



THE
POWER
TO KNOW.

SAS/STAT[®] 14.1 User's Guide

What's New in SAS/STAT 14.1

This document is an individual chapter from *SAS/STAT® 14.1 User's Guide*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2015. *SAS/STAT® 14.1 User's Guide*. Cary, NC: SAS Institute Inc.

SAS/STAT® 14.1 User's Guide

Copyright © 2015, SAS Institute Inc., Cary, NC, USA

All Rights Reserved. Produced in the United States of America.

For a hard-copy book: No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

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 License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication, or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a), and DFAR 227.7202-4, and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

July 2015

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 trademarks of their respective companies.

Chapter 1

What's New in SAS/STAT 14.1

Contents

Overview	2
New Procedures	2
Highlights of Enhancements	3
Highlights of Enhancements in SAS/STAT 13.2	3
Enhancements	4
BCHOICE Procedure	4
CALIS Procedure	4
FACTOR Procedure	5
FMM Procedure	5
FREQ Procedure	5
GEE Procedure	5
GENMOD Procedure	6
GLIMMIX Procedure	6
GLMSELECT Procedure	6
HPGENSELECT Procedure	6
HPLOGISTIC Procedure	6
HPPRINCOMP Procedure	7
HPSPLIT Procedure	7
ICLIFETEST Procedure	7
ICPHREG Procedure	7
IRT Procedure	8
LIFETEST Procedure	8
LOGISTIC Procedure	8
MCMC Procedure	8
MI Procedure	9
MIXED Procedure	9
NLMIXED Procedure	9
NPAR1WAY Procedure	9
PHREG Procedure	10
POWER Procedure	10
QUANTREG Procedure	10
QUANTSELECT Procedure	10
SPP Procedure	10
SURVEYLOGISTIC Procedure	10
SURVEYMEANS Procedure	11
SURVEYPHREG Procedure	11

What's Changed	11
References	12

Overview

SAS/STAT 14.1 includes two new procedures and many enhancements. 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 now documented in *SAS/STAT User's Guide* as well as in *SAS/STAT User's Guide: High-Performance Procedures*.

New Procedures

GAMPL Procedure

The GAMPL procedure is a high-performance procedure that fits generalized additive models by penalized likelihood estimation. Based on low-rank regression splines (Wood 2006), these models are powerful tools for nonparametric regression and smoothing. Generalized additive models are extensions of generalized linear models. They relax the linearity assumption in generalized linear models by allowing spline terms in order to characterize nonlinear dependency structures. Each spline term is constructed by the thin-plate regression spline technique (Wood 2003). A roughness penalty is applied to each spline term by a smoothing parameter that controls the balance between goodness of fit and the roughness of the spline curve. PROC GAMPL fits models for standard distributions in the exponential family, such as normal, Poisson, and gamma distributions.

Both the GAMPL procedure and the GAM procedure in SAS/STAT software fit generalized additive models. However, the GAMPL procedure uses different approaches for constructing spline basis expansions, fitting generalized additive models, and testing smoothing components. It focuses on automatic smoothing parameter selection by using global model-evaluation criteria to find optimal models. The GAM procedure focuses on constructing models by fitting partial residuals against each smoothing term. In general, you should not expect similar results from the two procedures.

SURVEYIMPUTE Procedure

The SURVEYIMPUTE procedure imputes missing values of an item in a sample survey by replacing them with observed values from the same item. Imputation methods include single and multiple hot-deck imputation and fully efficient fractional imputation (FEFI) (Fuller 2009, section 5.2). Donor selection techniques include simple random selection with or without replacement, probability proportional to weights selection (Rao and Shao 1992), and approximate Bayesian bootstrap selection (Rubin and Schenker 1986). When you use FEFI, PROC SURVEYIMPUTE produces replicate weights that appropriately account for the imputation. You can use these replicate weights in any survey analysis procedure to correctly estimate the variance of an estimator that uses the imputed data.

