Lesson 04: Interpreting Data (Center, Spread, Shape)

Lesson Objectives:

- Student will estimate and calculate central measures of data (mean, median, mode, & mid range)
- Student will estimate and calculate the spread of the data (range, variance, Standard deviation, & Inter quartile Range)
- Student will make interpretations based on the shape of the data (symmetric, skewed)

Central Measures of Data

Measures of Central Tendency describe the middle of the data set. Researchers use measures of central tendency to identify trends in the data. There are four measures of central tendency that we will explore in this course. Mean, median, mode, & mid range are described in the table below:

Measure of Central Tendency	Definition	How to calculate
Mean	The mean is the arithmetic average of all values in the data set.	Arrange values in the data set from least to greatest, then add all of the values together and divide by the number of values that you have in all.
Median	The median is the middle most value in the data set.	Arrange values in the data set from least to greatest, then select the middle most value. When you have an odd number of values it is simple to locate the median. When there is an even number of values then select the two middle most values and find the arithmetic average of the two to calculate the median.
Mode	The mode is the value in the data	Determine the frequency that each value in the data set occurs. The value that occurs the most is the mode. There may be one mode,

	set that occurs the most.	two modes, or three modes if different values occur in the data set the same number of times. In this case, list all modes and deem the data set multi modal.
Mid range	The mid range is the arithmetic average of the minimum and maximum values in the data set.	Arrange values in order from least to greatest. Add the minimum value to the maximum value and divide by 2 to calculate the mid range.

Example:

Torrie keeps up with her test grades in math class and wants to calculate the measures of central tendency for her data. Torrie scored 80, 79, 68, 90, 79, 92 on tests she took in math class.

First, arrange the values in the data set from least to greatest.

68 79 79 80 90 92

Find the Mean: (68 + 79 + 79 + 80 + 90 + 92)/6 = 488/6 = 81.3
Remember to divide by the number of values that you have in the data set.

Find the Mode: This data set has only one mode since 79 occurs twice and that's more times than any other value in the data set.

Find the Median: Since there are an even number of values in the data set, we have to find the arithmetic average of the two middle most values in the data set. (79 + 80)/2 = 159/2 = 79.5

Find the Mid Range: The minimum value in the data set it 68 and the maximum value is 92 so (68 + 92)/2 = 160/2 = 80

Here is a video: Mean Median and Mode

Now you try. Click on the Practice Problem buttons below to practice finding measures of central tendency:

Given this set of numbers: 84, 79, 89, 62, 45, 93, 89, 74, 78, 76, 80, 85

Spread of Data

Measures of central tendency use a single number to describe the data set. It is also important to use single numbers to describe the spread of the data. Range, variance, standard deviation, and inter quartile range all describe the spread of the data. Each descriptor of the spread of the data is described in the table below:

Spread of Data	Definition	How to Calculate
Range	The range determines the difference between the minimum and maximum values of the data set.	Arrange values from least to greatest and then subtract the minimum value from the maximum value to calculate the range.
Variance	Variance is a parameter since it describes the population but sample variance is a statistic since it describes the sample. Variance looks at how each value in the data set compares to the mean.	Calculate the mean of the data. Subtract the mean from each value in the data set and square those numbers. Then add all of the squared numbers together and divide by the number of values in the data set. This is the population variance. For sample variance instead of dividing by the number of values in the data set you divide by the number of values in the data set minus 1. For more information click Variance
Standard Deviation	The standard deviation is a statistic and it describes the sample.	The standard deviation is the square root of the variance.
Inter Quartile Range (IQR)	The inter quartile range is the difference between the third and first quartile.	Arrange data from least to greatest. Calculate the median. List the values that are less than the overall median and find the median of that data. This is the first quartile. List the values that are greater than the overall median and find the median of that data. This is the third quartile. Subtract the first quartile from the third quartile and this value is the Inter Quartile Range.

Here are some videos:

Example

Flossy recorded the length of time in minutes that it takes her to get dressed for school each morning for a week.

