
This page was adapted from a page titled PROC REG Summary.

(<http://www.psych.yorku.ca/lab/sas/sasregr.htm>) created by Professor Michael Friendly.

(<http://www.psych.yorku.ca/lab/>) of York University (<http://www.yorku.ca/web/index.htm>) . We thank Professor Friendly for permission to adapt and distribute this page via our web site.

The REG procedure fits least-squares estimates to linear regression models. The following statements are used with the REG procedure:

```
PROC REG (/sas/library/#REG). options;  
  MODEL (/sas/library/#MODEL). dependents=regressors / options;  
  VAR variables;  
  FREQ (/sas/library/#FREQ). variable;  
  WEIGHT (/sas/library/#WEIGHT). variable;  
  ID (/sas/library/#ID). variable;  
  OUTPUT (/sas/library/#OUTPUT). OUT=SASdataset keyword=names...;  
  PLOT (/sas/library/#PLOT). yvariable*xvariable = symbol ...;  
  RESTRICT linear_equation,...;  
  TEST (/sas/library/#TEST). linear_equation,...;  
  MTEST linear_equation,...;  
  BY variables;
```

The PROC REG statement is always accompanied by one or more MODEL statements to specify regression models. One OUTPUT statement may follow each MODEL statement. Several RESTRICT, TEST, and MTEST statements may follow each MODEL. WEIGHT, FREQ, and ID statements are optionally specified once for the entire PROC step. The purposes of the statements are:

- The MODEL statement specifies the dependent and independent variables in the regression model.
- The OUTPUT statement requests an output data set and names the variables to contain predicted values, residuals, and other output values.
- The ID statement names a variable to identify observations in the printout.
- The WEIGHT and FREQ statements declare variables to weight observations.
- The BY statement specifies variables to define subgroups for the analysis. The analysis is repeated for each value of the BY variable.

Proc REG Statement

```
PROC REG options;
```

These options may be specified on the PROC REG statement:

DATA=SASdataset

names the SAS data set to be used by PROC REG. If DATA= is not specified, REG uses the most recently created SAS data set.

OUTEST=SASdataset

requests that parameter estimates be output to this data set.

OUTSSCP=SASdataset

requests that the crossproducts matrix be output to this TYPE=SSCP data set.

NOPRINT

suppresses the normal printed output.

SIMPLE

prints the “simple” descriptive statistics for each variable used in REG.

ALL

requests many different printouts.

COVOUT

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

MODEL Statement

```
label: MODEL dependents = regressors / options;
```

After the keyword MODEL, the dependent (response) variables are specified, followed by an equal sign and the regressor variables. Variables specified in the MODEL statement must be variables in the data set being analyzed. The label is optional.

- General options:

NOPRINT

suppresses the normal printout of regression results.

NOINT

suppresses the intercept term that is normally included in the model automatically.

ALL

requests all the features of these options: XPX, SS1, SS2, STB, TOL, COVB, CORRB, SEQB, P, R, CLI, CLM.

- Options to request regression calculations:

XPX

prints the $X'X$ crossproducts matrix for the model.

I

prints the $(X'X)^{-1}$ matrix.

- Options for details on the estimates:

SS1

prints the sequential sums of squares (Type I SS) along with the parameter estimates for each term in the model.

SS2

prints the partial sums of squares (Type II SS) along with the parameter estimates for each term in the model.

STB

prints standardized regression coefficients.

TOL

prints tolerance values for the estimates.

VIF

prints variance inflation factors with the parameter estimates. Variance inflation is the reciprocal of tolerance.

COVB

prints the estimated covariance matrix of the estimates.

CORRB

prints the correlation matrix of the estimates.

SEQB

prints a sequence of parameter estimates as each variable is entered into the model.

COLLIN

requests a detailed analysis of collinearity among the regressors.

COLLINOINT

requests the same analysis as the COLLIN option with the intercept variable adjusted out rather than included in the diagnostics.

- Options for predicted values and residuals:

P

calculates predicted values from the input data and the estimated model.

R

R

requests that the residual be analyzed.

CLM

prints the 95% upper and lower confidence limits for the expected value of the dependent variable (mean) for each observation.

CLI

requests the 95% upper and lower confidence limits for an individual predicted value.

DW

calculates a Durbin-Watson statistic to test whether or not the errors have first-order autocorrelation. (This test is only appropriate for time-series data.)

INFLUENCE

requests a detailed analysis of the influence of each observation on the estimates and the predicted values.

PARTIAL

requests partial regression leverage plots for each regressor.

