Statistics/Forestry/Horticulture 571 Fall 2010

Instructors	Bret Larget (lecture 1)	Bret Hanlon (lecture 2)
Office	MSC 1275A	MSC 1241
email	brlarget@wisc.edu	hanlon@stat.wisc.edu

Lecture 1			Lecture 2		
TuTh	9:30 – 10:45am	Nutritional Science 290	TuTh	11:00am – 12:15pm	Russell Labs 184

Discussion sections			Discussion sections		
Tu	1:00 – 2:15pm	Chemistry 2307	Tu	1:00 – 2:15pm	Chemistry 2377
Th	8:00 – 9:15am	Social Science 4314	We	2:30 – 3:45pm	Sterling 1327
We	1:00 – 2:15pm	Noland 579	We	1:00 – 2:15pm	Psychology 134

As much as possible, the two lectures will cover approximately the same material at the same speed. Students may attend the alternative lecture when desired, if there is sufficient space in the lecture hall. The exams, homeworks, policies, and grading will be identical.

Students should select one discussion section to attend regularly; *this need not be the same discussion section in which the student is enrolled, space permitting.*

Discussion sections *will not meet* during the first week of classes or during the two weeks when midterm exams are given.

Course websites: See `http://www.stat.wisc.edu/courses/st571-larget/` for handouts, homework assignments and other course information. See `https://learnuw.wisc.edu/` for the gradebook and discussion forum.

Communication with instructors: Questions regarding course content should be directed to the instructors via the discussion forums in the Learn@UW website. The reasons are that responses from instructors to one student are shared and easily accessed by all, questions may be anonymous, the communication is automatically archived, and it provides an opportunity for students in the course who understand the material well to provide answers from a student perspective. Direct email to instructors should be used only for private correspondence.

Course objectives: The primary goal is to provide graduate students in the biological sciences with a thorough understanding of statistical methodology, so that they may apply it to their own research and answer questions like: Which statistical method is most appropriate? How can I check that the assumptions of the method are met? How do I interpret the results in the context of the biology? A secondary goal is to instruct students in the use of the software package R so that they become competent users for basic analyses by the end of the semester. We will use case studies and genuine data examples whenever possible to illustrate statistical concepts. Mathematical and computational complexity will be minimized when deriving/presenting methods and their justifications; however, students should expect to use both mathematics and computation as necessary for practical data analysis.

Textbook: We have selected the text *The Analysis of Biological Data* by Michael Whitlock and Dolph Schluter as a required text for the course. This text is readable for the audience, is reasonably complete for the topics we wish to cover, contains many interesting biological examples, contains many well-written *interleaf'* essays on important statistical concepts, and does a good job of presenting up-to-date statistical methods. We believe the text to be a valuable resource during the course and beyond. We will not, however, limit ourselves to the topics in the textbook or to the order/fashion in which the ideas are presented. In particular, we expect you to be able to read many early chapters on your own to review/learn the material therein and we will incorporate advanced methods from later chapters as the applications for which they are appropriate are introduced. In addition, we have opted for a *just-in-time* approach for teaching both statistical graphics and probability. Rather than presenting these ideas in a complete chunks (or as complete chapters in the textbook), we introduce methods and concepts from graphics and probability in the context of data analysis or model-based inference appropriate for the statistical topic at hand.

Computing: We will use the statistical package R, which is available for free download online at `http://cran.r-project.org/`. R is available on all commonly used platforms (Windows, Mac and Linux). R is the *de facto* standard statistical computing package used in graduate programs in statistics and among a majority of professional statisticians. Its influence and importance is growing in many other academic fields, including biology, and in major corporations. There have been over 100 books published on using R for statistical computing, including about 30 since 2009, and several are aimed at biological audiences. [1] The basic software has been extended with more than 1000 packages contributed by users, many of which are specifically tailored for use in biological applications.

The course assumes no prior experience with R. Developing proficiency with R requires practice and learning to interact with the computer by typing commands. R includes a high-level programming language. Advantages include an extensive code base that performs many tasks, flexibility to customize functions and graphs, and the ability to keep a commented history of steps of an analysis which allows one to reconstruct and redo analyses from scratch with a simple command.

Learning R does not come easily to all, and while existing help and documentation is quite plentiful, finding help at the appropriate level can be challenging. To help you, we will provide links to R tutorials on the course web page and each section of lecture notes will contain a supplement with R commands used for the graphs and statistical analyses from lecture. Some discussion section meetings will spend time teaching you to use R, so you are especially encouraged to bring your laptop to class these days, if you have one.

Exams: There will be two in-class midterm exams and a 2-hour final exam. All exams will be open book, open notes, and you will be allowed a calculator (but not a laptop). The first midterm exam will be on Tuesday, October 19, and the second midterm exam will be on Tuesday, November 23, the week of Thanksgiving. There will be no homework assigned the week before the midterms. The final exam will be on Sunday, December 19, from 7:45am to 9:45am, location to be announced. Notice of any conflict with these dates must be given to the instructor within the first week.

