

# Glossary for the Triola Statistics Series

**Absolute deviation** The measure of variation equal to the sum of the deviations of each value from the mean, divided by the number of values

**Acceptance sampling** Sampling items without replacement and rejecting the whole batch based on the number of defects obtained

**Actual odds against** The ratio  $P(\bar{A})/P(A)$ , usually expressed in the form of  $a:b$  (or “ $a$  to  $b$ ”)

**Actual odds in favor** The reciprocal of the actual odds against an event

**Addition rule** Rule for determining the probability that, on a single trial, either event  $A$  occurs, or event  $B$  occurs, or they both occur

**Adjusted coefficient of determination** Multiple coefficient of determination  $R^2$  modified to account for the number of variables and sample size

**Alpha ( $\alpha$ )** Symbol used to represent the probability of a type I error. *See also* Significance level.

**Alternative hypothesis** Statement that is equivalent to the negation of the null hypothesis; denoted by  $H_1$

**Analysis of variance** Method of analyzing population variances in order to test hypotheses about means of populations

**ANOVA** *See* Analysis of variance.

**Arithmetic mean** Sum of a set of values divided by the number of values; usually referred to as the mean

**Assignable variation** Type of variation in a process that results from causes that can be identified

**Attribute data** Data that can be separated into different categories distinguished by some nonnumeric characteristic

**Average** Any one of several measures designed to reveal the center of a collection of data

**Bar graph** Uses bars of equal width to show frequencies of categories of data

**Beta ( $\beta$ )** Symbol used to represent the probability of a type II error

**Bimodal** Having two modes

**Binomial experiment** Experiment with a fixed number of independent trials, where each outcome falls into exactly one of two categories

**Binomial probability formula** Expression used to calculate probabilities in a binomial experiment (see Formula 5-5 in Section 5-3)

**Bivariate data** Data arranged as matched pairs

**Bivariate normal distribution** Distribution of paired data in which, for any fixed value of one variable, the values of the other variable are normally distributed

**Blinding** Procedure used in experiments whereby the subject doesn't know whether he or she is receiving a treatment or a placebo

**Block** A group of subjects that are similar in the ways that might affect the outcome of an experiment

**Box-and-whisker diagram** See Boxplot.

**Boxplot** Graphical representation of the spread of a set of data

**Case-control study** Study in which data are collected from the past by going back in time (through examination of records, interviews, and so on).

**Categorical data** Data that can be separated into different categories that are distinguished by some nonnumeric characteristic

**Cell** Category used to separate qualitative (or attribute) data

**Census** Collection of data from every element in a population

**Centerline** Line used in a control chart to represent a central value of the characteristic measurements

**Central limit theorem** Theorem stating that sample means tend to be normally distributed with mean  $\mu$  and standard deviation  $\sigma/\sqrt{n}$

**Centroid** The point  $(\bar{x}, \bar{y})$  determined from a collection of bivariate data

**Chebyshev's theorem** Theorem that uses the standard deviation to provide information about the distribution of data

**Chi-square distribution** A continuous probability distribution (first introduced in Section 7-5)

**Class boundaries** Values obtained from a frequency distribution by increasing the upper class limits and decreasing the lower class limits by the same amount so that there are no gaps between consecutive classes

**Classical approach to probability** Approach in which the probability of an event is determined by dividing the number of ways the event can occur by the total number of possible outcomes

**Classical method of testing hypotheses** Method of testing hypotheses based on a comparison of the test statistic and critical values

**Class midpoint** In a class of a frequency distribution, the value midway between the lower class limit and the upper class limit

**Class width** The difference between two consecutive lower class limits in a frequency distribution

**Cluster sampling** Dividing the population area into sections (or clusters), then randomly selecting a few of those sections, and then choosing *all* the members from those selected sections

**Coefficient of determination** Amount of the variation in  $y$  that is explained by the regression line

**Coefficient of variation (or CV)** The ratio of the standard deviation to the mean, expressed as a percent

**Cohort study** Study of subjects in identified groups sharing common factors (called *cohorts*), with data collected in the future

