DOCUMENT RESUME

ED 404 361 TM 026 157

AUTHOR Barcikowski, Robert S.; Elliott, Ronald S.
TITLE Single Group Repeated Measures Analysis: Pairwise Multiple Comparisons under Bradley's Stringent Criterion.
PUB DATE Oct 96
NOTE 15p.; Paper presented at the Annual Meeting of the Mid-Western Education Research Association (Chicago, IL, October 2-5, 1996).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Comparative Analysis; *Educational Research; *Monte Carlo Methods; *Research Design; Research Methodology; Robustness (Statistics)
IDENTIFIERS Paired Comparisons; Power (Statistics); *Repeated Measures Design; *Single Group Design; Type I Errors; Type II Errors

ABSTRACT

A large number of pairwise multiple comparisons (P-MCPs) have been introduced recently to the educational research community. The use of these P-MCPs with single group repeated measures data was studied through an exploratory Monte Carlo study of P-MCPs that have been shown to control different types of Type 2 error and Type 1 familywise error under both no violations and violations of assumptions in other designs. A second purpose of the study was to recommend the P-MCPs based on ease of use. The stringent level of robustness developed by J. V. Bradley (1978) was used to examine the P-MCPs empirical rate of Type 1 error, and the range of sphericity was expanded to cover the values found in practice more realistically. Pairwise power among the P-MCPs was also compared. Nine P-MCPs were studied. Results indicate that all the new methods can not be recommended with single group repeated measures designs because their omnibus tests failed to control Type I error adequately. A familiar and easy-to-calculate method, the Dunn-Bonferroni procedure, successfully controlled familywise Type I error and may be recommended for use as a followup procedure with single group repeated measures designs. Further research with single group repeated measures designs through the Studentized maximum modulus statistic is recommended. (Contains 3 tables and 27 references.) (SLD)

Single Group Repeated Measures Analysis: Pairwise Multiple Comparisons Under Bradley's Stringent Criterion

Robert S. Barcikowski
Ronald S. Elliott
Ohio University

A paper presented at the annual meeting of the Mid-Western Education Research Association, Chicago, October, 1996.

Single Group Repeated Measures Analysis: Multiple Comparisons Under Bradley's Stringent Criterion

Objectives

The main purpose of this research was to provide educational researchers with a choice of pairwise multiple comparison procedures (P-MCPs) to use with single group repeated measures data. This was done through an exploratory Monte Carlo study of P-MCPs that have been shown to control different types of Type 2 error and Type 1 familywise error under both no violations and violations of assumptions in other designs. A second purpose, was to recommend one or more of the P-MCPs to educational researchers based on ease of use. This study expanded the previous work done in this area (e.g., Maxwell (1980), Boik (1981), Alberton and Hochberg (1984), Keselman, Keselman and Shaffer (1991), Keselman (1994), Keselman and Lix (1995)) by:

(a) using Bradley's(1978) stringent level of robustness to examine the P-MCPs empirical rate of Type I error ($\hat{\alpha}$) as compared with the nominal familywise level of significance (α);
(b) expanding the range of sphericity (as measured by ε) considered to more realistically cover those values found in practice (Green and Barcikowski, 1992);
(c) comparing per-pair power among the P-MCPs by finding the number of units (n's) necessary to reach per-pair power of .80.

Perspectives

P-MCPs Studied

A great deal of work has been done recently in the development of new and competing P-MCPs (Seaman, Levin, and Serlin, 1991). Many of these new P-MCPs have been adapted for use in split-plot repeated measures designs in papers written by the Keselmans and their colleagues (Keselman, Keselman and Shaffer (1991), Keselman Carriere and Lix (1993), Keselman (1994), Keselman and Lix (1995)). In this paper the following P-MCPs, described in detail by Maxwell (1980), Keselman (1994), and Keselman and Lix (1995) were examined for use with single group repeated measures data: 1) Tukey's T procedure (also known as the Studentized range procedure) (Tukey, 1953), 2) A modification of Tukey's T suggested by Keppel (1973) and studied by Maxwell (1980), 3) Dunn-Bonferroni controlled t-tests (DB), 4) Shaffer's (1986) sequentially rejective Bonferroni procedure (SB), 5) Hayter's (1986) two-stage modification of Fisher's Least Significant Difference test (HF), 6) A modified range procedure that combines the work of Shaffer(1979, 1986), Ryan(1960) and Welsch (1977) (SRW), 7) A multiple range procedure based on

