Psychological Bulletin
1972, Vol. 77, No. 6, 405–408

EXTENSION OF MULTIPLE-RANGE TESTS TO INTERACTION TABLES IN THE ANALYSIS OF VARIANCE: A RAPID APPROXIMATE SOLUTION

DOMENIC V. CICCHETTI[1]

Veterans Administration Hospital, West Haven, Connecticut

The research literature discusses the multiple-range tests applied to means derived from the one-way analysis of variance, in which none of the possible contrasts is confounded. Often, however, it is neccessary to compare means in an interaction table derived from a factorial analysis of variance. In this case, the only unconfounded comparisons are those made within rows and columns. The present study discusses an approximate solution that adjusts the number of treatments by basing the q statistic upon the number of unconfounded comparisons only. The solution is then applied, using actual data, (*a*) when only the $K(K-1)/2$ contrasts are desired (the method of Tukey) and (*b*) when all possible contrasts are desired (the method of Scheffé). The Duncan and Newman-Keuls tests are deemphasized, since research demonstrates these tests fail to control adequately for Type I error.

The purpose of this study is to describe a new method of applying the multiple-range tests of Tukey[2] and Scheffé (1953) to interaction tables from the analyses of variance. I discuss the problem encountered when one applies multiple-range tests to means derived from an interaction table in a factorial analysis of variance; the attempted solution of Winer (1962); and the advantages of the present author's new method (Cicchetti)[3] over that of Winer.[4]

Multiple Comparisons (Problems of Interpretation)

Recently, Petrinovich and Hardyck (1969) have shown that the Tukey and Scheffé multiple-range tests adequately hold the alpha and beta errors at the .05 level, whereas the *t* test, the Duncan test, and the test developed by Newman and Keuls do not. Petrinovich and Hardyck applied the above tests to means derived from a one-way analysis of variance. The situation is more complex when the multiple-range tests are applied to means based upon the interaction of two or more factors. In the simplest 2×2 interaction table, there are four cell means and six possible paired comparisons, two (or one-third) of which are *not* readily interpretable, as shown in Table 1.

It is clear that if one compared either Cells A_1B_1 and A_2B_2 or Cells A_2B_1 and A_1B_2, one could not determine how much of the difference to attribute to Factor A, and how much to attribute to Factor B, all other things being equal. This problem of interpretation does not occur in the remaining four paired contrasts.[5]

[1] Also at Yale University. Requests for reprints should be sent to Domenic V. Cicchetti, Veterans Administration Hospital, West Haven, Connecticut 06516.

[2] J. W. Tukey. The problem of multiple comparisons. Unpublished manuscript, Princeton University, 1953.

[3] D. V. Cicchetti. Extension of multiple-range tests to interaction tables in analysis of variance. Paper presented at the American Statistical Association, New York, August 20, 1969.

[4] The author wishes to express his deepest appreciation to John W. Tukey for his helpful comments concerning the multiple-range adjustments developed in this study.

[5] Although it is true that the only unconfounded comparisons in any given interaction table are those made within rows and columns, this author can conceive of certain instances, especially in nonpsychological research, in which one might wish to make paired comparisons that are confounded. For example, let us assume that in a 2×2 experimental design, the two factors, A and B, refer to temperature and pressure, respectively, and moreover, that A_1B_1 represents the speed with which a given chemical reaction occurs under standard operating conditions, whereas A_2B_2 represents the speed with which the same chemical reaction occurs under conditions which are not standard. One might conceivably be interested in comparing A_1B_1 with A_2B_2

TABLE 1

Illustration of Confounded Comparisons in a 2×2 $(A \times B)$ Interaction Table

Factor	Factor	
	B_1	B_2
A_1	$A_1\ B_1$	$A_1\ B_2$
A_2	$A_2\ B_1$	$A_2\ B_2$

Note.—The four unconfounded comparisons are $A_1\ B_1$ versus $A_1\ B_2$; $A_2\ B_1$ versus $A_2\ B_2$; $A_1\ B_1$ versus $A_2\ B_1$; and $A_1\ B_2$ versus $A_2\ B_2$. The two confounded comparisons are $A_1\ B_1$ versus $A_2\ B_2$ and $A_1\ B_2$ versus $A_2\ B_1$.

The number of confounded comparisons increases the greater the number of cells in a given interaction table. Thus, for the simple 3×4 analysis of variance, the proportion of confounded comparisons is about 55%, as shown in Table 2. The importance of this fact is that the more comparisons one is making in any given experiment, the greater the probability that some of these comparisons will be significantly different from each other, by chance alone. In order to correct for this, the difference between any set of means, to be judged statistically significant, must *increase*, the greater the number of means one is comparing. If one bases the q statistic on 12 means (allowing one to make 66 comparisons) and only 30 of these comparisons make any logical sense, then one is *penalized* by being forced to accept a minimal significant difference based upon 66 comparisons, when only 30 of these can be meaningfully interpreted.

