Chapter 8 PRODUCING DESCRIPTIVE STATISTICS

* To summarize descriptive statistics for **numeric** variables, use **PROC MEANS**. By default, this procedure computes the **n-count** (the number of non-missing values), **mean, standard deviation, minimum and maximum** **values** for every numeric variable in a data set.

Example. Age, and total shoulder scores pre-operatively and at one year are recorded for eight patients with repaired shoulders. Total scores are given on 100-point scale. They include evaluations of shoulder pain, rotation, and strength. The higher this score, the more functional the shoulder is. The raw data are

Age at Total Score Total Score

Pt ID gender procedure pre-operatively at 1 year

001 Female 53 62 89

002 Female 59 66 95

003 Female 56 68 95

004 Male 68 55 98

005 Male 56 53 98

006 Female 55 60 93

007 Female 64 69 93

008 Male 63 57 93

The following code computes the n-count, mean, standard deviation, min and max for the three numeric variables.

data shoulder;

input id $ gender $ age pre\_score year1\_score;

cards;

001 female 53 62 89

002 female 59 66 95

003 female 56 68 95

004 male 68 55 98

005 male 56 53 98

006 female 55 60 93

007 female 64 69 93

008 male 63 57 93

;

proc means;

run;

The output is

The MEANS Procedure

Variable N Mean Std Dev Minimum Maximum

ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ

age 8 59.2500000 5.2304056 53.0000000 68.0000000

pre\_score 8 61.2500000 6.0415230 53.0000000 69.0000000

year1\_score 8 94.2500000 2.9640706 89.0000000 98.0000000

ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ

* Other statistics may be computed if they are requested.

Example. In the shoulder example, the following code requests the n-count, mean, median, standard deviation, minimum, maximum, and range=maximum-minimum.

proc means n mean median std min max range;

run;

The output is

The MEANS Procedure

Variable N Mean Median Std Dev Minimum Maximum

ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ

age 8 59.2500000 57.5000000 5.2304056 53.0000000 68.0000000

pre\_score 8 61.2500000 61.0000000 6.0415230 53.0000000 69.0000000

year1\_score 8 94.2500000 94.0000000 2.9640706 89.0000000 98.0000000

ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ

Variable Range

ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ

age 15.0000000

pre\_score 16.0000000

year1\_score 9.0000000

ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ

* By default, each statistics has 7 decimal places. To limit the number of decimal places, option **MAXDEC=** may be used.

Example. In the shoulder example, the following code limits the number of decimal places to two.

proc means maxdec=2;

run;

The output is

The MEANS Procedure

Variable N Mean Std Dev Minimum Maximum

ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ

age 8 59.25 5.23 53.00 68.00

pre\_score 8 61.25 6.04 53.00 69.00

year1\_score 8 94.25 2.96 89.00 98.00

ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ

* A list of variables in PROC MEANS may be specified by using the **VAR statement**.

Example. In the shoulder example, the following code runs the MEANS procedure for Pre\_score and Year1\_score only.

proc means;

var pre\_score year1\_score;

run;

The output is

The MEANS Procedure

Variable N Mean Std Dev Minimum Maximum

ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ

pre\_score 8 61.2500000 6.0415230 53.0000000 69.0000000

year1\_score 8 94.2500000 2.9640706 89.0000000 98.0000000

ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ

* To produce separate analyses for different values of a **character** variable, specify this variable in the **CLASS statement**.

Example. In the shoulder example, separate analyses are produced for males and females.

proc means;

class gender;

run;

The output is

The MEANS Procedure

N

gender Obs Variable N Mean Std Dev Minimum Maximum

ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ

female 5 age 5 57.4000000 4.2778499 53.0000000 64.0000000

pre\_score 5 65.0000000 3.8729833 60.0000000 69.0000000

year1\_score 5 93.0000000 2.4494897 89.0000000 95.0000000

male 3 age 3 62.3333333 6.0277138 56.0000000 68.0000000

pre\_score 3 55.0000000 2.0000000 53.0000000 57.0000000

year1\_score 3 96.3333333 2.8867513 93.0000000 98.0000000

ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ

* Alternatively to the CLASS statement, a **BY statement** may be used. The data set must be sorted in the order of the BY variable. Contrary to the CLASS statement that produces a single table, the BY statement produces separate tables for different values of the BY variable.

