

# Appendix E

## Glossary

**adjusted  $R^2$**  used in multiple regression when  $n$  and  $k$  are approximately equal, to provide a more realistic value of  $R^2$

**alpha** the probability of a type I error, represented by the Greek letter  $\alpha$

**alternative hypothesis** a statistical hypothesis that states a difference between a parameter and a specific value or states that there is a difference between two parameters

**analysis of variance (ANOVA)** a statistical technique used to test a hypothesis concerning the means of three or more populations

**ANOVA summary table** the table used to summarize the results of an ANOVA test

**Bayes' theorem** a theorem that allows you to compute the revised probability of an event that occurred before another event when the events are dependent

**beta** the probability of a type II error, represented by the Greek letter  $\beta$

**between-group variance** a variance estimate using the means of the groups or between the groups in an  $F$  test

**biased sample** a sample for which some type of systematic error has been made in the selection of subjects for the sample

**bimodal** a data set with two modes

**binomial distribution** the outcomes of a binomial experiment and the corresponding probabilities of these outcomes

**binomial experiment** a probability experiment in which each trial has only two outcomes, there are a fixed number of trials, the outcomes of the trials are independent, and the probability of success remains the same for each trial

**boxplot** a graph used to represent a data set when the data set contains a small number of values

**categorical frequency distribution** a frequency distribution used when the data are categorical (nominal)

**central limit theorem** a theorem that states that as the sample size increases, the shape of the distribution of the sample means taken from the population with mean  $\mu$  and standard deviation  $\sigma$  will approach a normal distribution; the distribution will have a mean  $\mu$  and a standard deviation  $\sigma/\sqrt{n}$

**Chebyshev's theorem** a theorem that states that the proportion of values from a data set that fall within  $k$  standard deviations of the mean will be at least  $1 - 1/k^2$ , where  $k$  is a number greater than 1

**chi-square distribution** a probability distribution obtained from the values of  $(n - 1)s^2/\sigma^2$  when random samples are selected from a normally distributed population whose variance is  $\sigma^2$

**class boundaries** the upper and lower values of a class for a grouped frequency distribution whose values have one additional decimal place more than the data and end in the digit 5

**class midpoint** a value for a class in a frequency distribution obtained by adding the lower and upper class boundaries (or the lower and upper class limits) and dividing by 2

**class width** the difference between the upper class boundary and the lower class boundary for a class in a frequency distribution

**classical probability** the type of probability that uses sample spaces to determine the numerical probability that an event will happen

**cluster sample** a sample obtained by selecting a preexisting or natural group, called a cluster, and using the members in the cluster for the sample

**coefficient of determination** a measure of the variation of the dependent variable that is explained by the regression line and the independent variable; the ratio of the explained variation to the total variation

