

Data Input and Output

Statistics 426: SAS Programming

Module 2

2021

Basic approach to data input and output

- (1) If inputting raw data, set up data in form of text (.txt), comma separated value (.csv), etc. in the folder of choice on your computer or copy the website address of the file
- (2) If using a SAS dataset, create a library using the folder where the file is
- (3) Use DATA step with either SET, INFILE, or CARDS (if copying/)
- (4) Can use WHERE or other subsetting statements to create datasets
- (5) Analyze the data
- (6) Print results to the console or printer and/or store results in a file on your computer
- (7) Save your SAS program if desired, and close program if finished

Address input

Addresses of data files are input into SAS differently, depending on whether or not a website address is being used or a computer physical address is being used.

Website addresses: are input as they appear in the address bar of browsers, using / forward slash

Physical computer address: SAS needs the computer address to use \ back slashes to separate folder names

Data formats

Your uidaho OneDrive works great as the folder you have for your data.

I use .csv and .txt (or copy/paste with CARDS) primarily for data input rather than others for ease of use (they cause me less frustration). Additionally, there are other file types that require specific syntax to input them.

You can also read in datasets from websites with just a web address, which may just be my preferred way (if my data is on a website) when using INFILE.

Libraries

SAS needs to know where to read data files to and from, and always inserting a long computer address is not always fun nor easy. Changing the working directory to the location of choice makes the process of reading in the dataset more manageable. They are not necessary (as you can use the WORK library) but can make things more organized with a permanent library

General form of DATA step with INFILE

INFILE specifies an external file to read into SAS. The best file types are .csv and .txt files. Excel requires some extra SAS packages to be installed (purchased) to work correctly. Variables are assumed to be numeric unless otherwise specified in the INPUT statement with a \$ following the character variable name(s).

```
DATA libref.datasetname;
INPUT v1 v2 v3$ ... ;
INFILE 'address\filename.ext' DLM=',' firstobs=2;
RUN;
```

General form of FILENAME

FILENAME will save the address of the file you want, either from a website or other mediums.

Website:

```
FILENAME fileref URL 'address';
```

fileref: name of reference, limit 1-8 characters, starting with letter or _
'address': http protocol address, in quotes

Physical computer address:

```
FILENAME fileref 'address';
```

General form of DATA step with CARDS

The following will read in data that you either input manually or copy/paste. The CARDS statement is interchangeable with the DATALINES statement

```
DATA datasetname;
INPUT var1 var2 var3$ ...;
CARDS;
<input data here>
;
RUN;
```

General form of DATA step with SET

SET allows to read in an existing SAS dataset as the basis of the new dataset being created; the new dataset specified in the DATA statement is created using an existing SAS dataset specified in the SET statement, often by subsetting it based on conditions

```
DATA datasetname;
SET input-datasetname;
WHERE where-expression;
KEEP variable-list;
LABEL variable1='Label'
variable2='Label'
... ;
FORMAT variable1 format.
variable2 format.
... ;
RUN;
```

Analyses

Plot, calculate, subset, etc. your dataset

Printing and storing results

You can create multiple datasets based on a condition. We will go through this more in depth at in a later module but will cover a basic way to subset based on a single condition.

```
if cond='Y' then output data1;
else output data2;
```

Closing your SAS session

You can save your program file. Go to File to save as a .sas file (or copy and paste code into a text file).

GPA data: Addresses

The folder where I will have my files stored is: `"/Users/reanaes/OneDrive - University of Idaho/reanaes/Documents/Classes/R Class"`. I also have it at `http://webpages.uidaho.edu/~reanaes/Data/GPAdat2.csv` for .csv and `http://webpages.uidaho.edu/~reanaes/Data/GPAdat2.txt` for text file. Because address formats differ between OS, using a web address is simple, *if* it is available.

