eCornell

Glossary

Aggregate

When data have been combined at a level that causes a loss of detail.

Alpha

A numerical value between 0 and 1 that divides the non-rejection region from rejection region.

Alternative Hypothesis

In hypothesis testing, the alternative hypothesis is the statement you are trying to support by rejecting the null hypothesis.

Analysis

Examination of the elements or structure of something, typically as a basis for discussion or interpretation.

Binomial Distribution

A type of probability distribution, used when outcomes are 1 of 2 options (true or false, win or lose, etc.).

Categorical Variable

See "Qualitative Variable."

Chance

The likelihood or probability of something happening, a chance event happens with some level of uncertainty.

1

Chi-squared (test)

Refers to a probability distribution and also to a statistical test used to test relationships in two-way cross-tabular data tables.

Combination

A way of selecting items from a collection such that the order of selection does not matter. In smaller cases it is possible to count the number of combinations.

Confidence Interval

An estimate of the population parameter in the form of an interval with a lower and upper limit.

Critical Level

See Alpha

Data

A collection of observations of one or more variables.

Data Vector

Collection of observations of one variable; in a spreadsheet, a data vector is organized in one column or one row.

Degree of Freedom

The number of values that are free to vary.

Descriptive Statistics

Summary measures of a sample including average, variance, standard deviation, minimum, maximum, and range.

Experiment

A process that uses controlled conditions to study the effect on a variable of interest by varying the values of one or more other variables(s).

Goodness of Fit test

See Chi-squared.

Hypothesis

A testable statement.

Hypothesis Test

A method used to make inferences about a population parameter using sample statistics.

Inference

The practice of using sample statistics to make claims about a population.

Interval

A range of values with an upper and lower limit.

Left-tail Test

A statistical test used to calculate a p-value when the alternative hypothesis you're trying to support has a value greater than the null. Use a left-tail test when the alternative hypothesis includes a less-than operator.

Mean

The sum of all observations divided by the number of observations; a measure of a typical outcome.

Normal Distribution

A symmetric distribution which is used to test hypotheses and form confidence intervals for population means and proportions.

Null Hypothesis

A statement that sets a population parameter equal to a specific value, or the statement that no difference exists between parameters of more than one population.

One-tail Test

A statistical test used to calculate a p-value when the alternative hypothesis is that a population parameter is either bigger or smaller than a set number, but not both. One-tail tests are also used when the alternative hypothesis indicates the difference between population parameters is either bigger than a set number or smaller than a set number.

p-value

The probability of observing a test statistic at least as extreme as the one calculated based on your sample and null hypothesis.

Parameter

A measure of the population.

Pooled Sample

When summary statistics of two or more samples are combined into one composite value.

Population

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

Practical Significance

The generalization of statistical results into practical or managerially relevant insights.

Proportion

The ratio or fraction of one outcome out of all possible outcomes.

Qualitative Variable

A variable with no natural sense of ordering; qualitative variables may be coded with numbers, but the numbers are meaningless.

Quantitative Variable

A variable whose range is a measurable numeric scale.

Range

The difference between the largest and smallest value of a quantitative variable.

Right-tail Test

A statistical test used to calculate a p-value when the alternative hypothesis you're trying to support has a value less than the null. Use a right-tail test when the alternative hypothesis includes a greater-than operator.

Sample

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

Sampling

The process by which members of a population are selected to form a sample.

Sampling Distribution

A mathematical representation of the probabilities of outcomes for a sample statistic (such as a mean) for all possible samples of size n.

Simple Point Estimate

A non-interval estimate.

Standard Deviation

A measure of uncertainty for a quantitative variable; standard deviation has the same units as the variable.

Standard Error

A measure of uncertainty for a population parameter. A standard error is calculated by dividing the standard deviation by the square root of the sample size.

Statistics

Methods for processing and analyzing data to support rational decision-making

Statistic

A numerical measure that describes a characteristic of a sample.

Statistical Significance

A level of certainty that a relationship or difference exists. See also: alpha.

Symmetric Distribution

A probability distribution whose shape to the left of the maximum value (peak) is a mirror image of the shape to the right of the peak.

t-distribution

A distribution used to develop confidence intervals and test hypotheses about the mean of a population.

Test Statistic

The statistic used to determine whether or not to reject the null.

Treatment

The process or intervention evaluated in an experiment.

Trial

An experimental event.

Two-tail Test

A statistical test used to calculate a p-value when the alternative hypothesis is that a population parameter is different from a set number. Use a two-tail test when the alternative hypothesis includes a not-equal-to operator.

Type I Error

A false positive. The incorrect rejection of a null hypothesis. Type I errors can be minimized by setting alpha or the critical value lower.

Type II Error

A false negative. The failure to reject a null when in fact it is false. Type II errors can be minimized by improved sampling, such as larger and/or more representative samples.

Variance

A measure of uncertainty for a quantitative variable. Variance is expressed in units that are the variable units squared.