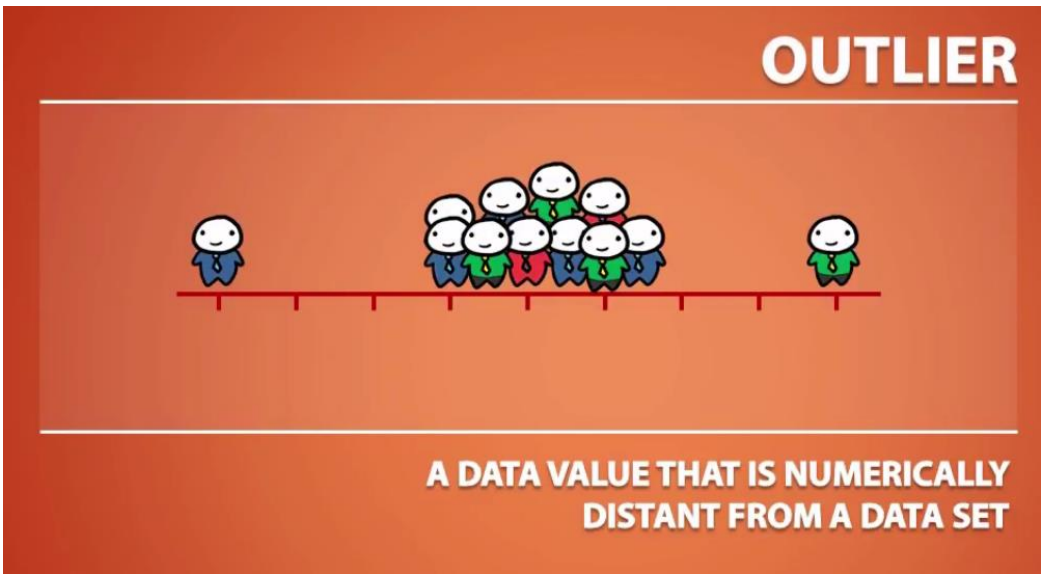
3. Maths - Statistics – PART – 3

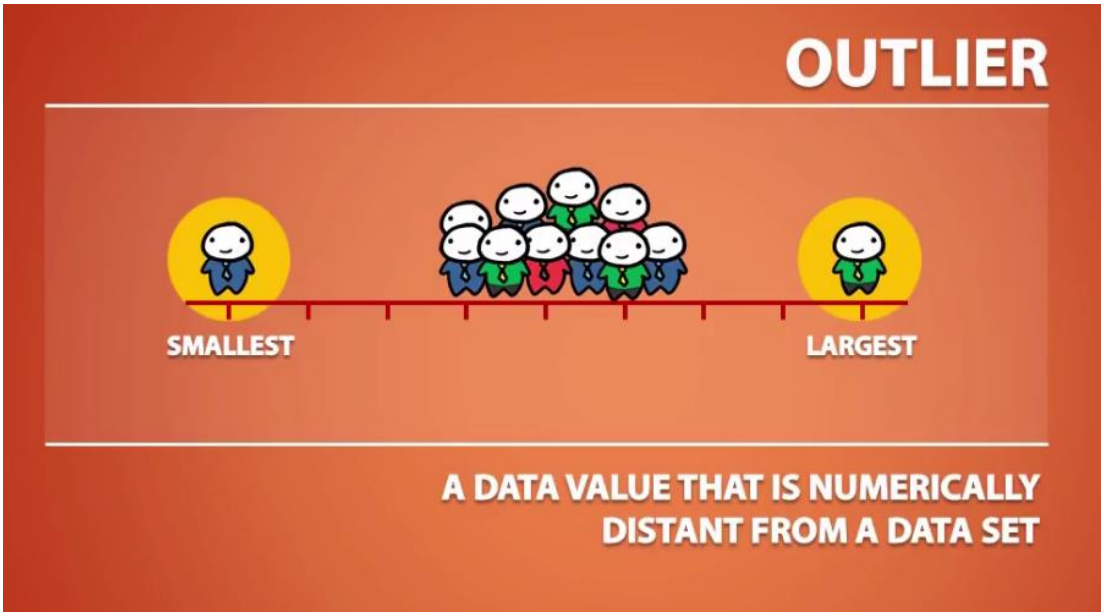
Contents

3. Maths - Statistics – PART – 3

1. What is an outlier?

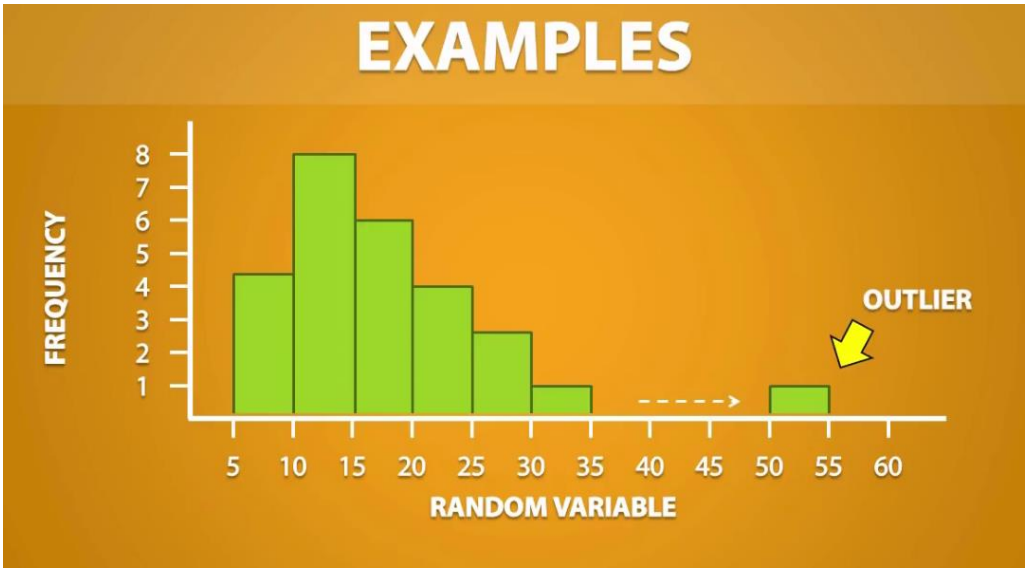
- ✓ An outlier is defined as a value which are very far from dataset
- ✓ An outlier is a data point that falls outside from main data points
- ✓ It can be largest value in dataset, smallest value in dataset





Examples

- ✓ Observe the below histogram, a point is far from value

