DIGITALCOMMONS
— @WAYNESTATE —

Journal of Modern Applied Statistical Methods

Volume 11 | Issue 1 Article 5

5-1-2012

Robust Modifications of the Levene and O'Brien Tests for Spread

Abdul R. Othman
Universiti Sains, Malaysia

The Sin Yan
Universiti Sains Malaysia, Georgetown, Penang, Malaysia

H. J. Keselman
University of Manitoba, Winnipeg, Manitoba, Canada

Rand R. Wilcox
University of Southern California, Los Angeles, rwilcox@usc.edu

James Algina
University of Florida

Robust Modifications of the Levene and O'Brien Tests for Spread

Cover Page Footnote
This research was supported by grants provided by the Social Sciences and Humanities Research Council of Canada and the Fundamental Research Grant Scheme of Malaysia.

Journal of Modern Applied Statistical Methods
May 2012, Vol. 11, No. 1, 54-68

1538 – 9472/12/$95.00

Robust Modifications of the Levene and O'Brien Tests for Spread

Abdul R. Othman Teh Sin Yin
Universiti Sains Malaysia,
Georgetown, Penang, Malaysia

H. J. Keselman
University of Manitoba,
Winnipeg, Manitoba Canada

Rand R. Wilcox
University of Southern California,
Los Angeles, CA

James Algina
University of Florida,
Gainesville, FL

Variants of Levene's and O'Brien's procedures not investigated by Keselman, Wilcox & Algina (2008) were examined. Simulations indicate that a new O'Brien variant provides very good Type I error control and is simpler for applied researchers to compute than the method recommended by Keselman, et al.

Key words: Levene test of spread, O'Brien test of spread, Type I error.

Introduction

Keselman, Wilcox, Algina, et al. (2008) compared a number of tests for spread that were based on either least squares or trimmed estimates of central tendency and variability. These estimators were based on either the original data or transformations suggested by Levene (1960) and O'Brien (1981). The adaptive trimming estimators they used were defined by Reed and Stark (1996), estimators which rely on procedures that determine whether data should be trimmed symmetrically, asymmetrically, or not at all. The transformed scores were used in an analysis of variance (ANOVA) F-test, a Welch (1951) test, and a robust ANOVA test due to Lee and Fung (1985). Based on their extensive simulation study, Keselman, et al. recommended a Levene-type transformation based on empirically determined 20% asymmetric trimmed means, involving a particular adaptive estimator, where the transformed scores are then used with an ANOVA F test.

In their investigation, Keselman, et al. only examined a limited number of variations of the Levene (1960) and O'Brien (1979) methods – variations where, by-in-large, the transformed variables were obtained via the application of asymmetrically trimmed means involving one of the seven hinge estimators defined by Reed and Stark (1996). However, there are many other ways in which the transformed variables may be created. For example, the transformed variables may be based on symmetrically trimmed means and then these transformed variables may be symmetrically/asymmetrically transformed with one of the seven hinge estimators. Thus, the purpose of this study was to examine other variants of the Levene and O'Brien methods not examined by Keselman, et al. (2008).

Abdul R. Othman is a Professor of Statistics. Email him at: arothman60@yahoo.com. Teh Sin Yin is a Senior Lecturer of Statistics. Email her at: syin.teh@gmail.com. H. J. Keselman is a Professor of Psychology. His research interests are in applied statistics. Email him at: kesel@ms.umanitoba.ca. Rand R. Wilcox is a Professor of Psychology. Email him at: rwilcox@usc.edu. James Algina is a Professor of Educational Psychology Research and Evaluation Methodology at the University of Florida. Email him at: algina@ufl.edu.

Background

As Keselman, et al. (2008), and others, have noted, the traditional test for equality of variances, e.g., $F = \frac{s_1^2}{s_2^2}$, where s_j^2 is the usual unbiased sample variance for the j^{th} group, is affected adversely when the data in the groups are not normally distributed (i.e., it is sensitive to kurtosis). That is, the actual level of significance can differ substantially from the nominal significance level. In addition, power can be low.

Levene (1960) suggested an alternative test statistic that can be used to assess equality of spread across independent treatment groups. For the one-way layout with model $X_{ij} = \mu_j + \varepsilon_{ij}$ $(i = 1,\ldots,n_j;\ j = 1,\ldots,J)$, where μ_j is the population mean for the j^{th} group and ε_{ij} is random error, Levene suggested that the scores could be modified with the transformation $z_{ij} = \left|X_{ij} - \bar{X}_{.j}\right|$, where $\bar{X}_{.j}$ is the j^{th} sample mean, and then these scores can be used in an ANOVA test. That is, the test suggested by Levene is

$$W_0 = \frac{\sum_{j=1}^{J} n_j (\bar{z}_{j.} - \bar{z}_{..})^2 \Big/ (J-1)}{\sum_{j=1}^{J}\sum_{i=1}^{n_j} (z_{ij} - \bar{z}_{j.})^2 \Big/ \sum_{j=1}^{J} (n_j - 1)}, \qquad (1)$$

where

$$\bar{z}_j = \sum_{i=1}^{n_j} \frac{z_{ij}}{n_j} \text{ and } \bar{z}_{..} = \frac{\sum_{j=1}^{J}\sum_{i=1}^{n_j} z_{ij}}{\sum_{j=1}^{J} n_j}.$$

Critical values for W_0 are obtained from the F-distribution based on $J - 1$ and $\sum_{j=1}^{J}(n_j - 1)$ degrees of freedom. Another statistic relevant to this article is $W_{\bar{X}_t}$. This statistic replaces the group mean in obtaining the transformed Z_{ij}s with the group trimmed mean.

