

# TOPIC: 1

## Introduction to statistics and frequency distributions

**Population:** The entire set of individuals or events that we are interested for our study. For ex. We are interested to find average height of Deakin university student, and in this case all students from university are our population group.

**Sample:** To do our study feasible it's very hard to reach entire population due to cost or availability or interest from population. In this case we make a subset from this population that called a sample.

**Representative Sample:** A sample that shares key characteristics of the population from which it has been taken. Eg. 20 psych students from Deakin University.

**Parameter:** A value that explain the key characteristics of population. E.g. Height of tallest or small person.

**Statistic:** A value that describes a key characteristic of the sample. Statistics are used to estimate values that exist in the population. Ex average height of students is between 165cm to 185cm. There are two main types of statistics:

1. **Descriptive statistics** are used first, when we want to describe the data. This might consist of calculating means (averages) and examining the data for extreme scores. Frequency tables and graphs are a way of organizing and simplifying the data we have collected. They provide two key pieces of information. The set of scores or range of categories that people could have either obtained or fallen into on the value of interest.
2. **Inferential statistics** are generally used when we want to answer the research questions. It allows us to make generalization from the sample to the population. Thus, in answering this we would be **inferring (conclude)** something about the population based on the results obtained with our sample.

**Sampling Error:** There will almost always be a discrepancy between your sample characteristics and the population characteristics due to sampling error (particularly in small samples). For eg. Average height of people would be 165cm to 185cm. so discrepancy in numbers is due to sampling error.

**Variable** is something about a construct/object/event that can take on different values (i.e. it can vary). Other variables can be gender, age, socioeconomic status, motivation and even personality.

**Discrete variables** have only a limited number of values (e.g. gender) whilst **continuous variables** can take on many different values (e.g. age, anxiety score). Scores on variables, such as anxiety, self-esteem, extraversion, etc.

**Independent variables:** are those that are manipulated by the researcher. In our height example, the researcher assigned participants as male or female. So in this example sex would be IV which we can manipulate.

**Dependent variable:** The variable we can actually measure and influence by independent variable. In our height example height will be our DV because that's what we are measuring.

**Measurement scale:** We measure variables using a variety of measurement scales. The type of measurement scale depends on the data we are dealing with. There are 4 main types of measurement scales that you will need to know as **Nominal, Ordinal, Interval, and Ratio.**

**Nominal scale:** These scales are simply categories with different names and have no underlying scale and order. For example, Religion is usually measured on a nominal scale like Christian, Muslim, Jewish, And Other. (1, 2, 3, 4). We can't say that one category is larger/higher/more than another. They are just labels for the different categories.

**Ordinal scale:** Ordinal scales are categories with different names AND are organised into an ordered sequence. Example: Degree of illness – None, Mild, Moderate, Severe. This allows us to determine the direction of the difference. We can say that severe is greater than moderate; and moderate is greater than mild. But distances between the categories is unknown.