

5-1-2006

Multiple Comparison Procedures, Trimmed Means And Transformed Statistics

Rhonda K. Kowalchuk

Southern Illinois University Carbondale, rkowal@siu.edu

H. J. Keselman

University of Manitoba, kesel@ms.umanitoba.ca

Rand R. Wilcox

University of Southern California, rwilcox@usc.edu

James Algina

University of Florida, algina@ufl.edu

James Algina

University of Florida, algina@ufl.edu

See next page for additional authors

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Kowalchuk, Rhonda K.; Keselman, H. J.; Wilcox, Rand R.; Algina, James; Algina, James; and Algina, James (2006) "Multiple Comparison Procedures, Trimmed Means And Transformed Statistics," *Journal of Modern Applied Statistical Methods*: Vol. 5 : Iss. 1 , Article 6.

Multiple Comparison Procedures, Trimmed Means And Transformed Statistics

Authors

Rhonda K. Kowalchuk, H. J. Keselman, Rand R. Wilcox, James Algina, James Algina, and James Algina

Multiple Comparison Procedures, Trimmed Means And Transformed Statistics

Rhonda K. Kowalchuk
Southern Illinois University Carbondale

H. J. Keselman
University of Manitoba

Rand R. Wilcox
University of Southern California

James Algina
University of Florida

A modification to testing pairwise comparisons that may provide better control of Type I errors in the presence of non-normality is to use a preliminary test for symmetry which determines whether data should be trimmed symmetrically or asymmetrically. Several pairwise MCPs were investigated, employing a test of symmetry with a number of heteroscedastic test statistics that used trimmed means and Winsorized variances. Results showed improved Type I error control than competing robust statistics.

Key words: Multiple comparison procedures, trimmed estimators, symmetric and asymmetric trimming, heteroscedastic test statistic, nonnormality, variance heterogeneity.

Introduction

Pairwise multiple comparison procedures (MCPs) are adversely affected by nonnormality, particularly when variances are heterogeneous and group sizes are unequal (Keselman, Lix, & Kowalchuk, 1998). Specifically, Type I errors are liberal, resulting in spurious rejections of null hypotheses. The deleterious effects of nonnormality on rates of Type I error are, for the most part, attributable to asymmetry of distributions, that is, to skewness (Westfall & Young, 1993). These results are predictable on theoretical grounds. Cressie and Whitford (1986) showed that Student's two-sample t test is not asymptotically correct when the group distributions have unequal third cumulants and sample sizes are unequal; therefore, Type I error

inflation is expected. In the one-way independent groups problem, Keselman, Lix, et al. (1998) found Type I error rates for popular pairwise MCPs approached .21 ($\alpha = .05$) when data were obtained from skewed distributions where group variances and sample sizes were unequal and negatively paired with one another.

One potential solution to this Type I error inflation is to replace the usual least squares estimators with estimates which are less influenced by the effects of nonnormality. Indeed, many investigators have shown that better results can be obtained by using statistics designed for heterogeneity combined with robust estimators of central tendency and variability (see Keselman, Kowalchuk, & Lix, 1998; Lix & Keselman, 1998; Wilcox, Keselman, & Kowalchuk, 1998; Yuen, 1974). For example, Keselman, Lix et al. (1998) found that the methods due to Ryan (1960), Welsch (1977), Peritz (1970), Shaffer (1979; 1986), Hayter (1986), and Hochberg (1988) provided much better Type I error control, typically having rates less than .075 when based on a heteroscedastic statistic with trimmed means and Winsorized variances. Though rates improved, these methods were, nonetheless, still occasionally affected when distributions were nonnormal, variances were heterogeneous, and group sizes

Rhonda K. Kowalchuk (rkowal@siu.edu) is an Assistant Professor of Educational Psychology. H. J. Keselman (kesel@ms.umanitoba.ca) is a Professor of Psychology. Rand R. Wilcox (rwilcox@usc.edu) is a Professor of Psychology. James Algina (algina@ufl.edu) is a Professor of Educational Psychology.

were unequal. That is, rates occasionally exceeded .075.

An approach that may provide improved Type I error control for tests of trimmed mean equality (pairwise) is to use a preliminary test for symmetry which determines whether data should be trimmed symmetrically or asymmetrically. Keselman, Wilcox, Othman, and Fradette (2002) found that by using a test for symmetry in conjunction with a test for equality of trimmed means, Type I error rates were well controlled when data were extremely heterogeneous and nonnormal in a one-way independent groups design. The test of symmetry investigated was first proposed by Hogg, Fisher, and Randles (1975) and later modified by Babu, Padmanabhan and Puri (1999). Specifically, two indices are computed, one that determines tail thickness and the other symmetry of the underlying distribution. The calculations determine whether a test of mean equality is based on symmetrically or asymmetrically trimmed means (see Othman, Keselman, Wilcox, Fradette, & Padmanabhan, 2002, for details of the test of symmetry).

Keselman, Lix, et al. (1998) symmetrically trimmed 20% of the data per group and used an approximate degrees of freedom Welch (1938) test statistic for the pairwise comparisons. Although, 20% symmetric trimming is recommended (Wilcox, 1995), theory would imply that asymmetric trimming would be more appropriate when data are skewed (Keselman et al., 2002; Othman et al., 2002). The rationale behind asymmetric trimming is to remove more of the offending data (i.e., data that does not represent the bulk of the observations, that is, the 'typical' score) from the tail containing more of the outlying values. Keselman et al. (2002) found other percentages of trimming, either symmetrically or asymmetrically, resulted in better Type I error control than uniformly adopting 20% symmetric trimming. For example, 15% symmetric trimming or 15% asymmetric trimming resulted in fewer non-robust values compared to always adopting 20% symmetric trimming.

In addition, Keselman et al. (2002) found that transformations (i.e., Johnson, 1978; Hall, 1992) of the Welch-James heteroscedastic statistic improved Type I error control. The

Johnson and Hall transformations are intended to remove the bias due to skewness. This is consistent with Guo and Luh (2000) and Luh and Guo (1999) who found that transformations of the Welch-James statistic improved its performance when trimmed means were used and distributions were skewed and heavy-tailed. As well, Keselman et al. (2002) found improved Type I error control when the transformed heteroscedastic statistics were preceded by a test of symmetry under extreme conditions of nonnormality and variance heterogeneity in a one-way independent groups design. Thus, the purpose of this article was to investigate whether these procedures would be beneficial in the pairwise multiple comparison problem.

Test of Symmetry

Othman et al. (2002) provided the details for the test of symmetry, a test based on the work of Hogg et al. (1975) and Babu et al. (1999). Essentially, two indices are computed, one index (Q_2) determines tail-weight (light or heavy) while the other index (Q_1) determines the symmetry of an underlying distribution. The value of the Q_2 index classifies a distribution as normal-tailed, heavy-tailed, or very heavy-tailed which then determines the number of sample points to be used in the computation of the Q_1 index. If the distribution is determined to be (a) normal-tailed, then all sample points are used, (b) heavy-tailed, then the top and bottom 10% of sample points are trimmed, or (c) very heavy-tailed, then the top and bottom 20% of sample points are trimmed. That is, the value of the Q_1 index determines the symmetry/asymmetry of a distribution (i.e., left skewed, symmetric or right skewed) which then determines the type of trimming (symmetric vs asymmetric). Keselman et al. (2002) provided a SAS/IML (1989) program to compute the test of symmetry.

Robust Estimation

Robust estimates of central tendency and variability were applied to heteroscedastic statistics. Specifically, trimmed means and Winsorized variances were used in order to test the hypothesis of the equality of population trimmed means in the pairwise multiple comparison problem. Let

$$Y_{(1)j} \leq Y_{(2)j} \leq \dots \leq Y_{(n_j)j}$$

represent the ordered observations associated with the j th ($j=1, \dots, J$) group, where n_j is the sample size in the j th group. Let

$$g_j = \lceil \gamma n_j \rceil$$

where γ represents the proportion of observations to be trimmed in each tail of the distribution and $\lceil x \rceil$ is the greatest integer $\leq x$. The effective sample size for the j th group becomes $h_j = n_j - 2g_j$. The j th sample trimmed mean is

$$\hat{\mu}_{tj} = \frac{1}{h_j} \sum_{i=g_j+1}^{n_j-g_j} Y_{(i)j}. \quad (1)$$

The sample Winsorized mean is necessary in order to compute the Winsorized variance. The j th sample Winsorized mean is

$$\hat{\mu}_{wj} = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ij}, \quad (2)$$

where

$$X_{ij} = \begin{cases} Y_{(g_j+1)j} & \text{if } Y_{ij} \leq Y_{(g_j+1)j} \\ Y_{ij} & \text{if } Y_{(g_j+1)j} < Y_{ij} < Y_{(n_j-g_j)j} \\ Y_{(n_j-g_j)j} & \text{if } Y_{ij} \geq Y_{(n_j-g_j)j}. \end{cases}$$

The sample Winsorized variance is required in order to get a valid estimate of the standard error of a trimmed mean. The sample Winsorized variance for the j th group is

$$\hat{\sigma}_{wj}^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (X_{ij} - \hat{\mu}_{wj})^2$$

and the estimated standard error of the trimmed mean is

$$\sqrt{(n_j - 1) \hat{\sigma}_{wj}^2 / [h_j (h_j - 1)]}.$$

Under asymmetric trimming, and assuming that the distribution is positively (right) skewed so that observations in the upper tail of the distribution are trimmed, the effective sample size for the j th group becomes $h_j = n_j - g_j$. The j th sample trimmed mean is

$$\hat{\mu}_{tj} = \frac{1}{h_j} \sum_{i=1}^{n_j-g_j} Y_{(i)j} \quad (3)$$

and the j th sample Winsorized mean is

$$\hat{\mu}_{wj} = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ij}, \quad (4)$$

where

$$X_{ij} = \begin{cases} Y_{ij} & \text{if } Y_{ij} < Y_{(n_j-g_j)j} \\ Y_{(n_j-g_j)j} & \text{if } Y_{ij} \geq Y_{(n_j-g_j)j}. \end{cases}$$

The sample Winsorized variance is computed based on the previous equation with the new definition of $\hat{\mu}_{wj}$ and the estimated standard error of the trimmed mean is also computed based on the previous equation with the new definitions of h_j and $\hat{\sigma}_{wj}^2$.

Definitions of the Heteroscedastic Statistics

Johanson's (1980) Welch-James (WJ)-type heteroscedastic statistic (see Lix & Keselman, 1995) with robust estimators has been found to obtain better Type I error control than the WJ statistic with least squares estimators in independent groups designs under nonnormality and variance heterogeneity (see Guh & Luh, 2000; Keselman, Kowalchuk, et al., 1998; Keselman, Lix, et al., 1998; Lix & Keselman, 1998; Luh & Guo, 1999; Wilcox et al. 1998). Guo and Luh (2000) found that two transformations of the WJ statistic combined with the use of trimmed means and Winsorized

variances resulted in better Type I error control than the WJ statistic with trimmed means and without a transformation for various skewed and heavy-tailed distributions. Specifically, Johnson's (1978) or Hall's (1992) transformations of the WJ statistic are intended to remove skewness. Hence, the transformations contend with skewness, trimmed means contend with heavy tails, and a heteroscedastic statistic contends with variance heterogeneity (Luh & Guo, 1999).

