Changes and Enhancements

Changes and Enhancements for All Operating Environments

This section describes the features of base SAS procedures that have been implemented or enhanced since Release 6.12. Version 8 changes and enhancements are preceded by \*8\. All other changes and enhancements described were included in Version 7. If you use SAS software in an OS/390, CMS, or OpenVMS VAX environment, then also see “Additional Version 7 Changes and Enhancements for OS/390, CMS, and OpenVMS VAX” on page xxiv.

For information about changes and enhancements to base SAS procedures that are relevant only to a particular operating environment, see the SAS documentation for that operating environment.

Changes That Affect Multiple Procedures

Output Delivery System

The Output Delivery System (ODS) is a new feature that enhances your ability to manage procedure output. Procedures that fully support ODS

- store links to each piece of output in the Results folder in the SAS Explorer
- can generate HTML output
- can generate output for a high-resolution printer
- can generate output data sets from procedure output
- provide a way for you to customize the procedure output by creating templates that you can use whenever you run the procedure.

For more information on ODS, see “Output Delivery System” on page 19.

Integrity Constraints

Integrity constraints are a set of rules that modifications to data sets (tables) must follow in order to guarantee the validity of the data. You can create integrity constraints with the DATASETS and SQL procedures. Integrity constraints have
Implications for other procedures, such as PROC SORT and PROC CPORT. (See the information on individual procedures.)

**Generation Data Sets**

Generation data sets enable you to keep historical versions of SAS data sets, SAS data views, SAS catalogs, and SAS/ACCESS files. In addition, a historical record of changes is provided so that you can audit the changes that have been made to these files.

PROC DATASETS manages the generation group — the set of generation data sets (see Chapter 14, “The DATASETS Procedure,” on page 329). Other procedures can use the GENNUM= data set option to read a specific version of the data.

**Changes to the Behavior of the WEIGHT Statement**

Prior to Version 7 of the SAS System, all base procedures except PROC REPORT used a value of 0 for missing weights. PROC REPORT and most SAS/STAT procedures, such as PROC GLM, excluded observations with missing weight from the analysis. Now all procedures exclude observations with missing weights from the analysis.

PROC REPORT and most SAS/STAT procedures have always excluded not only observations with missing weights but also observations with negative and zero weights from the analysis. Now, base procedures that do not, by default, exclude observations with negative and 0 weights support the EXCLNPWGT option in the PROC statement. EXCLNPWGT excludes observations with negative and 0 weights.

For more information, see “WEIGHT” on page 73.

**New Procedures**

- The DBCSTAB procedure produces conversion tables for the double-byte character sets that SAS supports.
- The FSLIST procedure on page 605 is now a part of base SAS software. You no longer need a SAS/FSP license to run the procedure.
- The EXPORT procedure on page 423 reads data from a SAS data set and writes it to an external data source.
- The IMPORT procedure on page 611 reads data from an external data source and writes it to a SAS data set.
- The REGISTRY procedure on page 849 maintains the SAS Registry.

**Changes to Existing Procedures**

**PROC CATALOG**

- The PROC CATALOG statement supports this new option:

  ```plaintext
  FORCE
  ```

  forces statements to execute when the catalog has been opened by a process other than the current one.

- The COPY statement on page 160 supports the NEW option, which overwrites the destination catalog (specified by OUT=) if it already exists.
PROC CIMPORT
- The CIMPORT procedure transports integrity constraints, passwords, and generation data sets.
- The PROC CIMPORT on page 212 statement supports these new options:
  - **MEMTYPE=** specifies to move only data sets or only catalogs during a library import.
  - **NEW** creates a new catalog for the imported transport file, and deletes any existing catalog with the same name.
- The SELECT statement on page 216 and the EXCLUDE statement on page 214 replace the SELECT= option and the EXCLUDE= option.

PROC CORR
- **v8** The WEIGHT statement now excludes observations with missing weight values from the analysis.
- **v8** The EXCLNPWGT option in the PROC CORR statement excludes observations with nonpositive weights from the analysis.

PROC CPORT
- The procedure transports integrity constraints, passwords, and generation data sets.
- The SELECT statement on page 319 and the EXCLUDE statement on page 318 replace the SELECT= option and the EXCLUDE= option.

