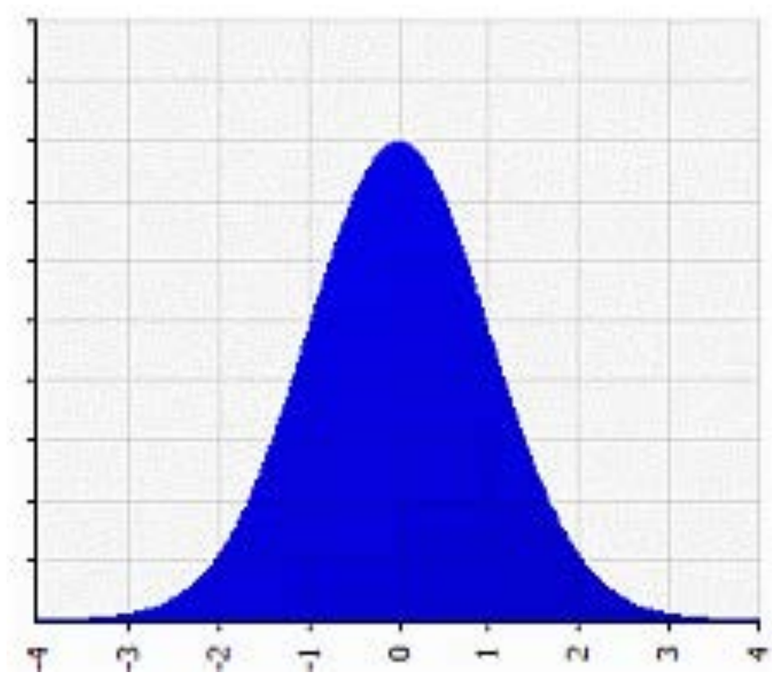
Lesson 8: Bell Curves and Standard Deviation

Opening Reading

1. Read over the description of a bell curve and then mark the picture with the characteristics of the curve. Which characteristic was confusing for you?

What is a Bell Curve or Normal Curve?

A bell curve is another name for a normal distribution curve (sometimes just shortened to "normal curve") or Gaussian distribution. The name comes from the fact it looks bell-shaped.

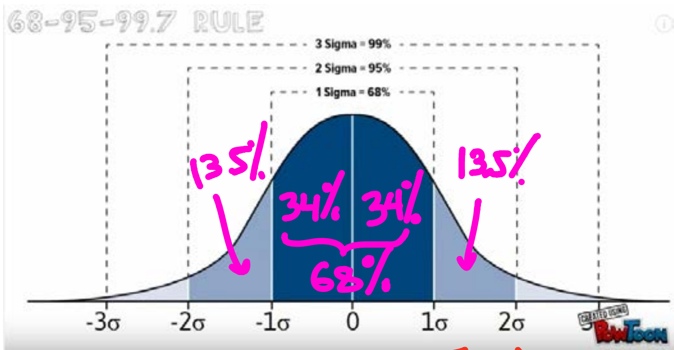


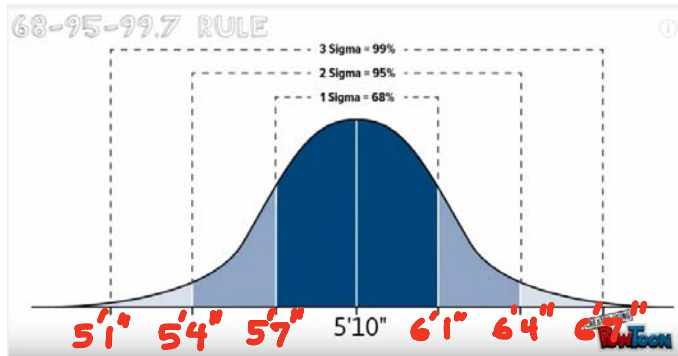
Characteristics of Bell Curves, Normal Curves

1. The mean (average) is always in the center of a bell curve or normal curve.
2. A bell curve / normal curve has only one mode, or peak. Mode here means "peak"; a curve with one peak is unimodal; two peaks is bimodal, and so on.
3. A bell curve / normal curve has predictable standard deviations that follow the 68 95 99.7 rule (see below).
4. A bell curve / normal curve is symmetric. Exactly half of data points are to the left of the mean and exactly half are to the right of the mean.

