CHAPTER 24

MRPP (Multi-response Permutation Procedures)

and Related Techniques

Background

MRPP is a nonparametric procedure for testing the hypothesis of no difference between two or more groups of entities. For example, one could compare species composition between burned and unburned plots to test the hypothesis of no treatment effect.

MRPP is just one of numerous independently developed nonparametric methods for testing for group differences. Two variants (ANOSIM and the Q_b method) can be found at the end of this chapter. A recent breakthrough (NPMANOVA, Anderson 2001) allowing the use of non-Euclidean distance measures in multifactorial designs is also described at the end of this chapter.

A good introduction to MRPP is the appendix in Biondini et al. (1985). More details can be found in Berry et al. (1983), Mielke (1984), and Mielke and Berry (2001).

The method requires pre-existing groups of entities or sample units. Commonly these groups are inherent in the sampling design or experimental design. For exploratory analysis they can be defined in any way you want. For example, two groups of sample units could be based on the presence-absence of a species of particular interest. You can also define groups based on a categorical environmental variable, such as whether the bedrock is limestone, granitic rock, or mafic rock.

MRPP provides little more than the test statistic, a measure of "effect size," and a p-value. Describing the differences among groups is usually done with one or more of the following methods:

1. **Overlay** the variable that defines group membership on an ordination of sample units. This gives a graphical representation of the relationships among groups. It is particularly useful when there are only two or three strong underlying patterns of variation.

2. For community data, describe the indicator value of individual species for separating the groups with Dufrêne and Legendre's (1997) **indicator species analysis** (Ch. 26).

3. Use **discriminant analysis** (Ch. 26) in a descriptive mode to identify variables most effective in predicting group membership. The size of the standardized discriminant function coefficient associated with each variable indicates the contribution of that variable to the discriminant function.

4. Use a **mean similarity dendrogram** (Van Sickle 1997) to graphically show relationships among groups. This is essentially a cluster analysis of within-group averages (or medians). Conceptually it is closely linked to MRPP, being based on mean similarities of replicate objects within the same class. The results are plotted on a scale of the original distance units rather than by Wishart's Objective Function (see Chapter 10). The end of each branch of the tree is clipped at the average within-group distance for the group represented by that branch.

When to use it

Discriminant analysis and multivariate analysis of variance (MANOVA) are parametric procedures that can be used on the same general class of questions as MRPP. However, MRPP has the advantage of not requiring distributional assumptions (such as multivariate normality and homogeneity of variances) that are seldom met with ecological community data. The default choice for community data should, therefore, be MRPP or a similar randomization technique.

With multivariate normal data, MRPP can be more powerful or less powerful than MANOVA (and a test based on Hotelling's T^2), depending on the nature of the data (Smith 1998). But with the nonlinear relationships and extremely skewed frequency distributions typical of community data, MANOVA is seldom a reasonable choice.

Complex experimental designs are difficult to analyze with MRPP and related techniques. If you must use a complex design, you can either use MANOVA, an MRPP-related technique based on Euclidean distance (e.g., see the Q_b method below), or analyze your data piecewise with MRPP, slicing it various ways to answer different questions. The last

approach sacrifices your ability to analyze the interaction terms. You can, however, visualize those interactions using your experimental factors in a joint plot overlay on an ordination diagram (Ch. 13). Independent, noninteracting factors will show up as perpendicular vectors.

Assumptions

It is sometimes said, though it is untrue, that nonparametric statistics require no assumptions. Nonparametric statistics such as MRPP do, however, avoid *distributional* assumptions. In some cases this comes with a loss in power. The following assumptions should be considered when applying MRPP and related techniques:

1. The distance measure chosen adequately represents the variation of interest in the data.
2. The sample units are independent. The usual problems with pseudoreplication, subsampling, and repeated measures are conceptually the same as with ANOVA.
3. The relative weighting of the variables has been controlled prior to calculating the distance measure, such that the weighting of variables is appropriate for assessing the ecological question at hand.

How it works

1. Calculate distance matrix, **D**. Various distance measures can be used. Euclidean distance and squared Euclidean distance were used most often in the early literature of MRPP. Zimmerman et al. (1985) recommended Euclidean distance over squared Euclidean distance. Proportional city-block distance measures (e.g., Sørensen distance) are increasingly used in published studies with MRPP and community data. If MRPP is used in conjunction with ordination, then it is often desirable to choose the same distance measure for both.

