

**Block 2:     Analysing one variable****Interval and ratio variables****2.2.1.2 [BSA86] Exercise - Reading in data for interval variables**

[4 December 2010]

**Previous session:**    2.2.1.1 [myclass] Tutorial and work-through - Frequencies**Exemplar:**    [British Social Attitudes](#) (1986 survey<sup>1</sup>)**Research question:**

What is the distribution of the number of people in a household, including the respondent?  
What is the average number of persons per household?

What is the age distribution of the sample and what is its average age? What shape does the distribution have? Where are the cutting points for the oldest 10% and 25% and the youngest 10% and 25%?

**Task:**            1: Read in raw data for number of people in a household and age of the respondent.  
                     2: Save contents of data editor in a **\*.sav** file and of syntax editor in a **\*.sps** file.

**File:**    [bsa86.txt](#)    (raw data from the 1986 British Social Attitudes survey)

**SPSS commands<sup>2</sup> used:****TITLE****DATA LIST****EXECUTE****LIST**

A bit of repetition won't hurt you and it's good for reinforcing what you've already learned, so we're going to do this from scratch on the raw data.

---

<sup>1</sup> Data files and user documentation obtainable via the UK Data Archive at Essex University as SN 2315. See <http://www.data-archive.ac.uk/findingData/snDescription.asp?sn=2315> for details and conditions for access.

<sup>2</sup> **General formats:**

<b>TITLE</b>	'<Any text>' .
<b>DATA LIST</b>	<b>FILE =</b> '<location and name of external data file>' <b>RECORDS =</b> <number of records per case> / <record> <varname(s)> <column(s)> <varname(s)> <column(s)> ~ ~ ~ / <record> <varname(s)> <column(s)> ~ ~ ~ .
<b>EXECUTE .</b>	[Executes all pending operations.]
<b>LIST</b>	<b>/CASES</b> <n> .

First we need to look at the questionnaire to find the relevant questions and know where to find the appropriate data. Below is a facsimile extract from page 43B of the interviewer-administered<sup>3</sup> questionnaire which has all the information we need.

- 43B -  
SECTION NINE

		Col./ Code	Skip to																				
113.a)	<u>ASK ALL</u> Finally, a few questions about you and your household. Including yourself, how many people live here regularly as members of this household? <u>INTERVIEWER:</u> CHECK INTERVIEWER MANUAL FOR DEFINITION OF HOUSEHOLD IF NECESSARY. <div style="text-align: right;">WRITE IN: <span style="border: 1px solid black; display: inline-block; width: 20px; height: 15px; vertical-align: middle;"></span> <span style="border: 1px solid black; display: inline-block; width: 20px; height: 15px; vertical-align: middle;"></span></div>	CARD 15	(1508-9)																				
b)	And can I just check your own marital status? At present are you ... <u>READ OUT</u> ... <div style="float: right; text-align: right;">             ... married,              living as married,              separated or divorced,              widowed,              or - not married?           </div>	1 2 3 4 5																					
114.	Now I'd like to ask for a few details about each person in your household. Starting with yourself, what was your age last birthday? WORK DOWN COLUMNS OF GRID FOR EACH HOUSEHOLD MEMBER.																						
	<table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th style="width: 10%;">Resp- ondent</th> <th style="width: 5%;">2</th> <th style="width: 5%;">3</th> <th style="width: 5%;">4</th> <th style="width: 5%;">5</th> <th style="width: 5%;">6</th> <th style="width: 5%;">7</th> <th style="width: 5%;">8</th> <th style="width: 5%;">9</th> <th style="width: 5%;">10</th> </tr> </thead> <tbody> <tr> <td>1511</td> <td>1515</td> <td>1520</td> <td>1525</td> <td>1530</td> <td>1535</td> <td>1540</td> <td>1545</td> <td>1550</td> <td>1555</td> </tr> </tbody> </table>	Resp- ondent	2	3	4	5	6	7	8	9	10	1511	1515	1520	1525	1530	1535	1540	1545	1550	1555		
Resp- ondent	2	3	4	5	6	7	8	9	10														
1511	1515	1520	1525	1530	1535	1540	1545	1550	1555														
a) <u>Sex:</u>	<table border="1" style="margin-left: auto; margin-right: auto;"> <tbody> <tr> <td style="width: 10%;">Male</td> <td style="width: 5%;">1</td> <td style="width: 5%;">1</td> <td style="width: 5%;">1</td> <td style="width: 5%;">1</td> <td style="width: 5%;">1</td> <td style="width: 5%;">1</td> <td style="width: 5%;">1</td> <td style="width: 5%;">1</td> <td style="width: 5%;">1</td> </tr> <tr> <td>Female</td> <td>2</td> <td>2</td> <td>2</td> <td>2</td> <td>2</td> <td>2</td> <td>2</td> <td>2</td> <td>2</td> </tr> </tbody> </table>	Male	1	1	1	1	1	1	1	1	1	Female	2	2	2	2	2	2	2	2	2		
Male	1	1	1	1	1	1	1	1	1														
Female	2	2	2	2	2	2	2	2	2														
b) <u>Age last birthday:</u>	<table border="1" style="margin-left: auto; margin-right: auto;"> <tbody> <tr> <td>1512-3</td> <td>1516-7</td> <td>1521-2</td> <td>1526-7</td> <td>1531-2</td> <td>1536-7</td> <td>1541-2</td> <td>1546-7</td> <td>1551-2</td> <td>1556-7</td> </tr> </tbody> </table>									1512-3	1516-7	1521-2	1526-7	1531-2	1536-7	1541-2	1546-7	1551-2	1556-7				
1512-3	1516-7	1521-2	1526-7	1531-2	1536-7	1541-2	1546-7	1551-2	1556-7														

Look closely at Q.113a)

- 43B -  
SECTION NINE

		Col./ Code
113.a)	<u>ASK ALL</u> Finally, a few questions about you and your household. Including yourself, how many people live here regularly as members of this household? <u>INTERVIEWER:</u> CHECK INTERVIEWER MANUAL FOR DEFINITION OF HOUSEHOLD IF NECESSARY. <div style="text-align: right;">WRITE IN: <span style="border: 1px solid black; display: inline-block; width: 20px; height: 15px; vertical-align: middle;"></span> <span style="border: 1px solid black; display: inline-block; width: 20px; height: 15px; vertical-align: middle;"></span></div>	CARD 15
		(1508-9)

The number of persons in the household is asked in question 113a) and the reply is to be written in as a 2-digit number (with leading 0 or blank for fewer than 10 persons) to be punched in columns 8-9 of card 15 (there's a small **(1508 - 9)** in the margin)

WRITE IN:   (1508-9)

