



## Lesson 8: Variability in a Data Distribution

### Student Outcomes

- Students interpret the mean of a data set as a “typical” value.
- Students compare and contrast two small data sets that have the same mean but different amounts of variability.
- Students see that a data distribution is not characterized only by its center. Its spread or variability must be considered as well.
- Students informally evaluate how precise the mean is as an indicator of the typical value of a distribution, based on the variability exhibited in the data.
- Students use dot plots to order distributions according to the variability around the mean for each of the data distributions.

### Classwork

#### Example 1 (5 minutes): Comparing Two Distributions

##### Example 1: Comparing Two Distributions

Robert’s family is planning to move to either New York City or San Francisco. Robert has a cousin in San Francisco and asked her how she likes living in a climate as warm as San Francisco. She replied that it doesn’t get very warm in San Francisco. He was surprised, and since temperature was one of the criteria he was going to use to form his opinion about where to move, he decided to investigate the temperature distributions for New York City and San Francisco. The table below gives average temperatures (in degrees Fahrenheit) for each month for the two cities.

City	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
New York City	39	42	50	61	71	81	85	84	76	65	55	47
San Francisco	57	60	62	63	64	67	67	68	70	69	63	58

Read through the introductory paragraph as a class. Give students a moment to examine the table and then ask:

- How would you describe the temperatures in New York City?
  - *The temperatures change a lot through the year.*
- How would you describe the temperatures in San Francisco?
  - *The temperatures do not change much.*

**Exercises 1–2 (5 minutes)**

Let students work independently and confirm their answer with a neighbor. Encourage calculator use when working with larger data sets.

**Exercises 1–2**

Use the table above to answer the following:

1. Calculate the annual mean monthly temperature for each city.

*New York City: 63 degrees*

*San Francisco: 64 degrees*

2. Recall that Robert is trying to decide to which city he wants to move. What is your advice to him based on comparing the overall annual mean monthly temperatures of the two cities?

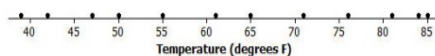
*Since the means are almost the same, it looks like Robert could move to either city.*

**Example 2 (5 minutes): Understanding Variability****Example 2: Understanding Variability**

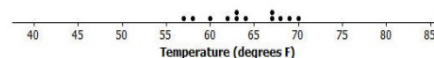
In Exercise 2, you found the overall mean monthly temperatures in both the New York City distribution and the San Francisco distribution to be about the same. That didn't help Robert very much in making a decision between the two cities. Since the mean monthly temperatures are about the same, should Robert just toss a coin to make his decision? Is there anything else Robert could look at in comparing the two distributions?

Variability was introduced in an earlier lesson. Variability is used in statistics to describe how spread out the data in a distribution are from some focal point in the distribution (such as the mean). Maybe Robert should look at how spread out the New York City monthly temperature data are from its mean and how spread out the San Francisco monthly temperature data are from its mean. To compare the variability of monthly temperatures between the two cities, it may be helpful to look at dot plots. The dot plots for the monthly temperature distributions for New York City and San Francisco follow.

Dot Plot of Temperature for New York City



Dot Plot of Temperature for San Francisco



MP.1

Read through the first paragraph as a class. Since the means are about the same, it would be helpful if Robert had more information as basis for a decision. He needs to go beyond comparing centers to incorporating variability into his decision-making process. Ask students:

- Should he just toss a coin to make a decision?
  - *Answers will vary.*
- What else do you think Robert could use to make a decision?
  - *He could consider the range or variety of temperatures in each city.*

Read though the second paragraph (above) and define variability. In this example, we want students to become familiar with the concept of variability by viewing how spread out the data are from their mean in dot plots. Give students a moment to examine the dot plots and ask:

- How are the two dot plots different?
  - *The temperatures for New York City are spread out, while the temperatures for San Francisco are clustered together.*

### Exercises 3–7 (5–7 minutes)

Let students work independently and compare answers with a neighbor.

#### Exercises 3–7

Use the dot plots above to answer the following:

3. Mark the location of the mean on each distribution with the balancing  $\Delta$  symbol. How do the two distributions compare based on their means?

*Place  $\Delta$  at 63 for New York City and at 64 for San Francisco. The means are about the same.*

4. Describe the variability of the New York City monthly temperatures from the mean of the New York City temperatures.

*The temperatures are widespread around the mean. From a low of around 40, to a high of 85.*

5. Describe the variability of the San Francisco monthly temperatures from the mean of the San Francisco monthly temperatures.

