DigitalCommons @WayneState

Journal of Modern Applied Statistical Methods

Volume 6 | Issue 2 Article 8

11-1-2007

Probability Coverage and Interval Length for Welch's and Yuen's Techniques: Shift in Location, Change in Scale, and (Un)Equal Sizes

S. Jonathan Mends-Cole
Walden University, sjmendscole@aim.com

Journal of Modern Applied Statistical Methods
November, 2007, Vol. 6, No. 2, 421- 442

1538 – 9472/07/$95.00

Probability Coverage and Interval Length for Welch's and Yuen's Techniques: Shift in Location, Change in Scale, and (Un)Equal Sizes

S. Jonathan Mends-cole
Walden University

Coverage for Welch's technique was less than the confidence-level when size was inversely proportional to variance and skewness was extreme. Under negative kurtosis, coverage for Yuen's technique was attenuated. Under skewness and heteroscedasticity, coverage for Yuen's technique was more accurate than Welch's technique.

Key words: Yuen's procedure, Welch's procedure, confidence interval, interval length, probability coverage, Monte Carlo simulation

Introduction

When assessing how well the sample effect ($\overline{X}_1 - \overline{X}_2$) estimates the population effect ($\mu_1 - \mu_2$), a confidence interval is the appropriate statistical technique. The interval-length conveys the magnitude of the standard error of the effect. When comparing intervals for measuring an effect, wider interval-lengths imply greater standard errors. The confidence-level expresses the long-run probability that the limits include the population parameter.

The use of confidence intervals has been strongly suggested in some disciplines (Cohen, 1994; Wilkinson & Task Force on Statistical Inference, 1999). Some spurious reasons include (a) they provide statistical inference without specifying an a priori threshold and (b) it is presumed that confidence intervals provide a degree of certainty about the population parameter that hypothesis tests do not. However, Sawilowsky (2003) was opposed to (a) as being contrary to the principles of the scientific method, and noted that the Type I and Type II probabilities of hypothesis tests are the same as for confidence intervals.

Type I and Type II errors do apply to confidence intervals as follows.

1. Is zero truly within the interval yet the interval does not enclose zero (Type I error)?

2. Is zero not truly within the interval yet the interval does enclose zero (Type II error)?

Monte Carlo simulations have been used to assess the extent to which the Type I and Type II error rates deviate from the α and β levels. Magnitudes of interval-length and probability-coverage $(1-\hat{\alpha})$ serve as criteria concerning the appropriateness of confidence intervals. The traditional test for bi-group comparisons is the independent samples t-test. The calculation of the confidence interval for the mean difference is outlined as follows. Where n_i is the sample size for group i, $\overline{X}_i$ is the mean for group i, and X_{ji} is the jth observation for group i, the standard error of the effect is given as follows:

$$SE_{\bar{x}1-\bar{x}2} = \sqrt{S^2(n_1^{-1}+n_2^{-1})} \qquad (1)$$

$$S^2 = \left[\sum\left(X_{j1}-\overline{X}_1\right)^2 + \sum\left(X_{j2}-\overline{X}_2\right)^2\right] \Big/ (n_1+n_2-2) \qquad (2)$$

S. Jonathan Mends-cole is an Instructor at Walden University, sjmendscole@aim.com. Mail: P.O. Box 07285, Detroit, MI 48207.

Where $t_{1-\alpha/2}$ refers to the critical value of the test distribution with $n_1 + n_2 - 2$ degrees of freedom, the confidence interval is:

$$\overline{X}_1 - \overline{X}_2 \mp t_{1-\alpha/2} SE_{\bar{x}1-\bar{x}2} \qquad (3)$$

Along with the assumptions that observations were randomly sampled from defined populations and that the samples were independent, some assumptions of parametric tests are homoscedasticity and normality (Wilcox, 1996). When heteroscedasticity and skewness are present in data, the error rates for a technique are inaccurate.

Violations of Parametric Test Assumptions

Skewness

Samples from skewed populations occur with some frequency as observed by Blair (1981) and Micceri (1989). Specifically, Micceri (1989) surveyed 440 published data sets. The p-value of the Kolmogorov-Smirnov test showed the distributions of each data set to be significantly different from a normal distribution ($p < .01$). Monte Carlo Type I error results (Sawilowsky & Blair, 1992) suggested that the probability-coverage would be greater than $1-\alpha$ for skewed distributions, i.e., for skewness ranging from 1.25 to 1.75. Setting the alpha level at 0.05, if Type I error rate is less than 0.05, probability-coverage is greater than 0.95. Sawilowsky and Blair observed that the independent samples t-test was robust: (a) if the test was two tailed rather than one tailed, (b) if sample sizes were about equal, and (c) if sample sizes were 25 or more.

Heteroscedasticity

Usually, when group means differ, group variances also differ (Sawilowsky & Blair, 1992, p. 358; Wilcox, 1996, p. 149). Why is heteroscedasticity likely to occur? Edwards (1972) attributed it to the absence of random assignment. If the variable for the treatment group exhibited greater variation before the application of the treatment after applying the treatment the difference is likely to remain unchanged. Another possibility is the multiplicative effect of the treatment. That is, if prior to the application of the treatment, $\sigma_2 / \sigma_1 < 2.0$, but after applying the treatment $\sigma_2 / \sigma_1 \geq 2.0$, the treatment may have acted multiplicatively to increase the variance.

