EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT
1991, 51

DEALING WITH NONNORMAL DATA: PARAMETRIC ANALYSIS OF TRANSFORMED DATA VS NONPARAMETRIC ANALYSIS

JEFFREY LEE RASMUSSEN
Indiana University-Purdue University at Indianapolis

WILLIAM P. DUNLAP
Tulane University

Researchers have typically employed parametric analysis of raw data to test experimental data for statistical significance. When the data are not normally distributed, data transformation or nonparametric analysis are often recommended. The present study compares parametric analysis of raw data to parametric analysis of transformed data and to nonparametric analysis when the tests are carried out under population nonnormality. The results of a Monte Carlo simulation indicate that when distributions depart markedly from normality, nonparametric analysis and parametric analysis of transformed data show superior power to parametric analysis of raw data. Furthermore, under the conditions studied, parametric analysis of transformed data appears to be somewhat more powerful than nonparametric analysis.

THE consequence of using parametric statistical tests on nonnormal data has been a concern of researchers and statisticians for over half a century. Mathematically, the assumption of underlying population normality is necessary for the calculated statistics to be distributed according to available tables. Researchers and statisticians have considered three major approaches to dealing with

Correspondence concerning this article should be addressed to Jeffrey Lee Rasmussen, Department of Psychology, Purdue School of Science, Indiana University-Purdue University at Indianapolis, 1125 East 38th Street, Indianapolis, Indiana 46205-2810.

nonnormal data: (a) parametric analysis of the raw data, (b) parametric analysis of transformed data, and (c) nonparametric analysis. There is a great deal of controversy surrounding the appropriateness of each approach. The purpose of this paper was to consider the advantages and disadvantages of these techniques. The present paper then carries out a comparison of the three techniques in terms of their Type I error rate and power.

Parametric Analysis of Raw Data

Some authors have recommended analysis of nonnormal data with parametric tests. For example, Glass, Peckham, and Sanders (1972) reviewed a number of studies that investigated the Type I error rate and power of parametric tests when they were calculated on samples drawn from nonnormal populations. Glass et al. (1972) concluded that parametric tests are essentially "distribution free" tests; they concluded that many parametric tests were not seriously affected by violation of assumptions. Bradley (1968, 1978), however, has challenged the robustness of parametric tests to assumption violation. He has shown, for example, that the Type I error rates of many parametric tests differ markedly from their nominal levels when normality and homogeneity assumptions are violated.

The studies by Bradley have been criticized as using "unrealistic" data distributions and overly stringent significance levels (Glass et al., 1972). Bradley (1978), however, has argued that his distributions are based on actual reaction time data, and that the empirical Type I error rates are unacceptable even for the traditional .01 and .05 significance levels.

It is clear that there is some controversy surrounding the use of parametric analysis of nonnormal data. In spite of this controversy, it appears that many researchers do not seem to show much concern about the distributional characteristics of their data. A popular statistical software user's manual offers these comments on this neglect:

> There is . . . little doubt that in many instances statistical techniques are being utilized by both students and researchers who understand neither the assumptions of the method nor their statistical or mathematical bases. There can be little doubt that this situation leads to some "garbage-in garbage-out" research (Nie, Hull, Jenkins, Steinbrenner, and Bent, 1970, p. 3).

Parametric Analysis of Transformed Data

Some researchers have advocated transforming data prior to analysis (Levine and Dunlap, 1982, 1983; Rasmussen, 1985b; Smith, 1976; Tabachnick and Fidell, 1983). There are several reasons why data transformation may be recommended. Some reasons—such as transformation to eliminate interaction—are not germane to the present investigation; therefore they are not discussed. The main benefit of data transformation, however, is that it may lead to a substantial increase in the power of parametric tests. For example, Levine and Dunlap (1982) demonstrated that a log transformation applied to log-normally distributed data could lead to a substantial increase in the power of the F test.

The value of data transformation has been questioned by Games (Games, 1983, 1984; Games and Lucas, 1966). The Games and Lucas (1966) article compared the power of the F ratio on nonnormal and normal data. They found that data transformation did not lead to much improvement in power; indeed in some cases, the power for the F ratio calculated on nonnormal data was higher than that for the F ratio calculated on the normal data.

