

## PROC MEANS: Introduction

The MEANS procedure provides data summarization tools to compute descriptive statistics for variables across all observations and within groups of observations. For example, PROC MEANS

- calculates descriptive statistics based on moments
- estimates quantiles, which includes the median
- calculates confidence limits for the mean
- identifies extreme values
- performs a **t** test.

## PROC MEANS: Syntax

**PROC MEANS** <option(s)> <statistic-keyword(s)>;

**BY** <DESCENDING> variable-1 <... <DESCENDING> variable-n><NOTSORTED>;

**CLASS** variable(s) </ option(s)>;

**FREQ** variable;

**ID** variable(s);

**OUTPUT** <OUT=SAS-data-set> <output-statistic-specification(s)>  
<id-group-specification(s)> <maximum-id-specification(s)>  
<minimum-id-specification(s)> </ option(s)> ;

**TYPES** request(s);

**VAR** variable(s) < / WEIGHT=weight-variable>;

**WAYS** list;

**WEIGHT** variable;

To do this	Use this statement
Calculate separate statistics for each BY group	BY
Identify variables whose values define subgroups for the analysis	CLASS
Identify a variable whose values represent the frequency of each observation	FREQ
Include additional identification variables in the output data set	ID
Create an output data set that contains specified statistics and identification variables	OUTPUT
Identify specific combinations of class variables to use to subdivide the data	TYPES
Identify the analysis variables and their order in the results	VAR
Specify the number of ways to make unique combinations of class variables	WAYS
Identify a variable whose values weight each observation in the statistical calculations	WEIGHT

## PROC MEANS Statement

**PROC MEANS** <option(s)> <statistic-keyword(s)>;

*Options* (only a few examples)

### ALPHA=value

specifies the confidence level to compute the confidence limits for the mean. The percentage for the confidence limits is  $(1-\text{value})\times 100$ . For example, ALPHA=.05 results in a 95% confidence limit.

<b>Default:</b>	.05
<b>Range:</b>	between 0 and 1
<b>Interaction:</b>	To compute confidence limits specify the <i>statistic-keyword</i> CLM, LCLM, or UCLM.

### DATA=SAS-data-set

identifies the input SAS data set.

### FW=field-width

specifies the field width to display the statistics in the output.

<b>Default:</b>	12
<b>Tip:</b>	If PROC MEANS truncates column labels in the output, increase the field width.

### MAXDEC=number

specifies the maximum number of decimal places to display the statistics in the output.

<b>Default:</b>	BEST. width for columnar format, typically about 7. (This does not apply to the PROBT statistic. The SAS system option PROBSIG= determines its format.)
<b>Range:</b>	0-8

### MISSING

considers missing values as valid values to create the combinations of class variables. Special missing values that represent numeric values (the letters A through Z and the underscore (\_) character) are each considered as a separate value.

<b>Default:</b>	If you omit MISSING, PROC MEANS excludes the observations with a missing class variable value from the analysis.
-----------------	------------------------------------------------------------------------------------------------------------------

### NONOBS

suppresses the column that displays the total number of observations for each unique combination of the values of the class variables. This column corresponds to the \_FREQ\_ variable in the output data set.

## NOPRINT | PRINT

specifies whether PROC MEANS displays the statistical analysis. NOPRINT suppresses all the output.

<b>Default:</b>	PRINT
<b>Tip:</b>	Use NOPRINT when you want to create only an OUT= output data set.

## VARDEF=divisor

specifies the divisor to use in the calculation of the variance and standard deviation.

<b>Default:</b>	DF
<b>Requirement:</b>	To compute the standard error of the mean, confidence limits for the mean, or the Student's <b>t</b> -test, use the default value of VARDEF=.

## statistic-keyword(s)

specifies which statistics to compute and the order to display them in the output. The available keywords in the PROC statement are:

### *Descriptive statistic keywords:*

CLM	MIN	STDERR
CSS	N	SUM
CV	NMISS	SUMWGT
KURTOSIS KURT	RANGE	UCLM
LCLM	SKEWNESS SKEW	USS
MAX	STDDEV STD	VAR
MEAN		

### *Quantile statistic keywords:*

MEDIAN P50	Q1 P25	P95
P1	Q3 P75	P99
P5	P90	QRANGE
P10		

### *Hypothesis testing keyword:*

PROBT	T
<b>Default:</b>	N, MEAN, STD, MIN, and MAX
<b>Requirement:</b>	To compute standard error, confidence limits for the mean, and the Student's <b>t</b> test you must use the default value of VARDEF= which is DF. To compute skewness or kurtosis you must use VARDEF=N or VARDEF=DF.
<b>Tip:</b>	Use CLM or both LCLM and UCLM to compute a two-sided confidence limit for the mean. Use only LCLM or UCLM, to compute a one-sided confidence limit.