**coefficient of variation** the standard deviation divided by the mean with the result expressed as a percentage

**combination** a selection of objects without regard to order

- complement of an event** the set of outcomes in the sample space that are not among the outcomes of the event itself
- compound event** an event that consists of two or more outcomes or simple events
- conditional probability** the probability that an event  $B$  occurs after an event  $A$  has already occurred
- confidence interval** a specific interval estimate of a parameter determined by using data obtained from a sample and the specific confidence level of the estimate
- confidence level** the probability that a parameter lies within the specified interval estimate of the parameter
- confounding variable** a variable that influences the outcome variable but cannot be separated from the other variables that influence the outcome variable
- consistent estimator** an estimator whose value approaches the value of the parameter estimated as the sample size increases
- contingency table** data arranged in table form for the chi-square independence test, with  $R$  rows and  $C$  columns
- continuous variable** a variable that can assume all values between any two specific values; a variable obtained by measuring
- control group** a group in an experimental study that is not given any special treatment
- convenience sample** sample of subjects used because they are convenient and available
- correction for continuity** a correction employed when a continuous distribution is used to approximate a discrete distribution
- correlation** a statistical method used to determine whether a linear relationship exists between variables
- correlation coefficient** a statistic or parameter that measures the strength and direction of a linear relationship between two variables
- critical or rejection region** the range of values of the test value that indicates that there is a significant difference and the null hypothesis should be rejected in a hypothesis test
- critical value (C.V.)** a value that separates the critical region from the noncritical region in a hypothesis test
- cumulative frequency** the sum of the frequencies accumulated up to the upper boundary of a class in a frequency distribution
- data** measurements or observations for a variable
- data array** a data set that has been ordered
- data set** a collection of data values
- data value or datum** a value in a data set
- decile** a location measure of a data value; it divides the distribution into 10 groups
- degrees of freedom** the number of values that are free to vary after a sample statistic has been computed; used when a distribution (such as the  $t$  distribution) consists of a family of curves
- dependent events** events for which the outcome or occurrence of the first event affects the outcome or occurrence of the second event in such a way that the probability is changed
- dependent samples** samples in which the subjects are paired or matched in some way; i.e., the samples are related
- dependent variable** a variable in correlation and regression analysis that cannot be controlled or manipulated
- descriptive statistics** a branch of statistics that consists of the collection, organization, summarization, and presentation of data
- discrete variable** a variable that assumes values that can be counted
- disordinal interaction** an interaction between variables in ANOVA, indicated when the graphs of the lines connecting the mean intersect
- distribution-free statistics** *see* nonparametric statistics
- double sampling** a sampling method in which a very large population is given a questionnaire to determine those who meet the qualifications for a study; the questionnaire is reviewed, a second smaller population is defined, and a sample is selected from this group
- empirical probability** the type of probability that uses frequency distributions based on observations to determine numerical probabilities of events
- empirical rule** a rule that states that when a distribution is bell-shaped (normal), approximately 68% of the data values will fall within 1 standard deviation of the mean; approximately 95% of the data values will fall within 2 standard deviations of the mean; and approximately 99.7% of the data values will fall within 3 standard deviations of the mean
- equally likely events** the events in the sample space that have the same probability of occurring
- estimation** the process of estimating the value of a parameter from information obtained from a sample
- estimator** a statistic used to estimate a parameter
- event** outcome of a probability experiment
- expected frequency** the frequency obtained by calculation (as if there were no preference) and used in the chi-square test
- expected value** the theoretical average of a variable that has a probability distribution

**experimental study** a study in which the researcher manipulates one of the variables and tries to determine how the manipulation influences other variables

**explanatory variable** a variable that is being manipulated by the researcher to see if it affects the outcome variable

**exploratory data analysis** the act of analyzing data to determine what information can be obtained by using stem and leaf plots, medians, interquartile ranges, and boxplots

**extrapolation** use of the equation for the regression line to predict  $y'$  for a value of  $x$  that is beyond the range of the data values of  $x$

**$F$  distribution** the sampling distribution of the variances when two independent samples are selected from two normally distributed populations in which the variances are equal and the variances  $s_1^2$  and  $s_2^2$  are compared as  $s_1^2 \div s_2^2$

**$F$  test** a statistical test used to compare two variances or three or more means

**factors** the independent variables in ANOVA tests

**finite population correction factor** a correction factor used to correct the standard error of the mean when the sample size is greater than 5% of the population size

**five-number summary** five specific values for a data set that consist of the lowest and highest values,  $Q_1$  and  $Q_3$ , and the median

**frequency** the number of values in a specific class of a frequency distribution

**frequency distribution** an organization of raw data in table form, using classes and frequencies

**frequency polygon** a graph that displays the data by using lines that connect points plotted for the frequencies at the midpoints of the classes

**goodness-of-fit test** a chi-square test used to see whether a frequency distribution fits a specific pattern

**grouped frequency distribution** a distribution used when the range is large and classes of several units in width are needed

**Hawthorne effect** an effect on an outcome variable caused by the fact that subjects of the study know that they are participating in the study

**histogram** a graph that displays the data by using vertical bars of various heights to represent the frequencies of a distribution

**homogeneity of proportions test** a test used to determine the equality of three or more proportions

**hypergeometric distribution** the distribution of a variable that has two outcomes when sampling is done without replacement

**hypothesis testing** a decision-making process for evaluating claims about a population

**independence test** a chi-square test used to test the independence of two variables when data are tabulated in table form in terms of frequencies

**independent events** events for which the probability of the first occurring does not affect the probability of the second occurring

