STATISTICS 12

# Definitions and Introduction

- **Statistics** is not about numbers, it is about **data. Statistics** is a science of collecting, analyzing, and drawing conclusions from data.
- **Data** = numbers in context.
- **Sample** is a subset of the population selected for study in some prescribed manner.
- **Population=** the entire collection of individuals or objects about which information is desired

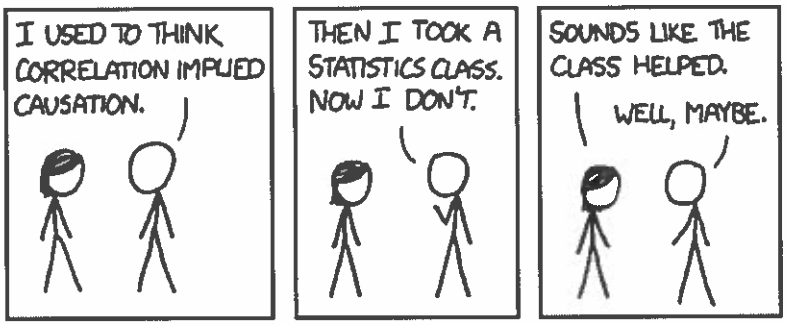
## Three reasons to study statistics

1. Being informed.
2. Making informed judgments.
3. Evaluating decisions that affect your life.

## Statistics and the Data Analysis Process

I. Descriptive Statistics = methods for organizing and summarizing data = tables, graphs, and numerical summaries. The branch of statistics that also interprets data.

II. Inferential Statistics = is the branch of statistics that involves generalizing from a sample to the population from which the sample was selected and assessing the reliability of such generalization. It also quantifies the chance of incorrect decision (probability) and describes variability that affects how we collect, summarize and draw conclusions.




## The Data Analysis Process

1. Understanding the nature of the problem.
2. Deciding what to measure and how to measure it.
3. Data collection.
4. Data summarization and preliminary analysis.
5. Formal data analysis.
6. Interpretation of results.

## Describing Data

**A variable** is any characteristics whose value may change from one individual or object to another (humans, animals, plants, objects, measurements).

**Data** result from making observation either on a single variable or simultaneously on two or more variables.

- A data set consisting of observations on a single variable attribute is a **univariate data set.**
- A data set is **multivariate** is a data set consisting of multiple variables observed or measured simultaneously.

- A univariate data set is **categorical** (or **qualitative**) if the individual observations are categorical responses (brand names, colour, yes/no). It places individuals or objects into one of several groups or categories.

- A univariate data set is **numerical** (or **quantitative**) is each observation is a number (a result of counting or measuring).

**Numerical Data**

There are two types of numerical data:

I. Discrete

II. Continuous

- A numeric variable results in discrete data if the possible values or the variable corresponds to isolated points on the number line (usually from counting).

- A numerical variable results in continuous data if the set of possible values forms an entire interval on the number line.

# Frequency Distributions and Bar Charts for Categorical Data

$$Relative\ frequency = \frac{frequency}{number\ of\ observations\ in\ the\ data\ set}$$

= an appropriate graphical or tabular display to effectively summarize and communicate information.

**A Frequency Distribution for Categorical Data**

A frequency distribution for categorical data is a table that displays the possible categories along with the associated frequencies and/or relative frequencies.

- The frequency for a particular category is the number of times the category appears in the data set.
- The relative frequency for a particular category is the fraction or proportion of the observations resulting the category. It is calculated as:

- If the table includes relative frequencies, it is sometimes referred to as a **relative frequency distribution.**