UCD School of Mathematics and Statistics

STAT40840: Data programming with SAS
Laura Kirwan

Lecture 2

1 Examining SAS Data Sets

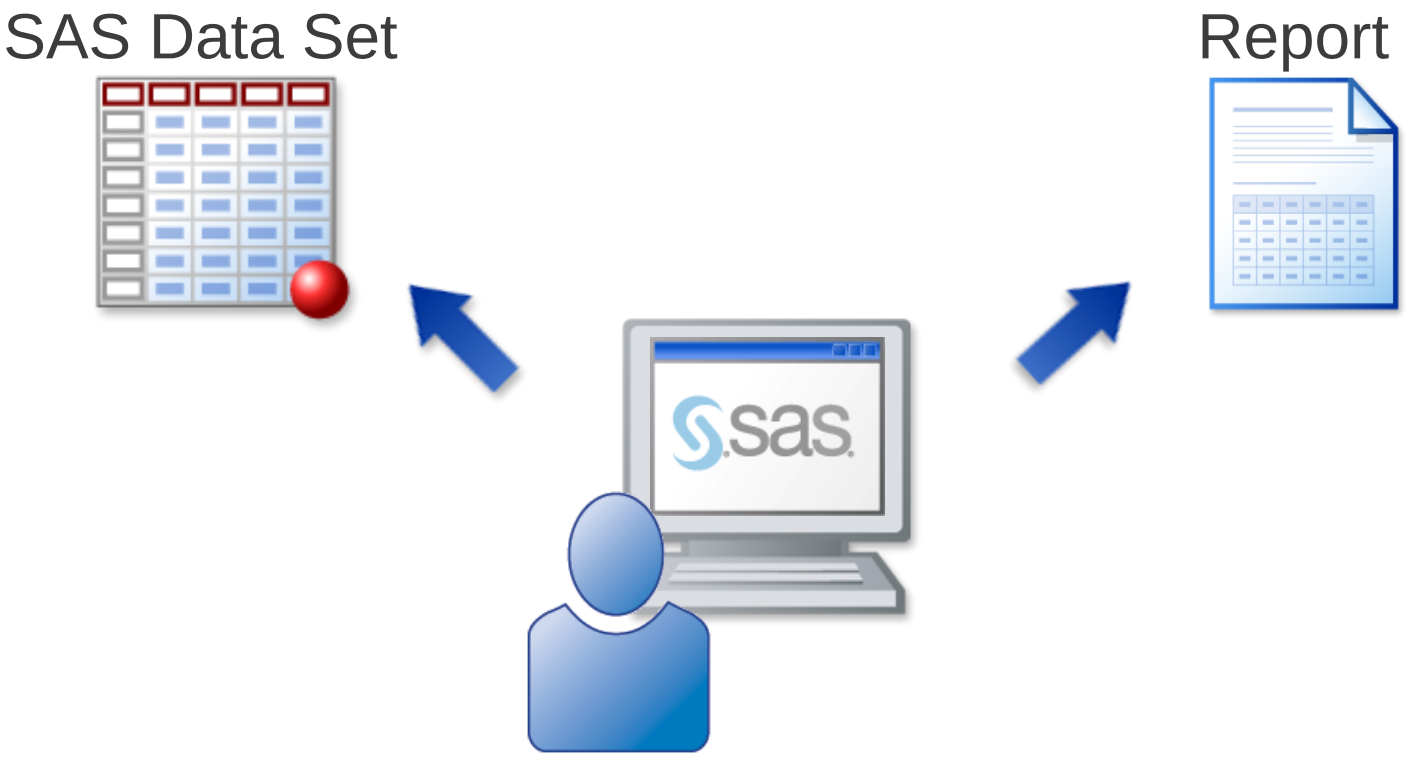
2 Accessing SAS Libraries

Examining SAS Data Sets

Objectives 1:

