Welcome to Statistics 111

Professor Shane Jensen

The goal of this course is to develop basic tools for data analysis, probability and statistical methods. Key topics covered in the course include exploratory data analysis, regression, probability, estimation, and hypothesis testing.

Syllabus notes: Website

- All handouts will be available on the website:
  http://stat.wharton.upenn.edu/~stjensen/stat111.html
- Website also contains contact information for myself and teaching assistants
- Link on website for getting Wharton class account if you are not a Wharton student
  Helpful if you want to use Wharton computer labs

Syllabus notes: Homeworks

- Homeworks will be handed out every week or two (around 8 HWs in all)
- Homeworks will be submitted during Friday recitation
- No late homeworks will be accepted!!
  - Late homeworks will get a score of zero, without exception
  - Your lowest homework grade is not included in final grade

Syllabus Notes: Midterm Exam

- Midterm is held on following date:
  Monday, February 29 (6-8pm)
- No makeup midterm examination!
  - A missing midterm exam counts as a zero score
  - Do not take this course if you can not attend midterm!

Syllabus Notes: Friday Recitations

- Friday recitations are mandatory: attendance will be taken by your TA. Recitation attendance worth 10% of final course grade
- No excuses for missing recitation are accepted: every missed recitation reduces your recitation score
- However, up to two missed recitations will be dropped from your recitation grade at end of the semester.
- No recitation this week...recitations start on Friday, January 22

Syllabus Notes: Textbook

- Moore and G. McCabe: Introduction to the Practice of Statistics (IPS), 8th Edition is required textbook
- Older editions of this textbook are fine for course material, but homeworks will be out of current edition, so you will need access to current edition for homework questions
Syllabus Notes: R

- The statistical software package R is the supported software for course. You will need to use statistical software for some homework questions
- R is free and can be downloaded for any operating system at: https://www.r-project.org/
- Friday recitations will be, in part, devoted to questions and issues about using R

Student Questionnaire

- Fill out a questionnaire and hand it in at the end of lecture
- I will try to incorporate some of the subjects that interest you into future lectures

Course Overview

- First few lectures will not involve much mathematics at all: focus on intuition and basic principles
- Then, we get to probability and inference, which most students find to be the difficult part of the course
- Second half of course is hypothesis testing and confidence intervals where all our methods come together

Out in public: You teach statistics ?!? 

- I hated that class in college!
- That was the most boring class ever!
- It didn’t connect with anything else I was doing!

Big Picture Ideas

- Statistics is all about uncertainty
  - Focus as much on what we don’t know (or haven’t observed) instead of what we know
- Formulating the question that we want to answer is often the most difficult part
- Unlike mathematics, no set of rules that you can just follow…you actually have to think!
Science and Skepticism

- We always need to be cautious about conclusions based on data
  - possible sources of bias and confounding
  - how might things have gone wrong?
- A little bit of skepticism is a good thing!

Statistical Modeling

- **Inference**: using mathematical models of uncertainty to answer questions
  - connecting probability concepts to our data
- Can not make claims without using models and making assumptions
  - Are the assumptions reasonable?

Next Class (Lecture 2)

- Collecting Data: Design of Experiments
- Sections 3.1 – 3.2 in Moore and McCabe