CrossMark

Effect of variance ratio on ANOVA robustness: Might 1.5 be the limit?

María J. Blanca[1] · Rafael Alarcón[1] · Jaume Arnau[2] · Roser Bono[2] · Rebecca Bendayan[1,3]

© Psychonomic Society, Inc. 2017

Abstract Inconsistencies in the research findings on *F*-test robustness to variance heterogeneity could be related to the lack of a standard criterion to assess robustness or to the different measures used to quantify heterogeneity. In the present paper we use Monte Carlo simulation to systematically examine the Type I error rate of *F*-test under heterogeneity. One-way, balanced, and unbalanced designs with monotonic patterns of variance were considered. Variance ratio (VR) was used as a measure of heterogeneity (1.5, 1.6, 1.7, 1.8, 2, 3, 5, and 9), the coefficient of sample size variation as a measure of inequality between group sizes (0.16, 0.33, and 0.50), and the correlation between variance and group size as an indicator of the pairing between them (1, .50, 0, −.50, and −1). Overall, the results suggest that in terms of Type I error a VR above 1.5 may be established as a rule of thumb for considering a potential threat to *F*-test robustness under heterogeneity with unequal sample sizes.

One-way analysis of variance (ANOVA) is one of the most common statistical techniques to test the equality of three or more means in educational and behavioral research (Keselman et al., 1998; Kieffer, Reese, & Thompson, 2001), although its use has decreased in recent years (Skidmore & Thompson, 2010). The *F*-test assumes that the outcome variable must be normally and independently distributed, and the samples must come from a population with common variances. However, the empirical evidence involving real data extracted from review of several scientific journals indicates that these assumptions are not always met (Blanca, Arnau, López-Montiel, Bono, & Bendayan, 2013; Micceri, 1989; Ruscio & Roche, 2012).

Specifically, with regard to homogeneity of variance, research reveals that group variances are often unequal (Erceg-Hurn & Miroservich, 2008; Grissom, 2000; Keselman et al., 1998; Ruscio & Roche, 2012; Wilcox, 1987). This inequality may be due to *a priori* differences in groups that are naturally formed or to an effect of experimental treatment that produces differences not only in means but also in variances (Bryk & Raudenbush 1988; Erceg-Hurn & Mirosevich, 2008; Grissom, 2000; Grissom & Kim, 2001; Sawilowsky, 2002; Sawilowsky & Blair, 1992). Several indexes have been proposed to measure the amount of heterogeneity, namely the coefficient of variance variation (Box, 1954; Rogan, Keselman, & Breen, 1977; Ruscio & Roche, 2012), the standardized variance heterogeneity index (Ruscio & Roche, 2012), and the variance ratio (Keselman et al., 1998; Ruscio & Roche, 2012). The variance ratio, which is the simplest measure of heterogeneity, is defined as the ratio of the largest variance to the smallest variance of the groups. This is the index most commonly used in Monte Carlo studies (e.g., Box, 1954; Cribbie, Wilcox, Bewell, & Keselman, 2007; Fan & Hancock, 2012; Hsu, 1938; Kang, Harring, & Li, 2015; Mendes & Pala, 2004; Mickelson, 2013; Moder, 2007, 2010; Scheffé, 1959; Tomarken & Serling, 1986; Wilcox, Charlin, &

✉ María J. Blanca
blamen@uma.es

[1] Department of Psychobiology and Behavioral Science Methodology, University of Malaga, Malaga, Spain

[2] Department of Social Psychology and Quantitative Psychology, University of Barcelona, Barcelona, Spain

[3] MRC Unit for Lifelong Health and Ageing, London, UK

 Springer

Thompson, 1986; Zijlstra, 2004). With real data, Keselman et al. (1998) found that the average value of the variance ratio was 2.0 (*SD* = 2.6), with a median of 1.5 and a maximum ratio of 23.8. Recently, Ruscio and Roche (2012) found variance heterogeneity in more than 50% of examined cases, with the mean variance ratio being equal to 2.51 when there were two groups, 3.95 when there were three groups, and 8.84 when there were four groups in the design.

As the abovementioned studies show, variance heterogeneity is frequently observed in real data. The question that follows logically from this is how heterogeneity affects the robustness of the *F*-test. Robustness, which has been extensively addressed in the literature, refers to a statistical test's insensitivity under violations of its assumptions, specifically in terms of its Type I error rates (Box, 1953). Type I error is the probability of rejecting a null hypothesis when it is actually true. The robustness of a statistical test can be evaluated via Monte Carlo simulation techniques, and in order to ensure the comparability of results from Monte Carlo studies a standard criterion to assess robustness must be established. Bradley's (1978) liberal criterion is considered the most appropriate (e.g., Keselman, Algina, Kowalchuk, & Wolfinger, 1999; Kowalchuk, Keselman, Algina, & Wolfinger, 2004). According to this criterion, a statistical test is considered robust if the empirical Type I error rate is between .025 and .075 for a nominal alpha level of .05. When the rate is above .075 the test is considered liberal, increasing the risk of declaring mean differences that do not exist. When the rate is below .025 the test is considered conservative, such that the researcher is assuming an alpha level below the nominal.

The first Monte Carlo studies that examined *F*-test robustness to violations of its assumptions were carried out between 1930 and 1960 and were summarized by Glass, Peckham, and Sanders (1972). With regard to variance heterogeneity, early studies (Box, 1954; David & Johnson, 1951; Horsnell, 1953; Norton, 1952, cit. Lindquist, 1953; Hsu, 1938; Scheffé, 1959) suggest two main conclusions: (1) *F*-test is robust when the groups have equal sample sizes and the group size is not very small (e.g., greater than 7; Kohr & Games, 1974); and (2) *F*-test tends not to be robust when the groups have unequal sample sizes, in which case the effect of heterogeneity on Type I error depends on the pairing of variance with group size. *F*-test tends to be conservative when the pairing is positive, that is, when the group with the largest sample size also has the largest variance and the group with the smallest sample size has the smallest variance. Conversely, it tends to be liberal when the pairing is negative, namely when the group with the largest sample size has the smallest variance and the group with the smallest variance has the largest variance. Based on these studies, many classical handbooks on research methods in education and psychology recommend using equal sample sizes as protection against the effect of heterogeneity (e.g., Glass & Stanley, 1970; Hays, 1981; Keppel, 1991; Maxwell & Delaney, 1990; Winner, 1971).

The issue of *F*-test robustness to variance heterogeneity has continued to be studied since 1970 until the present day (for a review, see Harwell, Rubinstein, Hayes, & Olds, 1992; Lix, Keselman, & Keselman, 1996). However, research to date with *equal sample sizes* provides contradictory results, there being both evidence that *F*-test is robust to variance heterogeneity (Lee & Ahn, 2003; Patrick, 2007; Yiğit & Gökpinar, 2010) and evidence against this (Alexander & Govern, 1994; Büning, 1997; Harwell et al., 1992; Lix et al., 1996; Moder, 2010; Rogan & Keselman, 1977; Tomarken & Serling, 1986; Wilcox et al., 1986). This inconsistency in the results may be due to several factors.

First, most of the cited studies did not use a standard criterion to assess robustness. Results were usually interpreted based on the comparison between empirical and nominal alpha without following any standard criterion: If the difference was small, *F*-test was said to be robust. The problem here is that the meaning of "small" is ambiguous and does not allow a clear decision to be made. Indeed, expressions such as "modest inflation" (Harwell et al., 1992) or "slightly increase" (Glass et al., 1972) are frequently used when referring to Type I error rates. Had Bradley's criterion of robustness been adopted, many of these results would have been interpreted differently.

Second, the studies in question used different measures to quantify variance heterogeneity, thus making it difficult to draw general conclusions. Some studies used the coefficient of variance variation (Lix et al., 1996; Rogan & Keselman, 1977), some used their own indexes (e.g., Patrick, 2007; Ruscio & Roche, 2012), and others used the variance ratio (e.g., Alexander & Govern, 1994; Box, 1954; Hsu, 1938; Moder, 2010; Scheffé, 1959; Tomarken & Serling, 1986; Wilcox et al., 1986; Zijlstra, 2004).

Third, the simulated conditions (e.g., variance values, number of groups, group sizes, pattern of variance, number of replications, etc.) were so varied that it is almost impossible to compare studies. In this context, the pattern of heterogeneity that is simulated appears to be the most relevant variable. The pattern of heterogeneity refers to the way in which the values of the group variances can be ordered. Thus, the group variances can monotonically increase (e.g., $\sigma_1^2 > \sigma_2^2 > \sigma_3^2$) or decrease (e.g., $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$) or follow another arbitrary pattern (e.g., $\sigma_1^2 = \sigma_2^2 > \sigma_3^2$). Research to date has included a wide variety of these patterns. In general, some studies have found that *F*-test is robust, according to Bradley's liberal criterion, with a monotonic pattern (Lee & Ahn, 2003; Tomarken & Serling, 1986; Wilcox et al., 1986), whereas others have found that it is liberal (Alexander & Govern, 1994; Büning, 1997). For example, Wilcox et al. (1986), who considered four

groups with a variance ratio equal to 4 and a monotonic pattern of variance of 1: 2: 3: 4 with equal sample sizes ($n = 11$), found that F-test was robust (Type I error rate = .068), whereas Alexander and Govern (1994) found it to be liberal with a pattern of 1: 2: 4: 6 (Type I error rate = .079). Büning (1997) found that F-test was robust with group size equal to 10 and a pattern of 1: 2: 4 (Type I error rate = .062), but liberal with a pattern of 1: 3: 7 (Type I error rate = .083). With arbitrary patterns of heterogeneity involving a set of groups with similar variances and one with extreme variance (e.g., 1: 1: 1: 6 and 1: 1: 30) the test has been found to be non-robust (Alexander & Govern, 1994; Lee & Ahn, 2003; Moder, 2010; Rogan & Keselman; 1977; Wilcox et al., 1986). Overall, these findings suggest that F-test robustness with equal group sizes is more affected by a pattern where the variance of one group is very different to that of the other groups. However, F-test robustness with monotonic patterns of variance is still unclear, and further research is needed to determine under which types of these patterns the test can be used.

The sensitivity of F-test to violations of the variance homogeneity assumption when *sample sizes are unequal* has been reported more consistently (Gamage & Weerahandi, 1998; Kohr & Games, 1974; Lee & Ahn, 2003; Moder, 2010; Patrick, 2007; Tomarken & Serling, 1986; Yiğit & Gökpinar, 2010; Zijlstra, 2004). The empirical evidence indicates that its robustness depends on the pairing of variance with group size, as was found in early studies. However, despite the large body of research the specific conditions under which F-test is robust have yet to be established, and a number of questions remained unanswered. For example, what values of the variance ratio are associated with correct/invalid inferences? How much inequality of group sizes can be assumed in order to ensure that F-test controls Type I error rate? What other types of pairing between variance and group size can be defined and how do they affect F-test robustness?

Regarding the first and second questions, some authors have suggested several rules of thumb, namely that variance homogeneity can probably be assumed when the variance ratio is not greater than 3 (Dean & Voss, 1999; Keppel, Saufley, & Tokunaga, 1992; Kirk, 2013), is less than 4 or 5 (Wuensch, 2017), or is even as high as 10 provided that the ratio of the largest to smallest sample size does not exceed 4 (Tabachnick & Fidell, 2007; 2013).

Regarding the pairing between variance and group size, previous Monte Carlo studies have usually included a perfect pairing with monotonic patterns of variance. For example, considering five groups with sample sizes equal to 32, 36, 40, 44, and 48 and variances equal to 1, 2, 3, 4, and 5, respectively, the pairing between these variables is perfect and positive. If the variances were 5, 4, 3, 2, and 1, respectively, the pairing would be perfect and negative. However, other types of pairing are also possible. If the pairing is defined by the correlation between group size and variance, then different

values of this variable can be obtained. For example, if the same group sample sizes were associated with variance values of 1, 4, 2, 5, and 3, respectively, the pairing would be equal to .50, while for values of 3, 5, 2, 4, and 1 it would be equal to −.50. Thus, different values of the pairing could be considered in Monte Carlo studies in order to extend our understanding of how F-test robustness is affected by the type of pairing. As mentioned, previous research does not provide consistent results about the robustness of F-test with monotonic patterns, and it does not consider other possible types of pairing.

In this context, the main aim of this study is to systematically examine the robustness of F-test, in terms of Type I error, to violations of variance heterogeneity, considering a wide range of conditions representative of real data in educational and psychological research (Golinski & Cribbie, 2009; Keselman et al., 1998; Ruscio & Roche, 2012). To this end, a series of Monte Carlo simulation studies are performed for a one-way design with equal and unequal sample sizes and monotonic patterns of variance. The variance ratio is used as a measure of heterogeneity, the coefficient of sample size variation as a measure of the amount of inequality in group size, and the correlation between variance and group sample size as an indicator of different values of pairing. Our goal, based on the results of this study, is to offer a guideline to help applied researchers decide whether they can use the F-test when their data do not meet the variance homogeneity assumption under certain conditions.

Method

With the aim of systematically examining the robustness of F-test to violations of variance heterogeneity we conducted a series of Monte Carlo simulation studies for a one-way design with equal and unequal sample sizes and monotonic patterns of variance. Simulation studies use computer-intensive procedures to assess the appropriateness and accuracy of a variety of statistical methods in relation to the known truth (Angelis & Young, 1998), and they are especially suitable for evaluating a test's robustness when the underlying assumptions are not fulfilled. For this reason, they are widely used by researchers in the health and social sciences (Burton, Altman, Royston, & Holder, 2006).

In order to examine the isolated effects of variance heterogeneity on F-test robustness, and considering a one-way design, data were assumed to be normally distributed. Normal data were generated using a series of macros created ad hoc in SAS 9.4 (SAS Institute, 2013). The group effect was set to zero in the population model. The following variables were manipulated:

1. Equal and unequal group sample sizes and number of groups. Data analytic practices for ANOVA show that unbalanced designs are more common than balanced

designs (Golinski & Cribbie, 2009; Keselman et al., 1998). We considered designs with three, four, five, and six groups with balanced cells, and three and five groups with unbalanced cells.

2. Group sample size and total sample size. A wide range of group sample sizes which can be frequently found in real research were considered, enabling us to study small, medium, and large sample sizes. With balanced designs, the group sizes were set to three, five, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, and 100. With unbalanced designs, group sizes were set between three and 170, with a mean group size from five to 100. Total sample size ranged from nine to 600, depending on the number of groups considered, this being the result of multiplying the number of groups by the minimum and maximum group sample size (e.g., with five groups the total sample size ranged from 15 to 500).

3. Coefficient of sample size variation (Δn), which represents the amount of inequality in group sizes. This was computed by dividing the standard deviation of the group sample size by its mean. Different degrees of variation were considered and were grouped as low, medium, and high. A low Δn was fixed at approximately 0.16 (0.141–0.178), a medium coefficient at 0.33 (0.316–0.334), and a high value at 0.50 (0.491–0.521). Keselman et al. (1998) showed that the ratio of the largest to the smallest group size was greater than 3 in 43.5% of cases. With $\Delta n = 0.16$ this ratio was equal to 1.5, with $\Delta n = 0.33$ it was equal to either 2.3 or 2.5, and with $\Delta n = 0.50$ it ranged from 3.3 to 5.7.

4. Ratio of the largest to the smallest variance. For one-way designs Keselman et al. (1998) found that the average value of the ratio of the largest to the smallest standard deviation was 2.0 ($SD = 2.6$), with a median of 1.5 and a maximum ratio of 23.8. Ruscio and Roche (2012), despite the enormous range in variance ratio found, showed that the ratio exceeded 3 in 23.18% of reviewed studies, with a median of 1.64 and a range for the middle 50% of cases of between 1.23 and 2.76. In addition, for three groups they found a mean value of 3.95. Based on these findings, the values of variance ratio selected for the present study were 1.5, 2, 3, 5, and 9 for balanced designs, and 1.5, 1.6, 1.7, 1.8, 2, 3, 5 and 9 for unbalanced designs.

