COMPARING DATA ANALYSIS TECHNIQUES FOR EVALUATION DESIGNS WITH NON-NORMAL POPULATIONS

Elaine S. Jeffers, University of Maryland, Eastern Shore*

The data collection phases for evaluation designs may involve analysis of data from distributions that are skewed, restricted in range, lacking in variability or otherwise possessing characteristics rendering it difficult to assume normality of the mean and variance of the parent population. When assumptions have been violated, the following are some examples of transformations that can be used to result in a closer approximation to the normal distribution:

1. The square-root transformation

This transformation is useable when data are in the form of frequency counts as in "yes" "no" responses. The transformed score (X') would be:

$$X' = \sqrt{X}$$

2. The arc sine transformation

This transformation is useable when scores are in proportions, as in "percentage correct responses." The transformed score (X') would be:

$$X' = \arcsin \sqrt{X}$$

3. The logarithmic transformation

This transformation is useable when the data are decidedly skewed. The transformed score (X') would be:

$$X' = \log X$$

If the measures are small, the transformation might be:

$$X' = \log (X + 1)$$

Transformations can be used for fairly complex designs. They are also useable when the researcher wants to present a more elaborate analysis which would extend beyond a classical analysis of variance designs. Transformed data, however, changes the relationships between the statistics observed from the original data. Hence, when generalizing results, caution may need to be used in relating the results of the converted measures to the original measures.

Non-parametric and other non-traditional tests are other alternatives that have been used when assumptions of a normal distribution have not been met. May and Konklin developed a simplified version of the nonparametric counterpart to the F test for trend (4). Cooper used a non-parametric test for increasing trend by making comparisons of all possible pairs of observations (1). Lewis and Johnson illustrated the use of Kendall's Coefficient of Concordance for the evaluation of the extent of agreement among a set of judges, each of whom ranks in entirety a set of objects (3). These authors demonstrated that critical values of "W" are given only by approximation and only for large numbers. It was shown that Kendall's Coefficient of Concordance can be used to test more specific hypothesis about agreement on the criteria of choice within a group of judges (3). Simms and Collons illustrated the use of an algorithm for the Coefficient of Concordance which allows for the identification of the particular combination of individuals in a group, and all possible combinations of such individuals who exhibit or display the greatest amount of consensus (7).

Reynolds has stressed that indiscriminate use of tests of significance contributes to theoretical inadequacies in terms of scope and explanatory power (5). Young has stressed that a large part of the data which social scientists must deal with is not distributed according to known mathematical functions; non-parametric techniques (order statistics) offer the opportunity to test hypothesis which require no assumption about the form of the distribution of the population (8). Siegel has offered a variety of such non-parametric techniques which are particularly useable at the nominal and ordinal level of measurement (6).

Non-parametric tests can be used for pilot testing of instruments within an evaluation design, where assumption of normality cannot be taken for granted. In an abbreviated pilot-tested needs analysis procedure, first and second year university students ranked their preferences for the following goals for an educational research center:

1) Develop skills in evaluation of Educational Programs.

2) Develop skills in behavioral objectives and criterion-referenced testing--applications and development.

3) Develop knowledge of Instructional Systems Development.

4) Improve personal proficiency with Instructional Systems Development.

5) Make contacts with happenings in Educational Research.

6) Earn a PHD.

7) Build an Academic Recommendation.

8) Obtain tools needed for Educational Research--math, statistics, computer operations and research design.

9) Become familiar with Professional Societies, and Journals of Educational Research.

10) Take courses in other centers to keep abreast of most recent trends in Education.

11) Understanding of knowledge of diffusion.

12) Developing Skills in Curriculum Development.

13) Further knowledge in concept of Creativity and Research.

14) Gain realistic and practical experience in Evaluation.

15) Relate Evaluation and Teaching Skills to Humanistic Education Field.

The sign-rank test was used to test the hypothesis of no differences in ranking of these goals by first and second year university students as follows:

Ho (Null Hypothesis)..There is no difference in the ranking of these goals as perceived by first and second year students; alpha level equal .05.