PROC DATASETS
The SAS System now supports the creation and maintenance of generations of SAS data sets. A generation group consists of the base version of the file and a set of historical versions. You can use PROC DATASETS to manage generation data sets.

The SAS System now supports the creation of integrity constraints. Integrity constraints are a set of rules that modifications to a data set (table) must follow in order to guarantee the validity of the data. You can use PROC DATASETS to create and delete integrity constraints.

For more information on using PROC DATASETS with generation data sets and integrity constraints, see Chapter 14, “The DATASETS Procedure,” on page 329.
- PROC DATASETS supports these new statements:
  - **v8** **AUDIT**“AUDIT Statement” on page 343
    - initiates and controls event logging to an audit file.
  - **IC CREATE**“IC CREATE Statement” on page 360
    - creates an integrity constraint.
  - **IC DELETE**“IC DELETE Statement” on page 361
    - deletes an integrity constraint.
  - **IC REACTIVATE**“IC REACTIVATE Statement” on page 362
    - reactivates a foreign key integrity constraint that has been set to inactive.
  - **INDEX CENTILES**“INDEX CENTILES” on page 362
    - updates centiles information for indexed variables.
- The APPEND statement on page 339 supports this new option:
APPENDVER=V6
uses the Version 6 behavior for appending observations to the BASE= data set.

- The CHANGE statement on page 345 supports this new option:
  GENNUM=
  restricts processing to a single generation file or to the entire generation group.

- The CONTENTS statement on page 346 supports these new options:
  CENTILES
  prints centiles information for indexed variables.
  OUT2=
  names an output data set to contain information about indexes and integrity constraints.

- The COPY statement on page 349 supports this new option:
  CONSTRAINT=
  specifies whether to copy all integrity constraints when copying a data set.

- The DELETE statement on page 354 supports this new option:
  GENNUM=
  restricts processing to the specified generation files

- The INDEX CREATE on page 363 statement supports this new option:
  UPDATECENTILES=
  specifies when the centiles are updated.

- The IC DELETE “IC DELETE Statement” on page 361 statement supports this new option:
  _ALL_
  deletes all constraints.

- The INDEX DELETE “INDEX DELETE Statement” on page 365 statement supports this new option:
  _ALL_
  deletes all indexes, except for indexes that are owned by an integrity constraint.

- The MODIFY statement on page 367 supports these new options:
  GENMAX=
  sets the maximum number of generation files in a generation group.
  GENNUM=
  restricts processing to the specified generation file.

- The REPAIR statement on page 371 supports this new option:
  GENNUM=
  restricts processing to the specified generation files.

**PROC FORMAT**

- The PICTURE statement on page 441 supports these new options:
  DATATYPE=
  specifies that the picture value is a template for formatting date, time, or datetime values.
DECSEP=
   specifies the separator character for the fractional part of a number. By
default, this character is a period (.)

DIG3SEP=
   specifies the character that separates each group of three digits in a number.
By default, this character is a comma (,).

MULTILABEL
   indicates that multiple labels for the same range are allowed. Secondary
   labels can be used by certain applications that are designed to handle
   multilabel formats. For all other applications, the secondary labels are
   ignored.

   When MULTILABEL is specified, overlapping ranges are also permitted.

□ The VALUE statement on page 450 supports this new option:

   MULTILABEL
   indicates that multiple labels for the same range are allowed. Secondary
   labels can be used by certain applications that are designed to handle
   multilabel formats. For all other applications, the secondary labels are
   ignored.

   When MULTILABEL is specified, overlapping ranges are also permitted.

PROC FREQ

□ In PROC FREQ, the method for calculating the standard error for common
   relative risks has been revised.

□ PROC FREQ supports a new statement, the TEST statement on page 523. It
   provides asymptotic tests for some measures of association and measures of
   agreement. The TEST statement supports the following statistics.

   MEASURES
   provides tests for all the measures of association. To select individual
   measures of association, use one or more of these options:

       GAMMA
       KENTB
       PCORR
       SCORR
       SMDCR
       SMDRC
       STUTC

   AGRE€€
   provides tests for all the measures of agreement. To select individual
   measures of agreement, use one or more of these options:

       KAPPA
       WTKAP

□ The TABLES statement on page 514 supports these new options:

   PRINTKWT
   prints the kappa coefficient weights.

