4-2001

An Analysis of the Impact of an Outlier on Correlation Coefficients Across Small Sample Data Where RHO is Non-Zero

Maria A. Suchowski
Western Michigan University

AN ANALYSIS OF THE IMPACT OF AN OUTLIER ON CORRELATION COEFFICIENTS ACROSS SMALL SAMPLE DATA WHERE RHO IS NON-ZERO

by

Maria A. Suchowski

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Educational Studies

Western Michigan University
Kalamazoo, Michigan
April 2001

AN ANALYSIS OF THE IMPACT OF AN OUTLIER ON CORRELATION COEFFICIENTS ACROSS SMALL SAMPLE DATA WHERE RHO IS NON-ZERO

Maria A. Suchowski, Ph.D.

Western Michigan University, 2001

This study addressed the problem of the probable effectiveness of the Pearson correlation coefficient (r) as an estimator of moderate or strong population correlation (rho) when that estimate is based on small sample data which contains an outlier. In such a situation, three components contribute to the size of a sample correlation coefficient, and so to the subsequent effectiveness of the resulting estimation decision. These components are 1) rho, 2) sample size, and 3) outlier. Considered in this study were: two conditions of rho (.5 and .8), three sample sizes (10, 30 and 50) and two outlier conditions (without outlier and with outlier).

The investigation was conducted by simulating the distribution of Pearson r's under each condition and observing its behavior. Each sample distribution was characterized by values of central tendency, dispersion and skew. Each distribution was also summarized in terms of a hit rate which indicated the percentage of times the confidence interval about its sample r's contained the known population rho. The nominal expected hit rate was 95%.

Results indicated that in the condition without outlier measures of central tendency were close to rho across all sample sizes and for both conditions of rho. Hit rate was very close to the expected 95% across all study conditions.

In the condition with outlier, measures of central tendency were not close to

rho, and were farther from rho as sample size became smaller. Hit rate was considerably smaller than the expected 95%, particularly when rho was .5 rather than .8. When rho was .5, the hit rate was 73% at sample size 10, 83% at sample size 30 and 87% at sample size 50. When rho was .8, the hit rate was 84% at sample size 10, 90% at sample size 30, and 92% at sample size 50.

The implication of these results for the practical investigator is that if an outlier appears in small study data, the risk of making an incorrect decision is substantially increased particularly when rho is moderate.

UMI Number: 3007026

UMI®

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

ACKNOWLEDGMENTS

A few years ago, I was sitting at the sunny edge of a glassy sea, and took it into my mind that I might like to cross the waters. So I built a little boat, squinted into the sun, and cast out onto the open sea.

I would like to thank the many who kept me afloat when the sun went down. Thank you to Dr. Uldis Smidchens for years of lighting the way. Thank you to Dr. Paul Lane for good spirits during the journey. Thank you to my committee Dr. Mary Anne Bunda, Dr. Jianping Shen, and Dr. Michael Stoline for navigating and bailing when the boat took on water and repairs were being made. Thank you, in particular, to Dr. Bunda for steering to land.

It's been quite a ride.

Maria A. Suchowski

TABLE OF CONTENTS

Table of Contents—continued

APPENDICES

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The Pearson product moment correlation coefficient (Pearson r) is the most widely used index of bivariate correlation (Sheskin, 1997; Chou, 1989; Liebetrau, 1983). Operating under the assumption that related random variables are normally distributed, when a random sample is drawn from this normal population, the sample correlation coefficient (r_{xy}) is an estimator of the population correlation coefficient (ρ_{xy}). When sample sizes are large and sample data consistent, the estimator is a robust one (McCallister, 1991; Trochim, 1997). Although large consistent samples are available to investigators in many situations, in some experimental or testing situations encountered in disciplines such as education, social analysis, and in the medical sciences, investigators may be limited to small samples. When sample sizes are small and data contains an anomaly, the effectiveness of the Pearson estimator is less certain. This uncertainty suggests the utility of an improved understanding of the estimator under conditions occurring in the small sample as well as the large one. One of these real world conditions is the occurrence of the outlying value in sample data. This study assesses the effectiveness of the Pearson sample correlation coefficient as an estimator of moderate or strong population correlation when the estimate is based on small sample data which contains an outlying value. The study context is correlational validity in measurement.

There are three components that are fundamental contributors to the size of a sample correlation coefficient, and so to the subsequent effectiveness of the resulting estimation decision. The first of these components is rho (ρ_{xy}), the size of the conceptual absolute correlation between two sets of measures x and y. The second of these components is sample size, the size of the sample on which the sample estimate r_{xy} was computed. Because any factor that affects the range of x or y affects the size of the resulting correlation coefficient, the third of these components is any outlying value, a data point within the sample that affects the range of sample points.

While in theory samples should be large and data free of anomalies, in practice, samples are customarily small to moderate and data are frequently less than perfect. In practical application, departures from normality are common (Hill & Dixon, 1982; Micceri, 1989) and random samples drawn from real world distributions can have deviant unusually small or large values among a sample of observations (Wilcox, 1997). Such an outlying value, lying apart from the rest of the sample observations is termed an outlier, as illustrated in Figure 1 below.

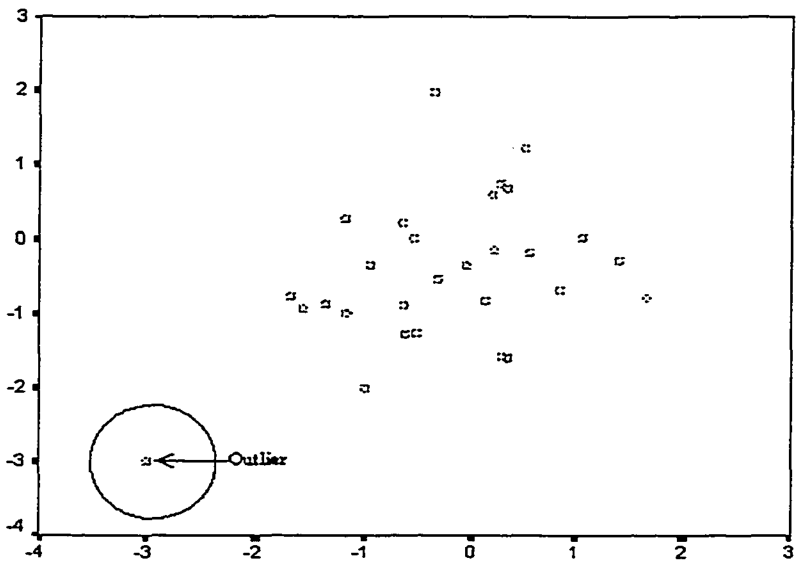


Figure 1. Example of Outlier in Sample Data.

Although in applied research, sample data sometimes contains an outlier (Barnett & Lewis, 1995; Statsoft, 1998), the source of this observation may be unknown to the investigator. This uncertainty results because any of several conditions may result in the appearance of an outlier in sample data. Sample data containing such an outlier may be a reflection of random variation in a normal distribution, reflect a non-normal distribution with heavy tails, reflect a distribution with larger variance than the one assumed, or may reflect a skewed or mixture distribution (Draper & Smith, 1981; Hamilton, 1992).

The effectiveness of estimation under non-ideal conditions has been a long standing problem for methodological researchers (Nevitt & Tam, 1997). A major topic in measurement research which has both applied and theoretical interest is the attenuation of coefficient values by a variety of conditions. The restrictions of range resulting from violations of the assumptions of linearity and homoscedasticity have received substantial attention in the measurement literature (Busby & Thompson, 1990; Fowler, 1987; McCallister, 1991). The problem of correcting correlation coefficients for various conditions has also received a good deal of attention (Alexander, Carson, Alliger & Carr, 1987). By altering the range of a data set, an outlier can cause a reduction or enhancement in the correlation coefficient (Armstrong & Frome, 1977; Rousseeuw & Leroy, 1987; Hubert & Rousseeuw, 1996; Rousseeuw & Hubert, 1996). The influence of such a point becomes larger as the sample size gets smaller (McCallister, 1991).

It is generally understood that when there is no correlation in the parent population and the sample size is large, the sample correlation coefficient is relatively insensitive to violations of underlying assumptions (Kowalski, 1972; Chou, 1989). An outlier may not be problematic. On the other hand, when there is some degree of association in the parent population and the sample size is small, the effectiveness of the coefficient as an estimator is less clear. Some writers suggest that the Pearson coefficient is robust and powerful even under extreme violations of normality assumptions (Fowler, 1987). Other researchers suggest that the effectiveness of the Pearson coefficient is reduced in the presence of a data anomaly (McCallister, 1991; Halperin, 1986).

These conflicting reports are not very meaningful for the practical investigator who is interested in detecting moderate or strong correlation in a population with imperfectly known characteristics. Specifically, the study of measurement validity presumes a correlation between two indices rather than no such correlation (Hopkins, 1998).

Measurement concerns can be exacerbated in small sample problems because much statistical theory is based on large sample proof (Mooney, 1997). Secondly, in correlation estimation, much statistical theory is based on conceptual population symmetry in which the population correlation is equal to zero. If the population correlation in a practical research situation is not equal to zero, evaluation may not be clear. Thirdly, in the situation in which a small sample contains an outlier, the outlier may influence commonly used tests for the correlation coefficient, further complicating the evaluation decision. The question of whether an outlier makes a difference to the effectiveness of estimation when appearing in small sample data drawn from a population with moderate or strong correlation remains to be assessed. For this evaluation, the sampling distribution of the sample correlation coefficient is needed.

The only general way through which to make this evaluation is through Monte Carlo simulation which assesses and compares estimators empirically under specified criteria (Borjas & Sueyoshi, 1994; Kleijnen, 1992; Yu & Dunn, 1982). The principal advantage of simulation is that this method provides information about both the population correlation parameter and the sampling distribution of the sample correlation coefficient under stipulated conditions (Davidson & MacKinnon, 1993).

The problem addressed in this simulation study was the probable effectiveness of the Pearson sample correlation coefficient (r_{xy}) as estimator of moderate or strong population correlation (ρ_{xy}) when sample size is small and when that sample data contains an outlier. These three components 1) population correlation, 2) sample size and 3) outlier condition, were systematically manipulated through the study. From a practical perspective, this study could have meaningful implications for both a) applied investigators concerned with assessment or analysis of research data and b) planning of a research design.

The study dealt with the following questions:

1. When ρ_{xy} is moderate (.5) what is the distribution of r_{xy} ?
   a. What is the effect of sample size on the distribution of r_{xy} ?
   b. What is the effect of an outlier on the distribution of r_{xy} ?
2. When ρ_{xy} is strong (.8) what is the distribution of r_{xy} ?
   a. What is the effect of sample size on the distribution of r_{xy} ?
   b. What is the effect of an outlier on the distribution of r_{xy} ?
3. When ρ_{xy} is moderate how is the efficacy of the estimate of ρ_{xy} affected by: (a) sample size and (b) an outlier?
4. When ρ_{xy} is strong how is the efficacy of the estimate of ρ_{xy} affected by: (a) sample size and (b) an outlier?

CHAPTER II

LITERATURE REVIEW

Overview

This chapter first discusses the correlation coefficient in the context of the estimation framework in which practical investigations are conducted. Second, the chapter reviews the influences of population correlation and sample size on the goodness of estimation in the normal case. Third, the chapter surveys empirical research conducted on the goodness of estimation in the non-normal case. Fourth, the chapter examines the related but separate estimation problem of an outlier occurring in sample data. Finally, the chapter closes with a summary of remaining needs and the question of interest.

Investigative Framework

Correlation Coefficient

Much work in the investigative sciences is devoted to the discovery of important associations or relationships. In addition to anticipating that some relationship exists, investigators seek to quantify that association. Quantification is a basic contributory step in the prediction process which is the central purpose of scientific inquiry (Liebetrau, 1983). Bivariate correlational procedures measure the degree of association between two variables. (Sheskin, 1997; Chou, 1989; Cohen & Cohen, 1983; Gibbons, 1993).

Among the oldest and most frequently used measures of association is the

Pearson product moment correlation coefficient, the most widely used measure of association (Liebetrau, 1983; Gibbons, 1993; Trochim, 1997). Use of this correlation coefficient depends on assumptions made about the variables of interest and about the population from which the sample is drawn. Under Pearson's correlation, the usual assumptions are that the underlying bivariate population is one in which both variables are normally distributed and the relationship between variables is linear (Chou, 1989; Tabachnik & Fidell, 1989). Under these assumptions, the size and effectiveness of the sample estimate then depends on the magnitude of population correlation, the adequacy of sample size drawn, and the consistency of sample data. These three components of the sample correlation estimate vary in practice.

Measurement research, particularly, is predicated on the investigation of anticipated relationship. The sample correlation coefficient (r_{xy}) is a validity coefficient between a test score (x) and a criterion score (y), providing a basis for the correctness or goodness of inference made on that measure. Similarly, the sample correlation coefficient between two sets of observations provides a coefficient of reliability for those observations (Brown, 1983; Crocker & Algina, 1986; Hopkins, 1998). Because correlational studies are concerned with the detection of an anticipated relationship between variables (Liebetrau, 1983), no correlation or very low correlation provides no useful information.

Correlation Strength

While the relationship between variables is a necessary one for correlational research, the useful magnitude of the relationship varies with the practical situation under investigation. The higher the correlation coefficient, the better the predictor, but the size of the correlation coefficient need not be high in order to provide useful information. Moderate correlation may be of use. The usefulness of this magnitude varies with the study.

The interpretation of this magnitude also varies with the study. There is no fixed definition of what constitutes a moderate or high degree of correlation. Because the scale of values for the correlation coefficient is ordinal, correlation values are relative not absolute. Depending on application and on interpretation, a correlation coefficient of .5 may be termed low, moderate or substantial. In general, however, a moderate correlation is considered to be in the range of coefficient values falling about .5. Similarly, a rule of thumb for interpreting correlation to be high is a correlation coefficient about .8 (Nunnally, 1967; Hinkle, Wiersma, & Jurs, 1988; Hopkins, 1998).

In validity studies, moderate correlation may provide good information. Similarly, correlation between various measurement forms is generally found to be moderate in practice. For example, in a study of the relationship between of test scores and other ratings, Bridgeman and Harvey (1998) report correlations of .5, .68, .57 and .53 across several samples. In a study of predictive validity, Simner (1992) reports an average correlation of .49. Powers (1986) reports between test correlations ranging from .34 to .60.

By contrast, strong correlation is more typically encountered in cognitive measurement (Hopkins, 1998). For example, in a study of a cognitive instrument, Drudge (1981) reports uniformly high indices of cognitive measures ranging from .78 to .94. Correlation is also typically strong in reliability studies, and in fact, should be as strong as possible (Hopkins, 1998). For example, in a literature search of 22 studies with 258 different reliability coefficients, Jiang (1997) found that median reported coefficients centered on a correlation of .8. Of 21 reported internal consistency coefficients, the median was .956; of 151 reported inter-rater reliability coefficients, the median was .806; and in one reported case, test-retest reliability was .801.

Sample Size

As with the question of what constitutes moderate or strong correlation, the question of what constitutes an adequate sample size has no fixed answer. Typically, correlational studies are not based on very large samples. It is assumed that if a relationship exists, it will be evident in a small or moderate sample (Liebetrau, 1983). In practice, what constitutes adequate sample size varies with the investigator. Correlation coefficients based on sample sizes as small as 10 or 20 are reported (Leknes, Lie, Boe, & Selvig, 1997; Smith & Knudtsen, 1990; Breen, Rogers, Lawless, Austin, & Johnson 1997). In a search of 87 ERIC studies containing the key words: tests, correlation and validity, approximately one third of these studies (30) reported results based on samples of a size smaller than 60. Of these, two predominant size groupings were 1) studies of size 20 to 36 (median size 27) and 2), studies of size 43 to 56 (median size 48). In addition, many larger studies reported correlations based on small subsets of larger samples.

