DOCUMENT RESUME

ED 371 023 | TM 021 654

AUTHOR Hsiung, Tung-Hsing; Olejnik, Stephen

TITLE Type I Error Rates and Statistical Power for the James Second-Order Test and the Univariate F Test in Two-Way Fixed-Effects ANOVA Models under Heteroscedasticity and/or Nonnormality.

PUB DATE Apr 94

NOTE 26p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).

PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS *Analysis of Variance; *Computer Simulation; *Estimation (Mathematics); Interaction; Models; *Robustness (Statistics); Sample Size; Statistical Distributions; Statistical Studies

IDENTIFIERS Factorial Designs; F Test; Nonnormal Distributions; *Power (Statistics); *Type I Errors; Univariate Analysis; Variance (Statistical)

ABSTRACT

This study investigated the robustness of the James second-order test (James 1951; Wilcox, 1989) and the univariate F test under a two-factor fixed-effect analysis of variance (ANOVA) model in which cell variances were heterogeneous and/or distributions were nonnormal. With computer-simulated data, Type I error rates and statistical power for the two tests were estimated. With data sampled from normal distributions, the F test was not robust to variance heterogeneity for equal or unequal sample sizes, but the James second-order test was robust in these situations. With normal distributions, equal variances, and equal sample sizes, the magnitude of power difference between the two tests was generally small when testing the main effects, but the magnitude of power difference between the two tests varied when testing the interaction effects. With data sampled from nonnormal distributions, although the James second-order test generally was liberal when the population distribution was skewed, the test was robust under several nonnormal distribution situations. Additionally, the robustness of the James second-order test in factorial designs may be affected by combinations of nonnormal distributions, sample sizes, and variance patterns. (Contains 22 references and 7 tables.) (Author/SLD)

Type I Error Rates and Statistical Power for the James Second-Order Test and the Univariate F Test in Two-Way Fixed-Effects ANOVA Models Under Heteroscedasticity and/or Nonnormality

Tung-Hsing Hsiung
Performance Assessment Laboratory
Stephen Olejnik
Department of Educational Psychology
The University of Georgia

Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA (April 5, 1994).

Abstract

This study investigated the robustness of the James second-order test (James, 1951; Wilcox, 1989) and the univariate F test under a two-factor fixed-effect ANOVA model where cell variances were heterogeneous and/or distributions were non-normal. Using computer simulated data (SAS/IML [1989]), Type I error rates and statistical power for the two tests were estimated. With data sampled from normal distributions, the F test was not robust to variance heterogeneity for equal or unequal sample sizes, but the James second order test was robust in these situations. With normal distributions, equal variances, and equal sample sizes, the magnitude of power difference between the two tests was generally small when testing the main effects, but the magnitude of power difference between the two tests varied when testing the interaction effects. With data sampled from non-normal distributions, although the James second-order test generally was liberal when the population distribution was skewed, the current study showed that the test was robust under several non-normal distribution situations. Additionally, the robustness of the James second-order test in factorial designs may be affected by combinations of non-normal distributions, sample sizes, and variance patterns. (The F test was not examined under non-normal distributions because the F test does not provide a valid test for many heterogeneous variance situations.)

A number of studies have investigated the robustness of omnibus tests when testing the equality of K means under variance heteroscedasticity and/or distribution non-normality. The univariate F-test, the Brown and Forsythe (1974) F^*-test, the Welch (1951) test, and the James (1951) second-order test are the omnibus tests most frequently considered. Earlier studies that dealt with the validity of omnibus tests under variance heterogeneity and/or distribution non-normality include Brown and Forsythe (1974), Clinch and Keselman (1982), Wilcox, Charlin, and Thompson (1986), and Oshima and Algina (1992a). These studies showed that neither the F-test nor the alternatives adequately control the Type I error rate under the nominal significance level when extreme violations of the variance equality and/or normality occur.

Wilcox (1988) proposed a new alternative, H, which was computationally simpler than the James second-order test. Wilcox showed that although the H test has properties comparable to the James second-order test, it was slightly less powerful the James second-order test. Wilcox (1989) proposed a modification of the H-test, H_m, which was shown to provide statistical power more comparable to the James second-order test. Oshima and Algina (1992a) pointed out that the Wilcox (1988) study focused on the effect of variance heterogeneity for both the James second-order test and the H test when sampling from normal distributions. Non-normality was studied but not in combination with variance heterogeneity. They argued that not considering the impact of the combined violations of variance homogeneity and distribution normality was an important omission. Their investigation of the robustness of the James second-order test and Wilcox H_m test under heteroscedasticity and non-normality revealed that the empirical Type I error rates for both tests were affected when variance homogeneity and distribution normality both were violated. They also indicated that the magnitude of difference between the empirical Type

I error rate and nominal α level is positively related to the degree of asymmetry; the greater the degree of asymmetry, the greater difference between the empirical Type I error rate and nominal α level.

While most of the investigations into the robustness of ANOVA have concentrated on the one-factor design, Milligan, Wong, and Thompson (1987) investigated the robustness properties of nonorthogonal two-way fixed-effect ANOVA models. They concluded that each of the standard computational routines of ANOVA for unequal cell size was not robust to the assumptions of variance homogeneity or normality. When sample sizes were equal, however, they found that violating the homogeneity of variance assumption had little effect on the actual Type I error rate. Although they suggested four alternatives for dealing with unbalanced designs with variance heterogeneity or non-normal distributions, Keppel (1991, p. 283) stated that none of these alternatives is as effective as avoiding unequal sample sizes in the first place. Nonetheless, this alternative is often not an option in applied research where unbalanced designs are common.

