**Unit 8: Inferential Statistics**

Welcome to week 8 where you will be introduced to the different analyses available for the different types of quantitative data. This will mean looking at ways to describe your data and how to create and perform hypothesis tests. This is called inference, as we are making inferences about a population from a sample of data. Inference is the process of extracting meaningful and useful business information from raw data. This process is known as ***statistical inference***, because it involves using the data to make valid inferences about the underlying population. As data is inherently variable, all such inferences will necessarily be subject to uncertainty. This uncertainty is quantified using probability.

Data consist of the observed values of one or more variables of interest. They are usually organised into **datasets**. A dataset can be thought of as a table, whose columns represent the different **variables**, and whose rows represent the individual **observations**. Each individual cell in the table contains the value that the variable takes for the given observation.

In order to extract meaningful and useful business intelligence from data, it is important that the methodology to be employed is valid for the variable or variables of interest. The first important distinction concerns the **level of measurement** of each variable.

When exploring data, it is helpful to try to summarise the data in some meaningful way. One approach is to construct a graphical summary. In this section, we (very briefly) explore some of the more useful graphical summaries.

It is often convenient to summarise numerical data using a few simple summary measures. Most commonly, two such measures are used. The first of these is a measure of **location**, and represents the value taken by “a typical observation” – that is, by an observation that falls “right in the middle of the data”. As well as knowing the magnitude of a typical observation, some idea of the “variability” or “spread” of the data is also useful. This is provided by a measure of **dispersion**.

Instead of *estimating* some population value of interest that underlies the data, an alternative form of inference is to use the data to provide evidence about whether some assumption of interest regarding that population value is likely to be true. Such a form of inference is known as **hypothesis testing** and is usually preferred to the estimation approach when interest lies in comparing the relevant values underlying two or more different populations.

These are all techniques we can employ to find patterns and meaning from our data.

**In this unit we shall:**

* Define the different levels of quantitative data.
* Define measures of location and spread.
* Introduce the concept of inference and hypothesis testing.

**On completion of this unit you will be able to:**

* Identify the different levels of measurement.
* Produce measures of location and spread.
* Perform appropriate hypothesis tests.