Business Statistics

Topic 1: Introduction to Business Statistics

By the end of this topic you will be able to:

- Define statistics
- Recognise the need for statistics in business
- Differentiate between descriptive & inferential statistics
- Identify different types of data and their sources
- Gain an awareness of different methods of collecting data
- Discuss some ethical issues

Contents

1 The need for statistics

Why use statistics? In today's businesses, managers and analysts have access to vast amounts of information and data. So it has become important for them to use statistics to be able to:

- Properly present and describe the available information
- Draw conclusions as sensible as possible based on the available information
- Obtain reliable estimations and forecasts from them

The following are some applications of Statistics in the fields of Economics and Business:

Accounting firms use them for audit purposes. Whenever the client accounts are huge, the common practise is it to look at a sample of it for making decisions. Here they use sampling techniques to pick a reasonable sample from the whole set of information available.

Financial analysts use statistics to choose between alternative portfolio investments and to track the trends of financial measures over time.

Management uses statistics to monitor and improve the quality of the product manufactured.

Marketing uses statistics for market research purposes such as:

- Identifying potential customers for their products using data collection techniques

- Estimating the proportion of customers who prefer one product over another

2 Statistics- A definition and two categories

What is statistics?
Statistics is a method of extracting useful information from a set of numerical data in order to make a more effective and informed decision.

There are two different types of statistics:
- •Descriptive
- •Inferential

Descriptive Statistics deals with methods of organising, summarising, and presenting numerical data in convenient forms such as graphs, charts and tables. The purpose of which is to describe the various features of the data that could be used when making decisions.

For example, “Kentucky Chickens” may want to compare the sales performance over past years at two of its outlets. Sales figures can be summarised using descriptive statistical methods such as average sales and spread of sales. Tables and charts can be used to enhance the presentation of the data so that the essential differences in sales performance at the two outlets can be highlighted.

Inferential Statistics is defined as statistical methods used for drawing conclusions about a population based on samples.

A ***population*** is a collection of all-possible individuals, objects, or measurements of interest.

A ***sample*** is a portion, or part, of the population of interest.

A summary measure of a population is called a **parameter.** A summary measure of a sample is a **statistic.**

For example, a manufacturer of computer chips claims that less than 10% of his products are defective. It is not practical to test each and every product, hence to clarify their claims 100 chips were drawn from a large production run and 7.5% were found defective.
In this example:
- Population is: All computer chips produced by the Production Unit
- Sample is: 100 chips drawn from the production unit
- Value of the parameter is: 10% (Population mean)
- Value of the sample statistic is: 7.5%(Sample mean)

The following are some of the key areas of application of statistical inference.

Estimation: this involves estimating the unknown quantities based on a depending quantity. For example estimating the sales figure based on the amount spent on advertising.

Forecasting: using past occurrences to predict the future. For example looking at the trend of unemployment rate of a country and forecasting for the future.

Testing of Hypotheses and decision making: this involves determining whether any apparent characteristics of a situation are unusual or whether they could have happened by chance. For example a Swim Centre may observe from a sample of their membership that there is relationship between the height and weight of female members in the swim centre. Testing of Hypothesis can be used to check whether this finding is by chance or not.

3 Data Collection

Often it is necessary to collect data when making decisions about a given situation. Data can be sourced either first hand or taken from other sources. If the data is collected by an organisation for its own research purposes, it is called a primary data. If you obtain data, which has been collected by other sources, it becomes secondary data.

Methods of Collecting Data include:

- **Survey**

Ask participants questions about their beliefs, behaviours and characteristics. Responses are then edited, coded and tabulated for analysis

- **Observation**

Observe participants' behavior directly, usually in its natural setting. Most knowledge of animal behaviour is developed in this way. One example is the use of a focus group discussion.

- **Experiment**

Make observations from results of treatments given to participants. For example studying the effectiveness of particular toothpaste.

- **Secondary sources**

Obtaining data from published articles in journals & magazines.

Primary data

Advantages:

- Results are up to date
- Can meet immediate needs

Disadvantages:

- Collection of data can be time-consuming
- It is an expensive process

Secondary Data

Advantages:

- Inexpensive

Disadvantages:

- Data may have been collected for other purposes and may focus on irrelevant characteristics.
- You might not know how the data was collected, over what time period and the source of the information.

4 Types of Data

There are two main categories of data:

- Categorical data
- Numerical data

Categorical data are nonnumeric responses. Examples include gender, religious affiliation and state of birth. An example of a question for a categorical response:

"Are you male or female?"
□ *Male* □ *Female*

"Do you smoke?"
□ *Yes* □ *No*

Numerical data are numerical responses. Examples include the balance in your bank account, or the number of children in your family.

Numerical variables are either:

- Discrete or
- Continuous

Discrete variables can only assume certain values. There are usually "gaps" between the values. For example the number bedrooms in your house cannot be over a continuous range between 1 and 10 because you can't have a fraction of a bedroom or bedrooms. Generally responses of **counting process** are **discrete**, whole numbers. Example:

How many children are there in your family?
How many subjects have you enrolled in for this semester?
How many times do you smoke in a day?

Continuous variables can assume any value within a specific range, such as the time taken for an exam. Responses arising from **measuring process** are **continuous** data. Examples:

How tall are you?

How much do you weigh?
How long did it take you to complete the exam?

5 Using Samples

If the data is collected for the entire population it is called a census. Sometimes it many not be possible/practical to collect the entire data. In such cases a sample is collected to make inferences about the entire population. The following are the main reasons for using a sample.

- Quicker to collect sample data
- Cheaper
- Easier to administer

There are various methods of sampling such as random sampling, systematic sampling, stratified sampling and cluster sampling. In this subject we will assume that we have used the simplest probability sample, which is called a **simple random sample**. In a simple random sample every item or person in the population has the same chance of being chosen as any other item or person in the population. To explain how we obtain a simple random sample we will consider the following example:

> An Internet gambling firm has a list of N = 1000 customers all of whom have a customer number. These numbers start at 1 and go to 1000. The firm wants to know whether the customers are happy with the web site that the firm is now using. Instead of interviewing all 1000 customers the firm selects a simple random sample of n = 10 customers and interviews these customers on the telephone.

6 Ethical Issues

When you use a sample even if you attempt to choose the sample in a random way it may not give you an accurate picture of the important features of the original population. The 4 main errors that arise when we use samples rather than taking a census are as follows:

a. **COVERAGE ERROR** occurs when the sample is chosen in such a way that a certain group or groups of people cannot be selected in the sample. If these people are different from those who can be selected this will lead to **selection bias**.

b. **NONRESPONSE ERROR** occurs when a certain group or groups of people do not respond to the questionnaire. If these people are different from those who do respond will lead to **nonresponse bias**..

c. **SAMPLING ERROR** occurs because any sample only contains part of the population. The exact results we obtain from a sample which has

been correctly selected will be affected by a large number of random factors and so will usually differ from the population results.

d. **MEASUREMENT ERROR** occurs when the responses we obtain do not actually measure what they are supposed to measure.

It is the ethical responsibility of anyone reporting the results of a sample survey to

1. Identify for their readers both the type of sample and the type of errors that may be associated with their survey.
2. Explain the possible implications for their conclusions of these errors.

7 Summary

In this topic you have:

- Reviewed the role of statistics in business
- Distinguished between descriptive & inferential statistics
- Identified different types of data
- Identified methods of data collection.