2. Calculate the average distance x_i within each group i.

3. Calculate delta (the weighted mean within-group distance)

$$delta = \delta = \sum_{i=1}^{g} C_i x_i$$

for g groups, where C is a weight that depends on the number of items in the groups (normally $C_i = n_i / N$, where n_i is the number of items in group i and N is the total number of items). Note that all $n_i \geq 2$. For a given mean overall distance, smaller values of δ indicate tighter clustering within groups.

Various other weightings have been used and are generally based on group size. Those available in PC-ORD are in Table 24.1. In each case, n_i is the number of items in group i, g is the number of groups, and C_i is the weight applied to each item in group i.

Table 24.1. Methods for weighting groups in MRPP.

Formula	Comments
$C_i = \frac{n_i}{\sum n_i}$	A natural weighting recommended by Mielke (1984) and used in most recent applications of MRPP.
$C_i = \frac{n_i - 1}{\sum (n_i - 1)}$	With squared Euclidean distance this weighting results in an MRPP statistic that is equivalent to a 2-sample t-test or one-way ANOVA F-test (Mielke et al. 1982; Zimmerman et al. 1985). While this option accounts for degrees of freedom, this is a foreign concept to permutation procedures.
$C_i = \frac{1}{g}$	Not recommended but available for experimentation.
$C_i = \frac{n_i * (n_i - 1)}{\sum (n_i * (n_i - 1))}$	Not recommended but available for experimentation. Used in some early applications of MRPP.

4. Determine probability of a δ this small or smaller. The brute force or exact method (applicable only to tiny data sets) is to calculate δ for all possible partitions of the same sizes. This is the same as the number of possible permutations of the values assigning cases (sample units) to groups (Fig. 24.1). The number of possible partitions (M) for two groups is

$$M = N!/(n_1! * n_2!)$$

For example, for $n_1 = n_2 = 15$, $M = 1.55 \times 10^8$

Assume all partitions could have occurred with equal chance. Then calculate the proportion of these that have δ smaller than the observed δ (Fig. 24.2).

$$p = \frac{1 + \text{no. smaller deltas}}{\text{total no. possible partitions}}$$

A more reasonable method that is also applicable to medium or large data sets is to approximate the distribution of δ from a continuous distribution (Pearson type III). This distribution accommodates the fact that the underlying permutation distribution is often substantially skewed. The Pearson type III distribution incorporates three parameters, the mean m, standard deviation s, and gamma g (skewness of δ under the null hypothesis).

The test statistic, T is

$$T = (\delta - m_\delta) / s_\delta$$

where m_δ and s_δ are the mean and standard deviation of δ under the null hypothesis. In words, the test statistic is the difference between the observed and expected deltas divided by the square root of the variance in delta. (Do you see the resemblance to the Student's t-test?)

$$T = \frac{\text{observed } \delta - \text{expected } \delta}{\text{s.dev. of expected } \delta}$$

The test statistic, T, describes the separation between the groups. The more negative is T, the stronger the separation. The observed delta (the average within-group distance) is compared to an expected delta, the latter calculated to represent the mean delta for all possible partitions of the data. The variance and skewness of delta are descriptors of the distribution of all possible deltas corresponding to the possible partitions of the items. The question then becomes, is the observed delta at all unusual considering the distribution of possible deltas? The probability value expresses the likelihood of getting a delta as extreme or more extreme than the observed delta, given the distribution of possible deltas. The p-value associated with T is determined by numerical integration of the Pearson type III distribution.

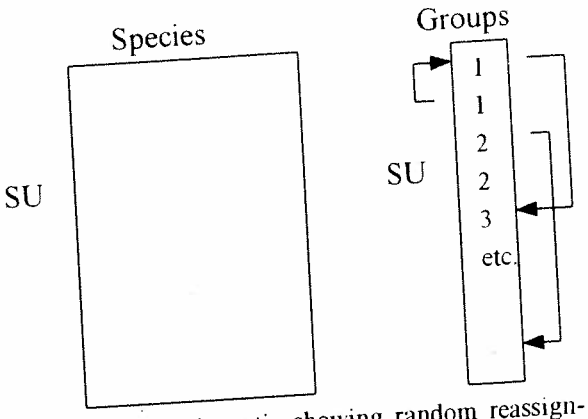

Figure 24.1. Schematic showing random reassignment of sample units to groups. The assignments of SUs to groups are shuffled, while the species matrix is held constant.

