

MultiResponse Constrained Permutation Procedure

Andrea Nibbe and Curtis Storlie
F. A. Graybill Statistical Laboratory
Colorado State University

I. Formal Presentation of Problem

The hypotheses we are interested in testing are of the following two forms:

Form 1: Subtest scores in Group I will have higher positive intercorrelations with each other than with the other subtest scores (Group II).

Form 2: Subtest scores in Group I will positively correlate more highly with subtest scores in Group II than with the other subtest scores (Group III).

There is no standard approach to analyze hypotheses of this type. A variation of Mielke and Berry's MultiResponse Blocked Permutation Procedure (Mielke & Berry, 2001) was developed to handle this case.

To define these hypotheses more concisely, the following notation will hold throughout the paper.

- a. The correlation between subtest i and subtest j will be written ρ_{ij} , the sample correlation will be written r_{ij} . Likewise, the squared distance between subtest i and subtest j will be written d_{ij} .
- b. Intercorrelation is defined as the average of the two-way correlations. Therefore, if ρ_{ijk} is the intercorrelation between subtests T_i, T_j, T_k , then $\rho_{ijk} = \frac{1}{3}(\rho_{ij} + \rho_{jk} + \rho_{ik})$.
- c. In the hypotheses of Form 1 analyzed, there are always three subtests in Group I. The subsequent explanations will use this information for ease in notation, but can easily be generalized.
- d. The number of subtests included in a given hypothesis will be denoted by P .

II. Overview of Permutation Testing

In order to motivate our method, a basic understanding of permutation procedures is necessary. The next two sections give a very brief introduction. For a more detailed description, see Mielke & Berry (2001).

Suppose we wish to test the hypotheses

$$H_0: \mu_1 = \mu_2$$

vs.

$$H_a: \mu_1 < \mu_2,$$

where we observed a response from two known groups. The observed data are of the form shown in Figure 1.

If normal theory applied, a straightforward t-test could be computed to assess whether the sample mean from Group 1 is significantly lower than the sample mean of Group 2. If the assumption of normality is not valid, however, a permutation test may be more appropriate. Since normality is not assumed in a permutation test, the null hypothesis is that both populations have the same distribution.

Thus the hypotheses being tested are now

$$\begin{aligned} H_0': F_1(y) &= F_2(y) \\ \text{vs.} \\ H_a': F_1(y) &\neq F_2(y), \end{aligned}$$

where $F_i(y)$ is the distribution of y for group i . Note that the original null hypothesis is a subset of this one. It will be useful later to recognize that the above hypotheses are equivalent to testing

$$\begin{aligned} H_0': F_{\underline{y}}(\underline{y}) &= F_{\underline{y}^*}(\underline{y}) \\ \text{vs.} \\ H_a': F_{\underline{y}}(\underline{y}) &\neq F_{\underline{y}^*}(\underline{y}), \end{aligned}$$

where $F_{\underline{y}}$ is the joint distribution of the observation vector \underline{y} and \underline{y}^* is any permutation of the y_{ij} .

Figure 1.

SUBJECT	GROUP	Y
1	1	y_{11}
2	1	y_{12}
3	1	y_{13}
⋮	⋮	⋮
n_1	1	y_{1n_1}
n_1+1	2	y_{21}
n_1+2	2	y_{22}
n_1+3	2	y_{23}
⋮	⋮	⋮
n_1+n_2	2	y_{2n_2}

The test statistic is

$$\begin{aligned} U &= \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} - \frac{1}{n_2} \sum_{k=1}^{n_2} y_{1k} \\ &= \bar{y}_1 - \bar{y}_2. \end{aligned}$$

The main idea of a permutation test is that all observations are "exchangeable" under the null hypothesis, since the distribution of Group 1 is the same as the distribution of Group 2. Thus it shouldn't matter which observations came from a given group if the null hypothesis is true. We want to determine whether the data we observed has a difference in means that is extreme with respect to all permutations of the data. This would indicate a deviation from the null hypothesis, and we would reject.

