Syntax

The following statements are available in PROC REG.

PROC REG < options > **;**
< label: > **MODEL** dependents=<regressors> < / options > **;**
BY variables **;**
FREQ variable **;**
ID variables **;**
VAR variables **;**
WEIGHT variable **;**
ADD variables **;**
DELETE variables **;**
< label: > **MTEST** <equation, … ,equation> < / options > **;**
OUTPUT < **OUT=**SAS-data-set > keyword=names
< … keyword=names > **;**
PAINT <condition | **ALLOBS**>
< / options > | < **STATUS | UNDO**> **;**
PLOT <yvariable*xvariable> <=symbol>
< …yvariable*xvariable> <=symbol> < / options > **;**
PRINT < options > < **ANOVA** > < **MODELDATA** > **;**
REFIT;
RESTRICT equation, … ,equation **;**
REWEIGHT <condition | **ALLOBS**>
< / options > | < **STATUS | UNDO**> **;**
< label: > **TEST** equation,<, …,equation> < / option > **;**

Although there are numerous statements and options available in PROC REG, many analyses use only a few of them. Often you can find the features you need by looking at an example or by scanning this section.

In the preceding list, brackets denote optional specifications, and vertical bars denote a choice of one of the specifications separated by the vertical bars. In all cases, *label* is optional.

The PROC REG statement is required. To fit a model to the data, you must specify the MODEL statement. If you want to use only the options available in the PROC REG statement, you do not need a MODEL statement, but you must use a VAR statement. (See the example in the "OUTSSCP= Data Sets" section.) Several MODEL statements can be used. In addition, several MTEST, OUTPUT, PAINT, PLOT, PRINT, RESTRICT, and TEST statements can follow each MODEL statement. The ADD,

DELETE, and REWEIGHT statements are used interactively to change the regression model and the data used in fitting the model. The ADD, DELETE, MTEST, OUTPUT, PLOT, PRINT, RESTRICT, and TEST statements implicitly refit the model; changes made to the model are reflected in the results from these statements. The REFIT statement is used to refit the model explicitly and is most helpful when it follows PAINT and REWEIGHT statements, which do not refit the model.

The BY, FREQ, ID, VAR, and WEIGHT statements are optionally specified once for the entire PROC step, and they must appear before the first RUN statement.

When TYPE=CORR, TYPE=COV, or TYPE=SSCP data sets are used as input data sets to PROC REG, statements and options that require the original data are not available. Specifically, the OUTPUT, PAINT, PLOT, and REWEIGHT statements and the MODEL and PRINT statement options P, R, CLM, CLI, DW, INFLUENCE, and PARTIAL are disabled.

You can specify the following statements with the REG procedure in addition to the PROC REG statement:

ADD
: adds independent variables to the regression model.

BY
: specifies variables to define subgroups for the analysis.

DELETE
: deletes independent variables from the regression model.

FREQ
: specifies a frequency variable.

ID
: names a variable to identify observations in the tables.

MODEL
: specifies the dependent and independent variables in the regression model, requests a model selection method, displays predicted values, and provides details on the estimates (according to which options are selected).

MTEST
: performs multivariate tests across multiple dependent variables.

OUTPUT
: creates an output data set and names the variables to contain predicted values, residuals, and other diagnostic statistics.

PAINT

paints points in scatter plots.

PLOT
: generates scatter plots.

PRINT
: displays information about the model and can reset options.

REFIT
: refits the model.

RESTRICT
: places linear equality restrictions on the parameter estimates.

REWEIGHT
: excludes specific observations from analysis or changes the weights of observations used.

TEST
: performs an *F* test on linear functions of the parameters.

VAR
: lists variables for which crossproducts are to be computed, variables that can be interactively added to the model, or variables to be used in scatter plots.

WEIGHT
: declares a variable to weight observations.

PROC REG Statement

PROC REG < *options* > **;**

The PROC REG statement is required. If you want to fit a model to the data, you must also use a MODEL statement. If you want to use only the PROC REG options, you do not need a MODEL statement, but you must use a VAR statement. If you do not use a MODEL statement, then the COVOUT and OUTEST= options are not available.

Table 55.1 lists the options you can use with the PROC REG statement. Note that any option specified in the PROC REG statement applies to all MODEL statements.

