Session 2 - SPSS[1]

Reading external raw data. Reading more complicated raw data – a hierarchical data set. LEAVE statement. Conditional input statements. Common errors in SPSS.

Time for victory dances – who was able to figure out how to read data from an outside file?

```
** do not type and run the following - read and understand .
title 'prog2.sps' .

DATA LIST  file = 'c:\all\spssclass\progs\prog2.data' FREE
   / ID * name (A9) AGE (f3) hinch (f3) weight (f3) gender (A) .
variable label
 name   'First name'
 age    'Age in years'
 hinch  'Height in inches'
 weight 'Weight in pounds' .
list var = all .
freq var = all .
```

You can tell SPSS to read external data by adding “file = ‘full path’ ” to the DATA LIST command line.

READING HIERARCHICAL DATA

Following is an example of a program that reads a hierarchical, or nested data set. This means that not all of the records are alike. In this example, there are household records, and person records. It is similar to census PUMS files. The variables in the household records are not the same as those in the person records. You may type or copy this program and run it. After the program, I explain each line.

```
* prog3.sps.
* Read more complicated data - a hierarchical data set.
* This is a PUMS-Like file - there are Household and Person records .
* The two different types of records have different variables .
** The data (as a comment) follow:
h011360010
p0101125
p0102222
p0103302
p0104301
h023130010
p0201145
p0202325
h034530010
p0301166     .

input program.

DATA LIST file = 'c:\all\spssclass\progs\prog3.data'   /RECTYPE 1 (a).

do if (RECTYPE eq 'h').
  reread.
     DATA LIST file = 'c:\all\spssclass\progs\prog3.data'
        / serial 2-3 region 4 statefip 5-6 county 7-10 stateco 5-10 .
         leave serial region statefip county stateco .

else if  (RECTYPE eq 'p').
  reread .
     DATA LIST file = 'c:\all\spssclass\progs\prog3.data'
         / serial 2-3 pernum 4-5 relate 6 age 7-8.
```

[1] This document has been prepared by Patty Glynn, University of Washington.

```
end case.
end if.

end input program.

execute .

variable label
 rectype  'Record Type'
 serial   'Number unique to HH'
 region   'Geographic Region'
 statefip 'State Fips Code'
 county   'County ICPSR Code'
 stateco  'Fips State + ICPSR County'
 pernum   'Person number in HH'
 relate   'Relationship to head'
 age      'Age in years' .

list var = all .
```

SYNTAX:

```
input program.
```

EXPLANATION:

THIS COMMAND TELLS SPSS THAT AN INPUT PROGRAM IS BEGINNING. THIS IS NOT NECESSARY READING SIMPLE FILES, BUT IS REQUIRED FOR READING THIS MORE COMPLEX FILE STRUCTURE.

SYNTAX:

```
DATA LIST file = 'c:\all\spssclass\progs\prog3.data'   /RECTYPE 1 (a).
```

EXPLANATION:

READ THE FIRST COLUMN OF THE FILE 'c:\all\spssclass\progs\prog3.data' AS A CHARACTER VARIABLE.

SYNTAX:

```
do if (RECTYPE eq 'h').
  reread.
     DATA LIST file = 'c:\all\spssclass\progs\prog3.data'
        / serial 2-3 region 4 statefip 5-6 county 7-10 stateco 5-10 .
```

EXPLANATION:

IF THE CHARACTER YOU JUST READ IS 'H', REREAD THE SAME LINE IN THE FILE, AND READ THE VARIABLE SERIAL IN COLUMNS 2-3, REGION IN COLUMN 4, STATEFIP IN COLUMNS 5-6, COUNTY IN COLUMNS 7-10, AND STATECO IN COLUMNS 5-10. NOTE THAT COLUMNS 5-10 ARE READ INTO TWO DIFFERENT VARIABLES. THIS IS OKAY WITH SPSS.

SYNTAX:

```
   leave serial region statefip county stateco .
```

EXPLANATION:

THE COMMAND "LEAVE" ASKS SPSS TO KEEP VALUES OF A VARIABLE UNTIL NEW VALUES ARE ASSIGNED. NORMALLY, VALUES IN A VARIABLE ARE NOT KEPT FOR THE NEXT CASE. THE PERSON RECORDS WOULD HAVE MISSING VALUES ON THESE HOUSEHOLD LEVEL VARIABLES IF THE LEAVE COMMAND WERE NOT INCLUDED.

SYNTAX:

```
else if  (RECTYPE eq 'p').
  reread .
     DATA LIST file = 'c:\all\spssclass\progs\prog3.data'
         / serial 2-3 pernum 4-5 relate 6 age 7-8.
```

EXPLANATION:
IF THE VARIABLE “RECTYPE” IS EQUAL TO ‘p’, REREAD THE SAME CASE IN THE SAME DATA FILE, AND READ THE VARIABLES AS SPECIFIED.

SYNTAX:
```
end case.
```

EXPLANATION:
I AM DONE READING THAT LINE OF DATA, GO ON TO THE NEXT CASE.

SYNTAX:
```
end if.
```

EXPLANATION:
THIS ENDS THE ‘DO IF” STARTED WITH do if (RECTYPE eq 'h').

