|  |  |  |
| --- | --- | --- |
| **MTH 101 Lab: Standard Deviation & Statistics** | | **Section:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** |
|  | | |
| **Names**: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   |  | | --- | | **Focus Question:**   * What does standard deviation tell us about our data set? | | | |

**Introduction:** In this lab, you will be computing standard deviations for a few different data sets using Excel. Then, you will figure out how the differences in datasets affect the standard deviation. In other words, we will figure out the attributes of a data set that affect the standard deviation and and in what way those attributes affect the standard deviation.

**Before you begin the lab:** Download “Lab 11 Excel Worksheet” from mth101.com.

As a refresher, to use Excel to calculate measures of center and other measures for a data set:

* Use “**=average**(data)” to calculate the mean
* Use “=**mode**(data)” to calculate the mode
* Use “=**median**(data)” to calculate the median
* Use “=**stdev.s**(data)” to calculate the standard deviation of a data set
* Use “**max**(data)-**min**(data)” to calculate the range. Think about why this works!

|  |
| --- |
| **Definition**: The **median** of a data set is the number that separates the higher half from the lower half of ordered data. You can think of the median as the center of the data set.  For example, in the data set {8,7,3,5,9} the median is 7. |

|  |
| --- |
| **Definition:** The middle quartile is another name for the median. The **lower quartile** is the median of the bottom half of the data set (all numbers below the center). The **upper quartile** is the median of the top half of the data set (all numbers above the center). |

|  |
| --- |
| **Definition**: The **standard deviation** is a measure that is used to quantify the amount of variation or dispersion of a set of data values. |

|  |
| --- |
| **Definition:** The **variation** of a data set is a measure of how spread out the numbers are from the center (median). If the numbers are very far from the median, we say there is high variation. If the numbers are very close to the median, we say there is low variation. |

|  |
| --- |
| **Definition:** The **range** of a data set is the largest value minus the smallest value. |

**Example One:** We asked 15 college students to rank their happiness (from 1 to 10 with 1 being completely unhappy and 10 being extremely happy). The values they gave are:  
  
 3,4,4,4,5,5,5,5,5,5,5,6,6,6,7

That means that one student said they have a happiness of 3, three students said they have a happiness of 4, seven students said they have a happiness of 5, and so on.

**E1Q1:** What does this data set suggest about students’ happiness?

Find the sample average, standard deviation, range, and median of this data set using Excel. This data is found on the first sheet of your workbook labeled “Example 1.”

|  |  |  |
| --- | --- | --- |
|  | **Formula in Excel** | **Calculated Value**  **(round to hundredths if needed)** |
| **Mean** | = |  |
| **Standard Deviation** | = |  |
| **Range** | = |  |
| **Median** | = |  |

**Example Two:** We asked 15 more different college students to rank their happiness (from 1 to 10 with 1 being completely unhappy and 10 being extremely happy). The values they gave us are

3,3,3,3,3,4,4,5,6,6,7,7,7,7,7

**E2Q1:** What seems to be the difference in this group’s happiness in comparison to that of the group in Example 1?

Find the sample average and standard deviation of this data set using Excel. This data is found on the second sheet of your workbook labeled “Example 2.”

|  |  |  |
| --- | --- | --- |
|  | **Formula in Excel** | **Calculated Value**  **(round to hundredths if needed)** |
| **Mean** | = |  |
| **Standard Deviation** | = |  |
| **Range** | = |  |
| **Median** | = |  |

**Problem 1:** Answer the following questions.

1. In the space below, draw a frequency chart (histogram) for the data from Example One and Example Two. You should have two frequency charts; one for each example. You can use excel and cut and paste into this space or draw by hand.
2. Make a statement that relates the shape of your frequency charts to the standard deviations you found in Example One and Example Two. If you are having a hard time relating standard deviation to the graphs, try overlaying the graphs you drew.

**Problem 4:** The heights of American adult men are normally distributed with a mean of 68 inches and a standard deviation of 3 inches.

1. What percentage of the American adult male population is taller than 6’2’’?
2. What percentage of the American adult male population is shorter than 6’2”?
3. How can you use your answer to part (a) to get an answer to part (b)?
4. What percentage of the American adult male population is shorter than 5’5’’?
5. What percentage of the American adult male population is between 4’11” and 7’2”?
6. What percentage of the American adult male population is between 5’5” and 5’8” tall?
7. We can say that 68% of adult males are between what two heights?

|  |  |  |
| --- | --- | --- |
| **Participation** | **/5** | **TOTAL** |
| **Correctness** | **/5** | **/10** |