5. Patterns of variance and pairing of variance with group sample size. Monotonic patterns of variance were considered, and are presented in Table 1 for each variance ratio. The type of pairing between variance and group size indicates the relationship or association between the two. Pairing is positive when the largest group size is associated with the largest value of the variance and the smallest group size is associated with the smallest value of variance. Pairing is negative when the largest group size is associated with the smallest value of variance, and vice-versa. Unpairing occurs when there is no association between group size and variance. This happens, for example, with

Table 1 Patterns of variance considered in relation to variance ratio and number of groups

Variance ratio	Number of groups	Variance pattern
1.5	3	1: 1.25: 1.5
	4	1: 1.17: 1.34: 1.5
	5	1: 1.125: 1.25: 1.375: 1.5
	6	1: 1.1: 1.2: 1.3: 1.4: 1.5
1.6	3	1: 1.3: 1.6
	5	1: 1.15: 1.3: 1.45: 1.6
1.7	3	1: 1.35: 1.7
	5	1: 1.175: 1.35: 1.525: 1.7
1.8	3	1: 1.4: 1.8
	5	1: 1.2: 1.4: 1.6: 1.8
2	3	1: 1.5: 2
	4	1: 1.33: 1.66: 2
	5	1: 1.25: 1.5: 1.75: 2
	6	1: 1.2: 1.4: 1.6: 1.8: 2
3	3	1: 1.5: 3
	4	1: 1.66: 2.33: 3
	5	1: 1.5: 2: 2.5: 3
	6	1: 1.4: 1.8: 2.2: 2.6: 3
5	3	1: 3: 5
	4	1: 2.33: 3.66: 5
	5	1: 2: 3: 4: 5
	6	1: 1.8: 2.6: 3.4: 4.2: 5
9	3	1: 5: 9
	4	1: 3.66: 6.32: 9
	5	1: 3: 5: 7: 9
	6	1: 2.6: 4.2: 5.8: 7.4: 9

equal group sizes and/or equal variances, but it can also appear with unequal group sizes. In order to consider conditions of pairing which can represent real data (Keselman et al., 1998; Ruscio & Roche, 2012), we calculated the correlation between group sample size and variance value. Correlations equal, approximately, to 1, .50, 0, −.50, and −1 were considered for unbalanced designs. The value of 0 was not included for three groups because it is a non-possible value. These correlation values were obtained by associating each group sample size with different values of variance for the monotonic pattern. Thus, if the groups are ordered as a function of their sample sizes, different values of this correlation are obtained by changing the value of their variance. Table 2 shows the order of variance associated with the group sample sizes, from the smallest sample size to the largest one.

To ensure reliable results 10,000 replications of each combination of the above conditions were performed at a significance level of .05, recording the empirical Type I error rate

Table 2 Order of variance associated with the groups according to pairing and number of groups

Pairing	J = 3	J = 5
1	1, 2, 3	1, 2, 3, 4, 5
.50	1, 3, 2	1, 4, 2, 5, 3
0	-	1, 5, 4, 2, 3
-.50	2, 3, 1	3, 5, 2, 4, 1
-1	3, 2, 1	5, 4, 3, 2, 1

Note. J = number of groups

(Bendayan, Arnau, Blanca, & Bono, 2014; Robey & Barcikowski, 1992).

Results

The empirical Type I error rates associated with the *F*-test of the group effect were analyzed for each combination. The results for equal and unequal sample size are shown in Appendices 1 and 2, respectively. Bradley's liberal criterion (1978) was used to assess the robustness of the procedure. To summarize the results, based on Bradley's criterion the empirical Type I error rates were dichotomized into a binary variable with two categories, robust (Type I error rate between .025 and .075) and not robust (Type I error rate below .025 or above .075). Chi-square tests were then performed to examine the association between robustness and the variables of interest. Results are presented according to equal and unequal sample sizes.

Equal sample sizes

As can be seen in Tables 11, 12, 13 and 14 (Appendix 1), all Type I error rates were inside the boundary of Bradley's liberal criterion. Thus, the results show that *F*-test is robust for three, four, five, and six groups in 100% of cases, regardless of the total sample size and variance ratio.

Unequal sample sizes

Total sample size The association between total sample size and categorical Type I error rate was not statistically significant for any condition of variance ratio and number of groups, considering 13 categories of the first variable and two of the second. Moreover, the association between group sample size mean and categorical Type I error rate, collapsed across all variance ratios, was not significant for either three groups, $\chi^2(12) = 1.47$, $p = .99$, or five groups, $\chi^2(12) = 0.38$, $p = .99$. The percentages of *F*-test robustness are shown in Table 3.

Variance ratio The relationship between variance ratio (eight categories) and categorical Type I error rate was significant

Table 3 Percentage of *F*-test robustness according to group sample size mean and number of groups

N/J	J = 3	J = 5
5	77.5	85.0
10	79.2	85.0
15	78.1	85.8
20	81.3	85.8
25	81.3	85.8
30	80.2	84.2
40	82.3	84.2
50	79.2	85.0
60	80.2	85.0
70	80.2	85.0
80	82.3	85.0
90	80.2	85.8
100	79.2	85.0

Note. N/J = mean of the group sample size; *J* = number of groups

both for three groups, $\chi^2(7) = 283.59$, $p < .001$, and for five groups, $\chi^2(7) = 288.57$, $p < .001$. In general, the percentage of robustness decreased as variance ratio increased, with *F*-test being more robust with five groups. Table 4 shows the percentages of *F*-test robustness.

Pairing of variance with group size Overall, for both three and five groups there was a significant association between categorical Type I error rate and pairing of variance with group size for ratios higher than 1.5. Tables 5 and 6 show the percentage of robustness according to variance ratio and pairing. With a ratio of 1.5, *F*-test was robust in all conditions. With a ratio from 1.6 to 2 it was robust except when the pairing was equal to −1. With a ratio of 3 or higher, *F*-test was robust with pairing equal to 0 or .50 and non-robust with pairing equal to 1, −.5, and −1. Negative pairing had more of an effect than did positive pairing, with the percentage of robustness decreasing as the amount of

Table 4 Percentage of *F*-test robustness according to variance ratio and number of groups

Variance ratio	J = 3	J = 5
1.5	100	100
1.6	97.4	99.0
1.7	92.2	95.4
1.8	92.2	94.4
2	88.3	92.3
3	69.5	77.4
5	54.5	64.6
9	46.8	57.9

Note. J = number of groups. Robustness: Type I error rate within range [.025–.075]

Table 5 Percentage of *F*-test robustness for three groups according to the pairing of variance with group size and the variance ratio

Pairing	Variance ratio							
	1.5	1.6	1.7	1.8	2	3	5	9
1 [a]	100	100	100	100	100	74.4	64.1	59
.50	100	100	100	100	100	100	100	100
-.50 [b]	100	100	100	100	100	71.1	47.4	28.9
-1 [b]	100	89.7	69.2	69.2	53.8	33.3	7.7	0
χ^2		12.11**	38.37**	38.37**	60.10**	41.20**	68.42**	84.70**
df		3	3	3	3	3	3	3

Note. ** *p* < .01. Robustness: Type I error rate within range [.025–.075]. When *F*-test is not robust: [a] Conservative; [b] Liberal

negative pairing increased; it even reached zero with pairing equal to −1 and a variance ratio of 9. In addition, when *F*-test was not robust with positive pairing it was always conservative, whereas with negative pairing it was always liberal.

Coefficient of sample size variation For both three and five groups there was a significant association between categorical Type I error rate and the coefficient of sample size variation (three categories) for each ratio higher than 1.5, with *F*-test being less robust with the highest values of this coefficient. Tables 7 and 8 show the percentage of robustness according to this coefficient. For ratios of 2 or higher, the more inequality between groups the less robust *F*-test was. The largest coefficient of sample size variation had an enormous effect on the percentage of robustness when the variance ratio was 3 or higher, decreasing it by as much as three-quarters (to 24%) in the cases for three groups.

All studied conditions As can be seen in Tables 15 and 16 (Appendix 2) there was a similar pattern of Type I error rates for three and five groups. The results are summarized in Table 9. In general, it appears that robustness

depends on the variance ratio, the pairing of variance with group size, and the coefficient of sample size variation, with the procedure being more robust when variance ratios were small, the pairing of variance was either zero or positive, and the coefficient of sample size variation was smaller. More specifically:

- *Variance ratio of 1.5.* As stated above, *F*-test was robust for all the studied conditions, regardless of the pairing or the coefficient of sample size variation.
- *Variance ratio ranged from 1.6 to 1.8.* *F*-test was robust for all the considered conditions, except when the pairing was equal to −1 and the coefficient of sample size variation was equal to 0.50, in which case it tended to be liberal.
- *Variance ratio of 2.* *F*-test was robust for all the considered conditions, except when the pairing was equal to −1 and the coefficient of sample size variation was equal to 0.33 or 0.50; in both these cases it was liberal, and in the 0.50 condition it was liberal in 100% of cases.
- *Variance ratio of 3.* *F*-test was robust for all the considered conditions, except when the pairing of variance with group size was:

Table 6 Percentage of *F*-test robustness for five groups according to the pairing of variance with group size and the variance ratio

Pairing	Variance ratio							
	1.5	1.6	1.7	1.8	2	3	5	9
1 [a]	100	100	100	100	100	87.2	66.7	66.7
.50	100	100	100	100	100	100	100	100
0	100	100	100	100	100	100	100	100
-.50 [b]	100	100	100	100	100	66.7	35.9	23.1
-1 [b]	100	94.9	76.9	71.8	61.5	33.3	20.5	0
χ^2		8.08	37.74**	46.63**	65.01**	70.85**	90.03**	131.02**
df		4	4	4	4	4	4	4

Note. ** *p* < .01. Robustness: Type I error rate within range [.025–.075]. When *F*-test is not robust: [a] Conservative; [b] Liberal

Table 7 Percentage of *F*-test robustness for three groups according to the coefficient of sample size variation and the variance ratio

Δn	Variance ratio							
	1.5	1.6	1.7	1.8	2	3	5	9
0.16	100	100	100	100	100	100	78.8	71.2
0.33	100	100	100	100	90.4	75	59.6	44.2
0.50	100	92	76	76	74	32	24	24
χ^2		8.54*	27.07**	27.07**	17.02**	56.71**	31.74**	22.96**
df		2	2	2	2	2	2	2

Note. Δn = Coefficient of sample size variation; * $p < .05$; ** $p < .01$. Robustness: Type I error rate within range [.025–.075]

– Equal to 1 and the coefficient of sample size variation was equal to 0.50, in which case it tended to be conservative.
– Equal to −.5 and the coefficient of sample size variation was equal to 0.50, in which case it was liberal in almost 100% of cases.
– Equal to −1 and the coefficient of sample size variation was equal to 0.33 or 0.50, in which case it was liberal in 100% of the considered conditions.

• *Variance ratios of 5 and 9*. The pattern of results here was similar to that for a variance ratio of 3, although robustness decreased. Specifically, *F*-test was not robust when the pairing of variance with group size was:

– Equal to 1 and the coefficient of sample size variation was equal to 0.33 or 0.50, the test being conservative in the latter condition in 100% of cases.
– Equal to −.5 and the coefficient of sample size variation was equal to 0.33, in which case it was liberal in fewer than 50% of cases for a variance ratio of 5 and in 100% of them for a variance ratio of 9. When the coefficient of sample size variation was equal to 0.50, *F*-test was liberal in 100% of cases.

– Equal to −1 and the coefficient of sample size variation was equal to 0.16, 0.33, or 0.50, the test being liberal in the latter two conditions in 100% of cases.

Discussion

The aim of this paper was to present a systematic examination of *F*-test robustness, in terms of Type I error, to violations of variance heterogeneity with monotonic patterns of variance in one-way balanced and unbalanced designs. We used the variance ratio as a measure of heterogeneity, the coefficient of sample size variation as a measure of the amount of inequality in group size, and the correlation between variance and group sample size as an indicator of different values of pairing of variance with group sizes. The studied variables cover a wide range of conditions (2,972 conditions), our goal being to provide a guideline that would help applied researchers decide whether they can trust *F*-test results under heterogeneity. Several main conclusions can be drawn from the results.

Table 8 Percentage of *F*-test robustness for five groups according to the coefficient of sample size variation and the variance ratio

Δn	Variance ratio							
	1.5	1.6	1.7	1.8	2	3	5	9
0.16	100	100	100	100	100	100	90.8	73.8
0.33	100	100	100	100	96.9	80	63.1	60
0.50	100	96.9	86.2	83.1	80	52.3	40	40
χ^2		4.04	18.87**	23.32**	21.23**	42.67**	36.73**	15.48**
df		2	2	2	2	2	2	2

Note. Δn = Coefficient of sample size variation; ** $p < .01$. Robustness: Type I error rate within range [.025–.075]

Table 9 Percentage of F-test robustness according to the coefficient of sample size variation, the pairing of variance with group size, and the variance ratio

Pairing	Variance ratio								
	Δn	1.5	1.6	1.7	1.8	2	3	5	9
					J = 3				
1[a]	0.16	100	100	100	100	100	100	100	100
	0.33	100	100	100	100	100	100	92.3	76.9
	0.50	100	100	100	100	100	23.1	0	0
.50	0.16	100	100	100	100	100	100	100	100
	0.33	100	100	100	100	100	100	100	100
	0.50	100	100	100	100	100	100	100	100
-.50[b]	0.16	100	100	100	100	100	100	92.3	84.6
	0.33	100	100	100	100	100	100	46.2	0
	0.50	100	100	100	100	100	8.3	0	0
-1[b]	0.16	100	100	100	100	100	100	23.1	0
	0.33	100	100	100	100	61.5	0	0	0
	0.50	100	69.2	7.7	7.7	0	0	0	0
					J = 5				
1[a]	0.16	100	100	100	100	100	100	100	100
	0.33	100	100	100	100	100	100	100	100
	0.50	100	100	100	100	100	61.5	0	0
.50	0.16	100	100	100	100	100	100	100	100
	0.33	100	100	100	100	100	100	100	100
	0.50	100	100	100	100	100	100	100	100
0	0.16	100	100	100	100	100	100	100	100
	0.33	100	100	100	100	100	100	100	100
	0.50	100	100	100	100	100	100	100	100
-.50[b]	0.16	100	100	100	100	100	100	92.3	69.2
	0.33	100	100	100	100	100	100	15.4	0
	0.50	100	100	100	100	100	0	0	0
-1[b]	0.16	100	100	100	100	100	100	61.5	0
	0.33	100	100	100	100	84.6	0	0	0
	0.50	100	84.6	30.8	15.4	0	0	0	0

Note. Δn: Coefficient of sample size variation; *J* = number of groups; Robustness: Type I error rate within range [.025–.075]. When *F*-test is not robust: [a] Conservative, [b] Liberal

First, *F*-test is robust with monotonic patterns of variance when the group sample sizes are equal, regardless of the number of groups, of the ratio between the largest and smallest variance, and of the total sample size. With a variance ratio as large as 9, *F*-test can, at least for the number of groups and sample sizes considered here, still be used without the Type I error rate being affected by heterogeneity when the design is balanced.