Wilcox (1989) generalized his H_m test for situations involving a factorial structure, the U test. After comparing the robustness properties of the U test and the James second-order test under various heterogeneous variance conditions, Wilcox (1989) concluded that both the U test and the James second-order tests (a) performed well under null conditions and that they generally controlled the Type I error rate under the nominal α level; (b) provided sufficient power under non-null conditions; and (c) can be extended to higher-order designs. Hsiung, Olejnik, and Huberty (1994), however, showed that the U test does not adequately control the Type I error rate when the sample sizes are unequal and population means differ from zero. Therefore, Hsiung et. al concluded that the U test is invalid for most practical situations and recommended the James second-order test for factorial designs.

After conducting a meta-analysis on the robustness of ANOVA to variance heterogeneity, Harwell, Rubinstein, Hayes, and Olds (1992) concluded that there is an absence of well-documented omnibus tests that can be applied to two-factor fixed-effects ANOVA cases. They advised that there is a need for an investigation into the robustness of available omnibus tests in two-factor ANOVA models. Responding to this call for further study of two-factor fixed-effect ANOVA models, the current investigation examines the robustness of the F-test and the James second-order test under heteroscedasticity and/or non-normality. Oshima and Algina (1992a) had shown that the James second-order test was affected by asymmetric distributions in a single factor design, but they only included two asymmetrical non-normal distributions (i.e., the Beta and the Exponential distributions). Moreover, the Exponential non-normal distribution is not common in applied research. Fleishman (1978) indicated that the "typical" non-normal empirical distributions are with the degree of skew less than 0.8 and the magnitude of kurtosis between -0.6 and +0.6. The current study, therefore, examines robustness of the univariate F-test and the James second-order test in two factor designs with data sampled from more typical non-normal distributions.

Method

The present study included five two-factor fixed-effect ANOVA models: 2×2, 2×3, 3×3, 3×4, and 4×4. Each model was studied under at least six conditions with each condition defined by sample sizes, population variances, and population distributions. Not all models were included for all conditions. Fifteen population distributions were considered; each population distribution was defined by the degrees of skew and kurtosis.

Twenty-six variance patterns were selected; each variance pattern consisted of different cell variances (Table 1 lists the characteristics of the 26 variance patterns). The sample sizes, variance patterns, magnitude of skew, and magnitude of kurtosis are reported in Tables 2 to 8 along with the results.

Insert Table 1 About Here

The present study used SAS/IML (SAS Inc, 1989) software to generate the data and compute the test statistics. Using the SAS-RANNOR function, scores for each cell were generated independently, $Y_{ijk} \sim (\mu_{jk}, \sigma_{jk}^2)$. Each population mean equaled 0 under the null conditions and the cell$_{11}$ mean equaled δ under the non-null conditions. Using the Fleishman (1978) transformation procedure, data were transformed to have a distribution with the target degrees of skew and kurtosis. For each condition, 10,000 replications were generated and the proportion of times the omnibus tests were rejected at the $\alpha = .05$ level was recorded. A test was concluded liberal if its empirical Type I error rate exceeded .0544 (i.e., greater than the two standard errors of the nominal significance level).

For each replication, the data were analyzed by using the univariate F-test and the James second-order test. For the James second-order test formula refer to Wilcox [1989]; for the univariate F-test the unweighted means solution (regression approach) was used.

Results

Tables 2 and 3 present the results for the F-test and the James second-order test based on small (average cell size equals 5, Table 2) and large (average cell size equals 25, Table 3) sample sizes when sampling from normal population distributions. Balanced, slightly unbalanced, and extremely unbalanced designs were considered. Each table includes three variance patterns with the coefficient of variance variation (Keselman & Rogan, 1978) ranging between 0 and 1.18.

Insert Tables 2 and 3 About Here

Results from Tables 2 and 3 reveal that, under heteroscedasticity, the F-test can have empirical Type I error rates greater than the nominal significance level even when sample sizes are equal. These results contradict Milligan, Wong, and Thompson (1987), who concluded that the F-test is valid under heterogenous variances when sample sizes are equal. The present results support Wilcox's (1987) cautionary note that, while equal sample sizes may reduce the effect of heterogeneous variance on the Type I error rate, the F-test may still be liberal if the degree of variance heterogeneous is great. In the present study, the small sample sizes variance ratio of 3:1 was sufficient to invalidate the univariate F-test.

With unequal sample sizes and unequal variances, the F-test can be either conservative or liberal depending upon the relationship between the patterns of heterogeneity and the sample sizes. This relationship has been shown repeatedly in previous research on the effect of variance heterogeneity on ANOVA Type I error rates. It has been suggested that the effect of variance heterogeneity in unbalanced designs can be reduced if sample sizes are large (e.g., Maxwell and Delaney [1990, p. 110]). The results presented here support that belief to degree, but even with relatively large samples with extreme sample size inequality, the F-test had empirical Type I error rates less than the nominal significance level. These results support Wilcox's (1987) position that it is difficult to know how large a sample size is needed to reduce the effects of unequal variances.

The James second-order test had the Type I error rates that ranged between .0456 and .0530 when sample sizes were large and ranged between .0420 and .0544 when sample sizes were small across both balanced and unbalanced designs. These results support the conclusion that the James second-order test is robust to variance heterogeneity for equal or unequal sample sizes when the population distributions are normal.

Table 4 presents the empirical power estimates for the univariate F-test and the James second-order test when sampling from normal population distributions with equal variance and equal sample sizes. Results show that for many of the hypotheses tests, the James second-order test is only slightly less powerful than the univariate F-test. The power difference between the two tests is in the range of magnitude from .000 to .052 when testing main effects and is in the range of magnitude from .000 to .202 when testing interaction effects.

