# Enhancements in SAS/STAT® 9.2 Software

Overview

The release of SAS® 9.2 brings exciting new enhancements to SAS/STAT software. The 9.1 experimental Web-downloadable procedures become production, and five new experimental procedures are introduced. ODS Statistical Graphics also become production, and numerous new graphs appear in many statistical procedures. In addition, SAS/STAT procedures have been updated with more than 200 new features. The following are the highlights of this new release.

Bayesian Software

The BAYES statement in three procedures—GENMOD, LIFEREG, and PHREG—enables you to perform Bayesian analyses for generalized linear models, parametric survival models, and proportional hazards models. Posterior samples are obtained via Gibbs sampling. Convergence diagnostics such as the Gelman-Rubin, Geweke, Heidelberger-Welch, and Raftery-Lewis tests are produced as well as diagnostic plots. A number of priors can be specified, such as the normal, gamma, and uniform prior, and the GENMOD procedure also provides Jeffreys' prior. You can output the posterior samples to a SAS data set for use in subsequent analyses.
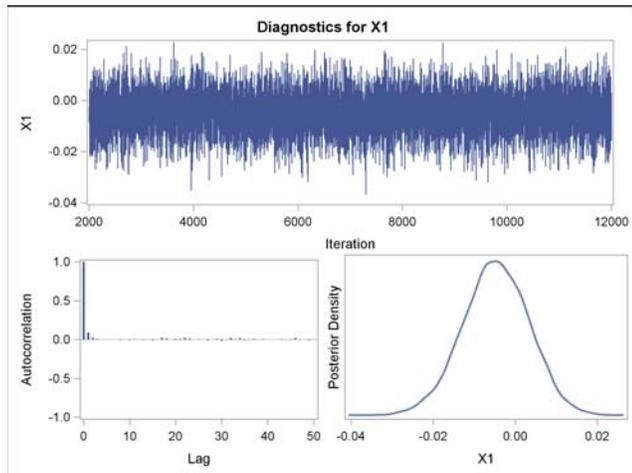

The experimental MCMC procedure is a flexible simulation-based procedure that is suitable for fitting a wide range of Bayesian models. You specify a likelihood function for the data and a prior distribution for the parameters by using SAS programming statements similar to those employed in the NLMIXED procedure. You can specify hyperprior distributions if you are fitting hierarchical models.

PROC MCMC then obtains samples from the corresponding posterior distributions, produces summary and diagnostic statistics, and saves the posterior samples in an output data set. The parameters can enter the model linearly or in any nonlinear functional form. By default, PROC MCMC uses an adaptive blocked random-walk Metropolis algorithm with a normal proposal distribution.

ODS Statistical Graphics

In SAS 9.2, forty SAS/STAT procedures support ODS graphics, as do the CORR, FREQ, and UNIVARIATE procedures in Base SAS. This functionality has been extended with the addition of new graph types, ODS styles designed for statistical work, and a point-and-click editor for enhancing titles, labels, and other graph features.
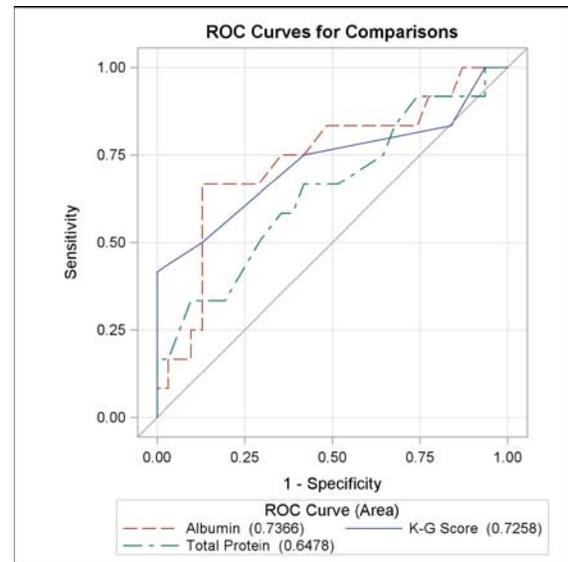
You can also modify graphs by changing their underlying templates, which are supplied by SAS and are written in the Graph Template Language (GTL). The LISTING destination is now supported by ODS Graphics. Note that a new family of SAS/GRAPH® procedures (SGPLOT, SGPANEL, SGSCATTER) uses ODS Graphics to create standalone plots, such as scatterplots overlaid with smoothers, which are particularly useful for exploratory data analysis. The new SGRENDER procedure in SAS/GRAPH provides a way to create customized displays by writing your own templates with the GTL.

