

DATA PROFILING

Descriptive Statistics

- Descriptive statistics refer to mean, median, and mode and allow you to take a range of numerical inputs and output a number that's descriptive of that range. They are also known as "summary statistics."
- **Mean** is the mathematical average of a set of numbers.
- **Median** is the middle value of a set of quantitative variables when arranged from least to greatest.
- **Mode** is the value that appears most frequently in a set of numbers.

Aggregate Functions

- Aggregate functions can be used to communicate information about what the data is measuring. They include average, sum, and count, and they help us group values across multiple rows or columns.
- **Average** is the same as "mean." Averages are more affected by the presence of outlier numbers.
 - Its syntax is: `=average(firstcell:last cell)`
- **Sum** is the total amount of a given set of numbers.
 - Its syntax is: `=sum(firstcell:last cell)`
- **Count** is the quantity of something.
 - If you're counting cells with numerical data, use: `=count(firstcell:last cell)`.
 - If you're counting cells with alphanumeric data, use: `=counta(firstcell:last cell)`.

Histograms

- Histograms display the frequency of a given value and can help when you're looking to contextualize a value in a straightforward, visual way.
- To create a histogram in Excel using the Analysis ToolPak, you need the input range, bin range, and output range.
 - Input range is the range of cells containing the data for the histogram.
 - Bin range is the range of cells that contain the predetermined bins.
 - Output range is where Excel will put the results.

Skewness

- Skewness is a distribution's degree of asymmetry.
- If the mean is larger than the median, the data set is positively skewed.
- If the mean is smaller than the median, the data set is negatively skewed.

Standard deviation

- Standard deviation is the value that represents how much a set of data departs from the mean.
- It can be calculated by using `STDEV.P` or `STDEV.S` in Excel.