3.2: Three variables

3.2.4 Income differences – Elaboration

[Draft only: 1 October 2014 16:39]

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

4: 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 downloadable from this site]

In session 3.1.4.5 Income differences for derived test variables we produced a set of zero order tables to investigate the different proportions of people earning £12,000 or more a year from paid work.

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 this session we shall be producing three-way contingency tables to see what happens to income differences between men and women when controlling for a third variable. The first example will compare the earnings of men and women controlling for mode of work (full-time or part-time).

For the logical structure (see Jim Ring's Statistics notes to accompany course. pp31-32)

Y, X or T	**Name**	**Label**
Y = Dependent	**incr3**	[income in three groups]
X = Independent	**sex**	[Men, Women]
T = Test	**workmode**	[Full-time, Part-time]

Sequence is:

Y, X, T	(incr3 sex workmode)
X by Y}	(sex workmode by incr3)
T by Y }	(sex workmode by incr3)
X by Y by T	(sex by incr3 by workmode)

SPSS output can get very cluttered using various options for cell contents or if both names and labels are used. For the following tables Edit >> Options >> Output has been set to Labels only.

Initial frequency counts

frequencies incr3 sex workmode.

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

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

Q901a: Sex of respondent

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Men	1393	46.0	46.0	46.0
	Women	1632	54.0	54.0	100.0
	Total	3025	100.0	100.0	

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		

Zero order tables

X → Y (sex by incr3)
T → Y (workmode by incr3)

[**NB:** Output from **CROSSTABS** can get very wide and has to be edited to fit on a portrait page.]

[Note to myself: Is there a way to control column widths in output?]

1: **crosstabs** sex workmode by incr3. (Default output, counts only)

Q901a: Sex of respondent * Q918b Gross income of R (if working) [3 groups] Crosstabulation

Count

		Q918b Gross income of R (if working) [3 groups]			
		<£6000	<£12000	£12000+	Total
Q901a: Sex of respondent	Men	86	362	426	874
	Women	383	231	72	686
Total		469	593	498	1560

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

Count

		Q918b Gross income of R (if working) [3 groups]			
		<£6000	<£12000	£12000+	Total
Mode of work	Parttime	257	31	9	297
	Fulltime	212	562	489	1263
Total		469	593	498	1560

2: **crosstabs** sex workmode by incr3 /cells row. (row % only)

Q901a: Sex of respondent * Q918b Gross income of R (if working) [3 groups] Crosstabulation

% within Q901a: Sex of respondent

		Q918b Gross income of R (if working) [3 groups]			
		<£6000	<£12000	£12000+	Total
Q901a: Sex of respondent	Men	9.8%	41.4%	48.7%	100.0%
	Women	55.8%	33.7%	10.5%	100.0%
Total		30.1%	38.0%	31.9%	100.0%

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]			
		<£6000	<£12000	£12000+	Total
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%

3: **crosstabs** sex workmode by incr3 /cells count row. (counts and row%)

Q901a: Sex of respondent * Q918b Gross income of R (if working) [3 groups] Crosstabulation

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

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

			Q918b Gross income of R (if working) [3 groups]			
			<£6000	<£12000	£12000+	Total
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%

First order tables

X → Y . T (sex by earnings, controlling for hours worked)

1: **crosstabs** sex by incr3 by workmode. (Default output, counts only)

Q901a: Sex of respondent * Q918b Gross income of R (if working) [3 groups] * Mode of work Crosstabulation

Count

			Q918b Gross income of R (if working) [3 groups]			
Mode of work			<£6000	<£12000	£12000+	Total
Parttime	Q901a: Sex of respondent	Men	12	4	5	21
		Women	245	27	4	276
	Total		257	31	9	297
Fulltime	Q901a: Sex of respondent	Men	74	358	421	853
		Women	138	204	68	410
	Total		212	562	489	1263
Total	Q901a: Sex of respondent	Men	86	362	426	874
		Women	383	231	72	686
	Total		469	593	498	1560

2: **crosstabs** sex by incr3 by workmode /cells row. (Row % only)

Q901a: Sex of respondent * Q918b Gross income of R (if working) [3 groups] * Mode of work Crosstabulation

% within Q901a: Sex of respondent

Mode of work			Q918b Gross income of R (if working) [3 groups] <£6000	<£12000	£12000+	Total
Parttime	Q901a: Sex of respondent	Men	57.1%	19.0%	23.8%	100.0%
		Women	88.8%	9.8%	1.4%	100.0%
	Total		86.5%	10.4%	3.0%	100.0%
Fulltime	Q901a: Sex of respondent	Men	8.7%	42.0%	49.4%	100.0%
		Women	33.7%	49.8%	16.6%	100.0%
	Total		16.8%	44.5%	38.7%	100.0%
Total	Q901a: Sex of respondent	Men	9.8%	41.4%	48.7%	100.0%
		Women	55.8%	33.7%	10.5%	100.0%
	Total		30.1%	38.0%	31.9%	100.0%

See what I mean about clutter?

