THE POWER TO KNOW.

SAS/STAT® 15.1 User's Guide

What's New in SAS/STAT 15.1

This document is an individual chapter from *SAS/STAT® 15.1 User's Guide*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2018. *SAS/STAT® 15.1 User's Guide*. Cary, NC: SAS Institute Inc.

SAS/STAT® 15.1 User's Guide

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

November 2018

Chapter 1
What's New in SAS/STAT 15.1

Contents

Overview

SAS/STAT 15.1 includes new procedures and many enhancements.

New Procedures

BGLIMM Procedure

The BGLIMM procedure provides full Bayesian inference for generalized linear mixed models (GLMMs). It models data from the exponential family distributions that have correlations or nonconstant variability. The features of PROC BGLIMM include the following:

- syntax similar to that of PROC MIXED and PROC GLIMMIX
- optimal sampling algorithms that are parallelized for performance
- multilevel nested and non-nested random-effects models
- the ability to fit models to multivariate data

PROC BGLIMM provides convenient access, with improved performance, to Bayesian analysis of complex mixed models that you can otherwise perform by using the MCMC procedure.

CAUSALGRAPH Procedure

The CAUSALGRAPH procedure examines the structure of graphical causal models and suggests statistical strategies that enable researchers to compute unbiased estimates of causal effects. Essentially, in situations where there are possible spurious association and unmeasured confounding between treatment and outcome variables, a causal effect cannot be determined by the observed association directly. Graphical causal models are a powerful and convenient tool that enable you to remove such noncausal influences and obtain valid causal inferences. This is particularly important in observational studies where the treatment conditions are not randomly assigned to individuals. You can use PROC CAUSALGRAPH to define graphical causal models in the form of directed acyclic graphs (DAGs). Causal effects within a DAG are interpreted in a counterfactual framework (Pearl 2009), which is consistent with the Neyman-Rubin potential outcome framework (Rubin 1990). PROC CAUSALGRAPH provides several tools that you can use to explore the formal properties of a causal graph, including the following:

- adjustment sets that can be used to remove or block noncausal associations between the treatment and outcome variables
- sets of variables that can be used as instruments to estimate a causal effect
- causal and noncausal (associative) paths between the treatment and outcome variables

You can use the CAUSALGRAPH procedure to test the validity of an existing causal effect identification criterion or to construct possible identification criteria. The procedure can also analyze multiple causal

models simultaneously to obtain common identification criteria. The identification results from PROC CAUSALGRAPH enable researchers to devise sound statistical strategies to estimate causal effects in confounding situations.

RMSTREG Procedure

The RMSTREG procedure analyzes time-to-event data by using regression with respect to the restricted mean survival time (RMST), using specialized methods such as those developed by Andersen, Hansen, and Klein (2004) and Tian, Zhao, and Wei (2014) . The RMST is defined as the expected value of the time-to-event variable up to a prespecified time point τ. It is an alternative measure that is often more reliably estimable than the mean and median survival time in certain situations. Also, it provides a summary of the whole survival curve up to a time horizon, in contrast to the survival rate at a specified time. Compared to Cox regression and other classical methods, the specialized methods in PROC RMSTREG enable you to model the RMST directly, facilitating model-based inference and prediction.

Highlights of Enhancements

Following are some highlights of the enhancements in SAS/STAT 15.1:

- In the ANOVA procedure, you can specify the EFFECTSIZE option in the MODEL statement to add measures of effect size to each analysis-of-variance table.
- The CAUSALMED procedure enables you to input observational weights in the WEIGHT statement and to input the standard deviations of continuous variables in the STD statement.
- The FMM procedure supports random starting values for maximum likelihood estimation.
- The FREQ and SURVEYFREQ procedures offer the SENSPEC option in the TABLES statement to provide estimates and confidence limits for sensitivity, specificity, positive predictive value, and negative predictive value.
- The ICPHREG procedure fits the semiparametric proportional hazards model to interval-censored data.
- The MCMC and NLMIXED procedures support a steady state option for one-, two-, and three-compartment models.
- The PHREG procedure performs Bayesian analysis of the proportional hazards spline model.
- The PSMATCH procedure provides a PSWEIGHT statement to compute weights for observations on the basis of propensity scores.
- The QUANTREG procedure enables you to perform observationwise conditional distribution analysis.
- The SURVEYFREQ, SURVEYLOGISTIC, SURVEYMEANS, and SURVEYREG procedures provide two methods of computing the deviations for the replication variance estimation methods.
- The TTEST procedure produces graphs of bootstrap distributions and confidence intervals.

