

Making Predictions Based on Random Samples



Objective

In this lesson, you will

use data from a random sample to draw conclusions about a population.

Inferences from Random Samples

Random samples should not be **biased** and should be **representative** of the entire population.

The statistics of a random sample are roughly **proportional** to the statistics of the population.

Lesson Activity

Gavin is doing a survey to predict the results of an upcoming school election. During each lunch period, he surveys every tenth student leaving the lunchroom. This table shows the results of his survey.

President	Votes
Henry	36
Christy	42
Celina	22

Based on his survey, Gavin will predict who the 1,500 students in his school will vote for in the race for president.

The data in the table shows that the sample size is **100** ($36 + 42 + 22$).

Set up a proportion to predict the number of votes each candidate for president will receive from the population (the 1,500 students in Gavin's school):

$$\frac{\text{number of votes from sample}}{\text{sample size}} = \frac{\text{predicted number of votes}}{\text{population size}}$$

Henry:

$$\frac{36}{100} = \frac{\text{predicted votes}}{1,500}$$

Christy:

$$\frac{42}{100} = \frac{\text{predicted votes}}{1,500}$$

Celina:

$$\frac{22}{100} = \frac{\text{predicted votes}}{1,500}$$

Based on Gavin's survey,

Henry will get **540** votes, Christy will get **630** votes, and Celina will get **330** votes.



These predictions are estimates.

The prediction might differ if a different random sample is used, but they should be fairly similar.

We can use statistics, such as measures of center, to make predictions about data.

Example:

Suppose that the Cedar Valley student council asked 50 students for their birthdates as well as their mascot choices. The mean age of the 22 students who voted for the gorilla was 16.37 years and the median age was 15.21.



What inferences can be drawn about students in the population who prefer the gorilla?

The median and mean of the population should be different from about the same as the median and mean of the sample:

- ✓ So, half of all the students who prefer the gorilla should be older (and half younger) than 15.21.
- ✓ So, the average age of all the students who prefer the gorilla should be close to 16.37.

Multiple Samples

When using the statistics of a random sample to predict the statistics of the entire population, a certain amount of variation or difference in the results is to be expected. This occurs because only a subset of the entire population is used and not all subsets are the same.



The larger the sample size, the larger smaller, the variation from the expected value.

So, it is a good practice to use a large sample size.

A large sample size tends to give more-accurate results for predicting the statistics of the population.

Example:

A toymaker produces 3,000 remote-controlled trucks each week. The company wants to predict the number of defective trucks in that week's production. Checking each truck could be time-consuming and costly, so the company decides to check a sample.

A random selection of 50 300 trucks will give an estimate that is closer to the expected result of the population than a random sample of 50 300 trucks.

Summary

Dana plans to use a survey of 60 randomly selected students at her school to predict the percentage of the 900 students at her school who will attend next week's school dance. How can Dana change her data collection plans to make her prediction closer to the actual result?

answers will vary