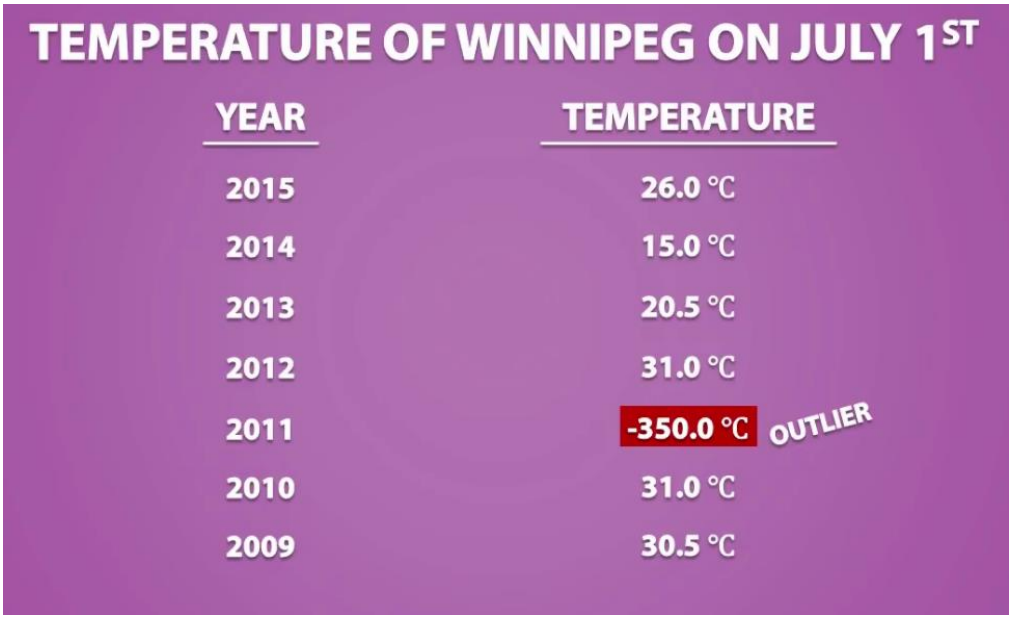


✓ Below example, 9000 is very larger value than other values

✓ Below example, 3 is very smaller value than other values

2. Surprising...!!!

- ✓ Outliers are data points but these are typical and surprising.
- ✓ These effects the measures of center and spread
- ✓ Observe the below example


TEMPERATURE OF WINNIPEG ON JULY 1ST
YEAR
TEMPERATURE
2015
26.0 ℃
2014
15.0 ℃
2013
20.5 ℃
2012
31.0 ℃
2011
-350.0 ℃
OUTLIER
2010
31.0 ℃
2009
30.5 ℃