In the present study, both transformations of the WJ statistic for removing skewness were investigated along with the nontransformed WJ statistic. Let $\hat{\mu}_{tj}$, $\hat{\mu}_{wj}$, $\hat{\sigma}_{wj}^2$ and h_j be the trimmed mean, Winsorized mean, Winsorized variance, and trimmed sample size, respectively, for group j . The third central Winsorized moment of the j th group is

$$\hat{\mu}_{3j} = \frac{1}{n_j} \sum_{i=1}^{n_j} (X_{ij} - \hat{\mu}_{wj})^3.$$

Let

$$\tilde{\sigma}_{wj}^2 = \frac{(n_j - 1)}{(h_j - 1)} \hat{\sigma}_{wj}^2, \quad \tilde{\mu}_{wj} = \frac{n_j}{h_j} \hat{\mu}_{3j},$$

$$q_j = \frac{\tilde{\sigma}_{wj}^2}{h_j}, \quad w_{tj} = \frac{1}{q_j}, \quad U_t = \sum_{j=1}^J w_{tj},$$

$$\text{and } \hat{\mu}_t = \frac{1}{U_t} \sum_{j=1}^J w_{tj} \hat{\mu}_{tj}.$$

Luh and Guo (1999) defined Johnson's (1978) transformed trimmed mean statistic as

$$T_{(Johnson)j} = (\hat{\mu}_{tj} - \hat{\mu}_t) + \frac{\tilde{u}_{wj}}{6\tilde{\sigma}_{wj}^2 h_j} + \frac{\tilde{u}_{wj}}{3\tilde{\sigma}_{wj}^4} (\hat{\mu}_{tj} - \hat{\mu}_t)^2. \quad (5)$$

From Guo and Luh (2000), Hall's (1992) transformed trimmed mean statistic can be defined as:

$$T_{(Hall)j} = (\hat{\mu}_{tj} - \hat{\mu}_t) + \frac{\tilde{u}_{wj}}{6\tilde{\sigma}_{wj}^2 h_j} + \frac{\tilde{u}_{wj}}{3\tilde{\sigma}_{wj}^4} (\hat{\mu}_{tj} - \hat{\mu}_t)^2 + \frac{\tilde{u}_{wj}}{27\tilde{\sigma}_{wj}^8} (\hat{\mu}_{tj} - \hat{\mu}_t)^3. \quad (6)$$

Keselman, Wilcox, and Lix (2003) indicated that sample trimmed means, sample Winsorized variances, and trimmed sample sizes can be used to compute the WJ statistic. That is,

$$WJ = \sum_{j=1}^J w_{tj} (\hat{\mu}_{tj} - \hat{\mu}_t)^2, \quad (7)$$

which, when divided by c , is distributed as an F variable with degrees of freedom equal to $J - 1$ and

$$v = (J^2 - 1) \left[3 \sum_{j=1}^J \frac{(1 - w_{tj}/U_t)^2}{h_j - 1} \right]^{-1},$$

where

$$c = (J - 1) \left(1 + \frac{2(J - 2)}{J^2 - 1} \sum_{j=1}^J \frac{(1 - w_{tj}/U_t)^2}{h_j - 1} \right).$$

Thus, the transformed WJ statistics may be defined as

$$JWJ = \sum_{j=1}^J w_{tj} (T_{(Johnson)j})^2 \quad (8)$$

and

$$HWJ = \sum_{j=1}^J w_{tj} (T_{(Hall)j})^2. \quad (9)$$

When Johnson's transformed WJ statistic (JWJ) and Hall's transformed WJ statistic (HWJ) are divided by c , they are also distributed as F variates with no change in degrees of freedom.

The WJ, JWJ, and HWJ statistics were used not only for the omnibus test, if one was required, but for the pairwise tests for each of the MCPs investigated.

Multiple Comparison Methods

The MCPs investigated, adopt stepwise testing for controlling the overall (familywise) rate of Type I error. Specifically, the MCPs examined were the: (a) Ryan (1960)-Welsch (1977) multiple range procedure, (b) Peritz (1970) procedure, (c) Shaffer (1986) sequentially rejective Bonferroni procedure, (d) Shaffer (1986) sequentially rejective Bonferroni procedure that begins with an omnibus test, (e) Hochberg (1988) step-up sequentially acceptive Bonferroni procedure, (f) multiple range procedure that begins with an omnibus test (see Shaffer 1979; 1986), and (g) Hayter (1986) two-stage modified least significant difference (LSD) procedure. These MCPs were previously investigated by Keselman, Lix, et al. (1998).

The Ryan (1960) and Welsch (1977) multiple range procedure begins by examining the J range, and steps down to examine successively smaller ranges only when a larger range test is declared significant. The designation q is used to denote this MCP. According to Ryan and Welsch, the overall rate of Type I error is controlled at α (when assumptions are satisfied) for a set of p ($p = 2, \dots, J$) means if each test is assessed for significance at a level equal to

$$\alpha_p = 1 - (1 - \alpha)^{p/J} \quad [2 \leq p \leq J - 2],$$

$$\alpha_{J-1} = \alpha_J = \alpha.$$

The Peritz (1970) procedure follows the same step-down logic of the usual range procedure, but assesses significance with Newman (1939), Keuls (1952), and/or Ryan-Welsch critical values. This MCP is designated PER. Shaffer's (1986) sequentially rejective Bonferroni procedure uses probability (p) values in assessing the pairwise hypotheses taking into account the number of hypotheses rejected at earlier stages in the sequence of testing in arriving at decisions regarding significance. The abbreviation for this MCP is SRB.

Shaffer's (1986) sequentially rejective Bonferroni procedure begins with an omnibus test (i.e., WJ, JWJ, HWJ), and if rejected, assesses significance of the pairwise comparisons by taking into account the number of true pairwise hypotheses remaining given previous rejections. Because three omnibus statistics are being investigated, there are three SRB MCPs and they are designated as WJ/SRB, JWJ/SRB, and HWJ/SRB.

Hochberg's (1988) step-up sequentially acceptive Bonferroni procedure uses the p values associated with the pairwise tests to arrive at accept-reject decisions; these are determined sequentially and hypotheses can be rejected by implication. Hochberg's MCP is designated as HOCH. Another set of MCPs were based on the modified range procedure due to Shaffer (1979; 1986), which starts with an omnibus test and only upon rejection, moves on to test range hypotheses with Ryan-Welsch critical values, modifying the J-range critical value to one based on J-1 means. The abbreviations of these three (stage 1 omnibus) Shaffer MCPs are WJ/q, JWJ/q, and HWJ/q. Lastly, Hayter's (1986) modified LSD begins with an omnibus test, which if rejected leads to the Stage 2 tests of the pairwise comparisons using a Studentized range critical value for J-1 means. The three MCPs based on Hayter's method are designated: WJ/HAY, JWJ/HAY, and HWJ/HAY. Detailed descriptions of all the pairwise MCPs can be found in the original references.

Methodology

Seven pairwise MCPs were compared in terms of Type I error control under conditions of nonnormality and variance heterogeneity in one-way independent groups designs. Variables that were examined by Keselman, Lix, et al. (1998) were chosen for investigation. Eight variables were manipulated in the present study: (a) number of groups (3 and 6), (b) sample size (equal or not equal), (c) degree/pattern of variance heterogeneity [moderate and large/all (mostly) unequal and all but one equal], (d) pairing of groups sizes and variances, (e) type of nonnormal population distribution, (f) method of computing a test of symmetry, (g) percentage of

trimming, and (h) type of heteroscedastic statistic.

One-way independent groups designs containing three and six groups to evaluate the effect of number of pairwise comparisons on Type I error were chosen for investigation. That is, for the former case, only three pairwise comparisons were tested, whereas, in the latter case, 15 pairwise comparisons were tested.

The sample sizes in each of the groups were either equal or unequal. When equal, $C = 0$, and when unequal, $C = .163$ and $.327$, where C denotes a coefficient of group size variation defined as

$$\left(\sum_j (n_j - \bar{n})^2 / J \right)^{1/2} / \bar{n}, \text{ where } \bar{n}$$

is the average group size. When equal, group sizes were set at 20 in both the $J = 3$ and $J = 6$ designs. When unequal, and for the $J = 3$ design, the two cases of group size inequality were 16, 20, 24 ($C = .163$) and 12, 20, 28 ($C = .327$), while for the $J = 6$ design, the group sizes were 16, 16, 20, 20, 24, 24 ($C = .163$) and 12, 12, 20, 20, 28, 28 ($C = .327$).

Two patterns of variance heterogeneity were examined: (a) all (most) variances unequal (Pattern 1) and (b) all variances equal but one (Pattern 2). When $J = 3$, Pattern 1 was 1, 9, 16 and Pattern 2 was 1, 1, 16. The patterns for $J = 6$ were, respectively, 1, 1, 4, 9, 9, 16, and 1, 1, 1, 1, 1, 16.

Seven cases of group sizes and variances pairings were investigated. Group sizes were both equal and unequal and paired with equal and unequal variances. Specifically, the combinations were: (a) equal n_j ; equal σ_j^2 , (b/b') equal n_j ; unequal σ_j^2 , (c/c') unequal n_j ; unequal σ_j^2 (positively paired), (d/d') unequal n_j ; unequal σ_j^2 (negatively paired). The b/c/d notation represents the Pattern 1 variance conditions, whereas the b'/c'/d' notation represents the Pattern 2 variance conditions. Considering the group size and variance inequalities, there were a total of eleven combinations.

To examine distributional shape, four nonnormal distributions with varying degrees of skewness (γ_1) and kurtosis (γ_2) were chosen for investigation. A chi-square (χ^2) distribution and three g- and h-distributions (Hoaglin, 1985) were selected. Specifically, the four nonnormal distributions were: (a) $\chi_{(3)}^2$ distribution ($\gamma_1 = 1.63$, $\gamma_2 = 4.00$); (b) $g = .5$ and $h = 0$ distribution ($\gamma_1 = 1.75$, $\gamma_2 = 8.9$); (c) $g = 1$ and $h = 0$ distribution ($\gamma_1 = 6.2$, $\gamma_2 = 114$); and (d) $g = .25$ and $h = .25$ distribution (γ_1 and γ_2 undefined). The three g- and h- distributions are hereafter notated as ($g = .5$, $h = 0$), ($g = 1$, $h = 0$), and ($g = .25$, $h = .25$), respectively. These nonnormal distributions were selected because educational and psychological research data are typically skewed and/or heavy-tailed (Micceri, 1989; Wilcox, 1990).

