Probability of Coverage and Interval Length for Two-Group Techniques Assessing the Median and Trimmed Mean

S. Jonathan Mends-Cole

Walden University, sjmendscole@aim.com

Probability of Coverage and Interval Length for Two-Group Techniques Assessing the Median and Trimmed Mean

S. Jonathan Mends-Cole
Walden University

The purpose of the present study was to assess the probability of coverage and interval length of selected statistical techniques that have a higher finite sample breakdown point than the mean and appropriate levels of probability of coverage when using Bradley's (1978) criterion. The techniques were examined using real education and psychology datasets (Sawilowsky & Fahoome, 2003, Sawilowsky & Blair, 1992). Welch's test exhibited appropriate coverage for the smooth symmetric, mass at zero, digit preference, and extreme bimodal distributions. Yuen's technique performed well under an extreme bimodal distribution. Results concerning the Maritz-Jarrett and the McKean-Schrader techniques are also presented.

Key words: Trimmed mean, median, confidence interval, interval length, probability of coverage

Introduction

A researcher may want to know a range of values that may enclose the population parameter with a given level of confidence. A confidence interval provides a range of values that one can be $(1-\alpha)100\%$ confident the population parameter is enclosed (Sawilowsky & Fahoome, 2003, p. 200-201). The adequacy of a confidence interval is assessed through probability of coverage, $1-\hat{\alpha}$. Within education and psychology, statistical techniques have been assessed through Type I ($\hat{\alpha}$) and Type II ($\hat{\beta}$) error rates (e.g., Wilcox & Charlin, 1986, Wilcox, 1994, Luh & Guo, 2000, Wilcox, Kowalchuk, & Olejnik, 2002). Although many studies have examined techniques using error rates, some studies have examined the techniques using probability of coverage and interval length. Examples would include the studies done by Bonett and Price (2002) and Price and Bonett (2002).

S. Jonathan Mends-Cole is a member of the part-time faculty at Walden University. He can be reached at sjmendscole@aim.com or at P.O. Box 07285, Detroit, MI 48207.

Many situations motivate the use of confidence intervals. First, if no hypothesis is to be tested and one wants to know the range of plausible values for the population parameter, Knapp (1999) recommended using a confidence interval. Second, retrospective power analysis employs statistical power (a) following the statistical analysis, and (b) with a sample estimate of effect size. Statistical power is the *a priori* probability of detecting an effect if it exists, that is $\mu_1 - \mu_2 \neq 0$ (Wilcox, 1996; Zumbo & Hubley, 1998). The use of statistical power in retrospective power analysis is untenable (Zumbo & Hubley, 1998; Knapp, 1999, Hoenig & Heisey, 2001). Hoenig and Heisey (2001), and Wilkinson and Taskforce on Statistical Inference (1999) recommended using confidence intervals instead of retrospective power analysis. Confidence intervals provide an indication of statistical precision. The interval is more precise if the length is narrower (Hinkle, Wiersma, & Jurs, 1998, p. 225).

Confidence intervals also provide a test of the null hypothesis. Values outside the interval limits provide evidence for rejecting a range of null hypotheses (Hinkle, Wiersma, & Jurs, 1998, p. 224, Bonett & Price, 2002, p. 372). When comparing different statistical techniques or when applied at different alpha levels, narrower interval lengths imply greater

statistical power. Narrower interval lengths imply a smaller standard error. A smaller standard error implies that it will be more likely that the test statistic will be rejected.

Here, a study was done on the probabilities of coverage and interval length of techniques selected on the minimum number of sample values that can be altered, thus making the measure of central tendency arbitrarily large or small – the finite sample breakdown point (FSBP) (Wilcox, 2001, p. 149). An FSBP of 1/n is given for the mean. One outlier can make the mean arbitrarily large or small and the outlier can increase the variance. The outlier effect on the mean and variance affects the Type I and Type II error rates of statistical techniques that depend on the mean.

Consider the 0.05 alpha level. When sampling from skewed distributions where outliers occur frequently, the Type I error rate exceeds the nominal level (α) and the probability of coverage is below the nominal level ($1-\alpha$) (Wilcox, 2003). For example, Mends-cole (2006) found that probability of coverage for Students' t was less than 0.925 under inverse heteroscedasticity. The standard of 0.925 was adopted from Bradley's (1978) criterion and recommended by Bunner (2003): $1-0.5\alpha \leq 1-\hat{\alpha} \leq 1-1.5\alpha$. At the extremes of skewness ($\gamma_3 > 1.25$), probabilities of coverage for Welch's t were below 0.925. The results were similar to Algina, Oshima, and Lin (1994) and Luh and Guo (2000). Under similar conditions, the probability of coverage for Welch's technique was less than 0.925.

Preference was given to measures of central tendency with an FSBP that was higher than the mean. An FSBP of 0.20 is given for the 20% trimmed mean; the median has an FSBP of 0.50 (Wilcox, 2001, p. 149). Selected procedures for evaluating the trimmed mean and median included the confidence interval obtained by inverting Yuen's trimmed *t*-statistic, the Maritz-Jarrett median *z*-statistic (M-J), and the McKean-Schrader median *z*-statistic (M-S). Beyond testing a measure of central tendency with a high FSBP, the techniques have acceptable levels of Type I error rates, $0.025 \leq \hat{\alpha} \leq 0.075$. In Table 1, γ_3 denotes

skewness, γ_4 denotes kurtosis ($\gamma_4=0$ for a normal distribution), skewness and kurtosis specifications for Wilcox (1994) represent the second group data. The first group is sampled from a standard normal distribution. Skewness and kurtosis specifications for Bonett and Price (2002) represent the second group data. The first group is sampled from an exponential distribution for the first two rows and from a standard normal distribution from the third row. Under conditions of skewness, each technique exhibited minor bias in terms of Type I error rate and probability of coverage, $0.925 \leq 1-\hat{\alpha} \leq 0.975$.

Some questions arise when considering the studies presented in Table 1. (a) The techniques were recommended based on random numbers generated using mathematical functions. Results from the mathematical functions may not represent the samples observed in applied situations in education and psychology. To the extent that Monte Carlo samples represent applied situations, the results generalize to like situations (Sawilowsky & Fahoome, 2003, p. 443). (b) The techniques were recommended based on Type I and Type II error rates. The probability-coverage and interval-length are specifications of the confidence interval.