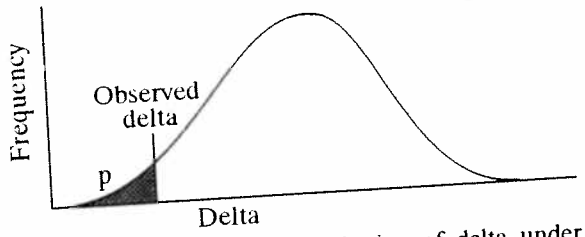

Figure 24.2. Frequency distribution of delta under the null hypothesis, compared to the observed delta. The area under the curve less than the observed delta is the probability of type I error under the null hypothesis of no difference between groups.

5. The p value is useful for evaluating how likely it is that an observed difference is due to chance, but we also need a description of the *effect size* that is independent of the sample size. This is provided by the **chance-corrected within-group agreement** (A):

$$A = 1 - \frac{\delta}{m_\delta} = 1 - \frac{\text{observed } \delta}{\text{expected } \delta}$$

The agreement statistic A describes within-group homogeneity, compared to the random expectation. (Originally shown as script R in Mielke's papers and R in older versions of PC-ORD, we have since adopted the symbol A to avoid confusion with regression

statistics.) When all items are identical within groups, then the observed delta = 0 and $A = 1$, the highest possible value for A. If heterogeneity within groups equals expectation by chance, then $A = 0$. On the other hand, if there is LESS agreement within groups than expected by chance, then $A < 0$.

In community ecology, values for A are commonly below 0.1, even when the observed delta differs significantly from the expected. An $A > 0.3$ is fairly high. Remember that for $A = 1$, all items must be identical within groups.

Statistical significance (small p) may result even when the "effect size" (A) is small, if the sample size is large. For example, you may find that $A = 0.01$ is associated with a statistically significant delta with $N = 200$. In such cases, you need to carefully consider the ecological significance of the result, not just the statistical significance. On the other hand, with a small sample size, a large effect size is needed to achieve statistical significance.

What to report

Because many readers will not be familiar with this method, you may wish to give a brief statement of the **purpose and general approach** of the method. Likewise, you should cite a paper that gives details of the method. Mielke (1984) and Mielke and Berry (2001) are good references for the method. Biondini et al. (1985) and Zimmerman et al. (1985) were some of the earliest applications to ecological problems.

You should also report:

- The software (because MRPP is not available in most statistical packages).
- Distance measure (consider including a justification for this).
- How groups were defined (this should directly or indirectly include a statement of the size of each group).
- Chance-corrected within-group agreement, A.
- p-value.

Examples

Simple example

A simple hypothetical example illustrates the basic concepts of MRPP. Assume we have three a priori groups of five plots each. In each plot we measured the abundance of two species (Table 24.2). The bivariate scatterplot (Fig. 24.3) shows a dust-bunny distribution typical of community data.

We first calculate a distance matrix (Table 24.2), but only a portion of these values are actually used. The between-group distances (shaded cells of Table 24.2) are ignored. Omitting the redundant cells above the diagonal leaves just the within-group distances, in this case ten distance values for each group. From these, we calculate the average within-group distance for each group (Table 24.3). For comparison we redid the calculation for Euclidean distance and the Sørensen distances converted to ranks. The latter yields a nonmetric MRPP, analogous to treatment of the distance matrix in nonmetric multidimensional scaling.

The average within-group distances (Table 24.3) show that Groups 1 and 3 have relatively high dispersions, while Group 2 is relatively tight. This accords with our intuition from Figure 24.3.

Clearly, we can reject the null hypothesis of no difference among groups (Table 24.4). The three groups occupy different regions of species space, as shown by the strong chance-corrected within-group agreement (A) and test statistic (T). In this case, we would have reached the same conclusion with any of the distance matrices. Selection of a distance measure should, however, be based on properties of the distance measures rather than comparison of results.

A particular pair of groups can be compared using the same methods as the comparison across all groups.

