SAS libraries

Statistics 426: SAS Programming

Module 3

2021

Overview of Datsets

Many of the data processing tasks performed with SAS is to access data in the form of a SAS dataset, and/or analyze, manage, or present the data

Conceptually, a SAS dataset is a file that consists of two parts: a *descriptor* portion and a *data* portion; periodically it also points to one or more indexes, which enable SAS to locate records in the dataset more efficiently

Descriptor Portion

The descriptor portion of a SAS dataset contains information about the dataset , including:
- the name of the dataset
- the date and time that the dataset was created
- the number of observations
- the number of variables

It can be accessed by PROC CONTENTS to see information about your created library and data files

General form of PROC CONTENTS

```
PROC CONTENTS data=libref._ALL_ NODS;
RUN;
```

`libref`: the library name

NODS option suppresses the descriptor portion of the datasets and is only used in conjunction with the *ALL* keyword. This allows the user to see all of the listing of the data files that are in the library. If you do want to see the descriptor portion of a dataset, rather than `_ALL_ NODS`, input the file name

Data Portion

The data portion of a SAS dataset is a collection of data values that are arranged in a rectangular table.

Observations (Rows)
Rows (called observations) in the dataset are collections of data values that usually relate to a single object

Variables (Columns)
Columns (called variables) in the dataset are collections of values that describe a particular characteristic

Variable Attributes

In addition to general information about the dataset, the descriptor portion contains information about the attributes of each variable in the dataset. The attribute information includes the variable's name, type, length, format, informat, and label.

Name

Each variable has a name that conforms to SAS naming conventions. Variable names follow exactly the same rules as SAS dataset names.

Like dataset names, variable names
- can be 1 to 32 characters long
- must begin with a letter (A–Z, either uppercase or lowercase) or an underscore (_)
- can continue with any combination of numbers, letters, or underscores

Type

A variable's type is either character or numeric
- Character variables, such as Name can contain any values
- Numeric variables, such as Salary, can contain only numeric values (the digits 0 through 9, +, -, ., and E for scientific notation)

A variable's type determines how missing values for a variable are displayed. For character variables a blank represents a missing value; numeric variables a period represents a missing value

Length

A variable's length (the number of bytes used to store it) is related to its type
- Character variables can be up to 32,767 bytes long
- All numeric variables have a default length of 8. Numeric values (no matter how many digits they contain) are stored as floating-point numbers in 8 bytes of storage, unless you specify a different length

Format

Formats are variable attributes that affect the way data values are written. SAS software offers a variety of character, numeric, and date and time formats. You can also create and store your own formats. To write values out using a particular form, you select the appropriate format.

Informat

Whereas formats write values out by using some particular form, informats read data values in certain forms into standard SAS values. Informats determine how data values are read into a SAS dataset . You must use informats to read numeric values that contain letters or other special characters.

This one was read in as using a comma to separate every 3rd digit and we want to display it as with the same commas, but as a dollar amount with a dollar sign and out to the nearest penny.

Label

A variable can have a label, which consists of descriptive text up to 256 characters long. By default, many reports identify variables by their names. You might want to display more descriptive information about the variable by assigning a label to the variable. You might even want to use labels to shorten long variable names in your reports (more on that later as well)

Libraries

SAS needs to know where to read data files to and from, and always inserting a long computer address is not always fun nor easy. Changing the working directory to the location of choice makes the process of reading in the dataset more manageable. They are not necessary (as you can use the WORK library) but can make things more organized with a permanent library

How Long Librefs Remain in Effect

The LIBNAME statement is global, which means that the librefs remain in effect until you modify them, cancel them, or end your SAS session. Therefore, the LIBNAME statement assigns the libref for the current SAS session only. Each time you begin a SAS session, you must assign a libref to each permanent SAS data library that contains files that you want to access in that session. When you end your SAS session or delete a libref, SAS no longer has access to the files in the library. However, the contents of the library still exist on your operating system

The DATASETS Procedure

In addition to PROC CONTENTS, you can also use PROC DATASETS with the CONTENTS statement to view the contents of a SAS library or a SAS dataset

General form of PROC DATASETS with CONTENTS statement

```
PROC DATASETS;
CONTENTS DATA=libref._ALL_ NODETAILS;
QUIT;
```

`CONTENTS`: describes the contents of one or more SAS datasets and prints the directory of the SAS data library
`libref` is the libref that has been assigned to the SAS data library
`_ALL_`: requests a listing of *all* files in the library, a period (.) to append *ALL* to the libref, NODS can *only* be specified when you specify *ALL*
`NODETAILS` (or `NODS`): suppresses the printing of detailed information about each dataset

