BOOK REVIEWS

ROY G. FRANCIS. *The Rhetoric of Science. A Methodological Discussion of the Two-by-Two Table.* Minneapolis: University of Minnesota Press, 1961. Pp. vi + 183.

At the end of his Introduction, the author asserts ". . . what follows is not a book for statisticians. It is a book directed at the bulk of social scientists. It begins with the assumption that the problem is the thing. It takes a simple statistical tool, the two-by-two table, and shows its relation to logic and to the intellectual history of the problems of cause."

Indeed, statisticians (at least those who are not yet full-blown Bayesians) will find much of what is presented familiar and comfortable. The notion of science as a language, i.e., as composed of "statements about the world"; the distinctions among the context of discovery (inquiry), of justification (research), and of technological application (analysis); the logical priority of theory to data, at least within the research context; the impossibility of error-free measurement and the inherent probabilistic character of propositions with this worldly referents—all this and much else within this little book will hardly seem new or questionable to statisticians—nor for that matter to methodologically sophisticated social scientists.

Nor will those readers who possess some knowledge of categorical logic and of methods for the analysis of qualitative data be surprised to discover that propositions may be related to one another in any of several modes or that alternative tests of association and estimation procedures are available for the analysis of two-by-two data tables.

What may be intriguing to some, however, is the manner in which the author attempts to tie together the metatheoretical position (essentially latter-day Logical Empiricism) with the treatment of propositional relations and the alternative techniques for statistical analysis. The various kinds of relations between propositions are mapped into the two-by-two format and are viewed as alternative models for research design and data analysis. A given cross-tabulation of data may be viewed as simply associational if neither of the dichotomous bases is *implied by theory* to be *logically antecedent* to the other. Or it may be seen to reflect one or another form of causal relation between the propositions depending on the epistemic status of the dichotomies. The choice of a measure of relationship, and of the interpretation made of its value in a given instance (granting the prior rejection of chance association) is seen thus to depend largely upon the structure of the argument from theory involving the two propositions.

Certain of the conventional measures (e.g., Yule's Q, Lambda, and Jahn's Index of Segregation, I_s) are criticized as essentially technical or ad hoc in orientation. These and other indices are further distinguished in terms of differences in the definitions of chance they employ (and hence the sense of causality assumed) and in terms of their relative computational simplicity.

There are two matters which, for this reviewer at least, detracted from the otherwise desirable features of this monograph. The first is the flippant style in which certain arguments are presented (for instance, the last two sentences in the second paragraph on p. 76). But even more distressing than the occasional lapses into cuteness are the numerous typographical mistakes and omissions that occur in the text and formulas. While making no systematic effort to check proof, I did discover several such mistakes. In the middle paragraph on p. 73 alone there are at least three. On p. 91 the reader will find that the "parenthetic term (sic!) in the denominator" of (5.6) referred to in the text lacks the needed parentheses and that the label, (5.7), has been omitted from the following equation. A very important 0 has been omitted from the A by C cell of the table for Case VI on p. 93,

the two row marginals are missing from the denominator of (5.8) on p. 98, and the exponent in the final line of (5.12) on p. 100 is omitted. Also missing is the factor N in the expression for maximum chi square under Case V on the following page and to top it off this same factor unaccountably seems to have become an exponent in the expression near the bottom of the same page.

At this point I stopped referring even casually to the technical formulations and read the remainder of the book for its general arguments. I urge others who pick up this book to do likewise or else be prepared to work through all of the symbolic material from the start.

University of Michigan WARREN T. NORMAN

R. L. THORNDIKE AND ELIZABETH HAGEN. *Measurement and Evaluation in Psychology and Education* (Second Edition). New York: John Wiley and Sons, 1961. Pp. viii + 602.

This book remains essentially unchanged from the first edition save for inclusion of a few recent references, omission of some material, and expansion or rearrangement of other topics. One reason why more revision was not undertaken is no doubt due in part to the care and skill with which the first edition was written. Nevertheless, there are some differences between the two editions and this review will concentrate on these.

In the first portion of the book, which deals with introductory material, teacher-made tests, elementary statistical concepts, and characteristics of a measurement procedure, the major revision is found in the section on validity. Three types of validity are suggested: validity as representing, validity as predicting, and validity as signifying or describing. Validity as representing is used synonymously with content validity.

Validity as predicting "... is primarily an empirical and statistical evaluation, and this aspect of validity has sometimes been spoken of as *empirical* or statistical validity" (p. 164). In the first edition, the authors refer to congruent validity as one type of empirical validity (p. 110, First Edition). Their description of validity as signifying, however, indicates that they intend to include congruent validity under this label. Indeed, they even use some of the examples found in the first edition under the heading congruent validity. This brings validity as signifying closer to construct validity, as described in the *APA Technical Recommendations*, than to "construct validity" as used in their first edition.

