CHAPTER

1

Choosing the Right Procedure

Functional Categories of Base SAS Procedures

Base SAS software provides a variety of procedures that produce reports, compute statistics, and perform utility operations.

Report Writing

These procedures display useful information, such as data listings (detail reports), summary reports, calendars, letters, labels, forms, multipanel reports, and graphical reports:

CALENDAR	MEANS*	SQL*
CHART*	PLOT	SUMMARY*
FORMS	PRINT	TABULATE*
FREQ*	REPORT*	TIMEPLOT

* These procedures produce reports and compute statistics.

Statistics

These procedures compute elementary statistical measures which include descriptive statistics based on moments, quantiles, confidence intervals, frequency counts, cross-tabulations, correlations, and distribution tests. They also rank and standardize data:

CHART	RANK	SUMMARY
CORR	REPORT	TABULATE
FREQ	SQL	UNIVARIATE
MEANS	STANDARD	

Utilities

These procedures perform basic utility operations. They create, edit, sort, and transpose data sets, create and restore transport data sets, create user defined formats, and provide basic file maintenance such as to copy, append, and compare data sets:

APPEND	EXPLODE	REGISTRY
BMDP[**]	EXPORT	RELEASE[**]
CATALOG	FORMAT	SORT
CIMPORT	FSLIST	SOURCE[**]
COMPARE	IMPORT	SQL
CONTENTS	OPTIONS	TAPECOPY[**]
CONVERT[**]	PDS[**]	TAPELABEL[**]
COPY	PDSCOPY[**]	TRANSPOSE
CPORT	PMENU	TRANTAB
DATASETS	PRINTTO	

** See the SAS documentation for your operating environment for a description of these procedures.

Report-Writing Procedures

Table 1.1 on page 4 lists report-writing procedures according to the type of report.

Table 1.1   Report-Writing Procedures by Task

To produce …	Use this procedure …	Which …
Detail reports	PRINT	produces data listings quickly; can supply titles, footnotes, and column sums.
	REPORT	offers more control and customization than PROC PRINT; can produce both column and row sums; has DATA step computation abilities.
	SQL	combines Structured Query Language and SAS features such as formats; can manipulate data and create a SAS data set in the same step that creates the report; can produce column and row statistics; does not offer as much control over output as PROC PRINT and PROC REPORT.

To produce …	Use this procedure …	Which …
Summary reports	PRINT	produces only one summary report: can sum the BY variables.
	REPORT	combines features of the PRINT, MEANS, and TABULATE procedures with features of the DATA step in a single report writing tool that can produce a variety of reports; can also create an output data set.
	TABULATE	produces descriptive statistics in a tabular format; can produce stub-and-banner reports (multidimensional tables with descriptive statistics); can also create an output data set.
	SQL	computes descriptive statistics for one or more SAS data sets or DBMS tables; can produce a printed report or create a SAS data set.
	MEANS or SUMMARY	computes descriptive statistics for numeric variables; can produce a printed report and create an output data set.
Miscellaneous highly formatted reports		
Calendars	CALENDAR	produces schedule and summary calendars; can schedule tasks around nonwork periods and holidays, weekly work schedules, and daily work shifts.
Labels, Forms	FORMS	produces labels, such as mailing and inventory, or other forms that have a repetitive format.
Name/address listings	FORMS	produces multicolumn name and address listings.
Multipanel reports (telephone book listings)	REPORT	produces multipanel reports.
Low-resolution graphical reports*		
	CHART	produces bar charts, histograms, block charts, pie charts, and star charts that display frequencies and other statistics.
	PLOT	produces scatter diagrams that plot one variable against another.
	TIMEPLOT	produces plots of one or more variables over time intervals.

* These reports quickly produce a simple graphical picture of the data. To produce high-resolution graphical reports, use SAS/GRAPH Software.

Statistical Procedures

Table 1.2 on page 6 lists statistical procedures according to task. Table A1.1 on page 1459 lists the most common statistics and the procedures that compute them.

