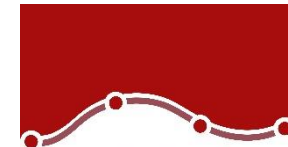




SPSS



**Statistics for
Data Analysis**

WHO WE ARE

SPS is an Italian center of statistical data analysis with more than 20 years of experience.

SPS was born in 1994 as SPSS Italia and it was the only reseller in Italy for SPSS software suite, authorised by SPSS inc.

Today SPS is an IBM Gold Business Partner, Software Support Provider and Expert Level in Data Science & Business Analytics.

CONTACTS

Registered office:
Via Antonio Zanolini, 36 A/B
40126 Bologna (BO)

Operational headquarters:
Via Isonzo, 55/2
40033 Casalecchio di Reno (BO)
P.I. 04222630370
Tel. 051-252573
www.spss.it

DATASHEET

Statistics Complex Samples



Statistics for Data Analysis

Organizations can solve a wide array of business and research problems with the solution Statistics for Data Analysis.

Compared to other statistical software, the solution is easier to use, has a lower total cost of ownership and more comprehensively addresses the entire analytical process, from planning to data collection to analysis, reporting and deployment.

Organizations of all types rely on Statistics for Data Analysis to help increase revenue, outmaneuver competitors, conduct research and make better decisions. With decades of built-in expertise and innovation, it's a leading choice for reliable statistical analysis.

Statistics Base is part of the solution Statistics for Data Analysis, which consists of:

- Software license
- Add-On
- SPS Service Program

This comprehensive, easy-to-use solution includes many different procedures and tests to help users solve complex business and research challenges.

Highlights Statistics for Data Analysis

- Get support through every step of the analytical process.
- Carry out essential analyses from an intuitive graphical interface.
- Select from more than a dozen integrated products to make specialized analyses faster and easier.



Statistics for Data Analysis

The solution analytical capabilities to meet the analysis requirements of any type of organization, from basic tools for solving common problems to advanced analytical techniques that enable all type of organization to address complex challenges.

Statistics for Data Analysis can help you:

- Analyze your data with new and advanced statistics, including a variety of new features within UNIANOVA methods
- Integrate better with third-party applications, including stronger integration with Microsoft Office
- Save time and effort with productivity enhancements:
 - More attractive and modern-looking charts in Chartbuilder
 - New groundbreaking features in Statistics Amos 25
 - Data and syntax editor enhancements
 - Accessibility improvements for the visually impaired
 - Updated merge user interface
 - Simplified toolbars

Statistics for Data Analysis can access quickly, manage and analyze any kind of dataset, including survey data, corporate databases or data downloaded from the web.

In addition, the software can process Unicode data. This eliminates variability in data due to language-specific encoding and enables your organization to view, analyze and share data written in multiple languages.

Business Benefit Statistics for Data Analysis

- Support business decisions with data-based analytics for improved outcomes.
- Be more confident in your results by incorporating data from many different sources, including geospatial information, in your analysis and using proven, tested techniques to perform your analysis.
- Save time and effort with capabilities that enable experienced analysts to develop procedures or dialogs that others can use to speed through repetitive tasks.
- Give results greater impact by using visualization capabilities that clearly show others the significance of your findings.



Statistics Complex Samples

Datasheet

Correctly compute complex samples statistics

When you conduct sample surveys, use a statistics package dedicated to producing correct estimates for complex sample data. Statistics Complex Samples provides specialized statistics that enable you to correctly and easily compute statistics and their standard errors from complex sample designs. You can apply it to:

- Highlights:
- Increase the precision of your sample or ensure a representative sample with stratified sampling.
 - Select groups of sampling units with clustered sampling.
 - Select an initial sample, then create a second-stage sample with multistage sampling.

- Survey research - Obtain descriptive and inferential statistics for survey data.
- Market research - Analyze customer satisfaction data.
- Health research - Analyze large public-use datasets on public health topics such as health and nutrition or alcohol use and traffic fatalities.
- Social science - Conduct secondary research on public survey datasets.
- Public opinion research - Characterize attitudes on policy issues.



Statistics Complex Samples

Datasheet

SPSS Complex Samples provides you with everything you need for working with complex samples. It includes:

- An intuitive Sampling Wizard that guides you step by step through the process of designing a scheme and drawing a sample
- An easy-to-use Analysis Preparation Wizard to help prepare public-use datasets that have been sampled, such as the National Health Inventory Survey data from the Centers for Disease Control and Prevention (CDC)
- Numerical outcome prediction through the Complex Samples General Linear Model (CSGLM)
- Ordinal outcome prediction through Complex Samples Ordinal Regression (CSORDINAL)
- Categorical outcome prediction through Complex Samples Logistic Regression (CSLOGISTIC)
- Time to an event prediction through Complex Samples Cox Regression (CSCOXREG)

From the planning stage and sampling through the analysis stage, Statistics Complex Samples makes it easy to obtain accurate and reliable results. Since Statistics Complex Samples takes up to three states into account when analyzing data from a multistage design, you'll end up with more accurate analyses.



