Survey Analysis Workshop   **© Copyright 2013   John F Hall**

Block 3: Analysing two variables (and sometimes three)

3.1.4.3: Income differences for test variables   [Draft only: 22 July 2013]

Research question:

Is there a difference between the incomes of men and women? What other variables might account for differences in income? What other variables might affect income regardless of gender? What effect do they have by themselves? What happens to any differences in income between men and women when controlling for these other variables?

Exemplar: British Social Attitudes 1989

Files: **e:weebly downloads\bsa89\3.1.4.2.sav** [Created in session 3.1.4.2]
                 If you don't have the file, download 3.1.4.2.sav from this site.

Previous: 3.1.4.2 Income differences – Build working file

Task: Recode gross earnings [**v1727**] into a new variable [**incr3**] with three categories: produce two-way contingency tables to investigate differences in earnings for each of the test variables.

In the previous exercises we produced a contingency table of **sex** * **v1727** in which epsilons (percentage point differences) were calculated between men and women for each earnings group.

sex Q901a Sex of respondent * v1727 Q.918b Income group of respondent (if working) Crosstabulation
% within sex Q901a Sex of respondent

	Q.918b Income group of respondent (if working)														Total
	Under £2000	£2000 < £2999	£3000 < £3999	£4000 < £4999	£5000 < £5999	£6000 < £6999	£7000 < £7999	£8000 < £9999	£10000 < £11999	£12000 < £1999	£15000 < £17999	£18000 < £19999	£20000 < £24000	£24000 or more	
	%	%	%	%	%	%	%	%	%	%	%	%	%	%	n = 100%
Total	5.2	5.7	5.8	6.0	7.4	7.2	8.1	11.6	11.2	12.2	7.1	3.7	1.9	7.0	1560
Men	0.3	0.8	0.9	2.4	5.4	5.3	8.7	13.4	14.1	17.4	10.9	5.7	3.2	11.6	**874**
Women	11.4	12.0	12.1	10.5	9.9	9.6	7.3	9.3	7.4	5.7	2.3	1.2	0.1	1.2	**686**
Epsilon	-11.1	-11.2	-11.2	-8.1	-4.5	-4.3	+1.4	+4.1	+6.7	+11.7	+8.6	+4.5	+3.1	+10.4	

In session 3.1.4.1 we looked at the distributions of our variables. In 3.1.4.2 we built up a working file with dependent variable earnings [**v1727**] independent variable sex [**v1411**) and selected test variables which might affect earnings. We then read in the raw data, added dictionary information, ran some file checks, saved the working file as 3.1.4.2.sav and the syntax file as 3.1.4.2.sps.

Dependent variable: Personal gross earnings   Q.918b   17   27   **v1727**
Independent variable: Sex   Q.901a   14   11   **v1411**

Test variables: Employee or self-employed   Q.23   2   71   **v271**
                      Hours worked, employee   Q.24   2   75   **v275**
                      Hours worked, self-employed Q.46a   4   61   **v461**
                      Public or private sector   Q.908f   16   17-18   **v1617**
                      Level of work   Q.908a   23   61   **v2361**
                      Terminal Education Age   Q.906a   15   30   **v1530**
                      Level of education [derived] Q.907b   23   74   **v2374**
                      Age last birthday   Q.901b   14   12-13   **v1412**

1

Step 1: Choosing cutting points

For demonstration purposes we now need to group earnings into fewer categories to make the tables easier to read when we produce conditional tables controlling for test variables. When we start controlling for test variables the number of cells rapidly gets very large and the cell counts consequently very small. Later we shall also need to reduce the number of categories for the test variables to keep the cell counts large enough to serve as a reasonable base for percentages. A general rule of thumb is that cell counts should be at least 40, at which level moving a single case from one category to another makes a net difference of 5 percentage points (it takes 2.5% from the source category adds 2.5% to the target category).

In this session we are going to produce two-way contingency tables to examine the income differences within each category of the test variables. For our current purpose, gross earnings [v1727] has too many categories to work with, so we need to reduce it to fewer categories. In session 3.1.4.1 we decided to use three categories.

For **three** groups, you can use the 33rd and 67th percentiles:

Statistics
v1727 Q.918b Income group of respondent (if working)

N	Valid	1560
	Missing	1465
Percentiles	33	**6.00**
	67	**9.00**

All

Statistics
v1727 Q.918b Income group of respondent (if working)

N	Valid	874
	Missing	519
Percentiles	33	**8.00**
	67	**10.00**

Men only

Statistics
v1727 Q.918b Income group of respondent (if working)

N	Valid	686
	Missing	946
Percentiles	33	**3.00**
	67	**7.00**

Women only

The 33rd percentile point for the whole sample is **6** but for men it is **8** and for women **3**: the 67th percentile point for the whole sample is **9** but for men it is **10** and for women **7**.

