THE POWER TO KNOW.

SAS/STAT® 14.3 User's Guide

What's New in SAS/STAT 14.3

This document is an individual chapter from *SAS/STAT® 14.3 User's Guide*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2017. *SAS/STAT® 14.3 User's Guide*. Cary, NC: SAS Institute Inc.

SAS/STAT® 14.3 User's Guide

Copyright © 2017, 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

September 2017

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.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to http://support.sas.com/thirdpartylicenses.

Chapter 1
What's New in SAS/STAT 14.3

Contents

Overview

SAS/STAT 14.3 includes one new procedure 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 documented both in *SAS/STAT User's Guide* and in *SAS/STAT User's Guide: High-Performance Procedures*.

New Procedure

The CAUSALMED procedure estimates causal mediation effects from observational data. In a causal mediation analysis, there are four main variables of interest:

- an outcome variable Y
- a treatment variable T that is hypothesized to have direct and indirect causal effects on the outcome variable Y
- a mediator variable M that is hypothesized to be causally affected by the treatment variable T and that itself has a direct effect on the outcome variable Y
- a set of background covariates C that confound the observed relationships among Y, T, and M

The main goal of the analysis is to obtain unbiased estimates of the direct causal effect of the treatment T on Y and the indirect causal effect of T on Y through the mediator M.

The CAUSALMED procedure uses a regression-based method of estimation (Valeri and VanderWeele 2013; VanderWeele 2014) within the counterfactual framework of Robins and Greenland (1992) and Pearl (2001). This framework provides a clear foundation for defining direct and indirect effects, a means for dealing with treatment and mediator interactions, and a unified approach for analyzing outcomes and mediators that can be continuous or binary.

In addition to estimating causal mediated effects and their percentage contributions to the total effect, you can use the CAUSALMED procedure to perform various two- and three-way decompositions and a four-way decomposition of the total effect (VanderWeele 2014). For data with binary outcomes, you can compute effects on the odds ratio and excess relative risk scales.

In general, causal mediation effects are defined within levels of the covariates. Some effects are also defined at specific levels of the mediator and treatment variables. The CAUSALMED procedure enables you to set the levels of these variables and estimate the causal effects that answer specific research questions. To make statistical inferences about causal mediation effects, the procedure computes standard errors by asymptotic and bootstrap methods.

Highlights of Enhancements

Following are some highlights of the enhancements in SAS/STAT 14.3:

- The GAMPL procedure now supports the Tweedie distribution.
- In PROC FREQ, the COMMONRISKDIFF option in the TABLES statement provides estimates, confidence limits, and tests for the overall risk (proportion) difference for multiway 2×2 tables.
- The IRT procedure now supports the nominal response model, which enables you to do item analysis of nominal responses.
- The NLMIXED and MCMC procedures add a CMPTMODEL statement that fits compartment models in pharmacokinetic analysis.
- The PHREG procedure provides the cause-specific proportional hazards analysis for competing-risks data.
- The QUANTREG and QUANTSELECT procedures provide fast quantile process regression.
- The VARMETHOD=BOOTSTRAP option provides variance estimation by the bootstrap method for the survey data analysis procedures.
- The TTEST procedure provides bootstrap standard error, bias estimates, and confidence limits.

For more information about the changes and enhancements, see the section "Enhancements" on page 4. Details can be found in the documentation for the individual procedures in *SAS/STAT User's Guide*.

Highlights of Enhancements in SAS/STAT 14.2

Some users might be unfamiliar with updates that were made in the previous release. SAS/STAT 14.2 introduced the CAUSALTRT and PSMATCH procedures. Following are highlights of the other enhancements in SAS/STAT 14.2:

- The FREQ and SURVEYFREQ procedures provide additional agreement
- The NLIN procedure provides ESTIMATE and CONTRAST statements.
- The NLMIXED procedure supports multithreading in models that have more than one RANDOM statement.
- The PHREG procedure provides time-dependent ROC analysis.
- The POWER procedure provides extensions of existing power analyses that are applicable to various generalized linear models.
- The SURVEYIMPUTE procedure provides two-stage fully efficient fractional imputation and fractional hot-deck imputation.
- The SURVEYSELECT procedure provides balanced bootstrap selection and sequential Poisson selection.