Example. In the shoulder example, separate tables are produced for males and females.

proc sort;

by gender;

run;

proc means;

by gender;

run;

The output is

---------------------------------------- gender=female ------------------------------------------

The MEANS Procedure

Variable N Mean Std Dev Minimum Maximum

ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ

age 5 57.4000000 4.2778499 53.0000000 64.0000000

pre\_score 5 65.0000000 3.8729833 60.0000000 69.0000000

year1\_score 5 93.0000000 2.4494897 89.0000000 95.0000000

ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ

----------------------------------------- gender=male -------------------------------------------

Variable N Mean Std Dev Minimum Maximum

ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ

age 3 62.3333333 6.0277138 56.0000000 68.0000000

pre\_score 3 55.0000000 2.0000000 53.0000000 57.0000000

year1\_score 3 96.3333333 2.8867513 93.0000000 98.0000000

ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ

* To write the statistics summary table to a file, use the **OUTPUT statement**.

The keyword for the statistic and the list of all the variables must be specified. The variables must be listed in the same order as they are in the data set or in the VAR statement. To suppress the report being created in the Output window, the **NOPRINT** option in the PROC MEANS may be used. The syntax is

Example. In the shoulder example, a data set named Work.Mean\_by\_gender is created that contains the means for age, pre\_score and year1\_score by gender. The report in the Output window is suppressed.

proc means noprint;

class gender;

output out=mean\_by\_gender

mean= Avg\_Age Avg\_Pre\_score Avg\_Year1\_score;

run;

The contents of the file Work.Mean\_by\_gender is

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| gender | \_TYPE\_ | \_FREQ\_ | Avg\_Age | Avg\_Pre\_score | Avg\_Year1\_score | |
|  | 0 | 8 | 59.25 | 61.25 | 94.25 |  |
| female | 1 | 5 | 57.4 | 65 | 93 |  |
| male | 1 | 3 | 62.3333333 | 55 | 96.33333 |  |

In addition to the specified variables, the procedure adds the \_TYPE\_ and \_FREQ\_ variables to the output data set.

* Alternatively to the PROC MEANS, the procedure **SUMMARY** may be used. The **PRINT** option and the **VAR statement** must be specified in PROC SUMMARY to create a report. To suppress the output in the Output window, omit the PRINT option.

Example. In the shoulder example, type

proc summary print;

var age pre\_score year1\_score;

run;

The output is

The SUMMARY Procedure

Variable N Mean Std Dev Minimum Maximum

ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ

age 8 59.2500000 5.2304056 53.0000000 68.0000000

pre\_score 8 61.2500000 6.0415230 53.0000000 69.0000000

year1\_score 8 94.2500000 2.9640706 89.0000000 98.0000000

ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ

To suppress the output and create a file with the report, type

proc summary;

var age pre\_score year1\_score;

class gender;

output out=mean\_by\_gender

mean=avg\_age avg\_pre\_score avg\_year1\_score;

run;

The contents of the file Work.Mean\_by\_gender is

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| gender | \_TYPE\_ | \_FREQ\_ | avg\_age | avg\_pre\_score | avg\_year1\_score | |
|  | 0 | 8 | 59.25 | 61.25 | 94.25 |  |
| female | 1 | 5 | 57.4 | 65 | 93 |  |
| male | 1 | 3 | 62.3333333 | 55 | 96.33333 |  |

* The PROC MEANS or PROC SUMMARY reports may be viewed as HTML output.

Example. In the shoulder example, to create the default PROC MEANS report as HTML output, type

ods html body=' C:\Users\mstudent\Desktop\outmean.html';

proc means;

class gender;

run;

ods html close;

The output is

