FORMING A BRACKETED INTERVAL AROUND THE TRIMMED MEAN: ALTERNATIVES TO S_W

by

JENNIFER M. BUNNER

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2003

MAJOR: EVALUATION AND RESEARCH

Approved by:

Shlomo S. Sawilowsky 6/25/03

Advisor Date

UMI Number: 3116481

INFORMATION TO USERS

DEDICATION

Dedicated to my mom for always inspiring, supporting, and encouraging me throughout my life, and my husband Scott, who has supported me in this and my many endeavors.

ACKNOWLEDGEMENTS

Grateful appreciation goes to:

my major advisor: Dr. Shlomo Sawilowsky- for all of his help with this dissertation, and for always taking the time to answer my many questions. Dr. Sawilowsky has inspired and motivated me since the day I started in the doctoral program.

my doctoral committee: Dr. Carmone, Dr. Fahoome, and Dr. Markman- for all of the support, advice, and input on this dissertation.

my friend Muhammad Rafiq- who helped on several occasions with SAS (initially coding was attempted using SAS) and FORTRAN questions.

my husband Scott- who also helped with programming, and with several questions and final formatting issues.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Background

Researchers have been concerned with nonnormally distributed data dating back to as early as 1805 (Pearson & Please, 1975; Stigler, 1973). Nonnormally distributed data is common in practice, and has been documented in many applied studies (Micceri, 1989; Pearson & Please, 1975). Nonnormal data is of concern to researchers as it has an effect on statistical procedures such as summary statistics and hypothesis tests. Several studies have documented the effect of nonnormality on measures of central tendency and Student's t (Blair, 1981; Blair & Higgins, 1980a, 1980b, 1985; Bradley, 1968, 1978).

In characterizing and summarizing data (i.e. measures of central tendency, dispersion), the main concern has been robustness of the statistical procedure to nonnormality. The term robust, when used to describe a statistical procedure, refers to the insensitivity of parametric statistics to violations of their assumptions. Nonnormality is an example of an assumption commonly violated. Nonnormality of data results for a variety of reasons. Some examples include growth or decay, where the underlying distribution of the variable is exponential, multimodal lumpy (Micceri, 1989), mass at zero with gap (Sawilowsky & Hillman, 1992), or some other non Gaussian shape. Micceri (1989) gives several examples of nonnormally distributed distributions in the social sciences.

One nonnormal distribution that is of particular interest when speaking of the effects of nonnormality in characterizing and describing data (i.e. measures of central tendency, dispersion) is the contaminated normal distribution. An example of a contaminated normal distribution is a distribution that has essentially a Gaussian structure, however the tails are contaminated by outliers. A normal model can be adopted for such distributions if the outliers can be assumed to have contaminated the model. Distributions such as this motivated the development of robust statistics.

An example of a statistic that is not robust to outliers is the arithmetic mean (Stigler, 1977). The arithmetic mean is a common measure of central tendency. It is a sample statistic that is used as a point estimate of the population parameter μ. The mean however, is not a robust measure of central tendency. With a finite sample breakdown point of 1/n, the mean can change drastically with a change in just one number in the sample.

An example of a robust measure of central tendency is the median. With a finite sample breakdown point of about ½, almost half of the values in the sample could be changed and the median would be unaffected. The median, however, never emerged as a popular measure of central tendency due to problems with estimating the population median from the sample median, difficulty in creating hypothesis tests (sampling distribution is intractable), and reliance on one to two points in the dataset in determining the median.

There are several robust measures of central tendency besides the median. One such measure is the trimmed mean. The trimmed mean first appeared in the literature in 1920 (Daniell, 1920). The trimmed mean can be used when a distribution is nonnormal, and the underlying structure of the data is Gaussian. The trimmed mean is a compromise between the mean (trim=0), and the median (trim=1/2). The trimmed mean of the population will lie somewhere between the population median and the population mean (Wilcox, 1996).

Trimming is essentially the deletion of outliers from the tails of data. Trimming is done as a percentage of observations. For example, a 2 x 10% trimmed mean means that 10% of the observations are deleted from both sides of an ordered data set, and the mean is calculated on the remaining scores.

Problem Statement

When working with the sample trimmed mean, one may want to determine how accurately the sample trimmed mean estimates the population mean μ. Modern textbooks (e.g. Wilcox, 1996) may ask the question: How well does the sample trimmed mean estimate the population trimmed mean? But in this case, the interest is in estimating the population mean from the sample trimmed mean. The reason being, a population may be essentially normal. However, in generating a sample, outliers may appear for a variety of reasons such as keyboard errors during data entry, cheating, etc. These outliers have an effect on sample statistics such as the mean and standard deviation, and the sample statistics may not estimate the population parameters well. Because the population is essentially normal, and a sample may have outliers, the focus question is how well does the sample trimmed mean estimate the population mean. To determine this, a bracketed interval can be formed around the sample trimmed mean to estimate the population mean, just as a bracketed interval can be formed around the mean to estimate the population mean. From a frequentist's perspective, a bracketed interval determines if one can be 95% sure that the population mean is contained within the interval built around the sample trimmed mean (α=.05). A Bayesian's perspective determines if many bracketed intervals were formed, would 95% of the intervals contain the population mean (α=.05). Modern textbooks (e.g., Wilcox, 1996) provide a formula similar to the following for calculating the bracketed interval around the trimmed mean:

$$\text{B.I.}_{1-\alpha}(\mu_t) = \bar{X}t \pm \left(t_{1-\alpha} \text{ x } \frac{1}{1-2\lambda} \text{ x } \frac{s_w}{\sqrt{n}} \right) \quad (1)$$

The three right-most expressions in the above formula constitute the interval of the trimmed mean. $\bar{X}_t$ plus the interval gives the bracketed interval around the trimmed mean. The problem under investigation is the use of s_w as a robust measure of dispersion in calculating the bracketed interval around the trimmed mean.

S_w, the winsorized sample standard deviation, is a robust measure of the population standard deviation, which is unbiased after being divided by the square root of the sample size. Winsorization is similar to trimming, although it is not as harsh on outliers. Instead of deleting or trimming a certain percentage of outliers, winsorizing recodes outliers at the ends of the distribution towards the median. The winsorized standard deviation is calculated by simply taking the standard deviation of the winsorized mean.

Due to the nature of the procedure, winsorizing results in a mass at the recoding points (in the tails of the distribution). There is a larger variance associated with heavy tailed distributions, which results in larger bracketed intervals (Wilcox, 1996). The width of the bracketed interval translates into the accuracy of the estimate of μ. Larger bracketed intervals provide a less accurate estimate of μ than smaller bracketed intervals. Tukey and McLaughlin (1963) is the primary source of support of s_w. There is however, no theoretical underpinning requiring Tukey and McLaughlin's selection of the winsorized procedure as a robust measure of dispersion. There are many robust measures of dispersion in the literature. Lax (1985) identified over 150 different robust measures of dispersion. The purpose of the study was to examine the properties of bracketed intervals formed using alternative measures of dispersion.

Important properties to investigate are bracketed interval coverage, and bracketed interval width. Bracketed interval coverage refers to the probability that the population mean is "covered" in the bracketed interval. If α is set at .05, then there is a 95% probability that the population mean should be contained in the bracketed interval. When used in a bracketed interval, α refers to the probability of an error, or the probability of μ falling outside of the bracketed interval range. Thus, α is related to the coverage. In a hypothesis test, α is the Type I error rate, and indicates the probability of rejecting the null hypothesis when it is true. With regards to hypothesis testing, Bradley, 1978 gave the following criteria for robustness to Type I error: $.5\alpha \leq$ Type I error $\leq 1.5\alpha$ (Liberal criterion); $.9\alpha \leq$ Type

I error ≤ 1.1α (Conservative criterion). These criteria can be extended to the α in bracketed intervals, and can be used to evaluate the robustness of the α, or error. For example, a bracketed interval formed using $\alpha = .05$ should have 95% coverage of the population mean. Bradley's criteria can be extended in the following way: (1 -1.5α) ≤ Coverage (1- α) ≤ (1 -.5α) (Liberal criterion) ;(1- .1.1α) ≤ Coverage (1- α) ≤ (1-.9α) (Conservative criterion)

Specifically, the following research questions were investigated:

1. What is the coverage for the resultant bracketed intervals around the trimmed mean using the alternative measures of dispersion, different trim percentages, and different distributions?

2. Does the coverage meet Bradley's (1978) liberal and conservative criteria for robustness as applied to coverage?

3. How do the widths of the bracketed intervals compare when using different measures of dispersion, different trim percentages, and different distributions in calculating the bracketed interval of the trimmed mean?

This study was a Monte Carlo study that evaluated the coverage and interval widths for the resultant bracketed intervals around the trimmed mean using alternative measures of dispersion. The sample winsorized standard deviation and the sample trimmed standard deviation were investigated for comparison purposes. In this study, measures of dispersion, alpha level, and distribution type were manipulated. The coverage of the trimmed mean bracketed interval and the bracketed interval width around the trimmed mean were observed.

Coverage was evaluated by repeatedly sampling from the distribution of interest, calculating the bracketed interval around the sample trimmed mean using the alternative measures of dispersion, and checking the interval to see if the population parameter (population mean) was found within the interval. The number of times that the population parameter is found within the interval, divided by the number of times the experiment is performed, will give the coverage. The resultant coverage rate was evaluated against Bradley's (1978) liberal and conservative criteria for Type I error rates as applied to coverage to determine if the coverage rates were inflated (the percentage of bracketed intervals in which the population parameter is found within the interval exceeds the 1-nominal alpha level), or underestimated (the percentage of bracketed intervals in which the population parameter is found within the interval is less than the 1-nominal alpha level).

The widths of the bracketed intervals were calculated by subtracting the lower value of the bracketed interval from the upper value of the bracketed interval. The Location Relative Efficiency (LRE) (Sawilowsky, 2002) was calculated for each interval. The LRE was used to compare the widths of the bracketed intervals for each alternative measure of dispersion. In this study the LRE was defined as follows:

$$LRE = \frac{Sw(U_{B.I.}) - Sw(L_{B.I})}{X(U_{B.I.}) - X(L_{B.I})} \quad (2)$$

The LRE is the ratio of the range of the bracketed interval for the sample winsorized standard deviation, and the bracketed interval for the competitor. In this case all competitors are compared to sample winsorized standard deviation because it is the measure of dispersion that is supported in calculating the bracketed interval around the trimmed mean. The ratio may be greater or less than one. A LRE greater then one indicates that the competitor yields narrower bracketed intervals than the sample winsorized standard deviation.

Definitions of Terms

Bradley's Criteria: Criteria to evaluate the robustness of a statistic in regards to Type I error rates. $.5\alpha \leq$ Type I error $\leq 1.5\alpha$ (Liberal criterion); $.9\alpha \leq$ Type I error $\leq 1.1\alpha$ (Conservative criterion) (Bradley, 1978).

Bracketed interval: An interval or range of values that contains μ with a probability of $1-\alpha$ (Wilcox, 1996).

Bracketed Interval Width: "The distance between the upper and lower end of the bracketed interval" (Wilcox, 1996).

Measures of Dispersion: "Numerical quantities intended to indicate how spread out a batch of numbers happens to be" (Wilcox, 1996).

Monte Carlo Methods: "Any procedure that involves statistical sampling techniques in obtaining a probabilistic approximation to the solution of a mathematical or physical problem (James & James, 1992).

Normal Distribution: "A purely theoretical continuous probability distribution in which the horizontal axis represents all possible values of a variable and the vertical axis represents the probability of those values occurring" (Vogt, 1993). "A random variable X is **normally distributed** or is a **normal random variable** if there are numbers m and σ for which X has the probability density function

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(t-m)^2/\sigma^2} \quad -\infty < t < \infty$$ (James & James, 1992)

Outlier: "Outliers are cases with scores that are very different from the rest…Although there is no absolute definition of "extreme", a common rule of thumb is that scores more than three standard deviations away from the mean may be outliers" (Kline, 1998)

Robustness: "A statistical procedure is said to be **robust**, with respect to a particular postulated assumption, if the procedure is relatively insensitive to (slight) departures from the assumption" (Hollander & Wolfe, 1973).

Sample Mean: The arithmetic average when otherwise unspecified (Vogt, 1993). "The **sample mean** is equal to the sum of the measurements divided by the number of measurements contained in the batch of numbers" (Wilcox, 1996).

Sample Median: The middle score (or average of two middle scores) in a set of ranked scores (Vogt, 1993)

Sample Size: The number of scores in a subset of the population (sample) (Wilcox, 1996).

Sample Trimmed Mean: A robust measure of central tendency which is a compromise between the sample mean and the sample median. To compute the trimmed mean, a percentage of scores is removed from the sample tails of ranked data, and the average of these numbers is calculated (Wilcox, 1996).

Standard Deviation: A statistic that shows the spread or dispersion of scores in a distribution of scores; in other words, a measure of dispersion. The more widely the scores are spread out, the larger the standard deviation. The standard deviation is calculated by taking the square root of the variance" (Vogt, 1993).

Type I error: The probability of rejecting the null hypothesis when it is in fact true (Wilcox, 1996).

Winsorized Sample Standard Deviation: A robust statistic that shows the spread of scores in a distribution of scores. To compute the winsorized standard deviation, a percentage of scores is replaced (g smallest values are replaced with $X_{(g+1)}$, and g largest values are replaced with $X_{(n-g)}$) in the sample tails of ranked data, and the sample variance of these numbers is computed. The standard deviation is calculated by taking the square root of the variance (Wilcox, 1996).

Assumptions and Limitations

Assumptions

A Monte Carlo study requires random sampling, thus the accuracy of the Monte Carlo study will depend on the random number generator used in the sampling process. In this study, observations were sampled independently with replacement from a Gaussian distribution, smooth symmetric distribution, digit preference distribution, and discrete mass at zero without gap distribution

Limitations

There were seven limitations to the study. All of these limitations have an impact on the generalizability of the study.

1. The study was focused on a limited number of distributions with an underlying symmetric structure.
    a. Distributions that have an underlying Gaussian structure but have outliers
        i. One wild score on the left of the distribution
        ii. Two wild scores on the left of the distribution
        iii. Three wild scores on the left of the distribution
        iv. Three wild scores on the left of the distribution and one wild score on the right of the distribution
    b. Smooth Symmetric Distribution (Micceri, 1989) with outliers: A real symmetric dataset of achievement test scores with added outliers
        i. One wild score on the left of the distribution
        ii. Two wild scores on the left of the distribution
        iii. Three wild scores on the left of the distribution
        iv. Three wild scores on the left of the distribution and one wild score on the right of the distribution
    c. Digit Preference (Micceri, 1989): A real symmetric dataset of achievement test scores
    d. Discrete Mass at Zero Without Gap (Micceri 1989): A real symmetric dataset of achievement test scores
2. The study was limited to a sample size of 30
3. The study was limited to the bracketed interval of the trimmed mean, and does not include hypothesis testing

4. The study was limited to the 2 X 10% and 2 x 20% trimming/winsorizing/recoding
5. The study was limited to symmetric trimming/winsorizing/recoding only. An applied researcher will likely not have an a priori knowledge of the symmetry/asymmetry of the population, and will not know what side to trim.
6. The study was limited to nominal alpha levels of .05 and .01
7. The study involved 10,000 replications

CHAPTER 2

REVIEW OF THE LITERATURE

Robust Statistics

When used to describe a statistical procedure, the term robust has been referred to as the insensitivity of parametric statistics to violations of their assumptions. Glass, Peckman and Sanders (1972) referred to robustness in terms of probability statements and Type I error. A statistical test is robust if the test preserves the validity of the probability statements applied to it when the assumptions upon which the test is based are violated (p. 274).

Bradley (1978) struggled with the definition of robustness as given by Glass et al. (1972) and other similar definitions. This struggle was due to the lack of a quantitative indication of how much deviation from alpha can occur for a test to be considered robust. Using the definition given by Glass et al. (1972), a researcher could make a claim such as "the test was robust to violations of the assumptions of normality as there was only a small distortion of the Type I error rate". The lack of a quantitative indication of the distortion in alpha levels led to an inconsistency in the categorization of the robustness of parametric statistics to departures from their assumptions (Bradley, 1978). Bradley (1978) proposed the following criteria as a quantitative indication of robustness: $1.5\alpha \leq$ Type I error $\leq .5\alpha$ (Liberal criterion); $1.1\alpha \leq$ Type I error $\leq .9\alpha$ (Conservative criterion).

Bradley (1978) also identified factors that affect the robustness of a procedure. These factors are: the size of α; the location of the rejection region; for the smallest sample, the absolute size of the sample and the absolute shape of the population from which it was drawn; and the absolute shape, relative shape and relative variance, of the population from which it was drawn. Bradley stated that all of these factors can have a strong interaction. Therefore, in order to make meaningful statements about the robustness of a statistical procedure, these factors should be qualified. This qualification should be combined with the above quantitative criteria for robustness when making claims of robustness for a statistical procedure.

In addition to the definition of the term robustness given by Glass et al. (1972), with the development of statistical procedures referred to as robust methods, the term robust has taken on another meaning. Wilcox (1997), in discussing robust methods, defined the term robustness as the following:

> For many years the term robust was taken to mean that a particular hypothesis-testing method controls the probability of Type I error.... Today, however, from a statistical point of view, the term robust means much more. Roughly, robust methods include finding population parameters, estimators, and hypothesis methods that are not affected by small changes in a distribution. (p. 5)

Nonnormality of data results for a variety of reasons, and nonnormal data can take on many shapes (Micceri, 1989). One nonnormal distribution in particular that is of importance to robust methods is the contaminated distribution. An example of a contaminated distribution is a distribution with an essentially Gaussian shape that is contaminated by outliers. Robust methods were motivated by distributions such as this, where outliers can be assumed to have contaminated the model. Outliers are values in the data that are considerably large or small compared to the other values in the dataset. The presence of outliers can bias common measures of central tendency and dispersion such as the sample mean, and sample variance.

When data is normally distributed, measures of central tendency such as the mean, median, and mode of the data fall in the same location. In the case where the data deviates slightly from the normal distribution, such as a distribution that is essentially normal but has outliers on the right (skewed to the right), the mean, median, and mode of the distribution do not fall in the same location. The mean of the data will fall towards the skewed end of the distribution. The mean is therefore not a good or robust estimate of the center of the distribution in this case as it is biased towards the skewed end of the distribution (towards the outliers).

The presence of outliers in the data not only has an effect on measures of central tendency, but also on measures of dispersion. One of the major concerns with distributions with outliers is that the variance of this type of distribution is larger than that of a normal distribution. Wilcox (1997) gives an example of a contaminated distribution with heavy tails and the implications of the contamination on the standard deviation of the distribution. In this example, two subpopulations are assumed, and a sample of each subpopulation is taken and combined into a new population. The first subpopulation has a standard normal distribution, while the other has a normal distribution with a mean of zero and standard deviation of ten. Ninety percent of the observations in the new population were taken from the standard normal distribution, and ten percent of the observations were taken from the normal distribution with a standard deviation of ten. The new population has a standard deviation of 3.3, while the normal distribution has a standard deviation of 1. The contaminated distribution has a standard deviation that is 3.3 times higher than the normal distribution. "Outliers and heavy-tailed distributions are serious practical problems because they inflate the standard error of the sample mean…Modern robust methods provide an effective way of dealing with this problem." (p. 2)

The Trimmed Mean: A Robust Estimator of Location

The trimmed mean was identified as a robust measure of central tendency by P.J. Daniell (Stigler, 1973). In his paper (Daniell, 1920), Daniell proposed the discard average (trimmed mean) as an alternative to the sample mean, where an outer fraction of measures was discarded from a sample, and the mean of the sample was then calculated. Daniell proposed the discard average as a way to deal with outliers.

The trimmed mean is a compromise between the sample mean and the sample median. The sample mean is calculated using all values in the sample (0% trim), and the sample median is calculated using one or two values (50% trim). The trimmed mean of the sample will lie somewhere between the sample median and the sample mean (Wilcox, 1996).

The sample mean is the most commonly used measure of central tendency (it is a point estimate of the population parameter μ); it is however, not a robust measure. The mean has a finite sample breakdown point of 1/n. This means that the mean of the sample will change with a change in only one observation. For a population with outliers, the outliers can vastly distort the value of the sample mean.

The median is a common measure of central tendency that can be found in many textbooks. Unlike the mean, the median is a robust measure of central tendency. The median has a finite sample breakdown point of about ½, which means that almost half of the values in the sample could be changed and the median would be unaffected. The median however, has been criticized for several reasons. These include: (1) intractability of the sampling distribution of the median (finding the sampling distribution of the median requires reliance on asymptotic variances or some other approach), which makes the construction of hypothesis testing difficult, (2) the sample median is usually not a good estimate of the population median, (3) the sample median relies on only one of 2 values in the sample, so in some sense information is lost (Wilcox, 1996).

The trimmed mean can be used when a distribution is nonnormal, and the underlying structure of the data is Gaussian. For example, a distribution that is Gaussian in nature with outliers. Trimming is essentially the deletion of outliers from the tails of data. To calculate the trimmed mean of a sample, a certain percentage of the smallest and largest values of an ordered dataset are removed, and the mean of this new dataset is calculated.

Trimming is done as a percentage from the tails of a distribution. For example, a 2 x 10% trim means that 10% of the values are deleted from both sides of the dataset. As an example, a 2 x 10% trim is calculated for the dataset in Table 1. The first step is to order the values in the dataset from low to high.

Next, a 2 x 10% trim is performed on the dataset. This means that a 1 x 10% trim is performed on each tail of the distribution. In this example, a 10% trim on each end of the distribution equates to 1 value, since there were 10 values before trimming. The amount to trim is represented by the symbol *g*. *g* in this case is equal to [.1n], the trim percent multiplied by the number of values in the dataset before trimming, or 1 value. The brackets mean to compute .1n and then round down to the nearest integer. Finally, the mean is calculated for the remaining 8 scores.

Table 1. Trimming of a dataset

Original	85	92	87	93	99	86	88	90	73	91
Ordered	73	85	86	87	88	90	91	92	93	99
Trimmed		85	86	87	88	90	91	92	93	

The trimmed mean is calculated as follows:

$\bar{X}_t = (85 + 86 + 87 + 88 + 90 + 91 + 92 + 93) / 8$

Forming a Bracketed Interval Around the Trimmed Mean

When working with the sample trimmed mean, one may want to form a bracketed interval around it to determine how well the sample trimmed mean is estimating the population mean. The population mean can be estimated by forming a bracketed interval around the sample trimmed mean, and can be calculated in a similar way as the bracketed interval of the population mean. The formula for forming a bracketed interval around the trimmed mean is as follows:

$$\mathrm{B.I.}_{1-\alpha}(\mu_t) = \bar{X}t \pm \left(t_{1-\alpha} \text{ x } \frac{1}{1-2\lambda} \text{ x } \frac{s_w}{\sqrt{n}} \right) \quad (1)$$

This formula can be found in many modern statistics textbooks (e.g., Wilcox, 1996).

The right side of the equation for calculating the bracketed interval around the trimmed mean consists of four expressions. The first expression $\bar{X}_t$ represents the sample trimmed mean.

The second expression $t_{1-\alpha}$, is the two sided Student's T distribution with v degrees of freedom. The degrees of freedom v is equal to n-2*g*-1, where n is the sample size before trimming, g is the number of values to trim (as previously defined [trim% x n]). Using the example in Table 1, the degrees of freedom would be calculated in the following manner; n = 10, g =.1 x 10 = 1, and v=10 – (2 x 1) – 1 = 7. For, α = .05, the $t_{1-\alpha}$, is equal to $t_{.975}$ = 2.365.

The third term is a multiplier that is used to adjust the standard error of the sample trimmed mean. If there has been no trimming, the multiplier will be equal to 1. If trimming has taken place, λ is equal to the trim percent. If a 2 x 10% trim is being applied to a dataset, λ is equal to .10, and the expression would be equal to 1.25, or 1/.8.

The fourth term is the standard error. S_w is the winsorized sample standard deviation. The winsorized sample standard deviation is a robust measure of dispersion, and is an estimate of the population standard deviation. The standard error of the sample trimmed mean is estimated by dividing s_w by the square root of the sample size before trimming.

Winsorization is a method for dealing with outliers that is not as harsh as trimming. While trimming results in the deletion of outliers, winsorization deals with outliers by recoding them back towards the median of the distribution. Like trimming, winsorization is focused on the tails of the distribution. Winsorization is done as a percentage from the tails of a distribution. For example, a 2 x 10% winsorization means that 10% of the values are winsorized or recoded from both sides of the dataset. As with trimming, the amount to winsorize is represented by the symbol *g*. For a sample of 10, *g* is equal to [.1n], the winsorization percent multiplied by the number of values in the dataset before winsorizing, or 1 value. As with trimming, the brackets mean to compute .1n and then round down to the nearest integer. The recoding scheme is as follows: the *g* smallest values are replaced with $X_{(g+1)}$, and the g largest values are replaced with $X_{(n-g)}$. An example of a 2 x 10% winsorization is presented in Table 2.

Table 2. Winsorization of a dataset

Original	85	92	87	93	99	86	88	90	73	91
Ordered	73	85	86	87	88	90	91	92	93	99
Winsorization	85	85	86	87	88	90	91	92	93	93

The two most extreme values were recoded back. For example, the 73 was recoded to an 85, and the 99 was recoded to a 93.

To calculate the winsorized variance the following equation is used:

$$s^2_w = SSD_w/(n-1)$$

SSD_w is the winsorized sum of squared deviations. SSD_w is calculated as follows:

$$SSD_w = (g+1)\,(X_{(g+1)} - \bar{X}_w)^2 + \ (X_{(g+2)} - \bar{X}_w)^2 + \ldots$$

$$+ (X_{(n-g-1)} - \bar{X}_w)^2 + (g+1)\,(X_{(n-g)} - \bar{X}_w)^2 \text{ (Wilcox, 1996)}$$

The square root of the winsorized variance gives the winsorized standard deviation. As an example of how the winsorized standard deviation is a robust measure of dispersion, the winsorized standard deviation of the dataset in Table 2 is 3.2, while the standard deviation without winsorization is 6.8.

