

## THE N. R. C. HANDBOOK ON STATISTICS

*Handbook of Mathematical Statistics.* By H. L. Rietz, editor in chief, and eight other members of the Committee on the Mathematical Analysis of Statistics of the Division of Physical Sciences of the National Research Council. Boston, Houghton Mifflin Company, 1924. viii + 221 pp.

The general purposes of this book may be inferred from statements in its preface and from its form. It deals "with the mathematical analysis of data", not with its collection or interpretation. "Special emphasis is laid on the limitations surrounding the proper applications of the various methods." It is not a treatise or a text-book. It is to be used as a reference book, probably also as a background for a course of lectures. At the same time, it provides sufficient explanation of the fundamental ideas, and especially sufficient illustration of the methods, and sufficient bibliography, to enable the reader to begin his acquaintance with statistical theory here, provided he has already a knowledge of collegiate mathematics. This last proviso is, in the mind of the reviewer, a distinct advantage that this book has over some others; it is possible to make the subject so much clearer with the use of mathematical language than without it. Apparently, also, having been written by a committee of nine authors, to each of whom a piece was allotted, the book is intended to be representative of expert knowledge in the several portions of its field.

This was an obvious and decided advantage of multiple authorship, and one cannot but recognize the appropriateness of choice of author in each case. There were also some obvious disadvantages, especially as the several authors were scattered about the country from the Atlantic seaboard to the Pacific. If the book is to serve well as a basis for lectures, it is important that there should be consistency both of notation and of language throughout. This is the more desirable because at best the notation of mathematical statistics is complicated. It was a difficult thing for a group of nine to achieve, and in fact it was not very well achieved. One is led to feel that, from a practical point of view, it would have been better to have had a smaller number, or at least a different division of labor among the nine. The chapters on multiple and simple correlation ought not to be widely different from each other in style, method of approach, notation and language. There would seem to be no good reason why the subject of frequency curves should be introduced by one writer and carried on by another, or why Bernoulli, Poisson, and Lexis distributions should be handled by a different author from the one who deals with probability. As a result

there are such inconsistencies as the following. In Chapter I, page 15, moments are defined as in mechanics, not as usually in statistics. This definition is in agreement, though the notation is slightly in disagreement, with the usage by the same author in Chapter IV, page 69. In Chapter VII, however, the statistical definition is used, but the notation of Chapter I. That is, there are two different definitions with the same notation, and two different notations with the same definition. There is also the so-called "Bernoulli Theorem" in two different notations, pages 16 and 72, so that the misprint in the first statement is not immediately obvious by comparison with the second. Further, there is a confusion of statements so that the reader cannot be sure that the expressions really include areas on both sides of the mode until he has examined the illustrative example on page 73.

Another considerable difficulty with which the committee had to deal was that of securing uniform adherence to their excellent ideal of stating and emphasizing the limitations of methods. It is more important that this be done meticulously in a handbook than either in an elementary text-book or in a treatise, for, were the proofs of the theorems given, the careful reader could discover these limitations for himself. Since the proofs are not usually given the reader has a right to demand extraordinary care in the statements. It is a joy to find the many places where special care is manifest, e. g., in the statements regarding the  $\chi$  test on page 80, and that generally, throughout the book, there is the spirit of mathematical rigor. There are some slips. In presenting the formulas for probable errors, for example, no author states that the parameters appearing on the right are, strictly, the parameters of the universe from which the sample is drawn, in some cases (p. 32; p. 77, probable errors of  $r$ ,  $\eta$ ,  $r \sigma_y / \sigma_x$ , and  $Y$ ) the condition that the universe from which the sample is drawn be normal is omitted, and there is little disposition to recognize that, especially for the higher moments, the expressions are derivable only on the assumption that the sample is very large. Indeed, neglect of this very important condition permits the author of Chapter VII to recommend a substitute for the  $\chi$  test of goodness of fit which cannot be accepted unless this difficulty has been rigorously dealt with. Not only do the probable errors of his high  $\alpha_s$ 's mount up, as he points out, with increasing index, but,  $n$  being fixed, his formulas for them become increasingly uncertain. This is due to the fact that, strictly, there should be more terms in his formula for  $\sigma_{\mu_s}$ , which involve higher powers of  $s$ . If  $n$  is fairly large, these terms are negligible when  $s$  is small, because they involve higher powers of  $1/n$  than the terms retained, but when  $s$  is large they are very troublesome. There is the more need of insisting on the use of large samples because one of the common errors of practical workers in statistics is the error of using

in a more or less wooden way book formulas with inadequate data. A more grievous instance of the Handbook's failure to warn its readers against the use of small samples occurs in the chapter on multiple correlation, cited below.

