Web-Based Supplementary Materials for "On Distance-Based Permutation Tests for Between-Group Comparisons"

Philip T. Reiss, M. Henry H. Stevens, Zarrar Shehzad,
Eva Petkova, and Michael P. Milham

Web Appendix A: Matrix Expressions for the Pseudo-F Statistic

As in McArdle and Anderson (2001), we begin by taking as given an $n \times p$ matrix $\boldsymbol{Y}$ of p-dimensional outcomes, whose columns have mean zero. In the slightly modified setup of our paper, we are also given projection matrices $\boldsymbol{H}_1$, $\boldsymbol{H}_2$ and $\boldsymbol{H}$ determining full and reduced models. The univariate F statistic for the ith component is

$$\frac{(\|\hat{\boldsymbol{y}}_i\|^2 - \|\hat{\boldsymbol{y}}_{i.1}\|^2)/m_2}{\|\hat{\boldsymbol{\varepsilon}}_i\|^2/(n-m)}, \tag{A.1}$$

where $\hat{\boldsymbol{y}}_i$ and $\hat{\boldsymbol{y}}_{i.1}$ are the vectors of fitted values under the full and reduced model for the ith component, respectively, and $\hat{\boldsymbol{\varepsilon}}_i$ denotes the residuals from the full model.

The pseudo-F statistic is then the sum, over i, of the numerators in (A.1), divided by the sum of the denominators, which equals

$$\frac{\left[\mathrm{tr}(\hat{\boldsymbol{Y}}^T\hat{\boldsymbol{Y}}) - \mathrm{tr}(\hat{\boldsymbol{Y}}_1^T\hat{\boldsymbol{Y}}_1)\right]/m_2}{\mathrm{tr}(\hat{\boldsymbol{R}}^T\hat{\boldsymbol{R}})/(n-m)} = \frac{\left[\mathrm{tr}(\hat{\boldsymbol{Y}}\hat{\boldsymbol{Y}}^T) - \mathrm{tr}(\hat{\boldsymbol{Y}}_1\hat{\boldsymbol{Y}}_1^T)\right]/m_2}{\mathrm{tr}(\hat{\boldsymbol{R}}\hat{\boldsymbol{R}}^T)/(n-m)}, \tag{A.2}$$

where

$$\hat{\boldsymbol{Y}} = \boldsymbol{H}\boldsymbol{Y}, \qquad \hat{\boldsymbol{Y}}_1 = \boldsymbol{H}_1\boldsymbol{Y}, \qquad \hat{\boldsymbol{R}} = (\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y}, \tag{A.3}$$

and where the equality in (A.2) follows from the identity $\mathrm{tr}(\boldsymbol{BC}) = \mathrm{tr}(\boldsymbol{CB})$. Substituting (A.3) into (A.2), and using the just-cited identity and the idempotence of $\boldsymbol{H}$, $\boldsymbol{H}_1$, and $\boldsymbol{I}-\boldsymbol{H}$, yields

$$\frac{\text{tr}\left[(\boldsymbol{H}-\boldsymbol{H}_1)\boldsymbol{Y}\boldsymbol{Y}^T\right]/m_2}{\text{tr}\left[(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y}\boldsymbol{Y}^T\right]/(n-m)} = \frac{\text{tr}(\boldsymbol{H}_2\boldsymbol{Y}\boldsymbol{Y}^T)/m_2}{\text{tr}\left[(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y}\boldsymbol{Y}^T\right]/(n-m)}. \tag{A.4}$$

As McArdle and Anderson (2001) note, $\boldsymbol{Y}\boldsymbol{Y}^T$ is equal to the matrix $\boldsymbol{G}$ defined at the beginning of Section 3.2 of our paper, whence (A.4) equals the pseudo-F statistic

$$\frac{\text{tr}(\boldsymbol{H}_2\boldsymbol{G})/m_2}{\text{tr}[(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{G}]/(n-m)}. \tag{A.5}$$

The permuted-data pseudo-F statistic is obtained by replacing $\boldsymbol{G}$ with $\boldsymbol{G}_\pi = \boldsymbol{E}_\pi\boldsymbol{G}\boldsymbol{E}_\pi^T$ in (A.5), where $\boldsymbol{E}_\pi$ is a permutation matrix as in the paper. In some settings, however, permuting the raw data is not the optimal way to perform a permutation test. Testing interactions between two factors, in particular, has been much discussed in the permutation test literature. Anderson and ter Braak (2003) recommend permuting residuals of the reduced model (Freedman and Lane, 1983). This means, in terms of our setup, that $\hat{\boldsymbol{Y}}_1$ denotes fitted values with the interaction excluded, and that we refit the model with pseudo-data

$$\hat{\boldsymbol{Y}}_\pi = \hat{\boldsymbol{Y}}_1 + \boldsymbol{E}_\pi(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_1).$$