3: **crosstabs** sex by incr3 by workmode /cells count row. (counts and row %)

Q901a: Sex of respondent * Q918b Gross income of R (if working) [3 groups] * Mode of work Crosstabulation

Mode of work				Q918b Gross income of R (if working) [3 groups] <£6000	<£12000	£12000+	Total
Parttime	Q901a: Sex of respondent	Men	Count	12	4	5	21
			% within Q901a: Sex of respondent	57.1%	19.0%	23.8%	100.0%
		Women	Count	245	27	4	276
			% within Q901a: Sex of respondent	88.8%	9.8%	1.4%	100.0%
	Total		Count	257	31	9	297
			% within Q901a: Sex of respondent	86.5%	10.4%	3.0%	100.0%
Fulltime	Q901a: Sex of respondent	Men	Count	74	358	421	853
			% within Q901a: Sex of respondent	8.7%	42.0%	49.4%	100.0%
		Women	Count	138	204	68	410
			% within Q901a: Sex of respondent	33.7%	49.8%	16.6%	100.0%
	Total		Count	212	562	489	1263
			% within Q901a: Sex of respondent	16.8%	44.5%	38.7%	100.0%
Total	Q901a: Sex of respondent	Men	Count	86	362	426	874
			% within Q901a: Sex of respondent	9.8%	41.4%	48.7%	100.0%
		Women	Count	383	231	72	686
			% within Q901a: Sex of respondent	55.8%	33.7%	10.5%	100.0%
	Total		Count	469	593	498	1560
			% within Q901a: Sex of respondent	30.1%	38.0%	31.9%	100.0%

Ouch!! **CROSSTABS** output displays both cell counts and row %%, so the table is now completely cluttered, unreadable and unusable: you certainly couldn't publish it like this.

CTABLES gives full control of output, but the syntax looks very complicated to the uninitiated (ie me!). The default output can still be a bit cluttered, but can be modified within the program. The default output can be very sparse for analysing one variable, but at least for a frequency distribution you don't get totally unnecessary cumulative totals for nominal variables.

Within the **CTABLES** command, tables have to be specified one at a time using:

```
/TABLE
```

1: Initial frequency counts

```
ctables /table incr3
        /table sex
        /table workmode.
```

		Count
Q918b Gross income of R (if working) [3 groups]	<£6000	469
	<£12000	593
	<£12000+	498

		Count
Q901a: Sex of respondent	Men	1393
	Women	1632

		Count
Mode of work	Parttime	317
	Fulltime	1365

2: Zero order tables (counts only)

```
ctables   /TABLE sex BY incr3
          /TABLE workmode BY incr3.
```

		Q918b Gross income of R (if working) [3 groups]		
		<£6000	<£12000	£12000+
		Count	Count	Count
Q901a: Sex of respondent	Men	86	362	426
	Women	383	231	72

		Q918b Gross income of R (if working) [3 groups]		
		<£6000	<£12000	£12000+
		Count	Count	Count
Mode of work	Parttime	257	31	9
	Fulltime	212	562	489

Note there are no column totals in the above tables. Also to compare groups we need row percentages, not counts, and the percentages need to be based on the row totals. In **CTABLES** these are specified by: **[ROWPCT.COUNT]**.