Ryan-Welsch critical values (MRW), 8) Peritz's (1970) procedure (P), and 9) Welsch's (1977) step-up procedure (W).

These P-MCPs were selected for study because they were found to be at least partially successful in controlling different types of Type 2 error and Type 1 familywise error in previous studies. The first three procedures were used by Maxwell (1980) in his study of this problem, and procedures 4 through 8 were found by Keselman and Lix (1995) to be robust to violations of normality, multisample sphericity and heterogeneity of variance-covariance matrices with unequal cell sizes in split-plot designs using Bradley's liberal criterion. Keselman and Lix (1995) examined procedures 4 through 8 using an overall Welch-James-Johansen (WJJ) overall multivariate test (Johansen, 1980). and Satterthwaite (1941) adjusted degrees of freedom (SDF) as described by Keselman, Keselman and Shaffer (1991). They also modified the range procedures (SRW, MRW, P) by using a process described by Duncan (1957). Keselman (1994) recommended the Welsh step-up procedure with SDF degrees of freedom for use with split-plot repeated measures designs over twenty-seven other methods that he studied. Therefore, the first three procedures are generally familiar to most educational researchers and they provided check points with Maxwell's study. The second six procedures were found to be effective under more severe violations of assumptions, and were expected to perform well in this study of a simpler design.

The T, K, DB, and W P-MCPs were studied without an overall test. The T, K, DB P-MCPs are called *simultaneous* procedures because they use a single critical value to test all pairwise differences. The SB, FH, SRW, MRW, P and W are referred to as *stepwise* or *sequential* procedures because they test stages of hypotheses in a stepwise fashion, usually using a different critical value at each stage. SB, FH, SRW, MRW, and P were to be examined after first being preceded the WJJ test. The FH procedure was to be studied after being preceded by Keppel's q-statistic based on the Studentized-range. The SRW, MRW, and P range procedures were to be conducted with the modification described by Duncan (1957).

Background Equations

The P-MCPs examined in this study may be better understood through the following set of equations. In the following equations we are comparing pairs of means from a set of J means where $i, j = 1, 2, ..., J$ and $i \neq j$. Then, S^2 is the mean square error (i.e., the mean square within, or residual) of the analysis of variance considered, and S_i^2 and S_j^2 are the variances of treatments or measures i and j, with sample sizes n_i and n_j, respectively. When all treatments or measures have an equal number of units, the treatment or measure sample size is denoted by n. The general form of these equations is found in Equation 1.

Equation 1: General Form.

$$TS_{ij} \geq CV_{ij,\alpha,v} * Con \qquad (1)$$

The term TS_{ij} is the calculated test statistic in the form of a *t* statistic for various situations, and the term $CV_{ij,\alpha,v}$ is a critical value with familywise error of α and error degrees of freedom v. The term Con is a constant which allows the equation to be valid. When the calculated test statistic TS_{ij} is greater than or equal to $CV_{ij,\alpha,v}$ times Con, mean i is said to differ significantly from mean j.