Other methods have also appeared in the literature in addition to Levene's (1960) procedure. Lee and Fung (1985) presented a robust ANOVA F-test based on trimmed means. Keselman, et al. (1979) and others (e.g., O'Brien, 1981) have indicated that a Welch statistic can be adopted instead of the usual ANOVA F-test to assess spread across independent groups. O'Brien (1979) also suggested that a Welch test can be used with his transformation of the data, X_{ij}, namely

$$r_{ij} = \frac{(n_j - 1.5) n_j (X_{ij} - \bar{X}_{.j})^2 - .5 s_j^2 (n_j - 1)}{(n_j - 1)(n_j - 2)}. \qquad (2)$$

Adaptive Trimming Methods

Keselman, et al. (2008) provided a detailed description of adaptive trimming methods. Reed and Stark (1996) defined seven adaptive location estimators based on measures of tail-length and skewness for a set of n observations based on the work of Hogg (1974, 1982). To define these estimators, measures of tail-length and skewness must first be defined. Using the notation of Hogg (1974, 1982) and Reed and Stark (1996) and based on the ordered values, let $L_\alpha =$ the mean of the smallest $[\alpha n]$ observations, where $[\alpha n]$ denotes the greatest integer less than αn and $U_\alpha =$ the mean of the largest $[\alpha n]$ observations. When $\alpha = .05$, $L_{(.05)}$ is the mean of the smallest $[0.05n]$ observations, B is the mean of the next largest $0.15n$ observations, C is the mean of the next largest $0.30n$ observations, D is the mean of the next largest $0.30n$ observations, E the mean of the next largest $0.15n$ observations, and $U_{(.05)}$ the mean of the largest $0.05n$ observations.

Tail-Length Measures

Hogg (1974) defined two measures of tail-length, Q and Q_1, where

$$Q = \left(U_{(.05)} - L_{(.05)}\right) \Big/ \left(U_{(.5)} - L_{(.5)}\right)$$
$$\text{and} \qquad (3)$$
$$Q_1 = \left(U_{(.2)} - L_{(.2)}\right) \Big/ \left(U_{(.5)} - L_{(.5)}\right).$$

Q and Q_1 are location free statistics, are uncorrelated with location statistics and can be used to classify symmetric distributions as light-tailed, medium-tailed or heavy-tailed (Reed & Stark, 1996). According to Hogg (1974) and Reed and Stark (1996), values of $Q < 2$ imply a light-tailed distribution, $2.0 \le Q \le 2.6$ a medium-tailed distribution, $2.6 \le Q \le 3.2$ a heavy-tailed distribution, and $Q > 3.2$ a very heavy-tailed distribution. The cutoffs for Q_1 are: $Q_1 < 1.81$ (light-tailed), $1.81 \le Q_1 \le 1.87$ (medium-tailed) and $Q_1 > 1.87$ (heavy-tailed).

Hogg (1982) introduced yet another measure of tail-length:

$$H_3 = \frac{\left(U_{(.05)} - L_{(.05)}\right)}{\left(E - B\right)}. \tag{4}$$

With this measure, values of $H_3 < 1.26$ suggest that the tails of the distribution are similar to a uniform distribution; values of 1.26 through 1.76 suggest a normal distribution, and values greater than 1.76 suggest the tails are similar to those of a double exponential distribution.

Measures of Skewness

Reed and Stark (1996) defined four measures of skewness as:

$$Q_2 = \left(U_{(.05)} - T_{(.25)}\right) / \left(T_{(.25)} - L_{(.05)}\right),$$
$$H_1 = \left(U_{(.05)} - D\right) / \left(C - L_{(.05)}\right),$$
$$SK_2 = \left(X_{(1)} - XMD\right) / \left(XMD - X_{(n)}\right)$$
and
$$SK_5 = \left(X_{(1)} - XM\right) / \left(XM - X_{(n)}\right), \tag{5}$$

where XMD is the median, XM is the arithmetic mean, $T_{(.25)}$ is the 0.25-trimmed mean (T_α) and $X_{(1)}$ and $X_{(n)}$ are the first and last ordered observations, respectively. Reed (1998) defined the α-trimmed mean as:

$$T_\alpha = \frac{1}{n(1-2\alpha)}\left[\sum_{i=k+1}^{n-k} X_i + (k - \alpha n)\left(X_k + X_{n-k+1}\right)\right]. \tag{6}$$

In this definition a proportion, α, has been trimmed from each tail and the accompanying Winsorized variance S^2 is defined as:

$$S^2 = \frac{1}{(n-1)(1-2\alpha)^2}\left[\sum_{i=k+1}^{n-k}\left(\begin{array}{l}\left(X_i - T_\alpha\right)^2 + k\left(X_k - T_\alpha\right)^2 \\ +k\left(X_{n-k+1} - T_\alpha\right)^2\end{array}\right)\right], \tag{7}$$

where $k = [\alpha n] + 1$.

Based on the definitions of tail-length and skewness, Reed and Stark proposed a set of adaptive linear estimators "that have the capability of asymmetric trimming" (1996, p. 13). They defined a general scheme for their approach as follows:

1. Set the value for the total amount of trimming from the sample, α.

2. Determine the proportion to be trimmed from the lower end of the sample (α_1) by the following proportion: $\alpha_1 = \alpha\left[UW_X / \left(UW_X + LW_X\right)\right]$, where UW_X and LW_X are the numerator and denominator of the defined selector statistics (i.e., tail-length and skewness).

3. The upper trimming proportion is: $\alpha_2 = \alpha - \alpha_1$.

Based on this general schema, Reed and Stark (1996) defined seven hinge estimators, which are trimmed means, as:

1. HQ $\quad \alpha_1 = \alpha\left[UW_Q / \left(UW_Q + LW_Q\right)\right]$,
2. $HQ_1 \quad \alpha_1 = \alpha\left[UW_{Q_1} / \left(UW_{Q_1} + LW_{Q_1}\right)\right]$,
3. $HH_3 \quad \alpha_1 = \alpha\left[UW_{H_3} / \left(UW_{H_3} + LW_{H_3}\right)\right]$,
4. $HQ_2 \quad \alpha_1 = \alpha\left[UW_{Q_2} / \left(UW_{Q_2} + LW_{Q_2}\right)\right]$,
5. $HH_1 \quad \alpha_1 = \alpha\left[UW_{H_1} / \left(UW_{H_1} + LW_{H_1}\right)\right]$,
6. $HSK_2 \quad \alpha_1 = \alpha\left[UW_{SK_2} / \left(UW_{SK_2} + LW_{SK_2}\right)\right]$,

and

7. $HSK_5 \quad \alpha_1 = \alpha\left[UW_{SK_5} / \left(UW_{SK_5} + LW_{SK_5}\right)\right]$. (8)

Keselman, et al. (2008), investigating Type I error rates of procedures for testing spread, examined the Reed and Stark (1996) procedure with various values for α because the literature varies on the amount of recommended (symmetric) trimming. Rosenberger and Gasko (1983) recommend 25% when sample sizes are small (although they state that 20% generally suffices), Wilcox (2005) recommends 20% and Mudholkar, Mudholkar and Srivastava (1991) suggest 15%. Ten percent has been considered by Hill and Dixon (1982), Huber (1977), Stigler (1977) and Staudte and Sheather (1990); results reported by Keselman, et al. (2002) also support 10% trimming. In addition, Keselman, et al. (2005) obtained good results with 5% symmetric trimming.