Method

Yuen's technique

Yuen's technique involves trimming and Winsorization to account for skewness. Trimming a group sample involves omitting a proportion of the largest scores and the same proportion of the smallest scores from the sample. Winsorization involves replacing a fixed proportion of the largest scores with the maximum score for the trimmed version of the same sample, and replacing an equivalent number of the smallest scores with the minimum score for the trimmed version of the same sample where [x] is the greatest integer $\leq$ x, $0 \leq \tau < 0.5$, and g=[τn], the Winsorized values (Z_i) for the scores (X_i) are given as follows: $Z = X_{(g+1)}$, if $X_i \leq X_{(g+1)}$; $Z = X_i$ if $X_{(g+1)} < X_i < X_{(n-g)}$; $Z = X_{(n-g)}$ if $X_i \geq X_{(n-g)}$. Wilcox (2003) suggested that 20% trimming is "a good choice for general use" (p. 251) and 20% trimming is applied here.

Table 1. Probability of Coverage, $\alpha = 0.05$, of Yuen's & Welch's Techniques Reported in the Literature

Technique	Author(s) (Date)	n_1	n_2	σ_2/σ_1	γ_3	γ_4	$1-\hat{\alpha}$
Yuen's	Wilcox (1994)	12	12	1	2.0	6.0	0.95
				1	3.9	42.2	0.95
	Luh &	12	24	4	6.2	111.0	0.95
	Guo (2000)			1/4	6.2	111.0	0.92
	Wilcox (1994)	40	12	1	2.0	6.0	0.95
				1	3.9	42.2	0.95
		80	20	1	2.0	6.0	0.94
				1	3.9	42.2	0.95
Maritz-	Wilcox &	11	11	1	0.0	6.0	0.97
Jarrett	Charlin (1986)			1	0.9	1.2	0.96
				1	2.0	6.0	0.96
		25	11	1	0.0	6.0	0.96
				1	0.9	1.2	0.95
				1	2.0	6.0	0.96
		25	19	1	0.0	6.0	0.96
				1	0.9	1.2	0.95
				1	2.0	6.0	0.96
		25	25	1	0.0	6.0	0.96
				1	0.9	1.2	0.95
				1	2.0	6.0	0.96
Maritz-	Bonett & Price	15	15	9	2	6	.95
Jarrett	(2002)			4.7	6	110	.97
				8	1.4	3	.94
McKean-	Bonett & Price	15	15	9	2	6	.96
Schrader	(2002)			4.7	6	110	.97
				8	1.4	3	.96

The technique is outlined as follows. (a) Where tau the sum of squared deviations from the Winsorized mean is given as SSD_{wi}, the Winsorized variance is estimated as $S_{wi}^2 = SSD_{wi}/(h_i - 1)$. The standard error of the trimmed mean is $S_{\bar{x}ti}^2 = S_{wi}^2/h_i$. The trimmed sample size is $h_i = n_i - 2(\tau_i)$. The degrees of freedom are calculated as follows.

$$df_{\bar{x}t1 - \bar{x}t2} =$$
$$[S_{\bar{x}t1}^2 + S_{\bar{x}t2}^2]^2 / [S_{\bar{x}t1}^2/(h_1 - 1) + S_{\bar{x}t2}^2/(h_2 - 1)] \quad (1)$$

The $1 - \alpha/2$ percentile of Student's t-distribution provides the critical value.

Maritz-Jarrett Technique

The equations for the M-J technique were provided in Wilcox (1996). Where p_{bi} is the probability that the value of a beta random variable is between $(i-1)/n$ and i/n, $i = 1,...,n$. The beta probability function depends on parameter values a and b; the parameter values for p_{bi} are given as $a = (n+1)/2$, $b = a$. The probability is obtained using the International Mathematical and Statistical Libraries (1998) function BETDF. The Harrell-Davis estimate of the population median is calculated as follows.

$$\hat{\theta} = \sum_{i=1}^{n} p_{bi} X_{(i)} \quad (2)$$

The variable $X_{(i)}$ is the ith ordered value of X_i. The Harrell-Davis estimate of the population median is a less biased estimate of the population median than the sample median (Wilcox, 1996, p. 73). The estimate of the variance of the median is given as follows.

$$S_{MJ}^2 = \sum_{i=1}^{n} p_{bi} X_{(i)}^2 - \hat{\theta}^2 \quad (3)$$

That is the variance of the median is difference between the probability of a beta random variable and the ordered value of X_i, less the square of the Harrell-Davis estimate of the population median.

McKean-Schrader Technique

The equations for the M-S technique were provided in Wilcox (2003, p. 134). The computations are given as follows as follows.

$$m = (n+1)/2 + z_{.995}\sqrt{n/4} \quad (4)$$

The estimate m is rounded to the nearest nonzero integer and $z_{.995}$ is the 99.5 percentile of the standard normal distribution. The estimate of the variance of the median is given as follows

$$S_{MS}^2 = [(X_{(n-m+1)} - X_{(m)})/2z_{.995}]^2 \quad (5)$$

The critical value for the M-J and M-S techniques is the $1 - \alpha/2$ percentile of the standard normal distribution. Both techniques provide confidence intervals using an estimate of the standard error of the median. Other methods of obtaining the confidence interval require the use of the bootstrap procedure (Wilcox, 1996, 2003). Modifications of the M-S variance have been recommended in Bonett and Price (2002) and Price and Bonett (2002). Yet, such modifications are not studied here.

The general form of the equation for the confidence interval for each statistic is given as follows.

$$D_{12} \mp \varphi_{1-\alpha/2} SE_D \quad (6)$$

Where D_{12} represents the difference between the trimmed means or between the medians; $\varphi_{1-\alpha/2}$ represents either the two-tailed critical z or t value of a test statistic and significance level $(\alpha/2)$, SE_D is the standard error for D_{12}.

