BOOK REVIEWS

QUINN MCNEMAR. *Psychological Statistics, Second Edition.* New York, John Wiley and Sons, Inc., 1955, pp. vi + 408.

The second edition of McNemar's well-known text has the major virtues and faults of the first edition. Though the revisions are considerable, much of the original text remains unchanged, but practically all of the revisions represent improvements.

The primary virtue of this book—not shared, unfortunately, with some others in the same field—is its relatively complete freedom from major errors. The instructor who uses it will not have to spend much time explaining wherein and why the author's discussions are incorrect. Its primary fault is still its unevenness of exposition: careful, detailed, and lucid on many topics, condensed on several others to the point where students are unable to acquire any real understanding. Apparently the author could not quite bring himself to omit entirely a number of topics which he evidently did not feel he had space to develop fully. If the instructor is willing to skip over these topics, he will find McNemar's text excellent, but if he feels he must teach them thoroughly he will find it necessary to supplement the text frequently in his lectures and additional assignments.

The first notable improvement in the revised edition is a rearrangement of the materials of the earlier chapters, with considerable re-writing at various points. The discussion of the t-test now precedes that of correlation, so that all the materials on the sampling theory of one variable follow consecutively; the treatment of the binomial distribution has been expanded and used more fully as a starting point for the ideas of statistical inference. Though seven chapters on the one-variable case now replace six, the total number of pages has increased only from 108 to 114; the reviewer still doubts that this treatment is sufficient for students taking their first course in statistics.

The second major improvement is in the treatment of analysis of variance. A chapter (still too brief) on testing variances for homogeneity has been added, and the chapter on complex analysis of variance has been improved considerably.

A four-page chapter on distribution-free methods presents the sign test, the median test, Mood's test of C correlated sets, and the Mann-Whitney U-test. This chapter seems more like the result of an afterthought than a serious effort to discuss a now important area of modern statistical inference.

The nine-page chapter entitled "Distribution Curves" describes only the normal distribution and standard scores. The binomial distribution, and the normal approximation to it, are considered in the following chapter. The Poisson distribution is not discussed.

A few minor errors and inconsistencies still remain. Thus on page 100 the standard error of the mean of a sample from a finite population is given as $\sigma \sqrt{1 - n/N} / \sqrt{n}$ (where σ is the population standard deviation, N the population number, and n the sample number), instead of as $\sigma \sqrt{N - n} / \sqrt{n(N - 1)}$. Also on page 133 the standard error of estimate is given as $\sigma_{x \cdot y} = \sigma_x \sqrt{1 - r^2}$, omitting the factor $\sqrt{(N - 1)/(N - 2)}$. Similarly on pp. 138–9 the expression $1 - \sigma^2_{y \cdot x}/\sigma^2_y$ is defined as the proportion of the y-variance determined by x, again without the factor $(N - 1)/(N - 2)$, though the correct formula is implied in the case of multiple correlation on page 186. The so-called "shrunken" multiple correlation is in fact the square root of the coefficient of non-determination, and in the two-variable case this becomes $r'^2 = 1 - (1 - r^2)(N - 1)/(N - 2)$. McNemar calls the "shrunken" multiple correlation the unbiased estimate of the multiple correlation in the population; if this is so, r' as defined above is the unbiased estimate of ρ. The reviewer has never seen a proof that this *is* an unbiased estimate, and in fact it is probably not. For if we write it in the form, $r'^2 = 1 - s^2_{y \cdot x}/s^2_y$, $s^2_{y \cdot x}$ *is* an unbiased estimate of $\sigma^2_{y \cdot x}$, and s^2_y *is* an unbiased estimate of σ^2_y; the unbiased estimate of a ratio is seldom the quotient of unbiased estimates of its numerator and denominator, and the unbiased estimate of a square root is seldom the square root of

the unbiased estimate of the number. Aside from these and a few other liberties with such ratios as $N/(N-1)$ and $(N-1)/(N-2)$, however, there are few errors.

The formula for the large-sample standard error of $(r_{12} - r_{13})$ has been replaced in the revised edition by Hotelling's t-test for this difference.

To the present reviewer, the author's deliberate lack of emphasis on efficient computing methods appears to be a deficiency, but this is admittedly a matter of opinion wherein many other authors agree with McNemar.

In summary, those instructors who liked the first edition of McNemar's book will like the second edition better; those who did not like the first edition will probably have the same objections (though somewhat weakened) to the second. This reviewer, it may be added, is one of the former.

University of Tennessee *Edward E. Cureton*

R. L. THORNDIKE and ELIZABETH HAGEN. *Measurement and Evaluation in Psychology and Education.* New York: Wiley, 1955, pp. vii + 575.