Statistics Complex Samples

Datasheet

In addition to giving you the ability to assess your design's impact, Statistics Complex Samples also produces a more accurate picture of your data because subpopulation assessments take other subpopulations into account.

You can use the following types of sample design information with Statistics Complex Samples:

- Stratified sampling - Increase the precision of your sample or ensure a representative sample from key groups by choosing to sample within subgroups of the survey population. For example, subgroups might be a specific number of males or females, or contain people in certain job categories or people of a certain age group.
- Clustered sampling - Select clusters, which are groups of sampling units, for your survey. Clusters can include schools, hospitals, or geographic areas with sampling units that might be students, patients, or citizens. Clustering often helps make surveys more cost-effective.
- Multistage sampling - Select an initial or first-stage sample based on groups of elements in the population, then create a second-stage sample by drawing a subsample from each selected unit in the first-stage sample. By repeating this option, you can select a higher-stage sample.



Statistics Complex Samples

Datasheet

More confidently reach results

As a researcher, you want to be confident about your results. Most conventional statistical software assumes your data arise from simple random sampling. Simple random sampling, however, is generally neither feasible nor cost-effective in most large-scale surveys.

Analyzing such sample data with conventional statistics risks incorrect results. For example, estimated standard errors of statistics are often too small, giving you a false sense of precision. Statistics Complex Samples enables you to achieve more statistically valid inferences for populations measured in your complex sample data because it incorporates the sample design into survey analysis.

Work efficiently and easily

Only Statistics Complex Samples makes understanding and working with your complex sample survey results easy. Through the intuitive interface, you can analyze data and interpret results. When you're finished, you can publish datasets and include your sampling or analysis plans. Each plan acts as a template and allows you to save all the decisions made when creating it. This saves time and improves accuracy for yourself and others who may want to use your plans with the data, either to replicate results or pick up where you left off.

Statistics Complex Samples

Datasheet

■ Accurate analysis of survey data

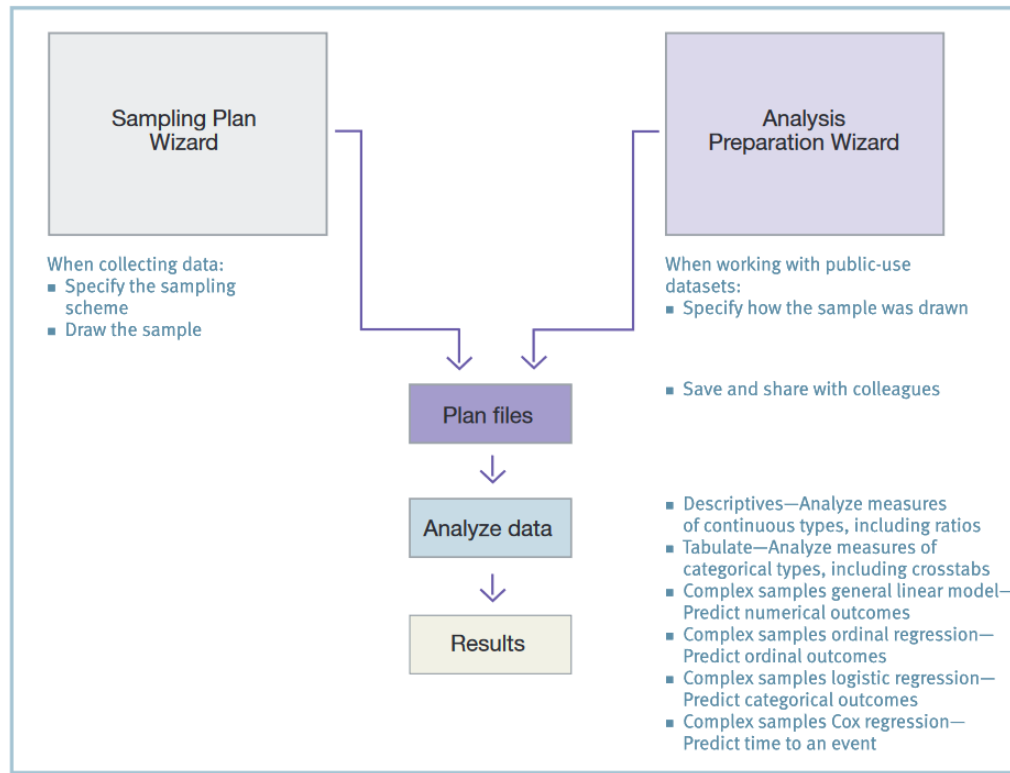


Figure 1: Accurate analysis of survey data is easy in Statistics Complex Samples. Start with one of the wizards (which one to select depends on your data source) and then use the interactive interface to create plans, analyze data and interpret results