Equation 2: Equal n, Homogeneous Variances.

$$TS_{ij} = (\bar{Y}_i - \bar{Y}_j) / (S^2 / n)^{1/2} \geq CV_{ij,\alpha,v} * CON \qquad (2)$$

The typical example for this equation is Tukey's HSD used to compare all pairs of means in a one-way ANOVA with J treatments. Then, $CV_{ij,\alpha,v}$ is the Studentized Range Statistic and Con = 1.0. For example, in a one-way ANOVA with J = 5, n =9 units (e.g., subjects) per treatment, and α =.05, we have that $CV_{ij,.05,40} = q_{\alpha,j,v} = q_{.05,5,40} = 4.04$ for all paired comparisons.

Equation 3: Unequal n, Homogeneous Variances.

$$TS_{ij} = (\bar{Y}_i - \bar{Y}_j) / (S^2 / n_i + S^2 / n_j)^{1/2} \geq CV_{ij,\alpha,v} * CON \qquad (3)$$

Equation 4: Unequal n, Heterogeneous Variances.

$$TS_{ij} = (\bar{Y}_i - \bar{Y}_j) / (S_i^2 / n_i + S_j^2 / n_j)^{1/2} \geq CV_{ij,\alpha,v} * CON \qquad (4)$$

Equation 5: Equal n, Heterogeneous Variances, correlated measures.

$$TS_{ij} = (\bar{Y}_i - \bar{Y}_j) / ((S_i^2 + S_j^2 - 2S_{ij}) / n)^{1/2} \geq CV_{ij,\alpha,v} * CON \qquad (5)$$

Where S_{ij} is the covariance between measures i and j and for single group repeated measures designs v = n-1.

Equation 5 may be used to illustrate all of the P-MCPs considered in this study, except the T procedure which uses Equation 2. This can be done with the assistance of Table 1 which provides information on the test statistics and how their levels of significance and "steps between means" degrees of freedom are determined in order to control familywise error rate. *Familywise error* is the probability of making at least one Type I error when testing a family of hypotheses.

An example of where Equation 5 might be used is in a single group repeated measures analysis with J = 7 measures on n = 25 subjects. Maxwell (1980) recommended the Dunn-Bonferroni approach to determine which pairs of means differed. Using the Dunn-Bonferroni approach, and the aid of Equation 5 and Table 1, we have that $CV_{ij,\ \alpha,\ v}$ is student's t-statistic with $\alpha' = 2\alpha/(J*(J-1)) = .00238$ and v = n-1 = 24 degrees of freedom. Then, $CV_{ij,\ .05,\ 24} = t_{.00238,24} = 3.396$ and Con = 1.0 for all paired comparisons.

Method

The complexity and number of conditions to be compared necessitated a Monte Carlo study. In order to investigate the Type 1 and Type 2 error rates the following characteristics of the single group design were manipulated: (1) the number of repeated measures (J = 3, 4, 5, 6, 8, 10), (2) the value of sphericity (for each J four values of ε were examined, ε = .50, .75, and 1.0 plus a value near the minimum for ε, i.e., for k = 3, ε = .51; J = 4, ε = .40; J = 5, ε = .30; J = 6, ε = .30; J = 8, ε = .20; J = 10, ε = .20;), and (3) the shape of the population (normal, nonnormal with skewness = 1.75, and kurtosis = 3.75). The number of repeated measures and the values of sphericity were based on a study by Green and Barcikowski (1992) and the shape of the nonnormal distribution was close to that chosen by Keselman (1994) (skewness = 1.633, and kurtosis = 4.0), based on an investigation by Micceri(1989). A FORTRAN program was used to generate the repeated measures normal data following procedures described by Keselman (1994). Nonnormal data were generated using procedures described by Fleishman (1978) and Vale and Maurelli (1983). Given a .05 level of significance, each condition was replicated 5,000 times for both power and Type 1 error rates.

Bradley's (1978) stringent criterion was used because past research, i.e., Seaman, Levin and Serlin (1991) and Keselman and Lix (1995) had indicated the potential for one or more of these P-MCPs to meet this criterion. Also, for reasons to be described when sample size is discussed, we were not as concerned with a P-MCP whose familywise α was less than Bradley's lower bound. Bradley's stringent criterion is to be considered robust when a P-MCP's empirical rate of Type 1 error ($\hat{\alpha}$) is contained in the interval $\alpha \pm 0.1\ \alpha$. For $\alpha = .05$, a P-MCP was considered robust if it fell in the interval $.04 \le \hat{\alpha} \le .06$.

