DOCUMENT RESUME

ED 271 499 TM 860 408

AUTHOR	Olejnik, Stephen F.; Algina, James
TITLE	Tests of Variance Equality When Distributions Differ in Form, Scale and Location.
PUB DATE	Apr 86
NOTE	25p.; Paper presented at the Annual Meeting of the American Educational Research Association (70th, San Francisco, CA, April 16-20, 1986).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Error of Measurement; Evaluation Methods; *Measurement Techniques; Nonparametric Statistics; *Research Design; Research Problems; Sample Size; *Sampling; Simulation; Statistical Analysis; Statistical Studies; *Testing
IDENTIFIERS	Power (Statistics); *Sampling Error; *Type I Errors; Variance (Statistical)

ABSTRACT

Sampling distributions for ten tests for comparing population variances in a two group design were generated for several combinations of equal and unequal sample sizes, population means, and group variances when distributional f.rms differed. The ten procedures included: (1) O'Brien's (OB); (2) O'Brien's with adjusted degrees of freedom; (3) Brown-Forsythe (BF); (4) Welch-James test on O'Brien's r transform (WJOB); (5) unaligned Klotz; (6) median-aligned Klotz; (7) mean-aligned Klotz; (8) unaligned Siegel-Tukey; (9) median-aligned Siegel-Tukey; and (10) mean-aligned Siegel-Tukey. Type 1 error rates and power estimates were compared using simulated data. The present results in conjuction with previous research indicates that no one procedure for comparing variances is uniformly the best strategy. With equal-sized samples a strategy that will yield appropriate Type 1 errors and reasonably powerful tests is to use OB unless there is evidence of heavy tails. In that case BF can be used to increase power while retaining appropriate Type 1 error rates. With unequal sample sizes and when the evidence suggests a direct relationship between the sample sizes and population variances the strategy might be modified to use WJCB in place of OB. (PN)

Tests of Variance Equality When Distributions Differ in Form, Scale and Location

Stephen F. Olejnik
University of Georgia

James Algina
University of Florida

Abstract

Sampling distributions of 10 tests for comparing population variances in a two group design were generated for several combinations of equal and unequal (a) sample sizes, (b) population means, and (c) group variances when distributional forms differed. Type I error rates and power estimates were compared. No one procedure provided the best solution for all conditions studied. The O'Brien procedure, however, generally had appropriate Type I error rates and provided power estimates as high or higher than alternatives for most conditions studied. Modifying O'Brien's procedure by adopting a Welch-type ANOVA test increased statstical power when sample size and group variance was positively related.

Tests of Variance Equality When Distributions Differ in Form, Scale, and Location

Although many statistical tests for comparing population variances have been developed, very few of these procedures are appropriate when the population distributions are non-normal. Conover, Johnson, and Johnson (1981) compared 56 tests of variance equality and could only recommend three tests which they felt were insensitive to non-normal distributions and had adequate statistical power. Of the parametric procedures studied by Conover et al., only the Brown-Forsythe (1974) procedure was recommended. This procedure is an ANOVA using $D_{ij} = |X_{ij} - m_j|$ as the dependent variable. Here, X_{ij} is the ith observation in the jth group and m_j is the median in the jth group.

An alternative parametric procedure, not considered by Conover et al. (1981), was suggested by O'Brien (1978). This procedure is an ANOVA using the following transformation of X_{ij}:

$$r_{ij} = (n_j-1.5)n_j(X_{ij}-X_{.j})^2 - .5S_j^2(n_j-1)/[(n_j-1)(n_j-2)].$$

In this expression n_j is the number of observations in the jth group, $X_{.j}$ is the mean score for the jth group, and S_j^2 is the variance of the observed scores in the jth group. It can be shown that for the jth population the mean of r_{ij} is equal to the population variance of X_{ij}. Similarly, for the jth sample the mean of r_{ij} is equal to the sample variance. O'Brien's approach has been shown to be robust to non-normal distributions and can be substantially more powerful than the Brown-Forsythe approach when the population distribution is light-tailed (O'Brien, 1978; Olejnik & Algina, 1985a).