There are a considerable number of places in the book where it seems to the reviewer that the statements should be modified or supplemented, and some of them will be noted as the several chapters are considered in turn. It is hoped that the reader will recognize, however, that the adverse criticisms are of details, and that, for the most part, they are offered with constructive purpose, rather than as indicative of general features. In fact, at first reading these details were seldom noticed. The general effect of the book was stimulating, and it was welcomed enthusiastically. Adverse criticism may probably be summarized in the statement that the book seems to have been written a bit too hastily. This book is such a useful contribution, and will occupy for some time to come such an important position in statistical literature, that it was well worth that extra effort which would have made that position permanent. It is much more than an elementary text, yet it just falls short of being a digest of a treatise, which one feels it might have been. The material is carefully selected and apportioned. The chapters differ markedly, but almost everywhere there is clearness, thoroughness, and simplicity of treatment, combined with mathematical exactness of statement which it is not usual to find in books on statistics. Numerical examples are used to illustrate almost every important idea.

The first chapter on "Mathematical Memoranda" comprises a collection of many mathematical formulas useful in statistics, and a very brief account of probability. The latter is rather too brief; at least it should include the point binomial theorem, and the probability theorem leading to the hypergeometric in finite form (Elderton, W. P., *Frequency Curves and Correlation*, p. 37). There are one or two mistakes noted later under "Errata." It should also be stated that the  $C$ 's, at the bottom of page 16, are mutually exclusive. There should be a reference, at least in the bibliography, for the important table of Burgess, mentioned on page 14, viz., TRANSACTIONS OF THE ROYAL SOCIETY OF EDINBURGH, vol. 39 (1897), pp. 257-321. A valuable reference on quadrature formulas is Irwin, J. O., *Tracts for Computers*, No. 10 (1923). Irwin gives the constants of the Euler-Maclaurin theorem in two useful forms. On page 12, line 4, the words "absolute error" are used in a different sense from that in which they were defined on page 2.

Chapter II is an almost perfect introduction to the subject of frequency distributions. There is an obvious slip on page 32. The author should insert after several formulas, as he does on page 77, the words, "normal distribution". This restriction has an importance here, for on the next page he says "preference should be given to the average which has

the smallest probable error." Glancing back at page 32, the reader would infer that the arithmetic mean is therefore always to be preferred, but this is not a desirable inference; for on page 155 a case is cited where the distribution is not normal and the median appears to have the smaller probable error.

The third chapter deals with interpolation, summation, and graduation, and is clearly the work of one who has had much experience in this field. Considering the small space this chapter occupies, it is remarkably readable and comprehensive. The notation  ${}_x C_k = x_k$  is confusing, since by  $u_k$  the author means the value of  $u$  at  $x = k$ . All his formulas are either successive averages or else based on interpolation by means of finite differences. As there are some functions to which the latter method does not well apply, a word of caution about the desirability of rapid convergence is needed. There is no mention of methods of interpolation on a surface (cf. Pearson, K., *Tracts for Computers*, No. 3, 1920). The author's formula for subdivision of an interval (p. 45) can be improved in form, provided the table on which it is to be used does not already contain its successive differences, and provided the computer is working with a machine which will subtract as well as it will add (Pearson, K., *Tracts for Computers*, No. 2, 1920). The author should give the dates in the references in his foot-notes. For example, at the bottom of page 45, it is misleading to cite the reference "Explanation of a New Formula for Interpolation" without the date, which is 1880!

