# INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

**The quality of this reproduction is dependent upon the quality of the copy submitted.** Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Order Number 9503924

Robustness of parametric and nonparametric tests when distances between points change on an ordinal measurement scale

Chen, Andrew Hwa-Fen, Ph.D.

University of North Texas, 1994

Chen, Andrew H., <u>Robustness of Parametric and</u>
<u>Nonparametric Tests When Distances between Points Change on</u>
<u>an Ordinal Measurement Scale</u>. Doctor of Philosophy
(Management Science), August, 1994, 143 pp., 14 tables, 10
figures, appendix, bibliography, 98 titles.

Ordinal measures currently enjoy widespread use.
The use of a limited number of categories with equal lengths
between points (the placements of categorical labels) on the
scale may cause a loss of information.  Many well-known
studies claim that only nonparametric tests are appropriate
for the nominal or ordinal level data, that parametric tests
are appropriate for the interval or ratio level data; and
ordinal numbers (equal lengths between points) cannot be
meaningfully added, subtracted, multiplied, and divided.
This citation has been one of the most controversial
statements in applied statistics.

The purpose of this research was to evaluate the
effect on parametric and nonparametric tests using ordinal
data when the distances between points changed on the
measurement scale.  The research examined the performance of
Type I and Type II error rates using selected parametric and
nonparametric tests.

Three experiments were conducted by generating
simulation data on a seven-point Likert-scale using one

uniform, three normal, and three gamma populations. Various
unequal distance changes were made between points in the
four phases of the experiments.

One and two random samples of simulation data were
selected from seven populations. Selected parametric tests
and nonparametric tests were used to examine the equality of
means, medians, deviations, and distributions between two
populations. Several computer programs were written in
FORTRAN 77 to implement the algorithm of data simulation,
parametric tests, and nonparametric tests. The simulation
data were tested by the programs.

The results were analyzed in terms of Type I and
Type II error rates with different sample sizes,
populations, and phases. In summary, the nonparametric
tests produced the same results when the distances between
points changed on the scale. However, parametric tests show
different results when the distances between points changed.
The power of parametric and nonparametric tests were
evaluated as underlying assumptions were violated in the
location parameters.

ROBUSTNESS OF PARAMETRIC AND NONPARAMETRIC TESTS

WHEN DISTANCES BETWEEN POINTS CHANGE
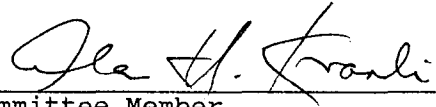
ON AN ORDINAL MEASUREMENT SCALE
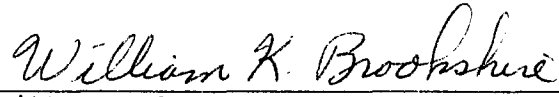
Andrew H. Chen, B.S., M.B.A., M.S.


APPROVED:

_____
Major Professor

_____
Minor Professor

_____
Committee Member

_____
Committee Member

_____
Chair of the Department of Business Computer
    Information Systems

_____
Dean of the College of Business Administration

_____
Dean of the Robert B. Toulouse School of
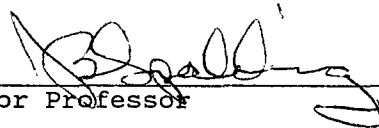    Graduate Studies

ROBUSTNESS OF PARAMETRIC AND NONPARAMETRIC TESTS

WHEN DISTANCES BETWEEN POINTS CHANGE

ON AN ORDINAL MEASUREMENT SCALE


DISSERTATION


Presented to the Graduate Council of the

University of North Texas in Partial

Fulfillment of the Requirements


For the degree of


DOCTOR OF PHILOSOPHY


By


Andrew H. Chen, B.S., M.B.A., M.S.

Denton, Texas

August, 1994

## TABLE OF CONTENTS

iii

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1

INTRODUCTION

## Purpose of the Study

The debate and research over the use of parametric
tests when underlying assumptions are not met have attracted
considerable attention for numerous years (Stevens 1951;
Siegel 1956; Anderson 1961; Lodge et al. 1976; Siegel and
Castellan 1988). The problem of using ordinal measures
(category scaling) or numeric estimates (magnitude scaling)
on applications has spanned almost the same time period as
the debate on the use of parametric tests (Guilford 1954;
Guilford and Dingman 1955; Goodman 1979; Agresti 1990).

Ordinal measures arise when observations from a
population fall into one of several ranked labels or points
(the placements of categorical labels) on the scale.
Respondents are often asked to use a scale with equal
distance between points to express attitudes regarding
stimulant objects in the behavioral, psychophysical,
psychological, medical sciences, and many other areas.
Numeric estimates are used to determine the direction and
strength of people's beliefs according to how respondents
rate an object by selecting one of a fixed number of options
(Lodge 1987). The distances between any two numeric
estimates may not have equal distances on the underlying

1

continuum.

Frequently, the ranked labels have equal distances between points (or labeled positions) on the scale. However, the descriptive stimuli and meanings that different respondents attach to the adjectives will influence the perceived distance between points (Stevens 1957; Cliff 1959; Labovitz 1967, 1970, 1972; Crask and Fox 1987). Such varieties of measurement may or may not affect the equal interval properties that the ordinal scale possesses.

The purpose of this research is to evaluate the effect on parametric and nonparametric statistical tests when the distances between points change on the measurement scale. The research examines the performance of the Type I and Type II error rates using selected parametric and nonparametric tests. The principal area of the research concerns the relationship between the level of measurement scales and the performance of the examined statistical tests.

### Problems Addressed by the Study

The main problems of research interest concerning parametric and nonparametric tests include the following:

1. Can one use ordinal data to do addition, subtraction, multiplication, or division?

2. Is it necessary to choose simple nonparametric tests to analyze ordinal data?

3

3. Do ranked ordinal data deny researchers legitimate access to more powerful parametric tests?

Harvard psychologist S. S. Stevens (1946), the pioneer of scaling (Bolanowski and Gescheider 1991; Teghtsoonian 1991), originally developed a set of techniques for the study of the intensity of sensations, and proposed four levels of measurement: nominal, ordinal, interval, and ratio scales. One of the major criticisms of Stevens' distinctions of measurement pertains to the definitions of ordinal and interval scales. Coombs (1952) and many other researchers have considered various scales that lie between the ordinal and interval levels. Peterson and Sharma (1977) claimed that the determination of measurement levels should consider information transmission from the objects to the respondents. Wildt and Mazis (1978) suggested that academic and industrial researchers investigating the issue of appropriate scale measures should be concerned with constructing instruments that contain unequal distances between labeled points.

In addition, Siegel and Castellan (1988) noted that independent and random observations, homogeneity of variance, normally distributed populations, and at least an interval scale were elements of the parametric statistical model. The Siegel and Castellan study contains one of the most controversial statements in the applied statistical community. Specifically, opposing researchers claim that

parametric tests can be performed on ordinal data as well as on interval or ratio data. The opposing group concludes that the use of parametric tests on ordinal data will not result in a biased evaluation (Borgatta 1968; Kim 1975; Gaito 1980; Bunger 1988; Gautam 1992).

Another group of researchers (Ramsay 1973; Peter 1979; Cox 1980; Gregoire and Driver 1987) favors the information transmission of ordinal measures and meaningful statistical tests. However, this group concludes that ordinal data must be categorized into an optimal number of response labels on the scale. The group also suggests using an appropriate number of labels ranging from as few as two classes to as many as fifty adjectives on the scale, depending on information transmission requirements (Guilford 1954; Garner 1960; Myers and Warner 1968; Cox 1980; Agresti 1990). Past studies generally indicate that the optimal number of response labels used is seven. The next most preferred numbers are five and nine labels. Whether the selection of a label number is appropriate or not depends on the performance (i.e., the result of error rates) of statistical tests using the ordinal data with selected optimal label numbers. The selection of label numbers determined by convenience and simpleness often causes both a loss of information between labeled points and poor statistical tests. Therefore, Gregoire and Driver (1987) noted that the view favoring an optimal number of labels and meaningful statistical tests was generally recommended, but

the view might not have a convincing argument if the statistical test results were significantly unacceptable, such as in the case of high error rates.

The actual positions of labeled points or ordinal scores on quantitative or qualitative categorizations are unclear (Agresti 1990). Previous views support the measures and statistical tests based on the main concerns of variation on the label numbers and equal distances between labeled points. However, little work has been done on the performance of the Type I and the Type II errors over empirical statistical tests when the sampling data are classified from equal distances to unequal (changing) distances between points on the scale.

Ordinal measurement and numeric estimation have achieved widespread use in recent decades. For example, Moses (1984) indicated that ordinal data analysis occurred in 32 of 168 papers published in volume 36 (1982) of the New England Journal of Medicine. Not surprisingly, the measurement techniques for unequally spaced scales have only recently been developed, but the techniques have ties to practical fields, such as marketing, consumer behavior, customer response, quality circles, employee performance, legal issues, and counseling (Perreault and Young 1980; Tull and Hawkins 1987; Markland 1989).

## Significance of the Study

A significant aspect of this research is presented in table 1, which illustrates category matching with real positions along a scale. Five categories (C1 to C5) of scaling represent the underlying continuum on which the stimuli of interest are equally spaced. Three positions of numeric estimates (P1, P2, and P3) are accurately assigned along the scale. However, P1 and P2 are classified in the same category, C2, using five-category scaling, but P1 and P2 are considered as different positions in terms of magnitude scaling. The category scaling process causes a loss of information when ordinal measures are used. A scale with unequal distances between points may be more appropriate. Therefore, changing the distance between points on the scale may provide superior information transmission to the traditional category scaling.

The current research investigated the robustness of empirical parametric and nonparametric tests when distances between points changed. The numeric estimation was used to avoid losing information. The conclusions provided insight and suggestions for the application of ordinal scaling.

TABLE 1

MATCHING OF CATEGORY AND REAL POSITION

```
                P1      P2                    P3
                1.2     1.8                   3.4
                 ↓       ↓                     ↓
I----------I-----------I-----------I-----------I----------I
      C1           C2          C3          C4          C5
0            1            2            3            4            5
```
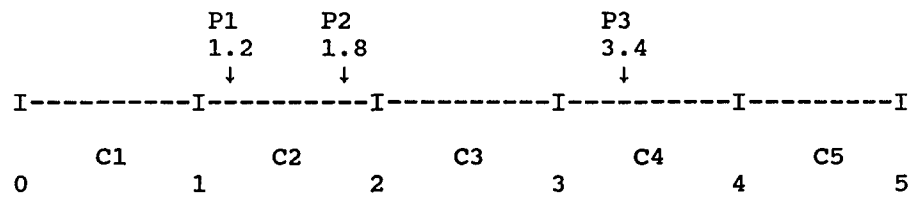
Notes: [1]P1, P2, and P3 represent three real positions of numeric estimates (1.2, 1.8, and 3.4) within the underlying continuum.

[2]C1, C2, C3, C4, and C5 represent five categories within the underlying continuum.

CHAPTER 2

LITERATURE REVIEW

Ordinal Rating Scales

Stevens (1957) indicated that an ordinal rating
scale was the measurement obtained when a subject evaluated
a set of stimuli in terms of categories labeled either by
adjectives or numbers.  To avoid the errors of time-order
and hysteresis (a lagging effect) during evaluation,
successive stimuli should be judged randomly.  With the
advent of Stevens' contribution, the range and strength of
responses became much easier to determine.  Proponents noted
that ordinal data could be performed with the admissible
operations of equivalence (=) and greater than (>) only; and
that any treatment of ordinal data in a manner not
consistent with the operation principle of the ordinal level
(namely, the class of admissible calculation
transformations) was unjustified (Champion 1967; Westermann
1983; Siegel and Castellan 1988).  Supporting views
concluded that ordinal measures could not be meaningfully
added, subtracted, multiplied, or squared (Schlinger 1979).
To compute a mean, standard deviation, or variance with
ordinal data was simply meaningless because summation,
division, and squaring were performed on rank values
(Townsend and Ashby 1984; Ross 1989).  Therefore, Siegel and

8

Castellan (1988) stated that parametric tests could be
applied only to interval and ratio scales of measurement.
This line of reasoning is presented implicitly or explicitly
in a majority of statistics textbooks today (Torgerson 1958;
Stockton 1962; Wasan 1970; Agresti 1984).

Since the ordinal measurement setting was proposed,
challenges to the appropriateness of mathematical
calculations on only interval and ratio data have occurred
(Traylor 1983). Researchers with opposing views maintain
that ordinal measures are similarly represented by a set of
numbers, so that ordinal numbers behave just the same as
cardinal numbers (Vidali 1975). In other words, one can
perform mathematical calculations on ordinal measures just
as on interval or ratio measures.

## Psychological Law

In the late nineteenth century, Fechner proposed
that sensation was logarithmically related to stimulus
intensity. However, the law was not found through other
researchers' experimentations. Few documents could support
Fechner's law (Stevens 1957). This history lesson appears
to leave extensive work to be done by twentieth century
researchers.

Stevens (1957) first confirmed a general
psychological law that equal stimulus ratios produced equal
subjective ratios. Stevens also conveyed a widely accepted
power law that the sensation $\delta$ was proportional to the

stimulus S raised to a power n, and that k was a constant. Thus,

$$\delta = kS^n.$$

When the equation was converted into logarithms, the equation became linear as

$$\log \delta = n\log S + \log k.$$

The equation could be plotted as a straight line on log-log coordinates. The slope of the line corresponded to the exponent n of the power law.

Furthermore, Stevens (1957) illustrated numeric estimation that required the respondent to assign numbers to a series of stimuli, using a standard, to make the estimates proportional to the magnitudes of the sensation or the perceived behavior dimensions. Both the psychological law and numeric estimation are essential to numerous subsequent studies. One of the earliest attempts to use numeric estimation was the evaluation of the seriousness of theft (Sellin and Wolfgang 1964). Respondents were instructed to assign numbers (0 to 10) to different descriptions of criminal offenses based on the seriousness of the crime. Similar studies have been replicated many times in the United States, Canada, and England (Figlio 1976, 1978). The subjects included court judges, police officers, college students, prison inmates, prosecuting attorneys, and many others.

## Magnitude Scaling

Adapting from Stevens' psychological law and numeric estimation (1957), Lodge (1987) presented a significant contribution in the scaling area: magnitude scaling. Magnitude scaling required numeric estimation (assigning numbers) associated with line production (drawing lines) to express the direction and strength of judgment. The evaluation rule was that, the stronger the judgment on an object, the larger the number and the longer the line. The judgment was based on a reference number or a reference line, that was given as a standard for all respondents. The responses of numeric estimation and line production generally were found to follow the power law (psychological law) and could be plotted on log-log (or ratio-ruled) graph paper. The premise of the log-log graph was that a linear relationship (represented by a solid line) existed between the pairs of numeric estimates and line lengths. The slope of the line on the log-log graph was the exponent of the power law. Theoretically, a slope of 1.0 indicated that a perfect ratio judgment was made between the numeric estimation and the line production. Any deviation from the perfect slope of 1.0 indicated that respondents could not make consistent ratio judgments over the perceived magnitudes of stimuli. Magnitude scaling was initially proposed in order to validate psychophysical scaling (i.e., the scales of sensory tests, light intensity, or numeric estimation), but it has since been widely used in marketing

and social behavior sciences (Bollen and Barb 1981; Anderson and Zeithaml 1984; Neibecker 1984; Gautam 1992). Teas (1987) recently extended magnitude scaling to examine the differences between individuals and groups. The subjects were asked to evaluate the features of a cash management account package service. A reference number 100 and a respective reference line were provided to aid in the rating. The regression analysis and correlation coefficient were used to test estimation validity and within-subject measurement reliability. The result strongly supported the validity of using numeric estimation when aggregate (group) data were analyzed. However, when the data were tested at the individual level, considerable heterogeneity among individuals was found. The exponents of the power law (psychological law) varied considerably from individual to individual.

Most studies of magnitude scaling have concluded that numeric estimation has definite advantages over ordinal measurement when used to explore rating scale values in applications. However, little has been done to investigate parametric and nonparametric test comparisons between ordinal measures (ordinal scale) and numeric estimates (interval scale).

<u>Parametric and Nonparametric Tests</u>
<u>on Ordinal Data</u>

Many studies have pointed out that parametric statistical tests provide satisfactory power and robustness

when compared to corresponding nonparametric tests in the evaluation of ordinal data. The use of ranked order data on parametric tests is not inconsistent insofar as significance level and relative efficiency are concerned.

Dixon and Massey (1957) found that ranked ordinal tests were almost as powerful as parametric tests when the data met the assumptions of selection as independent random samples from normally distributed populations with equal variances. The finding was in conflict with Siegel and Castellan's belief (1956 and 1988) that the use of a parametric test required at least an interval scale of data. In recent years, however, studies have often supported the use of more powerful parametric tests on ordinal measures (Gautam 1992; Kimeldorf, Sampson, and Whitaker 1992).

Anderson (1961) investigated parametric tests (i.e., $t$ test, $F$ test, analysis of variance, and regression analysis) and nonparametric tests (i.e., the Wilcoxon $T$, the Kruskal-Wallis $H$, and median tests) using different measurement data. Results indicated that both parametric and nonparametric tests were appropriate for ordinal, interval, and ratio data.

Woods (1972) took ranked data with equal and unequal sample sizes from normal and nonnormal distributions and compared the parametric analysis of variance results to the nonparametric alternatives. Woods concluded that, in nearly every experiment, the parametric analysis of variance

approach was more powerful and robust than the nonparametric
alternative.  In addition, Woods stated that one could also
use the $F$ test on ordinal data to investigate the equality
of two variances from two sampling populations.

More recently, Gregoire and Driver (1987) randomly
generated Likert-scale data with various symmetric and
nonsymmetric populations.  The Likert-scale scores had equal
distance between points on the scale.  With different sample
sizes, parametric and nonparametric tests were conducted to
investigate the effect on the Type I and the Type II errors.
Gregoire and Driver concluded a variety of important
outcomes as follows:

1.  Low error rates (the Type I and the Type II errors) were
    shown when testing (parametric) the equality of means
    using samples from symmetric populations.

2.  Parametric and nonparametric tests were equally powerful
    and had consistently low Type I and II error rates when
    the differences between medians were examined.

3.  Some Type I rates were higher (using nonparametric
    tests) on the test of equal variance when samples were
    selected from symmetric populations.

4.  The error rates (using both parametric and nonparametric
    tests) appeared to be sensitive and unreliable when
    populations changed from symmetric to nonsymmetric.

Thus, Gregoire and Driver (1987) noted that the
parametric test was as powerful as the nonparametric test

when samples were selected from symmetric populations. Both test results were less precise when samples were selected from nonsymmetric populations. Gregoire and Driver also concluded that a pretest using various stimuli with unknown variances between populations was advisable when testing the equality of central tendency.

If a valid statistical experiment is conducted and both parametric and nonparametric tests can be used in the evaluation of observed data (in the same level of measurement), then the parametric test is preferable for its popularity and convenience (Siegel and Castellan 1988). Strict adherence to this rule is essential to the effective selection and use of statistical tests. Numerous recent studies have employed parametric tests to analyze the effect on ordinal data, but most studies have nothing to do with the scaling of unequal distances between points on the measurement scale.

### Parametric and Nonparametric Tests on Equal and Unequal Distances

The ordinal measures which have been pervasively analyzed are obtained from equally spaced categories. In contrast, the numeric estimation (used in the current research) has unequal distances between points on the scale.

Crask and Fox (1987) completed an examination of the scale distances between points on the evaluation of consumer behavior characteristics: purchase intent, product importance, and overall product rating. Marketing

specialists usually used these three purchasing factors with equal distance between categories. Crask and Fox conducted a survey using numeric estimation associated with the reference number 100. The conclusions revealed that overall product rating showed nearly equal distances between points on the scale. Purchase intent and product importance had unequal distances between points. However, no further parametric and nonparametric tests were conducted to investigate the effect of changing distances.

Parasuraman and Varadarajan (1988) made significant progress in examining the robustness of ordinal data with changing distances between points. The samples were selected from a simulated population based on marketing strategy variables. The selected ordinal data were systematically transformed from equal distances to unequal distances between points by the following methods:

1. Uniform distance: equal distances between points
2. Bias toward middle: longer distance of positions on the midpoint of the ordinal scale
3. Bias toward extremes: longer distance of positions on the two ends of the ordinal scale
4. Bias toward high end: longer distance of positions on the upper side of the ordinal scale
5. Bias toward low end: longer distance of positions on the lower side of the ordinal scale

Parasuraman and Varadarajan (1988) used additive and multiplicative regression models to evaluate the variation of correlation coefficients when distances between points changed. They stated the following results:

1. Transforming the ordinal scale to the interval scale (by changing distances between points) failed to show a serious limitation on the regression analysis result as long as the ordinal data could be assumed to vary monotonically with systematic patterns (i.e., the five previous transformation methods).

2. The transformation might not be appropriate when the bias (changing distances between points) was subject to the nonlinear distortions of the assumed ordinal scale.

The conclusion failed to show an optimistic result when distances changed dramatically throughout the underlying continuum. No comparisons had been done between parametric and nonparametric tests. Parasuraman and Varadarajan (1988) left many problems and recommendations for further studies; for example, a noticeable change occurred in the regression analysis (i.e., slopes and correlation coefficients) as soon as the ordinal assumptions were not met. Further samples selected from multiple populations might provide useful insights into scaling techniques.

Although most studies are somewhat encouraging to applied statisticians engaged in the numeric estimation area

today, there is a need to enhance the experiment with changing distances made between points. The use of more sophisticated techniques to test the equality of central tendency, dispersions, or distributions would provide reliable results in the applications. In addition, respondents often hesitate to assign very unfavorable or negative numbers on the evaluation scale (Brown, Copeland, and Millward 1973). An unbalanced scale in which the distances between favorable labels are not equal to the distances between unfavorable labels (widely used by General Foods Corporation) may provide more useful information (Wildt and Mazis 1978).

Therefore, the current research indicates that a fruitful possibility for further studies would be to examine the performance (i.e., the Type I and the Type II errors) of parametric and nonparametric tests when distances between points change on the scale. Past studies have not analyzed the comparisons between parametric and nonparametric tests on the scale with unequal and unbalanced distances on two sides of the neutral point. The objective of this research is to investigate part of the unresolved areas.

CHAPTER 3

RESEARCH FRAMEWORK AND DESIGN

In the current research, seven different populations
(one uniform population, three normal populations, and three
gamma populations) were generated by simulation data.  A
mapping function was used to change each population into a
seven-point frequency population.  Then several unequal and
unbalanced distance changes were made between points on the
scale in the four phases.  The objective was to examine the
effect of changing distances between points using selected
parametric and nonparametric tests on simulation data.  The
Type I and the Type II error rates were used to evaluate the
test performance with a variety of population changes.


Hypotheses

Three experiments were conducted in the current
research.  Experiment 1 dealt with one sample selected from
one population;  experiment 2 dealt with two samples
selected from one population; and experiment 3 dealt with
two samples selected from two different populations.  Table
2 highlights the comparison of the three experiments.

Experiment 1, experiment 2 (four phases), and
experiment 3 (four phases) were tested throughout all seven

19

TABLE 2

HIGHLIGHTS OF THREE EXPERIMENTS

| Experiment | Number of Samples | Number of Sampling Populations | Distance Between Points | Number of Distance Changes |
|---|---|---|---|---|
| 1 | 1 | 1 | Equal | 0 |
| 2$_a$ | 2 | 1 | Equal | 0 |
| 2$_b$ | 2 | 1 | Unequal | 1 |
| 2$_c$ | 2 | 1 | Unequal | 2 |
| 2$_d$ | 2 | 1 | Unequal & Unbalanced | > 2 |
| 3$_a$ | 2 | 2 | Equal | 0 |
| 3$_b$ | 2 | 2 | Unequal | 1 |
| 3$_c$ | 2 | 2 | Unequal | 2 |
| 3$_d$ | 2 | 2 | Unequal & Unbalanced | > 2 |

Notes: [1]Experiment 2 consists of four phases. Different distance changes are made between points in each phase.

[2]Experiment 3 also consists of four phases. Different distance changes are made between points in each phase.

populations (one uniform population, three normal
populations, and three gamma populations) using parametric
and nonparametric tests. Accordingly, the following
hypotheses were formulated with respect to the three
experiments and respective four phases in the research:

1. Hypothesis in experiment 1:

   Hypothesis 1:  Equally changing distances between
                  points on the scale do not affect
                  the parametric and nonparametric test
                  results if one sample is selected
                  from one population.

2. Hypotheses of the four phases in experiment 2:

   Hypothesis $2_a$: Equally changing distances between
                  points on the scale do not affect
                  the parametric and nonparametric test
                  results if two samples are selected
                  from one population.

   Hypothesis $2_b$: Unequally changing distances between
                  points on the scale (one unequal
                  change made on both sides of the
                  neutral point) affect the parametric
                  test results but do not affect the
                  nonparametric test results if two
                  samples are selected from one
                  population.

Hypothesis $2_c$: Unequally changing distances between
points on the scale (two unequal
changes made on both sides of the
neutral point) affect the parametric
test results but do not affect the
nonparametric test results if two
samples are selected from one
population.

Hypothesis $2_d$: Unequally changing distances between
points on the scale (more than two
changes made on both sides of the
neutral point) affect the parametric
test results but do not affect the
nonparametric test results if two
samples are selected from one
population.

3. Hypotheses of the four phases in experiment 3:

Hypothesis $3_a$: Equally changing distances between
points on the scale do not affect
the parametric and nonparametric test
results if two samples are selected
from two populations.

Hypothesis $3_b$: Unequally changing distances between
points on the scale (one unequal
change made on both sides of the
neutral point) affect the parametric

test results but do not affect the
nonparametric test results if two
samples are selected from two
populations.

Hypothesis $3_c$: Unequally changing distances between
points on the scale (two unequal
changes made on both sides of the
neutral point) affect the parametric
test results but do not affect the
nonparametric test results if two
samples are selected from two
populations.

Hypothesis $3_d$: Unequally changing distances between
points on the scale (more than two
changes made on both sides of the
neutral point) affect the parametric
test results but do not affect the
nonparametric test results if two
samples are selected from two
populations.

## Simulation Data and Mapping Functions

The seven test populations (one uniform population,
three normal populations, and three gamma populations) were
simulated by the subroutines developed by IMSL, Inc (IMSL
1991). Table 3 lists the parameters used by each of the
seven populations to generate 10,000 element values.