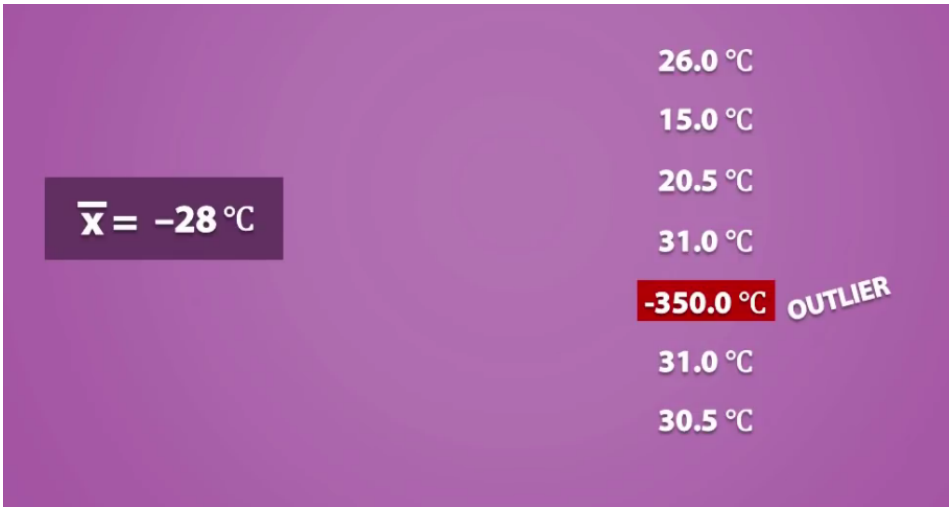
THE MEAN IS AFFECTED BY THE PRESENCE OF OUTLIERS

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

26.0 ℃
15.0 ℃
20.5 ℃
31.0 ℃
-350.0 ℃
31.0 ℃
30.5 ℃

$$\bar{x} = \frac{26 + 15 + 20.5 + 31 + (-350) + 31 + 30.5}{7}$$

