Jan 6, 2020 Stat 50+150

Chapter 1: Introduction to Statistics

1.1 Statistical and Critical Thinking

The SITUATION	The PROBLEM	The (IMPERFECT) SOLUTION
We want to know *SOMETHING* about a *CERTAIN GROUP*	Most of the time the *GROUP* we're interested in is TOO LARGE.	Don't ask *EVERYONE* about *WHAT YOU'RE INTERESTED IN*, just ask SOME OF THEM and infer.
Ex: Of UCLA undergraduate students, what proportion likes poke bowls? Ex red books in Lib.	Ex: As of 2017, UCLA had about ≈ 31,000 undergraduates. Ex library ≈ 110,000	Ex: Ask 100 UCLA undergraduates if they like poke bowls and use the collected DATA to infer about the total proportion of undergraduates who like poke bowls.

★ *Def* Statistics science of planning studies & experiments; obtaining data, organizing data, summarize data, analyze data & interpret data & draw conclusions!

POPULATION: the entire group to be studied.

• we pick!

SAMPLE: a subcollection (subset) of the population that is being studied.

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: PCC students Sample: the 50 students surveyed Data: 34/50 = 68%

Descriptive Statistics: organizing and summarizing the data.	**Inferential Statistics:** uses methods that take results from a sample, extend it to the population, and measure the reliability of the result. ie draw conclusion.
Ex "68% of PCC students return money"	Ex "I am 95% confident that between 64% and 72% of PCC students would return the money."
(Chapters 1- 6)	(Chapters 7 - 12)

(or ask Q) (or pop)

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

Def Variable the characteristic of the individual to be measured or observed.
Ex In Library Activity: color of book is red

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: teens in US ages 13 to 17

Sample: ↳ 1028 teens contacted

Variable: ↳ whether or not they had been prescribed meds for mental disorder.

1.2 Types of Data

Def **Data** are collections of observations.

Def A **parameter** is a numerical measurement describing some characteristic of a population. (PP)
Number is fixed based on the entire population.

Def A **statistic** is a numerical measurement describing some characteristic of a sample. (SS)
Number varies based on the sample.

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

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

QUALITATIVE (QL)
-Data consists of names or labels. "Quality"
* categorical
* can't do arithmetic w/ it

Ex:
- Country of origin
- student ID
- Eye color
- Yes / No Q
- favorite ice cream flavor

OR

QUANTITATIVE (QN)
- Data consists of numbers representing counts or measurements

DISCRETE ("count")
- has a countable (*finite*) number of values

CONTINUOUS ("measure")
- infinitely many possible values
- length of feet
- amount of rain in Jan

Ex: **Determine whether the following variables are qualitative or quantitative.**

a) Gender Qual (QL)

b) Temperature Quant (QN)

c) Number of days in the past week that you studied QN & discrete

d) Zip code QL

Ex: **Determine whether the quantitative variables are discrete or continuous.**

a) The number (count) of heads obtained after flipping a coin five times.
discrete

b) The number of cars (count) that arrive at McDonald's drive-thru between 12:00 pm and 1:00pm
discrete

c) The distance a 2014 Toyota Prius can travel in city driving conditions with a full tank of gas.
measure — continuous

d) The average test score on the first Statistics exam in a class of 35 students.
Continuous
measure

Statistically Significant is achieved in a study when we get a result that is very unlikely to occur by chance. RULE "less than 5%"	***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?

Data Gender Aide

- Boys: 980/2000 = 49%
- Girls: 1020/2000 = 51%

What would we expect by chance?
↳ expect 50% of girls to be born by chance

NO, IT'S A SCAM!

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.

Ex Eye color (category, no order)

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

Ex College Ranking Ex Grades (A, B, C, D, F)

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.

Ex Temperature, Years

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 Height, Distance

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

a) Nation of origin

Nomial

b) Movie ratings of one star through five stars (category)

Ordinal

c) Volume of water used by a household in a day

Ratio

d) Year of birth of college students

Interval

e) Highest degree conferred (high school, bachelor's, and so on)

ordinal

f) Eye Color

nominal

g) Assessed value of a house

ratio

h) Time of day measured in military time

ratio / interval
(use "test for 0")

Jan 7

1.3 Collecting Sample Data

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

OBSERVATIONAL

- **Retrospective** - Go back in time to collect data over some ***past*** time period.
- **Cross-Sectional** - Data are measured ***at one point in time***
- **Prospective** - Go forward in time and observe groups sharing common factors.

Survey

Ancestry Data

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.

Observational
↳ Retrospective

EXPERIMENTAL

Ideal

- **Completely Randomized** - Randomize who gets treatment or placebo.
- **Randomized Block** - Form different groups of blocks and w/in each block randomize
- **Matched-Pair** - Compare two groups that are related or have similar characteristics

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. Experimentals

Randomized Block

men vs women

also could be Matched Pair

Effect

men 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 RANDOMLY select individuals to stay UNBIASED.

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: Put names in a hat, shake it, & draw // Spin wheel // flip a coin

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: (SRS)

Randomly choose a few book cases in library, & randomly choose a few shelves in each case.

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

- Alaskan Airlines 1
- Akoa 2
- Ashland 3
- Bank of America 4
- BellSouth 5
- Chevron 6
- Citigroup 7
- Delta Airlines 8
- Disney 9
- DuPont 10
- ExxonMobil 11
- General Dynamics 12
- General Electric 13
- Clorox 14

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

Use Google Random # Generator: 7, 10, 14, 8, 11, 5

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

Ex:

→ we pick (not random)

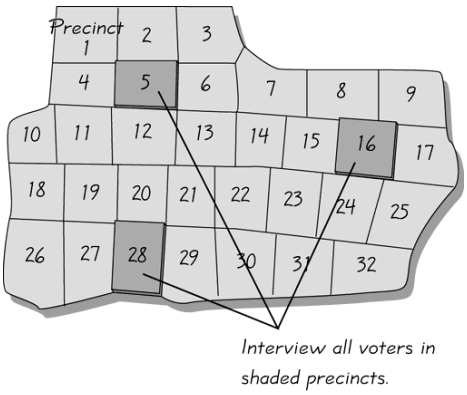
Def **Stratified** Sampling - Subdivide the population into at least two different subgroups that share the same characteristics, then draw a random sample from each subgroup, proportional to the population (or stratum). (cf. not SRS)

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:

Separate State into Counties
& Survey everyone from 3 counties



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

Ex:

- stand & wait for people to walk by you.
- go to a popular grocery store at busy time & ask.