Survey Analysis Workshop

Block 3: Analysing two variables (and sometimes three)

Section 3.2: Three (or more) variables   **Sub-section 3.2.1 Elaboration**

© Copyright 2019 John F Hall   [New tutorial 12 May 2019: **Draft only**]

3.2.1.6 Earnings differences 2009: Extracting and saving selected variables

Data source: **British Social Attitudes Survey, 2009[1]**   **(UKDS SN 6695)**

Previous session: 3.2.1.5 Earnings differences 2009: Download and check file

---

[1] National Centre for Social Research. (2011). *British Social Attitudes Survey, 2009.* [data collection]. UK Data Service. SN: 6695, http://doi.org/10.5255/UKDA-SN-6695-1

Previous research questions:

1: Is there a difference between the earnings (from paid work) of men and women?
See sessions: 2.3.1.6.2: Specimen answer for tasks 3 and 4
3.1.4.1 Income differences work-through

2: What other variables might account for differences in earnings?
See sessions: 3.1.4.2 Income differences - Build working file
3.1.4.3   Income differences for test variables
3.1.4.4 Income differences - Choose test variables and cutting points

3: What effect do they have by themselves?
See session: 3.1.4.5 Income differences for derived test variables

Current research question:

What happens to any differences in earnings between men and women when controlling for these other variables?

Elaboration [2] model

$X \rightarrow Y \cdot T$ (the effect of **X** on **Y** controlling for **T**) where:

Y = Dependent variable
X = Independent variable
T = Test variable(s)

Y (Dependent)	X (Independent)	T_n (Test or control)	
Gross earnings from paid work	Sex	T_1	Working full time or part time
		T_2	Employee or self employed
		T_3	Economic sector
		T_4	Socio-economic grade of work
		T_5	Years of full-time education
		T_6	Qualifications
		T_7	Age
		T_8	Geographical region

This session demonstrates how to:

a) extract the variables needed to test the above elaboration model

b) save them in a much smaller file.

---

[2] (See Elaboration (extract from Jim Ring's Statistical Notes specially written for this course)

[3.2.1.6 Earnings differences 2009] Extracting and saving selected variables

Variables to be extracted

Dependent variable

[REarn] [if working] R's own gross or total earnings,before income tax+national insurance

Valid values range from **1** to **20** denoting grouped earnings per calendar month.
Value **-1** "Item not applicable" is declared as missing. Values **97** "Refused information",
98 "Don't know" and **99** "Refused" are **not declared as missing**.

[REarnQ] [if working] Quartile groups pf R's gross earnings

Derived variable in which **[REarn]** has been regrouped into four categories: this helps to keep
contingency tables small and manageable. Values **7** " Refused information" and **8** " Don't
know" are **not declared as missing**.

Independent variable   **[Rsex]** "Sex of respondent"

[Rsex] " SEX OF respondent? :Q356" **[sic]**

[Rsex] is coded **1** "Male" **2** "Female" and has no missing values. Users may prefer to rename
it as **[sex]** or **[gender]** according to their preferences.

Weighting

Some analyses may also require the weighting factor **[Wtfactor]**   " Final BSA weights"

Test variables:

(Demographics)

[Rage]   Age of respondent last birthday

There are two existing groupings for age, but it may be preferable to create new groupings

[Ragecat]   Age of respondent (grouped into 7 categories"
[RAgeCat2]   Age of respondent (grouped into 6 categories

(Work-related)

If the respondent is working, several work-related variables are available:

[Remploye]   " Is R an employee or self-employed currently? "
[ROcsect2]   " Resp SEG <grouped> [pre-SOC2000]"
[RNSocCl]   " Respondent : social class [pre-SOC2000] "

There is no single variable for working part-time (Under 30 hours a week) and working full- time (30 or
more hours a week). There are two separate derived variables, one for employees and another for the
self-employed:

[EjbHrCal] "Hours R works per week, including overtime [employee].
[SJbHrCal] "Hours R works per week, including overtime [self-employed].

These variables[3] are mutually exclusive, so will have to be combined into a single variable**.**

---

[3] **[EjbHrCal]** and **[SJbHrCal]** are spelled **[ejbhrcai]** and **[sjbhrcai]**

(Education)

> **[Tea]** Age of R when completed continuous full-time education
> **[HEdqual2]** Highest educational qualification obtained (postgrad separate)

(Geography)

> **[GOR2]** Government Office Region (2003 version)
> **[Country]** Country of interview (England, Scotland or Wales)

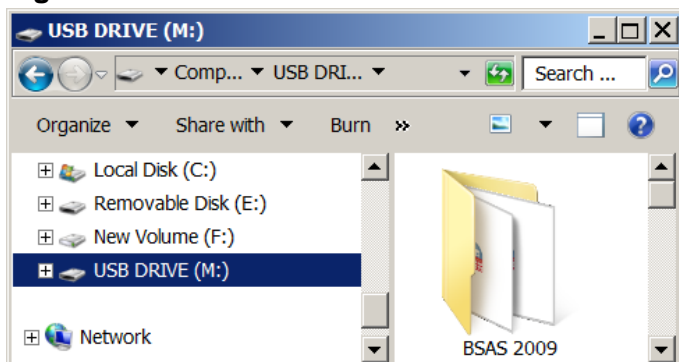In session 3.2.1.5 Earnings differences 2009: Download and check file we created a new folder

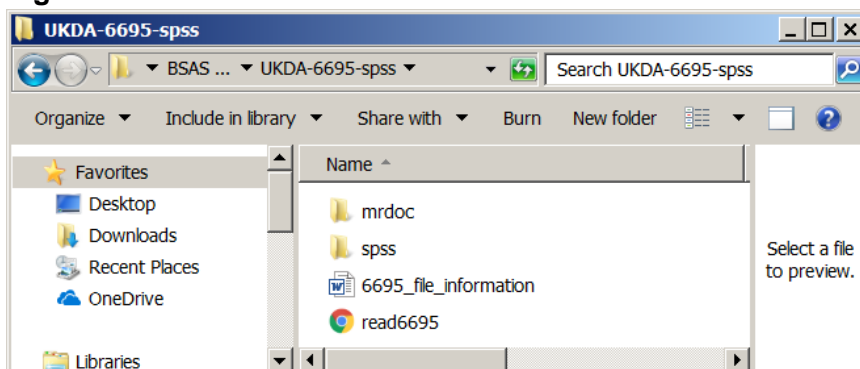📁 BSAS 2009 on USB Drive **M:**

Fig. 1:

. . copied zip file **6695spss_fa3ff1f37a5f7dd3c4ff6d62b3923ac4** (downloaded from UKDS) from

Downloads to folder 📁 BSAS 2009 and extracted the contents.

Fig. 2:

Name ▲	Date modified	Type
📁 UKDA-6695-spss	24/07/2018 00:50	File Folder
📄 6695spss_fa3ff1f37a5f7dd3c4ff6d62b3923ac4.zip	09/10/2018 09:47	7-Zip.zip

Double click on 📁 UKDA-6695-spss
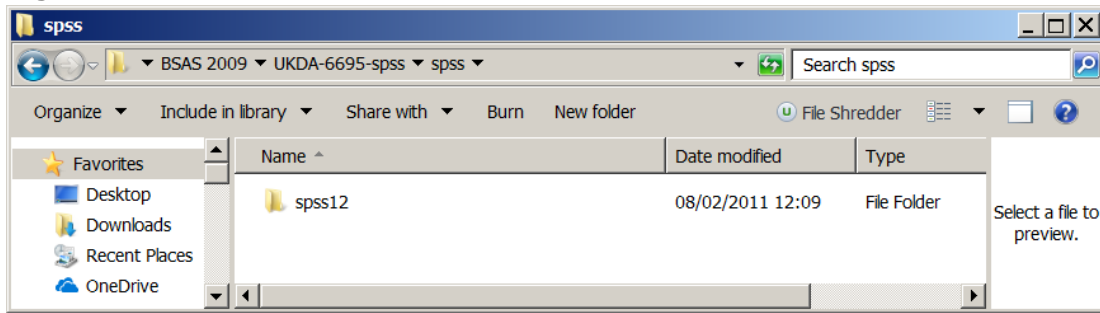
Fig. 3:

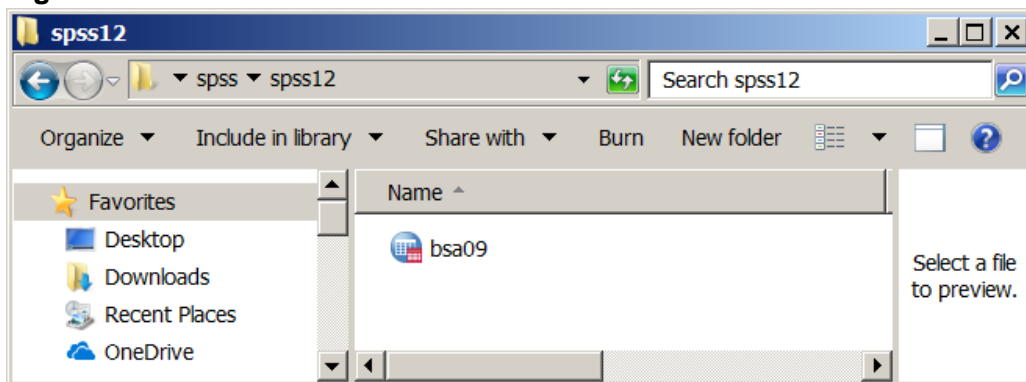
Double click 📁 spss

4

[3.2.1.6 Earnings differences 2009] Extracting and saving selected variables

Fig. 4:

Double click 📁 spss12

Fig. 5:

Double click 📊 bsa09

File 📊 bsa09 is the active file.

