SAMPLING PROCEDURES AND TYPE I ERROR RATES (FOR NONNORMAL POPULATIONS)

by

WILLIAM CADE

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

1998

MAJOR: EVALUATION AND RESEARCH

Approved by:

Shlomo S. Sawilowsky 12/15/97
Advisor Date

Donald R. Marcotte

Jo Anne Holbert

Cheryl C. Dawson

UMI Number: 9827190

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

Dedication

This paper is dedication to the memory of my mother. Her quiet dignity and devotion to family continue to inspire the best in me.

ACKNOWLEDGMENTS

It is with the deepest gratitude that I mention the names of the following people. This work would not have been possible without their support.

Patricia Cade, my wife, whose support in this project and all the other important areas of life was invaluable.

Travis Cade and Ashton Cade, my children, who never hesitated to offer encouragement.

Hiram Cade, my father, who taught me the meaning and the rewards of hard work.

Dr. Shlomo Sawilowsky, my major advisor, who gave me constant support in this project and in the doctoral program.

Dr. Donald Marcotte, who served on my committee and whose classes enriched me academically.

Dr. Cheryl Durwin, Dr. JoAnne Holbert, and Dr. Margret Smoller who found the time to served on my committee.

Table of Contents

List of Tables

Chapter One

Introduction

Sampling is an essential part of research in education and the social sciences. Researchers typically select subjects or participates from some finite group (a sample frame) and use the findings obtained from this group as an indication of how treatments might impact a larger group (target population). Researchers seek to generalize results obtained from experiments on samples to larger populations. The results obtained from sampling are generalizable to the extent that the sample reflects the population to which the results will be applied.

A sample estimate is biased if the average of all possible sample statistics are either larger or smaller than the population parameter. The amount of bias is defined as the difference between the average of all sample estimates and the value of the population parameter. The variance of an estimator is the averaged squared errors of estimation. The average is taken over all possible samples of a given size. The most desirable estimator is unbiased and the one with the smallest variance. For a given sample size an estimator that has a smaller mean squared error is more efficient. The sampling procedure that generates the smaller mean squared error is therefore the preferred method. Alternative sampling procedures can be used for error is therefore the preferred method. Alternative sampling procedures can be used for this purpose. Reducing error variance, and therefore increasing efficiency, is a primary concern of a researcher's sampling plan. The greater the efficiency the greater the chance the results obtained in research will generalize to the population of interest.

Underlying the generalizability theory is the assumption that the population being sampled is normally distributed. The t-test, the Analysis of Variance, and Regression Analysis assume a population variable that is normally distributed. Normality of distribution in education and social science research is not a given. Several researchers (Boneau, 1960; Peck, Glass et al. 1972; Micceri, 1989) considered the impact of violation of the assumption of normality. Their research on real world data has shown that distributions are often skewed, lumpy, discrete masses with gaps, bimodal, Leptokurtic, Platykurtic, exponential, or Hyper-geometric. Their collective findings suggest that many real world data sets are not normally distributed. Most researchers find that the violations most often encountered are not extreme enough to effect results. See for example, Boneau (1960) Glass Peckham et. al.(1972), Lindquist (1953), Scheffe (1959), Box (1953),Walker and Lev (1969) and Wright (1976). The alternatives offered by most researchers are:

1.) Increase sample size and the effects of the Central

Limit theorem will tend to correct for the problem

2.) Use a nonparametric method which essentially ignores the distribution at least in terms of Type I error.

The question in this study is how can sampling procedures minimize the effects of non-normal distributions when the violations are sufficiently large enough to affect results. In social science research two types of statistical inferences are made: hypotheses testing and estimation. This study will address issues related to hypotheses testing specifically, the effect on the t-test for independent means when sampling from non-normal distributions. The literature suggests that Type I errors are more likely when sampling from non-normal populations. Type I errors occur when the researcher mistakenly concludes that differences between mean are large enough to reject a null hypotheses of no difference. Type I errors in social science research usually suggest a treatment effect when no effect is present.

Type II errors occur when the researcher fails to reject a null hypotheses of no difference between measurements when there really is a difference.
Type II errors are a failure to find a treatment effect when one actually occurs.

This research will estimate the effect of non-normal distributions on Type I errors. The analysis will focus on quantifying the relationship between the degree of nonnormality, sample design and Type I error rates.

This study will consider Simple Random sampling as the standard to which other sampling procedures can be measured against. Cluster, Stratified and Systematic sampling, are alternatives to Simple Random sampling. Simple Random sampling procedures select subjects at random by assigning based on any method that provides each potential subject with an equal chance of selection such as social security number, or random number generators. No attempt is made to subdivide the population of interest.

Cluster sampling is a sampling procedure that sub-divides the population of interest into clusters or groups that represent sub-groups in a larger population. Groups such as city blocks are clusters when the population of interest is a city. Schools are a cluster when the population of interest is a school district. School districts are a cluster when the population of interest is schools within a state. After the clusters are selected each element in each cluster is sampled.

Stratified sampling divides the population into non-overlapping strata or groups. These groups are divided based on some variable that the researcher believes will create homogeneous groups that will be similar in their response to the variable of interest. Demographics such as gender, age, and race are variables that can be used to create stratas. Any variable that tends to group subjects based on their response can also be used to create strata. The strata are assigned a weight based on their proportion in the population of interest.

Systematic sampling is done by assigning a number to each element in the population of interest and selecting every nth element. For example, if the population is 600 and the sample size is determined to be 60, every tenth element is selected. These procedures are widely used in social science research. The question here is whether the use of Cluster, Stratified, Random, or Systematic Sampling lower error variance when sampling from non-normal populations.

The research methodology will be a Monte Carlo Simulation of non-normal population variables. These variables simulate population variables researchers typically encounter when sampling real world data. These distributions will be sampled by the random, systematic, stratified, and cluster sampling procedures. The *t* test will be used to compare the sample means and the population means. The rate of Type I error (rejection of the null hypotheses of no difference between sample means and population means) should be equal to α (.05 in this analysis). If the rejection rate is higher than 5%, inflated Type I errors are present.

This study compares plans based on the single sample t test. The t test will be used to compare the obtained means from sampling to the actual mean of the population distribution, since the mean and standard deviation are known.

Sample sizes will be determined by Cohen's (1988) power tables. Population variables will be introduced that have known levels of skew and kurtosis. This will give the researcher the ability to quantify the relationship between nonnormality (skew and Kurtosis) with specific Type I error rates. A table will be created to show this relationship.

One of the distributions will simulate the normal distribution. This distribution will be used to verify the internal validity of the simulation. When sampling this distribution the Random sampling procedure should yield results that approximate the true population parameter with approximately α (5% in this analysis) rejection of the null hypothesis.

Systematic Sampling will be done by sorting the population variable to achieve random order, and then numbering each element. Every nth element will be selected for the sample. Each successive sample will begin with random start based on random number selection.

Stratified sampling will be achieved by the use of a stratifying variable. This stratification creates groupings (stratas) that are as homogeneous as possible on measurement of the variable of interest. The more homogeneity between stratas the greater the precision (Kish, 1965).

Cluster Sampling will be conducted by creating groupings from the population variable and sampling 100 percent of the elements within each cluster. Cluster sampling precision is dependent on increasing heterogeneity within clusters and decreasing heterogeneity between clusters, Every element within these groups is sampled and the mean of the cluster means is the mean of the overall sample. In this analysis the clusters will be created by using the stratifying variable technique develop for stratified simply. However, the objective of this grouping is to make the clusters as heterogeneous as possible.

The study will consider the implications of assumption violations on the t test. The analysis will look specifically at inflated Type I error rates that result from nonnormally distributed variables. The impact of sample size and sample design on hypotheses testing will be considered. The relationship between sample size, power, and alpha level will be explored. Finally, four sample procedures (random, systematic, stratified, and cluster) will be reviewed. Each procedure will be analyzed in terms of its ability to increase sample precision based on the underlying structure of the design.

The results from this study should help to answer the following:

1. When the underlying population follows the normal curve does sample design effect Type I error rates?

2. As the underlying population becomes increasingly skewed and Kurtotic do Type I error rates increase? Which plans maintain the nominal Type I error rate?

3. When comparing sample estimates, how generalizable are the results?

Chapter Two

Review of Literature

This review will examine the relevant literature on sample design in the social sciences. Variations in population distributions will be reviewed to determine how closely social science variables follow the assumption of normality that underlies most parametric statistical tests.

Sampling is an essential part of research in education and the social sciences. Researchers typically select subjects from some finite group (a sample frame) and use the findings obtained from this group as an indication of how treatments might impact a larger group (target population). When samples are used in experimentation, hypothesis testing is often employed as a means determining the effectiveness of treatments. Researchers seek to generalize results obtained from experiments on samples to larger populations. The results obtained from sampling are generalizable to the extent that the sample reflects the population to which the results will be applied.

In any study that relies on sampling, there is a definable population to which the results of the study are to be applied, this is called generalizability. In educational research studies, populations are often composed of students or educational institutions. In social surveys, populations can consist of many types of elements, including persons, institutions, geographic units, or political units.

In behavioral science studies, the populations of interest are almost always animate but need not be human. Precise definition of the target population is essential in studies involving sampling. The researcher must be able to determine, unequivocally, whether a given element falls within or outside the population (Mansfield, 1973).

The theory of sampling provides a variety of methods for selecting groups from a finite population. There are at least two broad categories of sampling, purposive and probability. Purposive procedures select units into the sample based on how typical they are of the general population according to the personal judgment of those responsible (Cochran, 1977, Sukhatme and Sukhatme, 1970). These procedures are not objective and they are not based on the principles of probability theory.

Probability sampling procedures provide the possibility of estimating and controlling the magnitude of sampling errors. The probability sampling procedures that will be discussed in this study are cluster sampling, simple random sampling, stratified sampling and systematic sampling. These procedures have some defining characteristics (Jaeger, 1984):

1.) The population of interest is specifically defined.

2.) Every potential sample of a given size that could be drawn from the population, together with the probability of selecting each sample, must be specifiable prior to sampling taking place.

3.)Every sampling unit in the population must have a probability of selection that is greater than zero.

The most desirable unbiased estimator is the one with the smallest variance. Reducing error variance and therefore increasing efficiency, is a primary concern of a researchers sampling plan. Sample procedures can be compared for efficiency by comparing the variance of derived estimates. If the estimators are biased, the comparison is the mean squared error. The mean squared error is the sum of the variance and the bias. The bias is zero for all unbiased estimators. Therefore when comparing unbiased estimators the MSE=variance. For a given sample size an estimator that has smaller mean squared error is more efficient. In the case of zero variance and no bias every sample would yield an error-free estimate because every estimate would equal the population parameter. The sampling procedure that generates the smallest mean squared error is therefore the preferred method. Alternative sampling procedures can be used for this purpose.

Combinations of sample procedures are often used to reduce total error. For example, systematic selection can be used with cluster sampling to gain the benefits of proportional stratification if the clusters can be ordered using a stratification variable. Consideration must be given to total error, cost, feasibility, and implications for other choices when selecting a technique and an implementation strategy. Equal probability of selection guarantees all study population members an equal likelihood of being selected in the sample. Unequal probability of selection methods have known, but unequal, probabilities of selection for the members of the study population. Unequal probability methods require weights to correct for the bias that occurs from over-sampling some groups in the population relative to other groups (Cochran, 1977). Unequal probabilities are often needed to improve the precision of an estimate or to provide reliable estimates of sub-populations.

When comparing estimators (sampling procedures) Kish (1965) defined Deffs (design effects) as the ratio of the variance of the estimate obtained from the more complex sample to the variance of the estimate obtained from a simple random sample. A deff that is larger than one will result in an inflated probability of comitting a Type I error (Thomas, 1993).

The most commonly used statistical tests are those for comparing sample means and sample variances. The arithmetic mean is by far the most frequently used measure of location in the behavioral sciences, and hypothesis about them are the most frequently tested (Cohen, 1978). The test used for the equality of the means is usually the analysis of variance (ANOVA) (Box, 1952). The t test is the mathematical equivalent of ANOVA when the number of independent variables is two. One of the assumptions underlying this model is that the variable being measured has a normal distribution (Gaussian)and heterogeneous variance. The normal probability distribution has three properties:

1. A normal distribution histogram (has one mode or peak) and it is symmetrical
2. The normal distribution is continuous. For every value of x there is a value for y and the total area underneath the curve equals 100 percent
3. The normal distribution is asymptotic to the X axis

When these assumptions are not met the computed t statistic may not follow the standard t distribution, and Type I errors are more likely. These errors can occur when researchers compare a measurement taken on an individual member of an undefined sub-population and conclude that the results are significantly different from the overall population parameter. Type I errors can also result when distributions are Multi-modal. When comparing means from this distribution, a researcher may fail to realize that the two means come from two distinct distributions, each with its own peak and tails. The researcher could easily conclude that the means are significantly different.

