

11-1-2004

A Note On Extending Scheffé's Modified Multiple-Comparison Procedure to Other Analysis Situations

Xinyue Zhou
Fudan University, China

Joel R. Levin
University of Arizona, jrlevin@u.arizona.edu

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

Recommended Citation

Zhou, Xinyue and Levin, Joel R. (2004) "A Note On Extending Scheffé's Modified Multiple-Comparison Procedure to Other Analysis Situations," *Journal of Modern Applied Statistical Methods*: Vol. 3 : Iss. 2 , Article 15.
DOI: 10.22237/jmasm/1099268100

A Note On Extending Scheffé's Modified Multiple-Comparison Procedure to Other Analysis Situations

Xinyue Zhou

Department of Psychology
Fudan University, China

Joel R. Levin

Department of Educational Psychology
University of Arizona

This article extends Scheffé's modified (sequential) multiple-comparison procedure in one-way analysis-of-variance to other analysis situations, including interaction comparisons in factorial ANOVA designs, tests of partial regression coefficients in multiple-regression analysis, and comparisons of means in one-factor multivariate analyses of variance. Researchers who are concerned with maintaining familywise Type I error rates while increasing statistical power relative to the original (simultaneous) Scheffé-based procedures are encouraged to consider these improved multiple-comparison methods.

Key words: controlled multiple-test procedures, modified Scheffé method, Type I error and power

Introduction

A two-step modification of the original Scheffé (1953) multiple-comparison procedure was proposed by Scheffé (1970) and recently brought to researchers' attention by Klockars and Hancock (2000). Specifically, the statistical power of the original Scheffé procedure can be improved by conducting an initial omnibus F -test with a Type I error probability of α before proceeding to investigate any contrasts of interest. If the omnibus test is not rejected, the process stops. On the other hand, if the omnibus test is rejected the degrees of freedom associated with both Scheffé's original multiplier (v_1 , or $K-1$ in the one-way analysis of variance [ANOVA]) and critical F -value may be decreased by one. That is, v_1-1 (or $K-2$) may be employed to test all contrasts (rather than v_1 used in the original Scheffé procedure).

In a recent study, Meyers and Beretvas (2003) compared the familywise Type I error rates and power of the original and modified Scheffé procedures. The modified Scheffé procedure maintained its familywise Type I error at the nominal but less conservative level and, as a direct result, demonstrated greater power. However, Meyers and Beretvas' investigation was restricted to the one-way ANOVA situation.

As with Scheffé's (1953) multiple-comparison procedure, the Roy-Bose (1953) procedure is congruent with an omnibus test in a multivariate analysis of variance (MANOVA) context. Because a similar correspondence exists between Roy's θ criterion and the Roy-Bose procedure, it should be possible to improve the Roy-Bose procedure by adding an initial omnibus MANOVA test of Roy's θ . That is, if the omnibus test is not rejected, no subsequent multiple comparisons are conducted. However, if the omnibus test is rejected, all subsequent contrasts may be tested against a modified (reduced) Roy-Bose critical value.

The advantage of the modified Scheffé and Roy-Bose procedures over the original procedures is evident: similar control over the familywise Type I error rates, similar versatility, and similar computational ease, but greater statistical power. The major disadvantage of the modified procedures is that they do not permit the construction of probability-based confidence

Xinyue Zhou is a faculty member in the Department of Sociology at Fudan University, Shanghai, China, with specializations in cognitive processes and applied statistical methods. Joel R. Levin (jrlevin@u.arizona.edu) is Professor of Educational Psychology at the University of Arizona. His areas of expertise are cognitive strategies, applied statistical methods and pedagogy, and research methodology.

intervals. Thus, if such intervals are of interest or importance to a researcher, then these techniques are not recommended. Even though the modified Roy-Bose procedure is based on the same sequential hypothesis-testing logic (Seaman, Levin, & Serlin, 1991; Shaffer, 1986) as the modified Scheffé procedure, it has not yet been subjected to empirical test. The present Monte Carlo simulation study examines the familywise Type I error rates associated with both modified Scheffé and Roy-Bose procedures, along with two other commonly used analysis approaches.

