# Statistics: Background

## The Importance of Statistics

You will probably be aware of the extent of the use and 'influence' of statistics. You will probably have read about or heard of such statistics as "average earnings" and the "Retail Price Index" (or "Rate of Inflation"). These kinds of statistics are used in negotiations that lead to decisions that can have major effects on living standards, national as well as individual.

You have probably seen people standing around, usually in busy shopping centres, clutching clipboards and stopping the occasional passer-by to ask questions. You might even have had someone calling at your home, asking for your help with a survey of some kind, or you might have been asked to complete a questionnaire. The results of such surveys are often used to influence the types and qualities of goods and services that you are offered in the future.

Every ten years in the UK, all households must complete a census return. This is used to compile statistics on population patterns and developments. These statistics provide information for government planning of the provision of public services such as schools and hospitals, housing (both public and private sector) and even the development of industry.

In an environment where data and statistics play such an important role, it is in the interests of us all to know more about them, especially how they are collected, analysed and used. Moreover, and perhaps even more importantly, you might be called upon to carry out the collection and/or analysis of data which will lead to an important operational decision.

### Terminology

You will be introduced to some words that have specific meanings in the context of statistics. It is important to be clear about the terminology.

So far, we have used the word **statistics** without giving it a clear meaning. In this section of the module we can be content with two simple meanings that must be used in the correct context.

1. One or several collections of numerical facts. This is widely used when reference is made to information about such things as population distribution, incomes, and education. Statisticians refer to this sort of information as **data**.

2. The name for the general subject area which we are investigating in these sessions.

In addition, the word **statistic** is often used to refer to a single value which is somehow representative of the whole data set. One such statistic is the mean value, which we will meet later on.

The collection, analysis, and interpretation of data are referred to collectively as **statistical methods** and are detailed as:

1. **Descriptive Statistics:** This deals with the compilation and presentation of data in various forms such as tables, graphs and diagrams. The purpose of descriptive statistics is to display and pass on information in a 'simplified' form from which conclusions can be drawn and decisions made. Businesses, for example, use descriptive statistics when presenting their annual accounts and reports, and the Government is a particularly prolific provider of descriptive statistics.

2. **Inferential Statistics:** This deals with the tools of statistics, the techniques that are used to analyse the data, and to make inferences (draw conclusions) from the data.

## The Developing Importance of Statistics

Statistics have been collected for centuries. Governors needed to have information about the population and their possessions so that taxes could be levied. In addition, it has been essential for them to be aware of the military strength of the nation. In the sixteenth century, the word 'statist' was used to describe someone who dealt in facts about the state, its government and its people.

In this country the government collects data on many issues including:

1. production,
2. earnings,
3. expenditure,
4. imports and exports,
5. population growth or decline,
6. diseases and its incidence,
7. housing conditions,
8. sexuality and gender identity,
9. ethnic background.

For governments to make sensible decisions, however, these data need to be correctly collected, processed, and analysed.

## Statistics in Business

It is not only governments that have required more and more statistics. Companies have grown to such an extent that some have an annual turnover as great as the annual budget of a national economy. Large firms therefore have

to make decisions on the basis of available data. The statistics that are collected by the Government are vital to businesses, and, in the United Kingdom, the government's Office of National Statistics provides a wealth of information for businesses, etc. Businesses also collect their own statistics or pay specialist companies to collect the data for them. They require information on:

1. the reaction of other companies to their products;
2. the reaction of customers to their products;
3. the effect of other companies' activities on their own;
4. the need for new products.

## Misuse of Statistics

No doubt, you have heard an argument supported by the words 'statistics show that...'. Many people seem to believe that a case is proven if statistics can be produced to (supposedly) support it. So, we are bombarded with an endless stream of data in an attempt to impress, persuade or even coerce us into believing that this political party is wise and good, or that we should buy a particular product or that we should hold certain opinions. What we are seldom told is how the statistics, impressive though they may be, have been collected, where they were collected or from whom they were collected. The old quip about ''lies, damned lies, and statistics' still has relevance. Of course, the data themselves cannot mislead, but the people who present the data certainly can. Always remember:

1. Data can be collected correctly and then used for the wrong purpose;
2. Data can be collected incorrectly so that they are biased;
3. Data can be analysed incorrectly so that the results obtained from them are misleading.

A good statistician will be eager to show you how the data was obtained and dealt with, i.e., data collected correctly and analysed appropriately so that meaningful inferences can be obtained.

## The Collection of Data

Statistics is mostly concerned with the analysis of numerical data. Hence, the first stage in statistical method must be the collection of relevant data. It is an important stage, because the data must be collected correctly otherwise they will be practically useless.

## Populations

Usually, statistical data comprise a series of observations from a **population**. E.g., in the quality control of the production of batteries for use in laptops. A battery is picked at random from the production line, at regular intervals, and tested to see how long it will function correctly. The population here is

the lifetimes of all the batteries on that production line. The lifetimes of those batteries selected for testing are a **sample** from that population.

Another example might involve a network administrator who keeps a record of all network faults. The data might include information such as cause of fault and length of time for repair, possibly so that statistics can be produced after a time and the network service analysed. Here we have two populations : the number of faults in a shift and the repair times.

## Problems of Bias

When we collect data for analysis, we should endeavour to take a **representative sample**. That usually entails taking a **random sample**, i.e., the sampling plan is organised so that every member of the population has an equal chance of being selected as part of the sample. There must be no **bias** towards any individual or any group in the population. The sample should truly represent the population so that any conclusions drawn from the sample can be 'reliably' extended to the population as a whole. The sample must also be 'large enough'. The larger the sample the more representative it is likely to be of the whole population. Unfortunately, obtaining a random and unbiased sample is probably the biggest problem faced by statisticians and we have to rely on the skills of those who collect the data.