A permutation simply implies a reordering of the observed responses into the two groups. One possible permutation of the above data is shown in Figure 2. Subject n_1+3 was originally a member of Group 2, and is part of Group 1. Note that Group 1 will always have n_1 observations, and Group 2 will always have n_2 observations. The number of permutations resulting in distinct values of U is $\frac{(n_1+n_2)!}{n_1!n_2!}$.

Define $U_0 = \bar{y}_1 - \bar{y}_2$ *without permutation*. The proportion of permutations with test statistic values as extreme or more extreme than U_0 is the exact p-value (Mielke and Berry, 2001). Using this testing procedure, we will make a Type I error $(100 \times \alpha)\%$ of the time regardless of the distribution of the data.

It is important to note that by rejecting H_0 ' we are really only concluding that $F_{\underline{y}}(\underline{y}) \neq F_{\underline{y}^*}(\underline{y})$ for some \underline{y}^* . This does not necessarily mean that $\mu_1 < \mu_2$. However, keep in mind that we rejected H_0 ' because $\bar{y}_1 - \bar{y}_2$ was large negative. This fact makes it likely that $\mu_1 < \mu_2$ is the reason for $F_{\underline{y}}(\underline{y}) \neq F_{\underline{y}^*}(\underline{y})$.

Figure 2.

SUBJECT	GROUP	Y
1	1	y_{11}
n_1+3	1	y_{23}
3	1	y_{13}
.	.	.
n_1	1	y_{1n_1}
n_1+1	2	y_{21}
n_1+2	2	y_{22}
2	2	y_{12}
.	.	.
n_1+n_2	2	y_{2n_2}

III. MultiResponse Blocked Permutation Procedure

Suppose now that each subject gives multiple response measures. In our case, these responses are P subtest scores collected on each subject. This is illustrated in Figure 3, where N is the number of subjects, and T_{in} the n^{th} subject's score for the i^{th} subtest.

To test whether the first three subtest scores are higher in magnitude than the remaining scores, it makes sense to test the hypothesis $H_0: \frac{1}{3}(\mu_{T_1} + \mu_{T_2} + \mu_{T_3}) = \frac{1}{P-3}(\mu_{T_4} + \dots + \mu_{T_P})$ where μ_{T_i} is the mean of subtest I against the greater than alternative. Again, since we are formulating a permutation procedure, the hypothesis we would actually test is

$$H_0': F_{\underline{T}}(t) = F_{\underline{T}^*}(t)$$

vs.

$$H_a': F_{\underline{T}}(t) \neq F_{\underline{T}^*}(t),$$

for any permutation \underline{T}^* of the test scores within a subject. H_0' does not mean the subtest scores are independent, but does imply they have the same marginal distributions and equal joint moments including correlations. It will be important to note this implies that there is no subject by subtest interaction.

A test statistic that could used to test H_0' is defined as

$$V = \frac{1}{3}(\bar{T}_1 + \bar{T}_2 + \bar{T}_3) - \frac{1}{(P-3)} \sum_{i=4}^P \bar{T}_i,$$

where \bar{T}_i is the sample mean of subtest i scores across subjects.

When $H_0: \frac{1}{3}(\mu_{T_1} + \mu_{T_2} + \mu_{T_3}) = \frac{1}{P-3}(\mu_{T_4} + \dots + \mu_{T_P})$, is false, V will be large in absolute value. Again, we can really only conclude that $F_{\underline{T}}(t) \neq F_{\underline{T}^*}(t)$ for some \underline{T}^* . Since we would reject only when the test statistic

$\frac{1}{3}(\bar{T}_1 + \bar{T}_2 + \bar{T}_3) - \frac{1}{(P-3)} \sum_{i=4}^P \bar{T}_i$ is large negative, however, it is probable that a rejection will be due to our hypothesis of interest, H_0 .

Figure 3.