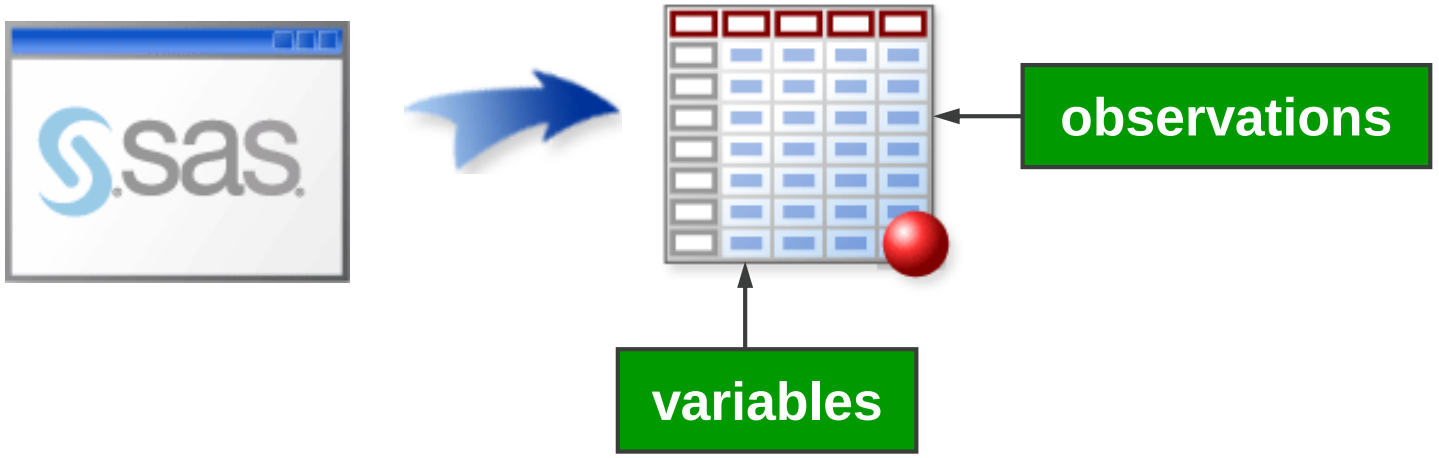
- Define the components of a SAS data set.
- Use the CONTENTS procedure to browse the descriptor portion of a SAS data set.
- Use the PRINT procedure to browse the data portion of a SAS data set.
- Define a SAS variable.
- Define a missing value.
- Define a SAS date value.

Many SAS data sets related to the Orion Star project already exist. We need to know how to display the structure and contents of the data sets.



What Is a SAS Data Set?

A *SAS data set* is a specially structured data file that SAS creates and that only SAS can read. A SAS data set is a table that contains observations and variables.