H_1 (Alternative Hypothesis)..There is a difference in the ranking of these goals as perceived by first and second year students.

Table I presents an application of the sign-rank test used to test differences for these purposes. The null hypothesis tested by the sign test is that:

$$p(XA > XB) = p(XA < XB) = \frac{1}{2}$$

Another way of stating the null hypothesis is that the median difference is zero. Focus is placed on the direction of the differences rather than the size of the differences. Reference is made to Table D in Siegel--Table of Probabilities Associated with Values as Small, as observed Values of x in the Binomial Test (B). The probability of getting this result for a two-tailed test is .180. This is outside the rejection region and the decision is to accept the null hypothesis.

Non-parametric tests can be used in the preliminary stages of validating an instrument that will later be used for evaluation purposes. Table II illustrates the use of the Coefficient of Concordance to test the degree of agreement among Freshman, Sophomore, Junior and Senior students, concerning the quality of library services. Comparable groups of students from each of these four levels were asked to rank the quality of library services in four different testing situations using different instruments. The results (W = .99) indicated a high degree of agreement among the four groups of students.

Table III lists a variety of non-parametric methods which can be used at various levels of measurement (nominal, ordinal and interval). A researcher who decides on the use of a non-parametric test must decide on the most appropriate test to use in terms of the level of measurement. This is an important consideration since the use of a non-parametric test yields less powerful results when a parametric tests can be used for the same purpose. The researcher must decide upon the relative consequence of the Type I or the Type II error. This is a decision that should be made based on the type of variables that are being investigated and the sampling environment. Hence, in researching the effects of a treatment to reduce the cause of cancer or in evaluating the effects of an educational program, it is important that Type II errors not be made.

The power of non-parametric methods relative to Type I and Type II errors have been summarized in detail by Festinger and Katz (2). A Type I error is made when the decision is made to reject the null hypothesis when it should be accepted; a Type II error is made when the decision is made to accept the null hypothesis when it should be rejected. Power = 1 minus the probability of a Type II error; stated alternatively, power is the probability of appropriately rejecting the null hypothesis.

Within an evaluation setting, generalizability of results is not necessarily a goal. However, it is very important that results be valid for a particular decision-making setting. Failure to adequately consider the importance of this factor results in the possibility of collecting data that lacks decision-maker validity. Hence, in an evaluation setting, the decision to use a parametric or a non-parametric test would depend on the relative consequences of the power of that test on decision-making accuracy.

*The author wishes to acknowledge appreciation of the University of Mass. and Dr. N. Nagarajan formerly of the University of Maryland, Eastern Shore, for data presented.

Table I

An Application of the Sign-Rank Test of Differences in
1st and 2nd Year Student Responses to Group Goals

2nd Year (a)	1st Year (B)	2nd yr-1st yr. (A-B)	Direction of Difference	Signs
6	5	+1	XA > XB	+
4	2	+2	XA > XB	+
4	2	+2	XA > XB	+
2	2	0	XA = XB	0
7	5	+2	XA > XB	+
6	2	+4	XA > XB	+
5	0	+5	XA > XB	+
7	8	-1	XA < XB	-
7	4	+3	XA > XB	+
7	2	+5	XA > XB	+
3	4	-1	XA < XB	-
2	3	-1	XA < XB	-
4	3	+1	XA > XB	+
5	6	-1	XA < XB	-
4	2	+2	XA > XB	+

x = number of minority signs = 4 N = Number of non-zero ranks = 14

Table II

Coefficient of Concordance

Groups	Conditions I	II	III	IV
A	4	2	3	1
B	4	3	2	1
C	4	3	2	1
D	4	3	2	1
R_j	16	11	9	4

Sum or ranks = 40 Mean of R_j = 10

$$W = \frac{s}{\frac{1}{12} K^2 (N^3 - N)} \qquad W = 0.99$$

s = Sum of squares of observed deviations from mean of R_j

k = # of sets of rankings--# of judges (i.e. A,B,C,D)

n = # of entities (objects or individuals) ranked (i.e. Conditions I-IV)

