Survey Analysis Workshop

3.2: Three variables

3.2.1.1 Earnings differences – Elaboration

[18 Oct 2014: Revised 2017 and 2 July 2021]

Exemplar: British Social Attitudes Survey 1989 [UKDS SN 2723]

File: **3.2.1.1 Elaboration 1 (BSAS 1989)**

Important Notice

New European General Data Protection Regulations (GDPR) mean that **no actual data can be uploaded to this site** from the British Social Attitudes Survey (BSAS) . Instead users must check the series list of available files and click the Access link to request downloads of individual source files direct from the UK Data Service (UKDDS). See: Downloading British Social Attitudes Survey (BSAS) data from the UK Data Service

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?

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 with gross earnings (before tax) of £12,000 or more a year from paid work.

Table 1: Zero-order summary

People earning £12,000 or more 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**
Type of work	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	

[Source: British Social Attitudes Survey, 1989]

In this session we shall be producing 1st-order (three-way) contingency tables to see what happens to any differences in gross earnings between men and women when **controlling for a third (test) variable**.

The first example compares the gross earnings (before tax) of men and women **controlling for mode of work** (full-time or part-time).

Status[1]	**Name**	**Label**
Y = Dependent	**earngrp**	[Gross annual earnings: <£6000 <£12000 £12000+]
X = Independent	**sex**	[Men, Women]
T = Test	**workmode**	[Full-time, Part-time]

Frequencies	**Y, X, T**	**frequencies earngrp sex workmode .**
Zero order tables	**X → Y** **T → Y**	**crosstabs sex workmode by earngrp .**
1st order table	**X → Y . T**	**crosstabs sex by workmode by earngrp .**

[1] For an explanation of the logic involved, see Jim Ring's Statistics notes to accompany course. pp31-32)
See also: Rosenberg M. ***The Logic of Survey Analysis*** (New York, Basic Books, 1968)

SPSS output can get quite cluttered if you display both names and labels or use too many options for cell contents. For the following tables Edit >> Options >> Output both variables and values have been set to display Labels·

Initial frequency counts

frequencies earngrp, sex, workmode .

Table 2:

Gross earnings 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		

Table 3:

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	

Table 4:

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 → earngrp**]
T → Y [**workmode → earngrp**]

[Default output: display counts only]

```
crosstabs sex workmode by earngrp .
```

Table 5: Effect of sex on earnings

Sex of respondent * Gross earnings of R (if working) [3 groups] Crosstabulation

Count

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

Table 6: Effect of workmode on earnings

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

Count

		Gross earnings 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

[Display row % only]

```
crosstabs sex workmode by earngrp /cells row.
```

Table 7: Effect of sex on earnings

Sex of respondent * Gross earnings of R (if working) [3 groups] Crosstabulation

% within Sex of respondent

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

Table 8: Effect of workmode on earnings

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

% within Mode of work

		Gross earnings 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%

[Display both counts and row %]

[**NB: CROSSTABS** output can get very wide [2] and needs editing to fit on an A4 portrait page.]

```
crosstabs sex workmode by earngrp /cells count row .
```

Table 9: Effect of sex on earnings

Sex of respondent * Gross earnings of R (if working) [3 groups] Crosstabulation

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

Table 10: Effect of workmode on earnings

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

			Gross earnings 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%

[**NB:** The tables above are beginning to look a bit cluttered.]

[2] [Is there a way to control column widths in output?]

First order table

X → Y . T [**sex** → **earnings** controlling for **mode of work**]

[Default output: display counts only]

```
crosstabs sex by earngrp by workmode .
```

Table 11: Effect of sex on earnings, controlling for workmode

Sex of respondent * Gross earnings of R (if working) [3 groups] * Mode of work Crosstabulation

Count

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

[Display row % only]

```
crosstabs sex by earngrp by workmode /cells row.
```

Table 12: Effect of sex on earnings, controlling for workmode

Sex of respondent * Gross earnings of R (if working) [3 groups] * Mode of work Crosstabulation

% within Sex of respondent

Mode of work			Gross earnings of R (if working) [3 groups] <£6000	<£12000	£12000+	Total
Parttime	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	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	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%

[Display both counts and row %]

```
crosstabs sex by earngrp by workmode /cells count row.
```

Table 12: Effect of sex on earnings, controlling for workmode

Sex of respondent * Gross earnings of R (if working) [3 groups] * Mode of work Crosstabulation

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

The table is now quite cluttered and difficult to interpret. Every cell in the output displays both counts and row %: you certainly couldn't publish it like this.