Note that a SAS/GRAPH license is now required to use ODS Graphics in SAS/STAT software.

Quantile Regression

Quantile regression extends the regression model to conditional quantiles of the response variable, such as the 90th percentile. Quantile regression is particularly useful when the rate of change in the conditional quantile, expressed by the regression coefficients, depends on the quantile. The main advantage of quantile regression over least squares regression is its flexibility for modeling data with heterogeneous conditional distributions. Data of this type occur in many fields, including biomedicine, econometrics, and ecology.

The QUANTREG procedure is production in SAS 9.2. It implements the simplex, interior point, and smoothing algorithms for estimation. In addition, it:

- Provides sparsity, rank, and resampling methods for computing confidence intervals for the regression quantile parameter
- Provides an asymptotic and bootstrap method to compute the covariance and correlation matrices of the estimated parameters
- Provides the Wald test and a likelihood ratio test for tests of the regression parameters
- Uses robust multivariate location and scale estimates for leverage point detection

Model Selection

The GLMSELECT procedure performs effect selection in the framework of general linear models. A variety of model selection methods are available, including the LASSO method of Tibshirani (1996) and the related LAR method of Efron et al. (2004). The procedure offers extensive capabilities for customizing the selection with a wide variety of selection and stopping criteria, from traditional and computationally efficient significance-level-based criteria to more computationally intensive validation-based criteria. The procedure also provides graphical summaries of the selection search.

PROC GLMSELECT focuses on the standard independently and identically distributed general linear model for univariate responses. It offers great flexibility in the model selection algorithm. PROC GLMSELECT provides results (displayed tables, output data sets, and macro variables) that make it convenient to analyze the selected model in a subsequent procedure such as REG or GLM. PROC GLMSELECT provides:

- Forward, backward, and stepwise selection
- Least angle regression (LAR) and LASSO
- Selection based on information criteria and predictive performance
- Models with crossed and nested effects

- Selection from very large number of effects (tens of thousands)
- Internal partitioning of data into training, validation, and testing roles

Generalized Linear Mixed Models

The GLIMMIX procedure fits statistical models to data with correlations or nonconstant variability and where the response is not necessarily normally distributed. These generalized linear mixed models (GLMM), like linear mixed models, assume normal (Gaussian) random effects. Conditional on these random effects, data can have any distribution in the exponential family. Originally available as a Web-download for Windows and several UNIX platforms for SAS 9.1, PROC GLIMMIX has been updated for SAS 9.2.
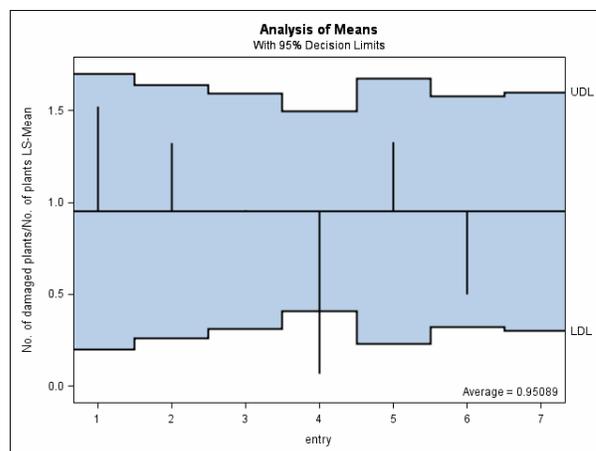
The GLIMMIX procedure now provides Laplace and adaptive quadrature estimation methods, and, with them, a likelihood-based empirical estimator. In addition, a new bias-corrected estimator is available. The COVTEST statement enables likelihood-based inference about the covariance parameters.

A number of additional covariance structures have been added, including heterogeneous AR(1), heterogeneous compound symmetry, linear structures, heterogeneous Toeplitz, penalized B-spline, spatial anisotropic, and the Matérn covariance structure. Step-down multiplicity adjustments are now supported for all ADJUST= methods in the LSMEANS, ESTIMATE, and LSMESTIMATE statements, except for ADJUST=NELSON in the LSMEANS statement. New graphs include boxplots of data and/or residuals with respect to classification effects as well as plots of odds ratios and their confidence limits. The diffogram, meanplot, anomplot, and controlplot have been enhanced.

