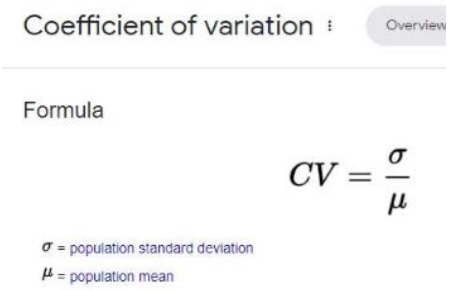
	POPULATION	SAMPLE
Measurement	"Parameter"	"Statistic"
Mean	μ "Mu" (myoo)	$\bar{x}$ "x-bar"
Variance	σ^2 "sigma-squared"	S^2, $\sigma^2_{\bar{x}}$
Standard Deviation	σ "sigma"	S, $\sigma_{\bar{x}}$

Glossary:

- Variance vs Standard deviation
    - Variance: The variance measures the average degree to which each point differs from the mean. The average of all data points within a group.
        - Formula: The average of the squared differences from the Mean.
            - Step 1: find the mean of the numbers (just plain and simple mean)
            - Step 2: For each number subtract the mean and then square the result.
            - Step 3: Find the average of those squared differences.
    - Standard deviation: Standard deviation is the spread of a group of numbers from the mean. Square root of the variance.
        - Formula: sqrt(variance)
- COV vs CV
    - coefficient of variation (CV): In statistical analysis, the coefficient of variation (COV) measures relative event dispersion. The COV is equal to the ratio between the standard deviation and the mean. Although COV is most commonly used in comparing relative risk, it may be applied to many types of probability distribution. Note: This applies to a single variable.
        - Example:
        - Formula:



    - Covariance (COV): A covariance refers to the measure of how two random variables will change when they are compared to each other. Note: This is when there is more than one variable.
        - Example:
        - Formula:

For Population

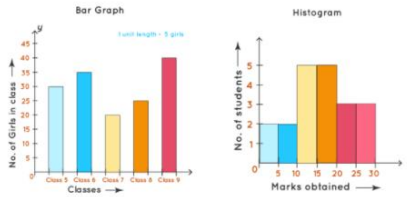
$$Cov(x,y) = \frac{\Sigma (x_i - \bar{x}) * (y_i - \bar{y})}{N}$$

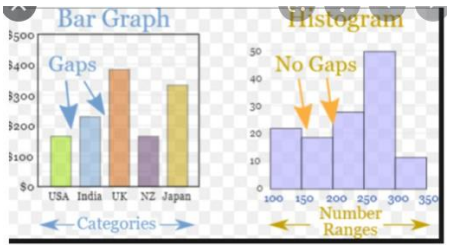
For Sample

$$Cov(x,y) = \frac{\Sigma (x_i - \bar{x}) * (y_i - \bar{y})}{(N - 1)}$$

024 x 526

- Discrete vs continuous data
    - Discrete: whole, concrete numbers with specific and fixed data values that can be determined by counting.
        - Example: number of students in a class, test questions answered correctly, etc.
    - Continuous: complex numbers and varying data values that are measured over a specific time interval.
        - Example: Age. Continuous variables would take forever to count. In fact, we would get to forever and never finish counting them. For example, take an age. We can't count "age". Because it would literally take forever. For example, it could be 37 years, 9 months, 6 days, 5 hours, 4 seconds, 5 milliseconds, 6 nanoseconds, 77 picoseconds…and so on.
- Categorical vs quantitative data
    - Categorical: Cannot be added, subtracted, etc.
        - Example: Zip code, color of shirts, etc.
    - Quantitative: Numerical and can be added, subtracted, etc.
        - Examples: Age, weight of products, etc.
- Histogram vs bar chart
    - Histogram: for quantitative data. No spaces between bars. X axis contains the continuous variable (numerical values). Usually requires "binning" logic where the observations are grouped into ranges.
    - Bar chart: Used for counting the number of observations in each categorial data group. X axis contains the categories / discreet variables.
    - Example:





- Frequency vs Relative frequency vs cumulative frequency
    - Frequency: The number of times a result occurs.

        - Formula: Count the observations in that group.
    - Relative frequency: How often that observation happens.
        - Formula: (frequency of the observation)/n number of observations in the population.
            - Relative frequency = 28/100 = 0.28 (assuming 28 people use windows and 100 people are in the survey).
    - Cumulative frequency: The cumulative frequency is calculated by adding each frequency from a frequency distribution table to the sum of its predecessors. The last value will always be equal to the total for all observations, since all frequencies will already have been added to the previous total.
        - Example: Each month you find the relative frequency. Such as profit made on mac n cheese. In the second month you add the profit from the first and second months to get the cumulative frequency. It rolls up from there each month.
        - Formula:

Table 1. Cumulative frequency of daily rock climber counts recorded in Lake Louise, Alberta, 30-day period

Stem	Leaf	Frequency (f)	Upper value	Cumulative frequency
0	4	1	4	1
1	8 9	2	19	1 + 2 = 3
2	3 4 6	3	26	3 + 3 = 6
3	1 5 5 7 9	5	39	6 + 5 = 11
4	0 1 2 3 5 9	6	49	11 + 6 = 17
5	0 1 1 2 4 4 5 6 7	9	57	17 + 9 = 26
6	0 2 3 5	4	65	26 + 4 = 30

- Nominal vs ordinal vs interval vs ratio data
    - Nominal: Categorical. Variables are labeled with no specific order. Typically nouns.
        - Example: gender, country, race
    - Ordinal: Categorical and ranked. Scale that has all its variables in a specific order, beyond just renaming them.
        - Example: Placement in a race (1st, 2nd, 3rd, etc)
    - Interval: Categorical, ranked, and evenly spaced.
    - Ratio: Categorical, ranked, evenly spaced, and natural zero.
- Python data types