However, a solution is available in SPSS.

Custom Tables

The SPSS command **CTABLES** gives much more control of output, but (unless you use the GUI) the syntax can get very complex to the uninitiated (i.e. me!). For analysing one variable, the default output can be very sparse, but at least frequency distributions don't contain totally meaningless cumulative percentages totals for nominal variables.

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

To display the variables in **columns**:

```
ctables /table <variable>
        /table <variable> .
```

To display the variables in **rows**:

```
ctables /table by <variable>
        /table by <variable> .
```

1: Initial frequency counts

```
ctables  /table earngrp
         /table sex
         /table workmode .
```

Table 13

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

Table 14

		Count
Sex of respondent	Men	1393
	Women	1632

Table 15

		Count
Mode of work	Parttime	317
	Fulltime	1365

Zero order tables (counts only)

For reasons which will become clear, my preference is to display the dependent variable **Y** in rows:

```
ctables   /table sex by earngrp .
```

Table 16

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

```
ctables   /table workmode by earngrp .
```

Table 17

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

Both the above tables can be specified within a single **CTABLES** command:

```
ctables   /table sex by earngrp
          /table workmode by earngrp .
```

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** row percentages based on row totals are specified by: **[ROWPCT.COUNT]**.

3: Zero order tables (with row %)

```
ctables   /table sex by earngrp  [rowpct.count] .
```

Table 18

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

```
ctables   /table workmode by earngrp  [rowpct.count ] .
```

Table 19

		Gross earnings 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%

Again both the above tables can be specified within a single **CTABLES** command:

```
ctables   /table sex by earngrp  [rowpct.count]
          /table workmode by earngrp  [rowpct.count ] .
```

Putting the dependent variable in the columns and displaying row % within the categories of the independent variable makes it much easier to compare men and women.

Unlike **CROSSTABS** the **CTABLES** command can also display the row totals used (**base n**) on the same row as the percentages. This is done by specifying **TOTALS [COUNT]** as an additional element inside the square brackets. However, to display totals you need an additional line for each **/TABLE** specification:

```
ctables   /table sex by earngrp  [rowpct.count totals [count]] .
              /categories variables =  earngrp total=yes .
```

Table 20

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

```
ctables   /table workmode by earngrp  [rowpct.count totals [count]]
              /categories variables=  earngrp total=yes .
```

Table 21

		Gross earnings 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 with superfluous information. They can be further improved by getting rid of all the % signs and by changing the column headers from Row N % to % and from Count to n = 100%.

The elements in the **/TABLES** specification can also be extended by adding a label in double primes:

[ROWPCT.COUNT "%"] and **[COUNT "n = 100%"]**

The default formats are **integer** for cell counts and **one decimal place** for percentages.

```
ctables
  /table sex by earngrp  [rowpct.count f5.1 "%"  totals [count "n= 100%"]]
        /categories variables=  earngrp total=yes
  /table workmode by earngrp  [rowpct.count f5.1 "%" totals [count "n= 100%"]]
        /categories variables=  earngrp total=yes .
```

Table 22

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

Table 23

		Gross earnings 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 column totals for the earnings groups: it's easier to compare the earnings groups of men/women and full-time/part-time workers without them and also to calculate the **epsilons** (percentage point differences).

[**NB**: The epsilons were produced separately by copying the tables [3] into Excel, performing the calculations and then copying the epsilons back into Word]

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.

[3] For a fully worked example, see Appendix 2 in 3.2.1.7 Earnings differences 2009: Elaboration

Model for elaboration

Dependent variable	**Y = earngrp**	Gross earnings (<£6000, <£12000, £12000+)
Independent variable	**X = sex**	(Men, Women)
Test variable 1	**T_1 = workmode**	Hours of work (Part-time, Full-time)
Test variable 2	**T_2 = worktype**	Type of work (Non-manual, manual)

Zero order tables

1: X → Y sex → earngrp
2: T_1 → Y workmode → earngrp
3: T_2 → Y worktype → earngrp

```
ctables
  /table sex by earngrp
        [rowpct.count f5.1 "%"  totals [count "n= 100%"]]
  /categories variables=  earngrp total=yes .
```