TABLE 3

POPULATION PARAMETERS FOR
THE GENERATION OF SIMULATION POPULATION

| Populations | Parameters |
| --- | --- |
| Uniform | Range = 0 to 100<br>$\mu = 50$ |
| Normal$_1$ | $\mu = 50$, $\sigma = 15$ |
| Normal$_2$ | $\mu = 40$, $\sigma = 15$ |
| Normal$_3$ | $\mu = 40$, $\sigma = 10$ |
| Gamma$_1$ | $\alpha = 11.1111$, $\beta = 4.5$<br>$\mu = 50$<br>$\sigma = 15$ |
| Gamma$_2$ | $\alpha = 7.1111$, $\beta = 5.625$<br>$\mu = 40$<br>$\sigma = 15$ |
| Gamma$_3$ | $\alpha = 16$, $\beta = 2.5$<br>$\mu = 40$<br>$\sigma = 10$ |

A frequency population of seven-point Likert-scale
was created by mapping each population element value into
one of the seven equal distance categories. Table 4
presents the mapping function. An interval including more
simulation data represented a higher frequency for the
class. The mapping procedure was applied by all seven test
populations (one uniform, three normal, and three gamma
populations). As an example, a simulation value of 78.13
from a population is mapped and produces a Likert-scale
score of 2, which is categorized as Agree. Therefore,
10,000 simulation values of each population were so mapped
and transformed into a Likert-scale equivalent population.
Then the distances between points were systematically
changed in the experiments, and selected parametric and
nonparametric tests were performed to evaluate the effect on
the distance changes on the scale.

### Experiment 1

Experiment 1 was a preliminary procedure to verify
the random number generation and related computer codes.
The target population observations were represented by
seven-point Likert-scale scores. One simple random sample
(Likert-scale scores) of size $n = 10$ was selected without
replacement from one population. No distance change between
points was made on the scale. Experiment 1 repeated
sampling one thousand times, with increasing sample sizes of
25, 50, and 100, to examine the effect on various sample

TABLE 4

MAPPING FUNCTION OF SEVEN CATEGORIES

| Label of Category | Class | Score of Likert-Scale |
|---|---|---|
| Strongly disagree | under 14.29 | -3 |
| Disagree | 14.29 and under 28.57 | -2 |
| Slightly disagree | 28.57 and under 42.86 | -1 |
| Neutral | 42.86 and under 57.14 | 0 |
| Slightly agree | 57.14 and under 71.43 | 1 |
| Agree | 71.43 and under 85.71 | 2 |
| Strongly agree | 85.71 and over | 3 |

Note: Using equal distances between labeled points.

sizes. This procedure was also applied to all seven test populations (one uniform, three normal, and three gamma populations).

Experiment 1 included four parametric and nonparametric tests: (1) confidence interval (or CI) for the population mean, (2) CI for the population median, (3) Kolmogorov-Smirnov (or KS) one-sample test, and (4) Chi-square one-sample test. The purpose was to determine whether the nominal 5% significance level (the Type I error) satisfactorily reflected the actual error rate when repeated sampling was performed. A Type I error rate of 5% was expected to be maintained for nearly all sample sizes obtained from symmetric populations (i.e., one uniform and three normal populations in experiment 1).

The anticipated 5% error rate was used to check whether simulation data were random and programs were accurate. If the error rate was substantially different from 5%, an unknown bias might exist in the simulation data, or logic errors might be found in the computer programs.

### Experiment 2

The goal of experiment 2 was to examine the effect on the Type I errors from parametric and nonparametric tests when the distances between points change using two samples selected from one population. The mapped populations of seven-point Likert-scale scores were the target sampling populations. One thousand pairs of equal-sized,

nonoverlapping samples were randomly and independently
selected without replacement from one population. The
experiment also repeated sampling, with increasing sample
sizes of 25, 50, and 100. This procedure was applied to all
seven test populations (one uniform, three normal, and three
gamma populations).

Experiment 2 consisted of four phases (phases $2_a$,
$2_b$, $2_c$, and $2_d$). Phases $2_b$, $2_c$, and $2_d$ were variations of
phase $2_a$ with alterations made in the distances of labeled
points. The Likert-scale scores were different among the
four phases. To avoid inconsistency, the scores were coded
by calculating the respective $Z$ values in each phase.
Therefore, the $Z$ scores were able to show relative distances
between points among different phases. Tables 5, 6, 7, and
8, respectively, present the four phases and required coding
procedure. The same sampling process was repeated and
tested for each of the four phases as follows:

## Phase $2_a$

Phase $2_a$ was a base model and was used as a paradigm
to compare with the results of phases $2_b$, $2_c$, and $2_d$. No
distance change between points was made on the scale.
Therefore, each of the seven mapped populations had its
respective frequency distribution of the seven-point Likert-
scale scores with equal distances between points. Table 5
presents the seven labels and the corresponding point
positions.

TABLE 5

LABELS VERSUS POINTS FOR EXPERIMENT $2_a$

| STD | D | SLD | N | SLA | A | STA |
|------|------|------|------|------|------|------|
| 1--------1----------1----------1----------1----------1----------1 | | | | | | |
| -3 | -2 | -1 | 0 | 1 | 2 | 3 |
| -1.5 | -1 | -0.5 | 0 | 0.5 | 1 | 1.5 |

Legend:  STD = Strongly disagree        STA = Strongly agree
         D   = Disagree                 A   = Agree
         SLD = Slightly disagree        SLA = Slightly agree
         N   = Neutral

Notes  [1]The first level of numeric values (-3 to 3)
represent the real positions of labeled points.

[2]The second level of numeric values represent the $Z$ scores
of labeled points,

$$Z = \frac{(x - \mu)}{\sigma}$$

where $\mu = 0$,   $\sigma = 2$.

[3]The distances between points have equal distances.

## Phase $2_b$

A numeric estimation scale with a balanced distortion in the middle area of the underlying continuum was examined in phase $2_b$. One distance change was made on both sides of the neutral position on the scale. Accordingly, the distances between Slightly disagree and Neutral as well as between Neutral and Slightly agree were subjectively extended five times longer than the distance of the same scale places in phase $2_a$. All other distances between points on the scale remained unchanged. Table 6 presents the seven labels and corresponding point positions (including $Z$ values).

## Phase $2_c$

Phase $2_c$ was to extend two scale ends of phase $2_a$. The basic assumption was that respondents tended to violate the rating pattern with a disproportionately wide range of estimates in the area of two extremes (Holdaway 1971; Galbraith and Schendel 1983; Parasuraman et al. 1988).

Two distance changes were made on both sides of the neutral position. Accordingly, the distances between Strongly disagree and Disagree as well as between Agree and Strongly agree were subjectively extended three times longer than the distances of the same scale places in phase $2_a$. The distances between Disagree and Slightly disagree as well as between Slightly agree and Agree were subjectively extended six times longer than the distances of the same

TABLE 6

LABELS VERSUS POINTS FOR EXPERIMENT $2_b$

| STD | D | SLD | N | SLA | A | STA |
|-----|---|-----|---|-----|---|-----|
| 1----1----1--------------------1--------------------1----1----1 |
| -7 | -6 | -5 | 0 | 5 | 6 | 7 |
| -1.249 | | -.892 | 0 | .892 | | 1.249 |
| | -1.07 | | | | 1.07 | |

Legend:  STD = Strongly disagree    STA = Strongly agree
         D   = Disagree             A   = Agree
         SLD = Slightly disagree    SLA = Slightly agree
         N   = Neutral

Notes: [1]The first level of numeric values (-7 to 7) represent the real positions of labeled points.

[2]The second level of numeric values represent the $Z$ scores of labeled points,

$$Z = \frac{(x - \mu)}{\sigma}$$

*where* $\mu = 0$, $\sigma = 5.606$.

[3]The distances between points have unequal lengths. One distance change between points is made at two sides of neutral. The lengths between Slightly disagree and Neutral as well as between Neutral and Slightly agree are extended five times longer than the original lengths. The others remain unchanged.

scale places in phase $2_a$. Table 7 presents the seven labels and corresponding point positions (including Z values). The other distances remained unchanged.

## Phase $2_d$

Bartram and Yielding (1973) indicated that people were more willing to grant positive values than negative values to categories and that, therefore, the scale was unbalanced between positive and negative sides. Worcester and Burns (1975) also found comparatively less strength on the negative end of the spectrum.

Therefore, phase $2_d$ was focused on forming a less-negative scale by making unequal and unbalanced changes on both sides of the neutral point. Distances between points on the positive side were extended more times than the distances between points on the negative side. The goal was to build an unbalanced scale on both sides of the neutral point. Table 8 presents the seven labels and corresponding point positions (including Z values).

## Experiment 3

The goal of experiment 3 was to select two samples from two populations and to examine the effect of the Type II error on parametric and nonparametric tests when the distances between points changed. Two types of errors may be made in decisions regarding the null hypothesis. Experiment 2 examined the Type I error using one population.

TABLE 7

LABELS VERSUS POINTS FOR EXPERIMENT $2_c$

| STD | D | | SLD | N | SLA | | A | STA |
|---|---|---|---|---|---|---|---|---|
| 1-------1----------------1---1---1----------------1-------1 |
| -10 | -7 | | -1 | 0 | 1 | | 7 | 10 |
| -1.527 | -1.069 | | -.153 | 0 | .153 | | 1.069 | 1.527 |

Legend:  STD = Strongly disagree      STA = Strongly agree
         D   = Disagree               A   = Agree
         SLD = Slightly disagree      SLA = Slightly agree
         N   = Neutral

Notes: [1]The first level of numeric values (-10 to 10)
represent the real positions of labeled points.

[2]The second level of numeric values represent the $Z$ scores
of labeled points,

$$Z = \frac{(x - \mu)}{\sigma}$$

where $\mu = 0$,  $\sigma = 6.547$.

[3]The distances between points have unequal lengths.  Two
changes between points are made at two sides of neutral.
The lengths between Strongly disagree and Disagree, between
Agree and Strongly agree are extended three times longer
than the originals, between Disagree and Slightly disagree,
and between Slightly agree and Agree are extended six times.
The others remain unchanged.

TABLE 8

LABELS VERSUS POINTS FOR EXPERIMENT $2_d$

| STD | D | SLD | N | SLA | A | STA |
|-----|---|-----|---|-----|---|-----|
| 1-------1---1---1----------1------------1-------------------1 | | | | | | |
| -4 | -2 | -1 | 0 | 4 | 10 | 20 |
| -.999 | -.618 | | | .018 | .781 | 2.053 |
| | -.745 | -.49 | | | | |

Legend:  STD = Strongly disagree        STA = Strongly agree
         D   = Disagree                A   = Agree
         SLD = Slightly disagree        SLA = Slightly agree
         N   = Neutral

Notes: [1]The first level of numeric values (-4 to 20) represent the real positions of labeled points.

[2]The second level of numeric values represent the $Z$ scores of labeled points,

$$Z = \frac{(x - \mu)}{\sigma}$$

where $\mu = 3.857$,  $\sigma = 7.864$.

[3]Multiple unequal and unbalanced changes between points are made at two sides of neutral. The length between Strongly disagree and Disagree is extended two times longer than the originals, between Neutral and Slightly agree four times, between Slightly agree and Agree six times, and between Agree and Strongly agree ten times.

In contrast, experiment 3 emphasized the importance of the
Type II error (Gregoire and Driver 1987), which was the
failure to reject the null hypothesis when it was false.
One test is more powerful and preferable than another when
the test has a lower Type II error rate (Kvanli, Guynes, and
Pavur 1992).

Experiment 3 also included four phases ($3_a$, $3_b$, $3_c$,
and $3_d$), which replicated each respective phase of
experiment 2 except that two samples were selected from two
different populations instead of one. Hence, there were 21
population combinations in experiment 3, as shown in table
9. The rationale was to investigate the robustness of the
four phases when differences in population characteristics
occurred. Experiment 3 looked at each population's changes
in location (mean and median); dispersion (standard
deviation); and distribution (frequency). As in experiment
2, one thousand pairs of equal-sized, nonoverlapping samples
(size $n$ = 10, 25, 50, and 100) were randomly and
independently selected, but from two different populations.
Again, the sampling procedure was applied to all seven test
populations (one uniform, three normal, and three gamma
populations).

## Statistical Testing Using Computer Programs

The data generation of uniform, normal, and gamma
populations, as well as the algorithm of selected parametric
and nonparametric tests were programmed in FORTRAN 77 (IMSL

TABLE 9

LIST OF POPULATION COMBINATIONS

| Uniform vs. Normal: | Uniform vs. Gamma: |
|---|---|
| 1 - 2 | 1 - 5 |
| 1 - 3 | 1 - 6 |
| 1 - 4 | 1 - 7 |

| Normal vs. Normal: | Gamma vs. Gamma: |
|---|---|
| 2 - 3 | 5 - 6 |
| 2 - 4 | 5 - 7 |
| 3 - 4 | 6 - 7 |

Normal vs. Gamma:

2 - 5
2 - 6
2 - 7
3 - 5
3 - 6
3 - 7
4 - 5
4 - 6
4 - 7

Note:  1 - Uniform population
       2 - $Normal_1$ population
       3 - $Normal_2$ population
       4 - $Normal_3$ population
       5 - $Gamma_1$ population
       6 - $Gamma_2$ population
       7 - $Gamma_3$ population

1991). In order to compare parametric tests and their nonparametric alternatives, the one sample selected from the populations in experiment 1 was used to perform all four parametric and nonparametric tests; the two samples selected from the populations in experiment 2 and experiment 3 were used to perform the other five parametric and nonparametric tests. The programs were designed using object-oriented modules (Barkakati 1990), and, thus, the programs could be reused by different populations, sample sizes, and phases with a change of parameters at the global level (seven populations) or respective classes (four phases and four sample sizes) on the basis of three objects (three experiments). Also, the validity of the tests and programs could be justified in experiment 1 and in phase 1 of experiments 2 and 3 because the results either were expected or could be verified by similar studies. The source codes of three computer programs respectively used by the three experiments are listed in the appendix.

CHAPTER 4

DATA ANALYSIS

## Parametric and Nonparametric Tests and Results in Experiment 1

Experiment 1 was used to verify random numbers and computer codes. The Type I error rates were accumulated to determine whether the nominal 5% was achieved by repeating sampling from seven simulated populations.

### Four Tests of Experiment 1

Four test statistics were selected to focus on the examination of population mean, median, standard deviation, and distribution (Plackett 1974; Nishisato 1980). The population mean $\mu$ of the CI test was based on the maximum likelihood estimator, the sample mean. The population standard deviation $\sigma$ of the chi-square test was based on the maximum likelihood estimator, the sample standard deviation (Hogg and Craig 1978; Thompson and Tapia 1990). The sample median was used as an estimator of the population median (Stockton 1962). Table 10 summarizes four parametric and nonparametric tests with the models of test statistics.

### Test Results of Experiment 1

In experiment 1, if one test failed to include the population mean or population median or rejected the null

Table 10

PARAMETRIC AND NONPARAMETRIC TESTS
IN EXPERIMENT 1

---

Confidence interval for $\mu$:

$$(\bar{x} - t_{\alpha/2}s_{\bar{x}}, \ \bar{x} + t_{\alpha/2}s_{\bar{x}})$$

---

Confidence interval for median[2]:

$$(median_{sample} + t_{\alpha/2}(1.2533)s_{\bar{x}}, \ median_{sample} - t_{\alpha/2}(1.2533)s_{\bar{x}})$$

---

Kolmogorov-Smirnov test on population distribution vs.
sample distribution:

$$D = \max |F_o(X_i) - S_N(X_i)|$$

where $F_0$ = population cumulative relative frequency
$S_N$ = sample cumulative relative frequency

---

Chi-square test on population standard deviation vs. sample
standard deviation:

$$\chi^2 = \frac{(n - 1)s^2}{\sigma^2}$$

---

Note: [1]$\alpha$ = 5%

[2]John R. Stockton. 1962. <u>Business statistics</u>. 2d ed.
(Cincinnati: South-Western Publishing) 249-50

hypothesis, 1 was added to the total error count. The
actual error rates, then, were computed by total error
count/1000. As such, the Type I error rates of parametric
and nonparametric tests in experiment 1 are summarized in
table 11.

## CI Test for Population Mean

As indicated in table 11, the Type I error rates of
the CI test for $\mu$ were nearly maintained at the nominal 5%
level for three normal populations. There was no major
difference between the large samples (n $\geq$ 30) and the small
samples (n < 30). Also, similar performance was exhibited
between the nonnormal populations and nonsymmetric
populations. The results had shed some light on the
selection of sample sizes and underlying populations in the
central limit theorem (Kvanli, Guynes, and Pavur 1992).
Thus, the expected error rates obtained in the repeating
tests revealed that the seven populations generated by
random numbers and associated computer codes were reliable
and could be reused for all other experiments in the
research.

## CI Test for Population Median

The purpose was to test the population median
location when the median was more demanding than the mean as
the primary statistic for describing the population. For
example, it was necessary when the underlying population was

TABLE 11

TYPE I ERROR RATES OF EXPERIMENT 1

| Population | n | $\mu$ | Median | KS | $\sigma$ |
|---|---|---|---|---|---|
| Uniform | 10 | 5.4 | 7.7 | 2.0 | 4.3 |
| | 25 | 4.8 | 16.2 | 1.1 | 20.1 |
| | 50 | 5.0 | 23.3 | 1.2 | 74.8 |
| | 100 | 5.6 | 15.0 | 1.1 | 99.7 |
| Normal$_1$ | 10 | 6.1 | 7.4 | 0.9 | 3.4 |
| | 25 | 4.7 | 5.5 | 0.8 | 5.4 |
| | 50 | 6.1 | 1.5 | 0.6 | 10.0 |
| | 100 | 4.8 | 0.1 | 0.5 | 16.9 |
| Normal$_2$ | 10 | 6.0 | 15.5 | 1.4 | 4.3 |
| | 25 | 5.0 | 22.6 | 0.8 | 4.9 |
| | 50 | 6.1 | 14.9 | 0.8 | 4.7 |
| | 100 | 5.0 | 7.1 | 0.9 | 4.8 |
| Normal$_3$ | 10 | 4.0 | 17.2 | 1.1 | 2.7 |
| | 25 | 4.6 | 12.3 | 0.6 | 4.2 |
| | 50 | 5.1 | 7.0 | 0.5 | 5.0 |
| | 100 | 4.8 | 1.8 | 0.5 | 5.3 |
| Gamma$_1$ | 10 | 5.5 | 8.9 | 1.4 | 4.6 |
| | 25 | 5.1 | 6.6 | 0.7 | 13.9 |
| | 50 | 5.0 | 1.5 | 0.6 | 25.2 |
| | 100 | 4.8 | 0.1 | 0.8 | 49.5 |
| Gamma$_2$ | 10 | 6.5 | 8.7 | 1.3 | 6.8 |
| | 25 | 5.4 | 10.5 | 0.4 | 9.5 |
| | 50 | 5.1 | 4.6 | 0.8 | 21.5 |
| | 100 | 5.1 | 1.0 | 0.9 | 31.2 |
| Gamma$_3$ | 10 | 4.2 | 12.7 | 0.4 | 3.4 |
| | 25 | 5.9 | 6.9 | 0.6 | 6.3 |
| | 50 | 6.1 | 3.0 | 0.5 | 7.5 |
| | 100 | 5.0 | 0.3 | 0.6 | 7.4 |

Note: [1]$\mu$: confidence interval test for the population mean

[2]Median: confidence interval test for the population median

[3]KS: Kolmogorov-Smirnov one-sample test for the population distribution

[4]$\sigma$: Chi-square test for the population standard deviation

skewed. The confidence interval for the median from a large

sample of normal population was a reasonably good

approximation for the parent population median (Stockton

1962). As expected, the Type I error rates of the $normal_1$

population retained around nominal 5%; the $normal_2$ and

$normal_3$ populations had higher error rates because the mean

was not equal to the median in the mapped populations. The

error rates, herein, had a 0%-18% range and decreased with

increasing sample size when sampling from three gamma

populations; the test provided the worst Type I error rates

and anomalous patterns when sampling from the uniform

population. Generally, the error rates reached the nominal

5% level or below in many instances with the sample size

increasing to 50 or higher. It was also essential to note

that the error rates resulting from gamma populations were

slightly lower than those from normal populations when both

the gamma and normal had identical means and variances in

the tests.

## KS One-Sample Test

The test results of the KS test (table 11) reported

the Type I error rates in the low 0%-2% range. It did not

matter whether the variable under consideration had a

continuous distribution (Siegel and Castellan 1988), whether

the distribution was normal (Puri and Sen 1985), or whether

the population was symmetric (Gregoire and Driver 1987).

Similar results were obtained by Gregoire and Driver (1987).

The KS test was the only nonparametric test in experiment 1, but it yielded lower Type I error rates than any other parametric test. The test was generally powerful in all sample sizes and populations of experiment 1. Hence, the KS test is able to provide extremely good protection against distribution heterogeneity.

## Chi-Square Test for Population Standard Deviation

The variation of populations was the primary concern in the chi-square test. The Type I error rates were almost retained at the nominal 5% when sampling from the normal populations and appeared expectable. The chi-square test was dramatically erratic when sampling from the nonnormal populations, uniform and gamma. Apparently, the chi-square test on $\sigma$ is highly sensitive to the assumption of normal populations.

It is of interest to note that the Type I error rates increased as did the sample size here. Normally, the opposite phenomenon occurs in sampling statistics because a larger sample could bring more adequate information and yield lower error rates than a smaller sample. As sampling observations increase so too will one's confidence in the population (Scheaffer, Mendenhall, and Ott 1990). However, this was not the case in experiment 1. These ambiguous Type I error patterns could be caused by the fact that the sample standard deviation was not an unbiased estimator of the population standard deviation, because the sample values

were selected from the seven populations with seven-point
Likert-scale values that were mapped by a mapping function
(table 4). Thus, the variation of the estimator decreased
when the sample size increased, the bias did not (Hogg and
Tanis 1977; Carr, Hafner, and Koch 1989).

### Parametric and Nonparametric Tests and Results in Experiment 2

Experiment 2 examined the tests based on the
equality of four parameters: the mean, median, dispersion,
and distribution. The four statistics were believed to have
conspicuous influences on the selection of statistical
tests. The Type I error rates were evaluated in experiment
2 (whereas the Type II error rates were evaluated in
experiment 3).

#### Equality of Means

The purpose was to examine the central location of
the population. All observed 95% confidence intervals

$$\left(\overline{x_1} - \overline{x_2}\right) \pm t\left(s_{\overline{x_1} - \overline{x_2}}\right)$$

were constructed from the samples and used to determine the
number of times (or percentage of time) the resulting
intervals failed to cover zero in repeated sampling of the
experiments.

#### Equality of Medians

Two populations may differ in central tendency. The
research used the Mann-Whitney/Wilcoxon (or MWW) test to

examine whether one sample tended to yield a larger median
than another sample. As the sample size increased, the rank
distribution of two samples rapidly approached a normal
distribution (Siegel and Castellan 1988). Therefore, in
experiment 2, the MWW test determined the mean and variance
by

$$\mu_{W_x} = \frac{n_1(N + 1)}{2}$$

and

$$\sigma^2_{W_x} = \frac{n_1 n_2}{N(N - 1)}\left(\frac{N^3 - N}{12} - \sum_{j=1}^{g} \frac{t_j^3 - t_j}{12}\right)$$

where $n_1$ = size of sample 1

$n_2$ = size of sample 2

$N = n_1 + n_2$

$i = 1, 2$

$W_x$ = ranks of sample 1

$g$ = number of tied groups

$t$ = number of tied ranks in the

$j$th grouping.

Although the MWW test assumes that the data are sampled from
a continuous distribution, the seven-point Likert-scale
values were used by a mapping function in the experiments.
Hence, the tied scores occurred frequently throughout the
tests. If several observed values were tied at the same
rank, the rank was assigned the average of the those tied

ranks. Then, the sum of the ranks (sample 1) was changed to a $Z$ value by

$$Z = \frac{W_x \pm 0.5 - \dfrac{n_1(N + 1)}{2}}{\sqrt{\left[\dfrac{n_1 n_2}{N(N - 1)}\right]\left[\dfrac{(N^3 - N)}{12} - \sum_{j=1}^{g} \dfrac{(t_j^3 - t_j)}{12}\right]}}$$

which was normally distributed with zero mean and unit variance. The $Z$ value was used to determine the significance of the Type I errors.

## Equality of Dispersions

The measure of variance was used to detect the variation of populations. As pointed out by Siegel and Tukey (1960), the $F$ test was very sensitive to the assumption of normality. The research used $s_1^2/s_2^2$ to test whether or not the variances of two samples followed the same $F$ distribution.

In addition, Duran (1976) used the value of absolute deviation from means to test whether two random variables from two samples had identical deviations. Thus, the nonparametric Squared Ranks (or SR) test (revised by Conover and Iman 1978) was used to examine the equality of dispersions with the following test statistics:

$$T_1 = \frac{T - n_1\overline{R^2}}{\left[\frac{n_1 n_2}{N(N-1)} \sum_{i=1}^{N} R^4 - \frac{n_1 n_2}{N-1} (\overline{R^2})^2\right]^{\frac{1}{2}}}$$

where $n_1$ = size of sample 1

$n_2$ = size of sample 2

$N$ = $n_1 + n_2$

$U_i = |X_i - \mu_1|$

$V_j = |Y_j - \mu_2|$

$R$ = ranks

$$T = \sum_{i=1}^{n} [R(U_i)]^2$$

$$\overline{R^2} = \frac{1}{N}\left[\sum_{i=1}^{n} [R(U_i)]^2 + \sum_{i=1}^{m} [R(V_i)]^2\right]$$

$$\sum_{i=1}^{N} R^4 = \sum_{i=1}^{n} [R(U_i)]^4 + \sum_{i=1}^{m} [R(V_i)]^4$$

The critical value of the decision to call for rejection of the null hypothesis was determined by the $Z$ value of normal distribution.

## Equality of Distributions

Sampling data invariably differed somewhat in distribution types. The identity of two distribution types was assessed by the KS test in the experiment. The KS test

was used to specify the cumulative frequency distribution to test samples coming either from the same population or from two identical populations (Siegel and Castellan 1988). The test also used more information contained in the observations than did the median test (Conover 1980).

For a sample, the value of $D_{m,n}$ in the KS two-sample test was

$$D_{m,n} = \max[S_m(X) - S_n(X)]$$

where m = size of sample 1

n = size of sample 2

X = observation

$S_m(X)$ = K/m

K = the number of data $\leq$ X in the

sample.

Here, the critical value of the decision to call for rejection of the null hypothesis was determined by

$$Critical\ Value\ T_{\alpha=.05} = 1.36\ \sqrt{\frac{m + n}{mn}}$$

## Five Tests of Experiment 2

In summary, the research was focused on the examination of the equality of means (the CI test/*t* test); the equality of medians (the MWW test); the equality of dispersions (the *F* test and the SR test); and the homogeneity of two population distributions (the KS test).

In total, two parametric and three nonparametric tests (as
listed in table 12) were used to determine whether to reject
the null hypotheses in experiment 2.


TABLE 12


PARAMETRIC AND NONPARAMETRIC TESTS
IN EXPERIMENT 2 AND EXPERIMENT 3

| Parametric Tests | Nonparametric Tests |
| --- | --- |
| Confidence interval: 5% Type I error rate | Mann-Whitney/Wilcoxon test |
| $F$ test | Squared Ranks test |
| | Kolmogorov-Smirnov test |


Test Results of Experiment 2

In experiment 2, if one test failed to include zero
or rejected the null hypothesis that the respective
population parameters (population means, medians, standard
deviations, or distributions) were identical, 1 was added to
the total error count.  The actual error rates were computed
by total error count/1000.  Moreover, different distance
changes between points were applied in the four phases.  The
results of the Type I error rates of parametric and
nonparametric tests in experiment 2 are exhibited in table
13.