3: Zero order tables (with row %)

```
ctables   /TABLE sex BY incr3  [ROWPCT.COUNT]
          /TABLE workmode BY incr3  [ROWPCT.COUNT ].
```

		Q918b Gross income of R (if working) [3 groups]		
		<£6000	<£12000	£12000+
		Row N %	Row N %	Row N %
Q901a: Sex of respondent	Men	9.8%	41.4%	48.7%
	Women	55.8%	33.7%	10.5%

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

Unlike **CROSSTABS**, **CTABLES** allows you to display the row totals in the same table by requesting TOTALS [COUNT] as an additional element inside the square brackets:

```
ctables   /TABLE sex BY incr3  [ROWPCT.COUNT TOTALS [COUNT]]
```

However, to display totals you need an additional line for each /TABLE specification:

```
          /CATEGORIES VARIABLES=  incr3 TOTAL=YES .
```

```
ctables   /TABLE sex BY incr3  [ROWPCT.COUNT TOTALS [COUNT]]
          /CATEGORIES VARIABLES=  incr3 TOTAL=YES
          /TABLE workmode BY incr3  [ROWPCT.COUNT TOTALS [COUNT]]
          /CATEGORIES VARIABLES=  incr3 TOTAL=YES .
```

		Q918b Gross income of R (if working) [3 groups]			
		<£6000	<£12000	£12000+	Total
		Row N %	Row N %	Row N %	Count
Q901a: Sex of respondent	Men	9.8%	41.4%	48.7%	874
	Women	55.8%	33.7%	10.5%	686

		Q918b Gross income of R (if working) [3 groups]			
		<£6000	<£12000	£12000+	Total
		Row N %	Row N %	Row N %	Count
Mode of work	Parttime	86.5%	10.4%	3.0%	297
	Fulltime	16.8%	44.5%	38.7%	1263

These tables are now much easier to read and interpret, but they are still slightly cluttered: they can be further improved by getting rid of all the % signs in the body of the table and by changing the column headers. The elements in the /TABLES specification can be extended by adding a label in double primes eg: [ROWPCT.COUNT "%"] and [COUNT "n = 100%"]

The default formats are integer for cell counts and one decimal place for percentages, but if needed the latter can be changed by adding a format eg:

[ROWPCT.COUNT f4.2]

However two decimal places for percentages seems a bit pointless for these data. We're trying to reduce the clutter, not add to it!

```
ctables
  /TABLE sex BY incr3  [ROWPCT.COUNT f5.1 "%"  totals [count "n= 100%"]]
        /CATEGORIES VARIABLES=  incr3 TOTAL=YES
  /TABLE workmode BY incr3  [ROWPCT.COUNT f5.1 "%" totals [count "n= 100%"]]
        /CATEGORIES VARIABLES=  incr3 TOTAL=YES .
```

		Q918b Gross income of R (if working) [3 groups]			
		<£6000	<£12000	£12000+	Total
		%	%	%	n= 100%
Q901a: Sex of respondent	Men	9.8	41.4	48.7	874
	Women	55.8	33.7	10.5	686
	Epsilon	-46.0	**7.7**	**38.2**	

		Q918b Gross income of R (if working) [3 groups]			
		<£6000	<£12000	£12000+	Total
		%	%	%	n= 100%
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	

The above tables do not have totals for the income groups as it's easier to compare the income groups of men/women and full-time/part-time workers without them (and to calculate the percentage point differences, epsilons).

[**NB**: The epsilons were produced separately by copying the tables into Excel, performing the calculations and then copying the epsilons back into Word]

I wonder if **CTABLES** can produce tables with epsilons?

For elaboration purposes you need to compare these conditional distributions with the original distribution to see how it has been partitioned when controlling for test variables. More test variables can be added at any stage.

Model

Dependent variable	Y = Earnings group	(<£6000, <£12000, £12000+)
Independent variable	X = Sex	(Men, Women)
Test variable 1	T_1 = Hours of work	(Part-time, Full-time
Test variable 2	T_2 = Type of work	(Non-manual, manual)

Zero order tables

1: X → Y Sex by earnings group
2: T_1 → Y Workmode by earnings group
3: T_2 → Y Type of work by earnings group

ctables

```
/TABLE sex BY incr3  [ROWPCT.COUNT f5.1 "%"  totals [count "n= 100%"]]
/CATEGORIES VARIABLES=  incr3 TOTAL=YES
/TABLE workmode BY incr3  [ROWPCT.COUNT f5.1 "%" totals [count "n= 100%"]]
/CATEGORIES VARIABLES=  incr3 TOTAL=YES .
```

		Q918b Gross income of R (if working) [3 groups]			
		<£6000	<£12000	£12000+	Total
		%	%	%	n= 100%
Q901a: Sex of respondent	Men	9.8	41.4	48.7	874
	Women	55.8	33.7	10.5	686
	Total	30.1	38.0	31.9	1560
	Epsilon	**-46.0**	**+7.7**	**+38.2**	