According to Keselman, et al. (2007), Reed and Stark's (1996) tail-length and skewness measures may be modified for the multi-group problem and applied to the modified multi-group measures to the hinge estimators. In particular, they indicated that each of the measures can be modified by taking weighted averages in a manner analogous to the modifications of tail-length and symmetry measures suggested by Babu, Padmanaban and Puri (1999) of each numerator and denominator term. For example, for the multi-group problem, where n_j represents the number of observations in each group, Q_1 and Q_2 can be defined as:

$$Q_1 = \frac{\left[\sum_j n_j\left(U_{(.2)} - L_{(.2)}\right) \Big/ \sum_j n_j\right]}{\left[\sum_j n_j\left(U_{(.5)} - L_{(.5)}\right) \Big/ \sum_j n_j\right]},$$

$$\text{and} \tag{9}$$

$$Q_2 = \frac{\left[\sum_j n_j\left(U_{(.05)} - T_{(.25)}\right) \Big/ \sum_j n_j\right]}{\left[\sum_j n_j\left(T_{(.25)} - L_{(.05)}\right) \Big/ \sum_j n_j\right]}.$$

The other measures would be similarly modified and it is these multi-group measures of tail-length and skewness that are applied to the general scheme proposed by Reed and Stark (1996), treating the transformed Z_{ij}s as the original variables

One could go a step further than merely applying the transformed Z_{ij}s in a Welch test. It is suggested that the transformed Z_{ij}s be treated as the original random variable in a test statistic that has been found to be generally insensitive to nonnormality, namely a Welch test based on trimmed means, that is, Yuen's (1974) test).

Thus, consider the following. The α trimmed means and Winsorized variances can be defined in a number of different ways (see Hogg, 1974; Reed, 1998; Keselman, et al., 2007; Wilcox, 2003). Let $Z_{(1)j} \le Z_{(2)j} \le \ldots \le Z_{(n_j)j}$ represent the ordered observations associated with the j[th] group. Reed's (1998) approach is based on the work of Hogg (1974). Hogg defined the α-trimmed mean as:

$$m(\alpha) = (1/h)\sum_{i=g+1}^{n_j - g} Z_{(i)}, \tag{10}$$

where α is selected so that $g = \left[n_j\alpha\right]$ and $h = n_j - 2g = n_j - 2[n_j\alpha]$. The standard error of $m(\alpha)$ Hogg suggests is based on the works of Tukey and McLaughlin (1963) and Huber (1970) and is estimated by:

$$S_{m(\alpha)} = \sqrt{\frac{SS(\alpha)}{h(h-1)}}, \tag{11}$$

where $SS(\alpha)$ is the Winsorized sum of squares defined as:

$$\begin{aligned} &(g+1)\left[Z_{(g+1)} - m(\alpha)\right]^2 \\ &+\left[Z_{(g+2)} - m(\alpha)\right]^2 \\ &+\ldots \\ &+\left[Z_{(n_j-g-1)} - m(\alpha)\right]^2 \\ &+(g+1)\left[Z_{(n_j-g)} - m(\alpha)\right]^2. \end{aligned} \tag{12}$$

When allowing for different amounts of trimming in each tail of the distribution, Hogg (1974) defines the trimmed mean as:

$$m(\alpha_1,\alpha_2)=\left(\frac{1}{h}\right)\sum_{i=g_1+1}^{n_j-g_2} z_{(i)}, \qquad (13)$$

where $g_1=[n_j\alpha_1]$, $g_2=[n_j\alpha_2]$ and $h_j=n_j - g_1 - g_2$. Hogg suggests that the standard deviation of $m(\alpha_1,\alpha_2)$ can be estimated as:

$$S_{m(\alpha_1,\alpha_2)}=\sqrt{\frac{SS(\alpha_1,\alpha_2)}{h(h-1)}}, \qquad (14)$$

where $SS(\alpha_1,\alpha_2)$ can be calculated as

$$SS(\alpha_1,\alpha_2)=\frac{\left(\begin{array}{c}(g_1+1)\left[Z_{(g_1+1)}-m(\alpha_1,\alpha_2)\right]^2\\ +\left[Z_{(g_1+2)}-m(\alpha_1,\alpha_2)\right]^2\\ +\ldots\\ +\left[Z_{(n_j-g_2-1)}-m(\alpha_1,\alpha_2)\right]^2\\ +(g_2+1)\left[Z_{(n_j-g_2)}-m(\alpha_1,\alpha_2)\right]^2\\ -\left\{\begin{array}{c}(g_1)\left[Z_{(g_1+1)}-m(\alpha_1,\alpha_2)\right]\\ +(g_2)\left[Z_{(n_j-g_2)}-m(\alpha_1,\alpha_2)\right]\end{array}\right\}^2\end{array}\right)}{n_j}$$

(15)

Test Statistics

Let $\hat{\mu}_j=\sum_{i=1}^{n_j} Z_{ij}/n_j$ and $s_j^2=\sum_{i=1}^{n}(Z_i-\bar{Z}_j)^2/(n_j-1)$ where $\hat{\mu}_j$ is the estimate of μ_j and s_j^2 is the unbiased estimate of the variance for population j. A heteroscedastic statistic (Welch, 1951) can be defined as:

$$F_W=\frac{\sum_{j=1}^{J} w_j(\hat{\mu}_j-\hat{\mu})^2/(J-1)}{1+\frac{2(J-2)}{(J^2-1)}\sum_{j=1}^{J}\frac{(1-w_j/W)^2}{n_j-1}}, \qquad (16)$$

where

$$\hat{\mu}=\sum_{j=1}^{J} w_j\hat{\mu}_j/W,$$

$$W=\sum_{j=1}^{J} w_j \ a,.$$

and

$$w_j=n_j/s_j^2.$$

The test statistic is approximately distributed as an F variate and is referred to the critical value $F[(1-\alpha);(J-1),\nu_W]$, the $(1-\alpha)$ quantile of the F distribution, where error degrees of freedom are obtained from

$$\upsilon_w=\frac{J^2-1}{3\sum_{j=1}^{J}\frac{\left(1-w_j/W\right)^2}{n_j-1}}. \qquad (17)$$

