THE POWER TO KNOW.

SAS/STAT® 14.2 User's Guide

What's New in SAS/STAT 14.2

This document is an individual chapter from *SAS/STAT® 14.2 User's Guide*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2016. *SAS/STAT® 14.2 User's Guide*. Cary, NC: SAS Institute Inc.

SAS/STAT® 14.2 User's Guide

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

November 2016

Chapter 1
What's New in SAS/STAT 14.2

Contents

Overview

SAS/STAT 14.2 includes two new procedures and many enhancements. The high-performance procedures that are available in SAS High-Performance Statistics software for distributed computing are also available in SAS/STAT software for use in single-machine mode. These procedures are documented both in *SAS/STAT User's Guide* and in *SAS/STAT User's Guide: High-Performance Procedures*.

New Procedures

CAUSALTRT Procedure

The CAUSALTRT procedure estimates the average causal effect of a binary treatment variable T on a continuous or discrete outcome Y. Depending on the application, the variable T can represent an intervention (such as smoking cessation versus control), an exposure to a condition (such as attending private versus public schools), or an existing characteristic of subjects (such as high versus low socioeconomic status). The CAUSALTRT procedure estimates two types of causal effects: the average treatment effect and the average treatment effect for the treated.

The causal inference methods that the CAUSALTRT procedure implements are designed primarily for use with data from nonrandomized trials or observational studies, where you observe T and Y without assigning subjects randomly to the treatment conditions. Instead, subjects select themselves into the treatment conditions according to their pretreatment (baseline) characteristics. If these characteristics are also associated with the outcome Y, they confound the relationship between T and Y, clouding the causal interpretation of T on Y. Therefore, estimating the causal effect of T usually requires adjustments that remove or reduce the effects of confounding.

You can adjust for confounding by using the CAUSALTRT procedure to model the treatment assignment or the outcome or both. Modeling the treatment assignment leads to inverse probability weighting methods, and modeling the outcome leads to regression adjustment methods. Modeling both leads to doubly robust methods that can provide unbiased estimates for the treatment effect even if one of the models is misspecified.

PSMATCH Procedure

The PSMATCH procedure provides a variety of tools for propensity score analysis, which is a general strategy for reducing the effects of confounding in observational studies, where the subjects are not randomly assigned to the treatment and control groups. Confounding can occur if some covariates are associated with both the treatment assignment and the outcome. If there are systematic differences between the distributions of the covariates in the treatment and control groups, a regression analysis will not produce valid inference.

The PSMATCH procedure computes or reads propensity scores, which estimate the probability that a subject is assigned to treatment given a set of pretreatment (baseline) covariates. The procedure provides the following methods for using the propensity scores to adjust the data for valid estimation of treatment effect:

- Inverse probability of treatment weighting and weighting by the odds. The procedure computes the weights for subsequent estimation of treatment effect that uses a weighted analysis.
- Stratification. The procedure creates strata of observations that have similar propensity scores. In a subsequent outcome analysis, the treatment effect can be estimated within each stratum, and the estimates can be combined across strata to compute an average treatment effect.
- Matching. The procedure matches each treated unit with one or more control units that have a similar value of the propensity score. In a subsequent outcome analysis, the treatment effect can be estimated by comparing outcomes between treatment and control groups in the matched sample.

The PSMATCH procedure also provides statistical and graphical methods for assessing how well the distributions of the propensity score and the covariates are balanced between the adjusted treatment and control groups.

Assuming that this balance is adequate, you can use the output data set that the procedure creates in a subsequent outcome analysis, which you can perform by using a broad variety of SAS/STAT procedures. The PSMATCH procedure does not carry out an outcome analysis, nor does it involve the outcome variable.

Highlights of Enhancements

Following are some highlights of the enhancements in SAS/STAT 14.2:

- The FREQ and SURVEYFREQ procedures provide additional agreement statistics.
- The NLIN procedure now provides ESTIMATE and CONTRAST statements.
- The NLMIXED procedure supports multithreading in models that have more than one RANDOM statement.
- The PHREG procedure now provides time-dependent ROC analysis.
- The POWER procedure now provides extensions of existing power analyses that are applicable to various generalized linear models.
- The SURVEYIMPUTE procedure provides two-stage fully efficient fractional imputation and fractional hot-deck imputation.
- The SURVEYSELECT procedure now provides balanced bootstrap selection and sequential Poisson selection.

For more information about the changes and enhancements, see the section "Enhancements" on page 4. Details can be found in the documentation for the individual procedures in *SAS/STAT User's Guide*.

