SAS Basics CFDR Fall Workshop, October 2006

Data Holdings
- CFDR affiliates are able to access many data sets, codebooks, and programs on the CFDR local area network online or CD-Rom.
- http://www.bgsu.edu/organizations/cfdr/dataresearch/data_holdings.htm
- Data holdings on the CFDR server: T:\Public\Data

SAS: Statistical Applications System
1. Open SAS
   - There are four main windows in SAS
     o Editor: type in, edit, and submit SAS programs

     o Log: gives notes, errors, or warnings associated with submitted programs

     o Output: contains any printable results generated from your program
       ▪ Tools, Options, Preferences, Results tab, Under HTML check Create HTML (check Use Work Folder), Style…Minimal

     o Results/Explorer: table of contents, allows you to maneuver through the libraries and the output

2. SAS programming
   - Written as "statements."

   - Two parts of a SAS program
     o DATA steps
       ▪ Create, read, and modify data
       ▪ Begin with DATA statements

     o PROC steps
       ▪ Perform specific analysis or function
       ▪ Begin with PROC statements

3. Getting to your data
   - Enter data directly into SAS data sets
   - Read other software's data files directly or convert into SAS data sets
   - Create SAS data sets from raw data files

   - Use a previously created SAS data set ☺
     o SAS program files (e.g. file with a .sas7bdat extension).

4. Using SAS program files
   - SAS data set names
     o Two-level name separated by a period
       ▪ 1st level: libref (SAS data library reference)
         • nickname that corresponds and points to a particular location where the data are stored (jump drive, cd, directory, folder, etc.)
         • Before you can use a libref, you have to define it…commonly done by using the libname statement (but there are other ways, p. 67 *LSB*)
           o LIBNAME libref 'drive:\directory';
           o This place must exist!

       ▪ 2nd level: member name that uniquely identifies the data set within the library (the filename of the sas data set without the extension).

       ▪ Both names must follow rules for valid SAS names

     o Examples…
       ▪ `libname NSFG6 'T:\Public\Data\NSFG\Wave 6\SAS datasets';`

         • Here, I have a libref called NSFG6. "NSFG6" is what I have named the location `'T:\Public\Data\NSFG\Wave 6\SAS datasets'`. Now, if I want to locate any files that are in `'T:\Public\Data\NSFG\Wave 6\SAS datasets'`, or create a file and store it in `'T:\Public\Data\NSFG\Wave 6\SAS datasets'`, I can refer to it with the libref NSFG6.

         • Within the `'T:\Public\Data\NSFG\Wave 6\SAS datasets'` folder, there are the following SAS data files:
           o ex3.sas7bdat
           o ex4.sas7bdat
           o female.sas7bdat
           o fempreg.sas7bdat
           o male.sas7bdat
           o preg.sas7bdat

       ▪ `libname mywork 'C:\Documents and Settings\mjanep\Desktop\SAS data';`
         • Here, I have a libref called mywork. "mywork" is what I have named the location `'C:\Documents and Settings\mjanep\Desktop\SAS data'` (a folder on the desktop called "SAS data". Now, if I want to locate any files that are in `'C:\Documents and Settings\mjanep\Desktop\SAS data'`, or create a file and

store it in `'C:\Documents and Settings\mjanep\Desktop\SAS data'`, I can refer to it with the libref mywork.

- Let's say I want to use the SAS data set called male.sas7bdat. I want to be able to save the data set and any changes I have made as a new data set called "example" in the folder "SAS data" my desktop. How do I do this??

```
libname NSFG6 'T:\Public\Data\NSFG\Wave 6\SAS datasets';

libname mywork 'C:\Documents and Settings\mjanep\Desktop\SAS
data';

data mywork.example; set nsfg6.male; run;
```

- Temporary versus permanent SAS data sets
  - Temporary: exists only during current job or session and is automatically erased by SAS when you close SAS

  - Permanent: exists somewhere (jump drive, folder, etc.) even after SAS session has ended

  - You don't explicitly tell SAS to make data set temporary or permanent…it is implied by the name you give it when you create or reference it.
    - Temporary data set: give it a two-level name with WORK for the *libref* OR give it a one-level name (member name only).
    - Permanent data set: give it a two-level name with anything but WORK for the *libref*.
      - REMEMBER…for a permanent data set, you must have defined the libref with a libname statement!

5. Some basic procedures
   - PROC CONTENTS: gives a description of the data set (data set name, number of observations, number of variables, date created, and information about each variable)
     - `proc contents; run;`

   - PROC PRINT: prints variables for all observations (or selected observations) in data set
     - `proc print data=mywork.female(obs = 100); run;`

   - PROC FREQ: produces variable frequency
     - `proc freq; tables <variables of interest>; run;`

- PROC MEANS: produces variable n, mean, standard deviation, minimum, and maximum
  - `proc means; var <variables of interest>; run;`

- PROC UNIVARIATE: produces variable distribution characteristics – mean, median, mode, standard deviation, skewness, and kurtosis
  - `proc univariate; var <variables of interest>; run;`

- PROC REG: OLS regression
  - `proc reg; model dv = iv1 iv2 iv3; run;`

6. Commenting out
   - This is used to make notations within a program that SAS will ignore. Any text that is commented out will turn green and will be ignored completely by SAS as is reads the rest of the program.
     - `/*Coding for family structure variables*/`
     - `*Coding for family structure variables;`

7. SAS Help
   - SAS Help Menu
   - CFDR website: http://www.bgsu.edu/organizations/cfdr/dataresearch/programming_research.htm
   - SAS Hours
   - Practice data on the CFDR server: T:\Public\Practice Data

Try it yourself!

1. Go to http://www.bgsu.edu/organizations/cfdr/events/workshop_series.htm
2. Click on sample.sas7bdat.
3. Save the data to your desktop.
4. Open SAS.
5. Type in the following (without the a, b, c…)

```
(a)     libname workshop "C:\Documents and Settings\user name\Desktop";

(b)     data tryit; set workshop.sample; run;

(c)     proc print data = tryit (obs = 100); run;

(d)     proc freq; tables age_a marstat numbabes; run;
```

Line (a) references the data we are using with the libref workshop. We are telling SAS that the data are located on the desktop.

Line (b) indicates that we want to create a temporary sas dataset called "tryit." This data is set to be exactly like the data set "sample." If we wanted to create a permanent data set, we would have to give "tryit" a two-level name (e.g. workshop.tryit). Because this data set is temporary, any changes we might make to it (recoding, creating new variables, creating subsamples, etc.) will not be saved when we close down SAS.

Line (c) is a procedure that allows us to see our data in a table format. Another way to see the data is to use the Explorer window.

Line (d) is a frequency procedure. We are telling SAS that we want to know the frequencies of the variables age_a, marstat, and numbabes. If we wanted to know the frequencies of all variables in the data set, we could just use proc freq; run;

Reference:
Delwiche, L. D. & Slaughter, S. J. (2003). *The Little SAS Book: A Primer, 3rd Ed.* Cary, NC: SAS Institute, Inc.