Random and Biased Sampling in a Population - 7.SP.1,2

suzanne fox

AUTHOR
suzanne fox

To access a customizable version of this book, as well as other interactive content, visit www.ck12.org

CK-12 Foundation is a non-profit organization with a mission to reduce the cost of textbook materials for the K-12 market both in the U.S. and worldwide. Using an open-content, web-based collaborative model termed the **FlexBook®**, CK-12 intends to pioneer the generation and distribution of high-quality educational content that will serve both as core text as well as provide an adaptive environment for learning, powered through the **FlexBook Platform®**.

Printed: March 1, 2013

flexbook
next generation textbooks

CONCEPT 1 Random and Biased Sampling in a Population - 7.SP.1,2

Students will learn about various sampling techniques and how the type of sample can change the results of the survey. They will see how sampling the population can give clues to the characteristics of interest of the entire population.

Have you ever played or watched Family Feud? The whole game is based on math! It is based on a specific part of math called statistics. Statistics is the study of data. Statistics has three functions. It collects data. It organizes the data into tables, charts, and graphs. It then analyzes it and makes conclusions about the data. This is all so people can understand what the data really means. Family Feud uses surveys in its game. Surveys are a type of collecting, organizing and making conclusions about data used in statistics.

Collecting Data From a Population

Data is information. It can be numbers or written responses. It can be collected using a written list of questions (questionaire), an interview, a log or diary, a chart, or a tally.

- Doctors may want to know if a new medicine is working to cure an illness.
- Researchers may want to know if a new golf ball goes farther.
- An advertising company wants to know if a new show on TV is popular.

Often data is collected from a population. A population is a group that is being studied. When every person must be included to get the data, it is called a census. However, sometimes the population that people want to study is much too large. For example, if we wanted to know how many people in the United States know the Pledge of Allegiance, that would be a population that is very large. When this happens, we use a sample of the population. A sample uses only part of the population and then the results are used to talk about the whole population.

A sample must be collected carefully and correctly. There are three things that people must be aware of when deciding on a sample:

1. **The sample must be fair.**
2. **The sample must have enough people or items in it.**
3. **The sample must be random.**

First let's understand what it is for a sample to be fair. Think about the following situation: a sample of students is to be selected and the height of each student is to be measured to find the average height of a student in high school. Are the following samples fair or unfair? Why or why not?

a. The basketball team
b. The first five students that enter each homeroom
c. All freshman girls
d. Every tenth boy in a list of all boys in the school

a. The basketball team is an example of an unfair sample. To be a basketball player , you have to be tall. That might make the average height taller than it really is.

b. The first five students that enter each homeroom would be an example of a fair sample. The people would be totally random and taken from the entire population.

c. All freshman girls is an example of an unfair sample. Girls are usually shorter than boys and also they are the youngest group, so many have not grown that much.

d. Every tenth boy in a list of all boys in the school. This may be a fair sample if we were only concerned with boys. The sample does not have any girls, which could be half the population of the school.

How many is "enough" people in a sample? If a doctor asked you to try a new medicine for your illness and you asked him how many other people tried it, what if he said 2? Would you still be willing to try the medicine? Maybe not. For a medicine to prove that it works, it has to be tried on more than 2 people.

Generally, the larger the sample size, the better the data will reflect the whole population. If the whole population is 100 people, which of the following would be a good number in the sample?

10 people 90 people 30 people 50 people

50 people would be half the population - this would be a good sample size. 90 people would be even better. Sample sizes of 10 and 30 may not give accurate data of the whole population. The bigger the sample size, the more accurate the data will be.

Random means by chance, with no purpose. In statistics, random sampling means to get a sample of the population equally from all parts of the population. A good random sample makes results more believeable. Why? Because the sample is representative of the whole population. Let's make this simple to understand with a model.

A survey is made to see which television programs are the most popular in a large city. Which of the following places would get the most accurate results? In other words, which would give a more random sample.