In addition to the issue of power, Games (1983, 1984) has also pointed out that the difficult aspect of data transformation is selecting the *appropriate* transformation. He cited Dixon and Massey (1969) who stated that "without prior information, the experimenter has little chance of proving or disproving normality or of finding a normalizing transformation from samples as small as 10 or 20" (p. 182). The Games and Lucas (1966) and the Levine and Dunlap (1982) studies had this prior information; as such their studies do not accurately reflect the researchers dilemma. Although there may be theoretical advantages to data transformation, it is possible that the practical difficulties may outweigh these advantages.

Nonparametric Analysis

Nonparametric tests have been favored over parametric tests by some (Bradley, 1968; Lehmann, 1975) because they purportedly require less stringent assumptions regarding the distribution of the underlying population. As such, nonparametric tests are sometimes recommended in place of parametric tests when parametric assumptions are not met (e.g., Runyon and Haber, 1984). Although nonparametric tests are often regarded as less powerful than parametric tests, Blair and Higgins have demonstrated that nonparametric tests may be much more powerful than parametric tests when data are sampled from certain nonnormal distributions (Blair and Higgins, 1980, 1985; Blair, Higgins, and Smitley, 1980). For example, Blair and Higgins (1980) demonstrated that the nonparametric Wilcoxon W test had a large power advantage over the parametric Student's t when they were calculated on samples drawn from a common mixed normal distribution.

Rasmussen (1985b), however, argued that as the mixed-normal distribution that Blair and Higgins used represented an outlier model, it was, therefore, inappropriate to calculate a parametric test on such data. He proposed instead that the data should first be corrected for outliers and then that a parametric test should be applied. He demonstrated that the power of the parametric test calculated on outlier-corrected data was superior to that of the nonparametric test.

Present Study

The present study is concerned with three approaches to dealing with nonnormal data. The first approach—parametric analysis of raw data—can perhaps be best thought of as the "baseline" or "modal" approach. That is, it is the method most commonly employed by researchers today. A quick glance at recent issues of psychology journals indicates that most researchers use parametric tests to analyze their data (or some linear transformation of the raw data, such as Z or T scores). As has been indicated earlier, such an analysis is usually carried out with little regard to the distributional characteristics of the data. Although in some cases this disregard may be justifiable, in other cases it clearly is not. For example, Rasmussen (1991) has pointed out that researchers continue to analyze raw absenteeism scores with parametric tests in spite of Hammer and Landau's (1981) caution that these scores are highly nonnormal.

If the first approach can be thought of as the modal approach, the other methods can be regarded as two contrasting alternatives to the modal approach. That is, researchers can deal with the lack of congruence between a test that assumes normality and a data set that is not normally distributed with either of two contrasting approaches. Researchers either can transform the data in order to meet the normality assumption, or can dispense with the assumption of normality by using a nonparametric test. In either case the incongruity is resolved.

On the basis of previous literature, it seems reasonable to state two hypotheses. First, the nonparametric technique should show superior power to the parametric technique when they are compared with a markedly nonnormal population. This prediction is consonant with the results of Blair and Higgins (1980, 1985). Second, the parametric technique on transformed data should show superior power to the parametric test on raw data when they are calculated on markedly nonnormal data. This hypothesis is suggested by the results of the Levine and Dunlap (1972) study. They found that the power of a parametric test was improved when the data were transformed toward normality. Games (1984), however, criticized their study as being unrealistic because they knew a priori what the transformation to normality was. The present study, however, does not use a priori information in selecting the appropriate data transformation; rather, the transformation, which is based upon sample characteristics, is selected by an algorithm by Rasmussen (1985a).

Additionally, and perhaps more importantly, the present study allows for a direct comparison of the two classic approaches to dealing with nonnormal data. That is, the Levine and Dunlap (1982) study indicates that researchers may benefit from data transformation, whereas the Blair and Higgins (1980, 1985) studies reveal that researchers may benefit from nonparametric tests. The present study compares the two alternatives.

Method

A FORTRAN program was written to simulate three approaches to analysis of skewed data: (a) parametric F ratio carried out on raw data (FR), (b) parametric F ratio applied to transformed data (FT), and (c) nonparametric Kruskal-Wallis H employed with the ranked data (NP). The Type I error rate and power of the three approaches was investigated under the effects of five parameters: (1) number of groups, (2) sample size, (3) effect size, (4) underlying population, and (5) significance level.

The study was concerned with a two-group condition and a three-group condition. The sample sizes per group for the two-group simulation were 3, 9; 6, 6; 9, 27; 18, 18; 27, 81; and 54, 54. For the three-group simulation they were 2, 4, 6; 4, 4, 4; 6, 12, 18; 12, 12, 12; 24, 48, 72; and 48, 48, 48.