Skewness & Heteroscedasticity

Heteroscedasticity has different effects on probability-coverage (Algina, Oshima, & Lin, 1994; Penfield, 1994). (a) If sizes are equal, the effect on probability-coverage is negligible, i.e., $0.925 \leq 1-\hat{\alpha} \leq 0.975$. (b) Small group sizes, e.g., $(n_1, n_2) = (5,15)$, skewness, and proportional heteroscedasticity augment probability-coverage (Penfield, 1994). (c) Small sizes, extreme skewness, and disproportional heteroscedasticity attenuate probability-coverage. If the confidence level was set at 0.95, the t-test displayed coverage-probabilities of 0.90 or less (Algina, Oshima, & Lin, 1994; Penfield, 1994). Although increasing sample sizes decreases the magnitude of separation between the Type I error rate and alpha level, Bradley (1978) observed that group sample sizes as large as 1,024 were needed for the independent samples t-test to maintain a 0.01 Type I error rate, if the application of the treatment increases the variance, heteroscedasticity increases interval-length. The larger group variance increases the standard error, thereby increasing the interval-length.

Use of Transformations

Using transformations to remedy the error rate problems of skewness and heteroscedasticity is problematic. The interpretation of statistical significance for the transformed scale no longer holds for the untransformed scale (Games, 1983). Yet, the untransformed scale was selected based upon an underlying rationale for doing the study.

Welch's and Yuen's Techniques

Both Welch's and Yuen's techniques have been recommended for amending the Type I and Type II error rate problems resulting from heteroscedasticity and skewness (Wilcox, 1996). 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

$$SE_{\bar{x}1-\bar{x}2} = \sqrt{s_{\bar{x}1}^2 + s_{\bar{x}2}^2} \qquad (4)$$

The degrees of freedom are calculated as

$$df_{wlch} = (s_{\bar{x}1}^2 + s_{\bar{x}2}^2)^2 \Big/ \left((s_{\bar{x}1}^2)^2 / (n_1 - 1) + (s_{\bar{x}2}^2)^2 / (n_2 - 1) \right) \qquad (5)$$

Yuen's technique assesses the difference between the trimmed means. The technique is outlined as follows. Trimming a group sample involves omitting a fixed proportion of the largest scores and an equivalent number 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. Wilcox (2003) suggested that 20% trimming is "a good choice for general use" (p. 251). (a) Where tau (τ_i) is the integer portion of $0.20(n_i)$, the trimmed sample size is $h_i = n_i - 2(\tau_i)$. The trimmed mean ($\overline{X}_{ti}$) is the mean of observations for the trimmed sample size. (b) The Winsorized mean ($\overline{X}_{wi}$) is the mean of observations for the Winsorized sample. The Winsorized sum of squared deviations is estimated as

$$SSD_{wi} = (\tau + 1)[(X_{(\tau+1)} - \overline{X}_w)^2 + (X_{(n-\tau)} - \overline{X}_w)^2] + \sum_{i=\tau+2}^{n-\tau-1} (X_{(i)} - \overline{X}_w)^2 \qquad (6)$$

Note that the subscripts in parentheses, e.g., ($\tau + 1$), ($n - \tau$), and (*i*) represent the ascending order of the *X* values. (c) Where the Winsorized variance is estimated as $S_{wi}^2 = SSD_{wi} / (h_i - 1)$ and the standard error of the trimmed mean is $S_{\bar{x}ti}^2 = S_{wi}^2 / h_i$, the standard error of the effect is estimated by:

$$SE_{\bar{x}t1-\bar{x}t2} = \sqrt{S_{\bar{x}t1}^2 + S_{\bar{x}t2}^2} \qquad (7)$$

The degrees of freedom is calculated as follows.

$$df_{yuen} = (S_{\bar{x}t1}^2 + S_{\bar{x}t2}^2) \Big/ (S_{\bar{x}t1}^2 / (h_1 - 1) + S_{\bar{x}t2}^2 / (h_2 - 1)) \qquad (8)$$

The confidence interval of trimmed means for bi-groups (Wilcox, 1996) is:

$$(\overline{X}_{t1} - \overline{X}_{t2}) \mp t_{1-\alpha/2} SE_{\bar{x}t1-\bar{x}t2} \qquad (9)$$

Welch's and Yuen's techniques exhibited appropriate coverage for extreme skewness, and homoscedasticity, i.e., $0.925 \le 1 - \hat{\alpha} \le 0.975$ (Algina et al., 1994; Wilcox, 1994). Under conditions of skewness and disproportional heteroscedasticity, Welch's coverage was less than 0.925 (Luh & Guo, 2000). Yuen's coverage was less than the confidence-level but to a lesser extent than Welch's technique was, i.e., $1 - \hat{\alpha} = 0.92$ versus 0.85. The probabilities of coverage were outlined in the table below.

Objections to the studies of Table 1 are related to the random samples assessed and the outcome measures used. The first objection is that the techniques were recommended based on random numbers generated using mathematical functions. The skewness and kurtosis properties of the random numbers may not generalize to the samples observed in applied situations in education and psychology. To the extent that Monte Carlo samples represent applied situations, the results are generalizable to similar situations (Sawilowsky & Fahoome, 2003).

The second objection with the manner in which the preceding studies were conducted is that the techniques were recommended based on Type I and Type II error rates alone. The Type I and the Type II error rates indirectly relate to confidence intervals; whereas, the probability-coverage and interval-length serve as outcome measures for confidence intervals. Though interval-length serves as an outcome measure for confidence intervals, journals in education and in psychology did not provide the interval-length for assessing Welch's and Yuen's techniques.

Table 1. Probability of Coverage of Yuen's & Welch's Techniques Reported in the Literature.