Statistics Complex Samples

Datasheet

To begin your work in Statistics Complex Samples, use the wizards, which prompt you for the many factors you must consider. If you are creating your own samples, use the Sampling Wizard to define the sampling scheme. If you're using public-use datasets that have been sampled, such as those provided by the CDC, use the Analysis Preparation Wizard to specify how the samples were defined and how to estimate standard errors. Once you create a sample or specify standard errors, you can create plans, analyze your data and produce results (see the diagram above for workflow)

Statistics Complex Samples makes it easy to learn and work quickly. Use the online help system, explore the interactive case studies, or run the online tutorial to learn more about using your data with the software. Statistics Complex Samples enables you to:

- Reach correct point estimates for statistics such as totals, means, and ratios
- Obtain the standard errors of these statistics
- Produce correct confidence intervals and hypothesis tests
- Predict numerical outcomes
- Predict ordinal outcomes
- Predict categorical outcomes
- Predict time to an event



Statistics Complex Samples Features

Complex Samples Plan (CSPLAN)

This procedure provides a common place to specify the sampling frame to create a complex sample design or analysis specification used by companion procedures in the Statistics Complex Samples module. CSPLAN does not actually extract the sample or analyze data. To sample cases, use a sample design created by CSPLAN as input to the CSSELECT procedure (described on the next page). To analyze sample data, use an analysis design created by CSPLAN as input to the CSDESCRIPTIVES, CSTABULATE, CSGLM, CSLOGISTIC, or CSORDINAL procedures (described on the following pages)

- Create a sample design: Use to extract sampling units from the active file.
- Create an analysis design: Use to analyze a complex sample
- When you create a sample design, the procedure automatically saves an appropriate analysis design to the plan file. A plan file is created for designing a sample, and therefore, can be used for both sample selection and analysis.
- Display a sample design or analysis design.
- Specify the plan in an external file.
- Name planwise variables to be created when you extract a sample or use it as input to the selection or estimation process with the PLANVARS subcommand.
 - Specify final sample weights for each unit to be used by Statistics Complex Samples analysis procedures in the estimation process.
 - Indicate overall sample weights that will be generated when the sample design is executed in the CSSELECT procedure.
 - Select weights to be used when computing final sampling weights in a multistage design.
- Control output from the CSPLAN procedure with the PRINT subcommand.
 - Display a plan specifications summary in which the output reflects your specifications at each stage of the design.
 - Display a table showing MATRIX specifications.
- Signal stages of the design with the DESIGN subcommand. You can also use this subcommand to define stratification variables and cluster variables or create descriptive labels for particular stages.
- Specify the sample extraction method using the METHOD subcommand. Select from a variety of equal- and unequal-probability methods, including simple and systematic random sampling. Methods for sampling with probability proportionate to size (PPS) are also available. Units can be drawn with replacement (WR) or without replacement (WOR) from the population.
 - SIMPLE_WOR: Select units with equal probability. Extract units without replacement.
 - SIMPLE_WR: Select units with equal probability. Extract units with replacement.
 - SIMPLE_SYSTEMATIC: Select units at a fixed interval throughout the sampling frame or stratum. A random starting point is chosen within the first interval.
 - SIMPLE_CHROMY: Select units sequentially with equal probability. Extract units without replacement.
 - PPS_WOR: Select units with probability proportional to size. Extract units without replacement.
 - PPS_WR: Select units with probability proportional to size. Extract units with replacement.
 - PPS_SYSTEMATIC: Select units by systematic random sampling with probability proportional to size. Extract units without replacement.