Fig. 6:
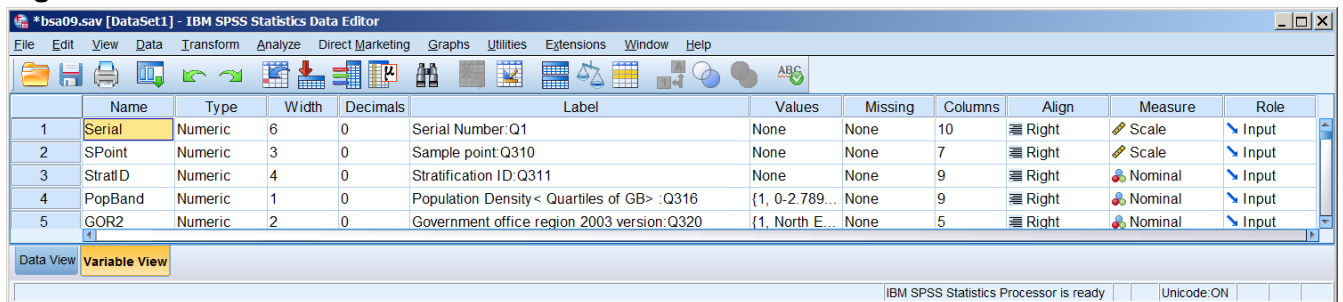
The file contains 847 variables, far too many for this exercise.

The SPSS icon in the task bar will show [4]

Fig. 7:
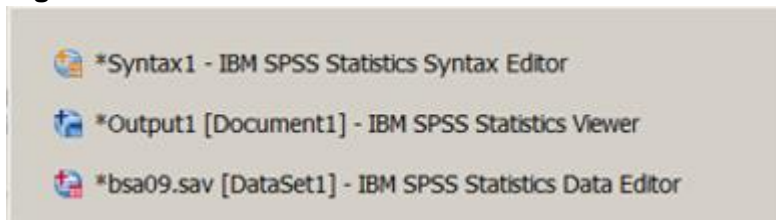
Intermediate stage 1: Making a copy of the file

Never work on an original file! Make a copy of the file

If your task bar does not show:

Fig. 8:
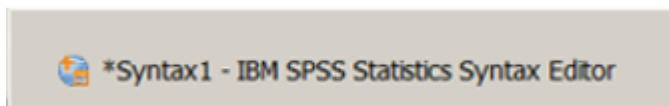
. . you must open a new **Syntax Editor** with:

File >> New >> Syntax

Fig. 9:
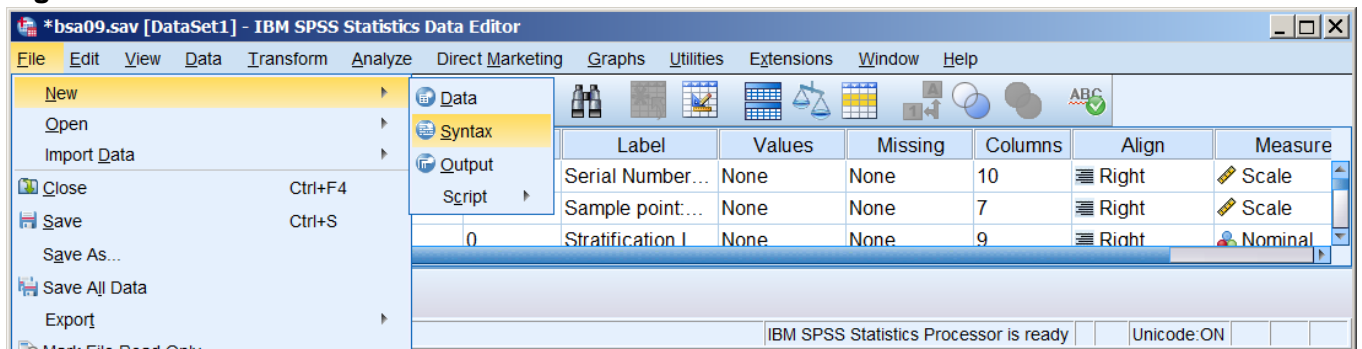
Fig. 10: Blank Syntax Editor

---
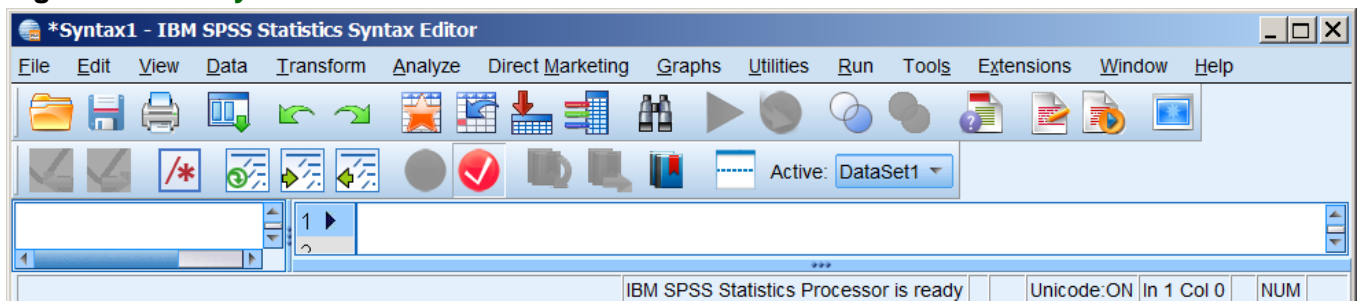
[4] The author has set SPSS to open a new **Syntax Editor** at startup.

6

In the **Syntax Editor** write:

title 'test1.sps'.   [Good practice is to give each *.sps file a **title**]
*file bsa09.sav open'.   [**Comment** (ignored by SPSS) to remind you what you are doing]

Fig. 11:
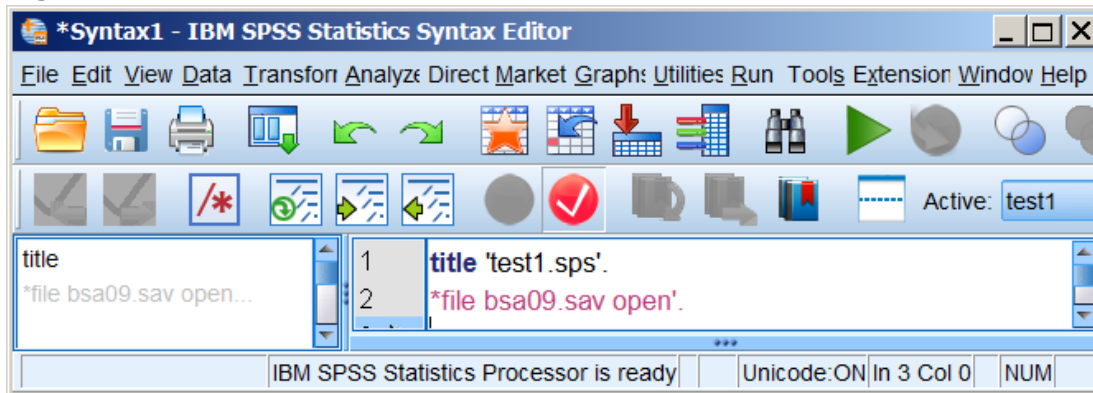
Now write:

dataset copy test1.   [Creates a copy of the fil]
dataset activate test1.   [Opens it as ***Untitled2 [test1]** which becomes the active file]

