Measures of Central Tendency & Variation

Finding the Measures of Central Tendency Mean = average median = middle when in order Mode = most

Example 1: Find the mean median, and mode for the data. Deer at a feeder each hour: ~~3~~, ~~0~~, ~~2~~, ~~0~~, ~~1~~, ~~2~~, 4

$\bar{x}$ mean = $\frac{3+0+2+0+1+2+4}{7}$ = 1.7 deer

median: 0, 0, 1, 2, 2, 3, 4 = 2 deer

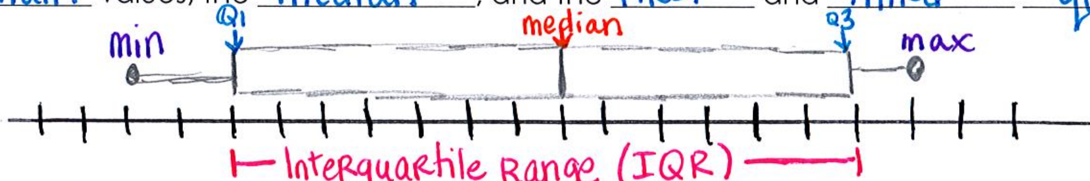
most: 0 & 2

Now you Try: Example 2:
Find the mean, median, and mode of the data set. $\{6,9,3,8\}$

mean = 6.5
median = 7
mode = none

on calculator
data L1: enter values
2nd data 1: 1-Vars Stat
$\bar{x}$ = mean
med = median

A box-and-whisker plot shows the spread of data set. It displays the 5 key points; the minimum and maximum values, the median, and the first and third quartiles.



The quartiles are the medians of the lower and upper halves of the data set. If there are an odd number of data, do not include the median in either half.

The interquartile range, or IQR, is the difference between the 1st and 3rd quartiles, or Q3-Q1. It represents the middle 50% of the data.

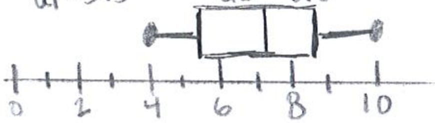
Example 3:
Make a box-and-whisker plot of the data. Find the interquartile range.
$\{6,8,7,5,10,6,9,8,4\}$

4 5, 6 6 | 7 | 8 8, 9 10

min med max

Q1 = 5.5 Q3 = 8.5

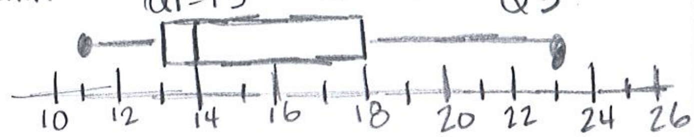
IQR = Q3 − Q1
8.5 − 5.5
= 3



Example 4: IQR = 18 − 13 = 5
Make a box-and-whisker plot of the data. Find the interquartile range.
$\{13,14,18,13,12,17,15,12,13,19,11,14,14,18,22,23\}$

11 12 12 13 13 13 14 14 | 14 15 17 18 18 19 22 23

min Q1 = 13 med = 14 Q3 max



The data sets $\{19,20,21\}$ and $\{0,20,40\}$ have the same mean and median, but the sets are very different. The way the data are spread out from the mean or median is important to the study of statistics.

A measure of variation is a value that describes the spread of a data set. The most commonly used measures of variation are the range, the interquartile range, the variance, and the standard deviation.

Symbols commonly used to represent **mean** is $\bar{x}$ or x bar. The symbol for **standard deviation** is the lowercase Greek letter Sigma, σ. The **variance**, denoted by σ^2, is the average of the squared differences from the mean. **Standard deviation**, denoted by σ, is the square root of the variance and is one of the most common and useful measures of variation.

Low standard deviation indicate data that are clustered near the measures of central tendency, whereas high standard deviations indicate data that are spread out form the center.

Finding Variance and Standard Deviation
Step 1: Find the mean of the data $\bar{X}$.
Step 2: Find the difference between the mean and each data value and square it.
Step 3: Find the variance, σ^2, by adding the squares of all of the differences from the mean and dividing by the number of data values.
Step 4: Find the standard deviation, σ , by taking the square root of the variance.

