SAS/STAT® 12.1 Software
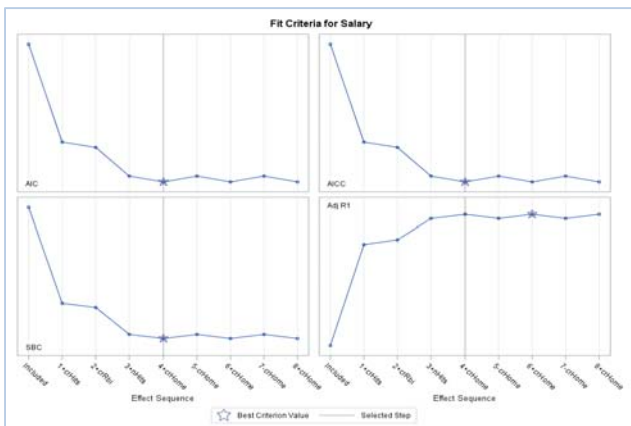
Comprehensive Statistical Capabilities

Overview

Heralded by a new release-numbering scheme, SAS/STAT 12.1 comes loaded with new statistical capabilities. New directions include model selection for quantile regression, quantile regression for censored data, and multivariate adaptive regression splines. Epidemiologists will like the STDRATE procedure for computing direct and indirect standardized rates and risks for study populations. The FMM procedure becomes production and includes new features such as support for additional distributions. Other notable enhancements include modeling missing covariates in the MCMC procedure and fitting Bayesian frailty models in the PHREG procedure.

QUANTSELECT Procedure

Quantile regression is a modern technique in which the quantiles of a response variable are modeled. It is especially useful for analyzing heterogeneous data in which the distribution of the response variable at the tails and central location vary conditionally on the explanatory covariates.
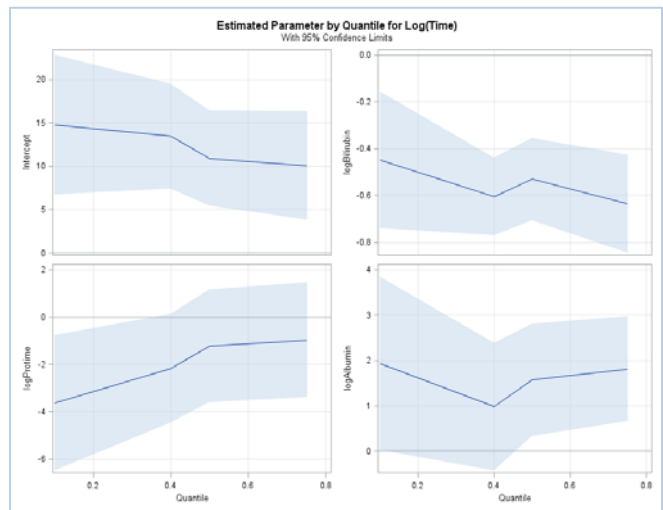
The new QUANTSELECT procedure performs model selection for quantile regression. It supports model selection for a single quantile and for a quantile process. The QUANTSELECT procedure offers capabilities similar to those offered by the GLMSELECT procedure. Selection methods include forward, backward, stepwise, and LASSO. PROC QUANTSELECT uses variable selection criteria such as AIC, SBC, and AICC. You can also partition your data into training, validation, and testing data sets and then use the validation data set to measure the fitness of candidate models.

QUANTLIFE Procedure

Quantile regression also provides an alternative and flexible technique for the analysis of survival data. You can apply the technique to right-censored responses, which allows you to explore the covariate effects on the quantiles of interest. Two approaches are implemented: one is based on the Kaplan-Meier estimator, and the other is based on the Nelson-Allen estimator of the cumulative hazard function.

The new QUANTLIFE procedure provides these methods in SAS. PROC QUANTLIFE employs interior point algorithms for estimation and produces Wald tests, survival plots, conditional quantile plots, and quantile process plots. PROC QUANTLIFE also supports the EFFECT statement so that it can fit regression quantile spline curves. The procedure is multithreaded to take advantage of multiple processors when they are available.

STDRATE Procedure

The new STDRATE procedure computes direct and indirect standardized rates and risks for study populations. Direct standardization uses weights such as population-time from a standard or reference population to compute the weighted average of stratum-specific estimates in the study population. For two study populations with the same reference population, PROC STDRATE compares directly standardized rates or risks. For two study populations without a reference population, the procedure computes Mantel-Haenszel effect estimates, such as the rate difference.

ADAPTIVEREG Procedure

The new ADAPTIVEREG procedure provides a nonparametric modeling approach for high-dimensional data. PROC ADAPTIVEREG fits multivariate adaptive regression splines as introduced by Friedman (1991). The approach is a nonparametric regression technique that combines both regression splines and model selection methods. It does not assume parametric model forms, and it does not require knot values for constructing regression spline terms. Instead, it constructs spline basis functions in an adaptive way by automatically selecting appropriate knot values for different variables, and it 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 distributions from the exponential family, such as binomial and Poisson. PROC ADAPTIVEREG is multithreaded to take advantage of multiple processors.

PROC ADAPTIVEREG enables you to force effects into the final model or restrict variables in linear forms; supports options for fast forward selection; supports partitioning of data into training, validation, and testing roles; and provides graphical representations of the selection process.

FMM Procedure

PROC FMM fits statistical models to data for which 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 FMM procedure becomes production in SAS/STAT 12.1.
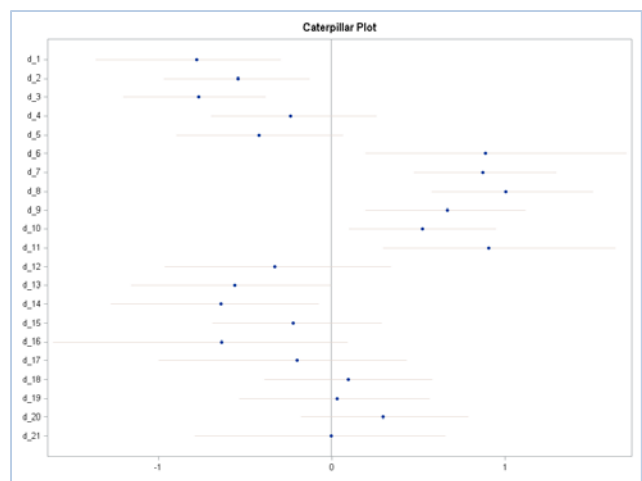
MCMC Procedure

The MCMC procedure provides many new capabilities in SAS/STAT 12.1, including the following:

- The MODEL statement augments missing values in the response variable by default. In addition, the MCMC procedure now models missing values for the covariates. 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.

- The RANDOM statement supports more distributions, such as the multivariate normal distribution with autoregressive structure, Poisson distribution, and general distribution (for the construction of nonstandard distributions).

- 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 provided as an alternative sampling algorithm for both the model parameters and random-effects parameters.

Caterpillar Plot

Other Enhancements

Other enhancements in SAS/STAT 12.1 include

- Mosaic plots and heat maps
- Partial proportional odds model
- Bayesian frailty models
- Poisson sampling in PROC SURVEYSELECT
- Poststratification estimation in PROC SURVEYMEANS
- Fleming-Harrington estimates in PROC PHREG
- ESTIMATE statement in PROC QUANTREG
- Additional postprocessing inference for the LIFEREG and PROBIT procedures

For More Information

SAS/STAT 12.1 is now available. For more information, see support.sas.com/statistics