A Robust ANOVA F-Test

Lee and Fung (1985) defined an ANOVA F-test based on trimmed means. Because the ANOVA F-test can be more powerful than the Welch F-test, this statistic was chosen for this investigation. The Lee and Fung (1985) statistic is defined as:

$$F_t=\frac{\sum_{j=1}^{J} h_j(\hat{m}_{tj}-\hat{M}_t)^2/(J-1)}{\sum_{j=1}^{J} SS_j(a_1, a_2)/(H-J)} \qquad (18)$$

where

$$H=\sum_{j=1}^{J} h_j,$$

$$\hat{M}_t=\sum_{j=1}^{J} h_j\hat{m}_{tj}/H,$$

and

$$SS_j(\alpha_1, \alpha_2)=\text{the } (\alpha_1, \alpha_2).$$

The Winsorized sum of squared deviations for the j^{th} group; h_j and $\hat{m}_{tj}$ are defined the same as previously. Note that, when $\alpha_1 = \alpha_2 = 0$, $F_t = F$ (O'Brien, 1979).

O'Brien (1979) indicated that the r_{ij}s can be used in the Welch test. Accordingly, the trimmed mean version is given by:

$$r_{t_{ij}} = \frac{\left(h_j - 1.5\right) n_j \left(X_{ij} - \hat{m}_{tj}\right)^2 - .5 s_{mj}^2 \left(h_j - 1\right)}{\left(h_j - 1\right)\left(h_j - 2\right)}. \tag{19}$$

where $h_j = n_j - g_1 - g_2$, $\hat{m}_{tj} =$ the (α_1, α_2) trimmed mean of the j^{th} group and $s_{mj}^2 = s_{m(\alpha_1,\alpha_2)}^2$ for group j. The r_{ij}s and $r_{t_{ij}}$s were also used with the Lee and Fung (1985) test.

Methodology

A total of 170 new Levene (1960) type procedures were created. These procedures were (see Table 2 for a summary of the Levene methods examined):

(A1) Let $Z_{ij} = \left|X_{ij} - \bar{X}_{.j}\right|$. The Z_{ij}s were then trimmed symmetrically $\alpha\%$ and the robust F-test, $F_{t\alpha}$ was computed. There are 4 variants with this designation because there are four symmetric trimming percentages: 5%, 10%, 15% and 20%. For example, $A_1 15$ signifies $Z_{ij} = \left|X_{ij} - \bar{X}_{.j}\right|$, the Z_{ij}s were trimmed symmetrically 15% and $F_{t\alpha}$ was computed.

(A2) Let $Z_{ij} = \left|X_{ij} - \bar{X}_{.j}\right|$, the Z_{ij}s were trimmed asymmetrically $\beta\%$ with a hinge estimator H and the robust F-test, $F_{t\beta}$ computed. $A_2 10HSK_5$ signifies transformation $Z_{ij} = \left|X_{ij} - \bar{X}_{.j}\right|$ and these values were subjected to 10% asymmetric trimming with hinge estimator HSK_5 before computing $F_{t\beta}$. Because there are four asymmetrical trimming percentages and seven hinge estimators, there are 28 variants with this designation.

(A3) Let $Z_{ij} = \left|X_{ij} - \bar{X}_{.j}\right|$. The Z_{ij}s were used with the Welch test, F_W.

(B3) Let $Z_{ij} = \left|X_{ij} - M_j\right|$. The Z_{ij}s were used with the Welch test, F_W.

(E1) These variants are designated $E_1\alpha H$, where $Z_{ij} = |X_{ij} - \bar{X}_{t_\alpha}|$, $\bar{X}_{t_\alpha}$ are group $\alpha\%$ symmetric trimmed means. The Z_{ij}s were used with the usual *F*-test, W, hence variant $E_1 20$ signifies transformed values $Z_{ij} = |X_{ij} - \bar{X}_{t_{20}}|$, where $\bar{X}_{t_{20}}$ are group 20% symmetrically trimmed at tail, computed with W. Because there are four symmetric trimming percentages (5%, 10%, 15% and 20%), there are four variants with this designation.

(E2) In variants $E_2\alpha_1\alpha_2$ let $Z_{ij} = |X_{ij} - \bar{X}_{t_{\alpha 1}}|$. The transformed values were trimmed symmetrically $\alpha_2\%$ and used with $F_{t\alpha_2}$. Therefore, variant $E_2 1520$ signifies transformation $Z_{ij} = |X_{ij} - \bar{X}_{t_{15}}|$, where $\bar{X}_{t_{15}}$ are group 15% symmetric trimmed means and these values were subjected to 20% symmetric trimming before being used with F_{ta_2}. Because there are four different $\bar{X}_{t_{\alpha 1}}$'s and four symmetric trimming percentages, there are 16 variants with this designation.

(E3) In variants $E_3\alpha\beta H$, let $Z_{ij} = |X_{ij} - \bar{X}_{t_\alpha}|$. The transformed values were asymmetrically trimmed at $\beta\%$ involving seven hinge estimators (HQ, HQ_1, HH_3, HQ_2, HH_1, HSK_2, HSK_5) and used with $F_{t\beta}$. Hence, $E_3 1025HH_3$

signifies transformation $Z_{ij} = |X_{ij} - \bar{X}_{t_{10}}|$, where $\bar{X}_{t_{10}}$ are group 10% symmetric trimmed means; these values were subjected to 25% asymmetric trimming calculated using the HH_3 hinge estimator before applying $F_{t\beta}$. Because there are four different $\bar{X}_{t_\alpha}$'s and four asymmetric trimming percentages with seven different hinge estimators, there are 112 variants for this designation.

(E4) Variants $E_4\alpha H$ use $\bar{X}_{t_\alpha}$ in place of $\bar{X}_{.j}$ (A_3) or M_j (B_3), in getting the Z_{ij} values. The Z_{ij}s were then used with the Welch test, F_W. Hence $E_4 10$ signifies transformed values $Z_{ij} = |X_{ij} - \bar{X}_{t_{10}}|$, where $\bar{X}_{t_{10}}$ are group 10% symmetric trimmed means and used with F_W. Since there are four symmetric trimming percentages, there are four variants with this designation.