Chapter IV on curve fitting is in Huntington's best style, simple, clear, and adequate. There follow two short chapters by Rietz, a thoroughly good account both of random sampling and of the distributions of Bernoulli, Poisson, and Lexis. The theorem attributed to Bernoulli on page 72 is not truly his. If it were, the normal law, which is implicit in it, and which the Handbook (following a bad custom) calls Gauss's, on page 11, would be Bernoulli's too. The *Ars Conjectandi* antedated Gauss by more than a century. This theorem is proved by Laplace, but subject to limitations which may make the second term worse than useless if  $c$  is not small. E. g., for the point binomial  $(1/3 + 2/3)^{450}$ , the approximation is better without the second term at  $c = 3\sigma$ . Somewhere in this book, probably in the chapter on sampling, there should be an explanation of the (different) formulas for probable errors commonly accepted by engineers, physicists, and astronomers (e. g., Chauvenet, *Practical Astronomy*, vol. 2, pp. 494, 506). The ordinary "theory of errors" is a special case of the theory of statistics, and is too frequently divorced from it. The desirability of severer warnings against the use of probable error formulas for small samples has already been mentioned. It is also desirable that there should be some warning against the usual *interpretation* of such formulas.

What is meant by the usual interpretation is indicated in the next chapter, on "frequency curves." At the top of page 100 there is a table giving the probability that a variable will differ in absolute value from its mean value by less than  $k$  times its probable error, provided the variable be normally distributed. Now, when the value of the probable error of a frequency constant has been determined, it is not surely correct to estimate by means of this table the probability that this constant will differ in absolute value from its mean value by less than  $k$  times its probable error. It is necessary to know first that this constant is normally distributed, and for this to be true it is not always sufficient to know that the distribution to which the constant belongs is normal. One may, of course, use the table to get an approximate estimate if one knows that the distribution of the constant is approximately normal, but this is frequently not true for small samples, as has been shown in the case of the coefficient of correlation by R. A. Fisher, (*BIOMETRIKA*, vol. 10 (1915), pp. 507-521). The author uses this table in connection with his high moment test of goodness of fit, to which an objection was made in an earlier paragraph. Waiving the question, there raised, as to whether, in a concrete case, he is finding the values of his probable errors correctly, it is now fair to raise the additional question: can he show that the distribution of his high  $\alpha_s$ 's is approximately normal? It is the opinion of the reviewer that the author's idea here, though interesting and original, is not as simple a matter as it appears, and that there was not space available in a handbook for the arguments which would be necessary to justify it. The author's account of Pearson's curves is an admirable simplification of the usual methods. His argument with regard to Sheppard's corrections (p. 94) is not very clear. Certainly the function  $f(x)$  used in his analysis on the preceding page is a graduating function, and so the conclusion that Sheppard's corrections do not demand the vanishing of the "chosen" graduating function does not seem to follow necessarily from the analysis. The author admittedly has a point here, but it is a delicate one, and requires more careful handling. There should also be a warning against the use of these corrections in "abrupt" cases, where they sometimes would make the approximation worse.

Some mention has already been made of the two chapters following on simple and on multiple correlation. The first is a brilliantly clear chapter, perhaps the most striking in the book. The second suffers from an attempt to compress too much into its short space. If the space was necessarily restricted, it would seem to have been better to have attempted first to show what a solid of frequency is (which is not described at all), perhaps by the analogy of distribution of mass in a physical solid, to show what are meant, geometrically, by total, partial, and multiple regression, and by the other constants to be used, and

then merely to recite the formulas with numerical illustration. There appears, however, no good reason for such a strict compression of a subject which is well known to be difficult to present. The author does not always define his terms before using them. The term, "total correlation coefficient" is used on page 140, and first defined on page 141; "partial alienation coefficients" are defined in terms of "partial correlation coefficients" on page 141, but these in turn are first defined on page 142. The reader must be careful to note that the word "cell" is used in this chapter in a different sense from that in which it is used in the preceding chapter, and that the words "assigned values" and "measures" are used in the sense of observed values or measurements. The example chosen to illustrate the methods of computation contains only 136 data. This would not be objectionable if it were carefully pointed out that it is for purposes of numerical illustration only, but the serious discussion of the meaning of the results gives the impression that it is to be taken as an adequate and approved piece of statistical investigation in the field of education. This is extremely unfortunate, even though accidental, since it was hoped that the Handbook would be of special value in improving the quality of current statistical practice. Roughly speaking, it is as futile to attempt to find the parameters of a frequency solid of three dimensions from 136 data as it would be to seek the parameters of a frequency curve from  $\sqrt[3]{136} = 5.14$  data. The situation is made worse in this case by the further fact that one of the variables does not represent a measurable character; there is only a division into three categories.