SUBJECT	Group I			Group II			
	T ₁	T ₂	T ₃	T ₄	T ₅	...	T _P
1	T ₁₁	T ₂₁	T ₃₁	T ₄₁	T ₅₁	...	T _{P1}
2	T ₁₂	T ₂₂	T ₃₂	T ₄₂	T ₅₂	...	T _{P2}
3	T ₁₃	T ₂₃	T ₃₃	T ₄₃	T ₅₃	...	T _{P3}
.
.
.
N	T _{1N}	T _{2N}	T _{3N}	T _{4N}	T _{5N}	...	T _{PN}

The permutation procedure is executed in the same manner as before, except that there are now N blocks (subjects) to permute independently. The number of possible permutations per subject that result in a unique V_n is $\frac{P!}{(P-3)!3!}$, so the number of total permutations is $\left(\frac{P!}{(P-3)!3!}\right)^N$. For a given permutation, the test statistic is computed and a p-value is computed as the proportion of test statistics more extreme than the one observed.

IV. MultiResponse Constrained Permutation Procedure

Form 1. As you'll recall, we are interested in testing whether the three subtest scores in Group I will have higher positive intercorrelations with each other than with the other P-3 subtest scores (Group II). We translated this into the following statistical hypotheses:

$$H_0: \frac{1}{3}(\rho_{12} + \rho_{23} + \rho_{13}) = \frac{1}{24} \sum_{i=1}^3 \sum_{j=4}^P \rho_{ij}$$

vs.

$$H_a: \frac{1}{3}(\rho_{12} + \rho_{23} + \rho_{13}) > \frac{1}{24} \sum_{i=1}^3 \sum_{j=4}^P \rho_{ij}$$

However, as in the first permutation test discussed, our formal hypothesis needs to be something of the form,

$$H_0': F_{\underline{T}}(\underline{t}) = F_{\underline{T}^*}(\underline{t})$$

vs.

$$H_a': F_{\underline{T}}(\underline{t}) \neq F_{\underline{T}^*}(\underline{t}),$$

for any permutation \underline{T}^* of the test scores within a subject. Again, H_0' does not imply that the subtest scores are independent. The assumptions of identical distribution, equal correlations (and other higher order moments) and hence no subtest by subject interaction *are* implied, however.

The test statistic we will use to test H_0' is

$$W_1 = \frac{1}{3}(r_{12} + r_{23} + r_{13}) - \frac{1}{24} \sum_{i=1}^3 \sum_{j=4}^P r_{ij}$$

where r_{ij} is the sample correlation coefficient between subtests i and j.

There can be a problem with using all possible permutations of the subtests in this setup. There is nothing wrong with the level of significance. It is still a size α test, but consider the possibility that subtests 4, 5, and 6 are highly correlated with each other and not with other tests. Under all possible permutations subtest 4, 5, and 6 will be exchangeable with subtests 1, 2, and 3. These types of

permutation could lead to a larger value of W_1 than the observed value and might hinder a rejection of H_0' in a situation that we should have rejected.

Since this type of situation was assumed by the researcher to be present in the problem of interest, we decided to make the following change to H_0' . It can be stated as

$$H_0': F_{\underline{T}}(t) = F_{\underline{T}^*}(t)$$

vs.

$$H_a': F_{\underline{T}}(t) \neq F_{\underline{T}^*}(t),$$

for any permutation \underline{T}^* that leaves two of the subtests 1, 2, and 3 in Group I. In other words, H_0' states that exchanging any one of subtests 1, 2, or 3, for any one of the other eight tests results in the same joint distribution. This again implies we have equal correlations and no interaction between subtests and subjects. There are $[3*(P-3)+1]^N$ permutations of this type that result in unique test statistic values.

Form 2. Here we are interested in testing whether subtest scores in Group I are more highly correlated with subtest scores in Group II than with those in Group III. We translated this into the following statistical hypotheses:

$$H_0: \frac{1}{P_1 P_2} \sum_{i=1}^{P_1} \sum_{j=p_1+1}^{P_1+P_2} \rho_{ij} = \frac{1}{P_1 P_3} \sum_{i=1}^{P_1} \sum_{j=p_1+1}^P \rho_{ik}$$

vs.

$$H_a: \frac{1}{P_1 P_2} \sum_{i=1}^{P_1} \sum_{j=p_1+1}^{P_1+P_2} \rho_{ij} > \frac{1}{P_1 P_3} \sum_{i=1}^{P_1} \sum_{j=p_1+1}^P \rho_{ik}$$

Our formal hypothesis in this case is

$$H_0': F_{\underline{T}}(t) = F_{\underline{T}^*}(t)$$

vs.

$$H_a': F_{\underline{T}}(t) \neq F_{\underline{T}^*}(t),$$

for any permutation \underline{T}^* that leaves the first p_1 subtests in Group I where they are. H_0' assumes all subtests in Group II and III are identically distributed, not necessarily independent but with equal correlations amongst one another. It also implies that any given subtest in Group I will have the same correlation with all subtests in Group II and III. This also prohibits most types of subject by subtest interaction.