26.0 ℃
15.0 ℃
20.5 ℃
31.0 ℃
-350.0 ℃
31.0 ℃
30.5 ℃

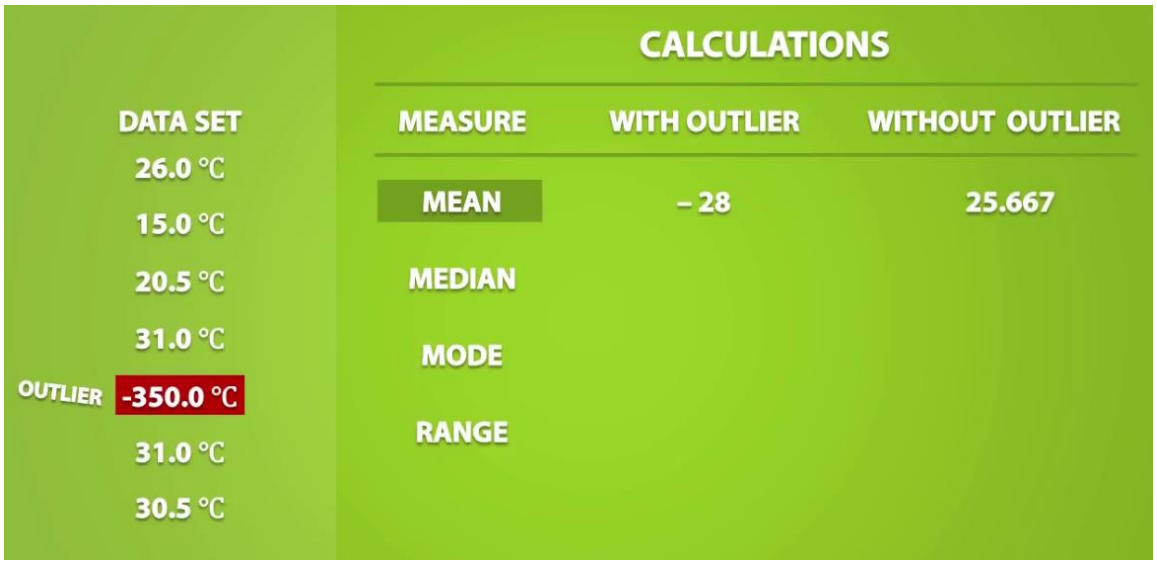

$\bar{x} = -28$ ℃
26.0 ℃
15.0 ℃
20.5 ℃
31.0 ℃
-350.0 ℃
OUTLIER
31.0 ℃
30.5 ℃

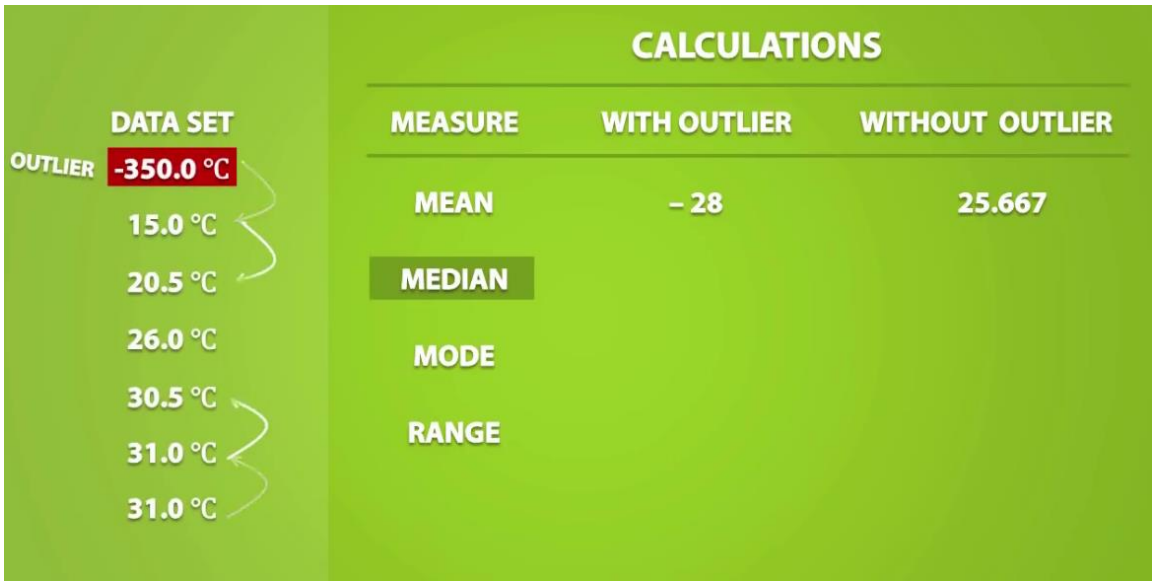
So

