

Biological Measurements.—*Report of Committee* (Prof. J. S. HUXLEY, *Chairman*; Dr. R. A. FISHER, *Secretary*; Dr. W. T. CALMAN, Mr. C. FORSTER-COOPER, Prof. J. W. NICHOLSON, Dr. E. S. PEARSON, Mr. O. W. RICHARDS, Mr. G. C. ROBSON, Dr. J. F. TOCHER) appointed to draw up recommendations for the taking and presentation of Biological Measurements, and to bring such before persons or bodies concerned.

THE Committee held six meetings during 1926 and 1927. Preliminary discussions showed that two obstacles ordinarily stood in the way of the satisfactory presentation of numerical data in the biological literature. In the first place editors showed a natural reluctance to printing extensive data in full detail, especially when every advantage had not been taken to arrange such data as compactly as possible; in the second place the methods available for providing a statistical summary, such as is essential wherever the original data are not presented in full, are neither sufficiently well known nor have been sufficiently standardised by accepted conventions for such summaries to have an exact and unambiguous meaning.

The Committee decided that these obstacles could be overcome by action along two lines: (a) by the establishment of centrally placed archives for the reception of original biological data, which were too extensive for complete publication, and (b) by the preparation of a leaflet for the guidance of contributors to biological journals who wish to conform to acceptable modern practice. It is anticipated that this leaflet will require periodical revision as need arises.

Negotiations with the Natural History Museum at South Kensington and with the Royal Society of Edinburgh have resulted in the establishment of the required archives for the reception of biological data, where they will be available to students, and in this sense will have secured effective publication. The thanks of the Committee are due to the authorities of these two institutions for undertaking a function which in the opinion of the Committee will be of increasing value to biological science.

The leaflet prepared by the Committee consists of a foreword illustrating the practical needs of modern biological work, followed by four sections (A) on general considerations in the planning and execution of research by metrical methods, (B) on the methods available for the compact presentation of data, and on the recognised methods by which it can be adequately summarised, (C) on the interpretation of statistical results and on tests of significance, and (D) giving detailed references to text books upon the several types of tests generally required. The leaflet is presented as an appendix to this report.

Seeing that the practical utility both of the archives and of the leaflet will depend on their existence becoming known to biological workers, the Committee have circularised the editors of the chief biological journals published in Great Britain asking for the incorporation of additional clauses in their permanent notices to contributors. The Committee are glad to report that a favourable reply has been given by the editors of a number of important journals.

Recommendations of the British Association Committee on Biological Measurements.

FOREWORD.

Biology is rapidly becoming more and more of a science in which exact mathematical methods are required. In all fields accurate measurements or quantitative data of some kind are being increasingly employed. On the other hand, such data are not infrequently rendered useless, or at least much less useful than they might have been, through neglect of simple precautions, either in the making, the recording, or the analysis of the data. Bearing these facts in mind, Section D of the British Association appointed the present Committee to draw up recommendations upon the presentation of biological measurements.

The chief fields in which statistical data, properly taken and analysed, can be of great service are perhaps the following:—

(a) *Genetics.*—Obviously, here all conclusions based upon ratios are valid only in so far as statistically significant. In the early days of Mendelianism much confusion was brought about through lack of proper statistical treatment. It has, however, recently become increasingly realised that a combination of Mendelian and statistical (biometric) methods is in many cases necessary for full analysis. In human genetics the statistical method is the main available weapon.

(b) *Variation.*—Here the achievements of biometrics are too well known to need statement or comment. It should be pointed out, however, that in many cases a technically perfect biometric analysis may tell us less than it ought owing to inadequate selection of material. *E.g.* without experiment, biometric methods cannot tell us how much of a given variation range is genotypic, how much phenotypic. Only properly directed work on variation can give us much needed information as to the differences between different species as regards variability, the reasons for the differences, and the bearing of the facts upon evolutionary theory.

(c) *Systematics.*—With increased delicacy of systematic determination, measurements are becoming more and more important as a criterion of the distinctness of closely related species, sub-species or races.

(d) *Development.*—Only by taking large numbers of measurements will it be possible to discover the laws of relative growth of parts.

(e) *Evolution.*—As more perfect palaeontological series are obtained, accurate measurements of absolute and relative sizes of parts may enable us to establish simple laws of evolutionary growth and development comparable to those which are being obtained by similar methods in ontogeny.

Naturally the taking of quantitative data constitutes the essence of much of physiology; but we have here been concerned mainly with data which may be called statistical.

It may be as well to begin by enumerating a few of the cases in which neglect of simple precautions has made laboriously taken measurements of much less value than they might otherwise have been; for such examples will serve better than anything else to convince the working zoologist of the need for improvement. The defect may have lain in the failure to take the most suitable measurements, to record them adequately when taken, or to analyse them in the most desirable way.