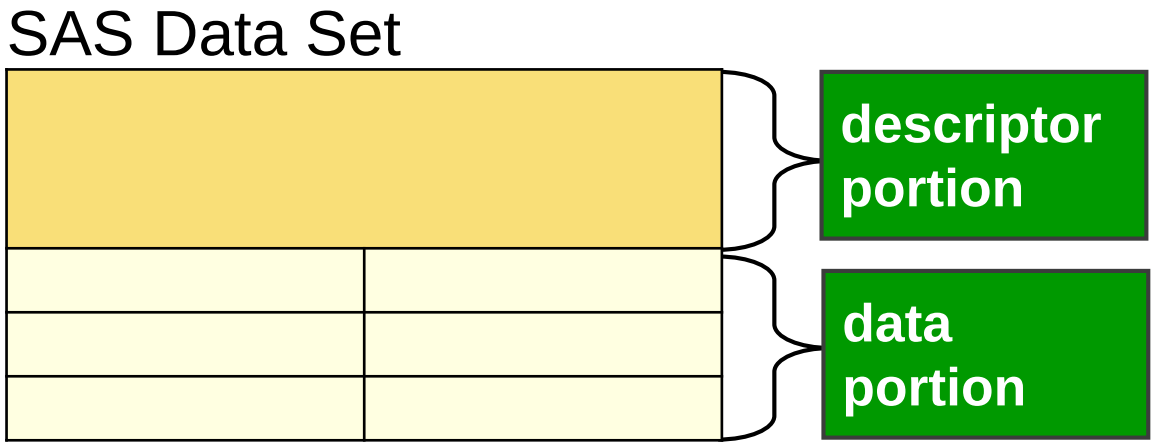


SAS Data Set Terminology

SAS Terminology		Database Terminology
SAS Data Set	⟷	Table
Observation	⟷	Row
Variable	⟷	Column

SAS Data Set Terminology

•A SAS data set contains a descriptor portion and a data portion.



Descriptor Portion

The *descriptor portion* contains the following metadata:

- general properties (such as data set name and number of observations)
- variable properties (such as name, type, and length)

Partial **work.newsalesemps**

Data Set Name	**WORK.NEWSALESEMPS**
Engine	**V9**
Created	**Mon, Feb 27, 2012 01:28 PM**
Observations	**71**
Variables	**4**
...	

First_Name	Last_Name	Job_Title	Salary
$ 12	$ 18	$ 25	N 8

general properties

variable properties

Browsing the Descriptor Portion

Use *PROC CONTENTS* to display the descriptor portion of a SAS data set.

```
proc contents data=work.newsalesemps;
run;
```

PROC CONTENTS DATA=*SAS-data-set*;
RUN;

L2_D1.sas

Viewing the Output

The CONTENTS Procedure

Data Set Name	WORK.NEWSALESEMPS	Observations	71
Member Type	DATA	Variables	4
Engine	V9	Indexes	0
Created	Mon, Feb 27, 2012 01:28:51 PM	Observation Length	64
Last Modified	Mon, Feb 27, 2012 01:28:51 PM	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO

Engine/Host Dependent Information

...

Alphabetic List of Variables and Attributes

#	Variable	Type	Len
1	First_Name	Char	12
3	Job_Title	Char	25
2	Last_Name	Char	18
4	Salary	Num	8

Exercise 1

Open program **L2_E1.sas**. Add a PROC CONTENTS step after the DATA step to view **work.donations**. Submit the program and review the results. How many observations are in the data set **work.donations**?

Data Portion

- The *data portion* of a SAS data set contains the data values, which are either character or numeric.

Partial **work.newsalesemps**

First_Name	Last_Name	Job_Title	Salary
Satyakam	Denny	Sales Rep. II	26780
Monica	Kletschkus	Sales Rep. IV	30890
Kevin	Lyon	Sales Rep. I	26955
Petrea	Soltau	Sales Rep. II	27440

variable names

data values

character values

numeric values

Browsing the Data Portion

Use *PROC PRINT* to display the data portion of a SAS data set.

```
proc print data=work.newsalesemps;
run;
```

PROC PRINT DATA=*SAS-data-set*;
RUN;

