OVERVIEW OF SELECTED MULTIVARIATE ANALYSES IN ECOLOGY

Ecologists are all too often faced with complexity. Indeed, ecological systems are comprised of a dazzling diversity of species and an overwhelming array of environmental factors, all of which interact in multiple and complex ways. Indeed, the species and ecological assemblages we observe today are the manifestation of these complex interactions as they play out in the proverbial ecological theater and evolutionary play\(^1\). Given the complex and multi-variable nature of ecological systems, it’s not surprising that ecologists have turned to multivariate approaches to help them tease apart these relationships. If properly used and understood, these methods are tremendously powerful tools in exploring underlying ecological patterns as well as testing multi-variable hypotheses.

The purpose of this module of Bio 221B is to introduce you to some of the types of multivariate methods in ecology and to give you a sense about how they work so you can decide whether they might be useful for you in your research. Although I will introduce you to the most common approaches, I also want your research interests to guide where we place our emphasis. Because of the sophistication of these methods and the limited time we have together, we won’t be able to go into a lot of theoretical or mathematical (thank goodness, right?) detail. However, my hope is that you will gain a basic understanding about the methods and their potential applications. In short, the course objectives are as follows:

1. Become familiar with selected multivariate analytical approaches and how they can be applied to answer important questions in ecology, evolution, and conservation.
2. Gain experience sampling multi-variable data and applying different kinds of multivariate analyses to them.
3. Gain experience presenting and interpreting the results of selected multivariate analyses using appropriate statistical and graphical methods.

So where do we start? Probably the best place to start is to clarify just what it is we’re talking about when referring to multivariate data in ecology. It’s simple really: a multivariate dataset is comprised of data collected on more than one variable for each sampling unit. In community ecology, for example, a data matrix often has taxa (usually species) as rows and samples as columns (Table 1) or vice versa. In community ecology, the term "sample" refers to the basic unit of observation. In studies of vegetation, the sample might be a quadrat (or plot), relevé, or transect. Samples in animal ecology can take the form of traps, seine sweeps, survey routes, etc. On the other hand, landscape scale studies might use grid cells on a GIS map or political units as samples.

Table 1. Example of a multivariate community data matrix. Rows are species (bryophytes in this case) and columns are locations of the samples. Data are measures of abundance.

<table>
<thead>
<tr>
<th>SPECIES</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th>S6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leucobryum albidum</td>
<td>3.1</td>
<td>0</td>
<td>0</td>
<td>44.6</td>
<td>0</td>
<td>35.9</td>
</tr>
<tr>
<td>Anomodon attenuatus</td>
<td>0.9</td>
<td>17.6</td>
<td>26.4</td>
<td>0</td>
<td>41.2</td>
<td>0</td>
</tr>
<tr>
<td>Isopterygium tenerum</td>
<td>16.1</td>
<td>4.6</td>
<td>2.2</td>
<td>30.9</td>
<td>1.4</td>
<td>26.8</td>
</tr>
<tr>
<td>Sematophyllum adnatum</td>
<td>11</td>
<td>7.9</td>
<td>6.2</td>
<td>6.4</td>
<td>3.9</td>
<td>7.4</td>
</tr>
<tr>
<td>Platygryrium repens</td>
<td>29.4</td>
<td>5.5</td>
<td>9.5</td>
<td>0</td>
<td>2.1</td>
<td>3.2</td>
</tr>
</tbody>
</table>

In general, community data matrices are made up of species abundance measures of some kind (Table 1). Abundance is a broad term and can refer to any of a variety of measures, including density, frequency, biomass, cover, or incidence (presence/absence) of a species. Which abundance measure is used depends on the focal taxa and the question(s) being asked. A shrub, mycorrhizal, or small mammal community will each require different abundance measures because of their dimensions, mobility, and how they occupy their habitat. We’ll talk more about that later. In addition, multivariate datasets can also include one or more sets of environmental variables (e.g., elevation, time since fire, soil moisture, predation risk) as possible explanatory variables for species abundances, and which can serve to test hypothesized species-environment relationships.

**Exploratory vs. Hypothesis Testing Methods**

Broadly speaking, multivariate methods are either exploratory – in that they allow you to search for underlying patterns in your data – or they involve testing of *a priori* hypotheses. Both have their place in ecological data analysis. For example, you may suspect but not know what environmental factors are responsible for the distribution and abundance of species in a community. Exploratory approaches allow you to identify plausible distribution/abundance patterns in your data and based on these patterns generate meaningful hypotheses. Hypothesis testing at the community level suggests that you have some knowledge of the key factors involved but you wish to explore possible mechanisms for observed patterns. Thus, you might apply a treatment to your community assemblage based on this prior knowledge and measure its effect on the structure of the component species. Indeed, these two approaches are complementary and often used in conjunction to answer key questions in ecology. Hybrid approaches exist as well and can also be explored. We will examine both approaches.

**Selected Multivariate Methods**

*Cluster analysis* is an exploratory technique that is commonly used to classify ecological groups or interpret patterns of species composition and abundance. Cluster analysis aims to group (or “cluster”) objects (e.g., sample plots) that are similar to each other but which differ among clusters. This method is commonly used in systematics to generate artificial phylogenies, bioinformatics to build groups of genes with related expression or functional attributes, and plant ecology to identify similar species assemblages or for vegetation mapping.

*Ordination* is an exploratory multivariate technique that involves the systematic “ordering” of objects such that similar objects are near to one another and dissimilar objects are far from one another. These relationships among objects are depicted along a set of axes and characterized graphically and numerically. With respect to ecological data, ordination separates out species and/or sample units along known or presumed environmental gradients (we’ll talk more about environmental gradients below) such that similar samples or species are near each other and dissimilar ones far apart. In this sense, ordination is a data reduction technique that simplifies the multi-variable dataset into a few axes that are presumed to represent real underlying gradients.

*Multivariate analysis of variance (MANOVA)* is a hypothesis testing method that relies on a testable hypothesis coupled with statistical inference. A MANOVA tests the effect of one or more independent variables on more than one dependent variable. If you’re familiar with ANOVA, MANOVA is simply an extension of ANOVA to multiple dependent variables. The technique is quite useful for testing hypotheses in community ecology because of the intrinsically multivariate nature of community data. As a particularly powerful use of MANOVA in community ecology, we will combine ordination with MANOVA as a means of first distilling out the fundamental patterns in the data (ordination) and then testing the effect of a treatment on these principal patterns.