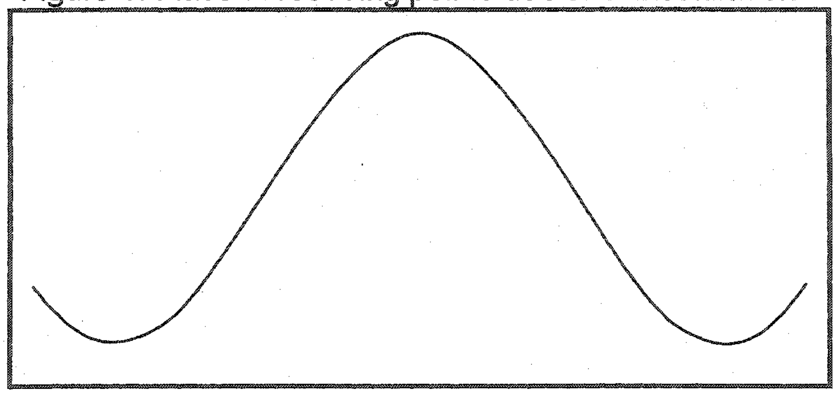
The difference between the formula for the bracketed interval of the of population mean as estimated by the trimmed mean, and the standard formula for the bracketed interval of the population mean, is the degrees of freedom for $t_{1-\alpha}$, the multiplier that is used to adjust the standard error, and the estimate of the population standard deviation s_w. The focus of this paper is the use of s_w as a robust measure of dispersion in calculating the bracketed interval around the trimmed mean.

When modern statistics textbooks give equation (1) for calculating the bracketed interval of the trimmed mean, Tukey and McLaughlin (1963) is cited as the source of support for using s_w. Gross (1976) claimed that Tukey and McLaughlin showed empirically that "the winsorized variance is a consistent estimator of the variance of the trimmed mean". This view is held by many authors of textbooks and journal articles, and the use of s_w in the equation for calculating the trimmed mean has received little challenge since it was proposed by Tukey and McLaughlin.

In their paper, Tukey and McLaughlin (1963) considered several measures of dispersion in the equation for the bracketed interval around the trimmed mean. They sought to use a more robust procedure than the sample standard deviation, as the sample standard deviation is not robust to outliers. As they tried different measures of dispersion, they sought to satisfy the primary condition that the average value of the denominator and numerator were "in constant proportion over as broad a spectrum of symmetrical distributions as is reasonably convenient". Tukey and McLaughlin tried the sample standard deviation, and the sample trimmed standard deviation, but these measures of dispersion did not satisfy the matching condition. The sample winsorized standard deviation was proposed because it is not as harsh on outliers as the sample trimmed standard deviation. Tukey and McLaughlin wanted to give more attention to outliers in the sample rather than simply deleting them. The sample winsorized standard deviation was found to satisfy the primary condition of matching, and the approach was adopted.

Although Tukey and McLaughlin did find the winsorized standard deviation to be a robust measure of dispersion in calculating the bracketed interval around the trimmed mean, their treatment of this subject is certainly not theoretical. In addition, due to the nature of the winsorization procedure, winsorizing results in a mass at the recoding points (in the tails of the distribution). This mass at the recoding points is shown in Figure 1. A mass at the recoding points results in a heavy tailed distribution. This can become problematic in creating bracketed intervals, as heavy tailed distributions commonly result in larger bracketed interval widths (Wilcox, 1996). The width of the bracketed interval translates into the accuracy of the estimate of μ, and larger bracketed intervals provide a less accurate estimate of μ than smaller bracketed intervals. Following the publication of Tukey and McLaughlin, 1963, there has not been much treatment of this subject in the literature.

Figure 1: Mass at recoding points due to Winsorization



Gross (1976) examined several 95% bracketed intervals for long tailed symmetric distributions using Monte Carlo techniques. In this study, Gross evaluated the bracketed intervals of 25 measures of central tendency to determine what estimator is best in terms of bracketed interval width (smaller widths being better than longer widths), and control of Type I error rates. Gross tested these 25 estimators with 7 distributions using sample sizes of 10 and 20. Each Monte Carlo experiment was repeated 1,000 or 640 times (repetitions depended on the distribution that was used).

The estimators in this study that are of interest with regards to the bracketed interval around the trimmed mean are: trimmed mean with winsorized standard deviation (10%, 25%, and 35% trimming), and trimmed mean with minimum estimated winsorized standard deviation. The trimmed mean with minimum estimated winsorized standard deviation is the only instance where the standard deviation of the trimmed mean bracketed interval was manipulated.

Two distributions in this study that are of interest with regard to the bracketed interval around the trimmed mean are the 10G/20, and 10G/4 distributions. These distributions both had simulated outliers. The 10G/20 was a

Gaussian distribution of sample size 20 with one observation multiplied by 10. The 10G/4 was a Gaussian distribution of sample size 20 with 5 observations multiplied by 10.

With respect to these distributions, the 2 x 10% trimmed mean with winsorized standard deviation had the shortest interval width for the 10G/20, and the 2 x 25% trimmed mean with winsorized standard deviation had the shortest interval width for the 10G/4 distribution. The 2 x 10% trimmed mean with winsorized standard deviation was most robust with respect to Type I error for the 10G/20 distribution, and the 2 x 25% trimmed mean with winsorized standard deviation was the most robust for the 10G/4 distribution.

The trimmed mean with minimum estimated winsorized standard deviation resulted in shorter bracketed interval widths than the 2 x 35% trimmed mean with winsorized standard deviation for the 10G/20 distribution, and shorter bracketed interval widths than the 2 x 10% trimmed mean with winsorized standard deviation, and the 2 x 35% trimmed mean with winsorized standard deviation for the 10G/4 distribution. With respect to Type I error, the trimmed mean with minimum estimated winsorized standard deviation was more robust than the 2 x 35% trimmed mean with winsorized standard deviation for the 10G/20 distribution, and more robust than the 2 x 10% trimmed mean with winsorized standard deviation, and the 2 x 35% trimmed mean with winsorized standard deviation for the 10G/4 distribution.

Although the findings of this study did not indicate that the trimmed mean with minimum estimated winsorized standard deviation performed better than a trimmed mean (2 x 10%, 25%, 35%) with winsorized standard deviation, there were instances where the trimmed mean with minimum winsorized standard deviation performed better than one of the trimmed means with winsorized standard deviation with respect to bracketed interval width or robustness to Type I error. This suggests that further study should be done concerning the use of other robust measures of dispersion in forming a bracketed interval around the trimmed mean.

Bunner and Sawilowsky (2002) examined 4 alternative measures of dispersion in forming the bracketed interval around the trimmed mean; the mean deviation (S_{md}), the median deviation (S_{mdd}), the median absolute deviation (S_{mad}), and the Bunner-Sawilowsky approach (S_{bs}). Of these four measures of dispersion considered, only the median absolute deviation and the Bunner-Sawilowsky approach are robust to outliers. The mean deviation and median absolute deviation all incorporate the mean in their calculation, and are therefore not robust to outliers. Each alternative measure was substituted for s_w in the equation for calculating the bracketed interval around the trimmed mean. Bracketed interval widths, and robustness to Type I error was evaluated for each approach.

The mean deviation is calculated as follows: subtract the mean from each score, take the absolute value of this result, sum the results, and divide by N-1.

The median deviation is calculated in a similar way as the mean deviation, except for instead of subtracting the mean from each score, the median is subtracted from each score.

The median absolute deviation is similar to the median deviation, except for instead of taking the mean of the sum of absolute values of the deviations from the median, the median of these values is taken.

The idea behind the Bunner-Sawilowsky approach is to perform a procedure similar to winsorization without creating a mass at the endpoints. An example of the Bunner-Sawilowsky approach is presented in Table 3. A 2 x 20% recoding scheme is used.

Table 3. The Bunner-Sawilowsky approach applied to a dataset

Original	85	92	87	93	99	86	88	90	73	91
Ordered	73	85	86	87	88	90	91	92	93	99
Bunner-Sawilowsky	86	87	86	87	88	90	91	92	91	92

In this example, two scores are recoded in the tails of the distribution. As an example, in the lower tails of the distribution, the 73 and 85 will be recoded. The 85 is recoded back two scores to 87, and the 73 is recoded back two scores to 86. The amount that the scores are recoded back, is equal to the percentage recoded (In this example 20% of the sample is 2, so each score to be recoded is pushed 2 scores towards the median).

This study was a Monte Carlo study with a sample size of 30. Each experiment was repeated 1,000 times. Four models of outliers were used: one wild score of the right of the distribution, two wild scores on the left of the distribution, three wild scores on the left of the distribution, and three wild scores on the left, and one wild score on the right of the distribution.

With respect to Type I error, the mean deviation, median deviation, and median absolute deviation resulted in Type I errors that were greatly inflated. These procedures did result in bracketed interval widths that were shorter than

s_w, but since the Type I error rates were inflated, these measures are not considered good alternatives.

The Bunner-Sawilowsky approach resulted in bracketed intervals that were shorter than s_w, and Type I error rates that were robust according to Bradley's (1978) liberal criterion. The bracketed interval widths were 5.13% shorter than bracketed interval widths using s_w, however the Type I error rates were not robust to Type I error according to Bradley's (1978) conservative criterion. This suggests that further study of this and other recoding schemes should be investigated.

With respect to the amount to trim, the literature on this subject is divided. Several studies have supported heavy trimming (e.g. 2 x 25% by Rosenberger & Gasko, 1983 or 2 x 20% by Wilcox, 1996) and lightly trimming (e.g. 2 x 10% or 2 x 5%, Stigler, 1977, Staudte & Sheather, 1990). Wilcox (1996) specified that "some trimming often gives substantially better results compared to no trimming", but did not support a particular level of trimming with certainty. Sawilowsky (2002) found that lightly trimmed means of 2 x 5% yielded narrower 95% bracketed intervals than heavily trimmed means of 2 x 25%. In this study, samples from applied social and behavioral datasets (Micceri, 1989) were used. Sample sizes of 10 in increments of 10 to 100 were used. Bracketed interval widths were compared using the Location Relative Efficiency.

Monte Carlo Methods

"Monte Carlo methods are any procedures that involve statistical sampling techniques in obtaining a probabilistic approximation to the solution of a mathematical or physical problem" (James & James, 1992). The random numbers that are generated in a Monte Carlo simulation are not truly random, rather they are pseudo random. The reason that these numbers are considered pseudo random is due to the nature of the random number generator. A seed number is supplied to the random number generator, and then a series of mathematical operations are applied to the seed number to generate the values in the random sample. The numbers that are generated are not truly random because the seed number is known, and all values generated in a sample can be determined.

CHAPTER 3

METHODS

Overview

This study used Monte Carlo methods to compare the performance of six alternative measures of dispersion in calculating the bracketed interval around the trimmed mean. The bracketed intervals for each measure of dispersion were compared in terms of bracketed interval width using Location Relative Efficiency (LRE) and robustness of the bracketed interval coverage (evaluated against Bradley's (1978) liberal and conservative criteria for Type I error rates as applied to coverage). Bracketed interval width and robustness of bracketed interval coverage were also calculated using the sample winsorized standard deviation, and the trimmed sample standard deviation.

The winsorized standard deviation was included in this study in order to test the performance of s_w across different symmetric distributions, different trim percentages (2 x 10%, 2 x 20%), and alpha levels (.05, .01).

S_t was included in this study because it is seen a "naïve" measure of dispersion for consideration in calculating the bracketed interval around the trimmed mean. Trimming does not create a buildup of data in the tails of the distribution, however trimming of the data set is rather extreme in the treatment of outliers. Trimming of the data set decreases the degrees of freedom of the sample resulting in a larger standard error than if the sample had a larger degree of freedom.

Measures of Dispersion

Six alternatives to s_w were considered in this study. Each alternative is described below.

Recoding Method 1

Method 1 is a combination of trimming and winsorization. Suppose that a 2 x 20% recode is desired for a sample that may contain outliers. 2 x 10% of the sample will be trimmed, and 2 x 10% of the sample will be winsorized (2 x 10%T x 10%W). This procedure first utilizes trimming to treat the outliers, and then winsorizes the remaining values to be treated. An example of recoding Method 1 is presented below in Table 4. A 2 x 20% recode is utilized to demonstrate this procedure. First the data are ordered. Next, 10% of the values on each side of the distribution are trimmed; in this case the values of 73 and 99 are trimmed. Finally, 10% (of the original data, in this case n = 1) of the values are winsorized; in this case the values of 85 and 93 are winsorized to 86 and 92 respectively.

Table 4. Recoding Method 1 applied to a dataset

Original	85	92	87	93	99	86	88	90	73	91
Ordered	73	85	86	87	88	90	91	92	93	99
Method 1		86	86	87	88	90	91	92	92	

The standard deviation of the recoded data using Method 1 and a 2 x 20% recode is 2.56.

$S_{recode1}$ was considered in this study because Recode Method 1 is a compromise between all out trimming or winsorizing. There is not as much of a buildup of data in the tails of the distribution as this recoding method does not rely only on winsorizing, and there isn't as extreme of a decrease in the degrees of freedom as this method does not rely only on trimming.

Recoding Method 2

Method 2 is a recoding procedure where the values to be recoded take on the next value in the data set towards the median. An example of recoding Method 2 is presented below in Table 5. A 2 x 20% recode is utilized to demonstrate this procedure. First the data are ordered. Next, 20% of the values on each side of the distribution are recoded to the next value towards the median; in this case the values of 73 and 99 are recoded to 85 and 93, and the values of 85 and 93 are recoded to 86 and 92 respectively.

Table 5. Recoding Method 2 applied to a dataset

Original	85	92	87	93	99	86	88	90	73	91
Ordered	73	85	86	87	88	90	91	92	93	99
Method 2	85	86	86	87	88	90	91	92	92	93

The standard deviation of the recoded data using Method 2 and a 2 x 20% recode is 2.94.

$S_{recode2}$ was considered in this study because Recode Method 2 is a compromise to winsorizing. Recode Method 2 does not involve any trimming, so there is no loss in the degrees of freedom. Also, there is not a buildup of data in the tails of the distribution as this recoding method deals with outliers by recoding them towards the median smoothly. The decrease in variance is not as extreme as winsorization.

Recoding Method 3

Method 3 is a modified winsorization. Using this method, the winsorized value is found, and the average of the winsorized value and the value to be recoded is the recoded value. An example of recoding Method 3 is presented below in Table 6. A 2 x 20% recode is utilized to demonstrate this procedure.

First the data are ordered. Next, the winsorized value is found for all values to be recoded. In this example the winsorized value is 86 for the values to recoded on the left side of the distribution, and 92 for the values to be recoded on the right side of the distribution. Finally, the average of the value to be recoded and the winsorized value of the value to be recoded is calculated. For example, the average of 73 and 86 is the recoded value for 73 (79.5).

Table 6. Recoding Method 3 applied to a dataset

Original	85	92	87	93	99	86	88	90	73	91
Ordered	73	85	86	87	88	90	91	92	93	99
Method 3	79.5	85.5	86	87	88	90	91	92	92.5	95.5

The standard deviation of the recoded data using Method 3 and a 2 x 20% recode is 4.52.

$S_{recode3}$ was considered in this study because Recode Method 3 is a compromise to winsorizing. Recode Method 3 does not involve any trimming, so there is no loss in the degrees of freedom. Also, there is not a buildup of data in the tails of the distribution as this recoding method deals with outliers by recoding them towards the median smoothly. The decrease in variance is not however as extreme as Recode Method 2.

Recoding Method 4

Method 4 is a variation of Method 3. Using this method, the winsorized value is found, and the median of the winsorized value and the value to be recoded is the recoded value. An example of recoding Method 4 is presented below in Table 7. A 2 x 20% recode is utilized to demonstrate this procedure. First the data are ordered. Next, the winsorized value is found for all values to be recoded. In this example the winsorized value is 86 for the values to recoded on the left side of the distribution, and 92 for the values to be recoded on the right side of the distribution. Finally, the median of the value to be recoded and the winsorized value of the value to be recoded is calculated. For example, the median of 73 and 86 is the recoded value for 73 (85).

Table 7. Recoding Method 4 applied to a dataset

Original	85	92	87	93	99	86	88	90	73	91
Ordered	73	85	86	87	88	90	91	92	93	99
Method 4	85	85.5	86	87	88	90	91	92	92.5	93

The standard deviation of the recoded data using Method 4 and a 2 x 20% recode is 3.06.

$S_{recode4}$ was considered in this study because Recode Method 4 is a compromise to winsorizing. Like Recode Methods 2 and 3, Recode Method 4 does not involve any trimming, so there is no loss in the degrees of freedom. Also, there is not a buildup of data in the tails of the distribution as this recoding method deals with outliers by recoding them towards the median smoothly. The decrease in variance is not however as extreme as Recode Method 2, but is more extreme than Recode Method 3. The variance will fall between the variance of Recode Methods 2 and 3.

Recoding Method 5

Method 5 is a modification to the Bunner-Sawilowsky approach. Using this method, the Bunner-Sawilowsky value is found. Next, the average of the Bunner-Sawilowsky value and the value to be recoded is taken to get the recoded value. An example of recoding Method 5 is presented below in Table 8. A 2 x 20% recode is utilized to demonstrate this procedure. First the data are ordered. Next, the Bunner-Sawilowsky value is found for all values to be recoded. In this example the Bunner-Sawilowsky values are 87 and 86 for the values of 85 and 73, and 91 and 92 for the values of 93 and 99 respectively. Finally, the average of the Bunner-Sawilowsky value and the value to be recoded is calculated to get the recoded value. Using the original value of 85 as an example, the Bunner-Sawilowsky value is found to be 87. Next, the average of the Bunner-Sawilowsky value and the value to be recoded (85) is taken. In this case the new recoded value is 86.

Table 8. Recoding Method 5 applied to a dataset

Original	85	92	87	93	99	86	88	90	73	91
Ordered	73	85	86	87	88	90	91	92	93	99
Method 5	79.5	86	86	87	88	90	91	92	92	95.5

The standard deviation of the recoded data using Method 5 and a 2 x 20% recode is 4.44.

$S_{recode5}$ was considered in this study as a variation of the Bunner-Sawilowsky that gives some weight to outliers. Like Recode Methods 2, 3, and 4, Recode Method 5 does not involve any trimming, so there is no loss in the degrees of freedom. Also, there is not a buildup of data in the tails of the distribution as this recoding method deals with outliers by recoding them towards the median smoothly. The decrease in variance is not however as extreme as Recode Methods 2 and 4, but is more extreme than Recode Method 3 as the Bunner-Sawilowsky recoding scheme brings outlying values closer to the median than winsorization.

Recoding Method 6

Method 6 is a variation of Method 5. Using this method, the Bunner-Sawilowsky value is found. Next, the median of the Bunner-Sawilowsky value and the value to be recoded is taken to get the recoded value. An example of recoding Method 6 is presented below in Table 9. A 2 x 20% recode is utilized to demonstrate this procedure. First the data are ordered. Next, the Bunner-Sawilowsky value is found for all values to be recoded. In this example the Bunner-Sawilowsky values are 87 and 86 for the values of 85 and 73, and 91 and 92 for the values of 93 and 99 respectively. Finally, the median of the Bunner-Sawilowsky value and the value to be recoded is calculated to get the recoded value. Using the original value of 85 as an example, the Bunner-Sawilowsky value is found to be 87. Next, the median of the Bunner-Sawilowsky value and the value to be recoded (85) is taken. In this case the new recoded value is 86.

Table 9. Recoding Method 6 applied to a dataset

Original	85	92	87	93	99	86	88	90	73	91
Ordered	73	85	86	87	88	90	91	92	93	99
Method 6	85	86	86	87	88	90	91	92	92	93

The standard deviation of the recoded data using Method 6 and a 2 x 20% recode is 2.94.

$S_{recode6}$ was considered in this study as a variation of the Bunner-Sawilowsky that gives some weight to outliers though not as much weight as $S_{recode5}$ as the median rather than the mean of the value to be recoded and the Bunner-Sawilowsky value is taken. Like Recode Methods 2, 3, 4, and 5, Recode Method 6 does not involve any trimming, so there is no loss in the degrees of freedom. Also, there is not a buildup of data in the tails of the distribution as this recoding method deals with outliers by recoding them towards the median smoothly. The decrease in variance is not however as extreme as Recode

Method 2, but is more extreme than Recode Method 3 and 4 as the Bunner-Sawilowsky recoding scheme brings outlying values closer to the median than winsorization, and more extreme than Recode Method 5 as the median rather than the average is used.

Summary

A summary of the standard deviations for the alternative are given in Table 10. All standard deviations are computed using a 2 x 20% trim/winsor/recode for example purposes.

Table 10. Summary of Alternative Measures of Dispersion

Ordered	73	85	86	87	88	90	91	92	93	99
(2 x 20% trimming/winsorizing/recoding)										

Recoding Method	**Standard Deviation**
Trimmed (s_t)	2.37
Winsorized (s_w)	2.67
Recode 1 ($s_{recode1}$)	2.56
Recode 2 ($s_{recode2}$)	2.94
Recode 3 ($s_{recode3}$)	4.52
Recode 4 ($s_{recode4}$)	3.06
Recode 5 ($s_{recode5}$)	4.44
Recode 6 ($s_{recode6}$)	2.94

Computer/Programs Used

Computer: A 900 MHz Hewlett Packard HP Notebook PC with mobile AMD Athlon(tm) 4 processor was used to conduct the simulations.

Programming: Essential Lahey FORTRAN 90 was used to write programs to produce random samples from the distributions of interest, calculate the alternative measures of dispersion, calculate trimmed means for the samples, and calculate standard deviations using the various recoding methods. To ensure that there was no sampling bias based on the seed number that is selected, a different seed number was used for each trial.

Functions in Microsoft Excel 2002 (Microsoft Corporation) were used to calculate the bracketed intervals around the trimmed means, the coverage rates for the resulting bracketed intervals, the widths of the bracketed intervals, and the Location Relative Efficiency (LRE).

Distributions

A total of 10 distributions were considered in this study.

1. Distributions that have an underlying Gaussian structure but have outliers
    a. One wild score on the left of the distribution
    b. Two wild scores on the left of the distribution
    c. Three wild scores on the left of the distribution
    d. Three wild scores on the left of the distribution and one wild score on the right of the distribution

The models with outliers were created by first generating a sample from the standard normal distribution N(0,1) using the NORMB1 subroutine from the RANGEN module (Blair, 1987). Next, to create a sample with 1 wild score on the left, 3.5 was subtracted from the lowest value. Since the normal distribution has a mean of zero and standard deviation of 1, subtracting or adding for example 3.5 is essentially moving a value 3.5 additional standard deviations from the mean. For a sample with 2 wild scores on the left, 3.0 was subtracted from the second lowest value, and for a sample with 3 wild scores on the left, 2.5 was subtracted from the third lowest value. For the sample with 1 wild score on the right, 3.5 was added to the largest value.

The following distributions come from applied social and behavioral science datasets.

2. Smooth Symmetric Distribution (Micceri, 1989) with outliers: A real symmetric dataset of achievement test scores with outliers added.
    a. One wild score on the left of the distribution
    b. Two wild scores on the left of the distribution
    c. Three wild scores on the left of the distribution
    d. Three wild scores on the left of the distribution and one wild score on the right of the distribution

The models with outliers were created by first generating a sample from the smooth symmetric distribution. Next, to create a sample with 1 wild score on the left, 3.5 times the standard deviation (4.906) was subtracted from the lowest value. Subtracting or adding for example 3.5 times the standard deviation is moving a value 3.5 additional standard deviations from the mean.

For a sample with 2 wild scores on the left, 3.0 times the standard deviation was subtracted from the second lowest value, and for a sample with 3 wild scores on the left, 2.5 times the standard deviation was subtracted from the third lowest value. For the sample with 1 wild score on the right, 3.5 times the standard deviation was added to the largest value.

3. Digit Preference (Micceri, 1989): A real symmetric dataset of achievement test scores
4. Discrete Mass at Zero Without Gap (Micceri, 1989): A real symmetric dataset of achievement test scores

Descriptive statistics for the applied social and behavioral science datasets are provided in Table 11. Histograms for these datasets are shown in Figures 2-4.

Table 11. Descriptive Statistics for Three Real Data Sets from Micceri (1989)

Distribution	N	Median	μ	σ	γ_1	γ_2
Achievement Smooth Symmetric	5,375	13	13.19	4.91	0.01	-0.34
Achievement Digit Preference	3,063	535	536.95	37.64	-0.07	-0.24
Achievement Discrete Mass At Zero	2,429	13	12.92	4.42	-0.03	0.31

Notes: μ=population mean, σ=population standard deviation, γ_1=skew, γ_2=kurtosis. Micceri (1989) assumed that these samples have large enough sample sizes to represent the population.