L2_D2.sas

Viewing the Output

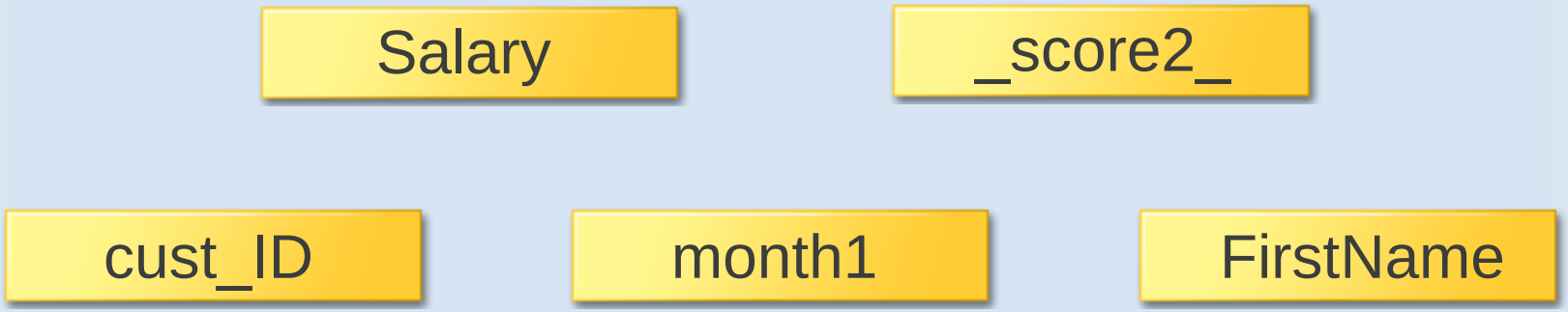
Partial PROC PRINT Output

Obs	First_Name	Last_Name	Job_Title	Salary
1	Satyakam	Denny	Sales Rep. II	26780
2	Monica	Kletschkus	Sales Rep. IV	30890
3	Kevin	Lyon	Sales Rep. I	26955
4	Petrea	Soltau	Sales Rep. II	27440
5	Marina	Iyengar	Sales Rep. III	29715

SAS Variable Names

SAS variable names

- can be 1 to 32 characters long.
- must start with a letter or underscore. Subsequent characters can be letters, underscores, or numerals.
- can be uppercase, lowercase, or mixed case.
- are not case sensitive.



Exercise 2

Which variable names are invalid?

a. data5mon
b. 5monthsdata
c. data#5
d. five months data
e. five_months_data
f. FiveMonthsData
g. fivemonthsdata

Data Types

A SAS data set supports two types of variables.

Character variables

- can contain any value: letters, numerals, special characters, and blanks
- range from 1 to 32,767 characters in length
- have 1 byte per character.

Numeric variables

- store numeric values using floating point or binary representation
- have 8 bytes of storage by default
- can store 16 or 17 significant digits.

Missing Data Values

- Missing values are valid values in a SAS data set.

Partial **work.newsalesemps**

First_Name	Last_Name	Job_Title	Salary
Monica	Kletschkus	Sales Rep. IV	.
Kevin	Lyon	Sales Rep. I	26955
Petrea	Soltau		27440

