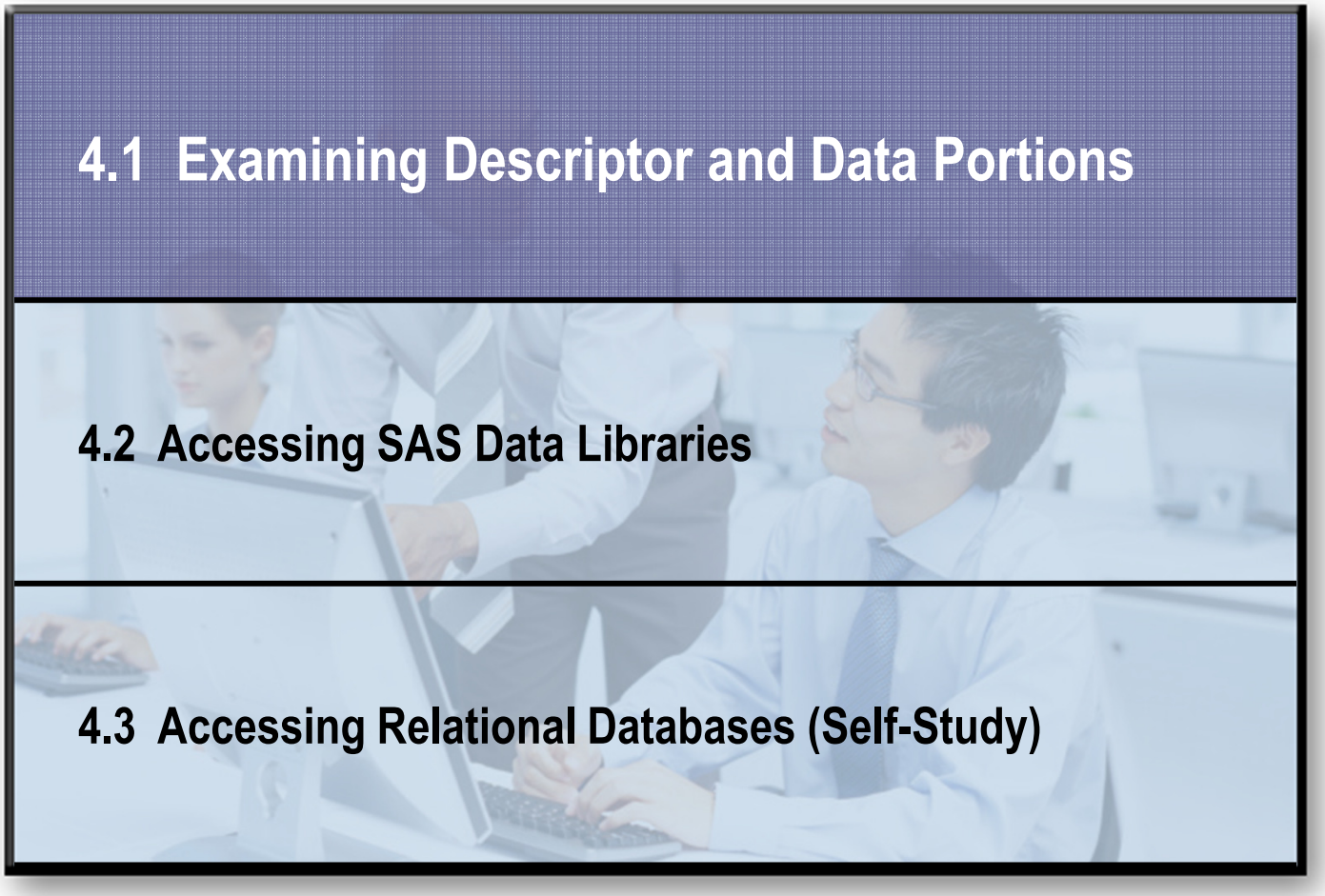
Chapter 4: Getting Familiar with SAS Data Sets



Objectives

- Define the components of a SAS data set.
- Define a SAS variable.
- Identify a missing value and a SAS date value.
- State the naming conventions for SAS data sets and variables.
- Browse the descriptor portion of SAS data sets by using the CONTENTS procedure.
- Browse the data portion of SAS data sets by using the PRINT procedure.

SAS Data Set

A *SAS data set* is a file that SAS creates and processes.

Partial `Work.NewSalesEmps`

Data Set Name	WORK.NEWSALESEMPS		
Engine	V9		
Created	Fri, Feb 08, 2008 01:40 PM		
Observations	71		
Variables	4		
...			
First_Name $ 12	Last_Name $ 18	Job_Title $ 25	Salary N 8
Satyakam	Denny	Sales Rep. II	26780
Monica	Kletschkus	Sales Rep. IV	30890
Kevin	Lyon	Sales Rep. I	26955
Petrea	Soltau	Sales Rep. II	27440

Descriptor Portion

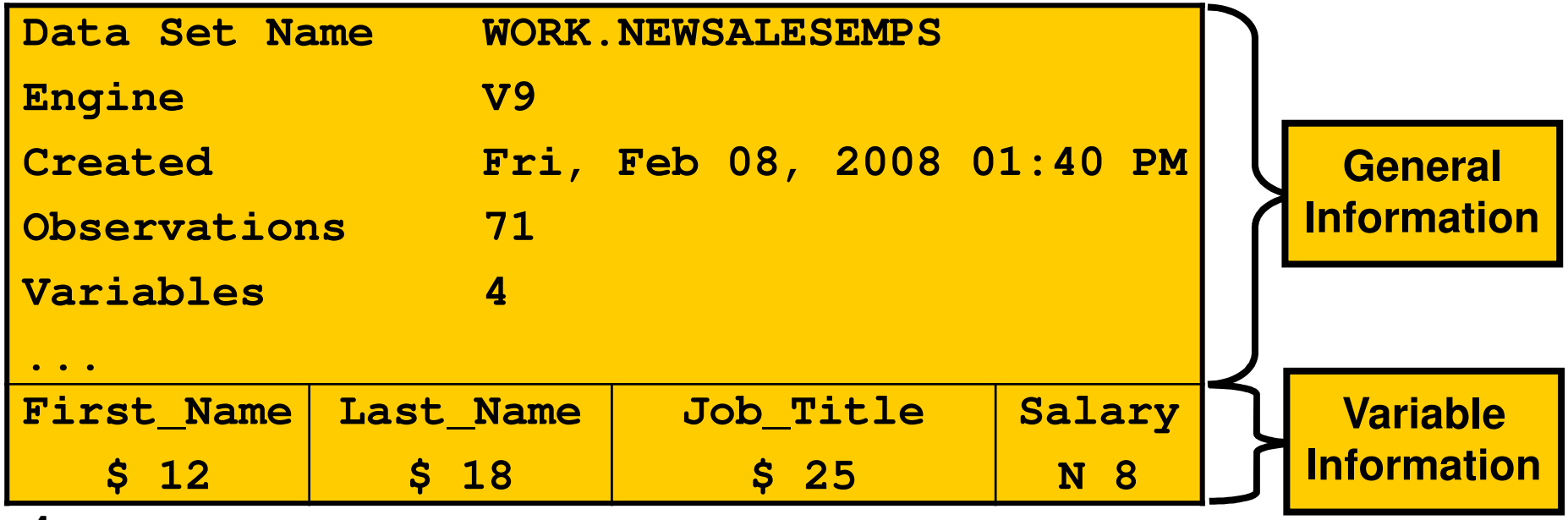
Data Portion

Descriptor Portion

The *descriptor portion* of a SAS data set contains the following:

- general information about the SAS data set (such as data set name and number of observations)
- variable information (such as name, type, and length)

Partial `Work.NewSalesEmps`



Browsing the Descriptor Portion

The *CONTENTS procedure* displays the descriptor portion of a SAS data set.