[source: http://www.statisticshowto.com/bell-curve/]

standard deviation - the average distance away from the mean.





σ sigma

Youtube video on bell curve and SD at https://www.youtube.com/watch?v=MRqtXL2WX2M

Bell Curve Probabability and Standard Deviation

To understand the probability factors of a normal distribution you need to understand the following "rules":

1. The total area under the curve is equal to 1 (100%).
2. About 68% of the area under the curves falls within 1 standard deviation.
3. About 95% of the area under the curve falls within 2 standard deviations.
4. About 99.7% of the area under the curve falls within 3 standard deviations.

Items 2, 3 and 4 are sometimes referred to as the "empirical rule" of the 68-95-99.7 rule.

1σ 2σ 3σ

[source: https://www.thoughtco.com/bell-curve-normal-distribution-defined-2312350]

2. A. If 200 people were in the data set above, about how many would you expect to be within 1 standard deviation of the mean?

$\frac{68}{100} = \frac{x}{200} \rightarrow x = 136$

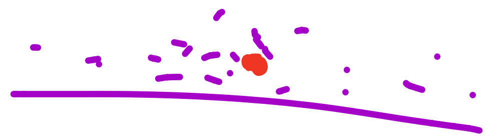
B. The standard deviation for men is about 3 inches. Label the graph above right with the heights of men at each standard deviation marking.

3. We've used the term Standard Deviation several times in the Opening Reading, but what does Standard Deviation mean? What the YouTube video ***How to Calculate Standard Deviation*** at https://www.youtube.com/watch?v=WVx3MYd-Q9w.

Answer the questions below as you watch the video.

A. What does standard deviation tell us about the data?

How spread the data is



B. Complete the steps for finding the standard deviation.

- First, find the mean of the data set. This is symbolized by $\bar{x}$.
- The second step is to subtract the mean from each data point. $x - \bar{x}$ (negative ok)
- The third step is to square each difference. This makes the difference positive so they don't cancel each other out.
- find the sum
- The fourth step is to calculate the mean of the squared diff.

Σ = sum

- The final step is to take the square root. This counteracts the squaring we did earlier.

$$\sigma = \sqrt{\frac{\Sigma (x-\bar{x})^2}{n}}$$

C. When do you divide by *n* and when do you divide by *n* – 1 in the fourth step?

(n-1) is used for a sample of the population.

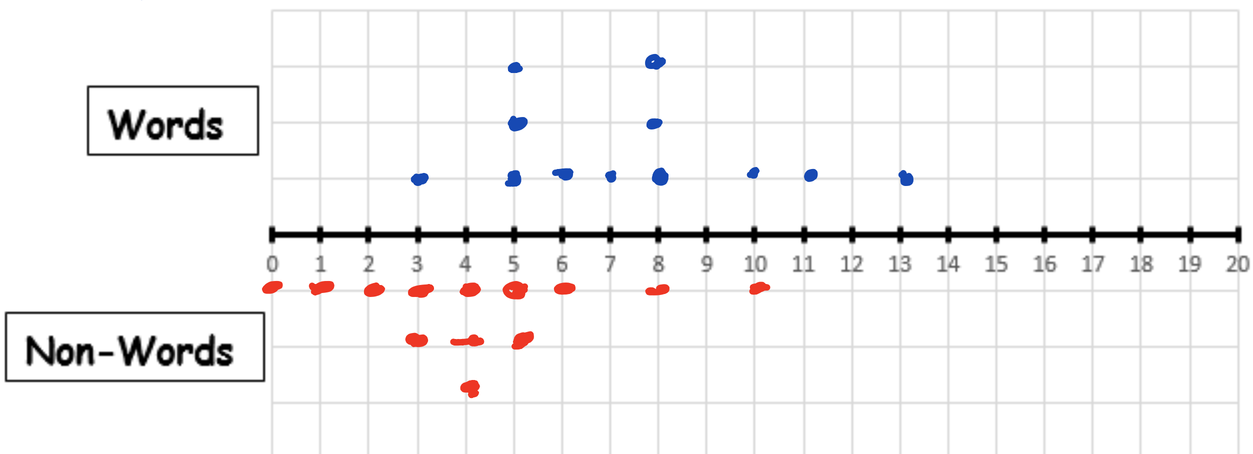
Exploratory Challenge 1 – Finding the Standard Deviation

You will need: the Lesson 1 data from **Station 8 – Test Your Memory – WQI versus MOP** handout

You'll be completing the table on the handout with the data from the **Test Your Memory** station. We are going to follow a multi-step process to calculate the **Standard Deviation**, which will help us measure how spread out the values are. Only the first row of each table is filled in with your class' data.