Test	Citation	n1	n2	σ1	σ2	Skew.	Kurt.	PC
Welch's	Yuen (1974)	10	10	1.00	0.71	0.00	-1.20	0.95
		20	10	1.00	0.71	0.00	-1.20	0.95
		20	20	1.41	0.71	0.00	-1.20	0.95
		10	20	4.00	1.00	0.00	0.00	0.95
		10	10	2.00	1.00	0.00	0.00	0.95
	Algina (1994)	33	67	3.00	1.00	6.10	np	0.88
	et al.	33	67	2.00	1.00	6.10	np	0.90
		33	67	1.00	1.00	6.10	np	0.94
	Penfield (1994)	10	20	1.00	1.00	0.00	np	0.96
		10	20	1.00	1.00	1.50	np	0.96
		20	20	1.00	2.00	0.00	np	0.95
		20	20	1.00	2.00	1.50	np	0.95
		10	20	1.00	2.00	1.50	np	0.95
Welch's		10	20	1.00	2.00	1.00	np	0.95
		10	20	2.00	1.00	0.00	np	0.95
	Penfield (1994)	10	20	2.00	1.00	1.50	np	0.96
	Luh & Guo (2000)	12	24	1.00	4.00	6.20	111.00	0.91
		12	24	4.00	1.00	6.20	111.00	0.85
	Guo & Luh (2000)	18	12	1.00	6.00	1.75	5.90	0.92
		18	12	1.00	6.00	6.20	111.00	0.85
Yuen's	Luh & Guo (2000)	12	24	1.00	4.00	6.20	111.00	0.95
		12	24	4.00	1.00	6.20	111.00	0.92
	Wilcox (1994)	12	12	1.00	1.00	2.00	6.00	0.95
		40	12	1.00	1.00	2.00	6.00	0.95
		80	20	1.00	1.00	2.00	6.00	0.94
		12	12	1.00	1.00	3.90	42.20	0.95
		40	12	1.00	1.00	3.90	42.20	0.95
		80	20	1.00	1.00	3.90	42.20	0.95

Purpose

The purpose of the study was to assess the probability-coverage and the interval-length for Welch's and Yuen's techniques. The techniques were assessed (a) using empirical data sets that were not normally distributed (i.e., Sawilowsky & Blair, 1992), (b) under conditions of heteroscedasticity, and (c) for unequal group sample sizes.

Methodology

Micceri (1986) identified eight distributions prevalent in educational and psychological research. Table 2 provides the means, standard deviations and third and fourth moment estimates of skewness and kurtosis of the eight distributions. The kurtosis was adjusted so that the value for a normal distribution would be 0.00. Estimates of interval-length and probability-coverage were obtained by sampling from the seven distributions. Random samples were obtained independently and with replacement using the International Mathematical and Statistical Libraries (1998): RNUND and RNSET subroutines. One million repetitions were performed.

The procedure involved obtaining random samples from the empirical distributions, standardizing the scores, modeling the effect and modeling heterogeneity, trimming and Winsorizing the dataset, computing the interval, summing values of interval length and probability-coverage, and averaging values of interval length and the values of probability-coverage.

Sample size ratios of 1:1, 3:1, and 1:3 were selected. The respective sample sizes were (n_1, n_2) = (13, 13), (13, 39), (39, 13), and (39, 39). Variance ratios of 1:1, 1:2 and 1:4 allowed for a comparison of the probability-coverage and interval-lengths for each technique under homoscedasticity and heteroscedasticity. Coverage-probabilities and interval-length were examined at the 0.01, and 0.05 alpha levels.

Where μ' is the mean for the transformed score, σ' is the standard deviation for the transformed score, and Z is a standard score, the transformed score was obtained as follows.

$$X' = \mu' + \sigma' Z \qquad (10)$$

The mean of the second group was set to one. The levels of skewness, size, variance, and effect under study represent a subset of conditions in applied situations.

The ratio of the average length for Student's technique divided by the average length for the comparison technique, i.e., Welch's or Yuen's technique, was calculated to compare interval lengths.

Results

Probability-coverage

The results showed inflated probability-coverage for Yuen's techniques was observed with extreme skewness. Probability-coverage was greater than the confidence-level when skewness was above 1.25, sample sizes were equal and less than 25 or sample sizes were unequal. The results were observed under homoscedasticity. In addition, probability-coverage was greater than the confidence-level when skewness was above 1.25 and heteroscedasticity was proportional to size or sample sizes were equal, less than 25, and heteroscedastic. The probability-coverage exceeded the upper bound of the Bradley-criterion, i.e., $(1-\hat{\alpha}) > (1-0.5\alpha)$. The results were not observed where σ_2 / σ_1 = 4. Results were presented in Table 3 through Table 9.

Welch's technique:

Attenuated coverage-probabilities were observed for both extreme skewness (i.e., absolute skewness greater than 1.25) and heteroscedasticity ($\sigma_2 / \sigma_1 =$ 4). That is, coverage-probabilities were less than 0.925 ($\alpha = 0.05$) or 0.985 ($\alpha = 0.01$). The results occurred where sample sizes were inversely proportional to variances; alternatively, the results occurred where group sample sizes were less than 25.

Table 2. Descriptive Information Pertaining to Eight Real World Distributions.

Distribution	M	SD	Skew.	Kurt.
Mass at Zero	12.92	4.42	-0.03	0.31
Extreme Asymmetry-Psychometric	13.67	5.75	1.64	1.52
Extreme Asymmetry-Achievement	24.5	5.79	-1.33	1.11
Extreme Bimodality	2.97	1.69	-0.08	-1.70
Multimodal & Lumpy	21.15	11.9	0.19	-1.20
Digit Preference	536.95	37.64	-0.07	-0.24
Smooth Symmetric	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

Table 3. Coverage-probabilities for Each Technique by Sizes, Standard Deviations, and Alpha Levels when Sampling from an Extreme Asymmetry – Achievement Distribution.