SYNTAX:
```
end input program.
```

EXPLANATION:
THIS TELLS SPSS THAT I AM FINISHED WITH THE COMPLEX INPUT PROGRAM THAT I TOLD IT I WAS GOING TO DO WHEN I SAID input program.

SYNTAX:
```
execute .
```

EXPLANATION:
THE EXECUTE COMMAND TELLS SPSS TO CREATE THE VARIABLES THAT YOU TOLD IT ABOUT. UNTIL YOU SAY EXECUTE, IT MAY NOT COMPLETE THE JOB.

SYNTAX:
```
variable label
 rectype  'Record Type'
 serial   'Number unique to HH'
 region   'Geographic Region'
 statefip 'State Fips Code'
 county   'County ICPSR Code'
 stateco  'Fips State + ICPSR County'
 pernum   'Person number in HH'
 relate   'Relationship to head'
 age      'Age in years' .
```

EXPLANATION:
GIVE VARIABLE LABELS TO THE VARIABLES. VARIABLE NAMES CAN HAVE ONLY UP TO 8 CHARACTERS. LABELS CAN BE UP TO 40 CHARACTERS, SO CAN GIVE MORE DESCRIPTIVE. THIS IS AN IMPORTANT PART OF DOCUMENTION PROGRAMS.

SYNTAX:
```
list var = all .
```

EXPLANATION:
LIST VARIABLES FOR ALL CASES.

Following is the output from this program.   Notes from SPSS included.

```
* prog3.sps.
title 'prog3.sps' .
```

prog3.sps

```
* Read more complicated data - a hierarchical data set.
* This is a PUMS-Like file - there are Household and Person records .
* The two different types of records have different variables .
** The data (as a comment) follow:
h011360010
p0101125
p0102222
p0103302
p0104301
h023130010
p0201145
p0202325
h034530010
p0301166     .

input program.

DATA LIST file = 'c:\all\spssclass\progs\prog3.data'   /RECTYPE 1 (a).

Data List will read 1 records from c:\all\spssclass\progs\prog3.data

Variable   Rec   Start     End         Format

RECTYPE      1       1       1         A1


do if (RECTYPE eq 'h').
  reread.
     DATA LIST file = 'c:\all\spssclass\progs\prog3.data'
        / serial 2-3 region 4 statefip 5-6 county 7-10 stateco 5-10 .

Data List will read 1 records from c:\all\spssclass\progs\prog3.data

Variable   Rec   Start     End         Format

SERIAL       1       2       3         F2.0
REGION       1       4       4         F1.0
STATEFIP     1       5       6         F2.0
COUNTY       1       7      10         F4.0
STATECO      1       5      10         F6.0

         leave serial region statefip county stateco .

else if  (RECTYPE eq 'p').
  reread .
     DATA LIST file = 'c:\all\spssclass\progs\prog3.data'
         / serial 2-3 pernum 4-5 relate 6 age 7-8.

Data List will read 1 records from c:\all\spssclass\progs\prog3.data

Variable   Rec   Start     End         Format

SERIAL       1       2       3         F2.0
PERNUM       1       4       5         F2.0
RELATE       1       6       6         F1.0
AGE          1       7       8         F2.0

end case.
end if.

end input program.

execute .

variable label
 rectype  'Record Type'
 serial   'Number unique to HH'
 region   'Geographic Region'
 statefip 'State Fips Code'
 county   'County ICPSR Code'
 stateco  'Fips State + ICPSR County'
 pernum   'Person number in HH'
 relate   'Relationship to head'
 age      'Age in years' .

list var = all .
```

```
List
```

```
Notes
                  Output              06-APR-
                 Created                 2001
                                     08:25:41
                Comments
                   Input      Filter   <none>
                              Weight   <none>
                          Split File   <none>
                           N of Rows        7
                          in Working
                           Data File
                  Syntax             list var =
                                          all .
               Resources  Elapsed0:00:00.00
                                Time
```

```
RECTYPE SERIAL REGION STATEFIP COUNTY STATECO PERNUM RELATE AGE

p           1      1       36       10   360010    1      1    25
p           1      1       36       10   360010    2      2    22
p           1      1       36       10   360010    3      3     2
p           1      1       36       10   360010    4      3     1
p           2      3       13       10   130010    1      1    45
p           2      3       13       10   130010    2      3    25
p           3      4       53       10   530010    1      1    66


Number of cases read:  7    Number of cases listed:  7
```

Homework for next session:

To get an idea of the kinds of data that SPSS can read, skim the following chapters. You can spend hours on this topic. Spend as much time as you want, but don't feel like you need to learn everything in it. But, it is good to get exposed to the variety of options so that you will have an idea of the tools available to you when you need them.

Read the chapter on **Data List** and **Missing Values** in: https://depts.washington.edu/csde1/spssbase.pdf

Also, to help prepare you for errors you will encounter as you run programs, read:

Debugging and Data Cleaning Techniques with SPSS. http://staff.washington.edu/glynn/debugsp.pdf

Until the hour ends, play with SPSS, make mistakes so that you can see what errors look like, look at documentation, ask questions, work together, etc.