Some Mathematical Notation

We will need to agree on some notation to express a few mathematical operations we will need along this course.

Variables and Constants

We will be referring to *variables*, names attached to quantities whose value may change as we work. Often these are denoted by upper case letters (many times chosen form the end of the alphabet), as $X, Y, Z, \ldots$. When referring to specific numerical values that these variables take, we might use corresponding lower case letters, as x, y, z, ... A collection of values that are somehow connected, may be indicated by using lower case letters with an *index*, as a subscript, as in $x_1, x_2, x_3..., x_k, ..., x_n$ (so, in this example, the index k was used as the generic value of the subscript, and the index n referred to the largest index: we are dealing with n numbers here.

Occasionally, we may have a *formula* expressing what number is the $k-$ th in this sequence. For example, if the numbers are the first 10 integers, we will have $x_1 = 1, x_2 = 2, ..., x_k = k, ..., x_{10} = 10$. If the numbers are, say, the reciprocals of the first n integers, we will have the generic formula $x_k = \frac{1}{k}$, for $k = 1$, $2, ..., n$

Sums

We will often have to consider the sum of several numbers. For small sums, no special notation is needed: if we have to add three numbers x_1, x_2, x_3, it's not a problem to write $x_1 + x_2 + x_3$. This becomes awkward when we are summing many (say, 100, or more) such numbers, or, even worse, when we are keeping the number of addends generic, as in “sum n numbers $x_1, x_2, ..., x_n$”.

For this, the common solution is to employ “sigma notation”, also called “summation notation”: a capital Greek letter $\sum$ (corresponding to the letter S in the Latin alphabet), with indexes expressing the range of the sum. For example, the sum of 10 numbers $x_1, x_2, ..., x_{10}$ is written in a compact way as

$$\sum_{k=1}^{10} x_k$$

The sum of n such numbers would be

$$\sum_{k=1}^{n} x_k$$

This notation helps write some famous formulas very neatly. For example, the sum of the first n integers is

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

and the sum of the squares of the first n integers turns out to be

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$$

Note that adding n times the same number a will result in $\sum_{k=1}^{n} a = n\,a$. Also note how, this notation being a shortcut for a sum, all the usual rules apply. For example,

$$a\sum_{k=1}^{n} x_k = \sum_{k=1}^{n} a\,x_k$$

is the usual distributive rule, while

$$\sum_{k=1}^{n} (x_k + y_k) = \sum_{k=1}^{n} x_k + \sum_{k=1}^{n} y_k$$

expresses the fact that we can add up numbers in any order we like (*commutativity*), and

$$\sum_{k=1}^{n} x_k = \sum_{k=1}^{m} x_k + \sum_{m+1}^{n} x_k$$

$(1 < m < n)$ reminds us that we can group the terms of a sum at will (*associativity*)

.

Applications in Statistical Formulas

We will see that we will need frequently to write a few formulas in statistics. The building blocks are n numbers obtained from observations. Let's call them, as before, $x_1, x_2, ..., x_n$. We will have often to compute

- their sum: $\sum_{k=1}^{n} x_k$
- the sum of their squares: $\sum_{k=1}^{n} x_k^2$

These are then used to compute

- Their *average* or *mean*: $\frac{1}{n}\sum_{k=1}^{n} x_k = \bar{x}$ (the overbar is a common notation). Note that $\sum_{k=1}^{n} x_k = n\bar{x}$
- Their *variance* (we'll get to this in detail later): $\frac{1}{n}\sum_{k=1}^{n} (x_k - \bar{x}\,)^2 = \frac{1}{n}\sum_{k=1}^{n} x_k^2 - \bar{x}^{\,2}$

The last equality follows from a little algebra, which is easy once you become fluent in utilizing summation notation:

$$\frac{1}{n}\sum_{k=1}^{n} (x_k - \bar{x}\,)^2 = \frac{1}{n}\sum_{k=1}^{n} \left(x_k^2 - 2\bar{x}x_k + \bar{x}^{\,2}\right) = \frac{1}{n}\left(\sum_{k=1}^{n} x_k^2 - 2\bar{x}\sum_{k=1}^{n} x_k + n\bar{x}\right) =$$
$$= \frac{1}{n}\left(\sum_{k=1}^{n} x_k^2 - 2\bar{x}\cdot n\bar{x} + n\bar{x}\right) = \frac{1}{n}\sum_{k=1}^{n} x_k^2 - 2\bar{x}^{\,2} + \bar{x}^{\,2} = \frac{1}{n}\sum_{k=1}^{n} x_k^2 - \bar{x}^{\,2}$$