THE EFFECTS OF DIFFERENTIAL BETWEEN-GROUP SKEWNESS ON HOMOSCEDASTIC, HETEROSCEDASTIC TRIMMED MEAN, AND RANK-BASED BETWEEN-GROUPS PROCEDRUES

LAURA J. MILLS

A DISSERTATION SUBMITTED TO

THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN PSYCHOLOGY

YORK UNIVERSITY

TORONTO, ONTARIO

April 2014

Abstract

The effect of differential between-group skewness was investigated for the traditional *t*- and ANOVA *F* tests, the Welch procedure without trimming (Welch, 1938) and with trimming and Winsorized variances (Yuen, 1974), the Welch-James (James, 1951) with trimming and transformation, the Yuen procedure with bootstrapping, trimming, and transformation (Keselman, Wilcox, Othman, & Fradette, 2002), and the Welch procedure with ranked data (Zimmerman & Zumbo, 1992). Empirical Type I error and power rates for these procedures were compared under varied conditions of non-normality, heterogeneity, group size imbalance, and positive and negative pairing of variance and group size. In particular, these conditions were combined with conditions of between-group skewness that was equal, dissimilar, and dissimilar and directionally opposite. Monte Carlo simulations revealed that when skewness across groups was unequal, there were deleterious effects on Type I error and power for models with two, four, and seven groups for the traditional *t*-test and ANOVA *F*, which had unacceptable rates of Type I error and power compared to other procedures. Further, procedures that accommodate heteroscedasticity fall short compared to those that can simultaneously accommodate heterogeneity and skewness. Finally, empirical power is highest for the Welch procedure on ranked data in most data conditions. It is recommended that investigators routinely investigate their data for violations and adopt robust procedures such as the Welch test on ranks to test differences of central tendency.

Acknowledgements

I would like to acknowledge Dr. Rob A Cribbie for his mentorship and guidance over the years that have culminated in this dissertation. Mostly, I must sincerely thank Dr. Cribbie for his patience, for without it I may have lost my resolve to complete this journey while I juggled the many roles I've taken on while completing my post-graduate studies. Secondly, I must thank my son, Logan, who, without knowing or trying, is my reminder that I will be a better mom if I model the commitment and dedication needed to complete post-secondary education. Moreover, I hope that he learns that someday he can use his own education to benefit the lives of others.

Table of Contents

List of Tables

The Effects of Differential Between-Group Skewness on Homoscedastic, Heteroscedastic Trimmed Mean, and Rank-Based Between-Groups Procedures

Researchers are often interested in comparing the central tendencies of a dependent variable's distribution across independent groups. For example, Leentjens, Wielaert, van Harksamp, and Wilmink (1998) examined differences in emotional expression between patients diagnosed with schizophrenia and those without a diagnosis. The authors found differences in three of four non-verbal expressions of emotion (prosody) tests, indicating that controls without a mental health diagnosis scored better than those diagnosed with schizophrenia. Findings such as these which shed light on group differences can be applied to the design of individual treatment plans, program planning, and policy decision-making. It is thus vital that scientists present accurate findings based on best practices for data analysis.

For researchers to have confidence in the results derived from a test statistic, the procedures that should be adopted should be the most powerful (Cohen, 1962, 1992; Sedlemeier & Gigerenzer, 1989) while maintaining Type I error (i.e., α) at the established level. In order to identify an appropriate test statistic for comparing the means of independent groups, the outcome variable must be investigated for evidence of nonnormality and unequal variances across populations because when one or both of these assumptions are violated, Type I error and power rates can be substantially affected. Often, however, the analytic protocol for psychological research does not include checking for and accommodating skewness or heterogeneity, as evidenced by countless studies having been published without recognition of these assumption violations and the use of appropriate analytic accommodations, even though the data are clearly skewed or heterogeneous. This was emphasized by Micceri (1989) who found that data which are not normally distributed is more the norm than the exception in psychological research.

Indeed, many quantitative methodologists regard normality as something of a mysterious oasis, using such terms as ‘unexpected’ (Garrett, 1926), ‘magic’ (Mosteller & Tukey, 1977), ‘unicorn’ (Micceri) and ‘the Holy Grail’ (Sartori, 2006). In terms of population variances, Golinski and Cribbie (2009) and Wilcox (1987) found that heterogeneous variances were common in psychological research. Clearly, normality and heterogeneity should not be cavalierly assumed; however, it seems they are, since so much published work does not include reports on issues that can affect the robustness of analyses, and often the results are based on procedures that may not accommodate problematic data.

In looking for ways to deal with skewed or heterogeneous data, quantitative researchers will simulate data conditions of skewness and heterogeneity to determine which procedures provide the best balance of Type I error control and statistical power. Until recently, much of the research that reflects the impact of skewness and heterogeneity on Type I error and power rates has focused on scenarios where both groups have equal skewness. In psychological data, however, within-group distributions are often not equally skewed. For example, in the Leentjens et al. (1998) example, the control group had a normal distribution on prosody scores, while the patient group had a skewed distribution. This type of scenario may present itself with many types of psychological measures. For example, Greer, Hunter, Dunlap, and Berman (2006) point out that intelligence tests provide a good scenario for differential between group skewness, as they have highly varied scores combined with a tendency toward skewness.

In the current study I investigate the effects of non-normality and heterogeneity on between-groups procedures when group variances are both equal and unequal, with a particular focus on differential between-groups skewness. In Study One, we set data from two groups to varied conditions of size, distribution shape, variability, and skewness and analyzed them with popular or traditional parametric procedures along with novel parametric and non-parametric procedures, some of which have previously been shown to be robust under certain data conditions. In Study Two, the procedures that were robust in Study One were tested with data from four and seven groups, again set to vary across conditions of normality and heterogeneous variability and skewness. We then recommend the use of the procedures found to maintain the best balance of Type I error and power for use in studies involving tests of differences between independent populations.

Prevalence of Variance Heterogeneity

Unequal variability often exists naturally between pre-existing groups and also exists as a result of experimental manipulation (Erceg-Hurn & Mirosevich, 2008). An example of naturally occurring heterogeneity in the Leentjens et al. (1998) study was that the variance of prosodic repetition for the schizophrenic group was more than twice that of the controls. The high prevalence of unequal variances in psychological data has been well noted, with variance ratios (VR; Erceg-Hurn & Mirosevich, 2008) as high as 282:1 (Grissom, 2000) and commonly 8:1 (Keselman et al., 1998). Golinski and Cribbie (2009) explored 489 articles from top-tier psychological journals published in 2000 and found that, of those that used a between-groups design, most (74%) had unequal group sizes, 42% had a VR of more than 2:1, but almost none (2%) reported having performed tests of variance homogeneity. Further, Wilcox (1987) found that 3 of 14 studies published in the *American Educational Research Journal* had VRs larger than 16:1.

Empirical Effects of Variance Heterogeneity

Alexander and Govern (1994) found that the empirical Type I error rate was not problematically affected by heterogeneity when sample sizes were equal but positive pairing of group sizes and variability – when the larger group size has the larger variance - resulted in Type I error rates as low as 1% (similar findings were reported by Coombs, Algina, & Oltman, 1996), with a parallel effect of reduced power. In the presence of negative pairing of group sizes and variability – when the larger group has the smaller variance – the authors found empirical Type I error rates of 18% when group sizes were 4 and 15 and VR = 2:1 and as high as 35% (VR = 6:1). Note that, in this example, and all cases below, results are based on nominal $\alpha = .05$.

These types of results pervade simulation studies that examine the accuracy of the traditional *t*-test (Yin & Othman, 2009; Zimmerman, 1987; Zimmerman & Zumbo, 1993). For example, Zimmerman (1987) showed that when VR = 1:25, empirical Type I error rates for the traditional *t* test were very low (down to .001 when n_1=16 and n_2=4) and very high (up to 35%) when the pairing of group size and variance was positive. In terms of power, Alexander and Govern (1994) showed that when variances were unequal and group sizes were equal, the traditional *t*-test had lower power than other tests investigated.

Prevalence of Skewed Distributions

Skewed distributions are very common in psychological research. Geary (1947) laboriously showed the extreme likelihood that distributions would be anything but symmetric and goes so far as to quip that normal distributions are a myth. Micceri (1989) explored the distributions of over 400 variables from large samples in education and psychology and found that 66% to 84% of the score distributions were at least moderately skewed. Micceri aptly notes that asymmetry is “more the rule than the exception” (p.161). Golinski and Cribbie (2009) found that, of 140 studies that used an independent groups design, only 8% reported having tested for normality and of those, 91% indicated departures from normality.

Empirical Effects of Skewed Distributions

Sawilowski and Blair (1992) found that Type I error control for the traditional *t*-test was not affected by moderate skewness when sample sizes were equal and large (i.e., over 120), but the *t*-test did not perform well under heavy skewness or when sample sizes were unequal or small (i.e., less than 60). Zimmerman & Zumbo (1992) also found that empirical Type I error was deflated for the traditional *t*-test when distributions were skewed and sample sizes were equal.

Skewed Distributions and Heterogeneous Variances Combined

The prevalence of data that has combined skewness and heterogeneity is not known (since it is rare for researchers to provide information on either type), but it can be surmised that the high prevalence of each data violation would naturally result in a high prevalence of their combined presence.

Empirical Effects of Simultaneous Nonnormality and Heterogeneity

When variance heterogeneity and skewness are combined, the deleterious effects on the performance of the traditional *t*-test tests are concerning. Zimmerman and Zumbo (1993) showed empirically that Type I error rates rose to unacceptably high levels (up to 15%) when variances (VR = 4:1) and group sizes (n_1=6; n_2=18) were negatively paired and unacceptably low levels (below 2%) when variances and group sizes were positively paired. In the same study, power was deleteriously affected for the traditional *t*-test in the presence of unequal group sizes (18:6) and VR = 4:1. Further, Algina, Oshima, and Lin (1994) found that the empirical Type I error rate for the traditional *t*-test was 23% when distributions were skewed, VR = 3:1, and group sizes (*N* = 100) were unequal (with a group size ratio of 1:3). When sample size was increased, empirical Type I error rates were still consistently over 20%, even when *N* = 700 and group size ratio was

1:3. Even when group sizes were equal, empirical Type I error rates were well above nominal (8%) when VR = 3:1, *N* = 100, and distributions were skewed; though Type I error rates tended more toward nominal as group sizes increased. For example, when *N* = 200 with equal group sizes, empirical Type I error when VR = 3:1 and distributions were log-normal was a more acceptable 7%, and continued to approach 5% as groups sizes increased. Yin and Othman (2006) also found that heterogeneity combined with non-normal distributions produced Type I error rates that strayed from the nominal level when the group sizes were unequal and that power was deleteriously affected.

Differential Between-Group Skewness

The primary objective of the current studies was to investigate the Type I error and power rates of independent-groups procedures in the presence of skewed distributions and heterogeneous variances and, in particular, when the level or direction of skewness differs across groups. The current studies were also proposed to advance the field by investigating empirical Type I error rates under conditions of differential skewness in combination with variance heterogeneity.

Although the prevalence of differential skewness in the published research is unknown, it has been discussed in the literature by Tiku (1964), who found that when variances were equal, and the long tails of two skewed distributions had tails pointed away from each other, empirical α was inflated to up to 10% when nominal α was 5%. When the long tails of skewed distributions were pointed toward each other, empirical α was deflated to 3%. Differential group skewness was more of a problem than skewness that is similar across groups, as shown by Wilcox (1990), who investigated the effects of differential between-group skewness and found that procedures typically understood as robust did not handle these conditions well. Although many of the procedures used in the Wilcox paper are not relevant to this project, one particular test, the Welch procedure (detailed below) had acceptable empirical Type I error rates in the presence of unbalanced sample size, heterogeneity, and unequal between-group skewness. In the same conditions, power rates varied depending on group size and degree of imbalance, degree of skewness, and degree of heterogeneity, with the lowest power associated with the furthest strays from normality and variance equality. Wilcox (1992) also showed that specific two-group conditions of unequal skewness combined with heterogeneity and unequal group size resulted in high empirical Type I error rate when testing differences in medians. Moreover, Wilcox showed that power is reduced across a host of procedures when differential skewness is combined with heterogeneity and unequal sample sizes. Further, procedures that were designed to deal with heterogeneity had lower power than other procedures even when group sizes were balanced and distributions were differentially skewed and heterogeneous. More recently, Greer et al. (2006) investigated opposite-direction skew and found that it reduced internal consistency alpha and inter-item correlation scores. Finally, Cribbie, Fiksenbaum, Keselman, and Wilcox (2012) investigated the effects of differential skewness in combination with heterogeneity and unbalanced group sizes when there were three and eight groups and found that Type I error control was compromised when distributions were differentially skewed and combined with unequal variances and group sizes, except in the case of one procedure that used trimming and bootstrapping for both three- and eight-group models. Power was also deleteriously affected in these data conditions across all procedures but was best maintained by the parametric bootstrap procedure. This study is of particular interest to the current project as it explores the effects of differential skewness. There has been little other work on the effects of differential skewness in psychological research and in particular on differential and directionally opposed skewness.

In the current studies I investigated the effects of differential between-group skewness on the Type I error rates and power of the traditional *t*- and ANOVA *F* tests, as well as the effects of differential skewness in combination with variance heterogeneity. Further, I examined the effects of these conditions on several other procedures (outlined below), including a procedure of particular interest, the Welch test (Welch, 1938) with ranked data. The objective of the project was to find a procedure that has a good balance of Type I and II error under myriad configurations of unbalanced group sizes, similar and dissimilar nonnormality, homo- and heterogeneity, and the combination of these three factors.

Vulnerability of Popular Parametric Procedures to Assumption Violation

The null hypothesis for the traditional *t*-test is that the population means are equal (H_0: $\mu_1 = \mu_2$) and H_0 is rejected if $|t| \geq t_{\alpha/2,df}$, (assuming a non-directional test), where $df = n_1 + n_2 - 2$. The test statistic is

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2},$$

with $\bar{x}_j$ denoting the mean of the j^{th} group, n_j the size of the j^{th} group, and s_j the standard deviation of the j^{th} group.

When the population variances are unequal, the pooled variance estimate may not accurately represent the variability of *either* group. When unequal variances exist in combination with unequal group sizes, the result is either a positive or negative pairing of group sizes and variability. When pairing of group sizes and variability exists, the larger group will have the greatest ‘weight’ on the *t*-test because of the pooled standard deviation (s_p^2). With positive pairing, the larger group will inflate the s_p^2, overestimate the denominator of *t* and thus underestimate the *t* statistic, thereby increasing the chances of a Type II error (thus lowering power). With negative pairing, the larger group will deflate the s_p^2 and thus overestimate the *t*-statistic, increasing the chances of a Type I error.

The one-way analysis of variance (ANOVA) *F* statistic is an extension of the traditional *t*-test to multiple groups. This is understood intuitively: variability between mean scores for two groups is compared to variability of the scores within each group, as shown below:

$$F_{(J-1,N-J)} = \frac{\frac{\sum_{j=1}^{J} n(\bar{x}_j - \bar{x})^2}{J-1}}{\frac{\sum_{j=1}^{J}\sum_{i=1}^{N}(x_{ij} - \bar{x}_j)^2}{N-J}}$$

Here, $\bar{x}$ is the grand mean of all scores. Similar to the traditional *t*-test, the ANOVA *F*-test is vulnerable to skewness and heterogeneous variances. First, the numerator is vulnerable to skewness, as the estimate of the variability of group means may be distorted by skewness in one or more groups. Second, the denominator is vulnerable to heterogeneity if group size imbalance is paired with heterogeneity, as the variances are weighted by their respective group size. When skewness and heterogeneity are combined, the deleterious effects on power and Type I error are compounded.

The vulnerabilities of the traditional *t*-test and ANOVA *F* test are well noted in statistics textbooks and journal articles and have been reviewed above. The exploration of the effects of differential skew on these tests, though, is novel. Obviously, if both distributions are equally skewed in the same direction, group mean differences and group variances are not affected. If the degree or direction of skewness differs between the groups, however, the mean difference may be either overestimated (increasing chances of a Type I error) or underestimated (increasing chances of a Type II error). Further, differential skew may render s_p^2 invalid because the pooled variance may be affected by the degree to which each group is skewed. This may decrease the ability of the traditional *t*-test to control for Type I error and power; the denominator for the *F* statistic will be likewise vulnerable. The effects of differential skewness is an area not well explored and one that needs to be clarified empirically

Existing Solutions for Nonnormality and Variance Heterogeneity

The continued popularity of traditional procedures makes it clear that dealing with skewness and heterogeneity has yet to pervade the culture of psychological research. Several solutions, however, are available and discussed below. These may not have been adopted simply because they are unfamiliar to researchers, cumbersome, or cannot be performed in popular statistical software packages.

Trimmed means. One common approach to dealing with problematic data is to apply a trimming technique, where a proportion of scores are eliminated from the extremes of a distribution. This is generally done in a symmetric manner, although asymmetric trimming has also been recommended in the case of outliers or specifically shaped asymmetric distributions (Hogg, Fisher, & Randles, 1975; Keselman, Wilcox, Othman, & Fradette, 2002). The amount of recommended trimming ranges from 5% to 25%; 20% is a common choice and recommended by Wilcox (2005) as the most effective at increasing statistical power. In order to obtain an accurate estimate of variability when the extreme observations have been trimmed from a distribution, a Winsorized variance is often computed by replacing the trimmed observations with the highest (upper-tail) or lowest (lower-tail) untrimmed observations. Detailed computations are provided below.

Using procedures that include trimmed means has the potential to improve Type I error rates and power substantially, but it should be noted that the null hypothesis no longer tests the equality between population means but rather trimmed means (i.e., $\mu_{(t1)} = \ldots = \mu_{(tJ)}$). However, this is consistent with the goal of most researchers, which is to compare the means of groups without those means being overly influenced by extreme observations (in other words, to compare the bulk of the scores in each population).

Adjusted degrees of freedom – heteroscedastic procedures. Welch (1938) addressed variance heterogeneity by 'unweighting' the influence of group size by adjusting the degrees of freedom (*df*), which is the number of values that are free to vary, based on the total sample size (*N*) and the number of groups (*J*). The Welch procedure (t_w) uses the traditional *t*-statistic but without pooling the sample variances. The t_w tends to deal well with moderate skewness when variances are unequal, but does not do well when extreme skewness is combined with variance heterogeneity, as found by a 11.6% empirical Type I error rate when $N = 20$, VR = 3:1, unequal (1:3) group sizes, and skewed distributions (Algina et al., 1994). The t_w also had poor Type I error control when unequal variances and group sizes were combined with skewed distributions, particularly when skewness was unbalanced (Cribbie et al., 2011).

Adjusted *df* combined with trimmed means. To deal better with skewed distributions (along with extreme scores), Yuen (1974) used a procedure with trimmed means and Winsorized variances. Winsorization is the process of replacing trimmed scores with the most extreme non-trimmed scores at each end of the distribution and computing the variance using the full complement of replacement scores. The denominator of the Yuen procedure includes Winsorized variances and thereby deals with variance heterogeneity and skewness, which has been shown empirically (Yuen, 1974; Wilcox, 1990). In particular, Cribbie et al. (2011) showed that Type I error for this procedure was acceptably between 4.8% and 6.7% under conditions of unequal group sizes and variability and when distributions were skewed, including when they were differentially skewed when there were three groups. When there were eight groups, however, Type I error for the Yuen procedure for multiple groups rose up to 9.2% when variances were unequal and distributions were differentially skewed. Power rates for this procedure were often as high as or higher than other procedures in the study both when there were three and eight groups.

Common Transformations. Transformations for positive skewness include the square root or logarithm and squaring or cubing for negative skewness. Transforming will adjust the shapes of distributions (towards normal, or possibly even towards normality and variance equality) such that popular parametric tests can be performed without violating assumptions. There are, however, limitations to using transformations. First, reporting results becomes cumbersome. Second, if skewness is unequal, a transformation might improve one distribution, but may do nothing to help with (or even make more nonnormal) the shape of the other group.

Ranking. Another transformation, ranking, is the substitution of ranked scores for raw data and has also been used to deal with skewed distributions. Zimmerman and Zumbo (1992) investigated ranking as a way to increase the power of tests of group differences. When distributions were skewed, the *t*-test using ranked data (i.e., the Wilcoxon-Mann-Whitney procedure; Mann & Whitney, 1947) showed better Type I error control and higher power than the traditional *t*-test on raw data. For example, when distributions were lognormal and group sizes were both 36, empirical Type I error for the *t*-test was 1%, while for the *t*-test on ranked data was 5%. The conditions of this study did not extend to unbalanced group sizes or heterogeneous variability. Similarly, Zimmerman and Zumbo (1993) found that transforming scores to ranks was beneficial in maintaining nominal α rates and achieving higher power than the traditional *t*-test in the case of non-normal and outlier-prone distributions, again with homogeneous data. Specifically, the authors demonstrated that the *t*-test on ranked data had empirical Type I error rates closer to the nominal α than the *t*-test across eight different distribution shapes. The authors note that simply applying ranks to the scores and performing a parametric test like the *t*-test does not protect against error when data are skewed and variances are unequal. Fagerland and Sandvic (2009) also tested the traditional *t*-test on ranks in conditions of skewness and heterogeneity and found that the test was robust to all degrees of skewness provided the variances were equal, but when VRs were very small (1.2:1) and skewness was moderate to extreme, the test had Type I error rates ranging from 7% to 17%. When the VR was 4:1 with moderate to extremely skewed distributions, Type I error rates soared as high as 63%. With these results evident, it is clear that ranking alone does not provide a good balance of Type I error and power.

Johnson Transformation. Other methods of transformation have been introduced to reduce or eliminate skewness. Often, transformation are applied to procedures known to deal with heterogeneity and this tends to improve the performance of the test when nonnormality is combined with heterogeneity. One popular transformation was introduced by Johnson (1978; defined below) as a way to address skewness. When applied to between groups procedures such as the Welch-James ($WJ_{(Jn)}$, detailed below), the transformation improves its ability to maintain acceptable Type I error rates. Luh and Guo (1999) found that for two-group designs, the $WJ_{(Jn)}$ showed Type I error between 2.5% and 7.5% under various conditions of skewness and variance heterogeneity,and power was about the same as for other procedures compared in the study. For one-way designs, Type I error control was maintained between 2.5% and 7.5% under all conditions of unbalanced group sizes, variance heterogeneity, and skewness when there were six groups and except in one condition of skewness, negative pairing and unequal samples sizes. In this case, the Type I error rose to 8.1%. In this same study, power for the $WJ_{(Jn)}$was similar to other procedures across the variety of data conditions. The *WJ* without the transformation, however, was not robust in the case of combined heterogeneity and skewness. Cribbie et al. (2011) examined procedures for two-groups differences when distributions were extremely skewed and variances and group sizes were negatively paired, and found that the Type I error rate was 11%. Regarding one-way designs with four and six groups under conditions of heterogeneity and unbalanced group sizes, Wilcox (1989) showed that the *WJ* had good control over Type I error with rates between 3.8% and 6.3%, but power rates declined considerably as data conditions worsened, which is expected, particularly with small effect sizes (δ). Specifically, under the worst data conditions (VR = 6:1 and *n'*s ranging from 11 to 30), the *WJ* had power rates as low as 6%. As expected, as δ increased, so did power. When δ = 2, power was 12% and when δ = 3 power was 21%.

Trimming and Transforming. Some researchers have combined trimming and transforming to deal with combined heterogeneity and skewness. Luh and Guo (2005) transformed data using the $WJ_{(Jn)}$ with trimming ($WJ_{t(Jn)}$), and found that empirical Type I error rates for all procedures tested with transformed data were closer to the nominal α when data were trimmed and transformed (6.8%) than when the data were just trimmed (7.2%) and when the data were not trimmed or transformed (12.7%). Further, power rates for the trimmed and transformed data were higher than power rates for trimmed-only or raw data. Likewise, when just trimming was applied, Keselman et al. (2002) found that when both skewness and heterogeneity existed, and when they trimmed 20% of the data, the $WJ_{t(Jn)}$ with both positive and negative pairing of group sizes and variability (VR of 1:36, $N = 70$) had an acceptable average Type I error rate of 5.8%.

Based on these findings, it can be concluded that the $WJ_{t(Jn)}$ is robust when data do not meet assumptions, but it is unclear whether this will be the case with unequal and opposed-direction skewness.