A. NEGLECT OF BIOMETRIC ANALYSIS. (See also No. 2.)

1. In the preparation of Witherby's 'Handbook of British Birds' (London, 1922) considerable numbers of accurate measurements were made both upon the skins (usually twelve specimens) and eggs (usually 100 specimens) of a large number of species of birds. However, in recording these valuable data only the mean and high and low extreme variants were set down (in the case of skins, the mean was omitted). Presumably the main purpose of such measurements is to give the systematist help in distinguishing between closely related forms (sub-species, races, &c.). Even for this purpose, however, and especially when the ranges of two forms overlap, this method of record is markedly inferior to one giving mean and standard deviation. In addition, the recording of standard deviation would have enabled a wholly different and very interesting problem to be attacked, namely, the suggestion originally made by Darwin ('Origin of Species,' chapter ii) that wide ranging and abundant species and genera are more variable than scarce and local ones.

B. NUMBERS INADEQUATE FOR THE STATISTICAL CONCLUSIONS DRAWN.

2. Examples of the failure to realise the statistical invalidity of small numbers are frequent. *E.g.* Krüger (1920 and 1924, *Zool. Jahrb. (Syst.)*, 42, 289, and 48, 1) distinguishes closely related 'species' of Humble-bees by means of certain relative proportions of parts. Considering, however, that the maximum number of any one species measured is 25, and is often below 10, that the ranges frequently overlap, and that only mean, maximum and minimum are recorded, it may be doubted whether these quantitative results are at all significant.

3. If data are properly taken and recorded, failure to use suitable analyses can be remedied by subsequent workers. Nevertheless, conclusions based on unsatisfactory analysis often, as a matter of fact, become generally accepted, and it is then difficult to correct the error. The most frequent source of error is failure to discount chance

or random sampling. A well-known case of this is afforded by the paper of Pearl and Parshley (1913, *Biol. Bull.* 24, 205), who believed that they had clear evidence that in cattle the relation of time of insemination to the cyclical events of oestrus influenced the sex-ratio. Later investigation of a larger body of material, however, convinced them that their first result had been wholly due to chance (Pearl, 1917, *Maine Agri. Exp. Station Bull.* 261, (3), 130).

C. INCOMPLETE RECORD OF ADEQUATE DATA.

4. Very often the investigator is so much preoccupied with the solution of a particular question that he is content to record his data incompletely, provided that this will suffice for his special problem. He fails to remember that complete record may make it possible for later investigators to use his original data for the solution of quite new problems. A good example of this is afforded by the classical paper of Bateson and Brindley (*Proc. Zool. Soc.* 1892, 285) upon dimorphic organs. The authors were concerned to prove that in the beetle *Xylotrupes*, while the frequency-curve for body-length was of normal type, that for cephalic horn-length was bimodal; but that in the stag-beetle *Lucanus* both body-length and mandible-length showed normal frequency distributions. The frequency distributions of these four characters were therefore given singly. Since, however, two measurements were taken and recorded for each individual, it would have been possible to present not only the information immediately required, but also all information bearing on the correlation between body-length and appendage-length, by means of two-way tables. This information was later required by another investigator: luckily the original data had been preserved, and so new conclusions could be drawn (*J. Genetics*, 1927, 18, 45). A precisely similar failure to record by means of two-way tables is found in the paper by Djakonov (*J. Genetics*, 1925, 15, 201) on the bimodality of forceps-length in male earwigs (*Forficula*). Here again only lucky chance preserved the original data, which were then found to yield new results (*J. Genetics*, 1927, 17, 309).

5. Frequently not merely are data published in an incomplete way, but owing to lack of space or for other reasons are not published at all. The danger of this procedure may be illustrated by the benefits accruing from its converse. Haldane (*J. Genetics*, 1920, 10, 47) was able to demonstrate from Nabour's data on heredity in the grasshopper *Paratettix* (*J. Genetics*, 3, 141, and 7, 1) that two factors which the original investigator had thought to segregate independently were in reality linked. He expressly states that this would not have been possible if it had not been for the exceptional fullness of Nabour's records.

6. Duncker (1903, *Biometrika*, 2, 307) re-analysed the figures of Yerkes (1901, *Proc. Amer. Soc. Arts and Sci.* 36, 417), which involved the careful measurement of a number of characters on eight hundred Fiddler-Crabs (*Gelasimus*). Neither author published the data in full; and since they were utilised only for certain special purposes, the very fundamental growth-relations between the various organs were not brought out. Duncker himself points out that asymmetry of all the organs on the side of the large chela increases with absolute body-size, but does not tabulate the figures by size classes. It is therefore impossible to arrive at the laws of growth underlying the phenomena. Duncker calculates a number of correlation coefficients from which he deduces certain conclusions. The conclusions would have been much more firmly based, however, if the underlying growth-laws had also been established, as only by so doing can we hope to understand the biological, as opposed to the statistical, meaning of the coefficients. This therefore represents a failure not only to publish the data in full, but also to analyse the data sufficiently fully even for the purpose envisaged by the author.