   SCOROUT
   displays the row and column scores when statistics are computed two-way
   tables.
BINOMIAL
computes binomial proportions, asymptotic standard error, and asymptotic
confidence bounds for one-way tables.

The AGREE option in the TABLES statement on page 514 supports a
parameter that specifies the type of weights that PROC FREQ uses to compute the
weighted kappa coefficient. You can specify Cicchetti-Allison weights or
Fleiss-Cohen weights.

The EXACT statement on page 507 now computes an exact chi-square
goodness-of-fit test for one-way tables as well as an exact chi-square test for
two-way tables. It also supports these new options:

* ALPHA=
specifies the confidence level for the confidence limits for the Monte Carlo
  p-value estimates.

* MAXTIME=
specifies the amount of time that PROC FREQ uses to compute an exact
  p-value before timing out.

* MC
  requests Monte Carlo estimation of exact p-values, instead of direct exact
  p-value computation.

* N=
specifies the number of samples for Monte Carlo estimation.

* SEED=
specifies the initial seed for random number generation for Monte Carlo
  estimation.

PROC MEANS

The PROC MEANS statement on page 627 supports these new options:

* TYPES
  specifies which combinations of class variables PROC MEANS uses to
  subgroup the data (see “TYPES Statement” on page 646).

* WAYS
  specifies how many class variables PROC MEANS combines to subgroup the
  data (see “WAYS Statement” on page 647).

The PROC MEANS statement on page 627 supports these new options:

* CHARTYPE
  specifies that the _TYPE_ variable contains character values.

* CLASSDATA=
specifies a data set that contains the combinations of class variable values to
  include in analysis.

* COMPLETETYPES
  creates all possible combinations of class variable values.

* EXCLNPWGT
  excludes observations with nonpositive weights from the analysis.

* EXCLUSIVE
  excludes from the analysis all class variable combinations that are not in the
  CLASSDATA= data set.
Changes and Enhancements

NOTRAP
  disables floating point exception (FPE) recovery during data processing.

PRINTALLTYPES
  displays all valid combinations of class variables in the output.

PRINTIDVARS
  prints the value of the ID variables.

QMARKERS=
  specifies the default number of markers to use for the $P_{\alpha}$ quantile estimation method.

QMETHOD=
  specifies the method to process the input data to compute quantiles.

QNTLDEF=
  specifies the mathematical definition used to compute quantiles.

SUMSIZE=
  specifies the amount of memory available for data summarization with class variables.

- PROC MEANS now supports multiple CLASS statements. The CLASS statement (only in the MEANS, SUMMARY, and TABULATE procedures) supports the following new options:

  ASCENDING
    specifies to sort the class variable levels in ascending order.

  DESCENDING
    specifies to sort the class variable levels in descending order.

  EXCLUSIVE
    excludes from the analysis all class variable values that are not found in the preloaded range of user-defined formats.

  GROUPINTERNAL
    specifies not to apply formats to the class variables when PROC MEANS sorts the values to create combinations of class variables.

  MISSING
    considers missing values as valid class variable levels.

  MLF
    enables PROC MEANS to use the primary and secondary format labels for a given range or overlapping ranges to create the subgroup combinations when a multilabel format is assigned to a class variable.

  ORDER=
    specifies the sort order for the levels of the class variables in the output.

  PRELOADFMT
    specifies to preload all the formats for the class variables.

- The OUTPUT statement on page 640: by default the statistics in the output data set automatically inherit the analysis variable’s format, informat, and label. However, statistics computed for N, NMISS, SUMWGT, USS, CSS, VAR, CV, T, PROBT, SKEWNESS, and KURTOSIS will not inherit the analysis variable’s format because this format may be invalid for these statistics (e.g. dollar or datetime formats).

  The OUTPUT statement supports these new options:
AUTOLABEL
specifies that PROC MEANS append the statistic name to the end of the variable label. If an analysis variable has no label, PROC MEANS creates a label by appending the statistic name to the analysis variable name.

AUTONAME
automatically resolves conflicts in the names of the variables in the OUT= data set.

IDGROUP
combines the features of the ID statement and the IDMIN option in the PROC MEANS statement.

INHERIT
specifies that statistics in the output data set inherit the attributes (label, length, and format) of the analysis variable that PROC MEANS uses to derive them.