**Combinations rule** Rule for determining the number of different combinations of selected items

**Complement of an event** All outcomes in which the original event does not occur

**Completely randomized design** Procedure in an experiment whereby each element is given the same chance of belonging to the different categories or treatments

**Compound event** Combination of simple events

**Conditional probability** The probability of an event, given that some other event has already occurred

**Confidence coefficient** Probability that a population parameter is contained within a particular confidence interval; also called confidence level or degree of confidence

**Confidence interval** Range of values used to estimate some population parameter with a specific confidence level; also called an interval estimate

**Confidence interval limits** Two numbers that are used as the high and low boundaries of a confidence interval

**Confidence level** Probability that a population parameter is contained within a particular confidence interval

**Confounding** A situation that occurs when the effects from two or more variables cannot be distinguished from each other

**Contingency table** Table of observed frequencies where the rows correspond to one variable of classification and the columns correspond to another variable of classification; also called a two-way table

**Continuity correction** Adjustment made when a discrete random variable is being approximated by a continuous random variable (Section 6-6)

**Continuous data** Data resulting from infinitely many possible values that correspond to some continuous scale that covers a range of values without gaps, interruptions, or jumps

**Continuous random variable** A random variable with infinite values that can be associated with points on a continuous line interval

**Control chart** Any one of several types of charts (Chapter 14) depicting some characteristic of a process in order to determine whether there is statistical stability

**Control group** A group of subjects in an experiment who are not given a particular treatment

**Control limit** Boundary used in a control chart for identifying unusual points

**Convenience sampling** Sampling in which data are selected because they are readily available

**Correlation** Statistical association between two variables

**Correlation coefficient** Measurement of the strength of the relationship between two variables

**Critical region** The set of all values of the test statistic that would cause rejection of the null hypothesis

**Critical value** Value separating the critical region from the values of the test statistic that would not lead to rejection of the null hypothesis

**Cross-sectional study** Study in which data are observed, measured, and collected at one point in time

**Cumulative frequency** Sum of the frequencies for a class and all preceding classes

**Cumulative frequency distribution** Frequency distribution in which each class and frequency represents cumulative data up to and including that class

**Data** Numbers or information describing some characteristic

**Degree of confidence** Probability that a population parameter is contained within a particular confidence interval; also called level of confidence

**Degrees of freedom** Number of values that are free to vary after certain restrictions have been imposed on all values

**Denominator degrees of freedom** Degrees of freedom corresponding to the denominator of the  $F$  test statistic

**Density curve** Graph of a continuous probability distribution

**Dependent events** Events for which the occurrence of any one event affects the probabilities of the occurrences of the other events

**Dependent sample** Sample whose values are related to the values in another sample

**Dependent variable**  $y$  variable in a regression or multiple regression equation

**Descriptive statistics** Methods used to summarize the key characteristics of known data

**Deviation** Amount of difference between a value and the mean; expressed as  $x - \bar{x}$

**Dichotomous variable** Variable which has two possible discrete values

**Discordant pairs** Pairs of categories in which the two categories are different; used in McNemar's test

**Discrete data** Data with the property that the number of possible values is either a finite number or a "countable" number, which results in 0 possibilities, or 1 possibility, or 2 possibilities, and so on

**Discrete random variable** Random variable with either a finite number of values or a countable number of values

**Disjoint events** Events that cannot occur simultaneously

**Distribution-free tests** Tests not requiring a particular distribution, such as the normal distribution. *See also* Nonparametric tests.

