Glossary

Acceptance criterion The maximum number of defective items that can be found in the sample and still indicate an acceptable lot.

Acceptance sampling A statistical method in which the number of defective items found in a sample is used to determine whether a lot should be accepted or rejected.

Addition law A probability law used to compute the probability of the union of two events. It is $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. For mutually exclusive events, $P(A \cap B) = 0$; in this case the addition law reduces to $P(A \cup B) = P(A) + P(B)$.

Adjusted multiple coefficient of determination A measure of the goodness of fit of the estimated multiple regression equation that adjusts for the number of independent variables in the model and thus avoids overestimating the impact of adding more independent variables.

Aggregate price index A composite price index based on the prices of a group of items.

Alternative hypothesis The hypothesis concluded to be true if the null hypothesis is rejected.

ANOVA table A table used to summarize the analysis of variance computations and results. It contains columns showing the source of variation, the sum of squares, the degrees of freedom, the mean square, and the F value(s).

Assignable causes Variations in process outputs that are due to factors such as machine tools wearing out, incorrect machine settings, poor-quality raw materials, operator error, and so on. Corrective action should be taken when assignable causes of output variation are detected.

Autocorrelation Correlation in the errors that arises when the error terms at successive points in time are related.

Autoregressive model A time series model whereby a regression relationship based on past time series values is used to predict the future time series values.

Bar graph, Bar chart A graphical device for depicting qualitative data that have been summarized in a frequency, relative frequency, or percentage frequency distribution.

Basic requirements for assigning probabilities Two requirements that restrict the manner in which probability assignments can be made: (1) for each experimental outcome E_i we must have $0 \leqslant P(E_i) \leqslant 1$; (2) considering all experimental outcomes, we must have $P(E_1) + P(E_2) + \cdots + P(E_n) = 1.0$.

Bayes' theorem A theorem that enables the use of sample information to revise prior probabilities.

Binomial experiment An experiment having the four properties stated at the beginning of Section 5.4.

Binomial probability distribution A probability distribution showing the probability of x successes trials of binomial experiences

Binomial probability function The function used to compute binomial probabilities.

Blocking The process of using the same or similar experimental units for all treatments. The purpose of blocking is to remove a source of variation from the error term and hence provide a more powerful test for a difference in population or treatment means.

Bound on the sampling error A number added to and subtracted from a point estimate to create an approximate 95 per cent confidence interval. It is given by two times the standard error of the point estimator.

Box plot A graphical summary of data based on a five-number summary.

Branch Lines showing the alternatives from decision nodes and the outcomes from chance nodes.

Causal forecasting methods Forecasting methods that relate a time series to other variables that are believed to explain or cause its behaviour.

Census A survey to collect data on the entire population.

Central limit theorem A theorem that enables one to use the normal probability distribution to approximate the sampling distribution of $\bar{x}$ when the sample size is large.

Chance event An uncertain future event affecting the consequence, or payoff, associated with a decision.

Chance nodes Nodes indicating points where an uncertain event will occur.

Chebyshev's theorem A theorem that can be used to make statements about the proportion of data values that must be within a specified number of standard deviations of the mean.

Class midpoint The value halfway between the lower and upper class limits in a frequency distribution.

Classical method A method of assigning probabilities that is appropriate when all the experimental outcomes are equally likely.

Cluster sampling A probabilistic method of sampling in which the population is first divided into clusters and then one or more clusters are selected for sampling. In single-stage cluster sampling, every element in each selected cluster is sampled; in two-stage cluster sampling, a sample of the elements in each selected cluster is collected.

Coefficient of determination A measure of the goodness of fit of the estimated regression equation. It can be interpreted as the proportion of the variability in the

dependent variable y that is explained by the estimated regression equation.

Coefficient of variation A measure of relative variability computed by dividing the standard deviation by the mean and multiplying by 100.

Common causes Normal or natural variations in process outputs that are due purely to chance. No corrective action is necessary when output variations are due to common causes.

Comparisonwise Type I error rate The probability of a Type I error associated with a single pairwise comparison.

Complement of A The event consisting of all sample points that are not in A.

Completely randomized design An experimental design in which the treatments are randomly assigned to the experimental units.

Conditional probability The probability of an event given that another event already occurred. The conditional probability of A given B is $P(A \mid B) = P(A \cap B)/P(B)$.

Confidence coefficient The confidence level expressed as a decimal value. For example, 0.95 is the confidence coefficient for a 95 per cent confidence level.