There are a variety of nonparametric procedures available (Duran, 1976). Two of the better known techniques were suggested by Siegel and Tukey (1960) and by Klotz (1962). In the Siegel-Tukey approach the data are pooled across groups and ranks are assigned as follows: 1 to the lowest score, 2 and 3 to the highest and second highest scores, respectively; 4 and 5 to the second and third lowest scores, respectively; and so forth. The ranks are then disaggregated and the mean ranks are compared among groups. In the two group case the Wilcoxon rank test can be used, whereas in the multiple group case the Kruskal-Wallis test is used (Puri, 1964). The Kruskal-Wallis test statistic is computed using

$$H = \frac{12}{N(N + 1)} \sum_j \frac{R_j^2}{n_j} - 3(N + 1)$$

where R_j is the sum of ranks in the jth group, n_j is the number of observations in the jth group and N is the total sample size in the study. The test statistic is asymptomatically distributed as chi-square with J - 1 degrees of freedom.

Klotz's (1962) test uses normal scores. After pooling the observed scores across groups, the data are ranked from lowest to highest; the ranks are replaced by their inverse normal scores

$$Z_{ij} = \phi^{-1} \frac{(i)}{N + 1} .$$

The test statistic is calculated using $T_{ij} = Z_{ij}^2$ in the following formula:

$$K = (N - 1) \frac{\sum_j n_j (T_{.j} - T_{..})^2}{\sum_i \sum_j (T_{ij} - T_{.j})^2}$$

(Puri, 1964). K is asymptomatically distributed as chi-square with J - 1 degrees of freedom. As expected from theoretical considerations, when applied to data sampled from populations with identical non-normal distributions, the Siegel-Tukey and Klotz tests have appropriate Type I error rates (Penfield & Koeffler, 1985; Olejnik & Algina, 1985a). When the sampled distributions differ only in variance, the Klotz test has power equal or greater than the power of the Brown-Forsythe or O'Brien procedure (Olejnik & Algina, 1985a).

The Siegel-Tukey and Klotz tests of scale are strongly affected by differences between population medians (Moses, 1963). When the sampled distributions have the same shape, the tests become increasingly insensitive to scale differences as differences in the location parameters increase. A suggested solution (e.g., Lehmann, 1975; Marascilo & McSweeney, 1977) to this problem is to align the data by computing the difference between each observation and its group mean, $(X_{ij} - X_{.j})$. In conducting the Siegel-Tukey or Klotz test, the deviation scores can then be used in place of the raw scores. Alternatively, the data can be aligned by computing the deviation of each score from its median. When applied to data sampled from two symmetric distributions that are identical except possibly in location, the mean-aligned Siegel-Tukey and the mean- and median-aligned Klotz tests tend to yield appropriate Type I error rates. With such symmetric distributions, the mean-aligned Klotz test has power comparable to the Brown-Forsythe and O'Brien tests (Olejnik & Algina, 1985a). Using the mean- or median-aligned data when the population distributions are asymmetric provides liberal Type I error rates for both the Klotz and Siegel-Tukey procedures.

From previous investigations it is clear that no one approach is uniformly superior for comparing variances. The best procedure for a given situation depends on the shape of the population distribution and the magnitude of the difference in the population location parameters. This conclusion is based on studies which have assumed that the distributional forms of the populations sampled were identical. The nonparametric and O'Brien's test also make this assumption. In the latter case, differences in distributional forms affects the independence between the mean square between and within group sources of variation in calculating the F-ratio (O'Brien, 1979). Since differences in distributional forms may occur in ex post facto studies as well as experiments where a treatment could affect the shape of the distribution, one of the purposes of the present study was to investigate the effect differences in distributional forms have on the Type I error rates and statistical power of several tests of variance equality.

The second purpose was to investigate two modifications suggested by O'Brien (1979) for improving the robustness and power of his test. Box and Andersen (1955) showed that when the population distributions have the same shape and only the normality assumption is violated, the ANOVA F statistic is approximately distributed as F with degrees of freedom δdf_1 and δdf_2. Here $\delta = (1 + g/N)$ and g is the kurtosis of the population distribution of the dependent variable. O'Brien suggested that if the distribution of r is known, the degrees of freedom for the test statistic can be adjusted accordingly. However, when the population distributional form is unknown, O'Brien recommended calculating δ under the assumption that the observed scored distribution is normal. Under that condition, the distribution of r is a chi-square with 1 degree of freedom; the kurtosis is 12. Thus, O'Brien suggested using (1 + 12/N) (J-1) and (1 + 12/N) (N-J) degrees of freedom for a one way analysis of variance. The effect of using the modified degrees of freedom when the population distribution is non-normal and unknown has not been studied.