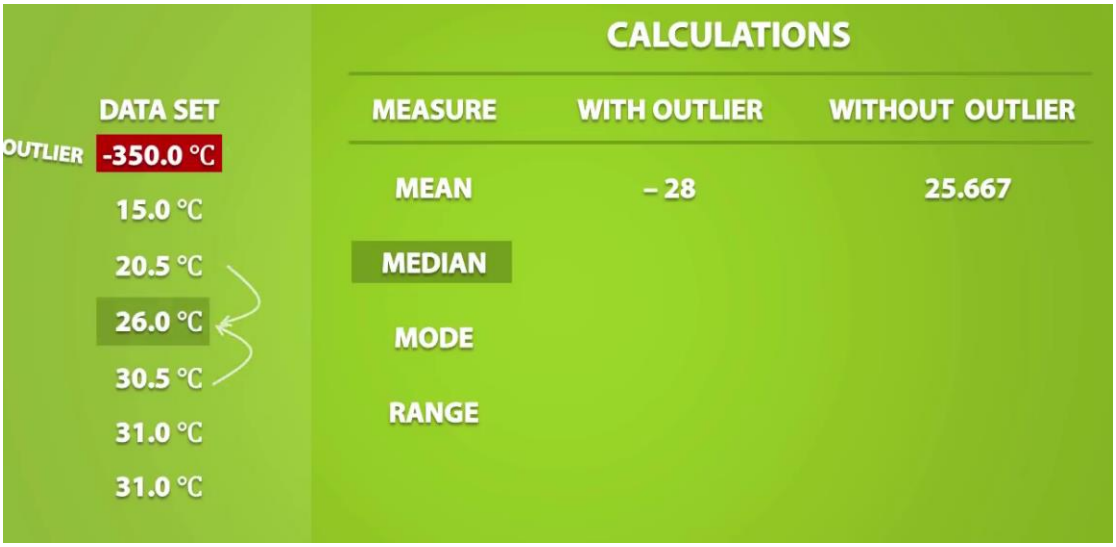

THE MEAN IS AFFECTED BY THE PRESENCE OF OUTLIERS

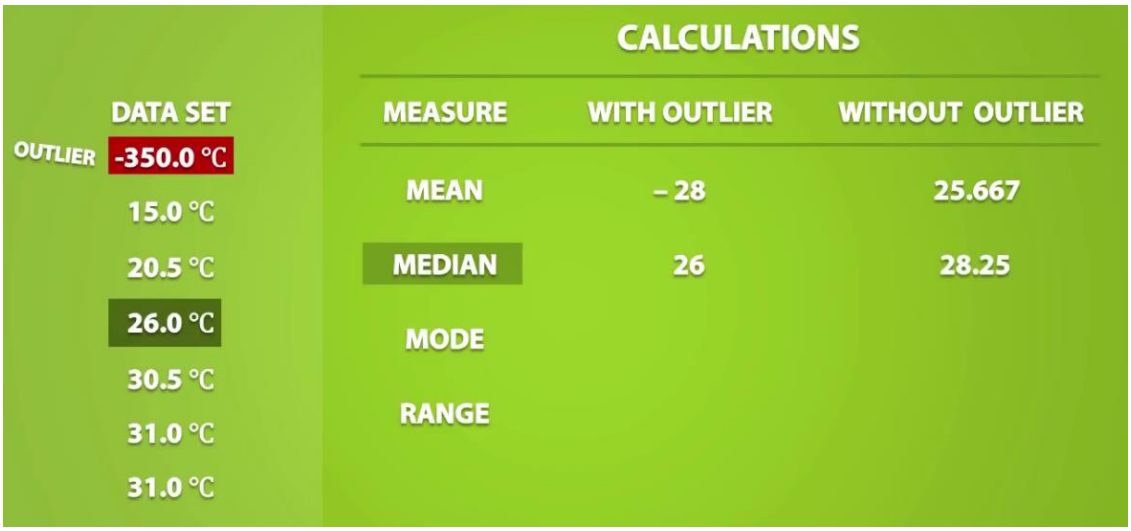
3. Checking mean, median, mode & Range

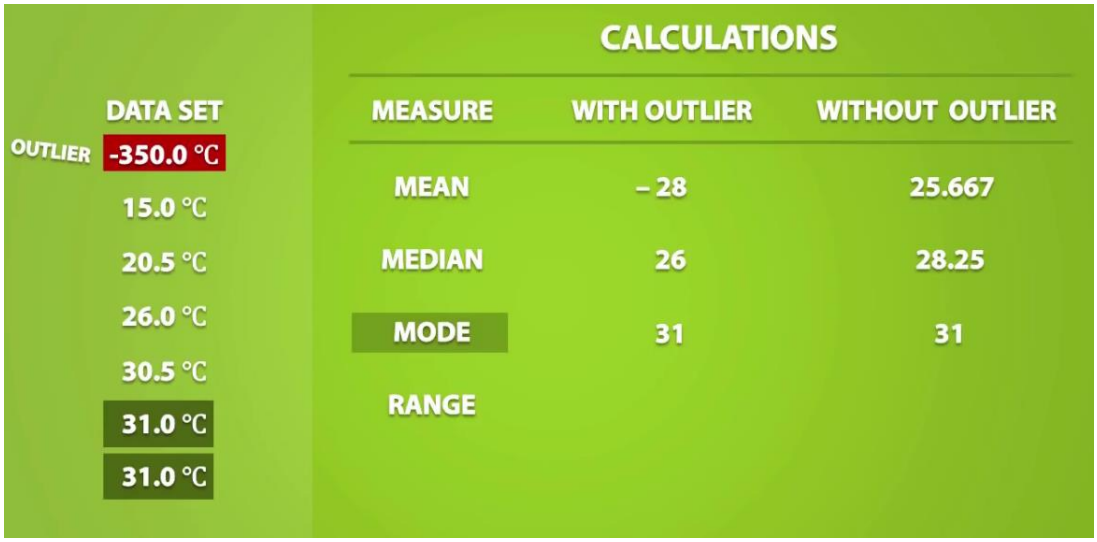
✓ Calculate mean, median, mode and Range for a dataset