FREQ Statement

```
FREQ variable;
```

If a variable in your data set represents the frequency of occurrence for the other values in the observation, include the variable's name in a FREQ statement. The procedure then treats the data set as if each observation appears *n* times, where *n* is the value of the FREQ variable for the observation. The total number of observations will be considered equal to the sum of the FREQ variable when the procedure determines degrees of freedom for significance probabilities.

WEIGHT Statement

```
WEIGHT variable;
```

A WEIGHT statement names a variable on the input data set whose values are relative weights for a weighted least-squares fit. If the weight value is proportional to the reciprocal of the variance for each observation, then the weighted estimates are the best linear unbiased estimates (BLUE).

ID Statement

```
ID variable;
```

The ID statement specifies one variable to identify observations as output from the MODEL options P, R, CLM, CLI, and INFLUENCE.

OUTPUT Statement

The OUTPUT statement specifies an output data set to contain statistics calculated for each observation. For each statistic, specify the keyword, an equal sign, and a variable name for the statistic on the output data set. If the MODEL has several dependent variables, then a list of output variable names can be specified after each keyword to correspond to the list of dependent variables.

```
OUTPUT OUT=SASdataset  
    PREDICTED=names or P=names  
    RESIDUAL=names or R=names  
    L95M=names  
    U95M=names  
    L95=names  
    U95=names  
    STDP=names  
    STDR=names  
    STUDENT=names  
    COOKD=names  
    H=names  
    PRESS=names  
    RSTUDENT=names  
    DFFITS=names  
    COVRATIO=names;
```

The output data set named with OUT= contains all the variables for which the analysis was performed, including any BY variables, any ID variables, and variables named in the OUTPUT statement that contain statistics.

These statistics may be output to the new data set:

PREDICTED=

P=

predicted values.

RESIDUAL=

R=

residuals, calculated as ACTUAL minus PREDICTED.

L95M=

lower bound of a 95% confidence interval for the expected value (mean) of the dependent variable.

U95M=

upper bound of a 95% confidence interval for the expected value (mean) of the dependent variable.

L95=

lower bound of a 95% confidence interval for an individual prediction. This includes the variance of the error as well as the variance of the parameter estimates.

U95=

upper bound of a 95% confidence interval for an individual prediction.

STDP=

standard error of the mean predicted value.

STDR=

standard error of the residual.

STUDENT=

studentized residuals, the residual divided by its standard error.

COOKD=

Cook's D influence statistic.

H=

leverage.

PRESS=

residual for estimates dropping this observation, which is the residual divided by $(1-h)$ where h is leverage above.

RSTUDENT=

studentized residual defined slightly differently than above.

DFFITS=

standard influence of observation on predicted value.

COVRATIO=

standard influence of observation on covariance of betas, as discussed with INFLUENCE option.

PLOT Statement

```
PLOT yvariable*xvariable=symbol / options
```

The PLOT statement prints scatter plots of the yvariables on the vertical axis and xvariables on the horizontal axis. It uses the symbol specified to mark the points. The yvariables and xvariables may be any variables in the data set or any of the calculated statistics available in the OUTPUT statement.

TEST Statement

TEST Statement

```
label:  TEST equation1,  
          equation2,  
          .  
          .  
          .  
          equationk;  
label:  TEST equation1,..., equationk / options;
```

The TEST statement, which has the same syntax as the RESTRICT statement except for options, tests hypotheses about the parameters estimated in the preceding MODEL statement. Each equation specifies a linear hypothesis to be tested.

One option may be specified in the TEST statement after a slash (/):

PRINT

prints intermediate calculations.

BY Statement

```
BY variables;
```

A BY statement may be used with PROC REG to obtain separate analyses on observations in groups defined by the BY variables. When a BY statement appears, the procedure expects the input data set to be sorted in order of the BY variables. If your input data set is not sorted in ascending order, use the SORT procedure with a similar BY statement to sort the data, or, if appropriate, use the BY statement options **NOTSORTED** or **DESCENDING**.

This page was adapted from a page titled [PROC REG Summary](http://www.psych.yorku.ca/lab/sas/sasregr.htm).

(<http://www.psych.yorku.ca/lab/sas/sasregr.htm>) created by [Professor Michael Friendly](http://www.psych.yorku.ca/lab/)

(<http://www.psych.yorku.ca/lab/>) of [York University](http://www.yorku.ca/web/index.htm) (<http://www.yorku.ca/web/index.htm>) . We thank Professor Friendly for permission to adapt and distribute this page via our web site.

[Click here to report an error on this page or leave a comment](#)

[How to cite this page \(https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-how-do-i-cite-web-pages-and-programs-from-the-ucla-statistical-consulting-group/\)](https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-how-do-i-cite-web-pages-and-programs-from-the-ucla-statistical-consulting-group/)

RESEARCH CENTER FOR THE STUDY OF THE HISTORY OF STATISTICS