*The temperatures are compact around the mean. From a low of 57, to a high of 70.*

6. Compare the amount of variability in the two distributions. Is the variability about the same, or is it different? If different, which monthly temperature distribution has more variability? Explain.

*The variability is different. The variability in New York City is much greater compared to San Francisco.*

7. If Robert prefers to choose the city where the temperatures vary the least from month to month, which city should he choose? Explain.

*He should choose San Francisco because the temperatures vary the least, from a low of 57 to a high of 70. New York City has temperatures with more variability, from a low of 40, to a high of 85.*

**Example 3 (7–9 minutes): Using Mean and Variability in a Data Distribution****Example 3: Using Mean and Variability in a Data Distribution**

The mean is used to describe the “typical” value for the entire distribution. Sabina asks Robert which city he thinks has the better climate? He responds that they both have about the same mean, but that the mean is a better measure or a more precise measure of a typical monthly temperature for San Francisco than it is for New York City. She’s confused and asks him to explain what he means by this statement.

Robert says that the mean of 63 degrees in New York City (64 in San Francisco) can be interpreted as the typical temperature for any month in the distributions. So, 63 or 64 degrees should represent all of the months’ temperatures fairly closely. However, the temperatures in New York City in the winter months are in the 40s and in the summer months are in the 80s. The mean of 63 isn’t too close to those temperatures. Therefore, the mean is not a good indicator of typical monthly temperature. The mean is a much better indicator of the typical monthly temperature in San Francisco because the variability of the temperatures there is much smaller.

**MP.3**

The concept in this example may be challenging for some students. When Robert talks about the precision of the mean, Sabina asks him to explain what he means by a mean being precise.

Although the means are about the same for the two distributions, Robert is suggesting that the mean of 64 degrees for San Francisco is a better indicator of the city’s typical monthly temperature, than the mean of 63 degrees is as an indicator of a typical monthly temperature in New York City. He bases this on the variability of the monthly temperatures in each city. He says that a mean is a only precise indicator of monthly temperatures if the variability in the data is very low. The higher the variability gets, the less precise the mean is as an indicator of typical monthly temperatures.

If there is still confusion, draw two dot plots similar to Example 3 on the board and ask the following:

- Which dot plot has greater variability?
- If data points have a lot of variability, is the mean a good indicator of a “typical” value in the data set?
  - No.
- If the data points are clustered around the mean, is the mean a good indicator of a “typical” value in the data set?
  - Yes.

**Exercises 8–11 (5 minutes)**

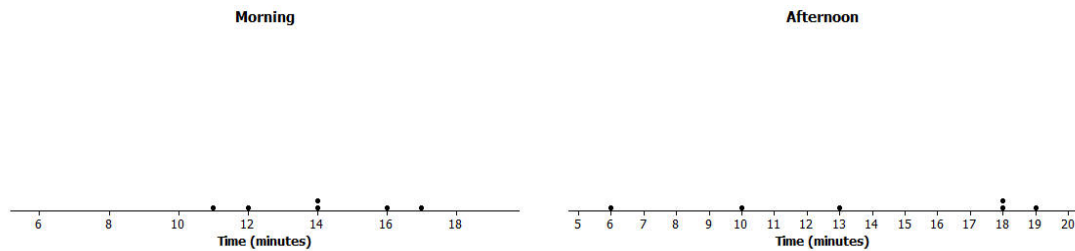
Let students work independently and confirm their answer with a neighbor.

**Exercises 8–14**

Consider the following two distributions of times it takes six students to get to school in the morning and to go home from school in the afternoon.

	Time (minutes)					
Morning	11	12	14	14	16	17
Afternoon	6	10	13	18	18	19

8. To visualize the means and variability, draw dot plots for each of the two distributions.



9. What is the mean time to get from home to school in the morning for these six students?

*The mean is 14 minutes. (Note: It is visible from the graphs.)*

10. What is the mean time to get from school to home in the afternoon for these six students?

*The mean is 14 minutes. (Note: The sum of the negative deviations is  $-13$ , and the sum of the positive deviations is  $+13$ .)*

11. For which distribution does the mean give a more precise indicator of a typical value? Explain your answer.

*The morning mean is a more precise indicator. The spread of the afternoon data is far greater around the mean.*

### Exercises 12–14 (7 minutes)

Let students work in pairs or small groups. If time allows, discuss Exercise 13 as a class.

