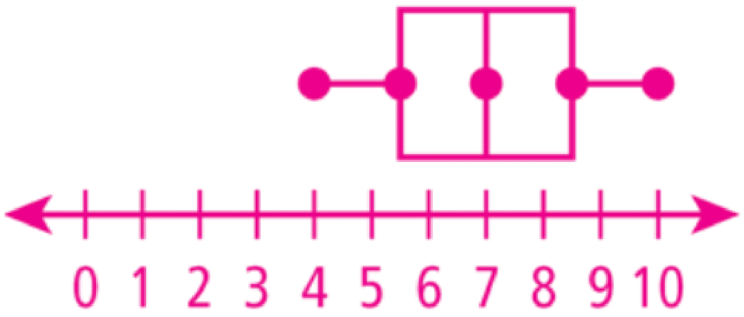
Warmup

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, $\boxed{7}$, 8, 8, 9, 10

min Q1 Med Q3 max



symmetrical.

IQR: $Q_3 - Q_1$

$= 8.5 - 5.5$

$= 3$

Range: Max − Min = 10 − 4 = 6

Calculate the mean and the median for each of the following data sets below:

Set A

19, 20, 21

$$\frac{19+20+21}{3}$$

Mean = 20

median = 20

Set B

0, 20, 40

mean = 20

median = 20

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

(spread)

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*.

Range = max − min

IQR = $Q_3 - Q_1$

Variance

↕

★ standard deviation.

M. A. D.

mean absolute deviation

Reading Math

The symbol commonly used to represent the mean is $\bar{x}$, or "*x* bar." The symbol for standard deviation is the lowercase Greek letter *sigma*, σ.

$\bar{x}$ = x - bar = mean

* σ = sigma = standard deviation

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.

$\sqrt{x^2} = x$

$\boxed{\text{variance} = \sigma^2}$

standard deviation $= \sqrt{\text{variance}} = \sqrt{\sigma^2} = \sigma$

Low standard deviations indicate data that are clustered near the measures of central tendency, whereas high standard deviations indicate data that are spread out from 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. $(x-\bar{x})^2$
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. avg. of all step 2 values.
Step 4.	Find the standard deviation, σ, by taking the square root of the variance. sq. root step 3 value.

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 Find the mean.

$$\bar{x} = \frac{63}{9} = 7$$

Step 2 Find the difference between the mean and each data value, and square it.

$$\bar{x} = 7$$

Data value x	6	8	7	5	10	6	9	8	4
$x - \bar{x}$	-1	1	0	-2	3	-1	2	1	-3
☆ $(x - \bar{x})^2$	1	1	0	4	9	1	4	1	9

Variance (find avg. of $(x-\bar{x})^2$ row) =

Find the variance.

$$\sigma^2 = \frac{1+1+0+4+9+1+4+1+9}{9} \approx 3.3$$

Find the standard deviation.

$$\sigma = \sqrt{3.3} \approx 1.83$$

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}

$\bar{x} = 1.4$

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

$(0 - 1.4)^2 = 1.96$
1.96
1.96
1.96

$(3 - 1.4)^2 = 2.56$
2.56

$(5 - 1.4)^2 = 12.96$

$(1 - 1.4)^2 = .16$
.16
.16

Variance = avg. of those values

= 2.64

σ = st. dev = $\sqrt{2.64}$ = 1.62

Using the Calculator:

Hours Slept

6.5	8	6.25	7.25	6.25	7.25
8.5	7	6.75	6.25		

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 of Outlier

In the 2003-2004 American League Championship Series, the New York 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 it affects the mean and standard deviation.

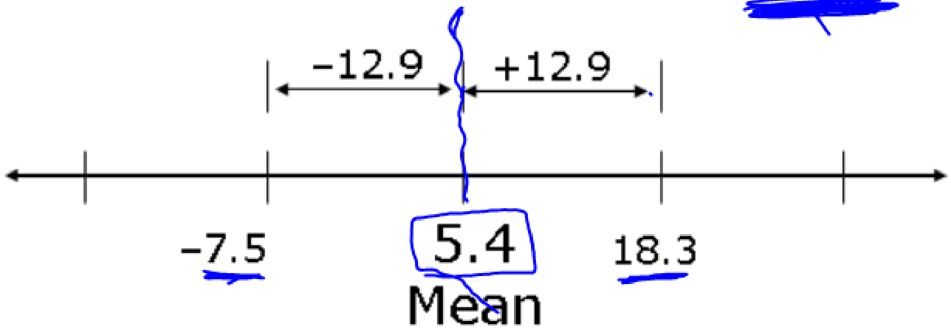
Step 1: Enter the data values into list L1 on a graphing calculator.

Step 2: Find the mean and standard deviation.

$\bar{x} = 5.4$ $\sigma = 4.32$

Step 3: Identify the outliers. Look for the data values that are more than 3 standard deviations away from the mean in either direction.

Three standard deviations is about 3(4.3) = 12.9.



Values less than -7.5 and greater than 18.3 are outliers, so 19 is an outlier.

Step 4: Remove the outlier to see the effect that it has on the mean and standard deviation.

w/outlier

mean: 5.4

σ: 4.3

w/out outlier

mean: 4.3

σ: 2.16}20 . . .

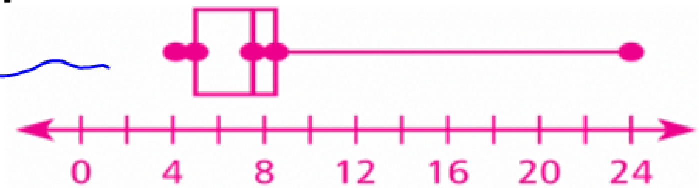
2.16

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 ran~~



IQR: 3.5

$\bar{x} = 8.75$

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

var: ≈ 35.94; std. dev: ≈ 5.99

$\sigma = 5.99$

$var = 35.9$

3. Are there any outliers in the data set?

No! Outliers less than -9.22, greater than 26.72

$\bar{x} + 3\sigma = \quad 8.75 + 3(5.99) = 26.72$

$\bar{x} - 3\sigma = \quad 8.75 - 3(5.99) = -9.22$

Practice with Statistics