Second, *F*-test is not robust with unequal sample sizes under certain conditions. The results showed that, in general, robustness depends on the variance ratio, the pairing

of variance with group size, and the coefficient of sample size variation, with the procedure being more robust when variance ratios are small, the pairing of variance is either zero or positive, and the coefficient of sample size variation is smaller. These conditions can be specified as follows:

1. The percentages of robustness tend to be lower for three groups than for five groups. This may indicate that the number of the groups is a variable that has to be considered: the smaller the number of groups, the greater the effect on *F*-test.
2. The total sample size does not influence *F*-test robustness under heterogeneity. The use of a large sample size does not, therefore, protect against the effect of heterogeneity.
3. When the pairing of variance with group size is equal to 0 for three groups and equal to 0 or .5 for five groups, *F*-test is not affected by heterogeneity under any considered condition. However, *F*-test tends to be conservative with positive pairing and liberal with negative pairing, the latter being the most influential variable. Consequently, researchers should pay particular attention when the pairing is negative in their data.
4. The ratio of the largest to the smallest variance, which represents the measure of heterogeneity, determines *F*-test robustness. Its robustness decreases as the variance ratio increases, in other words, robustness decreases as the homogeneity assumption is more violated. With a ratio of 1.5, *F*-test is robust in all studied conditions.
5. For a ratio higher than 1.5 there are two variables that have to be considered: The coefficient of sample size variation and the pairing of variance with group size. In general:

   • The coefficient of sample size variation, which represents the amount of inequality in group sizes, affects *F*-test robustness. In several cases its robustness decreases as the coefficient of variation increases, in other words, robustness decreases as the group sizes become more unequal.
   • When pairing is equal to 1, *F*-test tends to be conservative, whereas when pairing is negative (equal to −.5 or −1) the procedure tends to be liberal, depending on the variance ratio and the coefficient of sample size variation.
   • With a ratio higher than 1.5 and lower than 2, *F*-test is only affected by heterogeneity when pairing is equal

to −1 and the coefficient of sample size variation is 0.5.

- With a ratio equal to 2, *F*-test is only affected by heterogeneity when pairing is equal to −1 and the coefficient of sample size variation is as high as 0.33 or 0.5.
- With a ratio of 3 or higher, *F*-test tends to be conservative with pairing equal to 1 and a coefficient of sample size variation of 0.5. With a ratio of 5 or 9 it is conservative in 100% of the studied conditions. Likewise, *F*-test tends to be liberal with pairing equal to −.5 or −1 under several conditions of sample size variation. The more unequal the sample sizes, the less robust the *F*-test is.

In general, the results regarding equal sample sizes are consistent with early studies (e.g., Box, 1954; Glass et al., 1972; Hsu, 1938; Scheffé, 1959), as well as with more recent ones (Lee & Ahn, 2003; Patrick, 2007; Yiğit & Gökpinar, 2010). Specifically, our findings are consistent with the early research suggesting that balanced designs can be used as protection against the effect of variance heterogeneity. However, the results of the present study go further, since they show that this recommendation is accurate – even with small samples and with a variance ratio as high as 9 – when there is a monotonic pattern of variance in the groups, that is, when the values of group variance increase or decrease monotonically so that the groups can be ordered as a function of their respective variances. Other researchers have found that *F*-test is not robust with a balanced design when the pattern of heterogeneity involves a set of groups with similar variances and one with extreme variance (e.g., Alexander & Govern, 1994; Lee & Ahn, 2003; Moder, 2010; Rogan & Keselman, 1977; Wilcox et al., 1986). This finding highlights the relevance of knowing the pattern of variance in the data when performing *F*-test.

With regard to unequal sample sizes, our results appear to be consistent with previous findings, showing that Type I error rates vary depending on the degree of variance heterogeneity and the pairing of variance with group sample size (Box, 1954; Gamage & Weerahandi, 1998; Harwell et al., 1992; Horsnell, 1953; Hsu, 1938; Kohr & Games, 1974; Lee & Ahn, 2003; Moder, 2010; Patrick, 2007; Scheffé, 1959; Tomarken & Serling, 1986; Yiğit & Gökpinar, 2010; Zijlstra, 2004). Specifically, with positive pairing, *F*-test tends to be conservative, with the empirical level of alpha being less than the nominal. With negative pairing, *F*-test tends to be liberal, with the empirical level of alpha being higher than the nominal, such that the risk of declaring mean differences that do not exist is increased. However, the present study extends the findings of previous studies and provides further information about *F*-test robustness under heterogeneity in a

wide range of conditions that applied researchers may encounter in their data, taking into account specific variables such as different values of the pairing of variance with group size, several ratios of variance, and different values of the coefficient of sample size variation.

Furthermore, the results of this study enable us to offer researchers a specific guideline regarding whether or not *F*-test will be sensitive to departures from the homogeneity assumption that may be present in their data. When a monotonic pattern of variance is found in the groups, as was the case here, there are three steps that researchers can follow:

1. Calculate the variance ratio, dividing the value of the largest variance of the groups by the smallest variance. If this ratio is equal to or less than 1.5, *F*-test can be performed with confidence. If this ratio is higher than 1.5, then continue with step 2.
2. Calculate the correlation between group sample size and the values of variance in order to determine the amount of pairing of variance with group sample size. If this correlation is either 0 or 0.5, proceed with *F*-test. Otherwise, continue with step 3.
3. Calculate the coefficient of sample size variation, dividing the standard deviation of the group sample sizes by its mean in order to determine the amount of inequality in group sample sizes.

- If the pairing is equal to 1 and the coefficient of sample size variation is high (close to .50), it is not possible to trust the results of *F*-test for ratios higher than 2 because the actual Type I error may be much lower than the nominal alpha of .05, even reaching .01.

Table 10 Conditions under which *F*-test is not robust, in terms of Type I error, against violation of the homogeneity assumption, according to variance ratio, the pairing of variance with group sample size, and the coefficient of sample size variation

Variance ratio	Pairing	Coefficient of sample size variation	Type I error rate
1.6, 1.7, 1.8	−−1	0.50	Liberal
2	−1	0.33; 0.50	Liberal
3	1	0.50	Conservative
	−.50	0.50	Liberal
	−1	0.33; 0.50	Liberal
5 & 9	1	0.50	Conservative
	−.50	0.33; 0.50	Liberal
	−1	0.16; 0.33; 0.50	Liberal

Table 10 shows the specific conditions in which *F*-test is not robust.

- If the pairing is equal to −.50 and the coefficient of sample size variation is close to 0.33 or higher, results from *F*-test for ratios higher than 2 are not reliable because the actual Type I error may be much higher than the nominal alpha of .05. Thus, there is an increased likelihood of declaring mean differences that do not actually exist. The highest value of Type I error found in this condition was .10.
- If the pairing is equal to −1, and in this case for the majority of sample size coefficients for high variance ratios, results from *F*-test are distorted because the actual Type I error may be much higher than the nominal alpha of .05, even reaching .20 (see Table 10 for specific conditions).

One of the biggest advantages of following these steps is that applied researchers do not need to use any traditional homogeneity tests (e.g., Bartlett, 1937; Cochran, 1941; Hartley, 1950; Levene, 1960), which are known to rely on other assumptions that might not be met (Bhat, Badade, & Aruna Rao, 2002; Conover, Johnson, & Johnson, 1981; Harwell et al., 1992; Moder, 2007; Sharma & Kibria, 2013; Zimmerman, 2004). Moreover, researchers can locate the specific variance conditions and characteristics of their data in the tables provided and see directly if *F*-test is robust or not.

To sum up, this study has two main strengths. First, its systematic approach covers the largest variety of conditions simulated to date when exploring *F*-test robustness to variance heterogeneity, including conditions representative of real data in educational and psychological research. Second, the results yield an easy guideline that can be followed by applied researchers from any background, making it easier for them to decide whether *F*-test can reliably be used when variances are not equal between the groups. Moreover, the guideline provided makes this process fast and straightforward, avoiding the need for traditional homogeneity tests, which cannot be used in a number of conditions. It should be noted, however, that this study has only analyzed the effect of monotonic patterns of variance on the Type I error rate of *F*-test. Future studies should therefore aim to examine power and other patterns of variance besides those considered here. A further potential limitation of this paper is that it aimed to explore the isolated effect of heterogeneity on *F*-test, without considering other assumptions such as normality. An interesting line of future research would be to explore whether or not the violation of normality increases the effect of heterogeneity.

The results of this study suggest that the traditional variance ratio should be used as a measure of the degree of heterogeneity, and indicate that special attention should be paid when the design is unbalanced, the pairing is negative, and the ratio is higher than 1.5. Furthermore, a variance ratio higher than 1.5 may be established as a rule of thumb for considering a potential threat to *F*-test robustness under heterogeneity with unequal sample sizes. This rule of thumb is much more restrictive than the previously recommended maximums of three (Dean & Voss, 1999; Keppel et al., 1992; Kirk, 2013), four or five (Wuensch, 2017), or 10 (Tabachnick & Fidell, 2007, 2013). This paper shows that these criteria may lead, under certain conditions, to incorrect inferences.

The next problem to be tackled is how to address heterogeneity of variance when *F*-test is not robust. Although a detailed analysis of this issue is beyond the scope of the present study, we would like to offer some general recommendations. A first, practical recommendation is that researchers should, if possible, design their study with equal group sample sizes, or, at least, with low sample size variation. However, this is not always possible and there may be disagreement over whether the study design or the data collection procedure should be driven by the statistical analysis.

Some authors have also recommended using a more stringent alpha level in the condition under which an inflated alpha is expected, for example, .025 instead of .05 (Keppel et al., 1992; Keppel & Wickens, 2004; Tabachnick & Fidell, 2007, 2013), or .01 with severe violation (Tabachnick & Fidell, 2007, 2013). This is the simplest procedure for researchers since they may still use *F*-test while maintaining control of Type I error. For illustrative purposes, and in order to examine which alpha level may be used, we conducted simulations under those conditions for which *F*-test is liberal for three groups with a nominal alpha of .05, and considering other more restricted alpha levels (results are shown in Appendix 3). Overall, a nominal alpha level of .025 controls the Type I error rate within the bounds of Bradley's criterion for .05 in the conditions associated with Type I error rates around .10, while a nominal alpha level of .01 achieves this control in the conditions associated with Type I error rates above .10. However, in some conditions with Type I error rates above .15, the level of alpha has to be restricted to .005 to maintain empirical Type I error rates within the bounds of Bradley's criterion for .05. Consequently, researchers can adjust the nominal alpha level depending on the specific characteristics of their data, bearing in

mind that a severe violation of homogeneity requires a more restricted level of alpha.

Another common recommendation for meeting the assumption of variance homogeneity is to transform the response variable (e.g., Montgomery, 1991; Tabachnick & Fidell, 2007, 2013; Winer, Brown, & Michels, 1991). However, it is often difficult to determine which transformation is appropriate for a specific set of data, and results are usually difficult to interpret when data transformations are adopted.

The comparison of means using alternative statistical procedures which have been found to provide more robust results has also been proposed (e.g., Alexander & Govern, 1994; Brown & Forsythe, 1974; Brunner, Dette, & Munk, 1997; Chen & Chen, 1998; James, 1951; Krishnamoorthy, Lu, & Mathew, 2007; Kruskal & Wallis, 1952; Lee & Ahn, 2003; Li, Wang, & Liang, 2011; Lix & Keselman, 1998; Weerahandi, 1995; Welch, 1951; Wilcox, 1995; Wilcox, Keselman, & Kowalchuk, 1998). Below we focus on the most common ones.

The non-parametric Kruskal-Wallis test (Kruskal & Wallis, 1952) is one of the most widely recommended tests in classic handbooks on methodology and statistics. However, the Kruskal-Wallis test has several disadvantages: (1) It converts quantitative continuous data into rank-ordered data, with a consequent loss of information; (2) its null hypothesis differs from that of F-test, unless the distribution of groups has exactly the same shape (see Maxwell & Delaney, 2004); and (3) some Monte Carlo studies have shown that its Type I error is also affected by variance heterogeneity, being liberal (rates greater than .075) with negative pairing (Cribbie et al., 2007; Tomarken & Serling, 1986).

Another common proposal has been to use parametric modifications of F-test, such as Brown-Forsythe (1974) and Welch (1951) tests. Both seem to provide better control over Type I error rates than does F-test under heteroscedasticity. With variance patterns similar to those used here, Tomarken and Serlin (1986) recommended using the Welch test with normal populations, while Clinch and Keselman (1982) recommended the Brown-Forsythe test under both heterogeneity and non-normality. More recently, the results obtained by Parra-Frutos (2014) suggested that both tests perform well with normal data, although the Brown-Forsythe test offers better control of the Type I error rate under several non-normality conditions. Another recently proposed alternative is to use the F-test, Brown-Forsythe or Welch tests with bootstrapping in order to obtain distributions of the statistics instead of using their theoretical distribution (Krishnamoorthy et al., 2007; Parra-Frutos, 2014). Parra-Frutos (2014) showed that the bootstrapped F-test and the bootstrapped Brown-Forsythe test exhibit similar and exceptionally good behavior under heteroscedasticity and non-normality.

Finally, methods using robust estimators of location and robust measures of scale have also been proposed to compare trimmed means. For example, Lix and Keselman (1998), Wilcox (1995), and Wilcox et al. (1998) suggested that the best option was the Welch test on trimmed means and Winsorized variance, although the bootstrap procedure proposed by Krishnamoorthy et al. (2007), used in conjunction with a robust approach, has been shown to provide better control of Type I error under heteroscedasticity (Cribbie, Fiksenbaum, Keselman, & Wilcox, 2012).

Whatever the case, we encourage researchers to analyze the specific characteristics of their design and the data obtained and, if their data do not meet the assumption of variance homogeneity, to choose the best alternative in order to obtain valid results. To this end, the best approach is to perform a simulation study involving the specific conditions of the real data so as to determine whether or not F-test is robust in the situation being considered. We are aware, however, that applied researchers are not usually familiarized with this procedure.

Glossary

Coefficient of sample size variation	The amount of inequality in group sizes, calculated by dividing the standard deviation of the group sample size by its mean.
Conservative	Empirical Type I error rate below .025.
Liberal	Empirical Type I error rate above .075
Pairing of variance with group sample size	The correlation between the two. Pairing is positive when the largest group size is associated with the largest value of variance, and vice-versa. Pairing is negative when the largest group size is associated with the smallest variance, and vice-versa.
Percentage of robustness	Percentage of Type I error rate within the bounds of Bradley's criterion [.025-.075].
Variance ratio	Ratio of the largest to the smallest variance.

