

# Family Support Materials

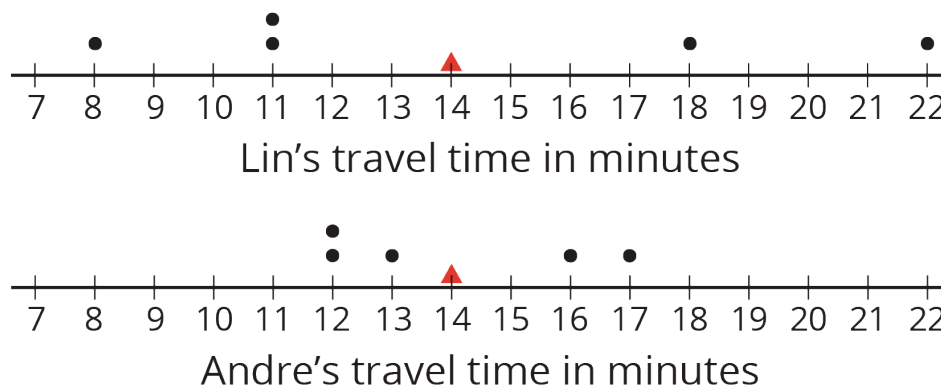
## Data Sets and Distributions

### Measures of Center and Variability

#### Family Support Materials 1

This week, your student will learn to calculate and interpret the **mean**, or the average, of a data set. We can think of the mean of a data set as a fair share—what would happen if the numbers in the data set were distributed evenly. Suppose a runner ran 3, 4, 3, 1, and 5 miles over five days. If the total number of miles she ran, 16 miles, was distributed evenly across five days, the distance run per day, 3.2 miles, would be the mean. To calculate the mean, we can add the data values and then divide the sum by how many there are.

If we think of data points as weights along a number line, the mean can also be interpreted as the balance point of the data. The dots show the travel times, in minutes, of Lin and Andre. The triangles show each mean travel time. Notice that the data points are “balanced” on either side of each triangle.



Your student will also learn to find and interpret the **mean absolute deviation** or the **MAD** of data. The MAD tells you the distance, on average, of a data point from the mean. When the data points are close to the mean, the distances between them and the mean are small, so the average distance—the MAD—will also be small. When data points are more spread out, the MAD will be greater.

We use mean and MAD values to help us summarize data. The mean is a way to describe the center of a data set. The MAD is a way to describe how spread out the data set is.

Here is a task to try with your student:

1. Use the data on Lin's and Andre's dot plots to verify that the mean travel time for each student is 14 minutes.

2. Andre says that the mean for his data should be 13 minutes, because there are two numbers to the left of 13 and two to the right. Explain why 13 minutes cannot be the mean.
3. Which data set, Lin's or Andre's, has a higher MAD (mean absolute deviation)? Explain how you know.

Solution:

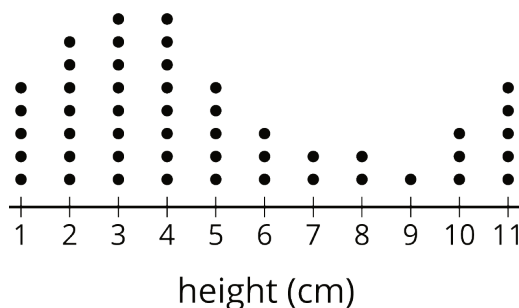
1. For Lin's data, the mean is  $\frac{8+11+11+18+22}{5} = \frac{70}{5}$ , which equals 14. For Andre's data, the mean is  $\frac{12+12+13+16+17}{5} = \frac{70}{5}$ , which also equals 14.
2. Explanations vary. Sample explanations:
  - The mean cannot be 13 minutes because it does not represent a fair share.
  - The mean cannot be 13 minutes because the data would be unbalanced. The two data values to the right of 13 (16 and 17) are much further away from the two that are to the left (12 and 12).
3. Lin's data has a higher MAD. Explanations vary. Sample explanations:
  - In Lin's data, the points are 6, 3, 3, 4, and 8 units away from the mean of 14. In Andre's data, the points are 2, 2, 1, 2, and 3 units away from the mean of 14. The average distance of Lin's data will be higher because those distances are greater.
  - The MAD of Lin's data is 4.8 minutes, and the MAD of Andre's data is 2 minutes.
  - Compared to Andre's data points, Lin's data points are farther away from the mean.