Fig. 12:
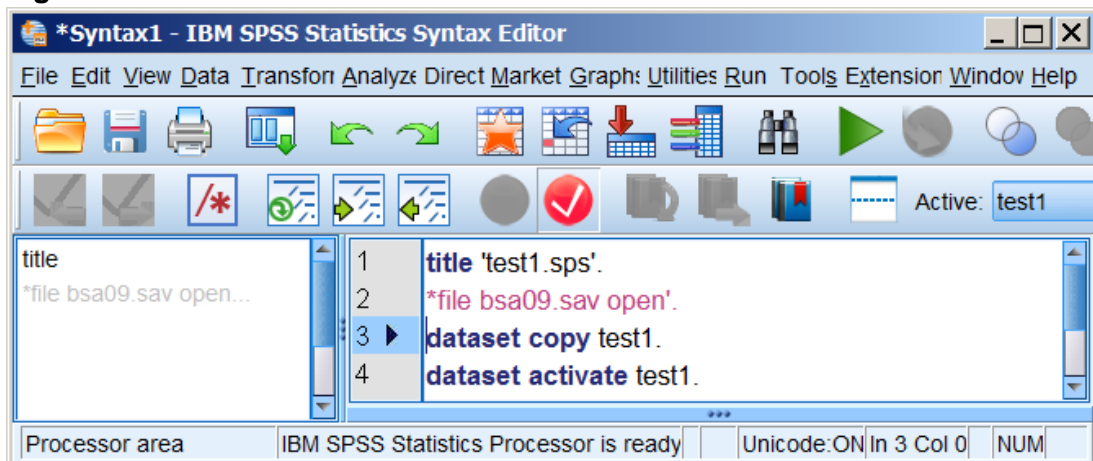
Either: Run >> All

Fig. 13:
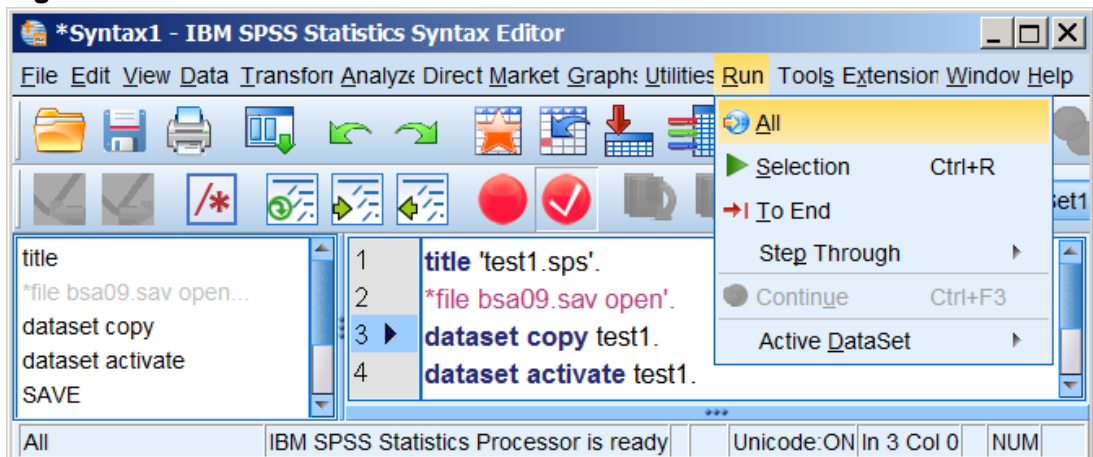
7

Or: Highlight all 4 lines:
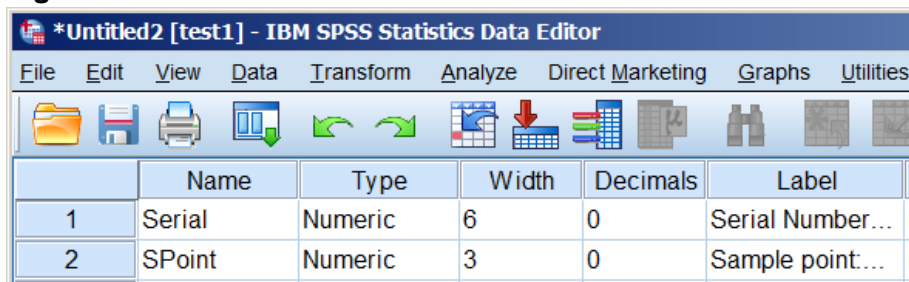
Fig. 14:
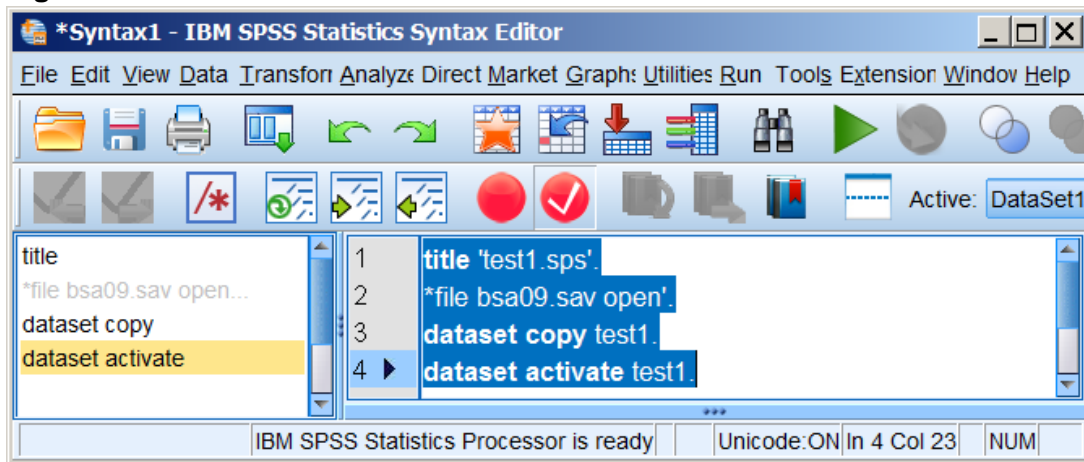
Press the green arrow: to make a copy of the file and open it.

The copy opens as **Untitled2 [test1]** and becomes the active file.

Fig. 15: New **Data Editor**

The task bar displays:

Fig. 16: SPSS icon in task bar
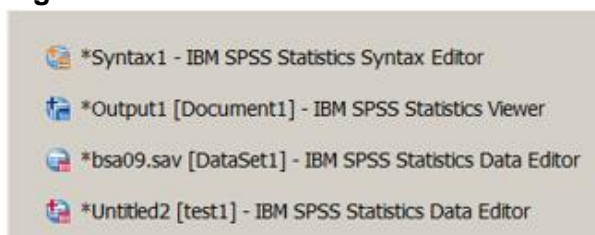
Without making any changes, save the file as **test1.sav**.

In the **Syntax Editor** add a line**:**

SAVE OUTFILE = 'M:\BSAS 2009**test1.sav**' .

Fig. 16:
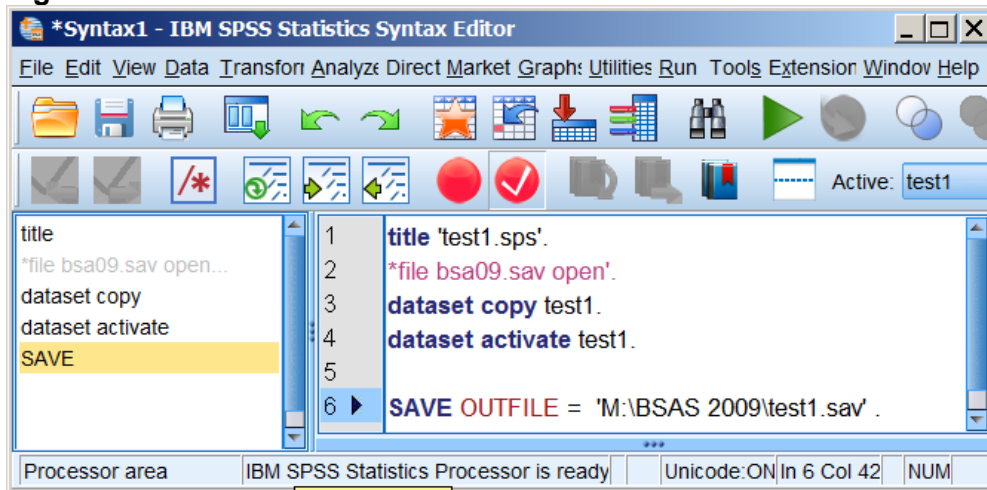
Make sure the cursor is on the same line: press the green arrow:

test1 is saved to folder BSAS 2009

Fig. 17:
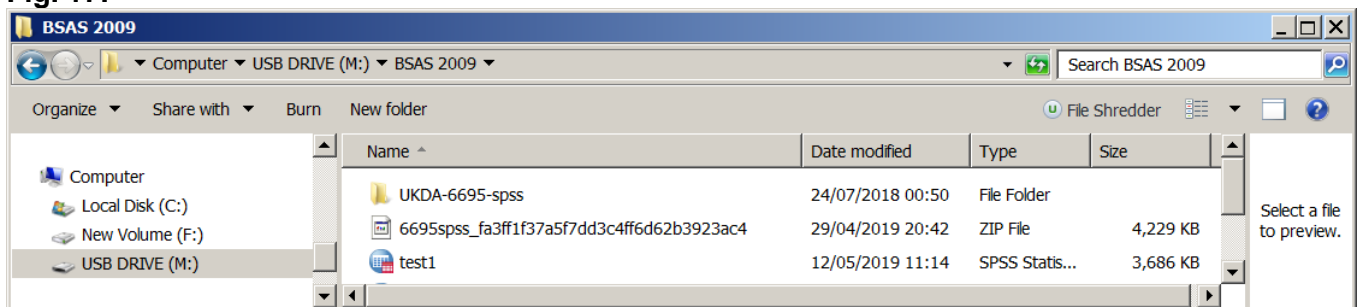
Name	Date modified	Type	Size
UKDA-6695-spss	24/07/2018 00:50	File Folder	
6695spss_fa3ff1f37a5f7dd3c4ff6d62b3923ac4	29/04/2019 20:42	ZIP File	4,229 KB
test1	12/05/2019 11:14	SPSS Statis...	3,686 KB

Locating the selected variables

The variables needed are:

Dependent: [REarn] [REarnQ]

Independent: [Rsex]

Test: Age: [Rage] [Ragecat] [RAgeCat2]
         Work: [Remploye] [EjbHrCal] [SJbHrCal] [rocsect2] [rnsoccl]
         Education: [Tea] [hedqual2]
         Geography: [GOR2] [country]
Other: [wtfactor]

How do we extract them and save them to a smaller file?