Enhancements

BCHOICE Procedure

The BCHOICE procedure allows varying weights among choice sets for allocation choice experiments to indicate how many times the allocated percentages of a choice set should be counted. The varying weight option does not apply to either the binary or MaxDiff choice response. The BCHOICE procedure also adds an option to create a SAS data set in which to save the estimates of all subject-specific random-effects (individual utilities).

CALIS Procedure

The CALIS procedure implements the "saturated correlates" method of Graham (2003) for the parameterization of auxiliary variables in structural equation modeling with missing data. The auxiliary variable method can reduce bias in missing data situations. It can also make the missing at random (MAR) assumption more plausible for carrying out the full information maximum likelihood (METHOD=FIML) estimation. You can specify the auxiliary variables in the AUXILIARY statement. PROC CALIS supports the use of auxiliary variables only in single-group analysis with one of the following model types: LINEQS, MSTRUCT, PATH, or RAM.

The COVDIAG option in the PROC CALIS statement outputs tables that count the occurrences of problematic variance and covariance estimates, which include the cases of zero or negative variance estimates and of zero or negative eigenvalues in predicted covariance matrices.

FACTOR Procedure

The PARALLEL option in the PROC FACTOR statement enables you to perform a parallel analysis (Horn 1965; Glorfeld 1995) to suggest the number of factors to retain. The MAP option in the PROC FACTOR statement enables you to perform a minimum average partial correlation analysis (Velicer 1976; Velicer, Eaton, and Fava 2000) to suggest the number of factors to retain. You can specify NFACTORS=PARALLEL or NFACTORS=MAP in the PROC FACTOR statement to extract the suggested number of factors.

The FACTOR procedure now provides multithreading for improved computational efficiency for the parallel analysis and for large factor rotation problems. You can use the NTHREADS= option in the FACTOR statement to control the multithreading behavior.

FREQ Procedure

The COMMONRISKDIFF option in the TABLES statement provides estimates, confidence limits, and tests for the overall risk (proportion) difference for multiway 2×2 tables. New statistics for the common risk difference include minimum risk confidence limits, the minimum risk test (Mehrotra and Railkar 2000), and stratified Newcombe confidence limits that are computed by using minimum risk weights.

The new METHOD=SCORE2 option is available for exact confidence limits for the risk difference and the relative risk (which you can request by specifying the RISKDIFF or RELRISK option, respectively, in the EXACT statement). METHOD=SCORE2 computes exact confidence limits by inverting a single two-sided exact test that is based on a score statistic (Agresti and Min 2001).

The EXACT SYMMETRY option provides an exact version of Bowker's symmetry test.

The "Kappa Details" table now includes the B_n measure (Bangdiwala 1988; Bangdiwala et al. 2008).

The POLYCHORIC(ADJUST) option adjusts zero-frequency table cells in the polychoric correlation computation.

GAMPL Procedure

The DIST=TWEEDIE option in the MODEL statement enables you to specify the Tweedie distribution.

GEE Procedure

The TYPE3 option in the MODEL statement enables you to test the hypothesis that the Type 3 contrast for a main effect is equal to 0. The default Type 3 analysis is to compute generalized score statistics for the contrasts. The WALD option in the MODEL statement enables you to request a Type 3 analysis that uses Wald statistics.

GLIMMIX Procedure

The MIXED and GLIMMIX procedures now include the TYPE=SP(LEAR) covariance structure for modeling a linear exponent autoregressive covariance, as proposed by Simpson et al. (2010). This can be viewed as an extension of TYPE=ARMA(1,1) to unequally-spaced repeated measurements, in the same way that TYPE=SP(EXP) and TYPE=SP(POW) extend TYPE=AR(1).

HPLOGISTIC Procedure

The STB option in the MODEL statement displays the standardized parameter estimates in the "Parameter Estimates" table.

HPMIXED Procedure

The HPMIXED procedure now includes the TYPE=TOEP(1) covariance structure, which specifies a Toeplitz structure with one band. This can be useful for specifying the same variance component for several effects.

IRT Procedure

The IRT procedure now supports the nominal response model, which enables you to do item analysis of nominal responses. You can specify the nominal response model by using the RESFUNC=NOMINAL option in the PROC IRT statement or in the MODEL statement.

LOGISTIC Procedure

The LACKFIT option in the MODEL statement now performs the Hosmer-Lemeshow test for polytomous response models and provides more suboptions for controlling the test.

The FIRTH option in the MODEL statement performs penalized-likelihood optimization for binary response models for all link functions.