Highlights of Enhancements

Following are some highlights of the enhancements in SAS/STAT 14.1:

- The BCHOICE procedure allows varying numbers of alternatives in choice sets for logit models.
- Exact mid- p , likelihood ratio, and Wald modified confidence limits are available for the odds ratio produced by the FREQ procedure.
- The GLIMMIX procedure provides the multilevel adaptive Gaussian quadrature algorithm of Pinheiro and Chao (2006) for multilevel models, which can greatly reduce the computational and memory requirements for these models with many random effects.
- The GLMSELECT procedure supports the group LASSO method.
- The IRT procedure fits generalized partial credit models.
- The LIFETEST procedure performs nonparametric analysis of competing-risks data.
- The LOGISTIC procedure fits an adjacent-category logit model to ordinal response data.
- The MCMC procedure adds an ordinary differential equation (ODE) solver and a general integration function, enabling the procedure to fit models that contain differential equations (for example, PK models) or models that require integration (for example, marginal likelihood models).
- The NPAR1WAY procedure performs stratified rank-based analysis for two-sample data.
- The POWER procedure supports Cox proportional hazards regression models.

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 13.2

Some users might be unfamiliar with updates that were made in the previous release. SAS/STAT 13.2 introduced the GEE, ICPHREG, and SPP procedures. Following are highlights of the other enhancements in SAS/STAT 13.2:

- The FACTOR procedure generates path diagrams.
- The FMM procedure fits multinomial models.
- The FREQ procedure 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 on which model selection is performed.
- The IRT procedure generates polychoric correlation matrices, item characteristic curves, and test information curve plots.

- The LOGISTIC procedure enables you to add or relax constraints on parameters in nominal response and partial proportional odds models.
- 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.

Enhancements

BCHOICE Procedure

The BCHOICE procedure allows varying numbers of alternatives in choice sets for logit models. The new RESTRICT statement enables you to specify boundary requirements and order constraints on fixed effects for logit models.

CALIS Procedure

The METHOD=MLSB option in the PROC CALIS statement computes maximum likelihood estimates, Satorra-Bentler scaled chi-square statistics for model fit, and standard error estimates that are based on a sandwich-type formula proposed by Satorra and Bentler (1994). This estimation method is suitable when you apply the normal-theory-based maximum likelihood estimation to either normal or nonnormal data. You can modify the formula of the Satorra-Bentler scaled statistic by using the SBNTW= option. SBNTW=PRED (the default) uses the estimated covariance matrix to compute the normal-theory weight matrix in the formula; SBNTW=OBS uses the observed covariance matrix.

The INFORMATION= option enables you to select either the expected or observed information matrix when you compute standard error estimates. The default is INFORMATION=OBSERVED when you use full information maximum likelihood (FIML) estimation. All other estimation methods use INFORMATION=EXPECTED as the default.

The STDERR=SBSW option enables you to compute standard errors that are based on a sandwich-type formula proposed by Satorra and Bentler (1994). The METHOD=MLSB option automatically sets the STDERR=SBSW option. All other estimation methods use STDERR=UNADJUSTED as the default.

The DESIGNWIDTH= and DESIGNHEIGHT= options in the PATHDIAGRAM statement enable you to change the default design dimensions of path diagrams. You can accommodate more nodes (variables) in the path diagram by setting a larger design dimension. The SCALE= option in the PATHDIAGRAM statement adjusts the relative node size. The TEXTSIZEMIN= option sets the minimum font size for text in path diagrams. These new options apply only to the GRIP layout algorithm (ARRANGE=GRIP).

FACTOR Procedure

The DESIGNWIDTH= and DESIGNHEIGHT= options in the PATHDIAGRAM statement enable you to change the default design dimensions of path diagrams. You can accommodate more nodes (variables) in the path diagrams by setting a larger design dimension. The SCALE= option in the PATHDIAGRAM statement adjusts the relative node size. The TEXTSIZEMIN= option sets the minimum font size for text in path diagrams. These new options apply only to the GRIP layout algorithm (ARRANGE=GRIP).

FMM Procedure

You can use the Dirichlet-multinomial distribution as the response distribution for a mixture component.

