### Discrimination Among Groups

<table>
<thead>
<tr>
<th>Id</th>
<th>Species</th>
<th>Canopy Cover</th>
<th>Snag Density</th>
<th>Canopy Height</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>

Are groups significantly different? (How valid are the groups?)
- Multivariate Analysis of Variance [(NP)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]
Tests of Among-Group Differences

2-group bird guilds example

<table>
<thead>
<tr>
<th>ODS</th>
<th>SPECIES</th>
<th>GTOTAL</th>
<th>LTOTAL</th>
<th>TOTOTAL</th>
<th>MTOTAL</th>
<th>OTOTAL</th>
<th>SNAGM1</th>
<th>SNAGM2</th>
<th>SNAGM3</th>
<th>SNAGM4</th>
<th>FHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AMRO</td>
<td>15.31</td>
<td>31.42</td>
<td>64.28</td>
<td>20.71</td>
<td>47.14</td>
<td>0.00</td>
<td>0.28</td>
<td>0.14</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>2</td>
<td>BHGR</td>
<td>5.76</td>
<td>24.77</td>
<td>73.18</td>
<td>22.95</td>
<td>61.59</td>
<td>0.00</td>
<td>0.00</td>
<td>1.09</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>3</td>
<td>BRGR</td>
<td>4.78</td>
<td>64.13</td>
<td>30.85</td>
<td>12.03</td>
<td>63.60</td>
<td>0.44</td>
<td>0.44</td>
<td>2.08</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>4</td>
<td>CBCH</td>
<td>3.08</td>
<td>58.52</td>
<td>39.69</td>
<td>15.47</td>
<td>62.19</td>
<td>0.31</td>
<td>0.28</td>
<td>1.52</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>5</td>
<td>DEJU</td>
<td>13.90</td>
<td>60.78</td>
<td>36.50</td>
<td>13.81</td>
<td>62.89</td>
<td>0.23</td>
<td>0.31</td>
<td>1.23</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>19</td>
<td>WIWR</td>
<td>8.05</td>
<td>41.09</td>
<td>55.00</td>
<td>18.62</td>
<td>53.77</td>
<td>0.09</td>
<td>0.18</td>
<td>0.81</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

Two-group clustering solution resulting from agglomerative hierarchical clustering (hclust) using Euclidean distance on standardized variables and Ward’s linkage method.
Tests of Among-Group Differences

Permutational Multivariate Analysis of Variance

(NP-MANOVA)

- Nonparametric procedure for testing the hypothesis of no difference between two or more groups of entities based on the analysis and partitioning sums of square distances (Anderson 2001).

![Fig. 2. The sum of squared distances from individual points to their group centroid is equal to the sum of squared inter-point distances divided by the number of points.](image)

Tests of Among-Group Differences

Permutational (nonparametric) MANOVA

- Calculate distance matrix (any distance metric can be used)
- Calculate average distance among all entities ($SS_T$)
- Calculate average distance among entities within groups ($SS_W$)
- Calculate average distance among groups ($SS_A = SS_T - SS_W$)
- Calculate $F$-ratio

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

\[
SS_W = \frac{1}{n} \sum_{i=1}^{N} \sum_{j=i+1}^{N} d_{ij}^2 \epsilon_{ij}
\]

\[
F = \frac{SS_A/(a-1)}{SS_W/(N-a)}
\]

$N = \text{total number of items} \quad a = \text{number of groups} \quad n = \text{number of items per group}$

$d_{ij} = \text{distance between } i^{th} \text{ and } j^{th} \text{ entity} \quad \epsilon_{ij} = 1 \text{ if in same group; } 0 \text{ otherwise}$
Tests of Among-Group Differences
Permutational (nonparametric) MANOVA

- Determine the probability of an F-ratio this large or larger through Monte carlo permutations.
  - Permutations involve randomly assigning sample observations to groups (within strata if nested design).
  - The significance test is simply the fraction of permuted F-ratios that are greater than the observed F-ratio.
Tests of Among-Group Differences

Permutational (nonparametric) MANOVA

2-group bird guilds example

\[
\delta = \sum_{i=1}^{g} C_i \bar{d}_i
\]

\[
C_i = \frac{n_i}{N}
\]

\* other options for calculating \( C_i \) exit

Conclude that the differences between the two bird species clusters is statistically significant and that 43% of the “variance” is accounted for by group differences.