Methodology

The multiple-comparison procedures examined here are MS (modified Scheffé), MRB (modified Roy-Bose), LSD (Fisher's least significant difference procedure – see Kirk, 1995), and U (Uncontrolled, or multiple t tests each conducted at α). The first step in LSD is to perform an omnibus α -level F test involving all means. Given a rejection of the omnibus test, pairwise differences are then tested using a per-contrast α . The U approach tests each comparison at a separate α without attending to familywise Type I error rate protection.

The study includes three common multiple-comparison situations: (1) interactions in two-factor ANOVA; (2) tests of partial regression coefficients in multiple-regression analysis; and (3) mean comparisons in one-factor MANOVA.

The SAS/IML program was used to simulate various experimental conditions for all specified situations, with the selection of samples from normally distributed populations accomplished using PROC RANNOR. The selection of samples from multivariate normal distributions was simulated using the pseudorandom number generator provided by PROC VNORMAL.

Ten thousand replications were conducted for each design specification. Each test was conducted using the algorithm prescribed by the corresponding multiple-comparison procedure (MS, MRB, LSD, U) based on a familywise Type I error probability of .05. The number of replications producing at least one Type I error for a comparison set was

tallied to yield an estimate of the traditionally defined familywise Type I error rate (i.e., the probability of making at least one Type I error in the set of comparisons). Decision rules proposed by Serlin (2000) were applied. According to his .25 α acceptability rule, with $\alpha = .05$ any familywise errors of 625 or fewer in 10,000 runs (.0625) are considered reasonable.

Interaction Comparisons in Two-Factor Analyses of Variance

Interactions in both 2 x 4 and 3 x 3 factorial designs were examined with $n = 20$ and $n = 100$ participants per cell. The sample sizes were selected so that the omnibus test would be rejected virtually all the time in the large-sample case and not all the time in the small-sample case. Both the complete null situation (no interactions associated with any contrasts) and various partial null situations (interactions associated with one or more contrasts) were examined. The cell means consisted of 1s and 0s designed to reflect various complete null and partial null patterns. The population variance for each variable was set to 1.00 for each simulation. To keep the analyses manageable, only tetrad (four-cell difference-in-difference) interaction comparisons were considered (Marascuilo & Levin, 1970). Accordingly, in the 2 x 4 layout, there are six tetrad contrasts; and in the 3 x 3 layout, there are nine.

Tests of Partial Regression Coefficients in Multiple-Regression Analysis

In multiple regression, various patterns were examined with varying parameters: P (number of predictor variables) = 2, 3, 4; and N (total number of participants) = 20, 100. In this study, all parameters, including the population variance of each predictor and the covariance between predictors, varied so that a single nonzero population partial regression coefficient (beta weight) was equal to 0, 0.2, 0.5, 0.8, 1, or 2 and the rest of the coefficients were equal to 0. Both the complete null situation (no independent variables have any unique contributions to the dependent variable) and various patterns of a partial null situation (the dependent variable shares some variance with only one independent variable) were included in the analysis.

Multivariate Analysis-of-Variance Comparisons

In MANOVA, various patterns of mean differences were studied by varying several parameters: K (number of groups) = 2, 3, 4, 5; P (number of outcome variables) = 2, 3, 4; n (number of participants per group) = 20, 100; and ρ (the common within-group correlation between all outcome variables) = 0.2, 0.5, 0.8. Both the complete null situation (no mean group differences on any variables) and various partial null situations (group mean differences on one or more variables) were included. The mean vectors consisted of 1s, 0s, and -1s to represent different complete null and partial null patterns. The within-group population variance for each variable was set at 1.00 for each simulation.

Special case (one-variable-at-a-time comparisons).

