Advanced Analyses Glossary

This document contains glossary information for the advanced analyses in STATGRAPHICS *Plus* Quality and Design, and STATGRAPHICS *Plus* Professional. To facilitate searching for terms while working with a specific procedure, the glossary entries are grouped by analysis in the order they are listed on the Special menu:

Quality Control

Acceptable Quality Level (AQL)
The poorest level of quality which the consumer finds acceptable on average.

Acceptance Charts
A charting method for instances where the specification limits for the process are wide compared to the process performance. Rather than placing the control limits at ± 3 sigma, the control limits are placed far enough within the specification limits to ensure that a signal is generated if the process is going out of control, but otherwise allows the process to vary around the long-term process mean.

Acceptance Sampling
Acceptance Sampling is a procedure for determining sample sizes and decision rules for incoming or outgoing product inspection.

Alpha Risk
The probability of a Type I error; determining the probability of declaring a process out-of-control when, in fact, it is not.

ARIMA Chart
An Autoregressive Individuals Moving Average (ARIMA) Chart is a modification of standard control charts to account for the serial correlation between consecutive periods with the objective of separating long-term changes from short-term dependencies.

Attribute
An inherent property usually measured by a finite number or a discrete set of values.

Attributes control chart
One of four control charts (*p*, *np*, *c*, or *u*) that allows you to construct control limits based on a particular attribute of a sample.

Average
A measure of the central tendency of the data. Also known as the mean.

Average Outgoing Quality (AOQ) Curve
This shows the average percent of defective items shipped as a function of the proportion of non-conforming items in the lot when subjected to the current sampling plan.

Average Outgoing Quality Limit (AOQL)
Based on the assumption that rejected lots are 100% inspected and that any non-conforming items are replaced by good items, this is the maximum fraction of non-conforming items which are accepted.

Beta Risk
The probability of a Type II error; determining the probability of declaring a process in control when it is not.

***c* Chart**
An attributes control chart, based on the Poisson distribution, that allows you to construct control limits based on an attribute related to frequency, such as the number of defects in an item.

Capability Indices
Statistics that indicate the inherent capability of a process; for example, C_p and C_{pk}.

Chi-Square Statistic
A statistic that lets you compare observed to expected frequencies. If none of the expected frequencies is too small, the statistic follows a chi-square distribution. A large chi-square statistic leads to a conclusion that two distributions are different. A small chi-square statistic leads to a conclusion that two distributions are similar. Also known as a Portmanteau statistic.

Confidence Interval
An interval that indicates the bounds between which it is likely that the real parameter lies.

Confidence Limits
The end points of a confidence interval. Common limits are 90, 95, and 99.

Control Chart
A graph of results from sampling the product of a manufacturing process; consists of a horizontal line that represents the expected mean of some characteristic of quality and two lines on either side that represent the extent you are allowed to sample random product deviations.

Controlled Process
A process for which a control chart shows no points outside the limits and no nonrandom variation within the limits.

Control Limits
On a control chart, the lines above and below the center horizontal line that designate the area into which a stated proportion of the sample task should fall.

Control Mean
The target value for the process mean.

Control-to-Standard
When generating control charts for attributes or variables, a type of study that produces a chart based on a process mean and sigma you specify. In this type of study, the data do not influence the control limits.

Cumulative Sum (CuSum) Chart (V-Mask and H-K)
One of three time-weighted charts based on subgroup statistics or observations. This chart helps to ensure that you do not specify a process as out of control when it is in control or one in which you do not fail to detect a marked change in the mean. It is useful for detecting small, persistent shifts in a process.

Density Function
A function that gives the rate at which the cumulative probability increases at a given point.

Exponentially Weighted Moving Average (EWMA)
The weighted average of the current observation and the estimate of the previous observation. To calculate the weighted average, the system multiplies the current observation by lambda and multiplies the estimate of the previous observation by 1 - lambda.

Exponentially Weighted Moving Average (EWMA) Chart
One of three time-weighted charts that plots the exponentially weighted moving average. This chart helps to ensure that you do not specify a process as out of control when it is in control, or one in which you do not fail to detect a marked change in the mean. It is useful for detecting small, persistent shifts in a process. This chart is not sensitive to the normality assumption and, therefore, is an ideal chart to use with individual data.

Fishbone Diagram
A cause and effect diagram to illustrate sources of nonconformance through a visualization of the interrelationships among factors.

Hotelling T-Squared Statistic
A statistic for testing multivariate normal mean vectors. In quality control, you use the Hotelling T-squared statistic to identify unusual observations or out-of-control points on a multivariate control chart.

Individuals Chart
A variables control chart that plots subgroups of size 1. You use this chart when measurements are difficult or expensive to obtain.

Initial Study
When generating control charts for attributes or variables, a type of study that produces a chart with control limits estimated from the data.

Kolmogorov-Smirnov One-Sample Test
A test that allows you to test the overall goodness-of-fit between the distribution of the data and the distribution you select.

Kurtosis
A measurement of how flat or steep the distribution of the data is with respect to a normal distribution. A positive kurtosis value usually represents a flatter then normal curve, whereas a negative value reflects a distribution with a higher than normal peak.

Lambda
The weighting factor for the exponentially weighted moving average. The larger lambda, the higher the weighting factor of the most recent data.

Lower Control Limit (LCL)
Control limit for points that plot below the centerline; used to determine if a process is in control.

Lower Specification Limit (LSL)
A specification limit that defines the lower conformance boundary for an individual unit of a manufacturing or service operation.

Mean
The sum of the observations divided by the number of observations. Also known as the average.

Mean Range
The average of subgroup ranges that is the centerline of a range chart. You use the mean range to estimate the process standard deviation.

Mean Sigma
The average of subgroup standard deviations that is the centerline of an S chart. You use the mean sigma to estimate the process standard deviation.

Mean Variance
The average of the variances for subgroups of data when the subgroup sizes are equal. You use the mean variance to determine the centerline for an S-squared chart.

Moving Average
The average of contiguous subgroup averages.

Moving Average (MA) Chart
One of three time-weighted charts that plots the moving averages of the subgroups. This chart helps to ensure that you do not specify a process as out of control when it is in control, or one in which you do not fail to detect a marked change in the mean. It is useful for showing trends in subgroup averages.

Moving Range
The absolute value of the difference between the range of a given subgroup and the range of a previous subgroup.

Moving Range (MR) Chart
A control chart that plots the moving ranges of the subgroups. This plot is necessary when plotting individual data (subgroup size equal to 1).

MR(2)
The moving range calculated by taking the difference between every pair of contiguous data values.

MR(2) Chart
A control chart that plots the moving ranges for each pair of subgroup values. This chart is preferred for individual data (subgroup size equal to 1).

Multivariate
An analysis that involves more than one variable.

Multivariate Control Chart
A control chart for multivariate samples in which you can correlate different measurements.

Nominal (target) Value
The preferred value for an item or data value that results from a process. Typically, the nominal value falls between the lower specification limit (LSL) and the upper specification limit (USL).

Normal Distribution
A common bell-shaped curve. A property of this distribution is that 68 percent of the data fall within plus or minus 1 sigma; 95 percent fall within plus or minus 2 sigma.

Normal Probability Plot
A plot with an arithmetic (interval) horizontal axis and a vertical axis scaled so the cumulative distribution function (cdf) of a normal distribution plots as a straight line. If observations fall close to the straight line, the residuals are normally distributed. You can also use this plot to identify outliers in the data.

***np* Chart**
An attributes control chart that allows you to analyze whether a process produces an unacceptably high proportion of defective items, based on the number of defective items in a sample or series of samples. You use a *c* or *u* chart if the proportion of defects is small.