Table 1
Each Pairwise Multiple Comparison Procedure Used in This Study, Its Abbreviation, Type I Error Similarity, Test Statistic, Critical Value α', and *q* Statistic Degrees of Freedom for Steps Between Means

Test	Letter ID	Type I Letter[c]	Test Statistic[d]	Critical Value α'	Df1[f]
Simultaneous Tests: No Omnibus Test					
(1) Tukey[a]	T	a	q	CT[e]	J[g]
(2) Keppel[b]	K	b	q	CT	J
(3) Dunn-Bonferroni	DB	c	t	$2\alpha/(J(J-1))$	–
Stepwise Tests: Preceded By Omnibus Test[h]					
(4) Schaffer-Bonferroni	SB	d	t	α/x[i]	–
(5) Hayter-Fisher	FH	d	q	CT	J-1
(6) Schaffer-Ryan-Welsch	SRW	d	q	Tukey-Welsch[j]	etc.[k]
(7) Multiple Range Ryan-Welsch	MRW	d	q	Tukey-Welsch[j]	etc.[l]
(8) Peritz	P	d	q	Tukey-Welsch[m]	etc.[l]
Stepwise Test: No Omnibus Test					
(9) Welsch	W	e	w	CT	etc.[l]

Note. When the Studentized Range Statistic, *q*, is the critical value, CON = $(2)^{-1/2}$ in Equation 5. When Student's *t* or Welsch's *w* are the critical values, CON = 1.0.
[a]Uses Equation 2 with pooled error term and degrees of freedom for error, $v = (n - 1)(J - 1)$. [b]Called SEP1 by Maxwell (1980) to indicate use of Equation 5 with $CV_{ij,\alpha,v} = q_{\alpha,J,n-1}$. Maxwell (1980) attributed this testing procedure to Keppel (1973). [c]Tests with the same letter have the same Type I error based on their first test. [d]The test statistics are the Studentized Range statistic *q*, Student's *t* statistic, and Welsch's *w* statistic. [e]CT (controlled by testing) indicates that the familywise level of significance (α) is controlled by the testing process and does not have to be modified by the user. [f]Df1 is the degrees of freedom for the *q* and *w* statistics based on the number of means or number of steps between means. [g]J is the number of repeated measures. [h]The possible omnibus tests considered here were: (1) Hotelling's T^2, (2) the Greenhouse-Geisser adjusted F test, (3) The Welch-James-Johansen multivariate test statistic, (4) the Keppel Studentized Range Test. [i]Values for x are tabled in Schaffer (1986). [j]The level of significance used at each step is found as $\alpha' = \alpha_p = 1-(1-\alpha)^{p/J}$ $(2 < p < J-2)$, $\alpha_{J-1} = \alpha_J = \alpha$ this and the testing process control the familywise error rate to be α. [k]Following the overall test the next two tests of means separated by J and J-1 steps are tested using Df1 = J-1 with an additional 1 subtracted from the Df1 from a previous step at the J-2 and subsequent steps. [l]Dfl = J at the first step and 1 is subtracted from the Df1 from a previous step at

the J-1 and subsequent steps. [m]The Peritz procedure makes use of the Tukey-Welsch and Newman-Keuls stepwise procedures as described by Hochberg and Tamhane (1987, pp.120-124).