It has been suggested that no definite answer to the question of adequate sample size is available because so many mediating factors exist (Hinckle, Wiersma & Jurs, 1988; Chou, 1989; Freedman, Pisani, Purves & Adhikari, 1991)). The sample size needed depends on many factors: the level of significance or confidence required, the power of the test, the probability of rejecting the null hypothesis (confidence and accuracy required), the probability of accepting the null hypothesis when it is false, and the population variance. Churchill (1991) notes that there is often a balance between precision, confidence and sample size in applied investigation, and that applied researchers often work with somewhat imprecise estimates.

It is commonly accepted that a sample size should be 25 or greater in order to have satisfactory approximations. A size of 30 is frequently recommended (Mendenhall, Wackerly, & Scheaffer, 1986; Chou, 1989). While a sample size of 30 is generally accepted as a size at which the central limit theorem begins to operate, providing a probability basis for estimation, confidence in obtained estimates is increased with sample size (Chou, 1989). In studies of confidence, McNemar (1962) and Hays (1981) indicate that approximation of standard error about the sample correlation coefficient begins to be accurate only when sample sizes are at 50, and recommend this somewhat larger sample size minimum for correlational work.

Outliers

In addition to consideration of rho and of sample size in the correlation estimation decision, investigators dealing with real world data may encounter an anomaly in an observed sample. Correlational study data may contain an outlier. Although there is no available true appraisal of degree to which data is consistent in practice, the large body of work devoted to the detection and treatment of outliers indicates the magnitude of interest in this source of uncertainty (Barnett & Lewis, 1995). In a set of observations, if one or more of those observations stands out in contrast to the other observations, this outlying observation is termed an outlier. Although the term outlier is often not rigorously defined (Motulsky, 1997), outliers are generally conceptualized as either observations which do not fit the pattern in the rest of the data, deviant observations, or extreme values which can bias the estimates of x and y (Barnett & Lewis, 1995). Apart from the rest of the data, outliers are seen as extreme cases on one variable or a combination of variables which can have an influence on the calculation of statistics (Wulder, 1996; SAS Institute, 1998). The term outlier is used collectively for discordant observations and for contamination (Iglewicz & Hoaglin, 1993).

Sources of Outliers

Outliers may arise in any of three general ways. (Anscombe 1960; Barnett 1978; Beckman & Cook 1983; Grubbs, 1969): 1) inherent data variability, 2) measurement error and 3) execution error. Data variability may include random variation in a normal distribution, distributions with heavy tails, distributions with larger variance than assumed, and skewed or mixture distributions. Measurement error and execution error are among possible data contaminants.

Some researchers distinguish between outliers caused by variation inherent to the population being studied and spurious observations caused by contaminants, measurement or execution errors (Anscombe 1960; Gideon & Hollister, 1987). Others make no such distinction (Barnett & Lewis, 1995), arguing that in practical applications it may not be possible to identify such causes. Both conceptual approaches tend to conclude that in no field of observation is it possible to entirely rule out the possibility that an observation is marred by investigator error.

Apart from outliers caused by outright mistakes in recording or coding, which can be eliminated or corrected, outliers resulting from other sources are dealt with in a wide variety of ways. In an analysis of 35 applied studies, Ulrich and Miller (1994) counted 17 different outlier treatment methods or criteria used by various investigators, and reported that none of the studies gave explicit reasons for the procedure selected. The distinction between erroneous and valid data, clear in theory, may become blurred in practice.

In the real world setting, the form of underlying distributions may be only imperfectly known. Measures of population characteristics are frequently normally distributed, but the distribution may have thicker or longer tails than expected. In such a case, an outlier may be generated from either the right or the left tail of the underlying distribution. Alternately, the underlying distribution may be somewhat skewed. Data may be characterized by a floor effect in which observations are grouped at the lower end of a distribution. For example, income data, with a floor of 0, in which most people have an average income and some few have a large income, may be right skewed. In such a case, a positive outlier may be generated from the right tail. Assessment data, for example test scores, may be left skewed, distributed as a compressed group of high scores with a long tail descending through lower scores. In such a case, a negative outlier may be generated from the left tail. The negative outlier may be particularly problematic because it may indicate the failure of a test to fully identify performance, or may reflect unanticipated variance, or may reflect simple random variation. For this reason, in a correlational study, the negative outlier is of particular interest.

Although an outlier in data presents an estimation problem, there is no uniformly accepted way in which to accommodate an outlier. While there is a variety of techniques suggested in the literature for the treatment of outliers, there is no mathematical or generally accepted resolution available to the investigator to determine cause or source of the outlier (SAS Institute, 1998). Additional evaluation of the data may be needed to determine whether the data in question are the result of common cause or special cause variation (DOE, 1997).

The practical problem for the investigator is not a trivial one since the presence of an outlier may cause bias (Motulsky, 1997). On the other hand, eliminating data may artificially raise the size of the experimental effect (error variance decreases while the difference between means remains constant), or produce a lower reliability of estimators through a smaller sample size, of particular importance when the sample is already small. Most outlier treatment procedures lead to a shift or bias about the mean (Kohnert, 1995).

Estimation in the Underlying Normal Case

Normal Assumption

While the practical investigator does not deal with populations in which there is no anticipated correlation, frequently relies on small samples, and may encounter an outlying value in small sample data, the theoretical framework around the Pearson r which guides estimation decisions is most exact when the population correlation is 0, when the sample size is large and when sample values are internally consistent. Under the assumption that the sample is drawn from a normal bivariate population, the sample correlation coefficient tends to be distributed approximately normally as: 1) the sample size increases, and 2) as values of the population correlation coefficient approach 0.

Sampling Distribution of the Estimator

Like any other statistic, the Pearson correlation coefficient (r) has a sampling distribution. If a large number of paired measurements were sampled and the Pearson r computed for each sample, the resulting Pearson coefficients would form a distribution of r's. When the absolute value of the correlation in the population is 0, and the sample is large, then the sampling distribution of Pearson r's is approximately normal. The normal distribution then furnishes a practical method of computing approximate probability values associated with arbitrarily distributed random variables (Chou, 1989). If the population correlation is 0 and sample size is large, then the resulting normal distribution of Pearson r's and a known testing statistic (t) is used to compute approximate probability that a given sample correlation coefficient provides a good estimate of population correlation.

If the value of the population correlation coefficient (ρ) is not 0, then the distribution of Pearson r's is no longer normally distributed and centered on 0, but tends to become skewed and peaked, centered on the value of ρ. With higher positive values of correlation in the population, the distribution of Pearson r's takes on a negative skew. With higher negative values of correlation in the population, the distribution of Pearson r's takes on a positive skew. If the population correlation is greater (or smaller) than 0, then inferences about the population correlation (ρ) are made by transforming the asymmetric distribution of Pearson r's into an approximately normal distribution. This z transformation then again provides a proven testing statistic (Fisher, 1915).

The pattern assumed by the distribution of Pearson r's is illustrated for correlations of $-\rho$, 0 or $+\rho$ in Figure 2 below. When there is no population correlation, the sampling distribution of the estimator, Pearson r's, is approximately normal, centered on $\rho = 0$. But when population correlation is positive (or negative), $\rho > 0$ (or $\rho < 0$), the sampling distribution of Pearson r's tends to move away from center (0), skew, and build over $+ \rho$ or $- \rho$ (Hinkle, Wiersma & Jurs, 1988).

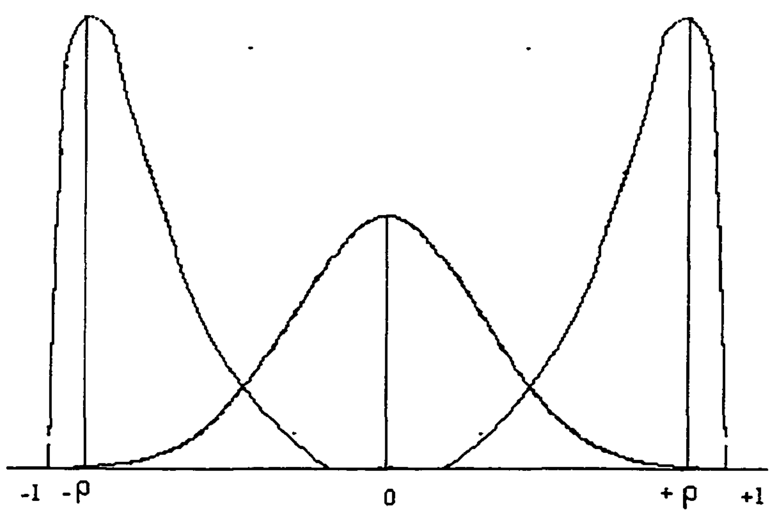


Figure 2. Example Sampling Distribution of the Correlation Coefficient When ρ is < 0, 0, and >0.

Approximations improve as sample size becomes larger. As the sample size becomes larger, the extremity of the tail becomes smaller with fewer observations appearing in the tail, while the kurtosis of the distribution is increased. Both sample size and ρ influence the shape of a particular sampling distribution of Pearson r's (Belsley, 1996; Chou, 1989; Edwards, 1976).

In order for estimates based on the sampling distribution of Pearson r's to be valid, the underlying assumption that the sample is drawn from a normal distribution must be met. Pearson (1929) showed that the Pearson r is stable in samples as small as 20 or 30 even if rho is not equal to 0 when samples are drawn from a normal distribution.

Estimation in the Non-Normal Case

Sampling from Non-Normal Distributions

The robustness of the Pearson r under anomalous conditions has been a topic of theoretical and applied interest for many years. Much initial interest revolved about the goodness of the Pearson r when sampling is from non-normal distributions, when the assumption of normality is violated. Hey (1928) reported that considerable non-normality will not affect the distribution of correlation coefficients. Gayen (1951) noted that these results are valid only when rho = 0. The investigation of robustness was extended by Cheriyan (1945) to the case where a sample is drawn from a non-normal distribution in which rho is moderate to large (.5, .75, .89). Cheriyan (1945) found close agreement with normal theory expectation for certain values of r, but found that as rho increased, the agreement with normal theory decreased. Haldane (1949) reported that if rho is small then non-normality does not affect the distribution of Pearson r's if non-normality is restricted to skewness, but showed by example that when rho was large a small change in kurtosis may affect the variance of Pearson r's. Extending the investigation to consideration of sample size as well as rho, Gayen (1951) showed that when rho = 0 the effect of non-normality is not serious even for samples as small as eleven, but found less agreement when rho is not equal to 0. Kowalski (1968, 1972) reported that there is general agreement that the distribution of Pearson r is quite robust to non-normality when rho = 0, but that there is good evidence that this is less so for increasing values of rho, especially if the non-normal distribution displays a high degree of kurtosis.

Sampling From Mixed Distributions

With the advent of Monte Carlo simulation and the computer, the investigation of the stability of the Pearson r under irregular circumstances was extended to the study of mixture distributions. In a simulation study considering samples drawn from populations which were either not normal or were mixtures, Devlin, Gnanadesikan and Kettenring (1975) considered the finite properties of Pearson r (and others) in two sample sizes (20 and 60) and three values of rho (0, .5 and .9). Results indicated that mean Pearson r's were very close to rho for both sample sizes and for all values of rho, although somewhat closer to rho in the larger sample size. Mixture contaminants tended to affect mean Pearson r's by a fraction of a percentage point, slightly more when rho was .5 as opposed to .9. Wainer and Thissen (1979), also investigated the robustness of estimators in simulation trials by using mixture distributions. Sample sizes of 50 and 100 were considered across rhos of 0, .5 and .9. When 10% and 20 % contaminants in which rho = 0 were introduced, the mean of r's for rho of .5 and .9 was substantially reduced. The reduction was somewhat greater when rho was .5. Results were identical for both sample sizes. The mean of Pearson r's was influence primarily by the extremity of the contaminant.

Kowalski (1972) illustrated that the distribution of Pearson r's in samples from mixtures of bivariate normal distributions may depart from the corresponding normal density even when rho = 0, if the mixture is extreme enough. Additional studies have continued to show that even a small proportion of extreme contaminants can influence computed estimates (Armstrong & Frome, 1977; Hubert & Rousseeuw (1996), Rousseeuw & Leroy, 1987; Rousseeuw & Hubert, 1996).

A number of studies also appeared on the robustness of tests of the correlation coefficient in a bivariate normal distribution under the prospect of contamination by outlying sample values. Duncan and Layard (1973) showed by the Monte Carlo method that the test of $\rho = 0$ is not robust under extreme contamination conditions, consistent with asymptotic theory. Srivastava and Lee (1984) studied the robustness of tests and estimators when the parent population is non-normal or when observation from a normal population are contaminated by observations from a normal population with different variance properties (mixing two bivariate normal distributions with zero means but with different variances). The test of $\rho = 0$ based on Student's t, Fisher's z, Arcsine, or Ruben's transformation was shown numerically to be non-robust when the proportion of contamination is between 5 and 50% and the contaminating population has a large variance compared to the bivariate normal population. Srivastava and Lee (1984) found that tests are very sensitive to even one contaminant if the variance of the contaminating population is very large, but the degree of seriousness in the distortion of tests varies with the extremity of the contaminant. This again implies that commonly used tests of $\rho = 0$ are very sensitive to the presence of extreme values if the values are large enough. Robust tests for the correlation coefficient are also described in Tiku and Balakrishnan (1986) who report that tests for rho = 0 are robust to small departures from normality but not robust to large departures.

Tiku (1987) extended the robust test for rho equal to 0 to the situation where rho is not equal to 0 (rho = .5 and .7) across various sample sizes (20, 30, 45, and 60), and reports that tests of rho ≠ 0 have the same drawbacks as for testing rho = 0, particularly where the mixing population has a high variance and the percentage of mix is large. These studies consistently show that very large outlying values in sample data drawn from non-normal or mixed distributions can have a negative effect on the goodness of estimation based the violation of normality. The effect of sample size and rho on estimation, although considered in some studies, is less certain, tending to be overshadowed by the extremity of the contaminant mixture.

Estimation in the Observed Data Case

Outlier in Sample

The effects of anomalies on the robustness of the Pearson r studied through non-normal or mixture models are typically based on the premise that the whole sample comes from a different distribution than the normal or from a distribution which is composed of mingled elements (Srivastava & Lee, 1984). A separate problem of practical interest is the case where there is an outlier in an observed sample rather than the sample being drawn from a population which is not normal or which is mixed with another population. Srivastava and Lee (1985) examined the problem of one outlier appearing in an observed sample by considering N independent observations in which N-1 of them were from a bivariate normal with a given mean and variance and one of them was from a bivariate normal with multiples of that variance. N's of 10, 20 and 40 were considered. Rho was zero. Srivastava and Lee (1985) found that if N was greater than 20, then the Pearson r was robust in the presence of a variance multiplier of 3, but was seriously affected by an outlier with a variance multiple of 9, even when N = 40. Results showed that when observations have an outlier, the distribution of Pearson r's has larger tails than that which would be expected under normal theory. Both sample size and the extremity of the outlier were found to combine in effect on the robustness of the Pearson r. The effect of rho was not considered.

Conception of Outlier

The distinction made by Srivastava and Lee (1985) between robustness of the Pearson r when sampling from non-normal distributions and robustness when an outlier observed in sample data is a meaningful one. In contrast to much theoretical work which has tended to be based on large outliers from non-normal distributions or from mixed distributions, the outlying value in an observed sample needs neither to be from a non-normal distribution nor to be very large in order to be defined as an outlier. The investigative problem extends to the question of how to characterize the outlier. An outlier may not be very large and still be apart from other sample data. The underlying distribution may be unknown.

Although concepts of outlier measurement vary, Hanneman (1998) suggests that outliers are usually at core defined in terms of deviation from an expected value, plus or minus three standard errors or three standard deviations. Three such deviations are highly improbable (Chaloner & Brant, 1988; Nevitt & Tam, 1997).

The Univariate Condition

Characterization of the outlier is straightforward in the univariate case, but is more complex in the bivariate case which is essential to correlation. In the univariate case, the univariate boxplot or box and whiskers plot developed by Tukey (1977) is a well known graphical tool for the isolation of possible outliers. In a data display, the boxplot describes data points included in the interquartile range. It consists of a box drawn from the lower quartile of data points to their upper quartile with a crossbar at the median. The box length is the interquartile range. The fence reaches from the lower quartile minus 1.5 times the interquartile range (IQR) to the upper quartile plus 1.5 times the interquartile range (IQR). Points lying outside of the fence are considered outliers. Under this 1.5(IQR) rule, values between 1.5 and 3 box lengths from the upper or lower edge of the box are flagged as outliers. Figure 3 below shows the Tukey boxplot which here identifies an outlier at three standard deviations from the mean when the mean is zero.

