Sec 1.1 – Inferences & Conclusions From Data
Statistical Studies & Definitions

Name:

1. POPULATION: A population is the entire group of objects being studied or it could he considered a collection of all outcomes, responses, measurements or counts that are of interest.

Parameter [Greek variables] A measure or count of the entire population.

2. SAMPLE: A subset of the population.

Statistic [English variables] **:** A measure or count of the a sample.

3. List reasons why you might use a **SAMPLE** study instead of a **POPULATION** study?

- USING A SAMPLE IS QUICKER
- USING A SAMPLE IS MORE COST EFFECTIVE
- SOMETIMES IT CAN BE MORE ACCURATE

4. A recent survey by the alumni of a major university indicated that the average salary of 8,500 of its 250,000 graduates was $123,000. Does this value describe a parameter or a statistic? WHY?

↑ SAMPLE STATISTIC B/C THE $123000 WAS A MEASURE OF THE SAMPLE (8500)

5. A survey of 976 American households found that 32% of the households own two cars. Identify the population and the sample.

SAMPLE: THE ACTUAL 976 AMERICAN HOUSEHOLDS THAT WERE INCLUDED IN THE STUDY.

POULATION: ALL AMERICAN HOUSEHOLDS

For # 6– 8 Identify each of the following data sets as either: (P) Population or (S) Sample

S 6. ….the age of a few randomly selected participants in a study about a race of runners

P 7. ….the annual salary of each full-time teacher in a study about Phoenix High School

S 8. ….a survey of 750 Georgia homeowners in a study about all of Georgia's homeowners.

For # 9 – 11 Identify each of the following numerical values as either: (P) Parameter or (S) Statistic

POPULATION SAMPLE

S 9. …. of a company's employees the opinion of just those that were there on time one morning about what they thought of a new training program.

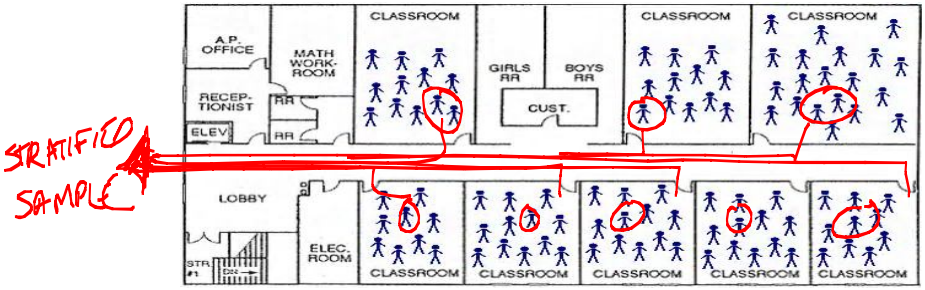
P 10. …. in a study about a small company of 25 employees, the range of their employee's salaries

S 11. …. in a study about the value of American homes in 2012, the average decrease of all the homes sold in Gwinnett.

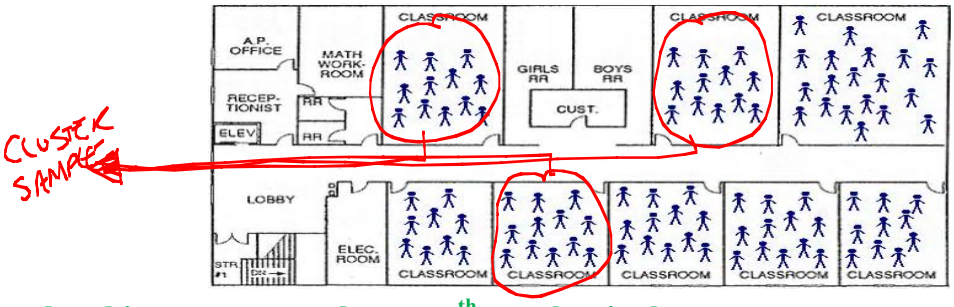
TYPES OF SAMPLES

12. Simple Random sample: A sampling method in which each member of the population has an equal chance of being selected to be a part of the sample. Usually a sampling frame is used in which each member of the population is listed and assigned a number. Then, numbers are randomly generated to select for the sample from the frame.

13. Stratified sample: A population is divided into at least two groups, then some members (but not all) are randomly selected from each group.



14. Cluster sample: A population is divided into groups, the all of the members in one or more (but not all) of the groups are selected



15. Systematic sample: The population is ordered in some way and every n^{th} member is chosen

16. Convenience sample: Sampling members from the population who are readily available or 'convenient'.

Choose which sampling technique is used.
(R) Random (STR) Stratified (CLS) Cluster (CON) Convenience (SYS) Systematic

STR 17. There are 250 seventh graders and 300 eighth graders at Generic Middle School. We ask 45 seventh graders and 50 eighth graders how many siblings they have to compare the two groups.