A total of 165 O'Brien-type procedures were created (see Table 2 for a summary of the O'Brien methods examined):

(J) O'Brien transformation based upon group means and variances used with the usual F-test, W (that is, r_{ij} in W).

(Q1) Variants are designated $Q_1\alpha$. The O'Brien transformation based upon symmetric trimmed means and Winsorized variances of X_{ij}. These trimmed means were calculated at $\alpha =$ 5%, 10%, 15% and 20%. The transformed values, r_{ij}, were used with the usual F-test (that is, r_{ij} in W). Because there are four symmetrical trimming percentages, there are four variants with this designation. Therefore, variants $Q_1 10$ signifies transformation of X_{ij} with 10% symmetric trimmed mean before used with W.

(Q2) Variants are designated $Q_2\alpha_1\alpha_2$. The O'Brien transformation based upon group symmetric trimmed means and Winsorized variances of X_{ij}. These trimmed means and Winsorized variances were calculated at symmetric trimming percentages (5%, 10%, 15% and 20%). The resultant transformed values, r_{ij}, were symmetrically trimmed based on the same percentages used for X_{ij} and used with the robust ANOVA F_t test (that is, symmetrically trimmed r_{ij} in F_t). Because there are four symmetric trimming percentages used twice, there are 16 variants with this designation. Hence, variant $Q_2 515$ signifies transformation of X_{ij} with 5% symmetric trimmed mean and 15% symmetric trimmed mean for the transform value, r_{ij}, before used with F_t.

(Q3) Variants are designated $Q_3\alpha\beta H$. The O'Brien transformation based upon symmetric trimmed means and Winsorized variances of X_{ij}. These trimmed means were calculated with the four symmetric trimming percentages (5%, 10%, 15% and 20%). The resultant transformed values, r_{ij}, were then asymmetrically trimmed at $\beta = 10\%$, 15%, 20% and 25% involving seven hinge estimators and used with the robust ANOVA F_t-test (that is, asymmetrically trimmed r_{ij} in F_t). Because there are four symmetric trimming percentages, on X_{ij}, four asymmetric trimming percentages on r_{ij} with seven hinge estimators, there are 112 variants with this designation.

(Q4) Because variants are designated $Q_4\alpha$. The O'Brien transformation based upon symmetric trimmed means Winsorized variances of X_{ij}. These trimmed means were calculated at the four symmetric trimming percentages (5%, 10%, 15% and 20%). The transformed values, r_{ij}, were then used with the Welch test, F_W (that is, r_{ij} in F_W). Because there are four symmetrical trimming percentages, there are four variants with this designation. Hence, variants $Q_4 05$ signifies transformation of X_{ij} with 5% symmetric trimmed mean before used with F_W.

Variants are designated $K\beta H$. The O'Brien transformation based upon asymmetric trimmed means Winsorized variances of X_{ij}. These trimmed means were calculated at $\beta =$ 10%, 15%, 20% and 25% involving seven hinge estimators (HQ, HQ_1, HH_3, HQ_2, HH_1, HSK_2, HSK_5). The transformed values, r_{ij}, were used with the usual F-test (that is, r_{ij} in W). Because there are four asymmetrical trimming percentages and seven hinge estimators, there are 28 variants with this designation. Hence, variants $K15HSK_2$ signifies transformation of X_{ij} with 15% asymmetric trimmed mean calculated using the HSK_2 hinge estimator before used with W.

Study Conditions

Four variables were employed in the J = 3 study: (a) total sample size; (b) degree of sample size inequality; (c) shape of the population distribution; and (d) type and amount of total trimming.

Total Sample Size

The effect of sample size on the performance of the various procedures was evaluated by varying the total sample size (N). The total sample size was manipulated, setting the average group size to $n_j = 20$ and 40. The average group-sizes correspond to total sample sizes of N = 60 and N = 120.

Degree of Sample Size Inequality

Three conditions of sample size equality/inequality were investigated which are referred to as: equal n_j, moderately unequal n_j and extremely unequal n_j (see below for values). These conditions were evaluated because Keselman, et al. (1998) found that unbalanced designs were more common than balanced designs.

Sample Size Values

n_js
20, 20, 20
15, 20, 25
10, 20, 30
40, 40, 40
35, 40, 45
30, 40, 50

Shape of the Population Distribution

This study investigated distributions ranging from symmetric to skewed and platykurtic to normal-tailed to leptokurtic distributions. In total, seven distributions were employed to compare the procedures. The distributions used were: (i) the Fleishman (1978) transformation of the standard normal distribution into a skewed platykurtic distribution with skewness, γ_1 = 0.5 and kurtosis, γ_2 = –0.5; (ii) a second Fleishman transformation of the standard normal distribution into a skewed normal-tailed distribution with γ_1 = 0.75 and γ_2 = 0; (iii) the Beta (0.5, 0.5) distribution representing symmetric platykurtic distributions; (iv) a g and h distribution (Hoaglin, 1985) where g = h = 0, which is the standard normal distribution with $\gamma_1 = \gamma_2 = 0$; (v) a g = 0 and h = 0.225 long-tailed distribution with $\gamma_1 = 0$ and $\gamma_2 = 154.84$, representing symmetric leptokurtic distributions; (vi) a g = 0.76 and h = –0.098 distribution with skew and kurtosis equal to that of an exponential

Table 1: Description of the Levene (1960) Transformations Used In the Simulations