Figure 2. Histogram of Smooth Symmetric Data Set (Micceri, 1989)

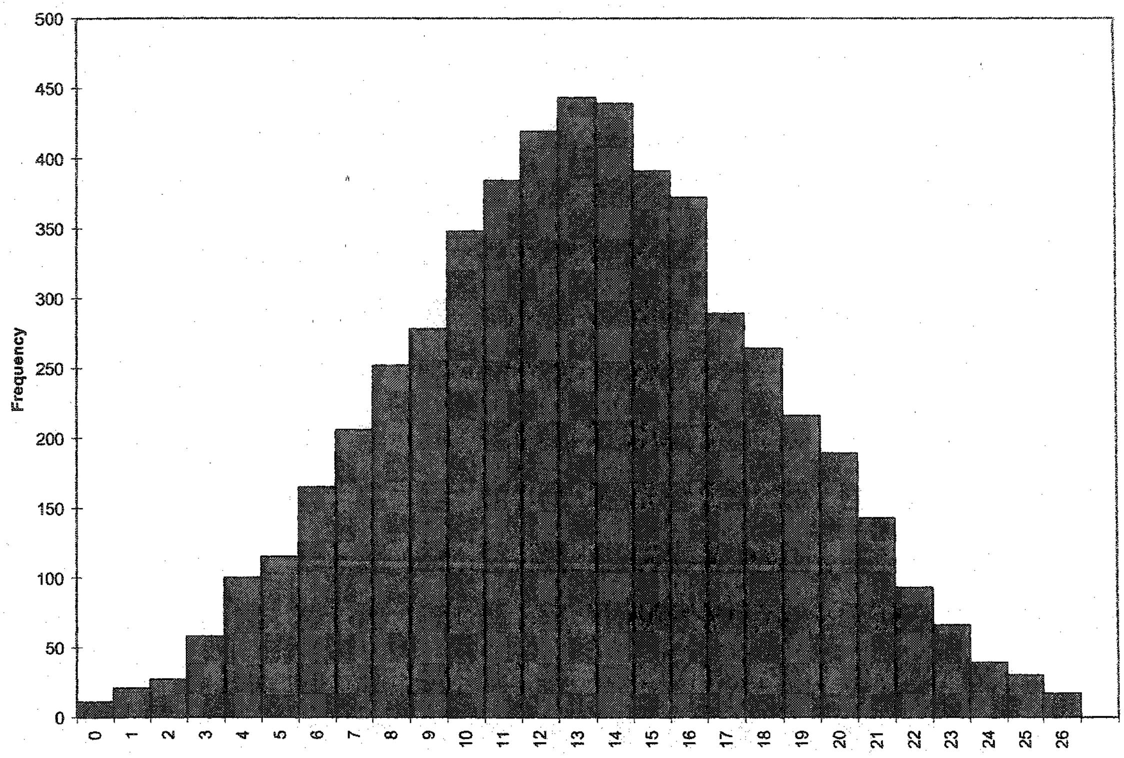


Figure 3. Histogram of Digit Preference Data Set (Micceri, 1989)

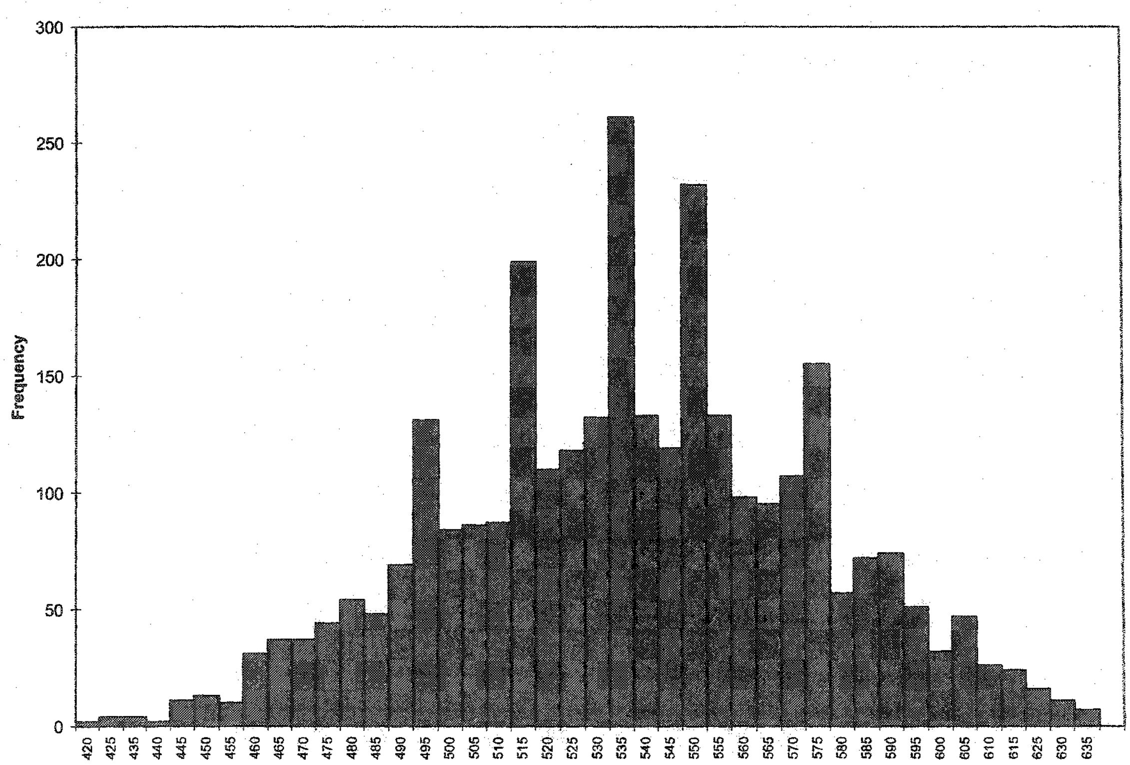
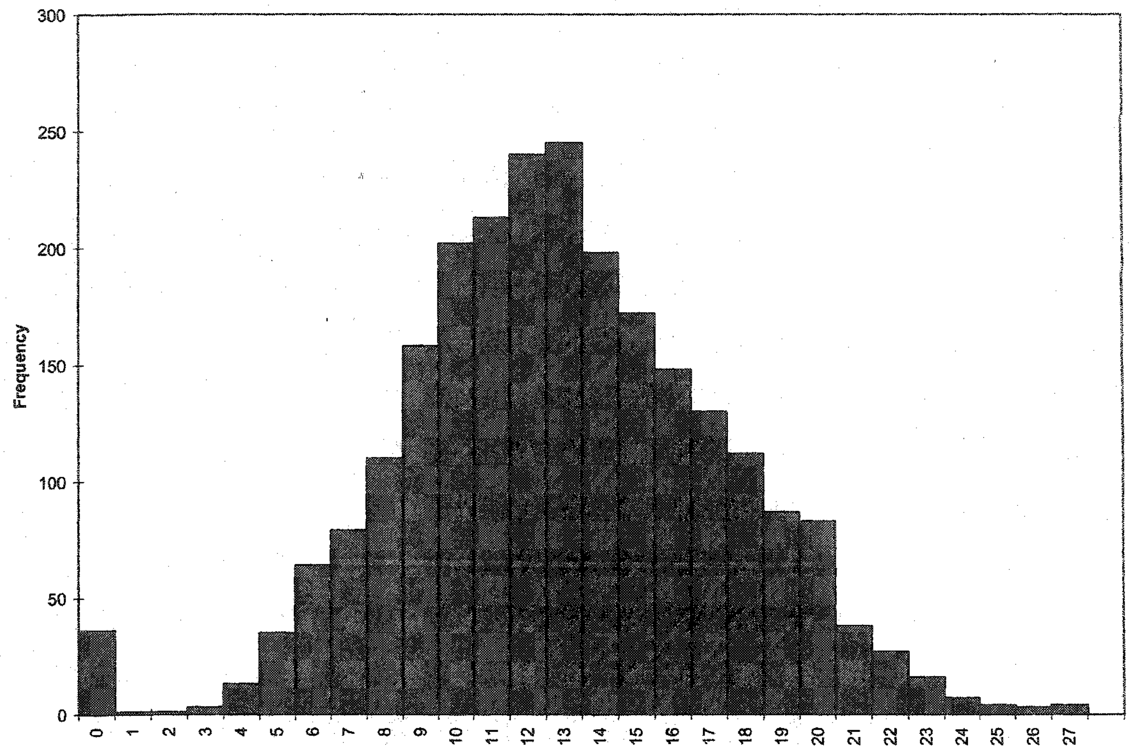


Figure 4. Histogram of Discrete Mass at Zero Data Set (Micceri, 1989)



Sample Size

A sample size of 30 was used in this study.

Alpha Levels

Alpha levels of .05 and .01 were used in this study

Trimming/Winsorizing/Recoding

The following trimming and recoding percentages were used in this study:

1. 2 x 10%
2. 2 x 20%

The study was limited to symmetric trimming/winsorizing/recoding only. An applied researcher will likely not have an a priori knowledge of the symmetry/asymmetry of the population, and will not know what side to trim.

Design

The coverage rates for the bracketed interval around the trimmed mean were evaluated for the six alternative measures of dispersion plus the winsorized sample standard deviation, and the trimmed sample standard deviation. Ten distributions were used, which include four models of outliers with an underlying Gaussian structure, four models of outliers from a real smooth symmetric dataset, and two symmetric applied datasets. A sample size of 30 was used in this study. Each experiment was repeated 10,000 times using alpha levels of .05 and .01. For each experiment, a sample was drawn, the measure of dispersion was calculated using a 2 x 10% and 2 x 20% recode, and the bracketed interval around the sample trimmed mean was calculated.

The interval was checked to see if the population parameter (population mean) was found within the interval. The number of times that the population parameter was found within the interval, divided by the number of times the experiment was performed, gives the coverage. The resultant coverage rate was be evaluated against Bradley's (1978) liberal and conservative criteria for Type I error rates as applied to coverage to determine if the coverage rates were inflated (the percentage of bracketed intervals in which the population parameter is found within the interval exceeds the 1-nominal alpha level), or underestimated (the percentage of bracketed intervals in which the population parameter is found within the interval is less than the 1-nominal alpha level). Bradley's criteria can be extended in the following way: (1 -1.5α) ≤ Coverage (1- α) ≤ (1 -.5α) (Liberal criterion) ;(1- 1.1α) ≤ Coverage (1- α) ≤ (1-.9α) (Conservative criterion).

The widths of the bracketed intervals were calculated by subtracting the lower value of the bracketed interval from the upper value of the bracketed interval. The Location Relative Efficiency (LRE) (Sawilowsky, 2002) was calculated for each interval. In this study the LRE was defined as follows:

$$LRE = \frac{Sw(U_{B.I.}) - Sw(L_{B.I})}{X(U_{B.I.}) - X(L_{B.I})}$$

The LRE is the ratio of the range of the bracketed interval for the sample winsorized standard deviation, and the bracketed interval for the competitor. In this case all competitors are compared to sample winsorized standard deviation because it is the measure of dispersion that is supported in calculating the bracketed interval around the trimmed mean. A LRE greater then one indicates that the competitor yields narrower bracketed intervals than the sample winsorized standard deviation.

Sample trimmed means and standard deviations for the various recoding methods were generated from the Essential Lahey FORTRAN 90 program. Microsoft Excel 2002 (Microsoft Corporation) functions were used to calculate the bracketed intervals around the trimmed means, the coverage rates for the resulting bracketed intervals, the widths of the bracketed intervals, and the Location Relative Efficiency (LRE).

CHAPTER 4

RESULTS

This chapter summarizes the coverage and Location Relative Efficiency (LRE) for the eight measures of dispersion that were tested in this study. For ease of comparison of the eight measures of dispersion, results are reported by alpha level and trim/recode percent. Results are first summarized for coverage, and then summarized by Location Relative Efficiency (LRE) for coverage rates that meet Bradley's liberal and conservative criteria for robustness.

Table 12. 95% Bracketed Interval Coverage For Various Robust Measures of Dispersion; 2x10% Trim

Sample	s_w	s_t	$s_{recode1}$	$s_{recode2}$	$s_{recode3}$	$s_{recode4}$	$s_{recode5}$	$s_{recode6}$
Gaussion 1WL	94.75%**	89.91%	95.20%**	96.75%*	98.97%	96.61%*	98.80%	95.95%*
Gaussian 2WL	95.33%**	90.59%	95.59%*	99.66%	99.48%	98.43%	98.72%	98.92%
Gaussian 3WL	94.68%**	90.18%	99.82%	99.92%	99.71%	99.94%	99.63%	99.60%
Gaussian 3WL1R	94.89%**	90.02%	99.85%	99.91%	99.91%	99.92%	99.87%	99.73%
Smooth Symmetric 1WL	94.63%**	89.68%	95.03%**	96.52%*	98.56%	96.31%*	98.39%	95.55%*
Smooth Symmetric 2WL	94.75%**	89.84%	95.09%**	99.57%	99.36%	98.33%	99.30%	98.03%
Smooth Symmetric 3WL	94.85%**	90.01%	99.77%	99.89%	99.61%	99.90%	99.57%	99.61%
Smooth Symmetric 3WL1WR	94.72%**	90.12%	99.74%	99.86%	99.80%	99.87%	99.77%	99.48%
Digit Preference	94.70%**	90.00%	95.16%**	96.54%*	96.68%*	96.41%*	96.23%*	95.70%*
Discrete Mass at Zero w/o Gap	95.05%**	90.11%	95.60%*	97.11%*	97.39%*	96.91%*	97.00%*	96.28%*

* Meets Bradley's Liberal Criterion

** Meets Bradley's Conservative Criterion

Figure 5. Number of Distributions Where Bradley's Liberal or Conservative Criteria for Robustness Was Met. 95% Bracketed Interval; 2x10% Trim

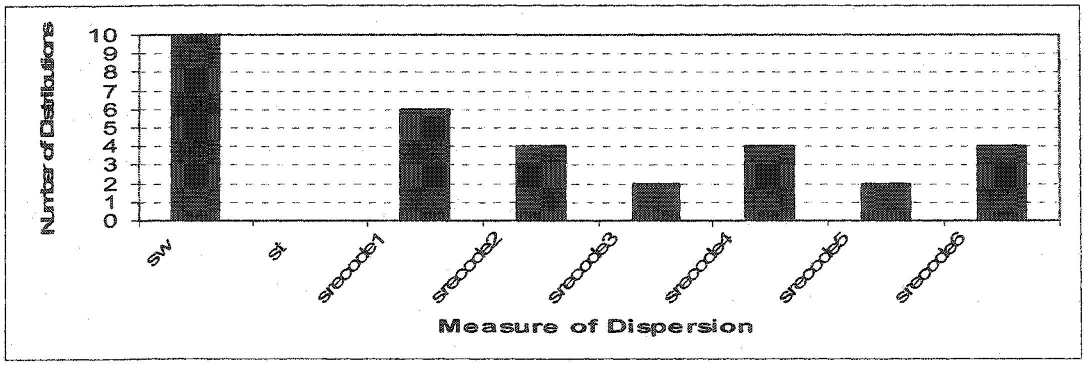


Table 13. 95% Bracketed Interval Coverage For Various Robust Measures of Dispersion; 2x20% Trim

Sample	S_w	S_t	$S_{recode1}$	$S_{recode2}$	$S_{recode3}$	$S_{recode4}$	$S_{recode5}$	$S_{recode6}$
Gaussion 1WL	94.33%*	87.03%	92.41%	99.18%	99.69%	98.13%	99.62%	96.45%*
Gaussian 2WL	94.53%**	86.65%	92.61%*	99.96%	99.87%	97.79%	99.78%	96.47%*
Gaussian 3WL	94.78%**	87.16%	92.69%*	100.00%	99.95%	98.03%	99.89%	96.87%*
Gaussian 3WL1R	94.06%*	86.23%	92.10%	99.99%	99.97%	97.79%	99.96%	96.10%*
Smooth Symmetric 1WL	94.69%**	86.88%	92.89%*	99.33%	99.72%	98.06%	99.61%	96.66%*
Smooth Symmetric 2WL	94.79%**	87.42%	92.87%*	99.98%	99.93%	98.29%	99.88%	96.90%*
Smooth Symmetric 3WL	94.57%**	86.92%	92.76%*	100.00%	99.94%	98.21%	99.93%	96.89%*
Smooth Symmetric 3WL1WR	94.47%*	86.99%	92.57%*	100.00%	99.99%	97.79%	99.88%	96.38%*
Digit Preference	94.77%**	86.49%	92.58%*	99.14%	98.55%	97.96%	97.84%	96.67%*
Discrete Mass at Zero w/o Gap	94.06%*	86.67%	91.93%	99.20%	98.76%	97.96%	98.05%	96.54%*

* Meets Bradley's Liberal Criterion
** Meets Bradley's Conservative Criterion

Figure 6. Number of Distributions Meeting Bradley's Liberal or Conservative Criteria for Robustness. 95% Bracketed Interval; 2x20% Trim