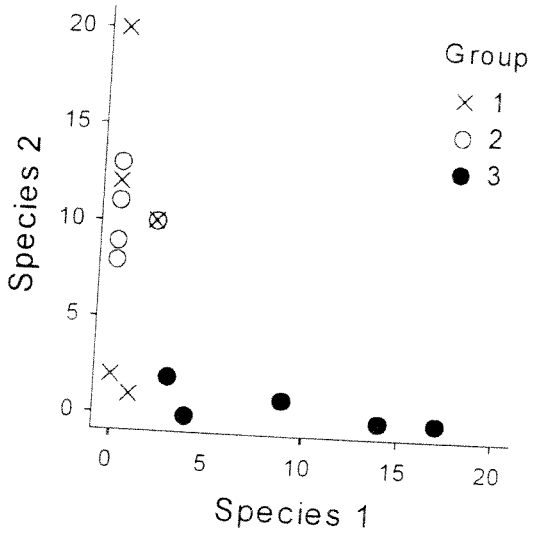

Figure 24.3. Fifteen sample units in species space, each sample unit assigned to one of three groups.

Table 24.2. A species data matrix of 15 plots by 2 species, their assignments to three groups, and Sørensen distances among plots. Shaded cells are between-group distances, ignored by MRPP.

	Species			Sørensen distance matrix													
Plot	Sp1	Sp2	Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	20	1	0.000	0.250	0.375	0.818	0.909	0.212	0.290	0.375	0.379	0.429	0.840	1.000	0.933	1.000	1.000
2	0	12	1	0.250	0.000	0.167	0.714	0.857	0.040	0.043	0.167	0.143	0.200	0.765	1.000	0.909	1.000	1.000
3	2	10	1	0.375	0.167	0.000	0.714	0.714	0.200	0.130	0.000	0.143	0.200	0.529	0.750	0.727	0.846	0.862
4	0	2	1	0.818	0.714	0.714	0.000	0.500	0.733	0.692	0.714	0.636	0.600	0.429	1.000	0.833	1.000	1.000
5	1	1	1	0.909	0.857	0.714	0.500	0.000	0.867	0.846	0.714	0.818	0.800	0.429	0.667	0.667	0.875	0.895
6	0	13	2	0.212	0.040	0.200	0.733	0.867	0.000	0.083	0.200	0.182	0.238	0.778	1.000	0.913	1.000	1.000
7	0	11	2	0.290	0.043	0.130	0.692	0.846	0.083	0.000	0.130	0.100	0.158	0.750	1.000	0.905	1.000	1.000
8	2	10	2	0.375	0.167	0.000	0.714	0.714	0.200	0.130	0.000	0.143	0.200	0.529	0.750	0.727	0.846	0.862
9	0	9	2	0.379	0.143	0.143	0.636	0.818	0.182	0.100	0.143	0.000	0.059	0.714	1.000	0.895	1.000	1.000
10	0	8	2	0.429	0.200	0.200	0.600	0.800	0.238	0.158	0.200	0.059	0.000	0.692	1.000	0.889	1.000	1.000
11	3	2	3	0.840	0.765	0.529	0.429	0.429	0.778	0.750	0.529	0.714	0.692	0.000	0.333	0.467	0.684	0.727
12	4	0	3	1.000	1.000	0.750	1.000	0.667	1.000	1.000	0.750	1.000	1.000	0.333	0.000	0.429	0.556	0.619
13	9	1	3	0.933	0.909	0.727	0.833	0.667	0.913	0.905	0.727	0.895	0.889	0.467	0.429	0.000	0.250	0.333
14	14	0	3	1.000	1.000	0.846	1.000	0.875	1.000	1.000	0.846	1.000	1.000	0.684	0.556	0.250	0.000	0.097
15	17	0	3	1.000	1.000	0.862	1.000	0.895	1.000	1.000	0.862	1.000	1.000	0.727	0.619	0.333	0.097	0.000

Table 24.3. Average within-group distance calculated from three different distance matrices. The average within-group distance is used as the test statistic.

	Average within-group distance		
Group	Sørensen	Ranked Sørensen	Euclidean
1	0.602	0.453	9.78
2	0.149	0.159	2.79
3	0.449	0.337	7.79
Average	0.400	0.316	6.79

Table 24.4. Summary statistics for MRPP of simple example. Results are given for three different distance matrices, comparing across all groups, as well as for multiple pairwise comparisons for the Sørensen distances. The pairwise comparisons were also made with MRPP.