- PPS_CHROMY: Select units sequentially with probability proportional to size. Extract units without replacement.
- PPS_BREWER: Select two units from each stratum with probability proportional to size. Extract units without replacement.
- PPS_MURTHY: Select two units from each stratum with probability proportional to size. Extract units without replacement.
- PPS_SAMPFORD: Extends Brewer's method to select more than two units from each stratum with probability proportional to size. Extract units without replacement.
- Control for the number or percentage of units to be drawn: Set at each stage of the design. You can also choose output variables, such as stagewise sampling weights, which are created upon the sample design execution.
- Estimation methods: With replacement, equal probability without replacement in the first stage, and unequal probability without replacement
- You can choose whether to include the finite population correction when estimating the variance under simple random sampling (SRS)
- Unequal probability estimation without replacement: Request in the first stage only
- Variable specification: Specify variables for input for the estimation process, including overall sample weights and inclusion probabilities
- Specify the number of sampling units drawn at the current stage using the SIZE subcommand.
- Specify the percentage of units drawn at the current stage. For example, specify the sampling fraction using the RATE subcommand.
- Specify the minimum number of units drawn when you specify RATE. This is useful when the sampling rate for a particular stratum is very small due to rounding

- Specify the measure of size for population units in a PPS design. Specify a variable that contains the sizes or request that sizes be determined when the CSSELECT procedure scans the sample frame.
- Obtain stagewise sample information variables when you execute a sample design using the STAGEVARS subcommand. You can obtain:
 - The proportion of units drawn from the population at a particular stage using stagewise inclusion (selection) probabilities
 - Prior stages using cumulative sampling weight for a given stage
 - Uniquely identified units that have been selected more than once when your sample is done with replacement, with a duplication index for units selected in a given stage
 - Population size for a given stage
 - Number of units drawn at a given stage
 - Stagewise sampling rate- Sampling weight for a given stage
- Choose an estimation method for the current stage with the ESTIMATOR subcommand. You can indicate:
 - Equal selection probabilities without replacement
 - Unequal selection probabilities without replacement
 - Selection with replacement
- Specify the population size for each sample element with the POPSIZE subcommand
- Specify the proportion of units drawn from the population at a given stage with the INCLPROB subcommand

Complex Samples Selection (CSSELECT)

CSSELECT selects complex, probability-based samples from a population. It chooses units according to a sample design created through the CSPLAN procedure.

- Control the scope of execution and specify a seed value with the CRITERIA subcommand
- Control whether or not user-missing values of classification (stratification and clustering) variables are treated as valid values with the CLASSMISSING subcommand



- Use the most updated Mersenne Twister random number generator to select the sample
- Specify general options concerning input and output files with the DATA subcommand
 - Opt to rename existing variables when the CSSELECT procedure writes sample weight variables and stagewise output variables requested in the plan file, such as inclusion probabilities
- Write sampled units to an external file using an option to keep/drop specified variables
- Automatically save first-stage joint inclusion probabilities to an external file when the plan file specifies a PPS_WR sampling method
- Opt to generate text files containing a rule that describes characteristics of selected units
- Control output display through the PRINT subcommand
 - Summarize the distribution of selected cases across strata. Information is reported per design stage.
 - Produce a case-processing summary

Complex Samples Descriptives (CSDESCRIPTIVES)

CSDESCRIPTIVES estimates means, sums, and ratios, and computes their standard errors, design effects, confidence intervals, and hypothesis tests for samples drawn by complex sampling methods. The procedure estimates variances by taking into account the sample design used to select the sample, including equal probability and PPS methods, and WR and WOR sampling procedures. Optionally, CSDESCRIPTIVES performs analyses for subpopulations.

- Specify the name of a plan file, which is written by the CSPLAN procedure, containing analysis design specifications with the PLAN subcommand
- Specify joint inclusion probabilities file names
- Specify the analysis variables used by the MEAN and SUM subcommands using the SUMMARY subcommand
- Request that means and sums be estimated for variables specified on the SUMMARY subcommand through the MEAN and SUM subcommands

- Request t tests of the population mean(s) and sums and give the null hypothesis value(s) through the TTEST keyword. If you define subpopulations using the SUBPOP subcommand, then null hypothesis values are used in the test(s) for each subpopulation, as well as for the entire population.
- Request that ratios be estimated for variables specified on the SUMMARY subcommand through the RATIO subcommand
 - Request t tests of the population ratios and give the null hypothesis value(s) through the TTEST keyword
- Associate syntax with the mean, sum, or ratio estimates, including:
 - The number of valid observations in the dataset for each mean, sum, or ratio estimate
 - The population size for each mean, sum, or ratio estimate
 - The standard error for each mean, sum, or ratio estimate
 - Coefficient of variation
 - Design effects
 - Square root of the design effects
 - Confidence interval
- Specify subpopulations for which analyses are to be performed using the SUBPOP subcommand
 - Display results for all subpopulations in the same or a separate table
- Specify how to handle missing data
 - Base each statistic on all valid data for the analysis variable(s) used in computing the statistic. Compute ratios using all cases with valid data for both of the specified variables. You may base statistics for different variables on different sample sizes.
 - Base only cases with valid data for all analysis variables when computing statistics. Always base statistics for different variables on the same sample size.
 - Exclude user-missing values among the strata, cluster, and subpopulation variables.
 - Include user-missing values among the strata, cluster, and subpopulation variables. Treat user-missing values for these variables as valid data.



