

# ST 516: Foundations of Data Analytics

## Sampling

## Selection of Units for a Study

- Examples

- Types of Sampling

- Random Samples

- Benefits of Random Sampling

- Examples of Samples

# Selection of Units for a Study

Recall that it's important to think about how the data you are analyzing came to be collected.

1. How did the study units get selected for study?
2. Is the study an observational study or a randomized experiment?

In this Lecture, you'll learn a little bit about different types of sampling, or the methods by which units are selected for a study.

## Some Examples

1. In the creativity case study from *The Statistical Sleuth*, subjects volunteered for the study.
2. Similarly, in the starting salaries case study, the employees essentially “volunteered” for their jobs.
3. In a study of numbers of visits to a particular website, a company may sample visits during 15 randomly selected 10 minute windows throughout a week.
4. In conducting an exit poll, an agency may ask opinions of the first 50 people who are willing to participate after casting their vote.

# Types of Sampling

There are many different methods by which units may be selected for study—to a statistician, these are called sampling methods, and they result in different types of samples:

- **Volunteer/Convenience Sample.** This is a non-random sample from some population of interest.
- **Random Sample.** These are samples that are obtained using some chance (probability!) mechanism:

# Types of Random Samples

There are many different types of random sampling. Here are a few of the most common ones:

- Simple random sample. Every unit in the population has the same chance of being selected for the sample.
- Stratified sample. The population is divided into non-overlapping sub-populations; within each sub-population—or stratum—units are chosen using simple random sampling.
- Cluster sample. The population is divided into non-overlapping groups (usually based upon geography), called clusters; clusters are chosen using simple random sampling, and each unit within the chosen cluster is measured.

## Benefits of Random Sampling

Obtaining a random sample from a population ensures us that the sample will be **representative** of the population in the sense that the characteristics of the units in the sample will be roughly the same as the characteristics of the units in the population.

On the next several slides, you'll see descriptions of different studies accompanied by a question. Try to answer the question on your own before clicking to the next slide, which will reveal the answer.

If you still don't entirely understand an answer, feel free to post a question to either of the weekly discussion boards.

## Sampling Example 1

A psychology professor asks her students to participate in an experiment. She will provide extra credit for participation. Twenty-three students sign up to participate.

Question: Is this an example of a random sample? Explain.



## Sampling Example 1

A psychology professor asks her students to participate in an experiment. She will provide extra credit for participation. Twenty-three students sign up to participate.

Question: Is this an example of a random sample? Explain.

- No; the students self-select into the study, and only those students that want or need extra credit may volunteer.

## Sampling Example 2

Suppose that a polling organization has telephone numbers of all registered voters in a state and a chance mechanism for selecting telephone numbers so that each number has equal chance of selection in their sample of size 300.

Question: Is the resulting sample of registered voters a simple random sample?

## Sampling Example 2

Suppose that a polling organization has telephone numbers of all registered voters in a state and a chance mechanism for selecting telephone numbers so that each number has equal chance of selection in their sample of size 300.

Question: Is the resulting sample of registered voters a simple random sample?

- Technically, it is. But, you should also consider that some registered voters may not answer their phone or they may refuse to participate in the survey. These things must be considered in any analysis of data resulting from this type of survey.

## Sampling Example 3

A wind power company must attempt to minimize the number of birds and bats that are killed by flying into active turbines. To monitor bird/bat mortalities, the company installs radar and infrared sensors on 10 of the 521 turbines it operates. The 10 turbines are chosen so that they are spread out evenly across the 521 turbines.

Do the 10 turbines represent a random sample of turbines?

## Sampling Example 3

A wind power company must attempt to minimize the number of birds and bats that are killed by flying into active turbines. To monitor bird/bat mortalities, the company installs radar and infrared sensors on 10 of the 521 turbines it operates. The 10 turbines are chosen so that they are spread out evenly across the 521 turbines.

Do the 10 turbines represent a random sample of turbines?

- No, the 10 turbines were not selected randomly. The company will have to make scientific arguments to assert that the 10 turbines are *representative* of all 521 turbines.