- a sporting event
- a rock concert
- the mall

If the survey were taken at a sporting event, there is a greater chance that more people would pick sports as the most popular shows. This would not be a random sample.

If the survey was taken at a rock concert, there would be a greater chance that more people would pick music shows. It might also have more people of a certain age, another factor that makes this a bad example of random sampling.

If the survey was taken at the mall, there would be all ages and a diversity of people. This would be a good example of random sampling.

A sample of a population that is not truely random is called biased sampling. If you were told that a survey said that more people liked deep dish pizza than thin crust pizza, and were not told where the survey was taken, you would

not know if it was random or biased. What if you found out that the survey was taken in Chicago, the home of deep dish pizza. What would you think about the results of the survey?

Here's more on biased sampling at http://learnzillion.com/lessons/1847-understanding-biased-samples

Example 1

The school food service wants to increase the number of students who eat hot lunch in the cafeteria. The student council has been asked to conduct a survey of the student body to determine the students' preferences for hot lunch. They have determined three ways to do the survey. The three methods are listed below. Determine if each survey option would produce a random sample. Which survey option should the student council use and why?

1. Write all of the students' names on cards and pull them out in a draw to determine who will complete the survey.
2. Survey the first 20 students that enter the lunchroom.
3. Survey every 3rd student who gets off a bus.

- *The first survey option would produce a random sample. No one would know ahead of time who the people were. Every person has an equal opportunity to be picked to be in the survey.*
- *The second survey option* **may** *produce a random sample. The first 20 students may be the ones who always want a hot lunch. There is no way to be sure.*
- *The third survey option* **may** *produce a random sample. Students who ride a bus may not be cold and may not want a hot lunch compared to students who walk. There is no way to be sure.*

The best survey option for the student council to use is the first one. Names on a card is totally random. A representative sample must have an equal opportunity for all members of the population to be a part of the survey.

Learn more about sampling from the video at http://virtualnerd.com/algebra-2/probability-statistics/unbiased-sample-definition.php

Interpreting Data From Random Samples

Martha wants to take a survey to show which sports students at her school like to play the most.

A sample of the population would include a random sample of the student population in Martha's school. A good strategy would be to randomly select students (using dice or a random number generator) as they walk into an all-school assembly.

This survey was done in a tally format. Below are the results of Martha's survey.

TABLE 1.1:

Sport	Tally	
baseball	𝍸 𝍸 𝍸 𝍸 𝍸 |	31
basketball	𝍸 𝍸 𝍸 ||	17
football	𝍸 𝍸 ||||	14

TABLE 1.1: (continued)

Sport	Tally	
soccer	𝍸 𝍸 𝍸 𝍸 𝍸 |||	28
volleyball	𝍸 ||||	9
swimming	𝍸 |||	8
gymnastics	|||	3
fencing	||	2
		Total: 112

When we look at data from a random sample, we can make conclusions as to what the entire population would most likely prefer. This is called making inferences.

What are some of the inferences we can make about the data in Martha's survey?

- The top two sports that students like to play the most are baseball and soccer.
- The sport students like to play the least is fencing.
- Students like to play volleyball and swim about the same.

Here is another form of recording data. It is in table format. Below is the data collected from two random samples of 100 students regarding student's school lunch preference. What inferences can be made based on the results?

Student Sample	Hamburgers	Tacos	Pizza	Total
#1	12	14	74	100
#2	12	11	77	100

Since these were two separate surveys, it is pretty clear that pizza is the number one choice of students. Even though the numbers are a little different, they are clearly way above the rest. Looking at the results for hamburgers and tacos. There really is not that much of a difference between the two in either survey. So students don't seem to have much of a preference of hamburgers to tacos. Lastly It seems that these two surveys must indeed be random, since their results are so similar. These surveys must be a good representation of the whole population.

Two random surveys were taken at Freemont High School. The purpose of the survey was to find out what students would want as their grades for courses taken. The data is displayed in the table below.

