Introduction to SAS: Lecture One

Shang Xue @ Purdue Statistics

▲□▶ ▲圖▶ ▲圖▶ ▲圖▶ = ● ● ●

Getting Started with SAS The SAS System SAS Components Overview of A SAS Program Running SAS Programs SAS Data Sets SAS Data Libraries

The SAS System

- SAS originally stands for Statistical Analysis Systems.
- Developed in 60's and 70's at North Carolina State University.
- Now an integrated system of software products provided by SAS institute.
- Widely used, enables Data Management, Report Writing & Graphics, Stat/Math Analysis etc.

- May run on Unix and Windows (we will only focus on Windows).
- Newest version: SAS 9.2 released in March, 2008

SAS Components

- Many components targeting Reporting and Graphics, Data Access and Management, User Interface, Analytical, Application Development, Visualization and Discovery, Business Solutions, Web Enhancement, such as:
- Base SAS The core of the SAS System, used to manage data, perform basic procedures.
- SAS/STAT Statistical Analysis. (ANOVA, regression, etc.)

- SAS/GRAPH Enhanced graphics
- SAS/OR Operations research
- SAS/ETS Econometrics and time series analysis
- SAS/QC Quality control
- etc.

A SAS program is a sequence of steps in a logical sequence that the user submits for execution. A SAS program is composed of three major parts:

- ▶ 1. The DATA STEP: Get the data set ready into SAS
- 2. The PROC STEP: Process data
- 3. Macro language: Enables flexible coding

SAS Programs

Below is an example of a SAS program:

```
DATA work.revenue;
INPUT City $ State $ Revenue;
CARDS:
LA
         CA 5000
Chicago IL 3000
.
PROC PRINT DATA=work.revenue;
RUN;
PROC MEANS data=work.revenue:
VAR Revenue;
RUN;
```

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

Step Boundaries

SAS steps begin with either one of the following:

- DATA statement
- PROC statement

SAS detects the end of a statement by checking one of the following:

- A RUN statement.
- A QUIT statement.
- The beginning of another step (DATA or PROC)

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

Now let's look at how to run a SAS program by doing the following in Windows environment:

Invoke the SAS system and include the SAS program into the session.

- Submit the program and browse the reuslts.
- Navigate the SAS windowng environment.

Starting a SAS session in the Windows Enviroment

- In the SAS session:
 - Explorer window shows the contents of the SAS environment.
 - Program Editor used to edit the SAS program.
 - SAS log window contains information about the SAS programs such as warning and error messages.
 - Output window contains reports generated by SAS DATA and PROC steps.
- Including and submitting a SAS program
 - ► Use File→Open Program or command "include" to include existing program. Write new progrmas directly in the Program Editor.
 - Selet the SAS codes you want to submit and use Run→Submit or command "submit" to submit programs.
- Examine the program results in the output window:
 - Becomes the active window each time it receives output.
 - ► Automatically accumulates output. You can use "clear" command or use Edit→Clear All.

SAS can work with SAS data sets, SAS data sets are data tables where:

- Columns are referred to as: Variables
- Rows are referred to as: Observations

There are two types of variables:

- Character variables, missing value is represented by a blank
- Numeric variables, missing value is represented by a period

SAS Data Libraries

- A SAS data library is a collection of SAS files that are regognized as a unit by SAS.
- At invocation, SAS automatically creates one temporary and at least one permanent SAS data library for user to access:
 - Work library: default temporary library
 - Sasuser: one of the permenant libarires
- Files in the temp library Work will be deleted after a SAS session ends, while files permanent libraries are saved.
- User can assign permanent library by using the LIBNAME statement:

```
LIBNAME libref 'library location';
```

For example:

```
LIBNAME mylib 'c:\SAS\project1';
```

Here libref refers to the library reference name.

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

- The first part libref refers to the library.
- The second part **filename** refers to a specific file in the library.
- The libref Work can be omitted when referring to a file in the temporary Work library.

Browsing SAS Libraries

User may browse a SAS library through the following ways:

- ▶ Using the Explorer window.
- Use "PROC CONTENTS" plus the keyword "_ALL_" to list all SAS files in a specific library: PROC CONTENTS DATA=libref._ALL_ NODS; RUN:

Here NODS is an option used to suppress the descriptor portion of the data sets.

Sample SAS Codes

Try the following codes in SAS:

```
LIBNAME mylib 'A local folder';
DATA mylib.revenue_LA_CHI;
INPUT City $ State $ Revenue;
CARDS;
T.A
          CA 5000
Chicago IL 3000
;
DATA mylib.revenue_DAL_BOS;
INPUT City $ State $ Revenue;
CARDS;
Dallas TX 4000
Boston
           MA 6000
;
PROC CONTENTS DATA=mylib._ALL_ NODS;
RUN;
```