Highlights of Enhancements in SAS/STAT 14.1

Some users might be unfamiliar with updates that were made in the previous release. SAS/STAT 14.1 introduced the GAMPL and SURVEYIMPUTE procedures. Following are highlights of the other enhancements in SAS/STAT 14.1:

- The BCHOICE procedure allows varying numbers of alternatives in choice sets for logit models.
- Exact mid-p, likelihood ratio, and Wald modified confidence limits are available for the odds ratio produced by the FREQ procedure.
- The GLIMMIX procedure provides the multilevel adaptive Gaussian quadrature algorithm of Pinheiro and Chao (2006) for multilevel models, which can greatly reduce the computational and memory requirements for these models when they include many random effects.
- The GLMSELECT procedure supports the group LASSO method.
- The IRT procedure fits generalized partial credit models.

- The LIFETEST procedure performs nonparametric analysis of competing-risks data.
- The LOGISTIC procedure fits an adjacent-category logit model to ordinal response data.
- The MCMC procedure adds an ordinary differential equation (ODE) solver and a general integration function, enabling the procedure to fit models that contain differential equations (for example, PK models) or models that require integration (for example, marginal likelihood models).
- The NPAR1WAY procedure performs stratified rank-based analysis for two-sample data.
- The POWER procedure supports Cox proportional hazards regression models.

Enhancements

BCHOICE Procedure

MaxDiff (maximum difference) and allocation (chip-sum) choice models have been implemented. Previously, PROC BCHOICE fit only the traditional discrete choice model, where the response takes a binary integer value: 1 for a chosen best alternative and 0 for any unchosen alternatives. In a MaxDiff choice experiment, each respondent is shown a set of possible items and chooses not only the best item (indicated by 1), but also the worst item (indicated by -1) in each choice set. In an allocation choice experiment, each respondent assigns a percentage of preference to each alternative, and the sum of percentages across all alternatives in each choice set is 100%.

FREQ Procedure

The new AGREE(PABAK) and AGREE(AC1) options produce the prevalence-adjusted, bias-adjusted kappa coefficient and Gwet's AC1 agreement coefficient, respectively. The AGREE(KAPPADETAILS) option provides the following statistics: observed agreement and chance-expected agreement components of the simple kappa coefficient, maximum possible kappa, prevalence index, and bias index. The AGREE(WTKAPDETAILS) option provides the observed agreement and chance-expected agreement components of the weighted kappa coefficient.

You can now specify nonzero null values for the simple and weighted kappa tests by using the AGREE(NULLKAPPA=) and AGREE(NULLWTKAP=) options. You can specify the degrees of freedom for Bowker's test of symmetry by using the AGREE(DFSYM=) option.

When you specify any of the AGREE options that are new in SAS/STAT 14.2, PROC FREQ displays all tables of AGREE statistics in tabular format (instead of factoid format); to display preexisting tables in their original factoid format, you can specify the AGREE(TABLES=RESTORE) option.

GEE Procedure

The GEE procedure now includes the EFFECTPLOT, LSMESTIMATE, and SLICE statements for additional postfitting inferences.

The GEE procedure also provides a STORE statement, which enables you to save the context and results of statistical analysis for further processing by the PLM procedure.

HPSPLIT Procedure

The PLOTS=CVCC(ASE) suboption has been added to the PROC HPSPLIT statement. You can use this suboption to request the display of ASE (average square error) in the cost-complexity plot that is produced when the response is a classification variable and cross validation is used to assess cost-complexity pruning. By default, the plot displays the misclassification rate.

IRT Procedure

The FIXVALUE statement enables you to specify fixed values for any types of parameters within or between independent samples.

The MEAN statement enables you to specify the factor means of the confirmatory IRT model that you specify in the FACTOR statement. You can specify these means as free, fixed, or constrained parameters.

The CEILPRIOR=, GUESSPRIOR=, and SLOPEPRIOR= options in the PROC IRT and MODEL statements enable you to specify prior distributions of the ceiling, guessing, and slope parameters for the three- and four-parameter models. These prior distributions help keep the ceiling and guessing parameter estimates within the boundaries of 0 and 1.

MCMC Procedure

The new NORMALCAR option in the RANDOM statement specifies a spatial conditional autoregressive (CAR) prior that can be used to model spatial correlations among sites and neighbors.

NLIN Procedure

The new ESTIMATE statement enables you to estimate arbitrary functions of model parameters. Multiple ESTIMATE statements are supported. The ECOV, ECORR, and EDER options in the PROC NLIN statement compute, respectively, the approximate covariance matrix, correlation matrix, and derivatives (with respect to model parameters) of the input estimates.

The new CONTRAST statement enables you to test a hypothesis that one or more expressions are simultaneously equal to 0. The expressions are typically contrasts—that is, differences whose expected values equal 0 under the null hypothesis. You can specify multiple CONTRAST statements.

