Topics:
1. Multivariate statistics
2. Finding groups - cluster analysis
3. Testing/describing group differences
4. Unconstrained ordination
5. Constrained ordination

Landscape of Statistical Methods...
The Landscape

The basic statistical model:
\[ Y = \text{deterministic part} + \text{stochastic part} \]

- Univariate
  - Multivariate
    - Linear
    - Nonlinear
    - Smoothed

- Distribution
- Heterogeneity
- Autocorrelation
- Multiple levels
- Random noise
Why do we need multivariate statistics?

- Reflect more accurately the true multidimensional nature of natural systems
- Provide a way to handle large data sets with large numbers of variables
- Provide a way of summarizing redundancy in large data sets
- Provide rules for combining variables in an "optimal" way

- Provide a means of detecting and quantifying truly multivariate patterns that arise out of the correlational structure of the variable set
- Provide a means of exploring complex data sets for patterns and relationships from which hypotheses can be generated and subsequently tested experimentally
Landscape of Statistical Methods...
What is multivariate statistics?

y = x1 + x2 + ... xj → Regression
Analysis of Variance
Contingency Tables, etc.

y1 + y2 + ... yi = x → Multivariate ANOVA
Discriminant Analysis
CART, MRPP, MANTEL

y1 + y2 + ... yi = x1 + x2 + ... xj → Canonical Corr. Analysis
Constrained ordination

y1 + y2 + ... yi → Unconstrained ordination
Cluster Analysis

Multivariate Statistics

Landscape of Statistical Methods...
Multivariate methods

- Finding groups (Cluster analysis)
- Testing for groups (e.g., MRPP, MANTEL)
- Discriminating among groups (e.g., DA, ISA, mCART)
- Unconstrained ordination (e.g., PCA, CA, NMDS)
- Constrained ordination (e.g., RDA, CCA, CAPS)

• Large family of techniques with similar goals; operating on data sets for which pre-specified, well-defined groups do "not" exist; characteristics of the data are used to assign entities into artificial groups
Can we organize sampling entities (e.g., sites) into discrete classes, such that within-group similarity is maximized and among-group similarity is minimized?

<table>
<thead>
<tr>
<th>Sites</th>
<th>Species</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

Nonhierarchical clustering:

- NHC methods merely assign each entity to a cluster, placing similar entities together in order to maximize within-cluster homogeneity.
Hierarchical clustering:

- HC methods combine similar entities into classes or groups and arrange these groups into a hierarchy that reveals relationships among the entities classified.

**Finding groups** – cluster analysis

**Mulivariate methods**

- Finding groups (Cluster analysis)
- Testing for groups (e.g., MRPP, MANTEL)
- Discriminating among groups (e.g., DA, ISA, mCART)
- Unconstrained ordination (e.g., PCA, CA, NMDS)
- Constrained ordination (e.g., RDA, CCA, CAPS)

- Family of different methods for testing and/or describing differences among pre-specified, well-defined groups based on a set of discriminating variables.
Discriminating among groups

- Are pre-defined groups of entities (e.g., species) significantly different from each other and, if so, how do they differ?

<table>
<thead>
<tr>
<th>Id</th>
<th>Species</th>
<th>Ccov</th>
<th>Snags</th>
<th>CHgt</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>80</td>
<td>1.2</td>
<td>35</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>75</td>
<td>0.5</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>72</td>
<td>2.8</td>
<td>28</td>
</tr>
<tr>
<td>...</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>B</td>
<td>35</td>
<td>3.3</td>
<td>15</td>
</tr>
<tr>
<td>32</td>
<td>B</td>
<td>75</td>
<td>4.1</td>
<td>25</td>
</tr>
<tr>
<td>60</td>
<td>B</td>
<td>15</td>
<td>5.0</td>
<td>3</td>
</tr>
<tr>
<td>...</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>C</td>
<td>5</td>
<td>2.1</td>
<td>5</td>
</tr>
<tr>
<td>62</td>
<td>C</td>
<td>8</td>
<td>3.4</td>
<td>2</td>
</tr>
<tr>
<td>90</td>
<td>C</td>
<td>25</td>
<td>0.6</td>
<td>15</td>
</tr>
</tbody>
</table>

Testing for group differences

- Are groups significantly different? (How valid are the groups?)
  - Multivariate Analysis of Variance (MANOVA)
  - Multi-Response Permutation Procedures (MRPP)
  - Analysis of Group Similarities (ANOSIM)
  - Mantel’s Test (MANTEL)

