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 20 May 2019: **Draft only**]

3.2.1.7 Earnings differences 2009: Elaboration

Previous session:

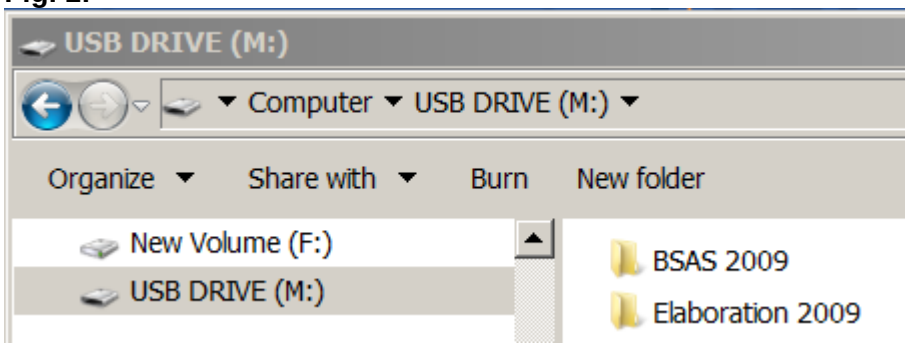
3.2.1.6 Earnings differences 2009: Extracting and saving selected variables

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

Exemplar test4 (Variables selected for our elaboration exercise: created in 3.2.1.6 above)

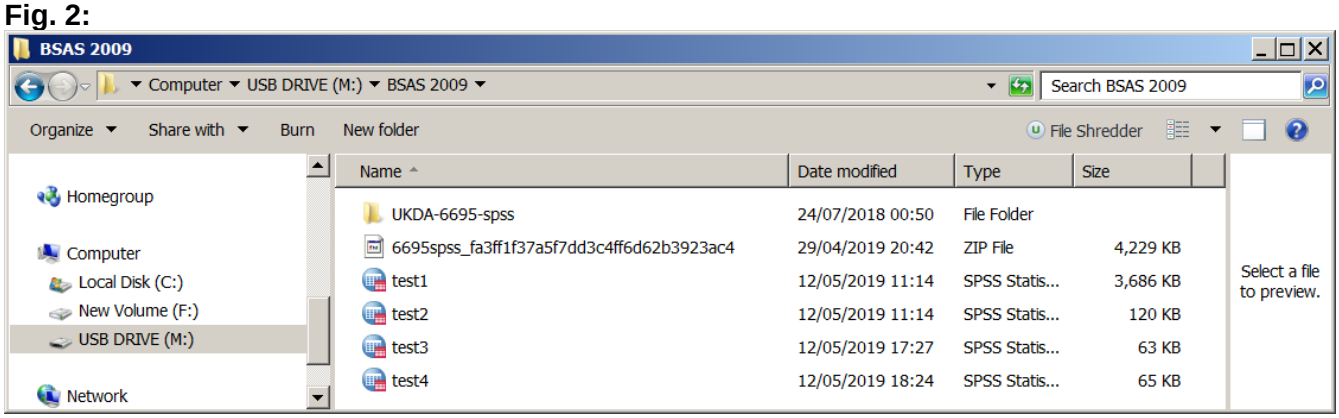
On **USB DRIVE M:**

Fig. 1:



Double click on BSAS 2009

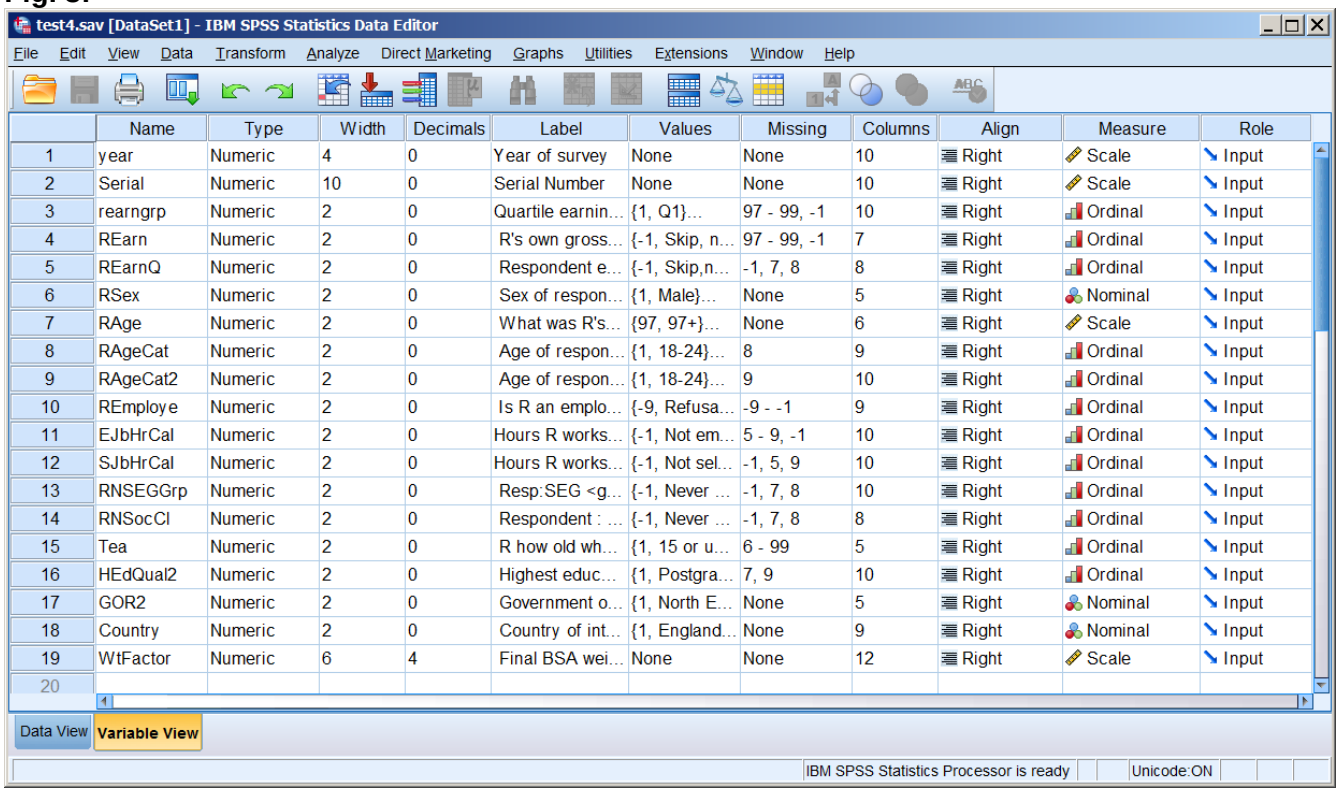
Fig. 2:



Double click on test4

[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

Fig. 3:



	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...	-1, 5, 9	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											

We have already produced frequency tables for the dependent, independent and test variables (See: 3.2.1.6 Earnings differences 2009: Extracting and saving selected variables pp. 21 - 24)

[**NB**: Cases with no earnings from paid work **have already been discarded**.]