**independent samples** samples that are not related

**independent variable** a variable in correlation and regression analysis that can be controlled or manipulated

**inferential statistics** a branch of statistics that consists of generalizing from samples to populations, performing hypothesis testing, determining relationships among variables, and making predictions

**influential observation** an observation that when removed from the data values would markedly change the position of the regression line

**interaction effect** the effect of two or more variables on each other in a two-way ANOVA study

**interquartile range**  $Q_3 - Q_1$

**interval estimate** a range of values used to estimate a parameter

**interval level of measurement** a measurement level that ranks data and in which precise differences between units of measure exist. *See also* nominal, ordinal, and ratio levels of measurement

**Kruskal-Wallis test** a nonparametric test used to compare three or more means

**law of large numbers** when a probability experiment is repeated a large number of times, the relative frequency probability of an outcome will approach its theoretical probability

**least-squares line** another name for the regression line

**left-tailed test** a test used on a hypothesis when the critical region is on the left side of the distribution

**level** a treatment in ANOVA for a variable

**level of significance** the maximum probability of committing a type I error in hypothesis testing

**lower class limit** the lower value of a class in a frequency distribution that has the same decimal place value as the data

**lurking variable** a variable that influences the relationship between  $x$  and  $y$ , but was not considered in the study

- main effect** the effect of the factors or independent variables when there is a nonsignificant interaction effect in a two-way ANOVA study
- marginal change** the magnitude of the change in the dependent variable when the independent variable changes 1 unit
- maximum error of estimate** the maximum likely difference between the point estimate of a parameter and the actual value of the parameter
- mean** the sum of the values, divided by the total number of values
- mean square** the variance found by dividing the sum of the squares of a variable by the corresponding degrees of freedom; used in ANOVA
- measurement scales** a type of classification that tells how variables are categorized, counted, or measured; the four types of scales are nominal, ordinal, interval, and ratio
- median** the midpoint of a data array
- midrange** the sum of the lowest and highest data values, divided by 2
- modal class** the class with the largest frequency
- mode** the value that occurs most often in a data set
- Monte Carlo method** a simulation technique using random numbers
- multimodal** a data set with three or more modes
- multinomial distribution** a probability distribution for an experiment in which each trial has more than two outcomes
- multiple correlation coefficient** a measure of the strength of the relationship between the independent variables and the dependent variable in a multiple regression study
- multiple regression** a study that seeks to determine if several independent variables are related to a dependent variable
- multiple relationship** a relationship in which many variables are under study
- multistage sampling** a sampling technique that uses a combination of sampling methods
- mutually exclusive events** probability events that cannot occur at the same time
- negative relationship** a relationship between variables such that as one variable increases, the other variable decreases, and vice versa
- negatively skewed or left-skewed distribution** a distribution in which the majority of the data values fall to the right of the mean
- nominal level of measurement** a measurement level that classifies data into mutually exclusive (nonoverlapping) exhaustive categories in which no order or ranking can be imposed on them. *See also* interval, ordinal, and ratio levels of measurement
- noncritical or nonrejection region** the range of values of the test value that indicates that the difference was probably due to chance and the null hypothesis should not be rejected
- nonparametric statistics** a branch of statistics for use when the population from which the samples are selected is not normally distributed and for use in testing hypotheses that do not involve specific population parameters
- nonrejection region** *see* noncritical region
- normal distribution** a continuous, symmetric, bell-shaped distribution of a variable
- normal quantile plot** graphical plot used to determine whether a variable is approximately normally distributed
- 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** a study in which the researcher merely observes what is happening or what has happened in the past and draws conclusions based on these observations
- observed frequency** the actual frequency value obtained from a sample and used in the chi-square test
- ogive** a graph that represents the cumulative frequencies for the classes in a frequency distribution
- one-tailed test** a test that indicates that the null hypothesis should be rejected when the test statistic value is in the critical region on one side of the mean
- one-way ANOVA** a study used to test for differences among means for a single independent variable when there are three or more groups
- open-ended distribution** a frequency distribution that has no specific beginning value or no specific ending value
- ordinal interaction** an interaction between variables in ANOVA, indicated when the graphs of the lines connecting the means do not intersect
- ordinal level of measurement** a measurement level that classifies data into categories that can be ranked; however, precise differences between the ranks do not exist. *See also* interval, nominal, and ratio levels of measurement
- outcome** the result of a single trial of a probability experiment
- outcome variable** a variable that is studied to see if it has changed significantly due to the manipulation of the explanatory variable
- outlier** an extreme value in a data set; it is omitted from a boxplot



