

# QUANTITATIVE METHODS FOR ECOLOGISTS

## FNR 647

**Instructor:** Patrick Zollner (102A FORS; email: pzollner@purdue.edu; phone: 49-43590)

**Objective and Rationale:** To provide a forum for examining the assumptions and utility of statistical techniques that are commonly encountered or increasingly being used in ecological research. Emphasis will be placed on the potential application of these quantitative methods in an ecological context. Courses offered in other departments often provide the theoretical foundation and alternative applications for some of these methods, but coverage of the methods' applied implications for ecologists receives diminished treatment as a consequence.

**Credits:** 3

**Prerequisites:** STAT 512 or equivalent, or consent of instructor. A course in linear algebra, knowledge of a programming language, and experience with Matlab, SAS, SPSS, or S-Plus desirable but not essential.

**Text:** Legendre, P., and L. Legendre. 1998. Numerical ecology, 2<sup>nd</sup> English edition. Elsevier, Amsterdam.

**Time and Location:** Spring 2004, TTh 9:00-10:15, in FORS 215

**Structure:** Principally a lecture format. Grading will be based on out-of-class homework assignments, attendance, and an oral presentation depicting application of software.

**Tentative Topics (for further description, see next section):**

1. Decision algorithm for selection of quantitative methods  
Overview of linear algebra
2. Multivariate methods: principal components analysis, factor analysis, other ordination techniques
3. Multivariate methods: parametric tests of significance
4. Multivariate methods: discriminant analysis, logistic regression
5. Multivariate methods: cluster analysis
6. Analysis of categorical data: log-linear models
7. An overview of common nonparametric methods
8. Randomization and Monte Carlo methods  
rationale, random number generation, jackknife, bootstrap, etc.

9. Analysis of resource selection: selection ratios, logistic, proportional hazards models, compositional analysis
10. Model selection: an information-theoretic approach
11. Spatial statistics
12. Probabilistic methods for estimating population attributes
13. Probabilistic methods for estimating community attributes
14. Analysis of time series and detection of density dependence
15. Analysis of community structure and dynamics
16. Meta-analysis: methods for quantifying results from independent studies
17. Topics in regression
18. Repeated measures analysis of variance
19. Sensitivity and elasticity analysis
20. Local stability analysis
21. Fourth corner problem

## OVERVIEW OF SELECTED TOPICS

### Multivariate tests of significance:

These tests permit simultaneous testing of data collected on several variables for each object in a study. Typically, they are relatively straightforward extensions of the univariate tests with which we are familiar (e.g., t-test, ANOVA, F-test). Additional information, related to multivariate distances (or similarities), also can be computed between objects for use in other applications.

### Principal Components Analysis:

The objective of PCA is to explain the structure (i.e., variation) in a data set containing a large number of correlated variables by using a smaller number of uncorrelated variables. The latter variables are linear combinations of the original variables. This reduced variable set can then be used in other applications, making PCA one of the more commonly used multivariate methods. Moreover, the procedure used in deriving the principal axes also is used in many other multivariate methods.

### Factor Analysis:

The objective of FA is similar to PCA. The methods differ in that FA is based on a specific statistical model whose aim is to account for the covariance among the original variables using a smaller number of uncorrelated variables. These “factors” may then be rotated to simplify the interpretation of the factors while retaining their explanatory power. FA has not received much use historically by ecologists, but that trend appears to be changing, as the reliability of PCA is questioned and the value of FA is heralded.

### Cluster Analysis:

The objective is to solve the following problem – given a sample of  $n$  objects, each of which has a score on  $p$  variables, devise a scheme for grouping the objects into classes so that “similar” ones are in the same class. The number of classes is unknown. Results often are presented as branching “trees” or dendograms. Cluster analysis is useful in a variety of contexts where the objects are not members of clearly defined groups. This technique was first used by systematists, but it’s now being used more commonly by ecologists.

### Multidimensional Scaling:

MDS is a technique that is designed to construct a “map” showing the relationships between a number of objects, given only a table of distances (or similarities) between them. So, the goal is to reduce the information in a multivariate data set into 2 or 3 dimensions that can then be plotted. Non-metric MDS preserves the ordering of objects with respect to their distances, but not the exact distances themselves. This approach permits data to be more easily collapsed into 2 or 3 dimensions than if the actual Euclidean distances are preserved (metric MDS).

### Ordination:

Ordination and scaling both describe the process of producing a small number of variables (from a larger number) that can be used to illustrate graphically the relationship between a group of objects. The new variables should represent a large fraction of the total variation in the original variables. Thus, PCA, FA and MDS all are ordination techniques. Correspondence analysis is another ordination technique commonly used on species abundance or presence-absence data.

### Discriminant Function Analysis:

The problem addressed by DFA is how well it is possible to separate two or more groups of objects, given measurements for these objects on several correlated variables. This is done by constructing linear combinations of the original variables that best discriminate among groups. These combinations are called discriminant functions. The information also can be expressed as classification functions, which can be used to assess errors made in classifying objects into groups.

