1.13 PERMANOVA versus ANOSIM

The analysis of similarities (ANOSIM), described by Clarke (1993) is also available within PRIMER and can be used to analyse multivariate resemblances according to one-way and some limited two-way experimental designs[19]. Not surprisingly, ANOSIM and PERMANOVA will tend to give very similar results for the one-way design on a given resemblance matrix. There are two essential differences, however, between ANOSIM and PERMANOVA. First, ANOSIM *ranks* the values in the resemblance matrix before proceeding with the analysis. The rationale behind the ranking procedure in ANOSIM is that the information of interest is the *relationships* among the dissimilarities (i.e., whether a given dissimilarity is larger or smaller than another) and not the *values* of the dissimilarities themselves. This is consistent with the philosophy of non-metric MDS ordination, which seeks to preserve only the rank order of the dissimilarities among samples. In contrast, PERMANOVA takes the point of view that the information of interest is in the dissimilarity values themselves, which describe a cloud of samples in multivariate space. This means that for PERMANOVA one must take special care to choose a measure of resemblance that is meaningful for the data and the goals of the analysis. For example, squared Euclidean distance may give different PERMANOVA results than Euclidean distance itself, whilst such a monotonic transform of the resemblances does not change the ranks and therefore cannot change ANOSIM.

The second essential difference is in the construction of the test statistic. The ANOSIM R statistic (

Clarke (1993)) is scaled to take a value between -1 and +1. This is a very useful feature, as it makes it possible to interpret the R statistic directly as an absolute measure of the strength of the difference between groups. R values are also directly comparable among different studies. In contrast, the value of pseudo-F (or pseudo-t) is, first of all, necessarily reliant on the degrees of freedom of the analysis, so cannot necessarily be compared in value across studies. A value of pseudo-F = 2.0 (like its univariate analogue) will generally provide much stronger evidence against the null hypothesis if the residual degrees of freedom are 98 than if they are 5. Although values of pseudo-F may be comparable across different tests where the degrees of freedom are equal (for a given dissimilarity measure and original number of variables, that is), it is also worth bearing in mind that the variability *among* groups (as measured by the numerator of the statistic) is always scaled against the variability *within* groups (as measured by the denominator). Thus, the within-group variability has an important role to play in the value of pseudo-F (or pseudo-t). An example is provided by the Victorian avifauna comparisons (Fig. 1.13), where, despite the pattern shown on the MDS plot (that samples from poor sites are farther away from the good sites than are the medium sites), pseudo-t is actually *larger* for the difference between good and medium sites than it is between good and poor sites, simply because the within-group variability between the poor sites is so high. ANOSIM, in contrast, yields an R statistic value of 1.0 (its maximum possible value) in both cases. In summary, while ANOSIM's R can be interpreted directly as a measure of the size of the between-group differences, PERMANOVA's pseudo-F (or pseudo-t) cannot necessarily be interpreted in this way. The sizes of effects in PERMANOVA are measured and compared in other ways: either by the average similarities (or dissimilarities) among pairs of groups (provided by the pair-wise routine) or by examining the estimated sizes of components of variation (see the section Estimating components of variation). In addition, in PERMANOVA it is the P-values (either 'P(perm)'

or, when necessary, 'P(MC)') which should be used as a measure of *strength of evidence* with respect to any particular null hypothesis. The Monte Carlo option available here also means that the power of the test need not especially rely on the number of possible permutations, as is the case for ANOSIM. Power in PERMANOVA will rely, however, on the number of replicates (more particularly, on the denominator degrees of freedom) available for the test.

Unlike ANOSIM, PERMANOVA achieves a partitioning of multivariate variability. As discussed in the introduction (page 0.3), this means PERMANOVA can be used to analyse much more complex experimental designs than ANOSIM. Although one could conceivably rank the dissimilarities before proceeding with a PERMANOVA analysis, this is not generally advisable when the goal is to achieve a partitioning of multivariate variability. The reason is that ranking dissimilarities loses information and therefore may result in less power[20]. Another reason is that ranking the dissimilarities will tend to make the multivariate system highly non-metric, which can result in *negative* sums of squares and thus negative values of pseudo-*F*! The concept of negative variance which arises in non-metric or semi-metric geometric systems is discussed in more detail by Legendre & Legendre (1998) and McArdle & Anderson (2001) and in chapter 3 below on principal coordinates analysis (PCO). Suffice it for now to state simply that such results are confusing and difficult to interpret. They usually result from a poor choice of resemblance measure, or from ranking resemblances unnecessarily, so should be avoided if possible. The partitioning of variability described by the resemblance matrix on the basis of most reasonable dissimilarity measures (that have *not* been ranked) will generally produce a result where all of the SS (and pseudo-*F* ratios) are positive.

---

[19] See chapter 12 in Clarke & Gorley (2006) and chapter 6 in Clarke & Warwick (2001) .

[20] This is analogous to the way that non-parametric univariate statistics are less powerful than their more traditional parametric counterparts when the assumptions of the latter are fulfilled. Interestingly enough, distance-based permutation tests (using Euclidean distance) can achieve even greater power than the traditional MANOVA test statistics in some situations, even when the assumptions of the traditional tests are true (Smith (1998) , Mielke & Berry (2001) , see also chapter 5 on CAP).