**Basic Statistical Procedures Using SAS**

**PROC MEANS**

This procedure is used to calculate arithmetic mean and standard deviation. For people who are new to statistics may find it difficult to understand these terms. So before we start coding and use this procedure. I will try to explain what these terms mean.

Let’s start with arithmetic mean and see how PROC MEANS is used in SAS programming to calculate it.

**Arithmetic Mean**

Sum of the value of numeric variables, divided by the number of variables gives you the **arithmetic mean**. It is also known as mean and is a measure of central tendency. A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data.

In SAS programming, you use PROC MEANS to calculate the arithmetic mean. This procedure lets you find mean of all variables or few variables of a data set. You can also form groups and calculate mean of variables specific to that group.

**Syntax:**

|  |  |
| --- | --- |
| 1  2  3 | PROC MEANS DATA = DATASET;  Class Variables ;  Var Variables; |

* **Variables**: Variables in the above syntax indicate variables from the data set whose mean is to be calculated.

**Mean Of A Dataset**

If you supply only the data set name without any variables, you can calculate the mean of all the variables in a data set.

Let us take a look at a sample code. I have considered a predefined SAS data set called as ‘cars’. The following command will display the data set.

|  |  |
| --- | --- |
| 1  2 | PROC PRINT data=sashelp.CARS;  Run; |

Now let us use this data set code and calculate the mean of each variable in the data set ‘cars’.

|  |  |
| --- | --- |
| 1  2 | PROC MEANS DATA = sashelp.CARS Mean SUM MAXDEC=2;  Run; |

**Mean Of Selected Variables**

By supplying the names in the Var option you can get the mean of the specified variables. Please refer the code below.

|  |  |
| --- | --- |
| 1  2  3 | PROC MEANS DATA = sashelp.CARS mean SUM MAXDEC=2;  var horsepower cylinders;  Run; |

**Mean By Class**

You can find the mean of the numeric variables by organizing them into groups by using some parameter to group them.Consider following sample code. Lets find out the mean of horsepower for different groups categorized by the classes  ‘make’ and ‘type’ of different cars.

|  |  |
| --- | --- |
| 1  2  3  4 | PROC MEANS DATA = sashelp.CARS MEANS SUM MAXDEC=2;  class make type;  var horsepower;  Run; |

**Standard Deviation**

Standard deviation (SD) is a measure of how varied is the data in a given data set. Mathematically, it tells you how close is each data point to the mean value of a data set. If the value of standard deviation is close to 0, it indicates that the data points are very close to the mean of the data set and a high standard deviation indicates that the data points are spread out over a wide range of values.

In SAS, you can calculate the value of Standard Deviation using two procedures. They are:

* PROC MEANS
* SURVEYMEANS

**Standard Deviation Using PROC MEANS**

You can measure the Standard Deviation using proc means, you have to choose the **STD** option in the PROC step. It will display the Standard Deviation values for each numeric variable in the data set.

**Syntax:**

|  |  |
| --- | --- |
| 1 | PROC MEANS DATA = dataset STD; |

Consider this sample code, let us create another data set CARS1 from the CARS data set in the SASHELP library. To do this we let us use PROC SQL procedure. Let us group the data using ‘type’ and ‘make’ of cars and calculate standard deviation for selected variables using the STD option with the PROC means step.

|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8  9  10 | PROC SQL;  create table CARS1 as  SELECT make,type,horsepower,cylinders,weight  FROM  SASHELP.CARS  WHERE make in ('Audi','BMW')  ;  RUN;  PROC MEANS DATA=CARS1 STD;  Run; |

**PROC SURVEYMEANS**

This procedure is used to measure Standard Deviation along with some advance features like measuring Standard Deviation for categorical variables and the variance.

**Syntax:**

|  |  |
| --- | --- |
| 1  2  3  4 | PROC SURVEYMEANS options statistic-keywords;  By variables;  Class variables;  Var variables; |

Following is the description of the parameters used:

* **By** is used to indicate the variables used to create groups of observations.
* **Class** indicates the variables used for categorical variables.
* **Var** indicates the variables for which SD will be calculated.

Let us take a look at this sample code which describes the use of the class parameter, that creates the statistics for each of the values in the class variable.

|  |  |
| --- | --- |
| 1  2  3  4  5  6  7 | PROC SURVEYMEANS DATA=CARS1 STD;  Class type;  Var type horsepower;  ods output statistics=rectangle;  Run;  PROC PRINT DATA=rectangle;  Run; |