Technology Short Courses: Fall 2008
(October, 2008)
Kentaka Aruga
Object of the course

- Sub-windows in SAS
- Basics of managing data files
- Basic commands in SAS
Introduction: What is SAS?

What is SAS?
- Originally an acronym for Statistical Analysis System
- Provided by SAS institute since the 1970s
- A software used for statistical analysis, graphing, and presenting data
Introduction: DATA Step

- Two distinct categories
  - DATA step
  - PROC step

- DATA Step
  - Provides data management
  - Use
    - Reading data
    - Data transformation
    - Creating or removing variables
Introduction: PROC Step

PROC Step

- Performs a wide variety of analysis on data those are retrieved and transformed from the DATA Step

Examples

- PROC MEANS, CONTENTS, SORT, FREQ, PRINT, PLOT etc.
Section 1
Learning About the Sub-windows
Opening SAS

- Start → All Programs → SAS → SAS 9.1
Three main windows: Program editor

On the top bar click ‘Window’ and then click ‘Tile Vertically.’ You will be able to see three sub-windows.

1. Program editor  
2. Log window  
3. Output window
Three main windows: Program editor

- Program editor
- Entering and editing SAS command lines

The extension of the saved file is .sas
Three main windows: Log window

- Log window
- This window keeps track of your command runs, and lists SAS notes and error messages (shown in red)

Commands written correctly

```
1 /importing direct data to SAS */
2 data direct;
3 input age weight gender $;
4 cards;
```

Commands with error

```
1 /importing direct data to SAS */
2 data direct;
3 input age weight gender $;
4 card;
```

Error message:

```
WARNING 14163: Assuming the symbol CARDS was misspelled as card.
```

Error message in red:

```
WARNING 14163: Assuming the symbol CARDS was misspelled as card.
```
Three main windows: Output window

- Output window
  - Shows the results of SAS procedures
  - The extension of the saved file is ``.lst``
The ‘Explorer’ and ‘Results’ Windows will appear on the left side of your screen.

- **Explorer window**
  - This window is used to explore various default libraries that contain a number of sample SAS data sets

- **Results window**
  - Organizes the information contained in the Output Window in a hierarchical fashion.

[Click]
Click ‘Libraries’ icon in the Explorer window. Then you will see several subfolders. You can find the raw SAS data in these subfolders.
‘Explorer’ window (Cont’d)

- To move backward from one folder to another in the Explorer Window, simply click the left most icon on the toolbar that looks like a folder.
‘Explorer’ window (Cont’d)

- To move backward from one folder to another in the Explorer Window, simply click the left most icon on the toolbar that looks like a folder.

Click
‘Results’ window

- Results window
  - This window allows you to view all the results of procedures you have executed in the program editor.

- Use the expansion icons (+ or - icons) next to the folder to open or hide its contents.
Points to Remember in SAS program

- All SAS statements begin with a keyword and end with a semicolon (;)
- Except for within the data section, SAS is not sensitive to spacing between words: the amount of space you put between words does not matter.
- Comments are entered in a SAS program using either the following formats:
  - /* comments */ (used for large comment blocks)
  - * comments ; (used for single line comments)
Section 2
Basics of managing data files:
DATA step, LIBNAME, PROC export, and PROC import, and data transformation
Practice Round: Getting data

- Go to
  http://www.uri.edu/its/instructional_on-line_materials/sasbasics.html

- Download the SAS program from
  http://www.uri.edu/its/research/basics.txt

- Download two data files from
  http://www.uri.edu/its/research/scores.txt
  http://www.uri.edu/its/research/scores2.txt

- After opening these files, select ‘Save As’ under File. Save these as C:\basics.txt, C:\scores.txt, and C:\scores2.txt.
Importing direct data

- Open basics.txt with ‘MS Word’ or ‘Notepad.’
- Drag lines shown below in the file and copy and paste it to the ‘editor ’ window in SAS.

```sas
data direct;
input age weight gender $;
cards;
21 134 F
33 167 M
45 157 M
;
run;
```

- 'cards' statement allows you to put raw data directly to SAS
You can copy and paste also with your keyboard. Copy is Ctrl-C and paste is Ctrl-V.
Importing direct data: Executing the commands

To execute the commands, highlight it and click the ‘submit’ icon or select ‘submit’ under the Run menu.
Data command

- data direct;
  - Allows SAS to create a temporary SAS data file.
  - In this example the file was named ‘direct’ but you can have your own name by renaming ‘direct.’
  - In the ‘Explorer’ window click Libraries.
Data command: How to see your data in the SAS library

- Now click and go into the ‘Work’ library.
  - You should see the ‘direct’ file you have just created in the library.