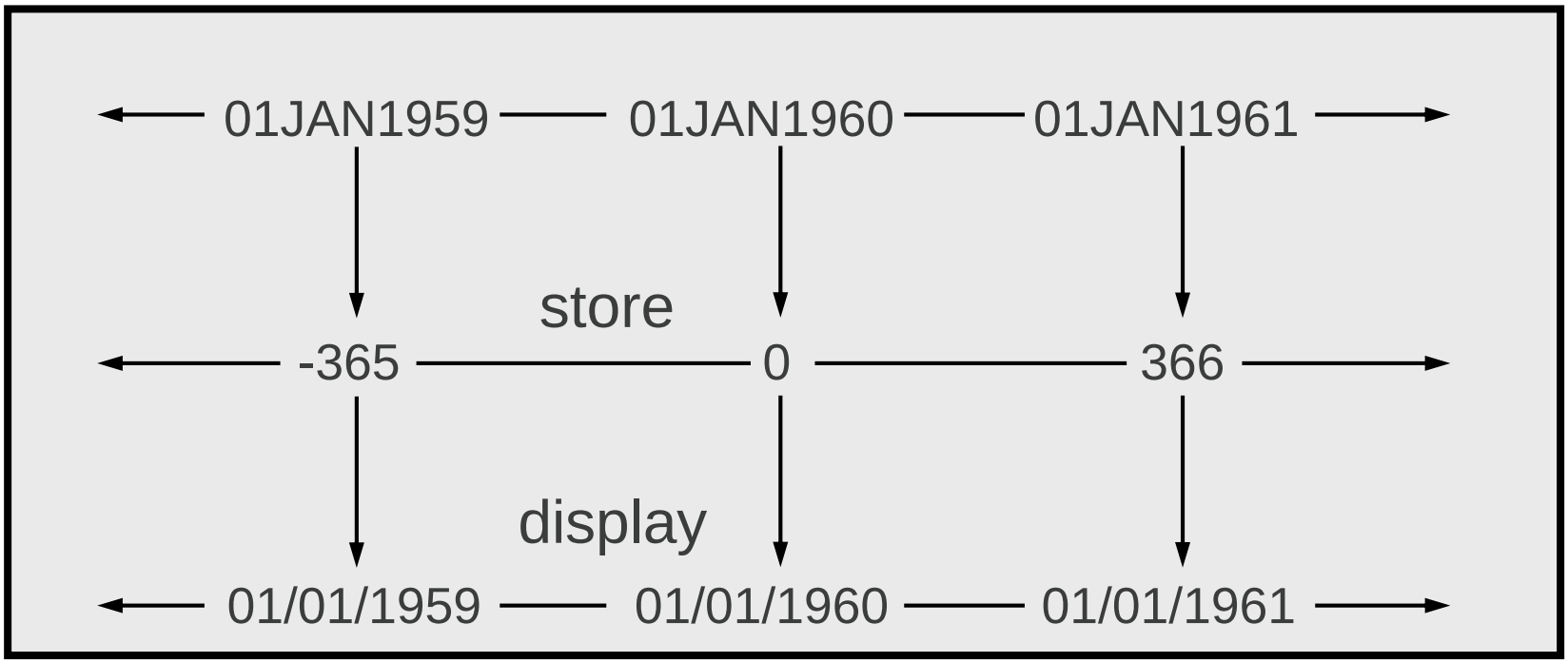
A blank represents a missing character value.

A period represents a missing numeric value.

A value must exist for every variable in every observation.

SAS Date Values

- SAS stores calendar dates as numeric values.



A *SAS date value* is stored as the number of days between January 1, 1960, and a specific date.

Exercise 3

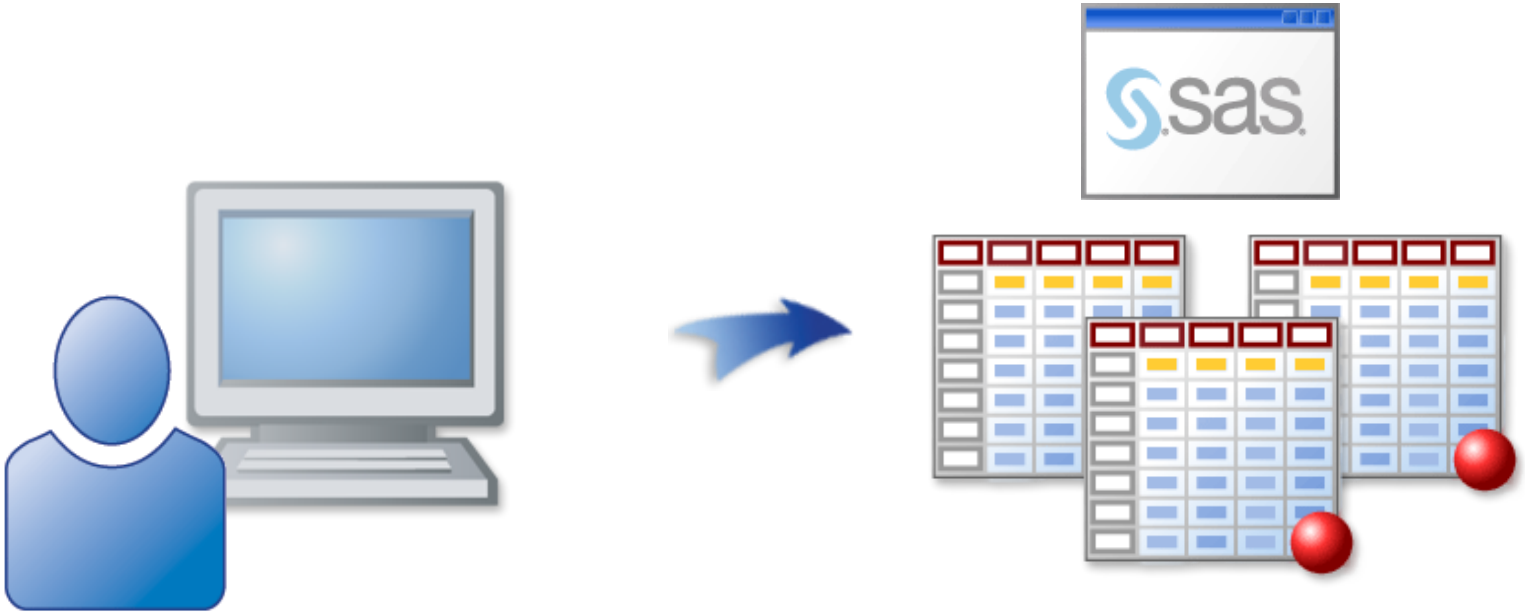
Submit program **L2_E3.sas**. View the output to retrieve the current date as a SAS date value (that is, a numeric value referencing January 1, 1960). What is the numeric value for today's date?