GPA data: Library

We could make a library to save things into but that will happen in a later module

Mac: `\Users\reanaes\OneDrive - University of Idaho\reanaes\Documents\Classes\R Class`

PC: `C:\reanaes\OneDrive - University of Idaho\reanaes\Documents\Classes\R Class`

One way in Windows to find the address is in the Windows Explorer, then you would have to add the filename and extension

GPA data read in with INFILE only

```
data gpa;
infile 'C:\reanaes\OneDrive - University of Idaho\reanaes\Documents\Classes\
R Class\GPAdat2.csv' dlm=',' firstobs=2;
input uigpa act sex$ housing$ college$;
run;
```

GPA data read in with FILENAME for computer address

```
FILENAME gpadata 'C:\reanaes\OneDrive - University of Idaho\reanaes\Documents\
Classes\R Class\GPAdat2.csv';
```

```
data gpa;
infile gpadata dlm=',' firstobs=2;
input uigpa act sex$ housing$ college$;
run;
```

GPA data read in with FILENAME for web address

```
FILENAME gdata URL 'http://webpages.uidaho.edu/~reanaes/Data/GPAdat2.csv';
```

```
data gpa;
infile gdata dlm=',';
input uigpa act sex$ housing$ college$;
run;
```

```
proc print data=gpa;
run;
```

GPA data step log

```
37 FILENAME gdata URL 'http://webpages.uidaho.edu/~reanaes/Data/GPAdata2.csv';
38
39 data gpa;
40 infile gdata dlm=',' firstobs=2;
41 input uigpa act sex$ housing$ college$;
42 run;
```

NOTE: The infile GDATA is:
Filename=http://webpages.uidaho.edu/~reanaes/Data/GPAdata2.csv,
Local Host Name=VLabApps04,
Local Host IP addr=::1,
Service Hostname Name=WEB01-webpages.its.uidaho.edu,
Service IP addr=129.101.105.230,
Service Name=N/A,Service Portno=443,
Lrecl=32767,Recfm=Variable

NOTE: 154 records were read from the infile GDATA.
The minimum record length was 18.
The maximum record length was 23.

NOTE: The data set WORK.GPA has 154 observations and 5 variables.

NOTE: DATA statement used (Total process time):
real time 0.07 seconds
...cpu time 0.01 seconds

GPA data proc print

The SAS System

Obs	uigpa	act	sex	housing	college
1	4.00	32	F	R	COS
2	4.00	29	F	R	CNR
3	3.00	26	M	T	COS
4	3.80	32	M	G	CLASS
5	3.80	27	F	G	CLASS
6	3.90	23	M	G	CLASS
7	3.90	32	F	T	COS
8	3.05	21	M	T	CALS
9	2.66	27	M	T	CBE
10	3.00	25	M	T	CALS

GPA data proc print log

```
63  proc sgplot data=gpa;  
64  scatter x=act y=uigpa;  
65  run;
```

```
NOTE: PROCEDURE SGPLOT used (Total process time):  
      real time          1.17 seconds  
      cpu time           0.35 seconds
```

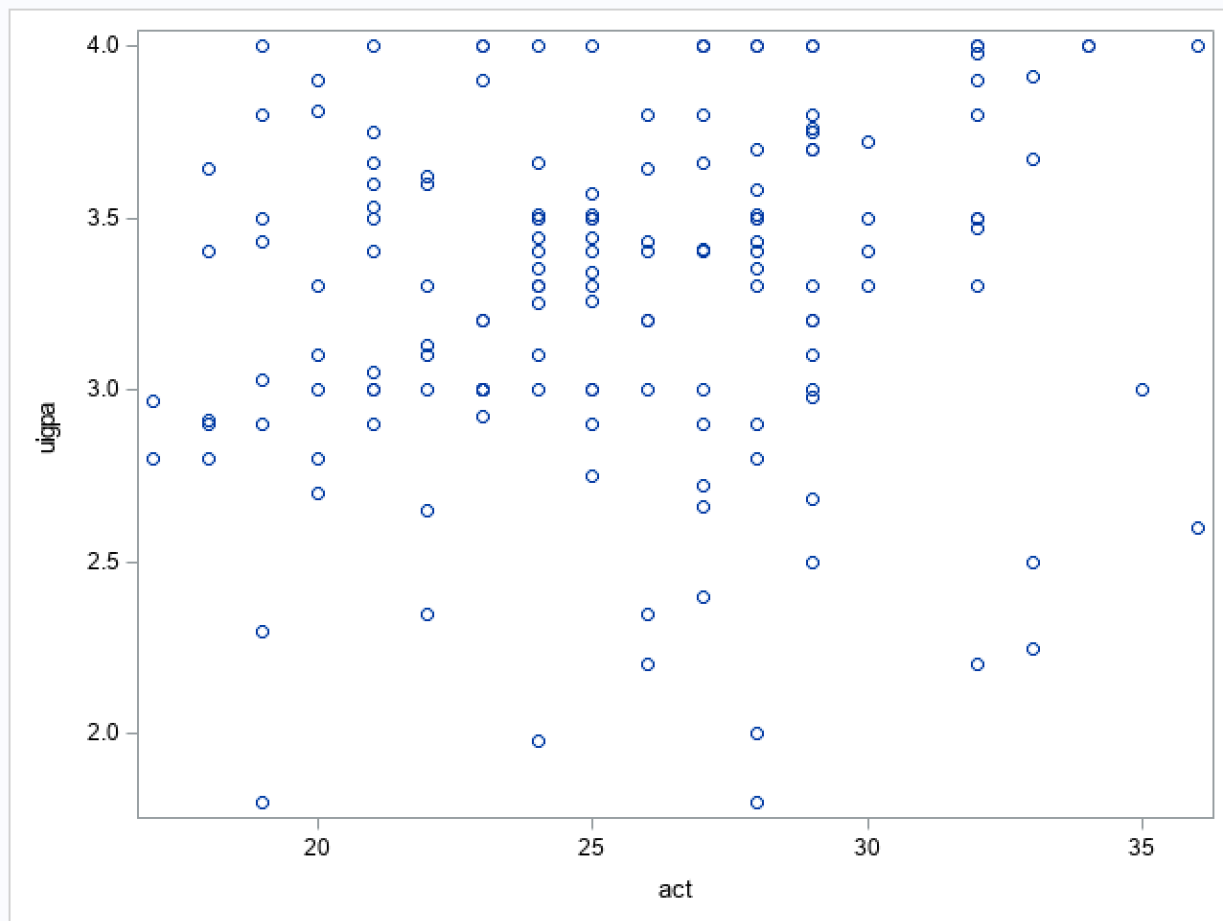
```
NOTE: There were 154 observations read from the data set WORK.GPA.
```