MCMC Procedure

The CMPTMODEL statement fits compartment models in pharmacokinetic analysis.

You can now model missing responses in the presence of missing covariates.

MIXED Procedure

The MIXED and GLIMMIX procedures now include the TYPE=SP(LEAR) covariance structure for modeling a linear exponent autoregressive covariance, as proposed by Simpson et al. (2010). This can be viewed as an extension of TYPE=ARMA(1,1) to unequally spaced repeated measurements, in the same way that TYPE=SP(EXP) and TYPE=SP(POW) extend TYPE=AR(1).

NLMIXED Procedure

The CMPTMODEL statement fits compartment models in pharmacokinetic analysis.

PHREG Procedure

The EVENTCODE(COX)= option in the MODEL statement provides the cause-specific analysis of competing-risks data by applying the Cox model to the cause-specific hazard for each event type separately. Estimates of the cumulative incidence function are based on a synthesis of all failure causes and can be obtained in the BASELINE statement.

The restriction that you can only use the TIMELIST= option (BASELINE statement) for a Bayesian analysis has been lifted. You can use this option to obtain prediction statistics such as the survivor function at specified time points.

PSMATCH Procedure

The PSMATCH procedure now provides variable assessment after stratification, and the new STRATUMWGT= option specifies the stratum weights to use to combine stratum-specific statistics.

The NLARGESTWGT= option displays a table of the observations that have the most extreme IPTW-ATE or ATT weights in the treated and control groups. The NMATCHMOST= option displays a table of the observations that have the greatest numbers of matches in the treated and control groups. Also, you can use the ID statement to specify one or more variables whose values identify the observations in these tables.

The new PLOTS=CDFPLOT option displays cumulative distribution function (CDF) plots for continuous variables in the treated and control groups for all observations and for observations in the support region. If you specify the MATCH statement, CDF plots are also created for matched observations. The new PLOTS=WGTCLOUDPLOT option displays cloud plots for IPTW-ATE or ATT weights in the treated and control groups for all observations and for observations in the support region.

QUANTREG Procedure

The QUANTREG procedure now supports the fast quantile process regression method of (Yao 2017). You can specify this analysis method by using the QUANTILE=FQPR option in the MODEL statement. The fast quantile process regression method can efficiently fit quantile regression models for a specified equally spaced quantile-level grid.

QUANTSELECT Procedure

The QUANTSELECT procedure now supports the fast quantile process regression method, as proposed by (Yao 2017). You can specify this analysis method by using the QUANTILE=FQPR option in the MODEL statement. You can use the QUANTILE=FQPR method with forward, backward, and stepwise model selection.

SURVEYFREQ Procedure

The VARMETHOD=BOOTSTRAP option in the PROC SURVEYFREQ statement provides variance estimation by the bootstrap method. You can specify bootstrap replicate weights for the analysis, or the procedure can construct bootstrap replicate weights.

When you specify a multiway table request, you can now select which two-way subtables (domains) to display by specifying level values in square brackets after the layer variable names.

The new DOMAIN=ROW option in the TABLES statement displays a separate one-way table for each row variable level (domain) in a two-way or multiway table request. This option enables you to produce one-way chi-square tests for the row-level domains by also specifying the CHISQ option.

PROC SURVEYFREQ now provides the AC1 agreement coefficient and the prevalence-adjusted bias-adjusted kappa coefficient (PABAK). The “Kappa Details” table now includes the B_n measure (Bangdiwala 1988; Bangdiwala et al. 2008).

The new NEGATIVE option in the REPWEIGHTS statement removes the restriction that the replicate weight values must be nonnegative. When you specify this option, PROC SURVEYFREQ accepts negative replicate weight values and includes them in the analysis.

SURVEYIMPUTE Procedure

PROC SURVEYIMPUTE now computes bootstrap replicate weights. The VARMETHOD=BOOTSTRAP option in the PROC SURVEYIMPUTE statement creates bootstrap replicate weights that are appropriate for complex surveys involving stratification, clustering, and unequal weights. In addition, if you use the METHOD=FEFI or METHOD=FHDI option in the PROC SURVEYIMPUTE statement then the bootstrap weights are further adjusted for imputation.

The new DISP=MEAN | SSCP option for METHOD=FHDI displays the weighted mean or the crossproduct of the weighted mean sum of squares for variables that are specified in the VAR statement but not in the CLASS statement.