Multi-Response Permutation Procedures (MRPP)

Nonparametric procedure for testing the hypothesis of no difference between two or more groups of entities based on permutation test of among- and within-group dissimilarities (Mielke 1984, 1991).

- Calculate distance matrix (Euclidean distance generally recommended, although proportional city-block measures often used with community data).
- Calculate average distance in each group \( = \bar{d}_i \)
- Calculate \( \delta \) (the weighted mean within-group distance) for \( g \) groups.


### Tests of Among-Group Differences

**Multi-Response Permutation Procedures (MRPP)**

#### 2-group bird guilds example

<table>
<thead>
<tr>
<th>Group</th>
<th>AMRO</th>
<th>BHGR</th>
<th>RUHU</th>
<th>SOSP</th>
<th>STJA</th>
<th>SWTH</th>
<th>WEFL</th>
<th>WIWA</th>
<th>WIWR</th>
<th>BRCR</th>
<th>CBCH</th>
<th>DEJU</th>
<th>EVGR</th>
<th>GCKI</th>
<th>GRJA</th>
<th>HAFL</th>
<th>HAWO</th>
<th>HEWA</th>
<th>VATH</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 AMRO</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 BHGR</td>
<td>6.83</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 RUHU</td>
<td>3.77</td>
<td>7.46</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 SOSP</td>
<td>5.60</td>
<td>8.97</td>
<td>3.41</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 STJA</td>
<td>4.88</td>
<td>7.26</td>
<td>3.14</td>
<td>4.95</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 SWTH</td>
<td>3.11</td>
<td>4.70</td>
<td>3.34</td>
<td>5.57</td>
<td>3.54</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 WEFL</td>
<td>3.61</td>
<td>6.57</td>
<td>3.27</td>
<td>4.85</td>
<td>2.66</td>
<td>2.51</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 WIWA</td>
<td>5.13</td>
<td>6.68</td>
<td>3.49</td>
<td>5.42</td>
<td>3.46</td>
<td>3.45</td>
<td>3.44</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 WIWR</td>
<td>3.71</td>
<td>6.78</td>
<td>2.40</td>
<td>4.25</td>
<td>2.02</td>
<td>2.49</td>
<td>1.50</td>
<td>2.56</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 BRCR</td>
<td>9.15</td>
<td>9.29</td>
<td>7.44</td>
<td>7.65</td>
<td>5.98</td>
<td>7.52</td>
<td>6.66</td>
<td>5.18</td>
<td>6.18</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 CBCH</td>
<td>7.04</td>
<td>7.68</td>
<td>5.14</td>
<td>5.66</td>
<td>3.84</td>
<td>5.28</td>
<td>4.50</td>
<td>3.14</td>
<td>3.85</td>
<td>2.70</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 DEJU</td>
<td>6.98</td>
<td>8.20</td>
<td>4.87</td>
<td>5.35</td>
<td>4.11</td>
<td>5.61</td>
<td>4.94</td>
<td>3.49</td>
<td>4.10</td>
<td>3.11</td>
<td>1.82</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 EVGR</td>
<td>10.73</td>
<td>10.57</td>
<td>8.53</td>
<td>8.57</td>
<td>7.70</td>
<td>9.14</td>
<td>8.66</td>
<td>6.43</td>
<td>7.96</td>
<td>4.19</td>
<td>4.82</td>
<td>5.24</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 GRJA</td>
<td>10.22</td>
<td>9.71</td>
<td>8.48</td>
<td>9.00</td>
<td>8.00</td>
<td>8.93</td>
<td>8.25</td>
<td>7.85</td>
<td>8.23</td>
<td>5.30</td>
<td>5.88</td>
<td>6.20</td>
<td>8.06</td>
<td>5.53</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 HAFL</td>
<td>11.21</td>
<td>10.18</td>
<td>9.15</td>
<td>9.32</td>
<td>8.07</td>
<td>9.44</td>
<td>9.00</td>
<td>6.97</td>
<td>8.48</td>
<td>3.98</td>
<td>5.10</td>
<td>5.46</td>
<td>3.98</td>
<td>2.96</td>
<td>7.29</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 HAWO</td>
<td>7.49</td>
<td>7.69</td>
<td>5.59</td>
<td>5.83</td>
<td>4.54</td>
<td>5.70</td>
<td>5.38</td>
<td>4.49</td>
<td>4.67</td>
<td>4.19</td>
<td>3.29</td>
<td>3.49</td>
<td>6.30</td>
<td>4.30</td>
<td>7.36</td>
<td>6.02</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 VATH</td>
<td>7.08</td>
<td>8.41</td>
<td>5.83</td>
<td>5.93</td>
<td>4.19</td>
<td>5.97</td>
<td>4.63</td>
<td>5.27</td>
<td>4.68</td>
<td>4.78</td>
<td>4.28</td>
<td>4.51</td>
<td>7.12</td>
<td>5.10</td>
<td>7.44</td>
<td>7.07</td>
<td>4.39</td>
<td>7.89</td>
<td>0</td>
</tr>
</tbody>
</table>