In exercise 3.1.4.1 we created a new variable **incr3** with **three** approximately equal-sized categories and tabulated it by **sex**:

*sex Q901a Sex of respondent * incr3 Q918b Gross income of R (if working) [3 groups] Crosstabulation*
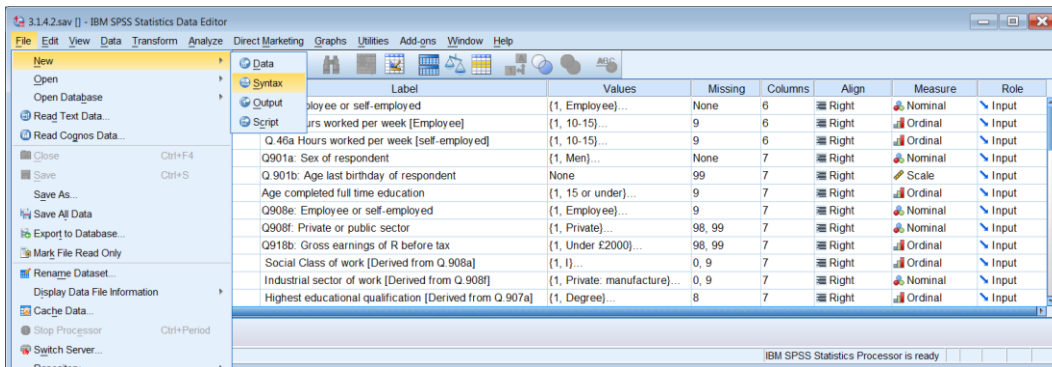
			Q918b Gross income of R (if working)			Total
			< £6000	< £12000	£12000+	
sex Q901a Sex of respondent	1 Men	Count	**86**	362	426	874
		% within sex Q901a Sex of respondent	9.8%	41.4%	48.7%	100.0%
	2 Women	Count	383	231	**72**	686
		% within sex Q901a Sex of respondent	55.8%	33.7%	10.5%	100.0%
Total		Count	469	593	498	1560
		% within sex Q901a Sex of respondent	30.1%	38.0%	31.9%	100.0%

. . to get **86** men earning under £6,000 and **72** women earning £12,000 or more. The numbers are quite small, but the cutting points divide earnings reasonably into Low (30%) Medium (38%) and High (32%) for the whole sample [*My rounding: 0.1% is only 2 cases out of 1560*].
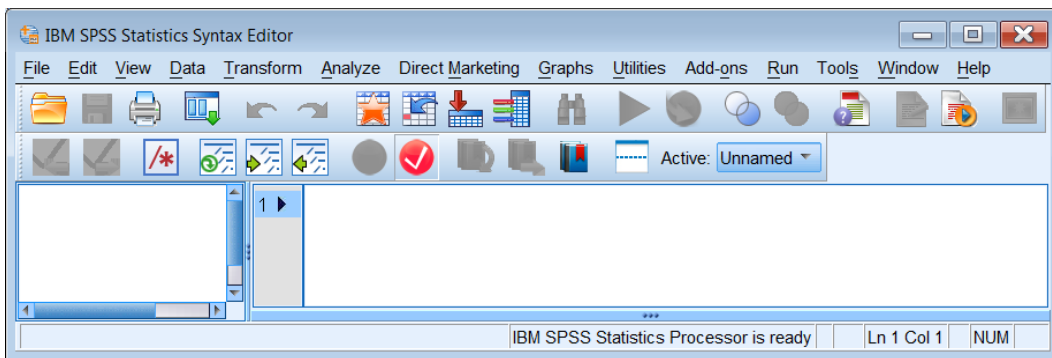
Open file 3.1.4.2 .sav and adjust the column separators to see the labels clearly.
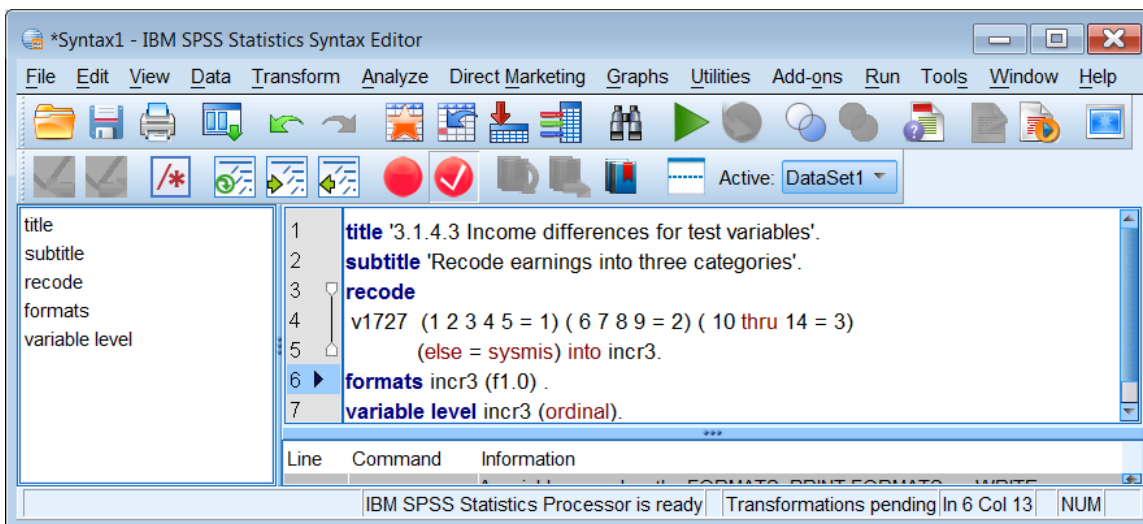
File > New > Syntax:

. . to open a blank **Syntax Editor**:

Step 1: Choose cutting points for earnings and create a new derived variable

```
title '3.1.4.3 Income differences for test variables'.
subtitle 'Recode earnings into three categories'.
recode
        v1727  (1 2 3 4 5 = 1) ( 6 7 8 9 = 2) ( 10 thru 14 = 3)
                (else = sysmis) into incr3.
formats incr (f1.0) .
variable level incr3 (ordinal).
```


We need the formats and variable level commands, otherwise **incr** will be set to SPSS defaults of **Decimals 2** and **Measurement** Nominal
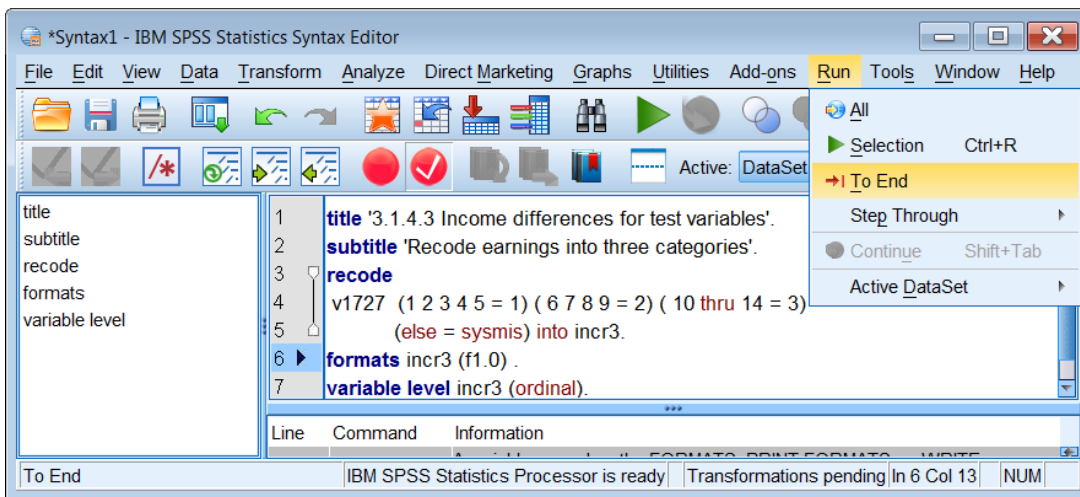
Place the cursor on the title command and click on Run > →| To End