**Dotplot** Graph in which each data value is plotted as a point (or dot) along a scale of values

**Double-blind** Procedure used in an experiment whereby the subject doesn't know whether he or she is receiving a treatment or placebo, and the person administering the treatment also does not know

**Dummy variable** A dichotomous variable with the two possible values of 0 and 1. Used in multiple regression

**Efficiency** Measure of the sensitivity of a nonparametric test in comparison to a corresponding parametric test

**Empirical rule** Rule that uses standard deviation to provide information about data with a bell-shaped distribution (Section 3-3)

**Estimate** Specific value or range of values used to approximate some population parameter

**Estimator** Sample statistic (such as the sample mean  $\bar{x}$ , used to approximate a population parameter

**Event** Result or outcome of an experiment

**Expected frequency** Theoretical frequency for a cell of a contingency table or multinomial table

**Expected value** For a discrete random variable, the mean value of the outcomes

**Experiment** Application of some treatment followed by observation of its effects on the subjects

**Experimental units** Subjects in an experiment

**Explained deviation** For one pair of values in a collection of bivariate data, the difference between the predicted  $y$  value and the mean of the  $y$  values

**Explained variation** Sum of the squares of the explained deviations for all pairs of bivariate data in a sample

**Exploratory data analysis (EDA)** Branch of statistics emphasizing the investigation of data

**Factor** In analysis of variance, a property or characteristic that allows us to distinguish the different populations from one another

**Factorial rule** Rule stating that  $n$  different items can be arranged  $n!$  different ways

**F distribution** Continuous probability distribution first introduced in Section 9-5

**Finite population correction factor** Factor for correcting the standard error of the mean when a sample size exceeds 5% of the size of a finite population

**Five-number summary** Minimum value, maximum value, median, and the first and third quartiles of a set of data

**Fractiles** Numbers that partition data into parts that are approximately equal in size

**Frequency distribution** Listing of data values (either individually or by groups of intervals), along with their corresponding frequencies (or counts)

**Frequency polygon** Graphical representation of the distribution of data using connected straight-line segments

**Frequency distribution** List of categories of values along with their corresponding frequencies

**Frequency table** See frequency distribution

**Fundamental counting rule** Rule stating that, for a sequence of two events in which the first event can occur  $m$  ways and the second can occur  $n$  ways, the events together can occur a total of  $m \cdot n$  ways

**Goodness-of-fit test** Test for how well some observed frequency distribution fits some theoretical distribution

**Histogram** Graph of vertical bars representing the frequency distribution of a set of data

**H test** The nonparametric Kruskal-Wallis test

**Hypothesis** Statement or claim about some property of a population

**Hypothesis test** Method for testing claims made about populations; also called test of significance

**Independent events** Events for which the occurrence of any one of the events does not affect the probabilities of the occurrences of the other events

**Independent sample** Sample whose values are not related to the values in another sample

**Independent variable** The  $x$  variable in a regression equation, or one of the  $x$  variables in a multiple regression equation

**Inferential statistics** Methods involving the use of sample data to make generalizations or inferences about a population

**Influential point** Point that strongly affects the graph of a regression line

**Interaction** In two-way analysis of variance, the effect when one of the factors changes for different categories of the other factor

**Interquartile range** The difference between the first and third quartiles

**Interval** Level of measurement of data; characterizes data that can be arranged in order and for which differences between data values are meaningful

**Interval estimate** Range of values used to estimate some population parameter with a specific level of confidence; also called a confidence interval

**Kruskal-Wallis test** Nonparametric hypothesis test used to compare three or more independent samples; also called an  $H$  test

**Least-squares property** Property stating that, for a regression line, the sum of the squares of the vertical deviations of the sample points from the regression line is the smallest sum possible

**Left-tailed test** Hypothesis test in which the critical region is located in the extreme left area of the probability distribution

**Level of confidence** Probability that a population parameter is contained within a particular confidence interval; also called degree of confidence

**Linear correlation coefficient** Measure of the strength of the relationship between two variables

**Logistic regression** Method used in multiple regression when the dummy variable is the response (  $y$  ) variable

**Longitudinal study** Study of subjects in identified groups sharing common factors (called *cohorts*), with data collected in the future

**Lower class limits** Smallest numbers that can actually belong to the different classes in a frequency distribution

**Lower control limit** Boundary used in a control chart to separate points that are unusually low

**Lurking variable** Variable that affects the variables being studied, but is not itself included in the study

**Mann-Whitney  $U$  test** Hypothesis test equivalent to the Wilcoxon rank-sum test for two independent samples

**Marginal change** For variables related by a regression equation, the amount of change in the dependent variable when one of the independent variables changes by one unit and the other independent variables remain constant