\[
\overline{d}_i = 4.34 \quad c_1 = \frac{36}{81} = .444 \\
\overline{d}_2 = 5.24 \quad c_2 = \frac{45}{81} = .556 \\
\delta = 4.84
\]

Note: only with-group dissimilarities are used.

---

### Tests of Among-Group Differences

**Multi-Response Permutation Procedures (MRPP)**

- Determine the **probability** of a delta this small or smaller through Monte Carlo permutations.
  - Permutations involve randomly assigning sample observations to groups.
  - The significance test is simply the fraction of permuted deltas that are less than the observed delta, with a small sample correction.
Tests of Among-Group Differences
Multi-Response Permutation Procedures (MRPP)

2-group bird guilds example

- Conclude that two clusters (bird species with similar niches) differ significantly in terms of the measured habitat variables.

Tests of Among-Group Differences
Multi-Response Permutation Procedures (MRPP)

- Determine the effect size independent of sample size (chance-corrected within-group agreement, $A$).

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

- The statistic $A$ is given as a descriptor of within-group homogeneity compared to the random expectation

  $A = 1$ when all items are identical within groups
  $A = 0$ when within-group heterogeneity equals expectation by chance
  $A < .1$ common in ecology
  $A > .3$ is fairly high in ecology (but see simulation results later)

Note: statistical significance (small p-value) may result even when the “effect size” ($A$) is small, if the sample size is large.
Conclude that the differences between the two bird species clusters is statistically significant and that the difference is moderately large and therefore probably ecologically significant as well.
Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

\[ e=0 \quad \text{Sample size} \quad e=1 \]
Tests of Among-Group Differences
Multi-Response Permutation Procedures (MRPP)

$p=1$ \hspace{1cm} \text{Sample size} \hspace{1cm} p=100$

Tests of Among-Group Differences
Multi-Response Permutation Procedures (MRPP)

$p=1, \ e=1$ \hspace{1cm} \text{Sample size} \hspace{1cm} p=10, \ e=10$
Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

Effect size

MRPP Simulation: $n = 100; p = 1; \epsilon = 0 - 50$

Effect size

MRPP Simulation: $n = 100; p = 10; \epsilon = 0 - 50$

Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

Effect size

MRPP Simulation: $n = 100; p = 1; \epsilon = 0 - 50$

Effect size

MRPP Simulation: $n = 20; p = 1; \epsilon = 0 - 50$
Tests of Among-Group Differences
Multi-Response Permutation Procedures (MRPP)

c=1 \quad \text{Noise Ratio} \quad c=10

Tests of Among-Group Differences
Multi-Response Permutation Procedures (MRPP)
n=100 \quad \text{Noise Ratio} \quad n=20
Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

\( n=100 \quad \text{Noise Ratio} \quad n=10 \)

Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

Column standardized \quad Data Standardization (sample size) \quad Raw (same range)
Tests of Among-Group Differences
Multi-Response Permutation Procedures (MRPP)

Column standardized  Data Standardization  Raw (same range)

Tests of Among-Group Differences
Multi-Response Permutation Procedures (MRPP)

Column standardized  Data Standardization  Raw (same range)
Test statistic is unreliable for \( n<20 \), and for larger sample sizes (e.g., \( n=40 \)) if the noise ratio is greater than \(~10:1\).