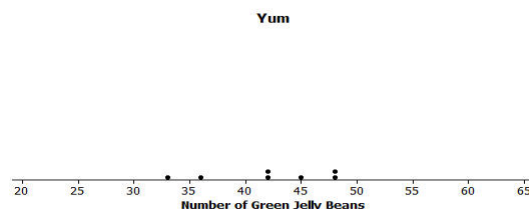
Distributions can be ordered according to how much the data values vary around their means.

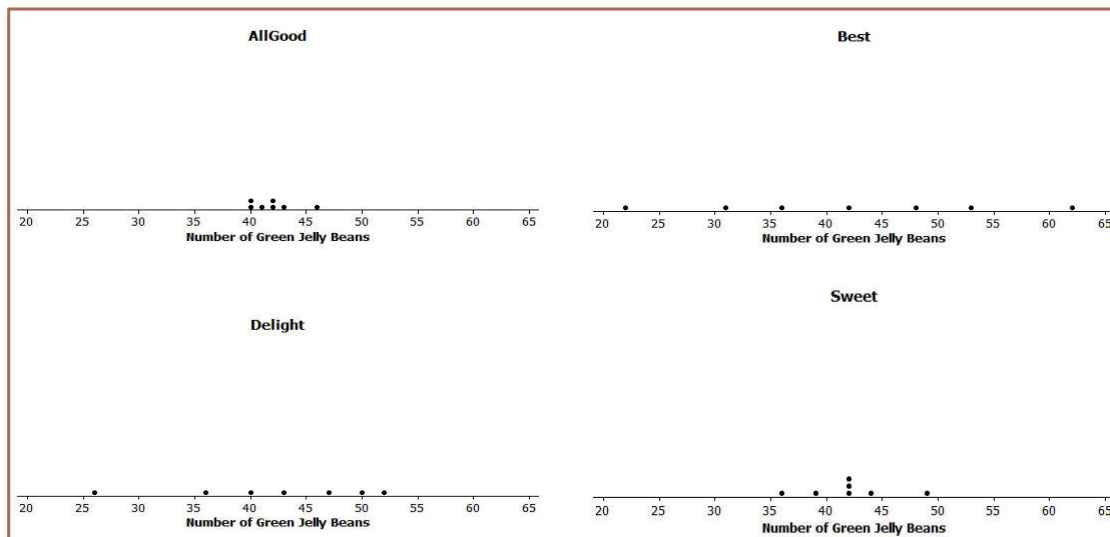
Consider the following data on the number of green jellybeans in seven bags of jellybeans from each of five different candy manufacturers (AllGood, Best, Delight, Sweet, Yum). The mean in each distribution is 42 green jellybeans.

	1	2	3	4	5	6	7
AllGood	40	40	41	42	42	43	46
Best	22	31	36	42	48	53	62
Delight	26	36	40	43	47	50	52
Sweet	36	39	42	42	42	44	49
Yum	33	36	42	42	45	48	48

12. Draw a dot plot of the distribution of number of green jellybeans for each of the five candy makers. Mark the location of the mean on each distribution with the balancing  $\Delta$  symbol.

*The dot plots should each have a balancing  $\Delta$  symbol located at 42.*





13. Order the candy manufacturers from the one you think has least variability to the one with most variability. Explain your reasoning for choosing the order.

*Note: Do not be critical, answers and explanations may vary. One possible answer:*

*In order from least to greatest: AllGood, Sweet, Yum, Delight, Best. The data points are all close to the mean for AllGood, which indicates it has the least variability, followed by Sweet and Yum. The data points are spread further from the mean for Delight and Best, which indicates they have the greatest variability.*

14. For which company would the mean be considered a better indicator of a typical value (based on least variability)?

*AllGood.*

#### Lesson Summary

We can compare distributions based on their means, but variability must also be considered. The mean of a distribution with small variability (not a lot of spread) is considered to be a better indication of a typical value than the mean of a distribution with greater variability (wide spread).

#### Exit Ticket (5 minutes)

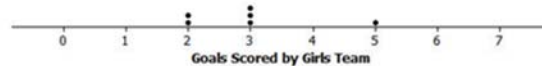
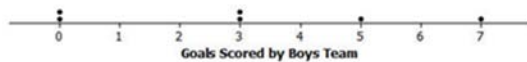
Name \_\_\_\_\_

Date \_\_\_\_\_

## Lesson 8: Variability in a Data Distribution

### Exit Ticket

1. Consider the following statement: Two sets of data with the same mean will also have the same variability. Do you agree or disagree with this statement? Explain.
2. Suppose the dot plot on the left shows the number of goals a boys' soccer team has scored in 6 games so far this season, and the dot plot on the right shows the number of goals a girls' soccer team has scored in 6 games so far this season.