Accessing SAS Libraries

Objectives 2

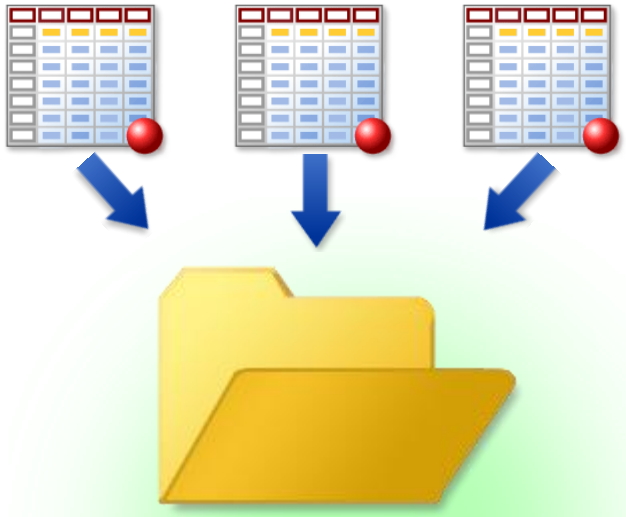
- Explain the concept of a SAS library.
- State the difference between a temporary library and a permanent library.
- Use a LIBNAME statement to assign a library reference name to a SAS library.
- Investigate a SAS library programmatically and interactively.

We need to access existing SAS data sets, so we need to understand how the data sets are stored in SAS.



SAS Libraries

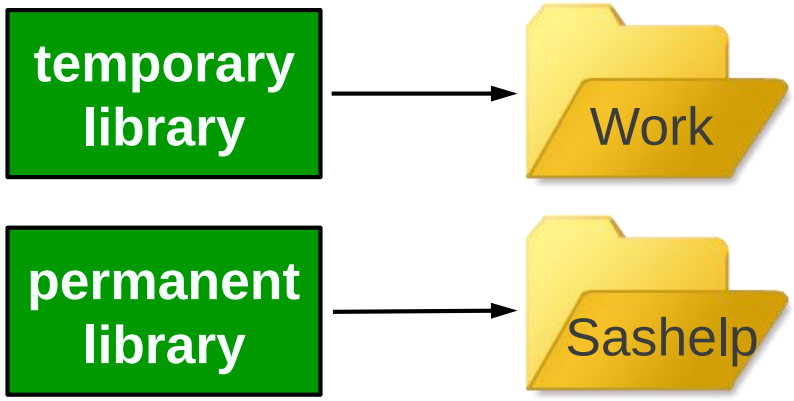
SAS data sets are stored in *SAS libraries*. A SAS library is a collection of SAS files that are referenced and stored as a unit.