To generate pseudorandom variates having a chi-square (χ^2) distribution with 3 degrees of freedom, three standard normal variates were squared and summed. The variates were transformed to $\chi_{(3)}^2$ variates having mean μ_j (population trimmed mean) and σ_j^2 (see Hastings & Peacock, 1975, p. 46-51, for further details). To generate data from a g- and h-distribution, standard unit normal variables (Z) were converted to the random variable

$$X_{ij} = \frac{\exp(gZ_{ij}) - 1}{g} \exp\left(\frac{hZ_{ij}^2}{2}\right), \quad (10)$$

according to the values of g and h selected for investigation. μ_j was subtracted from each observation. To obtain a distribution with standard deviation σ_j , each transformed X_{ij} ($j = 1, \dots, J$) was then multiplied by a value of σ_j . The standard deviation of a g- and h-distribution is not equal to one, and thus the values for the variances/standard deviations reflect the ratio of the variances/standard deviations between the groups (see Wilcox, 1994). Each population distribution was empirically generated and the indices of tail weight and symmetry were computed in order to determine whether the

population trimmed mean used for centering should be based on symmetric or asymmetric (e.g., right tailed) trimming for the percentage of trimming cases investigated.

Three approaches to computing the test of symmetry were examined by calculating the indices (Q_1 and Q_2) within each group and then: (a) using a weighted mean of the indices across all groups to determine the type of trimming for every group (average estimate; see Othman et al. 2002); (b) using the value for each particular group to determine the type of trimming for that group (individual estimate), and (c) using a weighted mean of the indices across two groups to determine the type of trimming for the groups involved in each particular comparison (pairwise estimate). The test of symmetry based on pairwise estimates could not be applied to an omnibus test, so only the MCPs that do not require an omnibus test were considered for this approach. In addition, the pairing of groups had to be predetermined in order to compute the weighted mean of the indices across the two groups in each pairwise comparison and this prevented the use of the approach with the range MCPs. Thus, the third approach was applied to only the SRB and HOCH procedures. The three approaches to symmetric/asymmetric trimming were compared to always adopting symmetric trimming. The Q_1 and Q_2 indices determine whether symmetrically/asymmetrically trimmed means for each group were used in the pairwise MCPs. For those MCPs that require an omnibus test, the same approach to trimming (i.e., average estimate, individual estimate or symmetric trimming) was adopted for the omnibus and the pairwise tests.

The following combinations of symmetric and asymmetric trimming percentages were investigated: (a) either 10% symmetric or 20% asymmetric trimming (10/20), (b) either 15% symmetric or 30% asymmetric trimming (15/30), (c) either 20% symmetric or 40% asymmetric trimming (20/40), (d) either 10% symmetric or 10% asymmetric trimming (10/10), (e) either 15% symmetric or 15% asymmetric trimming (15/15), and (f) either 20% symmetric or 20% asymmetric trimming (20/20). As well, symmetrically trimming 10%, 15%, and 20% of the data was investigated. Hence, the various

combinations of trimming percentages were chosen to evaluate whether there would be an optimal proportion of trimming.

Three heteroscedastic statistics were examined: (a) Welch-James statistic (WJ), (b) Johnson's (1978) transformation of WJ (JWJ), and (c) Hall's (1992) transformation of WJ (HWJ) (see Guo & Luh, 2000; Keselman et al. 2002; Luh & Guo, 1999). The seven pairwise MCPs were computed with each of the heteroscedastic statistics, resulting in a total of 21 pairwise MCPs.

Type I error rates were based on five thousand replications using a .05 level of significance for the complete null hypothesis.

Results

Bradley's (1978) liberal criterion of robustness to assess Type I error rates was chosen. That is, if an empirical estimate of Type I error ($\hat{\alpha}$) was contained within the interval of $.5\alpha \leq \hat{\alpha} \leq 1.5\alpha$, then the procedure was considered robust. For a significance level of .05, the interval is $.025 \leq \hat{\alpha} \leq .075$. If the Type I error was not contained in this interval, then a procedure was considered nonrobust for that particular condition. In the tables, bold entries correspond to these latter values.

Because of the large number of MCPs investigated and the form of assumption violations examined, only the mean Type I error rates (percentages), averaging across the eleven combinations of group sizes, and variances were tabled. Plus and minus symbols next to the tabled error rates are used to identify whether the minimum to maximum range of Type I error rates across the eleven combinations contained a conservative (-) value, a liberal (+) value, or both conservative and liberal (\pm) values. A conservative value is defined as an error rate below Bradley's lower limit (2.50%) and a liberal value is defined as an error rate above Bradley's upper limit (7.50%). Because of space considerations and the similar pattern of results for the chi-square and ($g = .5$, $h = 0$) distributions, only the latter are tabled.

$J = 3$

Tables 1 through 3 contain the summary percentages for the ($g = .5, h = 0$), ($g = 1, h = 0$), and ($g = .25, h = .25$) distributions, respectively. When the number of groups is equal to three, a few of the MCPs investigated are identical. Specifically, the Hayter (1986) two-stage and Shaffer (1986) sequentially rejective Bonferroni procedure that begins with an omnibus test are identical (denoted as WJ/*, JWJ/*, and HWJ/* in Tables 1 through 3). Additionally, the Ryan (1960)–Welsch (1977) and Peritz (1970) procedures are identical (denoted as q / PER in Tables 1 through 3).

$g = .5$ and $h = 0$ Distribution

When data were obtained from this particular nonnormal distribution, all MCPs were robust when preceded by the symmetry test with 10/10 symmetric/asymmetric trimming where the indices of tail weight and symmetry were averaged over all groups and under the 10%, 15%, and 20% symmetric trimming cases (see Table 1). The chi-square distribution had a similar pattern of results, however all MCPs were also robust under the 15/15 and 20/20 symmetric/asymmetric trimming where the indices were averaged over all groups. MCPs preceded by the test of symmetry generally had mean Type I error rates closer to the nominal 5% level compared to the strategy of always adopting symmetric trimming. For the symmetry test based on averaging (tail-weight and symmetry) indices across all groups, the mean error rates across robust MCPs were 4.83%, 4.80%, and 5.22% for the 10/10, 15/15, and 20/20 trimming cases, respectively and for the symmetry test based on the indices taken per group, the mean error rate across robust MCPs was 5.32% for the 10/10 trimming case. For the symmetric trimming conditions of 10%, 15%, and 20%, the mean error rates across MCPs were 4.75%, 4.68%, and 4.80%, respectively. In addition, the general pattern for MCPs preceded by a test for symmetry was for error rates to increase as the proportion of trimming increased (i.e., from 10/20 to 15/30 to 20/40 and from 10/10 to 15/15 to 20/20).

The MCPs based on the WJ statistic generally had more conservative error rates than the same MCPs based on the modified WJ

statistics (i.e., JWJ and HWJ), when preceded by a test of symmetry, a pattern opposite to that observed for the symmetric trimming cases. For example, under the 10/10 trimming case preceded by the test of symmetry based on indices (tail weight and symmetry) averaged across all groups, the mean error rates for the MCPs based on the WJ, JWJ, and HWJ statistics were equal to 4.70%, 4.87%, and 4.91%, respectively. However, when adopting 20% symmetric trimming, the mean error rates across MCPs based on the WJ, JWJ, and HWJ statistics were equal to 4.94%, 4.73%, and 4.74%, respectively. For the chi-square distribution, regardless of whether the MCPs were preceded by a test of symmetry, the MCPs based on the JWJ and HWJ statistics generally had more conservative Type I error rates than the corresponding MCPs based on the WJ statistic.

The mean error rates for the SRB and HOCH procedures based on symmetric trimming were more conservative than when the MCPs were preceded by a test of symmetry. When the test of symmetry was based on individual group estimates of tail weight and symmetry, the MCP's mean error rates were highest, and decreased when the test was based on pairwise estimates and further decreased when the symmetry test was based on average estimates across groups (a result consistent with that obtained for the chi-square distribution). Noteworthy is that the error rates for the SRB and HOCH MCPs fell within Bradley's (1978) limits for the 10/10 trimming percentage regardless of the method of computing the test for symmetry; a result consistent with that obtained for the chi-square distribution. An optimal strategy is to use a test of symmetry with either pairwise estimates or average estimates across groups with either 10/10 or 15/15 symmetric/asymmetric trimming.

$g = 1$ and $h = 0$ Distribution

The use of the test of symmetry resulted in improved Type I error control when data were obtained from the ($g = 1, h = 0$) nonnormal distribution (see Table 2). That is, the MCPs with conservative and/or liberal error rates based on symmetric trimming became either robust or closer to Bradley's (1978) limits when preceded

Table 1. Summary Percentages of Type I Error for Multiple Comparison Procedures (J = 3; g = .5, h = 0 Distribution)