D. FAILURE TO CHOOSE THE MOST SUITABLE MEASUREMENTS OR CONVENTIONS.

7. Sometimes data are less valuable than they should be because the points of reference used in making measurements are chosen arbitrarily instead of conforming to an accepted standard, or of being chosen with reference to their biological significance.

An example of the latter procedure is shown by two recent authors (Nomura, 1926, and Sasaki, 1926, *Sci. Report Tohoku Imp. Univ.* 2, 57 and 197) who have made elaborate measurements of a number of Molluscan shells, with a view to the analysis of relative growth of parts. The results, however, would have been more valuable if measurements had been made of the magnitudes needed for determining the mathematical growth relations of a Molluscan shell, as set forth for instance in D'Arcy Thompson's 'Growth and Form' (Cambridge, 1917), chapters xi and xii.

E. FAILURE TO MAKE THE MOST SUITABLE BIOLOGICAL ANALYSIS. (See also No. 6.)

8. A further example of failure to analyse data in the best way owing to lack of the most suitable preliminary biological method (the statistical method being wholly adequate) is afforded by a paper by Pearl, Gowen and Miner (1919, *Maine Agric. Exp. Sta. Ann. Rep.*), who in calculating the influence of bulls on the milk-production of their female descendants takes as a measure of the bull's performance: daughter's yield minus mother's yield. This clearly gives an undue advantage to bulls mated to cows of low milk-yield. The error here has practical consequences, since the market value of the bulls would be altered in relation to the verdict of the scientist.

9. As Klatt (1919, *Biol. Zentralb.* 39, 406) points out, failure to realise that other relations than that of simple proportionality may, and usually do, hold between the size of an organ and the size of the whole organism vitiates many discussions as to the relative size of organs in different types within one group. The usual plan is to express relative organ-size as a percentage of total size. Since, however, a frequent relation of organ to body is not $y = ax$, but $y = ax^b$, this is of no value. Parrot (1894, *Zool. Jahrb. (Syst.)* 7) had arranged a series of birds in a scale according to their percentage heart-weights. Klatt, having previously found that the heart-weight (h) of warm-blooded vertebrates was related to the body-weight (w) according to the exponential formula $h = a.w^b$ where a varied considerably, but b was always close to 0.83, re-analysed these figures, and was thus enabled to calculate the real relative heart-weight, which is given by the size of the fractional constant a in the above formula. Thus, for instance, the stork has a moderately low percentage heart-weight, but this is due to its large absolute size. When the value of a is calculated by the correct method, its true (physiological) relative heart-weight turns out to be one of the three highest in the list.

10. In general, measurements reveal the fact that in many groups there are no final fixed proportions of parts (e.g. many Crustacea), and that the only quantitative constants which are of value are not percentages but exponents. This is true even of certain Mammals ('Monograph of the Voles and Lemmings living and extinct,' M. A. C. Hinton, vol. i, 1926).

D'Arcy Thompson ('Growth and Form,' chapter ii) gives an historical and critical account of many similar cases where absolute size must be taken into consideration in assessing the functional meaning of particular relative sizes of parts.

A. General Considerations.

1. IDENTIFICATION.

The material under investigation should be examined by exact taxonomic methods and care should be taken that the series of specimens dealt with are, so far as possible, correctly identified. The advice of an expert in the group under consideration should be sought if necessary.

2. CHARACTERISTICS OF THE POPULATION SAMPLED.

The examination of a sample only supplies direct information respecting the population as sampled by the methods of collection employed, or, in other words, the population from which such a sample may be regarded as fairly drawn at random. This may often differ materially from:—

- (i) the whole population living at the time of capture, owing, for example, to selection of sex, age or size by the methods of capture;
- (ii) the average population ordinarily living in the same habitat, owing, for example, to seasonal or other periodic fluctuations; and
- (iii) the populations of different habitats in the same region.

The results of the examination of a sample should therefore be supplemented with all possible care by information designed to specify the population sampled, even though such specification is undoubtedly often difficult. The aim should be that any significant (see Section D) discrepancies between samples obtained by different investigators should be assignable to their true causes, whether age, sex, local variation, time, season, method of capture, &c.

These should always be specified where possible, but in every investigation special points will need to be considered.

3. CONFORMITY TO PREVIOUS MEASUREMENTS.

Whatever other measurements may be made, the value of the work for comparative purposes will often be much increased by the inclusion of measurements which are comparable, as strictly as possible, with those taken in the same or related species by previous workers.