KEEPLEN
specifies that statistics in the output data set inherit the length of the analysis variable that PROC MEANS uses to derive them.

LEVELS
includes a variable named _LEVEL_ in the output data set. This variable contains a value from 1 to n that indicates a unique combination of the values of class variables (the values of the _TYPE_ variable).

NOINHERIT
specifies that the variables in the output data set that contain statistics do not inherit the attributes (label and format) of the analysis variables which are used to derive them.

WAYS
includes a variable named _WAY_ in the output data set. This variable contains a value from 1 to the maximum number of CLASS variables that indicates how many CLASS variables PROC MEANS combines to create the TYPE value.

The VAR statement on page 646 supports this new option:

WEIGHT=
specifies a numeric variable whose values weight the variables that are specified in the VAR statement.

The WEIGHT statement on page 73 now excludes observations with missing weight values from the analysis.

PROC MEANS supports these new statistics:

MEDIAN
P1
P5
P10
P90
P95
P99
Q1
Q3
QRANGE
PROC OPTIONS

The OPTION= option in the PROC OPTIONS statement on page 687 supports two new suboptions, DEFINE and VALUE, that provide additional information about the specified option.

PROC PMENU

PROC PMENU supports submenus to enable multiple items to point to a common submenu.

PROC PRINT

- The PROC PRINT statement on page 786 supports this new option:
  OBS=
  specifies a column header for the column that identifies each observation by number.
- The N= option in the PROC PRINT statement on page 786 has been enhanced so that you can specify explanatory text to print with the value of N.

PROC REPORT

- The PROC REPORT statement on page 879 supports this new option:
  FORMCHAR=
  defines the characters to use as line-drawing characters in the report.
- The COMPUTE statement on page 905 supports this new option:
  _PAGE_
  places information at the top or bottom of each page.
- If you use the Output Delivery System to create HTML files or printer output from PROC REPORT, you can set the style that the procedure uses for various parts of the report. Styles determine attributes like font face, font weight, color, and so forth. Information on the attributes that you can set for a style is in “Customizing the Style Definition That ODS Uses” on page 42.
  You specify styles for PROC REPORT with the STYLE= option. You can use this option in the following statements:
  - PROC REPORT on page 879
  - BREAK on page 893
  - CALL DEFINE on page 899
  - COMPUTE on page 905(with a location and LINE statements)
  - DEFINE on page 908
  - RBREAK on page 918

PROC SQL

- Some PROC SQL views are now updateable. The view must be based on a single DBMS table or SAS data file and must not contain a join, an ORDER BY clause, or a subquery.
- Whenever possible, PROC SQL passes joins to the DBMS rather than doing the joins itself. This enhances performance.
- You can now store DBMS connection information in a view with the USING LIBNAME clause.
- A new option, DQUOTE=ANSI, enables you to use names that are not normally permissible in PROC SQL.
- A PROC SQL query can now reference up to 32 views or tables. PROC SQL can perform joins on up to 32 tables.
- PROC SQL can now create and update tables that contain integrity constraints. For more information, see Chapter 34, “The SQL Procedure,” on page 1021.

**PROC STANDARD**

- The PROC STANDARD statement supports this new option:
  
  `EXCLNPWGT`
  
  excludes observations with nonpositive weights from the analysis.

- The WEIGHT statement on page 73 now excludes observations with missing weight values from the analysis.

**PROC SUMMARY**

- The PROC SUMMARY statement supports this new option:

  `EXCLNPWGT`

  excludes observations with nonpositive weights from the analysis.

- The WEIGHT statement on page 73 now excludes observations with missing weight values from the analysis.

**PROC TABULATE**

- The PROC TABULATE statement on page 1158 supports these new options:

  `CLASSDATA=`
  
  specifies a data set that contains the combinations of class variable values to include in analysis.

  `V8 CONTENTS=`
  
  allows you to name the link in the HTML table of contents that points to the ODS output of the first table produced.

  `EXCLNPWGT`
  
  excludes observations with nonpositive weights from the analysis.

  `EXCLUSIVE`
  
  excludes from the analysis all class variable combinations that are not in the `CLASSDATA=` data set.

  `NOTRAP`
  
  disables trapping mathematical errors due to overflow.

  `OUT=`
  
  names the output data set.

  `QMARKERS=`
  
  specifies the default number of markers to use for the $P^2$ (fixed space) quantile estimation method.