This formulation presupposes an outcome matrix $\boldsymbol{Y}$. When $\boldsymbol{G}$ is not positive definite, the distance matrix does not correspond to Euclidean distances among rows of such a matrix, so there are no actual residuals to permute. However, one can define a pseudo-F statistic for "virtual" permutation of residuals, by adapting McArdle and Anderson's (2001) development as follows:

(1) substitute $\hat{\boldsymbol{Y}}_\pi$ for $\boldsymbol{Y}$ in (A.4) and write the resulting statistic as

$$\frac{\text{tr}(\boldsymbol{H}_2\hat{\boldsymbol{Y}}_\pi\hat{\boldsymbol{Y}}_\pi^T\boldsymbol{H}_2)/m_2}{\text{tr}\left[(\boldsymbol{I}-\boldsymbol{H})\hat{\boldsymbol{Y}}_\pi\hat{\boldsymbol{Y}}_\pi^T(\boldsymbol{I}-\boldsymbol{H})\right]/(n-m)}; \tag{A.6}$$

(2) observe that

$$\boldsymbol{H}_2\hat{\boldsymbol{Y}}_\pi = \boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{Y} + \boldsymbol{H}_2\boldsymbol{E}_\pi(\boldsymbol{I}-\boldsymbol{H}_1)\boldsymbol{Y} = \boldsymbol{H}_2\boldsymbol{E}_\pi(\boldsymbol{I}-\boldsymbol{H}_1)\boldsymbol{Y} \text{ and}$$

$$(\boldsymbol{I}-\boldsymbol{H})\hat{\boldsymbol{Y}}_\pi = (\boldsymbol{I}-\boldsymbol{H})\left[\boldsymbol{H}_1 + \boldsymbol{E}_\pi(\boldsymbol{I}-\boldsymbol{H}_1)\right]\boldsymbol{Y} = (\boldsymbol{I}-\boldsymbol{H})\boldsymbol{E}_\pi(\boldsymbol{I}-\boldsymbol{H}_1)\boldsymbol{Y},$$

where we have used the equalities $\boldsymbol{H}_2\boldsymbol{H}_1 = \boldsymbol{0}$ and $\boldsymbol{H}\boldsymbol{H}_1\boldsymbol{Y} = \boldsymbol{H}_1\boldsymbol{Y}$;

(3) substitute the above into (A.6); and

(4) replace $\boldsymbol{Y}\boldsymbol{Y}^T$ with $\boldsymbol{G}$ as above.

These steps result in the permuted-residuals pseudo-F statistic

$$\frac{\mathrm{tr}\left[\boldsymbol{H}_2\boldsymbol{E}_\pi(\boldsymbol{I}-\boldsymbol{H}_1)\boldsymbol{G}(\boldsymbol{I}-\boldsymbol{H}_1)\boldsymbol{E}_\pi^T\right]/m_2}{\mathrm{tr}\left[(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{E}_\pi(\boldsymbol{I}-\boldsymbol{H}_1)\boldsymbol{G}(\boldsymbol{I}-\boldsymbol{H}_1)\boldsymbol{E}_\pi^T\right]/(n-m)}.$$

We remark that, as one would expect, this expression reduces to that for the permuted-data pseudo-F statistic in the $m_1 = 0$ case.

Web Appendix B: Proofs

We shall require the following lemma, which provides two alternative forms for the pseudo-F statistic of equations (5) and (6).