General form of the CONTENTS procedure:

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

Example:

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

Browsing the Descriptor Portion

Partial PROC CONTENTS Output

The CONTENTS Procedure

Data Set Name	WORK.NEWSALESEMPS	Observations	71
Member Type	DATA	Variables	4
Engine	V9	Indexes	0
Created	Wed, Jan 16, 2008 02:14:20 PM	Observation Length	64
Last Modified	Wed, Jan 16, 2008 02:14:20 PM	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			

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

4.01 Quiz

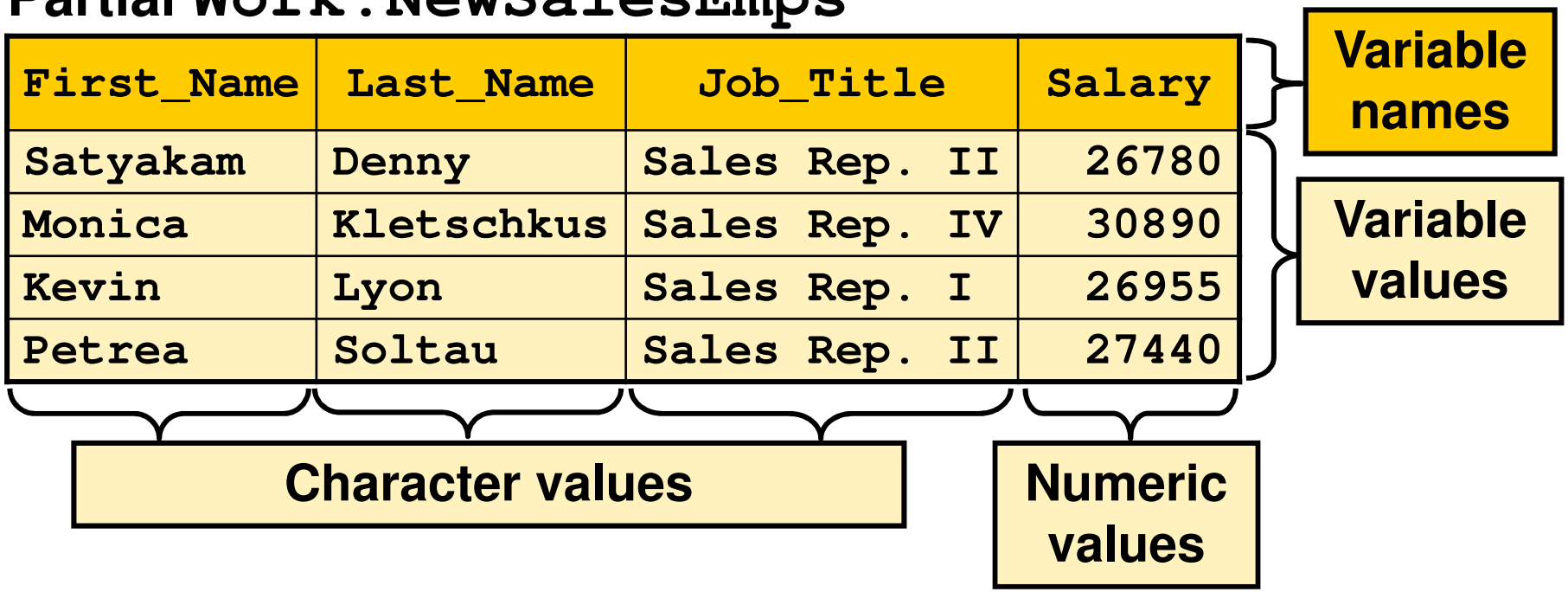
How many observations are in the data set **`Work.donations`**?

- Retrieve program **p104a01**.
- After the DATA step, add a PROC CONTENTS step to view the descriptor portion of **`Work.donations`**.
- Submit the program and review the results.

Data Portion

The *data portion* of a SAS data set is a rectangular table of character and/or numeric data values.

Partial `Work.NewSalesEmps`



The data values are organized as a table of observations (rows) and variables (columns).

SAS Variable Values

There are two types of variables:

character	Contain any value: letters, numbers, special characters, and blanks. Character values are stored with a length of 1 to 32,767 bytes. One byte equals one character.
numeric	Stored as floating point numbers in 8 bytes of storage by default. Eight bytes of floating point storage provide space for 16 or 17 significant digits. You are not restricted to 8 digits.

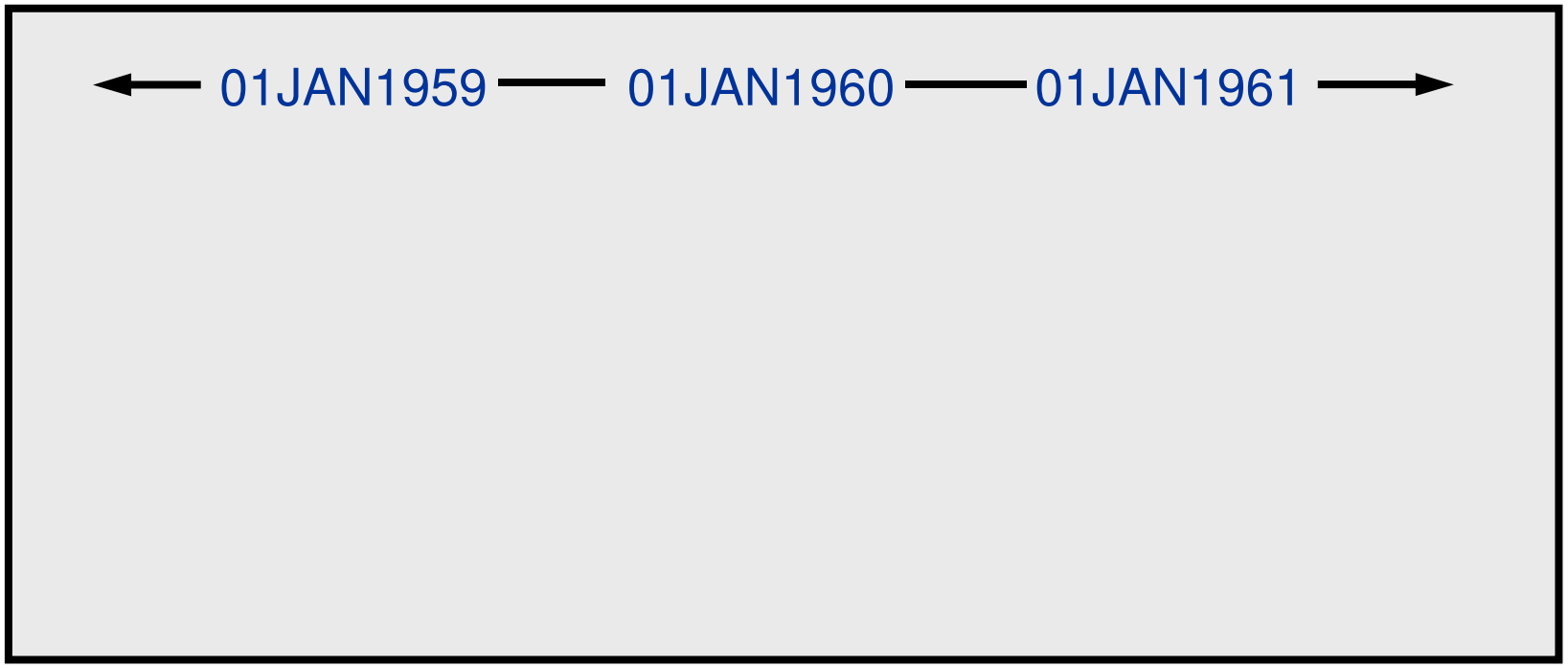
4.02 Multiple Choice Poll

Which variable type do you think SAS uses to store date values?