For these MANOVA simulations, $K = 3$ and $P = 2, 3, 4$ situations were investigated. Only one-variable-at-a-time comparisons were included in these analyses. It was assumed that the original Scheffé procedure could be employed to examine all one-variable-at-a-time comparisons by splitting the familywise α by the number of dependent variables (P) using the Bonferroni inequality (e.g., Kirk, 1995).

If the omnibus test is rejected, that means at least one of the variables is statistically different across groups. Then it is possible to modify the original Scheffé procedure by dividing the familywise α by $P-1$ instead of P (see Table 1). This modification was investigated in the simulation to see how it preserves the familywise Type I error rate.

General case: $K = 2, P > 1$ (both one-variable-at-a-time and multiple-variable comparisons).

When $K = 2$, the approach used for modifying (improving) the Roy-Bose procedure involved reducing by one the hypothesis degrees of freedom associated with the critical value of Roy's θ following a rejection of the initial omnibus test. The specifications included in this part of the MANOVA simulations were $K = 2, P = 2, 3, 4, 5$. This situation is the two-group MANOVA equivalent of multiple-regression analysis.

General case: $K > 2, P > 1$ (both one-variable-at-a-time and multiple-variable comparisons).

With $K > 2$ and multiple-variable comparisons included, three situations were investigated: $K-1 > P$ ($K = 5, P = 3$); $K-1 = P$ ($K = 4, P = 3$); and $K-1 < P$ ($K = 4, P = 4$). Two MRB approaches (one reducing K and the other reducing P) were considered. Comparisons based on both one variable at a time and multiple variables were included. To keep the analyses manageable, only four-mean "groups-by-variables interaction" comparisons (specifically, 2 groups by 2 variables) were included as multiple-variable comparisons.

Results

Factorial ANOVA Interaction Comparisons

The MS method maintained the nominal familywise Type I error rate for both 2×4 and 3×3 designs in both the complete and partial null situations. The maximum familywise Type I error rate of the MS method was .048 and the average familywise Type I error rate was .033 (see Table 2). All replications yielded empirical familywise Type I error rates below the $\alpha = 0.05$ criterion.

The LSD procedure preserved the familywise Type I error rate under the complete null situation, with a maximum error rate of .052 and an average error rate of .051. However, in the partial null situation, familywise Type I error rates were seriously inflated with the LSD procedure (average = .146, maximum = .182).

The U approach completely failed to preserve the familywise Type I error rate as long as there was more than one true null comparison. The proportion of times that at least one Type I error was made was as high as .280 and averaged .197.

Tests of Partial Regression Coefficients in Multiple-Regression Analysis

Similar patterns were observed in the multiple-regression analyses (see Table 3). With an increase in the number of predictors, the Type I error rates increased accordingly for the LSD and U methods. However, the reverse pattern was true for the MS method.

Table 1. Critical Values for Original Scheffé-Based Methods and Modifications

Analysis situations	Original Scheffé-based methods	Modified Scheffé-based methods
Interaction contrasts in ANOVA	$S = \sqrt{(I-1)(J-1)F_{(I-1)(J-1),IJ(n-1)}^{(1-\alpha)}}$	$MS = \sqrt{[(I-1)(J-1)-1]F_{(I-1)(J-1)-1,IJ(n-1)}^{(1-\alpha)}}$
Tests of partial regression coefficients in multiple regression	$S = \sqrt{PF_{P,N-P-1}^{(1-\alpha)}}$	$MS = \sqrt{(P-1)F_{P-1,N-P-1}^{(1-\alpha)}}$
MANOVA special case: one-variable-at-a-time comparisons only	$S = \sqrt{(K-1)F_{K-1,N-K}^{[1-\frac{\alpha}{P}]}}$	$MS = \sqrt{(K-1)F_{K-1,N-K}^{[1-\frac{\alpha}{P-1}]}}$
MANOVA general case ($K=2$): one-variable-at-a-time and multiple-variable comparisons	$RB = \frac{P(N-K)}{N-P-1} F_{P,N-P-1}^{(1-\alpha)}$	$MRB = \frac{(P-1)(N-K)}{N-P-1} F_{P-1,N-P-1}^{(1-\alpha)}$
MANOVA general case ($K > 2$): one-variable-at-a-time and multiple-variable comparisons	$RB = \sqrt{(\frac{\theta_{(s,m,n)}}{1-\theta_{(s,m,n)}})(N-K)}$ $s = \min(K-1, P);$ $m = \frac{abs[P-(K-1)]-1}{2}$ $n = \frac{(N-K-1)-P}{2}$	$MRB = \sqrt{(\frac{\theta_{(s,m,n)}}{1-\theta_{(s,m,n)}})(N-K)}$ based on reduced K or P

