

## Section 1.2

**Example:** Steve Kerr, the coach of the Golden State Warriors (a basketball team headquartered in San Francisco), wants to know the average height of people in San Francisco. In order to figure this out, he uses the heights of the members of his basketball team to make an estimate. He then concludes, based on the height of his players, that the average height of San Franciscans is 6 feet 6 inches. What went wrong?

- Basketball players are very tall.
- Warriors are not representative of population.
- Since Warriors are not representative of the population, the estimate was way off.

Even if Steve Kerr had used a more sensible sampling method (i.e. not picking basketball player-s), there is no guarantee that the estimate from the sample would have been accurate. Why do sampling at all? After all, isn't it more accurate to examine the **entire** population?

### Reasons for Sampling

- It saves time: if we wanted to know how every Democrat in Pennsylvania felt about Obama or Clinton during the 2008 primary season, asking each person would take quite a lot of time.
- It saves money: if a jeweler wanted to know how people in America felt about a particular diamond ring, it would cost a lot of money to produce diamond rings for every single American.
- It is easier to manage the data: it is easier to keep track of the opinions of 300 Americans than it is to keep track of the opinions of 300,000,000 Americans.

The most common form of sampling is simple random sampling.

**Definition:** A **simple random sample** of  $n$  measurements from a population is a subset of the population selected in such a manner that every sample of size  $n$  from the population has an equal chance of being selected.

**Exercise 1.** Let's say we have a class of 16 students and we wanted to take a simple random sample of 5 students from this class. How would we do this using the Random-Number Table?

### Procedure: How to draw a random sample

1. Number all members of the population sequentially.
2. Use a table, calculator, or computer to select random numbers from the numbers assigned to the population members.
3. Create the sample by using population members with numbers corresponding to those randomly selected.

**Definition:** A **simulation** is a numerical facsimile or representation of a real-world phenomenon.

**Exercise 2.** Use a random-number table to simulate the outcomes of tossing a quarter 25 times. Assume that the quarter is balanced (i.e., fair). (#14)

**Class Exercise 1.** Suppose there are 30 people at a party. Do you think any two share the same birthday? Let's use the random-number table to simulate the birthdays of the 30 people at the party. Ignoring leap year, let's assume that the year has 365 days. Number the days, with 1 representing January 1, 2 representing January 2, and so forth, with 365 representing December 31. Draw a random sample of 30 days (with replacement). These days represent the birthdays of the people at the party. Were any two of the birthdays the same? Compare your results with those obtained by other students in the class. Would you expect the results to be same or different? (#16)

Simple random sampling is not the only way to conduct a sample; in many situations, it is impractical to take a simple random sample. Here are some other sampling methods:

**Definition:** In systematic sampling, we number all members of the population sequentially. Then, from a starting point selected at random, include every  $k$ th member of the population in the sample.

**Example:** Suppose a supermarket wants to study buying habits of their customers, then using systematic sampling they can choose every 10th or 15th customer entering the supermarket and conduct the study on this sample.

**Definition:** With convenience sampling, we create a sample by using data from population members that are readily available.

**Example:** Suppose that we sell your company several crates of oranges each week. You examine a sample of oranges from each crate to determine the quality of our oranges. It is easy to inspect a few oranges from the top of each crate, but these oranges may not be representative of the entire crate. Those on the bottom are more often damaged in shipment. If we were less than honest, we might make sure that the rotten oranges are packed on the bottom with some good ones on top for you to inspect.

**Exercise 3.** A recent survey of 437 potential home buyers in Orange County, California, found, among other things that

*All but 2 percent of the buyers have at least one computer at home, and 62 percent have two or more. Of those with a computer, 99 percent are connected to the Internet (Jennifer Hieger, "Portrait of Home-buyer Household: 2 Kids and a PC," Orange County Register, 27 July 2011).*

The survey was conducted via the Internet. Comment on this.

**Definition:** With stratified sampling, we subdivide the population into distinct subgroups called strata. The strata are based on a specific characteristic such as age, income, education level, and so on. All members of a stratum share the specific characteristic. Draw random samples from each stratum.

**Exercise 4.** Suppose a farmer wishes to work out the average milk yield of each cow type in his herd which consists of Ayrshire, Friesian, Galloway, and Jersey cows. He divides up his herd into four sub-groups and take samples from these. This is an example of stratified sampling. What are the strata in this case?

