RESEARCH METHODOLOGY

Edited by Dev S. Pathak, Ph.D.

PARAMETRIC OR NONPARAMETRIC STATISTICS

Part I: A Dilemma for the Clinical Researcher

This article inaugurates the "Research Methodology" column, to be published in this journal at least four times a year. The column will be devoted to the various elements of research methodology as they are used in investigating problems and questions faced by clinical pharmacists and other health professionals. Those readers who are interested in contributing to this column are invited and encouraged to submit manuscripts. Potential contributors may write or call Dr. Pathak for further information.

A LETTER TO THE EDITOR[1] in the February, 1979, issue of the *American Journal of Hospital Pharmacy* contained a critical evaluation of the data analysis and research design of a previous article.[2] The original article, "Changes in Physicians' Attitudes Toward Pharmacists as Drug Information Consultants Following Implementation of Clinical Pharmacy Services," appeared in the January, 1979, issue of the same journal. One of the comments made in the letter by Bootman and Hammel was that, since the responses for the study were obtained on a "0 to 5" scale, usage of the t-test or F-test on the data was inappropriate. They argued: "In using statistical tests such as the t-test or F-test, the researchers are assuming that the data are interval when, in fact, they are ordinal. A more conservative analytical approach would require that they use nonparametric procedures . . ."[1] The authors of the original study, Nelson, Meinhold, and Hutchinson, responded with the assertion that the F-test is documented to be robust, and hence, "given the general nature of behavioral research, we believe minor violations of the basic mathematical assumptions of the ANOVA model do not discount the general findings of the study."[3]

The disagreement described above is not new. It has persisted for many years between those who are called the "school of 'weak measurement' theorists" and those who might be classified as the "school of 'strong statistics' theorists."[4] The purpose of this article is to provide an overview of this controversy and to offer some guidelines for clinical researchers who are constrained by "weak" data.

DEV S. PATHAK, Ph.D., is Professor of Pharmacy Administration, College of Pharmacy, University of Illinois at the Medical Center, 833 South Wood Street, Chicago, Illinois 60612, 312/996-6300.

Part II of this article will appear in the September 1979 issue of *DICP*.

In 1946, Stevens published a very influential paper entitled, "On the Theory of Scales of Measurement."[5] In that article he identified four distinct scales of measurement — nominal, ordinal, interval, and ratio — and specified the arithmetic operations (and thus, the statistics) which are permissible for each scale.

Nominal-scale data consist of the numbers or letters assigned to observations or to a group of objects. The letters or numbers serve only as labels for categories. For example, a clinical researcher may classify each member of his patient population as "hypertensive," "cardiac," "renal" or "other," and refer to these categories as 1, 2, 3 and 4, respectively. The only requirements for nominal-scale categorization are: (1) that all the categories are mutually exclusive, and (2) that all those objects assigned a number or letter are homogeneous with respect to some attribute. Thus, the only permissible operations on nominal-scale data are frequency statistics and contingency correlation.

The ordinal scale arises from a rank-ordering of observations on a given attribute. However, no indication is given of "how much" of the attribute is possessed by each object or "how far apart" the objects are with respect to the attribute. For example, consider a competency rating on a five-point scale where 1 indicates "not at all competent" and 5 indicates "very competent." For the purpose of descriptive statistics, this type of data should not be added, subtracted, multiplied, or divided. Only the algebra of inequality is applicable to this type of data: this is the reason why the structure of the ordinal scale is called "isotonic," or "order-preserving." This type of scale is most widely used in behavioral science research. The permissible statistical operations for ordinal-scale data are median, percentile, rank-order correlation, and other statistics that do not disturb the order-preserving quality of the data.