Table 2. Empirical Familywise Type I Error Rates of Three Methods at Nominal Alpha = .05 for Testing ANOVA Interaction Contrasts ($n = 20/n = 100$).

		LSD	U	MS
2 x 4 design	Complete null	.050/.052	.199/.203	.039/.048
	Partial null	.107/.128	.119/.128	.023/.038
3 x 3 design	Complete null	.051/.049	.275/.280	.034/.033
	Partial null	.167/.182	.187/.182	.024/.025

Table 3. Empirical Familywise Type I Error Rates of Three Methods at Nominal Alpha = .05 for Testing Partial Regression Coefficients in Multiple Regression ($N = 20/N = 100$)

$P = 2, B_2 = 0:$			
	LSD*	MS	U
$B_1 = 0.2$	.029/.040	.029/.040	.045/.047
$B_1 = 0.5$	.042/.053	.042/.053	.045/.053
$B_1 = 0.8$	.052/.053	.052/.053	.052/.054
$B_1 = 1$	.047/.050	.047/.050	.051/.050
$B_1 = 2$	.050/.051	.050/.051	.054/.052
$B_1 = 0$	.048/.046	.048/.046	.096/.097

$P = 3, B_2 = 0, B_3 = 0:$			
	LSD	MS	U
$B_1 = 0.2$	.046/.072	.027/.026	.094/.095
$B_1 = 0.5$	.076/.100	.028/.029	.091/.100
$B_1 = 0.8$	.093/.099	.030/.030	.093/.099
$B_1 = 1$	.081/.099	.030/.028	.096/.099
$B_1 = 2$	.076/.096	.029/.029	.089/.096
$B_1 = 0$	.046/.047	.035/.032	.133/.140

$P = 4, B_2 = 0, B_3 = 0, B_4 = 0:$			
	LSD	MS	U
$B_1 = 0.2$	.051/.092	.015/.015	.132/.135
$B_1 = 0.5$	.102/.135	.018/.015	.131/.135
$B_1 = 0.8$	.131/.143	.019/.015	.131/.143
$B_1 = 1$	.104/.145	.018/.017	.132/.145
$B_1 = 2$	.100/.134	.019/.015	.131/.134
$B_1 = 0$	.042/.042	.016/.014	.135/.144

* LSD in the $P = 2$ situation is the same procedure as MS.Table 4. Special Case: Empirical Familywise Type I Error Rates of Three Methods at Alpha = .05 for Testing One-Variable-At-a-Time Comparisons in MANOVA ($n = 20/n = 100$)

MS

	Complete null	Partial null
$K=3, P = 2, \rho = 0.2$	.038/.042	.054/.053
$K=3, P = 2, \rho = 0.5$	.037/.039	.055/.052
$K=3, P = 2, \rho = 0.8$	.030/.029	.053/.050
$K=3, P = 3, \rho = 0.2$	.033/.033	.049/.045
$K=3, P = 3, \rho = 0.5$	.027/.030	.050/.043
$K=3, P = 3, \rho = 0.8$	.022/.020	.042/.035
$K=3, P = 4, \rho = 0.2$	.025/.029	.040/.041
$K=3, P = 4, \rho = 0.5$	.024/.023	.045/.040
$K=3, P = 4, \rho = 0.8$	.016/.014	.036/.030

Table 4 Continued.