FREQ Procedure

Noninferiority, equivalence, and equality tests are available for the relative risk. Test methods include score (Farrington-Manning), Wald, Wald modified, and likelihood ratio. You can specify these tests and methods by using the RELRISK option in the TABLES statement.

Likelihood ratio and Wald modified confidence limits are available for the relative risk. (You can request confidence limits for the relative risk by using the RELRISK(CL=) option.) These confidence limits can be displayed in the relative risk plot.

Exact mid- p , likelihood ratio, and Wald modified confidence limits are available for the odds ratio. (You can request confidence limits for the odds ratio by using the OR(CL=) option.) These confidence limits can be displayed in the odds ratio plot.

The score (Farrington-Manning) and Hauck-Anderson methods are available for the risk difference equality test, which you can request by using the RISKDIFF option in the TABLES statement. You can specify a null value of the risk difference for equality tests.

You can specify a null value of the ratio of discordant proportions for McNemar's test by using the AGREE(MNULLRATIO=) option.

GEE Procedure

The GEE procedure is production in this release.

The GEE procedure supports the ESTIMATE, LSMEANS, and OUTPUT statements.

The DIST=MULTINOMIAL option in the MODEL statement enables you to specify the multinomial distribution for polytomous responses. Responses that follow a multinomial distribution can be treated as either nominal or ordinal data.

The LOGOR= option in the REPEATED statement enables you to perform alternating logistic regression (ALR) analysis for ordinal and binary data.

GENMOD Procedure

An overdispersion diagnostic plot is available for zero-inflated models. You can use it to examine the level of overdispersion that has been accounted by the fitted zero-inflated model relative to a standard Poisson regression model.

GLIMMIX Procedure

You can apply the multilevel adaptive Gaussian quadrature algorithm of Pinheiro and Chao (2006) to multilevel models. This algorithm reduces the number of random effects over which the marginal likelihood needs to be integrated, significantly reducing the computational and memory requirements.

GLMSELECT Procedure

The `SELECTION=GROUPLASSO` option specifies the group LASSO method, a variant of LASSO that estimates parameters by using a penalty on the sum of the Euclidean norms of groups of regression coefficients. This enables you to constrain groups of parameters to enter the model together, such as all the parameters that correspond to the effect of a classification variable.

HPGENSELECT Procedure

The `INEST=` option in the `PROC HPGENSELECT` statement enables you to input your own starting values for the optimization, as well as bounds and equality constraints for single parameters. The `OUTEST` option adds a column that contains the parameter names to the “Parameter Estimates” ODS OUTPUT data set.

The `RESTRICT` statement enables you to specify general linear equality or inequality constraints on regression parameters.

The `METHOD=LASSO` option in the `SELECTION` statement enables you to perform model selection by using the LASSO method.

The `CHOOSE=VALIDATE` option enables you to choose a final model by using the minimum Sawa Bayesian information criterion (BIC) criterion that is computed for validation data when you specify LASSO model selection.

The `BY` statement enables processing with BY groups.

HPLOGISTIC Procedure

The `ALLSTATS` option in the `OUTPUT` statement outputs all valid statistics.

HPPRINCOMP Procedure

The METHOD= option in the PROC HPPRINCOMP statement specifies which of the principal component extraction methods to use. METHOD=EIG requests the eigenvalue decomposition method, METHOD=NIPALS requests the nonlinear iterative partial least squares (NIPALS) method, and METHOD=ITERGS requests the iterative method based on Gram-Schmidt orthogonalization (ITERGS).

The OUTPUT statement creates a data set that contains observationwise statistics, which are computed after the model is fit. The OUTPUT statement causes the OUT= option in the PROC HPPRINCOMP statement to be ignored.

HPSPLIT Procedure

The HPSPLIT procedure supports the MODEL and CLASS statements, making its syntax comparable to that of other SAS/STAT modeling procedures.

Cost-complexity pruning is supported, with k -fold cross validation as the default method of selecting the penalty parameter.

ODS tables provide confusion matrices for training and validation data; fit statistics for the selected tree; and information about cross validation, nodes, and variable importance.