22 mins., 28 mins., 35 mins., 30 mins., 15 mins

Calculate the range, variance, standard deviation, & inter quartile range

First, arrange the values from least to greatest.

15 22 28 30 35

Find the Range: maximum value - minimum value = 35 - 15 = 20

Find the Sample Variance:

Value in data set	Mean of data set	Difference of data set value and mean	(Difference of data set value and mean)^2
15	26	-11	121
22	26	-4	16
28	26	2	4
30	26	4	16
35	26	9	81

Now we total the values in the last column and divide that by the number of values minus 1: (121 + 16 + 4 + 16 + 81)/4 = 238/4 = 59.5

Find the Standard Deviation: $\sqrt{59.5}$ = 7.71

Find the Inter Quartile Range: The median is 28. 15 and 22 are less than the median so to find the first quartile we find the median(in this case the mean

since there are an even number of values below the median) of those two numbers (15 + 22)/2 = 37/2 = 18.5= first quartile. To find the third quartile, 30 and 35 are greater than the median so we find the median of those two numbers (30 + 35)/2 = 65/2 = 32.5 = third quartile. Now we subtract the first quartile from the third to determine the inter quartile range 32.5 - 18.5 = 14.

Now you try:

Click on the Practice Problem buttons below to practice finding the spread of data.

Given this set of numbers   32, 14, 19, 28

Find the range:

Practice Problem	5

Find the variance:

Practice Problem	6

To do some more practice, click the links below:

[Variance](#)

[Standard deviation](#)

Making Interpretations based on the Shape of Data

Recall from the previous lesson that a histogram is a graphical representation of data that separates the data into classes or groups to make the details of the data set specific, but at the same time this graph shows the pattern of the data overall. On the x-axis, the classes or groups are arranged, and on the y-axis the number of occurrences for each class or group are arranged.

A histogram is considered to be symmetric if it has an equal amount of values below and above the mean. This type of graph is also known as a normal distribution and its shadow resembles that of a bell curve.

When a normal distribution is not present, then the graph is considered to be skewedeither negatively or positively depending upon where the tail of the distribution lies.

More on Skewed Graphs

A skewed histogram is not a normal curve like the normal distribution that you will learn about later in this course. If a graph is skewed positive then that means the tail of the curve is toward the right and that there are a greater number of low values in the data set with outliers on the high end. This is shown in the first graph below. A graph that is skewed negative is just the opposite meaning that the data set has a greater number of high values with outliers on the low end. This is shown in the second graph below. The third graph is a normal distribution meaning that the values are symmetrically positioned around the mean of the data.

Grading Rubric:

To get a 10: All answers are correct the first time, or within first revision.
To get a 9: All answers are correct after more than 1 revision.
To get an 8: You can have 1 incorrect answer.
To get a 7: You can have 2 incorrect answers.

To get a 5: Cheating- Plagiarism - purposeful or mistaken, which will lower your final grade for the course (so be very careful when posting your work!); lack of effort, disrespect, or attitude (we are here to communicate with you if you don't understand something);

Note: For this class it is necessary to post the questions over each answer. Failure to do so will result in asking for a revision.   No grade will be given for incomplete work.

Assignment:

Use the following data set to answer questions 1-8 and show all work:

2, 5, 8, 21, 9, 18, 21, 14, 15, 10, 8, 10

1. Calculate the mean of the data set.

2. Calculate the median of the data set.

3. Calculate the mode of the data set.

4. Calculate the mid range of the data set.

5. Calculate the range of the data set.

6. Subtract the mean from each value in the data set.

7. Square the difference of the mean and each value in the data set.

8. Add all of the squared differences together.

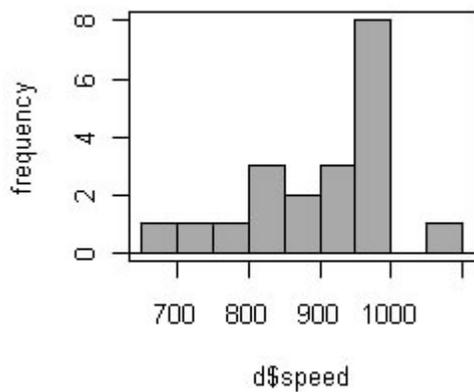9. Calculate the sample variance of the data set.

