

KEY CONCEPTS: This section summarizes three levels of concepts taught in this course. The first section contains ideas I would like you to remember throughout your life-long career. The second gives concepts that are important to know and use in this course, and the third section lists concepts that I would like you to become familiar with as we work through the course.

Enduring understandings:

1. Most physical, biological, and social processes, produce variable results. This variability can often be quantified in ways that enhance our understanding of the process and facilitate decision making.
2. Statisticians use four steps to quantify and interpret process variability. The four steps are:
 - (1) formulate a statistical question about the process;
 - (2) design a data collection procedure to answer the question;
 - (3) collect and analyze the data;
 - (4) interpret and report the results.

Important to know/be able to do:

1. Distinguish between categorical and quantitative data.
2. Construct and interpret appropriate graphical and numeric summaries of data.
3. Construct and interpret coverage and confidence intervals.
4. Construct and interpret hypothesis tests.
5. Construct and interpret simple and multiple linear statistical models.
6. Carry out, document, and explain a randomization for a survey or an experiment.

Be familiar with:

1. Observational and experimental studies, their advantages and disadvantages.
2. Experimental designs including factorial designs and randomized block designs.
3. Conditional statistics and their interpretations.
4. Data analysis using R statistical software

CLASSES: Classes meet Monday, Wednesday, and Friday in Sabin-Reed 301 at 11 a.m. Lab meetings are in the same classroom. I expect you to attend class and lab. Your participation is an important part of the process of learning statistics, and we need you in class, prepared to participate in discussion. If you cannot attend a particular class I would appreciate the courtesy of advanced notice and an explanation for your absence.

I hope it goes without saying that while the class is in session, you should not use your computer or cell phone for personal email, web browsing, Facebook, or any activity that's not related to the class. Please be sure to bring your book and laptop to class (we'll need to have sufficient numbers of both to allow students to work in pairs).

Policies

LABS: We have two lab sections. Section L01 meets on Mondays at 1:10-2:30, and Section L02 meets on Wednesdays at 1:10-2:30. You have registered for one of these labs. Please attend the lab section you enrolled in. Laboratory activities form an integral part of this course. We use the lab for learning R statistical software, work on projects, and explore statistical concepts in greater depth. This may include collecting and analyzing your own data and writing reports that interpret the results of your analyses. We emphasize writing in non-technical language. We will make homework assignments in lab that will usually be due at the next lab period.

POLICIES:

Attendance: Your attendance in class is crucial, as is your punctuality. We are all going to learn this material together, so we need to have everyone present and working. We will make accommodations for an unavoidable absence if you notify us. One necessary absence during the semester is not unusual; having more than two is uncommon.

Collaboration: Much of this course will operate on a collaborative basis, and you are expected and encouraged to work together with a partner or in small groups to study, complete homework assignments, and prepare for exams. However, every word that you write must be your own. Copying and pasting sentences, paragraphs, or blocks of code from another student is not acceptable and will receive no credit. No interaction with anyone but the instructors is allowed on any exams or quizzes. All students, staff and faculty are bound by the Smith College Honor Code, which Smith has had since 1944. Cases of dishonesty, plagiarism, etc., will be reported to the Academic Honor Board.

Academic Honor Code Statement

Smith College expects all students to be honest and committed to the principles of academic and intellectual integrity in their preparation and submission of course work and examinations. Students and faculty at Smith are part of an academic community defined by its commitment to scholarship, which depends on scrupulous and attentive acknowledgement of all sources of information, and honest and respectful use of college resources.

Assignments [and percent of final grade]:

1. Homework [20%]: Homework is the most effective way to reinforce concepts learned in class. There will be daily and weekly homework assignments. The daily homework will consist of one or two exercises related to the reading you were assigned. It will be due in class and discussed with your group. The weekly homework will have 4 or 5 exercises that typically focus on the material covered in the previous week or so, but occasionally can come from earlier material. Weekly homework will generally be due on Monday by 4 pm in the envelope on my office door (Burton 306). Late homework will NOT be accepted.

We strongly encourage you to complete the assigned exercises and try others from the textbook if you need more practice. You may discuss the homework problems with your colleagues, but you must write up your own solutions. Homework papers that contain several pages should have *your name at the top of each page* and the *pages should be stapled* together. If they are not stapled and a page gets lost, the grader will not be held responsible. Please note that numerically correct answers, alone, are not sufficient. Satisfactory solutions contain an explanation of the reasoning required to obtain the correct answer. I value neatness and clear exposition in your work. *If I cannot read your work, I cannot give you credit for it.*

2. Labs [10%]: Labs provide the opportunity to delve into real data sets and build your computational and analytical skills. All labs will utilize the statistical programming language R. Lab reports will be written in R Markdown and the resulting HTML files will be submitted via Moodle, generally one week after the lab assignment was made.
3. Project & Presentation [20%]: Each group of three students will complete a research project during the term, and you will present your results in a final report and oral presentation. Your group will conduct a statistical investigation of a question of interest to you. Rather than collect primary data, you will use data available on the Internet or from faculty research. You will prepare a project proposal describing your study and obtain approval from your instructor before you begin the investigation. During the last week of class, you (and your group) will give an 8-10 minute oral presentation of your study. The project will give you experience planning a

statistical study, acquiring data, creating and testing a linear model, and writing a technical report. We'll talk a lot more about the project as the semester proceeds.