Significance (p-value) of test statistic insensitive to degree of noise (nondiscriminating variables); i.e., a single effective discriminating variable will produce a significant test result for \( n>20 \), as long as noise ratio \(<~10:1\).

Effect size (A) sensitive to both magnitude of effect on at least one discriminating variable and the noise ratio (i.e., ratio of nondiscriminating variables to true discriminators); noise ratio relatively more important.

Data standardization (col z-score) has moderate effect on effect size (A).

Simulation Results Summary

Nonparametric procedure for testing the hypothesis of no difference between two or more groups of entities based on permutation test of among- and within-group similarities (Clark 1993).

- Calculate dissimilarity matrix.
- Calculate rank dissimilarities (smallest dissimilarity is given a rank of 1).
- Calculate mean among- and within-group rank dissimilarities.
- Calculate test statistic \( R \) (an index of relative within-group dissimilarity).

\[
R = \frac{-r_A - r_W}{M/2}
\]

\( M = N(N-1)/2 \)

= number of sample pairs
**Tests of Among-Group Differences**

*Analysis of Group Similarities (ANOSIM)*

R is interpreted like a correlation coefficient and is a measure of ‘effect size’, like $A$ in MRPP:

- $R = 1$ when all pairs of samples within groups are more similar than to any pair of samples from different groups.
- $R = 0$ expected value under the null model that among- and within-group dissimilarities are the same on average.
- $R < 0$ numerically possible but ecologically unlikely.

$$R = \frac{\bar{r}_A - \bar{r}_W}{\sqrt{\frac{M}{2}}}$$

$M = N(N-1)/2 = $ number of sample pairs

---

**Tests of Among-Group Differences**

*Analysis of Group Similarities (ANOSIM)*

2-group bird guilds example

$\bar{r}_a = 111.50$  \hspace{1cm} $\bar{r}_w = 57.68$

$R = 0.629$

Note: both within- and among-group dissimilarities are used.
Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

- Determine the probability of an $R$ this large or larger through Monte carlo permutations.
  - Permutations involve randomly assigning sample observations to groups.
  - The significance test is simply the fraction of permuted $R$’s that are greater than the observed $R$.

![R Permutation Distribution](image)

$-1 \leq R \leq 1$

Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

2-group bird guilds example

Call:
anosim(dis = y.eucl, grouping = grp)

Dissimilarity: euclidean

ANOSIM statistic $R$: 0.6234
Significance: $< 0.001$

Based on 1000 permutations

Empirical upper confidence limits of $R$: 90% 95% 97.5% 99%
0.0887 0.1397 0.1536 0.2573

Dissimilarity ranks between and within classes:
0% 25% 50% 75% 100% N
Between 15 80.25 121.5 148.75 171 90
1 1 16.75 29.5 74.75 145 36
2 2 36.00 63.0 103.00 136 45

- Conclude that two clusters differ significantly in terms of the measured habitat variables.
Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

Simulation Study

n=20, p=1, e=1

\[
\begin{array}{c|cc|c}
\text{grp} & y & l \\
[1] & 0.78172641 & 0.758466880 \\
[2] & 0.73212922 & 0.89764596 \\
[3] & 0.22718943 & 0.012008126 \\
[4] & 0.60753563 & 0.448890944 \\
[5] & 0.03840249 & 0.225436849 \\
[6] & 0.98346911 & 0.006262736 \\
[7] & 0.39250403 & 0.397803540 \\
[8] & 0.03179824 & 0.316845145 \\
[9] & 0.63247471 & 0.410760431 \\
[10] & 0.73152952 & 0.178422687 \\
[11] & 1.10265749 & 0.318450545 \\
[12] & 1.10405233 & 0.61288473 \\
[13] & 1.72804012 & 0.631067666 \\
[14] & 1.37381699 & 0.80144922 \\
[15] & 1.19468304 & 0.65631478 \\
[16] & 1.52813037 & 0.47549311 \\
[17] & 1.30526785 & 0.656755662 \\
[18] & 1.29030220 & 0.209625201 \\
[19] & 1.36418721 & 0.099046791 \\
[20] & 1.37266570 & 0.253065173 \\
\end{array}
\]