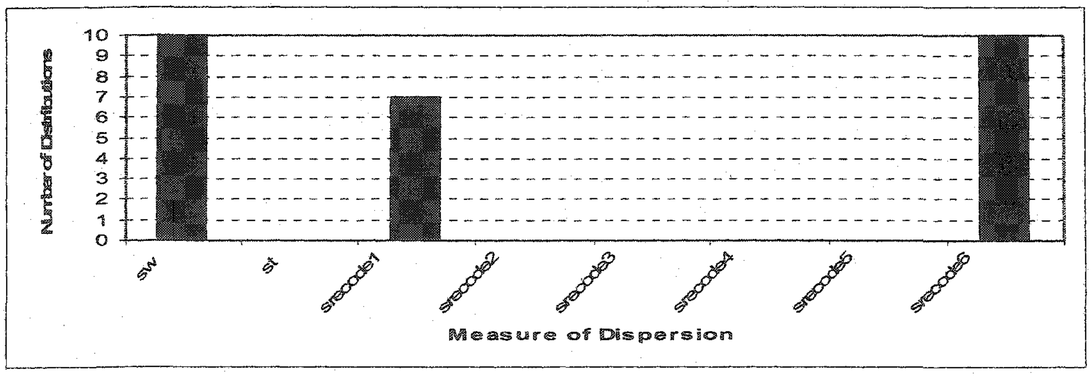


Figure 7. 95% Coverage Rates for s_w ; 2x10% and 2x20% Trim

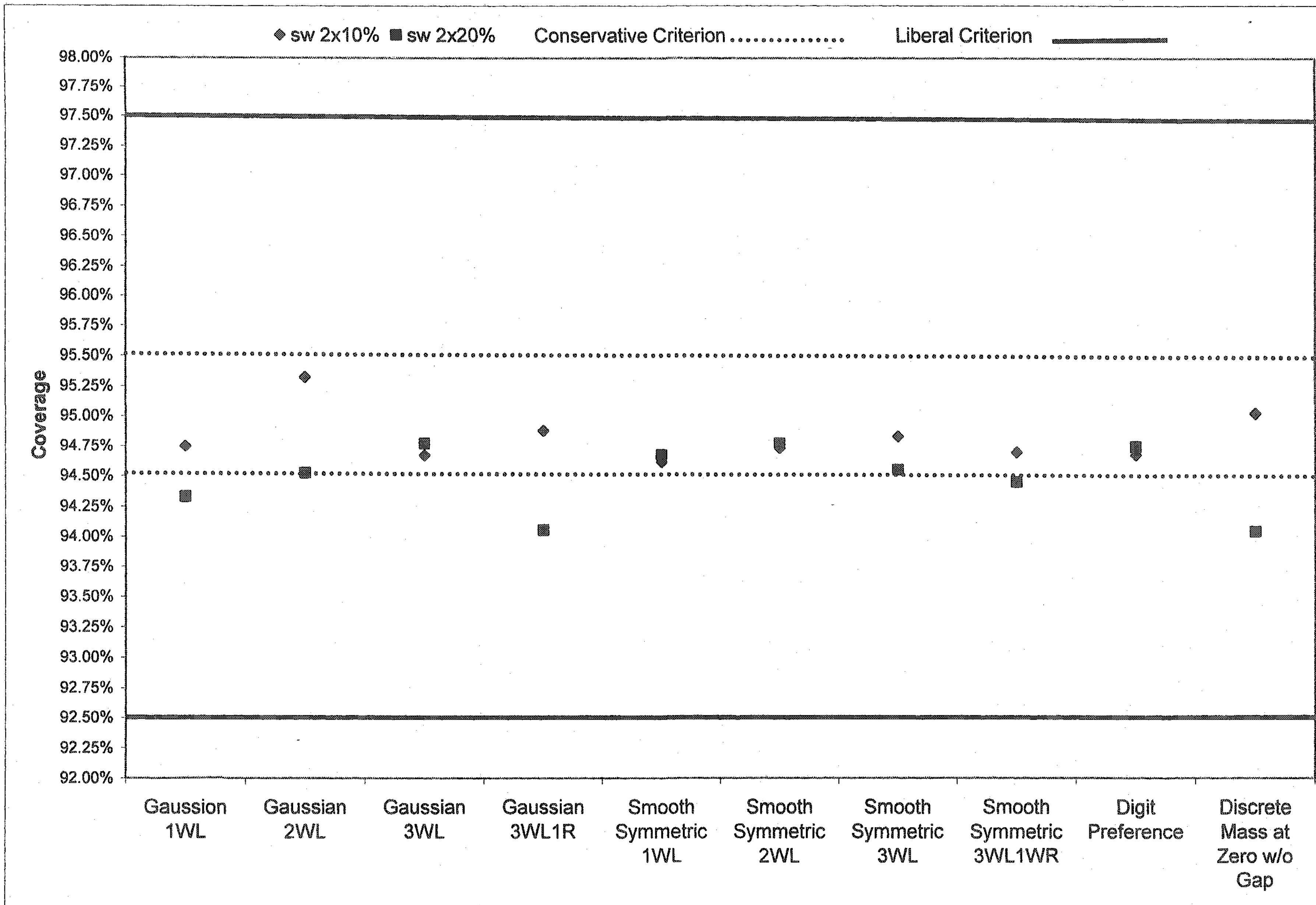


Figure 8. 95% Coverage Rates for s_t; 2x10% and 2x20% Trim

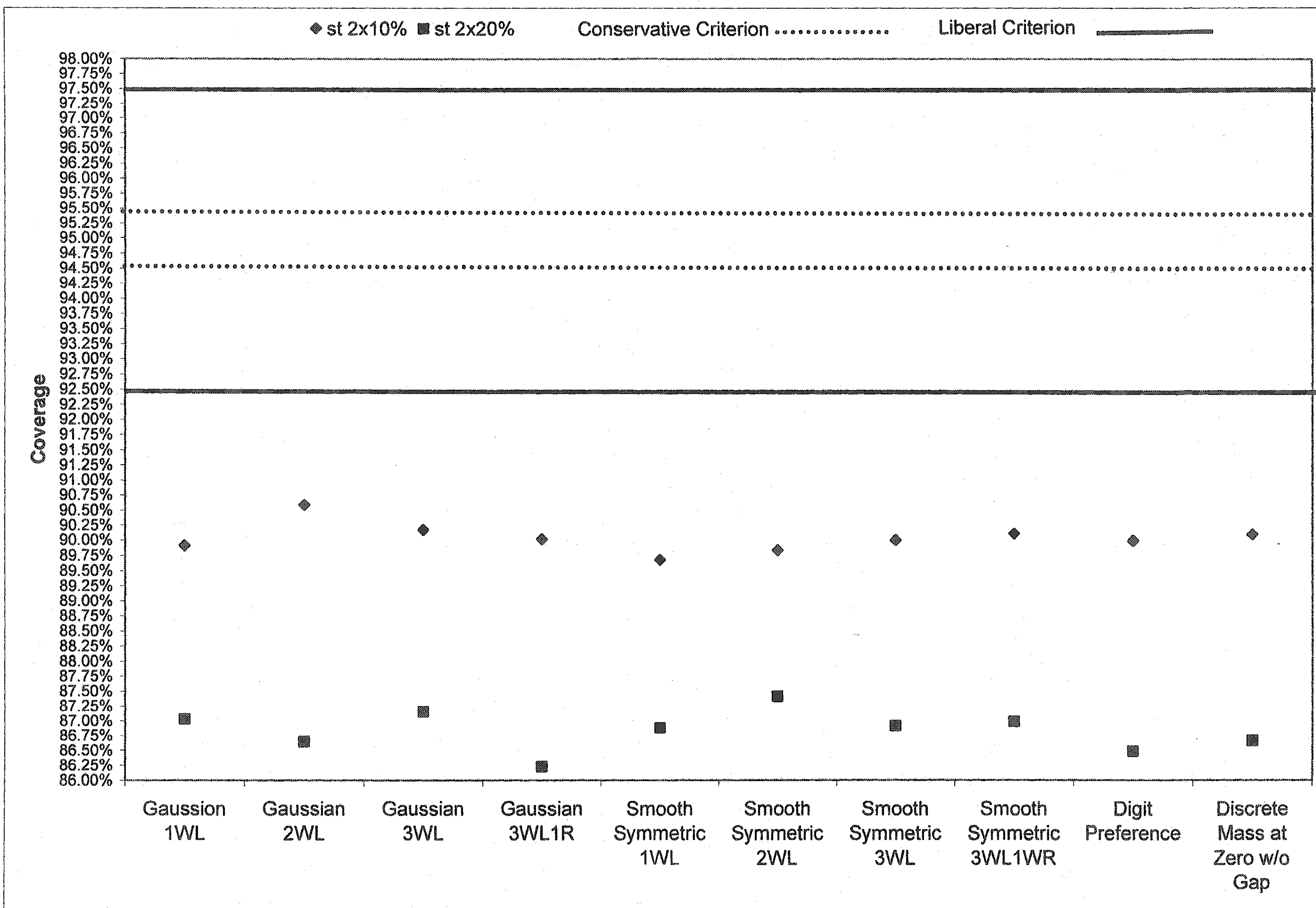


Figure 9. 95% Coverage Rates for $s_{recode1}$; 2x10% and 2x20% Trim

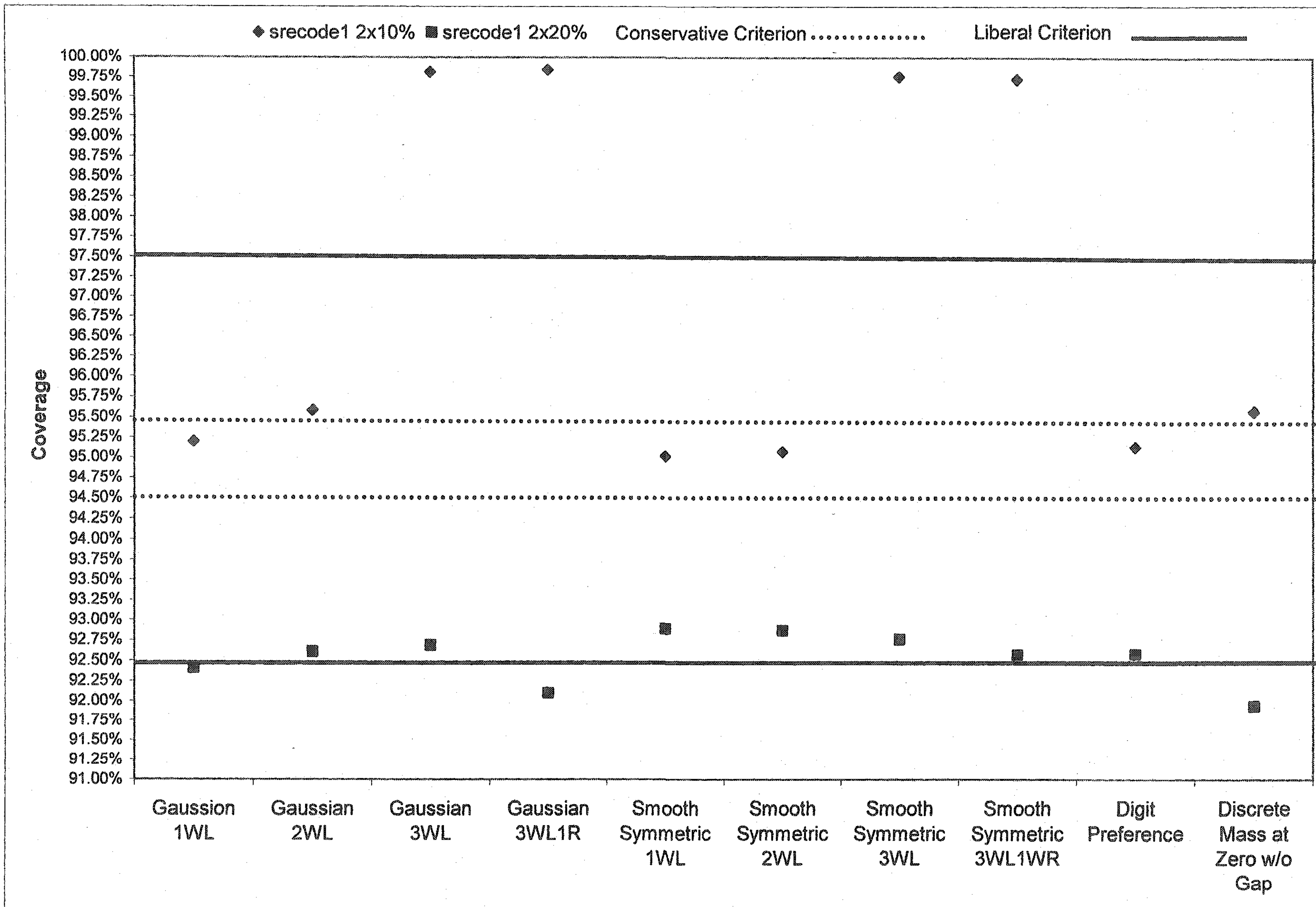


Figure 10. 95% Coverage Rates for $s_{recode2}$; 2x10% and 2x20% Trim

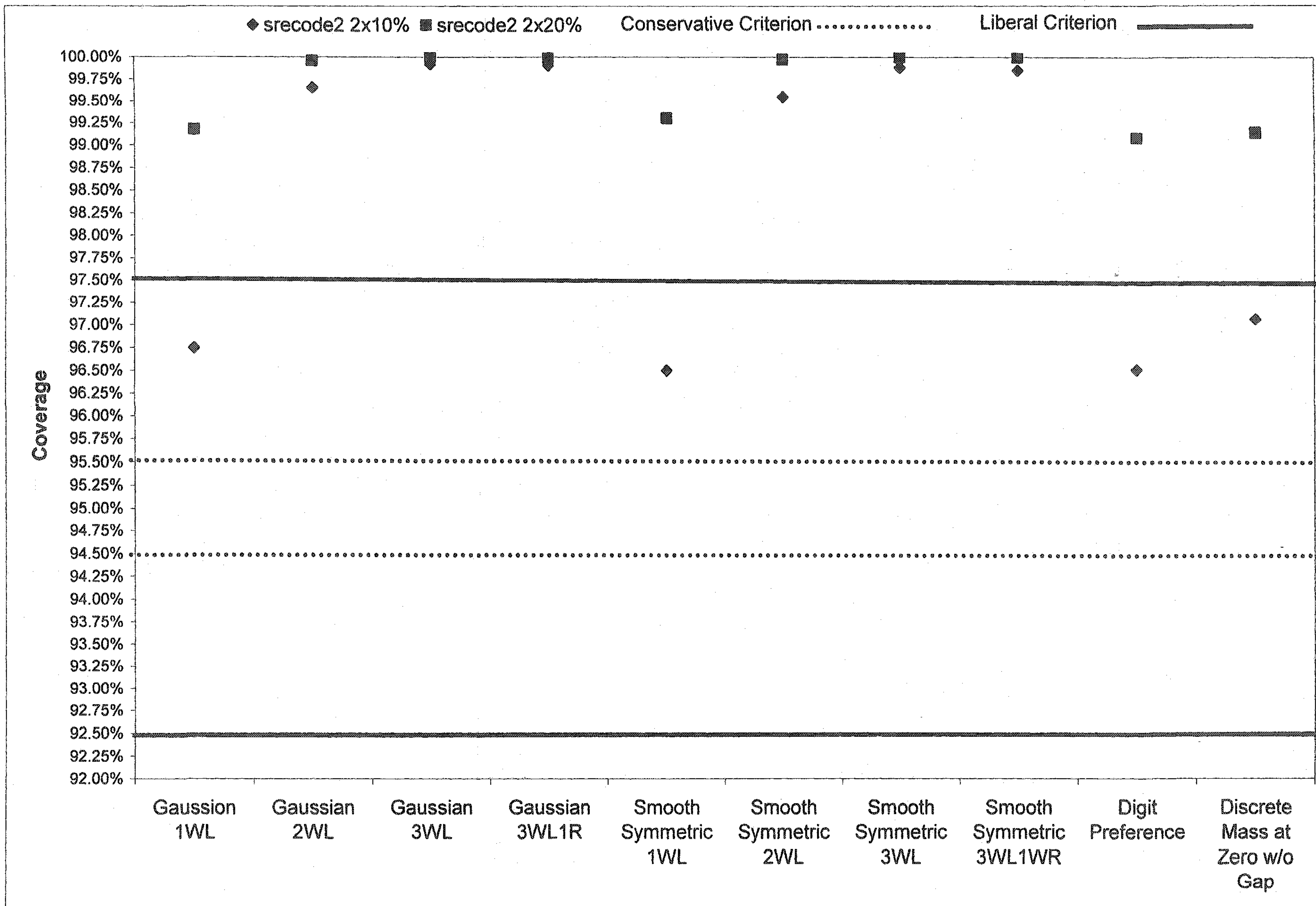


Figure 11. 95% Coverage Rates for $s_{recode3}$; 2x10% and 2x20% Trim

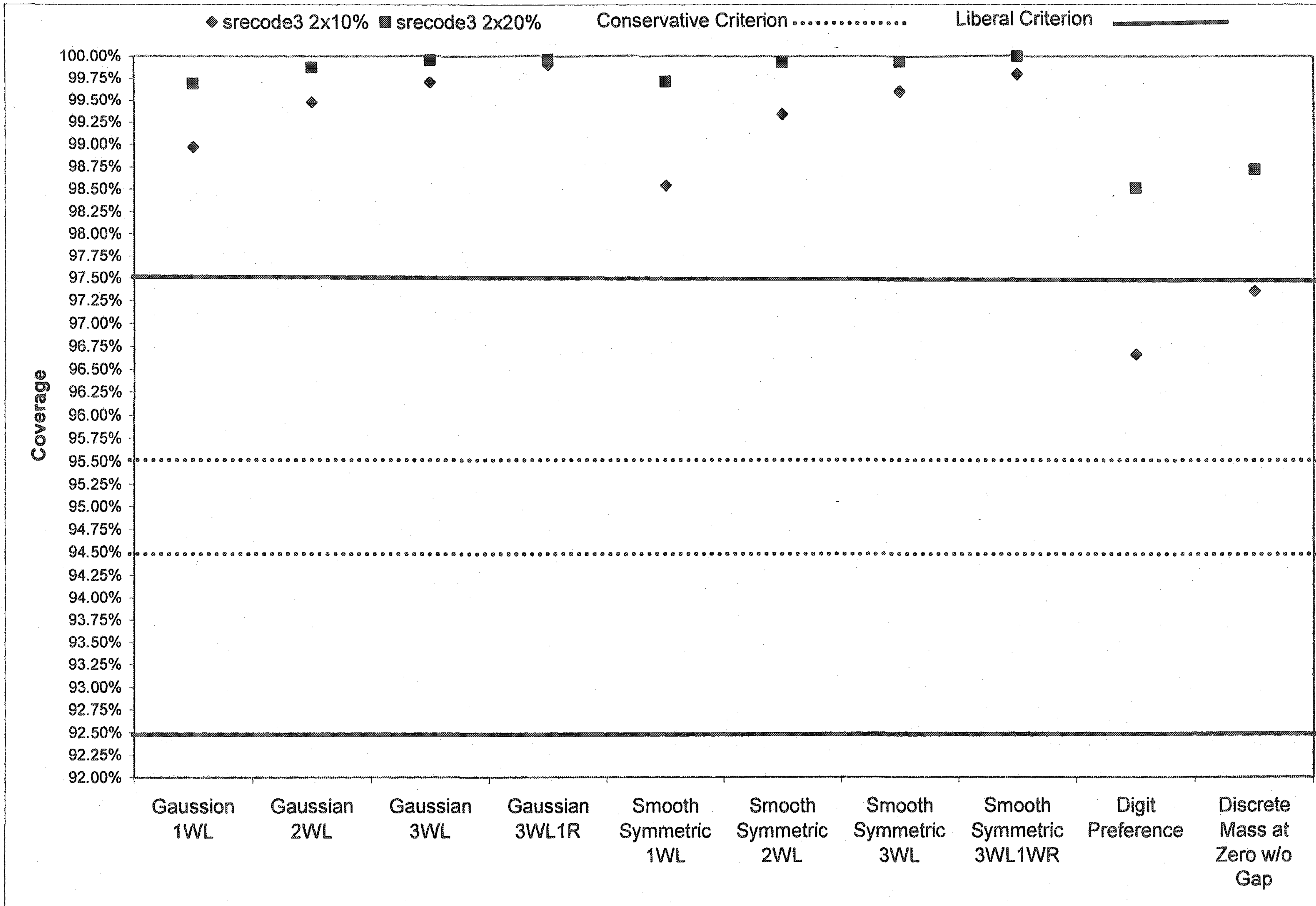


Figure 12. 95% Coverage Rates for $s_{recode4}$; 2x10% and 2x20% Trim

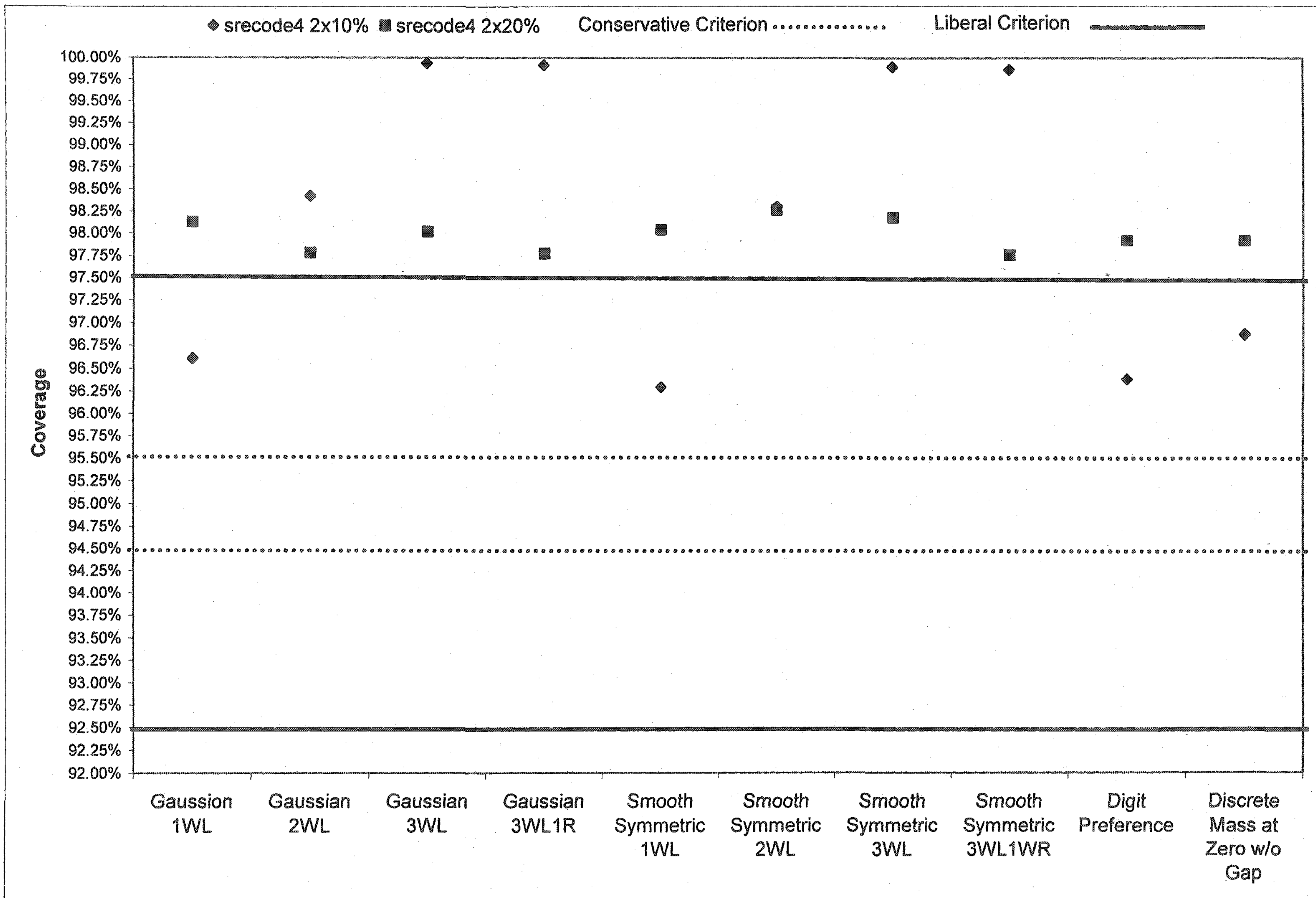


Figure 13. 95% Coverage Rates for $s_{recode5}$; 2x10% and 2x20% Trim

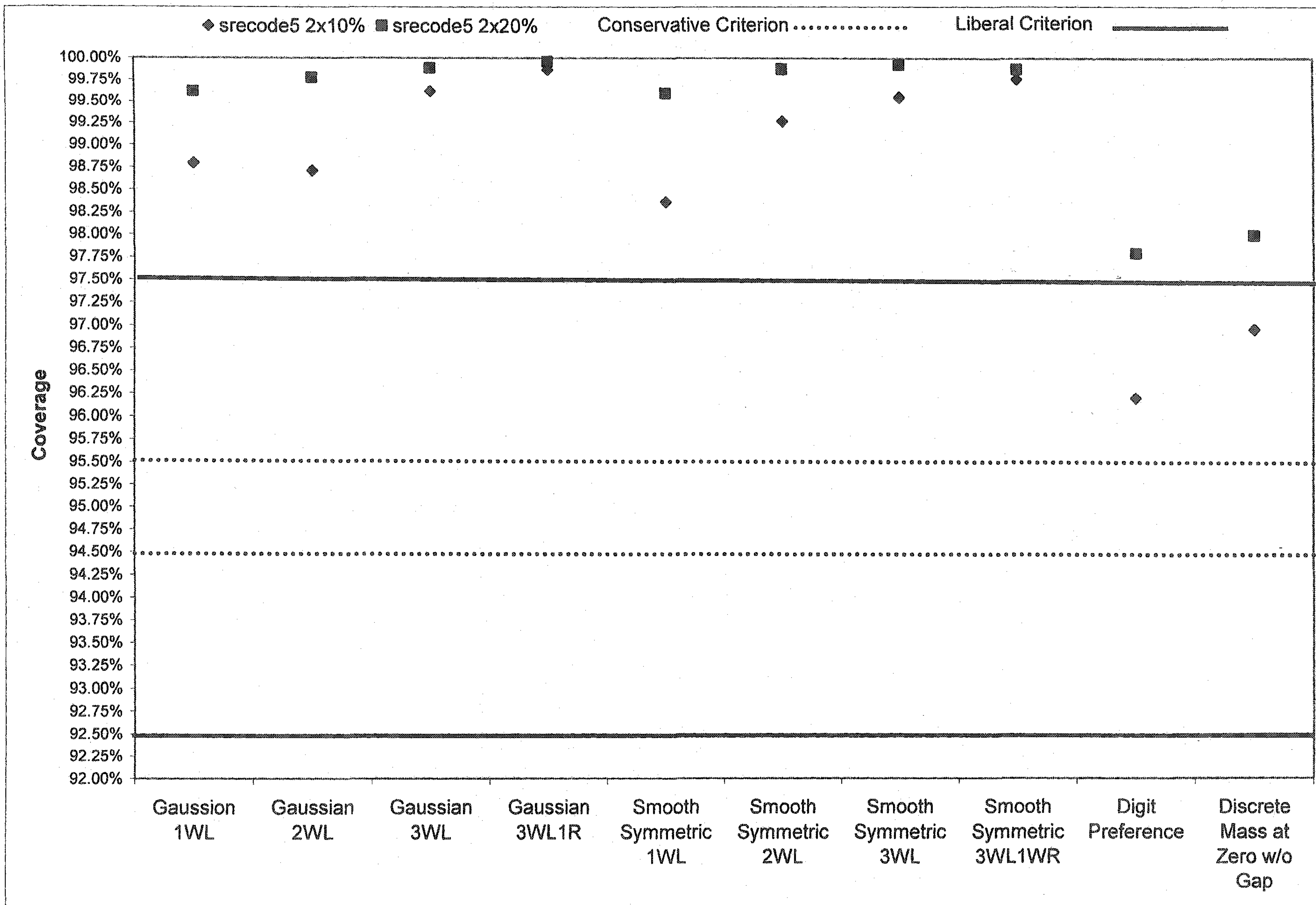


Figure 14. 95% Coverage Rates for $s_{recode6}$; 2x10% and 2x20% Trim

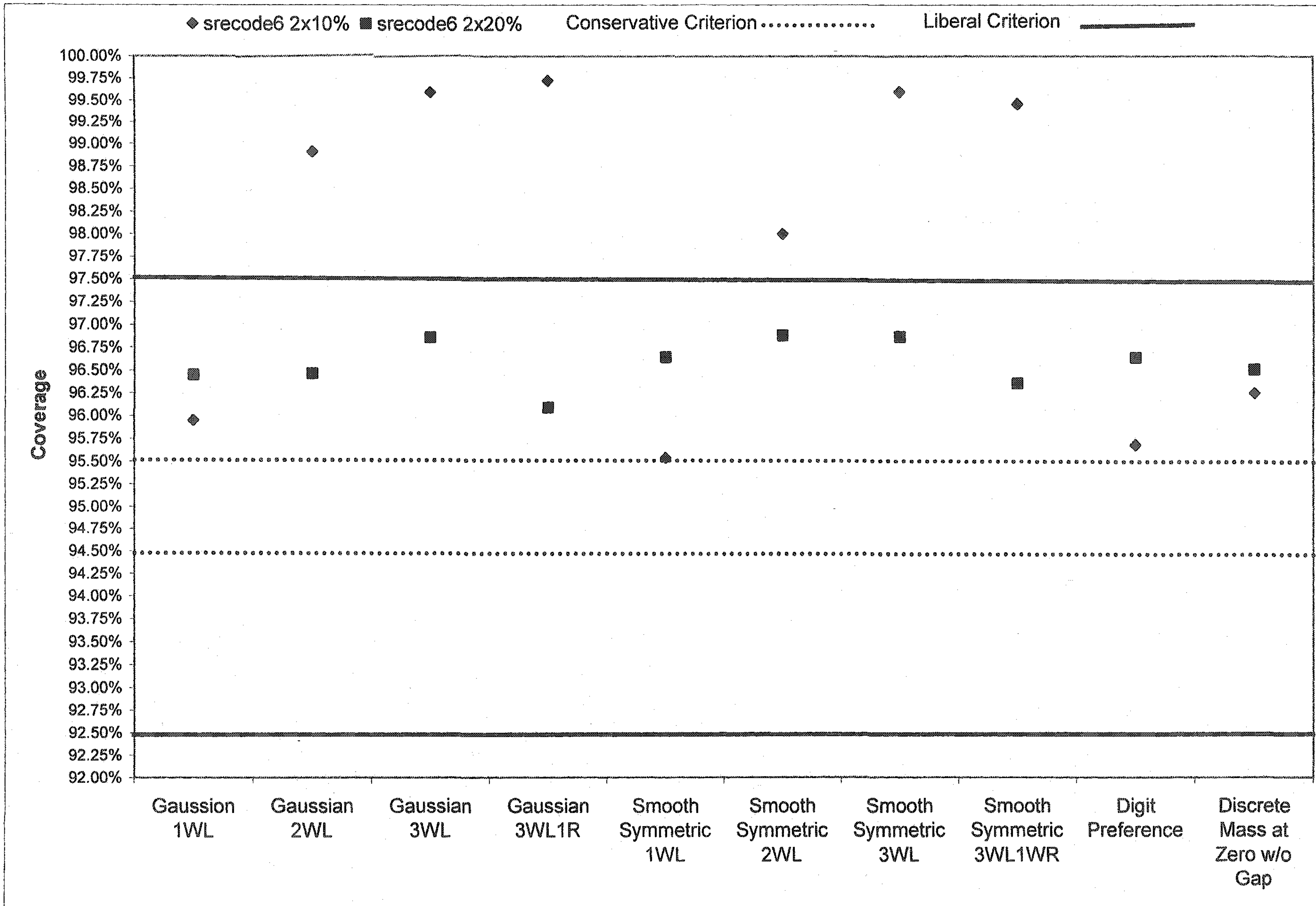


Table 14. 99% Bracketed Interval Coverage For Various Robust Measures of Dispersion; 2x10% Trim

Sample	s_w	s_t	$s_{recode1}$	$s_{recode2}$	$s_{recode3}$	$s_{recode4}$	$s_{recode5}$	$s_{recode6}$
Gaussion 1WL	99.01%**	97.07%	99.16%*	99.57%	99.94%	99.53%	99.91%	99.37%*
Gaussian 2WL	99.03%**	97.24%	99.21%*	99.99%	99.95%	99.82%	99.94%	99.79%
Gaussian 3WL	98.75%*	96.81%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Gaussian 3WL1R	99.01%**	97.29%	100.00%	100.00%	100.00%	100.00%	100.00%	99.97%
Smooth Symmetric 1WL	98.72%*	96.74%	98.94%**	99.41%*	99.92%	99.30%*	99.90%	99.14%*
Smooth Symmetric 2WL	98.98%**	97.09%	99.07%**	100.00%	99.95%	99.79%	99.94%	99.74%
Smooth Symmetric 3WL	98.80%*	96.81%	100.00%	100.00%	100.00%	100.00%	99.99%	99.99%
Smooth Symmetric 3WL1WR	98.77%*	96.95%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Digit Preference	98.90%**	97.06%	99.10%**	99.51%	99.54%	99.50%*	99.47%*	99.31%*
Discrete Mass at Zero w/o Gap	98.96%**	97.28%	99.10%**	99.51%	99.58%	99.44%*	99.50%*	99.27%*

* Meets Bradley's Liberal Criterion
** Meets Bradley's Conservative Criterion

Figure 15. Number of Distributions Meeting Bradley's Liberal or Conservative Criteria for Robustness. 99% Bracketed Interval; 2x10% Trim

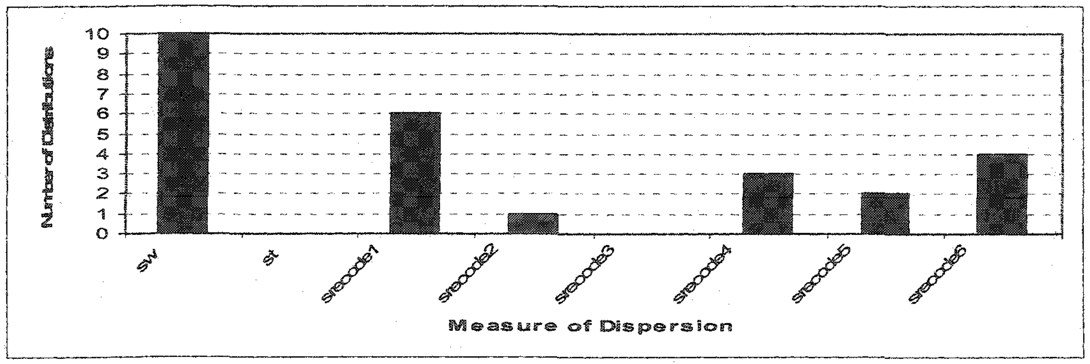


Table 15. 99% Bracketed Interval Coverage For Various Robust Measures of Dispersion; 2x20% Trim

