

Some basic terminology and concepts

- t A *population* is a finite or infinite group of individuals or objects. Often it is impossible or impractical to get data on all of the population so a small part of the group or *sample* is examined instead
- t If the sample is representative of the population important conclusions can be inferred by analysing the sample. *Statistical inference* using *probability* deals with the use and validity of samples to infer properties of the population.
- t *Descriptive* statistics deals with the analysis or description of a given set of data, without considering whether it is a population or a sample.
- t A *variable* is an entity that varies. It can have a number of possible values as well as a number of actual or actually observed values
- t A *continuous* variable is generally one where based on measurements such as people's heights, which has no breaks or jumps in the values it can take.
- t A *discrete* variable is usually based upon enumeration or countings which have breaks or jumps. For example if you throw a die it can only take one of six values.
- t If one or more values of a variable Y corresponds to each value of a particular variable X we say that Y is a function of X and write:

$$Y = f(X)$$

Y is the dependent variable

X is the independent variable

we can represent this function in a table or as an

equation and we can extend the concept to two or more variables.

t Equalities and inequalities:

$a = b$	means	a	equals	b	$3 = 3$
$a < b$	means	a	is less than	b	$2 < 5$
$a \leq b$	means	a	is less than or equal to	b	$a \leq 10$
$a > b$	means	a	is greater than	b	$15 > 2$
$a \geq b$	means	a	is greater than or equal to	b	$a \geq 10$

we can combine these $a \leq b < c$ means a is less than or equal to b and b is less than c

t A valid inequality remains valid if:

- ✕ the same number is added or subtracted from both side
eg $2 < 5$; $2 + 3 < 5 + 3 \Rightarrow 5 < 8$
- ✕ Both sides are multiplied or divided by the same positive number
eg $2 \times 2 < 5 \times 2 \Rightarrow 4 < 10$
- ✕ Both sides are multiplied or divided by the same negative number and the inequality is reversed
eg $2 \times -2 > 5 \times -2 \Rightarrow -4 > -10$

t Rounding: In rounding data a convention has to be followed to minimise cumulative rounding errors. For example for 7.389 to two decimal places, this is closer to 7.39 than to 7.38; but for 7.385 we round to 7.38 and for 7.375 we round to 7.37 by rounding the odd number up but the even number down

t Summation notation:

X_j is X subscript j and denotes a particular value of X which is indexed by the value of j

So X_j is $X_1, X_2, X_3, \dots, X_N$ where N is the total number of them

$$\sum_{j=1}^{j=N} X_j \text{ or } \sum_{j=1}^N X_j$$

denotes the sum of the X_j s where j takes the value $1, 2, 3, 4, 5, 6, \dots, N$

$$\sum_{j=1}^{10} X_j = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + X_9 + X_{10}$$

t Note

$$\sum_{j=1}^4 X_j Y_j = X_1 Y_1 + X_2 Y_2 + X_3 Y_3 + X_4 Y_4$$

$$\sum_{j=1}^4 a X_j = a \sum_{j=1}^4 X_j = a (X_1 + X_2 + X_3 + X_4)$$

where a is a constant

$$\sum_{j=1}^4 (X_j + Y_j) = \sum_{j=1}^4 X_j + \sum_{j=1}^4 Y_j$$