Spring 2013 BIOL B215: Experimental Design and Statistics

instructor: Joshua Shapiro
office: 112 Park Science
phone: 610-526-5099
email: jshapiro01@brynmawr.edu

Overview & Goals

We have learned huge amounts in biology by doing experiments which generate clear and simple results. But many parts of biology are governed by smaller effects, which are sometimes hard to measure exactly. For example: Does an amino acid substitution change the efficiency of an enzyme? Do female frogs prefer mates with lower-pitched croaks? Does the chance of developing diabetes depend on your genotype at a specific locus? To answer these kinds of questions, we have to use statistics, not only to decide on our answer, but also to make some determination of how much we believe it. To actually use those statistics, you need two things: understanding of the theory, and the ability to apply them. To that end, the primary goals of the course are:

- to introduce to the field of statistics and its uses for describing data and testing hypothesis, with a special emphasis on biological examples.
- to develop an understanding of the considerations of designing experiments to answer biological questions clearly and efficiently.
- to learn how to implement statistical tests and calculations using classical formulae, as well as through computational tools, specifically the R language.

Meeting time & Location

Class meetings MWF 10-11AM; Lab Wednesday 1-4PM
10 Park Science Center

Office Hours & Contact info

Monday & Friday 2–3:30PM, and by appointment.

Feel free to email me any time. I will try to answer as quickly as possible, usually the same day for emails sent before about 7PM.

Textbook & Readings

The textbook for the class is *The Analysis of Biological Data*, by Whitlock and Schluter. It should be available at the bookstore.

You are **required** to complete the assigned reading for the week before class on Wednesday. This is vital for the class to function as a time in which we can go beyond basics in the book to clarify things the book might not explain to your satisfaction, and to explore how the methods and tests described are actually used. There will be occasional brief quizzes on the readings during class.

The book covers most of the statistical methods and theory that we will be covering in class, but it does not cover R at all. There are many good books about R, but some of the best resources are available for free online. I will post links to some of these on Moodle, as well as to notes and tutorials that I have written.

Class Structure

We will spend class time on three main activities: The first is to work through example problems, exploring real biological data to illustrate the statistical concepts presented in the books. In the course of these examples, we will likely encounter areas where it is necessary to expand on the textbook's presentation of a topic, either to clarify or extend the reading.

The lab periods will be primarily devoted to learning R. R is (or is quickly becoming) the standard tool for performing statistics in biology and many other fields. It powerful and flexible, but it is notoriously challenging to learn at first. To get over that hump, we will spend a fair amount of class time working with R, learning our way around the software. This means that you will need to have a laptop at every lab, and I would *highly recommend* bringing one to every class. If you do not have your own laptop, you can borrow one from the department.

Some class and lab time will be devoted to working on problem sets. While there will probably not be enough time for you to fully complete the problem sets within class hours, my hope is to provide the opportunity to get started, and to identify any challenges that I can help with before you get stuck.

Collaboration & Online Resources

Problem sets are designed to improve your understanding of the material, and so all materials you hand in should represent your own work. However, you are allowed and encouraged to discuss the problem sets with your classmates, both during class time and outside of class. In addition, your ability to find and evaluate information on the web is one of the most important skills you can develop. Therefore, you are strongly encouraged to seek out resources online to help you with any aspect of the course. If there is something in the book or about R that you do not understand, chances are somebody else out there didn't understand it either, and as a result there may well be a website out there that can help. *My only requirement is that you acknowledge classmates who you worked with on each assignment and cite any online resources that you used.*

Evaluation & Grading

Problem sets 30%
Quizzes 10%
Exam I 15%
Exam II 15%
Final 20%

Class Participation.....10%

Accommodations

Students who think they may need accommodations in this course due to the impact of a learning, physical, or psychological disability are encouraged to meet with me privately early in the semester to discuss their concerns. Students should also contact Stephanie Bell, Coordinator of Access Services (610-526-7351 or sbell@brynmawr.edu), as soon as possible, to verify their eligibility for reasonable academic accommodations. Early contact will help to avoid unnecessary inconvenience and delays.

Tentative Schedule

(Subject to change; see Moodle for updates)

Week	Dates	Topic	Reading
1	Jan 23-25	Introduction to Data and Sampling Getting started with R	Chapter 1
2	Jan 28-Feb 1	Displaying and Describing Data Calculating descriptive statistics and generating basic plots	Chapters 2 & 3
3	Feb 4-8	Probability Writing R functions and simulating data	Chapter 5 & Sections 7.1 & 7.4
4	Feb 11-15	Estimation and Hypothesis Testing Organizing and manipulating data in R; getting started with plyr	Chapters 4 & 6
5	Feb 18-22	Proportions and probability models	Chapter 7
		Exam I	
6	Feb 25-Mar 1	Probability models & contingency tables Categorical and sampling-based tests in R	Chapters 8 & 9
7	Mar 4-8	The Normal Distribution Simulating data from distributions	Chapters 10 & 11
		Spring Break	
8	Mar 18-20	Comparing samples Fancier plotting with ggplot2 *Dr. Shapiro away Mar 22; no class*	Chapters 12 & 13
9	Mar 25-29	Experimental Design Calculating power and simulating experiments	Chapter 14
10	Apr 1-5	Analysis of Variance	Chapter 15
		Exam II	
11	Apr 8-12	Correlation and Regression Parametric and nonparametric tests of correlation	Chapters 16 & 17
12	Apr 15-19	Bootstrapping and permutation tests	Chapter 19
13	Apr 22-26	Bayesian statistics	TBA (not in the textbook)
14	Apr 29-Jun 3	Catch up and review	