SAS/STAT® 13.2 User's Guide
What's New in SAS/STAT 13.2

This document is an individual chapter from *SAS/STAT® 13.2 User's Guide*.

The correct bibliographic citation for the complete manual is as follows: SAS Institute Inc. 2014. *SAS/STAT® 13.2 User's Guide*. Cary, NC: SAS Institute Inc.

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

August 2014

Chapter 1
What's New in SAS/STAT 13.2

Contents

Overview

SAS/STAT 13.2 includes three new procedures, two new high-performance procedures, and many enhancements.

New Procedures

GEE Procedure (Experimental)

The experimental GEE procedure fits generalized linear models for longitudinal data by using the generalized estimating equations (GEE) estimation method of Liang and Zeger (1986). The GEE method fits a marginal model to longitudinal data and is commonly used to analyze longitudinal data when the population-average effect is of interest. The GEE procedure's syntax is similar to that of the GENMOD procedure. You specify the response variable and the explanatory variables in a MODEL statement, and you specify the correlation structure of multivariate responses in a REPEATED statement.

Because missing data are common in longitudinal studies and can lead to biased parameter estimates when missing responses depend on previous responses, PROC GEE also implements observation-specific and subject-specific weighted estimating equations. Both weighted estimators provide unbiased and consistent estimates when data are missing at random.

ICPHREG Procedure

The ICPHREG procedure fits proportional hazards regression models to interval-censored data. You can fit models that have a variety of configurations with respect to the baseline hazard function, including the piecewise constant model and the cubic spline model. PROC ICPHREG maximizes the full likelihood instead of the Cox partial likelihood to estimate the regression coefficients. Standard errors of the estimates are obtained by inverting the observed information matrix that is derived from the full likelihood.

The ICPHREG procedure compares most closely to the PHREG and LIFEREG procedures. All three procedures can fit proportional hazards models. The distinction lies in the types of data that they are designed to handle and whether the baseline hazard function is parameterized. PROC PHREG deals exclusively with right-censored data, and it adopts a semiparametric approach by leaving the baseline hazard function unspecified. Both PROC LIFEREG and PROC ICPHREG can handle interval-censored data, which is a generalization of right-censored data. PROC LIFEREG focuses on parametric analysis by using accelerated failure time models, and it fits proportional hazards models only by assuming a Weibull baseline hazard function. PROC ICPHREG offers several options to parameterize the baseline hazard function. PROC ICPHREG generalizes PROC LIFEREG to fit more flexible parametric models and also generalizes PROC PHREG to handle interval-censored data.

SPP Procedure

The SPP procedure analyzes spatial point patterns. The broad goal of spatial point pattern analysis is to describe the occurrence of events (observations) that compose the pattern. The event locations are a discrete realization of a random spatial process. Therefore, the analysis goal is to investigate and characterize the original spatial process that generated the events in the spatial point pattern.

The SPP procedure enables you to specify the study area as a window, or you can rely on the input data coordinates to automatically compute a suitable study area by using the Ripley-Rasson window estimator. You can perform exploratory analysis of spatial point patterns by using the F, G, J, K, L, and PCF distance functions, which compare the empirical function distributions to the theoretical homogeneous Poisson process. PROC SPP enables you to perform nonparametric intensity estimation by using different types of kernels, and it supports adaptive kernel estimation. In addition, PROC SPP enables you to fit parametric inhomogeneous Poisson process models to perform model validation by using goodness-of-fit testing and a variety of residual diagnostics.

Highlights of Enhancements

Following are highlights of the enhancements in SAS/STAT 13.2:

- The FACTOR procedure generates path diagrams.
- The FMM procedure fits multinomial models.
- The IRT procedure generates polychoric correlation matrices, item characteristic curves, and test information curve plots.
- The MCMC procedure supports a categorical distribution in the MODEL, RANDOM, and PRIOR statements.
- The NLMIXED procedure enables you to specify more than one RANDOM statement in order to fit hierarchical nonlinear mixed models.
- The SEQDESIGN procedure enables you to create a ceiling-adjusted design that corresponds to integer-valued sample sizes at the stages for nonsurvival data.
- The LOGISTIC procedure enables you to add or relax constraints on parameters in nominal response and partial proportional odds models.
- The FREQ procedure now provides score confidence limits for the odds ratio and the relative risk.
- The GLMSELECT procedure enables you to apply safe screening and sure independence screening methods to reduce a large number of regressors to a smaller subset from which model selection is performed.