Researchers such as Boneau (1960), Glass Peckham et. al.(1972), Lindquist (1953), Scheffe (1959), Box (1953), Walker and Lev (1969), Wright (1976) have stated that the t-test is robust to violations of the normality assumption. Blair (1981) Bradley (1977,1978), Sawilowsky and Blair (1992), have argued that the t-test used on non-normal population distributions may result in inflated Type I error rates(rejecting a true null Hypotheses), and decreased comparative statistical power. These authors suggest non-parametric statistics such as the Mann-Whitney U test as a substitute for the t-test when the population distribution is non-normal.

Thomas (1993) suggested that sampling procedures in complex data (non-normal data) can be used to address the problem of inflated Type I error rates.

Researchers have investigated violations of the normality dating back to Pearson in 1929. Box (1953) concluded that the practice of ignoring the assumption underlying statistical test was justified thanks to the work of Pearson (1931), Bartlett (1935), Geary (1947), Gayen (1950), and David and Johnson (1951). Box stated that tests for comparing sample means and variances are remarkably insensitive to general non-normality of the parent populations. Scheffe stated the effect of violations of the normality assumption is slight on inferences about means(1959).

In the late 1950s and 1960 serious questions were raised about the appropriateness of most of the common methods of statistical inference, the t-test and ANOVA. These tests rely on the assumption of normality of the parent population. This movement in statistics culminated in a study by Glass, Peckham, and Sanders, 1972. This study concluded that even though the parametric tests assume normality, rarely has there been any serious practical consequences on the accuracy of the probability statements regarding population means when conventional significance tests are used. The use of ANOVA (and the t-test)again became the statistical method of choice because of their statistical power.

In the 1970s and 1980s there has been a renewed emphasis on the shape of distributions. John Tukey (1977) illustrated that important descriptive information is ignored when the shape of a distribution is not reported.

In the 1980s and 1990s, increased consideration was given to computer-intensive methods such as the bootstrap (Efron, 1977), jackknifing (Tukey, 1950), cross-validation and balanced repeated replications. These procedures ignore the shape of the distribution entirely. These methods generate simulated data sets from original data and assess the actual variability of a statistic from its variability over all possible sets of simulated data. The bootstrap method uses the original sample of data and mimics the process of selecting many samples. The resulting distribution of samples reflect the magnitude of the deviation of the estimate from the true value of the parameter. The jackknife procedure uses a similar methodology. A sample is drawn and observations are removed one at a time. An estimate is recalculated as succeeding observations are removed. The jackknife calls for fewer calculations than the bootstrap but it also seems less flexible. Cross validation splits the original data set in half and fits a curve fitted to the first set of data. Then each observation is tested for best fit to the second half of data. The final testing is the cross validation, it gives a reliable indication of how well fitted the curve would predict the values of new data. The balanced repeated-replication method makes splits in the data systematically in order to assess the variability of surveys and samples. None of these methods make any assumptions about the shape of the population variable of interest (Bradley 1977). Some population variables are very different from the bell shaped curve assumed by most statistical procedures.

Bradley has identified the L and U shaped distributions as fairly common in behavioral research. Bradley stated that the L shape results typically from behavior influenced by uncontrolled conformity.

The U shape distribution can result from measures of taste sensitivity to certain compounds (Bradley 1977). In referring to his study of 440 distributions of educational and psychological measures Micceri (1989) states " The current inquiry shows that even among the bounded measures of psychometry and achievement, extremes of asymmetry and lumpiness are more the rule than the exception". Games and Lucas (1966) suggested that skewed distributions are a greater threat to robustness than leptokurtic or platykurtic distributions, Pearson's 1929 power analysis suggest the opposite. Norton (1952 cited in Glass et al. 1972) examined the degree of skewness(non-symetrical) in data distributions and found that a moderately skewed distribution has a skew value around .5. Runyan and Haber (1991) also agree with this view. The skew value for an extremely skewed distribution was around 1.0. A perfectly mesokurtic distribution has a Kurtosis of 0. Distributions with Kurtosis significantly greater than zero are Leptokurtic, (more observations toward the center of the distribution) while those significantly less than zero are platokurtic (more observations toward the tails of the distribution).

In addition to these measurable deviations from normal distribution real world data are often lumpy and discrete rather than smooth (Micceri, 1989). Lumpiness and discreteness occurs as a result of populations that are the sum of many distinct and non-overlapping sum-population. Populations sub-divided by race, gender , ethnicity, age, and social economic status are examples of such populations. Fleishman (1978)concurred with this view stating: ...many, if not most of the psychological variables seen are skewed and/or Kurtotic to various degrees. When sampling from these populations without any prior knowledge the researcher receives no indication as to the true shape of the underlying distribution.

A sampling distribution of an estimator is the frequency distribution that would result if estimates of a population parameter were computed for all potential samples of a given size. The Central Limit Theorem states: given a population with mean equal to U and variance equal to alpha, as sample size increases, the sampling distribution of the mean for simple random samples will approach the normal distribution (Hinkle 1994). Therefore even a skewed parent population will result in a normal sample distribution (Harnett, 1975).

Non-normal populations in the social sciences are common and some authors have argued that they are the norm. A more pertinent question is what are the effects of nonnormal data on sample design and hypotheses testing. When researchers encounter these distributions what are the implications for sample design, hypotheses testing and generalizability.

When studying complex data, ignoring the survey design can result in biased estimates (Thomas, 1993). When a distribution is non-normal and the researcher ignores this fact, inflated Type I errors can occur (Scott & Holt, 1982;

Rao & Scott, 1981; Rao & Thomas, 1988; Zumbo & Zimmerman, 1991). Type I errors (false indication of a treatment effect) occur when the research falsely determines that the means of two samples are significantly different, too different to be the result of chance error (Gay, 1992). The probability level selected determines the probability of committing a Type I error (rejecting the null hypothesis of no difference between means). If the significance (Alpha)level is set at .05 there is a 5% probability of making a Type I error. Whereas, if the significance level is set at .01 there is a 1% probability of committing a Type I error (Gay, 1992). As the significance level becomes smaller the probability of committing a Type I error decreases and the probability of Type II error increases. Type II errors occur when the researcher fails to reject a false null hypothesis (fails to report a difference when there actually is a difference between means). Inflated Type II errors rates occur as a result of a lack of power(Cohen, 1988).

Cohen provides tables for assessing power levels and needed sample sizes. For any given effect size, alpha and sample size, the tables provide the probability of a Type II error (β). Power is 1-β, the probability of rejecting a false Null Hypothesis. The researcher's challenge is to select alpha prior to execution of the research and determine the seriousness of committing a Type I versus a

Type II error. Research convention seems to view Type I errors as a greater threat than Type II. A conservative tradition among researchers tends to guard against false positives, (Type I errors) indicating a treatment effect when there really is none. Type I errors are seen as more harmful than rejecting an effective treatment in error (Runyon and Haber, 1991). The researcher must determine the consequences of each type of error. An Alpha of .05 is common in social science research (Cohen, 1977). Cohen emphasized the importance of sample size when selecting alpha and effect size.

Determination of the appropriate Sample size has implications for sample design as well as hypotheses testing. As samples get larger, power is increased and sample precision is increased. However, the threat of Type I error is also increased. Small samples reduce power and sample precision and increase the threat of Type II errors.

The size of the sample needed to estimate a population mean for a population total with a specified level of precision is one of the first issues that arise in a practical sampling problem involving parameters(Snedecor and Cochran, 1967). All probability sampling procedures will result in some amount of error. The allowable level of error must be specified prior to the determination of sample size. The following mathematical statement from Jaeger (1984) captures this requirement.

$$\Pr ob\{-\varepsilon \le y - \bar{Y} \le \varepsilon\} \ge 1 - \alpha$$

The statement indicates that the estimate will differ from the population parameter by the amount of ε with the probability of 1-α (usually 99%, 95%, or 90%). If the sample means or sample totals calculated for successive samples of elements are assumed to be normally distributed, sample size requirements can be determined from the standard error.

Perhaps the most powerful attribute of any sampling technique is sample size. Statistics calculated from large samples are more accurate, other things equal, than those calculated from small samples. Large samples are preferred because they have greater probability of achieving randomization. Random selection and random assignment justify the theory that underlie tests of significance (Heinsman and Shadish, 1996). With small samples, the probability of selecting deviant samples is greater than with large samples (Kerlinger, 1986). Increases in sample size increases the power of a statistical test. Power efficiency, is the sample size that is required to make one test as powerful as a competing test (Runyan and Haber, 1991). Cohen (1988)states,:

> The larger the sample size, other things being equal, the smaller the error and the greater the reliability or precision of the results. The further relationship with power is also intuitively evident: the greater the precision of the sample results, other things being equal, the greater the probability of detecting a none null state of affairs, i.e.,

> the more clearly the phenomenon under test can manifest itself against the background of (experimentally irrelevant) variability. As is intuitive obvious, increases in sample size increase statistical power, the probability of detecting the phenomenon under test. (P. 7)

As power increases the probability of a Type I error increases as well, in cases where the null hypothesis of no difference between means is not true. Critics of hypotheses testing have noted that large sample sizes will result in rejection of the null hypotheses even for trivial treatment effects (Serlin, 1992). Serlin suggested a "range null" hypotheses that would reflect the magnitude of the population effect (This principle is born out by the properties of the central Limit theorem. As sample size increases the mean of the sample better approximates the population mean. In addition, the variance of the estimate decreases as the sample size increases. The formula for the standard error of the of the sample mean illustrates this point.

$$V = \frac{N^2 S^2}{n}\left(1 - f\right)$$

In this formula *V* is the variance, *N* is the total finite population, *S* is the standard deviation of the sample elements and *F* is the fraction of the population that was sampled and *n* is the size of the sample. No matter what the size of population and no matter what the standard deviation, the variance becomes smaller as the sample size

(n) increases.

Cohen's (1988) power tables can be used to set sample size requirements, once power, alpha, and effect size are determined. Cohen suggest .8 as convention for power when alpha is set at .05. Cohen's "d" (effect size) is calculated as follows for independent sample means:

$$d = \frac{M_a - M_b}{\sigma}$$

M_a is the mean of the first group, M_b is the mean of the second group and σ is the standard deviation of the population. "d" is operationally defined as small, medium, and large, with numeric values of .20,.50, and .80 respectively. In this way "d" becomes a measure of the magnitude of the difference between means expressed in terms of standard deviation units. Cohen has been critical of psychological experiments that lack power. His power tables suggest that experiments with power levels around .5 have only a 50-50 chance of producing statistical significance in the presence of a real treatment effect (Hammond, 1996).

Cohen concedes that this approach will sometimes lead to sample size requirements that are beyond the limits of cost and practicality (Cohen, 1992). The researcher must decide how large an error is acceptable, which makes the sample size requirement somewhat arbitrary when making comparisons between samples (Snedecor and Cochran, 1967).

Each sampling procedure (random, systematic, stratified, and cluster) has its own unique calculations for estimation of population means and variances. These calculations reflect the underlying structure of each procedure. Each plan will be evaluated to determine how precision can be increased when calculating sample means. This analysis will focus on the sampling procedures that must be followed in each sample design. Methods designed to reduce variance and increase precision and efficiency will be examined. Advantages and disadvantages of each procedure will be addressed. Issues of cost will not be addressed explicitly.

SIMPLE RANDOM SAMPLING

With Simple Random Sampling every member of the population has an equal probability of selection. With simple random sampling differences between the sample and the population should be small and unsystematic (Gay, 1992). Simple random sampling is used in situations where simplicity is the main concern. To select a simple random sample, researchers need a complete listing of the members of the study population. The sample can be selected by the following procedure as outlined by Jaeger (1984).

1) assigning each member of the population a unique identification number.

2) selecting a random start in a table of random numbers,

3) using the number of digits in the random number table that is equal to the number of digits in the highest

identification number,

4) selecting each population member that has a number that corresponds to the random number selected.