			$\alpha = 0.05$			$\alpha = 0.01$		
n1	n2	σ2/σ1	Student	Welch	Yuen	Student	Welch	Yuen
13	13	1	0.952	0.956a	0.964a	0.992a	0.994a	0.995a
13	39	1	0.952	0.937c	0.948	0.991	0.980d	0.988c
39	13	1	0.952	0.937c	0.948	0.991	0.980d	0.988c
39	39	1	0.950	0.950	0.955	0.990	0.991	0.993a
13	13	2	0.933c	0.935c	0.947	0.978d	0.979d	0.988c
13	39	2	0.986b	0.952	0.958a	0.997b	0.992a	0.994a
39	13	2	0.838d	0.922d	0.932c	0.934d	0.965d	0.974d
39	39	2	0.945	0.946	0.948	0.986c	0.986c	0.988c
13	13	4	0.914d	0.921d	0.931c	0.961d	0.965d	0.974d
13	39	4	0.992b	0.945	0.948	0.998b	0.986c	0.988c
39	13	4	0.753d	0.918d	0.929c	0.863d	0.962d	0.972d
39	39	4	0.938c	0.941c	0.943c	0.981d	0.982d	0.983d

a. $1-\hat{\alpha} > 0.955, \alpha = 0.05$ or $1-\hat{\alpha} > 0.991, \alpha = 0.01$
b. $1-\hat{\alpha} > 0.975, \alpha = 0.05$ or $1-\hat{\alpha} > 0.995, \alpha = 0.01$
c. $1-\hat{\alpha} < 0.945, \alpha = 0.05$ or $1-\hat{\alpha} < 0.989, \alpha = 0.01$
d. $1-\hat{\alpha} < 0.925, \alpha = 0.05$ or $1-\hat{\alpha} < 0.985, \alpha = 0.01$

Table 4. Coverage-probabilities for Each Technique by Sizes, Standard Deviations, and Alpha Levels when Sampling from an Extreme Bimodal Distribution.

			$\alpha = 0.05$			$\alpha = 0.01$		
n1	n2	σ2/σ1	Student	Welch	Yuen	Student	Welch	Yuen
13	13	1	0.962a	0.962a	0.961a	0.994a	0.994a	0.993a
13	39	1	0.959a	0.958a	0.952	0.994a	0.993a	0.988c
39	13	1	0.959a	0.958a	0.952	0.994a	0.993a	0.988c
39	39	1	0.950	0.950	0.949	0.990	0.990	0.989
13	13	2	0.958a	0.961a	0.953	0.993a	0.994a	0.989
13	39	2	0.993b	0.955	0.953	0.999b	0.991	0.990
39	13	2	0.857d	0.960a	0.948	0.949d	0.994a	0.984d
39	39	2	0.948	0.950	0.947	0.989	0.989	0.987c
13	13	4	0.953	0.961a	0.949	0.991	0.994a	0.984d
13	39	4	0.997b	0.951	0.948	1.000b	0.990	0.987c
39	13	4	0.781d	0.961a	0.949	0.893d	0.994a	0.982d
39	39	4	0.946	0.949	0.946	0.988c	0.989	0.985c

a. $1-\hat{\alpha} > 0.955, \alpha = 0.05$ or $1-\hat{\alpha} > 0.991, \alpha = 0.01$
b. $1-\hat{\alpha} > 0.975, \alpha = 0.05$ or $1-\hat{\alpha} > 0.995, \alpha = 0.01$
c. $1-\hat{\alpha} < 0.945, \alpha = 0.05$ or $1-\hat{\alpha} < 0.989, \alpha = 0.01$
d. $1-\hat{\alpha} < 0.925, \alpha = 0.05$ or $1-\hat{\alpha} < 0.985, \alpha = 0.01$

Table 5. Coverage-probabilities for Each Technique by Sizes, Standard Deviations, and Alpha Levels when Sampling from a Digit Preference Distribution.

			$\alpha = 0.05$			$\alpha = 0.01$		
n1	n2	σ2/σ1	Student	Welch	Yuen	Student	Welch	Yuen
13	13	1	0.950	0.951	0.951	0.990	0.991	0.990
13	39	1	0.950	0.949	0.947	0.990	0.989	0.988c
39	13	1	0.950	0.949	0.947	0.990	0.989	0.988c
39	39	1	0.950	0.950	0.949	0.990	0.990	0.990
13	13	2	0.946	0.950	0.948	0.988c	0.990	0.989
13	39	2	0.991b	0.950	0.949	0.999b	0.990	0.989
39	13	2	0.846d	0.949	0.945	0.940d	0.989	0.988c
39	39	2	0.948	0.950	0.948	0.989	0.990	0.989
13	13	4	0.941c	0.949	0.945	0.985c	0.989	0.988c
13	39	4	0.998b	0.950	0.949	1.000b	0.990	0.989
39	13	4	0.765d	0.949	0.946	0.881d	0.989	0.988c
39	39	4	0.947	0.950	0.948	0.989	0.990	0.989

a. $1-\hat{\alpha} > 0.955, \alpha = 0.05$ or $1-\hat{\alpha} > 0.991, \alpha = 0.01$
b. $1-\hat{\alpha} > 0.975, \alpha = 0.05$ or $1-\hat{\alpha} > 0.995, \alpha = 0.01$
c. $1-\hat{\alpha} < 0.945, \alpha = 0.05$ or $1-\hat{\alpha} < 0.989, \alpha = 0.01$
d. $1-\hat{\alpha} < 0.925, \alpha = 0.05$ or $1-\hat{\alpha} < 0.985, \alpha = 0.01$

Table 6. Coverage-probabilities for Each Technique by Sizes, Standard Deviations, and Alpha Levels when Sampling from a Mass at Zero Distribution.