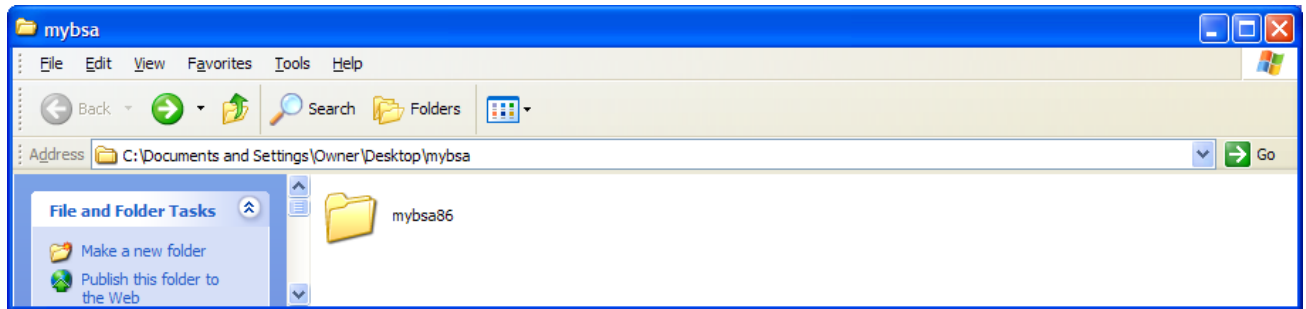
Similarly the age of the respondent last birthday is asked at question 106b) and is entered as a 2-digit number to be punched in columns 12-13 of card 15 (**1512 - 3** in margin)

b) Age last birthday:   1512-3

Without a lot of additional scripting or macro facilities not covered in this course, SPSS cannot easily be made to read files from internet addresses. You will therefore need to download and save raw data and SPSS files from this site.

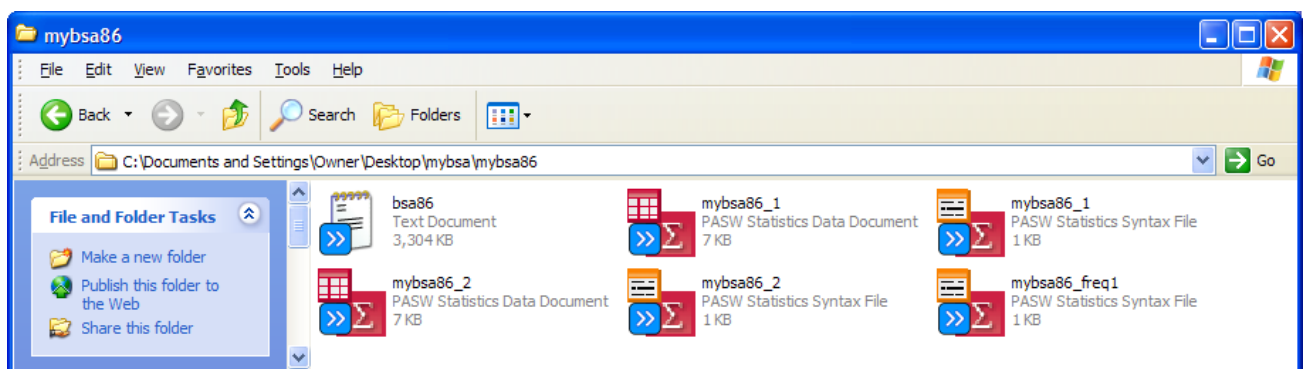
<sup>3</sup> There were two versions of the interviewer administered questionnaire (with a common core) each covering a different set of topics. This is from version B. Respondents also filled in a self-completion questionnaire to be collected later or returned by post.

Go to folder **mybsa**. If you have done the previous exercises your folder should look like this:



[If you don't already have folder **mybsa86**, go back to the [Block 2 menu](#) and do exercises **2.1.2.3 to 2.1.2.7** and the housekeeping exercise **2.1.2.8**.]

Double-click on **mybsa86**



The full pathway name for the raw data file is:

**'C:\Documents and Settings\Owner\Desktop\mybsa\mybsa86\bsa86.txt'**

which is quite long, but it's the best solution. When you need it for SPSS you only have type the pathway in once (or you can also copy and paste it from here).

### Work-through exercise

External data file **'C:\Documents and Settings\Owner\Desktop\mybsa\mybsa86\bsa86.txt'** has 23 records per case.

Write a set of SPSS commands to read in the number of persons from interviewer questionnaire Q105a (record 15 columns 8-9) and age of respondent from Q106b (record 15 columns 12-13). There are no missing values for number of persons in the household, but code 99 has been entered where age has been refused or is not known. Use the **positional** convention for variable names and specify missing values and variable labels. Again, it helps to prepare a table with your variable names and locations: the question number and summary question text can also be used later as variable labels.

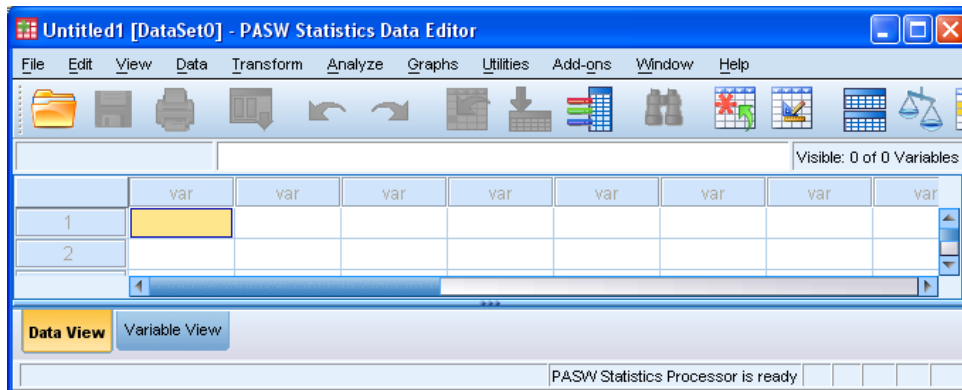
Question	Record number	Variable name	Start column	End column
Q105a Persons in household	15	v1508	8	9
Q106b Age of respondent	15	v1512	12	13

Open SPSS:

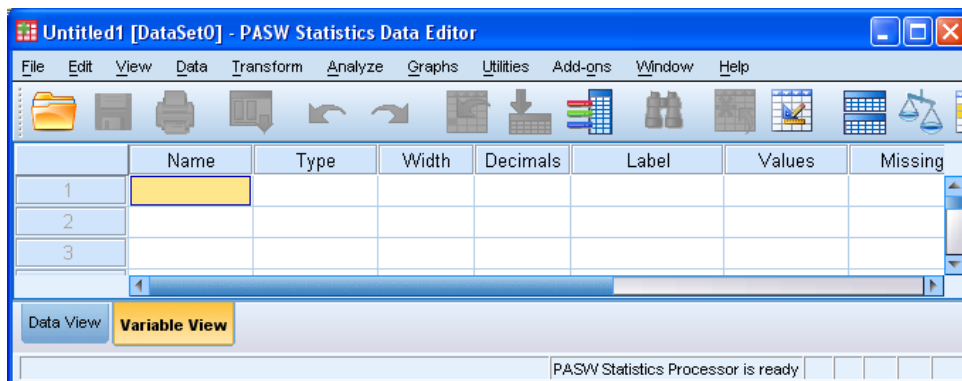


... and click **Cancel** to display a blank Data Editor (behind this pane). It will be called **\*Untitled1**.

