SAS

Wajdi Ben Saad
ESSAI, Sept , 2015

Today we are learning :

- SAS libraries
- Reading data from external files
  - txt and csv
  - Datalines
- SET statement
- PROC Print
- PROC Freq
- PROC Means

SAS libraries

SAS libraries

- LIBNAME statement assigns a libref

- Libref (short for "Library Reference") is an alias or nickname for a directory or folder for **SAS datasets**

SAS libraries

- LIBNAME statement assigns a libref
- Libref (short for "Library Reference") is an alias or nickname for a directory or folder for **SAS datasets**
- Dataset references contain two parts:
  - libref
  - dataset-name
  - Looks like: *libref.dataset-name*
- If libref is blank, the default is the Work library

SAS work library

- Work is a temporary library
- SAS datasets created in Work only exist during SAS session
- Once SAS session ends, datasets are erased
- Do not need to assign a libref for Work or specify it in dataset references
  ```
  data Test_Scores;
  ```
  is the same as
  ```
  data work.Test_Scores;
  ```

SAS libraries

```
libname mozart 'c:\books\learning';

data mozart.test_scores;
    length ID $ 3 Name $ 15;
    input ID $ Score1-Score3 Name $;
datalines;
1 90 95 98 Cody
2 78 77 75 Preston
3 88 91 92 Russell
;
run;
```
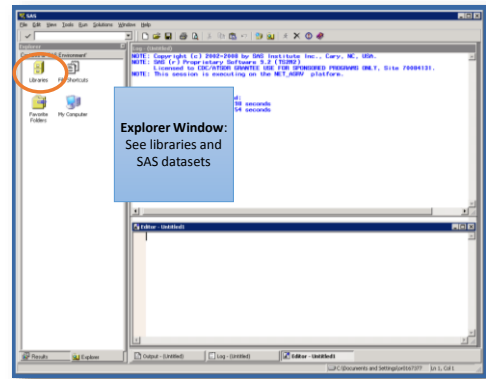
LIBNAME statement: Assigns a libref
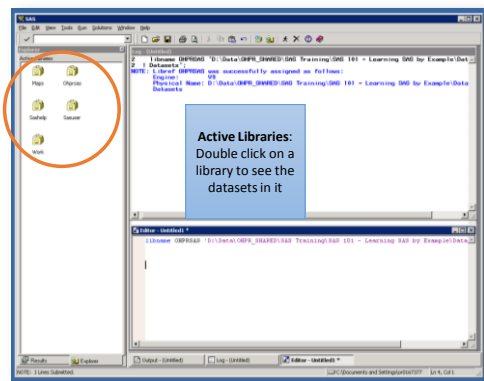
Use the libref for saving data and for retrieving data

```
libname OHPRSAS 'D:\Data\OHPR_SHARED\SAS Training\SAS 101 - Learning SAS by Example\Data\SAS Datasets';

data ohprsas.testdat; set mozart.test_scores;
run;
```

Wajdi Ben Saad

Explorer Window: See libraries and SAS datasets

Wajdi Ben Saad

Active Libraries: Double click on a library to see the datasets in it

Wajdi Ben Saad

LIBNAME examples

```
libname mozart 'c:\books\learning';
```

```
libname OHPRSAS 'D:\Data\OHPR_SHARED\SAS Training\SAS 101 - Learning SAS by Example\Data\SAS Datasets';
```

```
libname locald '\\tsclient\D\SAS\data';
```

Wajdi Ben Saad

Reading external data

Wajdi Ben Saad

Reading data from a text file

```
data demographics;
    infile 'D:\Data\OHPR_SHARED\SAS Training\SAS 101 - Learning SAS by Example\Data\Text and CSV\mydata.txt';
    input  Gender $
           Age
           Height
           Weight;
run;
```

INFILE – where to find the data

INPUT – variable names to associate with each data value
(**$** indicates character variable. Otherwise numeric.)

Wajdi Ben Saad

Reading data from a text file

Results of PROC Print of "Demographics"

Obs – short for "observation" (part of PROC Print output)
Numbers observations from 1 to N

Wajdi Ben Saad

Reading data from a csv file

Four variables: Gender, Age, Height (in inches), Weight (in pounds)

Variables separated by **commas**

Wajdi Ben Saad

Reading data from a csv file

dsd option (delimiter-sensitive data):
❑ Changes default delimiter from blank to comma
❑ If two delimiters in a row, assumes missing value between
❑ Quotes stripped from character values

Wajdi Ben Saad

Reading data from a csv file

Results of PROC Print of "Demographics"

SAS data results are the same

Wajdi Ben Saad

Other delimiters

- `dlm=` (or `delimiter=`) option to specify data delimiters other than blanks or commas
  - Example: `infile 'D:\Data\mydata.txt' dlm=':';`
- `dsd` and `dlm=` options together
  - Performs all functions of dsd, but overrides default delimiter

