexity of each, becomes the important point. The number of operations can be considerably reduced by working directly from the raw data, and there are special forms of all the more important formulas adapted to this purpose.

In cases where deviations enter into the usual formulas, the raw-data formulas for machine calculation take advantage of a special form of the "short method" explained in the text. As in that method, they use an assumed mean and then apply a correction, but the mean they assume is always zero. The raw data are then themselves the pertinent deviations, since any ordinary number is a deviation from zero, and the correction factor, which is the square of the distance from the assumed mean to the real mean, becomes simply the square of the mean, M^2.

The following raw-data forms of the most used deviation formulas show the way in which this special machine procedure is applied:

$$\sigma = \sqrt{\frac{\Sigma(X^2)}{N} - M^2} = \sqrt{\frac{N\Sigma(X^2) - [\Sigma(X)]^2}{N^2}}$$

$$r = \frac{\Sigma(XY) - \Sigma(X)M_Y}{\sqrt{\{\Sigma(X^2) - [\Sigma(X)]^2\}\{\Sigma(Y^2) - [\Sigma(Y)]^2\}}}$$

Note that $\Sigma(XY) - \Sigma(X)M_Y = \Sigma(XY) - \Sigma(Y)M_X$, so that either may be used as numerator.

t for comparing means of small samples:

$$t = \frac{M_1 - M_2\sqrt{\frac{N_1N_2}{N_1 + N_2}}}{\sqrt{\frac{\Sigma(X_1^2) - N_1M_1^2 + \Sigma(X_2^2) - N_2M_2^2}{N_1 + N_2 - 2}}}$$

In machine calculation it is an unnecessary complication to record the separate values of X^2 or Y^2. If each of the values of the variate X is squared successively without clearing the windows, when the whole column has been run through, the upper window will record $\Sigma(X)$ and the lower $\Sigma(X^2)$. Repeat

to check. Similarly each value of XY need not be separately recorded. Each X is entered on the board and multiplied by the corresponding Y. When the series has been run through without clearing, the upper window will show $\Sigma(Y)$ and the lower window $\Sigma(XY)$. Repeat with Y entered on the board and X used as multiplier to check.

SYMBOLS

The principal symbols consistently used with definite meanings are here gathered and defined for reference. Some other symbols are used for occasional purposes without standardization; and a few of those here defined are used in more than one way, as made clear in each specific instance by context or note. No two workers agree exactly in notation, but the symbols here given and used throughout this book are in common use, with very few exceptions, and so are about as nearly standard as any. They are listed nearly in the order of appearance in the text, which is a reasonably natural sequence.

X = any individual value of a variate. If two variates are involved, X is usually taken as the independent variate.

Y = any individual value of a second variate, usually a dependent variate.

f = any frequency within a single class.

N = the total frequency of a given sample.

n = any number, also used in several special senses clear from the context, *e.g.*, in a binomial to represent its power, in obtaining the median to represent the serial number of the desired observation, or in the χ^2 technique to represent the degrees of freedom.

p = in a binomial, the probability of occurrence.

q = in a binomial, the probability of nonoccurrence.

Σ = the summation of all quantities designated by a symbol following in parentheses; *e.g.*, $\Sigma(X)$ is the sum of all the values of a variate symbolized by X.

M = the arithmetic mean.

A = (1) an assumed value, *e.g.*, an assumed mean in the short method of calculating M, σ, etc.; (2) the arithmetic growth rate.

d = any deviation from a given value, usually, if not otherwise specified, the deviation of a single observation from the arithmetic mean for the sample of which it is part.

c = used alone or with various superscripts and subscripts for corrections to be applied in any operation, especially in the short method of finding the mean, etc. Also used rarely and when

obvious from the context for a third parameter of a mathematical equation.

i = the class interval.

1, 2, etc. = numerical subscripts ordinarily distinguishing data and parameters derived from each of two or more samples.

L = sometimes used for class limits. L_l is the lower and L_u the upper limit.

M.D. = mean deviation.

σ = standard deviation. With some subscripts it represents standard errors; *e.g.*, σ_M symbolizes the standard error of an arithmetic mean.

v = variance. Subscripts designate the point around which the variance is taken.

Q_1, Q_3 = the first and third quartiles.

Q.D. = quartile deviation or semi-interquartile range.

V = coefficient of variation.

π = the mathematical constant, ratio of the circumference of a circle to its diameter, 3.14159. . . .

e = the mathematical constant, base of Napierian logarithms, limit of the series $1 + \frac{1}{1} + \frac{1}{1 \times 2} + \frac{1}{1 \times 2 \times 3} \cdots$, = 2.7183. . . .

x, y = here occasionally used in the usual mathematical sense of the unknowns in equations, not the same as X, Y, which are values of variates, even when used in equations.

y = also used in the special sense of the second unknown in the normal equation, hence to symbolize ordinates of a normal curve. y_0 is the ordinate at the mean, the maximum ordinate.

S_k = an approximate measure of skewness.

K = a measure of kurtosis.

P.E. = probable error.

r = a measure of correlation, the product moment coefficient.

z = a measure of correlation, transformed from r.

t = (1) a value distributed in a definite way (as tabled), calculated in various ways, and used to estimate probabilities in comparisons of several parameters estimated from small samples; (2) time in the study of growth.

ρ = coefficient of rank correlation.

η = correlation ratio. η^1 symbolizes the corrected value.

log = logarithm, on base 10 unless otherwise specified; $\log_e$, Napierian logarithm.

antilog = antilogarithm.

D = a theoretical range of logarithms of observed values as estimated from a small sample or single specimen.

a = any constant to be inserted in an equation, especially the first constant of the rectilinear regression equation.

b = usually the second constant of the rectilinear regression equation, the coefficient of regression. b_{YX}, regression of Y on X. b_{XY}, regression of X on Y.

S_Y = standard error of estimate of Y (in its regression on X).
χ^2 = a value distributed in a definite way (as tabled), calculated in any of several ways, and used, especially in the study of association, to estimate the probable agreement between an observed series of frequencies and a theoretical series.
G = geometric growth rate.
k = the coefficient of heterogony or constant differential growth rate.

FORMULAS

Such operations as can be reduced to formulas are here classified and listed for reference. Once the meaning of a procedure has been grasped, its use can generally be guided by the formulas alone and they are here given in such a form as to make unnecessary, in most cases, a search through the explanatory text. The sequence is approximately that of the text with some rearrangement to bring together formulas similar in nature or purpose. Symbols sufficiently defined in the preceding list (page 380) are not again explained here. There are a few special formulas adapted for use with calculating machines only that are not given in this list but in the section on machine calculation (page 379).

Distributions

The normal curve:

$$y = \frac{N}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}}$$

Binomial distribution: Expansion of $(p + q)^n$. The frequency of any given class X is given by the expression

$$f = N\left(\frac{n(n-1)(n-2)\cdots(n-X+1)}{X(X-1)(X-2)\cdots 1}\right)p^X q^{n-X}$$

Poisson series: The frequency of any given class X is given by the expression

$$f = Ne^{-M}\left(\frac{M^X}{X(X-1)(X-2)\cdots 1}\right)$$

Averages

Arithmetic mean: Ungrouped or primarily grouped data—

$$M = \frac{\Sigma(X)}{N}$$

Secondarily grouped data—

$$M = \frac{\Sigma(fX)}{N}$$

Short method—

$$M = A + c,$$

where $c = \Sigma(fd_A)/N$. If class intervals are not unity

$$M = A + c_1$$

where

$$c_1 = \frac{\Sigma(fd_A)}{N} i.$$

Weighted arithmetic mean of arithmetic means:

$$M = \frac{N_1M_1 + N_2M_2 + \cdots}{N_1 + N_2 + \cdots}$$

Median: $L_l + \frac{(n - .5)i}{f}$, where n = the serial number of the desired observation within the median class and f = the frequency of that class, or $L_u - \frac{(n - .5)i}{f}$.

Mode: $M - 3$ (M − median), or 3 median $- 2M$.

Geometric mean: $\sqrt[N]{X_1X_2X_3 \cdots X_N}$, or antilog $\frac{\Sigma(\log X)}{N}$

Harmonic mean: The reciprocal of $\frac{1}{N}\Sigma\left(\frac{1}{X}\right)$.

Quadratic mean: $\sqrt{\frac{\Sigma(X^2)}{N}}$.

Measures of Dispersion

Mean deviation: $\frac{\Sigma(fd)}{N}$, in which all values of fd are arbitrarily made positive.

Standard deviation:

$$\sigma = \sqrt{\frac{\Sigma(fd^2)}{N}}$$

For small samples,

$$\sigma' = \sqrt{\frac{\Sigma(fd^2)}{N - 1}}$$

σ^2 and σ'^2 (the same expressions without the radicals) are the variances about the mean v_M. With an assumed mean, the variance is $v_A = v_M + d_A^2$, in which $d_A = M - A$.

The standard deviation by the short method is

$$\sigma = i\sqrt{\frac{\Sigma(fd_A^2)}{N} - c_1^2}$$

in which $c_1 = \sqrt{\frac{\Sigma(fd_A)}{N}}$, d_A being $X - A$ in terms of class intervals.