  `QMETHOD`
  
  specifies the method to process the input data to compute quantiles.

  `QNTLDEF=`
  
  specifies the mathematical definition used to compute quantiles.
TRAP enables trapping mathematical errors due to overflow.

PROC TABULATE now supports multiple CLASS statements. For a discussion of the options that the CLASS statement supports, see the discussion of PROC MEANS and the CLASS statement on page xvii. The CLASS statement has this new option:

MLF allows you to make use of multiple labels when a multilabel format is assigned to a class variable in PROC FORMAT.

In the TABLE statement, the following options have been enhanced:

CONDENSE prints multiple logical pages on a physical page.

CONTENTS= allows you to name the link in the HTML table of contents that points to the ODS output of the table produced using the TABLE statement.

NOCONTINUED suppresses the printing of the "(Continued)" continuation message for tables that span physical pages.

PROC TABULATE supports these new statistics:

COLPCNT
COLPCTSUM
MEDIAN
P1
P5
P10
P90
P95
P99
PAGEPCTN
PAGEPCTSUM
Q1
Q3
QRANGE
REPPCTN
REPPCTSUM
ROWPCTN
ROWPCTSUM

If you use the Output Delivery System to create HTML files or printer output from PROC TABULATE, you can set the style that the procedure uses for various parts of the report. Styles determine attributes like font face, font weight, color, and so forth. Information on the attributes that you can set for a style is in “Customizing the Style Definition That ODS Uses” on page 42.

You specify styles for PROC TABULATE with the STYLE= option. You can use this option in several locations in the procedure. For details see Chapter 37, “The TABULATE Procedure,” on page 1151. In addition, there are two new styles:
PROC TABULATE supports multiple VAR statements. The VAR statement on page 646 supports this new option:

```
WEIGHT=
```

specifies a numeric variable whose values weight the variables that are specified in the VAR statement.

The WEIGHT statement on page 73 now excludes observations with missing weight values from the analysis.

**PROC UNIVARIATE**

PROC UNIVARIATE now supports high resolution graphical displays. You can generate histograms and comparative histograms and optionally superimpose fitted probability density curves for various distributions and kernel density estimates. You can generate quantile-quantile plots (Q-Q plots) and probability plots to compare a data distribution with various theoretical distributions. You also have the ability to inset summary statistics in the graphical displays.

The WEIGHT statement on page 73 now excludes observations with missing weight values from the analysis.

The output from PROC UNIVARIATE has been reorganized and includes some new tables:

- The Moments table displays only those statistics that are related to sample moments.
- A new table, Tests for Location, shows the Student’s $t$ test, the sign test, and the signed rank test.
- A new table, Basic Statistical Measures, provides basic measures of location and variability.

The UNIVARIATE procedure supports these new statements:

**CLASS statement**

specifies one or two class variables for the analysis.

You can use the CLASS statement with a HISTOGRAM, PROBPLOT, or QQPLOT statement to create one-way and two-way high resolution comparative plots. When you use a single class variable, PROC UNIVARIATE displays an array of component plots (stacked or side-by-side), for each level of the class variable. When you use two class variables, PROC UNIVARIATE displays a matrix of component plots, one for each combination of levels of the class variables.

**HISTOGRAM statement**

creates a high resolution graph of a histogram and optionally includes parametric and nonparametric density curve estimates. You can use the HISTOGRAM statement to specify the midpoints for histogram intervals, display density curves for fitted theoretical distributions (beta, exponential, gamma, lognormal, normal, and Weibull) on histograms, request goodness-of-fit tests for fitted distributions, display kernel density estimates on histograms, save histogram intervals and parameters of fitted distributions in output data sets, and request graphical enhancements.

**INSET statement**

Places a box or table of summary statistics, called an *inset*, directly in the graphical display. The inset can display statistics that PROC UNIVARIATE
calculates or display values that you provide in a SAS data set. The INSET statement does not produce the graphical display. You must specify a HISTOGRAM, PROBPLOT, or QQPLOT statement. You can use options in the INSET statement to specify the position of the inset, to specify a header for the inset, and to specify graphical enhancements, such as background colors, text colors, text height, text font, and drop shadows.

