1.17 Methods of permutations

As for the one-way case, the distribution of each of the pseudo-*F* ratios in a multi-way design is generally unknown. Thus, a permutation test (or some other approach using re-sampling methods) is desirable. When there is more than one factor, situations commonly arise which prevent the possibility of obtaining an exact test of individual terms in the model using permutations. For example, there is no exact permutation test for an interaction (but see Pesarin (2001) , who describes a synchronised permutation method for testing interactions). In addition, restricted permutation methods for testing main effects in ANOVA models generally have low power (Anderson & ter Braak (2003)). However, several good approximate permutation methods can be used instead to get accurate *P*-values (e.g., Anderson & Legendre (1999)). PERMANOVA provides three general options regarding the method of permutation to be used: (i) unrestricted permutation of raw data, (ii) permutation of residuals under a reduced model, or (iii) permutation of residuals under the full model. These methods and their properties are described in detail elsewhere (e.g., Anderson & Legendre (1999) , Anderson (2001b) , Anderson & Robinson (2001) , Anderson & ter Braak (2003) , Manly (2006)). Although these methods do not give exact *P*-values for complex designs in all cases, they are *asymptotically* exact[26] and give very reliable results. In practice, these three approaches will give very similar results, so there is (thankfully) no need to agonise much about making a choice here. All three of the methods are implemented in PERMANOVA so as to ensure that the correct exchangeable units (identified by the denominator of the pseudo-*F* ratio) are used for each individual test (see Anderson & ter Braak (2003) for details). Some of the known properties of the three methods are outlined below.

- *(i) Unrestricted permutation of raw data.* This is a good approximate test proposed for complex ANOVA designs by Manly (1997) . It will generally have type I error close to α, although with larger sample sizes it tends to be more conservative (less powerful) than the tests that permute residuals (Anderson & ter Braak (2003))[27]. However, this method does not need large sample sizes to work well (Gonzalez & Manly (1998)). It is also, computationally, the fastest option. The method does suffer from a few problems, however, if there happen to be outliers in covariables (if present, see Kennedy & Cade (1996)), so should not be used for such cases.

- *(ii) Permutation of residuals under a reduced model.* This approach was first described for linear models by Freedman & Lane (1983) and is the default option in PERMANOVA because it has excellent empirical and theoretical properties. Empirically, it yields the best power and the most accurate type I error for multi-factorial designs in the widest set of circumstances (Anderson & Legendre (1999) , Anderson & ter Braak (2003)). Also, this method is theoretically the closest to the conceptually exact test (Anderson & Robinson

(2001)). The idea is to isolate the term of interest in the model for each test by fitting the other terms (the *reduced* model), obtaining residuals from that reduced model, and permuting those. In other words, the entities that are exchangeable under the null hypothesis for a particular term are the *errors* (estimated by the residuals) obtained after *removing* the terms in the model that are not of interest for that test. The definition of the reduced model (and therefore the residuals arising from them) thus depends on which term is being tested[28].

- *(iii) Permutation of residuals under the full model.* This method was described by ter Braak (1992) . The idea is to obtain residuals of the full model by subtracting from each replicate the mean corresponding to its particular cell (combination of factor levels). These residuals are estimating the errors associated with each replicate. These are then permuted and the statistic is re-calculated *for all terms using these residuals* (as if they were the data) under permutation. This method mostly gives results highly comparable to method (ii). It relies somewhat more than (ii), however, on large within-cell sample sizes for precision. It has the advantage, however, of being faster than method (ii) for the analysis of the entire design, as the same residuals are permuted for all terms in the model under test.

In general, we recommend using method (ii), which is the default. Method (i), however, does provide an exact test for the one-way case, so should be used for one-way ANOVA models. Otherwise, note that methods (ii) and (iii) both require estimation of parameters (means) in order to calculate residuals as deviations from fitted values. When sample sizes are small, these estimates are not very precise (i.e., they may not be very close to their "true" values), so the residuals being permuted, in turn, may not be good representatives of the "true" errors (Anderson & Robinson (2001)). Thus, in the case of relatively small sample sizes (say, $n < 4$ replicates per cell), method (i) is also recommended (provided there are no outliers in covariables, as mentioned above). Method (iii) is probably only advisable if you wish to use method (ii), but the time required is getting overly burdensome.

---

[26] An *asymptotically exact test* is a test for which the type I error (probability of rejecting the null hypothesis when it is true) asymptotically approaches (converges on) the *a priori* chosen significance level (α) with increases in the sample size (N).

[27] Note that "less powerful" does not necessarily mean that using (i) will give you a smaller *P*-value than (ii) or (iii) for any particular data set. It means that, in repeated simulations, the empirical power (estimated probability of rejecting the null hypothesis when it is false) was, on average, smaller for method (i) than for either of the other two methods in most situations.

[28] For unbalanced designs, it also depends on which Type of SS is chosen for the test. For Type I SS, the order in which the terms are fitted will also matter here.