SAS: Creating and Using Permanent Data Sets

Platform: UNIX

Level of Difficulty: Beginner

Overview

A SAS Data Set is a special type of file maintained by the SAS system which contains both data values and descriptive information (e.g., variable names, formats, labels) about the data. All SAS PROCs (e.g., PROC PRINT, PROC GLM) use SAS Data Sets as input.

SAS Data Sets are created during the Data Step of a SAS job. Data Steps begin with a SAS Data statement (e.g., data test;) and end with (a) another Data statement, (b) a PROC Statement (e.g., PROC MEANS), or (c) a Run statement (i.e., run;).

By default, all SAS Data Sets are deleted at the end of a batch job or Display Manager session. In order to use the SAS Data Set in future jobs/sessions, it is necessary to re-run the Data Step that created the SAS Data Set. This consumes both time and computer resources.

As an alternative to re-running a SAS Data Step at the beginning of each SAS batch job or Display Manager session, it is possible to make a SAS Data Set permanent. Permanent SAS Data Sets exist independently of the SAS job/session in which they are created. Future SAS jobs/sessions may then simply refer to the existing (i.e., permanent) SAS Data Set rather than re-create it. This saves time and resources.

Conventions

Within this document, UNIX commands that you will enter at your terminal are in **`Bold Courier`** font**.** Information intended to be typed into files or windows appears in **boldface**. Filenames and (sub)directory names are in *italics*.

Sample Run - Creating the Permanent SAS Data Set

Step 1:
Create a directory called *mywork* within your RCI or other UNIX account using the UNIX **`mkdir`** command.

```
> mkdir mywork
```

Step 2:
Type and submit the following SAS statements using either batch mode or the SAS Display Manager.

```
libname proj1 'mywork';
data proj1.test;
input id var1 var2;
cards;
01 2 3
02 4 4
03 4 2
04 2 1
05 2 3
06 3 9
;
run;
```

The Libname statement associates the directory *mywork* with the name *proj1*. *Proj1*is known as a Libref. A Libref is a nickname for a physical location such as a directory (e.g., *mywork*).

The Data Statement instructs SAS to create a SAS Data Set *test* in the directory referenced by the Libref *proj1* (i.e., in the *mywork* directory).

Step 4:
Use the UNIX **ls** command to look for a file named *test.ssd01* in the *proj1* directory. This is the permanent SAS Data Set created with the SAS statements above.

```
> ls mywork
```

Two Sample Runs - Using the Permanent SAS Data Set

Sample Run #1:
Later (i.e., in a different batch job, or after quitting and re-starting the SAS Display Manager), type and submit the following SAS statements.

```
libname proj1 'mywork';
proc means data=proj1.test;
run;
```

The above statements again assign *proj1* as a Libref for the *mywork* directory and then instruct SAS to run PROC MEANS using the SAS Data Set *test* from the *proj1* Library (i.e., from the *mywork* directory).

Sample Run #2
In a later SAS batch job or Display Manager session, create a new permanent SAS Data Set called *newtest* using the data in *test*.

```
libname proj1 'mywork';
data proj1.newtest;
set proj1.test;
var3=var1+var2;
run;
```

The new permanent SAS Data Set *newtest* will contain the same data as *test* plus a new variable called "var3". Use the **ls** command to verify that your *mywork* directory now contains *test.ssd01* and *newtest.ssd01*.

Notes

All Libref assignments (created with a Libname statement) are cancelled at the end of a SAS batch job or Display Manager session. Later SAS jobs/sessions must re-assign Librefs to the directories containing permanent SAS Data Sets before these Data Sets may be accessed.

Multiple Librefs may be established during a SAS job/session (i.e., multiple Libname statements are permitted, each linking a Libref to a directory). This is useful for storing your permanent SAS Data Sets in different locations.

Multiple SAS Data Sets may be assigned to each Libref/Library. For example, the following statements read and sort a permanent SAS Data Set called *presort* and save the sorted data in a permanent SAS Data Set called *postsort*. Both SAS Data Sets are assigned to the *proj1* Library.

```
libname proj1 'mywork';
proc sort data=proj1.presort out=proj1.postsort;
by name;
```

Learning More

Visit the RUCS Information Center (128 Hill Center, Busch Campus) for copies of other SAS-related and UNIX-related documents. The Information Center also has reference copies of SAS manuals available for use.