- Finally click the ‘direct’ file in the work library.
  - You should be able to see the ‘viewtable window’
‘Work’ library

- The data in the ‘Work’ library is not stored permanently in SAS. The work folder store files only **temporarily**. Once you exit the SAS program the file will be erased from the folder.
  - End SAS session.
  - Open SAS again and look in the Work library. NO DATA FILES!
LIBNAME statement

- To store the data permanently, you need to create and reference a library
  ⇒ Use LIBNAME statement
- Drag the lines shown below from the file ‘basics.txt’ and copy and paste it to the ‘editor’ window in SAS.

```sas
libname test 'C:\' ;
data test.direct;
input age weight gender $;
cards;
21 134 F
33 167 M
45 157 M
;
run;
```

Name of the library

Name of the file
After pasting the commands to the ‘editor’ window of SAS, highlight the commands and then click submit.
The command submitted has created a new library named ‘test’ on SAS, and saved data file ‘direct’ in this library and in the ‘C:\’ folder of your computer.

- In the ‘Explorer’ window click Libraries. Then go into the ‘test’ library.
- Click the ‘test library’
LIBNAME statement (cont’d)

- You will now see the ‘Direct’ file in the ‘test’ library.

- To view the ‘Direct’ data file click ‘direct’

- You will also find the same file in the ‘C:\’ folder of your computer.
LIBNAME statement (cont’d)

Once you stored your data file into your C:/ drive with the LIBNAME statement, you can refer to the file without importing the raw data again.

Example:

- Close SAS session, re-open it.
- Then copy and paste the following commands from ‘basics.txt’ to the ‘editor’ window in SAS.

  libname test ‘C:\’;
  proc print data=test.direct;
  run;

- Click the submit icon to execute the command
/* Storing permanent data on SAS */
libname test 'c:\';
data test.direct;
input age weight gender $;
cards;
21 134 F  
33 167 M  
45 157 M  
;
run;

libname test 'c:\';
proc print data=test.direct;
run;

/* Importing data */
data score;
infile 'c:\' input height weight;
data score;
infile 'c:\' input height name $ 8-1 run;
/* To put output on one line */
data linecontrol:
You will see the same data as before!
Forms of INPUT statement

Example 1

- input age weight gender $;
  - This statement allows SAS to read the variables used for the raw data.
  - In this example three variables (age, weight, and gender) were put into SAS.
  - SAS initially only reads numeric variable so in order to read character values you need to use modifiers:
    - The variable ‘gender’ is a character variable. You need to use ‘$’
      - $: enables SAS to read character values with default size of eight characters with no embedded blanks
      - &: enables SAS to read character values with embedded blanks
INPUT statement: Example 2

- input height 1-3 weight 4-6 gender 7 name $ 8-14 score 15-16;

- If the data contain the followings you need to set up a column input mode to specify the column positions of the pointer
  - Standard character and numeric data
  - Values entered in fixed column positions
  - Character values longer than eight characters
  - Character values with embedded blanks
Importing external data

Open scores.txt, and scores2.txt from c:/ drive and compare.
Importing external data (Cont’d)

- Open basics.txt with ‘MS Word’ or ‘Notepad.’
- Drag the lines shown below on the file, copy and paste it to ‘editor ’ window in SAS, and execute the commands.

```sas
data scores;
infile 'C:\scores.txt';
input height weight gender name $ score;
run;

data scores2;
infile 'C:\scores2.txt';
input height 1-3 weight 4-6 gender 7
name $ 8-14 score 15-16;
run;
```
Importing external data (Cont’d)

- Go to the ‘Explorer’ window, click the work folder, and open ‘scores’ and ‘scores2.’ You will see exactly the same file.
INPUT statement: Example 3