The dependent variable **[REarn]** has too many categories*:* **[REarnQ]** has only four.

frequencies rearnq. [2]

Table 1:

REarnQ Respondent earnings quartiles

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 less than 11999	395	23.4	23.4	23.4
	2 12000- 19999	414	24.5	24.5	47.9
	3 20000- 31999	467	27.6	27.6	75.5
	4 32000 or more	413	24.5	24.5	100.0
	Total	1689	100.0	100.0	

[NB: No **£** sign available in SPSS 12]

The value labels of **[REarnQ]** denote specific earnings intervals. Rather than change the labels from earnings intervals to quartile groupings, it is better to create a new variable **[rearngrp]**.

To create a new variable **[rearngrp]** with four approximately equal groups and different value labels:

```
compute rearngrp = rearnq.
variable level rearngrp (ordinal).
variable labels rearngrp 'Quartile earnings group'.
value labels rearngrp 1 'Q1' 2 'Q2' 3 'Q3' 4 'Q4'.
```

[2] See Appendix 1 for a full listing of all the syntax in this tutorial: **lower case** is used throughout as it's quicker and easier.

In this session **[rearngrp]** will be the dependent variable.

frequencies rearngrp.

Table 2:

rearngrp Quartile earnings group

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Q1	395	23.4	23.4	23.4
	2 Q2	414	24.5	24.5	47.9
	3 Q3	467	27.6	27.6	75.5
	4 Q4	413	24.5	24.5	100.0
	Total	1689	100.0	100.0	

crosstabs rsex by rearngrp /cells count row.

Table 3:

RSex Sex of respondent * rearngrp Quartile earnings group Crosstabulation

			rearngrp Quartile earnings group				Total
			Q1	Q2	Q3	Q4	
RSex Sex of respondent	Male	Count	102	186	247	296	831
		% within RSex Sex of respondent	12.3%	22.4%	29.7%	35.6%	100.0%
	Female	Count	293	228	220	117	858
		% within RSex Sex of respondent	34.1%	26.6%	25.6%	13.6%	100.0%
Total		Count	395	414	467	413	1689
		% within RSex Sex of respondent	23.4%	24.5%	27.6%	24.5%	100.0%

Table 3 is our starting point for elaboration [3], a method which compares percentages of cases falling into (specified) categories of the dependent variable, within categories of independent and test variables. Independent and test variables should preferably be regrouped into two categories (dichotomised) or at most three (trichotomized).

Elaboration model

X → Y . T (the effect of **X** on **Y** controlling for **T**) where:

Y = Dependent variable
X = Independent variable
T_n = 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

[3] 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***, Basic Books 1968

We are now ready to produce:

a) **zero-order** (two-way) tables comparing the gross earnings (before tax and National Insurance) of:

1: Men and women
2: Categories within the selected test variables

b) **1st order** (three-way) contingency tables to see what happens to differences in gross earnings between men and women when controlling for a third (test) variable

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

There is no single variable denoting part- and full- time work. There are two separate variables, one for employees **[EjbHrCal]** and another **[SJbHrCal]** for the self-employed:

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

frequencies ejbhrcai sjbhrcai.

Table 4:

EJbHrCal Hours R works per week, including overtime [employee]. DV:Q1008

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 10-15 hours a week	82	2.4	5.1	5.1
	2 16-23 hours a week	183	5.3	11.5	16.6
	3 24-29 hours a week	90	2.6	5.6	22.3
	4 30 or more hours a week	1228	35.9	77.0	99.2
	5 Varies too much to say	7	0.2	0.4	99.7
	8 Don't know	4	0.1	0.3	99.9
	9 Refusal	1	0.0	0.1	100.0
	Total	1595	46.6	100.0	
Missing	-1 Not employee	1826	53.4		
Total		3421	100.0		

For statistical analysis codes **5**, **8** and **9** should be treated as missing:

Table 5:

SJbHrCal Hours R works per week, including overtime [self-employed]. DV:Q1010

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 10-15 hours a week	16	0.5	6.2	6.2
	2 16-23 hours a week	38	1.1	14.6	20.8
	3 24-29 hours a week	16	0.5	6.2	26.9
	4 30 or more hours a week	183	5.3	70.4	97.3
	5 Varies too much to say	5	0.1	1.9	99.2
	9 Refusal	2	0.1	0.8	100.0
	Total	260	7.6	100.0	
Missing	-1 Not self-employed	3161	92.4		
Total		3421	100.0		

For statistical analysis codes **5** and **9** should be treated as missing:

```
missing values ejbhrcai (-1 5 thru 9) sjbhrcai (-1 5 9).
frequencies   ejbhrcai sjbhrcai.
```

Table 6:

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 7:

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		

Variables **[EjbHrCal]** and **[SJbHrCal]** are mutually exclusive, but a new variable **[workhours]** combining information from both variables can be generated with:

```
compute workhours = max (ejbhrcai, sjbhrcai).
value labels workhours
        1 " 10-15 hours a week " 2 " 16-23 hours a week "
        3 " 24-29 hours a week " 4 " 30 or more hours a week ".
frequencies workhours.
```

Table 6:

workhours

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 10-15 hours a week	89	5.3	5.3	5.3
	2 16-23 hours a week	199	11.8	11.9	17.2
	3 24-29 hours a week	98	5.8	5.8	23.0
	4 30 or more hours a week	1292	76.5	77.0	100.0
	Total	1678	99.3	100.0	
Missing	System [4]	11	0.7		
Total		1689	100.0		