Confidence interval The interval estimate of the mean value of y for a given value of x.

Confidence level The confidence associated with an interval estimate. For example, if an interval estimation procedure provides intervals such that 95 per cent of the intervals formed using the procedure will include the population parameter, the interval estimate is said to be constructed at the 95 per cent confidence level.

Consequence The result obtained when a decision alternative is chosen and a chance event occurs. A measure of the consequence is often called a payoff.

Consumer Price Index A price index that uses the price changes in a market basket of consumer goods and services to measure the changes in consumer prices over time.

Consumer's risk The risk of accepting a poor-quality lot; a Type II error.

Contingency table A table used to summarize observed and expected frequencies for a test of independence.

Continuity correction factor A value of 0.5 that is added to or subtracted from a value of x when the continuous normal distribution is used to approximate the discrete binomial distribution.

Continuous random variable A random variable that may assume any numerical value in an interval or collection of intervals.

Control chart A graphical tool used to help determine whether a process is in control or out of control.

Convenience sampling A non-probabilistic method of sampling whereby elements are selected on the basis of convenience.

Cook's distance measure A measure of the influence of an observation based on both the leverage of observation i and the residual for observation i.

Correlation coefficient A measure of association between two variables that takes on values between -1 and $+1$. Values near $+1$ indicate a strong positive relationship, values near -1 indicate a strong negative relationship. Values near zero indicate the lack of a relationship. Pearson's product-moment correlation coefficient measures linear association between two variables.

Covariance A measure of linear association between two variables. Positive values indicate a positive relationship; negative values indicate a negative relationship.

Critical value A value that is compared with the test statistic to determine whether H_0 should be rejected.

Cross-sectional data Data collected at the same or approximately the same point in time.

Cross-tabulation A tabular summary of data for two variables. The classes for one variable are represented by the rows; the classes for the other variable are represented by the columns.

Cumulative frequency distribution A tabular summary of quantitative data showing the number of items with values less than or equal to the upper class limit of each class.

Cumulative percentage frequency distribution A tabular summary of quantitative data showing the percentage of items with values less than or equal to the upper class limit of each class.

Cumulative relative frequency distribution A tabular summary of quantitative data showing the fraction or proportion of items with values less than or equal to the upper class limit of each class.

Cyclical component The component of the time series that results in periodic above-trend and below-trend behaviour of the time series lasting more than one year.

Data The facts and figures collected, analyzed, and summarized for presentation and interpretation.

Decision nodes Nodes indicating points where a decision is made.

Data set All the data collected in a particular study.

Decision strategy A strategy involving a sequence of decisions and chance outcomes to provide the optimal solution to a decision problem.

Decision tree A graphical representation of the decision problem that shows the sequential nature of the decision-making process.

Degrees of freedom A parameter of the t distribution. When the t distribution is used in the computation of an interval estimate of a population mean, the appropriate t distribution has $n - 1$ degrees of freedom, where n is the size of the simple random sample. (Also a parameter of the χ^2 distribution.)

Delphi method A qualitative forecasting method that obtains forecasts through group consensus.

Dependent variable The variable that is being predicted or explained. It is denoted by y.

Descriptive statistics Tabular, graphical, and numerical summaries of data.

Deseasonalized time series A time series from which the effect of season has been removed by dividing each original time series observation by the corresponding seasonal index.

Discrete random variable A random variable that may assume either a finite number of values or an infinite sequence of values.

Discrete uniform probability distribution A probability distribution for which each possible value of the random variable has the same probability.

Dot plot A graphical device that summarizes data by the number of dots above each data value on the horizontal axis.

Dummy variable A variable used to model the effect of qualitative independent variables. A dummy variable may take only the value zero or one.

Durbin-Watson test A test to determine whether first-order correlation is present.

Element The entity on which data are collected.

Empirical rule A rule that can be used to compute the percentage of data values that must be within one, two and three standard deviations of the mean for data that exhibit a bell-shaped distribution.

Estimated logistic regression equation The estimate of the logistic regression equation based on sample data; that is $\hat{y}$ = estimate of

$$P(y = 1 \mid x_1, x_2, \ldots x_p) = \frac{e^{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p}}{1 + e^{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p}}$$

Estimated logit An estimate of the logit based on sample data; that is,

$$\hat{g}(x_1, x_2 \ldots, x_p) = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p$$

Estimated multiple regression equation The estimate of the multiple regression equation based on sample data and the least squares method; it is $\hat{y} = b_0 + b_1x_1 + b_2x_2 + \cdots\cdots + b_px_p$.