Table 1.2 Elementary Statistical Procedures by Task

To produce …	Use this procedure …	Which …
Descriptive statistics	CORR	computes simple descriptive statistics.
	MEANS or SUMMARY	computes descriptive statistics; can produce printed output and output data sets. By default, PROC MEANS produces printed output and PROC SUMMARY creates an output data set.
	REPORT	computes most of the same statistics as PROC TABULATE; allows customization of format.
	SQL	computes descriptive statistics for data in one or more DBMS tables; can produce a printed report or create a SAS data set.
	TABULATE	produces tabular reports for descriptive statistics; can create an output data set.
	UNIVARIATE	computes the broadest set of descriptive statistics; can create an output data set.
Frequency and cross-tabulation tables	FREQ	produces one-way to n-way tables; reports frequency counts; computes chi-square tests; computes tests and measures of association and agreement for two-way to n-way cross-tabulation tables; can compute exact tests and asymptotic tests; can create output data sets.
	TABULATE	produces one-way and two-way cross-tabulation tables; can create an output data set.
	UNIVARIATE	produces one-way frequency tables.
Correlation analysis	CORR	computes Pearson's, Spearman's, and Kendall's correlations and partial correlations; also computes Hoeffding's D and Cronbach's coefficient alpha.
Distribution analysis	UNIVARIATE	computes tests for location and tests for normality.
	FREQ	computes a test for the binomial proportion for one-way tables; computes a goodness-of-fit test for one-way tables; computes a chi-square test of equal distribution for two-way tables.
Robust estimation	UNIVARIATE	computes robust estimates of scale, trimmed means, and Winsorized means.
Data transformation		
Computing ranks	RANK	computes ranks for one or more numeric variables across the observations of a SAS data set and creates an output data set; can produce normal scores or other rank scores.
Standardizing data	STANDARD	creates an output data set that contains variables that are standardized to a given mean and standard deviation.

Low-resolution graphics[*]

To produce …	Use this procedure …	Which …
	CHART	produces a graphical report that can show one of the following statistics for the chart variable: frequency counts, percentages, cumulative frequencies, cumulative percentages, totals, or averages.
	UNIVARIATE	produces descriptive plots such as stem and leaf, box plot, and normal probability plot.

* To produce high-resolution graphical reports, use SAS/GRAPH Software.

Efficiency Issues

Quantiles

The calculation of quantiles, including the median, requires time proportional to $n\log(n)$ for large sample sizes. Therefore, a procedure, like UNIVARIATE, that automatically calculates quantiles may require more time than other data summarization procedures. Furthermore, because data are held in memory, the procedure also requires more storage space to perform the computations. By default, the report procedures PROC MEANS, PROC SUMMARY, and PROC TABULATE require less memory because they do not automatically compute quantiles. These procedures also provide an option to use a new fixed-memory quantiles estimation method that is usually less memory intense. See "Quantiles" on page 653 for more information.

Computing Statistics for Groups of Observations

To compute statistics for several groups of observations, you can use any of the previous procedures with a BY statement to specify BY-group variables. However, BY-group processing requires that you previously sort or index the data set, which for very large data sets may require substantial computer resources. A more efficient way to compute statistics within groups without sorting is to use a CLASS statement with one of the following procedures: MEANS, SUMMARY, or TABULATE.

Additional Information on the Statistical Procedures

Appendix 1, "SAS Elementary Statistics Procedures," on page 1457 lists standard keywords, statistical notation, and formulas for the statistics that base SAS procedures compute frequently. The individual statistical procedures discuss the statistical concepts that are useful to interpret the output of a procedure.

Utility Procedures

Table 1.3 on page 8 groups utility procedures according to task.

Table 1.3 Utility Procedures by Task

To perform these utility tasks…	Use this procedure …	Which …
Supply information	COMPARE	compares the contents of two SAS data sets.
	CONTENTS	describes the contents of a SAS data library or specific library members.
	OPTIONS	lists the current values of all SAS system options.
	SQL	supplies information through dictionary tables on an individual SAS data set as well as all SAS files active in the current SAS session. Dictionary tables can also provide information about macros, titles, indexes, external files, or SAS system options.
Affect printing	EXPLODE	produces oversized text on printed output; can produce displays such as posters, flip charts, and header pages.
	FORMAT	creates user-defined formats to display and print data.
	PRINTTO	routes procedure output to a file, a SAS catalog entry, or a printer; can also redirect the SAS log to a file.
Create, browse, and edit data	FSLIST	browses external files such as files that contain SAS source lines or SAS procedure output.
	SQL	creates SAS data sets using Structured Query Language and SAS features.
Transform data	FORMAT	creates user defined informats to read data and user defined formats to display data.
	SORT	sorts SAS data sets by one or more variables.
	SQL	sorts SAS data sets by one or more variables.
	TRANSPOSE	transforms SAS data sets so that observations become variables and variables become observations.
	TRANTAB	creates, edits, and displays customized translation tables.
Manage SAS files	APPEND	appends one SAS data set to the end of another.
	BMDP*	invokes a BMDP program to analyze data in a SAS data set.
	CATALOG	manages SAS catalog entries.
	CIMPORT	restores a transport sequential file that PROC CPORT creates (usually under another operating environment) to its original form as a SAS catalog, a SAS data set, or a SAS library.
	CONVERT*	converts BMDP system files, OSIRIS system files, and SPSS portable files to SAS data sets.
	COPY	copies a SAS data library or specific members of the library.