a. character
b. numeric

SAS Date Values

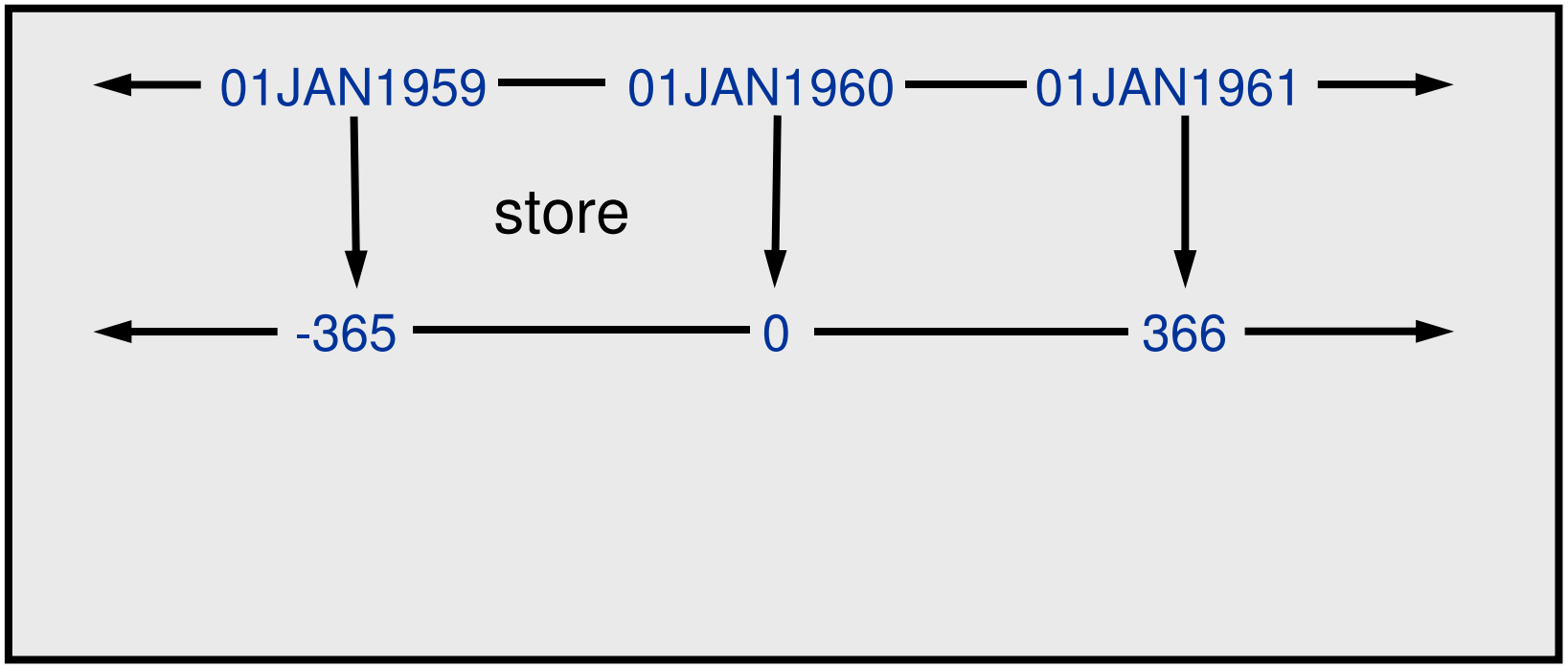
SAS stores date values as numeric values.



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

SAS Date Values

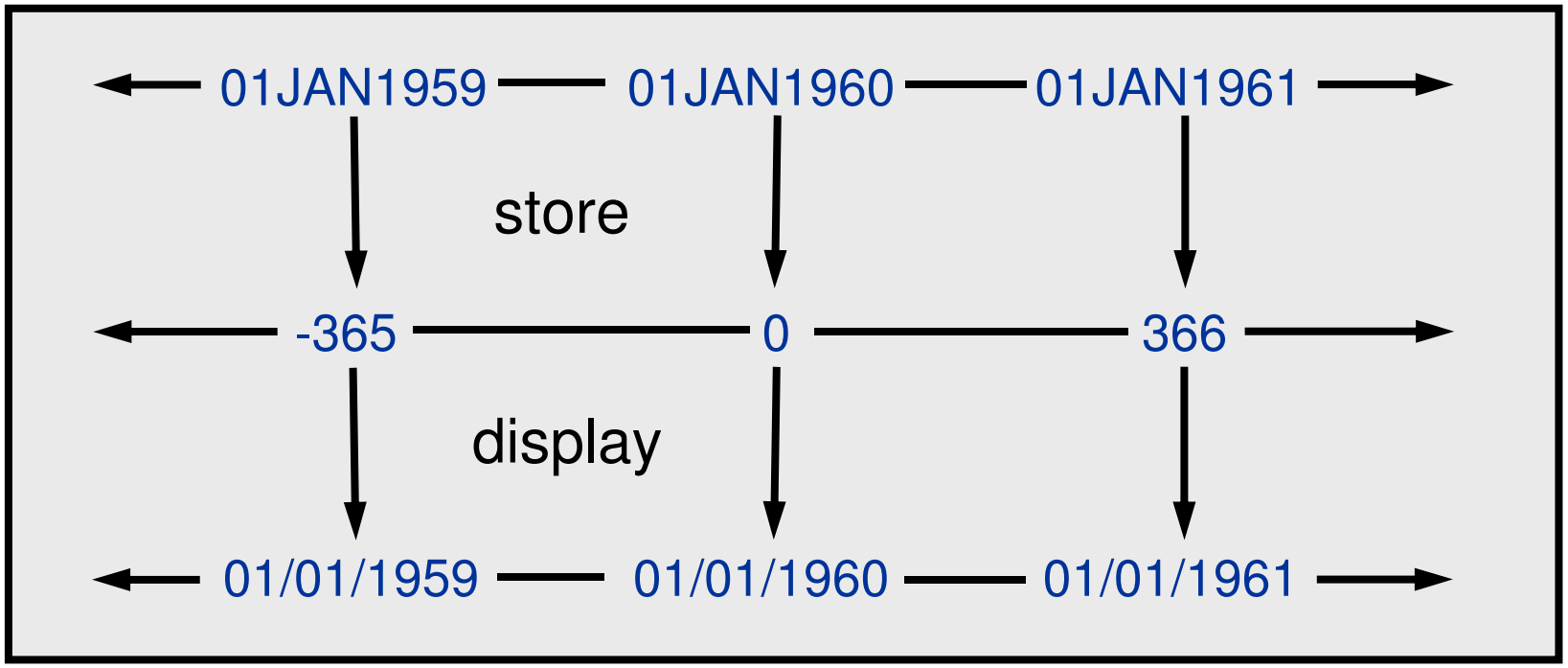
SAS stores date values as numeric values.



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

SAS Date Values

SAS stores date values as numeric values.



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

4.03 Quiz

What is the numeric value for today's date?

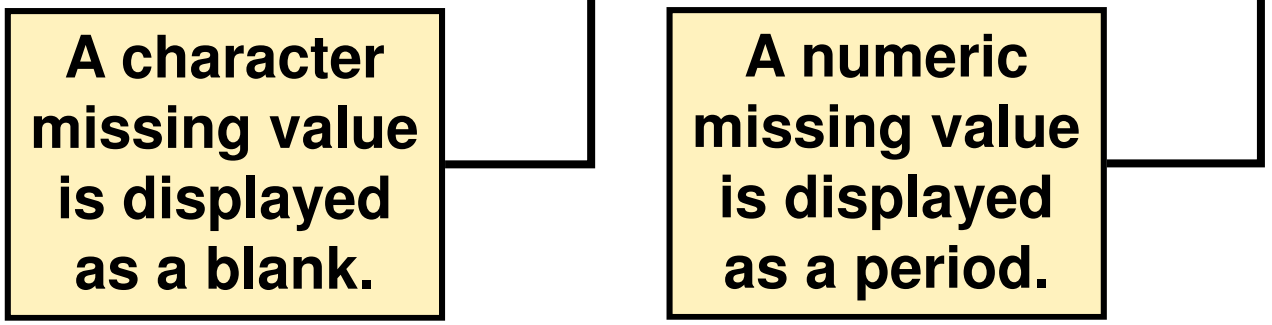
- Submit program **p104a02**.
- View the output to retrieve the current date as a numeric value referencing January 1, 1960.

Missing Data Values

A value must exist for every variable for each observation. Missing values are valid values in a SAS data set.

Partial `Work.NewSalesEmps`

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



SAS Data Set and Variable Names

SAS names have these characteristics:

- can be 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.