- How to put observations in more than one line
- #n: moves the pointer to record n.
- Example
  ```
data linecontrol;
  input #1 name $ height weight #2 country & $24.
  #3 score1 score2;
  cards;
  Ken 5.9 158
  Great Britain
  44 36
  Pete 6.2 180
  United States of America
  32 29
  ;
  run;
  ```
INPUT statement: Example 4

- How to put several observations in one line
  - @@: Used when each input line contains values for several observations
- Example
  ```
  data oneline;
  input name $ score @@;
  cards;
  Joanne 23 John 34 Jimmy 45
  Katrina 0 Chris 20
  ;
  run;
  ```
/*Exporting data to MS Excel data*/
  proc export data=scores
    outfile="C:\scores.xls"
    dbms=excel2000 replace;
    sheet="scores";
  run;

/*Importing data from MS Excel*/
  proc import out=impscores
    datafile="C:\scores.xls"
    dbms=excel2000 replace;
    sheet="scores";
    getnames=yes;
    mixed=yes;
  run;
Click file under the main tab and open "Import Data"
Importing MS Excel data 2-2
Data Transformation

How to transform data in SAS

```sas
data trans;
set scores;
* 'Set' command allows reusing created SAS data;
lnheight=log(height);
logheight=log10(height);
index=height/weight;
run;
```
Data Transformation (Cont’d)

- Note
  - LOG(x) : the natural logarithm of x
  - LOG10(x) : the log base ten of x
  - LOG2(x) : the log base two of x
# Arithmetic and Comparison Operators

## Arithmetic Operators

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>**</td>
<td>exponentiation</td>
<td>a**3</td>
</tr>
<tr>
<td>*</td>
<td>multiplication</td>
<td>2*y</td>
</tr>
<tr>
<td>/</td>
<td>division</td>
<td>var/5</td>
</tr>
<tr>
<td>+</td>
<td>addition</td>
<td>num+3</td>
</tr>
<tr>
<td>-</td>
<td>subtraction</td>
<td>sale-discount</td>
</tr>
</tbody>
</table>

## Comparison Operators

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>=</td>
<td>equal to</td>
<td>a = 3</td>
</tr>
<tr>
<td>^= or NE</td>
<td>not equal to</td>
<td>a ne 3</td>
</tr>
<tr>
<td>&lt;= or LE</td>
<td>less than or equal to</td>
<td>num &lt;= 8</td>
</tr>
<tr>
<td>&gt;= or GE</td>
<td>greater than or equal to</td>
<td>sales &gt;= 300</td>
</tr>
<tr>
<td>&lt; or LT</td>
<td>less than</td>
<td>num &lt; 8</td>
</tr>
<tr>
<td>&gt; or GT</td>
<td>greater than</td>
<td>num &gt; 5</td>
</tr>
</tbody>
</table>
How to delete certain observations from data

Example: The following command deletes observations having weight more than 160

```sas
data modify;
set trans; /*'Set' command allows reusing created SAS data;
if weight > 160 then delete;
run;
```

Open the created data file `modify` in the `Work` folder of your library and compare that from the data file `trans`.
You can see that observations for ‘Mark,’ ‘Eric,’ and ‘Bruce,’ have been deleted in ‘modify.’
Section 3
Basic commands in SAS:
PROC step
Proc Steps: proc print

- Use: to see the SAS data file in the output window
  proc print data=scores;
  run;
Proc Steps: proc contents

- Use: to see the contents of SAS data file
proc contents data=scores;
run;
Proc Steps: proc sort

- Use: to sort SAS data file

**proc sort** data=scores out=name by name; *Sorts the data by name in alphabetical orders*
**run**;

**proc sort** data=scores out=height by height; *Sorts the data by height in ascending orders; Run;

**proc sort** data=scores out=height2 by **descending** height; *Sorts the data by height in descending orders; run;
Proc Steps: proc means

- Use: to see basic simple statistics of data

```
proc means data=scores;
run;
```

*This provides the number of obvs, mean, std, min, and max of all numeric variables;
Proc means: How to see other simple statistics

To find out the commands for other simple statistics click the help icon and then click index. Then type in 'keywords' in the search box and enter. Finally, click 'for statistics';
Proc Steps: How to see other simple statistics (Cont’d)

- Insert the commands for the simple statistics you want to calculate with SAS before the command ‘data=“file name” ’:

  ```sas
  proc means nmiss range kurt skew data=scores;
  run;
  ```
Proc Steps: proc freq