Insert Table 4 About Here

Results show that when testing the main effects, the magnitude of power difference between the two tests was generally small. However, when testing the interaction effects, the magnitude of power difference between the two tests varied. The magnitude of the power difference depends on the number of interaction contrasts that are conducted. Wilcox (1989) suggested using the Bonferroni procedure to adjust the nominal α level for each contrast (i.e., $\alpha' = \alpha$ / [min(J, K) - 1]). This approach reduces the statistical power of the James second-order test if the minimum of (J, K) $\geq$ 3. Using the Holland-

Copenhaver (1987) enhancement to the Bonferroni procedure would likely reduce the power difference between the F-test and the James second-order test.

Results for the James second-order test when data were sampled from non-normal distributions are reported in Tables 5 through 7. Univariate F-test results are not included in these tables since, as shown previously, the F-test does not provide a valid test for many situations where variances are heterogeneous.

Table 5 presents the results for the James second-order test for the two by four fixed-effects ANOVA model. A total of 72 conditions were considered; each condition was defined by the sample size (design type), distribution type, and variance pattern. Three distributions were considered. They were (a) normal distribution, (b) positively skewed-leptokurtic non-normal distribution (skew = 1.75 and kurtosis = 3.75), and (c) platykurtic non-normal distribution (skew = 0, kurtosis = -1.0).

Consistent with Tables 2 and 3, the James second-order test was valid when the assumption of normality was met. But when data were sampled from a population distribution that was skewed, the James second-order test frequently had Type I error rates greater than the nominal significance level. With the same non-normal distribution, the test was more liberal when sample sizes were extremely unequal than when sample sizes were equal or slightly unequal.

When data were sampled from the platykurtic non-normal distribution, the James second-order test was robust when sample sizes were equal or slightly unequal, but appeared to be liberal when sample sizes were extremely unequal.

Although the James second-order test may be liberal when the assumption of normality is violated, the current results show that the test is robust under several non-normal distribution situations. It appears that in a factorial design, the robustness of the

James second-order test may be affected by combinations of non-normal distributions, sample sizes, and variance patterns.

Table 6 presents the results for the James second-order test under a four by four fixed-effect balanced factorial design. A total of 72 conditions were included with each condition being defined by the sample size, variance pattern, degree of skew, and degree of kurtosis.

Insert Table 6 About Here

When the assumption of variance homogeneity was met, but normality was violated, the James second-order test had empirical Type I error rates that did not exceed two standard errors above the nominal significance level. However, the James second-order test was conservative when sample sizes were small - empirical Type I error rates were generally less than two standard errors below the nominal significance level.

The James second-order test appeared to be liberal when the degree of skew was equal to or greater than 1.0. Yet, the patterns of sample size, variance, and degree of kurtosis also had some effect on the robustness of the test.

With the same degree of skew, the current results show that the James second-order test had greater Type I error rates when the degree of kurtosis was small than when the degree of kurtosis was large.

Table 7 presents the results for the James second-order test under two by four balanced and unbalanced fixed-effect designs. Data were sampled from 12 non-normal distributions; each distribution was defined by the degrees of skew and kurtosis.

Insert Table 7 About Here

When the variances were equal and the degree of skew was less than 1.50, the James second-order test generally controlled the Type I error rate under the nominal criterion α level. But the James second-order test appeared to be liberal when the degree of skew was equal to 1.50. With the same degree of skew, the test generally had greater Type I error rate when the degree of kurtosis was small than when the degree of kurtosis was large. Finally, it appears that a balanced design might reduce the effect of skewed distributions somewhat. However, as demonstrated in Tables 6 and 7, a balanced design cannot be relied on to provide a valid test when distributions are skewed.

Conclusions

Contrary to what some believe (Milligan, Wong, & Thompson, 1987), the univariate F-test for a factorial design is not robust to the violation of the equal variance assumption when sample sizes are equal. The present study shows that the actual Type I error rate for the F-test can exceed the nominal significance level when sample sizes are equal but cell variances differ by as small as a 3 to 1 ratio. The James second-order test, on the other hand, control the actual risk of a Type I error under the nominal significance level ($\alpha = .05$) when sampled populations have normal distributions. Further, the study provides some evidence indicating that when all parametric assumptions are met, the James second-order test provides statistical power comparable to the univariate F-test at least for hypotheses on main effects. Considerably lower power might be obtained for the interaction test depending on the dimensions of the factorial structure. The lower power can be attributed to the use, in the present study, of the Bonferroni adjustment for multiple hypothesis tests. If one of the enhancements to the Bonferroni method was used, the power difference between the univariate test and the James second-order test, however, would be reduced. In addition, if an omnibus test is of interest and it is reasonable to assume normal population distributions, the F-test should be abandoned in favor of the James second-order test.

Micceri (1989) reported that sampling from normal population distributions may be the exception rather than the rule in educational research. With both skewed-leptokurtic and platykurtic distributions, the James second-order test may not adequately control the risk of a Type I error to the nominal significance level. The degree of non-normality, variance heterogeneity, and the inequality of sample sizes all can affect the actual risk of a Type I error rate. The results of the present study did not make clear the exact relationship among these three factors. However, it did appear that having equal sample sizes can reduce the effect of non-normal distributions and heterogeneous variances on the Type I error rate.

Finally, Keppel (1991, p. 105) indicated that the James second-order test, the currently favored procedure, is "simply too complicated for general use." Recently, Oshima and Algina (1992b) developed a SAS/IML program for one-factor designs and Hsiung, Olejnik, and Oshima (1994) developed a SAS/IML program for two-factor fixed-effect designs. Lix and Keselman (1994) have developed a more general program to compute approximate degrees of freedom tests for both univariate and multivariate omnibus tests as well as tests for contrasts. With these computer programs available for application of alternative tests to the univariate F-test, the "disadvantage" of being computationa" · intense should not be a limitation.

References

Brown, M. B., & Forsythe, A. B. (1974). The small sample behavior of some statistics which test the equality of several means. Technometrics, 16, 129-132.