5) discarding any random number that does not have a corresponding number in the population, and

6) repeating the process until the desired number (n) of members have been selected.

The advantages of simple random sampling are the ease of selection and the ease of use of the data. Once the sampling frame is assembled, no other information about the population is needed for sampling. A major disadvantage of this method is that a listing of the entire study population is required. This method does not use any previous knowledge of the target population. Subgroups and special populations within the larger population are not explicitly addressed. Randomization is achieved by way of the random selection process. Simple random sampling is used with and without modification through out the social sciences. For example pre-existing groups such as classrooms are often chosen and random sampling is conducted on these groups. Results will be generalized to larger groups even though the actual sample frame was the smaller pre-existing group. Random sampling is the standard against which all other sampling procedures are compared (Kish, 1965). However, it is usually not the most efficient procedure when compared to cluster and stratified sampling. Simple random sampling is not as efficient, in terms of standard error per unit sampled, as techniques using stratified sampling. The formula for the sample mean when using simple random sampling is:

$$\bar{y} = \frac{1}{n}\sum_{i} y_i$$

$\bar{y}$ is the sample mean, n is the number of elements in the sample, y_i is the observed value (Frankel, 1983). The formula for the standard error of estimate when using simple random sampling is

$$s^2 = \frac{1}{n-1}\sum_{i}(y_i - \bar{y})^2$$

In this formula s is the standard error , y_i is the observed value, $\bar{y}$ is the mean of the sample and n is the sample size. This formula is used when the population variance is unknown. The only variable under the control of the researcher in this formula is n, the sample size. Therefore, the only way to reduce variance from a sample design stand point is to increase sample size. The formula for the sample size required for a given interval estimate is:

$$n = \frac{(tS/\varepsilon)^2}{1+(1/N)(tS/\varepsilon)^2}$$

In this formula t is a value from the distribution based degrees of freedom, S is the standard error of the sample, ε is the allowable error, N is the population size.

SYSTEMATIC SAMPLING

Systematic Sampling can offer researchers benefits over simple random sampling. Ease of selection is one benefit of Systematic Sampling. Another advantage is that systematic sampling can be used for detecting stratification. To achieve de facto stratification, the units to be sampled must be arranged by the stratifying variable prior to the selection plan. The variable must be available for the entire study population. The stratification variable can be either a continuous variable such as the age of respondents used to order the population, or a discrete variable, such as grade level for a student. The variable is used to group the population to obtain proportional representation on the stratification variable. One caution is necessary in using systematic sampling, the sampling frame must be well mixed or purposefully arranged for stratification. If a cyclical arrangement of the frame is inadvertently used, the sample can be biased.

A systematic sample is often viewed as being essentially the same as random sample. It is important to recognize that this is true only if the elements of the population are in random order on the list (from the sampling frame) (Mansfield, 1973). The major objection to systematic sampling of a non-random list is the possibility that certain subgroups of the population can be systematically excluded from the sample (Gay, 1992). If on-site fieldwork is required, a disadvantage of this technique is the dispersal of the fieldwork across the entire geographic spread of the study population. Costs of travel and supporting the fieldwork may be prohibitive. The formula to compute standard errors is (Cochran, 1977):

$$\overline{Y} = \frac{N-1}{N}S^2 - \frac{k(n-1)}{N^2}S^2_{wsy}$$

linear systematic sampling yields more precise estimates than simple random sampling only when $S^2_{wsy} > S^2$ and the two methods are equally efficient only when $S^2_{wsy} = S^2$. Linear systematic simply is more efficient than sample random sampling only when the average variance among elements (S^2_{wsy}) in the systematic sample is larger than the variance among elements in the entire population (S^2). Whenever adjacent elements in the sampling frame are more similar than those some distance apart, the within sample variance will be larger than the overall variance, and linear systematic sampling will be more efficient than simple random sampling.

Systematic sampling presents a unique problem in sample determination. The sampling variance depends much on the arrangement of units in the population. Cochran (1977) states:

> The performance of systematic sampling in relation to that of stratified or simple random sampling is greatly dependent on the properties of the population...For some populations and some values of (sample size), the variance of the sample mean even increases when a larger sample is taken, a startling departure from good behavior. (P.212)

Stratified Sampling

Stratified sampling involves grouping the study population into strata and selecting a random sample within each stratum. The basic idea in formulating strata is to subdivide the population so that these subdivisions differ greatly with regard to the characteristic being measured and so that there is as little variation as possible within each stratum (or sub-division) with regard to the characteristic being measured (Mansfield, 1973). Kish explains the way gains in precision are achieved from stratified sampling as follows:

(1) Variability Between *Stratas:* the greater the difference between the means of the strata and the overall means, the greater the gain; and

(2) Homogeneity Within Stratum: the greater the similarity within the stratum the greater the gain.

For example the views of teachers can be sampled by subdividing teachers by grade level. They can be stratified as high school teachers, middle school teachers, and elementary school teachers. Views on political issues can be stratified in several ways. Stratas could be based on age, race, political party affiliation, income, or region of the country.

Several methods for creating stratas have been proposed. The *cum* $\sqrt{f}$ method of Delenius and Hodge (1959), creates stratum of approximately equal size by using a stratification variable. The following conditions must be satisfied to use this method:

1 The stratification variable is identical to the characteristic of interest.

2. The stratification variable is approximately uniformly distributed between adjacent endpoints.

3. Neyman allocation is used. (to be discussed later)

4. The allocation variable is the characteristic of interest.

When determining the number of stratas to create, Cochran (1967) concludes that there is very little additional variance reduction from having more than six stratas unless the correlation between the stratification variable and the variable of interest exceeds .95. Stratified sampling is often used to ensure proportional representation for each stratum, decrease the sampling variability, or to yield a sufficient number of a sub-population in the sample for reliable analysis. Stratas may be formed to provide separate estimates for sub-universes or domains such as ethnic groups or particular age groups (Wilburn, 1984). Stratified sampling takes advantage of the existence of different sub-groups in the population. With simple random sampling within stratas assuring proportional representation of various sub-groups is a problem when sample sizes are small (Wilburn, 1984).

Stratified sampling can be done proportionately or disproportionately. Proportional allocation makes the sample size in each stratum proportional to the total number of elements in the population. Optimal allocation requires that the sample in each stratum be proportional to the product of the number of elements in the stratum and the standard deviation of the characteristic being measured. This formula for allocation is ascribed to Neyman (1934) and the method is sometimes called "Neyman allocation". The methodology ensures proportionate representation of the stratas by using the same sampling fraction in each stratum. The sample mean when using stratified sampling is calculated as follows:

$$\bar{y}_w = \sum_{h=1}^{H} W_h \bar{y}_h = \sum_{h=1}^{H} \frac{1}{n_h} W_h \sum_{i=1}^{n_h} y_{hi}$$

In this formula $\bar{y}_w$ is the overall estimate of the population mean, W_h is the weight associated with the *hth* stratum, $\bar{y}_h$ is the mean associated with the *hth*, n_h is the number of elements in the *h*th stratum and y_{hi} is the value of variable y in the *i*th sample element in the *h*th stratum and $\bar{y}_h$ is the sample mean from the *h*th stratum (Frankel, 1983). The variance of the sample means in each stratum can be combined to yield the following formula for the variance of the estimator of the overall population mean:

$$v = \sum_{k=1}^{K} W_k^2 \frac{s_k^2}{n_k}\left(1 - f_k\right)$$

In this formula *W* is the weight associated with strata *K*, S is the standard error of the sample, *n* is the sample size in strata *K*, $1-f$ is the percent not sampled in the *k*th strata. The calculation for *S* in the above formula is as follows

$$S_k^2 = \frac{\sum_{i=1}^{n_k}\left(y_{ik} - \bar{y}_k\right)^2}{n_k - 1}$$

y_{ik} is an element in the *k*th strata $\bar{y}_k$ is the mean of the *k*th strata, *n* is the sample size.

When a stratifying variable is used to create stratas the greater the explanatory power of the stratifying variable, the greater the gain from stratification. (Kish, 1965) The gain is proportional to the amount of variance explained (R). Kish (1965) pointed out that the proportional size of the strata has a bearing on the relative gain.

The advantages of proportional stratification are improved precision of estimates and insuring proportional representation of stratifying groups. The formula to determine the appropriate sample size when using proportional allocation,

$$n = \frac{\left(\sum_{k=1}^{k} W_k s_k^2\right) / (\varepsilon / t)^2}{1 = (1/N)(t/\varepsilon)^2 \sum_{k=1}^{k} W_k s_k^2}$$

In this formula *W* is the weight associated with the *k*th strata, *S* is the standard error of the *k*th strata, *N* is the sample size, ε is the allowable error, and t is the *t* value from the *t* distribution.

Disproportional stratification is used when researchers are faced with the situation where the overall sample precision (variance of the sample estimate) is not sufficient or where the representation of a sub-population is not sufficient. The researcher can turn to disproportional stratification (Gay, 1992). Disproportional stratification results from using different sampling fractions in different strata. Employing different sampling fractions causes unequal probabilities of selection and disproportional representation in the final sample. Thus, weighting is required to adjust for the selection bias. The benefit from disproportional stratification results from lowering the sampling variability in the stratum where the standard deviation is relatively high by increasing the number of sampling units allocated to that stratum (Groves, 1989). The formula for sample size requirements when using disproportional or optimal allocation is,

$$n = \frac{\left(\sum_{k=1}^{k} W_{k^{n}}\right)^2}{(\varepsilon / t)^2 + (1/N)\sum_{k=1}^{k} W_{k^{n^2}}}$$

The term used in this equation are defined on the previous page. The formula can be used for either proportional or disproportional stratification. Since standard errors are computed for each stratum and then combined as a weighted average, the strata with the largest standard errors and the largest weights have the most influence on the standard error.

Increasing the sample size within the stratum that has the highest sampling error will reduce the sampling error within the stratum and, reduce the overall sampling error.

The standard deviation of the population and strata are seldom known. Therefore, the allocation of sampling units to strata is usually imprecise but can result in improved precision if the allocation is done accurately. The standard deviations, or more usually their relative sizes, can be estimated. Each stratum can be estimated by the same methods used in estimating the standard deviation. Prior studies or pilot studies can be used for this purpose.

Stratification allows increasing the sample size for the sub-population without increasing the entire sample size proportionally. Stratas can be defined in such a way that sub-populations can be included within a particular strata to increase the effect of stratification. Stratas created in this way can result in the loss of some efficiency. The central drawback in using disproportional stratification is the use of weights in the calculation of standard errors (Sudman, 1976). The calculations become large and more complex and the strata identification must be maintained in weights included as a variable in the data set.

CLUSTER SAMPLING

Cluster sampling is the random selection of groupings, referred to as clusters, from which all members are chosen for the sample. The major advantage of cluster sampling is that it is cheaper to sample elements that are physically or geographically close (Mansfield, 1973). Any location with intact groups of similar characteristics is a cluster (Gay, 1992).

Geographic groupings and intact groups are commonly used for clusters. Counties are frequently used as clusters. Schools or, alternatively, classrooms are often used as clusters for education-related studies. Regional offices or local clinics are potential clusters for management or evaluation research. When clusters such as these are used, the researcher forgoes the opportunity to provide information on individual regions or schools, other than those units in the selected clusters (Jaeger, 1984).

As a result Cluster sampling can greatly reduce travel and training expense. Cluster sampling can be done in stages involving selection of clusters within clusters.

This is multi-stage sampling (Gay, 1992).

In cluster samples, the selection of each cluster is random and, likely to be independent. However, the selection of each sampling unit is not independent. That is, the sampling units included in the sample are determined by the selection of the clusters. This results in a loss of independence in selection. This loss of independence in selection can lead to inflated Type I errors (Thomas, 1993) (Zumbo and Zimmerman, 1991). The formula for the cluster mean when using Cluster sampling is as follows (Frankel, 1983):

$$\bar{y}_{\alpha} = \frac{1}{b}\sum_{\beta=1}^{b} y_{\alpha\beta}$$

This formula for the cluster mean is the same calculation as simple random sampling.

The estimated population mean is:

$$\bar{y}_{cl} = \frac{1}{a}\sum_{\alpha=1}^{a} \bar{y}_{\alpha}$$

The formula for the estimation of variance for cluster sampling is,

$$v = \frac{N^2(1-f)}{(N-1)nM_0^2}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)^2$$

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M_j} y_{ij}$$

N is the sample size, *n* is the elements in cluster *i*, M_o is the total elements in the population.