	Average Estimate						Individual Estimate					
	10/20	15/30	20/40	10/10	15/15	20/20	10/20	15/30	20/40	10/10	15/15	20/20
q / PER (WJ)	5.45+	7.29+	9.56+	4.07	4.50	5.00	11.58+	19.96+	29.43+	4.90	6.58+	9.10+
WJ / q (WJ)	6.37+	8.50+	11.03+	4.64	5.06	5.64	13.28+	23.15+	33.42+	5.57	7.51+	10.43+
WJ / * (WJ)	7.48+	9.73+	12.42+	5.70	6.10+	6.73+	15.50+	26.68+	38.03+	6.72+	8.99+	12.33+
SRB (WJ)	5.65+	7.39+	9.50+	4.46	4.73	5.26	12.31+	21.63+	31.96+	5.24	6.95+	9.61+
HOCH (WJ)	5.87+	7.67+	9.83+	4.62	4.90	5.43	12.68+	22.17+	32.65+	5.43	7.19+	9.91+
q / PER (JWJ)	5.72+	7.31+	9.28+	4.25	4.43	4.87	12.58+	21.57+	31.46+	5.05	6.67+	9.30+
JWJ / q (JWJ)	6.55+	8.48+	10.69+	4.80	5.09	5.44	14.28+	24.71+	35.19+	5.70+	7.63+	10.56+
JWJ / * (JWJ)	7.66+	9.71+	12.15+	5.83	6.15+	6.52+	16.62+	28.38+	40.00+	6.80+	9.14+	12.51+
SRB (JWJ)	5.86+	7.43+	9.43+	4.66	4.72	5.08	13.53+	23.41+	34.29+	5.36	7.13+	9.80+
HOCH (JWJ)	6.06+	7.68+	9.74+	4.82	4.89	5.25	13.87+	23.97+	35.03+	5.54	7.34+	10.08+
q / PER (HWJ)	5.75+	7.36+	9.37+	4.29	4.46	4.89	12.62+	21.63+	31.53+	5.10	6.71+	9.33+
HWJ / q (HWJ)	6.58+	8.52+	10.79+	4.83	5.11	5.46	14.31+	24.77+	35.30+	5.75+	7.67+	10.58+
HWJ / * (HWJ)	7.69+	9.76+	12.25+	5.87	6.17+	6.54+	16.66+	28.44+	40.12+	6.87+	9.17+	12.53+
SRB (HWJ)	5.90+	7.49+	9.55+	4.70	4.76	5.10	13.58+	23.47+	34.40+	5.42	7.17+	9.83+
HOCH (HWJ)	6.09+	7.73+	9.86+	4.86	4.92	5.27	13.90+	24.01+	35.12+	5.59	7.39+	10.10+
Pairwise Estimate												
SRB (WJ)	6.39+	9.03+	12.41+	4.46	4.91	5.70						
HOCH (WJ)	6.59+	9.24+	12.68+	4.62	5.08	5.86						
SRB (JWJ)	6.85+	9.31+	12.53+	4.68	4.96	5.62						
HOCH (JWJ)	7.00+	9.51+	12.79+	4.83	5.10	5.77						
SRB (HWJ)	6.89+	9.37+	12.63+	4.73	4.99	5.64						
HOCH (HWJ)	7.03+	9.57+	12.88+	4.88	5.13	5.79						
No Preliminary Test (symmetric trimming)												
	10	15	20									
q / PER (WJ)	4.24	4.26	4.43									
WJ / q (WJ)	4.65	4.65	4.80									
WJ / * (WJ)	5.60	5.59	5.84									
SRB (WJ)	4.62	4.65	4.75									
HOCH (WJ)	4.74	4.77	4.87									
q / PER (JWJ)	4.23	4.09	4.29									
JWJ / q (JWJ)	4.54	4.47	4.54									
JWJ / * (JWJ)	5.52	5.43	5.56									
SRB (JWJ)	4.55	4.46	4.57									
HOCH (JWJ)	4.71	4.58	4.69									
q / PER (HWJ)	4.26	4.11	4.30									
HWJ / q (HWJ)	4.60	4.50	4.55									
HWJ / * (HWJ)	5.59	5.47	5.57									
SRB (HWJ)	4.61	4.49	4.59									
HOCH (HWJ)	4.77	4.61	4.71									

Notes: 10/20 = 10% symmetric/20% asymmetric trimming; 15/30 = 15% symmetric/30% asymmetric trimming; 20/40 = 20% symmetric/40% asymmetric trimming; 10/10 = 10% symmetric/10% asymmetric trimming; 15/15 = 15% symmetric/15% asymmetric trimming; 20/20 = 20% symmetric/20% asymmetric trimming; q/PER indicates that q and Peritz procedures are equivalent; /* indicates that the SRB and Hayter procedures are equivalent; HOCH is the Hochberg procedure; 10 = 10% symmetric trimming; 15 = 15% symmetric trimming; 20 = 20% symmetric trimming; bold entries indicate values that exceeded Bradley's (1978) lower and upper limits; + indicates a liberal value, - indicates a conservative value, and ± indicates both conservative and liberal values in the minimum to maximum range of error rates.

Table 2. Summary Percentages of Type I Error for Multiple Comparison Procedures ($J = 3$; $g = 1$, $h = 0$ Distribution)

	Average Estimate						Individual Estimate					
	10/20	15/30	20/40	10/10	15/15	20/20	10/20	15/30	20/40	10/10	15/15	20/20
q / PER (WJ)	4.46	4.55	4.83	4.33	4.09	4.31	6.77+	10.28+	14.18+	4.25	4.47	5.34
WJ / q (WJ)	5.01	5.14	5.42	4.97	4.78	4.84	7.55+	11.35+	15.64+	4.82	5.14	6.06+
WJ / * (WJ)	6.26+	6.31+	6.55+	6.46+	6.12+	6.06+	9.11+	13.36+	18.05+	6.16+	6.46+	7.46+
SRB (WJ)	5.06-	5.04	5.22	5.30±	4.85-	4.93-	7.42+	10.99+	15.02+	5.05	5.14	5.89+
HOCH (WJ)	5.17-	5.18	5.38	5.42±	4.97-	5.03-	7.63+	11.32+	15.49+	5.15	5.29	6.06+
q / PER (JWJ)	4.42	4.22	4.29	4.74	4.18	4.20	7.24+	10.92+	14.85+	4.61	4.61	5.33
JWJ / q (JWJ)	5.01	4.78	4.78	5.23	4.87	4.76	8.03+	11.96+	16.08+	5.07	5.30	6.12+
JWJ / * (JWJ)	6.22+	5.95+	6.02+	6.70+	6.22+	5.94+	9.66+	14.12+	18.68+	6.43	6.62+	7.48+
SRB (JWJ)	4.97-	4.81-	4.90-	5.53	4.89	4.78-	7.94+	11.82+	15.92+	5.25	5.15	5.89+
HOCH (JWJ)	5.11-	4.92-	5.04-	5.67	5.01	4.91-	8.12+	12.16+	16.36+	5.39	5.28	6.06+
q / PER (HWJ)	4.43	4.24	4.35	4.83	4.23	4.21	7.26+	10.96+	14.92+	4.70	4.66	5.34
HWJ / q (HWJ)	5.03	4.82	4.86	5.33	4.90	4.78	8.06+	12.03+	16.17+	5.15	5.34+	6.14+
HWJ / * (HWJ)	6.24+	5.98+	6.11+	6.81+	6.25+	5.97+	9.69+	14.15+	18.75+	6.53+	6.66+	7.50+
SRB (HWJ)	4.99-	4.84-	4.97-	5.64	4.93	4.80-	7.97+	11.86+	16.00+	5.36	5.20	5.90+
HOCH (HWJ)	5.14-	4.95-	5.11-	5.78	5.05	4.94-	8.16+	12.20+	16.44+	5.51	5.32	6.09+
Pairwise Estimate												
SRB (WJ)	5.22	5.39+	5.82+	5.25±	4.80-	4.93						
HOCH (WJ)	5.33	5.53+	5.97+	5.35±	4.91-	5.02						
SRB (JWJ)	5.31	5.31±	5.62±	5.50	4.84	4.81-						
HOCH (JWJ)	5.43	5.41±	5.74±	5.62	4.95	4.93-						
SRB (HWJ)	5.33	5.34±	5.68±	5.61	4.89	4.83-						
HOCH (HWJ)	5.46	5.44±	5.81±	5.74	5.00	4.96-						
No Preliminary Test (symmetric trimming)												
	10	15	20									
q / PER (WJ)	4.61	4.41	4.60-									
WJ / q (WJ)	5.22	5.00	5.00									
WJ / * (WJ)	6.64+	6.31+	6.31+									
SRB (WJ)	5.57±	5.24-	5.31-									
HOCH (WJ)	5.66±	5.36-	5.40±									
q / PER (JWJ)	4.80	4.41	4.50									
JWJ / q (JWJ)	5.24	4.94	4.89									
JWJ / * (JWJ)	6.66+	6.18+	6.09+									
SRB (JWJ)	5.59	5.11	5.15-									
HOCH (JWJ)	5.71+	5.23	5.24									
q / PER (HWJ)	4.91	4.45	4.52									
HWJ / q (HWJ)	5.34	5.01	4.91									
HWJ / * (HWJ)	6.78+	6.24+	6.12+									
SRB (HWJ)	5.68+	5.17	5.17-									
HOCH (HWJ)	5.82+	5.30	5.26									

Note. See note from Table 1

Table 3. Summary Percentages of Type I Error for Multiple Comparison Procedures (J = 3; g = .25, h = .25 Distribution)

	Average Estimate						Individual Estimate					
	10/20	15/30	20/40	10/10	15/15	20/20	10/20	15/30	20/40	10/10	15/15	20/20
q / PER (WJ)	4.72	6.11+	7.64+	3.66	4.11-	4.69	6.62+	11.73+	17.69+	3.36-	4.49	6.38+
WJ / q (WJ)	5.17	6.83+	8.62+	3.94	4.48	5.09	7.28+	12.92+	19.32+	3.71	4.95	6.88+
WJ / * (WJ)	6.11+	7.92+	9.93+	4.77	5.40	6.07+	8.63+	15.25+	22.51+	4.47	5.90	8.23+
SRB (WJ)	4.92	6.29+	7.86+	3.91	4.35	4.88	7.15+	13.03+	19.75+	3.59	4.82	6.83+
HOCH (WJ)	5.07	6.45+	8.05+	4.03	4.48	5.03	7.32+	13.30+	20.12+	3.70	4.95	7.03+
q / PER (JWJ)	5.95+	7.87+	10.06+	4.47	4.81	5.47	8.20+	15.19+	22.75+	3.95	5.29	7.71+
JWJ / q (JWJ)	6.49+	8.57+	10.90+	4.84	5.22	5.95+	8.85+	16.47+	24.37+	4.23	5.72	8.18+
JWJ / * (JWJ)	7.57+	9.68+	12.16+	5.84	6.22+	6.99+	10.54+	19.24+	28.16+	5.11	6.77+	9.75+
SRB (JWJ)	6.31+	8.11+	10.40+	4.79	5.10	5.74+	9.06+	16.97+	25.65+	4.15	5.64	8.40+
HOCH (JWJ)	6.45+	8.29+	10.58+	4.92	5.25	5.89+	9.25+	17.21+	25.99+	4.28	5.80	8.61+
q / PER (HWJ)	6.02+	7.97+	10.21+	4.51	4.85	5.53	8.29+	15.37+	23.00+	3.98	5.33	7.80+
HWJ / q (HWJ)	6.57+	8.66+	11.02+	4.88	5.26	5.98+	8.96+	16.65+	24.60+	4.27	5.76+	8.26+
HWJ / * (HWJ)	7.67+	9.80+	12.28+	5.91	6.27+	7.02+	10.66+	19.46+	28.42+	5.16	6.82+	9.84+
SRB (HWJ)	6.38+	8.23+	10.57+	4.85	5.17	5.80+	9.17+	17.22+	26.01+	4.19	5.69	8.51+
HOCH (HWJ)	6.52+	8.41+	10.74+	4.97	5.32	5.95+	9.36+	17.48+	26.32+	4.33	5.85	8.70+
Pairwise Estimate												
SRB (WJ)	5.33	7.24+	9.42+	3.99	4.55	5.20						
HOCH (WJ)	5.44	7.36+	9.57+	4.10	4.66	5.31						
SRB (JWJ)	6.88+	9.71+	13.14+	4.93	5.42	6.24+						
HOCH (JWJ)	7.01+	9.86+	13.28+	5.05	5.54	6.36+						
SRB (HWJ)	6.98+	9.86+	13.45+	4.99	5.47	6.31+						
HOCH (HWJ)	7.11+	10.00+	13.58+	5.10	5.60	6.42+						
No Preliminary Test (symmetric trimming)												
	10	15	20									
q / PER (WJ)	3.42-	3.50-	3.72									
WJ / q (WJ)	3.64	3.82	3.98									
WJ / * (WJ)	4.42	4.62	4.84									
SRB (WJ)	3.61	3.72	3.94									
HOCH (WJ)	3.72	3.81	4.05									
q / PER (JWJ)	4.07	3.81	3.84									
JWJ / q (JWJ)	4.39	4.15	4.11									
JWJ / * (JWJ)	5.32	5.00	4.98									
SRB (JWJ)	4.29	4.04	4.03									
HOCH (JWJ)	4.43	4.16	4.14									
q / PER (HWJ)	4.10	3.83	3.85									
HWJ / q (HWJ)	4.44	4.17	4.11									
HWJ / * (HWJ)	5.37	5.03	5.00									
SRB (HWJ)	4.34	4.06	4.05									
HOCH (HWJ)	4.47	4.18	4.15									

Note. See note from Table 1

by a test of symmetry, particularly for the MCPs based on the modified WJ statistic (i.e., JWJ or HWJ).