The program simulated a "zero difference between group means" condition in order to estimate the Type I error rates of the three approaches. The power of the tests was estimated by simulating 15

TABLE 1
Parameters of Four Populations Used in the Simulation

	Normal deviates			Nonnormal deviates			
Pop	lmbd	var	mean	mean	var	skew	kur
1	0.5	0.75	4.00	1.990	0.191	−0.3	0.4
2	−1.5	0.50	2.75	0.234	0.075	2.3	13.9
3	−1.0	0.75	3.00	0.360	0.142	4.7	56.5
4	−3.0	1.00	5.00	0.011	0.017	8.6	147.3

Note. Pop = population number, lmbd = lambda, var = variance, kur = kurtosis.

effect sizes. The effect size was set to $e = b / \sqrt{N}$, where b was varied from 0.5 to 7.5 in 15 increments of 0.5; and where N was the total sample size. The square root of N divisor was used for convenience. Using this divisor made the results easier to table than they would have been had this term not been used. For example, a quick glance at the results will show that the range of power values displayed for the 6/6 simulation is fairly close to the range of values for the 54/54 simulation. If the square root of N divisor were not used, then the larger N simulations would have reached full power at very small effect sizes, or conversely the smaller N simulations would have displayed only little power at large effect sizes.

The three data analysis strategies were evaluated on four different nonnormal populations. The populations were selected to have skew and kurtosis values in line with previous computer simulation and empirical studies (Rasmussen, 1991). Mean, variance, skew, and kurtosis for the nonnormal populations, as well as the underlying parameters, are given in Table 1.

Each population was generated in the following manner. First, normal deviates with mean and variance as shown in Table 1 were generated by using the Ahren and Dieter algorithm (in Lehman, 1977). Next, the normal deviates were exponentiated by the lambda value shown in the table—a procedure which resulted in the indicated mean, variance, skew and kurtosis for the nonnormal population. For example, with the first population, normal deviates with a mean of 5.0 and a variance of 1.0 were generated, and the inverse cube root of these values was taken. This procedure resulted in a distribution with skew of approximately 8.6 and kurtosis of 147.3. The values for the nonnormal population were estimated by calculating the average values of these statistics for 50 data sets of 1000 deviates per set.

The three techniques were compared at the approximately .01 and .05 two-tailed significance levels. Because there are not H critical values that correspond exactly to the .01 and the .05 significance levels for all sample sizes, the H and the F critical values were obtained as follows. First, the values that came closest to without exceeding the .01 and .05 probability levels for H were located in Fix and Hodges (1955), Iman, Quade, and Alexander (1975), and Wilcoxon, Katti, and Wilcox (1975). These references contain critical values for the 3, 9; 6, 6; 9, 27; 2, 4, 6; and 4, 4, 4 conditions. The critical values for the larger N simulations were derived by using the chi square approximation. Next, the critical values of F that corresponded with these H probability values were calculated by employing a program by Dunlap and Duffy (1975).

The five parameters were factorially crossed to give a 2 (number of groups) $\times$ 12 (sample sizes) $\times$ 16 (effect sizes) $\times$ 4 (underlying populations) $\times$ 2 (significance levels) design with a total of 3072 conditions.

The Type I error rate was calculated as follows. For each of the null hypothesis true conditions 2000 replications were carried out. For each replication, a total of N values was generated and distributed among the g groups. The F and H statistics were calculated on the data. The data were then transformed by using an algorithm by Rasmussen (1985a). This algorithm selects a transformation that optimally meets the assumptions of within-group normality and homogeneity of variance. An F ratio was then calculated on the transformed data. The Type I error rate for each statistic was estimated by the proportion of the 2000 statistics that exceeded the appropriate .01 and .05 critical values.

The power of the tests was estimated in a manner similar to the estimation of the Type I error rate. For the power estimations, 500 replications were carried out. For the 15 $e > 0$ conditions, data were generated for the groups, and an effect was added to the first group (the effect was added prior to the exponentiation, as is recommended by Levine and Dunlap, [1982]). The F and H statistics were calculated and compared with the appropriate critical values. The power of a test, at a given effect size, was the proportion of the 500 replications in which the obtained statistic exceeded the critical value.

