Measurement Scales and Statistics: Resurgence of an Old Misconception

John Gaito
York University, Downsview, Ontario, Canada

A number of years ago, Stevens stated that there was a relationship between psychological measurement scales and statistical procedures such that parametric techniques (e.g., r, F, t procedures) required the presence of at least interval scale data. This idea was used by Siegel and by Senders as a framework for their statistics books. This conception was attacked by a number of statisticians and was shown to be a fallacy. Recently, four books have appeared (by Blalock, by Schmidt, by Sharp, and by Twaite and Monroe), again using the Stevens–Siegel–Senders misconception. This problem is reviewed, and measurement scales and statistical aspects are considered. The misconception was previously and is presently based on a confusion between measurement theory and statistical theory. For statistical tests of null hypothesis, as Lord stated, "the numbers do not know where they came from."

Some recent elementary statistics books show a resurgence of an old misconception, namely, that specific measurement scales (nominal, ordinal, interval, ratio) are included as requirements in the use of statistical procedures. This notion was first suggested by Stevens (1946, 1951) and was later used by textbook writers such as Siegel (1956) and Senders (1958). Many elementary statistics books discuss the four types of scales; however, some statisticians apparently do not realize that there is no relationship between type of scale and statistical techniques used. Recently, some books have again proposed this misconception. Twaite and Monroe (1979), with a classification matrix, Blalock (1979), and Schmidt (1979) remind one of the Siegel and Senders textbooks. This misconception reaches the height of absurdity in the book by Sharp (1979), who proposes a "street map" schema that uses terms such as "level of measurement boulevard," "nominal avenue," "ordinal avenue," and "interval avenue."

These writers apparently do not read the statistical journal literature, inasmuch as a number of articles on this topic showed clearly that measurement scales are not related to statistical techniques. Furthermore, they show little understanding of the mathematical statistics underlying the various statistical procedures. Scale properties do not enter into any of the mathematical requirements for the various statistical procedures. I have not known of any mathematical statistician who agreed with the Stevens misconception. A number have indicated in print that this suggestion is erroneous.

Measurement–Statistics Relationship

Proponents

Stevens (1946) introduced the idea of a measurement scale–statistics relationship. His classification of psychological scales into nominal, ordinal, interval, and ratio categories was a significant contribution to psychophysical and measurement theory. The classification was widely accepted and applied by psychologists in conversation, in teaching, and in publication. Stevens specified the appropriate statistical measures for use with each scale. Thus, nonparametric procedures would be appropriate with nominal and ordinal scales, whereas para-

metric procedures would be required for interval and ratio scales.

This classification led to some misunderstanding with regard to the use of various statistical techniques and to an overemphasis on the utility of nonparametric techniques in psychological research. The book by Siegel (1956) on nonparametric techniques and that by Senders (1958) used Stevens's notion as a framework for a discussion of the various statistical procedures.

The recent books by Blalock (1979), Schmidt (1979), Sharp (1979), and Twaite and Monroe (1979) follow the Siegel–Senders approach in their own unique styles. .

Antagonists

One early criticism of Stevens's idea was the excellent article by Lord (1953). In an entertaining fashion, Lord made the essential point that "the numbers do not know where they came from" (p. 751).

Burke (1953) compared measurement scales and statistical operations; he stated that although the psychological interpretation given to experimental results does take into account the origin of the numbers, this aspect is irrelevant for statistical purposes. Thus, he · concluded that "the properties of a set of numbers as a measurement scale should have no effect upon the choice of statistical techniques for representing and interpreting the numbers" (p. 74). .

Gaito (1959) pointed to possible inconsistencies in the approach. For example, Siegel (1956) listed the Binomial Test as a nominal scale and the Sign Test as an example for ordinal scale data. However, both rely on the binomial distribution. Thus, why should there be the need for two different scales in the case of one underlying distribution? Sharp (1979) is guilty of this same inconsistency.

Gaito (1960) indicated that interval scale aspects were not important for use of analysis of variance (ANOVA) procedures and that the assumptions follow from the mathematical model. The only assumption that resembles the interval scale aspect is the normality one. If the data follow a normal distribution, then the data would be of interval scale nature because the intervals between any data points

are known (in terms of probabilities, i.e., areas under the curve).