TABLE 13

TYPE I ERROR RATES OF EXPERIMENT 2

| Population | Phase | n | $\mu$ | MWW | F | SR | KS |
|---|---|---|---|---|---|---|---|
| Uniform | 1 | 10 | 5.3 | 4.9 | 1.0 | 4.7 | 0.5 |
| | | 25 | 4.7 | 4.8 | 0.4 | 4.8 | 0.9 |
| | | 50 | 5.3 | 4.8 | 0.1 | 5.3 | 1.0 |
| | | 100 | 4.5 | 4.4 | 0.0 | 5.0 | 0.8 |
| | 2 | 10 | 6.0 | 4.9 | 0.8 | 6.1 | 0.5 |
| | | 25 | 5.0 | 5.1 | 0.0 | 4.0 | 0.8 |
| | | 50 | 5.5 | 4.7 | 0.0 | 4.8 | 1.5 |
| | | 100 | 5.0 | 4.9 | 0.0 | 3.7 | 1.4 |
| | 3 | 10 | 5.1 | 5.0 | 1.5 | 4.9 | 0.6 |
| | | 25 | 5.3 | 5.1 | 0.8 | 5.3 | 1.1 |
| | | 50 | 5.4 | 4.8 | 0.4 | 5.4 | 1.8 |
| | | 100 | 5.5 | 5.2 | 0.2 | 5.5 | 0.4 |
| | 4 | 10 | 5.3 | 5.4 | 11.5 | 17.9 | 0.4 |
| | | 25 | 5.4 | 5.3 | 5.8 | 21.1 | 1.1 |
| | | 50 | 5.0 | 5.1 | 4.5 | 22.5 | 1.5 |
| | | 100 | 6.1 | 5.0 | 4.0 | 20.3 | 1.2 |
| Normal$_1$ | 1 | 10 | 4.9 | 4.7 | 4.8 | 5.1 | 0.4 |
| | | 25 | 4.9 | 4.9 | 4.5 | 6.3 | 0.3 |
| | | 50 | 4.8 | 4.4 | 4.7 | 6.7 | 0.7 |
| | | 100 | 5.4 | 4.9 | 2.9 | 9.6 | 0.7 |
| | 2 | 10 | 4.6 | 4.2 | 0.9 | 6.4 | 0.4 |
| | | 25 | 5.1 | 5.3 | 0.4 | 5.5 | 0.7 |
| | | 50 | 4.8 | 4.6 | 0.1 | 4.8 | 0.6 |
| | | 100 | 5.0 | 5.1 | 0.1 | 6.5 | 0.4 |
| | 3 | 10 | 3.5 | 3.7 | 33.5 | 17.7 | 0.1 |
| | | 25 | 4.8 | 5.1 | 25.5 | 14.2 | 0.4 |
| | | 50 | 5.4 | 5.6 | 21.2 | 14.2 | 0.5 |
| | | 100 | 4.9 | 5.5 | 22.6 | 16.5 | 0.7 |
| | 4 | 10 | 4.0 | 4.9 | 23.2 | 21.5 | 0.7 |
| | | 25 | 4.1 | 4.6 | 30.0 | 24.3 | 0.3 |
| | | 50 | 5.5 | 5.2 | 31.7 | 22.8 | 0.9 |
| | | 100 | 5.8 | 5.7 | 31.3 | 18.0 | 0.8 |

Table 13--<u>Continued</u>

| Population | Phase | n | $\mu$ | MWW | F | SR | KS |
|---|---|---|---|---|---|---|---|
| Normal$_2$ | 1 | 10 | 4.7 | 4.0 | 4.9 | 6.2 | 0.3 |
| | | 25 | 5.0 | 5.2 | 4.2 | 5.9 | 0.8 |
| | | 50 | 4.8 | 4.3 | 3.0 | 5.7 | 0.4 |
| | | 100 | 6.1 | 6.3 | 3.4 | 7.0 | 1.1 |
| | 2 | 10 | 5.9 | 5.6 | 4.8 | 14.7 | 0.2 |
| | | 25 | 5.4 | 5.4 | 3.5 | 17.4 | 0.5 |
| | | 50 | 5.2 | 5.1 | 2.5 | 20.3 | 0.8 |
| | | 100 | 5.4 | 4.9 | 1.8 | 21.7 | 0.9 |
| | 3 | 10 | 5.4 | 5.1 | 14.2 | 24.3 | 0.1 |
| | | 25 | 4.9 | 4.9 | 9.8 | 30.4 | 0.2 |
| | | 50 | 5.6 | 5.0 | 9.7 | 28.5 | 0.7 |
| | | 100 | 5.4 | 4.9 | 7.7 | 32.0 | 0.7 |
| | 4 | 10 | 4.0 | 5.2 | 37.0 | 20.1 | 0.3 |
| | | 25 | 5.6 | 5.7 | 38.2 | 22.2 | 0.5 |
| | | 50 | 4.8 | 4.8 | 38.0 | 24.9 | 0.9 |
| | | 100 | 4.5 | 4.7 | 38.3 | 36.6 | 0.5 |
| Normal$_3$ | 1 | 10 | 5.9 | 6.0 | 5.2 | 7.3 | 0.4 |
| | | 25 | 5.4 | 5.1 | 5.2 | 8.1 | 0.4 |
| | | 50 | 4.8 | 4.3 | 5.0 | 9.0 | 0.5 |
| | | 100 | 5.6 | 5.3 | 4.8 | 8.6 | 0.5 |
| | 2 | 10 | 5.1 | 4.8 | 3.7 | 12.2 | 0.2 |
| | | 25 | 5.4 | 5.1 | 4.0 | 14.9 | 0.4 |
| | | 50 | 5.2 | 4.7 | 3.7 | 11.3 | 0.4 |
| | | 100 | 5.4 | 5.6 | 3.5 | 7.6 | 0.4 |
| | 3 | 10 | 3.8 | 5.3 | 38.3 | 33.3 | 0.3 |
| | | 25 | 4.4 | 5.2 | 24.3 | 41.3 | 0.4 |
| | | 50 | 6.0 | 5.5 | 22.7 | 68.7 | 0.4 |
| | | 100 | 5.1 | 5.5 | 18.4 | 91.7 | 0.5 |
| | 4 | 10 | 3.4 | 4.9 | 42.4 | 15.9 | 0.3 |
| | | 25 | 4.2 | 4.6 | 47.3 | 13.6 | 0.3 |
| | | 50 | 4.2 | 5.2 | 43.0 | 19.7 | 0.6 |
| | | 100 | 5.5 | 5.6 | 42.5 | 20.6 | 0.6 |

Table 13--<u>Continued</u>

| Population | Phase | n | $\mu$ | MWW | F | SR | KS |
|---|---|---|---|---|---|---|---|
| Gamma$_1$ | 1 | 10 | 4.4 | 4.2 | 4.4 | 6.3 | 0.7 |
| | | 25 | 6.1 | 6.2 | 5.6 | 7.9 | 0.5 |
| | | 50 | 5.8 | 5.5 | 5.0 | 10.1 | 0.7 |
| | | 100 | 4.4 | 4.3 | 4.9 | 16.1 | 0.5 |
| | 2 | 10 | 5.0 | 4.6 | 1.2 | 6.0 | 0.2 |
| | | 25 | 4.7 | 4.6 | 0.6 | 6.2 | 0.5 |
| | | 50 | 5.6 | 5.4 | 0.1 | 6.8 | 0.7 |
| | | 100 | 5.5 | 5.5 | 0.1 | 7.6 | 0.8 |
| | 3 | 10 | 3.3 | 5.3 | 36.1 | 17.9 | 0.4 |
| | | 25 | 5.0 | 5.0 | 28.9 | 21.2 | 0.6 |
| | | 50 | 6.3 | 5.6 | 27.1 | 25.2 | 0.8 |
| | | 100 | 5.0 | 4.8 | 24.8 | 35.2 | 0.5 |
| | 4 | 10 | 4.0 | 4.6 | 34.4 | 30.2 | 0.1 |
| | | 25 | 5.1 | 5.0 | 40.9 | 38.2 | 0.4 |
| | | 50 | 4.6 | 4.6 | 36.7 | 36.4 | 0.4 |
| | | 100 | 4.4 | 4.5 | 35.2 | 39.7 | 0.5 |
| Gamma$_2$ | 1 | 10 | 3.8 | 4.0 | 7.6 | 7.6 | 0.1 |
| | | 25 | 4.5 | 4.8 | 7.3 | 8.7 | 0.5 |
| | | 50 | 4.6 | 4.1 | 6.4 | 11.0 | 0.8 |
| | | 100 | 4.2 | 3.9 | 6.4 | 15.2 | 0.4 |
| | 2 | 10 | 4.7 | 4.2 | 8.2 | 19.2 | 0.2 |
| | | 25 | 5.7 | 5.8 | 5.0 | 28.1 | 0.5 |
| | | 50 | 4.9 | 4.6 | 4.7 | 28.9 | 0.7 |
| | | 100 | 5.0 | 4.9 | 3.4 | 33.2 | 0.6 |
| | 3 | 10 | 6.0 | 5.4 | 12.4 | 16.3 | 0.3 |
| | | 25 | 4.8 | 4.6 | 9.6 | 21.7 | 0.4 |
| | | 50 | 4.7 | 4.9 | 9.5 | 18.6 | 0.9 |
| | | 100 | 5.1 | 5.1 | 8.8 | 20.5 | 0.3 |
| | 4 | 10 | 3.3 | 4.9 | 51.8 | 32.0 | 0.4 |
| | | 25 | 3.9 | 4.9 | 54.6 | 35.5 | 0.5 |
| | | 50 | 4.7 | 5.4 | 56.2 | 47.9 | 0.8 |
| | | 100 | 4.8 | 5.0 | 55.8 | 70.1 | 0.6 |

Table 13--_Continued_

| Population | Phase | n | $\mu$ | MWW | F | SR | KS |
|-----------|-------|-----|-----|-----|------|------|-----|
| Gamma$_3$ | 1 | 10 | 5.5 | 4.9 | 5.9 | 8.1 | 0.2 |
| | | 25 | 5.0 | 5.3 | 7.8 | 11.3 | 0.3 |
| | | 50 | 5.7 | 5.4 | 5.4 | 11.4 | 0.3 |
| | | 100 | 5.7 | 5.5 | 5.2 | 12.8 | 0.3 |
| | 2 | 10 | 5.5 | 5.3 | 7.8 | 18.1 | 0.3 |
| | | 25 | 5.7 | 6.1 | 7.4 | 20.2 | 0.4 |
| | | 50 | 5.3 | 5.2 | 5.5 | 21.1 | 0.8 |
| | | 100 | 4.3 | 4.4 | 5.1 | 14.4 | 0.5 |
| | 3 | 10 | 4.2 | 5.5 | 38.5 | 30.4 | 0.1 |
| | | 25 | 4.5 | 5.0 | 25.9 | 38.1 | 0.4 |
| | | 50 | 5.4 | 4.7 | 23.6 | 61.5 | 0.4 |
| | | 100 | 5.6 | 4.8 | 21.8 | 86.7 | 0.6 |
| | 4 | 10 | 2.7 | 4.5 | 48.6 | 20.2 | 0.2 |
| | | 25 | 3.5 | 4.7 | 51.1 | 27.9 | 0.4 |
| | | 50 | 5.6 | 5.5 | 51.5 | 37.5 | 0.4 |
| | | 100 | 4.2 | 4.5 | 52.6 | 56.7 | 0.2 |

Note: [1]$\mu$: confidence interval test for the difference of population means

[2]MWW: Mann-Whitney/Wilcoxon two-sample test for medians

[3]F: F test for variances

[4]SR: Squared Ranks two-sample test for variances

[5]KS: Kolmogorov-Smirnov two-sample test for distributions

[6]$\alpha = 5\%$

[7]All entry values are the percentages of the Type I error rates

CI Test for the Difference of Population Means

The Type I error rates generally retained to attest
the nominal 5% level, regardless of sample size in
experiment 2. The CI test, herein, was not sensitive to the
assumption of normal distribution. No particularly
different error rates or error patterns could be found when
underlying populations were not normal. However, it cannot
be concluded that the CI test does not need an assumption of
normality in the tests; a coarser nonsymmetric population
may show serious impact on the CI test results (Burdick and
Graybill 1992). Within each population, a few unequal
distance changes between points were made in phases 2, 3,
and 4. The Type I error rates revealed little differences
when comparing phases 2, 3, and 4 against phase 1 where no
distance change was made between points on the scale.
Surprisingly, the parametric CI test performed awkwardly in
detecting the unequal lengths between points in experiment
2. Inevitably, some instances were liberal (the actual
error rates were a little higher than the nominal 5%), and
some were conservative (the actual error rates were a little
lower than the nominal 5%). Similar results were obtained
by Baker et al. (1966) and Gregoire and Driver (1987).

MWW Test for the Difference of Population Medians

As expected, no remarkable differences were shown
among the four phases, because the MWW test was a
nonparametric test. The MWW test performed as well as the

CI test in experiment 2. Hence, when samples are selected
from symmetric populations (where the population mean equals
the population median), there is no noticeable gain or loss
by using one of these two tests over the other. Of course,
the mean and median are two different location parameters;
one may make the choice based on the test locations when
underlying populations are nonsymmetric.

When compared with the more powerful $t$ test, the
power-efficiency of the MWW test approaches $3/\pi = 95.4927\%$
as the sample size increases (Siegel and Castellan 1988).
The Type I error rates showed that the nominal 5% was nearly
maintained in the MWW test under each sample size, phase,
and population. Therefore, when the assumption of normal
distribution is unwarranted, the MWW test should be used.
In other words, the MWW test is a very appropriate
alternative to the parametric CI test (or $t$ test) when one
tries to avoid the CI test's assumptions. The sample size
has no major effect on the performance of the MWW test.

The values of populations in experiment 2 were
mapped and transformed into seven points (seven-point
Likert-scale), and populations were consequently represented
by the frequencies of various intervals. Thus, there were
many ties in the samples during the repeating tests.
However, the consistently low error rates in experiment 2
indicated that the use of the MWW test was also quite
appropriate when the populations were not continuous and
when many ties were found in the samples (Iman and Conover

1983; Rao 1983).

## F Test for the Difference of Population Variances

The parametric $F$ test presented results with variety. Many studies have indicated that the $F$ test is highly sensitive to the assumption of normal population (Jarjoura 1983; Kvanli, Guynes, and Pavur 1992). This was confirmed in experiment 2. The $F$ test performed well and appeared expectable in phase 1 of three normal populations (as shown in figure 1) where no distance change was made between points on the scale (the frequencies of each population resembled a distribution of normal shape). The Type I error rates gradually decreased in the 5.2%-2.9% range with increasing sample size.

A few distance changes between points were made in phases 2, 3, and 4 of three normal populations so that the shapes of frequencies reflected different degrees of deviation from a normal distribution. The Type I error of phases 2, 3, and 4 of three normal populations showed either much lower or higher rates than in phase 1. Increasing the sample size did not decrease the Type I error rates.

In addition, the other nonnormal populations (three gamma populations) also reported anomalous Type I error rates; but the performance was almost perfect for the uniform population. Increasing the sample size might either deteriorate or ameliorate the Type I error rates under different circumstances (sample sizes or phases). Phase 4
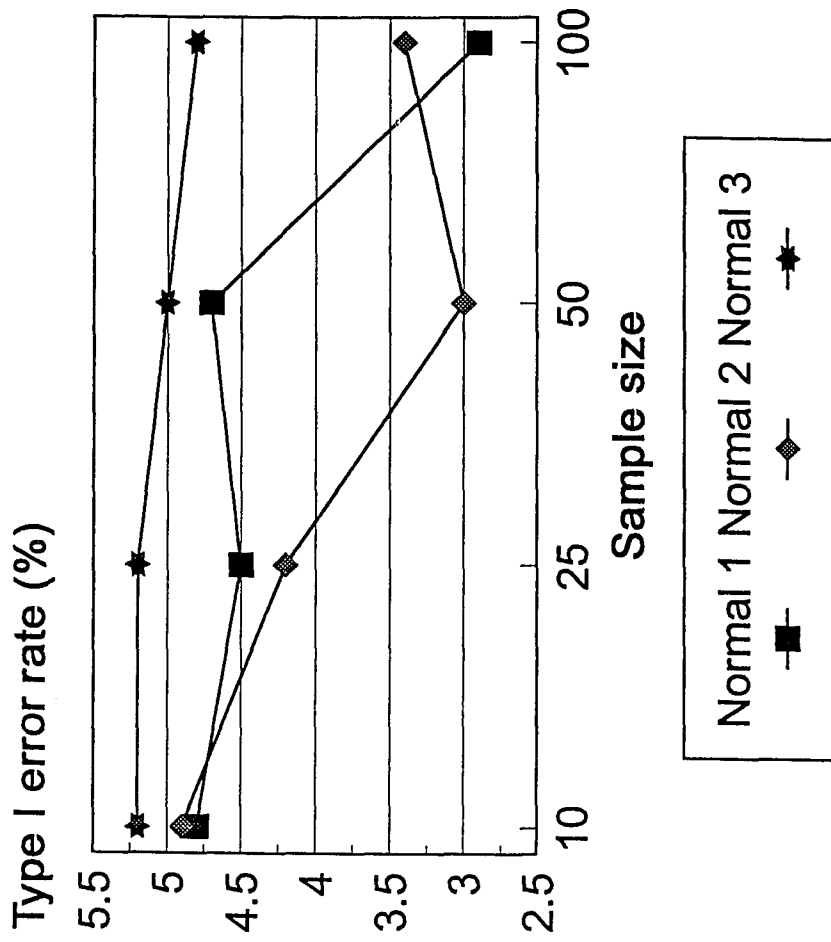
Fig. 1. ERROR PATTERNS OF F TEST IN PHASE 1 OF THREE POPULATIONS

(that had unequal and unbalanced distance changes between
points on the scale) generally showed the worst error rates
among the four phases.

The performance of the $F$ test was extremely good in
phase 2 of the $gamma_1$ population as shown in figure 2.
Apparently, the situation was a haphazard snapshot in one
phase of the experiment. It was impossible to benefit from
the good performance of coincidence based on one case
without considering other knowledge. Some nearby evidence
could reveal the truth, e.g., the results in phase 3, phase
4, and phase 1 of $gamma_1$ population (also shown in figure
2). The Type I error rates were much higher in phase 3 of
the same population, $gamma_1$. Here, the true population
resembled the skewed distribution and was somewhat
nonsymmetric, yet the error rate performance was poorly
displayed. Hence, the $F$ test is not strongly recommended
to use unless the normal population is justified.

## SR Test for the Difference of Population Variances

The Type I error rates uniformly retained at the 4%-
9% range in phase 1 of uniform and three normal populations.
The SR test performed erratically in most of other cases
without consistent patterns.

Unequally changing distances between points on the
scale have no effect on the nonparametric test. So the
inconsistent error performance among the four phases did not
appear expectable in the nonparametric SR test of experiment
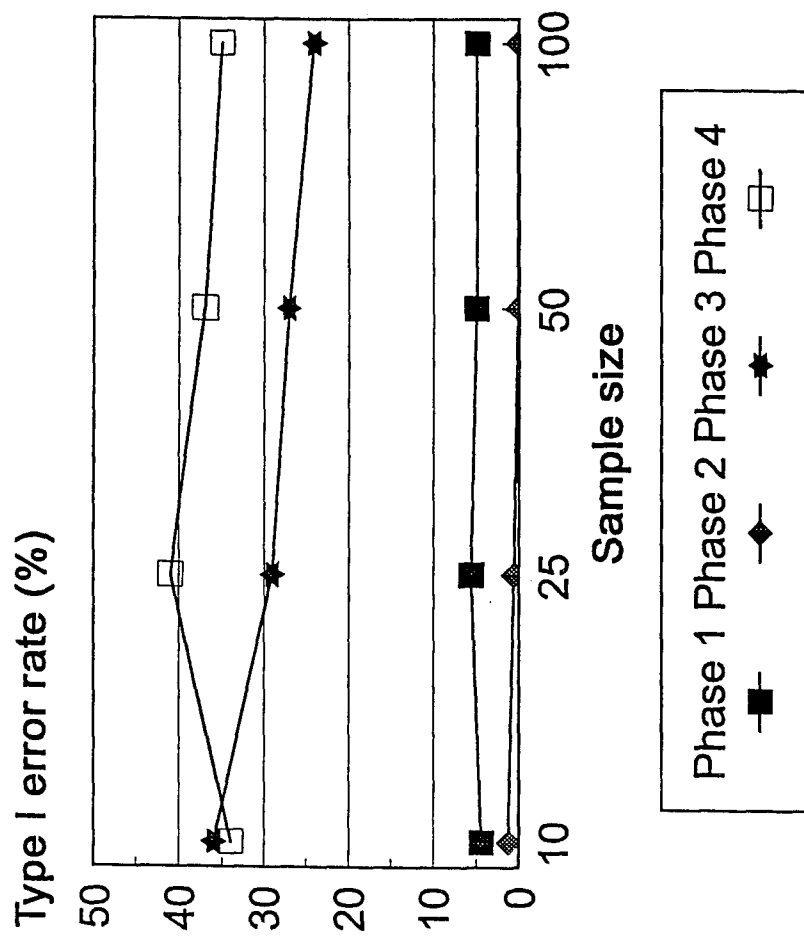
Fig. 2. ERROR PATTERNS OF F TEST IN GAMMA 1 POPULATION

2. The disturbing outcome displayed here was because the SR test ranked each sample observation by its absolute deviation from the sample mean. Unequally changing distances between points on the scale would rearrange the rank order in phases 2, 3, and 4, because the sample means were dramatically shifted along the underlying continuum. Therefore, the distance changes had an effect on the SR test.

The SR test was included as an alternative nonparametric test of the $F$ test in order to study their usefulness and power. The Type I error rates indicated that, with very few exceptions, the SR test performed consistently worse than the $F$ test across populations in experiment 2. In summary, the SR test should be used when one is not sure whether the population is normal. If the SR test is used instead of the $F$ test when the population is normal, the asymptotic relative efficiency (or A.R.E.) is as low as $15/(2\pi^2) = 75.99\%$ (Conover 1980). However, the sensitivity of the $F$ test to the assumption of normality, coupled with its lack of power in some common nonnormal populations, indicates that the analogous nonparametric test, the SR test, is a good second choice.

KS Test for the Difference of Population Distributions

As indicated in table 13, the KS test performed extremely well in all cases of experiment 2. The Type I error rates uniformly reported in the 0%-1.5% range. As the

sample size increased, the Type I error rates slightly decreased in many instances. Apparently, the KS test could easily distinguish distribution equality as long as the populations had identical means and variances; it did not matter whether the distributions were normal or nonnormal, symmetric or nonsymmetric in experiment 2.

The $t$ test and the MWW test may also be appropriate in determining whether the two population distributions are identical or not, but they are sensitive to differences between the two means and medians, respectively; they may not be able to detect the differences of other location parameters, such as variances. The KS test appeared to be dependable on examining all types of location disparity (differences in central tendency, dispersion, skewness, etc.).

The KS test has high power-efficiency (about 95%) for small samples when compared with the $t$ test (Thompson and Tapia 1990). Some evidence has shown that, for small samples, the KS test is slightly more efficient than the MWW test, whereas for large samples the converse holds (Siegel and Castellan 1988). However, as shown in table 13, the Type I error of the KS test presented lower rates in either small or large samples when compared with the MWW test or even the CI test. Hence, the KS test could provide very reliable protection against false claims of distribution heterogeneity (Puri and Sen 1985).

## Parametric and Nonparametric Tests and Results
## in Experiment 3

### Five Tests of Experiment 3

The Type II error was the main object of concern in
experiment 3. The same five tests outlined in experiment 2
were used once more in experiment 3 (table 12) except that
two samples were selected from two different populations.
Inasmuch as several populations had identical location
parameters (population means, medians, and standard
deviations), the Type I error was evaluated instead.

### Test Results of Experiment 3

In experiment 3, if one test included zero or failed
to reject the null hypothesis when the respective location
parameters (population means, medians, standard deviations,
and distributions) were not identical, 1 was added to the
total error count. The actual error rates were computed by
total error count/1000. As such, the Type II error rates of
parametric and nonparametric tests in experiment 3 are
reported in table 14.