Specifically, all the MCPs based on the 10/10 trimming case with the test of symmetry based on individual group estimates of tail weight and symmetry had rates of Type I error within Bradley's (1978) limits except the Hayter (1986) two-stage and Shaffer (1986) sequentially rejective Bonferroni procedure that begins with an omnibus test utilizing the WJ statistic (denoted WJ/*) and the HWJ statistic (denoted HWJ/*) with liberal rates of 8.28% and 7.52%, respectively. Interestingly, this particular condition had the largest number of MCPs that fell within Bradley's lower and upper limits. The mean error rates across robust MCPs based on the JWJ and HWJ heteroscedastic statistics for the 10/10 and 15/15 trimming cases were 5.34% and 4.76%, respectively for the test of symmetry based on average estimates across groups and 5.27% and 5.07%, respectively for the test of symmetry based on individual group estimates.

The MCPs based on the WJ statistic generally had more conservative error rates than the same MCPs based on the modified WJ statistic (i.e., JWJ and HWJ) when preceded by a test of symmetry except under the 10/20, 15/30, 20/40, and 20/20 trimming cases for the test of symmetry based on average estimates across groups where the opposite pattern was observed (i.e., WJ based MCPs had higher mean error rates). Additionally, Type I error rates for the MCPs tended to decrease with an increase in the percentage of trimming (i.e., from 10/20 to 15/30 to 20/40 and from 10/10 to 15/15 to 20/20), except for the MCPs preceded by a test of symmetry based on individual group estimates where the pattern was reversed, that is, error rates tended to increase as the proportion of trimming increased.

The mean error rates for the SRB and HOCH procedures indicate that an optimal strategy is to use a test of symmetry based either on indices of tail weight and symmetry averaged across the pairwise comparisons or averaged across all groups with 15/15 symmetric/asymmetric trimming (i.e., mean error rates closer to the nominal 5% level). A result consistent with the ($g = .5$, $h = 0$) distribution.

$g = .25$ and $h = .25$ Distribution

When nonnormal data were obtained from the ($g = .25$, $h = .25$) distribution, the use of the symmetry test based on the individual group indices resulted in all MCPs having liberal Type I error rates, for the 10/20, 15/30, 20/40, and 20/20 trimming cases (see Table 3). However, improved Type I error control was obtained when the test of symmetry was based on indices averaged across all groups or averaged across the two groups defining the pairwise comparison. Interestingly, all MCPs had rates below Bradley's (1978) upper limit for the 10/10 trimming case when preceded by the preliminary test of symmetry, regardless of the method of computing the test. In addition, all MCPs had rates of Type I error below Bradley's upper limit when always adopting 10%, 15%, or 20% symmetric trimming.

The use of the averaged over all groups tail weight and symmetry indices resulted in Type I error rates closer to the nominal level compared to always adopting symmetric trimming. For example, the 10/10 and 15/15 trimming cases had mean rates of Type I error across non-liberal MCPs equal to 4.69% and 4.91%, respectively, whereas the 10%, 15%, and 20% symmetric trimming cases had mean error rates, across MCPs equal to 4.27%, 4.13%, and 4.19%, respectively.

The MCPs based on the JWJ or HWJ heteroscedastic statistics had rates of Type I error closer to the nominal level compared to MCPs based on the WJ statistic. For example, (a) with the symmetry test based on average estimates across groups, the mean rates of Type I error across all five MCPs when based on the WJ, JWJ, and HWJ test statistics for the 10/10 trimming condition equaled 4.06%, 4.97%, and 5.02%, respectively, (b) with the symmetry test based on individual group estimates, the mean error rates for the 10/10 trimming condition equaled 3.77%, 4.34%, and 4.39%, respectively, and (c) with symmetric trimming, the mean rates for 20% trimming equaled 4.11%, 4.22%, and 4.23%, respectively.

Mean rates of Type I error for the SRB and HOCH procedures, when preceded by a test of symmetry based on tail weight and symmetry estimates from the two groups forming the pairwise comparison, were higher than when the

symmetry test was based on the average estimate of the indices across all groups for a given trimming condition, with the highest rates occurring when individual group indices of tail weight and symmetry were used. The optimal level of trimming occurs under the 10/10 symmetric/asymmetric trimming case when the MPCs were based on the JWJ or HWJ statistics (i.e., mean error rates closest to the nominal 5% level).

J = 6

Tables 4 through 6 contain the summary percentages of Type I error for the MPCs for the ($g = .5, h = 0$), ($g = 1, h = 0$), and ($g = .25, h = .25$) distributions, respectively. The SRB and HOCH procedures had identical error rates across the eleven pairings of groups sizes and variances, thus they have been combined into one row in the tables (denoted as SRB/HOCH).

$g = .5$ and $h = 0$ Distribution

All MPCs had Type I error rates below Bradley's (1978) upper limit (i.e., 7.50%) when based on the test of symmetry with indices of tail weight and symmetry averaged over groups except Hayter's (1986) two-stage and Shaffer's (1986) sequentially rejective Bonferroni procedure that begins with an omnibus test (i.e., WJ/HAY, JWJ/HAY, HWJ/HAY, WJ/SRB, JWJ/SRB, HWJ/SRB) under the 20/40 symmetric/asymmetric trimming case (see Table 4). Unlike when $J = 3$, some MPCs had error rates below Bradley's lower limit (i.e., 2.50%). Specifically, the effected MPCs were the range procedures [(PER (WJ), q (WJ), WJ/ q , PER (JWJ), q (JWJ), JWJ/ q , PER (HWJ), q (HWJ), and HWJ/ q)] when they were based on the test of symmetry using an average estimate of tail weight and symmetry across all of the groups and symmetric trimming (a result consistent with that obtained for the chi-square distribution).

The mean error rate across MPCs for the 10/20, 15/30, 10/10, 15/15, and 20/20 trimming cases when preceded by the test of symmetry based on average estimates of tail weight and symmetry across all groups was equal to 3.57%,

3.94%, 3.39%, 3.31%, and 3.39%, respectively and for the 10/10 trimming case, when preceded by the test of symmetry based on individual group estimates of tail weight and symmetry, the mean error rate was equal to 4.17%. Thus, an optimal strategy and level of trimming is to use 10/10 symmetric/asymmetric trimming with the test of symmetry based on individual group estimates (a result consistent with that obtained for the chi-square distribution).

The pattern of error rates differed with the type of heteroscedastic statistic. Error rates tended to increase as the proportion of trimming increased for the 10/20, 15/30 and 20/40 trimming cases and for the 10/10, 15/15, and 20/20 trimming cases. However, MPCs based on the JWJ and HWJ statistics, had rates that tended to decrease as the proportion of trimming increased for the 10/10, 15/15, and 20/20 conditions with the test of symmetry based on average group estimates (a result consistent with that obtained for the chi-square distribution).

The MPCs based on the WJ statistic generally had more conservative rates of error than the same MPCs based on the modified WJ statistics (i.e., JWJ and HWJ), when preceded by a test of symmetry based on individual group estimates or pairwise estimates of tail weight and symmetry, a pattern opposite to that observed for the symmetry test based on average estimates across groups (except under the 10/10 trimming case) or when always adopting symmetric trimming.

For example, under the 10/10 trimming case with the test of symmetry based on indices (tail weight and symmetry) for individual groups, the mean error rates for the MPCs based on the WJ, JWJ, and HWJ statistics were equal to 4.07%, 4.20%, and 4.25%, respectively and when based on average indices across groups, the mean error rates were equal to 3.29%, 3.43%, and 3.46%, respectively. On the other hand, when adopting 20% symmetric trimming the mean error rates across MPCs based on the WJ, JWJ, and HWJ statistics were equal to 3.63%, 3.45%, and 3.47%, respectively. This pattern is consistent with the results obtained for the chi-square distribution.

Table 4. Summary Percentages of Type I Error for Multiple Comparison Procedures (J = 6; g = .5, h = 0 Distribution)