An interval scale exists when observations are not only ranked on the basis of a particular attribute, but when the exact distances between the points on the scale are known.* However, no information is available regarding the absolute amount of attribute possessed by each object. For example, the difference between 60° F and 20° F is equal to the difference between 80° F and 120° F; however, it cannot be stated that 60° F is three times as hot as 20° F. In short, an interval scale requires that some sort of unit of measurement be used by the researcher with an *arbitrary* zero point defined for the measurements on the scale. Because of the arbitrary zero, additions and subtractions can be performed on the actual scale value as well as on the intervals; however, multiplication and division can be performed only on "intervals" obtained from the data.[6] In other words, interval-scale data are invariant only under any linear transformation. For this reason, statistics such as the mean, the standard deviation, and the product-moment correlation can be calculated for interval-scale data.

A ratio scale has the same properties as the interval scale with the additional requirement that there be a "rational" or "absolute" zero point defined for the data. For example, consider the height of individuals in inches or feet. The absolute zero for a ratio scale of this type can be implicit in the measurement — i.e., it is impossible to actually measure zero inches in height for any individual, but that zero point anchors the scale nevertheless. Any type of descriptive statistic, including logarithms, can be calculated for ratio-scale data.

In most cases, the distinction between interval and ratio scales is purely academic. Thus, in all instances where the size of the unit can be established, "it is legitimate to use all of the operations of arithmetic, square roots, powers, and logarithms"[7] on the available interval-scale data.

Although Stevens's article discusses only the relationships of measurement scales and descriptive statistics, he indirectly refers to the usage of inferential statistical procedures in his discussion of the invariance of results under scale transformation.[7] He argues that a statistical procedure is considered appropriate for the data only when the appropriate statistic calculated from the original scale remains invariant under the transformation that also leaves the scale of measurement invariant. In other words, "it means that if a statistic is computed from a set of scale values and this statistic is then transformed, the identical results will be obtained as when the separate scale values are transformed and the statistic is then computed from these transformed scale values."[8]

This argument of invariance under transformation forms the basis for a researcher's decision to use parametric or nonparametric statistics. Because no arithmetic operation can be performed on ordinal-scale data without leaving the original scale of measurement invariant, no t-tests, F-tests, or other parametric tests should be performed on ordinal-scale data. Since permissible transformations for parametric tests are linear, the data required for parametric tests should at least meet the criteria established for interval-scale measurement. Furthermore, parametric tests also assume that the population from which samples are drawn is normally distributed, and they require the estimation of at least one parameter (i.e., the population value). Nonparametric tests,** in comparison, do not require interval-scale measurement, estimation of population values, or any assumptions regarding the shape of the population distribution curve.

If the above arguments are accepted as discussed, the conclusion for the clinical researcher is obvious: parametric statistics should be used only when there is evidence that the data collected represent interval- or ratio-scale measurement. But before rushing to any conclusion, the researcher must answer the following questions: (1) Do the data collected have at least interval-scale properties? (2) If interval-scale properties can be established for the data collected, are the assumptions of the statistical procedure selected to analyze the data satisfied?

Unless the answers to both questions are "yes," the researcher should not use parametric tests to analyze the data. To a novice clinical researcher, it would probably seem that most of the studies involving measurement of some attribute of an object will result in a negative answer to one of the two questions, and hence, he should use only nonparametric tests. But in light of the evidence accumulated in the last 20 years, the decision to abandon parametric techniques in favor of nonparametric techniques may be premature.

*In this paper, the term "interval" scales is considered synonymous with "equal-interval" scales. However, it is possible to construct unequal-interval scales.

**Many authors distinguish between "nonparametric" and "distribution-free" statistics. Although there is a specific difference between the two, the distinction does not affect the discussion in this paper. Hence, when the term "nonparametric" is used in this article, it can be considered synonymous with "distribution-free."

