22S:166
Computing in Statistics

Introduction to SAS

Lect 16
Oct. 22, 2008

Kate Cowles
374 SH, 335-0727
kcowles@stat.uiowa.edu

SAS

- SAS is the statistical software package most commonly used in business, government, and many other applied research settings
  - has the best data management capabilities
  - can handle huge datasets
  - not as user-friendly or well-designed as R or S-PLUS in some other ways
  - R or S-PLUS are often preferred for research in theoretical and methodological statistics
- *preparing* data for analysis usually takes more time and work than doing the actual analysis

Two ways to run SAS

- batch mode
  - use text editor to write a file containing SAS code
  - extension ".sas"; e.g. filename might by "conadp.sas"
  - run it in batch mode by entering "sas <filename>"; e.g. "sas conadp"
  - SAS produces two new output files
    - log file with .log extension
    - output file with .lst extension
- interactive mode
  - just enter "sas"
  - program editor, log, and output are three windows
  - online help available

How SAS programs and commands are organized

- Use a *DATA step* to organize your data by creating a SAS dataset.
- Then use *PROC steps* to analyze your data using SAS procedures.
  - Once you have created a SAS dataset, you may apply any SAS procedures to it during the current SAS session without recreating the dataset.
- DATA and PROC steps consist of SAS *statements*.
  - Each statement must end with a semicolon.
  - Most statements include one or more *keywords* that must be spelled exactly as shown.

The DATA step: Creating a SAS dataset

Before it can process data, SAS must read in the data in the form of a table with

- a row for each *observation*
- a column for each *variable*

You must choose a name for the entire dataset and a name for each variable.

SAS distinguishes between two types of variables:

- *numeric variables*, which contain only digits and decimal points and with which arithmetic operations may be done
- *character variables* (all other kinds of data).

SAS data step

- *data* statement
  - must begin the data step
  - gives name by which to refer to dataset during this SAS run
- *infile* statement
  - tells SAS in what physical file to find data
  - may optionally give SAS information about file and how to read it

- *input* statement
  - gives names to all variables
  - variables are assumed to be numeric unless you put a $ after name in input statement
  - may contain informats (goes *after* variable name to which it applies)
  - may contain @ <column-number> (goes *before* variable name to which it applies)
- may contain additional statements
  - formats for printing values of variables
  - calculating new variables
  - etc.
- to embed data right into the SAS program, *datalines* statement is used instead of *infile*
- ends with *run* statement

Starting SAS in different environments

- Running SAS in the ITC

  Click on Start/All Programs/SAS/SAS 9.1 (English). SAS will come up, but it will be running on the university's "Virtual Desktop," not on your local computer. That means that SAS will not be able to find files on the local drives. It should be able to find files in your H: drive, however.

  You can access SAS on the university's "Virtual Desktop" from other Windows computers, including from home. Go to itc.uiowa.edu and select "Virtual Desktop" on the left panel.

- Running SAS on the Linux network

  The Division of Mathematical Sciences has a SAS licence for only one of the computers on the Linux network. If you type "sas" at the Linux prompt, you will automatically be logged into that computer (you will be prompted for your password). SAS will come up in interactive mode.

To make Linux SAS work more like Windows SAS, do the following:

- Select the Program Editor window.
- Click "Tools" and "Options," and select "Preferences"
- Click the "Toolbox" tab and DEPRESS the three toggles (Display Tools Window, Display Command Window, Combine Windows)
- Click the "Editing" tab and UNPRESS the toggle "Automatically store selection"
- Click "OK"

Entering commands and programs

However you access SAS, you will get a screen that shows:

- a menu bar
- a log window
- a program editor window

Click in the program editor window. You may now type commands and programs in this window.

How SAS programs and commands are organized

Use a *DATA step* to organize your data by creating a SAS dataset. Then use *PROC steps* or automated features to analyze your data. Once you have created a SAS dataset, you may apply any SAS procedures or automated features to it during the SAS session without recreating the dataset.

DATA and PROC steps consist of SAS *statements*. Each statement must end with a semicolon. Most statements include one or more *keywords* that must be spelled exactly as shown.

The DATA step: Creating a SAS dataset

Before it can process data, SAS must read in the data in the form of a table with

- a row for each *observation*
- a column for each *variable*

You must choose a name for the entire dataset and a name for each variable. SAS has the following rules for names:

- SAS names must begin with a letter or an underscore. The remaining characters in a SAS name can be letters, numbers, or underscores. There must be no embedded blanks.

SAS distinguishes between two types of variables: *numeric variables*, which contain only digits and decimal points and with which arithmetic operations may be done, and *character variables* (all other kinds of data).