	Average Estimate						Individual Estimate					
	10/20	15/30	20/40	10/10	15/15	20/20	10/20	15/30	20/40	10/10	15/15	20/20
PER (WJ)	3.05-	3.59-	4.43-	2.67-	2.77-	2.97-	9.35+	17.72+	27.66+	3.29-	4.83±	7.00+
q (WJ)	2.97-	3.47-	4.23-	2.58-	2.69-	2.84-	9.06+	17.20+	26.59+	3.18-	4.70±	6.64+
WJ / q (WJ)	2.61-	3.13-	4.04-	2.26-	2.41-	2.50-	8.90+	17.70+	27.83+	2.89-	4.34±	6.41+
WJ / SRB (WJ)	4.33	4.84	6.00+	3.97	4.06	4.25	14.31+	28.11+	44.63+	4.95	7.20+	11.11+
WJ / HAY (WJ)	5.13	5.69	7.18+	4.67	4.73	5.01	16.69+	32.57+	50.30+	5.79	8.41+	13.06+
SRB/HOCH (WJ)	3.84	4.19	5.07	3.57	3.53	3.79	12.09+	23.83+	38.70+	4.33	6.19	9.42+
PER (JWJ)	2.86-	3.21-	3.98-	2.82-	2.71-	2.67-	10.74+	20.23+	30.90+	3.50-	4.98±	7.36+
q (JWJ)	2.75-	3.08-	3.78-	2.75-	2.63-	2.56-	10.45+	19.63+	29.87+	3.39-	4.83±	7.02+
JWJ / q (JWJ)	2.52-	2.80-	3.45-	2.39-	2.29-	2.29-	10.32+	20.13+	31.07+	2.99-	4.45+	6.73+
JWJ / SRB (JWJ)	4.22	4.45	5.49+	4.11	3.95	3.99	16.08+	31.60+	49.35+	4.99	7.29+	11.52+
JWJ / HAY (JWJ)	5.01	5.30	6.51+	4.76	4.57	4.71	18.72+	36.12+	54.93+	5.83	8.56+	13.53+
SRB/HOCH (JWJ)	3.70	3.91	4.88	3.74	3.46	3.54	13.87+	27.27+	43.70+	4.48	6.37	9.83+
PER (HWJ)	2.91-	3.27-	4.09-	2.85-	2.73-	2.70-	10.80+	20.32+	31.03+	3.55-	5.03+	7.40+
q (HWJ)	2.78-	3.14-	3.91-	2.77-	2.66-	2.58-	10.51+	19.70+	30.01+	3.43-	4.85±	7.04+
HWJ / q (HWJ)	2.56-	2.85-	3.58-	2.41-	2.33-	2.31-	10.38+	20.21+	31.22+	3.03-	4.48+	6.78+
HWJ / SRB (HWJ)	4.28	4.55	5.68+	4.15	3.99	4.03	16.18+	31.76+	49.57+	5.04	7.35+	11.58+
HWJ / HAY (HWJ)	5.06	5.40	6.71+	4.80	4.61	4.76	18.82+	36.24+	55.13+	5.89	8.61+	13.59+
SRB/HOCH (HWJ)	3.75	4.00	5.03	3.79	3.49	3.58	13.97+	27.42+	43.96+	4.54	6.42	9.89+
Pairwise Estimate												
SRB/HOCH (WJ)	4.58	6.21+	8.98+	3.46	3.72	4.35						
SRB/HOCH (JWJ)	4.97	6.33+	9.16+	3.77	3.75	4.23						
SRB/HOCH (HWJ)	5.01	6.42+	9.29+	3.81	3.78	4.25						
No Preliminary Test (symmetric trimming)												
	10	15	20									
PER (WJ)	2.69-	2.81-	2.89-									
q (WJ)	2.60-	2.71-	2.69-									
WJ / q (WJ)	2.36-	2.38-	2.43-									
WJ / SRB (WJ)	4.14	4.23	4.53									
WJ / HAY (WJ)	4.81	4.86	5.21									
SRB/HOCH (WJ)	3.66	3.77	4.01									
PER (JWJ)	2.85-	2.77-	2.81-									
q (JWJ)	2.76-	2.67-	2.58-									
JWJ / q (JWJ)	2.35-	2.28-	2.27-									
JWJ / SRB (JWJ)	4.11	4.06	4.25									
JWJ / HAY (JWJ)	4.72	4.67	4.97									
SRB/HOCH (JWJ)	3.78	3.62	3.81									
PER (HWJ)	2.90-	2.79-	2.82-									
q (HWJ)	2.80-	2.70-	2.60-									
HWJ / q (HWJ)	2.38-	2.30-	2.28-									
HWJ / SRB (HWJ)	4.19	4.08	4.27									
HWJ / HAY (HWJ)	4.77	4.70	5.00									
SRB/HOCH (HWJ)	3.81	3.67	3.85									

Notes: 10/20 = 10% symmetric/20% asymmetric trimming; 15/30 = 15% symmetric/30% asymmetric trimming; 20/40 = 20% symmetric/40% asymmetric trimming; 10/10 = 10% symmetric/10% asymmetric trimming; 15/15 = 15% symmetric/15% asymmetric trimming; 20/20 = 20% symmetric/20% asymmetric trimming; PER is the Peritz procedure; HAY is the Hayter procedure; SRB/HOCH indicates that SRB and Hochberg procedures had equivalent rates; 10 = 10% symmetric trimming; 15 = 15% symmetric trimming; 20 = 20% symmetric trimming; bold entries indicate values that exceeded Bradley's (1978) lower and upper limits; + indicates a liberal value, - indicates a conservative value, and ± indicates both conservative and liberal values in the minimum to maximum range of error rates.

Table 5. Summary Percentages of Type I Error for Multiple Comparison Procedures (J = 6; g = 1, h=0 Distribution)

	Average Estimate						Individual Estimate					
	10/20	15/30	20/40	10/10	15/15	20/20	10/20	15/30	20/40	10/10	15/15	20/20
PER (WJ)	2.43-	2.43-	2.53-	2.39-	2.27-	2.43-	5.05	8.49+	12.78+	2.35-	2.74-	3.45-
q (WJ)	2.33-	2.32-	2.38-	2.32-	2.23-	2.32-	4.88	8.16+	12.13+	2.29-	2.67-	3.30-
WJ / q (WJ)	2.17-	2.06-	2.17-	2.18-	2.11-	2.17-	4.58	7.91+	12.02+	2.14-	2.55-	3.15-
WJ / SRB (WJ)	4.46-	3.99	4.28	4.77±	4.36-	4.45-	7.85+	12.51+	18.40+	4.52±	4.72	5.88+
WJ / HAY (WJ)	5.11+	4.71	5.07	5.44±	5.01	5.08±	9.16+	14.66+	21.50+	5.21+	5.59	6.92+
SRB/HOCH (WJ)	3.90-	3.53-	3.68-	4.14±	3.84-	3.88-	6.76+	10.57+	15.68+	3.87-	4.07-	4.96
PER (JWJ)	2.36-	1.89-	1.88-	2.93-	2.50-	2.34-	5.84+	9.52+	13.99+	2.81-	3.00-	3.50-
q (JWJ)	2.25-	1.82-	1.76-	2.85-	2.44-	2.23-	5.62+	9.19+	13.32+	2.73-	2.93-	3.30-
JWJ / q (JWJ)	2.10-	1.66-	1.56-	2.51-	2.26-	2.08-	5.31+	8.99+	13.18+	2.40-	2.67-	3.19-
JWJ / SRB (JWJ)	4.28-	3.56-	3.65-	5.08+	4.37-	4.24-	8.80+	13.92+	19.98+	4.76	4.77	5.78+
JWJ / HAY (JWJ)	4.93-	4.13-	4.24-	5.86+	5.07	4.91-	10.20+	16.19+	22.90+	5.54	5.64	6.90+
SRB/HOCH (JWJ)	3.79-	3.10-	3.22-	4.72±	3.84-	3.77-	7.82+	12.02+	17.40+	4.32-	4.11-	4.98
PER (HWJ)	2.37-	1.93-	1.94-	3.00-	2.52-	2.35-	5.89+	9.58+	14.10+	2.87-	3.03-	3.52-
q (HWJ)	2.27-	1.85-	1.82-	2.92-	2.47-	2.25-	5.68+	9.27+	13.43+	2.80-	2.96-	3.32-
HWJ / q (HWJ)	2.11-	1.68-	1.63-	2.58-	2.29-	2.09-	5.36+	9.04+	13.28+	2.47-	2.70-	3.22-
HWJ / SRB (HWJ)	4.30-	3.59-	3.74-	5.21+	4.41-	4.26-	8.87+	14.03+	20.11+	4.88	4.82	5.81+
HWJ / HAY (HWJ)	4.96-	4.17-	4.36-	6.01+	5.12	4.93-	10.27+	16.29+	23.07+	5.68+	5.70	6.94+
SRB/HOCH (HWJ)	3.80-	3.13-	3.31-	4.83±	3.88-	3.77-	7.87+	12.09+	17.51+	4.41-	4.16-	5.00
Pairwise Estimate												
SRB/HOCH (WJ)	3.99-	3.88-	4.30-	3.95±	3.65-	3.83-						
SRB/HOCH (JWJ)	4.10-	3.60-	3.99-	4.55±	3.73-	3.78-						
SRB/HOCH (HWJ)	4.12-	3.64-	4.07-	4.66±	3.77-	3.79-						
No Preliminary Test (symmetric trimming)												
	10	15	20									
PER (WJ)	2.52-	2.44-	2.49-									
q (WJ)	2.43-	2.37-	2.34-									
WJ / q (WJ)	2.28-	2.18-	2.17-									
WJ / SRB (WJ)	4.84±	4.53-	4.68-									
WJ / HAY (WJ)	5.51±	5.14±	5.42+									
SRB/HOCH (WJ)	4.23±	3.95-	4.12-									
PER (JWJ)	3.00-	2.61-	2.54-									
q (JWJ)	2.90-	2.52-	2.37-									
JWJ / q (JWJ)	2.52-	2.25-	2.10-									
JWJ / SRB (JWJ)	5.01+	4.37-	4.48-									
JWJ / HAY (JWJ)	5.78+	5.08	5.22+									
SRB/HOCH (JWJ)	4.65±	3.99-	4.03-									
PER (HWJ)	3.07-	2.65-	2.56-									
q (HWJ)	2.96-	2.56-	2.37-									
HWJ / q (HWJ)	2.60-	2.28-	2.12-									
HWJ / SRB (HWJ)	5.13+	4.45-	4.53									
HWJ / HAY (HWJ)	5.92+	5.15	5.25+									
SRB/HOCH (HWJ)	4.78±	4.05-	4.07-									

Note. See note from Table 4

Table 6. Summary Percentages of Type I Error for Multiple Comparison Procedures (J = 6; g = .25, h = .25 Distribution)