References

1. Bootman, J. L. and Hammel, R. J.: Data Analysis in Study of Physicians' Attitudes (letter), *Am. J. Hosp. Pharm. 36*:161, 164 (Feb.) 1979.
2. Nelson, A. A., Jr., Meinhold, J. M. and Hutchinson, R. A.: Changes in Physicians' Attitudes Toward Pharmacists as Drug Information Consultants Following Implementation of Clinical Pharmacy Services, *Am. J. Hosp. Pharm. 35*: 1201-1206 (Oct.) 1978.
3. Nelson, A. A., Jr., Meinhold, J. M. and Hutchinson, R. A.: Reply to Bootman-Hammel Letter (letters section), *Am. J. Hosp. Pharm. 36*:164 (Feb.) 1979.
4. Baker, B. O., Hardyck, Curtis D. and Petrinovich, Lewis F.: Weak Measurements vs. Strong Statistics: An Empirical Critique of S. S. Stevens's Proscriptions on Statistics, *Educational and Psychological Measurements 26*:291-309 (March) 1966.
5. Stevens, S. S.: On the Theory of Scales of Measurement, *Science 103*:677-680 (June) 1946.
6. Nunnally, J. C.: *Psychometric Theory*, McGraw-Hill Book Company, Inc., New York, N.Y., 1967, p. 16.
7. Blalock, H. M.: Social Statistics, 2nd Ed., McGraw-Hill Book Company, Inc., New York, N. Y., 1972, p. 18.
8. Anderson, N. H.: Scales and Statistics: Parametric and Nonparametric, *Psychological Bull. 58*:309 (July) 1961.

RESEARCH METHODOLOGY

Edited by Dev S. Pathak, D.B.A.

PARAMETRIC OR NONPARAMETRIC STATISTICS

A Dilemma for the Clinical Researcher, Part II

This article inaugurates the "Research Methodology" column, to be published in this journal at least four times a year. The column will be devoted to the various elements of research methodology as they are used in investigating problems and questions faced by clinical pharmacists and other health professionals. Those readers who are interested in contributing to this column are invited and encouraged to submit manuscripts. Potential contributors may write or call Dr. Pathak for further information.

ADVOCATES OF THE USE OF NONPARAMETRIC STATISTICS have argued that parametric tests must not be used for purposes of statistical inference, unless the following two conditions are satisfied:

1. The data must exhibit at least interval-scale properties.
2. The assumptions of the selected parametric test must not be violated.

The purpose of this article is to evaluate this controversial issue and its implications for the pharmacy researcher.

DEV S. PATHAK, D.B.A., is Professor of Pharmacy Administration, Department of Pharmacy Practice, College of Pharmacy, University of Illinois at the Medical Center, 833 South Wood Street, Chicago, Illinois 60612, Tel: (312) 996-6300.

To maintain continuity, review the first part of this article, published in the July-August issue, pp. 441-442.

Appreciation is extended to Mr. Lee Simon for his thoughtful comments on the first draft of this paper.

Scale Properties

It is commonly accepted that data exhibiting nominal-scale properties should be analyzed with nonparametric techniques only, whereas either nonparametric or parametric tests may be used for statistical inference based on ratio scale data. Thus, the controversy regarding the selection of an appropriate statistical technique applies primarily to interval- and ordinal-scale data. Siegel, for example, argues that ". . . parametric statistical tests, which use mean and standard deviations [i.e., which require the operations of arithmetic on the original scores] ought not to be used with data in an ordinal scale. The properties of an ordinal scale are *not* isomorphic to the numerical system known as arithmetic."[1] The problem of isomorphism with ordinal-scale data arises because ordinal scales lack two characteristics found in interval scales: (1) an implied zero point and (2) equal intervals.