LSD

	Complete null	Partial null
$K=3, P=2, \rho=0.2$	.044/.051	.151/.165
$K=3, P=2, \rho=0.5$	.043/.047	.157/.158
$K=3, P=2, \rho=0.8$	.036/.038	.145/.148
$K=3, P=3, \rho=0.2$	.049/.048	.235/.260
$K=3, P=3, \rho=0.5$	.040/.046	.233/.235
$K=3, P=3, \rho=0.8$	.036/.035	.204/.196
$K=3, P=4, \rho=0.2$	.048/.052	.299/.351
$K=3, P=4, \rho=0.5$	.045/.046	.295/.316
$K=3, P=4, \rho=0.8$	.033/.033	.248/.244

U

	Complete null	Partial null
$K=3, P=2, \rho=0.2$	.227/.232	.160/.165
$K=3, P=2, \rho=0.5$	.211/.217	.161/.158
$K=3, P=2, \rho=0.8$	.180/.184	.145/.148
$K=3, P=3, \rho=0.2$	.313/.325	.263/.260
$K=3, P=3, \rho=0.5$	.290/.284	.243/.235
$K=3, P=3, \rho=0.8$	.230/.222	.205/.196
$K=3, P=4, \rho=0.2$	.400/.398	.354/.351
$K=3, P=4, \rho=0.5$	.356/.337	.309/.316
$K=3, P=4, \rho=0.8$	.259/.263	.248/.244

Table 5. Empirical Familywise Type I Error Rates of Three Methods at Alpha = .05 for Both One-Variable-At-a-Time and Multiple-Variable Comparisons in $K=2$ MANOVA ($n=20/n=100$)

MRB

	Complete null	Partial null
$P=2, \rho=0.2$	.040/.043	.048/.052
$P=2, \rho=0.5$	.044/.046	.046/.051
$P=2, \rho=0.8$	.040/.043	.043/.045
$P=3, \rho=0.2$	.039/.036	.042/.043
$P=3, \rho=0.5$	.035/.032	.040/.042
$P=3, \rho=0.8$	.031/.033	.040/.042
$P=4, \rho=0.2$	.026/.033	.023/.023
$P=4, \rho=0.5$	.026/.035	.022/.026
$P=4, \rho=0.8$	.030/.032	.022/.020
$P=5, \rho=0.2$	.019/.023	.014/.017
$P=5, \rho=0.5$	.016/.025	.012/.020
$P=5, \rho=0.8$	.018/.021	.013/.015

Table 5 Continued.

LSD

	Complete null	Partial null
$P = 2, \rho = 0.2$	.045/.054	.049/.056
$P = 2, \rho = 0.5$	.047/.051	.050/.053
$P = 2, \rho = 0.8$	.041/.049	.051/.052
$P = 3, \rho = 0.2$	.044/.045	.101/.098
$P = 3, \rho = 0.5$	.047/.047	.116/.101
$P = 3, \rho = 0.8$	.050/.049	.121/.093
$P = 4, \rho = 0.2$	.048/.051	.177/.197
$P = 4, \rho = 0.5$	.044/.052	.198/.204
$P = 4, \rho = 0.8$	.054/.048	.198/.188
$P = 5, \rho = 0.2$	.052/.053	.231/.285
$P = 5, \rho = 0.5$	.048/.050	.288/.289
$P = 5, \rho = 0.8$	.050/.049	.283/.274

