Biol 310-Statistical Notation

Statistical notation appears complex but is easily understood with a little practice. The elementary formulas involved in statistical computing involve summation notation. Before introducing statistical formulas for descriptive statistics which use summation notation, we will define it and illustrate the common forms of occurrence.

The collection of **N** values of a quantitative variable in a population will be denoted by:

$$x_1, x_2, \ldots.. x_N$$

Where the value of the variable is denoted by "x" and the subscripts are used to indicate different unit. Thus, X_1 denotes the value of the variable on the first unit, X_2 is the value on the second unit,.....(the string of periods indicates a continuation up to), X_N is the value on the last (N^{th}) unit. The sum of the N values is indicated by:

$$\sum_{i=1}^{i=N} x_i = x_1 + x_2 + \ldots + x_N$$

This gives the the definition of the summation notation: The capital Greek letter $\sum$ (sigma) denotes summation of the quantity following it; what is added are the values of the quantity for each value of the index of summation "i" from 1 to N; "1" and "N" are the lower and upper limits, respectively, of the summation. When it is clear that the summation includes all values, the limits of summation are omitted and the notation for the sum of all the x's will be denoted as $\sum x_i$.

Example. For the 8 (=N) values: 2, 3, 5, 1, 4, 3, 2, 4, then: (x_1=2, x_2=3, x_3=5, x_4=1, x_5=4, x_6=3, x_7=2, x_8=4)

The sum of the 8 values would be:

$$\sum_{i=1}^{i=8} x_i = 2 + 3 + 5 + 1 + 4 + 3 + 2 + 4 = 24$$

The sum of only the second through the fifth values would be:

$$\sum_{i=2}^{i=5} x_i = 3 + 5 + 1 + 4 = 13$$

The summation instruction terminates at a "+" or "-" operator. For example,

$$\sum x_i - 5 = x_1 + x_2 + \ldots + x_N - 5$$

indicates that all x's are added, then 5 is subtracted from the sum. In terms of the example,

$$\sum x_i - 5 = 2 + 3 + 5 + 1 + 4 + 3 + 2 + 4 - 5 = 19$$

If we want to indicate that 5 is subtracted from each x and the results are summed, then the notation is

$$\sum (x_i - 5)$$

which is equal to (2-5) + (3-5) + (5-5) + (1-5) + (4-5) + (3-5) + (2-5) + (4-5) = -16

When the quantity being summed is a constant, that is, its value is the same for every value of the summation index i, then the value of the sum is the constant multiplied by the number of terms in the summation. Thus,

$$\sum_{i=1}^{i=N} 3 = 3+3+\ldots+3 = 3N$$

From the above results, we see that

$$\sum(x_i-5) = \sum x_i - \sum 5 = \sum x_i - 5N$$

In words, the left-hand side indicates that 5 is subtracted from each x and the results are added; the right-hand side is the sum of the x's minus N times 5. The right-hand side is usually the easier way to do the arithmetic when the value of the left-hand side is wanted.

Adding the squares of the x's would be denoted by: $\sum_{i=1}^{i=N} x_i^2 = x_1^2 + x_2^2 + \ldots + x_N^2$

For the above example, $\sum x_i^2$ = 4 + 9 + 25 + 1 + 16 + 9 + 4 + 16 = 84

A different operation than this is indicated by $(\sum x_i)^2$ which indicates squaring the sum of the x values. For example, this would be $(2+3+5+1+4+3+2+4)^2 = 576$

We close with three algebraic properties of summation:

For the constant c (i.e., c does not change as i changes),

1) $\sum c = Nc$
2) $\sum cx_i = c\sum x_i$
3) $\sum(x_i \pm c) = \sum x_i \pm Nc$

The second property, which was not illustrated above, says that adding the products of the constant c and each x is the same as the product of the constant and the sum of the x's. Because division by a number, a form similar to the second property would be

$\sum(x_i/c) = (\sum x_i)/c$. The third property indicates that the subtraction illustrated above also holds when the operation is addition.