If it opens in **Data View** :



... click on **Variable View** in the bottom left corner to display it in **Variable View** (this one has been adjusted by shrinking it to three rows):

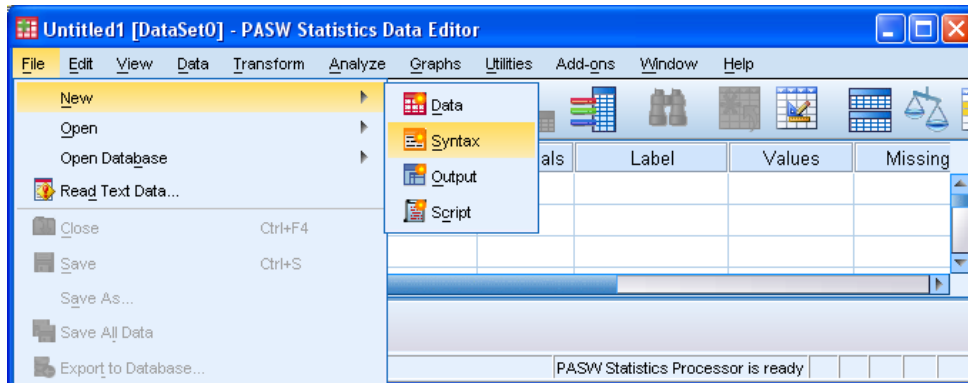


Like the magician's black velvet bag, you can see it is empty on both views.

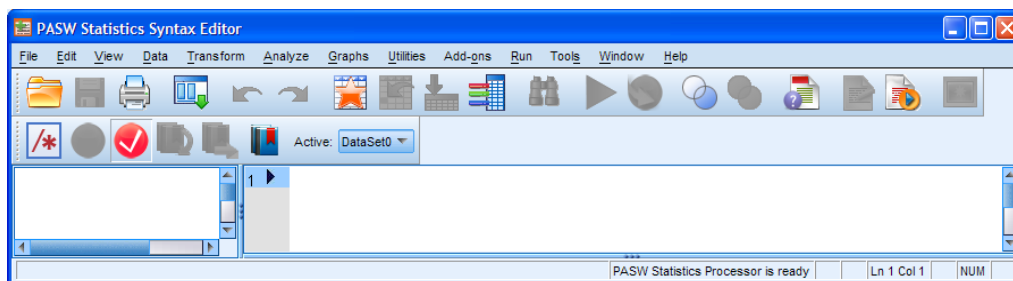
For most exercises, but not all, we are going to use SPSS in syntax mode in preference to drop-down menus with point-and-click, so we need to write out a set of commands which SPSS can then execute. For long and complex jobs we could do this in MS Word (or in a **.txt** file using other word-processing software, and then drag the text into SPSS, but it's easier for very short jobs to do it inside SPSS using the syntax editor.

Look in the top left corner of the Data Editor where you will see File. Open a new SPSS syntax editor by clicking on:

**File** > **New** > **Syntax**



. . . to display a blank syntax editor:



Before reading on, try writing out the syntax in the space below for **TITLE** and **DATA LIST**

**Don't forget the primes and full stops!**

**title**      ' \_\_\_\_\_ ' .

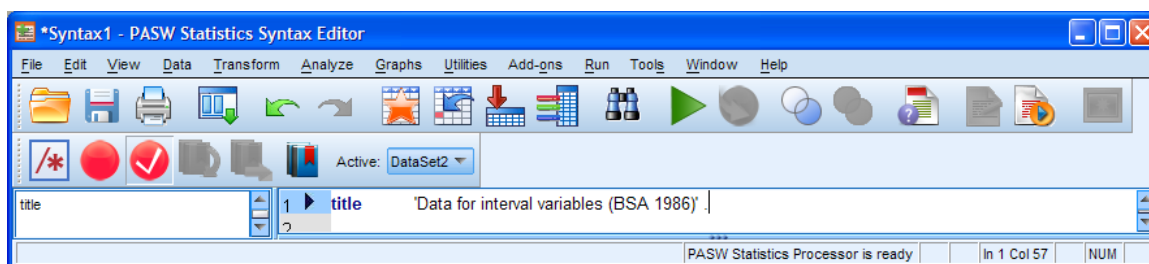
**data list**      \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_ .

**execute .**

## Work-through exercise

Using the right hand pane give your job a **title** (any text in primes) and type it into the syntax editor:

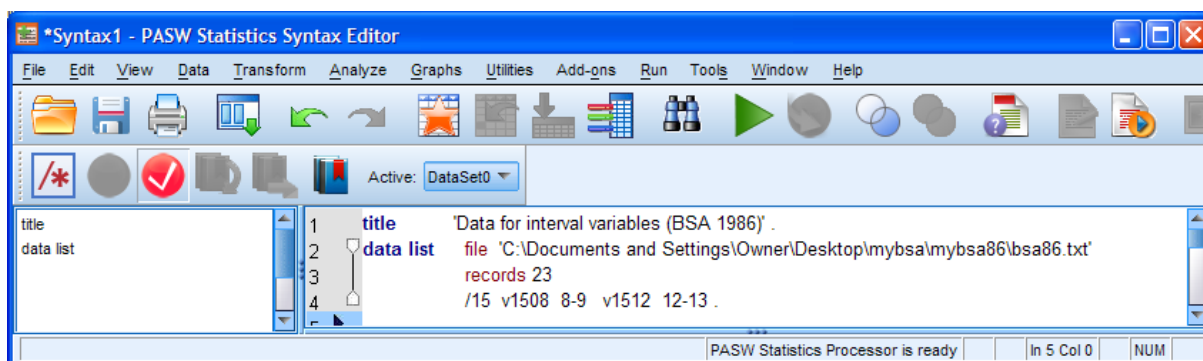
**title**      **'Data for interval variables (BSA 1986)' .**



Don't worry about the colour change for the **TITLE** command. Release 18 of SPSS now uses colour coding for commands, subcommands and keywords. This is very helpful as we shall see.

