

Lesson 24: Collecting Data

TERMS

In order to better understand the characteristics of a population, statisticians and researchers often use a sample from that population and **make inferences** based on the summery results from the sample.

Population: entire group of individuals about which we want information.

Sample: the part of the population from which we actually collect information.

Sampling Hardwood and Humans

Populations and samples

PROBLEM: Identify the population and the sample in each of the following settings.

(a) A furniture maker buys hardwood in large batches. The supplier is supposed to dry the wood before shipping (wood that isn't dry won't hold its size and shape). The furniture maker chooses five pieces of wood from each batch and tests their moisture content. If any piece exceeds 12% moisture content, the entire batch is sent back.

(b) Each week, the Gallup Poll questions a sample of about 1500 adult U.S. residents to determine national opinion on a wide variety of issues.

Polling is an example of sampling from the population in order to get a better idea of the characteristics of a population. Because we make inferences about a population from the sample, it is very important that the sample is collected appropriately and that it is **representative** of the population being studied.

Type of sampling	Example	Advantage	Disadvantage
Convenience Sampling	In order to get an idea of how students think of the new school policy, the principal stands outside the library and asks a few students their opinions.	Easy and Cheap	Not representative of the population
Voluntary Response Sample	After the State of the Union speech, ABC tells its audience to call a 1-900-555-1234 if they thought the speech was good and 1-900-555-7890 if they thought the speech was bad (there is a \$0.50 charge for the call).	Easy	Biased toward people with strong opinions.
Systematic Random Sampling	HP Selects every 200 th computer off the assembly line and inspects it for quality control.	Every member has an equal probability of being selected.	Not every sample of size n is possible.

Illegal Immigrants and Driver's Licenses

Online polls

Former CNN commentator Lou Dobbs doesn't like illegal immigration. One of his shows was largely devoted to attacking a proposal to offer driver's licenses to illegal immigrants. During the show, Mr. Dobbs invited his viewers to go to loudobbs.com to vote on the question "Would you be more or less likely to vote for a presidential candidate who supports giving drivers' licenses to illegal aliens? The result: 97% of the 7350 people who voted by the end of the show said "Less likely."

PROBLEM: What type of sample did Mr. Dobbs use in his poll? Explain how this sampling method could lead to bias in the poll results.

CHECK YOUR UNDERSTANDING

For each of the following situations, identify the sampling method used. Then explain how the sampling method could lead to bias.

1. A farmer brings a juice company several crates of oranges each week. A company inspector looks at 10 oranges from the top of each crate before deciding whether to buy all the oranges.
 2. The ABC program *Nightline* once asked whether the United Nations should continue to have its headquarters in the United States. Viewers were invited to call one telephone number to respond "Yes" and another for "No." There was a charge for calling either number. More than 186,000 callers responded, and 67% said "No."
-

Randomization Methods

Method 1: Roll a Die or 2 Dice

Method 2: Put all the items in a hat and draw them out randomly

Method 3: Use a random number table (Table D in AP Guide)

Method 4: Use technology to randomize items

AP Stats Calculator Program > "3 random numbers"

Option 1: Random Number given a number range to stay within. Results in 1 number.

Option 2: Random List given a number range to stay within AND the total number of random digits desired. Results in as many numbers as you desire.

Option 3: Shuffle a set of numbers within a range.

Answering a free response question on how to conduct a poll:

1. Describe marking the population
2. Describe what randomization tool you will use (table, die, hat, calculator, etc.)
3. Describe the method of selection (interpret how the random numbers become subjects)
4. State how repeats will be handled.
5. Answer the question and ensure that the selection is a Simple Random Sample.

★★★ The Simple Random Sample (SRS) ★★★

This consists of n individuals from the population chosen in such a way that every set of n individuals has an equal chance of being the sample actually selected. This is often the best and most appropriate way to collect data for a sample.

Example: In order to determine how happy students are with their education at NWHHS, the principal assigns each student a number from 1 to 850 (the number of students at the school) and then uses a random number generator to choose 50 numbers between 1 and 850. He then surveys all the students with the chosen numbers.

The easiest way to design a SRS of size 30:

- Place the names of all members of the population on slips of paper.
- Place the slips in a bag and shake it up.
- Select 30 slips at random without replacement.

Note: Try to keep the sample size over 30. The reason for this will be described later in this course.

Example Question:

The school newspaper is planning an article on family-friendly places to stay over spring break at a nearby beach town. The editors intend to call 5 randomly chosen hotels to ask about their amenities for families with children. They have an alphabetized list of all 28 hotels in the town.

Population:

Sample:

Describe how to create an SRS of the hotels to be called:

Daily Data Collection

In groups: Create one question on a topic of interest and describe how to survey a sample using each sampling technique. Volunteers will be called upon to share methods to the class.

The class will decide on the most interesting question and use it to perform a survey in class.