Advanced Biometrics in Marine Science

MBF 615 -- Fall 2011

Instructor: Dr. Jerald S. Ault
CIMAS 210, University of Miami RSMAS campus
305-421-4884 ph; 305-421-4791 fax; jault@rsmas.miami.edu

TA: Mr. William Harford
Collier 201
305-421-4611 office; wharford@rsmas.miami.edu

Hours: T-TH 0900-1015, RSMAS S/A-120
with computer-intensive laboratory


Brief Course Description.- An introduction to advanced multivariate statistical analysis of empirical observations with primary emphasis on applications in the assessment and interpretation of the dynamics of populations and communities in marine biology, fisheries, biomedical sciences, and biological oceanography. Advanced methods in generalized linear models, multiple and nonlinear regression model analysis, probability and estimation theory, multiple partial correlation, ANCOVA, generalized additive models, nonlinear optimization, multivariate statistics (classification and ordination), and sampling techniques. Exploratory data analysis and modeling will be emphasized using the software: R Project for Statistical Computing, SAS and MATLAB.

Grades.- Grades are based on: 55% computer-intensive laboratories; 15% Mid-term; 25% Final project; and, 5% for class participation.

Course Textbooks:
Computer-Intensive Laboratories

Lab 1: Mathematical Refresher: Principles of Advanced Biometrics

Lab 2: Linear Statistical Models

Lab 3: Generalized Linear Models: Logistic Regression

Lab 4: Bayesian Linear Regression

Lab 5: Multivariate Regression Analysis

Lab 6: Analysis of Covariance

Lab 7: Nonlinear Regression

Lab 8: Classification: Cluster Analysis

Lab 9: Ordination: Principal Components Analysis

Lab 10: Nonparametric Regression and Other Advanced GLMs
# Advanced Biometrics in Marine Science

## Table of Contents

### GENERALIZED LINEAR MODELS (GLMs)

1.0  **Linear Regression with One Predictor Variable**
- Functional and Statistical Relations between Two Variables
- Regression Models and Their Uses
- Simple Linear Regression Model
- Data for Regression Analysis
- Method of Least Squares Estimators: Development of the Normal Equations
- Assumptions in Regression
- Properties of Least Squares Estimators
- Estimation of Error Terms Variance
- ANOVA Table for Straight-Line Regression
- Fundamental Equation of Regression Analysis
- Equivalence of $F$ and $t$ tests

2.0  **Linear Regression Model in Matrix Terms**
- Least Squares Estimation of Regression Parameters
- Fitted Values and Residuals
- Analysis of Variance Results
- Sums of Squares as Quadratic Forms
- Inferences Concerning the Regression Line
- Prediction of a New Value of $Y$ at $X_0$
- Confidence Ellipses for the Bivariate Normal Distribution

3.0  **Regression Model Diagnostics**
- Diagnostics for Predictor Variables
- Properties of the Residuals
- Assessing the Assumption of Normality
- Presence of Outliers
- Nonlinearity of Regression Function
- Nonnormality of Error Terms
- Multicollinearity & Omission of Important Predictor Variables
- Tests of Constancy of Error Variance
- Appropriateness of the Straight-Line Model: $F$ Test for Lack of Fit
- Overview of Remedial Measures
- Multivariate Unconditional Box-Cox Transformations
- Other Regression Diagnostics: Leverage Points
- Standardized Residuals
- Assessing the Influence of Certain Cases
# Advanced Biometrics in Marine Science

## Table of Contents

### 4.0 Probability and Estimation Theory
- Basic Concepts of Probability
- Samples and Statistical Expectation
- Parameterization of Discrete and Continuous Probability Density Functions
- Maximum Likelihood Estimation (MLE): normal & exponential distributions
- Desirable Properties of Estimators
- Probability distributions for inference
- Comparison of ordinary least squares and maximum likelihood estimation
- Logistic Regression
- Bernoulli Distribution, Binomial Regression and MLE Computations
- Numerical Optimization: Logit Model
- Newton-Raphson’s algorithm
- Computational Example: The Generalized Linear Model (GLM)
- Iterative Algorithms for Model-Fitting
- Logistic, Exponential and Gamma error distributions
- Multinomial Logit Model

### 5.0 Multivariate Regression Analysis
- Polynomial Regression
- Graphical Look at the Problem
- First-Order Model with More than Two Predictor Variables
- Assumptions of Multiple Regression
- Determining the Best Estimate: Least-Squares Approach
- Analysis-of-Variance Table, Overall $F$ Test, and Partial $F$ Test

#### Generalized Linear Multiple Regression Model in Matrix Terms
- Analysis of Variance Results
- Coefficient of Multiple Determination
- Estimation of Regression Parameters and Mean Response
- Inferences about Regression Parameters
- Analysis of Appropriateness of Model
- Selecting the Best Regression Equation