[4] When generating new numeric variables, SPSS by default assigns zero decimal places. Whenever missing values were found for **[EjbHrCal]** (-1, 5 thru 9) " Not employee" and **[SJbHrCal]** (-1 5 9) " Not self-employed" they have been automatically assigned the value **sysmis** in **[workhours]**

For elaboration purposes we need only two categories "Full-time" and "Part-time":

```
recode workhours (2 3=1)( 4 =2)(else = copy) into workmode.
variable labels workmode 'R works full- or part- time'.
value labels workmode 1  'Part-time' 2 'Full-time' .
frequencies workmode.
```

Table 7:

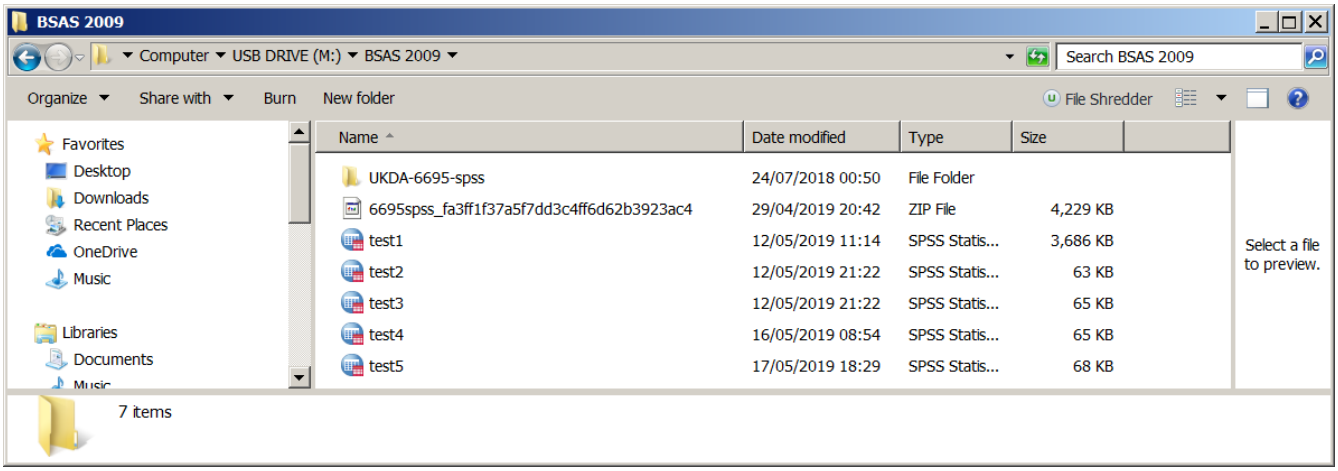
workmode (R works full- or part- time)

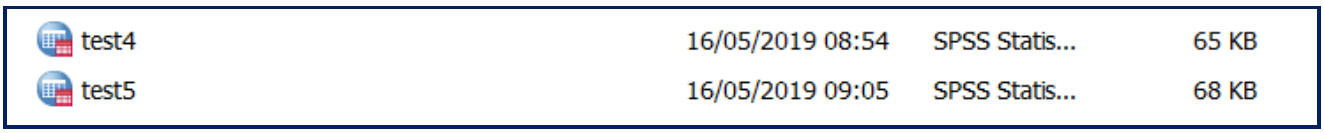
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Part-time	386	22.9	23.0	23.0
	2 Full-time	1292	76.5	77.0	100.0
	Total	1678	99.3	100.0	
Missing	System	11	0.7		
Total		1689	100.0		

All the variables we need have now been extracted, generated and checked. **Save the file** as **test5.sav** on USB Drive **M:**

```
save outfile = 'M:\BSAS 2009\test5.sav'
  /keep year serial
        rearn rearnq rearngrp
        rsex rage ragecat ragecat2
        remploye workhours workmode
        ejbhrcai sjbhrcai rnseggrp rnsoccl
        tea hedqual2
        gor2 country
        wtfactor.
```

File test5 is saved to folder BSAS 2009 on USB Drive **M:**





Elaboration

In this and following sessions the dependent variable will be **[rearngrp]**

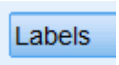
Status	**Name**	**Label**
Y = Dependent	**rearngrp**	[Gross annual earnings: 4 groups based on quartiles]
X = Independent	**rsex**	[Men, Women]
T = Test	**workmode**	[Working full-time or part-time]

Our first test variable is **[workmode]** "R works full -or part-time?".

Frequencies	**Y, X, T**	**frequencies** rearngrp rsex workmode **.**
Zero order tables[5]	**X →Y** **T → Y**	**crosstabs** rsex workmode by rearngrp .
1st order table[6]	**X → Y . T**	**crosstabs** rsex **by** rearngrp **by** workmode **.**

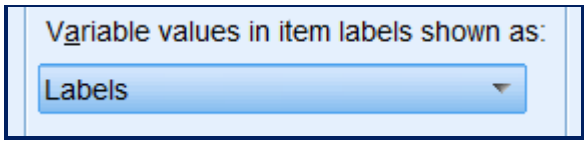
Cluttered output

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, SPSS has been set to display variable values as 

only.

Edit >> Options >> Output



[5] SPSS command **CROSSTABS** produces very cluttered output: **CTABLES** is more complex to use, but the output is far less cluttered. (See: 3.2.1.4 Elaboration 4 (Income differences 2009 – 2014 CTABLES)

Initial frequency counts

frequencies rearngrp, rsex, workmode **.**

Table 8:

rearngrp Quartile earnings group

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Q1	395	23.4	23.4	23.4
	Q2	414	24.5	24.5	47.9
	Q3	467	27.6	27.6	75.5
	Q4	413	24.5	24.5	100.0
	Total	1689	100.0	100.0	

Table 9:

RSex Sex of respondent

		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	

Table 10:

workmode R works full- or part- time

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Part-time	386	22.9	23.0	23.0
	Full-time	1292	76.5	77.0	100.0
	Total	1678	99.3	100.0	
Missing	System	11	0.7		
Total		1689	100.0		

Zero-order tables

1: Effect of sex on earnings

crosstabs rsex by rearngrp .