It is desirable that some quantitative measurements should always be presented with general biological data. Length measurements are the most usual. However, a frequent 'failure to record' is seen in microscopical figures to which no record of magnification is appended. For example, in the article on *Rotifera* in the Cambridge Natural History and in the Encyclopædia Britannica no magnification is given in any of the figures, nor are any measurements given in the text, so that the reader (*inter alia*) will not be told, nor able to find out for himself, the interesting biological fact that the *Rotifera* have the lowest average size and the smallest size range of any considerable Metazoan group.

The most satisfactory way of giving magnifications is to reproduce with the figure some unit of length magnified to the same scale. This obviates the error which frequently creeps in when figures from one source are reproduced in another publication on a different scale, but without altering the statement as to magnification in the legend.

In addition measurements of weight or volume should be made whenever possible as a matter of routine, since they provide the best standard of quantitative comparison between differently shaped organisms or organs.

4. SPECIFICATION OF PRECISE CONVENTIONS.

It is essential to specify the conventions, including any points of reference adopted, by which each measurement is defined. This can often best be done by the aid of a diagram. When satisfactory standard terms, conventions or points of reference already exist they should be adopted whenever possible. The aim should be to ensure that a second observer, working over the identical material, and guided only by such specifications, should normally obtain significantly similar results.

The state of preservation of the material may often affect the measurements, especially in the case of soft parts. Accordingly the method of preservation and the degree of contraction or relaxation of the parts should be noted.

Observations of colour should when possible be referred to one of the standard scales in general use, e.g. 'Nomenclature of colours for naturalists,' R. Ridgeway, U.S. National Museum, 1912; 'Code des couleurs à l'usage des naturalistes, artistes, commerçants et industriels,' P. Klinksieck, Paris, 1908.

5. TESTS OF SIGNIFICANCE.

The critical stages of the statistical examinations of a body of data are reached in the application of what are known as tests of significance. (See Section D, 3-8.) These are essentially tests whether the difference between two (or the variance among several) groups is or is not greater than can with reasonable probability be ascribed to the variability found within each group. The ultimate value of the conclusions to be drawn from any data depends upon the precision and validity with which such tests can be carried out; consequently it is advisable that investigators, whether or not they undertake the work of statistical analysis, should have a general acquaintance with the nature of such tests, and, where the case does not seem clear, should seek the advice or co-operation of a statistician.

B. Presentation of Data.

1. An incomplete specification of a sample is never to be preferred to a complete specification, e.g. greatest, least and mean length is an incomplete specification (see below).

2. SINGLE-VARIATE DATA.

For a single measurement a complete specification of a sample may be given by recording the number of cases observed to fall in successive intervals of magnitude.

ON BIOLOGICAL MEASUREMENTS.

Example: Length of cuckoo's egg (after O. W. Lister).

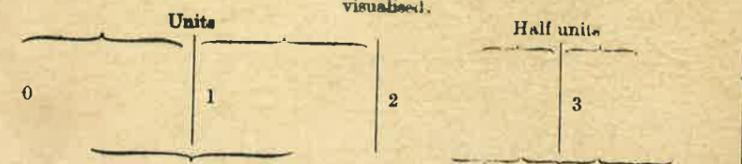
Length class, mm.	19.0	19.5	20.0	20.5	21.0	21.5	22.0	22.5	23.0
Frequency	1	3	21	76	152	152	392	288	286
Length class, mm.	23.5	24.0	24.5	25.0	25.5	26.0	26.5		
Frequency	100	86	21	12	2				
									Total
									1572

A series of numbers arranged in this way form what is called the frequency distribution of the sample.

The total of 1572 eggs is distributed in 16 length classes, each with a range of half a millimetre, each class being specified by its central length. Thus the entry under 21.5 mm. indicates that 152 of the eggs measured were judged to lie between the precise limits 21.25 and 21.75 mm. The class range need not be equal to the unit of measurement, but should be (either one unit or) an integral number of such units; the table above was condensed from a record giving the length to 0.1 mm.

A fruitful source of bias is avoided, at the time the measurements are actually taken, by using length classes bounded by the divisions marked on the measuring instrument used, instead of the more common practice of using length classes centred on visible divisions, and bounded by imaginary ones. The effect of the latter procedure appears to be especially noticeable in micro-measurements. If working with length classes of 1 mm. adopt class boundaries of 0-1, 1-2, 2-3 mm., &c., with class centres at 0.5, 1.5, 2.5 mm., &c. If working with length classes of 0.5 mm. adopt class boundaries of 0-0.5, 0.5-1.0 mm. &c., with class centres at 0.25, 0.75, 1.25 mm., &c.

Measurement groups free from bias, bounded by divisions which can be accurately visualised.



Groups usually employed, centred on divisions which can be accurately visualised, but bounded by imaginary divisions.

The use of small units is less important than accuracy of the class boundaries, and it is above all essential that these boundaries should be clearly indicated. For example, headings such as these are ambiguous:

Age	6 years	7 years	8 years
Frequency	15	38	62

It is impossible to tell whether the 38 individuals were between 7.0 and 8.0, or between 6.5 and 7.5; the former interpretation adhering to the popular convention of age, the latter to the scientific convention of specifying the central measurement of each class.