Wajdi Ben Saad

Datalines

- Allows dataset to be created within SAS program
- Can be useful for creating a quick set of test data
- Use either `datalines` or `cards` options
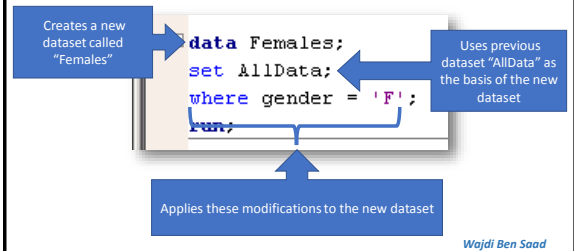- Follow with semi-colon **after** last line of data

Wajdi Ben Saad

SET statement

SET statement

• After you've brought your data into a SAS dataset, most of your DATA steps will look like this:

```
data Females;
set AllData;
where gender = 'F';
run;
```

Creates a new dataset called "Females"

Uses previous dataset "AllData" as the basis of the new dataset

Applies these modifications to the new dataset

SET statement

• The SET statement is similar to an INPUT statement
  • Except instead of a raw data file, you are reading observations from a SAS dataset
• Can read in temporary or permanent SAS datasets

```
libname mozart 'c:\books\learning';

data testdata;
set mozart.test_scores;
where score ge 90
    and gender='F';
run;

data winners;
set testdata;
if score ge 90 then top_perc = 'Top 2 Percent ';
else if score ge 95 then top_perc = 'Top 5 Percent ';
else if score ge 90 then top_perc = 'Top 10 Percent';
run;
```

PROC Print

PROC Print

• PROC Print can be used to list the data in a SAS dataset

```
data demographics;
infile datalines;
input Gender $ Age Height Weight;
datalines;
M 50 68 155
F 23 60 101
M 65 72 220
F 35 65 133
M 15 71 166
;

proc print; run;
```

PROC Print

```
              The SAS System      13:41 Wednesd
Obs   Gender   Age   Height   Weight
 1      M       50     68       155
 2      F       23     60       101
 3      M       65     72       220
 4      F       35     65       133
 5      M       15     71       166
```

Results of PROC Print of "Demographics"

PROC Print

- Many options to control output of PROC Print
  - noobs – Suppresses "OBS" column in output
  - (obs=2) – Only prints the first two observations
    - Can put in any number: 1 through N
    - Must be placed in parentheses after data= option
  - var statement – Only prints listed variables

```
proc print noobs data=demographics (obs=2);
var Gender Height;
run;
```

PROC Print

We'll discuss other PROC Print options in later chapters

PROC Freq

PROC Freq

- PROC Freq can be used to run simple frequency tables on your data

```
data demographics;
infile datalines;
input Gender $ Age Height Weight;
datalines;
M 50 69 155
F 23 60 101
M 65 72 220
F 35 65 133
M 15 71 166
;
proc freq; run;
```

PROC Freq

Results of PROC Freq of "Demographics"

PROC Freq

- Use the table statement to only print selected variables
- Use the nocum option to suppress cumulative statistics
- Use the nopercent option to suppress percent statistics
- Can use options together or separately

```
proc freq data=demographics;
table Gender / nocum nopercent;
run;
```

PROC Freq

• Can create simple cross-tabulations

Wajdi Ben Saad

PROC Freq

• Use the `list` option to display cross-tab tables in a list format

Wajdi Ben Saad

PROC Means

Wajdi Ben Saad

PROC Means

• PROC Means can be used to run simple summary statistics on your data

Wajdi Ben Saad

PROC Means

Results of PROC Means of "Demographics"

Wajdi Ben Saad

PROC Means

• Many options to control output of PROC Means
  • `NMiss Mean Median` – Examples of statistics that can be specified in PROC Means
    • (see later slide for list of statistical keywords)
  • `class` statement – Allows for grouping by categorical variables
  • `var` statement – Only provides statistics for listed analysis variables

Wajdi Ben Saad

6

PROC Means

The MEANS Procedure

Analysis Variable : Height

Gender	N Obs	N Miss	Mean	Median
F	2	0	62.5000000	62.5000000
M	3	0	70.3333333	71.0000000

We'll discuss other PROC Freq and PROC Means options in later chapters

Wajdi Ben Saad

PROC Means

Descriptive statistics keyword	
CSS	RANGE
CV	SKEWNESS | SKEW
KURTOSIS | KURT	STDDEV | STD
LCLM	STDERR
MAX	SUM
MEAN	SUMWGT
MIN	UCLM
MODE	USS
N	VAR
NMISS	
Quantile statistics keyword	
MEDIAN | P50	Q3 | P75
P1	P90
P5	P95
P10	P99
Q1 | P25	QRANGE
Hypothesis testing keyword	
PROBT | PRT	T

Examples of statistics that can be run with PROC Means

Wajdi Ben Saad