The Winer (1962) Approximation

A "fragmentized" approximate solution to this problem has been offered by Winer (1962), who presents the interaction table from a 2×3 factorial $(A \times B)$ analysis of variance with six independent groups, each containing three subjects. Winer's solution is to make all possible comparisons within each of the two separate, "logical groupings" (Winer, 1962, p. 238) of rows a_1 and a_2, respectively. Thus, within row a_1, the comparisons would be $a_1 b_1$ versus $a_1 b_2$; $a_1 b_1$ verus $a_1 b_3$; and $a_1 b_2$ versus $a_1 b_3$. The same logic would apply to row a_2, in which the comparisons would be $a_2 b_1$ versus $a_2 b_2$; $a_2 b_1$ versus $a_2 b_3$; and $a_2 b_2$ versus $a_2 b_3$. Winer's approach is partial in that it makes no provision for comparing the three pairs of logical groupings of unconfounded column means, that is, $a_1 b_1$ versus $a_2 b_1$; $a_1 b_2$ versus $a_2 b_2$; and $a_1 b_3$ versus $a_2 b_3$. One would guess he might consider each of the three column comparisons as a separate, two-mean contrast, in the same sense that each of the two sets of row means was considered as a set of three separate, paired-mean contrasts.

Even granting the above, however, one must regard Winer's solution as a fragmented one which treats a single experiment with six cell means as if it were five separate experiments, that is, one based upon the means within a_1; another containing the means in row a_2; and three experiments based upon the two means within each of the three columns, b_1, b_2, and b_3. Winer's attempted solution appears inadequate for the application of the Tukey or Scheffé methods, whose very logic depends upon basing the alpha error and the q statistic on the whole experiment, rather than upon fragmentized portions of it.

The Cicchetti Approximation

The logic of the author's adjustment is based upon the relationship between the number of K treatments and the number of $K(K-1)/2$ paired comparisons. An inspection of Table 3 reveals that any given, consecutive, increase in K results in an arithmetic

TABLE 2

Illustration of the 12 Cells in a 3×4 $(A \times B)$ Interaction Table

Factor	Factor			
	B_1	B_2	B_3	B_4
A_1	$A_1\ B_1$	$A_1\ B_2$	$A_1\ B_3$	$A_1\ B_4$
A_2	$A_2\ B_1$	$A_2\ B_2$	$A_2\ B_3$	$A_2\ B_4$
A_3	$A_3\ B_1$	$A_3\ B_2$	$A_3\ B_3$	$A_3\ B_4$

Note.—The total number of possible paired comparisons is given by $K(K-1)/2 = (12)(11)/2 = 66$. The number of unconfounded comparisons is given by $R[K(K-1)/2] = 3[(4)(3)/2] = 18$, for the rows and $C[K(K-1)/2] = 4[(3)(2)/2] = 12$, for the columns, giving a total of 30 unconfounded comparisons. By subtraction, there are $66 - 30$ or 36 or 55% confounded comparisons.

regardless of how much of the difference is attributable to the change in temperature and how much to the change in pressure. For comparisons of this type, the adjustments discussed in this study are not necessitated.

increase in the total possible number of $K(K-1)/2$ paired comparisons.

If one examines each successive number of K treatments in Table 3, it is seen that as K increases from 3, to 4, to 5, to 6, to 7, for example, the number of $K(K-1)/2$ comparisons increases from 3 to 6 (3), to 10 (4), to 15 (5), to 21 (6). This arithmetic progression continues to 190, an increase of 19 over the previous 171 $K(K-1)/2$ comparisons, based upon 19 treatments. For comparing means from a one-way analysis of variance, one should base the q value upon the number of K treatments and then perform the Tukey test if one is interested in the $K(K-1)/2$ paired contrasts only, or the Scheffé test if one wishes to perform the $K(K-1)/2$ comparisons, as well as other possible comparisons, based upon various combinations of the K treatments.

In order to solve the problem of confounded comparisons of treatments from interaction tables, in a factorial analysis of variance, the solution proposed here bases the number of treatments (K'), for the q statistic, upon the number of *unconfounded* comparisons only. Table 4 shows the relationship between the number of unconfounded comparisons and K'. Here, it can be seen that when the number of unconfounded comparisons is 3, the number of K' treatments is 3; when the number is 6, $K' = 4$; when there are 190 unconfounded comparisons, $K' = 20$. Table 4 is analogous to Table 3, except that Table 3 is based upon the total number of $K(K-1)/2$ comparisons, whereas Table 4 is based upon the number of unconfounded paired comparisons only. Table 4 also includes the interpolated (to the nearest whole number) intermediate K' values, that is, between 3 and 4, 5 and 8, 9 and 12 . . . , 181–200 $K'(K'-1)/2$ comparisons. Given this information, it is now possible to apply the proposed solution to both the Tukey and Scheffé adjustments. The data are provided by Winer (1962) from the results of a 2 × 3 analysis of variance, presented below in Tables 5 and 6.

