

OUTPUT Statement

OUTPUT < **OUT**=SAS-data-set > keyword=names
< ... keyword=names > ;

The OUTPUT statement creates a new SAS data set that saves diagnostic measures calculated after fitting the model. The OUTPUT statement refers to the most recent MODEL statement. At least one *keyword=names* specification is required.

All the variables in the original data set are included in the new data set, along with variables created in the OUTPUT statement. These new variables contain the values of a variety of statistics and diagnostic measures that are calculated for each observation in the data set. If you want to create a permanent SAS data set, you must specify a two-level name (for example, *libref.data-set-name*). For more information on permanent SAS data sets, refer to the section "SAS Files" in *SAS Language Reference: Concepts*.

The OUTPUT statement cannot be used when a TYPE=CORR, TYPE=COV, or TYPE=SSCP data set is used as the input data set for PROC REG. See the ["Input Data Sets"](#) section for more details.

The statistics created in the OUTPUT statement are described in this section. More details are contained in the ["Predicted and Residual Values"](#) section and the ["Influence Diagnostics"](#) section. Also see [Chapter 3, "Introduction to Regression Procedures,"](#) for definitions of the statistics available from the REG procedure.

You can specify the following options in the OUTPUT statement.

OUT=SAS data set

gives the name of the new data set. By default, the procedure uses the DATA_n convention to name the new data set.

keyword=names

specifies the statistics to include in the output data set and names the new variables that contain the statistics. Specify a keyword for each desired statistic (see the following list of keywords), an equal sign, and the variable or variables to contain the statistic.

In the output data set, the first variable listed after a keyword in the OUTPUT statement contains that statistic for the first dependent variable listed in the MODEL statement; the second variable contains the statistic for the second dependent variable in the MODEL statement, and so on. The list of variables following the equal sign can be shorter than the list of dependent variables in the MODEL statement. In this case, the procedure creates the new names in order of

the dependent variables in the MODEL statement.

For example, the SAS statements

```
proc reg data=a;
  model y z=x1 x2;
  output out=b
    p=yhat zhat
    r=yresid zresid;
run;
```

create an output data set named b. In addition to the variables in the input data set, b contains the following variables:

- yhat, with values that are predicted values of the dependent variable y
- zhat, with values that are predicted values of the dependent variable z
- yresid, with values that are the residual values of y
- zresid, with values that are the residual values of z

You can specify the following keywords in the OUTPUT statement. See the ["Model Fit and Diagnostic Statistics"](#) section for computational formulas.

Keyword	Description
COOKD= <i>names</i>	Cook's <i>D</i> influence statistic
COVRATIO= <i>names</i>	standard influence of observation on covariance of betas, as discussed in the "Influence Diagnostics" section
DFFITS= <i>names</i>	standard influence of observation on predicted value
H= <i>names</i>	leverage, $x_i(\mathbf{X}'\mathbf{X})^{-1}x_i'$
LCL= <i>names</i>	lower bound of a $100(1 - \alpha)$ % confidence interval for an individual prediction. This includes the variance of the error, as well as the variance of the parameter estimates.
LCLM= <i>names</i>	lower bound of a $100(1 - \alpha)$ % confidence interval for the expected value (mean) of the dependent variable
PREDICTED P= <i>names</i>	predicted values
PRESS= <i>names</i>	<i>i</i> th residual divided by $(1-h)$. where <i>h</i> is the leverage.

	and where the model has been refit without the i th observation
RESIDUAL R= <i>names</i>	residuals, calculated as ACTUAL minus PREDICTED
RSTUDENT= <i>names</i>	a studentized residual with the current observation deleted
STDI= <i>names</i>	standard error of the individual predicted value
STDP= <i>names</i>	standard error of the mean predicted value
STDR= <i>names</i>	standard error of the residual
STUDENT= <i>names</i>	studentized residuals, which are the residuals divided by their standard errors
UCL= <i>names</i>	upper bound of a $100(1 - \alpha)$ % confidence interval for an individual prediction
UCLM= <i>names</i>	upper bound of a $100(1 - \alpha)$ % confidence interval for the expected value (mean) of the dependent variable