Bootstrapping. Bootstrapping is the process of generating a sampling distribution by drawing continuously from a set of sample data with replacement a pre-determined number of times. The upper and lower tails of this sampling distribution at $\alpha/2$ denote the critical cutoff values for the test statistic used to determine differences between groups. Keselman, Algina, Lix, Wilcox, and Deering (2008) adopted a bootstrap Yuen procedure with trimmed means and transformed data that maintained empirical Type I error rates closer to nominal α, ranging from .045 to .059, than non-bootstrapped and non-transformed data in the presence of skewed distributions and unequal variances with VR as high as 36:1. Further, Keselman et al. (2002) applied a boostrap Welch-James procedure when data from four groups were skewed, with and without also applying trimming and transforming. In all cases, bootstrapped procedures had better control over Type I error rates (between 2.5% and 5.3%) than non-bootstrapped procedures. Unfortunately, power results for the two- and four-group bootstrap procedures were not provided by the authors but they note that by trimming and Winsorizing, power is increased when using the Yuen procedure.

Adjusted *df* with Rank Transformed Data. Of particular interest to this study is the use of the heteroscedastic t_w with rank transformed data to accommodate skewed distributions. The t_w is detailed above and simple ranking is applied to the dependent variable prior to analyses. Ranking data and performing this procedure is hypothesized to address co-occurring nonnormality and heterogeneity. Zimmerman and Zumbo (1993) showed that this procedure was more robust to combined heterogeneity and skewness than the traditional *t* test. Later, Cribbie, Wilcox, Bewell, and Keselman (2007) found that when ranking data and using a Welch procedure for one-way designs with four or seven groups, in the presence of moderate or heavy skewness, and equal or unequal variances, W_r showed good Type I error control, with rates between 5.7% and 6.4%, and power rates were higher than the other procedures compared in the study (including t_w and t_y). These results point to a promising way to deal with violations of the variance homogeneity and normality assumptions when all distributions are equally skewed. It is not yet clear how the W_r will perform when the distributions are differentially skewed or how the W_r will compare to the trimmed, transformed, and bootstrapped procedures.

Procedures Selected for Analyzing Two Groups

The traditional *t*- and ANOVA *F*-tests were examined for the sake of comparison, along with the W_r and the following procedures.

Welch (t_w).

$$t_w = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

And the adjusted $df = \frac{\left(\frac{s_1^2}{n_1}+\frac{2_2^2}{n_2}\right)}{\frac{s_1^4}{n_1^2(n_1-1)}+\frac{s_2^4}{n_2^2(n_2-1)}}$.

Welch Procedure on Trimmed Means (t_y). If $X_1,\ldots,X_n$ is a random score from a group, let $X_{(1)} \leq \ldots \leq X_{(n)}$ be ordered scores. Let $e = [\beta n]$ where $[x]$ (score rounded down to integer less than or equal to x) is the highest score less than or equal to x and β is the proportion of trimmed scores. Then the trimmed sample size is then $f = n - 2e$ and the trimmed sample mean is

$$\bar{X}_t = \sum_{i=e+1}^{n-e} \frac{X_{(i)}}{fX}.$$

The two group Welch procedure with trimmed means is:

$$t_y = \frac{\bar{x}_{t1} - \bar{x}_{t2}}{\sqrt{d_1 + d_2}} \text{, where}$$

$$d_j = \frac{(n_j - 1)s_{tj}^2}{n_{tj}(n_{tj} - 1)},$$

Where n_{tj} is the group size after trimming, $\bar{x}_{tj}$ is the trimmed mean for the j^{th} group. Here, *df* is:

$$v_y = \frac{(d_1 + d_2)^2}{\frac{d_1^2}{n_{t1} - 1} + \frac{d_2^2}{n_{t2} - 1}}$$

and the null hypothesis $\mu_{1t} = \mu_{2t}$ is rejected based on the *t*-distribution using α and v_y.

Welch-James with Trimming and Johnson Transformation ($WJt_{(Jn)}$). The sample Winsorized mean is

$$\bar{X}_w = \sum \frac{Y_i}{n} \text{, with}$$

$$y = \begin{cases} x_{(e+1)} & \text{if } x_i \le x_{(e+1)}, \\ x_i & \text{if } X_{(e+1)} < x_i < x_{(n-e)}, \\ x_{(n-e)} & \text{if } x_i \ge x_{(n-e).} \end{cases}$$

The Winsorized variance for the sample is

$$S_w^2 = \sum \frac{(y_i - \bar{x}_w)^2}{(n-1)}.$$

Let

$$\hat{\mu}_w = {}^{n\mu_3}/_{f}$$

and the squared standard error for the (trimmed) sample is

$$\vartheta_w^2 = \frac{(n-1)S_w^2}{(f-1)}(f-1).$$

Further,

$$q_j = \frac{\vartheta_{wj}^2}{f_j},$$

$$w_{tj} = \frac{1}{q_j},$$

$$U_t = \sum w_{tj}, \text{ and}$$

$$\bar{X}_t = U_t^{-1} \sum w_{tj} \bar{x}_{tj}.$$

Now, the trimmed mean test statistic for Johnson's transformation is

$$T_j^* = (\bar{x}_{tj} - \tilde{x}_t) + \frac{\hat{\mu}_{wj}}{\left(6\hat{\sigma}_{wj}^2 f_j\right)} + \frac{\hat{\mu}_{wj}(\bar{x}_{tj} - \tilde{x}_t)^2}{(3\hat{\sigma}_{wj}^4)}$$

and the Welch-James procedure with Johnson transformation is determined by:

$$WJ_{(Jn)} = \sum w_{tj} \left(T_j^*\right)^2.$$

The null hypothesis for $WJ_{t(Jn)}$ is rejected if $WJ_{t(Jn)} > h_t$, where

$$h_t = c + \left(\frac{1}{2}\right)(3C_4 + C_2)\frac{\sum\left(1 - w_{tj}/U_t\right)^2}{(f_j - 1)} + \left(\frac{1}{16}\right)(3C_4 + C_2)^2\left(1 - \frac{(J-3)}{c}\right)\left\{\frac{\sum\left(\frac{1-w_{tj}}{U_t}\right)^2}{(f_j - 1)}\right\}^2 +$$

$$\left(\frac{1}{2}\right)(3C_4 + C_2)\Big\{(8R_{t23} - 10R_{t22} + 4R_{t21} - 6R_{t12}^2 + 8R_{t12}R_{t11} - 4R_{t11}^2) + (2R_{t23} - 4R_{t22} +$$

$$2R_{t21} - 2R_{t12}^2 + 4R_{t12}R_{t11} - 2R_{t11}^2)(C_2 - 1) + \left(\frac{1}{4}\right)(-R_{t12}^2 + 4R_{t12}R_{t11} - 2R_{t12}R_{t10} -$$

$$4R_{t11}^2 + 4R_{t11}R_{t10} - 4R_{t10}^2) \times (3C_4 - 2C_2 - 1)\Big\} + (R_{t23} - 3R_{t22} + 3R_{t21} - R_{t20})(5C_6 +$$

$$2C_4 + C_2) + \left(\frac{3}{16}\right)(R_{t12}^2 - 4R_{t23} + 6R_{t22} - 4R_{t21} + R_{t20}) \times (35C_8 + 15C_6 + 9C_4 + 5C_2) +$$

$$\left(\frac{1}{16}\right)(-2R_{t22} + 4R_{t21} - R_{t20} + 2R_{t12}R_{t10} - 4R_{t11}R_{t10} + R_{t10}^2) \times (9C_8 - 3C_6 - 5C_4 - C_2) +$$

$$\frac{1}{4}(-R_{t22} + R_{t11}^2)(27C_8 + 3C_6 + C_4 + C_2) + \left(\frac{1}{4}\right)(R_{t23} - R_{t12}R_{t11})(45C_8 + 9C_6 + 7C_4 + 3C_2),$$

where for any positive integers *a* and *b*, let

$$R_{tab} = \sum_{j=1}^{J} \frac{w_{tj}^{b}}{(f_j - 1)^2 U_t^b}$$

and

$$C_{2a} = \frac{c^a}{(J-1)(J+1)\dots(J+2a-3)}$$

where c is the $1 - \alpha$ quantile of the chi-square distribution with $J - 1$ degrees of freedom.

Bootstrapped Yuen Procedure with Trimming and Johnson Transformation (*Bty(Jn)*). Let $C_{ij} = Y_{ij} - \hat{\mu}_{ij}$ be the empirical (centered) values of the j^{th} group. For each j, β random samples of n_j replacement scores are drawn (with replacement), giving $Y_1^*, \dots, Y_{nj}^*$. Now t_y^* is the value of Yuen's test statistic based on the C_{ij} values calculated β times and put in ascending order. Given $= {}^{\alpha\beta}/_{2}$, rounded to the nearest integer, and $u = \beta - 1$, the null hypothesis (H_0: $\mu_{t1} = \mu_{t2}$) is rejected when $t_y \le t_{y(l)}^*$ or $t_y \ge t_{y(u)}^*$ where t_y is Yuen's test statistic on original data. The 100(1-α)/2 percentile interval for $\mu_1 = \mu_2$ is

$$\left[(\hat{\mu}_{t1} - \hat{\mu}_{t2}) - t_{y(u)}^* S_w^2, (\hat{\mu}_{t1} - \hat{\mu}_{t2}) - t_{y(l+1)}^* S_w^2\right].$$

With Johnson's transformation applied to the interval, it becomes

$\left[\left\{(\hat{\mu}_{t1} - \hat{\mu}_{t2}) + \frac{\tilde{\mu}_w}{6S_w^2}\right\} - t_{y(Johnson)(u)}^* S_w^2, \left\{(\hat{\mu}_{t1} - \hat{\mu}_{t2}) + \frac{\tilde{\mu}_w}{6S_w^2}\right\} - t_{y(Johnson)(l+1)}^* S_w^2\right]$ (Keselman et al., 2008).

Overview of the Current Studies

The above procedures, t_y, $WJ_{t(Jn)}$, $Bt_{y(Jn)}$, and W_r , present themselves as viable alternatives to the *t*-test and t_w when data are skewed and variances are unequal. My goal was to find a suitable procedure for independent groups designs that performs well in terms of empirical Type I error control and power when the data do and do not meet all of the underlying assumptions of the traditional parametric *t* and *F* tests. Most importantly, I sought to find a procedure that has all of the above characteristics and can also deal effectively with differential between-group skewness under conditions of variance heterogeneity. To determine this, I compared the $WJ_{t(Jn)}$, t_y, B$t_{y(Jn)}$, W_r, and the traditional t, and t_w tests under these conditions.

Study One

Method

A Monte Carlo simulation study was conducted using 5000 replications of each condition. R-project software (R Development Core Team, 2010) was used, with random normal data generated using the *rnorm* function. The variables manipulated were sample size, degree of group size imbalance, variance inequality, pairing of unequal sample sizes and variances (positive/negative), group distribution shapes, and means.

Total sample sizes of N = 60 and N = 100 were produced for power estimates of 80% and 95% for normally distributed, homogeneous data with equal sample sizes and the means presented below, respectively. When N = 60, group sizes were set to 30,30; 25,35; and 20,40. When N = 100, group sizes were set to 50,50; 40,60; and 30,70.

Variance ratios were set to 1:1, 1:4 and 1:8, which were found by Keselman et al. (1998) to be common in psychological testing. The unequal variance ratios were also reversed to test for negative pairings of unequal group sizes and variances.

Distribution shapes were manipulated to be normal, skewed, and to have differential between-group skewness. The data were drawn from Hoaglin's (1985) *g*- and *h*- distributions, where g is manipulated to create varying levels of skewness and h to create varying levels of kurtosis. In the current study, distributions were normal (g = 0, h = 0), negatively or positively moderately skewed (g = .5, h = 0; skewness = 1.75), and negatively or positively heavily skewed (g = 1, h = 0; skewness = 6.18). Standard normal variables (Z) were converted to

$\varepsilon = g^{-1}[\exp(gZ)-1] \exp(hZ^2/2)$. Note that when $g = 0$, $= Zexp(\frac{hZ^2}{2})$. The mean of the g- and h-distribution equals 0 when g = 0 and as such, multiplying each ε_{ij} by σ_j to get unequal variances did not affect the null hypothesis. For $g > 0$, the mean of the g- and h- distribution was

$$\mu_{gh} = \frac{\left(\exp\left\{\frac{g^2}{[2(1-h)]}\right\}-1\right)}{\left[g(1-h)^{\frac{1}{2}}\right]}$$

(Luh & Guo, 1999). For untrimmed data, the population mean (μ_{gh}) was subtracted from ε before multiplying by the group standard deviation (σ_j) to maintain the validity of the null hypothesis. For trimmed data, the population trimmed mean (μ_{tgh}) was subtracted from ε. Thus, the ij^{th} observation for untrimmed data is then $X_{ij} = \mu_j + \sigma_j \times (\varepsilon_{ij} - \mu_{gh})$ and for trimmed data is $X_{ij} = \mu_j + \sigma_j \times (\varepsilon_{ij} - \mu_{tgh})$. It is important to note that the variance of ε_{ij} is not equal to one and therefore the variances of the distributions are not equal to σ_j but are proportional to σ_j (see Lix & Keselman, 1998).

For the W_r, the population mean rank is not equal across cells when the distribution shapes are skewed and the variances are unequal. Therefore, for each condition of skewness and variance heterogeneity, the distribution of each group was first adjusted so that the population mean ranks were equal. Specifically, the empirically derived population mean rank for each condition was subtracted from X_{ij}.

The nominal Type I error rate was set at $\alpha = .05$. Empirical Type I error rates were considered acceptable if they fell between 2.5% and 7.5% (Bradley's, 1978, liberal limits). All procedures were tested under all conditions: two sample size conditions ($N = 60$; $N = 100$), 17 different pairings of skewness (two normal distributions, normal with heavy and moderate skewness, two similarly skewed distributions, one moderate and one heavy skewness, one positive one negative skewness with all combinations of moderate and heavy), five variance ratios (1:1; 1:4; 1:8; 4:1; 8:1), and six group sizes (30,30; 25,35; 20,40; 50,50; 40,60; 30,70) to a total of 1020 conditions for both power and Type I error rate. To assess power, I set means to 0 and 0.736 in data conditions of normality, homogeneity and equal group sizes. These means would yield 80% power under these data conditions when N = 60 and 95% power when N = 100. Given the volume of results, it was not feasible to discuss every condition. In most cases, the patterns of results were similar across several conditions but varied slightly in the degree of departure from nominal α, or in the degree of disparity between the power of the different procedures. Therefore, I have summarized the results below in general terms, providing examples of results from selected conditions. For those interested in the full results, the complete complement of tables for every two-group condition is provided in Appendix A.

Results

Similar Distribution Shapes

Two normal distributions. I demonstrate Type I error (Table 1) when the assumption of normality was met, and show that the t_w, t_y, $WJ_{t(Jn)}$, the $Bt_{y(Jn)}$, W_r, maintained empirical Type I error rates very close to α, but rates for the traditional t were inflated to almost 15% when VR = 8:1 and the group size and variance pairing was negative. The empirical Type I error rate for the traditional t was deflated to approximately 1% when the pairing of unequal group size and variance was positive.

In the case of N = 100 (Table 2), again the traditional t-test is the only procedure that cannot deal with heterogeneity and again in this case the vulnerability of the t-test to positive pairing of group size and variance is evident, as Type I error rates drop below the acceptable

2.5%. It can be seen here that the differences between $N = 100$ and $N = 60$ are marginal. Thus, only results for $N = 60$ are displayed and discussed below in terms of Type I error.

When distributions were both normal, empirical power rates for all procedures were fairly similar, predictably decreasing as VR and group sizes became more extreme. This can be seen in Table 3 for $N = 60$, which shows that the traditional *t*-test has the lowest power of all procedures any time variances are unequal. For example, when variances are 1 and 4 and group sizes are 20 and 40 respectively, the power of the traditional *t*-test is 30% while the lowest power among all other procedures is 41% (the highest is 47%).

Similar to the smaller group sizes, when $N = 100$, the traditional *t*-test showed the lowest power (Table 4), with rates as low as 14% in the case of positive pairing of group size and variance. Other procedures in this condition had power rates between 27% (t_y) and 32% (W_r). Again, since results for $N = 60$ and $N = 100$ are similar in pattern, power results will be shown for $N = 60$.

Two equally skewed distributions. For results for every condition of two-group equally skewed distributions, see Appendix A Tables 1 - 6. When both distributions were heavily positively skewed, empirical Type I error rate when $N = 60$ for the t_y, $WJ_{t(Jn)}$, $\mathrm{B}t_{y(Jn)}$, and W_r were all very close to α (Table 5). With negative pairing of group sizes and variability, empirical Type I error rates for the traditional *t* were 14% when VR = 4:1 and 18% when VR = 8:1. Empirical Type I error rates for the t_w were also inflated to 10% when VR = 4:1 and 11% when VR = 8:1.

When there were two heavily skewed distributions, all procedures declined in power as the data conditions worsened. The W_r maintained the highest empirical power in most of the scenarios, but the t_y showed higher power in certain cases such as when group sizes were equal and variances were not. All procedures performed as expected, with decreasing power as group size differences and variance ratios increased (Table 6).

Dissimilar Distribution Shapes

The primary interest of this study was the investigation of unequal skewness across distributions. The following results highlight the robustness of the investigated procedures under various conditions in which the distribution shapes are not the same across groups. For results for every condition tested with two groups and dissimilar distribution shapes, see Appendix A Tables 7 – 26.

One normal, one skewed distribution. When one distribution was normal and one heavily negatively skewed, with N = 60, empirical Type I error rates for the t_y, $WJ_{t(Jn)}$, $Bt_{y(Jn)}$, and W_r were all acceptably close to nominal α, whether or not variances and group sizes were unequal. The traditional t and the t_w, however, had empirical Type I error rates of up to 11% in cases of positive pairing of group sizes and variability and the traditional t had empirical rates of 9% in the case of negative pairing of group sizes and variability when VR=8:1 and sample sizes n_1 = 20, n_2 = 40 (Table 7).

Power was similar across procedures, with the W_r generally having slightly higher than the other procedures in most conditions; the traditional t was unpredictable in that power rates varied from the lowest to slightly higher than others, depending on the condition (Table 8).

Dissimilar and Directionally Opposed Distributions

One moderately positively skewed, one heavily positively skewed distribution. For full results when there were two groups and distributions were dissimilar and directionally opposite, see Appendix A Tables 27 – 54. In this case of directionally similar but differing degrees of skewness, and N = 60, Type I error rates for the traditional t and the t_w again strayed far from nominal α. When VR was 1:4 or 1:8 and group sizes were equal, Type I error rates for the traditional *t*-test and the t_w were high at 10% and 11%, respectively. All other procedures maintained acceptable Type I error rates, with slight inflations in the presence of positive pairing of group sizes and variability, but none that exceeded 7.5% (Table 9). The traditional *t*-test also had high Type I error rates when group sizes and variances were negatively paired, with rates as high as 10%.

In terms of power, the W_r maintained the highest power over all conditions, except when VR = 8:1 and group sizes were 30 and 70, where power for the W_r was about the same as all other procedures at 18% (Table 10) except the traditional *t*, which was slightly lower than the others at 27%.

First distribution positively skewed, second distribution negatively skewed. When the first distribution was heavily positively skewed and the second distribution was heavily negatively skewed and N = 60, the t_y, $WJ_{t(Jn)}$, $Bt_{y(Jn)}$, and W_r all maintained empirical Type I error close to nominal α. The traditional *t* and t_w had unacceptable Type I error rates under all conditions of variance heterogeneity and unbalanced sample sizes, with rates as high as 15% when VR = 4:1 and group sizes were 20 and 40 and 17% when VR = 8:1 and group sizes were 20 and 40. The t_w had consistently high Type I error rates across all conditions, ranging from 9% to 13% (Table 11).

Empirical power for all tests was similar across all conditions, and again deflated for all as variance heterogeneity and group size differences became more extreme (Table 12).

First distribution negatively skewed, second distribution positively skewed. When the first distribution was heavily negatively skewed and the second distribution moderately positively skewed with N = 60, empirical Type I error rates for the traditional *t* and the t_w were again unacceptable. Even with equal variances, Type I error rates for these the traditional *t* ranged from 9% to 10%, whether group sizes were equal or not, and when there was negative pairing of group sizes and variability. In this condition, Type I error rate was as high as 18%. The t_w type I error rates were also high across all conditions, ranging from 9% to 12%. (Table 13). All other procedures maintained acceptable Type I error control.

Power rates for the t_y, $WJ_{t(Jn)}$, the, the $Bt_{y(Jn)}$, and W_r were similar when variances were equal or when there was positive pairing of group sizes and variability. All four of these procedures had higher empirical power than the traditional *t* and the t_w. Indeed, these two procedures in all cases had power rates below 20%, even with equal variances and group sizes. In the case of negative pairing of group sizes and variability, the W_r maintained the highest power of all procedures, followed by the $Bt_{y(Jn)}$ (Table 14).

Study One Discussion

The primary goal of Study One was to investigate the effects of differential skewness in the presence of heterogeneous variances and unbalanced sample sizes to determine the procedure that provided the best balance of Type I error and power. The results highlight the vulnerability of popular traditional procedures and indicate the need for researchers to explore the nature of their data and understand the deleterious effects that 'imperfect' data may have. It is clear from the results of this study that the traditional *t*-test should not be used in any case where group sizes and variances are not equal, particularly when at least one of the distributions is not normal. Type I error rates stray too far from the nominal α. This is likewise the case for the Welch (t_w). Although the Welch procedure performs well under conditions of heterogeneity, or heterogeneity combined with mild nonnormality, it does not fare well in the presence of extreme skewness or dissimilar distribution shapes. From a Type I error control perspective, any of the other four procedures, the WJ_t, the t_y, the B$t_{y(Jn)}$, and W_r perform well even in the most extreme cases of heterogeneity, group size imbalance, and differential skewness, and also when data are homogeneous and normally distributed. In terms of power, however, the W_r generally had higher power than other procedures when data violated the assumptions, and was about as powerful as other tests when there were no assumption violations.

It is important to note that the current study was limited as not all possible data conditions were tested. Indeed, there are infinite possibilities for scenarios of degree and direction of skewness, presence and degree of heterogeneity, and degree of difference between group sizes. However, I applied a very wide range of conditions and am confident that the general pattern of results obtained in this study would extend to many conditions of assumption violation.

In light of the findings, I recommend that researchers begin to incorporate the W_r in their toolbox of analytic resources. It is robust to sample size imbalance, heterogeneity, and a variety of conditions involving nonnormality, is generally more powerful than the available procedures, and can be easily implemented in almost any software package (e.g., SPSS, R, SAS) since all that is required is to rank the data and perform a standard Welch test. None of the other robust procedures investigated in this study is currently available in mainstream software packages.

Study Two

In Study One, I demonstrated robust alternatives that offer a good balance of Type I error control and power and recommended the use of the Welch procedure on ranked data, as it was beneficial in terms of Type I error and consistently had higher or similar empirical power rates than other procedures. However, Study One is limited to designs with only two groups. Therefore, the purpose of Study Two was to expand the investigations to multiple-group designs such that the findings can be more broadly applicable in psychological research.