You can generate tree plots, cross validation plots, and ROC curves.

You can save observationwise results, including node and leaf assignments, predicted levels, posterior probabilities for classification trees, and predicted response values for regression trees in an output data set.

ICLIFETEST Procedure

You can specify the OUTSCORE= option in the TEST statement to output the subject-specific scores that are derived under the permutation form of the generalized log-rank test statistic.

ICPHREG Procedure

You can fit stratified proportional hazards models by using a STRATA statement. Supported models include the piecewise constant hazards model (default) and the cubic spline model.

You can fit models to data that might be left-truncated by specifying the ENTRY= option in the MODEL statement. The variable that you specify is treated as the left-truncation time for each subject of the data set.

You can produce hazard plots by specifying the PLOT=HAZARD option in the PROC ICPHREG statement. By default, a plot of the baseline hazard function is created. If you specify an input data set by using the COVARIATES= option in the BASELINE statement, the procedure plots the estimated hazard functions conditional on the input covariate patterns.

The Lagakos and deviance residuals (Farrington 2000) are available for assessing the fitted proportional hazards model. You can display these residuals against their observational index in the input data by default (INDEX option) or against the estimated linear predictor by specifying the XBETA option.

IRT Procedure

The generalized partial credit (GPC) model is available for ordinal items.

The ITEMSTAT option in the PROC IRT statement displays the classical item statistics table.

The IRT procedure supports the OUTMODEL= and INMODEL= options in the PROC IRT statement. The OUTMODEL= option enables you to create an output data set that contains the model specification. The INMODEL= option enables you to specify an input data set that contains information about the analysis model. The INMODEL(SCORE)= option enables you to score subjects in a new data set by using the parameter estimates from a previous analysis.

The XVIEWMAX and XVIEWMIN suboptions in the PLOTS= option specify the maximum and minimum values for the X axis. They can also be used as suboptions in the PLOTS=ICC, PLOTS=IIC, and PLOTS=TIC options.

LIFETEST Procedure

The FAILCODE option in the TIME statement enables you to perform a nonparametric analysis of competing-risks data. PROC LIFETEST estimates the cumulative incidence function (CIF), which is the probability subdistribution of failure from a specific cause. If you have more than one sample of competing-risks data, PROC LIFETEST performs Gray's test (Gray 1988) to compare the CIFs of the samples.

LOGISTIC Procedure

The LINK=ALOGIT option fits an adjacent-category logit model to ordinal response data.

The EQUALSLOPES and UNEQUALSLOPES options are available with all polytomous response models.

The ROCOPTIONS(CROSSVALIDATE) option in the PROC LOGISTIC statement creates ROC curves and computes the AUC by using cross validated predicted probabilities.

The ROCCI option in the MODEL statement produces the ROC association table for a single model.

You can control ROC titles and axis labels by using macro variables.

The ORPVALUE option in the MODEL statement displays p -values for odds ratios.

The METHOD=MCMC option in the EXACTOPTIONS statement implements exact logistic regression according to the method of Forster, McDonald, and Smith (2003).

MCMC Procedure

Two default sampling algorithms for continuous parameters have been added to the procedure: Hamiltonian Monte Carlo (HMC) and No-U-Turn Sampler (NUTS). You can select them as replacements for the normal- or t -distribution-based random-walk Metropolis algorithm to draw posterior samples.

PROC MCMC supports lagging and leading variables. This enables the procedure to fit dynamic linear models, state space models, autoregressive models, or other models that have a conditionally dependent structure on either the random-effects parameters or the response variable.

PROC MCMC adds an ordinary differential equation (ODE) solver and a general integration function, which enable the procedure to fit models that contain differential equations (for example, PK models) or models that require integration (for example, marginal likelihood models).

The PREDDIST statement makes predictions from marginal random-effects models. For example, you can make predictions for new observations that do not have group membership information in a random-effects model.

MI Procedure

The NIMPUTE= option supports the PCTMISSING suboption for specifying the number of imputations as the percentage of incomplete cases. The new default is NIMPUTE=25.