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

Both variables can be included in the same table if X and T_1 are linked with a **+** sign.

```
/TABLE sex [c] + workmode [c] by incr3 [c]
        [ROWPCT.count  f8.1 "%" TOTALS[validn f8.0  "n= 100%"]]
/CATEGORIES VARIABLES= sex edlevel incr3 TOTAL=YES POSITION=AFTER.
```

	<£6000	<£12000	£12000+	Total
	%	%	%	n= 100%
Men	9.8	41.4	48.7	874
Women	55.8	33.7	10.5	686
Total	30.1	38.0	31.9	1560
Parttime	86.5	10.4	3.0	297
Fulltime	16.8	44.5	38.7	1263
Total	30.1	38.0	31.9	1560

First order tables

1: X → Y . T_1
2: X → Y . T_2

To produce three-way contingency tables in CTABLES, the specification of variables is slightly different. One pair of variables has to linked by **>** (variable on the right of **>** is nested in categories of the variable on the left). There three ways of producing such tables:

X > T by Y **T > X by Y** **X > Y by T**

1: X > T_1 by Y

/TABLE sex > workmode by incr3 (nests workmode within sex):

		<£6000	<£12000	£12000+	Total
		%	%	%	n= 100%
Men	Parttime	57.1	19.0	23.8	21
	Fulltime	8.7	42.0	49.4	853
	Total	9.8	41.4	48.7	874
Women	Parttime	88.8	9.8	1.4	276
	Fulltime	33.7	49.8	16.6	410
	Total	55.8	33.7	10.5	686
Total	Parttime	86.5	10.4	3.0	297
	Fulltime	16.8	44.5	38.7	1263
	Total	30.1	38.0	31.9	1560

2: T_1 > X by Y

/TABLE workmode > sex by incr3 (nests sex within workmode):

		<£6000	<£12000	£12000+	Total
		%	%	%	n= 100%
Parttime	Men	57.1	19.0	23.8	21
	Women	88.8	9.8	1.4	276
	Total	86.5	10.4	3.0	297
Fulltime	Men	8.7	42.0	49.4	853
	Women	33.7	49.8	16.6	410
	Total	16.8	44.5	38.7	1263
Total	Men	9.8	41.4	48.7	874
	Women	55.8	33.7	10.5	686
	Total	30.1	38.0	31.9	1560

However it's easier to compare men and women when the table is spread out using

3: **X by T_1 > Y**

/TABLE sex by workmode > incr3

The full **ctables** syntax looks like this:

```
ctables
    /VLABELS VARIABLES=sex incr3 workmode DISPLAY=NONE
    /TABLE sex by workmode > incr3
    [ROWPCT.COUNT f5.1 "%"  TOTALS [COUNT "n= 100%"]]
    /CATEGORIES VARIABLES=  sex workmode incr3  TOTAL=YES POSITION=after.
```

Earnings from paid work of men and women controlling for hours worked

	Parttime				Fulltime				Total			
	<£6000	<£12000	£12000+	Total	<£6000	<£12000	£12000+	Total	<£6000	<£12000	£12000+	Total
	%	%	%	n= 100%	%	%	%	n= 100%	%	%	%	n= 100%
Men	57.1	19.0	23.8	21	8.7	42.0	49.4	853	9.8	41.4	48.7	874
Women	88.8	9.8	1.4	276	33.7	49.8	16.6	410	55.8	33.7	10.5	686
Total	86.5	10.4	3.0	297	16.8	44.5	38.7	1263	30.1	38.0	31.9	1560
Epsilon	**-31.7**	**+9.2**	**+22.4**		**-25.0**	**-7.8**	**+32.8**		**-46.0**	**+7.7**	**+38.2**	

From this table it is possible to construct a summary table to show what happens to differences in earnings of men and women when controlling for hours worked (Full time is 30 or more hours per week). Taking a criterion category of £12,000 or more per annum as an indicator of "high earnings" the figure of 31.9% for the whole sample of 1560 can be broken down into 48.7% of 874 men and 10.5% of 686 women. For hours worked the 31.9% is broken down into 3.0% of 297 working part-time and 38.7% of 1263 working full-time.