Sample	s_w	s_t	$s_{recode1}$	$s_{recode2}$	$s_{recode3}$	$s_{recode4}$	$s_{recode5}$	$s_{recode6}$
Gaussion 1WL	98.90%**	95.40%	98.08%	99.94%	100.00%	99.71%	100.00%	99.50%*
Gaussian 2WL	98.89%*	95.52%	98.13%	100.00%	100.00%	99.72%	100.00%	99.41%*
Gaussian 3WL	98.82%*	95.73%	98.27%	100.00%	100.00%	99.75%	100.00%	99.58%
Gaussian 3WL1R	98.65%*	95.13%	97.91%	100.00%	100.00%	99.73%	100.00%	99.41%*
Smooth Symmetric 1WL	98.80%*	95.99%	98.29%	99.99%	100.00%	99.81%	99.99%	99.64%
Smooth Symmetric 2WL	99.04%**	95.88%	98.43%	100.00%	100.00%	99.85%	99.99%	99.56%
Smooth Symmetric 3WL	99.04%**	95.97%	98.44%	100.00%	100.00%	99.82%	100.00%	99.62%
Smooth Symmetric 3WL1WR	98.67%*	95.61%	98.04%	100.00%	100.00%	99.76%	100.00%	99.46%*
Digit Preference	98.84%*	95.84%	98.28%	99.95%	99.91%	99.76%	99.73%	99.42%*
Discrete Mass at Zero w/o Gap	98.75%*	95.46%	98.03%	99.94%	99.88%	99.75%	99.75%	99.42%*

* Meets Bradley's Liberal Criterion

** Meets Bradley's Conservative Criterion

Figure 16. Number of Distributions Meeting Bradley's Liberal or Conservative Criteria for Robustness. 99% Bracketed Interval; 2x20% Trim