- parameter** a characteristic or measure obtained by using all the data values for a specific population
- parametric tests** statistical tests for population parameters such as means, variances, and proportions that involve assumptions about the populations from which the samples were selected
- Pareto chart** chart that uses vertical bars to represent frequencies for a categorical variable
- Pearson product moment correlation coefficient (PPMCC)** a statistic used to determine the strength of a relationship when the variables are normally distributed
- Pearson's index of skewness** value used to determine the degree of skewness of a variable
- percentile** a location measure of a data value; it divides the distribution into 100 groups
- permutation** an arrangement of  $n$  objects in a specific order
- pie graph** a circle that is divided into sections or wedges according to the percentage of frequencies in each category of the distribution
- point estimate** a specific numerical value estimate of a parameter
- Poisson distribution** a probability distribution used when  $n$  is large and  $p$  is small and when the independent variables occur over a period of time
- pooled estimate of the variance** a weighted average of the variance using the two sample variances and their respective degrees of freedom as the weights
- population** the totality of all subjects possessing certain common characteristics that are being studied
- population correlation coefficient** the value of the correlation coefficient computed by using all possible pairs of data values  $(x, y)$  taken from a population
- positive relationship** a relationship between two variables such that as one variable increases, the other variable increases or as one variable decreases, the other decreases
- positively skewed or right-skewed distribution** a distribution in which the majority of the data values fall to the left of the mean
- power of a test** the probability of rejecting the null hypothesis when it is false
- prediction interval** a confidence interval for a predicted value  $y$
- probability** the chance of an event occurring
- probability distribution** the values a random variable can assume and the corresponding probabilities of the values
- probability experiment** a chance process that leads to well-defined results called outcomes
- proportion** a part of a whole, represented by a fraction, a decimal, or a percentage
- P-value** the actual probability of getting the sample mean value if the null hypothesis is true
- qualitative variable** a variable that can be placed into distinct categories, according to some characteristic or attribute
- quantiles** values that separate the data set into approximately equal groups
- quantitative variable** a variable that is numerical in nature and that can be ordered or ranked
- quartile** a location measure of a data value; it divides the distribution into four groups
- quasi-experimental study** a study that uses intact groups rather than random assignment of subjects to groups
- random sample** a sample obtained by using random or chance methods; a sample for which every member of the population has an equal chance of being selected
- random variable** a variable whose values are determined by chance
- range** the highest data value minus the lowest data value
- range rule of thumb** dividing the range by 4, given an approximation of the standard deviation
- ranking** the positioning of a data value in a data array according to some rating scale
- ratio level of measurement** a measurement level that possesses all the characteristics of interval measurement and a true zero; it also has true ratios between different units of measure. *See also* interval, nominal, and ordinal levels of measurement
- raw data** data collected in original form
- regression** a statistical method used to describe the nature of the relationship between variables, that is, a positive or negative, linear or nonlinear relationship
- regression line** the line of best fit of the data
- rejection region** *see* critical region
- relative frequency graph** a graph using proportions instead of raw data as frequencies
- relatively efficient estimator** an estimator that has the smallest variance from among all the statistics that can be used to estimate a parameter
- residual** the difference between the actual value of  $y$  and the predicted value  $y'$  for a specific value of  $x$
- residual plot** plot of the  $x$  values and the residuals to determine how well the regression line can be used to make predictions
- resistant statistic** a statistic that is not affected by an extremely skewed distribution
- right-tailed test** a test used on a hypothesis when the critical region is on the right side of the distribution
- run** a succession of identical letters preceded by or followed by a different letter or no letter at all, such as the beginning or end of the succession

**runs test** a nonparametric test used to determine whether data are random

**sample** a group of subjects selected from the population

**sample space** the set of all possible outcomes of a probability experiment

**sampling distribution of sample means** a distribution obtained by using the means computed from random samples taken from a population