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 14.3

Some users might be unfamiliar with updates that were made in the previous release. SAS/STAT 14.3 introduced the CAUSALMED procedure. Following are highlights of the other enhancements in SAS/STAT 14.3:

- The GAMPL procedure supports the Tweedie distribution.
- In the FREQ procedure, the COMMONRISKDIFF option in the TABLES statement provides estimates, confidence limits, and tests for the overall risk (proportion) difference for multiway 2×2 tables.
- The IRT procedure supports the nominal response model, which enables you to do item analysis of nominal responses.
- The NLMIXED and MCMC procedures add the CMPTMODEL statement, which fits compartment models in pharmacokinetic analysis.
- The PHREG procedure provides cause-specific proportional hazards analysis for competing-risks data.
- The QUANTREG and QUANTSELECT procedures provide fast quantile process regression.
- The VARMETHOD=BOOTSTRAP option provides variance estimation by the bootstrap method for the survey data analysis procedures.
- The TTEST procedure provides bootstrap standard error, bias estimates, and confidence limits.

Enhancements

ANOVA Procedure

In PROC ANOVA, you can specify the EFFECTSIZE option in the MODEL statement to add measures of effect size to each analysis-of-variance table that is displayed by the procedure, except for those displayed by the MANOVA, REPEATED, and TEST statements. The effect size measures include the intraclass correlation and both estimates and confidence intervals for the noncentrality for the F test, as well as for the semipartial and partial correlation ratios. For more information about the computation and interpretation of these measures, see the section "Effect Size Measures for F Tests in GLM" on page 4032 in Chapter 50, "The GLM Procedure."

BOXPLOT Procedure

The BOXPLOT procedure's PLOT statement has three new options:

- The BLOCKREF option produces vertical reference lines that delineate blocks of box plots that are defined by the block variables.

- The BLOCKREFFILL option fills the graph walls of blocks with colors that you specify in the BLOCKVAR= option.
- The BLOCKVALUEPOS= option specifies the alignment of block variable labels in the block legend.

These options are available only when ODS Graphics is enabled.

CAUSALMED Procedure

The CAUSALMED procedure supports the WEIGHT and STD statements. You can input observational weights in the WEIGHT statement and the standard deviations of continuous variables in the STD statement.

The bootstrapping results include a new table that shows the number of converged samples and diagnostic information for the nonconvergent bootstrap samples. The new IGNORECC option in the BOOTSTRAP statement ignores the case-control design and performs bootstrap sampling from the entire pool of observations without distinguishing the treatment conditions of the observations.

FMM Procedure

The FMM procedure supports random starting values for maximum likelihood (ML) estimation. This feature enables you to generate multiple sets of random starting values for ML estimation and automatically finds the best ML estimates from a two-stage fitting process. You use the RANDSTART statement to request this new random starting method and fine-tune the two-stage fitting process.

FREQ Procedure

The SENSPEC option in the TABLES statement provides estimates and confidence limits for sensitivity, specificity, positive predictive value, and negative predictive value.

The COMMONRISKDIFF(CL=KLINGENBERG) option provides Klingenberg confidence limits for the common risk difference. The COMMON= option for the risk difference plot specifies the confidence limit type for the common value. Available confidence limit types include Mantel-Haenszel, minimum risk, summary score, Newcombe, Newcombe with minimum risk weights, and Klingenberg. The COMMONRISKDIFF(PRINTWTS=SCORE) option includes the summary score weights in the "Stratum Risk Differences and Weights" table.

The CMH(I2) option provides Higgins's I-square measure of heterogeneity for odds ratios in multiway 2×2 tables.

GLM Procedure

In PROC GLM, if you specify a model that has two CLASS variables, and one variable is nested within the other, the procedure produces a nested box plot of the response values, where horizontal position represents one CLASS variable nested within the other CLASS variable.

GLMSELECT Procedure

In PROC GLMSELECT, there are two new SELECTION= method options for displaying the details of the selection process for elastic net and LASSO selection.

- For SELECTION=ELASTICNET, the SHOWL2SEARCH method option displays an elastic net parameter summary table, which shows the chosen values for the LASSO regularization parameter L1 and the ridge regularization parameter L2 (in the ratio scaling).
- For SELECTION=LASSO or SELECTION=GROUPLASSO, the SHOWSTEPL1 option adds to the selection summary table a column that shows the value of the LASSO regularization parameter L1 for each step of the LASSO selection process, with the selected step identified.

ICPHREG Procedure

In the ICPHREG procedure, the BASE=UNSPECIFIED option in the MODEL statement fits the semiparametric proportional hazards model to interval-censored data. The ID statement enables you to fit the semiparametric model that contains time-dependent covariates. The EM and EMICM algorithms can fit these models efficiently.