Results and Discussion

The results are shown in Tables 2 to 5. To conserve space, only the two-group and the .05 simulations are shown. (The .01 and three-group simulations are available from the first author.) In general the results indicated that the F ratio calculated on trans-

TABLE 2
Type I Error Rates and Power of F Test on Raw Data (FR), F Test on Transformed Data (FT) and Kruskal-Wallis Test (NP) for Population 1

			Effect															
*n*2	*n*1	test	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75
3	9	FR	35	3	5	7	14	21	32	35	47	60	71	69	83	88	94	95
		FT	42	4	7	8	15	23	32	36	50	62	74	72	84	90	95	95
		NP	18	2	5	4	8	14	20	26	34	45	56	56	69	76	82	86
6	6	FR	38	4	7	14	23	29	41	53	65	74	83	89	94	96	98	99
		FT	43	4	8	14	23	29	42	55	66	75	83	90	94	97	99	**
		NP	29	3	5	9	14	20	31	42	54	65	74	83	88	92	96	97
9	27	FR	52	6	11	14	22	27	41	48	61	73	79	88	90	94	96	99
		FT	53	6	12	14	21	27	41	48	61	74	79	87	90	94	96	98
		NP	45	7	12	13	20	25	40	45	60	71	76	85	88	93	95	98
18	18	FR	48	6	10	15	26	36	50	61	76	86	91	95	98	99	**	**
		FT	46	6	10	14	25	35	49	60	76	85	90	95	98	99	**	**
		NP	46	6	8	13	23	34	48	58	73	85	89	93	96	99	**	**
27	81	FR	51	7	7	13	20	29	42	53	66	73	83	89	91	95	98	99
		FT	48	6	7	13	19	28	40	52	65	71	82	89	91	95	98	99
		NP	45	7	8	13	20	29	42	53	65	71	81	90	91	95	97	99
54	54	FR	44	4	8	15	24	36	49	64	77	86	92	94	98	99	**	**
		FT	41	4	7	14	23	34	48	62	75	85	91	94	98	99	**	**
		NP	43	4	8	15	24	36	49	64	76	85	91	94	98	99	99	**

Note. For Tables 2 to 5, decimal places have been omitted: Type I error rates carried out to three places (e.g. 8 = .008); power rates carried out to two places.
** = 1.00.

formed data (FT) showed superior power to the *F* ratio based upon raw data (FR). Likewise, the nonparametric Kruskal-Wallis test (NP) demonstrated superior power to the *F* ratio on raw data. The extent of superiority of the FT and NP approaches appeared to be strongly related to the extent of nonnormality. That is, the more nonnormal the underlying populations were (i.e., the larger the skew and kurtosis values) the greater the superiority of the alternative approaches to the modal approach.

The two alternative approaches showed similar power in many instances. In some cases, however, the data transformation approach showed superior power to the nonparametric approach. This outcome was especially true with the smaller sample sizes.

The results are consistent with those of Blair and Higgins (1980, 1985). The present study used nonnormal populations somewhat different from those that they had employed; however, the same results were found that the nonparametric test was superior to the parametric test on nonnormal data. The current investigation also demonstrates that the results of the latter study generalize to the three group case using the Kruskal-Wallis test.

The findings also are consistent with those of Levine and Dunlap

TABLE 3

Type I Error Rates and Power of F Test on Raw Data (FR), F Test on Transformed Data (FT) and Kruskal-Wallis Test (NP) for Population 2

			Effect															
*n*2	*n*1	test	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75
3	9	FR	36	3	5	8	14	23	36	45	58	66	74	77	85	87	88	90
		FT	47	6	10	20	30	50	63	77	84	92	98	97	99	**	**	**
		NP	18	2	6	9	18	29	42	58	69	80	89	90	96	98	99	99
6	6	FR	31	4	11	22	37	49	66	79	87	94	96	97	98	99	99	99
		FT	46	5	13	26	44	57	72	88	93	99	99	**	**	**	**	**
		NP	29	3	9	17	34	45	61	79	88	94	98	99	99	**	**	**
9	27	FR	44	6	11	17	28	41	62	71	83	93	96	97	99	99	99	**
		FT	52	7	18	24	39	55	74	85	91	96	99	**	**	**	**	**
		NP	45	7	17	22	35	53	70	81	90	96	98	99	**	**	**	**
18	18	FR	43	7	14	25	47	64	81	89	95	98	99	**	**	**	**	**
		FT	45	8	17	30	50	67	85	90	97	99	**	**	**	**	**	**
		NP	46	7	16	29	48	62	83	91	97	99	99	**	**	**	**	**
27	81	FR	56	5	10	18	31	49	62	78	89	95	97	99	**	**	**	**
		FT	49	6	13	22	41	56	69	82	93	98	99	99	**	**	**	**
		NP	47	5	14	25	42	56	71	85	93	98	99	99	**	**	**	**
54	54	FR	44	9	14	25	50	64	80	90	96	98	**	**	**	**	**	**
		FT	37	9	14	27	52	70	83	93	97	99	**	**	**	**	**	**
		NP	39	11	12	29	52	71	81	93	98	99	**	**	**	**	**	**