Estimated regression equation The estimate of the regression equation developed from sample data by using the least squares method. For simple linear regression, the estimated regression equation is $\hat{y} = b_0 + b_1x$.

Event A collection of sample points.

Expected value A measure of the central location of a random variable. For a chance node, it is the weighted average of the payoffs. The weights are the state-of-nature probabilities.

Expected value approach An approach to choosing a decision alternative that is based on the expected value of each decision alternative. The recommended decision alternative is the one that provides the best expected value.

Expected value of perfect information (EVPI) The expected value of information that would tell the decision-maker exactly which state of nature is going to occur (i.e., perfect information).

Expected value of sample information (EVSI) The difference between the expected value of an optimal strategy based on sample information and the 'best' expected value without any sample information.

Experiment A process that generates well-defined outcomes.

Experimental units The objects of interest in the experiment.

Experimentwise Type I error rate The probability of making a Type I error on at least one of several pairwise comparisons.

Exploratory data analysis Methods that use simple arithmetic and easy-to-draw graphs to summarize data quickly.

Exponential probability distribution A continuous probability distribution that is useful in computing probabilities for the time it takes to complete a task.

Exponential smoothing A forecasting technique that uses a weighted average of past time series values as the forecast.

Factor Another word for the independent variable of interest.

Factorial experiment An experimental design that allows statistical conclusions about two or more factors.

Finite population correction factor The term $\sqrt{(N-n)/(N-1)}$ that is used in the formulae for $\sigma_{\bar{x}}$ and σ_p when a finite population, rather than an infinite population, is being sampled. The generally accepted rule of thumb is to ignore the finite population correction factor whenever $n/N \leq 0.05$.

Five-number summary An exploratory data analysis technique that uses five numbers to summarize the data: smallest value, first quartile, median, third quartile, and largest value.

Forecast A prediction of future values of a time series.

Frame A list of the sampling units for a study. The sample is drawn by selecting units from the frame.

Frequency distribution A tabular summary of data showing the number (frequency) of items in each of several non-overlapping classes.

General linear model A model of the form $y = \beta_0 + \beta_1z_1 + \beta_2z_2 + \cdots + \beta_pz_p + \varepsilon$, where each of the independent variables z_j $(j = 1, 2, \ldots, p)$ is a function of $x_1, x_2, \ldots, x_k$, the variables for which data have been collected.

Goodness of fit test A statistical test conducted to determine whether to reject a hypothesized probability distribution for a population.

Grouped data Data available in class intervals as summarized by a frequency distribution. Individual values of the original data are not available.

High leverage points Observations with extreme values for the independent variables.

Histogram A graphical presentation of a frequency distribution, relative frequency distribution, or percentage frequency distribution of quantitative data constructed by placing the class intervals on the horizontal axis and the frequencies, relative frequencies, or percentage frequencies on the vertical axis.

Hypergeometric probability distribution A probability distribution showing the probability of x successes in n trials from a population with r successes and $N - r$ failures.

Hypergeometric probability function The function used to compute hypergeometric probabilities.

Independent events Two events A and B where $P(A \mid B) = P(A)$ or $P(B \mid A) = P(B)$; that is, the events have no influence on each other.

Independent samples Where, e.g., two groups of workers are selected and each group uses a different method to collect production time data.

Independent variable The variable that is doing the predicting or explaining. It is denoted by x.

Influential observation An observation that has a strong influence or effect on the regression results.

Interaction The effect of two independent variables acting together.

Interquartile range (IQR) A measure of variability, defined to be the difference between the third and first quartiles.

Intersection of *A* and *B* The event containing the sample points belonging to both A and B. The intersection is denoted $A \cap B$.

Interval estimate An estimate of a population parameter that provides an interval believed to contain the value of the parameter.

Interval scale The scale of measurement for a variable if the data demonstrate the properties of ordinal data and the interval between values is expressed in terms of a fixed unit of measure. Interval data are always numeric.

Irregular component The component of the time series that reflects the random variation of the time series values beyond what can be explained by the trend, cyclical, and seasonal components.

***i*th residual** The difference between the observed value of the dependent variable and the value predicted using the estimated regression equation; for the ith observation the ith residual is $y_i - \hat{y}_i$.