O'Brien showed that when the sampled populations are identically distributed, his test will tend to be liberal if the sample sizes are unequal. To overcome this problem, O'Brien suggested using a Welch-James analysis in place of the ANOVA of the transformed variable r. It can be shown that when there are between population differences in raw score (X) kurtoses, there will also be between population differences in the variances of r. The Welch-James analysis is also a solution to potential problems that may arise in this situation. Finally, regardless of whether there are between population differences in kurtoses, when there are between population differences in the variances of X there will also be between population differences in the variances of the r. In this situation it can be predicted from research comparing ANOVA to the Welch-James procedure for testing hypothesis on means (Olejnik & Algina, 1985b), that applying the Welch-James procedure to r will result in a more powerful test than O'Brien's when there is a direct relationship between the sample sizes and the population variances of r. This relationship will occur when there is a direct relationship between the sample sizes and the population variances of X. Because of these considerations the present study investigated applying the Welch-James procedure to r. The Welch-James (Welch, 1951) statistic is calculated using

$$F^* = \frac{\sum_j w_j (r_{.j} - \sum_j w_j r_{.j} / \sum_j w_j)/J - 1}{1 + \frac{2(J-2)}{J^2 - 1}\left[\sum_j \frac{1}{n_j - 1} \quad \frac{(1 - w_j)^2}{\sum_j w_j}\right]}$$

where $w_j = n_j / S_j^2$. The statistic F* is approximately distributed as F with (J - 1) and

$$\left[\frac{3}{(J^2 - 1)} \quad \sum_j \frac{1}{n_j - 1} \quad \left(1 - \frac{w_j}{\sum w_j}\right)\right]^{-1}$$

degrees of freedom.

In summary, the present study investigated the impact of between population differences in shape, variance, and mean on the Type I error rates and power of 10 procedures. In addition, the impact of between sample differences in sample size were investigated. The 10 procedures included were: O'Brien's (OB); O'Brien's with adjusted degrees of freedom (AOB); Brown-Forsythe (BF); Welch-James test on O'Brien's r transform (WJOB); the unaligned, median-, and mean-aligned Klotz (K, KMD, and KM); and the unaligned, median- and mean-aligned Siegel-Tukey (ST, STMD, and STM).

Method

Type I error rates and power were estimated using simulated data. Sample size, population variances, population means, and the shapes of the population distributions were all manipulated. All comparisons were based on data generated for a two group design, although the parametric tests are appropriate for multiple group designs. The nonparametric approaches, however, are not appropriate for factorial designs.

Three sample-size combinations were included: (20,20); (23,17); and (17,23). To compare Type I error rates, data were generated from populations having equal variances. Statistical power estimates were made based on samples generated from two populations with variances in the following ratios: (1.5:1); (2.0:1); (2.5:1); (3.5:1); and (4.0:1). Population means were equal or differed by .5 pooled standard deviation units. Finally, eight distributional shapes were considered: normal; platykurtic; skewed, with three levels of skewness; and leptokurtotic, with three levels of kurtosis. Table 1 presents skewness and kurtosis figures for the distributions from which the data were generated.

Insert Table 1 about here

Data for the study were generated and analyzed using the SAS computing package. Scores on the dependent measure were created using the linear model $X_{ij} = \mu_{..} + \alpha_{.j} + \sigma\varepsilon_{ij}$. The grand mean, $\mu_{..}$, was set equal to 10. The effect size parameter for the jth group, $\alpha_{.j}$, was set to either 0 or .5 pooled standard deviation units. The random error component for the model ε_{ij} was generated using the RANNOR normal random number generating function in SAS. A standard normal random variable Y_{ij} was generated with the RANNOR function and ε_{ij} was set equal to it when studying the normal distribution. To study the non-normal distributions, Y_{ij} was transformed using the power function suggested by Fleisman (1978): $\varepsilon_{ij} = a + bY_{ij} + cY_{ij}^2 + dY_{ij}^3$. (The constants a, b, c, and d are chosen to transform the normally distributed variable to a standardized variable with known skewness and kurtosis.)