#### Multiple Partial Correlation and Analysis of Covariance
- Correlations: Multiple, Partial and Multiple Partial
- Correlation Matrix
- Relationship of $R_{YX_1, X_2, \ldots, X_k}$ to the Multivariate Normal Distribution
- Multiple Correlation Coefficient
- Partial Correlation Coefficient
- Relating Partial Correlation Test to the Partial $F$ Test
Advanced Biometrics in Marine Science

Table of Contents

Multiple Partial Correlation

**Bayesian Statistical Regression Methods**
- Frequentist versus Bayesian
- Bayes Rule and Parameter Estimation
- Estimating the Posterior Distributions
- Example Applications

***6.0 Analysis of Covariance***
- Concept of Interaction
- Analysis of Covariance (ANCOVA): Comparing Two Straight Lines
- Dummy Variables in Regression
- Tests Concerning Slopes and Intercepts
- ANCOVA: Several Groups and Several Covariates

***7.0 Nonlinear Regression and Nonlinear Least Squares***
- Nonlinear Regression
- Numerical Optimization: theory and practice
- Nonlinear regression and the Gauss-Newton algorithm
- Distinctions between OLS and NLS
- Nonlinear Regression Examples
- Initial Values for Nonlinear Regression Estimation
- Specifying the Gradient

**MULTIVARIATE STATISTICAL ANALYSES**

***8.0 Clustering and Distance Methods***
- Introduction to Multivariate Statistics
- Classification methods and statistical distance
- The Organization of Data: arrays and descriptive statistics
- Distances and Similarity Coefficients for Pairs of Items
- Computation of a (statistical) Distance Matrix
- Hierarchical Clustering Methods (e.g., single linkage, nearest-neighbor, centroid)
- Final Comments – Hierarchical Procedures

***9.0 Population Principal Components (EOF) Analysis***
- Basic Data Requirements for Ordination in Reduced Space
- Principal Variances for Standardized Data Sets
- Multivariate Normal Distributions
- Computation of the EOF (Empirical orthogonal functions)
- Principal Components: Principal directions as basis vectors
- Spectral Decomposition: Factoring the Scatter matrix
# Advanced Biometrics in Marine Science

Table of Contents

- Principal Variances, Covariances and Angles
- Exploratory Data Analysis: Interpretation of Principal Components
- “Rule N” Selection of Principal Components
- Factor Analysis: relationship to PCA
- Multidimensional Scaling
- Non-metric Multidimensional Scaling

**NONLINEAR AND NONPARAMETRIC MULTIVARIATE REGRESSION MODELS**

## 10.0 Advanced Multivariate Statistical Models

- Extending the Linear Model: generalized Linear Models Redux
  - The Structure of Generalized Linear Models
  - Examples of Generalized Linear Models
  - Fishing power estimation model
- Nonparametric Regression
  - Kernal Estimators
  - Smoothing Splines
  - Regression Splines
  - Local Polynomials
  - Wavelets
- Generalized Additive Models
  - Alternating Conditional Expectations
  - Additivity and Variance Stabilization
  - Generalized Additive Mixed Models
  - Multivariate Adaptive Regression Splines
- Neural Networks
  - Statistical Models as NNs
  - Feed-Forward Neural Networks with One Hidden Layer
  - NN Applications
- Tree Models
  - Background
  - Regression Tress
  - Tree Models as Regressions
  - Model Simplifications
  - Classification Trees with Categorical Explanatory Variables
  - Testing for the Existence of Humps
- Spatial Statistics
  - Point Processes
  - Nearest Neighbors
  - Tests for Spatial Randomness
# Advanced Biometrics in Marine Science

## Table of Contents

- Ripley’s K
- Aggregated Data and Quadrat Count Data
- Geostatistical Data
- Regression Models with Spatially Correlated Errors
- **Mixed-Effects Models**
- Replication and Pseudoreplication
- Best Linear Unbiased Predictors
- Hierarchical Sampling and Variance Components Analysis
- Model Simplification in Hierarchical Sampling
- Mixed Effects Models with Temporal Pseudoreplication
- Time-Series Analysis in Mixed-Effects Models
- Random Effects in Designed Experiments
- Generalized Linear Mixed Models
- Fixed Effects in Hierarchical Sampling
- **Modern Sampling Techniques: Probability-based Sampling**
  - Sampling survey theory
  - Simple random sampling
  - Stratified random sampling
  - Design and performance criteria
  - Optimal allocation
  - Model-based versus design-based approaches to analyses
  - From HSIs to GAMs to optimal sampling designs for abundance estimation