Without an absolute — implied or real — zero point, addition or subtraction of ordinal-scale data has no

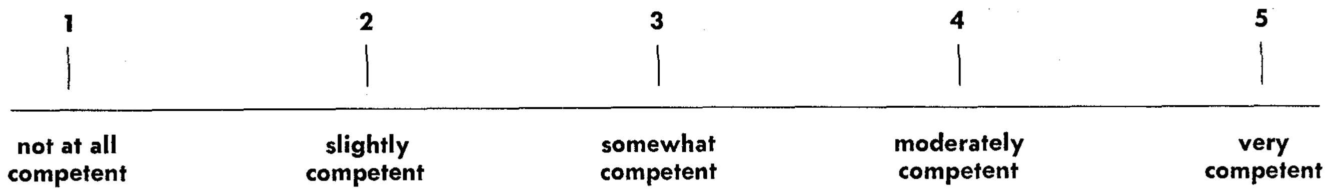

Figure 1. Pharmacist's competency scale

meaning. For example, consider a 5-point scale that is used by physicians to rate a pharmacist's competence (see Figure 1). One extreme of this scale, "not at all competent," is assigned a value of "1," whereas the other extreme, "very competent," is assigned a value of "5." If we assume that the ratings by a physician of the pharmacist's knowledge of adverse drug reactions and contraindications were "3" and "4" respectively, then the equation "3 + 4 = 7" might be thought to represent a point that is 7 scale units (or the average of 3.5 units) above the zero point. However, the physicians might actually consider a "1" rating, "not at all competent," to be a zero value for the pharmacist's competence, and might therefore be assigning the values of 2 and 3, and *not* 3 and 4, as viewed by the researcher. Thus, since the arbitrary zero of the researcher is 1 point below the "real" zero, the total should be 5 (or the average of 2.5) and not 7 (or the average of 3.5).

The absence of an established zero point for ordinal-scale data, however, is not as serious a problem as it seems, for two reasons. First, many psychological scaling techniques have some type of implied zero included in the measurement scales. For example, semantic differential scales[2] that are anchored by two bipolar adjectives or phrases (e.g., good-bad, clean-dirty, kind-cruel) include neutral points that can be considered implied zeros for those scales. Secondly, even with the absence of a zero point on an ordinal scale, addition or subtraction on the "intervals" within the scale is permissible as long as these intervals are equal.[3] However, a researcher's assumption of equality of intervals within an ordinal scale could be in error.

As indicated in Figure 2, the "real" intervals for a physician's rating of a pharmacist's competence may be different from the "equal" intervals assumed by the researcher for his ordinal-scale measurement. Many examples similar to the one used here can be obtained from the literature of behavioral disciplines. Nevertheless, most behavioral scientists continue to use parametric statistics on data exhibiting ordinal-scale properties, thereby assuming that they have achieved equality of intervals. This problem is compounded by the fact that many of these researchers fail either to explicitly recognize this assumption or to attempt to compensate for the obvious inequality of intervals for their data.

With regard to this issue of correcting for unequal intervals, it is possible to approach the "equal interval" condition by using various observation techniques, transformation methods, and scaling procedures, as discussed by Guilford in his book, *Psychometric Methods*.[4] Guilford also suggests that rank-order judgments can be converted to interval scales by using the normalized rank approach or the comparative judgment approach.[5]

Although Guilford's suggestions are useful in approaching the "equal interval" condition in behavioral research, the necessity for their use is debatable. It has been argued, for example, that "most psychological and educational scales approximate interval equality fairly well"[3] and that "there is tolerable error"[6] in applying various parametric statistics to ordinal-scale data. The basis for this argument is the assumption that any researcher using a scaling instrument will follow the proper procedure for designing it. A well-designed instrument will be highly reliable and valid. The high reliability, in turn, may guarantee that each variable is measured by scales that are "substantially and linearly related, and thus equal intervals can be assumed"[3] for these scales. Kerlinger states that "this assumption is valid because the more nearly a relation approaches linearity, the more nearly equal are the intervals of the scales."[3] It is the researcher's responsibility, however, to provide proof of the assumption of "equal intervals" for his data. This is the reason that adherence to established psychometric procedures for the development of survey instruments, and the reporting of internal consistency measures such as coefficient alpha, should be considered absolutely essential for any research that uses scaling or judgmental devices.