U

	Complete null	Partial null
$P = 2, \rho = 0.2$	.108/.106	.060/.056
$P = 2, \rho = 0.5$	.095/.097	.051/.053
$P = 2, \rho = 0.8$	.084/.090	.057/.052
$P = 3, \rho = 0.2$	.185/.170	.101/.098
$P = 3, \rho = 0.5$	.164/.171	.099/.101
$P = 3, \rho = 0.8$	.169/.171	.088/.093
$P = 4, \rho = 0.2$	.275/.271	.202/.197
$P = 4, \rho = 0.5$	.293/.288	.207/.204
$P = 4, \rho = 0.8$	.283/.277	.199/.188
$P = 5, \rho = 0.2$	.362/.367	.279/.285
$P = 5, \rho = 0.5$	.375/.366	.288/.289
$P = 5, \rho = 0.8$	.370/.361	.283/.274

Note. Only the worst-case scenario (specifications with most null contrasts) is included in the table under the partial null situation.

Table 6. Empirical Familywise Type I Error Rates of Three Methods at Alpha = .05 for Multiple-Variable Comparisons in MANOVA ($n = 20/n = 100$).MRB (reducing P)

	Complete null	Partial null
$K = 4, P = 3, \rho = 0.2$ ($K-1 = P$)	.028/.025	.034/.028
$K = 4, P = 3, \rho = 0.5$	.030/.027	.034/.024
$K = 4, P = 3, \rho = 0.8$	.027/.028	.033/.023
$K = 4, P = 4, \rho = 0.2$ ($K-1 < P$)	.019/.014	.020/.013
$K = 4, P = 4, \rho = 0.5$	.019/.013	.019/.013
$K = 4, P = 4, \rho = 0.8$	.017/.016	.018/.012
$K = 5, P = 3, \rho = 0.2$ ($K-1 > P$)	.027/.023	.028/.016
$K = 5, P = 3, \rho = 0.5$	.021/.023	.028/.018
$K = 5, P = 3, \rho = 0.8$	.022/.018	.029/.016

Table 6 Continued.

MRB (reducing K)

	Complete null	Partial null
$K = 4, P = 3, \rho = 0.2$ ($K-1 = P$)	.028/.025	.034/.028
$K = 4, P = 3, \rho = 0.5$	.030/.027	.034/.024
$K = 4, P = 3, \rho = 0.8$	.027/.028	.033/.023
$K = 4, P = 4, \rho = 0.2$ ($K-1 < P$)	.020/.016	.024/.016
$K = 4, P = 4, \rho = 0.5$	.023/.015	.024/.015
$K = 4, P = 4, \rho = 0.8$	.018/.019	.022/.014
$K = 5, P = 3, \rho = 0.2$ ($K-1 > P$)	.021/.019	.024/.012
$K = 5, P = 3, \rho = 0.5$	.017/.019	.022/.015
$K = 5, P = 3, \rho = 0.8$	.019/.017	.023/.013

LSD

	Complete null	Partial null
$K = 4, P = 3, \rho = 0.2$ ($K-1 = P$)	.049/.048	.462/.522
$K = 4, P = 3, \rho = 0.5$	.051/.049	.513/.543
$K = 4, P = 3, \rho = 0.8$	.048/.049	.528/.526
$K = 4, P = 4, \rho = 0.2$ ($K-1 < P$)	.048/.042	.558/.691
$K = 4, P = 4, \rho = 0.5$	.054/.047	.664/.702
$K = 4, P = 4, \rho = 0.8$	.052/.047	.678/.690
$K = 5, P = 3, \rho = 0.2$ ($K-1 > P$)	.051/.045	.582/.688
$K = 5, P = 3, \rho = 0.5$	.052/.050	.680/.707
$K = 5, P = 3, \rho = 0.8$	.050/.049	.691/.704