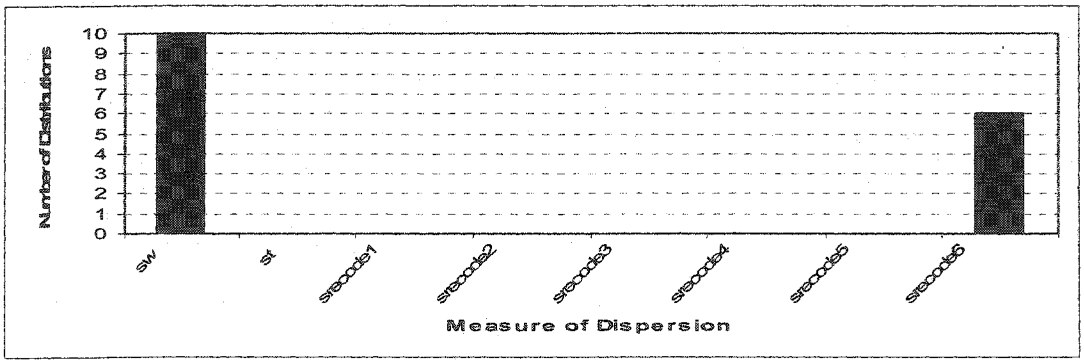


Figure 17. 99% Coverage Rates for s_w; 2x10% and 2x20% Trim

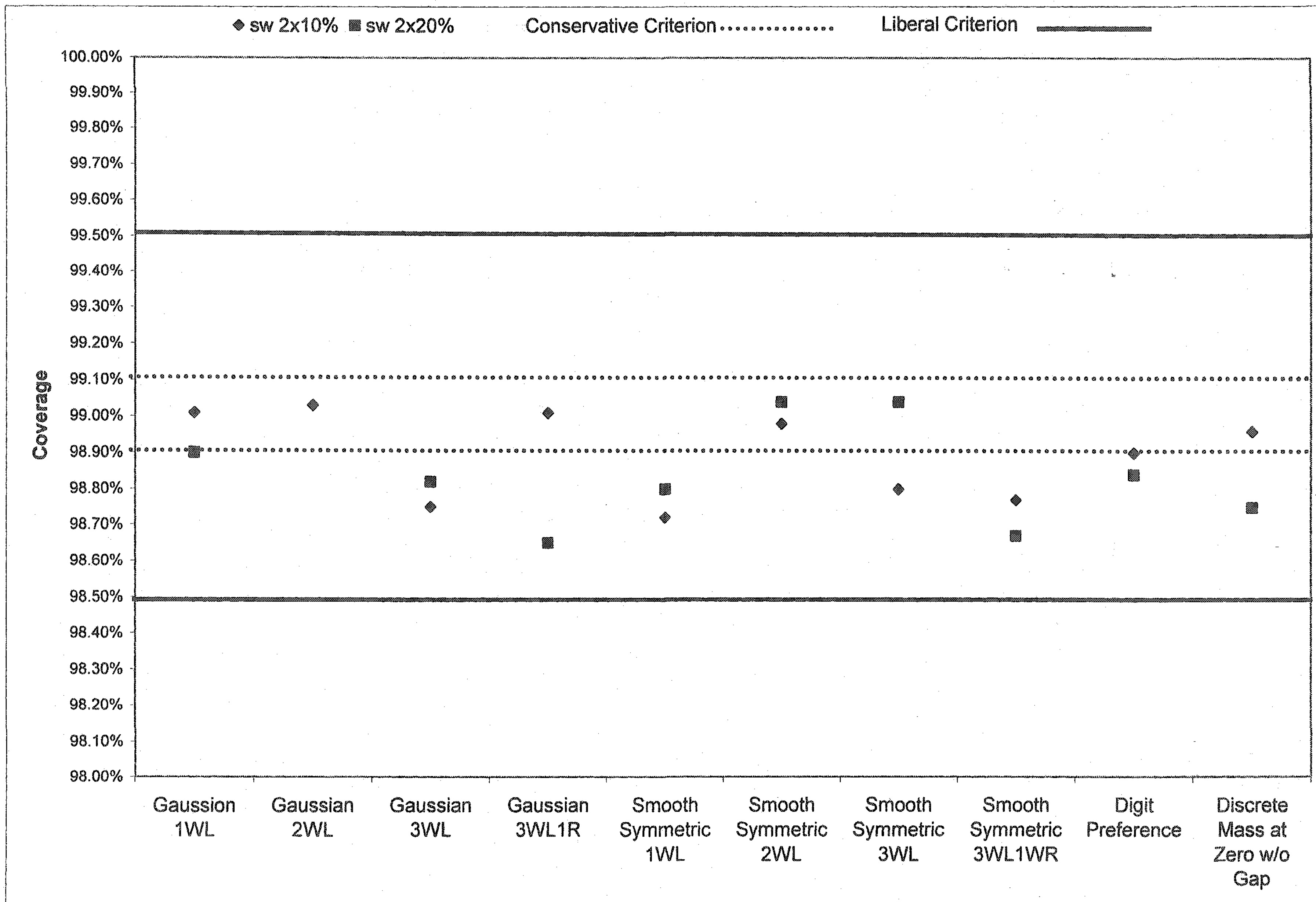


Figure 18. 99% Coverage Rates for s_t; 2x10% and 2x20% Trim

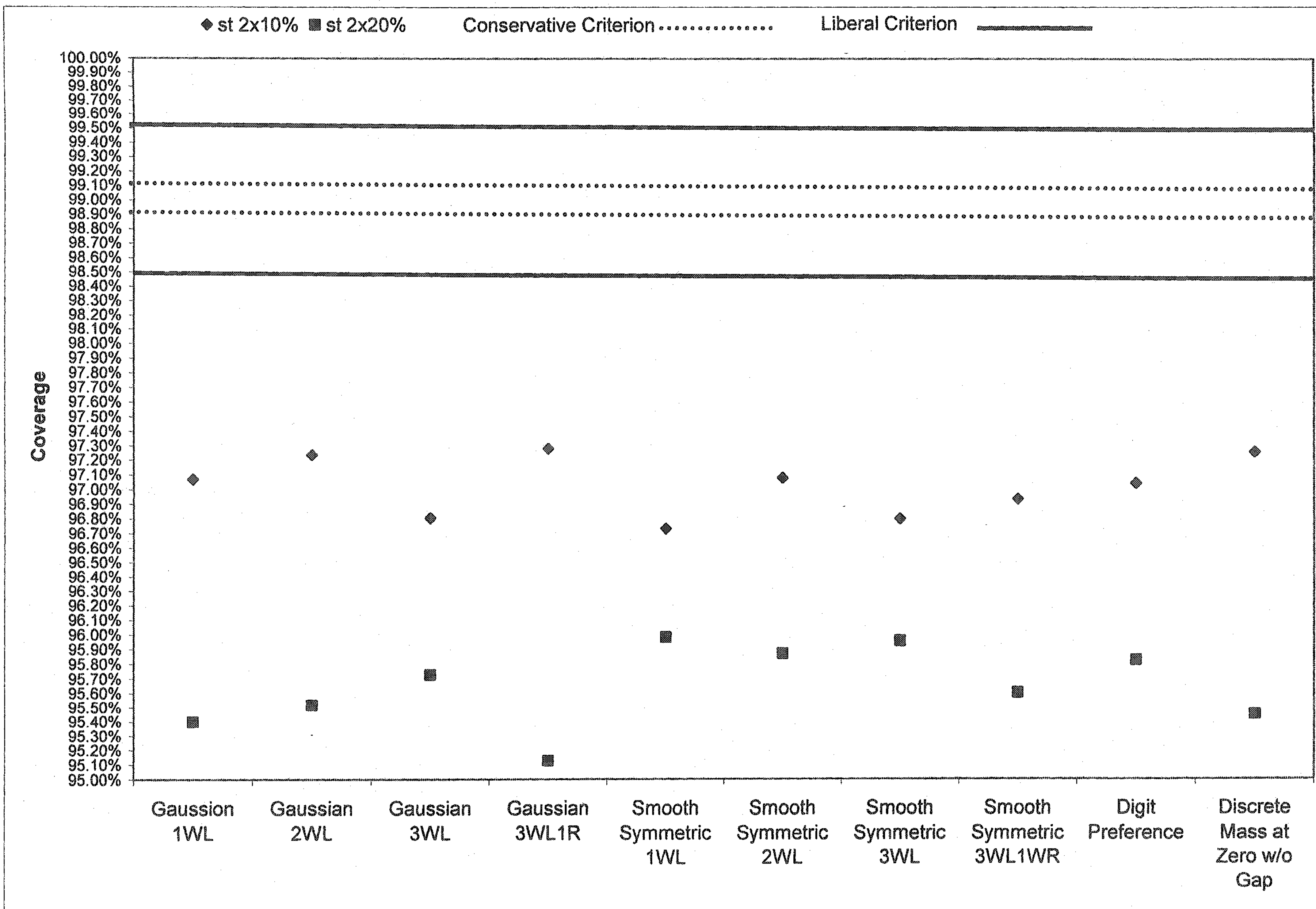


Figure 19. 99% Coverage Rates for $s_{recode1}$; 2x10% and 2x20% Trim

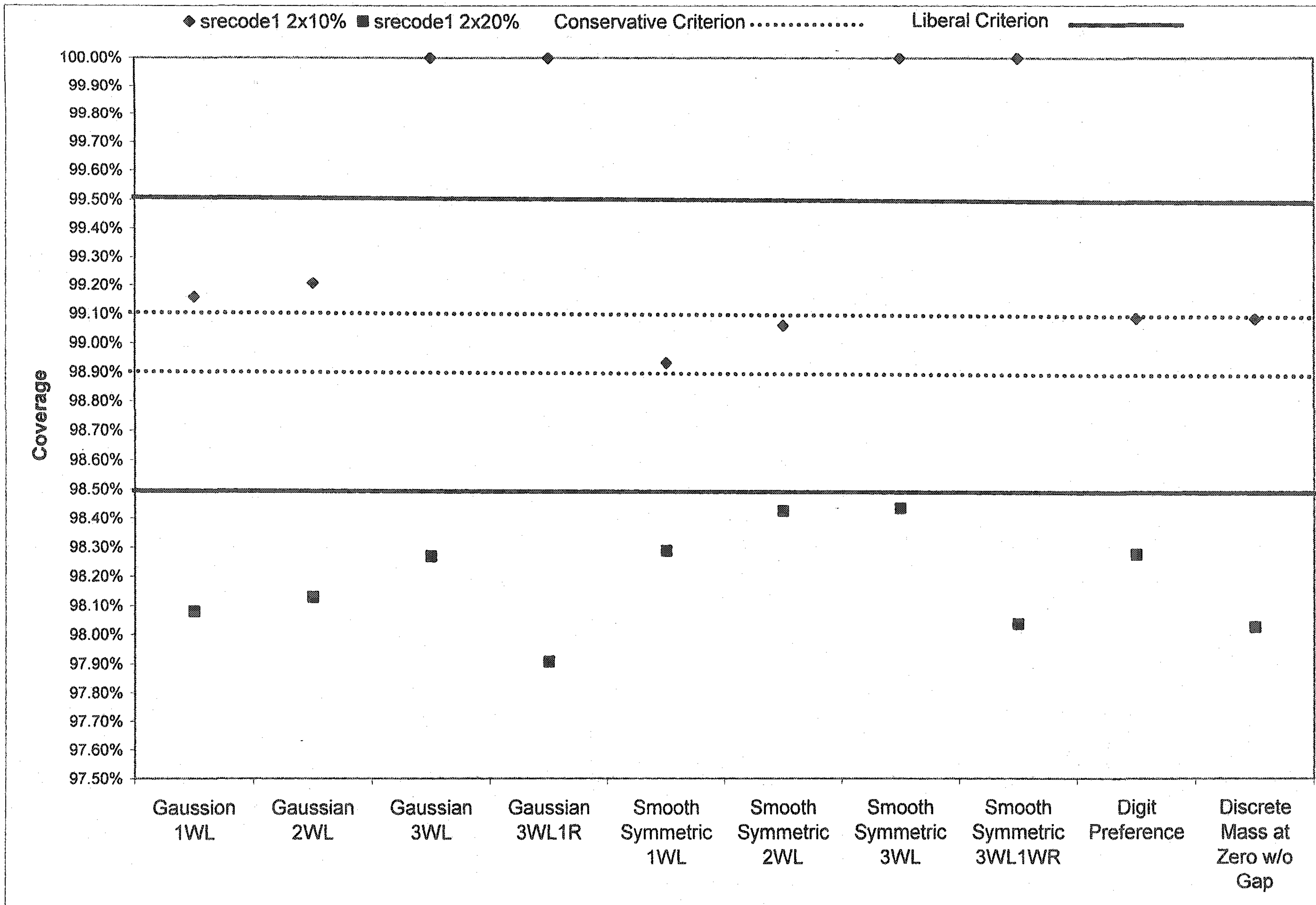


Figure 20. 99% Coverage Rates for $s_{recode2}$; 2x10% and 2x20% Trim

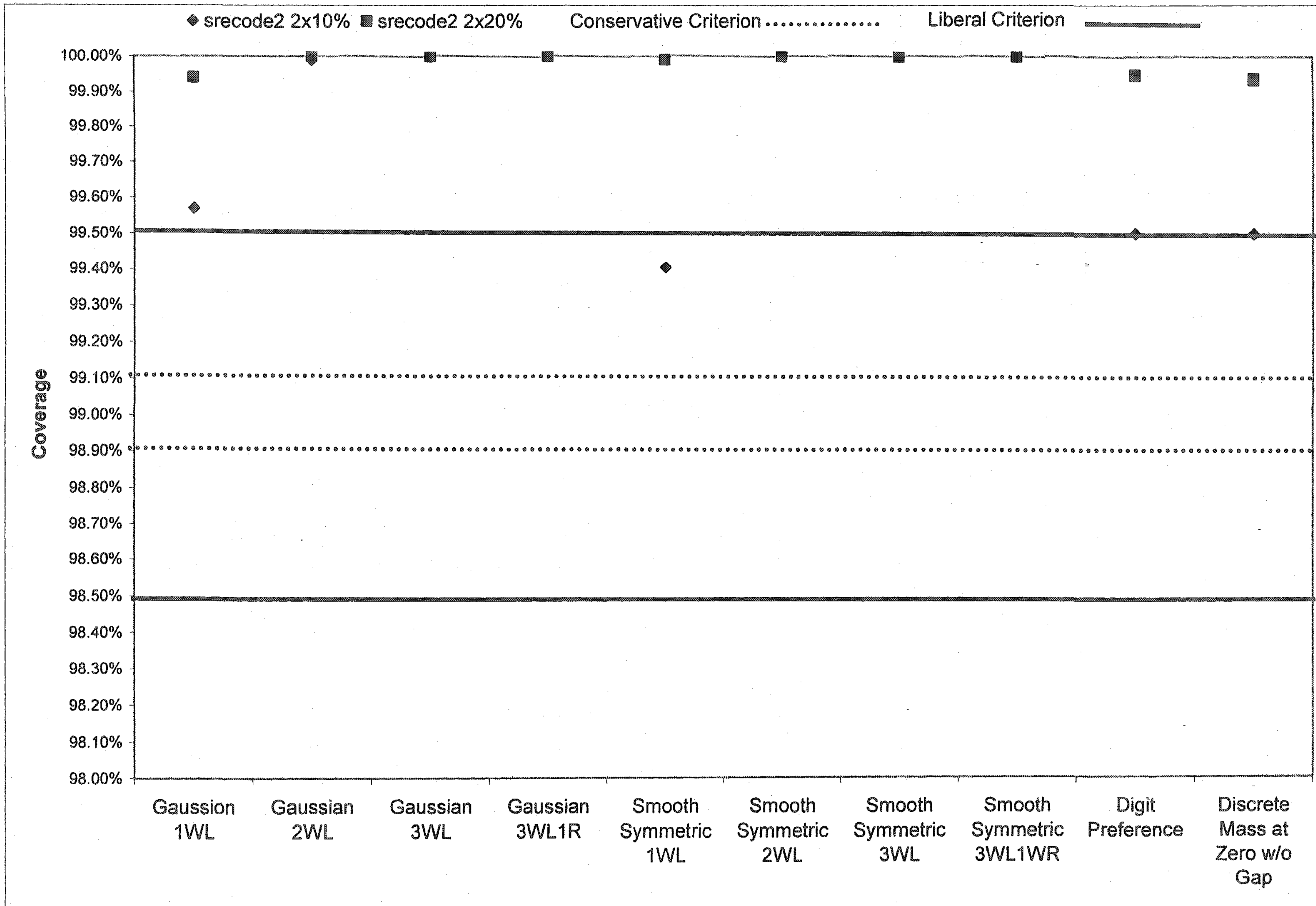


Figure 21. 99% Coverage Rates for $s_{recode3}$; 2x10% and 2x20% Trim

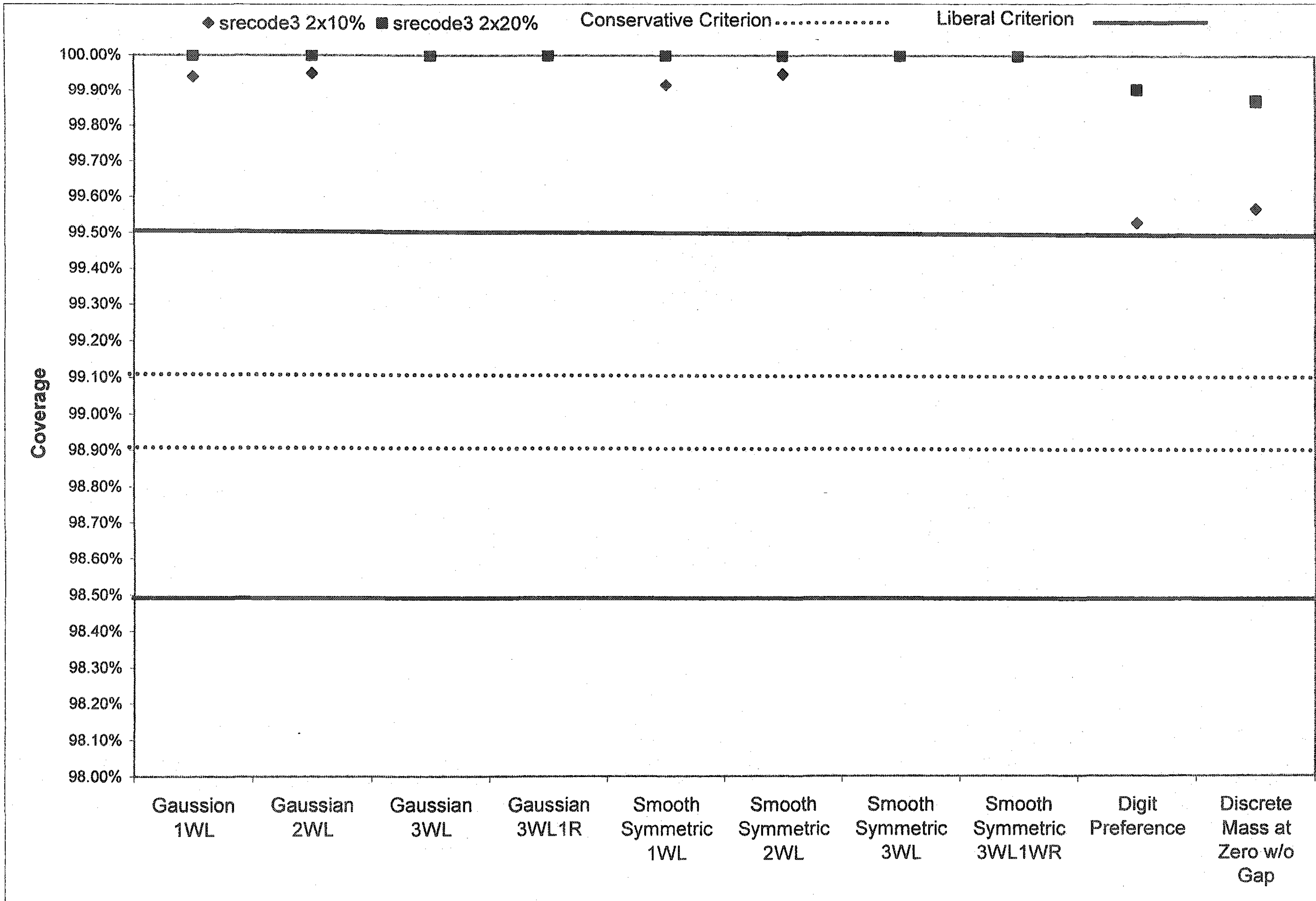


Figure 22. 99% Coverage Rates for $s_{recode4}$; 2x10% and 2x20% Trim

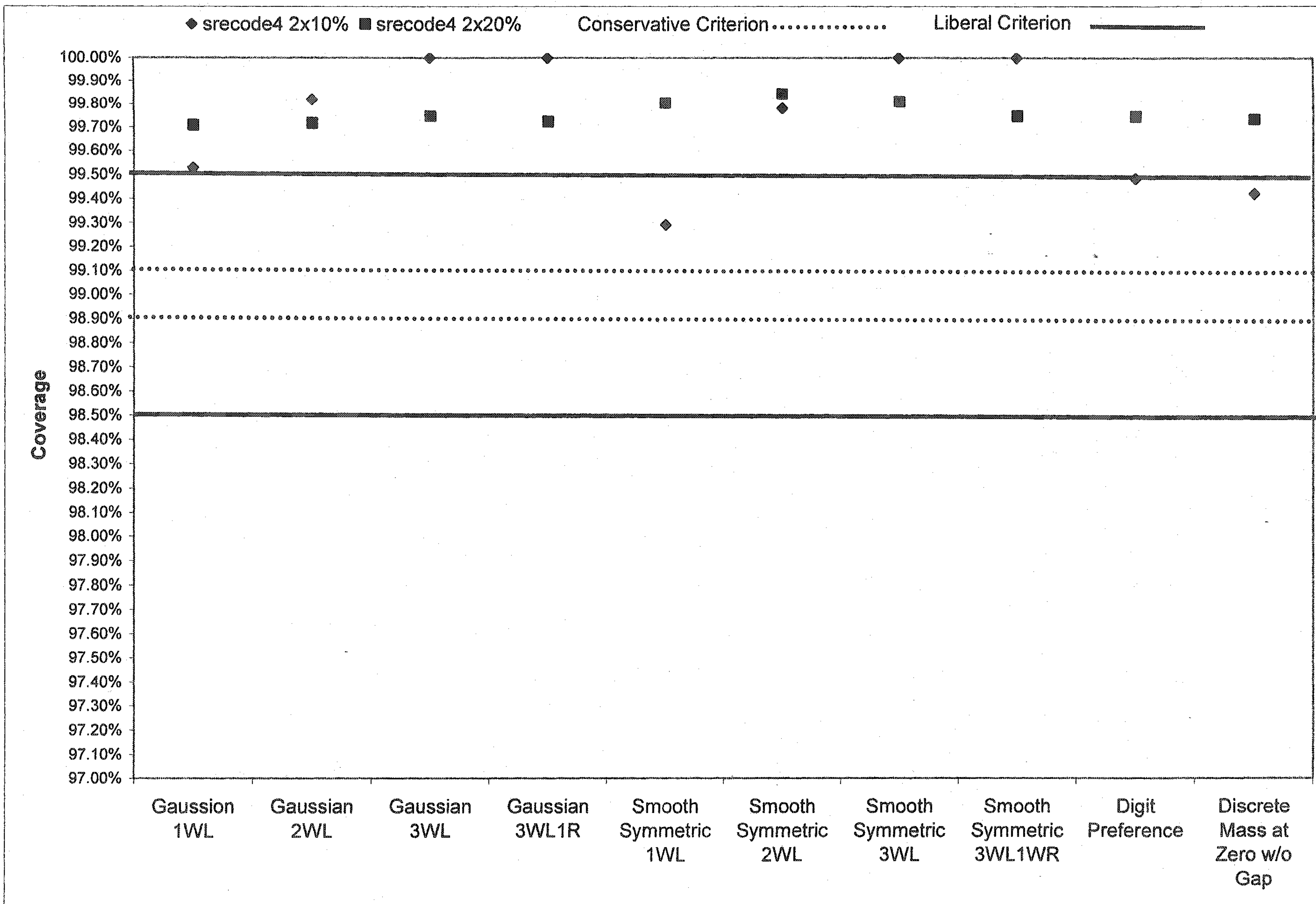


Figure 23. 99% Coverage Rates for $s_{recode5}$; 2x10% and 2x20% Trim

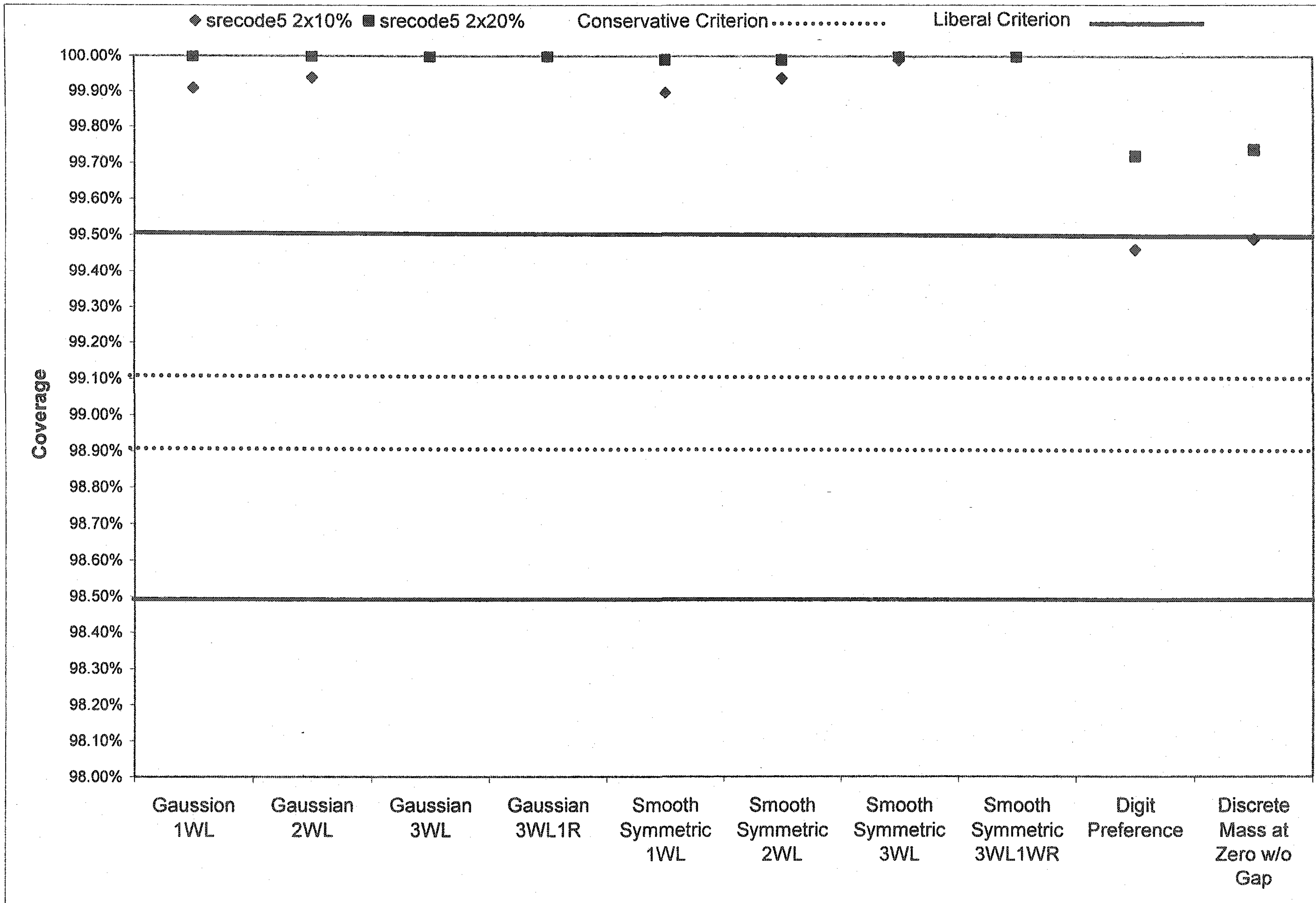


Figure 24. 99% Coverage Rates for $s_{recode6}$; 2x10% and 2x20% Trim

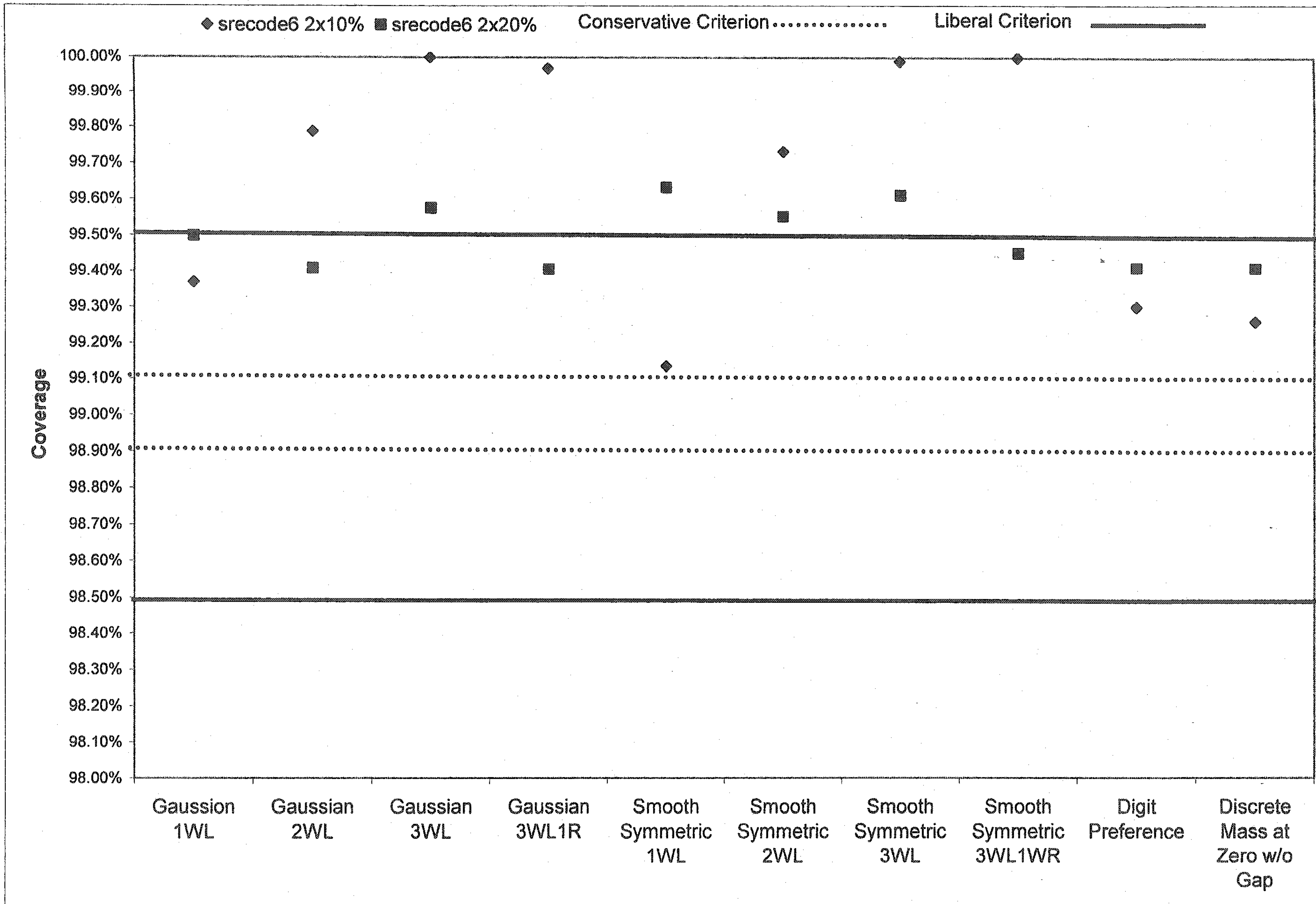


Table 16. Median Location Relative Efficiency For Various Robust Measures of Dispersion; α=.05; 2x10% Trim

Sample	S_w	S_t	$S_{recode1}$	$S_{recode2}$	$S_{recode3}$	$S_{recode4}$	$S_{recode5}$	$S_{recode6}$
Gaussion 1WL	1.00	*	1.00	0.92	*	0.93	*	0.96
Gaussian 2WL	1.00	*	1.00	*	*	*	*	*
Gaussian 3WL	1.00	*	*	*	*	*	*	*
Gaussian 3WL1R	1.00	*	*	*	*	*	*	*
Smooth Symmetric 1WL	1.00	*	0.99	0.93	*	0.94	*	0.96
Smooth Symmetric 2WL	1.00	*	1.00	*	*	*	*	*
Smooth Symmetric 3WL	1.00	*	*	*	*	*	*	*
Smooth Symmetric 3WL1WR	1.00	*	*	*	*	*	*	*
Digit Preference	1.00	*	1.00	0.93	0.92	0.94	0.94	0.96
Discrete Mass at Zero w/o Gap	1.00	*	1.00	0.92	0.90	0.93	0.93	0.96

*Does not meet Bradley's liberal or conservative criteria

Table 17. Median Location Relative Efficiency For Various Robust Measures of Dispersion; α=.05; 2x20% Trim

Sample	s_w	s_t	$s_{recode1}$	$s_{recode2}$	$s_{recode3}$	$s_{recode4}$	$s_{recode5}$	$s_{recode6}$
Gaussion 1WL	1.00	*	*	*	*	*	*	0.92
Gaussian 2WL	1.00	*	1.08	*	*	*	*	0.92
Gaussian 3WL	1.00	*	1.08	*	*	*	*	0.92
Gaussian 3WL1R	1.00	*	*	*	*	*	*	0.92
Smooth Symmetric 1WL	1.00	*	1.08	*	*	*	*	0.92
Smooth Symmetric 2WL	1.00	*	1.08	*	*	*	*	0.92
Smooth Symmetric 3WL	1.00	*	1.08	*	*	*	*	0.92
Smooth Symmetric 3WL1WR	1.00	*	1.08	*	*	*	*	0.92
Digit Preference	1.00	*	1.08	*	*	*	*	0.92
Discrete Mass at Zero w/o Gap	1.00	*	*	*	*	*	*	0.92

*Does not meet Bradley's liberal or conservative criteria

Table 18. Median Location Relative Efficiency For Various Robust Measures of Dispersion; α=.01; 2x10% Trim

Sample	s_w	s_t	$s_{recode1}$	$s_{recode2}$	$s_{recode3}$	$s_{recode4}$	$s_{recode5}$	$s_{recode6}$
Gaussion 1WL	1.00	*	1.00	*	*	*	*	0.96
Gaussian 2WL	1.00	*	1.00	*	*	*	*	*
Gaussian 3WL	1.00	*	*	*	*	*	*	*
Gaussian 3WL1R	1.00	*	*	*	*	*	*	*
Smooth Symmetric 1WL	1.00	*	0.99	0.93	*	0.94	*	0.96
Smooth Symmetric 2WL	1.00	*	1.00	*	*	*	*	*
Smooth Symmetric 3WL	1.00	*	*	*	*	*	*	*
Smooth Symmetric 3WL1WR	1.00	*	*	*	*	*	*	*
Digit Preference	1.00	*	1.00	*	*	0.94	0.94	0.96
Discrete Mass at Zero w/o Gap	1.00	*	1.00	*	*	0.93	0.93	0.96

*Does not meet Bradley's liberal or conservative criteria

Table 19. Median Location Relative Efficiency For Various Robust Measures of Dispersion; α=.01; 2x20% Trim

Sample	s_w	s_t	$s_{recode1}$	$s_{recode2}$	$s_{recode3}$	$s_{recode4}$	$s_{recode5}$	$s_{recode6}$
Gaussion 1WL	1.00	*	*	*	*	*	*	0.92
Gaussian 2WL	1.00	*	*	*	*	*	*	0.92
Gaussian 3WL	1.00	*	*	*	*	*	*	*
Gaussian 3WL1R	1.00	*	*	*	*	*	*	0.92
Smooth Symmetric 1WL	1.00	*	*	*	*	*	*	*
Smooth Symmetric 2WL	1.00	*	*	*	*	*	*	*
Smooth Symmetric 3WL	1.00	*	*	*	*	*	*	*
Smooth Symmetric 3WL1WR	1.00	*	*	*	*	*	*	0.92
Digit Preference	1.00	*	*	*	*	*	*	0.92
Discrete Mass at Zero w/o Gap	1.00	*	*	*	*	*	*	0.92

*Does not meet Bradley's liberal or conservative criteria

Winsorized Standard Deviation

S_w met Bradley's liberal or conservative criteria for robustness for all distributions for 95% and 99% bracketed intervals, and 2 x 10% and 2 x 20% trims.

S_w met Bradley's conservative criterion for all distributions for a 95% bracketed interval using a 10% trim, and for six distributions (Gaussian 2 Wild Left, Gaussian 3 Wild Left, Smooth Symmetric 1 Wild Left, Smooth Symmetric 2 Wild Left, Smooth Symmetric 3 Wild Left, Digit Preference) for a 95% bracketed interval using a 20% trim. Bradley's liberal criterion was met for the remaining four distributions (Gaussian 1 Wild Left, Gaussian 3 Wild Left 1 Wild Right, Smooth Symmetric 3 Wild Left 1 Wild Right, Discrete Mass at Zero Without Gap). In the cases where Bradley's conservative criterion was not met, the bracketed interval coverage was underestimated. The most extreme deviation was a coverage of 94.06%; only .5% below the cutoff of 94.50% for Bradley's conservative criterion.

S_w met Bradley's conservative criterion for six distributions for a 99% bracketed interval using a 10% trim (Gaussian 1 Wild Left, Gaussian 2 Wild Left, Gaussian 3 Wild Left 1 Wild Right, Smooth Symmetric 2 Wild Left, Digit Preference, Discrete Mass at Zero Without Gap). Bradley's liberal criterion was met for the remaining four distributions for a 99% bracketed interval using a 10% trim (Gaussian 3 Wild Left, Smooth Symmetric 1 Wild Left, Smooth Symmetric 3 Wild Left, Smooth Symmetric 3 Wild Left 1 Wild Right).

S_w met Bradley's conservative criterion for three distributions for a 99% bracketed interval using a 20% trim (Gaussian 1 Wild Left, Smooth Symmetric 2 Wild Left, Smooth Symmetric 3 Wild Left). Bradley's liberal criterion was met for the remaining seven distributions for a 99% bracketed interval using a 20% trim (Gaussian 2 Wild Left, Gaussian 3 Wild Left, Gaussian 3 Wild Left 1 Wild Right, Smooth Symmetric 1 Wild Left, Smooth Symmetric 3 Wild Left 1 Wild Right, Digit Preference, Discrete Mass at Zero Without Gap).

The widths of the 95% bracketed intervals for a 2 x 10% trim using s_w were equal to or shorter than the widths of the bracketed intervals using other measures of dispersion for situations where Bradley's liberal or conservative criteria was met. All bracketed interval widths were compared to the bracketed interval widths computed using s_w. Therefore the LRE for s_w is the width of s_w compared to itself, and will always equal 1.00. $S_{recode1}$ produced some intervals that were equal in width to those calculated using s_w. However, Bradley's liberal or conservative criteria were only met for six of the tested distributions.

The widths of the 95% bracketed intervals for a 2 x 20% trim using s_w were both longer and shorter than the widths of the bracketed intervals using other measures of dispersion for situations where Bradley's liberal or conservative criteria were met. $S_{recode1}$ produced some intervals that were shorter in width to those calculated using s_w. The LRE for those intervals that were shorter than the intervals computed using s_w were 1.08. This means that these intervals computed using $s_{recode1}$ were 8 % shorter. However, Bradley's liberal or conservative criteria were only met for seven of the tested distributions using $s_{recode1}$.

The widths of the 99% bracketed intervals for a 2 x 10% trim using s_w were equal to or shorter than the widths of the bracketed intervals using other measures of dispersion for situations where Bradley's liberal or conservative criteria were met. $S_{recode1}$ produced some intervals that were equal in width to those calculated using s_w. However, Bradley's liberal or conservative criteria were only met for six of the tested distributions.

The widths of the 99% bracketed intervals for a 2 x 20% trim using s_w were shorter than the widths of the bracketed intervals using other measures of dispersion for situations where Bradley's liberal or conservative criteria were met.

Trimmed Standard Deviation

S_t did not meet Bradley's liberal or conservative criteria for robustness for any of the distributions for 95% and 99% bracketed intervals, or 2 x 10% and 2 x 20% trims.

S_t 95% bracketed interval coverage was underestimated for both 2 x 10% and 2 x 20% trims for all distributions. For 2 x 10% trimming, the coverage ranged from 89.68% to 90.59%; 4.3%-5.2% below the cutoff of 94.50% for Bradley's conservative criterion, and 2.15%-3.1% below the cutoff of 92.50% for Bradley's liberal criterion. For 2 x 20% trimming, the coverage ranged from 86.23%-87.42%; 8.1%-9.6% below the cutoff of 94.50% for Bradley's conservative criterion, and 5.8%-7.2% below the cutoff of 92.50% for Bradley's liberal criterion.

S_t 99% bracketed interval coverage was underestimated for both 2 x 10% and 2 x 20% trims for all distributions. For 2 x 10% trimming, the coverage ranged from 96.74% to 97.29%; 2.2%-2.9% below the cutoff of 98.90% for Bradley's conservative criterion, and 1.2%-1.8% below the cutoff of 98.50% for Bradley's liberal criterion. For 2 x 20% trimming, the coverage ranged from 95.13% to 95.99%; 3.0%-4.0% below the cutoff of 98.90% for Bradley's conservative criterion, and 2.6%-3.5% below the cutoff of 98.50% for Bradley's liberal criterion.

LRE was not calculated for s_t as none of the bracketed intervals computed using s_t met Bradley's liberal or conservative criteria for robustness.

Recode Method 1 Standard Deviation

$S_{recode1}$ met Bradley's liberal or conservative criteria for robustness for six distributions for 95% bracketed intervals with a 2 x 10% trim, seven distributions for 95% bracketed intervals with a 2 x 20% trim, six distributions for 99% bracketed intervals with a 2 x 10% trim, and zero distributions for 99% bracketed intervals with a 2 x 20% trim. Overall, $s_{recode1}$ performed better for .05 alpha than for .01 alpha, and for 2 x 10% trimming versus 2 x 20% trimming.

$S_{recode1}$ met Bradley's conservative criterion for four distributions for a 95% bracketed interval using a 10% trim (Gaussian 1 Wild Left, Smooth Symmetric 1 Wild Left, Smooth Symmetric 2 Wild Left, Digit Preference). Bradley's liberal criterion was met for two distributions (Gaussian 2 Wild Left, Discrete Mass at Zero Without Gap) for a 95% bracketed interval using a 10% trim. 95% bracketed intervals were inflated beyond Bradley's liberal criterion cutoffs for four distributions (Gaussian 3 Wild Left, Gaussian 3 Wild Left 1 Wild Right, Smooth Symmetric 3 Wild Left, Smooth Symmetric 3 Wild Left 1 Wild Right). The coverage ranged from 99.74% to 99.85% in cases where coverage was inflated; 4.3%-4.4% above the cutoff of 95.50% for Bradley's conservative criterion, and 2.2%-2.4% above the cutoff of 97.50% for Bradley's liberal criterion.

$S_{recode1}$ did not meet Bradley's conservative criterion for a 95% bracketed interval using a 20% trim. Bradley's liberal criterion was met for seven distributions (Gaussian 2 Wild Left, Gaussian 3 Wild Left, Smooth Symmetric 1 Wild Left, Smooth Symmetric 2 Wild Left, Smooth Symmetric 3 Wild Left, Smooth Symmetric 3 Wild Left 1 Wild Right, Digit Preference) for a 95% bracketed interval using a 20% trim. 95% bracketed intervals were underestimated beyond Bradley's liberal criterion cutoffs for three distributions (Gaussian 1 Wild Left, Gaussian 3 Wild Left 1 Wild Right, Discrete Mass at Zero Without Gap). The coverage ranged from 91.93% to 92.41%; 2.3%-2.8% below the cutoff of 94.50% for Bradley's conservative criterion, and .1%-.6% below the cutoff of 92.50% for Bradley's liberal criterion.

$S_{recode1}$ met Bradley's conservative criterion for four distributions for a 99% bracketed interval using a 10% trim (Smooth Symmetric 1 Wild Left, Smooth Symmetric 2 Wild Left, Digit Preference, Discrete Mass at Zero Without Gap). Bradley's liberal criterion was met for two distributions (Gaussian 1 Wild Left, Gaussian 2 Wild Left) for a 99% bracketed interval using a 10% trim. 99% bracketed intervals were inflated beyond Bradley's liberal criterion cutoffs for four distributions (Gaussian 3 Wild Left, Gaussian 3 Wild Left 1 Wild Right, Smooth

Symmetric 3 Wild Left, Smooth Symmetric 3 Wild Left 1 Wild Right). The coverage for these distributions was 100.00%; .9% above the cutoff of 99.10% for Bradley's conservative criterion, and .5% above the cutoff of 99.50% for Bradley's liberal criterion.

$S_{recode1}$ did not meet Bradley's liberal or conservative criteria for a 99% bracketed interval using a 20% trim. 99% bracketed intervals were underestimated for all distributions. The coverage ranged from 97.91% to 98.44%; .5%-1.0% below the cutoff of 98.90% for Bradley's conservative criterion, and .06%-.6% below the cutoff of 98.50% for Bradley's liberal criterion.

The widths of the 95% bracketed intervals for a 2 x 10% trim using $s_{recode1}$ were equal to the widths of the bracketed intervals using s_w for five distributions where Bradley's liberal or conservative criteria was met. These distributions were Gaussian 1 Wild Left, Gaussian 2 Wild Left, Smooth Symmetric 2 Wild Left, Digit Preference, and Discrete Mass at Zero Without Gap. The LRE for these distributions was 1.00. The bracketed interval for a 2 x 10% trim using $s_{recode1}$ was 1% greater than the width of the bracketed interval using s_w for the Smooth Symmetric 1 Wild Left distribution. $S_{recode1}$ does not have an advantage over s_w in terms of bracketed interval width for a 95% bracketed interval with 2 x 10% trimming. Also, $s_{recode1}$ does not result in stable bracketed interval estimates as Bradley's liberal or conservative criteria were not met for all distributions tested.

The widths of the 95% bracketed intervals for a 2 x 20% trim using s_w were shorter than the widths of the bracketed intervals using other measures of dispersion for situations where Bradley's liberal or conservative criteria were met.

The LRE for those intervals that were shorter than the intervals computed using s_w were 1.08. This means that these intervals computed using $s_{recode1}$ were 8 % shorter. $S_{recode1}$ does not result in stable bracketed interval estimates as Bradley's liberal or conservative criteria was only met for seven of the tested distributions using $s_{recode1}$.

The widths of the 99% bracketed intervals for a 2 x 10% trim using $s_{recode1}$ were equal to the widths of the bracketed intervals using s_w for five distributions where Bradley's liberal or conservative criteria was met. These distributions were Gaussian 1 Wild Left, Gaussian 2 Wild Left, Smooth Symmetric 2 Wild Left, Digit Preference, and Discrete Mass at Zero Without Gap. The LRE for these distributions was 1.00. The bracketed interval for a 2 x 10% trim using $s_{recode1}$ was 1% greater than the width of the bracketed interval using s_w for the Smooth Symmetric 1 Wild Left distribution. $S_{recode1}$ does not have an advantage over s_w in terms of bracketed interval width for a 99% bracketed interval with 2 x 10% trimming. Also, $s_{recode1}$ does not result in stable bracketed interval estimates as Bradley's liberal or conservative criteria were not met for all distributions tested.

LRE was not calculated for $s_{recode1}$ for a 99% bracketed interval with 2 x 20% trimming as none of the bracketed intervals computed using $s_{recode1}$ met Bradley's liberal or conservative criteria for robustness.

Recode Method 2 Standard Deviation

$S_{recode2}$ met Bradley's liberal or conservative criteria for robustness for four distributions for 95% bracketed intervals with a 2 x 10% trim, zero distributions for 95% bracketed intervals with a 2 x 20% trim, one distribution for 99%

bracketed intervals with a 2 x 10% trim, and zero distributions for 99% bracketed intervals with a 2 x 20% trim. Overall, $s_{recode2}$ performed better for .05 alpha than for .01 alpha, and for 2 x 10% trimming versus 2 x 20% trimming.

$S_{recode2}$ met Bradley's liberal criterion met for four distributions (Gaussian 1 Wild left, Smooth Symmetric 1 Wild Left, Digit Preference, Discrete Mass at Zero Without Gap) for a 95% bracketed interval using a 10% trim. Bradley's conservative criterion was not met for any 95% bracketed intervals using a 10% trim for $s_{recode2}$. 95% bracketed intervals for the 2 x 10% trim that did not meet Bradley's liberal or conservative criteria were inflated for six distributions (Gaussian 2 Wild Left, Gaussian 3 Wild Left, Gaussian 3 Wild Left 1 Wild Right, Smooth Symmetric 2 Wild Left, Smooth Symmetric 3 Wild Left, Smooth Symmetric 3 Wild Left 1 Wild Right). The coverage for these distributions ranged from 99.57% to 99.92%; 4.1%-4.4% above the cutoff of 95.50% for Bradley's conservative criterion, and 2.1%-2.4% above the cutoff of 97.50% for Bradley's liberal criterion.