4. First we'll look at the shape of the data using a dot plot. Create dot plots for each set of data (real words versus non-words) in the space at the below.

7, 5, 5, 13, 7, 5, 11, 10, 8, 3, 8, 6



3, 4, 2, 10, 4, 5, 6, 1, 0, 3, 8, 4, 5

5. Do the number of 3-letter "words" differ more when the letters are random or when they form a word in English? Explain your thinking.

6. Did either graph form a bell curve? Why do you think this happened?

7. Even if the data did not form a bell curve, we'll use it to find the standard deviation of each set of data. Complete the steps below.

   A.. Calculate the mean of each data set.

   B. Calculate the deviations from the mean (actual value minus the mean) for the remaining values, and write your answers in the appropriate places in the table. Remember the mean for the "words" and the mean for the "non-words" are probably not the same.

   C. Square the deviations from the mean for each data set.

   D. Add up the squared deviations. This result is the *sum* of the squared deviations.

   For the "words": ________________

   For the "non-words": ________________

   E. The number of values in the data set is denoted by n. You divide the sum of the squared deviations by $n - 1$.

   For the "words": ________________

   For the "non-words": ________________

   F. Finally, you take the square root of the value you found in Part E. Units for Standard Deviation are the same as the units for the data set (number of words, in this case). What does this answer mean?

   For the "words": ________________

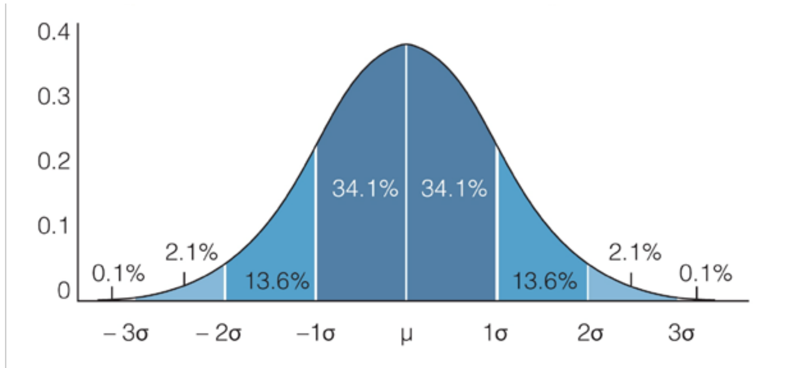
   For the "non-words": ________________

8. Which set of data had the greatest standard deviation? How can you see this in the graph?

9. Mark off the mean, 1 standard deviation, 2 standard deviations, and e standard deviations on the graph. Does this follow the 68-95-99.7 rule?

Lesson Summary

Bell Curve



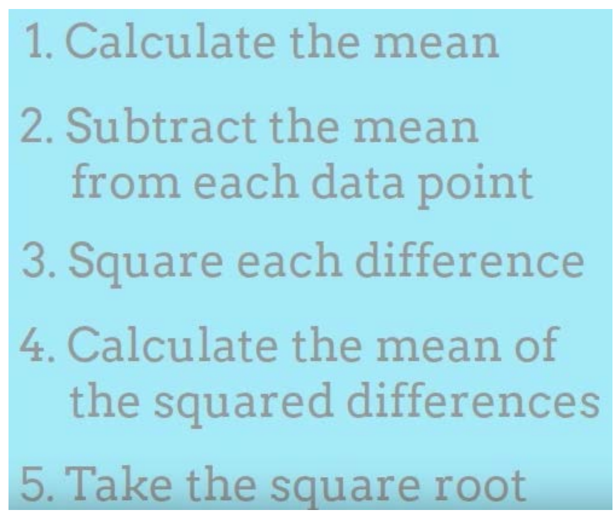
Standard Deviation

The formula for the standard deviation:

- x is a value from the original data set;
- $x - \bar{x}$ is a deviation of the value, x, from the mean, $\bar{x}$;
- $(x - \bar{x})^2$ is a squared deviation from the mean;
- $\sum(x - \bar{x})^2$ is the sum of the squared deviations;
- $\frac{\sum(x-\bar{x})^2}{n-1}$ is the result of dividing the sum of the squared deviations by $n - 1$;

So, $\sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$ is the standard deviation.

