



Chapter 1: Introduction to Statistics

1.1 Statistical and Critical Thinking

The SITUATION	The PROBLEM	The (IMPERFECT) SOLUTION
We want to know <i>SOMETHING</i> about a <i>CERTAIN GROUP</i>	Most of the time the <i>GROUP</i> we're interested in is _____.	Don't ask <i>EVERYONE</i> about <i>WHAT YOU'RE INTERESTED IN</i> , just ask _____ and infer.
Ex: Of UCLA undergraduate students, what proportion likes poke bowls?	Ex: As of 2017, UCLA had about _____ undergraduates.	Ex: Ask _____ UCLA undergraduates if they like poke bowls and use the collected _____ to infer about the total proportion of undergraduates who like poke bowls.

Def Statistics

<p>POPULATION: the _____ group to be studied.</p> 	<p>SAMPLE: a _____ of the population that is being studied.</p> 
--	--

EX: You are walking down the street and notice that a person walking in front of you drops \$100. Nobody notices the \$100 except you. Since you could keep the money without anyone knowing, would you keep the money or return it to the owner?

Let's say you want to do a study to gauge the morality of the students at Pasadena City College by determining the percent of students who would return the money. You survey fifty students, and thirty-four of them say they would return the money.

In this example, what is our population and what is the sample?

Population:

Sample:

Data:

<p>Descriptive Statistics: organizing and summarizing the data.</p> <p>(Chapters 1- 6)</p>	<p>Inferential Statistics: uses methods that take results from a sample, extend it to the population, and measure the reliability of the result.</p> <p>(Chapters 7 - 12)</p>
---	--

We said statistics begins with wanting to know *SOMETHING* about a *CERTAIN GROUP*. Well, the *CERTAIN GROUP* is our _____ and the *SOMETHING* is the _____.

Def Variable

EX: The Gallup Organization contacts 1028 teenagers who are 13 to 17 years of age and live in the United States and asks whether or not they had been prescribed medications for any mental disorders, such as depression.

Population:

Sample:

Variable:

1.2 Types of Data

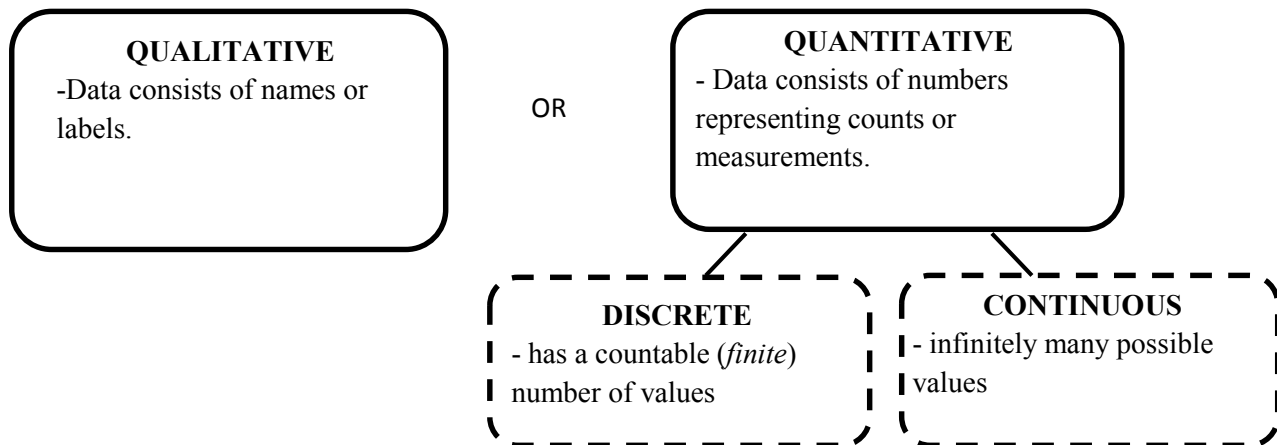
Def **Data** are _____ of observations.

Def A **parameter** is a _____ describing some characteristic of a _____.
Number is _____ based on the entire population.

Def A **statistic** is a _____ describing some characteristic of a _____.
Number _____ based on the sample.

EX: Identify whether the underlined value is a parameter or a statistic.

<p>a) Following the 2018 national midterm election, <u>23.4%</u> of the representatives in the U.S. House of Representatives are female.</p>	<p>b) In a 2015 national survey of high school students (grades 9 to 12), <u>15.5%</u> of the respondents reported that they had been cyber-bullied.</p>
<p>c) Only 12 people have walked on the moon. The average time these people spent on the moon was <u>43.92 hours</u>.</p>	<p>d) A study of 6076 adults in public restrooms (in Atlanta, Chicago, New York City, and San Francisco) found that <u>23%</u> did not wash their hands before exiting.</p>



EX: Determine whether the following variables are qualitative or quantitative.

- | | |
|---|----------------|
| a) Gender | b) Temperature |
| c) Number of days in the past week that you studied | d) Zip code |

EX: Determine whether the quantitative variables are discrete or continuous.

- The number of heads obtained after flipping a coin five times.
- The number of cars that arrive at McDonald's drive-thru between 12:00 pm and 1:00pm
- The distance a 2014 Toyota Prius can travel in city driving conditions with a full tank of gas.
- The average test score on the first Statistics exam in a class of 35 students.

***Statistically Significant** is achieved in a study when we get a result that is very unlikely to occur by chance.

Practical Significance looks at whether the difference is large enough to be of value in a practical sense.

EX: In a study of the Gender Aide method of gender selection used to increase the likelihood of a baby being born a girl, 2000 users of the method gave birth to 980 boys and 1020 girls. Would you pay \$50,000 to use this method?

