1 The Summation Operator

Because many elementary propositions in econometrics involve the use of sums of numbers, it should be useful to review the summation operator, i.e. Σ. Assume a random variable (henceforth r.v) denoted X from which a sample of n quantities are observed, i.e., $X_i, i = 1, ..., n$.

Then the total sum of the observations $(X_1 + X_2 + \cdots + X_n)$ can be represented as

$$\sum_{i=1}^{n} X_i = X_1 + X_2 + \cdots + X_n \tag{1}$$

The following summation operator rules are useful.

Rule 1. The summation of a constant k times a r.v X_i is equal to the constant times the summation of that r.v.

$$\sum_{i=1}^{n} kX_i = k\sum_{i=1}^{n} X_i \tag{2}$$

PROOF)

$$\sum_{i=1}^{n} kX_i = kX_1 + kX_2 + \cdots + kX_n = k(X_1 + X_2 + \cdots + X_n) = k\sum_{i=1}^{n} X_i$$

Rule 2. The summation of the sum of observations on two r.v's is equal to the sum of their summations.

$$\sum_{i=1}^{n}(X_i + Y_i) = \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} Y_i \tag{3}$$

PROOF)

$$\begin{aligned}\sum_{i=1}^{n}(X_i + Y_i) &= (X_1 + Y_1) + (X_2 + Y_2) + \cdots + (X_n + Y_n) \\ &= (X_1 + X_2 + \cdots + X_n) + (Y_1 + Y_2 + \cdots + Y_n) = \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} Y_i\end{aligned}$$

Rule 3. The summation of a constant over n observations equals the product of the constant and n.

$$\sum_{i=1}^{n} k = nk \tag{4}$$

PROOF)

$$\sum_{i=1}^{n} k = \underbrace{k + k + \cdots + k}_{\text{n times}} = nk$$

Using the above 3 rules, useful results concerning the mean, variance, and covariance of r.v's can be obtained. First, let's define the average of n observations of the r.v X as

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{5}$$

Using this definition, we can prove Rule 4.

Rule 4. The summation of the deviations of observations on X about its mean is zero.

$$\sum_{i=1}^{n}(X_i - \overline{X}) = 0 \tag{6}$$

Lowercase letters is used to represent deviations form, that is, $x_i = X_i - \overline{X}$. Hence,

$$\sum_{i=1}^{n} x_i = 0 \tag{7}$$

PROOF)

$$\overline{x} = \frac{\sum x_i}{n} = \frac{\sum(X_i - \overline{X})}{n} = \frac{\sum X_i}{n} - \overline{X} = \overline{X} - \overline{X} = 0$$

Now, let's define the variance of X to be

$$Var(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{8}$$

and the covariance of X and Y to be

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) \tag{9}$$

Using these definitions and earlier results, we can prove two more summation rules.

Rule 5. The covariance between X and Y is equal to the mean of the products of observations on X and Y minus the product of their means.

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) = \frac{1}{n}\sum_{i=1}^{n} X_iY_i - \overline{XY} \tag{10}$$

PROOF)

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) &= \frac{1}{n}\sum_{i=1}^{n} X_iY_i - \frac{1}{n}\sum_{i=1}^{n}\overline{X}Y_i \\ &\quad - \frac{1}{n}\sum_{i=1}^{n} X_i\overline{Y} + \frac{1}{n}\sum_{i=1}^{n}\overline{XY}\end{aligned}$$

and using Rule 1, we get

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n} X_iY_i - \frac{1}{n}\overline{X}\sum_{i=1}^{n} Y_i - \frac{1}{n}\overline{Y}\sum_{i=1}^{n} X_i + \frac{1}{n}\sum_{i=1}^{n}\overline{XY}$$

Now, recalling the definition of the mean of X and the mean of Y,

$$\begin{aligned}
Cov(X,Y) &= \frac{1}{n}\sum_{i=1}^{n} X_iY_i - \overline{XY} - \overline{YX} + \frac{1}{n}\sum_{i=1}^{n}\overline{XY} \\
&= \frac{1}{n}\sum_{i=1}^{n} X_iY_i - 2\overline{XY} + \overline{XY} \\
&\text{since } \sum_{i=1}^{n}\overline{XY} = n\overline{XY} \text{ by Rule 3} \\
&= \frac{1}{n}\sum_{i=1}^{n} X_iY_i - \overline{XY}
\end{aligned}$$

Rule 6 follows easily from Rule 5, since it applies to the case in which X and X again are two variables.

Rule 6. The variance of X is equal to the mean of the squares of observations on X minus its mean squared.

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 \tag{11}$$

Note, incidentally, that when X and Y happen to have zero means (as occurs when they are measured in deviations about their means), the definitions of covariance and variance become

$$Cov(x,y) = \frac{1}{n}\sum_{i=1}^{n} x_iy_i \qquad \text{and} \qquad Var(x) = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

In certain situations it will be necessary to use summations which apply to two random variables, called *double summations*. Specifically, let X_{ij} be a r.v which takes on n values for each outcome of i and j. There will, of course, be n^2 total outcomes. Now we define the double summation of these n^2 outcomes as

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij} &= \sum_{i=1}^{n}(X_{i1} + X_{i2} + \cdots + X_{im}) \\
&= (X_{11} + X_{12} + \cdots + X_{1m}) + (X_{21} + X_{22} + \cdots + X_{2m}) \\
&+ \cdots + (X_{n1} + X_{n2} + \cdots + X_{nm})
\end{aligned}$$

The following two double-summation rules will be useful.

Rule 7.

$$\sum_{i=1}^{n}\sum_{j=1}^{m} X_iY_j = \left(\sum_{i=1}^{n} X_i\right)\left(\sum_{j=1}^{m} Y_j\right) \tag{12}$$

Note that the double summation in Rule 7 is very different from the single summation $\sum_{i=1}^{n} X_i Y_i$, which contains n (rather than n^2) terms.

Rule 8.

$$\sum_{i=1}^{n}\sum_{j=1}^{m}(X_{ij} + Y_{ij}) = \sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{m} Y_{ij} \tag{13}$$