Statistics for Data Science
Week 1 Live Session Plan

May 7, 2018

1. **What's statistics? What's this class?**

2. **Instructor Introduction**

    (a) Office hours Tuesdays 6:00pm - 7:00pm

3. **Polls**

    (a) How much experience do you have with Statistics?

    (b) What are you most apprehensive about in this class?

    (c) What's the most advanced math class you've taken?

4. **Student Introductions** Please take just 2 minutes to tell us a tiny bit about yourself.

    (a) Your job / career interest / reason you're learning data science

    (b) Something very few people know about you

5. **Weekly Workflow**
   A typical week of the course proceeds as follows:

    (a) Before live session: Watch all async content

    (b) In live session: We will build upon the async to test and extend your understanding

    (c) After live session: complete the homework for the unit (due 24 hours before next live session)

6. **Homework**
   Homework is graded only with a check, check-minus, or a zero. Each week, we will have one or two students present their homework solution from the previous unit. Please take just 10 minutes to present your work. A further 5 minutes will be reserved for questions. Your homework does not need to be correct, but we expect you to be prepared to lead the class discussion, and this will count as part of your participation score. Even though homework is not graded closely, it is the best preparation for labs and exams, so we encourage you to work hard on it.

   Choose a week to be a homework presenter and sign up on our googledoc.

7. **How to Succeed in this Class**
   Here are some strategies we think will help.

(a) Get to know your readings, especially the Devore and Wooldridge textbooks. (We may occasionally ask questions that you can only answer if you keep up with the readings)

(b) Do the best job on the homework you can.

(c) Strategize about the homework exercises with friends as much as you want, but write the final proofs and final scripts by yourself. (However, note that you are not allowed to discusses quizzes or labs with anyone else unless otherwise specified)

(d) Form study groups!

8. **Discussion Questions** Take 5 minutes to discuss these in breakouts, before our class-wide discussion.

(a) Why do you need to know classical statistics, when we have fancy machine learning algorithms that can classify and predict with amazing accuracy?

(b) Why isn't descriptive statistics enough for your career?

(c) What mistake might result from confusing a sample mean for a population mean?

9. **Software**

(a) Apart from today I will use jupyter notebook with an R kernel to instruct you. In order so that you can participate in class you will need to install applications and packages in order to use these notebooks.

(b) Every system is different and in the life of a data scientist figuring out how to get and install useful software packages from tutorial and question/answers on the internet is an extremely valuable skill.

(c) Please do this by next week.

(d) One possible way to do it.
   - Install R from CRAN: https://cran.r-project.org/
   - Install Anaconda: https://www.anaconda.com/download/#macos
   - Update conda using conda: "conda update conda"
   - Install R-essentials package using conda: "conda install -c r r-essentials"
   - Install any other packages from CRAN or from Anaconda Cloud

(e) You may find R studio (https://www.rstudio.com/products/rstudio/download/ or in Anaconda) useful to complete your homework.

10. **Reminders**
Before next week:

(a) Complete Unit 1 homework (due 24 hours before next live session).

(b) Pre Class Exercises

(c) Watch all unit 2 async content.

Lab 1 will be distributed after Week 2 live session, and you will have 2 weeks (minus 24 hours) to complete it.