Common Goals of Multivariate Procedures

- Classify observations into groups
- Test for differences or trends based on multiple measures
- Classify and relate variables to one another
Methods for dealing with nondetects

• NOT Substitution! This adds invasive data
• For one RL:
  – Rank data and run procedure on the ranks
  – Use a nonparametric method
For 1 RL, run procedure on the ranks

- PCA on ranks
- Discriminant Function on ranks
- All nondetects are tied with each other at the lowest rank:

<1 <1 <1 3 7 10 18 data
2 2 2 4 5 6 7 ranks

- MANOVA, Cluster Analysis have nonparametric analogues, so no reason to do this approach
For 1 RL, use a nonparametric method

- ANOSIM (Nonparametric MANOVA)
- Test for seriation (Nonparametric Trend Test)
- Nearest Neighbor Cluster Analysis
- Nonmetric Multidimensional Scaling
General Method: ANOSIM and test for seriation

Explanatory Variables

Environmental Data

0/1 Pattern or Similarity Matrix

Kendall's tau similarity matrix

Element by element Kendall's tau correlation between similarity matrices
Test for seriation

For trend, the explanatory matrix contains the number of time steps between measurements.
For tests between groups, the similarity matrix is a 1 if two values are in the same group, and a 0 otherwise.
General Method

For other, regression type problems, the similarity matrix uses something like Kendall's tau.
Permutation p-values

Productive vs. Non-productive observations

Observed test statistic
\[ R = 0.758 \]

Permutation results when there is no difference between groups

Observed test statistic higher than all 1000 results in permutations, so p-value = 0.001

\[ R = \frac{\bar{r}_{\text{between}} - \bar{r}_{\text{within}}}{n(n-1)/4} \]

\( r \) is a rank similarity between or within groups
Nearest-neighbor clustering

- ranks of distances between locations in multivariate space
Methods for dealing with nondetects

- For more than one RL:
  1. Compute Kaplan-Meier percentiles and run a procedure on these 'scores'
  2. Use maximum likelihood (MLE) versions of methods, where those exist
For >1 RL, compute procedure on K-M percentiles

- existing tests take multivariate measures and convert to a univariate composite score. Test the scores for differences between groups. Scores are combinations of K-M percentiles for each variable.
Kaplan-Meier
(nonparametric method)

- K-M estimates the survival function $S$, the probability of $\leq$ each detected value
- $S$ estimates the empirical CDF (percentile function) of the original data
- Percentiles only estimated for detects, not nondetects
- Percentile values are affected, however, by the number of and DL value
Kaplan-Meier survival curve

Same as a cdf
(plotted left to right)
MLE methods for >1 RL

- Multiple regression analogue: predicting a Y with nondetects from multiple explanatory variables.
- Use MLE with an assumed distribution (normal, lognormal, other) to determine the best predictors.
- Censoring only allowed for the Y variable
- Has been extended to factor analysis, but software not readily available
2nd Edition will be even better!

Non-detects

And

Data Analysis

Statistics for Censored Environmental Data

Helsel (2005)

www.PracticalStats.com / nada

Stop by our booth in the right corner of the exhibit hall