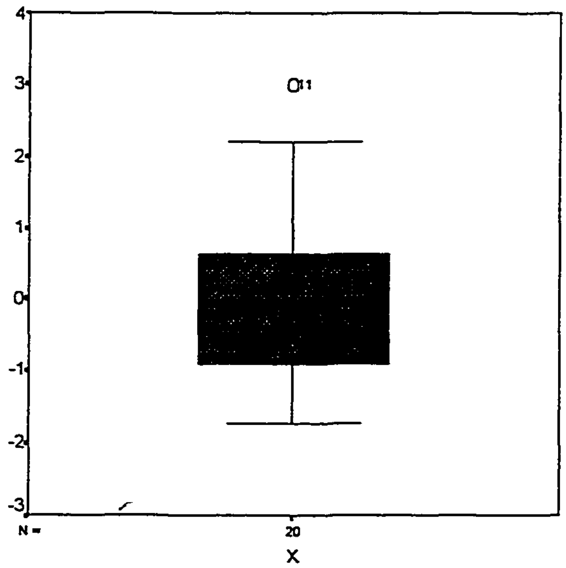


Figure 3. Boxplot Identifying Outlier at 3 Standard Deviations From Mean.

The Bivariate Condition

Identification is exacerbated in the bivariate case. There has been less development in estimation methods for bivariate distributions and correlation than for univariate methods (Johnson, Kotz, & Balakrishnan, 1997; Kocherlakota & Kocherlakota, 1997). The univariate boxplot is based on ranks since the box is described by the value with rank n/4 to that with rank 3n/4. In the bivariate case, there is no unique ranking of values. The degree of data extremeness is not intuitive.

Because there is no unique ranking system in bivariate data, a bivariate outlier does not need to be extreme on both measures. It is possible for a measure to be extreme on the x measurement but not on the y measurement. While in bivariate data there is no unique form of total ordering, however, several different types of ordering principles have been defined and employed. (Barnett & Lewis, 1995). When extremeness is relative to the normal correlation model, bivariate outliers will be governed by the population correlation. Because of this governing relationship, bivariate sample measures will tend to associate in the same direction and to be related in size. As an outlier is extreme on the x measurement, it will also tend to be correspondingly extreme on the y measurement. This relationship is illustrated in Figure 4 below. Although the outlier is apart from the mass of the data, this point still tends to conform to the operating relationship or shape, and to be circumscribed by this relationship.

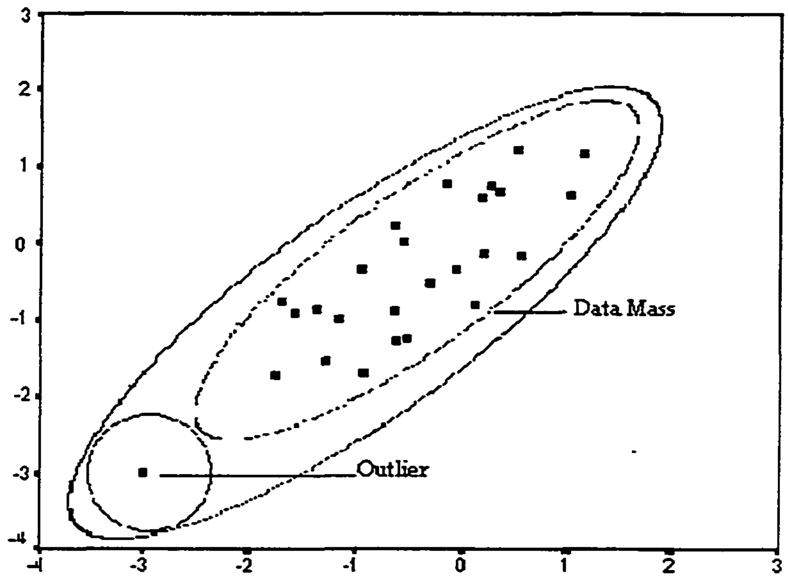


Figure 4. Example of an Outlier Relative to Mass of Data Points.

Several bivariate plots have been proposed for the description of outliers (Rousseeuw & Ruts, 1997). A bivariate boxplot has been suggested by Becketti and Gould (1987), and modified by Lenth (1988). This suggested extension of the univariate boxplot is based on the median and the quartiles for the two variables. These numbers are used to draw horizontal and vertical lines on the scatterplot, forming a cross and rectangle, which replaces the familiar Tukey box. Goldberg and Iglewicz (1992) suggest a bivariate generalization of the boxplot in which the box is an ellipse, based on data which is assumed to be elliptically symmetric.

A more general approach which does not rely on elliptical models has been suggested by Rousseeuw and Ruts (1997) who propose a generalization of Tukey's univariate box and whiskers plot in which the box is a bag. The Tukey median is the deepest data location surrounded by a bag which contains the n/2 observations closest to the median. These observations are described as having the greatest depth. Magnifying the bag by a factor of 3 yields the fence. Observations outside the fence are considered outliers. Under this conceptualization, an outlier is a data point that is away from the mass of the data.

In a somewhat different approach, Hyndman (1996) suggested a plot based on data density. The central data mass is described by a contour that encompasses 50% of the mass. Typically, the 50% region and the 99% region are superimposed on the scatterplot of the data. In this approach, an outlier is a point lying in an empty area.

For the boxplot, ellipse and bagplot, the outlier is a point lying far away from the bulk of the data. Based on the Tukey median, the boxplot and bagplot (Rousseeuw & Ruts, 1997; Rousseeuw, Ruts, & Tukey, 1999) are stable, which makes them particularly useful for describing outliers based on a concept of distance from the data mass. All of these approaches, including that of Hyndman (1996) data density, are based on a graphical core which describes 50% of the data and a fence beyond which data points are considered to be improbable, and are described as outliers.

The box- or bag-plot fence, which provides a boundary beyond which data points are considered to be outliers, is a probability based rule in a variety of distributions. Assuming that a data set is approximately normally distributed, about 1 data point in 100 would be classified as an outlier using the 1.5(IQR) rule. Assuming a standard normal distribution with a mean 0 and standard deviation 1, this outlying point fall outside a z value of +/- 2.68.

Similar reasoning can be employed using the Rousseau & Ruts (1997) bagplot, or Hyndman 99% rule. At core, an outlier is a value at an improbable distance from the center of the mass, and can be measured in terms of z scores, or population standard deviation in an infinitely large population. By all rules, in an infinite normally distributed population with mean 0 and standard deviation 1, a data value falling beyond 3 population standard deviations is highly improbably. If a sample is drawn from this distribution, a sample data point with a value beyond 3 population standard deviations is also highly improbable. The probability that a value exceeds 3z is very small and such a value can be regarded as an outlier.

Limitations in Empirical Research

While in practical data evaluation a value at three standard deviations from data center is typically identified as an outlier, empirical research conducted on outlier effects tends to be more extreme in its definitions. The characterization is an important one since it has been shown that contaminants will damage the robustness of the Pearson r, irrespective of sample and population conditions, if the mixture is extreme enough (Kowalski, 1972). More materially, while population correlation, sample size and data consistency combine together in practical importance for the investigator, empirical research which considers the combined effect of all three variables is limited.

It has been well demonstrated that the Pearson r is robust if rho is 0 and sample size is large (Havlicek & Peterson, 1977; Edgell & Noon, 1984). It has also been shown that the Pearson r is robust even if rho is not 0 and sample size is relatively small if the underlying population distribution is normal, that is, if there is no violation of the normality assumption (Pearson, 1929).

The focus of robustness research with respect to the effect of an anomaly has been on the effects of non-normality either on the distribution of the sample correlation coefficients (Kowalski, 1972; Fowler, 1987) or on formal inference concerning the population correlation coefficient (Duncan & Layard, 1973; Devlin, Gnanadesikan, & Kettenring 1975). There is general agreement that the Pearson r sample correlation coefficient is relatively insensitive to violations of normality assumptions when the population correlation is equal to 0 (Gayen, 1951; Kowalski, 1972).

The robustness of the Pearson's r when the population correlation is not equal to 0 is less extensive. (Kraemer, 1980). Evidence tends to show that the Pearson r is less robust to non-normality for increasing values of rho (Kowalski, 1972). The robustness of the correlation coefficient in a mixture model has been variously examined by Devlin, Gnanadesikan and Kettenring (1975), Wainer and Thissen (1979), Muirhead (1980), Srivastava and Awan (1980), Gnanadesikan and Kettenring (1981) and Srivastava and Lee (1984) with major emphasis on multivariate methods which are recalculated until some convergence criterion is reached. These studies provide evidence that mixture contaminants tend to affect mean Pearson r's more for moderate than for strong values of rho.

The specific problem of an outlier appearing in an observed sample rather than an entire sample being drawn from a non-normal distribution has been little considered. Srivastava and Lee (1985) investigating this problem when rho = 0 found the Pearson r robust to a small outlier but not robust to a large one. The effect of rho ≠ 0 is an unanswered question.

A number of published Monte Carlo simulation studies concerning outlier methods (Bush, Hess & Wolford, 1993; Miller, 1988, 1991; Ulrich & Miller, 1994; Van Selst & Joliceour, 1994) have been conducted, but most focus on specific cases and special methods. Results are presented either in the framework of statistical theory or by examining estimator performance on limited exemplary data sets (Nevitt & Tam, 1997). Additional comparative studies assess estimators of correlation but do not consider anomalies in sample data (Yu & Dunn, 1982). The optimality and robustness of various estimators and designs used in these studies often hold only asymptotically. Less is known about the behavior or robust estimators and tests for finite samples (Muller, 1997).

In addition, the small sample distribution of Pearson's r is, in general, little investigated (Belsley, 1996). When sample size is small, the sampling distribution of coefficients tends to be biased downwards, be skewed, and have varying degrees of kurtosis making estimation less certain. Belsley (1996) found that in a small sample relative to the 5% expected in each tail when using a 10% two tailed test for correlation, more typically, 5.5% might appear in the lower tail and 4.5% in the upper. The combination of sample size, rho and outlier is essentially little understood.

Remaining Needs and Rationale

While there has been a large amount of interest and published work on the subject of the treatment of outliers in a wide range of data contexts, there has been little published work on outliers in finite samples (Barnett & Lewis, 1995; Barnett, 1992, 1993). If outliers arise in infinite populations whose properties are known, then the probability of the occurrence of a given value can be assessed through the use of probabilistic procedures. If outliers occur in sampling from an unknown population then how to reflect a value which is atypical in the sample but is merely a facet of the finite population is a fundamental dilemma with undefined principles (Barnett, 1992).

Computer methods and simulations are important to this study (Muller, 1997). In particular, small and intermediate sample properties of robust procedures almost always have to be determined by empirical sampling. In addition, ordinary variance is not an adequate measure of performance of a robust estimator because it too is sensitive to extreme tail behavior. In order to assess robustness of an estimator for finite sample sizes, it is necessary to use Monte Carlo techniques to obtain accurate categories (Huber, 1964; Andrews et al., 1972). The effect of an outlier in the context of anticipated correlation and small sample size is a largely unexplored question.

Study Focus

In the situation in which an investigator is working with large sample data drawn from a population with known characteristics, there is a large body of theoretical work to support the correlation estimation decision. If data is homogenous, then the estimation decision is also made in accordance with known behavior based on compliance with underlying assumptions.

In the situation in which an extreme data point occurs in data, if this point is deviant enough to be considered a contaminant from another distribution, or is considered to be an error, then inclusion or exclusion of that data point becomes the practical problem for the investigator. In such a case, the estimation decision may be of less practical importance than the treatment decision.

If, on the other hand, a point is deviant in relationship to the sample data, but is not readily identifiable as a contaminant from outside the model under consideration, then direction from existing estimation studies is limited. If, in addition, the sample is a small one from a population in which a moderate or large relationship between two variables is anticipated, the estimation decision is still less clear. This was the situation of interest in this study.

Available research indicated that there are three characteristics which influence the Pearson estimate of rho. The first of these is the strength of rho itself. When values of rho considered in non-normality studies were moderate, typically defined at .5, and strong, generally defined to fall in a range from .7 (Tiku, 1987) to .9 (Devlin, Gnanadesikan, & Kettenring, 1975), Pearson estimates were dissimilar. The second influencing characteristic is sample size, variously investigated in samples as small as 10 (Belsley, 1996) and as large as 100 (Thissen, 1979). Study results tend to indicate that there was little influence on estimation due to sample size in samples over 50, but that the influence of sample size on estimation was less clear in small sample sizes. The third characteristic which affected estimation was the presence of an outlier, which was shown to influence the Pearson estimator when rho was 0, particularly in the small sample (Srivastava & Lee, 1985).

CHAPTER III

DESIGN AND METHOD

Overview

This study used a simulation methodology to empirically investigate the problem of an outlier appearing in an observed sample by assessing the behavior of the sampling distribution of Pearson r's under three sets of conditions. These conditions incorporated the three characteristics of interest: rho, sample size and outlier.

As discussed, particular assumptions were proscribed on the variables of interest. Population correlation was assumed to be moderate or strong, moderate defined as .5, strong defined as .8. Sample size was assumed to be small, defined as 10, 30 or 50. An outlier, defined to fall at about 3 SD below the mean, occurred in study data. Under these circumstances the study considered the following questions:

1. When $\rho_{xy} = .5$ what is the distribution of r_{xy} ?
   a. What is the effect of sample size on the distribution characteristics of r_{xy} ?
   b. What is the effect of an outlier on the distribution characteristics of r_{xy}?
2. When $\rho_{xy} = .8$ what is the distribution of r_{xy} ?
   a. What is the effect of sample size on the distribution characteristics of r_{xy} ?
   b. What is the effect of an outlier on the distribution characteristics of r_{xy}?
3. When $\rho_{xy} = .5$ how is the efficacy of the estimate of ρ_{xy} affected by: (a) sample size and (b) an outlier?

4. When ρ_{xy} = .8 how is the efficacy of the estimate of ρ_{xy} affected by: (a) sample size and (b) an outlier?

As a result of the posed questions, several types of data were needed. These data included 1) sets of correlated data pairs from distributions with known characteristics, 2) Pearson r's calculated on each set of pairs, and 3) distributions of these calculated Pearson r's.

Because these data are not readily available in the real world, data used in this study were computer generated, data which were the product of a simulation model. A simulation model is a computer program representing a process or system, written for the purpose of experimenting (Chisman, 1992; Watkins, 1993; Widman, Loparo & Nielsen, 1989; Law & Kelton, 1991; Fishman, 1996; Mooney, 1997). Simulation provided an empirical alternative when there was little or no real world data available. Simulation also provided a way to systematically design and control for each of the characteristics of interest, then to replicate those conditions over multiple trials. (Please see Appendix A for a description of the computer procedure.)

Data Pairs

Overview

First was needed a correlated data pair from a known distribution with a fixed rho. Computer syntax was written for the generation of such data pairs based on a bivariate normal distribution: x, N=0,1; y, N=0,1. The number of data pairs generated depended on the condition of sample size. The kind of data pair generated depended on the condition of outlier influence. (Please see Appendix B for an illustration of the computer syntax used in the generation of data pairs.)

Influence of Size on Pairs

The condition of sample size was specified to be 10, 30 or 50. The 10 case condition represented the very small sample size, based on minimal sizes sometimes encountered in practice. The primary purpose for inclusion of this very small size was to provide a range of observable effects due to sample size (May & Hittner, 1997). The 30 case condition was consistent with both a small sample typically found in practice and with the minimal requirements of the central limit theorem. The 50 case condition was consistent with both a small sample typically found in practice and with the suggestion of McNemar (1962) and Hays (1981) who indicate that approximation of standard error about the sample correlation coefficient begins to be accurate only when sample sizes are at 50.