### CI Test for the Difference of Two Population Means

The Type II error rates were considerably high in
experiment 3. The CI test performed erratically in every
population combination (sampling from two populations).
However, the lower Type II error rates when sampling from
two normal populations with heterogeneous means offered
strong support to the assumption of normality. The Type II

TABLE 14

TYPE II ERROR RATES OF EXPERIMENT 3

| Population | Phase | n | $\mu$ | MWW | F | SR | KS |
|---|---|---|---|---|---|---|---|
| Uniform & | 1 | 10 | [5.2] | [5.5] | 59.8 | 55.1 | 98.8 |
| Normal$_1$ | | 25 | [5.4] | [6.6] | 10.2 | 7.9 | 89.2 |
| | | 50 | [5.6] | [6.4] | 0.3 | 0.2 | 59.8 |
| | | 100 | [4.9] | [5.9] | 0.0 | 0.0 | 9.6 |
| | 2 | 10 | [4.4] | [5.7] | 95.3 | 70.8 | 98.7 |
| | | 25 | [5.6] | [6.9] | 88.0 | 27.5 | 89.5 |
| | | 50 | [5.2] | [6.7] | 50.7 | 4.0 | 56.1 |
| | | 100 | [5.3] | [6.8] | 5.3 | 0.2 | 10.3 |
| | 3 | 10 | [4.9] | [6.1] | 39.2 | 43.6 | 98.7 |
| | | 25 | [6.5] | [6.9] | 4.8 | 7.0 | 89.9 |
| | | 50 | [5.9] | [6.8] | 0.0 | 0.2 | 56.3 |
| | | 100 | [5.1] | [6.4] | 0.0 | 0.0 | 9.8 |
| | 4 | 10 | [9.9] | [7.0] | 33.7 | 40.1 | 97.7 |
| | | 25 | [25.1] | [6.1] | 7.9 | 6.5 | 89.4 |
| | | 50 | [53.6] | [7.1] | 0.8 | 0.3 | 57.8 |
| | | 100 | [83.9] | [5.7] | 0.0 | 0.0 | 10.0 |
| Uniform & | 1 | 10 | 85.2 | 87.3 | 57.8 | 52.1 | 96.6 |
| Normal$_2$ | | 25 | 68.4 | 75.2 | 8.8 | 6.9 | 73.4 |
| | | 50 | 42.7 | 56.8 | 0.2 | 0.2 | 26.4 |
| | | 100 | 12.0 | 29.9 | 0.0 | 0.0 | 1.4 |
| | 2 | 10 | 80.1 | 86.7 | 83.2 | 58.4 | 95.9 |
| | | 25 | 54.5 | 74.5 | 54.5 | 10.3 | 73.3 |
| | | 50 | 26.6 | 57.5 | 13.7 | 0.5 | 27.1 |
| | | 100 | 5.6 | 32.2 | 0.4 | 0.0 | 2.3 |
| | 3 | 10 | 90.1 | 88.1 | 52.8 | 56.0 | 96.7 |
| | | 25 | 79.3 | 75.6 | 6.8 | 24.2 | 75.5 |
| | | 50 | 61.3 | 59.1 | 0.3 | 6.5 | 28.3 |
| | | 100 | 28.0 | 27.4 | 0.0 | 0.6 | 1.0 |
| | 4 | 10 | 73.3 | 86.9 | 12.6 | 17.4 | 96.2 |
| | | 25 | 27.1 | 73.1 | 1.1 | 0.5 | 73.0 |
| | | 50 | 3.7 | 58.7 | 0.1 | 0.0 | 27.9 |
| | | 100 | 0.0 | 28.4 | 0.0 | 0.0 | 0.9 |

Table 14--<u>Continued</u>

| Population | Phase | n | $\mu$ | MWW | F | SR | KS |
|---|---|---|---|---|---|---|---|
| Uniform & | 1 | 10 | 83.5 | 85.6 | 14.7 | 22.3 | 93.5 |
| Normal$_3$ |  | 25 | 63.4 | 72.0 | 0.1 | 0.6 | 50.9 |
|  |  | 50 | 34.6 | 52.4 | 0.0 | 0.0 | 3.6 |
|  |  | 100 | 8.7 | 27.9 | 0.0 | 0.0 | 0.0 |
|  | 2 | 10 | 72.6 | 84.4 | 5.3 | 29.6 | 92.0 |
|  |  | 25 | 39.2 | 75.4 | 5.8 | 0.8 | 51.8 |
|  |  | 50 | 9.4 | 58.8 | 0.0 | 0.0 | 6.4 |
|  |  | 100 | 0.1 | 30.4 | 0.0 | 0.0 | 0.0 |
|  | 3 | 10 | 89.4 | 84.9 | 15.4 | 26.9 | 93.3 |
|  |  | 25 | 82.6 | 71.3 | 0.1 | 4.7 | 51.3 |
|  |  | 50 | 69.0 | 52.9 | 0.0 | 0.4 | 3.6 |
|  |  | 100 | 48.5 | 31.8 | 0.0 | 0.0 | 0.0 |
|  | 4 | 10 | 70.1 | 85.9 | 1.6 | 5.4 | 93.8 |
|  |  | 25 | 18.5 | 76.1 | 0.1 | 0.1 | 54.3 |
|  |  | 50 | 0.5 | 57.5 | 0.0 | 0.0 | 3.6 |
|  |  | 100 | 0.0 | 34.7 | 0.0 | 0.0 | 0.0 |
| Uniform & | 1 | 10 | [5.6] | [6.5] | 54.9 | 50.8 | 98.4 |
| Gamma$_1$ |  | 25 | [5.2] | [6.4] | 8.7 | 6.9 | 89.4 |
|  |  | 50 | [4.8] | [7.0] | 0.2 | 0.2 | 52.4 |
|  |  | 100 | [5.6] | [6.5] | 0.0 | 0.0 | 4.4 |
|  | 2 | 10 | [6.3] | [6.5] | 95.1 | 70.5 | 98.4 |
|  |  | 25 | [4.7] | [5.5] | 87.9 | 26.8 | 88.6 |
|  |  | 50 | [5.9] | [6.6] | 46.8 | 3.9 | 52.3 |
|  |  | 100 | [7.2] | [6.6] | 2.9 | 0.2 | 3.6 |
|  | 3 | 10 | [6.1] | [6.9] | 34.5 | 40.2 | 97.6 |
|  |  | 25 | [5.6] | [6.4] | 3.2 | 6.9 | 88.9 |
|  |  | 50 | [4.9] | [6.6] | 0.0 | 0.5 | 51.8 |
|  |  | 100 | [6.5] | [6.0] | 0.0 | 0.0 | 4.7 |
|  | 4 | 10 | [8.6] | [6.5] | 38.3 | 41.2 | 98.4 |
|  |  | 25 | [25.4] | [6.1] | 20.4 | 8.5 | 89.7 |
|  |  | 50 | [45.8] | [7.7] | 3.5 | 0.9 | 50.7 |
|  |  | 100 | [82.1] | [7.2] | 0.1 | 0.0 | 7.3 |

Table 14--Continued

| Population | Phase | n | $\mu$ | MWW | F | SR | KS |
|---|---|---|---|---|---|---|---|
| Uniform & Gamma$_2$ | 1 | 10 | 83.7 | 86.3 | 57.5 | 52.0 | 96.2 |
| | | 25 | 65.3 | 73.1 | 11.1 | 7.8 | 71.7 |
| | | 50 | 41.7 | 58.6 | 0.3 | 0.1 | 29.5 |
| | | 100 | 10.9 | 25.8 | 0.0 | 0.0 | 1.2 |
| | 2 | 10 | 78.1 | 87.1 | 81.2 | 55.9 | 95.7 |
| | | 25 | 51.3 | 73.8 | 58.1 | 12.0 | 72.1 |
| | | 50 | 20.9 | 57.5 | 17.4 | 0.2 | 28.1 |
| | | 100 | 2.7 | 28.9 | 0.3 | 0.0 | 1.7 |
| | 3 | 10 | 91.3 | 88.6 | 52.5 | 57.2 | 96.7 |
| | | 25 | 79.6 | 74.7 | 8.5 | 19.5 | 72.8 |
| | | 50 | 62.0 | 55.4 | 0.4 | 3.6 | 27.4 |
| | | 100 | 32.8 | 27.9 | 0.0 | 0.5 | 2.2 |
| | 4 | 10 | 75.6 | 86.5 | 18.2 | 19.9 | 95.9 |
| | | 25 | 33.9 | 75.3 | 5.4 | 1.5 | 75.1 |
| | | 50 | 6.3 | 55.5 | 0.2 | 0.0 | 26.8 |
| | | 100 | 0.2 | 28.3 | 0.0 | 0.0 | 1.6 |
| Uniform & Gamma$_3$ | 1 | 10 | 83.9 | 85.8 | 16.4 | 21.8 | 94.7 |
| | | 25 | 66.9 | 74.5 | 0.1 | 0.2 | 54.5 |
| | | 50 | 35.5 | 55.4 | 0.0 | 0.0 | 5.0 |
| | | 100 | 8.6 | 28.5 | 0.0 | 0.0 | 0.0 |
| | 2 | 10 | 71.5 | 85.2 | 57.1 | 32.5 | 93.2 |
| | | 25 | 37.7 | 76.5 | 11.2 | 0.5 | 56.5 |
| | | 50 | 7.8 | 58.4 | 0.1 | 0.0 | 6.7 |
| | | 100 | 0.5 | 34.4 | 0.0 | 0.0 | 0.0 |
| | 3 | 10 | 90.9 | 85.4 | 12.8 | 24.6 | 93.9 |
| | | 25 | 84.7 | 73.7 | 0.2 | 4.1 | 53.6 |
| | | 50 | 75.3 | 57.8 | 0.0 | 0.2 | 5.3 |
| | | 100 | 54.4 | 32.0 | 0.0 | 0.0 | 0.1 |
| | 4 | 10 | 69.5 | 85.9 | 2.7 | 5.4 | 93.2 |
| | | 25 | 18.7 | 75.0 | 0.0 | 0.0 | 55.8 |
| | | 50 | 1.0 | 60.0 | 0.0 | 0.0 | 6.1 |
| | | 100 | 0.0 | 31.6 | 0.0 | 0.0 | 0.0 |

Table 14--<u>Continued</u>

| Population | | Phase | n | $\mu$ | MWW | F | SR | KS |
|---|---|---|---|---|---|---|---|---|
| Normal$_1$ | & | 1 | 10 | 71.2 | 72.3 | [4.8] | [5.5] | 96.9 |
| Normal$_2$ | | | 25 | 40.2 | 40.4 | [4.5] | [6.9] | 83.4 |
| | | | 50 | 12.3 | 13.9 | [3.2] | [6.4] | 50.1 |
| | | | 100 | 0.1 | 0.5 | [2.8] | [6.5] | 13.5 |
| | | 2 | 10 | 74.2 | 74.3 | [3.6] | [13.2] | 96.2 |
| | | | 25 | 47.6 | 45.0 | [3.0] | [24.9] | 85.1 |
| | | | 50 | 14.8 | 13.9 | [4.6] | [45.2] | 51.9 |
| | | | 100 | 0.7 | 0.6 | [9.0] | [75.9] | 11.6 |
| | | 3 | 10 | 76.9 | 73.9 | [25.5] | [29.1] | 96.6 |
| | | | 25 | 49.4 | 41.1 | [22.2] | [52.4] | 82.5 |
| | | | 50 | 16.3 | 11.7 | [23.7] | [76.1] | 48.1 |
| | | | 100 | 1.5 | 0.7 | [31.7] | [93.3] | 13.9 |
| | | 4 | 10 | 78.2 | 74.2 | [38.9] | [35.4] | 96.7 |
| | | | 25 | 49.8 | 41.5 | [54.6] | [65.6] | 82.9 |
| | | | 50 | 19.4 | 13.1 | [67.3] | [85.4] | 49.5 |
| | | | 100 | 3.0 | 0.7 | [81.6] | [94.6] | 14.0 |
| Normal$_1$ | & | 1 | 10 | 65.7 | 66.0 | 83.5 | 83.2 | 95.3 |
| Normal$_3$ | | | 25 | 27.2 | 27.7 | 58.0 | 67.2 | 71.7 |
| | | | 50 | 3.8 | 5.0 | 30.9 | 43.7 | 23.8 |
| | | | 100 | 0.0 | 0.1 | 6.6 | 16.6 | 0.9 |
| | | 2 | 10 | 57.9 | 63.4 | 89.1 | 71.1 | 94.3 |
| | | | 25 | 20.2 | 27.6 | 64.3 | 33.5 | 71.7 |
| | | | 50 | 2.3 | 5.4 | 29.9 | 10.5 | 25.9 |
| | | | 100 | 0.0 | 0.1 | 4.8 | 0.4 | 0.9 |
| | | 3 | 10 | 79.6 | 65.4 | 61.5 | 71.2 | 94.6 |
| | | | 25 | 59.0 | 28.8 | 62.1 | 63.9 | 73.5 |
| | | | 50 | 29.4 | 5.2 | 50.9 | 53.3 | 26.0 |
| | | | 100 | 4.5 | 0.1 | 35.8 | 35.5 | 1.1 |
| | | 4 | 10 | 67.9 | 66.4 | 22.8 | 30.7 | 95.4 |
| | | | 25 | 21.1 | 30.9 | 6.4 | 6.9 | 72.3 |
| | | | 50 | 0.4 | 4.6 | 0.2 | 0.2 | 23.9 |
| | | | 100 | 0.1 | 0.2 | 0.0 | 0.0 | 1.7 |

Table 14--<u>Continued</u>

| Population | | Phase | n | $\mu$ | MWW | F | SR | KS |
|---|---|---|---|---|---|---|---|---|
| Normal$_2$ | & | 1 | 10 | [5.1] | [4.9] | 85.9 | 85.2 | 99.7 |
| Normal$_3$ | | | 25 | [4.9] | [4.9] | 61.9 | 66.9 | 99.1 |
| | | | 50 | [5.3] | [5.6] | 35.3 | 43.6 | 98.1 |
| | | | 100 | [5.7] | [5.5] | 7.1 | 14.4 | 95.5 |
| | | 2 | 10 | [5.9] | [5.3] | 92.8 | 79.4 | 99.7 |
| | | | 25 | [6.5] | [5.1] | 80.9 | 60.1 | 99.4 |
| | | | 50 | [9.2] | [4.5] | 64.3 | 41.4 | 98.5 |
| | | | 100 | [15.5] | [5.4] | 38.2 | 19.2 | 95.3 |
| | | 3 | 10 | [4.9] | [4.5] | 65.9 | 57.7 | 99.7 |
| | | | 25 | [8.1] | [5.7] | 52.9 | 41.9 | 99.2 |
| | | | 50 | [11.7] | [5.8] | 32.4 | 23.8 | 98.0 |
| | | | 100 | [25.2] | [4.2] | 6.8 | 9.2 | 95.1 |
| | | 4 | 10 | [3.7] | [4.7] | 43.3 | 71.1 | 99.9 |
| | | | 25 | [7.2] | [5.4] | 26.8 | 47.9 | 99.6 |
| | | | 50 | [9.8] | [5.7] | 15.0 | 29.8 | 97.8 |
| | | | 100 | [16.7] | [5.6] | 3.0 | 9.8 | 96.7 |
| Normal$_1$ | & | 1 | 10 | [5.3] | [5.4] | [4.3] | [5.3] | 99.9 |
| Gamma$_1$ | | | 25 | [4.9] | [5.3] | [4.9] | [6.8] | 99.4 |
| | | | 50 | [5.7] | [5.4] | [4.7] | [8.9] | 98.9 |
| | | | 100 | [5.3] | [5.5] | [4.5] | [12.7] | 98.5 |
| | | 2 | 10 | [5.5] | [5.5] | [1.2] | [5.4] | 99.9 |
| | | | 25 | [5.5] | [5.6] | [0.3] | [6.0] | 99.5 |
| | | | 50 | [5.6] | [5.2] | [0.2] | [6.8] | 99.2 |
| | | | 100 | [7.7] | [6.4] | [0.0] | [6.9] | 98.9 |
| | | 3 | 10 | [4.0] | [5.1] | [34.9] | [19.0] | 99.7 |
| | | | 25 | [4.0] | [5.1] | [27.6] | [16.4] | 99.4 |
| | | | 50 | [6.0] | [6.8] | [25.1] | [19.3] | 98.9 |
| | | | 100 | [6.8] | [4.8] | [24.2] | [24.0] | 99.1 |
| | | 4 | 10 | [3.4] | [3.7] | [29.6] | [26.1] | 99.9 |
| | | | 25 | [4.5] | [5.4] | [35.5] | [32.8] | 99.3 |
| | | | 50 | [6.1] | [5.5] | [41.2] | [31.6] | 99.0 |
| | | | 100 | [4.3] | [5.6] | [38.5] | [29.2] | 99.1 |

Table 14--<u>Continued</u>

| Population | Phase | n | $\mu$ | MWW | F | SR | KS |
|---|---|---|---|---|---|---|---|
| Normal₁ & Gamma₂ | 1 | 10 | 69.0 | 69.5 | [5.5] | [6.0] | 95.4 |
| | | 25 | 38.9 | 33.3 | [6.1] | [7.7] | 73.1 |
| | | 50 | 11.3 | 9.7 | [6.1] | [8.9] | 37.1 |
| | | 100 | 0.4 | 0.1 | [4.9] | [9.1] | 4.7 |
| | 2 | 10 | 69.7 | 69.0 | [6.3] | [15.3] | 94.8 |
| | | 25 | 41.3 | 37.9 | [4.1] | [26.1] | 75.8 |
| | | 50 | 8.0 | 7.5 | [6.8] | [51.9] | 32.7 |
| | | 100 | 0.4 | 0.2 | [6.9] | [76.7] | 4.9 |
| | 3 | 10 | 78.2 | 70.2 | [26.8] | [25.1] | 94.7 |
| | | 25 | 55.4 | 36.8 | [23.1] | [42.4] | 75.5 |
| | | 50 | 27.1 | 9.2 | [28.9] | [65.5] | 33.1 |
| | | 100 | 4.2 | 0.4 | [30.2] | [85.4] | 5.5 |
| | 4 | 10 | 81.8 | 71.5 | [43.6] | [38.2] | 94.7 |
| | | 25 | 60.5 | 37.3 | [50.4] | [63.0] | 75.9 |
| | | 50 | 39.8 | 10.9 | [51.1] | [73.7] | 35.8 |
| | | 100 | 8.8 | 0.2 | [57.3] | [90.4] | 3.9 |
| Normal₁ & Gamma₃ | 1 | 10 | 63.9 | 64.6 | 82.1 | 81.8 | 93.8 |
| | | 25 | 28.2 | 27.5 | 55.7 | 64.4 | 67.7 |
| | | 50 | 4.6 | 4.7 | 31.9 | 45.1 | 21.8 |
| | | 100 | 0.0 | 0.0 | 7.3 | 20.0 | 0.9 |
| | 2 | 10 | 59.4 | 64.9 | 88.5 | 72.2 | 94.0 |
| | | 25 | 20.8 | 28.2 | 68.8 | 37.6 | 68.7 |
| | | 50 | 1.9 | 5.5 | 36.9 | 9.8 | 23.0 |
| | | 100 | 0.0 | 0.0 | 8.9 | 0.8 | 1.2 |
| | 3 | 10 | 80.4 | 64.9 | 60.9 | 73.9 | 94.1 |
| | | 25 | 59.7 | 25.6 | 60.6 | 63.5 | 64.5 |
| | | 50 | 33.6 | 4.2 | 47.9 | 48.9 | 21.1 |
| | | 100 | 5.1 | 0.0 | 30.2 | 39.1 | 0.8 |
| | 4 | 10 | 70.8 | 64.6 | 25.4 | 31.2 | 93.9 |
| | | 25 | 27.4 | 30.4 | 11.9 | 9.0 | 69.1 |
| | | 50 | 2.7 | 4.4 | 2.1 | 2.3 | 19.9 |
| | | 100 | 0.1 | 0.1 | 0.3 | 0.1 | 1.0 |

Table 14--<u>Continued</u>

| Population | | Phase | n | $\mu$ | MWW | F | SR | KS |
|---|---|---|---|---|---|---|---|---|
| Normal$_2$ | & | 1 | 10 | 72.8 | 75.0 | [5.2] | [5.9] | 96.6 |
| Gamma$_1$ | | | 25 | 39.6 | 43.9 | [5.0] | [6.7] | 87.1 |
| | | | 50 | 12.2 | 16.9 | [4.5] | [7.2] | 61.8 |
| | | | 100 | 0.4 | 1.1 | [3.9] | [7.8] | 23.5 |
| | | 2 | 10 | 77.3 | 75.2 | [3.3] | [13.1] | 96.9 |
| | | | 25 | 51.5 | 45.2 | [3.5] | [28.3] | 88.1 |
| | | | 50 | 19.6 | 14.2 | [4.0] | [43.2] | 60.1 |
| | | | 100 | 2.3 | 1.0 | [7.6] | [74.8] | 21.9 |
| | | 3 | 10 | 77.9 | 78.1 | [27.5] | [30.7] | 97.9 |
| | | | 25 | 39.2 | 45.8 | [25.0] | [53.9] | 87.9 |
| | | | 50 | 8.8 | 14.6 | [27.5] | [74.9] | 59.2 |
| | | | 100 | 0.4 | 1.5 | [35.9] | [91.5] | 23.4 |
| | | 4 | 10 | 84.1 | 75.7 | [46.5] | [40.4] | 97.5 |
| | | | 25 | 54.3 | 44.2 | [63.8] | [66.0] | 86.9 |
| | | | 50 | 19.8 | 14.3 | [79.9] | [86.0] | 59.4 |
| | | | 100 | 2.1 | 0.5 | [92.9] | [94.9] | 22.1 |
| Normal$_2$ | & | 1 | 10 | [5.2] | [5.1] | [5.3] | [7.5] | 99.8 |
| Gamma$_2$ | | | 25 | [5.0] | [4.9] | [6.5] | [7.4] | 99.4 |
| | | | 50 | [5.9] | [6.3] | [6.5] | [8.8] | 98.8 |
| | | | 100 | [5.2] | [6.3] | [4.6] | [7.8] | 98.1 |
| | | 2 | 10 | [5.8] | [4.9] | [6.5] | [17.2] | 99.7 |
| | | | 25 | [5.9] | [5.2] | [4.8] | [24.7] | 99.3 |
| | | | 50 | [5.0] | [5.0] | [3.0] | [25.1] | 99.2 |
| | | | 100 | [6.9] | [6.9] | [3.3] | [25.9] | 98.5 |
| | | 3 | 10 | [4.2] | [4.4] | [13.6] | [21.1] | 99.7 |
| | | | 25 | [4.6] | [4.5] | [10.4] | [25.5] | 99.5 |
| | | | 50 | [4.9] | [5.8] | [9.6] | [25.1] | 99.0 |
| | | | 100 | [5.2] | [6.5] | [9.5] | [27.1] | 98.3 |
| | | 4 | 10 | [3.3] | [5.2] | [43.4] | [23.7] | 99.7 |
| | | | 25 | [4.9] | [5.0] | [49.1] | [30.5] | 99.3 |
| | | | 50 | [4.7] | [5.6] | [56.5] | [38.1] | 98.3 |
| | | | 100 | [4.8] | [8.1] | [54.8] | [55.5] | 98.0 |

Table 14--<u>Continued</u>

| Population | | Phase | n | $\mu$ | MWW | F | SR | KS |
|---|---|---|---|---|---|---|---|---|
| Normal$_2$ | & | 1 | 10 | [4.9] | [5.1] | 83.4 | 82.5 | 99.8 |
| Gamma$_3$ | | | 25 | [4.7] | [5.1] | 59.7 | 62.7 | 98.5 |
| | | | 50 | [6.3] | [7.0] | 31.6 | 37.4 | 96.3 |
| | | | 100 | [5.6] | [6.3] | 6.9 | 10.5 | 92.6 |
| | | 2 | 10 | [6.0] | [5.1] | 90.9 | 78.7 | 99.9 |
| | | | 25 | [8.7] | [5.4] | 82.7 | 57.6 | 98.4 |
| | | | 50 | [10.5] | [5.0] | 70.9 | 39.7 | 97.2 |
| | | | 100 | [18.4] | [4.5] | 48.0 | 14.1 | 92.8 |
| | | 3 | 10 | [4.2] | [3.9] | 66.9 | 60.5 | 99.7 |
| | | | 25 | [8.9] | [5.4] | 49.1 | 37.2 | 98.8 |
| | | | 50 | [15.9] | [4.5] | 26.4 | 19.9 | 97.8 |
| | | | 100 | [24.3] | [5.9] | 7.1 | 10.1 | 92.6 |
| | | 4 | 10 | [4.2] | [5.4] | 44.5 | 67.9 | 99.5 |
| | | | 25 | [5.9] | [5.8] | 33.8 | 51.0 | 98.5 |
| | | | 50 | [9.3] | [6.0] | 23.7 | 26.2 | 96.9 |
| | | | 100 | [12.5] | [4.8] | 12.5 | 14.5 | 91.9 |
| Normal$_3$ | & | 1 | 10 | 67.5 | 70.2 | 82.8 | 83.4 | 96.5 |
| Gamma$_1$ | | | 25 | 25.9 | 32.9 | 58.5 | 67.4 | 80.9 |
| | | | 50 | 2.8 | 5.4 | 30.2 | 45.9 | 35.9 |
| | | | 100 | 0.0 | 0.1 | 7.9 | 20.9 | 4.6 |
| | | 2 | 10 | 64.9 | 68.1 | 91.1 | 72.8 | 96.7 |
| | | | 25 | 27.1 | 32.5 | 65.2 | 34.9 | 80.4 |
| | | | 50 | 3.8 | 6.4 | 29.0 | 8.3 | 38.7 |
| | | | 100 | 0.0 | 0.1 | 3.5 | 0.6 | 3.3 |
| | | 3 | 10 | 80.6 | 68.9 | 60.9 | 71.7 | 96.4 |
| | | | 25 | 46.2 | 32.6 | 63.9 | 60.8 | 80.2 |
| | | | 50 | 15.4 | 7.4 | 56.7 | 46.2 | 41.6 |
| | | | 100 | 0.6 | 0.2 | 46.9 | 25.4 | 4.2 |
| | | 4 | 10 | 70.5 | 65.5 | 18.0 | 28.6 | 96.1 |
| | | | 25 | 21.5 | 30.1 | 4.7 | 6.2 | 79.2 |
| | | | 50 | 1.4 | 6.4 | 0.7 | 1.3 | 38.9 |
| | | | 100 | 0.0 | 0.1 | 0.0 | 0.1 | 4.6 |

Table 14--<u>Continued</u>

| Population | | Phase | n | $\mu$ | MWW | F | SR | KS |
|---|---|---|---|---|---|---|---|---|
| Normal$_3$ | & | 1 | 10 | [5.3] | [5.9] | 84.2 | 84.5 | 99.9 |
| Gamma$_2$ | | | 25 | [5.5] | [6.9] | 55.4 | 64.9 | 99.2 |
| | | | 50 | [4.9] | [7.2] | 33.0 | 45.1 | 97.6 |
| | | | 100 | [4.8] | [9.0] | 7.0 | 23.2 | 93.7 |
| | | 2 | 10 | [4.8] | [5.2] | 91.5 | 77.2 | 99.9 |
| | | | 25 | [5.3] | [6.3] | 77.8 | 63.7 | 99.4 |
| | | | 50 | [7.0] | [6.7] | 59.9 | 48.8 | 99.7 |
| | | | 100 | [7.4] | [10.1] | 36.4 | 37.9 | 93.9 |
| | | 3 | 10 | [5.7] | [5.1] | 65.2 | 62.2 | 99.7 |
| | | | 25 | [8.0] | [6.2] | 47.4 | 46.3 | 99.5 |
| | | | 50 | [10.1] | [6.8] | 27.7 | 31.5 | 97.9 |
| | | | 100 | [15.3] | [8.0] | 9.4 | 19.2 | 94.1 |
| | | 4 | 10 | [3.4] | [4.9] | 39.7 | 62.6 | 99.9 |
| | | | 25 | [6.4] | [6.1] | 24.0 | 39.8 | 99.1 |
| | | | 50 | [14.8] | [6.2] | 9.5 | 17.7 | 97.9 |
| | | | 100 | [27.8] | [8.5] | 1.5 | 6.1 | 94.6 |
| Normal$_3$ | & | 1 | 10 | [3.7] | [3.8] | [5.3] | [8.7] | 99.9 |
| Gamma$_3$ | | | 25 | [4.9] | [5.0] | [5.6] | [8.9] | 99.7 |
| | | | 50 | [6.4] | [6.4] | [5.3] | [10.5] | 99.4 |
| | | | 100 | [5.2] | [5.5] | [4.4] | [11.4] | 99.5 |
| | | 2 | 10 | [5.0] | [4.9] | [6.3] | [15.4] | 99.9 |
| | | | 25 | [4.7] | [5.2] | [4.9] | [20.3] | 99.4 |
| | | | 50 | [5.1] | [4.8] | [4.5] | [16.2] | 99.5 |
| | | | 100 | [5.2] | [5.4] | [4.7] | [11.1] | 99.4 |
| | | 3 | 10 | [2.9] | [4.6] | [38.6] | [31.9] | 99.7 |
| | | | 25 | [5.7] | [5.4] | [27.5] | [43.1] | 99.4 |
| | | | 50 | [5.7] | [4.6] | [24.6] | [64.8] | 99.4 |
| | | | 100 | [5.9] | [5.3] | [21.9] | [89.5] | 99.2 |
| | | 4 | 10 | [2.3] | [4.2] | [44.6] | [17.8] | 99.9 |
| | | | 25 | [3.8] | [4.5] | [49.9] | [17.9] | 99.6 |
| | | | 50 | [5.2] | [4.1] | [46.2] | [30.5] | 99.6 |
| | | | 100 | [4.3] | [4.9] | [47.9] | [36.8] | 99.0 |