		δ under null hypothesis					
	Observed δ	Expected	Variance	Skewness	*T*	*p*	*A*
Sørensen distances	0.400	0.625	0.0019	-1.24	-5.14	0.0007	0.359
Ranked Sørensen	0.316	0.495	0.0012	-1.26	-5.18	0.0007	0.361
Euclidean	6.79	9.97	0.5177	-1.18	-4.43	0.0017	0.320
Multiple comparisons (Sørensen)							
1 vs 2	0.376	0.397	0.0005	-0.31	-0.93	0.1730	0.055
1 vs 3	0.526	0.699	0.0019	-1.26	-3.89	0.0039	0.248
2 vs 3	0.299	0.627	0.0036	-2.11	-5.49	0.0017	0.523

The pairwise comparisons show that Groups 1 and 2 are broadly overlapping, and we cannot reject the null hypothesis (Table 24.4). The comparisons of Groups 1 vs. 3 and 2 vs. 3 yielded statistics comparable to the overall comparisons. The A values are similar in size, but the p-values are somewhat smaller than for the overall test, reflecting the smaller sample size. Note that the hazards of multiple comparisons apply here. A conservative approach would be to apply a Bonferroni procedure to control the "experimentwise" error rate.

Examples in the literature

Lesica et al. (1991) tested the null hypothesis of no difference in species composition between old-growth and second-growth forests. The test was separately applied to four different groups of species. The average within-group distance allowed the reader to assess how tightly the old-growth and second-growth sample units grouped in species space for each of the species groups.

That study and most of the other early applications of MRPP used Euclidean distance between sample units in species space. Informal comparisons of MRPP based on Sørensen distances vs. Euclidean distances have often shown similar results. Sørensen distances are, however, less prone to exaggerate the influence of outliers. Use of MRPP with Sørensen distances or ranked Sorensen distances is increasingly frequent (McCune et al. 2000).

Variations

Rank-transformed MRPP

One can rank transform the distance matrix, then conduct an otherwise normal MRPP. Any of the distance measures can be rank transformed. The rank transformation can help to correct the loss of sensitivity of distance measures as community heterogeneity increases. It also subtly changes the null hypothesis being answered from "average within-group distance no smaller than expected by chance" to "no difference in average within-group ranked distances." It also makes the MRPP results more analogous in theory to those from nonmetric multidimensional scaling. In theory, MRPP on ranked distances is more similar to ANOSIM (Clarke & Green 1988, Clarke 1993) than is MRPP with raw distances.

With community data, the test statistic, skewness of the test statistic under the null hypothesis, and the resulting p-value are often similar, whether the data are ranked or not. The chance-corrected within-group agreement, however, is often considerably higher after the distance measure is converted to ranks.

The ranking procedure operates on nonredundant distances (i.e., one triangle of the square, symmetrical distance matrix). Ties are assigned the average rank of the tied elements. For example, the values 1, 3, 3, 9, 10 would receive ranks 1, 2.5, 2.5, 4, 5. After elements are assigned initial ranks, they are adjusted by subtracting the rank given to zero distances. This results in all raw distances of zero being assigned a rank distance of zero. For example, if there were five zero distances in the matrix, they would each initially be assigned a rank of 3, taking into account the five-way tie. Then 3 would be subtracted from each element in the matrix.

McCune et al. (2000) used nonmetric MRPP (based on a rank-transformed Sørensen distance matrix) to compare species composition among groups of sample units in a data set with high beta diversity. This choice of distance matrix enhanced the correspondence of the MRPP results with the nonmetric multidimensional scaling that they used to illustrate the relationships among sample units.

Blocked MRPP (MRBP)

Randomized block experiments, paired-sample data, and simple repeated-measures designs are very common in ecology. These designs can be analyzed with a variant of MRPP called MRBP or blocked MRPP. The method is explained by Mielke (1984, 1991), Mielke and Berry (1982), Mielke and Iyer (1982), and Biondini et al. (1988). It has been applied to ecological data by Biondini et al. (1988), Peterson & McCune (2001), Ponzetti & McCune (2001), and Zimmerman et al. (1985).