Description of the standard deviation steps:

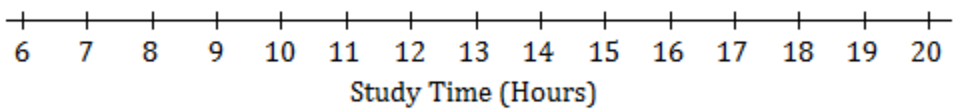


[source: https://www.youtube.com/watch?v=WVx3MYd-Q9w]

Homework Problem Set

1. Ten members of a high school girls' basketball team were asked how many hours they studied in a typical week. Their responses (in hours) were 20, 13, 10, 6, 13, 10, 13, 11, 11, 10.

   a. Using the axis given below, draw a dot plot of these values. (Remember, when there are repeated values, stack the dots with one above the other.)



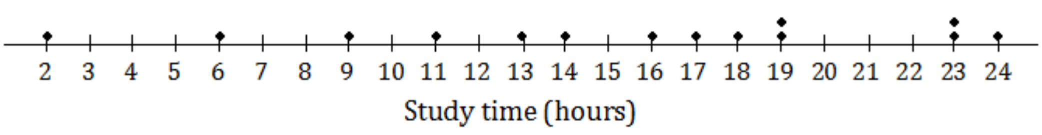
Girls' Basketball Team

   b. Calculate the mean study time for these students.

   c. Calculate the deviations from the mean for these study times, and write your answers in the appropriate places in the table below.

Number of Hours Studied	20	13	10	6	13	10	13	11	11	10
Deviation from the Mean										

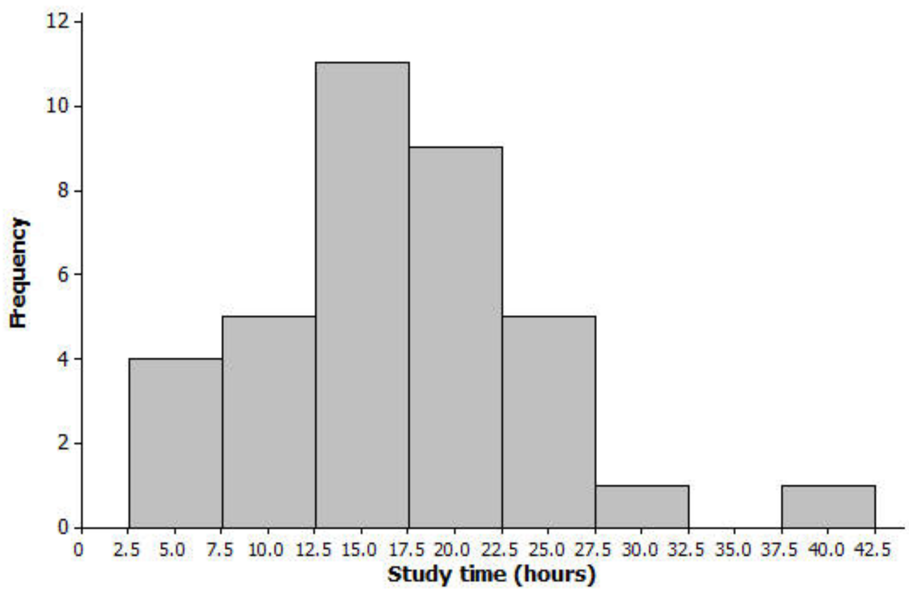
   d. The study times for fourteen girls from the soccer team at the same school as the one above are shown in the dot plot below.



Girls' Soccer Team

Based on the data, would the deviations from the mean (ignoring the sign of the deviations) be greater or less for the soccer players than for the basketball players?

2. All the members of a high school softball team were asked how many hours they studied in a typical week. The results are shown in the histogram below.
(The data set in this question comes from NCTM Core Math Tools, http://www.nctm.org/Classroom-Resources/Core-Math-Tools/Data-Sets/)



Softball Team

a. We can see from the histogram that four students studied around 5 hours per week. How many students studied around 15 hours per week?

b. How many students were there in total?

c. Suppose that the four students represented by the histogram bar centered at 5 had all studied exactly 5 hours, the five students represented by the next histogram bar had all studied exactly 10 hours, and so on. If you were to add up the study times for all of the students, what result would you get?

d. What is the mean study time for these students?

e. What would you consider to be a typical deviation from the mean for this data set?

sample

3. A small car dealership tests the fuel efficiency of sedans on its lot. It chooses 12 sedans for the test. The fuel efficiency (mpg) values of the cars are given in the table below. Complete the table as directed below.