Table 14--<u>Continued</u>

| Population | | Phase | n | μ | MWW | F | SR | KS |
|---|---|---|---|---|---|---|---|---|
| Gamma₁ | & | 1 | 10 | 71.6 | 72.0 | [5.6] | [6.6] | 95.9 |
| Gamma₂ | | | 25 | 39.1 | 36.5 | [6.8] | [8.5] | 77.9 |
| | | | 50 | 12.6 | 11.5 | [5.8] | [9.9] | 45.1 |
| | | | 100 | 0.6 | 0.5 | [5.2] | [9.7] | 8.9 |
| | | 2 | 10 | 75.5 | 73.2 | [5.7] | [15.1] | 96.1 |
| | | | 25 | 44.4 | 38.9 | [4.2] | [24.9] | 79.4 |
| | | | 50 | 14.7 | 9.5 | [5.4] | [48.7] | 42.0 |
| | | | 100 | 1.3 | 0.3 | [6.9] | [78.8] | 11.6 |
| | | 3 | 10 | 76.3 | 72.0 | [26.9] | [28.9] | 95.7 |
| | | | 25 | 41.7 | 36.1 | [25.1] | [49.3] | 78.3 |
| | | | 50 | 14.3 | 10.1 | [30.6] | [68.7] | 42.3 |
| | | | 100 | 1.1 | 0.6 | [47.2] | [86.8] | 10.0 |
| | | 4 | 10 | 83.6 | 71.9 | [45.4] | [41.3] | 96.2 |
| | | | 25 | 64.2 | 38.8 | [55.5] | [65.2] | 80.5 |
| | | | 50 | 36.5 | 10.3 | [63.8] | [77.3] | 44.9 |
| | | | 100 | 10.5 | 0.8 | [70.3] | [89.6] | 7.0 |
| Gamma₁ | & | 1 | 10 | 64.8 | 66.9 | 84.2 | 82.8 | 94.9 |
| Gamma₃ | | | 25 | 25.1 | 30.5 | 58.3 | 64.3 | 74.4 |
| | | | 50 | 3.4 | 6.1 | 32.6 | 46.2 | 31.3 |
| | | | 100 | 0.0 | 0.1 | 8.3 | 20.4 | 3.2 |
| | | 2 | 10 | 62.9 | 66.2 | 88.9 | 72.4 | 94.3 |
| | | | 25 | 23.1 | 27.9 | 69.5 | 40.3 | 71.6 |
| | | | 50 | 3.8 | 5.5 | 44.9 | 13.0 | 32.6 |
| | | | 100 | 0.0 | 0.1 | 8.0 | 0.7 | 2.8 |
| | | 3 | 10 | 77.2 | 68.0 | 61.1 | 72.1 | 95.0 |
| | | | 25 | 47.1 | 27.7 | 62.8 | 61.9 | 71.9 |
| | | | 50 | 17.2 | 5.7 | 52.3 | 50.4 | 31.8 |
| | | | 100 | 1.3 | 0.0 | 37.2 | 34.1 | 2.6 |
| | | 4 | 10 | 74.9 | 66.7 | 24.2 | 31.6 | 94.7 |
| | | | 25 | 29.1 | 29.2 | 10.1 | 8.7 | 74.7 |
| | | | 50 | 3.9 | 6.0 | 2.3 | 2.1 | 32.9 |
| | | | 100 | 0.0 | 0.1 | 0.1 | 0.1 | 5.2 |

Table 14--<u>Continued</u>

| Population | Phase | n | $\mu$ | MWW | F | SR | KS |
|---|---|---|---|---|---|---|---|
| Gamma$_2$ & | 1 | 10 | [5.7] | [5.9] | 83.0 | 81.6 | 99.6 |
| Gamma$_3$ | | 25 | [5.1] | [6.0] | 59.9 | 65.4 | 99.2 |
| | | 50 | [4.7] | [5.7] | 31.0 | 43.9 | 97.1 |
| | | 100 | [5.7] | [7.9] | 8.3 | 22.1 | 91.5 |
| | 2 | 10 | [5.6] | [5.6] | 88.2 | 74.7 | 99.5 |
| | | 25 | [5.6] | [5.2] | 80.8 | 58.9 | 99.0 |
| | | 50 | [8.4] | [6.0] | 68.3 | 44.3 | 97.4 |
| | | 100 | [10.3] | [7.7] | 45.9 | 26.0 | 92.5 |
| | 3 | 10 | [6.2] | [4.7] | 66.8 | 60.8 | 99.7 |
| | | 25 | [9.3] | [6.0] | 48.4 | 43.9 | 99.3 |
| | | 50 | [14.3] | [6.5] | 24.3 | 22.9 | 96.8 |
| | | 100 | [22.3] | [6.5] | 5.4 | 11.2 | 92.4 |
| | 4 | 10 | [3.3] | [5.6] | 38.7 | 61.1 | 99.6 |
| | | 25 | [5.6] | [6.0] | 28.5 | 40.8 | 99.0 |
| | | 50 | [9.5] | [5.5] | 15.2 | 20.1 | 97.9 |
| | | 100 | [21.5] | [6.7] | 5.8 | 6.9 | 93.5 |

Note: [1]$\mu$: confidence interval test for the difference of population means

[2]MWW: Mann-Whitney/Wilcoxon two-sample test for medians

[3]F: F test for variances

[4]SR: Squared Ranks two-sample test for variances

[5]KS: Kolmogorov-Smirnov two-sample test for distributions

[6]$\alpha$ = 5%

[7]All entry values are the percentages of the Type II error rates except those values in brackets that are Type I error rates.

error rates of the CI test uniformly deteriorated with
smaller samples when a true disparity of two population
means existed.  The deterioration could be caused by
increased sample variances as the sample size shrank.
Increasing the sample size remarkably improved the
performance in all population combinations, regardless of
either normal or nonnormal, symmetric or nonsymmetric
distributions.  In many instances, the rates reached 5% or
below when the sample size increased to 100.  The Type II
error with distance changes between points on the scale
(phases 2, 3, and 4) reported both better and worse rates
than without distance change (phase 1).  Therefore, the
various error rate patterns in phases 2, 3, and 4 indicated
that the parametric CI test had an effect on the scale with
unequal distances between points.

In experiment 2, the CI test was not sensitive to
distributional heterogeneity and distance changes between
points.  This was not the case in experiment 3.  The Type I
error rates were not uniformly maintained at the nominal 5%
when the two population means were identical in experiment
3.  The Type I error rates were drastically augmented when
distances were unequally changed between points on the scale
in phase 4.  It was essential to note that the Type I error
rates deteriorated with increasing sample size.  Usually,
the smaller sample provides less information and produces
larger error rates (Scheaffer, Mendenhall, and Ott 1990).
Here, the disturbing outcome might result from the fact that

the sample mean was not an unbiased estimator of the
population mean, because all seven populations were mapped
by a mapping function (table 4) and transformed into seven-
point Likert-scale values. Therefore, although the
variation of the estimator decreased when the sample size
increased, the bias did not (Hogg and Tanis 1977; Gregoire
and Driver 1987).

## MWW Test for the Difference of Two Population Medians

As expected, the error rates showed similar patterns
among the four phases in each population comparison, since
unequally changing distances between points should not have
an effect on the nonparametric test. The Type II error
rates of the MWW test in experiment 3 reported in a 26%-90%
range when comparing the uniform population against other
populations. The rates were much lower when comparing
normal or gamma population against each other; when the
sample size increased to 100, the MWW test performed nearly
perfectly in all instances. Figure 3 displays the two error
patterns where the rates are the averages of all specified
population combinations. Thus, the MWW test appeared better
able to detect median equality when comparing the normal and
gamma populations against each other than when comparing the
uniform population against normal or gamma populations.

The MWW test performed consistently well when
sampling from populations with homogeneous medians; the Type
I error rates exhibited slightly above the nominal 5%,

Type II error rate (%)

Sample size

Pattern 1   Pattern 2

Fig. 3.   ERROR PATTERNS OF MWW TEST (AVERAGE ERROR RATES)
Pattern 1 - comparing uniform against others
Pattern 2 - comparing normal & gamma against each other

regardless of sample size, in all population combinations. The circumstances showed that the MWW test was less sensitive to distribution differences when the populations had identical medians.

### F Test for the Difference of Two Population Variances

The parametric $F$ test with smaller samples appeared to be sensitive, not only to departure from normality but also to changing distances between points on the scale. There was a universal deterioration in the Type II error rates when sample size shrank. The $F$ test presented slightly better overall performance when comparing the uniform population against normal or gamma populations than comparing the normal and gamma populations against each other; the test reached excellent performance when the sample size increased to 50 or above, regardless of phase. Figures 4 and 5 display the error patterns where the rates are the averages of all specified population combinations. It is of interest to note that the $F$ test has serious difficulty in distinguishing two population variances when true variance and mean differences exist; the $F$ test becomes not sensitive to normality and the scale with unequal distances between points when large samples are selected.

In experiment 3, the $F$ test was erratic with anomalous patterns in terms of the Type I error rates. No consistent pattern could indicate that the circumstances
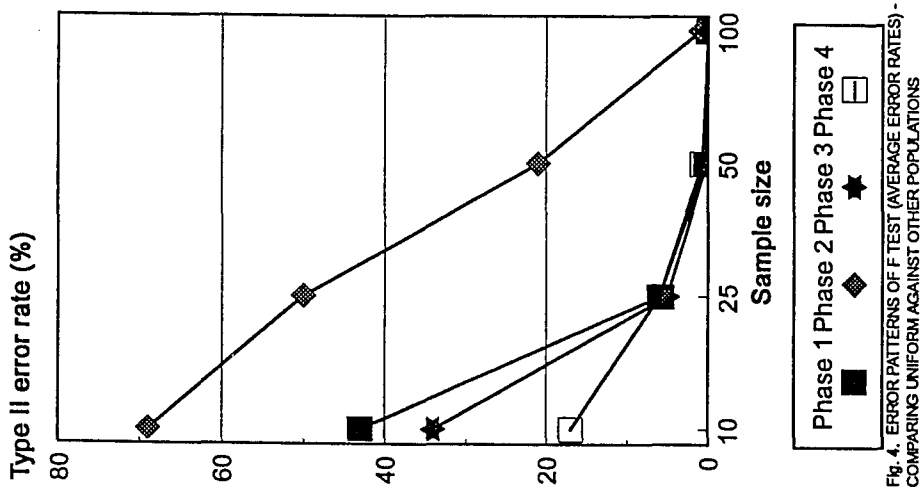
Type II error rate (%)

Sample size

Phase 1 Phase 2 Phase 3 Phase 4

Fig. 5. ERROR PATTERNS OF F TEST (AVERAGE ERROR RATES) -
COMPARING NORMAL & GAMMA AGAINST EACH OTHER



Type II error rate (%)

Sample size

Phase 1 Phase 2 Phase 3 Phase 4
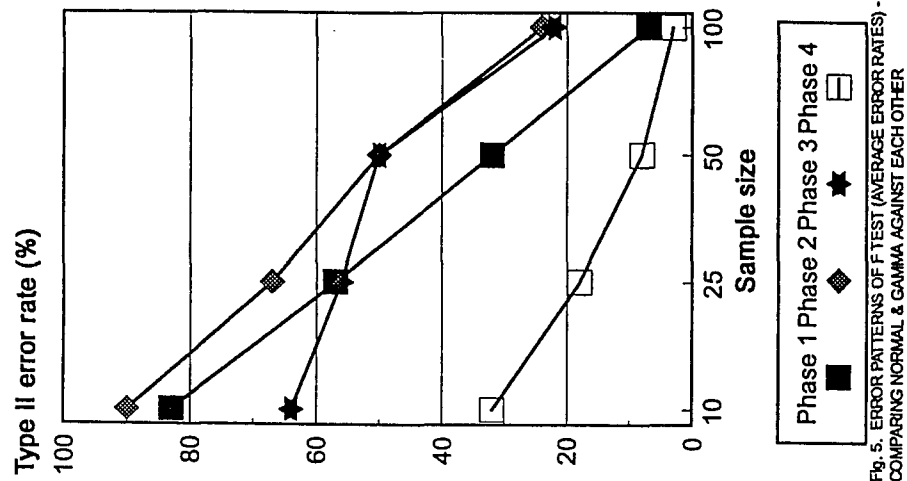
Fig. 4. ERROR PATTERNS OF F TEST (AVERAGE ERROR RATES) -
COMPARING UNIFORM AGAINST OTHER POPULATIONS

under which the $F$ test might be expected to perform well or poorly. The Type I error rates revealed both positive and negative correlations with increasing sample size among the four phases. In addition to the impact of departure from normality, the Type I error rates were evidently distorted by the distance changes between points on the scale. The $F$ test performed almost perfectly in phase 2 when comparing $normal_1$ population against $gamma_1$ populations as exhibited in figure 6. The surprisingly good result was one of the haphazard instances among all test conditions. No further knowledge could be learned or benefited from this population combination or the distances between points because the nearby phases 3 and 4 of the same population comparison ($normal_1$ and $gamma_1$) presented a crooked error pattern. Nor did the excellent performance result from the identical means and variances, because the result of phase 2 was even worse when comparing $normal_2$ against $gamma_2$ populations (having equal means and variances) as shown in figure 7.

## SR Test for the Difference of Two Population Variances

The SR test was as erratic as the $F$ test in each population comparison. The Type II error rates apparently decreased with increasing sample size and were close to perfect when the sample size reached 100 in many instances. A better error rate performance was exhibited when comparing the uniform population against normal or gamma populations than when comparing the normal or gamma populations against

Type II error rate (%)

Phase 1  Phase 2  Phase 3  Phase 4

Sample size

Fig. 7. ERROR PATTERNS OF F TEST - COMPARING NORMAL 2 AGAINST
GAMMA 2 POPULATIONS



Type II error rate (%)

Phase 1  Phase 2  Phase 3  Phase 4

Sample size
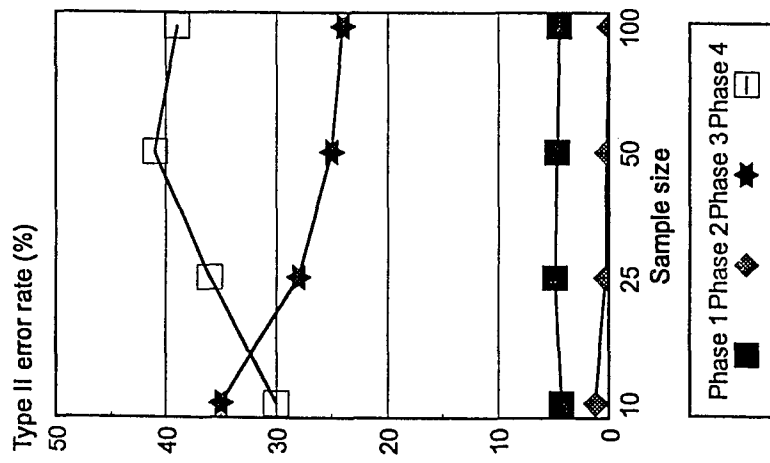
Fig. 6.  ERROR PATTERNS OF F TEST - COMPARING NORMAL 1
AGAINST GAMMA 1 POPULATIONS
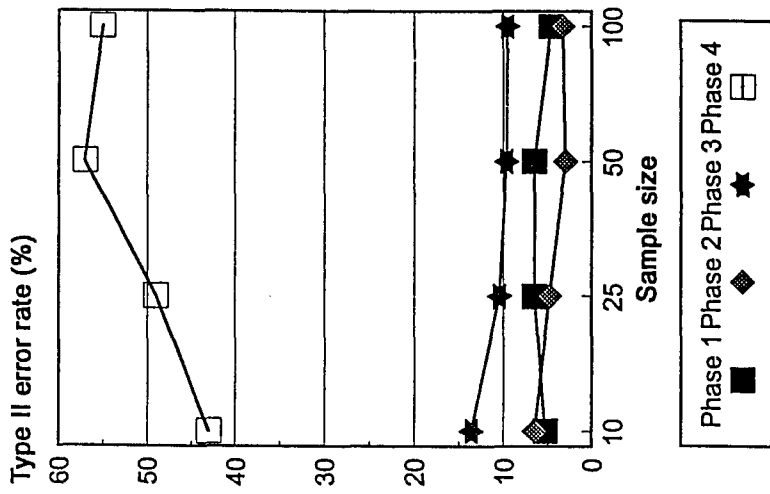
80

each other. Figure 8 displays the error patterns where the
rates are the averages of all specified population
combinations. Therefore, the SR test could easily detect
the heterogeneous variances with large samples (n > 30) when
comparing the uniform population against other populations.

The Type I error rates uniformly augmented with
increasing sample size. More distance changes between
points (phase 4) presented less reliable power than fewer
distance changes between points (phase 3 and 2) when
homogeneous variances existed. The Type I error rates also
indicated that the SR test tended to work slightly better
when sampling from two populations with identical population
variances and means than otherwise. Whether or not the
populations are normal is not an important issue in this
matter.

KS Test for the Difference of Two Population Distributions

All seven populations had different distributions,
so only the Type II error was of concern here. Two error
patterns could be found in the KS test when sampling from
normal or gamma populations. First, the Type II error rates
were consistently higher than 90% when sampling from two
populations with identical means. Increasing the sample
size could decrease only negligibly the error rates. The KS
test almost always erroneously declared dissimilar
distributions when the two populations had equal means.
Apparently the KS test has difficulty in distinguishing the
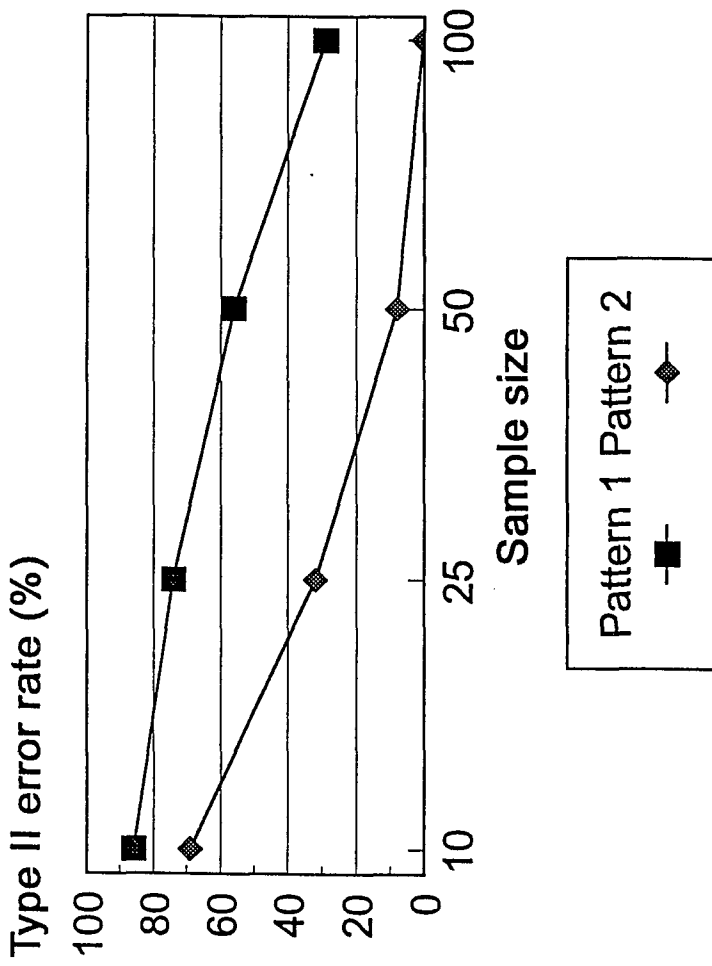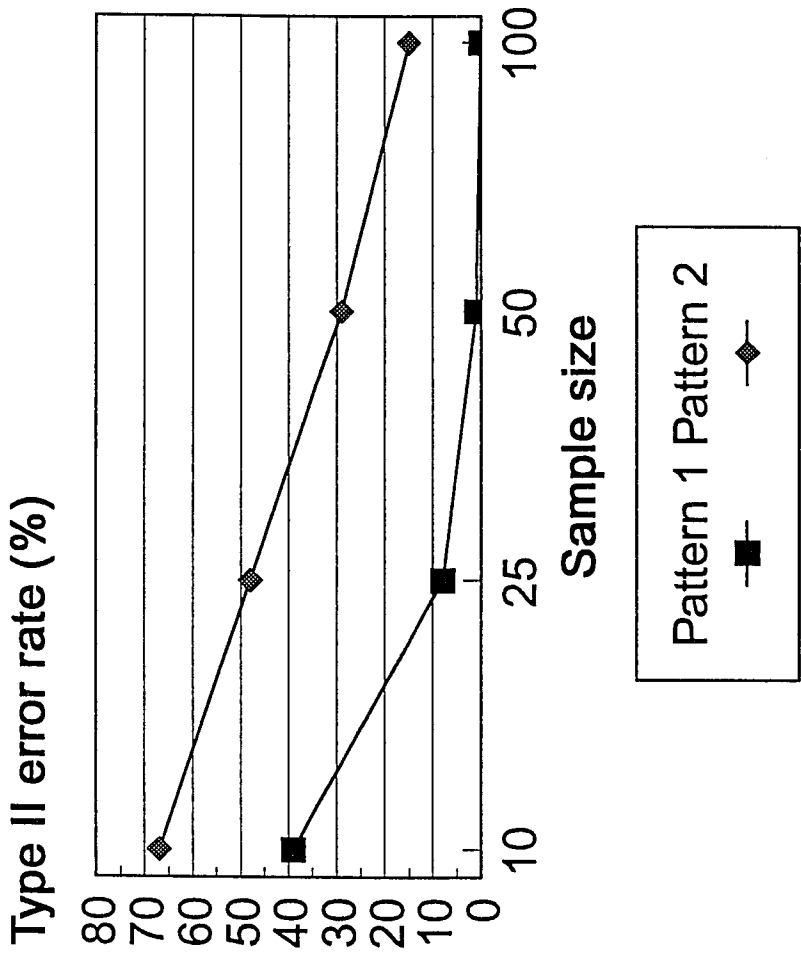
Fig. 8. ERROR PATTERNS OF SR TEST (AVERAGE ERROR RATES)
Pattern 1 - comparing uniform against others
Pattern 2 - comparing normal & gamma against each other

two populations with equal means. Second, the Type II error
rates dramatically decreased with increasing sample size
when sampling from two populations without identical means.
In many instances, the dissimilar distributions were capably
detected by the KS test when the sample size increased to
100. Figure 9 displays the two error patterns where the
rates are the averages of all specified population
combinations. Both error patterns contained the population
comparisons with equal and unequal variances. Surprisingly,
the population variance had no critical effect on the
determination of error patterns in the KS test. Also, both
symmetric and nonsymmetric populations were alternatively
used by the tests of two error patterns. Hence, the issue
of whether or not the populations are symmetric is not very
meaningful in examining the power-efficiency of the KS test.

The Type II error rates revealed the third error
pattern (similar to the second error pattern) when comparing
the uniform population against normal or gamma populations,
regardless of population means, variances, or symmetry.
Figure 10 displays the third pattern where the rates are the
averages of all specified population combinations. The KS
test could easily detect the distributional heterogeneity
when the sample size increased to 100.

Type II error rate (%)

100

80

60

40

20

0

10          25          50          100

Sample size

┌─────────────┐
│  Pattern 3  │
│      ■      │
└─────────────┘

Fig. 10.  ERROR PATTERNS OF KS TEST (AVERAGE ERROR RATES)
COMPARING UNIFORM AGAINST OTHER POPULATIONS
Pattern 3 - using uniform

Type II error rate (%)

100

80

60

40

20

0

10          25          50          100

Sample size

┌──────────────────────┐
│  Pattern 1  Pattern 2│
│      ■          ◈    │
└──────────────────────┘

Fig. 9.  ERROR PATTERNS OF KS TEST (AVERAGE ERROR RATES) - COMPARING
NORMAL & GAMMA AGAINST EACH OTHER
Pattern 1 - with equal means    Pattern 2 - with unequal means

CHAPTER 5

SUMMARY AND IMPLICATIONS

## Discussion

In the current research, three experiments were
conducted using normal and nonnormal distributions to
examine the robustness of parametric and nonparametric tests
(Huber 1977). The normal distribution included three normal
populations (with various central locations or dispersion),
whereas the nonnormal distribution included a symmetric
population (uniform) and three nonsymmetric populations
(gamma).

### Test of Central Tendency

With equal distances between points on the scale,
the Type I error of the CI test (for mean equality) and the
MWW test (for median equality) consistently occurred at the
nominal 5% in experiments 1 and 2. Yet the Type I error
rates of the MWW test slightly deteriorated when sampling
from populations with heterogeneous variances in experiment
3, whereas the CI test still remained at expected level.
Increasing the sample size could even worsen the case in the
MWW test. As evidenced by the Type I error rates summarized
from three experiments, the MWW test had more limitations
than the CI test with regard to location parameters and

convenience. For this reason, a pretest of variance
equality using the $F$ test or the SR test might be a
safeguard before heavily relying on either the CI test or
the MWW test to examine the central locations.

## Test of Dispersion

The Types I and II error rates of both the $F$ test
and the SR test appeared similarly wayward to the distance
changes between points on the scale. If the population
variance differences indeed existed, there was no rule to
determine that one test is preferred to another. The Type
II error rates of both tests was severely erratic with
smaller samples, but the two performed extremely well with
larger samples; whether the two populations had identical
means was not important. For the Type I error rates of both
tests, it was also interesting to note that the best
underlying population was uniform, followed by normal and
gamma. In addition, the symmetric populations apparently
overperformed the nonsymmetric populations in the
experiments.

## Test of Distribution

The results of experiments 2 and 3 together
indicated that the power of the KS test was justified when
sampling from identical distributions or when using large
samples with unequal population means. The distance changes
between points had no effect on this nonparametric test.
Yet the Type II error performance was noticeably poor when

comparing the normal or gamma populations each other with equal means. Hence, if no prior information confirmed mean homogeneity between populations, a pretest was advisable in order to examine the underlying differences in central location (mean) before the KS test was selected for use. The CI test for mean equality, which performed extremely well in experiment 2, could provide reliable protection in this regard.

## Conclusions

The current research examined the validity of experiment 1 that paved the way for the next two experiments. Then experiments 2 and 3 compared parametric and nonparametric tests under two performance indicators: the Type I error and the Type II error rates.