$S_{recode2}$ did not meet Bradley's liberal or conservative criteria for a 95% bracketed interval using a 20% trim. 95% bracketed intervals were inflated and the coverage ranged from 99.14% to 100.00%; 3.7%-4.5% above the cutoff of 95.50% for Bradley's conservative criterion, and 1.7%-2.5% above the cutoff of 97.50% for Bradley's liberal criterion.

$S_{recode2}$ met Bradley's liberal criterion for one distribution for a 99% bracketed interval using a 10% trim (Smooth Symmetric 1 Wild Left). Bradley's conservative criterion was not met for any distributions for a 99% bracketed interval using a 10% trim. 99% bracketed intervals were inflated for the nine other distributions. The coverage for these distributions ranged from 99.51%-100.00%; .4%-.9% above the cutoff of 99.10% for Bradley's conservative criterion, and .01%-.5% above the cutoff of 99.50% for Bradley's liberal criterion.

$S_{recode2}$ did not meet Bradley's liberal or conservative criteria for a 99% bracketed interval using a 20% trim. 99% bracketed intervals were inflated for all distributions. The coverage ranged from 99.94% to 100.00%; .8%-.9% above the cutoff of 99.10% for Bradley's conservative criterion, and .4%-.5% above the cutoff of 99.50% for Bradley's liberal criterion.

The widths of the 95% bracketed intervals for a 2 x 10% trim using $s_{recode2}$ were wider than the widths of the bracketed intervals using s_w for four distributions where Bradley's liberal or conservative criteria were met. These distributions were Gaussian 1 Wild Left, Smooth Symmetric 1 Wild Left, Digit Preference, and Discrete Mass at Zero Without Gap. The LRE for these distributions was .92-.93. The bracketed intervals for a 2 x 10% trim using $s_{recode2}$ were 7%-8% greater than the width of the bracketed interval using $s_{w.}$ $S_{recode2}$ does not have an advantage over s_w in terms of bracketed interval width for a 95% bracketed interval with 2 x 10% trimming. Also, $s_{recode2}$ does not result in stable bracketed interval estimates as Bradley's liberal or conservative criteria were not met for all distributions tested.

LRE was not calculated for $s_{recode2}$ for a 95% bracketed interval with 2 x 20% trimming as none of the bracketed intervals computed using $s_{recode2}$ met Bradley's liberal or conservative criteria for robustness.

The width of the 99% bracketed interval for a 2 x 10% trim using $s_{recode1}$ was wider than the width of the bracketed interval using s_w for the one distribution where Bradley's liberal or conservative criteria were met. This distribution was Smooth Symmetric 1 Wild Left. The LRE for this distribution was .93. The bracketed interval for a 2 x 10% trim using $s_{recode2}$ was 7% greater than the width of the bracketed interval using s_w for the Smooth Symmetric 1 Wild Left distribution. $S_{recode2}$ does not have an advantage over s_w in terms of bracketed interval width for a 99% bracketed interval with 2 x 10% trimming. Also, $s_{recode2}$ does not result in stable bracketed interval estimates as Bradley's liberal or conservative criteria were not met for all distributions tested.

LRE was not calculated for $s_{recode2}$ for a 99% bracketed interval with 2 x 20% trimming as none of the bracketed intervals computed using $s_{recode2}$ met Bradley's liberal or conservative criteria for robustness.

Recode Method 3 Standard Deviation

$S_{recode3}$ met Bradley's liberal or conservative criteria for robustness for two distributions for 95% bracketed intervals with a 2 x 10% trim, zero distributions for 95% bracketed intervals with a 2 x 20% trim, zero distributions for 99% bracketed intervals with a 2 x 10% trim, and zero distributions for 99% bracketed intervals with a 2 x 20% trim.

$S_{recode3}$ met Bradley's liberal criterion for two distributions (Digit Preference, Discrete Mass at Zero Without Gap) for a 95% bracketed interval using a 10% trim. Bradley's conservative criterion was not met for any 95% bracketed intervals using a 10% trim for $s_{recode3}$. 95% bracketed intervals for the

2 x 10% trim that did not meet Bradley's liberal or conservative criteria were inflated for eight distributions. The coverage ranged from 98.56% to 99.91%; 3.1%-4.4% above the cutoff of 95.50% for Bradley's conservative criterion, and 1.1%-2.4% above the cutoff of 97.50% for Bradley's liberal criterion.

$S_{recode3}$ did not meet Bradley's liberal or conservative criteria for a 95% bracketed interval using a 20% trim. 95% bracketed intervals were inflated and the coverage ranged from 98.55% to 99.99%; 3.1%-4.5% above the cutoff of 95.50% for Bradley's conservative criterion, and 1.1%-2.5% above the cutoff of 97.50% for Bradley's liberal criterion.

$S_{recode3}$ did not meet Bradley's liberal or conservative criterion for a 99% bracketed interval using a 10% trim. 99% bracketed intervals were inflated and the coverage for these distributions ranged from 99.54%-100.00%; .4%-.9% above the cutoff of 99.10% for Bradley's conservative criterion, and .04%-.5% above the cutoff of 99.50% for Bradley's liberal criterion.

$S_{recode3}$ did not meet Bradley's liberal or conservative criteria for a 99% bracketed interval using a 20% trim. 99% bracketed intervals were inflated for all distributions. The coverage ranged from 99.88% to 100.00%; .7%-.9% above the cutoff of 99.10% for Bradley's conservative criterion, and .4%-.5% above the cutoff of 99.50% for Bradley's liberal criterion.

The widths of the 95% bracketed intervals for a 2 x 10% trim using $s_{recode3}$ were wider than the widths of the bracketed intervals using s_w for two distributions where Bradley's liberal or conservative criteria was met. These distributions were Digit Preference, and Discrete Mass at Zero Without Gap. The

LRE for these distributions was .90-.92. The bracketed intervals for a 2 x 10% trim using $s_{recode3}$ were 8%-10% greater than the width of the bracketed intervals using s_w. $S_{recode3}$ does not have an advantage over s_w in terms of bracketed interval width for a 95% bracketed interval with 2 x 10% trimming. Also, $s_{recode3}$ does not result in stable bracketed interval estimates as Bradley's liberal or conservative criteria were not met for all distributions tested.

LRE was not calculated for $s_{recode3}$ for a 95% bracketed interval with 2 x 20% trimming, 99% bracketed interval with 2 x 10% trimming, or 99% bracketed interval with 2 x 20% trimming as none of the bracketed intervals computed using $s_{recode3}$ met Bradley's liberal or conservative criteria for robustness.

Recode Method 4 Standard Deviation

$S_{recode4}$ met Bradley's liberal or conservative criteria for robustness for four distributions for 95% bracketed intervals with a 2 x 10% trim, zero distributions for 95% bracketed intervals with a 2 x 20% trim, three distributions for 99% bracketed intervals with a 2 x 10% trim, and zero distributions for 99% bracketed intervals with a 2 x 20% trim.

$S_{recode4}$ met Bradley's liberal criterion for four distributions (Gaussian 1 Wild Left, Smooth Symmetric 1 Wild Left, Digit Preference, Discrete Mass at Zero Without Gap) for a 95% bracketed interval using a 10% trim. Bradley's conservative criterion was not met for any 95% bracketed intervals using a 10% trim for $s_{recode4}$. 95% bracketed intervals for the 2 x 10% trim that did not meet Bradley's liberal or conservative criteria were inflated for six distributions. The coverage ranged from 98.33% to 99.94%; 2.9%-4.4% above the cutoff of

95.50% for Bradley's conservative criterion, and .8%-2.4% above the cutoff of 97.50% for Bradley's liberal criterion.

$s_{recode4}$ did not meet Bradley's liberal or conservative criteria for a 95% bracketed interval using a 20% trim. 95% bracketed intervals were inflated and the coverage ranged from 97.79% to 98.29%; 2.3%-2.8% above the cutoff of 95.50% for Bradley's conservative criterion, and .3%-.8% above the cutoff of 97.50% for Bradley's liberal criterion.

$s_{recode4}$ met Bradley's liberal criterion for three distributions (Smooth Symmetric 1 Wild Left, Digit Preference, Discrete Mass at Zero Without Gap) for a 99% bracketed interval using a 10% trim. Bradley's conservative criterion was not met for any 99% bracketed intervals using a 10% trim for $s_{recode4}$. 99% bracketed intervals for the 2 x 10% trim that did not meet Bradley's liberal or conservative criteria were inflated for seven distributions. Coverage for these distributions ranged from 99.53%-100.00%; .4%-.9% above the cutoff of 99.10% for Bradley's conservative criterion, and .03%-.5% above the cutoff of 99.50% for Bradley's liberal criterion.

$s_{recode4}$ did not meet Bradley's liberal or conservative criteria for a 99% bracketed interval using a 20% trim. 99% bracketed intervals were inflated for all distributions. The coverage ranged from 99.71% to 99.85%; .6% above the cutoff of 99.10% for Bradley's conservative criterion, and .2%-.4% above the cutoff of 99.50% for Bradley's liberal criterion.

The widths of the 95% bracketed intervals for a 2 x 10% trim using $s_{recode4}$ were wider than the widths of the bracketed intervals using s_w for four distributions where Bradley's liberal or conservative criteria was met. These distributions were Gaussian 1 Wild Left, Smooth Symmetric 1 Wild Left, Digit Preference, and Discrete Mass at Zero Without Gap. The LRE for these distributions was .93-.94. The bracketed intervals for a 2 x 10% trim using $s_{recode4}$ were 6%-7% greater than the width of the bracketed intervals using $s_{w.}$ $S_{recode4}$ does not have an advantage over s_w in terms of bracketed interval width for a 95% bracketed interval with 2 x 10% trimming. Also, $s_{recode4}$ does not result in stable bracketed interval estimates as Bradley's liberal or conservative criteria were not met for all distributions tested.

LRE was not calculated for $s_{recode4}$ for a 95% bracketed interval with 2 x 20% trimming as none of the bracketed intervals computed using $s_{recode4}$ met Bradley's liberal or conservative criteria for robustness.

The widths of the 99% bracketed intervals for a 2 x 10% trim using $s_{recode4}$ were wider than the widths of the bracketed intervals using s_w for three distributions where Bradley's liberal or conservative criteria was met. These distributions were Smooth Symmetric 1 Wild Left, Digit Preference, and Discrete Mass at Zero Without Gap. The LRE for these distributions was .93-.94. The bracketed intervals for a 2 x 10% trim using $s_{recode4}$ were 6%-7% greater than the width of the bracketed intervals using $s_{w.}$ $S_{recode4}$ does not have an advantage over s_w in terms of bracketed interval width for a 99% bracketed interval with 2 x 10% trimming. Also, $s_{recode4}$ does not result in stable bracketed interval estimates as Bradley's liberal or conservative criteria were not met for all distributions tested.

LRE was not calculated for $s_{recode4}$ for a 99% bracketed interval with 2 x 20% trimming as none of the bracketed intervals computed using $s_{recode4}$ met Bradley's liberal or conservative criteria for robustness.

Recode Method 5 Standard Deviation

$S_{recode5}$ met Bradley's liberal or conservative criteria for robustness for two distributions for 95% bracketed intervals with a 2 x 10% trim, zero distributions for 95% bracketed intervals with a 2 x 20% trim, two distributions for 99% bracketed intervals with a 2 x 10% trim, and zero distributions for 99% bracketed intervals with a 2 x 20% trim.

$S_{recode5}$ met Bradley's liberal criterion for two distributions (Digit Preference, Discrete Mass at Zero Without Gap) for a 95% bracketed interval using a 10% trim. Bradley's conservative criterion was not met for any 95% bracketed intervals using a 10% trim for $s_{recode5}$. 95% bracketed intervals for the 2 x 10% trim that did not meet Bradley's liberal or conservative criteria were inflated for eight distributions. The coverage ranged from 98.39% to 99.87%; 2.9%-4.4% above the cutoff of 95.50% for Bradley's conservative criterion, and .9%-2.4% above the cutoff of 97.50% for Bradley's liberal criterion.

$S_{recode5}$ did not meet Bradley's liberal or conservative criteria for a 95% bracketed interval using a 20% trim. 95% bracketed intervals were inflated and the coverage ranged from 97.84% to 99.96%; 2.4%-4.5% above the cutoff of 95.50% for Bradley's conservative criterion, and .3%-2.5% above the cutoff of 97.50% for Bradley's liberal criterion.

$S_{recode5}$ met Bradley's liberal criterion for two distributions (Digit Preference, Discrete Mass at Zero Without Gap) for a 99% bracketed interval using a 10% trim. Bradley's conservative criterion was not met for any 99% bracketed intervals using a 10% trim for $s_{recode5}$. 99% bracketed intervals for the 2 x 10% trim that did not meet Bradley's liberal or conservative criteria were inflated for eight distributions. Coverage for these distributions ranged from 99.90%-100.00%; .8%-.9% above the cutoff of 99.10% for Bradley's conservative criterion, and .4%-.5% above the cutoff of 99.50% for Bradley's liberal criterion.

$S_{recode5}$ did not meet Bradley's liberal or conservative criteria for a 99% bracketed interval using a 20% trim. 99% bracketed intervals were inflated for all distributions. The coverage ranged from 99.73% to 100.00%; .6%-.9% above the cutoff of 99.10% for Bradley's conservative criterion, and .2%-.5% above the cutoff of 99.50% for Bradley's liberal criterion.

The widths of the 95% bracketed intervals for a 2 x 10% trim using $s_{recode5}$ were wider than the widths of the bracketed intervals using s_w for two distributions where Bradley's liberal or conservative criteria was met. These distributions were Digit Preference, and Discrete Mass at Zero Without Gap. The LRE for these distributions was .93-.94. The bracketed intervals for a 2 x 10% trim using $s_{recode5}$ were 6%-7% greater than the width of the bracketed intervals using $s_{w.}$ $S_{recode5}$ does not have an advantage over s_w in terms of bracketed interval width for a 95% bracketed interval with 2 x 10% trimming. Also, $s_{recode5}$

does not result in stable bracketed interval estimates as Bradley's liberal or conservative criteria were not met for all distributions tested.

LRE was not calculated for $s_{recode5}$ for a 95% bracketed interval with 2 x 20% trimming as none of the bracketed intervals computed using $s_{recode5}$ met Bradley's liberal or conservative criteria for robustness.

The widths of the 99% bracketed intervals for a 2 x 10% trim using $s_{recode5}$ were wider than the widths of the bracketed intervals using s_w for two distributions where Bradley's liberal or conservative criteria was met. These distributions were Digit Preference, and Discrete Mass at Zero Without Gap. The LRE for these distributions was .93-.94. The bracketed intervals for a 2 x 10% trim using $s_{recode5}$ were 6%-7% greater than the width of the bracketed intervals using $s_{w.}$ $S_{recode5}$ does not have an advantage over s_w in terms of bracketed interval width for a 99% bracketed interval with 2 x 10% trimming. Also, $s_{recode5}$ does not result in stable bracketed interval estimates as Bradley's liberal or conservative criteria were not met for all distributions tested.

LRE was not calculated for $s_{recode5}$ for a 99% bracketed interval with 2 x 20% trimming as none of the bracketed intervals computed using $s_{recode5}$ met Bradley's liberal or conservative criteria for robustness.

Recode Method 6 Standard Deviation

$S_{recode6}$ met Bradley's liberal or conservative criteria for robustness for four distributions for 95% bracketed intervals with a 2 x 10% trim, ten distributions for 95% bracketed intervals with a 2 x 20% trim, four distributions for 99%

bracketed intervals with a 2 x 10% trim, and six distributions for 99% bracketed intervals with a 2 x 20% trim.

$s_{recode6}$ met Bradley's liberal criterion for four distributions (Gaussian 1 Wild Left, Smooth Symmetric 1 Wild Left, Digit Preference, Discrete Mass at Zero Without Gap) for a 95% bracketed interval using a 10% trim. Bradley's conservative criterion was not met for any 95% bracketed intervals using a 10% trim for $s_{recode6}$. 95% bracketed intervals for the 2 x 10% trim that did not meet Bradley's liberal or conservative criteria were inflated for six distributions. The coverage ranged from 98.03% to 99.73%; 2.6%-4.2% above the cutoff of 95.50% for Bradley's conservative criterion, and .5%-2.2% above the cutoff of 97.50% for Bradley's liberal criterion.

$s_{recode6}$ met Bradley's liberal criterion for all ten distributions for a 95% bracketed interval using a 20% trim. Bradley's conservative criterion was not met for any 95% bracketed intervals using a 20% trim for $s_{recode6}$. The coverage ranged from 96.10% to 96.90%; .6%-1.4% above the cutoff of 95.50% for Bradley's conservative criterion.

$s_{recode6}$ met Bradley's liberal criterion for four distributions (Gaussian 1 Wild Left, Smooth Symmetric 1 Wild Left, Digit Preference, Discrete Mass at Zero Without Gap) for a 99% bracketed interval using a 10% trim. Bradley's conservative criterion was not met for any 99% bracketed intervals using a 10% trim for $s_{recode6}$. 99% bracketed intervals for the 2 x 10% trim that did not meet Bradley's liberal or conservative criteria were inflated for six distributions. Coverage for these distributions ranged from 99.74%-100.00%; .6%-.9% above the cutoff of 99.10% for Bradley's conservative criterion, and .2%-.5% above the cutoff of 99.50% for Bradley's liberal criterion.

$S_{recode6}$ met Bradley's liberal criterion for six distributions (Gaussian 1 Wild Left, Gaussian 2 Wild Left, Gaussian 3 Wild Left 1 Wild Right, Smooth Symmetric 3 Wild Left 1 Wild Right, Digit Preference, Discrete Mass at Zero Without Gap) for a 99% bracketed interval using a 20% trim. Bradley's conservative criterion was not met for any 99% bracketed intervals using a 20% trim for $s_{recode6}$. 99% bracketed intervals for the 2 x 20% trim that did not meet Bradley's liberal or conservative criteria were inflated for four distributions. Coverage for these distributions ranged from 99.56%-99.64%; .4%-.5% above the cutoff of 99.10% for Bradley's conservative criterion, and .06%-.1% above the cutoff of 99.50% for Bradley's liberal criterion.

The widths of the 95% bracketed intervals for a 2 x 10% trim using $s_{recode6}$ were wider than the widths of the bracketed intervals using s_w for four distributions where Bradley's liberal or conservative criteria was met. These distributions were Gaussian 1 Wild Left, Smooth Symmetric 1 Wild Left, Digit Preference, and Discrete Mass at Zero Without Gap. The LRE for these distributions was .96. The bracketed intervals for a 2 x 10% trim using $s_{recode6}$ were 4% greater than the width of the bracketed intervals using $s_{w.}$ $S_{recode6}$ does not have an advantage over s_w in terms of bracketed interval width for a 95% bracketed interval with 2 x 10% trimming. Also, $s_{recode6}$ does not result in stable bracketed interval estimates as Bradley's liberal or conservative criteria were not met for all distributions tested.

The widths of the 95% bracketed intervals for a 2 x 20% trim using $s_{recode6}$ were wider than the widths of the bracketed intervals using s_w for the ten distributions where Bradley's liberal or conservative criteria was met. The LRE for these distributions was .92. The bracketed intervals for a 2 x 20% trim using $s_{recode6}$ were 8% greater than the width of the bracketed intervals using $s_{w.}$ $S_{recode6}$ does not have an advantage over s_w in terms of bracketed interval width for a 95% bracketed interval with 2 x 10% trimming.

The widths of the 99% bracketed intervals for a 2 x 10% trim using $s_{recode6}$ were wider than the widths of the bracketed intervals using s_w for four distributions where Bradley's liberal or conservative criteria was met. These distributions were Gaussian 1 Wild Left, Smooth Symmetric 1 Wild Left, Digit Preference, and Discrete Mass at Zero Without Gap. The LRE for these distributions was .96. The bracketed intervals for a 2 x 10% trim using $s_{recode6}$ were 4% greater than the width of the bracketed intervals using $s_{w.}$ $S_{recode6}$ does not have an advantage over s_w in terms of bracketed interval width for a 99% bracketed interval with 2 x 10% trimming. Also, $s_{recode6}$ does not result in stable bracketed interval estimates as Bradley's liberal or conservative criteria were not met for all distributions tested.

The widths of the 99% bracketed intervals for a 2 x 20% trim using $s_{recode6}$ were wider than the widths of the bracketed intervals using s_w for six distributions where Bradley's liberal or conservative criteria was met. These distributions were Gaussian 1 Wild Left, Gaussian 2 Wild Left, Gaussian 3 Wild Left 1 Wild Right, Smooth Symmetric 3 Wild Left 1 Wild Right, Digit Preference, and Discrete Mass at Zero Without Gap. The LRE for these distributions was .92. The bracketed intervals for a 2 x 20% trim using $s_{recode6}$ were 8% greater than the width of the bracketed intervals using s_w. $S_{recode6}$ does not have an advantage over s_w in terms of bracketed interval width for a 99% bracketed interval with 2 x 20% trimming. Also, $s_{recode6}$ does not result in stable bracketed interval estimates as Bradley's liberal or conservative criteria were not met for all distributions tested.

CHAPTER 5

CONCLUSION

The purpose of this study was to examine the properties of bracketed intervals around the trimmed mean using alternatives to s_w as the measure of dispersion. The impetus for this study was the lack of a true theoretical treatment in the literature of the use of s_w in computing the bracketed interval around the trimmed mean, and the mass at the recoding points due to winsorization.

This study evaluated bracketed interval coverage and widths for the alternative measures of dispersion. These measures of dispersion are not an inclusive list. In fact Lax (1985) identified over 150 different robust measures of dispersion. The measures of dispersion that were tested in this study were chosen because they would deal with outliers without creating a mass in the tails of the distribution. The alternative measures of dispersion were used to create bracketed intervals around the trimmed mean for ten different symmetric distributions. Four of the distributions were theoretical Gaussian distributions with added outliers, and six of the distributions were prevalent educational and psychological distributions. Of the six prevalent educational and psychological distributions, four were smooth symmetric with added outliers, and the remaining two were digit preference and discrete mass at zero without gap. The six educational and psychological distributions have practical significance.

The results of this study will be discussed with two major emphases: (a) bracketed interval coverage associated with each alternative measure of dispersion for a 2 x 10% trim and 2 x 20% trim, .05 and .01 alpha, and for the 10 different distributions (b) Location Relative Efficiency (LRE) for each alternative measure of dispersion meeting Bradley's (1978) criteria for robustness for each alternative measure of dispersion with a 2 x 10% trim and 2 x 20% trim, .05 and .01 alpha, and for the 10 different distributions.

Study Conclusions

Winsorized Standard Deviation (s_w)

The winsorized standard deviation was included in this study in order to test the performance of s_w across different symmetric distributions, different trim percentages (2 x 10%, 2 x 20%), and alpha levels (.05, .01).

The results of this study support the use of s_w in calculating the bracketed interval around the trimmed mean. S_w is the only measure of dispersion that produced bracketed intervals with coverage rates that met Bradley's liberal or conservative criteria for both alpha levels (.05 and .01), and across all of the distributions tested.

Although s_w met Bradley's liberal or conservative criteria for both alpha levels (.05, .01) for all distributions tested, s_w performed better at the .05 alpha level. The 99% bracketed interval coverage rates were slightly underestimated, and there were eleven cases where Bradley's conservative criterion was not met, versus four cases for the 95% bracketed interval.

Another interesting finding was that s_w performed better when a 2 x 10% trim was used versus a 2 x 20% trim. For the 95% bracketed interval, s_w met Bradley's conservative criterion for all distributions tested when a 2 x 10% trim was used, versus 6 distributions when a 2 x 20% trim was used. The same trend was observed for the 99% bracketed interval, where Bradley's conservative criterion was met for six distributions when a 2 x 10% trim was used, versus three distributions when a 2 x 20% trim was used. The results of the trial of s_w as a measure of dispersion support the use of lightly trimming.

In most cases, s_w produced bracketed intervals with widths that were equal to or shorter than the bracketed interval widths of the competing measures of dispersion. The only case where s_w produced bracketed intervals that were longer than a competitor was for the 95% bracketed interval with a 2 x 20% trim. However, for the competing measure of dispersion that resulted in shorter intervals ($s_{recode1}$), Bradley's liberal or conservative criteria were only met for 70% of the distributions tested. Therefore, s_w has an advantage of being stable over the different distributions tested.

It is surprising that s_w resulted in some bracketed intervals where nominal alpha was slightly inflated. This suggests that while s_w was the best performer in this study, further research is necessary to determine the conditions under which nominal alpha becomes inflated.

Trimmed Standard Deviation (s_t)

S_t was included in this study because it is seen a "naïve" measure of dispersion for consideration in calculating the bracketed interval around the trimmed mean. The results for s_t were not favorable, however the findings were interesting and consistent with the findings of s_w. S_t was unstable for both alpha levels (.05, .01), and trim percentages, failing to meet Bradley's liberal or conservative criteria for any of the distributions tested.

Alpha levels were underestimated in all cases, with the largest underestimation occurring with heavy trimming. Although s_t is obviously not a good alternative to s_w for forming a bracketed interval around the trimmed mean, the findings with regards to heavy vs. light trimming are consistent with the findings of s_w; in support of lighter trimming.

The reason for the poor performance of s_t is probably due to the extreme treatment of outliers. Trimming does not create a buildup of data in the tails of the distribution; however trimming of the data set decreases the degrees of freedom of the sample resulting in a larger standard error than if the sample had a larger degree of freedom.