Complex Samples Tabulate (CSTABULATE)

CSTABULATE displays one-way frequency tables or two-way crosstabulations and associated standard errors, design effects, confidence intervals, and hypothesis tests for samples drawn by complex sampling methods. The procedure estimates variances by taking into account the sample design used to select the sample, including equal probability and PPS methods, and WR and WOR sampling procedures. Optionally, CSTABULATE creates tables for subpopulations.

- Specify the name of an XML file, written by the CSPLAN procedure, containing analysis design using the PLAN subcommand
- Specify the joint inclusion probabilities file name
- Use the following statistics within the table:
 - Population size: Estimate the population size for each cell and marginal in a table
 - Standard error: Calculate the standard error for each population size estimate
- Row and column percentages: Express the population size estimate for each cell in a row or column as a percentage of the population size estimate for that row or column. This functionality is available for two-way crosstabulations
 - Table percentages: Express the population size estimate in each cell of a table as a percentage of the population size estimate for that table
 - Coefficient of variation
 - Design effects
 - Square root of the design effects
 - Confidence interval: Specify any number between zero and 100 as the confidence interval
 - Unweighted counts: Use unweighted counts as the number of valid observations in the dataset for each population size estimate
 - Cumulative population size estimates: Use cumulative population size estimates for one-way frequency tables only
 - Cumulative percentages: Use cumulative percentages corresponding to the population size estimates for one-way frequency tables only
- Expected population size estimates: Use expected population size estimates if the population size estimates of each cell in the two variables in the crosstabulation are statistically independent. This functionality is available for two-way crosstabulations only
- Residuals: Show the difference between the observed and expected population size estimates in each cell. This functionality is available for two-way crosstabulations only
- Pearson residuals: This functionality is available for two-way crosstabulations only
- Adjusted Pearson residuals: This functionality is available for two-way crosstabulations only
- Use the following statistics and tests for the entire table:
 - Test of homogeneous proportions
 - Test of independence
 - Odds ratio- Relative risk
 - Risk difference
- Specify subpopulations for which analyses are to be performed using the SUBPOP subcommand
 - Display results for all subpopulations in the same or a separate table
- Specify how to handle missing data
 - Base each table on all valid data for the tabulation variable(s) used in creating the table. You may base tables for different variables on different sample sizes
 - Use only cases with valid data for all tabulation variables in creating the tables. Always base tables for different variables on the same sample size
 - Exclude user-missing values among the strata, cluster, and subpopulation variables
 - Include user-missing values among the strata, cluster, and subpopulation variables. Treat user-missing values for these variables as valid data



Complex Samples General Linear Model (CSGLM)

This procedure enables you to build linear regression, analysis of variance (ANOVA), and analysis of covariance (ANCOVA) models for samples drawn using complex sampling methods. The procedure estimates variances by taking into account the sample design used to select the sample, including equal probability and PPS methods, and WR and WOR sampling procedures. Optionally, CSGLM performs analyses for subpopulations.

- Models
 - Main effects
 - All n-way interactions
 - Fully crossed
 - Custom, including nested terms
- Statistics
 - Model parameters: Coefficient estimates, standard error for each coefficient estimate, t test for each coefficient estimate, confidence interval for each coefficient estimate, design effect for each coefficient estimate, and square root of the design effect for each coefficient estimate
 - Population means of dependent variable and covariates
 - Model fit- Sample design information
- Hypothesis tests
 - Test statistics: Wald F test, adjusted Wald F test, Wald Chi-square test, and adjusted Wald Chi-square test
 - Adjustment for multiple comparisons: Least significant difference, Bonferroni, sequential Bonferroni, Sidak, and sequential Sidak
 - Sampling degrees of freedom: Based on sample design or fixed by user
- Estimated means: Requests estimated marginal means for factors and interactions in the model
 - Contrasts: Simple, deviation, Helmert, repeated, or polynomial
- Model variables can be saved to the active file and/or exported to external files that contain parameter matrices
 - Variables: Predicted values and residuals
 - Parameter covariance matrix and its other statistics, as well as parameter correlation matrix and its other statistics, can be exported as a SPSS Statistics data file
 - Parameter estimates and/or the parameter covariance matrix can be exported to an XML file
- Output
 - Sample design information (such as strata and PSUs)
 - Regression coefficient estimates and t tests
 - Summary information about the dependent variable, covariates, and factors
 - Summary information about the sample, including the unweighted count and population size
 - Confidence limits for parameter estimates and user-specified confidence levels
 - Wald F test for model effects
 - Design effects
 - Multiple R²
 - Set of contrast coefficients (L) matrices
 - Variance-covariance matrix of regression coefficient estimates
 - Root mean square error
 - Covariance and correlation matrices for regression coefficients
- Missing data handling- Listwise deletion of missing values
- Other
 - User-specified denominator, df, used in computing p values for all test statistics
 - Collinearity diagnostics
 - Model can be fitted for subpopulations



