

Calculate

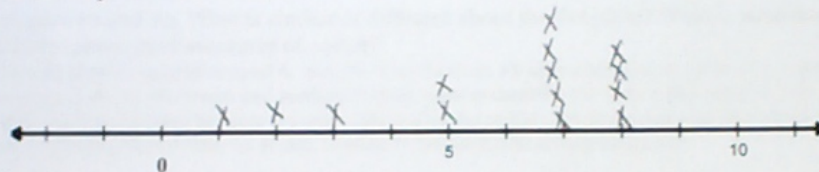
- Mean 6.4
- Median 6

Which measure of center would be the best when representing most of the data? Why?  
 Either measure of center seem to describe most of the data since they are both in the "middle" or the average of the data points.

2. Below is a survey of the number of hours slept each night by medical students during mid-terms.

Average Hours Slept Each Night								
Hours	1	2	3	4	5	6	7	8
Frequency	1	1	1	0	2	0	5	4

Make a dot plot of the data above.



Calculate

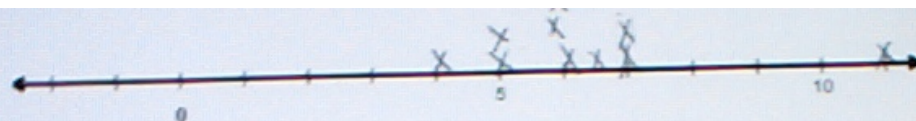
- Mean 5.9
- Median 7

Which measure of center would be the best when representing most of the data? Why?  
 The median seems to describe the data set better than the mean because the mean isn't close to "most" of the data. Most of the data points are 7's and 8's; so the median better describes "most" of the data than the mean.

3. Below is a survey of the number of hours slept each night of several junior high students.

Average Hours Slept Each Night										
Shane	Joelle	Rich	Sally	Abby	Reagan	Ali	Bobby	George	Jane	Sam
4	6	6.5	6	6	5	7	5	6	11	7

Make a dot plot of the data above.



- Calculate
- Mean 6.6
  - Median 6

Which measure of center would be the best when representing most of the data? Why?  
 The median still seems slightly better than the mean because there are eight values of 6 or less, and only four values greater than 6.

- Compare #1 and #2. What is similar or different about the dot plots? What is similar or different about your measures of center?  
 The data distribution on #2 is skewed left, but the distribution for the data set on #1 is mostly symmetric. On #1, both the mean and the median seemed to describe most the data well; but on #2 we would want the median because the median didn't seem to be affected by the few lower values of 1, 2, and 3 (causing the data to be skewed). Hmm... I wonder if this always happens??
- Compare #1 and #3. What is similar or different about the dot plots? What is similar or different about your measures of center?  
 Both are mostly centered around 6. But the data set from #3 seems to have an outlier or extreme value at 11. On #1, the mean and median both seemed to describe the data well, but on #3 we would want the median because it wasn't affected by the outlier like the mean was (the mean ended up being higher than 6). Hmm... again, I wonder if this always happens?

### Spread

- For a data set, the **variation** describes how widely the data varies or answers the question "how far apart are the numbers?" Two common measures of variation are IQR (interquartile range) and standard deviation.

Determine which measure of variation is most appropriate for the given data set. **Explain your answer.**

- Using the following data points, calculate the mean, standard deviation, median, and interquartile range.  
 48, 52, 55, 49, 53, 55, 51, 50, 54, 53, 54, 55, 50, 51, 60, 52, 57, 56, 58, 55
  - Mean and Standard Deviation  
 Mean = 53.4 and Standard Deviation = 3.10
  - Median and IQR  
 Median = 53.5 and IQR = 4
  - What do each of *i.* and *ii.* mean above?
    - Using 53.4 as the average data point; 68% of the data is within 3.10 (one standard deviation) from the mean.
    - Using 53.5 as the middle data point; 50% of the data is within an IQR of 4 centered about the median.



b. Create a dot plot from the data on the number line below:



Which measure of variation would be the best when describing the spread of the data?

**Why?**

Either standard deviation or IQR both seem to represent the spread of the data well. Both seem to encompass almost the same amount of data.

7. Below is a slightly different list of data points than #6. Using the data points, calculate the mean, standard deviation, median, and interquartile range.

48, 52, 55, 49, 53, 55, 51, 50, 54, 53, 54, 55, 50, 51, 60, 52, 57, 56, 58, 55, 105

i. Mean and Standard Deviation

Mean = 55.85 and Standard Deviation = 11.66

ii. Median and IQR

Median = 54 and IQR = 4.25

a. What do each of i. and ii. mean above?

b. Using 55.85 as the average data point; 68% of the data is within 11.66 (one standard deviation) from the mean

c. Using 54 as the middle data point; 50% of the data is within an IQR of 4.25 centered about the median.

d. Create a dot plot from the data on the number line below:



Which measure of variation would be the best when describing the spread of the data?

**Why?**

The IQR and the median best describe the data because the mean and standard deviation are too high with the data point 105 (outlier) added to the data set. The mean and standard deviation do not actually fit the spread of the data very well. However, the IQR was not affected much by the new data point, and thus, still represents the spread of most of the data.

8. Compare #6 and #7. What do you notice about their dot plots? What is the same? Different? What do you notice about your choice of measure for the variation? What do you think might cause the difference?  
#7 has an outlier while #6 does not. Otherwise, the rest of the data is clumped together the same between the two. When the data was clumped and somewhat symmetric, both measures of variation seemed to be fine. However, once there was an outlier, the median and IQR better represented the spread of the data.

### Modifying Data

9. Suppose you take the data from #1 above:
- What if you wanted to raise the mean? What data point could you add? Why would that point work? Calculate the new mean with your new data point.  
Possible new data point: 10 (the original mean = 6.4, so adding a value higher will raise the mean). The new mean = 6.8
  - What if you wanted to raise the standard deviation of the original data set? What data point could you add? Explain your reasoning. Calculate the new standard deviation with the new data point.  
Possible new data point: 2 (The original standard deviation was 1.87 from the mean = 6.4. Since 2 is a distance of 4 away from the mean, it will increase the average distance from the data points to the mean, or the variation). The new standard deviation = 2.26
  - What if you wanted to raise the mean of the original data set, but keep the median the same by adding two data points, what could two possible values be? How do you know?  
Possible new data points: 11 and 5 (If we add a point on each side of the median = 6, it will stay in the middle. But since we want the mean to become larger, we want to add large values on each side of the median).
  - Suppose we added one student to the data who only sleeps 1 hour per night. What would you expect to happen to the original mean and standard deviation?  
The mean would become lower since 1 is a value below the mean = 6.4, so it will bring the average down. The standard deviation would become larger because the current standard deviation = 1.87 and 1 is 5.4 units away from the mean, which will thus increase the average distance from the data points to the mean, or variation.