**Chapter I - Base Programming**

**SAS programs**

Sample SAS program contains two steps: DATA step & PROC step (alone or combined)

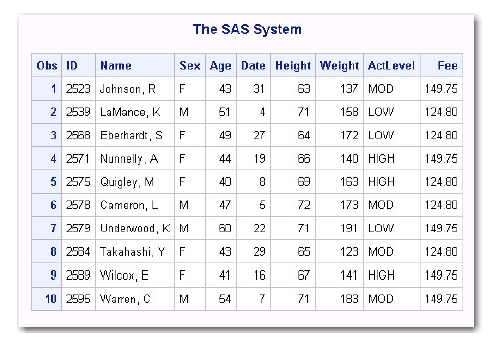
DATA step: 1. create or modify SAS data set

2. used to produce custom-designed report.

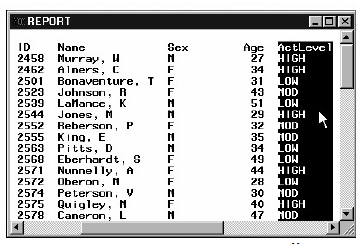
PROC step: 1. invoke or call pre-written routines that enable you to analyze process the data in a SAS data set

2. present the data in the form of a report

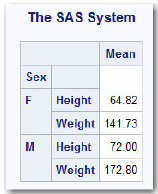
PROC PRINT creates a report and displayed in HTML



PROC REPORT open an interactive window, a window that you can use to directly modify data



PROC TABULATE invoke procedure that create output in the form of a report (相当于一个整理过的表格with 一些整合过的data?)



DATA & PROC is the signal of the beginning of a new step, and end of the previous step. RUN and QUIT mark step boundaries

Log Massage: A log of the processing activities and the result from each time a step is executed. (Contain processing and any error occurred)

Some programs such as sorting and managing data are invisible from log

SAS program statement: (Quiz 9)

1. Begins with a SAS keyword
2. Ends with a semicolon

**SAS Libraries (**the highest level of organization for information within SAS (包含所有相关的SAS文件)**)**

Sort SAS files: (Prc. 10)

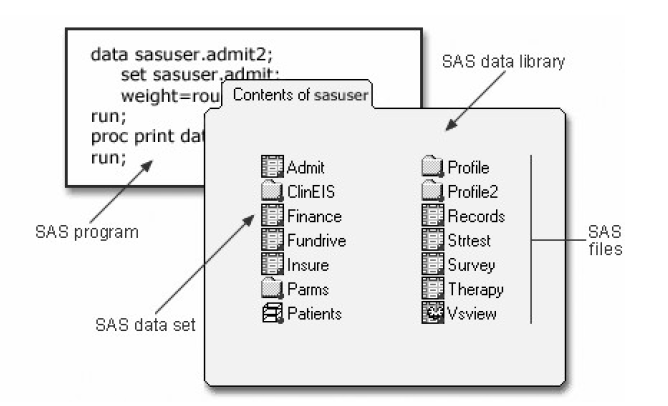
1. Every SAS file is stored in a SAS library, which is a collection of SAS files, such as SAS data sets and catalogs.
2. In some operating environments, a SAS library is a physical collection of files.
3. In others, the files are only logically related.
4. In the Windows and UNIX environments, a SAS library is typically a group of SAS files in the same folder or directory.

Temporarily Sort: (list only for current SAS session, deleted when session ends)

The library name does not specify, or specified the library name Work

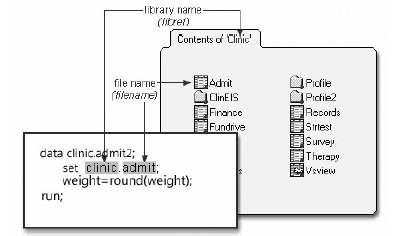
Permanently Sort: (available during the subsequent SAS session, exist until it has been deleted manually)

The library name is specified other than the default library name Work



**Referencing SAS file** (must assign a libref to each permanent SAS data library, each time begin a SAS session)

Two-Level Name (Temp. or Perm.): Library name (libref) + filename or data set name (Prac. 7)

E.g.: libref. filename

When reference a temporary file, the file can be named as Work. Test, so the data set named Test is sorted in the temporary library Work

One-Level Name (Temp.): used to reference a file in a temporary SAS library, the default libref Work is assumed.

Roles for SAS Names (Quiz 6)

Data set names and variable name:

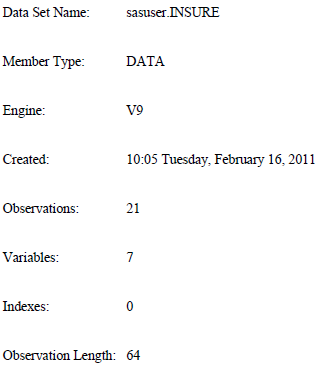
1. From 1 to 32 characters long
2. Begin with a letter or an underscore
3. Can continue with any combination of numbers, letters, or underscores

Under the filename portion of a SAS data set name, a libref can have a length of only 8 characters

**Data Set:**

SAS data set is a file that consists 2 or 3 parts: descriptor portion, data portion, and indexes (Optional, can be one or more, enable SAS to locate rows in the data set more efficiently)

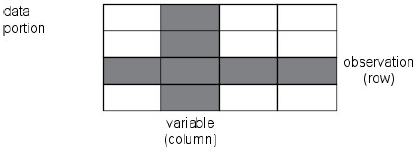
Descriptor Portion: (Contains information about the data set: name, date and time created, number of observations (row) and variables (columns))



Descriptor Portion:

Contains information about the data set: name, date and time that the data set was created, the number of observation and variables

Data Portion: (A table contains all data)



1. Observations (Row - relate to a single object)
2. Variable (Columns - Describe a particular characteristic)

Variable attributes:

The properties information includes the variable's name, type, length, format, informat, and label.

Name (SAME AS DATA SET)

Type:

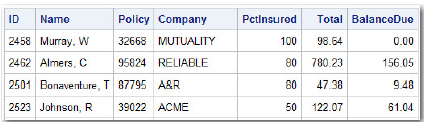
Character variable – can contain any value

Numerical variable – contain only numerical value: number, 运算符号, and E for scientific notation

Type attribute :

Character variable – a blank represents a missing value

Numerical variable – a period represents a missing value

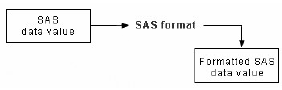


1. Missing value: Every variable and observation in a SAS data set must have a value. If a data value is unknow for a particular observation, a missing value is recorded in the SAS data set
2. Length:

Character variable – up to 32767 bytes

Numerical variable – only 8 bytes, no matter how many digits it contains

1. Format (variable attributes that affect the way data values are written):



Usually the maximum width (w) has to be specified, the specification of the number of decimal places (d) is optional

Format can be assign to a variable in SAS data set as permanent, or temporarily by specifying a

format in a PROC step to determine the way the data values appear in output.

1. Informat (Determine how data values are read into a SAS data set)

You mustuse informats to read numeric values that contain letters or other special characters.



1. Label (display more descriptive information about the variable by assigning a label to the variable)

Consists of descriptive text up to 256 characters long

**Points to Remember**

* Before referencing SAS files, you must assign a name (libref, or library reference) to the library in which the files are stored (or specify that SAS is to assign the name automatically).
* You can store SAS files either temporarily or permanently.
* Variable names follow the same rules as SAS data set names. However, your site may choose to restrict variable names to those valid in Version 6 SAS, to uppercase variable names automatically, or to remove all restrictions on variable names.