

# FUNDAMENTAL STATISTICS.

## 1. Introduction

We can summarize an important feature of a set of data using just one number i.e. a *statistic*. Basic summary statistics can be used to gain a fundamental knowledge of the central tendency, most frequently occurring or spread of the data. Such information are referred to as summary statistics and are simple to either calculate by hand or in any basic statistical software.

## 2. Summary Statistics

Recalling that all data can be classified as either, **Categorical** or **Quantitative** there exist different ways to summarise each type of data; for example an average measure of eye colour is nonsensical.

- **Mean** A measure of central tendency for Quantitative data i.e. the long term average value.
- **Median** A measure of central tendency for Quantitative data i.e. the half-way point.
- **Mode** The most frequently occurring (discrete), or where the probability density function peaks (continuous).
- **Minimum** The smallest value.
- **Maximum** The largest value.
- **Inter quartile range** Can be thought of as the *middle 50* of the (Quantitative) data, used as a measure of spread.
- **Variance** - Used as a measure of spread, may be thought of as the *moment of inertia*.
- **Standard deviation** - A measure of spread, the square root of the variance.

## 3. Data shape

Continuous data can be;

- **Symmetrical** Perfectly symmetrical about the mean, i.e. **Normally distributed** data.
- **Skewed** Right/Positively skewed i.e. long tail to the right of the *peak*, Left/Negatively skewed i.e. long tail to the left of the *peak*.
- **Multimodal** Multiple *peaks* in the data.

In addition when dealing with continuous data the shape of the data effects which summary statistic are most appropriate to use i.e heavily skewed data will pull the mean toward the tail of the data.

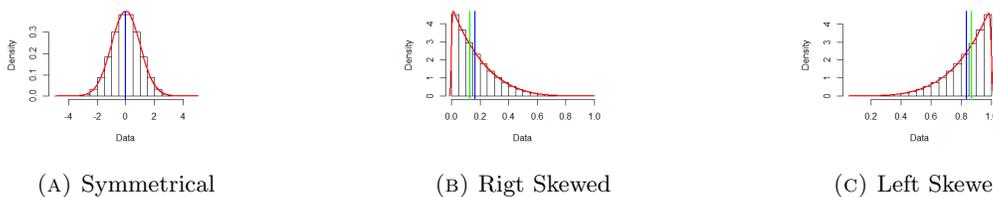


FIGURE 1. Displaying different data shapes with **median**, **mean**, **density** respectively.