Grading: Semester grades will be based on homework (20%), two midterms (20% each), and the final exam (40%). Letter grades will not be given for midterm exams, but information will be provided to let you know how you are doing in the class.

[1] See `http://www.r-project.org/doc/bib/R-publications.html` for a complete list.

H1N1 influenza: We are all encouraged to stay home when sick (including instructors and TAs). Colleague coverage will be used as much as possible to continue instruction in the case that instructors become ill. Students who are sick will be responsible for getting class notes that they have missed and for making up assignments or exams within a reasonable period of time. Students do not need to communicate with instructors if they miss class or discussion for sickness. They do need to contact the instructor by email *in a timely manner* in case they need to miss an exam. Students will not need to provide medical excuses for absences from flu-like symptoms. The situation calls for trust and responsibility among all of us.

Homework: Assignments will be assigned and posted on the course webpage on Thursdays, to be handed in by the following Friday by 4pm to your TA's mailbox. Solutions will be posted the following Monday. The mailboxes are in the hallway just inside the main University Avenue entrance to the Medical Sciences Center. You must show your work and organize it to get full credit. Homework should be neat, clearly legible (typed or written legibly), order the problems in the order assigned, and use complete sentences in proper English. Assignment that ask for the use of R should include *short* R commands and output used to answer questions, *but should not include large samples of R output.* Including R output does not preclude the necessity to write out answers: do not expect the grader to find the answers somewhere in an included excerpt of R output.

Late homework will be accepted only under extenuating circumstances and: (1) only with prior notice to and permission from your instructor; and (2) only if it is turned in before solutions are posted. Unexcused late homework will receive no credit. Each student is allowed to drop two assignments (which might have been skipped due to illness or other factors). The lowest two scores will be dropped if a student completes more than the minumum number of assignments.

Academic honesty: You are encouraged to work together with classmates and talk to your teaching assistant or instructor about your homework. We are convinced it is very beneficial to share and discuss ideas. However, you may not present other people's work as your own. If you work with other students solving problems, you still have to write up your own solution independently, run your own R computations independently, and produce your graphs independently. You must work independently during exams. You may not share calculators or pass notes during exams.

Laptop policy: You may enjoy the wireless capability of the classroom so long as you stay on task. Advantages to using a laptop include: taking notes, viewing lecture notes rather than printing them, experimenting with R or other computing software, and so on. There are also limitations; figures and sketches drawn on the board cannot easily be replicated on a notebook in the classroom, for instance. In addition, activities such as emailing, web surfing, and gaming are not allowed in class. These activities are a distraction to classmates: be respectful of others. Be sure the sound is off at the beginning of the class.

Tentative schedule

Day	Date	Topics	Chapter/Section
R	Sept. 2	Introduction; Data	1.1, 1.3, 1.5
T	Sept. 7	Samples and Populations	1.2, Interleaf 2
R	Sept. 9	Proportions: Graphs and the Binomial Distribution, Discrete Distributions	3.4, 5.1–5.6, 7.1, 7.4
T	Sept. 14	Probability: Sampling distributions; Mean; Standard error	4.1–4.2
R	Sept. 16	Proportions: Estimation; Likelihood; Confidence Intervals	7.3
T	Sept. 21	Proportions: Hypothesis Testing; R; Cautions; Extensions	6.1–6.7, 7.2, Interleaf 3
R	Sept. 23	Proportions: Comparing two or more samples	
T	Sept. 28	Probability: Conditional probability; independence probability trees; Bayes' Theorem	5.7–5.10
R	Sept. 30	Contingency tables:	9.1–9.7
T	Oct. 5	Probability: The normal and t distributions continuous distributions	10.1–10.4, 11.1
R	Oct. 7	Probability: Sampling distribution of the sample mean The central limit theorem	10.5–10.6
T	Oct. 12	One sample inference: sampling distributions and estimation	11.2
R	Oct. 14	One sample inference: hypothesis testing	11.1–11.7, 19.1
T	Oct. 19	Midterm 1	
R	Oct. 21	Two sample inference: randomization/permutation tests	19.2
T	Oct. 26	Two sample inference: confidence intervals and testing	12.1–12.9
R	Oct. 28	Power and sample size determination	14.7
T	Nov. 2	Elements of Experimental Design	14.1–14.9
R	Nov. 4	ANOVA: Randomization	19.1
T	Nov. 9	ANOVA: The F-test and estimation of variance	15.1–15.8
R	Nov. 11	ANOVA: Inference for contrasts	
T	Nov. 16	Multiple Comparisons; Bonferroni and false discovery rate	
R	Nov. 18	Probability: Correlation	Interleaf 4
T	Nov. 23	Midterm 2	
R	Nov. 25	Thanksgiving Break	
T	Nov. 30	Regression	17.1–17.10
R	Dec. 2		
T	Dec. 7	ANOVA: Multiple explanatory variables; interaction	18.1–18.4
R	Dec. 9	The Poisson distribution and Poisson regression	
T	Dec. 14	Probability: Goodness-of-fit to a probability distribution	8.1–8.8
Sun	Dec. 19	Final Examination, 7:45–9:45am	