Given b blocks and g groups (treatments), the MRPP statistic is modified to:

$$\delta = \left[g\binom{b}{2} \right]^{-1} \sum_{i=1}^{g} \sum_{j<k} \Delta(x_{ij}, x_{ik})$$

where $\Delta(x,y)$ is the distance between points x and y in the p-dimensional space. Each of the p dimensions should be measured on the same scale or standardized to a common scale. In community ecology, each of the p dimensions would normally correspond to a different species or a different environmental variable. Note that for paired-sample data, b is the number of pairs of observations and $g = 2$. The combinatoric term is simply the number of items represented in the double summation. Thus δ is the average distance between

blocks within treatments, including only nonredundant off-diagonal elements of the distance matrix. Normally δ is calculated after alignment of the blocks.

The null hypothesis assigns equal probabilities to each of the $M = (g!)^b$ possible allocations of the g, r-dimensional measurements to the g treatments within each of the b blocks. In other words, the observed values are randomly reassigned to different treatments in each block.

Like MRPP, small values of δ imply a concentration of treatments in the p dimensional space. The added twists of MRBP are that these distances are summed with respect to the blocks, and that the user has the option of aligning blocks so that all treatments in a given block have a median of zero.

Equations for estimating the standard deviation and skewness of δ are relatively complex (see Mielke 1984) and are not reproduced here. The test statistic, T, is based on these values as in MRPP and the p-value is again approximated from a Pearson type III distribution.

Note that this analysis requires a balanced design: there must be one sample unit for each combination of block and treatment. The number of treatments must be the same among blocks and each treatment must be present in each block.

Consider the following application to a paired-sample design. Imagine a study of predation effects on rocky intertidal communities. We paired treatment and control plots, the treatment being exclosure of a predator. The purpose of the pairing was to control for other variables, such as wave exposure, that influenced the communities. Each pair of plots was considered a block, and groups were defined by the treatments. Using blocked MRPP focuses the analysis on within-block differences, presumably due to the treatment alone. In contrast, randomly placed plots without blocking would have added the variation due to wave exposure, reducing our ability to detect the treatment effect.

Several options provide flexibility with MRBP:

Average distance function commensuration. This option equalizes the contribution of each variable to the distance function. For each variable m the sum of deviations (Dev_m) is calculated:

$$Dev_m = \sum_{i=1}^{g}\sum_{j=1}^{b}\sum_{k=1}^{g}\sum_{l=1}^{b} |x_{mij} - x_{mkl}|^V$$

V is set to 2 for squared Euclidean distance or 1 for Euclidean distance. Then each element x of the data matrix is divided by the sum of the deviations for the corresponding variable to produce the transformed value y:

$$y_{mij} = x_{mij} / Dev_m$$

Distance function commensuration relativizes your variables. If you have already relativized your data then it is unnecessary. If you have not already relativized your data, you may or may not want to at this point. The same criteria apply as discussed under relativization in Chapter 9.

Median alignment within blocks. If the median for each variable in each block is subtracted from the raw data for each block, then the medians are said to be aligned to zero for all blocks (Table 24.5). Usually, alignment in a randomized block design is desirable, focusing the analysis on within-block differences among treatments. But if the problem is conceptualized as paired agreement, say between model predictions and observed data, then alignment is not used. In this case, each event or observation represents a group, one block contains the observed data, and one block contains the predicted data. Because the goal is an exact match of predicted and observed, rather than the two just being correlated, then the medians should not be aligned. If, however, the goal is to determine whether or not the two sets of numbers are correlated apart from any exact agreement in value, then the blocks should be aligned.

Distance measure. Euclidean distance has been used most often with MRBP. If squared Euclidean distance is chosen, then the resulting statistics are comparable to permutation tests of Pearson correlation. Methods dependent on the normal distribution require sample statistics based on squared Euclidean distances. Because MRPP, MRBP, and related statistics are based on a permutation distribution, there is no necessity for choosing squared Euclidean distance. Mielke (1991) and Mielke & Berry (2001) argued convincingly against using squared Euclidean distances. Mielke (1991, pp. 56-57) stated that "absolute deviations" (meaning Euclidean distances) are more logical and intuitive than squared deviations (squared Euclidean distance). Although Mielke's papers did not consider proportional city-block distance measures, Sørensen distance is incompatible with median alignment. Alignment results in both positive and negative values, but Sørensen distances require nonnegative data. A nonproportional city-block distance could, however, be used in conjunction with median alignment.