A number of pair sets was generated for each condition of sample size: 10, i.e. 10 (x,y) pairs were generated in the set, 30 or 50,. Pearson's r was calculated on each sample of pair sets. The Pearson r's were saved.

Influence of Outlier on Pairs

The outlier condition was one of two conditions. The first condition was the simple random condition in which no outlier was specified. The second condition was a deterministic condition in which one outlier was specified to occur.

In the no outlier condition a Pearson's r was directly calculated on each pair set. In the outlier condition, one random pair was first removed from each pair set. The pair was replaced with a randomly generated outlier pair. For that outlier pair, x was randomly generated to be an outlier that had a mean of –3 and a standard deviation of .1 across all x's that were generated, and y was generated with respect to the rho in the condition. (Please see Appendix C for an illustration of the computer syntax used to generate the outlier pair.) The outlier was defined to occur on the negative end of the distribution in conformance with the leftward test skew encountered in a measurement situation. The size and deviation of the outlier was designed to fall outside a z value of +/- 2.7 in agreement with the discussed classification of outlier using the 1.5IQR A Pearson's r was then calculated on each sample pair set in the outlier condition. The Pearson r's were saved.

Formation of Pearson r Distributions

Pair Set Size 10, 30, 50

Pair sets were re-generated 1,000 times in each unique condition. Although there are no general guidelines for the number of trials needed in order for experimental results to be valid, since simulation results are unbiased for any number of trials (Hope, 1968; Mooney, 1997), the power of any assessment increases as the number of cases or trials increases. Many trials are commonly used in estimator simulation studies because a large number of trials increases power and because large computer capacity is increasingly available. 1,000 trials were run, consistent with current usage (Mooney, 1997).

Distributions of Pearson r's were formed. Each distribution had 1,000 cases but varied by the size of the pair sets in each condition of sample size: 10, 30 or 50.

Pair Set for Two Outlier Conditions

Pair sets were re-generated 1,000 times. Three distributions of Pearson r's were formed in the no outlier condition. Each distribution had 1,000 cases where pair sets varied by the condition of sample size, 10, 30 or 50. Three additional distributions of Pearson r's were formed in the no outlier condition. Each distribution had 1,000 cases where pair sets varied by the condition of sample size, 10, 30 or 50, and each pair set had one determined outlier.

Pair Set of r = .5 or r = .8

Moderate correlation was defined to be .5 and strong correlation was defined to be .8. Each of the three distributions of Pearson r's, no outlier (10, 30, 50) and three distributions of Pearson r's, deterministic outlier (10, 30, 50) was generated for each of two levels of rho (.5 and .8). Resulting was a total of 12 distributions of Pearson r's.

Characterization of Pearson r Distributions

The distributions of Pearson r's represented each of 12 conditions (3X2X2). Considered were three (3) levels of sample size, two (2) levels of outlier, and two (2) levels of rho.

Descriptors of Effect

Certain descriptive aspects of the Pearson r distributions were relevant to study questions 1 and 2. For each distribution, calculated was the mean, median, range, interquartile range, and skewness. (Skewness was computed on the mean as n/(n-1)(n-

2)sum(x- $\bar{x}$)^3/s^3).

Study Questions 1. and 2. with Corresponding Analytical Techniques

1. When $\rho_{xy} = .5$ what is the distribution of r_{xy} ?
   a. What is the effect of sample size on the distribution of r_{xy}?

      (1) Display measures of central tendency (mean, median) and skewness across sample size for each of 10, 30, 50 in each outlier condition.

      (2) Display measures of dispersion (range, interquartile range) across sample size for each of 10, 30, 50 in each outlier condition.
   b. What is the effect of an outlier on the distribution of r_{xy} ?

      (1) Display measures of central tendency (mean, median) and skewness in the no outlier condition with measures of central tendency (mean, median) and skewness in the outlier condition across sample size for each of 10, 30, 50.

      (2) Display measures of dispersion (range, interquartile range) in the no outlier condition with measures of dispersion (range, interquartile range) in the outlier condition across sample size for each of 10, 30, 50.
2. When $\rho_{xy} = .8$ what is the distribution of r_{xy} ?
   a. What is the effect of sample size on the distribution of r_{xy} ?

      (1) Display measures of central tendency (mean, median) and skewness across sample size for each of 10, 30, 50 in each outlier condition.

      (2) Display measures of dispersion (range, interquartile range) across sample size for each of 10, 30, 50 in each outlier condition.
   b. What is the effect of an outlier on the distribution of r_{xy} ?

      (1) Display measures of central tendency (mean, median) and skewness

in the no outlier condition with measures of central tendency (mean, median) and skewness in the outlier condition across sample size for each of 10, 30, 50.

(2) Display measures of dispersion (range, interquartile range) in the no outlier condition with measures of dispersion (range, interquartile range) in the outlier condition across sample size for each of 10, 30, 50.

Efficacy of Estimate

Overview

The primary purpose of this study was an assessment of the efficacy of the Pearson estimate of rho under described conditions, the problem of an outlier occurring in an observed sample. This assessment was made in terms of the probability of making an incorrect decision. The probability of making an incorrect decision, in turn, was the error rate, computed as the difference between the sample correlation estimate, Pearson r, and the rho being estimated (Chou, 1989; Mooney 1997).

Confidence Intervals are Built on Individual r's

Error rate estimation was done by testing whether or not the known rho value was excluded from the confidence interval for each Pearson r, when a nominal α was set for the confidence interval. In this procedure, α was set at .05. Upper and lower limits of the confidence intervals for each Pearson r in the distribution of 1,000 generated within a specific condition were calculated using Fisher's (1915) transformation of the Pearson r distribution. This method of transformation was such as to give the distribution of z a close approximation to a normal curve with a standard deviation depending only on the size of the sample and not on rho; as the size of the sample increases, deviation becomes nearly independent of rho, accurate for even very small sample sizes provided that rho is not too near unity (Pearson, 1929; David, 1938; Kraemer, 1973). (Please see Appendix D for detailed description of the z transformation and confidence interval calculation.)

Hit Rate

If rho was captured in the confidence interval about a Pearson r, the capture constituted a hit. If rho was not captured in the confidence interval about r, a miss occurred. The number of hits and misses occurring for each distribution of 1,000 Pearson r's was recorded. The number of recorded hits in each distribution was divided by 1,000 to produce a percentage hit rate for that distribution developed under given rho, sample size and outlier condition. This hit rate constituted the efficacy of the estimator under a given condition. If there were no effects of size or outlier, the hit rate under the specified nominal α would have been 95%. The hit rate was used to operationally define efficacy of the estimator in questions 3 and 4.

Study Questions 3. and 4. with Corresponding Analytical Techniques

3. When $\rho_{xy} = .5$ how is the efficacy of the estimate of ρ_{xy} affected by:
   a. sample size. Display hit rate across sample size for each of 10, 30, 50 in each outlier condition.
   b. an outlier. Display hit rate in the no outlier condition with hit rate in the outlier condition across sample size 10, 30, 50.
4. When $\rho_{xy} = .8$ how is the efficacy of the estimate of ρ_{xy} affected by:

a. sample size. Display hit rate across sample size for each of 10, 30, 50 in each outlier condition.

b. an outlier. Display hit rate in the no outlier condition with hit rate in the outlier condition across sample size 10, 30, 50.

Summary

Using simulation, sets of data pairs were computer generated from a known distribution with a fixed rho (.5 or .8). The number of data pairs in a set was 10, 30 or 50. The kind of data pair in a set was either with or without outlier. Pearson's r was calculated on each pair set. Pair sets were regenerated 1,000 times; Pearson r's were calculated. Distributions of 1,000 r's each were formed. Each of the twelve distribution of Pearson r's represented one of twelve specified conditions.

Descriptors of central tendency, dispersion and skewness were computed on each distribution. Descriptors were displayed and synthesis of results shown.

Confidence intervals were computed on Pearson r's in each distribution of 1,000 Pearson r's, using a nominal α of .05. The number of times the known rho was captured or was not captured by the confidence interval was counted. In each distribution, the number of times the known rho was not captured when divided by 1,000 represented the effectiveness of the estimator under conditions of that distribution. Effectiveness was displayed, and synthesis of results shown.

CHAPTER IV

RESULTS

Overview

Descriptors and rates were calculated on each of twelve simulated distributions of Pearson r's. Each distribution was composed on 1,000 Pearson r's. Each distribution represented one of 12 conditions (3X2X2). These conditions were three (3) levels of sample size, two (2) levels of outlier, and two (2) levels of rho.

The following chapter presents results of the simulation. Results are presented in terms of the measures which define the characteristics of each distribution. These measures are central tendency, dispersion and skewness. Measures of Pearson's r_{xy} are presented.in tabular form. Tables show central tendency (mean, median), skewness (on mean) and dispersion (range and interquartile range). Additional tables show the effectiveness of the Pearson estimator, measured by the hit rate, under conditions of each distribution.

Each distribution of Pearson r's was characterized by particular values of central tendency, dispersion and skewness. As an introduction to the tables which display particular values for each of the distributions, the following histogram shows one such distribution of r's as an example in order to illustrate the location of each of the measures of interest appearing on the histogram. (All twelve histograms are displayed organized by sample size in Appendix E, and are also presented organized by outlier condition in Appendix F.) The following example histogram shows the simulated distribution of Pearson r's in which rho is .5 and sample size is 10.

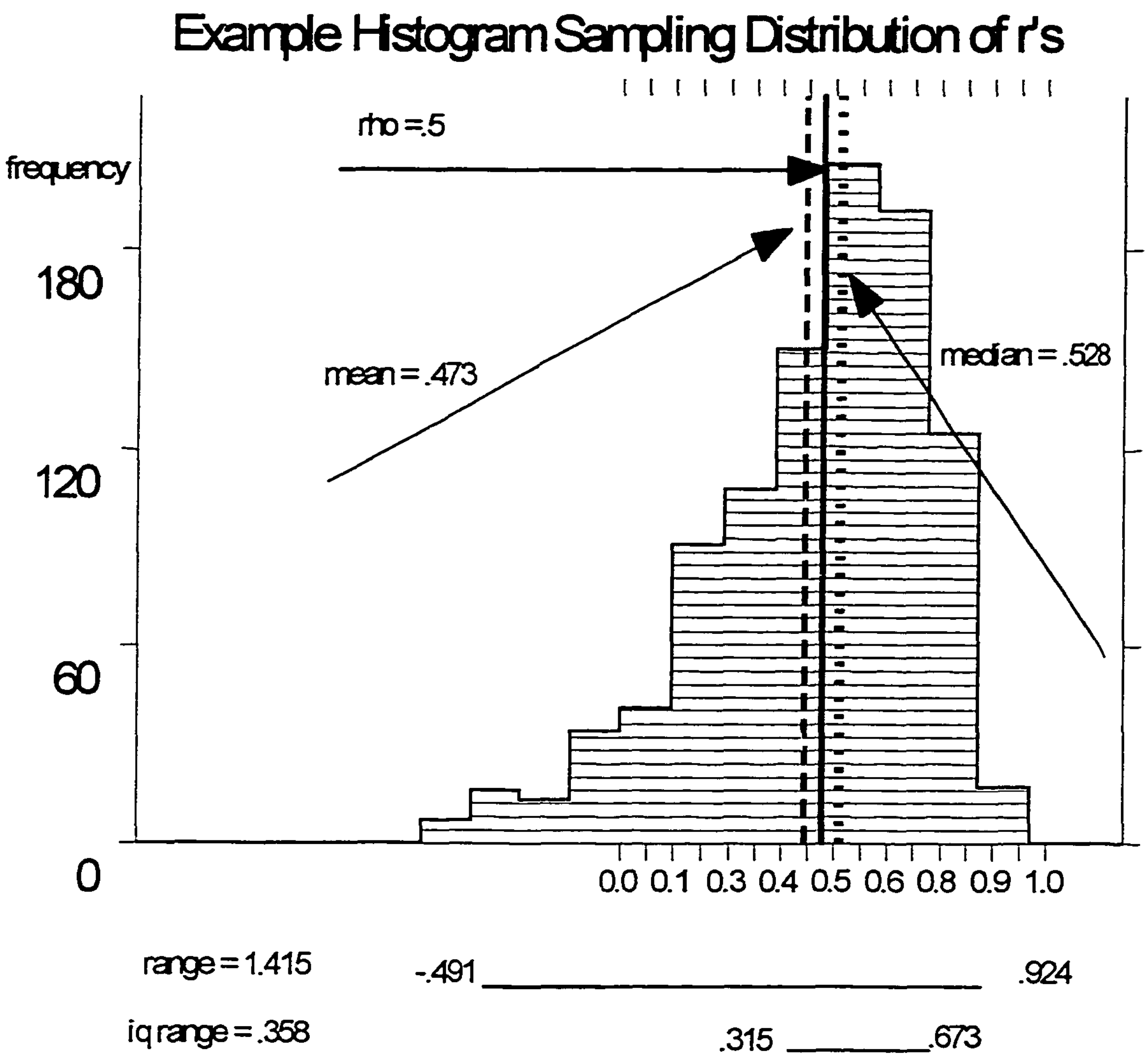


Figure 5. Simulated Sampling Distribution of Pearson r's: Sample Size is 10, rho is .5.

In this illustration the range of values which characterize this distribution of Pearson r's is 1.415. r's fell from a low of -.491 to a high of .924. While no value of Pearson r can be greater than 1, some values were negative. 50% of all r's, the interquartile range, were between .358 and .673, quite closely centered on the rho of .5. The median value was .528. The mean was .473, a value lower than the median, indicating a negative skew, in this example calculated to be -.898.

The following tables display the characterizing values of the distributions of r's across sample size for each rho and outlier condition. First are shown values for distributions without outlier. Then are shown values for distributions with outlier.

Displays of Characterizing Measures:
Central Tendency, Skewness and Dispersion

Measures When rho is .5, Without Outlier

Table 1 displays the mean, median and skewness of Pearson r's across sample size for each of 10, 30, and 50 when rho is .5. It can be seen that the mean and median of Pearson r's closely approximate rho across all sample sizes. As sample size gets bigger, there is an almost trivial convergence on rho. The degree of skewness in

Table 1

Mean, Median and Skewness of Pearson r's Across Sample Size for Each of 10, 30, 50, Without Outlier, When rho is .5

Sample Size	ρ	Mean	Median	Skewness
10	.5	.473	.528	-.898
30	.5	.487	.492	-.428
50	.5	.496	.509	-.408

the distribution of r's decreases dramatically between the sample size of 10 and the sample size of 30, but there is little additional change in skewness at sample size 50.

Table 2 displays the range and interquartile range of Pearson r's across sample size for each of 10, 30, and 50 when rho is .5. As shown, the dispersion of the distribution of Pearson r's dramatically shifts downward as sample size gets bigger. Both range and interquartile range are essentially halved between sample size 10 and sample size 50.

Table 2

Range and Interquartile Range of Pearson r's Across Sample Size for Each of 10, 30, 50, Without Outlier, When rho is .5

Sample Size	ρ	Range	Interquartile Range
10	.5	1.415	.358
30	.5	.909	.200
50	.5	.721	.142

Measures When rho is .8, Without Outlier

Table 3 displays the mean, median and skewness of Pearson r's across sample size for each of 10, 30, and 50 when rho is .8. It can be seen, as with rho of .5, that the mean and median of Pearson r's closely approximate rho across all sample sizes. As sample size gets bigger, there is only slight convergence on rho. The degree of skewness decreases as sample size gets bigger. A dramatic contraction in skewness occurs between the sample size of 10 and the sample size of 30. While lesser, there is still substantial contraction in the skewness of the distribution of r's between sample size 30 and 50.

Table 3

Mean, Median and Skewness of Pearson r's Across Sample Size for Each of 10, 30, 50, Without Outlier, When rho is .8

Sample Size	ρ	Mean	Median	Skewness
10	.8	.780	.821	-1.771
30	.8	.791	.803	- .948
50	.8	.797	.804	- .623

Table 4 displays the range and interquartile range of Pearson r's across sample size for each of 10, 30, and 50 when rho is .8. As shown, the dispersion of the distribution of Pearson r's dramatically shifts downward as sample size gets bigger, with a particularly large jump between sample size 10 and sample size 30.

