Correlation

\[ \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}} \]

Sometimes called correlation coefficient.
Sometimes denoted \( \rho(X, Y) \) or \( p \).

**Idea**

Covariance measures strength of relationship.
Suppose you have \( X, Y \) with \( \text{Cov}(X, Y) = 5 \)
Define \( U = 10X \), \( V = 10Y \)
Then \( \text{Cov}(U, V) = \text{Cov}(10X, 10Y) = 100 \text{Cov}(X, Y) = 500 \)

Are \( U, V \) are 100 times more strongly related than \( X \) and \( Y \)? No.

Covariance is influenced by size of random variables, not just strength of relationship.

**Solution:** Normalize.

Note: \( \text{Corr}(X, Y)^2 = \frac{\text{Cov}(X, Y)^2 \text{Cov}(X, Y)}{\text{Cov}(X, X) \text{Cov}(Y, Y)} \)

**Properties of Correlation**

1. \( -1 \leq \text{Corr}(X, Y) \leq 1 \)
2. \( \text{Corr}(X, Y) = 1 \) if and only if \( Y = aX + b \) for some \( a, b \in \mathbb{R} \) with \( a > 0 \)
3. \( \text{Corr}(X, Y) = -1 \) if and only if \( Y = aX + b \) for some \( a, b \in \mathbb{R} \) with \( a < 0 \)
Example

Consider tossing a coin n times. \( P(\text{heads}) = p \).

\( X = \text{# of heads}, \ Y = \text{# of tails} \).

\( X + Y = n \)

\( \therefore \ E[X] + E[Y] = n \)

\( \therefore \ E[X] = - (Y - E[Y]) \)

\[ \begin{align*}
\text{Cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] \\
&= E[- (Y - E[Y])(Y - E[Y])] \\
&= -\text{Var}(Y) \\
\text{Corr}(X, Y) &= \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}} \\
&= - \frac{\text{Var}(Y)}{\sqrt{\text{Var}(Y) \cdot \text{Var}(Y)}} \\
&= -1
\end{align*} \]

We knew that it must be \(-1\) because

\( X + Y = n \), i.e., \( Y = -X + n \)