- proc freq
  - Use: to analyze frequency of the variables and to create frequency tables for variables
  ```
  proc freq data=scores;
  run; *shows one-way frequencies;
  
  proc freq data=scores;
  tables gender*weight;
  run; *creates cross-tabulation table;
  ```
### Output - (Untitled)

#### The SAS System

**The FREQ Procedure**

<table>
<thead>
<tr>
<th>height</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Frequency</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.3</td>
<td>1</td>
<td>12.50</td>
<td>1</td>
<td>12.50</td>
</tr>
<tr>
<td>5.4</td>
<td>1</td>
<td>12.50</td>
<td>2</td>
<td>25.00</td>
</tr>
<tr>
<td>5.6</td>
<td>1</td>
<td>12.50</td>
<td>3</td>
<td>37.50</td>
</tr>
<tr>
<td>5.7</td>
<td>2</td>
<td>25.00</td>
<td>5</td>
<td>62.50</td>
</tr>
<tr>
<td>5.9</td>
<td>1</td>
<td>12.50</td>
<td>6</td>
<td>75.00</td>
</tr>
<tr>
<td>6.5</td>
<td>1</td>
<td>12.50</td>
<td>7</td>
<td>87.50</td>
</tr>
<tr>
<td>6.4</td>
<td>1</td>
<td>12.50</td>
<td>8</td>
<td>100.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>weight</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Frequency</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>122</td>
<td>1</td>
<td>14.29</td>
<td>1</td>
<td>14.29</td>
</tr>
<tr>
<td>125</td>
<td>1</td>
<td>14.29</td>
<td>2</td>
<td>28.57</td>
</tr>
<tr>
<td>145</td>
<td>1</td>
<td>14.29</td>
<td>3</td>
<td>42.86</td>
</tr>
<tr>
<td>156</td>
<td>1</td>
<td>14.29</td>
<td>4</td>
<td>57.14</td>
</tr>
<tr>
<td>165</td>
<td>1</td>
<td>14.29</td>
<td>5</td>
<td>71.43</td>
</tr>
<tr>
<td>170</td>
<td>1</td>
<td>14.29</td>
<td>6</td>
<td>85.71</td>
</tr>
<tr>
<td>200</td>
<td>1</td>
<td>14.29</td>
<td>7</td>
<td>100.00</td>
</tr>
</tbody>
</table>

**Gender Missing = 1**

<table>
<thead>
<tr>
<th>gender</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Frequency</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>50.00</td>
<td>4</td>
<td>50.00</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>50.00</td>
<td>8</td>
<td>100.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>name</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Frequency</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Output - (Untitled)

#### The SAS System

**The FREQ Procedure**

**Table of gender by weight**

<table>
<thead>
<tr>
<th>gender</th>
<th>weight</th>
<th>Frequency</th>
<th>Percent</th>
<th>Row Pct</th>
<th>Col Pct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>122</td>
<td>165</td>
<td>156</td>
<td>170</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>14.29</td>
<td>14.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>25.00</td>
<td>25.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>14.29</td>
<td>14.29</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>33.33</td>
<td>33.33</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100.00</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100.00</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Gender Missing = 1**
Proc Steps: proc reg

proc reg

One of a general-purpose procedures for regression analysis in SAS

```
proc reg data=scores;
model height=weight / dw alpha=0.01 clb ;
plot height*weight / cframe=ligr conf pred ;
run;
```

\[
height = \alpha + \beta weight + \epsilon
\]
Proc Steps: proc gplot

proc gplot

Use: to plot the values of two or more variables on a set of coordinate axes

```
proc gplot data=scores;
plot height*weight;
*height=vertical axis,
weight=horizontal axis;
run;
```
Using advanced options in SAS

**proc gplot** data=scores;
plot height*weight
/skipmiss haxis=120 to 200 by 10 hminor=1
vaxis=5.0 to 7.0 by 1.0 vminor=1
Regeqn cframe=gold; *Options for the plot statement;

**title** font=arial c=blue box=3 bcolor=yellow 'Study of Height vs Weight'; *Putting a title for your graph;

**symbol** i=rcclm95 value=dot height=1
  cv=green ci=blue co=red width=2;
  *Setting the colors and size for the plot symbol and lines. i= can be also expressed as interpol=;
**run;**
Study of Height vs Weight