Table 4

Range and Interquartile Range of Pearson r's Across Sample Size for Each of 10, 30, 50, Without Outlier, When rho is .8

Sample Size	ρ	Range	Interquartile Range
10	.8	1.085	.172
30	.8	.555	.098
50	.8	.358	.072

Measures When rho is .5, With Outlier

Table 5 displays the mean, median and skewness of Pearson r's across sample size for each of 10, 30, and 50 when rho is .5 in the outlier condition. The mean and median of Pearson r's become closer to rho as sample size gets bigger. However, the disparity between rho and the mean or median of Pearson r's is substantial for the sample size of 10 and still not trivial for the sample sizes of 30 and 50. The degree of skewness becomes smaller as sample size moves from 10 to 30, then stabilizes.

Table 5

Mean, Median and Skewness of Pearson r's Across Sample Size for Each of 10, 30, 50, With Outlier, When rho is .5

Sample Size	ρ	Mean	Median	Skewness
10	.5	.784	.800	-.818
30	.5	.638	.641	-.367
50	.5	.595	.602	-.400

Table 6 displays the range and interquartile range of Pearson r's across sample size for each of 10, 30, and 50 when rho is .5 in the outlier condition. As seen, the dispersion of the distribution of Pearson r's is only slightly decreased as sample size gets bigger.

Table 6

Range and Interquartile Range of Pearson r's Across Sample Size for Each of 10, 30, 50, With Outlier, When rho is .5

Sample Size	ρ	Range	Interquartile Range
10	.5	.601	.142
30	.5	.572	.129
50	.5	.556	.110

Measures When rho is .8, With Outlier

Table 7 displays the mean, median and skewness of Pearson r's across sample size for each of 10, 30, and 50 when rho is .8 in the outlier condition. It can be seen that the mean and median of Pearson r's approximate rho across all sample sizes but

Table 7

Mean, Median and Skewness of Pearson r's Across Sample Size for Each of 10, 30, 50, With Outlier, When rho is .8

Sample Size	ρ	Mean	Median	Skewness
10	.8	.905	.913	-1.083
30	.8	.847	.852	- .702
50	.8	.832	.835	- .546

the approximation improves somewhat as sample size gets bigger. The degree of skewness decreases markedly as sample size gets bigger.

Table 8 displays the range and interquartile range of Pearson r's across sample size for each of 10, 30, and 50 when rho is .8 in the outlier condition. As shown, the dispersion of the distribution of Pearson r's is narrow across all sample sizes, with little difference between sample size 10 and 30 and some reduction at size 50.

Table 8

Range and Interquartile Range of Pearson r's Across Sample Size for Each of 10, 30, 50, With Outlier, When rho is .8

Sample Size	ρ	Range	Interquartile Range
10	.8	.322	.064
30	.8	.300	.061
50	.8	.253	.052

Displays Showing Efficacy of Estimate:
Hit Rates Across Sample Sizes

The following set of tables presents the values which represent the effectiveness of the Pearson estimate under conditions of each distribution. These values, the hit rates, are displayed across sample size for each rho and outlier condition. First are shown hit rate values for distributions without outlier. Then are shown hit rate values for distributions with outlier.

Hit Rates When rho is .5 and .8, Without Outlier

As is shown in Table 9, the hit rate of the distribution of r's without outlier closely approximates the expected 95% rate across all sample sizes and across both conditions of rho. This is the case even in the smallest sample size.

Table 9

Hit Rates of Pearson r's Across Sample Sizes When rho is .5 and .8, Without Outlier

Sample Size	Expected %	Hit Rate r's rho = .5	Hit Rate r's rho = .8
10	95.0%	94.1%	94.6%
30	95.0%	94.7%	94.3%
50	95.0%	94.4%	94.6%

Hit Rates When rho is .5 or .8, With Outlier

As is shown in Table 10, the hit rate of the distribution of r's with outlier approximates the expected 95% rate poorly although proximity improves as sample size increases. Approximation is substantially poorer when rho is .5 than when rho is .8. The divergence is particularly noticeable in the sample size of 10 for both conditions of rho, and still marked in the sample size of 30 when rho is .5.

Table 10

Hit Rates of Pearson r's Across Sample Sizes When rho is .5 and .8, With Outlier

Sample Size	Expected %	Hit Rate r's rho = .5	Hit Rate r's rho = .8
10	95.0%	72.9%	83.9%
30	95.0%	82.9%	90.0%
50	95.0%	87.1% .	91.8%

Displays Organized by Outlier Condition

The purpose of the following tables is to provide a ready basis for comparison between outlier conditions across sample size for each rho. Displayed are characteristic values of the distributions both without and with outlier, followed by hit rates of those distributions both without and with outlier.

Measures of Central Tendency, Skewness and Dispersion Under Both Outlier Conditions When rho is .5

Table 11 displays measures of central tendency and skewness across sample sizes when rho is .5 both without and with outlier. Measures of central tendency closely approximate rho across all sample sizes in the without outlier condition, but not in the with outlier condition, particularly when sample size is 10. The degree of skewness drops dramatically when sample size is higher than 10 in both the without outlier condition and in the with outlier condition.

Table 11

Mean, Median and Skewness of Pearson r's Across Sample Size for Each of 10, 30, 50, Without Outlier and With Outlier, When rho is .5

Sample Size	ρ	Mean r's		Median r's		Skewness r's	
		w/o Outlier	-w Outlier	w/o Outlier	-w Outlier	w/o Outlier	-w Outlier
10	.5	.473	.784	.528	.800	-.898	-.818
30	.5	.487	.638	.492	.641	-.428	-.367
50	.5	.496	.595	.509	.602	-.408	-.400

As shown in Table 12, the dispersion of the distribution of r's without outlier decreases substantially with increased sample size. In the distributions with outlier,

Table 12

Range and Interquartile Range of Pearson r's Across Sample Size for Each of 10, 30, 50, Without Outlier and With Outlier, When rho is .5

Sample Size	ρ	Range r's		Interquartile Range r's	
		w/o Outlier	-w Outlier	w/o Outlier	-w Outlier
10	.5	1.415	.601	.358	.142
30	.5	.909	.572	.200	.129
50	.5	.721	.556	.142	.110

dispersion is overall smaller than in the distributions without outlier, but is little altered with increased sample size.

Measures of Central Tendency, Skewness and Dispersion Under Both Outlier Conditions When rho is .8

As can be seen in Table 13, when rho is .8, measures of central tendency closely approximate rho across sample sizes in the without outlier condition, but not closely in the with outlier condition when sample size is 10. In the with outlier condition, both mean and median become closer to rho as sample size increases. Skewness is smaller in the with outlier condition than in the without outlier condition. Skewness decreases as sample size increases in both outlier conditions.

Table 13

Mean, Median and Skewness of Pearson r's Across Sample Size for Each of 10, 30, 50, Without Outlier and With Outlier, When rho is .8

Sample Size	ρ	Mean r's		Median r's		Skewness r's	
		w/o Outlier	w Outlier	w/o Outlier	w Outlier	w/o Outlier	w Outlier
10	.8	.780	.905	.821	.913	-1.771	-1.083
30	.8	.791	.847	.803	.852	- .948	- .702
50	.8	.797	.832	.804	.835	- .623	- .546

Table 14 shows range and interquartile across sample sizes when rho is .8 both without and with outlier. Dispersion becomes substantially smaller with increased sample size in the without outlier condition, but is little affected by increased sample size in the with outlier condition. Range is noticeably smaller in the outlier condition.

Table 14

Range and Interquartile Range of Pearson r's Across Sample Size for Each of 10, 30, 50, Without Outlier and With Outlier, When rho is .8

Sample Size	ρ	Range r's		Interquartile Range r's	
		w/o Outlier-	w Outlier	w/o Outlier-	w Outlier
10	.8	1.085	.322	.172	.064
30	.8	.555	.300	.098	.061
50	.8	.358	.253	.072	.052

Hit Rates Under Both Outlier Conditions When rho is .5 or .8

Table 15 displays the hit rate of r's across study conditions. As can be seen, distributions of r's in the no outlier condition closely approximated rho and effectiveness, as measured by hit rate, paralleled expected values. In the outlier condition, however, the hit rate of r's decreased. The outlier reduced effectiveness of the Pearson r as sample size decreased. The reduction was more substantial when rho was .5 than when rho was .8, tangible when sample size was 10 and even as large as 30.

Table 15

Comparison of the Hit Rates of Pearson r's Across Sample Sizes, Conditions of rho and Outlier Condition

		rho = .5		rho = .8	
Sample Size		w/o Outlier	with Outlier	w/o Outlier	with Outlier
	Expected %	Hit Rate r's	Hit Rate r's	Hit Rate r's	Hit Rate r's
10	95.0%	94.1%	72.9%	94.6%	83.9%
30	95.0%	94.7%	82.9%	94.3%	90.0%
50	95.0%	94.4%	87.1%	94.6%	91.8%

Synthesis of Findings Across Conditions of Study

Without Outlier

In the without outlier condition, there was little risk associated with the Pearson estimator across sample sizes and conditions of rho. The hit rate was consistently high. As sample size increased, mean r's and median r's, which were already very close to rho in the smallest sample size, became even closer to rho and closer to one another. With increase in sample size skewness of the distributions decreased. With increase in sample size, range and interquartile range become much smaller.

With Outlier

In the with outlier condition, risk associated with the Pearson estimator increased considerably as sample size and rho became smaller. Hit rate was lowered. Mean and median r's were not close to rho in the smallest sample size, particularly when rho was .5. With increase in sample size, mean r's and median r's became somewhat closer to rho, but were still not close approximations of rho. With increase in sample size, skewness of the distributions became smaller but range and interquartile range decreased little with increase in sample size.

Contrast

Overall, in the without outlier condition, there was little risk associated with the Pearson estimator across study conditions and there were some benefits from increased sample size, little in terms of central tendency but striking in terms of reduced dispersion. By contrast, with the outlier, the degree of risk associated with the Pearson estimator was substantially increased, particularly for moderate rho. Increased sample size provided some reduction in risk but only moderate benefit in terms of central tendency and very little benefit in terms of dispersion.

CHAPTER V

DISCUSSION

Introduction to Discussion

The purpose of this study was to address the probable effectiveness of the Pearson sample correlation coefficient (r_{xy}) as estimator of moderate or strong population correlation (ρ_{xy}) when sample size was small and when that sample data contained an outlier. This purpose was addressed by observing central tendency, dispersion and skewness of simulated distributions of Pearson r's under moderate (.5) and strong (.8) conditions of rho, in three sample sizes (10, 30, 50), both without and with outlier. In addition, the hit rate, or percentage of times in which the Pearson estimator captured the known rho, was used to measure effectiveness of the Pearson estimator under conditions of that distribution.

Overview of Findings

Without Outlier

In the without outlier condition, sample size had little effect on central tendency. Mean and median of Pearson r's approximated rho across sample size. Hit rate was consistently high. Both range and skewness of the distribution of r's decreased with increasing sample size. All measures indicated that the Pearson r was robust across all sample sizes and conditions of rho.

With Outlier

Addition of the outlier affected values of both central tendency and dispersion, but had little additional effect on skewness of the distributions. With the outlier, mean and median values or r's were shifted farther from rho. Range of the distributions became smaller, irrespective of sample size or rho. Increased sample size had little effect on dispersion of the with outlier distributions.

Of most practical interest, in the outlier condition, the degree of risk associated with the estimator, as measured by the hit rate, was reduced across all sample sizes and conditions of rho. This reduction in hit rate of the Pearson estimator was most pronounced when rho was moderate and sample size was smallest.

When rho was moderate and sample size was 10, the hit rate of the estimator was 73% as contrasted to the 94% achieved in the same sample size without outlier. When sample size was 30 hit rate was increased to 83%, still well below the expected rate of 95% found in the same sample size without outlier. When sample size was 50, hit rate was at 87% as contrasted with the expected 95%.

When rho was strong, degradation in effectiveness of the Pearson estimator was less marked than when rho was moderate, but was still substantial. When rho was strong, the hit rate of the Pearson estimator in the very small sample size was 84% rather than the expected 95%. Increasing sample size to 30 improved the hit rate of the estimator to 90%, but still short of the expected 95%. Little further gain was accomplished by raising sample size to 50.

The primary implication of this finding is that if an outlier occurs in small sample data the risk of making an incorrect decision based on the Pearson sample coefficient is greatly increased. If rho is moderate rather than strong, the degree of risk is further increased. While increasing sample size does tend to reduce the degree of risk somewhat, the remaining degree of risk is still well higher than that which would be expected in the no outlier situation, even when sample size is as large as 50.

Implications for Practice

Without Outlier

The results of this study have particular implications for investigators. First, when there is no outlier, findings indicated that the Pearson r is robust even in the small sample. The indication of this finding, consistent with studies dating back to Pearson (1929) and summarized by Kowalski (1972), is that sample size is of inconsequential practical importance to the estimation problem. The practical implication is that use of the Pearson estimator is essentially secure even in very small samples.

With Outlier

Results indicated that the presence of an outlier tended to increase the value of the Pearson estimate in the small sample and to materially affect uncertainty related to the estimation decision. The implication of these results is twofold. 1) The practical investigator should carefully attend the interpretation of results if an outlier occurs in small sample data. 2) The practical investigator should be aware that since the likelihood of estimation accuracy is diminished, the likelihood of making a correct decision is likewise diminished if an outlier is observed in small sample data. Affected are both data assessment and study design.

Results indicate that at the time of data assessment the likelihood of increased risk of an incorrect decision is a given. If an outlier appears in small study data, the value of the Pearson coefficient will tend to be high. The degree of risk attending the estimation decision will tend to be high. If small study data contains an outlier, and the cause of that outlier cannot be determined, depending on the gravity of the decision involved, the best course of action may be to discard the data and begin again. Study results show that increasing the amount of data collected once the outlier has appeared is of little benefit. Sample size will not totally overcome the outlier problem.

At the time of study design, control of risk may still be possible. In study design, if the investigator is confronted with a new situation where a population is unknown, the sample variance from a pilot study is typically used in determining desired sample size for further assessment. Confidence about the correlation coefficient is based on the sampling distribution of the correlation coefficient with its associated confidence or probability of making an error at a given sample size.

In the presence of an outlier, measures of central tendency are shifted, variance is affected and the sampling distribution is altered. The probability of making a correct decision based on the Pearson r, particularly about a moderate rho, may fall from an expected 95% to as little as 73%. This result will be of particular importance in measurement studies where moderate correlation is typically encountered. Once it is understood that if an outlier should occur in small study data, either sample size needs to be well increased or a greater risk than expected will be assumed, it may still be possible to increase study size which in turn will reduce risk.

When the underlying probability is considered, the appropriate course of action selected will depend on a number of factors since applied investigation tends to be problem specific (Clemen, 1996). The relevant outcomes of conducting the research, the payoff of the research, may need to be balanced against the expense of increased sample size. Expected value of the decision, or expected opportunity loss of the decision may be weighed against the cost of making the decision. Availability of study subjects may be a limiting or uncontrollable constraint. Time may be a factor. Whether or not the assumption of a higher degree of risk about the Pearson r is acceptable in a particular situation remains in the hands of the investigator. Understanding that the probability of making an incorrect decision is increased and that this uncertainty might be controlled, if not totally eliminated, is an advantage in weighing potential courses of action.

Conclusions

This study was conducted in the expectation that it could contribute to 1) the assessment or analysis of research data and to 2) planning of research design. The study has provided a fuller explanation of the effect an outlier in small sample data under conditions which might be encountered in a measurement situation. Secondly, the study has demonstrated that the effect of an outlier is better explained by a combination of sample size and rho than by either characteristic alone. Thirdly, the study has provided a fuller assessment of the degree of risk associated with the presence of an outlier which is extreme relative to the body of sample data but not readily excludable from that data.