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

Highlights of Enhancements in SAS/STAT 13.1

Some users might be unfamiliar with updates made in the previous releases. SAS/STAT 13.1 introduced the BCHOICE, ICLIFETEST, and IRT procedures. Following are highlights of the other enhancements in SAS/STAT 13.1:

- The MI procedure added the MNAR statement to facilitate sensitivity analysis.
- The Tweedie distribution is supported by the GENMOD procedure.
- The competing-risks model of Fine and Gray (1999) is available in the PHREG procedure.
- The NLIN procedure enables you to generate both bootstrap estimates of confidence intervals for the parameters and bootstrap estimates of the covariance matrix and correlation matrix of the parameter estimates.
- The MCMC procedure is multithreaded.

- Path diagrams are available in the CALIS procedure.
- You can compute power for PROC GLM–type MANOVA and repeated measurements in the GLMPOWER procedure.
- The SURVEYMEANS procedure produces domain quantile estimates.

Enhancements

BCHOICE Procedure

The BCHOICE procedure is production in SAS/STAT 13.2. The Gamerman sampling algorithm is improved and now runs up to 30% faster for some models.

CALIS Procedure

All output tables for parameter estimates now show the *p*-values of the estimates. The new CI option in the PROC CALIS statement adds confidence intervals in these tables. The new ALPHA=α option enables you to request that interval estimation of parameters be performed at the (1–α)100% confidence level instead of the default 95% confidence level.

FACTOR Procedure

The new PATHDIAGRAM statement enables you to generate and fine-tune path diagrams. You can also use the new PLOTS=PATHDIAGRAM option in the PROC FACTOR statement to request the default path diagrams, which show the links between factors and variables, the factor correlations, and the error variances in the model.

FMM Procedure

The DIST=MULTINOMIAL option in the MODEL statement enables you to specify the multinomial distribution to model a discrete response that has multiple levels. The DIST=MULTINOMIALCLUSTER option enables you to fit a multinomial cluster model to address overdispersion in multinomial models.

FREQ Procedure

The following types of binomial confidence limits are now available: Blaker, exact mid-*p*, likelihood ratio, and logit. (You can request binomial confidence limits in the BINOMIAL(CL=) option.)

The new RELRISK(CL=SCORE) option provides score confidence limits for the relative risk. Score confidence limits can be displayed in the relative risk plot.

The RISKDIFF(COMMON) option provides stratified Newcombe confidence limits for the common risk (proportion) difference. Stratified Newcombe confidence limits can be displayed in the risk difference plot.

GLMPOWER Procedure

PROC GLMPOWER now supports multiple factors in the REPEATED statement.

GLMSELECT Procedure

The SCREEN= option in the MODEL statement enables you to apply either a safe screening method or a sure independence screening method to reduce a large number of regressors to a smaller subset from which model selection is performed.

IRT Procedure

The IRT procedure is at production status in this release.

The POLYCHORIC option in the PROC IRT statement displays the polychoric correlation matrix.

The PLOTS=POLYCHORIC option in the PROC IRT statement displays a heat map for the polychoric correlation matrix.

The PLOTS=IIC option in the PROC IRT statement displays item characteristic curves.

The PLOTS=TIC in the PROC IRT statement displays a test information curve plot.

LOGISTIC Procedure

The EQUALSLOPES and UNEQUALSLOPES options in the MODEL statement enable you to add or relax constraints on parameters in nominal response models and partial proportional odds models.

The INCLUDE=EQUALSLOPES specification in the MODEL statement facilitates model selection by enabling you to include all the equal slope effects in every model and perform the selection process on the unequal slope effects.

The START=EQUALSLOPES specification in the MODEL statement facilitates model selection by enabling you to begin the model selection process with all the equal slope effects in the model.

MCMC Procedure

The PRIOR, RANDOM, and MODEL statements now support a categorical distribution.

The RANDOM statement now supports a uniform prior distribution.

All conjugate sampling algorithms are now multithreaded.

NLMIXED Procedure

PROC NLMIXED now enables you to specify more than one RANDOM statement in order to fit hierarchical nonlinear mixed models.

The OUTR= option in the PROC NLMIXED statement requests an output data set that contains all the random-effects variable estimates from all RANDOM statements.

The default degrees of freedom for the fixed parameters is calculated for random-effects models as the total number of subjects minus the number of random-effects variables that are specified in a RANDOM statement. If this calculation produces a nonpositive number, the default degrees of freedom is replaced by the total number of observations.

PHREG Procedure

PROC PHREG now supports the following features for the competing-risks analysis of Fine and Gray (1999):

- survival times with left truncation or start/stop syntax
- a WEIGHT statement

PLS Procedure

You can now display correlation loading plots for all pairs of factors, not just the first two.

POWER Procedure

The TEST=FM option in the TWOSAMPLEFREQ statement specifies the score test of Farrington and Manning (1990).