**Margin of error** Maximum likely (with probability  $1 - \alpha$ ) difference between the observed sample statistic and the true value of the population parameter

**Matched pairs** With two samples, there is some relationship so that each value in one sample is paired with a corresponding value in the other sample

**Mathematical model** Mathematical function that “fits” or describes real-world data

**Maximum error of estimate** See Margin of error.

**McNemar’s test** Uses frequency counts from matched pairs of nominal data from two categories to test the null hypothesis that the frequencies from discordant pairs occur in the same proportion

**Mean** The sum of a set of values divided by the number of values

**Mean absolute deviation** Measure of variation equal to the sum of the deviations of each value from the mean, divided by the number of values

**Measure of center** Value intended to indicate the center of the values in a collection of data

**Measure of variation** Any of several measures designed to reflect the amount of variation or spread for a set of values

**Median** Middle value of a set of values arranged in order of magnitude

**Midquartile** One-half of the sum of the first and third quartiles

**Midrange** One-half the sum of the highest and lowest values

**Mode** Value that occurs most frequently

**Modified boxplot** Boxplot constructed with these modifications: (1) A special symbol (such as an asterisk or point) is used to identify, and (2) the solid horizontal line extends only as far as the minimum data value that is not an outlier and the maximum data value that is not an outlier

**MS(error)** Mean square for error; used in analysis of variance

**MS(total)** Mean square for total variation; used in analysis of variance

**MS(treatment)** Mean square for treatments; used in analysis of variance

**Multimodal** Having more than two modes

**Multinomial experiment** Experiment with a fixed number of independent trials, where each outcome falls into exactly one of several categories

**Multiple bar graph** Bar graph with two or more sets of bars used to compare two or more data sets

**Multiple coefficient of determination** Measure of how well a multiple regression equation fits the sample data

**Multiple comparison procedures** Procedures for identifying which particular means are different, after concluding that three or more means are not all equal

**Multiple regression** Study of linear relationships among three or more variables

**Multiple regression equation** Equation that expresses a linear relationship between a dependent variable  $y$  and two or more independent variables  $(x_1, x_2, \dots, x_k)$

**Multiplication rule** Rule for determining the probability that event  $A$  will occur on one trial and event  $B$  will occur on a second trial

**Mutually exclusive events** Events that cannot occur simultaneously

**Negatively skewed** Skewed to the left

**Nominal** Level of measurement of data; characterizes data that consist of names, labels, or categories only

**Nonparametric tests** Statistical procedures for testing hypotheses or estimating parameters, where there are no required assumptions about the nature or shape of population distributions; also called distribution-free tests

**Nonsampling errors** Errors from external factors not related to sampling

**Normal distribution** Bell-shaped probability distribution described algebraically by Formula 6-1 in Section 6-1

**Normal quantile plot** Graph of points  $(x, y)$ , where each  $x$  value is from the original set of sample data, and each  $y$  value is a  $z$  score corresponding to a quantile value of the standard normal distribution

***np* chart** Control chart in which numbers of defects are plotted so that a process can be monitored

**Null hypothesis** Claim made about some population characteristic, usually involving the case of no difference; denoted by  $H_0$

**Numerator degrees of freedom** Degrees of freedom corresponding to the numerator of the  $F$  test statistic

**Numerical data** Data consisting of numbers representing counts or measurements

**Observational study** Study in which we observe and measure specific characteristics, but don't attempt to manipulate or modify the subjects being studied

**Observed frequency** Actual frequency count recorded in one cell of a contingency table or multinomial table

**Odds against** Ratio of the probability of an event not occurring to the event occurring, usually expressed in the form of  $a:b$  where  $a$  and  $b$  are integers having no common factors

**Odds in favor** Ratio of the probability of an event occurring to the event not occurring, usually expressed as the ratio of two integers with no common factors