Appendix 1. Empirical Type I error rates for *F*-test and equal sample sizes

Table 11 Empirical Type I error for *F*-test with three groups and equal sample sizes

N	n	Variance ratio				
		1.5	2	3	5	9
9	3	.0486	.0513	.0540	.0603	.0699
15	5	.0468	.0547	.0584	.0652	.0606
30	10	.0506	.0529	.0533	.0579	.0629
45	15	.0480	.0518	.0531	.0581	.0627
60	20	.0490	.0507	.0517	.0533	.0586
75	25	.0479	.0533	.0515	.0559	.0576
90	30	.0485	.0508	.0510	.0600	.0610
120	40	.0477	.0495	.0554	.0541	.0567
150	50	.0519	.0507	.0510	.0567	.0569
180	60	.0479	.0506	.0529	.0583	.0607
210	70	.0508	.0523	.0525	.0591	.0607
240	80	.0479	.0482	.0533	.0558	.0553
270	90	.0520	.0590	.0538	.0561	.0646
300	100	.0511	.0476	.0523	.0561	.0569

Table 12 Empirical Type I error for *F*-test with four groups and equal sample sizes

N	n	Variance ratio				
		1.5	2	3	5	9
12	3	.0481	.0580	.0564	.0605	.0749
20	5	.0526	.0535	.0579	.0616	.0642
40	10	.0525	.0560	.0574	.0611	.0590
60	15	.0510	.0546	.0558	.0581	.0610
80	20	.0534	.0561	.0604	.0605	.0602
100	25	.0499	.0515	.0504	.0613	.0594
120	30	.0501	.0531	.0520	.0614	.0580
160	40	.0547	.0514	.0568	.0591	.0606
200	50	.0508	.0518	.0568	.0572	.0611
240	60	.0488	.0497	.0547	.0577	.0630
280	70	.0489	.0563	.0525	.0541	.0627
320	80	.0531	.0535	.0574	.0544	.0599
360	90	.0534	.0472	.0480	.0610	.0613
400	100	.0551	.0453	.0577	.0569	.0619

Table 13 Empirical Type I error for *F*-test with five groups and equal sample sizes

N	n	Variance ratio				
		1.5	2	3	5	9
15	3	.0540	.0544	.0563	.0608	.0721
25	5	.0529	.0546	.0552	.0632	.0689
50	10	.0509	.0514	.0574	.0586	.0622
75	15	.0521	.0513	.0544	.0595	.0667
100	20	.0480	.0501	.0512	.0612	.0670
125	25	.0479	.0536	.0521	.0555	.0608
150	30	.0530	.0477	.0534	.0570	.0640
200	40	.0487	.0528	.0583	.0574	.0629
250	50	.0486	.0572	.0563	.0587	.0598
300	60	.0476	.0498	.0529	.0599	.0600
350	70	.0513	.0502	.0517	.0551	.0626
400	80	.0485	.0486	.0551	.0587	.0607
450	90	.0461	.0513	.0556	.0561	.0614
500	100	.0486	.0533	.0584	.0565	.0633

Table 14 Empirical Type I error for *F*-test with six groups and equal sample sizes

N	n	Variance ratio				
		1.5	2	3	5	9
25	5	.0521	.0523	.0543	.0585	.0632
50	10	.0523	.0506	.0532	.0619	.0624
75	15	.0522	.0517	.0543	.0639	.0658
100	20	.0497	.0497	.0568	.0616	.0670
125	25	.0499	.0492	.0534	.0585	.0616
150	30	.0494	.0516	.0522	.0603	.0675
200	40	.0489	.0489	.0545	.0573	.0639
250	50	.0542	.0560	.0574	.0621	.0630
300	60	.0507	.0487	.0541	.0604	.0630
350	70	.0469	.0518	.0577	.0608	.0596
400	80	.0517	.0556	.0508	.0605	.0607
450	90	.0468	.0547	.0492	.0581	.0602
500	100	.0493	.0550	.0548	.0592	.0622

Appendix 2. Empirical Type I error rates for *F*-test with unequal sample sizes

Table 15 Empirical Type I error for *F*-test with three groups

N	N/J	$\frac{n_{Largest}}{n_{Smallest}}$	Δn	n	Pairing	Order of variance	Variance ratio							
							1.5	1.6	1.7	1.8	2	3	5	9
15	5	1.5	0.16	4, 5, 6	1	1, 2, 3	.0469	.0494	.0505	.0455	.0443	.0438	.0460	.0466
					.50	1, 3, 2	.0487	.0448	.0450	.0446	.0529	.0489	.0510	.0544
					−.50	2, 3, 1	.0552	.0521	.0558	.0586	.0538	.0644	**.0787**	**.0864**
					−1	3, 2, 1	.0551	.0472	.0427	.0454	.0660	.0713	**.0816**	**.0948**
		2.3	0.33	3, 5, 7	1	1, 2, 3	.0451	.0417	.0404	.0355	.0344	.0343	.0260	.0282
					.50	1, 3, 2	.0478	.0415	.0386	.0393	.0487	.0445	.0428	.0441
					−.50	2, 3, 1	.0549	.0633	.0683	.0652	.0506	.0677	**.0873**	**.0977**
					−1	3, 2, 1	.0570	.0565	.0566	.0572	**.0791**	**.0897**	**.1122**	**.1319**
		3	0.50	3, 3, 9	1	1, 2, 3	.0360	.0415	.0370	.0320	.0283	.0254	.0201	.0188
					−1	3, 2, 1	.0688	.0719	**.0760**	**.0786**	**.0794**	**.1171**	**.1624**	**.2077**
30	10	1.5	0.16	8, 10, 12	1	1, 2, 3	.0436	.0483	.0470	.0448	.0448	.0449	.0414	.0400
					.50	1, 3, 2	.0447	.0476	.0464	.0492	.0511	.0477	.0461	.0525
					−.50	2, 3, 1	.0498	.0574	.0570	.0544	.0507	.0632	.0668	.0754
					−1	3, 2, 1	.0581	.0607	.0577	.0626	.0601	.0708	**.0788**	**.0867**
		2.3	0.33	6, 10, 14	1	1, 2, 3	.0428	.0408	.0408	.0351	.0371	.0331	.0292	.0260
					.50	1, 3, 2	.0416	.0445	.0447	.0437	.0470	.0414	.0416	.0459
					−.50	2, 3, 1	.0585	.0586	.0598	.0614	.0580	.0654	**.0870**	**.0913**
					−1	3, 2, 1	.0628	.0698	.0625	.0673	.0734	**.0863**	**.1128**	**.1245**
		4	0.50	4, 10, 16	1	1, 2, 3	.0385	.0362	.0364	.0324	.0320	.0260	.0209	.0201
					.50	1, 3, 2	.0440	.0425	.0408	.0444	.0464	.0370	.0369	.0387
					−.50	2, 3, 1	.0609	.0604	.0651	.0633	.0597	**.0809**	**.0926**	**.1099**
					−1	3, 2, 1	.0725	**.0761**	**.0757**	**.0761**	**.0881**	**.1185**	**.1486**	**.1788**
45	15	1.5	0.16	12, 15, 18	1	1, 2, 3	.0444	.0442	.0407	.0445	.0430	.0425	.0417	.0396
					.50	1, 3, 2	.0476	.0463	.0469	.0463	.0510	.0497	.0469	.0502
					−.50	2, 3, 1	.0544	.0524	.0513	.0543	.0532	.0589	.0653	.0690
					−1	3, 2, 1	.0599	.0592	.0614	.0589	.0614	.0699	**.0807**	**.0861**
		2.3	0.33	9, 15, 21	1	1, 2, 3	.0404	.0396	.0370	.0377	.0339	.0310	.0282	**.0245**
					.50	1, 3, 2	.0426	.0425	.0500	.0419	.0428	.0434	.0400	.0424
					−.50	2, 3, 1	.0540	.0601	.0570	.0553	.0554	.0645	**.0794**	**.0876**
					−1	3, 2, 1	.0607	.0621	.0651	.0631	**.0761**	**.0921**	**.1074**	**.1302**
		4	0.50	6, 15, 24	1	1, 2, 3	.0342	.0373	.0345	.0317	.0270	**.0231**	**.0174**	**.0178**
					.50	1, 3, 2	.0406	.0445	.0414	.0393	.0459	.0348	.0344	.0374
					−.50	2, 3, 1	.0567	.0596	.0661	.0639	.0613	**.0785**	**.0914**	**.1082**
					−1	3, 2, 1	.0727	.0747	**.0788**	**.0803**	**.0916**	**.1132**	**.1497**	**.1737**
60	20	1.5	0.16	16, 20, 24	1	1, 2, 3	.0476	.0444	.0454	.0440	.0423	.0407	.0399	.0399
					.50	1, 3, 2	.0464	.0453	.0449	.0472	.0489	.0468	.0491	.0527
					−.50	2, 3, 1	.0502	.0533	.0527	.0541	.0548	.0612	.0641	.0735
					−1	3, 2, 1	.0599	.0611	.0597	.0579	.0637	.0697	**.0761**	**.0905**
		2.3	0.33	12, 20, 28	1	1, 2, 3	.0409	.0433	.0393	.0382	.0344	.0300	.0286	.0270
					.50	1, 3, 2	.0454	.0446	.0471	.0398	.0458	.0420	.0418	.0425
					−.50	2, 3, 1	.0569	.0595	.0585	.0559	.0553	.0707	.0748	**.0873**
					−1	3, 2, 1	.0601	.0626	.0647	.0672	**.0765**	**.0949**	**.1072**	**.1208**
		4	0.50	8, 20, 32	1	1, 2, 3	.0399	.0320	.0325	.0352	.0292	**.0248**	**.0191**	**.0182**
					.50	1, 3, 2	.0452	.0436	.0441	.0386	.0431	.0353	.0309	.0348
					−.50	2, 3, 1	.0590	.0592	.0646	.0611	.0532	**.0764**	**.0885**	**.0953**
					−1	3, 2, 1	.0646	.0710	.0733	**.0767**	**.0911**	**.1137**	**.1452**	**.1739**
75	25	1.5	0.16	20, 25, 30	1	1, 2, 3	.0450	.0448	.0447	.0463	.0409	.0440	.0378	.0391
					.50	1, 3, 2	.0479	.0487	.0459	.0475	.0521	.0489	.0501	.0491
					−.50	2, 3, 1	.0490	.0558	.0539	.0574	.0545	.0605	.0665	.0708
					−1	3, 2, 1	.0560	.0581	.0601	.0615	.0616	.0712	.0748	**.0885**
		2.3	0.33	15, 25, 35	1	1, 2, 3	.0433	.0390	.0392	.0397	.0366	.0287	.0281	.0264
					.50	1, 3, 2	.0415	.0465	.0438	.0422	.0491	.0420	.0420	.0442
					−.50	2, 3, 1	.0537	.0542	.0605	.0597	.0544	.0702	.0729	**.0873**
					−1	3, 2, 1	.0565	.0669	.0670	.0679	.0738	**.0829**	**.1069**	**.1224**
		4	0.50	10, 25, 40	1	1, 2, 3	.0366	.0323	.0315	.0326	.0270	**.0236**	**.0197**	**.0186**
					.50	1, 3, 2	.0431	.0448	.0414	.0428	.0453	.0354	.0361	.0363
					−.50	2, 3, 1	.0579	.0601	.0665	.0708	.0575	**.0783**	**.0902**	**.1009**
					−1	3, 2, 1	.0691	**.0759**	**.0764**	**.0806**	**.0900**	**.1101**	**.1420**	**.1737**
90	30	1.5	0.16	24, 30, 36	1	1, 2, 3	.0419	.0421	.0421	.0441	.0395	.0397	.0395	.0427

Table 15 (continued)

N	N/J	$\frac{n_{\text{Largest}}}{n_{\text{Smallest}}}$	Δn	n	Pairing	Order of variance	Variance ratio							
							1.5	1.6	1.7	1.8	2	3	5	9
					.50	1, 3, 2	.0454	.0476	.0508	.0403	.0515	.0475	.0488	.0510
					−.50	2, 3, 1	.0553	.0563	.0557	.0577	.0560	.0577	.0664	.0695
					−1	3, 2, 1	.0557	.0592	.0509	.0619	.0641	.0681	**.0762**	**.0859**
		2.3	0.33	18, 30, 42	1	1, 2, 3	.0391	.0413	.0379	.0346	.0347	.0300	.0329	.0287
					.50	1, 3, 2	.0453	.0432	.0424	.0411	.0435	.0405	.0422	.0427
					−.50	2, 3, 1	.0549	.0556	.0591	.0602	.0573	.0719	**.0755**	**.0812**
					−1	3, 2, 1	.0582	.0632	.0689	.0679	.0731	**.0880**	**.1032**	**.1242**
		4	0.50	12, 30, 48	1	1, 2, 3	.0365	.0315	.0371	.0318	.0260	**.0232**	**.0199**	**.0128**
					.50	1, 3, 2	.0403	.0444	.0416	.0403	.0432	.0381	.0351	.0328
					−.50	2, 3, 1	.0577	.0632	.0640	.0677	.0601	**.0755**	**.0890**	**.1041**
					−1	3, 2, 1	.0640	.0729	**.0756**	**.0796**	**.0930**	**.1085**	**.1400**	**.1748**
120	40	1.5	0.16	32, 40, 48	1	1, 2, 3	.0423	.0455	.0437	.0426	.0430	.0399	.0381	.0408
					.50	1, 3, 2	.0494	.0469	.0479	.0448	.0492	.0445	.0524	.0541
					−.50	2, 3, 1	.0556	.0543	.0534	.0519	.0539	.0566	.0653	.0697
					−1	3, 2, 1	.0576	.0560	.0567	.0586	.0601	.0662	**.0753**	**.0846**
		2.3	0.33	24, 40, 56	1	1, 2, 3	.0360	.0383	.0401	.0368	.0352	.0312	.0294	.0253
					.50	1, 3, 2	.0415	.0450	.0453	.0439	.0453	.0426	.0415	.0457
					−.50	2, 3, 1	.0576	.0541	.0559	.0589	.0589	.0726	**.0779**	**.0840**
					−1	3, 2, 1	.0627	.0660	.0742	.0663	.0732	**.0903**	**.1061**	**.1206**
		4	0.50	16, 40, 64	1	1, 2, 3	.0353	.0310	.0283	.0349	.0268	.0259	**.0186**	**.0179**
					.50	1, 3, 2	.0423	.0416	.0400	.0422	.0482	.0368	.0340	.0413
					−.50	2, 3, 1	.0643	.0620	.0662	.0649	.0552	**.0774**	**.0933**	**.0967**
					−1	3, 2, 1	.0711	.0726	**.0807**	**.0734**	**.0880**	**.1161**	**.1460**	**.1646**
150	50	1.5	0.16	40, 50, 60	1	1, 2, 3	.0444	.0473	.0446	.0450	.0391	.0409	.0401	.0390
					.50	1, 3, 2	.0464	.0459	.0452	.0445	.0494	.0464	.0453	.0490
					−.50	2, 3, 1	.0522	.0515	.0538	.0535	.0527	.0602	.0603	.0699
					−1	3, 2, 1	.0563	.0530	.0564	.0597	.0609	.0683	**.0764**	**.0868**
		2.3	0.33	30, 50, 70	1	1, 2, 3	.0399	.0370	.0368	.0356	.0341	.0279	.0269	.0259
					.50	1, 3, 2	.0458	.0449	.0459	.0440	.0476	.0417	.0450	.0398
					−.50	2, 3, 1	.0540	.0589	.0596	.0593	.0521	.0673	**.0803**	**.0772**
					−1	3, 2, 1	.0610	.0618	.0634	.0685	.0727	**.0879**	**.1061**	**.1258**
		4	0.50	20, 50, 80	1	1, 2, 3	.0341	.0345	.0327	.0295	.0251	**.0224**	**.0180**	**.0174**
					.50	1, 3, 2	.0455	.0387	.0444	.0414	.0437	.0367	.0375	.0343
					−.50	2, 3, 1	.0614	.0582	.0598	.0660	.0557	**.0793**	**.0881**	**.0987**
					−1	3, 2, 1	.0685	**.0759**	**.0808**	**.0831**	**.0880**	**.1071**	**.1424**	**.1731**
180	60	1.5	0.16	48, 60, 72	1	1, 2, 3	.0427	.0449	.0415	.0404	.0433	.0402	.0390	.0394
					.50	1, 3, 2	.0444	.0499	.0451	.0468	.0453	.0475	.0452	.0488
					−.50	2, 3, 1	.0529	.0540	.0534	.0613	.0512	.0582	.0650	.0705
					−1	3, 2, 1	.0557	.0585	.0567	.0587	.0606	.0691	**.0774**	**.0823**
		2.3	0.33	36, 60, 84	1	1, 2, 3	.0376	.0373	.0387	.0372	.0327	.0322	.0262	.0261
					.50	1, 3, 2	.0426	.0436	.0448	.0465	.0420	.0394	.0404	.0445
					−.50	2, 3, 1	.0561	.0592	.0554	.0623	.0563	.0690	.0748	**.0834**
					−1	3, 2, 1	.0595	.0643	.0627	.0687	.0716	**.0907**	**.1039**	**.1190**
		4	0.50	24, 60, 96	1	1, 2, 3	.0349	.0316	.0340	.0329	.0282	**.0228**	**.0206**	**.0182**
					.50	1, 3, 2	.0410	.0424	.0435	.0399	.0421	.0351	.0409	.0347
					−.50	2, 3, 1	.0573	.0638	.0624	.0613	.0548	**.0762**	**.0888**	**.0991**
					−1	3, 2, 1	.0686	**.0753**	**.0775**	**.0787**	**.0825**	**.1167**	**.1445**	**.1697**
210	70	1.5	0.16	56, 70, 84	1	1, 2, 3	.0454	.0464	.0422	.0426	.0452	.0408	.0427	.0440
					.50	1, 3, 2	.0478	.0451	.0449	.0497	.0496	.0502	.0475	.0520
					−.50	2, 3, 1	.0530	.0565	.0561	.0552	.0548	.0584	.0630	.0699
					−1	3, 2, 1	.0588	.0591	.0611	.0560	.0610	.0666	**.0767**	**.0855**
		2.3	0.33	42, 70, 98	1	1, 2, 3	.0410	.0367	.0353	.0337	.0356	.0302	.0251	**.0244**
					.50	1, 3, 2	.0442	.0451	.0457	.0454	.0438	.0407	.0380	.0417
					−.50	2, 3, 1	.0550	.0571	.0584	.0599	.0562	.0665	.0750	**.0759**
					−1	3, 2, 1	.0603	.0676	.0635	.0715	.0749	**.0842**	**.1074**	**.1224**
		4	0.50	28, 70, 112	1	1, 2, 3	.0361	.0312	.0337	.0329	.0284	**.0225**	**.0184**	**.0186**
					.50	1, 3, 2	.0406	.0400	.0436	.0384	.0395	.0340	.0370	.0373
					−.50	2, 3, 1	.0592	.0579	.0604	.0603	.0570	**.0764**	**.0866**	**.1019**
					−1	3, 2, 1	.0694	.0734	**.0769**	**.0762**	**.0843**	**.1141**	**.1470**	**.1639**
240	80	1.5	0.16	64, 80, 96	1	1, 2, 3	.0455	.0457	.0414	.0405	.0400	.0386	.0391	.0363
					.50	1, 3, 2	.0464	.0470	.0457	.0485	.0465	.0422	.0456	.0532
					−.50	2, 3, 1	.0490	.0527	.0528	.0569	.0522	.0601	.0629	.0671
					−1	3, 2, 1	.0520	.0580	.0576	.0587	.0618	.0668	.0698	**.0821**