9

[3.2.1.6 Earnings differences 2009] Extracting and saving selected variables

Double click on 🖫 test1

Fig. 18:
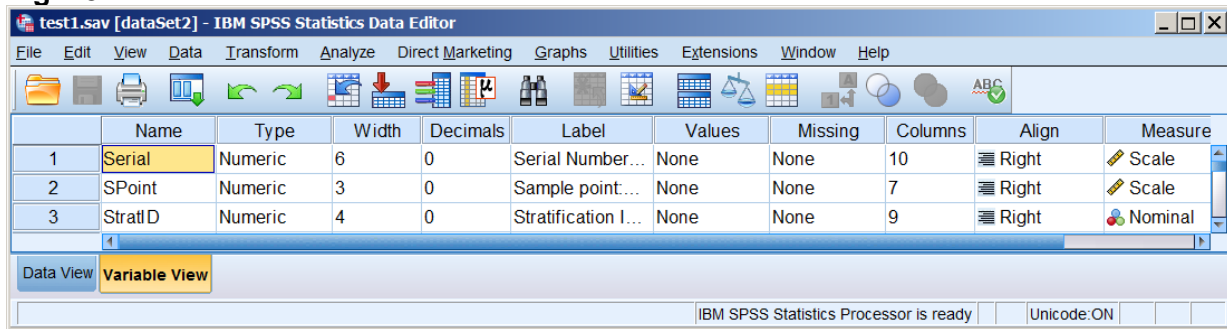
Intermediate stage 2: Extracting the subset of variables

To extract only the 17 variables needed for this exercise and save them in a new file **test2.sav**

Open a new **Syntax Editor**

Fig. 19:
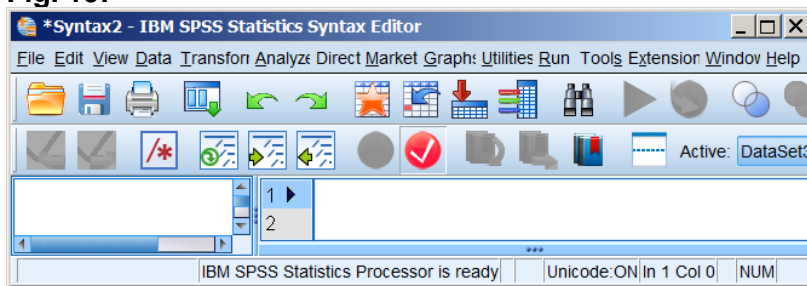
. . and write the following:

```
title 'test2: Extract variables'.
SAVE OUTFILE =
'M:\BSAS 2009\test2.sav'
   /KEEP serial GOR2 WtFactor country
        REarn REarnQ
        RSex Rage RAgeCat RAgeCat2
        REmploye EjbHrCaI  SJbHrCaI RNSEGGrp RNSocCl
        Tea HEdQual2
   /COMPRESSED.
```

Fig. 20:
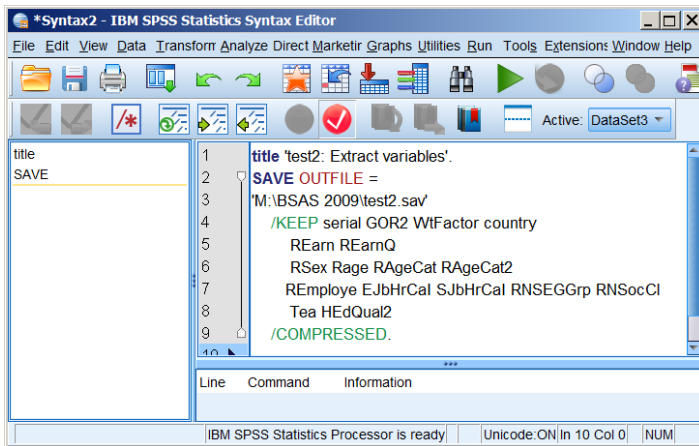
Make sure the cursor is somewhere inside the block of syntax and press the green arrow

File **test2** will be saved to folder **BSAS 2009**

Fig. 21:

Fig. 22:

Using **/KEEP** has reduced the file size from 3,656 kb to 120 kb.

Double click on **test2**
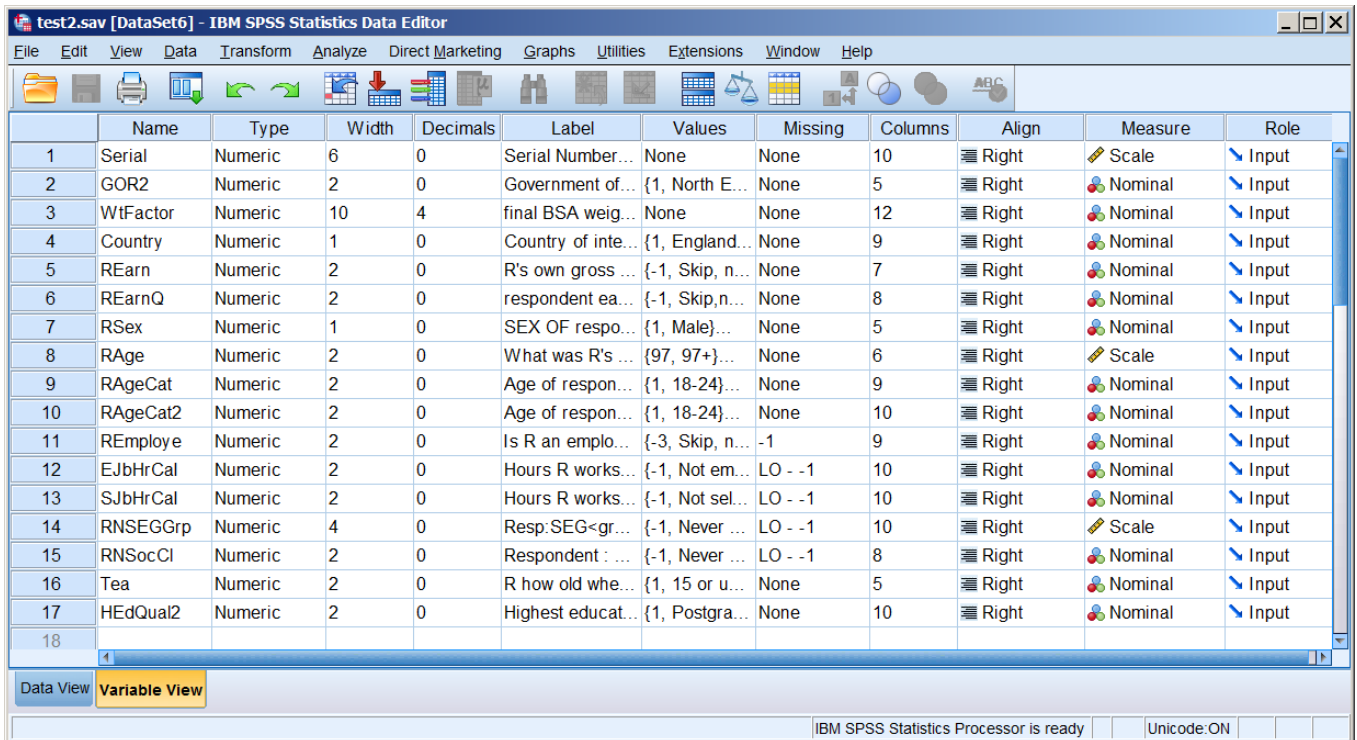
File **test2** is now the active file.

Fig. 23:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Serial	Numeric	6	0	Serial Number...	None	None	10	Right	Scale	Input
2	GOR2	Numeric	2	0	Government of...	{1, North E...	None	5	Right	Nominal	Input
3	WtFactor	Numeric	10	4	final BSA weig...	None	None	12	Right	Nominal	Input
4	Country	Numeric	1	0	Country of inte...	{1, England...	None	9	Right	Nominal	Input
5	REarn	Numeric	2	0	R's own gross ...	{-1, Skip, n...	None	7	Right	Nominal	Input
6	REarnQ	Numeric	2	0	respondent ea...	{-1, Skip,n...	None	8	Right	Nominal	Input
7	RSex	Numeric	1	0	SEX OF respo...	{1, Male}...	None	5	Right	Nominal	Input
8	RAge	Numeric	2	0	What was R's ...	{97, 97+}...	None	6	Right	Scale	Input
9	RAgeCat	Numeric	2	0	Age of respon...	{1, 18-24}...	None	9	Right	Nominal	Input
10	RAgeCat2	Numeric	2	0	Age of respon...	{1, 18-24}...	None	10	Right	Nominal	Input
11	REmploye	Numeric	2	0	Is R an emplo...	{-3, Skip, n...	-1	9	Right	Nominal	Input
12	EJbHrCal	Numeric	2	0	Hours R works...	{-1, Not em...	LO - -1	10	Right	Nominal	Input
13	SJbHrCal	Numeric	2	0	Hours R works...	{-1, Not sel...	LO - -1	10	Right	Nominal	Input
14	RNSEGGrp	Numeric	4	0	Resp:SEG<gr...	{-1, Never ...	LO - -1	10	Right	Scale	Input
15	RNSocCl	Numeric	2	0	Respondent : ...	{-1, Never ...	LO - -1	8	Right	Nominal	Input
16	Tea	Numeric	2	0	R how old whe...	{1, 15 or u...	None	5	Right	Nominal	Input
17	HEdQual2	Numeric	2	0	Highest educat...	{1, Postgra...	None	10	Right	Nominal	Input
18											

Using **/KEEP** has reduced the number of variables from 847 to 17

Checking the data file

Open a new **Syntax Editor**

Fig. 24:
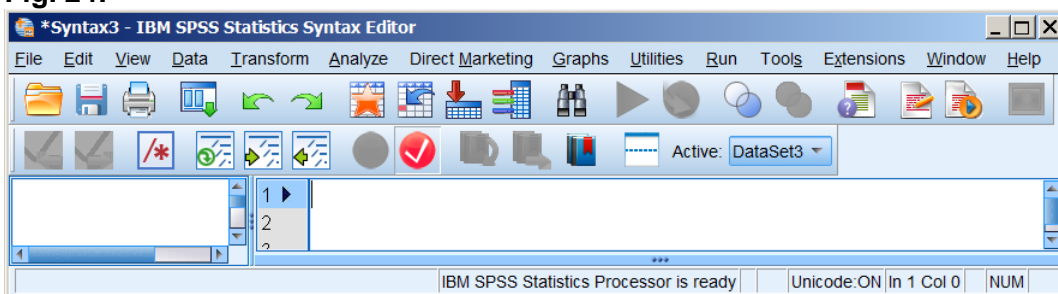
Write in:

> **title** 'test3: Check data file'**.**
> **display** dictionary.
> **display** labels.