Joint probability The probability of two events both occurring; that is, the probability of the intersection of two events.

Judgement sampling A non-probabilistic method of sampling whereby element selection is based on the judgement of the person doing the study.

Kruskal-Wallis test A non-parametric test for identifying differences among three or more populations on the basis of independent samples.

Laspeyres price index A weighted aggregate price index in which the weight for each item is its base-period quantity.

Least squares method The method used to develop the estimated regression equation. It minimizes the sum of squared residuals (the deviations between the observed values of the dependent variable, y_i, and the estimated values of the dependent variable, $\hat{y}_i$).

Level of significance The probability of making a Type I error when the null hypothesis is true as an equality.

Leverage A measure of how far the values of the independent variables are from their mean values.

Logistic regression equation The mathematical equation relating $E(y)$, the probability that $y = 1$, to the values of the independent variables; that is,

$$E(y) = P(y = 1 \mid x_1, x_2, \ldots x_p) = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p}}$$

Logit The natural logarithm of the odds in favour of $y = 1$; that is, $g(x_1, x_2, \ldots x_p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \cdots + \beta_p x_p$

Lot A group of items such as incoming shipments of raw materials or purchased parts as well as finished goods from final assembly.

Mann-Whitney-Wilcoxon (MWW) test A non-parametric statistical test for identifying differences between two populations based on the analysis of two independent samples.

Margin of error The $\pm$ value added to and subtracted from a point estimate in order to construct an interval estimate of a population parameter.

Marginal probability The values in the margins of a joint probability table that provide the probabilities of each event separately.

Matched Samples Where, e.g., only are sample of workers is selected and each worker uses first one and then the other method, with each worker providing a pair of data values.

Mean A measure of central location computed by summing the data values and dividing by the number of observations.

Mean squared error (MSE) A measure of the accuracy of a forecasting method. This measure is the average of the sum of the squared differences between the forecast values and the actual time series values.

Median A measure of central location provided by the value in the middle when the data are arranged in ascending order.

Mode A measure of location, defined as the value that occurs with greatest frequency.

Moving averages A method of forecasting or smoothing a time series that uses the average of the most recent n data values in the time series as the forecast for the next period.

Multicollinearity The term used to describe the correlation among the independent variables.

Multinomial population A population in which each element is assigned to one and only one of several categories. The multinomial distribution extends the binomial distribution from two to three or more outcomes.

Multiple coefficient of determination A measure of the goodness of fit of the estimated multiple regression equation. It can be interpreted as the proportion of the variability in the dependent.

Multiple comparison procedures Statistical procedures that can be used to conduct statistical comparisons between pairs of population means.

Multiple regression analysis Regression analysis involving two or more independent variables.

Multiple regression equation The mathematical equation relating the expected value or mean value of the dependent variable to the values of the independent variables; that is $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$.

Multiple regression model The mathematical equation that describes how the dependent variable y is related to the independent variables $x_1, x_2, \ldots x_p$ and an error term ε.

Multiple sampling plan A form of acceptance sampling in which more than one sample or stage is used. On the basis of the number of defective items found in a sample, a decision will be made to accept the lot, reject the lot, or continue sampling.

Multiplication law A probability law used to compute the probability of the intersection of two events. It is $P(A \cap B) = P(B)P(A \mid B)$ or $P(A \cap B) = P(A)P(B \mid A)$. For independent events it reduces to $P(A \cap B) = P(A)P(B)$.

Multiplicative time series model A model whereby the separate components of the time series are multiplied together to identify the actual time series value. When the four components of trend, cyclical, seasonal, and irregular are assumed present, we obtain $Y_t = T_t \times C_t \times S_t \times I_t$. When the cyclical component is not modelled, we obtain $Y_t = T_t \times S_t \times I_t$.

Mutually exclusive events Events that have no sample points in common; that is, $A \cap B$ is empty and $P(A \cap B) = 0$.

Node An intersection or junction point of an influence diagram or a decision tree.

Nominal scale The scale of measurement for a variable when the data use labels or names to identify an attribute of an element. Nominal data may be non-numeric or numeric.

Non-parametric methods Statistical methods that require relatively few assumptions about the population probability distributions and about the level of measurement. These methods can be applied when nominal or ordinal data are available.

Non-probabilistic sampling Any method of sampling for which the probability of selecting a sample of any given configuration cannot be computed.

Non-sampling error All types of errors other than sampling error, such as measurement error, interviewer error, and processing error.