LIFETEST Procedure

In the LIFETEST procedure, you can analyze the restricted mean survival time (RMST) and the restricted mean time lost (RMTL). The RMST and RMTL options in the PROC LIFETEST statement compute the RMST and RMTL at specified times. The STRATA statement supports comparison of the estimated RMST and RMTL; the DIFF= and ADJ= options enable you to compute pairwise differences and perform various multiple-comparison adjustments. The PLOT=RMST and PLOT=RMTL options in the PROC LIFETEST statement plot the RMST and RMTL curves.

LOGISTIC Procedure

In the LOGISTIC procedure, the GOF option performs goodness-of-fit tests that remain valid even when you have continuous covariates or sparse data. This option also computes a larger assortment of model-fit statistics, including a set of pseudo-R-squares.

A plot that displays the loess fit of the observed response against the model predicted probability is available when you specify the PLOTS=CALIBRATION option in the PROC LOGISTIC statement.

MCMC Procedure

The CMPTMODEL statement in PROC MCMC fits one-, two-, and three-compartment models to subjects who are in steady-state conditions.

NLMIXED Procedure

The CMPTMODEL statement in PROC NLMIXED fits one-, two-, and three-compartment models to subjects who are in steady-state conditions.

NPAR1WAY Procedure

In the NPAR1WAY procedure, the RMD option in the PROC NPAR1WAY or EXACT statement provides the ratio mean deviations exact test for two-sample data. The VARINFO option in the PROC NPAR1WAY statement displays a data summary table for each analysis variable. The ADJUST option, which adjusts by class medians before performing the analysis, is available for SCORES=DATA analysis.

PHREG Procedure

In the PHREG procedure, the SPLINE option in the BAYES statement performs a Bayesian analysis of the proportional hazards spline model. This is a proportional hazards model in which the logarithm of the baseline cumulative hazard function is a cubic spline of the log of time.

PSMATCH Procedure

The PSMATCH procedure provides a PSWEIGHT statement to compute weights for the observations on the basis of propensity scores.

QUANTREG Procedure

The new CONDDIST statement in PROC QUANTREG enables you to perform observationwise conditional distribution analysis that can estimate and visualize cumulative distribution functions, probability density functions, and conditional percentages.

SURVEYFREQ Procedure

In PROC SURVEYFREQ, the SENSPEC option in the TABLES statement provides estimates and confidence limits for sensitivity, specificity, positive predictive value, and negative predictive value.

The CENTER= option in the PROC SURVEYFREQ statement specifies how to compute the deviations for replication variance estimation methods (BOOTSTRAP, BRR, and JACKKNIFE). CENTER=REPLICATES computes the deviations of the replicates from the replicate average, and CENTER=FULLSAMPLE computes the deviations of the replicates from the full sample estimate. By default, CENTER=FULLSAMPLE, which is the method used in releases before SAS/STAT 15.1.

The DEFF(FPC=) option in the PROC SURVEYFREQ statement controls whether to include the finite population correction in the SRS variance component of the design effect.

SURVEYIMPUTE Procedure

The SURVEYIMPUTE procedure provides the RATE= or TOTAL= option in the PROC SURVEYIMPUTE statement to specify the sampling rate, which is used to compute a finite population correction for bootstrap replicate weights.

The new LIST option in the STRATA statement displays a "Stratum Information" table, which provides the number of observations, the number of clusters, the sum of weights, and the sampling fraction in each stratum.

SURVEYLOGISTIC Procedure

In the SURVEYLOGISTIC procedure, the CENTER= option in the PROC SURVEYLOGISTIC statement defines how to compute the deviations for the replication variance estimation methods. You can use either the full sample estimate (default) or the average of the replicate estimates.

SURVEYMEANS Procedure

In the SURVEYMEANS procedure, the CENTER= option in the PROC SURVEYMEANS statement defines how to compute the deviations for the replication variance estimation methods. You can use either the full sample estimate (default) or the average of the replicate estimates.

The procedure computes the design effect for the mean or proportion estimates when you specify the DEFF statistic keyword in the PROC SURVEYMEANS statement.

SURVEYPHREG Procedure

PROC SURVEYPHREG provides Firth's penalized likelihood method for proportional hazards regression models. The new FIRTH option in the MODEL statement is useful when the likelihood is monotone, indicating that the likelihood converges to a finite value but at least one estimate diverges to infinity.

The new NAMELEN= option in the PROC SURVEYPHREG statement specifies the length of effect names in tables and output data sets.