The coefficient σ_1 was chosen so that the variance of the first group increased from 1 to 4 in increments of .5 units. The coefficient σ_2 was set equal to 1 for all observations in group two. Each condition was replicated 1,000 times. Test statistics for the 10 procedures were computed and the frequencies of rejecting the null hypothesis of equal variances were recorded for the nominal .01, .05, and .10 significance levels.

The present study investigated the Type I error rates and power of each of the tests when the sampled population distributions differed in shape. For the eight distributions in Table 1, there are 56 possible combinations not including combinations involving negatively skewed distributions. Because it was not possible to consider all possible combinations, only a subset of the distribution combinations were selected for investigation. Table 2 summarizes the conditions investigated. In Table 2 the symbol -S indicates a negatively skewed distribution.

Insert Table 2 about here

Results

Only the results obtained when the nominal significance level was equal to .05 are reported here. Patterns of results similar to those reported here were obtained when the nominal significance level was .01 and .10.

To evaluate the adequacy of the procedures with regard to Type I error rates, estimated Type I error rates more than two standard errors above or below the nominal significance level were judged as unacceptable. Based on 1,000 replications the standard error for the nominal .05 significance level is .0069, so observations outside the interval (.036, .064) were considered either less than or greater than the nominal significance level.

Type I Error Rates

Estimated Type I error rates for the conditions in which population means were equal are reported in Tables 3 and 4. The former results are for equal-sized samples and the latter are for unequal-sized samples. With equal-sized samples the following seven tests had inappropriate Type I error rates for no more than two distribution combinations: OB, BF, WJOB, ST, K, STMD, KMD. The remaining tests exhibited a strong tendency to be liberal. Among the seven tests with appropriate Type I error rates, only ST and STMD tended to become markedly liberal with unequal sample sizes. The remaining tests (OB, BF, WJOB, K, and KMD) maintained their appropriate Type I error rates. The tests that were liberal with equal-sized samples were also liberal with unequal-sized samples.

Insert Tables 3 and 4 about here

With two notable exceptions, the same pattern of results tended to occur for the conditions in which the population means were unequal. First, ST did not become more liberal when the sample sizes were unequal. Second, K exhibited more of a conservative tendency for both equal- and unequal-sized samples.

Power

Power estimates are reported in Table 5 for the conditions in which the population means were equal. The power estimates are reported only for tests that had appropriate Type I error rates for both equal- and unequal-sized samples. However, power estimates for K are not reported because although it had good power in conditions with equal means, it had poor power in conditions with unequal means.

Insert Table 5 about here

For conditions involving two equal-sized samples the power comparisons depended on the distribution combination. OB, BF, WJOB, and KMD were approximately equivalent in power for the following distribution combinations: N/N, SS/MS, N/SS, N/MS, N/S, N/-S and N/SL. For the N/P distribution combination, BF had somewhat less power than the other procedures. The KMD test had less power than the other three procedures when applied to data from the S/-S distribution combination. When at least one of the distributions was leptokurtic, BF and KMD had approximately equivalent power and more power than either OB or WJOB.

The pattern of results was similar for conditions in which there were between population mean differences. Because these results would not lead to any radically different conclusions about power, they are not reported here.

When the sample sizes were unequal, the power estimates depended on the relationship between the population variances and sample sizes.

With a direct relationship, WJOB had power superior to or approximately equal to the power of the other three tests. This was true for every distribution combination. WJOB was superior to OB, BF, and KMD for the N/N, N/P, SS/MS, N/SS, N/MS, N/S, N/-S, and S/-S distribution combinations. Only for the S/-S combination were there any notable differences among the power estimates for OB, BF, and KMD; KMD had less power than the other two tests. WJOB, BF, and KMD were about equally powerful and more powerful than OB when applied to data from the N/SL, N/ML, N/L, or ML/SL distribution combinations.

When the population variances and sample sizes were inversely related, [the (17,23) sample-size combination], the OB test had power superior to or approximately equal to the power of the other three tests. With all but one distribution combination, WJOB had inferior power to the other three tests. The exception occurred for the S/-S combination. Here KMD and WJOB were approximately equal in power, and less powerful than BF. BF was, in turn, less powerful than OB. With the N/N combination, KMD and BF were about equally powerful and less powerful than OB. With the N/P, SS/MS, N/SS, N/MS, N/S, and N/-S combinations OB had more power than KMD which, in turn, was somewhat more powerful than BF. OB, BF, and KMD had approximately equal power when applied to the N/ML, N/L, and ML/SL condition combinations.