TABLE 3

TOTAL NUMBER OF $K(K-1)/2$ PAIRED COMPARISONS AS A FUNCTION OF THE NUMBER OF K TREATMENTS

Number of K treatments	Number of $K(K-1)/2$ comparisons
3	3
4	6
5	10
6	15
7	21
8	28
9	36
10	45
11	55
12	66
13	78
14	91
15	105
16	120
17	136
18	153
19	171
20	190

TABLE 4

NUMBER OF ADJUSTED K' TREATMENTS AS A FUNCTION OF THE NUMBER OF UNCONFOUNDED PAIRED COMPARISONS IN A GIVEN INTERACTION TABLE[a]

Number of unconfounded comparisons	Number of adjusted K' treatments
3–4	3
5–8	4
9–12	5
13–17	6
18–24	7
25–32	8
33–40	9
41–50	10
51–60	11
61–72	12
73–84	13
85–98	14
99–112	15
113–128	16
129–144	17
145–162	18
163–180	19
181–200	20

[a] D. V. Cicchetti. Extension of multiple-range tests to interaction tables in analysis of variance. Paper presented at the American Statistical Association, New York, August 20, 1969.

The Tukey Application

The test used here is described in Snedecor (1956). The number of unconfounded paired contrasts is nine, as given by three within each of two rows (yielding six comparisons) plus one within each column (yielding another three comparisons).

TABLE 5

RESULTS OF ANALYSIS OF VARIANCE
(Winer, 1962, p. 234)

Source of variation	*SS*	*df*	*MS*	*F*
A	18.00	1	18.00	2.04
B	48.00	2	24.00	2.72
AB	144.00	2	72.00	8.15*
Within cell	106.00	12	8.83	
Total	316.00	17		

* $p < .001$.

The number of adjusted K' treatments, based upon nine *unconfounded* comparisons, is given in Table 4 as 5. The $q_{.05}$ statistic–based upon five adjusted treatments and $df = 12$ in the error term from the results of the analysis of variance in Table 5—is 4.51 (given in Snedecor, 1956, Table 10.6.1, p. 252). The standard error of the mean is given by $S_{\bar{x}} = \sqrt{8.83/3} = 1.71$. The smallest mean difference required for statistical significance at the .05 level, for the unconfounded contrasts only, is given by multiplying $q_{.05}$ by $S_{\bar{x}}$, that is, $(4.51)(1.71) = 7.71$, which is used for each of the nine unconfounded paired comparisons. Here a_2b_1 (with $\bar{x} = 10$) is statistically greater than a_2b_2 (with $\bar{x} = 2$), and a_2b_3 (with $\bar{x} = 12$) is statistically greater than a_2b_2 (with $\bar{x} = 2$), since *both* these mean differences exceed 7.71, required for significance at the .05 level.

The Scheffé Application

For the Scheffé test, the critical value of $F_{.95}$ is given by $(K - 1)$ multiplied by $F_{1-\alpha}\ (K - 1,\ df)$. The number of adjusted or K' treatments from Table 4 is 5, once again. Substituting in the formula above, we obtain

TABLE 6

MEANS OF A × B INTERACTION
(Winer, 1962, p. 235)

Factor	Factor		
	B_1	B_2	B_3
A_1	4	8	6
A_2	10	2	12

$(5 - 1)\ [F_{.95}\ (4,\ 12)] = 4\ (3.26) = 13.04$. Since $q = \sqrt{2F\ (\text{ratio})}$, the corresponding $q_{.05}$ statistic is $\sqrt{2\ (13.04)} = 5.11$. The $S_{\bar{x}}$ is once again 1.71. Therefore, the mean difference required for the .05 level is $(5.11)(1.71)$ or 8.74. The only paired contrast that reaches statistical significance, at the .05 level of confidence, is the a_2b_3 mean (value = 12), compared to the a_2b_2 mean (value = 2), since $12 - 2 = 10$, which is greater than the required 8.74 (also used for the combination K' treatment contrasts as well).

This study has attempted to extend the multiple-range tests developed by Tukey and Scheffé to interaction tables in the analysis of variance, ***without disturbing the logical principles upon which each test rests when applied to the one-way analysis of variance.*** Although the illustrations were based upon a factorial analysis with only one measurement per subject, the adjustment can just as easily be applied to a repeated-measures design, providing one uses the appropriate error terms (between or within) to obtain the standard error of the mean. Also, the example chosen contained equal numbers of subjects in each experimental condition. In analyses of variance employing ***unequal*** numbers of subjects, the procedures outlined in this article can be used in conjunction with the formulas provided by Kramer (1956). Thus, the adjustment described here has a complete range of applicability to the various possible analysis of variance designs.

REFERENCES

KRAMER, C. Y. Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics,* 1956, **13**, 307–310.

PETRINOVICH, L. F., & HARDYCK, C. D. Error rates for multiple comparison methods: Some evidence concerning the frequency of erroneous conclusions. *Psychological Bulletin,* 1969, **71**, 43–54.

SCHEFFÉ, H. A. A method for judging all possible contrasts in the analysis of variance. *Biometrika,* 1953, **40**, 87–104.

SNEDECOR, G. W. *Statistical methods.* (5th ed.) Ames: Iowa State College Press, 1956.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received August 24, 1970)