$n = 12$

Fuel Efficiency (miles per gallon)	29	35	24	25	21 (21−25.5)	21	18	28	31	26	26	22
Deviation from the Mean	3.5	9.5	-1.5	-0.5	-4.5	-4.5	-7.5	2.5	5.5	0.5	0.5	-3.5
Squared Deviation from the Mean	3.5^2 12.25	90.25	2.25	0.25	20.25	20.25	56.25	6.25	30.25	0.25	0.25	12.25

a. Calculate the mean fuel efficiency for these cars.

$\bar{x} = 25.5 \leftarrow \frac{306}{12}$

b. Calculate the deviations from the mean, and write your answers in the second row of the table.

$x - \bar{x}$

c. Square the deviations from the mean, and write the squared deviations in the third row of the table.

$(x - \bar{x})^2$

d. Find the sum of the squared deviations.

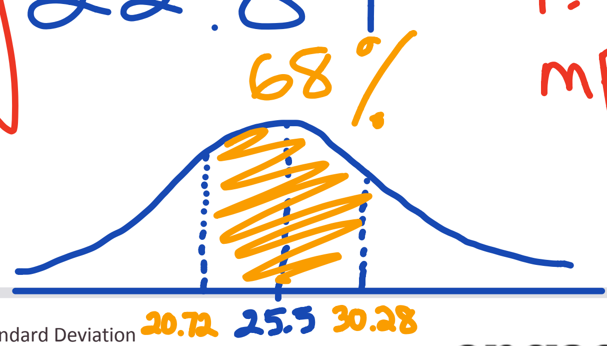
$\sum (x - \bar{x})^2 = 251.25$

e. What is the value of n for this data set? Divide the sum of the squared deviations by $n - 1$.

$n = 12 \rightarrow n - 1 = 11$

f. Take the square root of your answer to part (e) to find the standard deviation of the fuel efficiencies of these cars. Round your answer to the nearest hundredth.

$\frac{251.25}{11} \approx \sqrt{22.84} \rightarrow 4.78$ mpg



4. The same dealership decides to test fuel efficiency of SUVs. It selects six SUVs on its lot for the test. The fuel efficiencies (in miles per gallon) of these cars are shown below.

$$21 \quad 21 \quad 21 \quad 30 \quad 28 \quad 24$$

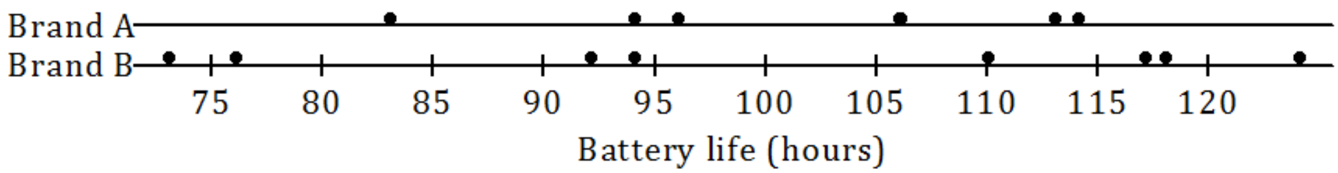
Calculate the mean and the standard deviation of these values. Be sure to show your work, and include a unit in your answer.

5. Consider the following questions regarding the cars described in Problems 3 and 4.

   a. What is the standard deviation of the fuel efficiencies of the cars in Problem 3? Explain what this value tells you.

   b. You also calculated the standard deviation of the fuel efficiencies for the cars in Problem 4. Which of the two data sets (Problem 3 or Problem 4) has the larger standard deviation? What does this tell you about the two types of cars (sedans and SUVs)?

A consumers' organization is planning a study of the various brands of batteries that are available. They are interested in how long a battery can be used before it must be replaced. As part of its planning, it measures lifetime for each of six batteries of Brand A and eight batteries of Brand B. Dot plots showing the battery lives for each brand are shown below.