			$\alpha = 0.05$			$\alpha = 0.01$		
n1	n2	σ2/σ1	Student	Welch	Yuen	Student	Welch	Yuen
13	13	1	0.950	0.951	0.951	0.990	0.991	0.990
13	39	1	0.950	0.950	0.947	0.990	0.990	0.988c
39	13	1	0.950	0.950	0.947	0.990	0.990	0.988c
39	39	1	0.950	0.950	0.950	0.990	0.990	0.990
13	13	2	0.947	0.951	0.948	0.989	0.990	0.989
13	39	2	0.991b	0.950	0.950	0.999b	0.990	0.990
39	13	2	0.847d	0.950	0.945	0.941d	0.990	0.987c
39	39	2	0.949	0.950	0.948	0.990	0.990	0.989
13	13	4	0.942c	0.950	0.945	0.986c	0.990	0.987c
13	39	4	0.998b	0.950	0.948	1.000b	0.990	0.989
39	13	4	0.765d	0.950	0.945	0.882d	0.990	0.988c
39	39	4	0.947	0.950	0.948	0.989	0.990	0.989

a. $1-\hat{\alpha} > 0.955, \alpha = 0.05$ or $1-\hat{\alpha} > 0.991, \alpha = 0.01$
b. $1-\hat{\alpha} > 0.975, \alpha = 0.05$ or $1-\hat{\alpha} > 0.995, \alpha = 0.01$
c. $1-\hat{\alpha} < 0.945, \alpha = 0.05$ or $1-\hat{\alpha} < 0.989, \alpha = 0.01$
d. $1-\hat{\alpha} < 0.925, \alpha = 0.05$ or $1-\hat{\alpha} < 0.985, \alpha = 0.01$

Table 7. Coverage-probabilities for Each Technique by Sizes, Standard Deviations, and Alpha Levels when Sampling from a Smooth Symmetric Distribution.

			$\alpha = 0.05$			$\alpha = 0.01$		
n1	n2	σ2/σ1	Student	Welch	Yuen	Student	Welch	Yuen
13	13	1	0.950	0.950	0.950	0.990	0.990	0.990
13	39	1	0.950	0.948	0.946	0.990	0.989	0.988c
39	13	1	0.950	0.949	0.947	0.990	0.989	0.988c
39	39	1	0.950	0.950	0.950	0.990	0.990	0.990
13	13	2	0.945	0.949	0.947	0.988c	0.989	0.988c
13	39	2	0.991b	0.950	0.950	0.999b	0.990	0.990
39	13	2	0.846d	0.949	0.945	0.939d	0.989	0.987c
39	39	2	0.949	0.950	0.949	0.989	0.990	0.989
13	13	4	0.941c	0.949	0.946	0.985c	0.989	0.988c
13	39	4	0.998b	0.950	0.949	1.000b	0.990	0.989
39	13	4	0.765d	0.949	0.946	0.881d	0.989	0.988c
39	39	4	0.947	0.950	0.948	0.988c	0.990	0.989

a. $1-\hat{\alpha} > 0.955, \alpha = 0.05$ or $1-\hat{\alpha} > 0.991, \alpha = 0.01$
b. $1-\hat{\alpha} > 0.975, \alpha = 0.05$ or $1-\hat{\alpha} > 0.995, \alpha = 0.01$
c. $1-\hat{\alpha} < 0.945, \alpha = 0.05$ or $1-\hat{\alpha} < 0.989, \alpha = 0.01$
d. $1-\hat{\alpha} < 0.925, \alpha = 0.05$ or $1-\hat{\alpha} < 0.985, \alpha = 0.01$

Table 8. Coverage-probabilities for Each Technique by Sizes, Standard Deviations, and Alpha Levels when Sampling from a Multimodal Lumpy Distribution.

			$\alpha = 0.05$			$\alpha = 0.01$		
n1	n2	σ2/σ1	Student	Welch	Yuen	Student	Welch	Yuen
13	13	1	0.949	0.949	0.949	0.989	0.989	0.989
13	39	1	0.950	0.947	0.939c	0.990	0.987c	0.981d
39	13	1	0.950	0.947	0.939c	0.990	0.987c	0.981d
39	39	1	0.950	0.950	0.950	0.990	0.990	0.989
13	13	2	0.944c	0.947	0.937c	0.986c	0.987c	0.980d
13	39	2	0.991b	0.950	0.948	0.999b	0.989	0.989
39	13	2	0.845d	0.947	0.930c	0.937d	0.986c	0.972d
39	39	2	0.948	0.949	0.947	0.989	0.989	0.987c
13	13	4	0.938c	0.946	0.929c	0.982d	0.986c	0.971d
13	39	4	0.997b	0.950	0.948	1.000b	0.989	0.987c
39	13	4	0.767d	0.946	0.929c	0.880d	0.986c	0.971d
39	39	4	0.947	0.949	0.946	0.988c	0.989	0.986c

a. $1-\hat{\alpha} > 0.955, \alpha = 0.05$ or $1-\hat{\alpha} > 0.991, \alpha = 0.01$

b. $1-\hat{\alpha} > 0.975, \alpha = 0.05$ or $1-\hat{\alpha} > 0.995, \alpha = 0.01$

c. $1-\hat{\alpha} < 0.945, \alpha = 0.05$ or $1-\hat{\alpha} < 0.989, \alpha = 0.01$

d. $1-\hat{\alpha} < 0.925, \alpha = 0.05$ or $1-\hat{\alpha} < 0.985, \alpha = 0.01$

Table 9. Coverage-probabilities for Each Technique by Sizes, Standard Deviations, and Alpha Levels when Sampling from a Extreme Asymmetry - Psychometric Distribution.