CLS 18. I ask all freshmen, no sophomores, no juniors, and all seniors if they prefer Vanilla or Cherry Coke (these four groups are my only four groups) to create a study of what should be in the vending machines.

CON 19. I ask everyone in my 5^{th} period class who has more than one computer at home in a study about all of my students for the year.

SYS 20. I collect data from every 15^{th} student on my list of the entire school population.

RAND 21. After using a random number table to generate two-digit numbers, I decide on 10 people to choose from the population.

Rank the sampling types in order from what would usually be the WORST to BEST representation of a POPULATION. Provide brief explanations (especially if the ranking depends on the study).
(R) Random (STR) Stratified (CLS) Cluster (CON) Convenience (SYS) Systematic

CONVENIENCE → STRATIFIED / CLUSTER → SYSTEMATIC → RANDOM

TYPES of STUDIES and DATA COLLECTION METHODS

1. **Observational Study:** This type of study attempts to understand a cause and effect relationship but the researcher is not able to control how the groups are assigned or the treatment each group receives. The researcher also usually attempts to minimize any influence the study may have on the subjects.

2. **Experimental Study:** This type of study also attempts to understand a cause and effect relationship. The researcher usually selects groups from the population and assigns treatment to one or more groups. The researcher also commonly uses a control group that receives a placebo.

   Treatment Group: The group that actually receives real treatment.

   Control Group & Placebo: The group that receives a fake treatment, called a placebo.

3. **Simulations:** A way to model random events in a statistical study, such that simulated outcomes closely match real-world outcomes in a safer or more efficient way.

4. **Census:** A study that uses counts or measures of the entire population.

5. **Sampling:** A study that selects a subset of the population to estimate the characteristics of the whole population.

Choose the type of Study that is most likely to be used (each is used just once).
(E) Experimental (SIM) Simulation (C) Census (SMP) Sampling (O) Observational

C 6. You want to know how many pets the teachers at Phoenix High School own.

E 7. A drug is given to 15 patients and a placebo to another group to determine its effect on an illness.

O 8. You are doing a study at a mall in which you are counting the number of men that wash their hands after using the restroom.

SIM 9. You want to know the g-forces a person would experience during a fall from a 90 foot high bridge into a lake.

SMP 10. You need data on the average number of hours worked per week by an American teenager with a part-time job.

11. Define Data Types.

a. Qualitative: **is information about qualities; information that can't actually be measured. Usually it is described as a categorical description or name.**

b. Quantitative: **is information about quantities; information that can actually be measured. It is a numerical description where the numbers provide measurement.**

For numbers 12 - 20 choose (QL) Qualitative or (QN) Quantitative

QL 12. The colors of automobiles on a used car lot. CATEGORY NOMINAL

QN 13. The number of seats in a movie theater RATIO

QL 14. Numbers on shirts of a girls' soccer team. NUMBER HERE ACTS MORE LIKE A NAME. NOMINAL

QN 15. Ages of the students at North High School. RATIO

QN 16. The temperatures of 30 refrigerators. INTERVAL

QN 17. The amount of fat grams of 24 different cookies. RATIO

QN 18. The years the Olympics were held in the United States. (1984, 1996, ...) INTERVAL

QL 19. Marriage status (married, single, divorced). CATEGORY ORDINAL

QL 20. Social Security Numbers of the employees of a school. AGAIN THIS NUMBER ACTS MORE LIKE A NAME. NOMINAL

11. Bias: **refers to the tendency of a measurement process to over- or under-estimate the value of a population parameter. It is commonly the reason your sample data may not represent the population very well**

- **Sampling Bias:** **is a bias in which a sample is collected in such a way that some members of the intended population are less likely to be included than others.**
  - TO BEST AVOID SAMPLE BIAS USE RANDOM SAMPLING
  - SAMPLE OF CONVENIENCE LIKELY TO HAVE SAMPLE BIAS.

  Which would most likely be the **best representative sample** and which would be the **worst sample** to use in determining the voting preference for the next president in the city of Lawrenceville?

NOT GOOD A. A reporter asks everyone in front of the court house who they plan on voting for and keeps a record.

BEST (B) An analyst gets a spreadsheet list from public records of a telephone number of each resident of the city and has the computer randomly sort the list and calls the first 100 residents to ask their preference.

NOT GOOD C. A surveyor leaves a survey at the front of all of the restaurants in the city to ask customers their preference.

WORST (D) A surveyor asks all of the students at the local middle school their preference. MEMBERS AREN'T EVEN A PART OF THE POPULATION

- **Non Response Bias:** **is the bias that results when respondents differ in meaningful ways from nonrespondents. Nonresponse is often problem with mail surveys, where the response rate can be very low.**

  Explain why looking on the internet at reviews of a product may suffer from a Non-Response Bias.

  MOST SELF-SELECTED SURVEYS HAVE A "NON-RESPONSE" FROM THE MIDDLE MAJORITY OF THE POPULATION. USUALLY JUST THE EXTREMES GOOD AND BAD RESPOND.