**Ogive** Graphical representation of a cumulative frequency distribution

**One-way analysis of variance** Analysis of variance involving data classified into groups according to a single criterion only

**Ordinal** Level of measurement of data; characterizes data that may be arranged in order, but differences between data values either cannot be determined or are meaningless

**Outliers** Values that are very unusual in the sense that they are very far away from most of the data

**Paired samples** Two samples that are dependent in the sense that the data values are matched by pairs

**Parameter** Measured characteristic of a population

**Parametric tests** Statistical procedures, based on population parameters, for testing hypotheses or estimating parameters

**Pareto chart** Bar graph for qualitative data, with the bars arranged in order according to frequencies

**Payoff odds** Ratio of net profit (if you win) to the amount bet

**$p$  chart** Control chart used to monitor the proportion  $p$  for some attribute in a process

**Pearson's product moment correlation coefficient** *See* Linear correlation coefficient.

**Percentile** The 99 values that divide ranked data into 100 groups with approximately 1% of the values in each group

**Permutations rule** Rule for determining the number of different arrangements of selected items

**Pie chart** Graphical representation of data in the form of a circle containing wedges

**Placebo effect** Effect that occurs when an untreated subject incorrectly believes that he or she is receiving a real treatment and reports an improvement in symptoms

**Point estimate** Single value that serves as an estimate of a population parameter

**Poisson distribution** Discrete probability distribution that applies to occurrences of some event over a specified interval of time, distance, area, volume, or some similar unit

**Pooled estimate of  $p_1$  and  $p_2$**  Probability obtained by combining the data from two sample proportions and dividing the total number of successes by the total number of observations

**Pooled estimate of  $\sigma^2$**  Estimate of the variance  $\sigma^2$  that is common to two populations, found by computing a weighted average of the two sample variances

**Population** Complete and entire collection of elements to be studied

**Positively skewed** Skewed to the right

**Power of a test** Probability  $(1 - \beta)$  of rejecting a false null hypothesis

**Predicted values** Values of a dependent variable found by using values of independent variables in a regression equation



**Prediction interval** Confidence interval estimate of a predicted value of  $y$

**Predictor variables** Independent variables in a regression equation

**Probability** Measure of the likelihood that a given event will occur; expressed as a number between 0 and 1

**Probability distribution** Collection of values of a random variable along with their corresponding probabilities

**Probability histogram** Histogram with outcomes listed along the horizontal axis and probabilities listed along the vertical axis

**Probability sample** Sample selected so that each member of the population has a know (but not necessarily the same) chance of being selected

**Probability value** See *P*-value.

**Process data** Data, arranged according to some time sequence, that measure a characteristic of goods or services resulting from some combination of equipment, people, materials, methods, and conditions

**Prospective study** Study of subjects in identified groups sharing common factors (called *cohorts*), with data collected in the future

***P*-value** Probability that a test statistic in a hypothesis test is at least as extreme as the one actually obtained

**Qualitative data** Data that can be separated into different categories distinguished by some nonnumeric characteristic

**Quantitative data** Data consisting of numbers representing counts or measurements

**Quartiles** The three values that divide ranked data into four groups with approximately 25% of the values in each group

**Randomized block design** Design in which a measurement is obtained for each treatment on each of several individuals matched according to similar characteristics

**Random sample** Sample selected in a way that allows every member of the population to have the same chance of being chosen

**Random selection** Selection of sample elements in such a way that all elements available for selection have the same chance of being selected

**Random variable** Variable (typically represented by  $x$ ) that has a single numerical value (determined by chance) for each outcome of an experiment

**Random variation** Type of variation in a process that is due to chance; the type of variation inherent in any process not capable of producing every good or service exactly the same way every time

**Range** The measure of variation that is the difference between the highest and lowest values

**Range chart** Control chart based on sample ranges; used to monitor variation in a process

**Range rule of thumb** Rule based on the principle that for typical data sets, the difference between the lowest typical value and the highest typical value is approximately 4 standard deviations ( $4s$ )