Per-pair power (the probability that a true difference between two specified means will be detected) was investigated by setting two means at .3 and -.3 with the other means set at zero. Sample size (n) for each case was then found such that power was as close to .80 as possible (at n-1 power was less than .80). Per-pair power was investigated because of results and reasoning given by Seaman, Levin and Serlin (1991). All pairs power (the probability that all true pairwise mean differences will be detected) was found by Seaman et al. (1991) to be highly correlated with per-pair power ($r > .90$), and any-pair power (the probability that at least one true pairwise mean difference will be detected) *was found to differ comparatively little among procedures, generally centering around the theoretical omnibus-test powers* (p. 581).

We found the sample size necessary for per-pair power to be .80 because, based on the results of Keselman and Lix (1995), we expected these n's to differ by only a few units across P-MCPs. This would be an important finding if a P-MCP failed to meet Bradley's stringent criterion only at its lower bound, but could reach power of .80 with only one or two more units than the n needed for a P-MCP that failed to reach Bradley's criterion at the upper bound or the n needed for a P-MCP that was much more difficult to calculate.

Results

Type I Error

As a check on our procedures, we replicated Maxwell's (1980) results for WSD, Dunn-Bonferroni, and Keppel. We found that our results (not shown here) were consistent with Maxwell's to within $\hat{\alpha} \pm .005$. Our results when we tested the full null hypothesis (i.e., that all of the means for a given single group repeated measures design were equal) are presented in Table 2 for Wilks's overall multivariate test, WJJ, T, K, W, and DB. We included Wilks's tests as a further check on our process, because it should have found (and did find) empirical error rates that were within Bradley's stringent criteria.

Welch-James-Johansen. The results for the WJJ test indicated that with a sample size of fifteen units, the $\hat{\alpha}$'s became too liberal (i.e., $\hat{\alpha} > .06$) when the ratio of number of units to the number of measures became less than or equal to 3 to 1, i.e., $n/J \leq 3$, and that this situation became worse as sphericity dropped. These results are similar to those found by Keselman, Carriere, and Lix (1993) for repeated measures main effects in unequal n split-plot designs. The latter authors found...*that, for normally distributed data, the number of subjects in the smallest of the unequal groups should be 2 to 3 times the number of repeated measurements minus one in order to achieve reasonable Type I error protection. (p. 311)*

Table 2
Empirical Type I error rates ($\hat{\alpha}$'s) for the full null hypothesis.

J	n	ε	Wilks	Welch-James Johansen	Tukey WSD	Keppel	Welsch	Dunn-Bonferroni
3	15	.51	0490	0500	0854*	0408	0788*	0356**
		.75	0532	0542	0686*	0476	0654*	0394**
		1.00	0496	0504	0496	0492	0514	0414
4	15	.40	0552	0598	0994*	0504	1028*	0382**
		.50	0482	0540	0822*	0532	0928*	0396**
		.75	0520	0542	0658*	0588	0722*	0440
		1.00	0466	0530	0460	0602*	0508	0464
5	15	.30	0462	0592	1178*	0552	1188*	0370**
		.50	0540	0662*	0980*	0606*	0948*	0404
		.75	0488	0604*	0680*	0660*	0698*	0436
		1.00	0474	0600	0460	0672*	0532	0454
6	15	.30	0508	0748*	1204*	0554	1270*	0328**
		.50	0456	0666*	0946*	0596	0970*	0352**
		.75	0590	0838*	0698*	0628*	0734*	0384**
		1.00	0494	0704*	0482	0646*	0482	0380**
8	15	.20	0520	1272*	1542*	0594	1622*	0324**
		.50	0486	1252*	1100*	0644*	1088*	0356**
		.75	0514	1262*	0762*	0676*	0764*	0380**
		1.00	0470	1168*	0458	0712*	0496	0398**
10	15	.20	0456	2092*	1852*	0730*	1940*	0398**
		.50	0544	2346*	1136*	0776*	1210*	0428
		.75	0482	2160*	0902*	0776*	0826*	0436
		1.00	0526	2212*	0542	0832*	0534	0442

Note. An * indicates that the empirical error rate was greater than Bradley's upper confidence value of .06, and an ** indicates that the empirical error rate was less than Bradley's lower confidence value of .04.