Table 15 (continued)

N	N/J	$\frac{n_{Largest}}{n_{Smallest}}$	Δn	n	Pairing	Order of variance	1.5	1.6	1.7	1.8	2	3	5	9
		2.3	0.33	48, 80, 112	1	1, 2, 3	.0415	.0400	.0372	.0368	.0331	.0297	.0265	.0279
					.50	1, 3, 2	.0470	.0477	.0439	.0459	.0532	.0418	.0406	.0440
					−.50	2, 3, 1	.0566	.0573	.0571	.0599	.0560	.0686	.0742	**.0866**
					−1	3, 2, 1	.0629	.0701	.0658	.0692	**.0773**	**.0859**	**.1052**	**.1256**
		4	0.50	32, 80, 128	1	1, 2, 3	.0313	.0301	.0332	.0309	.0284	**.0223**	**.0183**	**.0175**
					.50	1, 3, 2	.0413	.0409	.0452	.0395	.0437	.0361	.0333	.0367
					−.50	2, 3, 1	.0567	.0632	.0619	.0610	.0568	.0741	**.0880**	**.0960**
					−1	3, 2, 1	.0634	.0736	**.0782**	**.0762**	**.0908**	**.1133**	**.1429**	**.1687**
270	90	1.5	0.16	72, 90, 108	1	1, 2, 3	.0438	.0462	.0439	.0388	.0414	.0429	.0426	.0381
					.50	1, 3, 2	.0460	.0425	.0438	.0457	.0451	.0412	.0468	.0458
					−.50	2, 3, 1	.0522	.0554	.0552	.0571	.0534	.0608	.0634	.0703
					−1	3, 2, 1	.0541	.0589	.0575	.0593	.0582	.0662	**.0837**	**.0878**
		2.3	0.33	54, 90, 126	1	1, 2, 3	.0385	.0353	.0378	.0398	.0331	.0295	.0250	**.0242**
					.50	1, 3, 2	.0479	.0439	.0481	.0448	.0447	.0428	.0417	.0424
					−.50	2, 3, 1	.0543	.0579	.0587	.0597	.0532	.0429	.0741	**.0827**
					−1	3, 2, 1	.0596	.0672	.0665	.0673	.0743	**.0873**	**.1012**	**.1140**
		4	0.50	36, 90, 144	1	1, 2, 3	.0359	.0343	.0325	.0318	.0256	**.0230**	**.0188**	**.0169**
					.50	1, 3, 2	.0362	.0424	.0411	.0401	.0440	.0334	.0363	.0373
					−.50	2, 3, 1	.0569	.0599	.0589	.0666	.0548	**.0782**	**.0863**	**.0995**
					−1	3, 2, 1	.0679	.0708	**.0786**	**.0816**	**.0882**	**.1075**	**.1397**	**.1629**
300	100	1.5	0.16	80, 100, 120	1	1, 2, 3	.0429	.0458	.0399	.0424	.0399	.0383	.0408	.0400
					.50	1, 3, 2	.0496	.0442	.0482	.0475	.0489	.0462	.0454	.0510
					−.50	2, 3, 1	.0530	.0550	.0541	.0528	.0513	.0581	.0659	.0621
					−1	3, 2, 1	.0531	.0568	.0610	.0562	.0607	.0729	.0746	**.0853**
		2.3	0.33	60, 100, 140	1	1, 2, 3	.0395	.0385	.0370	.0358	.0330	.0276	**.0246**	.0278
					.50	1, 3, 2	.0430	.0441	.0420	.0408	.0493	.0431	.0442	.0426
					−.50	2, 3, 1	.0579	.0571	.0585	.0592	.0587	.0696	**.0752**	**.0867**
					−1	3, 2, 1	.0628	.0638	.0708	.0645	**.0755**	**.0857**	**.1106**	**.1196**
		4	0.50	40, 100, 160	1	1, 2, 3	.0340	.0319	.0307	.0308	.0283	**.0228**	**.0203**	**.0166**
					.50	1, 3, 2	.0432	.0444	.0380	.0383	.0454	.0343	.0358	.0366
					−.50	2, 3, 1	.0599	.0616	.0677	.0667	.0572	**.0781**	**.0866**	**.0981**
					−1	3, 2, 1	.0648	.0698	**.0782**	**.0821**	**.0816**	**.1078**	**.1375**	**.1713**

Note. N = total sample size; N/J = mean of group sample size; $\frac{n_{Largest}}{n_{Smallest}}$ = ratio of the largest to the smallest group sample size; Δn = coefficient of sample size variation; n = number of units per group. The Type I error rates outside the boundary of Bradley's liberal criterion are in bold (conservative: < .025; liberal: > .075)

Table 16 Empirical Type I error for *F*-test with five groups

N	N/J	$\frac{n_{Largest}}{n_{Smallest}}$	Δn	n	Pairing	Order of variance	1.5	1.6	1.7	1.8	2	3	5	9
25	5	1.5	0.16	4, 4, 5, 6, 6	1	1, 2, 3, 4, 5	.0489	.0454	.0473	.0454	.0473	.0430	.0435	.0446
					.50	1, 4, 2, 5, 3	.0506	.0475	.0516	.0525	.0448	.0469	.0507	.0522
					0	1, 5, 3, 4, 2	.0492	.0531	.0534	.0529	.0551	.0528	.0592	.0654
					−.50	3, 5, 2, 4, 1	.0529	.0518	.0552	.0544	.0565	.0634	**.0764**	**.0829**
					−1	5, 4, 3, 2, 1	.0548	.0599	.0550	.0578	.0650	.0733	**.0873**	**.0994**
		2.3	0.33	3, 3, 6, 6, 7	1	1, 2, 3, 4, 5	.0421	.0424	.0415	.0411	.0402	.0354	.0348	.0312
					.50	1, 4, 2, 3, 5	.0444	.0439	.0436	.0439	.0467	.0409	.0469	.0478
					0	1, 5, 4, 2, 3	.0483	.0463	.0520	.0525	.0480	.0528	.0601	.0622
					−.50	5, 3, 2, 1, 4	.0550	.0584	.0558	.0598	.0585	.0675	.0799	.0861
					−1	5, 4, 3, 2, 1	.0653	.0659	.0687	.0594	.0726	**.0908**	**.1088**	**.1285**
		3.3	0.50	3, 3, 4, 5, 10	1	1, 2, 3, 4, 5	.0414	.0358	.0397	.0359	.0341	.0268	**.0221**	**.0211**
					.50	1, 2, 5, 3, 4	.0441	.0441	.0458	.0500	.0309	.0373	.0363	.0331
					0	1, 5, 4, 2, 3	.0551	.0557	.0610	.0605	.0546	.0557	.0622	.0659

Table 16 (continued)

N	N/J	$\frac{n_{Largest}}{n_{Smallest}}$	Δn	n	Pairing	Order of variance	Variance ratio							
							1.5	1.6	1.7	1.8	2	3	5	9
					−.50	4, 2, 3, 5, 1	.0561	.0496	.0505	.0492	.0721	.0889	.1088	.1296
					−1	5, 4, 3, 2, 1	.0669	.0733	.0749	.0712	**.0791**	**.1103**	**.1511**	**.1810**
50	10	1.5	0.16	8, 9, 10, 11, 12	1	1, 2, 3, 4, 5	.0471	.0458	.0483	.0441	.0414	.0454	.0416	.0466
					.50	1, 4, 2, 5, 3	.0495	.0464	.0514	.0479	.0479	.0466	.0534	.0537
					0	1, 5, 4, 2, 3	.0476	.0478	.0510	.0529	.0502	.0544	.0573	.0640
					−.50	3, 5, 2, 4, 1	.0522	.0494	.0546	.0525	.0547	.0619	.0693	**.0778**
					−1	5, 4, 3, 2, 1	.0549	.0534	.0570	.0576	.0609	.0702	**.0803**	**.0855**
		2.5	0.33	6, 8, 9, 12, 15	1	1, 2, 3, 4, 5	.0432	.0387	.0388	.0393	.0372	.0349	.0315	.0295
					.50	1, 4, 2, 5, 3	.0470	.0481	.0448	.0454	.0477	.0441	.0464	.0435
					0	1, 5, 4, 2, 3	.0514	.0488	.0502	.0547	.0497	.0508	.0586	.0634
					−.50	3, 5, 2, 4, 1	.0560	.0533	.0573	.0591	.0615	.0687	**.0791**	**.0864**
					−1	5, 4, 3, 2, 1	.0599	.0660	.0685	.0707	.0722	**.0872**	**.1088**	**.1309**
		5.7	0.50	3, 6, 10, 14, 17	1	1, 2, 3, 4, 5	.0358	.0371	.0316	.0315	.0317	.0252	**.0228**	**.0206**
					.50	1, 4, 2, 5, 3	.0448	.0425	.0371	.0453	.0418	.0340	.0374	.0351
					0	1, 5, 4, 2, 3	.0511	.0514	.0491	.0497	.0502	.0550	.0537	.0638
					−.50	3, 5, 2, 4, 1	.0590	.0599	.0610	.0697	.0674	**.0822**	**.0983**	**.1124**
					−1	5, 4, 3, 2, 1	.0700	.0711	.0740	**.0801**	**.0852**	**.1148**	**.1511**	**.1918**
75	15	1.5	0.16	12, 13, 15, 17, 18	1	1, 2, 3, 4, 5	.0481	.0484	.0435	.0447	.0465	.0472	.0415	.0464
					.50	1, 4, 2, 5, 3	.0471	.0489	.0489	.0465	.0488	.0473	.0498	.0531
					0	1, 5, 4, 2, 3	.0550	.0495	.0482	.0528	.0549	.0544	.0616	.0665
					−.50	3, 5, 2, 4, 1	.0529	.0517	.0529	.0558	.0583	.0589	.0687	.0727
					−1	5, 4, 3, 2, 1	.0580	.0582	.0529	.0611	.0604	.0692	.0746	**.0867**
		2.5	0.33	9, 12, 13, 18, 23	1	1, 2, 3, 4, 5	.0405	.0422	.0385	.0410	.0365	.0360	.0332	.0304
					.50	1, 4, 2, 5, 3	.0421	.0456	.0446	.0451	.0471	.0449	.0427	.0447
					0	1, 5, 4, 2, 3	.0512	.0510	.0526	.0554	.0529	.0516	.0586	.0603
					−.50	3, 5, 2, 4, 1	.0567	.0533	.0582	.0594	.0617	.0690	**.0798**	**.0882**
					−1	5, 4, 3, 2, 1	.0640	.0653	.0686	.0680	.0693	**.0875**	**.1059**	**.1277**
		5	0.50	5, 9, 15, 21, 25	1	1, 2, 3, 4, 5	.0341	.0375	.0336	.0309	.0308	.0252	**.0204**	**.0193**
					.50	1, 4, 2, 5, 3	.0437	.0433	.0427	.0429	.0480	.0405	.0370	.0389
					0	1, 5, 4, 2, 3	.0540	.0479	.0460	.0512	.0536	.0563	.0589	.0607
					−.50	3, 5, 2, 4, 1	.0603	.0635	.0556	.0594	.0670	**.0820**	**.0920**	**.1096**
					−1	5, 4, 3, 2, 1	.0670	.0711	**.0778**	**.0834**	**.0848**	**.1117**	**.1462**	**.1759**
100	20	1.5	0.16	16, 18, 20, 22, 24	1	1, 2, 3, 4, 5	.0484	.0457	.0438	.0467	.0467	.0483	.0442	.0466
					.50	1, 4, 2, 5, 3	.0460	.0459	.0432	.0481	.0458	.0486	.0527	.0542
					0	1, 5, 4, 2, 3	.0491	.0536	.0492	.0511	.0526	.0527	.0573	.0643
					−.50	3, 5, 2, 4, 1	.0520	.0532	.0585	.0548	.0579	.0579	.0632	.0728
					−1	5, 4, 3, 2, 1	.0577	.0586	.0585	.0548	.0611	.0688	.0714	**.0824**
		2.5	0.33	12, 16, 18, 24, 30	1	1, 2, 3, 4, 5	.0413	.0421	.0388	.0402	.0364	.0350	.0331	.0339
					.50	1, 4, 2, 5, 3	.0465	.0472	.0459	.0437	.0472	.0415	.0465	.0421
					0	1, 5, 4, 2, 3	.0516	.0533	.0496	.0537	.0510	.0557	.0611	.0604
					−.50	3, 5, 2, 4, 1	.0560	.0561	.0583	.0569	.0634	.0665	**.0786**	**.0911**
					−1	5, 4, 3, 2, 1	.0622	.0597	.0661	.0664	.0704	**.0900**	**.1059**	**.1235**
		5.7	0.50	6, 12, 20, 28, 34	1	1, 2, 3, 4, 5	.0319	.0349	.0329	.0333	.0300	.0252	**.0208**	**.0214**
					.50	1, 4, 2, 5, 3	.0440	.0397	.0420	.0452	.0387	.0386	.0333	.0377
					0	1, 5, 4, 2, 3	.0513	.0483	.0489	.0481	.0512	.0606	.0551	.0621
					−.50	3, 5, 2, 4, 1	.0560	.0593	.0611	.0626	.0676	**.0806**	**.0970**	**.1122**
					−1	5, 4, 3, 2, 1	.0672	**.0769**	.0716	**.0809**	**.0880**	**.1090**	**.1471**	**.1797**

Table 16 (continued)