Method

In Study Two, I investigated empirical Type I error and power when distributions were skewed and variances were heterogeneous and negatively and positively paired for one-way designs with four and seven groups. After amending procedures to multiple-group designs, I examined the procedures from Study One that were robust. Specifically, along with the ANOVA *F*, I investigated three other procedures. We included the Yuen procedure (t_y) which needed to be amended for multiple groups and as such has a null hypothesis $\mu_{t1} = \ldots = u_{tj}$ ($j = 1, \ldots, J$) and is rejected if $t_y \geq F_{\alpha J-1, v_w}$. The procedure is defined as:

$$t_y = \frac{\sum_j w_j(\bar{x}_j - \bar{x}'_{.})^2}{J-1}\left(1 + \frac{2(J-2)}{J^2-1}\sum_j\left(\frac{1}{n_j-1}\right)\left(1 - \frac{w_j}{\sum_j w_j}\right)^2\right)^{-1} \text{ with}$$

$$v_w = \frac{J^2 - 1}{3\sum_j\left(\frac{1}{(n_j-1)}\right)\left(1 - \left(\frac{w_j}{\sum_j w_j}\right)\right)^2}$$

$$w_j = \frac{n_j}{s_j^2} \text{ and}$$

$$\bar{X}'_{.} = \frac{\sum_j w_j \bar{x}_j}{\sum_j w_j}$$

where s is the standard deviation, $w_j = \frac{n_j}{s_j^2}$, and $F_{\alpha J-1, v_w}$ is the F critical value for α with $J - 1$ *and* v_w degrees of freedom. The third procedure I used was the Welch-James with Johnson transformation $WJ_{t(Jn)}$, and the Welch procedure on ranked data (W_r). The Welch procedure also needed to be adjusted for multiple groups, and as such can be defined as F_w. The null hypothesis $\mu_1 = \ldots = u_j$ ($j = 1, \ldots, J$) is rejected if $F_w \geq F_{\alpha J-1, v_w}$ and is defined as:

$$F_w = \frac{\frac{\sum_j w_j \left(\bar{x}_j - \bar{x}_j^*\right)^2}{(J-1)}}{1 + \left\{\frac{2(J-2)}{J^2-1}\right\} \left\{\frac{\sum_j \left(1 - w_j / \sum_j w_j\right)^2}{(n_j - 1}\right\}} \text{ where}$$

$$x_j^* = \frac{(\sum_j w_j \bar{x}_j)}{(\sum_j w_j)} \text{ and}$$

$$w_j = \frac{n_j}{s_j^2},$$

s_j^2 is the sample group variance, and there are v_w degrees of freedom, where

$$v_w = \frac{(J^2 - 1)}{\left\{\frac{3 \sum_j \left(1 - w_j / \sum_j w_j\right)^2}{(n_j - 1)}\right\}}.$$

These four procedures were chosen as they showed the best balance of Type I error control and power in Study One. I did not include the bootstrap procedure $Bt_{y(Jn)}$, as the results for the procedure were almost identical to those of the $WJ_{t(Jn)}$, and it would be repetitive to show two such similar procedures here.

I reduced the number of conditions investigated in Study Two to only those that were the most problematic in terms of empirical Type I error and power for the two-group designs. Specifically, for four group designs, I set variances to equal (1,1,1,1) or unequal (1,3,5,7), group sizes to equal (30,30,30,30) or unequal (20,27,33,40) and reversed the order of group sizes to simulate negative pairings of group sizes and variability (40,33,27,20). Distribution shapes were set to all normal, all skewed, moderately skewed with heavy skewed (same direction), normal with heavy skew, and heavy positive skew with heavy negative skew. There was a total of eight different shape configurations. In the seven-group designs, I simply extended these conditions to variances equal at one or unequal (1,2,3,4,5,6,7), group sizes equal (30,30,30,30,30,30,30) or unequal (20,24,27,30,33,36,40) and reversed the latter to simulate negative pairings of group size and variances. Again, the eight different distribution shape configurations were: 1) normal; 2) three normal and four heavily skewed; 3) two normal, three positively skewed and two negatively skewed; 4) three moderately skewed and four heavily skewed (same direction); 5) all heavily skewed; and 6) three heavily positively skewed with four heavily negatively skewed.

To assess power, I set means to 0, 0.271, 0.542, and 0.813 in data conditions of normality, homogeneity ($\sigma^2 = 1$) and equal group sizes ($N = 120$) in the four group model and to 0, 0.1295, 0.259, 0.3885, 0.518, 0.6475, 0.777 ($N = 210$) in the seven group model. These means yield 80% power under these data conditions.

Results

The deleterious effects of skew and heteroscedasticity found for two group designs also emerged in multiple group one-way designs. In general, the ANOVA *F*-test was vulnerable to variance heterogeneity and skewed data in all scenarios of distribution shape variation. The other procedures were typically more robust in terms of Type I error control, and power tended to decline as data strayed further from equal group sizes and variability and normality and in particular when normality was not equal across groups. The specific results are below.

Identical Distribution Shapes

All groups normally distributed. When all group distributions were normal, the t_y, $WJ_{t(Jn)}$, and W_r procedures had acceptable empirical Type I error, even when variances and group sizes were heterogeneous, with four groups (Table 15) and seven groups (Table 16). The ANOVA F test, however, had Type I error rate of 11% in the case of negative pairing of group size when there were four groups and 10% in the case of negative pairing of group sizes and variability when there were seven groups.

Empirical power for all procedures was similar at around 80% when distributions were normal with homogeneous variances in the four and seven group designs. Power rates declined when variances were heterogeneous, ranging from 17% (for the ANOVA F) to 32% (for the Wr) in the four-group conditions (Table 17) and from 16% (ANOVA F) to 30% (Wr) in the seven-group conditions (Table 18).

All groups heavily positively skewed. When the data for all groups were heavily skewed, empirical Type I error rates were well-maintained by all procedures when variances were equal. When variances were heterogeneous, in the four group design, the t_y, $WJ_{t(Jn)}$, and W_r, maintained acceptable Type I error rates but the ANOVA F had rates of 10% when there was negative pairing of variances and group sizes (Table 19). In the seven group designs, the $WJ_{t(Jn)}$, and W_r maintained acceptable Type I error rates but the ANOVA F and the t_y had rates of 8% (Table 20).

In terms of power for the four-group conditions, the W_r maintained the highest power of between 82% and 89% when variances were equal and between 37% and 41% when heterogeneous. The $WJ_{t(Jn)}$ and t_y had rates between 50% and 60% when variances were equal and between 11% and 22% when they were not. Finally, the ANOVA F had the lowest power, between 26% and 27% when variances were equal and between 4% and 10% when they were not (Table 21). In the seven-group conditions, the W_r again showed the highest power whether variances were equal (63% - 69%) or not (48% - 59%) and the ANOVA *F* showed the lowest at 20% to 24% when variances were equal and 4% to 12% when they were not (Table 22).

Dissimilar Distribution Shapes

Normal distributions combined with identically skewed distributions. In the four-group conditions, I set two distribution shapes to normal and two to heavily skewed. For these, empirical Type I error rates were acceptable for the t_y, $WJ_{t(Jn)}$, and W_r procedures but for the ANOVA *F* were 10% when variances were positively paired with group sizes and 17% when they were negatively paired (Table 23). In the seven-group conditions, I set three shapes to normal and four to heavily positively skewed and here, Type I error was adequately maintained by the $WJ_{t(Jn)}$, and W_r, but were 9% for the ANOVA *F* when variances were equal. When there were unequal variances, the t_y had Type I error rates of 8% to 9% and the ANOVA *F* had rates of 14% when there was negative pairing of group sizes and variability (Table 24).

In terms of power for the four-group conditions, when variances were equal, the W_r maintained power between 79% and 83%, the t_y and $WJ_{t(Jn)}$ had power rates between 57% and 67% and the ANOVA *F* had rates from 23% to 40%. When variances were heterogeneous, the W_r again maintained the highest power, ranging from 29% to 32% while the other procedures ranged from 3% (in the case of ANOVA *F*) to 22% (in the case of $WJt_{(Jn)}$) (Table 25). With seven-group conditions, power rates followed the same pattern as in the four-group conditions; the W_r maintained power at between 80% and 84% when variances were equal and at 29% when they were not. The t_y and $WJ_{t(Jn)}$ had lower power rates, similar to each other, at between 50% and 66% when variances were equal and between 9% and 19% when they were not. The

ANOVA *F* had the lowest power between 20% and 36% when variances were equal and between 8% and 15% when they were not (Table 26).

Normal distributions combined with differentially skewed distributions. In the four-group conditions, I set one group to normal, two to heavily positively skewed, and one to heavily negatively skewed. In the seven-group conditions, I set two groups to normal, three to heavily positively skewed, and two to heavily negatively skewed. Empirical Type I error rates for the four-group conditions were adequately maintained by all procedures except the ANOVA *F*, which had empirical Type I error rates of 10% when variances were positively paired with group sizes and as high as 14% when variances were negatively paired with group sizes (Table 27). In the seven-group conditions (Table 28), the W_r maintained empirical Type I error in the acceptable range, but all other procedures had rates outside this range. Specifically, in the case of positive pairing of group sizes and variability, the t_y had error rates of up to 8% and the ANOVA *F* had rates of up to 9%. When there was negative pairing of variances and group sizes, the t_y and the $WJ_{t(Jn)}$ and the ANOVA *F* all had unacceptable Type I error rates (9%, 8%, and 14%, respectively) when variances and group sizes were negatively paired. The W_r maintained acceptable Type I error rates in across all of these seven-group designs.

Power rates for the four-group conditions were maintained similarly by the t_y, $WJ_{t(Jn)}$, and W_r, ranging from 54% to 67% when variances were equal and between 20% and 29% when variances were not equal. The ANOVA *F* showed lower rates; when variances were equal, power was between 32% and 39% and when variances were heterogeneous, power was between 13% and 23% (Table 29). In the seven-group conditions, the ANOVA *F* consistently showed the lowest power, and all other procedures were approximately the same, ranging from 50% to 69% when variances were equal and from 17% to 30% when they were not (Table 30).

Directionally similar with different degree of skewness. For the four-group conditions, I set one distribution shape to moderately positively skewed and three to heavily positively skewed. Empirical Type I error rates were well-maintained by all procedures when variances were equal but when they were not, the ANOVA *F* had Type I error rates between 7% and 11% (Table 31). In the seven-group condition, I set three distributions to moderately positively skewed and four to heavily positively skewed. Here, empirical Type I error was acceptable for all procedures except the ANOVA *F*, which had rates up to 7.5% when variances were equal and 13% when there was negative pairing of group sizes and variability (Table 32).

Power rates in the four group condition were highest for the W_r, ranging from 82% to 88% when variances were equal and from 38% to 41% when they were not. For the $WJ_{t(Jn)}$, and t_y, power ranged from 54% to 65% when variances were equal and from 11% to 22% when they were not. For the ANOVA *F*, power ranged from 22% to 29% when variances were equal and from 5% to 12% when they negatively paired (Table 33). For the seven-group conditions, similar patterns emerged with the W_r having the highest power, the $WJ_{t(Jn)}$, and t_y similar to each other but lower than W_r, and the ANOVA *F* consistently having the lowest power (Table 34).

Directionally dissimilar distributions. In the four-group conditions, I set two groups to heavily positively skewed and two to heavily negatively skewed**.** Acceptable Type I error rates for the four-group conditions when variances were equal were maintained by all procedures. When variances were heterogeneous, Type I error rates were maintained for the W_r, and $WJ_{t(Jn)}$, but when variances were negatively paired the t_y had rates of 8% and the ANOVA *F* of 12% (Table 35). In the seven-group conditions (Table 36), I set three groups to be heavily positively skewed and four to heavily negatively skewed. W_r maintained acceptable Type I error rates across all conditions. The $WJt_{(Jn)}$ had rates of 8% when variances and group sizes were positively paired, t_y had rates between 8% and 9% across all conditions, the ANOVA *F* had rates of 9% when variances and group sizes were negatively paired.

Power rates for the four-group condition were similar for all procedures except ANOVA *F*, which had rates between 33% and 34% when variances were equal and between 13% and 22%. The W_r had power rates slightly higher than the other procedures, but they all had power rates between 48% and 60% when variances were equal and between 21% and 19% when they were not (Table 37). Similarly, in the seven-group conditions, the ANOVA *F* had the lowest power across all conditions, ranging from 10% to 30% and the other procedures showed similar Type I error rates, with the W_r slightly higher than the other two procedures (Table 38).

Study Two Discussion

It is clear that the ANOVA *F* procedure is vulnerable to skewness and heterogeneity, and this vulnerability is extended to cases when skewness is differential across distributions. When data fit this description, Type I error is unacceptably high for the ANOVA *F*, particularly in the case of negative pairing of group size and variability. In some of these cases in this study, Type I error for the *F* test was as high as 16% when there were four groups and 14% when there were seven. In the same scenarios, the W_r held Type I error rates to 6% when there were four or seven groups. The $WJt_{(Jn)}$ and t_y also maintained Type I error rates close to nominal levels in these situations, but the W_r had higher power than these two procedures in both the seven- and four-group conditions. In general, Type I error is well maintained by the procedures designed to deal with both skewness and heterogeneity. Power is deleteriously affected across all procedures when data are skewed and heterogeneous, but is generally highest for the W_r, with the t_y and *WJt(Jn)* typically showing power rates somewhere between the *F* and the W_r.

General Discussion

It has long been understood that the traditional parametric *t* and *F* tests are vulnerable when data do not meet their underlying assumptions. It has also been consistently shown that procedures for heteroscedastic variances which that deal with skewness (either through transformation, trimming, or some combination of the two) will yield a better balance of Type I error and power. I have now extended these findings to show that procedures that deal with heteroscedastic data and skewness are also preferred when skewness is differential across groups and in particular the Welch test on ranked data does well to maintain Type I error and statistical power. These findings have implications across research in the behavioural and social sciences, education, and other relevant fields. If researchers endeavor to utilize procedures that deal well with skewness and variance heterogeneity, their results and subsequent recommendations are more likely to have a meaningful impact on the populations that they study.

It is primarily recommended that researchers survey the landscape of their data to understand the extent to which skewness and heterogeneity exist before conducting any analyses that determine differences between treatment groups. They can then proceed to utilize statistical tests that indicate genuine differences, and not accidentally find differences when they do not exist or miss differences when groups are truly different. As an adjunct to this recommendation, I suggest that the skewness and within-group variance of research data be reported in published studies as a part of the results of the investigation of group differences.

I have made a solid case for the adoption of the Welch procedure on ranks in terms of empirical robustness, that is, an excellent balance of Type I error control and power. My second recommendation is for researchers to consider the use of the Welch on ranks as a replacement of the *t* and *F* procedures, as the Welch on ranks performs well not only when data meet the assumptions of the traditional parametric procedures but also when variances are unequal, distributions are skewed (similarly or differentially) and also when data meet the assumptions of the traditional parametric procedures.

As an additional benefit to the Welch on ranks, it is a procedure that can be carried out in popular statistical software packages. Analysts can simply rank their raw data and perform the Welch two-group or Welch multiple-group procedures (often denoted by ‘variance homogeneity not assumed). As such, there is no need to learn new software.

Limitations

This study is not without its limits. First, the degree of skewness, skewness differential and direction, and group variability do not necessarily represent all possible combinations of data configurations which are found naturally in applied research. As such, there may be data conditions that would cause different types of results than those found in this study. However, the breadth of conditions and array of possible data configurations explored here indicate general patterns of Type I error and power findings that can be at least loosely assumed in cases where data are not normal or homogeneous and in particular when skewness is unbalanced. Further, in cases where data stray more extremely from normality and variance equality than examined here, it stands to reason that the robustness of traditional parametric tests (*t* and *F*) will be the lowest, while that of the presented alternatives, and in particular the W_r, will allow for more confident evaluation of differences in central tendency between groups.

The second limitation is the potential exclusion of procedures that have been presented as robust to detect differences between groups. Indeed, there are many extant procedures, and I selected the most robust and promising procedures available at the time of this study. The procedures chosen represented well-tested, novel, and often recently developed procedures across different approaches to dealing with skewness and heterogeneity (e.g., transforming, trimming, adjusting the *df*, and bootstrapping).

The final limitation is a more recently discovered issue with the *g* and *h* skewness distribution. The variances of these distributions are not equal to one, as has been informally assumed by many quantitative researchers. The variance ratios are all still accurate as ratios, but readers may assume that the variance ratios of, for example, 2:4 represent *variances* of two and four, respectively, but this is not the case. In fact, the variances are more extreme than two and four. Since the results are for actual variances higher than the shown variance ratios, the findings indicate that the robust alternatives to the traditional parametric tests are robust when variances are even more extreme than the variances implied by the variance ratios. This recent discovery thereby does not imply that the results of this study or the recommendations are erroneous.

Future Research Directions

The goal of this paper was to identify the vulnerability of commonly used procedures that test differences in central tendency between two or more groups and to develop information on procedures which provide robust ways to measure these differences, particularly in the face of data that are nonnormal or heteroscedastic. These types of quantitative methodology studies are becoming more prevalent in the psychology and educational literature but the adoption of robust procedures is slow, as evidenced by continued use of the traditional *t* and *F* procedures in recent applied studies. Indeed, it is often unknown wiether a robust procedure was needed, as skewness and heterogeneity of the data are typically not reported. While quantitative methodologists continue to try to encourage the uptake of robust statistical procedures in psychology, education, and other social sciences, more work can be done to expand the findings of this paper. For example, further study of the Welch procedure on ranked data for factorial designs is recommended. In particular, we published an earlier paper that investigated this procedure on 2 × 2 designs, but did not include skewness imbalance (Mills, Cribbie, & Luh, 2008). Future work could investigate the impact of heterogeneity combined with differential skewness on 2 × 2 and higher order factorial designs. This work would allow researchers to understand whether the procedure offers good balance of Type I error and power under various conditions of unequal variances and myriad configurations of skewness degree and direction on almost any study designed to measure the differences in central tendency between groups.

References

Alexander, R. A. & Govern, D. M. (1994). A new and simpler approximation for ANOVA under variance heterogeneity. *Journal of Educational Statistics, 19,* 91-101.

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

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

Cohen, J. (1962). The statistical power of abnormal social-psychological research: A review. *Journal of Abnormal and Social Psychology, 65*, 145-153.

Cohen, J. (1992). Power primer. *Psychological Bulletin, 112*, 155-159.

Coombs, W. T., Algina, J., & Oltman, D. O. (1996). Univariate and multivariate omnibus hypothesis tests selected to control Type I error rates when population variances are not necessarily equal. *Review of Educational Research, 66*, 137-179.

Cribbie, R. A., Fiksenbaum, L., Keselman, H. J., & Wilcox, R. R. (2012). Effect of non-normality on test statistics for one-way independent groups designs. *British Journal of Mathematical and Statistical Psychology*, 65, 56-73.

Cribbie, R. A., Wilcox, R. R., Bewell, C., & Keselman, H. J. (2007). Tests for treatment group equality when data are nonnormal and heteroscedastic. *Journal of Modern Applied Statistical Methods, 6*, 117-132.

Erceg-Hurn, D. M. & Mirosevich, V. M. (2008). Modern Robust Statistical Methods: An easy way to maximize the accuracy and power of your research. *American Psychologist, 63*, 591-601.

Fagerland, M. W. & Sandvik, L. (2009). The Wilcoxon-Mann-Whitney test under scrutiny. *Statistics in Medicine, 28*, 1487-1497.

Fisher, R. A. (1935). *The design of experiments*. Oxford, England: Oliver & Boyd.

Garrett, H. E. (1926). *Statistics in psychology and education*. Oxford, England: Longmans, Green.

Geary, R. C. (1947). Testing for normality. *Biometrika*, *34*, 209-242.

Golinski, C. & Cribbie, R. A. (2009). The expanding role of quantitative methodologists in advancing psychology. *Canadian Psychological Association, 50*, 83-90.

Greer, T., Hunter, S. T. Dunlap, W. P., & Berman, M. E. (2006). Skew and internal consistency. *Journal of Applied Psychology, 91*, 1351-1358.

Grissom, R. J. (2000). Heterogeneity of variance in clinical data. *Journal of Consulting and Clinical Psychology, 68*, 155-165.

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.

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.

James, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika, 38*, 324-329.

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

Keselman, H. J., Algina, J., Lix, L. M., Wilcox, R. R., & Deering, K. (2008). Supplemental materials to 129a. A SAS program to implement a general approximate degrees of freedom solution for inference and estimation. http://dx.doi.org/10.1037/1082-989X.13.2.110.supp

Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R., Donahue, B., et al. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research, 68,* 350-386.

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,* 281-287.

Leentjens, A. F. G., Wielaert, S. M., van Harskamp, F., & Wilmink, F. W. (1998). Disturbances of affective prosody in patients with schizophrenia; a cross sectional study. *Journal of Neurology and Neurosurgical Psychiatry, 64*, 375-378.

Lix, L. M., & Keselman, H. J. (1998). To trim or not to trim. Tests of mean equality under heteroscedasticity and nonnormality. *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.

Luh, W-M. & Guo, J-H. (2005). Heteroscedastic test statistics for one-way Analysis of Variance: The trimmed means and Hall's transformation conjunction. *Journal of Experiential Education, 74,* 75-100.

Mann, H. B., Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, *18*, 50-60.

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

Mills, L., Cribbie, R.A., & Luh, W-M. (2008). A heteroscedastic, rank-based approach for analyzing 2 x 2 independent groups designs. *Journal of Modern Applied Statistical Methods, 8,* 322-336.

Mosteller, F., & Tukey, J. (1977). *Data analysis and regression*. Reading, MA: Addison-Wesley.

R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Sartori, N. (2006). Bias prevention of maximum likelihood estimates for scalar skew normal and skew *t* distributions. *Journal of Statistical Planning and Inference, 136*, 4259-4275.

Sawilowski, S., & Blair, R. C. (1992). A more realistic look at the robustness and type II error properties of the t test to departures from population normality. *Psychological Bulletin, 111*, 353-360.

Sedlemeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin, 105*, 309-316

Tiku, M. L. (1964). Approximating the general non-normal variance-ratio sampling distributions. *Biometrika*, 51, 83-95.

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

Wilcox, R. R. (1987). New designs in analysis of variance. *Annual Review of Psychology, 38,* 29-60.

Wilcox, R. R. (1989). Adjusting for unequal variances when comparing means in one-way and two-way fixed effects ANOVA models. *Journal of Educational Statistics, 14*, 269-278.

Wilcox, R. R. (1990). Comparing the means of two independent groups. *Biometrika, 82*, 771-780.

Wilcox, R. R. (1992). Comparing one-step M-Estimators of location corresponding to two independent groups. *Psychometrika, 57*, 141-154.

Wilcox, R. R. (2005). New methods for comparing groups. *Current Directions in Psychological Science, 14*, 272-275.

Yin, T. S. & Othman, A. R. (2009). When does the pooled variance *t*-test fail? *African Journal of Mathematics and Computer Science Research, 2,* 56-62.

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

Zimmerman, D. W. (1987). Comparative power of Student *T* test and Mann-Whitney *U* test for unequal sample sizes and variances. *Journal of Experimental Education, 55,* 171-174.

Zimmerman, D. W. & Zumbo, B. D. (1992). Increasing the power of tests of location and correlation by transforming scores to ranks and indicators. *Journal of General Psychology, 119*, 409-418.

Zimmerman, D. W. & Zumbo, B. D. (1993). Rank transformations and the power of the student *t* test and Welch *t* test for non-normal populations with unequal variances. *Canadian Journal of Experimental Psychology, 47*, 523-539.

Appendix A

Results for Two-Group Conditions Not Included in Study Results
Similar Distribution Shapes

Table 1.

Empirical Type I Error with Two Moderately Positively Skewed Distributions, N = 60.

Variance Ratio	$n1,n2$	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*30 30*	0.0477	0.0468	0.0466	0.0494	0.0536	0.0495
	25 35	0.0449	0.0470	0.0456	0.0534	0.0512	0.0493
	20 40	0.0472	0.0508	0.0502	0.0576	0.0539	0.0491
1,4	*30 30*	0.0583	0.0561	0.0520	0.0480	0.0565	0.0602
	25 35	0.0346	0.0498	0.0475	0.0506	0.0522	0.0579
	20 40	**0.0231**	0.0493	0.0478	0.0502	0.0551	0.0619
1,8	*30 30*	0.0602	0.0579	0.0515	0.0540	0.0511	0.0637
	25 35	0.0354	0.0576	0.0502	0.0452	0.0520	0.0655
	20 40	**0.0194**	0.0547	0.0511	0.0616	0.0531	0.0719
4,1	*30 30*	0.0568	0.0556	0.0486	0.0526	0.0528	0.0580
	25 35	**0.0832**	0.0612	0.0528	0.0536	0.0558	0.0576
	20 40	**0.1186**	0.0688	0.0624	0.0656	0.0570	0.0606
8,1	*30 30*	0.0662	0.0628	0.0568	0.0568	0.0560	0.0706
	25 35	**0.1004**	0.0656	0.0540	0.0614	0.0496	0.0622
	20 40	**0.1450**	0.0628	0.0548	0.0644	0.0472	0.0592

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bold values reflect empirical Type I error outside the acceptable criterion (2.5% - 7.5%).