A file can be stored in a temporary or permanent library.

How SAS Libraries Are Defined

When a SAS session starts, SAS creates one temporary and at least one permanent SAS library. These libraries are open and ready to be used.



You refer to a SAS library by a logical name called a library reference name, or *libref*.

Temporary Library

Work is a temporary library where you can store and access SAS data sets for the duration of the SAS session. It is the default library.

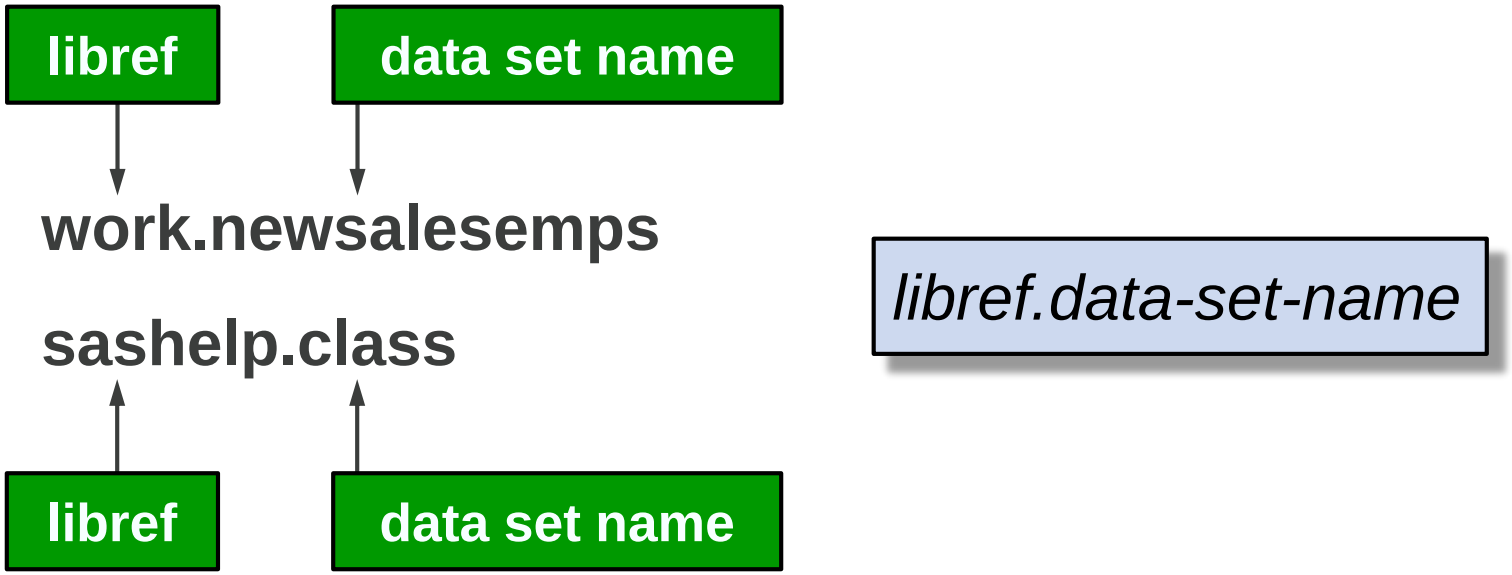
SAS deletes the **Work** library and its contents when the session terminates.

Permanent Libraries

Sashelp is a permanent library that contains sample SAS data sets you can access during your SAS session.

Accessing SAS Data Sets

All SAS data sets have a two-level name that consists of the libref and the data set name, separated by a period.



When a data set is in the temporary **Work** library, you can use a one-level name (for example, **newsalesemps**).

User-defined Permanent Libraries

We need to access and view SAS data sets that are stored in a permanent user-defined library.

User-Defined Libraries

A user-defined library

- is created by the user.
- is permanent. Data sets are stored until the user deletes them.
- is not automatically available in a SAS session.
- is implemented within the operating environment's file system.