Observation
The result of determining the presence or absence of attributes, or the result of measuring a variable.

Operating Characteristic (OC) Curve
A curve that gives the probability of accepting a process as a function of the process parameter level.

Out-of-Control Process
A process that is not in statistical control. Indications of this condition include subgroup values that plot beyond control limits on a control chart, or subgroup values that form nonrandom patterns.

***p* Chart**
An attributes control chart, based on a binomial distribution, that allows you to analyze the proportion of defective items that a process produces.

Pareto Chart
A frequency histogram of attribute data arranged by category. You use this chart to identify the defects that most frequently occur.

Process Average (mean)
The average fraction nonconforming, or mean value of a quality characteristic, of the output of a process.

Process Capabilities
The actual or potential capabilities of a process with respect to meeting specifications. Interest may center on the standard deviation of the process, ultimately on the percentage that does not conform.

Process Mean
The assumed or target average for a process.

Process Sigma
The assumed or target standard deviation for a process.

Quality Control
The operational techniques and activities that sustain the quality of a product or service that satisfies given needs.

Range
The difference between the largest and smallest observed values in a sample.

Range (R) Chart
A control chart you use to maintain a check on the variability of the quality of a particular product or process.

Repeatability
The variation in measurement obtained with one gage when one operator uses it several times to measure the identical characteristic on the same part. Also known as gage repeatability.

Reproducibility
The variation in measurements obtained when several operators use one gage to measure the identical characteristic on the same part. Also known as gage reproducibility.

Runs Test
A statistical test you use to check for patterns in subgroup statistics.

S Chart
A variables control chart with control limits based on the standard deviations of the subgroups.

S-Squared Chart
A variables control chart with control limits based on the variances of the subgroups.

Sample
A set of objects or things from a larger set called a universe.

Sample Mean
The average from a set of data, usually noted as X-bar.

Shewhart Control Chart
Variables and attributes control charts with superimposed upper and lower control limits.

Sigma
The standard deviation of a population.

Sigma Chart
A quality control chart of the subgroup standard deviations.

Sigma Limits
Boundaries you determine by adding and subtracting one or more standard deviation from the average.

Specification Limits
Limits that define the conformance boundaries for an individual unit of a manufacturing or service operation.

Standard Deviation
The measurement of the spread or dispersion of the data. A statistic that defines the average size of a deviation of a single observation from its expected value. Also known as a sigma.

Standard Mean
The assumed or target average.

Standard Sigma
The assumed or target standard deviation.

Time-Weighted Charts
Control charts that incorporate the past history of a process to estimate the charted point. They are generally more sensitive to small, gradual changes in the process than Shewhart charts.

Tolerance Limits
Limits that confine the conformance boundaries for an individual unit of a manufacturing or service operation.

Toolwear Chart
A modification of the standard control chart to address situations where the mean of the process follows a linear trend rather than remaining constant. These charts were originally developed to monitor the wear of tools, where the measurements were expected to follow a natural trend.

Type I Error
The error of rejecting a hypothesis when it is true. Also known as an alpha error.

Type II Error
The error of accepting a hypothesis when it is not true. Also called a beta error.

***u* Chart**
An attributes control chart that allows you to construct control limits based on an attribute that is related to frequency, such as the number of defects per item. This chart is similar to a *c* chart except that a *u* chart expresses rates.

Upper Control Limit (UCL)
The control limit for points that plot above the centerline; used to determine if a process is in control.

Upper Specification Limit (USL)
A specification limit that defines the upper conformance boundary for an individual unit of a manufacturing or service operation.

V-mask
A V-shaped mask you use with a CuSum Chart. When cumulative sums cross the "arms" of the V-mask, you consider the process to be out of control. You should apply the V-mask to each new observation after it is placed on a graph.

Variables Control Chart
A control chart for plotting central tendency and variability. This type of control chart provides more information about process performance and establishes more efficient control procedures that attributes control charts.

Variation
A quantifiable measure of the deviations from a target value.

Warning Limits
The limits on a control chart, usually drawn at 2 sigma. If a process exceeds the warning limits, you should include more data for study.

Weibull Distribution
A distribution that is an appropriate model for product failures because its failure rate curves can take various shapes. This distribution is a generalization of the exponential distribution.

Weighted Average
The result of multiplying each data value in a variable by its associated weighting factor and dividing the total by the sum of the weighting factors.

X-Bar
The mean of a sample.

X-Bar and R Chart
A variables control chart with control limits based on the means and ranges of the subgroups in the data.

X-Bar and S Chart
A variables control chart with control limits based on the means and sigmas (standard deviations) of the subgroups in the data.

X-Bar and S-squared Chart
A variables control chart with control limits based on the means and variances of the subgroups in the data.

X-Bar Chart
A quality control chart that plots the averages of collected subgroup data.

z-Score
The most commonly used standard score; the mean is 0, the standard deviation is 1.

Design of Experiments

2-Level Design
An experiment where all the factors are set at one of two levels: low or high. For example, -1, +1 or 1, 2.

3-Level Design
An experiment where all the factors are set at one of three levels: low, medium, or high. For example, -1, 0, 1 or 1, 2, 3.

Aliased Effects
The effects of two or more factors that cannot be separated and assigned. See also *confounding effects*.

Aliases
Two or more main or interaction factors whose effects are estimated by the same contrast; because of this the effects cannot be estimated separately.

Aliasing
Confusing two or more factors so their effects cannot be separated; when two factors are set at the same levels throughout the experiment; that is, the columns are 100 percent correlated. See also *confounding*.

ANOVA
Analysis of Variance. A statistical tool based on F-ratios that measures whether a factor contributes significantly to the variance of a response. Also determines the amount of variance that is due to pure error.

Arrays
Inner/Outer Arrays are the STATGRAPHICS implementation of designs that follow the Taguchi approach to solving problems by identifying levels of the controllable factors where the response variables are insensitive to normal fluctuations in the uncontrollable noise to which every process is subjected.

Average Run Length (ARL)
ARL is the average number of points which must be obtained until a point falls outside the control limits.

Axial Distance
In a central composite design, the distance from the center of a cube to a point in the star portion of the design.

Balanced Design
A two-level experimental design that is balanced if each factor is run the same number of times at the high and low ends.

Balanced Incomplete Block (BIB) Designs
Balanced Incomplete Block (BIB) Designs are designs in which only *k* levels of a factor can be run within each block, where *k*<*q*. STATGRAPHICS creates two types of BIB designs: The combinatoric designs involve blocks where all possible combinations of *k* treatments are run; and small BIB designs with fewer blocks where all pairs of the treatments occur together in the same block the same number of times.

Balanced One-Way Design
A one-factor experimental design where each factor level is run an equal number of times.

Balanced Randomized Block Design
An experimental design where two kinds of effects are considered: the treatments, which are of most interest, and a nuisance factor, which is blocked to eliminate its effect. The nuisance factor can be run randomly within each block to protect against unknown sources of bias. The design is balanced when all the treatment factor levels are run the same number of times within each block.

Block
A homogenous group of tests or experiments. Blocks might be created when it is necessary to include batches of raw material, different laboratories, or different experimenters in the design.

Block effect
The change in the average responses between blocks. The influence (variability) of the blocking can be removed from the estimates of the other effects by estimating any of the block effects. See also *main effect*.

Blocked Design
An experimental design where the respondents are grouped into categories (blocks). Each block can then be treated as a unit in the analysis.

Blocking Variable
A variable (factor) that cannot be randomized. The experiment is usually run in blocks for each level of the blocking variable and randomization is performed within the blocks.