The specifications for the equation for the confidence interval were outlined in the table below. In summary, the respective measures of central tendency have a higher FSBP than the mean. Yuen's method adjusts for skewness by trimming extreme scores. The method adjusts for heteroscedasticity by the manner in which the degrees of freedom are calculated. The Maritz-Jarret technique is based on an estimate of the median that is less biased than the sample median. Further, results by Bonett and Price (2002) show that the M-S technique maintained

Table 2. Equations for the Difference Estimate, Critical Value and Hypothesis for the Selected Techniques

Technique	Hypothesis	D_{12}	$\varphi_{1-\alpha/2}$	SE_D
Yuen's	$H_o : \mu_{t1} = \mu_{t2}$	$\overline{X}_{t1} - \overline{X}_{t2}$	$t_{1-\alpha/2}$	$SE_{\overline{x}t1 - \overline{x}t2} = \sqrt{S^2_{\overline{x}t1} + S^2_{\overline{x}t2}}$
Maritz-Jarrett	$H_o : \theta_1 = \theta_2$	$X_{M1} - X_{M2}$	$z_{1-\alpha/2}$	$SEMJ_{X_{M1} - X_{M2}} = \sqrt{S^2_{MJ1} + S^2_{MJ2}}$
McKean-Schrader	$H_o : \theta_1 = \theta_2$	$X_{M1} - X_{M2}$	$z_{1-\alpha/2}$	$SEMS_{X_{M1} - X_{M2}} = \sqrt{S^2_{MS1} + S^2_{MS2}}$

Table 3. Descriptive Information Pertaining to Eight Real World Distributions

Distribution	μ	σ	γ_3	γ_4
Mass at Zero (MZ)	12.92	4.42	-0.03	0.31
Extreme Asymmetry-Psychometric (EAP)	13.67	5.75	1.64	1.52
Extreme Asymmetry-Achievement (EAA)	24.5	5.79	-1.33	1.11
Extreme Bimodality (EB)	2.97	1.69	-0.08	-1.70
Multimodal & Lumpy (ML)	21.15	11.9	0.19	-1.20
Digit Preference (DP)	536.95	37.64	-0.07	-0.24
Smooth Symmetric (SS)	13.19	4.91	0.01	-0.34

Note. Adapted from "A More Realistic Look at the Robustness and Type II Error Properties of the t Test to Departures From Population Normality", by S. S. Sawilowsky and R. C. Blair, 1992, *Psychological Bulletin*, *2*, p. 353. Copyright 1992 by the American Psychological Association

good levels of probability of coverage under heteroscedasticity. As such, the methods were selected for consideration in the present study.

Random samples were obtained independently and with replacement using the IMSL routines: RNUND and RNSET subroutines. Each sample was obtained using the real world sampling distributions provided in Sawilowsky and Fahoome (2003). The probabilities of coverage for each interval were evaluated using samples from seven non-normal distributions that are prevalent in educational and psychological research (Micceri, 1986). The table below provides the means (μ), standard deviations (σ) and third and fourth moment estimates of skewness and kurtoses of the seven

of the eight distributions. For the purpose of study, the distribution serves as a proxy for the population. Estimates of interval-length and probability-coverage were obtained by sampling from the seven distributions. The kurtosis was adjusted so that the value for a normal distribution would be 0.00.

Sample size and standard deviation ratios of 1:1, 1:3, and 1:9 were specified. The group sizes were $n_i = 5$, 15, and 45. Sample size pairs included the following: (5,5), (5,15), (5,45), (15,15), (15,45), and (45,45). Sample size pairs were crossed with each level of heterogeneity - proportionately and disproportionately. Coverage-probabilities and interval-length were examined at the 0.05 alpha

level. As in other Monte Carlo studies, results at the 0.01 alpha level will require a larger sample size to achieve appropriate levels of Type I error.

Let k be the ratio of the standard deviation for the second group compared with the first; and X is the untransformed score; the variance for the second group was modified as follows.

$$X' = kX \qquad (7)$$

Multiplying a variable by k resulted in multiplying the measure of central tendency by a factor of k. However, the Monte Carlo study requires that the variance of the second group increase while the measures of central tendency remain equal. For k greater than one, the measure of central tendency was readjusted by subtracting out the population value of central tendency (η) as follows.

$$X'' = X' - (k-1)\eta \qquad (8)$$

That is, multiplying the scores in one group by a factor increased the measure of central tendency by the same factor. Further, the measures of central tendency were adjusted to their original value. One million repetitions were done. Note that the levels of skewness, size, variance, and effect under study represent a subset of conditions that may occur in an applied situation.

The method involved (1) generating n_i random samples per group, (2) modeling heterogeneity. (3) For Yuen's technique, one had to trim and Winsorize the sample values. (4) Evaluating Equation 6 to obtain the limits of the two-sided confidence interval. (5) The location relative efficiency (LRE) for the interval width (Sawilowsky, 2002) was obtained. Welch's technique was used to calculate the LRE. The confidence interval for Welch's technique uses a separate variance estimate of the standard error. Where s_i^2 is the variance for group i; and $s_{\bar{x}i}^2 = s_i^2/n_i$, the standard error is estimated as follows.

$$SE_{\bar{x}1-\bar{x}2} = \sqrt{s_{\bar{x}1}^2 + s_{\bar{x}2}^2} \qquad (9)$$

The degrees of freedom are calculated as follows.

$$df_{wlch} = \qquad (10)$$
$$\left[s_{\bar{x}1}^2 + s_{\bar{x}2}^2 \right]^2 \Big/ \left[(s_{\bar{x}1}^2)^2 \big/ (n_1 - 1) + (s_{\bar{x}2}^2)^2 \big/ (n_2 - 1) \right]$$

The interval length for Welch's technique was used to make comparisons with that of other intervals. The equation for the LRE is obtained as follows.

$$LRE = (UL_W - LL_W) \big/ (UL_{\{T\}} - LL_{\{T\}}) \quad (11)$$

The subscript W denotes Welch's technique and $\{T\}$ denotes either the use of Yuen's, M-J's, or M-S's technique. An LRE above one shows that the interval for the selected technique is narrower than the interval for Welch's technique; an LRE below one shows the reverse is true. Here, the standard that the comparison interval was 50% wider (narrower) than Welch's interval was adopted. (6) Estimates of probability of coverage and interval length were obtained over the number of repetitions. The average LRE served as the outcome measure for interval length comparisons. In summary, a random sample was obtained and the confidence interval was computed. The location relative efficiency was obtained and the coverage was totaled. After 1,000,000 repetitions, the average LRE and probability of coverage were obtained.