Study findings supports several theoretical underpinnings. First, findings support the premise that the Pearson r is robust even in the small sample (Pearson, 1929; Kowalski, 1972). There was no indication that sample size alone diminished the effectiveness of the estimator to any noticeable degree. Second, findings support the premise that the outlier affects the mean and the range of the sample distribution of the estimator (Armstrong & Frome, 1977; Barnett & Lewis, 1995). The presence of the outlier increased the mean of Pearson r's and decreased the range. Third, findings support the premise that the effect is greatest when sample size is small (Srivastava &

Lee, 1985; McCallister, 1991).

Results of this study also have additional implications for the theoretical. Historical approaches to outlier study have been based on the violation of normal assumptions through the use of non-normal and mixture models. The findings of this study support the suggestion of Srivastava & Lee (1985) who submit that the case in which the whole sample comes from a distribution which is not normal is not identical to the case in which an outlier occurs in an observed sample. This study supports the usefulness of studying this separate problem in a different way. Although several mixture studies find that the correlation coefficient to be depressed in the presence of an outlier (Devlin, Gnanadesikan, & Kettenring, 1975; Wainer & Thissen, 1979), the results of this study indicated that the estimated correlation coefficient was greater than that expected in the presence of an outlier. The difference in conception and definition of the outlier may be a substantive one.

A second related implication is that the consideration of rho, in addition to the consideration of sample size, is useful to the study of the case in which an outlier occurs in an observed sample. Findings indicated that rho made a material difference in outcome. The outlier was found to have the greatest overall effect on the Pearson r under the combination of moderate rho and small sample size.

In addition, an implication for another stream of study exists. Historically, study and discussion of outlier treatment, including removal or abridgement of the outlier, has been considerable (Kohnert, 1995; Barnett & Lewis, 1995). By contrast, work on risk assessment associated with estimation as it affects decision making has been less substantial. The results of this research indicate that in addition to developing a clearer understanding of the shapes of underlying distributions under outlier conditions and an understanding of the possible treatment of an outlier, it is also useful expand understanding of the ways in which changes in distributions might affect investigative practice such as the amount of risk investigators are likely to assume under the prospect of an outlier.

Implications for Further Research

As an addendum to the preceding discussion several observations about the limitations to the generalizability of these findings should be noted. The simulation used in this study was limited to consideration of specific sample size, rho and outlier conditions. Although these conditions were chosen to limit the scope of the study in accordance with particular measurement conditions, this restriction necessarily limits generalizability of findings to conditions which were not considered. While selected restrictions limit generalizability of findings, they at once suggest several directions for future research. Because the Pearson coefficient was found to be most sensitive to the presence of an outlier when sample size was smallest (10) and rho lowest (.5 rather than .8), the opportunity exists to consider the effect of other sample size and rho combinations on the effectiveness of the Pearson r. Similarly, the opportunity exists to consider the effect of an outliers arising from different locations, the affect of larger outliers, or combinations of outliers on the Pearson r.

In addition, because this study found a loss of effectiveness in the Pearson r under specified outlier conditions, further study might be extended to a comparison of the efficiency of the Pearson r to the efficiency of non-parametric estimators under the conditions of this study. It is possible that a non-parametric estimator such as the Spearman, which is based on ranks, might provide a better estimate since the outlier will become an extreme rank rather than extreme value.

In closing, it is recommended that practical investigation be concerned with the possible risk attending the occurrence of an outlier in small study data. It is also recommended that further study continues to address the quantification of decision risk associated with estimation under practical investigative conditions.

Appendix A

Computer Specific Procedure

First, for each of six conditions (3 sizes X 2 ρ), a script was written as a syntax file in SPSS 8.0 to generate base data. Each segment of script specified a number of cases (10, 30 or 50). In each case, two variables, x' and y', were created as random numbers from a normal distribution (standard normal: x, N=0,1; y, N=0,1). The data list of independent variables x' and y' was saved in a data matrix, X. To create the correlated data list, new correlated variables were created from the independent variables. The data matrix X was post multiplied by the desired correlation matrix, the Cholesky decomposition of R, where R was computed as [1, .5; .5, 1} or [1, .8; .8, 1]. The matrix conversion generated paired variables created as if they were sampled from a distribution with the specified correlation of .5 or .8. The new correlated variables were saved as x and y. This process was replicated 1,000 times, for each of six scripts, one for each set of six conditions yielding 6,000 correlated samples, 1,000 samples for each of 6 size and rho combinations.(Please see Appendix B for a detailed description of script). Following the generation of the six sets of samples, Pearson's r was calculated on each sample. 6 X 1,000 values of Pearson's r were saved.

In a second program, for each of two conditions (2 ρs), a script was written in SPSS to generate outliers. As in the first set of scripts, two variables, x' and y', were created as random numbers from a normal distribution but with a mean of -3 (x, N=-3, .1 y, N=3,.1). The random numbers were then post multiplied by the correlation matrix used in the original program, either [1, .5; .5, 1} or [1, .8; .8, 1]. The resulting variables x and y were saved. The matrix conversion generated correlated numbers as if they were drawn from a distribution with the same correlation (.5 or .8) as the original sample sets, but with a mean of –3 rather than a mean of 0. In this manner, correlated variables were shifted to occur in a circumscribed area pivoting on a value of –3 standard deviations away from the mean of the base data sample sets. These pairs of variables represent the outliers. Two scripts were run yielding two groups of outliers. (Please see Appendix C for a detailed description of script).

As a sub-procedure, one of the data pairs in each of the six sample sets was replaced at random by one of the outliers from the corresponding correlation. The process was replicated 6,000 times yielding 6,000 correlated samples, 1,000 samples for each of 6 size and rho combinations, in which each sample contained an outlier. Following the generation of the six sets of outlier samples, Pearson's r was calculated on each sample. 6 X 1,000 values of Pearson's r were saved.

A third process calculated the upper and lower limits of the confidence intervals for each sample statistic using the z transformation. For each condition, the proportion of times that the population correlation (.5 or .8) was captured in the interval provided the hit rate. These proportions were saved. Please see Appendix D for description of the z transformation.)

Appendix B

Generating Data as if Drawn From a Population With a Given Correlation, Base Sets

*In SPSS, two variables are created as (pseudo) random numbers. Using these (pseudo) random numbers, correlated numbers are created using the property that for a set of random variables, a given correlation matrix can be imposed by post-multiplying the data matrix by the upper triangular Cholesky decomposition of the correlation matrix. The matrix conversion generates variables created as if they were sampled from a population with a given correlation. *

*Script generates one bivariate data set of desired sample size, desired population distribution (here standard normal, mean 0, SD 1), and desired population correlation.
Example for: Sample Size 30, Population Correlation .5

```
new file.
input program.
* Draw 30 cases (modify for number of cases desired)
SET SEED 2469768.
loop #i = 1to 30.
* Draw data for two variables
do repeat response=r1 to r2.
* Specify desired distribution characteristics
COMPUTE response = rv.normal(0,1).
end repeat.
end case.
end loop.
end file.
end input program.
* Save variable list as file
Save outfile = "DataOut.sav"

Matrix.
Get X/var = r1 to r2.
*Replace the variable list r1 to r2 with nr1 to nr2
* to produce correlation matrix.
*Define desired pattern of correlations.
Compute R = {1, .5;
.5, 1}.
Compute NewX = X*chol(R).
* Save NewX as the working file.
Save NewX /outfile = */variables = nr1 to nr2.
End matrix.
*Save correlated variable list as file
```

```
Save outfile = "data1.sav"

*Script generates one bivariate data set as desired, and adds correlated
variable list to previous file
new file.
input program.
* Draw 30 cases (modify for number of cases desired)
SET SEED 2469769.
loop #i = 1to 30.
* Draw data for two variables
do repeat response=r3 to r4.
* Specify desired distribution characteristics
COMPUTE response = rv.normal(0,1).
end repeat.
end case.
end loop.
end file.
end input program.
* Save variable list as file
Save outfile = "DataOut.sav"

Matrix.
Get X/var = r3 to r4.
*Replace the variable list r1 to r2 with nr1 to nr2
* to produce correlation matrix.
*Define desired pattern of correlations.
Compute R = {1, .5;
.5, 1}.
Compute NewX = X*chol(R).
* Save NewX as the working file.
Save NewX /outfile = */variables = nr3 to nr4.
End matrix.

* Add correlated variable list to file
Match files file=data1.sav /file=*.
Save outfile = "data1.sav".

* Repeat for as many variable sets (trials) as desired.
*1,000 trials (sets) were independently generated for each combination of
sample size and population correlation.*
```

(modified from Howell, 1998; Nichols, 1996; SPSS, 1997 and 1998)

Appendix C

Generating Data as if Drawn From a Population With a Given Correlation, Outliers

**Script generates one set of 1,000 bivariate outliers, of desired population distribution (here standard normal, mean -3, SD .1), and desired population correlation.
Example for: Population Correlation .5

```
new file.
input program.
* Draw 1,000 cases (modify for number of trials desired)
SET SEED 2469768.
loop #i = 1to 1000.
* Draw data for two variables
do repeat response=r1 to r2.
* Specify desired distribution characteristics
COMPUTE response = rv.normal(-3,.1).
end repeat.
end case.
end loop.
end file.
end input program.
* Save variable list as file
Save outfile = "DataOut.sav"

Matrix.
Get X/var = r1 to r2.
*Replace the variable list r1 to r2 with nr1 to nr2
* to produce correlation matrix.
*Define desired pattern of correlations.
Compute R = {1, .5;
.5, 1}.
Compute NewX = X*chol(R).
* Save NewX as the working file.
Save NewX /outfile = */variables = nr1 to nr2.
End matrix.
```

*Using this script, 1,000 bivariate outliers were generated for the population correlation = .5. Using a second script of the same form, 1,000 bivariate outliers were generated for the population correlation = .8.

(modified from Howell, 1998; Nichols, 1996; SPSS, 1997 and 1998)

Appendix D

Calculating Confidence Interval About Derived Correlation Coefficients

CONFIDENCE INTERVAL ABOUT SAMPLE CORRELATION COEFFICIENT

.

A 95% confidence interval about each sample correlation coefficient (r) is developed using the Fisher's Z which transforms the asymmetric distribution of rs into an approximately normal distribution as follows.

a. Each r (and r_o) is converted to a Fisher's Z using the formula:

$$Zr = 1/2 \ln \frac{1+r}{1-r}$$

Zr is approximately normally distributed with an estimated standard error of:

$$SE = \frac{1}{\sqrt{n-3}}$$

where n= sample size

b. Upper and lower bounds in the *Zr* scale are calculated using:

Upper bound = *Zr* + *1.96(SE)* and Lower bound = *Zr* - *1.96(SE)*

c. *Zr* is returned to the original scale using:

$$r = \frac{e^{2Z} - 1}{e^{2Z} + 1}$$

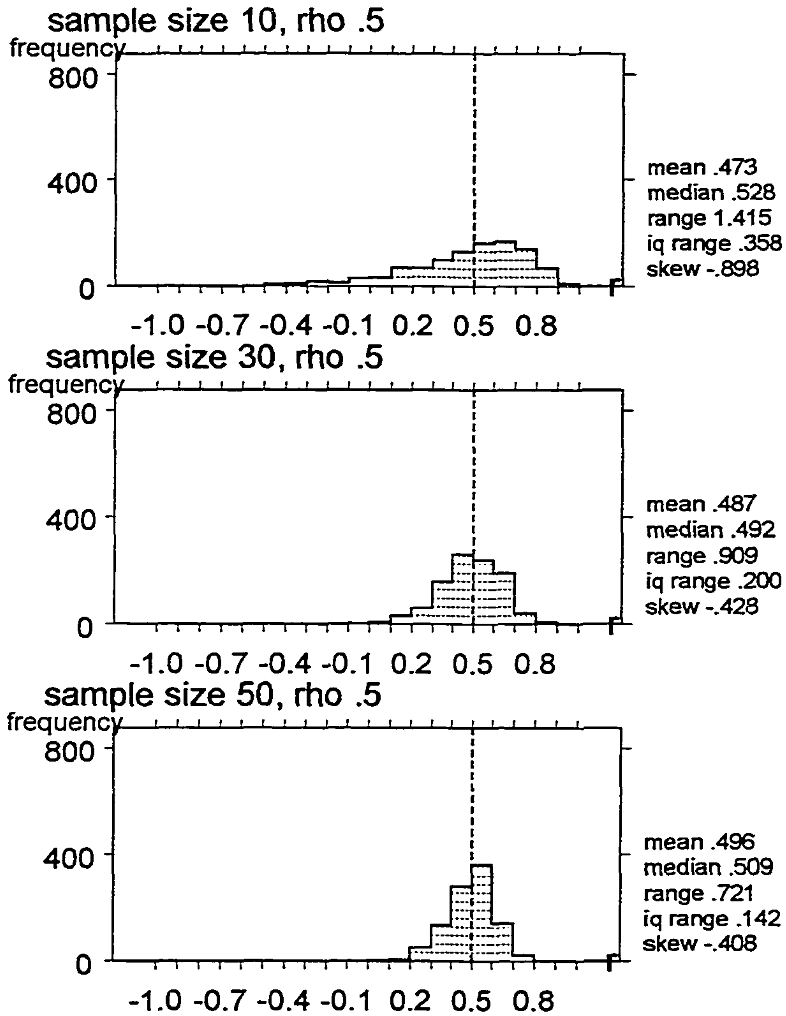
where *e* is the base of the natural log.

(see Fisher, 1915; Chou, 1969; Mendenhall, Wackerly & Scheaffer, 1990)