Using the **positional** convention for naming variables, write a **DATA LIST** command to read the data for number of persons in the household and age of respondent from external file:  
**'C:\Documents and Settings\Owner\Desktop\mybsa\bsa86.txt'**

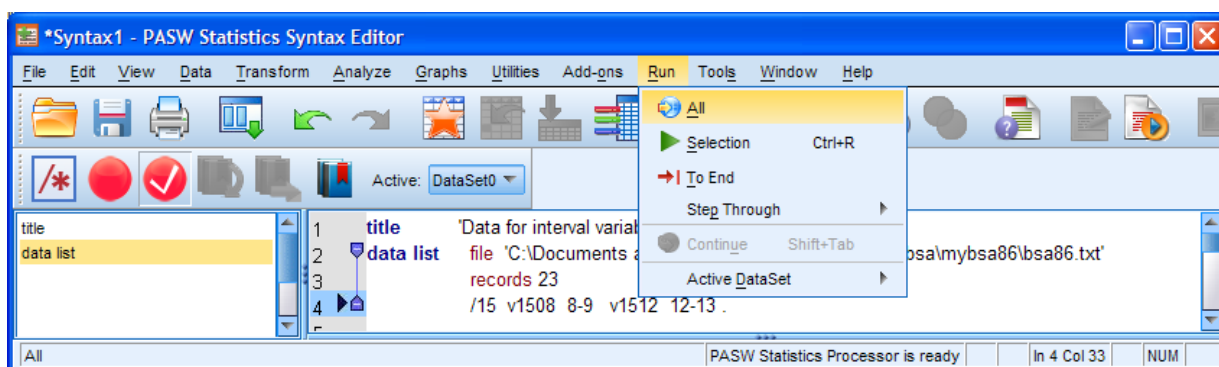
**data list**      **file 'C:\Documents and Settings\Owner\Desktop\mybsa\mybsa86\bsa86.txt'**  
                 **records 23**  
                 **/15 v1508 8-9 v1512 12-13 .**



[Note the colour coding for the **keywords**]

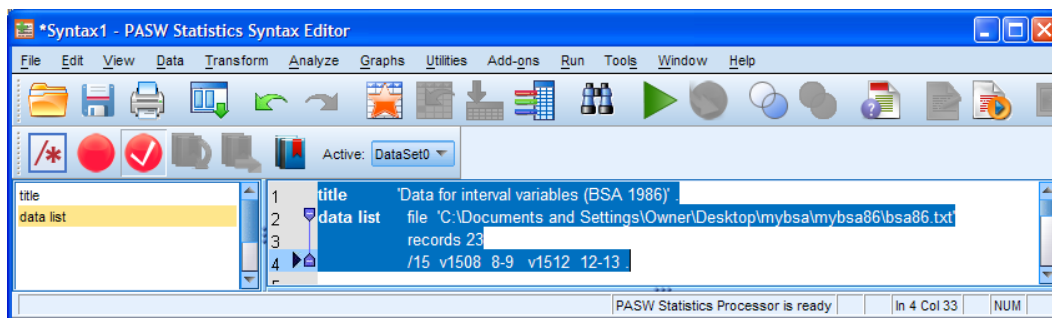
As you type, or if you make a syntax mistake, SPSS will display part or all of the text in **red** in the syntax pane until the command is complete and ends with a full stop (period).

When you are happy with your command syntax, click on **Run > All**



... or highlight the whole text:





... and run the job by pressing **[CTRL]+R** or by clicking on the green ► icon in the toolbar.

*[If you make any non-syntax mistakes, SPSS will display one or more error messages in the output file. **Do not print your results if this happens.** Try to work out what caused the error (usually the first one reported causes all or most of the others). Close the output file without saving it, go back and edit your syntax file and run it again. A new output file will be generated with an incremental number (ie if the first output file was called **output1.spo**, the new one will be **output2.spo** and so on).]*

The SPSS syntax commands for **title** and **data list** are repeated and SPSS tells you what variables it is about to read and where it expects to find them. The output file will repeat your syntax and will also display the way it has been interpreted. The output viewer will display:

```
title 'Data for interval variables (BSA 1986)' .
```

```
>> Data for interval variables (BSA 1986)
```

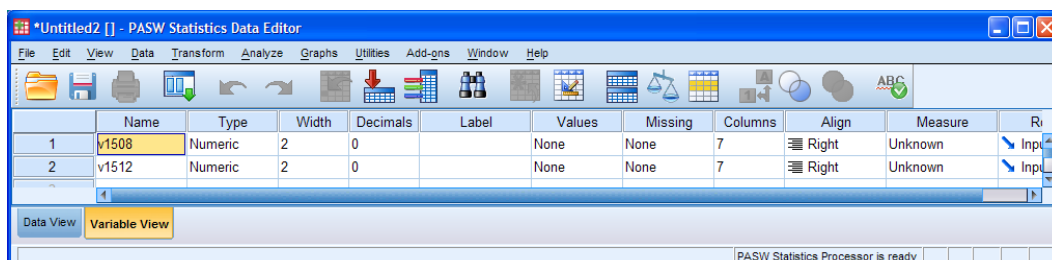
```
data list file 'C:\Documents and Settings\Owner\Desktop\mybsa\bsa86.txt'
records 23
/15 v1508 8-9 v1512 12-13 .
```

```
Data List will read 23 records from C:\Documents and
Settings\Owner\Desktop\mybsa\mybsa86\bsa86.txt
```

Variable	Rec	Start	End	Format
v1508	15	8	9	F2.0
v1512	15	12	13	F2.0

This is a useful check, especially when using the **positional** naming convention<sup>4</sup> as the variable names should tally exactly with the record number and start column as in the display above.

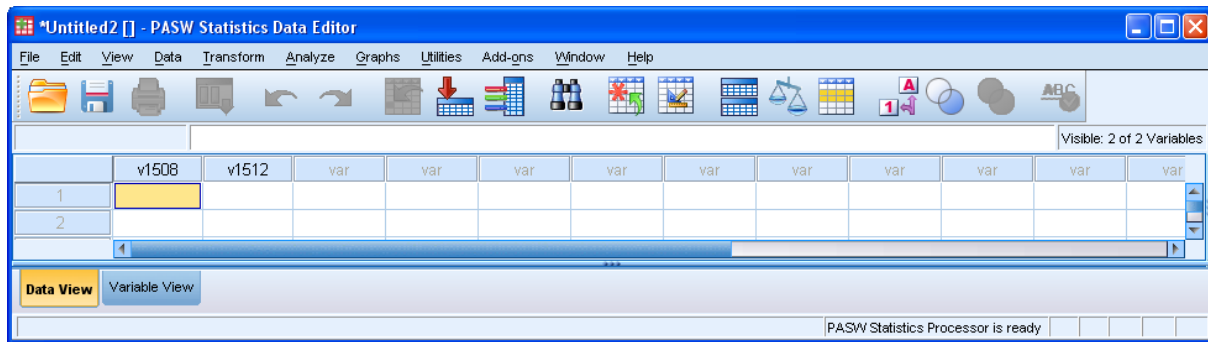
If you bring up the SPSS Data Editor again you will see the name has changed to **\*Untitled2**



Your own variable names are displayed in the **Name** column together with, in the other columns, various technical information about them. *[NB: **Measure** is "Unknown" because the data have not yet been read in.]*

