A PERMUTATION TEST FOR MULTIVARIATE MATCHED-PAIR ANALYSES: COMPARISONS WITH HOTELLING'S MULTIVARIATE MATCHED-PAIR T^2 TEST [1]

PAUL W. MIELKE, JR., KENNETH J. BERRY, AND CHARLES O. NEIDT

Colorado State University

---

[1] Request reprints from Paul W. Mielke, Jr., Department of Statistics, Colorado State University, Fort Collins, CO 80523-1877. E-mail address: mielke@lamar.colostate.edu

Summary.—A permutation alternative for Hotelling's multivariate matched-pair T^2 test is introduced. The permutation test allows for analyses when the number of subjects is less than or equal to the number of measurements, which is not possible with Hotelling's multivariate matched-pair T^2 test. For the data analyzed the permutation test is shown to provide improved discrimination over Hotelling's multivariate matched-pair T^2 test.

It is often important in psychological research to compare the scores of matched subjects at two time periods such as pre-treatment and post-treatment. When a single measurement is obtained on each subject at each of the two time periods the usual test of difference employed is Student's matched-pair t test. When multiple measurements are obtained on each subject at each of the two time periods the usual test of difference employed is Hotelling's multivariate matched-pair T^2 test (Hotelling, 1931), which is simply a generalization of Student's matched-pair t test. Because Hotelling's multivariate matched-pair T^2 test (hereafter, the T^2 test) obtains a vector of measurements on each subject at each time period, it lends itself to two analyses, as follows. Consider n subjects and c judges. It is possible to block on the n subjects and examine the multivariate difference between the c judges at the two time periods. Alternatively, it is possible to block on the c judges and examine the multivariate difference between the n subjects at the two time periods.

For the first analysis the T^2 test statistic is distributed under the null hypothesis as an F distribution with c and $n - c$ degrees-of-freedom (*df*) in the numerator and denominator, respectively. Under the second analysis the T^2 test statistic is distributed as an F distribution with n and $c - n$ *df* in the numerator and denominator, respectively. Consequently, one of the two analyses will yield a *df* in the denominator which is less than or equal to zero. Moreover, when $n = c$ neither analysis is possible. If only one of the two analyses is desired and the *df* in the denominator is less than or equal to zero for this analysis, then the T^2 test cannot be evaluated. In this paper a permutation test is developed which eliminates this problem and for the data analyzed is shown to be more discriminating than the T^2 test. This permutation test is a multivariate extension of univariate permutation tests for matched pairs (Mielke & Berry, 1982).

PERMUTATION TEST

Let n subjects be associated with a multivariate pre-treatment and post-treatment matched-pair permutation test. Let $(x_{11r},\ldots,x_{c1r})$ and $(x_{12r},\ldots,x_{c2r})$ denote c-dimensional row vectors with elements comprised of the c measurements on the r^{th} subject from the pre-treatment and post-treatment, respectively, where $r = 1,\ldots,n$. Let

$$\mathbf{d}_{1r} = \begin{pmatrix} d_{11r} \\ \vdots \\ d_{c1r} \end{pmatrix} \qquad [1]$$

where $d_{h1r} = x_{h1r} - x_{h2r}$ for $h = 1,\ldots,c$, be the c-dimensional column vector of differences between pre-treatment and post-treatment measurements for the r^{th} subject, and let $\mathbf{d}_{2r} = -\mathbf{d}_{1r}$ be the c-dimensional origin reflection of $\mathbf{d}_{1r}$ $(r = 1,\ldots,n)$. The probability (P), under the null hypothesis (H_0) of the matched-pair experiment, is $P(\mathbf{d}_{1r}) = P(\mathbf{d}_{2r}) = 0.5$, for $r = 1,\ldots,n$. Consider the test statistic given by

$$\delta = \binom{n}{2}^{-1} \sum_{r<s} \Delta(\mathbf{d}_{1r}, \mathbf{d}_{1s}) \qquad [2]$$

where

$$\Delta(\mathbf{d}_{1r}, \mathbf{d}_{1s}) = \left[(\mathbf{d}_{1r} - \mathbf{d}_{1s})^T (\mathbf{d}_{1r} - \mathbf{d}_{1s})\right]^{1/2} \qquad [3]$$

is the c-dimensional Euclidean distance between the r^{th} and s^{th} subjects' differences, and the sum $\Sigma_{r<s}$ is over all r and s such that $1 \le r < s \le n$. This approach includes the multivariate one-sample permutation test in which $d_{h1r} = x_{hr} - \mu_h$ for $h = 1,\ldots,c$, where $(x_{1r},\ldots,x_{cr})$ denotes the c-dimensional row vector with elements comprised of the c measurements for the r^{th} subject $(r = 1,\ldots,n)$, and $(\mu_1,\ldots,\mu_c)$ is the c-dimensional row vector of the c hypothesized central values under H_0.