Results

Probability of Coverage

Probabilities of coverage for the M-S technique were less than 0.925 under more of the conditions than for the other techniques examined. This finding was observed for each distribution studied. Low probability of coverage occurred consistently for the inverse pairing of size and variance and for total samples sizes less than sixty, $N < 60$. However, low probability of coverage did not occur if the ratio of variances (σ_2 / σ_1) equal to

Table 4. Probabilities of Coverage for the Welch, Yuen, Maritz-Jarrett, & McKean-Schrader Techniques for Equal & Unequal Sample Sizes & Variances when Sampling is from a Smooth Symmetric Distribution

σ_2/σ_1	n_2/n_1	N	Welch	Yuen	Maritz-Jarrett	McKean-Schrader
1	1	10	0.956	0.961	0.955	0.885b
1	3	20	0.945	0.934	0.950	0.919b
1	9	50	0.943	0.919b	0.940	0.882b
1	1	30	0.950	0.950	0.956	0.970
1	3	60	0.949	0.946	0.951	0.963
1	1	90	0.950	0.950	0.954	0.953
3	1	10	0.948	0.945	0.941	0.867b
3	3	20	0.951	0.951	0.953	0.954
3	1/3	20	0.946	0.928	0.934	0.869b
3	9	50	0.948	0.941	0.953	0.932
3	1/9	50	0.949	0.935	0.930	0.859b
3	1	30	0.949	0.945	0.946	0.963
3	3	60	0.951	0.950	0.951	0.959
3	1/3	60	0.949	0.944	0.942	0.962
3	1	90	0.950	0.949	0.945	0.950
9	1	10	0.949	0.941	0.930	0.854b
9	3	20	0.949	0.945	0.941	0.957
9	1/3	20	0.950	0.945	0.929	0.853b
9	9	50	0.950	0.950	0.945	0.942
9	1/9	50	0.951	0.949	0.929	0.854b
9	1	30	0.950	0.945	0.939	0.958
9	3	60	0.950	0.948	0.942	0.944
9	1/3	60	0.949	0.945	0.938	0.959
9	1	90	0.950	0.948	0.941	0.945

a. $1-\hat{\alpha} > 0.975$
b. $1-\hat{\alpha} < 0.925$

Table 5. Probabilities of Coverage for the Welch, Yuen, Maritz-Jarrett, & McKean-Schrader Techniques for Equal & Unequal Sample Sizes & Variances when Sampling is from an Extreme Asymmetry-Achievement Distribution

σ_2/σ_1	n_2/n_1	N	Welch	Yuen	Maritz-Jarrett	McKean-Schrader
1	1	10	0.975	0.968	0.960	0.885b
1	3	20	0.927	0.943	0.957	0.921b
1	9	50	0.886b	0.897b	0.944	0.883b
1	1	30	0.954	0.963	0.965	0.973
1	3	60	0.939	0.945	0.959	0.966
1	1	90	0.951	0.955	0.960	0.962
3	1	10	0.899b	0.938	0.947	0.865b
3	3	20	0.951	0.956	0.961	0.954
3	1/3	20	0.881b	0.897b	0.936	0.865b
3	9	50	0.952	0.955	0.960	0.936
3	1/9	50	0.882b	0.904b	0.929	0.849b
3	1	30	0.928	0.930	0.954	0.964
3	3	60	0.949	0.952	0.958	0.964
3	1/3	60	0.923b	0.922b	0.948	0.959
3	1	90	0.944	0.942	0.952	0.954
9	1	10	0.882b	0.913b	0.930	0.842b
9	3	20	0.923b	0.925	0.949	0.955
9	1/3	20	0.883b	0.918b	0.927	0.841b
9	9	50	0.945	0.946	0.954	0.947
9	1/9	50	0.884b	0.927	0.927	0.839b
9	1	30	0.922b	0.920b	0.946	0.956
9	3	60	0.941	0.938	0.949	0.950
9	1/3	60	0.921b	0.921b	0.945	0.955
9	1	90	0.941	0.938	0.946	0.950

a. $1-\hat{\alpha} > 0.975$
b. $1-\hat{\alpha} < 0.925$

Table 6. Probabilities of Coverage for the Welch, Yuen, Maritz-Jarrett, & McKean-Schrader Techniques for Equal & Unequal Sample Sizes & Variances when Sampling is from a Extreme Bimodality Distribution

σ_2/σ_1	n_2/n_1	N	Welch	Yuen	Maritz-Jarrett	McKean-Schrader
1	1	10	0.980a	0.971	0.917b	0.648b
1	3	20	0.955	0.931	0.912b	0.698b
1	9	50	0.952	0.879b	0.913b	0.636b
1	1	30	0.950	0.946	0.901b	0.757b
1	3	60	0.948	0.933	0.905b	0.709b
1	1	90	0.949	0.948	0.916b	0.768b
3	1	10	0.954	0.942	0.868b	0.604b
3	3	20	0.958	0.952	0.866b	0.723b
3	1/3	20	0.952	0.903b	0.858b	0.602b
3	9	50	0.961	0.951	0.902b	0.770b
3	1/9	50	0.952	0.958	0.840b	0.598b
3	1	30	0.948	0.928	0.859b	0.714b
3	3	60	0.950	0.947	0.894b	0.781b
3	1/3	60	0.948	0.927	0.850b	0.698b
3	1	90	0.949	0.945	0.886b	0.778b
9	1	10	0.952	0.969	0.817b	0.595b
9	3	20	0.949	0.929	0.825b	0.671b
9	1/3	20	0.952	1.000a	0.813b	0.596b
9	9	50	0.953	0.948	0.856b	0.771b
9	1/9	50	0.952	1.000a	0.808b	0.597b
9	1	30	0.948	0.928	0.823b	0.671b
9	3	60	0.949	0.944	0.852b	0.770b
9	1/3	60	0.948	0.928	0.819b	0.672b
9	1	90	0.949	0.943	0.846b	0.766b

a. $1-\hat{\alpha} > 0.975$
b. $1-\hat{\alpha} < 0.925$

Table 7. Probabilities of Coverage for the Welch, Yuen, Maritz-Jarrett, & McKean-Schrader Techniques for Equal & Unequal Sample Sizes & Variances when Sampling is from a Multimodal & Lumpy Distribution

