



# Correlation

## Covariance

Covariance is a quantitative measure of the extent to which the deviation of one variable from its mean matches the deviation of the other from its mean. It is a mathematical relationship that is defined as:

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

That is a little hard to wrap your mind around (but worth pushing on a bit). The outer expectation will be a weighted sum of the inner function evaluated at a particular  $(x, y)$  weighted by the probability of  $(x, y)$ . If  $x$  and  $y$  are both above their respective means, or if  $x$  and  $y$  are both below their respective means, that term will be positive. If one is above its mean and the other is below, the term is negative. If the weighted sum of terms is positive, the two random variables will have a positive correlation. We can rewrite the above equation to get an equivalent equation:

$$\text{Cov}(X, Y) = E[XY] - E[Y]E[X]$$

### ***Lemma: Correlation of Independent Random Variables:***

If two random variables  $X$  and  $Y$  are independent, then their covariance must be 0.

$\text{Cov}(X, Y) = E[XY] - E[Y]E[X]$	Def of Cov
$= E[X]E[Y] - E[Y]E[X]$	Lemma Product of Expectation
$= 0$	