EBIO 5460

Exercise 2

Analyses of Dissimilarity Matrices

In this exercise, we will introduce you to some of the analyses used to understand how predictor variables relate to multivariate dissimilarities and distances.  These analyses are not described in your textbook, Numerical Ecology with R, but further information can be found in the package descriptions and online tutorials (ex. vegan: cc.oulu.fi/~jarioksa/opetus/metodi/vegantutor.pdf), as well as in the primary literature.  While we will expose you to several different tests that do similar things, current evidence indicates that Permutational Multivariate ANOVA (PERMANOVA) is the most robust analysis available at present.  Most of the analyses that follow can be conducted using PERMANOVA, which is generally the preferable analysis. However, it is important to know how to use all these analyses to appreciate differences among them and to help you evaluate studies that use these techniques.

To start, load the following packages:

library(ade4)
library(vegan)

We will primarily use the vegan package for this exercise, which comes with several datasets that provide useful examples for analyzing dissimilarities.  Here, we will mainly use the "dune" dataset, which describes the cover of 30 plant species at 20 sites in Dutch dune meadows.  These data are accompanied by an environmental dataset, "dune.env".  We will focus on understanding how plant community composition relates to two of these environmental variables:

Management = a categorical factor with levels: BF (biological farming), HF (hobby farming), NM (nature conservation management), and SF (standard farming)

A1 = a continuous numeric vector describing the thickness of the soil A1 horizon

To access these data, use:

data(dune)
data(dune.env)

Because we want to analyze community composition in a relative sense, let's relativize the data and focus on using Bray-Curtis dissimilarities:

dune.rel <- decostand(dune, "total")
dune.rel.bray <- vegdist(dune.rel, method="bray")

Analysis of Similarities - ANOSIM

The first analysis we will conduct is ANOSIM, which is used to test whether there is a significant difference between two or more groups.  This analysis is based on ranks of dissimilarities and is therefore similar to NMDS ordination, which we will use later in the course.  Here, we will focus on the environmental factor for dune management:

Management = dune.env[,3]

To conduct the test, we use the function "anosim" from vegan:

dune.anosim1 <- anosim(dune.rel.bray, Management, permutations = 999)
summary(dune.anosim1)

The first term in the model is the dissimilarity matrix, which is followed by the categorical predictor variable and the number of permutations used to evaluate significance.  Alternatively, we can specify a dataframe instead of a dissimilarity matrix, and ask the model to calculate dissimilarities based on these data.  The following model is equivalent to above.

dune.anosim2 <- anosim(dune.rel, Management, permutations = 999, distance = "bray")
summary(dune.anosim2)

The models indicate that there are significant differences in community composition across the types of land management.  In the output for both models, notice that the ANOSIM statistic R is the same for both models, but the significance and quantiles of permutations are slightly different.  Why did this happen?


Multi Response Permutation Procedure - MRPP

Instead of analyzing rank orders of dissimilarities, we might want to use MRPP to directly analyze the dissimilarities.  Similar to ANOSIM, this is another test for evaluating if there are significant differences between two or more treatment groups.  Here we use vegan's "mrpp" function:

dune.rel.mrpp <- mrpp(dune.rel, Management)
dune.rel.mrpp

Notice that the default model uses Euclidean distances instead of Bray-Curtis dissimilarities! This is an important point, because different models do not use the same dissimilarity/distance measure by default.  Thus, it is important to always specify the type of measure you are using. To compare the MRPP results with those from ANOSIM, we have to tell the model to use Bray-Curtis dissimilarities.  The following two models are equivalent, the first of which specifies a dissimilarity matrix, while the second a dataframe:

dune.rel.bray.mrpp1 <- mrpp(dune.rel.bray, Management)
dune.rel.bray.mrpp1

dune.rel.bray.mrpp2 <- mrpp(dune.rel, Management, distance = "bray")
dune.rel.bray.mrpp2

These models use a test statistic called delta, and the results indicate that differences in community composition across the types of land management are significantly different than random.

Another function used in conjunction with MRPP is "meandist", which calculates a matrix of mean within-group dissimilarities. This helps us evaluate differences among groups.

dune.rel.bray.md1 <- meandist(vegdist(dune.rel, method = "bray"), Management)
dune.rel.bray.md1

dune.rel.bray.md2 <- meandist(dune.rel.bray, Management)
dune.rel.bray.md2

Both ANOSIM and MRPP can be strongly affected by different dispersions in dissimilarities across treatment groups. These models may indicate that there are significant differences between groups simply because the dissimilarities are greater within one group than another. A more robust alternative is demonstrated next…

Permutational Multivariate Analysis of Variance - PERMANOVA

The currently preferred analysis for evaluating differences among groups is PERMANOVA. This analysis partitions sums of squares using dissimilarities, evaluating differences in the centroids of groups in multivariate space. The vegan functions "adonis" and "adonis2" are used to compute PERMANOVA in R. The following models are equivalent:

adonis(dune.rel ~ Management, data = dune.env, permutations = 999, method = "bray")

adonis2(dune.rel ~ Management, data = dune.env, permutations = 999, method = "bray")

Here, we are asking the model to analyze community dissimilarities as a function of land management, sourcing the predictor variable from the dataframe "dune.env". Once again, we find that there are differences in community composition across land management practices. Notice that adonis provides $R^2$ values, whereas these are not provided by adonis2. The above models specify dataframes for analysis, but we can alternatively specify a dissimilarity matrix:

adonis2(dune.rel.bray ~ Management, data = dune.env, permutations = 999)