10. Calculate the sample standard deviation of the data set.

11. Calculate the first quartile of the data set

12. Calculate the third quartile of the data set.

13. Calculate the inter quartile range of the data set.

Look at the graphs below to determine whether the graph is a normal distribution, negatively skewed, or positively skewed, and give one reason to support your answer.

14.

Kibbie Extended DNBR
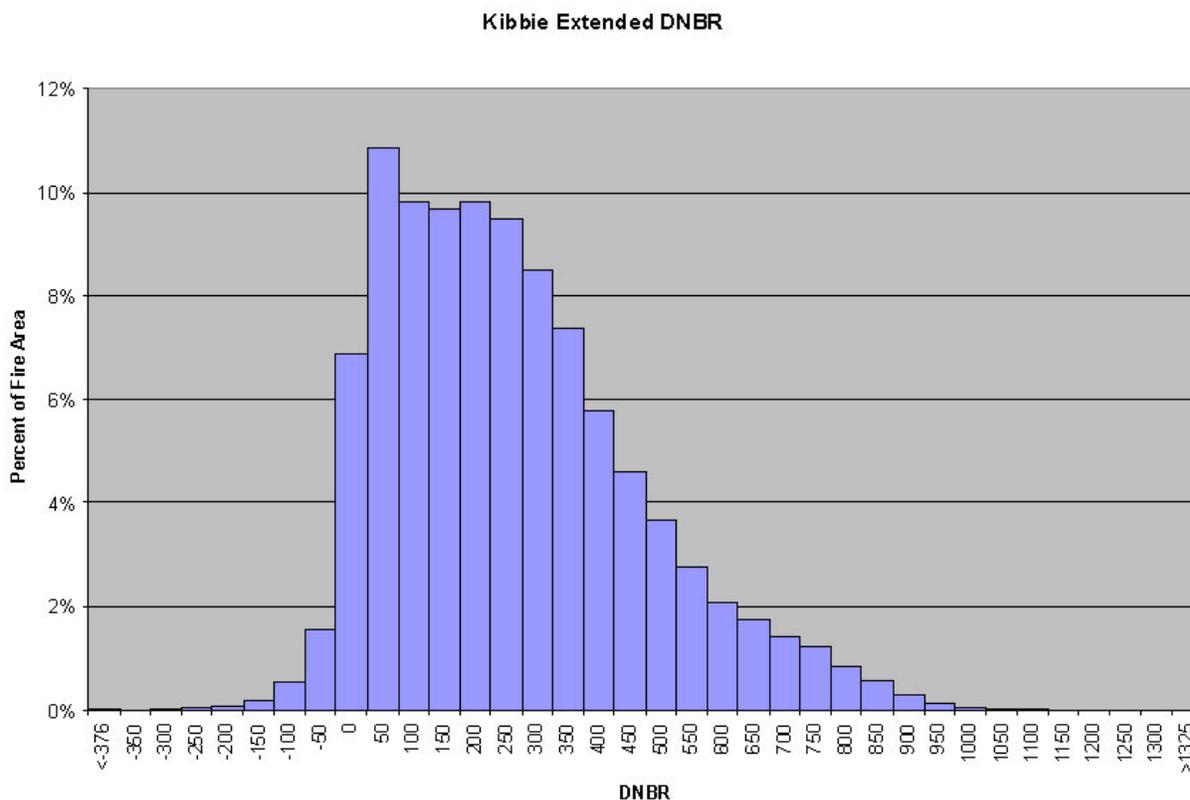
15. Interpret what this means as far as the number of high and low values in the data set.

16.

17. Interpret what this means as far as the number of high and low values in the data set.