To perform these utility tasks…	Use this procedure …	Which …
	CPORT	converts a SAS catalog, a SAS data set, or a SAS library to a transport sequential file that PROC CIMPORT can restore (usually under another operating environment) to its original form.
	DATASETS	manages SAS files.
	EXPORT	reads data from a SAS data set and writes them to an external data source.
	IMPORT	reads data from an external data source and writes them to a SAS data set.
	PDS*	lists, deletes, and renames the members of a partitioned data set.
	PDSCOPY*	copies partitioned data sets from disk to tape, disk to disk, tape to tape, or tape to disk.
	REGISTRY	imports registry information to the USER portion of the SAS Registry.
	RELEASE*	releases unused space at the end of a disk data set under the OS/390 environment.
	SOURCE*	provides an easy way to back up and process source library data sets.
	SQL	concatenates SAS data sets.
	TAPECOPY*	copies an entire tape volume or files from one or more tape volumes to one output tape volume.
	TAPELABEL*	lists the label information of an IBM standard-labeled tape volume under the OS/390 environment.
Control windows	PMENU	creates customized pull-down menus for SAS applications.

* See the SAS documentation for your operating environment for a description of these procedures.

Brief Descriptions of Base SAS Procedures

APPEND procedure
adds observations from one SAS data set to the end of another SAS data set.

BMDP procedure
invokes a BMDP program to analyze data in a SAS data set. See the SAS documentation for your operating environment for more information.

CALENDAR procedure
displays data from a SAS data set in a monthly calendar format. PROC CALENDAR can display holidays in the month, schedule tasks, and process data for multiple calendars with work schedules that vary.

CATALOG procedure
manages entries in SAS catalogs. PROC CATALOG is an interactive, nonwindowing procedure that enables you to display the contents of a catalog,

copy an entire catalog or specific entries in a catalog, and rename, exchange, or delete entries in a catalog.

CHART procedure
produces vertical and horizontal bar charts (also called *histograms*), block charts, pie charts, and star charts. These charts provide a quick visual representation of the values of a single variable or several variables. PROC CHART can also display a statistic associated with the values.

CIMPORT procedure
restores a transport file created by the CPORT procedure to its original form (a SAS data library, catalog, or data set) in the format appropriate to the operating environment. Coupled with the CPORT procedure, PROC CIMPORT enables you to move SAS data libraries, catalogs, and data sets from one operating environment to another.

COMPARE procedure
compares the contents of two SAS data sets. You can also use PROC COMPARE to compare the values of different variables within a single data set. PROC COMPARE produces a variety of reports on the comparisons that it performs.

CONTENTS procedure
prints descriptions of the contents of one or more files in a SAS data library.

CONVERT procedure
converts BMDP system files, OSIRIS system files, and SPSS portable files to SAS data sets. See the SAS documentation for your operating environment for more information.

COPY procedure
copies an entire SAS data library or specific members of the library. You can limit processing to specific types of library members.

CORR procedure
computes Pearson product-moment and weighted product-moment correlation coefficients between variables and descriptive statistics for these variables. In addition, PROC CORR can compute three nonparametric measures of association (Spearman's rank-order correlation, Kendall's tau-b, and Hoeffding's measure of dependence, D), partial correlations (Pearson's partial correlation, Spearman's partial rank-order correlation, and Kendall's partial tau-b), and Cronbach's coefficient alpha.

CPORT procedure
writes SAS data libraries, data sets, and catalogs in a special format called a transport file. Coupled with the CIMPORT procedure, PROC CPORT enables you to move SAS libraries, data sets, and catalogs from one operating environment to another.

DATASETS procedure
lists, copies, renames, and deletes SAS files and SAS generation groups, manages indexes, and appends SAS data sets in a SAS data library. The procedure provides all the capabilities of the APPEND, CONTENTS, and COPY procedures. You can also modify variables within data sets, manage data set attributes, such as labels and passwords, or create and delete integrity constraints.

EXPLODE procedure
produces oversized printing of text to generate displays such as posters, flip charts, and header pages.

EXPORT procedure
reads data from a SAS data set and writes it to an external data source.

