

Sunday, October 14, 2018
5:13 PM

KEY

Boxplots Day 2

People say that it takes a long time to get to work from New York State due to the heaving traffic near big cities. What does the data say? Here are the travel times in minutes of 23 randomly chosen New York workers:

10	5	15	15	30	45	40	15	10	20	20	25
15	25	25	25	30	40	65	20	85	60	60	

The **mean of a sample**, \bar{x} :

$$\bar{x} = \frac{\sum x_i}{n}$$

\sum

“sigma” means to add all of the data.

x_i

each different piece of data; x_1, x_2, x_3, \dots

\bar{x}

“x-bar” is the sample mean of the data

n

sample size (how many data pieces there are)

From last night’s HW...

What did one *outstanding* time (of 85 minutes) increase the **mean** (\bar{x}) time count by? **2.4803**

This is because the **mean is sensitive** to the influence of a few extreme outliers.

It is **NOT A RESISTANT MEASURE OF CENTER**.

From last night’s HW...

What did one *outstanding* time (of 85 minutes) do to the **median**? **nothing!**

One outstanding value **DOES NOT** affect the median.

The median **IS RESISTANT** to outliers.

**** Summary of measures of center:**

The **mean and median** are measures of **center**. Therefore, when choosing the best measure of center to describe your data, we would choose the **MEAN** if we have symmetric data with no outliers and we would choose the **MEDIAN** if we have skewed data or data with outliers.

Box and Whisker Plots

Use the 5 number summary: minimum, Q1, median, Q3, and maximum to create boxplots.

- We will be using the New Yorkers work travel time data from last night's HW (also on page 1) to construct a boxplot.

First write your numbers in order below:

5 10 10 15 15 15 15 20 20 20 25 25
 25 25 30 30 40 40 45 60 60 65 85

The five numbers in a *five number summary* are:

- Minimum: **smallest # in data set = 5**
- 1st Quartile: The median of all data **to the left of the median = 15**
- Median: The middle number of all data (if there are two numbers in the middle, average them!) **= 25**
- 3rd Quartile: The median of all data **to the right of the median = 40**
- Maximum: **largest # in data set = 85**

Now, we will mathematically check for outliers. From now on, we will always use the formulas to check for outliers, so we will know for sure if it is one. *Therefore, you no longer have to say "appears to be"*.

IMPORTANT: All formulas are on your formula sheet! You do not have to memorize!

We start by finding the **IQR (interquartile range)**. The **IQR** is another measure of **spread**. It tells us the distance from **the third and first quartile** (or the middle half of the data).

$$IQR = Q_3 - Q_1 \quad IQR = 40 - 15 = 25$$

Then, we use the following 2 formulas to find the outliers:

Are there any data values < -22.5?

$$< Q_1 - 1.5(IQR) \\ < 15 - 1.5(25) \\ < -22.5$$

$$> Q_3 + 1.5(IQR) \\ > 40 + 1.5(25) \\ > 77.5$$

85 is an outlier

Are there values > 77.5?

To draw a boxplot, we draw a modified boxplot that show **any outliers plotted individually**. Therefore, on your graph, you plot your 5 number summary (5 points) and **you will have additional points if there are outliers**. Don't forget to always write a title, and label the number line with an appropriate scale.



Partner Practice

Find the 5 number summary, check for outliers, and draw a modified boxplot.

2. Here are the number of homeruns Babe Ruth hit in each of his 15 years with the New York Yankees:
 (notice the data is not given in numerical order)

54 59 35 41 46 25 47 60 54 46 49 46 41 34 22

#'s in order: 22 25 34 35 41 41 46 46 46 47 49 54 54 59 60

5 # summary: min: 22 Q1: 35 med: 46 Q3: 54 max: 60

Outliers (show all formulas/work):

$$IQR = Q3 - Q1 \left\{ \begin{array}{l} < Q_1 - 1.5 IQR \\ < 35 - 1.5(19) \\ < 6.5 \end{array} \right\} \left\{ \begin{array}{l} > Q_3 + 1.5 IQR \\ > 54 + 1.5(19) \\ > 82.5 \end{array} \right.$$

NO outliers

Boxplot:

Home Runs Hit By Babe Ruth in 15 years with NY Yankees



- 3.) Roger Maris had these homerun totals in 10 years in the American League:

13 23 26 16 33 61 28 39 14 8

#'s in order: 8 13 14 16 23 26 28 33 39 61

5 # summary: min: 8 Q1: 14 med: 24.5 Q3: 33 max: 61
 (23+26)÷2

Outliers (show all formulas/work):

$$IQR = Q3 - Q1 \left\{ \begin{array}{l} < Q_1 - 1.5 IQR \\ < 14 - 1.5(19) \\ < -14.5 \end{array} \right\} \left\{ \begin{array}{l} > Q_3 + 1.5 IQR \\ > 33 + 1.5(19) \\ > 61.5 \end{array} \right.$$

NO outliers

Boxplot:

Home Runs Hit By Roger Maris in 10 years with American League