σ_2/σ_1	n_2/n_1	N	Welch	Yuen	Maritz-Jarrett	McKean-Schrader
1	1	10	0.955	0.947	0.919b	0.779b
1	3	20	0.931	0.903b	0.919b	0.833b
1	9	50	0.921b	0.846b	0.910b	0.815b
1	1	30	0.949	0.948	0.929	0.905b
1	3	60	0.948	0.933	0.930	0.912b
1	1	90	0.950	0.949	0.946	0.934
3	1	10	0.925	0.897b	0.898b	0.769b
3	3	20	0.950	0.945	0.918b	0.888b
3	1/3	20	0.922b	0.846b	0.890b	0.770b
3	9	50	0.943	0.932	0.933	0.898b
3	1/9	50	0.923b	0.856b	0.883b	0.758b
3	1	30	0.947	0.923b	0.914b	0.896b
3	3	60	0.950	0.948	0.935	0.924b
3	1/3	60	0.946	0.918b	0.907b	0.890b
3	1	90	0.949	0.945	0.933	0.925
9	1	10	0.923b	0.865b	0.879b	0.749b
9	3	20	0.947	0.920b	0.903b	0.885b
9	1/3	20	0.924b	0.869b	0.878b	0.749b
9	9	50	0.950	0.946	0.927	0.917b
9	1/9	50	0.924b	0.875b	0.877b	0.748b
9	1	30	0.947	0.918b	0.902b	0.884b
9	3	60	0.949	0.945	0.926	0.918b
9	1/3	60	0.947	0.918b	0.900b	0.882b
9	1	90	0.949	0.944	0.925	0.917b

a. $1-\hat{\alpha} > 0.975$
b. $1-\hat{\alpha} < 0.925$

Table 8. Probabilities of Coverage for the Welch, Yuen, Maritz-Jarrett, & McKean-Schrader Techniques for Equal & Unequal Sample Sizes & Variances when Sampling is from a Mass at Zero Distribution

σ_2/σ_1	n_2/n_1	N	Welch	Yuen	Maritz-Jarrett	McKean-Schrader
1	1	10	0.958	0.962	0.958	0.893b
1	3	20	0.948	0.936	0.953	0.923b
1	9	50	0.946	0.923b	0.944	0.888b
1	1	30	0.951	0.950	0.957	0.970
1	3	60	0.950	0.946	0.953	0.963
1	1	90	0.950	0.950	0.956	0.955
3	1	10	0.951	0.948	0.946	0.875b
3	3	20	0.953	0.952	0.953	0.954
3	1/3	20	0.949	0.932	0.937	0.874b
3	9	50	0.949	0.943	0.954	0.929
3	1/9	50	0.951	0.941	0.934	0.864b
3	1	30	0.950	0.945	0.945	0.962
3	3	60	0.950	0.949	0.951	0.955
3	1/3	60	0.950	0.943	0.940	0.959
3	1	90	0.950	0.947	0.947	0.939
9	1	10	0.952	0.945	0.934	0.859b
9	3	20	0.951	0.945	0.939	0.954
9	1/3	20	0.953	0.950	0.932	0.859b
9	9	50	0.951	0.949	0.945	0.937
9	1/9	50	0.954	0.953	0.932	0.860b
9	1	30	0.950	0.943	0.937	0.955
9	3	60	0.950	0.947	0.944	0.942
9	1/3	60	0.951	0.943	0.936	0.956
9	1	90	0.950	0.947	0.945	0.945

a. $1-\hat{\alpha} > 0.975$
b. $1-\hat{\alpha} < 0.925$

Table 9. Probabilities of Coverage for the Welch, Yuen, Maritz-Jarrett, & McKean-Schrader Techniques for Equal & Unequal Sample Sizes & Variances when Sampling is from an Extreme Asymmetry-Psychometric Distribution

σ_2/σ_1	n_2/n_1	N	Welch	Yuen	Maritz-Jarrett	McKean-Schrader
1	1	10	0.990a	0.985a	0.974	0.894b
1	3	20	0.955	0.984a	0.969	0.920b
1	9	50	0.893b	0.971	0.960	0.888b
1	1	30	0.962	0.988a	0.987a	0.983a
1	3	60	0.937	0.964	0.985a	0.980a
1	1	90	0.951	0.968	0.980a	0.973
3	1	10	0.907b	0.984a	0.969	0.888b
3	3	20	0.952	0.962	0.963	0.941
3	1/3	20	0.878b	0.973	0.967	0.896b
3	9	50	0.968	0.972	0.932	0.864b
3	1/9	50	0.878b	0.972	0.964	0.872b
3	1	30	0.920b	0.913b	0.972	0.974
3	3	60	0.949	0.952	0.937	0.935
3	1/3	60	0.913b	0.877b	0.953	0.953
3	1	90	0.940	0.918b	0.882b	0.845b
9	1	10	0.876b	0.984a	0.963	0.866b
9	3	20	0.912b	0.900b	0.937	0.930
9	1/3	20	0.878b	0.985a	0.962	0.863b
9	9	50	0.940	0.933	0.848b	0.764b
9	1/9	50	0.881b	0.990a	0.960	0.858b
9	1	30	0.911b	0.867b	0.927	0.934
9	3	60	0.937	0.905b	0.813b	0.762b
9	1/3	60	0.911b	0.866b	0.921b	0.929
9	1	90	0.936	0.902b	0.798b	0.739b

a. $1-\hat{\alpha} > 0.975$
b. $1-\hat{\alpha} < 0.925$

Table10. Probabilities of Coverage for the Welch, Yuen, Maritz-Jarrett, & McKean-Schrader Techniques for Equal & Unequal Sample Sizes & Variances when Sampling is from a Digit Preference Distribution