Designation	Description	Cases
A1	$X_{ij} \to Z_{ij}$: use group means from X_{ij}. Z_{ij}: symmetrically trimmed at tail proportions: 0.05, 0.10, 0.15, 0.20, apply robust ANOVA *F*-test.	4
A2	$X_{ij} \to Z_{ij}$: use group means from X_{ij}. Z_{ij}: asymmetrically trimmed at total proportions: 0.10, 0.15, 0.20, 0.25 and 7 hinge estimators: Q, Q1, H3, Q2, H1, SK2, SK5, apply robust ANOVA *F*-test.	28
A3	$X_{ij} \to Z_{ij}$: use group means from X_{ij}. Z_{ij}: apply Welch F-test.	1
B3	$X_{ij} \to Z_{ij}$: use group medians from X_{ij}. Z_{ij}: apply Welch F-test.	1
E1	$X_{ij} \to Z_{ij}$: use group symmetric trimmed means from X_{ij}. X_{ij} symmetrically trimmed at tail proportions: 0.05, 0.10, 0.15, 0.20, Z_{ij}: apply usual ANOVA *F*-test.	4
E2	$X_{ij} \to Z_{ij}$: use group symmetric trimmed means from X_{ij}. X_{ij} symmetrically trimmed at tail proportions: 0.05, 0.10, 0.15, 0.20, Z_{ij}: symmetrically trimmed at tail proportions: 0.05, 0.10, 0.15, 0.20, apply robust ANOVA *F*-test.	16
E3	$X_{ij} \to Z_{ij}$: use group symmetric trimmed means from X_{ij}. X_{ij} symmetrically trimmed at tail proportions: 0.05, 0.10, 0.15, 0.20, Z_{ij}: asymmetrically trimmed at total proportions: 0.10, 0.15, 0.20, 0.25, keeping hinge estimator constant, apply robust ANOVA F-test.	112
E4	$X_{ij} \to Z_{ij}$: use group symmetric trimmed means from X_{ij}. X_{ij} symmetrically trimmed at tail proportions: 0.05, 0.10, 0.15, 0.20, Z_{ij}: apply Welch F-test.	4
	Total	170

Table 2: Description of the O'Brien (1979) Designations Used In the Simulations

Designation	Description	Cases
J	$X_{ij} \rightarrow R_{ij}$: use group means and variances from X_{ij}. R_{ij}: apply usual ANOVA F-test.	1
Q1	$X_{ij} \rightarrow R_{ij}$: use group symmetric trimmed means and Winsorized variances from X_{ij}. X_{ij} symmetrically trimmed at tail proportions: 0.05, 0.10, 0.15, 0.20, R_{ij}: apply usual F-test.	4
Q2	$X_{ij} \rightarrow R_{ij}$: use group symmetric trimmed means and Winsorized variances from X_{ij}. X_{ij} symmetrically trimmed at tail proportions: 0.05, 0.10, 0.15, 0.20, R_{ij}: symmetrically trimmed at tail proportions: 0.05, 0.10, 0.15, 0.20, apply robust ANOVA F-test.	16
Q3	$X_{ij} \rightarrow R_{ij}$: use group symmetric trimmed means and Winsorized variances from X_{ij}. X_{ij} symmetrically trimmed at tail proportions: 0.05, 0.10, 0.15, 0.20, R_{ij}: asymmetrically trimmed at total proportions: 0.10, 0.15, 0.20, 0.25 and 7 hinge estimators: Q, Q1, H3, Q2, H1, SK2, SK5. Apply robust ANOVA F-test.	112
Q4	$X_{ij} \rightarrow R_{ij}$: use group symmetric trimmed means and Winsorized variances from X_{ij}. X_{ij} symmetrically trimmed at tail proportions: 0.05, 0.10, 0.15, 0.20, R_{ij}: apply Welch F-test.	4
K	$X_{ij} \rightarrow R_{ij}$: use group asymmetric trimmed means and Winsorized variances from X_{ij}. X_{ij} asymmetrically trimmed at total proportions: 0.10, 0.15, 0.20, 0.25 and 7 hinge estimators: Q, Q1, H3, Q2, H1, SK2, SK5. R_{ij}: apply usual F-test.	28
	Total:	165

distribution ($\gamma_1 = 2, \gamma_2 = 6$); and (vii) a g = 0.225 and h = 0.225 distribution, which is also a long-tailed skewed distribution $(\gamma_1 = 4.90, \gamma_2 = 4673.80)$. The last two distributions represent skewed leptokurtic distributions, with (vii) more severe than (vi). These distribution conditions were selected in order to evaluate the operating characteristics of the procedures across a variety of distributions and because they have been examined in other studies (e.g., Algina, Keselman & Penfield, 2007).

The Fleishman (1978) power transformation is of the form Y = a + bZ + cZ^2 + dZ^3, where Z are standard normal variates. Fleishman provided a table of values for the coefficients, b, c, and d that enables the standard normal distribution to be transformed into a nonnormal distribution, also having mean zero and variance one, but with different degrees of skewness and kurtosis. The extra coefficient a is obtained through the relation $a = -c$ as a direct result of constraining $E(Y) = 0$. Two sets of coefficients (b, c, d) were selected from Fleishman (1978) and used in the preceding equation to generate Z_{ij}s from the RANDGEN function (SAS, 2006) with the normal distribution option to produce distributions (i) and (ii). This RANDGEN SAS subroutine allows a user to generate 20 known distributions, both discrete and continuous. Data from the third distribution was also generated using the RANDGEN function but with the beta distribution option. Beta (0.5, 0.5) is a symmetric u-shaped distribution, hence the negative kurtosis.

To generate data from a g and h distribution, standard unit normal variables (Z_{ij}) were converted to g and h distributed random variables via

$$Y_{ij} = \frac{\exp\left(g Z_{ij}\right) - 1}{g} \exp\left(\frac{h Z_{ij}^2}{2}\right), \qquad (20)$$

where both g and h are non-zero. When g is zero

$$Y_{ij} = Z_{ij} \exp\left(\frac{h Z_{ij}^2}{2}\right). \qquad (21)$$

The Z_{ij} scores were generated using the generator RANDGEN with the normal distribution option.

Observations generated for distributions (iii), (v), (vi) and (vii), where the variances were not equal to one, were standardized so that they were one, to reflect the null hypothesis, H$_0$: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$; that is, in the simulations, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$.

Percentages of Total Trimming

Four values of total trimming, namely 10%, 15%, 20% and 25% were examined when data were asymmetrically trimmed, whether to obtain the values used in the transformation of the X_{ij} data or when trimming was carried out on the Levene transformed values Z_{ij} or O'Brien's transformed values, r_{ij} and $r_{t_{ij}}$. Symmetric trimming values of 5%, 10%, 15% and 20% were also investigated. As noted, the literature varies on the amount of recommended (symmetric) trimming and thus these values were chosen to cover the range of values recommended. For each condition 5,000 replications were conducted and the nominal levels of significance for all tests were 0.05 and 0.10.