Standard Deviation

Example 5: Find the mean and standard deviation for the data set of the number of people getting on and off a bus for several stops. $\{6,8,7,5,10,6,9,8,4\}$

step 1: $\bar{X} = 7$

step 2: $(6-7)^2, (8-7)^2, (7-7)^2, (5-7)^2, (10-7)^2, (6-7)^2, (9-7)^2, (8-7)^2, (4-7)^2$

$1, 1, 0, 4, 9, 1, 4, 1, 9$

step 3: $\frac{1+1+0+4+9+1+4+1+9}{9} = 3.\bar{3} = \sigma^2$

step 4: $\sqrt{\sigma^2} = \sqrt{3.\bar{3}}$ $\sigma = 1.83$ people

Example 6: Find the mean and standard deviation for the data set of the number of elevator stops for several rides. $\{0,3,1,1,0,5,1,0,3,0\}$

step 1: $\bar{X} = 1.4$

step 2: $(0-1.4)^2, (3-1.4)^2, (1-1.4)^2, (1-1.4)^2, (0-1.4)^2, (5-1.4)^2, (1-1.4)^2, (0-1.4)^2, (3-1.4)^2, (0-1.4)^2$

step 3: $\frac{(-1.4)^2+(1.6)^2+(-.4)^2+(-.4)^2+(-1.4)^2+(3.6)^2+(-.4)^2+(-1.4)^2+(1.6)^2+(-1.4)^2}{10}$

$\sigma^2 = 2.64$

step 4: $\sqrt{\sigma^2} = \sqrt{2.64}$ $\sigma = 1.6$ stops

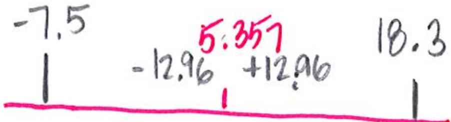
An **outlier** is an extreme value that is much less than or much greater than the other data values. Outliers have a strong effect on the mean and standard deviation. If an outlier is the result of measurement error or represents data from the wrong population it is usually removed.

There are different ways to determine whether a value is an outlier. One is to look for data values that are more than 3 standard deviations from the mean.

Example 7: in the 2003-2004 American League Championship Series, the New Your Yankees scored the following numbers of runs against the Boston Red Sox: 2, 6, 4, 2, 4, 6, 6, 10, 3, 19, 4, 4, 2, 3. Identify the outlier, and describe how if affects the mean and standard deviation.

In L_1 Plug in all 14 data points

$\bar{X} = 5.357$ (mean) $\sigma = 4.32$ (standard deviation)



$\sigma(3)$

$(4.32)(3)$

12.96

19 is yes an outlier

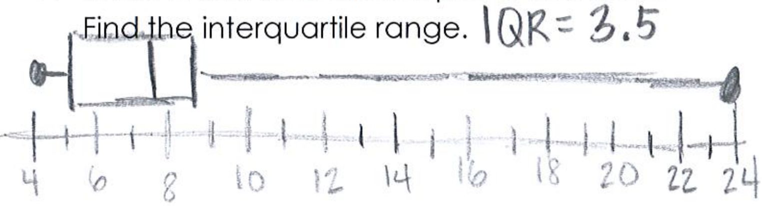
In L_1 Plug all 13 data points (not the outlier)

$\bar{X} = 4.31$ $\sigma = 2.2$

Try It: Use the data set for 1-3

$\{9,4,7,8,5,8,24,5\}$

1. Make a box-and-whisker plot of the data. Find the interquartile range. IQR = 3.5



2. Find the variance and the standard deviation of the data set.

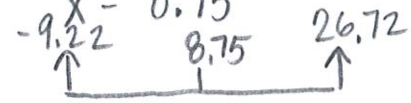
$\sigma = 5.99$ (standard deviation)

$\sigma^2 = 35.94$ (variance)

3. Are there any outliers in the date set?

$3\sigma = 3(5.99) = 17.97$ No!!

$\bar{X} = 8.75$



All data values are between (-9.22, 26.72)