Environmental Analysis and Modeling
BIOL 650
R. C. Jones

I. Nature of Environmental Analysis
    A. Types of Analysis
        1. Empirical vs. Theoretical
        2. Physiological
        3. Population
        4. Community
        5. Ecosystem
    B. Objectives of Analysis
        1. Description
        2. Generalization
        3. Hypothesis testing
    C. Basic Concepts
        1. Variation
        2. Surveys vs. Experiments
        3. Population (universe) vs. Sample
        4. Stratified vs. Completely Randomized Designs
        5. Univariate vs. Multivariate
        6. Graphical vs. Numerical
        7. Scale
        8. Replication
    D. Study Designs (Green 1979)
II. Graphical Analysis of Environmental Data
    A. Uses of Graphing
        1. Exploratory Data Analysis
        2. Presentation of Data
    B. Criteria for Good Graphs
        1. Ellison (1993) in Scheiner and Gurevitch [on reserve]
        2. Cleveland (1985) cited in Ellison
    C. Data Types and their Graphs
        1. Univariate data - one sample "Looking at distributions"
            a. One-dimensional scatterplots
            b. Histograms
            c. Stem and Leaf Plots
            d. Quantile Plots
            e. Box and Whisker Plots
        2. Univariate data - multiple samples/categories
            a. Bar graphs
            b. Multiple one-dimensional representations arranged side-by-side
                -scatterplots
                -back-to-back stem and leaf
                -multiple box plots: time series, treatment comparisons
        3. Univariate data classified by 2 variables (Fig. 2.12, Ellison)

a. Grouped bar graphs (with or without error bars)
b. Stacked bar graph or pie chart
c. Grouped mean and error bar graphs
d. Grouped box plots

4. Bivariate data
    a. Scatterplot
    b. Influence plot
    c. Convex hulls
    d. Vertical strips
    e. Smoothing
    f. Jittering, sunflowers, cellulation, and sharpening
5. Multivariate data - Three dimensions
    a. Three-dimensional scatterplots using perspective
    b. Multiple 2-D views -Draftsman's Display
    c. Multiple 2-D views - Pasted on a cube
    d. Cube cut at a common vertex
    e. Symbolic variation
6. Multivariate data - Four or more dimensions
    a. Draftsman's Display
    b. Multple code symbols = Glyphs

III. Overview of Multivariate Data

A. Definition: Each entity (e.g. sample) is represented by multiple attributes (e.g. species)
B. Example: Thorp et al. SAV macroinvertebrate data
C. Approaches to using multivariate data
    1. Graphical
    2. Indices: extraction of a single value for each entity (conversion of multivariate to univariate)
    3. Pairwise comparisons among entities (similarity/distance measures)
    4. Multivariate analysis: classification and ordination

IV. Indices - Condensing Multivariate Data into Univariate

A. Diversity
    1. Richness
    2. Simpson's Index
    3. Shannon's Index
B. Other simple indices
    1. Value of dominant species (attributes)
    2. Weighted composite of selected species (attributes)
C. Composite Indices
    1. Rapid Bioassessment Index for macroinvertebrate
        -composed of several distinct subindices or “metrics”
        -e.g. taxa richness, tolerance index, % dominance
        -each index is calculated separately on each sample
        -each is standardized as a percentage of reference value and scored according to its own distribution relative to the desired condition

-then they are summed to create a composite index of macroinvertebrate community condition
-thus you have a single number (RBP index) arising from a multivariate observation (community composition)
-similar to “Index of Leading Economic Indicators”

V. Pairwise Comparisons: Similarity/Dissimilarity Measures
- A. Quantitative Data
  1. Metric vs. non-metric: triangle inequality
  2. Euclidean Distance and related measures
  3. Absolute Distance (aka city block, Manhattan, taxicab) and related measures
  4. Percentage Similarity (Bray-Curtis Distance) and related measures
  5. Relativizing and Normalized Distances
  6. Relativized Euclidean and Absolute Distance
  7. Chord and geodesic distance
- B. Binary Data
  1. Advantages and disadvantages
  2. Percentage similarity
  3. Ochiai
  4. Dice
  5. Jaccard

V. Cluster Analysis - Using Multivariate Data to Create a Hierarchical Classification
- A. Overview
  1. Choose a Similarity/Dissimilarity/Distance Measure
  2. Choose a Clustering/Linkage Algorithm
- B. Similarity/Dissimilarity/Distance Measures (see IV.C. above)
- C. Clustering Algorithms
  1. Methods using original distance matrix
     a. Nearest neighbor (Single linkage) clustering
     b. Farthest neighbor (Complete linkage) clustering
  2. Methods requiring recalculation of distance after each linkage
     a. Centroid likage ("unweighted" and "weighted")
     b. Average distance linkage ("unweighted" and "weighted")
     c. Minimum variance clustering
- D. Pros and Cons of each approach
- E. Papers using Cluster Analysis

Ross, S.T. 1977. Patterns of resource partioning in sea robins (Pisces: Triglidae). *Copeia* 3: 561-571.

Jones, R.C., J. Cairns, Jr., and W.H. Yongue. 1976. Vertical gradients in artificial substrate-associated protozoan community structure in a stratified freshwater lake. *Journal of the Elisha Mitchell Scientific Society* 92: 1-8.

Thorp, A.G., R.C. Jones, and D.P. Kelso. 1997. A comparison of water-column macroinvertebrate communities in beds of differing submersed aquatic vegetation in the tidal freshwater Potomac River. *Estuaries* 20: 86-95.