Normal probability distribution A continuous probability distribution. Its probability density function is bell shaped and determined by its mean μ and standard deviation σ.

Normal probability plot A graph of the standardized residuals plotted against values of the normal scores. This plot helps determine whether the assumption that the error term has a normal probability distribution appears to be valid.

***np* chart** A control chart used to monitor the quality of the output of a process in terms of the number of defective items.

Null hypothesis The hypothesis tentatively assumed true in the hypothesis testing procedure.

Observation The set of measurements obtained for a particular element.

Odds in favour of an event occurring The probability the event will occur divided by the probability the event will not occur.

Odds ratio The odds that $y = 1$ given that one of the independent variables increased by one unit (odds_1) divided by the odds that $y = 1$ given no change in the values for the independent variables (odds_0); that is, Odds ratio $= \text{odds}_1/\text{odds}_0$.

Ogive A graph of a cumulative distribution.

One-tailed test A hypothesis test in which rejection of the null hypothesis occurs for values of the test statistic in one tail of its sampling distribution.

Operating characteristic curve A graph showing the probability of accepting the lot as a function of the percentage defective in the lot. This curve can be used to help determine whether a particular acceptance sampling plan meets both the producer's and the consumer's risk requirements.

Ordinal scale The scale of measurement for a variable if the data exhibit the properties of nominal data and the order or rank of the data is meaningful. Ordinal data may be non-numeric or numeric.

Outlier A data point or observation that does not fit the trend shown by the remaining data, often unusually small or unusually large.

***p* chart** A control chart used when the quality of the output of a process is measured in terms of the proportion defective.

***p*-value** A probability, computed using the test statistic, that measures the support (or lack of support) provided by the sample for the null hypothesis. For a lower tail test, the p-value is the probability of obtaining a value for the test statistic at least as small as that provided by the sample. For an

upper tail test, the *p*-value is the probability of obtaining a value for the test statistic at least as large as that provided by the sample. For a two-tailed test, the *p*-value is the probability of obtaining a value for the test statistic at least as unlikely as that provided by the sample.

Paasche price index A weighted aggregate price index in which the weight for each item is its current-period quantity.

Parameter A numerical characteristic of a population, such as a population mean μ, a population standard deviation σ, a population proportion π, and so on.

Partitioning The process of allocating the total sum of squares and degrees of freedom to the various components.

Payoff A measure of the consequence of a decision such as profit, cost, or time. Each combination of a decision alternative and a state of nature has an associated payoff (consequence).

Payoff table A tabular representation of the payoffs for a decision problem.

Percentage frequency distribution A tabular summary of data showing the percentage of items in each of several non-overlapping classes.

Percentile A value such that at least p per cent of the observations are less than or equal to this value and at least $(100 - p)$ per cent of the observations are greater than or equal to this value. The 50th percentile is the median.

Pie chart A graphical device for presenting data summaries based on subdivision of a circle into sectors that correspond to the relative frequency for each class.

Point estimate The value of a point estimator used in a particular instance as an estimate of a population parameter.

Point estimator The sample statistic, such as $\bar{x}$, s, or p, that provides the point estimate of the population parameter.

Poisson probability distribution A probability distribution showing the probability of x occurrences of an event over a specified interval of time or space.

Poisson probability function The function used to compute Poisson probabilities.

Pooled estimator of π A weighted average of p_1 and p_2.

Population The set of all elements of interest in a particular study.

Population parameter A numerical value used as a summary measure for a population (e.g. the population mean μ, the population variance σ^2, and the population standard deviation σ).

Posterior probabilities Revised probabilities of events based on additional information.

Posterior (revised) probabilities The probabilities of the states of nature after revising the prior probabilities based on sample information.

Power The probability of correctly rejecting H_0 when it is false.

Power curve A graph of the probability of rejecting H_0 for all possible values of the population parameter not satisfying the null hypothesis. The power curve provides the probability of correctly rejecting the null hypothesis.

Prediction interval The interval estimate of an individual value of y for a given value of x.

Price relative A price index for a given item that is computed by dividing a current unit price by a base-period unit price and multiplying the result by 100.

Prior probabilities The probabilities of the states of nature prior to obtaining sample information.

Probabilistic sampling Any method of sampling for which the probability of each possible sample can be computed.

Probability A numerical measure of the likelihood that an event will occur.