Semi-interquartile range:

$$\text{Q.D.} = \frac{Q_3 - Q_1}{2}$$

Coefficient of variation:

$$V = \frac{100\sigma}{M}$$

Correlation and Regression

Coefficient of correlation (product moment):

$$r = \frac{\Sigma(d_X d_Y)}{\sqrt{\Sigma(d_X^2)\Sigma(d_Y^2)}} = \frac{\Sigma(d_X d_Y)}{N\sigma_X\sigma_Y}$$

By the short method,

$$r = \frac{\Sigma(d_{A_X} d_{A_Y})}{N\left(\frac{\sigma_X}{i_X}\frac{\sigma_Y}{i_Y}\right)} - c_{1_X}c_{1_Y}$$

Transformed correlation coefficient:

$$z = r + \frac{r^3}{3} + \frac{r^5}{5} + \frac{r^7}{7} \cdots$$

$$z = \frac{\log_e (1 + r) - \log_e (1 - r)}{2}$$

Partial correlation:

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{(1 - r_{13}^2)(1 - r_{23}^2)}$$

$$r_{12.34} = \frac{r_{12.4} - r_{13.4}r_{23.4}}{(1 - r_{13.4}^2)(1 - r_{23.4}^2)}$$

Rank correlation:

$$\rho = 1 - \frac{6\Sigma(d^2)}{N(N^2 - 1)}$$

Rectilinear regression:

$$Y = a_Y + b_{YX}X$$

$$b_{YX} = \frac{\Sigma(d_X d_Y)}{\Sigma(d_X^2)} = r\frac{\sigma_Y}{\sigma_X}$$

$$a_Y = M_Y - b_{YX}M_X$$

$$r = \sqrt{b_{YX}b_{XY}}$$

$$b_{YX} = \frac{r^2}{b_{XY}}$$

Standard error of estimate:

$$S_Y = \sigma_Y\sqrt{1 - r^2}\sqrt{\frac{N}{N-2}}$$

For large samples the second radical can be omitted.

Correlation ratio:

Uncorrected—

$$\eta = \frac{\sigma_{M_X}}{\sigma_X}$$

where σ_{M_X} is the standard deviation of the distribution of the means of the arrays of X.

Corrected—

$$\eta' = \sqrt{\frac{\eta^2 - \frac{(k-3)}{N}}{1 - \frac{(k-3)}{N}}}$$

where k is the number of arrays.

Standard Errors

(All probable errors are .6745 times the corresponding standard errors.)

Of arithmetic mean: $\sigma_M = \frac{\sigma}{\sqrt{N}}$.

Of median: $1.2533\sigma_M$

Of first or third quartile: $1.3626\sigma_M$

Of standard deviation: $\sigma_\sigma = \sigma/\sqrt{2N}$

Of mean deviation: $.6028\sigma_\sigma$.

Of coefficient of variation: $\frac{V}{\sqrt{2N}}\left[\sqrt{1+2\left(\frac{V}{100}\right)^2}\right]$. In almost all cases the expression in brackets can be ignored.

Of coefficient of correlation: $\sigma_r = \frac{1-r^2}{\sqrt{N}}$ or, for small samples, $\frac{1-r^2}{\sqrt{N-1}}$, but it is better to transform to z and use its standard error.

Of transformed correlation coefficient: $\sigma_z = 1/\sqrt{N-3}$.

Of regression coefficient: $\sigma_{b_{yx}} = S_y\sqrt{\frac{1}{\Sigma(d_x^2)}}$ where S_y = the standard error of estimate (page 385).

$$\sigma_{b_{yx}} = \frac{\sigma_y}{\sigma_x}\sqrt{\frac{1-r^2}{N-2}}$$

Comparisons

The formulas here given are in part related to the distribution of t and in part are standard errors of differences to be used in the form d/σ which, because of their distinct character and purpose, are placed here rather than with the standard errors of parameters, etc.

Comparisons of or with large samples: Use various forms of d/σ_d with table d/σ and P.

Comparisons of or with small samples: Use t with table of t and P.

Comparisons of single specimens with means:

$$\frac{d}{\sigma} = \frac{X-M}{\sigma}$$

$$t = \frac{d}{\sigma'} = \frac{d}{\sqrt{\frac{\Sigma(fd^2)}{N-1}}}$$

Comparisons of observed and hypothetical or assumed values of means:

$$\frac{d}{\sigma} = \frac{M-M_A}{\sigma_M}$$

$$t = \frac{M-M_A}{\sqrt{\frac{\Sigma(d^2)}{N(N-1)}}}$$

Comparisons of two observed means: For large samples use d/σ_d, calculating σ_d as follows:

On the hypothesis of derivation from one population or from populations with equal means and variances—

$$\sigma_d = \sqrt{\frac{N_1}{N_2}\sigma^2_{M_1} + \frac{N_2}{N_1}\sigma^2_{M_2}} = \sqrt{\frac{\Sigma(d_1^2) + \Sigma(d_2^2)}{N_1 N_2}} = \sqrt{\frac{\sigma_1^2}{N_2} + \frac{\sigma_2^2}{N_1}}$$

On the hypothesis of derivation from different samples with equal means but with any variances, as estimated by the samples—

$$\sigma_d = \sqrt{\sigma^2_{M_1} + \sigma^2_{M_2}} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

For small samples:

$$t = \frac{(M_1 - M_2)\sqrt{\frac{N_1 N_2}{N_1 + N_2}}}{\sqrt{\frac{\Sigma(d_1^2) + \Sigma(d_2^2)}{N_1 + N_2 - 2}}} = \frac{(M_1 - M_2)\sqrt{\frac{N_1 N_2}{N_1 + N_2}}}{\sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2}{N_1 + N_2 - 2}}}$$

$$= \frac{(M_1 - M_2)\sqrt{\frac{N_1 N_2}{N_1 + N_2}}}{\sqrt{\frac{N_1^2\sigma^2_{M_1} + N_2^2\sigma^2_{M_2}}{N_1 + N_2 - 2}}}$$

Comparison of correlation coefficient with zero: Instead of using r/σ_r, use

$$t = \frac{r\sqrt{N - 2}}{\sqrt{1 - r^2}}$$

Or transform to z and use z/σ_z.

Comparison of two correlation coefficients: Transform r to z and use $\frac{d}{\sigma} = \frac{z_1 - z_2}{\sigma_{d_z}}$, in which

$$\sigma_{d_z} = \sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}$$

Comparison of two regression coefficients:

$$\sigma_{d_b} = \sqrt{\frac{N_1\sigma^2_{Y_1}(1 - r_1^2) + N_2\sigma^2_{Y_2}(1 - r_2^2)}{N_1 + N_2 - 4}\left(\frac{1}{N_1\sigma^2_{X_1}} + \frac{1}{N_2\sigma^2_{X_2}}\right)}$$

$$\sigma_{d_b} = \sqrt{\frac{\Sigma(d_{Y_1}^2)(1 - r_1^2) + \Sigma(d_{Y_2}^2)(1 - r_2^2)}{N_1 + N_2 - 4}\left(\frac{1}{\Sigma(d_{X_1}^2)} + \frac{1}{\Sigma(d_{X_2}^2)}\right)}$$

For large samples use $\frac{b_{YX_1} - b_{YX_2}}{\sigma_{d_b}}$ as $\frac{d}{\sigma}$, and for small samples use the same figure to enter a table of t.

Comparison of η and r: For large samples only—

$$\frac{d}{\sigma} = \frac{\eta^2 - r^2}{2\sqrt{\frac{(\eta^2 - r^2)[(1 - \eta^2)^2 - (1 - r^2)^2 + 1]}{N}}}$$

Association

Theoretical frequencies: Using A, B, C, D for theoretical and a, b, c, d for observed frequencies, in a 2 × 2 table, the relationships are

$$A = \frac{(a + b)(a + c)}{N}$$

$$B = \frac{(a + b)(b + d)}{N}$$

$$C = \frac{(c + d)(a + c)}{N}$$

$$D = \frac{(c + d)(b + d)}{N}$$

χ^2: The sum of the cell contributions = χ^2. To get the contributions of each cell, take the difference between the observed and theoretical frequencies, square this difference, and then divide by the theoretical frequency. Or in a 2 × 2 table—

$$\chi^2 = \frac{N(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$$

By the ratio method—

$$\chi^2 = \frac{\sum\left[a'\left(\frac{a'}{a' + b'}\right)\right] - a\left(\frac{a}{a + b}\right)}{\frac{a}{a + b}\left[1 - \left(\frac{a}{a + b}\right)\right]}$$

where a' and b' = individual values in the first and second rows, respectively.

a and b = the corresponding totals.

Growth

Arithmetic growth rate:

A = arithmetic growth rate.
Y_t = value of variate at end of period of observation.
Y_0 = value of variate at beginning of period of observation.
t = period of observation.

$$A = \frac{Y_t - Y_0}{t}$$

Geometric (or so-called instantaneous) growth rate:

G = geometric growth rate.
Other symbols as for arithmetic rate.

$$Y_t = Y_0 e^{Gt}.$$

$$\log_e Y_t = \log_e Y_0 + Gt.$$

$$G = \frac{\log_e Y_t - \log_e Y_0}{t} = 2.3026 \frac{\log Y_t - \log Y_0}{t}$$

Heterogony:

$$Y = bX^k.$$

$\log Y = \log b + k \log X$. k is calculated as a regression coefficient in $Y = a_Y + b_{YX}X$, by the following equivalences—

$$\begin{aligned} \log Y &= Y \\ \log b &= a_Y \\ k &= b_{YX} \\ \log X &= X \\ b &= \text{antilog } (\log Y - k \log X) \end{aligned}$$

For rough estimate of k—

$$k = \frac{\log Y_t - \log Y_0}{\log X_t - \log X_0}$$

Miscellaneous

Ordinates of the normal curve:

$y = ay_0$,
where y_0, the maximum ordinate, $= Ni/(\sigma\sqrt{2\pi})$, or $Ni/(2.507\sigma)$.
$a = d/\sigma$.