Kaiser (1960) reviewed Senders's (1958) book and criticized her attempt to consider both measurement and statistics because "it is clearly a matter of fact that assumption about scales of measurement are irrelevant to *statistical* hypotheses" (italics in original) and because "the book's consideration of scales of measurement seems to muddy the treatment of statistical problems" (p. 413). Furthermore, the book is "confounded by a naive devotion to Stevens' scales of measurement, and apparently written in relatively thoroughgoing ignorance of modern statistical theory" (p. 413).

Boneau's (1961) treatment of the subject was similar to that of Burke (1953); he maintained that the numbers assigned by the measurement operation is a measurement problem and not a statistical one; that is, "the numbers do not known where they came from."

Anderson (1961) addressed Stevens's (1951) statement that a statistic is appropriate for a specific scale if it remains invariant under transformations that leave that scale invariant. He showed that although use of "permissible" statistics for a given scale may guarantee invariance over the class of permissible transformations of that scale, it does not guarantee invariance over the class of transformations that might be used by the investigator. Furthermore, the invariance obtained by permissible transformations of a scale was of relatively minor importance in comparison with other types of invariance, and thus invariance of a statistic under permissible transformations was not a suitable criterion for choosing between statistical procedures. His basic conclusion was that psychological meaning was not a statistical matter, and thus the type of measuring scale used had little relevance to the question of whether to use parametric or nonparametric procedures.

Baker, Hardyck, and Petrinovich (1966) tested Stevens's permissible transformations notion. They constructed three types of distributions (normal, rectangular, exponential) that were of interval scale nature. Then they performed a number of nonpermissible transformations that produced subinterval-type data. Pairs of random samples were selected from the original and transformed distribu-

tions, and t values were obtained for each pair. In most cases the resulting sampling t distributions were similar to the theoretical t distributions. They concluded that "probabilities estimated from the t distribution are little affected by the kind of measurement scale used" (p. 308).

These articles essentially indicated that "the numbers do not know where they came from." Thus, the scale requirements for statistical techniques appeared to be a figment of the imagination of a number of psychologists.

Measurement Scales and Statistics

The idea of a relationship between measurement scales and statistical procedures seems to be based on a confusion between measurement theory and statistical theory (Anderson, 1961; Boneau, 1961; Burke, 1953; Gaito, 1960).

Measurement Theory

In the development of measuring instruments (e.g., tests of intelligence, personality, interests, etc.), one is concerned with their reliability and validity. The validity or authenticity aspect brings into focus the meaning underlying the numbers that are used to indicate amounts of the characteristics of concern. Thus, the numbers used must be meaningful relative to the characteristics of concern.

Stevens, as a psychophysicist, would naturally pay attention to the meaning of numbers. He did a service for measurement theory in his discussions on the four types of scales. Also, it is easy to understand how Stevens might be misled into relating these measurement scales to statistical procedures because the two areas are superficially similar.

Statistical Theory

In the statistical procedures and especially in the tests of null hypotheses, differences and relatedness of numbers are of concern. Thus, meaning of numbers does not enter the picture because, as Lord (1953) stated, "the numbers do not know where they came from." For example, an interval scale assumption was suggested for ANOVA procedures. However, this assumption cannot be found if one looks to the mathematical bases of assumptions (Eisenhart, 1947). Savage (1957), a mathematical statisti-

cian, wrote, "I know of no reason to limit statistical procedures to those involving authentic operations consistent with the scale of observed quantities" (p. 340). Another noted statistician, Kempthorne (1955), after showing mathematically that the normal theory ANOVA test can approximate the randomization test, stated that

this serves as some theoretical basis for the fact which has been noticed by most statisticians, that the level of significance of the analysis of variance test for differences between treatments *is little affected by the choice of a scale of measure for analysis.* (p. 965, italics added for emphasis)

It should be clear that scale properties do not enter into ANOVA assumptions. The mathematical or structural model for each design shows a statement of NID (0, σ_e^2), which explicitly indicates that errors are normally, independently distributed with a mean of zero and one variance, σ_e^2, showing the assumptions of normality, independence, and homogeneity of errors; an assumption of interval scale is nowhere to be found. The only requirements for the use of ANOVA or of any statistical procedure are that the mathematical assumptions underlying the procedure be met or approximated.

