Chapter 1: Basics For Conducting Statistical Studies

Statistics → A collection of procedures and principles for gaining and analyzing information in order to help people make decisions when faced with uncertainty.

ex. Choosing which route to take to school

In a statistical study we are looking to determine relationships or patterns in population, the group of interest in the stat. study. However, we notice that in most cases it is not plausible to study all of the population, this would be too time consuming. We must study a sample of the population.

ex. Do teenage drivers have more accidents than elderly drivers within any given year?

* 3 Basics for conducting a statistical study:

1. Representative sample
   Subjects in the sample must represent the larger population of interest. We will talk more on Monday about how to select a proper sample.
   Non rep sample: Students of this class and senior citizens in a nearby nursing home

2. Sample size
   Know that looking at just one teenager and one elderly person would not be a "large enough" sample. However, would looking at 100 of each be "large enough"? How do we define what is large enough.
   The sample’s size depends on the variability of the information being measured or how diverse the measurements are within each group.
   Our example: If we knew that the range of accidents for teenagers was between 1 and 2 and the range for elderly was between 8-10 it would not take a very large sample to detect a difference. However, if the range for teens was 1-11 and elderly was 3-15, it would take more samples to detect this subtle difference.

3. Observation versus randomized experiment
   In an observational study we simply observe things about our sample. As in our example where we observe how many accidents the teens and elderly have within a given year. Notice in observational studies we can not conclude anything about causation. Ex. Smoking and lung cancer
   In a randomized experiment when we want to imply causation we randomly assign people to one of two groups where the two groups differ only by one aspect... what we generally call the "treatment". Ex. Does caffeine cause an increase heart rate. Here, the difference is that the increase heart rate would only be due to the caffeine consumption since the two groups would be chosen so that nothing else differed between them.

Causation A study shows strong positive correlation in children between shoe size and vocabulary. Does this mean that feet cause children to do better in school? Data shows that students with tutors have lower test scores than those without. Is the tutor causing the student to perform poorly in class? Both these cases were observational studies: we can clearly see that we can not imply causation from an observational study.
Chapter 2: News Reports of Statistical Studies

When we are presented with a study in the newspaper we never actually get to see all of the data from the study we only get to see either a summary of the data or a conclusion based on the data. However, in both situations we should be skeptical of the results. Present 7 key components that should be presented in all newspaper reports and discuss what we should infer from these components.

★ 7 Key Components of a Statistical Report

1. Source of Research and of Funding
   This is important to detect if the study was funded by an organization that would be biased to a particular outcome.

2. Researchers who had contact with participants
   Participants give answers based on the questioners implied expectations. Therefore, we would want all people that are in contact with the participants to be as blind as possible to the details of the study.

3. Individuals studied and how they were selected
   So that we can know if the sample is representative of the entire population. Also, if the study claims to be a randomized experiment, was the sampling method really random? We will learn more about random sampling next week.

4. Exact nature of measurements made
   We want to precisely define what we want to measure.
   Ex. With the teen and elderly drivers: we would want to define specifically what an accident would mean. Would it be an minor collision or only those that caused major damage.

5. Setting in which measurements were taken
   This can help us see if there were any bias in the measurement process.
   Ex. With the teen and elderly drivers: we would not want to only sample say the teens at night and the elderly during the day. Or sample the teens in a city where that is pron to bad whether and driving conditions and elderly in a city that has better whether.

6. Differences in groups being compared
   What other factors may have influenced the comparison (we call these confounding factors)
   This is the key in the difference between observational studies and randomized experiments. In observational studies such as the ones mentioned above because of confounding factors in the observational studies we can NOT imply causation. In randomized experiments we hope to be able to control for these confounding factors and thus we can imply causation if the study is done correctly.
   Ex. With the teen and elderly drivers: other such factors would be how often or how many miles each group drives.

7. Extent or size of any claimed effect or differences
   Are the results of practical importance.
Ex. Teens are 3 times more likely to be in a car accident than seniors ... versus ... Teens are less safe drivers than seniors.