Lesson 3.3 - February 18, 2021

Conventions

- The median of a dataset is the center value when the numbers are sorted from smallest to largest
- If there are an even number of values in the dataset, the median is the average of the middle two values
- If the data is skewed right, we would expect the mean to be larger than the median
- if the data is skewed to the left, the opposite would be true

Symmetric Distributions

- Standard Deviation tells us about the variance of the dataset, or how spread out the data is
- Mean is the average of the dataset
- Mathematically, the mean refers to the sum of all the numbers, divided by the total number of terms
- The mean could be imagined as the fulcrum of a see-saw, being dynamic based on the relative weights of the objects on either side
- Outliers have a very noticeable effect on the mean of a specific dataset
- In StatCrunch Histograms, you can choose to display the mean in the graph.

Conventions

- The mean of a specific dataset it either represented as $\bar{x}$ or as μ
- The standard deviation of a specific dataset is represented as S
- The formal mean formula is:

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

- The sample size is usually referred to as n

- Standard Deviation is calculated as follows:

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

Symmetric Distribution Rules

- If the data is unimodal and symmetric, then the following rules apply:
- Approximately 68% of all data points are within 1 standard deviation from the mean value
- Approximately 95% are within 2 standard deviations from the mean value
- Approximately 99.7% (pretty much all of the values) are withing 3 standard deviations from the mean value