Block 3: Analysing two variables (and sometimes three)

3.1.4.5 Income differences for derived test variables [Draft only: 19 August 2013]

Research questions:

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

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

What effect do they have by themselves?

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

Exemplar: British Social Attitudes 1989

Files: 3.1.4.4.sav
[Created in session 3.1.4.4 and saved to **e:weebly downloads\bsa89**]

In session 3.1.4.4 Income differences - Choose test variables and cutting points we selected possible test variables from the full data set, then created derived variables with fewer categories to make the data easier to work with.

Cutting points were chosen to keep category counts large enough to act as a base for percentages, but also bearing in mind the need for the resulting categories to make sociological sense.

We finished up with the following test variables:

Mode of work

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Parttime	317	10.5	18.8	18.8
	Fulltime	1365	45.1	81.2	100.0
	Total	1682	55.6	100.0	
Missing	System	1343	44.4		
Total		3025	100.0		

Social class of work

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Non-manual	1487	49.2	52.2	52.2
	Manual	1359	44.9	47.8	100.0
	Total	2846	94.1	100.0	
Missing	System	179	5.9		
Total		3025	100.0		

Highest qualification level

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A-level or above	944	31.2	31.4	31.4
	O-level or CSE	778	25.7	25.9	57.3
	None	1283	42.4	42.7	100.0
	Total	3005	99.3	100.0	
Missing	System	20	.7		
Total		3025	100.0		

Age completed full-time education

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	15 or under	1421	47.0	48.0	48.0
	16 or 17	972	32.1	32.8	80.8
	18 or over	568	18.8	19.2	100.0
	Total	2961	97.9	100.0	
Missing	System	64	2.1		
Total		3025	100.0		

Age group if working

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	18 – 29	668	22.1	22.1	22.1
	30 – 49	1124	37.2	37.2	59.2
	50 or over	538	17.8	17.8	77.0
	Pensioner	695	23.0	23.0	100.0
	Total	3025	100.0	100.0	

What effect on earnings from paid work do these variables have by themselves?

Homework exercise:

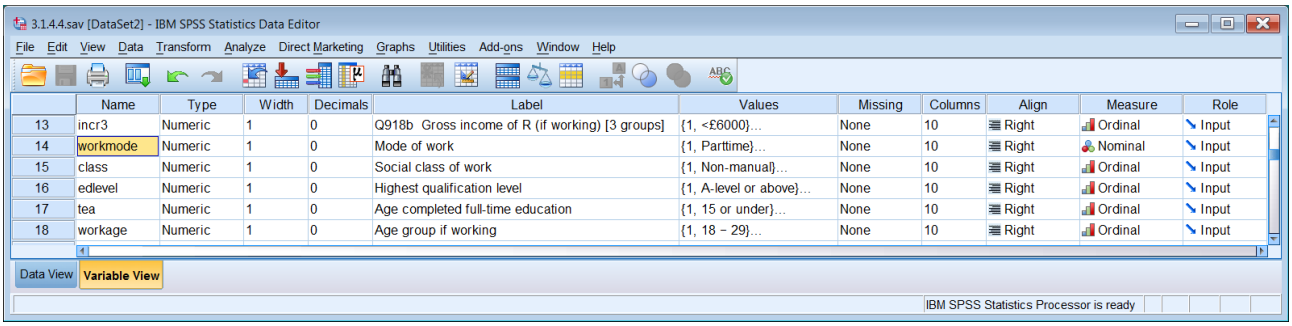
Taking grouped personal income **incr3** as the **dependent** variable and **workmode**, **class**, **edlevel**, **tea** and **workage** as the **independent** variables, produce joint frequency distributions (contingency tables) with appropriate percentages to compare the earnings of different groups.