Fig. 25:
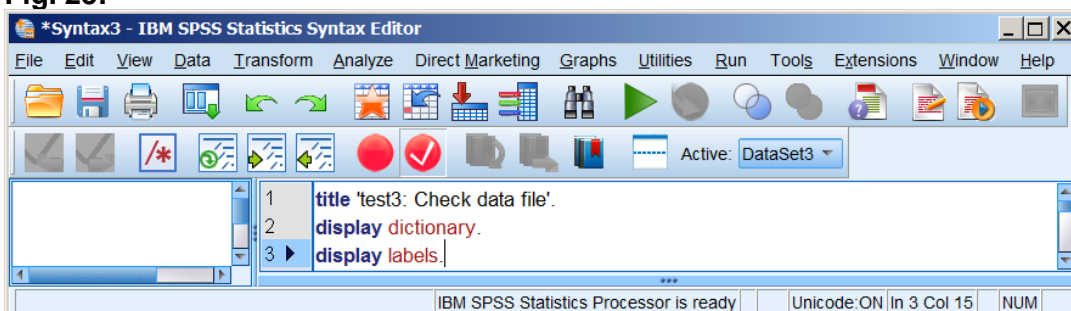
12

Run >> All

Fig. 26:
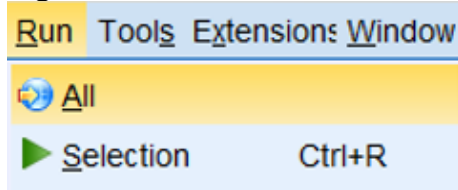
Table 1: Data dictionary for selected variables (condensed)

Variable	Line	Label	Level	Format	Missing
Serial	1	Serial Number:Q1	Scale	F6	
GOR2	2	Government office region 2003 version:Q320	Nominal	F2	
WtFactor	3	final BSA weights	Nominal	F10.4	
Country	4	Country of interview England, Scotland or Wales? :Q333	Nominal	F1	
REarn	5	R's own gross or total earnings,before income tax+national insurance?:Q1376	Nominal	F2	
REarnQ	6	respondent earnings quartiles (dv):Q1377	Nominal	F2	
RSex	7	SEX OF respondent? :Q356	Nominal	F1	
RAge	8	What was R's age last birthday? :Q357	Scale	F2	
RAgeCat	9	Age of respondent(grouped)<7 category> dv:Q446	Nominal	F2	
RAgeCat2	10	Age of respondent(grouped)<6 category> dv:Q447	Nominal	F2	
REmploye	11	Is R an employee or self-employed currently? dv:Q987	Nominal	F2	-1
EJbHrCal	12	Hours R works per week, including overtime [employee]. DV:Q1008	Nominal	F2	Lo thru -1
SJbHrCal	13	Hours R works per week, including overtime [self-employed]. DV:Q1010	Nominal	F2	Lo thru -1
RNSEGGrp	14	Resp:SEG<grouped>[pre-SOC2000]best est.Q	Scale	F4	Lo thru -1
RNSocCl	15	Respondent : social class[pre-SOC2000]best estimate dv:Q1022	Nominal	F2	Lo thru -1
Tea	16	R how old when completed continuous fulltime education?[compressed]dv:Q1196	Nominal	F2	
HEdQual2	17	Highest educational qual obtained (postgrad separate) - dv:Q1262	Nominal	F2	

[**NB:** The above table has been heavily edited to delete columns **Role**, **Width**, **Align** and to combine **Print Format** and **Write Format** in a single column **Format**.]

Table 2: Variable labels

Variable Labels

Variable	Position	Label
Serial	1	Serial Number:Q1
GOR2	2	Government office region 2003 version:Q320
WtFactor	3	final BSA weights
Country	4	Country of interview England, Scotland or Wales? :Q333
REarn	5	R's own gross or total earnings,before income tax+national insurance?:Q1376
REarnQ	6	respondent earnings quartiles (dv):Q1377
RSex	7	SEX OF respondent? :Q356
RAge	8	What was R's age last birthday? :Q357
RAgeCat	9	Age of respondent(grouped)<7 category> dv:Q446
RAgeCat2	10	Age of respondent(grouped)<6 category> dv:Q447
REmploye	11	Is R an employee or self-employed currently? dv:Q987
EJbHrCal	12	Hours R works per week, including overtime [employee]. DV:Q1008
SJbHrCal	13	Hours R works per week, including overtime [self-employed]. DV:Q1010
RNSEGGrp	14	Resp:SEG<grouped>[pre-SOC2000]best est.Q*
RNSocCl	15	Respondent : social class[pre-SOC2000]best estimate dv:Q1022
Tea	16	R how old when completed continuous fulltime education?[compressed]dv:Q1196
HEdQual2	17	Highest educational qual obtained (postgrad separate) - dv:Q1262

Variables in the working file

The **question numbers** at the end of each label are useful for checking against the original questionnaire, but are effectively superfluous and could be deleted: so could **[compressed]** and **dv:** (derived variable).

* There is no question number in the variable label for **[RNSEGGrp]**

Across all waves of the survey, the same question or item always has the same variable name, but the question numbers will be different in each year and consequently be incompatible between years. Similarly, variable types, formats, widths, labels and value labels also vary between years. When combining data for different years, all variable properties will need to be standardised.

Further file modifications

To show screenshots of the **Syntax Editor** on the following pages would be repetitive and tedious. From this point only the **commands** and **output** are shown.

Discarding cases with no earnings

We need to discard cases who have no earnings from paid work.

> **frequencies** rearnq.

Table 3: Respondent earnings quartiles

REarnQ respondent earnings quartiles (dv):Q1377

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	-1 Skip,not in paid work	1558	45.5	45.5	45.5
	1 less than 11999	395	11.5	11.5	57.1
	2 12000- 19999	414	12.1	12.1	69.2
	3 20000- 31999	467	13.7	13.7	82.8
	4 32000 or more	413	12.1	12.1	94.9
	7 Refused information	147	4.3	4.3	99.2
	8 Don't know	27	0.8	0.8	100.0
	Total	3421	100.0	100.0	

[**NB:** The **£** sign was not available in SPSS 12]

Values **-1** " Skip,not in paid work" **7** " Refused information" and **8** " Don't know" need to be treated as missing.

> **missing values** REarn (-1 97 thru 99) REarnQ (-1, 7, 8)
> **frequencies** rearnq.

Table 4: Respondent earnings quartiles

REarnQ Respondent earnings quartiles

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 less than 11999	395	11.5	23.4	23.4
	2 12000- 19999	414	12.1	24.5	47.9
	3 20000- 31999	467	13.7	27.6	75.5
	4 32000 or more	413	12.1	24.5	100.0
	Total	1689	49.4	100.0	
Missing	-1 Skip,not in paid work	1558	45.5		
	7 Refused information	147	4.3		
	8 Don't know	27	0.8		
	Total	1732	50.6		
Total		3421	100.0		

To select only those cases with non-missing values for **[REarnQ]:**

> **select if** (not (missing (rearnq))).
> **frequencies** rearnq.

Table 5: Respondent earnings quartiles (non-earners excluded)

REarnQ

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	less than 11999	395	23.4	23.4	23.4
	12000- 19999	414	24.5	24.5	47.9
	20000- 31999	467	27.6	27.6	75.5
	32000 or more	413	24.5	24.5	100.0
	Total	1689	100.0	100.0	

The number of cases in the file has been reduced from **3421** to **1689**.

There is no variable denoting the year of survey:

> **compute** year = 2009.
> **variable labels** year 'Year of survey'.

Variable **[year]** is appended to the file:

Fig. 27:

	Name	Type	Width	Decimals	Label
16	Tea	Numeric	2	0	R how old whe…
17	HEdQual2	Numeric	2	0	Highest educat…
18	year	Numeric	8	0	Year of survey

Tidy up the data dictionary

The data dictionary now needs editing to add or correct measurement levels, missing values and variable and value labels.