### Ordinal Measurement

Although the seven observed populations did not provide conclusive proof about what to expect for all required populations, results of the three experiments could reassure the researchers that the ordinal measures (mapped from seven populations) were justified for both the parametric CI test and the $F$ test. Moreover, little research has been completed on the examination of statistical test performance with unequal and unbalanced distance changes between points on the scale. Here, the seven-point Likert-scale data and changing distances with an unequal and unbalanced scale had revealed some degree of

success in describing the disturbing relationships of ordinal and interval measurement. Also, the parametric tests had performed well in spite of misgivings that, from the viewpoint of the measurement scale typology, only nonparametric tests were appropriate to use for the ordinal data.

### Comparison between Parametric and Nonparametric Tests

In the three experiments, treating ordinal data with three underlying distributions (uniform, normal, and gamma) as interval data for the parametric tests did not lead to inappropriate results. In many instances, both the parametric CI test and the $F$ test appeared to be more robust than their nonparametric alternatives as long as the distance between points changed monotonically on the scale, as in phases 1 and 2. There was a distortion in the error rate performance when the underlying ordinality assumptions were seriously violated, as in phase 3 and 4; both parametric and nonparametric tests performed erratically in examining mean, median, or variance equality, because the bias and inconsistency of the Likert-scale mean and variance were respectively used as an estimator of the population mean and variance when having unequal and unbalanced distances between points on the scale.

### Decisions of Hypothesis in Experiment 1

Hypothesis 1 was tested via one sample selected from one population with no distance change between points on the

scale.  The hypothesis is reiterated as follows:

H1:  Equally changing distances between points on the scale
     do not affect the parametric and nonparametric test
     results if one sample is selected from one population.

        The sample data, as conceived, constituted an

ordinal scale.  The Type I error of the CI test for the mean

and the KS test frequently occurred at, or below, the

nominal 5%.  Both the CI test for the median and the chi-

square test for the standard deviation were sensitive to the

nonnormal distribution and sample size, rather than the

ordinal data.  No serious impact could be found for the

equal distances between points on the underlying continuum.

In experiment 1, therefore, equally changing distances

between points on the scale do not affect the parametric and

nonparametric test results if one sample is selected from

one population.

Decisions of Hypothesis in Experiment 2

## Phase 1 Tests

        Hypothesis $2_a$ was tested via two samples selected

from one population with no distance change between points

on the scale.  The hypothesis is reiterated as follows:

H2$_a$:  Equally changing distances between points on the scale
       do not affect the parametric and nonparametric test
       results if two samples are selected from one
       population.

The sample data constituted an ordinal scale. The Type I error rates of CI test and MWW test uniformly retained at nominal 5% level for all populations. The error rates of $F$ test and SR test also retained at 5%-9% level for three normal populations, but not for the other nonnormal populations. Both the CI test and the $F$ test consistently transcended their nonparametric alternatives, the MWW test and the SR test. The KS test performed extremely well in either symmetric or nonsymmetric, normal or nonnormal populations. Thus, equally changing distances between points on the scale do not affect the parametric and nonparametric test results if two samples are selected from one population.

## Phase 2 Tests

Hypothesis $2_b$ was tested via two samples selected from one population with one distance change between points on the scale. The hypothesis is reiterated as follows:

$H2_b$: Unequally changing distances between points on the scale (one unequal change made on both sides of the neutral point) affect the parametric test results but do not affect the nonparametric test results if two samples are selected from one population.

One distance change was made on both sides of the neutral position. The scale was violated with two longer distances in the middle area. The CI test in phase 2 had the Type I error rates about 1% higher than in phase 1 to distinguish the means. Both the $F$ test and the SR test in

phase 2 had either more (normal and uniform populations) or less (gamma populations) power than in phase 1 to detect the variances. The performance of the MWW test and the KS test remained the same as in phase 1. Therefore, unequally changing distances between points on the scale (one change on both sides of neutral) affect the parametric test results, but not the nonparametric test (without the SR test) results if two samples are selected from one population. The SR test is a hybrid of parametric and nonparametric tests, so it has some effect on the distance changes in both normal and nonnormal populations.

## Phase 3 Tests

Hypothesis 2$_c$ was tested via two samples selected from one population with two distance changes between points on the scale. The hypothesis is reiterated as follows:

H2$_c$: Unequally changing distances between points on the scale (two unequal changes made on both sides of the neutral point) affect the parametric test results but do not affect the nonparametric test results if two samples are selected from one population.

Two distance changes were made on both sides of the neutral position. The scale was violated with four longer distances in the two extreme areas. The CI test in phase 3 had the Type I error rates about 1%-2% higher than in phase 1 to distinguish the means, regardless of distribution. Both the $F$ test and the SR test in phase 3 (mostly error rates higher than 20%) appeared to have less power than in

phase 1 to detect the variances, except that the underlying
population was uniform where the performance was
surprisingly good.  As expected, the performance of both the
MWW test and the KS test remained the same as in phase 1.
Thus, unequally changing distances between points on the
scale (two changes on both sides of neutral) affect the
parametric test results, but not the nonparametric test
(without the SR test) results if two samples are selected
from one population.  The SR test has a serious effect on
the distance changes in both normal and nonnormal
populations.

## Phase 4 Tests

Hypothesis $2_d$ was tested via two samples selected
from one population with more than two distance changes
between points on the scale.  The hypothesis is reiterated
as follows:

$H2_d$:  Unequally changing distances between points on the
scale (more than two changes made on both sides of the
neutral point) affect the parametric test results but
do not affect the nonparametric test results if two
samples are selected from one population.

More than two distance changes were made on both
sides of neutral position.  The underlying continuum was
violated to constitute an unbalanced scale.  The Type I
error rates of the CI test were equal to, or below, the
nominal 5% but deteriorated with increasing sample size.
The $F$ test and the SR test had the least power among the

four phases to detect the variances, especially in the
nonsymmetric distribution, three gamma populations (error
rates higher than 30%). The performance of the MWW test and
the KS test remained the same as in phase 1 and appeared
expectable. Therefore, unequally changing distances between
points on the scale (more than two changes on both sides of
neutral) affect the parametric test results, but not the
nonparametric test (without the SR test) results if two
samples are selected from one population. The SR test has
an erratic effect on the distance changes in both normal and
nonnormal populations.

<div align="center">Decisions of Hypothesis in Experiment 3</div>

## Phase 1 Tests

Hypothesis $3_a$ was tested via two samples selected
from two populations with no distance change between points
on the scale. The hypothesis is reiterated as follows:

$H3_a$: Equally changing distances between points on the scale
do not affect the parametric and nonparametric test
results if two samples are selected from two
populations.

Both parametric and nonparametric tests reported
high Type II error rates (mostly greater than 50%) and did
not capably detect the true location differences of means,
medians, variances, and distributions, respectively. The
performance was affected by the sample size, rather than the
distance changes between points. The Type I error rates

were frequently at the nominal 5% for the CI test and the $F$ test. No serious impact could be found in the parametric tests. Therefore, unequally changing distances between points on the scale do not affect the parametric and nonparametric test results if two samples are selected from two populations.

## Phase 2 Tests

Hypothesis $3_b$ was tested via two samples selected from two populations with one distance change between points on the scale. The hypothesis is reiterated as follows:

H3$_b$:  Unequally changing distances between points on the scale (one unequal change made on both sides of the neutral point) affect the parametric test results but do not affect the nonparametric test results if two samples are selected from two populations.

The CI test, the $F$ test, and the SR test had either better or worse performance than in phase 1 in distinguishing the true location differences; the MWW test and the KS test remained the same as in phase 1. The CI test, the $F$ test, and the SR test also reported slightly higher Type I error rates than in phase 1, regardless of distribution. Therefore, unequally changing distances between points on the scale (one change on both sides of neutral) affect the parametric test results, but not the nonparametric test (without the SR test) results if two samples are selected from two populations. The SR test has as erratic an effect on the distance changes as do the

parametric tests.

## Phase 3 Tests

Hypothesis $3_c$ was tested via two samples selected
from two populations with two distance changes between
points on the scale. The hypothesis is reiterated as
follows:

$H3_c$: Unequally changing distances between points on the
scale (two unequal changes made on both sides of the
neutral point) affect the parametric test results but
do not affect the nonparametric test results if two
samples are selected from two populations.

The CI test, the $F$ test, and the SR test had either
better or worse performance than in phase 1 in
distinguishing the true location differences, but the $F$
test and the SR test unexpectedly had better performance
than phase 2 in distinguishing the true variation
differences. The MWW test and the KS test remained the same
as phases 1 and 2. The CI test, the $F$ test, and the SR
test also reported consistently higher Type I error rates
than phases 1 and 2, regardless of distribution. Therefore,
unequally changing distances between points on the scale
(two changes on both sides of neutral) affect the parametric
test results, but not the nonparametric test (without the SR
test) results if two samples are selected from two
populations. The SR test has as erratic an effect on the
distance changes as do the parametric tests.

<u>Phase 4 Tests</u>

Hypothesis $3_d$ was tested via two samples selected from two populations with more than two distance changes between points on the scale. The hypothesis is reiterated as follows:

$H3_d$: Unequally changing distances between points on the scale (more than two changes made on both sides of the neutral point) affect the parametric test results but do not affect the nonparametric test results if two samples are selected from two populations.

The CI test, the $F$ test, and the SR test surprisingly had better performance than any other phases in distinguishing the true location differences, except when comparing gamma populations to each other. The MWW test and the KS test remained the same as in phases 1, 2, and 3. The CI test, the $F$ test, and the SR test, however, reported the worst Type I error rates among the four phases, regardless of distribution. Therefore, unequally changing distances between points on the scale (more than two changes on both sides of neutral) affect the parametric test results, but not the nonparametric test (without the SR test) results if two samples are selected from two populations. The SR test has as erratic an effect on the distance changes as do the parametric tests in both normal and nonnormal populations.

Selection between Parametric and Nonparametric Tests

It would be premature to conclude that the type of measurement scale had little relevance to the question of

whether to use parametric or nonparametric tests (Lord 1953;
Neibecker 1984; Thompson and Tapia 1990). However, the
parametric tests consistently transcended their
nonparametric alternatives in phases 1 and 2 of experiments
2 and 3. In regard to practical requirements, it is thus
concluded that magnitude scaling (numeric estimates) can be
accepted as a valid and reliable alternative to category
scaling (ordinal data). Use of parametric tests can have
the advantage of achieving more reliable performance (in
terms of the Type I and the Type II errors) when one tests
the location equality with steadily changing distances
between points on the scale; the nonparametric tests,
however, would be informative alternatives, or at least
useful complements, to parametric tests in many instances.

### Limitations of Research

The results of the three experiments offered strong
support for the validity of the magnitude scaling technique.
However, the same scale was used one thousand times in each
phase, so the error rate performance was analyzed by the
aggregate data that was simulated from the four phases of
scales. Teas (1987) indicated that the power function
exponents varied considerably from individual to individual.
Apparently, the respondents might not express their
preferred labeled points and relative strengths in a
consistent manner. If the error rate performance was
analyzed at an individual level, dramatic heterogeneity

across individuals could be revealed.

The normal distribution was one of the underlying assumptions for the parametric tests chosen in the current research. Both parametric and nonparametric tests were focused on the violation of normality in the populations. However, there were only seven populations generated in three experiments. The nonnormal (symmetric and nonsymmetric) distributions were restricted to uniform and gamma populations. Apparently, the uniform and gamma populations did not typically represent all types of nonnormal distributions. Also, the use of only one mapping function in the experiments was another limitation to generating more symmetric and nonsymmetric populations for the tests.

### Directions for Future Research

This research investigates samples selected from three types of distributions (uniform, normal, and gamma). It would be informative to rigorously generate a series of populations with coarser nonnormal shapes for the statistical tests. Much would be gained by a further investigation of the error rates of sampling from a greater variety of populations. Such an extension would ensure that the conclusions are not unique to the seven populations chosen (one uniform, three normal, and three gamma populations).

The changes in both the direction and the magnitude

of the Type I and II error rates are the only two measures
used in this research to examine the effect on parametric
and nonparametric tests. Two samples with identical sample
size were selected each time in experiment 2 and 3. A
worthwhile study would be to test the effect of using two
samples with different sample sizes in the experiments.

Another challenge to researchers is to explore the
change of regression performance when the number of points
(with unequal and unbalanced distances) steadily increases.
Also, it is essential to investigate the systematic
progression of test performance with a population
distribution that digresses steadily in central location or
dispersion.

Furthermore, prior studies (Stevens 1975; Lodge
1987; Teas 1987) have directed the labeled distance
fluctuations across from individual to individual. It would
be interesting to look into the problem and develop a
within-subject power function to solve within-subject
heterogeneity as well.

APPENDIX

SOURCE CODES OF COMPUTER PROGRAMS

1. EXP1.F (Experiment 1)

```
ccccccccccccccccccccccccccccccccccccccccccccccccc
c         Program name: EXP1.F                            c
c         Programmer  : Andrew H. Chen                    c
c         Date        : 12-20-1993                        c
ccccccccccccccccccccccccccccccccccccccccccccccccc
          integer    ido, index(25), ldpop, ldsamp
          integer    nvar, npop, nrow, nsamp
          integer    i, j
          integer    iseed, nout, nr
          integer    n, npass, m, qpass
          integer    aerrct, berrct, cerrct, derrct
          real       a, b
          real       psum, pmean, ssize, tt
          real       r(10000,1), q(10000,1), samp(25,1)
          real       sum, sumdev, mean, median, stddev, psmsq
          real       pstddev
          real       aerrrate, berrrate, cerrrate, derrrate
          real       acil, acih, bcil, bcih, ccil, ccih, dcil,
     $               dcih
          real       pfq1, pfq2, pfq3, pfq4, pfq5, pfq6, pfq7
          real       pcfq1, pcfq2, pcfq3, pcfq4
          real       pcfq5, pcfq6, pcfq7
          real       sfq1, sfq2, sfq3, sfq4, sfq5, sfq6, sfq7
          real       scfq1, scfq2, scfq3, scfq4
          real       scfq5, scfq6, scfq7
          real       ksd, ksd1, ksd2, ksd3, ksd4, ksd5, ksd6,
     $               ksd7
          real       kscrit, chicrit
          external   rnsrs
          external   rnset, rnun, umach
ccccccccccccccccccccccccccccccccccccccccccccccccc
c         Initialize all variables                        c
ccccccccccccccccccccccccccccccccccccccccccccccccc
          call umach(2, nout)
          a = 16
          b = 2.5
          ssize = 25.0
          nrow = 10000
          nvar = 1
          ldpop = 10000
          nsamp = ssize
          ldsamp = ssize
          ido = 0
          nr = 10000
          ctneg1 = 0
          ctneg2 = 0
          ctneg3 = 0
          ctneu0 = 0
          ctpos1 = 0
          ctpos2 = 0
          ctpos3 = 0
          aerrct = 0
          aerrrate = 0.0
```

```fortran
      berrct = 0
      berrrate = 0.0
      cerrct = 0
      cerrrate = 0.0
      derrct = 0
      derrrate = 0.0
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Generate a uniform population            c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
      call rnset(iseed)
      call rnun(nr, r)
      call sscal(nr, 100.00, r, 1)
      call sadd(nr, 0.00, r, 1)
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Generate a normal population             c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         call rnset(iseed)
c         call rnnoa(nr, r)
c         call sscal(nr, 15.00, r, 1)
c         call sadd(nr, 50.00, r, 1)
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Generate a gamma population              c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         call rnset(iseed)
c         call rngam(nr, a, r)
c         call sscal(nr, b, r, 1)
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Find population mean                      c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
      psum = 0.0
      pmean = 0.0
      pfq1 = 0.0
      pfq2 = 0.0
      pfq3 = 0.0
      pfq4 = 0.0
      pfq5 = 0.0
      pfq6 = 0.0
      pfq7 = 0.0
      pcfq1 = 0.0
      pcfq2 = 0.0
      pcfq3 = 0.0
      pcfq4 = 0.0
      pcfq5 = 0.0
      pcfq6 = 0.0
      pcfq7 = 0.0
      i = 1
      do 0100 i=1,10000
      if (r(i,1) .lt. 14.29) then
          psum = psum + -3
          pfq1 = pfq1 + 1
      elseif (r(i,1) .lt. 28.57) then
          psum = psum + -2
          pfq2 = pfq2 + 1
```

```
         elseif (r(i,1) .lt. 42.86) then
            psum = psum + -1
            pfq3 = pfq3 + 1
         elseif (r(i,1) .lt. 57.14) then
            psum = psum + 0
            pfq4 = pfq4 + 1
         elseif (r(i,1) .lt. 71.43) then
            psum = psum + 1
            pfq5 = pfq5 + 1
         elseif (r(i,1) .lt. 85.71) then
            psum = psum + 2
            pfq6 = pfq6 + 1
         else
            psum = psum + 3
            pfq7 = pfq7 + 1
         endif
0100     continue
         pmean = psum / 10000
         psmsq = 0.0
         pstddev = 0.0
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c        Find population standard deviation and pcfq       c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
         i = 1
         do 0200 i=1,10000
         if (r(i,1) .lt. 14.29) then
            psmsq = psmsq + (-3 - pmean)**2
         elseif (r(i,1) .lt. 28.57) then
            psmsq = psmsq + (-2 - pmean)**2
         elseif (r(i,1) .lt. 42.86) then
            psmsq = psmsq + (-1 - pmean)**2
         elseif (r(i,1) .lt. 57.14) then
            psmsq = psmsq + (0  - pmean)**2
         elseif (r(i,1) .lt. 71.43) then
            psmsq = psmsq + (1  - pmean)**2
         elseif (r(i,1) .lt. 85.71) then
            psmsq = psmsq + (2  - pmean)**2
         else
            psmsq = psmsq + (3  - pmean)**3
         endif
0200     continue
         pstddev = sqrt(psmsq/10000)
         pcfq1 = pfq1 / 10000.0
         pcfq2 = (pfq1+pfq2) / 10000.0
         pcfq3 = (pfq1+pfq2+pfq3) / 10000.0
         pcfq4 = (pfq1+pfq2+pfq3+pfq4) / 10000.0
         pcfq5 = (pfq1+pfq2+pfq3+pfq4+pfq5) / 10000.0
         pcfq6 = (pfq1+pfq2+pfq3+pfq4+pfq5+pfq6) / 10000.0
         pcfq7 = (pfq1+pfq2+pfq3+pfq4+pfq5+pfq6+pfq7) /
     $           10000.0
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c        Find population median                            c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

```
          i = 1
          do 0300 i=1,10000
          if (r(i,1) .lt. 14.29) then
             q(i,1) = -3
          elseif (r(i,1) .lt. 28.57) then
             q(i,1) = -2
          elseif (r(i,1) .lt. 42.86) then
             q(i,1) = -1
          elseif (r(i,1) .lt. 57.14) then
             q(i,1) = 0
          elseif (r(i,1) .lt. 71.43) then
             q(i,1) = 1
          elseif (r(i,1) .lt. 85.71) then
             q(i,1) = 2
          else
             q(i,1) = 3
          endif
0300      continue
          do 0310 qpass = 1, 10000 - 1
             do 0320 m = 1, 10000 - qpass
                if (q(m, 1) .gt. q(m + 1, 1)) then
                   temp = q(m, 1)
                   q(m, 1) = q(m + 1, 1)
                   q(m + 1, 1) = temp
                endif
0320         continue
0310      continue
          pmedian = (q(5000,1) + q(5001,1))/2
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Select a sample from the gamma population       c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
          j = 1
          do 9999 j=1,1000
          call rnsrs(ido,nrow,nvar,r,ldpop,nsamp,npop,
     $    samp,ldsamp,index)
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Map the sample into a score of seven likert-scale   c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
          sfq1 = 0.0
          sfq2 = 0.0
          sfq3 = 0.0
          sfq4 = 0.0
          sfq5 = 0.0
          sfq6 = 0.0
          sfq7 = 0.0
          scfq1 = 0.0
          scfq2 = 0.0
          scfq3 = 0.0
          scfq4 = 0.0
          scfq5 = 0.0
          scfq6 = 0.0
          scfq7 = 0.0
          i = 1
```

```
          do 1000 i=1,ssize
          if (samp(i,1) .lt. 14.29) then
              samp(i,1) = -3
              sfq1 = sfq1 + 1
          elseif (samp(i,1) .lt. 28.57) then
              samp(i,1) = -2
              sfq2 = sfq2 + 1
          elseif (samp(i,1) .lt. 42.86) then
              samp(i,1) = -1
              sfq3 = sfq3 + 1
          elseif (samp(i,1) .lt. 57.14) then
              samp(i,1) = 0
              sfq4 = sfq4 + 1
          elseif (samp(i,1) .lt. 71.43) then
              samp(i,1) = 1
              sfq5 = sfq5 + 1
          elseif (samp(i,1) .lt. 85.71) then
              samp(i,1) = 2
              sfq6 = sfq6 + 1
          else
              samp(i,1) = 3
              sfq7 = sfq7 + 1
          endif
1000      continue
          scfq1 = sfq1 / ssize
          scfq2 = (sfq1+sfq2) / ssize
          scfq3 = (sfq1+sfq2+sfq3) / ssize
          scfq4 = (sfq1+sfq2+sfq3+sfq4) / ssize
          scfq5 = (sfq1+sfq2+sfq3+sfq4+sfq5) / ssize
          scfq6 = (sfq1+sfq2+sfq3+sfq4+sfq5+sfq6) / ssize
          scfq7 = (sfq1+sfq2+sfq3+sfq4+sfq5+sfq6+sfq7) / ssize
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Perform tests                                    c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
          sum = 0.0
          sumdev = 0.0
          mean = 0.0
          stddev = 0.0
          acil = 0.0
          acih = 0.0
          bcil = 0.0
          bcih = 0.0
          ccil = 0.0
          ccih = 0.0
          dcil = 0.0
          dcih = 0.0
          tt = 0.0
          do 2000 i=1,ssize
          sum = sum + samp(i,1)
2000      continue
          mean = sum / ssize
          do 2100 i=1,ssize
          sumdev = sumdev + (samp(i,1) - mean)**2
```

```fortran
2100      continue
          stddev = sqrt(sumdev / (ssize - 1))
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         CI on mean                                         c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
          if (ssize .eq. 10.0 ) then
              tt = 2.262
          endif
          if (ssize .eq. 25.0 ) then
              tt = 2.064
          endif
          if (ssize .gt. 30.0 ) then
              tt = 1.96
          endif
          acil = mean - tt * (stddev / (sqrt(ssize)))
          acih = mean + tt * (stddev / (sqrt(ssize)))
          if (acil .gt. pmean) then
              aerrct = aerrct + 1
          endif
          if (acih .lt. pmean) then
              aerrct = aerrct + 1
          endif
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         CI on median                                       c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
          do 3000 npass = 1, ssize - 1
              do 3100 n = 1, ssize - npass
                  if (samp(n, 1) .gt. samp(n + 1, 1)) then
                      temp = samp(n, 1)
                      samp(n, 1) = samp(n + 1, 1)
                      samp(n + 1, 1) = temp
                  endif
3100          continue
3000      continue
          median = 0.0
          if (ssize .eq. 10.0) then
              median = (samp(5,1) + samp(6,1))/2
          endif
          if (ssize .eq. 25.0) then
              median = samp(13,1)
          endif
          if (ssize .eq. 50.0) then
              median = (samp(25,1) + samp(26,1))/2
          endif
          if (ssize .eq. 100.0) then
              median = (samp(50,1) + samp(51,1))/2
          endif
          tt = 0.0
          if (ssize .eq. 10.0) then
              tt = 2.262
          endif
          if (ssize .eq. 25.0) then
              tt = 2.064
```

```
      endif
      if (ssize .ge. 30.0) then
          tt = 1.96
      endif
      bcil = median - tt * (1.2533)*(stddev /
   $  (sqrt(ssize)))
      bcih = median + tt * (1.2533)*(stddev /
   $  (sqrt(ssize)))
      if (bcil .gt. pmedian) then
          berrct = berrct + 1
      endif
      if (bcih .lt. pmedian) then
          berrct = berrct + 1
      endif
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Kolmogorov-Smirnov Test                       c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
      ksd = 0.0
      ksd1 = 0.0
      ksd2 = 0.0
      ksd3 = 0.0
      ksd4 = 0.0
      ksd5 = 0.0
      ksd6 = 0.0
      ksd7 = 0.0
      kscrit = 0.0
      ksd1 = abs(pcfq1 - scfq1)
      ksd2 = abs(pcfq2 - scfq2)
      ksd3 = abs(pcfq3 - scfq3)
      ksd4 = abs(pcfq4 - scfq4)
      ksd5 = abs(pcfq5 - scfq5)
      ksd6 = abs(pcfq6 - scfq6)
      ksd7 = abs(pcfq7 - scfq7)
      ksd = max(ksd1, ksd2, ksd3, ksd4, ksd5, ksd6, ksd7)
          if (ssize .eq. 10.0) then
          kscrit = .41
      endif
      if (ssize .eq. 25.0) then
          kscrit = .27
      endif
      if (ssize .eq. 50.0) then
          kscrit = .1923
      endif
      if (ssize .eq. 100.0) then
          kscrit = .136
      endif
      if (ksd .gt. kscrit) then
          cerrct = cerrct + 1
      endif
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Chi-Square Test                               c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
      chicrit = 0.0
```

```fortran
      if (ssize .eq. 10.0) then
          chicrit1 = 19.0228
          chicrit2 = 2.70039
      endif
      if (ssize .eq. 25.0) then
          chicrit1 = 39.3641
          chicrit2 = 12.4011
      endif
      if (ssize .eq. 50.0) then
          chicrit1 = 71.4202
          chicrit2 = 32.3574
      endif
      if (ssize .eq. 100.0) then
          chicrit1 = 129.561
          chicrit2 = 74.2219
      endif
      dcil = sqrt(((ssize - 1)*((stddev)**2))/chicrit1)
      dcih = sqrt(((ssize -
     $      1)*((stddev)**2))/chicrit2)
      if (dcil .gt. pstddev) then
          derrct = derrct + 1
      endif
      if (dcih .lt. pstddev) then
          derrct = derrct + 1
      endif
9999  continue
c
      aerrrate = aerrct / 1000.0
      write (nout, 99991) aerrct, aerrrate
99991 format ('    mean: err count = ', i5,
     $ '    err rate = ', f10.4)
c
      berrrate = berrct / 1000.0
      write (nout, 99992) berrct, berrrate
99992 format (' median: err count = ', i5,
     $ '    err rate = ', f10.4)
c
      cerrrate = cerrct / 1000.0
      write (nout, 99993) cerrct, cerrrate
99993 format ('    k-s: err count = ', i5,
     $ '    err rate = ', f10.4)
c
      derrrate = derrct / 1000.0
      write (nout, 99994) derrct, derrrate
99994 format ('    chi: err count = ', i5,
     $ '    err rate = ', f10.4)
      end
```

2. EXP2.F (Experiment 2)