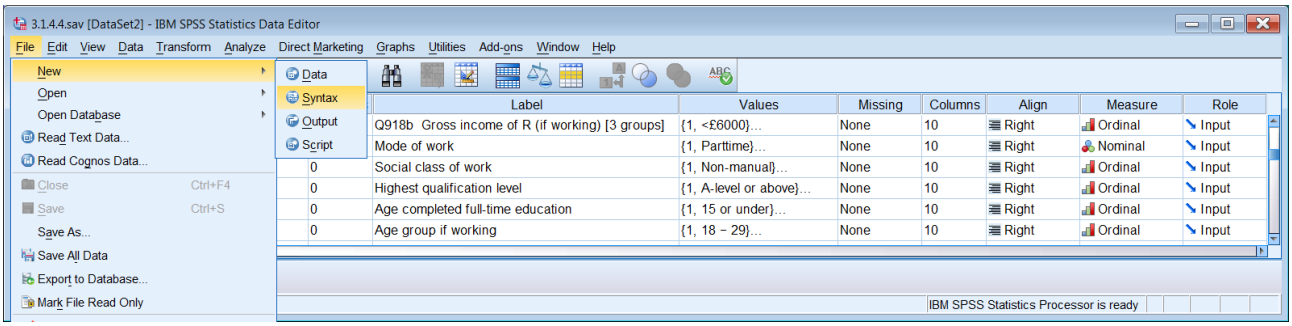
To get tables which are in fact easier to interpret, put the independent variables down the side of the tables (the rows) and the dependent variable across the top (the columns) using my preferred sociological rather than the statistical convention.

Which percentages do you need, and why?

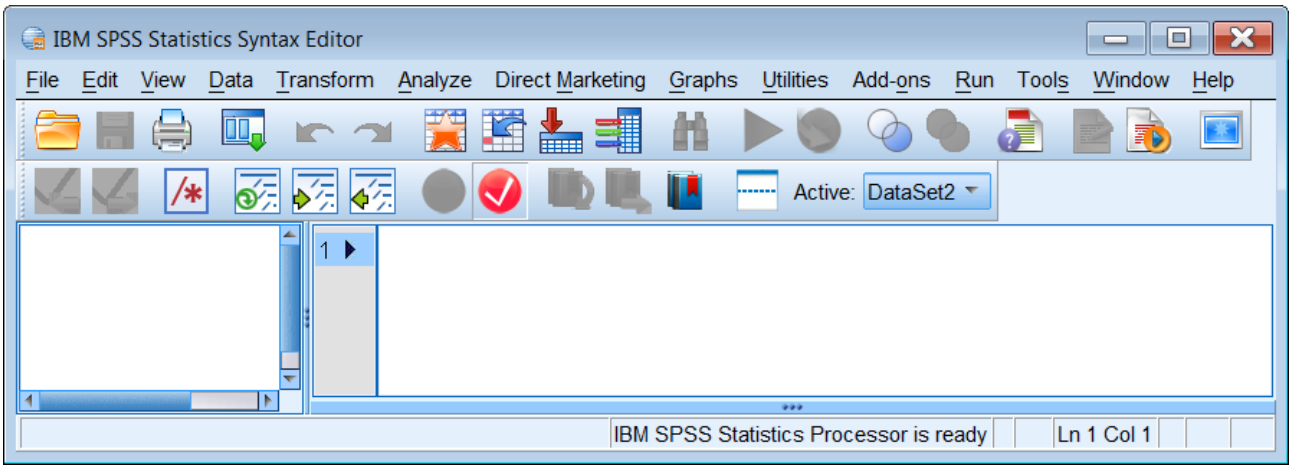
There's a specimen answer on the next page, but try to do this without peeping!

Open file [3.1.4.4.sav]



