Summation Notation

The most frequently appearing notation in the mathematical descriptions of different quantities or procedures used in data analysis involves the application of the summation "operator," represented by the upper-case Greek letter sigma, or Σ.

The way to think of the summation operator is as a shorthand indication that the values represented by the symbols to the right of the operator should be added together. More specifically, the notation indicates that a sequence of numbers (usually represented by symbols, such as x_i (standing for the "i-th" value of x in a list of numbers of length n)), should be added together in the sequence defined by the subscripts (the i's).

An example should make this clearer:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \ldots + x_n.$$

The expression $\sum_{i=1}^{n} x_i$ is read as "the sum of x_i (x sub i) as i goes from 1 to n," and the right-hand side of the equation shows the individual x's being added together.

For example, if there are five values of the variable, x (i.e. $n = 5$), e.g. $x_1 = 3$, $x_2 = 2$, $x_3 = 6$, $x_4 = 4$, and $x_5 = 1$, then

$$\begin{aligned}\sum_{i=1}^{n} x_i &= x_1 + x_2 + x_3 + \ldots + x_n. \\ &= 3 + 2 + 6 + 2 + 1 \\ &= 14.\end{aligned}$$

There are a few basic rules that the summation operator follows:

$$\sum_{i=1}^{n} a = a + a + a + \ldots + a \text{ (n times)} = n \times a.$$

$$\sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n.$$

$$\sum_{i=1}^{n} x_i^m = x_1^m + x_2^m + \ldots + x_n^m.$$

$$(\sum_{i=1}^{n} x_i)^m = (x_1 + x_2 + x_3 + \ldots + x_n)^m.$$

$$(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i) = (x_1 + x_2 + \ldots + x_n) \times (y_1 + y_2 + \ldots + y_n).$$

$$\sum_{i=1}^{n} ax_i = a\sum_{i=1}^{n} x_i.$$

$$\sum_{i=1}^{n} (ax_i + b) = (a\sum_{i=1}^{n} x_i) + (n \times b).$$

$$\sum_{i=1}^{n} (x_i + b) = \sum_{i=1}^{n} x_i + (n \times b).$$

$$\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i.$$

But note that

$$\sum_{i=1}^{n} x_i^2 \neq (\sum_{i=1}^{n} x_i)^2$$

and

$$\sum_{i=1}^{n} x_i y_i \neq \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i$$

(Double summation) In the case of a variable that has two subscripts (e.g. x_{ij}, where i might stand for location and j for the time of an observation):

$$\begin{aligned}\sum_{j=1}^{m}\sum_{i=1}^{n} X_{ij} &= \sum_{j=1}^{m} (x_{1j} + x_{2j} + \ldots + x_{nj}) \\ &= (x_{11} + x_{21} + \ldots + x_{n1}) + \\ &\quad (x_{12} + x_{22} + \ldots + x_{n2}) + \ldots + \\ &\quad (x_{1m} + x_{2m} + \ldots + x_{nm})\end{aligned}$$

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

(see Griffiths et al., 1991, *Statistical Analysis for Geographers*).