	Average Estimate						Individual Estimate					
	10/20	15/30	20/40	10/10	15/15	20/20	10/20	15/30	20/40	10/10	15/15	20/20
PER (WJ)	2.57-	3.25-	3.99-	2.17-	2.43-	2.62-	4.84	10.70+	17.47+	2.16-	3.12-	4.55
q (WJ)	2.48-	3.13-	3.79-	2.08-	2.33-	2.47-	4.66	10.33+	16.82+	2.07-	3.00-	4.30
WJ / q (WJ)	2.16-	2.85-	3.60-	1.74-	1.96-	2.17-	4.25	9.86+	16.52+	1.69-	2.52-	3.87
WJ / SRB (WJ)	3.57	4.46	5.55+	3.04	3.37	3.81	7.14+	16.38+	27.48+	3.00-	4.42	6.81+
WJ / HAY (WJ)	4.21	5.29	6.67+	3.58	3.97	4.54	8.25+	18.37+	30.36+	3.45	5.13	7.91+
SRB/HOCH (WJ)	3.04	3.83	4.71	2.66-	2.95-	3.31-	6.40+	14.60+	24.66+	2.67-	3.93	6.11
PER (JWJ)	3.74-	4.79±	6.06+	3.00-	3.16-	3.29-	7.08+	16.04+	25.79+	2.63-	3.96-	6.44+
q (JWJ)	3.63-	4.65±	5.81+	2.90-	3.05-	3.11-	6.88+	15.71+	25.15+	2.54-	3.83-	6.12+
JWJ / q (JWJ)	3.16-	4.24-	5.56+	2.44-	2.57-	2.72-	6.30+	14.98+	24.86+	2.11-	3.28-	5.52+
JWJ / SRB (JWJ)	5.10	6.36+	7.95+	4.25	4.32	4.58	10.40+	24.27+	39.41+	3.69	5.66	9.44+
JWJ / HAY (JWJ)	5.85	7.26+	9.05+	4.89	4.96	5.35	11.63+	26.25+	41.98+	4.20	6.40+	10.61+
SRB/HOCH (JWJ)	4.56	5.65+	7.14+	3.85	3.81	4.12	9.58+	22.47+	36.87+	3.36	5.19	8.81+
PER (HWJ)	3.80-	4.92+	6.24+	3.04-	3.21-	3.33-	7.25+	16.46+	26.33+	2.68-	4.02-	6.58+
q (HWJ)	3.69-	4.76±	5.99+	2.93-	3.10-	3.17-	7.04+	16.08+	25.70+	2.59-	3.90-	6.25+
HWJ / q (HWJ)	3.24-	4.33-	5.74+	2.49-	2.63-	2.79-	6.42+	15.35+	25.38+	2.13-	3.34-	5.64+
HWJ / SRB (HWJ)	5.23	6.51+	8.17+	4.34	4.39	4.66	10.65+	24.80+	40.10+	3.73	5.74	9.62+
HWJ / HAY (HWJ)	5.70+	7.39+	9.26+	4.98	5.03	5.46	11.86+	26.74+	42.66+	4.24	6.50+	10.81+
SRB/HOCH (HWJ)	4.66	5.80+	7.35+	3.92	3.89	4.19	9.82+	23.13+	37.70+	3.42	5.31	9.06+
Pairwise Estimate												
SRB/HOCH (WJ)	3.86	5.50+	7.42+	2.84	3.30	3.84						
SRB/HOCH (JWJ)	6.14+	9.45+	13.92+	4.14	4.56	5.36						
SRB/HOCH (HWJ)	6.31+	9.82+	14.51+	4.22	4.69	5.51+						
No Preliminary Test (symmetric trimming)												
	10	15	20									
PER (WJ)	2.06-	2.18-	2.24-									
q (WJ)	1.97-	2.10-	2.10-									
WJ / q (WJ)	1.65-	1.74-	1.81-									
WJ / SRB (WJ)	2.90	3.06	3.35									
WJ / HAY (WJ)	3.41	3.62	3.99									
SRB/HOCH (WJ)	2.53-	2.70-	2.95-									
PER (JWJ)	2.67-	2.53-	2.39-									
q (JWJ)	2.58-	2.44-	2.21-									
JWJ / q (JWJ)	2.19-	2.03-	1.91-									
JWJ / SRB (JWJ)	3.87	3.60	3.52									
JWJ / HAY (JWJ)	4.44	4.17	4.15									
SRB/HOCH (JWJ)	3.46	3.17	3.15									
PER (HWJ)	2.70-	2.56-	2.40-									
q (HWJ)	2.61-	2.46-	2.23-									
HWJ / q (HWJ)	2.22-	2.05-	1.92-									
HWJ / SRB (HWJ)	3.92	3.63	3.54									
HWJ / HAY (HWJ)	4.50	4.21	4.18									
SRB/HOCH (HWJ)	3.50	3.21	3.17									

Note. See note from Table 4

The SRB and HOCH methods had mean error rates closest to the nominal level when preceded by the test of symmetry based on average group estimates for the 20/40 trimming case or pairwise estimates for the 10/20 trimming case. Specifically, the mean error rates for the procedures based on the WJ, JWJ, and HWJ statistics were 5.07%, 4.88%, and 5.03%, respectively when using the average group estimates of tail weight and symmetry and 4.58%, 4.97%, and 5.01%, respectively for the pairwise estimate indices. It is worth noting that under the 20/40 trimming case, the SRB/HOCH procedures were the only MCP to have robust error rates when preceded by a test of symmetry.

$g = 1$ and $h = 0$ Distribution

All MCPs had rates of Type I error below Bradley's (1978) upper limit when preceded by a test of symmetry based on indices averaged across all groups for the 15/30, 20/40, and 15/15 trimming conditions and when the test of symmetry was based on individual group indices for the 15/15 trimming condition (see Table 5). Few trimming conditions resulted in MCPs with error rates within Bradley's limits. The condition with the most robust MCPs occurred with a test of symmetry based on tail weight and symmetry estimates from the individual groups with 15/15 symmetric/asymmetric trimming. For this particular trimming condition, the mean error rates were closer to the nominal 5% level for MCPs preceded with the symmetry test based on the individual group estimates (average rate equal to 3.82%) compared to MCPs preceded with the test of symmetry based on average estimates across all groups (average rate equal to 3.39%). Furthermore, MCPs based on the JWJ and HWJ statistics generally had error rates closer to the nominal level compared to MCPs based on the WJ statistic. For example, under the 15/15 trimming case with the test of symmetry based on tail weight and symmetry estimates from individual groups, the mean error rates across the MCPs based on WJ, JWJ, and HWJ statistics were equal to 3.72%, 3.85%, and 3.90%, respectively.

Noteworthy is that the form of the heteroscedastic statistic had an influence on Type I error rates regardless of whether a test of

symmetry was used. For example, under the 15% symmetric trimming condition, the liberal error rate for the Hayter (1986) procedure based on the WJ statistic became nonliberal when based on the JWJ or HWJ statistic. This follows the general pattern that error rates tended to be smaller (more conservative) for MCPs based on the JWJ or HWJ statistics compared to when the MCPs were based on the WJ statistic. However, under the 10/10 and 15/15 symmetric/asymmetric trimming cases when preceded by the test of symmetry, the opposite pattern was observed, that is, the MCPs based on the WJ statistic were more conservative than the same MCPs based on the modified WJ statistics (i.e., JWJ and HWJ), a result consistent with the ($g = .5$, $h = 0$) distribution under the 10/10 trimming case. In addition, Type I error rates for the MCPs tended to decrease with an increase in the proportion of trimming cases (i.e., from 10/20 to 15/30 to 20/40 and from 10/10 to 15/15 to 20/20), except for the MCPs preceded by a test of symmetry based on individual group estimates where the pattern was reversed, that is, error rates tended to increase as the proportion of trimming increased (i.e., a pattern consistent with the results for $J = 3$).

Type I error rates for the SRB and HOCH procedures indicated that a test of symmetry based on the individual group indices provided mean error rates closer to the nominal 5% level compared to always adopting symmetric trimming or trimming symmetrically/asymmetrically based on the pairwise or across all groups average indices. For example, the mean error rates for SRB/HOCH, based on the WJ, JWJ, and HWJ statistics, were 4.96%, 4.98%, and 5.00%, respectively, under the 20/20 trimming case when using individual group indices of tail weight and symmetry, and were 4.12%, 4.03%, and 4.07%, respectively, for the 20% symmetric trimming case.

$g = .25$ and $h = .25$ Distribution

All MCPs had rates of Type I error below Bradley's (1978) upper limit for the 10/10, 15/15, and 20/20 trimming cases when preceded by the test of symmetry with average estimates across groups and the 10/10 trimming case when preceded by the test of symmetry

with individual group estimates (see Table 6). Under the 10%, 15%, and 20% symmetric trimming cases, all MCPs had non-liberal error rates. The MCPs based on the range statistic tended to have conservative error rates, whereas under these trimming cases, the MCPs with rates within Bradley's limits were the WJ/SRB, WJ/HAY, JWJ/SRB, JWJ/HAY, SRB/HOCH (JWJ), HWJ/SRB, HWJ/HAY, and SRB/HOCH (HWJ).

The mean error rates, however, were more conservative under the symmetric trimming cases compared to the rates obtained for the MCPs when a symmetric/asymmetric strategy based on indices of tail weight and symmetry was adopted. Specifically, the mean error rates across non-liberal MCPs for the 10/10, 15/15, and 20/20 trimming cases when preceded by the test of symmetry with average group estimates were equal to 3.24%, 3.40%, and 3.65%, respectively and the mean rate for the 10/10 and 15/15 trimming case when preceded by the test of symmetry with individual group estimates were equal to 2.91% and 4.15%, respectively. Whereas, under the 10%, 15%, and 20% symmetric trimming cases, the mean error rates across MCPs were equal to 2.95%, 2.86%, and 2.85%, respectively.

MCPs based on the WJ statistic tended to have more conservative rates than when based on the JWJ or HWJ statistic. For example, under the 20/20 trimming case with the test of symmetry based on average group estimates, the mean error rates for the MCPs based on the WJ, JWJ, and HWJ statistics were 3.15%, 3.86%, and 3.93%, respectively and under the 15/15 trimming case with the test of symmetry based on individual group estimates, the mean error rates for non-liberal MCPs based on the WJ, JWJ, and HWJ statistics were 3.69%, 4.38%, and 4.46%, respectively. The general pattern was for error rates to increase as the proportion of trimming increases when the MCPs were preceded by a test of symmetry. However, this pattern only occurred for the MCPs based on a WJ statistic when always adopting symmetric trimming.

The SRB and HOCH procedures had higher mean error rates when based on symmetric/asymmetric trimming obtained from pairwise estimates than when based on indices

obtained from all the groups. For example, liberal rates under the 10/20 trimming case based on pairwise estimates became robust when symmetric/asymmetric trimming was based on indices of tail weight and symmetry averaged over all groups. The data suggests that an optimal strategy was 10/20 symmetric/asymmetric trimming based on Q_1 and Q_2 obtained from all groups in the design. Specifically, the mean error rates for the SRB/HOCH procedures, based on the JWJ and HWJ statistics, were 4.56% and 4.66%, respectively.

Conclusion

In the present study, the strategy of computing a test of symmetry in order to determine whether to trim nonnormal data symmetrically (from both tails of the empirical distributions) or asymmetrically (from one tail of the empirical distributions) was compared to always utilizing an *a priori* symmetric trimming strategy, an approach previously investigated by Keselman, Lix et al. (1998) and typically recommended in the empirical literature (e.g., see Wilcox, 2003). We investigated the utility of testing for symmetry within the context of pairwise multiple comparison testing in a one-way independent groups design.

Three variations of a test of symmetry were investigated, each utilizing indices of tail weight and symmetry. The first variation obtains the indices of tail weight and symmetry by computing them within each group of a one-way completely randomized layout and then averages these values across the groups to obtain a summary measure of tail weight and symmetry. A second variation also takes an average of group indices, but only from the two groups comprising a particular pairwise comparison. The third variation, does no averaging across groups but measures tail weight and symmetry within each group of the pairwise comparison, using this information to determine whether data should be trimmed symmetrically or asymmetrically within each particular group.

The rationale behind all three approaches is to obtain an estimate of the typical score, that is, an estimate that represents the bulk of the observations, and accordingly outlying

values are not wanted, found in the tail(s) of the nonnormal distributions, to adversely affect the score to be selected as typical – selecting a score that is not central to the distribution (e.g., the usual mean can be very far away from the central portion of a distribution of scores for skewed data). Though the rationale is the same for these three approaches, they respond to the need to obtain a good representation of the typical score in different ways.