Box-Behnken Design
A 3-level design used for quantitative factors and designed to estimate all the main, quadratic, and two-way interaction effects. The design is a combination of two-level factorials and incomplete block designs.

Brainstorming
A group activity that generates a list of possible factors and levels, and the method by which the results can be evaluated.

Causality
The assertion that changes to an input factor will directly result in a specified change in an output.

Centerpoint
A design point at which all the factors are run halfway between their high and low levels; all the factors must be continuous variables. Centerpoints can be used to add additional runs to experiments, to check for curvature in screening designs, and, if replicated, to estimate pure error. Also known as the *design centerpoint*.

Central Composite Design
A 3-level design that starts with a 2-level factorial and some centerpoints. Used typically for quantitative factors and designed to estimate all the main effects plus the desired quadratics and two-way interactions. The design has two parts: a factorial and a star. The star portion of the design consists of an additional set of points arranged at equal distances from the center of the cube on radii that pass through the centerpoint in the face of the cube.

Confidence Interval
A range of values based on a sample mean and standard deviation that has a given probability of containing the true population parameter.

Confounded Effects
The effects of two or more factors that cannot be separated and assigned; the result of the factors being set at the same level throughout the experiment. Also known as *aliased effects*.

Confounding
Confusing two or more factors so their effects cannot be separated. Also known as *aliasing*.

Contour Plot
A plot that represents a two-dimensional grid surface similar to a topographical map. In experimental design, the contours represent the estimated level of the response variable.

Controllable Factors
Factors the experimenter has control of during all phases of the experiment; that is, experimental, production, and operational phases.

Cube Plot
A plot of the estimated responses for three factors.

Curvature
The degree of curving for a line or surface.

Curvature Check
A test to determine whether curvature (quadratic) terms should be included in the response model.

D-Optimal Designs
D-Optimal Designs are procedures for reducing the number of runs in an experimental design. The process takes a set of *n* candidate runs and reduces it to the minimum number of runs required to fit a specific model.

Daniel Plot
A plot that identifies estimated effects that are large relative to the noise in an experiment. Large effects appear on the right side of the plot. Also known as a *half-normal plot.*

Defining Relationship
A statement of one or more factor word equalities used to determine the aliasing structure in a fractional factorial design.

Degrees of Freedom (df)
For a fitted model, the number of independent observations minus the number of estimated parameters. For a factor, the number of levels minus 1.

Design Matrix
A matrix that represents the experimental settings. Usually contains values that range from -1 to 1 but that could be wider. The rows represent the runs; columns represent the factors.

Design
The complete specification of experimental runs, including blocking, randomization, replication, and the assignment of factor-level combinations.

Design Generators
Equations that indicate the columns that must be multiplied to produce the last columns in a fractional factorial design.

Design Point
An intended experimental run.

Design Resolution (III, IV, V, or V+)
The degree of confounding in a two-level fractional design (screening design). A design of resolution III does not confound main effects with one another, but does confound main effects with two-factor interactions. A resolution IV design does not confound main effects and two-factor interactions, but does confound two-factor interactions with other two-factor interactions. A resolution V design does not confound main effects and two-factor interactions with each other, but does confound two-factor interactions with three-factor interactions.

Draper-Lin Small Composite Design
A central composite design that runs a fraction rather than a full factorial. Also known as a *Draper-Lin design*.

Effect
The change in the average of the responses between two factor-level combinations or two experimental settings. For a two-level factor, the effect of the factor is the average response at the high level of the factor minus the average of the responses at the low level of the factor.

Empirical Data
Data that are based on observation or experience.

Empirical Model
A model that uses sequential experimentation techniques to survey a domain of interest and to focus on the most important variables and their effects.

Experimental Condition
A specific combination of factors and levels that is studied in an experiment. Also known as *experimental trials.* See also *Runs*.

Experimental Design
The complete specification of experimental runs, including blocking, randomization, replication, and the assignment of factor-level combinations.

Experimental Factor
An independent variable set at different levels in an experimental design.

Experimental Region
All the factor-level combinations for which experimentation is possible.

Extreme Vertices Designs
Extreme Vertices Designs are

F-Ratio (F)
A ratio of the variance explained by a factor (MS of the factor) to the unexplained variance (MSE). If there is no effect, the associated p-value is close to 1.

Face-Centered Design
A central composite design with three levels rather than five, and with star points (other than centerpoints) in the center of the faces of the factorial cube.

Factor
An input to a process that can be manipulated during experimentation.

Factor Effects
A factor's main effect and all the interactions that involve that factor.

Factor Level
A given value, a specification of procedure, or specific setting of a factor.

Factorial Design
A design that combines the levels for each factor with all the levels for every other factor. The number of experimental runs can be reduced by fractioning the design.

Fitted Model
Any model type that is used to estimate response values from specified independent variable values.

Foldover Design
A way to obtain a resolution IV design based on two designs of resolution III (R_{III}). Used when confirmation runs from a R_{III} design differ substantially from their prediction, and the experimenter wants to de-alias the two-way interactions from the main effects.

Fractional Factorial Design
Instead of using a full factorial, the experimenter uses some subset of it assuming that some interactions will not occur and that a factor has been assigned to that interaction column of the design. A design that does not specify all the combinations of the factors.

Full Factorial Design
A design that combines the levels for each factor with all the levels for every other factor.

Graeco-Latin Square Designs
Designs in which treatments are balanced across three blocking factors.

Half-normal Plot
A plot that identifies estimated effects that are large relative to the noise in an experiment. Large effects appear on the right side of the plot. Also known as the *Daniel plot*.

Inner/Outer Arrays
Inner/Outer Arrays are the STATGRAPHICS implementation of designs that follow the Taguchi approach to solving problems by identifying levels of the controllable factors where the response variables are insensitive to normal fluctuations in the uncontrollable noise to which every process is subjected.

Interaction
A change in the response due to the combination of two or more factors. An interaction involving two factors is known as a *two-factor interaction;* three factors as a *three-factor interaction,* and so on.

Interaction Effect
The influence of two or more interacting factors on the response when they are changed from one level to another.

Latin Square Designs
Designs in which treatments are balanced across two blocking factors, but where the number of runs is less than would be required by a full multi-factorial design.

Level
The setting or value of a factor.

Main Effect
A measure that estimates the influence of a single factor on a response when the factor is changed from one level to another.

Metric Variable
A variable that measures an interval or ratio scale.

Mixture Experiment
An experiment where it is assumed that the response depends only on the relative proportions of the ingredients (components) in the mixture and not on the amount of the mixture.

Model Fitting
The method of using an equation to quantify a mathematical relationship between observed data and variables or combinations of variables.

Multi-Factor Categorical Designs
A design for experiments where several categorical factors are to be varied across multiple levels. The design incorporates all combinations of the factors.

Multilevel Factorial Design
Designs involving between 2 and 8 quantitative factors, each of which is set at an arbitrary number of levels. Among the uses of this design is providing candidate runs for augmenting *undesigned* experiments where the experimenter may take previous test data generated in a non-structured manner and incorporate them into a formal experimental design.

Multiple Regression
A Taylor series model that uses several independent variables to predict one dependent variable.

Noise
Any unexplained or random variability in the response.

Nonmetric Variable
A variable that contains data that are not measured on an interval or ratio scale; a variable that contains ordinal data.

Normal Distribution
The “bell-shaped” curve distribution used to calculate probabilities of events that tend to occur around a mean value and trail off with decreasing likelihood.

Normal Plot
A plot that identifies estimated effects that are large relative to the noise in an experiment. STATGRAPHICS *Plus* for Windows builds this plot as if there were no important effects so all the points fall on a straight line. Any points that fall away from the line indicate real effects. A way to identify significant effects when ANOVA is not possible due to unreplicated data.