Variable **incr3** has been appended to the file with Decimals set at 0 and Measure set at Ordinal, but no values have been calculated because SPSS has not yet made a pass through the data.
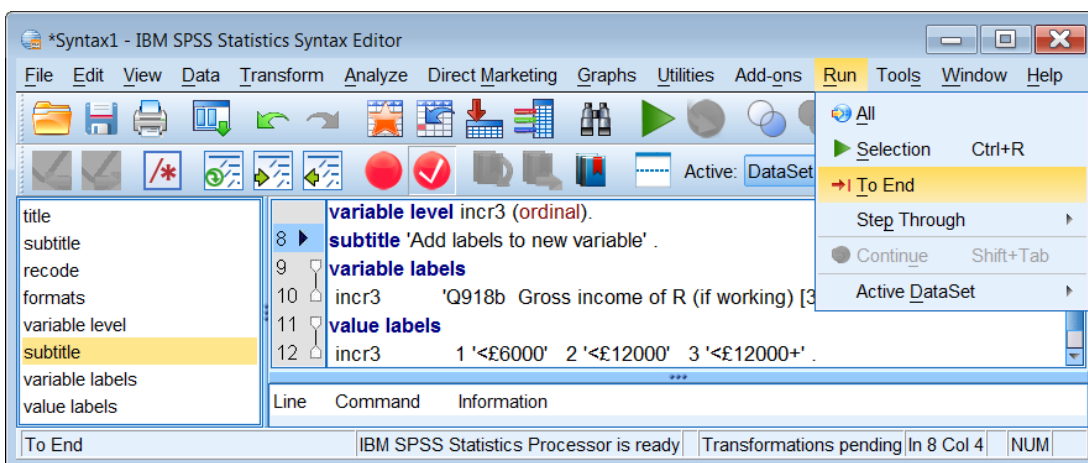

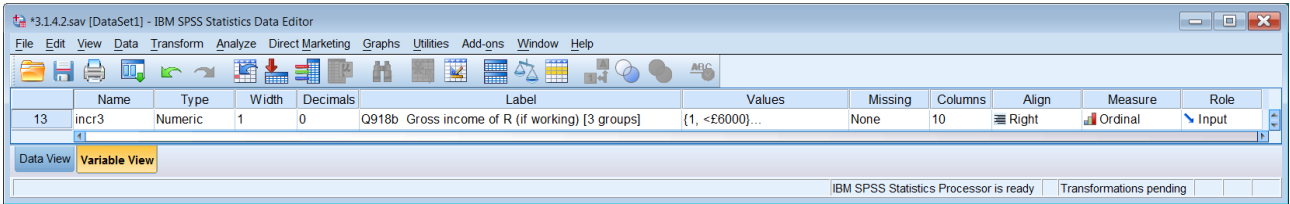**Step 2: Add labels to the new variable**

> subtitle 'Add labels to new variable' .
> variable labels
>   incr3   'Q918b Gross income of R (if working) [3 groups]'.
> value labels
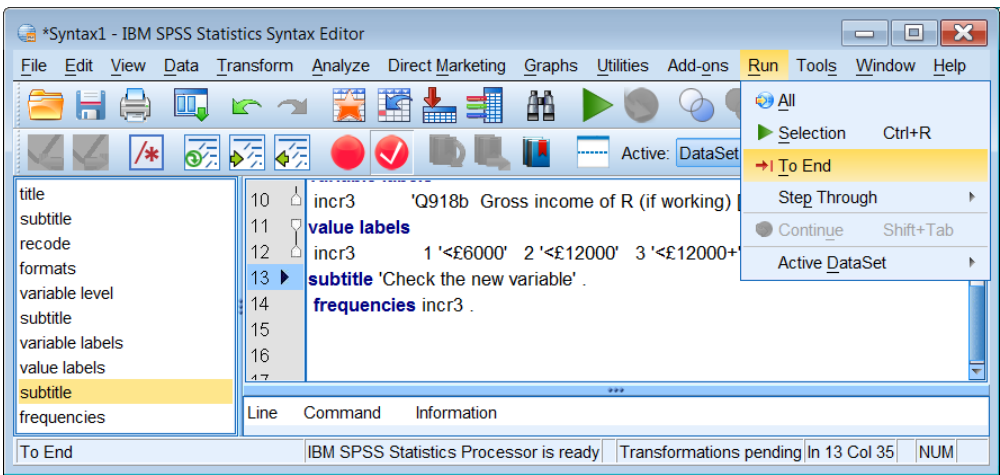>   incr3   1 '<£6000'   2 '<£12000'   3 '£12000+' .

Place the cursor on the subtitle command and click on Run > →| To End

4

Variable and value labels have been added to **incr**:

Step 3: Check the new variable

> **subtitle 'Check the new variable' .**
> **frequencies incr3 .**

Place the cursor on the next **subtitle** command and click on **Run** > **→| To End**

Statistics

incr3 Q918b Gross income of R (if working) [3 groups]

N	Valid	1560
	Missing	1465

incr3 Q918b Gross income of R (if working) [3 groups]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 <£6000	**469**	15.5	30.1	30.1
	2 <£12000	**593**	19.6	38.0	68.1
	3 <£12000+	**498**	16.5	31.9	100.0
	Total	1560	51.6	100.0	
Missing	System	1465	48.4		
Total		3025	100.0		

This seems fine, with three reasonably sized groups (469, 593 and 498).

