Unit 17: Samples and Surveys

SUMMARY OF VIDEO

Listen to a news broadcast, read a newspaper, open a magazine and you'll find an item related to the results of a poll on some topic. But how do pollsters survey a population as large and diverse as the United States and wind up with a complete and unbiased picture of attitudes on a particular topic? Before selecting the sample to be surveyed, pollsters need to identify the population – the group of interest to the study.

For example, that could be all registered voters, people in a particular age group, or households with a certain income. The characteristic of a population that we are interested in is called a parameter. We can't determine the true value of a parameter unless we can examine the entire population, which isn't usually possible. However, we can estimate the unknown parameter based on information collected from a sample, a subset of the population. Such an estimate is called a statistic. (Remember **P**arameters are for **P**opulations and **S**tatistics are for **S**amples.)

The pollsters at the University of New Hampshire's Survey Center conduct everything from academic research to political polls. They are experts at selecting samples to represent the attitudes and opinions of the whole population. For a public opinion survey they use random digit dialing to select households; they start with a random sample of households. However, another sampling stage is required because they talk to an individual at each house and not to everyone in the house. So, for example, they might ask to speak to the person in the house who has had the most recent birthday.

Convenience sampling, where pollsters survey a convenient group such as their friends, or voluntary sampling, where data are collected from those who volunteer for the survey, often create an unrepresentative sample and produce biased results. The same issue arises when a sample draws from a list that excludes a portion of the population. This happened in the 1936 presidential election when a *Literary Digest* poll predicted Alf Landon would be the winner. However, Franklin Roosevelt went on to win in a landslide. The *Literary Digest* drew their samples from lists of car and telephone owners – items that, at the time, were indicative of wealth. The poll had omitted the largely pro-Roosevelt poor from their survey, causing bias in favor of Landon.

Getting a representative sample is the cornerstone of accurate sampling. But just as important as a representative sample is the design of the questionnaire. Some practical advice on questionnaires includes using simple words, not asking people about things they are not likely to know about, and keeping the questions short. Even small things such as changing the order in which you read the choice of responses to survey participants can change how they answer a question.

In a simple random sample, each individual of the population has an equal chance of being selected. This can be hard to achieve in a real-life survey since it can be nearly impossible to get a complete list that includes every single member of a large population. Another way of ensuring a representative sample is by doing a multistage sample. For example, the Survey Center might begin with a random sample of counties in New Hampshire. Then they would take a random sample of towns within those counties. Finally they would select random households within those towns.

The problem with multistage sampling is that it could leave out groups of interest merely by chance. To solve this problem, the Survey Center might decide to use a stratified random sample. For this type of sampling design, the entire population is divided into groups with similar characteristics called “strata.” For example, census tracts might first be classified as urban, rural, or suburban, and then a separate random sample is selected from each stratum. In New Hampshire this ensures that cities are represented in the sample even though most counties are rural.