Covariance

This is a measure of how two random variables change together, or the strength of their correlation.

Consider two random variables, X and Y, each with n values (i.e., $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$). The covariance of X and Y can be found using either of the following equivalent formulas:

$$\texttt{cov}(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})$$

$$\texttt{cov}(X,Y) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{2}(x_i - x_j) \cdot (y_i - y_j) = \frac{1}{n^2}\sum_{i}\sum_{j>i}(x_i - x_j) \cdot (y_i - y_j)$$

Here, $\bar{x}$ is the mean of X (or μ_X) and $\bar{y}$ is the mean of Y (or μ_Y).

Pearson Correlation Coefficient

The Pearson correlation coefficient, $\rho_{X,Y}$, is given by:

$$\rho_{X,Y} = \frac{\texttt{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{n\sigma_X \sigma_Y}$$

Here, σ_X is the standard deviation of X and σ_Y is the standard deviation of Y. You may also see $\rho_{X,Y}$ written as $r_{X,Y}$.