**sampling error** the difference between the sample measure and the corresponding population measure due to the fact that the sample is not a perfect representation of the population

**scatter plot** a graph of the independent and dependent variables in regression and correlation analysis

**Scheffé test** a test used after ANOVA, if the null hypothesis is rejected, to locate significant differences in the means

**sequence sampling** a sampling technique used in quality control in which successive units are taken from production lines and tested to see whether they meet the standards set by the manufacturing company

**sign test** a nonparametric test used to test the value of the median for a specific sample or to test sample means in a comparison of two dependent samples

**simple event** an outcome that results from a single trial of a probability experiment

**simple relationship** a relationship in which only two variables are under study

**simulation techniques** techniques that use probability experiments to mimic real-life situations

**Spearman rank correlation coefficient** the nonparametric equivalent to the correlation coefficient, used when the data are ranked

**standard deviation** the square root of the variance

**standard error of the estimate** the standard deviation of the observed  $y$  values about the predicted  $y'$  values in regression and correlation analysis

**standard error of the mean** the standard deviation of the sample means for samples taken from the same population

**standard normal distribution** a normal distribution for which the mean is equal to 0 and the standard deviation is equal to 1

**standard score** the difference between a data value and the mean, divided by the standard deviation

**statistic** a characteristic or measure obtained by using the data values from a sample

**statistical hypothesis** a conjecture about a population parameter, which may or may not be true

**statistical test** a test that uses data obtained from a sample to make a decision about whether the null hypothesis should be rejected

**statistics** the science of conducting studies to collect, organize, summarize, analyze, and draw conclusions from data

**stem and leaf plot** a data plot that uses part of a data value as the stem and part of the data value as the leaf to form groups or classes

**stratified sample** a sample obtained by dividing the population into subgroups, called strata, according to various homogeneous characteristics and then selecting members from each stratum

**subjective probability** the type of probability that uses a probability value based on an educated guess or estimate, employing opinions and inexact information

**sum of squares between groups** a statistic computed in the numerator of the fraction used to find the between-group variance in ANOVA

**sum of squares within groups** a statistic computed in the numerator of the fraction used to find the within-group variance in ANOVA

**symmetric distribution** a distribution in which the data values are uniformly distributed about the mean

**systematic sample** a sample obtained by numbering each element in the population and then selecting every  $k$ th number from the population to be included in the sample

**$t$  distribution** a family of bell-shaped curves based on degrees of freedom, similar to the standard normal distribution with the exception that the variance is greater than 1; used when you are testing small samples and when the population standard deviation is unknown

**$t$  test** a statistical test for the mean of a population, used when the population is normally distributed and the population standard deviation is unknown

**test value** the numerical value obtained from a statistical test, computed from  $(\text{observed value} - \text{expected value}) \div \text{standard error}$

**time series graph** a graph that represents data that occur over a specific time

**treatment group** a group in an experimental study that has received some type of treatment

**treatment groups** the groups used in an ANOVA study

**tree diagram** a device used to list all possibilities of a sequence of events in a systematic way

**Tukey test** a test used to make pairwise comparisons of means in an ANOVA study when samples are the same size

**two-tailed test** a test that indicates that the null hypothesis should be rejected when the test value is in either of the two critical regions

**two-way ANOVA** a study used to test the effects of two or more independent variables and the possible interaction between them

**type I error** the error that occurs if you reject the null hypothesis when it is true

**type II error** the error that occurs if you do not reject the null hypothesis when it is false

**unbiased estimator** an estimator whose value approximates the expected value of a population parameter, used for the variance or standard deviation when the sample size is less than 30; an estimator whose expected value or mean must be equal to the mean of the parameter being estimated

**unbiased sample** a sample chosen at random from the population that is, for the most part, representative of the population

**ungrouped frequency distribution** a distribution that uses individual data and has a small range of data

**uniform distribution** a distribution whose values are evenly distributed over its range

**upper class limit** the upper value of a class in a frequency distribution that has the same decimal place value as the data

**variable** a characteristic or attribute that can assume different values

**variance** the average of the squares of the distance that each value is from the mean

**Venn diagram** a diagram used as a pictorial representative for a probability concept or rule

**weighted mean** the mean found by multiplying each value by its corresponding weight and dividing by the sum of the weights