SURVEYLOGISTIC Procedure

The VARMETHOD=BOOTSTRAP option in the PROC SURVEYLOGISTIC statement provides variance estimation by the bootstrap method.

You can request that PROC SURVEYLOGISTIC display specific domain levels in the output by listing quoted formatted-level values in parentheses after each variable name in the DOMAIN statement.

SURVEYMEANS Procedure

The VARMETHOD=BOOTSTRAP option in the PROC SURVEYMEANS statement provides variance estimation by the bootstrap method.

You can specify the NEGATIVE option in the REPWEIGHTS statement to remove the restriction that the REPWEIGHTS variable values must be nonnegative. When you specify this option, PROC SURVEYMEANS accepts negative replicate weight values and includes them in the analysis.

SURVEYPHREG Procedure

PROC SURVEYPHREG now allows two time variables in the MODEL statement for the counting process style of input. This new form of the MODEL statement identifies a pair of failure time variables. Their names are enclosed in parentheses, and they signify the endpoints of a semiclosed interval $(t1, t2]$ during which the subject is at risk.

The new HAZARDRATIO statement enables you to request hazard ratios for any variable in the model at customized settings.

The VARMETHOD=BOOTSTRAP option in the PROC SURVEYPHREG statement provides variance estimation by the bootstrap method.

The ATRISK option in the PROC SURVEYPHREG statement displays a table that contains the sum of weights for the number of units and the sum of weights for the corresponding number of events in the risk sets.

The DOMAIN statement now allows domain analyses for specified levels of a DOMAIN variable by listing formatted-level values in parentheses after the variable name.

The VARRATIO=SRSWOR | SRSWR option in the MODEL statement displays the generalized design effect matrix for estimating the regression parameters. PROC SURVEYPHREG now computes the design-adjusted likelihood-ratio test for the full model.

SURVEYREG Procedure

The VARMETHOD=BOOTSTRAP option in the PROC SURVEYREG statement provides variance estimation by the bootstrap method.

You can request that PROC SURVEYREG display specific domain levels in the output by listing quoted formatted-level values in parentheses after each variable name in the DOMAIN statement.

SURVEYSELECT Procedure

The OUTORDER=RANDOM option in the PROC SURVEYSELECT statement randomly orders the selected observations in the output data set. By default for most sample selection methods, the order of the observations in the output data set corresponds to the order in the input data set.

The new REPNAME= option for replicate sampling (which you can request by specifying the REPS= option in the PROC SURVEYSELECT statement) names the replicate identification variable in the output data set; by default, the variable is named Replicate.

TTEST Procedure

The new BOOTSTRAP statement in the TTEST procedure provides bootstrap standard error, bias estimates, and confidence limits for means and standard deviations in one-sample, paired, and two-sample designs. The following bootstrap confidence intervals are available:

- bias-corrected percentile intervals
- bootstrap *t* intervals, which use a traditional standard error estimate and quantiles of the bootstrap distribution of the *t* statistic
- percentile-based confidence intervals that include a narrowness bias adjustment
- uncorrected percentile-based confidence intervals
- *t*-based confidence intervals that use the bootstrap standard error estimate
- normal-based confidence intervals that use the bootstrap standard error estimate

In addition to displaying the standard errors of the within-class means and the pooled standard error of the mean difference, the TTEST procedure now also displays the unpooled (Satterthwaite) standard error of the mean class difference.

What's Changed

The following sections describe changes in software behavior from SAS/STAT 14.2 to SAS/STAT 14.3.

FREQ Procedure

The default method for computing exact confidence limits for the risk difference and the relative risk has changed. Beginning in SAS/STAT 14.3, the default is METHOD=SCORE, which inverts two separate one-sided exact tests that are based on the score statistic (Chan and Zhang 1999). In releases before SAS/STAT 14.3, the default method is based on the unstandardized statistic (risk difference or relative risk), which you can now request by specifying RISKDIFF(METHOD=NOSCORE) or RELRISK(METHOD=NOSCORE), respectively, in the EXACT statement.

Beginning in SAS/STAT 14.3, PROC FREQ displays all agreement tables (which are produced by the AGREE option in the TABLES statement) in tabular format. You can specify the AGREE(TABLES=RESTORE) option to display these tables in factoid (label-value) format, which is their format in releases before SAS/STAT 14.2.