Skewness:

S_k = (mean − mode)/σ = 3 (mean − median)/σ.

Kurtosis:

$$K = \frac{\frac{\Sigma(fd^4)}{N}}{\sigma^4} - 3$$

or by the short method

$$K = \frac{\frac{\Sigma(fd_A^4)}{N} - c_1^4}{\sigma_1^4} - 3$$

Theoretical logarithmic range:
$D = \log (50 + V) - \log (50 - V)$.
With two observations assumed to be at range limits—

$$D = \log X_1 - \log X_2$$

or

$$\text{antilog } D = \frac{X_1}{X_2}$$

Degrees of freedom: In correlation, regression, and the comparison of two distributions, the degrees of freedom are one less than the number of pairs of observations. In a contingency table, the degrees of freedom are one less than the number of rows multiplied by one less than the number of columns.

GLOSSARY

Included in this brief glossary are the principal words used in a technical sense in this book but not purely or primarily zoological in connotation.

Abscissa.—Horizontal distance on a graph, values on the horizontal or x-axis, usually taken as values of the independent variate.

Accurate.—When used in a technical sense, a range that includes the precise value, or a class midpoint designation that implies such a range, as opposed to *exact*, which designates the precise value, without implication of range—a practical impossibility in observations on continuous variates.

Additive growth.—Growth by addition of increments that do not themselves proceed to grow.

Arithmetic increment.—The absolute change in value of a variate over a period of time.

Arithmetic rate.—The absolute change in value of a variate per unit of time.

Array.—Among paired observations on two variates, as in regression, all the values of one variate corresponding to any one given class of the distribution of the other variate.

Attribute.—A qualitative character not primarily recorded by a number.

Bias.—Any tendency in sampling or in observations on samples for the recorded values to depart from those that would arise purely at random or from chance. Inaccuracy is not bias if the directions of the inaccuracies are themselves random and inconsistent. Bias is by implication entirely unconscious on the part of the investigator.

Bimodal.—In a frequency distribution, the presence of two distinctly separate points or classes of high frequency separated by lower frequencies.

Binomial distribution.—A frequency distribution approximating the form given by the expansion of the expression $(p + q)^n$.

Cell.—A rectangular space in a numerical table or diagram defined by its position in a certain row and a certain column.

Class.—Any one category of a frequency distribution (grouped or ungrouped) of a variate.

Class limits.—Distinguished as *conventional* or *real* (see page 37).

Coefficient.—A conventional calculated numerical representation of the intensity of some characteristic. The most important are the coefficients of variation, regression, and correlation.

Column.—In any numerical table, a vertical series of numbers, contrasted with row.

Constant.—An observed or calculated numerical value, taken as invariable for any one set of observations, that may be inserted in an equation to give an approximate formalization of an observed frequency distribution.

For instance, the mean and the standard deviation are constants of the normal distribution (*see* Parameter).

Constant differential growth ratio.—The coefficient of heterogony (*see* Heterogony).

Contingency table.—A table in which frequencies of a sample are distributed by their pertinence to a certain category as regards each of two variables. The categories of one variable are represented in the table by the rows and of the other by the columns, each cell thus being defined by a definite combination of values of the two variables.

Coordinates.—Any system of lines or any field imagined as containing lines used for the graphic expression of numerical values. *Rectangular coordinates* have the lines at right angles. If equal distances express equal absolute values, the coordinates are *arithmetic;* if they express equal ratios, the coordinates are *logarithmic. Semilogarithmic* or *arithlog* coordinates are logarithmic in one direction and arithmetic in the other. In *angular coordinates*, lines radiate from a point, and numbers are represented by angles. *Polar coordinates* combine angular coordinates with concentric circles or the expression of numbers also by distance from a center.

Correlation.—The measurement of the degree in which two variates tend to vary with each other or, in graphic terms, of the intensity of clustering along a line of trend of observations on two variates. Usually measured by the coefficient of correlation r.

Correlation ratio.—A special coefficient of correlation η, not assuming rectilinearity.

Cumulative distribution.—A distribution in which each specified value of the variate is accompanied by the number of observations with this or a greater value, or with this or a lesser value.

Decile.—One of nine points in a frequency distribution dividing it into 10 categories of equal observed frequency. The first decile is the point below which one-tenth of the observations lie, etc.

Degree of freedom.—In a contingency table, frequency distribution, correlation table, etc., the number of spaces that can be filled in at random without affecting the pertinent totals.

Deviation.—The difference between any two values of a variate, especially between an observed and a calculated or hypothetical value, and among these especially between an observed value and the arithmetic mean for a sample.

Efficiency.—As applied to numerical observations on samples, the ratio of the least number of observations from which equally good results would have been obtained to the number of observations actually used.

Frequency.—The number of observations in a given sample belonging to any one category of record.

Frequency distributions.—In general, any classification of frequencies. Usually and in particular, a list of values of a variate accompanied by the numbers of observations of each value.

Geometric increment.—The change in a variate over a given period of time expressed as a percentage of the value at the beginning of that time.

Geometric rate.—The percentage change in a variate per unit of time that would, if acting constantly and continuously over a given period, have produced the observed total geometric increment.

Grouping.—The classification of values of a variate in such a way that more than one exact possible numerical value is included in a single category. Observations on continuous variates are always grouped. If recorded by the original measurements, the grouping is primary (by most other authors erroneously called ungrouped). If recorded in groups of larger scope than the original measurements, the grouping is secondary (by most other authors simply called grouped).

Harmonious growth.—Growth in which a part maintains a constant ratio to a whole (represented by the equation $Y = bX$).

Heterogony.—Growth in which one part grows at a rate different from the whole or from another part. If the two growth rates are unequal but maintain a constant ratio, the relationship is *simple heterogony.* If the part grows more rapidly than the whole (or standard of comparison), the heterogony is *positive;* if more slowly, it is *negative.*

Histogram.—A graphic representation formed by plotting rectangles the widths of which are the class intervals (with values of the variate as abscissas) and the heights (ordinates) are the frequencies.

Index.—(*a*) In morphology, a special sort of ratio, usually a dimension multiplied by 100 and divided by a larger dimension of the same anatomical element. (*b*) In statistics, especially economic statistics, any one of many proposed special measurements of relationships involving ratios.

Instantaneous growth rate.—Same as geometric growth rate.

Interval.—In a distribution, the difference between corresponding points in successive classes, as between their lower limits or their midpoints.

Isogony.—Growth in which one part grows at the same rate as another or as the whole. The rates need not be constant so long as they remain equal to each other at all times. Harmonious growth is necessarily isogonic; growth may be isogonic but disharmonious.

Kurtosis.—The sharpness or peakedness of a curve representing a frequency distribution.

Least squares.—A method of fitting a given form of mathematical equation (such as a linear equation) to an observed distribution in such a way as to give the least possible sum of the squares of the differences between the observed values and those theoretically demanded by the equation.

Leptokurtic.—Of a frequency distribution, one giving a graph sharper than the most nearly equivalent normal curve.

Mean.—One of several averages. If used without qualification, reference is to the *arithmetic mean,* the sum of observed values divided by the number of observations. For others, see Chap. V.

Median.—A value of a variate such that in an observed sample half the observations lie above and half below that value.

Mode.—A value of a variate corresponding with the peak of the frequency curve best fitting a given series of observations. Also any observed high frequency point in a distribution.

Module.—One of several proposed measures, especially an arithmetic mean of two different linear dimensions of a single individual.

Moments.—Average powers of deviations from a fixed value, usually the mean. The first moment about the mean is the average of the deviations and is always zero. The second moment is the average of the squares of the deviations, the third moment the average of their cubes, etc. (Some authors define moments as the sums, rather than the averages, of powers of deviations.)

Multiple correlation.—A correlation involving more than two variates.

Multiplicative growth.—Growth in which the increments themselves begin to grow as soon as formed. Also called *compound-interest* growth.

Normal curve.—A symmetrical complex mathematical curve that approximates the fluctuating variation of many zoological characters and of the probability of random sampling under many conditions.

Normal distribution.—Theoretically a frequency distribution of a sample from a population in which the same distribution would follow the normal curve. In practice, in zoology, literally normal distributions do not occur; but most distributions of variates approximate this form.

Ogive.—An oblique S-shaped curve that approximates the graphic form of a cumulative frequency distribution.

Ordinate.—Vertical distance on a graph, values on the vertical or y-axis, usually taken as values of the dependent variate or as frequencies.

Origin.—The point $x = 0$, $y = 0$, in a graph; the intersection of the axes in coordinates.

Parameter.—Same as the special use of "constant" as defined above. Strictly, parameters are the constants of purely mathematical abstractions. In practice and as usually used in this work they are estimates or approximations of such constants calculated from the concrete data of given samples.

Partial correlation.—In a multiple correlation of three or more variates, the procedure of making allowance for the influence of the variation of the third or later variates and estimating the correlation strictly between two.

Percentile.—One of the 99 points dividing a frequency distribution into 100 parts all of equal frequency. Thus the fiftieth percentile is the same as the median, the tenth is tne same as the first decile, etc.

Platykurtic.—Of a frequency distribution, one giving a graph flatter than the most nearly equivalent normal curve.