### Logistic Regression:

This tool can be used to address the discrimination problem described above under DFA. In addition, logistic regression is used to model the probability of a “success” or “failure” as a function of several explanatory variables. Thus, it provides an appropriate alternative to multiple linear regression when the response variable is binary (i.e., either 0 or 1). Explanatory variables may be of mixed precision (qualitative or quantitative). Logistic regression is useful in any situation where the response variable is of the “either-or” type (e.g., presence-absence of a species at many different sites).

### Multiway Contingency Analysis:

This tool is a multivariate extension of the  $r \times c$  contingency analysis covered in introductory statistics courses for analyzing categorical data for 2 variables. Thus, the structure of a multivariate data consisting of qualitative data (data separated into categories) can be analyzed. When the qualitative response variable is binary, logistic regression can be used.

### Topics in Regression:

Traditional statistics courses devote most of their time to simple linear regression (model I) and to multiple linear regression. However, many other types of regression are useful, including logistic regression (see above). We will consider model II linear regression (major axis regression), polynomial regression, partial linear regression, nonlinear regression, and smoothing techniques.

### Nonparametric Statistics:

Traditional statistics courses deal principally with tests based on one of a few families of probability distributions:  $z$ ,  $F$ ,  $t$ ,  $\chi^2$ . When data do not meet the distributional requirements of these parametric tests, nonparametric statistics are useful alternatives. Nonparametric statistics are “distribution-free”. They often rely on ranks of original data. Instead of inference on the mean and variance, these statistics focus on the median and range, respectively.

### Randomization and Monte Carlo Methods:

The role of a statistical test is to decide whether some parameter of the reference population may take a value assumed by hypothesis, given the fact that the corresponding statistic, whose value is estimated from a sample of objects, may have a somewhat different value. In standard statistical tests, the test statistic is compared to a critical value from one of the usual distributions; e.g.,  $z$ ,  $F$ ,  $t$ ,  $\chi^2$ . The validity of these tests depends on certain assumptions. Common assumptions include normality of the variable(s) and homoscedasticity. These assumptions may be violated when sample sizes are small. In addition, some the distributional

properties of some statistics are poorly known. In these instances, randomization or permutation tests can be performed. The idea is to generate the reference distribution to which the test statistic is compared from the data themselves, rather than relying on a standard statistical distribution.

#### Sensitivity Analysis:

Analytical and simulation models often depend on several variables as input. Sensitivity and related analyses address the question: How sensitive is the response variable to changes in each of the input variables? In answering this question, one can determine the relative importance of various life history traits on, e.g., population growth.

#### Stability Analysis:

Models in ecology often exhibit nonlinear response functions, or response functions that are related to several variables. Two tasks incumbent on a modeler are to assess the conditions necessary for equilibrium (no change in rate of change of response variable) and to determine how the system responds to small disturbances from equilibrium. Stability analysis is the method by which the latter task is achieved.

#### Fourth-Corner Problem:

How do the biological and behavioral characteristics of species determine their relative locations in a landscape or region? To answer this question requires a way of detecting relationships between species traits and habitat characteristics, and then of testing the significance of these relationships. Given data on presence-absence of species at a variety of sites, as well as supplementary information on aspects of the species' biology and habitat characteristics at each of the sites, the 4<sup>th</sup>-corner problem provides a means of answering this question. Note that the 4<sup>th</sup>-corner problem also can be used to address how prey use and availability influence the relative locations of predators in a landscape, among other things.

#### Analysis of Resource Selection:

Resource selection can be determined when information is available on resource use and availability. Data may come from behavioral observation, remote sensing, analysis of stomach or scat contents in conjunction with resource assessment, or other means. Analysis can take several forms, including calculation and testing of selection coefficients, as well as logistic regression and proportional hazards models.

#### Spatial Analysis:

Virtually all ecological data are collected within a heterogeneous spatial context. Spatial analysis has as its goal the determination of spatial patterns associated with ecological phenomena. This is done with a variety of tools, including correlograms and variograms to assess spatial autocorrelation, as well as Mantel correlograms for multivariate data and Mantel tests for spatial gradients and modeling species-environment relationships. Temporal autocorrelation (correlation of observations through time instead of through space) can be examined using similar tools (e.g., autocorrelation functions).

### Estimation of Population Attributes:

Size and density are fundamental attributes of populations. In turn, these are influenced by vital rates, namely dispersal, birth, and death. Several approaches are available for estimating population attributes. We will focus only on a few, such as distance sampling and mark-recapture methods, for estimating abundance, density, and vital rates.

### Estimation of Community Attributes:

Estimation of community metrics, such as species richness and diversity, niche breadth, and niche overlap, are seldom dealt with in a statistical context. In addition, little attention has been devoted to other community properties of interest from a conservation perspective, such as species turnover (which can be thought of as the population equivalent of survival). We will review previous work and examine some recent advances in these areas.

### Model Selection – An Information-Theoretic Approach:

Ecologists commonly use ad hoc model-fitting procedures with no underlying conceptual foundation. Whenever possible, models should be selected from an a priori candidate set, formulated from hypotheses regarding the mechanisms underlying the system under study. Here, we will discuss Kullback-Liebler information and the use of Akaike's Information Criterion to compare the performance of competing models in approximating "truth". We also will examine ways of determining model goodness-of-fit, including likelihood ratio tests and bootstrapping methods.