Table 11:

RSex Sex of respondent * rearngrp Quartile earnings group Crosstabulation

Count

		rearngrp Quartile earnings group				Total
		Q1	Q2	Q3	Q4	
RSex Sex of respondent	Male	102	186	247	296	831
	Female	293	228	220	117	858
Total		395	414	467	413	1689

Because there are almost equal numbers of men and women it's clear that the gradients for quartile earnings run in opposite directions.

It's easier to compare percentages than cell counts:

crosstabs rsex by rearngrp /cells count row.

Table 12:

RSex Sex of respondent * rearngrp Quartile earnings group Crosstabulation

			rearngrp Quartile earnings group				
			Q1	Q2	Q3	Q4	Total
RSex Sex of respondent	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%

It's now easier to compare men and women, but the table is a bit cluttered.

crosstabs rsex by rearngrp /cells row.

Table 13:

RSex Sex of respondent * rearngrp Quartile earnings group Crosstabulation

% within RSex Sex of respondent

		rearngrp Quartile earnings group				
		Q1	Q2	Q3	Q4	Total
RSex Sex of respondent	Male	12.3%	22.4%	29.7%	35.6%	100.0%
	Female	34.1%	26.6%	25.6%	13.6%	100.0%
Total		23.4%	24.5%	27.6%	24.5%	100.0%

It's now even easier to compare men and women, but we've now lost the base **n** for percentages.

2: Effect of workmode on earnings

crosstabs workmode by rearngrp .

Table 14:

workmode R works full- or part- time * rearngrp Quartile earnings group Crosstabulation

Count

		rearngrp Quartile earnings group				
		Q1	Q2	Q3	Q4	Total
workmode R works full- or part- time	Part-time	241	86	41	18	386
	Full-time	150	328	423	391	1292
Total		391	414	464	409	1678

Even with raw counts, the gradients for quartile earnings seem to run in opposite directions.

crosstabs workmode by rearngrp /cells count row.

Table 15:

workmode R works full- or part- time * rearngrp Quartile earnings group Crosstabulation

			rearngrp Quartile earnings group				Total
			Q1	Q2	Q3	Q4	
workmode R works full- or part- time	Part-time	Count	241	86	41	18	386
		% within workmode	62.4%	22.3%	10.6%	4.7%	100.0%
	Full-time	Count	150	328	423	391	1292
		% within workmode	11.6%	25.4%	32.7%	30.3%	100.0%
Total		Count	391	414	464	409	1678
		% within workmode	23.3%	24.7%	27.7%	24.4%	100.0%

You can compare the percentages, but the table is a bit cluttered.

crosstabs workmode by rearngrp /cells row.

Table 16:

workmode R works full- or part- time * rearngrp Quartile earnings group Crosstabulation

% within workmode R works full- or part- time

		rearngrp Quartile earnings group				Total
		Q1	Q2	Q3	Q4	
workmode R works full- or part- time	Part-time	62.4%	22.3%	10.6%	4.7%	100.0%
	Full-time	11.6%	25.4%	32.7%	30.3%	100.0%
Total		23.3%	24.7%	27.7%	24.4%	100.0%

It's easier to compare people working full-time or part-time, but there is now no base **n** for percentages.

1st order tables

1: Effect of sex on earnings, controlling for workmode

crosstabs rsex by rearngrp by workmode**.**

Table17:

RSex Sex of respondent * rearngrp Quartile earnings group * workmode R works full- or part- time Crosstabulation

Count

workmode R works full- or part- time			rearngrp Quartile earnings group				Total
			Q1	Q2	Q3	Q4	
Part-time	RSex Sex of respondent	Male	44	23	9	13	89
		Female	197	63	32	5	297
	Total		241	86	41	18	386
Full-time	RSex Sex of respondent	Male	56	163	237	279	735
		Female	94	165	186	112	557
	Total		150	328	423	391	1292
Total	RSex Sex of respondent	Male	100	186	246	292	824
		Female	291	228	218	117	854
	Total		391	414	464	409	1678

Useful for checking the structure of the sample, but difficult to interpret.

crosstabs rsex by rearngrp by workmode /cells count row**.**

Table18:

RSex Sex of respondent * rearngrp Quartile earnings group * workmode R works full- or part-time Crosstabulation

workmode R works full- or part- time				rearngrp Quartile earnings group				Total
				Q1	Q2	Q3	Q4	
Part-time	RSex	Male	Count	44	23	9	13	89
			% within RSex	49.4%	25.8%	10.1%	14.6%	100.0%
		Female	Count	197	63	32	5	297
			% within RSex	66.3%	21.2%	10.8%	1.7%	100.0%
	Total		Count	241	86	41	18	386
			% within RSex	62.4%	22.3%	10.6%	4.7%	100.0%
Full-time	RSex	Male	Count	56	163	237	279	735
			% within RSex	7.6%	22.2%	32.2%	38.0%	100.0%
		Female	Count	94	165	186	112	557
			% within RSex	16.9%	29.6%	33.4%	20.1%	100.0%
	Total		Count	150	328	423	391	1292
			% within RSex	11.6%	25.4%	32.7%	30.3%	100.0%
Total	RSex	Male	Count	100	186	246	292	824
			% within RSex	12.1%	22.6%	29.9%	35.4%	100.0%
		Female	Count	291	228	218	117	854
			% within RSex	34.1%	26.7%	25.5%	13.7%	100.0%
	Total		Count	391	414	464	409	1678
			% within RSex	23.3%	24.7%	27.7%	24.4%	100.0%

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.

crosstabs rsex by rearngrp by workmode /cells row**.**

Table19:

RSex Sex of respondent * rearngrp Quartile earnings group * workmode R works full- or part- time Crosstabulation

% within RSex Sex of respondent

workmode R works full- or part- time			rearngrp Quartile earnings group				
			Q1	Q2	Q3	Q4	Total
Part-time	RSex Sex of respondent	Male	49.4%	25.8%	10.1%	14.6%	100.0%
		Female	66.3%	21.2%	10.8%	1.7%	100.0%
	Total		62.4%	22.3%	10.6%	4.7%	100.0%
Full-time	RSex Sex of respondent	Male	7.6%	22.2%	32.2%	38.0%	100.0%
		Female	16.9%	29.6%	33.4%	20.1%	100.0%
	Total		11.6%	25.4%	32.7%	30.3%	100.0%
Total	RSex Sex of respondent	Male	12.1%	22.6%	29.9%	35.4%	100.0%
		Female	34.1%	26.7%	25.5%	13.7%	100.0%
	Total		23.3%	24.7%	27.7%	24.4%	100.0%

This table is easier to interpret, but is still overly cluttered. It could be manually edited to get rid of the **%** signs in the body of the table and substitute base **n** for **100%** in the column headings, but that would take a long time and is possibly error-prone.