Assumptions of Parametric Tests

Apart from the issue of scale properties, a researcher should not use parametric statistical tests when the assumptions permitting the use of these tests are violated. The major assumptions required for the use of parametric techniques such as the t- and F-tests, as described by Siegel,[7] are:[a]

[a] The numerical order of these assumptions as presented in this paper is not the same as outlined by Siegel.

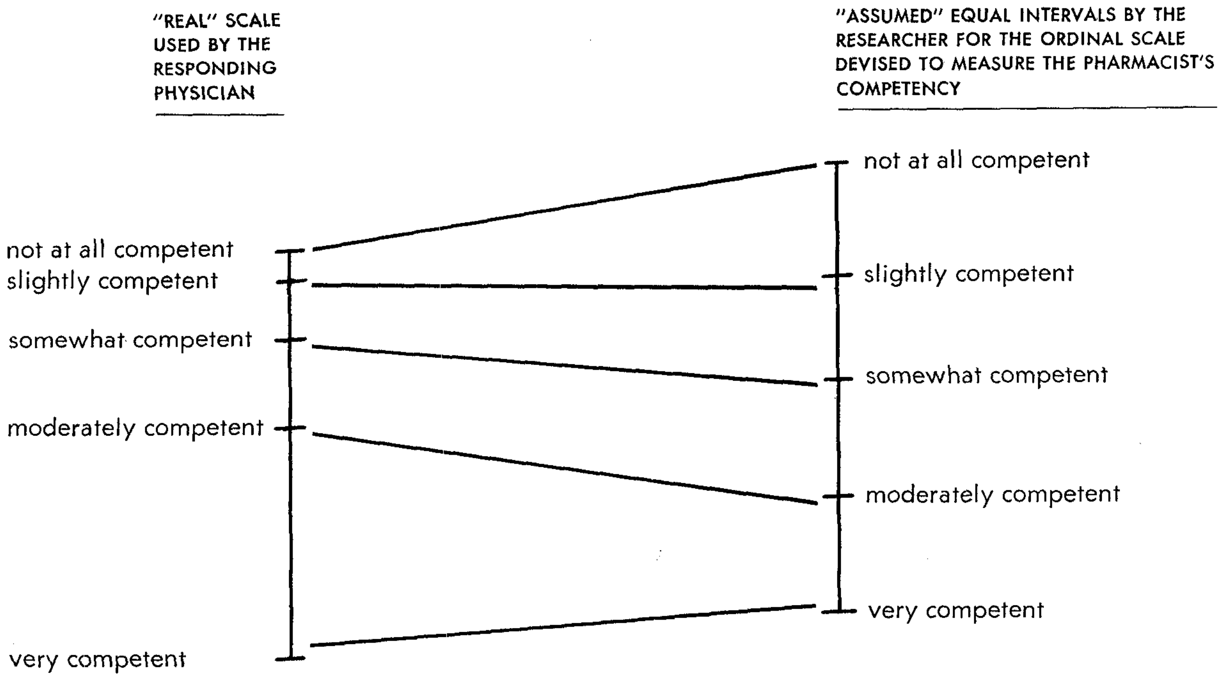

Figure 2. "Real" vs "assumed" intervals for the competency rating of the pharmacist

1. The observations must be drawn from normally distributed populations.
2. The variances of these populations must be equal (or, in special cases, must have a known ratio of variances).
3. The variable(s) involved must have been measured on at least an interval scale.
4. The observations must be independent.
5. For the F-test, the means of these normal and homoscedastic (equal variances) populations must be linear combinations of effects due to columns and/or rows — that is, the effects must be additive.[b]

The first two assumptions — normality and homoscedasticity — are frequently referred to as conditions of equinormality. A statistical test is considered robust when conclusions derived by the application of the test remain consistent even though one or more assumptions behind the test have been violated.

