

Row-less SET Statements

Craig Kasper, Manitoba Health

Winnipeg SAS User Group

May 17th, 2013

What is a row-less SET statement?

- A row-less SET statement is a SET statement in a DATA step which is designed to not return any rows.
 - One easy way to write a SET statement to do this is
`SET libname.dataset (WHERE= (1=0))`
- SET statements do not have to return data to be useful.
 - SET statements provide information that SAS uses to set up the data structure which holds the data for the current data row in a DATA step program.
 - This data structure is called the Program Data Vector (PDV).

The Program Data Vector

- The Program Data Vector for a DATA step program holds all information that a DATA step will need for processing, for every data field in a data step program.
 - This includes information such as field names, field lengths, labels, formats, and informats.
- Because a DATA step program may output data to a data set at any time, the Program Data Vector's data structure must be created before the DATA step program code is run.

SET and the Program Data Vector

- When SAS creates the Program Data Vector before running the program, it scans the program looking for program code which will require data fields to be present. This program code may include:
 - variable assignments
 - statements which create variables such as RETAIN
 - statements which access data from an external dataset or filename, such as SET
- When SAS analyzes a SET statement while creating a PDV, SAS examines the variables in the data set, and not the rows themselves.
 - As a result, Even a SET statement which returns no rows will contribute information about its variables to the output data set.

When to use a Row-less SET

- SETting a row-less data set can be used instead of program code which would create or modify variables within the DATA step.
 - Note that doing so ties the output format to a specific data set, rather than the program itself.
- SETting a row-less data set can make it easier to write programs in cases where the input data sets for a program haven't been fully specified yet, or may be incomplete.
 - This can be particularly useful for writing macros which write DATA step programs, though it can be useful for writing DATA step programs as well.

Another Use of WHERE=(1=0)

- SAS's hash object is an object which can be used load a table into memory for random access within a DATA step.
 - A table cannot be loaded into a hash object, however, unless the variables for that table have been added to the PDV before the table is loaded.
 - Using a row-less SET statement is often used in these cases to avoid having to write the program code to create and set up these variables.

Conditionally SETting a data set

- SET statements do not need to be executed on every pass through a DATA step.
- The row-less set statements discussed here are normally executed during the first pass through the DATA step only.

```
IF _N_ = 1 THEN . . .
```

- Note that when a SET statement is not executed during every DATA step pass, that data set's variables are RETAINED between each SET, though your program may change their values after they are read.

Any Questions?