Power estimates were made for conditions in which the mean for the second population exceeded the mean for the first, and for conditions in which the size order of the means was reversed. For these conditions WJOB tended to be the procedure of choice with an indirect relationship between sample sizes and population variances, and OB tended to be the procedure of choice for indirect relationships. Because these results parallel the results for the conditions with equal means, the power estimates are not reported here.

Conclusions

Based on the results reported in the present paper the following conclusions were drawn.

1. The ST procedure frequently led to a liberal test when the populations had equal means and variances but dissimilar shapes and the sample sizes were unequal. When population means differed by a half standard deviation, ST had Type I error rates similar to the nominal significance level. The Type I error rate for K was not seriously affected by differences in distributional shape when population means were equal. The empirical Type I error rates tended to be less than the nominal significance level when population means were unequal. Previously, Conover et al. (1981) and Olejnik and Algina (1985a) found that when distributions had unequal means but were otherwise identical, both ST and K tended to be conservative. These results are consistent with the results reported in this paper for K, but are not consistent with the results reported for ST.

2. STM and KM tended to be liberal both for conditions with equal population means and for conditions with unequal population means. They were also liberal with both equal and unequal sample sizes. When the sample sizes were unequal the STMD was liberal for many distribution combinations. Only KMD had empirical Type I error rates similar to the

nominal level for most of the distribution combinations, for equal and unequal sample sizes, and w .h both equal and unequal population means. The results for STMD are consistent with previous research indicating that STMD becomes liberal with unequal sample sizes (Olejnik & Algina, 1985a). The results for STM and KM are not consistent with this previous research indicating that these tests have appropriate Type I error rates only with symmetric distributions (Conover et al., 1981; Olejnik & Algina, 1985a). In a sense, the results for KMD are not consistent with previous research indicating that KMD becomes liberal when all populations are skewed to the same degree.

3. Using AOB frequently led to a liberal test. This result was found for equal and unequal sample sizes and for equal and unequal population means. Even when both distributions were normal, a slightly inflated Type I error rate was observed.

4. Using OB, BF, and WJOB resulted in empirical Type I error rates similar to the nominal significance level. These results were consistent across equal and unequal sample sizes and across equal and unequal population means.

5. The relative power of the tests depended on the sample-size conditions being investigated. With equal sample sizes and either equal or unequal population means, the four tests considered in the power comparisons (OB, BF, WJOB, and KMD) tended to have similar power. The exceptions to this generalization were that KMD had lower power with the S/-S distribution combination, BF had lower power with the N/P combination, and BF and WJOB had lower power with the N/ML, N/L, and ML/SL combinations.

6. WJOB was at least as powerful as OB, BF, and KMD when sample sizes and population variances were positively related. It was more powerful except when at least one distribution was leptokurtic. In those conditions BF and KMD had approximately the same power as WJOB. These results were consistent for conditions involving equal population means and for conditions involving unequal population means. When sample sizes and population variances were negatively related, OB was at least as powerful as the other tests. OB was the most powerful test except with distribution combinations involving at least one leptokurtic distribution. Then, BF and KMD had power approximately equivalent to the power of OB. Again these results were consistent across conditions with equal population means and conditions with unequal population means.

7. The results of the present study in conjunction with previous research indicates that no one procedure for comparing variances is uniformly the best strategy. With equal-sized samples a strategy that will yield appropriate Type I errors and reasonably powerful tests is to use OB unless there is evidence of heavy tails. In that case BF can be used to increase power while retaining appropriate Type I error rates. With unequal sample sizes and when the evidence suggests a direct relationship between the sample sizes and population variances the strategy might be modified to use WJOB in place of OB. The major question about this modification is whether the relationship between the sample sizes and sample variances is a reliable guide to the relationship between the sample sizes and population variances. If the former relationship is an unreliable guide, then the modification may entail a fairly large loss of power.

References

Box, G.E.P., & Andersen, S.L. (1955). Permutation theory in the derivation of robust criteria and the study of departures from assumption. Journal of the Royal Statistical Society, Ser. B., 17, 1-26.