```
cccccccccccccccccccccccccccccccccccccccccccccccccc
c          Program name: EXP2.F                        c
c          Programmer  : Andrew H. Chen                 c
c          Date        : 01-10-1994                    c
cccccccccccccccccccccccccccccccccccccccccccccccccc
           integer     ido, index(50), ldpop, ldsamp
           integer     nvar, npop, nrow, nsamp
           integer     i, j, k
           integer     iseed, nout, nr
           integer     n, npass, m
           integer     aerrct, berrct, cerrct, derrct, eerrct
           integer     tie(14,1)
           real        a, b
           real        pt1, pt2, pt3, pt4, pt5, pt6, pt7
           real        psum, pmean, ssize, tt
           real        r(10000,1)
           real        samp(50,1), absdevo(50,1), absdevn(50,1)
           real        sqrank(50,1)
           real        aerrrate, berrrate, cerrrate, derrrate
           real        eerrrate
           real        acil, acih, dcil, dcih
           real        sum1, sum2, sumdev1, sumdev2, mean1,
     $                 mean2
           real        stddev1, stddev2
           real        lgr, sml, ff, fcrit, var1, var2
           real        sqroot, varp, varp1, varp2
           real        ranksum, ranktie(14,1)
           real        sumsqrk, rsqbar, rsqbarsz, sumr4
           real        srt1, t1crit
           real        rksum
           real        t1, t2, t3, t4, t5, t6, t7
           real        r1, r2, r3, r4, r5, r6, r7
           real        wx, tsum
           real        z, z1, z2, z3, z4
           real        ksd ,dmn
           real        m1, m2, m3, m4, m5, m6, m7
           real        n1, n2, n3, n4, n5, n6, n7
           real        sm1, sm2, sm3, sm4, sm5, sm6, sm7
           real        sn1, sn2, sn3, sn4, sn5, sn6, sn7
           real        d1, d2, d3, d4, d5, d6, d7
           external    rnsrs
           external    rnset, rnun, umach
cccccccccccccccccccccccccccccccccccccccccccccccccc
c          Initialize all variables                    c
cccccccccccccccccccccccccccccccccccccccccccccccccc
           call umach(2, nout)
           a = 16
           b = 2.5
           ssize = 50.0
           nrow = 10000
           nvar = 1
           ldpop = 10000
           nsamp = ssize
```

```
          ldsamp = ssize
          ido = 0
          nr = 10000
          pt1 = -0.999
          pt2 = -0.745
          pt3 = -0.618
          pt4 = -0.49
          pt5 =  0.018
          pt6 =  0.781
          pt7 =  2.053
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Generate a uniform population                 c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
          call rnset(iseed)
          call rnun(nr, r)
          call sscal(nr, 100.00, r, 1)
          call sadd(nr, 0.00, r, 1)
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Generate a normal population                  c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         call rnset(iseed)
c         call rnnoa(nr, r)
c         call sscal(nr, 10.00, r, 1)
c         call sadd(nr, 40.00, r, 1)
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Generate a gamma population                   c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         call rnset(iseed)
c         call rngam(nr, a, r)
c         call sscal(nr, b, r, 1)
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Find population mean                          c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
          psum = 0.0
          pmean = 0.0
          i = 1
          do 0100 i=1,10000
          if (r(i,1) .lt. 14.29) then
             psum = psum + pt1
          elseif (r(i,1) .lt. 28.57) then
             psum = psum + pt2
          elseif (r(i,1) .lt. 42.86) then
             psum = psum + pt3
          elseif (r(i,1) .lt. 57.14) then
             psum = psum + pt4
          elseif (r(i,1) .lt. 71.43) then
             psum = psum + pt5
          elseif (r(i,1) .lt. 85.71) then
             psum = psum + pt6
          else
             psum = psum + pt7
          endif
0100      continue
```

```
        2pmean = psum / 10000
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c       Select 2 samples from a gamma population          c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
        j = 1
        do 9999 j=1,1000
        call rnsrs(ido,nrow,nvar,r,ldpop,nsamp,npop,
     $  samp,ldsamp,index)
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c       Map the sample into one of seven-point likert-scale c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
        i = 1
        t1 = 0.0
        t2 = 0.0
        t3 = 0.0
        t4 = 0.0
        t5 = 0.0
        t6 = 0.0
        t7 = 0.0
        m1 = 0.0
        m2 = 0.0
        m3 = 0.0
        m4 = 0.0
        m5 = 0.0
        m6 = 0.0
        m7 = 0.0
        do 1000 i=1,ssize/2
        if (samp(i,1) .lt. 14.29) then
           samp(i,1) = pt1
           t1 = t1 + 1
           m1 = m1 + 1
        elseif (samp(i,1) .lt. 28.57) then
           samp(i,1) = pt2
           t2 = t2 + 1
           m2 = m2 + 1
        elseif (samp(i,1) .lt. 42.86) then
           samp(i,1) = pt3
           t3 = t3 + 1
           m3 = m3 + 1
        elseif (samp(i,1) .lt. 57.14) then
           samp(i,1) = pt4
           t4 = t4 + 1
           m4 = m4 + 1
        elseif (samp(i,1) .lt. 71.43) then
           samp(i,1) = pt5
           t5 = t5 + 1
           m5 = m5 + 1
        elseif (samp(i,1) .lt. 85.71) then
           samp(i,1) = pt6
           t6 = t6 + 1
           m6 = m6 + 1
        else
           samp(i,1) = pt7
```

```
            t7 = t7 + 1
            m7 = m7 + 1
         endif
1000     continue
c
         i = 1
         n1 = 0.0
         n2 = 0.0
         n3 = 0.0
         n4 = 0.0
         n5 = 0.0
         n6 = 0.0
         n7 = 0.0
         do 1001 i=(ssize/2) + 1, ssize
         if (samp(i,1) .lt. 14.29) then
            samp(i,1) = pt1
            t1 = t1 + 1
            n1 = n1 + 1
         elseif (samp(i,1) .lt. 28.57) then
            samp(i,1) = pt2
            t2 = t2 + 1
            n2 = n2 + 1
         elseif (samp(i,1) .lt. 42.86) then
            samp(i,1) = pt3
            t3 = t3 + 1
            n3 = n3 + 1
         elseif (samp(i,1) .lt. 57.14) then
            samp(i,1) = pt4
            t4 = t4 + 1
            n4 = n4 + 1
         elseif (samp(i,1) .lt. 71.43) then
            samp(i,1) = pt5
            t5 = t5 + 1
            n5 = n5 + 1
         elseif (samp(i,1) .lt. 85.71) then
            samp(i,1) = pt6
            t6 = t6 + 1
            n6 = n6 + 1
         else
            samp(i,1) = pt7
            t7 = t7 + 1
            n7 = n7 + 1
         endif
1001     continue
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c        Confidence Interval on Mean                       c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
         sum1 = 0.0
         sum2 = 0.0
         sumdev1 = 0.0
         sumdev2 = 0.0
         mean1 = 0.0
         mean2 = 0.0
```

```
        stddev1 = 0.0
        stddev2 = 0.0
        sqroot = 0.0
        varp = 0.0
        varp1 = 0.0
        varp2 = 0.0
        cil = 0.0
        cih = 0.0
c
        do 1100 i=1,ssize/2
        sum1 = sum1 + samp(i,1)
1100    continue
c
        do 1200 i=ssize/2 + 1, ssize
        sum2 = sum2 + samp(i,1)
1200    continue
c
        mean1 = sum1 / (ssize/2)
        do 1300 i=1,ssize/2
        sumdev1 = sumdev1 + (samp(i,1) - mean1)**2
1300    continue
        mean2 = sum2 / (ssize/2)
        do 1400 i=ssize/2 + 1, ssize
        sumdev2 = sumdev2 + (samp(i,1) - mean2)**2
1400    continue
        stddev1 = sqrt(sumdev1 / (ssize/2 - 1))
        stddev2 = sqrt(sumdev2 / (ssize/2 - 1))
c
        varp1 = ((ssize/2 - 1)*stddev1**2) +
     $          ((ssize/2 - 1)*stddev2**2)
        varp2 = ssize - 2
        varp = varp1 / varp2
        sqroot = (varp/(ssize/2)) + (varp/(ssize/2))
        tt = 0.0
        if (ssize .eq. 20.0) then
            tt = 2.101
        endif
        if (ssize .eq. 50.0) then
            tt = 2.021
        endif
        if (ssize .gt. 50.0) then
            tt = 1.96
        endif
        acil = (mean1 - mean2) - tt*(sqrt(sqroot))
        acih = (mean1 - mean2) + tt*(sqrt(sqroot))
        if (acil .gt. 0) then
            aerrct = aerrct + 1
        endif
        if (acih .lt. 0) then
            aerrct = aerrct + 1
        endif
```

```
ccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c        Wilcoxon-Mann-Whitney Test                    c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccc
          r1 = 0.0
          r2 = 0.0
          r3 = 0.0
          r4 = 0.0
          r5 = 0.0
          r6 = 0.0
          r7 = 0.0
c
          rksum = 0.0
          if (t1 .gt. 0 ) then
               do 2010 i=1, t1
               rksum = rksum + i
2010           continue
               r1 = rksum / t1
          endif
c
          rksum = 0.0
          if (t2 .gt. 0 ) then
               do 2020 i=t1 + 1, t1 + t2
               rksum = rksum + i
2020           continue
               r2 = rksum / t2
          endif
c
          rksum = 0.0
          if (t3 .gt. 0 ) then
               do 2030 i=t1 + t2 + 1, t1 + t2 + t3
               rksum = rksum + i
2030           continue
               r3 = rksum / t3
          endif
c
          rksum = 0.0
          if (t4 .gt. 0 ) then
               do 2040 i=t1 + t2 + t3 + 1, t1 + t2 + t3 + t4
               rksum = rksum + i
2040           continue
               r4 = rksum / t4
          endif
c
          rksum = 0.0
          if (t5 .gt. 0 ) then
               do 2050 i=t1 + t2 + t3 + t4 + 1,
     $         t1 + t2 + t3 + t4 + t5
               rksum = rksum + i
2050           continue
               r5 = rksum / t5
          endif
c
          rksum = 0.0
```

```fortran
            if (t6 .gt. 0 ) then
                  do 2060 i=t1 + t2 + t3 + t4 + t5 + 1,
     $                t1 + t2 + t3 + t4 + t5 + t6
                  rksum = rksum + i
2060              continue
                  r6 = rksum / t6
            endif
c
            rksum = 0.0
            if (t7 .gt. 0 ) then
                  do 2070 i=t1 + t2 + t3 + t4 + t5 + t6 + 1,
     $                t1 + t2 + t3 + t4 + t5 + t6 + t7
                  rksum = rksum + i
2070              continue
                  r7 = rksum / t7
            endif
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Find wx                                         c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
            wx = 0.0
            do 2100 i=1, ssize/2
            if (samp(i,1) .eq. pt1)      then
               wx = wx + r1
            elseif (samp(i,1) .eq. pt2) then
               wx = wx + r2
            elseif (samp(i,1) .eq. pt3) then
               wx = wx + r3
            elseif (samp(i,1) .eq. pt4) then
               wx = wx + r4
            elseif (samp(i,1) .eq. pt5) then
               wx = wx + r5
            elseif (samp(i,1) .eq. pt6) then
               wx = wx + r6
            else
               wx = wx + r7
            endif
2100        continue
c
            tsum = 0.0
            tsum = ((t1**3-t1)+(t2**3-t2)+(t3**3-t3)+(t4**3-t4)+
     $           (t5**3-t5)+(t6**3-t6)+(t7**3-t7))/12
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Calculate z statistics                          c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
            z1 = 0.0
            z2 = 0.0
            z3 = 0.0
            z4 = 0.0
            z  = 0.0
            z1=wx - (ssize/2)*(ssize + 1)/2
            z3=(ssize/2)**2/(ssize*(ssize-1))
            z4=(ssize**3-ssize)/12-tsum
            z2= sqrt(z3*z4)
```

```
      z = z1/z2
      if ((z .gt. 1.96) .or. (z .lt. -1.96)) then
          berrct = berrct + 1
      endif
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         F Test                                          c
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
      var1 = 0.0
      var2 = 0.0
      lgr = 0.0
      sml = 0.0
      ff = 0.0
      fcrit = 0.0
c
      var1 = stddev1**2
      var2 = stddev2**2
      lgr = max(var1,var2)
      sml = min(var1,var2)
      ff = lgr/sml
      if (ssize .eq. 20.0) then
          fcrit = 4.03
      endif
      if (ssize .eq. 50.0) then
          fcrit = 2.27
      endif
      if (ssize .eq. 100.0) then
          fcrit = 1.78
      endif
      if (ssize .eq. 200.0) then
          fcrit = 1.51
      endif
      if (ff .gt. fcrit) then
          cerrct = cerrct + 1
      endif
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Squared Ranks Test                               c
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
      do 4000 i=1,ssize/2
      absdevo(i,1) = abs(samp(i,1) - mean1)
      absdevn(i,1) = abs(samp(i,1) - mean1)
4000      continue
      do 4100 i=ssize/2 + 1, ssize
      absdevo(i,1) = abs(samp(i,1) - mean2)
      absdevn(i,1) = abs(samp(i,1) - mean2)
4100      continue
      do 4200 npass = 1, ssize - 1
          do 4210 n = 1, ssize - npass
             if (absdevn(n, 1) .gt. absdevn(n + 1, 1)) then
                 temp = absdevn(n, 1)
                 absdevn(n, 1) = absdevn(n + 1, 1)
                 absdevn(n + 1, 1) = temp
             endif
4210         continue
```

```
4200    continue
c
        k = 1
        tie(k,1) = 1
        ranksum = 1
        do 4300 i=1, ssize - 1
            if (absdevn(i,1) .ne. absdevn(i+1,1)) then
                if (i .eq. 1) then
                    ranktie(1,1) = ranksum / i
                else
                    ranktie(k,1) = (ranksum / tie(k,1))**2
                endif
                do 4400 l = 1,ssize
                    if (absdevo(l,1) .eq. absdevn(i,1)) then
                        sqrank(l,1) = ranktie(k,1)
                    endif
4400            continue
                ranksum = i + 1
                k = k + 1
                tie(k,1) = 1
                if (i .eq. (ssize - 1)) then
                    ranktie(k,1) = (ranksum / tie(k,1))**2
                    do 4500 l = 1,ssize
                        if (absdevo(l,1) .eq. absdevn(i+1,1))
                            then
                            sqrank(l,1) = ranktie(k,1)
                        endif
4500                continue
                endif
            else
                tie(k,1) = tie(k,1) + 1
                ranksum = ranksum + i + 1
                if (i .eq. (ssize - 1)) then
                    ranktie(k,1) = (ranksum / tie(k,1))**2
                    do 4600 l = 1,ssize
                        if (absdevo(l,1) .eq. absdevn(i,1))
then
                            sqrank(l,1) = ranktie(k,1)
                        endif
4600                continue
                endif
            endif
4300    continue
        sumsqrk = 0.0
        rsqbar = 0.0
        sumr4 = 0.0
        do 4700 m = 1,ssize/2
            sumsqrk = sumsqrk + sqrank(m,1)
4700    continue
        do 4800 m = 1,ssize
        rsqbar = rsqbar + sqrank(m,1)
        sumr4 = sumr4 + (sqrank(m,1))**2
4800    continue
```

```
          rsqbarsz = rsqbar / ssize
          srt1 = ((sumsqrk -
     $    (ssize/2)*rsqbarsz)/(sqrt(((((ssize/2)**2)/
     $    (ssize*(ssize-1))*sumr4)-(((ssize/2)**2)/(ssize-1)*
     $    (rsqbarsz**2))))))
c
          t1crit = 0.0
          if (ssize .eq. 20.0) then
              t1crit = 2.101
          endif
          if (ssize .eq. 50.0) then
              t1crit = 2.011
          endif
          if (ssize .gt. 50.0) then
              t1crit = 1.96
          endif
          dcil = -t1crit
          dcih = t1crit
          if (dcil .gt. srt1) then
              derrct = derrct + 1
          endif
          if (dcih .lt. srt1) then
              derrct = derrct + 1
          endif
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c          Kolmogorov-Smirnov Test                          c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
          sm1 = m1/(ssize/2)
          sm2 = m1/(ssize/2)+m2/(ssize/2)
          sm3 = m1/(ssize/2)+m2/(ssize/2)+m3/(ssize/2)
          sm4 =
     $    m1/(ssize/2)+m2/(ssize/2)+m3/(ssize/2)+m4/(ssize/2)
     $    sm5 = m1/(ssize/2)+m2/(ssize/2)+m3/(ssize/2)+
     $          m4/(ssize/2)+m5/(ssize/2)
          sm6 = m1/(ssize/2)+m2/(ssize/2)+m3/(ssize/2)+
     $          m4/(ssize/2)+m5/(ssize/2)+m6/(ssize/2)
          sm7 = m1/(ssize/2)+m2/(ssize/2)+m3/(ssize/2)+
     $
     $    m4/(ssize/2)+m5/(ssize/2)+m6/(ssize/2)+m7/(ssize/2)
          sn1 = n1/(ssize/2)
          sn2 = n1/(ssize/2)+n2/(ssize/2)
          sn3 = n1/(ssize/2)+n2/(ssize/2)+n3/(ssize/2)
          sn4 =
     $    n1/(ssize/2)+n2/(ssize/2)+n3/(ssize/2)+n4/(ssize/2)
          sn5 = n1/(ssize/2)+n2/(ssize/2)+n3/(ssize/2)+
     $          n4/(ssize/2)+n5/(ssize/2)
          sn6 = n1/(ssize/2)+n2/(ssize/2)+n3/(ssize/2)+
     $          n4/(ssize/2)+n5/(ssize/2)+n6/(ssize/2)
          sn7 = n1/(ssize/2)+n2/(ssize/2)+n3/(ssize/2)+
     $          n4/(ssize/2)+n5/(ssize/2)+
     $          n6/(ssize/2)+n7/(ssize/2)
c
          d1 = abs(sm1-sn1)
```

```
           d2 = abs(sm2-sn2)
           d3 = abs(sm3-sn3)
           d4 = abs(sm4-sn4)
           d5 = abs(sm5-sn5)
           d6 = abs(sm6-sn6)
           d7 = abs(sm7-sn7)
c
           dmn = 0.0
           if (ssize .gt. 50.0) then
                 dmn = max(d1,d2,d3,d4,d5,d6,d7)
              else
                 dmn =
      $          (ssize/2)*(ssize/2)*
      $          (max(d1,d2,d3,d4,d5,d6,d7))
           endif
ccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c          Calculate ksd statistics                        c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccc
           ksd = 0.0
           if (ssize .eq. 20.0) then
               ksd = 70
           endif
           if (ssize .eq. 50.0) then
               ksd = 250
           endif
           if (ssize .gt. 50.0) then
               ksd = 1.36*(sqrt(ssize/((ssize/2)*(ssize/2))))
           endif
           if (dmn .ge. ksd) then
               eerrct = eerrct + 1
           endif
9999       continue
c
           aerrrate = aerrct / 1000.0
           write (nout, 99991) aerrct, aerrrate
99991      format ('   Mean: err count = ', i5,
      $    '   err rate = ', f10.4)
c
           berrrate = berrct / 1000.0
           write (nout, 99992) berrct, berrrate
99992      format (' Man-Wh: err count = ', i5,
      $    '   err rate = ', f10.4)
c
           cerrrate = cerrct / 1000.0
           write (nout, 99993) cerrct, cerrrate
99993      format ('      F: err count = ', i5,
      $    '   err rate = ', f10.4)
c
           derrrate = derrct / 1000.0
           write (nout, 99994) derrct, derrrate
99994      format (' Sq Rnk: err count = ', i5,
      $    '   err rate = ', f10.4)
c
```

```
        eerrrate = eerrct / 1000.0
        write (nout, 99995) eerrct, eerrrate
99995   format ('     K-S: err count = ', i5,
   $    '    err rate = ', f10.4)
        end
```

3. EXP3.F (Experiment 3)

```
ccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Program name: EXP3.F                        c
c         Programmer  : Andrew H. Chen                c
c         Date        : 01-22-1994                    c
ccccccccccccccccccccccccccccccccccccccccccccccccccc
          integer    ido, index(50), ldpop, ldsamp
          integer    nvar, npop, nrow, nsamp
          integer    i, j, k
          integer    iseed, nout, nr
          integer    n, npass, m
          integer    aerrct, berrct, cerrct, derrct, eerrct
          integer    tie(14,1)
          real       a1, b1, a2, b2
          real       pt1, pt2, pt3, pt4, pt5, pt6, pt7
          real       ssize, tt
          real       p1(10000,1), p2(10000,1)
          real       samp1(50,1), samp2(50,1)
          real       samp(100,1)
          real       sqrank(100,1), absdevo(100,1),
      $              absdevn(100,1)
          real       aerrrate, berrrate, cerrrate, derrrate
          real       eerrrate
          real       acil, acih, dcil, dcih
          real       sum1, sum2, sumdev1, sumdev2, mean1,
      $              mean2
          real       stddev1, stddev2
          real       lgr, sml, ff, fcrit, var1, var2
          real       sqroot, varp, varp1, varp2
          real       ranksum, ranktie(14,1)
          real       sumsqrk, rsqbar, rsqbarsz, sumr4
          real       srt1, t1crit
          real       rksum
          real       t1, t2, t3, t4, t5, t6, t7
          real       r1, r2, r3, r4, r5, r6, r7
          real       wx, tsum
          real       z, z1, z2, z3, z4
          real       ksd ,dmn
          real       m1, m2, m3, m4, m5, m6, m7
          real       n1, n2, n3, n4, n5, n6, n7
          real       sm1, sm2, sm3, sm4, sm5, sm6, sm7
          real       sn1, sn2, sn3, sn4, sn5, sn6, sn7
          real       d1, d2, d3, d4, d5, d6, d7
          external   rnsrs
          external   rnset, rnun, umach
ccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Initialize all variables                    c
ccccccccccccccccccccccccccccccccccccccccccccccccccc
          call umach(2, nout)
          a1 = 16
          b1 = 2.5
          a2 = 16
          b2 = 2.5
          ssize = 100.0
```

```
          nrow = 10000
          nvar = 1
          ldpop = 10000
          nsamp = ssize/2
          ldsamp = ssize/2
          ido = 0
          nr = 10000
          pt1 = -1.249
          pt2 = -1.07
          pt3 = -0.892
          pt4 =  0.0
          pt5 =  0.892
          pt6 =  1.07
          pt7 =  1.249
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Generate a uniform population                  c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
          call rnset(iseed)
          call rnun(nr, p1)
          call sscal(nr, 100.00, p1, 1)
          call sadd(nr, 0.00, p1, 1)
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Generate a normal population                   c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
          call rnset(iseed)
          call rnnoa(nr, p1)
          call sscal(nr, 10.00, p1, 1)
          call sadd(nr, 40.00, p1, 1)
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Generate a gamma population                    c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c          call rnset(iseed)
c          call rngam(nr, a1, p2)
c          call sscal(nr, b1, p2, 1)
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Select 1 sample from a uniform population      c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
          j = 1
          do 9999 j=1,1000
          call rnsrs(ido,nrow,nvar,p1,ldpop,nsamp,npop,
     $    samp1,ldsamp,index)
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Select 1 sample from a normal population       c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
          call rnsrs(ido,nrow,nvar,p2,ldpop,nsamp,npop,
     $    samp2,ldsamp,index)
c
          do 0200 i=1,ssize/2
          samp(i,1) = samp1(i,1)
0200      continue
          do 0300 i=ssize/2 + 1, ssize
          samp(i,1) = samp2(i - ssize/2,1)
0300      continue
```

```
cccccccccccccccccccccccccccccccccccccccccccccccccccccc
c       Map the sample into one of seven-point likert-scale c
cccccccccccccccccccccccccccccccccccccccccccccccccccccc
          i = 1
          t1 = 0.0
          t2 = 0.0
          t3 = 0.0
          t4 = 0.0
          t5 = 0.0
          t6 = 0.0
          t7 = 0.0
          m1 = 0.0
          m2 = 0.0
          m3 = 0.0
          m4 = 0.0
          m5 = 0.0
          m6 = 0.0
          m7 = 0.0
          do 1000 i=1,ssize/2
          if (samp(i,1) .lt. 14.29) then
             samp(i,1) = pt1
             t1 = t1 + 1
             m1 = m1 + 1
          elseif (samp(i,1) .lt. 28.57) then
             samp(i,1) = pt2
             t2 = t2 + 1
             m2 = m2 + 1
          elseif (samp(i,1) .lt. 42.86) then
             samp(i,1) = pt3
             t3 = t3 + 1
             m3 = m3 + 1
          elseif (samp(i,1) .lt. 57.14) then
             samp(i,1) = pt4
             t4 = t4 + 1
             m4 = m4 + 1
          elseif (samp(i,1) .lt. 71.43) then
             samp(i,1) = pt5
             t5 = t5 + 1
             m5 = m5 + 1
          elseif (samp(i,1) .lt. 85.71) then
             samp(i,1) = pt6
             t6 = t6 + 1
             m6 = m6 + 1
          else
             samp(i,1) = pt7
             t7 = t7 + 1
             m7 = m7 + 1
          endif
1000      continue
c
          i = 1
          n1 = 0.0
          n2 = 0.0
```

```
          n3 = 0.0
          n4 = 0.0
          n5 = 0.0
          n6 = 0.0
          n7 = 0.0
          do 1001 i=(ssize/2) + 1, ssize
          if (samp(i,1) .lt. 14.29) then
              samp(i,1) = pt1
              t1 = t1 + 1
              n1 = n1 + 1
          elseif (samp(i,1) .lt. 28.57) then
              samp(i,1) = pt2
              t2 = t2 + 1
              n2 = n2 + 1
          elseif (samp(i,1) .lt. 42.86) then
              samp(i,1) = pt3
              t3 = t3 + 1
              n3 = n3 + 1
          elseif (samp(i,1) .lt. 57.14) then
              samp(i,1) = pt4
              t4 = t4 + 1
              n4 = n4 + 1
          elseif (samp(i,1) .lt. 71.43) then
              samp(i,1) = pt5
              t5 = t5 + 1
              n5 = n5 + 1
          elseif (samp(i,1) .lt. 85.71) then
              samp(i,1) = pt6
              t6 = t6 + 1
              n6 = n6 + 1
          else
              samp(i,1) = pt7
              t7 = t7 + 1
              n7 = n7 + 1
          endif
1001      continue
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Confidence Interval on Mean                    c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
          sum1 = 0.0
          sum2 = 0.0
          sumdev1 = 0.0
          sumdev2 = 0.0
          mean1 = 0.0
          mean2 = 0.0
          stddev1 = 0.0
          stddev2 = 0.0
          sqroot = 0.0
          varp = 0.0
          varp1 = 0.0
          varp2 = 0.0
          cil = 0.0
          cih = 0.0
```

