Lesson 8: Variability in a Data Distribution

Classwork

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

Exercises 1–2

Use the table above to answer the following:

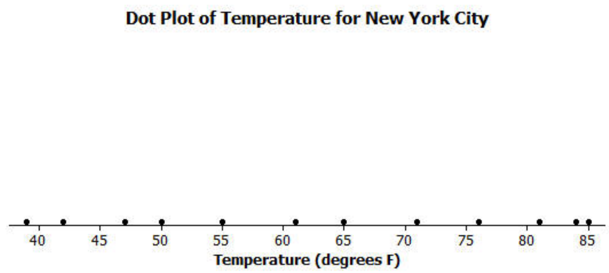
1. Calculate the annual mean monthly temperature for each city.

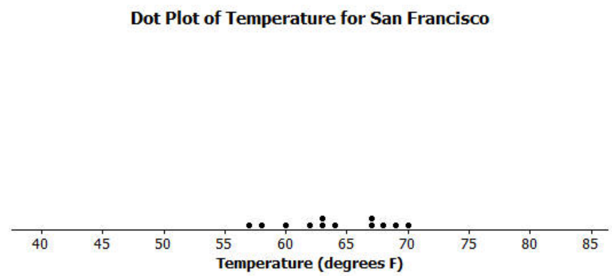
2. Recall that Robert is trying to decide to which city he wants to move. What is your advice to him based on comparing the overall annual mean monthly temperatures of the two cities?

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.





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 Δ symbol. How do the two distributions compare based on their means?

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

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

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.

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

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.

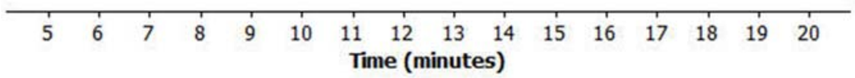
Exercises 8–14

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.

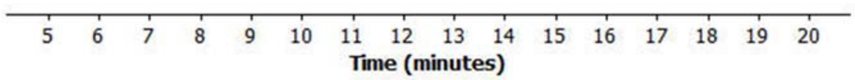
	Time (minutes)					
Morning	11	12	14	14	16	17
Afternoon	6	10	13	18	18	19

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

Morning



Afternoon



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

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

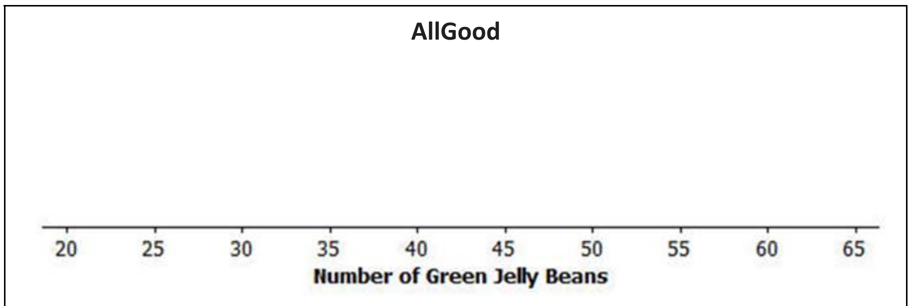
11. For which distribution does the mean give a more precise indicator of a typical value? Explain your answer.

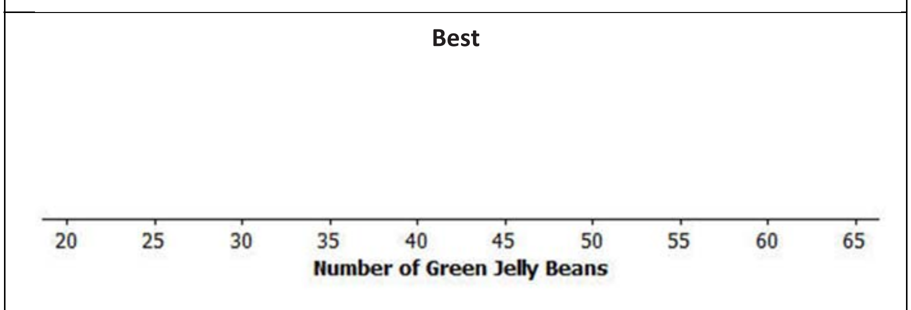
Distributions can be ordered according to how much the data values vary around their means.