# Sampling

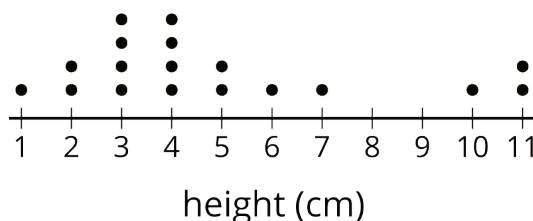
## Family Support Materials 2

This week your student will be working with data. Sometimes we want to know information about a group, but the group is too large for us to be able to ask everyone. It can be useful to collect data from a **sample** (some of the group) of the **population** (the whole group). It is important for the sample to resemble the population.

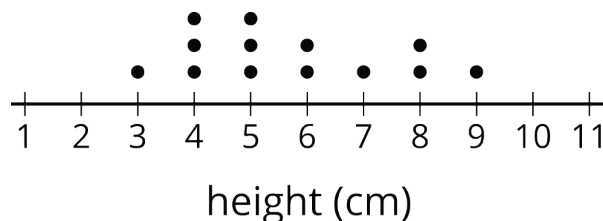
- For example, here is a dot plot showing a population: the height of 49 plants in a sprout garden.



- This sample is **representative** of the population, because it includes only a part of the data, but it still resembles the population in shape, center, and spread.



- This sample is not representative of the population. It has too many plant heights in the middle and not enough really short or really tall ones.



A sample that is selected at random is more likely to be representative of the population than a sample that was selected some other way.

Here is a task to try with your student:

A city council needs to know how many buildings in the city have lead paint, but they don't have enough time to test all 100,000 buildings in the city. They want to test a sample of buildings that will be representative of the population.

1. What would be a *bad* way to pick a sample of the buildings?
2. What would be a *good* way to pick a sample of the buildings?

Solution:

1. There are many possible answers.
  - Testing all the same type of buildings (like all the schools, or all the gas stations) would not lead to a representative sample of all the buildings in the city.
  - Testing buildings all in the same location, such as the buildings closest to city hall, would also be a bad way to get a sample.
  - Testing all the newest buildings would *bias* the sample towards buildings that don't have any lead paint.
  - Testing a small number of buildings, like 5 or 10, would also make it harder to use the sample to make predictions about the entire population.
2. To select a sample at random, they could put the addresses of all 100,000 buildings into a computer and have the computer select 50 addresses randomly from the list. Another possibility could be picking papers out a bag, but with so many buildings in the city, this method would be difficult.

## Probabilities of Single Step Events

### Family Support Materials 3

This week your student will be working with probability. A **probability** is a number that represents how likely something is to happen. For example, think about flipping a coin.

- The probability that the coin lands somewhere is 1. That is certain.
- The probability that the coin lands heads up is  $\frac{1}{2}$ , or 0.5.
- The probability that the coin turns into a bottle of ketchup is 0. That is impossible.

Sometimes we can figure out an exact probability. For example, if we pick a random date, the chance that it is on a weekend is  $\frac{2}{7}$ , because 2 out of every 7 days fall on the weekend. Other times, we can estimate a probability based on what we have observed in the past.

Here is a task to try with your student:

People at a fishing contest are writing down the type of each fish they catch. Here are their results:

- Person 1: bass, catfish, catfish, bass, bass, bass
- Person 2: catfish, catfish, bass, bass, bass, bass, catfish, catfish, bass, catfish
- Person 3: bass, bass, bass, catfish, bass, bass, catfish, bass, catfish

1. Estimate the probability that the next fish that gets caught will be a bass.
2. Another person in the competition caught 5 fish. Predict how many of these fish were bass.
3. Before the competition, the lake was stocked with equal numbers of catfish and bass. Describe some possible reasons for why the results do not show a probability of  $\frac{1}{2}$  for catching a bass.

Solution:

1. About  $\frac{15}{25}$ , or 0.6, because of the 25 fish that have been caught, 15 of them were bass.
2. About 3 bass, because  $\frac{3}{5} = 0.6$ . It would also be reasonable if they caught 2 or 4 bass, out of their 5 fish.

3. There are many possible answers. For example:
- Maybe the lures or bait they were using are more likely to catch bass.
  - With results from only 25 total fish caught, we can expect the results to vary a little from the exact probability.