

Chapter 1 – Introduction to Statistics

1.1 Basic Statistical Concepts

- A population is a collection of objects (often called units or subjects) of interest. Collection of data on a whole population is called a census.
- A sample is a subset of the units in a population and can be expected to be representative of the whole population. Sampling is cheaper and simpler.

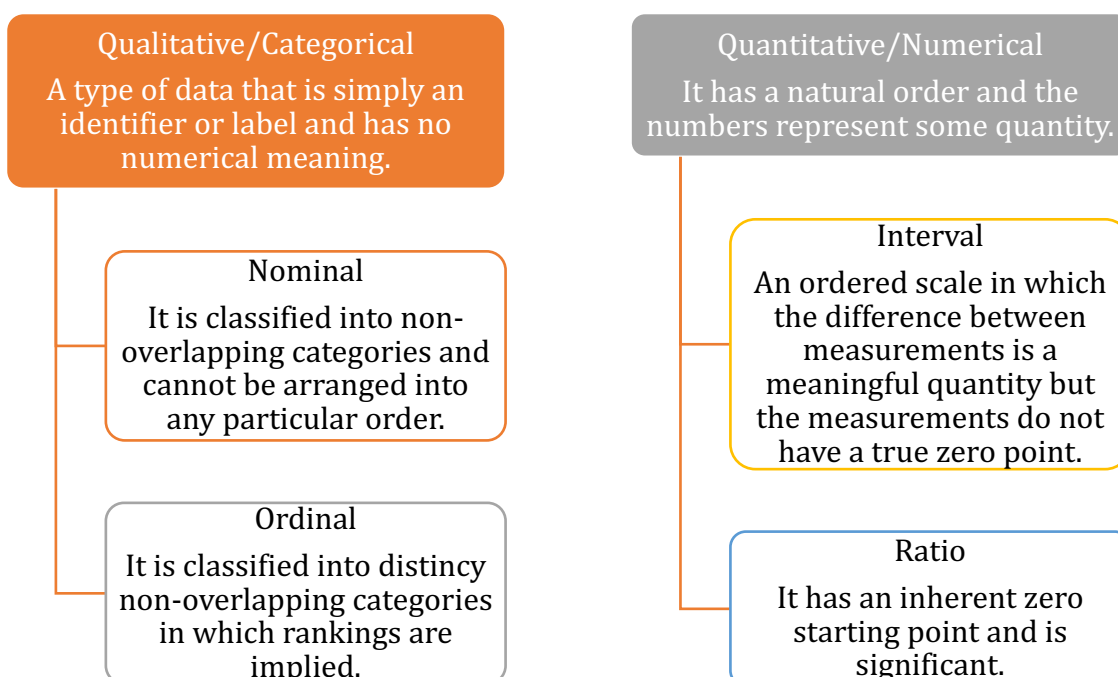
Steps to Analysing Data from a Sample

1. Exploratory Data Analysis (EDA) in which numerical, tabular and graphical summaries of data are produced to summaries and highlight key aspects or features of the data.
2. Statistical Inference uses sample data to reach conclusions about the population from which the sample is drawn. This is usually the main aim of any statistical exercise and involves more formal data analysis techniques. An inference is a conclusion. That patterns observed in the data sample are present in the wider population and clearly assumes that the sample is representative of the population.

A descriptive measure of the population is called a parameter which is denoted by Greek letters. Examples of parameters are population mean (μ), population standard deviation (σ) and population variance (σ^2). A descriptive measure of a sample is called a statistic which is denoted by roman letters. Examples of statistics are sample mean (\bar{x}), sample standard deviation (s) and sample variance (s^2).

Distinction between parameter and statistic is important since a business researcher often wants to estimate the value of a parameter or draw inferences about the parameter. However, the calculation of parameters is impossible and infeasible in the given time and money and therefore business researchers take sample to represent the population.

1.2 Types of Data It can be argued that less than 5 points scales are considered ordinal, not interval.



Cross sectional data are collected at a fixed point in time. They give a snapshot of measured variables at that point in time. **Time Series** data are collected over time.

1.3 Obtaining Data – Data collected to address a specific need are known as **primary data**. Data that were collected for some other purpose and are already available are called **secondary data**.

1.5 When Things Go Wrong – Analysing data correctly is important since incorrect data can have disastrous consequences.

Chapter 2 – Charts and Graphs

2.1 Frequency Distributions

Raw data or data that has not be summarised is referred to as ungrouped data which can be difficult to interpret and obtain insights from. Frequency distributions group continuous data and is a summary of the data presented as non-overlapping class intervals covering the entire range of data and their corresponding frequencies.

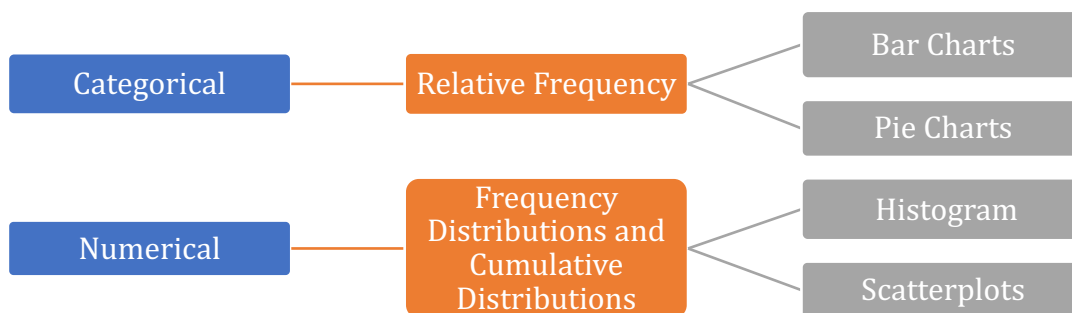
Range – The range is defined as the difference between the smallest and largest data values.

Class Midpoint – This is calculated by taking the midpoint of the class interval.

Relative Frequency – This is the ratio of the frequency of the class interval to the total frequency and gives the proportion of the total data that lie within the class interval.

Cumulative Frequency – This is the running total of frequencies through the classes of a frequency distribution and is found by adding the frequency of that interval to the cumulative frequency of the previous class interval.

2.2 Graphical Display of Data



Histograms – They are the most useful and common graphs for displaying continuous data. It can show the shape of distribution, spread or variability, central location of data and any outliers that appear outside the main body of observations.