People earning £12,000 or more per annum from paid work

%
(n = 100%)

		Part time	Full time	Zero order epsilon	First order epsilon
All	**31.9** (1560)	3.0 (297)	38.7 (1263)	**-35.7**	
Men	48.7 (874)	23.8 (21)	49.4 (853)		**-25.6**
Women	10.5 (686)	1.4 (276)	16.6 (410)		**-15.2**
Zero order epsilon	**+38.2**				
First order epsilon		**+22.4**	**+32.8**		

Now do the same controlling for class (type of work)**: X → Y . T_2**

1: X > T_2 by Y

/TABLE sex > class by incr3 (nests class within sex)

		<£6000	<£12000	£12000+	Total
		%	%	%	n= 100%
Men	Non-manual	5.9	25.9	68.3	410
	Manual	13.5	55.4	31.1	444
	Total	9.8	41.2	48.9	854
Women	Non-manual	43.4	40.5	16.0	449
	Manual	79.6	20.4	.0	235
	Total	55.8	33.6	10.5	684
Total	Non-manual	25.5	33.5	41.0	859
	Manual	36.4	43.3	20.3	679
	Total	30.3	37.8	31.9	1538

2: T_2 > X by Y

/TABLE class > sex by incr3 (nests sex within class):

		<£6000	<£12000	£12000+	Total
		%	%	%	n= 100%
Non-manual	Men	5.9	25.9	68.3	410
	Women	43.4	40.5	16.0	449
	Total	25.5	33.5	41.0	859
Manual	Men	13.5	55.4	31.1	444
	Women	79.6	20.4	.0	235
	Total	36.4	43.3	20.3	679
Total	Men	9.8	41.2	48.9	854
	Women	55.8	33.6	10.5	684
	Total	30.3	37.8	31.9	1538

As before it's easier to compare men and women when the table is spread out using

3: **X by T_2 > Y**

```
/TABLE sex by class > incr3
```

The full **ctables** syntax looks like this:

```
ctables
  /VLABELS VARIABLES=sex incr3 class DISPLAY=NONE
  /TABLE sex by class ] > incr3
        [ROWPCT.COUNT f8.1 "%"    TOTALS [COUNT f8.0  "n= 100%"]]
 /CATEGORIES VARIABLES= sex class incr3 TOTAL=YES POSITION=AFTER.
```

Earnings from paid work of men and women controlling for type of work

	Non-manual				Manual				Total			
	<£6000	<£12000	£12000+	Total	<£6000	<£12000	£12000+	Total	<£6000	<£12000	£12000+	Total
	%	%	%	n= 100%	%	%	%	n= 100%	%	%	%	n= 100%
Men	5.9	25.9	68.3	410	13.5	55.4	31.1	444	9.8	41.2	48.9	854
Women	43.4	40.5	16.0	449	79.6	20.4	.0	235	55.8	33.6	10.5	684
Total	25.5	33.5	41.0	859	36.4	43.3	20.3	679	30.3	37.8	31.9	1538
Epsilon	**-37.5**	**-14.6**	**+52.3**		**-66.1**	**+35.0**	**+31.1**		**-46.0**	**+7.6**	**+38.4**	

From this table it is possible to construct another summary table, this time to show what happens to differences in earnings of men and women when controlling for type of work (non-manual/manual). The counts are slightly different because type of work couldn't be classified for some people. Again taking the criterion value of £12,000 or more per annum as an indicator of "high earnings" the figure of 31.9% for the whole sample of 1538 can be broken down into 48.9% of 854 men and 10.5% of 684 women. For type of work the 31.9% is broken down into 41.0% of 859 non-manual and 20.3% of 679 manual workers.

People earning £12,000 or more p.a. from paid work

% (n = 100%)		Non-manual	Manual	Zero order epsilon	First order epsilon
All	**31.9** (1538)	41.0 (859)	20.3 (679)	**+20.7**	
Men	48.9 (854)	68.3 (410)	31.1 (444)		**+37.2**
Women	10.5 (684)	16.0 (449)	0.0 (235)		**+16.0**
Zero order epsilon	**+38.4**				
First order epsilon		**+52.3**	**+31.1**		

Second order table

X → Y . T_1 T_2

Four-way tables can be produced in SPSS, but they are complex to read and interpret: it's preferable when controlling for a second test variable **T_2** (in this case type of work: non-manual/manual) to select only those working full time. The selection has to temporary otherwise all other cases will be lost from the working file.

```
temp.
select if workmode = 2.
ctables
  /VLABELS VARIABLES=sex incr3 class DISPLAY=NONE
 /table sex by class > incr3 [ROWPCT.COUNT f8.1 "%" TOTALS[COUNT f8.0  "n= 100%"]]
 /CATEGORIES VARIABLES= sex class incr3 TOTAL=YES POSITION=AFTER.
```