4.04 Multiple Answer Poll

Which variable names are valid?

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

SAS Data Set Terminology

Comparable Terminology:

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

- The terminology of data set, observation, and variable is specific to SAS.
- The terminology of table, row, and column is common among databases.

Browsing the Data Portion

The *PRINT procedure* displays the data portion of a SAS data set.

By default, PROC PRINT displays the following:

- all observations
- all variables
- an Obs column on the left side

Browsing the Data Portion

General form of the PRINT procedure:

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

Example:

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

Browsing the Data Portion

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
6	Shani	Duckett	Sales Rep. I	25795
7	Fang	Wilson	Sales Rep. II	26810
8	Michael	Minas	Sales Rep. I	26970
9	Amanda	Liebman	Sales Rep. II	27465
10	Vincent	Eastley	Sales Rep. III	29695
11	Viney	Barbis	Sales Rep. III	30265
12	Skev	Rusli	Sales Rep. II	26580
13	Narelle	James	Sales Rep. III	29990
14	Gerry	Snellings	Sales Rep. I	26445
15	Leonid	Karavdic	Sales Rep. II	27860

Browsing the Data Portion

Options and statements can be added to the PRINT procedure.

```
PROC PRINT DATA=SAS-data-set NOOBS;
      VAR variable(s);
RUN;
```

- The NOOBS option suppresses the observation numbers on the left side of the report.
- The VAR statement selects variables that appear in the report and determines their order.

Browsing the Data Portion

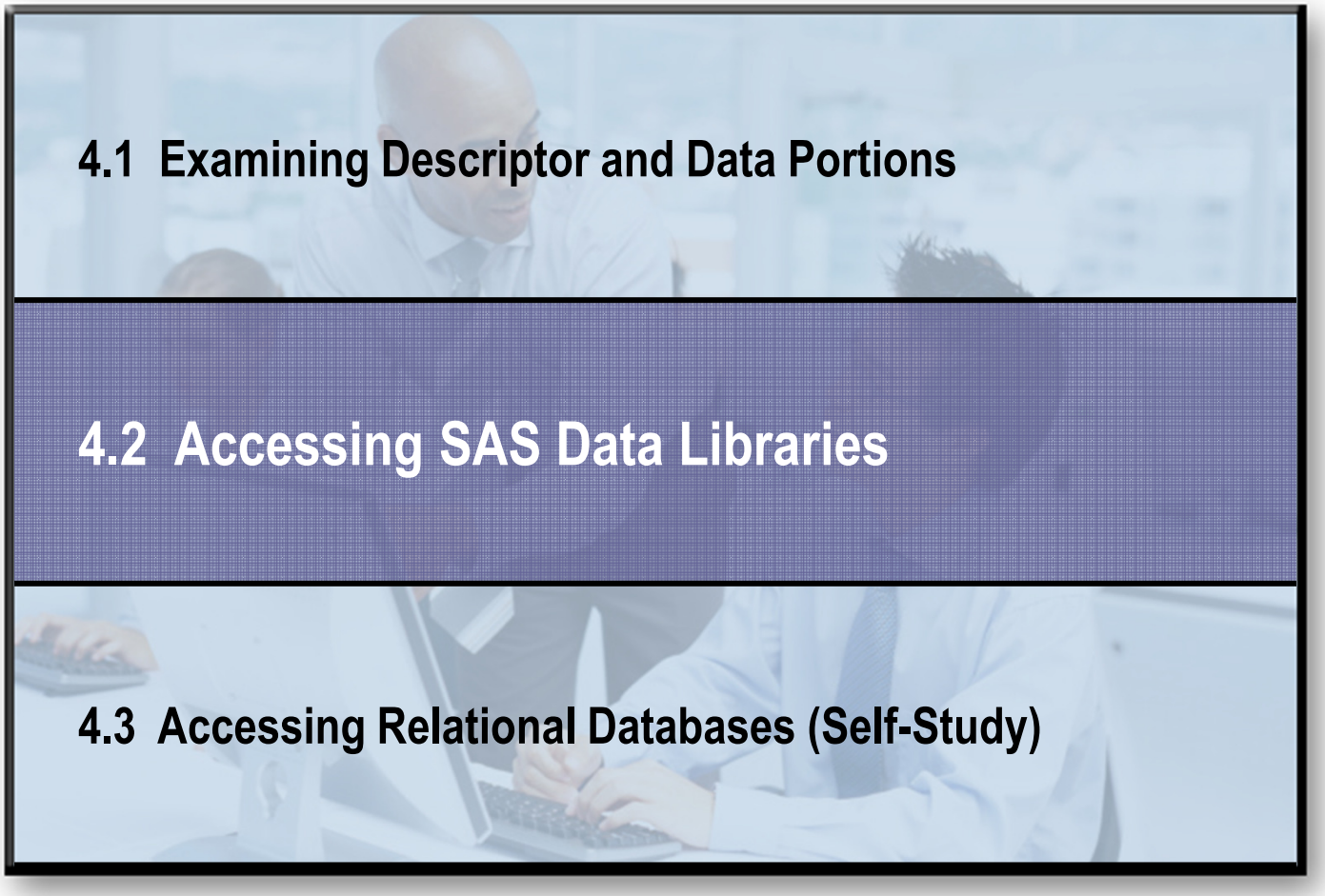
```
proc print data=work.NewSalesEmps noobs;
    var Last_Name First_Name Salary;
run;
```

Partial PROC PRINT Output

Last_Name	First_Name	Salary
Denny	Satyakam	26780
Kletschkus	Monica	30890
Lyon	Kevin	26955
Soltau	Petrea	27440
Iyengar	Marina	29715
Duckett	Shani	25795
Wilson	Fang	26810
Minas	Michael	26970
Liebman	Amanda	27465
Eastley	Vincent	29695

p104d03

Chapter 4: Getting Familiar with SAS Data Sets



Objectives

- Explain the concept of a SAS data library.
- Assign a library reference name to a SAS data library by using the LIBNAME statement.
- State the difference between a permanent library and a temporary library.
- Browse the contents of a SAS data library by using the SAS Explorer window.
- Investigate a SAS data library by using the CONTENTS procedure.

SAS Data Libraries

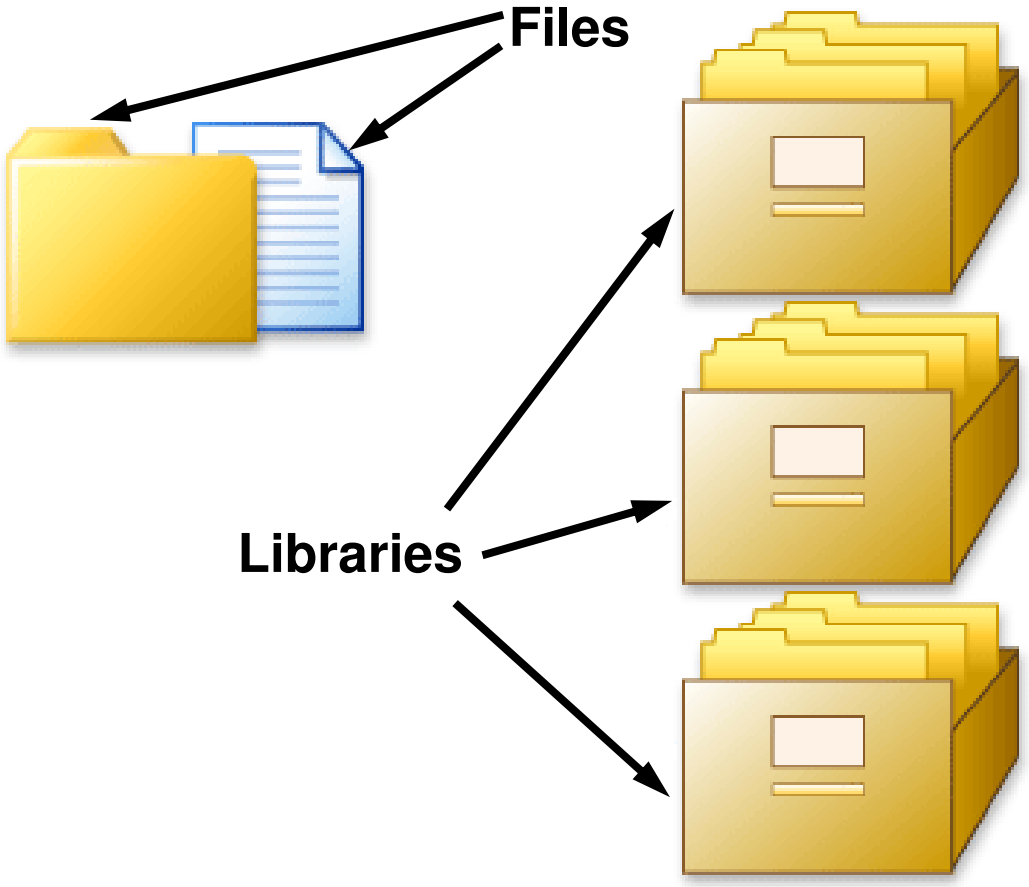
A *SAS data library* is a collection of SAS files that are recognized as a unit by SAS.