- a. Compute the mean number of goals for each distribution.
- b. For which distribution, if either, would the mean be considered a better indicator of a typical value? Explain your answer.

## Exit Ticket Sample Solutions

1. Consider the following statement: Two sets of data with the same mean will also have the same variability. Do you agree or disagree with this statement? Explain.

*Students should disagree with this statement. There were many examples in this lesson that could be used as the basis for an explanation.*

2. Suppose the dot plot on the left shows the number of goals a boys' soccer team has scored in 6 games so far this season, and the dot plot on the right shows the number of goals a girls' soccer team has scored in 6 games so far this season.



- a. Compute the mean number of goals for each distribution.

*The mean for each is 3 goals. If your students found the mean by the formula, have them verify the answer by summing the negative and positive deviations.*

- b. For which distribution, if either, would the mean be considered a better indicator of a typical value? Explain your answer.

*Variability in the girls' distribution is less than in the boys', so the mean of 3 goals for the girls' is more precise.*

## Problem Set Sample Solutions

1. The number of pockets in the clothes worn by seven students to school yesterday were 4, 1, 3, 4, 2, 2, 5. Today those seven students each had three pockets in their clothes.

- a. Draw one dot plot for what the students wore yesterday, and another dot plot for what the students wore today. Be sure to use the same scales. Show the means by using the balancing  $\Delta$  symbol.



- b. For each distribution, find the mean number of pockets worn by the seven students.

*Students should not need to use the formula to calculate the means for either of these distributions. The first is clearly balanced around 3 pockets, and the second only has 3 as its data value.*

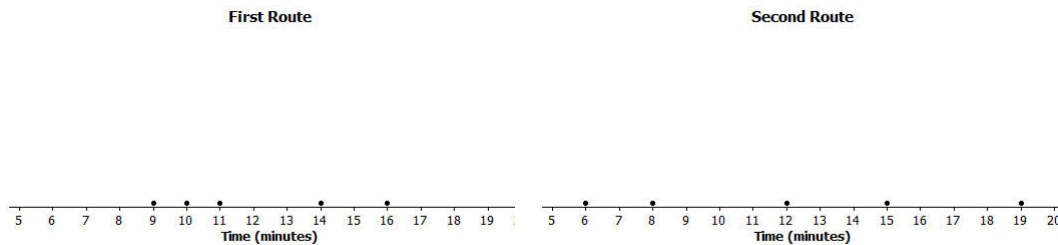


- c. For which distribution is the mean number of pockets a better indicator of what is “typical?” Explain.

*There is certainly variability in the data for the yesterday’s distribution, whereas today’s distribution has none. The mean of 3 pockets is a better indicator (more precise) for today’s distribution.*

2. The number of minutes (rounded) it took to run a certain short cross-country route was recorded for each of five students. The resulting data were 9, 10, 11, 14, and 16 minutes. The number of minutes (rounded to the nearest minute) it took the five students to run a different cross-country route was also recorded, resulting in the following data: 6, 8, 12, 15, and 19 minutes.

- a. Draw dot plots for the two distributions of the time it takes to run a cross-country route. Be sure to use the same scale on both dot plots.



- b. Do the distributions have the same mean?

*Yes, both distributions have the same mean, 12 minutes.*

- c. In which distribution is the mean a better indicator of the typical amount of time taken to run its cross-country route? Explain.

*Looking at the dot plots, the times completing the second route are more varied than those in the first route. So, the mean in the first route is a better indicator (more precise) of a typical value.*

3. The following table shows the prices per gallon of gasoline (in cents) at five stations across town as recorded on Monday, Wednesday, and Friday of a certain week.

Day	R&C	Al's	PB	Sam's	Ann's
Monday	359	358	362	359	362
Wednesday	357	365	364	354	360
Friday	350	350	360	370	370

- a. The mean price per day over the five stations is the same for the three days. Without doing any calculation and simply looking at Friday’s prices, what must the mean price be?

*Friday’s prices are symmetric around 360 cents. So, the mean is 360 cents.*

- b. In which daily distribution is its mean a better indicator of the typical price per gallon for the five stations? Explain.

*Note that the data are not in numerical order across the stations for Monday and Wednesday prices. So, encourage students to draw dot plots to help them answer this question. From the dot plots, the mean for Monday is the most precise (least variability), and the mean for Friday is the least precise (most variability).*