The new LIKELIHOOD=AUGMENT suboption of the LOGISTIC option in the FCS and MONOTONE statements adds observations to the likelihood function in the computation of maximum likelihood parameter estimates for logistic regression. This suboption is applicable in situations where the likelihood estimates do not otherwise exist because the data points exhibit a complete separation pattern or a quasi-complete separation pattern. The method of augmenting the data is from White, Daniel, and Royston (2010).

MIXED 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 reduces the precision estimator bias of the fixed and random effects under nonlinear covariance structures.

NLMIXED Procedure

The NLMIXED procedure is multithreaded and can take advantage of multiple processors. The NTHREADS= option in the PROC NLMIXED statement specifies the number of threads to use for marginal log-likelihood calculations when you specify the METHOD=GAUSS option. PROC NLMIXED allocates data to different threads and calculates the objective function by accumulating values from each thread when it calculates a marginal log likelihood. The NTHREADS= option is applicable only when you specify METHOD=GAUSS (the default) in the PROC NLMIXED statement.

NPAR1WAY Procedure

The new STRATA statement provides stratified rank-based analysis of two-sample data. The following score types are available: Wilcoxon, median, normal (Van der Waerden), Savage, and raw data scores. Rank-based scores can be computed by using within-stratum ranks or overall ranks; strata can be equally weighted or weighted by stratum size. The new ALIGN=STRATA option in the PROC NPAR1WAY statement aligns

responses by location measure within strata before performing unstratified analyses; location measures include the stratum median, mean, and Hodges-Lehmann shift.

PHREG Procedure

The FAST option in the PROC PHREG statement can speed up the fitting of the Breslow and Efron partial likelihoods with counting process style of input. The algorithm is especially efficient when the data have a large number of observations with many event times. It requires only one pass through the data to compute the likelihood, gradient, and Hessian.

POWER Procedure

The POWER procedure supports Cox proportional hazards regression models and Farrington-Manning noninferiority tests of relative risk.

QUANTREG Procedure

The QUANTREG procedure supports a new algorithm for quantile regression model fitting, an alternative interior point algorithm that can be more efficient for large data. You request this new algorithm by specifying the ALGORITHM=IPM option in the PROC QUANTREG statement.

QUANTSELECT Procedure

The QUANTSELECT procedure can output the parameter estimates from quantile process regression to a data set. You request this output data set by using an appropriate ODS statement to select it and to name the data set.

Furthermore, when you specify the QUANTILE=PROCESS option, PROC QUANTSELECT can also compute the estimated quantile level for each observation, based on the conditional cumulative distribution function. You request observation quantiles by specifying the QUANTLEVEL option in the OUTPUT statement, in addition to specifying QUANTILE=PROCESS in the MODEL statement.

SPP Procedure

You can diagnose overdispersion in the event counts when you specify a negative binomial model.

SURVEYLOGISTIC Procedure

The DF= option in the MODEL statement enables you to perform a wider variety of adjusted and customized Wald F tests.

SURVEYMEANS Procedure

Replication variance estimates for quantiles are based on the approach of Fuller (2009) by default. This method uses a smooth representation of the quantiles and can therefore provide an estimated variance with better stability than that of the naïve jackknife variance estimate that was previously employed.

SURVEYPHREG Procedure

The `DF=` option in the `MODEL` statement enables you to perform a wider variety of adjusted and customized Wald F tests.

What's Changed

The following sections describe changes in software behavior from SAS/STAT 13.2 to SAS/STAT 14.1.

Random Number Generator Change

Users might occasionally see notes in the SAS log about random number generation changes in the `BCHOICE`, `GENMOD`, `MCMC`, `MULTTEST`, `NLIN`, `PHREG`, `SURVEYIMPUTE`, and `SURVEYSELECT` procedures. For a very small number of seeds, the Mersenne twister algorithm that is used is changed as discussed in Matsumoto and Nishimura (2002). To revert to the former (1998) version of the Mersenne twister algorithm, submit the following statement:

```
OPTIONS SET=SAS_RNG_METHOD=MT1998;
```

Note that in the case of the `SURVEYSELECT` procedure, this message might be printed as a result of an internal stratum seed, not the initial seed that the user specifies. See *SAS Language Reference: Dictionary* for a description of the `OPTIONS` statement.