Complex Samples Ordinal (CSORDINAL)

CSORDINAL performs regression analysis on a binary or ordinal polytomous dependent variable using the selected cumulative link function for samples drawn by complex sampling methods. The procedure estimates variances by taking into account the sample design used to select the sample, including equal probability and PPS methods, as well as WR and WOR sampling procedures. Optionally, CSORDINAL performs analyses for a subpopulation.

- Models
 - Main effects
 - All n-way interactions
 - Fully crossed
 - Custom, including nested terms
- Statistics:
 - Model parameters: Coefficient estimates, exponentiated estimates, standard error for each coefficient estimate, t test for each coefficient estimate, confidence interval for each coefficient estimate, design effect for each coefficient estimate, square root of the design effect for each coefficient estimate, covariances of parameter estimates, and correlations of the parameter estimates
 - Model fit: Pseudo R2 and classification table
 - Parallel lines tests: Wald tests of equal slopes, parameter estimates for generalized (unequal slopes) model, and covariances of parameter estimates for generalized unequal slopes) model
 - Summary statistics for model variables
 - Sample design information
- Hypothesis tests
 - Test statistics: Wald F test, adjusted Wald F test, Wald Chi-square test, and adjusted Wald Chi-square test
 - Adjustment for multiple comparisons: Least significant difference, Bonferroni, sequential Bonferroni, Sidak, and sequential Sidak
 - Sampling degrees of freedom: based on sample design or fixed by user
- Model variables can be saved to the active file and/or exported to external files that contain parameter matrices
 - Variables: Predicted category, probability of predicted category, probability of observed category, cumulative probabilities (one variable per category), predicted probabilities (one variable per category)
 - Export as SPSS Statistics data file: Parameter covariance matrix and other statistics, parameter correlation matrix and other statistics
 - Export as XML: Parameter estimates and/or the parameter covariance matrix to an XML file
- Three estimation methods: Newton-Raphson, Fisher Scoring, and Fisher Scoring followed by Newton-Raphson
- Cumulative link function to specify the model: Cauchit, complementary log-log, logit, negative log-log, and probit
- Cumulative odds ratios for the specified factor(s) or covariate(s). The subcommand is available only for LOGIT link.
- Output
 - Sample design information (such as strata and PSUs)
 - Summary information about the dependent variable, covariates, and factors
 - Summary information about the sample, including the unweighted count and the population size
 - Confidence limits for parameter estimates and user-specified confidence levels
 - Model summary statistics
 - Wald F test, adjusted Wald F test, Wald Chi-square, and adjusted Wald Chi-square for model effects
 - Design effects
 - Classification table
 - Set of contrast coefficients (L) matrices
 - Variance-covariance matrix of regression coefficient estimates
 - General estimable function table
 - Correlation matrix for regression coefficients



- Missing data handling
 - Listwise deletion of missing values
- Other
 - User-specified denominator, df, used in computing p values for all test statistics
 - Collinearity diagnostics
 - Fits model for a subpopulation

Complex Samples Logistic Regression (CSLOGISTIC)

This procedure performs binary logistic regression analysis, as well as multinomial logistic regression (MLR) analysis, for samples drawn by complex sampling methods. CSLOGISTIC estimates variances by taking into account the sample design used to select the sample, including equal probability and PPS methods, and WR and WOR sampling procedures. Optionally, CSLOGISTIC performs analyses for subpopulations