Introducing scale aspects as a requirement can be awkward. For example, Siegel (1956) and Sharp (1979) indicated that the Binomial Test was an example for nominal data and that the Sign Test was an example for ordinal data. However, both of these procedures use the binomial distribution. Twaite and Monroe (1979) describe a similarly awkward situation. They classify the normal approximation to the binomial distribution as a procedure for use with nominal or ordered scale data. However, if one takes the scale aspects seriously, then the normal approximation to the binomial distribution should be classified as of interval scale nature. When the normal distribution is applied, the intervals between any two points are well defined in terms of area under the curve, or probabilities.

Conclusions

In mathematical statistics literature one will not find scale properties as a requirement for the use of the various statistical procedures. This requirement was merely a figment of the imagination of a number of psychologists be-

cause of a confusion of measurement theory and statistical theory. Statistical procedures do not require specific scale properties. The assumptions for the use of statistical procedures can be clearly stated and are based on the mathematical aspects underlying the procedures. For example, with ANOVA problems, the statistical requirements are succinctly stated in the mathematical or statistical model. In a simple randomized design, the model states that $y_{ij} = \mu + \alpha_j + e_{ij}$, where the e_{ij}s are NID $(0, \sigma_e^2)$. (y_{ij} is the observed scores; μ is the population mean, estimated by the general mean; α_j is the treatment effect, whose presence or absence is indicated by the difference between groups; e_{ij} is the random or error source, which is represented by the differences within each group; and the es are normally, independently distributed with a mean of 0 and one variance $[\sigma_e]$.) This model indicates the assumptions of normality, independence, and homogeneity of variance. As long as these three assumptions are met, the presence or absence of α can be ascertained by the use of the central F distribution; this distribution is the one that is present when $\alpha = 0$ and when the three assumptions hold. Similar ideas apply for other procedures also.

Thus, although Stevens did a service for measurement theory in developing scale ideas, his notion led to a misconception that has been difficult to eliminate.

References

Anderson, N. H. Scales and statistics: Parametric and nonparametric. *Psychological Bulletin*, 1961, *58*, 305–316.

Baker, B. O., Hardyck, C. D., & Petrinovich, L. F. Weak measurement vs. strong statistics: An empirical critique of S. S. Stevens' proscriptions on statistics. *Educational and Psychological Measurement*, 1966, *26*, 291–309.

Blalock, H. M., Jr. *Social statistics*. New York: McGraw-Hill, 1979.

Boneau, C. A. A note on measurement scales and statistical tests. *American Psychologist*, 1961, *16*, 260–261.

Burke, D. J. Additive scales and statistics. *Psychological Review*, 1953, *60*, 73–75.

Eisenhart, C. The assumptions underlying the analysis of variance. *Biometrics*, 1947, *3*, 1–21.

Gaito, J. Nonparametric methods in psychological research. *Psychological Reports*, 1959, *5*, 115–125.

Gaito, J. Scale classification and statistics. *Psychological Review*, 1960, *67*, 277–278.

Lord, F. M. On the statistical treatment of football numbers. *American Psychologist*, 1953, *8*, 750–751.

Kaiser, H. Review of V. Senders' *Measurement and statistics*. *Psychometrika*, 1960, *25*, 411–413.

Kempthorne, P. The randomization theory of experimental inference. *Journal of the American Statistical Association*, 1955, *50*, 946–967.

Savage, I. R. Nonparametric statistics. *Journal of the American Statistical Association*, 1957, *52*, 331–344.

Schmidt, J. J. *Understanding and using statistics: Basic concepts*. Lexington, Mass.: Heath, 1979.

Senders, V. L. *Measurement and statistics*. New York: Oxford University Press, 1958.

Sharp, V. F. *Statistics for the social sciences*. Boston: Little, Brown, 1979.

Siegel, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

Stevens, S. S. On the theory of scales of measurement. *Science*, 1946, *103*, 677–680.

Stevens. S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

Twaite, J. A., & Monroe, J. A. *Introductory statistics*. Glenview, Ill.: Scott, Foresman, 1979.