n=10, p=1, e=1

\[
\begin{array}{c|cc|c}
\text{grp} & y & l \\
[1] & 0.24804322 & 0.67330864 \\
[2] & 0.3473905 & 0.232588895 \\
[3] & 0.5006673 & 0.55771587 \\
[4] & 0.0891539 & 0.79033950 \\
[5] & 0.4879530 & 0.3760798 \\
[6] & 1.8644004 & 0.69309441 \\
[7] & 1.9746505 & 0.81360903 \\
[8] & 1.1038381 & 0.57232166 \\
[9] & 1.0238596 & 0.79687679 \\
[10] & 1.3103284 & 0.036707041 \\
\end{array}
\]

n=10, p=3, e=1

\[
\begin{array}{c|cc|c}
\text{grp} & y & l \\
[1] & 0.2384904 & 0.9270140 & 0.1637725 & 0.04744548 \\
[2] & 0.5240702 & 0.4438856 & 0.8387992 & 0.73323831 \\
[3] & 0.4846718 & 0.0880790 & 0.5378692 & 0.66965015 \\
[4] & 0.7693610 & 0.1904204 & 0.8995969 & 0.26179045 \\
[5] & 0.1217240 & 0.5295393 & 0.6956824 & 0.79223646 \\
[6] & 1.7356405 & 0.7803704 & 0.7213608 & 0.17149252 \\
[7] & 1.4519720 & 0.1926339 & 0.8744577 & 0.48857552 \\
[8] & 1.8991477 & 0.8965194 & 0.5275195 & 0.17056135 \\
[9] & 1.7226480 & 0.3448639 & 0.0082422 & 0.2937219 \\
[10] & 1.1669151 & 0.3416431 & 0.9159500 & 0.63253065 \\
\end{array}
\]

Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

c=0

Sample size e=1
Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

$e=1$  Sample size  $e=2$

Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

$e=1$  Sample size  $e=10$
Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

\( e=1 \) \hspace{1cm} \text{Sample size} \hspace{1cm} \( e=10 \)

\begin{align*}
\text{ANOSIM Simulation: } n=10-100 \ ; \ p=1 \ ; \ e=1 \\
\text{Effect size} &\quad \text{P-value} \\
\end{align*}

\begin{align*}
\text{ANOSIM Simulation: } n=10-100 \ ; \ p=10 \ ; \ e=1 \\
\text{Effect size} &\quad \text{P-value} \\
\end{align*}

\begin{align*}
\text{ANOSIM Simulation: } n=10-100 \ ; \ p=1 \ ; \ e=10 \\
\text{Effect size} &\quad \text{P-value} \\
\end{align*}

\begin{align*}
\text{ANOSIM Simulation: } n=10-100 \ ; \ p=10 \ ; \ e=10 \\
\text{Effect size} &\quad \text{P-value} \\
\end{align*}

Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

\( p=1 \) \hspace{1cm} \text{Sample size} \hspace{1cm} \( p=10 \)

\begin{align*}
\text{ANOSIM Simulation: } n=10-100 \ ; \ p=1 \ ; \ e=1 \\
\text{Effect size} &\quad \text{P-value} \\
\end{align*}

\begin{align*}
\text{ANOSIM Simulation: } n=10-100 \ ; \ p=10 \ ; \ e=1 \\
\text{Effect size} &\quad \text{P-value} \\
\end{align*}

\begin{align*}
\text{ANOSIM Simulation: } n=10-100 \ ; \ p=1 \ ; \ e=10 \\
\text{Effect size} &\quad \text{P-value} \\
\end{align*}

\begin{align*}
\text{ANOSIM Simulation: } n=10-100 \ ; \ p=10 \ ; \ e=10 \\
\text{Effect size} &\quad \text{P-value} \\
\end{align*}
Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