			$\alpha = 0.05$			$\alpha = 0.01$		
n1	n2	σ2/σ1	Student	Welch	Yuen	Student	Welch	Yuen
13	13	1	0.969a	0.973a	0.989b	0.996b	0.998b	0.999b
13	39	1	0.961a	0.949	0.979b	0.991	0.985c	0.998b
39	13	1	0.960a	0.948	0.979b	0.991	0.985c	0.998b
39	39	1	0.952	0.953	0.969a	0.992a	0.992a	0.997b
13	13	2	0.943c	0.946	0.982b	0.983d	0.984d	0.999b
13	39	2	0.983b	0.960a	0.976b	0.996b	0.995a	0.998b
39	13	2	0.861d	0.928c	0.958a	0.950d	0.965d	0.994a
39	39	2	0.944c	0.945	0.947	0.985c	0.985c	0.989
13	13	4	0.921d	0.925c	0.952	0.961d	0.963d	0.995a
13	39	4	0.989b	0.944c	0.944c	0.997b	0.984d	0.987c
39	13	4	0.779d	0.923d	0.940c	0.880d	0.960d	0.987c
39	39	4	0.936c	0.938c	0.927c	0.977d	0.979d	0.968d

a. $1-\hat{\alpha} > 0.955, \alpha = 0.05$ or $1-\hat{\alpha} > 0.991, \alpha = 0.01$
b. $1-\hat{\alpha} > 0.975, \alpha = 0.05$ or $1-\hat{\alpha} > 0.995, \alpha = 0.01$
c. $1-\hat{\alpha} < 0.945, \alpha = 0.05$ or $1-\hat{\alpha} < 0.989, \alpha = 0.01$
d. $1-\hat{\alpha} < 0.925, \alpha = 0.05$ or $1-\hat{\alpha} < 0.985, \alpha = 0.01$

Table 10. Ratio of the Average Lengths for Student's Technique to that for Welch's and Yuen's Techniques when Sampling from an Extreme Asymmetry – Achievement Distribution.

			$\alpha = 0.05$		$\alpha = 0.01$	
n_1	n_2	σ_2/σ_1	Welch	Yuen	Welch	Yuen
13	13	1	0.993a	0.733a	0.989a	0.712a
13	39	1	0.975c	0.727	0.954d	0.696c
39	13	1	0.975c	0.727	0.954d	0.696c
39	39	1	0.999	0.763	0.999	0.757a
13	13	2	0.979c	0.724	0.966d	0.695c
13	39	2	1.359	1.023a	1.353a	1.005a
39	13	2	0.694d	0.514c	0.667d	0.480d
39	39	2	0.994	0.760	0.991c	0.751c
13	13	4	0.960d	0.709c	0.935d	0.670d
13	39	4	1.610	1.224	1.608c	1.211c
39	13	4	0.573d	0.424c	0.547d	0.390d
39	39	4	0.987c	0.755c	0.980d	0.742d

a. $1-\hat{\alpha} > 0.955, \alpha = 0.05$ or $1-\hat{\alpha} > 0.991, \alpha = 0.01$

b. $1-\hat{\alpha} > 0.975, \alpha = 0.05$ or $1-\hat{\alpha} > 0.995, \alpha = 0.01$

c. $1-\hat{\alpha} < 0.945, \alpha = 0.05$ or $1-\hat{\alpha} < 0.989, \alpha = 0.01$

d. $1-\hat{\alpha} < 0.925, \alpha = 0.05$ or $1-\hat{\alpha} < 0.985, \alpha = 0.01$

Table 11. Ratio of the Average Lengths for Student's Technique to that for Welch's and Yuen's Techniques when Sampling from an Extreme Bimodal Distribution.

			$\alpha = 0.05$		$\alpha = 0.01$	
n_1	n_2	σ_2/σ_1	Welch	Yuen	Welch	Yuen
13	13	1	0.999a	0.886a	0.999a	0.870a
13	39	1	0.964a	0.840	0.943a	0.809c
39	13	1	0.964a	0.840	0.944a	0.809c
39	39	1	1.000	0.827	1.000	0.822
13	13	2	0.981a	0.864	0.969a	0.833
13	39	2	1.360	1.162	1.356	1.148
39	13	2	0.682a	0.594	0.655a	0.556d
39	39	2	0.994	0.820	0.991	0.811c
13	13	4	0.959a	0.837	0.934a	0.790d
13	39	4	1.613	1.343	1.612	1.330c
39	13	4	0.568a	0.493	0.542a	0.455d
39	39	4	0.987	0.811	0.980	0.797c

a. $1-\hat{\alpha} > 0.955, \alpha = 0.05$ or $1-\hat{\alpha} > 0.991, \alpha = 0.01$

b. $1-\hat{\alpha} > 0.975, \alpha = 0.05$ or $1-\hat{\alpha} > 0.995, \alpha = 0.01$

c. $1-\hat{\alpha} < 0.945, \alpha = 0.05$ or $1-\hat{\alpha} < 0.989, \alpha = 0.01$

d. $1-\hat{\alpha} < 0.925, \alpha = 0.05$ or $1-\hat{\alpha} < 0.985, \alpha = 0.01$

Table 12. Ratio of the Average Lengths for Student's Technique to that for Welch's and Yuen's Techniques when Sampling from a Digit Preference Distribution.