Levels of measurement:

> **variable level**
> RSEx GOR2 country (nominal)
> REarn REarnQ RAgeCat RAgeCat2
> REmploye EJbHrCaI SJbHrCaI RNSEGGrp to HEdQual2 (ordinal)
> year WtFactor (scale).

Formats

> **formats**
> year (f4.0) serial (f10.0)
> REarn GOR2 (f3.0)
> REarnQ RAgeCat RAgeCat2
> REmploye EJbHrCaI SJbHrCaI RNSEGGrp RNSocCl
> Tea HEdQual2 country (f2.0)
> wtfactor (f6.4) .

Variable labels

Variable labels can be edited manually in the **Data Editor**, but it's much safer to use syntax:

> **variable labels**
> year   'Year of survey'
> Serial   'Serial Number'
> GOR2   ' Government office region: 2003 version '
> WtFactor   ' Final BSA weights '
> Country   ' Country of interview: England, Scotland or Wales '
> REarn   " R's own gross earnings,before Income Tax and National Insurance "
> REarnQ   ' Respondent earnings quartiles '
> RSex   ' Sex of respondent '
> RAge   " What was R's age last birthday? "
> RAgeCat   ' Age of respondent (grouped: 7 categories) '
> RAgeCat2 ' Age of respondent (grouped: 6 categories) '
> REmploye ' Is R an employee or self-employed currently? '
> EjbHrCaI   'Hours R works per week, including overtime [employee].'
> SJbHrCaI   'Hours R works per week, including overtime [self-employed].'
> RNSEGGrp ' Resp:SEG <grouped> [pre-SOC2000] '
> RNSocCl   ' Respondent : social class [pre-SOC2000] '
> Tea   ' R how old when completed continuous full time education '
> HEdQual2 ' Highest educational qualification obtained '.

[**NB:** Double quotes are needed for **[REarn]** and **[Rage]** because the labels contain single quotes].

Missing values

No missing values are declared for **[REarn]** and **[REarnQ]**

Fig. 28:

618	REarn	Numeric	2	0	R's own gross ...	{-98, Don't ...	None
619	REarnQ	Numeric	2	0	respondent ea...	{-1, Skip,n...	None

These two variables have already been checked (see frequency tables on page 13 above and on pages 13 to 16 of 3.2.1.5 Earnings differences 2009: Download and check file)

Variable **[REmploye]** has two negative values declared as missing: **-3** " Skip, not now in paid work" and **-1** " Skip, never had a job".

It also has two other values which we need to treat as missing: **8** "Don't know" and **9** "Refusal".

SPSS only allows 3 missing values eg **(7, 8, 9)** two of which can be lower and upper limits of a value range eg **(-1, 97 thru 99)**. Where there are more than 3 values to be treated as missing, one method is to recode all positive missing values to negative and specify missing values as **(lo thru -1)**. If you do this you must use the **ADD VALUE LABELS** command, otherwise the existing labels will be deleted.

> **recode** REmploye (8 = -8)(9 = -9) .
> **add value labels** REmploye -8 " Don't know" -9 " Refusal".
>
> **missing values**
>   REarn (-1 97 thru 99) REarnQ (-1, 7, 8) RAgeCat (8) RAgeCat2 (9)
>   REmploye (-9 thru -1) EJbHrCaI SJbHrCaI (-1, 5 thru 9) RNSEGGrp RNSocCI (-1, 7, 8)
>   Tea (6 thru 99) HEdQual2 (7, 9).

Variable **[year]** needs to be at the beginning of the file.

To make files easier to navigate, experienced researchers often re-order the variables:

> **[Dependent >> Independent >> Test]**

. . and arrange the test variables in blocks defined by their source topic, in this case:

> **Age >> Work >> Education >> Geography**

Save the reduced file

> **SAVE** OUTFILE =
> 'M:\BSAS 2009\test4.sav'
>   /KEEP year serial
>   Rearngrp REarn REarnQ
>   RSex
>   Rage RAgeCat RAgeCat2
>   REmploye EjbHrCaI SJbHrCaI RNSEGGrp RNSocCI
>   Tea HEdQual2
>   GOR2 Country WtFactor
>   /COMPRESSED.

File 🗐 test3 will be saved to folder 📁 BSAS 2009 on USB Drive **M:**

[3.2.1.6 Earnings differences 2009] Extracting and saving selected variables

Fig. 28:

Fig. 29:

Checking the variable properties

Double click on [test3 icon] **test3** which becomes the active file [test3.sav [DataSet icon]

Fig. 30:

Correct measurement levels have now been assigned.

The variable labels are also clearer:

Fig. 31:

display labels.

Table 6: Variable labels

Variable Labels

Variable	Position	Label
year	1	Year of survey
Serial	2	Serial Number
REarn	3	R's own gross earnings (before Income Tax and National Insurance)
REarnQ	4	Respondent earnings quartiles
RSex	5	Sex of respondent
RAge	6	What was R's age last birthday?
RAgeCat	7	Age of respondent (grouped: 7 categories)
RAgeCat2	8	Age of respondent (grouped: 6 categories)
REmploye	9	Is R an employee or self-employed currently?
RNSEGGrp	10	Resp:SEG <grouped> [pre-SOC2000]
RNSocCl	11	Respondent : social class [pre-SOC2000]
Tea	12	R how old when completed continuous full time education
HEdQual2	13	Highest educational qualification obtained
GOR2	14	Government office region: 2003 version
Country	15	Country of interview: England, Scotland or Wales
WtFactor	16	Final BSA weights

Variables in the working file

Check variable properties

Check 1: Dependent variable

For the purposes of elaboration our dependent variable will be **[REarnQ]**

> **frequencies** rearnq.

Table 7:

REarnQ

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	less than 11999	395	23.4	23.4	23.4
	12000- 19999	414	24.5	24.5	47.9
	20000- 31999	467	27.6	27.6	75.5
	32000 or more	413	24.5	24.5	100.0
	Total	1689	100.0	100.0	

[**NB**: The **£** sign was not available in SPSS 12]

Check 2: Independent variable

> **frequencies** rsex.

Table 8:

RSex

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	831	49.2	49.2	49.2
	Female	858	50.8	50.8	100.0
	Total	1689	100.0	100.0	

20

Check 3: Earnings of men and women

crosstabs rsex by rearnq /cells count row.

Table 9:

RSex * REarnQ Crosstabulation

			REarnQ				Total
			less than 11999	12000-19999	20000-31999	32000 or more	
RSex	Male	Count	102	186	247	296	831
		% within RSex	12.3%	22.4%	29.7%	35.6%	100.0%
	Female	Count	293	228	220	117	858
		% within RSex	34.1%	26.6%	25.6%	13.6%	100.0%
Total		Count	395	414	467	413	1689
		% within RSex	23.4%	24.5%	27.6%	24.5%	100.0%

Check 4: Test variables

Person related

> **frequencies** RageCat RAgeCat2 .

Table 10:

RAgeCat Age of respondent (grouped: 7 categories)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 18-24	109	6.5	6.5	6.5
	2 25-34	357	21.1	21.1	27.6
	3 35-44	489	29.0	29.0	56.5
	4 45-54	425	25.2	25.2	81.7
	5 55-59	174	10.3	10.3	92.0
	6 60-64	94	5.6	5.6	97.6
	7 65+	41	2.4	2.4	100.0
	Total	1689	100.0	100.0	

Table11:

RAgeCat2 Age of respondent (grouped: 6categories)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 18-24	109	6.5	6.5	6.5
	2 25-34	357	21.1	21.1	27.6
	3 35-44	489	29.0	29.0	56.5
	4 45-54	425	25.2	25.2	81.7
	5 55-64	268	15.9	15.9	97.6
	6 65-97	41	2.4	2.4	100.0
	Total	1689	100.0	100.0	

We shall need to group **Rage** into fewer categories, but using different cutting points.

Work related

> **frequencies** REmploye EJbHrCai SJbHrCai RNSEGGrp RNSocCl .

Table 12:

REmploye Is R an employee or self-employed currently?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Emp	1467	86.9	86.9	86.9
	2 SEmp	222	13.1	13.1	100.0
	Total	1689	100.0	100.0	

Table 13:

EJbHrCal Hours R works per week, including overtime [employee].

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 10-15 hours a week	76	4.5	5.2	5.2
	2 16-23 hours a week	163	9.7	11.2	16.4
	3 24-29 hours a week	84	5.0	5.8	22.1
	4 30 or more hours a week	1137	67.3	77.9	100.0
	Total	1460	86.4	100.0	
Missing	-1 Not employee	222	13.1		
	5 Varies too much to say	5	0.3		
	8 Don't know	2	0.1		
	Total	229	13.6		
Total		1689	100.0		

Table 14:

SJbHrCal Hours R works per week, including overtime [self-employed].

