Fall 2015 BIOL B215: Experimental Design and Statistics

instructor: Joshua Shapiro
office: 210 Park Science
email: joshua.shapiro@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 behind the statistics, 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.
- to develop the ability to communicate biological, statistical, and computational concepts, results, and techniques through prose and statistical graphics.

Meeting Times & Locations

Lecture: MWF 11:10AM-12PM; 100 Park Science Center
Lab: Tuesday 1-4PM; 10 Park Science Center

Class Communication

You are welcome to visit me with any questions you might have or just to say hello during my office hours, which are Tuesdays 10 - 11:30AM, and Fridays 2–3:30PM. I will also be happy to meet at other times, but I would request that you set up an appointment by email at least 24 hours in advance.

Feel free to email me any time, though for most course questions I would prefer that you post to Piazza (see below). I will try to answer as quickly as possible during normal business hours, but generally less quickly in evenings and weekends.

We will be using Piazza.com for out-of class questions and discussion, so any questions that might be of general interest should be asked there, rather than by email. This will give you the opportunity to see the questions that other students have already asked, and to answer each other's questions. Active participation on Piazza (questions and/or answers) will count toward your class participation grade. I would also encourage you to use Piazza to post links to articles that you might think are of general interest to the class. There are usually at least a few little statistical controversies that appear over the course of a semester, and if you come across an article about one, it would be great if you shared it with the class.

Textbook & Readings

The textbook for the class is *The Analysis of Biological Data*, 2nd edition by Whitlock and Schluter. It should be available at the bookstore (though cheaper online). The 1st edition is acceptable for readings (though page numbers may have changed), but problems and problem numbers have

changed, so **you must use the second edition for assignments**. A copy of the new edition will be on reserve in the library.

You are expected to complete any assigned readings 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 may 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 (I have suggested O'Reilly's *R Cookbook* by Teetor), 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 try to spend class time on three main activities: The first is some traditional lecture on course topics. We will also spend time working on 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.

Evaluation & Grading

Problem sets 25%
Writing assignments/Lab 30%
Exam I.....10%
Exam II 10%
Final Exam.....15%
Class Participation.....10%

Late assignments will incur a 10% penalty per day unless arrangements are made for an extension at least 24 hours prior to the due date. *All exams and projects are required; failure to complete any component will result in failure in the course.*

Problem Sets

The problem sets are a critical component of the course, and you should expect to devote a substantial amount of time to each one. You should show all work so that partial credit can be assigned if necessary. Problems that involve calculations may be completely by hand or using R, though in some cases I may require that you not use certain R functions, and complete the calculations with intermediate steps shown. When R is used, you will be expected to include all R code with your assignment. I greatly prefer (but do not require) that all assignments be typed and submitted as PDFs or Word documents, and using R (and in particular R Markdown, which we will discuss) will greatly facilitate that for you. Whether handwritten or typed, your assignments *must* be legible and neatly formatted, with all problems in order, and final results and conclusions clearly marked.

Writing Assignments

When performing statistical analysis, it is critical that you not only perform the correct analysis, but also communicate your results effectively. To that end, there will be a variety of writing assignments throughout the course, including tutorials in the use of R, brief descriptions of analyses and results, and a complete lab report with integrated analysis. You will also have the opportunity to read and critique each other's writing, and to revise and improve your work.

Exams

There are three closed-book exams in the course: two during the semester and a final. The exams during the semester will be given during the normally scheduled lab periods, and the final exam is self-scheduled. Just like in life, all exams are cumulative, but will tend to focus on more recently covered material. Any scheduling conflicts must be addressed well in advance of the exam. Make-up exams will not be given, except in cases of illness, or with prior arrangement in the case of extremely unusual circumstances beyond your control.

Collaboration & Online Resources

Problem sets and labs are designed to improve your understanding of the course content, 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. Use Piazza! 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.*

Support Services and Accommodations

Bryn Mawr offers a variety of resources to help students thrive in their academic endeavors while managing stress and maintaining mental health. For further information on those services, consult the Support Services website at http://www.brynmawr.edu/academicsupport/StudentSupportServices.html or contact Rachel Heiser, Academic Support and Learning Resources Specialist in the Dean's Office, with questions: (rheiser@brynmawr.edu, 610-526-5275)

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 Deb Alder, Student Access Coordinator (dalder@brynmawr.edu, 610–526–7351), 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)

Dates	Topic	Reading
Aug 31– Sep 4	Data and Sampling *Lab:* Install R and RStudio	Chapter 1
Sept 7–11	Displaying and Describing Data *Lab*: Getting started with R: descriptive statistics and plots	Chapters 2 & 3
Sept 14–18	Probability *Lab*: Organizing and manipulating data	Chapter 5 & Sections 7.1 & 7.4
Sept 21–25	Estimation and Hypothesis Testing *Lab*: Advanced data manipulation and plotting: Composing an R tutorial.	Chapters 4 & 6
Sept 28 – Oct 2	Proportions and Probability Models **Exam I**	Chapter 7
Oct 5–9	Probability Models & Contingency Tables *Lab*: R tutorial critiques and revision	Chapters 8 & 9
	Fall Break	
Oct 19–23	The Normal Distribution *Lab:* Lizard capture-recapture Part 1	Chapters 10 & 11
Oct 26–30	Comparing Samples *Lab:* Lizard capture-recapture Part 2	Chapters 12 & 13
Nov 2–6	Experimental Design & ANOVA *Lab:* Lizard capture-recapture Part 3	Chapters 14 & 15
Nov 9–13	Correlation and Regression *Lab:* Lab report critiques and revision	Chapter 16 & 17
Nov 16–19	Multivariable Statistics **Exam II: Nov 17**	Chapter 18
Nov 23–25	Bootstrapping & Permutation tests *Lab:* Data cleaning/carpentry	Chapter 19
	Thanksgiving	
Nov 30–Dec 4	Bayesian Perspectives *Lab:* Fitting models to data	TBA
Dec 7–9	Catch up and review	