Custom Tables

To obtain tables that are less cluttered, and thus much easier to interpret, SPSS has a facility for producing **Custom Tables**.

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 it can be modified within the program. The default output can be very sparse for tabulating a single variable, but at least the frequency distribution doesn't display totally unnecessary cumulative percentages for nominal variables.

Within the **CTABLES** command, tables must be specified one at a time, e.g:

```
CTABLES
        /TABLE <variable>
```

1: Initial frequency counts

```
ctables /table rearngrp
        /table rsex
        /table workmode.
```

Table20:

		Count
rearngrp Quartile earnings group	Q1	395
	Q2	414
	Q3	467
	Q4	413

Table21:

		Count
RSex Sex of respondent	Male	831
	Female	858

Table22:

		Count
workmode (R works full- or part- time)	Part-time	386
	Full-time	1292

Zero-order tables

X → Y **Effect of sex on earnings**

```
ctables   /table rsex by rearngrp
          /table workmode by rearngrp.
```

Table23:

		rearngrp Quartile earnings group			
		Q1	Q2	Q3	Q4
		Count	Count	Count	Count
RSex Sex of respondent	Male	102	186	247	296
	Female	293	228	220	117

T_1 → Y **Effect of workmode on earnings**

Table24:

		rearngrp Quartile earnings group			
		Q1	Q2	Q3	Q4
		Count	Count	Count	Count
workmode (R works full- or part- time)	Part-time	241	86	41	18
	Full-time	150	328	423	391

[Note there are no column totals in the above tables.]

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]**.

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

Table25:

		rearngrp Quartile earnings group			
		Q1	Q2	Q3	Q4
		Row N %	Row N %	Row N %	Row N %
RSex Sex of respondent	Male	12.3%	22.4%	29.7%	35.6%
	Female	34.1%	26.6%	25.6%	13.6%

Table26:

		rearngrp Quartile earnings group			
		Q1	Q2	Q3	Q4
		Row N %	Row N %	Row N %	Row N %
workmode (R works full- or part- time)	Part-time	62.4%	22.3%	10.6%	4.7%
	Full-time	11.6%	25.4%	32.7%	30.3%

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

```
ctables   /table sex by rearngrp3 [rowpct.count totals [count]]
```

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

```
/categories variables = rearngrp3 total=yes.
```

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

Table27:

		rearngrp Quartile earnings group				
		Q1	Q2	Q3	Q4	Total
		Row N %	Row N %	Row N %	Row N %	Count
RSex Sex of respondent	Male	12.3%	22.4%	29.7%	35.6%	831
	Female	34.1%	26.6%	25.6%	13.6%	858

Table28:

		rearngrp Quartile earnings group				
		Q1	Q2	Q3	Q4	Total
		Row N %	Row N %	Row N %	Row N %	Count
workmode (R works full- or part- time)	Part-time	62.4%	22.3%	10.6%	4.7%	386
	Full-time	11.6%	25.4%	32.7%	30.3%	1292

These tables are now much easier to read and interpret, but they are still slightly cluttered. They can be further improved by changing the column headers from **Row N %** to a simple **%** and **Count** to **(n=100%)** and then by getting rid of all the **%** signs in the body of the table.

To do this, the elements in the /TABLES specification need to be extended by adding labels 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 rsex by rearngrp  [rowpct.count f5.1 "%"  totals [count "n= 100%"]]
        /categories variables=  rearngrp total=yes
  /table workmode by rearngrp  [rowpct.count f5.1 "%" totals [count "n= 100%"]]
       /categories variables=  rearngrp total=yes .
```

Table29:

		rearngrp Quartile earnings group				
		Q1	Q2	Q3	Q4	Total
		%	%	%	%	n= 100%
RSex Sex of respondent	Male	12.3	22.4	29.7	35.6	831
	Female	34.1	26.6	25.6	13.6	858
	Epsilon [7]	**-21.9**	**-4.2**	**+4.1**	**+22.0**	

I wonder if **CTABLES** can be modified to produce tables with epsilons?

[7] The epsilons (percentage point difference between men and women in each column) were produced separately by copying the body of the table into Excel, subtracting the female percentage from the male percentage, then copying an extract from Excel back into Word] See Appendix 2 for a worked example.

Table30:

		rearngrp Quartile earnings group				
		Q1	Q2	Q3	Q4	Total
		%	%	%	%	n= 100%
workmode (R works full- or part- time)	Part-time	62.4	22.3	10.6	4.7	386
	Full-time	11.6	25.4	32.7	30.3	1292
	Epsilon	**50.8**	**-3.1**	**-22.1**	**-25.6**	

The above tables do not have column 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).

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

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

```
ctables
   /table rsex [c] + workmode [c] by rearngrp [c]
          [rowpct.count  f8.1 "%" totals[validn f8.0  "n= 100%"]]
   /categories variables= sex workmode rearngrp total=yes position=after.
```

Table31:

		rearngrp Quartile earnings group				
		Q1	Q2	Q3	Q4	Total
		%	%	%	%	n= 100%
RSex Sex of respondent	Male	12.3	22.4	29.7	35.6	831
	Female	34.1	26.6	25.6	13.6	858
workmode (R works full- or part- time)	Part-time	62.4	22.3	10.6	4.7	386
	Full-time	11.6	25.4	32.7	30.3	1292
	Total	23.3	24.7	27.7	24.4	1678

First order nested 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 are three ways of producing such tables:

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

1: X > T_1 by Y

