BIOL 680: Data Analysis in Biology  
Fall 2010

Instructor: Amy Blanchard  
Class hours: Lecture 9:15-10:15, M W; Lab 2-5 pm Thursday.

Office: O’Neill 127  
Phone: 474-1123, Email: arny@ims.uaf.edu,  
Office Hours: Monday and Wednesday, 1 to 2, by arrangement, as needed.  
Text: Quinn and Keough, Experimental Design and Data Analysis for Biologists.  
Supplementary texts: McCune and Grace, Analysis of Ecological Communities.  
Everitt and Hothorn, A Handbook of Statistical Analyses Using R. Available as a PDF in the R library HSAUR.  
Published papers will be assigned to be read for some topics.

Course Description:  
Data Analysis in Biology is a course into the methods and logic of the wide range of mathematical tools included in the field of statistics as applied to biological questions. My goal for individuals finishing this class is to understand general statistical methods used in research, when to apply them, and how to interpret results. The intent of this course is for you to gain the skills you need to analyze and present your thesis data. The skills you learn here are applicable to your future careers as well. I will place an emphasis on how to selected procedures, assessing statistical assumptions, and professionally interpreting and presenting results. This class will not be mathematically oriented in that the goal is not to teach statistical theory but how to apply and understand results from statistical procedures. The statistical program R will be used. R is a free statistical software program with many libraries added by professional statisticians. MYSTAT, a free program from SYSTAT can also be used.

Computers are used in the workplace and the ability to perform basic statistical tasks on computers will be expected. Completion of this course will contribute to these needs. I use powerpoint for my lectures and work through many examples on the board. What you learn from this class depends on your effort and everyone is capable of passing this class with a good grade.

Prerequisites:  
STAT 200 or equivalent. Expected proficiencies: understanding basic statistics, ability to use computer technology, familiarity with Microsoft Excel or another spreadsheet program.

Course Objectives:  
This class will expand student skills in the application of statistics necessary for starting careers in research. As students are exposed to a variety of statistical methods, they will learn how to correctly apply statistical tests, how to diagnose assumptions, which tests to use, and how to interpret output from statistical procedures. My belief is that everyone can learn how to use statistical methods and understand statistical output if they are willing to do the work. The material can be challenging so it is important to attend lectures, keep up with class reading assignments and complete the homeworks.
Course Readings:
I REQUIRE that you will have read the material for each lecture PRIOR to coming to class. As a result, lectures can focus on the student questions and areas of uncertainty rather than lecturing on topics that students already know. I anticipate asking questions via email or blackboard prior to class to which students will respond providing insights into the topics understood and those that are less clear.

Learning measurement and grading policy:
Success in the class will be measured by computer-based homeworks, exams, and projects. Coursework will be evaluated based on these criteria:
1) Neatness of work and ability to summarize statistical procedures (minimize the output presented).
2) Accuracy of work performed.
3) Completeness of work performed.
4) Accuracy and adequacy of the interpretation of statistical procedures applied.
5) On time.

Grading
A 90 and up
B 80 – 89
C 70 – 79
D 60 – 69
F Below 60

Tentatively, there will be 14 computer lab assignments/homeworks @ 10 pts. each, 4 in-class quizzes @ 10 pts. each, 20 points for class and lab participation, a take-home/lab data analysis project as part of the final @ 100 pts, and three exams @ 100 pts each including the final. The total points for the class are therefore 600 points. One portion of the final may be replaced by a project and presentation based on your research. The actual number of homeworks and quizzes will depend on the flow of the class and the amount of material covered. Homeworks will cover applications of methods discussed, quizzes will cover assigned readings, and exams will seek to evaluate your learning by challenging your statistical decision-making abilities.

Homework
Homeworks will be computer based and are essential to passing the class. Homeworks are assigned on Thursday during lab and then collected the next Thursday. Homework must be legible and stapled. Homework turned in up to 1 class day late, the start of class Monday, will be docked 50%. Assignments turned in after that will be given no score. Exceptions to this rule will be granted if you become ill (proof may be required), have to travel for legitimate work purposes, etc. Make every effort to turn your work in on time! Students may work together, but you must write out your own assignment. Missed homeworks with a valid exception are due within 1 week of returning to class.

Copying of another student’s homework (plagiarism) will not be tolerated. Working together is encouraged but DO NOT copy another student’s work or let someone else copy your work.
Options include a failing grade for the assignment, a failing grade for the class, or removal from the class.

Messy homeworks are difficult to grade and will lose points. Assignments will be posted on Blackboard usually Friday morning and no later than Friday 5 pm.

**Class participation**
I presume that each student is responsible for their time and actions. It is important to note, however, that in general, failure to attend class will be reflected in your class grade. Individuals not attending class regularly miss important information that is usually reflected as poor performance on tests, particularly on the final where information is cumulative. The grade you receive will depend on your effort in learning the material in class, class discussions, and on homework.

Chronically late or 2 zero scores for homework assignments, very poor effort on assignments, or skipping an exam prior to the last drop/add date will result in a faculty withdrawal of the student.

Talking in class, answering cell phones, reading books, or doing homework assignments are disruptive to the class. Class and lab participation points may be subtracted from your score if this becomes a problem. Class participation points are also given for class participation (or negatively, the lack thereof).