Directory-based System	A SAS data library is a directory.
Windows Example:	`s:\workshop`
UNIX Example:	`/users/userid`

z/OS (OS/390)	A SAS data library is an operating system file.
z/OS (OS/390) Example:	`userid.workshop.sasdata`

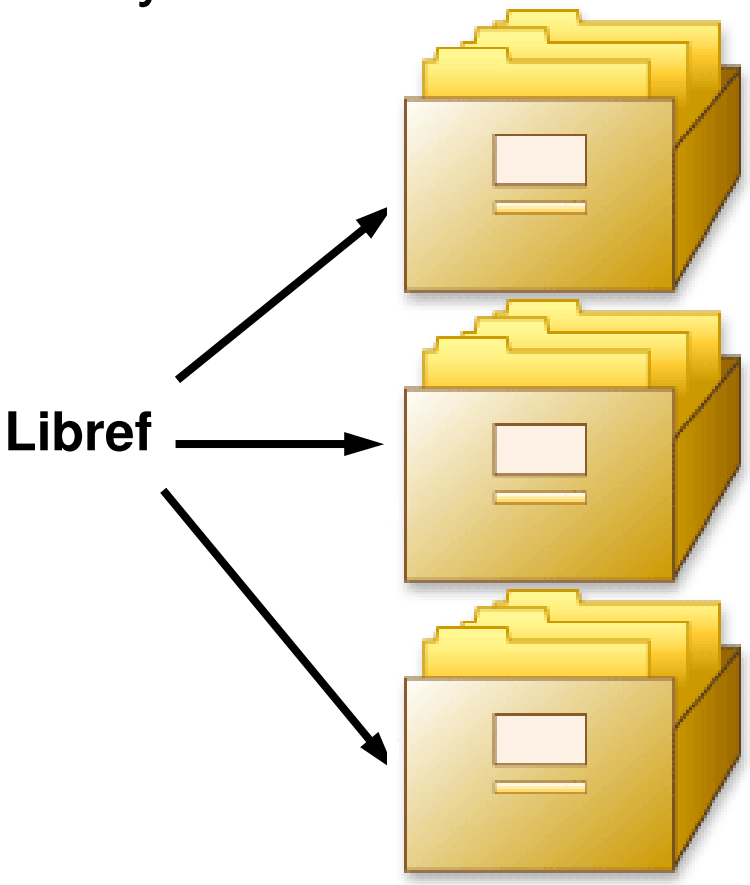
SAS Data Libraries

You can think of a SAS data library as a drawer in a filing cabinet and a SAS data set as one of the file folders in the drawer.



Assigning a Libref

Regardless of which host operating system you use, you identify SAS data libraries by assigning a *library reference name (libref)* to each library.



SAS Data Libraries

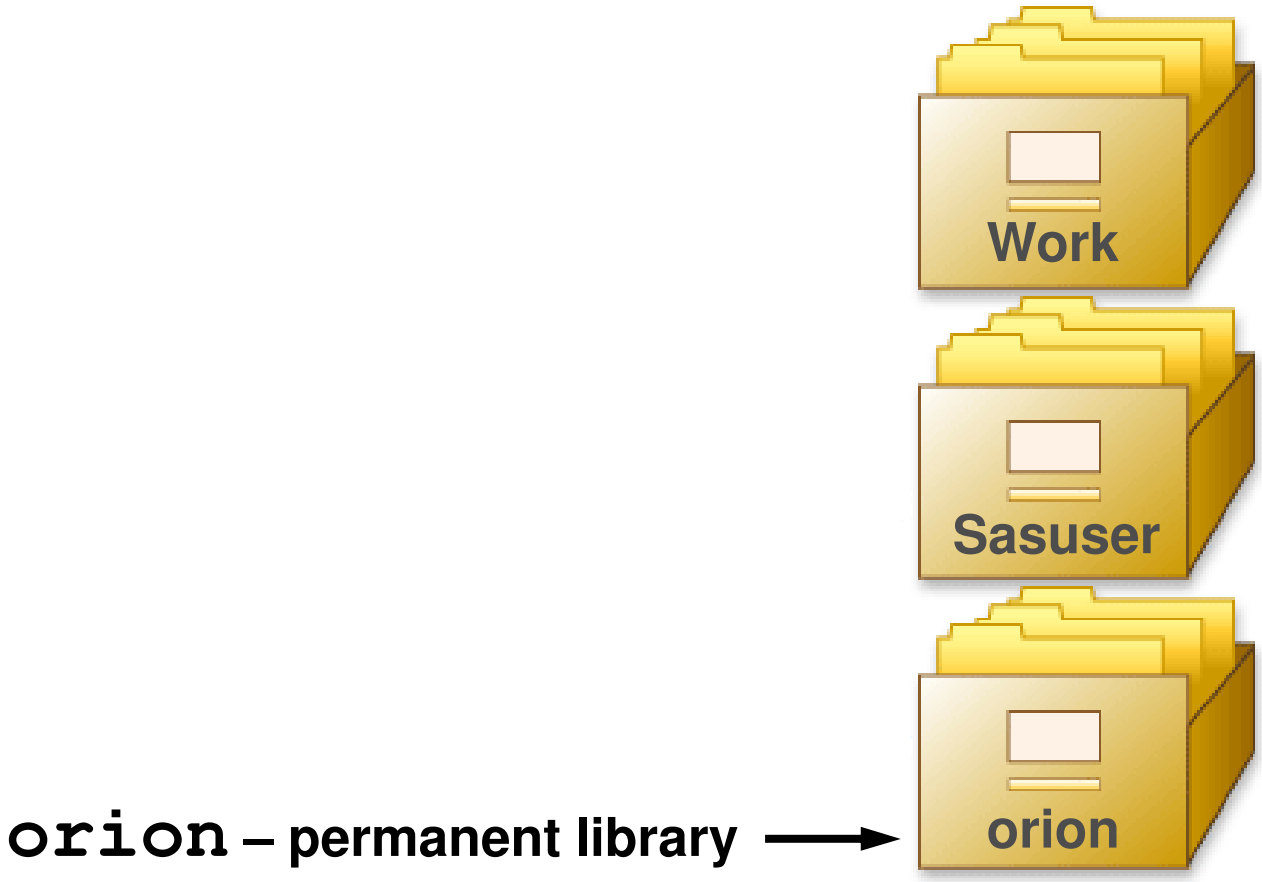
When a SAS session starts, SAS automatically creates one temporary and at least one permanent SAS data library that you can access.

Work – **temporary library**

Sasuser – **permanent library**

SAS Data Libraries

You can also create and access your own permanent libraries.



Assigning a Libref

You can use the *LIBNAME statement* to assign a library reference name (libref) to a SAS data library.

General form of the LIBNAME statement:

LIBNAME *libref 'SAS-data-library' <options>*;

Rules for naming a libref:

- The name must be 8 characters or less.
- The name must begin with a letter or underscore.
- The remaining characters must be letters, numerals, or underscores.