```
c
        do 1100 i=1,ssize/2
        sum1 = sum1 + samp(i,1)
1100    continue
c
        do 1200 i=ssize/2 + 1, ssize
        sum2 = sum2 + samp(i,1)
1200    continue
c
        mean1 = sum1 / (ssize/2)
        do 1300 i=1,ssize/2
        sumdev1 = sumdev1 + (samp(i,1) - mean1)**2
1300    continue
        mean2 = sum2 / (ssize/2)
        do 1400 i=ssize/2 + 1, ssize
        sumdev2 = sumdev2 + (samp(i,1) - mean2)**2
1400    continue
        stddev1 = sqrt(sumdev1 / (ssize/2 - 1))
        stddev2 = sqrt(sumdev2 / (ssize/2 - 1))
c
        varp1 = ((ssize/2 - 1)*stddev1**2) +
     $          ((ssize/2 - 1)*stddev2**2)
        varp2 = ssize - 2
        varp = varp1 / varp2
        sqroot = (varp/(ssize/2)) + (varp/(ssize/2))
        tt = 0.0
        if (ssize .eq. 20.0) then
            tt = 2.101
        endif
        if (ssize .eq. 50.0) then
            tt = 2.021
        endif
        if (ssize .gt. 50.0) then
            tt = 1.96
        endif
        acil = (mean1 - mean2) - tt*(sqrt(sqroot))
        acih = (mean1 - mean2) + tt*(sqrt(sqroot))
        if (acil .gt. 0) then
            aerrct = aerrct + 1
        endif
        if (acih .lt. 0) then
            aerrct = aerrct + 1
        endif
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c       Wilcoxon-Mann-Whitney Test                         c
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
        r1 = 0.0
        r2 = 0.0
        r3 = 0.0
        r4 = 0.0
        r5 = 0.0
        r6 = 0.0
        r7 = 0.0
```

```
c
        rksum = 0.0
        if (t1 .gt. 0 ) then
             do 2010 i=1, t1
             rksum = rksum + i
2010         continue
             r1 = rksum / t1
        endif
c
        rksum = 0.0
        if (t2 .gt. 0 ) then
             do 2020 i=t1 + 1, t1 + t2
             rksum = rksum + i
2020         continue
             r2 = rksum / t2
        endif
c
        rksum = 0.0
        if (t3 .gt. 0 ) then
             do 2030 i=t1 + t2 + 1, t1 + t2 + t3
             rksum = rksum + i
2030         continue
             r3 = rksum / t3
        endif
c
        rksum = 0.0
        if (t4 .gt. 0 ) then
             do 2040 i=t1 + t2 + t3 + 1, t1 + t2 + t3 + t4
             rksum = rksum + i
2040         continue
             r4 = rksum / t4
        endif
c
        rksum = 0.0
        if (t5 .gt. 0 ) then
             do 2050 i=t1 + t2 + t3 + t4 + 1,
     $       t1 + t2 + t3 + t4 + t5
             rksum = rksum + i
2050         continue
             r5 = rksum / t5
        endif
c
        rksum = 0.0
        if (t6 .gt. 0 ) then
             do 2060 i=t1 + t2 + t3 + t4 + t5 + 1,
     $       t1 + t2 + t3 + t4 + t5 + t6
             rksum = rksum + i
2060         continue
             r6 = rksum / t6
        endif
c
        rksum = 0.0
        if (t7 .gt. 0 ) then
```

```
            do 2070 i=t1 + t2 + t3 + t4 + t5 + t6 + 1,
     $      t1 + t2 + t3 + t4 + t5 + t6 + t7
            rksum = rksum + i
2070        continue
            r7 = rksum / t7
        endif
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c       Find wx                                             c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
        wx = 0.0
        do 2100 i=1, ssize/2
        if (samp(i,1) .eq. pt1)      then
            wx = wx + r1
        elseif (samp(i,1) .eq. pt2) then
            wx = wx + r2
        elseif (samp(i,1) .eq. pt3) then
            wx = wx + r3
        elseif (samp(i,1) .eq. pt4) then
            wx = wx + r4
        elseif (samp(i,1) .eq. pt5) then
            wx = wx + r5
        elseif (samp(i,1) .eq. pt6) then
            wx = wx + r6
        else
            wx = wx + r7
        endif
2100        continue
c
        tsum = 0.0
        tsum = ((t1**3-t1)+(t2**3-t2)+(t3**3-t3)+(t4**3-t4)+
     $  (t5**3-t5)+(t6**3-t6)+(t7**3-t7))/12
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c       Calculate z statistics                             c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
        z1 = 0.0
        z2 = 0.0
        z3 = 0.0
        z4 = 0.0
        z  = 0.0
        z1=wx - (ssize/2)*(ssize + 1)/2
        z3=(ssize/2)**2/(ssize*(ssize-1))
        z4=(ssize**3-ssize)/12-tsum
        z2= sqrt(z3*z4)
        z = z1/z2
        if ((z .gt. 1.96) .or. (z .lt. -1.96)) then
            berrct = berrct + 1
        endif

ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c       F Test                                             c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
        var1 = 0.0
        var2 = 0.0
```

```
          lgr = 0.0
          sml = 0.0
          ff = 0.0
          fcrit = 0.0
c
          var1 = stddev1**2
          var2 = stddev2**2
          lgr = max(var1,var2)
          sml = min(var1,var2)
          ff = lgr/sml
          if (ssize .eq. 20.0) then
              fcrit = 4.03
          endif
          if (ssize .eq. 50.0) then
              fcrit = 2.27
          endif
          if (ssize .eq. 100.0) then
              fcrit = 1.78
          endif
          if (ssize .eq. 200.0) then
              fcrit = 1.51
          endif
          if (ff .gt. fcrit) then
              cerrct = cerrct + 1
          endif
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Squared Ranks Test                               c
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
          do 4000 i=1,ssize/2
          absdevo(i,1) = abs(samp(i,1) - mean1)
          absdevn(i,1) = abs(samp(i,1) - mean1)
4000      continue
          do 4100 i=ssize/2 + 1, ssize
          absdevo(i,1) = abs(samp(i,1) - mean2)
          absdevn(i,1) = abs(samp(i,1) - mean2)
4100      continue
          do 4200 npass = 1, ssize - 1
             do 4210 n = 1, ssize - npass
                if (absdevn(n, 1) .gt. absdevn(n + 1, 1)) then
                   temp = absdevn(n, 1)
                   absdevn(n, 1) = absdevn(n + 1, 1)
                   absdevn(n + 1, 1) = temp
                endif
4210         continue
4200      continue
c
          k = 1
          tie(k,1) = 1
          ranksum = 1
          do 4300 i=1, ssize - 1
             if (absdevn(i,1) .ne. absdevn(i+1,1)) then
                if (i .eq. 1) then
                    ranktie(1,1) = ranksum / i
```

```
                  else
                     ranktie(k,1) = (ranksum / tie(k,1))**2
               endif
               do 4400 l = 1,ssize
                  if (absdevo(l,1) .eq. absdevn(i,1)) then
                     sqrank(l,1) = ranktie(k,1)
                  endif
4400           continue
               ranksum = i + 1
               k = k + 1
               tie(k,1) = 1
               if (i .eq. (ssize - 1)) then
                  ranktie(k,1) = (ranksum / tie(k,1))**2
                  do 4500 l = 1,ssize
                     if (absdevo(l,1) .eq. absdevn(i+1,1))
     $                  then
                        sqrank(l,1) = ranktie(k,1)
                     endif
4500              continue
               endif
             else
               tie(k,1) = tie(k,1) + 1
               ranksum = ranksum + i + 1
               if (i .eq. (ssize - 1)) then
                  ranktie(k,1) = (ranksum / tie(k,1))**2
                  do 4600 l = 1,ssize
                     if (absdevo(l,1) .eq. absdevn(i,1))
     $                  then
                        sqrank(l,1) = ranktie(k,1)
                     endif
4600              continue
               endif
             endif
         endif
4300     continue
         sumsqrk = 0.0
         rsqbar = 0.0
         sumr4 = 0.0
         do 4700 m = 1,ssize/2
            sumsqrk = sumsqrk + sqrank(m,1)
4700     continue
         do 4800 m = 1,ssize
         rsqbar = rsqbar + sqrank(m,1)
         sumr4 = sumr4 + (sqrank(m,1))**2
4800     continue
         rsqbarsz = rsqbar / ssize
         srt1 = ((sumsqrk -
     $   (ssize/2)*rsqbarsz)/(sqrt(((((ssize/2)**2)/
     $   (ssize*(ssize-1))*sumr4)-(((ssize/2)**2)/(ssize-1)*
     $   (rsqbarsz**2))))))
c
         t1crit = 0.0
         if (ssize .eq. 20.0) then
             t1crit = 2.101
```

```
      endif
      if (ssize .eq. 50.0) then
          t1crit = 2.011
      endif
      if (ssize .gt. 50.0) then
          t1crit = 1.96
      endif
      dcil = -t1crit
      dcih = t1crit
      if (dcil .gt. srt1) then
          derrct = derrct + 1
      endif
      if (dcih .lt. srt1) then
          derrct = derrct + 1
      endif
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c          Kolmogorov-Smirnov Test                       c
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
      sm1 = m1/(ssize/2)
      sm2 = m1/(ssize/2)+m2/(ssize/2)
      sm3 = m1/(ssize/2)+m2/(ssize/2)+m3/(ssize/2)
      sm4 =
    $ m1/(ssize/2)+m2/(ssize/2)+m3/(ssize/2)+m4/(ssize/2)
    $ sm5 = m1/(ssize/2)+m2/(ssize/2)+m3/(ssize/2)+
    $       m4/(ssize/2)+m5/(ssize/2)
      sm6 = m1/(ssize/2)+m2/(ssize/2)+m3/(ssize/2)+
    $       m4/(ssize/2)+m5/(ssize/2)+m6/(ssize/2)
      sm7 = m1/(ssize/2)+m2/(ssize/2)+m3/(ssize/2)+
    $
    $ m4/(ssize/2)+m5/(ssize/2)+m6/(ssize/2)+m7/(ssize/2)
      sn1 = n1/(ssize/2)
      sn2 = n1/(ssize/2)+n2/(ssize/2)
      sn3 = n1/(ssize/2)+n2/(ssize/2)+n3/(ssize/2)
      sn4 =
    $ n1/(ssize/2)+n2/(ssize/2)+n3/(ssize/2)+n4/(ssize/2)
      sn5 = n1/(ssize/2)+n2/(ssize/2)+n3/(ssize/2)+
    $       n4/(ssize/2)+n5/(ssize/2)
      sn6 = n1/(ssize/2)+n2/(ssize/2)+n3/(ssize/2)+
    $       n4/(ssize/2)+n5/(ssize/2)+n6/(ssize/2)
      sn7 = n1/(ssize/2)+n2/(ssize/2)+n3/(ssize/2)+
    $       n4/(ssize/2)+n5/(ssize/2)+
    $       n6/(ssize/2)+n7/(ssize/2) c
      d1 = abs(sm1-sn1)
      d2 = abs(sm2-sn2)
      d3 = abs(sm3-sn3)
      d4 = abs(sm4-sn4)
      d5 = abs(sm5-sn5)
      d6 = abs(sm6-sn6)
      d7 = abs(sm7-sn7)
c
      dmn = 0.0
      if (ssize .gt. 50.0) then
          dmn = max(d1,d2,d3,d4,d5,d6,d7)
```

```
          else
              dmn =
   $          (ssize/2)*(ssize/2)*
   $          (max(d1,d2,d3,d4,d5,d6,d7))
          endif
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Calculate ksd statistics                         c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
          ksd = 0.0
          if (ssize .eq. 20.0) then
              ksd = 70
          endif
          if (ssize .eq. 50.0) then
              ksd = 250
          endif
          if (ssize .gt. 50.0) then
              ksd = 1.36*(sqrt(ssize/((ssize/2)*(ssize/2))))
          endif
          if (dmn .lt. ksd) then
              eerrct = eerrct + 1
          endif
9999      continue
c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         Display error rate                               c
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c
          aerrrate = aerrct / 1000.0
          write (nout, 99991) aerrct, aerrrate
99991     format ('   Mean: err count = ', i5,
   $    '  Ierr rate = ', f10.4)
c
          berrrate = berrct / 1000.0
          write (nout, 99992) berrct, berrrate
99992     format (' Man-Wh: err count = ', i5,
   $    '  Ierr rate = ', f10.4)
c
          cerrrate = cerrct / 1000.0
          write (nout, 99993) cerrct, cerrrate
99993     format ('      F: err count = ', i5,
   $    '  Ierr rate = ', f10.4)
c
          derrrate = derrct / 1000.0
          write (nout, 99994) derrct, derrrate
99994     format (' Sq Rnk: err count = ', i5,
   $    '  Ierr rate = ', f10.4)
c
          eerrrate = eerrct / 1000.0
          write (nout, 99995) eerrct, eerrrate
99995     format ('    K-S: err count = ', i5,
   $    '  IIerr rate = ', f10.4)
c
c
```

```
cccccccccccccccccccccccccccccccccccccccccccccccccccccc
c         End of program                              c
cccccccccccccccccccccccccccccccccccccccccccccccccccccc
          end
```

REFERENCES

Agresti, A. 1984. <u>Analysis of ordinal categorical data</u>. New
    York: John Wiley & Sons.

_____. 1990. <u>Categorical data analysis</u>. New York: John
    Wiley & Sons.

Anderson, Carl R., and Carl P. Zeithaml. 1984. Stage of the
    product life cycle, business strategy and business
    performance. <u>Academy of Management Journal</u> 27 (March):
    25-41.

Anderson, Norman H. 1961. Scales and statistics: Parametric
    and nonparametric. <u>Psychological Bulletin</u> (July):
    305-16.

Andrews, F. M. 1984. Construct validity and error components
    of survey measures. <u>Public Opinion Quarterly</u> 48
    (Summer): 409-42.

Baker, B. O., C. D. Hardyck, and L. F. Pertinovich. 1966.
    Weak measurements vs. strong statistics: An empirical
    critique of S. S. Stevens' proscriptions on statistics.
    <u>Educational and Psychological Measurement</u> 26 (Summer):
    291-309.

Barkakati, Nabajyoti. 1990. <u>Turbo C++ Bible</u>. Carmel, IN: The
    Waite Group.

Bartram, Peter, and David Yielding. 1973. The development of
    an empirical method of selecting phrases used in verbal
    rating scales: A report on a recent experiment.
    <u>Journal of the Market Research Society</u> 15 (July):
    151-56.

Bolanowski, Stanley J., and George A. Gescheider. 1991.
    <u>Ratio scaling of psychological magnitude</u>. ed. Erlbaum,
    Hillsdale. Breiman, L. and J. Friedman. 1985.
    Estimating optimal transformations for multiple
    regression and correlation. <u>Journal of the American
    Statistical Association</u> 80 (September): 580-98.

Bollen, K. A., and K. H. Barb. 1981. Pearson's R and
    coarsely categorized measures. <u>American Sociological
    Review</u> 46 (April): 232-239.

Borgatta, E. F. 1968. My student, the purist: A lament. _Sociological Quarterly_ 9 (March): 29-34.

Box, George E. P., William G. Hunter, and J. Stuart Hunter. 1978. _Statistics for experimenters_. New York: John Wiley & Sons.

Brown, Gordon, Tony Copeland, and Maurice Millward. 1973. Monadic testing of new product: An old problem and some partial solutions. _Journal of the Market Research Society_ 15 (April): 112-31.

Bunger, R. C. 1988. Derivation of probability density functions for the relative differences in the standard and Poor's 100 stock index. Ph.D. diss. University of North Texas, Denton.

Burdick, Richard K., and Franklin A. Graybill. 1992. _Confidence intervals on variance components_. New York: Marcel Dekker.

Carr, Gregory J., Kerry B. Hafner, and Gary G. Koch. 1989. Analysis of rank measures of association for ordinal data from longitudinal studies. _Journal of the American Statistical Association_ 84 (September): 797-804.

Champion, D. J. 1967. Some observations on measurement and statistics. _Social Forces_ 46 (August): 541-44.

Churchill, G. A., and J. P. Peter. 1984. Research design effects on the reliability of rating scales. _Journal of Marketing Research_ 21 (November): 365-66.

Cliff, Norman. 1959. Adverbs as multipliers. _The Psychological Review_ 66 (April): 27-45.

Conover, W. J. 1980. _Practical nonparametric statistics_. 2d ed. New York: John Wiley & Sons.

Conover, W. J., and R. L. Iman. 1978. Some exact tables for the squared ranks test. _Communications in Statistics_ B7 (October): 491-513.

Coombs, C. H. 1952. A theory of psychological scaling. _Bulletin of Engineering Research Institute No. 34_ Ann Arbor: University of Michigan.

Cox, Eli P. 1980. The optimal number of response alternatives for a scale: A review. _Journal of Marketing Research_ 17 (November): 407-22.

Crask, Melvin R., and Richard J. Fox. 1987. An exploration
    of the interval properties of three commonly used
    marketing research scales: A magnitude estimation
    approach. Journal of the Market Research Society 24
    (July): 317-39.

Dixon, W. J., and F. J. Massey. 1957. Introduction to
    statistics analysis. 2d ed. New York: McGraw-Hill.

Duran, B. S. 1976. A survey of nonparametric tests for
    scale. Communications in Statistics A5 (October):
    1287-1312.

Figlio, R. M. 1976. The seriousness of offenses: An
    evaluation by offenders and nonoffenders. Journal
    of Criminal Law and Criminology 66 (June): 189-200.

_____. 1978. The national survey of crime severity:
    Result of the pretest. Monograph. Philadelphia:
    University of Pennsylvania, Department of Criminology.

Gaito, J. 1980. Measurement scales and statistics:
    Resurgence of an old misconception. Psychological
    Bulletin 87 (May): 564-7.

Galbraith, Craig, and Dan E. Schendel. 1983. An empirical
    analysis of strategy types. Strategic Management
    Journal 4 (April): 153-173.

Garner, W. R. 1960. Rating scales, discriminability, and
    information transmission. Psychological Review 67
    (November): 343-52.

Gautam, Shiva Prasad. 1992. Applications of the t, $T_2$ and F
    statistics to ordinal categorical data. Ph.D. Diss.
    University of Texas, Dallas.

Goodman, L. A. 1979. Simple models for the analysis of
    association in cross-classifications having ordered
    categories. Journal of the American Statistical
    Association 74 (September): 537-52.

_____. 1983. The analysis of dependence in cross-
    classifications having ordered categories, using
    log-linear models for frequencies and log-linear models
    for odds. Biometrics 39 (March): 149-60.

Green, Paul E., and Vithala R. Rao. 1970. Rating scales and
    information recovery: How many scales and response
    categories to use? Journal of Marketing 34 (July):
    33-39.

Gregoire, T. G., and B. L. Driver. 1987. Analysis of ordinal data to detect population differences. Psychological Bulletin 101 (Spring): 159-65.

Guilford, J. P. 1938. The computation of psychological values from judgments in absolute categories. Journal of Experimental Psychology 22 (January): 32-42.

_____. 1954. Psychometric methods. 2d ed. New York: McGraw-Hill.

Guilford, J. P., and H. F. Dingman. 1955. A modification of the method of equal-appearing intervals. American Journal of Psychology 68 (September): 450-54.

Hogg, Robert V., and Allen T. Craig. 1978. Introduction to mathematical statistics. New York: MacMillan.

Hogg, Robert V., and Elliot A. Tanis. 1977. Probability and statistical inference. New York: MacMillan.

Holdaway, E. A. 1971. Different response categories and questionnaire response patterns. Journal of Experimental Education 40 (Winter): 59.

Hollander, Myles, and Douglas A. Wolfe. 1973. Nonparametric statistical methods. New York: John Wiley & Sons.

Huber, Peter J. 1977. Robust statistical procedures. Philadelphia: Society for Industrial and Applied Mathematics.

Iman, Ronald L., and W. J. Conover. 1983. Modern business statistics. New York: John Wiley & Sons.

IMSL Stat/Library. 1991. FORTRAN Subroutines for Statistical Analysis. Ver. 2.0. Houston: IMSL.

Jarjoura, David. 1983. Confidence and tolerance intervals for true scores. ACT Technical Bulletin 42 (July): 1-57.

Johnson, Richard A., and Dean W. Wichern. 1988. Applied multivariate statistical analysis. 2d ed. Englewood Cliffs, NJ: Prentice-Hall.

Kim, J. 1975. Multivariate analysis of ordinal variables. American Journal of Sociology 81 (September): 261-298.

Kimeldorf, George, Allan R. Sampson, and Lyn R. Whitaker. 1992. Min and max scorings for two-sample ordinal data. Journal of the American Statistical Association 87 (March): 241-47.

Kvanli, Alan H., C. Stephen Guynes, and Robert J. Pavur. 1992. _Introduction to business statistics_. 3d ed. St. Paul: West Publishing.

Labovitz, S. 1967. Some observations on measurement and statistics. _Social Forces_ 46 (September): 151-60.

_____ . 1970. The assignment of numbers to rank order categories. _American Sociological Review_ 35 (June): 515-24.

_____ . 1972. Statistical usage in sociology. _Sociological Methods and Research_ 1 (August): 13-37.

Lodge, Milton, D. Cross, B. Tursky, M. A. Foley, and H. Foley. 1976. The calibration and cross-modal validation of ratio scales of political opinion in survey research. _Social Science Research_ 5 (March): 325-47.

Lodge, Milton. 1987. _Magnitude scaling: Quantitative measurement of opinions_. Beverly Hills: Sage Publications.

Lord, Frederic M. 1953. On the statistical treatment of football numbers. _American Psychologist_ 8 (December): 750-51.

Manly, Bryan F. J. 1990. _Multivariate statistical methods_. London: Chapman and Hall.

Markland, Robert E. 1989. _Topics in management science_. 3d ed. New York: John Wiley & Sons.

Mendenhall, William, and Terry Sincich. 1989. _A second course in business statistics: regression analysis_. San Francisco: Dellen Publishing.

Moses, A. 1984. _New England Journal of Medicine_ 311 (May): 1141-45.

Myers, James H., and W. Gregory Warner. 1968. Semantic Properties of selected evaluation adjectives. _Journal of Marketing Research_ 5 (November): 409-12.

Neibecker, Bruno. 1984. The validity of computer-controlled magnitude scaling to measure emotional impact of stimuli. _Journal of Marketing Research_ 21 (August): 325-31.

Nishisato, Shizuhiko. 1980. _Analysis of categorical data: Dual scaling and its applications_. Toronto: University of Toronto Press.

Parasuraman, A., and P. R. Varadarajan. 1988. Robustness of ordinal measures of competitive strategy variables employed in business research: A PIMS data-based analysis. Journal of Business Research 17 (August): 101-13.

Perreault, William D., and Forrest W. Young. 1980. Alternating least squares optimal scaling: Analysis of nonmetric data in marketing research. Journal of Marketing Research 17 (February): 1-13.

Peter, J. Paul. 1979. Reliability: A review of Psychometric basics and recent marketing practices. Journal of Marketing Research 16 (February): 6-17.

Peterson, Robert A., and Subhash Sharma. 1977. A note on the information content of rating scales. Proceedings of the American Marketing Association 16 (November): 324-6.

Plackett, R. L. 1974. The analysis of categorical data. London: Charles Griffin.

_____. 1981. The analysis of categorical data. 2d ed. London: Charles Griffin.

Puri, Madan L., and Pranab K. Sen. 1985. Nonparametric methods in general linear models. New York: John Wiley & Sons.

Ramsay, J. O. 1973. The effect of number of categories in rating scales on precision of estimation of scale values. Psychometrika 37 (December): 513-32.

Rao, B. L. S. Prakasa. 1983. Nonparametric functional estimation. London: Academic Press.

Ross, Sheldon M. 1989. Introduction to probability models. 4th ed. San Diego: Academic Press.

Scheaffer, Richard L., William Mendenhall, and Lyman Ott. 1990. Elementary survey sampling. 4th ed. Boston: PWS-Kent Publishing.

Schlinger, M. J. 1979. A profile of responses to commercials. Journal of Advertising Research 19 (April): 37-46.

Sellin, J. T., and M. E. Wolfgang. 1964. The Measurement of Delinquency. New York: John Wiley.

Siegel, Sidney. 1956. Nonparametric statistics. New York: McGraw-Hill.

_____, Sidney, and J. W. Tukey. 1960. A nonparametric sum of ranks procedure for relative spread in unpaired samples. <u>Journal of the American Statistical Association</u> 55 (September): 429-44.

_____, Sidney, and N. John Castellan, Jr. 1988. <u>Nonparametric statistics</u>. 2d ed. New York: McGraw-Hill.

Smith, Pamela A. 1993. The early adoption of accounting standards as an earnings management tool. Ph.D. diss. proposal. University of North Texas, Denton.

Stevens, S. S. 1946. On the theory of scales of measurement. <u>Science</u> 103 (June): 677-80.

_____. 1951. Mathematics, measurement, and psychophysics. <u>Handbook of experimental psychology</u>. New York: John Wiley & Sons.

_____. 1957. On the psychological law. <u>The Psychological Review</u> 64 (April): 153-81.

Stockton, John R. 1962. <u>Business statistics</u>. 2d ed. Cincinnati: South-Western Publishing.

Teas, R. Kenneth. 1987. Magnitude scaling of the dependent variable in decompositional multiattribute preference models. <u>Journal of the Academy of Marketing Science</u> 15 (Fall): 64-73.

Teghtsoonian, Robert. 1991. The career of a methodology. <u>Science</u> 251 (March): 1258-59.

Thompson, James R., and Richard A. Tapia. 1990. <u>Nonparametric function estimation, modeling, and simulation</u>. Philadelphia: Society for Industrial and Applied Mathematics.

Torgerson, Warren J. 1958. <u>Theory and methods of scaling</u>. New York: John Wiley & Sons.

Townsend, J., and F. G. Ashby. 1984. Measurement scales and statistics: The misconception misconceived. <u>PsychologicalBulletin</u> 96 (September): 394-401.

Traylor, Mark. 1983. Ordinal and interval scaling. <u>Journal of the Market Research Society</u> 25 (October): 297-302.

Tull, Donald S., and Del I. Hawkins. 1987. <u>Marketing research: Measurement and methods</u>. 4th ed. New York: Macmillan Publishing.

Vidali, J. J. 1975. Context effects on scaled evaluatory adjective meaning. Journal of the Market Research Society 17 (January): 21-25.

Wasan, M. T. 1970. Parametric estimation. New York: McGraw-Hill Books.

Westermann, R. 1983. Interval-scale measurement of attitudes: Some theoretical conditions and empirical testing methods. British Journal of Mathematical and Statistical Psychology 36 (April): 228-39.

Wildt, Albert R., and Michael B. Mazis. 1978. Determinants of scale response: Label versus position. Journal of Marketing Research 15 (May): 261-67.

Woods, Donaldson G. 1972. Studies okay classic tests for ranked data. Journal of Marketing 35 (July): 33-39.

Worcester, Robert M., and Timothy R. Burns. 1975. A statistical examination of the relative precision of verbal scales. Journal of the Market Research Society 17 (July): 181-97.