### Selected References:

- Boulinier, T., J. D. Nichols, J. R. Sauer, J. E. Hines, and K. H. Pollock. 1998. Estimating species richness: the importance of heterogeneity in species detectability. *Ecology* 79:1018-1028.
- Buckland, S. T., D. R. Anderson, K. P. Burnham, and J. L. Laake. 1993. Distance sampling: Estimating abundance of biological populations. Chapman and Hall, New York.
- Burnham, K. P., and D. R. Anderson. 1998. Model selection and inference: a practical information-theoretic approach. Springer-Verlag, New York.
- Caswell, H. 2001. Matrix population models, 2<sup>nd</sup> edition. Sinauer Associates, Sunderland, MA.
- Cox, D. R., V. Isham, and P. Northrop. 2000. Statistical modeling and analysis of spatial patterns. Pages 65-88 in U. Dieckmann, R. Law, and J. A. J. Metz (eds.), *The geometry of ecological interactions*. Cambridge Studies in Adaptive Dynamics, Cambridge University Press.
- Crawley, P. H. 1992. Resampling methods for computation-intensive data analysis in ecology and evolution. *Annual Review of Ecology and Systematics* 23:405-447.
- Dale, M. R. T. 1999. Spatial pattern analysis in plant ecology. Cambridge University Press.
- De'ath, G., and K. E. Fabricius. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81:3178-3192.

- Dennis, B., and M. L. Taper. 1994. Density dependence in time series observations of natural populations: estimation and testing. *Ecol. Monogr.* 64:204-224.
- Gardner, R. H., W. M. Kemp, V. S. Kennedy, and J. E. Petersen. 2001. Columbia University Press, New York.
- Graham, M. H. 2003. Confronting multicollinearity in ecological multiple regression. *Ecology* 84:2809-2815.
- Hilborn, R., and M. Mangel. 1997. *The ecological detective: confronting models with data.* Monographs in Population Biology Number 28, Princeton University Press, Princeton, NJ.
- Hof, J., and M. Bevers. 1998. *Spatial optimization for managed ecosystems.* Columbia University Press, New York.
- Legendre, P., and L. Legendre. 1998. *Numerical ecology*, 2<sup>nd</sup> English edition. Elsevier, Amsterdam.
- Legendre, P., R. Galzin, and M. Harmelin-Vivien. 1997. Relating behavior to habitat: solutions to the fourth-corner problem. *Ecology* 78:547-562.
- Koenig, W. D. 1998. Spatial autocorrelation in California land birds. *Conservation Biology* 12:612-620.
- McGarigal, K., S. Cushman, and S. Stafford. 2000. *Multivariate statistics for wildlife and ecology research.* Springer.
- Manly, B. J. F. 1997. *Randomization, bootstrap and Monte Carlo methods in biology*, 2<sup>nd</sup> edition. Chapman and Hall, London.
- Manly, B. J. F. 1992. *The design and analysis of research studies.* Cambridge University Press, Cambridge.
- Manly, B. J. F. 1994. *Multivariate statistical methods: a primer.* 2<sup>nd</sup> edition. Chapman and Hall, London.
- Manly, B. J. F., L. McDonald, and D. Thomas. 1993. *Resource selection by animals: statistical design and analysis for field studies.* Chapman and Hall, London.
- Meyer, J. S., L. L. Irwin, and M. S. Boyce. 1998. Influence of habitat abundance and fragmentation on northern spotted owls in western Oregon. *Wildlife Monographs* Number 139.

- Nichols, J. D., T. Boulinier, J. E. Hines, K. H. Pollock, and J. R. Sauer. 1998. Estimating rates of local species extinction, colonization, and turnover in animal communities. *Ecological Applications* 8:1213-1225.
- Nichols, J. D., J. E. Hines, J-D. Lebreton, and R. Pradel. 2000. Estimation of contributions to population growth: a reverse-time capture-recapture approach. *Ecology* 81:3362-3376.
- Phillips, P. C., and S. J. Arnold. 1999. Hierarchical comparison of genetic variance-covariance matrices. I. Using the Flury hierarchy. *Evolution* 53:1506-1515.
- Pielou, E. C. 1984. *The interpretation of ecological data: a primer on classification and ordination*. John Wiley and Sons, New York.
- Ripley, B. D. 1988. *Statistical inference for spatial processes*. Cambridge University Press, Cambridge, UK.
- Roughgarden, J. 1998. *Primer of ecological theory*. Prentice-Hall, Upper Saddle River, New Jersey.
- Roxburgh, S. H., and P. Chesson. 1998. A new method for detecting species associations with spatially autocorrelated data. *Ecology* 79:2180-2192.
- Trexler, J. C., and J. Travis. 1993. Nontraditional regression analyses. *Ecology* 74:1629-1637.
- White, G. C., and K. P. Burnham. 1999. Program MARK: survival estimation from populations of marked animals. *Bird Study* 46(suppl.) S120-S139.
- Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. *Analysis and management of animal populations*. Academic Press.