Student Survey	letter grades	number grades	Pass/fail	no grades	Total
#1	72	28	45	55	200
#2	68	65	47	20	200

a. Is there a correlation between the two surveys? If so, explain the correlation. If not, show where there is no correlation.
b. Do the surveys achieve the same results? Explain your answer.

a. Correlation means relationship. Is there a relationship between the two surveys? Looking at each of the catagories, some do have a relationship. In both surveys the number of students that prefer letter grades is the highest. Also, the number of students that prefer courses to be pass/fail is about the same. There seems to be no relationship between students that want number grades or no grades between the two surveys. In fact they are almost opposite in these two catagories.

b. The surveys achieve the same results if you want to know what students want the most in grading. The surveys achieve different results in two of the 4 catagories. There is no agreement between wanting number grades or no grades.

Word Problems and Population Sampling

A survey from random sampling can also be used to predict an entire population. It can be very difficult to get a census to know the actual results. Surveys can give us an estimate of an entire population. Here is a model that fits this purpose. If we wanted to know a lizard population in a certain area, how hard would it be to catch and count **every** lizard? Almost impossible. Here is a way to estimate the lizard population.

To track the lizard population around a lake, researchers marked 36 small lizards.
A few weeks later, they counted 91 lizards, of which 21 were marked.
To the nearest whole number, what is the best estimate for the lizard population?

To get an estimate of the whole population of lizards, we can use a proportion. The colors in the proportion match the part of the word proplem we will use. Since we are looking for the total estimated population of lizards, we use a variable for that part of the proportion.

$$\frac{\text{number of marked lizards counted}}{\text{total number of lizards counted}} = \frac{\text{total number of marked lizards}}{\text{total estimated population}}$$

a sample population — estimate of the whole population

We put the numbers or variable in the proportion and solve. Let the total estimated population = p.

$$\begin{aligned} \frac{21}{91} &= \frac{36}{p} \\ (21)(p) &= (91)(36) \\ 21p &= 3276 \\ \frac{21p}{21} &= \frac{3276}{21} \\ p &= 156 \end{aligned}$$

So even though we cannot count every lizard, a sample can give us a good idea to the total population. We can estimate thate there are about 156 lizards around the lake.

Example 3

Rafael and his team study the deer population and want to estimate the population in northern New York. Last year, his team estimated that the deer population would grow to about 280 deer. To test his estimate, Rafael's team catches 80 deer, tags them, and releases them. Later that month, Rafael and his team catch 64 deer and record that 22 of them are tagged. How close was Rafael's team estimate last year to his estimate this year?

Set up a proportion to estimate the total deer population.

$$\frac{\textit{number of marked deer counted}}{\textit{total number of deer counted}} = \frac{\textit{total number of marked deer}}{\textit{total estimated deer population}}$$

What we do and don't know:

- *number of marked deer counted = 22*
- *total number of deer counuted = 64*
- *total number of marked deer = 80*
- *total estimated population = p*

Put in the information into the proportion and solve for the variable.

$$\begin{aligned}
\frac{22}{64} &= \frac{80}{p} \\
(22)(p) &= (64)(80) \\
22p &= 5120 \\
\frac{22p}{22} &= \frac{5120}{22} \\
p &= 232.7272727\ldots
\end{aligned}$$

Since deer are in unit of 1, we can round the estimated population to 233 deer.

The estimate from last year was 280 deer. Subtract the two to find the difference between the two estimates.

$280 - 233 = 47$. *The estimate from last year was over by 47 deer.*

A great video to watch is http://science360.gov/obj/video/dc8e0737-ceac-4582-bee9-5f195951a01a/science-behind-news-opinion-polls-random-sampling. It talks all about how random sampling is used in many real life situations.

http://www.learner.org/courses/learningmath/video/data/wmp/dat_09_ch4.html shows how random sampling and estimation is helping to save endangered fish in Lake Victoria.

sample, random sample, fair sample, biased sample, statistics, data, population, censussurvey, inference, estimate