Chapter 3: Covariance, regression, and correlation

Covariance

Covariance is a measure of association and the covariance between x and y would be denoted by $\sigma(x, y)$. If x and y are independent then $\sigma(x, y) = 0$, BUT if $\sigma(x, y) = 0$, x and y aren't necessarily independent.

Useful identities for cov

Covariance of x with itself = variance of x:

$$\sigma(x, x) = \sigma^2(x, x)$$

For constants (here represented by a):

$$\sigma(a, x) = 0$$

$$\sigma(ax, y) = a\sigma(x, y)$$

$$\sigma^2(a, x) = a^2\sigma^2(x)$$

$$\sigma[(a + x), y] = \sigma(x, y)$$

The covariance of 2 sums can be written as the sum of covariances, i.e. just multiply out the brackets (I've left this blank, do it yourself or check book):

$$\sigma[(x + y), (w + z)] = ...$$

Variance of a sum is sum of variances and covariances (figure this out):

$$\sigma^2(x + y) = ...$$

Least squares linear regression

Linear model:

$$y = \alpha + \beta x + e$$

Continuing on, α and β will be the true population values and a and b will be the intercept and slope for the line of best fit derived from observed data. The derivation of a and b using the least-squares model can be found on pages 39-41. Buuut, who cares about that, here are the results:

$$a = \bar{y} - b\bar{x}$$

$$b = \frac{Cov(x, y)}{Var(x)}$$