**Class Exercise 2.** An article in the journal *Science* looks at attitudes toward genetically modified foods in Europe. A sample is conducted by taking samples from 17 European countries and then combining them. Is the sample a stratified one? If so, what are the strata.

There are three main reasons for stratifying a population.

- *Convenience* in selecting the sample is enhanced. It is easier to sample in smaller, more compact groups (countries) than in one large group spread out over a huge area (the world).
- *Coverage* of each stratum is assured. The company may want data from each country in which it sells products; a simple random sample from the frame does not guarantee that this would happen.
- *Precision* of the results may be improved. That is, stratification tends to give estimates that are closer to the value for the entire population than does an SRS. This is the fundamental reason for stratification.

**Exercise 5.** To see how well U.S. third graders do an arithmetic test, you might choose a simple random sample of children enrolled in the third grade and give each child a standardized test. In theory, this is a reasonable plan, but it is not very practical. For one thing, how would you go about making a complete list of all the third graders in the United States. For another, imagine the work required to track down each child in your sample and get him or her to take the test. What would be a better approach?

The above example illustrates cluster sampling, in which each elementary school is a cluster of third-grade students.

**Definition:** In **cluster sampling**, we first divide the population area into sections (or clusters), then randomly select some of those clusters, and then choose *all* the members from those selected clusters.

**Example:** Suppose you want to survey opinions of AP Statistics students to see if they feel they have enough time on the free-response section of the AP exam. One option would be to begin with a list of all schools that had students who took the AP Statistics exam. A number of schools are randomly selected from the list of all schools represented, and every student at selected schools who took the AP exam is asked if he or she thought the time allowed to complete the free-response portion of the exam was adequate. The students interviewed would constitute a cluster sample.

What's the difference between cluster sampling and stratified sampling? We stratify to ensure that our sample represents different groups in the population, and sample randomly within each

stratum. Strata are homogenous, but differ from one another. By contrast, clusters are more or less alike, each heterogeneous and resembling the overall population. We select clusters to make sampling more practical or affordable.

**Class Exercise 3.** For each of the following exercises, identify the sampling technique used:

- (a) Chosen at random, 500 rural and 500 urban persons age 65 or older were asked about their health and experience with prescription drugs.
- (b) Questioning teachers as they left a faculty lounge, a researcher asked 32 teachers about their teaching styles and grading methods.
- (c) A list of managers is compiled and ordered. After a starting number is randomly chosen, every ninth name is selected until 1000 managers are selected. The managers are questioned about the use of digital media.
- (d) A company packaging snack foods maintains quality control by randomly selecting 10 cases from each day's production and weighing the bags. Then they open one bag from each case and inspect the contents.

**Class Exercise 4.** Modern Managed Hospitals (MMH) is a national for-profit chain of hospitals. Management wants to survey patients discharged this past year to obtain patient satisfaction profiles. They wish to use a sample of such patients. Several sampling techniques are described below. Categorize each technique as simple random sample, *stratified sample*, *systematic sample*, *cluster sample*, or *convenience sample*. (#20)

- (a) Obtain a list of patients discharged from all MMH facilities. Divide the patients according to length of hospital stay (2 days or less, 3-7 days, 8-14 days, more than 14 days). Draw simple random samples from each group.
- (b) Obtain lists of patients discharged from all MMH facilities. Number these patients, and then use a random-number table to obtain the sample.
- (c) Randomly select some MMH facilities from each of five geographic regions, and then include all the patients on the discharge lists of the selected hospitals.
- (d) At the beginning of the year, instruct each MMH facility to survey every 500th patient discharged.
- (e) Instruct each MMH facility to survey 10 discharged patients this week and send in the results.

Sometimes, we use a variety of sampling methods together.

**Definition:** With **Multistage sampling**, we use a variety of sampling methods to create successively smaller groups at each stage. The final sample consists of clusters.

Most surveys conducted by professional polling organizations use some combination of stratified and cluster sampling as well as simple random samples.

## Student Feedback

My teaching methods are (I hope) continually subject to improvement. If you have any comments, suggestions, or ideas, please email them to me at [Sithparan.Vanniasegaram@evc.edu](mailto:Sithparan.Vanniasegaram@evc.edu) .

### Homework

#### C problems

Section 1.2: 11 - 19 ODD

#### B problems

Section 1.2: 1, 3, 5

#### A problems

Section 1.2: 7, 9