Nuisance Variables
Factors that are not included or blocked in a design that, if not held constant or controlled through randomization, will distort the results.

One-Factor-at-a-Time Experimentation
A highly inefficient method where one factor is changed while all the others are kept constant. The method ignores the possibility of interactions.

One-way Design
An experimental design that studies the effect of one factor that has two or more levels.

Orthogonal design
A design where the correlation between factors is zero.

P-Value
The p-value of a hypothesis test is the probability of observing a value of the test statistic that is at least as inconsistent with the null hypothesis as the value of the test statistic actually observed.

Pareto Chart
A graph that shows the amount of influence each factor has on the response in order of decreasing influence.

Partial confounding
An experimental design where a comparison (main effect or interaction) is confounded with block effects in one but not all the replications of the experiment.

Path of Steepest Ascent
The direction at right angles to the contours of a response surface plot that is used to locate the region in which the response increases; this area includes a local (or global) maximum or minimum on the response surface.

Path of Steepest Descent
The direction at right angles to the contours of a response surface plot that is used to locate the region in which the response decreases.

Plackett-Burman Design
An orthogonal, balanced design that has a multiple of four runs. Used for estimating main effects only; that is, no interactions.

Pure Error
The sums of squares from replicated environmental runs. Pure error provides an opportunity to test for lack-of-fit in the fitted model.

Randomized Block Design
A design where treatments are randomly assigned to compare positions in each of several blocks of replications. In a complete block experiment, each block contains all the treatments.

Randomization
A system of using random numbers to evenly spread the effects of factors not included in an experiment.

Replication
The number of times a specific combination of factor levels is run during an experiment.

Response Surface Methodology
A process of locating an optimal value in a higher-order model. The methodology utilizes regression, contour plots and/or method of steepest ascent/descent.

Response Surface Plot
A plot of the predicted response for one or more factors that uses a model derived from experimental observations.

Response Variable
A variable that is influenced by factors.

Rotatability
A design characteristic that implies that the estimated parameter variability is not a function of direction from the center of a design matrix.

Rotatable Design
A design used in the mapping response surfaces in which fitted models estimate the response with equal precision at all points in the experimental region that are equidistant from the center of the design.

Runs
An experimental condition determined by a row of settings in the design matrix. Also known as a trial. See also *Experimental condition*.

Screening
Testing a set of factors for main effects in hopes of reducing their numbers.

Second-Order Model
An equation that includes squared terms or two-factor interaction terms. A complete second-order model includes all the possible second-order terms (squared terms and two-factor interactions).

Signal-to-Noise Ratio
Signal-to-Noise Ratio is a straight forward way to incorporate the mean and variance into a single performance measure.

Significance Level
The probability of making a wrong conclusion about the importance of a factor, based on statistics. The value is used to test the significance of the factor. Also known as the *p-value*.

Simple Linear Regression
A model where one independent variable is used to predict one dependent variable.

Single Factor Categorical Designs
Designs used to evaluate the experiment in which data is collected at different levels of a single non-quantitative factor. This encompasses designs including one or more blocking factors, such as Latin Squares.

Square Plot
A response plot of the estimated responses for two factors in an experimental design.

Standard Order
The order of the experimental runs of the design matrix for a two-level factorial design where the first column consists of successive low and high settings, the second column consists of successive pairs of low and high settings, the third column consists of four low settings followed by four high settings, and so on.

Taguchi Designs/Robust Designs
See Inner/Outer Arrays

Three-Level Factorial Design.
See *3-Level design*

Two-Level Factorial Design
See *2-Level design*.

Two-Way Design
A design in which the effects of two factors on a response variable are analyzed.

Unbalanced Design
A design that has an unequal number of observations for different factors, or whose cells contain an unequal number of subjects.

Unblocked Design
A design that does not include a blocking factor.

Unsaturated Model
A model that does not contain all the possible factors or independent variables.

User-Defined Design
An experimental design in which the user specifies and enters the factor levels and runs.

Variance Components Designs
Designs based on the assumption that the various error components are independent, so the overall process variability is the sum of the variability due to the different components.

Yates' Order
The standard order of a two-level factorial design.

Time Series

3RSR
A nonlinear smoothing technique that includes a median for a value and three points around that value, resmoothing (R), a splitting operation to eliminate flat segments in the data (S), and a second resmoothing (R).

3RSS
A nonlinear smoothing technique that includes a median for a value and three points around that value, resmoothing (R), and two splitting operations to eliminate flat segments in the data (SS).

3RSSH
A nonlinear smoothing technique that includes a median for a value and three points around that value, resmoothing (R), two splitting operations to eliminate flat segments in the data (SS), and a "Hanning" weighted average with weights .25, .5, and .25 (H).

5RSS
A nonlinear smoothing technique that includes a median for a value and five points around that value, resmoothing (R), and two splitting operations to eliminate flat segments in the data (SS).

5RSSH
A nonlinear smoothing technique that includes a median for a value and five points around that value, resmoothing (R), two splitting operations to eliminate flat segments in the data (SS), and a "Hanning" weighted average with weights .25, .5, and .25 (H).

ARIMA
An abbreviation for Autoregressive Integrated Moving Average. This class of forecasting models, popularized by Box and Jenkins, is comprised of an AR parameter that represents the order of the autoregressive portion of the model, and an MA parameter that represents the order of the moving average in the model.

Autocorrelated Residuals
An indication that the forecasting method has not removed all of the pattern from the data when residual or error terms remain after a forecasting method has been applied. See also *Autocorrelation*.

Autocorrelation
The association of the same time series at different time periods. The pattern of autocorrelation coefficients is used to identify seasonality in a time series and to determine an appropriate model for the data.

Autoregressive (AR)
A form of regression; instead of the dependent variable being related to independent variables, it is related to past values of itself at varying time lags.

Autoregressive/Moving Average (ARMA) Scheme
A type of time-series forecasting model that can be autoregressive (AR) in form, moving average (MA) in form, or a combination of the two (ARMA). In an ARMA model, the series to be forecast is expressed as a function of both previous values of the series and previous error values from forecasting.

Backforecasting
A technique used to estimate initial values of error terms used in applying forecasting techniques. Backforecasting applies the forecasting method to the data starting at the end of the time series and working backward. This provides a set of starting values for the errors that can then be applied to the forecasting technique in the standard forward sequence.

Box-Cox Transformation
A method of transforming time-series data to stabilize the variance of the series and to make the distribution closer to normal. There are two parameters to the Box-Cox transformation: the *power* (or strength) of the transformation, and the *addend*, which is the value that is added to the data before the power is raised.

Box-Jenkins Methodology
The application of autoregressive/moving average schemes to time-series forecasting problems and popularized by George E. Box and G. M. Jenkins. Box and Jenkins suggested the general methodology of applying ARIMA models to time-series analysis, forecasting, and control; in time-series forecasting, the methodology is known as the Box-Jenkins method.

Box-Pierce Test
A test for randomness that helps to determine if a set of autocorrelations differs from the null set; also known as the Portmanteau Test.

Brown's Linear Exponential Smoothing
A one-parameter, double exponential smoothing technique that places greater weight on the most recent observations.

Classical Decomposition
An approach to forecasting that breaks down the underlying pattern of a time series into cyclical, seasonal, trend, and random subpatterns. The subpatterns can then be analyzed individually and recombined to obtain forecasts of the original series.

Crosscorrelation
A measure of association between two time-series variables. Crosscorrelations range in value from -1 to 1 and have similar interpretations as regular correlation coefficients. This determines whether two time series are correlated.

