# Module Summary: Accessing Data

## Understanding SAS Data

- SAS data sets have a data portion and a descriptor portion.

- SAS columns must have a name, type, and length.

- Column names can be 1-32 characters, must start with a letter or underscore and continue with letters, numbers or underscores, and can be in any case.

- Columns are either character or numeric.

- Character columns can have a length between 1 and 32,767 bytes (1 byte = 1 character).

- Numeric columns are stored with a length of 8 bytes.

- Character columns can consist letters, numbers, special characters, or blanks.

- Numeric columns can consist of the digits 0-9, minus sign, decimal point, and E for scientific notation.

- SAS date values are a type of numeric value and represent the number of days between January 1, 1960, and a specified date.

> **PROC CONTENTS DATA=**_table-name_**;**
> **RUN;**

## Accessing Data through Libraries

- A _libref_ is the name of the library that can be used in a SAS program to read data files.

- The _engine_ provides instructions for reading SAS files and other types of files.

- The _path_ provides the directory where the collection of tables is located.

- The libref remains active until you clear it, delete it, or shut down SAS.

> **LIBNAME** _libref_ **engine "**_path_**";**
> **LIBNAME** _libref_ **CLEAR;**

### Automatic Libraries

- Tables stored in the **Work** library are deleted at the end of each SAS session.

- **Work** is the default library, so if a table name is provided in the program without a libref, the table will be read from or written to the Work library.

- The **Sashelp** library contains a collection of sample tables and other files that include information about your SAS session.

## Using a Library to Read Excel Files

- The XLSX engine enables us to read data directly from Excel workbooks. The XLSX engine requires the SAS/ACCESS to PC Files license.

- The VALIDVARNAME=V7 system option forces table and column names read from Excel to adhere to recommended SAS naming conventions. Spaces and special symbols are replaced with underscores, and names greater than 32 characters are truncated.

- Date values are automatically converted to numeric SAS date values and formatted for easy interpretation.

- Worksheets or named ranges from the Excel workbook can be referenced in a SAS program as *libref.spreadsheet-name*.

- When you define a connection to a data source such as Excel or other databases, it's a good practice to delete the libref at the end of your program with the CLEAR option.

> **OPTIONS VALIDVARNAME=V7;**
> **LIBNAME** *libref* **XLSX "***path/file***.xlsx";**

## Importing Data

- The DBMS option identifies the file type. The CSV value is included with Base SAS.

- The OUT= option provides the library and name of the SAS output table.

- The REPLACE option is necessary to overwrite the SAS output table if it exists.

- SAS assumes that column names are in the first line of the text file and data begins on the second line.

- Date values are automatically converted to numeric SAS date values and formatted for easy interpretation.

- The GUESSINGROWS= option can be used to increase the number of rows SAS scans to determine each column's type and length from the default 20 rows up to 32,767.

### Importing a Comma-Delimited (CSV) File

> **PROC IMPORT DATAFILE=***"file.csv"* **DBMS=CSV**
> **OUT=***output-table* **<REPLACE>;**
> **<GUESSINGROWS=***n***>;**
> **RUN;**

### Importing an Excel (XLSX) File

> **PROC IMPORT DATAFILE="***file.xlsx***" DBMS=XLSX**
> **OUT=***output-table* **<REPLACE>;**
> **<SHEET=***sheet-name***;>**
> **RUN;**