Since the authors clearly intend status validity to be included in the framework of validity as predicting, it is unfortunate that all the examples found in the section introducing this type of validity involve a future criterion of success. This section begins with the statement "Frequently we are interested in using a test to *predict* some specific future outcome" (p. 163), and the summary statement for the chapter reads "For other tests, especially aptitude tests, we may evaluate how well the test *predicts* some measure that serves as a later criterion of success" (p. 203). An incorrect concept of validity as predicting may result in the reader's mind.

Another revision in the first portion of the book is an expanded discussion defining objectives and planning a test. The essential points, however, were made in the first edition. The chapter on norms and units of measurement now appear before the discussion on validity and reliability. In a review of the first edition, appearing in *Psychometrika*, the statement that standard-score scales (a linear transformation of the raw scores) have units which "... have the same meaning throughout its whole range of values" (p. 165, First Edition) was criticized. Unfortunately, this statement can be found on page 137 of the second edition.

In the second portion of the book, eight chapters cover the topics: sources of information about tests, the various kinds of traits for which tests have been designed, and

kinds of measuring instruments in current use. Again, these chapters remain essentially unchanged with only 12 references out of the 113 cited in these chapters having a publication date of 1955 or later. Mention is made, however, of Form L-M of the Stanford-Binet and three pages are devoted to results from Thorndike's *10,000 Careers*. In contrast, material about Buros' *Tests in Print* and Guilford's work on the structure of the intellect was, evidently, too recent to be included.

The last group of chapters in the book deals with planning a school testing program, marking and reporting, measurement in educational and vocational guidance, and tests in the selection of personnel. The chapter on measurement in diagnosis and therapy found in the first edition has been eliminated. Seven pages devoted to recording school test results have been reduced to less than one page in the new edition. The discussion of the mechanics of administering and scoring school tests has been expanded and a new section on large-scale testing programs introduced. The authors take the position that tests external to the educational program of the school influence strongly the curriculum of the school and that it is important that these testing programs emphasize the important goals of education. In the chapter on the selection of personnel, Venn diagrams have now been used to provide an intuitive feeling for partial and multiple correlation.

The appendix, which includes a brief description of widely used tests and a list of test publishers and their addresses, has been up-dated.

A teacher's manual is available which contains over 500 keyed items, many of which test for understanding of principles rather than recall of fact. These items can be of value to any instructor of a measurement and evaluation course, regardless of the textbook he chooses.

While the revisions found in the second edition do not contribute substantially to the value of the book, *Measurement and Evaluation in Pyschology and Education* remains an outstanding introductory textbook in its field.

JASON MILLMAN

School of Education, Cornell University

PATRICK BILLINGSLEY. *Statistical Inference for Markov Processes*. Chicago: University of Chicago Press, 1961. Pp. 75.

"In this monograph a theory is developed for problems of statistical inference on Markov processes." Part I presents procedures for estimation of parameters, testing hypotheses about parameters, and testing whether two samples come from Markov processes which are structurally related in some way, all for time-discrete Markov processes. Part II deals with time-continuous Markov process of the completely discontinuous type. There is a Mathematical Appendix and a Bibliography of 35 titles.

G. J. LIEBERMAN and D. E. OWEN. *Tables of the Hypergeometric Probability Distribution*. Stanford: Stanford University Press, 1961. Pp. 726 + vi.

"The Introduction to the tables presents methods of calculation and detailed applications of the tables to a sequential procedure, tests of equality of two proportions, the distribution of number of exceedances, the binomial distribution (Bayes' prediction), and sampling inspection."

"In the first section of the tabulation, point and cumulative probabilities are given for values of N (lot size) from 2 through 50 in steps of 1. Values from 60 through 100 are given in steps of 10. The second section of the tabulation gives the probabilities for $N = 1000$ and n (sample size) $= 500$. A third table of probabilities is given for N from 100

through 2000 in intervals of 100, with $n = N/2$ and k (number of defectives in lot) $= n - 1$ and n. A tabulation of logarithms of factorials is provided in an appendix."

J. H. B. KEMPERMAN. *The Passage Problem for a Stationary Markov Chain.* Chicago: University of Chicago Press, 1961. Pp. 127.

"The purpose of this monograph is to develop into systematic methods certain procedures which in special cases have been found useful in studying the phenomena of first passage and absorption for a stationary Markov chain . . ." There is a bibliography of 56 titles.