```
*Nest workmode within sex.
ctables
          /table rsex > workmode by rearngrp
           [rowpct.count  f8.1 "%" totals[validn f8.0  "n= 100%"]]
   /categories variables= sex workmode rearngrp total=yes position=after.
```

Table32:

				rearngrp Quartile earnings group				
				Q1	Q2	Q3	Q4	Total
				%	%	%	%	n= 100%
RSex Sex of respondent	Male	workmode	Part-time	49.4	25.8	10.1	14.6	89
			Full-time	7.6	22.2	32.2	38.0	735
			Total	12.1	22.6	29.9	35.4	824
	Female	workmode	Part-time	66.3	21.2	10.8	1.7	297
			Full-time	16.9	29.6	33.4	20.1	557
			Total	34.1	26.7	25.5	13.7	854

2: T_1 > X by Y

```
*Nest sex within workmode.
ctables
        /table  workmode > rsex by  rearngrp
         [rowpct.count  f8.1 "%" totals[validn f8.0  "n= 100%"]]
   /categories variables= rsex workmode rearngrp total=yes position=after.
```

Table33:

				rearngrp Quartile earnings group				
				Q1	Q2	Q3	Q4	Total
				%	%	%	%	n= 100%
workmode (R works full- or part- time)	Part-time	RSex Sex of respondent	Male	49.4	25.8	10.1	14.6	89
			Female	66.3	21.2	10.8	1.7	297
			Total	62.4	22.3	10.6	4.7	386
	Full-time	RSex Sex of respondent	Male	7.6	22.2	32.2	38.0	735
			Female	16.9	29.6	33.4	20.1	557
			Total	11.6	25.4	32.7	30.3	1292
	Total	RSex Sex of respondent	Male	12.1	22.6	29.9	35.4	824
			Female	34.1	26.7	25.5	13.7	854
			Total	23.3	24.7	27.7	24.4	1678

3: **X by T_1 > Y**

```
ctables
        /vlabels variables=rsex rearngrp workmode display=none
        /table rsex by workmode > rearngrp
        [rowpct.count f5.1 "%"  totals [count "n= 100%"]]
        /categories variables=  rsex workmode rearngrp  total=yes position=after.
```

Table34:

	Part-time					Full-time					Total				
	Q1	Q2	Q3	Q4	Total	Q1	Q2	Q3	Q4	Total	Q1	Q2	Q3	Q4	Total
	%	%	%	%	n= 100%	%	%	%	%	n= 100%	%	%	%	%	n= 100%
Male	49.4	25.8	10.1	14.6	89	7.6	22.2	32.2	38.0	735	12.1	22.6	29.9	35.4	824
Female	66.3	21.2	10.8	1.7	297	16.9	29.6	33.4	20.1	557	34.1	26.7	25.5	13.7	854
Total	62.4	22.3	10.6	4.7	386	11.6	25.4	32.7	30.3	1292	23.3	24.7	27.7	24.4	1678

Perhaps not!

To get separate tables for **workmode**:

```
temporary.
select if workmode = 1.
ctables
        /vlabels variables=rsex rearngrp display=none
        /table rsex by rearngrp
        [rowpct.count f5.1 "%"  totals [count "n= 100%"]]
        /categories variables=  rsex rearngrp  total=yes position=after.
```

Table35: Part time workers

	Q1	Q2	Q3	Q4	Total
	%	%	%	%	n= 100%
Male	49.4	25.8	10.1	14.6	89
Female	66.3	21.2	10.8	1.7	297
Total	62.4	22.3	10.6	4.7	386

Epsilon	**-9.3**	**-7.4**	**-1.2**	**17.9**

```
temporary.
select if workmode = 2.
ctables
        /vlabels variables=rsex rearngrp display=none
        /table rsex by rearngrp
        [rowpct.count f5.1 "%"  totals [count "n= 100%"]]
        /categories variables=  rsex rearngrp  total=yes position=after.
```

Table36: Full time workers

	Q1	Q2	Q3	Q4	Total
	%	%	%	%	n= 100%
Male	7.6	22.2	32.2	38.0	735
Female	16.9	29.6	33.4	20.1	557
Total	11.6	25.4	32.7	30.3	1292

Epsilon	**-16.9**	**4.6**	**-0.7**	**12.9**

We should now discard the part-time workers and restrict future analysis to those working full time.

```
select if workmode = 2.
save outfile = 'M:\BSAS 2009\test6.sav' .
```

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
test2	12/05/2019 21:22	SPSS Statis...	63 KB
test3	12/05/2019 21:22	SPSS Statis...	65 KB
test4	16/05/2019 08:54	SPSS Statis...	65 KB
test5	17/05/2019 18:29	SPSS Statis...	68 KB
test6	20/05/2019 05:25	SPSS Statis...	54 KB

End of session: **3.2.1.7: Earnings differences 2009: Elaboration**

Back to: 3.2.1.6 Earnings differences 2009: Extracting and saving variables
Back to: 3.2: Three (or more) variables

Appendix 1: SPSS syntax for 3.2.1.7

```
* Encoding: UTF-8.
frequencies rearnq.

compute rearngrp = rearnq.
variable level rearngrp (ordinal).
variable labels rearngrp 'Quartile earnings group'.
value labels rearngrp 1 'Q1' 2 'Q2' 3 'Q3' 4 'Q4'.

frequencies rearngrp.

crosstabs rsex by rearngrp /cells count row.

frequencies   EJbHrCal SJbHrCal.
missing values EJbHrCal (-1 5 thru 9) SJbHrCal (-1 5 9).
frequencies   EJbHrCal SJbHrCal.

compute  workhours = max (EJbHrCal, SJbHrCal).
missing values workhours (5 8 9).
value labels workhours
        1 " 10-15 hours a week " 2 " 16-23 hours a week "
        3 " 24-29 hours a week " 4 " 30 or more hours a week "
        5 " Varies too much to say" 8 " Don't know" 9 " Refusal".
frequencies workhours.

recode workhours (2 3=1)( 4 =2)(else = copy) into workmode.
variable labels workmode (R works full- or part- time).
value labels workmode
          1  'Part-time' 2 'Full-time' 5 "Varies too much to say"
          8 "Don't know" 9 "Refusal".
missing values workmode (-1, 5 thru 9).
frequencies workmode.