**Rank** Numerical position of an item in a sample set arranged in order

**Rank correlation coefficient** Measure of the strength of the relationship between two variables; based on the ranks of the values

**Rare event rule** If, under a given assumption, the probability of a particular observed result is extremely small, we conclude that the assumption is probably not correct

**Ratio** Level of measurement of data; characterizes data that can be arranged in order, for which differences between data values are meaningful, and there is an inherent zero starting point

**R chart** Control chart based on sample ranges; used to monitor variation in a process

**Regression equation** Algebraic equation describing the relationship among variables

**Regression line** Straight line that best fits a collection of points representing paired sample data

**Relative frequency** Frequency for a class, divided by the total of all frequencies

**Relative frequency approximation of probability** Estimated value of probability based on actual observations

**Relative frequency distribution** Variation of the basic frequency distribution in which the frequency for each class is divided by the total of all frequencies

**Relative frequency histogram** Variation of the basic histogram in which frequencies are replaced by relative frequencies

**Replication** Repetition of an experiment

**Residual** Difference between an observed sample  $y$  value and the value of  $y$  that is predicted from a regression equation

**Response variable**  $y$  variable in a regression or multiple regression equation

**Retrospective study** Study in which data are collected from the past by going back in time (through examination of records, interviews, and so on)

**Right-tailed test** Hypothesis test in which the critical region is located in the extreme right area of the probability distribution

**Rigorously controlled design** Design of experiment in which all factors are forced to be constant so that effects of extraneous factors are eliminated

**Run** Sequence of data exhibiting the same characteristic; used in runs test for randomness

**Run chart** Sequential plot of individual data values over time, where one axis (usually the vertical axis) is used for the data values and the other axis (usually the horizontal axis) is used for the time sequence

**Runs test** Nonparametric method used to test for randomness

**Sample** Subset of a population

**Sample size** Number of items in a sample

**Sample space** Set of all possible outcomes or events in an experiment that cannot be further broken down

**Sampling distribution of proportion** The probability distribution of sample proportions, with all samples having the same sample size  $n$

**Sampling distribution of sample means** Distribution of the sample means that is obtained when we repeatedly draw samples of the same size from the same population

**Sampling error** Difference between a sample result and the true population result; results from chance sample fluctuations

**Sampling variability** Variation of a statistic in different samples

**Scatter diagram** Graphical display of paired  $(x, y)$  data

**Scatterplot** Graphical display of paired  $(x, y)$  data

**s chart** Control chart, based on sample standard deviations, that is used to monitor variation in a process

**Self-selected sample** Sample in which the respondents themselves decide whether to be included; also called voluntary response sample

**Semi-interquartile range** One-half of the difference between the first and third quartiles

**Significance level** Probability of making a type I error when conducting a hypothesis test

**Sign test** Nonparametric hypothesis test used to compare samples from two populations

**Simple event** Experimental outcome that cannot be further broken down

**Simple random sample** Sample of a particular size selected so that every possible sample of the same size has the same chance of being chosen

**Simulation** Process that behaves in a way that is similar to some experiment so that similar results are produced

**Single factor analysis of variance** *See* One-way analysis of variance.

**Skewed** Not symmetric and extending more to one side than the other

**Slope** Measure of steepness of a straight line

**Sorted data** Data arranged in order

**Spearman's rank correlation coefficient** *See* Rank correlation coefficient.

**SS(error)** Sum of squares representing the variability that is assumed to be common to all the populations being considered; used in analysis of variance

**SS(total)** Measure of the total variation (around  $\bar{\bar{x}}$ ) in all of the sample data combined; used in analysis of variance

**SS(treatment)** Measure of the variation between the sample means; used in analysis of variance

**Standard deviation** Measure of variation equal to the square root of the variance

**Standard error of estimate** Measure of spread of sample points about the regression line

**Standard error of the mean** Standard deviation of all possible sample means  $\bar{\bar{x}}$

**Standard normal distribution** Normal distribution with a mean of 0 and a standard deviation equal to 1

**Standard score** Number of standard deviations that a given value is above or below the mean; also called  $z$  score

**Statistic** Measured characteristic of a sample

**Statistically stable process** Process with only natural variation and no patterns, cycles, or unusual points

**Statistical process control (SPC)** Use of statistical techniques such as control charts to analyze a process or its outputs so as to take appropriate actions to achieve and maintain a state of statistical control and to improve the process capability

