Introduction to SAS

Ryne Estabrook

Department of Psychology
University of Virginia

December 17, 2007
CDHRM Lab Meeting
Outline

Overview

Data

Summarizing Data

Statistical Analyses

ODS

Graphing

Closing
Basics of SAS

- Integrated Statistical Software Package
- Pros:
  - Incredibly large
  - Flexible input and output
  - Relatively intuitive syntax
- Cons:
  - Incredibly large
  - Can’t/shouldn’t do everything
  - Graphics in particular are lacking relative to other programs
Major Components of the SAS Window

- Editor
- Log
- Output
- Other:
  - Explorer
  - Graphics Window
Basic Structure of a SAS session

- **DATA Steps**
  - Opening and importing data files
  - Creating or transforming new datasets from existing datasets
- **PROC (Process) Steps**
  - Analytic component of system
  - Make systematic changes to datasets (sort, for instance)
  - Create output, including datasets and graphs
Other Important Information

- Format of most SAS procedures:
  - DATA or PROC statement, including data source
  - Statements identifying processes and output
  - RUN; or RUN; QUIT;

- All statements end with a semicolon

- "SAS is always right"

- Documentation is useful
Other Important Information

- Comment something out /* */
- TITLE ’title’; will title all output until replaced or cleared
- Default system missing is ‘.’
- Raising variables to powers requires **
Libraries: How SAS deals with data

- A library is a location or folder where files are stored
- All files are referenced by their library and file name
- Default libraries:
  - Sashelp
    - Includes help and example files
  - Sasuser
    - Default permanent library
  - Work
    - Default temporary library
File Names

- All files have a two part name
  - The first part is the library name
  - The second part is the file name
  - These parts are separated by a period

- Rules for file names
  - All characters must be alphanumeric or underscores
  - First character cannot be a number
  - Must be between 1 and 32 characters long
File Name Examples

sashelp.class
Dataset 'class' in 'sashelp' library

work.test
test
Dataset 'test' in 'work' library
Assigning a Library

- Assigning a library is easy using the 'libname' command.
- Once assigned, SAS recognizes all SAS files in that folder.

```sas
LIBNAME Ryne 'C:/SASintro/';

LIBNAME HEADACHE SPSS 'C:/SASintro/SPSS/'; //
LIBNAME MIGRAINE EXCEL 'C:/SASintro/Excel/';
```
Other Ways of Importing Files

- Data statements can input files directly
  - Good for text files
  - Doesn’t require using the library system
  - Does require knowing the names of your variables

- Copy and Paste (small) data sets with CARDS command

- PROC IMPORT and point-click Import Wizard exist

- Create new datasets from existing ones in any library
Overview of Data Statements

- Call the program (DATA)
- Identify the resulting dataset (newdata)
- Identify the data source (SET existingdata)
- Create, KEEP or DROP variables as needed

```sas
DATA newdata;
   SET existingdata;
   total=V1+V2;
   KEEP id total V1 V2;
RUN;
```
Other Data Topics
For Discussion or Future Sessions

- Moving between ’tall’ and ’long’ data formats
- Importing data with multiple lines per observation
- Simulating data
- Using arrays to manage data
Your data’s in. Now what?

- Reports and univariate analyses are often the first step of an analysis
- Create extensive output about distribution of data
- Output datasets for other analyses
- Good intro to the SAS PROC programs
Common Univariate Procs

- Many different PROCs exist for summarizing data
  - PROC UNIVARIATE
  - PROC MEANS
  - PROC FREQ
  - Other: PROC CORR, PROC SORT
- All have very similar structures
- All can be extended or iterated over multivariate data
Overview of PROC Statements

- Call the program (PROC WHATEVER)
- Identify the input dataset (DATA=dataset)
- Identify the variables for analysis (VAR or TABLES)
- Give additional statements for extra output

```
PROC UNIVARIATE data=example;
  VAR x y;
  HISTOGRAM;
  OUTPUT OUT=summary pctlpts=5 to 95 by 5;
RUN;
```
PROC UNIVARIATE capabilities

- Univariate summaries
  - Normal moments
  - Quantiles and Extreme observations
- Tests of normality
- Histograms, QQ plots and boxplots (basic)
PROC MEANS and PROC FREQ

- PROC MEANS
  - Less output than UNIVARIATE
  - Output statistics by group
  - Mirrors other procedures (PROC SURVEYMEANS)

- PROC FREQ
  - Categorical analyses of variables
  - Chi-square tests
Other things to know

- **PROC SORT**
  - Required before some PROCs that run on multiple groups
  - Does not require a RUN; command if given before one of these PROCs

- **PROC CORR**
  - Can output any number of covariance matrices
  - Outputs mean and variance information as well
  - These rows may get in the way, especially when exporting out of SAS
Other Report and Univariate Topics
For Discussion or Future Sessions

- PROC PRINT, REPORT and TABULATE
- PROC CAPABILITY
  - Creates a number of complex reports and output
  - Output lowess-smoothed data for graphics PROCs
PROCs for Statistical Analysis