Brown, M.B., & Forsythe, A.B. (1974a). Robust tests for the equality of variances. Journal of the American Statistical Association, 69, 364-367.

Conover, W.J., Johnson, M.E., & Johnson, M.M. (1981). A comparative study of tests for homogeneity of variances, with applications to the outer continental self bidding data. Technometrics, 23, 351-361.

Duran, B.S. (1976). A survey of nonparametric tests for scale. Commun. Statist. Theor. Meth., A5, 1287-1312.

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

Klotz, J. (1962). Nonparametric tests for scale. Annals for Mathematical Statistics, 33, 495-512.

Kruskal, W.H., & Wallis, W.A. (1962). Use of ranks in one criterion variance analysis. Journal of the American Statistical Association, 47, 583-621.

Lehmann, E.L. (1975). Nonparametrics: Statistical methods based on ranks. San Francisco: Holden-Day.

Marascuilo, L.A., & McSweeney, M. (1977). Nonparametric and distribution free methods for the social sciences. Monteray, CA: Brooks/Cole.

Moses, L.E. (1963). Rank tests of disperson. Annals of Mathematical Statistics, 34, 973-983.

O'Brien, R.G. (1978). Robust techniques for testing heterogeneity of variance effects in factorial designs. Psychometrika, 43, 327-342.

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

Olejnik, S.F., & Algina, J. (1985a, April). Power analysis of selected parametric and nonparametric tests for heterogeneous variances in non-normal distributions. Paper presented at the meetings of the American Educational Research Association, Chicago, Illinois.

Olejnik, S.F., & Algina, J. (1985b). Type I error rate and power of the rank transform ANOVA when populations are non-normal and have unequal variances. Florida Journal of Educational Research, 27, 61-81.

Penfield, D.A., & Koffler, S. (1985, April). A power study of selected nonparametric k-sample tests. Paper presented at the meeting of the American Educational Research Association, Chicago, Illinois.

Puri, M.L. (1964). On some tests of homogeneity of variances. Annals of the Institute of Statistical Mathematics, 17, 323-330.

Siegel, S., & Tukey, J.W. (1960). A nonparametric sum of ranks procedure for relative spread in unpaired samples. American Statistical Association Journal, 55, 429-445.

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

Table 1

Skewness and Kurtosis of the Distributions

Distribution		Skewness	Kurtosis
Normal	(N)	0	0
Platykurtic	(P)	0	-1.00
Slightly Skewed	(SS)	.25	0
Moderately Skewed	(MS)	.50	0
Skewed	(S)	.75	0
Slightly Leptokurtic	(SL)	0	.50
Moderately Leptokurtic	(ML)	0	1.75
Leptokurtic	(L)	0	3.75

Table 2

Summary of Conditions Simulated

Distributions	$\sigma_1^2 > \sigma_2^2$ $\alpha_1 = \alpha_2$	$\sigma_1^2 > \sigma_2^2$ $\alpha_1 < \alpha_2$	$\sigma_1^2 > \sigma_2^2$ $\alpha_1 > \alpha_2$
N/N	*	*	*
N/P	*	*	*
N/SS	*	*	*
N/MS	*	*	*
N/S	*	*	*
N/SL	*	*	*
N/ML	*	*	*
N/L	*	*	*
SS/or S	*	*	*
ML/SL	*	*	*

Note: The following sample-size combinations were employed with each condition: 20/20, 23/17, 17/23.

Table 3

Estimated Type I Error Rates: Conditions with Equal Effect Sizes and Sample Sizes

Distributions	AOB	OB	BF	WJOB	ST	K	STM	KM	STMD	KMD
N/N	64[a]	51	56	47	51	55	56	71	45	50
N/P	64	52	42	52	57	52	63	65	48	46
SS/MS	53	46	39	40	47	44	53	64	35	38
N/SS	49	36	41	36	55	52	60	60	46	44
N/MS	58	44	41	41	54	49	54	59	41	47
N/S	64	50	39	47	56	61	65	69	50	41
N/-S	76	57	52	56	54	57	79	74	54	54
S/-S	72	53	43	47	57	43	62	48	48	27
N/SL	54	40	48	35	39	41	51	54	45	40
N/ML	60	44	51	39	65	48	74	71	56	52
N/L	68	61	66	56	82	70	86	79	72	58
ML/SL	59	47	46	44	50	52	68	68	54	51

Note: Underlined figures are liberal estimated Type I error rates.