CALIS Procedure

The `CALIS` procedure computes standard error estimates based on the observed information matrix when you use the full information maximum likelihood method (`METHOD=FIML`). Before this release, standard errors are computed based on the expected information matrix for all estimation methods. You can use the `INFORMATION=` option to specify which information matrix to use.

LOGISTIC Procedure

The `BINWIDTH` option in the `MODEL` statement is set to 0 by default; this affects the statistics that are displayed in the “Association” table.

MI Procedure

The default number of imputations is changed from 5 to 25. The previous default value of 5 was based on relative efficiency considerations. However, recent studies, which consider aspects such as confidence intervals and p -values, recommend a larger number of imputations (Allison 2012; Van Buuren 2012, pp. 49–50).

SPP Procedure

The SPP procedure uses the normal distribution instead of the t distribution for inference on the parameters of the nonhomogeneous intensity model.

SURVEYLOGISTIC Procedure

A Rao-Scott first- or second-order adjustment is applied to the likelihood ratio test of the global null hypothesis.

SURVEYMEANS Procedure

The estimated standard error for quantiles is now modified by a smoothing method when you use replication variance estimation methods. The new option `VARMETHOD=BRR(NAIVEQUANTILE)` or `VARMETHOD=JK(NAIVEQUANTILE)` produces the previous results.

References

- Allison, P. (2012). “Why You Probably Need More Imputations Than You Think.” Accessed February 20, 2015. <http://www.statisticalhorizons.com/more-imputations>.
- Farrington, C. P. (2000). “Residuals for Proportional Hazards Models with Interval-Censored Survival Data.” *Biometrics* 56:473–482.
- Forster, J. J., McDonald, J. W., and Smith, P. W. F. (2003). “Markov Chain Monte Carlo Exact Inference for Binomial and Multinomial Logistic Regression Models.” *Statistics and Computing* 13:169–177.
- Fuller, W. A. (2009). *Sampling Statistics*. Hoboken, NJ: John Wiley & Sons.
- Gray, R. J. (1988). “A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk.” *Annals of Statistics* 16:1141–1154.
- 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.
- Matsumoto, M., and Nishimura, T. (2002). “Mersenne Twister with Improved Initialization.” Accessed April 10, 2015. <http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/MT2002/emt19937ar.html>.
- Pinheiro, J. C., and Chao, E. C. (2006). “Efficient Laplacian and Adaptive Gaussian Quadrature Algorithms for Multilevel Generalized Linear Mixed Models.” *Journal of Computational and Graphical Statistics* 15:58–81.

- Rao, J. N. K., and Shao, J. (1992). “Jackknife Variance Estimation with Survey Data under Hot Deck Imputation.” *Biometrika* 79:811–822.
- Rubin, D. B., and Schenker, N. (1986). “Multiple Imputation for Interval Estimation from Simple Random Samples with Ignorable Nonresponse.” *Journal of the American Statistical Association* 81:366–374.
- Satorra, A., and Bentler, P. M. (1994). “Corrections to Test Statistics and Standard Errors in Covariance Structure Analysis.” In *Latent Variables Analysis: Applications for Developmental Research*, edited by A. von Eye, and C. C. Clogg, 399–419. Thousand Oaks, CA: Sage.
- Van Buuren, S. (2012). *Flexible Imputation of Missing Data*. Boca Raton, FL: Chapman & Hall/CRC.
- White, I. R., Daniel, R., and Royston, P. (2010). “Avoiding Bias Due to Perfect Prediction in Multiple Imputation of Incomplete Categorical Variables.” *Computational Statistics and Data Analysis* 54:2267–2275.
- Wood, S. (2003). “Thin Plate Regression Splines.” *Journal of the Royal Statistical Society, Series B* 65:95–114.
- Wood, S. (2006). *Generalized Additive Models*. Boca Raton, FL: Chapman & Hall/CRC.