<sup>4</sup> See 1.3.1 [Conventions for Naming Variables in SPSS](#)

Click on **Data View** :

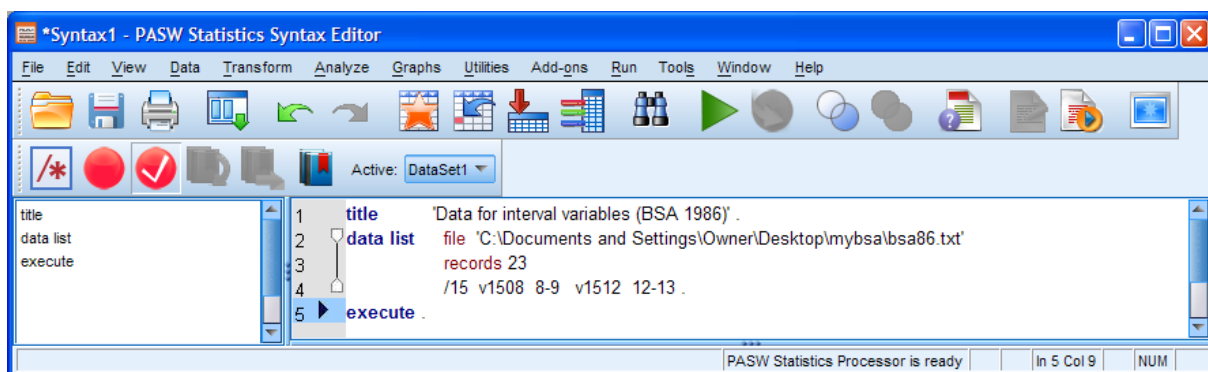


The names of the two variables you are about to create ( **v1508** and **v1512** ) appear in the headers of the first two columns, but there are no actual data. **Do not panic!** SPSS hasn't yet read the raw data. It only actually makes a pass through the data when asked to perform some sort of analysis or when given a special **EXECUTE** command.

## EXECUTE

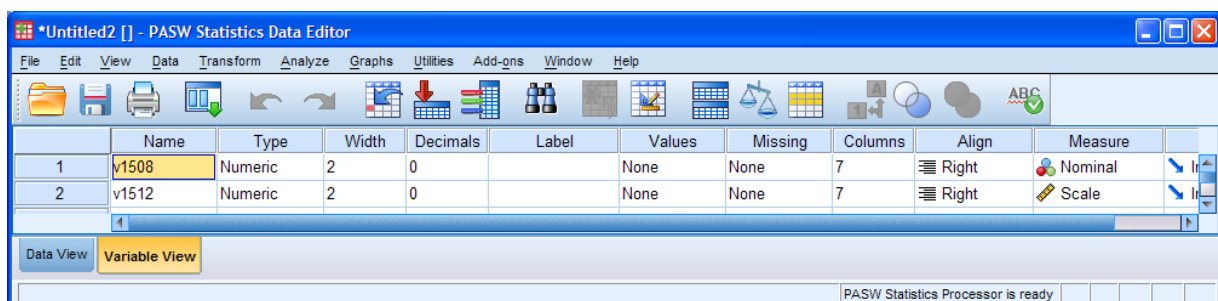
The **EXECUTE** command forces SPSS to read the data without waiting for a statistical command. Go back to your syntax editor and type:

**execute .**



Click on the green ► to run the command:

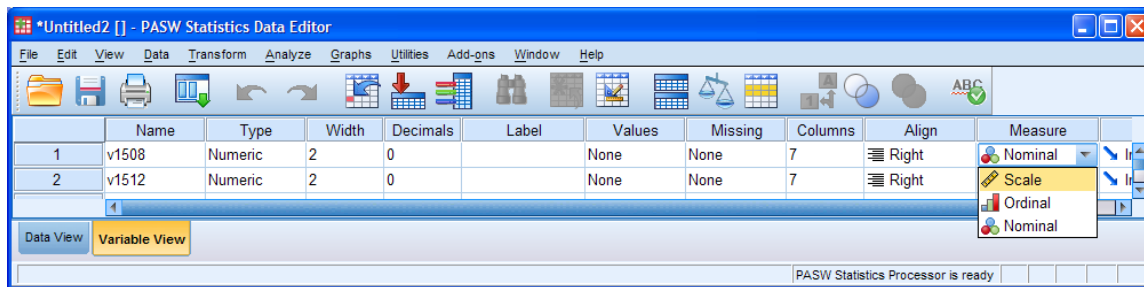
In the data editor the previous "Unknown" levels of measurement in the **Measure** column will change to "Nominal" for **v1508** and to "Scale" for **v1512**.



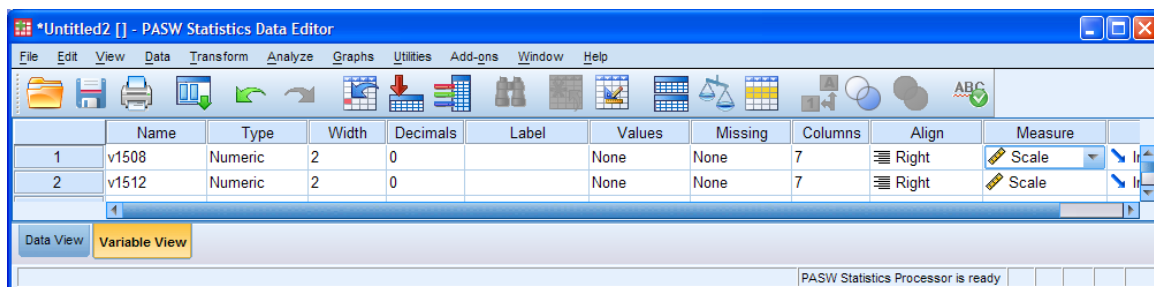
But, you may ask, the number of persons in the household is an interval variable, so why isn't it displayed as "Scale"? It's because SPSS has default settings for the number of unique values found (in this case 7) before it declares them as "Scale", but **v1508** has fewer values than the default so it is left as nominal.