[a] Number are rounded to the thousandth place.

Table 4

Estimated Type I Error Rates: Conditions with Equal Effect Sizes and Unequal Sample Sizes

Distribution	Sample Size	AOB	OB	BF	WJOB	ST	K	STM	KM	STMD	KMD
N/N	23/17	58[a]	41	41	43	53	50	56	71	45	50
	17/23	55	43	41	42	50	58	55	56	76	45
N/P	23/17	66	51	34	46	72	61	69	56	114	48
	17/23	67	51	50	53	73	59	84	63	103	46
SS/MS	23/17	53	52	37	40	47	44	53	64	35	38
	17/23	68	39	33	59	57	56	56	68	92	43
N/SS	23/17	69	60	43	48	63	59	66	73	94	51
	17/23	48	39	43	43	60	43	59	58	86	46
N/MS	23/17	51	43	33	43	43	42	43	49	54	31
	17/23	63	52	51	52	57	56	58	72	109	63
N/S	23/17	64	44	36	46	65	50	69	72	96	43
	17/23	60	44	39	42	44	57	79	74	54	54
N/-S	23/17	54	41	39	51	50	43	61	52	77	38
	17/23	63	45	35	49	50	60	68	69	80	45
S/-S	23/17	73	57	40	65	69	70	75	53	84	27
	17/23	77	65	50	57	69	67	88	63	102	37
N/SL	23/17	62	50	44	50	54	51	63	56	91	51
	17/23	69	55	48	53	65	55	70	66	109	57
N/ML	23/17	64	46	53	58	60	60	77	72	92	48
	17/23	54	43	38	44	69	52	70	59	101	49
N/L	23/17	61	51	50	64	78	53	82	64	115	53
	17/23	75	62	63	38	85	81	85	84	141	69
ML/SL	23/17	62	45	43	42	62	68	65	70	90	53
	17/23	48	30	42	52	49	35	61	60	83	38

Note: Underlined figures are liberal estimated Type I error rates.

[a]Numbers are rounded to the thousandth place.

Table 5

Estimated Power: Conditions with Equal Differences

Distribution	Variance Ratio	Sample Size 20/20 OB	BF	WJOB	KND	23/17 OB	BF	WJOB	KMD	17/23 OB	BF	WJOB	KMD
N/N	1.5:1	12[a]	12	11	11	10	10	14	10	12	9	7	10
	2.0:1	24	23	23	23	21	21	29	21	26	22	16	23
	2.5:1	36	37	34	31	33	38	44	34	45	38	32	38
	3.0:1	51	50	50	52	43	46	55	44	57	51	38	52
	3.5:1	62	62	60	61	55	60	69	58	68	62	51	63
	4.0:1	70	70	67	71	61	68	75	68	78	72	60	71
N/P	1.5:1	10	8	5	8	7	6	11	7	11	5	5	9
	2.0:1	25	20	22	23	17	18	27	20	28	17	16	23
	2.5:1	40	31	37	38	29	26	44	31	44	29	28	37
	3.0:1	53	45	50	52	42	39	58	43	62	44	40	53
	3.5:1	65	55	61	60	54	53	69	58	72	58	51	65
	4.0:1	75	66	71	70	67	66	80	70	79	65	59	70
SS/MS	1.5:1	11	11	11	12	11	10	14	11	14	9	8	11
	2.0:1	28	23	23	24	22	22	30	24	28	22	18	27
	2.5:1	39	40	37	40	33	35	45	37	43	34	27	40
	3.0:1	51	50	50	53	43	47	56	51	59	47	40	54
	3.5:1	64	60	61	63	57	59	70	62	69	62	51	65
	4.0:1	70	68	67	71	62	66	74	68	77	70	58	72
N/SS	1.5:1	12	12	12	12	8	9	11	9	12	9	7	11
	2.0:1	24	22	23	23	21	22	28	21	28	21	18	24
	2.5:1	40	38	38	39	31	33	42	33	42	34	27	36
	3.0:1	52	50	50	49	42	47	57	47	57	51	40	52
	3.5:1	60	59	57	59	54	56	66	56	67	59	49	60
	4.0:1	70	71	68	71	63	68	76	67	74	70	57	69
N/MS	1.5:1	11	9	10	10	10	11	14	11	14	10	8	11
	2.0:1	25	23	23	24	19	20	29	20	27	19	16	22
	2.5:1	38	36	36	38	34	35	44	34	47	36	28	41
	3.0:1	53	51	51	51	46	50	61	50	56	49	40	52
	3.5:1	62	61	59	62	55	60	69	60	68	59	49	62
	4.0:1	71	72	68	72	58	65	72	65	73	67	55	70
N/S	1.5:1	11	10	11	13	10	10	15	13	12	8	7	13
	2.0:1	23	20	21	23	20	22	28	23	28	21	18	27
	2.5:1	40	36	38	40	34	34	44	36	44	36	31	42
	3.0:1	52	49	50	53	43	46	56	49	58	47	39	54
	3.5:1	64	63	61	68	53	55	66	58	68	59	48	64
	4.0:1	70	65	67	70	62	64	76	67	75	69	56	73