Cumulative Forecasting
Forecasting the cumulative level of a variable over several periods.

Decomposition
See *classical decomposition*.

Differencing
When a nonstationary time series can be made stationary by taking first differences of the series. If first differences do not make the series stationary, then first differences of first differences can be created. This is known as second-order differencing.

Exponential Smoothing
See *Brown's linear exponential smoothing, Holt's linear exponential smoothing, Quadratic exponential smoothing*, and *Simple exponential smoothing*.

Exponentially Weighted Moving Average (EWMA)
A one-parameter smoothing technique that places greater weight on the most recent observations.

Forecasting
The prediction of values of a variable based on known past values of that variable or other related variables. Forecasts can also be based on expert judgments, which in turn are based on historical data and experience.

Henderson's Weighted Moving Average
A high-order smoothing technique used with time series that exhibit substantial randomness.

Holt's Linear Exponential Smoothing
A two-parameter, double exponential smoothing technique that places greater weight on the most recent observations. This method is useful in that in allows any trend in the data to be smoothed with a different smoothing parameter than that used on the original data.

Integrated Time-Series Models
An element of time-series models where the model includes one or more of the differences of the time series.

Lag
A period of time.

Mean Absolute Error (MAE)
A measure of forecast accuracy calculated by summing the absolute values of the individual forecast errors of the time series and dividing by the number of observations.

Mean Absolute Percentage Error (MAPE)
A measure of forecast accuracy calculated by taking the absolute value of the individual forecast errors of the time series, dividing by the actual value of the time series, multiplying by 100, and calculating the average of the resulting values. The MAPE is not related to the scale of the observations, so it is useful in comparing the errors between distinct variables that are not measured in the same units.

Mean Error (ME)
A measure of forecast accuracy calculated by averaging the individual forecast errors across all of the time-series observations.

Mean Percentage Error (MPE)
A measure of forecast accuracy calculated by taking the individual forecast errors of the time series forecasts, dividing by the actual value of the time series, multiplying by 100, and calculating the average of the resulting values. The MPE is typically small because positive and negative errors tend to offset each other.

Mean Squared Error (MSE)
A measure of forecast accuracy calculated by squaring the individual forecast error for each time-series observation and then finding the average of those errors. The MSE gives larger weight to large errors than to other measures of accuracy because those values are squared before they are averaged.

Nonstationary
Not having a constant mean or variance.

Partial Autocorrelation
A measure of correlation used to identify the extent of relationship between current values of a variable with earlier values of that same variable while holding the effects of all the other time lags constant.

Periodogram
A graph that breaks a time series into a set of sine waves of various frequencies. A periodogram is useful in identifying randomness, seasonality, or autocorrelation. Also known as the line spectrum.

Smoothing
A methodology combining two or more observations, taken from periods during which the same causal factors were in effect, that provides a smoothed value or estimate. The term *smoothed* is used because these types of combinations tend to reduce randomness by allowing positive and negative random effects to partially offset each other.

Quadratic Exponential Smoothing
A one-parameter, triple exponential smoothing technique that places greater weight on the most recent observations. The forecasting equation that results is quadratic and may be appropriate to use when the underlying time-series data do not follow a nonlinear pattern.

Random Walk
A forecasting model that randomly forecasts the next observation based on the current observation and the mean and standard deviation of the difference of the values.

Residual
An error that is calculated by subtracting the forecasted value from the actual value.

Rough
The values outside a smoothed curve.

Sampling Interval
The frequency with which time-series data are collected. For example, temperature data might be collected once each hour, which makes *hourly* the sampling interval. Possible values for the sampling interval include daily, monthly, quarterly, hourly, each minute, each second, and so on.

Seasonality
A pattern that repeats itself over fixed intervals of time. For example, temperature data are high in summer and low in winter, following a 12-month seasonal pattern.

Simple Exponential Smoothing
A one-parameter smoothing technique that places greater weight on the most recent observations.

Simple Moving Average
A smoothing technique that calculates smoothed values by taking the average of the previous *N* values of the series. The value of *N* is known as the *order* (or length) of the moving average.

Spencer's 15-term/21-term Moving Averages
A high-order smoothing technique used to deal with time series that exhibit substantial randomness. In general, the greater the randomness, the larger the number of terms in the moving average method.

Stationary
A time-series is stationary when it is based on a constant mean and variance, and when its statistical properties are independent of the time period during which it is observed.

Taper
A bias-reducing process that damps or tapers a time-series variable at its two ends. STATGRAPHICS *Plus* for Windows uses a cosine-bell tapering method to calculate the transformed values.

Time Series
The values of a variable, in ordered sequence, that are observed at equally spaced time intervals.

Trading Day
An active day in a month; that is, a business day rather than a holiday or weekend. Trading-day adjustments are made to time-series data to reflect that similar periods may not include the same number of trading days.

Trend
Values of a variable that have movement in one direction. Trend can follow any of various functional forms such as linear, quadratic, exponential, S-curve, and so on.

Validation
A process of testing a model's ability to make forecasts where some of the data are “held out” of the model-fitting process and used to test the model’s forecast accuracy.

Multivariate Methods

Agglomerative Methods
A hierarchical clustering procedure that begins with each variable or observation in separate clusters. In subsequent steps, clusters that are closest together are combined to build a new aggregate cluster(s).

Algorithm
A set of rules or procedures; similar to an equation.

Analysis Sample
The group used to compute the discriminant function when constructing classification matrices by dividing the original sample randomly into two groups, one for developing the discriminant function and the other for validating it.

A Priori Criterion
A technique used to extract factors from a matrix when you know how many factors to extract before performing the analysis.

Average Linkage
An agglomerative method that uses the average distance from individuals in one cluster to individuals in another cluster as the clustering criterion. This approach tends to combine clusters with small variances.

Categorical Variable
A nonmetric, nominal, binary, or qualitative variable. When a number or value is assigned to a categorical variable it serves merely as a label or a means of identification. An example is the number on a football player's uniform.

Centroid
The point whose coordinates are the mean values of the coordinates of the points in the configuration.

Centroid Method
An agglomerative method in which the distance between two clusters is the distance (typically Euclidean) between their centroids (means). Each time objects are grouped, a new centroid is computed; therefore, cluster centroids move as clusters are merged.

Chance Models
The procedure used to determine the percentage of individuals that would be correctly classified by chance.

City-Block Distance
A method for calculating distances based upon the sum of the absolute differences of the coordinates for the objects. This method assumes that the variables are uncorrelated and that unit scales are compatible.

Classical Factor Analysis
A factor model in which the communalities are replaced by the multiple R-squared values.

Classification
The placing of objects into more or less homogeneous groups in a manner so that the relation between groups is revealed.

Classification Matrix
A matrix that contains numbers that reveal the predictive ability of the discriminant function. The numbers on the diagonal of the matrix represent correct classifications; the off-diagonal numbers represent incorrect representations. Also referred to as a *confusion, assignment,* or *prediction* matrix.

Cluster Analysis
A technique for grouping individuals or objects into clusters so objects in the same cluster are more like each other than they are like objects in other clusters. Also known as *Q-analysis*, *typology*, *classification analysis*, and *numerical taxonomy*.

Cluster Centroid
The average value of the objects in a cluster on all of the variables in an analysis.

Cluster Seeds
The initial center or starting point of a cluster. This value is used to begin nonhierarchical clustering procedures. Clusters are built around these pre-selected seeds.

Communality
The amount of variance an original variable shares with all other variables in an analysis.