GPA data analyses: plot

Looking at the relationship between college GPA and ACT scores.

```
proc sgplot data=gpa;  
scatter x=act y=uigpa;  
run;
```

GPA data plot



GPA data plot log

```
84 data m_gpa f_gpa;
85 set gpa;
86 if sex='M' then output m_gpa;
87 else output f_gpa;
88 run;
```

```
NOTE: There were 154 observations read from the data set WORK.GPA.
NOTE: The data set WORK.M_GPA has 87 observations and 5 variables.
NOTE: The data set WORK.F_GPA has 67 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time          0.03 seconds
      cpu time           0.01 seconds
```

GPA data analyses: prepping data

Here we will look at separating the dataset into males and females. This will be revisited in more detail in a later module

```
data m_gpa f_gpa;
set gpa;
if sex='M' then output m_gpa;
else output f_gpa;
run;
proc print data=m_gpa;
run;
proc print data=f_gpa;
run;
```

gpa subset data step log

```
84 data m_gpa f_gpa;
85 set gpa;
86 if sex='M' then output m_gpa;
87 else output f_gpa;
88 run;
```

```
NOTE: There were 154 observations read from the data set WORK.GPA.
NOTE: The data set WORK.M_GPA has 87 observations and 5 variables.
NOTE: The data set WORK.F_GPA has 67 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time          0.03 seconds
      cpu time           0.01 seconds
```

m_gpa proc print

The SAS System

Obs	uigpa	act	sex	housing	college
1	3.00	26	M	T	COS
2	3.80	32	M	G	CLASS
3	3.90	23	M	G	CLASS
4	3.05	21	M	T	CALS
5	2.66	27	M	T	CBE

m_gpa log

```
89 proc print data=m_gpa;  
90 run;
```

```
NOTE: There were 87 observations read from the data set WORK.M_GPA.  
NOTE: PROCEDURE PRINT used (Total process time):  
      real time           0.03 seconds  
      cpu time            0.01 seconds
```

f_gpa proc print

The SAS System

Obs	uigpa	act	sex	housing	college
1	4.00	32	F	R	COS
2	4.00	29	F	R	CNR
3	3.80	27	F	G	CLASS
4	3.90	32	F	T	COS
5	2.80	18	F	G	CALS

f_gpa log

```
91  proc print data=f_gpa;  
92  run;
```

NOTE: There were 67 observations read from the data set WORK.F_GPA.

NOTE: PROCEDURE PRINT used (Total process time):

real time	0.03 seconds
cpu time	0.01 seconds