Recode Method 1 Standard Deviation ($s_{recode1}$)

Recode Method1 is a combination of trimming and winsorizing outliers. $S_{recode1}$ was considered in this study because Recode Method 1 is a compromise between all out trimming or winsorizing. There is not as much of a buildup of data in the tails of the distribution as this recoding method does not rely only on winsorizing, and there isn't as extreme of a decrease in the degrees of freedom as this method does not rely only on trimming.

$S_{recode1}$ did meet Bradley's liberal or conservative criteria for several distributions. However the performance of this measure of dispersion cannot be considered stable, as the coverage rates did not meet Bradley's liberal or conservative criteria for all distributions tested under a particular alpha level or trim/winsor percentage. Overall, $s_{recode1}$ performed better at a .05 alpha level, and for 2 x 10% trimming.

$S_{recode1}$ performed better when a 2 x 10% trim was used versus a 2 x 20% trim. For the 95% bracketed interval, $s_{recode1}$ met Bradley's liberal or conservative criteria for 6 distributions when a 2 x 10% trim was used (4 met conservative criterion), versus 7 distributions when a 2 x 20% trim was used (zero met conservative criterion). The same trend was observed for the 99% bracketed interval, where Bradley's liberal criterion was met for 6 distributions when a 2 x 10% trim was used (4 met conservative criterion), versus zero distributions when a 2 x 20% trim was used. Although $s_{recode1}$ did not perform as well as $s_{w,}$ the findings with regards to heavy vs. light trimming are consistent with the findings of s_w; in support of light trimming.

$S_{recode1}$ produced intervals that were shorter than s_w (8% shorter in some cases) for 95% bracketed intervals for a 2 x 20% trim. However as mentioned above, the bracketed interval estimates are not stable, in this case only 70% of intervals tested using this trim and alpha level met Bradley's liberal or conservative criteria.

$S_{recode1}$ did not perform as well or better than s_w. As in the case of all out trimming, the trimming involved in calculating this measure of dispersion decreases the degrees of freedom of the sample resulting in a larger standard error than if the sample had a larger degree of freedom. Although this decrease in degrees of freedom is not as substantial as in all out trimming, it may have had an effect on the instability of this procedure in calculating the bracketed interval around the trimmed mean.

Recode Method 2 Standard Deviation ($s_{recode2}$)

Recode Method 2 is a recoding procedure in which the value to be recoded takes on the next value in the data set towards the median. $S_{recode2}$ was considered in this study because Recode Method 2 is a compromise to winsorizing. Recode Method 2 does not involve any trimming, so there is no loss in the degrees of freedom. Also, there is not a buildup of data in the tails of the distribution as this recoding method deals with outliers by recoding them towards the median smoothly.

$S_{recode2}$ did meet Bradley's liberal or conservative criteria for several distributions. However the performance of this measure of dispersion cannot be considered stable, as the coverage rates did not meet Bradley's liberal or conservative criteria for all distributions tested under a particular alpha level or trim/recode percentage. Overall, $s_{recode2}$ performed better at a .05 alpha level, and for 2 x 10% trimming.

$S_{recode2}$ performed better when a 2 x 10% trim was used versus a 2 x 20% trim. For the 95% bracketed interval, $s_{recode2}$ met Bradley's liberal or conservative criteria for 4 distributions when a 2 x 10% trim was used (zero met conservative criterion), versus zero distributions when a 2 x 20% trim was used. The same trend was observed for the 99% bracketed interval, where Bradley's liberal criterion was met for 1 distribution when a 2 x 10% trim was used (zero met conservative criterion), versus zero distributions when a 2 x 20% trim was used. Although $s_{recode2}$ did not perform as well as s_w, the findings with regards to heavy vs. light trimming are consistent with the findings of s_w; in support of light trimming.

$S_{recode2}$ did not produce intervals that were shorter than s_w for the cases where Bradley's liberal or conservative criteria were met. $S_{recode2}$ therefore does not hold any advantages over s_w, as the bracketed interval estimates are not stable, and the recoding method does not result in bracketed intervals that are narrower than those produced using s_w.

$S_{recode2}$ did not perform as well or better than s_w. The difference between this recoding procedure and winsorizing was that this recoding procedure attempted to smooth the tails of the distribution when recoding values towards the median. This procedure did not treat outliers as extremely as winsorizing, and perhaps did not treat the outliers enough.

Recode Method 3 Standard Deviation ($s_{recode3}$)

Recode Method 3 is a recoding procedure in which the value to be recoded is averaged with the winsorized value to get the recoded value. $S_{recode3}$ was considered in this study because Recode Method 3 is a compromise to winsorizing. Recode Method 3 does not involve any trimming, so there is no loss in the degrees of freedom. Also, there is not a buildup of data in the tails of the distribution as this recoding method deals with outliers by recoding them towards the median smoothly. The decrease in variance is not however as extreme as Recode Method 2.

$S_{recode3}$ did meet Bradley's liberal or conservative criteria for 2 distributions for a 2 x 10% trim, and .05 alpha level. However the performance of this measure of dispersion cannot be considered stable, as the coverage rates did not meet Bradley's liberal or conservative criteria for all distributions tested under a particular alpha level or trim/recode percentage. Overall, $s_{recode3}$ performed better at a .05 alpha level (none of 99% bracketed intervals met Bradley's liberal or conservative criteria), and for 2 x 10% trimming.

$S_{recode3}$ performed better when a 2 x 10% trim was used versus a 2 x 20% trim. For the 95% bracketed interval, $s_{recode3}$ met Bradley's liberal or conservative criteria for 2 distributions when a 2 x 10% trim was used (zero met conservative criterion), versus zero distributions when a 2 x 20% trim was used. Although $s_{recode3}$ did not perform as well as s_w, the findings with regards to heavy vs. light trimming are consistent with the findings of s_w; in support of light trimming.

$S_{recode3}$ did not produce intervals that were shorter than s_w for the cases where Bradley's liberal or conservative criteria were met. $S_{recode3}$ therefore does not hold any advantages over s_w as the bracketed interval estimates are not stable, and the recoding method does not result in bracketed intervals that are narrower than those produced using s_w.

$S_{recode3}$ did not perform as well or better than s_w. The difference between this recoding procedure and winsorizing was that this recoding procedure attempted to smooth the tails of the distribution when recoding values towards the median. This procedure did not treat outliers as extremely as winsorizing, and perhaps did not treat the outliers enough as in the case of $s_{recode2}$.

Recode Method 4 Standard Deviation ($s_{recode4}$)

Recode Method 4 is a recoding procedure in which the median of the value to be recoded and the winsorized value is taken to get the recoded value. $S_{recode4}$ was considered in this study because Recode Method 4 is a compromise between winsorizing. Like Recode Methods 2 and 3, Recode Method 4 does not involve any trimming, so there is no loss in the degrees of freedom. Also, there is not a buildup of data in the tails of the distribution as this recoding method deals with outliers by recoding them towards the median smoothly. The decrease in variance is not however as extreme as Recode Method 2, but is more extreme than Recode Method 3. The variance will fall between the variance of Recode Methods 2 and 3.

$S_{recode4}$ did meet Bradley's liberal or conservative criteria for several distributions. However the performance of this measure of dispersion cannot be considered stable, as the coverage rates did not meet Bradley's liberal or conservative criteria for all distributions tested under a particular alpha level or trim/recode percentage. Overall, $s_{recode4}$ performed better at a .05 alpha level, and for 2 x 10% trimming.

$S_{recode4}$ performed better when a 2 x 10% trim was used versus a 2 x 20% trim. For the 95% bracketed interval, $s_{recode4}$ met Bradley's liberal criterion for 4 distributions when a 2 x 10% trim was used (zero met conservative criterion), versus zero distributions when a 2 x 20% trim was used. The same trend was observed for the 99% bracketed interval, where Bradley's liberal criterion was met for 3 distributions when a 2 x 10% trim was used (zero met conservative criterion), versus zero distributions when a 2 x 20% trim was used. Although $s_{recode4}$ did not perform as well as s_w, the findings with regards to heavy vs. light trimming are consistent with the findings of s_w; in support of light trimming.

$S_{recode4}$ did not produce intervals that were shorter than s_w for the cases where Bradley's liberal or conservative criteria were met. $S_{recode4}$ therefore does not hold any advantages over s_w as the bracketed interval estimates are not stable, and the recoding method does not result in bracketed intervals that are narrower than those produced using s_w.

$S_{recode4}$ did not perform as well or better than s_w. The difference between this recoding procedure and winsorizing was that this recoding procedure attempted to smooth the tails of the distribution when recoding values towards the median. This procedure did not treat outliers as extremely as winsorizing, and perhaps did not treat the outliers enough as in the case of $s_{recode3}$ and $s_{recode2}$.

Recode Method 5 Standard Deviation ($s_{recode5}$)

Recode Method 5 is a recoding procedure in which the average of the value to be recoded and the Bunner-Sawilowsky value is taken to get the recoded value. $S_{recode5}$ was considered in this study as a variation of the Bunner-Sawilowsky approach that gives some weight to outliers. Like Recode Methods 2, 3, and 4, Recode Method 5 does not involve any trimming, so there is no loss in the degrees of freedom. Also, there is not a buildup of data in the tails of the distribution as this recoding method deals with outliers by recoding them towards the median smoothly. The decrease in variance is not however as extreme as

Recode Methods 2 and 4, but is more extreme than Recode Method 3 as the Bunner-Sawilowsky recoding scheme brings outlying values closer to the median than winsorization.

$S_{recode5}$ did meet Bradley's liberal or conservative criteria for a few distributions. However the performance of this measure of dispersion cannot be considered stable, as the coverage rates did not meet Bradley's liberal or conservative criteria for all distributions tested under a particular alpha level or trim/recode percentage. Overall, $s_{recode5}$ performed the same at both alpha levels, but performed better for 2 x 10% trimming.

$S_{recode5}$ performed better when a 2 x 10% trim was used versus a 2 x 20% trim. For the 95% bracketed interval, $s_{recode5}$ met Bradley's liberal criterion for 2 distributions when a 2 x 10% trim was used (zero met conservative criterion), versus zero distributions when a 2 x 20% trim was used. The same trend was observed for the 99% bracketed interval, where Bradley's liberal criterion was met for 2 distributions when a 2 x 10% trim was used (zero met conservative criterion), versus zero distributions when a 2 x 20% trim was used. Although $s_{recode5}$ did not perform as well as $s_{w,}$ the findings with regards to heavy vs. light trimming are consistent with the findings of s_w; in support of light trimming.

$S_{recode5}$ did not produce intervals that were shorter than s_w for the cases where Bradley's liberal or conservative criteria were met. $S_{recode5}$ therefore does not hold any advantages over s_w as the bracketed interval estimates are not stable, and the recoding method does not result in bracketed intervals that are narrower than those produced using s_w.

$S_{recode5}$ did not perform as well or better than s_w. The difference between this recoding procedure and winsorizing was that this recoding procedure attempted to smooth the tails of the distribution when recoding values towards the median. This procedure did not treat outliers as extremely as winsorizing, and perhaps did not treat the outliers enough as in the case of $s_{recode4}$, $s_{recode3}$, and $s_{recode2}$.

Recode Method 6 Standard Deviation ($s_{recode6}$)

Recode Method 6 is a recoding procedure in which the median of the value to be recoded and the Bunner-Sawilowsky value is taken to get the recoded value. $S_{recode6}$ was considered in this study as a variation of the Bunner-Sawilowsky that gives some weight to outliers though not as much weight as $s_{recode5}$ as the median rather than the mean of the value to be recoded and the Bunner-Sawilowsky value is taken. Like Recode Methods 2, 3, 4, and 5, Recode Method 6 does not involve any trimming, so there is no loss in the degrees of freedom. Also, there is not a buildup of data in the tails of the distribution as this recoding method deals with outliers by recoding them towards the median smoothly. The decrease in variance is not however as extreme as Recode Method 2, but is more extreme than Recode Method 3 and 4 as the Bunner-Sawilowsky recoding scheme brings outlying values closer to the median than winsorization, and more extreme than Recode Method 5 as the median rather than the average is used.

$S_{recode6}$ did meet Bradley's liberal or conservative criteria for many distributions. $S_{recode6}$, unlike the other recoding procedures, did produce stable bracketed intervals for all distributions tested under the .05 alpha level with a 2 x 20% trim. This measure of dispersion however was not stable under the other alpha level/trim combinations as the coverage rates did not meet Bradley's liberal or conservative criteria for all distributions tested under a particular alpha level or trim/recode percentage. Overall, $s_{recode6}$ performed better for the .05 alpha level, and for 2 x 20% trimming.

$S_{recode6}$ performed better when a 2 x 20% trim was used versus a 2 x 10% trim. This was not consistent with the findings of the other measures of dispersion that were investigated in this study. For the 95% bracketed interval, $s_{recode6}$ met Bradley's liberal criterion for 4 distributions when a 2 x 10% trim was used (zero met conservative criterion), versus 10 distributions when a 2 x 20% trim was used (zero met conservative criterion). The same trend was observed for the 99% bracketed interval, where Bradley's liberal criterion was met for 4 distributions when a 2 x 10% trim was used (zero met conservative criterion), versus 6 distributions when a 2 x 20% trim was used (zero met conservative criterion).

$S_{recode6}$ did not produce intervals that were shorter than s_w for the cases where Bradley's liberal or conservative criteria were met. $S_{recode6}$ therefore does not hold any advantages over s_w as the bracketed interval estimates are not stable (except for the 95% bracketed interval with 2 x 20% trim), and the recoding method does not result in bracketed intervals that are narrower than those produced using s_w.

$S_{recode6}$ did not perform as well or better than s_w. The difference between this recoding procedure and winsorizing was that this recoding procedure attempted to smooth the tails of the distribution when recoding values towards the median. This procedure did not treat outliers as extremely as winsorizing, and perhaps did not treat the outliers enough as in the case of $s_{recode5}$, $s_{recode4}$, $s_{recode3}$, and $s_{recode2}$.

Summary and Future Direction

The results of this study support the use of light trimming vs. heavy trimming. With the exception of $s_{recode6}$, 2 x 10% trimming resulted in more robust bracketed intervals according to Bradley's liberal or conservative criteria. The literature is divided with regards to heavy trimming (2 x 25% or 2 x 20%) and lightly trimming (2 x 5%, 2 x 10%). Further study could focus on lightly trimming (2 x 5% vs. 2 x 10%) to determine if lightly trimming improves coverage, especially for alpha levels less than .05, where the coverage rates start to break down. This would be an interesting study using s_w, since s_w performed well (met Bradley's liberal or conservative criterion for all distributions and alpha levels) in this study, however with an advantage when the 2 x 10% trim was used.

Some other starting points for further research could include investigation of different models of outliers (including asymmetric distributions, and highly/lightly contaminated samples), and different sample sizes. Looking into these variables could help to determine the conditions under which nominal alpha becomes inflated. Again, this would be an interesting study using s_w, since s_w performed well (met Bradley's liberal or conservative criterion for all distributions and alpha levels) in this study, however, performance was better for alpha of .05, vs. alpha of .01.

Finally, this study was focused on eight measures of dispersion. Lax (1985) identified over 150 different measures of dispersion. There are clearly many other measures of dispersion that could be included in a study such as this.

Study Limitations

While the limitations of this study were spelled out in Chapter 1, it is important to again focus on these limitations in light of the results of this study. This study had several limitations:

1. Symmetric distributions: These results cannot not be extrapolated to other symmetric distributions or to nonsymmetric distributions without further study.
2. Sample Size: A sample size of 30 was used in this study. The results are specific to this sample size and cannot be extrapolated to larger or smaller sample sizes.
3. Trim Percent: In this study, 2 x 10% and 2 x 20% symmetric trimming was used. The results cannot be extrapolated to other trim percentages or to nonsymmetric trimming.
4. Alpha Level: In this study an alpha levels of .05 and .01 were examined. Bracketed intervals may behave differently for different alpha levels. Thus the results cannot be extrapolated to alpha levels outside of this study.

APPENDIX

Sample Excel Calculations: Samples Drawn From Standard Normal Distribution; Outliers Created; 2x10% Trim; 95% Bracketed Interval; s_w= Measure of Dispersion

	A	B	C	D	E	F	G	H	I	J
1	Sample Trimmed Mean	Sample SD	Lower BI	Upper BI	BI Width	Pop Mean	Mean in BI?	LRE	Coverage	Median LRE
2	-0.223215	0.783986	-0.59339946	0.14696946	0.74036892	0	TRUE	1	0.9475000000	1
3	5.44E-02	0.743849	-0.296854872	0.405610072	0.702464943	0	TRUE	1		
4	-0.367159	0.522211	-0.613737889	-0.120580111	0.493157779	0	FALSE	1		
5	0.224238	0.86615	-0.184742862	0.633218862	0.817961724	0	TRUE	1		
6	2.94E-02	0.791756	-0.344428518	0.403278118	0.747706636	0	TRUE	1		
7	-6.71E-02	0.761351	-0.42660891	0.29238431	0.718993219	0	TRUE	1		
8	-9.70E-02	0.844347	-0.495699667	0.301672067	0.797371735	0	TRUE	1		
9	0.176746	0.845554	-0.222509792	0.576001792	0.798511583	0	TRUE	1		
10	7.27E-02	0.721595	-0.268068923	0.413380123	0.681449045	0	TRUE	1		
11	0.195986	0.665314	-0.11816362	0.51013562	0.62829924	0	TRUE	1		
12	0.116078	0.829861	-0.275767832	0.507923832	0.783691664	0	TRUE	1		
13	2.79E-02	0.863638	-0.37989684	0.43569264	0.815589479	0	TRUE	1		
14	-0.291417	0.705047	-0.624327847	0.041493847	0.665821693	0	TRUE	1		
15	-4.13E-02	0.836755	-0.436424958	0.353777158	0.790202116	0	TRUE	1		
16	-0.230601	0.746248	-0.582966237	0.121764237	0.704730475	0	TRUE	1		
17	1.74E-02	0.735268	-0.329762374	0.364598974	0.694361347	0	TRUE	1		
18	-0.115783	0.578421	-0.388903267	0.157337267	0.546240534	0	TRUE	1		
19	-9.40E-02	0.834545	-0.488082635	0.300032435	0.788115069	0	TRUE	1		
20	-0.192213	0.773378	-0.557388548	0.172962548	0.730351097	0	TRUE	1		

Sample Trimmed Mean: Pasted output from Fortran
Sample SD (Standard Deviation): Pasted output from Fortran
Lower BI (Bracketed Interval): Sample calculation for C2
=A2-(2.069*(1/0.8)*(B2/SQRT(30)))
Upper BI: Sample calculation for D2
=A2+(2.069*(1/0.8)*(B2/SQRT(30)))
BI Width: Sample calculation for E2
=D2-C2
Pop (Population) Mean: Population mean is zero for the Standard Normal distribution
Mean in BI (Does the population mean fall in the bracketed interval?): Sample calculation for G2
=IF(F2>=C2,F2<=D2,FALSE)
LRE: Sample calculation for H2
=E2/E2
Coverage: Sample calculation for I2
=(COUNTIF(G2:G10001,"True"))/10000
Median LRE: Sample calculation for J2
=MEDIAN(H2:H10001)

References

Blair, R. C. (1981). A reaction to "Consequences of failure to meet assumptions underlying the fixed effects Analysis of Variance and Covariance". *Review of Educational Research, 51*, 499-507.

Blair, R.C. (1987). RANGEN. Boca Raton: IBM Corporation

Blair, R. C., & Higgins, J. J. (1980a). A comparison of the power of the t test and Wilcoxon statistics when samples are drawnfrom a certain mixed normal distribution. *Evaluation Review, 4*, 645-656.

Blair, R. C., & Higgins, J. J. (1980b). A comparison of the power of the Wilcoxon's rank-sum statistic to that of the Student's t under various normal distributions. *Journal of Educational Statistics, 5*(309-355).

Blair, R. C., & Higgins, J. J. (1985). A comparison of the power of the paired samples t test to that of Wilcoxon's signed-ranks test under population shapes. *Psychological Bulletin, 97*, 119-128.

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

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

Bunner, J., & Sawilowsky, S. (2002). Alternatives to s_w in the bracketed interval of the trimmed mean. *Journal of Modern Applied Statistical Methods, 1*(1), 176-181.

Daniell, P. J. (1920). Observations weighted according to order. *American Journal of Mathematics, 42*(4), 222-236.

Glass, G.V., Peckman, P.D., & Sanders, J.R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research, 42*(30), 237-288.

Gross, A.M. (1976). Bracketed interval robustness with long tailed symmetric distributions. *Journal of the American Statistical Association, 71*(354), 409-416.

Hollander, M., & Wolfe, D. (1973). *Nonparametrical statistical methods*. New York: John Wiley & Sons.

James, G., & James, R. C. (1992). *Mathematic Dictionary* (5th ed.). New York: Van Nostrand Reinhold.

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford Press.

Lax, D. L. (1985). Robust estimators of scale: Finite-sample performance in long tailed symmetric distributions. *Journal of the American Statistical Association, 80*(391), 736-741.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*(1), 156-166.

Pearson, E. S., & Please, N. W. (1975). Relation between the shape of population distribution and robustness of four simple test statistics. *Biometrika, 62*(2), 223-241.

Rosenberger, J.L. & Gasko, M. (1983). Comparing locationestimators: trimmed means, medians, and trimean. In D.C. Hoaglin, F. Mostellar, and J.W.

Tukey (Eds.) *Understanding robust and exploratory data analysis,* Chapter 10. NY: Wiley.

Sawilowsky, S. S., & Hillman, S. B. (1992). Power of the independent samples t test under a prevalent psychometric distribution. *Journal of Consulting and Clinical Psychology, 60*, 240-243.

Sawilowsky, S.S. (2002). A measure of relative efficiency for location of a single sample. *Journal of Modern Applied Statistical Methods, 1(1), 52-60.*

Staudte, Jr., R.G., & Sheather, S.J. (1990). *Robust estimation and testing.* NY: Wiley.

Stigler, S. M. (1973). Simon Newcomb, Percy Daniell, and the history of robust estimation 1885-1920. *Journal of the American Statistical Association, 68*(344), 872-879.

Stigler, S. M. (1977). Do robust estimators work with real data. *Annals of Statistics, 5*(6), 1055-1098.

Tukey, J. W., & McLaughlin, D. H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/winsorization 1. *Sankhya, 25*, 331-352.

Vogt, W. P. (1993). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences.* Newbury Park: Sage Publications.

Wilcox, R. R. (1996). *Statistics for the social sciences*. San Diego: Academic Press.

Wilcox, R.R. (1997). *Introduction to Robust Estimation and Hypothesis Testing.* San Diego, CA: Academic Press.

ABSTRACT

FORMING A BRACKETED INTERVAL AROUND THE TRIMMED MEAN: ALTERNATIVES TO S_W

by

JENNIFER M. BUNNER

December 2003

Advisor: Dr. Shlomo Sawilowsky

Major: Evaluation and Research

Degree: Doctor of Philosophy

Nonnormally distributed data is common in practice and is often a concern to researchers due to the effects on statistical procedures such as measures of central tendency and dispersion. Robust methods were motivated by nonnormal distributions where outliers can be assumed to have contaminated the model, and are methods that are not affected by small deviations from normality. The trimmed mean ($\bar{X}_t$) is an example of a robust method, and is a robust estimator of the population mean that can be used when the underlying distribution of the data is normal. When working with the trimmed mean ($\bar{X}_t$), one may want to determine how accurately the sample trimmed mean estimates the population mean by forming a bracketed interval around the sample trimmed mean. In forming a bracketed interval around $\bar{X}_t$, the sample winsorized standard deviation (s_w) is commonly used as a robust measure of dispersion.

This study investigated alternatives to s_w in forming a bracketed interval around the sample trimmed mean. A Monte Carlo study was used to evaluate six

alternative measures of dispersion on ten distributions with an underlying symmetric structure using both a 2x10% and 2x20% symmetric trim. Resulting bracketed interval coverage and widths were compared to the performance of s_w. The results of this study support the use of s_w in calculating bracketed intervals around the sample trimmed mean. S_w was the only measure of dispersion with robust coverage rates across all distributions, alpha levels, and trim percentages tested. Generally bracketed interval widths were shorter for s_w. One measure of dispersion ($s_{recode1}$, a combination of trimming and winsorizing) resulted in shorter bracketed intervals using a 2x20% trim. However, coverage was robust for only 70% of the distributions tested. In general, the results of this study support light vs. heavy trimming.

AUTOBIOGRAPHICAL STATEMENT

JENNIFER M. BUNNER

Education:

2003 **Ph.D., Wayne State University**
Major: Evaluation and Research **Cognate:** Business Administration

Awarded the Graduate Professional Scholarship – 2000-2003

1998 **B.A., University of Rochester**
Major: Psychology (Magna Cum Laude, Highest Honors in Research)

Elected to Phi Beta Kappa, Golden Key National Honor Society, Psi Chi National Psychology Honor Society, and Deans List.

Professional Experience:

2001-2003 **Y&R Group-Wunderman, Dearborn, MI**
Manager of Campaign Analytics
Senior Analyst
Analyst

2000-2001 **Vector Research, Ann Arbor, MI**
Analyst

Peer Reviewed Publications:

1. **Bunner J**, Sawilowsky S. Alternatives to S_w in the Bracketed Interval of the Trimmed Mean. Journal of Modern Applied Statistical Methods 2002; 1: 176-181.

2. Yoshigi M, **Ettel JM**, Keller BB. Developmental Changes in Flow Wave Propagation Velocity in Embryonic Chick Vascular System. American Journal of Physiology 1997; 273: H1523-H1529.

3. **Ettel JM**, Yoshigi M, Keller BB. Fourier Analysis of Flow Wave Propagation Velocity in the Embryonic Chick Vascular System. Journal of the University of Rochester Medical Center 1997(Spring); 9: 51.