MULTIPLE COMPARISON TESTS FOR CONTRASTS AMONG CORRELATED CORRELATION COEFFICIENTS

Bradley E. Huitema, Western Michigan University

An important use of tests of the equality of two correlated correlation coefficients has been pointed out by Campbell (1971) and Kenny (1975). They suggest that an appropriate method of analyzing the nonequivalent groups pretest-posttest design is to compare the correlation between pretest scores and a group membership dummy variable with the correlation between posttest scores and the dummy variable. Hence, the pre and post scores are X (predictor) variables and the dummy variable is Y. If the difference between r_{yx_1} and r_{yx_2} is significant, it is concluded that treatment effects exist. If more than two observation periods are involved (e.g., pre, post #1 and post #2) it is necessary to analyze the difference among three or more correlated correlation coefficients.

It appears that multiple comparison tests for the case of correlated correlation coefficients have not yet been suggested. The purpose of this paper is to propose three methods of dealing with several correlated correlations.

Method A

Method A is suggested for tests on all pairwise comparisons. This method involves two stages:

Stage I: Test the overall hypothesis

$$H_0: \rho_{yx_1} = \rho_{yx_2} = \cdots = \rho_{yx_m}$$

(where ρ is the population correlation coefficient, y is the dependent variable and x_1 through x_m are the predictor variables) using the generalization of Hotelling's well known test of the equality of two correlated correlation coefficients (Hotelling, 1940). If this test is significant, proceed to stage II.

Stage II: Compute Hotelling's two predictor test for each pairwise contrast.

Computation Procedure

The computational formula required for Stage II tests is presented in several texts (e.g., McNemar, 1968; Walker and Lev, 1953) and will not be repeated here. The computational steps required for the m predictor generalization of Hotelling's two predictor test (i.e., Stage I) are provided below.

Step 1

Compute $\underline{R}_x$, the intercorrelation matrix of all predictor variables.

Step 2

Compute $\underline{R}_x^{-1}$, the inverse of $\underline{R}_x$.

Step 3

Compute $\underline{r}_{xy}$, the column vector of correlations between the dependent variable and each predictor, i.e.,

$$\underline{r}_{xy} = \begin{bmatrix} r_{x_1 y} \\ r_{x_2 y} \\ \cdot \\ \cdot \\ \cdot \\ r_{x_m y} \end{bmatrix}$$

Step 4

Compute the coefficient of multiple determination $R^2_{yx_1,x_2,\ldots x_m}$.

Step 5

Obtain $\sum\sum c_{ij}$, the sum of the elements of $\underline{R}_x^{-1}$.

Step 6

Obtain $c_1, c_2, \ldots c_m$, the sums of rows 1 through m of $\underline{R}_x^{-1}$.

Step 7

Compute the weights $w_1, w_2, \cdots w_m$ where

$$w_i = \frac{c_i}{\sum\sum c_{ij}}$$

The sum of the weights = 1. The column vector of weights is denoted $\underline{w}$.

Step 8

Compute h where

$$h = \underline{w}' \underline{r}_{xy}$$

Step 9

Compute the product $h^2(\sum\sum c_{ij})$.

Step 10

Compute the F statistic using

$$F = \frac{[R^2_{yx_1,x_2,\cdots x_m} - h^2(\sum\sum c_{ij})]/m-1}{(1 - R^2_{yx_1,x_2\cdots x_m})/N-m-1}.$$

The obtained F is evaluated with $F_{\alpha,m-1,N-m-1}$ where N is the total number of subjects and m is the number of predictor variables.

Method B

Method B is appropriate when there is interest in testing the m-1 control versus treatment correlations. Suppose $r_{x_1 y}$ is the sample control correlation and r_{x_2y} and r_{x_3y} are the sample treatment correlations. The m-1 null hypotheses are

$$H_0: \rho_{x_1y} = \rho_{x_2y} \text{ and}$$

$$H_0: \rho_{x_1y} = \rho_{x_3y}.$$

Hotelling's conventional two predictor formula is applied to each contrast. The obtained t statistics are not, however, evaluated using student's t distribution. Rather, each obtained t is evaluated using Dunnett's t statistic (Dunnett, 1955). The critical value is based on m and N-3 degrees of freedom. No preliminary (Stage I) test is required with this method.

Method C

Method C is most appropriate if a relatively large number of correlations are involved but the researcher has interest in making only a few planned comparisons. As with Method B, no preliminary test is employed. Hotelling's two predictor formula is employed for each planned contrast. The obtained t statistics are compared with the critical value of the Bonferroni t statistic (Dunn, 1961) associated with C (the number of planned contrasts) and N-3 degrees of freedom.

Example

Data:

X_1	X_2	X_3	Y
4	4	8	1
5	8	6	1
7	9	7	1
6	7	7	1
6	8	8	1
8	7	7	1
6	2	4	0
5	3	5	0
5	2	6	0
5	1	5	0
4	3	6	0
5	2	7	0

Method A

Stage I

The sample correlations are:

$$r_{x_1y} = .45$$

$$r_{x_2y} = .90$$

$$r_{x_3y} = .71$$

The overall test on the differences among these coefficients yields an obtained F value of 8.31. The critical value is $F_{.05,2,8} = 4.46$. Since $H_0: \rho_{x_1y}=\rho_{x_2y}=\rho_{x_3y}$ is rejected we are entitled to proceed to Stage II.

