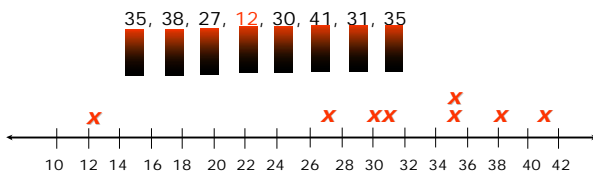# Effects of Outliers

- The **mean** is a good measure to use to describe data that are close in value.
- The **median** more accurately describes data with an outlier.
- The **mode** is a good measure to use when you have categorical data; for example, if each student records his or her favorite color, the color (a category) listed most often is the mode of the data.

In the data set below, the value 12 is much less than the other values in the set. An extreme value such as this is called an <u>outlier</u>.

35, 38, 27, 12, 30, 41, 31, 35

Additional Example 2: Exploring the Effects of Outliers on Measures of Central Tendency

The data shows Sara's scores for the last 5 math tests: 88, 90, 55, 94, and 89. Identify the outlier in the data set. Then determine how the outlier affects the mean, median, and mode of the data.

55, 88, 89, 90, 94

outlier ▮ 55

1

Additional Example 2 Continued
With the Outlier

55, 88, 89, 90, 94

outlier ⬛ 55

mean:	median:	mode:
55+88+89+90+94 = 416 416 ÷ 5 = 83.2 The mean is 83.2.	55, 88, (89) 90, 94 The median is 89.	 There is no mode.

Additional Example 2 Continued
Without the Outlier

~~55~~, 88, 89, 90, 94

mean:	median:	mode:
88+89+90+94 = 361 361 ÷ 4 = 90.25 The mean is 90.25.	88, (89+90), 94 ‾‾2‾‾ = 89.5 The median is 89.5.	 There is no mode.

Additional Example 2 Continued

	Without the Outlier	With the Outlier
mean	90.25	83.2
median	89.5	89
mode	no mode	no mode

The outlier _decreased_ the _mean_ by _7.05_ .

The outlier _decreased_ the _median_ by _0.5_.

The mode did not change/ There is no mode.

Effects of Outliers….

- <u>Mean</u>: Significant change
  - Mean increases with high outlier
  - Mean decreases with low outlier

- <u>Median</u>: Little change

- <u>Mode</u>: No change

- <u>Range</u>: Spread of data increases

Try This: Example 2

Identify the outlier in the data set. Then determine how the outlier affects the mean, median, and mode of the data.

63, 58, 57, 61, 42

42, 57, 58, 61, 63

outlier ▮ 42

Try This: Example 2 Continued
With the Outlier

42, 57, 58, 61, 63

outlier ▮ 42

mean:	median:	mode:
42+57+58+61+63 = 281	42, 57, (58) 61, 63	
281 ÷ 5 = 56.2		
The mean is 56.2.	The median is 58.	There is no mode.

Try This: Example 2 Continued
Without the Outlier

~~42~~, 57, 58, 61, 63

mean:	median:	mode:
57+58+61+63 = 239	57, (58+61), 63	
239 ÷ 4 = 59.75	$\frac{}{2}$	
The mean is 59.75.	= 59.5	
	The median is 59.5.	There is no mode.

Try This: Example 2 Continued

	Without the Outlier	With the Outlier
mean	59.75	56.2
median	59.5	58
mode	no mode	no mode

The outlier <u>decreased</u> the <u>mean</u> by <u>3.55</u>.
The outlier <u>decreased</u> the <u>median</u> by <u>1.5</u>.

The outlier decreased the mean by 3.55 and the median by 1.5

The mode did not change./ There is no mode.

Best Measure of Central Tendency

You should use the….

- **MODE**- If there is an outlier AND about ½ or more of the data is the same
  - 3, 4, 4, 4, 4, 17
    - » NO MEAN!

- **MEDIAN-** If there is an outlier
  - 2, 7, 8, 12, 12, 17, 31

- **MEAN**- No outliers