Math 247: Surveys (Section 7.1)

(Target) Population: The entire group that we want to know something about.

Sample: A subset of a population group, containing the items/individuals who are actually observed and whose data (the sample data) is included in our analysis.

A parameter: A number (mean, median, standard deviation, proportion, etc.) that describes a population. Parameters are **unknown** unless we have **census** data.

A statistic: A number (mean, median, standard deviation, proportion, etc.) that describes a sample. This is also called an "estimator". Sample statistics are **known** because we have **sample** data.

Parameters:

Statistics:
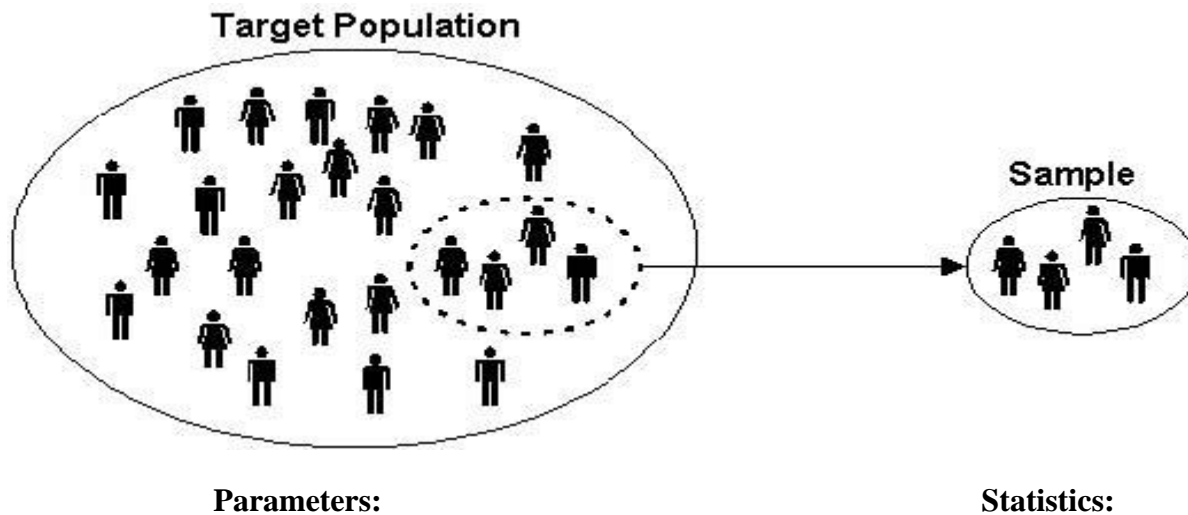
Representative sample: This is a sample that <u>reflects the characteristics</u> of the target population, including considerations such as sex, health, age, political views, religious views, etc.

Sampling methods…what could go wrong?

Bias: A sampling method is <u>biased</u> if it *systematically produces a sample statistic that is not the same as the true population value.*

    <u>Positive</u> bias creates an <u>overestimation</u> of the parameter in question.

    <u>Negative</u> bias creates an <u>underestimation</u> of the parameter.

Types of bias: **Sampling Bias**   vs.   **Response Bias**

Bad sample = SAMPLING BIAS
This type of bias comes from taking a sample that is <u>not representative</u> of the population.

Common survey methods that lead to this type of bias are …

- **Phone surveys**. Using a landline for a survey is going to leave out younger people. Likewise using only cell-phones will leave out some older people.

- Sampling from **non-representative group**.

- **Self-Selection Bias**. These are polls where a <u>sample isn't selected by the researcher</u> but instead a question is posed by, say, a radio talk-show host or an internet poll, and anyone can respond.

Example: Suppose you want to find out the proportion of **<u>all</u>** <u>adults in San Luis Obispo</u> who support a ban on having a cannabis shop in town.

Parameter:

Determine what type of bias the following scenarios would likely have (type of bias and whether it would be positive or negative):

(a) You take a poll outside a church on Sunday morning.

(b) You call 25 people out of a paper phone book (do those even exist anymore???).

(c) You set up a poll on the "New Times" website that anyone can respond to.

So, if there are so many ways to get a BAD sample, how do we get a GOOD sample?

We take a **random sample** so our sampling method won't create bias.

Simple Random Sample (SRS): This a sample taken where **every individual in the target population is equally likely to be chosen**. (*Memorize this definition!*)

Suppose you want to find out the proportion of Cuesta students who work more than 20 hours per week.

Determine which of the following scenarios would give a SRS of 25 Cuesta students and explain why it is or is not a random sample.

    (a)   You gather a sample by walking around campus and randomly asking 25 people.

    (b)   You go to the Cuesta library at noon and ask the first 25 people who walk in the door.

    (c)   You number the students at Cuesta from 1 to 11,019 (the current enrollment), then randomly select 25 numbers and contact those 25 students.

Okay, suppose you finally got a perfect simple random sample. Does this mean you'll now be guaranteed to get unbiased results? Sadly, no.

Good sample with bad results = RESPONSE BIAS

Here are some of the many types of Response Bias:

- **Leading Question Bias:** This type of bias comes from asking questions that are meant to <u>influence</u> the person to answer in a particular way.

- **Social Acceptability Bias:** People may answer questions untruthfully because they're uncomfortable or embarrassed about the issue.

- **Non-Response Bias:** This type of bias comes from the fact that some people in the sample won't respond to a survey. There may be strong characteristics of people who do choose to respond which could bias the result.

Okay, let's suppose you have an SRS of 25 Cuesta students. Suppose now you want to find out the proportion of Cuesta students who have used a drug (not for a medical purpose) in the last two months.

Parameter:

(a) Make up a leading question that would probably result in a …

positive bias

negative bias

(b) What would a neutral question be?

(c) Would all students answer the neutral question honestly? Describe how this would bias the results (mention positive or negative bias).