GLM Procedure

In PROC GLM and most other procedures that compute LS-means, the method of representing mean comparisons by connecting groups using lines to indicate insignificant differences is now displayed graphically instead of as a table. This makes comparisons between a large number of groups much easier to interpret.

Market Research Application

The Market Research Application is no longer documented. Beginning with the next release of SAS/STAT software, it will no longer be available.

PSMATCH Procedure

When you use matching with replacement and k control units are matched with each treated unit where $k > 1$, you now specify a single MATCHID= variable in the OUTPUT statement to identify the mutually exclusive matched sets of treated and control units in the output data set. In SAS/STAT 14.2, you specify k MATCHID= variables, and instead of identifying the matched sets, these variables provide identification numbers for the k control units that are matched with each treated unit.

The Propensity Score Information table now displays propensity score mean differences between the treated and control groups for the sets of all observations, support region observations, and matched observations. The Variable Information and Strata Variable Information tables now display variable mean differences between the treated and control groups for these same sets.

The Standardized Mean Differences table has been rearranged. In SAS/STAT 14.2, statistics such as the mean difference for different sets of observations are displayed in the same row for each variable. In SAS/STAT 14.3, these statistics are displayed in multiple rows, one for each set of observations.

References

Agresti, A., and Min, Y. (2001). "On Small-Sample Confidence Intervals for Parameters in Discrete Distributions." *Biometrics* 57:963–971.

Bangdiwala, S. I. (1988). *The Agreement Chart*. Technical report, Department of Biostatistics, University of North Carolina at Chapel Hill.

Bangdiwala, S. I., Haedo, A. S., Natal, M. L., and Villaveces, A. (2008). "The Agreement Chart as an Alternative to the Receiver-Operating Characteristic Curve for Diagnostic Tests." *Journal of Clinical Epidemiology* 61:866–874.

Chan, I. S. F., and Zhang, Z. (1999). "Test-Based Exact Confidence Intervals for the Difference of Two Binomial Proportions." *Biometrics* 55:1202–1209.

Glorfeld, L. W. (1995). "An Improvement on Horn's Parallel Analysis Methodology for Selecting the Correct Number of Factors to Retain." *Educational and Psychological Measurement* 55:377–393.

Graham, J. W. (2003). "Adding Missing-Data-Relevant Variables to FIML-Based Structural Equation Models." *Structural Equation Modeling* 10:80–100.

Horn, J. L. (1965). "A Rationale and Test for the Number of Factors in Factor Analysis." *Psychometrika* 30:179–185.

Mehrotra, D. V., and Railkar, R. (2000). "Minimum Risk Weights for Comparing Treatments in Stratified Binomial Trials." *Statistics in Medicine* 19:811–825.

Pearl, J. (2001). "Direct and Indirect Effects." In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, edited by J. Breese, and D. Koller, 411–420. San Francisco: Morgan Kaufmann.

Robins, J. M., and Greenland, S. (1992). "Identifiability and Exchangeability for Direct and Indirect Effects." *Epidemiology* 3:143–155.

Simpson, S. L., Edwards, L. J., Muller, K. E., Sen, P. K., and Styner, M. A. (2010). "A Linear Exponent AR(1) Family of Correlation Structures." *Statistics in Medicine* 29:1825–1838.

Valeri, L., and VanderWeele, T. J. (2013). "Mediation Analysis Allowing for Exposure-Mediator Interactions and Causal Interpretation: Theoretical Assumptions and Implementation with SAS and SPSS Macros." *Psychological Methods* 18:137–150.

VanderWeele, T. J. (2014). "A Unification of Mediation and Interaction: A 4-Way Decomposition." *Epidemiology* 25:749–761.

Velicer, W. F. (1976). "Determining the Number of Components from the Matrix of Partial Correlations." *Psychometrika* 41:321–327.

Velicer, W. F., Eaton, C. A., and Fava, J. L. (2000). "Construct Explication through Factor or Component Analysis: A Review and Evaluation of Alternative Procedures for Determining the Number of Factors or Components." In *Problems and Solutions in Human Assessment: Honoring Douglas N. Jackson at Seventy*, edited by R. D. Goffin, and E. Helmes, 41–71. Boston: Kluwer.

Yao, Y. (2017). "Fast Quantile Process Regression." Paper presented at the 2017 International Conference on Robust Statistics, Wollongong, NSW, Australia.