Table 2.

Empirical Power with Two Moderately Positively Skewed Distributions, N = 60.

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	30 30	0.7121	0.7110	0.6989	0.7192	0.6763	0.8589
	25 35	0.7064	0.6814	0.6736	0.6822	0.6451	0.8325
	20 40	0.6860	0.6373	0.6242	0.6398	0.5821	0.7830
1,4	30 30	0.3004	0.2928	0.3350	0.3470	0.3880	0.4724
	25 35	0.2329	0.3193	0.3645	0.3772	0.4034	0.4875
	20 40	0.1749	0.3455	0.3783	0.3934	0.4103	0.4840
1,8	30 30	0.1540	0.1463	0.1869	0.2008	0.2233	0.2937
	25 35	0.1025	0.1739	0.2124	0.2266	0.2492	0.3186
	20 40	0.0509	0.1988	0.2342	0.2406	0.2720	0.3338
4,1	30 30	0.3650	0.3618	0.3628	0.3698	0.3274	0.3844
	25 35	0.4090	0.3484	0.3370	0.3368	0.2990	0.3436
	20 40	0.4260	0.3204	0.2928	0.3222	0.2458	0.2858
8,1	30 30	0.2618	0.2554	0.2502	0.2394	0.2142	0.2574
	25 35	0.3110	0.2408	0.2328	0.2296	0.1974	0.2234
	20 40	0.3504	0.2550	0.2006	0.2142	0.1608	0.1828

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 3.

Empirical Type I Error with Two Moderately Positively Skewed Distributions, N = 100.

Variance Ratio	$n1,n2$		t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*50*	*50*	0.0503	0.0500	0.0492	0.0456	0.0552	0.0520
	40	*60*	0.0478	0.0491	0.0473	0.0518	0.0518	0.0480
	30	*70*	0.0474	0.0504	0.0488	0.0496	0.0524	0.0487
1,4	*50*	*50*	0.0586	0.0575	0.0533	0.0546	0.0559	0.0610
	40	*60*	0.0314	0.0509	0.0502	0.0522	0.0550	0.0591
	30	*70*	**0.0158**	0.0483	0.0496	0.0512	0.0550	0.0602
1,8	*50*	*50*	0.0572	0.0558	0.0521	0.0512	0.0523	0.0655
	40	*60*	0.0298	0.0534	0.0490	0.0564	0.0505	0.0681
	30	*70*	**0.0128**	0.0554	0.0534	0.0534	0.0583	0.0746
4,1	*50*	*50*	0.0512	0.0502	0.0502	0.0506	0.0522	0.0588
	40	*60*	**0.0852**	0.0554	0.0554	0.0508	0.0562	0.0596
	30	*70*	**0.1260**	0.0548	0.0556	0.0628	0.0502	0.0546
8,1	*50*	*50*	0.0522	0.0506	0.0482	0.0512	0.0480	0.0594
	40	*60*	**0.1032**	0.0616	0.0504	0.0514	0.0502	0.0658
	30	*70*	**0.1684**	0.0680	0.0586	0.0606	0.0530	0.0628

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 4.

Empirical Power with Two Moderately Positively Skewed Distributions, N = 100.

Variance Ratio	n_1	n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	50	50	0.8839	0.8835	0.8928	0.8970	0.8826	0.9785
	40	60	0.8721	0.8479	0.8663	0.8730	0.8500	0.9654
	30	70	0.8370	0.7840	0.8026	0.8110	0.7685	0.9255
1,4	50	50	0.4855	0.4808	0.5467	0.5508	0.5921	0.6867
	40	60	0.4074	0.5218	0.5853	0.5908	0.6167	0.7069
	30	70	0.2987	0.5374	0.5894	0.5868	0.6084	0.6939
1,8	50	50	0.2667	0.2604	0.3194	0.3340	0.3669	0.4366
	40	60	0.1825	0.3075	0.3636	0.3760	0.4097	0.4864
	30	70	0.0938	0.3420	0.3932	0.3760	0.4311	0.5067
4,1	50	50	0.5108	0.5074	0.5454	0.5326	0.5066	0.5832
	40	60	0.5300	0.4568	0.4764	0.4764	0.4256	0.5026
	30	70	0.5422	0.3954	0.4012	0.4056	0.3472	0.4004
8,1	50	50	0.3382	0.3348	0.3548	0.3450	0.3114	0.3660
	40	60	0.3966	0.3078	0.3074	0.3450	0.2534	0.3088
	30	70	0.4428	0.2734	0.2550	0.2686	0.2030	0.2412

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 5.

Empirical Type I Error with Two Heavily Positively Skewed Distributions, N = 100.

Variance Ratio	$n1$,$n2$		t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*50*	*50*	0.0440	0.0437	0.0465	0.0480	0.0526	0.0481
	40	*60*	0.0445	0.0473	0.0478	0.0488	0.0547	0.0514
	30	*70*	0.0438	0.0557	0.0512	0.0562	0.0536	0.0500
1,4	*50*	*50*	0.0686	0.0679	0.0571	0.0498	0.0576	0.0592
	40	*60*	0.0446	0.0556	0.0557	0.0476	0.0594	0.0590
	30	*70*	0.0302	0.0471	0.0542	0.0476	0.0583	0.0594
1,8	*50*	*50*	**0.0792**	0.0785	0.0653	0.0620	0.0612	0.0657
	40	*60*	0.0487	0.0676	0.0613	0.0522	0.0581	0.0663
	30	*70*	0.0276	0.0567	0.0603	0.0492	0.0605	0.0720
4,1	*50*	*50*	0.0674	0.0665	0.0586	0.0500	0.0579	0.0585
	40	*60*	**0.0925**	**0.0757**	0.0633	0.0508	0.0597	0.0604
	30	*70*	**0.1382**	**0.0980**	0.0615	0.0638	0.0553	0.0571
8,1	*50*	*50*	**0.0812**	**0.0802**	0.0629	0.0500	0.0599	0.0654
	40	*60*	**0.1226**	**0.0900**	0.0601	0.0500	0.0655	0.0615
	30	*70*	**0.1806**	**0.1052**	0.0640	0.0678	0.0556	0.0578

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 6.

Empirical Power with Two Heavily Positively Skewed Distributions, N = 100.

Variance Ratio	n_1	n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	50	50	0.4846	0.4836	0.8370	0.8228	0.8328	0.9682
	40	60	0.4856	0.4956	0.7902	0.7752	0.7660	0.9336
	30	70	0.4360	0.4758	0.7124	0.7268	0.6688	0.8612
1,4	50	50	0.1554	0.1506	0.4758	0.4906	0.5188	0.8357
	40	60	0.1044	0.1947	0.5114	0.5232	0.5461	0.8199
	30	70	0.0480	0.2333	0.5130	0.5218	0.5341	0.7761
1,8	50	50	0.0742	0.0710	0.2378	0.2658	0.2769	0.5231
	40	60	0.0320	0.0896	0.2723	0.3200	0.3113	0.5578
	30	70	0.0090	0.1093	0.2927	0.3420	0.3257	0.5419
4,1	50	50	0.3287	0.3271	0.5977	0.4664	0.5593	0.5806
	40	60	0.3568	0.3177	0.5271	0.4262	0.4764	0.4929
	30	70	0.3888	0.3158	0.4425	0.3620	0.3806	0.3900
8,1	50	50	0.2605	0.2580	0.4285	0.3058	0.3835	0.3760
	40	60	0.3141	0.2605	0.3765	0.2918	0.3223	0.3105
	30	70	0.3566	0.2554	0.3066	0.2518	0.2496	0.2423

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Differential Distribution Shapes

Table 7. *Empirical Type I Error with One Normal, One Moderately Positively Skewed Distribution, N = 60.*

Variance Ratio	$n1,n2$	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*30 30*	0.0512	0.0508	0.0487	0.0484	0.0500	0.0495
	25 35	0.0474	0.0504	0.0487	0.0570	0.0504	0.0515
	20 40	0.0425	0.0497	0.0531	0.0616	0.0552	0.0539
1,4	*30 30*	0.0598	0.0579	0.0503	0.0468	0.0505	0.0563
	25 35	0.0361	0.0555	0.0487	0.0430	0.0491	0.0592
	20 40	**0.0217**	0.0547	0.0507	0.0502	0.0506	0.0607
1,8	*30 30*	0.0652	0.0623	0.0529	0.0554	0.0499	0.0665
	25 35	0.0362	0.0573	0.0509	0.0466	0.0510	0.0678
	20 40	**0.0173**	0.0583	0.0512	0.0544	0.0507	0.0672
4,1	*30 30*	0.0516	0.0505	0.0521	0.0516	0.0522	0.0574
	25 35	0.0699	0.0485	0.0486	0.0560	0.0481	0.0555
	20 40	**0.0969**	0.0474	0.0515	0.0540	0.0498	0.0542
8,1	*30 30*	0.0517	0.0497	0.0505	0.0486	0.0495	0.0651
	25 35	**0.0850**	0.0484	0.0453	0.0560	0.0515	0.0627
	20 40	**0.1221**	0.0482	0.0523	0.0622	0.0501	0.0594

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 8. *Empirical Power with One Normal, One Moderately Positively Skewed Distribution, N = 60.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	30 30	0.7454	0.7440	0.7514	0.7554	0.7704	0.8140
	25 35	0.7184	0.7320	0.7278	0.7440	0.7378	0.7844
	20 40	0.6704	0.6990	0.6792	0.7060	0.6814	0.7370
1,4	30 30	0.2998	0.2909	0.3429	0.3634	0.3879	0.4622
	25 35	0.2313	0.3245	0.3735	0.3814	0.4515	0.4829
	20 40	0.1551	0.3518	0.3942	0.4020	0.4308	0.4866
1,8	30 30	0.1546	0.1447	0.1908	0.1998	0.2220	0.2895
	25 35	0.0947	0.1693	0.2149	0.2276	0.2513	0.3159
	20 40	0.0441	0.1873	0.2321	0.2504	0.2697	0.3298
4,1	30 30	0.3917	0.3869	0.3694	0.3754	0.3721	0.4054
	25 35	0.4218	0.3574	0.3321	0.3468	0.3251	0.3638
	20 40	0.4282	0.3098	0.2848	0.3084	0.2743	0.3098
8,1	30 30	0.2507	0.2446	0.2292	0.2374	0.2233	0.2556
	25 35	0.2980	0.2197	0.2052	0.2062	0.1930	0.2253
	20 40	0.3326	0.1855	0.1716	0.1956	0.1592	0.1866

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 9. *Empirical Type I Error with One Normal, One Moderately Positively Skewed Distribution, N = 100.*

Variance Ratio	*n*1,*n*2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*50 50*	0.0529	0.0526	0.0497	0.0542	0.0527	0.0535
	40 60	0.0434	0.0494	0.0498	0.0516	0.0517	0.0510
	30 70	0.0398	0.0518	0.0524	0.0560	0.0525	0.0526
1,4	*50 50*	0.0577	0.0563	0.0514	0.0504	0.0522	0.0585
	40 60	0.0319	0.0544	0.0507	0.0494	0.0519	0.0595
	30 70	**0.0142**	0.0522	0.0515	0.0498	0.0533	0.0607
1,8	*50 50*	0.0603	0.0586	0.0505	0.0532	0.0513	0.0656
	40 60	0.0272	0.0543	0.0530	0.0502	0.0517	0.0695
	30 70	**0.0105**	0.0517	0.0492	0.0516	0.0520	0.0706
4,1	*50 50*	0.0531	0.0527	0.0511	0.0544	0.0524	0.0596
	40 60	**0.0791**	0.0527	0.0523	0.0574	0.0531	0.0569
	30 70	**0.1058**	0.0495	0.0500	0.0526	0.0489	0.0545
8,1	*50 50*	0.0531	0.0512	0.0524	0.0496	0.0502	0.0676
	40 60	**0.0880**	0.0466	0.0506	0.0548	0.0487	0.0602
	30 70	**0.1436**	0.0494	0.0528	0.0578	0.0500	0.0605

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 10. *Empirical Power with One Normal, One Moderately Positively Skewed Distribution, N = 100.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	50 50	0.9368	0.9368	0.9382	0.9324	0.9422	0.9608
	40 60	0.9168	0.9234	0.9140	0.9204	0.9202	0.9438
	30 70	0.8748	0.8908	0.8706	0.8724	0.8726	0.9028
1,4	50 50	0.5028	0.4977	0.5572	0.5830	0.6008	0.6700
	40 60	0.4150	0.5501	0.6002	0.6060	0.6340	0.7003
	30 70	0.2785	0.5729	0.6054	0.6226	0.6306	0.6953
1,8	50 50	0.2727	0.2654	0.3255	0.3406	0.3733	0.4305
	40 60	0.1813	0.3165	0.3755	0.3972	0.4157	0.4802
	30 70	0.0845	0.3566	0.4107	0.4182	0.4477	0.5126
4,1	50 50	0.5941	0.5923	0.5675	0.5696	0.5693	0.6039
	40 60	0.6160	0.5355	0.5033	0.5088	0.5006	0.5315
	30 70	0.5973	0.4511	0.4098	0.4278	0.3951	0.4352
8,1	50 50	0.3899	0.3858	0.3569	0.3498	0.3541	0.3846
	40 60	0.4385	0.3315	0.3010	0.3066	0.2927	0.3166
	30 70	0.4615	0.2647	0.2378	0.2574	0.2275	0.2503

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 11. *Empirical Type I Error with One Normal, One Moderately Negatively Skewed Distribution, N = 60.*

Variance Ratio	$n1,n2$	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*30 30*	0.0494	0.0494	0.0450	0.0430	0.0458	0.0446
	25 35	0.0428	0.0472	0.0446	0.0478	0.0444	0.0514
	20 40	0.0496	0.0524	0.0548	0.0550	0.0542	0.0546
1,4	*30 30*	0.0662	0.0642	0.0578	0.0568	0.0564	0.0644
	25 35	0.0390	0.0566	0.0506	0.0510	0.0526	0.0588
	20 40	**0.0206**	0.0496	0.0494	0.0496	0.0518	0.0598
1,8	*30 30*	0.0668	0.0626	0.0524	0.0566	0.0518	0.0652
	25 35	0.0396	0.0622	0.0488	0.0526	0.0460	0.0630
	20 40	**0.0168**	0.0532	0.0502	0.0554	0.0500	0.0682
4,1	*30 30*	0.0510	0.0490	0.0530	0.0480	0.0518	0.0584
	25 35	0.0654	0.0446	0.0512	0.0576	0.0518	0.0576
	20 40	**0.0992**	0.0530	0.0570	0.0572	0.0520	0.0572
8,1	*30 30*	0.0472	0.0462	0.0504	0.0530	0.0484	0.0600
	25 35	**0.0796**	0.0470	0.0526	0.0566	0.0506	0.0620
	20 40	**0.1260**	0.0492	0.0566	0.0648	0.0546	0.0606

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 12. *Empirical Power with One Normal, One Moderately Negatively Skewed Distribution, N = 60.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	30 30	0.7052	0.7038	0.6934	0.6846	0.6694	0.7354
	25 35	0.6752	0.6964	0.6798	0.6802	0.6596	0.7246
	20 40	0.6522	0.6896	0.6486	0.6608	0.6356	0.7112
1,4	30 30	0.3950	0.3886	0.3640	0.3594	0.3250	0.3832
	25 35	0.3372	0.4002	0.3960	0.3750	0.3644	0.4220
	20 40	0.2760	0.4124	0.4004	0.3878	0.3712	0.4418
1,8	30 30	0.2766	0.2706	0.2502	0.2312	0.2124	0.2552
	25 35	0.2182	0.2906	0.2642	0.2566	0.2296	0.2886
	20 40	0.1594	0.2786	0.2786	0.2702	0.2480	0.3164
4,1	30 30	0.4080	0.4036	0.3636	0.3420	0.3522	0.3940
	25 35	0.4346	0.3642	0.3198	0.3186	0.3074	0.3534
	20 40	0.4396	0.3126	0.2758	0.3040	0.2638	0.3062
8,1	30 30	0.2546	0.2498	0.2218	0.2208	0.2170	0.2604
	25 35	0.3020	0.2216	0.1908	0.1944	0.1884	0.2280
	20 40	0.3400	0.1860	0.1644	0.1874	0.1582	0.1910

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 13. *Empirical Type I Error with One Normal, One Moderately Negatively Skewed Distribution, N = 100.*

Variance Ratio	$n1,n2$	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*50 50*	0.0510	0.0508	0.0458	0.0504	0.0492	0.0468
	40 60	0.0458	0.0518	0.0468	0.0594	0.0472	0.0464
	30 70	0.0402	0.0508	0.0498	0.0548	0.0522	0.0526
1,4	*50 50*	0.0574	0.0564	0.0496	0.0500	0.0496	0.0574
	40 60	0.0348	0.0568	0.0506	0.0548	0.0512	0.0590
	30 70	**0.0148**	0.0514	0.0482	0.0536	0.0518	0.0592
1,8	*50 50*	0.0594	0.0516	0.0516	0.0458	0.0506	0.0666
	40 60	0.0318	0.0516	0.0516	0.0524	0.0524	0.0752
	30 70	**0.0104**	0.0554	0.0554	0.0500	0.0546	0.0732
4,1	*50 50*	0.0508	0.0502	0.0488	0.0470	0.0492	0.0570
	40 60	**0.0774**	0.0512	0.0506	0.0524	0.0470	0.0578
	30 70	**0.1120**	0.0554	0.0538	0.0612	0.0530	0.0574
8,1	*50 50*	0.0530	0.0508	0.0500	0.0562	0.0516	0.0670
	40 60	**0.0920**	0.0558	0.0538	0.0530	0.0502	0.0650
	30 70	**0.1434**	0.0496	0.0480	0.0606	0.0466	0.0578

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 14. *Empirical Power with One Normal, One Moderately Negatively Skewed Distribution, N = 100.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	50 50	0.8882	0.8878	0.8958	0.8870	0.8882	0.9174
	40 60	0.8742	0.8938	0.8938	0.8844	0.8880	0.9152
	30 70	0.8262	0.8676	0.8482	0.8526	0.8404	0.8876
1,4	50 50	0.5210	0.5174	0.5338	0.5214	0.4958	0.5566
	40 60	0.4620	0.5634	0.5706	0.5746	0.5456	0.6118
	30 70	0.3648	0.5778	0.5844	0.5786	0.5664	0.6436
1,8	50 50	0.3576	0.3516	0.3516	0.3444	0.3132	0.3764
	40 60	0.2848	0.3920	0.3920	0.3776	0.3588	0.4318
	30 70	0.1894	0.4076	0.4076	0.3912	0.3768	0.4654
4,1	50 50	0.6044	0.6020	0.5578	0.5402	0.5652	0.5934
	40 60	0.6100	0.5414	0.4852	0.4978	0.4812	0.5194
	30 70	0.5976	0.4452	0.3940	0.4308	0.3884	0.4266
8,1	50 50	0.3892	0.3850	0.3398	0.3456	0.3408	0.3820
	40 60	0.4400	0.3328	0.2894	0.2928	0.2848	0.3264
	30 70	0.4752	0.2742	0.2370	0.2534	0.2336	0.2658

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 15. *Empirical Type I Error with One Normal, One Heavily Positively Skewed Distribution, N = 60.*

Variance Ratio	*n*1,*n*2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*30 30*	**0.0822**	**0.0814**	0.0552	0.0442	0.0533	0.0514
	25 35	0.0678	0.0740	0.0524	0.0464	0.0554	0.0520
	20 40	0.0474	0.0614	0.0508	0.0540	0.0500	0.0470
1,4	*30 30*	**0.1062**	**0.1036**	0.0633	0.0488	0.0586	0.0590
	25 35	0.0745	**0.0980**	0.0650	0.0542	0.0593	0.0587
	20 40	0.0500	**0.0915**	0.0634	0.0508	0.0606	0.0621
1,8	*30 30*	**0.1129**	**0.1089**	0.0695	0.0542	0.0600	0.0646
	25 35	**0.0780**	**0.1049**	0.0674	0.0580	0.0606	0.0683
	20 40	0.0480	**0.0985**	0.0672	0.0530	0.0610	0.0708
4,1	*30 30*	0.0557	0.0547	0.0525	0.0458	0.0527	0.0591
	25 35	0.0626	0.0538	0.0540	0.0600	0.0524	0.0586
	20 40	0.0690	0.0510	0.0547	0.0646	0.0532	0.0570
8,1	*30 30*	0.0550	0.0530	0.0520	0.0496	0.0499	0.0653
	25 35	0.0687	0.0489	0.0513	0.0596	0.0483	0.0601
	20 40	**0.0926**	0.0515	0.0538	0.0660	0.0532	0.0607

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 16. *Empirical Power with One Normal, One Heavily Positively Skewed Distribution, N = 60.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	30 30	0.4074	0.4024	0.7358	0.7436	0.7780	0.8124
	25 35	0.3372	0.4300	0.7076	0.7218	0.7404	0.7846
	20 40	0.2512	0.4362	0.6522	0.6854	0.6768	0.7222
1,4	30 30	0.0766	0.0710	0.2573	0.3110	0.2947	0.5105
	25 35	0.0378	0.0858	0.2805	0.3442	0.3180	0.5304
	20 40	0.0143	0.1006	0.2801	0.3852	0.3184	0.5147
1,8	30 30	0.0546	0.0493	0.1281	0.1614	0.1506	0.3280
	25 35	0.0217	0.0497	0.1422	0.1846	0.1693	0.3555
	20 40	0.0217	0.0497	0.1422	0.2130	0.1693	0.3555
4,1	30 30	0.2519	0.2500	0.3266	0.3706	0.3242	0.3881
	25 35	0.2510	0.2480	0.2913	0.3474	0.2854	0.3457
	20 40	0.2358	0.2312	0.2460	0.3176	0.2324	0.2893
8,1	30 30	0.1805	0.1775	0.2026	0.2296	0.1957	0.2461
	25 35	0.2049	0.1721	0.1819	0.2144	0.1718	0.2177
	20 40	0.2186	0.1551	0.1513	0.1950	0.1396	0.1794

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 17. *Empirical Type I Error with One Normal, One Heavily Positively Skewed Distribution, N = 100.*

Variance Ratio	*n*1,*n*2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*50 50*	**0.0770**	**0.0764**	0.0596	0.0482	0.0570	0.0506
	40 60	0.0555	0.0676	0.0610	0.0480	0.0638	0.0686
	30 70	0.0334	0.0602	0.0548	0.0486	0.0532	0.0506
1,4	*50 50*	**0.0917**	**0.0915**	0.0694	0.0494	0.0639	0.0592
	40 60	0.0541	**0.0828**	0.0708	0.0490	0.0674	0.0612
	30 70	0.0274	**0.0779**	0.0667	0.0528	0.0652	0.0603
1,8	*50 50*	**0.0964**	**0.0944**	0.0747	0.0482	0.0643	0.0645
	40 60	0.0548	**0.0887**	0.0739	0.0472	0.0688	0.0694
	30 70	0.0254	**0.0825**	0.0714	0.0458	0.0673	0.0686
4,1	*50 50*	0.0564	0.0560	0.0543	0.0510	0.0546	0.0602
	40 60	0.0589	0.0511	0.0541	0.0568	0.0542	0.0574
	30 70	0.0648	0.0527	0.0561	0.0586	0.0545	0.0603
8,1	*50 50*	0.0521	0.0512	0.0523	0.0472	0.0532	0.0650
	40 60	0.0725	0.0506	0.0522	0.0524	0.0507	0.0634
	30 70	**0.0963**	0.0503	0.0524	0.0620	0.0492	0.0583

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 18. *Empirical Power with One Normal, One Heavily Positively Skewed Distribution, N = 100.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	50 50	0.6712	0.6678	0.9258	0.9312	0.9424	0.9620
	40 60	0.5990	0.7078	0.9168	0.9152	0.9268	0.9458
	30 70	0.4464	0.6812	0.8532	0.8734	0.8622	0.8926
1,4	50 50	0.1472	0.1408	0.4143	0.5240	0.4584	0.7331
	40 60	0.0673	0.1792	0.4623	0.5724	0.4993	0.7533
	30 70	0.0187	0.2137	0.4678	0.5892	0.4937	0.7267
1,8	50 50	0.0737	0.0699	0.2051	0.2924	0.2418	0.4946
	40 60	0.0255	0.0808	0.2394	0.3376	0.2734	0.5323
	30 70	0.0039	0.0971	0.2594	0.3730	0.2910	0.5485
4,1	50 50	0.4213	,41995	0.5011	0.5774	0.5022	0.5818
	40 60	0.4015	0.4017	0.4428	0.5164	0.4377	0.5097
	30 70	0.3641	0.3623	0.3562	0.4156	0.3468	0.4154
8,1	50 50	0.2906	0.2890	0.3110	0.3652	0.3063	0.3688
	40 60	0.3137	0.2722	0.2649	0.3100	0.2581	0.3113
	30 70	0.2874	0.2858	0.3083	0.2606	0.3063	0.3673