\( p = 1 \) \hspace{1cm} \text{Sample size} \hspace{1cm} \text{\( p = 100 \)}

\( p = 1, e = 1 \) \hspace{1cm} \text{Sample size} \hspace{1cm} \text{\( p = 10, e = 10 \)}
Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

Effect size

p=1

Effect size

p=10

n=20

Effect size

n=100

Effect size

n=20
Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

$e=1$    Noise Ratio    $e=10$

$\text{ANOSIM Simulation: } n = 100; p = 1-20; e = 1$

$\text{ANOSIM Simulation: } n = 100; p = 1-20; e = 10$

$\text{Tests of Among-Group Differences}$

$\text{Analysis of Group Similarities (ANOSIM)}$

$n=100$    Noise Ratio    $n=20$

$\text{ANOSIM Simulation: } n = 100; p = 1-20; e = 1$

$\text{ANOSIM Simulation: } n = 20; p = 1-20; e = 1$
Tests of Among-Group Differences
Analysis of Group Similarities (ANOSIM)

n=100  Noise Ratio  n=10

Tests of Among-Group Differences
Analysis of Group Similarities (ANOSIM)

Column standardized  Data Standardization  Raw (same range)  (sample size)

ANOSIM Simulation: n = 10-100 ; p = 1 ; e = 10

ANOSIM Simulation: n = 10-100 ; p = 1 ; e = 10
Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

Column standardized
Data Standardization
Raw (same range)

Data Standardization

(Effect size)

Data Standardization

(Noise ratio)
Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

Simulation Results Summary

- Test statistic is unreliable for $n<20$, and for larger sample sizes (e.g., n=40) if the noise ratio is greater than ~10:1.

- **Significance** (p-value) of test statistic insensitive to degree of noise (nondiscriminating variables); i.e., a single effective discriminating variable will produce a significant test result for $n>20$, as long as noise ratio $<\sim 10:1$.

- **Effect size** ($R$) sensitive to both magnitude of effect on at least one discriminating variable and the noise ratio (i.e., ratio of nondiscriminating variables to true discriminators); noise ratio relatively more important.

- **Data standardization** (col z-score) has moderate effect on effect size ($R$).

Tests of Among-Group Differences

Variations on a Good Theme

- **Blocked MRPP (MRBP)** – For randomized blocked designs, paired-sample data, and simple repeated measures (haven’t found R function; but BLOSSOM software can do it).

- **$Q_b$ method** – for partitioning variance in the distance matrix into sums of squares for multiple factors, including interactions (limited to euclidean distance metric).

- **NPMANOVA** – for non-euclidean distance measures in multifactor designs, including nested and factorial designs. [adonis(vegan)]
Tests of Among-Group Differences
Considerations with MRPP & Related Methods

- MRPP and related methods (ANOSIM, MRBP, Qb, NPMANOVA) test for differences among groups in either *location* (differences in mean) or *spread* (differences in within-group distance). That is, they may find a significant difference between groups simply because one of the groups has greater dissimilarities (dispersion) among its sampling units.

- Differences in spread (‘heterogeneity of dispersions’) can be either an asset or an assumption that must be met.

Tests of Among-Group Differences
Considerations with MRPP & ANOSIM
Simulated Data Sets

*Equal location(e)*
*Equal spread(d)*

![Box plots](image)

\[ n=100, p=1, e=0, d=1 \]

*Unequal location(e)*
*Equal spread(d)*

![Box plots](image)

\[ n=100, p=1, e=1, d=1 \]
Tests of Among-Group Differences
Considerations with MRPP & ANOSIM
Simulated Data Sets

Equal location(e) Unequal spread(d)

Unequal location(e) Unequal spread(d)

n=100, p=1, e=0, d=2

n=100, p=1, e=1, d=2

n=100, p=1, e=0, d=2

n=100, p=1, e=0, d=4
Tests of Among-Group Differences
Considerations with MRPP & ANOSIM

Simulated Data Sets

**Unequal location**
equal spread

MRPP

ANOSIM

**Real data set example**

MRPP
ANOSIM
Tests of Among-Group Differences

Assumptions of MRPP & Related Methods

- **Independent samples** -- Usual problems associated with pseudoreplication, subsampling and repeated measures apply here.