Table 55.1: PROC REG Statement Options

Option	Description
Data Set Options	
DATA=	names a data set to use for the regression
OUTEST=	outputs a data set that contains parameter estimates and other model fit summary statistics
OUTSSCP=	outputs a data set that contains sums of squares and crossproducts
COVOUT	outputs the covariance matrix for parameter estimates to the OUTEST= data set
EDF	outputs the number of regressors, the error degrees of freedom, and the model R^2 to the OUTEST= data set
OUTSTB	outputs standardized parameter estimates to the OUTEST= data set. Use only with the RIDGE= or PCOMIT= option.
OUTSEB	outputs standard errors of the parameter estimates to the OUTEST= data set
OUTVIF	outputs the variance inflation factors to the OUTEST= data set. Use only with the RIDGE= or PCOMIT= option.
PCOMIT=	performs incomplete principal component analysis and outputs estimates to the OUTEST= data set
PRESS	outputs the PRESS statistic to the OUTEST= data set
RIDGE=	performs ridge regression analysis and outputs estimates to the OUTEST= data set

RSQUARE	same effect as the EDF option
TABLEOUT	outputs standard errors, confidence limits, and associated test statistics of the parameter estimates to the OUTEST= data set
High Resolution Graphics Options	
ANNOTATE=	specifies an annotation data set
GOUT=	specifies the graphics catalog in which graphics output is saved
Display Options	
CORR	displays correlation matrix for variables listed in MODEL and VAR statements
SIMPLE	displays simple statistics for each variable listed in MODEL and VAR statements
USCCP	displays uncorrected sums of squares and crossproducts matrix
ALL	displays all statistics (CORR, SIMPLE, and USSCP)
NOPRINT	suppresses output
LINEPRINTER	creates plots requested as line printer plot
Other Options	
ALPHA=	sets significance value for confidence and prediction intervals and tests
SINGULAR=	sets criterion for checking for singularity

Following are explanations of the options that you can specify in the PROC REG statement (in alphabetical order).

Note that any option specified in the PROC REG statement applies to all MODEL statements.

ALL

requests the display of many tables. Using the ALL option in the PROC REG statement is equivalent to specifying ALL in every MODEL statement. The ALL option also implies the CORR, SIMPLE, and USSCP options.

ALPHA=*number*

sets the significance level used for the construction of confidence intervals. The value must be between 0 and 1; the default value of 0.05 results in 95% intervals. This option affects the PROC REG option TABLEOUT; the MODEL options CLB, CLI, and CLM; the OUTPUT statement keywords LCL, LCLM, UCL, and

UCLM; the PLOT statement keywords LCL., LCLM., UCL., and UCLM.; and the PLOT statement options CONF and PRED.

ANNOTATE=*SAS-data-set*
ANNO= *SAS-data-set*

specifies an input data set containing annotate variables, as described in *SAS/GRAPH Software: Reference*. You can use this data set to add features to plots. Features provided in this data set are applied to all plots produced in the current run of PROC REG. To add features to individual plots, use the ANNOTATE= option in the PLOT statement. This option cannot be used if the LINEPRINTER option is specified.

CORR

displays the correlation matrix for all variables listed in the MODEL or VAR statement.

COVOUT

outputs the covariance matrices for the parameter estimates to the OUTEST= data set. This option is valid only if the OUTEST= option is also specified. See the "OUTEST= Data Set" section.

DATA=*SAS-data-set*

names the SAS data set to be used by PROC REG. The data set can be an ordinary SAS data set or a TYPE=CORR, TYPE=COV, or TYPE=SSCP data set. If one of these special TYPE= data sets is used, the OUTPUT, PAINT, PLOT, and REWEIGHT statements and some options in the MODEL and PRINT statements are not available. See Appendix A, "Special SAS Data Sets," for more information on TYPE= data sets. If the DATA= option is not specified, PROC REG uses the most recently created SAS data set.

EDF

outputs the number of regressors in the model excluding and including the intercept, the error degrees of freedom, and the model R^2 to the OUTEST= data set.

GOUT=*graphics-catalog*

specifies the graphics catalog in which graphics output is saved. The default *graphics-catalog* is WORK.GSEG. The GOUT= option cannot be used if the LINEPRINTER option is specified.

LINEPRINTER | LP

creates plots requested as line printer plots. If you do not specify this option, requested plots are created on a high resolution graphics device. This option is required if plots are requested and you do not have SAS/GRAPH software.