Results

Tables 4 and 5 summarize the ten best results for the modified Levene (1960) tests for spread. Table 4 shows the average rates of Type I error, the absolute values of the difference between the average rates and 0.05, and the percent of cases falling in three intervals – (0.025, 0.050), (0.045, 0.055), and (0.045, 0.050). The last column indicates total percentage of cases falling in (0.025, 0.055); using simple set theory algebra, this is just percent of cases in (0.025, 0.050) and in (0.045, 0.055) minus percent of cases in (0.045, 0.055). Based on these findings the following are noted:

1. All ten methods examined provided very good Type I error control. Indeed, the

empirical rates ranged from 0.046 to 0.0579; and

2. In order to identify the best method(s) the percentages reported in the last column were relied upon. From this information, B_3 was identified as the best of the Levene (1960) modifications defined and examined.

Table 5 presents Type I error rates for each characteristic of the distributions investigated, as well as the overall rate, indicates that the method that selected as best, contains average Type I errors of 0.048.

The same information is presented in Tables 6 and 7 for the ten best modified O'Brien (1981) tests for spread. Based on these findings, the ten best O'Brien variants provided tight Type I error control ranging from 0.490 to 0.0508. The last column of Table 6 identifies two of the modified procedures, $Q_3 1025HQ_1$ and $K20HH_3$, as the best of the O'Brien (1981) modifications. Table 7 presents Type I error rates for each characteristic of the distributions investigated, as well as the overall rate, and indicates that the both methods that selected as best contain Type I errors averaging 0.050.

Conclusion

This study examined the Type I error rate (for $\alpha = 0.05$) of various modifications of Levene's (1960) and O'Brien's (1981) procedures that could be used to compare variability across groups in independent groups designs, specifically variations not examined by Keselman, et al. (2008). The procedures examined used Levene (1960) or O'Brien (1981) type transformations of the original scores or transformed scores, except as opposed to using the measures of central tendency and variability suggested by Levene and O'Brien, robust measures of central tendency and/or variability were adopted.

The robust values of central tendency and variability (i.e., the trimmed means and Winsorized variances) were based on symmetric or asymmetric trimming rules, that is, rules that either set *a priori* the amount of total trimming or determined empirically the amount to be trimmed from the tails (if at all) based on varied recommendations for total trimming. These approaches were also applied to various test statistics: the ANOVA F-test, a robust F-test (Lee & Fung, 1985), the Welch (1951) test, and bootstrapped versions of these statistics. The procedures were compared under seven distributions when group sizes were equal, moderately, or very unequal. The skewness and kurtosis of the distributions examined varied from the normal distribution ($\gamma_1 = 0$, $\gamma_2 = 0$ respectively) to distributions that were nonnormal, $\gamma_1 = 4.9$ and $\gamma_2 = 4673.80$, respectively).

The procedures were compared on four measures: the average rate of Type I error across the 42 conditions examined, the percentage of empirical Type I errors that fell within the intervals (0.025, 0.05), (0.045, 0.055) and (0.045, 0.05), and the absolute value of the difference between the mean Type I error rate and 0.05. Finally, it should be noted that though it was intended to examine bootstrapped versions of these procedures, this was not pursued because very good Type I error control was achieved without resorting to bootstrapping. Results indicated that the results reported by Keselman, et al. (2008) could not be improved upon with respect to the Levene (1960) test. That is, though the new Levene modifications all worked very well in controlling Type I error rates, they did not result in as many cases falling into the three intervals defined for good Type I error control as reported by Keselman, et al. (2008).

Conversely, two of the O'Brien (1960) modifications did perform well, at least as well as the variants examined by Keselman, et al. (2008) and their recommended Levene variant. These were $Q_3 1025HQ_1$ and $K20HH_3$ with tighter Type I error control and a decent number of cases falling into the three intervals defined for good Type I error control by Keselman, et al. (2008).

Acknowledgements

This research was supported by grants provided by the Social Sciences and Humanities Research Council of Canada and the Fundamental Research Grant Scheme of Malaysia.

Table 4: Type I Error Rates for the 10 Best Performing New Levene's Variants and Percentages of Type I Error Rates within Various Intervals

No.	Variant	Average p-Values	|Mean-0.05|	Percent Within 2.5,5.0	Percent Within 4.5,5.5	Percent Within 4.5,5.0	Total %
1	B_3	.0476	.0024	54.76	30.95	14.29	71
2	E_120	.0531	.0031	47.62	40.48	23.81	64
3	$E_30525HH_1$	.0555	.0055	33.33	28.57	14.29	48
4	$E_31025HH_1$	.0564	.0064	26.19	21.43	4.76	43
5	$E_30525HH_3$	.0566	.0066	38.10	23.81	16.67	45
6	$E_30525HHQ_2$	.0566	.0066	28.57	26.19	9.52	45
7	$E_30525HHQ$	.0568	.0068	33.33	23.81	9.52	48
8	$E_31025HHQ_2$	.0577	.0077	28.57	19.05	7.14	41
9	E_115	.0579	.0079	35.71	47.62	28.57	55
10	$E_31025HHQ$	.0579	.0079	26.19	40.48	21.43	45

Table 5: Type I Error Rates for the 10 Best Performing Variants of Levene's Procedure

No.	Variant	Overall	Skewed	Symmetric	Leptokurtic	Normal Tailed	Platykurtic
1	B_3	.048	.053	.042	.055	.048	.038
2	E_120	.053	.057	.048	.052	.055	.052
3	$E_30525HH_1$	.056	.060	.050	.052	.058	.058
4	$E_31025HH_1$	.056	.059	.053	.052	.058	.062
5	$E_30525HH_3$	.057	.061	.051	.055	.058	.058
6	$E_30525HHQ_2$	.057	.061	.051	.053	.059	.059
7	$E_30525HHQ$	.057	.062	.050	.054	.060	.058
8	$E_31025HHQ_2$	.058	.061	.053	.053	.059	.063
9	E_115	.058	.064	.050	.058	.059	.057
10	$E_31025HHQ$	.058	.063	.051	.058	.058	.058

Table 6: Type I Error Rates for the 10 Best Performing O'Brien (1979) Variants and Percentages of Type I Error Rates within Various Intervals