This book is an outgrowth of, and was developed for, a Teachers College course for teachers, administrators, guidance workers, and majors in several branches of psychology. "It undertakes to provide the foundations that these workers in different branches of education and psychology will need in order to use and interpret tests, move ahead into more specialized testing courses, and go ahead independently to study their own practical testing problem." (Preface, p. v). It seems admirably suited to these purposes; it should also be an excellent handbook for the school administrator who is not an expert on testing to peruse and to have available for reference.

Many value-judgments are made throughout the book, as is inevitable if it is to be practically serviceable, but I think most measurement people would agree with most of the judgments. Those to which I take most violent exception are noted below. The general effect, however, is a sound evaluation of tests in terms of what they can and cannot contribute to the making of decisions, principally in education.

In the first chapter, on historical and philosophical orientation, the authors depart from the usual nominal, ordinal, interval, and ratio categories of measurement, as proposed by Stevens, in favor of four other classes: either-or, qualitatively described degrees, rank in a group, and amount expressed in uniform established units. The first is identical with the Stevens nominal scale (one of the examples given refers to a man being either single, married, widowed, or divorced); the second and third are ordinal scales; and the fourth describes an interval scale, although the examples given—weight, height, and age—are ratio scales. The propaedeutic function of the book might have been better served by the use of standard terminology.

After a chapter cataloging the different measurement options with respect to what is measured and how it is to be done, there follow two on teachers' tests and on preparing objective tests. These chapters present a detailed plan for constructing a classroom test, built around a sample unit of instruction. Rules for item construction are given, together with examples of good and poor items.

The chapter on elementary statistical concepts reviews the usual topics of central tendency, variability, and relationship; in addition it points out the dependence of the usual interpretations of standard deviation upon the presence of a normal distribution. The discussion of the correlation coefficient and its interpretation is particularly good for the level of sophistication the authors have chosen.

The sixth chapter deals with test desiderata: validity, reliability, and practicality, and

culminates in a Schedule for Evaluating a Test. The discussion of reliability is especially good, but the treatment of validity differs from that of the recent APA Technical Recommendations, without any particular gain, it seems to me. Furthermore, the term "construct validity" is used with a meaning different from that of the APA committee.

The four common types of norms—age, grade, percentile, and standard-score—are next described and evaluated. In the discussion of standard scores the authors seem to go completely off the track, and are actually perpetuating the prevalent fallacy that there is some magic in the process of subtracting the mean and dividing by the standard deviation. To quote (p. 165): "Because the units of a score system based on percentiles are so clearly not equal, we are led to look for some other unit that does have the same meaning throughout the whole range of values. *Standard-score* scales have been developed to serve this purpose." An example is given in which Mary gets a score 1.50 above the mean on test A and Johnny, a member of the same class, gets the same standard score on test B. "Thus, we may say that Mary did as well on test A as Johnny did on test B . . ." (p. 166), To aid in interpreting the "degree of excellence represented by a standard score" the reader is referred to a table of percentile equivalents in a *normal* distribution. This puts the authors in the peculiar position of beginning the section on standard scores with a statement on the inadequacy of percentile scores, and then midway in the discussion using them to interpret the supposedly superior standard scores.

It is only several paragraphs beyond this that they take up the matter of normalized standard scores. In a summary, however, they say that standard scores are ". . . presumably equal units. The basic unit is the number of standard deviation units above or below the mean of the group" (p. 168).

It would seem to be too easy for the naive reader to get a false picture of the virtues of standard scores from this presentation. I should think it entirely within the grasp of the audience to whom the book is directed to understand the rationale for using a unit based on an assumed normal distribution of ability, and that non-normalized standard scores, being only linear transformations on raw scores, are no better than raw scores unless there is some kind of reference distribution or unless it is desired to compare performance across tests which have the same form of distribution.

This same chapter has an excellent discussion of profile interpretation, emphasizing the necessity for taking into account the reliabilities of difference scores.

Chapters 8 through 15 cover the topics which form the heart of the usual tests and measurements course: sources of information about tests, the various kinds of traits for which tests have been designed, and kinds of measuring instruments in current use. The general approach here has been to try to set forth principles governing various measurement techniques so as to give the reader a background for evaluation. Illustrative tests are discussed briefly, and findings of some of the research studies in relevant areas are summarized.

The last five chapters deal with planning a school testing program, marking and reporting, educational and vocational guidance, personnel selection, and diagnosis and therapy. In the chapter on Marking and Reporting, Thorndike and Hagen take the position that course marks can be only a relative appraisal, with respect to some reference group. They ignore the alternative of assigning marks based on the extent to which the students have achieved the operationally defined objectives of instruction. Sufficient progress has been made in this direction, certainly, to make it a functional alternative, and, for me at least, a preferable one.