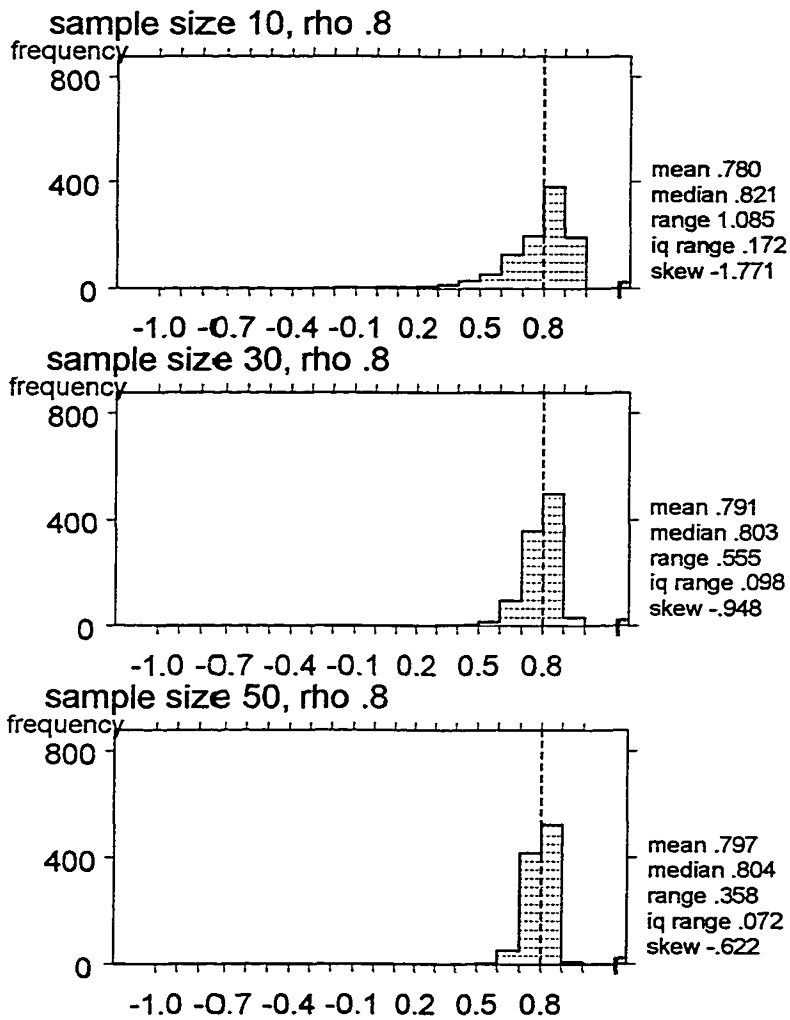
Appendix E

Sampling Distributions of r's Organized by Sample Size

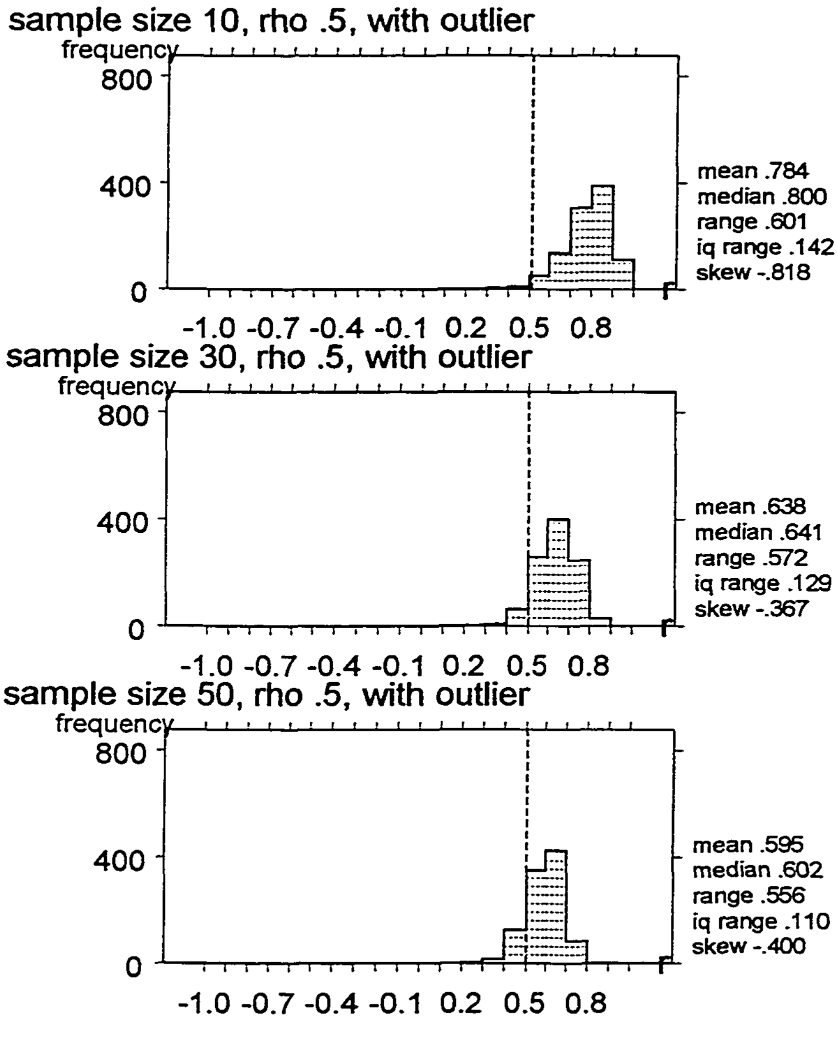
E1: Distributions of r's across 3 samples sizes when rho is .5



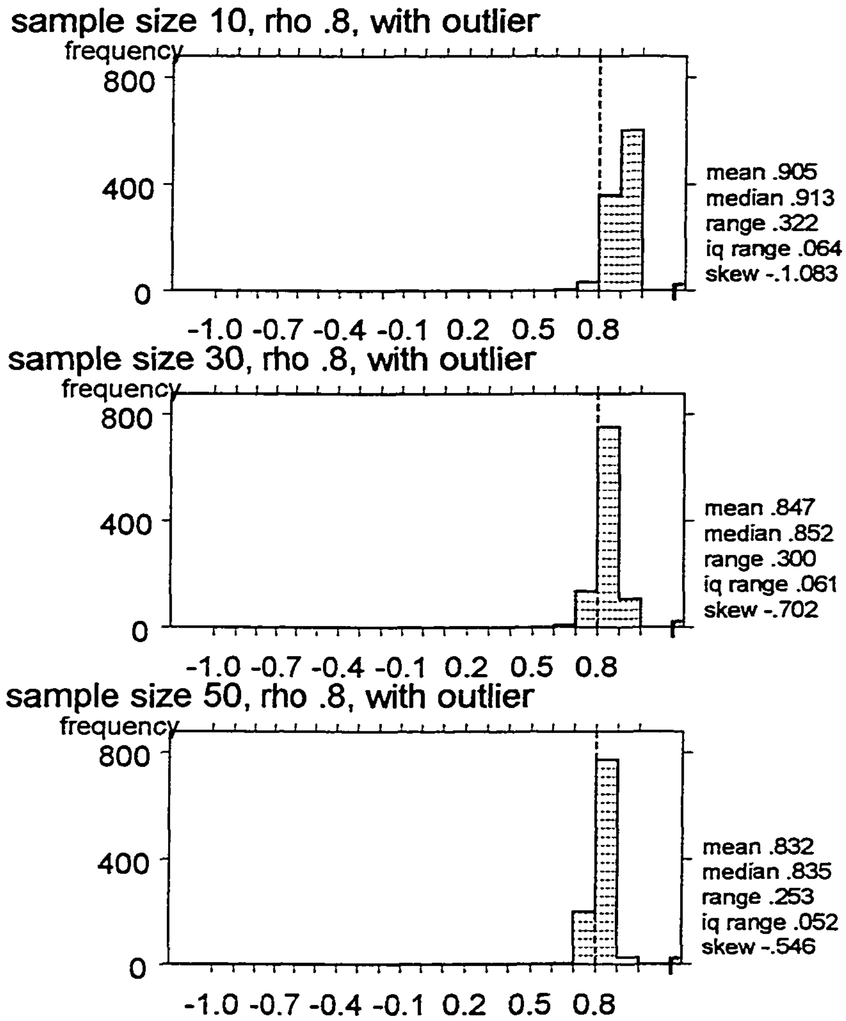
E2: Distributions of r's across 3 samples sizes when rho is .8



E3: Distributions of r's across 3 samples sizes when rho is .5, with outlier



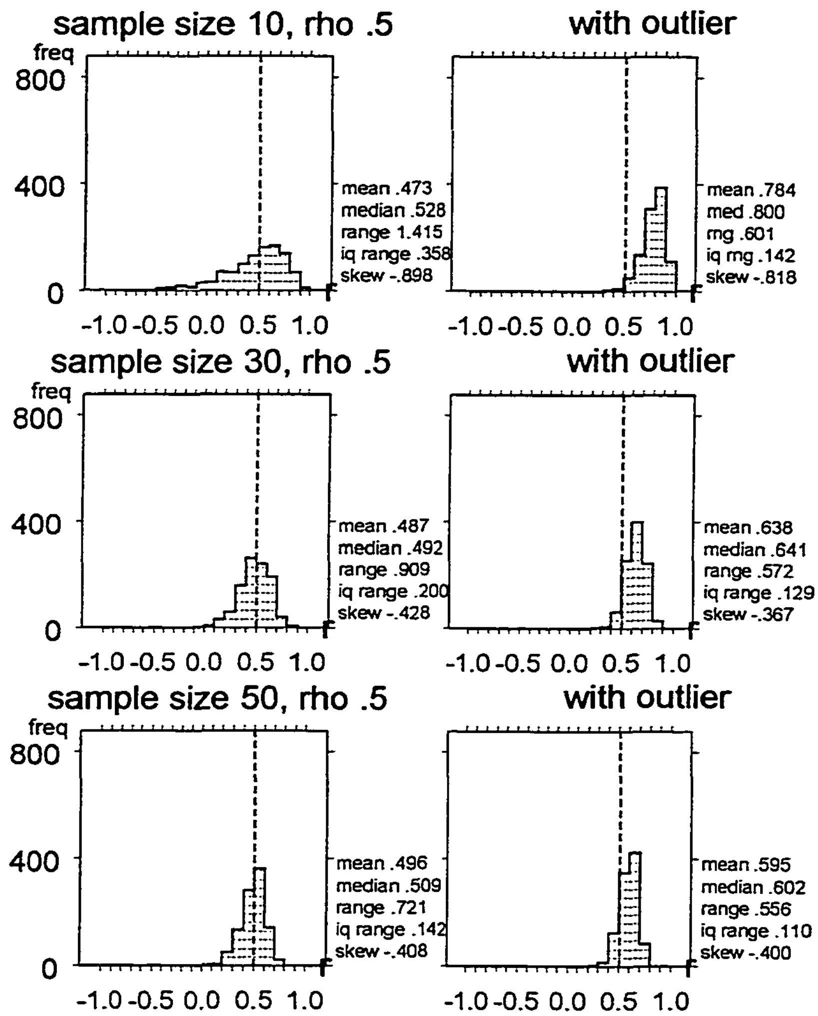
E4: Distributions of r's across 3 samples sizes when rho is .8, with outlier



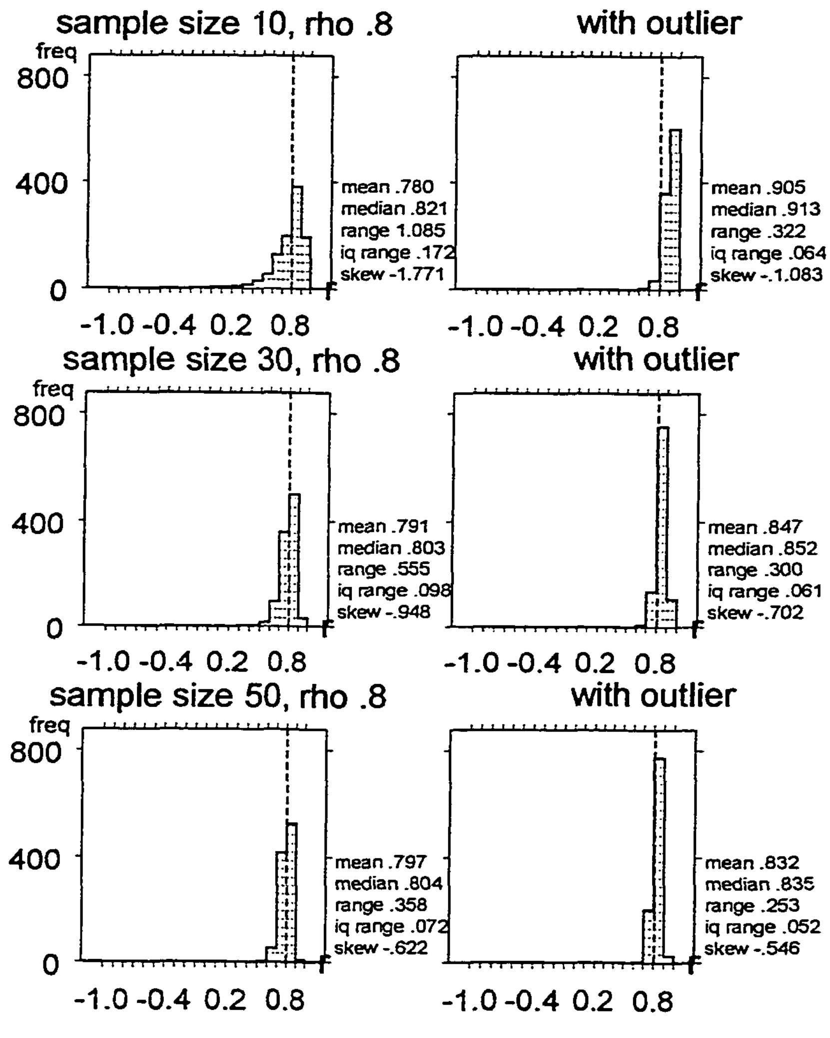
Appendix F

Sampling Distributions of r's by Outlier Condition

F1. Distributions of r's without and with outlier when rho is .5



F2. Distributions of r's without and with outlier when rho is .8



BIBLIOGRAPHY

Alexander, R. L., Carson, K. P., Alliger, G. M., & Carr, L. (1987). Correlating doubly truncated correlations: An improved approximation for correcting the bivariate normal correlation when truncation has occurred on both variables. Educational and Psychological Measurement, 47(2), 309-315.

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., & Tukey, J. W. (1972). Robust estimation of location. Princeton: Princeton University Press.

Andrews, D. F. & Pregibon, D. (1978). Finding the outliers that matter. Journal of the Royal Statistical Society B, 1, 85-93.

Anscombe, F. J. (1960). Rejection of outliers. Technometrics, 2, 123-147.

Armstrong, R. D., & Frome, E. L. (1977). A special purpose linear programming algorithm for obtaining least absolute value estimators in a linear model with dummy variables. Communications in Statistics: Simulation and Computation B, 1, 383-398.

Banerjee, K. K., Bishayee, A. & Marimutha, P. (1997). Evaluation of cyanide exposure and its effects on thyroid function of workers in a cable industry. Journal of Occupational and Environmental Medicine, 39 (3), 258-260.

Barnett, V. (1978). The study of outliers: Purpose and model. Applied Statistics, 27, 242-250.

Barnett, V. (1992). Unusual outliers. In Schach, S. & Trenkler, G. (Eds.), Data analysis and statistical inference. Koln: Verlag.

Barnett, V. (1993). Outliers in sample surveys. Journal of applied statistics.

Barnett, V., & Lewis, T. (1995). Outliers in statistical data (3rd ed.). New York: John Wiley & Sons.

Beckman, R. J., & Cook, R. D. (1983). Outliers. Technometrics, 25, 119-163.

Becketti, S., & Gould, W. (1987). Rangefinder box plots. The American Statistician, 41, 149.

Belsley, D. A. (1996). A small-sample correction for testing for gth-order serial correlation with artificial regressions (Boston College Working Papers in Economics, No. 331). Boston, MA: Boston College Department of Economics.

Bickel, P. J., & Lehman, E. L. (1975-6). Descriptive statistics for nonparametric models. I, II, and III Annals of Statistics, 3, 1038-1069; 4, 1139-1158.

Borjas, G. J., & Sueyoshi, G. T. (1994). A two-state estimator for probit models with structural group effects. Journal of Econometrics, 64(1), 165-182.

Breen, H. J., Rogers, P. A., Lawless, H. C., Austin, J. S., & Johnson, N. W. (1997). Important differences in clinical data from third, second, and first generation periodontal probes. Periodontal Journal, 68, 335-345.

Bridgeman, B., & Harvey, A. (1998, April). Validity of the English language proficiency test. Paper presented at the annual meeting of the National Council on Measurement in Education. San Diego, CA.

Brown, F. G. (1983). Principles of educational and psychological testing (3rd ed.) Orlando, FL: Harcourt, Brace, Jovanovich.

Busby, D., & Thompson, B. (1990, January). Factors attenuating Pearson's r: A review of the basics and some corrections. Paper presented at the annual meeting of the Southwestern Educational Research Association, Austin, TX.

Bush, L.K., Hess, U., & Wolford, G. (1993). Transformations for within subject designs: A monte-carlo investigation. Psychological Review, 113, 566-579.

Campbell, N. A. (1980). Robust procedures in multivariate analysis I: Robust covariance estimation. Applied Statistics, 29, 231-237.

Chaloner, K., & Brant, R. (1988). A Bayesian approach to outlier detection and residual analysis. Biometrika, 75(4), 651-659.

Cheriyan, K. C. (1945). Distributions of certain frequency constants in samples from non-normal populations. Sankhya, 7, 159-166.

Chisman, J. (1992). Introduction to simulation modeling using gpss/pc. Englewood Cliffs, NJ: Prentice Hall.

Chou, Y. (1969). Statistical analysis with business and economic applications. New York: Holt, Rinehart and Winston.

Chou, Y. (1989). Statistical analysis for business and economics. NY: Elsevier.

Churchill, G. A., Jr. (1991). Marketing research: Methodological foundations (5th ed.). Orlando, FL: Dryden Press.

Clemen, R. T. (1996). Making hard decisions (2nd.ed.). Pacific Grove, CA: Duxbury Press.

Cohen, J., & Cohen, P. (1983). Applied multiple regression and correlation analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Conover, W. J. (1980). Practical nonparametric statistics (2nd ed.). NY: John Wiley and Sons.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. Orlando, FL: Harcourt, Brace, Jovanovich.

Daniel, W. W. (1990). Applied nonparametric statistics (2nd ed.). Boston, MA: PWS – Kent.

David, F. N. (1938). Tables of the distribution of the correlation coefficient. Biometrika Office.