Earnings from paid work of men and women working full time (30 or more hours a week)

	Non-manual				Manual				Total			
	<£6000	<£12000	£12000+	Total	<£6000	<£12000	£12000+	Total	<£6000	<£12000	£12000+	Total
	%	%	%	n= 100%	%	%	%	n= 100%	%	%	%	n= 100%
Men	5.0	26.2	68.8	401	12.0	56.4	31.6	433	8.6	41.8	49.5	834
Women	22.8	54.4	22.8	298	62.7	37.3	.0.0	110	33.6	49.8	16.7	408
Total	12.6	38.2	49.2	699	22.3	52.5	25.2	543	16.8	44.4	38.7	1242
Epsilon	**-17.8**	**-28.2**	**+46.0**		**-50.7**	**+19.1**	**+31.6**		**-25.0**	**-8.0**	**+32.8**	

As before it is possible to construct a summary table to show what happens to differences in earnings from paid work of men and women controlling simultaneously for hours worked and type of work, in this case by selecting only those working full time. Again, taking the criterion category of £12,000 or more per annum for earnings of men and women in full time work, the figure of 38.7% for the sub-sample of 1242 can be broken down into 49.5% of 543 men and 16.7% of 408 women. For type of work the 38.7% is broken down into 49.2% of 699 non-manual and 25.2% of 543 manual workers.

People earning £12,000 or more per annum from full time paid work

% **(n = 100%)**	All	Non-manual	Manual	First order epsilon	Second order epsilon
All	**38.7** (1242)	49.2 (699)	25.2 (543)	**+24.0**	
Men	49.5 (834)	68.8 (401)	31.6 (433)		**+37.2**
Women	16.7 (408)	22.8 (298)	0.0 (110)		**+22.8**
First order epsilon	+32.8				
Second order epsilon		**+46.0**	**+31.6**		

Elaboration relies on epsilon (percentage point difference) and is best used with dichotomised data, but can be used to compare any two categories of variables with three or more categories. It's not particularly sophisticated as it loses information when categories are condensed, but it was good enough for Rosenberg. It's easily understood by beginners, simple to specify tables in SPSS **CROSSTABS** and very useful for making students think about effects and interactions. **CTABLES** is perhaps too complex to specify for beginners, but the tables are much more useful.

This tutorial involved creating the following variables by grouping some variables into fewer categories or by combining two variables into one.

display labels /variables = sex incr3 to workage.

Variable Labels

Variable	Position	Label
sex	5	Q901a: Sex of respondent
incr3	14	Q918b Gross income of R (if working) [3 groups]
workmode	15	Mode of work
class	16	Social class of work
edlevel	17	Highest qualification level
tea	18	Age completed full-time education
workage	19	Age group if working

Variables included in the analysis so far are:

Dependent variable	Y = Earnings group	(Ordinal <£6000, <£12000, £12000+)
Independent variable	X = Sex	(Dichotomy Men, Women)
Test variable 1	T_1 = Hours of work	(Dichotomy Part-time, Full-time
Test variable 2	T_2 = Type of work	(Dichotomy Non-manual, manual)

Other test variables not yet considered include (all Ordinal):

Test variable 3	T_3 = edlevel	Highest qualification level (A-level+, O-level/CSE+, None)
Test variable 4	T_4 = tea	Age completed full-time education (15, 16-17, 18+)
Test variable 5	T_5 = workage	Age group if working (18-29, 30-49, 50+)

Some of the test variables will be correlated to some degree (in statistical jargon, there will be inter-action). Age will be correlated with educational level and age of completion of full time education: educational level will be correlated with type of work. These inter-actions can be neutralised if they are all controlled simultaneously. To do this with tabulation makes for some seriously complex specifications for tables, in which the cell counts used as a base for % soon become too small to be reliable.

As well as using epsilons, it would also be possible to use appropriate measures of association such as *gamma* or *phi*, to see how they change under different conditions, but that belongs to a different tutorial.

Another way of dealing with this problem is to use a statistical technique called multi-level modelling using categorical variables, but this is beyond the scope of this tutorial (and well above my statistical competence).

End of session: **3.2.4 Income differences – Elaboration**

Back to: 3.1.4.5 Income differences for derived test variables
3.2 Three variables
Block 3: Analysing two variables