Many empirical studies have been conducted to test the robustness of parametric tests under conditions such as variable shapes of the population distribution, unequal variances, and unequal sizes.[3,8-11] In general, these studies indicate that both the t-test and the F-test are highly robust to both nonnormality and unequal variances.[c] Labovitz, after reviewing several studies, concluded that "[especially with large samples,] the population distribution can be markedly skewed or platykurtic with negligible effects on the t-values."[9] Similarly, Gaito[10] and Cochran[11] cite several sources that support the contention that the F-test is also highly robust to departure from equinormality. Testimony of this type has led Kerlinger to conclude: "The evidence to date is that the importance of normality and homogeneity is overrated, a view that is shared by the author. Unless there is good evidence to believe that populations are rather seriously non-normal and that variances are heterogeneous, it is usually unwise to use a nonparametric statistical test in place of a parametric one."[12]

Despite such enthusiastic endorsements for parametric tests, researchers using the t-test or the F-test may find the following observations useful:

---

[b] The evaluation of these assumptions in this paper borrows extensively from Gaito's work (see reference 13).

[c] The F-test is found satisfactory even under the conditions in which one variance was approximately 45 times as large as the other variance in the study. For details of this study, see: Lindquist, E., reference 3, pp. 78-86. Similar results are reported for the t-test by Boneau, C.: The Effects of Violation of Assumptions underlying the t-test, *Psychological Bulletin 57*:49-64 (Jan.) 1960, and Baker, B. O., Hardyck, C. D., and Petrinovich, L. F. (see reference 13). Since it can be shown that the square of the t-statistic with (n-2) degrees of freedom is the same as the F-statistic with (1, n-2) degrees of freedom under the linearity condition, the evidences of robustness of the t-test are equally applicable to the F-test. For proof of $t^2_{(n-2)} = F_{(1,n-2)}$, see Draper, M. R., and Smith, H.: *Applied Regression Analysis,* New York, John Wiley and Sons, Inc., 1968, p. 25.

1. Although violation of one assumption of a parametric test does not seriously affect the results, violation of two or more assumptions may have a marked effect.[8]
2. Parametric tests are sensitive to heterogeneity of variance when it coexists with groups of unequal sizes. Under these conditions, the probability level becomes two or three times as great as expected.[10]
3. The one-tailed t-test should not be performed when samples are of unequal sizes and are drawn from a badly skewed population.[13]
4. In order to compensate for violations of assumptions of the t-test, the researcher may use two conservative rules: (1) select groups of equal sizes, and (2) use a two-tailed test.[13]
5. When extreme heterogeneity occurs, the researcher may still use normal theory by imposing more stringent significance requirements. For example, he may use alpha = 0.025 rather than 0.05; or alpha = 0.005 rather than 0.01.[14]
6. Various transformation techniques can be used when deviations from equinormality are identified. "Fortunately, if a suitable transformation is chosen, both non-normality and heterogeneity may be reduced in as much as these tend to vary together."[15] The most commonly used transformation techniques[16] are: square-root transformations for Poisson distributions or for data in which the variance is proportional to the mean; arcsin transformations for proportions; and logarithmic transformations for data in which the coefficient of variation is found constant.
7. Every researcher must remember that the "normal distribution" assumption in the analysis of variance refers "to the distribution of those proportions of the data which are used as the appropriate estimates of error and not necessarily to the distribution of the error. This is an important consideration because in multivariable designs the variation within each group may not distribute normally but the portions of certain interaction terms may."[17]
8. Finally, the essential ingredient basic to statistical inference is randomization, but most of the discussion in research methodology literature focuses on the assumption of equinormality. The fact is that the mere act of randomization assures that the usual analysis of variance significance tests are, to a good approximation, nonparametric. Fisher, the developer of the modern approach to the experimental design, investigated the effects of randomization and concluded that "the physical act of randomization . . . affords the means . . . of examining the wider hypothesis in which no normality of distribution is implied."[18] Similarly, Kempthorne argues that: "Tests of significance in the randomization experiment have frequently been presented by way of normal law theory whereas their validity stems from randomization theory."[19] This is the reason that sample selection through the randomization process is vital to any clinical research.