- Chosen **distance measure** adequately represents the variation of interest in the data.

- The **relative weighting of variables** has been controlled prior to calculating distance, such that the weighting of variables is appropriate for the ecological question at hand.

Tests of Among-Group Differences

**Mantel’s Test (MANTEL)**

- Mantel statistic tests for differences between two distance matrices (e.g., between ecological and geographic distances between points), but can also be used to test for differences among groups.

- **Calculate** *dissimilarity matrices*.

- **Calculate** *test statistic* $Z$.

\[
Z = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} x_{ij} \cdot y_{ij}
\]

Contrast Matrix for Groups:

\[
X = \begin{bmatrix}
1 & 0 & 1 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 1
\end{bmatrix}
\]

Ecological Distance:

\[
Y = \begin{bmatrix}
0 & 1 & 1 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]
Tests of Among-Group Differences

Mantel’s Test (MANTEL)

- Determine the probability of a Z this large or larger through Monte carlo permutations.
  - Permutations involve randomly shuffling one of the two distance matrices.
  - The significance test is simply the fraction of permuted Z’s that are greater than the observed Z.

Note, if ecological distances are rank transformed, this test is the same as ANOSIM and similar to rank-transformed MRPP.

Tests of Among-Group Differences

Mantel’s Test (MANTEL)

- Determine the strength of the relationship between the two matrices by computing their correlation, r (standardized Mantel statistic).

\[
 r = \frac{\sum \sum \left( \frac{x_{ij} - \bar{x}}{s_x} \right) \left( \frac{y_{ij} - \bar{y}}{s_y} \right)}{n - 1}
\]

- \( r > 0 \) indicates that the average within-group distance between samples is less than the average overall distance between samples; the larger the r, the more tightly clustered the groups are.

- \( r < 0 \) numerically possible, but highly unlikely in ecology.
Tests of Among-Group Differences

*Mantel’s Test (MANTEL)*

2-group bird guilds example

- Conclude that two clusters differ *significantly* in terms of the measured habitat variables.

```
Mantel statistic based on Pearson's product

Call:
mantel2(xdis = y.eucl, ydis = contrast)

Mantel statistic r: 0.5404
Significance: < 0.001

Empirical upper confidence limits of r:
  99%  95%  97.5%  99%
0.084 0.122 0.176 0.211

Based on 1000 permutations
```

Variations on a Really Good Theme

- Because Mantel’s test is merely a correlation between distance matrices and the distance matrices can be variously defined, the test can assume a variety of forms as special cases.

- These are, in fact, variants of the same case but are interpreted somewhat differently. There are at least six variants.

\[
\begin{bmatrix}
Y \\
\text{Dissimilarity matrix}
\end{bmatrix}
\begin{bmatrix}
X \\
\text{Dissimilarity matrix}
\end{bmatrix}
\begin{bmatrix}
Z \\
\text{Dissimilarity matrix}
\end{bmatrix}
\]
Case 1. Simple Mantel’s Test on Geographic Distance.

If the dependent distance matrix is species similarity and the predictor matrix is geographic distance (“spatial dissimilarity”), the research question is “Are samples that are close together also compositionally similar?” This is equivalent to testing for overall autocorrelation in the dependent matrix (i.e., averaged over all distances).

\[
\begin{bmatrix}
\text{Species Dissimilarity} \\
\text{Geographic Distance}
\end{bmatrix}
\]

Tests of Among-Group Differences

**Mantel’s Test (MANTEL)**

- **Case 1. Simple Mantel’s Test on Geographic Distance.**

**Xmatrix** = euclidean distance between samples) [24 sites, 2 variables]

**Ymatrix** = Mass pine barrens moth abundances [24 x 10 species]
Case 2. Simple Mantel’s Test on a Predictor Matrix.

If the dependent matrix is again species similarity and the predictor matrix is a dissimilarity matrix based on a set of environmental variables, then the simple test is for correlation between the two matrices. Such correlation would indicate that locations that are similar environmentally tend to be similar compositionally. This, of course, is one of the fundamental questions in ecology.