σ_2/σ_1	n_2/n_1	N	Welch	Yuen	Maritz-Jarrett	McKean-Schrader
1	1	10	0.955	0.960	0.955	0.888b
1	3	20	0.945	0.935	0.951	0.922b
1	9	50	0.943	0.920b	0.941	0.887b
1	1	30	0.951	0.950	0.958	0.973
1	3	60	0.949	0.946	0.955	0.967
1	1	90	0.950	0.949	0.958	0.966
3	1	10	0.948	0.945	0.941	0.869b
3	3	20	0.952	0.951	0.954	0.956
3	1/3	20	0.947	0.928	0.934	0.871b
3	9	50	0.948	0.941	0.952	0.935
3	1/9	50	0.949	0.934	0.931	0.859b
3	1	30	0.949	0.945	0.948	0.965
3	3	60	0.950	0.949	0.947	0.958
3	1/3	60	0.949	0.945	0.943	0.962
3	1	90	0.950	0.948	0.938	0.946
9	1	10	0.949	0.938	0.930	0.855b
9	3	20	0.949	0.946	0.942	0.959
9	1/3	20	0.950	0.941	0.929	0.855b
9	9	50	0.950	0.949	0.932	0.934
9	1/9	50	0.951	0.944	0.928	0.853b
9	1	30	0.950	0.945	0.940	0.961
9	3	60	0.950	0.948	0.927	0.936
9	1/3	60	0.950	0.945	0.939	0.960
9	1	90	0.950	0.947	0.925	0.933

a. $1-\hat{\alpha} > 0.975$
b. $1-\hat{\alpha} < 0.925$

Table 11. Length Ratios for the Yuen, Maritz-Jarrett, & McKean-Schrader Techniques Compared with Welch's Technique for Equal & Unequal Sample Sizes & Variances when Sampling is from a Smooth Symmetric Distribution

σ_2/σ_1	n_2/n_1	N	Yuen	Maritz-Jarrett	McKean-Schrader
1	1	10	0.803	0.882	1.140b
1	3	20	0.904	0.875	1.002b
1	9	50	0.946b	0.941	1.157b
1	1	30	0.906	0.782	0.712
1	3	60	0.916	0.794	0.732
1	1	90	0.913	0.760	0.743
3	1	10	0.856	0.955	1.230b
3	3	20	0.890	0.812	0.794
3	1/3	20	0.922	0.994	1.245b
3	9	50	0.902	0.804	0.888
3	1/9	50	0.894	1.020	1.301b
3	1	30	0.917	0.823	0.741
3	3	60	0.912	0.770	0.738
3	1/3	60	0.919	0.843	0.756
3	1	90	0.916	0.783	0.763
9	1	10	0.868	1.021	1.312b
9	3	20	0.916	0.843	0.763
9	1/3	20	0.853	1.031	1.320b
9	9	50	0.910	0.784	0.786
9	1/9	50	0.834	1.035	1.329b
9	1	30	0.918	0.851	0.759
9	3	60	0.916	0.792	0.768
9	1/3	60	0.918	0.855	0.761
9	1	90	0.916	0.796	0.774

a. $1 - \hat{\alpha} > 0.975$
b. $1 - \hat{\alpha} < 0.925$

Table 12. Length Ratios for the Yuen, Maritz-Jarrett, & McKean-Schrader Techniques Compared with Welch's Technique for Equal & Unequal Sample Sizes & Variances when Sampling is from an Extreme Asymmetry-Achievement Distribution

σ_2/σ_1	n_2/n_1	N	Yuen	Maritz-Jarrett	McKean-Schrader
1	1	10	0.895	0.898	1.207b
1	3	20	0.999	0.893	1.046b
1	9	50	1.069b	0.941	1.188b
1	1	30	0.988	0.854	0.773
1	3	60	1.008	0.879	0.806
1	1	90	0.995	0.867	0.839
3	1	10	0.954	0.948	1.271b
3	3	20	0.974	0.873	0.870
3	1/3	20	1.053b	0.979	1.265b
3	9	50	0.982	0.863	0.972
3	1/9	50	1.038b	1.007	1.331b
3	1	30	1.020	0.904	0.810
3	3	60	0.995	0.869	0.825
3	1/3	60	1.038b	0.936	0.836
3	1	90	1.007	0.899	0.869
9	1	10	0.986b	1.006	1.346b
9	3	20	1.026	0.929	0.845
9	1/3	20	0.981b	1.019	1.355b
9	9	50	0.993	0.888	0.894
9	1/9	50	0.952	1.025	1.370b
9	1	30	1.040b	0.946	0.840
9	3	60	1.011	0.910	0.877
9	1/3	60	1.041b	0.953	0.845
9	1	90	1.013	0.920	0.888

a. $1-\hat{\alpha} > 0.975$
b. $1-\hat{\alpha} < 0.925$

Table 3. Length Ratios for the Yuen, Maritz-Jarrett, & McKean-Schrader Techniques Compared with Welch's Technique for Equal & Unequal Sample Sizes & Variances when Sampling is from a Extreme Bimodality Distribution

σ_2/σ_1	n_2/n_1	N	Yuen	Maritz-Jarrett	McKean-Schrader
1	1	10	0.564	0.746b	1.281b
1	3	20	0.631	0.673b	1.043b
1	9	50	0.650b	0.725b	1.124b
1	1	30	0.623	0.463b	0.626b
1	3	60	0.622	0.428b	0.563b
1	1	90	0.608	0.341b	0.430b
3	1	10	0.594	0.830b	1.408b
3	3	20	0.617	0.527b	0.717b
3	1/3	20	0.608b	0.850b	1.405b
3	9	50	0.623	0.471b	0.626b
3	1/9	50	0.575	0.872b	1.442b
3	1	30	0.630	0.495b	0.648b
3	3	60	0.613	0.377b	0.471b
3	1/3	60	0.633	0.498b	0.644b
3	1	90	0.608	0.363b	0.442b
9	1	10	0.568	0.891b	1.499b
9	3	20	0.631	0.519b	0.668b
9	1/3	20	0.554a	0.895b	1.503b
9	9	50	0.610	0.392b	0.471b
9	1/9	50	0.547a	0.897b	1.509b
9	1	30	0.629	0.517b	0.662b
9	3	60	0.608	0.382b	0.453b
9	1/3	60	0.629	0.519b	0.662b
9	1	90	0.607	0.383b	0.450b

a. $1-\hat{\alpha} > 0.975$
b. $1-\hat{\alpha} < 0.925$

Table 14. Length Ratios for the Yuen, Maritz-Jarrett, & McKean-Schrader Techniques Compared with Welch's Technique for Equal & Unequal Sample Sizes & Variances when Sampling is from a Multimodal & Lumpy Distribution