Table 24.5. Example comparing results from raw data versus data aligned within blocks to zero as input to Blocked MRPP.

	Raw Data		Aligned Data	
	Block 1	Block 2	Block 1	Block 2
Group 1	4	9	1.5	1.5
Group 2	2	7	-0.5	-0.5
Group 3	3	8	0.5	0.5
Group 4	1	2	-1.5	-5.5
Median	2.5	7.5	0	0
Observed δ	5 = (5+5+5+1)/4		1 = (0+0+0+4)/4	
Expected δ	4.375		2.225	
Agreement (A)	0.086		0.556	
p	0.184		0.016	

Permutation test for bivariate Pearson correlation coefficient. This method offers a nonparametric test of the null hypothesis that a bivariate correlation coefficient is equal to zero. Assume we are looking at the relationship between N measurements of variables X and Y. If there is only one response variable, squared Euclidean distance is selected, each variable is selected as a block, and there are N groups, then MRBP yields a permutation test for the Pearson correlation coefficient, r. Under the null hypothesis, δ is equivalent to r (Mielke 1984, p. 824). See that same paper for a description of the relationship between δ and the standard F statistic.

Analysis of similarity (ANOSIM)

ANOSIM (Clarke & Green 1988; Clarke 1993) is similar to MRPP in concept but uses a different test statistic. Elements of a similarity matrix among all sample units are ranked. The highest similarity is given a rank of 1.

$$R = \left(\bar{r}_B - \bar{r}_W\right) / \left(M / 2\right)$$

where:

r_B = rank similarity for each between-group similarity

r_W = rank similarity for each within-group similarity

$M = n(n\text{-}1)/2$

n = the total number of sample units

The denominator constrains R to the range -1 to 1. Positive values indicate differences among groups. The interpretation is thus similar to the A statistic of MRPP. We know of no formal comparisons of the results of ANOSIM and MRPP.

The Q_b method

Randomization tests can be devised for almost any problem. One difficult group of problems is testing for interactions in multivariate experiments (multiple response variables and complex experimental designs). Unfortunately, this is one of the most common classes of experiments in ecology, although ecologists usually dodge the multivariate nature of their experiments. A method has been proposed, however, that allows randomization tests from such experiments (Pillar and Orloci 1996).

Pillar and Orloci (1996) present a method for partitioning variance in a distance matrix into sums of squares for multiple factors, including interaction terms and multiple contrasts. This promising method has at least one serious shortcoming in that it is limited to Euclidean distance measures (they use the "chord distance;" see Chapter 6). The method is apparently not available in the large statistical packages.

The test criterion is the sum of the squared distances between groups:

$$Q_b = Q_t - Q_w$$

Note that the additivity of these terms depends on Euclidean metric properties of the distance matrix.

The component terms are calculated as follows. The total sum of squares (Q_t) is based on one triangle of the distance matrix, the triangle having $n(n\text{-}1)/2$ terms, each term being a squared distance between two entities j and k:

$$Q_t = \frac{1}{n} \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} d_{kj}^2$$

The within-group sum of squares Q_{wg} is summed across all g groups:

$$Q_w = \sum_{i=1}^{g} Q_{wg}$$

where

$$Q_{wg} = \frac{1}{n_g} \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} d_{jk}^2 \ \delta(j,k,g)$$

The last term, $\delta(j,k,g)$, is an indicator variable that takes the value of one if the entities belong to the same group g; otherwise the value is zero.

The problem is more complex if there are two or more factors. Calculation of the interaction terms is based on a joint classification by the two factors. For example, in a 2 × 2 factorial design, there are four groups defined by the joint classification of the two factors X and Y. The sum of squares is partitioned as follows:

$$Q_{b|X+Y} = Q_{b|X} + Q_{b|Y} + Q_{b|XY}$$

$Q_{b|X}$ and $Q_{b|Y}$ are the sums of squares between groups specific to factors X and Y alone. The last term is the interaction term. The calculations are given in detail in Pillar and Orloci (1996) but are too extensive to repeat here. The results can be summarized in a table much like a traditional ANOVA table.