U

	Complete null	Partial null
$K = 4, P = 3, \rho = 0.2$ ($K-1 = P$)	.607/.613	.532/.522
$K = 4, P = 3, \rho = 0.5$	.626/.620	.539/.543
$K = 4, P = 3, \rho = 0.8$	.611/.621	.528/.526
$K = 4, P = 4, \rho = 0.2$ ($K-1 < P$)	.747/.735	.690/.691
$K = 4, P = 4, \rho = 0.5$	.763/.751	.710/.702
$K = 4, P = 4, \rho = 0.8$	.743/.726	.678/.690
$K = 5, P = 3, \rho = 0.2$ ($K-1 > P$)	.753/.756	.701/.688
$K = 5, P = 3, \rho = 0.5$	.782/.774	.693/.707
$K = 5, P = 3, \rho = 0.8$	.782/.769	.691/.704

The more predictors in the multiple-regression analysis, the less was the familywise Type I error rate. The MS procedure was successful in maintaining the nominal familywise Type I error rate, with a maximum of .053.

The LSD method exhibited control over the familywise Type I error rate when there were only two predictors (average = .047, maximum = .053), which is consistent with Levin et al. (1994). With two predictors, the LSD and MS procedures are equivalent and so both of them produced the same results. The LSD method also performed well under the complete null situation no matter how many predictors (maximum familywise Type I error rate = .048, average = .045). However, that method was not acceptable in partial null situations with more than two predictors (maximum familywise Type I error rate = .145, average = .099).

Not surprisingly, the U method maintained the familywise Type I error rate of .05 only when there was just one true null regression coefficient. In other situations, the familywise Type I error rate increased as the number of null regression coefficients increased. With one null coefficient, the average familywise Type I error rate was .050; with two null coefficients, the average familywise Type I error rate was .095; and with three null coefficients, the average familywise Type I error rate was .135.

MANOVA Comparisons

Special case ($K = 3$ one-variable-at-a-time comparisons only).

Univariate contrasts in $K = 3$, $P = 2, 3, 4$ designs were examined with $n = 20$ and $n = 100$ participants per group. The MS method maintained the nominal familywise Type I error rate for both complete and partial null situations within an acceptable level (see Table 4). In the $K = 3$, $P = 2$ situation, the maximum proportion of familywise Type I errors was .055 and the average familywise Type I error rate was .044. In the $K = 3$, $P = 3$ situation, the maximum familywise Type I error rate was .050 and the average familywise Type I error rate was .036. In the $K = 3$, $P = 4$ situation, the maximum familywise Type I error rate was .045 and the average was .030.

The LSD method preserved the familywise Type I error rate only under the complete null situation, with a maximum Type I error rate of .052. In the partial null situation, the familywise Type I error rate was seriously inflated (maximum = .351). The U method failed to protect familywise Type I error rate as long as there was more than one true null comparison. The proportion of times there was at least one Type I error was as high as .400.

General case for $K = 2$ (both one-variable-at-a-time and multiple-variable contrasts).

When both one-variable-at-a-time and multiple-variable contrasts were analyzed in the simulation, the MRB method preserved the familywise Type I error rate at the desired level, with a maximum error rate of .052 (see Table 5). The LSD method maintained the nominal familywise Type I error rate only under the complete null situation, with a maximum familywise Type I error rate of .054. In the partial null situation, the error rate inflated to as much as .289. The U method completely failed to preserve the familywise Type I error rate, with a maximum of .375.

General case for $P > 1$, $K > 2$ (both one-variable-at-a-time and multiple-variable contrasts).

When multiple-variable (2 groups by 2 variables) and one-variable-at-a-time comparisons were considered following a significant omnibus test, either P or K could be reduced by 1 in the MRB critical value. Reducing either one of these was found to be adequate for preserving the familywise Type I error rate (see Table 6). Reducing the minimum of $K-1$ and P produced lower critical values and, therefore, and greater power.

For example, when $K = 4$, $P = 4$ ($K-1 < P$), reducing K yielded an acceptable familywise error rate (maximum = .024) that was higher than that associated with reducing P . Conversely, when $K = 5$, $P = 3$ ($K-1 > P$), reducing P yielded an acceptable familywise error rate (maximum = .029) that was higher than that associated with reducing K . When $K-1 = P$ (e.g., $K = 4$, $P = 3$), reducing either $K-1$ or P produced the same critical values. Both of these

were acceptable for preserving the familywise Type I error rate (maximum = .034).