6. Does one brand of battery tend to last longer, or are they roughly the same? What calculations could you do in order to compare the battery lives of the two brands?

7. Do the battery lives tend to differ more from battery to battery for Brand A or for Brand B? Explain your thinking.

8. Would you prefer a battery brand that has battery lives that do not vary much from battery to battery? Why or why not?

The table below shows the lives (in hours) of the Brand A batteries. We are going to follow a multi-step process to calculate the **Standard Deviation**, which will help us measure how spread out the values are – how long can we expect an average Brand A Battery to last. The first value has been done for you.

de·vi·a·tion

1. the action of departing from an established course or accepted standard.
2. Statistics: the amount by which a single measurement differs from a fixed value such as the mean.

Life (Hours)	83	94	96	106	113	114
Deviation from the Mean	-18					
Squared Deviations from the Mean	324					

9. A. Calculate the mean.

B. Calculate the deviations from the mean (actual value minus the mean) for the remaining values, and write your answers in the appropriate places in the table. The first one has been done for you.

C. Square the deviations from the mean. For example, when the deviation from the mean is -18 the squared deviation from the mean is $(-18)^2 = 324$.

D. Add up the squared deviations (This result is the *sum* of the squared deviations).

E. The number of values in the data set is denoted by n. In this example, n is 6. You divide the sum of the squared deviations by $n - 1$, which here is $6 - 1 = 5$.

F. Finally, you take the square root of the value you found in Part E. Units for Standard Deviation are the same as the units for the data set (hours, in this case). What does this answer mean?

The table below shows the battery lives and the deviations from the mean for Brand B.

Life (Hours)	73	76	92	94	110	117	118	124
Deviation from the Mean	-27.5							
Squared Deviation from the Mean								

10. Follow the steps to find and interpret the Standard Deviation for Brand B.

A. Calculate the mean.

B. Calculate the deviations from the mean (actual value minus the mean) for the remaining values, and write your answers in the appropriate places in the table.

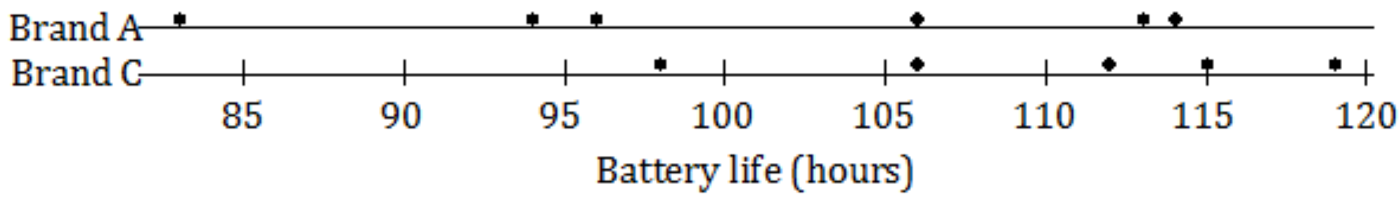
C. Square the deviations from the mean.

D. Add up the squared deviations (This result is the *sum* of the squared deviations).

E. Divide the sum of the squared deviations by $n-1$.

F. Finally, you take the square root of the value you found in Part E. Units for Standard Deviation are the same as the units for the data set (hours, in this case). What does this answer mean?

The lives of five batteries of a third brand, Brand C, were determined. The dot plot below shows the lives of the Brand A and Brand C batteries.



11. Which brand has the greater mean battery life? (You should be able to answer this question without doing any calculations.)

12. Which brand shows greater **variability**?

Variability: The extent to which data points diverge from the mean and from each other.

13. Which brand would you expect to have the greater standard deviation?

14. Why would we want to know this information? (Quality control, reliability, etc.)

REVIEW – Slope

15. For each slope and point given below, determine a point that is on the graph of the line determined by the slope and point. For example, if the slope is $\frac{1}{3}$ and the point is (1, 2), then (4, 3) or (-2, 1) are also on the line with slope $\frac{1}{3}$ and point (1, 2). There are many possible answers to each one.

A. slope = $\frac{4}{5}$ and point A (0, 3)

B. slope = $-\frac{1}{2}$ and point B (2, 0)

C. slope = 3 and point C (2, 3)

D. slope = -4 and point D (-1, 2)

16. Explain the method you used to complete Exercise 15.