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 19. *Empirical Type I Error with One Normal, One Heavily Negatively Skewed Distribution, N = 100.*

Variance Ratio	*n*1,*n*2	*t*	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*50 50*	**0.0796**	**0.0792**	0.0490	0.0516	0.0546	0.0540
	40 60	0.0480	0.0592	0.0484	0.0496	0.0512	0.0496
	30 70	0.0384	0.0628	0.0552	0.0556	0.0572	0.0592
1,4	*50 50*	0.1000	**0.0970**	0.0514	0.0554	0.0552	0.0590
	40 60	0.0548	**0.0826**	0.0520	0.0482	0.0550	0.0596
	30 70	0.0306	**0.0758**	0.0482	0.0524	0.0506	0.0596
1,8	*50 50*	**0.1058**	**0.1022**	0.0536	0.0586	0.0492	0.0624
	40 60	0.0530	**0.0840**	0.0548	0.0538	0.0542	0.0748
	30 70	0.0262	**0.0844**	0.0534	0.0480	0.0558	0.0736
4,1	*50 50*	0.0534	0.0526	0.0476	0.0468	0.0536	0.0536
	40 60	0.0610	0.0510	0.0464	0.0528	0.0468	0.0544
	30 70	0.0642	0.0486	0.0490	0.0626	0.0458	0.0554
8,1	*50 50*	0.0484	0.0480	0.0510	0.0508	0.0508	0.0616
	40 60	**0.0770**	0.0532	0.0538	0.0596	0.0540	0.0662
	30 70	**0.0906**	0.0458	0.0502	0.0636	0.0468	0.0552

Notes: *t*=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 20. *Empirical Power with One Normal, One Heavily Negatively Skewed Distribution, N = 100.*

Variance Ratio	n_1 n_2	*t*	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	50 50	0.6474	0.6430	0.8282	0.8214	0.7972	0.8694
	40 60	0.6012	0.6636	0.8384	0.8382	0.8198	0.8850
	30 70	0.5272	0.6616	0.8046	0.8382	0.7936	0.8556
1,4	50 50	0.3936	0.3896	0.4882	0.4700	0.4152	0.5470
	40 60	0.3200	0.3948	0.5238	0.4194	0.4698	0.6004
	30 70	0.2222	0.3976	0.5338	0.5140	0.4926	0.6312
1,8	50 50	0.2846	0.2814	0.3342	0.3098	0.2682	0.3636
	40 60	0.2168	0.2922	0.3552	0.3386	0.2934	0.4092
	30 70	0.1378	0.2866	0.3726	0.3622	0.3192	0.4508
4,1	50 50	0.4906	0.4904	0.5170	0.5042	0.5234	0.5420
	40 60	0.4598	0.4438	0.4574	0.4746	0.4520	0.4786
	30 70	0.4266	0.4022	0.3900	0.4056	0.3812	0.4208
8,1	50 50	0.3584	0.3564	0.3388	0.3334	0.3430	0.3586
	40 60	0.3658	0.3072	0.2784	0.2946	0.2812	0.3116
	30 70	0.3562	0.2478	0.2320	0.2504	0.2274	0.2544

Note: *t*=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 21. *Empirical Type I Error with One Moderately Positively, One Heavily Positively Skewed Distribution, N = 100.*

Variance Ratio	$n1,n2$	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*50 50*	0.0671	0.0663	0.0537	0.0496	0.0608	0.0499
	40 60	0.0503	0.0560	0.0505	0.0536	0.0592	0.0500
	30 70	0.0354	0.0487	0.0523	0.0546	0.0589	0.0523
1,4	*50 50*	**0.0933**	**0.0918**	0.0711	0.0436	0.0679	0.0568
	40 60	0.0549	**0.0817**	0.0683	0.0484	0.0584	0.0600
	30 70	0.0293	0.0689	0.0638	0.0476	0.0681	0.0584
1,8	*50 50*	**0.0968**	**0.0949**	0.0749	0.0490	0.0666	0.0652
	40 60	0.0557	**0.0857**	0.0733	0.0570	0.0690	0.0699
	30 70	0.0299	**0.0845**	**0.0765**	0.0512	0.0740	0.0728
4,1	*50 50*	0.0499	0.0492	0.0498	0.0480	0.0571	0.0600
	40 60	0.0579	0.0479	0.0494	0.0560	0.0544	0.0555
	30 70	0.0734	0.0510	0.0518	0.0578	0.0543	0.0578
8,1	*50 50*	0.0505	0.0495	0.0505	0.0488	0.0534	0.0654
	40 60	**0.0766**	0.0523	0.0531	0.0588	0.0529	0.0632
	30 70	**0.1078**	0.0566	0.0521	0.0610	0.0496	0.0599

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 22. *Empirical Power with One Moderately Positively, One Heavily Positively Skewed Distribution, N = 100.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	50 50	0.6242	0.6222	0.8950	0.8966	0.9014	0.9668
	40 60	0.5554	0.6330	0.8538	0.8672	0.8518	0.9386
	30 70	0.4470	0.6202	0.7974	0.8068	0.7758	0.8820
1,4	50 50	0.1451	0.1391	0.4086	0.5156	0.4500	0.8051
	40 60	0.0770	0.1828	0.4507	0.5534	0.4829	0.8099
	30 70	0.0249	0.2140	0.4534	0.5462	0.4761	0.7807
1,8	50 50	0.0703	0.0659	0.2053	0.2680	0.2446	0.5235
	40 60	0.0236	0.0805	0.2272	0.3232	0.2641	0.5666
	30 70	0.0044	0.0941	0.2481	0.3670	0.2801	0.5750
4,1	50 50	0.3882	0.3871	0.4874	0.5398	0.4533	0.5788
	40 60	0.3890	0.3715	0.4253	0.4560	0.3825	0.4980
	30 70	0.3694	0.3441	0.3579	0.4050	0.3038	0.3990
8,1	50 50	0.2843	0.2825	0.3186	0.3376	0.2801	0.3701
	40 60	0.3193	0.2741	0.2766	0.2982	0.2331	0.3127
	30 70	0.3369	0.2501	0.2314	0.2544	0.1853	0.2419

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 23. *Empirical Type I Error with One Moderately Negatively, One Heavily Negatively Skewed Distribution, N = 60.*

Variance Ratio	n1,n2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*30 30*	0.0596	0.0668	0.0434	0.0428	0.0550	0.0480
	25 35	0.0682	0.0524	0.0446	0.0518	0.0540	0.0488
	20 40	0.0530	0.0446	0.0494	0.0562	0.0598	0.0524
1,4	*30 30*	**0.0994**	**0.0972**	0.0530	0.0456	0.0572	0.0582
	25 35	**0.0702**	**0.0896**	0.0514	0.0548	0.0564	0.0564
	20 40	0.0440	0.0738	0.0454	0.0554	0.0534	0.0552
1,8	*30 30*	**0.1124**	**0.1092**	0.0596	0.0598	0.0474	0.0636
	25 35	**0.0792**	**0.1042**	0.0578	0.0556	0.0510	0.0696
	20 40	0.0464	**0.0926**	0.0588	0.0530	0.0564	0.0712
4,1	*30 30*	0.0454	0.0448	0.0468	0.0522	0.0556	0.0560
	25 35	0.0574	0.0452	0.0510	0.0496	0.0560	0.0570
	20 40	0.0660	0.0490	0.0490	0.0620	0.0538	0.0554
8,1	*30 30*	0.0524	0.0504	0.0536	0.0528	0.0530	0.0644
	25 35	0.0728	0.0502	0.0574	0.0600	0.0562	0.0650
	20 40	**0.1024**	0.0584	0.0614	0.0702	0.0522	0.0630

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 24. *Empirical Power with One Moderately Negatively, One Heavily Negatively Skewed Distribution, N = 60.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	30 30	0.5154	0.5140	0.6464	0.6416	0.6236	0.7904
	25 35	0.4826	0.5144	0.6476	0.6310	0.6492	0.8216
	20 40	0.4306	0.4916	0.6250	0.6316	0.6402	0.8262
1,4	30 30	0.3194	0.3148	0.3556	0.3376	0.3014	0.3858
	25 35	0.2742	0.3194	0.3670	0.3376	0.3264	0.4332
	20 40	0.2144	0.3178	0.3640	0.3376	0.3356	0.4614
1,8	30 30	0.2724	0.2672	0.2560	0.2384	0.2026	0.2530
	25 35	0.2144	0.2602	0.2604	0.2420	0.2130	0.2746
	20 40	0.1570	0.2614	0.2748	0.2578	0.2332	0.3136
4,1	30 30	0.2888	0.2856	0.3154	0.3176	0.3570	0.4604
	25 35	0.3016	0.2550	0.2792	0.2914	0.3288	0.4424
	20 40	0.2902	0.1956	0.2252	0.2542	0.2536	0.3888
8,1	30 30	0.0538	0.0500	0.1394	0.1874	0.1456	0.3372
	25 35	0.1032	0.0520	0.1102	0.1802	0.1016	0.2872
	20 40	0.1770	0.0530	0.0812	0.1534	0.0640	0.2356

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 25. *Empirical Type I Error with One Moderately Negatively, One Heavily Negatively Skewed Distribution, N = 100.*

Variance Ratio	*n*1,*n*2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*50 50*	0.0404	0.0620	0.0498	0.0426	0.0600	0.0524
	40 60	0.0624	0.0520	0.0514	0.0490	0.0628	0.0534
	30 70	0.0478	0.0438	0.0512	0.0572	0.0580	0.0534
1,4	*50 50*	**0.0894**	**0.0884**	0.0536	0.0510	0.0578	0.0570
	40 60	0.0536	**0.0792**	0.0504	0.0516	0.0572	0.0638
	30 70	0.0302	0.0700	0.0480	0.0452	0.0582	0.0608
1,8	*50 50*	**0.0972**	**0.0952**	0.0532	0.0542	0.0534	0.0674
	40 60	0.0512	**0.0802**	0.0580	0.0522	0.0612	0.0732
	30 70	0.0276	**0.0794**	0.0492	0.0526	0.0564	0.0674
4,1	*50 50*	0.0472	0.0464	0.0524	0.0486	0.0600	0.0636
	40 60	0.0606	0.0472	0.0496	0.0520	0.0554	0.0606
	30 70	**0.0758**	0.0562	0.0524	0.0572	0.0552	0.0620
8,1	*50 50*	0.0510	0.0506	0.0558	0.0502	0.0580	0.0678
	40 60	**0.0762**	0.0554	0.0572	0.0608	0.0564	0.0642
	30 70	**0.1106**	0.0540	0.0512	0.0630	0.0502	0.0612

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 26. *Empirical Power with One Moderately Negatively, One Heavily Negatively Skewed Distribution, N = 100.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	50 50	0.6230	0.6208	0.8436	0.8340	0.8246	0.9486
	40 60	0.5888	0.6448	0.8576	0.8516	0.8528	0.9620
	30 70	0.5150	0.6332	0.8274	0.8490	0.8376	0.9598
1,4	50 50	0.3850	0.3814	0.4834	0.4774	0.4166	0.5708
	40 60	0.3064	0.3872	0.5286	0.5056	0.4774	0.6370
	30 70	0.2322	0.3894	0.5280	0.5290	0.5028	0.6656
1,8	50 50	0.2804	0.2778	0.3240	0.3162	0.2542	0.3652
	40 60	0.2246	0.2904	0.3680	0.3370	0.3058	0.4300
	30 70	0.1506	0.2906	0.3730	0.3714	0.3294	0.4694
4,1	50 50	0.4078	0.4064	0.5198	0.5258	0.5574	0.6696
	40 60	0.4092	0.3674	0.4524	0.4734	0.5048	0.6292
	30 70	0.3836	0.2844	0.3522	0.3970	0.4066	0.5462
8,1	50 50	0.0710	0.0664	0.2522	0.3324	0.3310	0.4986
	40 60	0.1456	0.0556	0.2064	0.2814	0.2582	0.4110
	30 70	0.2408	0.0476	0.1560	0.2352	0.1692	0.3226

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Differential and Opposed Skewness

Table 27. *Empirical Type I Error with One Moderately Positively, One Moderately Negatively Skewed Distribution, N = 60.*

Variance Ratio	$n1,n2$	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
	30 30	0.0592	0.0590	0.0558	0.0474	0.0540	0.0560
1,1	*25 35*	0.0564	0.0550	0.0460	0.0520	0.0456	0.0486
	20 40	0.0506	0.0570	0.0480	0.0570	0.0488	0.0464
	30 30	0.0638	0.0618	0.0510	0.0562	0.0492	0.0594
1,4	*25 35*	0.0430	0.0580	0.0560	0.0518	0.0544	0.0676
	20 40	**0.0244**	0.0582	0.0522	0.0512	0.0506	0.0562
	30 30	0.0668	0.0648	0.0524	0.0532	0.0482	0.0648
1,8	*25 35*	0.0392	0.0626	0.0534	0.0480	0.0516	0.0664
	20 40	**0.0216**	0.0618	0.0526	0.0534	0.0508	0.0694
	30 30	0.0646	0.0628	0.0532	0.0562	0.0518	0.0582
4,1	*25 35*	**0.0926**	0.0660	0.0564	0.0582	0.0534	0.0602
	20 40	**0.1204**	0.0686	0.0608	0.0636	0.0578	0.0638
	30 30	0.0674	0.0640	0.0544	0.0500	0.0516	0.0696
8,1	*25 35*	**0.1094**	**0.0718**	0.0534	0.0628	0.0506	0.0672
	20 40	**0.1476**	0.0680	0.0554	0.0732	0.0468	0.0628

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 28. *Empirical Power with One Moderately Positively, One Moderately Negatively Skewed Distribution, N = 60.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
	30 30	0.6380	0.6370	0.6554	0.6400	0.6174	0.6290
1,1	25 35	0.6348	0.6368	0.6488	0.6314	0.5968	0.6254
	20 40	0.6228	0.6124	0.6170	0.6088	0.5622	0.5866
	30 30	0.3642	0.3608	0.3546	0.3380	0.3082	0.3364
1,4	25 35	0.3212	0.3764	0.3692	0.3746	0.3274	0.3630
	20 40	0.2678	0.3938	0.3742	0.3590	0.3388	0.3730
	30 30	0.2634	0.2570	0.2352	0.2258	0.1912	0.2364
1,8	25 35	0.2106	0.2736	0.2548	0.2316	0.2248	0.2668
	20 40	0.1536	0.2730	0.2702	0.2512	0.2344	0.2772
	30 30	0.3836	0.3804	0.3670	0.3454	0.3164	0.3542
4,1	25 35	0.3962	0.3376	0.3232	0.3228	0.2802	0.3058
	20 40	0.4216	0.3222	0.2904	0.2922	0.2452	0.2760
	30 30	0.2658	0.2606	0.2386	0.2274	0.2024	0.2390
8,1	25 35	0.3228	0.2610	0.2302	0.2198	0.1908	0.2236
	20 40	0.3536	0.2272	0.1920	0.2022	0.1534	0.1786

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 29. *Empirical Type I Error with One Moderately Positively, One Moderately Negatively Skewed Distribution, N = 100.*

Variance Ratio	$n1,n2$	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*50 50*	0.0540	0.0538	0.0462	0.0502	0.0472	0.0486
	40 60	0.0536	0.0538	0.0508	0.0532	0.0538	0.0536
	30 70	0.0530	0.0586	0.0472	0.0500	0.0474	0.0488
1,4	*50 50*	0.0542	0.0530	0.0516	0.0496	0.0512	0.0564
	40 60	0.0326	0.0564	0.0538	0.0522	0.0534	0.0612
	30 70	**0.0146**	0.0512	0.0430	0.0466	0.0468	0.0662
1,8	*50 50*	0.0588	0.0564	0.0470	0.0480	0.0474	0.0606
	40 60	**0.0298**	0.0602	0.0560	0.0480	0.0588	0.0740
	30 70	**0.0126**	0.0568	0.0500	0.0526	0.0484	0.0700
4,1	*50 50*	0.0568	0.0560	0.0504	0.0460	0.0500	0.0564
	40 60	**0.0912**	0.0616	0.0510	0.0502	0.0508	0.0578
	30 70	**0.1318**	0.0616	0.0544	0.0556	0.0490	0.0560
8,1	*50 50*	0.0586	0.0562	0.0542	0.0516	0.0552	0.0660
	40 60	**0.1022**	0.0598	0.0509	0.0528	0.0494	0.0618
	30 70	**0.1674**	0.0620	0.0602	0.0630	0.0550	0.0674

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 30. *Empirical Power with One Moderately Positively, One Moderately Negatively Skewed Distribution, N = 100.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	50 50	0.8116	0.8112	0.8596	0.8580	0.8390	0.8392
	40 60	0.8128	0.8086	0.8428	0.8448	0.8190	0.8268
	30 70	0.7736	0.7574	0.7870	0.7908	0.7534	0.7696
1,4	50 50	0.5078	0.5040	0.5310	0.4972	0.4920	0.5190
	40 60	0.4478	0.5260	0.5542	0.5480	0.5208	0.5506
	30 70	0.3536	0.5310	0.5492	0.5508	0.5216	0.5484
1,8	50 50	0.3542	0.3512	0.3420	0.3184	0.3048	0.3522
	40 60	0.2798	0.3672	0.3646	0.3608	0.3364	0.3858
	30 70	0.1790	0.3882	0.3912	0.3856	0.3626	0.4134
4,1	50 50	0.5114	0.5080	0.5196	0.5126	0.4830	0.5128
	40 60	0.5270	0.4544	0.4570	0.4638	0.4122	0.4504
	30 70	0.5502	0.4126	0.3962	0.4030	0.3436	0.3774
8,1	50 50	0.3570	0.3522	0.3458	0.3438	0.3102	0.3510
	40 60	0.4010	0.3160	0.3038	0.3048	0.2614	0.2954
	30 70	0.4334	0.2702	0.2492	0.2614	0.2050	0.2298

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 31. *Empirical Type I Error with One Moderately Positively, One Heavily Negatively Skewed Distribution, N = 60.*

Variance Ratio	*n*1,*n*2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*30 30*	**0.0788**	**0.0780**	0.0508	0.0458	0.0504	0.0552
	25 35	0.0692	**0.0772**	0.0484	0.0526	0.0484	0.0542
	20 40	0.0596	**0.0800**	0.0556	0.0568	0.0554	0.0594
1,4	*30 30*	**0.1056**	**0.1042**	0.0534	0.0510	0.0464	0.0560
	25 35	0.0740	**0.0938**	0.0532	0.0488	0.0470	0.0566
	20 40	0.0478	**0.0804**	0.0510	0.0520	0.0504	0.0602
1,8	*30 30*	**0.1066**	**0.1044**	0.0564	0.0514	0.0398	0.0682
	25 35	0.0740	**0.1034**	0.0590	0.0570	0.0478	0.0682
	20 40	0.0472	**0.0942**	0.0544	0.0584	0.0462	0.0708
4,1	*30 30*	0.0696	0.0684	0.0490	0.0534	0.0466	0.0580
	25 35	**0.0798**	0.0664	0.0510	0.0558	0.0488	0.0598
	20 40	**0.0922**	0.0676	0.0556	0.0596	0.0522	0.0586
8,1	*30 30*	0.0700	0.0690	0.0530	0.0556	0.0496	0.0708
	25 35	**0.0908**	0.0646	0.0506	0.0508	0.0460	0.0614
	20 40	**0.1072**	0.0606	0.0508	0.0654	0.0440	0.0634

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 32. *Empirical Power with One Moderately Positively, One Heavily Negatively Skewed Distribution, N = 60.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	30 30	0.4958	0.4934	0.5946	0.5836	0.5372	0.5290
	25 35	0.4618	0.4882	0.6038	0.5926	0.5360	0.5432
	20 40	0.4124	0.4710	0.5494	0.5616	0.4798	0.5036
1,4	30 30	0.3306	0.3262	0.3450	0.3154	0.2772	0.3214
	25 35	0.2714	0.3164	0.3506	0.3294	0.2854	0.3410
	20 40	0.2170	0.3274	0.3426	0.3374	0.2924	0.3398
1,8	30 30	0.2644	0.2584	0.2364	0.2276	0.1754	0.2126
	25 35	0.2088	0.2518	0.2466	0.2264	0.1870	0.2414
	20 40	0.1560	0.2650	0.2604	0.2496	0.2020	0.2710
4,1	30 30	0.3320	0.3298	0.3324	0.3180	0.2868	0.2958
	25 35	0.3310	0.3086	0.3166	0.3100	0.2704	0.2860
	20 40	0.3396	0.2994	0.2802	0.2828	0.2290	0.2510
8,1	30 30	0.2444	0.2414	0.2208	0.2160	0.1860	0.2056
	25 35	0.2702	0.2282	0.2134	0.2094	0.1764	0.1932
	20 40	0.2866	0.2190	0.1880	0.1984	0.1508	0.1684

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 33. *Empirical Type I Error with One Moderately Positively, One Heavily Negatively Skewed Distribution, N = 100.*

Variance Ratio	$n1,n2$	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*50 50*	**0.0764**	**0.0758**	0.0514	0.0474	0.0536	0.0576
	40 60	0.0594	0.0704	0.0490	0.0496	0.0506	0.0534
	30 70	0.0476	0.0752	0.0528	0.0550	0.0542	0.0560
1,4	*50 50*	**0.0930**	**0.0914**	0.0534	0.0534	0.0564	0.0584
	40 60	0.0530	**0.0808**	0.0578	0.0544	0.0566	0.0678
	30 70	0.0288	0.0736	0.0456	0.0458	0.0448	0.0574
1,8	*50 50*	**0.0870**	**0.0864**	0.0568	0.0514	0.0534	0.0684
	40 60	0.0566	**0.0870**	0.0536	0.0476	0.0542	0.0724
	30 70	0.0300	**0.0822**	0.0522	0.0460	0.0482	0.0688
4,1	*50 50*	0.0608	0.0602	0.0520	0.0524	0.0530	0.0652
	40 60	0.0690	0.0572	0.0532	0.0572	0.0516	0.0556
	30 70	**0.0878**	0.0614	0.0532	0.0564	0.0510	0.0582
8,1	*50 50*	0.0622	0.0610	0.0528	0.0542	0.0524	0.0718
	40 60	**0.0940**	0.0658	0.0580	0.0544	0.0566	0.0684
	30 70	**0.1216**	0.0646	0.0554	0.0616	0.0496	0.0622

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 34. *Empirical Power with One Moderately Positively, One Heavily Negatively Skewed Distribution, N = 100.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	50 50	0.6172	0.6156	0.8098	0.8034	0.7702	0.7620
	40 60	0.5774	0.6234	0.7924	0.7870	0.7574	0.7464
	30 70	0.5090	0.6194	0.7628	0.7526	0.7174	0.7090
1,4	50 50	0.3686	0.3636	0.4782	0.4426	0.4036	0.4692
	40 60	0.2888	0.3634	0.5028	0.4816	0.4402	0.4952
	30 70	0.2198	0.3778	0.5174	0.4974	0.4610	0.5126
1,8	50 50	0.2946	0.2900	0.3392	0.2870	0.2704	0.3440
	40 60	0.2194	0.2896	0.3504	0.3320	0.2910	0.3674
	30 70	0.1436	0.3006	0.3658	0.3510	0.3092	0.3894
4,1	50 50	0.4410	0.4404	0.5120	0.4760	0.4698	0.4646
	40 60	0.4340	0.4006	0.4460	0.4524	0.3982	0.4070
	30 70	0.4064	0.3634	0.3878	0.3880	0.3422	0.3478
8,1	50 50	0.3082	0.3048	0.3246	0.3214	0.2898	0.3070
	40 60	0.3458	0.2884	0.2874	0.2822	0.2474	0.2682
	30 70	0.5264	0.6580	0.8376	0.2656	0.8016	0.9934

