Glossary

Box-and-whisker plot

A graph used to depict a range of data that shows a line running from the minimum to the first quartile, a box from the first quartile to the median, another box from the median to the third quartile, and a line running from the third quartile to the maximum.

Case

An object in a collection of data (e.g. if the data are a collection of shells, a case is an individual shell).

Categorical variable

A variable that can take on a limited, or fixed, number of possible values (e.g. a ball's color would be a categorical variable if the ball can only be red, blue, green, and purple).

Central tendency

The measurable clustering of values in a statistical distribution.

Cluster sampling

A sampling technique in which the population is divided into smaller groups, or clusters, and a simple random sample is taken of the clusters. This is different from a stratified random sample because the whole cluster is sampled, rather than a sample from within the cluster.

Confound

An extraneous variable that could also cause a correlation between the dependent variable and the independent variable (e.g. shark attacks are more common when beach ice cream sales increase. Is this because ice cream causes shark attacks, or because hot weather leads to both increased ice cream

consumption and increased swimming? In this case, hot weather is the confound).

Convenience sample

A sample made up of the portion of the population that is easiest to reach (e.g. a poll on the O'Reilly Factor website, where Bill O'Reilly asks all of his listeners to go to his website and answer the poll).

Data validation

Verifying that each datum in a set of data is an accepted value (e.g. checking to make sure that all answers to "how many children do you have" are non-negative integers because you cannot have -5 children, or half a child, etc.)

Dependent variable

The variable in an experiment that is being measured (e.g. in an experimental drug trial, the dependent variable is the effect of the drug because experimenters are measuring the effects of the drug).

First quartile

The middle value in a set of data between the minimum and the median (e.g. if the data is {1,2,3,4,5}, then 2 is the first quartile because it is between 1 and 3).

Grouped bar graph

Charts designed to show different sub-groups within a category (e.g. in a bar chart of the different types of recycling used in locations around the city, the categories would be the locations in the city, and the sub-groups would be the types of recycling used). It is a bar chart in which each bar is a histogram of the different sub-groups of that bar.

Histogram

A diagram that uses rectangles to show the frequency of data values within successive numerical intervals.

Independent variable

The variable in an experiment being adjusted by the experimenter to measure its effects on the dependent variable (e.g. in an experimental drug trial, the drug is the independent variable because experimenters are controlling how much is taken).

Inference

A claim about the properties of a population based on statistical analysis.

Interquartile range (IQR)

The range of values between the first and third quartiles.

Mean

The central value of data, calculated by adding up all the values and dividing the sum by the number of different values (e.g. {1, 4, 5}, the mean is 3.33).

Median

The central value of data, calculated by taking the middle value of the data (e.g. {1,4,5} the median is 4).

Multistage sampling

A sampling method in which sampling is carried out in stages (e.g. taking a cluster sample of the elements chosen by a preliminary cluster sample).

Pie graph

A circular chart that is divided into slices or radiuses, in which the angle created by each radius is proportional to the percentage of the population that it represents.

Pivot table

A data summarization tool that automatically sorts and gives summary statistics for data within a table.

Population

The pool from which a statistical sample is drawn and about which inferences are made.

Quantitative variable

A variable that can be measured in terms of numbers (e.g. the age of children in a household).

Random sample

A sample from a population in which members are chosen randomly.

Sample

A set of data selected from a population, from which inferences are being drawn.

Sensitivity testing

Sensitivity is the proportion of correctly identified positive results in a test (for example, if 100 people are known to have AIDS and 96 test positive, then the test has 96% sensitivity). Sensitivity testing is the measuring of the sensitivity of a test.

Slicer

A quick method of filtering data on a pivot table by the type of data needed (e.g. on a pivot table showing the different types of vegetable oils, a slicer would narrow it down to olive oil).

Spread

A measure of how much variation exists within a sample.

Stacked column graph

A bar graph in which each bar is a tower of smaller bars, in which the height of each smaller bar represents the proportion of the total bar that the smaller bar represents (a combination between a bar graph and a pie chart).

Stratified random sample

A sample in which the population is divided into homogeneous clusters, or stratums, and a random sample is taken from within each stratum. This is different from a cluster sample because a random sample is taken from within each stratum, rather than of a whole stratum.

Summary statistics

Numbers derived from a set of data that give a quick description, or summary, of the data.

Systematic sampling

A sampling method in which members of the sample are selected from the population at a specific interval (e.g. going through a phone book and selecting the first, fourth, seventh, tenth, and so forth, names for the sample).

Third quartile

The middle value between the median and the maximum (e.g. in a set of {1,2,3,4,5}, the third quartile is 4 because it is between 3 and 5).

Two-way table

A table showing the distribution of one variable in rows and another in columns, used to visualize the association between the two variables.

Type

A specific instance within a categorical variable (e.g. for the categorical variable automobiles, one type could be minivan).