\[
\begin{bmatrix}
\text{Species Dissimilarity} \\
\text{Environmental Dissimilarity}
\end{bmatrix}
\]

Tests of Among-Group Differences
Mantel’s Test (MANTEL)

Xmatrix =
Oregon riparian environment
(geomorphology, landscape composition and configuration)
[164 sites, 48 variables]

Ymatrix =
Oregon riparian birds abundances [164 sites, 49 species]
Case 3. Simple Mantel’s Test between an Observed Matrix and One Posed by a Model.

As a formal hypothesis test, Mantel’s test can be used to compare an observed dissimilarity matrix to one posed by a conceptual or numerical model. Here, the model is provided as a user-provided matrix of similarities or distances, and the test is to summarize the strength of the correspondence between the two matrices. The model distance matrix might be provided as a simple binary matrix of 0’s and 1’s, or a matrix derived from a more complicated model.

\[
\begin{bmatrix}
\text{Observed Dissimilarity} \\
\text{Model Dissimilarity}
\end{bmatrix}
\]

Tests of Among-Group Differences

**Mantel’s Test (MANTEL)**

\[\text{Xmatrix} = \]
Hammond’s flycatcher presence/absence in Oregon Coast Range
[96 sites, 1 indicator variable]

\[\text{Ymatrix} = \]
Oregon habitat variables
[96 sites, 20 habitat variables]
Case 4. The Mantel Correlogram.

A special case of case 1 (above) is to partition or subset the analysis into a series of discrete distance class. That is, a first distance matrix is evaluated for all pairs of points within the first distance class; then a second matrix is scored for all pairs of points within the second distance interval, and so on. The result of this analysis is a Mantel’s correlogram, completely analogous to an autocorrelation function but performed on a (possibly multivariate) distance matrix.

\[
\begin{bmatrix}
\text{Observed Dissimilarity} \\
\text{Geographic Distance Class}
\end{bmatrix}
\]

Tests of Among-Group Differences

Mantel’s Test (MANTEL)

Case 4. The Mantel Correlogram.

\[
\begin{bmatrix}
\text{X} \\
\text{Ecological distance}
\end{bmatrix}
\quad
\begin{bmatrix}
\text{Y} \\
\text{Geographic distance}
\end{bmatrix}
\quad
\begin{bmatrix}
1 & 0 & 1 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
0 & 1 & 0 \\
0 & 0 & 1 & 0 & 1
\end{bmatrix}
\]

Etc.
Tests of Among-Group Differences
Mantel’s Test (MANTEL)

Case 5. Partial Mantel’s Test on Three Distance Matrices.

The idealized Mantel’s test is a partial regression on three distance matrices. Here, the research question is, “How much of the variability in the dependent matrix is explained by the independent matrix after removing the effects of a third constraining matrix?” The analysis in this case is partial regression, and both partial correlation (or regression) coefficients are of interest: \( r_{YX \mid Z} \) and \( r_{YZ \mid X} \).

\[
\begin{bmatrix}
Y \\
\text{Dissimilarity matrix}
\end{bmatrix} = \begin{bmatrix}
X \\
\text{Dissimilarity matrix}
\end{bmatrix} \begin{bmatrix}
Z \\
\text{Dissimilarity matrix}
\end{bmatrix}
\]

Path Diagram (Oregon streamside bird communities)
Case 5. Partial Mantel’s Test on Multiple Distance Matrices.

Path Diagram (Mass. pine barrens moth community)

Case 6. Partial Mantel’s Test on Multiple Predictor Variables.

Often, knowing that the environment has some relationship with the dependent variable of interest is not sufficiently satisfying: we wish to know which variables are actually related to the dependent variable. The logical extension of Mantel’s test is multiple regression, in which the predictor variables are entered into the analysis as individual distance matrices. As a partial regression technique, Mantel’s test provides not only an overall test for the relationships among distance matrices, but also tests the contribution of each predictor variable for its pure partial effect on the dependent variable.