Poisson series.—A type of frequency distribution, essentially a special case of the binomial distribution in which the probability of occurrence is very small but the total frequency is great.

Population.—The whole existing or possible group of individuals or of observations fulfilling the specifications of a problem being studied on the basis of a more limited group, the sample belonging to that population. The universe of discourse in investigation, in mathematical theory infinite, in zoological practice finite, but invariably more extensive than the concrete data available.

Probable error.—A value such that a single observation drawn at random from a normal distribution is as likely as not to lie within this distance (plus or minus) of the mean. .6745 times any standard error.

Product moment.—In correlation, the average product of corresponding deviations from the means of the two variates. r is usually called the product moment coefficient of correlation.

Quadrat.—In faunal sampling, a rectangular field area selected for collecting or observation of its whole faunule.

Quartile.—Any of three points dividing a frequency distribution into four parts of equal frequency. The median is the second quartile, as it is the fifth decile and the fiftieth percentile.

Quartile deviation.—Half the difference between first and third quartiles, also called the semi-interquartile range.

Quetelet's principle.—The empirical rule that most variations in nature are such that frequencies tend to pile up on a central point and to fall away more or less symmetrically on both sides of this.

Range.—The distance between extreme values of a variate, usually recorded by giving these terminal values. The *observed range* is the range actually seen in a given sample, as opposed to the *real range* in the population, always greater than the observed range but never exactly determinable. The *theoretical range* is an estimate of the real range based on an actual sample, usually taken as $(M + 3\sigma) - (M - 3\sigma)$.

Rank correlation.—A method of estimating correlation not on the actual values of observations but on their serial order. The coefficient of rank correlation is ρ.

Raw.—Data as derived and recorded from observation, without calculation. Rawness does not technically imply any inaccuracy, inadequacy, or crudity.

Refinement.—In measurement and calculation, the number of significant figures given. Refinement and accuracy are technically quite distinct: a highly refined measurement may be and is quite likely to be inaccurate, whereas one of little refinement may be and is more likely to be entirely accurate.

Regression.—The average change in value of one variate accompanying a unit change in another related variate; also the general phenomenon of such consequent changes. The *coefficient of regression* of Y on X, *i.e.*, the average change in Y for a unit change in X, is b_{YX}.

Reliability.—Aside from the usual vernacular meaning, used technically to mean the probable degree of approximation to a population parameter given by calculation of the parameter from a given sample.

Row.—In any numerical table, a horizontal series of numbers; contrasted with column.

Sample.—Any series of specimens or of observations actually in hand pertinent to a given problem and derived from a specified population which is to be studied on the basis of the sample.

Sampling.—The obtaining of individual observations by means of which the population is to be studied. *Qualitative* sampling seeks to obtain at least one observation on each of the principal pertinent variations of the population. *Quantitative* sampling seeks to obtain these approximately in the same proportions as in the population.

Sampling limits.—Given a population in which a character occurs in a specified proportion and given a sample of a specified size, the sampling

limits are the least and the greatest number of individuals in the sample that might have the character according to theories of probability.

Scatter diagram.—A graph on rectangular coordinates in which each pair of observations on two correlated variates is represented by a dot.

Semi-interquartile range.—Same as quartile deviation.

Significance.—The general result of operations showing a figure to be significant in the technical sense defined below. Thus the significance of the difference between parameters calculated from two samples is based on an estimate of the probability that two samples of the given sizes and characters could have been drawn at random from a single population.

Significant.—In statistical usage the statement that a result is significant implies that an actual numerical test has been made and has shown that it is highly improbable that the given result can have arisen by chance. Usually the sense is that this result differs from some hypothetical or expected value more than is likely by chance alone.

Significant figures.—In this phrase the word "significant" is used in a different sense from that of statistical significance or significant differences. Significant figures are those digits in a number that do approximate the real value of the thing measured. *Strictly significant* figures are digits known to be accurate in a technical sense. *Broadly significant* figures are those not surely known to be accurate but known to make the approximation to true values closer than if these digits were omitted.

Skew, skewness.—The property of bilateral asymmetry in a distribution or in a curve based on it. Differences from the normal curve in bilateral symmetry are usually measured by the coefficient of skewness S_k. The skewness is *positive* if the distribution tails off farther among the higher values, or to the right in a curve, and *negative* if it tails off among lower values, to the left in a curve.

Spurious correlation.—A correlation significant numerically but in fact caused by relationship of both of two variates to a third, not perceived or measured, and not by a relationship directly between the two correlated variates.

Standard deviation.—An essential parameter of the normal curve, symbolized by σ, determining its dispersion, and hence the most important estimate of dispersion that can be made from a sample.

Standard error.—The standard deviation not of a primary frequency distribution but of any parameter of such a distribution. The standard error of the mean, for instance, is the standard deviation of a distribution composed of many means from many samples drawn from the same population of specified character. The standard error thus permits some judgment as to the probable closeness of the parameters as calculated in a given sample to the real, but unknown, parameters of the population. The standard error is simply a special and secondary sort of standard deviation, a measure of potential dispersion, and has nothing to do with "error" in the usual vernacular sense.

Standard error of estimate.—In correlation, a measure of the average dispersion of values of one variate within the arrays defined by the other variate.

Statistic.—A calculated, collective measure of some characteristic of a whole group of markedly varying raw data. Estimates of population parameters.

Statistics.—The science of the mathematical study of variation, of the characters of groups, and of the reduction of numerical data.

Total correlation.—A measured correlation between two variates in which the possible or certain influence of other variates is not taken into account or excluded from the measurement.

Trend.—In correlation, regression, time series, etc., as represented graphically, the general direction, axis, or line along which the individual sets of observations tend to cluster.

Variability.—The theoretical, biological and not statistical, capacity for variation, the innate tendency of which variation is in part the expression.

Variable.—Any thing or category of things that exhibits variations. Variables include variates and attributes.

Variance.—In any distribution, the mean of the squares of deviations from a given value, such as the mean.

Variate.—A variable the variations of which are expressed numerically. A variate is *continuous* if it can take any of an infinite number of exact values within its range, and *discontinuous* if it must take one of a definite, finite series of fixed values, such as the integral values.

Variation.—The phenomena of differences between observations classed as belonging to a single category, *e.g.*, between observations on a single thing over a period of time or on a group of biologically homologous things. In numerical practice in zoology, variation usually includes any such differences, without causal definition, embracing differences expressing innate variability or genetic mutation along with those caused by extrinsic factors. Variation is also used to mean any one defined condition assumed by a variable.

BIBLIOGRAPHY

The purpose of the following references is (1) to give the sources of data used in our examples and of other special points cited by the author in the text, (2) to facilitate finding more detailed information on various phases of the subject, and (3) to give a few more or less typical examples of current usages. There has been no previous attempt to review, even in an elementary way, the whole subject of numerical and related procedures in zoology; and the available studies fall into two quite distinct groups, here made separate sections of the bibliography. On one hand there are essentially mathematical or statistical texts and reference works which contain material that is useful in zoology, generally with a great deal that is not, but which are not written from a zoological point of view, do not cover many important points in that field, and do not show the zoological uses of the material that they do give. On the other hand are zoological papers which discuss or exemplify numerical methods but do not in any one case cover much of this field and which are generally deficient from a more definitely mathematical or statistical point of view.

MATHEMATICAL AND STATISTICAL REFERENCES

For general mathematical background any of almost innumerable textbooks are suitable, and none is cited except Hogben's work which is unusual in giving the basic concepts and operations of almost the whole field of mathematics in a single volume comprehensible to any intelligent person with no previous mathematical training. As regards specifically statistical works, there are dozens of available texts, all of which deal in an adequate way with the more important parts of this broad field. Several of the best of these are cited. The student who has sufficient mathematical training and wishes to become proficient in statistics would do well to work through Yule, then Fisher, and thence will readily progress to more specialized and detailed studies. Fisher's excellent book is written especially for biologists (chiefly experimental biologists, not zoologists) and is perhaps the most

logical and, within its scope, the best available work on statistics. Although it is intended as an introduction, experience has suggested that it is almost completely incomprehensible to anyone not already well grounded in statistics. A small but adequate selection of the many available sets of tables and other works of reference is also here given.

ARKIN, H., and R. R. COLTON. 1934. An outline of statistical methods. Barnes and Noble, Inc., New York. [A useful review outline, with references to many standard texts.]

BARLOW, 1919. Barlow's tables of squares, cubes, square roots, cube roots, reciprocals of all integer numbers up to 10,000. Spon and Chamberlain, New York. [Or stereotype ed., 1927. Invaluable arithmetical tables.]

BLAKEMAN, J. 1905. On tests for linearity of regression in frequency distributions. *Biometrika*, 4, pp. 332–350. [Including comparison of r and η.]

BURGESS, R. W. 1927. Introduction to the mathematics of statistics. Houghton Mifflin Company, Boston.

CAMP, B. H. 1931. Mathematical part of elementary statistics. D. C. Heath & Company, Boston.

COOLIDGE, J. L. 1925. An introduction to mathematical probability. Clarendon Press, Oxford.

DAVENPORT, C. B. 1904. Statistical method with special reference to biological variations. John Wiley & Sons, Inc., New York.

DICE, L. R., and H. J. LERAAS. 1936. A graphic method for comparing several sets of measurements. *Cont. Lab. Vert. Genetics, Univ. Michigan*, No. 3, pp. 1–3.

FISHER, R. A. 1935. Statistical methods for research workers. Biological Monographs and Manuals, V, 1 vol. Oliver and Boyd; Edinburgh and London. [A standard biometric text, with special reference to small samples, which are inadequately considered in most other works, especially useful to zoologists.]

