



THE
POWER
TO KNOW.

SAS/STAT[®] 12.1 User's Guide

What's New in SAS/STAT

12.1

(Chapter)

This document is an individual chapter from *SAS/STAT® 12.1 User's Guide*.

The correct bibliographic citation for the complete manual is as follows: SAS Institute Inc. 2012. *SAS/STAT® 12.1 User's Guide*. Cary, NC: SAS Institute Inc.

Copyright © 2012, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

For a Web download or e-book: Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

The scanning, uploading, and distribution of this book via the Internet or any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions and do not participate in or encourage electronic piracy of copyrighted materials. Your support of others' rights is appreciated.

U.S. Government Restricted Rights Notice: Use, duplication, or disclosure of this software and related documentation by the U.S. government is subject to the Agreement with SAS Institute and the restrictions set forth in FAR 52.227-19, Commercial Computer Software-Restricted Rights (June 1987).

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

Electronic book 1, August 2012

SAS® Publishing provides a complete selection of books and electronic products to help customers use SAS software to its fullest potential. For more information about our e-books, e-learning products, CDs, and hard-copy books, visit the SAS Publishing Web site at support.sas.com/publishing or call 1-800-727-3228.

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

Chapter 1

What's New in SAS/STAT 12.1

Contents

Overview	2
New Procedures	2
Highlights of Enhancements	3
Highlights of Enhancements in SAS/STAT 9.3	4
New Macros	4
Enhancements	5
CALIS Procedure	5
EFFECTPLOT Statement	5
FMM Procedure	5
FREQ Procedure	5
GLIMMIX Procedure	6
LIFEREG Procedure	6
LIFETEST Procedure	6
LOESS Procedure	6
LOGISTIC Procedure	7
MCMC Procedure	7
MULTTEST Procedure	7
NLIN Procedure	8
NPAR1WAY Procedure	8
ODS Graphics	8
PHREG Procedure	8
POWER Procedure	8
PROBIT Procedure	9
QUANTREG Procedure	9
REG Procedure	9
ROBUSTREG Procedure	9
SEQDESIGN Procedure	9
SEQTEST Procedure	9
SURVEYFREQ Procedure	10
SURVEYMEANS Procedure	10
SURVEYPHREG Procedure	10
SURVEYREG Procedure	10
SURVEYSELECT Procedure	10
What's Changed	10
References	11

Overview

SAS/STAT 12.1 includes four new procedures and many enhancements.

In previous years, SAS/STAT[®] software was updated only with new releases of Base SAS[®] software, but this is no longer the case. This means that SAS/STAT software can be released to customers when enhancements are ready, and the goal is to update SAS/STAT every 12 to 18 months. To mark this newfound independence, the release numbering scheme for SAS/STAT is changing with this release. This new numbering scheme will be maintained when new versions of Base SAS and SAS/STAT ship at the same time. For example, when Base SAS 9.4 is released, SAS/STAT 13.1 will be released.

New Procedures

New Experimental ADAPTIVEREG Procedure

The ADAPTIVEREG procedure fits multivariate adaptive regression splines. The method is a nonparametric regression technique that combines both regression splines and model selection methods. It constructs spline basis functions in an adaptive way by automatically selecting appropriate knot values for different variables. The procedure performs model reduction by applying model selection techniques. Thus, the ADAPTIVEREG procedure is both a nonparametric regression procedure and a predictive modeling procedure.

The ADAPTIVEREG procedure supports models with classification variables, and it provides options for improving modeling speed. PROC ADAPTIVEREG extends the method to data with response variables that are distributed in the exponential family, including the binomial, Poisson, negative binomial, gamma, and inverse Gaussian distributions. PROC ADAPTIVEREG is multithreaded, enabling it to take advantage of multiple processors.

New Experimental QUANTLIFE Procedure

The QUANTLIFE procedure implements two quantile regression approaches that have been developed to account for right-censoring and provide valid estimates. Portnoy (2003) proposed a method to estimate conditional quantile functions from survival data by generalizing the idea of the Kaplan-Meier estimator of the survival function, and Peng and Huang (2008) developed a quantile regression approach motivated by the Nelson-Aalen estimator of the cumulative hazard function. Both methods are implemented with linear programming algorithms in the QUANTLIFE procedure. Like the standard quantile regression method for uncensored data, these two methods are distribution-free and are applicable to heteroscedastic data.

The QUANTLIFE procedure produces survival plots, conditional quantile plots, and quantile process plots. It performs semiparametric quantile regression when you specify spline effects. PROC QUANTLIFE takes advantage of parallel computing when multiple processors are available.