Lemma 1:
$$\frac{m_2}{n-m}F^* = \frac{tr(\boldsymbol{H}_2\boldsymbol{A})}{tr(-\boldsymbol{H}\boldsymbol{A})} = \frac{\frac{1}{2n}\sum_{i=1}^n\sum_{j=1}^n d_{ij}^2 - tr(\boldsymbol{H}_1\boldsymbol{A})}{tr(-\boldsymbol{H}\boldsymbol{A})} - 1. \qquad \text{(A.}$$

If $m_1 = 0$ then

$$\frac{m-1}{n-m}F^* = \frac{\frac{1}{2n}\sum_{i=1}^n\sum_{j=1}^n d_{ij}^2}{tr(-\boldsymbol{H}\boldsymbol{A})} - 1. \qquad \text{(A.8)}$$

Proof. It is easy to show that the $\boldsymbol{H}_k$'s, as well as $\boldsymbol{I}-\boldsymbol{H}$, are all idempotent and mutually orthogonal. By repeated use of these properties, and of the identity $\mathrm{tr}(\boldsymbol{B}\boldsymbol{C}) = \mathrm{tr}(\boldsymbol{C}\boldsymbol{B})$, we can infer from (5) that

$$\frac{m_2}{n-m}F^* = \frac{\mathrm{tr}(\boldsymbol{H}_2\boldsymbol{A})}{\mathrm{tr}[(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{A}]}.$$

Since $\mathrm{tr}(\boldsymbol{A}) = 0$, this quantity equals the second expression in (A.7). The third expression in (A.7) equals the second since

$$\mathrm{tr}(\boldsymbol{H}_2\boldsymbol{A}) = \mathrm{tr}[(\boldsymbol{H}-\boldsymbol{H}_0-\boldsymbol{H}_1)\boldsymbol{A}] = \mathrm{tr}(\boldsymbol{H}\boldsymbol{A}) + \frac{1}{2n}\sum_{i=1}^n\sum_{j=1}^n d_{ij}^2 - \mathrm{tr}(\boldsymbol{H}_1\boldsymbol{A}). \qquad \text{(A.9)}$$

(A.8) follows easily from (A.7).

Proof of Theorem 1

Let F_π^* denote the value of F^* for the permuted outcomes, i.e.,

$$F_\pi^* = \frac{\mathrm{tr}(\boldsymbol{H}_2 \boldsymbol{G}_\pi \boldsymbol{H}_2)/m_2}{\mathrm{tr}[(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{G}_\pi(\boldsymbol{I}-\boldsymbol{H})]/(n-m).} \tag{A.10}$$

To prove the equivalence of F^* and $\mathrm{tr}(-\boldsymbol{HA})$ it suffices to show that, given (11) and (12), $\mathrm{tr}(-\boldsymbol{HA}_\pi)$ is a decreasing function of F_π^*. But this follows since, by (A.7),

$$\frac{m_2}{n-m}F_\pi^* + 1 = \frac{\frac{1}{2n}\sum_{i=1}^n \sum_{j=1}^n d_{\pi(i)\pi(j)}^2 - \mathrm{tr}(\boldsymbol{H}_1\boldsymbol{A}_\pi)}{\mathrm{tr}(-\boldsymbol{HA}_\pi)} = \frac{\frac{1}{2n}\sum_{i=1}^n \sum_{j=1}^n d_{ij}^2 - K}{\mathrm{tr}(-\boldsymbol{HA}_\pi)}.$$

Proof of Proposition 1

Given that $m_1 = 0$, (11) holds automatically (with $K = 0$), so it suffices to prove (12). In case (i), for any $\pi \in \Pi$, $\mathrm{tr}(-\boldsymbol{HA}_\pi) = \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n h_{ij} d_{\pi(i)\pi(j)}^2 \geqslant 0$. In case (ii), for any π, $\boldsymbol{G}_\pi$ is positive semidefinite by (10) and the positive semidefiniteness of $\boldsymbol{G}$. Since $\boldsymbol{G}_\pi$ is positive semidefinite, so are the two matrices whose traces appear in the numerator and denominator in (A.10). Thus each of these matrices has a nonnegative trace. It follows that $F_\pi^* \geqslant 0$; by (A.8) this implies that again $\mathrm{tr}(-\boldsymbol{HA}_\pi) \geqslant 0$. We conclude that Theorem 1 applies in either of the two cases.

Proof of Proposition 3

We shall (1) verify the conditions of Theorem 1, and infer that the pseudo-F statistic is equivalent to $\mathrm{tr}(-\boldsymbol{H}_2\boldsymbol{A})$; then (2) show that $\mathrm{tr}(-\boldsymbol{H}_2\boldsymbol{A})$ is equivalent to the MRBP statistic.

Step 1. We first prove (11). By (14),

$$\mathrm{tr}(\boldsymbol{H}_1\boldsymbol{A}) = \frac{1}{2bg}\sum_{i=1}^{bg}\sum_{j=1}^{bg} d_{ij}^2 - \frac{1}{2g}\sum_{i=1}^{bg}\sum_{j\sim_b i} d_{ij}^2. \tag{A.11}$$

For the randomized block design, only within-block permutations are permitted. Expression (A.11) is invariant to such permutations, and is less than $\frac{1}{2bg}\sum_{i=1}^{bg}\sum_{j=1}^{bg} d_{ij}^2$. Thus (11) holds.