- How do groups differ? (Which variables best distinguish among the groups?)
  - Discriminant Analysis (DA)
  - Classification and Regression Trees (CART)
  - Logistic Regression (LR)
  - Indicator Species Analysis (ISA)
Landscape of Statistical Methods...
Mulivariate methods

- Finding groups (Cluster analysis)
- Testing for groups (e.g., MRPP, MANTEL)
- Discriminating among groups (e.g., DA, ISA, mCART)
- Unconstrained ordination (e.g., PCA, CA, NMDS)
- Constrained ordination (e.g., RDA, CCA, CAPS)

- A family of different methods for organizing sampling entities (e.g., species, sites, observations, etc.) along continuous gradients based on a set of interdependent variables

Landscape of Statistical Methods...
Unconstrained ordination

- Can we organize entities (e.g., sites) along one or more gradients based on their relationships among the interdependent variables?

<table>
<thead>
<tr>
<th>Sites</th>
<th>Species A</th>
<th>Species B</th>
<th>Species C</th>
<th>Species D</th>
<th>Species E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 (1)</td>
<td>5 (1)</td>
<td>1 (1)</td>
<td>10 (4)</td>
<td>10 (4)</td>
</tr>
<tr>
<td>2</td>
<td>2 (3)</td>
<td>8 (3)</td>
<td>4 (3)</td>
<td>12 (6)</td>
<td>20 (6)</td>
</tr>
<tr>
<td>3</td>
<td>8 (6)</td>
<td>20 (6)</td>
<td>10 (6)</td>
<td>1 (2)</td>
<td>3 (2)</td>
</tr>
<tr>
<td>4</td>
<td>4 (5)</td>
<td>11 (5)</td>
<td>8 (5)</td>
<td>11 (5)</td>
<td>14 (5)</td>
</tr>
<tr>
<td>5</td>
<td>1 (2)</td>
<td>6 (2)</td>
<td>2 (2)</td>
<td>2 (3)</td>
<td>6 (3)</td>
</tr>
<tr>
<td>6</td>
<td>3 (4)</td>
<td>10 (4)</td>
<td>6 (4)</td>
<td>0 (1)</td>
<td>0 (1)</td>
</tr>
</tbody>
</table>
Unconstrained ordination

<table>
<thead>
<tr>
<th>Obs</th>
<th>Canopy Cover</th>
<th>Snag Density</th>
<th>Canopy Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>80</td>
<td>1.2</td>
<td>35</td>
</tr>
<tr>
<td>2</td>
<td>75</td>
<td>0.5</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>72</td>
<td>0.8</td>
<td>28</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>N</td>
<td>25</td>
<td>0.6</td>
<td>15</td>
</tr>
</tbody>
</table>

PC1 = .8x1 - .4x2 + .1x3
PC2 = -.1x1 - .1x2 + .9x3

Landscape of Statistical Methods...

Unconstrained ordination

2d ordi bubble plot

3d ordi scatter plot
Unconstrained ordination

- Principal components analysis (PCA)
- Factor analysis (FA)
- Multidimensional scaling (MDS/PCO)
- **ML**-Unconstrained linear ordination (ULO)
- Correspondence analysis (CA & DCA)
- **ML**-Unconstrained quadratic ordination (UQO)
- **ML**-Unconstrained additive ordination (UAO)
- Nonmetric multidimensional scaling (NMDS)
Landscape of Statistical Methods...
Mulivariate methods

- Finding groups (Cluster analysis)
- Testing for groups (e.g., MRPP, MANTEL)
- Discriminating among groups (e.g., DA, ISA, mCART)
- Unconstrained ordination (e.g., PCA, CA, NMDS)
- Constrained ordination (e.g., RDA, CCA, CAPS)

A family of different methods for extending unconstrained ordination in which the solution is constrained to be expressed by ancillary variables.

Landscape of Statistical Methods...
Constrained ordination

- Can bird community patterns be explained by measured environmental variables?
The triplot displays the major patterns in the species data with respect to the environmental variables.

Tri = (1) Samples  
(2) Species  
(3) Environment
Landscape of Statistical Methods...  
Constrained ordination

Variance partioning

- Constrained analysis of principal coordinates (CAP)
- Redundancy analysis (RDA)
- Canonical correspondence analysis (CCA)