**Wilcoxon rank sum test** a nonparametric test used to test independent samples and compare distributions

**Wilcoxon signed-rank test** a nonparametric test used to test dependent samples and compare distributions

**within-group variance** a variance estimate using all the sample data for an  $F$  test; it is not affected by differences in the means

**z distribution** *see* standard normal distribution

**z score** *see* standard score

**z test** a statistical test for means and proportions of a population, used when the population is normally distributed and the population standard deviation is known

**z value** same as  $z$  score

# Glossary of Symbols

$a$	$y$ intercept of a line	$F_S$	Scheffé test value
$\alpha$	Probability of a type I error	GM	Geometric mean
$b$	Slope of a line	$H$	Kruskal-Wallis test value
$\beta$	Probability of a type II error	$H_0$	Null hypothesis
$C$	Column frequency	$H_1$	Alternative hypothesis
cf	Cumulative frequency	HM	Harmonic mean
${}_n C_r$	Number of combinations of $n$ objects taking $r$ objects at a time	$k$	Number of samples
C.V.	Critical value	$\lambda$	Number of occurrences for the Poisson distribution
CVar	Coefficient of variation	$s_D$	Standard deviation of the differences
$D$	Difference; decile	$s_{est}$	Standard error of estimate
$\bar{D}$	Mean of the differences	$SS_B$	Sum of squares between groups
d.f.	Degrees of freedom	$SS_W$	Sum of squares within groups
d.f.N.	Degrees of freedom, numerator	$s_B^2$	Between-group variance
d.f.D.	Degrees of freedom, denominator	$s_W^2$	Within-group variance
$E$	Event; expected frequency; maximum error of estimate	$t$	$t$ test value
$\bar{E}$	Complement of an event	$t_{\alpha/2}$	Two-tailed $t$ critical value
$e$	Euler's constant $\approx 2.7183$	$\mu$	Population mean
$E(X)$	Expected value	$\mu_D$	Mean of the population differences
$f$	Frequency	$\mu_{\bar{x}}$	Mean of the sample means
$F$	$F$ test value; failure	$w$	Class width; weight
$F'$	Critical value for the Scheffé test	$r$	Sample correlation coefficient
MD	Median	$R$	Multiple correlation coefficient
MR	Midrange	$r^2$	Coefficient of determination
$MS_B$	Mean square between groups	$\rho$	Population correlation coefficient
$MS_W$	Mean square within groups (error)	$r_s$	Spearman rank correlation coefficient
$n$	Sample size	$S$	Sample space; success
$N$	Population size	$s$	Sample standard deviation
$n(E)$	Number of ways $E$ can occur	$s^2$	Sample variance
$n(S)$	Number of outcomes in the sample space	$\sigma$	Population standard deviation
$O$	Observed frequency	$\sigma^2$	Population variance
$P$	Percentile; probability	$\sigma_{\bar{x}}$	Standard error of the mean
$p$	Probability; population proportion	$\Sigma$	Summation notation
$\hat{p}$	Sample proportion	$w_s$	Smaller sum of signed ranks, Wilcoxon signed-rank test
$\bar{p}$	Weighted estimate of $p$	$X$	Data value; number of successes for a binomial distribution
$P(B A)$	Conditional probability	$\bar{X}$	Sample mean
$P(E)$	Probability of an event $E$	$x$	Independent variable in regression
$P(\bar{E})$	Probability of the complement of $E$	$\bar{X}_{GM}$	Grand mean
${}_n P_r$	Number of permutations of $n$ objects taking $r$ objects at a time	$X_m$	Midpoint of a class
$\pi$	Pi $\approx 3.14$	$\chi^2$	Chi-square
$Q$	Quartile	$y$	Dependent variable in regression
$q$	$1 - p$ ; test value for Tukey test	$y'$	Predicted $y$ value
$\hat{q}$	$1 - \hat{p}$	$z$	$z$ test value or $z$ score
$\bar{q}$	$1 - \bar{p}$	$z_{\alpha/2}$	Two-tailed $z$ critical value
$R$	Range; rank sum	!	Factorial