Summarizing Your Data Using PROC MEANS

One of the first things people usually want to do with their data, after reading it and making sure it is correct, is look at some simple statistics. Statistics such as the mean value, standard deviation, and minimum and maximum values give you a feel for your data. The MEANS procedure provides simple statistics for numeric variables.

The MEANS procedure starts with the keywords PROC MEANS, followed by options:

PROC MEANS *options*;

Some options control how your data are summarized:

MAXDEC = n	specifies the number of decimal places to be displayed
MISSING	treats missing values as valid summary groups

Other options request specific summary statistics:

MAX	maximum value
MIN	minimum value
MEAN	mean
MEDIAN	median
MODE	mode
N	number of non-missing values
NMISS	number of missing values
RANGE	range
STDDEV	standard deviation
SUM	sum

If you do not specify any summary statistics, SAS will print the number of non-missing values, the mean, the standard deviation, and the minimum and maximum values for each variable.

If you use the PROC MEANS statement with no other statements, then you will get statistics for all numeric variables in your data set. Here are some of the optional statements for controlling which variables are used:

BY variable-list; The BY statement performs separate analyses for each level of the variables in the list. The data must first be sorted by these variables. (You can use PROC SORT to do this.)

CLASS variable-list; The CLASS statement also performs separate analyses for each level of the variables in the list, but its output is more compact than with the BY statement, and the data do not have to be sorted first.

VAR variable-list; The VAR statement specifies which numeric variables to use in the analysis. If it is absent, then SAS uses all numeric variables.