Table 24:

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

```
ctables
  /table workmode by earngrp
        [rowpct.count f5.1 "%" totals [count "n= 100%"]]
  /categories variables=  earngrp total=yes .
```

Table 25:

		Gross earnings 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**	

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

Table 26:

		Gross annual earnings (if working) [3 groups]			
		<£6000	<£12000	£12000 +	Total
		%	%	%	n= 100%
Mode of work (Full-time: Part-time)	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**	

The independent variable **X** and test variables $\mathbf{T_1}$ or $\mathbf{T_2}$ can both be included in the same or table if **X** and $\mathbf{T_1}$ or **X** and $\mathbf{T_1}$ are linked with a **+** sign.

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

Table 27:

		Gross annual earnings (if working) [3 groups]			
		<£6000	<£12000	£12000 +	Total
		%	%	%	n= 100%
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
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

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

		Gross annual earnings (if working) [3 groups]			
		<£6000	<£12000	£12000 +	Total
		%	%	%	n= 100%
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
Type of work	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

First order[4] tables

1: X → Y . T_1 **sex → earngrp** (controlling for) **workmode**

2: X → Y . T_2 **sex → earngrp** (controlling for) **worktype**

To produce three-way contingency tables in **CTABLES**, the specification of variables is slightly different. One pair of variables has to linked by a **>** sign: the variable on the right of **>** will be nested within the categories of the variable on the left of it.

We are comparing men and women, so we need to keep **sex** nested within the categories of the test variables **workmode** and **worktype**.

1: T_1 > X by Y [Displays **sex** nested within **workmode**]:

```
ctables
        /table  workmode > sex by  earngrp
```

		<£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

2: X by T_1 > Y [Displays **sex** nested within **workmode**]:

```
ctables
        /VLABELS VARIABLES=sex earngrp workmode DISPLAY=NONE
        /TABLE sex by workmode > earngrp
        [ROWPCT.COUNT f5.1 "%"  TOTALS [COUNT "n= 100%"]]
        /CATEGORIES VARIABLES=  sex workmode earngrp  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 the original difference between men and women in gross earnings when **controlling for hours worked** (full-time = 30 or more hours per week).

[4] First order = one test variable

Taking a **criterion category** of £12,000 or more per annum as an indicator of "high earnings" the overall figure of **31.9%** of 1560 for the whole sample can be partitioned into **48.7%** of 874 men and **10.5%** of 686 women to yield an epsilon of **+38.2** points in favour of men. For **workmode** this is partitioned into **+22.4** points for those working part-time and **+32.8** for those working full-time. Thus, the original difference between men and women is reduced when controlling for hours worked.

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 let's do the same controlling for **worktype** (type of work)**: X → Y . T_2**

1: T_2 > X by Y [Displays **sex** nested within **worktype**]:

```
ctables /table worktype > sex by  earngrp .
```

		<£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

2: X by T_2 > Y

```
ctables
  /VLABELS VARIABLES=sex earngrp worktype DISPLAY=NONE
  /TABLE sex by worktype ] > earngrp
        [ROWPCT.COUNT f8.1 "%"    TOTALS [COUNT f8.0  "n= 100%"]]
 /CATEGORIES VARIABLES= sex worktype earngrp 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 the initial difference between of men and women in gross earnings when controlling for type of work.

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 partitioned into 48.9% of 854 men and 10.5% of 684 women. For **worktype** the 31.9% is partitioned 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 very 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 earngrp worktype DISPLAY=NONE
 /table sex by worktype > earngrp [ROWPCT.COUNT f8.1 "%" TOTALS[COUNT f8.0  "n= 100%"]]
 /CATEGORIES VARIABLES= sex worktype earngrp 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 partitioned into 49.5% of 543 men and 16.7% of 408 women. For type of work the 38.7% is partitioned 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 also 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 the logic is powerful and 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 earngrp to workage.

Variable Labels

Variable	Position	Label
sex	5	Sex of respondent
earngrp	14	Gross earnings of R (if working) [3 groups]
workmode	15	Mode of work
worktype	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 = workage1	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 logistic regression modelling (which can use ordinal and nominal variables) but this is way beyond the scope of this tutorial (and well above my statistical competence).

End of session: **3.2.1.1: Earnings differences – Elaboration**

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