Section 8
Introduction to Statistics

In this section we'll introduce the topic of statistics. We consider how statistics experiments should be set up, how data should be collected, how the various variable and data types are defined, what is meant by a significant digits, and how calculation performed with them are affected. Statistics is concerned with scientific methods for collecting, organizing, summarizing, presenting and analysing data, as well as drawing conclusions which are valid and making decisions which are reasonable on the basis of such an analysis.

Population & Sample, Descriptive & Inductive Statistics

In collecting data concerning characteristics of a group of objects or individuals it is often impossible or impractical to analyse the entire group especially when it is large. Instead of examining the entire group, referred to as the *population* or *universe*, we examine just a small and usually representative part of the group called a *sample; e.g.*

> You are required to conduct a survey in which you wish to determine the average height and mass of an adult in a target population. Then, you wouldn't base your conclusions on results from, say, a basketball team as they would, more than likely, be taller than the average person. Similarly you wouldn't sample only obese people for mass determination. Instead, you would seek to find a sample whose height and mass characteristics would be representative of the entire population they are taken from.

This is at the heart of statistical experiment preparation; you ensure, to the best of your ability, that the statistics gathered are a true reflection of the population at large in the areas you're interested in.

Populations can be *finite* or *infinite*. If a sample is representative of the population then from the analysis of the sample we can infer important conclusions about the population at large. We call dealing with such cases where inferences from the sample are valid to the population as a whole as *inductive statistics* or *statistical inference.* Due to the inherent presence of uncertainties in the analysis of samples over populations we often use *probabilities* when stating conclusions.

That part of statistics which deals with merely describing a given group without seeking to draw any inductive or inferred conclusions is called *descriptive* or *deductive statistics.*

Discrete & Continuous Variables

A *variable* is a symbol which can take on any value in a defined *domain* or set of prescribed values. It is referred to as a *constant* if it can only take on one value. A variable which is capable of assuming any value between two stated values is called *continuous* otherwise it is called *discrete; e.g.*

(1) The age of a person is theoretically capable of assuming any age,within the bounds placed by their life span, depending on the accuracy of the measurement. Therefore it is continuous; i.e.
32.1 years, 32.15 years, 32.1578 years …… .

(2) The number of students in a class is discrete. You can't have fractions of people; i.e.

20, 22, 25 but not 22.1.

For continuous variables which describe data the data is called *continuous data* whereas data described by discrete variables is called *discrete data.* In general, *measurements* give rise to continuous data whereas *countings* or *enumerations* give rise to discrete data.

Rounding of Data

The rounding of a number like 32.7 to the nearest unit is 33 because 0.7 is nearer to 1 then 0. Similarly 32.7146 rounded to the nearest hundredth is 32.71 as 32.7146 is nearer to 32.71 then 32.72. In rounding 32.715 to the nearest hundredth we have a dilemma since 32.715 is halfway between 32.71 and 32.72. However the standard practice is to round the number is such cases to the nearest *even* digit. Thus 32.715 rounds to 32.72.

Significant Figures

If a mass is recorded accurately as being 65.4 kg then this means that the true mass lies between 65.35 kg and 65.45 kg. The accurate digits, apart from the zeroes needed to locate the decimal point, are called the *significant digits* or *significant figures* of the number.

Examples:

(1) 65.4 has 3 significant figures.
(2) 4.5300 has 5 significant figures
(3) 0.0018 has 2 significant figures
(4) 0.001800 has 4 significant figures

Numbers associated with enumerations or countings are exact and have an unlimited number of significant figures. However, in some cases it may be difficult to decide which figures are significant without further information. If, for example, we took the number 256 000 000 we could have 3, 4, …., 9 significant figures. If we knew it had 6 significant places then it could be written as 256.000 million or 2.56000×10^8 .

Computations

In performing calculations using multiplication, additions, division or subtraction the result can have no more significant figures then the number with the least significant figures;

e.g.

$$73.24 \times 4.53 = (73.24)(4.53) = 331$$

$$\sqrt{38.7} = 6.22$$

$$\frac{1.648}{0.023} = 72$$

$$3.16 + 2.7 = 5.9$$

$$83.42 - 72 = 11$$

$$47.816 - 25 = 22.816 \quad \Leftrightarrow 25 \equiv 25.000 \text{ ; i.e. 25 is exact.}$$