There are four appendices, the first two of which are computational (square root and correlation coefficient). The third is a listing and evaluative description of some of the more widely used tests, and the fourth is a list of seven prominent test distributing agencies, with a description of the kinds of services they offer.

University of Michigan *John E. Milholland*

Decision Processes, THRALL, R. M., COOMBS, C. H., AND DAVIS, R. L., Editors, John Wiley and Sons, New York, 1954, viii + 332, $5.00.

Decision Processes, while nominally a book, is in fact a one-issue journal consisting of nineteen mathematical and experimental papers on statistical decision theory, game theory, learning theory, and measurement theory (including utility measurement)—all parts of an area well described by the title. The work stems from an eight-week summer conference on "The Design of Experiments in Decision Processes" held at the RAND Corporation in 1952. On the grounds that such a book will not be definitive and that research activity in the area is lively, the editors felt that "an informal and relatively speedy method of printing" was justified. While agreeing with their conclusion, two questions can be raised: Did these considerations actually force the publisher to employ such an unattractive format? And do not these same reasons, plus the desirability of the lowest possible price for a volume soon to be antedated, suggest paper, not cloth, covers?

The volume begins with an introduction by R. L. Davis, which outlines the area, cites a bit of its history, and sketches the major focus and results of each paper. A clear notion of the relevance of this book to one's interests can be obtained by reading these eighteen pages. The next article, also introductory in nature, "Some Views on Mathematical Models and Measurement Theory" by C. H. Coombs, Howard Raiffa, and R. M. Thrall is divided into two parts. The first offers a highly idealized scheme of scientific research with particular emphasis on the role of mathematical models. The second part, on measurement models, is presented as an exemplification of the general scheme; it should serve as a handy reference of possible scales which, by being more complete, supplements Stevens' widely known classification. Definitions and social science illustrations are given of transitive relation, partial order, weak order, lattice, vector space, etc.; the interrelations among them are discussed and neatly summarized in a diagram.

The remaining articles are grouped in four sections: individual and social choice, learning theory, theory and applications of utility, and experimental studies. Since it is impossible to discuss them all in detail, attention will be restricted to those the reviewer found particularly satisfying or stimulating; as it happens all four divisions of the book are represented.

L. A. Goodman's paper "On Methods of Amalgamation," John Milnor's "Games Against Nature," and the "Note on Some Proposed Decision Criteria" by Roy Radner and Jacob Marschak are all concerned with decision criteria for the selection of a strategy in a game against nature. Goodman offers a new criterion which, simultaneously, generalizes those of LaPlace, Bayes, and Copeland. Radner and Marschak present an example which suggests that both the Hurwicz generalization of the Wald minimax criterion and the Savage minimax regret criterion may be inadequate, and, as we shall see, Milnor's work raises similar doubts. The Hurwicz criterion leads to a decision distinctly at variance with common sense, and the Savage criterion depends on irrelevant alternatives, in a sense analogous to Arrow's usage. Milnor's paper, the most interesting and elegant of the three, overlaps the others, covering the LaPlace, Wald, Hurwicz, and Savage criteria. Milnor lists eleven axioms a criterion might meet, and he shows which are met by the four criteria mentioned, and which characterize each of the four. It is striking that all but the LaPlace criterion fail to meet a Pareto condition on strategies (domination), and that the LaPlace criterion fails on another axiom, which, while not so basic, seems desirable. Furthermore, no criterion can meet all eleven axioms, so one is led to consider classes of criteria defined by subsets of axioms which seem intuitively necessary. Milnor selects five as essential and three others as desirable; he shows that the class so defined is non-empty. Finding a simple characterization of this class of criteria, or indeed of any member of the class, remains an unsolved problem.

The first paper of part II, "A Formal Structure for Multiple-Choice Situations"

by R. R. Bush, Frederick Mosteller, and G. L. Thompson, is a welcome concise statement of the mathematical structure of the Bush-Mosteller stochastic learning model. As is well known, the model can be stated in very general terms, but most of the results and applications assume linear operators. A major and controversial part of the paper is an attempt by means of the "combining of classes and condition," to give a more respectable basis for this assumption than the intriguing observation that it works. Roughly, this condition requires that the model yield the same results whether or not two alternatives with the same set of outcome probabilities are combined. At first glance this seems to have the same status and intuitive necessity as, say, the requirement that the laws of physics shall be independent of the position of the observer; to the extent it has this status and necessity it is exciting. Careful inquiry, however, suggests otherwise, for the probabilities relating outcomes to alternatives are under the arbitrary control of the experimenter; hence, the model must allow for *any* possible combining of classes. It appears to the reviewer that this is too demanding to be considered intuitively necessary, and thus is not really a justification for the linearity assumption. Still a persuasive justification is needed, for the linear model fits an impressive collection of data. An example of such data is presented in "Individual Behavior in Uncertain Situations: An Interpretation in Terms of Statistical Association Theory" by W. K. Estes.