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 35. *Empirical Type I Error with One Heavily Positively, One Moderately Negatively Skewed Distribution, N = 60.*

Variance Ratio	$n1,n2$	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*30 30*	**0.0842**	**0.0838**	0.0474	0.0466	0.0482	0.0548
	25 35	**0.0994**	**0.0904**	0.0526	0.0488	0.0514	0.0586
	20 40	**0.1078**	**0.0934**	0.0544	0.0566	0.0522	0.0504
1,4	*30 30*	0.0748	0.0734	0.0520	0.0568	0.0496	0.0602
	25 35	0.0590	0.0686	0.0466	0.0546	0.0466	0.0582
	20 40	0.0490	**0.0764**	0.0460	0.0492	0.0462	0.0606
1,8	*30 30*	0.0626	0.0610	0.0558	0.0510	0.0520	0.0704
	25 35	0.0426	0.0618	0.0512	0.0530	0.0490	0.0688
	20 40	0.0358	0.0694	0.0456	0.0452	0.0438	0.0664
4,1	*30 30*	**0.1022**	**0.1000**	0.0536	0.0570	0.0478	0.0632
	25 35	**0.1336**	**0.1034**	0.0556	0.0608	0.0476	0.0622
	20 40	**0.1976**	**0.1280**	0.0658	0.0632	0.0518	0.0594
8,1	*30 30*	**0.1086**	**0.1060**	0.0520	0.0538	0.0426	0.0634
	25 35	**0.1622**	**0.1268**	0.0592	0.0590	0.0404	0.0642
	20 40	**0.2158**	**0.1294**	0.0614	0.0678	0.0542	0.0632

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 36. *Empirical Power with One Heavily Positively, One Moderately Negatively Skewed Distribution, N = 60.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	30 30	0.4994	0.4978	0.6018	0.5904	0.5316	0.5424
	25 35	0.5266	0.4984	0.5822	0.5636	0.4182	0.5256
	20 40	0.5294	0.4790	0.5516	0.5384	0.4770	0.4824
1,4	30 30	0.3340	0.3326	0.3380	0.3308	0.2902	0.3070
	25 35	0.3200	0.3462	0.3484	0.3426	0.3088	0.3130
	20 40	0.2748	0.3488	0.3564	0.3372	0.3054	0.3044
1,8	30 30	0.2538	0.2508	0.2330	0.2132	0.1990	0.2194
	25 35	0.2074	0.2508	0.2364	0.2316	0.2074	0.2238
	20 40	0.1798	0.2758	0.2482	0.2492	0.2196	0.2432
4,1	30 30	0.3380	0.3328	0.3394	0.3188	0.2716	0.3126
	25 35	0.3666	0.3148	0.3146	0.3120	0.2444	0.2816
	20 40	0.4136	0.3152	0.2938	0.2976	0.2182	0.2414
8,1	30 30	0.2706	0.2666	0.2502	0.2160	0.1906	0.2260
	25 35	0.3158	0.2608	0.2224	0.2090	0.1632	0.1866
	20 40	0.3658	0.2618	0.2112	0.2126	0.1544	0.1692

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 37. *Empirical Type I Error with One Heavily Positively, One Moderately Negatively Skewed Distribution, N = 100.*

Variance Ratio	$n1,n2$	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*50 50*	0.0682	0.0682	0.0492	0.0466	0.0522	0.0556
	40 60	**0.0976**	**0.0860**	0.0500	0.0518	0.0522	0.0542
	30 70	**0.1174**	**0.0998**	0.0510	0.0566	0.0526	0.0534
1,4	*50 50*	0.0594	0.0592	0.0546	0.0448	0.0532	0.0628
	40 60	0.0538	0.0678	0.0500	0.0460	0.0532	0.0646
	30 70	0.0364	0.0634	0.0526	0.0510	0.0566	0.0604
1,8	*50 50*	0.0646	0.0636	0.0570	0.0496	0.0574	0.0736
	40 60	0.0370	0.0580	0.0500	0.0534	0.0474	0.0646
	30 70	**0.0246**	0.0672	0.0506	0.0566	0.0502	0.0688
4,1	*50 50*	**0.0876**	**0.0858**	0.0546	0.0564	0.0568	0.0646
	40 60	**0.1318**	**0.0910**	0.0524	0.0578	0.0536	0.0600
	30 70	**0.2124**	**0.1122**	0.0624	0.0624	0.0542	0.0640
8,1	*50 50*	**0.0928**	**0.0906**	0.0528	0.0528	0.0502	0.0624
	40 60	**0.1450**	**0.1002**	0.0536	0.0560	0.0466	0.0618
	30 70	**0.2280**	**0.1052**	0.0568	0.0640	0.0398	0.0564

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 38. *Empirical Power with One Heavily Positively, One Moderately Negatively Skewed Distribution, N = 100.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	50 50	0.6154	0.6140	0.8084	0.7858	0.7080	0.7672
	40 60	0.6340	0.5966	0.7786	0.7650	0.7272	0.7322
	30 70	0.6456	0.5728	0.7146	0.7038	0.6438	0.6588
1,4	50 50	0.4302	0.4298	0.4956	0.4888	0.4610	0.4496
	40 60	0.4056	0.4392	0.5342	0.5154	0.4962	0.4732
	30 70	0.3748	0.4558	0.5290	0.5062	0.4874	0.4666
1,8	50 50	0.3088	0.3074	0.3312	0.3182	0.3016	0.3096
	40 60	0.2676	0.3282	0.3662	0.3528	0.3324	0.3488
	30 70	0.2082	0.3468	0.3728	0.3678	0.3494	0.3522
4,1	50 50	0.3692	0.3656	0.4844	0.4616	0.4122	0.4756
	40 60	0.4210	0.3494	0.4218	0.4098	0.3426	0.3988
	30 70	0.4732	0.3398	0.3674	0.3552	0.2858	0.3336
8,1	50 50	0.2848	0.2800	0.3142	0.2986	0.2448	0.3192
	40 60	0.3468	0.2760	0.2958	0.2734	0.2232	0.2814
	30 70	0.4214	0.2728	0.2608	0.2516	0.1908	0.2236

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 39. *Empirical Type I Error with One Heavily Positively, One Heavily Negatively Skewed Distribution, N = 100.*

Variance Ratio	$n1,n2$	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*50 50*	**0.0792**	**0.0792**	0.0558	0.0516	0.0568	0.0620
	40 60	**0.0842**	**0.0860**	0.0554	0.0576	0.0528	0.0566
	30 70	**0.0762**	**0.0842**	0.0500	0.0598	0.0484	0.0584
1,4	*50 50*	**0.0870**	**0.0866**	0.0492	0.0482	0.0486	0.0624
	40 60	0.0634	**0.0826**	0.0492	0.0494	0.0506	0.0596
	30 70	0.0394	**0.0760**	0.0558	0.0522	0.0552	0.0616
1,8	*50 50*	**0.0942**	**0.0934**	0.0492	0.0554	0.0480	0.0620
	40 60	0.0594	**0.0844**	0.0458	0.0490	0.0452	0.0630
	30 70	0.0348	**0.0832**	0.0522	0.0530	0.0526	0.0668
4,1	*50 50*	**0.0836**	**0.0826**	0.0472	0.0552	0.0460	0.0588
	40 60	**0.1196**	**0.0956**	0.0596	0.0632	0.0566	0.0598
	30 70	**0.1530**	**0.1018**	0.0598	0.0572	0.0496	0.0632
8,1	*50 50*	**0.0920**	**0.0904**	0.0502	0.0494	0.0456	0.0696
	40 60	**0.1304**	**0.0954**	0.0572	0.0586	0.0472	0.0640
	30 70	**0.1874**	**0.1002**	0.0548	0.0590	0.0422	0.0554

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 40. *Empirical Power with One Heavily Positively, One Heavily Negatively Skewed Distribution, N = 100.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	50 50	0.4950	0.4938	0.7476	0.7336	0.6970	0.6424
	40 60	0.4918	0.4956	0.7422	0.7248	0.6826	0.6320
	30 70	0.4580	0.4816	0.6822	0.6798	0.6114	0.5774
1,4	50 50	0.3392	0.3382	0.4564	0.4324	0.3948	0.3924
	40 60	0.2788	0.3260	0.4666	0.4630	0.4086	0.4048
	30 70	0.2246	0.3346	0.4732	0.4508	0.4218	0.3950
1,8	50 50	0.2696	0.2666	0.3146	0.2958	0.2546	0.2824
	40 60	0.2176	0.2768	0.3434	0.3100	0.2880	0.3042
	30 70	0.1488	0.2656	0.3440	0.3090	0.2920	0.3184
4,1	50 50	0.3324	0.3302	0.4386	0.4264	0.3718	0.3814
	40 60	0.3646	0.3220	0.4076	0.4004	0.3316	0.3440
	30 70	0.3978	0.3146	0.3650	0.3406	0.2834	0.2952
8,1	50 50	0.2628	0.2610	0.3156	0.2826	0.2570	0.2778
	40 60	0.3084	0.2500	0.2800	0.2720	0.2166	0.2416
	30 70	0.3512	0.2572	0.2476	0.2618	0.1786	0.1972

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 41. *Empirical Type I Error with One Moderately Negatively, One Moderately Positively Skewed Distribution, N = 60.*

Variance Ratio	*n*1,*n*2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*30 30*	0.0548	0.0538	0.0520	0.0506	0.0510	0.0528
	25 35	0.0636	0.0636	0.0544	0.0514	0.0520	0.0558
	20 40	0.0598	0.0606	0.0534	0.0560	0.0534	0.0580
1,4	*30 30*	**0.0918**	**0.0914**	0.0490	0.0490	0.0484	0.0536
	25 35	**0.0972**	**0.1000**	0.0554	0.0538	0.0536	0.0654
	20 40	**0.0880**	**0.0940**	0.0502	0.0498	0.0458	0.0576
1,8	*30 30*	0.0656	0.0618	0.0568	0.0520	0.0514	0.0640
	25 35	0.0396	0.0626	0.0474	0.0484	0.0458	0.0674
	20 40	**0.0194**	0.0608	0.0512	0.0572	0.0540	0.0732
4,1	*30 30*	0.0638	0.0614	0.0516	0.0536	0.0502	0.0606
	25 35	**0.0986**	0.0724	0.0564	0.0512	0.0574	0.0634
	20 40	**0.1094**	0.0638	0.0514	0.0614	0.0464	0.0544
8,1	*30 30*	0.0660	0.0628	0.0540	0.0490	0.0536	0.0668
	25 35	**0.0946**	0.0618	0.0558	0.0568	0.0506	0.0664
	20 40	**0.1530**	0.0700	0.0584	0.0638	0.0482	0.0620

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 42. *Empirical Power with One Moderately Negatively, One Moderately Positively Skewed Distribution, N = 60.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	30 30	0.6728	0.6718	0.7580	0.7672	0.7580	0.7950
	25 35	0.6564	0.6544	0.7384	0.7748	0.7384	0.7846
	20 40	0.6085	0.6042	0.6824	0.7202	0.6824	0.7462
1,4	30 30	0.2704	0.2642	0.3464	0.3710	0.3950	0.4232
	25 35	0.2070	0.2870	0.3760	0.4048	0.4160	0.4388
	20 40	0.1588	0.3042	0.3944	0.4132	0.4388	0.4580
1,8	30 30	0.1576	0.1520	0.1954	0.2084	0.2290	0.2806
	25 35	0.0974	0.1626	0.2238	0.2270	0.2506	0.2982
	20 40	0.0480	0.1748	0.2406	0.2518	0.2764	0.3138
4,1	30 30	0.2792	0.2734	0.3462	0.3648	0.3878	0.4208
	25 35	0.3128	0.2236	0.2912	0.3286	0.3370	0.3760
	20 40	0.3404	0.1806	0.2424	0.2886	0.2740	0.3222
8,1	30 30	0.1390	0.1318	0.1872	0.1976	0.2124	0.2642
	25 35	0.1952	0.1114	0.1504	0.1986	0.1690	0.2304
	20 40	0.2390	0.0816	0.1176	0.1658	0.1242	0.1874

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 43. *Empirical Type I Error with One Moderately Negatively, One Moderately Positively Skewed Distribution, N = 100.*

Variance Ratio	*n*1,*n*2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*50 50*	0.0564	0.0562	0.0560	0.0510	0.0534	0.0558
	40 60	0.0564	0.0554	0.0488	0.0526	0.0504	0.0489
	30 70	0.0578	0.0608	0.0552	0.0530	0.0552	0.0558
1,4	*50 50*	**0.0908**	**0.0900**	0.0558	0.0502	0.0546	0.0620
	40 60	**0.0778**	**0.0810**	0.0456	0.0520	0.0474	0.0512
	30 70	**0.0796**	**0.0938**	0.0538	0.0490	0.0526	0.0598
1,8	*50 50*	0.0566	0.0550	0.0496	0.0546	0.0484	0.0658
	40 60	0.0284	0.0578	0.0518	0.0548	0.0514	0.0712
	30 70	**0.0152**	0.0582	0.0484	0.0490	0.0506	0.0704
4,1	*50 50*	0.0586	0.0570	0.0512	0.0526	0.0522	0.0584
	40 60	**0.0948**	0.0614	0.0514	0.0564	0.0490	0.0598
	30 70	**0.1306**	0.0646	0.0544	0.0614	0.0510	0.0596
8,1	*50 50*	0.0632	0.0610	0.0526	0.0518	0.0504	0.0652
	40 60	**0.1056**	0.0588	0.0550	0.0578	0.0554	0.0720
	30 70	**0.1702**	0.0684	0.0612	0.0614	0.0502	0.0656

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 44. *Empirical Power with One Moderately Negatively, One Moderately Positively Skewed Distribution, N = 100.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	50 50	0.8978	0.8974	0.9426	0.9442	0.9426	0.9490
	40 60	0.8854	0.8892	0.9324	0.9396	0.9324	0.9436
	30 70	0.8300	0.8364	0.8894	0.9194	0.8894	0.9148
1,4	50 50	0.4662	0.4636	0.5484	0.5836	0.5942	0.6126
	40 60	0.3842	0.5066	0.6054	0.6166	0.5408	0.6492
	30 70	0.2706	0.5136	0.6034	0.6302	0.6388	0.6420
1,8	50 50	0.2594	0.2540	0.3314	0.3444	0.3726	0.4234
	40 60	0.1860	0.3054	0.3764	0.3988	0.4194	0.4578
	30 70	0.0926	0.3172	0.4036	0.4278	0.4344	0.4700
4,1	50 50	0.4728	0.4668	0.5664	0.5790	0.6108	0.6290
	40 60	0.5180	0.4074	0.5012	0.5192	0.5484	0.5704
	30 70	0.5106	0.3036	0.3716	0.4288	0.4214	0.4592
8,1	50 50	0.2522	0.2470	0.3182	0.3356	0.3626	0.3954
	40 60	0.3356	0.2084	0.2682	0.2948	0.3120	0.3470
	30 70	0.3734	0.1446	0.1946	0.2386	0.2322	0.2718

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 45. *Empirical Type I Error with One Moderately Negatively, One Heavily Positively Skewed Distribution, N = 60.*

Variance Ratio	$n1,n2$	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*30 30*	**0.0856**	**0.0848**	0.0498	0.0460	0.0488	0.0524
	25 35	**0.0752**	**0.0820**	0.0534	0.0508	0.0536	0.0592
	20 40	0.0628	**0.0784**	0.0500	0.0524	0.0494	0.0546
1,4	*30 30*	**0.0764**	**0.0762**	0.0448	0.0518	0.0454	0.0502
	25 35	0.0620	0.0718	0.0480	0.0480	0.0502	0.0514
	20 40	0.0440	0.0680	0.0544	0.0534	0.0532	0.0588
1,8	*30 30*	**0.1046**	**0.1026**	0.0534	0.0462	0.0514	0.0616
	25 35	**0.0750**	**0.0940**	0.0514	0.0514	0.0470	0.0632
	20 40	0.0502	**0.0876**	0.0502	0.0502	0.0496	0.0616
4,1	*30 30*	**0.0874**	**0.0870**	0.0554	0.0510	0.0558	0.0642
	25 35	0.0610	**0.0882**	0.0500	0.0566	0.0544	0.0620
	20 40	0.0294	**0.0758**	0.0516	0.0580	0.0496	0.0636
8,1	*30 30*	0.0678	0.0658	0.0522	0.0532	0.0496	0.0654
	25 35	**0.0920**	0.0692	0.0542	0.0590	0.0500	0.0588
	20 40	**0.1182**	0.0676	0.0536	0.0526	0.0474	0.0588

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 46. *Empirical Type I Error with One Moderately Negatively, One Heavily Positively Skewed Distribution, N = 60.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	30 30	0.3500	0.3432	0.7372	0.7560	0.8056	0.7792
	25 35	0.2950	0.3584	0.7172	0.7570	0.7780	0.7500
	20 40	0.2364	0.3594	0.6718	0.7252	0.7388	0.7098
1,4	30 30	0.0776	0.0720	0.2678	0.3136	0.3362	0.4216
	25 35	0.0364	0.0816	0.3174	0.3476	0.3966	0.4476
	20 40	0.0154	0.0874	0.3374	0.3752	0.4098	0.4288
1,8	30 30	0.0478	0.0442	0.1434	0.1550	0.1418	0.2940
	25 35	0.0233	0.0486	0.1636	0.1928	0.1874	0.3006
	20 40	0.0104	0.0558	0.1780	0.2148	0.2180	0.3138
4,1	30 30	0.1624	0.1600	0.3170	0.3548	0.3760	0.3574
	25 35	0.1942	0.1704	0.2996	0.3350	0.3440	0.3510
	20 40	0.1936	0.1470	0.2414	0.2800	0.2732	0.2980
8,1	30 30	0.1086	0.1040	0.1880	0.2800	0.2144	0.2410
	25 35	0.1458	0.0940	0.1666	0.3278	0.1840	0.2178
	20 40	0.1732	0.0870	0.1356	0.2786	0.1338	0.1816

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 47. *Empirical Type I Error with One Moderately Negatively, One Heavily Positively Skewed Distribution, N = 100.*

Variance Ratio	$n1,n2$	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*50 50*	0.0628	0.0620	0.0524	0.0556	0.0498	0.0642
	40 60	**0.0876**	0.0592	0.0508	0.0480	0.0500	0.0634
	30 70	**0.1188**	0.0618	0.0500	0.0554	0.0474	0.0558
1,4	*50 50*	0.0664	0.0648	0.0514	0.0538	0.0492	0.0590
	40 60	0.0732	0.0626	0.0500	0.0594	0.0474	0.0720
	30 70	**0.0870**	0.0624	0.0462	0.0652	0.0426	0.0510
1,8	*50 50*	0.0580	0.0572	0.0494	0.0482	0.0514	0.0580
	40 60	**0.0772**	0.0648	0.0588	0.0562	0.0562	0.0656
	30 70	**0.0922**	0.0708	0.0598	0.0594	0.0532	0.0636
4,1	*50 50*	0.0678	0.0658	0.0522	0.0520	0.0496	0.0654
	40 60	**0.0920**	0.0692	0.0542	0.0560	0.0500	0.0588
	30 70	**0.1182**	0.0676	0.0536	0.0604	0.0474	0.0588
8,1	*50 50*	0.0628	0.0620	0.0524	0.0498	0.0498	0.0642
	40 60	**0.0876**	0.0592	0.0501	0.0538	0.0500	0.0634
	30 70	**0.1188**	0.0618	0.0500	0.0562	0.0474	0.0558

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 48. *Empirical Power with One Moderately Negatively, One Heavily Positively Skewed Distribution, N = 100.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	50 50	0.6216	0.6192	0.9396	0.9460	0.9562	0.9392
	40 60	0.5398	0.6340	0.9260	0.9310	0.9418	0.9286
	30 70	0.4198	0.6112	0.8850	0.9036	0.9126	0.8882
1,4	50 50	0.1474	0.1430	0.5014	0.5380	0.5920	0.6328
	40 60	0.0712	0.1602	0.5420	0.5820	0.6228	0.6442
	30 70	0.0204	0.1828	0.5594	0.6030	0.6234	0.6280
1,8	50 50	0.0704	0.0670	0.2638	0.2982	0.3436	0.4418
	40 60	0.0262	0.0782	0.3032	0.3298	0.3868	0.4630
	30 70	0.0050	0.0966	0.3558	0.3776	0.4292	0.4836
4,1	50 50	0.3158	0.3130	0.5450	0.5620	0.5924	0.5686
	40 60	0.3288	0.2956	0.4766	0.5106	0.5214	0.5052
	30 70	0.3064	0.2418	0.3692	0.4104	0.4214	0.4166
8,1	50 50	0.1976	0.1954	0.3158	0.3404	0.3572	0.3662
	40 60	0.2386	0.1696	0.2654	0.2892	0.2998	0.3142
	30 70	0.2538	0.1264	0.1892	0.2426	0.2208	0.2388

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 49. *Empirical Type I Error with One Heavily Negatively, One Moderately Positively Skewed Distribution, N = 60.*

Variance Ratio	$n1,n2$	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*30 30*	**0.0892**	**0.0888**	0.0512	0.0552	0.0488	0.0532
	25 35	**0.0954**	**0.0900**	0.0474	0.0526	0.0502	0.0508
	20 40	**0.1096**	**0.0968**	0.0534	0.0558	0.0506	0.0566
1,4	*30 30*	0.0662	0.0654	0.0488	0.0496	0.0500	0.0382
	25 35	0.0684	**0.0804**	0.0538	0.0508	0.0534	0.0602
	20 40	0.0526	**0.0820**	0.0534	0.0490	0.0520	0.0660
1,8	*30 30*	0.0646	0.0634	0.0536	0.0572	0.0560	0.0632
	25 35	0.0470	0.0652	0.0562	0.0474	0.0526	0.0638
	20 40	0.0330	0.0612	0.0544	0.0488	0.0502	0.0760
4,1	*30 30*	**0.1120**	**0.1096**	0.0586	0.0580	0.0480	0.0624
	25 35	**0.1532**	**0.1196**	0.0638	0.0512	0.0526	0.0654
	20 40	**0.1946**	**0.1160**	0.0698	0.0654	0.0522	0.0622
8,1	*30 30*	**0.1116**	**0.1118**	0.0626	0.0560	0.0464	0.0686
	25 35	**0.1610**	**0.1070**	0.0594	0.0588	0.0436	0.0646
	20 40	**0.2224**	**0.1330**	0.0672	0.0666	0.0516	0.0664

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 50. *Empirical Power with One Heavily Negatively, One Moderately Positively Skewed Distribution, N = 60.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	30 30	0.3426	0.3380	0.7412	0.7474	0.8000	0.7748
	25 35	0.3696	0.2944	0.6986	0.7548	0.7752	0.7570
	20 40	0.3764	0.2342	0.6564	0.7118	0.7352	0.7388
1,4	30 30	0.1818	0.1796	0.3438	0.3578	0.3916	0.3824
	25 35	0.1612	0.1780	0.2694	0.4012	0.4182	0.4050
	20 40	0.1326	0.1532	0.3596	0.4076	0.4062	0.3924
1,8	30 30	0.1104	0.1078	0.1948	0.2070	0.2212	0.2426
	25 35	0.0884	0.1256	0.2248	0.2448	0.2576	0.2730
	20 40	0.0564	0.1154	0.2406	0.2516	0.2686	0.2782
4,1	30 30	0.0718	0.0666	0.2822	0.3150	0.3482	0.4352
	25 35	0.1194	0.0572	0.2272	0.2742	0.2726	0.3900
	20 40	0.1896	0.0530	0.1684	0.2348	0.1828	0.3444
8,1	30 30	0.0528	0.0486	0.1388	0.1586	0.1448	0.2814
	25 35	0.0988	0.0506	0.1120	0.1368	0.1016	0.2546
	20 40	0.1738	0.0550	0.0838	0.1216	0.0680	0.2178

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 51. *Empirical Type I Error with One Heavily Negatively, One Moderately Positively Skewed Distribution, N = 100.*