Assigning a Libref

Examples:

Windows

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

UNIX

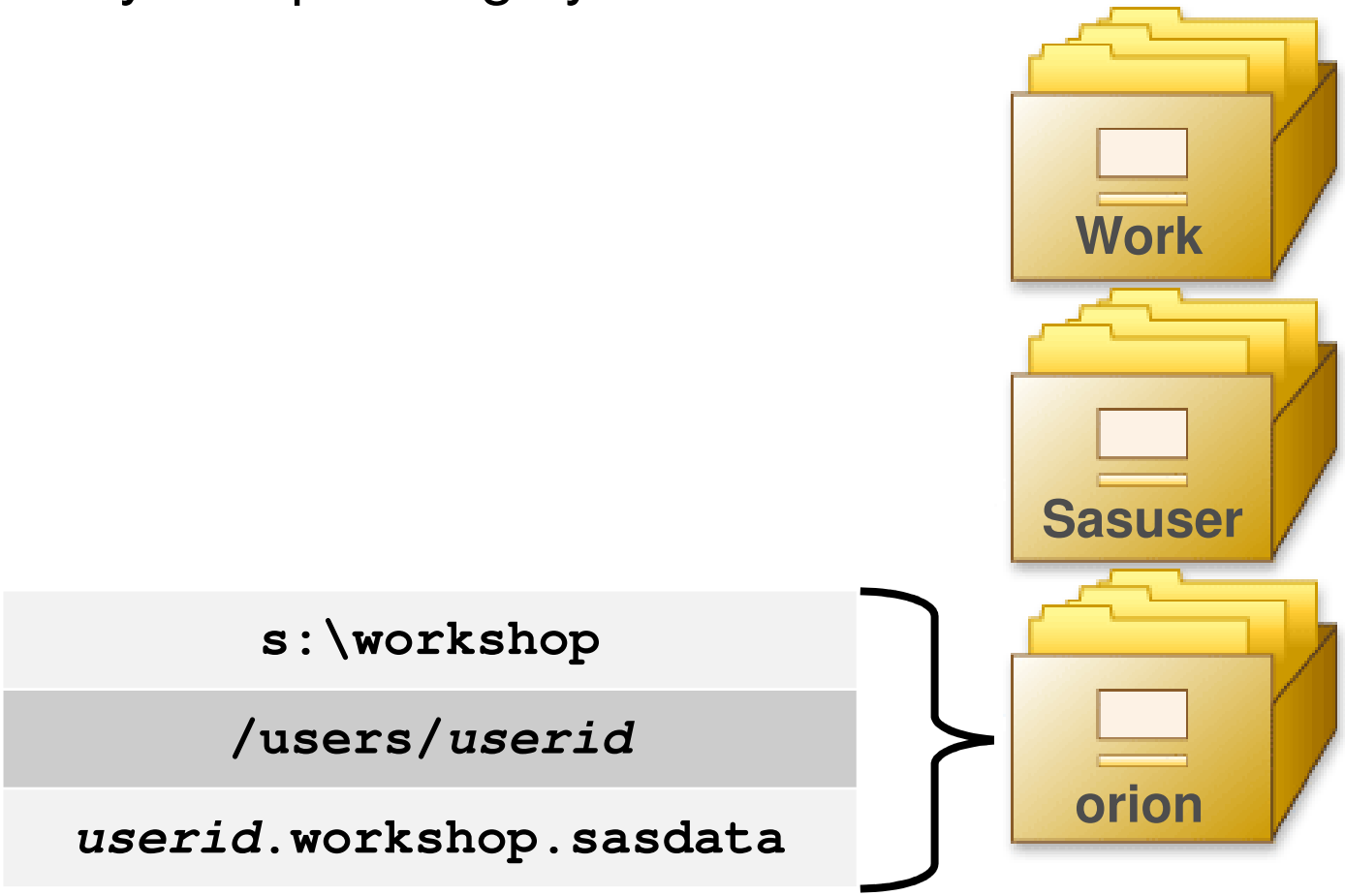
```
libname orion '/users/userid';
```

z/OS (OS/390)

```
libname orion 'userid.workshop.sasdata';
```

Making the Connection

When you submit the LIBNAME statement, a connection is made between a libref in SAS and the physical location of files on your operating system.



4.05 Poll

During an interactive SAS session, every time that you submit a program you must also resubmit the LIBNAME statement.

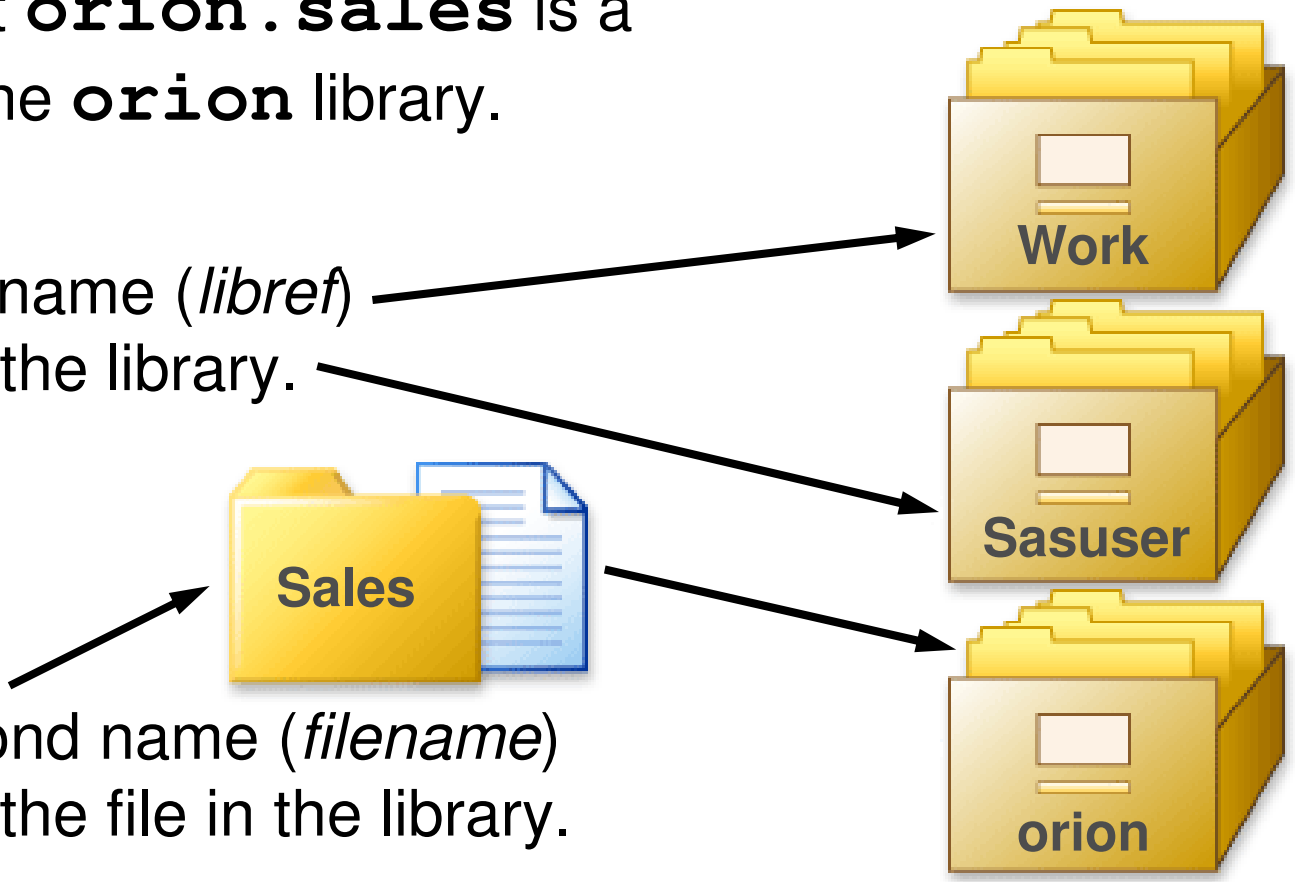
- ❍ True
- ❍ False

Two-Level SAS Filenames

Every SAS file has a two-level name: *libref.filename*

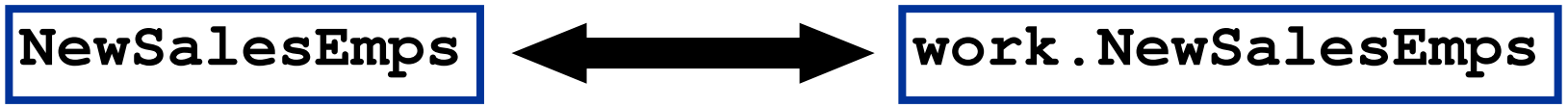
The data set **`orion.sales`** is a SAS file in the **`orion`** library.