In the choice of the class interval, which should be uniform throughout, little additional information is supplied by a very fine classification; for material which is apparently homogeneous a class interval equal to a quarter of the Standard Deviation is sufficiently small; this will usually be provided for by dividing the material into about 20-25 classes. Coarser groupings are by no means valueless. To bring out the peculiarities of heterogeneous material finer grouping will sometimes be required. Small samples should not be grouped more coarsely than large samples. Extreme measurements should never be pooled as, e.g., 'more than 25 mm.'; since in the statistical treatment the precise determination of these is of particular importance.

3. SUMMARY OF SINGLE MEASUREMENT DATA.

If space does not allow a complete specification of the observations, these may be summarised by means of a few quantities calculated from them; each of these quantities is technically termed a *statistic*. If this course be taken, great care and some additional knowledge will be needed to make the summary adequate. For instance, the mean and range of the lengths of the individuals of a sample contain only a small

class w_6 , in the largest number of cases (101.5) the 'second egg' is also in this class and the next largest numbers are in the neighbouring classes w_5 (95.5) and w_7 (75.75).

' Value ' of ground-colour in first egg.								
	w_2	w_3	w_4	w_5	w_6	w_7	w_8	Totals
w_2	3	8	2	1.75	1.5	.75	—	17
w_3	8	44	40	19	6	6	—	123
w_4	2	40	135	68	29.5	12	2	288.5
w_5	1.75	19	68	153	95.25	41	9	387
w_6	1.5	6	29.5	95.25	101.5	75.5	13	322.5
w_7	.75	6	12	41	75.75	126.5	28	290
w_8	—	—	2	9	13	28	30	82
	17	123	288.5	387	322.5	290	82	1510

The mere arrangement therefore in the table brings out the similarity in colour-value between eggs in the same clutch.

With pairs of measurements the same considerations as to class interval should be applied, as with single variates, save that with close associations a finer grouping may be required to make the class interval as small as a quarter of the average standard deviation of the arrays (separate rows or columns).

5. SUMMARY OF A TWO-WAY TABLE.

In all cases where two characters are considered the results can be displayed most compactly in a two-way table, and this forms a convenient basis for the calculation required if it is wished to study the relationship between them. Where the data are in the form of numerical measurements, as in the first example, and if they conform to what is termed a Normal Correlation distribution, the contents of the table may be described by five quantities. These, in addition to being useful in themselves, will serve as a statistical summary of the data when the two-way table cannot be presented in full. These are (i) and (ii) the means and (iii) and (iv) the variances of the two marginal distributions, and (v) the 'product-moment coefficient' which is calculated like the variance from the deviations from the means, using the products of the deviations of the two variates (having regard to the positive and negative sign of these deviations) instead of the squares of the deviations of a single variate. From the three latter statistics any of the following may be at once obtained, one or other of which will in almost all cases be of importance in the interpretation of the data. Denoting the two variates by x and y ,

(a) the regression coefficient of y on x is the ratio of the product moment coefficient to the variance of x ; here x is regarded as the independent variate, and y as dependent upon it;

(b) the regression coefficient of x on y is the ratio of the product moment coefficient to the variance of y ;

(c) the coefficient of correlation is the geometric mean of the two regressions, and may be found either from them, or by dividing the product moment by the geometric mean of the two variances, or by the product of the standard deviations. (Section D, 2.)

This method of description becomes inadequate if the material differs markedly from Normal in the form of its distribution; in such cases the two-way distribution table should not be replaced by a summary.

6. MORE THAN TWO VARIATES.

There is no compact form for the complete publication of sets of three or more measurements. When the number of individuals is not too great, these may be set out *seriatim*, each occupying a line of the table. If the number of class combinations is sufficiently small, which can only occur if very broad classes are employed, it may happen that the class combinations and the corresponding frequencies of occurrence can be compactly listed. For storage in a form ready for immediate use, cards are recommended, each card representing an individual with its numerical measures

entered in corresponding positions on the different cards. A key card should always be prepared giving the significance and units of the several entries on the individual cards.

An incomplete but valuable record of a large number of individuals measured in more than two characters is provided by the preparation of every possible two-way table. Thus with seven variates, twenty-one tables will specify the simultaneous distribution of the samples for every pair of variates. Such a record, though incomplete (because it does not specify which values of all seven characters were associated together in an individual, but only considers them in pairs), will yet provide a basis for all calculations ordinarily conducted.

7. GRAPHIC METHODS.

Diagrams should be freely used in exploring the character of the relationship between two closely related variates. In plotting two sets of values against each other, we may take absolute values, or the reciprocals of the absolute values of one or both, or the logarithms of one or both, and so forth. If a straight graph is obtained by any one of these methods, it suggests a particular type of mathematical relationship, the recognition of which may facilitate the detection of the biological process or mechanism involved.