You can change the default settings for scale variables in SPSS from inside the data editor with **Edit > Options > Data** , but it's easier for now just to click on the cell:

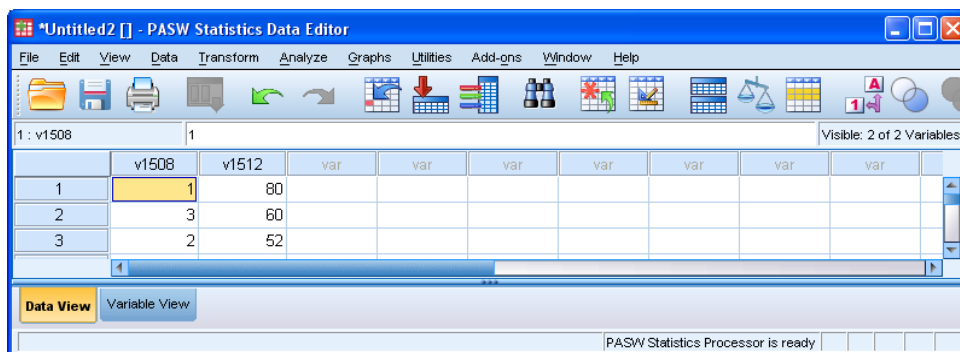


and change it by clicking on **Scale**

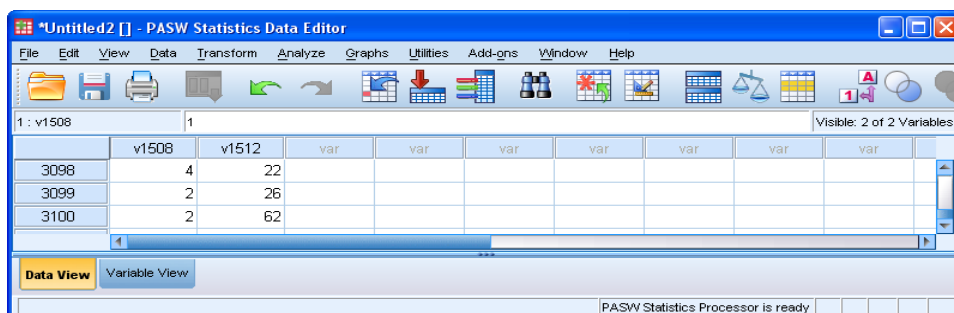


## Checking your data

Before proceeding to extend your data dictionary or do any statistical analysis, it is good practice to check the actual data to see if they look like what you'd expect. Switch the data editor to **Data View** and you will see that the first two columns have now been filled with data.



If you scroll down the whole file (or more quickly press **[CTRL]+[end]** to go to the end of the file) you will see that there are 3,100 lines of data ie 3,100 cases.

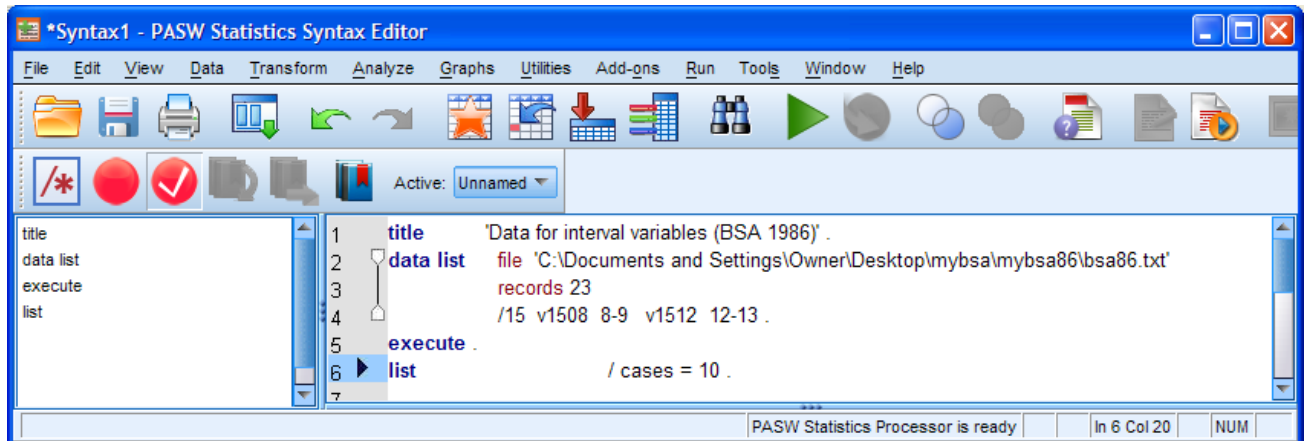


## LIST

As a data check you can also use the **LIST** command *[not available as a drop-down menu]*.

It has to be limited to a few cases, otherwise you will get a listing of all 3,100 lines of data. That's a lot of trees if you print a hard copy!

**list**                      **/ cases = 10 .**

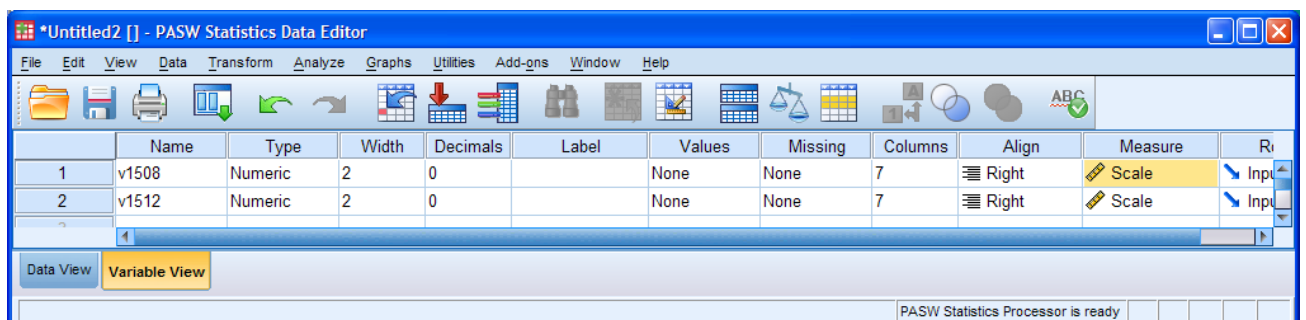


...and run the command by clicking on the green ► icon. The output file will now display:

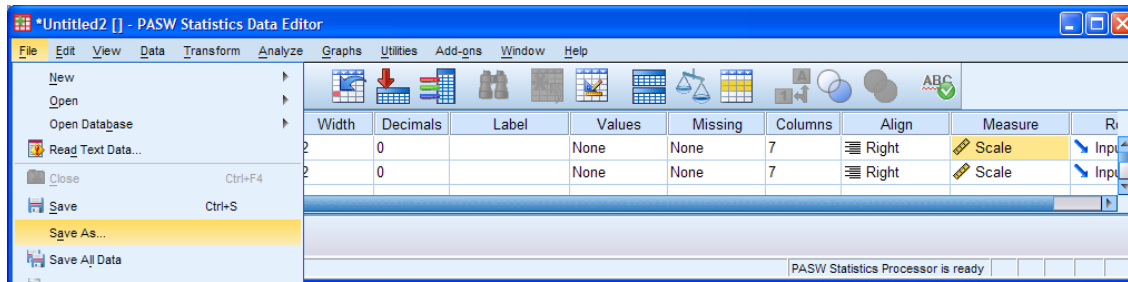
```
v1508 v1512
  1    80
  3    60
  2    52
  5    31
  3    20
  5    43
  2    26
  3    37
  4    48
  2    49

Number of cases read:  10
Number of cases listed: 10
```

Whilst not strictly necessary for this exercise, it's also good practice to get into the habit of regularly saving your work as you complete each stage of file-building. Go back to the data editor and switch to **Variable View**