Regression Equation:
height = -0.219354 + 0.090498*weight - 0.000498*weight^2 + 1.054E-6*weight^3
Useful supports

- In the tool bar click the help menu or the help icon
Useful supports: using the Help in SAS

Example: click index and type ‘reg.’ Then double click ‘REG procedure’
Useful supports: other useful sites

- Online SAS manuals
  
  http://www.uri.edu/sasdoc
  
  This will automatically link you to
  
  http://support.sas.com/documentation/onlinedoc/sas9_doc.html

- Statbookstore: useful site for finding program examples
  
  http://www.geocities.com/statbookstore/
Exercise

- Import the following data and use the libname statement to save the data to your ‘c:/’ drive of the computer.
- Use SAS to determine the mean and variance of ‘height’ and ‘score’ of the data.
- Determine the intercept (b1) and the coefficient (b2) of the model, height = b1 + b2 * weight + e using the data.

<table>
<thead>
<tr>
<th>height</th>
<th>weight</th>
<th>gender</th>
<th>name</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.4</td>
<td>125</td>
<td>2</td>
<td>JAUNITA</td>
<td>65</td>
</tr>
<tr>
<td>5.3</td>
<td>122</td>
<td>2</td>
<td>SALLY</td>
<td>77</td>
</tr>
<tr>
<td>5.7</td>
<td>145</td>
<td>2</td>
<td>SABRINA</td>
<td>36</td>
</tr>
<tr>
<td>5.9</td>
<td>150</td>
<td>2</td>
<td>KATE</td>
<td>55</td>
</tr>
<tr>
<td>5.7</td>
<td>156</td>
<td>1</td>
<td>JOHN</td>
<td>84</td>
</tr>
<tr>
<td>6.0</td>
<td>170</td>
<td>1</td>
<td>MARK</td>
<td>56</td>
</tr>
<tr>
<td>6.4</td>
<td>200</td>
<td>1</td>
<td>ERIC</td>
<td>34</td>
</tr>
<tr>
<td>5.9</td>
<td>165</td>
<td>1</td>
<td>BRUCE</td>
<td>72</td>
</tr>
<tr>
<td>6.2</td>
<td>160</td>
<td>1</td>
<td>TOM</td>
<td>88</td>
</tr>
</tbody>
</table>
libname test 'c:\';

data test.scores3;
  input height weight gender name $ score;
cards;
  5.4  125  2  JAUNITA   65
  5.3  122  2  SALLY     77
  5.7  145  2  SABRINA  36
  5.9  150  2  KATE     55
  5.7  156  1  JOHN     84
   6   170  1  MARK     56
  6.4  200  1  ERIC     34
  5.9  165  1  BRUCE    72
  6.2  160  1  TOM     88
;
run;

proc means mean var data=test.scores3;
run;

proc reg data=test.scores3;
  model height = weight;
run;
The MEANS Procedure

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>height</td>
<td>5.833333</td>
<td>0.1250000</td>
</tr>
<tr>
<td>weight</td>
<td>154.777778</td>
<td>561.1944444</td>
</tr>
<tr>
<td>gender</td>
<td>1.4444444</td>
<td>0.2777778</td>
</tr>
<tr>
<td>score</td>
<td>63.0000000</td>
<td>378.7500000</td>
</tr>
</tbody>
</table>

The REG Procedure

Model: MODEL1
Dependent Variable: height

Number of Observations Read: 9
Number of Observations Used: 9

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>0.85529</td>
<td>0.85529</td>
<td>41.37</td>
<td>0.0004</td>
</tr>
<tr>
<td>Error</td>
<td>7</td>
<td>0.14471</td>
<td>0.02067</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>8</td>
<td>1.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE: 0.14378
R-Square: 0.8553
Adj R-Sq: 0.8346

Coefficient Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > |t| |
|----------|----|--------------------|----------------|---------|------|---|
| intercept| 1  | 3.89703            | 0.33557        | 11.62   | 0.0001|
| weight   | 1  | 0.01380            | 0.00215        | 6.43    | 0.0004|
For further Questions:
kentaka@mail.uri.edu