SAS Procedures exist to carry out all the forms of statistical analysis. A procedure is invoked in a "PROC step" which starts with the keyword PROC, such as:

PROC MEANS DATA=CLASS;

VAR HEIGHT WEIGHT;

The VAR or VARIABLES statement can be used with all procedures to indicate which variables are to be analyzed. If this statement is omitted, the default is to include ***all variables***of the appropriate type (character or numeric) for the given analysis.

Some other statements that can be used with most SAS procedure steps are:

BY variable(s);

Causes the procedure to be repeated automatically for ***each***different value of the named variable(s). The data set must first be sorted by those variables.

ID variable(s);

Give the name of a variable to be used as an observation **ID**entifier.

LABEL var='label';

Assign a descriptive label to a variable.

WHERE (expression);

Select only those observations for which the expression is true.

For example, the following lines produce separate means for males and females, with the variable SEX labelled 'Gender'. (An ID statement is not appropriate, because PROC MEANS produces only summary output.)

PROC SORT DATA=CLASS;

BY SEX;

RUN;

PROC MEANS DATA=CLASS;

VAR HEIGHT WEIGHT;

BY SEX;

LABEL SEX='Gender';

RUN;

If the DATA= option is not used, SAS procedures process the most recently created dataset. In the brief summaries below, the required portions of a PROC step are shown in **bold**. Only a few representative options are shown.

**Descriptive statistics**

**PROC CORR**

**Correlations among a set of variables.**

**PROC CORR** DATA=*SASdataset options***;**

*options:*NOMISS ALPHA

VAR *variable(s)*;

WITH *variable(s)*;

PROC FREQ

**PROC FREQ** DATA=*SASdataset***;**

TABLES *variable(s) / options*;

*options:*NOCOL NOROW NOPERCENT

OUTPUT OUT=*SASdataset*;

PROC MEANS

**Means, standard deviations, and a host of other univariate statistics for a set of variables.**

**PROC MEANS** DATA=*SASdataset options***;**

*options:*N MEAN STD MIN MAX SUM VAR CSS USS

VAR *variable(s)*;

BY *variable(s)*;

OUTPUT OUT=*SASdataset keyword=variablename ...* ;

Statistical options on the PROC MEANS statement determine which statistics are printed. The (optional) OUTPUT statement is used to create a SAS dataset containing the values of these statistics.

PROC UNIVARIATE

**Univariate statistics and displays for a set of variables.**

**PROC UNIVARIATE** DATA=*SASdataset options***;**

*options:*PLOT

VAR *variable(s)*;

BY *variable(s)*;

OUTPUT OUT=*SASdataset keyword=variablename ...* ;

**Linear models**

SAS statements and options for regression **(PROC REG)** are described in more detail in the document PROC REG Summary. SAS statements and options for analysis of variance **(PROC ANOVA and PROC GLM)** described in the document PROC ANOVA and PROC GLM.

PROC ANOVA

**Analysis of variance (balanced designs)**

**PROC ANOVA** DATA=*SASdataset options***;**

**CLASS *variable(s) a b c*** ;

**MODEL *dependent(s)***= ***effect(s) b c***;

PROC GLM

**General linear models, including ANOVA, regression and analysis of covariance models.**

**PROC GLM** DATA=*SASdataset options***;**

CLASS *variable(s)*;

**MODEL** ***dependent(s)***= ***effect(s)***;

OUTPUT OUT=*SASdataset keyword=variablename ...* ;

PROC REG

**Regression analysis**

**PROC REG** DATA=*SASdataset options***;**

**MODEL** ***dependent(s) = regressors***

*/ options*;

PLOT *variable | keyword.* \*

*variable | keyword.* = symbol ;

OUTPUT OUT=*SASdataset* P=name R=name ... ;

**Plots and charts**

PROC CHART

**Histograms and bar charts**

**PROC CHART** DATA=*SASdataset options***;**

VBAR *variable / options*;

HBAR *variable / options*;

*options:* MIDPOINTS= GROUP= SUMVAR=

PROC PLOT

Scatter plots

**PROC PLOT** DATA=*SASdataset options***;**

*options:* HPERCENT= VPERCENT=

**PLOT*****yvariable*** \*

***xvariable*** =  *symbol / options*;

PLOT *(yvariables)* \*

*(xvariables)* =  *symbol / options* ;

*PLOT options:* BOX OVERLAY VREF= HREF=

BY  *variable(s)* ;

Note that the parenthesized form in the PLOT statement plots **each** y-variable listed against each x-variable.

**Utility procedures**

PROC PRINT

Print a SAS data set

**PROC PRINT** DATA= *SASdataset options***;**

*options:* UNIFORM LABEL SPLIT='char'

VAR *variable(s)*;

BY *variable(s)*;

SUM *variable(s)*;

PROC SORT

Sort a SAS data set according to one or more variables.

**PROC SORT** DATA=*SASdataset options***;**

*options:* OUT=

**BY** ***variable(s)***;