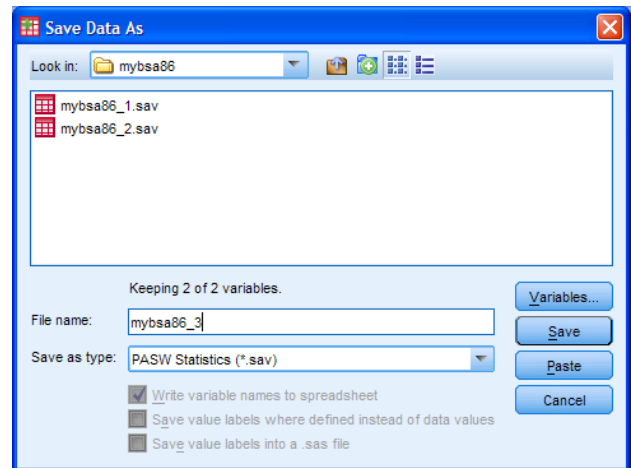
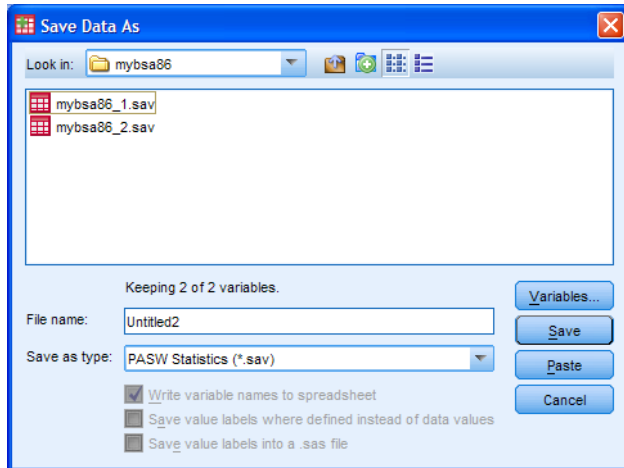


Click on **File** > **Save As ..**



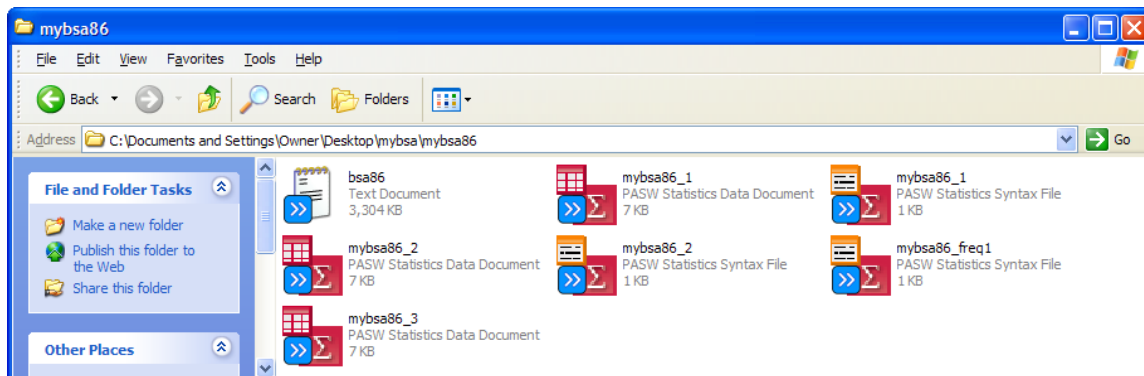
Navigate to folder **mybsa86**

... and write **mybsa86\_3** in the **File name:** box

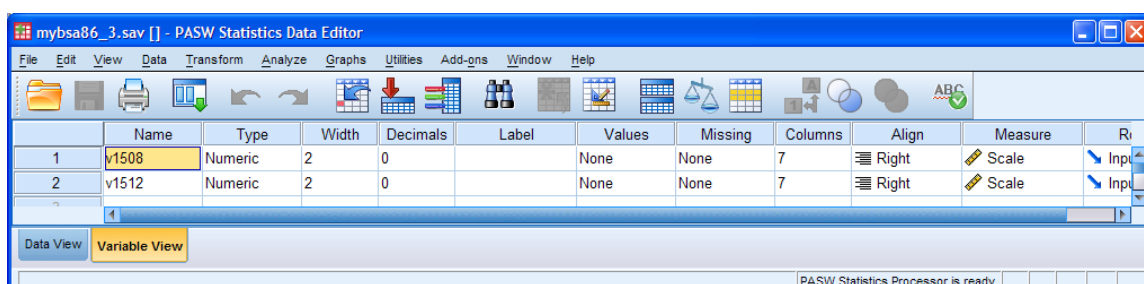


Make sure the **Save as type:** box displays **PASW Statistics (\*.sav)** and click on **Save**

A copy of the data editor will be saved as **mybsa86\_3.sav** in folder **mybsa86**.



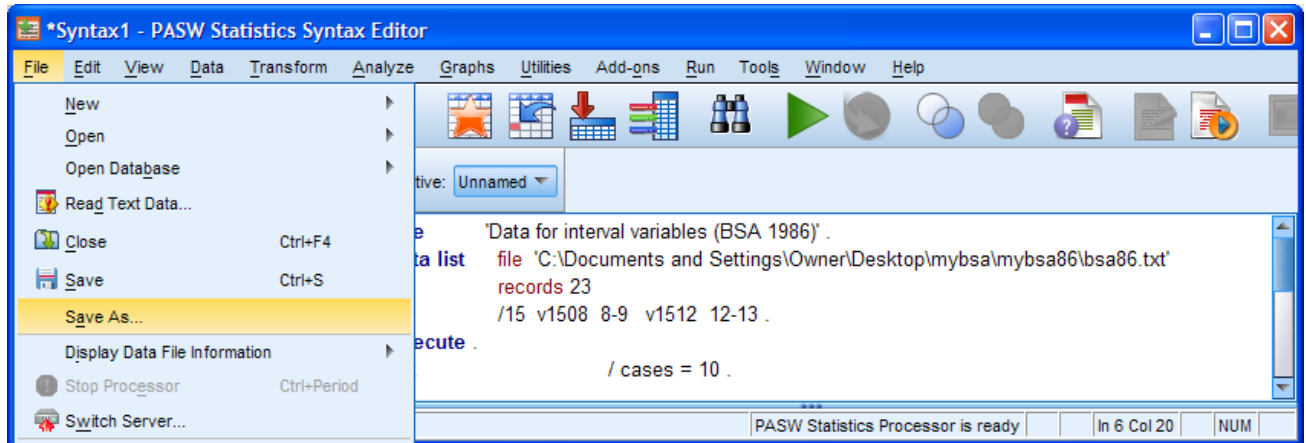
The data editor remains open as the **active file** in SPSS, but the header will change from **\*Untitled2.sav** to **bsa86\_3.sav**.