If the c measurements are in different units, then the measurements must be made commensurate (i.e., standardized to a common unit of measurement). The replacement of d_{hir} with $d^*_{hir} = d_{hir}/\Phi_h$, where

$$\Phi_h = \sum_{r<s} \left| |d_{h1r}| - |d_{h1s}| \right|, \qquad [4]$$

for $h = 1,\ldots,c$, ensures that each measurement makes a similar contribution in the c-dimensional Euclidean space since

$$\sum_{r<s} \left| |d^*_{h1r}| - |d^*_{h1s}| \right| = 1, \qquad [5]$$

for $h = 1,\ldots,c$. This commensuration is invariant relative to any permutation under H_0 and is termed Euclidean commensuration (Berry & Mielke, 1992).

Hotelling's multivariate matched-pair T^2 test statistic (Anderson, 1958, pp. 101-108) is given by

$$T^2 = n\bar{\mathbf{d}}_1^{\mathrm{T}} \mathbf{S}_d^{-1} \bar{\mathbf{d}}_1 \qquad [6]$$

where $\mathbf{S}_d$ is a $c \times c$ matrix given by

$$\mathbf{S}_d = \frac{1}{n-1} \sum_{r=1}^{n} (\mathbf{d}_{1r} - \bar{\mathbf{d}}_1)(\mathbf{d}_{1r} - \bar{\mathbf{d}}_1)^{\mathrm{T}}, \qquad [7]$$

and

$$\bar{\mathbf{d}}_1 = \frac{1}{n} \sum_{r=1}^{n} \mathbf{d}_{1r}. \qquad [8]$$

If the observed value of δ is δ_o, then the exact P-value is given by $P(\delta \le \delta_o | H_0)$, i.e., the proportion of the 2^n possible δ values which are less than or equal to δ_o. If the observed value of T^2 is T_o^2, then the analogous exact P-value is given by $P\left(T^2 \ge T_o^2 | H_0\right)$, i.e., the proportion of the 2^n possible T^2 values which are greater than or equal to T_o^2. When n is large, e.g., $n > 20$ since $2^{20} = 1,048,576$, then a method to approximate the P-value is essential. One such method involves calculating the first three exact cumulants of δ under H_0, equating the obtained exact cumulants of δ to the corresponding three cumulants which characterize the Pearson type III distribution, and obtaining an approximate P-value by numerically integrating the resulting Pearson type III distribution (Mielke, 1984, 1991). The first three exact cumulants of δ under H_0 are given by

$$\kappa_1(\delta) = \mu_\delta = [2n(n-1)]^{-1} \sum_{r<s} \sum_{i=1}^{2} \sum_{j=1}^{2} \Delta(\mathbf{d}_{ir}, \mathbf{d}_{js}), \qquad [9]$$

$$\kappa_2(\delta) = \sigma_\delta^2 = [n(n-1)]^{-2} \sum_{r<s} \sum_{i=1}^{2} \sum_{j=1}^{2} [D(i,r;j,s)]^2, \qquad [10]$$

and

$$\kappa_3(\delta) = \sigma_\delta^3 \gamma_\delta = 6[n(n-1)]^{-3}$$
$$\times \sum_{r<s<t} \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{2} D(i,r;j,s)\, D(i,r;k,t)\, D(j,s;k,t) \qquad [11]$$

where

$$D(i,r;j,s) = \Delta(\mathbf{d}_{ir}, \mathbf{d}_{js}) - \frac{1}{2} \sum_{i=1}^{2} \Delta(\mathbf{d}_{ir}, \mathbf{d}_{js})$$
$$- \frac{1}{2} \sum_{j=1}^{2} \Delta(\mathbf{d}_{ir}, \mathbf{d}_{js}) + \frac{1}{4} \sum_{i=1}^{2} \sum_{j=1}^{2} \Delta(\mathbf{d}_{ir}, \mathbf{d}_{js}), \qquad [12]$$

and μ_δ, σ_δ^2, and γ_δ are the exact mean, variance, and skewness of δ under H_0. While the multivariate permutation test is applicable to any combination of $c \geq 1$ and n, although $n \geq 10$ might be desired simply to ensure at least $2^{10} = 1{,}024$ permutations, any application of the T^2 test under its null hypothesis and the assumption of multivariate normality requires that $min(c, n-c) \geq 1$ since the distribution of the modified statistic given by

$$\frac{(n-c)\,T^2}{(n-1)\,c} \qquad [13]$$

is F with c and $n - c$ df in the numerator and denominator, respectively.

EXAMPLE

The data are from a study in which students majoring in mathematics education were enrolled in a course on teaching discrete mathematics to secondary school students. One of the objectives of the course was to enhance the skills of these prospective teachers in writing in the discipline of mathematics. On the first day of the course the students were given a mathematics writing exercise. The same exercise was repeated on the last day of the semester-long course. A total of 11 students completed both writing exercises. The paired, but randomly arranged, pre-training and post-training writing samples of the 11 students were presented blindly to 13 experienced teachers of mathematics and language arts for grading. Each of the

---

Insert Tables 1 and 2 about here

---

13 judges scored each of the 22 writing samples on a scale from 0 to 10. The pre-training grades are given in Table 1 and the post-training grades are given in Table 2. The analyses for two cases are considered.

