

Chapter 1

ST 511 - Introduction

Readings: Chapter 1 (for 1.3 just read the 2 that interest you the most)

_____ - the science of designing studies or experiments, collecting data and modeling/analyzing data for the purpose of decisions making and scientific discovery when the available information is both limited and variable.

Why learn statistics?

- We live in a society that collects volumes upon volumes of data.
- Are people looking at the data?
- Are they interpreting the data properly?
- How do we turn raw data into information?
 - to make new policy
 - to make better product
 - to increase yield

Statistics is often called the ‘science of learning from data.’

ex: Gas mileage

Suppose we fill 20 of the same model of car with a full tank of gas. Each car will have a different miles per gallon.

Why?

Factors that affect gas mileage:

To summarize the information from the 20 cars we might look at the **average** gas mileage of the 20 cars.

Questions to answer:

- How do we obtain an overall average miles per gallon for this model of car? (Not just for these 20.)
- overall average when driving in city?
- overall average when driving on a highway?
- overall average with low tire pressure?
- overall average when in a city with low tire pressure?

Statistics provides a framework for solving this type of problem!

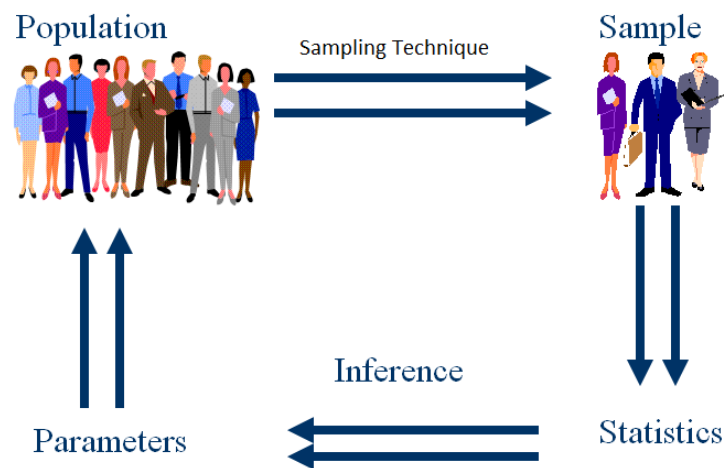
Method of statistics often follows a 4 step process

- Step 1: Identify the research objective
- Step 2: Collect the information needed to answer the questions
- Step 3: Organize and summarize the information.
- Step 4: Draw conclusions from the information.

(Repeat as necessary to answer research objective.)

Big ideas in stats:

- _____ - all the values, items, or individuals of interest
- _____ - a (usually) unknown summary value about the population
- _____ - a subset of the population we observe data on
- _____ - a summary value calculated from the sample observations



Gas Example:

What is the population, sample, parameter of interest, and statistic (most likely to be used)?

Common Notation in statistics:

Name	Parameter	Statistic	Quantity Measured
Mean	μ	\bar{Y} or \bar{y} or \bar{X} or \bar{x}	Center or Location
Proportion	p or π	\hat{P} or \hat{p} or $\hat{\pi}$	Location or Frequency
Standard Deviation (SD)	σ	S or s	Variability or spread
Variance (Var)	σ^2	S^2 or s^2	Variability or spread

Note: $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ and $S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$ where n is the sample size (or number of observed values in the sample).

Many, many, many, more to come!

Question of interest will lead you to which parameter you have interest in. This will also most likely lead you to which type of data you will collect.

Scales (Types) of Data:

- _____ - A variable that is described by attributes or labels
Subscales:
Nominal - categories have no ordering (Male, Female) (zip codes)
Ordinal - can order categories (Lickert scale data) (college football rankings)
- _____ - A variable that is described by numerical measurements where arithmetic can be performed
Subscales:
Discrete - finite or countable finite number of values (# of flowers on a plant, 0, 1, 2, ...)
Continuous - any value in an interval is possible (Temperature, $(-459.67 \text{ deg } F, \infty)$)
(Some lump these together and call them interval.)

How we summarize and analyze the data will depend on which type of data we have.

ex: SAT (get to know each other a little!)

- 50 total students (16 males and 34 females) where matched on socio-economic background (all had similar income).
- A study was done to examine the effect of classroom atmosphere on SAT scores.
- Two types of classrooms were investigated (strict vs easy going).
- Students were divided into two groups of 25 (12 males and 13 females in strict class and 4 males 21 females in the easy going class).
- After the class the SAT was taken (although 1 in the strict class did not take the exam and 5 in the easy going class did not take the exam).

With a partner or two (introduce yourselves):

1. Determine the research question.
2. Define the population and sample.
3. Define possible parameter(s) of interest.
4. Define possible statistics that might be calculated.
5. Why might the students have been matched on socio-economic background?
6. What issues might you see with the design of this study?
7. What other variables might you collect?