Tukey and Welsch. The T and W procedures yielded very similar results. In Table 2 both procedures yielded empirical error rates within Bradley's stringent confidence bounds only when sphericity was equal to one ($\varepsilon = 1.00$). Both procedures were too liberal ($\hat{\alpha} > .06$) when sphericity was less than one, having higher $\hat{\alpha}$'s as sphericity decreased.

Keppel and Dunn-Bonferroni. In Table 2, the K procedure yielded $\hat{\alpha}$'s that became too liberal ($\hat{\alpha} > .06$) as the number of measures increased and as the measure of sphericity increased. The DB procedure yielded error rates that averaged .04, and that dropped below .04 at levels of sphericity that were close to our minimum values.

Sample Size For Power Of .80

As a result of the liberal $\hat{\alpha}$'s values found for WJJ, T, and W, these procedures were not considered further in our sample size calculations. This caused the SB, FH, SRW, MRW, and P procedures to also be eliminated because they are dependent on the overall WJJ and K tests.

We decided to investigate sample size for power of .80 for the DB procedure because it controlled $\hat{\alpha}$ below, but close to, Bradley's lower limit. We also decided to reconsider Type I error for the K procedure because its error rate seemed to be related to the unit/measure (n/J) ratio, and because the $\hat{\alpha}$'s reported in Table 5 where within Bradley's liberal criterion of robustness (i.e., $.025 < \hat{\alpha} < .075$ for $\alpha = .05$) for all values except those with $J = 10$ and $\varepsilon > .20$. We considered both K's and DB's Type I error rate under both normality and nonnormality, using the sample size found to have power of .80 for the DB procedure. This process was used because if the n needed for K to have power of .80 did not control Type I error, the DB procedure would be a better choice.

The results for the latter analyses are shown in Table 3. In Table 3 the sample sizes needed for power of the DB procedure to reach .80 under normality are the same in most cases as the n's found under the nonnormal situation, requiring an additional unit for $J = 3$, $\varepsilon = .51$ and $J=4$, $\varepsilon = .40$. For these sample sizes the Type I error shown in Table 3 was similar to that found with 15 cases in Table 2 under normality, but is more conservative (approximately .02) for the nonnormal cases. The K procedure was too liberal ($\hat{\alpha} > .06$) for several cases when the n/J ratio was less than 3 and ε approached 1.0. The K procedure was conservative, with $\hat{\alpha}$ approximately equal to .04 under nonnormality.

Discussion

This study was an exploratory look at P-MCP's that had been found to control familywise Type I error in more complex designs, and therefore, were expected to also be similarly effective in the simpler single group repeated measures design. This was not found to be true. The reason for this may be that in the single group

Table 3
Sample size for power of .80 with the Dunn-Bonferroni procedure and empirical Type I error rates (full hypothesis) for this sample size.

