

CRITICAL NOTICE.

By G. UDNY YULE.

The Essentials of Mental Measurement. By WILLIAM BROWN and GODFREY H. THOMSON. Cambridge, at the University Press. 1921. pp. 216. Price 21s. net.

This is a revised and expanded edition of Dr Brown's well-known little work, bearing the same title, which was published in 1911 and has now been for some time out of print. The format has been increased from demy to royal 8vo; the number of pages from 154 to 216, and the price (alas!) from 3s. 6d. to 21s. net. Dr Brown, owing to his time being taken up with Army medical work during the war, was unable to take any further part in the development of applications of statistics to psychology after 1914 and the revision of the work, we gather, is almost wholly due to Professor Thomson.

The additions and revisions are very considerable, and for the benefit of those who know the old edition it may be convenient to describe them. In *Part I. Psychophysics* two new statistical chapters have been interpolated. Chapter I on Mental Measurement is now followed by a chapter on "The Elementary Theory of Probability" which covers a very wide field—frequency distributions, averages and measures of dispersion, the standard error of the mean, the binomial distribution, and the normal curve. The former Chapter II on the psychophysical methods succeeds this, and then comes the second new chapter on "Skewness and Heterogeneity in Psychophysical Data," in which Professor Pearson's test of goodness of fit is described, his system of skew curves, and the analysis of a compound distribution into two normal curves. The former *Part II. Correlation* has been even more largely expanded. The Introduction and the chapter that followed on the Mathematical Theory of Correlation remain the same in title though the matter has been revised. Chapter VII, on the Influence of Selection, covers not only the effect of selection on standard deviations and correlations, but also multiple correlation, 'spurious' correlation ('spurious' is an unfortunate term

which is better avoided) and a short account of the variate-difference method. In Chapter VIII the authors consider the correction of raw correlation coefficients for irrelevant factors and for observational errors, and in the two final chapters the theory of general ability and the hierarchy. There are some useful tables at the end of the book and excellent bibliographies.

The exposition, it may be said at once, is admirably lucid, as one would expect from both authors, though it is at times extremely compressed; the student will often have to refer to other volumes or to the original memoirs if he wishes adequately to study some of the methods of which hardly more than an outline is given. Certainly the volume is one which should be in the hands of every investigator who is using the methods considered. And yet—if I were asked whether the volume was entirely a satisfactory one I should feel bound to answer, No. My objection is not to the manner of exposition, but to the matter chosen. This may seem an unfair objection: an objection merely to the authors (for I take it that both authors must be held responsible for the matter included) having given us one sort of book when I would rather that they had given us another. But I do not think that my particular objection is unfair. A book, a good book, should be written with one aim, one outlook. The present book is heterogeneous: it is in part a text-book and in part a polemical pamphlet. If the reader skips the prefaces, to avoid the initial note of controversy, he will find himself reading an ordinary peaceful treatise until he gets to Chapter VI or perhaps Chapter VIII. In Chapter VI the trumpets begin to sound, though gently and musically, about the method of ranks. But with Chapter VIII, and the correction of correlation coefficients for errors of observation, the *mêlée* is fairly started and Chapters IX and X continue the fight on the field of the hierarchy. It should not have been impossible to write an exposition in a less controversial spirit. It would have been perfectly possible, for example, to exhibit the methods of correction of a correlation coefficient for uncorrelated errors without criticism of Professor Spearman's use of the formulae, to emphasise the assumptions made, to show what happened when errors were not uncorrelated, and to illustrate the inapplicability of the assumptions (or of certain of them) in the few cases that have been tested. It would have been perfectly possible, in the same way, to have shown how different systems of correlations arise from specific assumptions as to the nature of the causation—variables which are linear functions of general, specific and group factors; the dice-throwing cases on which Dr Thomson has made

himself a specialist; and so on—and to have shown the student how difficult an aggregate of correlations might be to interpret, without entering into controversy at all. References at the end of the discussion to show the reader how questions of interpretation had arisen would have been quite sufficient. In spite of the controversy it is pleasant to note the statement by Dr Brown in his preface that as an opponent he has learnt to admire the work of Professor Spearman to a very great degree, and regards his work in correlational psychology as epoch-making in its significance.

