

Terminology

Sample - A group of objects, individuals, or values selected from a population. The idea is for this sample to provide estimates for the population.

Sample Size - The number of objects, individuals, or values in a sample.

Survey - A systematic collection of data taken by questioning a sample of people.

Census - A study that attempts to measure every unit in a population.

Experiment - A process or study that results in the collection of data, the outcome of which is unknown.

Measurement (Continuous Numerical) Data - Data that can take any value in an interval of numbers. Eg: Height, Weight.

Counting (Discrete Numerical) Data - Data that can take only distinct values, usually whole numbers. Eg: number of left handed people in the class.

Group or Category (Categorical) Data - Values can be organised into groups. These groups (or categories) must be chosen so that they do not overlap and that every value belongs to one and only one group, and there should be no doubt as to which one. Eg: Eye colour

Time Series Data - A data set gathered over time. For one object, such as climate in Auckland, the values of a variable (or several variables) are obtained over time. Usually there are equal intervals between the times. Eg: Temperature, attendance in class.

Symmetry - When numerical data is distributed the same way either side of the centre.

Skew - A lack of symmetry in a distribution of a numerical distribution in which the values are stretched out in one direction.

Average - A number that is representative or typical of the centre of a set of numerical values. In this sense, the number used could be the mean or the median or mode.

Mean - Calculated by adding all the values together and dividing by the number of values.

Median - The middle number when a set of data is ordered numerically.

Mode - A value in a set of numerical data that occurs most often.

Lower Quartile - A number that is a quarter of the way through the data when it is ordered, from the lower end.

Upper Quartile - A number that is a quarter of the way through the data when it is ordered, from the upper end.

Minimum - The smallest number when a set of data is ordered numerically.

Maximum - The largest number when a set of data is ordered numerically.

Outlier - A data point whose values for the variable we are looking at are much bigger or smaller than most of the other data points. This can be either on a dot plot or a scatter graph