Probability density function A function used to compute probabilities for a continuous random variable. The area under the graph of a probability density function over an interval represents probability.

Probability distribution A description of how the probabilities are distributed over the values of the random variable.

Probability function A function, denoted by $p(x)$, that provides the probability that x assumes a particular value for a discrete random variable.

Producer Price Index A price index designed to measure changes in prices of goods sold in primary markets (i.e. first purchase of a commodity in non-retail markets).

Producer's risk The risk of rejecting a good-quality lot; a Type I error.

Qualitative data Labels or names used to identify an attribute of each element. Qualitative data use either the nominal or ordinal scale of measurement and may be non-numeric or numeric.

Qualitative independent variable An independent variable with qualitative data.

Qualitative variable A variable with qualitative data.

Quality control A series of inspections and measurements that determine whether quality standards are being met.

Quantitative data Numeric values that indicate how much or how many of something.

Quantitative variable A variable with quantitative data.

Quantity index An index designed to measure changes in quantities over time.

Quartiles The 25th, 50th and 75th percentiles, referred to as the first quartile, the second quartile (median), and third quartile, respectively. The quartiles can be used to divide a data set into four parts, with each part containing approximately 25 per cent of the data.

***R* chart** A control chart used when the quality of the output of a process is measured in terms of the range of a variable.

Random variable A numerical description of the outcome of an experiment.
Randomized block design An experimental design employing blocking.
Range A measure of variability, defined to be the largest value minus the smallest value.
Ratio scale The scale of measurement for a variable if the data demonstrate all the properties of interval data and the ratio of two values is meaningful. Ratio data are always numeric.
Regression equation The equation that describes how the mean or expected value of the dependent variable is related to the independent variable; in simple linear regression, $E(y) = \beta_0 + \beta_1 x$
Regression model The equation describing how y is related to x and an error term; in simple linear regression, the regression model is $y = \beta_0 + \beta_1 x + \varepsilon$
Relative frequency distribution A tabular summary of data showing the fraction or proportion of data items in each of several non-overlapping classes.
Relative frequency method A method of assigning probabilities that is appropriate when data are available to estimate the proportion of the time the experimental outcome will occur if the experiment is repeated a large number of times.
Replications The number of times each experimental condition is repeated in an experiment.
Residual analysis The analysis of the residuals used to determine whether the assumptions made about the regression model appear to be valid. Residual analysis is also used to identify outliers and influential observations.
Residual plot Graphical representation of the residuals that can be used to determine whether the assumptions made about the regression model appear to be valid.

Sample A subset of the population.
Sample information New information obtained through research or experimentation that enables an updating or revision of the state-of-nature probabilities.
Sample point An element of the sample space. A sample point represents an experimental outcome.
Sample space The set of all experimental outcomes.
Sample statistic A numerical value used as a summary measure for a sample (e.g. the sample mean $\bar{x}$, the sample variance s^2, and the sample standard deviation s).
Sample survey A survey to collect data on a sample.
Sampled population The population from which the sample is taken.
Sampling distribution A probability distribution consisting of all possible values of a sample statistic.
Sampling error The error that occurs because a sample, and not the entire population, is used to estimate a population parameter.
Sampling unit The units selected for sampling. A sampling unit may include several elements.
Sampling with replacement Once an element has been included in the sample, it is returned to the population. A previously selected element can be selected again and therefore may appear in the sample more than once.
Sampling without replacement Once an element has been included in the sample, it is removed from the population and cannot be selected a second time.
Scatter diagram A graphical presentation of the relationship between two quantitative variables. One variable is shown on the horizontal axis and the other variable is shown on the vertical axis.
Scenario writing A qualitative forecasting method that consists of developing a conceptual scenario of the future based on a well-defined set of assumptions.
Seasonal component The component of the time series that shows a periodic pattern over one year or less.
Serial correlation Same as autocorrelation.
σ (sigma) known The condition existing when historical data or other information provide a good estimate or value for the population standard deviation prior to taking a sample. The interval estimation procedure uses this known value of σ in computing the margin of error.
σ (sigma) unknown The condition existing when no good basis exists for estimating the population standard deviation prior to taking the sample. The interval estimation procedure uses the sample standard deviation s in computing the margin of error.
Sign test A non-parametric statistical test for identifying differences between two populations based on the analysis of nominal data.
Simple linear regression Regression analysis involving one independent variable and one dependent variable in which the relationship between the variables is approximated by a straight line.
Simple random sampling Finite population: a sample selected such that each possible sample of size n has the same probability of being selected. Infinite population: a sample selected such that each element comes from the same population and the elements are selected independently.
Simpson's paradox Conclusions drawn from two or more separate cross-tabulations that can be reversed when the data are aggregated into a single cross-tabulation.
Single-factor experiment An experiment involving only one factor with k populations or treatments.
Skewness A measure of the shape of a data distribution. Data skewed to the left result in negative skewness; a symmetric data distribution results in zero skewness; and data skewed to the right result in positive skewness.
Smoothing constant A parameter of the exponential smoothing model that provides the weight given to the most recent time series value in the calculation of the forecast value.
Spearman rank-correlation coefficient A correlation measure based on rank-ordered data for two variables.