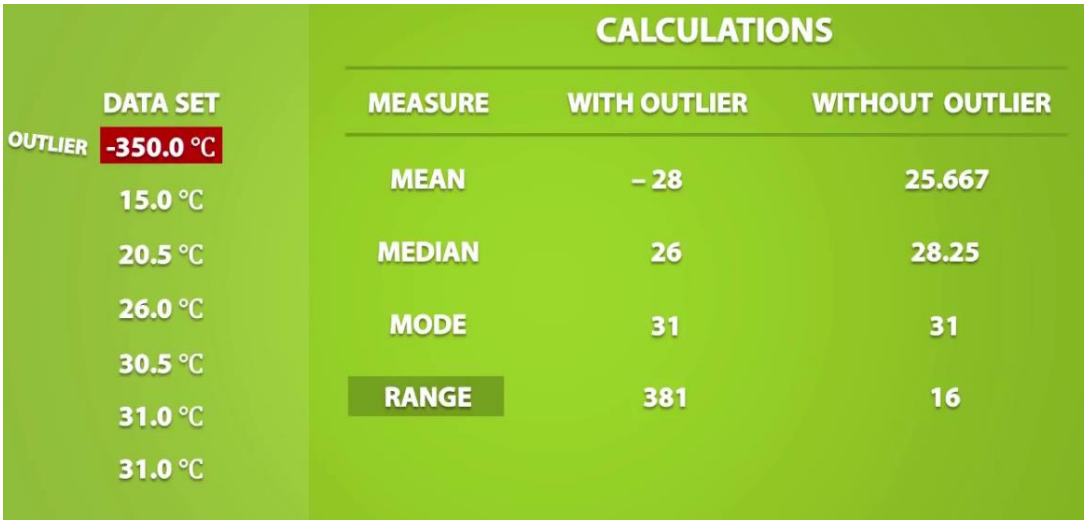





DATA SET
OUTLIER
-350.0 ℃
15.0 ℃
20.5 ℃
26.0 ℃
30.5 ℃
31.0 ℃
31.0 ℃
CALCULATIONS
MEASURE
WITH OUTLIER
WITHOUT OUTLIER
MEAN
– 28
25.667
MEDIAN
MODE
RANGE


DATA SET
OUTLIER
-350.0 ℃
15.0 ℃
20.5 ℃
26.0 ℃
30.5 ℃
31.0 ℃
31.0 ℃
CALCULATIONS
MEASURE
WITH OUTLIER
WITHOUT OUTLIER
MEAN
– 28
25.667
MEDIAN
26
28.25
MODE
RANGE

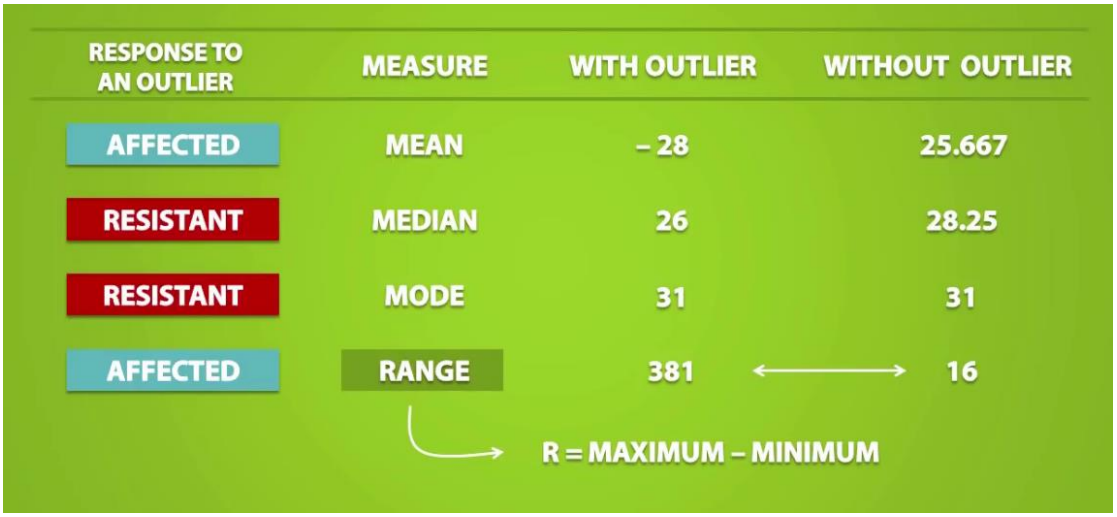

DATA SET
OUTLIER
-350.0 °C
15.0 °C
20.5 °C
26.0 °C
30.5 °C
31.0 °C
31.0 °C
CALCULATIONS
MEASURE
WITH OUTLIER
WITHOUT OUTLIER
MEAN
– 28
25.667
MEDIAN
26
28.25
MODE
31
31
RANGE