- Models
 - Main effects
 - All n-way interactions
 - Fully crossed
 - Custom, including nested terms
- Statistics
 - Model parameters: Coefficient estimates, exponential estimates, standard error for each coefficient estimate, t test for each coefficient estimate, confidence interval for each coefficient estimate, design effect for each coefficient estimate, square root of the design effect for each coefficient estimate, covariances of parameter estimates, and correlations of the parameter estimates
 - Model fit: Pseudo R2 and classification table
 - Summary statistics for model variables
 - Sample design information
- Hypothesis tests
 - Test statistics: Wald F test, adjusted Wald F test, Wald Chi-square test, and adjusted Wald Chi-square test
- Adjustment for multiple comparisons: Least significant difference, Bonferroni, sequential Bonferroni, Sidak, and sequential Sidak
- Sampling degrees of freedom: Based on sample design or fixed by user
- Model variables can be saved to the active file and/or exported to external files that contain parameter matrices
 - Variables: Predicted category and predicted probabilities
 - Parameter covariance matrix and its other statistics, as well as parameter correlation matrix and its other statistics, can be exported as an SPSS Statistics data file
 - Parameter estimates and/or the parameter covariance matrix can be exported to an XML file
- Output
 - Sample design information (such as strata and PSUs)
 - Summary information about the dependent variable, covariates, and factors
 - Summary information about the sample, including the unweighted count and population size
 - Confidence limits for parameter estimates and user-specified confidence levels
 - Model summary statistics
 - Wald F test for model effects
 - Design effects
 - Classification table
 - Set of contrast coefficients (L) matrices
 - Variance-covariance matrix of regression coefficient estimates- Root mean square error- Covariance and correlation matrices for regression coefficients
- Missing data handling
 - Listwise deletion of missing values
- Other
 - User-specified denominator, df, used in computing p values for all test statistics
 - Collinearity diagnostics
 - Model can be fitted for subpopulations



Complex Samples Cox Regression (CSCOXREG)

This procedure applies Cox proportional hazards regression to analysis of survival times - that is, the length of time before the occurrence of an event for samples drawn by complex sampling methods. CSCOXREG supports continuous and categorical predictors, which can be time-dependent. CSCOXREG provides an easy way of considering differences in subgroups as well as analyzing effects of a set of predictors. Also, the procedure handles data where there are multiple cases (such as patient visits, encounters, and observations) for a single subject

- Time and Event: specify survival time variables and values that indicate that the event of interest has occurred
 - Survival time°
 - Start of interval (onset of risk)
 - Time 0
 - Varies by subject°
 - End of interval
 - Event as individual values or a range of values
- Predictors:
 - Factors
 - Covariates
 - Time-dependent predictors
- Subgroups: stratify the analysis and/or limit it to a particular subpopulation.
- Models
 - Main effects
 - All n-way interactions
 - Custom, including nested terms
- Statistics:
 - Sample design information
 - Event and censoring summary
 - Risk set at event time
 - Model parameters: Coefficient estimates, exponentiated estimates, standard error for each coefficient estimate, t test for each coefficient estimate, confidence interval for each coefficient estimate, design effect for each coefficient estimate, square root of the design effect for each coefficient estimate, covariances of parameter estimates, and correlations of the parameter estimates
- Model assumptions
 - Test of proportional hazards
 - Parameter estimates for alternative model
 - Covariance matrix for alternative model- Baseline survival and cumulative hazard functions
 - Test of proportional hazards
 - Parameter estimates for alternative model
 - Covariance matrix for alternative model
 - Baseline survival and cumulative hazard functions
- Plots:
 - Survival function
 - Hazard function
 - Log minus log of the survival function
 - One minus survival function
 - Option to display confidence intervals
 - Plot factors and covariates at specified levels
- Hypothesis tests
 - Test Statistics: F test, Adjusted F test, Chi-square test. Adjusted Chi-square test
 - Adjustment for multiple comparisons: Least significant difference, Bonferroni, Sequential Bonferroni, Sidak, and sequential Sidak
 - Sampling degrees of freedom: based on sample design or fixed by user
- Save model variables to the active file and/or export external files that contain parameter matrices
 - Variables: Survival function, lower bound of confidence interval for survival function, upper bound of confidence interval for survival function, cumulative hazard function, lower bound of confidence interval for cumulative hazard function, upper bound of confidence interval for cumulative hazard function, predicted value of linear predictor, Schoenfeld residual (one variable per model parameter), Martingale residual, deviance residual, Cox-Snell residual, score residual (one variable per model parameter), DFBeta residual (one variable per model parameter), aggregated Martingale residual, aggregated deviance residual, aggregated Cox Snell residual, aggregated Score residual (one variable per model parameter), and aggregated DFBETA residual (one variable per model parameter)



- Export the model and/or the survival function- Export as SPSS Statistics data file
- Export survival function as SPSS Statistics data file
- Export model as XML file
- Options to specify estimation criteria, methods for computing survival functions and confidence intervals, and handling of user-missing values
 - Estimation: Maximum iterations, maximum step halving, limit iterations based on change in parameter estimates, limit iterations based on change in log-likelihood, display iteration history, and tie breaking method for parameter estimation (Efron or Breslow)
 - Survival functions: method for estimating baseline survival functions (Efron, Breslow or product-limit), and confidence intervals for survival functions transformed or original units)
 - Specify level of confidence interval
 - Missing Data Handling (treat as valid or invalid)



Statistics for Data Analysis solution

Add more analytical power, as you need it, with optional modules and stand-alone software from the Statistics for Data Analysis family.