Part III, on utility, includes two papers on the existence of utility functions; these papers are interesting but mathematically the most difficult in the book. The first, "Representation of a Preference Ordering by a Numerical Function" by Gerard Debreu, is concerned with topological conditions on a weakly-ordered set which are sufficient to insure the existence of a utility function. If certain sets are closed, he shows that either separability and connectedness or perfect separability are sufficient. No algebra of probability-combining is assumed as in the von Neumann and Morgenstern theory, but no unique results are obtained. In "Multidimensional Utilities" Melvin Hausner examines the effect of dropping the Archimedean axiom from the von Neumann and Morgenstern axioms. Let ApB denote a probability combination of A and B; the axiom requires that if A is preferred to B, and B to C, then ApC and B are indifferent for some p. The possible objection to the axiom is seen when one lets A = five cents, B = two cents, and C = death. Hausner obtains the elegant results that any non-Archimedean "mixture" space satisfying the other von Neumann and Morgenstern axioms can be imbedded in an ordered vector space, and that any ordered vector space is lexicographically ordered in some basis. Some interesting applications of this theory are suggested by R. M. Thrall in "Application of Multidimensional Utility Theory."

"Towards an Economic Theory of Organization and Information" by Jacob Marschak initiates a fascinating normative study of decision-making by communicating "teams," where teams are defined to be groups with identical individual and group utility functions. A team may collect data, transmit information over a communication network at some cost, and take actions based on a decision rule. Three classes of problems are considered for a team which completes all observation before making any decisions. 1) *Procedural*: given a network and cost of communication, to select the best rules for governing information transmission and actions. 2) *Network*: given rules and a cost function over networks, to select the best communication network. 3) *Constitutional*: to select the best procedural-network pair. Several simple special cases are solved, but as Davis notes (p. 13): "The relatively difficult manipulations required even for these simple cases show for one thing how desirable further development and simplification of the theory would be, while on the other hand they serve to emphasize how difficult would be any analysis at all without the machinery of this formalization."

In the final experimental section, two of the four papers deal with coalition formation in the game-theory sense; both emphasize that psychological rather than "objective"

utilities are necessary for a descriptive theory. In "Tendencies Toward Group Comparability in Competitive Bargaining," Paul Hoffman, Leon Festinger, and Douglas Lawrence employ Festinger's psychological theories of group behavior to predict that those who are perceived as superior in an ability relevant to the conflict of interest involved tend to be excluded from effective bargaining. The confirming experiment was based on a symmetric 3-person game. One player was always a stooge who, in one variation, appeared to be of similar intelligence to the subjects, but who, in the second variation, was evidently of superior ability. In the latter case he was excluded from coalitions more often than in the former, the degree increasing with the importance subjects placed on the game situation. These results strongly suggest that utility functions are subject to modification by psychological manipulations—an unfortunate complication. More directly related to game theory itself is the paper "Some Experimental n-Person Games" by G. Kalisch, J. W. Milnor, J. Nash, and E. D. Nering. Several n-person games ($n = 4, 5, 7$) were run in characteristic function form, i.e., payments were stated for each possible coalition. In each case subjects bargained for 10 minutes, and they reported their agreements to an umpire who enforced them. Considering the rationality assumptions of the theory, the time limit seems questionable. The principal results appear to be: contrary to theory, strategically equivalent games were treated differently; the Shapley value tended to be similar, though by no means identical, to the experimental payments; no satisfactory method was devised to check the von Neumann and Morgenstern theory of solutions. If the authors intended to show that objective payments rather than subjective utilities are sufficient for descriptive purposes, the first result is most disturbing. The failure of the subjects to respond to the objective situation is further confirmed by the authors' observation that the subjects tended to form coalitions having large payments without regard to benefits resulting from other *apparently* less impressive coalitions. While the prospects of positive findings are not great, the experiment probably should be replicated under more carefully controlled conditions and using many more subjects. At that time data could be collected from the subjects prior to each run as to their perceptions of relative coalition strength per coalition member. We do not expect these to be the same as the "rational" ordering derived from the objective characteristic function, but it might be possible to establish that their bargaining behavior is consistent with their orderings. Certainly these two experiments reinforce the contention of von Neumann and Morgenstern that an individual's utility function need not be simply related to any objective measure arising from the situation.

In summary, we may agree with the editors that the book is not definitive and yet recommend it as stimulating and useful for those working in the area. Anyone attracted by any one of the papers will surely be interested in several others, and he may very well have a passing curiosity about most of them.

Center for Advanced Study in the Behavioral Sciences *R. Duncan Luce*