**Quizzes**
If you need to take a quiz outside of the scheduled time, you must contact me in advance. Make-up quizzes are usually more difficult than the original quiz and should be taken within a week of the quiz. A zero score will be given for missed quiz without a legitimate reason, missing a scheduled make-up, or failing to make the quiz up within the week. Quiz dates scheduled in the syllabus are TENTATIVE so keep an eye on Blackboard. Quizzes will cover the discussion and lecture materials.

**Projects**
If you have data at hand, you can analyze it as a project that can replace the lab portion of your final. For the project, you need a data set, preferably your thesis data, and you will follow the data through with various statistical methods. The project will be graded on quality of writing, appropriateness of methods chosen, and the overall interpretation.

**The final will be in the regular classroom on the date and time scheduled by UAF and available online.**

**Software**
We will learn to use the R, a FREE program for statistics with many contributed libraries. The class will be very technology oriented. An additional FREE stats program that is likely to be useful is the student version of SYSTAT called MySTAT available at www.systat.com. If you are interested in purchasing a single statistics program for your graduate career and beyond, I highly recommend STATISTICA which can be purchased at an academic discount. You can purchase the base program with the linear models and multivariate add-ons for about $500. If you don’t have a copy of Excel or word, you might try OpenOffice at openoffice.org.
**Blackboard**
This class will make use of Blackboard web-based software. All grades, announcements, handouts, data sets, and lab/homework assignments will be posted there. Everyone in the class should have a Blackboard account ASAP.

**Prerequisites**
STAT 200, STAT 401.

**Attendance, Absences, etc.**
If you have to travel for work or school, become sick, or have another valid reason for missing a homework or quiz, talk to me and we can work something out. I can be flexible when work or life concerns are involved. The university allows instructors to initiate withdrawals for absences and/or lack of performance by the last add/drop date. Any one of these, no attendance, zero scores for two or more assignments, or to miss a quiz and not following up, will lead to instructor initiated withdrawal from the course.

**Ethics**
Students are expected to adhere to the University code of conduct (see UAF catalog). Unethical behavior includes reviewing another person’s homework without permission, copying homeworks, and using unauthorized notes during a test in class. At my discretion, I may take points off or give a zero for the homework, zero for the test, or have the student removed from the class. Repeated offenses of plagiarism by students will result in all students involved being penalized.

**Additional Accommodations**
I will work with the Office of Disabilities Services (203 WHIT, 474-7043) to provide reasonable accommodation to students with disabilities.
<table>
<thead>
<tr>
<th>Week of</th>
<th>Topic</th>
<th>Chapter and sections</th>
</tr>
</thead>
<tbody>
<tr>
<td>September 8-10</td>
<td>Introduction to statistics, Intro to R, descriptive measures in R.</td>
<td>Chp. 1 (1.1-1.3, 1.5) and 2 (2.2-2.3).</td>
</tr>
<tr>
<td>September 13-17</td>
<td>Topics in regression: multiple regression, polynomial regression, crossvalidation, errors in X.</td>
<td>Chp. 5 (5.1-5.4),</td>
</tr>
<tr>
<td>September 20-24</td>
<td>Topics in ANOVA: nested, blocked, repeated measures, permutation tests.</td>
<td>Chp 6 (6.1), Chp. 8, Chp. 9, Topics from Chps. 10-12.</td>
</tr>
<tr>
<td>September 27 – October 1</td>
<td>Nonparametric statistics/Bootstrap methods. (Exam on ANOVA and Regression.)</td>
<td>Sects. 5.1.2, 5.5, 8.5.</td>
</tr>
<tr>
<td>October 18-22</td>
<td>Categorical Regression Trees (CART), randomForest.</td>
<td>Sect. 6.2, papers on BB.</td>
</tr>
<tr>
<td>October 25-29</td>
<td>Introduction to Multivariate Statistics. (Quiz)</td>
<td>Chap. 15 (all), papers on BB.</td>
</tr>
<tr>
<td>October 29</td>
<td>Last day for student or faculty initiated withdrawals.</td>
<td></td>
</tr>
<tr>
<td>November 1-5</td>
<td>Cluster Analysis and ordination (PCA, MDS, CCA, Discriminant Analysis). (Exam on logistic regression, nonparametrics, categorical data analysis, CART and random forest.)</td>
<td>Chap. 17, Chap. 18.</td>
</tr>
<tr>
<td>November 8-12</td>
<td>Ordination (PCA, MDS, CCA, Discriminant Analysis)</td>
<td>Chap. 18.</td>
</tr>
<tr>
<td>November 15-19</td>
<td>Ordination, Multivariate permutation tests. (Quiz)</td>
<td>Papers on BB.</td>
</tr>
<tr>
<td>November 22-26</td>
<td>Methods for correlated data: geostatistics.</td>
<td>Papers on BB.</td>
</tr>
<tr>
<td>November 29-December 3</td>
<td>Nonlinear Regression. (Quiz and Take home assigned.)</td>
<td>Papers on BB.</td>
</tr>
<tr>
<td>December 4 – 10</td>
<td>Remaining Topics.</td>
<td></td>
</tr>
<tr>
<td>December 13</td>
<td>Wrap up.</td>
<td></td>
</tr>
<tr>
<td>December 15</td>
<td>Final Exam: 8 - 10 a.m., Wednesday, Dec. 15.</td>
<td></td>
</tr>
</tbody>
</table>