The third assumption of parametric techniques, as suggested by Siegel, is that data should be interval-scaled. In his enthusiasm to make a positive case for nonparametric statistics, Siegel seems to have erred in declaring interval-scale measurement to be a requirement for the performance of parametric tests. The mathematical derivations of parametric tests require no assumptions regarding the properties of the measurement scale of the data investigated. Kempthorne, for example, investigated the mathematical bases for the use of randomization in the analysis of variance tests and stated 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 measurement for analysis."[20] Similarly, using axioms of probability and the axiomatic basis of measurement, Burke 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."[21] In short, statistics as a tool does not distinguish between properties of the data to which it is applied. Hence, the validity of statistical inference from the data is affected by the assumptions of the test and not of the measurement scale.

Even though statistics does not recognize the empirical meaning assigned to numbers, the researcher should. The violation of the interval-scale assumption basically reflects the degree of risk involved in using parametric statistics when the data fail to meet the test of invariance under permissible scale transformation. Since only linear transformations are permissible for the t- and F- tests, should the researcher use these tests on ordinal-scale data that do not remain invariant under linear transformations? Baker et al. have investigated this question by measuring the effects of nonlinear transformations of the unit-interval score on the t-test. They concluded that "strong statistics such as the t-test are more than adequate to cope with weak measurements — and, with some minor reservations, probability estimates from the t-distribution are little affected by the kind of measurement scale used."[22] The minor reservations referred to by Baker et al. can be overcome by using equal sample sizes and a two-tailed test.

The fourth assumption of "independence" between observations is not unique to parametric tests such as the t- and F-tests. It is equally applicable to comparable nonparametric tests.

Finally, the additivity assumption of the analysis of variance is not considered necessary by some writers. Banks argues that "when the differences among treatment effects do not exceed 20 percent of their overall mean, this problem need not be serious, since within that range, the additivity relationship is likely to be a good approximation to almost any type of relationship that may arise."[23] However, if the presence of interaction is found to be consistently additive, researchers can either use transformations or account for these interactions by selecting appropriate ANOVA designs.

Summary

The implications of the arguments presented in this article regarding the usage of parametric tests can be summarized as follows:

1. Even though there is only a tolerable error in applying parametric tests to ordinal-scale data, the researcher interested in using scaling techniques should follow the normal rules of psychometric procedures to obtain acceptable reliability coefficients for his instrument. Nunnally, for example, suggests that in the early stages of research, reliability coefficients of 0.5 to 0.6 are considered "moderately sufficient"; however, in the applied setting, such as recruiting and promotion of personnel in a hospital, a reliability coefficient of 0.9 is considered desirable.[24]

2. There is no statistical requirement prohibiting the use of parametric tests on ordinal-scale data. A researcher who uses equal sample sizes and a two-tailed test need not worry about using parametric tests on ordinal-scale data.

3. To be conservative, a researcher should exercise caution when two or more assumptions of the t- or F-test are violated. In this situation he may use transformations, subdivide the error variance, or omit part of the experiment.

4. Use of randomization in designing an experiment allows the researcher to examine differences between the means without the assumption of normal distribution.

5. The simplest approach to handling violations of the assumptions for parametric tests is to use a lower probability level — i.e., use 0.025 for 0.05, or 0.005 for 0.01, significance level.