SURVEYREG Procedure

In the SURVEYREG procedure, the CENTER= option in the PROC SURVEYREG statement defines how to compute the deviations for the replication variance estimation methods. You can use either the full sample estimate (default) or the average of the replicate estimates.

TTEST Procedure

The TTEST procedure produces graphs of bootstrap distributions and confidence intervals. The new BOOTDATA= option creates an output data set of statistics across bootstrap samples, and the FREQ statement is supported in bootstrap analyses.

Modeling Procedures

This new feature applies to the following procedures: GLIMMIX, GLMSELECT, HPMIXED, LOGISTIC, ORTHOREG, PHREG, PLS, QUANTLIFE, QUANTREG, QUANTSELECT, ROBUSTREG, SURVEYLOGISTIC, and SURVEYREG.

For spline effects that you incorporate in your linear model by using a statement of the form

EFFECT *effect-name* = **SPLINE** (*variable* < / *effect-options* >) **;**

you can use the KNOTMETHOD=PERCENTILELIST(*percentiles*) *effect-option* to place the internal spline knots at the percentiles of the variable that you specify. For example, the following statement specifies five internal knots at the particular percentiles of the variable x that are suggested by Harrell (2001):

```
EFFECT spl = spline(x / knotmethod=percentilelist(5 27.5 50 72.5 95));
```

What's Changed

The following sections describe changes in software behavior from SAS/STAT 14.3 to SAS/STAT 15.1.

CAUSALMED Procedure

For bootstrap sampling that you specify using the CASECONTROL option, PROC CAUSALMED draws bootstrap samples from the treatment and control conditions separately. This change affects analyses with binary treatment variables only. Previously, bootstrap samples were drawn from the entire pool of observations without distinguishing the treatment conditions of the observations. You can use the new IGNORECC option in the BOOTSTRAP statement to restore the previous bootstrap sampling scheme.

LOGISTIC Procedure

In PROC LOGISTIC, the maximum number of response levels that are allowed for polytomous response models is 100. You can now allow more levels by specifying the MAXRESPONSELEVELS= option in the PROC LOGISTIC statement.

The CTABLE option displays the positive and negative predictive values. The previous FALSPOS and FALSNEG values are still available from the ODS Output table, but they were replaced because of confusing terminology.

NPAR1WAY Procedure

PROC NPAR1WAY displays all score test and analysis tables in tabular format. You can specify the TABLES=RESTORE option in the PROC NPAR1WAY statement to display these tables in factoid (label-value) format, which is their format in releases before SAS/STAT 15.1.

PSMATCH Procedure

In PROC PSMATCH, the NLARGESTWGT=, WEIGHT=ATEWGT, and WEIGHT=ATTWGT options in the ASSESS statement move to the new PSWEIGHT statement.

The following options in the ASSESS statement move to other statements:

- The NLARGESTWGT= option moves to the new PSWEIGHT statement.
- The NMATCHMOST= option moves to the MATCH statement.
- The STRATUMWGT= option moves to the STRATA statement.

In addition, the WEIGHT=ATEWGT and WEIGHT=ATTWGT options move to the PSWEIGHT statement, and the WEIGHT=MATCHATEWGT, WEIGHT=MATCHATTWGT, WEIGHT=MATCHWGT, and WEIGHT=NONE options move to the MATCH statement.

The OUTPUT statement has the following changes:

- The ATEWGT= option is replaced by the WEIGHT= option in the OUTPUT statement and by the accompanying WEIGHT=ATEWGT option in the PSWEIGHT statement.
- The ATTWGT= option is replaced by the WEIGHT= option in the OUTPUT statement and by the accompanying WEIGHT=ATTWGT option in the PSWEIGHT statement.
- The MATCHATEWGT= option is replaced by the WEIGHT= option in the OUTPUT statement and by the accompanying WEIGHT=ATEWGT option in the MATCH statement.
- The MATCHATTWGT option is replaced by the WEIGHT= option in the OUTPUT statement and by the accompanying WEIGHT=ATTWGT option in the MATCH statement.

Note that the original syntax is still supported.

References

Andersen, P. K., Hansen, M. G., and Klein, J. P. (2004). "Regression Analysis of Restricted Mean Survival Time Based on Pseudo-observations." *Lifetime Data Analysis* 10:335–350.

Harrell, F. E., Jr. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge: Cambridge University Press.

Rubin, D. B. (1990). "Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies." *Statistical Science* 5:472–480.

Tian, L., Zhao, L., and Wei, L. J. (2014). "Predicting the Restricted Mean Event Time with the Subject's Baseline Covariates in Survival Analysis." *Biostatistics* 15:222–233.