NPMANOVA

Anderson (2001) solved the problem of using non-Euclidean distance measures (such as Sørensen distance) in multifactor designs. The key to the solution is that "the sum of squared distances between points and their centroid is equal to (and can be calculated directly from) the sum of squared interpoint distances divided by the number of points [Fig. 24.4]... The

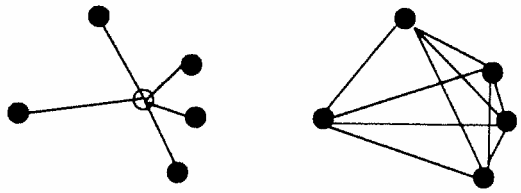

Figure 24.4. The sum of squared distances from points to the centroid (left) can be calculated from the average squared interpoint distance (right).

relationship between distances to centroids and interpoint distances for the Euclidean measure has been known for a long time... What is important is the implication this has for analyses based on non-Euclidean distances. Namely, an additive partitioning of sums of squares can be obtained for any distance measure directly from the distance matrix, without calculating the central locations of groups." Avoiding calculating the position of the centroid is important, because the centroid does not represent the central tendency in a non-Euclidean space.

The total sum of squares of a distance matrix **D** with N rows and N columns is

$$SS_T = \frac{1}{N} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij}^2$$

The residual (within-group) sum of squares for a one-way classification is

$$SS_R = \frac{1}{n} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij}^2 \varepsilon_{ij}$$

where n is the number of observations per group, N is the number of sample units, and $\varepsilon_{ij} = 1$ if i and j are in the same group, but $\varepsilon_{ij} = 0$ if in different groups.

The sum of squares between groups is then $SS_A = SS_T - SS_R$ so we can calculate a pseudo-F-ratio:

$$F = \frac{SS_A / (a-1)}{SS_R / (N-a)}$$

where a is the number of groups. If the distance matrix contains Euclidean distances, then this gives the traditional parametric univariate F ratio.

Evaluating the significance of this F requires a permutation test, even if the distance measure is Euclidean and the variables are normally distributed (Anderson 2001). As with MRPP, the number of possible permutations becomes enormous, even with small samples sizes. Anderson (2001) recommended

at least 1000 permutations for a test with $\alpha = 0.05$ and 5000 permutations for a test with $\alpha = 0.01$.

The permutation is simple with a one-way design: randomly reassign group labels. Nested and factorial designs are more complicated. Equations for these sums of squares are given in Anderson (2001). For a two-factor design (say factors A and B), one calculates the following terms:

SS_A = within-group sum of squares for A, ignoring any influence of B

SS_B = within-group sum of squares for B, ignoring any influence of A

SS_R = residual sum of squares, pooling the sum of squares within groups defined by each of the combinations of factors A and B

SS_{AB} = interaction sum of squares for AB, by subtraction:

$SS_{AB} = SS_T - SS_A - SS_B - SS_R$

If factor B is nested within A, then

$$SS_{B(A)} = SS_T - SS_A - SS_R$$

and there is no interaction term. Calculation of the mean-squared errors and pseudo-F-ratios for the nested and factorial designs are analogous to parametric ANOVA and the formula for the one-way NPANOVA above.

Permutation tests for the nested and factorial designs are, unfortunately, not straightforward. Decisions must be made for each study, based on sample sizes and the nature of the design. Anderson (2001) gave two examples. A factorial design with small sample sizes was handled by unrestricted permutation of the raw data. This weakens the test, but there is little choice if the sample sizes are so small that the number of possible permutations is small. Exact permutation tests were used for a design with three factors, one of them nested. Variability in a higher level factor was tested against the nested factor by permuting the subsamples but not mixing replicates within subsamples.

Like the other permutation tests discussed above, one of the appeals of NPMANOVA for community data is that we can avoid assumptions of linear species responses and normally distributed errors. Anderson (2001) stated that "the only assumption of the test is that the observations (rows of the original data matrix) are exchangeable under a true null hypothesis." Exchangeability requires that the observations (rows) be independent and that they have similar dispersions. ANOSIM, MRPP, and NPMANOVA are all sensitive to different dispersions. Imagine two clouds of points, one much larger than another, but with the same centroid. If you consider these two clouds different, then this sensitivity is an asset; if you consider them the same, then homogeneity of dispersions is another assumption you must meet.