The last three chapters of the book are written by economic statisticians, on correlation in time series, periodogram analysis, and index numbers. They clearly set forth and illustrate methods that have been used by them, and will be a welcome addition to the earlier mathematical theory, especially in a country where the theory of statistics has been so largely identified with business.

*Errata* (partly supplied by the Editor). Page 16, line 12: *For* at least read some. Page 16, line 14: *For* at least one 2-spot with two dice read a 1-spot or a 2-spot with one die. Page 16, line 7 from bottom: *For*  $\sqrt{2\pi npq}$  read  $\sqrt{2npq}$ . Page 54, line 5: *For* coefficients read coefficient. Page 56, line 5 from bottom: The minus sign (—) in  $\frac{1}{2m}(u_0 - u_\infty)$  should be changed to +. Page 70, line 7 from bottom: *For*  $m^2$  read  $m_1^2$ . Page 73, line 15: *For* 688 read 668. Page 74, line 6 of Ex. 2: *After* number insert of women's numbers; *to avoid ambiguity from the use of number in two senses*. Page 78, line 4 of second paragraph: *For* fit read fitted. Page 92: The integrals are definite and should have upper limits. Page 97, line 15: *For*  $\frac{\sigma}{N}y$  read  $\frac{\sigma}{N}y$ . Page 103, line 7:

For 80 read 81. Page 105, line 3 from bottom: Omit  $x$  and the two commas following. Page 107, line 4: For V read VI and for VI read VII. Page 108, line 1: For VII read VIII. Page 121, line 4 from bottom: For  $\frac{y'_i}{\sigma_y}$  read  $\frac{y_i}{\sigma_y}$ ; i. e., delete the prime. Page 126, formula (9): For  $\frac{2}{x}$  read  $\sigma_x^2$ . Page 142: The last line should begin with  $\beta_{12.34}$ . Page 149, line 2: For (13) read (19). Page 187, formula (8): Insert  $\Sigma$  before parenthesis in the denominator.

B. H. CAMP

---

## FOUR BOOKS ON PROBABILITIES

*Éléments de la Théorie des Probabilités*. Third edition. By Emile Borel. Paris, J. Hermann, 1924. vii + 266 pp.

*Probabilités, Erreurs*. By Emile Borel and Robert Deltheil. Paris, Librairie Armand Colin, 1923. vi + 197 pp.

*Wahrscheinlichkeitsrechnung*. By Otto Knopf. Sammlung Götschen. Berlin, Walter de Gruyter and Co., 1923. Two volumes, 112 and 112 pp.

*Grundlagen der Wahrscheinlichkeitsrechnung und der Theorie der Beobachtungsfehler*. By F. M. Urban. Leipzig, B. G. Teubner, 1923. vi + 274 pp.

The present time is a time of decided activity in the publication of books on the theory of probability. While two or three works of considerable originality and merit have appeared, the majority are merely text-book rearrangements of the traditional course in probability adapted to various classes of readers.

The first of the books in this review is the third edition of Borel's *Théorie des Probabilités*. This well known work has been before the public since 1909. The chief change in this third edition is the addition of four notes at the end of the book on applications to radio-activity, on a problem leading to a Stieltjes integral, on games of chance in which the ability of the players is taken into consideration, and on what Borel calls a differential method in statistics.

The book by Borel and Deltheil is number 34 of the series of little books known as the "Collection Armand Colin". The purpose of the series is to present to the educated person readable monographs on special topics in philosophy, history, science, mathematics, literature and other branches of learning. As might be expected, the book covers much the same ground as the more extended work of Borel.