HASKELL, A. C. 1919. How to make and use graphic charts. Codex Book Co., New York.

HEDRICK, E. R. 1929 [or other issue]. Logarithmic and trigonometric tables, revised ed. The Macmillan Company, New York. [Also contains squares, cubes, and roots and other useful tables and is one of the most convenient compilations of nonstatistical numerical tables.]

HOGBEN, L. 1937. Mathematics for the million. W. W. Norton & Company, Inc., New York. [Somewhat biased in its brief treatment of statistics but perhaps the most painless means of reviewing or acquiring knowledge of mathematics more than adequate for present purposes.]

KELLEY, T. L. 1923. Statistical method. The Macmillan Company, New York. [An excellent general text.]

MILLS, F. C. and D. H. DAVENPORT. 1925. A manual of problems and tables in statistics with notes on statistical procedure. Henry Holt and Co., New York. [Not a text, and the problems are commercial and economic, but it contains convenient tables and formulas.]

MINER, J. R. 1922. Tables of $\sqrt{1 - r^2}$ and $1 - r^2$ for use in partial correlation and trigonometry. Johns Hopkins Press, Baltimore.

PEARL, R. 1930. Introduction to medical biometry and statistics. 2d. ed., revised. W. B. Saunders Company, Philadelphia. [Although primarily medical, is also an excellent, unusually readable introduction to biological statistics in general.]

PEARSON, K. (editor). 1914. Tables for statisticians and biometricians. Cambridge University Press, London.

RIETZ, H. L. 1924. Handbook of mathematical statistics. Houghton Mifflin Company, New York.

ROSAHN, P. D. 1935. A graphic method for representing the significance of the difference between means. *Human Biology*, VII, pp. 267–271.

WALKER, H. M. 1929. Studies in the history of statistical method with special reference to certain educational problems. Williams and Wilkins Company, Baltimore. [Of wider than educational interest.]

YULE, G. U. 1924. An introduction to the theory of statistics, 7th ed., revised. Charles Griffin & Co., Ltd., London; J. B. Lippincott Company, Philadelphia. [Or other editions. Standard detailed general text. Few tables but numerous references.]

ZOOLOGICAL REFERENCES

A very large proportion of the hundreds of papers on zoology that now appear annually contain some numerical data or imply the use of some essentially numerical concepts. The brief list of references here given is therefore obviously only a very small sample and is selected not so much to represent the best or worst that has been or can be done as to give more or less typical examples of such studies scattered over the whole field of vertebrate zoological and paleontological research.

ALDOUS, C. M. 1937. Notes on the life history of the snow-shoe hare. *Jour. Mammalogy*, XVIII, pp. 46–57. [Unusual distributions of trapping data.]

ANTHONY, H. E. 1925. The capture and preservation of small mammals for study. *Amer. Mus. Guide Leaflet*, No. 61, 53 pp. [Includes standard mensuration.]

BLANCHARD, F. N. 1921. A revision of the king snakes: Genus *Lampropeltis*. *U. S. Nat. Mus., Bull.* 114. [Many numerical data and ingenious graphs of geographic variation.]

BLANCHARD, F. N., and F. C. BLANCHARD. 1931. Size groups and their characteristics in the salamander *Hemidactylium scutatum* (Schlegel). *Amer. Nat.*, LXV, pp. 149–164. [Statistical and biological age grouping.]

BROUN, M. 1933. Some live weights and measurements of small birds. *Bird-Banding*, IV, pp. 52–54.

BRUUN, A. FR. 1935. Flying Fishes (Exocoetidae) of the Atlantic. Dana Report, No. 6. The Carlsberg Foundation, Copenhagen and London. [Many interesting frequency distributions.]

BUBNOFF, S. VON. 1919. Über einige grundlegende Prinzipien der paläontologischen Systematik. *Zeits. indukt. Abstam.—u. Vererbungslehre*, XXI, pp. 158–168. [Usefulness of statistics in taxonomy of ammonites. Use of standard deviation as a measure of dispersion urged. Species distinguished from larger genetic groups by the positive correlation of characters in the latter which vary independently in the former.]

COLLINS, H. H. 1923. Studies of the pelage phases and of the nature of color variations in mice of the genus *Peromyscus*. *Jour. Exper. Zool.*, XXXVIII, pp. 45–107.

DICE, L. R. 1931. Methods of indicating the abundance of mammals. *Jour. Mammalogy*, XII, pp. 376–381. [Quadrat, traverse, and station sampling.]

DICE, L. R. 1932. Variation in a geographic race of the deermouse, *Peromyscus maniculatus bairdii*. *Occ. Papers Mus. Zool., Univ. Michigan*, No. 239.

DICE, L. R. 1937. Additional data on variation in the prairie deermouse *Peromyscus maniculatus bairdii*. *Occ. Papers Mus. Zool., Univ. Michigan*, No. 351.

DIXON, J. 1925. Food predilections of predatory and fur-bearing mammals. *Jour. Mammalogy*, VI, pp. 34–46. [Includes graphic presentation.]

DONALDSON, H. H. 1924. The rat. Wistar Inst., Philadelphia. [Innumerable data, many reduced to empirical equations.]

DUERST, V. V. 1926. Vergleichende Untersuchungsmethoden am Skelett bei Säugern. *Handb. biol. Arbeitsmeth.*, Abt. VII, Heft 2, pp. 125–530. [Very detailed system of mammalian osteometry, including synonymy with many previously published systems.]

EATON, R. J. 1924. The migratory movements of certain colonies of herring gulls in Eastern North America. *Bird-Banding*, V, pp. 70–84. [Typical birdbanding data.]

EHRENBERG, K. 1928. Betrachtungen über den Wert variationsstatistischer Untersuchungen in der Paläozoologie nebst einigen Bemerkungen über eiszeitliche Bären. *Pal. Zeits.*, X, pp. 235–257. [A critique of Soergel on Pleistocene bears, concluding that numerical methods are of little or no value and emphasizing morphological methods. The criticism is partly a misunderstanding of the real nature and use of statistical concepts, partly a valid objection to the use of poor statistical methods (*e.g.*, the range as a means of comparison).]

FROST, F. H. 1927. Statistical identification. I. As applied to *Parapavo*. *Univ. California Pub., Bull. Dept. Geol. Sci.*, XVII, pp. 57–62. [Comparisons of means by a method analogous to Fisher's *t*-test. Exemplified in paper by H. Howard, of which Frost's note is an appendix, on fossil bird material.]

GREEN, C. V. 1933. Differential growth in the crania of mature mice. *Jour. Mammalogy*, XIV, pp. 122–131. [An excellent study of heterogony.]

GREGORY, W. K. 1933. Fish Skulls: A study of the evolution of natural mechanisms. *Trans. Amer. Phil. Soc.*, XXIII, Article 2. [No important numerical data, but an excellent osteometric system.]

GREGORY, W. K., and H. J. COOK. 1928. New material for the study of evolution. A series of primitive rhinoceros skulls (*Trigonias*) from the Lower Oligocene of Colorado. *Proc. Colorado Mus. Nat. Hist.*, VIII, No. 1, pp. 1–32. [Graphic presentation of frequencies of variates and attributes, without, however, any objective analysis of the distributions.]

HATT, R. T. 1936. Hyraxes collected by the American Museum Congo Expedition. *Bull. Amer. Mus. Nat. Hist.*, LXXII, pp. 117–141. [Series of measurements on typical, well defined samples.]

HICKS, L. E. 1934. Individual and sexual variations in the European starling. *Bird-Banding*, V, pp. 103–118.

HILDEBRAND, S. F., and W. C. SCHROEDER. 1927. Fishes of Chesapeake Bay. *Bull. U. S. Bureau of Fisheries*, XLIII, Part 1. [Many tables of numerical data.]

HOWARD, H. 1927. A review of the fossil bird, *Parapavo californicus* (Miller), from the Pleistocene asphalt beds of Rancho la Brea. *Univ. California Pub., Bull. Dept. Geol. Sci.*, XVII, No. 1. [With an appendix Statistical identification as applied to *Parapavo*, by F. H. Frost.]

HUE, E. 1907. Musée ostéologique—Étude de la faune quaternaire: ostéométrie des mammifères. Schleicher Frères, Paris. [A system of mensuration and its applications.]

HUESTIS, R. R. 1931. Seasonal pelage differences in *Peromyscus*. *Jour. Mammalogy*, XII, pp. 372–375.

HUXLEY, J. S. 1932. Problems of relative growth. Methuen & Co., Ltd., London. [A very able and important work on a phase of numerical zoology, the pioneer and standard study of heterogony, with extensive tables, statistics, and graphs.]

JENKINSON, J. W. 1912. Growth, variability and correlation in young trout. *Biometrika*, VIII, pp. 444–455.

KÄLIN, J. A. 1933. Beiträge zur vergleichenden Osteologie des Crocodilidenschädels. *Zool. Jahrb., Abt. Anat. Ont. Tiere*, LVII, pp. 535–714. [A good osteological system and many measurements, but not well analyzed and evaluated.]

KLÄHN, H. 1920. Der Wert der Variationsstatistik für die Paläontologie. *Ber. naturf. Ges. Freiburg i. Br.*, XXII, Heft 2, pp. 7–224. [Elaborate consideration of statistical concepts used in genetics and selection of those believed applicable to invertebrate paleontology, the author's judgment of which is, however, not supported by later and more critical studies. Also an interesting but abortive attempt to deal with relative growth.]