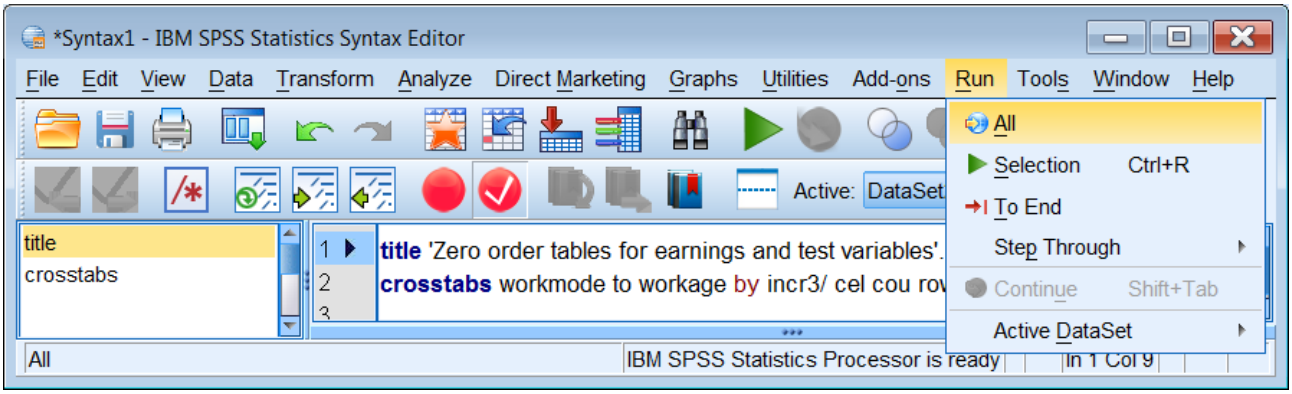


File > New > Syntax to open a new Syntax Editor:




```
title 'Zero order tables for earnings and test variables'.
crosstabs workmode to workage by incr3/ cel cou row.
```

Place the cursor on the title command and press Run > All .



Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Mode of work * Q918b Gross income of R (if working) [3 groups]	1560	51.6%	1465	48.4%	3025	100.0%
Social class of work * Q918b Gross income of R (if working) [3 groups]	1538	50.8%	1487	49.2%	3025	100.0%
Highest qualification level * Q918b Gross income of R (if working) [3 groups]	1554	51.4%	1471	48.6%	3025	100.0%
Age completed full-time education * Q918b Gross income of R (if working) [3 groups]	1556	51.4%	1469	48.6%	3025	100.0%
Age group if working * Q918b Gross income of R (if working) [3 groups]	1560	51.6%	1465	48.4%	3025	100.0%

Mode of work * Q918b Gross income of R (if working) [3 groups] Crosstabulation

			Q918b Gross income of R (if working) [3 groups]			Total
			<£6000	<£12000	£12000 +	
Mode of work	Parttime	Count	257	31	9	297
		% within Mode of work	86.5%	10.4%	3.0%	100.0%
	Fulltime	Count	212	562	489	1263
		% within Mode of work	16.8%	44.5%	38.7%	100.0%
Total		Count	469	593	498	1560
		% within Mode of work	30.1%	38.0%	31.9%	100.0%

The tables produced are quite cluttered (see example above) and not easy to interpret, so it's easier to just ask for row percent, but then we lose the row totals:

```
crosstabs workmode to workage by incr3 / cel row.
```

Mode of work * Q918b Gross income of R (if working) [3 groups] Crosstabulation

% within Mode of work

		Q918b Gross income of R (if working) [3 groups]			Total
		<£6000	<£12000	<£12000+	
Mode of work	Parttime	86.5%	10.4%	3.0%	100.0%
	Fulltime	16.8%	44.5%	38.7%	100.0%
Total		30.1%	38.0%	31.9%	100.0%

We already know that the category counts within the derived variables are large enough to serve as a base n for % and the 100% in the above table tells us nothing, so I've modified it below by substituting n for 100%, getting rid of the % signs in the table body, moving the row total to the top of the table and calculating epsilons (percentage point differences) for the dichotomies.

The effect of part-time working is so marked that we should think about leaving the part-time workers out and restricting our analysis to those working full-time.

Mode of work * Q918b Gross income of R (if working) [3 groups] Crosstabulation

% within Mode of work

		Q918b Gross income of R (if working) [3 groups]			n=100%
		<£6000 %	<£12000 %	£12000+ %	
Total		30.1	38.0	31.9	1560
Mode of work	Parttime	86.5	10.4	3.0	297
	Fulltime	16.8	44.5	38.7	1263
	Epsilon	**-69.7**	**+34.1**	**+35.7**	

There is a clear gradient in favour of non-manual work:

Social class of work * Q918b Gross income of R (if working) [3 groups] Crosstabulation

% within Social class of work

		Q918b Gross income of R (if working) [3 groups]			n=100%
		<£6000 %	<£12000 %	£12000+ %	
Total		30.3	37.8	31.9	1538
Social class of work	Non-manual	25.5	33.5	41.0	859
	Manual	36.4	43.3	20.3	679
	Epsilon	**-10.9**	**-9.8**	**+20.7**	

Educational qualifications [**edlevel**] and terminal education age [**tea**]show clear gradients in both lower and upper earnings categories, but these two variables will be correlated, so their effect will be to some extent confounded: perhaps we should just use one of them?