**PROBPLOT statement**
Creates a high-resolution graphics display of a probability plot, which compares ordered variable values with the percentiles of a specified theoretical distribution.

**QQPLOT statement**
Creates a graphical display of a quantile-quantile plot (Q-Q plot), which compares ordered variable values with quantiles of a specified theoretical distribution.

- The PROC UNIVARIATE statement on page 1325 statement supports these new options:
  - **ANNOTATE=** specifies an input data set that contains annotate variables as described in SAS/GRAPH documentation. You can use this data set to add features to your high resolution graphics. PROC UNIVARIATE adds the features in this data set to every high resolution graph produced in the PROC step.
  - **EXCLNPWGT** excludes observations with nonpositive weights from the analysis.
  - **GOUT=** specifies the SAS catalog in which to save the high resolution graphics output that the UNIVARIATE procedure produces.

PROC UNIVARIATE also supports the following enhancements:
- The **ALL** option requests tables of confidence limits, frequency, modes, and extreme values, and plots. For unweighted analysis variables, it also requests location counts, tests for normality, and robust estimators of scale and location.
- The **CIBASIC** option produces the Basic Statistical Measures table. This table provides basic measures of location and variability.
- The **EXCLNPWGT** option excludes observations with nonpositive weights from the analysis.
- The **LOCCOUNT** option produces the Location Counts table. This table shows the numbers of observations greater than, less than, and equal to the value of the new MU0= option. Formerly, this information was in the Moments table. The default value for MU0= is 0.
- Mode is no longer shown for continuous data. If there are multiple modes, the lowest mode is still shown. The **MODES** option in the PROC UNIVARIATE statement displays a table of all modes.
- The **NORMAL** option produces a table of Tests for Normality. These tests include empirical distribution function (EDF) goodness-of-fit tests. The Shapiro-Wilk test is included only if the sample size is less than or equal to 2000. The Kolmogorov test is always included.
- The **ROBUSTSCALE** option produces a table with robust estimates of scale.
- The **TRIM=** option produces a table of trimmed means.
- The **WINSOR=** option produces a table of winsorized means.
A new table, Basic Confidence Limits, provides basic confidence intervals for the mean, the standard deviation, and the variance.

You can request confidence intervals for quantiles with these options:

- **CIPCTLNORMAL** requests confidence intervals that are based on the assumption that the data are normal.
- **CIPCTLDFF** requests confidence intervals that are distribution free.

The Extremes table is replaced by two tables:

- **Extreme Observations table**
  
  Displays the $n$ lowest observations and $n$ highest observations, where $n$ is specified with the new option NEXTROBS= in the PROC UNIVARIATE statement. FREQ and ID variables change the table.

- **Extreme Values table**
  
  Displays the $n$ lowest unique and $n$ highest unique values, where $n$ is specified with the new option NEXTRVAL= in the PROC UNIVARIATE statement.

---

### Additional Version 7 Changes and Enhancements for OS/390, CMS, and OpenVMS VAX

For OS/390 (MVS), CMS, and OpenVMS VAX, the last release of base SAS software was the 6.09 Enhanced Release. Some changes and enhancements that were implemented for the other operating environments in the 6.10, 6.11, and 6.12 releases were not implemented for the OS/390, CMS, or OpenVMS VAX until Version 7. This section describes those additional features for the base procedures.

#### PROC FREQ

- The EXACT statement on page 507 now provides statistical keywords that request exact $p$-values for the simple kappa coefficient, the weight kappa coefficient, and the odds ratio for $2 \times 2$ tables. The new statistic keywords are KAPPA, WTKAP, and OR, respectively.
- The TABLES statement on page 514 supports these new options:
  - **CL**
    
    Requests confidence bounds for measures of association.
  - **RELRISK**
    
    Requests just the relative risk measures for $2 \times 2$ tables.
  - **RISKDIF**
    
    Requests column 1 and column 2 risk (or binomial proportions), risk differences, and confidence bounds for $2 \times 2$ tables.

- **TESTF=**
  
  Requests a chi-square statistic to test for equal or specified frequencies for one-way tables.

- **TESTP=**
  
  Requests a chi-square statistic to test for equal or specified proportions for one-way tables.
PROC REPORT

A new window, the EXPLORE window on page 934, lets you experiment with your data. For example, you can subset your data or suppress the display of a column. If you like the results, you can apply them to the report.