KLÄHN, H. 1929. Was nützt die Variationsstatistik der Paläontologie? *Zeits. deutsch. geol. Ges.*, LXXXI, Heft 1/2, pp. 23–35. [Important as one of the few serious attempts to apply special statistical methods to paleontological data, but unfortunate in choice of methods and inadequate in their application.]

LOW, S. H. 1933. Further notes on the nesting of the tree swallows. *Bird-Banding*, IV, pp. 76–87.

LOW, S. H. 1934. Nest distribution and survival ratio of tree swallows. *Bird-Banding*, V, pp. 24–30.

MACDONELL, W. R. 1902. On criminal anthropometry and the identification of criminals. *Biometrika*, I, pp. 177–227. [Statistical data on numerous dimensions of large groups of criminals and of Cambridge undergraduates.]

MATTHEW, W. D. 1909. The Carnivora and Insectivora of the Bridger Basin, Middle Eocene. *Mem. Amer. Mus. Nat. Hist.*, IX, Part VI, pp. 291–567. [Includes analysis of habitats in a fossil fauna.]

MATTHEW, W. D. 1924. Third contribution to the Snake Creek fauna. *Bull. Amer. Mus. Nat. Hist.*, L, pp. 59–210. [Pages 163–166, age groups in a fossil horse *Merychippus primus*, recognized on biological rather than statistical criteria.]

MEEK, S. W., and S. F. HILDEBRAND. 1923. Marine fishes of Panama. *Field Mus. Pub. Zool.*, XV, Part 1. [From the present point of view, chiefly interesting as an extensive sampling operation.]

MOOK, C. C. 1921. Skull characters of recent Crocodilia with notes on the affinities of the recent genera. *Bull. Amer. Mus. Nat. Hist.*, XLIV, pp. 123–268. [Many unreduced measurements.]

NICE, M. M. 1933. Nesting success during three seasons in a song sparrow population. *Bird-Banding*, IV, pp. 119–131.

NICHOLS, J. T. 1918. Fishes of the vicinity of New York City. *Amer. Mus. Nat. Hist.*, Handbook Series, No. 7. New York. [Includes suggestions on mensuration.]

NICHOLS, J. T. 1935. The Hawaiian "Ulua." *Copeia*, pp. 192–193. [Distributions and grouping.]

OSBORN, H. F. 1912. Craniometry of the Equidae. *Mem. Amer. Mus. Nat. Hist.*, n.s., I, pp. 57–100. [Review of several systems, and mensuration essentially adapted to fossil and recent horses but in part of much wider application.]

OSBORN, H. F. 1929. The titanotheres of ancient Wyoming, Dakota and Nebraska. *U. S. Geol. Survey, Monograph* 55. [Systems and numerous applications and examples of mensuration, ratios, etc.]

PARKER, G. H. 1922. The breathing of the Florida manatee (*Trichechus latirostris*). *Jour. Mammal.*, III, pp. 127–135. [Typical numerical data on a physiological variate.]

PEARSON, H. S. 1928. Chinese fossil Suidae. *Palaentologica Sinica*, V, Fascicle 5, Series C. [Excellent, carefully explained use of M, σ, V, Fisher's t-test, and d/σ with fossil vertebrate material.]

PHILLIPS, J. C. 1920. Skull measurements of the northern Virginia deer. *Jour. Mammalogy*, I, pp. 130–133. [An unusually good study of variation, especially by the use of V, although some of the samples are poorly specified and statistical definitions are not entirely correct.]

RIDGWAY, R. 1901. The birds of North and Middle America . . . Part I. Family Fringillidae. *U. S. Nat. Mus., Bull.* 50, pp. i–xxxii, 1–716. [Standard system of mensuration of birds.]

RIDGWAY, R. 1912. Color standards and color nomenclature. Published by the author, Washington.

ROBB, R. C. 1929. On the nature of hereditary size limitation: I. Body growth in giant and pigmy rabbits. II. The growth of parts in relation to the whole. *Brit. Jour. Exper. Biology*, VI, pp. 293–324.

ROBB, R. C. 1935. A study of mutations in evolution. I. Evolution in the equine skull. II. Ontogeny in the equine skull. *Jour. Genetics*, XXXI, pp. 39–52. [An outstanding study of heterogony and of numerical data on relationships between phylogeny and ontogeny, size and proportions.]

RUTHVEN, A. 1908. Variations and genetic relationships of the garter snakes. *U. S. Nat. Mus., Bull.* 61. [Excellent numerical data and sane, helpful suggestions on the place of such data in zoology.]

SCHOOLEY, J. P. 1934. A summer breeding season in the eastern chipmunk, *Tamias striatus*. *Jour. Mammalogy*, XV, pp. 194–196. [Physiological numerical data.]

SHAW, W. T. 1926. Age of the animal and slope of the ground surface, factors modifying the structure of hibernation dens of ground squirrels. *Jour. Mammalogy*, VII, pp. 91–96. [Interesting data on a problem in correlation or association, insufficiently digested to reveal its real bearing on the questions considered.]

SIMPSON, G. G. 1937. The Fort Union of the Crazy Mountain Field, Montana, and its mammalian faunas. *U. S. Nat. Mus., Bull.* 169. [Examples of applications of many numerical methods to the study of a fossil fauna.]

SNYDER, L., and H. YINGLING. 1935. Studies in inheritance. XII. *Human Biology*, VII, pp. 608–615. [Genetic diagrams.]

SOERGEL, W. 1925. Die Fährten der Chirotheria. Eine paläobiologische studie. Gustav Fischer, Jena. [A system of mensuration and of mathematical deduction applied to a very specialized paleontological problem.]

SOERGEL, W. 1930. Die Bedeutung variationsstatistischer Untersuchungen für die Säugetier—Paläontologie. *Neues Jahrb. Min. Geol. Pal.*, Beil. Bd. LXIII, Abt. B, pp. 349–450. [A reply to Ehrenberg (1928), emphasizing the value of statistical methods for certain paleontological problems. Demonstrates the invalidity of Ehrenberg's objections to Soergel's first statistical work on bears and discusses some abstract principles of classification. "Metrical and statistical studies, which are indeed merely morphological studies with more exact means, go far beyond purely morphological investigation in their possible results."]

SPERRY, C. 1934. Winter food habits of Coyotes: A report of progress, 1933. *Jour. Mammalogy*, XV, pp. 287–290.

STOCK, C. 1929. A census of the Pleistocene mammals of Rancho La Brea, based on the collections of the Los Angeles Museum. *Jour. Mammalogy*, X, pp. 281–289. [Graphic presentation of relative and absolute frequencies of individuals belonging to various taxonomic groups in a single fauna.]

SUMNER, F. B. 1927. Linear and colorimetric measurements of small mammals. *Jour. Mammalogy*, VIII, pp. 177–206.

SVIHLA, A. 1934. Development and growth of deer mice (*Peromyscus maniculatus artemisiae*). *Jour. Mammalogy*, XV, pp. 99–104. [Growth curves, the true nature of which would be clearer with more and better analysis.]

THOMPSON, D'ARCY. 1917. Growth and form. Cambridge University Press, Cambridge. [A classic study of numerical and other relationships in growth.]

TROXELL, D. L. 1915. The vertebrate fossils of Rock Creek, Texas. *Amer. Jour. Sci.*, XXXIX, pp. 613–638. [Comparisons of fragmentary materials to more complete related specimens, by means of ratios.]

WEDEKIND, R. 1916. Über die Grundlagen und Methoden der Biostratigraphie. Berlin. [One of the first applications of strictly statistical concepts to paleontology. The statistical methodology proposed is, however, very largely emended or superseded by later work. For review and criticism see, *e.g.*, Klähn 1920.]

ZEUNER, F. 1934. Die Beziehungen zwischen Schädelform und Lebensweise bei rezenten und fossilen Nashörnern. *Ber. naturf. Ges. Freiburg i. Br.*, XXXIV, pp. 21–80. [Biological conclusions drawn from statistical treatment of a detailed system of cranial angles. Distributions given by quartiles and median, which are less adapted to this problem than are the sigma and mean. Dispersion and variability are confused. No explicit criterion of significance is used.]

INDEX

(Symbols, formulas, and definitions are not indexed separately as such, but may be found in the text as indexed by subjects, and also in the compilations in the Appendix, pages 380–397.)