save outfile =  'M:\BSAS 2009\test5.sav'
   /keep year serial
        rearn rearnq rearngrp
        rsex rage ragecat ragecat2
        remploye workhours workmode
        ejbhrcai sjbhrcai rnseggrp rnsoccl
        tea hedqual2
        gor2 country
        wtfactor.

frequencies     rearngrp, rsex, workmode .
crosstabs rsex by rearngrp .
crosstabs rsex by rearngrp /cells count row.
crosstabs rsex by rearngrp /cells row.

crosstabs workmode by rearngrp .
crosstabs workmode by rearngrp /cells count row.
crosstabs workmode by rearngrp /cells row.

crosstabs rsex by rearngrp by workmode.
crosstabs rsex by rearngrp by workmode /cells count row.
crosstabs rsex by rearngrp by workmode /cells row.

ctables /table rearngrp
        /table rsex
        /table workmode.
```

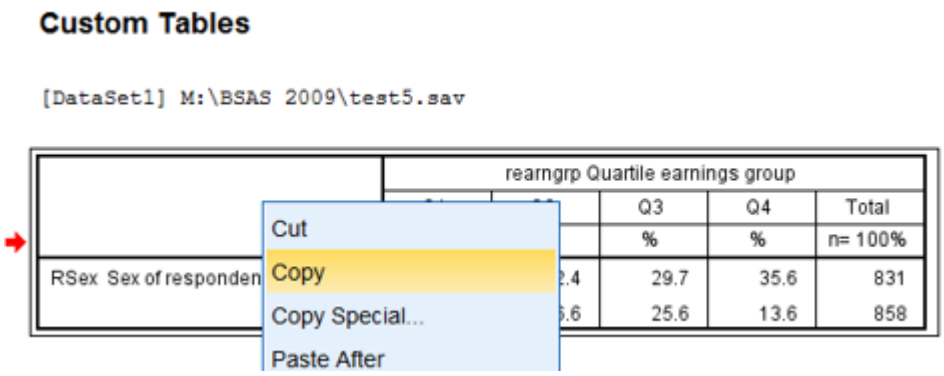
```
ctables    /table rsex by rearngrp
             /table workmode by rearngrp.

ctables    /table sex by rearngrp  [rowpct.count]
             /table workmode by rearngrp  [rowpct.count ].
ctables    /table sex by rearngrp3  [rowpct.count totals [count]].
ctables    /table rsex by rearngrp [rowpct.count totals [count]]
                 /categories variables=  rearngrp total=yes
           /table workmode by rearngrp  [rowpct.count totals [count]]
                 /categories variables =  rearngrp total=yes.

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

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

ctables
          /table rsex > workmode by rearngrp
           [rowpct.count  f8.1 "%" totals[validn f8.0  "n= 100%"]]
         /categories variables = sex workmode rearngrp
          total=yes position=after.

ctables
          /table workmode > rsex by  rearngrp
           [rowpct.count  f8.1 "%" totals[validn f8.0  "n= 100%"]]
          /categories variables= rsex workmode rearngrp
          total=yes position=after.

ctables
          /table variables =rsex rearngrp workmode display=none
          /table rsex by workmode > rearngrp
          [rowpct.count f5.1 "%"  totals [count "n= 100%"]]
         /categories variables =  rsex workmode rearngrp  total=yes position=after.

temporary.
select if workmode = 1.
ctables
          /categories variables=rsex rearngrp display=none
          /categories rsex by rearngrp
          [rowpct.count f5.1 "%"  totals [count "n= 100%"]]
          /categories variables =  rsex rearngrp  total=yes position=after.

temporary.
select if workmode = 2.
ctables
          /categories variables =rsex rearngrp display=none
          /table rsex by rearngrp
          [rowpct.count f5.1 "%"  totals [count "n= 100%"]]
          /categories variables=  rsex rearngrp  total=yes position=after.