New Experimental QUANTSELECT Procedure

The QUANTSELECT procedure performs model selection for linear quantile regression. It provides capabilities similar to those offered by the GLMSELECT procedure (provides model selection for univariate linear models) including forward, backward, stepwise, LASSO, and adaptive LASSO selection methods. You can specify criteria such as AIC, SBC, AICC, adjusted R², and significance levels to compare the fit of

models, to determine when to stop the model selection process, and to choose the final selection model. The QUANTSELECT procedure supports constructed effects (such as smoothing splines and polynomials) and the SPLIT option for its CLASS statement. PROC QUANTREG produces parameter estimates progression plots and a variety of criterion progression plots.

PROC QUANTSELECT supports constructed effects such as regression spline, and it enables you to partition your data into training, validation, and testing roles. It is also multithreaded so that it can take advantage of multiple processors.

PROC QUANTSELECT is very efficient and can handle hundreds of variables and thousands of observations.

New STDRAE Procedure

The STDRAE procedure computes direct and indirect standardized rates and risks for study populations. With direct standardization, you compute the weighted average of stratum-specific estimates in the study population, using weights such as population-time from a standard or reference population. With indirect standardization, you compute the weighted average of stratum-specific estimates in the reference population by using weights from the study population. The procedure provides summary statistics such as rate and risk estimates (and their confidence limits) for each stratum, as well as graphs.

Highlights of Enhancements

The following are highlights of other enhancements in SAS/STAT 12.1:

- The MCMC procedure now models missing values by default. The RANDOM statement supports multilevel hierarchy to an arbitrary depth. The procedure also implements faster and more efficient sampling algorithms.
- The PHREG procedure supports Bayesian frailty models.
- The FMM procedure for finite mixture models is now production and adds several truncated distributions.
- The LIFEREG and PROBIT procedures include additional postprocessing statements. They now support the TEST, LSMEANS, LSMESTIMATE, ESTIMATE, SLICE, and EFFECTPLOT statements.
- The FREQ procedure produces mosaic plots.
- The SURVEYSELECT procedure provides Poisson sampling.
- The SURVEYMEANS procedure now performs poststratification estimation.
- The GLM, MIXED, GLIMMIX, and ORTHOREG procedures support the REF= option in the CLASS statement.

More information about the changes and enhancements follows. Details can be found in the documentation for the individual procedures in the *SAS/STAT 12.1 User's Guide*.

Highlights of Enhancements in SAS/STAT 9.3

Some users might be unfamiliar with updates made in SAS/STAT 9.3. The following are some of the major enhancements that were introduced in SAS/STAT 9.3:

- The experimental FMM procedure fits statistical models to data where the distribution of the response is a finite mixture of univariate distributions. These models are useful for applications such as estimating multimodal or heavy-tailed densities, fitting zero-inflated or hurdle models to count data with excess zeros, modeling overdispersed data, and fitting regression models with complex error distributions.
- The EFFECT statement became production. This statement is available in the HPMIXED, GLIMMIX, GLMSELECT, LOGISTIC, ORTHOREG, PHREG, PLS, QUANTREG, ROBUSTREG, SURVEYLOGISTIC, and SURVEYREG procedures.
- The MCMC procedure added a RANDOM statement, which simplifies the specification of hierarchical random-effects models and significantly reduces simulation time while improving convergence.
- The METHOD=FIML option in the CALIS procedure became production. This option specifies the full information maximum likelihood method.
- The SURVEYPHREG procedure became production and handles time-dependent covariates.
- The HPMIXED procedure added a REPEATED statement and additional covariance structures.
- The MI procedure added fully conditional specification methods for multiple imputation.
- The NLIN procedure was updated with features for diagnosing the nonlinear model fit.

New Macros

Bayesian Analysis Postprocessing Macros

Postprocessing macros are now available in the SAS autocall library. These macros duplicate the corresponding summary and diagnostic capabilities provided by the MCMC procedure and can be used with any SAS data set that contains posterior samples. These macros are documented with the MCMC procedure.

%CIF Macro

The %CIF macro implements nonparametric methods for estimating cumulative incidence functions with competing risks data. The macro can also be used to test the hypothesis that cumulative incidence functions are identical across groups. The %CIF macro is available from the SAS autocall library.

Enhancements

CALIS Procedure

The CALIS procedure now provides the following features:

- Residual analysis at the case level or observation level is provided when you input raw data.
- Robust estimation is now available. You can request either direct robust estimation based on residual weighting or two-stage robust estimation based on analyzing the robust mean and covariance matrices.
- The BASEFUNC= and BASEFIT= options enable you to input the model fit information of the customized baseline model of your choice. With this input information, PROC CALIS computes various fit indices (mainly the incremental fit indices) based on your customized model fit rather than the fit of the default uncorrelatedness model.