- The first name (*libref*) refers to the library.
- The second name (*filename*) refers to the file in the library.



Temporary SAS Filename

The default libref is Work if the libref is omitted.




```
data NewSalesEmps;
   length First_Name $ 12
          Last_Name $ 18 Job_Title $ 25;
   infile 'newemps.csv' dlm=',';
   input First_Name $ Last_Name $
         Job_Title $ Salary;
run;

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

Browsing a SAS Data Library

The *SAS Explorer* enables you to manage your files in the windowing environment.

In the SAS Explorer, you can do the following:

- view a list of all the libraries available during your current SAS session
- navigate to see all members of a specific library
- display the descriptor portion of a SAS data set

Browsing a SAS Data Library

The CONTENTS procedure with the _ALL_ keyword produces a list of all the SAS files in the data library.

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

- The NODS option suppresses the descriptor portions of the data sets.
- NODS is only used in conjunction with the keyword _ALL_.

5. Reviewing Concepts

a. SAS statements usually begin with an _____.

b. Every SAS statement ends with a _____.

c. The descriptor portion of a SAS data set can be viewed using the _____ procedure.

d. Character variable values can be up to _____ characters long and use _____ byte(s) of storage per character.

e. By default, numeric variables are stored in ___ bytes of storage.

continued...

5. Reviewing Concepts

f. The internally stored SAS date value for January 3, 1960, is _____.

g. A SAS variable name has _____ to _____ characters and begins with a _____ or an _____.

h. A missing character value is displayed as a _____.

i. A missing numeric value is displayed as a _____.

j. When a SAS session starts, SAS automatically creates the temporary library called _____.

continued...

5. Reviewing Concepts

k. A libref name must be ____ characters or less.

l. What are the two kinds of steps? _____

m. What are the three primary windows in the SAS windowing environment? _____

n. What are the two portions of every SAS data set?

_____

o. What are the two types of variables?

_____

continued...

5. Reviewing Concepts

p. True or False: If a SAS program produces output, then the program ran successfully and there is no need to check the SAS log. _____

q. True or False: There are two methods for commenting in a SAS program. _____

r. True or False: Omitting a semicolon never causes errors. _____

continued...

