

proc sql

Introduction

- SAS proc sql implements most of the functionality of standard SQL (Standard Query Language).
- proc sql has added some statements and other modifications that are not standard SQL.
- proc sql also captures much of the capability of Base SAS programming statements.
- This table summarizes the differences in terminology for the two styles of programming:

Base SAS	proc sql	Data Processing
Dataset	Table	File
Observation	Row	Record
Variable	Column	Field
Merge	Join	Join
Subsetting	Query	

-
- A table created or used by proc sql is exactly the same as a standard SAS dataset.
- See the [Popul Example](#).
- The SAS run statement is not needed to force proc sql to run.
- The datatypes used in the Popul Example are num (numeric) and char(15) (character data of length 15).
- The [Finance Example](#) has a column formatted as a date. However, this column is actually defined as numeric, just as in a SAS dataset.
- In proc sql, informats and formats can be specified for variables. This is non-standard SQL.

Modifying a Single Table

- Here are some commonly used proc sql statements for modifying tables. Each of these statements generates statements in the SAS log.
 - ***1. Create table according to specifications:**
 -
 - `create table sql.pop(`
 - `continent char(15), population num`
 - `);`
 -
 - ***2. Create table with same column attributes as an existing table:**
 -
 - `create table new_table like sql.pop;`
 -
 - ***3. Copy entire table;**
 -
 - `create table new_table as`
 - `select * from sql.pop;`
 -
 - ***4. Delete rows from table:**
 -
 - `delete from sql.pop where population < 1000000;`
 -

- *5. Insert rows into table;
-
- * Method 1
- insert into sql.pop
- values('Africa', 706611183)
- values('Asia', 3379469458);
-
- * Method 2
- insert into sql.pop
- set continent='Africa', population=706611183
- set continent='Asia', population=3379469458;
-
- * Method 3
- insert into new_table
- select * from sql.pop;
-
- *6. Write the create table statement showing the column attributes for an existing table to the SAS log;
-
- describe table sql.pop;
-
- *7. Add column to table:
-
- alter table sql.pop add largest_city char(15);
-
- *8. Modify table column:
-
- alter table sql.pop modify population format 15.;
-
- *9. Update table fields:
-
- update sql.pop set density = population / area;
-
- *10. Delete table column:
-
- alter table sql.pop drop continent;
-
- *11. Drop entire table:
-
- drop table sql.pop;
-
-
- **Problems.** Use the dataset [drugstore1.txt](#) from the [DrugStores](#) Example. Write code to do each of the following using proc sql and also using Base SAS. Use these [data step statements](#) to input the data from drugstore1.txt. Here are the [SQL answers](#) and [Base SAS answers](#).
 1. Remove the store_id column from the table drugstore1.
 2. Rename the store_id column to storeid in the table drugstore1.
 3. Add this row to the table drugstore1:


```
product='cough medicine' chain='CVS' store_id=105 price=6.11
```

Querying a Single Table

- An SQL query is expressed as a select statement.
- The simplest select statement looks like this:
 - **select * from table_name;**
- In general, a select statement is composed by clauses containing these keywords in this order:
 - **select from where**
 -

- **group by** **having** **order by**

Only the select and from clauses are required.

- The meanings of these keywords:

Key Word	Meaning
select	Specify Columns for Query
from	Source Table(s) for Query
where	Specify Rows for Query
group by	Specify Groupings for Summary Statistics
having	Filter Grouped Data
order by	Sort Table Rows

-
- **Problems.** Use the input dataset [drugstore1.txt](#). Write code to do each of the following using proc sql and also using Base SAS. Here are the [SQL answers](#) and [Base SAS answers](#).

4. Find the price of cough medicine at Walgreens store 305.
5. List all of the store numbers for which the price of aspirin was collected.
6. List all of the CVS store numbers. Remove duplicate numbers.

- At first glance group by and order by may seem similar. Here are the major differences:

having Clause	where Clause
specify conditions for including or excluding groups.	specify conditions for including or excluding individual observations.
must follow group by clause if group by is used.	must precede group by if group by is used.
affected by group by clause; if there is no group by, having is treated as group by.	is not affected by group by clause.
processed after group by and any aggregate function.	processed before group by clause, if there is one, and before any aggregate functions.

-
- proc sql provides these aggregate functions:
 - **count, freq, n** **css** **cv** **max** **mean, avg**
 -
 - **min** **nmiss** **pct** **range** **std** **stderr**
 -
 - **sum** **sumwgt** **t** **uss** **var**
- **Problems.** Use the input dataset [drugstore1.txt](#). Write code to do each of the following using proc sql and also using Base SAS. Here are the [SQL answers](#) and [Base SAS answers](#).

7. Sort the rows by descending price.
8. Sort the rows by chain and product.
9. Find the minimum, maximum and average prices of all products.
10. Find the minimum, maximum and average prices by chain and product.
11. Find the minimum, maximum and average prices of aspirin, grouped by chain.

Queries Involving More than One Table

- A proc sql select statement can act on more than one table.
- A join of two tables contains rows with fields from two tables.
- If two tables are used in the from clause and there is no where clause, the cartesian product of the two tables is output.
- A where clause restricts the rows that are output from the query. See the [Patients Example](#).
- **Problems.** Write code to do each of the following using proc sql and also using Base SAS. Use the [drugstore2.txt](#) and [drugstore-info.txt](#) datasets. Here is a [script for creating the SAS datasets](#) and the Here are the [SQL answers](#) and [Base SAS answers](#).
 12. Find the phone numbers of all stores in Chicago.
 13. Find the phone numbers of all Walgreens stores in Chicago.
 14. Find the average price of aspirin at CVS.
 15. At how many stores was price information collected? Give a breakdown by product.

Reference

SAS® 9.1 SQL Procedure: User's Guide, SAS Institute Inc., 2004.
http://support.sas.com/documentation/onlinedoc/91pdf/sasdoc_91/base_sqlproc_6992.pdf