EFFECTPLOT Statement

The CLUSTER option in the EFFECTPLOT statement modifies the INTERACTION plot-type by displaying the levels of the SLICEBY= effect in a side-by-side fashion. The CONNECT option in the EFFECTPLOT statement modifies the BOX and INTERACTION plot-types by connecting the predicted values with a line.

FMM Procedure

The FMM procedure is now production and adds the truncated exponential, truncated lognormal, truncated negative binomial, and truncated normal distributions.

FREQ Procedure

The Agresti-Caffo and Miettinen-Nurminen confidence limits for the risk (proportion) difference are now available, and any confidence limit type can now be displayed in the risk difference plot. You can also request a continuity correction for the Wilson confidence limits for a binomial proportion.

The PLCORR option in the TEST statement produces Wald and likelihood ratio tests for the polychoric correlation coefficient.

The new DF= chi-square option specifies or adjusts the degrees of freedom for chi-square tests. The TESTF= chi-square option now permits you to provide null frequencies for a one-way chi-square test by using a secondary input data set. Similarly, the TESTP= chi-square option now permits you to provide null proportions by using a secondary input data set.

The LRCHISQ chi-square option in the TABLES statement produces a likelihood ratio chi-square test for one-way tables. This test can be based on a null hypothesis of equal proportions, specified proportions, or specified frequencies. The LRCHISQ option in the EXACT statement produces an exact likelihood ratio chi-square test for one-way tables.

The EXACT statement can now produce Barnard's exact unconditional test for the risk difference as well as an exact likelihood ratio goodness-of-fit test.

The MAXLEVELS=*n* option in the TABLES statement specifies the maximum number of variable levels to display in one-way frequency tables; it also applies to one-way frequency plots.

The CROSSLIST(STDRES) option in the TABLES statement displays standardized residuals in the CROSSLIST table for two-way crosstabulation.

Mosaic plots are now produced with the PLOTS=MOSAICPLOT option. The GROUPBY= plot option specifies the primary grouping for two-way frequency plots. The new TWOWAY=CLUSTER plot option provides a cluster layout for two-way frequency plots that are displayed as bar charts.

GLIMMIX Procedure

The new DDFM=KENWARDROGER2 option applies the (prediction) standard error and degrees-of-freedom correction that are detailed by Kenward and Roger (2009). This correction further reduces the precision estimator bias for the fixed and random effects under nonlinear covariance structures.

The odds ratio, relative risk, and kappa plots now display the common (overall) statistic in addition to the statistics for each two-way table (stratum).

LIFEREG Procedure

The LIFEREG procedure now supports numerous postprocessing statements including the ESTIMATE, EFFECTPLOT, LSMEANS, LSMESTIMATE, SLICE, STORE, and TEST statements.

LIFETEST Procedure

The LIFETEST procedure now supports a WEIGHT statement. For survival plot enhancements, the MAXLEN= option of the ATRISK option specifies a maximum number of characters that can be used in displaying stratum labels, and the OUTSIDE option of the ATRISK option specifies that the at-risk table be drawn outside the plot area. If you assign a label to a strata variable, that label is used in all tables and graphs.

LOESS Procedure

The LOESS procedure now supports an OUTPUT statement.

LOGISTIC Procedure

PROC LOGISTIC provides partial proportional odds logistic regression with the UNEQUALSLOPES option in the MODEL statement. The ESTIMATE, LSMEANS, LSMESTIMATE, SLICE, and STORE statements can now be used for a stratified analysis. The PCORR option in the MODEL statement computes the partial correlation statistic for each model parameter (excluding the intercept).

The ID statement specifies variables in the DATA= data set that are used for labeling ROC curves and influence diagnostic plots.

The NLOPTIONS statement controls the optimization process for conditional analyses (specified with a STRATA statement) and for constrained optimization (specified with the UNEQUALSLOPES option in the MODEL statement).

The EFFECTPLOT statement and the PLOTS=EFFECT option have two new options for displaying plots with a CLASS effect on the X axis. The CLUSTER option displays the levels of the SLICEBY= effect in a side-by-side fashion. The CONNECT option connects the predicted values with a line.