We next prove (12). Using (16) and the triangle inequality, we obtain

$$\begin{aligned} 2\mathrm{tr}(-\boldsymbol{HA}_\pi) &= \frac{1}{g}\sum_{i=1}^{bg}\sum_{\pi(j)\sim_g \pi(i)} d_{\pi(i)\pi(j)}^2 + \frac{1}{b}\sum_{i=1}^{bg}\sum_{j\sim_b i} d_{\pi(i)\pi(j)}^2 - \frac{1}{bg}\sum_{i=1}^{bg}\sum_{j=1}^{bg} d_{\pi(i)\pi(j)}^2 \\ &\geqslant \frac{1}{g}\sum_{i=1}^{bg}\sum_{\pi(j)\sim_g \pi(i)} d_{\pi(i)\pi(j)}^2 + \frac{1}{b}\sum_{i=1}^{bg}\sum_{j\sim_b i} d_{\pi(i)\pi(j)}^2 \end{aligned} \tag{A.12}$$

$$-\frac{1}{bg}\sum_{i=1}^{bg}\sum_{j=1}^{bg} d^2_{\pi(i)k_{ij,\pi}} - \frac{1}{bg}\sum_{i=1}^{bg}\sum_{j=1}^{bg} d^2_{k_{ij,\pi}\pi(j)},$$

where $k_{ij,\pi}$ is the unique observation belonging to the same group as observation $\pi(i)$ and the same block as observation $\pi(j)$. For given i and π, the bg values of $k_{ij,\pi}$ comprise b copies of $\pi(j)$ for each j such that $\pi(j) \sim_g \pi(i)$. Thus the third term of (A.12) cancels with the first. Similarly the fourth term cancels with the second. Hence (12) holds. By Theorem 1, the pseudo-F statistic is equivalent to $\mathrm{tr}(-\boldsymbol{H}\boldsymbol{A})$, which is in turn equivalent to $\mathrm{tr}(-\boldsymbol{H}_2\boldsymbol{A})$, since by (A.9) and (11), these two expressions differ by a quantity that is permutation-invariant.

Step 2. By (15), $2\mathrm{tr}(-\boldsymbol{H}_2\boldsymbol{A}) = \frac{1}{b}\sum_{i=1}^{bg}\sum_{j\sim_g i} d^2_{ij} - \frac{1}{bg}\sum_{i=1}^{bg}\sum_{j=1}^{bg} d^2_{ij}$. The second term on the right side is permutation-invariant, so $\mathrm{tr}(-\boldsymbol{H}_2\boldsymbol{A})$ is equivalent to the first term, which is proportional to the MRBP statistic (4) with dissimilarity $\Delta_{ij} = d^2_{ij}$.

Combining steps 1 and 2, we conclude that the pseudo-F and MRBP tests are equivalent.

Web Appendix C:

A non-metric squared distance for which the conclusion of Proposition 3 does not hold

Consider a randomized block design with two groups and two blocks, with distance matrix $\boldsymbol{D} = \begin{pmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 2 & 1 \\ 1 & 2 & 0 & 1 \\ 2 & 1 & 1 & 0 \end{pmatrix}$, where block 1 consists of the first two observations and group 1 consists of observations 1 and 3. This distance function is a metric but its square is not. There exist within-block permutations such that both the pseudo-F statistic and the MRBP statistic (based on squared distances as in Proposition 3) are larger for the permuted data than for the real data. In other words, applying such a permutation results in a pseudo-F statistic that is nominally less consistent with the null, but an MRBP statistic that is more consistent. We say "nominally" because in this (admittedly artificial) instance the pseudo-

F statistic is negative, and hence difficult to interpret. At any rate the two tests are not equivalent for the given distances.

References

Anderson, M. J., and ter Braak, C. J. F. (2003). Permutation tests for multi-factorial analysis of variance. *Journal of Statistical Computation and Simulation* **73,** 85–113.

Freedman, D., and Lane, D. (1983). A nonstochastic interpretation of reported significance levels. *Journal of Business and Economic Statistics* **1,** 292–298.

McArdle, B. H., and Anderson, M. J. (2001). Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* **82,** 290–297.