σ_2/σ_1	n_2/n_1	N	Yuen	Maritz-Jarrett	McKean-Schrader
1	1	10	0.715	0.839b	1.173b
1	3	20	0.802b	0.790b	0.995b
1	9	50	0.884b	0.862b	1.115b
1	1	30	0.753	0.617	0.658b
1	3	60	0.758	0.607	0.634b
1	1	90	0.747	0.538	0.550
3	1	10	0.800b	0.920b	1.273b
3	3	20	0.750	0.676b	0.740b
3	1/3	20	0.911b	0.949b	1.278b
3	9	50	0.768	0.651	0.734b
3	1/9	50	0.938b	0.978b	1.324b
3	1	30	0.763b	0.658b	0.682b
3	3	60	0.748	0.566	0.579b
3	1/3	60	0.767b	0.674b	0.686b
3	1	90	0.746	0.564	0.565
9	1	10	0.880b	0.988b	1.352b
9	3	20	0.765b	0.686b	0.701b
9	1/3	20	0.913b	0.996b	1.358b
9	9	50	0.748	0.586	0.593b
9	1/9	50	0.893b	1.001b	1.365b
9	1	30	0.767b	0.688b	0.696b
9	3	60	0.746	0.580	0.574b
9	1/3	60	0.767b	0.691b	0.697b
9	1	90	0.745	0.582	0.573b

a. $1-\hat{\alpha} > 0.975$
b. $1-\hat{\alpha} < 0.925$

Table 15. Length Ratios for the Yuen, Maritz-Jarrett, & McKean-Schrader Techniques Compared with Welch's Technique for Equal & Unequal Sample Sizes & Variances when Sampling is from a Mass at Zero Distribution

σ_2/σ_1	n_2/n_1	N	Yuen	Maritz-Jarrett	McKean-Schrader
1	1	10	0.834	0.892	1.143b
1	3	20	0.937	0.889	1.015b
1	9	50	0.974b	0.949	1.162b
1	1	30	0.948	0.818	0.743
1	3	60	0.958	0.829	0.765
1	1	90	0.956	0.797	0.781
3	1	10	0.882	0.960	1.228b
3	3	20	0.929	0.840	0.822
3	1/3	20	0.939	0.998	1.246b
3	9	50	0.940	0.828	0.914
3	1/9	50	0.904	1.024	1.301b
3	1	30	0.959	0.859	0.772
3	3	60	0.955	0.807	0.775
3	1/3	60	0.962	0.878	0.788
3	1	90	0.958	0.821	0.803
9	1	10	0.882	1.024	1.311b
9	3	20	0.957	0.878	0.795
9	1/3	20	0.864	1.035	1.320b
9	9	50	0.952	0.819	0.825
9	1/9	50	0.847	1.038	1.329b
9	1	30	0.961	0.887	0.791
9	3	60	0.959	0.830	0.809
9	1/3	60	0.960	0.891	0.793
9	1	90	0.959	0.835	0.817

a. $1 - \hat{\alpha} > 0.975$
b. $1 - \hat{\alpha} < 0.925$

Table 16. Length Ratios for the Yuen, Maritz-Jarrett, & McKean-Schrader Techniques Compared with Welch's Technique for Equal & Unequal Sample Sizes & Variances when Sampling is from an Extreme Asymmetry-Psychometric Distribution

σ_2/σ_1	n_2/n_1	N	Yuen	Maritz-Jarrett	McKean-Schrader
1	1	10	1.008a	0.894	1.277b
1	3	20	1.156a	0.935	1.150b
1	9	50	1.265	0.977	1.328b
1	1	30	1.225a	1.047a	0.898a
1	3	60	1.253	1.151a	1.026a
1	1	90	1.175	1.255a	1.283
3	1	10	1.080a	0.938	1.336b
3	3	20	1.171	0.997	1.027
3	1/3	20	1.238	0.975	1.329b
3	9	50	1.117	1.008	1.256b
3	1/9	50	1.243	0.997	1.405b
3	1	30	1.350b	1.147	0.979
3	3	60	1.191	1.197	1.148
3	1/3	60	1.456b	1.254	1.079
3	1	90	1.239b	1.360b	1.404b
9	1	10	1.164a	0.987	1.406b
9	3	20	1.377b	1.178	1.073
9	1/3	20	1.171a	1.001	1.415b
9	9	50	1.177	1.216b	1.378b
9	1/9	50	1.131a	1.008	1.439b
9	1	30	1.508b	1.270	1.083
9	3	60	1.268b	1.391b	1.412b
9	1/3	60	1.544b	1.307b	1.120
9	1	90	1.293b	1.488b	1.548b

a. $1-\hat{\alpha} > 0.975$
b. $1-\hat{\alpha} < 0.925$

Table 17. Length Ratios for the Yuen, Maritz-Jarrett, & McKean-Schrader Techniques Compared with Welch's Technique for Equal & Unequal Sample Sizes & Variances when Sampling is from a Digit Preference Distribution

σ_2/σ_1	n_2/n_1	N	Yuen	Maritz-Jarrett	McKean-Schrader
1	1	10	0.812	0.886	1.139b
1	3	20	0.915	0.882	1.002b
1	9	50	0.964b	0.950	1.157b
1	1	30	0.915	0.797	0.716
1	3	60	0.925	0.815	0.739
1	1	90	0.919	0.798	0.756
3	1	10	0.872	0.958	1.227b
3	3	20	0.900	0.824	0.797
3	1/3	20	0.953	0.998	1.243b
3	9	50	0.912	0.827	0.896
3	1/9	50	0.945	1.025	1.298b
3	1	30	0.927	0.839	0.745
3	3	60	0.919	0.803	0.749
3	1/3	60	0.930	0.861	0.760
3	1	90	0.922	0.825	0.776
9	1	10	0.905	1.024	1.309b
9	3	20	0.926	0.859	0.767
9	1/3	20	0.905	1.035	1.317b
9	9	50	0.917	0.822	0.798
9	1/9	50	0.881	1.039	1.326b
9	1	30	0.929	0.869	0.763
9	3	60	0.923	0.834	0.781
9	1/3	60	0.929	0.873	0.766
9	1	90	0.923	0.841	0.788

a. $1-\hat{\alpha} > 0.975$
b. $1-\hat{\alpha} < 0.925$

nine with the ratio of sample sizes (n_2/n_1) equal to three or one, or for total sample sizes greater than sixty ($N \geq 60$). Results were presented in the tables. Where absolute kurtosis was above 1.15, the probability of coverage for each technique was less than 0.925. Low probability of coverage occurred consistently for the inverse pairing of size and variance and for total sample sizes less than sixty ($N < 60$).