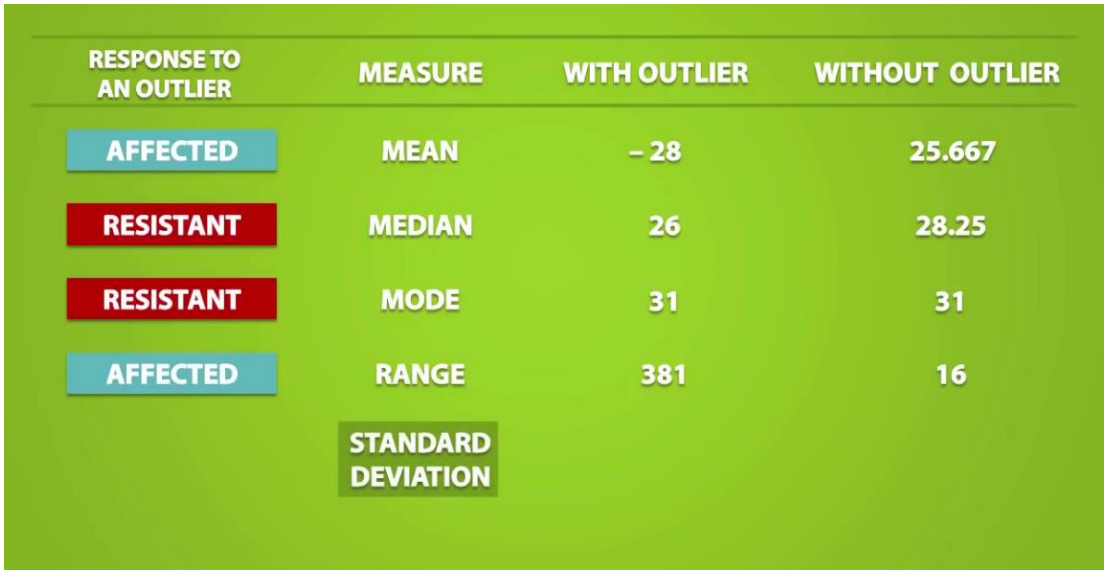

DATA SET
OUTLIER
-350.0 °C
15.0 °C
20.5 °C
26.0 °C
30.5 °C
31.0 °C
31.0 °C
CALCULATIONS
MEASURE
WITH OUTLIER
WITHOUT OUTLIER
MEAN
– 28
25.667
MEDIAN
26
28.25
MODE
31
31
RANGE
381
16

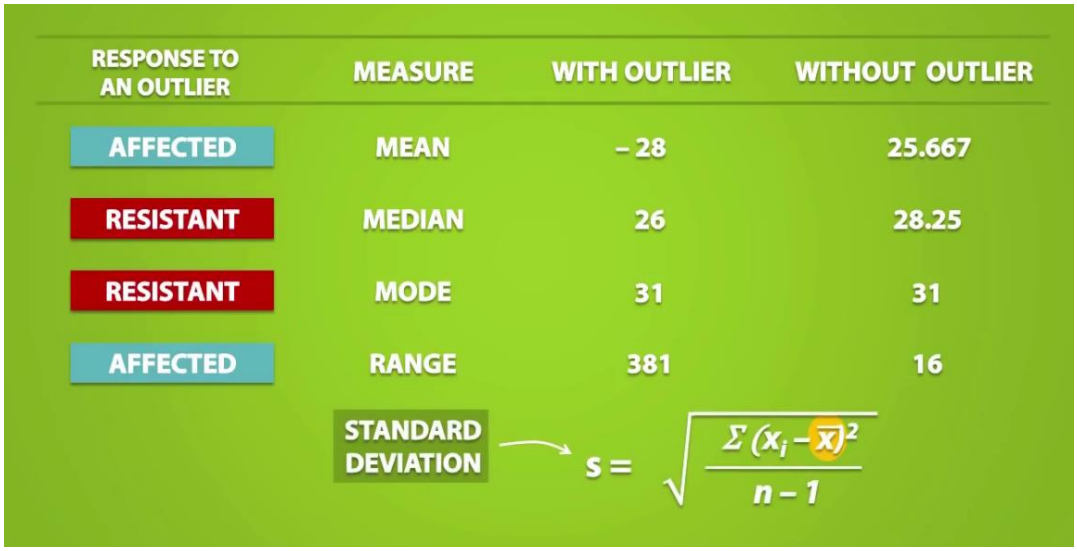
✓ Is outlier affects the calculations: Yes then observe the below table

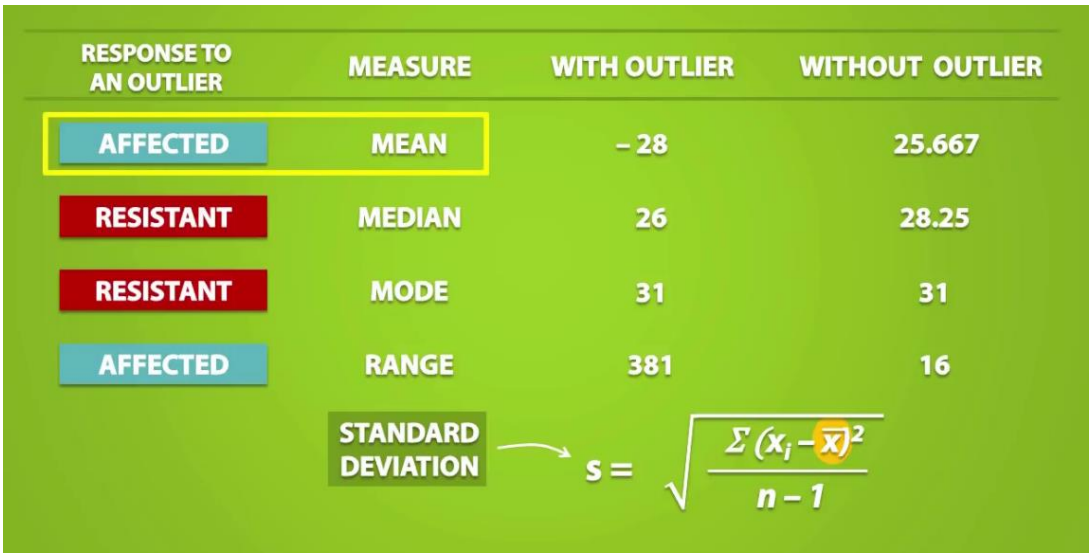
RESPONSE TO AN OUTLIER	MEASURE	WITH OUTLIER	WITHOUT OUTLIER
AFFECTED	MEAN	– 28	25.667
	MEDIAN	26	28.25
	MODE	31	31
	RANGE	381	16

