Math 221: Statistical Inference
MWTh 9:50-11, Science Center E211
Spring 2017 Syllabus

Instructor:
Cassandra Pattanayak, cpattanayak@wellesley.edu
Guthman Director of the Quantitative Analysis Institute
QR and Math Departments

Office Hours:
Mondays 3-4:30, Tuesdays, 2:30-3:30, and Wednesdays 2:30-4, or by appointment. My office, Clapp 238, is in the back of the main floor of Clapp Library, off of Brackett Reading Room, behind the Sanger Room.

My experience is that the concepts in this course are best explained in person. I strongly encourage you to take advantage of office hours, and I look forward to discussing your questions and reactions to the material. Office hours are my favorite!

This course hasn't been offered at Wellesley before. There are no students available as graders or attached tutors, though I hope to provide some student support for R. So, we will depend on each other: I hope that you will come to me with all questions and plan ahead to attend office hours, and I will rely on you for feedback as the course progresses.

Description and Goals: This course introduces the theory of statistical inference: given a data set, how do we estimate the parameters of probabilistic models like those introduced in Math 220? What is the optimal way to make use of the information in our data? Topics include the theories that underly traditional hypothesis testing and confidence intervals, such as maximum likelihood inference and sufficiency. The course will also cover Bayesian techniques for point and interval estimation and resampling approaches, such as the bootstrap.

The goal of this course is for you to gain an understanding of the most important ideas and theories underlying common statistical methods. These ideas are crucial if you continue to study statistics - this course completes the one-year theoretical sequence (220-221) often required for further coursework in statistics - and also very valuable for anyone conducting an applied data analysis, in any context.

At various institutions, this course is called "statistical inference," "mathematical statistics" ("math-stats"), "theoretical statistics," or simply "statistics." At MIT, there is even a course called "Statistics for Applications" using the same textbook we are using, though different chapters are emphasized. In my view, Math 221 differs importantly from applied statistics courses that focus on techniques for analyzing data in real-life contexts: in QR 260 or an intro stats course like Econ 103 or Psych 205 or AP Stats, and in real life, the assumptions underlying the methods we use are never perfectly true for a particular example. In Math 221, we are focused on the theoretical underpinnings of methods, and so we do assume that methods' assumptions are

true. Understanding the theory allows you to fully appreciate conversations about whether and when various methods are appropriate for real-life data sets.

After this course, you should be able to:

- Estimate and generate intervals reflecting uncertainty of parameters from probability distributions
- Understand the conceptual motivation for and differences between method of moments, maximum likelihood, Bayesian, and resampling approaches to estimation
- Apply these approaches to inference to a variety of types of data
- Use the statistical software R to run simulations and implement inference methods
- Make connections between the content of this course and topics in applied statistics from previous or future courses

Prerequisite: Math 220. We will rely heavily on this probability prerequisite. I think of Math 221 as the second half of a year-long sequence. The first half of our textbook covers the prerequisite material, and you should refer to it as needed.

Google Group and Drive: I will use the class google group for announcements. I will post all course documents in a google drive folder that will be shared with the google group. Also, you will submit your assignments electronically by scanning/photographing (or typing) your work and placing the files in a folder you'll share with me.

Computing: We will use the statistical software R. R is popular among statisticians and other researchers because it is free, downloadable, open source, field-neutral, and powerful. In this course, we will primarily use R for simulations that demonstrate the ideas discussed in class, though we will use R for data analysis as well. Because this course is about theory, learning R will not be our focus, and I do not expect you to achieve the same fluency in R that you would in an applied statistics course.

You will receive an invitation to view online modules that introduce R. Computing questions (like conceptual questions) are usually easier to answer in person than by email, so plan ahead and attend office hours.

Required Textbook: *Mathematical Statistics and Data Analysis, 3rd Edition*, John A. Rice.
It is important that you have a copy of the 3rd rather than 2nd edition. I expect you to buy or rent a hard copy or e-book. Buying a new hard copy is $200+; buying a used hard copy or an e-book is $100-$150; renting a hard copy or e-book is $50 or less. Check the publisher's website (cengage.com) in addition to Amazon. A copy of the book is on reserve at the Science Library.

I expect that you will read the book along with the lectures. You may decide that you learn better from reading the text just before or just after we cover the material in lecture. Reading a statistical text is not like reading a novel - you should have paper and pencil on hand and work through the content as you go.

