STAT 5110/6110: SAS Programming and Applications

1-B. Create a SAS Data Set

Peng Zeng

Department of Mathematics and Statistics

Auburn University

Different Ways of Creating a SAS Data Set

Different ways of creating a SAS data set

- Use DATA step
  - datalines statement: include data in the SAS program
  - set statement: from an existing SAS data set (discuss it later)
  - infile statement: from an external data file (discuss it later)
- Use Import Wizard
  - Read external data files (Excel, csv, etc)
  - Generate a SAS program automatically
- Use proc import
  - Read external data files (Excel, csv, etc)

DATA Step

A DATA step creates or modifies data sets

- Input can be raw data or an existing SAS data set.
- Create a new SAS data set
- Produce messages in the SAS log, no report or other output

For example, you can use DATA steps to do the following:

- create a SAS data set from raw data
- compute values
- check and correct errors in your data
- produce new SAS data sets by subsetting, concatenating, merging, and updating existing data sets

Create SAS Data Sets

```
data name-of-dataset;
   input var1 $ var2 var3;
   /* more statments; */
datalines;
   put-data-here
;
```

The input statement

- list the names of variables

The datalines statement

- should be placed after the input statement
- should be placed toward the end of the DATA step
- one row for one observation, values separated by space
- end with a semicolon at the beginning of a new line

Type of Variables

There are two major types of variables:

- character variables
  - can contain any values (letters, numbers, symbols, etc).
  - add $ after variable names in the input statement
  - The length of a character variable is 8 by default. So long characters may be chopped.
- numeric variables can contain only numeric values.
  - standard numbers such as 123, 95.67, or 34E+5
  - no formatted numbers such as 12,345 or $345.00

A value must exist for every variable for each observation. Missing values are valid values.

- Use period (.) to indicate a missing value in datalines.
- A character missing value is displayed as a blank.
- A numeric missing value is displayed as a period.

CSV Files

The data (`blood-pressure.csv`) are the systolic and diastolic blood pressure readings for 22 patients. (patient, gender, systolic, diastolic)

CK	M	120	50	SS	F	96	60	FR	F	100	70
CP	F	120	75	BL	M	140	90	ES	F	120	70
CP	M	165	110	JI	F	110	40	MC	M	119	66
FC	M	125	76	RW	F	133	60	KD	M	108	54
DS	M	110	50	JW	M	130	80	BH	F	120	65
JW	F	134	80	SB	M	118	76	NS	F	122	78
GS	F	122	70	AB	M	122	78	EC	F	112	62
HH	F	122	82								

For a csv (comma separated values) file:

- ASCII file (can be edited using Notepad)
- One row for each observation
- The names of variables on the first row
- Values are separated by comma (,)

Create SAS Data Sets from External Files

Following the steps below.

- Upload your data file to SAS Studio
- Right click the name of the data file and select “Import Data” (It also works if you double click the name of the data file)
- SAS Studio will create SAS codes for you

Procedure Import

Import data from a csv file.

```
proc import out = mySASdata datafile = "link-to-mydatafile.csv"
   dbms = csv replace;
run;
```

Import data from an Excel file. (dbms = xls or dbms = xlsx)

```
proc import out = mySASdata datafile = "link-to-mydatafile.xlsx"
   dbms = xlsx replace;
run;
```

- Only one semicolon (;) from proc to replace;

Option datafile

The option datafile = tells SAS the location of the external file.

- in a local drive (not applicable for SAS Studio in the cloud)

```
datafile = "C:\myfolder\data.csv"
```

- in the SAS cloud space

```
datafile = "/home/username/datasets/data.csv"
```

- a weblink: need to create a fileref first using filename statement

Statement filename

We can create an alias (fileref) for an external file.

```
filename myfile "C:\myfolder\data.csv";
```

```
filename myfile "/home/username/datasets/data.csv";
```

```
filename myfile url "http://www-weblink.com/data.csv";
```

Then use datafile = myfile in the import procedure.

- The myfile is a name (at most 8 characters) you choose.
- Pay attention to the extra option url for a weblink.
- It is slow to import data directly from a weblink.
- The filename url does not work for xlsx files.

In-Class Exercise

- Create a SAS data set using three different methods
  - copy the contents in `blood-pressure.txt` and use data step with datalines statement.
  - load data from `blood-pressure.csv` using proc import
  - load data from `blood-pressure.xlsx` using proc import
- If you copy the data from `blood-pressure.xlsx` and use data step with datalines statement, it does not work. Why?
- Use proc print to check the values in the SAS data set and compare them with the raw data