		Normality				Nonnormality			
				Type I Error				Type I Error	
J	ε	n	Power	DB	K	n	Power	DB	K
3	.51	32	7748			33	7976	___[a]	___
		33	8048	0314**	0366**	34	8100	___	___
	.75	8	7684			8	7782		
		9	8574	0440	0544	9	8410	0260**	0380**
	1.00	8	7476			8	7622		
		9	8402	0452	0578	9	8288	0268**	0364**
4	.40	8	6960			9	7946		
		9	8023	0392**	0566	10	8482	0280**	0414
	.50	9	7598			9	7668		
		10	8356	0432	0586	10	8222	0242**	0374**
	.75	9	7372			9	7474		
		10	8140	0480	0630*	10	8058	0212**	0350**
	1.00	9	7298			9	7432		
		10	8090	0482	0684*	10	8018	0199**	0324**
5	.30	10	7946			10	7908		
		11	8648	0348**	0556	11	8396	0368**	0368**
	.50	10	7420			10	7532		
		11	8206	0370**	0596	11	8026	0240**	0342**
	.75	10	7304			11	7932		
		11	8114	0399**	0620*	12	8402	0186**	0380**
	1.00	10	7264			11	7890		
		11	8058	0432	0680*	12	8356	0194**	0368**
6	.30	11	7744			11	7696		
		12	8448	0358**	0588	12	8316	0256**	0440
	.50	11	7568			11	7606		
		12	8314	0368**	0638*	12	8184	0240**	0380**
	.75	11	7512			11	7560		
		12	8258	0386**	0654*	12	8150	0344**	0046
	1.00	11	7512			11	7560		
		12	8258	0418	0680*	12	8150	0170**	0340**
8	.20	12	7760			12	7740		
		13	8416	0320**	0618*	13	8244	0240**	0402
	.50	12	7512			12	7522		
		13	8202	0354**	0672*	13	8088	0156**	0368**
	.75	12	7518			12	7476		
		13	8202	0374**	0696*	13	8056	0148**	0340**
	1.00	12	7470			12	7476		
		13	8148	0392**	0734*	13	8056	0142**	0330**
10	.20	13	7480			13	7548		
		14	8218	0354**	0764*	14	8046	0224**	0428
	.50	13	7410			13	7492		
		14	8148	0388**	0808*	14	8004	0172**	0384**
	.75	13	7362			14	7958		
		14	8078	0390**	0834*	15	8394	0154**	0386**
	1.00	13	7362			14	7960		
		14	8078	0406	0868*	15	8394	0132**	0370**

Note. An * indicates that the empirical error rate was greater than Bradley's upper confidence value of .60, and an ** indicates that the empirical error rate was less than Bradley's lower confidence value of .40.

[a] The variance covariance was singular under nonnormality.

design the adjusted degrees of freedom (SDF) reduce to n-1 and do not involve the treatment variances as is true in more complex designs.

Based on our results we recommend that further research with single group repeated measures P-MCP's be done using the Studentized maximum modulus statistic recommended by Alberton and Hochberg (1984). The Studentized maximum modulus statistic yields critical values that fall between the DB t statistic and the K q statistic. If the Studentized maximum modulus statistic proves to be successful, it could be used as the test statistic with the SB, FH, SRW, MRW, and P procedures. Also, power should be studied under a wide variety of mean patterns and variance-covariance structures because past studies (e.g., Klockars and Hancock, 1992; Seaman, Levin, and Serlin, 1991) have indicated that different MCP's are more powerful with different mean patterns and this will probably be exacerbated with different variance-covariance structures.

Recommendations for Practitioners

Recently, a large number of pairwise multiple comparison procedures were introduced to the educational research community. This study considered the use of some of the more robust of these new methods with a single group repeated measures design over a range of nonsphericity values. The results indicated that all of the new methods could not be recommended for use with single group repeated measures designs because their omnibus tests failed to adequately control Type I error. However, a familiar and easy to calculate method, the Dunn-Bonferroni procedure, did successfully control familywise Type I error and may be recommended for use as a follow-up procedure with single group repeated measures designs.

References

Alberton, Y., & Hochberg, Y. (1984). Approximations for the distribution of a maximal pairwise t in some repeated measures designs. Communications in Statistics, 13(Series A), 2847-2854.

Boik, R. J. (1981). A priori tests in repeated measures designs: Effects of nonsphericity. Psychometrika, 46, 241-255.

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

Duncan, D. B. (1957). Multiple range tests for correlated and heteroscedastic means. Biometrics, 13, 164-176.

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

Green, S. & Barcikowski, R. S. (1992, April). Sphericity in the repeated measures univariate mixed-model design. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Hayter, A. J. (1986). The maximum familywise error rate of Fisher's least significant difference test. Journal of the American Statistical Association, 81, 1000-1004.

Hochberg, Y. & Tamhane, A. C. (1987). Multiple comparison procedures. New York, NY: Wiley.