LEVELS OF MEASUREMENT

Def **Nominal** Level of Measurement - Data that consists of names, labels, or categories only. However, data cannot be arranged in an ordering scheme or hierarchy.

Def **Ordinal** Level of Measurement - Categorical data that can be arranged in some order, but differences cannot be determined or are meaningless.

Def **Interval** Level of Measurement - Numerical data in which the difference between any two data values is meaningful. However, there is no natural zero starting point and ratios are meaningless.

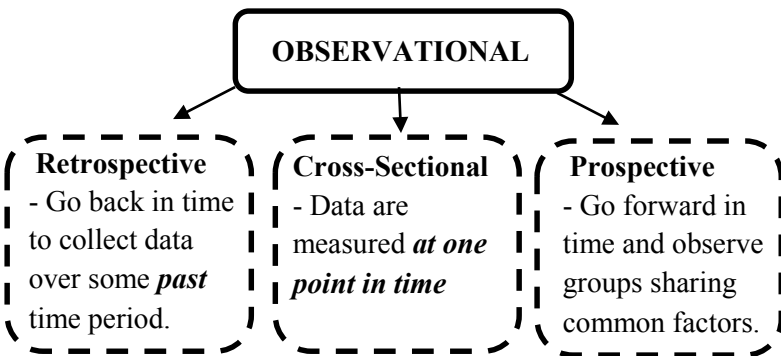
Def **Ratio** Level of Measurement - Numerical data with a natural zero starting point and ratios are meaningful. Zero indicates that none of the quantity is present.

EX: **Determine the level of measurement of each variable.**

- a) Nation of origin
- b) Movie ratings of one star through five stars
- c) Volume of water used by a household in a day
- d) Year of birth of college students
- e) Highest degree conferred (high school, bachelor's, and so on)
- f) Eye Color
- g) Assessed value of a house
- h) Time of day measured in military time

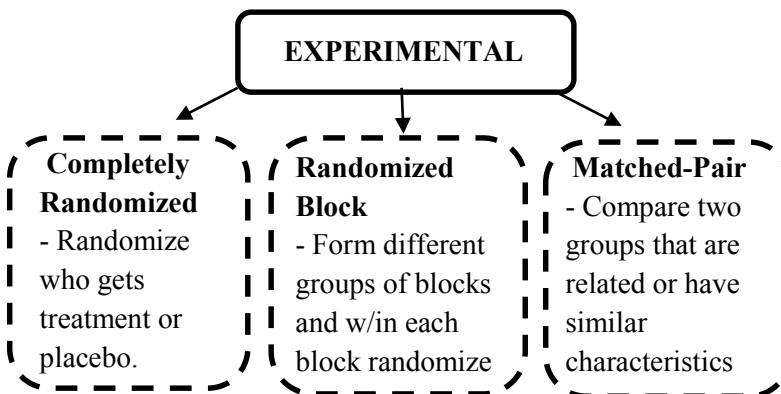
1.3 Collecting Sample Data

OBSERVATIONAL STUDY	DESIGNED EXPERIMENTAL STUDY
<p>*No intervention (Researcher has no control)</p> <p>-Study that looks at data that has already been collected or as it occurs naturally</p> <p><u>Ex:</u></p> <p>(+) Less Expensive, (-) Can't claim causation, only association.</p>	<p>*Intervention (Researcher has control)</p> <p>-Study that applies some treatment and then proceeds to observe its effects on the individual.</p> <p><u>Ex:</u></p> <p>(+) Controls unknown variables (-) More costly and sometimes unethical</p>



Example: Determine which type of observational study is shown below:

Samples of subjects with and without heart disease were selected, then researchers looked at what the subject did ten years ago to determine whether they took aspirin on a regular basis.



Example: Determine which type of experimental study is shown below:

A clinical trial of Lipitor treatments is being planned to determine whether its effects on diastolic blood pressure are different for men and women.

RANDOM SAMPLING METHODS

Regardless of whether or not a researcher decides to use an observational study or a designed experiment, a sample group needs to be chosen to best represent the population. The **BEST** way to choose a sample is to _____ select individuals to stay _____.

Def **Random Sample** - Members from the population are selected in such a way that each individual member in the population has an equal chance of being selected.

Ex:

Def **Simple Random Sample** - A sample of n subjects is selected in such a way that every possible sample of the same size n has the same chance of being chosen.

Ex:

Ex: Randomly sample six of the following companies for a survey on profit margins.

- Alaskan Airlines
- Akoa
- Ashland
- Bank of America
- BellSouth
- Chevron
- Citigroup
- Delta Airlines
- Disney
- DuPont
- ExxonMobil
- General Dynamics
- General Electric
- Clorox

Using a random number generator, the six randomly sampled companies are:

Def **Systematic** Sampling - Select some starting point and then select every k^{th} element in the population.

Ex:



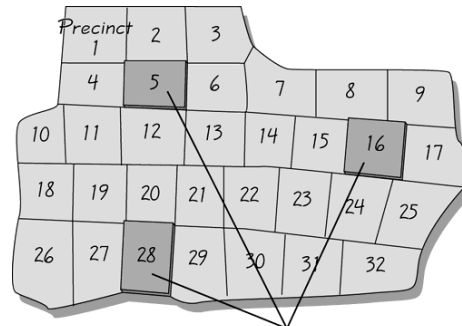
Def **Stratified** Sampling - Subdivide the population into at least two different subgroups that share the same characteristics, then draw a random sample from each subgroup, proportional to the population (or stratum).

Ex:



Def **Cluster** Sampling - Divide the population area into sections (or clusters). Then randomly select some of those clusters. Now choose all members from selected clusters.

Ex:



Interview all voters in shaded precincts.

Def **Convenience** Sampling (**non-random**) - Use results that are easy to get.

Ex:

