The standard measure of the covariation between two variables is the population covariance:

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^{N} (x_i - \mu_x)(y_i - \mu_y)}{N},$$

where $\mu_x$ and $\mu_y$ are the arithmetic population means of $X$ and $Y$, respectively.

Note that the variance on $X$, i.e., $\sigma^2$, is simply the covariance of $X$ with itself:

$$\text{var}(X) = \text{cov}(X, X) = \frac{\sum_{i=1}^{N} (x_i - \mu_x)(x_i - \mu_x)}{N}.$$

The standard measure of interaction within a sample is the sample covariance

$$s_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n - 1}.$$  

Note that the sample variance on $X$, i.e., $s^2$, is simply the sample covariance of $X$ with itself:

$$s_x^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})}{n - 1} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}.$$
The standard parameter of interaction is the **population correlation coefficient**:

$$\rho_{x,y} = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y}.$$  

The standard estimator of the population parameter of interaction is the **sample correlation coefficient**:

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{s_{x,y}}{s_x s_y}.$$  

The sample correlation coefficient is very informative because it measures both **strength and direction**. Note that $r$ is always between -1 and 1. If $r$ is close to either -1 or 1, then the variables $X$ and $Y$ are **strongly** correlated. If $r$ is less than 0, then the variables move in opposite directions and are **negatively correlated** (e.g., interest rates and bond prices); and if $r$ is greater than 0, then the variables move in the same direction and are **positively correlated**. (e.g., height and weight). If $r$ is close to 0, then the variables are **uncorrelated**, and the variables move independently of each other (e.g., height and IQ).

**INTERACTIVE VISUAL DISPLAY OF THE CORRELATION COEFFICIENT**

[GO TO the Correlation Simulator](<http://SunSITE.univie.ac.at/spreadsite/statexamp/>)
EXAMPLE

Suppose we two samples: \{1, 2, 3, 4, 5\} and \{-3, -1, 5, 7, 12\}.

The sample means are 3 and 4, respectively. Then the sample covariance is

\[
\begin{align*}
    s_{xy} &= \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \\
    &= \frac{1}{5-1} \left( (1-3)(-3-4) + (2-3)(-1-4) + (3-3)(5-4) + (4-3)(7-4) + (5-3)(12-4) \right) \\
    &= \frac{38}{4} = 9.5
\end{align*}
\]

Recall that the sample variance for the first sample was 2.5; the sample variance for the second sample is 37. Therefore, the sample correlation coefficient is

\[
    r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{(n-1)sxsy} = \frac{s_{x,y}}{sxsy} = \frac{9.5}{\sqrt{2.5} \sqrt{37}} = 0.98776297.
\]

PROBLEMS:
Ch. 2, #28, 29