# Stat 170 – Introductory Statistics

## Lecture 1 – Introduction to Statistics

### Intro

Statistics is the learning from data. This involves collecting presenting, analysing and interpreting data.

### Objectives

The objective is to obtain information about a target population using a sample.

Target Population: The relevant subjects of interest.

Sample: Manageable group that represents the target population.

### Research Questions

First find the target population and take a sample that represents the population. This sample is used to answer the research questions about the target population.

### A Data Set

A data set is a set of variables that can be categorised as numerical or categorical. E.g.; Location, life expectancy, infant mortality rate, unemployment.

### Data Classification

1. Categorical (Observed in groups):
    a. Nominal – There is no set order. E.g. Sex, Colour
    b. Ordinal – There is an order set. E.g. Age group, Size
2. Numerical (Observed as integers):
    a. Discrete – When you can count exact. E.g. Students in a class
    b. Continuous – Is always changing and can be rounded. E.g. Height, weights

## Graphing Data

### Displaying Data

- List of Charts/Graphs
    - Clustered Bar Chart [2 Variables (Categorical & Categorical)]
    - Comparative Box Plot [2 Variables (Categorical & Numerical)]
    - Scatter Plots [2 Variables (Numerical & Numerical)]
    - Bar Chart/Pie Chart [1 Variable (Categorical)]
    - Histograms [1 Variable (Numerical)]

| Data | Categorical | Numerical | |
|---|---|---|---|
| Categorical | Clustered Bar Charts | Comparative Box Plot | Bar Chart/Pie Chart |
| Numerical | Comparative Box Plot | Scatter Plots | Histograms |
| | Bar Chart/Pie Chart | Histograms | |

### Graphs Need

1. A title
2. Clearly Labelled Axes
3. An Explanatory Comment
4. To be clear and uncluttered

# Lecture 2 – Numerical Summaries`

## Population Distributions

### Definitions

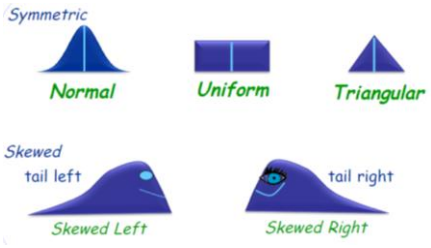
- Measure of centre: where the 'middle' of the data is
- Measure of spread: tells us how spread out the histogram is

### Distribution

Numerical Data: Centre, Spread and Shape of Histogram

The centre is a nice summary but does not bring all the answers to the data.

The shape of the distributions can be in many shapes: symmetric, skewed, bimodal or multimodal




A normal distribution is the shape of a bell curve.

## Sample Statistics and Population Parameters

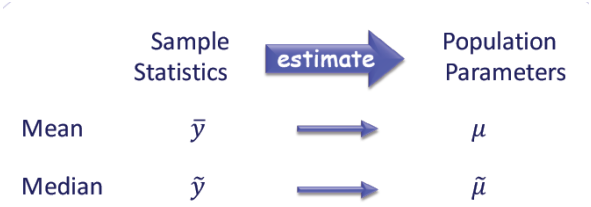
Using sample statistics, we can estimate population parameters

### Summarising Numerical Data

2 major variables to measure the centre is median and the mean.

The median: Is the measure if the direct centre of the set of data. Also, known as the middle value of the data set.

The mean: Is the arithmetic average of the data set




The spread of a data set expresses the inter quartile range (iqr) which is the difference between the upper and lower quartiles giving a rough estimate of the middle 50% of the data. However, the range is the difference from the largest outlier to the smallest outlier, however this can be misleading data due to random outliers.

Standard Deviation

The standard deviation is how much does the data deviate from the data from the mean. These numbers are called residuals.