(Table continues)

Table 5 (continued)

Distribution	Variance Ratio	20/20 OB	20/20 BF	20/20 WJOB	20/20 KND	23/17 OB	23/17 BF	23/17 WJOB	23/17 KMD	17/23 OB	17/23 BF	17/23 WJOB	17/23 KMD
N/-S	1.5:1	12	9	10	12	9	9	13	10	13	9	7	12
	2.0:1	28	24	26	26	19	19	27	22	28	21	18	27
	2.5:1	40	36	37	39	30	34	43	35	42	34	27	39
	3.0:1	52	49	49	53	44	46	56	50	57	49	41	55
	3.5:1	62	59	58	63	57	58	69	58	67	58	49	65
	4.0:1	70	69	68	72	62	65	75	68	76	69	59	72
S/-S	1.5:1	12	10	11	7	12	11	16	7	14	9	8	7
	2.0:1	25	21	24	14	23	23	30	14	28	21	19	14
	2.5:1	42	37	39	25	34	33	44	21	43	34	29	26
	3.0:1	53	49	50	35	42	46	55	30	57	47	42	37
	3.5:1	62	59	59	44	58	60	70	45	70	60	55	47
	4.0:1	69	67	66	51	63	68	75	49	75	69	59	56
N/SL	1.5:1	12	13	11	12	10	10	13	10	15	12	10	14
	2.0:1	24	24	23	24	22	24	30	23	27	23	17	25
	2.5:1	40	39	37	39	32	37	43	36	45	40	30	40
	3.0:1	51	54	48	51	44	49	55	47	57	51	40	53
	3.5:1	60	63	57	61	55	60	66	60	69	64	50	63
	4.0:1	67	69	64	67	63	73	76	71	73	70	54	69
N/ML	1.5:1	14	15	14	13	12	14	16	13	17	15	12	15
	2.0:1	25	26	23	25	21	25	29	23	29	28	18	29
	2.5:1	40	43	39	43	34	40	46	38	39	39	28	41
	3.0:1	53	58	50	54	46	54	56	59	55	54	41	53
	3.5:1	63	67	61	65	52	61	63	56	65	64	48	62
	4.0:1	67	71	64	70	61	72	75	71	73	73	57	72
N/L	1.5:1	18	19	17	19	16	19	21	18	17	17	11	18
	2.0:1	33	39	31	35	25	33	33	32	33	34	23	33
	2.5:1	42	49	40	46	38	47	50	45	47	48	35	47
	3.0:1	54	63	52	59	48	60	60	54	56	59	41	56
	3.5:1	60	68	58	65	56	68	66	63	66	69	50	66
	4.0:1	66	75	64	72	61	74	72	71	73	77	61	75
ML/SL	1.5:1	6	6	6	7	4	6	8	8	7	5	3	6
	2.0:1	14	16	13	17	10	15	18	15	16	14	7	15
	2.5:1	22	25	20	27	14	22	24	21	28	27	15	28
	3.0:1	32	37	30	37	26	36	39	37	41	37	22	38
	3.5:1	41	47	38	49	33	47	49	47	50	48	30	49
	4.0:1	48	55	44	55	39	51	55	46	56	55	35	55

[a]Numbers are rounded to the hundredth place.