- Hundreds of SAS PROCs
- For today, we’ll focus on common GLM derivations
  - Commonly used and versatile
  - Very similar in structure
Overview of PROC Statements

- Call the program (PROC WHATEVER)
- Identify the input dataset (DATA=dataset)
- Identify the model for analysis (dependent=independent)
- Give additional statements for extra output

```sas
PROC REG data=example;
  MODEL y=x/STB COVB CORRB;
  OUTPUT OUT=model P=predicted R=residual;
RUN;
```
Common GLM-based PROCs

- PROC REG
- PROC GLM (GLMSELECT)
- PROC ANOVA
- Related models:
  - PROC LOGISTIC, for logistic regression
  - PROC GENMOD, for GEE estimation
  - PROC NLIN, for non-linear models
  - PROC MIXED, NLMIXED, for mixed effects models
PROC REG

- OLS Regression
- Stepwise selection and interactive sessions
- Output:
  - Estimates, ANOVA tables, fit statistics
  - Predicted and residual value plots
  - Datasets
- Limitations:
  - Polynomial terms must be defined as variables beforehand
PROC GLM

- Allows for broader model inputs, esp. interactions
- Fewer output options
- PROC GLMSELECT
  - Incorporates many model selection criteria into GLM
  - Experimental release available for download from support.sas.com
Common Statements for GLM PROCs

- CLASS statement
  - Defines variables as categorical
- BY statement
  - Creates separate analyses for each level of a variable
- OUTPUT statement
  - Creates output datasets, often requiring that variables be named
Other topics in SAS PROCs
For Discussion or Future Sessions

- Other univariate PROCs
- PROCs for multivariate analyses
  - PROC FACTOR for EFA, CFA (complex)
- Any number of specific analyses can be covered
The Output Delivery System

- SAS output is monospace and boring
- Up until version 7.0, output was meant for line printers
- Output Delivery System outputs to a number of formats
  - Supports most common file types and markup languages
  - User defined formats and styles with PROC TEMPLATE
ODS: Destinations

- **SAS Formatted**
  - LISTING (Monospace Output Listing)
  - OUTPUT (SAS data set)
  - DOCUMENT (ODS Document)

- **Third-Party**
  - HTML
  - MARKUP (including LATEX)
  - PRINTER (PCL, PDF, PS)
  - RTF
ODS: Destinations

To view the tagsets and files supported by your system:

```sas
PROC TEMPLATE;
   LIST tagsets;
RUN;
```

- SAS comes default with about 50 tagsets
- User defined tagsets are listed first, and can be found in the SASUSER library
Basic Structure of an ODS command

- Preface programs by calling the ODS, including:
  - The tagset or template to be used
  - The name of the destination file (FILE=)
  - The path where the file should be written (PATH=)
- Include an 'ODS close' statement after said program

```sas
ODS rtf FILE='file.rtf' PATH='C:/';
PROC REG DATA=example;
  MODEL y=x;
RUN;
ODS rtf CLOSE;
```
Using ODS for LaTeX

▶ LaTeX is included as a default ODS destination
▶ This LaTeX format made to simulate HTML3, and is "verbose"
▶ SAS RnD provides "user defined templates" for three other LaTeX tagsets
Using ODS for LaTeX

- tagsets.latex
  - HTML3 based, built into SAS 9
- tagsets.colorlatex
  - Same as above, but with color
- tagsets.simplelatex
  - Simpler version, using the longtable command
- tagsets.tablesonlylatex
  - Simplelatex, but with only the tables
Other ODS topics

For Discussion or Future Sessions

- Other filetypes using ODS
- ODS for graphics files
- Using PROC TEMPLATE to create custom formats and tagsets
  - Very complex, not for the faint of heart
  - ODS improvements some of the primary features in 9.2, the next SAS release.
Brief Overview of Graphing

PROC G PLOT

- Features
  - 2D plotting of one or several series
  - Plots by individuals, or overlay multiple graphs
  - Simple smoothing

- Overview of syntax
  - Call the program and identify the data source
  - Specify graphics options, particularly symbols and axes
  - Specify variables with the PLOT statement

- ODS graphics (experimental)
  - Online document quote: "DO NOT USE FOR PRODUCTION JOBS"
Closing and Future Topics

- SAS is a powerful tool, especially with use of ODS
- Understanding the library system simplifies file management
  - LIBNAME statements precede saved codes
- PROCs provide very detailed and very customizable output
- Graphics generally lag behind other programs
Future Topics

- Macros
- Detailed tutorials on specific PROCs/analyses (MIXED)
- Dataset management, particularly arrays
- Customizing output
- Whatever there is interest in
Resources

- SAS online document: v9doc.sas.com
  - Similar to SAS help document
  - Larger/includes experimental procedures
- Applied Statistics and the SAS Programming Language
  - Cody and Smith, 5th edition
  - Well tailored to types of problems faced in psychology
- Little SAS book
  - Delwiche and Slaughter
  - Chapter 5: ODS
  - Otherwise, good overview of basic processes