

## Ch. 19: Sample Surveys

Frequently, we like to make some statements about a class of individuals, the \_\_\_\_\_.

Ex: In predicting the results of a U.S. presidential election, the population are all the voters.

Ex: When asking about student interest in basketball at USU, the population are all USU students (and not only those attending a particular game).

It is usually impractical to measure or interview the whole population. Instead, we take a \_\_\_\_\_ from the population.

We use the sample to make \_\_\_\_\_ about the population, i.e., we generalize from the sample to the population as a whole.

We are interested in numerical values describing the population, called \_\_\_\_\_.

We estimate these parameters by \_\_\_\_\_, i.e., numbers calculated from the sample.

The sample should be \_\_\_\_\_ for the population.

Unfortunately, to check this requires knowledge of the population – exactly what we don't have!

So we must choose the sample very carefully.

## The *Literary Digest* Poll of 1936

In the 1936 presidential election, Franklin Delano Roosevelt (a Democrat), was running for his second term against Alf Landon, the Republican governor of Kansas.

*Literary Digest* magazine did a poll and received 2.4 million responses. They predicted Landon would win 57% to 43%. On election day, however, FDR won 62% to 38%. This was the largest error ever made by a major poll.

How did the *Digest* go so wrong? Their sample was more than big enough. Unfortunately, their selection method showed a \_\_\_\_\_, where some groups were more likely to be counted than others.

The *Digest* mailed questionnaires to 10 million people from phone books, club membership lists, and magazine subscription lists.

What was wrong here?

The poll also suffered from a \_\_\_\_\_ (only 24% responded). Why was this a problem?

## The Presidential Election of 1948

In 1948, three major polls all predicted Thomas Dewey would beat Harry Truman by about 5% in the presidential election. Instead, Truman won, (50% to 45%).

The polls used a \_\_\_\_\_: each interviewer must interview a certain number of subjects in each of several categories (broken down by gender, race, income, etc.), selected to match the country as a whole. But within those constraints, the interviewers can freely choose whom they interview.

This choice resulted in more Republicans being interviewed in each group.

## Probability Methods

When using \_\_\_\_\_, interviewers cannot choose whom they interview. The sample is selected randomly.

The most basic example is called \_\_\_\_\_ (SRS). All sets of the proper size are equally likely to be selected.

A SRS is like putting all individuals' names in a box and drawing without replacement until we reach the desired sample size.

Ex: Ask every fifth student that enters the TSC whether they support women's basketball at USU.  
SRS – yes/no?

Ex: Ask all students in randomly selected classes whether they support women's basketball at USU.  
SRS – yes/no?

Ex: Obtain all student SSNs from the Registrar's Office and let the computer randomly select 500 of these numbers. Ask students with these SSNs whether they support women's basketball at USU.  
SRS – yes/no?

Even with a well-designed sampling scheme, there are many problems which make taking surveys difficult.

\_\_\_\_\_ is a problem even with face-to-face interviews. This happens more frequently with telephone surveys and mailed questionnaires.

We must always worry about the biases that the wording of the questions may bring.

Ex: Three different ways to ask about abortion:

- "The U.S. Supreme Court has ruled that a woman may go to a doctor to end pregnancy at any time during the first three months of pregnancy. Do you favor or oppose this ruling?" (47% favor, 44% oppose).
- "The U.S. Supreme Court has ruled that a woman may go to a doctor for an abortion at any time during the first three months of pregnancy. Do you favor or oppose this ruling?" (41% favor, 48% oppose).
- "As far as you yourself are concerned, would you say that you are for or against abortion, or what do you think?" (36% favor, 59% oppose).

Ex: College students watched a film of a car crash, and were asked:

- "About how fast were the cars going when they contacted each other?"  
(Average answer: 31.8 mph)
- "About how fast were the cars going when they collided with each other?"  
(Average answer: 40.8 mph)

We must always be aware of the possibility of respondents lying, or simply not knowing the correct answer.

Ex: In 1991, the National Survey of Men was conducted. 3,321 respondents were interviewed. These men (aged 20–39) were questioned in their homes by female interviewers. The questions concerned the men's sexual practices, and 30% refused to be interviewed. How trustworthy are such results as that the median number of sexual partners is 7.3, or that only about 1% of men are exclusively homosexual?

Ex: A political pollster once asked people what they thought of the Pepper-Johnson bill being debated by Congress. Most people expressed strong opinions for or against the bill, even though it didn't exist! Apparently, giving an uninformed opinion is easier than admitting ignorance.

Ex: \_\_\_\_\_:

Who do you think answers the "Consumer Product Survey of America?"

Are those people who answer representative for the entire US population?