**Statistics** Collection of methods for planning experiments, obtaining data, organizing, summarizing, presenting, analyzing, interpreting, and drawing conclusions based on data

**Stem-and-Leaf plot** *See* stemplot

**Stemplot** Method of sorting and arranging data to reveal the distribution

**Stepwise regression** Process of using different combinations of variables until the best model is obtained; used in multiple regression

**Stratified sampling** Sampling in which samples are drawn from each stratum (class)

**Student  $t$  distribution** *See*  $t$  distribution.

**Subjective probability** Guess or estimate of a probability based on knowledge of relevant circumstances

**Symmetric** Property of data for which the distribution can be divided into two halves that are approximately mirror images by drawing a vertical line through the middle

**Systematic sampling** Sampling in which every  $k$ th element is selected

**$t$  distribution** Bell-shaped distribution usually associated with sample data from a population with an unknown standard deviation.

**10–90 percentile range** Difference between the 10th and 90th percentiles

**Test of homogeneity** Test of the claim that different populations have the same proportion of some characteristic

**Test of independence** Test of the null hypothesis that for a contingency table, the row variable and column variable are not related

**Test of significance** *See* Hypothesis test.

**Test statistic** Sample statistic based on the sample data; used in making the decision about rejection of the null hypothesis

**Time-series data** Data that have been collected at different points in time

**Total deviation** Sum of the explained deviation and unexplained deviation for a given pair of values in a collection of bivariate data

**Total variation** Sum of the squares of the total deviation for all pairs of bivariate data in a sample

**Traditional method of testing hypotheses** Method of testing hypotheses based on a comparison of the test statistic and critical values

**Treatment** Property or characteristic that allows us to distinguish the different populations from one another; used in analysis of variance

**Treatment group** Group of subjects given some treatment in an experiment

**Tree diagram** Graphical depiction of the different possible outcomes in a compound event

**Two-tailed test** Hypothesis test in which the critical region is divided between the left and right extreme areas of the probability distribution

**Two-way analysis of variance** Analysis of variance involving data classified according to two different factors

**Two-way table** *See* Contingency table.

**Type I error** Mistake of rejecting the null hypothesis when it is true

**Type II error** Mistake of failing to reject the null hypothesis when it is false

**Unbiased estimator** Sample statistic that tends to target the population parameter that it is used to estimate

**Unexplained deviation** For one pair of values in a collection of bivariate data, the difference between the y coordinate and the predicted value

**Unexplained variation** Sum of the squares of the unexplained deviations for all pairs of bivariate data in a sample

**Uniform distribution** Probability distribution in which every value of the random variable is equally likely

**Upper class limits** Largest numbers that can belong to the different classes in a frequency distribution

**Upper control limit** Boundary used in a control chart to separate points that are unusually high

**Variance** Measure of variation equal to the square of the standard deviation

**Variance between samples** In analysis of variance, the variation among the different samples

**Variation due to error** *See* Variation within samples.

**Variation due to treatment** *See* Variance between samples.

**Variation within samples** In analysis of variance, the variation that is due to chance

**Voluntary response sample** Sample in which the respondents themselves decide whether to be included

**Weighted mean** Mean of a collection of values that have been assigned different degrees of importance

**Wilcoxon rank-sum test** Nonparametric hypothesis test used to compare two independent samples

**Wilcoxon signed-ranks test** Nonparametric hypothesis test used to compare two dependent samples

**Within statistical control** *See* Statistically stable process.

**$\bar{x}$  chart** Control chart used to monitor the mean of a process

**y-intercept** Point at which a straight line crosses the y-axis

**z score** Number of standard deviations that a given value is above or below the mean