The test statistic we used is

$$W_2 = \frac{1}{P_1 P_2} \sum_{i=1}^{P_1} \sum_{j=p_1+1}^{P_1+P_2} r_{ij} - \frac{1}{P_1 P_3} \sum_{i=1}^{P_1} \sum_{j=p_1+1}^P r_{ik}.$$

There are $\left[\binom{P-p_1}{p_2} \right]^N$ permutations that result in unique test statistic values. The p-value is calculated as usual.

V. Notes on Execution of Procedure

- 1) For ease of calculation, a measure equivalent to the sample correlation was used. It saved computation time to use the average squared distance between the standardized data in place of sample correlation, as the former can be calculated on each observation, while the sample correlation is only defined for multiple observations. Also in section VI we mention a bootstrap procedure. Having these distances on each subject allowed us to perform a bootstrap t which is more robust than

the usual bootstrap. The proof of the equivalence of the sample correlation and the average of squared distances between standardized data can be found in Section VII.

- 2) The total number of permutations is large. For example, the Form 1 hypotheses had $p_1=3$, $p_2=8$, and $N=18$. The total number of permutations that resulted in a unique test statistic was therefore $[3 \times 8 + 1]^{18} = 25^{18} = 1.455 \times 10^{25}$. Thus to actually implement these MRCP procedures, a Monte Carlo simulation was used. Four runs using different random seeds, each with 1,000,000 permutations, resulted in nearly identical p-values. The average of these four p-values was reported.

VI. Validity

Recall that the MRCP is assuming no interaction between subject and groups. A bootstrap procedure was also computed in case this assumption was not met. For all hypotheses tested on the researcher's data, the results were in agreement with those of the permutation test. Thus we can be fairly certain that the effect of interaction between subject and subtest on correlation structure, if present, was not large enough to have an effect on the analysis.

VII. Proof of Equivalence of Sample Correlation and Squared Distance of Standardized Values

The sample correlation is defined as

$$r_{xy} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}, \text{ where } s_x = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}.$$

Let $w_i = \frac{x_i - \bar{x}}{s_x}$ and $z_i = \frac{y_i - \bar{y}}{s_y}$ denote the standardized values.

Note that $r_{xy} = \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \bar{x})}{s_x} \frac{(y_i - \bar{y})}{s_y} = \frac{1}{n} \sum_{i=1}^n w_i z_i$

Let $d_{w_i z_i} = (w_i - z_i)^2$. Then using the fact that $\bar{w} = 0$, $\bar{z} = 0$, $s_w = 1$, and $s_z = 1$, we have

$$\begin{aligned} \bar{d}_{wz} &= \frac{1}{n} \sum_{i=1}^n d_{w_i z_i} \\ &= \frac{1}{n} \sum_{i=1}^n (w_i - z_i)^2 \\ &= \frac{1}{n} \sum_{i=1}^n w_i^2 - \frac{2}{n} \sum_{i=1}^n w_i z_i + \frac{1}{n} \sum_{i=1}^n z_i^2 \\ &= 1 - \frac{2}{n} \sum_{i=1}^n w_i z_i + 1 \\ &= 2 - \frac{2}{n} \sum_{i=1}^n w_i z_i \\ &= 2 - 2r_{xy} \end{aligned}$$

Since a one-to-one function exists between r_{xy} and \bar{d}_{wz} , a test statistic using \bar{d}_{wz} in place of r_{xy} in a permutation test will give equivalent p-values.

References:

Mielke and Berry. Permutation Methods. 2001.