TABLE 4

Type I Error Rates and Power of F Test on Raw Data (FR), F Test on Transformed Data (FT) and Kruskal-Wallis Test (NP) for Population 3

			Effect															
*n*2	*n*1	test	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75
3	9	FR	34	2	3	4	6	8	15	17	23	30	41	46	48	58	59	66
		FT	45	5	7	11	15	21	31	42	51	59	70	80	82	90	95	96
		NP	18	2	4	5	9	11	18	27	33	40	52	62	66	75	79	87
6	6	FR	30	3	6	10	17	26	36	46	52	66	70	80	85	88	89	94
		FT	39	4	8	15	22	33	44	56	65	76	81	88	93	97	98	**
		NP	24	2	5	9	16	22	34	44	52	62	71	80	86	94	94	99
9	27	FR	41	4	5	6	11	18	26	31	42	53	64	74	78	85	89	90
		FT	48	5	7	11	17	27	37	47	61	66	78	88	92	96	99	99
		NP	46	5	6	10	15	27	34	44	54	65	77	85	90	94	98	98
18	18	FR	39	5	8	15	21	31	43	54	65	74	84	93	93	97	98	99
		FT	42	6	10	15	23	35	48	61	73	81	89	95	97	99	99	**
		NP	46	6	8	15	23	35	47	59	70	79	88	94	96	99	99	99
27	81	FR	55	5	6	9	15	21	29	37	46	57	71	78	83	89	92	94
		FT	49	6	8	13	18	27	40	49	57	70	82	89	91	97	97	98
		NP	60	6	8	14	21	30	42	49	60	71	81	89	91	96	98	98
54	54	FR	46	7	8	15	23	33	43	56	70	79	84	90	95	98	98	**
		FT	37	6	8	16	25	37	48	65	76	82	90	95	98	**	99	**
		NP	41	7	9	16	28	38	51	65	79	82	91	95	99	99	99	**

TABLE 5
Type I Error Rates and Power of F Test on Raw Data (FR), F Test on Transformed Data (FT) and Kruskal-Wallis Test (NP) for Population 4

			Effect															
$n2$	$n1$	test	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75
3	9	FR	36	2	4	3	5	6	12	15	19	24	28	32	39	47	47	54
		FT	47	5	8	9	18	27	36	42	55	64	78	75	85	90	96	95
		NP	18	2	5	4	8	14	20	26	34	45	56	56	69	76	82	86
6	6	FR	30	3	5	11	19	22	30	43	52	60	68	76	84	87	88	90
		FT	46	4	8	14	25	31	43	56	67	77	83	90	94	97	99	**
		NP	29	3	5	9	14	20	31	42	54	65	74	83	88	92	96	97
9	27	FR	44	5	6	8	12	15	25	31	42	48	60	69	72	83	84	90
		FT	52	6	12	14	21	27	41	48	62	74	79	88	90	95	96	99
		NP	45	7	12	13	20	25	40	45	60	71	76	85	88	93	95	98
18	18	FR	40	6	9	13	21	31	43	53	67	78	85	89	96	97	98	99
		FT	45	6	9	14	25	35	49	59	75	85	90	94	97	99	**	**
		NP	46	6	8	13	23	34	48	58	73	85	89	93	96	99	**	**
27	81	FR	54	6	5	8	15	19	30	39	46	56	68	79	80	89	94	94
		FT	46	6	6	13	20	27	39	51	63	71	80	88	90	96	98	99
		NP	45	7	8	13	20	29	42	53	65	71	81	90	91	95	97	99
54	54	FR	46	5	7	13	21	27	43	52	67	78	87	88	95	98	98	99
		FT	39	3	7	13	23	33	48	62	73	84	91	94	98	99	**	**
		NP	43	4	8	15	24	36	49	64	76	85	91	94	98	99	99	**

(1982). In addition, the present study indicates that data transformation becomes increasingly useful when data are sampled from increasingly more nonnormal populations. Unlike the Levine and Dunlap (1982) study, in which data transformation values were known a priori, the investigators in the present study selected the appropriate data transformation on the basis of sample characteristics. Games (1983, 1984) criticized Levine and Dunlap's (1982) investigation as being unrealistic because the appropriate data transformation was known a priori, a criticism that does not apply to the present study.