The loss of independent information from each sampling unit brings a loss in precision. The impact of the loss of information is evident in the formula used to estimate the sampling error for cluster samples, in this case with approximately equal size clusters:

$$\bar{y}_\alpha = \frac{1}{b}\sum_{\beta=1}^{b} y_{\alpha\beta}$$

This formula is equivalent to the formula for simple random sampling. The number of clusters selected replaces the sample size, and the number of clusters in the population replaces the population size in the formula for simple random samples. This is because the number of clusters selected is the number of independent selections. The greater the within-cluster differences, the greater the sample precision (Sukhatme and Sukhatme, 1970).

The number of clusters affects the precision of the sample estimates. This is analogous to the effect of increasing the sample size in simple random samples. Standard error varies as a function of the square root of the number of clusters. Increasing the number of clusters increases the precision of the sample. When using cluster sampling, selecting more clusters with less between-group variation improves precision (Sukhatme and Sukhatme, 1970). However, increasing the clusters can also increase costs of collecting data.

The impact of increasing the standard error from cluster sampling can be partially overcome by stratification of the clusters. This occurs because combining the standard errors of weighted cluster means within strata improves sample precision(Hansen and Hurwitz, 1943).

Chapter Three

Methodology

The research methodology will be a Monte Carlo Simulation of non-normal population variables. These variables simulate population variables researchers typically encounter when sampling real world data. These distributions will be sampled by the random, systematic, stratified, and cluster sampling procedures. The *t* test will be used to compare the sample means and the population means. The t statistic is calculated as follows:

$$t = \sqrt{[\frac{SS_1 + SS_2}{N_1 + N_2 - 2}][\frac{1}{N_1} + \frac{1}{N_2}]}$$

The rate of Type I error (rejection of the null hypotheses of no difference between sample means and population means) will be equal to α (.05 in this analysis). If the rejection rate is higher than 5%, inflated Type I errors are present. If the rate of rejection falls below .05, Type II errors are indicated. Type II errors also indicate a loss of statistical power. This study compares plans based on the single sample t test.

Sample sizes will be determined by Cohen's power tables. The effect size will be assumed to be moderate. Alpha will be set at .05. Cohen assumes .8 as a convention for power. With effect size, power, and Alpha set as indicated the appropriate sample size is 20 (Cohen, 1988).

Additional sampling will be conducted using sample sizes of 15 and 25. This additional sampling will further define the relationship between sample procedures and sample size. The distributions will be non-normal with specified degrees of kurtosis and skew.

Fleishman's (1978) Power method with the Headrick and Sawilowsky correction will be used to generate the non-normal distributions. This methodology offers advantages when compared to other procedures for producing non-normal distributions. Fleishman (1978) identifies four alternative methods to generate non-normality.

1.) The outlier method involves adding a small proportion of data having variance much greater than the remainder of the population. This procedure creates highly Kurtotic but symmetrical distributions.

2.) The extreme non-normality method samples from extreme distributions by using inverse probability function on uniform deviates.

3.) The transformation method transforms a random normal deviate by a skewing function, usually without knowledge of the degree of skew.

4.) The tabular method uses a table of non-normal population The power method allows the analysis to specify the level of non-normality by the use of a polynomial transformation. The power method as modified by Headrick and Sawilowsky is of the form:

$$Y = a + bX + cX^2 + dX^3$$

In this formula a, b, c, and d are constants. When this formula is applied to a normal distribution the non-normal distributions are created based on the constants. This method requires a normal distribution with a mean of zero and a standard deviation of one as a starting point. Rangen will be used to generate the normal deviates. Rangen is a is a collection Fortran source code and subroutines that generate pseudo-random deviates.

Table 3.1 gives the constant for solved values of skew and kurtosis. When these constants are applied to the random deviates generated by Rangen, non-normal distribution with specific skew and kurtosis are produced.

Skew and Kurtosis as measures of nonnormality were chosen because they are the most important indicators of the extent to which nonnormality affects the usual inferences made in the analysis of means (Bradley, 1982). E.S. Pearson (cited by Runyon and Haber, 1991) proposed the following coefficient of skew (sk):

$$sk = \frac{3(\bar{X} - median)}{s}$$

Runyon and Haber conclude that data sets with indices of skew ranging between +/-.5 may be considered sufficiently symmetrical for most practical applications. Skewed data sets often result from floor or ceiling effects of measurements in social sciences. Therefore, for this analysis skew values larger than +/-.5 will used in non-normal data sets. This analysis will generate distributions with skew ranging from .5 to 3.00 in increments of approximately .25. Distributions will be generated with minimum, mid-point, and maximum Kurtosis for each skew value (see table 3.1 for specific values).

The purpose of selecting this range of values is to show how increasing levels of nonnormality effect Type I errors and the role that sample design can play. Additional analysis will include Chi-Square distribution with 1 degree of freedom, the Exponential distribution and the Uniform distribution. The normal distribution will be simulated to validate the internal validity of the study methodology. When sampling this distribution Random sampling procedure should yield results that approximate the true population parameter with approximately 5% rejection of the null hypothesis. Each sampling procedure will be used to sample the population at a fixed sample size. Ten Thousand repetitions of each procedure and each sample size will be completed. The results from these samples should yield 500 rejections of the null hypothesis of no difference between means.

Random sampling

Two variables will be generated by Rangen. Random sampling will done at sample sizes of 15, 20, and 25. Ten thousand repetitions will be conducted at each sample size

Table 3.1. Skew and Kurtosis For Normal and Nonnormal Distributions and Fleishman power constants

Distribution	Skew	Kurtosis	a	b	c	d
Normal	0.00	0.00	0.0000	1.0000	0.0000	0.0000
Uniform	0.00	-1.15	0.0000	1.3320	0.0000	0.0000
	0.26	-1.04	-0.0912	1.3174	0.0912	-0.1246
	0.26	6.34	-0.0250	0.6531	0.0250	0.1044
	0.26	13.71	-0.0428	1.5780	0.0428	-0.3170
	0.50	0.74	-0.1652	1.2742	0.1652	-0.1124
	0.50	7.23	-0.1104	1.5520	0.1104	-0.2825
	0.50	13.71	-0.0836	1.5700	0.0836	-0.3140
	0.76	-0.23	-0.2354	1.2093	0.2354	-0.0995
	0.76	6.97	-0.0730	0.6450	0.0730	0.1052
	0.76	13.71	-0.0598	-0.4679	0.0598	-0.1521
	1.00	0.42	-0.2936	1.1488	0.2936	-0.0884
	1.00	7.07	-0.0977	0.6522	0.0977	0.1020
	1.00	13.71	-0.0795	-0.4736	0.0795	-0.1499
	1.26	1.34	-0.3517	1.0781	0.3517	-0.0771
	1.26	7.53	-0.1247	0.6522	0.1247	0.1003
	1.26	13.71	-0.1018	-0.4818	0.1018	-0.1467
	1.50	2.37	-0.4015	1.0072	0.4015	-0.0669
	1.50	8.04	-0.1506	-0.6518	0.1506	-0.0984
	1.50	13.71	-0.1235	-0.4914	0.1235	-0.1430
	1.76	3.71	-0.4533	0.9230	0.4533	-0.0560
	1.76	8.71	-0.1804	-0.6515	0.1804	-0.0956
	1.76	13.71	-0.1488	-0.5045	0.1488	-0.1377
	2.00	5.15	-0.4995	0.8358	0.4995	-0.0456
	2.00	9.43	-0.2102	-0.6511	0.2102	-0.0922
	2.00	13.71	-0.4815	1.1480	0.4815	-0.2046
Exponential	2.00	6.00	-0.3137	0.8263	0.3137	0.0227
	2.26	6.96	-0.5496	0.7256	0.5496	-0.0340
	2.26	10.34	-0.2461	-0.6502	0.2461	-0.0875
	2.26	13.71	-0.5607	0.9525	0.5607	-0.1668
	2.50	8.87	-0.5982	0.5972	0.5982	-0.0223
	2.50	11.29	-0.2852	-0.6492	0.2852	-0.0813
	2.50	13.71	-0.2435	-0.5635	0.2435	-0.1116
	2.76	11.28	-0.6645	0.3619	0.6645	-0.0068
	2.76	12.50	-0.5738	0.4467	0.5738	0.0251
	2.76	13.71	-0.6779	0.4581	0.6779	0.0846
Chi-Square 1df	2.83	12.00	-0.5207	0.6146	0.5207	0.0201
	3.00	13.71	-0.5047	0.5993	0.5047	0.0350

Note: a=-c

and each level of skew and kurtosis. Two tail t test will be calculated for each sample. Rejections in the upper and lower tail should total 5% (500 rejections for each 10,000 repetitions of sampling).

Systematic Sampling

Systematic sampling will done by selecting sample elements for two variables using a specified interval of selection. Every nth element will be selected for both variables. The means will be compared using the independent samples t test. Ten thousand repetitions will be completed for each distribution and each sample size.

Stratified sampling

Stratified sampling will done by selecting groups for two variables. The means will be compared using independent samples t test. This process would continue until all groups are created.

Random sampling will be conducted within each strata. Researchers can use proportional or disproportional stratification. If the stratas are proportional in size weights are the same for each strata. If the stratas are different in size, the weights must reflect this difference. The strata means will be weighted based on size.

Cluster Sampling

Cluster sampling creates groupings of the population variable based on geographic consideration or a stratifying variable. Sampling is conducted on 100 percent of the elements within each cluster. Cluster sampling precision is dependent on increasing heterogeneity within clusters and decreasing heterogeneity between clusters. Researchers typically select pre existing groups as clusters. Groups such as classrooms, city blocks, political precincts of such groupings. Every element within these groups are sampled and the mean of the cluster samples is the mean of the overall sample.

Cluster sampling will be done by creating three clusters of different sizes and comparing the weighted means to the mean of a group of 400. This comparison should be mathematically equivalent to cluster sampling. Cluster sizes will vary so that different sample sizes can be obtained. The following combinations will create sample sizes of 15, 20, and 25: 9+3+3=15, 10+5+5=20 and, 15+5+5=25. The sampling plan will be similar to the previous plan used in random, systematic, and stratified sampling. The distribution will be varied from .50 skew to 3.00 skew, and kurtosis will vary from the minimum to a mid-point to high for each level of skew. Sample sizes will be 15, 20 and 25 for each distribution.

The results from this analysis should answer the following:

1. When the underlying population follows the normal curve does sample design effect Type I error rates?

2. As the underlying population becomes increasingly skewed

and Kurtotic do Type I error rates increase? Which plans maintain the nominal Type I error rate?

3. When comparing sample estimates, how generalizable are the results ?

Additional sampling analysis will be conducted on populations that have natural clustering and stratification. These populations have distinct groupings within a larger population. The groups can differ in terms of means or symmetry. This area of the study will focus on stratified populations where the stratas represent groups based on mean differences and clustered populations with distinctly different skew and kurtosis within each cluster.

A clustered population will be created by grouping cluster with skew and kurtosis as follows:

	skew	Kurtosis
cluster I	.76	-.23
cluster II	1.50	8.04
cluster III	2.50	11.29

The stratified population will be created by grouping sample elements by z score. Elements in the lower third of the z scale will distributed across all three stratas, with 80% in the strata I, 10% in the strata II and 10% in the strata III. Similarly elements in the middle of the z scale will be distributed as follows, 10% in strata I, 80% in strata II and 10% in the strata III. The elements in the higher end of the z scale will distributed with 10% in strata I, 10% in strata II and 80% in strata III. The distribution of will be as follows:

Z scale	lower 1/3	middle 1/3	upper 1/3
strata I	80%	10%	10%
strata II	10%	80%	10%
strata III	10%	10%	80%

When researchers either ignore or are unaware of these special grouping, what are the implications for Type I error rates?

Chapter Four

Results

The Monte Carlo simulation was completed for the four sample procedures. Each procedure was completed thousand times for each sample size. Sample sizes of fifteen, twenty, and twenty-five were used in each case.

The computer simulation work was done on an ACER computer using an Intel Pentium processor. The computer programming work was done using Lahey Fortran 77 code.

The following tables present Type I error rates for each sampling plan and each distribution. The samples were taken over a range of skew values from 0 to 3.00. The distributions were sampled at increments of approximately .25 levels of skew. Simple random sampling was used to sample the clustered and stratified distributions.

