Biology 495, Spring 2015

Undergraduate Seminar- Advanced Biostatistics

Prerequisites: Stat 401 or other introductory statistics course

Course Format: There will be one lecture each week, followed by a computer laboratory. In laboratory we will implement the methods discussed in lecture to provide 'hands on' experiences designed to reinforce the statistical concepts learned that day. The laboratory exercises will be performed in R.

Preliminaries and Course Preparation: Students should bring their laptops to class. **Prior to the first class period,** please download and install a copy of R on your laptop (found at: http://cran.r-project.org/). All lecture notes and lab materials will be made available before class. They can be found on the internet at: http://www.public.iastate.edu/~dcadams/BioStatMaterials/.

    NOTE: laboratory sessions alone are not possible for one to become an expert in R! One hour is simply not enough time for this. Thus, students are strongly encouraged to go through the scripts again on your own to better understand how and why the functions operate. **ONLY THROUGH PRACTICE AND REPETITION WILL YOU BECOME PROFICIENT IN R!**

Recommended Readings: There is no formal text for this seminar course. The sequence of topics does not follow a particular textbook, but rather is one that I have found to be successful in teaching biologists the basics of biological statistics. However, there are several textbooks that students may borrow from me at any time. These are books that I keep on my shelf, and have found to be useful reference material. They are:

NUMERICAL ECOLOGY WITH R.

Legendre, P., and L. Legendre. 1998. *Numerical Ecology.* Second Edition. Elsevier Science, Amsterdam.
Manly, B. F. J. 1994. *Multivariate Statistical Methods: A Primer.* Second Edition. Chapman & Hall, London.
Sokal, R. R. and F. J. Rohlf. 2012. *Biometry.* 4th Edition. Freeman Associates, San Francisco.

Course Description and Objectives: All biological research requires the use of statistics to determine whether or not the data collected in an experiment deviate significantly from null expectations. Unfortunately, biology students spend far too little time investigating what statistical tools they have at their disposal, and instead invest their limited time in learning more biology. While it is certainly necessary to learn as much biology as possible, an undesirable consequence of this choice is that a student's rudimentary knowledge of statistics relegates their hypothesis testing to methods that they already know: if a student only knows a t-test, **the design of their experiments begins to resemble that of a t-test**! This is extremely limiting, as the range of biological questions they want to ask is essentially hindered by lack of tools, not lack of biological knowledge and intuition. This course is designed as the first-step in alleviating this difficulty.

    This course presents a survey of many univariate and multivariate statistical methods commonly used in evolutionary and ecological research. It motivates this topic from the conceptual and logical foundations of matching one's data and hypotheses with analytical techniques, so that biological patterns can be properly assessed. The goal is to provide a working knowledge of statistical methods, so that one may better determine which analytical approaches are most appropriate for different types of data. While this seminar is **NOT** intended to be a substitute for multivariate courses in statistics, it **IS** intended to provide sufficient detail so that students may pick up a recent issue of *Evolution, Ameraican Naturalist, Ecology,* etc. and be able to understand the statistical methods used in the articles. This way, students may more critically evaluate the literature, and hopefully, be able to use these methods in their own research as the need arises.

    To make the mathematics accessible, I will use a 'think-first, implement methods second' approach; beginning with data types and hypotheses, progressing to the choice of statistical method, and finally ending with equations and necessary practical details. We first cover basic statistical concepts and resampling procedures for assessing significance, followed by univariate statistics (ANOVA, regression, correlation), and the 'jump' to multivariate general linear models (GLM: MANOVA, MANCOVA, multivariate regression). We then

review exploratory methods, including ordination approaches (CVA, PCA, PCoA, MDS), clustering (UPGMA, WPGMA, K-means), and multivariate 'correlation' (canonical correlation, and PLS). We will also discuss several more specialized topics, including: Mantel tests, the phylogenetic comparative method, spatial statistics, meta-analysis, and model selection.

Disability Accommodations: Iowa State University complies with the Americans with Disabilities Act and Section 504 of the Rehabilitation Act. Please address any special needs or special accommodations with Dr. Adams no later than the end of the first week of class. Those seeking accommodations based on disabilities should obtain a Student Academic Accommodation Request (SAAR) form from the Disability Resources (DR) office, 515-294-6624. DR is located 1200 Hixson-Lied Student Success Center. No retroactive accommodations will be provided in this class.

Non-Discrimination Policy: Iowa State University does not discriminate on the basis of race, color, age, religion, national origin, sexual orientation, gender identity, sex, marital status, disability, or status as a U.S. veteran. Inquiries can be directed to the Director of Equal Opportunity and Diversity, 3680 Beardshear Hall, 515-294-7220.

Schedule of Lectures and Laboratories

Week	Topic	Material
Jan 12	Introduction and Basic Concepts	Moment statistics, distributions, transformations, t-test, correlation, chi-square, Dean's roadmap of inferential statistics
Jan 26	Resampling Methods	Randomization, bootstrap, jackknife, Monte Carlo methods
Feb 2	Univariate ANOVA Models	1-way, factorial, and nested models, Type I vs. Type III SS
Feb 9	Univariate Regression	Regression, multiple regression, ANCOVA
Feb 16	Matrix Algebra and GLM	Deriving ANOVA and regression in matrix form, GLM
Feb 23	Multivariate & GLM	MANOVA, MANCOVA, regression, pairwise comparisons with Mahalanobic distance, etc.; permutational-MANOVA
Mar 2	Interactions & Patterns of Change	Interpreting interaction terms in GLM, why significant interaction terms exist, dissecting patterns of change
Mar 9	Ordination Methods	Principal components analysis (PCA), principal coordinates analysis (PCoA), multidimensional scaling (MDS), Canonical Variates/Discriminant Function Analysis (CVA, DFA), correspondence analysis (CA)
Mar 16	**SPRING BREAK**	SAHN methods (UPGMA, WPGMA,etc), K-means clustering
Mar 23	Clustering Methods	SAHN methods (UPGMA, WPGMA,etc), K-means clustering
Mar 30	Multivariate Association and Canonical Ordination Methods	Mantel tests, canonical correlation, partial least squares, RV coefficient, canonical variates analysis (CVA), canonical correspondence analysis (CCA), redundancy analysis (RDA)
Ap 6	Model Selection	AIC, BIC, LRTmethods
Ap 13	Phylogenetic Comparative Method	Phylogenetic non-independence, accounting for phylogeny in hypothesis testing, phylogenetic signal, rates of evolution
Ap 20	Spatial Statistics	Point patterns, autocorrelation, correlograms, semivariograms, spatial non-independence
Ap 27	**NO CLASS: DCA out of town**	
May 4	Meta-Analysis	Combining results across studies, effect sizes, summary analyses, exploratory and graphical approaches, funnel plots, etc.