Diagrams provide no adequate substitute for the tabular presentation of data, or for the critical tests necessary to examine their conformity with the hypotheses they suggest. In the publication of results their purpose is to illustrate and make plain particular facts selected for emphasis by the author, and not to establish such facts. It is not necessary to publish every diagram which has proved useful in studying the data.

C. The Interpretation of Results and Tests of Significance.

1. In carrying out any statistical analysis it is necessary to bear in mind the distinction between the following:—

- (1) The population which has been sampled.
- (2) The true measurements of the sample available.
- (3) The measurements of these individuals as recorded.

Provided that the specification is adequate and that the errors of measurement are small compared with the real biological variation among the individuals of the sample, it may be assumed that (3) provides no adequate description of (2). The problem that remains is to consider what may be inferred legitimately from the measurements (3) regarding the population (1). It needs little experience to realise that the average measure of some character found in a sample, or the percentage of individuals falling into certain groups, may often differ considerably from the values in the population sampled, and further that two samples will themselves often differ considerably from one another. The problem is therefore to obtain criteria which will enable a judgment to be formed as to whether the variation in a sample is of statistical significance (see Section D, 3-8); or is not more than might be expected to arise from the chance fluctuations of random sampling.

By mathematical analysis it has been found possible to determine the variation due to random sampling of some of the most important descriptive measures or statistics, such as the mean or the standard deviation of a series of observations. A definite measure of probability can therefore be assigned to the occurrence of a particular value of the statistic in a random sample. In general, the procedure is to calculate the ratio of (a); the difference between the statistic and the quantitative character of the population of which it is the estimate or between the corresponding statistics in two samples, to (b), the Standard Error or an estimate of the Standard Error of that difference, and then to obtain the probability from the appropriate table.

The nature of the problem can be indicated most readily by considering two typical examples.

(1) Suppose there to be a population of individuals whose frequency distribution for measurements of a single character is Normal (Section D, No. 4); and that the mean measurement is known to be 22.56 cm. while the standard deviation is 1.54 cm. Then it is possible to state by reference to the appropriate table that only in about two cases out of one thousand should we expect to find a mean of 23.56 cm. or more in a random sample of twenty individuals. Or supposing that the only available

information to be contained in the sample of twenty, with mean of 23.56 cm. and a standard deviation of 1.44 cm. by the use of the appropriate probability table, we can assign a measure of probability of 7 in 1000 for the mean in the population sampled lying outside the range 22.56 cm. to 24.56 cm. (Section D, No. 5b). It follows that, in whichever way the problem is presented, one is justified in concluding that it is most unlikely that the difference between the sample mean (23.56) and the population mean (22.56) can be due to the chance fluctuation of sampling. It is therefore what is termed a significant difference, the cause of which must be sought elsewhere.

(2) Another type of illustration is as follows:—

Suppose that in a Mendelian experiment there are theoretical reasons for expecting ratios of 2:1:1 in three frequency groups. In a sample of forty the following frequencies are observed: 22, 7, 11. There it is possible to say that a divergence from the 'expected' frequencies of 20, 10, 10 as great or greater than that observed will occur in the long run on fifty-five random samples out of one hundred, or in other words that the divergence is not at all exceptional (Section D, No. 8).

2. THE DISADVANTAGE OF SMALL SAMPLES.

In both the examples that have been given, the samples contained only a small number of observations. If the distribution of the character or characters in the population is known it is possible to obtain a measure of the probability of drawing a given sample, however small that sample may be. But when the sample only is known, the nature of the population can be inferred with far less precision from small than from large samples. The difficulties in interpretation to which this may lead can again be illustrated by the previous examples.

(1) Suppose that in addition to the sample of twenty observations with mean of 23.56 cm. and standard deviation of 1.54 cm., there is a second sample of twenty-five with mean 23.14 cm. and standard deviation 1.61 cm. Then in neither case is it possible to estimate the mean of the population sampled with sufficient precision to conclude that the two samples have been drawn from different populations. The position may be put into exact terms by stating that on the evidence available a difference, one way or the other, between the means as great or greater than the observed 0.42 cm. would occur in thirty-seven cases out of one hundred in the random drawings of two samples from the same population.

If, however, the samples had been each ten times as large, viz. 200 and 250, it would have been possible to obtain a more precise estimate of the populations sampled, and to infer that they had almost certainly different means. This can be expressed by saying that for these larger samples a difference in means of 0.42 cm. or more would be expected to be found in only five cases out of one thousand (Section D, No. 6).

(2) If two possible hypotheses existed as to the Mendelian ratios, viz. either 2:1:1 or 9:3:4, the evidence provided by the sample figures 22, 7, 11 would be quite inadequate to distinguish between them. It has been seen that the odds are 55 to 45 in favour of obtaining so great a divergence from the expected numbers on the first hypothesis, and for the second hypothesis the corresponding odds are 925 to 75 in favour. These figures would not justify the acceptance of one hypothesis rather than the other, for the observations are not improbable on either hypothesis.