Davidson, R., & MacKinnon, J. G. (1993). Estimation and inference in econometrics. New York: Oxford University Press.

Devlin, S. J., Gnanadesikan, R., & Kettenring, J. R. (1975). Robust estimation and outlier detection with correlation coefficients. Biometrika, 62 (3), 531-545.

Dowd, S. B. (1991). The relationship between method of clinical instruction in radiography and scores on the American registry of radiologic technologists certification examination. (U.S. Illinois. Issue Number: RIEJAN92). Eric Document Reproduction Service No. ED 335503.

Draper, N., & Smith, H. (1981). Applied regression analysis (2nd ed.). New York: Wiley.

Drudge, O. W. (1981). A comparison of the WISC-R, McCarthy Scales, Woodcock-Johnson, and academic achievement: Concurrent and predictive validity. (ERIC Document Reproduction Service No. ED 244 996).

Duncan, G. T., & Layard, M. W. J. (1973). A Monte-Carlo study of asymptotically robust tests for correlation coefficients. Biometrika, 60, 551-558.

Edgell, S. E., & Noon, S. M. (1984). Effect of violation of normality on the t test of the correlation coefficient. Psychological Bulletin, 95, 576-583.

Edwards, A. L. (1976). An introduction to linear regression and correlation. San Francisco, CA: W.H. Freeman and Company.

Fallon, A. & Spada, C. (1996). Detection and accommodation of outliers in normally distributed data sets. In D. Gallagher Environmental sampling & monitoring primer. Blacksburg, VA: Virginia Tech: Department of Civil Engineering.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an infinitely large population. Biometrika, 10, 507-521.

Fishman, G. S. (1996). Monte Carlo concepts, algorithms, and applications: Springer series in operations research. New York: Springer-Verlag.

Fowler, R. L. (1987). Power and robustness in product-moment correlation. Applied Psychological Measurement, 11(4), 419-428.

Freedman, D., Pisani, R., Purves, R., & Adhikari, A. (1991). Statistics (2nd ed.). New York: W. W. Norton.

Gayen, A. K. (1951). The frequency distribution of the product-moment correlation coefficient in random samples of any size drawn from non-normal universes. Biometrika, 38, 219-247.

Gibbons, J. D. (1993). Nonparametric measures of association (Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-091). Newbury Park, CA: Sage.

Gibbons, R. D. (1994). Statistical methods for groundwater monitoring. New York: John Wiley & Sons.

Gideon, R. A., & Hollister, R. A. (1987). A rank correlation coefficient resistant to outliers. Journal of the American Statistical Association, 82(398), 656-666.

Goldberg, K. M. & Iglewicz, B. (1992). Bivariate extensions of the boxplot. Technometrics, 34, 307-320.

Green, R. F. (1976). Outlier prone and outlier-resistant distributions. Journal of the American Statistical Association, 71, 502-505.

Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. Technometrics, 11, 1-21.

Haldane, J. B. S. (1949). A note on non-normal correlation. Biometrika, 36, 467-468.

Halperin, S. (1986). Spurious correlations: Causes and cures. Psychoneuroendocrinology, 11(1), 3-13.

Hamilton, L. C. (1992). Regression with graphics. A second course in applied statistics. Pacific Grove, CA: Brooks/Cole.

Hampel, F. (1971). A general qualitative definition of robustness. Annals of Mathematical Statistics, 42, 1887-1896.

Hanneman, R. (1998). Basic statistics for social scientists. [Slide show]. Riverside: University of California, Department of Sociology.

Havlicek, L. L., & Peterson, N. L. (1977). Effect of the violation of assumption upon significance levels of the Pearson r. Psychological Bulletin, 84, 373-377.

Hawkins, D. M. (1980). Identification of outliers. New York: Chapman and Hall.

Hays, W. L. (1981). Statistics for psychologists. New York: Holt, Rinehart and Winston.

Hey, G. B. (1928). A new method of experimental sampling illustrated on certain non-normal populations. Biometrika, 30, 68-80.

Hill, M., & Dixon, W. J. (1982). Robustness in real life: A study of clinical laboratory data. Biometrics, 38, 377-396.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1988). Applied statistics for the behavioral sciences (2nd ed.). Boston: Houghton, Mifflin.

Hope, A. C. A., (1968). A simplified Monte Carlo significance test procedure. Journal of the Royal Statistical Society B. 30, 582-598.

Hopkins, K. D. (1998). Educational and psychological measurement and evaluation (8th ed.). Needham Heights, MA: Allyn & Bacon.

Howell, D. (1998, March). Generating correlated data. Available http://www.uvm.edu/~dhowell/statpages.

Huber, P. J. (1964). Robust estimation of a location parameter. Annals of Mathematical Statistics, 35, 73-101.

Hubert, M., & Rousseeuw, P. J. (1996). Robust regression with a categorical covariable. In H. Rieder (Ed.), Robust statistics, data analysis, and computer intensive methods (pp. 215-224). New York: Springer-Verlag.

Hyndman, R. J. (1996). Computing and graphing highest density regions. The American Statistician, 50, 120-126.

Iglewicz, B., & Hoaglin, D. C. (1993). How to detect and handle outliers. Milwaukee, WI: American Society for Quality Control.

Jarrell, M. G. (1991, November). Multivariate outliers: Review of the literature. Paper presented at the annual meeting of the Mid-South Educational Research Association. Lexington, KY.

Jiang, Y. H. (1997, March). Error sources influencing performance assessment reliability or generalizability: A meta analysis. Paper presented at the annual meeting of the American Educational Research Association. Chicago, IL.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1997). Discrete multivariate distributions. New York: John Wiley & Sons.

Kendall, M. G., & Gibbons, J. D. (1990). Rank correlation methods (5th ed.). NY: Oxford University Press.

Kleijnen, J. P. (1992). Regression metamodels for simulation with common random numbers: Comparison of validation tests and confidence intervals. Management Science, 38, 1164-1185.

Kocherlakota, S., & Kocherlakota, K. (1997). Bivariate discrete distributions. In Encyclopedia of Statistical Sciences – Update Volume 1 (Eds. S. Kotz, C. B. Read, and D. L. Banks). New York: John Wiley & Sons.

Kohnert, A. (1995). RTSIM: A program to simulate the effect of outlier procedures on reaction time data. Marburg, Germany: University of Marburg: Unpublished paper.

Kowalski, C. J. (1968). Coordinate Transformations to Normality in the Theory of Correlation. Doctoral Dissertation, University of Michigan, Ann Arbor.

Kowalski, C. J. (1972). On the effects on non-normality on the distribution of the sample product-moment correlation coefficient. Applied Statistics, 21(1), 1-12.

Kraemer, H. C. (1973). Improved approximation to the non-null distribution of the correlation coefficient. Journal of the American Statistical Association, 68(344), 1004-1008.

Kraemer, H. C. (1980). Robustness of the distribution theory of the product-moment correlation coefficient. Journal of Educational Statistics, 5, 115-128.

Krzanowski, W. J. and Marriott, F. H. C. (1994). Multivariate analysis, Part I: Distributions, ordination and inference. Kendall's Library of Statistics I. New York: Halstead Press.

Law, A. M., & Kelton, W. D. (1991). Simulation modeling & analysis (2nd ed.). New York: McGraw Hill.

Leknes, K. N., Lie, T, Boe, O., & Selvig, K. (1997). A correlation study of inflammatory cell mobilization in response to subgingival microbial colonization. Periodontal Journal, 68, 67-72.

Liebetrau, A. M. (1983). Measures of association (Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-032). Newbury Park, CA: Sage.

Lenth, R. (1988). Comment on Becketti and Gould (with reply by Becketti). The American Statistician, 42, 87-88.

Maracuilo, L. A. & McSweeney, M. (1977). Nonparametric and distribution-free methods for the social sciences. Monterey, CA: Brooks/Cole Publishing Company.

May, K. O. & Hittner, J. B. (1997). Tests for comparing dependent correlations revisited: A Monte Carlo study. The Journal of Experimental Education, 65, 257-69.

McCallister, C. (1991 January). Phi, Rho, P.M., Biserial and Point-Biserial "r": A review of linkages. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, TX.

McNemar, Q. (1962). Psychological statistics. New York: John Wiley.

Mendenhall, W., Wackerly, R. L., & Scheaffer, R. L. (1990). Mathematical statistics with applications (4th ed.). Boston: PWS-KENT Publishing.

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

Miller, J. (1988). A warning about median reaction time. Journal of Experimental Psychology: Human Perception and Performance, 14, 539-543.

Miller, J. (1991). Reaction time analysis with outlier exclusion: Bias varies with sample size. Quarterly Journal of Experimental Psychology, 43A, 907-912.

Mohr, L. B. (1990). Understanding Significance Testing. Newbury Park, CA: Sage.

Mooney, C. Z. (1997). Monte Carlo Simulation. Thousand Oaks, CA: Sage Publications, Inc.

Motulsky, H. (1997, Spring). Detecting outliers. GraphPad Insight, 12.

Muirhead, R. J. (1980). The effects of elliptical distributions on some standard procedures involving correlation coefficients: A review. In R. P. Gupta (Ed.) Multivariate statistical analysis (pp. 143-159). Place: Publisher.

Muller, C. H. (1997). Robust planning and analysis of experiments: Lecture notes in statistics. New York: Springer-Verlag.

Nevitt, J., & Tam, H. P. (1997). A comparison of robust and nonparametric estimators under the simple linear regression model. Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).

Nichols, D. (1996, September). Generating correlated numbers. Available E-mail: nichols@spss.com.

Nunnally, J. C. (1967). Psychometric Theory. New York: McGraw Hill.

O'Hagan, A. (1994). Kendall's Advanced Theory of Statistics – Bayesian Inference. London, England: Edward Arnold.

Pearson, E. S. (1929). Some notes on sampling tests with two variables. Biometrika, 21(1/4), 337-360.

Powers, S. (1986). Validity of the Standard Progressive Matrices as a predictor of achievement of sixth and seventh grade students. Educational and Psychological Measurement, 46(3), 719-22.

Ratcliff, R. (1993). Methods for dealing with outliers. Psychological Review, 114, 510-532.

Rosenberger, J. L. & Gasco, M. (1983). Comparing location estimators: Trimmed means, medians and trimean. In D.C. Hogan, F. Mostellar, & J.W. Tukey, Understanding robust and exploratory data analysis (pp. 297-338). New York: John Wiley & Sons.

Rousseeuw, P. J. & Leroy, A. M. (1987). Robust regression and outlier detection, New York: John Wiley & Sons.

Rousseeuw, P. J. & Hubert, M. (1996). Regression-free and robust estimation of scale for bivariate data. Computational Statistics and Data Analysis, 21, 67-85.

Rousseeuw, P. J. & Ruts, I. (1997). The bagplot: A bivariate box-and-whiskers plot. Technical Report. University of Antwerp.

Rousseeuw, P. J., Ruts, I., & Tukey, J. W. (1999). The bagplot: A bivariate boxplot. Technical Report. University of Antwerp.

Rubien, B. (1996). A test-retest study of the mental scale of the Bayley scales of infant development – second edition in the assessment of school-age children with significant developmental disabilities. Philadelphia, PA: Temple University.

Samo, D. G., Chen, S. C., Crampton, A. R., Chen, E. H., Conrad, K. M., Egan, L. & Mitton, J. (1997). Validity of three lumbar sagittal motion measurement methods: Surface inclinometers compared with radiographs. Journal of Occupational and Environmental Medicine, 39(3), 209-216.

SAS Institute (1998). Experiment infected by creatures from the sixth dimension. Cary, NC: Author.

Sheskin, D. J. (1997). Handbook of parametric and non parametric statistical procedures. Boca Raton, FL: CRC Press.

Siegel, S. (1956). Nonparametric statistics. New York: McGraw Hill.

Siegel, S. & Castellan, N. J. Jr. (1988). Nonparametric statistics for the behavioral sciences (2nd ed.). NY: McGraw Hill.

Simner, M. L. (1992, June). Predictive validity of the caregiver's school readiness inventory. Paper presented at the annual meeting of the Canadian Psychological Association. Quebec City, Quebec, Canada.

Smith, D. K., & Knudtson, L. S. (1990 August). K-ABC and S-B:FE relationships in an at-risk preschool sample. Paper presented at the annual meeting of the American Psychological Association, Boston, MA.

SPSS. (1997). SPSS Base 7.5 for Windows "users" guide. Chicago, IL: Author.

SPSS. (1998). SPSS Base 8.0 applications guide. Chicago, IL: Author.

Srivastava, M. S., & Awan, H. M. (1984). On the robustness of the correlation coefficient in sampling from a mixture of two bivariate normals. Communications in statistics, Theoretical methods.

Srivastava, M. S., & Lee, G. C. (1984). On the distribution of the correlation coefficient when sampling from a mixture of two bivariate normal densities: robustness and the effect of outliers. The Canadian Journal of Statistics, 12 (2), 119-133.

Srivastava, M. S., & Lee, G. C. (1985). On the robustness of tests of correlation coefficient in the presence of an outlier. Communications in Statistics: Theory and Methods, 14 (1), 25-40.

Statsoft. (1998). Basic statistics. Tulsa, OK: Author.

Staudte, R. G. Jr. (1980). Robust estimation. Queen's Papers in Pure and Applied Mathematics – no. 53. Kingston, Ontario, Canada: Queens University.

Stuart, A. and Ord, J. K. (1991) Kendall's Advanced Theory of Statistics, Vol. II, (5th ed.). London, England: Edward Arnold.

Tabachnik, B. G. & Fidell, L. S. (1989). Using multivariate statistics. NY: Harper Collins.

Taylor, J. K. (1987). Quality assurance of chemical measurements. Cheslea, MI: Lewis Publishers.

Tiku, M. L. (1987). A robust procedure for testing an assumed value of the population correlation coefficient. Communications in Statistics: Simualtion and Computation, 16(4), 907-924.

Tiku, M. L., & Balakrishnan, N. (1986). A robust test for testing the correlation coefficient. Communications in Statistics: Simulation and Computation, 15(4), 945-971.

Trochim, W. M. (1997). Research methods knowledge base. Ithaca, NY: Cornell Custom Publishing.

Tukey, J. W. (1975). Mathematics and the picturing of data. Proceedings of the International Congress of Mathematicians, 2, 523-531. Vancouver.

Tukey, J. W. (1977). Exploratory Data Analysis. Reading, MA: Addison-Wesley.

Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. Journal of Experimental Psychology: General, 123, 34-80.

U.S. Department of Energy. (1997). Performance based management handbook. Washington, D.C.: DOE Information, Performance Based Management Interest Group.

Van Selst, M., & Joliceour, P. (1994). A solution to the effect of sample size on outlier elimination. Quarterly Journal of Experimental Psychology, 47A, 631-650.

Wainer, H., & Thissen, D. (1976). Three steps towards robust regression. Psychometrica, 41, 9-34.

Wainer, H., & Thissen, D. (1979). On the robustness of a class of naïve estimators. Applied psychological measurement, 3 (4), 543-551.

Watkins, K. (1993). Discrete event simulation in C. Berkshire, England: McGraw Hill.

Widman, L. E., Loparo, K. A., & Nielsen, N. R. (Eds.). (1989). Artificial intelligence, simulation & modeling. New York: John Wiley and Sons.

Wilcox, R. R. (1992). Why can methods for comparing means have relatively low power and how to cope with the problem. Current Directions in Psychological Science, 1, 101-105.

Wilcox, R. R. (1997). Introduction to robust estimation and hypothesis testing. San Diego, CA: Academic Press.

Wulder, M. (1996). Multivariate statistics: A practical guide. Waterloo, Ontario, Canada: University of Waterloo, Waterloo Laboratory for Earth Observations.

Yashima, N., Nasu, M., Kawazoe, K., Hiramori, K. (1997). Serial evaluation of atrial function by Doppler echocardiography after the maze procedure for chronic atrial fibrillation. European Heart Journal, 18(3), 496-502.

Yu, M. C., & Dunn, O. J. (1982). Robust tests for the equality of two correlation coefficients: A monte carlo study. Educational and Psychological Measurement, 42, 987-1004.