

1. Pass out syllabus
2. Read through syllabus
 - explain extra credit
 - read example article
3. Statistics is not math. Success in this class is dependent on your doing the homework. Don't get frustrated if you don't understand the material initially, it may take a while. If you don't understand anything, come up to the loft and talk to me, I'll be around most of the time (and definitely during office hours).
4. Why do we need statistics?
 - (a) Interpret results from medical studies
 - (b) Read and critique scholarly research
 - (c) Understand business processes and make important decisions
 - (d) Book focuses mostly on formal research and experiments. Statistics can also be used to analyze data gathered in other ways (for example: sports data, student enrollment, bank account info, etc.).
5. Three aspects to statistics
 - (a) Design
 - How to conduct an experiment where the results are trustworthy.
 - How to select subjects of an experiment
 - Determine goals of analysis
 - (b) Description
 - Summarizing data
 - (c) Inference
 - Making decisions or predictions from data
 - Voter example : In 2003 an exit poll survey of 3160 voters in California. 54% voted to recall 46% voted not to. Stating the results is description, predicting the outcome for all 8 million voters is inference. In order to predict the outcome, we need to ask the question "If the recall were actually supported by more than half of the population, how likely is 54% in the exit poll?"
 - (d) Description and inference are types of analyzing data. Long before the data is analyzed, we need to give careful attention to what questions the analysis is supposed to answer.
6. Population vs Sample

Population the total set of subjects in which we are interested

Sample the subset of the population for whom we have (or will have) data

Use exercise 1.1 to identify the sample, the population and the inference. In this study 22,000 physicians were chosen to participate. Aspirin takers were chosen by flipping a coin, 11,000 were chosen to take aspirin, the other 11,000 were not to take aspirin. The results showed that 1.7% of the non-aspirin takers had a heart attack during the trial and 0.9% of the aspirin takers had a heart attack. Define the population, sample and inference of this study.

Parameter a numerical summary of the population

Statistic a numerical summary of a sample taken from the population

Random Sampling subjects chosen from a population at random such that each subject has the same probability of being included (follow up on aspirin example)

Now we have a new definition for inference.

Inference methods of making decisions or predictions about the population from the sample.

7. We will cover most of the text. I am thinking of skipping ch 4 so that we can cover chapter 11 on regression.
8. There is a section in chapter 1 about computers. Statistics cannot be done today without computers. You will see why when we start calculating things like means, medians, standard deviations and correlations. Data sets tend to be extremely large and dealing with these large data sets is impossible without modern computers. The book suggests a few software packages such as MiniTab, SPSS and Excel. You'll find that the most often used (though not the best) is Excel. It will handle all of the techniques that we will talk about (although Excel may not be the best choice for advanced statistics). MiniTab is almost never used. SPSS is used (at the U at least) mostly by social scientists and occasionally by people in marketing (also social scientists). SAS is another popular package used in science and industry. There is an open-source package called R that is used in academia and the sciences occasionally. I suggest that at the minimum you learn to use Excel (we may have some computer based exercises later in the semester). All of you have access to Excel in the Multimedia center in the library.