FORMAT procedure
  creates user-defined informats and formats for character or numeric variables. PROC FORMAT also prints the contents of a format library, creates a control data set to write other informats or formats, and reads a control data set to create informats or formats.

FORMS procedure
  produces labels for envelopes, mailing labels, external tape labels, file cards, and other printer forms that have a regular pattern.

FREQ procedure
  produces one-way to *n*-way frequency tables and reports frequency counts. PROC FREQ can compute chi-square tests for one-way to *n*-way tables, tests and measures of association and of agreement for two-way to *n*-way cross-tabulation tables, risks and risk difference for $2{\times}2$ tables, trends tests, and Cochran-Mantel-Haenszel statistics. You can also create output data sets.

FSLIST procedure
  displays the contents of an external file or copies text from an external file to the SAS Text Editor.

IMPORT procedure
  reads data from an external data source and writes them to a SAS data set.

MEANS procedure
  computes descriptive statistics for numeric variables across all observations and within groups of observations. You can also create an output data set that contains specific statistics and identifies minimum and maximum values for groups of observations.

OPTIONS procedure
  lists the current values of all SAS system options.

PDS procedure
  lists, deletes, and renames the members of a partitioned data set. See the SAS documentation for your operating environment for more information.

PDSCOPY procedure
  copies partitioned data sets from disk to tape, disk to disk, tape to tape, or tape to disk. See the SAS documentation for your operating environment for more information.

PLOT procedure
  produces scatter plots that graph one variable against another. The coordinates of each point on the plot correspond to the two variables' values in one or more observations of the input data set.

PMENU procedure
  defines menus that you can use in DATA step windows, macro windows, and SAS/ AF windows, or in any SAS application that enables you to specify customized menus.

PRINT procedure
  prints the observations in a SAS data set, using all or some of the variables. PROC PRINT can also print totals and subtotals for numeric variables.

PRINTTO procedure
  defines destinations for SAS procedure output and the SAS log.

RANK procedure
  computes ranks for one or more numeric variables across the observations of a SAS data set. The ranks are output to a new SAS data set. Alternatively, PROC RANK produces normal scores or other rank scores.

REGISTRY procedure
imports registry information into the USER portion of the SAS Registry.

RELEASE procedure
releases unused space at the end of a disk data set in the OS/390 environment.
See the SAS documentation for this operating environment for more information.

REPORT procedure
combines features of the PRINT, MEANS, and TABULATE procedures with
features of the DATA step in a single report-writing tool that can produce both
detail and summary reports.

SORT procedure
sorts observations in a SAS data set by one or more variables. PROC SORT stores
the resulting sorted observations in a new SAS data set or replaces the original
data set.

SOURCE procedure
provides an easy way to back up and process source library data sets. See the SAS
documentation for your operating environment for more information.

SQL procedure
implements a subset of the Structured Query Language (SQL) for the SAS System.
SQL is a standardized, widely used language that retrieves and updates data in
SAS data sets, SQL views, and DBMS tables, as well as views based on those
tables. PROC SQL can also create tables and views, summaries, statistics, and
reports and perform utility functions such as sorting and concatenating.

STANDARD procedure
standardizes some or all of the variables in a SAS data set to a given mean and
standard deviation and produces a new SAS data set that contains the
standardized values.

SUMMARY procedure
computes descriptive statistics for the variables in a SAS data across all
observations and within groups of observations and outputs the results to a new
SAS data set.

TABULATE procedure
displays descriptive statistics in tabular form. The value in each table cell is
calculated from the variables and statistics that define the pages, rows, and
columns of the table. The statistic associated with each cell is calculated on values
from all observations in that category. You can output the results to a SAS data set.

TAPECOPY procedure
copies an entire tape volume or files from one or more tape volumes to one output
tape volume. See the SAS documentation for your operating environment for more
information.

TAPELABEL procedure
lists the label information of an IBM standard-labeled tape volume under the OS/
390 environment. See the SAS documentation for this operating environment for
more information.

TIMEPLOT procedure
produces plots of one or more variables over time intervals.

TRANSPOSE procedure
transposes a data set that changes observations into variables and vice versa.

TRANTAB procedure
creates, edits, and displays customized translation tables.

UNIVARIATE procedure
  computes descriptive statistics (including quantiles), confidence intervals, and robust estimates for numeric variables. Provides detail on the distribution of numeric variables, which include tests for normality, plots to illustrate the distribution, frequency tables, and tests of location.

SAS® Procedures Guide, Version 8