User-Defined Libraries

Operating Environment	A SAS library is…	Example
Microsoft Windows	a folder	s:\workshop
UNIX	a directory	~/workshop
z/OS (OS/390)	a sequential file	*userid*.workshop.sasdata

The user must submit a SAS LIBNAME statement to associate a libref with the physical location of the library.

LIBNAME Statement

The SAS LIBNAME statement is a *global* SAS statement.

```
libname orion "s:\workshop";
```

LIBNAME *libref* "*SAS-library*" <*options*>**;**

- It is not required to be in a DATA step or PROC step.
- It does not require a RUN statement.
- It executes immediately.
- It remains in effect until changed or canceled, or until the session ends.

Use the location of ***your*** course data in your LIBNAME statement.

LIBNAME Statement

Partial SAS Log

```
47   libname orion "s:\workshop";
NOTE: Libref ORION was successfully assigned as follows:
      Engine:        V9
      Physical Name: s:\workshop
```

SAS files in **s:\workshop** are referenced using the **orion** libref.

orion.*data-set-name*

Changing or Canceling a Libref

A libref remains in effect until you change or cancel it, or until you end your SAS session.

To change a libref, submit a LIBNAME statement with the same libref but a different path.

```
libname orion "c:\myfiles";
```

To cancel a libref, submit a LIBNAME statement with the CLEAR option.

```
libname orion clear;
```

Exercise 4

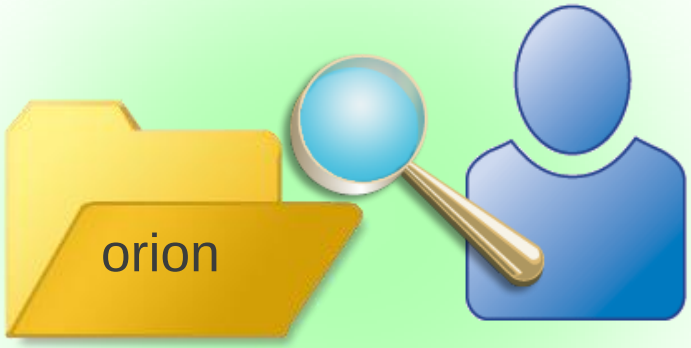
Which of the following correctly assigns the libref **myfiles** to a SAS library in the **c:\mysasfiles** folder?

a. libname orion myfiles "c:\mysasfiles";
b. libname myfiles "c:\mysasfiles";
c. libref orion myfiles "c:\mysasfiles";
d. libref myfiles "c:\mysasfiles";

Browsing a Library

You can browse a library

- programmatically using the CONTENTS procedure
- interactively in SAS Studio, SAS Enterprise Guide, or the SAS windowing environment.



Browsing a Library Programmatically

Use PROC CONTENTS with the _ALL_ keyword to generate a list of all SAS files in a library.

```
proc contents data=orion._all_ nods;
run;
```

PROC CONTENTS DATA=*libref*._ALL_ NODS;
RUN;

- _ALL_ requests all the files in the library.
- The NODS option suppresses the individual data set descriptor information.
- NODS can be used only with the keyword _ALL_.

L2_D3.sas

Viewing the Output

Partial PROC CONTENTS Output

The CONTENTS Procedure

Directory

Libref	ORION
Engine	V9
Physical Name	S:\workshop
Filename	S:\workshop

#	Name	Member Type	File Size	Last Modified
1	CHARITIES	DATA	9216	23Aug12:15:58:39
2	CONSULTANTS	DATA	5120	23Aug12:15:58:39
3	COUNTRY	DATA	17408	13Oct10:19:04:39
	COUNTRY	INDEX	17408	13Oct10:19:04:39
4	CUSTOMER	DATA	33792	04Nov11:09:52:27
5	CUSTOMER_DIM	DATA	33792	04Nov11:09:52:27