This review of the evidence available during the past 20 years indicates that the clinical researcher should not worry about the possibility of errors that will discount his results, when using parametric tests on "weak" measurement, even under conditions of moderate departure from the assumptions of these tests. This is the reason why Nelson et al.[25] are correct in their reply to the Bootman-Hammel[26] criticism — i.e., "given the general nature of behavioral research, we believe minor violations of the basic mathematical assumptions of the ANOVA model do not discount the general findings of the study."[25]

In the final analysis, however, it is the researcher's responsibility to justify the selection of appropriate statistical tools. Otherwise, as Bootman-Hammel[26] pointed out in their letter to the editor, the researcher should clearly identify for the reader the limitations of his statistical analysis. Although parametric tests are highly robust, this does not excuse the investigator ". . . from being alert for intolerable approximations and for results and conclusions that are essentially a function of his faulty application of statistics."[27]

References

1. Siegel, S.: *Nonparametric Statistics for the Behavioral Sciences,* New York, McGraw-Hill Book Company, 1956, p. 26.
2. Osgood, C. E., Tannenbaum, P.H., and Suci, G. J.: *The Measurement of Meaning,* Champaign, IL, University of Illinois Press, 1957.
3. Kerlinger, F. N.: *Foundations of Behavioral Research,* 2nd ed, New York, Holt, Rinehart and Winston, Inc., 1973, pp. 439, 440.
4. Guilford, J. P.: *Psychometric Methods,* 2nd ed, New York, McGraw-Hill Book Company, 1954, pp. 15, 178-222.
5. *Ibid.,* pp. 181-188.
6. Guilford: *op.cit.,* p. 15.
7. Siegel: *op.cit.,* p. 19.
8. Anderson, N. H.: Scales and Statistics: Parametric and Nonparametric, *Psychological Bulletin 58*:305-316 (July) 1961.
9. Labovitz, S.: Some Observations on Measurement and Statistics, *Social Forces 46*:151-160 (Dec.) 1967.
10. Gaito, J.: Nonparametric Methods in Psychological Research, *Psychological Reports 5*:115-125 (June) 1959.
11. Cochran, W. G.: Some Consequences When the Assumptions for the Analysis of Variance Are Not Satisfied, *Biometrics 3*:22-38 (Mar.) 1947.
12. Kerlinger: *op.cit.,* p. 287.
13. Baker, B. O., Hardyck, C. D., and Petrinovich, L. F.: Weak Measurements vs. Strong Statistics: An Empirical Critique of S. S. Stevens's Proscription on Statistics, *Educational and Psychological Measurements 26*:294 (Mar.) 1966.
14. Lindquist, E. F.: *Design and Analysis of Experiments in Psychology and Education,* Houghton Mifflin, New York, N.Y., 1953, p. 86.
15. Gaito, J.: *op.cit.,* p. 118.
16. Snedecor, G. W., and Cochran, W. G.: *Statistical Methods,* 6th ed, Ames, IA, Iowa State University Press, 1972, pp. 325-338.
17. Gaito, J.: Scale Classification and Statistics, *Psychological Review 67*:277-278 (July) 1960.
18. Fisher, R. A.: *The Design of Experiments,* 7th ed, New York, Hafner, 1960, p. 45.
19. Kempthorne, O.: The Randomization Theory of Experimental Inference, *Journal of American Statistical Association 50*:947 (Sept.) 1955.
20. *Ibid,* p. 958.
21. Burke, C. J.: Additive Scales and Statistics, *Psychological Review 60*:75 (Jan.) 1953.
22. Baker, et al.: *op.cit.,* p. 309.
23. Banks, S.: *Experimentation in Marketing,* New York, McGraw-Hill Book Company, 1965, p. 76.
24. Nunnally, J. C.: *Psychometric Theory,* New York, McGraw Hill Book Company, 1967, p. 226.
25. Nelson, A. A., Jr., Meinhold, J. M., and Hutchinson, R. A.: Reply to Bootman-Hammel Letter (letters section), *Am. J. Hosp. Pharm. 36*:164 (Feb.) 1979.
26. Bootman, J. L., and Hammel, R. J.: Data Analysis in Study of Physicians' Attitudes, (letters section), *Am. J. Hosp. Pharm. 36*:161, 164 (Feb.) 1979.
27. Guilford, *op. cit.,* p. 16.