Go back to the **Data Editor** and click on **File** > **Save as**

5
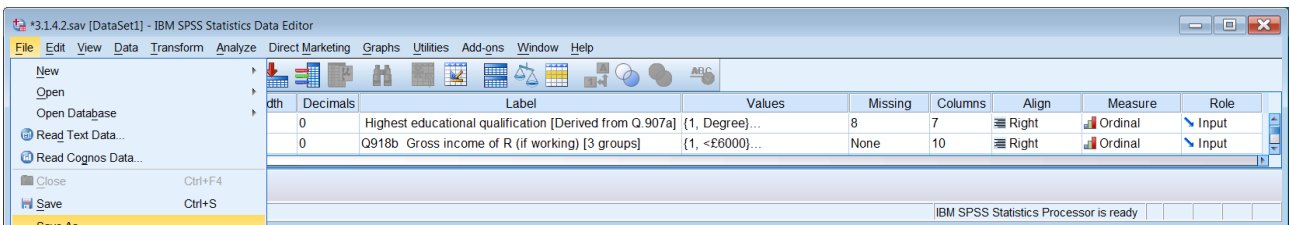
Change 3.1.4.2.sav to: 3.1.4.3.sav   and click on Save
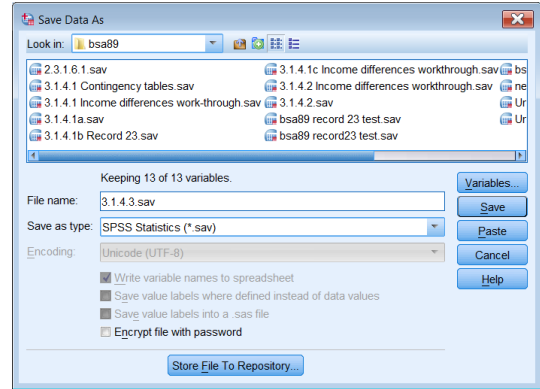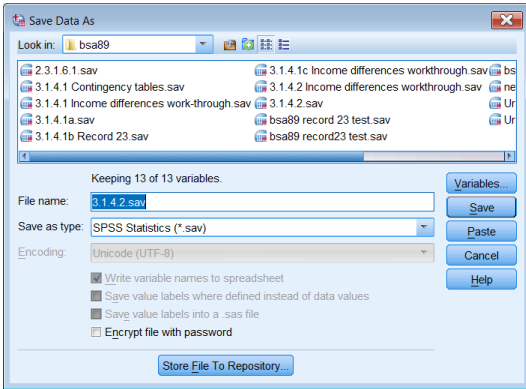
Step 4: Tabulate test variables by earnings

Tabulate all test variables (except age [v1412]) by incr3 (earnings in three groups):

> subtitle ' Tabulate test variables by earnings' .
> crosstabs v271 to v463, v1530 to v1617 v2363 v2374 by incr3 .
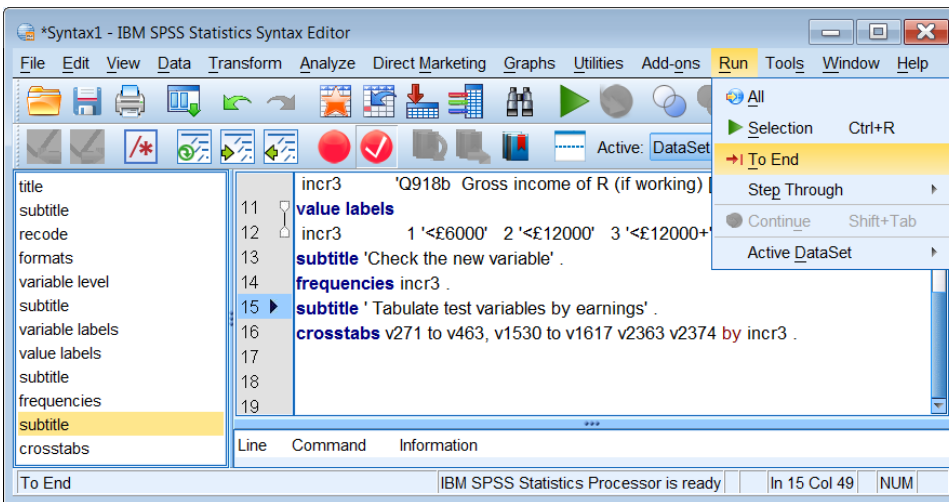
Place the cursor on the next subtitle command and click on Run > →| To End