The output file will display the following syntax generated automatically by SPSS:

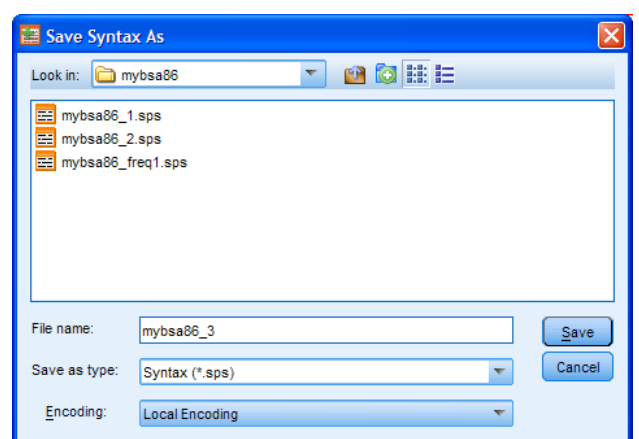
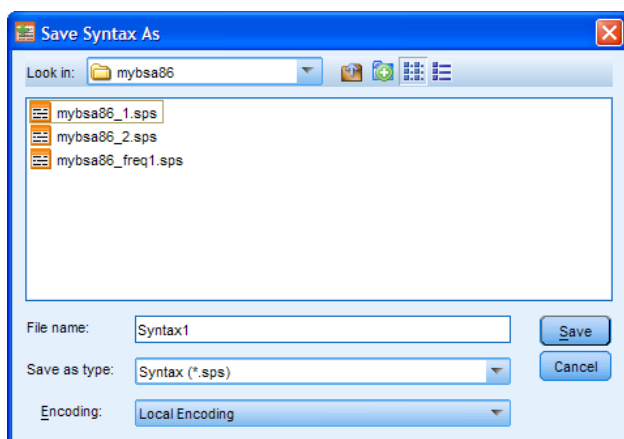
```
SAVE OUTFILE='C:\Documents and Settings\Owner\Desktop\mybsa\mybsa86\mybsa86_3.sav'  
/COMPRESSED.
```

You should also get in to the habit of saving your file-building syntax. Good practice is to have the same name as the **\*.sav** file it creates, but with a **\*.sps** extension, in this case **bsa86\_3.sps**. Go back to the syntax editor and click on **File** > **Save As** ..



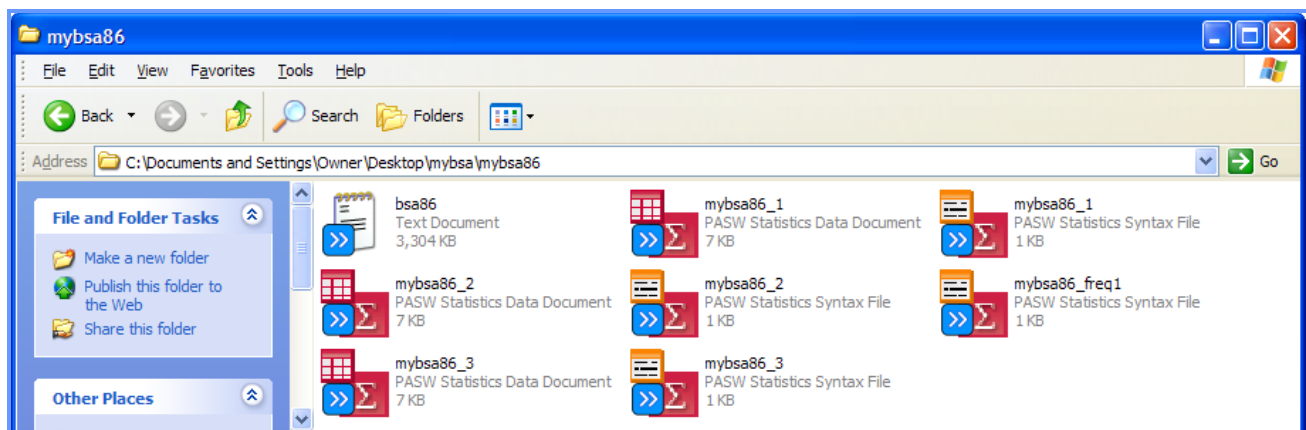
Navigate to folder **mybsa86**

... and write **mybsa86\_3** in the **File name:** box:

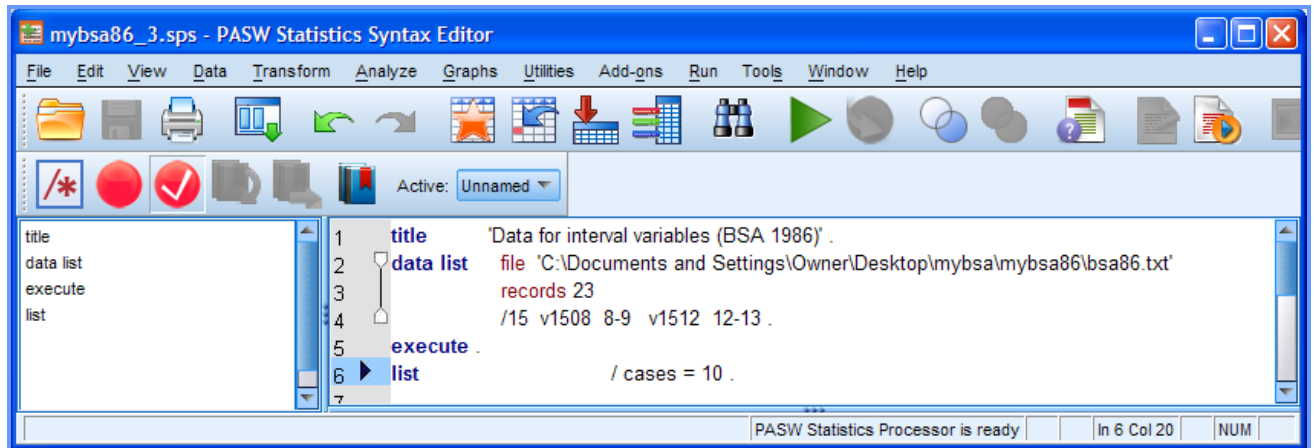


Make sure the **Save as type:** box displays **Syntax (\*.sps)** and click on **Save**

Your **mybsa86** folder should now look like this:



The syntax editor header will have changed from **\*Syntax1.sps** to **mybsa86\_3.sps**.



The first editions of your SPSS data and syntax files are now safely stored for you to proceed to the next stage. This will be to extend your data dictionary by adding **MISSING VALUES**, **VARIABLE LABELS** and **VALUE LABELS**.

You can save the output file as well if you like, but you don't really need to as you can always run the job again.

**End of session**

**Next session: 2.2.1.3 [BSA86] Extending your data dictionary**

[\[Back to Block 2 menu\]](#)