Preliminary Statistics September 2012

Example: Expectation, Variance, and Covariance

	1.	2.	3.	4.	5.
X	7	5	8	3	4
Y	3	1	5	2	7

The expected values of the random variables X and Y are their respective means:

$$E(X) = \frac{1}{N}\sum x_i = \frac{7+5+8+3+4}{5} = 5.4 = \mu_x$$

$$E(Y) = \frac{3+1+5+2+7}{5} = 3.6 = \mu_Y$$

The variance can be estimated by either using the well known formula for the population variance (assuming that the five observations cover the entire population):

$$Var(X) = \frac{1}{N}\sum (x_i - \mu_x)^2 = \frac{(7-5.4)^2 + (5-5.4)^2 + \cdots}{5} = 3.44 = \sigma_X^2$$

$$Var(Y) = 4.64 = \sigma_Y^2$$

Or one can use the relationship:

$$Var(X) = E(X-\mu_X)^2 = E(X^2) - \mu_X^2 = \frac{1}{N}\sum x_i^2 - \mu_X^2 = \frac{49+25+64+\cdots}{5} - 5.4 = 3.44$$

You can also think of this in terms of probabilities. Assuming that all the above values for X occur with the same probability i.e. $P(x_i) = f(x_i) = 1/5 = 0.2$:

$$Var(X) = \sum f(x_i)\,(x_i - \mu_x)^2 = 0.2 * (7-5.4)^2 + 0.2 * (5-5.4)^2 + \cdots = 3.44$$

The covariance of the two variables is defined as:

$$Cov(X,Y) = \frac{1}{N}\sum (x_i - \mu_x)(y_i - \mu_Y) = \frac{(7-5.4)*(3-3.6) + (5-5.4)*(1-3.6) + \cdots}{5}$$
$$= 0.56$$

Alternative:

$$Cov(X,Y) = \frac{1}{N}\sum x_i y_i - \mu_X \mu_Y = 20 - 19.44 = 0.56$$

In terms of expectation operators one can use:

$$Cov(X,Y) = E(X - \mu_x)(Y - \mu_Y) = E(XY) - E(X)E(Y) = \frac{1}{N}\sum x_i y_i - \mu_X \mu_Y$$
$$= 20 - 19.44 = 0.56$$

In terms of PDF one can use:

$$Cov(X,Y) = \sum\sum (x_i - \mu_x)(y_i - \mu_Y) f(x,y) = 0.2 * (7 - 5.4) * (3 - 3.6) + 0.2$$
$$* (5 - 5.4) * (1 - 3.6) + \cdots = 0.56$$

If the probabilities for each of the values are not equal i.e. are not 0.2 the weighting would be done according to the respective probabilities.

NOTE: This is an example in which it is assumed that with the sample of 5 observations for each of the variables the entire population is observed. Hence, the true parameters for the mean and variance can be estimated without a bias. This is not the case as these parameters are estimated based on a sample.