REG Procedure

The SRT option in the MODEL statement displays a column in the "Output Statistics" table that shows the magnitude of studentized residuals by using a series of asterisks.

SEQDESIGN Procedure

The CEILADJDESIGN=INCLUDE suboption of the MODEL= option in the SAMPLESIZE statement creates a ceiling-adjusted design that corresponds to integer-valued sample sizes at the stages for nonsurvival data, and to integer-valued times or sample sizes at the stages for survival data.

SEQTEST Procedure

The INFOVAR= suboption of the DATA= option in the PROC SEQTEST statement enables you to specify the information variable in the DATA= data set.

The INFOVAR= suboption of the PARMS= option in the PROC SEQTEST statement enables you to specify the information variable in the PARMS= data set.

SURVEYFREQ Procedure

The COV option in the TABLES statement displays the covariance matrix of the total frequency estimates.

The COVP option in the TABLES statement displays the covariance matrix of the proportion estimates.

SURVEYLOGISTIC Procedure

The default hypothesis tests are changed from chi-square tests to F tests.

The default confidence limits for parameter estimates and odds ratio estimates are changed from Wald confidence limits to t confidence limits.

A new DF= option in the MODEL statement specifies the denominator degrees of freedom for F statistics in hypothesis testing, and the degrees of freedom in t tests for parameter estimates, odds ratio estimates, and their t percentiles for confidence limits.

The score test for the global null hypothesis now takes survey design information into account.

Enhancements to the High-Performance Procedures

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 in *SAS/STAT User's Guide: High-Performance Procedures*.

New Procedures

HPPLS Procedure

The new HPPLS procedure fits models by using any of several linear predictive methods, including partial least squares (PLS), to optimally address one or both of these two goals: explaining response variation and explaining predictor variation.

HPQUANTSELECT Procedure

The new HPQUANTSELECT procedure performs high-performance quantile regression analysis. PROC HPQUANTSELECT not only fits quantile regression models but also offers extensive capabilities for quantile regression model selection, and it supports statistical inferences with or without the assumption of independently and identically distributed (iid) errors.

Procedure Enhancements

HPGENSELECT Procedure

The PARTITION statement specifies how observations in the input data set are to be logically partitioned into disjoint subsets for model training, validation, and testing. Models are fit and selected based on the training data. After you fit a model, you can use the validation and test sets to assess how the selected model generalizes on data that played no role in selecting the model.

HPLOGISTIC Procedure

- The PARTITION statement divides the observations in the input data set into disjoint subsets for model training, validation, and testing. Various fit statistics are displayed in the new "Partition Fit Statistics" table.
- The new CUTPOINT= option in the MODEL statement enables you to control the classification of events and nonevents.
- The new CHOOSE=VALIDATE and STOP=VALIDATE options in the SELECTION statement use the validation data set during the selection process.
- The AIC, BIC, and AICC criteria are added to the SELECT= option in the SELECTION statement.
- The INEST= option in the PROC statement enables you to input your own starting values for the optimization. The OUTEST option adds a column that contains the parameter names to the "Parameter Estimates" ODS OUTPUT data.
- The CTABLE option in the MODEL statement creates data for receiver operating characteristic (ROC) curves. The PRIOR= option in the MODEL statement specifies population prevalences that are used to adjust statistics displayed by the CTABLE option and by the PARTITION statement.
- The POST keyword in the OUTPUT statement outputs the posterior probabilities that are specified in the PRIOR= option, and the ROLE keyword outputs the partition to which the observation is assigned.

What's Changed

The following sections describe changes in software behavior from SAS/STAT 13.1 to SAS/STAT 13.2.

The NLMIXED Procedure

The default degrees of freedom for the fixed parameters is calculated for random-effects models as the total number of subjects minus the number of random-effects variables that are specified in a RANDOM statement. If this calculation produces a nonpositive number, the default degrees of freedom is replaced by the total number of observations.

The SURVEYLOGISTIC Procedure

The default hypothesis tests are changed from chi-square tests to F tests.

The default confidence limits for parameter estimates and odds ratio estimates are changed from Wald confidence limits to t confidence limits.

References

Farrington, C. P. and Manning, G. (1990), "Test Statistics and Sample Size Formulae for Comparative Binomial Trials with Null Hypothesis of Non-zero Risk Difference or Non-unity Relative Risk," *Statistics in Medicine*, 9, 1447–1454.

Fine, J. P. and Gray, R. J. (1999), "A Proportional Hazards Model for the Subdistribution of a Competing Risk," *Journal of the American Statistical Association*, 94, 496–509.

Liang, K.-Y. and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13–22.