

Glossary

Alternative Hypothesis | A statistical hypothesis that states that there is a difference between a parameter and a specific value, or that there is a difference between two parameters.

Blind Study | The individual does not know which treatment they are getting or if they are getting the treatment or a placebo.

Class Boundary | The halfway points that separate the class limits.

Class Limit | The smallest and largest value that could go into a class.

Cluster Sample | The populations are split into groups called clusters, then one or more clusters are randomly selected and all individuals in the chosen clusters are sampled.

Coefficient of Determination | The coefficient of determination is the fraction (or percent) of the variation in the values of y that is explained by the least-squares regression of y on x .

Coefficient of Variation | The standard deviation divided by the mean, which allows you to compare variability among data sets when the units or scale is different.

Combination Rule | The number of ways to select r out of n objects where order does not make a difference.

Completely Randomized Experiment | In this experiment, individuals are randomly placed into the two or more groups. One group gets either nothing or a placebo; this group is called the control group. The groups getting the treatment are called the treatment groups.

Confidence Interval | A range of potential values for the unknown population parameter.

Continuous Random Variable | A variable that has an infinite number of possible values in an interval of numbers.

Continuous Variable | Can take on any value. Continuous variables are usually things you measure.

Convenience Sample | Picking a sample that is conveniently at hand. For example, asking other students in your statistics course or using social media. Most convenience sampling will give biased views and is not encouraged.

Correlation Coefficient | A measure of the strength and direction of the linear relationship between two variables.

Critical Value | The number of standard errors added and subtracted in order to achieve a desired confidence level.

Cross-sectional Study | Data observed, measured, or collected at one point in time.

Descriptive statistics | The process where you collect, organize and describe data.

Discrete Random Variable | A variable that has a countable number of possible values.

Discrete Variable | Can only take on particular values like integers. Discrete variables are usually things you count.

Double-blind Study | Neither the individual nor the researcher knows who is getting which treatment or who is getting the treatment and who is getting the placebo.

Event | A set of certain outcomes of an experiment that you want to have happen.

Experiment | An activity or process that has specific results that can be repeated indefinitely which has a set of well-defined outcomes.

Extrapolation | The use of a regression line for prediction far outside the range of values of the independent variable x .

Factorial Design | This design has two or more independent categorical variables called factors. Each factor has two or more different treatment levels.

Factorial Rule | The number of ways to arrange n objects is $n!$

Five-number Summary | Minimum, first quartile, second quartile, third quartile and maximum value in a data set.

Fundamental Counting Rule | The number of ways to do task 1, 2, ..., n together would be to multiply the number of ways for each task: $m_1 \cdot m_2 \cdot \dots \cdot m_n$.

Hypothesis Testing | The scientific method used to evaluate claims about population parameters.

Independent Events | Two events that are not related and the probability of one event does not affect the probability of the other event.

Individual | A person or object that you are interested in finding out information about.

Inferential statistics | The process of using sample data to make inferences about the population.

Influential point | A point in a scatter plot that is positioned far away from the main cluster of data points on the y -axis.

Interquartile Range (IQR) | The distance between the first and third quartile.

Intersection | Where two events overlap and happen at the same time.

Interval data | Numeric data where there is a known difference between values, but zero does not mean "nothing."

Law of Large Numbers | As n increases, the relative frequency tends toward the theoretical probability.

Leverage Point | A point in a scatter plot that is positioned far away from the main cluster of data points on the x -axis.

Longitudinal Study | Data collected in the future from groups sharing common factors.

Lurking Variable | A variable other than the independent or dependent variables that may influence the regression equation

Margin of Error | Half the width of the confidence interval and is a statistic that represents the amount of random sampling error.

Matched Pairs Design | This is a subset of the randomized block design where the treatments are given to two groups that can be matched up with each other in some way.

Mean | The mean is the arithmetic average of the numbers. This is the center that most people call the average.

Median | The median is the data value in the middle of the data that has 50% of the data below that point and 50% of the data above that point.

Midrange | The average of the smallest and largest data values.

Modal Class | The modal class is the class with the highest frequency.

Mode | The mode is the data value that occurs the most frequently in the data.

Mutually Exclusive Events | Disjoint events that cannot occur at the same time.

Nominal data | Categorical data that has no order or rank. For example, the color of your car, ethnicity, race, or gender.

Null Hypothesis | A statistical hypothesis that states that there is no difference between a parameter and a specific value, or that there is no difference between two parameters.

Observational Study | An observational study is when the investigator collects data by observing, measuring, counting, watching or asking questions. The investigator does not change anything.

Ordinal data | Categorical data that has a natural order to it.

Outcomes | The results of an experiment.

Outlier | An outlier is a data value that is very different from the rest of the data and is very far from the center.

Parameter | Any characteristic or measure from a population. This number is a fixed, unknown number that you want to estimate.

Pareto Chart | A bar graph that is ordered from the most frequent down to the least frequent category.

Permutation Rule | The number of ways to arrange r out of n objects where order does make a difference.

Point Estimate | A sample statistic used to estimate the population parameter.

Population | The total set of all the observations that are the subject of a study.

Prediction Interval | The confidence interval for the predicted value of y .

Probability Distribution | An assignment of probabilities to all the possible values of the random variable.

Qualitative or Categorical Variable | A word or name that describes a quality of the individual

Quantitative or Numerical Variable | A number (quantity), something that can be counted or measured from the individual.

Quartile | Numbers that divide a data set into fourths. One fourth (or a quarter) of the data falls between consecutive quartiles.

Random Variable | If you have a variable, and can find a probability associated with that variable, it is called a random variable.

Randomized Block Design | A block is a group of subjects that are similar or the same subject measured multiple times, but the blocks differ from each other. Then randomly assign treatments to subjects inside each block.

Range | The difference between the largest and smallest data values.

Ratio Data | Numeric data that has a true zero, meaning when the variable is zero nothing is there. Most measurement data are ratio data.

Replication | Repetition of an experiment on more than one subject so you can make sure that the sample is large enough to distinguish true effects from random effects. It is also the ability for someone else to duplicate the results of the experiment.

Residual | The vertical distance between the actual value of y and the predicted value of y .

Retrospective Study | Data collected from the past using records, interviews, and other similar artifacts.

Sample | A subset from the population.

Sample Space | Collection of all possible outcomes of the experiment.

Scatterplot | A scatterplot shows the relationship between two quantitative variables measured on the same individuals.

Simple Random Sample (SRS) | Selecting a sample size of n objects from the population so that every sample of the same size n has equal probability of being selected as every other possible sample of the same size from that population.

Standard Deviation | The average distance of the data points from the mean.

Standard Error of Estimate | The standard deviation of the residuals.

Statistic | Any characteristic or measure from a sample.

Statistical Hypothesis | An educated conjecture about a population parameter.

Statistics | Statistics is the study of how to collect, organize, analyze, and interpret data collected from a group.

Stratified Sample | The population is split into groups called strata, then a random sample is taken from each stratum.

Systematic Sample | We list the entire population, then randomly pick a starting point n and then take every n^{th} value until the sample size is reached.

Type I Error | Rejecting the null hypothesis when it was actually true.

Type II Error | Failing to reject the null hypothesis when it is actually false.

Union | The junction of two events including their intersection.

Variable (also known as a random variable) | The measurement or observation of the individual.

Variance | The average squared distance from the data points to the mean.

Z-Score | The number of standard deviations a data point is from the mean.