MCMC Procedure

The MCMC procedure provides the following new capabilities:

- The MODEL statement augments missing values in the response variable by default. PROC MCMC treats missing values as unknown parameters and incorporates the sampling of the missing data as part of the Markov chain.
- The RANDOM statement supports multilevel hierarchical modeling to an arbitrary depth; a random effect can appear in the distributional hierarchy of other random effects.
- More distributions, such as multivariate normal distribution with autoregressive structure, Poisson distribution, and general distribution (for the construction of nonstandard distributions), are made available for the RANDOM statement.
- Direct sampling and more conjugate sampling algorithms are available for all parameters in the model (including model parameters, random-effects parameters, and missing data variables) when appropriate.
- A slice sampler is an alternative sampling algorithm for both the model parameters and random-effects parameters.

MULTTEST Procedure

The ID statement names one or more variables for identifying observations in the output and in the plots. In addition, the MANHATTAN plot option generates a Manhattan plot.

NLIN Procedure

The PROFILE statement selects parameters for profiling for the assessment of their nonlinear characteristics. It can also gauge the influence of each observation on the selected parameters.

NPAR1WAY Procedure

The following options are added to the PROC NPAR1WAY statement:

- The DSCF option requests the Dwass, Steel, Critchlow-Fligner multiple comparison procedure, which is based on pairwise two-sample rankings.
 - The FP option requests the Fligner-Policello test for two-sample data.
 - The ADJUST option adjusts for location differences among classes before tests for scale differences are performed.
 - The REFCLASS= option of the HL option specifies which of the two CLASS variable levels (samples) to use as the reference class X for the location shift, Y-X.
-

ODS Graphics

You can use the new ODS GRAPHICS statement option BYLINE=FOOTNOTE or BYLINE=TITLE to display BY-group lines in a footnote or title in graphs.

PHREG Procedure

The DIST= option in the RANDOM statement enables you to specify either a gamma or lognormal distribution for the shared frailty. Bayesian frailty models are now supported. The DISPERSIONPRIOR= option in the BAYES statement specifies the prior distribution of the dispersion parameter.

Fleming-Harrington estimates can be requested with the METHOD=FM option in the BASELINE statement and in the OUTPUT statement. The DIRADJ option in the BASELINE statement specifies direct adjusted survival curves.

POWER Procedure

You can specify the multiple correlation between the tested predictor and the covariates with the CORR= option in the LOGISTIC statement.

PROBIT Procedure

The PROBIT procedure now supports numerous postprocessing statements including the ESTIMATE, EFFECTPLOT, LSMEANS, LSMESTIMATE, SLICE, STORE, and TEST statements.

QUANTREG Procedure

The QUANTREG procedure now supports the ESTIMATE statement. The EFFECT statement now supports the effect-types COLLECTION, LAG, MULTIMEMBER, and POLYNOMIAL in addition to SPLINE.

REG Procedure

The REG procedure creates heat maps for residual and fit plots when the MAXPOINTS= threshold is exceeded.

ROBUSTREG Procedure

The STDI= option in the OUTPUT statement specifies a variable to contain the estimates of the standard errors of the individual predicted values.

SEQDESIGN Procedure

The MODEL=INPUTNEVENTS option in the SAMPLESIZE statement specifies the number of events from a fixed-sample study of survival data. There are two new INPUTEVENTS options for the sample size computation: ACCRUAL= and LOSS=. The ACCRUAL= option specifies the method for individual accrual. The LOSS= option specifies the individual loss to follow-up in the sample size computation.

The STOP=BOTH option in the DESIGN statement specifies the condition of early stopping for the design. The new BETABOUNDARY=BINDING suboption computes the Type I error probability with the acceptance boundary, and the new BETABOUNDARY=NONBINDING suboption computes the Type I error probability without the acceptance boundary.

SEQTEST Procedure

The BETABOUNDARY option in the PROC SEQTEST statement specifies whether the β boundary is used in the computation of the Type I error level α . The BETABOUNDARY=BINDING option computes the Type I error probability with the β (acceptance) boundary, and the BETABOUNDARY=NONBINDING suboption computes the Type I error probability without the β boundary.

SURVEYFREQ Procedure

The SURVEYFREQ procedure now produces mosaic plots for crosstabulation tables. The GROUPBY= plot option specifies the primary grouping for two-way weighted frequency plots.

SURVEYMEANS Procedure

You can now estimate geometric means for finite populations with the GEOMEAN keyword in the PROC SURVEYMEANS statement. The new POSTSTRATA statement provides poststratification analysis.

SURVEYPHREG Procedure

The SERATIO and VARRATIO options in the MODEL statement compute the ratio of two standard errors for the regression coefficients and the ratio of two variances for the regression coefficients, respectively.

SURVEYREG Procedure

The STB option in the MODEL statement produces standardized regression coefficients.