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

Harwell, M. R., Rubinstein, E. N., Hayes, W. S., and Olds, C. C. (1992). Summarizing Monte Carlo results in methodological research: The one- and two-factor fixed-effects ANOVA cases. Journal of Educational Statistics, 17, 297-313.

Holland, B. S., & Copenhaver, M. D. (1987). An improved sequentially rejective Bonferroni test procedure. Biometrics, 43, 417-423.

Hsiung, T., & Olejnik, S. (1993). Contrast analyses for the additive nonorthogonal two-factor design in the unequal variance case. Paper presented at the annual meeting of the American Educational Research Association, Atlanta.

Hsiung, T., Olejnik, S., Huberty, C. (1994). Comment on a Wilcox test statistic for comparing means when variances are unequal. Journal of Educational Statistics (in press).

Hsiung, T., Olejnik, S., Oshima, T. C. (1994). A SAS/IML program for applying the James second-order test in two-factor fixed-effect ANOVA models. Manuscript submitted for publication.

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

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

Keselman, H. J., & Rogan, J. C. (1978). A comparison of the modified Tukey and Scheffé methods of multiple comparisons for pairwise contrasts. Journal of the American Statistical Association, 73, 47-52.

Lix, L. M., & Keselman, H. J. (1994). Approximate degrees of freedom tests: A unified perspective on testing for mean equality. Manuscript submitted for publication.

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

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

Milligan, G. W., Wong, D. S., & Thompson P. A. (1987). Robustness properties of nonorthogonal analysis of variance. Psychological Bulletin, 101, 464-470.

Oshima, T. C., & Algina, J. (1992a). Type I error rates for James's second-order test and Wilcox's Hm test under heteroscedasticity and non-normality. British Journal of Mathematical and Statistical Psychology, 45, 225-263.

Oshima, T. C., & Algina, J. (1992b). A SAS program for testing the hypothesis of the equal means under heteroscedasticity: James's second-order test. Educational and Psychological Measurement, 52, 117-118.

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

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

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

Wilcox, R. R. (1988). A new alternative to the ANOVA F and on James's second-order method. British Journal of Mathematical and Statistical Psychology, 41, 109-117.

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

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. Communication in Statistics - Simulation and Computation, 15, 933-943.

Table 1.

Summary Table for the Characteristics Variance Patterns Considered.

Variance Pattern				
Pattern	Within Row	Cross Average Row	Within Column	Cross Average Column
1	Equal	Equal	Equal	Equal
2[a]	Equal	Unequal	Unequal	Equal
3[b]	Unequal	Equal	Equal	Unequal
4[c]	Unequal	Equal	Unequal	Equal
5[d]	Unequal	Equal	Unequal	Unequal
6[e]	Unequal	Unequal	Unequal	Equal
7[f]	Unequal	Unequal	Unequal	Unequal

Note. [a]For examples, Tables 2 and 3: all unequal variance patterns, Tables 5 and 8: (1, 1, 1, 1 ; 9, 9, 9, 9), and Table 6: (1, 1, 1, 1 ; 4, 4, 4, 4 ; 16, 16, 16, 16).

[b]For examples, Table 5: (16, 9, 4, 1 ; 16, 9, 4, 1) and Tables 6 and 7: (1, 4, 9, 16 ; 1, 4, 9, 16 ; 1, 4, 9, 16).

[c]For examples, Table 5: (4, 4, 1, 1 ; 1, 1, 4, 4) and Table 7: (1, 16, 9, 4 ; 4, 1, 16, 9 ; 9, 4, 1, 16 ; 16, 9, 4, 1).

[d]For example, Tables 5 and 8: (16, 9, 4, 1 ; 1, 4, 9, 16).

[e]For example, Table 5: (16, 14, 12, 10 ; 2, 4, 6, 8).

[f]For examples, Table 5: (16, 14, 12, 10 ; 8, 6, 4, 2) and Table 6: (1, 4, 9, 16 ; 16, 13, 8, 1 ; 4, 9, 16, 1).

Table 2

Type I Error Rate for the Univariate F test and the James Second-Order Test in Balanced/Unbalanced Two by Two, Two by Three, or Three by Three Fixed-Effect Factorial Designs with Small Sample Sizes