A

Abmodality, coefficient of, 188*n*.
Absolute growth, 344–346, 350–354
Accretive growth (*see* Arithmetic growth)
Accuracy, 25–30
Acropithecus rigidus, 58, 60, 121, 247
Additive growth (*see* Arithmetic growth)
Age groups, 336–342
  biological, 338–341
  in fossils, 341–342
Aldous, C. M., 68, 401
Alligator mississippiensis, 364–365, 370
Angular coordinates, 307–308
Anisonchus sectorius, 311
Anthony, H. E., 22, 401
Area diagrams, 306
Areas, 4, 15
  of normal curve, 135–140
    table of, 137
Arithlog coordinates, 306–308
Arithmetic coordinates, 306–310
Arithmetic growth, 345–346, 350–353
  rate of, 346, 350–353, 355, 358
Arkin, H., 142, 400
Association, 285–304
  hypotheses of dependence in, 300–304
  multiple, 299–301
  ratio method of, 294–296
  significance of, 289–304
  small samples, 296–298
Association, spurious, 299–301
Attributes, 2, 182
  graphs of, 310, 312–314
  sampling for, 170–171
Auxanodifferentiation, 371
Averages, 85–105
  meaning of word, 103–105

B

Baboon, 369
Bar diagrams, 306, 312–314, 324–325, 329–330
Barlow, P., 400
Becher, A., 368
Bias, 30–33, 45
  in paleontological sampling, 180–181
  in sampling, 171–174
    for growth, 343
Binomial distribution, 75–78
  standard deviation of, 183
Birds, mensuration of, 22
Blakeman, J., 284, 400
Blanchard, F., 22, 47, 49, 226, 228, 242, 279, 339–340, 401
Broun, M., 208, 401
Brown, B., 342
Bruun, A. Fr., 65, 69, 293, 401
Bubnoff, S. von, 402
Burgess, R. W., 400

C

Calculating machines, use of, 377–380
Calculation, 375–380

Camp, B. H., 400
Caranx melampygus, 67–68
Central tendency, measures of, 85–105
Chipmunks, 35, 329
Chleuastochoerus, 190
Citellus columbianus columbianus, 297
Class limits, 37–44
Coefficient, of abmodality, 188*n*.
of consistency, 33
of correlation, 224–254
of heterogony, 363–373
of regression, 262–274
of reliability, 33
of skewness, 143–145
of variability, 122–128
of variation, 122–128, 135, 146
Colbert, E. C., 334
Collins, H. H., 24, 402
Color, mensuration of, 24–25, 325–326
Color top, 24
Colton, R. R., 142, 400
Comparisons, of correlation coefficients, 242–243, 284
of means of two samples, 192–197
of regression coefficients, 277–280
of samples, 186–202
of single specimens, with large samples, 188–192
with small samples, 205–209
Consistency, 30–33, 158–159
coefficient of, 33
Constant differential growth ratio (*see* Coefficient, of heterogony)
Contingency classifications, 286–289
Contingency tables, 286–289, 295–296
Continuous variates (*see* Variates)
Cook, H. J., 403
Coolidge, J. L., 400
Coordinates, 306–310
angular, 307–308
arithlog, 306–308
arithmetic, 306–310
deformed, 333–335
logarithmic, 306–308
Coordinates, polar, 307–308
rectangular, 306–310
semilogarithmic, 306–308
Correction factor, for arithmetic mean, 90
Sheppard's, for fourth moment, 147*n*.
for standard deviation, 116–117
Correlation, 222–254
coefficient of, 224–254
comparisons of coefficients, 242–243
partial, 245–250
product moment, 224–237
rank, 250–254
relationship to regression, 270–272
significance of, 233–237
spurious, 243
standard error of, 234, 241–242
transformation of coefficients, 237–240
Correlation ratio, 281–284
comparison with correlation coefficient, 284
Correlation surfaces, 306, 312
Coyote, 314
Crab, 371
Crocodiles, mensuration of, 22
Cumulative distributions, 82–84
Curve fitting, 140–142
Curves, asymmetrical, 64–66
of distributions, 59–75
growth, 351–354
logistic, 354*n*.
ogive, 82, 84
skewed, 64–66
(*See also* Normal curve)
Curvilinear regression, 280–284

D

Data, numerical, from direct observation, 3–7
Daucus carota, 368
Davenport, C. B., 400
Davenport, D. H., 400
Deciles, 120

Deer, white-tailed, 124
Deer-mouse, 108
Deformed coordinates, 333–335
Dendrohydrax dorsalis emini, 125
Dependence, hypotheses of, 300–304
Deviations, significance of, 191
  in small samples, 206–209
  (*See also* Mean deviation; Standard deviation)
Diagrams, types of, 305–310
  area, 306
  bar, 306, 312–314, 324–325, 329–330
  line, 305
  pictorial, 306, 330–335
  pie, 306, 325–328
  point, 305, 329–330
  pyramid, 314
  scatter, 232, 255–258, 265, 305, 312, 329–330
  sector, 325*n.*
  three-dimensional, 306
Dice, L. R., 108, 178*n.*, 195, 201, 316–317, 321*n.*, 326, 400, 402
Dice-Leraas method of comparisons of means, 316–321
Dicerorhinus sumatrensis, 334
Dimensions, linear, 4
Diodon, 333
Discontinuous variates (*see* Variates)
Dispersion, measures of, 106–128
  relative, 121–128
Distributions, binomial, 75–78
  cumulative, 82–84
  J-shaped, 66–68
  patterns of, 59–75
  qualitative, 34–35
  quantitative, 35–37
  U-shaped, 69
  (*See also* Curves; Frequency distribution)
Dixon, J., 314, 402
Donaldson, H. H., 354*n.*, 402
Dot diagrams, 56–57
Duerst, V. V., 24, 402

E

Eaton, R. J., 309, 402
Efficiency, 158–161
Ehrenberg, K., 402
Elephant, 332
Equus, 369
Error (*see* Probable error; Standard error)
Exocoetus obtusirostris, 293

F

Faunal sampling, 174–179
Fisher, R. A., 78, 113*n.*, 132*n.*, 280*n.*, 290*n.*, 296, 400
Fishes, mensuration of, 21
Flying-fish, 293
Formulas, 382–390
  for machine calculation, 379
Freedom, degrees of, 290–292
Frequencies, 2, 4, 6–7
Frequency distributions, 34–37, 45–48
  graphic representation of, 56–62
  patterns of, 56–84
  tests of agreement between observed and theoretical, 301–302
  (*See also* Curves; Distributions)
Frequency polygon, 57–61, 305, 310
Frequency surface, 312
Frost, F. H., 196, 402

G

Gaindatherium browni, 334
Galton, F., 256*n.*,
Gauss, K. F., 70*n.*
Gaussian curve, 70
Geissler, 78
Geometric growth, 346–353, 355–358
Geometric mean, 102, 146
Gleason, H. A., 178*n.*
Glossary, 391–397
Glut-herring, 64

Graphic methods, 305–335
(*See also* Graphs)
Graphs, of attributes, 310, 312–314
of comparisons of ranges and means, 314–321
of frequency distributions, 56–62
of more than one variable, 310–314
of percentages, 321–328
of periodic phenomena, 328–330
of proportions, 321–328
of relative frequencies, 310, 321–328
of time series, 328–330
of tint photometry, 325–326
Green, C. K., 369, 402
Gregory, W. K., 21, 402, 403
Grinnell, G. B., 178*n.*
Ground squirrel, 297
Group limits, 37–44
Grouping, 34–55
by ages, 36–37, 336–342
criteria for secondary, 50–55
numerical qualitative, 48–50
secondary, 37, 50–55
Growth, 336–373
absolute, 344–346
accretive or additive, 345–346, 350–353
arithmetic, 345–346, 350–353
compound-interest, 346–353
geometric, 346–353
heterogonic, 358
isogonic, 358
multiplicative, 346–353
of populations, 354*n.*
principal elements of, 344–345
relative, 354–373
sampling for, 342–344
Growth curves, 322–325, 351–354
Growth gradients, 372–373
Growth rates, 344–354
arithmetic or additive, 346, 350–353, 355, 358
geometric or multiplicative, 346–353, 355–358
instantaneous, 348

H

Harmonic mean, 102
Haskell, A. C., 400
Hatt, R. T., 125, 403
Hedrick, E. R., 400
Hemidactylium scutatum, 339–340
Herring gulls, 309
Heterogeneity of samples, 197–202
Heterogony, 367–373
coefficient of, 367–373
calculation, 363–366
(*See also* Growth)
Hicks, L. E., 35, 403
Hildebrand, S. F., 64, 175, 313, 403–404
Histodifferentiation, 370
Histograms, 57–61, 306, 311–312
Hogben, L., 400
Homogeneity of samples, 168–170
association tests of, 303–304
Horse, 332
Howard, H., 196, 403
Hue, E., 10*n.*, 403
Huestis, R. R., 304, 403
Huxley, J. S., 355, 367–373, 403
Hydractinia echinata, 368
Hyracotherium, 369

I

Increments, in growth, 346–353
Indices, 10–13
Instantaneous growth rate, 348
Iridoprocne bicolor, 46, 294
Isogony, 367
(*See also* Growth)

J

J-shaped distributions, 66–68
Jenkinson, J. W., 350–352, 403

K

Kälin, J. A., 22, 403
Kelley, T. L., 400
Kemp, S., 371

Kenoyer, L. A., 178*n*.
King-snake, 47, 226
Klähn, H., 403
Kurtosis, 75, 147–148
  coefficient of, 147

L

Lampropeltis elapsoides elapsoides, 49, 242
Lampropeltis getulus getulus, 47, 54
Lampropeltis polyzona, 226, 228, 231–232, 242, 279
Laplace, P. S., 70*n*.
Laplace's normal curve, 70
Lemoine, V., 216
Leptokurtic distribution, 147
Leraas, H. J., 316, 321*n*., 400
Limits, of groups or classes, 37–44
Line diagrams, 305
Linsdale, J., 178*n*.
Litolestes notissimus, 80–81, 126, 145
Lizards, mensuration of, 21–22
Logarithmic coordinates, 306–308
Logistic curve, 354*n*.
Loomis, F. B., 342
Low, S. H., 46, 294, 403