4. Exams [45%]: There will be three self-scheduled, closed-book exams. You will want to bring a scientific calculator (one that can take logs and exponents), and you may bring a double-sided, handwritten piece of paper with your notes. You may not share a calculator with another student during a test, and you may NOT use your cell phone as your calculator during in-class tests so make sure you have a separate calculator.
5. Participation [5%]: Active participation in class, engagement with the annotated reading notes, and regular attendance will comprise the remainder of your grade.
6. Extra Credit [?]: Extra credit is available in several ways: attending an out-of-class lecture (as will be announced) and writing a short review of it; pointing out a substantial mistake in the book or a homework exercise; drawing our attention to an interesting data set or news article; etc. The extra credit is applied when a student is near the boundary of a letter grade.

Grading: When grading your written work, we are looking for solutions that are technically correct and reasoning that is clearly explained. Numerically correct answers, alone, are not sufficient in reports, homework, tests or quizzes. Neatness and organization are valued, with brief, clear answers that explain your thinking. If we cannot read or follow your work, we cannot give you full credit for it.

RESOURCES:

Moodle: I will regularly update our course Moodle site with lecture handouts, project information, assignments, and other course resources. Lab assignments will be submitted via Moodle. You should check Moodle regularly.

Computing: The use of the R statistical computing environment with the RStudio interface is thoroughly integrated into the course. You have two options for using RStudio:

1. The server version of RStudio on the web. The advantage of using the server version is that all of your work will be stored in the cloud, where it is automatically saved and backed up. This means that you can access your work from any computer with a web browser (Firefox is recommended) and an Internet connection.
2. A desktop version of RStudio installed on your machine. The downside to this approach is that your work is only stored locally, and you will have to manage your own installation.

Note that you do not have to choose one or the other – you may use both. However, it is important that you understand the distinction so that you can keep track of your work. Both R and RStudio are free and open-source, and are installed on most computer labs on campus. Please see the Resources section on Moodle for help with R. We will occasionally use other software available on the Smith network or on the internet, and when we do I will show you how to use it.

Unless otherwise noted, you should assume that it will be helpful to bring a laptop to class. It is not required, but since there are only three workstations in the classroom, we will need a critical mass (i.e. at least 15) computers in the classroom pretty much everyday.

Writing: Your ability to communicate results, which may be technical in nature, to your audience, which is likely to be non-technical, is critical to your success as a data analyst. The assignments in this class will place an emphasis on the clarity of your writing.

Extra Help: There are Statistics TAs available from 7:00-9:00pm on Sunday–Thursday evenings in Burton 301. I strongly encourage you to work with a study group in Burton 301 during TA hours where

you can get answers easily if you are stuck. In addition, the Spinelli Center for Quantitative Learning (2nd Level of Neilson Library) supports students doing quantitative work across the curriculum, and has a Statistics Counselor available for appointments. Your fellow students are also an excellent source for explanations, tips, etc.

LEARNING STATISTICS: is like learning a new language. There are new ideas, new vocabulary and new rules. The pace of this course is moderate but **relentless**, so it is essential that you do not fall behind. You will have something to prepare for class and hand in for almost every class period. The exams, homework and project assignments are scheduled to encourage steady, consistent application on your part.

It's crucially important that you read ahead in the book as well as review the material after we've discussed it in class. There's no better way to learn statistics than to tackle problems in the text, working examples and wrestling with the exercises (many of which have solutions at the back of the book).

We recommend that you form study groups of 3 or 4 students from the class, and get together outside of class to discuss the homework. Each of you should try the problems on your own and then get together to discuss your work. This allows you to develop your own way of thinking about statistics problems before hearing how others think. You will start to gain some self-confidence this way. Some of our most successful students have used the TA hours in the evenings to meet with their study groups when the stats TAs are available to answer questions.

We will spend class time discussing your questions, looking at other examples, and doing activities. We will also show you how to use statistical software to do the extensive computations most statistical problems require. The computer is usually faster and more accurate than we are at doing arithmetic and graphs, but we have to know what arithmetic and which graphs will be useful.

Finally, I recommend that you create a 3-ring binder devoted to this course, so that you can collect all of the course notes, lab notes, hand-outs, project drafts, homework papers, returned tests, and related materials in one place. **Please bring your binder with you when you come to office hours.**

Text last revised: 1-24-16

TENTATIVE COURSE SCHEDULE: The following outline lists each week of class and gives the topics that will be discussed during that week. The reading assignment from the textbook is also given for each week. Please do the reading before coming to class so that you can participate fully in the discussion.

WEEK	TOPIC	READING
1 Jan 25-29	Introduction to Data, Statistical Questions, Collecting Data, Sampling, Study Design	Ch 1
2 Feb 1-5	Examining univariate and bivariate data; both numerical and categorical data,	Ch 1
3 Feb 8-12	Simple linear regression	Ch 5.1-3
4 Feb 15-19	Randomization tests, hypothesis testing EXAM 1 (Ch1 and Ch 5.1-3)	Ch. 2.1-4
5 Feb 22-26	CLT and the Normal Distribution	Ch. 2.5-7
6 Feb 29 – Mar 4	Probability	Appendix A 1-3
7 Mar 7-11	Inference for proportions	Ch 3
Mar 14-18	SPRING BREAK	
8 Mar 21-25	Inference for means EXAM 2 (Ch 2 and Appendix A 1-3)	Ch. 4
9 Mar 28 – Apr 1	Inference for means	Ch. 4
10 Apr 4-8	Inference for SLR	Ch. 5
11 Apr 11-15	Multiple Regression	Ch. 6
12 Apr 18-22	Leave open for new material or catching up EXAM 3 (Ch 3 – 6)	
13 Apr 25-29	Project Presentations	