Table 4.1A. Type I error rate when nominal rate is .05 **Normal Distribution**

	Distribution		Sample Size								
Sampling Plan	Skew	Kurtosis	*n=15*			*n=20*			*n=25*		
			Lower Tail	Upper Tail	Total	Lower Tail	Upper Tail	Total	Lower Tail	Upper Tail	Total
Random	0.00	0.00	.022	.027	.049	.025	.027	.052	.025	.029	.054
Systematic	0.00	0.00	.028	.026	.054	.023	.024	.047	.025	.026	.051
Stratified	0.00	0.00	.025	.024	.049	.026	.024	.050	.026	.025	.051
Cluster	0.00	0.00	.026	.025	.051	.029	.026	.055	.025	.026	.051

Table 4.1B. Type I error rate when nominal rate is .05 **Uniform Distribution**

	Distribution		Sample Size								
Sampling Plan	Skew	Kurtosis	*n=15*			*n=20*			*n=25*		
			Lower Tail	Upper Tail	Total	Lower Tail	Upper Tail	Total	Lower Tail	Upper Tail	Total
Random	0.00	1.15	.023	.027	.050	.025	.024	.049	.025	.029	.054
Systematic	0.00	1.15	.027	.027	.054	.022	.026	.048	.025	.025	.050
Stratified	0.00	1.15	.026	.023	.049	.028	.024	.052	.026	.026	.052
Cluster	0.00	1.15	.026	.024	.050	.028	.025	.053	.025	.027	.052

Table 4.1C. Type I error rate when nominal rate is .05 **Exponential Distribution**

	Distribution		Sample Size								
Sampling Plan	Skew	Kurtosis	*n=15*			*n=20*			*n=25*		
			Lower Tail	Upper Tail	Total	Lower Tail	Upper Tail	Total	Lower Tail	Upper Tail	Total
Random	2.00	6.00	.020	.024	.044	.023	.024	.047	.024	.025	.049
Systematic	2.00	6.00	.024	.023	.047	.021	.024	.045	.023	.024	.047
Stratified	2.00	6.00	.022	.022	.044	.024	.025	.049	.025	.024	.049
Cluster	2.00	6.00	.016	.033	.049	.017	.031	.048	.014	.029	.043

Table 4.1D. Type I error rate when nominal rate is .05 **Chi Square Distribution (1df)**

	Distribution		Sample Size								
Sampling Plan	Skew	Kurtosis	*n=15*			*n=20*			*n=25*		
			Lower Tail	Upper Tail	Total	Lower Tail	Upper Tail	Total	Lower Tail	Upper Tail	Total
Random	2.83	12.00	.021	.020	.041	.020	.021	.041	.023	.023	.046
Systematic	2.83	12.00	.022	.022	.044	.022	.023	.045	.023	.023	.046
Stratified	2.83	12.00	.019	.020	.039	.023	.023	.046	.024	.021	.045
Cluster	2.83	12.00	.009	.035	.044	.010	.034	.044	.014	.031	.045

Table 4.2A. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	n=15 Lower Tail	n=15 Upper Tail	n=15 Total	n=20 Lower Tail	n=20 Upper Tail	n=20 Total	n=25 Lower Tail	n=25 Upper Tail	n=25 Total
Random	0.26	-1.04	.023	.025	.048	.025	.025	.050	.024	.028	.052
Systematic	0.26	-1.04	.028	.026	.054	.023	.026	.049	.026	.026	.052
Stratified	0.26	-1.04	.028	.025	.053	.026	.024	.050	.026	.024	.050
Cluster	0.26	-1.04	.025	.025	.050	.026	.025	.051	.024	.027	.051

Table 4.2B. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	n=15 Lower Tail	n=15 Upper Tail	n=15 Total	n=20 Lower Tail	n=20 Upper Tail	n=20 Total	n=25 Lower Tail	n=25 Upper Tail	n=25 Total
Random	0.26	6.34	.025	.022	.047	.024	.021	.045	.022	.023	.045
Systematic	0.26	6.34	.028	.026	.054	.023	.026	.049	.026	.026	.052
Stratified	0.26	6.34	.022	.023	.045	.023	.021	.044	.027	.025	.052
Cluster	0.26	6.34	.024	.025	.049	.026	.026	.052	.023	.024	.047

Table 4.2C Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	n=15 Lower Tail	n=15 Upper Tail	n=15 Total	n=20 Lower Tail	n=20 Upper Tail	n=20 Total	n=25 Lower Tail	n=25 Upper Tail	n=25 Total
Random	0.26	13.71	.023	.024	.047	.023	.022	.045	.024	.026	.050
Systematic	0.26	13.71	.028	.026	.054	.025	.027	.052	.025	.025	.050
Stratified	0.26	13.71	.027	.025	.052	.023	.024	.047	.023	.026	.049
Cluster	0.26	13.71	.026	.024	.050	.024	.024	.048	.025	.026	.051

Table 4.3A. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	n=15 Lower Tail	n=15 Upper Tail	n=15 Total	n=20 Lower Tail	n=20 Upper Tail	n=20 Total	n=25 Lower Tail	n=25 Upper Tail	n=25 Total
Random	0.50	-0.74	.023	.025	.048	.025	.025	.050	.024	.027	.051
Systematic	0.50	-0.74	.028	.025	.053	.023	.027	.050	.026	.025	.051
Stratified	0.50	-0.74	.029	.025	.054	.025	.024	.049	.026	.025	.051
Cluster	0.50	-0.74	.024	.027	.051	.026	.026	.052	.024	.026	.050

Table 4.3B. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	n=15 Lower Tail	n=15 Upper Tail	n=15 Total	n=20 Lower Tail	n=20 Upper Tail	n=20 Total	n=25 Lower Tail	n=25 Upper Tail	n=25 Total
Random	0.50	7.23	.023	.024	.047	.024	.022	.046	.023	.025	.048
Systematic	0.50	7.23	.029	.027	.056	.025	.026	.051	.025	.025	.050
Stratified	0.50	7.23	.027	.026	.053	.024	.024	.048	.024	.026	.050
Cluster	0.50	7.23	.026	.024	.050	.023	.024	.047	.023	.026	.049

Table 4.3C. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	n=15 Lower Tail	n=15 Upper Tail	n=15 Total	n=20 Lower Tail	n=20 Upper Tail	n=20 Total	n=25 Lower Tail	n=25 Upper Tail	n=25 Total
Random	0.50	13.71	.023	.024	.047	.024	.023	.047	.024	.025	.049
Systematic	0.50	13.71	.028	.026	.054	.025	.027	.052	.025	.025	.050
Stratified	0.50	13.71	.026	.025	.051	.024	.024	.048	.023	.026	.049
Cluster	0.50	13.71	.025	.023	.048	.023	.024	.047	.024	.026	.050

Table 4.4A. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Kurtosis	Sample Size: n=15 Lower Tail	n=15 Upper Tail	n=15 Total	n=20 Lower Tail	n=20 Upper Tail	n=20 Total	n=25 Lower Tail	n=25 Upper Tail	n=25 Total
Random	0.76	-0.23	.022	.026	.048	.025	.026	.051	.024	.026	.050
Systematic	0.76	-0.23	.027	.025	052	.022	.026	.048	.026	.026	.052
Stratified	0.76	-0.23	.026	.024	.050	.024	.024	.048	.026	.025	.051
Cluster	0.76	-0.23	.022	.028	.050	.024	.027	.051	.023	.027	.050

Table 4.4B. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Kurtosis	Sample Size: n=15 Lower Tail	n=15 Upper Tail	n=15 Total	n=20 Lower Tail	n=20 Upper Tail	n=20 Total	n=25 Lower Tail	n=25 Upper Tail	n=25 Total
Random	0.76	6.97	.025	.022	047	.023	.022	.045	.023	.022	.045
Systematic	0.76	6.97	.025	.024	.049	.021	.025	.046	.022	.025	.047
Stratified	0.76	6.97	.022	.022	.044	.023	.022	.045	.027	.024	.051
Cluster	0.76	6.97	.023	.026	.049	.024	.028	.052	.022	.025	.047

Table 4.4C. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Kurtosis	Sample Size: n=15 Lower Tail	n=15 Upper Tail	n=15 Total	n=20 Lower Tail	n=20 Upper Tail	n=20 Total	n=25 Lower Tail	n=25 Upper Tail	n=25 Total
Random	0.76	13.71	.020	.024	.044	.020	.024	.044	.022	.021	.043
Systematic	0.76	13.71	.022	.025	.047	.022	.022	.044	.023	.021	.044
Stratified	0.76	13.71	.022	.020	.042	.019	.021	.040	.023	.024	.047
Cluster	0.76	13.71	.023	.028	.051	.025	.026	.051	.021	.025	.046

Table 4.5A. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	n=15 Lower Tail	n=15 Upper Tail	n=15 Total	n=20 Lower Tail	n=20 Upper Tail	n=20 Total	n=25 Lower Tail	n=25 Upper Tail	n=25 Total
Random	1.00	0.42	.022	.025	.047	.024	.025	.049	.024	.026	.050
Systematic	1.00	0.42	.028	.025	.053	.024	.026	.050	.026	.026	.052
Stratified	1.00	0.42	.025	.023	.048	.024	.025	.049	.026	.024	.050
Cluster	1.00	0.42	.020	.028	.048	.023	.029	.052	.021	.028	.049

Table 4.5B. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	n=15 Lower Tail	n=15 Upper Tail	n=15 Total	n=20 Lower Tail	n=20 Upper Tail	n=20 Total	n=25 Lower Tail	n=25 Upper Tail	n=25 Total
Random	1.00	7.07	.025	.022	.047	.024	.022	.046	.023	.022	.045
Systematic	1.00	7.07	.025	.024	.049	.020	.025	.045	.022	.025	.047
Stratified	1.00	7.07	.022	.022	.044	.023	.022	.045	.027	.024	.051
Cluster	1.00	7.07	.021	.027	.048	.024	.029	.053	.021	.025	.046

Table 4.5C. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	n=15 Lower Tail	n=15 Upper Tail	n=15 Total	n=20 Lower Tail	n=20 Upper Tail	n=20 Total	n=25 Lower Tail	n=25 Upper Tail	n=25 Total
Random	1.00	13.71	.020	.024	.044	.020	.024	.044	.022	.021	.043
Systematic	1.00	13.71	.022	.025	.047	.022	.022	.044	.023	.021	.044
Stratified	1.00	13.71	.022	.021	.043	.019	.021	.040	.023	.024	.047
Cluster	1.00	13.71	.023	.029	.052	.024	.027	.051	.021	.025	.046

Table 4.6A. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	Sample Size: n=15 Lower Tail	n=15 Upper Tail	n=15 Total	n=20 Lower Tail	n=20 Upper Tail	n=20 Total	n=25 Lower Tail	n=25 Upper Tail	n=25 Total
Random	1.26	1.34	.022	.023	.045	.024	.026	.050	.024	.025	.049
Systematic	1.26	1.34	.027	.024	.051	.024	.025	.049	.025	.026	.051
Stratified	1.26	1.34	.026	.025	.051	.023	.024	.047	.027	.024	.051
Cluster	1.26	1.34	.018	.029	.047	.021	.028	.049	.021	.028	.049

Table 4.6B. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	Sample Size: n=15 Lower Tail	n=15 Upper Tail	n=15 Total	n=20 Lower Tail	n=20 Upper Tail	n=20 Total	n=25 Lower Tail	n=25 Upper Tail	n=25 Total
Random	1.26	7.53	.024	.022	.046	.024	.022	.046	.022	.022	.044
Systematic	1.26	7.53	.024	.024	.048	.021	.025	.046	.023	.026	.049
Stratified	1.26	7.53	.027	.026	.053	.024	.024	.048	.024	.025	.049
Cluster	1.26	7.53	.020	.028	.048	.023	.030	.053	.020	.027	.047

Table 4.6C. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	Sample Size: n=15 Lower Tail	n=15 Upper Tail	n=15 Total	n=20 Lower Tail	n=20 Upper Tail	n=20 Total	n=25 Lower Tail	n=25 Upper Tail	n=25 Total
Random	1.26	13.71	.020	.023	.043	.019	.024	.043	.022	.021	.043
Systematic	1.26	13.71	.022	.025	.047	.021	.022	.043	.023	.022	.045
Stratified	1.26	13.71	.022	.021	.043	.020	.022	.042	.023	.023	.046
Cluster	1.26	13.71	.021	.029	.050	.023	.028	.051	.020	.026	.046