			F test			James Test		
			Type of Design			Type of Design		
Sample Size	Variance	Factor	BA[a]	SU[b]	EU[c]	BA[a]	SU[b]	EU[c]
2 × 2	1 , 1 1 , 1	A_{Row}	.0479	.0552	.0499	.0440	.0453	.0555
[a]BA: Balanced Design (5, 5 ; 5, 5)		B_{Col}	.0479	.0534	.0481	.0445	.0460	.0544
		A×B	.0471	.0575	.0501	.0439	.0476	.0529
[b]SU: Slightly Unbalanced (6, 4 ; 4, 6)	3 , 1 3 , 1	A_{Row}	.0518	.0552	.0136	.0458	.0481	.0531
[c]EU: Extremely Unbalanced (8, 2 ; 7, 3)		B_{Col}	.0531	.0534	.0142	.0465	.0473	.0496
		A×B	.0545	.0575	.0112	.0476	.0487	.0499
	16, 1 16, 1	A_{Row}	.0628	.0629	.0045	.0493	.0479	.0481
		B_{Col}	.0572	.0601	.0029	.0421	.0445	.0480
		A×B	.0626	.0655	.0047	.0453	.0493	.0496
2 × 3	1, 1, 1 1, 1, 1	A_{Row}	.0506	.0499	.0505	.0472	.0481	.0538
[a]BA: Balanced Design (5, 5, 5 ; 5, 5 ,5)		B_{Col}	.0486	.0498	.0581	.0426	.0428	.0532
		A×B	.0479	.0523	.0557	.0423	.0448	.0537
[b]SU: Slightly Unbalanced (6, 5, 4 ; 7, 5, 3)	3, 1, 1 3, 1, 1	A_{Row}	.0507	.0277	.0125	.0450	.0451	.0463
[c]EU: Extremely Unbalanced (8, 5, 2 ; 9, 4, 2)		B_{Col}	.0572	.0278	.0108	.0422	.0449	.0483
		A×B	.0607	.0258	.0092	.0445	.0452	.0446
	16, 1, 1 16, 1. 1	A_{Row}	.0628	.0193	.0014	.0467	.0495	.0440
		B_{Col}	.0869	.0259	.0019	.0437	.0416	.0431
		A×B	.0850	.0276	.0020	.0430	.0434	.0450
3 × 3	1, 1, 1 1, 1, 1 1, 1, 1	A_{Row}	.0445	.0488	.0464	.0395	.0421	.0407
[a]BA: Balanced Design (5, 5, 5; 5, 5, 5; 5, 5 ,5)		B_{Col}	.0473	.0517	.0558	.0436	.0442	.0420
		A×B	.0496	.0503	.0532	.0377	.0390	.0514
[b]SU: Slightly Unbalanced (6, 5, 4; 7, 5, 3; 7, 4, 4)	3, 1, 1 3, 1. 1 3, 1, 1	A_{Row}	.0524	.0226	.0091	.0449	.0439	.0408
[c]EU: Extremely Unbalanced (8, 5, 2; 9, 4, 2; 8, 4, 3)		B_{Col}	.0578	.0240	.0092	.0450	.0392	.0423
		A×B	.0606	.0212	.0061	.0365	.0370	.0484
	16, 1, 1 16, 1, 1 16, 1, 1	A_{Row}	.0665	.0104	.0005	.0494	.0470	.0453
		B_{Col}	.0845	.0175	.0013	.0451	.0427	.0370
		A×B	.0966	.0151	.0009	.0427	.0379	.0449

Note. Data were sampled from normal distributions. Shading indicates that the value is greater than the criterion .0544 and the test has an inflated Type I error rate.

Table 3

Type I Error Rate for the Univariate F test and the James Second-Order Test in Balanced/Unbalanced Two by Two, Two by Three, or Three by Three Fixed-Effect Factorial Designs with Large Sample Sizes

			F Test			James Test		
			Type of Design			Type of Design		
Sample Size	Variance	Factor	BA[a]	SU[b]	EU[c]	BA[a]	SU[b]	EU[c]
2 × 2	1 , 1 / 1 , 1	A_{Row}	.0517	.0516	.0497	.0514	.0518	.0484
[a]BA: Balanced Design (25, 25 ; 25, 25)		B_{Col}	.0528	.0527	.0522	.0526	.0525	.0520
		A×B	.0463	.0519	.0495	.0462	.0515	.0512
[b]SU: Slightly Unbalanced (27, 23 ; 26, 24)	3 , 1 / 3 , 1	A_{Row}	.0463	.0452	.0132	.0456	.0518	.0479
[c]EU: Extremely Unbalanced (35, 15 ; 37, 13)		B_{Col}	.0511	.0429	.0150	.0505	.0485	.0475
		A×B	.0529	.0448	.0152	.0518	.0501	.0536
	16, 1 / 16, 1	A_{Row}	.0520	.0410	.0029	.0490	.0499	.0481
		B_{Col}	.0526	.0428	.0039	.0497	.0528	.0529
		A×B	.0539	.0394	.0024	.0514	.0483	.0488
2 × 3	1, 1, 1 / 1, 1, 1	A_{Row}	.0459	.0497	.0531	.0458	.0498	.0532
[a]BA: Balanced Design (25, 25, 25 ; 25, 25 ,25)		B_{Col}	.0510	.0506	.0489	.0500	.0512	.0484
		A×B	.0520	.0485	.0513	.0525	.0492	.0489
[b]SU: Slightly Unbalanced (28, 25, 23 ; 27, 26, 22)	3, 1, 1 / 3, 1, 1	A_{Row}	.0476	.0459	.0293	.0466	.0527	.0511
[c]EU: Extremely Unbalanced (32, 25, 18 ; 30, 26, 17)		B_{Col}	.0544	.0503	.0275	.0466	.0527	.0511
		A×B	.0554	.0416	.0292	.0485	.0459	.0488
	16, 1, 1 / 16, 1, 1	A_{Row}	.0528	.0385	.0165	.0500	.0527	.0485
		B_{Col}	.0760	.0555	.0287	.0491	.0496	.0476
		A×B	.0824	.0563	.0299	.0525	.0470	.0524
3 × 3	1, 1, 1 / 1, 1, 1 / 1, 1, 1	A_{Row}	.0492	.0548	.0505	.0488	.0533	.0480
[a]BA: Balanced Design 20, 20, 20 / 20, 20, 20 / 20, 20, 20		B_{Col}	.0515	.0467	.0526	.0495	.0468	.0492
		A×B	.0505	.0521	.0562	.0491	.0477	.0460
[b]SU: Slightly Unbalanced 23, 19, 18 / 21, 22, 17 / 22, 19, 19	3, 1, 1 / 3, 1, 1 / 3, 1, 1	A_{Row}	.0506	.0418	.0126	.0487	.0515	.0487
		B_{Col}	.0543	.0451	.0131	.0491	.0526	.0476
		A×B	.0590	.0426	.0081	.0481	.0462	.0487
[c]EU: Extremely Unbalanced 32, 16, 12 / 29, 20, 11 / 30, 21, 9	16, 1, 1 / 16, 1, 1 / 16, 1, 1	A_{Row}	.0555	.0321	.0012	.0536	.0498	.0524
		B_{Col}	.0756	.0546	.0049	.0483	.0484	.0500
		A×B	.0857	.0560	.0022	.0478	.0453	.0474

Note. Data were sampled from normal distributions. Shading indicates that the value is greater than the criterion .0544 and the test has an inflated Type I error rate.