Standard deviation A measure of variability computed by taking the positive square root of the variance.
Standard error The standard deviation of a point estimator.
Standard error of the estimate The square root of the mean square error, denoted by s. It is the estimate of σ, the standard deviation of the error term ε.
Standard normal probability distribution A normal distribution with a mean of zero and a standard deviation of one.
Standardized residual The value obtained by dividing a residual by its standard deviation.
States of nature The possible outcomes for chance events that affect the payoff associated with a decision alternative.
Statistical inference The process of using data obtained from a sample to make estimates or test hypotheses about the characteristics of a population.
Statistics The art and science of collecting, analyzing, presenting, and interpreting data.
Stem-and-leaf display An exploratory data analysis technique that simultaneously rank orders quantitative data and provides insight about the shape of the distribution.
Stratified random sampling A probabilistic method of selecting a sample in which the population is first divided into strata and a simple random sample is then taken from each stratum.
Studentized deleted residuals Standardized residuals that are based on a revised standard error of the estimate obtained by deleting observation i from the data set and then performing the regression analysis and computations.
Subjective method A method of assigning probabilities on the basis of judgement.
Systematic sampling A method of choosing a sample by randomly selecting the first element and then selecting every kth element thereafter.

***t* distribution** A family of probability distributions that can be used to develop an interval estimate of a population mean whenever the population standard deviation σ is unknown and is estimated by the sample standard deviation s.
Target population The population about which inferences are made.
Test statistic A statistic whose value helps determine whether a null hypothesis can be rejected.
Time series A set of observations on a variable measured at successive points in time or over successive periods of time.
Time series data Data collected over several time periods.
Treatments Different levels of a factor.
Tree diagram A graphical representation that helps in visualizing a multiple-step experiment.
Trend The long-run shift or movement in the time series observable over several periods of time.
Trend line A line that provides an approximation of the relationship between two variables.
Two-tailed test A hypothesis test in which rejection of the null hypothesis occurs for values of the test statistic in either tail of its sampling distribution.
Type I error The error of rejecting H_0 when it is true.
Type II error The error of accepting H_0 when it is false.

Unbiasedness A property of a point estimator that is present when the expected value of the point estimator is equal to the population parameter it estimates.
Uniform probability distribution A continuous probability distribution for which the probability that the random variable will assume a value in any interval is the same for each interval of equal length.
Union of *A* and *B* The event containing all sample points belonging to A or B or both. The union is denoted $A \cup B$.

Variable A characteristic of interest for the elements.
Variable selection procedures Methods for selecting a subset of the independent variables for a regression model.
Variance A measure of variability based on the squared deviations of the data values about the mean.
Variance inflation factor A measure of how correlated an independent variable is with all other independent predictors in a multiple regression model.
Venn diagram A graphical representation for showing symbolically the sample space and operations involving events in which the sample space is represented by a rectangle and events are represented as circles within the sample space.

Weighted aggregate price index A composite price index in which the prices of the items in the composite are weighted by their relative importance.
Weighted mean The mean obtained by assigning each observation a weight that reflects its importance.
Weighted moving averages A method of forecasting or smoothing a time series by computing a weighted average of past data values. The sum of the weights must equal one.
Wilcoxon signed-rank test A non-parametric statistical test for identifying differences between two populations based on the analysis of two matched or paired samples.

$\bar{x}$ chart A control chart used when the quality of the output of a process is measured in terms of the mean value of a variable such as a length, weight, temperature and so on.

***z*-score** A value computed by dividing the deviation about the mean $(x_i - \bar{x})$ by the standard deviation s. A z-score is referred to as a standardized value and denotes the number of standard deviations x_i is from the mean.