

Using SQL to Summarize Data Down Rows and Across Columns

Kirk Paul Lafler, Software Intelligence Corporation

Abstract

The SQL procedure provides SAS users with the ability to use SAS-supplied summary functions. These functions permit the creation of averages as well as the set of all quantities. Without the ability to summarize data in PROC SQL, users would be forced to write complicated formulas and/or routines, or even write and test DATA step programs to summarize data. This paper shows how summary functions can be used in PROC SQL to summarize data down rows as well as across columns of data.

Introduction

PROC SQL gives users the ability to summarize data down rows (observations) as well as across columns (variables). This flexibility gives SAS users an incredible range of power, and the ability to take advantage of several SAS-supplied (or built-in) summary functions. These summarization techniques permit the average of quantities rather than the set of all quantities. Without the ability to summarize data in SQL, users would be forced to write complicated formulas and/or routines, or even write and test DATA step programs to summarize data. Two examples will be illustrated to show how SQL can be constructed to summarize data: 1) Summarizing data down rows and 2) Summarizing data across columns.

1. Summarizing data down rows

The SQL procedure can be used to produce a single aggregate value by summarizing data down rows (or observations). The advantages of using a summary function in PROC SQL is that it will generally compute the aggregate quicker than if a user-defined equation were constructed and it saves the effort of having to construct and test a program containing the user-defined equation in the first place.

Suppose you wanted to know the distribution of movies by their movie rating in the MOVIES table. The following query computes the number of movies within each rating using the **COUNT** function and **GROUP BY** clause.

SQL Code

```
PROC SQL;
  SELECT RATING,
         COUNT(RATING) AS
            Movie_Count
  FROM MOVIES
  GROUP BY RATING;
QUIT;
```

Results

Rating	Movie_Count
G	1
PG	6
PG-13	7
R	8

2. Summarizing data across columns

When a computation is needed on two or more columns in a row, PROC SQL can be used to summarize data across columns. Suppose you wanted to convert the length of each movie from minutes to hours in the MOVIES table. The next example rounds each row's movie length by using the ROUND function to the closest hour, and once computed displays the hour rounded to the nearest whole number as Movie_Length.

SQL Code

```
PROC SQL;
  SELECT TITLE,
         ROUND((LENGTH / 60),1) AS
            Movie_Length
  FROM MOVIES;
QUIT;
```

Results

Title	Movie_Length
Brave Heart	3
Casablanca	2
Christmas Vacation	2
Coming to America	2
Dracula	2
Dressed to Kill	2
Forrest Gump	2
Ghost	2
Jaws	2
Jurassic Park	2
Lethal Weapon	2
Michael	2
National Lampoon's Vacation	2
Poltergeist	2

Rocky	2
Scarface	3
Silence of the Lambs	2
Star Wars	2
The Hunt for Red October	2
The Terminator	2
The Wizard of Oz	2
Titanic	3

Conclusion

PROC SQL provides SAS users with the ability to use SAS-supplied summary functions. These functions permit the creation of averages as well as the set of all quantities. Without these functions, users would be forced to write complicated formulas and/or routines, or even write and test DATA step programs to summarize data. This paper has showed how summary functions can be used in PROC SQL to summarize data down rows as well as across columns of data.

Acknowledgments

The author would like to thank Beth Worrall and Kim LeBouton (Coders Corner Co-Chairs) for accepting my paper, as well as the WUSS Leadership for their support of a great Conference.

References

- Lafler, Kirk, PROC SQL for Beginners; Software Intelligence Corporation, Spring Valley, CA, USA; 1992-2003.
- Lafler, Kirk, Intermediate PROC SQL; Software Intelligence Corporation, Spring Valley, CA, USA; 1998-2003.
- Lafler, Kirk, Advanced PROC SQL; Software Intelligence Corporation, Spring Valley, CA, USA; 2001-2003.
- Lafler, Kirk, PROC SQL Programming Tips; Software Intelligence Corporation, Spring Valley, CA, USA; 2002.
- SAS® Guide to the SQL Procedure: Usage and Reference, Version 6, First Edition; SAS Institute, Cary, NC, USA; 1990.
- SAS® SQL Procedure User's Guide, Version 8; SAS Institute Inc., Cary, NC, USA; 2000.

Trademark Citations

SAS, SAS Alliance Partner, and SAS Certified Professional are registered trademarks of SAS Institute Inc. in the USA and other countries. The ® symbol indicates USA registration.

Bio

Kirk Paul Lafler, a SAS Certified Professional® and SAS Alliance Partner® (1996 – 2002), with more than 25 years of SAS software experience provides consulting services and hands-on SAS training around the world. Kirk has written four books including Power SAS by Apress and more than one hundred articles for professional journals and SAS User Group proceedings. His popular SAS Tips column appears regularly in the BASAS, HASUG, SANDS, and SESUG Newsletters. His expertise includes application design and development, training, and programming using base-SAS, SQL, ODS, SAS/FSP, SAS/AF, SCL, FRAME, and SAS/EIS software.

Comments and suggestions can be sent to:

Kirk Paul Lafler
 Software Intelligence Corporation
 P.O. Box 1390
 Spring Valley, California 91979-1390
 E-mail: KirkLafler@cs.com
 Voice: 619.660.2400