Highest qualification level * Q918b Gross income of R (if working) [3 groups] Crosstabulation

% within Highest qualification level

		Q918b Gross income of R (if working) [3 groups]			n=100%
		<£6000 %	<£12000 %	£12000+ %	
Total		29.9	38.0	32.0	1554
Highest qualification level	A-level or above	12.2	33.7	54.1	615
	O-level or CSE	35.0	45.1	19.9	472
	None	48.2	36.6	15.2	467

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

% within Age completed full-time education

		Q918b Gross income of R (if working) [3 groups]			n=100%
		<£6000 %	<£12000 %	£12000+ %	
Total		30.1	38.0	31.9	1556
Age completed full-time education	15 or under	40.3	38.9	20.8	573
	16 or 17	28.5	40.7	30.8	600
	18 or over	17.2	32.6	50.1	383

The education variables are not dichotomies, but we can instead calculate the percentage difference between the highest and lowest educational categories:

	Q918b Gross income of R (if working) [3 groups]		
	<£6000 %	<£12000 %	£12000+ %
edlevel (A level - None)	-36.0	-2.9	+38.9
tea (18 or over - 15 or under)	-23.1	-6.3	+29.3

The effects of age are less clear. There is little difference between age groups at the lower end of the earnings scale, but a marked leap of 19.8 percentage points after age 30.

Age group if working * Q918b Gross income of R (if working) [3 groups] Crosstabulation

% within Age group if working

		Q918b Gross income of R (if working) [3 groups]			n=100%
		<£6000 %	<£12000 %	£12000+ %	
Total		30.1	38.0	31.9	1560
Age group if working	18 – 29	28.8	51.4	19.8	420
	30 – 49	29.9	30.4	39.6	815
	50 or over	28.7	40.7	30.7	300
	Pensioner	72.0	28.0		25

We could try a different grouping of 18 -39 and 40+, but this is inconclusive and again there may also be some interaction with other variables.

workage2 * Q918b Gross income of R (if working) [3 groups] Crosstabulation

% within workage2

		Q918b Gross income of R (if working) [3 groups]			n = 100%
		<£6000 %	<£12000 %	£12000 + %	
Total		30.1	38.0	31.9	1560
workage2	18 – 39	28.3	41.0	30.7	809
	40 or over	30.6	35.0	34.4	726
	Epsilon	+2.3	-6.0	+3.7	

Working with contingency tables for three or more variables can quickly lead to masses of indigestible information, especially when we start investigating combinations of independent variables. We also need to be aware of possible interactions between variables. For instance many older respondents will have completed their full time education at a much younger age. Access to better paid non-manual work will be dependent on educational qualifications which older respondents have not had the opportunity to acquire.

We also need to think about simplifying vast amounts of information by refining our analysis. One way of doing this is to restrict the sample to more homogeneous categories such as people working full-time rather than part-time, or who are employees rather than self-employed. Given that political arguments about gender discrimination in earnings are normally concerned only with employees, this makes sense. At the very least we also need to analyse part-time and full-time workers separately.

Another research trick is to choose a single category of the dependent variable, treat that as a criterion value then simply tabulate that by categories of the independent variable. It doesn't matter whether this is the low-earning or high-earning category, but we'll take the latter and summarise the results of this session by tabulating the percentage earning £12,000 or more per annum.

For the whole sample this was 31.9%, but this figure is only a weighted average of all the possible sub-samples earning £12,000 pa or more.

People earning £12,000+ from paid work

Variable	Category	%	n = 100%	Zero order epsilon
	All	**31.9**	1560	
Sex	Men	48.7	874	
	Women	10.5	686	+38.2
Work mode	Parttime	3.0	297	
	Fulltime	38.7	1263	+35.7
Social class	Non-manual	41.0	859	
	Manual	20.3	679	+20.7
Educational quals	A-level or above	54.1	615	
	O-level or CSE	19.9	472	
	None	15.2	467	+38.9
Terminal education age	15 or under	20.8	573	
	16 or 17	30.8	600	
	18 or over	50.1	383	+29.3
Age group	18 – 29	19.8	420	
	30 – 49	39.6	815	+19.8
	50 or over	30.7	300	

In the next session we'll be producing three-way contingency tables to see what happens to income differences between men and women when controlling for the newly derived test variables.

End of session: 3.1.4.5 Income differences for derived test variables

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

Forward to: 3.2 Three variables

[Back to Block 3: Analysing two variables]