Other advantages of using PERMANOVA are that we can test for interactions between predictor variables, and we can use both categorical and continuous predictor variables. Next we are testing for the statistical interaction between land management and soil depth (A1):

adonis(dune.rel ~ Management*A1, data = dune.env, permutations = 999, method = "bray")

adonis2(dune.rel ~ Management*A1, data = dune.env, permutations = 999, method = "bray")

These models provide the significance for statistical interactions and the main effects. An advantage of adonis2 is that we can also test for overall model fit, using the "by" command:

adonis2(dune.rel ~ Management*A1, data = dune.env, method = "bray", by = NULL)

Now, let us assume for a moment that our data are pseudo-replicated and we are comparing four areas, each corresponding to a type of land management. These data are nested, with multiple observations within each land management block. In this case, we may want to ask whether soil depth has a significant effect and whether this relationship varies across blocks. We can account for nested block effects using the "strata" command. This command tells the model to permute the data within blocks and not across the entire dataset. Compare the unnested and nested models using adonis:

adonis(dune.rel ~ Management*A1, data = dune.env, permutations = 999, method = "bray")

adonis(dune.rel ~ Management*A1, data = dune.env, permutations = 999, method = "bray", strata = dune.env$Management)

Now compare the unnested and nested models using adonis2:

adonis2(dune.rel ~ Management*A1, data = dune.env, permutations = 999, method = "bray")

adonis2(dune.rel ~ Management*A1, data = dune.env, permutations = 999, method = "bray", strata = dune.env$Management)

Why are the nested models different for adonis and adonis2? Even though adonis2 does not return an error message, this function does not use the strata command, so watch out! Blocks need to be specified ahead of time to conduct a nested analysis with adonis2:

perm <- how(nperm = 999)
setBlocks(perm) <- with(dune.env, Management)
adonis2(dune.rel ~ Management*A1, data = dune.env, permutations = perm, method = "bray")


While PERMANOVA tests differences in group means (analogous to MANOVA), a related test called PERMDISP can be used to evaluate homogeneity of group dispersions (analogous to Levene's test for equal variances). The vegan function for this test is "betadisper":

```
dune.disper <- betadisper(dune.rel.bray, Management)
dune.disper
```

Significant differences in homogeneities can be tested using either parametric or permutational tests, and parametric post hoc contrasts can also be investigated:

```
anova(dune.disper)
permutest(dune.disper)
TukeyHSD(dune.disper)
```


Mantel test

A Mantel test is another way to test whether community composition is related to continuous variables.  This test compares two matrices of equal size, calculating the correlation between them.  Here, we use the "mantel" function from vegan to ask whether dune community composition is related to soil depth:

```
A1 = dune.env[,1]
A1
dune.A1.dist <- vegdist(A1, "euclid")
mantel(dune.rel.bray, dune.A1.dist)
```

This test evaluates the Pearson correlation between matrices.  We can also specify other types of correlations, such as the Spearman correlation:

```
mantel(dune.rel.bray, dune.A1.dist, method="spear")
```

Often, we have more than one environmental variable that we want to test in relation to a community matrix.  The following example tests the relationship between plant community composition and an environmental matrix calculated from an environmental data frame (rather than a vector, as above).  Here, we are going to use different datasets found in the vegan package: "varespec" describes cover of 44 species in lichen pastures, while "varechem" describes the corresponding soil characteristics at each of 24 sites.  The next test asks whether community composition is related to soil characteristics:

```
data(varespec)
data(varechem)
veg.rel <- decostand(varespec, "total")
veg.rel.bray <- vegdist(veg.rel, method = "bray")
env.dist <- vegdist(scale(varechem), method = "euclid")
mantel(veg.rel.bray, env.dist)
mantel(veg.rel.bray, env.dist, method="spear")
```

Analysis of Molecular Variance - AMOVA

Another multivariate analysis that may be useful is AMOVA, which tests differences in alleles among populations and groups of populations. Note that PERMANOVA can also be adapted to do the same thing. However, package ade4 contains the function "amova" specifically for this purpose. For this example, we will use data contained in the ade4 package called "humDNAm". These data are for the frequencies of haplotypes of human mitochondrial DNA restriction sites in ten populations across the world. It has three components: distances among haplotypes (the number of differences in restriction sites between two haplotypes), samples (haplotype abundance in a given population), and structures (the name of the continent from which populations were sampled).

```
data(humDNAm)
amovahum <- amova(humDNAm$samples, sqrt(humDNAm$distances), humDNAm$structures)
amovahum
```

To test for differences between populations and regions, we use the "randtest" function:

```
randtesthum <- randtest(amovahum, permutations = 999)
randtesthum
plot(randtesthum)
```

_____

Application

Now that you have been exposed to a variety of multivariate tests, apply at least two of the above analyses to your own data. Also, see how your results change when using two different types of dissimilarities or distances.

If you do not have a dataset to work with, then feel free to use the additional data included with this assignment, named "CarrizoVeg09.csv" and "CarrizoSoil.csv". The first dataset is the same as was supplied for Exercise 1: plant cover data at each of 20 sites in annual grassland across the Carrizo Plain, CA, in 2009. The second dataset includes 13 soil variables (continuous data) and a categorical variable called "PriorGraze" that corresponds to the history of cattle grazing at each site (ie. grazed or ungrazed). Try conducting a PERMANOVA evaluating the interaction between grazing history and a soil variable. What happens to your results if you assume that sites are nested within grazing history?

Also, if you would like to learn more about differences among these tests, check out the supplemental reading included with this exercise:

Anderson, MJ, and DCI Walsh. 2013. PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? Ecological Monographs 83:557-574.