Mixed Models Analysis

The experimental HPMIXED procedure uses a number of specialized high-performance techniques to fit linear mixed models with variance component structure. The HPMIXED procedure is specifically designed to cope with estimation problems that involve a large number of fixed effects, a large number of random effects, or a large number of observations. While the HPMIXED procedure fits only a subset of the models fit by the MIXED procedure and it does not provide the breadth of confirmatory inference that is available with the MIXED procedure, it can have considerably better performance in terms of memory requirements and computational speed.

Structural Equations Modeling

The TCALIS procedure is experimental in SAS 9.2. It enables you to perform the same analyses that you can do with PROC CALIS. In addition, PROC TCALIS provides functionality such as multiple-group analysis, enhanced mean structure analysis, path-like model specification, support of LISREL-type models, customizable effect analysis, general parametric function testing, customizable Lagrange multiplier tests, and so on. Currently, you can specify COSAN models only in PROC CALIS, but not in PROC TCALIS. In future releases, PROC TCALIS will become PROC CALIS.

Group Sequential Design and Analysis

A group sequential trial provides for interim analyses before the formal completion of a clinical trial while maintaining the specified overall Type I and Type II error probability levels. A group sequential trial is most useful in situations where it is important to monitor the trial to prevent unnecessary exposure of patients to an unsafe new drug, or alternatively to a placebo treatment if the new drug shows significant improvement. If a group sequential trial stops early, then usually it requires fewer participants than a corresponding fixed-sample trial.

The experimental SEQDESIGN procedure designs interim analyses for clinical trials. PROC SEQDESIGN computes the boundary values and required sample sizes for the trial. The boundary values are derived in such a way that the overall Type I and Type II error probability levels are maintained at the levels specified in the design. Available methods include fixed boundary shape methods, which include unified family methods such as the O'Brian-Fleming method, Whitehead methods, and error spending methods. In addition to the boundary values, the SEQDESIGN procedure

computes quantities such as average sample sizes and stopping probabilities.
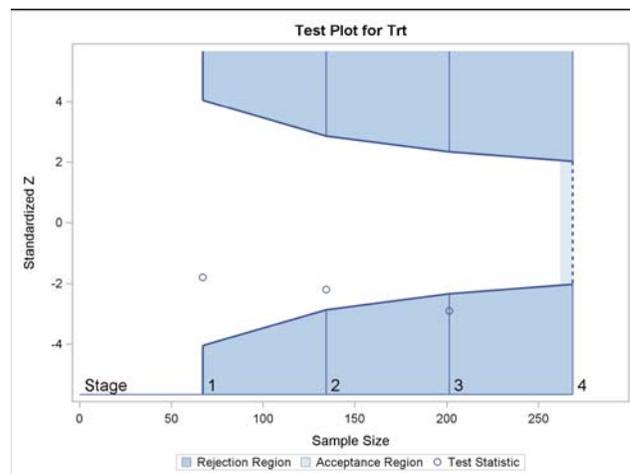
The experimental SEQTEST procedure is used in conjunction with the SEQDESIGN procedure to carry out interim analyses for clinical trials. At each stage, you analyze the data with a statistical procedure and compute test statistics. You then use the SEQTEST procedure to compare the test statistic with the corresponding boundary values computed by the SEQDESIGN procedure.

PSS Application

The PSS application has been converted to a Java client application and no longer requires a Web server. It now offers a relative risk parameterization for the two proportions analysis. New analyses covered include equivalence and noninferiority for proportions, confidence interval for one proportion, Wilcoxon-Mann-Whitney for two distributions, logistic regression, and GLM contrasts for interactions.

Survival Analysis

The LIFETEST procedure now produces Nelson-Aalen estimates of the cumulative hazard function. The number of subjects at risk can be displayed for Kaplan-Meier survival curves. Comparison methods are available for the *k*-sample test, and you can now specify a smoother hazard function by using the kernel method.

The CLASS statement, previously available only in the TPHREG procedure, is now included with the PHREG procedure. PROC PHREG now fits the piecewise exponential model, which is specified in the BAYES statement. Bayesian baseline survival prediction becomes available with SAS 9.2 as well. The HAZARDRATIO statement provides a new facility for computing hazard ratios, including hazard ratios in the presence of interactions. The PLOTS option in the PROC PHREG statement produces

baseline survival function plots. Profile-likelihood confidence limits are now available for hazard ratios produced in classical analyses.