NOPRINT

suppresses the normal display of results. Using this option in the PROC REG statement is equivalent to specifying NOPRINT in each MODEL statement. Note that this option temporarily disables the Output Delivery System (ODS); see Chapter 15, "Using the Output Delivery System," for more information.

OUTEST=*SAS-data-set*

requests that parameter estimates and optional model fit summary statistics be output to this data set. See the "OUTEST= Data Set" section for details. If you want to create a permanent SAS data set, you must specify a two-level name (refer to the section "SAS Files" in *SAS Language Reference: Concepts* for more information on permanent SAS data sets).

OUTSEB

outputs the standard errors of the parameter estimates to the OUTEST= data set. The value SEB for the variable _TYPE_ identifies the standard errors. If the RIDGE= or PCOMIT= option is specified, additional observations are included and identified by the values RIDGESEB and IPCSEB, respectively, for the variable _TYPE_. The standard errors for ridge regression estimates and IPC estimates are limited in their usefulness because these estimates are biased. This option is available for all model selection methods except RSQUARE, ADJRSQ, and CP.

OUTSSCP=*SAS-data-set*

requests that the sums of squares and crossproducts matrix be output to this TYPE=SSCP data set. See the "OUTSSCP= Data Sets" section for details. If you want to create a permanent SAS data set, you must specify a two-level name (refer to the section "SAS Files" in *SAS Language Reference: Concepts* for more information on permanent SAS data sets).

OUTSTB

outputs the standardized parameter estimates as well as the usual estimates to the OUTEST= data set when the RIDGE= or PCOMIT= option is specified. The values RIDGESTB and IPCSTB for the variable _TYPE_ identify ridge regression estimates and IPC estimates, respectively.

OUTVIF

outputs the variance inflation factors (VIF) to the OUTEST= data set when the RIDGE= or PCOMIT= option is specified. The factors are the diagonal elements of the inverse of the correlation matrix of regressors as adjusted by ridge regression or IPC analysis. These observations are identified in the output data set by the values RIDGEVIF and IPCVIF for the variable _TYPE_.

PCOMIT=*list*

requests an incomplete principal components (IPC) analysis for each value *m* in the list. The procedure computes parameter estimates using all but the last *m* principal components. Each value of *m* produces a set of IPC estimates, which are output to the OUTEST= data set. The values of *m* are saved by the variable

PCOMIT, and the value of the variable _TYPE_ is set to IPC to identify the estimates. Only nonnegative integers can be specified with the PCOMIT= option.

If you specify the PCOMIT= option, RESTRICT statements are ignored.

PRESS

outputs the PRESS statistic to the OUTEST= data set. The values of this statistic are saved in the variable _PRESS_. This option is available for all model selection methods except RSQUARE, ADJRSQ, and CP.

RIDGE=*list*

requests a ridge regression analysis and specifies the values of the ridge constant *k* (see the "Computations for Ridge Regression and IPC Analysis" section). Each value of *k* produces a set of ridge regression estimates that are placed in the OUTEST= data set. The values of *k* are saved by the variable _RIDGE_, and the value of the variable _TYPE_ is set to RIDGE to identify the estimates.

Only nonnegative numbers can be specified with the RIDGE= option. Example 55.10 illustrates this option.

If you specify the RIDGE= option, RESTRICT statements are ignored.

RSQUARE

has the same effect as the EDF option.

SIMPLE

displays the sum, mean, variance, standard deviation, and uncorrected sum of squares for each variable used in PROC REG.

SINGULAR=*n*

tunes the mechanism used to check for singularities. The default value is machine dependent but is approximately 1E-7 on most machines. This option is rarely needed. Singularity checking is described in the "Computational Methods" section.

TABLEOUT

outputs the standard errors and $100(1-\alpha)$% confidence limits for the parameter estimates, the *t* statistics for testing if the estimates are zero, and the associated *p*-values to the OUTEST= data set. The _TYPE_ variable values STDERR, L*n*B, U*n*B, T, and PVALUE, where $n = 100(1-\alpha)$, identify these rows in the OUTEST= data set. The α-level can be set with the ALPHA= option in the PROC REG or MODEL statement. The OUTEST= option must be specified in the PROC REG statement for this option to take effect.

USSCP

displays the uncorrected sums-of-squares and crossproducts matrix for all variables used in the procedure.