The first method uses all of the data, across groups, to measure symmetry in the data and applies the results across all groups, that is, trims in a consistent fashion across all groups. The second and third approaches measure symmetry, or the lack thereof, by only looking at the data involved in the pairwise comparison. The logic here is to ignore the type of nonsymmetry that may exist in groups that are not involved in a particular comparison. This rationale is similar to the approach of using a nonpooled error term, rather than a pooled error term, in order to avoid the biasing effects of variance heterogeneity in tests of mean equality. The third approach takes this rationale to its logical completion by finding the typical score in each group of the pairwise comparison by assessing symmetry/asymmetry within each individual group, rather than averaging over the two groups and applying the same form of trimming to both groups. That is, with this approach we are comparing the typical score from one group with the typical score from a second group, even though these typical scores were developed through different methods of trimming.

In addition to the use of a test of symmetry, the type of heteroscedastic statistic used in the computation of the MCPs was also investigated. The WJ statistic was investigated by Keselman, Lix et al. (1998) and the Johnson (1978) and Hall (1992) transformed WJ statistics investigated by Keselman et al. (2002). The MCPs with transformed WJ statistics [i.e., Hall (1992) or Johnson (1978)] based on a test of symmetry provided better Type I error control when distributions were nonnormal in form and had heterogeneous variances compared to the use of the WJ statistic with 20% symmetric trimming, the approach investigated by

Keselman, Lix et al. (1998) and generally recommended in the literature.

Specifically, MCPs showed improved Type I error control, that is, nonrobust MCPs became robust and mean Type I error rates were closer to the nominal 5% level when data were first checked for symmetry and the MCPs were computed based on modified WJ statistics (i.e., JWJ or HWJ). A test of symmetry based on each individual group's indices of tail weight and symmetry generally provided mean Type I error rates closer to the nominal level for the MCPs than when the symmetry test was based on indices averaged over all groups in the design or just the groups in a particular pairwise comparison, particularly for the more extreme non-normal distributions. Across all nonnormal distributions investigated, optimal percentages of trimming in terms of controlling Type I error rates within Bradley's (1978) limits were the 10/10 and 15/15 symmetric/asymmetric trimming conditions. Interestingly, these proportions are less than the recommended 20% symmetric trimming.

The magnitude of Type I error rates changed as the pattern and percentage of trimming changed. Across the nonnormal distributions investigated, Type I error rates generally increased for the MCPs as the proportion of trimming increased over the 10/20, 15/30, and 20/40 trimming cases and for the 10/10, 15/15, and 20/20 trimming cases when preceded by a test of symmetry. However, under the following conditions the opposite pattern occurred when the MCPs were preceded by a symmetry test where the indices of tail weight and symmetry were obtained by averaging across the indices within each group of the design (a) for the chi-square distribution, Type I error rates decreased as the proportion of trimming increased (10/10, 15/15, and 20/20) for MCPs based on the JWJ and HWJ statistics, (b) for the ($g = .5$, $h = 0$) distribution, Type I error rates decreased as the proportion of trimming increased (10/10, 15/15, and 20/20) for MCPs based on the JWJ and HWJ statistics only for $J = 6$, and (c) for the ($g = 1$, $h = 0$) distribution, Type I error rates generally decreased as the proportion of trimming increased (from 10/20 to 15/30 to 20/40 and from 10/10 to 15/15 to 20/20).

The Type I error rates for the MCPs based on the JWJ or HWJ statistics were generally more conservative than the same MCPs based on the WJ statistic for the chi-square distribution. However, as the degree of nonnormality increased, this pattern reversed itself, firstly for the $J = 3$ condition and smaller percent trimming condition (10/10) for $J = 6$ for the ($g = .5, h = 0$) distribution, the smaller percent trimming conditions (10/10 and 15/15) for the ($g = 1, h = 0$) distribution, and across all trimming cases for the most extreme non-normal distribution ($g = .25, h = .25$) investigated. As the population distribution became more non-normal (e.g., skewed), the advantage of the transformed WJ statistics in terms of providing more robust MCPs was evident. This is not surprising given that the JWJ and HWJ statistics were developed to deal with the skewness bias. The Type I error rates for MCPs based on the JWJ statistic were slightly smaller (i.e., more conservative) than the rates for the same MCPs based on the HWJ statistic across the non-normal distributions investigated.

Taking into consideration the trimming cases that resulted in non-liberal error rates across most MCPs preceded by a test of symmetry with the pattern of error rates across trimming percentages and the generally superior performance of the MCPs with either the JWJ or HWJ statistics, the following general recommendations are provided for a strategy to achieve good Type I error control in a one-way independent groups design: (a) for distributions with skewness less than 2, adopt the 10% symmetric or 10% asymmetric trimming condition based on a test of symmetry where the indices of tail weight and symmetry are obtained by averaging over all groups when $J = 3$, whereas for $J = 6$, use a test of symmetry based on individual group indices of tail weight and symmetry and (b) for distributions with skewness greater than 2, adopt the 15% symmetric or 15% asymmetric trimming condition based on a test of symmetry using individual group indices of tail weight and symmetry.

As an overall recommendation, researchers may adopt any one of the MCPs with either the JWJ or HWJ statistic with trimmed means and Winsorized variances

preceded by a test of symmetry in order to deal with nonnormal data and heterogeneous variances, conditions likely to be encountered in applied research. The importance of this finding is that educational researchers will be assured that the method will provide good Type I error control with generally more modest amounts of trimming compared to the generally recommended strategy of uniformly adopting 20% symmetric trimming.

References

- Babu, J. G., Padmanabhan, A. R., & Puri, M. P. (1999). Robust one-way ANOVA under possibly non-regular conditions. *Biometrical Journal*, *41*, 321-339.
- Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*, 144-152.
- Cressie, N. A. C., & Whitford, H. J. (1986). How to use the two-sample *t*-test. *Biometrical Journal*, *28*, 131-148.
- Guo, J. H., & Luh, W. M. (2000). An invertible transformation two-sample trimmed *t*-statistic under heterogeneity and nonnormality. *Statistics & Probability Letters*, *49*, 1-7.
- Hall, P. (1992). On the removal of skewness by transformation. *Journal of the Royal Statistical Society, Series B*, *54*, 221-228.
- Hastings, N. A. J., & Peacock, J. B. (1975). *Statistical distributions: A handbook for students and practitioners*. New York: Wiley.
- Hayter, A. J. (1986). The maximum familywise error rate of Fisher's least significant difference test. *Journal of the American Statistical Association*, *81*, 1000-1004.
- Hoaglin, D. C. (1985). Summarizing shape numerically: The *g*- and *h*-distributions. In D. Hoaglin, F. Mosteller, & J. Tukey (Eds.), *Exploring data tables, trends, and shapes*, (pp. 461-513). New York: Wiley.
- Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, *75*, 800-802.
- Hogg, R. V., Fisher, D. M., & Randles, R. H. (1975). A two-sample adaptive distribution free test. *Journal of the American Statistical Association*, *70*, 656-661.

Johansen, S. (1980). The Welch-James approximation to the distribution of the residual sum of squares in a weighted linear regression. *Biometrika*, *67*, 85-92.

Johnson, N. J. (1978). Modified t tests and confidence intervals for asymmetrical populations. *Journal of the American Statistical Association*, *73*, 536-544.

Keselman, H. J., Kowalchuk, R. K., & Lix, L. M. (1998). Robust nonorthogonal analyses revisited: An update based on trimmed means. *Psychometrika*, *63*, 145-163.

Keselman, H. J., Lix, L. M., & Kowalchuk, R. K. (1998). Multiple comparison procedures for trimmed means. *Psychological Methods*, *3*, 123-141.

Keselman, H. J., Wilcox, R. R., & Lix, L. M. (2003). A generally robust approach to hypothesis testing in independent and correlated groups designs. *Psychophysiology*, *40*, 586-596.

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.

Keuls, M. (1952). The use of the "Studentized range" in conjunction with an analysis of variance. *Euphytica*, *1*, 112-122.

Lix, L. M., & Keselman, H. J. (1995). Approximate degrees of freedom tests: A unified perspective on testing for mean equality. *Psychological Bulletin*, *117*, 547-560.

Lix, L. M., & Keselman, H. J. (1998). To trim or not to trim: Tests of location equality under heteroscedasticity and non-normality. *Educational and Psychological Measurement*, *58*, 409-429.

Luh, W. M., & Guo, J. H. (1999). A powerful transformation trimmed mean method for one-way fixed effects ANOVA model under non-normality and inequality of variances. *British Journal of Mathematical and Statistical Psychology*, *52*, 303-320.

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

Newman, D. (1939). The distribution of the range in samples from a normal population expressed in terms of an independent estimate of standard deviation. *Biometrika*, *31*, 20-30.

Othman, A. R., Keselman, H. J., Wilcox, R. R., Fradette, K., & Padmanabhan, A. R. (2002). A test of symmetry. *Journal of Modern Applied Statistical Methods*, *1*, 310-315.

Peritz, E. (1970). *A note on multiple comparisons*. Unpublished manuscript, Hebrew University, Jerusalem, Israel.

Ryan, T. A. (1960). Significance tests for multiple comparison proportions, variances and other statistics. *Psychological Bulletin*, *57*, 318-328.

SAS Institute Inc. (1989). *SAS/IML software: Usage and reference, version 6* (1st ed.). Cary, NC: Author.

Shaffer, J. P. (1979). Comparison of means: An F test followed by a modified multiple range procedure. *Journal of Educational Statistics*, *4*, 14-23.

Shaffer, J. P. (1986). Modified sequentially rejective multiple test procedures. *Journal of the American Statistical Association*, *81*, 826-831.

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, *29*, 350-362.

Welsch, R. E. (1977). Stepwise multiple comparison procedures. *Journal of the American Statistical Association*, *72*, 566-575.

Westfall, P. H., & Young, S. S. (1993). *Resampling-based multiple testing*. New York: Wiley.

Wilcox, R. R. (1990). Comparing the means of two independent groups. *Biometrical Journal*, *32*, 771-780.

Wilcox, R. R. (1994). A one-way random effects model for trimmed means. *Psychometrika*, *59*, 289-306.

Wilcox, R. R. (1995). ANOVA: A paradigm for low power and misleading measures of effect size? *Review of Educational Research*, *65*, 51-77.

Wilcox, R. R. (2003). *Applying contemporary statistical techniques*. New York: Academic Press.

Wilcox, R. R., Keselman, H. J., & Kowalchuk, R. K. (1998). Can tests for treatment group equality be improved?: The bootstrap and trimmed means conjecture. *British Journal of Mathematical and Statistical Psychology*, *51*, 123-134.

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