Stage II

The t values associated with the three pairwise contrasts are as follows

Population Contrasts	Sample Coefficients	t_{obt}
ρ_{x_1y} vs ρ_{x_2y}	.45 vs .90	3.37
ρ_{x_1y} vs ρ_{x_3y}	.45 vs .71	.97
ρ_{x_2y} vs ρ_{x_3}	.90 vs .71	1.70

Since the critical value is 2.262, we conclude that $\rho_{x_1y} \neq \rho_{x_2y}$; the null hypothesis is retained for the other contrasts.

Method C

The critical values of the Bonferroni t statistic for one, two and three planned contrasts are shown below along with the contrasts and obtained t values.

Population Contrasts	Sample Coefficients	t_{obt}	C=1 $t_B=t$	C=2 t_B	C=3 t_B
ρ_{x_1y} vs ρ_{x_2y}	.45 vs .90	3.37	2.262	2.69	2.93
ρ_{x_1y} vs ρ_{x_3y}	.45 vs .71	.97	2.262	2.69	2.93
ρ_{x_2y} vs ρ_{x_3y}	.90 vs .71	1.70	2.262	2.69	2.93

If all three pairwise contrasts are planned, the critical value is 2.93 and, as with Methods A and B, the $H_0: \rho_{x_1y} = \rho_{x_2y}$ is rejected while

$H_0: \rho_{x_1y} = \rho_{x_3y}$ and $H_0: \rho_{x_2y} = \rho_{x_3y}$ are retained.

When only two contrasts are planned, the critical value is 2.69.

Discussion

Methods A, B and C may be viewed as analogs to the Protected LSD, Dunnett and Dunn-Bonferroni procedures for multiple comparisons among means. Recent work by Carmer and Swanson (1973) and Bernhardson (1975) on the protected LSD procedure suggests that this two stage approach has two very desirable characteristics. The first is good control of Type I error and the second is high power. These characteristics should also be associated with Method A. Monte Carlo work is needed, however, to support this supposition.

Method B, which is limited to m-1 control versus treatment contrasts, is simpler to compute than Method A.

Method C requires the researcher to plan the contrasts of interest. If the number of planned contrasts relative to the total number of pairwise contrasts is not small, Method A is more powerful. On the other hand, if a very large number of correlations are involved along with relatively few error d.f. Method C may be much more powerful if only a few contrasts are planned.

It should be pointed out that the error rate for these procedures is defined somewhat differently in each case. Under Method A the error rate is experimentwise where the experiment is the whole collection of pairwise contrasts. Under Method B the error rate is also experimentwise but the experiment is defined as the collection of m-1 contrasts. The error rate under Method C is the Per Experiment error rate which is defined as the number of comparisons falsely declared significant over the total number of experiments. These are theoretical error rates and, unfortunately, it has not been demonstrated empirically that the rates for these tests are the same as the nominal rates. A possible fly in the ointment has just been published by Neill and Dunn (1975). They found that the two predictor Hotelling test has unacceptably high empirical error rates. For $\alpha = .05$ and N=10 and 50, the empirical error rates were .18 and .19. For $\alpha = .01$ and N=10 and 50, the average empirical error rates were .13 and .14 (45 cases in each study). In contrast, William's modification of Hotelling's test (which Dunn claims is essentially the same test Hotelling discarded before 1940 when he published his well-known test) maintains the empirical error rate very close to the nominal level. Neill and Dunn do not point out, however, that Hotelling's test was derived under the assumption of fixed values on the predictor variables whereas their simulation was carried out with trivariate normal distributions. Whether or not this difference is sufficient to explain the unusually high error rates for Hotelling's test remains to be seen.

The application suggested here (i.e., X variables random and Y fixed) meets neither the assumptions of fixed predictor values nor multivariate normality. Until this issue is investigated a conservative approach is to substitute William's modification of Hotelling's test where Hotelling's conventional test has been suggested in Methods A, B and C.

Under the more typical situation encountered in typical correlation studies the multivariate normal assumption is probably reasonable and William's modification should be substituted for Hotelling's conventional test in Stage II of Method A and in Methods B and C.

In general, Method A is recommended for any number of pairwise comparisons unless computation is a problem. Method C is suggested for the situation in which Stage I of Method A presents computation problems or if comparisons are planned.

References

Bernhardson, C. S. Type I error rates when multiple comparison procedures follow a significant F test of ANOVA. Biometrics, 1975, 31, 229-232.

Campbell, D. T. Temporal changes in treatment-effect correlations: A quasi-experimental model for institutional records and longitudinal studies. In G. V. Glass (Ed.); Proceedings of the 1970 Invitational Conference on Testing Problems. Princeton, N.J.: Educational Testing Service, 1971.

Carmer, S. G., and Swanson, M. A. An evaluation of ten pairwise multiple comparisons procedures by Monte Carlo. Journal of the American Statistical Association, 1973, 68, 66-74.

Dunn, O. J. Multiple comparisons among means. Journal of the American Statistical Association, 1961, 56, 52-64.

Dunnett, C. W. A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Association, 1955, 50, 1096-1121.

Hotelling, H. The selection of variates for use in prediction with some comments on the general problems of nuisance parameters. Annals of Mathematical Statistics. 1940, 11, 271-283.

Kenny, D. A. A quasi-experimental approach to assessing treatment effects in the nonequivalent control group design. Psychological Bulletin. 1975, 82, 345-362.

McNemar, Psychological Statistics. (4th ed.) New York: Wiley, 1968.

Neill, J. J., and Dunn, O. J. Equality of dependent correlation coefficients. Biometrics, 1975, 31, 531-543.

Walker, H. M., and Lev, J. Statistical Inference. New York: Holt, 1953.