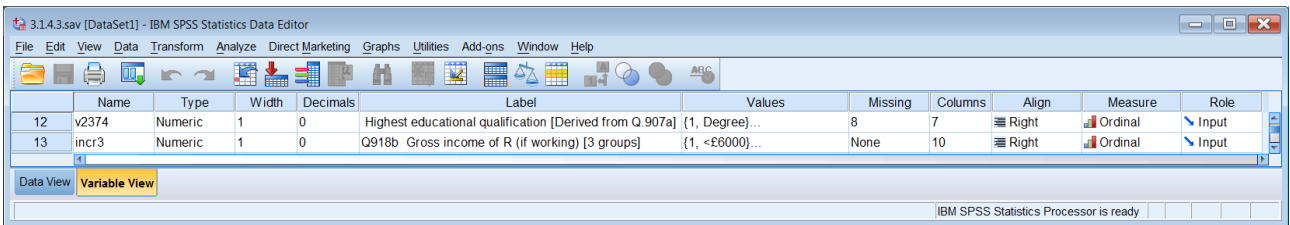**v271 Q23: Employee or self-employed * incr3 Q918b Gross income of R (if working) [3 groups] Crosstabulation**

Count

		incr3 Q918b Gross income of R (if working) [3 groups]			Total
		1 <£6000	2 <£12000	3 £12000+	
v271 Q23: Employee or self-employed	1 Employee	419	540	416	1375
	2 Self-employed	50	53	82	185
Total		469	593	498	1560

v275 Q24: Hours worked per week [Employee] * incr3 Q918b Gross income of R (if working) [3 groups] Crosstabulation

Count

		incr3 Q918b Gross income of R (if working) [3 groups]			Total
		1 <£6000	2 <£12000	3 £12000+	
v275 Q24: Hours worked per week [Employee]	1 10-15	81	4	1	86
	2 16-23	100	10	1	111
	3 24-29	55	12	4	71
	4 30 or more	183	514	410	1107
Total		419	540	416	1375

v463 Q.46a Hours worked per week [self-employed] * incr3 Q918b Gross income of R (if working) [3 groups] Crosstabulation

Count

		incr3 Q918b Gross income of R (if working) [3 groups]			Total
		1 <£6000	2 <£12000	3 £12000+	
v463 Q.46a Hours worked per week [self-employed]	1 10-15	7	2	1	10
	2 16-23	10	3	2	15
	3 24-29	4	0	0	4
	4 30 or more	29	48	79	156
Total		50	53	82	185

v1530 Age completed full time education * incr3 Q918b Gross income of R (if working) [3 groups] Crosstabulation

Count

		incr3 Q918b Gross income of R (if working) [3 groups]			Total
		1 <£6000	2 <£12000	3 £12000+	
v1530 Age completed full time education	1 15 or under	231	223	119	573
	2 16	139	193	131	463
	3 17	32	51	54	137
	4 18	24	60	46	130
	5 19 or 0ver	42	65	146	253
	6 Still at school	0	0	2	2
	7 Still at college	1	1	0	2
Total		469	593	498	1560

v1616 Q908e: Employee or self-employed * incr3 Q918b Gross income of R (if working) [3 groups] Crosstabulation

Count

		incr3 Q918b Gross income of R (if working) [3 groups]			Total
		1 <£6000	2 <£12000	3 £12000+	
v1616 Q908e: Employee or self-employed	1 Employee	421	540	415	1376
	2 Self-employed	48	53	83	184
Total		469	593	498	1560

v1617 Q908f: Private or public sector * incr3 Q918b Gross income of R (if working) [3 groups] Crosstabulation

Count

		incr3 Q918b Gross income of R (if working) [3 groups]			Total
		1 <£6000	2 <£12000	3 £12000+	
v1617 Q908f: Private or public sector	1 Private	265	355	246	866
	2 Nationalised	14	42	37	93
	3 Local Government	80	53	52	185
	4 Health Authority	35	33	22	90
	5 Civil Service	10	38	35	83
	6 Charity or Trust	6	7	7	20
	7 Other	10	12	16	38
Total		420	540	415	1375

v2363 Social Class of work [Derived from Q.908a] * incr3 Q918b Gross income of R (if working) [3 groups] Crosstabulation