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 10-15 hours a week	13	0.8	6.0	6.0
	2 16-23 hours a week	36	2.1	16.5	22.5
	3 24-29 hours a week	14	0.8	6.4	28.9
	4 30 or more hours a week	155	9.2	71.1	100.0
	Total	218	12.9	100.0	
Missing	-1 Not self-employed	1467	86.9		
	5 Varies too much to say	3	0.2		
	9 Refusal	1	0.1		
	Total	1471	87.1		
Total		1689	100.0		

Table 15:

RNSEGGrp Resp SEG <grouped> [pre-SOC2000]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Professional/employers/managers	359	21.3	21.4	21.4
	2 Intermediate non-manual	501	29.7	29.9	51.3
	3 Junior non-manual	227	13.4	13.5	64.8
	4 Supervisor/skilled manual	308	18.2	18.4	83.1
	5 Semi-skilled/personal services	239	14.2	14.2	97.4
	6 Unskilled manual	44	2.6	2.6	100.0
	Total	1678	99.3	100.0	
Missing	7 Armed forces	5	0.3		
	8 Inadequately described/not stated	6	0.4		
	Total	11	0.7		
Total		1689	100.0		

[Values **7** and **8** will be treated as missing. Only 11 cases, so hardly any effect on statistical analysis]

Table 16:

RNSocCl Respondent : social class [pre-SOC2000]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 I (SC=1)	108	6.4	6.4	6.4
	2 II (SC=2)	657	38.9	39.2	45.6
	3 III (non-manual) (SC=3)	343	20.3	20.4	66.0
	4 III (manual) (SC=4)	312	18.5	18.6	84.6
	5 IV (SC=5)	204	12.1	12.2	96.8
	6 V (SC=6)	54	3.2	3.2	100.0
	Total	1678	99.3	100.0	
Missing	7 Armed forces	5	0.3		
	8 Insufficient information	6	0.4		
	Total	11	0.7		
Total		1689	100.0		

[Values **7** and **8** will be treated as missing. Only 11 cases, so hardly any effect on statistical analysis]

Education related

   frequencies Tea HEdQual2.

Table 16:

Tea R how old when completed continuous full time education

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 15 or under	908	26.5	27.2	27.2
	2 16	965	28.2	28.9	56.1
	3 17	305	8.9	9.1	65.2
	4 18	423	12.4	12.7	77.9
	5 19 or over	740	21.6	22.1	100.0
	Total	3341	97.7	100.0	
Missing	6 Still at school	12	0.4		
	7 Still at college or university	57	1.7		
	97 Other answer (WRITE IN)	5	0.1		
	98 Don't know	3	0.1		
	99 Refusal	3	0.1		
	Total	80	2.3		
Total		3421	100.0		

[Treating values **6** and **7** as missing makes sample more homogenous and **[Tea]** smoothly ordinal.]

Table 17:

HEdQual2 Highest educational qualification obtained

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Postgraduate degree	174	5.1	5.6	5.6
	2 First degree	449	13.1	14.5	20.1
	3 Higher educ below degree	357	10.4	11.5	31.7
	4 A level or equiv	521	15.2	16.8	48.5
	5 O level or equiv	603	17.6	19.5	68.0
	6 CSE or equiv	244	7.1	7.9	75.9
	8 No qualification	746	21.8	24.1	100.0
	Total	3094	90.4	100.0	
Missing	7 Foreign or other	34	1.0		
	9 DK/Refusal/NA	293	8.6		
	Total	327	9.6		
Total		3421	100.0		

[Treating values **7** and **9** as missing makes **[hedqual2]** smoothly ordinal.]

Geographic

> **frequencies** GOR2 Country.

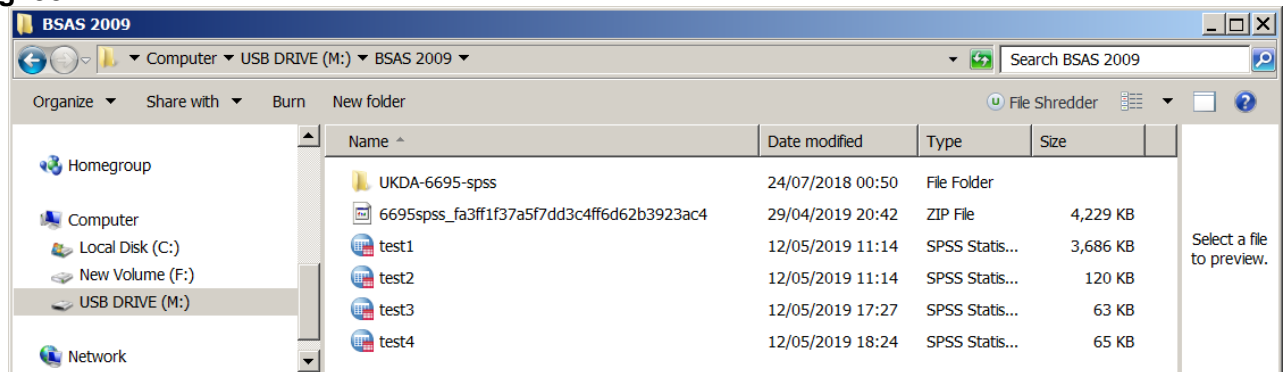**Table 18:**

GOR2 Government office region 2003 version:Q320

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 North East	100	5.9	5.9	5.9
	2 North West	188	11.1	11.1	17.1
	3 Yorkshire and Humberside	136	8.1	8.1	25.1
	4 East Midlands	119	7.0	7.0	32.1
	5 West Midlands	164	9.7	9.7	41.9
	6 SW	156	9.2	9.2	51.1
	7 Eastern	183	10.8	10.8	61.9
	8 Inner London	80	4.7	4.7	66.7
	9 Outer London	115	6.8	6.8	73.5
	10 South East	228	13.5	13.5	87.0
	11 Wales	75	4.4	4.4	91.4
	12 Scotland	145	8.6	8.6	100.0
	Total	1689	100.0	100.0	

Table 19:

Country Country of interview England, Scotland or Wales? :Q333

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 England	1469	87.0	87.0	87.0
	2 Scotland	145	8.6	8.6	95.6
	3 Wales	75	4.4	4.4	100.0
	Total	1689	100.0	100.0	

Fig. 33:

Fig. 34:
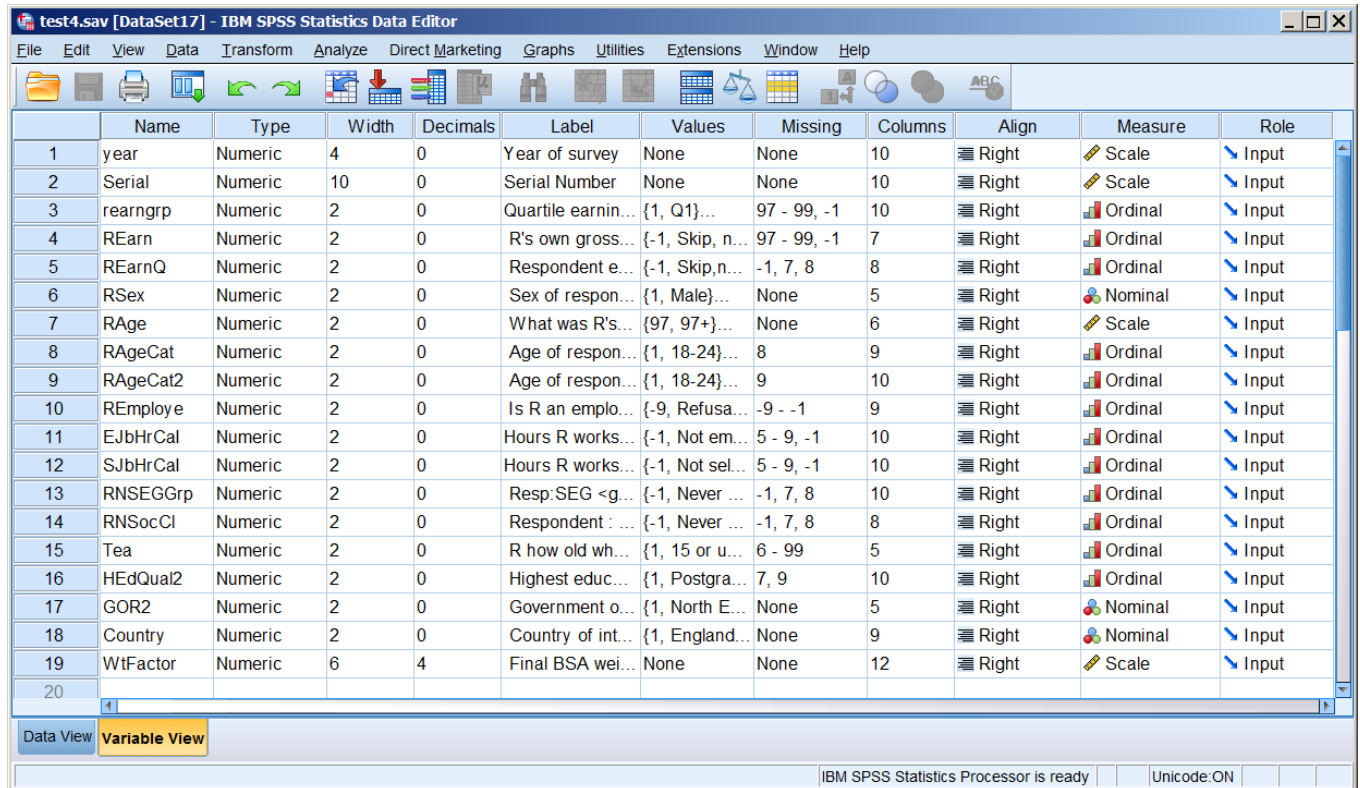
 test4

Double click on

File 🔲 test4 is now the active file.