Finally, the current investigation allows for a direct comparison of the results of the Blair and Higgins (1985) and the Levine and Dunlap (1982) studies. In general, data transformation appears to be superior to nonparametric analysis—at least for the two tests investigated.

REFERENCES

Blair, R. C. and Higgins, J. J. (1980). The power of t and Wilcoxon statistics, a comparison. *Evaluation Review*, 4, 645–655.

Blair, R. C. and Higgins, J. J. (1985). Comparison of the power of the paired samples t test to that of the Wilcoxon's signed rank test under various population shapes. *Psychological Bulletin*, 97, 119–128.

Blair, R. C., Higgins, J. J., and Smitley, W. D. S. (1980). On the relative power of the *U* and *t* tests. *British Journal of Mathematical and Statistical Psychology*, 33, 114–120.

Bradley, J. V. (1968). *Distribution-free statistical tests*. Englewood Cliffs, NJ: Prentice-Hall.

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31, 144–152.

Dixon, W. J. and Massey, F. J. (1969). *Introduction to statistical analysis*, (3rd ed.). New York: McGraw-Hill.

Dunlap, W. P. and Duffy, J. A. (1975). FORTRAN IV functions for calculating exact probabilities associated with *z*, Chi-square, *t*, and *F* values. *Behavior Research Methods and Instrumentation*, 7, 59–60.

Fix, E. and Hodges, J. L. (1955). Significance probabilities of the Wilcoxon test. *Annals of Mathematical Statistics*, 26, 301–312.

Games, P. A. (1983). Curvilinear transformations of the dependent variable. *Psychological Bulletin*, 93, 382–387.

Games, P. A. (1984). Data transformation, power, and skew: A rebuttal to Levine and Dunlap. *Psychological Bulletin*, 95, 345–347.

Games, P. A. and Lucas, P. A. (1966). Power of the analysis of variance of independent groups on nonnormal and normally transformed data. EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 26, 311–327.

Glass, G. V, Peckham, P. D., and Sanders, J. R. (1972). Consequences of failure to meet the assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research*, 42, 237–288.

Hammer, T. H. and Landau, J. (1981). Methodological issues in the use of absence data. *Journal of Applied Psychology*, 66, 574–581.

Iman, R. L., Quade, D., and Alexander, D. A. (1975). Exact probability levels for the Kruskal-Wallis test. In H. L. Harter and D. B. Owen (Eds.), *Selected tables in mathematical statistics, Vol. 3*. Providence Rhode Island: American Mathematical Society.

Lehman, R. S. (1977). *Computer simulation and modeling: An introduction*. Hillsdale, NJ: Erlbaum.

Lehmann, E. L. (1975). *Nonparametrics*. San Francisco: Holden Day.

Levine, D. W. and Dunlap, W. P. (1982). Power of the *F* test with skewed data: Should one transform or not? *Psychological Bulletin*, 92, 272–280.

Levine, D. W. and Dunlap, W. P. (1983). Data transformation, power, and skew: A rejoinder to Games. *Psychological Bulletin*, 93, 596–599.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., and Bent, D. H. (1970). *SPSS statistical package for the social sciences, second edition*. New York: McGraw-Hill.

Rasmussen, J. L. (1985a). Data transformation maximizing homoscedasticity and within-group normality. *Behavior Research Methods, Instruments and Computers*, 17, 411–412.

Rasmussen, J. L. (1985b). The power of student's *t* and Wilcoxon *W* statistics: A comparison. *Evaluation Review*, 9, 505–510.

Rasmussen, J. L. (1991). Data transformation and absenteeism. *Methodika*.

Runyon, R. P. and Haber, A. (1984). *Fundamentals of behavioral statistics*, Fifth Edition. Reading, MA: Addison-Wesley.

Smith, J. E. K. (1976). Data transformation in analysis of variance. *Journal of Verbal Learning and Verbal Behavior*, 15, 339–346.

Tabachnick, B. G. and Fidell, L. S. (1983). *Using multivariate statistics*. New York: Harper and Row.

Wilcoxon, F., Katti, S. K., and Wilcox, R. A. (1975). Critical values and probability level for the Wilcoxon rank sum test and the Wilcoxon signed ranks test. In H. L. Harter and D. B. Owen (Eds.). *Selected tables in mathematical statistics, Vol. 1*. Providence Rhode Island: Mathematical Society.