Complete Linkage
An agglomerative method in which the clustering criterion is based on the maximum distance between objects. All the objects in a cluster are linked to each other at some maximum distance or minimum similarity.

Component Analysis
A factor model in which the factors are based upon the total variance. In component analysis, unities are used in the diagonal of the of the correlation matrix, which implies that all of the variance is common or shared.

Computational Method
Two methods that can be utilized to derive a discriminant function: the simultaneous (direct) method and the stepwise method. See also *simultaneous method* and *stepwise method*.

Correlation Matrix
A table that shows the intercorrelations among all the variables.

Cutting Score
The score against which each individual's discriminant score is judged to determine into which group the individual should be classified. When the analysis involves two groups, the hit ratio is determined by computing a single “cutting” score. Those entities whose Z-scores are below this score are assigned to one group, while those whose scores are above it are classified in the other group.

Dendrogram
A graphical representation of the results of a clustering procedure. The vertical axis is made up of the objects or individuals; the horizontal axis represents the number of clusters formed at each step of the clustering procedure. Also known as a *tree graph*.

Discriminant Analysis
The appropriate statistical technique when the dependent variables are metric. In most cases the dependent variable consists of two groups or classifications; in others, more than two groups are involved, such as a three-group classification that involves low, medium, and high classifications.

Discriminant Function
A linear equation of the form

$$Z = W_1X_1 + W_2X_2 + ... + W_n X_n$$

where

$Z =$ discriminant score
$W_i =$ discriminant weight
$X_i =$ independent variable

Discriminant Loadings
Used to measure the simple linear correlation between the independent variables and the discriminant function. Also referred to as *structure correlations*.

Discriminant Score
Referred to as a *Z-score*. See Discriminant Loadings.

Discriminant Weight
Also known as a *discriminant coefficient*, its size is determined by the variance structure of the original variables. Independent variables with large discriminatory power usually have large weights; those with little discriminatory power usually have small weights. Collinearity among the independent variables cause an exception to this rule.

Divisive
A clustering procedure that begins with all the objects in a single cluster. It is opposite of agglomerative procedures. The procedure begins with a single large cluster that is divided into separate clusters based on the the most dissimilar objects.

Eigenvalue
The column sum of squares for a factor that represents the amount of variance.

Entropy Group
The few observations that are independent of either cluster.

Factor
A linear combination of original variables. Factors also represent the underlying dimensions that summarize or account for the original set of observed variables.

Factor Analysis
A generic name given to a class of multivariate statistical methods whose primary purpose is to reduce and summarize data. A technique particularly suitable for analyzing complex, multidimensional problems.

Factor Loadings
The correlation between original variables and factors, and the key to understanding the nature of a specific factor.

Factor Matrix
A table that displays the factor loadings of all variables on each factor.

Factor Rotation
The process of manipulating or adjustment the factor axes to achieve a simpler and more meaningful factor solution.

Factor Score
A measure of the composite of all the original variables that make up a new factor.

Hierarchical Clustering
A method used to cluster by joining the most similar observations, then successively connecting the next most similar observations to these.

Hit Ratio
The percentage of statistical units (individuals, respondents, objects, and so forth), that are correctly classified by the discriminant function.

Hold-out Sample
The group of subjects held out of the total sample when the function is computed. Also known as the *validation sample*.

Icicle Plot
A graphical representation that resembles a row of hanging icicles.

Interpretation Stage
The point in a process where the characteristics of each cluster are understood and at which a name or label is developed that defines the nature of each characteristic. See also *Partitioning Stage* and *Profiling Stage*.

K-Means Procedure
Probably the most widely used clustering procedure; uses a limited number of arbitrary cluster centers that offset computational difficulty.

Linear Combination
Representations of the weighted sum of two or more variables. Also known as *linear composites, linear compounds,* and *discriminant variables*.

Metric Variable
A variable with a constant unit of measurement. For example, if a variable is scaled from 1 to 9, the difference between 1 and 2 is the same as that between 8 and 9.

Multiple Discriminant Analysis
A methodology to help explain research problems that involve a single categorical dependent variable and several metric independent variables.

Mutual Similarity Procedures
A method used to cluster by grouping together observations that have a common similarity to other observations.

Nonhierarchical Procedures
Cluster seeds are used instead of a tree-like construction to group objects that are within a pre-specified distance of the seeds.

Normalized Distance Function
A process that converts each raw data score to a standardized variate with zero mean and unit standard deviation. The purpose of the process is to remove the bias introduced by differences in scales of several variables.

Oblique Factor Solutions
A factor solution that is computed in a way that the extracted factors are correlated.

Orthogonal
The mathematical independence of factor axes to each other; that is, at right angles or 90 degrees.

Orthogonal Factor Solutions
A factor solution in which the factors are extracted so that the factor axes are maintained at 90 degrees thereby causing each factor to be independent of or orthogonal from all other factors.

Partitioning Stage
The point in a process that determines if and how clusters could be developed. See also *Interpretation Stage* and *Profile Stage*.

Percentage of Variance Criterion
A technique used in hard sciences and in social sciences where the criterion is the cumulative percentages of the variance extracted by successive factors. In the hard sciences, the factoring procedure is usually stopped when the extracted factors account for at least 95 percent of the variance. In the social sciences, the factoring procedure is usually stopped when the extracted factors account for at least 60 percent of the total variance.

Principal Components
A method for reducing the dimensionality of a set of variables by constructing uncorrelated linear combinations of them. The combinations are computed in a way that the first component accounts for the major part of the variance; that is, it is the major axis of the points in the p-dimensional space.

Profiling Stage
The point in a process that describes the characteristics of each cluster to explain how they may differ on relevant dimensions. See also *Partitioning Stage* and *Interpretation Stage*.

Rotation
A methodology used to preserve the original structure and reliability of discriminant models while at the same time making them substantively easier to interpret.

Scree Test Criterion
A technique that uses the scree tail test to extract factors. The scree tail test is an approach that identifies the optimum number of factors that can be extracted before the amount of unique variance dominates the common variance structure.

Simultaneous Method
A methodology used to compute the discriminant function so all the independent variables are considered concurrently. Therefore, the discriminant function(s) is computed based upon the entire set of independent variables, regardless of the discriminating power of each independent variable. This method is appropriate to use when all of the independent variables are included in the analysis and it is not important to see intermediate results based only on the most discriminating variables.

Single Linkage Method
A hierarchical clustering procedure based on minimum distance. The procedure locates the two objects with the shortest distance and places them in the first cluster. The process continues until all the objects are in one cluster.

Stepwise Method
An alternative to the simultaneous method. Involves entering the independent variables into the discriminant function one at a time, based on the basis of their discriminating power. The single best variable is chosen first; the initial variable is then paired with each of the other independent variables, one at a time, and a second variable is chosen, and so on.

Tolerance
The proportion of the variation in the independent variables that is not explained by the variables already in the model.

Trace
The sum of the square of the numbers on the diagonal of the correlation matrix used in the factor analysis. The trace represents the total amount of variance on which the factor solution is based. In component analysis, the trace is equal to the number of variables based on the assumption that the variance in each variable is equal to 1. In common factor analysis, the trade is equal to the sum of the communalities on the diagonal of the reduced correlation matrix.

Vector
A straight line drawn from the center of a graph to the coordinates of a specific variable vector. The length of each vector indicates the relative importance of each variable in discriminating among groups.

Ward's Method
A hierarchical cluster procedure that calculates the distance between two clusters as the sum of squares between the two clusters summed over all the variables.

Advanced Regression

Adjusted R-Squared Statistic
A statistic that is suitable for comparing models that have different numbers of independent variables; indicates the percentage of variability for which the model accounts.

