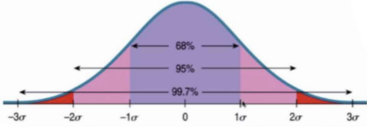Standard Deviation is Spread (ex: any Z Score that is 2 or more standard devation away from the Mean (avg) is an outlier)

The 68 – 95 – 99.7 Rule

- In a Normal model:
  - About 68% of the values fall within 1 standard deviation of the mean
  - About 95% of the values fall within 2 standard deviations of the mean
  - About 99.7% of the values fall within 3 standard deviations of the mean

FAKE DATA SET						
Notes	Market Data	Survey	Delta Δ	Z Score σ	IQR Outliers	Notes
The wider the spread, the less accurate the data is	60,000	Cthra	▼-73.82%	-1	Yes	Need to revist survey job leveling as this job code might have been mapped incorrectly
	75,000	Croner	▼-67.27%	-1	Yes	Need to revist survey job leveling as this job code might have been mapped incorrectly
Any # greater than Upper Quartile + 1.5 x IQR = Outlier*	275,000	Mercer	▲20.00%	0	No	
Any # less than the Lower Quartile - 1.5 x IQR = Outlier*	290,000	Culpepper	▲26.55%	0	No	
	310,000	Aon	▲35.27%	1	No	
You can use either Delta Δ or Z Score or IQR to identify data integrity risks.						
	365,000	Radford	▲59.27%	1	No	

Median   **282,500**
Mean (avg)   **229,167**

Mean (50th P)	$	229,167
Standard Deviation (sample dataset)	$	128,974
Upper Quartile (75th P)	$	305,000
Lower Quartile (25th P)	$	125,000
IQR	$	180,000
Outlier UQ	$	575,000
Outlier LQ	$	145,000
Sample Size		6

Footnotes**

1a) The **median** answers the question "Where is the midpoint of the data?" Also known as the 50th percentile, the median is the value that splits the data into two equal halves: 50% of the data is lower than the median, and 50% of the data is higher. To calculate the median, sort the data points in order, and then locate the point in the middle.

1b) The **mean** is the sum of the data divided by the number of data points. You can think of the mean as answering the question "If every data point contributed the same amount, what would that amount be?"

2) It is important to circle "outliers" as those are red flags to look out for. We can identify outliers with the following formulas. Remember to use "PEMDAS" order whenever you are "+,-,/,or *" in excel

3) **Standard Deviation** - identify the spread

$$\sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}}$$

- σ , lowercase sigma, is the symbol for standard deviation.
- Σ , uppercase sigma, is the symbol for summation or the sum.
- X , represents each point of data in the dataset.
- $\bar{X}$, represents the mean of the dataset and is pronounced "x-bar."
- n , represents the total number of points in the dataset.

Z Score- How many standard deviations is the value away from the mean?

Z Score Formula

$$Z = \frac{X - \mu}{\sigma}$$

4) **Delta** identify the change % (variance or differences)

Annualized Rate of Return

$$\text{Annualized Rate of Return} = \left(\frac{\text{Ending Value}}{\text{Beginning Value}}\right)^{\frac{1}{n}} - 1$$

M
Variance
▲25.00%
▲3.77%
▼-88.37%
▲2.38%
▲7.32%
▲2.86%
▲15.15%
▲21.21%