Table 4

Statistical Power for the Univariate F test and the James Second-Order Test in Balanced Two by Two, Two by Three, or Three by Three Fixed-Effect Factorial Designs

Sample Size	Factor	F test	James Test	Power Difference (F test - James Test)
5, 5 5, 5	A_{Row}	.7458	.7314	.0144
	B_{Col}	.7438	.7317	.0121
	A×B	.7483	.7347	.0136
5, 5, 5 5, 5. 5	A_{Row}	.5979	.5855	.0124
	B_{Col}	.7772	.7245	.0527
	A×B	.7742	.7250	.0492
5, 5, 5 5, 5, 5 5, 5, 5	A_{Row}	.6136	.5723	.0413
	B_{Col}	.6133	.5703	.0430
	A×B	.8167	.6146	.2021
25, 25 25, 25	A_{Row}	.5924	.5919	.0005
	B_{Col}	.6096	.6091	.0005
	A×B	.5997	.5992	.0005
25, 25, 25 25, 25, 25	A_{Row}	.4492	.4489	.0003
	B_{Col}	.6292	.6245	.0047
	A×B	.6214	.6187	.0027
20. 20. 20 20, 20, 20 20. 20. 20	A_{Row}	.3768	.3720	.0048
	B_{Col}	.3739	.3671	.0068
	A×B	.5254	.4078	.1176

Note. Data were sampled from normal distributions. The true group mean difference was created by adding a constant to each observation of the first cell (i.e., $Cell_{\_11}$); the constant δ was set equal to 2.5 for n_{jk} = 5 and was set equal to 0.9 for n_{jk} = 25.

Table 5

Type I Error Rate for the James Second-Order Test in Balanced/Unbalanced Two by Four Fixed-Effect Factorial Designs with Normal/Non-Normal Distributions and Homogeneous/Heterogeneous Variances

		Balanced Design 15, 15, 15, 15 15, 15, 15, 15			Slightly Unbalanced 18, 16, 14, 12 17, 16, 14, 13			Extremely Unbalanced 22, 18, 12, 8 24, 20, 10, 6		
		Distribution			Distribution			Distribution		
Variance Pattern	Factor	Normal	Skew[a]	Platy[b]	Normal	Skew	Platy	Normal	Skew	Platy
1,1,1,1 1,1,1,1	A_{Row}	.0471	.0473	.0501	.0500	.0458	.0497	.0447	.0447	.0489
	B_{Col}	.0515	.0529	.0545	.0480	.0559	.0517	.0492	.0657	.0533
	A×B	.0502	.0450	.0518	.0482	.0381	.0498	.0482	.0378	.0516
1,1,9,9 1,1,9,9	A_{Row}	.0510	.0461	.0508	.0481	.0455	.0518	.0492	.0460	.0503
	B_{Col}	.0466	.0738	.0514	.0465	.0741	.0503	.0472	.0913	.0523
	A×B	.0497	.0392	.0522	.0482	.0432	.0492	.0486	.0328	.0539
4,4,1,1 1,1,4,4	A_{Row}	.0511	.0493	.0507	.0512	.0488	.0498	.0517	.0530	.0551
	B_{Col}	.0499	.0564	.0504	.0510	.0563	.0493	.0494	.0709	.0577
	A×B	.0488	.0679	.0521	.0490	.0684	.0515	.0521	.0761	.0584
16,9,4,1 16,9,4,1	A_{Row}	.0495	.0528	.0500	.0523	.0486	.0505	.0505	.0518	.0483
	B_{Col}	.0459	.0765	.0545	.0463	.0689	.0486	.0499	.0632	.0518
	A×B	.0476	.0415	.0504	.0474	.0403	.0491	.0500	.0411	.0496
16, 9, 4, 1 1, 4, 9, 16	A_{Row}	.0490	.0499	.0508	.0484	.0491	.0522	.0496	.0559	.0576
	B_{Col}	.0492	.0579	.0493	.0534	.0569	.0549	.0496	.0819	.0597
	A×B	.0506	.0776	.0515	.0504	.0747	.0543	.0497	.1002	.0609
16,14,12,10 2, 4, 6, 8	A_{Row}	.0512	.0498	.0508	.0508	.0545	.0512	.0467	.0439	.0507
	B_{Col}	.0486	.0586	.0526	.0490	.0565	.0507	.0524	.0654	.0500
	A×B	.0484	.0525	.0528	.0497	.0455	.0517	.0501	.0406	.0552
16,14,12,10 8, 6, 4, 2	A_{Row}	.0479	.0533	.0502	.0517	.0520	.0522	.0521	.0489	.0486
	B_{Col}	.0502	.0567	.0538	.0481	.0553	.0578	.0503	.0620	.0544
	A×B	.0480	.0438	.0536	.0525	.0491	.0516	.0487	.0453	.0503
1,1,1,1 9,9,9,9	A	.0493	.0567	.0509	.0491	.0521	.0518	.0526	.0635	.0548
	B	.0504	.0612	.0546	.0463	.0543	.0514	.0501	.0782	.0608
	A×B	.0471	.0528	.0577	.0464	.0521	.0495	.0498	.0704	.0630

Note. Shading indicates that the value is greater than the criterion .0544; the test has an inflated Type I error rate.

[a]Skew: Skewed-leptokurtic non-normal distribution (skew = 1.75, kurtosis = 3.75).

[b]Platy: Platykurtic non-normal distribution (skew = 0, kurtosis = -1.0).

