

# Release 8.2 Enhancements to SAS/STAT Software

Release 8.2 updates SAS/STAT software with new procedures and enhancements to existing procedures. The following are the highlights of the new release:

## Multiple Imputation

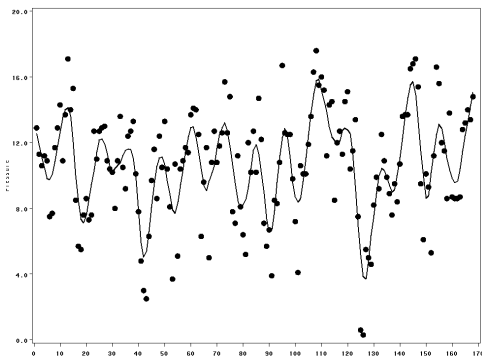
Multiple imputation provides a useful strategy for dealing with data sets with missing values. Instead of filling in a single value for each missing value, Rubin's (1987) multiple imputation procedure replaces each missing value with a set of plausible values that represent the uncertainty about the right value to impute. These multiply imputed data sets are then analyzed by using standard procedures for complete data and combining the results from these analyses. This results in statistically valid inferences.

Release 8.2 of SAS/STAT software includes the second experimental releases of the MI and MIANALYZE procedures for creating and analyzing multiply imputed data sets for incomplete multivariate data. The MI procedure is a multiple imputation procedure that creates multiply imputed data sets for incomplete  $p$ -dimensional multivariate data. It uses methods that incorporate appropriate variability across  $m$  imputations. Once the  $m$  complete data sets are analyzed using standard SAS/STAT procedures, PROC MIANALYZE can be used to generate valid statistical inferences about these parameters by combining the results. The MI procedure provides three methods for imputing missing values. For monotone missing data patterns, either a parametric regression method that assumes multivariate normality or a nonparametric method that uses propensity scores is appropriate. For an arbitrary missing data pattern, a Markov chain Monte Carlo (MCMC) method that assumes multivariate normality can be used.

Additions to PROC MI in Release 8.2 include the TRANSFORM statement to transform variables before performing the imputation, autocorrelation and iteration plots, a monotone-data MCMC method to impute just enough values to achieve a monotone missing pattern for the imputed data, and the EM statement to derive the MLE and related EM results.

## Generalized Additive Models

The GAM procedure, production in Release 8.2, fits generalized additive models as defined by Hastie and Tibshirani (1990), providing a set of powerful tools for data analysis that are based on nonparametric and smoothing techniques. These techniques relax the usual linearity assumptions and allow you to discern structure in the relationship between independent variables and dependent variables that you might otherwise not find.



The generalized additive models combine an additive assumption that allows many nonparametric relationships to be explored simultaneously plus provides the distributional flexibility of the generalized linear model. You can use the GAM procedure when you have multiple independent variables that you want to model nonparametrically, or when the dependent variable is not distributed normally. The GAM procedure provides nonparametric estimates for additive models, supports the use of multidimensional data, fits both generalized semiparametric additive models and generalized additive models, and enables you to choose a model by specifying model degrees of freedom or smoothing parameter.

Enhancements for Release 8.2 include the addition of a local regression smoother (LOESS) for univariate smoothing. In addition, several new distributions are available with the DIST= option in the MODEL statement, including the gamma, Poisson, and binomial distributions. Additional statistics are available with the OUTPUT statement, and multiple SCORE statements are now supported.

The following are some of the key enhancements to existing procedures in SAS/STAT software in Release 8.2.

### **BOXPLOT Procedure**

Several new options in the PLOT statement enable you to overlay plots of additional variables on a box plot and to create links in a plot when graphics output is directed into HTML.

### **LIFEREG Procedure**

The LIFEREG procedure has been updated to support interaction terms among any of the explanatory variables in the same manner as in the GLM procedure. You can construct probability plots for complete or censored data with the new PROBLOT statement, displaying the fitted distribution and confidence intervals along with nonparametric estimates of the CDF on the probability plots. The NEW XDATA= input data set can be used to provide values for the effects when the predicted probability and quantile and their confidence limits are computed. In addition, you can now obtain standardized and Cox-Snell residuals in the OUTPUT= data set.

### **LOESS Procedure**

The LOESS procedure now performs automatic smoothing parameter selection by setting (as near as possible) an approximate measure of model degrees of freedom to a specified target value. You can use these methods to make meaningful comparisons between loess fits and other parametric and nonparametric fits. You request this selection by using the SELECT=DFTYPE options in the MODEL statement.

### **LOGISTIC Procedure**

You can now fit a generalized logit model when you have nominal response data. The generalized logit model produces a set of intercepts and slope parameters for each response level other than the reference level. Exact logistic regression is also available for the generalized logits model. You specify the generalized logits model by specifying the option LINK=GLOGIT in the MODEL statement. In addition, you can now specify options that are specific to the response variable by enclosing them in a set of parentheses right after the response variable in the MODEL statement. Such response variable specific options include the EVENT=, REFERENCE=, ORDER=, and DESCENDING= options.

### **NPAR1WAY Procedure**

You can now produce the exact Kolmogorov-Smirnov test for two-sample data and you can also now produce exact point probabilities for the statistics produced by the EXACT statement.

### **PROBIT Procedure**

The PROBIT procedure has been updated to support interaction terms among any of the explanatory variables in the same manner as in the GLM procedure. You can construct plots of predicted probabilities, inverse predicted probabilities, and cumulative probabilities for binomial models. You can construct plots of predicted probabilities, linear predictors, and cumulative probabilities for ordinal multinomial models. The new XDATA= input data set can be used to provide values for the effects other than the single continuous dosage effect when the predicted dosage and their fiducial limits are computed. The OUTEST= data set is extended to both binomial and multinomial models. The new AGGREGATE option in the MODEL statement defines the subpopulations for calculating the Pearson chi-square statistic and the deviance.

### **SURVEYMEANS Procedure**

The new RATIO statement enables you to perform ratio analysis for means or proportions of the analysis variables. The numerator and denominators of your ratio can be either continuous or categorical variables, and you can submit as many RATIO statements as you want.

### **TRANSREG Procedure**

You can now request a Box-Cox transformation of the dependent variable with the new BOXCOX option in the MODEL statement. You can specify a list of power parameters using the LAMBDA= transformation option.