If, however, the sample had been of 400 and the group frequencies 220, 70, and 110, it is found that samples with as or more divergent frequencies would only occur:

- (a) in 25 cases in 10,000 if hypothesis 2:1:1 were true;
- (b) in 49 cases in 100 if hypothesis 9:3:4 were true.

It follows that now the evidence is sufficient to show that the first hypothesis is quite improbable, while the second is still in reasonable accordance with the facts (Section D, No. 8).

The statistical methods available thus allow one to test the validity of various hypotheses in relation both to the nature and to the extent of the data presented. An increase in the number of observations will usually increase the precision of all tests, and may justify conclusions which would otherwise be doubtful. The size of the sample is not, however, always a mere matter of the number of individuals measured. Each unit may be a district, a season, or a complete and lengthy experiment, and for such cases the more exact methods appropriate for small samples will be particularly necessary.

D. Notes on Methods and References.

The greater part of the procedure and methods of analysis indicated below was given in memoirs which appeared in journals such as *The Philosophical Transactions of the Royal Society*, *The Philosophical Magazine*, *Biometrika*, *The Journal of the Royal Statistical Society*, *Metron*, &c. The fullest collected information is probably contained in the following books, for which detailed page references are given below:—

- (A) A. L. Bowley. 'The Elements of Statistics.' King & Son, London. 1920.
- (B) R. A. Fisher. 'Statistical Methods for Research Workers.' Oliver & Boyd, Edinburgh. 1925.
- (C) T. L. Kelley. 'Statistical Method.' Macmillan, New York. 1923.
- (D) G. U. Yule. 'An Introduction to the Theory of Statistics.' Griffin & Co., London. 1927.

(1) Computation of the Arithmetic Mean, Variance and Standard Deviation.

- (A) pp. 251-255. (B) pp. 48-54. (C) pp. 45-48; 77-82.
- (D) pp. 108-113; 134-141.

(2) Definition and Computation of the Product Moment, Coefficient of Correlation, and Regression Coefficients.

- (A) pp. 350-355; 380-383. (B) pp. 114-125 for regression.
pp. 146-150 for correlation.
- (C) pp. 161-164. (D) pp. 157-188.

(3) The Standard Error and the Probable Error.

The standard error of a descriptive constant or statistic measures the amount of variability to be expected between the values of that quantity found in different samples of the same size drawn at random from the same material. If σ^2 represent the variance in the population, then σ/\sqrt{N} is the standard error of the arithmetic mean for samples of N individuals. Since in the study of natural variation the variance of the population is unknown, we must use instead the variance as estimated from a sample or group of samples, taking care to employ methods which made due allowance for the sampling errors so introduced.

Owing to the fact that if the distribution of a variate be Normal, 50 per cent. of the observations will lie within the range taken from $.6745 \times$ standard deviation below to a corresponding distance above the mean; this multiple of the standard deviation has been termed the Probable Error. Thus $\pm .6745 \sigma/\sqrt{N}$ is termed the probable error of a mean, by which it is implied that the means in samples of N will as often fall inside as outside limits greater and less than the population mean by $.6745 \frac{\sigma}{\sqrt{N}}$. As it is always necessary to find σ or an estimate of σ before the probable error can be calculated, it is always simpler, and more conformable to modern practice, to measure variation by the standard error rather than the probable error.

(4) Tests for Normality.

The distributions given above of the length of Cuckoo's egg and the length and breadth of Tern's egg may be taken as roughly representing the Normal form. This is typified by a central concentration about the mean and a symmetrical tailing off of the frequency for positive or negative deviations, in accordance with a definite mathematical law.

In dealing with samples containing only a few observations, it is only possible to detect wide deviations from the normal form. The test, which is sensitive only for large samples, involves the calculation of the third and higher moments followed by a comparison of the values obtained from the sample with those to be expected from a normal distribution having the same variance.

- (B) pp. 54-56 or *Phil. Trans.* vol. 198A, 1902, p. 278.

(5) The Significance of a Mean.

(a) Population variance known, or estimated from a large sample.
The test consists in entering the tables of the Normal Probability Function with the ratio of (a), the deviation of the sample mean from the population mean, to (b), the standard error of the mean.

(A) pp. 415-416. (B) pp. 101-103. (C) pp. 82-83. (D) pp. 344-345.
Tables. 'Tables for Statisticians and Biometricians,' Cambridge University Press. Table II.

(B) Table I. (C) Appendix C.

(b) Sample variance only known.
Here the ratio of (i) the deviation of the sample mean from the population mean to (ii) the estimated standard error of the mean is taken, and the tables of the 't' distribution are used.