No.	Variant	Average p-Values	|Mean-0.05|	Percent Within 2.5, 5.0	Percent Within 4.5, 5.5	Percent Within 4.5, 5.0	Total %
1	$Q_32015HH_3$	.0499	.0001	35.71	11.90	9.52	38
2	$Q_31515HH_3$	.0498	.0002	35.71	19.05	9.52	45
3	$K10HQ_1$	.0503	.0003	54.76	23.81	16.67	62
4	$Q_31025HQ_1$	.0503	.0003	57.14	26.19	11.90	71
5	Q_20515	.0504	.0004	52.38	19.05	9.52	62
6	$K20HH_3$	.0496	.0004	54.76	28.57	14.29	69
7	$K10HH_3$	.0506	.0006	54.76	23.81	16.67	62
8	$Q_31010HQ$	.0507	.0007	42.86	28.57	16.67	55
9	$Q_32015HQ$	.0508	.0008	33.33	16.67	9.52	40
10	$K25HQ_1$	.0490	.0010	54.76	9.52	9.52	55

Table 7: Type I Error Rates for the 10 Best Performing Variants of O'Brien's (1979) Procedure

No.	Variant	Overall	Skewed	Symmetric	Leptokurtic	Normal Tailed	Platykurtic
1	$Q_3 2015 HH_3$	.050	.051	.049	.035	.056	.066
2	$Q_3 1515 HH_3$	.050	.051	.049	.036	.058	.063
3	$K10HQ_1$	.050	.056	.043	.045	.056	.053
4	$Q_3 1025 HQ_1$	.050	.052	.048	.045	.059	.051
5	$Q_2 0515$	.050	.054	.046	.043	.059	.053
6	$K20HH_3$	.050	.055	.042	.043	.058	.052
7	$K10HH_3$	.051	.057	.043	.046	.056	.053
8	$Q_3 1010 HQ$	.051	.052	.049	.040	.058	.059
9	$Q_3 2015 HQ$	.051	.053	.048	.038	.059	.062
10	$K25HQ_1$	.049	.058	.038	.042	.058	.052

References

Algina, J., Keselman, H. J., & Penfield, R. D. (2007). Confidence intervals for an effect size measure in multiple linear regression. *Educational and Psychological Measurement*, *67*, 207-218.

Babu, J. G., Padmanabhan A. R., & Puri, M. P. (1999). Robust one-way ANOVA under possibly non-regular conditions. *Biometrical Journal*, *41*(*1978*), 321-339.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, *43*, 521-532.

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

Hoaglin, D. C. (1985). Summarizing shape numerically: The g- and h distributions. In D. Hoaglin, F. Mosteller, and J. Tukey (Eds), *Exploring data tables, trends, and shapes*, 461-513. New York, NY: Wiley.

Hogg, R. V. (1974). Adaptive robust procedures: A partial review and some suggestions for future applications and theory. *Journal of the American Statistical Association*, *69*, 909-927.

Hogg, R. V. (1982). On adaptive statistical inferences. *Communications in Statistics: Theory and Methods*, *11*, 2531-2542.

Huber, P. J. (1970). Studentizing robust estimates. In M. L. Puri, (Ed.) *Nonparametric techniques in statistical inference*. London, England: Cambridge University Press.

Huber, P. J. (1977). Discussion. *The Annals of Statistics*, *5*, 1090-1091.

Keselman, H. J., Algina, J., & Fradette, K. (2005). Robust confidence intervals for effect size in the two-group case. *Journal of Modern Applied Statistical Methods*, *4*, 353-371.

Keselman, H. J., Games, P. A., & Clinch, J. J. (1979). Tests for homogeneity of variance. *Communications in Statistics, Simulation and Computation*, *B8*(*2*), 113-129.

Keselman, H. J., et al. (1998). Statistical practices of Educational Researchers: An analysis of their ANOVA, MANOVA and ANCOVA analyses. *Review of Educational Research*, *68*(*3*), 350-386.

Keselman, H. J., Wilcox, R. R., Algina, J., Othman, A. R., & Fradette, K. (2008). A comparative study of robust tests for spread: Asymmetric trimming strategies. *British Journal of Mathematical and Statistical Psychology*, *61*, 235-253.

Keselman, H. J., Wilcox, R. R., Lix, L. M., Algina, J., & Fradette, K. (2007). Adaptive robust estimation and testing. *British Journal of Mathematical and Statistical Psychology*, *60*, 267-293.

Keselman, H. J., Wilcox, R. R., Othman, A. R., & Fradette, K. (2002). Trimming, transforming statistics, and bootstrapping: Circumventing the biasing effects of heteroscedasticity and nonnormality. *Journal of Modern Applied Statistical Methods*, *1*, 288-309.

Lee, H., & Fung, K. Y. (1985). Behaviour of trimmed F and sine-wave F statistics in one-way ANOVA. *Sankhya*: *The Indian Journal of Statistics*, *47*, *Series B*, 186-201.

Levene, H. (1960). Robust tests for the equality of variances. In I. Olkin (Ed.), *Contributions to probability and statistics*. Palo Alto, CA: Stanford University Press.

Mudholkar, A., Mudholkar, G. S., & Srivastava, D. K. (1991). A construction and appraisal of pooled trimmed-t statistics. *Communications in Statistics: Theory and Methods*, *20*, 1345-1359.

O'Brien, R. G. (1979). A general ANOVA method for robust tests of additive models for variances, *Journal of the American Statistical Association*, *74*, 877-880.

O'Brien, R. G. (1981). A simple test for variance effects in experimental designs. *Psychological Bulletin*, *89*, 570-574.

Olejnik, S. F., & Algina, J. (1987). Type I error rates and power estimates of selected parametric and nonparametric tests of scale. *Journal of Educational Statistics*, *12*, 45-61.

Reed III, J. F. (1998). Contributions to adaptive estimation. *Journal of Applied Statistics*, *25*, 651-669.

Reed III, J. F., & Stark, D. B. (1996). Hinge estimators of location: Robust to asymmetry. *Computer Methods and Programs in Biomedicine*, *49*, 11-17.

Rosenberger, J. L., & Gasko, M. (1983). Comparing location estimators: Trimmed means, medians and trimean. In D. Hoaglin, F. Mosteller and J. Tukey, (Eds.), *Understanding robust and exploratory data analysis*, 297-336. New York, NY: Wiley.

Rousseeuw, P. J., & van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, *85*, 633-639.

SAS Institute Inc. (2006). *SAS OnlineDoc® 9.1.3*. Cary, NC: SAS Institute Inc.

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

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

Tukey, J. W., & McLaughlin, D. H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/Winsorization. *Sankhya*: *The Indian Journal of Statistics*, *25*, *Series A*, 331-352.

Welch, B. L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika*, *38*, 330-336.

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

Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (*2nd Ed*.). San Diego, CA: Academic Press.

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