Count

		incr3 Q918b Gross income of R (if working) [3 groups]			Total
		1 <£6000	2 <£12000	3 £12000+	
v2363 Social Class of work [Derived from Q.908a]	1 I	3	15	80	98
	2 II	67	119	196	382
	3 III non-manual	149	154	76	379
	4 III manual	63	187	111	361
	5 IV	135	92	22	249
	6 V	49	15	5	69
Total		466	582	490	1538

v2367 Industrial sector of work [Derived from Q.908f] * incr3 Q918b Gross income of R (if working) [3 groups] Crosstabulation

Count

		incr3 Q918b Gross income of R (if working) [3 groups]			Total
		1 <£6000	2 <£12000	3 £12000+	
v2367 Industrial sector of work [Derived from Q.908f]	1 Private: manufacture	96	147	112	355
	2 Private: non-manufacture	180	202	134	516
	3 Public: service	132	135	120	387
	4 Public: Manufacture + transport	4	36	26	66
Total		412	520	392	1324

v2374 Highest educational qualification [Derived from Q.907a] * incr3 Q918b Gross income of R (if working) [3 groups] Crosstabulation

Count

|		incr3 Q918b Gross income of R (if working) [3 groups]			Total
		1 <£6000	2 <£12000	3 £12000+	
v2374 Highest educational qualification [Derived from Q.907a]	1 Degree	21	34	109	164
	2 HE below degree	35	104	124	263
	3 A-level or equiv.	19	69	100	188
	4 O-level or equiv	109	154	76	339
	5 CSE or equiv	56	59	18	133
	6 Foreign and other	4	1	0	5
	7 None	225	171	71	467
Total		469	592	498	1559