As briefly as possible, a few comments may be made on the points of controversy. First as regards the method of ranks, that is of ranging individuals in order as regards any quality, numbering them off, and taking the number assigned to each individual as the measure of his quality for purposes of correlation. The alternative is to assume that if the two qualities correlated could be measured the correlation between them would be normal, and to use an approximate formula based on this assumption. Dr Thomson summarises his views on this point most clearly in his preface. He makes the, possibly very valuable, suggestion that it may be practicable to approximate to the true distribution by estimating not only ranks, but the relative order of differences between the ranked individuals, of the differences between these and so on. Failing this, "the question is whether it is more probable that the assumed or the measured distribution corresponds to the actual unknown mental distribution" (p. vii). On this I would make two comments. In the first place, as it seems to me, the mental distribution is not merely unknown but totally indeterminate and it is impossible to speak of an 'actual' distribution, or of two intervals on the scale being equal or unequal, until the scale has been agreed on. Secondly, it is always possible to choose such a scale that the distribution of any variable taken singly shall be normal—though of course a scale so fixed is arbitrary and may be unnatural: but the assumption made in using Professor Pearson's formula for obtaining the 'true' correlation from the correlation for ranks is one that it is *not* necessarily possible to secure by any choice of scale. It is assumed not only that the distribution for each variable taken singly is normal, but also that *the correlation is normal*. A choice of scales such that the distribution for both variables is normal will only make the correlation normal, if it was originally of the type that I have elsewhere termed 'strained normal'.¹ Given

¹ Cf. my paper on Association in the *Journal of the Royal Statistical Society*, May 1912, Sections 39 et seq.

sufficient observations the existence of this type of distribution can perfectly well be tested in either of two ways. (1) A contingency table can be drawn up for the ranks, using at least three rows and three columns, and by the formulae of pp. 124-5 it can be seen whether this can be regarded as a normal distribution with the assigned correlation. (2) Each rank can be replaced by the equivalent normal deviation and the correlation then worked out between these deviations. As the validity of the assumption can be tested, there is no justification for using it without the test. It may be added that an objection may be brought against the use of any approximation which assumes normality, precisely of the same kind as is brought against one of Professor Spearman's methods by Dr Thomson—namely that, even if the assumption may be nearly true for a large aggregate of observations, it cannot be true for small samples. The method (2) suggested above is always applicable and may, perhaps, be of service—but it is not so brief as the pure ranks method. Given, however, the same number of observations, the standard deviations can be determined once for all.

On the question of correction of correlations for errors of observation, I would like to point out that the formulae for correction only assume that errors and quantities measured are uncorrelated—that the correlation coefficient is zero—not that they are strictly independent, and the fact that the *correlation-ratio* in any given case is large is of no importance at all: the correlation-ratio on p. 160 ought not to be cited in this connection. In that case the correlation *coefficient* between errors and variables is practically zero: that is all that is required, and surely, in spite of the authors' scepticism, what ought to be expected? No experimenting would be of any value at all if there were a liability to biased errors, which could not be eliminated by averaging, at different points of the scale. The correlation between simultaneous errors in the observed magnitudes of two variables is another matter; but here again to allege that such correlation cannot be eliminated by proper methods of experimenting seems to me to be taking a very hopeless attitude. Moreover I cannot help thinking that some blunder has exaggerated the evidence as to the existence of the latter type of correlation. I must decline to believe that with 43 observations the probable error of a product-sum can be of the order of one-half of one per cent. of the quantity dealt with, as alleged by the figures on p. 159: and I note that the formula given, which I do not recognise, is of the wrong order in n (the number of observations). On the general principle I agree, however, with the authors that no formulae assuming lack of correlation can be trusted

for small samples and that “it is an essential of good work to use such samples that corrections to the raw values obtained are unimportant” (p. 161).

Now as regards the ‘hierarchy’ controversy, let me first take the points with which I find myself in agreement with the authors. I agree that the Hart-Spearman criterion is unsatisfactory, and I agree that the use of the ‘correction’ formulae has been dangerous and possibly misleading. But both these conclusions leave me rather cold, for in the very little work that I have personally done on the hierarchy question I used Burt’s uncorrected coefficients and did not employ the criterion, and the results left no doubt at all in my own mind that the assumption of a general and of specific factors gave an extraordinarily good fit over the part of the table that I tested. Where Dr Thomson’s work seems to me most important is in his proof that quite other assumptions may lead to true hierarchies, and yet other assumptions again to distributions which may in some degree ape the form of the hierarchy and, even with a sufficient number of observations, be mistaken for that form under careless or incorrect methods of examination. But I part company altogether with him in his view that hierarchical order “is the natural relationship among these coefficients, on any theory whatever of the cause of the correlations, excepting only theories specially designed to prevent its occurrence. It is the *absence* of hierarchical order which would be a remarkable phenomenon requiring special explanation; its presence requires none beyond what is termed chance” (pp. 183–4). Let us be clear, in the first place, as to what is meant by hierarchical arrangement: its accepted meaning—I think I am right in saying—is that

$$\frac{r_{km}}{r_{kn}} = \frac{r_{sm}}{r_{sn}}$$

for all values of k , m , n , and s . Dr Thomson seems at times to degrade the meaning of the phrase to imply no more than positive correlation between r_{km} and r_{sm} in any pair of columns. I use the phrase in the first sense, and cannot regard any table as hierarchical unless the rule holds—strictly for theoretical coefficients, or within the limits of probable error for observed coefficients. Now it is not necessarily true that “coefficients of correlation are themselves correlated” if by this is meant that r_{km} and r_{sm} are positively correlated for any pair of columns k and s . Fluctuations of sampling in the r ’s are so correlated and this may give rise to some trouble in interpreting observed tables, but not necessarily to a hierarchy in the strict sense. Nor—so far as I can see—