Library example

```
libname cloud 'S:\Courses\stat-renaes\Stat426\data2';

proc print data=cloud.us_newhire;
run;
```

cloud.us_newhire proc print

The SAS System

Obs	ID	Telephone	Birthday
1	120-012-40-4928	5467887	05DEC1968
2	120-012-83-3816	6888321	03MAY1965
3	120-341-44-0781	9418123	23NOV1972
4	120-423-01-7721	7839191	28JUN1967
5	120-448-23-8111	9428122	30NOV1960
6	120-723-14-8422	8280911	12FEB1964
7	120-819-32-1294	3878181	01SEP1968
8	120-831-34-2411	9677810	24DEC1972
9	120-837-33-8374	9927615	06OCT1971
10	120-877-22-7731	2337449	08JUL1969

cloud.us_newhire library log

```
Log - (Untitled)
3     libname cloud 'S:\Courses\stat-renaes\Stat426\data2';
NOTE: Libref CLOUD was successfully assigned as follows:
      Engine:        V9
      Physical Name: S:\Courses\stat-renaes\Stat426\data2
```

cloud.us_newhire proc print log

```
Log - (Untitled)
4      proc print data=cloud.us_newhire;
NOTE: Writing HTML Body file: sashtml.htm
5      run;

NOTE: There were 10 observations read from the data set CLOUD.US_NEWHIRE.
NOTE: PROCEDURE PRINT used (Total process time):
      real time             0.65 seconds
      cpu time              0.45 seconds
```

proc contents `cloud` library

On whole library

```
proc contents data=cloud._ALL_ NODS;
run;
```

proc contents output

The SAS System

The CONTENTS Procedure

Directory	
Libref	CLOUD
Engine	V9
Physical Name	S:\Courses\stat-renaes\Stat426\data2
Filename	S:\Courses\stat-renaes\Stat426\data2
Owner Name	AD\su-donm
File Size	16KB
File Size (bytes)	16384

#	Name	Member Type	File Size	Last Modified
1	APRSALES	DATA	5KB	07/12/2012 13:54:53
2	APRSALES2	DATA	5KB	07/12/2012 13:54:53
3	AU_SALESFORCE	DATA	9KB	07/12/2012 13:54:53
4	BANKS	DATA	5KB	07/12/2012 13:54:53
5	BIZ_LIST	DATA	5KB	07/12/2012 13:54:53
6	CLEAN_UP	DATA	5KB	07/12/2012 13:54:53
7	CONTACTS	DATA	9KB	07/12/2012 13:54:53
8	CONVERT	DATA	5KB	07/12/2012 13:54:53
9	COUNTRY	DATA	17KB	07/12/2012 13:54:53

proc contents log

```
6      proc contents data=cloud._ALL_ NODS;
7      run;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.15 seconds
      cpu time            0.07 seconds
```

proc contents `cloud.us_newhire`

On one dataset in the library

```
proc contents data=cloud.us_newhire;
run;
```

proc contents `cloud.us_newhire` output a

The SAS System

The CONTENTS Procedure

Data Set Name	CLOUD.US_NEWHIRE	Observations	10
Member Type	DATA	Variables	3
Engine	V9	Indexes	0
Created	03/09/2008 11:05:09	Observation Length	32
Last Modified	03/09/2008 11:05:09	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_32		
Encoding	wlatin1 Western (Windows)		
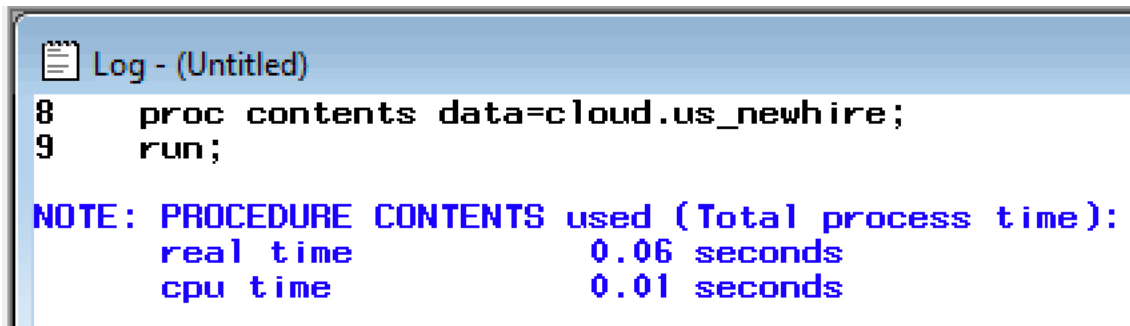
proc contents `cloud.us_newhire` output b

Engine/Host Dependent Information	
Data Set Page Size	4096
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	126
Obs in First Data Page	10
Number of Data Set Repairs	0
Filename	S:\Courses\stat-renaes\Stat426\data2\us_newhire.sas7bdat
Release Created	9.0101M3
Host Created	XP_PRO
Owner Name	AD\su-donm
File Size	5KB
File Size (bytes)	5120

proc contents `cloud.us_newhire` output c

Alphabetic List of Variables and Attributes			
#	Variable	Type	Len
3	Birthday	Char	9
1	ID	Char	15
2	Telephone	Num	8

proc contents `cloud.us_newhire` log

```
Log - (Untitled)
8       proc contents data=cloud.us_newhire;
9       run;

NOTE: PROCEDURE CONTENTS used (Total process time):
        real time           0.06 seconds
        cpu time            0.01 seconds
```

proc datasets `cloud library`

```
proc datasets;
contents data=cloud._all_ nodetails;
run;
```

proc datasets output

The SAS System

The DATASETS Procedure

Data Set Name	CLOUD.WEB_PRODUCTS2	Observations	20
Member Type	DATA	Variables	3
Engine	V9	Indexes	0
Created	03/11/2008 22:26:23	Observation Length	56
Last Modified	03/11/2008 22:26:23	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_32		
Encoding	wlatin1 Western (Windows)		

Engine/Host Dependent Information	
Data Set Page Size	8192
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	145

proc datasets log

Log - (Untitled)

```
25    proc datasets;
26    contents data=cloud._all_ nodetails;
27    run;
```