NLMIXED Procedure

PROC NLMIXED supports multithreading in models that have more than one RANDOM statement. Prior to SAS/STAT 14.2, the NTHREADS= option is ignored in models that have more than one RANDOM statement.

PHREG Procedure

The PHREG procedure now provides concordance statistics and time-dependent ROC curves for assessing predictive accuracy.

The ROC statement specifies a model to be used in computing concordance statistics or ROC curves.

The CONCORDANCE option in the PROC PHREG statement produces concordance statistics for the model that is specified in the MODEL statement and for a model that is specified in an ROC statement.

The PLOTS=ROC option in the PROC PHREG statement plots time-dependent ROC curves for the model that is specified in the MODEL statement and for a model that is specified in an ROC statement. The PLOTS=AUC option displays the areas under the ROC curves, and the PLOTS=AUCDIFF option displays the difference of the AUC curves between models. You can use the ROCOPTIONS option in the PROC PHREG statement to specify the time points at which the ROC curves are computed and to create tables and output data sets for the analysis of the ROC curves.

POWER Procedure

The new CUSTOM statement performs power and sample size analyses for extensions of existing analyses that involve the chi-square, F, t, normal, or correlation coefficient distribution. Use cases include logistic regression with classification variables, Poisson regression, zero-inflated models, and other generalized linear models.

SURVEYFREQ Procedure

The new AGREE(KAPPADETAILS) option displays the following information: observed agreement and chance-expected agreement components of the simple kappa coefficient, maximum possible kappa, prevalence index, and bias index. The AGREE(WTKAPDETAILS) option displays the observed agreement and chance-expected agreement components of the weighted kappa coefficient.

The DISCORDDIFF option provides the estimate, confidence limits, and test of the difference between discordant proportions in 2×2 tables.

The CL, CV, NOSTD, and VAR suboptions are now available for the ROW and COLUMN options. You can use these suboptions to control display of crosstabulation table statistics for row and column percentages (separately from statistics for overall percentages). The RISKDIFF(COLUMN=) option specifies the table column for which to compute risk difference statistics, and the RISKDIFF(ONLY) option displays only the risk difference (and suppresses display of the risks).

SURVEYIMPUTE Procedure

PROC SURVEYIMPUTE now includes two-stage fully efficient fractional imputation and fractional hot-deck imputation. These methods are especially useful for imputing missing values in continuous variables such as age or salary. The CLEVELS= and CLEVVAR= options in the VAR statement along with the METHOD=FEFI option in the PROC SURVEYIMPUTE statement produce two-stage fully efficient fractional imputation. The new METHOD=FHDI option in the PROC SURVEYIMPUTE statement provides fractional hot-deck imputation.

The new REPWTADJ= option for METHOD=FHDI provides three replicate weight adjustment methods, and the new SELECTION= option for METHOD=FHDI provides two donor selection methods for fractional hot-deck imputation.

SURVEYMEANS Procedure

You can request that PROC SURVEYMEANS display specific domain levels in the output by listing quoted formatted-level values in parentheses after each variable name in the DOMAIN statement.

You can now request a covariance matrix of domain means by specifying the COV option in the DOMAIN statement.

You can now compare domain means in PROC SURVEYMEANS for continuous variables by specifying the DIFFMEANS options in the DOMAIN statement. PROC SURVEYMEANS provides differences between domain means for pairwise levels of a defined domain, the standard error of the difference, and the *t* test. You can also specify ADJUST=BON option to request a Bonferroni multiple comparison adjustment of the *p*-values for the *t* tests.

SURVEYSELECT Procedure

The SURVEYSELECT procedure now provides balanced bootstrap selection and sequential Poisson selection. The new ROUND= option provides additional rounding methods to use when converting the sampling rate to an integer-valued sample size.

What's Changed

The following section describes a change in software behavior from SAS/STAT 14.1 to SAS/STAT 14.2.

HPSPLIT Procedure

In SAS/STAT 14.1, when cross validation is used to assess cost-complexity pruning and the response is a classification variable, PROC HPSPLIT displays the average square error (ASE) by default on the cost-complexity plot. In SAS/STAT 14.2, PROC HPSPLIT displays the misclassification rate by default in this

case. If you want to display the ASE instead, specify the new PLOTS=CVCC(ASE) option in the PROC HPSPLIT statement.

Power and Sample Size Application

The functionality of the Power and Sample Size application has been replaced by tasks in SAS Studio.

References

Pinheiro, J. C., and Chao, E. C. (2006). "Efficient Laplacian and Adaptive Gaussian Quadrature Algorithms for Multilevel Generalized Linear Mixed Models." *Journal of Computational and Graphical Statistics* 15:58–81.