Variance Ratio	$n1,n2$	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*50 50*	0.0672	0.0668	0.0470	0.0442	0.0474	0.0524
	40 60	**0.0966**	**0.0828**	0.0524	0.0510	0.0540	0.0566
	30 70	**0.1234**	**0.0978**	0.0512	0.0554	0.0516	0.0558
1,4	*50 50*	0.0618	0.0624	0.0482	0.0520	0.0584	0.0588
	40 60	0.0508	0.0624	0.0512	0.0544	0.0544	0.0586
	30 70	0.0394	0.0672	0.0502	0.0584	0.0528	0.0620
1,8	*50 50*	0.0652	0.0648	0.0568	0.0532	0.0572	0.0662
	40 60	0.0428	0.0650	0.0542	0.0486	0.0574	0.0630
	30 70	0.0256	0.0630	0.0532	0.0486	0.0518	0.0680
4,1	*50 50*	**0.0894**	**0.0866**	0.0500	0.0526	0.0490	0.0622
	40 60	**0.1452**	**0.1036**	0.0600	0.0554	0.0584	0.0622
	30 70	**0.2134**	**0.1174**	0.0554	0.0582	0.0500	0.0598
8,1	*50 50*	**0.0946**	**0.0992**	0.0556	0.0518	0.0502	0.0626
	40 60	**0.1474**	**0.1030**	0.0530	0.0588	0.0446	0.0610
	30 70	**0.2266**	**0.1060**	0.0600	0.0652	0.0474	0.0580

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 52. *Empirical Power with One Heavily Negatively, One Moderately Positively Skewed Distribution, N = 100.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	50 50	0.6282	0.6260	0.9384	0.9476	0.9552	0.9382
	40 60	0.6328	0.5426	0.9246	0.9408	0.9514	0.9304
	30 70	0.6074	0.4354	0.8762	0.9122	0.9240	0.9060
1,4	50 50	0.3250	0.3230	0.5520	0.5640	0.6000	0.5686
	40 60	0.2776	0.3140	0.5810	0.6220	0.6252	0.5770
	30 70	0.2376	0.2886	0.6030	0.6154	0.6520	0.5862
1,8	50 50	0.2072	0.2044	0.3312	0.3494	0.3726	0.3708
	40 60	0.1516	0.2084	0.3646	0.3960	0.4038	0.3948
	30 70	0.1102	0.2128	0.4038	0.4200	0.4428	0.4200
4,1	50 50	0.1368	0.1320	0.4924	0.5448	0.5818	0.6248
	40 60	0.2212	0.1030	0.4200	0.4482	0.5234	0.5786
	30 70	0.2806	0.0708	0.2974	0.3686	0.3862	0.4790
8,1	50 50	0.0746	0.0708	0.2578	0.2860	0.3298	0.4338
	40 60	0.1438	0.0584	0.2078	0.2440	0.2642	0.3858
	30 70	0.2308	0.0522	0.1348	0.1812	0.1498	0.2908

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 53. *Empirical Type I Error with One Heavily Negatively, One Heavily Positively Skewed Distribution, N = 100.*

Variance Ratio	*n*1,*n*2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*50 50*	**0.0832**	**0.0830**	0.0580	0.0522	0.0570	0.0678
	40 60	**0.0846**	**0.0860**	0.0530	0.0552	0.0522	0.0576
	30 70	**0.0828**	**0.0892**	0.0506	0.0564	0.0528	0.0604
1,4	*50 50*	**0.0894**	**0.0884**	0.0480	0.0498	0.0484	0.0586
	40 60	0.0628	**0.0790**	0.0512	0.0462	0.0490	0.0644
	30 70	0.0388	**0.0788**	0.0602	0.0476	0.0566	0.0616
1,8	*50 50*	**0.0908**	**0.0896**	0.0528	0.0548	0.0474	0.0642
	40 60	0.0614	**0.0898**	0.0532	0.0510	0.0520	0.0688
	30 70	0.0314	**0.0788**	0.0462	0.0446	0.0464	0.0708
4,1	*50 50*	**0.0828**	**0.0826**	0.0558	0.0518	0.0570	0.0656
	40 60	**0.1140**	**0.0916**	0.0524	0.0530	0.0490	0.0618
	30 70	**0.1508**	**0.1026**	0.0568	0.0620	0.0480	0.0612
8,1	*50 50*	**0.0966**	**0.0960**	0.0558	0.0528	0.0528	0.0738
	40 60	**0.1392**	**0.1028**	0.0602	0.0544	0.0538	0.0638
	30 70	**0.1832**	**0.1044**	0.0564	0.0590	0.0422	0.0596

Notes: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded entries indicate results outside the acceptable range (2.5% - 7.5%).

Table 54. *Empirical Power with One Heavily Negatively, One Heavily Positively Skewed Distribution, N = 100.*

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	50 50	0.3630	0.3616	0.9278	0.9466	0.9578	0.9118
	40 60	0.3600	0.3528	0.9308	0.9404	0.9560	0.9046
	30 70	0.2972	0.3852	0.8662	0.8980	0.9152	0.8480
1,4	50 50	0.1112	0.1094	0.5006	0.5236	0.5812	0.5484
	40 60	0.0700	0.1224	0.5434	0.5738	0.6192	0.5682
	30 70	0.0424	0.1242	0.5448	0.5888	0.6226	0.5432
1,8	50 50	0.0654	0.0626	0.2522	0.2944	0.3276	0.3622
	40 60	0.0288	0.0732	0.3194	0.3376	0.3944	0.4022
	30 70	0.0110	0.0778	0.3476	0.3856	0.4124	0.3970
4,1	50 50	0.1166	0.1122	0.4796	0.5196	0.5712	0.5408
	40 60	0.1152	0.0868	0.3970	0.4658	0.4950	0.4822
	30 70	0.2108	0.0674	0.2992	0.3824	0.3702	0.4020
8,1	50 50	0.0712	0.0680	0.2616	0.2892	0.3316	0.3558
	40 60	0.1240	0.0584	0.2058	0.2386	0.2552	0.3162
	30 70	0.1794	0.0486	0.1442	0.1838	0.1548	0.2470

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 1.

Empirical Type I Error with Two Normal Distributions, N = 60.

Variance Ratio	n_1	n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*30*	*30*	0.0495	0.0493	0.0515	0.0482	0.0519	0.0500
	25	*35*	0.0495	0.0493	0.0499	0.0568	0.0499	0.0504
	20	*40*	0.0494	0.0504	0.0564	0.0614	0.0550	0.0540
1,4	*30*	*30*	0.0532	0.0516	0.0524	0.0476	0.0514	0.0584
	25	*35*	0.0323	0.0506	0.0511	0.0486	0.0509	0.0600
	20	*40*	**0.0185**	0.0500	0.0490	0.0508	0.0485	0.0588
1,8	*30*	*30*	0.0499	0.0476	0.0494	0.0536	0.0484	0.0590
	25	*35*	0.0299	0.0504	0.0517	0.0534	0.0515	0.0647
	20	*40*	**0.0119**	0.0513	0.0541	0.0478	0.0546	0.0674
4,1	*30*	*30*	0.0528	0.0511	0.0471	0.0594	0.0477	0.0552
	25	*35*	**0.0795**	0.0512	0.0562	0.0544	0.0531	0.0507
	20	*40*	**0.1111**	0.0496	0.0542	0.0618	0.0524	0.0563
8,1	*30*	*30*	0.0519	0.0494	0.0512	0.0490	0.0492	0.0647
	25	*35*	**0.0874**	0.0472	0.0524	0.0572	0.0491	0.0627
	20	*40*	**0.1387**	0.0487	0.0548	0.0630	0.0508	0.0626

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded values reflect empirical Type I error outside the acceptable criterion (2.5% - 7.5%).

Table 2.

Empirical Type I Error with Two Normal Distributions, N = 100.

Variance Ratio	n_1	n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*50*	*50*	0.0489	0.0489	0.0496	0.0542	0.0508	0.0493
	40	*60*	0.0494	0.0492	0.0506	0.0542	0.0482	0.0478
	30	*70*	0.0489	0.0494	0.0508	0.0470	0.0511	0.0504
1,4	*50*	*50*	0.0496	0.0481	0.0479	0.0464	0.0495	0.0554
	40	*60*	0.0277	0.0488	0.0498	0.0482	0.0509	0.0578
	30	*70*	**0.0126**	0.0490	0.0481	0.0526	0.0508	0.0591
1,8	*50*	*50*	0.0540	0.0522	0.0532	0.0574	0.0522	0.0682
	40	*60*	**0.0247**	0.0519	0.0519	0.0462	0.0514	0.0701
	30	*70*	**0.0111**	0.0510	0.0504	0.0520	0.0525	0.0622
4,1	*50*	*50*	0.0534	0.0525	0.0508	0.0512	0.0519	0.0600
	40	*60*	**0.0846**	0.0497	0.0517	0.0516	0.0509	0.0575
	30	*70*	**0.1302**	0.0522	0.0554	0.0614	0.0517	0.0587
8,1	*50*	*50*	0.0514	0.0495	0.0522	0.0552	0.0518	0.0685
	40	*60*	**0.0965**	0.0490	0.0523	0.0522	0.0478	0.0639
	30	*70*	**0.1584**	0.0507	0.0519	0.0590	0.0493	0.0600

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded values reflect empirical Type I error outside the acceptable criterion (2.5% - 7.5%).

Table 3.

Empirical Power with Two Normal Distributions, N = 60.

Variance Ratio	n_1	n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	30	30	0.8004	0.8001	0.7358	0.7444	0.7352	0.7808
	25	35	0.7871	0.7856	0.7179	0.7184	0.7164	0.7674
	20	40	0.7528	0.7454	0.6740	0.6846	0.6687	0.7317
1,4	30	30	0.4289	0.4224	0.3625	0.3740	0.3653	0.4098
	25	35	0.3743	0.4505	0.3979	0.3974	0.3966	0.4438
	20	40	0.2995	0.4667	0.4074	0.4212	0.4051	0.4570
1,8	30	30	0.2619	0.2542	0.2240	0.2158	0.2216	0.2631
	25	35	0.2047	0.2827	0.2547	0.2434	0.2494	0.2946
	20	40	0.1361	0.3034	0.2677	0.2730	0.2670	0.3189
4,1	30	30	0.4269	0.4211	0.3747	0.3746	0.3667	0.4139
	25	35	0.4796	0.3207	0.2727	0.3414	0.2641	0.3112
	20	40	0.4802	0.3209	0.2795	0.3034	0.2696	0.3154
8,1	30	30	0.2682	0.2597	0.2305	0.2338	0.2250	0.2709
	25	35	0.3206	0.2290	0.2032	0.2128	0.1907	0.2326
	20	40	0.3559	0.1882	0.1678	0.1918	0.1575	0.1935

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 4.

Empirical Power with Two Normal Distributions, N = 100.

Variance Ratio	n_1	n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	50	50	0.9537	0.9536	0.9215	0.9272	0.9216	0.9435
	40	60	0.9454	0.9439	0.9100	0.9143	0.9108	0.9348
	30	70	0.9164	0.9121	0.8644	0.8688	0.8638	0.9008
1,4	50	50	0.6330	0.6296	0.5635	0.5634	0.5641	0.6080
	40	60	0.5723	0.6689	0.6045	0.6118	0.6074	0.6504
	30	70	0.4646	0.6826	0.6125	0.6012	0.6177	0.6653
1,8	50	50	0.4070	0.4009	0.3571	0.3522	0.3545	0.3982
	40	60	0.3308	0.4562	0.4058	0.3926	0.4046	0.4574
	30	70	0.2148	0.4804	0.4261	0.4144	0.4296	0.4804
4,1	50	50	0.6362	0.6332	0.5695	0.5808	0.5677	0.6092
	40	60	0.6570	0.5636	0.4984	0.5042	0.4965	0.5427
	30	70	0.6430	0.4594	0.3981	0.4298	0.3895	0.4361
8,1	50	50	0.4029	0.3983	0.3482	0.3458	0.3488	0.3887
	40	60	0.4553	0.3388	0.3001	0.3108	0.2934	0.3310
	30	70	0.4919	0.2737	0.2380	0.2614	0.2203	0.2658

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 5.

Empirical Type I Error with Two Heavily Positively Skewed Distributions, N = 60.

Variance Ratio	$n1,n2$		t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*30*	*30*	0.0441	0.0411	0.0470	0.0444	0.0535	0.0497
	25	*35*	0.0422	0.0406	0.0487	0.0390	0.0522	0.0509
	20	*40*	0.0427	0.0512	0.0502	0.0546	0.0555	0.0517
1,4	*30*	*30*	0.0735	0.0722	0.0586	0.0532	0.0586	0.0596
	25	*35*	0.0459	0.0593	0.0536	0.0474	0.0575	0.0570
	20	*40*	0.0393	0.0490	0.0520	0.0496	0.0545	0.0576
1,8	*30*	*30*	**0.0908**	**0.0893**	0.0614	0.0536	0.0570	0.0648
	25	*35*	0.0632	**0.0781**	0.0640	0.0582	0.0601	0.0689
	20	*40*	0.0416	0.0657	0.0587	0.0516	0.0595	0.0694
4,1	*30*	*30*	0.0734	0.0726	0.0582	0.0490	0.0590	0.0591
	25	*35*	**0.0955**	**0.0840**	0.0555	0.0572	0.0554	0.0547
	20	*40*	**0.1251**	**0.0998**	0.0639	0.0672	0.0595	0.0584
8,1	*30*	*30*	**0.0899**	**0.0876**	0.0628	0.0616	0.0574	0.0666
	25	*35*	**0.1249**	**0.1013**	0.0641	0.0632	0.0563	0.0628
	20	*40*	**0.1720**	**0.1170**	0.0664	0.0722	0.0560	0.0608

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded values reflect empirical Type I error outside the acceptable criterion (2.5% - 7.5%).

Table 6.

Empirical Power with Two Heavily Positively Skewed Distributions, N = 60.

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	30 30	0.2614	0.3594	0.6360	0.6318	0.6470	0.8342
	25 35	0.3592	0.3836	0.6058	0.6046	0.5966	0.7782
	20 40	0.3426	0.3926	0.5584	0.5718	0.5340	0.6936
1,4	30 30	0.0939	0.0882	0.2845	0.2928	0.3265	0.6017
	25 35	0.0585	0.1097	0.3147	0.3218	0.3545	0.6105
	20 40	0.0327	0.1433	0.3338	0.3384	0.3623	0.5777
1,8	30 30	0.0499	0.0457	0.1402	0.1436	0.1701	0.3480
	25 35	0.0280	0.0582	0.1640	0.1722	0.1958	0.3735
	20 40	0.0121	0.0603	0.1746	0.1950	0.2047	0.3719
4,1	30 30	0.2698	0.2678	0.4090	0.3310	0.3708	0.3890
	25 35	0.3035	0.2784	0.3741	0.3238	0.3289	0.3388
	20 40	0.3311	0.2835	0.3268	0.3026	0.2809	0.2808
8,1	30 30	0.2334	0.2301	0.3023	0.2344	0.2477	0.2562
	25 35	0.2789	0.2375	0.2706	0.2404	0.2211	0.2177
	20 40	0.3243	0.2482	0.2396	0.2220	0.1907	0.1845

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 7.

Empirical Type I Error with One Normal and One Heavily Negatively Skewed Distribution, N = 60.

Variance Ratio	$n1,n2$	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*30 30*	**0.0800**	**0.0876**	0.0506	0.0516	0.0594	0.0598
	25 35	0.0656	0.0718	0.0468	0.0544	0.0546	0.0508
	20 40	0.0458	0.0614	0.0448	0.0574	0.0510	0.0510
1,4	*30 30*	**0.1062**	**0.1042**	0.0562	0.0540	0.0546	0.0614
	25 35	**0.0788**	**0.1008**	0.0552	0.0404	0.0572	0.0624
	20 40	0.0484	**0.0904**	0.0504	0.0522	0.0514	0.0628
1,8	*30 30*	**0.1140**	**0.1102**	0.0580	0.0506	0.0442	0.0654
	25 35	**0.0778**	**0.1036**	0.0542	0.0522	0.0490	0.0700
	20 40	0.0534	**0.1020**	0.0566	0.0554	0.0518	0.0740
4,1	*30 30*	0.0548	0.0536	0.0518	0.0522	0.0542	0.0560
	25 35	0.0644	0.0542	0.0502	0.0508	0.0514	0.0526
	20 40	0.0698	0.0494	0.0550	0.0608	0.0542	0.0594
8,1	*30 30*	0.0504	0.0486	0.0514	0.0480	0.0510	0.0600
	25 35	0.0734	0.0526	0.0520	0.0540	0.0530	0.0642
	20 40	**0.0874**	0.0536	0.0534	0.0598	0.0522	0.0624

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded values reflect empirical Type I error outside the acceptable criterion (2.5% - 7.5%).

Table 8.

Empirical Power with One Normal and One Heavily Negatively Skewed Distribution, N = 60.

Variance Ratio	n_1	n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	30	30	0.5322	0.5290	0.6316	0.6196	0.5874	0.6672
	25	35	0.4918	0.5366	0.6334	0.6370	0.5980	0.6846
	20	40	0.4416	0.5342	0.6142	0.5970	0.5876	0.6734
1,4	30	30	0.3368	0.3318	0.3410	0.3228	0.2806	0.3614
	25	35	0.2686	0.3262	0.3458	0.3458	0.2966	0.3808
	20	40	0.2174	0.3250	0.3540	0.3622	0.3062	0.4054
1,8	30	30	0.2624	0.2560	0.2444	0.2330	0.1860	0.2426
	25	35	0.2162	0.2680	0.2560	0.2430	0.2004	0.2788
	20	40	0.1524	0.2624	0.2688	0.2504	0.2200	0.3068
4,1	30	30	0.3714	0.3690	0.3432	0.3348	0.3348	0.3684
	25	35	0.3608	0.3318	0.3064	0.3042	0.2996	0.3364
	20	40	0.3422	0.2990	0.2604	0.2796	0.2526	0.2982
8,1	30	30	0.2518	0.2462	0.2110	0.2196	0.2102	0.2358
	25	35	0.2688	0.2088	0.1754	0.1936	0.1768	0.2078
	20	40	0.2870	0.1884	0.1666	0.1808	0.1624	0.1934

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 9.

Empirical Type I Error with One Moderately Positively Skewed and One Heavily Positively Skewed Distribution, N = 60.

Variance Ratio	*n*1,*n*2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*30 30*	0.0687	0.0681	0.0534	0.0472	0.0605	0.0519
	25 35	0.0566	0.0573	0.0522	0.0516	0.0596	0.0519
	20 40	0.0454	0.0491	0.0498	0.0480	0.0577	0.0511
1,4	*30 30*	**0.1027**	**0.1010**	0.0658	0.0548	0.0636	0.0619
	25 35	0.0732	**0.0920**	0.0641	0.0544	0.0644	0.0602
	20 40	0.0463	**0.0825**	0.0607	0.0530	0.0635	0.0611
1,8	*30 30*	**0.1138**	**0.1111**	0.0712	0.0554	0.0617	0.0654
	25 35	0.0725	0.0969	0.0680	0.0510	0.0631	0.0660
	20 40	0.0477	0.0929	0.0674	0.0576	0.0655	0.0699
4,1	*30 30*	0.0485	0.0470	0.0519	0.0512	0.0570	0.0604
	25 35	0.0566	0.0467	0.0512	0.0526	0.0547	0.0592
	20 40	0.0697	0.0496	0.0524	0.0656	0.0523	0.0586
8,1	*30 30*	0.0627	0.0603	0.0534	0.0526	0.0526	0.0661
	25 35	**0.0981**	0.0636	0.0551	0.0562	0.0489	0.0636
	20 40	**0.1017**	0.0569	0.0531	0.0668	0.0486	0.0608

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded values reflect empirical Type I error outside the acceptable criterion (2.5% - 7.5%).

Table 10.

Empirical Power with One Moderately Positively Skewed and One Heavily Positively Skewed Distribution, N = 60.

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	30 30	0.3850	0.3788	0.6764	0.6850	0.7078	0.8232
	25 35	0.3360	0.3980	0.6630	0.6702	0.6692	0.7806
	20 40	0.2750	0.4278	0.6234	0.6378	0.6168	0.7250
1,4	30 30	0.0759	0.0706	0.2379	0.2834	0.2748	0.5791
	25 35	0.0400	0.0867	0.2681	0.3410	0.3064	0.5954
	20 40	0.0163	0.1085	0.2796	0.3680	0.3079	0.5681
1,8	30 30	0.0542	0.0502	0.1242	0.1552	0.1477	0.3457
	25 35	0.0245	0.0548	0.1416	0.1832	0.1694	0.3799
	20 40	0.0101	0.0579	0.1499	0.2054	0.1789	0.3851
4,1	30 30	0.2605	0.2594	0.3305	0.3636	0.2971	0.3882
	25 35	0.2701	0.2634	0.2962	0.3410	0.2602	0.3365
	20 40	0.2716	0.2590	0.2687	0.3120	0.2297	0.2840
8,1	30 30	0.2148	0.2123	0.2278	0.2348	0.1961	0.2534
	25 35	0.2327	0.2024	0.2032	0.2300	0.1666	0.2145
	20 40	0.2659	0.2030	0.1890	0.2104	0.1510	0.1857

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 11.

Empirical Type I Error with One Heavily Positively Skewed and One Heavily Negatively Skewed Distribution, N = 60.

Variance Ratio	$n1,n2$	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*30 30*	**0.0940**	**0.0924**	0.0534	0.0490	0.0544	0.0600
	25 35	**0.0944**	**0.0968**	0.0512	0.0500	0.0490	0.0570
	20 40	**0.0934**	**0.1006**	0.0536	0.0552	0.0514	0.0612
1,4	*30 30*	**0.1104**	**0.1090**	0.0544	0.0638	0.0478	0.0604
	25 35	**0.0844**	**0.1004**	0.0486	0.0508	0.0448	0.0626
	20 40	0.0604	**0.0928**	0.0496	0.0546	0.0460	0.0628
1,8	*30 30*	**0.1076**	**0.1050**	0.0602	0.0592	0.0442	0.0738
	25 35	0.0732	**0.0960**	0.0520	0.0542	0.0456	0.0686
	20 40	0.0496	**0.0952**	0.0522	0.0504	0.0388	0.0676
4,1	*30 30*	**0.1082**	**0.1064**	0.0582	0.0450	0.0510	0.0640
	25 35	**0.1186**	**0.0994**	0.0554	0.0506	0.0466	0.0616
	20 40	**0.1482**	**0.1092**	0.0556	0.0584	0.0454	0.0558
8,1	*30 30*	**0.1044**	**0.1026**	0.0602	0.0574	0.0452	0.0716
	25 35	**0.1364**	**0.1088**	0.0554	0.0648	0.0434	0.0646
	20 40	**0.1734**	**0.1182**	0.0620	0.0676	0.0448	0.0628

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded values reflect empirical Type I error outside the acceptable criterion (2.5% - 7.5%).

Table 12.

Empirical Power with One Heavily Positively Skewed and One Heavily Negatively Skewed Distribution, N = 60.

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	30 30	0.4216	0.4198	0.5540	0.5120	0.4676	0.4474
	25 35	0.4012	0.4052	0.5364	0.5242	0.4520	0.4334
	20 40	0.3880	0.4026	0.5120	0.5120	0.4258	0.4092
1,4	30 30	0.2894	0.2864	0.3056	0.2994	0.2388	0.2586
	25 35	0.2496	0.2824	0.3114	0.3004	0.2466	0.2656
	20 40	0.2108	0.2904	0.3240	0.3168	0.2632	0.2858
1,8	30 30	0.2466	0.2432	0.2334	0.2146	0.1726	0.1896
	25 35	0.2094	0.2502	0.2472	0.2286	0.1908	0.2162
	20 40	0.1588	0.2486	0.2388	0.2198	0.1854	0.2144
4,1	30 30	0.3022	0.2998	0.3144	0.2966	0.2498	0.2558
	25 35	0.2114	0.2808	0.3022	0.2930	0.2282	0.2360
	20 40	0.3530	0.2966	0.2862	0.2880	0.3116	0.2138
8,1	30 30	0.2570	0.2514	0.2300	0.2058	0.1700	0.1986
	25 35	0.2904	0.2486	0.2184	0.2144	0.1610	0.1782
	20 40	0.3240	0.2428	0.2120	0.2150	0.1552	0.1626

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 13.