Chapters 1-6 approximately cover the material from Math 220. We will cover chapters 7, 8, 9, 10, 11, and 13. We are skipping chapters 12 and 14 for two reasons:

- We will spend more time on Bayesian techniques than the textbook does. Bayesian methods are a relatively new addition to undergraduate curricula - this topic used to be reserved for graduate-level courses - but the methods are fascinating, increasingly widely used, and closely tied to the content of this class. Therefore, we extend past the Rice book on that topic.
- The topics in Chapters 12 and 14 are covered in other statistics courses at Wellesley. One of the primary purposes of this course is to cover important topics that are not already taught at Wellesley, so this time is better spent on Bayes.

Course Requirements and Grading:

- Assignments (25%): There will be 10 assignments generally due Wednesdays at 11:59 pm, on the dates specified on the Outline. *The lowest assignment score will be dropped when your grade is calculated.*
- Quizzes (5%): There will be 10 untimed, electronic quizzes generally due Mondays at 9 am, on the dates specified on the Outline. I intend that you look over the quiz questions throughout the week and discuss them with your classmates and me. This is a way to emphasize key concepts and make sure that we are ready to move forward on Monday mornings. *Your lowest two quiz scores will be dropped when your grade is calculated.*
- Midterm Exams (20% each): Two in-class exams, **Wed, March 1** and **Thu, April 20**. Please mark these dates on your calendars, as I expect everyone will be present.
- Final Exam (25%): Self-scheduled.
- Participation (5%): I expect that you will be present and engaged at all class meetings and upload your notes for three class meetings in a timely manner.

Policies:

- Quizzes and assignments should be submitted electronically by the due dates and times. Submission instructions will be posted. I will be able to see the timestamp on the files you share with me. You are not expected to type your problem sets (though you are welcome to - let me know if you're interested in learning LaTeX) - you can scan or photograph your work and upload the files. I prefer that you combine the pages into one electronic document. When an assignment has an R component, you will be required to submit your R code in addition to your assignment write-up.
- I will deduct 20% of the possible points from graded assignments for each day of lateness: if the deadline is 11:59 pm on Wed, then assignments received by 11:59 pm Thu can score no higher than 80%; assignments received by 11:59 pm Fri can score no higher than 60%; etc. Similarly, I will deduct 20% of possible points from quizzes for each day of lateness, so that a quiz submitted within 24 hours after the deadline can score no higher than an 80%, etc.
- Extensions on assignments and quizzes and postponement of in-class midterms will be granted only in exceptional circumstances, such as serious illness or a family emergency, with documentation from a dean. If you think you may miss a deadline

because of travel, interviews, a thesis, or other scheduling conflicts, submit the assignment early.

- You may bring two 2-sided pages of notes (on 8.5" x 11" paper) to Midterm 1, four such pages to Midterm 2, and five such pages to the final. You can bring a calculator to each exam. No other resources are allowed.
- Please sign up for three dates to share your class notes. You can scan your hand-written notes or type them. I encourage you to use LaTeX, but you will receive full credit for submitting complete notes in any format. Submit your notes by uploading them to the class notes google drive.

Laptop/Phones: I expect that you will actively engage in class meetings. This means that laptops are used only for note-taking or class exercises and phones are away. My experience is that it is easier to take notes by hand in a quantitative class, and screens can be distracting to others. Please let me know if you plan to take notes electronically.

Honor Code and Collaboration: The Honor Code will be strictly enforced. Except when specified otherwise, all assignments, quizzes, and exams should be submitted individually. However, you are encouraged to (orally) discuss the assignments and quizzes with your classmates. Each student must write up solutions separately. Be sure that you have worked through each problem yourself and that the answers you submit are the results of your own efforts. You also may not share another student's computer code, submit output from another student's computer session, or allow another student to share your code or output. A good rule of thumb: if a classmates asks you if you would like to discuss a problem on an assignment or quiz, you are encouraged to say "yes"; if a fellow student asks to see your answer to a problem or your R code, the answer is "no." **Please explicitly acknowledge collaborators by writing their names at the top of your asssignments.**

Accommodations: If you have documentation from Disability Services or anticipate conflicts with the course due to religious observance or a Wellesley-sponsored activity (such as an athletic team), please let me know early in the semester so that arrangements can be made.