N	N/J	$\frac{n_{\text{Largest}}}{n_{\text{Smallest}}}$	Δn	n	Pairing	Order of variance	Variance ratio							
							1.5	1.6	1.7	1.8	2	3	5	9
125	25	1.5	0.16	20, 22, 25, 28, 30	1	1, 2, 3, 4, 5	.0470	.0492	.0437	.0469	.0436	.0446	.0481	.0459
					.50	1, 4, 2, 5, 3	.0469	.0465	.0492	.0459	.0467	.0476	.0519	.0520
					0	1, 5, 4, 2, 3	.0480	.0533	.0521	.0534	.0518	.0569	.0598	.0667
					−.50	3, 5, 2, 4, 1	.0540	.0523	.0582	.0518	.0545	.0621	.0674	.0750
					−1	5, 4, 3, 2, 1	.0550	.0571	.0576	.0599	.0616	.0676	**.0791**	**.0813**
		2.5	0.33	15, 20, 22, 30, 38	1	1, 2, 3, 4, 5	.0398	.0465	.0419	.0396	.0357	.0350	.0313	.0325
					.50	1, 4, 2, 5, 3	.0453	.0480	.0454	.0428	.0443	.0429	.0454	.0501
					0	1, 5, 4, 2, 3	.0506	.0523	.0521	.0527	.0517	.0486	.0571	.0620
					−.50	3, 5, 2, 4, 1	.0572	.0579	.0591	.0528	.0650	.0697	**.0771**	**.0865**
					−1	5, 4, 3, 2, 1	.0595	.0599	.0652	.0707	.0713	**.0883**	**.1058**	**.1281**
		5.3	0.50	8, 15, 25, 35, 42	1	1, 2, 3, 4, 5	.0325	.0369	.0312	.0319	.0291	.0279	**.0248**	**.0207**
					.50	1, 4, 2, 5, 3	.0415	.0412	.0430	.0405	.0400	.0373	.0371	.0380
					0	1, 5, 4, 2, 3	.0481	.0486	.0497	.0466	.0545	.0615	.0544	.0607
					−.50	3, 5, 2, 4, 1	.0576	.0501	.0653	.0610	.0645	**.0768**	**.0908**	**.1042**
					−1	5, 4, 3, 2, 1	.0684	.0745	**.0804**	**.0738**	**.0867**	**.1112**	**.1447**	**.1803**
150	30	1.5	0.16	24, 27, 30, 33, 36	1	1, 2, 3, 4, 5	.0452	.0440	.0459	.0493	.0475	.0414	.0440	.0431
					.50	1, 4, 2, 5, 3	.0486	.0450	.0499	.0491	.0474	.0449	.0545	.0506
					0	1, 5, 4, 2, 3	.0460	.0490	.0512	.0473	.0487	.0526	.0574	.0608
					−.50	3, 5, 2, 4, 1	.0506	.0542	.0534	.0566	.0542	.0616	.0715	**.0766**
					−1	5, 4, 3, 2, 1	.0519	.0568	.0583	.0573	.0586	.0707	**.0804**	**.0914**
		2.5	0.33	18, 24, 27, 36, 45	1	1, 2, 3, 4, 5	.0408	.0414	.0390	.0390	.0384	.0337	.0326	.0289
					.50	1, 4, 2, 5, 3	.0453	.0453	.0413	.0493	.0403	.0458	.0456	.0452
					0	1, 5, 4, 2, 3	.0503	.0484	.0534	.0537	.0513	.0548	.0569	.0591
					−.50	3, 5, 2, 4, 1	.0548	.0544	.0588	.0524	.0608	.0703	**.0785**	**.0858**
					−1	5, 4, 3, 2, 1	.0615	.0677	.0667	.0664	.0682	**.0853**	**.1043**	**.1221**
		5.7	0.50	9, 18, 30, 42, 51	1	1, 2, 3, 4, 5	.0320	.0351	.0316	.0312	.0287	.0252	**.0224**	**.0195**
					.50	1, 4, 2, 5, 3	.0438	.0419	.0409	.0398	.0410	.0362	.0355	.0361
					0	1, 5, 4, 2, 3	.0503	.0523	.0486	.0486	.0496	.0535	.0540	.0599
					−.50	3, 5, 2, 4, 1	.0611	.0618	.0640	.0630	.0700	**.0784**	**.0966**	**.1041**
					−1	5, 4, 3, 2, 1	.0705	.0732	**.0801**	**.0789**	**.0905**	**.1172**	**.1451**	**.1783**
200	40	1.5	0.16	32, 36, 40, 44, 48	1	1, 2, 3, 4, 5	.0428	.0410	.0433	.0452	.0397	.0418	.0468	.0431
					.50	1, 4, 2, 5, 3	.0474	.0469	.0452	.0503	.0460	.0512	.0508	.0521
					0	1, 5, 4, 2, 3	.0499	.0481	.0461	.0472	.0550	.0528	.0539	.0641
					−.50	3, 5, 2, 4, 1	.0535	.0501	.0506	.0543	.0534	.0635	.0663	**.0753**
					−1	5, 4, 3, 2, 1	.0543	.0577	.0573	.0588	.0559	.0648	.0741	**.0848**
		2.5	0.33	24, 32, 36, 48, 60	1	1, 2, 3, 4, 5	.0399	.0422	.0360	.0394	.0329	.0331	.0313	.0288
					.50	1, 4, 2, 5, 3	.0466	.0466	.0439	.0447	.0430	.0451	.0416	.0436
					0	1, 5, 4, 2, 3	.0519	.0555	.0530	.0512	.0517	.0550	.0584	.0620
					−.50	3, 5, 2, 4, 1	.0551	.0537	.0558	.0575	.0655	.0711	**.0799**	**.0844**
					−1	5, 4, 3, 2, 1	.0576	.0628	.0636	.0634	**.0769**	**.0805**	**.1080**	**.1177**
		5.7	0.50	12, 24, 40, 56, 68	1	1, 2, 3, 4, 5	.0348	.0369	.0327	.0363	.0290	.0254	**.0221**	**.0198**
					.50	1, 4, 2, 5, 3	.0439	.0408	.0420	.0393	.0382	.0368	.0356	.0379
					0	1, 5, 4, 2, 3	.0499	.0497	.0474	.0491	.0480	.0524	.0600	.0615
					−.50	3, 5, 2, 4, 1	.0627	.0591	.0585	.0627	.0722	**.0782**	**.0775**	**.1003**
					−1	5, 4, 3, 2, 1	.0715	.0737	**.0756**	**.0789**	**.0907**	**.1136**	**.1480**	**.1874**
250	50	1.5	0.16	40, 45, 50, 55, 60	1	1, 2, 3, 4, 5	.0466	.0437	.0455	.0422	.0421	.0449	.0495	.0452
					.50	1, 4, 2, 5, 3	.0481	.0465	.0454	.0453	.0428	.0520	.0511	.0556

Table 16 (continued)

N	N/J	$\frac{n_{Largest}}{n_{Smallest}}$	Δn	n	Pairing	Order of variance	Variance ratio							
							1.5	1.6	1.7	1.8	2	3	5	9
					0	1, 5, 4, 2, 3	.0497	.0449	.0505	.0509	.0513	.0496	.0576	.0622
					−.50	3, 5, 2, 4, 1	.0549	.0511	.0505	.0530	.0506	.0650	.0605	.0697
					−1	5, 4, 3, 2, 1	.0552	.0586	.0596	.0614	.0600	.0679	.0733	**.0818**
		2.5	0.33	30, 40, 45, 60, 75	1	1, 2, 3, 4, 5	.0397	.0408	.0385	.0391	.0350	.0309	.0296	.0308
					.50	1, 4, 2, 5, 3	.0433	.0478	.0482	.0473	.0443	.0447	.0410	.0481
					0	1, 5, 4, 2, 3	.0504	.0492	.0497	.0538	.0509	.0571	.0596	.0609
					−.50	3, 5, 2, 4, 1	.0531	.0566	.0573	.0569	.0626	.0627	.0738	**.0840**
					−1	5, 4, 3, 2, 1	.0605	.0604	.0660	.0643	.0707	**.0878**	**.1017**	**.1200**
		5.7	0.50	15, 30, 50, 70, 85	1	1, 2, 3, 4, 5	.0352	.0324	.0342	.0297	.0303	**.0232**	**.0207**	**.0188**
					.50	1, 4, 2, 5, 3	.0410	.0429	.0442	.0405	.0419	.0353	.0371	.0354
					0	1, 5, 4, 2, 3	.0480	.0486	.0526	.0512	.0527	.0520	.0552	.0601
					−.50	3, 5, 2, 4, 1	.0577	.0513	.0612	.0596	.0677	**.0778**	**.0944**	**.1036**
					−1	5, 4, 3, 2, 1	.0711	**.0774**	**.0777**	**.0772**	**.0853**	**.1184**	**.1427**	**.1828**
300	60	1.5	0.16	48, 54, 60, 66, 72	1	1, 2, 3, 4, 5	.0491	.0459	.0424	.0428	.0403	.0414	.0448	.0464
					.50	1, 4, 2, 5, 3	.0497	.0516	.0482	.0453	.0474	.0479	.0494	.0538
					0	1, 5, 4, 2, 3	.0484	.0538	.0498	.0502	.0523	.0541	.0587	.0637
					−.50	3, 5, 2, 4, 1	.0540	.0530	.0508	.0547	.0578	.0594	.0670	.0737
					−1	5, 4, 3, 2, 1	.0535	.0540	.0571	.0558	.0603	.0708	.0722	**.0872**
		2.5	0.33	36, 48, 54, 72, 90	1	1, 2, 3, 4, 5	.0371	.0385	.0432	.0392	.0360	.0317	.0320	.0330
					.50	1, 4, 2, 5, 3	.0473	.0499	.0436	.0429	.0472	.0426	.0416	.0485
					0	1, 5, 4, 2, 3	.0476	.0531	.0555	.0530	.0486	.0516	.0581	.0655
					−.50	3, 5, 2, 4, 1	.0531	.0557	.0595	.0546	.0661	.0723	**.0765**	**.0851**
					−1	5, 4, 3, 2, 1	.0588	.0651	.0646	.0721	.0732	**.0890**	**.1031**	**.1185**
		5.7	0.50	18, 36, 60, 84, 102	1	1, 2, 3, 4, 5	.0351	.0378	.0322	.0289	.0283	**.0246**	**.0225**	**.0188**
					.50	1, 4, 2, 5, 3	.0413	.0408	.0398	.0408	.0411	.0370	.0363	.0360
					0	1, 5, 4, 2, 3	.0515	.0443	.0464	.0516	.0490	.0515	.0560	.0553
					−.50	3, 5, 2, 4, 1	.0607	.0578	.0603	.0654	.0679	**.0790**	**.0881**	**.1124**
					−1	5, 4, 3, 2, 1	.0714	.0725	**.0786**	**.0817**	**.0853**	**.1154**	**.1437**	**.1802**
350	70	1.5	0.16	56, 63, 70, 77, 84	1	1, 2, 3, 4, 5	.0459	.0451	.0414	.0416	.0454	.0444	.0447	.0447
					.50	1, 4, 2, 5, 3	.0465	.0488	.0444	.0499	.0469	.0498	.0514	.0509
					0	1, 5, 4, 2, 3	.0477	.0497	.0516	.0506	.0520	.0559	.0562	.0567
					−.50	3, 5, 2, 4, 1	.0503	.0532	.0566	.0506	.0585	.0597	.0637	.0709
					−1	5, 4, 3, 2, 1	.0551	.0548	.0542	.0570	.0630	.0722	**.0804**	**.0850**
		2.5	0.33	42, 56, 63, 84, 105	1	1, 2, 3, 4, 5	.0419	.0424	.0437	.0405	.0334	.0337	.0310	.0279
					.50	1, 4, 2, 5, 3	.0466	.0444	.0427	.0461	.0380	.0436	.0408	.0446
					0	1, 5, 4, 2, 3	.0473	.0539	.0537	.0474	.0506	.0527	.0572	.0659
					−.50	3, 5, 2, 4, 1	.0570	.0562	.0586	.0629	.0596	.0681	.0742	**.0874**
					−1	5, 4, 3, 2, 1	.0580	.0629	.0612	.0629	.0700	**.0818**	**.1031**	**.1216**
		5.7	0.50	21, 42, 70, 98, 119	1	1, 2, 3, 4, 5	.0356	.0389	.0332	.0296	.0282	**.0245**	**.0237**	**.0196**
					.50	1, 4, 2, 5, 3	.0414	.0418	.0410	.0401	.0384	.0383	.0345	.0356
					0	1, 5, 4, 2, 3	.0492	.0542	.0491	.0453	.0510	.0544	.0586	.0614
					−.50	3, 5, 2, 4, 1	.0588	.0574	.0581	.0631	.0731	**.0827**	**.0900**	**.1085**
					−1	5, 4, 3, 2, 1	.0666	.0721	**.0812**	**.0780**	**.0870**	**.1148**	**.1583**	**.1811**
400	80	1.5	0.16	64, 72, 80, 88, 96	1	1, 2, 3, 4, 5	.0469	.0428	.0437	.0478	.0455	.0403	.0424	.0440
					.50	1, 4, 2, 5, 3	.0455	.0487	.0487	.0484	.0491	.0486	.0530	.0523
					0	1, 5, 4, 2, 3	.0515	.0517	.0493	.0464	.0478	.0527	.0597	.0591
					−.50	3, 5, 2, 4, 1	.0495	.0555	.0492	.0535	.0554	.0571	.0638	.0675

Table 16 (continued)

N	N/J	$\frac{n_{Largest}}{n_{Smallest}}$	Δn	n	Pairing	Order of variance	1.5	1.6	1.7	1.8	2	3	5	9
					−1	5, 4, 3, 2, 1	.0538	.0529	.0558	.0582	.0600	.0657	.0717	**.0845**
		2.5	0.33	48, 64, 72, 96, 120	1	1, 2, 3, 4, 5	.0402	.0405	.0431	.0382	.0372	.0330	.0343	.0314
					.50	1, 4, 2, 5, 3	.0475	.0456	.0451	.0441	.0406	.0420	.0426	.0398
					0	1, 5, 4, 2, 3	.0500	.0487	.0476	.0499	.0550	.0524	.0611	.0669
					−.50	3, 5, 2, 4, 1	.0581	.0532	.0585	.0542	.0632	.0683	**.0820**	**.0839**
					−1	5, 4, 3, 2, 1	.0628	.0630	.0673	.0693	.0710	**.0905**	**.1024**	**.1257**
		5.7	0.50	24, 48, 80, 112, 136	1	1, 2, 3, 4, 5	.0333	.0363	.0325	.0316	.0289	**.0224**	**.0202**	**.0186**
					.50	1, 4, 2, 5, 3	.0426	.0455	.0401	.0404	.0386	.0364	.0357	.0403
					0	1, 5, 4, 2, 3	.0493	.0501	.0490	.0499	.0511	.0506	.0582	.0634
					−.50	3, 5, 2, 4, 1	.0619	.0529	.0620	.0620	.0701	**.0762**	**.0966**	**.1055**
					−1	5, 4, 3, 2, 1	.0661	.0688	**.0769**	**.0828**	**.0818**	**.1172**	**.1506**	**.1840**
450	90	1.5	0.16	72, 81, 90, 99, 108	1	1, 2, 3, 4, 5	.0441	.0439	.0473	.0484	.0460	.0431	.0464	.0432
					.50	1, 4, 2, 5, 3	.0494	.0483	.0463	.0462	.0475	.0507	.0525	.0521
					0	1, 5, 4, 2, 3	.0501	.0510	.0490	.0482	.0541	.0510	.0557	.0599
					−.50	3, 5, 2, 4, 1	.0507	.0567	.0531	.0505	.0572	.0641	.0660	.0739
					−1	5, 4, 3, 2, 1	.0550	.0599	.0574	.0605	.0599	.0711	.0737	**.0833**
		2.5	0.33	54, 72, 81, 108, 135	1	1, 2, 3, 4, 5	.0435	.0376	.0404	.0397	.0375	.0349	.0303	.0301
					.50	1, 4, 2, 5, 3	.0439	.0459	.0450	.0425	.0429	.0439	.0467	.0416
					0	1, 5, 4, 2, 3	.0456	.0493	.0555	.0552	.0503	.0530	.0609	.0647
					−.50	3, 5, 2, 4, 1	.0541	.0512	.0557	.0565	.0601	.0687	**.0794**	**.0874**
					−1	5, 4, 3, 2, 1	.0646	.0624	.0645	.0652	.0719	**.0900**	**.0995**	**.1215**
		5.7	0.50	27, 54, 90, 126, 153	1	1, 2, 3, 4, 5	.0353	.0352	.0329	.0349	.0300	**.0244**	**.0208**	**.0231**
					.50	1, 4, 2, 5, 3	.0443	.0446	.0430	.0381	.0386	.0336	.0391	.0390
					0	1, 5, 4, 2, 3	.0485	.0485	.0492	.0511	.0525	.0549	.0569	.0606
					−.50	3, 5, 2, 4, 1	.0585	.0611	.0584	.0585	.0684	**.0762**	**.0937**	**.1083**
					−1	5, 4, 3, 2, 1	.0666	.0686	.0743	**.0809**	**.0906**	**.1107**	**.1429**	**.1794**
500	100	1.5	0.16	80, 90, 100, 110, 120	1	1, 2, 3, 4, 5	.0460	.0434	.0428	.0432	.0472	.0454	.0450	.0408
					.50	1, 4, 2, 5, 3	.0477	.0462	.0470	.0481	.0497	.0490	.0503	.0546
					0	1, 5, 4, 2, 3	.0519	.0527	.0506	.0502	.0504	.0492	.0553	.0608
					−.50	3, 5, 2, 4, 1	.0533	.0579	.0516	.0542	.0506	.0603	.0679	.0747
					−1	5, 4, 3, 2, 1	.0529	.0550	.0574	.0560	.0625	.0707	.0724	**.0867**
		2.5	0.33	60, 80, 90, 120, 150	1	1, 2, 3, 4, 5	.0414	.0432	.0420	.0399	.0385	.0325	.0315	.0300
					.50	1, 4, 2, 5, 3	.0461	.0430	.0495	.0486	.0462	.0419	.0411	.0465
					0	1, 5, 4, 2, 3	.0507	.0478	.0483	.0525	.0531	.0517	.0591	.0589
					−.50	3, 5, 2, 4, 1	.0548	.0489	.0603	.0604	.0636	.0647	**.0779**	**.0792**
					−1	5, 4, 3, 2, 1	.0648	.0675	.0680	.0699	**.0753**	**.0866**	**.1042**	**.1180**
		5.7	0.50	30, 60, 100, 140, 170	1	1, 2, 3, 4, 5	.0377	.0361	.0315	.0322	.0294	.0302	**.0213**	**.0197**
					.50	1, 4, 2, 5, 3	.0428	.0448	.0410	.0385	.0434	.0352	.0375	.0331
					0	1, 5, 4, 2, 3	.0502	.0478	.0529	.0536	.0523	.0534	.0553	.0632
					−.50	3, 5, 2, 4, 1	.0630	.0639	.0623	.0642	.0651	**.0810**	**.0962**	**.1042**
					−1	5, 4, 3, 2, 1	.0694	.0738	**.0755**	**.0766**	**.0869**	**.1139**	**.1461**	**.1776**