can 'chance' give rise to a hierarchy. Dr Thomson's mechanism of pp. 175-7 in the first place does not give rise to a hierarchy at all, but only to positive correlation between columns, and in the second place surely cannot be termed a 'chance' mechanism—all the important conditions were predetermined by Dr Thomson himself¹. The result can no more be termed a 'chance' result than a cubical parabola can be termed a chance law if the values of the coefficients are determined by drawing cards from a pack. As a matter of experience, I may say, no one of the few cases that I have examined from non-psychological fields yields a hierarchical arrangement.

It seems to me impossible to agree then that the hierarchical arrangement found by some observers is a 'chance' result or does not call for explanation. The difficulty is that there may be more than one explanation. From the statistical standpoint Dr Spearman's explanation seems to me to be by far the simplest, but the judgment as to its validity must be based on other grounds.

It is not, or certainly not wholly, the fault of the authors, but the reading of this volume has produced in my own mind a feeling of extreme depression. "The Essentials of Mental Measurement": a fine, rhythmic, magniloquent title! A heartening picture it calls up too: *Essentials*... the real root of the matter here! *Mental*... we are really going to get at the *mind* now, look you! *Measurement*... absence of personal bias... precision... splendid! But what do we find? *Essentials*—a few experimental methods, rather difficult to interpret: a mass of statistical methods, which ought to be termed ancillary, not essential: and a great deal of controversy. *Mental*? well it is no good attempting to pretend that many of the data (or *any* of the data?) can be termed mental. Even the authors, with their title before them, can only hope that their arguments have sufficed to "show that purely psychical measurement is a *conceivable possibility*" (p. 9, my italics) and that their data are "indirect psychical measurements" (p. 10). And *measurement*! O dear! isn't it almost an insult to the word to term some of these numerical data *measurements*? They are of the nature of estimates, most of them, and outrageously bad estimates often at that. I have begun to visualise the title of this book as "The Essentials(?) of Mental(?) Measurement(?)."

Statistical methods, I say, should be regarded as ancillary, not essential. They are only essential where the subject of investigation is itself an aggregate, as a swarm of atoms, or a crowd. But here the subject

¹ Compare Dr Garnett's article in this *Journal*, x. 242.

is the individual, not the aggregate of individuals as such. This being the case, statistical methods are only necessary in so far as experiment fails to attain its ideal, the ideal of only permitting one causal circumstance to vary at a time. And it should always be the aim of the experimenter not to revel in statistical methods (when he does revel and not swear) but steadily to diminish, by continual improvement of his experimental methods, the necessity for their use and the influence they have on his conclusions. Statistical methods are not only ancillary; they are, to the experimenter, a warning of failure.

During the last year or two I have made it a practice to begin a short course of lectures to agricultural students, on some elements of statistical methods, by a lecture devoted entirely to showing how exceedingly bad agricultural experiments are, how this introduces the necessity for statistical methods in discussing the results, but does not obviate the primary necessity for improving the experimental methods. Is the latter warning unnecessary in the case of experimental psychology? I can only speak as a statistician with an exceedingly limited and sporadic knowledge of what is being done, but it seems to me that it is urgently necessary. The experiment (if it can be called an experiment at all) is often not only bad, but its badness seems to be accepted merely as a rather interesting fact, calling perhaps for the exercise of more elaborate statistical methods—a few partial coefficients of correlation—a few rather elaborate probable errors—and that is all. It does not seem to be recognised as primarily a condemnation of the experimenter, perhaps because the experimenter has never been exercised in the sciences where really accurate experiment is possible, so that he has no high ideal before him. Let me cite again from the volume before me one instance that I have mentioned already. On p. 160 it is seriously suggested that we cannot safely assume the error of measurement to be uncorrelated with the quantity measured. It does not seem to have occurred to the authors that if the suggestion were true (it was made perhaps under the domination of the Spearman-complex) it would be an utter condemnation of the science on which they were writing. Unless it were possible to assume a great deal more, namely that the average error was zero for all values of the quantity measured, it would hardly be legitimate to speak of the data as measurements at all: their value for any useful purpose would be nil.

The New Psychology that *was* the new psychology some five-and-twenty years ago has had to surrender the title to a yet newer successor. What has happened in the interval? Measurement does not necessarily

mean progress. Failing the possibility of your measuring that which you desire, the lust for measurement may, for example, merely result in your measuring something else—and perhaps forgetting the difference—or in your ignoring some things merely because they cannot be measured. *Has* measurement in this field led to progress? The impression left by “The Essentials of Mental Measurement” is not cheering on the point.