			$\alpha = 0.05$		$\alpha = 0.01$	
n_1	n_2	σ_2/σ_1	Welch	Yuen	Welch	Yuen
13	13	1	0.996	0.768	0.994	0.750
13	39	1	0.971	0.749	0.950	0.718c
39	13	1	0.971	0.749	0.950	0.718c
39	39	1	1.000	0.775	0.999	0.770
13	13	2	0.980	0.753	0.968	0.725
13	39	2	1.361	1.054	1.356	1.038
39	13	2	0.688	0.527	0.662	0.492c
39	39	2	0.994	0.769	0.991	0.761
13	13	4	0.959	0.733	0.935	0.692c
13	39	4	1.612	1.247	1.610	1.234
39	13	4	0.570	0.435	0.544	0.400c
39	39	4	0.987	0.763	0.980	0.749

a. $1-\hat{\alpha} > 0.955, \alpha = 0.05$ or $1-\hat{\alpha} > 0.991, \alpha = 0.01$

b. $1-\hat{\alpha} > 0.975, \alpha = 0.05$ or $1-\hat{\alpha} > 0.995, \alpha = 0.01$

c. $1-\hat{\alpha} < 0.945, \alpha = 0.05$ or $1-\hat{\alpha} < 0.989, \alpha = 0.01$

d. $1-\hat{\alpha} < 0.925, \alpha = 0.05$ or $1-\hat{\alpha} < 0.985, \alpha = 0.01$

Table 13. Ratio of the Average Lengths for Student's Technique to that for Welch's and Yuen's Techniques when Sampling from a Mass at Zero Distribution.

			$\alpha = 0.05$		$\alpha = 0.01$	
n_1	n_2	σ_2/σ_1	Welch	Yuen	Welch	Yuen
13	13	1	0.995	0.812	0.993	0.793
13	39	1	0.972	0.793	0.951	0.761c
39	13	1	0.972	0.794	0.951	0.761c
39	39	1	1.000	0.824	0.999	0.819
13	13	2	0.980	0.796	0.967	0.767
13	39	2	1.360	1.119	1.355	1.102
39	13	2	0.690	0.558	0.663	0.521c
39	39	2	0.994	0.819	0.991	0.809
13	13	4	0.959	0.775	0.935	0.731c
13	39	4	1.611	1.325	1.610	1.312
39	13	4	0.571	0.460	0.545	0.423c
39	39	4	0.987	0.811	0.980	0.797

a. $1-\hat{\alpha} > 0.955, \alpha = 0.05$ or $1-\hat{\alpha} > 0.991, \alpha = 0.01$

b. $1-\hat{\alpha} > 0.975, \alpha = 0.05$ or $1-\hat{\alpha} > 0.995, \alpha = 0.01$

c. $1-\hat{\alpha} < 0.945, \alpha = 0.05$ or $1-\hat{\alpha} < 0.989, \alpha = 0.01$

d. $1-\hat{\alpha} < 0.925, \alpha = 0.05$ or $1-\hat{\alpha} < 0.985, \alpha = 0.01$

Table 14. Ratio of the Average Lengths for Student's Technique to that for Welch's and Yuen's Techniques when Sampling from a Smooth Symmetric Distribution.

			$\alpha = 0.05$		$\alpha = 0.01$	
n_1	n_2	σ_2/σ_1	Welch	Yuen	Welch	Yuen
13	13	1	0.996	0.735	0.994	0.718
13	39	1	0.971	0.716	0.950	0.687c
39	13	1	0.971	0.716	0.950	0.687c
39	39	1	1.000	0.742	0.999	0.738
13	13	2	0.980	0.721	0.968	0.694c
13	39	2	1.361	1.009	1.356	0.994
39	13	2	0.688	0.504	0.661	0.470c
39	39	2	0.994	0.737	0.991	0.729
13	13	4	0.959	0.702	0.935	0.662c
13	39	4	1.612	1.195	1.611	1.183
39	13	4	0.570	0.416	0.544	0.383c
39	39	4	0.987	0.731	0.980	0.718

a. $1-\hat{\alpha} > 0.955, \alpha = 0.05$ or $1-\hat{\alpha} > 0.991, \alpha = 0.01$

b. $1-\hat{\alpha} > 0.975, \alpha = 0.05$ or $1-\hat{\alpha} > 0.995, \alpha = 0.01$

c. $1-\hat{\alpha} < 0.945, \alpha = 0.05$ or $1-\hat{\alpha} < 0.989, \alpha = 0.01$

d. $1-\hat{\alpha} < 0.925, \alpha = 0.05$ or $1-\hat{\alpha} < 0.985, \alpha = 0.01$

Table 15. Ratio of the Average Lengths for Student's Technique to that for Welch's and Yuen's Techniques when Sampling from a Multimodal Lumpy Distribution.

.			$\alpha = 0.05$		$\alpha = 0.01$	
n_1	n_2	σ_2/σ_1	Welch	Yuen	Welch	Yuen
13	13	1	0.998	0.808	0.997	0.790
13	39	1	0.968	0.777c	0.947c	0.747d
39	13	1	0.968	0.777c	0.947c	0.747d
39	39	1	1.000	0.784	1.000	0.779
13	13	2	0.980	0.790c	0.969c	0.762d
13	39	2	1.362	1.085	1.357	1.071
39	13	2	0.685	0.548c	0.658c	0.512d
39	39	2	0.994	0.778	0.991	0.769c
13	13	4	0.959	0.768c	0.935c	0.725d
13	39	4	1.613	1.268	1.612	1.256c
39	13	4	0.569	0.454c	0.543c	0.418d
39	39	4	0.987	0.770	0.980	0.756c

a. $1-\hat{\alpha} > 0.955, \alpha = 0.05$ or $1-\hat{\alpha} > 0.991, \alpha = 0.01$

b. $1-\hat{\alpha} > 0.975, \alpha = 0.05$ or $1-\hat{\alpha} > 0.995, \alpha = 0.01$

c. $1-\hat{\alpha} < 0.945, \alpha = 0.05$ or $1-\hat{\alpha} < 0.989, \alpha = 0.01$

d. $1-\hat{\alpha} < 0.925, \alpha = 0.05$ or $1-\hat{\alpha} < 0.985, \alpha = 0.01$

Table 16. Ratio of the Average Lengths for Student's Technique to that for Welch's and Yuen's Techniques when Sampling from an Extreme Asymmetry – Psychometric Distribution.