M

Macdonell, W. R., 144, 404
Mammals, mensuration of, 22–23
Manatee, 46
Matthew, W. D., 35, 341–342, 404
Mean, arithmetic, 85–94, 146
  geometric, 102, 146
  graphic comparisons of, 314–321
  harmonic, 102
  quadratic, 102–103
  significance of, 209–210
  standard error of, 151–152, 154
    differences of, 192–197, 317–321
Mean deviation, 109–112, 117–118
  standard error of, 155
Measurement, requirements of good, 16–21
Median, 94–99
  standard error of, 154
Meek, S. W., 175, 404
Meleagris gallopavo, 196
Melospiza melodia beata, 58, 61
Mendelian ratios, association as test of, 363
Mensuration, 16–33
  of birds, 22
  of color, 24–25
  of crocodiles, 22
  of fishes, 21
  of lizards, 21–22
  of mammals, 22–23
  osteological, 24
  paleontological, 23–24
  of snakes, 21–22
  systems of, 21–25
    (*See also* Measurement)
Merychippus primus, 341
Midpoints, of groups, 37–43
Mills, F. C., 400
Miner, J. R., 401
Mode, 99–101, 146
Modules, 14–15
Moments, 147*n*.
Mook, C. C., 365–370, 404
Multiple association, 299–300
Multiplicative growth (*see* Geometric growth)
Mus bactrianus, 369

N

Nice, M. M., 58, 404
Nichols, J. T., 21, 68, 404
Nondecimal units of measurement, 42–44
Normal, meaning of word, 103–105
Normal curve, 70–75, 129–148
  areas of, 135–140
  mathematical equation of, 132
  ordinates of, 135–142
  parameters of, 132–135
  table, of areas, 137
    of ordinates, 139
Notostylops, 212
Numbers, continuous, 162–164
  discontinuous, 162–164
  meaning of, in zoology, 1–3
  sorts of, 1–2

Numerical data, types and properties, 1–15

O

Odocoileus virginianus borealis, 124
Ogive curve, 82, 84
Ordinates, of normal curve, 135–142
  table of, 139
Orthagoristus, 333
Osborn, H. F., 24, 331–332, 335, 404
Osteology, mensuration, 24

P

Palaemon bengalensis, 371
Paleontology, mensuration, 23–24
  sampling in, 179–181
Papio porcarius, 369
Parameters, families of, 134–135
  of normal curve, 132–135
  of populations, 149–153
  rule for significant figures of, 165
  of samples, 149–153
Parapavo californicus, 196
Parexocoetus brachypterus hillianus, 65–66, 69
Parker, G. H., 46, 404
Partial correlation, 245–250
Pearl, R., 184*n.*, 315, 354*n.*, 401
Pearsall, W. H., 368
Pearson, H. S., 190, 404
Pearson, K., 122, 281, 284*n.*, 401
Percentages, graphic representation of, 321–328
Percentiles, 120
Periodic phenomena, 328–330
Perissodactyls, 331
Peromyscus maniculatus, 201
Peromyscus maniculatus artemisiae, 322
Peromyscus maniculatus bairdii, 108, 140, 195, 317, 326
Peromyscus maniculatus gambelii, 20, 32
Peromyscus maniculatus rubidus, 304
Phenacodus primaevus, 157, 230, 259, 261
Phillips, J. C., 124, 404
Photometry, tint, 25
  graphic representation of, 325–326
Pictorial diagrams, 306, 330–335
Pie diagrams, 306, 325–328
Pisum sativum, 368
Platykurtic distribution, 147
Plesiadapis gidleyi, 215
Plesiadapis tricuspidens, 215–216
Point diagrams, 305, 329–330
Poisson, S. D., 79*n.*
Poisson series, 78–83, 161
Polar coordinates, 307–308
Pomolobus aestivalus, 63–64, 82, 90
Pooecetes gramineus gramineus, 208
Populations, 166
  growth of, 354*n.*
Probability, 149–165
Probable error, 134, 153
  significance of deviations in terms of, 191
  (*See also* Standard error)
Product moment correlation, 224–237
Proportions, anatomical, graphs of, 332–335
  graphic representation of, 321–328
  in growth, 354–373
Ptilodontidae, 198
Ptilodus, 201, 287
Ptilodus montanus, 41, 46, 86, 289, 303
Pyramid diagrams, 314

Q

Quadrat sampling, 176–177
Quadratic mean, 102–103
Quartile deviation, 120
Quartiles, 119–120
  first, 119
  standard error of, 155
  third, 119
Quetelet, L. A. J., 62*n.*, 73
Quetelet's principle, 62–63, 85
Quintiles, 120

R

Rainbow trout, 350–352
Range, graphic comparisons of, 314–321
  observed, 107–109, 135
  real, 109
  semi-interquartile, 119–120
  theoretical, 216–218
Range midpoint, 102
Rank correlation, 250–254
Ratio method in association, 294–296
Ratios, 8–14, 272–273, 357–363
  dependence on growth, 363
Raunkiaer, C., 178*n*.
Rectangular coordinates, 306–310
Refinement, 25–30
Regression, 255–284
  coefficients of, 262–274
  curvilinear, 280–284
  difference between coefficients, 277–280
  rough approximation of, 258–262
  significance of, 274–280
  standard error of, 276–277
Relative dispersion, 121–128
Reliability, 149–165
  coefficient of, 33
Rhea, 69
Rhinoceros unicornis, 334
Ridgway, R., 22, 24, 404
Rietz, H. L., 401
Robb, R. C., 354*n*., 358, 369, 404–405
Rosahn, P. D., 321*n*., 401
Rounding figures, 44–45
Ruthven, A., 315, 405

S

Salamanders, 339–340
Samples, characteristics of good, 167–174
  comparisons of, 186–202
  heterogeneity of, 197–202
  homogeneity of, 168–170
  small, 203–221, 296–298
    (*See also* Sampling)
Sampling, 166–185
  criteria of good, 167–174
  faunal, 174–179
  for growth, 342–344
  paleontological, 179–181
  quadrat, 176–177
  qualitative, 174–175
  quantitative, 175–179
  skimming, 175
  specifications, for fossil, 174
    for vertebrate, 174
  station, 176, 178–179
  traverse, 176–178
    (*See also* Samples)
Sampling limits, 182–185
  table of, 184
Scatter diagrams, 232, 255–258, 265, 305, 329–330
  of three variables, 312
Schooley, J. P., 35, 329, 405
Schroeder, W. C., 64, 313, 403
Sector diagrams, 325*n*.
Semi-interquartile range, 119–120
Semilogarithmic coordinates, 306–308
Shaw, W. T., 297, 405
Sheep, 372
Sheep dog, 368
Sheppard, W. F., 147*n*.
Sheppard's correction, 147*n*.
Significance, of association, 289–304
  of correlation coefficient, 233–237
  criteria of, 190–192, 206–209
  of difference between correlation coefficient and correlation ratio, 284
  of means, 209–210
  of regression, 274–280
  of single observations, 212–221
Significant figures, 161–165
Simpson, G. G., 405
Single specimens, 212–221
Skewness, 64–66, 75, 143–146
  coefficient of, 143–145
  negative, 64–66
  positive, 64–66
Skimming sampling, 175
Snakes, mensuration of, 21–22
Snowshoe-hares, 68

Snyder, L., 331, 405
Soergel, W., 24, 405
Song-sparrow, 58
Spearman, C., 250
Sperry, C., 327, 405
Spurious association, 299–301
Spurious correlation, 243
Standard deviation, 112–119
  of binomial, 183
  in small samples, 205
  standard error of, 152, 155
Standard error, 134, 149–160
  calculation of, 154–158
  of coefficient of variation, 152–155
  of correlation coefficient, 234
  of difference, between correlation coefficients, 242–243
    between means, 192–197, 317–321
    between regression coefficients, 277–280
  of estimate, 274–276
  of mean, 151–152, 154
    in small samples, 210–212
  of mean deviation, 155
  of median, 154
  of quartiles, 155
  of regression coefficient, 276–277
  of standard deviation, 152, 155
Starling, 35
Station sampling, 176, 178–179
Stenomylus, 342
Stock, C., 405
Storer, T. I., 178*n*.
Sturnis vulgaris, 35
Subhyracodon occidentalis, 334
Sufficiency, 158, 161
Sumner, F. B., 20, 22, 32–33, 325, 405
Svihla, A., 322, 405
Symbols, 380–382

T

Tables (*see list in front of book*)
Tamias striatus, 35
Teilhard de Chardin P., 216
Teissier, G., 368
Temperature, 5–6
Thamnophis megalops, 315
Thompson, D'Arcy, 333, 406
Three-dimensional diagrams, 306
Time, periods, 6
Time series, 328–330
Traverse sampling, 176–178
Tree-hydrax, 125
Tree swallows, 46
Trend, 255–258, 305
Troxell, E. L., 10*n*., 12*n*., 406
Turkeys, 196
Typical, meaning of word, 103–105

U

U-shaped distributions, 69
Uca pugnax, 368

V

Variability, 121–128
  coefficient of, 122–128
Variables, 2
  dependent, 258–262
  independent, 258–262
Variance, 113–115
Variates, 2
  continuous, 3–6
  discontinuous, 3, 6–7, 45–48
  sampling for, 170
Variation, coefficient of, 122–128, 135, 146
  relationship, to percentage differences in observations, 219
    to theoretical range, 217
  standard error of, 152, 153
Vesper sparrow, 208
Volume, 4–5

W

Walker, H. M., 401
Wedekind, R., 406
Weight, 5

Y

Yates, F., 296
Yingling, H., 331, 405
Yule, G. U., 133*n*., 154, 312*n*., 401

Z

Zeuner, F., 24, 406
Zuckerman, S., 369

7 DAYS ONLY.