Table 4.7A. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	Sample Size $n=15$: Lower Tail	$n=15$: Upper Tail	$n=15$: Total	$n=20$: Lower Tail	$n=20$: Upper Tail	$n=20$: Total	$n=25$: Lower Tail	$n=25$: Upper Tail	$n=25$: Total
Random	1.50	2.37	.022	.024	.046	.023	.025	.048	.024	.025	.049
Systematic	1.50	2.37	.026	.024	.050	.024	.026	.050	.025	.025	.050
Stratified	1.50	2.37	.025	.024	.049	.024	.024	.048	.027	.023	.050
Cluster	1.50	2.37	.017	.030	.047	.018	.029	.047	.020	.028	.048

Table 4.7B. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	Sample Size $n=15$: Lower Tail	$n=15$: Upper Tail	$n=15$: Total	$n=20$: Lower Tail	$n=20$: Upper Tail	$n=20$: Total	$n=25$: Lower Tail	$n=25$: Upper Tail	$n=25$: Total
Random	1.50	8.04	.020	.024	.044	.020	.025	.045	.023	.022	.045
Systematic	1.50	8.04	.024	.026	.050	.022	.023	.045	.023	.023	.046
Stratified	1.50	8.04	.023	.022	.045	.020	.024	.044	.023	.026	.049
Cluster	1.50	8.04	.019	.031	.050	.023	.030	.053	.021	.027	.048

Table 4.7C. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	Sample Size $n=15$: Lower Tail	$n=15$: Upper Tail	$n=15$: Total	$n=20$: Lower Tail	$n=20$: Upper Tail	$n=20$: Total	$n=25$: Lower Tail	$n=25$: Upper Tail	$n=25$: Total
Random	1.50	13.71	.020	.023	.043	.018	.024	.042	.022	.022	.044
Systematic	1.50	13.71	.022	.024	.046	.020	.022	.042	.022	.022	.044
Stratified	1.50	13.71	.021	.021	.042	.020	.022	.042	.023	.024	.047
Cluster	1.50	13.71	.020	.030	.050	.022	.028	.050	.020	.026	.046

Table 4.8A. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	Sample Size *n=15* Lower Tail	*n=15* Upper Tail	*n=15* Total	*n=20* Lower Tail	*n=20* Upper Tail	*n=20* Total	*n=25* Lower Tail	*n=25* Upper Tail	*n=25* Total
Random	1.76	3.71	.022	.024	.046	.022	.024	.046	.024	.023	.047
Systematic	1.76	3.71	.025	.024	.049	.024	.026	.050	.025	.026	.051
Stratified	1.76	3.71	.022	.022	.044	.023	.022	.045	.026	.023	.049
Cluster	1.76	3.71	.016	.031	.047	.016	.030	.046	.018	.028	.046

Table 4.8B. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	Sample Size *n=15* Lower Tail	*n=15* Upper Tail	*n=15* Total	*n=20* Lower Tail	*n=20* Upper Tail	*n=20* Total	*n=25* Lower Tail	*n=25* Upper Tail	*n=25* Total
Random	1.76	8.71	.020	.024	.044	.020	.024	.044	.023	.023	.046
Systematic	1.76	8.71	.024	.026	.050	.022	.023	.045	.023	.023	.046
Stratified	1.76	8.71	.023	.022	.045	.020	.024	.044	.023	.026	.049
Cluster	1.76	8.71	.017	.032	.049	.021	.031	.052	.020	.027	.047

Table 4.8C. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	Sample Size *n=15* Lower Tail	*n=15* Upper Tail	*n=15* Total	*n=20* Lower Tail	*n=20* Upper Tail	*n=20* Total	*n=25* Lower Tail	*n=25* Upper Tail	*n=25* Total
Random	1.76	13.71	.020	.023	.043	.019	.023	.042	.022	.021	.043
Systematic	1.76	13.71	.023	.024	.047	.021	.021	.042	.022	.021	.043
Stratified	1.76	13.71	.021	.021	.042	.019	.022	.041	.022	.024	.046
Cluster	1.76	13.71	.018	.031	.049	.022	.030	.052	.020	.026	.046

Table 4.9A. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	Sample Size n=15: Lower Tail	n=15: Upper Tail	n=15: Total	n=20: Lower Tail	n=20: Upper Tail	n=20: Total	n=25: Lower Tail	n=25: Upper Tail	n=25: Total
Random	2.00	5.15	.022	.023	.045	.023	.024	.047	.024	.023	.047
Systematic	2.00	5.15	.024	.024	.048	.023	.025	.048	.025	.025	.050
Stratified	2.00	5.15	.021	.020	.041	.024	.023	.047	.026	.022	.048
Cluster	2.00	5.15	.014	.032	.046	.014	.030	.044	.017	.030	.047

Table 4.9B. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	Sample Size n=15: Lower Tail	n=15: Upper Tail	n=15: Total	n=20: Lower Tail	n=20: Upper Tail	n=20: Total	n=25: Lower Tail	n=25: Upper Tail	n=25: Total
Random	2.00	9.43	.020	.023	043	.019	.023	.042	.024	.023	.047
Systematic	2.00	9.43	.023	.025	048	.022	.023	.045	.023	.023	.046
Stratified	2.00	9.43	.023	.021	.044	.020	.024	.044	.023	.026	.049
Cluster	2.00	9.43	.015	.032	.047	.019	.031	.050	.020	.028	.048

Table 4.9C. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	Sample Size n=15: Lower Tail	n=15: Upper Tail	n=15: Total	n=20: Lower Tail	n=20: Upper Tail	n=20: Total	n=25: Lower Tail	n=25: Upper Tail	n=25: Total
Random	2.00	13.71	.025	.027	.052	.024	.025	.049	.022	.026	.048
Systematic	2.00	13.71	026	.025	.051	.025	.025	.050	.026	.027	.053
Stratified	2.00	13.71	027	026	053	.021	.025	.046	.023	.025	.048
Cluster	2.00	13.71	017	029	046	017	.030	.047	.020	.027	.047

Table 4.10A. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	n=15 Lower Tail	n=15 Upper Tail	n=15 Total	n=20 Lower Tail	n=20 Upper Tail	n=20 Total	n=25 Lower Tail	n=25 Upper Tail	n=25 Total
Random	2 26	6.96	.022	.022	.044	.020	.023	.043	.023	.023	.046
Systematic	2.26	6 96	.023	.023	.046	.023	.024	.047	.024	.024	.048
Stratified	2 26	6 96	.021	.019	.040	.023	.023	.046	.025	.021	.046
Cluster	2.26	6.96	.013	.034	.047	.014	.032	.046	.017	.030	.047

Table 4.10B. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	n=15 Lower Tail	n=15 Upper Tail	n=15 Total	n=20 Lower Tail	n=20 Upper Tail	n=20 Total	n=25 Lower Tail	n=25 Upper Tail	n=25 Total
Random	2 26	10 34	.019	.023	.042	.020	.023	.043	.023	.023	.046
Systematic	2 26	10 34	.023	.025	048	.022	.022	.044	.023	.023	.046
Stratified	2 26	10 34	022	021	043	020	.024	.044	.024	.025	.049
Cluster	2 26	10 34	.013	.034	047	.018	.031	.049	.019	.028	.047

Table 4.10C. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	n=15 Lower Tail	n=15 Upper Tail	n=15 Total	n=20 Lower Tail	n=20 Upper Tail	n=20 Total	n=25 Lower Tail	n=25 Upper Tail	n=25 Total
Random	2 26	13 71	.026	.027	.053	.024	.025	.049	.024	.026	.050
Systematic	2 26	13 71	.025	.025	.050	.024	.025	.049	.025	.025	.050
Stratified	2 26	13 71	025	025	050	023	.025	.048	.025	.024	.049
Cluster	2 26	13 71	015	031	046	.016	030	.046	.019	.028	.047

Table 4.11A. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	n=15 Lower Tail	n=15 Upper Tail	n=15 Total	n=20 Lower Tail	n=20 Upper Tail	n=20 Total	n=25 Lower Tail	n=25 Upper Tail	n=25 Total
Random	2 50	8 87	.021	.020	.041	.020	.023	.043	.022	.023	.045
Systematic	2 50	8 87	.021	023	.044	.023	.024	.047	.024	.024	.048
Stratified	2 50	8 87	.020	019	.039	.022	.023	.045	.023	.022	.045
Cluster	2 50	8 87	.010	034	.044	.013	.033	.046	.016	.031	.047

Table 4.11B. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	n=15 Lower Tail	n=15 Upper Tail	n=15 Total	n=20 Lower Tail	n=20 Upper Tail	n=20 Total	n=25 Lower Tail	n=25 Upper Tail	n=25 Total
Random	2 50	11 29	.019	.022	.041	.019	.022	.041	.024	.023	.047
Systematic	2 50	11 29	.023	.023	.046	.022	.021	.043	.022	.023	.045
Stratified	2 50	11 29	022	019	.041	.021	.023	.044	.024	.025	.049
Cluster	2 50	11 29	.011	.036	.047	.016	.032	.048	.018	.029	.047

Table 4.11C. Type I error rate when nominal rate is .05

Sampling Plan	Distribution: Skew	Distribution: Kurtosis	n=15 Lower Tail	n=15 Upper Tail	n=15 Total	n=20 Lower Tail	n=20 Upper Tail	n=20 Total	n=25 Lower Tail	n=25 Upper Tail	n=25 Total
Random	2 50	13 71	.019	.022	.041	.019	.022	.041	.023	.022	.045
Systematic	2 50	13 71	022	.024	.046	.022	.021	.043	.023	.023	.046
Stratified	2 50	13 71	.021	.020	.041	.019	.023	.042	.023	.024	.047
Cluster	2 50	13 71	.013	.034	.047	.017	.031	.048	.018	.028	.046

Table 4.12A. Type I error rate when nominal rate is .05

Distribution			Sample Size								
Sampling Plan	Skew	Kurtosis	n=15			n=20			n=25		
			Lower Tail	Upper Tail	Total	Lower Tail	Upper Tail	Total	Lower Tail	Upper Tail	Total
Random	2.76	11.28	.019	.021	.040	.021	.022	.043	.021	.022	.043
Systematic	2.76	11.28	.021	.022	.043	.023	.023	.046	.025	.024	.049
Stratified	2.76	11.28	.019	.019	.038	.023	.022	.045	.024	.022	.046
Cluster	2.76	11.28	.009	.035	.044	.011	.034	.045	.015	.032	.047

Table 4.12B. Type I error rate when nominal rate is .05

Distribution			Sample Size								
Sampling Plan	Skew	Kurtosis	n=15			n=20			n=25		
			Lower Tail	Upper Tail	Total	Lower Tail	Upper Tail	Total	Lower Tail	Upper Tail	Total
Random	2.76	12.50	.022	.023	.045	.023	.024	.047	.024	.025	.049
Systematic	2.76	12.50	.021	.022	.043	.024	.023	.047	.025	.026	.051
Stratified	2.76	12.50	.019	.022	.041	.021	.022	.043	.022	.022	.044
Cluster	2.76	12.50	.006	.034	.040	.008	.032	.040	.013	.030	.043

Table 4.12C. Type I error rate when nominal rate is .05

Distribution			Sample Size								
Sampling Plan	Skew	Kurtosis	n=15			n=20			n=25		
			Lower Tail	Upper Tail	Total	Lower Tail	Upper Tail	Total	Lower Tail	Upper Tail	Total
Random	2.76	13.71	.020	.021	.041	.021	.021	.042	.022	.021	.043
Systematic	2.76	13.71	.018	.020	.038	.021	.021	.042	.022	.021	.043
Stratified	2.76	13.71	.019	.020	.039	.020	.021	.041	.021	.021	.042
Cluster	2.76	13.71	.019	.055	.074	.023	.050	.073	.026	.048	.074

Table 4.13. Type I error rate when nominal rate is .05

	Distribution		Sample Size								
Plan	**Skew**	**Kurtosis**	***n=15***			***n=20***			***n=25***		
			L-Tail	U-Tail	Total	L-Tail	U-Tail	Total	L-Tail	U-Tail	Total
Random	3.00	13.71	.020	.020	.040	.019	.021	.040	.023	.022	.045
Systematic	3.00	13.71	.021	.021	.042	.021	.023	.044	.023	.022	.045
Stratified	3.00	13.71	.021	.019	.040	.023	.023	.046	.024	.021	.045
Cluster	3.00	13.71	.009	.036	.045	.010	.035	.045	.014	.031	.045

Table 4.14. Sampling from a Clustered Distribution Type I error rate when nominal rate is .05

			Sample Size *n=30*		
	skew	**kurtosis**	lower tail	upper tail	Total
Cluster I	0.76	-0.23			
Cluster II	1.50	8.04	.018	.018	.036
Cluster III	2.50	11.29			

Table 4.15. Sampling from a Stratified Normal Distribution

	Stratification of Distribution			Sample Size *n=30*		
	lower 1/3	middle 1/3	upper 1/3	lower tail	upper tail	Total
Stratum I	80%	10%	10%	0.000	0.891	0.891
Stratum II	10%	80%	10%	0.026	0.022	0.048
Stratum III	10%	10%	80%	0.892	0.000	0.892

Table 4.16. Sampling from a Stratified Chi-Square Distribution

	Stratification of Distribution			**Sample Size *n=30***		
	lower 1/3	middle 1/3	upper 1/3	lower tail	upper tail	Total
Stratum I	80%	10%	10%	0.000	0.891	0.891
Stratum II	10%	80%	10%	0.111	0.003	0.114
Stratum III	10%	10%	80%	0.754	0.000	0.754

Chapter Five

Conclusion

The results obtained from the sampling of nonnormal distributions indicate the importance of sampling in hypotheses testing. As the distributions varied from normal to a skew of 3.00 and kurtosis of 13.71, all of the sampling procedures became more conservative with respect to Type I error rates.