Note. N = total sample size; N/J = mean of group sample size; $\frac{n_{Largest}}{n_{Smallest}}$ = ratio of the largest to the smallest group sample size; Δn = coefficient of sample size variation; n = number of units per group. The Type I error rates outside the boundary of Bradley's liberal criterion are in bold (conservative: <.025; liberal: >.075)

Appendix 3. Empirical Type I error rates for *F*-test with unequal sample sizes for several nominal alpha levels in the conditions under which it is not robust against heterogeneity for a nominal alpha level equal to .05

Table 17

N	N/J	$\frac{n_{Largest}}{n_{Smallest}}$	Δn	n	Pairing	Order of variance	Alpha	1.6	1.7	1.8	2	3	5	9
15	5	1.5	0.16	4, 5, 6	−.50	2, 3, 1	.05						**.0787**	**.0864**
							.025						.0390	.0590
							.01						**.0240**	**.0200**
					−1	3, 2, 1	.05						**.0816**	**.0948**
							.025						.0540	.0540
							.01						.0260	.0310
		2.3	0.33	3, 5, 7	−.50	2, 3, 1	.05						**.0873**	**.0977**
							.025						.0570	.0600
							.01						.0250	**.0230**
					−1	3, 2, 1	.05				**.0791**	**.0897**	**.1122**	**.1319**
							.025				.0413	.0520	**.0730**	**.1000**
							.01				**.0200**	**.0180**	.0430	.0450
		3	0.50	3, 3, 9	−1	3, 2, 1	.05		**.0760**	**.0786**	**.0794**	**.1171**	**.1624**	**.2077**
							.025		.0350	.0420	.0458	.0650	**.1080**	**.1390**
							.01		**.0140**	.0250	**.0170**	.0306	.0510	.0660
30	10	1.5	0.16	8, 10, 12	−1	3, 2, 1	.05						**.0788**	**.0867**
							.025						.0500	.0550
							.01						.0270	**.0200**
		2.3	0.33	6, 10, 14	−.50	2, 3, 1	.05						**.0870**	**.0913**
							.025						.0590	.0510
							.01						.0270	.0310
					−1	3, 2, 1	.05					**.0863**	**.1128**	**.1245**
							.025					.0535	.0630	**.0830**
							.01					**.0230**	.0310	.0430
		4	0.50	4, 10, 16	−.50	2, 3, 1	.05					**.0809**	**.0926**	**.1099**
							.025					.0466	.0610	.0550
							.01					.0280	.0310	.0380
					−1	3, 2, 1	.05	**.0761**	**.0757**	**.0761**	**.0881**	**.1185**	**.1486**	**.1788**
							.025	.0400	.0390	.0500	.0515	.0620	**.1040**	**.1300**
							.01	**.0190**	**.0230**	**.0170**	**.0190**	.0379	.0580	.0660
45	15	1.5	0.16	12, 15, 18	−1	3, 2, 1	.005						**.0807**	**.0861**
							.025						.0460	.0570
							.01						**.0220**	**.0190**
		2.3	0.33	9, 15, 21	−.50	2, 3, 1	.05						**.0794**	**.0876**
							.025						.0540	.0570
							.01						.0250	.0290
					−1	3, 2, 1	.05				**.0761**	**.0921**	**.1074**	**.1302**
							.025				.0399	.0502	**.0760**	**.1000**
							.01				**.0130**	**.0240**	.0310	.0340
		4	0.50	6, 15, 24	−.50	2, 3, 1	.05					**.0785**	**.0914**	**.1082**
							.025					.0481	.0680	.0720
							.01					**.0240**	.0340	.0380
					−1	3, 2, 1	.05		**.0788**	**.0803**	**.0916**	**.1132**	**.1497**	**.1737**
							.025			.0460	.0530	.0506	**.0830**	**.0900**	**.1100**

Table 17 (continued)

N	N/J	$\frac{n_{\text{Largest}}}{n_{\text{Smallest}}}$	Δn	n	Pairing	Order of variance	Alpha	1.6	1.7	1.8	2	3	5	9
							.01		**.0160**	**.0180**	.0250	.0374	.0660	.0520
60	20	1.5	0.16	16, 20, 24	−1	3, 2, 1	.05						**.0761**	**.0905**
							.025						.0540	.0630
							.01						**.0230**	**.0230**
		2.3	0.33	12, 20, 28	−.50	2, 3, 1	.05							**.0873**
							.025							.0490
							.01							.0330
					−1	3, 2, 1	.05				**.0765**	**.0949**	.1072	.1208
							.025				.0384	.0497	.0700	.0630
							.01				**.0190**	**.0200**	.0470	.0530
		4	0.50	8, 20, 32	−.50	2, 3, 1	.05				**.0764**	.0885	.0953	
							.025				.0462	.0680	.0630	
							.01				**.0200**	.0530	.0420	
					−1	3, 2, 1	.05			**.0767**	**.0911**	**.1137**	**.1452**	**.1739**
							.025			.0520	.0483	.0769	.1140	.1170
							.01			**.0200**	**.0270**	.0340	.0570	.0800
							.005							.0520
75	25	1.5	0.16	20, 25, 30	−1	3, 2, 1	.05							**.0885**
							.025							.0590
							.01							.0270
		2.3	0.33	15, 25, 35	−.50	2, 3, 1	.05							**.0873**
							.025							.0580
							.01							**.0200**
					−1	3, 2, 1	.05					**.0829**	**.1069**	**.1224**
							.025					.0494	.0620	.0730
							.01					.0250	.0450	.0550
		4	0.50	10, 25, 40	−.50	2, 3, 1	.05					**.0783**	**.0902**	**.1009**
							.025					.0489	.0670	.0570
							.01					.0270	.0400	.0390
					−1	3, 2, 1	.05	**.0759**	**.0764**	**.0806**	**.0900**	**.1101**	**.1420**	**.1737**
							.025	.0480	.0420	.0500	.0475	.0672	**.0880**	**.1340**
							.01	**.0140**	**.0180**	.0310	**.0230**	.0370	.0690	.0720
90	30	1.5	0.16	24, 30, 36	−1	3, 2, 1	.05						**.0762**	**.0859**
							.025						.0470	.0560
							.01						.0290	.0290
		2.3	0.33	18, 30, 42	−.50	2, 3, 1	.05						**.0755**	**.0812**
							.025						.0530	.0450
							.01						.0340	.0320
					−1	3, 2, 1	.05					**.0880**	**.1032**	**.1242**
							.025					.0523	**.0780**	**.0780**
							.01					.0220	.0320	.0460
		4	0.50	12, 30, 48	−.50	2, 3, 1	.05					**.0755**	**.0890**	**.1041**
							.025					.0463	**.0760**	.0590
							.01					**.0203**	.0400	.0410
					−1	3, 2, 1	.05		**.0756**	**.0796**	**.0930**	**.1085**	**.1400**	**.1748**
							.025		.0530	.0460	.0504	.0649	**.1120**	**.1100**
							.01		**.0240**	**.0190**	**.0230**	.0370	.0660	.0540

Table 17 (continued)

N	N/J	$\frac{n_{Largest}}{n_{Smallest}}$	Δn	n	Pairing	Order of variance	Alpha	1.6	1.7	1.8	2	3	5	9
120	40	1.5	0.16	32, 40, 48	−1	3, 2, 1	.05						**.0753**	**.0846**
							.025						.0450	.0530
							.01						.0270	.0290
		2.3	0.33	24, 40, 56	−.50	2, 3, 1	.05						**.0779**	**.0840**
							.025						.0590	.0480
							.01						.0300	.0270
					−1	3, 2, 1	.05					**.0903**	**.1061**	**.1206**
							.025					.0513	.0680	**.0770**
							.01					.0280	.0400	.0370
		4	0.50	16, 40, 64	−.50	2, 3, 1	.05					**.0774**	**.0933**	**.0967**
							.025					.0422	**.0770**	.0510
							.01					**.0210**	.0400	.0390
					−1	3, 2, 1	.05		**.0807**	**.0734**	**.0880**	**.1161**	**.1460**	**.1646**
							.025		.0470	.0340	.0518	.0643	**.1030**	**.1190**
							.01		**.0180**	**.0240**	**.0190**	.0290	.0540	.0630
150	50	1.5	0.16	40, 50, 60	−1	3, 2, 1	.05						**.0764**	**.0868**
							.025						.0460	.0440
							.01						**.0170**	.0260
		2.3	0.33	30, 50, 70	−.50	2, 3, 1	.05						**.0803**	**.0772**
							.025						.0510	.0530
							.01						.0270	.0260
					−1	3, 2, 1	.05					**.0879**	**.1061**	**.1258**
							.025					.0507	.0680	**.0900**
							.01					.0310	.0480	.0480
		4	0.50	20, 50, 80	−.50	2, 3, 1	.05					**.0793**	**.0881**	**.0987**
							.025					.0436	.0660	**.0820**
							.01					**.0190**	.0460	.0350
					−1	3, 2, 1	.05	**.0759**	**.0808**	**.0831**	**.0880**	**.1071**	**.1424**	**.1731**
							.025	.0410	.0400	.0370	.0471	.0706	**.1110**	**.1200**
							.01	**.0140**	**.0210**	**.0160**	**.0180**	.0330	.0510	.0660
180	60	1.5	0.16	48, 60, 72	−1	3, 2, 1	.05						**.0774**	**.0823**
							.025						.0520	.0410
							.01						**.0210**	.0250
		2.3	0.33	36, 60, 84	−.50	2, 3, 1	.05							**.0834**
							.025							.0600
							.01							.0270
					−1	3, 2, 1	.05					**.0907**	**.1039**	**.1190**
							.025					.0482	.0570	.0650
							.01					.0270	.0280	.0480
		4	0.50	24, 60, 96	−.50	2, 3, 1	.05					**.0762**	**.0888**	**.0991**
							.025					.0380	.0680	.0650
							.01					.0261	.0430	.0330
					−1	3, 2, 1	.05	**.0753**	**.0775**	**.0787**	**.0825**	**.1167**	**.1445**	**.1697**
							.025	.0410	.0370	.0460	.0503	**.0760**	**.0890**	**.1140**
							.01	**.0120**	**.0200**	.0250	**.0200**	.0420	.0610	.0380
210	70	1.5	0.16	56, 70, 84	−1	3, 2, 1	.05						**.0767**	**.0855**
							.025						.0460	.0430

Table 17 (continued)

N	N/J	$\frac{n_{\text{Largest}}}{n_{\text{Smallest}}}$	Δn	n	Pairing	Order of variance	Alpha	1.6	1.7	1.8	2	3	5	9
							.01						.0310	.0320
	2.3	0.33	42, 70, 98	−.50	2, 3, 1	.05							**.0759**
						.025							.0530
						.01							**.0220**
				−1	3, 2, 1	.05					**.0842**	.1074	.1224
						.025					.0590	.0580	**.0840**
						.01					.0290	.0420	.0380
	4	0.50	28, 70, 112	−.50	2, 3, 1	.05					**.0764**	.0866	.1019
						.025					.0402	.0590	.0680
						.01					**.0160**	.0430	.0440
				−1	3, 2, 1	.05		**.0769**	**.0762**	**.0843**	**.1141**	**.1470**	**.1639**
						.025		.0410	.0330	.0497	**.0780**	**.1130**	**.1000**
						.01		.0250	**.0160**	.0250	.0320	.0670	**.0840**
						.005							.0460
240	80	1.5	0.16	64, 80, 96	−1	3, 2, 1	.05							**.0821**
						.025							.0470
						.01							.0310
	2.3	0.33	48, 80, 112	−.50	2, 3, 1	.05							**.0866**
						.025							.0540
						.01							.0310
				−1	3, 2, 1	.05				**.0773**	**.0859**	**.1052**	**.1256**
						.025				.0429	.0600	.0660	**.0770**
						.01				**.0230**	.0300	.0470
	4	0.50	32, 80, 128	−.50	2, 3, 1	.05						**.0880**	**.0960**
						.025						.0720	.0640
						.01						.0390	.0390
				−1	3, 2, 1	.05		**.0782**	**.0762**	**.0908**	**.1133**	**.1429**	**.1687**
						.025		.0400	.0410	.0503	.0660	**.1010**	**.1110**
						.01		**.0150**	**.0130**	.0250	.0420	.0560	**.0770**
						.005							.0440
270	90	1.5	0.16	72, 90, 108	−1	3, 2, 1	.05						**.0837**	**.0878**
						.025						.0550	.0440
						.01						.0250	.0250
	2.3	0.33	54, 90, 126	−.50	2, 3, 1	.05							**.0827**
						.025							.0550
						.01							.0290
				−1	3, 2, 1	.05					**.0873**	**.1012**	**.1140**
						.025					.0620	**.0720**	**.0910**
						.01					.0330	.0340	.0520
	4	0.50	36, 90, 144	−.50	2, 3, 1	.05					**.0782**	**.0863**	**.0995**
						.025					.0470	.0704	.0630
						.01					.0390	.0390	.0490
				−1	3, 2, 1	.05		**.0786**	**.0816**	**.0882**	**.1075**	**.1397**	**.1629**
						.025		.0410	.0600	.0493	.0600	.0930	**.1250**
						.01		**.0200**	**.0180**	.0320	.0340	.0590	.0640
300	100	1.5	0.16	80, 100, 120	−1	3, 2, 1	.05							**.0853**
						.025							.0450

Table 17 (continued)

N	N/J	$\frac{n_{Largest}}{n_{Smallest}}$	Δn	n	Pairing	Order of variance	Alpha	1.6	1.7	1.8	2	3	5	9
							.01							**.0160**
300	100	2.3	0.33	60, 100, 140	−.50	2, 3, 1	.05						.0752	**.0867**
							.025						.0420	.0550
							.01						.0310	.0300
					−1	3, 2, 1	.05				**.0755**	**.0857**	**.1106**	**.1196**
							.025				.0377	.0500	.0670	.0640
							.01				**.0120**	.0250	.0400	.0510
		4	0.50	40, 100, 160	−.50	2, 3, 1	.05					**.0781**	**.0866**	**.0981**
							.025					.0430	.0610	**.0840**
							.01					**.0190**	.0450	.0360
					−1	3, 2, 1	.05		**.0782**	**.0821**	**.0816**	**.1078**	**.1375**	**.1713**
							.025		.0410	.0390	.0489	.0740	**.0780**	**.0830**
							.01		**.0180**	**.0200**	.0290	.0450	.0580	.0560