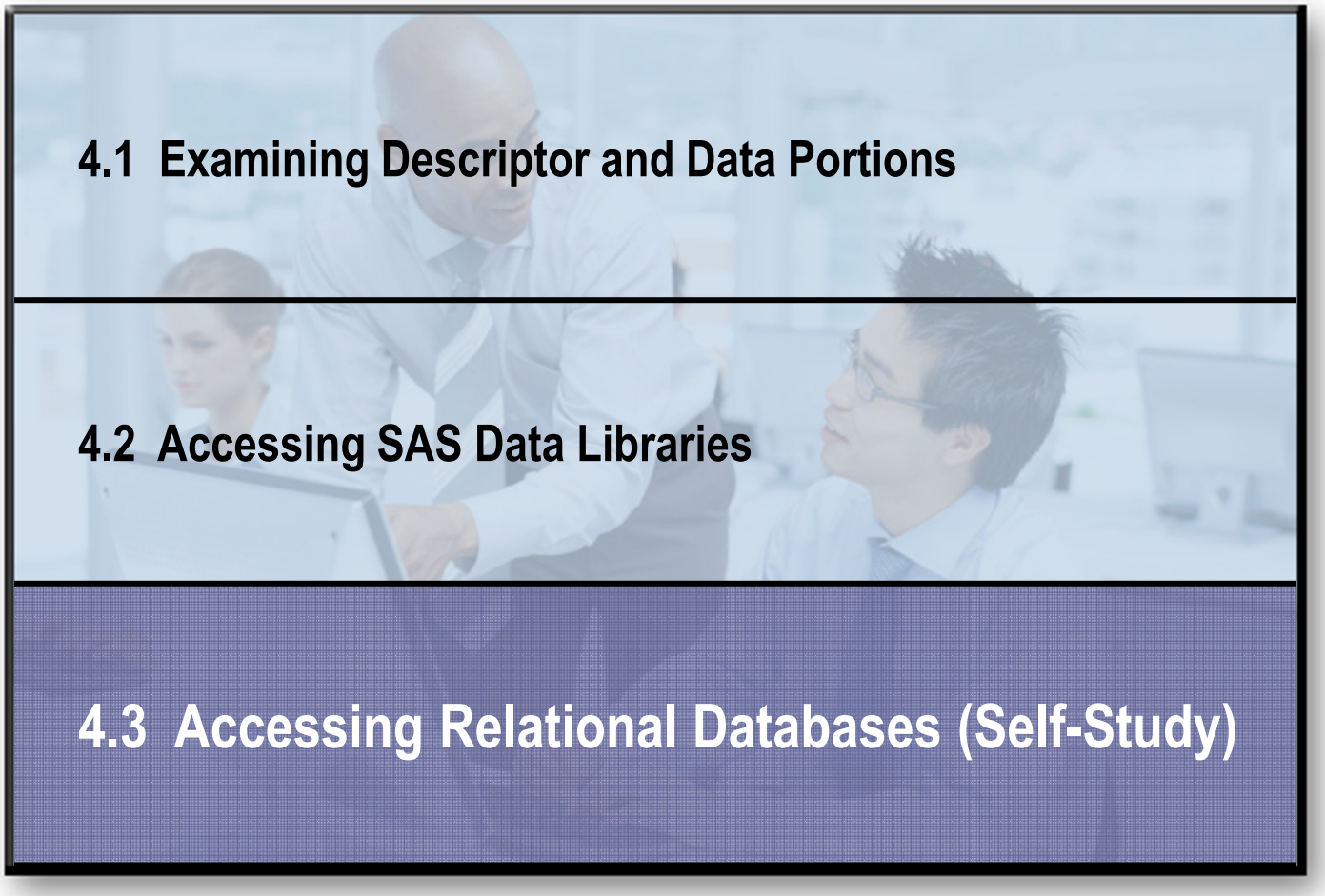
5. Reviewing Concepts

s. True or False: A library reference name (libref) references a particular data set. _____

t. True or False: If a data set is referenced with a one level name, **`Work`** is the implied libref. _____

u. True or False: The _ALL_ keyword is used with the PRINT procedure. _____

Chapter 4: Getting Familiar with SAS Data Sets



Objectives

- Assign a library reference name to a relational database by using the LIBNAME statement.
- Reference a relational database table using a SAS two-level name.

The LIBNAME Statement (Review)

The *LIBNAME statement* assigns a library reference name (libref) to a SAS data library.

General form of the LIBNAME statement:

LIBNAME *libref 'SAS-data-library' <options>*;

The SAS/ACCESS LIBNAME Statement

The *SAS/ACCESS LIBNAME statement* assigns a library reference name (libref) to a relational database.

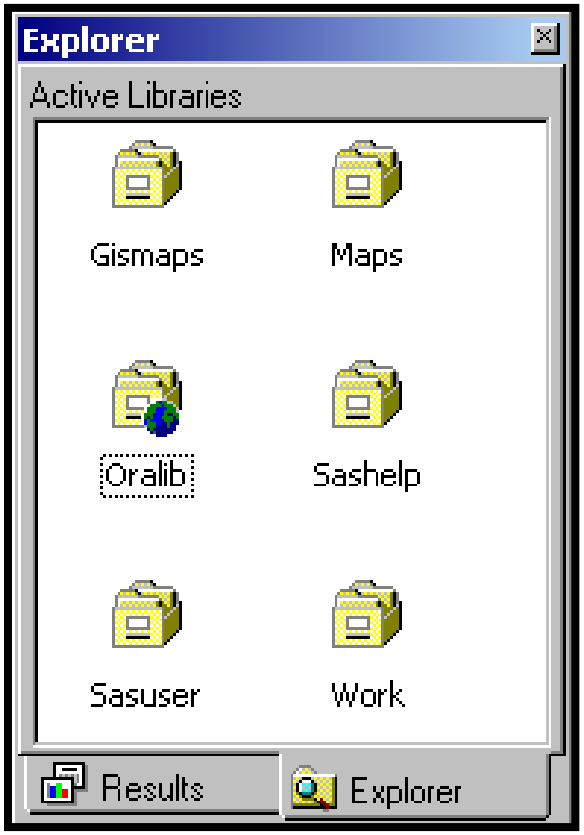
General form of the SAS/ACCESS LIBNAME statement:

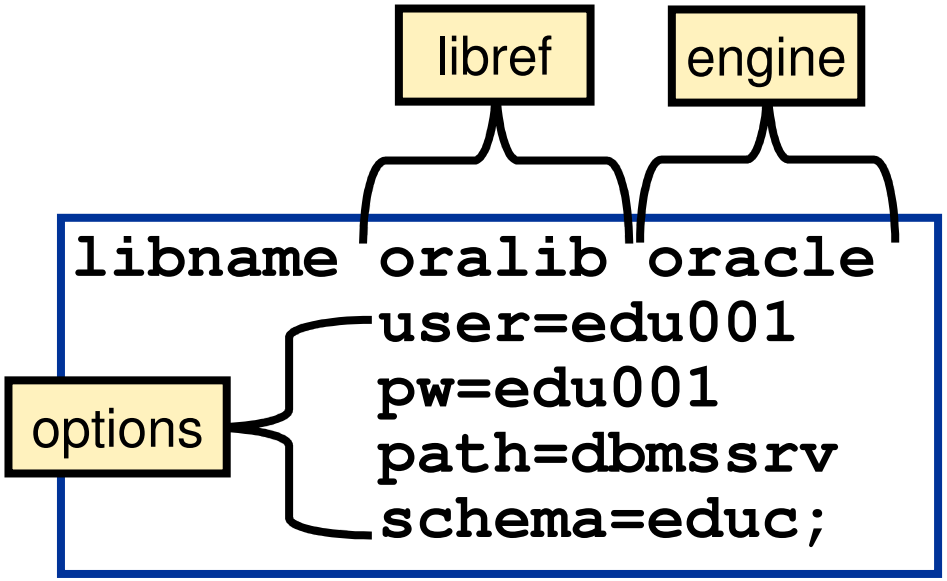
LIBNAME *libref engine-name <SAS/ACCESS-options>*;

After a database is associated with a libref, you can use a SAS two-level name to specify any table in the database and then work with the table as you would with a SAS data set.

Oracle Example

This example uses the LIBNAME statement as supported in the SAS/ACCESS interface to Oracle.

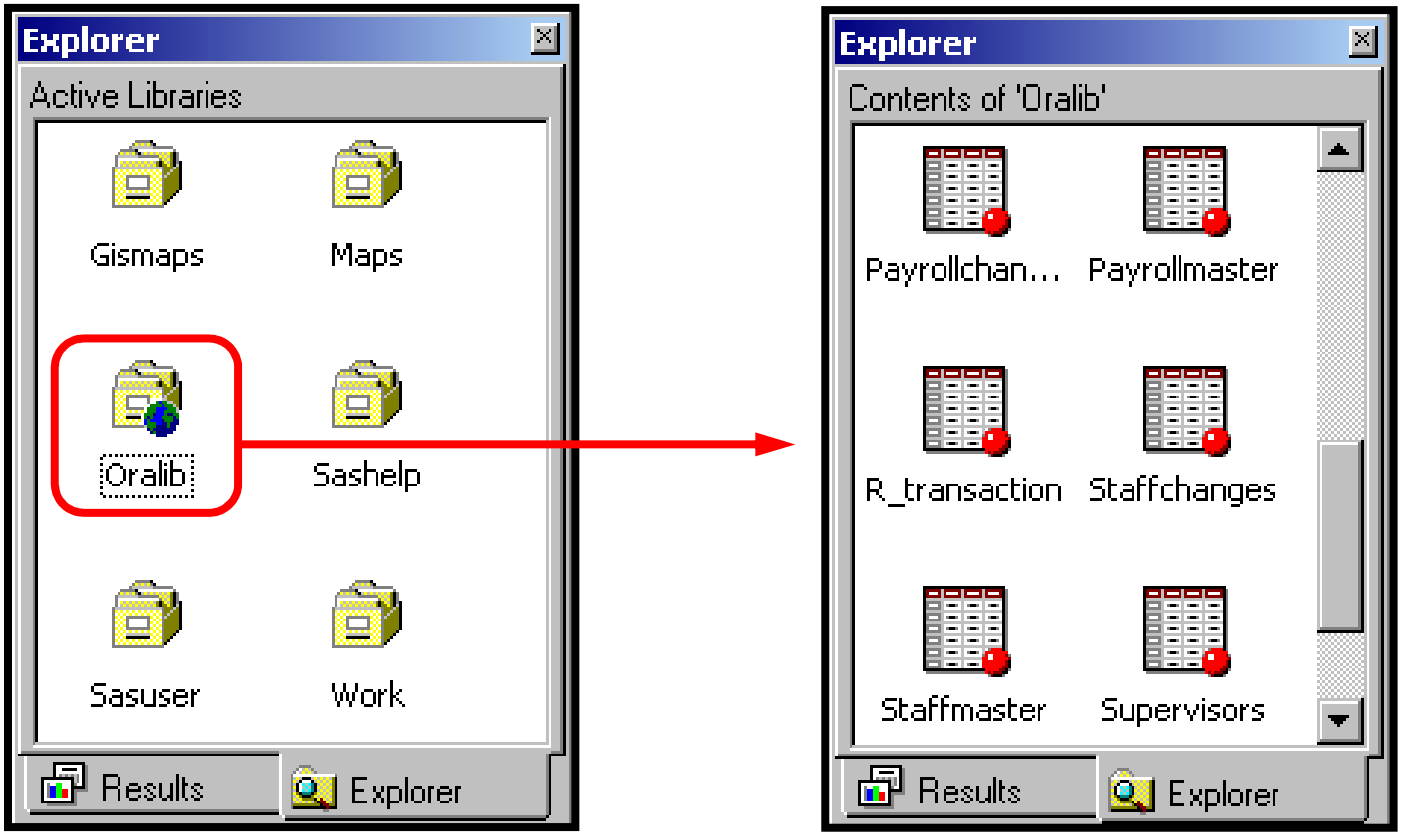




p104d05

Oracle Example

Any table in this Oracle database can be referenced using a SAS two-level name.



Oracle Example

```
libname oralib oracle
        user=edu001 pw=edu001
        path=dbmssrv schema=educ;

proc print data=oralib.supervisors;
run;

data work.staffpay;
   merge oralib.staffmaster
         oralib.payrollmaster;
   by empid;
run;

libname oralib clear;
```

4.06 Quiz

Which option in the LIBNAME statement specifies a user's password when accessing an Informix database?

Documentation on SAS/ACCESS for Informix can be found in the SAS Help and Documentation from the Contents tab (**SAS Products** ⇨ **SAS/ACCESS** ⇨ **SAS/ACCESS 9.2 for Relational Databases Reference** ⇨ **DBMS-Specific Reference** ⇨ **SAS/ACCESS for Informix** ⇨ **LIBNAME Statement Specifics for Informix**).

Chapter Review

1. What structure is a SAS data library in UNIX or Windows?

2. What structure is a SAS data library in z/OS (OS/390)?

3. What is the name of the permanent SAS data library that SAS creates for you?

4. What can you do with the SAS Explorer?

5. What window enables you to interactively browse a SAS data set?