Fig. 35:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	year	Numeric	4	0	Year of survey	None	None	10	≡ Right	◆ Scale	↘ Input
2	Serial	Numeric	10	0	Serial Number	None	None	10	≡ Right	◆ Scale	↘ Input
3	rearngrp	Numeric	2	0	Quartile earnin...	{1, Q1}...	97 - 99, -1	10	≡ Right	▦ Ordinal	↘ Input
4	REarn	Numeric	2	0	R's own gross...	{-1, Skip, n...	97 - 99, -1	7	≡ Right	▦ Ordinal	↘ Input
5	REarnQ	Numeric	2	0	Respondent e...	{-1, Skip,n...	-1, 7, 8	8	≡ Right	▦ Ordinal	↘ Input
6	RSex	Numeric	2	0	Sex of respon...	{1, Male}...	None	5	≡ Right	⬥ Nominal	↘ Input
7	RAge	Numeric	2	0	What was R's...	{97, 97+}...	None	6	≡ Right	◆ Scale	↘ Input
8	RAgeCat	Numeric	2	0	Age of respon...	{1, 18-24}...	8	9	≡ Right	▦ Ordinal	↘ Input
9	RAgeCat2	Numeric	2	0	Age of respon...	{1, 18-24}...	9	10	≡ Right	▦ Ordinal	↘ Input
10	REmploye	Numeric	2	0	Is R an emplo...	{-9, Refusa...	-9 - -1	9	≡ Right	▦ Ordinal	↘ Input
11	EJbHrCal	Numeric	2	0	Hours R works...	{-1, Not em...	5 - 9, -1	10	≡ Right	▦ Ordinal	↘ Input
12	SJbHrCal	Numeric	2	0	Hours R works...	{-1, Not sel...	5 - 9, -1	10	≡ Right	▦ Ordinal	↘ Input
13	RNSEGGrp	Numeric	2	0	Resp:SEG <g...	{-1, Never ...	-1, 7, 8	10	≡ Right	▦ Ordinal	↘ Input
14	RNSocCl	Numeric	2	0	Respondent : ...	{-1, Never ...	-1, 7, 8	8	≡ Right	▦ Ordinal	↘ Input
15	Tea	Numeric	2	0	R how old wh...	{1, 15 or u...	6 - 99	5	≡ Right	▦ Ordinal	↘ Input
16	HEdQual2	Numeric	2	0	Highest educ...	{1, Postgra...	7, 9	10	≡ Right	▦ Ordinal	↘ Input
17	GOR2	Numeric	2	0	Government o...	{1, North E...	None	5	≡ Right	⬥ Nominal	↘ Input
18	Country	Numeric	2	0	Country of int...	{1, England...	None	9	≡ Right	⬥ Nominal	↘ Input
19	WtFactor	Numeric	6	4	Final BSA wei...	None	None	12	≡ Right	◆ Scale	↘ Input
20											

[3.2.1.6 Earnings differences 2009] Extracting and saving selected variables

Changing the display of variable attributes

The author's preferred display of variable attributes is:

Name   Measure   Label   Values   Missing   Decimals

View >> Customize Variable View
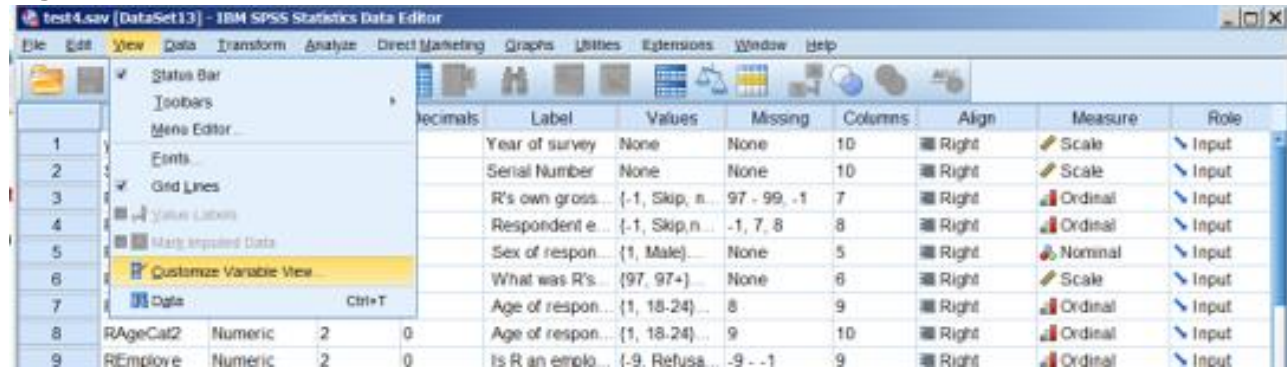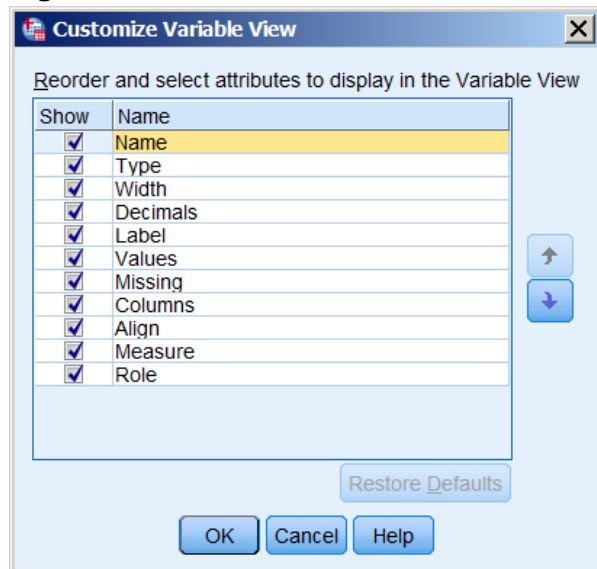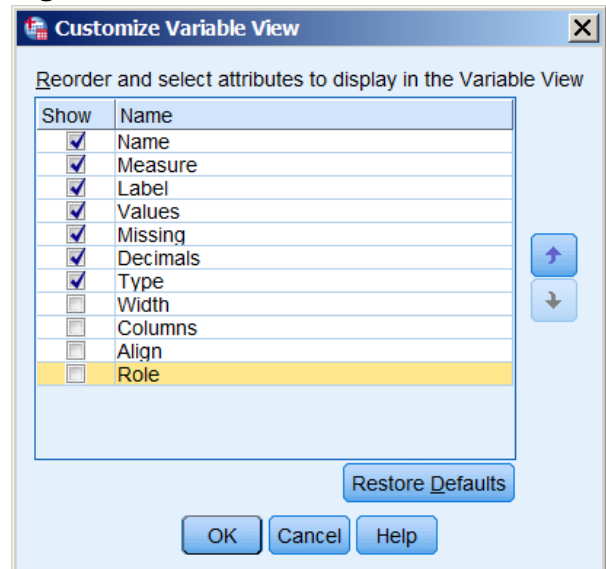
Fig. 35:

Fig. 36a:

Fig. 36b:

Use the blue arrows to move variables up or down the priority list. Uncheck to hide attributes you don't really need. Press OK

[3.2.1.6 Earnings differences 2009] Extracting and saving selected variables

Fig. 37:

	Name	Measure	Label	Values	Missing	Decimals	Type
1	year	Scale	Year of survey	None	None	0	Numeric
2	Serial	Scale	Serial Number	None	None	0	Numeric
3	rearngrp	Ordinal	Quartile earnin...	{1, Q1}...	97 - 99, -1	0	Numeric
4	REarn	Ordinal	R's own gross...	{-1, Skip, n...	97 - 99, -1	0	Numeric
5	REarnQ	Ordinal	Respondent e...	{-1, Skip,n...	-1, 7, 8	0	Numeric
6	RSex	Nominal	Sex of respon...	{1, Male}...	None	0	Numeric
7	RAge	Scale	What was R's...	{97, 97+}...	None	0	Numeric
8	RAgeCat	Ordinal	Age of respon...	{1, 18-24}...	8	0	Numeric
9	RAgeCat2	Ordinal	Age of respon...	{1, 18-24}...	9	0	Numeric
10	REmploye	Ordinal	Is R an emplo...	{-9, Refusa...	-9 - -1	0	Numeric
11	EJbHrCal	Ordinal	Hours R works...	{-1, Not em...	5 - 9, -1	0	Numeric
12	SJbHrCal	Ordinal	Hours R works...	{-1, Not sel...	-1, 5, 9	0	Numeric
13	RNSEGGrp	Ordinal	Resp:SEG <g...	{-1, Never ...	-1, 7, 8	0	Numeric
14	RNSocCl	Ordinal	Respondent : ...	{-1, Never ...	-1, 7, 8	0	Numeric
15	Tea	Ordinal	R how old wh...	{1, 15 or u...	6 - 99	0	Numeric
16	HEdQual2	Ordinal	Highest educ...	{1, Postgra...	7, 9	0	Numeric
17	GOR2	Nominal	Government o...	{1, North E...	None	0	Numeric
18	Country	Nominal	Country of int...	{1, England...	None	0	Numeric
19	WtFactor	Scale	Final BSA wei...	None	None	4	Numeric
20							

Since the variables are all **Numeric**, we don't really need **Type** either, but this format is much easier to navigate.

Ctrl S to save the file.

End of session: **3.2.1.6 Earnings differences 2009: Extracting and saving selected variables**

Next session: **3.2.1.7: Earnings differences 2009: Elaboration**