Note. N = total sample size; N/J = mean of group sample size; $\frac{n_{Largest}}{n_{Smallest}}$ = ratio of the largest to the smallest group sample size; Δn = coefficient of sample size variation; n = number of units per group. The Type I error rates outside the boundary of Bradley's liberal criterion for nominal alpha level of .05 are in bold (conservative: < .025; liberal: > .075)

References

Alexander, R. A., & Govern, D. M. (1994). A new and simpler approximation of ANOVA under variance heterogeneity. *Journal of Educational Statistics, 19,* 91–101. doi:10.2307/1165140

Angelis, D., & Young, G. A. (1998). Bootstrap method. *Encyclopedia of Biostatistics.*

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society, Series A, 160,* 268–282. doi:10.1098/rspa.1937.0109

Bendayan, R., Arnau, J., Blanca, M. J., & Bono, R. (2014). Comparison of the procedures of Fleishman and Ramberg et al. for generating non-normal data in simulation studies. *Anales de Psicología, 30,* 364–371. doi:10.6018/analesps.30.1.135911

Bhat, B. R., Badade, M. N., & Aruna Rao, K. (2002). A new test for equality of variances for *k* normal populations. *Communication in Statistics-Simulation and Computation, 31,* 567–587. doi:10.1081/SAC-120004313

Blanca, M. J., Arnau, J., López-Montiel, D., Bono, R., & Bendayan, R. (2013). Skewness and kurtosis in real data samples. *Methodology, 9,* 78–84. doi:10.1027/1614-2241/a000057

Box, G. E. P. (1953). Non-normality and tests on variances. *Biometrika, 40,* 318–335. doi:10.1093/biomet/40.3-4.318

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problem. I. Effect of inequality of variance in one-way classification. *The Annals of Mathematical Statistics, 25,* 290–302. doi:10.1214/aoms/1177728786

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology, 31,* 144–152. doi:10.1111/j.2044-8317.1978.tb00581.x

Brown, M. B., & Forsythe, A. B. (1974). The small sample behaviour of some statistics which test the equality of several means. *Technometrics, 16,* 129–132. doi:10.1080/00401706.1974.10489158

Brunner, E., Dette, H., & Munk, A. (1997). Box-type approximations in nonparametric factorial designs. *Journal of the American Statistical Association, 92,* 1494–1502. doi:10.1080/01621459.1997.10473671

Bryk, A. S., & Raudenbush, S. W. (1988). Heterogeneity of variance in experimental studies: A challenge to conventional interpretations. *Psychological Bulletin, 104,* 396–404. doi:10.1037/0033-2909.104.3.396

Büning, H. (1997). Robust analysis of variance. *Journal of Applied Statistics, 24,* 319–332. doi:10.1080/02664769723710

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine, 25,* 4279–4292.

Chen, S. Y., & Chen, H. J. (1998). Single-stage analysis of variance under heteroscedasticity. *Communications in Statistics – Simulation and Computation, 27,* 641–666. doi:10.1080/03610919808813501

Clinch, J. J., & Keselman, H. J. (1982). Parametric alternatives to the analysis of variance. *Journal of Educational Statistics, 7,* 207–214. doi:10.2307/1164645

Cochran, W. (1941). The distribution of the largest of a set of estimated variances as a fraction of their total. *Annals of Human Genetics, 11,* 47–52. doi:10.1111/j.1469-1809.1941.tb02271.x

Conover, W. J., Johnson, M. E., & Johnson, M. M. (1981). A comparative study of test for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics, 23,* 351–361. doi:10.2307/1268225

Cribbie, R. A., Fiksenbaum, L., Keselman, H. J., & Wilcox, R. R. (2012). Effect of non-normality on test statistics for one-way independent group designs. *British Journal of Mathematical and Statistical Psychology, 65,* 56–73. doi:10.1111/j.2044-8317.2011.02014.x

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.

David, F. N., & Johnson, N. L. (1951). The effect of non-normality on the power function of the *F*-test in the analysis of variance. *Biometrika, 38,* 43–57. doi:10.1093/biomet/38.1-2.43

Dean, A., & Voss, D. (1999). *Design and analysis of experiments.* New York: Springer-Verlag. doi:10.1007/b97673

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. doi:10.1037/0003-066X.63.7.591

Fan, W., & Hancock, G. R. (2012). Robust means modeling: An alternative for hypothesis testing of independent means under variance heterogeneity and nonnormality. *Journal of Educational and Behavioral Statistics, 37,* 137–156. doi:10.3102/1076998610396897

Gamage, J., & Weerahandi, S. (1998). Size performance of some tests in one-way ANOVA. *Communications in Statistics–Simulation and Computation, 27,* 625–640. doi:10.1080/03610919808813500

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research, 42,* 237–288. doi:10.3102/00346543042003237

Glass, G. V., & Stanley, J. C. (1970). *Statistical Methods in Education and Psychology.* Englewood Cliffs: Prentice-Hall.

Golinski, C., & Cribbie, R. A. (2009). The expanding role of quantitative methodologists in advancing psychology. *Canadian Psychology, 50,* 83–90. doi:10.1037/a0015180

Grissom, R. J. (2000). Heterogeneity of variance in clinical data. *Journal of Consulting and Clinical Psychology, 68,* 155–165. doi:10.1037/0022-006X.68.1.155

Grissom, R. J., & Kim, J. J. (2001). Review of assumptions and problems in the appropriate conceptualization of effect size. *Psychological Methods, 6,* 135–146. doi:10.1037//1082-989X.6.2.135

Hartley, H. O. (1950). The use of range in analysis of variance. *Biometrika, 37,* 271–280. doi:10.1093/biomet/37.3-4.271

Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing Monte Carlo results in methodological research: The one- and two-factor fixed effects ANOVA cases. *Journal of Educational and Behavioral Statistics, 17,* 315–339. doi:10.3102/10769986017004315

Hays, W. L. (1981). *Statistics* (3rd ed.). New York: Holt, Rinehart & Winston.

Horsnell, G. (1953). The effect of unequal group variances on the *F*-test for the homogeneity of group means. *Biometrika, 40,* 128–136. doi:10.2307/2333104

Hsu, P. L. (1938). Contribution to the theory of "Student's" *t*-test as applied to the problem of two samples. *Statistical Research Memoirs, 2,* 1–24.

James, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika, 38,* 324–329. doi:10.1093/biomet/38.3-4.324

Kang, Y., Harring, J. R., & Li, M. (2015). Reexamining the impact of nonnormality in two-group comparison procedures. *The Journal of Experimental Education, 83,* 147–174. doi:10.1080/00220973.2013.876605

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs: Prentice-Hall.

Keppel, G., Saufley, W. H., & Tokunaga, H. (1992). *Introduction to design and analysis: A student's handbook* (2nd ed.). New York: W. H. Freeman.

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Englewood Cliffs: Prentice Hall.

Keselman, H. J., Algina, J., Kowalchuk, R. K., & Wolfinger, R. D. (1999). A comparison of recent approaches to the analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology, 52,* 63–78. doi:10.1348/000711099158964

Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., … & Levin, J. R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research, 68,* 350–386. doi:10.3102/00346543068003350

Kieffer, K. M., Reese, R. J., & Thompson, B. (2001). Statistical techniques employed in *AERJ* and *JCP* articles from 1988 to 1997: A methodological review. *The Journal of Experimental Education, 69,* 280–309. doi:10.1080/00220970109599489

Kirk, R. E. (2013). *Experimental design. Procedures for the behavioral sciences* (4th ed.). Thousand Oaks: Sage Publications.

Kohr, R. L., & Games, P. A. (1974). Robustness of the analysis of variance, the Welch procedure and a Box procedure to heterogeneous variances. *The Journal of Experimental Education, 43,* 61–69. doi:10.1080/00220973.1974.10806305

Kowalchuk, R. K., Keselman, H. J., Algina, J., & Wolfinger, R. D. (2004). The analysis of repeated measurements with mixed-model adjusted *F* tests. *Educational and Psychological Measurement, 64,* 224–242. doi:10.1177/0013164403260196

Krishnamoorthy, K., Lu, F., & Mathew, T. (2007). A parametric bootstrap approach for ANOVA with unequal variances: Fixed and random models. *Computational Statistics & Data Analysis, 51,* 5731–5742. doi:10.1016/j.csda.2006.09.039

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association, 47,* 538–621. doi:10.1080/01621459.1952.10483441

Lee, S., & Ahn, C. H. (2003). Modified ANOVA for unequal variances. *Communications in Statistics–Simulation and Computation, 32,* 987–1004. doi:10.1081/SAC-120023874

Levene, H. (1960). Robust tests for equality of variance. In I. Olkin & H. Hotelling (Eds.), *Contributions to probability and statistics* (pp. 278–292). Palo Alto: Stanford University Press.

Li, X., Wang, J., & Liang, H. (2011). Comparison of several means: A fiducial based approach. *Computational Statistics and Data Analysis, 55,* 1993–2002. doi:10.1016/j.csda.2010.12.009

Lindquist, E. F. (1953). *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin.

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. doi:10.1177/0013164498058003004

Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance *F* test. *Review of Educational Research, 66,* 579–619. doi:10.3102/00346543066004579

Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison perspective.* Belmont: Wadsworth.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah: Lawrence Erlbaum Associates.

Mendes, M., & Pala, A. (2004). Evaluation of four tests when normality and homogeneity of variance assumptions are violated. *Journal of Applied Sciences, 4,* 38–42. doi:10.3923/jas.2004.38.42

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105,* 156–166. doi:10.1037/0033-2909.105.1.156

Mickelson, W. T. (2013). A Monte Carlo simulation of the robust rank-order test under various population symmetry conditions. *Journal of Modern Applied Statistical Methods, 12,* 21–33.

Moder, K. (2007). How to keep the Type I error rate in ANOVA if variances are heteroscedastic. *Austrian Journal of Statistics, 36,* 179–188.

Moder, K. (2010). Alternatives to *F*-test in one way ANOVA in case of heterogeneity of variances (a simulation study). *Psychological Test and Assessment Modeling, 52,* 343–353.

Montgomery, D. C. (1991). *Design and analysis of experiments* (3rd ed.). New York: Wiley.

Norton, D. W. (1952). *An empirical investigation of the effects of nonnormality and heterogeneity upon the F-test of analysis of*

variance. Unpublished doctoral dissertation, University of Iowa, Iowa City.

Parra-Frutos, I. (2014). Controlling the Type I error rate by using the nonparametric bootstrap when comparing means. *British Journal of Mathematical and Statistical Psychology, 67,* 117–132. doi:10.1111/bmsp.12011

Patrick, J. D. (2007). *Simulations to analyze Type I error and power in the ANOVA F test and nonparametric alternatives* (Master's thesis, University of West Florida). Retrieved from http://etd.fcla.edu/WF/WFE0000158/Patrick_Joshua_Daniel_200905_MS.pdf

Robey, R. R., & Barcikowski, R. S. (1992). Type I error and the number of iterations in Monte Carlo studies of robustness. *British Journal of Mathematical and Statistical Psychology, 45,* 283–288. doi:10.1111/j.2044-8317.1992.tb00993.x

Rogan, J. C., & Keselman, H. J. (1977). Is the ANOVA *F*-test robust to variance heterogeneity when sample sizes are equal? An investigation via a coefficient of variation. *American Educational Research Journal, 14,* 493–498. doi:10.3102/00028312014004493

Rogan, J. C., Keselman, H. J., & Breen, L. J. (1977). Assumption violations and rates of Type I error for the Tukey multiple comparison test: A review and empirical investigation via a coefficient of variance variation. *Journal of Experimental Education, 46,* 20–25. doi:10.1080/00220973.1977.11011605

Ruscio, J., & Roche, B. (2012). Variance heterogeneity in published psychological research: A review and a new index. *Methodology, 8,* 1–11. doi:10.1027/1614-2241/a000034

SAS Institute Inc. (2013). *SAS® 9.4 guide to software Updates*. Cary: SAS Institute Inc.

Sawilowsky, S. S. (2002). Fermat, Shubert, Einstein, and Behrens-Fisher: The probable difference between two means when $\sigma_1^2 \neq \sigma_2^2$. *Journal of Modern Applied Statistical Methods, 1,* 461–472.

Sawilowsky, S. S., & Blair, R. C. (1992). A more realistic look at the robustness and Type II error properties of the *t* test to departures from population normality. *Psychological Bulletin, 111,* 352–360. doi:10.1037/0033-2909.111.2.352

Scheffé, H. (1959). *The analysis of variance*. New York: Wiley.

Sharma, D., & Kibria, B. M. G. (2013). On some test statistics for testing homogeneity of variances: A comparative study. *Journal of Statistical Computation and Simulation, 83,* 1944–1963. doi:10.1080/00949655.2012.675336

Skidmore, S. T., & Thompson, B. (2010). Statistical techniques used in published articles: A historical review of reviews. *Educational and Psychological Measurement, 70,* 777–795. doi:10.1177/0013164410379320

Tabachnick, B. G., & Fidell, L. S. (2007). *Experimental design using ANOVA*. Belmont: Thomson Brooks/Cole.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston: Pearson Education, Inc.

Tomarken, A. J., & Serling, R. C. (1986). Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. *Psychological Bulletin, 99,* 90–99. doi:10.1037/0033-2909.99.1.90

Weerahandi, S. (1995). ANOVA under unequal error variances. *Biometrics, 51,* 589–599. doi:10.2307/2532947

Welch, B. L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika, 38,* 330–336. doi:10.1093/biomet/38.3-4.330

Wilcox, R. R. (1987). New designs in analysis of variance. *Annual Review of Psychology, 38,* 29–60. doi:10.1146/annurev.ps.38.020187.000333

Wilcox, R. R. (1995). ANOVA: The practical importance of heteroscedastic methods, using trimmed means versus means, and designing simulation studies. *British Journal of Mathematical and Statistical Psychology, 48,* 99–114. doi:10.1111/j.2044-8317.1995.tb01052.x

Wilcox, R. R., Charlin, V. L., & Thompson, K. L. (1986). New Monte Carlo results on the robustness of the ANOVA *F*, *W* and *F* statistics. *Communications in Statistics–Simulation and Computation, 15,* 933–943. doi:10.1080/03610918608812553

Wilcox, R. R., Keselman, H. J., & Kowalchuk, R. H. (1998). Can tests for treatment group equality be improved? The bootstrap and trimmed means conjecture. *British Journal of Mathematical and Statistical Psychology, 51,* 123–143. doi:10.1111/j.2044-8317.1998.tb00670.x

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York: McGraw-Hill.

Winner, B. J. (1971). *Statistical principles in experimental designs* (2nd ed.). New York: McGraw-Hill.

Wuensch, K. L. (2017). *One-way independent samples analysis of variance*. Retrieved from http://core.ecu.edu/psyc/wuenschk/docs30/anova1.pdf

Yiğit, E., & Gökpınar, F. (2010). A simulation study on tests for one-way ANOVA under the unequal variance assumption. *Communications Faculty of Sciences University of Ankara, Series, A1*(59), 15–34. doi:10.1501/Commua1_0000000660

Zijlstra, W. (2004). Comparing the Student's t and the ANOVA contrast procedure with five alternative procedures (Master's thesis, Rijksuniversiteit Groningen). Retrieved from http://www.ppsw.rug.nl/~kiers/ReportZijlstra.pdf

Zimmerman, D. W. (2004). A note on preliminary tests of equality of variances. *British Journal of Mathematical and Statistical Psychology, 57,* 173–181. doi:10.1348/000711004849222