Statistics Base

Statistics Base includes the core capabilities to take the analytical process from start to finish. It is easy to use and includes a broad range of procedures and techniques to increase revenue, outperform competitors, conduct research and make better decisions.

Statistics Advanced

Statistics Advanced includes these powerful multivariate techniques: generalized linear models (GENLIN), generalized estimating equations (GEE), mixed level models, general linear mixed models (GLMM), variance component estimation, MANOVA, Kaplan-Meier estimation, Cox regression, hiloglinear, loglinear and survival analysis.

Statistics Bootstrapping

Statistics Bootstrapping enables researchers and analysts to use bootstrapping techniques on a number of tests contained in Statistics for Data Analysis modules. This provides an efficient way to ensure that your models are stable and reliable. With Statistics Bootstrapping, you can reliably estimate the standard errors and confidence intervals of a population parameter like a mean, median, proportion, odds ratio, correlation coefficient, regression coefficient and numerous.

Statistics Categories

Unleash the full potential of your categorical data through perceptual maps with optimal scaling and dimension reduction techniques. This add-on module provides you with everything you need to analyze and interpret multivariate data and their relationships more completely.

Statistics Complex Samples

Incorporate complex sample designs into data analysis for more accurate analysis of complex sample data. Statistics Complex Samples, with specialized planning tools and statistics, reduces the risk of reaching incorrect or misleading inferences for stratified, clustered or multistage sampling.

Statistics Conjoint

Statistics Conjoint helps market researchers develop successful products. By performing conjoint analysis, you learn what product attributes are important in the consumer's mind and what the most preferred attribute levels are, and can perform pricing studies and brand equity studies.

Statistics Tables

Use Statistics Tables to present survey, customer satisfaction, polling and compliance reporting results. Features such as a table builder preview, included inferential statistics and data management capabilities make it easy to clearly communicate your results.



Statistics Preparation

With Statistics Preparation, you gain several procedures that facilitate the data preparation process. This add-on module enables you to easily identify suspicious and invalid cases, variables and data values; view patterns of missing data; summarize variable distributions to get your data ready for analysis; and more accurately work with algorithms designed for nominal attributes.

Statistics Decision Trees

Create highly visual classification and decision trees directly within Statistics for Data Analysis for segmentation, stratification, prediction, data reduction and variable screening, interaction identification, category merging and discretizing continuous variables. Highly visual trees enable you to present results in an intuitive manner.

Statistics Direct Marketing

Statistics Direct Marketing helps marketers perform various kinds of analyses easily and confidently, without requiring a detailed understanding of statistics. They can conduct recency, frequency and monetary value (RFM) analysis, cluster analysis, and prospect profiling. They can also improve marketing campaigns through postal code analysis, propensity scoring, and control package testing. And they can easily score new customer data and access pre-built models.

Statistics Exact Tests

Statistics Exact Tests always provides you with correct p values, regardless of your data structure, even if you have a

small number of cases, have subset your data into fine breakdowns or have variables where 80 percent or more of the responses are in one category.

Statistics Forecasting

Improve forecasting with complete time-series analyses, including multiple curve-fitting, smoothing models, methods for estimating autoregressive functions and temporal causal modeling. Use the Expert Modeler to automatically determine

which ARIMA (autoregressive integrated moving average) process or exponential smoothing model best fits your time-series and independent variables, eliminating selection through trial and error.

Statistics Missing Values

If values are missing from your data, this module may find some relationships between the missing values and other variables. In addition, the missing values module can estimate what the value would be if data weren't missing.

Statistics Neural Networks

Use the Statistics Neural Networks module to model complex relationships between inputs and outputs or to discover patterns in your data. Choose from algorithms that can be used for classification (categorical outcomes) and prediction (numerical outcomes). The two available algorithms are Multilayer Perceptron and Radial Basis Function.



Statistics Regression

Predict behavior or events when your data go beyond the assumptions of linear regression techniques. Perform multinomial or binary logistic regression and nonlinear regression, weighted least squares, two-stage least squares and probit analysis.

Complementary product

Use these products with Statistics for Data Analysis to enhance your analytical results.

Statistics Amos

Support your research and theories by extending standard multivariate analysis methods when using this stand-alone software package for structural equation modeling (SEM). Build attitudinal and behavioral models that more realistically reflect complex relationships, because any numeric variable, whether observed or latent, can be used to predict any other numeric variable. The latest release includes a new nongraphical method of model specification that improves accessibility for users who need scripting capabilities and enables large, complicated models to be run more quickly.