Location Relative Efficiency

The conditions of the inverse pairing of size and variance and total sample sizes less than 60 that resulted in probability of coverage less than 0.925 also results in *LRE*s at or above 1.25 using the M-S technique. That is, the interval length for the M-S technique was 80% of the length for Welch's technique. The results were presented in the tables below. If absolute kurtosis was less than 1.15, the interval length for Yuen's technique and the interval length for the M-J technique approximated the length for Welch's technique. The length ratios were within the following range: 0.75-1.10. Further, the M-J technique displayed appropriate levels of coverage for each size and variance pairing.

If absolute kurtosis exceeded 1.15, specified results were observed for the EB distribution and for the multimodal lumpy distribution. (a) Under an EB distribution, the following was observed. The length for Yuen's technique exceeded that for Welch's technique by 50% or more. Length ratios for Yuen's technique were below 0.66. Yuen's technique also displayed appropriate levels of coverage for more of the size and variance conditions than either the Maritz-Jarrett or McKean-Schrader techniques. (b) Under a multimodal lumpy distribution, the following was observed. If total sample sizes exceeded 50, were equal or were proportional to variances, the width for the M-J technique was 50% wider than for Welch's technique. *LRE*s for the M-J technique were less than 0.66. On the other hand, the interval lengths for Yuen's technique approximated that for Welch's technique. *LRE*s were within the range 0.715-0.940. For both Yuen's and the M-J techniques, the probability of coverage was still less than 0.925 under several size and variance conditions. Low probability of coverage occurred consistently with the inverse pairing of size and variance and for total sample sizes less than 60.

Conclusion

The purpose of the present study was to assess the probability of coverage and interval length of selected statistical techniques that have a higher FSBP than the mean and appropriate levels of probability of coverage when using Bradley's (1978) criterion. The techniques were examined using real education and psychology datasets (Sawilowsky & Fahoome, 2003, Sawilowsky & Blair, 1992). Welch's test appears to be robust to minor violations involving heteroscedasticity. Welch's test exhibited appropriate coverage for the smooth symmetric, mass at zero, digit preference, and extreme bimodal distributions. In general, the M-S technique exhibits narrow interval lengths that do not provide accurate coverage. The M-J technique does not perform well when kurtosis is at or below −1.25. However, it does perform well otherwise. Yuen's technique does perform well when kurtosis is below −1.25. However, its performance approximated that of the M-J technique under an ML distribution. A tradeoff was observed between coverage and length for Yuen's technique. Adequate coverage is often observed with wider interval lengths and vice versa.

References

References marked with an asterisk were included in Table 1.

Bonett, D. G., & Price, R. M. (2002). Statistical inference for a linear function of medians: Confidence intervals, hypothesis testing, and sample size requirements. *Psychological Methods*, *7*, 370-383.

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

Bunner, J. M. (2003). *Forming a bracketed interval around the trimmed mean: Alternatives to Sw* (Doctoral Dissertation, Wayne State University, 2003). Dissertation Abstracts International, *64*, 6147.

Cohen, J. (1994). The earth is round (p<.05). *American Psychologist, 49*, 997-1003.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1998). *Applied statistics for the behavioral sciences*. Boston, MA: Houghton Mifflin Company.

Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. *American Statistician, 55*, 19-23.

International Mathematical and Statistical Libraries (1998). *IMSL F90 MP library* 3.0. Incline Village, NV: Lahey Computer Systems, Inc.

Knapp, T. R. (1998). Comments on statistical significance testing articles. *Research in the Schools, 5*, 39-41.

Knapp, T. R., & Sawilowsky, S. S. (2001). Constructive criticism of methodological and editorial practices. *Journal of Experimental Education, 70*, 65-79.

Lahey Computer Systems. (1995-2000). Essential lahey FORTRAN 90. Incline Village, NV: Lahey Computer Systems, Inc.

*Luh, W., & Guo, J. (2000). Johnson' transformation two-sample trimmed t and its bootstrap method for heterogeneity and non-normality. *Journal of Applied Statistics, 27*, 965-973.

Mends-cole, S. J. (2006). Coverage & interval length for Welch's and Yuen's procedures for shift in location and change in scale for (Un)equal sample sizes (Doctoral Dissertation, Wayne State University, 2006). Dissertation Abstracts International.

Price, R. M., & Bonett, D. G. (2002). Distribution-free confidence intervals for difference and ratio of medians. *Journal of Statistical Computation and Simulation, 72,* 119-124.

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

Sawilowsky, S. S., & Fahoome, G. (2003). *Statistics through Monte Carlo experimentation with FORTRAN*. Oak Park, MI: JMASM, Inc.

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

*Wilcox, R. R. (1994). Some results on the Tukey-McLaughlin and Yuen methods for trimmed means when distributions are skewed. *Biometrical Journal, 3*, 259-273.

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

Wilcox, R. R. (2001). *Fundamentals of modern statistical methods*. New York, NY: Springer-Verlag.

Wilcox, R. R. (2003). *Applying contemporary statistical techniques*. San Diego, CA: Academic Press.

*Wilcox, R. R., & Charlin, V. L. (1986). Comparing medians: A Monte Carlo study. *Journal of Educational Statistics, 11*, 263-274.

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist. 54*,594-604.

Zumbo, B. D., & Hubley, A. M. (1998). A note on misconceptions concerning prospective and retrospective power. *The Statistician, 47*, 385-388.