Table 6

Type I Error Rate for the James Second-Order Test in a Balanced Four by Four Fixed-Effect Factorial Design with Non-Normal Distribution Data

Sample Size	Variance	Skew	.75			1.0			1.25			1.50		
		Kurtosis	.50	1.0	3.75	.50	1.0	3.75	1.50	2.75	3.75	2.50	3.00	3.75
5, 5, 5, 5 5, 5, 5, 5 5, 5, 5, 5 5, 5, 5, 5	1, 1, 1, 1 1, 1, 1, 1 1, 1, 1, 1 1, 1, 1, 1	A_{Row}	.0427	.0445	.0368	.0490	.0462	.0346	.0450	.0413	.0375	.0470	.0429	.0438
		B_{Col}	.0448	.0435	.0362	.0470	.0442	.0376	.0413	.0414	.0387	.0476	.0440	.0425
		A×B	.0304	.0326	.0239	.0309	.0285	.0224	.0286	.0245	.0232	.0210	.0233	.0247
	1, 4, 9, 16 1, 4, 9, 16 1, 4, 9, 16 1, 4, 9, 16	A_{Row}	.0437	.0436	.0370	.0496	.0456	.0390	.0475	.0426	.0370	.0397	.0448	.0439
		B_{Col}	.0480	.0452	.0383	.0549	.0514	.0404	.0619	.0510	.0473	.0650	.0614	.0566
		A×B	.0389	.0342	.0262	.0315	.0332	.0252	.0271	.0275	.0220	.0208	.0236	0251
	1, 16, 9, 4 4, 1, 16, 9 9, 4, 1, 16 16, 9, 4, 1	A_{Row}	.0480	.0415	.0241	.0457	.0457	.0365	.0434	.0401	.0361	.0434	.0447	.0419
		B_{Col}	.0426	.0425	.0382	.0469	.0476	.0365	.0441	.0425	.0396	.0431	.0455	.0411
		A×B	.0548	.0462	.0264	.0756	.0609	.0320	.0758	.0534	.0412	.0810	.0731	.0581
25, 25, 25, 25 25, 25, 25, 25 25, 25, 25, 25 25, 25, 25, 25	1, 1, 1, 1 1, 1, 1, 1 1, 1, 1, 1 1, 1, 1, 1	A_{Row}	.0530	.0477	.0485	.0477	.0486	.0479	.0525	.0536	.0503	.0536	.0532	.0527
		B_{Col}	.0535	.0510	.0458	.0505	.0519	.0508	.0527	.0504	.0482	.0529	.0515	.0528
		A×B	.0445	.0445	.0391	.0448	.0471	.0402	.0453	.0411	.0407	.0380	.0416	.0424
	1, 4, 9, 16 1, 4, 9, 16 1, 4, 9, 16 1, 4, 9, 16	A_{Row}	.0478	.0483	.0455	.0520	.0502	.0477	.0560	.0478	.0517	.0490	.0530	.0537
		B_{Col}	.0513	.0523	.0486	.0505	.0507	.0530	.0546	.0520	.0562	.0583	.0517	.0573
		A×B	.0477	.0446	.0388	.0424	.0456	.0400	.0380	.0409	.0415	.0408	.0386	.0373
	1, 16, 9, 4 4, 1, 16, 9 9, 4, 1, 16 16, 9, 4, 1	A_{Row}	.0509	.0508	.0471	.0469	.0498	.0491	.0542	.0497	.0534	.0541	.0543	.0528
		B_{Col}	.0482	.0489	.0456	.0499	.0491	.0493	.0521	.0483	.0510	.0500	.0489	.0549
		A×B	.0531	.0497	.0425	.0545	.0549	.0441	.0599	.0532	.0484	.0611	.0626	.0597

Note. Shading indicates that the value is greater than the criterion .0544; the test has an inflated Type I error rate.

Table 7

Type I Error Rate for the James Second-Order Test in a Two by Four Fixed-Effect Factorial Design with Non-Normal Distribution Data

Sample Size	Variance	Skew	0.50		0.75		1.00		1.25			1.50		
		Kurtosis	0.50	3.75	0.50	3.75	0.50	3.75	2.50	3.50	3.75	2.50	3.50	3.75
15, 15, 15, 15 15, 15, 15, 15	1, 1, 1, 1 1, 1, 1, 1	A_{Row}	.0480	.0500	.0492	.0462	.0482	.0516	.0492	.0503	.0517	.0533	.0506	.0463
		B_{Col}	.0517	.0486	.0485	.0438	.0546	.0458	.0544	.0528	.0473	.0590	.0554	.0501
		A×B	.0466	.0465	.0478	.0452	.0480	.0381	.0486	.0439	.0480	.0443	.0431	.0394
	16, 9, 4, 1 1, 4, 9, 16	A_{Row}	.0499	.0496	.0510	.0501	.0510	.0468	.0507	.0483	.0460	.0468	.0509	.0498
		B_{Col}	.0512	.0415	.0539	.0457	.0562	.0452	.0533	.0497	.0517	.0578	.0529	.0563
		A×B	.0515	.0464	.0571	.0470	.0645	.0502	.0606	.0589	.0572	.0717	.0656	.0672
	1, 1, 1, 1 9, 9, 9, 9	A_{Row}	.0480	.0509	.0509	.0502	.0509	.0531	.0514	.0522	.0495	.0513	.0517	.0524
		B_{Col}	.0526	.0465	.0516	.0446	.0511	.0449	.0609	.0499	.0499	.0592	.0517	.0569
		A×B	.0531	.0455	.0563	.0447	.0528	.0459	.0562	.0476	.0465	.0534	.0515	.0540
18, 16, 14, 12 17, 16, 14, 13	1, 1, 1, 1 1, 1, 1, 1	A_{Row}	.0504	.0461	.0505	.0483	.0515	.0540	.0501	.0499	.0478	.0468	.0481	.0453
		B_{Col}	.0476	.0428	.0502	.0448	.0535	.0474	.0510	.0504	.0496	.0552	.0535	.0542
		A×B	.0478	.0446	.0493	.0449	.0432	.0433	.0433	.0441	.0460	.0418	.0445	.0428
	16, 9, 4, 1 1, 4, 9, 16	A_{Row}	.0510	.0468	.0510	.0507	.0459	.0485	.0501	.0472	.0491	.0499	.0473	.0501
		B_{Col}	.0505	.0457	.0545	.0462	.0578	.0484	.0531	.0500	.0498	.0592	.0520	.0534
		A×B	.0533	.0452	.0556	.0439	.0601	.0491	.0632	.0529	.0506	.0742	.0637	.0628
	1, 1, 1, 1 9, 9, 9, 9	A_{Row}	.0519	.0482	.0465	.0497	.0513	.0506	.0540	.0534	.0511	.0607	.0561	.0531
		B_{Col}	.0528	.0414	.0518	.0455	.0559	.0472	.0536	.0504	.0498	.0627	.0520	.0591
		A×B	.0516	.0453	.0502	.0458	.0563	.0468	.0509	.0514	.0493	.0553	.0516	.0540