Case 1

The first analysis blocks on the $n = 11$ students and compares the pre-training and post-training grades of the $c = 13$ judges. This analysis evaluates the following question: Are the judges consistent in their grading or are there significant pre-training/post-training differences in grading among the judges?

The multivariate permutation test based on Euclidean commensuration yields an exact P-value of 190/2048 = 0.09277 and a Pearson type III approximate P-value of 0.07448. Because the $\mathbf{S}_d$ matrix is singular in this context, the T^2 test is not defined. The *df* in the denominator of the F distribution associated with Expression 13 is negative, i.e., $n - c = 11 - 13 = -2$. Obviously, this is a serious limitation of the T^2 test.

Case 2

The second analysis blocks on the $n = 13$ judges and compares the pre-training and post-training grades of the $c = 11$ students. This analysis evaluates the following question: Did the coursework result in significant pre-training/post-training differences in writing among the students? The multivariate permutation test based on Euclidean commensuration yields an exact P-value of

2/8192 = 0.00024 and the Pearson type III approximate P-value is 0.00011. The permutation version of the T^2 test under H_0 that $P(\mathbf{d}_{1r}) = P(\mathbf{d}_{2r}) = 0.5$ yields an exact P-value of 634/8192 = 0.07739 and the T^2 test yields a P-value of 0.08193 under the assumption of multivariate normality.

DISCUSSION

The results of the analysis of the data for Case 1 indicate that the T^2 test is undefined for $n \geq c$. However, the multivariate permutation test employing Euclidean commensuration is unaffected by this constraint. The results of the analysis of the data for Case 2 indicate that the T^2 test yields results which differ substantially from the results of the multivariate permutation test based on Euclidean commensuration when $2 \leq c < n$.

While the analyses of the present example indicate that substantial advantages exist for the multivariate permutation test based on Euclidean commensuration, other data sets will indicate advantages for the T^2 test. Not only is the multivariate permutation test applicable to situations where the T^2 test cannot be used, it may be more discriminating than the classical Hotelling's multivariate matched-pair T^2 test.

REFERENCES

Anderson, T. W. (1958) *An introduction to multivariate statistical analysis*. New York: John Wiley & Sons.

Berry, K. J., & Mielke, P. W. (1992) A family of multivariate measures of association for nominal independent variables. *Educational and Psychological Measurement*, 52, 41-55.

Hotelling, H. (1931) The generalization of Student's ratio. *Annals of Mathematical Statistics*, 2, 360-378.

Mielke, P. W. (1984) Meteorological applications of permutation techniques based on distance functions. In P. R. Krishnaiah & P. K. Sen (Eds.), *Handbook of Statistics, Vol. 4: Nonparametric methods*. Amsterdam: Elsevier North-Holland.

Mielke, P. W. (1991) The application of multivariate permutation methods based on distance functions in the earth sciences. *Earth-Science Reviews*, 31, 55-71.

Mielke, P. W., & Berry, K. J. (1982) An extended class of permutation techniques for matched pairs. *Communications in Statistics-Theory and Methods*, 11, 1197-1207.

TABLE 1

PRE-TRAINING SCORES ASSIGNED BY 13 JUDGES TO WRITING SAMPLES OF 11 STUDENTS

Judge	Student										
	1	2	3	4	5	6	7	8	9	10	11
1	1	6	1	1	8	1	5	8	6	3	1
2	3	4	6	2	8	3	6	9	9	7	3
3	1	6	2	3	7	3	3	5	5	2	4
4	2	5	5	1	8	2	4	7	8	6	4
5	3	6	5	2	8	2	3	5	9	4	2
6	0	4	3	0	8	0	3	9	7	3	0
7	1	5	0	1	7	0	1	2	8	5	1
8	5	8	4	0	2	0	2	10	2	2	0
9	1	7	2	5	9	2	6	6	9	6	3
10	2	3	2	0	6	1	5	7	5	3	3
11	1	5	2	1	7	1	2	8	8	7	4
12	0	4	1	0	9	0	2	5	3	2	1
13	4	9	2	2	5	3	3	9	8	4	3

TABLE 2

Post-Training Scores Assigned by 13 Judges to Writing Samples of 11 Students

Judge	Student										
	1	2	3	4	5	6	7	8	9	10	11
1	9	5	3	1	8	1	7	6	6	4	5
2	8	5	5	2	9	2	6	6	7	5	5
3	5	6	2	3	3	3	6	8	7	5	8
4	7	6	3	2	9	4	5	6	6	4	7
5	8	7	4	2	8	4	7	8	6	3	5
6	6	7	2	0	6	0	5	7	6	5	4
7	5	5	2	1	5	3	5	5	4	0	7
8	4	9	6	0	3	3	10	8	5	3	5
9	9	5	5	7	8	3	8	8	8	7	8
10	4	4	1	0	4	3	4	5	6	6	6
11	6	3	3	2	9	2	9	7	7	5	9
12	6	2	3	1	5	1	6	9	6	5	6
13	9	6	4	4	7	6	9	7	6	7	5