Power and Sample Size Software

The POWER procedure now performs power and sample size analyses for the likelihood ratio chi-square test of a single predictor in binary logistic regression, possibly in the presence of one or more covariates (where all predictors are independent of each other). It also performs power and sample size analyses for the Wilcoxon-Mann-Whitney test for two independent groups. The ONESAMPLEFREQ statement now covers equivalence, noninferiority, and confidence interval precision for a proportion. The GLMPOWER procedure now supports continuous variables, and the noncentrality parameter is computed.

Survey Data Analysis Software

The SURVEYFREQ, SURVEYMEANS, SURVEYLOGISTIC, and SURVEYREG procedures now provide variance estimation by balanced repeated replication (BRR) and jackknife methods, in addition to the Taylor series method. You can provide replicate weights for the new replication methods with a REPWEIGHTS statement, or the procedures can construct the replicate weights. The new NOMCAR option in these procedures requests a subpopulation analysis of the set of respondents for Taylor series variance estimation.

In addition, PROC SURVEYFREQ now computes odds ratio and relative risk estimates. The OUTPUT and DOMAIN statements are now available in the SURVEYLOGISTIC and SURVEYREG procedures. PROC SURVEYMEANS now computes percentiles (Woodruff variance estimation only). The SURVEYSELECT procedure now provides methods to allocate the total sample size among the strata. Allocation methods include proportional, Neyman, and optimal allocation.

TTEST Procedure Updated

The TTEST procedure now performs TOST equivalence analyses, analyses of treatment and period in AB/BA crossover designs, weighted Satterthwaite tests and confidence intervals, analyses of ratios, and one-sided analyses. It supports both normal and lognormal data. Sasabuchi tests and Fieller confidence intervals are computed for normal ratios. PROC TTEST now provides graphs, including histograms, densities, box plots, profiles, agreement plots, Q-Q plots, and interval plots.
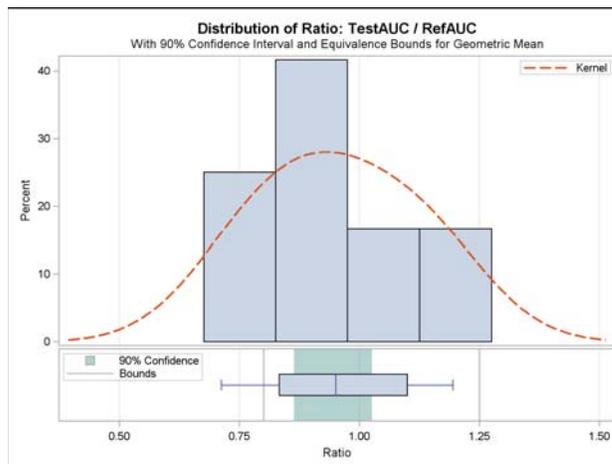
Other Highlights

The experimental EFFECT statement provides for the creation of splines and other special effects. It is available in the QUANTREG, GLIMMIX, and GLMSELECT procedures. The LOGISTIC procedure now provides the ROCCONTRAST statement for comparing different ROC models. PROC LOGISTIC also computes odds ratios in the presence of interactions, as well as providing odds ratio plots. The GAM procedure is production and offers the usual SAS/STAT options for response and classification variables.

The GENMOD procedure now fits zero-inflated Poisson regression models, as well as computes AIC and QIC model fit statistics. PROC GENMOD also produces deletion diagnostics and plots for GLMs and GEEs.

SAS® Stat Studio

SAS/STAT users will be interested in SAS Stat Studio, which is new software for data exploration and analysis. It provides a highly flexible programming environment in which you can run SAS/STAT or SAS/IML® analyses and display the results with dynamically linked graphics and data tables. Stat Studio is intended for data analysts who write SAS programs to solve statistical problems but need more versatility for model building and for implementing their own innovative methods. The programming language in SAS Stat Studio, which is called IMLPlus, is an enhanced version of the IML programming language. IMLPlus extends SAS/IML to provide new language features, including the ability to create and manipulate customized dynamic graphics, call SAS procedures as functions, and call computational programs written in C, C++, Java, and FORTRAN. SAS Stat Studio runs on a PC in the Microsoft Windows operating environment. It is distributed with the SAS/IML product.