5: Save your work:
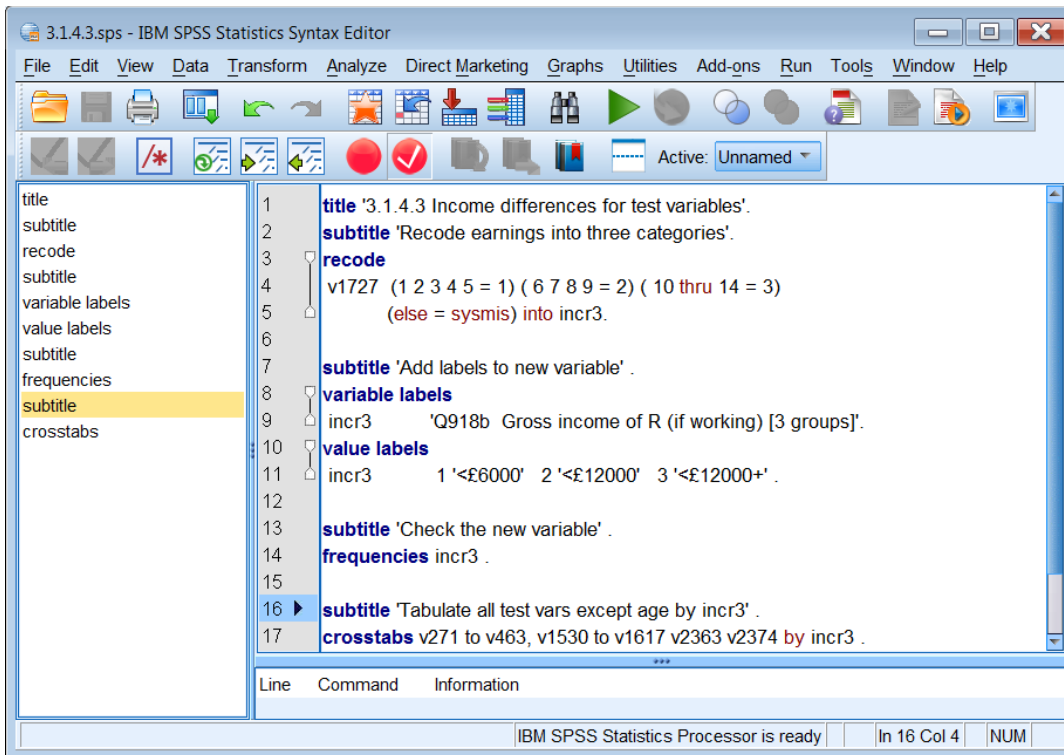
You've already saved your **Data Editor** as 3.1.4.3.sav and you haven't made any changes, but you should also save your syntax. Go back to the **Syntax Editor** and click on File > Save as
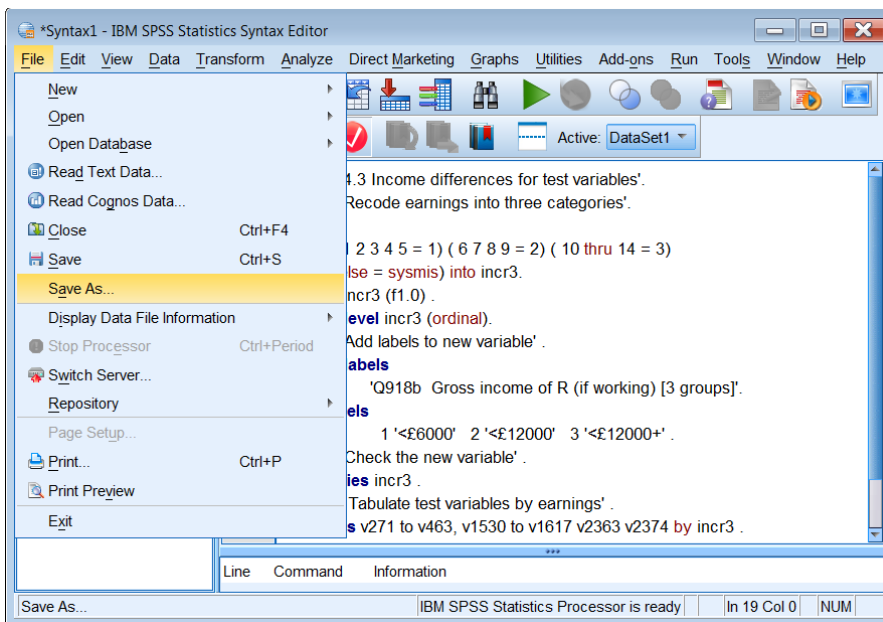
Change Syntax1.sps to: 3.1.4.3.sps and click on Save
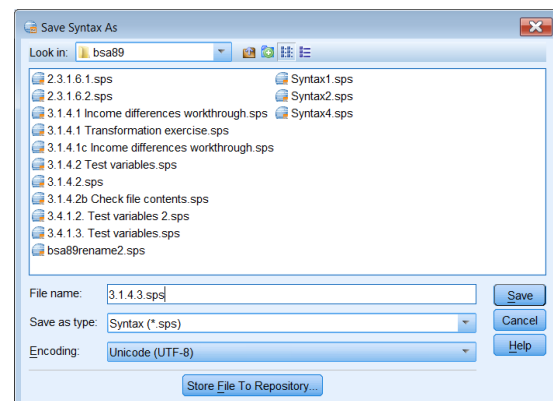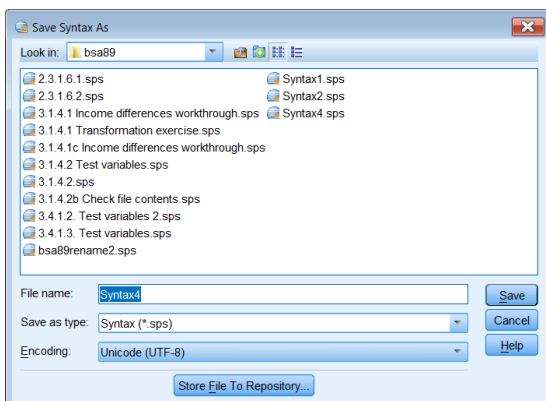
For tutorial purposes we need to group the test variables into fewer categories, preferably dichotomies.

Age is problematic because of different retirement ages for men (65) and women (60). To keep things manageable we shall need to reduce the data by restricting the sample to a more homogenous group of people below pensionable age working 30 hours or more per week, which we define as working full-time. We'll do that in 3.1.4.4

End of session   3.1.4.3: Income differences for test variables

Next sessions:

3.1.4.4: Income differences - Choose test variables and cutting points
> Decide which test variables to use and choose cutting points; recode test variables into derived variables with fewer categories; produce two-way contingency tables to investigate income differences for the derived test variables.

3.2.: Income differences - Elaboration
> Three-way contingency tables to see what happens to income differences between men and women when controlling for test variables.

Back to: Block 3 Analysing two variables (and sometimes three)
              3.1 Two variables (**CROSSTABS**)
              3.1.4.2 Income differences – Build a working file

Forward to: 3.1.4.4 Income differences - Choose test variables and cutting points [b]