Alternative Hypothesis
A hypothesis, H_A, against which an assumption or null hypothesis is tested.

Analysis Summary
A tabular option in STATGRAPHICS *Plus* that displays the initial results of a specific analysis.

ANOVA
Analysis of variance.

Autocorrelation Function Plot
A plot of the autocorrelation estimates for the residuals.

Backward Selection
An option that allows the program to select variables by performing a backward stepwise regression, which begins with all the variables in the model and removes them one at a time.

Biased Estimates
An over- or underestimation of the expected value of a sample statistic.

Bonferroni
A way of estimating the probability of making at least one Type I Error when conducting a series of *t* tests on the means of three or more groups.

Calibration Line
The equation of the line that best fits through the calibration points.

Calibration Problem
The process of using a regression equation to predict in reverse; for example, predict *X* given *Y*.

Categorical Variables
Variables that contain the level codes. When an independent variable is entered as a categorical variable, the program sets up indicator variables for each level except the first one.

Comparison of Regression Lines
Using hypothesis tests to compare two or more regression lines for significantly different slopes and/or intercepts.

Conditional Sums of Squares
An option that allows you to test the statistical significance of the terms in a linear model.

Confidence Intervals
Intervals that show the precision of the estimated coefficients given the amount of available data and noise in the model.

Confidence Limits for Forecast Means
The end points for the confidence interval for the means of each estimated value.

Confidence Limits for Individual Forecasts
The end points for the prediction interval for the individual estimated values.

Confidence Level
A value, usually 90 to 99 percent, the program uses to calculate the confidence intervals for the estimated parameters.

Constants
(a) A measure or value that is the same for all units of analysis. (b) A quantity that does not change value in a particular context. (c) In a regression equation, another name for the intercept.

Contour Plot
A two-dimensional plot that traces the contours of the estimated dependent variable as a function of the other variables.

Cook's Distance
A statistic that measures the distance between the estimated coefficients with and without each observation.

Collinear
Having a common line.

Correlation Matrix
A table of correlation coefficients that shows all the pairs of correlations for a set of variables.

Cross Operator
An operator that instructs the program to multiply two factors together or to form a crossed effect.

Cube Plot
A plot of the estimated effects for the high and low settings of three factors.

Dependent Variable
The presumed response in a study; known as dependent because it “depends” on other variables; a variable whose values are characterized by the independent variable(s), whether or not a causal relationship.

DFITS
A statistic that measures the amount of change for each estimated coefficient if the observation is removed from the data.

Double Reciprocal Model
A calibration model that fits the reciprocal of the dependent variable and the reciprocal of the independent variable; defined by the equation $Y = 1/(a + b/X)$.

Dummy Variable
A dichotomous variable, usually coded 1 to indicate the presence of an attribute and 0 to indicate its absence.

Duncan
A test used after an ANOVA to determine which sample means differ significantly from one other, which allows you to protect the results from Type I errors for a small number of comparisons.

Dunnett
An option in the General Linear Models Analysis that allows you to use a control treatment as a benchmark for making comparisons.

Durbin-Watson Statistic
A test for autocorrelation or serial correlation in the residuals of a least squares regression analysis. As the autocorrelation increases, the Durbin-Watson goes down. The larger the autocorrelation, the less reliable the results of the regression analysis.

Error Term
The part of an equation that indicates what is unexplained by the independent variables.

Estimator
A sample statistic that is used to determine a probable value for a population parameter.

Exclude Command
A command in STATGRAPHICS *Plus* that allows you to select the terms you want to exclude from a model.

Exponential Model
A model that is equivalent to taking the natural log of Y defined by the equation $Y = exp(a + bX)$.

Factors
In analysis of variance, an independent variable; that is, a variable presumed to affect the value of another variable.

Forecast
A prediction generated by calculating the value of a function given a new set of values.

Forward Selection
In STATGRAPHICS *Plus*, a method for selecting variables for a multiple regression, which begins with no variables in the model and adds them one at a time.

Gauss-Newton
A method of estimation for nonlinear regression that uses a Taylor series expansion to approximate a model with linear terms and then uses ordinary least squares to estimate the parameters.

General Linear Model (GLM)
A common set of statistical assumptions upon which are based a full range of methods used to study one or more continuous dependent variables and one or more independent variables, whether they are continuous or categorical. The basic concept of GLM is that the relationship between the dependent variable and the independent variables is expressed as an equation that contains a term for the weighted sum of the values of the independent variables, plus a term for everything that is unknown (an error term). The least-squares criterion determines the weight for each independent variable.

Goodness-of-Fit
How well observed data match expected values under a specific assumption.

Hold... Command
A command that allows you to set the level of the quantitative factors when you are plotting a fitted model.

Independent Variable
A variable you can use to explain or to predict the values of another variable.

Interaction
A multiplicative effect.

Intercept
The point at which a regression line crosses the vertical (Y) axis; that is, when the value on the horizontal (X) axis is zero.

Lag
In STATGRAPHICS *Plus*, a user-entered integer value used to estimate the autocorrelation values.

Least Squares Regression
A method that fits a regression by finding the coefficients that minimize the sum of the squared effects.

Leverage
A statistic that measures the amount each estimated coefficient would change if each observation was removed from the data.

Linear Regression Model
A model that fits a dependent variable *Y* to a linear function of an independent variable *X*.

Log Probit Model
A model that fits a response of observed proportions or probabilities at each level of an independent variable; defined by the equation *Y* = *normal* (a + *b*(*LogX*)). The model is appropriate when the response variable tends to follow a cumulative normal distribution.

Logarithmic-X Model
A calibration model defined by the equation *Y* = *a* + *b*(*LogX*).

Logistic Model
A model that fits a response of observed proportions or probabilities at each level of an independent variable; defined by the equation *Y* = *exp*(*a* + *bX*)/(*1* + *exp*(*a* + *bX*)). The response function is a nonlinear S-shaped curve with asymptotes at 0 and 1.

Logistic Regression Analysis
An analysis that allows analysts to estimate multiple regression models when the response being modeled is dichotomous and can be scored 0,1; that is, the outcome must be one of two choices.

Lower Bound Interval
An option in STATGRAPHICS *Plus* that calculates the one-sided, lower confidence limit for each coefficient.

LSD (Least Significant Differences)
A multiple range test in STATGRAPHICS *Plus* that, when the F-ratio is significant, allows you to make planned comparisons.

MAE (Mean Absolute Error)
The average of the absolute values of the residuals; appropriate for linear and symmetric data. If the result is a small value, you can predict performance more precisely; if the result is a large value, you may want to use a different model.

Mallows' C_p
A measure of the bias in a model based on a comparison of total Mean Squared Error to the true error variance. Unbiased models have an expected C_p value of approximately *p*, where *p* is the number of coefficients in the fitted model. C_p is based on the assumption that the model that contains all the candidate variables is unbiased; therefore, the full model will always have $C_p = p$. Look for models that have C_p values close to *p*.

MANOVA
Multivariate analysis of variance.

MAPE (Mean Absolute Percentage Error)
The mean or average of the sum of all the percentage errors for a given set of data without regard to sign (that is, their absolute values are summed and the average is computed).

Marquardt Algorithm
An algorithm that uses the best features of the Gauss-Newton method and the method of steepest descent, which results in a middle ground between the two methods.

Maximum Likelihood
A statistical method that estimates the population parameters that will most likely result in observed sample data.

ME (Mean Error)
The average of the residuals. The closer the ME is to 0, the less biased, or more accurate, the prediction.

Means Plot
A plot of the mean for a dependent variable for each level of an independent categorical variable, if there is one. The plot displays the confidence intervals for each of the means separately.