- **Response Bias:** **is the tendency of a person to answer questions on a survey untruthfully or misleadingly.**

  Are there any concerns of Response Bias in the following survey questions?

  - What is wrong with your current school? IMPLIES SOMETHING IS WRONG WITH YOUR SCHOOL.
  - To improve education, should taxes be raised to fund building more schools? IMPLIES BUILDING MORE SCHOOLS WILL IMPROVE EDUCATION
  - Why are teenage drivers dangerous? IMPLIES TEENAGE DRIVERS ARE DANGEROUS
  - How long does it take you to get to school? ← GOOD QUESTION. NO LEADING OPINIONS.

Primary Data Source: When the researcher collects and controls data collection themselves.

Secondary Data Source: When the researcher uses data that was collected by another person or group other than themselves.

22. Additional Types of Data: Primary vs. Secondary

a. A researcher for a local school system was researching the accuracy of how well the Scholastic Aptitude Test (SAT) predicts college success based on a student's college grade point average (GPA) and used the data provided by College Board the company that designed the SAT. Which best describes the data set?

a. Primary Data: The collected data comes straight from the company that developed the SAT.
b. Secondary Data: The collected data used was not collected directly by the researcher or research team.
c. Secondary Data: The collected data should be scrutinized since it could influence the company's status and profits and therefore have been manipulated.

b. An administrator at a school wishes to compare attendance rates to student's GPAs to see if there is a connection between grades and attendance. The administrator looks up each student grade and attendance on the school's database. Which best describes the data?

a. Primary Data: The collected data is determined by the person conducting the research.
b. Secondary Data: The collected data is ultimately determined by the student's actions.
c. Secondary Data: The collected data should be scrutinized since it would represent the school.

c. A researcher for an insurance company was conducting a study to determine on average how often a person in the county goes to the hospital. The county had recently conducted a government census survey which had specifically asked the question in survey, "How many times have you visited a medical facility in the last year?". Rather than conduct his own research, the researcher decided to use the data from the county census. Which best describes the data set?

a. Primary Data: The collected data should be accurate.
b. Secondary Data: The collected data is accurate since it was conducted by the government.
c. Secondary Data: The collected data used was not collected directly by the researcher or research team.

Ethics & Privacy in data collection:

a. A researcher is feeding mice food that has been sprayed with a herbicide to see if it is poisonous to humans. Do you think this could be a violation of ethics or ~~privacy~~? Why or why not?

Privacy has probably not been violated but ethics may have been depending on your view point.

b. In a confidential study, a researcher shares detailed information such as identity, personal symptoms, habits, and other data about patients with Parkinson's Disease involved in a study with a pharmaceutical company because the researcher truly believes sharing this could improve the lives of thousands suffering from Parkinson's. Do you think this could be a violation of ethics or privacy? Why or why not? What is informed consent?

Privacy was definitely violated. Identities of participants should never be shared and an explicit informed consent policy should be used.

24. Experimental Study Variable Types Consider the study:

A doctor is conducting a study on a new medicine to treat patients with diabetes to lower their A1C level. She has 40 patients that have agreed to be part of the clinical trial. So, a doctor asked an independent company to send her 20 containers of the new medicine in tablets and 20 containers of a placebo medicine in tablets that look identical to the new medicine but have essentially no effect. The medicine is put into a container labeled "A" and the placebo into a container labeled "B" but the doctor conducting the study does **not** know which label represents the actual medicine and which is the placebo. She gives it to the patients to be taken daily for 3 months. At the end of the study, it was determined that the group that took "A" consistently had a slightly lower A1C level which showed the new medicine may have been effective. She also learned that a large number of patients in the group that took the medicine also significantly changed their diet to vegan. So, the doctor isn't certain if the improvement was due to the medicine or the change in diet.

A. Which type of variable best describes the **A1C level**?

The variable to be observed.

a. Confounding Variable b. Independent Variable c. Dependent Variable

OBSERVING

B. Which type of variable best describes the amount of **new medicine take**?

a. Confounding Variable b. Independent Variable c. Dependent Variable

The variable the researcher controls

C. Which type of variable best describes the **Diet**?

a. Confounding Variable b. Independent Variable c. Dependent Variable

A variable not originally accounted for that could invalidate the study.

D. Which best describes the Group that took the actual new medicine?

a. Treatment Group b. Control Group ~~c. Regular Group~~ ~~d. Placebo Group~~

The group that actually receives the real treatment

E. Which best describes the Group that took the placebo medicine?

a. Treatment Group b. Control Group ~~c. Regular Group~~ ~~d. Placebo Group~~

The group that receives the placebo or no treatement

F. Which type of study was conducted?

a. Blind Study b. Double Blind Study

When the researcher knows which group got the real treatment and which got the placebo but the participant doesn't

When both the researcher/doctor and the participant/patient both don't know who got the real treatment and who got the placebo until the very end of the study.