Johansen, S. (1980). The Welch-James approximation of the distribution of the residual sum of squares in weighted linear regression. Biometrika, 67, 85-92.

Keppel, G. (1973). Design and analysis: A researcher's handbook. Englewood Cliffs, N.J.: Prentice-Hall.

Keselman, H. J. (1994). Stepwise and simultaneous multiple comparison procedures of repeated measures' means. Journal of Educational Statistics, 19(2), 127-162.

Keselman, H. J., Keselman, J. C., & Shaffer, J. P. (1991). Multiple pairwise comparisons of repeated measures means under violation of multisample sphericity. Psychological Bulletin, 111(1), 162-170.

Keselman, H. J. & Lix, L. M. (1995). Improved repeated measures stepwise multiple comparison procedures, Journal of Educational Statistics, 20(1), 83-99.

Keselman, H. J., Carriere K. C., & Lix L. M. (1993). Testing repeated measures hypotheses when covariance matrices are heterogeneous. Journal of Educational Statistics, 18, 305-319.

Klockars, A. J. & Hancock G. R. (1992). Power of recent comparison procedures as applied to a complete set of planned orthogonal contrasts. Psychological Bulletin, 111, 505-510.

Lix, M. L. & Keselman, H. J. (1995). Approximate degrees of freedom tests: A unified perspective on testing for mean equality. Psychological Bulletin, 117(3), 547-560.

Maxwell, S. E. (1980). Pairwise multiple comparisons in repeated measures designs. Journal of Educational Statistics, 5, 269-287.

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

Peritz, E. (1970). A note on multiple comparisons. Unpublished manuscript, Hebrew University, Israel.

Ryan T. A. (1960). Significance tests for multiple comparisons of proportions, variances, and other statistics. Psychological Bulletin, 57, 318-328.

Satterthwaite, F. E. (1941). Synthesis of variance. Psychometrika, 6, 309-316.

Shaffer, J. P. (1979). Comparison of means: An F test followed by a modified multiple range procedure. Journal of Educational Statistics, 4, 14-23.

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

Seaman, M. A. Levin, J. R. & Serlin, R. C. (1991). New developments in pairwise multiple comparisons: Some powerful and practicable procedures. Psychological Bulletin, 110(3), 577-586.

Tukey, J. W. (1953). The problem of multiple comparison procedures. Unpublished manuscript, Princeton University.

Vale, C. D. & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. Psychometrika, 48, 465-471.

Welsch, R. E. (1977). Stepwise multiple comparison procedures. Journal of the American Statistical Association, 72, 566-575.

TM026157

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC®

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Single Group Repeated Measures Analysis: Pairwise Multiple Comparisons Under Bradley's Stringent Criterion

Author(s): Robert S. Barcikowski and Ronald S. Elliott

Corporate Source:

Publication Date: 10-5-96

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following two options and sign at the bottom of the page.

[X] Check here **For Level 1 Release:** Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical) ***and*** paper copy.

The sample sticker shown below will be affixed to all **Level 1** documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 1

The sample sticker shown below will be affixed to all **Level 2** documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 2

[ ] Check here **For Level 2 Release:** Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical), but ***not*** in paper copy.

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at **Level 1**.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Sign here→ please

Signature: Ronald S. Elliott

Printed Name/Position/Title: Ronald S. Elliott, Assoc. Prof. of Computer Science

Organization/Address: Ohio University - Chillicothe, 571 W. Fifth St., Chillicothe, OH 45601

Telephone: 614-774-7256

FAX: 614-774-7214

E-Mail Address: Elliott@oak.cats.ohiou.edu

Date: 10/16/96

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:

Address:

Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:

Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1301 Piccard Drive, Suite 100
Rockville, Maryland 20850-4305

Telephone: 301-258-5500
FAX: 301-948-3695
Toll Free: 800-799-3742
e-mail: ericfac@inet.ed.gov