# Business Statistics Notes

## *Chapter 1 – What is statistics*

**Types of statistics**

Statistics is a branch of mathematics that transforms data into useful information for decision makers. There are two types of statistics: descriptive and inferential

### Descriptive statistics

Descriptive statistics refer to method of organizing, summarizing and presenting data in ways that are useful, attractive and informative to the reader

### Inferential statistics

Inferential statistics refers to methods used to draw conclusions about characteristics of a population, based on information available in a sample taken from the population

**Key statistical concepts**

1. Population – refers to all of the items (individuals, objects, other measurements) of interest

2. Sample – refers to a portion, part, or subset of the population of interest

3. Parameters – refers to a descriptive measure of a population

4. Statistic – refers to a descriptive measure of a sample

5. Statistical inference – the process of making an estimate, forecast or decision about a population parameter, based on the sample data. Each statistical inference must include a measure of reliability of inference – two such measures include the confidence level and the significance level

6. Confidence level – the proportion of times that an estimating procedure would be correct

7. Significance level – the relative frequency of a wrong conclusion

# Chapter 2 – Types of data, data collection and sampling

**Data**

The objective of statistics is to ==extract information from data==. Data refers to the observations of variables. There are some key terms that we need to know before we can properly examine data
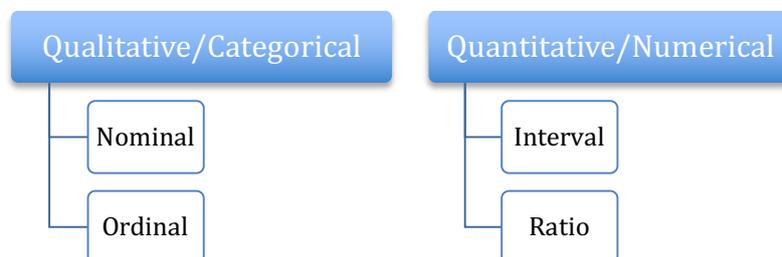
<u>Key terms</u>

1. Estimate – refers to an ==approximate value== of a parameter based on a sample statistic

2. Variable – any characteristic of a sample or population

**Types of data**

Data can either be categorical/qualitative or numerical/quantitative

1. Categorical (Qualitative) data – refers to when observations fall into ==categories==

2. Numerical (Quantitative) data - refers to when observations are ==real numbers==

**Measurements of data**

```
Qualitative/Categorical        Quantitative/Numerical
        |                              |
     Nominal                       Interval
        |                              |
     Ordinal                         Ratio
```

<u>Qualitative/Categorical data</u>

1. Nominal measurement

   - Non overlapping categories
   - Cannot be ranked

2. Ordinal measurement

   - Non overlapping categories
   - Can be ranked

<div align="center">Quantitative/ Numerical data</div>

1. Interval measurement

   - Non overlapping categories
   - Can be ranked
   - Differences are meaningful
   - No true zero point

Ie: the difference between 10 degrees and 20 degrees is the same as the difference 30 degrees and 40 degrees

2. Ratio measurement

   - Non overlapping categories
   - Can be ranked
   - Differences are meaningful
   - True zero point

Ie: weight

## Methods of collecting data

Collecting data involves systematically collecting, recording and analysing information concerning a specific statistical problem. This includes collecting data from both primary and secondary sources

Primary data refers to the facts and figures collected from original sources for the purpose of the specific research problem. There are three main methods of collecting primary data.

Secondary data refers to information that has already been collected by some other person or organization.

<div align="center">Primary Data</div>

1. Observation method

The observation method involves the recording of customer's behaviors with regards to a specific research problem. Ie: to test whether aspirin reduces the incidence of heart attacks, a sample of men and women could be selected, and they could be asked whether or not they have taken aspirin over the last two years, and if they have suffered any heart attacks.

The observation method is beneficial as it is relatively inexpensive, however it is limited in its effectiveness as the information provided through the observation method is often limited in its usefulness. Ie: the aspirin example above does not rule out that people who take aspirin might be healthier, and thus that's why they experience less heart attacks.

2. Experimental method

The experimental method involves gathering data by ==altering factors under tightly controlled conditions, and then evaluating causes and effects==. This method of data collection is generally ==more expensive in both time and money== than the observation method. However this method is usually more accurate than data collected via the observation method.

3. Survey method

Surveys involve ==collecting information== from people concerning a specific research problem. Some methods of surveying include interviewing someone via a phone call, interviewing a focus group, having a personal interview with an individual or via an online poll. For the survey method, the response rate for the survey is crucial, as a low response rate can destroy the validity of any conclusion resulting from the statistical analysis.

<u>Secondary Data</u>

1. Published data

Published data refers to data from ==secondary sources==, and the data is readily available. This method of collecting data is preferred due to its ==convenience, relatively low cost, and its reliability== (assuming its been collected by a reputable organization). However when using secondary data, care needs to be taken as errors may have been introduced as a result of a false transcription or a misinterpretation.

**Sampling**

Sampling refers to ==examining a sample rather than a population==. The main reason for examining a sample rather than a population is due to ==cost and practicability.==

==Statistical inference== permits us to draw conclusions about a population parameter based on a sample that is quite small in comparison to the population.

When sampling, it is essential that we identify a ==target population== and a ==sample population==

1. Target population – refers to the population about which we want to draw inferences

2. Sampled population – refers to the actual population from which the sample has been drawn