Empirical Type I Error with One Heavily Negatively Skewed and One Heavily Positively Skewed Distribution, N = 60.

Variance Ratio	$n1,n2$	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	*30 30*	**0.0956**	**0.0954**	0.0546	0.0450	0.0516	0.0580
	25 35	**0.1032**	**0.1020**	0.0532	0.0536	0.0508	0.0582
	20 40	**0.0960**	**0.1006**	0.0606	0.0584	0.0538	0.0624
1,4	*30 30*	**0.0996**	**0.0980**	0.0536	0.0540	0.0436	0.0666
	25 35	**0.0778**	**0.0942**	0.0520	0.0516	0.0478	0.0670
	20 40	**0.0696**	**0.1010**	0.0556	0.0518	0.0490	0.0636
1,8	*30 30*	**0.1026**	**0.1014**	0.0566	0.0506	0.0426	0.0736
	25 35	0.0748	**0.0978**	0.0580	0.0488	0.0434	0.0718
	20 40	0.0508	**0.0952**	0.0504	0.0484	0.0414	0.0700
4,1	*30 30*	**0.0940**	**0.0934**	0.0490	0.0564	0.0424	0.0638
	25 35	**0.1234**	**0.1070**	0.0524	0.0540	0.0424	0.0606
	20 40	**0.1584**	**0.1192**	0.0644	0.0668	0.0508	0.0598
8,1	*30 30*	**0.1126**	**0.1106**	0.0548	0.0544	0.0418	0.0668
	25 35	**0.1390**	**0.1080**	0.0598	0.0554	0.0478	0.0662
	20 40	**0.1852**	**0.1202**	0.0596	0.0652	0.0446	0.0654

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data. Bolded values reflect empirical Type I error outside the acceptable criterion (2.5% - 7.5%).

Table 14.

Empirical Power with One Heavily Negatively Skewed and One Heavily Positively Skewed Distribution, N = 60.

Variance Ratio	n_1 n_2	t	t_w	t_y	$WJ_{t(Jn)}$	$Bt_{y(Jn)}$	W_r
1,1	30 30	0.1788	0.1750	0.6930	0.7586	0.7820	0.7124
	25 35	0.1712	0.1608	0.6838	0.7410	0.7716	0.7040
	20 40	0.1728	0.1532	0.6306	0.7096	0.7180	0.6754
1,4	30 30	0.0700	0.0668	0.2858	0.3110	0.3450	0.3786
	25 35	0.0390	0.0642	0.3122	0.3454	0.3872	0.3864
	20 40	0.0260	0.0654	0.3170	0.3668	0.3872	0.3712
1,8	30 30	0.0490	0.0458	0.1304	0.1570	0.1304	0.2314
	25 35	0.0304	0.0526	0.1672	0.1930	0.1840	0.2550
	20 40	0.0152	0.0512	0.1754	0.2076	0.1982	0.2528
4,1	30 30	0.0632	0.0606	0.2746	0.3138	0.3350	0.3590
	25 35	0.0978	0.0550	0.2384	0.2864	0.2722	0.3436
	20 40	0.1338	0.0508	0.1776	0.2382	0.1860	0.2914
8,1	30 30	0.0588	0.0550	0.1488	0.1714	0.1490	0.2480
	25 35	0.0954	0.0522	0.1170	0.1442	0.0976	0.2244
	20 40	0.1374	0.0526	0.0954	0.1192	0.0702	0.1876

Note: t=Traditional t test; t_w=Welch; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); $Bt_{y(Jn)}$=Bootstrapped Yuen with Johnson Transformation; W_r=Welch on ranked data.

Table 15.

Empirical Type I Error Rates with All Normal Distributions, N = 120 and K = 4.

Variance Ratio	*n*1,*n*2,*n*3,*n*4	*F*	t_y	$WJ_{t(Jn)}$	W_r
1,1,1,1	30,30,30,30	0.0458	0.0530	0.0490	0.0492
	20,27,33,40	0.0458	0.0542	0.0586	0.0520
1,3,5,7	30,30,30,30	0.0592	0.0574	0.0524	0.0614
	20,27,33,40	0.0374	0.0518	0.0556	0.0590
	40,33,27,20	**0.1062**	0.0570	0.0430	0.0584

Note: *F*=ANOVA *F* test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data. Highlighted values reflect empirical Type I error outside the acceptable criterion (2.5% - 7.5%).

Table 16. Empirical Type I Error with All Normal Distributions, N = 210 and K = 7.

Variance Ratio	*n*1,*n*2,*n*3,*n*4,*n*5,*n*6,*n*7	*F*	t_y	$WJ_{t(Jn)}$	W_r
1,1,1,1,1,1,1	30,30,30,30,30,30,30	0.0480	0.0504	0.0448	0.0486
	20,24,27,30,33,36,40	0.0498	0.0558	0.0596	0.0566
1,2,3,4,5,6,7	30,30,30,30,30,30,30	0.0586	0.0606	0.0502	0.0588
	20,24,27,30,33,36,40	0.0404	0.0558	0.0582	0.0562
	40,36,33,30,27,24,20	**0.0994**	0.0564	0.0290	0.0576

Note: *F*=ANOVA *F* test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data. Bolded values reflect empirical Type I error outside the acceptable criterion (2.5% - 7.5%).

Table 17.

Empirical Power with All Normal Distributions, N = 120 and K = 4.

Variance Ratio	$n1,n2,n3,n4$	F	t_y	$WJ_{t(Jn)}$	W_r
1,1,1,1	30,30,30,30	0.8012	0.7140	0.7078	0.7764
	20,27,33,40	0.7662	0.6700	0.6936	0.7324
1,3,5,7	30,30,30,30	0.2458	0.2658	0.2600	0.3406
	20,27,33,40	0.1686	0.2746	0.2850	0.3564
	40,33,27,20	0.3072	0.2332	0.1964	0.2948

Note: F=ANOVA F test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data.

Table 18. Empirical Power with All Normal Distributions, N = 210 and K = 7.

Variance Ratio	$n1,n2,n3,n4,n5,n6,n7$	F	t_y	$WJ_{t(Jn)}$	W_r
1,1,1,1,1,1,1	30,30,30,30,30,30,30	0.8024	0.7000	0.6836	0.7680
	20,24,27,30,33,36,40	0.7814	0.6802	0.6998	0.7486
1,2,3,4,5,6,7	30,30,30,30,30,30,30	0.2458	0.2654	0.2430	0.3016
	20,24,27,30,33,36,40	0.1582	0.2534	0.2634	0.3016
	40,36,33,30,27,24,20	0.3134	0.2440	0.1630	0.2782

Note: F=ANOVA F test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data.

Table 19.

Empirical Type I Error with All Distributions Heavily Positively Skewed, N = 120 and K = 4.

Variance Ratio	*n*1,*n*2,*n*3,*n*4	F	t_y	$WJ_{t(Jn)}$	W_r
	30,30,30,30	0.0372	0.0382	0.0406	0.0446
1,1,1,1	20,27,33,40	0.0384	0.0484	0.0594	0.0532
	40,33,27,20	0.0416	0.0554	0.0414	0.0530
	30,30,30,30	0.0592	0.0626	0.0598	0.0572
1,3,5,7	20,27,33,40	0.0336	0.0490	0.0566	0.0534
	40,33,27,20	**0.1036**	0.0710	0.0528	0.0562

Note: F=ANOVA F test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data. Bolded values reflect empirical Type I error outside the acceptable criterion (2.5% - 7.5%)

Table 20.

Empirical Type I Error with All Distributions Heavily Positively Skewed, N = 210 and K = 7.

Variance Ratio	*n*1,*n*2,*n*3,*n*4,*n*5,*n*6,*n*7	F	t_y	$WJ_{t(Jn)}$	W_r
	30,30,30,30,30,30,30	0.0326	0.0550	0.0502	0.0516
1,1,1,1,1,1,1	20,24,27,30,33,36,40	0.0398	0.0596	0.0674	0.0524
	40,36,33,30,27,24,20	0.0460	0.0624	0.0342	0.0586
	30,30,30,30,30,30,30	0.0562	0.0688	0.0596	0.0550
1,2,3,4,5,6,7	20,24,27,30,33,36,40	0.0386	0.0634	0.0706	0.0592
	40,36,33,30,27,24,20	**0.0850**	**0.0778**	0.0430	0.0610

Note: F=ANOVA F test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data. Bolded values reflect empirical Type I error outside the acceptable criterion (2.5% - 7.5%).

Table 21.

Empirical Power with All Distributions Heavily Positively Skewed, N = 120 and K = 4.

Variance Ratio	*n*1,*n*2,*n*3,*n*4	F	t_y	$WJ_{t(Jn)}$	W_r
1,1,1,1	30,30,30,30	0.2736	0.5950	0.5942	0.8776
	20,27,33,40	0.2586	0.5808	0.6032	0.8168
	40,33,27,20	0.2632	0.5404	0.4964	0.8858
1,3,5,7	30,30,30,30	0.0590	0.1492	0.1632	0.4052
	20,27,33,40	0.0360	0.1898	0.2156	0.3844
	40,33,27,20	0.1050	0.1248	0.1114	0.3706

Note: F=ANOVA F test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data.

Table 22.

Empirical Power with All Distributions Heavily Positively Skewed, N = 210 and K = 7.

Variance Ratio	*n*1,*n*2,*n*3,*n*4,*n*5,*n*6,*n*7	F	t_y	$WJ_{t(Jn)}$	W_r
1,1,1,1,1,1,1	30,30,30,30,30,30,30	0.2442	0.6222	0.6256	0.6868
	20,24,27,30,33,36,40	0.2012	0.5988	0.6310	0.6828
	40,36,33,30,27,24,20	0.2570	0.5844	0.4810	0.6290
1,2,3,4,5,6,7	30,30,30,30,30,30,30	0.0698	0.1428	0.1428	0.5600
	20,24,27,30,33,36,40	0.0374	0.1566	0.1924	0.4816
	40,36,33,30,27,24,20	0.1188	0.1230	0.0758	0.5864

Note: F=ANOVA F test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data.

Table 23.

Empirical Type I Error with Two Normal and Two Heavily Positively Skewed Distributions, N = 120 and K = 4.

Variance Ratio	*n*1,*n*2,*n*3,*n*4	*F*	t_y	$WJ_{t(Jn)}$	W_r
1,1,1,1	30,30,30,30	0.0746	0.0580	0.0536	0.0564
	20,27,33,40	0.0508	0.0554	0.0616	0.0574
	40,33,27,20	**0.1068**	0.0654	0.0468	0.0588
1,3,5,7	30,30,30,30	**0.0864**	0.0658	0.0584	0.0606
	20,27,33,40	0.0502	0.0568	0.0590	0.0600
	40,33,27,20	**0.1664**	0.0716	0.0504	0.0620

Note: *F*=ANOVA *F* test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data. Bolded values reflect empirical Type I error outside the acceptable criterion (2.5% - 7.5%)

Table 24.

Empirical Type I Error with Three Normal and Four Heavily Positively Skewed Distributions, N = 210 and K = 7.

Variance Ratio	*n*1,*n*2,*n*3,*n*4,*n*5,*n*6,*n*7	*F*	t_y	$WJ_{t(Jn)}$	W_r
1,1,1,1,1,1,1	30,30,30,30,30,30,30	0.0650	0.0710	0.0600	0.0596
	20,24,27,30,33,36,40	0.0410	0.0650	0.0712	0.0620
	40,36,33,30,27,24,20	**0.0946**	0.0730	0.0372	0.0608
1,2,3,4,5,6,7	30,30,30,30,30,30,30	**0.0808**	**0.0760**	0.0632	0.0582
	20,24,27,30,33,36,40	**0.0141**	0.0652	0.0636	0.0620
	40,36,33,30,27,24,20	**0.1402**	**0.0758**	0.0428	0.0572

Note: *F*=ANOVA *F* test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data. Bolded values reflect empirical Type I error outside the acceptable criterion (2.5% - 7.5%)

Table 25.

Empirical Power with Two Normal and Two Heavily Positively Skewed Distributions, N = 120 and K = 4.

Variance Ratio	*n*1,*n*2,*n*3,*n*4	F	t_y	$WJ_{t(Jn)}$	W_r
1,1,1,1	20,27,33,40	0.2322	0.6234	0.6800	0.7908
	40,33,27,20	0.3996	0.5924	0.5692	0.8102
1,3,5,7	30,30,30,30	0.0746	0.1580	0.1688	0.3242
	20,27,33,40	0.0322	0.1848	0.2186	0.3182
	40,33,27,20	0.1472	0.1322	0.1132	0.2872

Note: F=ANOVA F test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data.

Table 26.

Empirical Power with Three Normal and Four Heavily Positively Skewed Distributions, N = 210 and K = 7.

Variance Ratio	*n*1,*n*2,*n*3,*n*4,*n*5,*n*6,*n*7	F	t_y	$WJ_{t(Jn)}$	W_r
1,1,1,1,1,1,1	30,30,30,30,30,30,30	0.2858	0.6118	0.6204	0.8362
	20,24,27,30,33,36,40	0.2014	0.6008	0.6598	0.7978
	40,36,33,30,27,24,20	0.3590	0.5938	0.4956	0.8118
1,2,3,4,5,6,7	30,30,30,30,30,30,30	0.0764	0.1366	0.1388	0.2930
	20,24,27,30,33,36,40	0.0444	0.1614	0.1934	0.2896
	40,36,33,30,27,24,20	0.1478	0.1426	0.0902	0.2880

Note: F=ANOVA F test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data.

Table 27.

Empirical Type I Error with One Normal, Two Heavily Positively Skewed and One Heavily Negatively Skewed Distributions, N = 120 and K = 4.

Variance Ratio	$n1,n2,n3,n4$	F	t_y	$WJ_{t(Jn)}$	W_r
1,1,1,1	30,30,30,30	0.0610	0.0596	0.0620	0.0606
	20,27,33,40	0.0732	0.0560	0.0512	0.0546
	40,33,27,20	**0.0978**	0.0682	0.0496	0.0584
1,3,5,7	30,30,30,30	**0.0974**	0.0734	0.0646	0.0622
	20,27,33,40	0.0704	0.0634	0.0642	0.0592
	40,33,27,20	**0.1372**	0.0728	0.0522	0.0608

Note: F=ANOVA F test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data. Bolded values reflect empirical Type I error outside the acceptable criterion (2.5% - 7.5%).

Table 28.

Empirical Type I Error with Three Normal, Three Heavily Positively Skewed and Two Heavily Negatively Skewed Distributions, N = 210 and K = 7.

Variance Ratio	$n1,n2,n3,n4,n5,n6,n7$	F	t_y	$WJ_{t(Jn)}$	W_r
1,1,1,1,1,1,1	30,30,30,30,30,30,30	0.0664	**0.0758**	0.0616	0.0542
	20,24,27,30,33,36,40	0.0522	0.0698	0.0702	0.0572
	40,36,33,30,27,24,20	**0.0868**	0.0706	0.0366	0.0602
1,2,3,4,5,6,7	30,30,30,30,30,30,30	**0.0830**	**0.0820**	0.0738	0.0630
	20,24,27,30,33,36,40	0.0528	**0.0790**	**0.0778**	0.0622
	40,36,33,30,27,24,20	**0.1364**	**0.0880**	0.0492	0.0650

Note: F=ANOVA F test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data. Bolded values reflect empirical Type I error outside the acceptable criterion (2.5% - 7.5%).

Table 29.

Empirical Power with One Normal, Two Heavily Positively Skewed, and One Heavily Negatively Skewed Distributions, N = 120 and K = 4.

Variance Ratio	*n*1,*n*2,*n*3,*n*4	F	t_y	$WJ_{t(Jn)}$	W_r
1,1,1,1	30,30,30,30	0.3688	0.6282	0.6102	0.6666
	20,27,33,40	0.3248	0.5980	0.6060	0.6156
	40,33,27,20	0.3914	0.5996	0.5446	0.6502
1,3,5,7	30,30,30,30	0.1748	0.2516	0.2360	0.2846
	20,27,33,40	0.1347	0.2544	0.2550	0.2894
	40,33,27,20	0.2274	0.2430	0.1984	0.2488

Note: F=ANOVA F test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data.

Table 30.

Empirical Power with Three Normal, Three Heavily Positively Skewed and Two Heavily Negatively Skewed Distributions, N = 210 and K = 7.

Variance Ratio	*n*1,*n*2,*n*3,*n*4,*n*5,*n*6,*n*7	F	t_y	$WJ_{t(Jn)}$	W_r
1,1,1,1,1,1,1	30,30,30,30,30,30,30	0.3500	0.6708	0.6272	0.6932
	20,24,27,30,33,36,40	0.3072	0.6360	0.6326	0.6496
	40,36,33,30,27,24,20	0.3842	0.6222	0.4964	0.6746
1,2,3,4,5,6,7	30,30,30,30,30,30,30	0.1490	0.2710	0.2394	0.2870
	20,24,27,30,33,36,40	0.1120	0.2752	0.2704	0.2982
	40,36,33,30,27,24,20	0.2094	0.2532	0.1664	0.2594

Note: F=ANOVA F test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data.

Table 31.

Empirical Type I Error with One Moderately Positively Skewed and Three Heavily Positively Skewed Distributions, N = 120 and K = 4.

Variance Ratio	$n1,n2,n3,n4$	F	t_y	$WJ_{t(Jn)}$	W_r
	30,30,30,30	0.0488	0.0490	0.0490	0.0524
1,1,1,1	20,27,33,40	0.0376	0.0474	0.0584	0.0502
	40,33,27,20	0.0590	0.0502	0.0384	0.0520
	30,30,30,30	0.0666	0.0656	0.0604	0.0600
1,3,5,7	20,27,33,40	0.0332	0.0514	0.0552	0.0542
	40,33,27,20	**0.1146**	0.0712	0.0564	0.0592

Note: F=ANOVA F test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data. Bolded values reflect empirical Type I error outside the acceptable criterion (2.5% - 7.5%).

Table 32.

Empirical Type I Error with Three Moderately Positively and Four Heavily Positively Skewed Distributions, N = 210 and K = 7.

Variance Ratio	$n1,n2,n3,n4,n5,n6,n7$	F	t_y	$WJ_{t(Jn)}$	W_r
	30,30,30,30,30,30,30	0.0542	0.0590	0.0510	0.0524
1,1,1,1,1,1,1	20,24,27,30,33,36,40	0.0404	0.0588	0.0688	0.0562
	40,36,33,30,27,24,20	**0.0752**	0.0650	0.0354	0.0554
	30,30,30,30,30,30,30	0.0736	0.0666	0.0574	0.0526
1,2,3,4,5,6,7	20,24,27,30,33,36,40	0.0398	0.0662	0.0692	0.0534
	40,36,33,30,27,24,20	**0.1306**	0.0730	0.0392	0.0564

Note: F=ANOVA F test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data. Bolded values reflect empirical Type I error outside the acceptable criterion (2.5% - 7.5%).

Table 33.

Empirical Power with One Moderately Positively Skewed and Three Heavily Positively Skewed Distributions, N = 120 and K = 4.

Variance Ratio	*n*1,*n*2,*n*3,*n*4	*F*	t_y	$WJ_{t(Jn)}$	W_r
	30,30,30,30	0.2834	0.6398	0.6492	0.8758
1,1,1,1	20,27,33,40	0.2232	0.6036	0.6298	0.8210
	40,33,27,20	0.295	0.5786	0.5442	0.8754
	30,30,30,30	0.0680	0.1416	0.1574	0.4070
1,3,5,7	20,27,33,40	0.0342	0.1852	0.2196	0.4050
	40,33,27,20	0.1172	0.1262	0.1096	0.3758

Note: *F*=ANOVA *F* test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data.

Table 34.

Empirical Power with Three Moderately Positively and Four Heavily Positively Skewed Distributions, N = 210 and K = 7.

Variance Ratio	*n*1,*n*2,*n*3,*n*4,*n*5,*n*6,*n*7	*F*	t_y	$WJ_{t(Jn)}$	W_r
	30,30,30,30,30,30,30	0.2724	0.6266	0.6202	0.8742
1,1,1,1,1,1,1	20,24,27,30,33,36,40	0.2056	0.6094	0.6436	0.8376
	40,36,33,30,27,24,20	0.3262	0.5894	0.4876	0.8714
	30,30,30,30,30,30,30	0.0762	0.1410	0.1376	0.3324
1,2,3,4,5,6,7	20,24,27,30,33,36,40	0.0382	0.1778	0.2072	0.3492
	40,36,33,30,27,24,20	0.1382	0.1372	0.0826	0.3242

Note: *F*=ANOVA *F* test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data.

Table 35.

Empirical Type I Error with Two Heavily Positively Skewed and Two Heavily Negatively Skewed Distributions, N = 120 and K = 4.

Variance Ratio	*n*1,*n*2,*n*3,*n*4	F	t_y	$WJ_{t(Jn)}$	W_r
1,1,1,1	30,30,30,30	0.0622	0.0634	0.0576	0.0550
	20,27,33,40	0.0626	0.0644	0.0664	0.0636
	40,33,27,20	0.0654	0.0578	0.041	0.0552
1,3,5,7	30,30,30,30	**0.0832**	0.0688	0.0618	0.0662
	20,27,33,40	0.0476	0.0574	0.0566	0.0628
	40,33,27,20	**0.1208**	**0.0782**	0.0552	0.0636

Note: F=ANOVA F test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data. Bolded values reflect empirical Type I error outside the acceptable criterion (2.5% - 7.5%).

Table 36.

Empirical Type I Error with Three Heavily Positively and Four Heavily Negatively Skewed Distributions, N = 210 and K = 7.

Variance Ratio	*n*1,*n*2,*n*3,*n*4,*n*5,*n*6,*n*7	F	t_y	$WJ_{t(Jn)}$	W_r
1,1,1,1,1,1,1	30,30,30,30,30,30,30	0.0562	**0.0800**	0.0628	0.0560
	20,24,27,30,33,36,40	0.0496	**0.0808**	**0.0778**	0.0654
	40,36,33,30,27,24,20	0.0536	**0.085**	0.0424	0.0646
1,2,3,4,5,6,7	30,30,30,30,30,30,30	0.0600	**0.0808**	0.0640	0.0658
	20,24,27,30,33,36,40	0.0400	**0.0824**	**0.0784**	0.0706
	40,36,33,30,27,24,20	**0.0926**	**0.0932**	0.0482	0.0658

Note: F=ANOVA F test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data. Bolded values reflect empirical Type I error outside the acceptable criterion (2.5% - 7.5%).

Table 37.

Empirical Power with Two Heavily Positively Skewed and Two Heavily Negatively Skewed Distributions, N = 120 and K = 4.

Variance Ratio	$n1,n2,n3,n4$	F	t_y	$WJ_{t(Jn)}$	W_r
1,1,1,1	30,30,30,30	0.3364	0.5990	0.5566	0.4948
	20,27,33,40	0.3266	0.5760	0.5610	0.4818
	40,33,27,20	0.3354	0.5760	0.4984	0.4904
1,3,5,7	30,30,30,30	0.1822	0.2880	0.258	0.2408
	20,27,33,40	0.1312	0.2690	0.2528	0.2402
	40,33,27,20	0.2226	0.2718	0.2118	0.2184

Note: F=ANOVA F test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data.

Table 38.

Empirical Power with Three Heavily Positively and Four Heavily Negatively Skewed Distributions, N = 210 and K = 7.

Variance Ratio	$n1,n2,n3,n4,n5,n6,n7$	F	t_y	$WJ_{t(Jn)}$	W_r
1,1,1,1,1,1,1	30,30,30,30,30,30,30	0.3060	0.6772	0.6158	0.5450
	20,24,27,30,33,36,40	0.3016	0.6284	0.5958	0.5414
	40,36,33,30,27,24,20	0.2908	0.6246	0.4944	0.5274
1,2,3,4,5,6,7	30,30,30,30,30,30,30	0.1332	0.3356	0.2800	0.2354
	20,24,27,30,33,36,40	0.1016	0.3406	0.3130	0.2624
	40,36,33,30,27,24,20	0.1826	0.3202	0.1986	0.2328

Note: F=ANOVA F test; t_y=Yuen (trimmed), $WJ_{t(Jn)}$=Welch-James with Johnson transformation (trimmed); W_r=Welch on ranked data.