In all situations, the LSD method preserved the familywise Type I error rate only under the complete null situation, with a familywise Type I error rate of .054. However, in partial null situations, familywise Type I error rates were enormously inflated under the LSD method (maximum = .707). The U approach failed to preserve the familywise Type I error rate as long as there was more than one true null comparison. The proportion of times there was at least one Type I error was as high as .782.

Conclusion

The choice of multiple-test procedures should be based on considerations of Type I and Type II error characteristics instead of either tradition or increasing power at the risk of obtaining spurious statistical significance (Seaman et al., 1991). The results of this study offer insights for both educational statisticians and researchers. What we have shown is that there are valid reasons for replacing traditional Scheffé-based methods with an improved sequential version of the test when the researcher is not interested in constructing confidence intervals. These modified procedures are not only able to preserve the familywise Type I error rate, but they are also easy to perform, being based on the same test statistics as in the simultaneous case and merely requiring reducing the critical value at the second step. The modified Scheffé-based procedures are recommended if researchers need a more powerful and adequate alternative to the original procedures.

The choice of procedures depends on the parameter specifications. When v_1 is equal to 2, Fisher LSD-like methods are the most powerful while providing adequate familywise Type I error protection. Thus, for 3 x 2 designs in factorial ANOVA, when there are only 2 predictors in multiple regression, or for any situation in which v_1 equals 2 in MANOVA, LSD is recommended. However, beyond $v_1 = 2$ situations, the familywise Type I error rate is seriously inflated in partial null situations.

The results of this study demonstrate that indiscriminate use of LSD-like methods or U methods lead to inflated familywise Type I

error rates. Researchers are cautioned about applying these two classes of procedure in general multiple-regression analysis and MANOVA situations.

There should be an investigation of the power of the modified Scheffé-based methods relative to other commonly applied multiple-test procedures. This article was restricted to considering only ideal specifications with normal distributions and balanced designs. It is important to determine how robust modified Scheffé-based methods are in preserving the familywise Type I error rate under less than ideal distributional and design conditions.

References

- Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.
- Klockars, A. J., & Hancock, G. R. (2000). Scheffé's more powerful *F*-protected post hoc procedure. *Journal of Educational and Behavioral Statistics*, 25, 13-19.
- Marascuilo, L. A., & Levin, J. R. (1970). Appropriate post hoc comparisons for interaction and nested hypotheses in analysis of variance designs: The elimination of Type IV errors. *American Educational Research Journal*, 7, 397-421.
- Meyers, J. L., & Beretvas, S. N. (2003, Apr.). *Type I error rates and power of variations of Scheffé's MCP*. Paper presented at the annual meeting of the American Educational Association, Chicago.
- Roy, S. N., & Bose, R. C. (1953). Simultaneous confidence interval estimation. *Annals of Mathematical Statistics*, 24, 513-536.
- Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika*, 40, 87-104.
- Scheffé, H. (1970). Multiple testing versus multiple estimation: Improper confidence sets, Estimation of directions and ratios. *Annals of Mathematical Statistics*, 41, 1-29.
- Seaman, M. A., Levin, J. R., & Serlin, R. C. (1991). New developments in pairwise multiple comparisons: Some powerful and practicable procedures. *Psychological Bulletin*, 110, 577-586.

Serlin, R. C. (2000). Testing for robustness in Monte Carlo studies. *Psychological Methods*, 5, 230-240.

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

Author Note

This study was conducted as part of the Xinyue Zhou's doctoral research in the Department of Educational Psychology at the University of Arizona. We are grateful to Jerome D'Agostino, Rosemary Rosser, Janice Streitmatter, and Rebecca Gomez for serving on the dissertation committee, and to Patricia Jones for her assistance with the simulations.