MPE (Mean Percentage Error)
The average of the absolute values of the residuals divided by the corresponding estimates.

MSE (Mean Square Error)
A measure of accuracy computed by squaring the individual error for each item in the set of data, then finding the average or mean value for the sum of those squares.

Multicollinearity
Exists in multiple regression analysis when two or more independent variables are highly correlated, which makes it difficult if not impossible to determine their separate effects on the dependent variable.

Multiple Regression Analysis
An analysis that involves fitting a random variable to a set of explanatory variables to provide regression coefficients for a linear model.

Multiplicative Model
This calibration model is equivalent to taking the natural logs of *Y* and *X*; defined by the equation *Y* = *aX^b*.

Multivariate t
A multiple range test that tests only a few comparisons (those of most importance).

Nest Operator
An operator that adds the left (first) and right (second) parentheses to an expression, which instructs the program to create a nested effect.

Newman-Keuls
An option in STATGRAPHICS *Plus*, that when the F-ratio is significant, allows you to test multiple hypotheses. The Type I error rate is smaller when you use this method than it is when you use the Duncan method.

Nonlinear Regression
Regression analysis that finds a least squares solution for a nonlinear model, which cannot be done using matrix algebra as it is in linear regression.

Normal Probability Plot
A plot that displays the residuals to determine if the errors follow a normal distribution. The plot consists of an arithmetic (interval) horizontal axis scaled for the data and a vertical axis scaled so the cumulative distribution function of a normal distribution plots as a straight line.

Observed versus Predicted Plot
A plot of the observed values of *Y* versus the values predicted by the predicted model.

***p*-value**
A component of the ANOVA table that serves as a measure of significance. When the *p*-value is less than 0.01, there is a statistically significant relationship between two variables at the 99 percent confidence level.

Pairwise Means Comparison
A comparison of the difference between two treatment group means.

Plot of Fitted Model
A plot of the fitted model where a separate line is shown for each level code.

Prediction Capability Plot
A plot that summarizes the prediction capability of a fitted logistic model: The percentages of correct values versus the cutoff values for a dependent variable that was Total, True, or False.

Prediction Histogram
A plot that demonstrates the ability of a fitted logistic model to distinguish between cases when a dependent variable is True or False.

Quantitative Variables
Numeric variables that contain continuous data.

Quartiles
Divisions of the total cases or observations in a study into four groups of equal size.

Reciprocal-X Model
A calibration model that is defined by the equation $Y = a + b/X$.

Reciprocal-Y
A calibration model that is defined by the equation $Y = 1/(a + bX)$.

Regression Analysis
A mathematical tool that quantifies the relationship between a dependent variable and one or more independent variables.

Regression Coefficient
A parameter or its estimate, for a regression model; often denoted by the Greek letter beta.

Regression Model Selection Analysis
The Regression Model Selection Analysis in STATGRAPHICS *Plus* ranks the best subsets of explanatory variables based on criterion you select, calculates the statistics for all possible linear regression models, and sorts the values so you can choose the "best" model.

Residuals
The difference between the observed and the fitted values.

Residuals versus Level Codes Plot
A plot that displays the residuals or the studentized residuals versus the level codes.

Residuals versus Predicted Plot
A plot that displays the residuals or the studentized residuals versus the predicted values for the observed variable (Y).

Residuals versus Row Number Plot
A plot that displays the residuals or the studentized residuals versus the row number.

Residuals versus *X* (where *X* indicates the name of each independent *X* variable) Plot
A plot that displays the residuals or studentized residuals versus the independent variable (*X*), where *X* equals the name of each independent variable you selected.

Ridge Parameter
A parameter that controls the extent of the bias that is introduced in ridge regression. A good value for the ridge parameter is the smallest value that occurs before the estimates slowly change.

Ridge Regression
One of several methods proposed to remedy multicollinearity problems by modifying the method of least squares to allow biased estimators of the regression coefficients.

Ridge Trace Plot
A plot of the values for the standardized or unstandardized coefficients versus the values for the ridge parameter. As the ridge parameter increases from 0.0, the coefficients at first change dramatically, then become relatively stable.

R-Squared Statistic
A statistic that indicates the percentage of variability in the dependent variable for which the model accounts.

S-Curve Model
A model that takes the natural log of *Y* and the reciprocal of *X*; defined by the equation $Y = exp(a + b/X)$.

Scatterplot
A plot of one variable versus another.

Scheffe
A test of statistical significance used for post-hoc multiple comparisons after a regression analysis or an analysis of variance.

Second Factor Plot
A plot of the dependent variable on the *X*-axis using the second factor in the interaction. A line is drawn for each level of the first factor.

Slope
The rate at which a line or curve rises or falls when covering a given horizontal distance. The most common reference is to the steepness or angle of a regression line.

Square Plot
A plot of the predicted values of the dependent variable versus the high and low values of a factor.

Square Root-X Model
A calibration model defined by the equation $Y = a + b\ sqrt(X)$. The reciprocal-*X*, logarithmic-*X*, and square root-*X* transformations linearize a nonlinear relationship between the dependent and independent variables.

Square Root-Y Model
A calibration model defined by the equation $Y = (a + bX)$^2. The exponential, reciprocal-*Y*, and square root-*Y* transformations can help remedy error terms that are not normal and that do not have constant variance.

Standard Error of the Estimate
This statistic explains the value for the standard deviation of the residuals. You can use this value to construct prediction limits for new observations.

Standard Errors for Forecasts
This statistic includes the standard deviation for the distribution of the estimated values in a report.

Standardized Regression Coefficient
A statistic that provides a way to compare the relative importance of different variables in a multiple regression analysis.

Steepest Descent
A method of nonlinear regression analysis that searches for the minimum least squares criterion measure by iteratively determining the direction in which the regression coefficients should be changed.

Stopping Criterion
A number used to stop the estimation process; that is, the estimation stops when this value is less than the estimated values (as a proportion) for all the unknown parameters.

Studentized Residuals
The residuals scaled by an estimate of the variance from the data with the observation removed.

Surface Plot
A three-dimensional plot of the relationship between the estimated dependent variable and two selected variables.

Target Variables
The names of the variables in which you can save the values generated during an analysis. The variables are shown on the Save Results Dialog Box where you can enter new names or accept the defaults.

Tukey HSD (Honestly Significant Differences)
A test used to determine which means are significantly different after an analysis of variance of the differences in group means is performed.

Two-Sided Interval
An option in STATGRAPHICS *Plus* that calculates the upper and lower confidence limits for each coefficient.

Type I Sums of Squares
An option in STATGRAPHICS *Plus* that computes the sums of squares for each factor in the order you enter the factors into the Effects list box on the GLM Model Specification dialog box.

Type III Sums of Squares
An option in STATGRAPHICS *Plus* that computes additional sums of squares for each factor as though that factor was the last one added to the model.

Unbalanced Design
Factorial designs that contain unequal numbers of observations for different factor levels or combinations, or cells that contain unequal numbers of subjects.

Upper Bound Interval
In STATGRAPHICS *Plus*, an option that calculates the one-sided, upper confidence limit for each coefficient.

User-Specified Comparison
In STATGRAPHICS *Plus*, an option that allows you to select contrasts for multiple range tests in the General Linear Models Analysis.

Variance Inflation Factors (VIFs)
Factors that measure the extent to which explanatory variables are correlated among themselves. VIF values above 10 usually indicate serious multicollinearity, which greatly increases the estimation error of the model coefficients when compared with an orthogonal sample.

Weights
The values applied to residuals when the program estimates the model coefficients in a weighted least squares regression.