Correlation

Covariance

is a measure that tells you how two variables move together. It considers both the direction and the magnitude of their linear relationship.

- **Direction:**
  - Positive covariance: When one variable increases, the other generally increases too, and vice versa.
  - Negative covariance: An increase in one variable is accompanied by a decrease in the other.
- **Magnitude:** Covariance doesn't tell you the strength of the relationship, just the direction. The value it produces can be positive or negative infinity, depending on how much the variables tend to move together in the same or opposite directions.

Covariance (Cov(X, Y))

The covariance formula calculates the average product of the deviations from the mean for two variables (X and Y).

$$Cov(X,Y) = (\Sigma(Xi - \bar{X}) * (Yi - \bar{Y}))/N$$

Where:

- Σ (sigma) represents the sum across all N data points.
- Xi and Yi are the individual values of variables X and Y.
- X̄ (X bar) and Ȳ (Y bar) represent the mean of X and Y respectively.
- N is the total number of data points.

Hint : Covariance wouldn't tell you how strong that relationship is

Correlation

Correlation, like covariance, is a statistical concept used to understand the relationship between two variables. But correlation builds on covariance to give a clearer picture. Measures the **strength** and direction of the linear relationship. It takes the covariance and standardizes it by considering the standard deviations of both variables. This results in a value between -1 and +1, independent of the original units.

Values:

- +1: Perfect positive correlation (as one variable increases, the other always increases proportionally).
- 0: No linear relationship (the variables move independently).
- 1: Perfect negative correlation (as one variable increases, the other always decreases proportionally).

correlation does not imply causation

Why Correlation Doesn't Equal Causation?

- **Third Variable:** A hidden third variable might be influencing both variables, creating a false sense of correlation. For example, imagine an ice cream stand sees a rise in sales alongside an increase in raincoat purchases. The correlation doesn't mean rain **causes** ice cream sales to go up. It's more likely that hot weather (the third variable) is driving both factors (people buy ice cream to cool down and need raincoats for the sudden downpour).
- **Coincidence:** Sometimes, random chance can create a correlation that doesn't reflect a true underlying relationship.

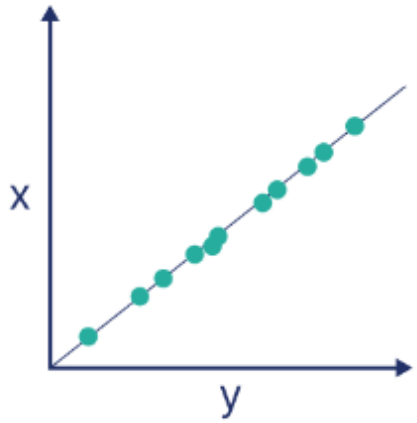
Correlation Coefficient (r)

The correlation coefficient formula takes the covariance and standardizes it by dividing by the product of the standard deviations of X (σX) and Y (σY). This gives a value between -1 and 1 that is independent of the units of the original variables.

$$r = Cov(X, Y)/(\sigma X * \sigma Y)$$

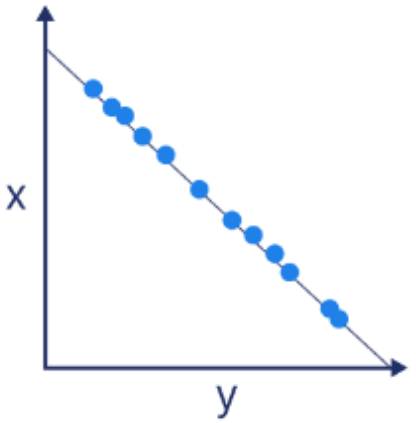
Perfect positive correlation

$r = 1$



Perfect negative correlation

$r = -1$



No correlation

$r = 0$