

Kirk's Korner



Quick & Simple Tips

Kirk Paul Lafler, Software Intelligence Corporation

Tip: Fun with PROC SQL Summary Functions

The SQL procedure is a wonderful tool for summarizing (or aggregating) data. It provides a number of useful summary (or aggregate) functions to help perform calculations, descriptive statistics, and other aggregating operations in a SELECT statement or HAVING clause. These functions are designed to summarize information and not display detail about data.

Without the availability of summary functions, you would have to construct the necessary logic using somewhat complicated SQL programming constructs. When using a summary function without a GROUP BY clause, all the rows in a table are treated as a single group. Consequently, the results are often a single row value.

A number of summary functions are available including facilities to count non-missing values; determine the minimum and maximum values in specific columns; return the range of values; compute the mean, standard deviation, and variance of specific values; and other aggregating functions. The next table shows an alphabetical listing of the available PROC SQL summary functions and, when multiple names for the same function are available, the ANSI-approved name appears first.

Summary Function	Description
AVG, MEAN	Average or mean of values
COUNT, FREQ, N	Aggregate number of non-missing values
CSS	Corrected sum of squares
CV	Coefficient of variation
MAX	Largest value
MIN	Smallest value
NMISS	Number of missing values
PRT	Probability of a greater absolute value of Student's t
RANGE	Difference between the largest and smallest values
STD	Standard deviation
STDERR	Standard error of the mean
SUM	Sum of values
SUMWGT	Sum of the weight variable values which is 1
T	Testing the hypothesis that the population mean is zero
USS	Uncorrected sum of squares
VAR	Variance

The next example uses the **COUNT** function with the (*) argument to produce a total number of rows, regardless if data is missing. The asterisk (*) is specified as the argument to the COUNT function to count all rows in the MOVIES table.

PROC SQL Code

```
PROC SQL;
  SELECT COUNT(*) AS Row_Count
  FROM MOVIES;
QUIT;
```

Results

```
Row_Count
fffffff
      22
```

Unlike the COUNT(*) function syntax that counts all rows, regardless if data is missing or not, the next example uses the **COUNT** function with the (column-name) argument to produce a total number of non-missing rows based on the column, RATING.

SQL Code

```
PROC SQL;
  SELECT COUNT(RATING) AS Non_Missing_Row_Count
  FROM MOVIES;
QUIT;
```

Results

```
Non_Missing_
Row_Count
fffffff
      22
```

The **MIN** summary function can be specified to determine what the shortest running movie is in the MOVIES table.

SQL Code

```
PROC SQL;
  SELECT MIN(length) AS Shortest
  Label='Shortest Length'
  FROM MOVIES;
QUIT;
```

Results

```
Shortest
```

Length
fffffffff
97

In the next example, the RANGE function is specified to determine the difference between the largest and smallest values for a selected column. Suppose you wanted to determine the range of movie lengths for all movies by their movie rating (e.g., G, PG, PG-13, and R). You could construct the following query as follows.

SQL Code

```
PROC SQL;  
  SELECT RANGE(length)  
         AS Range_Length  
  FROM MOVIES  
  GROUP BY RATING;  
QUIT;
```

Results

Range_
Length
fffff
0
32
97
72

CONTACT INFORMATION

If you would like more information or have any questions about this tip, please contact: Kirk Paul Lafler, Software Intelligence Corporation at KirkLafler@cs.com. Kirk has been working with the SAS System since 1979 and is a SAS Certified Professional® and SAS Alliance Partner® (1996 - 2002). His company provides custom SAS programming, application design and development, consulting services, and hands-on SAS training to clients around the world. Kirk is the author of four books including PROC SQL: Beyond the Basics Using SAS by SAS Institute, Power SAS and Power AOL by Apress, and more than one hundred peer-reviewed articles and papers for professional journals and SAS User Group proceedings. His popular SAS Tips column, Kirk's Korner, appears regularly in several SAS user group Newsletters and websites. Kirk can be reached at:

Kirk Paul Lafler
Software Intelligence Corporation
P.O. Box 1390
Spring Valley, California 91979-1390
E-mail: KirkLafler@cs.com



SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.