			$\alpha = 0.05$		$\alpha = 0.01$	
n_1	n_2	σ_2/σ_1	Welch	Yuen	Welch	Yuen
13	13	1	0.992a	0.698b	0.988b	0.673b
13	39	1	0.967	0.702b	0.946c	0.669b
39	13	1	0.967	0.702b	0.945c	0.668b
39	39	1	0.999	0.772a	0.999a	0.765b
13	13	2	0.978	0.696b	0.965d	0.666b
13	39	2	1.349a	0.998b	1.343a	0.974b
39	13	2	0.691c	0.501a	0.664d	0.467a
39	39	2	0.994	0.774	0.990c	0.765
13	13	4	0.960c	0.692	0.935d	0.654a
13	39	4	1.605c	1.230c	1.604d	1.213c
39	13	4	0.573d	0.416c	0.547d	0.383c
39	39	4	0.988c	0.777c	0.980d	0.763d

a. $1-\hat{\alpha} > 0.955, \alpha = 0.05$ or $1-\hat{\alpha} > 0.991, \alpha = 0.01$

b. $1-\hat{\alpha} > 0.975, \alpha = 0.05$ or $1-\hat{\alpha} > 0.995, \alpha = 0.01$

c. $1-\hat{\alpha} < 0.945, \alpha = 0.05$ or $1-\hat{\alpha} < 0.989, \alpha = 0.01$

d. $1-\hat{\alpha} < 0.925, \alpha = 0.05$ or $1-\hat{\alpha} < 0.985, \alpha = 0.01$

Yuen's technique

For Yuen's techniques, attenuated coverage-probabilities were observed for extreme negative kurtosis. For the population defined by the trimmed mean, kurtosis values less than –1.25 were observed with coverage-probabilities less than 0.985. Given extreme bimodality, coverage-probabilities that were within the range of 0.925-0.975 ($\alpha = 0.05$) were below 0.985 ($\alpha = 0.01$). The kurtosis of the extreme bimodal distribution after trimming was –1.454. The results occurred under both homoscedastic and heteroscedastic conditions. Where size was inversely paired with variance, under a multimodal lumpy distribution, coverage-probabilities were within the range 0.925-0.975 at the 0.05 alpha level. At the 0.01 alpha level, coverage-probabilities were less than 0.985. The kurtosis of the multimodal lumpy distribution after trimming was -1.269.

Interval Length

The results for interval length showed where $S^2_{\min}$ was divided by $n_{\min}$, the interval-lengths for Welch's and Yuen's techniques were less than the interval-lengths for Student's technique. If $S^2_{\max}$ was divided by $n_{\min}$, the reverse was true. Results were presented in Table 10 through Table 16. Second, interval-lengths for Yuen's technique were wider than the interval-lengths for Welch's technique. The interval-length ratios for Yuen's technique were smaller than the ratios of Welch's technique. Larger interval-lengths were observed for the heteroscedastic than for the homoscedastic condition.

Conclusion

Similar to findings by Sawilowsky and Blair (1992, p. 359) showing that skewness attenuated the Type I error rates for the t-test, the results of the present study showed that if skewness was above 1.25, e.g., skewness of the extreme asymmetric - psychometric distribution was 1.417 after trimming, coverage-probabilities were augmented (i.e., $(1-\hat{\alpha}) > (1-0.5\alpha)$).

Similar to findings by Luh and Guo (2000) and Algina et al. (1994) showing that when size was inversely proportional to heteroscedasticity and skewness was greater or equal to 2.00, Welch's technique displayed coverage-probabilities less than the confidence-level when size was inversely proportional to heteroscedasticity and skewness was –1.33 or 1.64.

Finally, the augmentation or attenuation of probability-coverage for both techniques occurred more at 0.01 than at 0.05 alpha levels; this finding was consistent with results from Bradley (1978, p. 147) showing that larger sample sizes were required for the t-test to exhibit robustness at the 0.01 level than at the 0.05 level.

References

References marked with an asterisk indicate studies included in Table 1.

*Algina, J., Oshima, T. C., & Lin, W-Y. (1994). Type I error rates for Welch's tests and James's second-order test under nonnormality and inequality of variance when there are two groups. *Journal of Educational and Behavioral Statistics. 19*, 275-291.

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.

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

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

Edwards, A. L. (1972). *Experimental design in psychological research*. New York: Holt, Rinehart, and Winston, Inc.

Games, P. A. (1983). Curvilinear transformations of the dependent variable. *Psychological Bulletin, 93,* 382-387.

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

*Guo, J-H. & Luh, W-M. (2000). An invertable tranformation two-sample trimmed tstatistic under heterogeneity and nonnormality. *Statistics & Probability Letters, 49*, 1-7.

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

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

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

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

Markowski, C. A., & Markowski, E. P. (1990). Conditions for the effectiveness of a preliminary test of variance. *American Statistician, 44,* 322-326.

Micceri, T. (1986). A futile search for that statistical chimera of normality. Paper presented at the 31st Annual Convention of the Florida Educational Research Association, Tampa.

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

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

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. (2003). A different future for the social and behavioral sciences. *Journal of Modern Applied Statistical Methods, 2*, 128-132.

*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.

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

*Yuen, K. K. (1974)/ The two-sampled trimmed *t* for unequal population variances. *Biometrika,* 61, 165-170.