SURVEYSELECT Procedure

The SURVEYSELECT procedure now provides Bernoulli and Poisson sampling.

What's Changed

What follows are changes in software behavior from SAS/STAT 9.3 to SAS/STAT 12.1.

LIFETEST Procedure

If you assign a label to a strata variable, the procedure now uses the label instead of the variable name in all tables and graphs.

FREQ Procedure

The appearance of the default bar chart for two-way frequency plots has changed. The row level labels have been moved outside the plot so that the row grouping appears less dominant.

For two-way dot plots (TYPE=DOTPLOT) in nonstacked layouts, the default positions of the row and column variables are reversed to group graph cells by the column variable. You can specify GROUPBY=ROW to group graph cells by the row variable.

MCMC Procedure

Random-effects parameter names are constructed using the formatted values rather than the unformatted values.

If the `MISSING=` option is not specified, PROC MCMC samples all missing values (including partial missing in some cases) by default. Observations for which the procedure failed to identify proper sampling algorithms are discarded prior to the simulation. If the `MISSING=` option is explicitly specified (`AC` or `CC`), the option is honored.

PROC MCMC avoids performing an optimization prior to the start of the simulation if the only sampling algorithms used in the program are conjugate or direct.

PROC MCMC now permits a model specification that has only `RANDOM` and `MODEL` statements; `PRIOR` and `PARMS` statements are no longer required in that case.

MULTTEST Procedure

By default, the `AFDR` and `PFDR` are constrained to be greater than or equal to the raw p -value. The `UNRESTRICT` option of the PROC MULTTEST statement's `AFDR` and `PFDR` options estimates the `AFDR` and `PFDR` as defined in Benjamini and Hochberg (2000), which allows the adjustment to reduce the raw p -value.

SURVEYSELECT Procedure

PROC SURVEYSELECT now uses the Mersenne-Twister random number generator by default. In previous releases, PROC SURVEYSELECT used the `RANUNI` random number generator. To reproduce samples that PROC SURVEYSELECT selected in releases prior to SAS/STAT 12.1, you can use the `RANUNI` option with the `SEED=` option (for the same input data set and selection parameters).

GLIMMIX, GLM, HPMIXED, and MIXED Procedures

PCs running Windows 64-bit operating systems can now allocate more than 2Gb of memory to fit a model. This change affects the `GLIMMIX`, `GLM`, `HPMIXED`, and `MIXED` procedures.

References

- Benjamini, Y. and Hochberg, Y. (2000), "On the Adaptive Control of the False Discovery Rate in Multiple Testing with Independent Statistics," *Journal of Educational and Behavioral Statistics*, 25, 60–83.
- Kenward, M. G. and Roger, J. H. (2009), "An Improved Approximation to the Precision of Fixed Effects from Restricted Maximum Likelihood," *Computational Statistics and Data Analysis*, 53, 2583–2595.
- Peng, L. and Huang, Y. (2008), "Survival Analysis with Quantile Regression Models," *Journal of the American Statistical Association*, 103, 637–649.
- Portnoy, S. (2003), "Censored Regression Quantiles," *Journal of the American Statistical Association*, 98, 1001–1012.

Your Turn

We welcome your feedback.

- If you have comments about this book, please send them to **yourturn@sas.com**. Include the full title and page numbers (if applicable).
- If you have comments about the software, please send them to **suggest@sas.com**.

SAS® Publishing Delivers!

Whether you are new to the work force or an experienced professional, you need to distinguish yourself in this rapidly changing and competitive job market. SAS® Publishing provides you with a wide range of resources to help you set yourself apart. Visit us online at support.sas.com/bookstore.

SAS® Press

Need to learn the basics? Struggling with a programming problem? You'll find the expert answers that you need in example-rich books from SAS Press. Written by experienced SAS professionals from around the world, SAS Press books deliver real-world insights on a broad range of topics for all skill levels.

support.sas.com/saspress

SAS® Documentation

To successfully implement applications using SAS software, companies in every industry and on every continent all turn to the one source for accurate, timely, and reliable information: SAS documentation. We currently produce the following types of reference documentation to improve your work experience:

- Online help that is built into the software.
- Tutorials that are integrated into the product.
- Reference documentation delivered in HTML and PDF – free on the Web.
- Hard-copy books.

support.sas.com/publishing

SAS® Publishing News

Subscribe to SAS Publishing News to receive up-to-date information about all new SAS titles, author podcasts, and new Web site features via e-mail. Complete instructions on how to subscribe, as well as access to past issues, are available at our Web site.

support.sas.com/spn



sas

**THE
POWER
TO KNOW®**