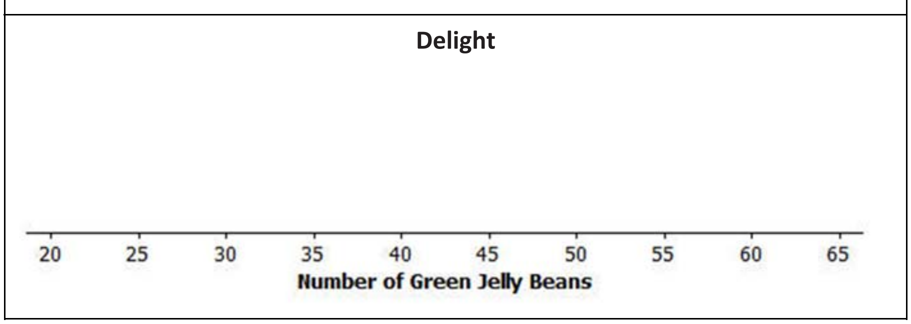
Consider the following data on the number of green jellybeans in seven bags of jellybeans from each of five different candy manufacturers (AllGood, Best, Delight, Sweet, Yum). The mean in each distribution is 42 green jellybeans.

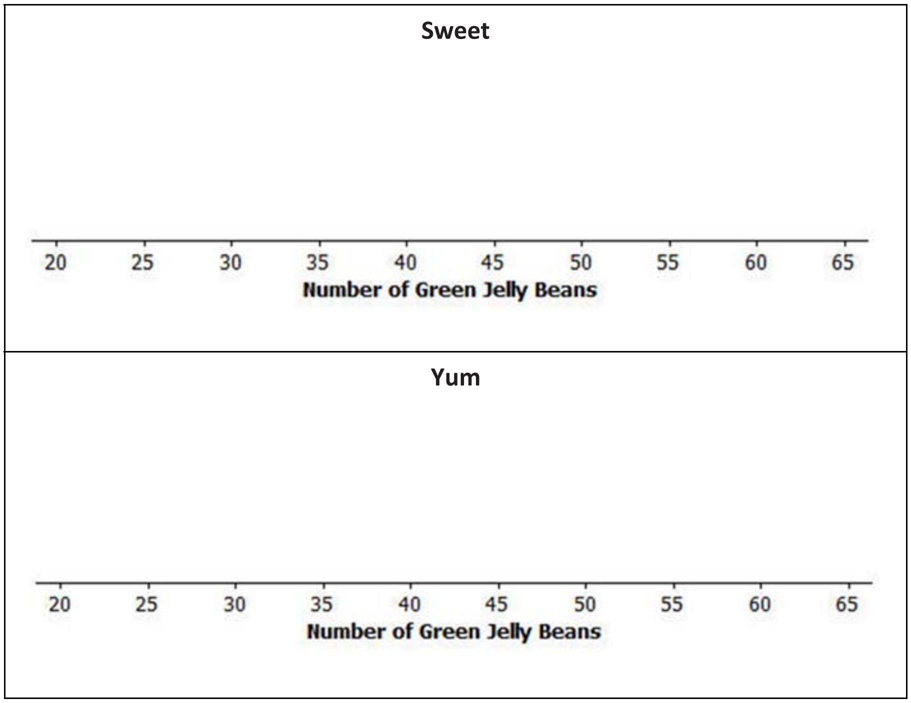
	1	2	3	4	5	6	7
AllGood	40	40	41	42	42	43	46
Best	22	31	36	42	48	53	62
Delight	26	36	40	43	47	50	52
Sweet	36	39	42	42	42	44	49
Yum	33	36	42	42	45	48	48

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 Δ symbol.









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.

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

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).

Problem Set

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 Δ symbol.
   b. For each distribution, find the mean number of pockets worn by the seven students.
   c. For which distribution is the mean number of pockets a better indicator of what is "typical?" Explain.

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

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