Larger sample sizes maintained power and preserved Type II error rates better than smaller samples. In almost all examples larger samples sizes (25) produced Type I error rates that were closer to nominal rates when compared to smaller sample sizes. As the distributions varied farther from normality , samples of size 25 maintained nominal alpha throughout. Larger samples almost always performed better than smller samples.

The additional analysis of stratified populations variables and clustered population variables showed even more striking results. When researchers use simple random sampling procedures on stratified and clustered populations, the t test becomes even more liberal with respect to Type I error rates in clustered populations. In stratified populations Type I error rates reached 89%. The analysis of these distributions indicate that lumpy, non-overlapping, or bi-modal populations are far greater threats to power than smooth continuous population forms (see Micceri, 1989).

When skew reached 2.00 the loss of power was significant for all plans when sample sizes were small (15). The combination of small sample sizes (15) and nonnormality created a significant lose of statistical power. Cluster sampling and to a lesser degree stratified sampling procedures, show uneven Type I error rates between upper and lower tails. Bradley (1980) found that robustness (insensitivity to assumption violations) is worse (less insensitive) in two tail t test when the population is skewed. When populations are symmetric, robustness for a two-tail test is either superior to or intermediate between the robustness of upper or lower tail test at the same alpha level. The study results also indicates that the t-test is more sensitive to Skew nonnormality than it is to kurtosis nonnormality. As the distributions became more positively skewed, lower tail Type I error rates became more conservative. This was especially true in the case of cluster sampling.

The cluster sampling and stratified sampling results reported in tables 4.1 through 4.13 lack the design considerations researchers would normally apply in field studies. Researchers would normally establish clusters or stratas based on some pre-existing knowledge of the population. Clusters would typically be formed based on some pre-existing geographic, political or social grouping.

Stratas would normally be the result of grouping by race, income ethnicity, or gender. This would have added greater precision to both cluster and stratified sampling.

Systematic sampling maintained nominal alpha better than simple random sampling. As greater positive skew was added, random sampling became more conservative. Type I error rates dropped to 3.96%, when skew was 3.00 and sample sizes were 15. Systematic sampling had Type I error rates of 4.2% for the same skew and sample size. While these procedures seem very similar, systematic sampling provided a slight advantage with greater power and near perfect balance between upper and lower tail rejections rates.

Overall all sampling plans provided randomization and preserved power when skew was below 2.00. Kurtosis ranged from the absolute minimum for each level of skew to the maximum (13.71). Sampling was done for each level of skew at the minimum kurtosis to a mid-point then to the maximum kurtosis. Varying the level of kurtosis seems to have only minor effect on Type I or Type II error (compare tables 4.12A, 4.12B and 4.12C).

Research design should include consideration for sampling plan. It should also include considerations for power, alpha, sample size and effect size. In order to make these decisions researchers must have some indication of the shape of the population variable of interest. They need to know a priori if the distribution of the variable of interest is so nonnormal that it poses a threat to nominal Type I error rate. This can be determined from pilot studies, prior research, or research on correlated variables. Hopkins and Weeks (1990) suggest that measures of skew and Kurtotsis should be a routine part of data analyses and reporting. When this data is reported the reader can better assess the statistical conclusions of the research.

Cohen's power tables and sample design options provide a framework for research planning and design. The researcher can determine the appropriate level of power based on experimental data versus data from quasi experiments or field studies. Effect sizes can be based the expected treatment effect size. The researcher should have some hypothesized treatment effect size. Alpha can be set based on convention or consideration for the risk of Type I versus Type II error consequences. Sampling procedures should be selected based on some knowledge of the population of interest.

When Hypotheses testing is conducted using sample data the results are totally dependent on the value of the sample elements. Ronald Carver (1978, 1993) is a severe critic of hypotheses testing. He suggest that hypotheses testing be completely abandoned in social science research. This conclusion might be extreme, however He offers the following summary of the t test when used in test of statistical significance.

$$(X_1 - X_2) / SD_{(\bar{x}_1 - \bar{x}_2)}$$

The numerator is the mean difference or effect size. The denominator is a measure of sampling error. This equation could be written as follows:

Effect size/Standard error of the sample

Carver suggest that the t test is a ratio of effect size to standard error. This equations reveals a flaw in hypotheses testing when researchers fail to report effect size separate from significance results. When effect sizes are large and the standard error of the sample is large the resulting t is small. If the same effect size is obtained with a small sample standard error, the resulting t value is larger. If the standard error of the sample is large as a result of poor sampling procedures and not a reflection of the variability in the population, Type II error rates will result. The resulting loss of power is a direct result of a poor sample plan.

The relationship between effect size and sample standard error demonstrates the importance sampling has on hypotheses testing. When a researcher conducts Hypotheses testing he in effect accepts the sample as a reflection of the population to which results will be generalized. Hypotheses testing makes no attempt at verifying or connecting sample data with real world statistics. The researchers must accept the data at face value. The results obtained from a single test of hypotheses are therefore ambiguous. This problem is made more compelling when the underlying assumption of normality is violated. Sample design considerations are therefore critical to hypotheses testing.

When hypotheses tests are based on samples, results may not generalize to the population or replicate in other samples or hypotheses tests. When the researcher comes across nonnormal data he must be guided by considerations for the impact on hypotheses testing as well as generalizability. Researchers need to always be aware that hypothesized results are only valid to the extend that they reflect the target population. Replication in other samples and other experiments confirm the theoretical results obtained from test of hypotheses.

When researchers are unaware or ignore clustering or stratification in population variables, the implications for hypotheses testing can be disastrous. Tables 4.14. and 4.15. present sample results based on sampling from stratified populations (4.15.) and clustered populations (4.14.).

Table 4.14. presents results obtained from sampling a clustered population with different skew and kurtosis in each cluster. The results indicate that the t test becomes more conservative when samples are from these types of populations. Actual Type I error rate is 3.6% when the

nominal rate is 5.00%. This distribution is similar to distributions actually encountered in real world populations (Micceri, 1989). Distributions that are lumpy and multi-modal, could yield similar results when sampled.

When researchers sample stratified populations and ignore strata important information is over looked and a significant opportunity to increase sample precision is lost. This study stratified the population in thirds. This was done to illustrate the effect of stratification. If the stratification was done to create stratas of vastly different sizes, the results could have been even more dramatic. A researcher conducting sampling on true population variables could mistakenly ignore stratas and sample from either strata I or strata III as in table 4.15. For example, if a researcher wanted to sample psychologists and only sampled psychologists in private practice, the sample results would achieve randomization, but only within one strata. The sample would include urban and rural psychologists, psychologists with large medium and small practices, different ethnic groups would be included and different income levels would be reflected. However, significant information in different strata would be ignored. The research would exclude psychologists working in clinics or hospitals, industrial psychologists working in private industry, psychologists working in public schools, psychologists teaching at universities, and all of the licensed psychologists who work in areas totally unrelated to psychology. If this occurred the Type I error rate could be similar to the results presented in table 4.15.

Bibliography

Alsawalmeh, Y.M., Feldt, L.S. (1992). Test of the hypotheses that the interclass reliability coefficient is the same for two measurement procedures. *Applied Psychological Measurement* , 16, 195-205.

Sadooghi-Alvandi, M. (1996). The choice of subsample size in two-stage sampling. *Journal of the American Statistical Association,* 81, no. 394, 555-558.

Armitage, B.A. (1947). A comparison of stratified with unrestricted random sampling from a finite population. *Biometrika,* 34, 273-280.

Bankier, M.D. (1986). Estimators based on several stratified samples with applications to multiple frame surveys. *Journal of the American Statistical Association,* 81, no. 396, 1074-1079.

Blair, R. C. (1981). A reaction to "Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance." *Review of Education Research,* 51, 499-509.

Boneau, A.C.(1960). The effects of violations of assumptions underlying the t test. *Psychological Bulletin, 57,* 49-64.

Box, G.E.P. (1953). Non-normality and tests for variance. *Biometrika,* 40, 318-335.

Bradley, J.V. (1977). A common situation conducive to bizarre distribution shape. *American Statistician,* 31, 147-150.

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

Bradley, J.V. (1980). Nonrobustness in z, t, and f tests at large sample sizes. *Bulletin of the Psychonomic Society,* 16, 333-336.

Bradley, J.V. (1982). The Insidious L-shaped distribution. *Bulletin of the Psychometrics Society,* 20, 85-88.

Campbell, D.T. and Stanley, J.C. (1963). Experimental and quasi-experimental designs for research. Massachusetts:Houghton Mifflin.

Carver, R.P. (1978). The case against statistical significance testing. *Harvard Educational Review,* 48, no.

3, 378-399.

Carver, R.P. (1993). The case against statistical significance testing, revisited. *Journal of Experimental Education,* 61, 287-292.

Chalmer, B.J. (1987). *Understanding Statistics*. New York:Marcel Dekker.

Cochran, W.G. (1977). *Sampling Techniques*(3rd ed.). New York:Wiley.

Cochran, W,G. (1947). Some consequences when the assumption for the analysis of variance are not satisfied. *Biometrics,* 3, 22-39.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York:McGraw-Hill.

Cohen, J. (1992). A power primer. *Psychological Bulletin,* 112, no. 1, 155-159.

Cohen, S.A. and Hyman, J.S. (1979). How come so many hypotheses in educational research are supported? (a modest proposal). *Educational Research,* 50, 12-16.

Coren, S. and Hakstain, A.R. (1990). Methodological implications of interaural correlation: Count heads not ears. *Perception and Psychophysics,* 48, 291-294.

Dar, R. Serlin, R.C. and Omer, H.(1994). Misuse of statistical tests in three decades of psychotherapy research. *Journal Consulting Clinical Psychology,* 62, 75-82.

Diaconis, P. and Bradley, E. (1983). Computer intensive methods in statistics. *Scientific American, 248,*116-130.

Fay, R.E. (1990). Alternative paradigms for the analysis of imputed survey data. *Journal of the American Statistical Association,* 91, no. 434, 490-498.

Fleishman, A.I. (1978). A Method for Simulating Non-Normal Distributions. *Psychometrika,* 43, 521-532.

Fowler, R.L. (1992). Using the extreme groups strategy when measures are not normally distributed. *Applied Psychological Measurement,* 16, 249-259.

Frankel, M. (1983). Sampling theory. *Handbook of Survey Research*. 21-67.

Games, P.A. and Lukas, P.A. (1966). Power of the analysis of variance of independent groups on nonnormal and normally transformed data. *Educational and Psychological Measurement,* 26, 311-327.

Garret, H.E. (1966). *Statistics in Psychology and Education.* New York:McKay.

Gay, L.R. (1992).*Educational Research*(4th ed.). New York: Macmillian.

Gayen, A.K. (1949). The distribution of 'student' *t* in random samples of any size drawn from non-normal universes. *Biometrika,* 36, 353-369.