$\frac{1}{12} K^2 (N^3 - n)$ = max. possible sum of squared deviations , i.e. sum "s" which would occur with perfect agreement among K ranks

TABLE III (From Siegel's Non-Parametric Statistics)

	NONPARAMETRIC STATISTICAL TEST**				NONPARAMETRIC MEASURE OF CORRELATION
	Two-sample case		k-sample case		
One-sample case (Chap. 4)	Related samples (Chap. 5)	Independent samples (Chap. 6)	Related Samples (Chap. 7)	Independent Samples (Chap. 8)	(Chap. 9)
Binomial test, pp. 36-42 x^2 one-sample test, pp. 42-47 (Nominal measurement)	McNemar test for the significance of changes, pp. 63-67 (Nominal measurement)	Fisher exact probability test, pp. 96-104 x^2 test for two independent samples, pp. 104-111 (Nominal measurement)	Cochran Q test, pp. 161-166 (Nominal measurement)	x^2 test for k independent samples, pp. 175-179 (Nominal measurement)	Contingency coefficient C, pp. 196-202 (Nominal measurement)
Kolmogorov-Smirnov one-sample test, pp. 47-52 One-sample runs test, pp. 52-58 (Ordinal measurement)	Sign test, pp. 68-75 Wilcoxon matched-pairs signed-ranks tests,*** pp. 75-83 (Ordinal measurement)	Median test, pp. 111-116 Mann-Whitney U test, pp. 116-127 Kolmogorov-Smirnov two-sample test, pp. 127-136 Wald-Wolfowitz runs test, pp. 136-145 Moses test of extreme reactions, pp. 145-152 (Ordinal measurement)	Friedman two-way analysis of variance, pp. 166-127 (Ordinal measurement)	Extension of the median test, pp. 179-184 Kruskal-Wallis one-way analysis of variance, pp. 184-193 (Ordinal measurement)	Spearman rank correlation coefficient: r_s, pp. 202-213 Kendall rank correlation coefficient: r, pp.213-223 Kendall partial rank correlation coefficient: $r_{zy,z}$, pp. 223-229 Kendall coefficient of concordance: W, pp. 229-238 (Ordinal measurement
	Walsh test, pp. 83-87 Randomization test for matched pairs, pp. 88-92 (Interval measurement)	Randomization test for two independent samples, pp. 152-156 (Interval measurement)			

**Each column lists, cumulatively downward, the tests applicable to the given level of measurement. For example, in the case of k related samples, when ordinal measurement has been achieved both the Friedman two-way analysis of variance and the Cochran Q test are applicable.

***The Wilcoxon test requires ordinal measurement not only within pairs, as is required for the sign test, but also of the differences between pairs. See the discussion on pp. 75-76.

References

(1) Cooper, M. "A Non-Parametric Test for Increasing Trend". Educational and Psychological Measurement, Vol. 35, 1975, pp. 303-306.

(2) Festinger, L. and Katz, D. Research Methods in the Behavioral Sciences. New York: Holt, Rinehart and Winston, Inc., 1966, pp. 536-577.

(3) Lewis, G. H. and Johnson, R. G. "Kendall's Coefficient of Concordance for Sociometric Rankings with Self Excluded". Sociometry, 1971, Vol. 34, No. 4, December 1971, pp. 496-503.

(4) May, R. and Konkin, P. "A Non-Parametric Test of an Ordered Hypothesis for K Independent Samples." Educational and Psychological Measurement, Vol. 30, 1970, pp. 251-258.

(5) Reynolds, R. Jr. "Replication and Substantive Import: A Critique on the Use of Statistical Inference in Social Research". Sociology and Social Research, Vol. 53, No. 3, April 1969, pp. 299-310.

(6) Siegel, S. Non-Parametric Statistics for The Behavioral Sciences. New York: McGraw-Hill Book Co., 1956.

(7) Simms, C. and Collons, R. "Sociological Research Applications of an Algorithm for the Coefficient of Concordance". American Sociologist, Vol. 4, No. 4, November 1969, pp. 321-324.

(8) Young, P. Scientific Social Surveys and Research. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1966. pp. 340-341.