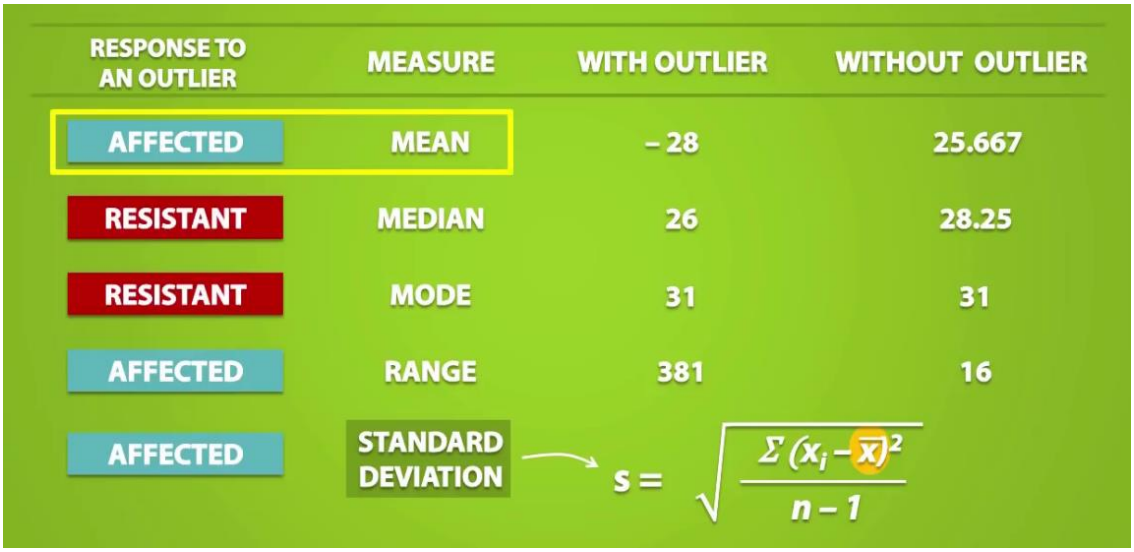
RESPONSE TO AN OUTLIER	MEASURE	WITH OUTLIER	WITHOUT OUTLIER
AFFECTED	MEAN	– 28	25.667
RESISTANT	MEDIAN	26	28.25
RESISTANT	MODE	31	31
	RANGE	381	16


RESPONSE TO AN OUTLIER
MEASURE
WITH OUTLIER
WITHOUT OUTLIER
AFFECTED
MEAN
– 28
25.667
RESISTANT
MEDIAN
26
28.25
RESISTANT
MODE
31
31
AFFECTED
RANGE
381
16
R = MAXIMUM – MINIMUM


RESPONSE TO AN OUTLIER
MEASURE
WITH OUTLIER
WITHOUT OUTLIER
AFFECTED
MEAN
– 28
25.667
RESISTANT
MEDIAN
26
28.25
RESISTANT
MODE
31
31
AFFECTED
RANGE
381
16
STANDARD DEVIATION


RESPONSE TO AN OUTLIER
MEASURE
WITH OUTLIER
WITHOUT OUTLIER
AFFECTED
MEAN
– 28
25.667
RESISTANT
MEDIAN
26
28.25
RESISTANT
MODE
31
31
AFFECTED
RANGE
381
16
STANDARD DEVIATION
$s = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}}$


RESPONSE TO AN OUTLIER
MEASURE
WITH OUTLIER
WITHOUT OUTLIER
AFFECTED
MEAN
– 28
25.667
RESISTANT
MEDIAN
26
28.25
RESISTANT
MODE
31
31
AFFECTED
RANGE
381
16
STANDARD DEVIATION
$s = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}}$


RESPONSE TO AN OUTLIER
MEASURE
WITH OUTLIER
WITHOUT OUTLIER
AFFECTED
MEAN
– 28
25.667
RESISTANT
MEDIAN
26
28.25
RESISTANT
MODE
31
31
AFFECTED
RANGE
381
16
AFFECTED
STANDARD DEVIATION
$s = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}}$