For large samples this test becomes the same as (a) for all practical purposes.

(B) pp. 106-108.

Tables: (B) Table IV. *Metron*, vol. v, 3, 1925.

(6) The Significance of the Difference between the Means of Two Samples.

The procedure is to enter the appropriate probability tables with the ratio of the difference between the means to the standard error, or estimated standard error, of that difference.

(a) Test applicable in the case of large samples from two populations in which the variance may differ.

(B) pp. 103-105. (D) pp. 345-346. Tables as for 5 (a).

(b) Test, sensitive for difference between the means, as to whether two samples can be regarded as drawn from the same population, accurate for small samples.

(B) pp. 109-113. Tables as for 5 (b).

(7) The Significance of the Difference between the Variance of Two Samples.

Test whether two samples can be regarded as drawn from populations of equal variance.

(B) pp. 192-200. Table VI., p. 210.

(8) The χ^2 Test for Goodness of Fit.

(a) Comparison of the observed frequencies with those expected theoretically in the corresponding groups or classes.

(A) pp. 426-433. (B) pp. 77-84. (C) pp. 370-387.

Tables. 'Tables for Statisticians and Biometricians,' Table XII.

(B) Table III.

(b) Analogous test of agreement between two or more observed frequency series.

(B) pp. 94-95. *Biometrika*, vol. viii, pp. 250-254.

Tables as for 8 (a).

Geography Teaching.—*Report of Committee* (Prof. T. P. NUNN, *Chairman*; Mr. W. H. BARKER, *Secretary*; Mr. L. BROOKS, Prof. H. J. FLEURE, Mr. O. J. R. HOWARTH, Mr. J. MCFARLANE, Sir H. J. MACKINDER, Prof. J. L. MYRES, Dr. MARION NEWBIGIN, Mr. A. G. OGILVIE, Mr. A. STEVENS, and Prof. J. F. UNSTEAD, *from Section E*; Mr. D. BERRIDGE, Mr. C. E. BROWNE, Sir R. GREGORY, Mr. E. R. THOMAS, Miss O. WRIGHT, *from Section L*) appointed to formulate suggestions for a syllabus for the teaching of Geography both to Matriculation Standard and in Advanced Courses; to report upon the present position of the geographical training of teachers, and to make recommendations thereon; and to report, as occasion arises, to Council through the Organising Committee of Section E, upon the practical working of Regulations issued by the Board of Education and by the Scottish Education Department affecting the position of Geography in Training Colleges and Secondary Schools.

The Report consists primarily of a statement prepared by the Scottish members of the Committee on the position of geography in the schools of Scotland. The syllabuses and regulations governing the subject in England remain substantially as given in a previous report. The Oxford and Cambridge Schools Examination Board has the whole question of the status of geography under consideration.

In the Oxford Local Examinations the syllabus in the School Certificate Examination of 1927 differs from the corresponding syllabus for 1926 only in one important particular, viz.: in the omission of common Map Projections. The Delegates have come to the conclusion as a result of the 1926 and previous Examinations that it is inadvisable to expect a knowledge of Map Projections from candidates of about sixteen years of age. The Delegates have also endeavoured, while giving adequate choice to all candidates, to limit the range of the Regional Geography. Examiners and many teachers had come to the conclusion that some limitation was essential. For 1927 and future years the Higher School Certificate syllabus has been enlarged by the inclusion of a practical paper in response to a recommendation received from a number of geographers and teachers of geography in schools.

In the Cambridge Local Examinations no change is contemplated in Geography Syllabus in the School Certificate Examinations: in the Higher School Certificate Examination the Syndicate have decided, on the recommendation of the investigators, to introduce a new paper on the geography of France and Germany, which may be taken as a subsidiary subject. A list of books will be printed for the guidance of teacher and candidates, and some of these books will be in French and German. It is thought that such a subsidiary subject will appeal to Group II candidates taking French or German, and possibly also to candidates taking History.

The University of Durham School Examinations Board have made certain modifications in the Syllabus, the subject counting in strict parity with the other subjects of Group C (Science and Mathematics).

The Report of the Scottish members is attached.

Report by Dr. M. I. Newbigin and Messrs. J. McFarlane, A. G. Ogilvie, and A. Stevens, Scottish Members of the Committee.

Leeds, September 1927.

The Scottish members of the Committee wish to report as follows upon the present position of Geography as a subject of higher study in Scottish schools.

Geography was recognised in the school curriculum by the Scottish Education Department as a subject for the Higher Grade Leaving Certificate in 1914. New regulations were promulgated in 1924, retaining the subject, but resulting in the discouragement of its study.

In Scottish schools the leaving certificate (Group Leaving Certificate) is awarded by the Scottish Education Department as evidence of the satisfactory completion of a 'Secondary Course.' The award is made on the basis of teachers' estimates checked by official examinations, and the certificate covers a number of subjects. The