Continued

(Table 7 continued)

Sample Size	Variance	Skew	0.50		0.75		1.0		1.25			1.50		
		Kurtosis	0.5	3.75	.50	3.75	.50	3.75	2.50	3.50	3.75	2.50	3.50	3.75
18, 16, 12, 14 12, 14, 18, 16	1, 1, 1, 1 1, 1, 1, 1	A_{Row}	.0482	.0488	.0500	.0483	.0482	.0482	.0491	.0480	.049[illegible]	.0464	.0476	.0475
		B_{Col}	.0476	.0471	.0484	.0459	.0547	.0445	.0511	.0501	.0493	.0545	.0531	.0540
		A×B	.0448	.0463	.0455	.0419	.0465	.0446	.0431	.0426	.0446	.0463	.0445	.0456
	16, 9, 4, 1 1, 4, 9, 16	A_{Row}	.0500	.0485	.0515	.0491	.0473	.0511	.0498	.0479	.0508	.0487	.0474	.0488
		B_{Col}	.0450	.0483	.0465	.0434	.0557	.0468	.0510	.0498	.0511	.0550	.0553	.0523
		A×B	.0439	.0500	.0497	.0475	.0596	.0502	.0513	.0517	.0502	.0634	.0617	.0571
	1, 1, 1, 1 9, 9, 9, 9	A_{Row}	.0468	.0524	.0465	.0472	.0530	.0527	.0467	.0485	.0521	.0546	.0553	.0532
		B_{Col}	.0441	.0508	.0497	.0428	.0584	.0473	.0523	.0512	.0450	.0588	.0521	.0536
		A×B	.0423	.0473	.0471	.0427	.0545	.0461	.0461	.0464	.0483	.0568	.0511	.0566
22, 18, 12, 8 24, 20, 10, 6	1, 1, 1, 1 1, 1, 1, 1	A_{Row}	.0474	.0459	.0495	.0465	.0487	.0497	.0414	.0473	.0495	.0435	.0488	.0482
		B_{Col}	.0471	.0418	.0511	.0454	.0564	.0428	.0547	.0497	.0525	.0598	.0584	.0553
		A×B	.0467	.0386	.0440	.0408	.0438	.0453	.0425	.0466	.0406	.0431	.0415	.0408
	16, 9, 4, 1 1, 4, 9, 16	A_{Row}	.0430	.0451	.0537	.0454	.0591	.0464	.0529	.0509	.0494	.0611	.0531	.0513
		B_{Col}	.0547	.0426	.0580	.0459	.0703	.0498	.0625	.0532	.0516	.0794	.0658	.0640
		A×B	.0522	.0434	.0576	.0483	.0777	.0531	.0677	.0567	.0634	.0958	.0780	.0777
	1, 1, 1, 1 9, 9, 9, 9	A_{Row}	.0493	.0488	.0556	.0535	.0605	.0498	.0582	.0542	.0512	.0638	.0588	.0619
		B_{Col}	.0493	.0423	.0577	.0438	.0655	.0464	.0618	.0534	.0533	.0735	.0690	.0666
		A×B	.0514	.0446	.0617	.0458	.0662	.0428	.0591	.0513	.0478	.0719	.0655	.0615
22, 12, 10, 6 8, 18, 20, 24	1, 1, 1, 1 1, 1, 1, 1	A_{Row}	.0480	.0505	.0477	.0449	.0419	.0466	.0467	.0495	.0483	.0464	.0475	.0516
		B_{Col}	.0433	.0490	.0504	.0444	.0557	.0435	.0517	.0499	.0482	.0524	.0530	.0533
		A×B	.0400	.0492	.0491	.0436	.0586	.0456	.0520	.0469	.0513	.0615	.0545	.0539
	16, 9, 4, 1 1, 4, 9, 16	A_{Row}	.0462	.0493	.0489	.0457	.0498	.0526	.0496	.0458	.0508	.0492	.0523	.0459
		B_{Col}	.0396	.0513	.0524	.0427	.0525	.0482	.0546	.0503	.0451	.0560	.0557	.0503
		A×B	.0414	.0495	.0486	.0457	.0510	.0518	.0518	.0503	.0489	.0559	.0538	.0530
	1, 1, 1, 1 9, 9, 9, 9	Row	.0460	.0547	.0481	.0520	.0526	.0492	.0572	.0497	.0533	.0598	.0532	.0542
		B_{Col}	.0407	.0477	.0521	.0449	.0616	.0447	.0584	.0515	.0536	.0684	.0575	.0552
		A×B	.0442	.0491	.0504	.0407	.0609	.0448	.0584	.0502	.0521	.0669	.0589	.0588

Note. Shading indicates that the value is greater than the criterion .0544; the test has an inflated Type I error rate.