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

Glass, G.V. ,Hopkins, K.D. (1996). *Statistical methods in education and psychology* (3rd ed.). Massachusetts:Allyn & Bacon.

Groves, R.M. (1989). *Survey Errors and Survey Costs.* New York:Wiley.

Hammond, G.(1996). The objections to null hypothesis testing as a means of analysing Psychological data. *Austrialian Journal of Psychology,* 48, 104-106.

Hinkle, D.E. Wiersma, W., and Jurs, S.G. (1994). *Applied Statistics for the behavioral sciences.* Boston:Houghton Mifflin.

Hopkins, K.D. and Weeks, D.L. (1990). Tests for normality and measures of skewness and kurtosis: their place in research reporting. *Educational and Psychological Measurement,* 50, 717-729.

Hsu, L.M. (1989). Random sampling, randomization, and equivalence of contrasted groups in psychotherapy outcomes research. *Journal of Consulting and Clinical Psychology,* 57, no. 1,131-137.

Jaeger, R.M. (1984). *Sampling in education and the social sciences.* New York:Longman.

Kazdin, A.E. and Bass, D. (1989). Power to detect differences between alternative treatments in comparative psychotherapy outcomes research. *Journal of Consulting and Clinical Psychology*, 57, no. 1, 138-147.

Kerlinger, F.N.(1986). *Foundations of Behavioral Research* (3rd ed.). Fort Worth:Harcourt Brace Jovanovich.

Kish, L. (1965). Survey Sampling. New York:Wiley.

Kish, L. and Frankel, M.R. (1974). Inference from Complex Samples *Journal of the Royal Statistical Society,* B36, 1-22.

Kish, L. (1989). Deffs: why, when and how? a review. From the proceedings of the survey research section of the American statistical Association, Washington.

Lindguist, E.F. (1952). *Design and analysis of experiments in psychology and education*. Boston:Houghton Mifflin.

Luh, W. and Olejnik, S. (1990). Two-stage sampling procedures for comparing means when population distributions are non-normal. Paper presented at the American Educational Research Association meeting.

Mansfield, E. (1973). *Statistics for Business and Economics* (2nd. ed). New York:Norton.

McGraw, K.O. and Wong, S.P. (1992). A common language effect size statistic. *Psychological Bulletin,* 111, no. 2, 361-365.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin,* 105, 156-166.

Neyman, J. (1934). On two different aspects of the representative method. *Journal of the Royal Statistical Society,* 97, 558-625.

Pearson, E.S. (1929). The distribution of frequency constants in small samples from nonnormal symmetrical and skew populations. *Biometrika,* 21, 259-286.

Pearson, E.S. and Please, N.W. (1975). Relation between the shape of population distribution and the robustness of four simple test statistics. *Biometrika,* 63, 231-241.

Penfield, D.A. (1994). Choosing a two-sample location test. *Journal of Experimental Education. 62,* 343-360.

Prentice, D.A. and Miller, D.T. (1992). When small effects are impressive. *Psychological Bulletin,* 112, no. 1, 160-164.

Rao, J.N.K. and Scott, A.J. (1992). A simple method for the analysis of clustered binary data. *Biometrics,* 48, 577-585.

Rasmussen, J.L. and Dunlap, W.P. (1991). Dealing with nonnormal data: Parametric analysis of transformed data vs. nonparametric analysis. *Educational and Psychological Measurement,* 51, 809-819.

Rogan, J.C., Keselman, H.J. and Breen, L.J. (1976). Assumption violations and rates of type I error for the tukey multiple comparison test; a review and empirical investigation via a coefficient of variance variation. *Journal of Experimental Education,* 46, 20-25.

Roger, R.S. (1978) Two-stage sampling to set sample size for post hoc tests in anova with decision-based error rates. *British Journal of Mathematics and Statistics,* 31, 153-178.

Rogers, W.T. and Hopkins, K.D. (1988). Power estimates in the presence of a covariate and measurement error. *Educational and Psychological Measurement,* 48, 647-656.

Royall, R.M. (1986). The prediction approach to robust variance estimation in two-stage cluster sampling. *Journal of the American Statistical Association,* 81, no. 393, 119-123.

Runyon, R.P. and Haber, A. (1991). *Fundamentals of Behavioral Statistics* (7th ed.). New York: McGraw Hill.

Sawilowsky, S.S. (1990). Nonparametric tests of interaction in experimental design. Review of Educational Research, 60, no. 1, 91-126.

Sawilowsky, S.S. and Blair, R.C. (1992). A more realistic look at the robustness and type II error properties of the t test to departures from population normality. *Psychological Bulletin,* 111, 352-360.

Sawilowsky, S.S. and Hillman, S.B. (1992). Power of the independent samples *t* test under a prevalent psychometric measure distribution. *Journal of Consulting and Clinical Psychology,* 60, no. 2, 240-243.

Scheffe, H.(1959). *The Analysis of variance*. New York:Wiley.

Scott, A.J. and Holt, D. (1982). The effect of two-stage sampling on ordinary least squares methods. *Journal of the American Statistical Association,* 77, 848-854.

Serlin, R.C.(1993). Confidence intervals and the scientific method: a case for holm on the range. *Journal of Experimental Education, 61,* 350-360.

Shao, J. (1996). Bootstrap model selection. *Journal of the American Statistical Association,* 91, no. 434, 655-659.

Shaver, J.P. (1993). What statistical significance testing is, and what it is not. *Journal of Experimental Education,* 61, 293-316.

Sigel, S. (1956). *Non-parametric statistics for the behavioral sciences.* New York:McGraw Hill.

Snedecor, G.W. and Cochran, W.G. (1967). *Statistical Methods* (6th ed.). Ames Iowa: Iowa State University Press.

Strube, M.J. (1991). Small sample failure of random assignment: a further examination. *Journal of Consulting and Clinical Psychology,* 59, 346-350.

Student, (1908). The probable error of a mean. *Biometrika,* 6, 1-25.

Student, (1927). Errors of routine analysis. *Biometrika,* 19, 151-163.

Sukhatme, P.V. and Sukhatme, B.V. (1970). Sampling theory of surveys with applications (2nd ed.) Ames Iowa:Iowa State University Press.

Tadikamalla, P.R. (1980). On simulating non-normal distributions. *Psychometrika,* 45, no. 2. 273-279.

Thomas, D.R. (1993). Inference using complex data from surveys and experiments. *Canadian Psychology,* 34, 415-431.

Walker, H.M. and Lev, J. (1969). *Elementary Statistical Methods,* (3rd ed.). New York:Holt Rinehart and Winston.

Wendell, J.P. and Schmee, J. (1996). Exact inference for proportions from a stratified finite population. *Journal of the American Statistical Association,* 91, no. 434, 825-830.

Wilburn, A.J. (1977) *Practical Statistical Sampling for Auditors.* New York:Dekker.

Wright, R.L.D. (1976). *Understanding Statistics.* New York:Harcourt Brace Jovanovich.

Zumbo, B.D. and Zimmerman, D.W. (1991). Further evidence for Coren and Hakstain's "Methodological implications of interval correlation: Count heads not ears" and an alternative correction formula. *Perception and Psychophysics,* 50, 290-301.

ABSTRACT

SAMPLING PROCEDURES AND TYPE I ERROR RATES
(FOR NONNORMAL POPULATIONS)

by

WILLIAM CADE

May 1998

Advisor: Dr. Shlomo Sawilowsky

Major: Evaluation and Research

Degree: Doctor of Philosophy

This study focuses on sampling as an essential part of research in education and the social sciences. The results obtained from sampling are generalizable to the extent that the sample reflects the population to which the results will be applied. Underlying the generalizability theory is the assumption that the population of interest is normally distributed. Many researchers contend that distributions in the social sciences do not conform the to assumptions of normality. These researchers find that population variables that violate the underlying assumption of normality can effect results when real data distributions are encountered.

The question in this research is how can sampling procedures minimize the effects of non-normal distributions when the violations are sufficiently large enough to affect results. This study addresses issues related to hypotheses testing specifically, the effect on the t-test for

independent means when sampling from non-normal distributions. Random sampling, systematic sampling, stratified sampling and cluster sampling procedures are examined. A Monte Carlo simulation is used to compare these procedures. Type I error rates are computed for each procedure. Sample sizes of fifteen, twenty, and twenty-five are used for each sampling plan. The results from sample simulations are presented in tables for comparison.

AUTOBIOGRAPHICAL STATEMENT
WILLIAM CADE

Work Experience:

Ameritech/ Michigan Bell Telephone Company 1979-1996

Manager-Regulatory and Pricing (1990-1996)

Manager-Strategic Planning, Financial Assurance (1984-1990)

Marketing Account Representative (1979-1984)

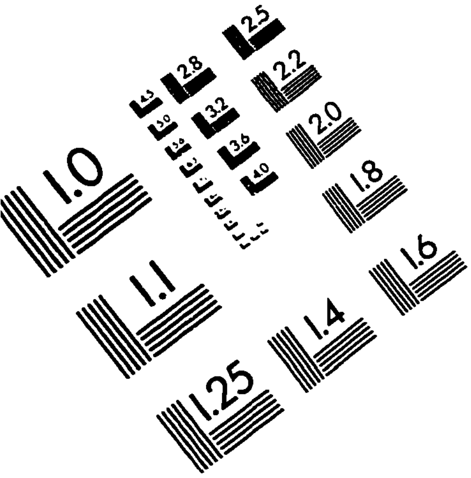
Education:

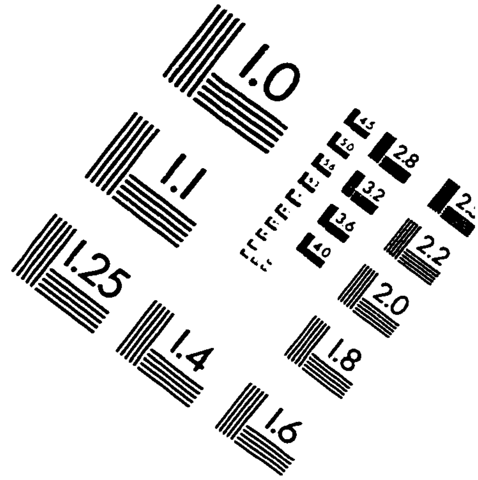
Ph.D., Evaluation and Research, Wayne State University
(dissertation to be completed 12/97)

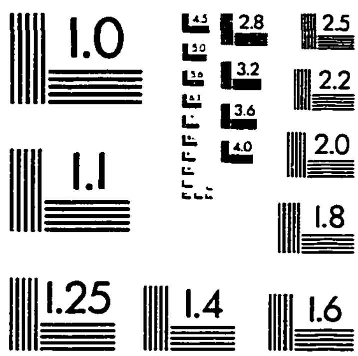
MSA, Management Science, Central Michigan University, 1989

BGS, Political Science, University of Michigan, 1971

IMAGE EVALUATION TEST TARGET (QA-3)

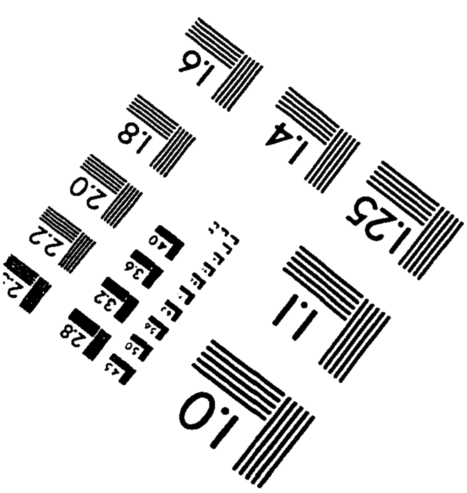


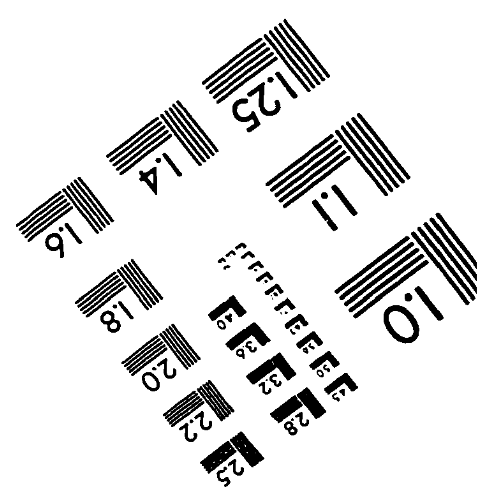




150mm

6"





APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989