select if workmode = 2.
save outfile = 'M:\BSAS 2009\test6.sav' .
```

Appendix 2: Epsilons via Excel

Table as in viewer:

		rearngrp Quartile earnings group				
		Q1	Q2	Q3	Q4	Total
		%	%	%	%	n= 100%
RSex Sex of respondent	Male	12.3	22.4	29.7	35.6	831
	Female	34.1	26.6	25.6	13.6	858

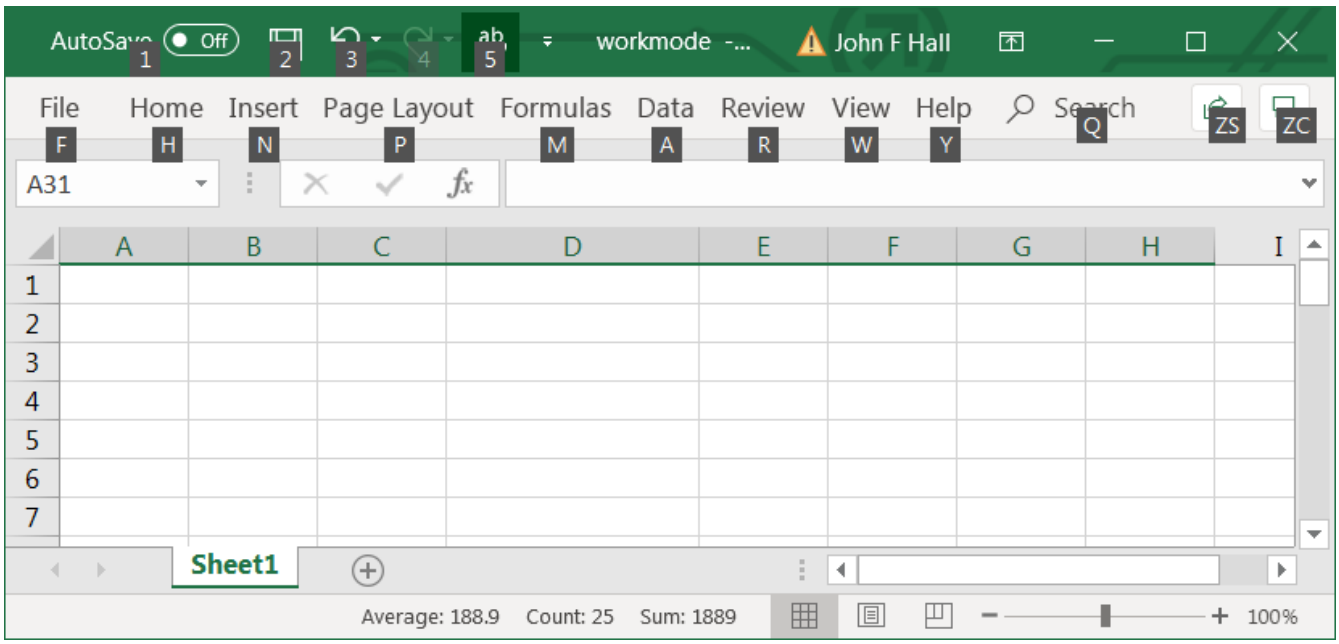
Right click the table



Double click on the table to

Click on Copy

Open a new Excel sheet:



With the cursor in cell A1 press Ctrl + V or Right click > Paste

	A	B	C	D	E	F	G
1			rearngrp Quartile earnings group				
2			Q1	Q2	Q3	Q4	Total
3			%	%	%	%	n= 100%
4	RSex Sex of respondent	Male	12.3	22.4	29.7	35.6	831
5		Female	34.1	26.6	25.6	13.6	858
6							
7							

Write "Epsilon" in B7:

B7 | Epsilon

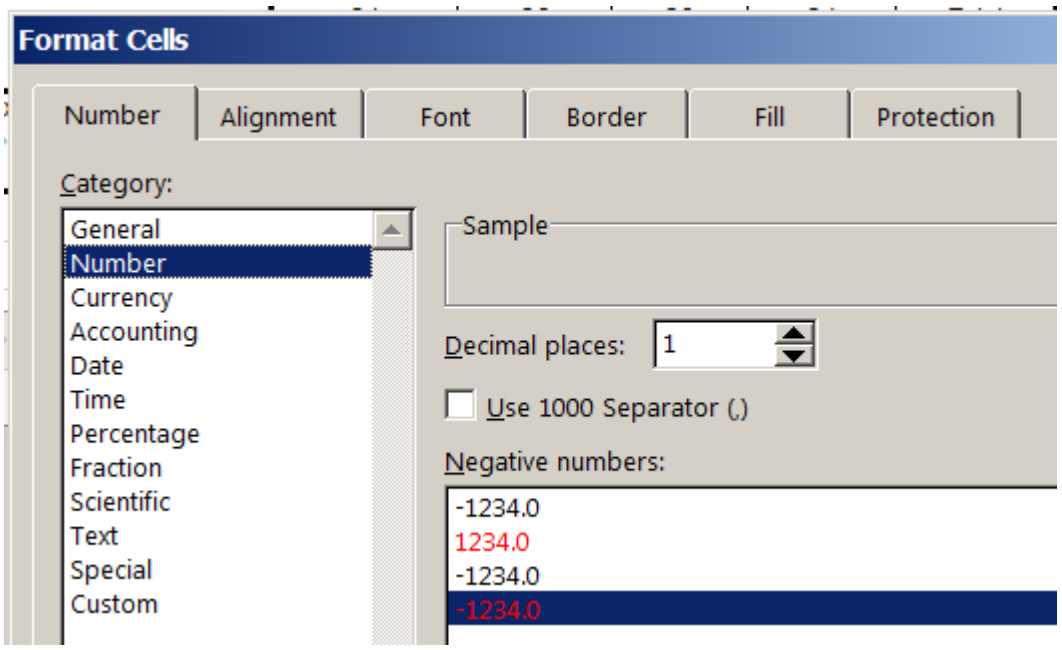
	A	B	C	D	E	F	G
1			rearngrp Quartile earnings group				
2			Q1	Q2	Q3	Q4	Total
3			%	%	%	%	n= 100%
4	RSex Sex of respondent	Male	12.3	22.4	29.7	35.6	831
5		Female	34.1	26.6	25.6	13.6	858
6							
7		Epsilon					

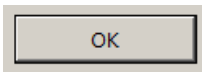
Highlight cell C7

	A	B	C	D	E	F	G
1			rearngrp Quartile earnings group				
2			Q1	Q2	Q3	Q4	Total
3			%	%	%	%	n= 100%
4	RSex Sex of respondent	Male	12.3	22.4	29.7	35.6	831
5		Female	34.1	26.6	25.6	13.6	858
6							
7		Epsilon					

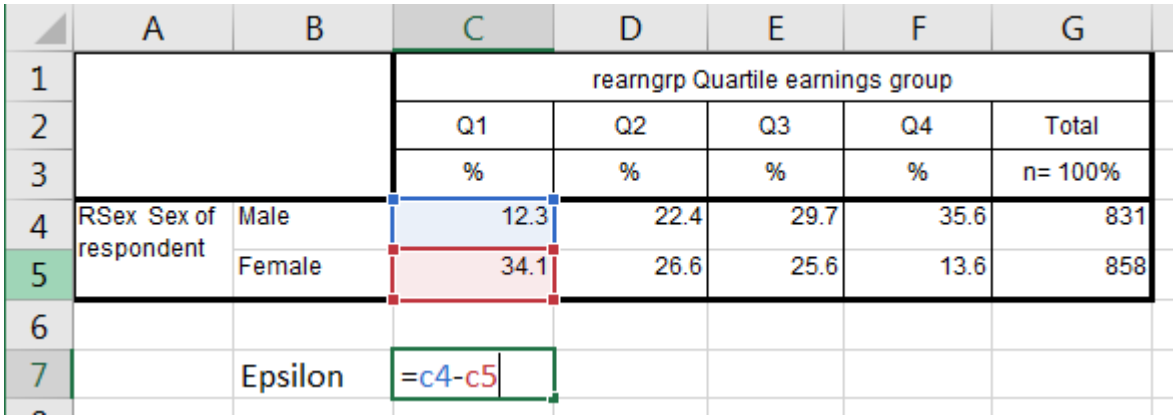
Right click >> Format cells

Click on Number choose Negative numbers 1234.0 change Decimal places from **2** to **1**



Click 

In cell C7 write **=c4-c5**



	A	B	C	D	E	F	G
1			rearngrp Quartile earnings group				
2			Q1	Q2	Q3	Q4	Total
3			%	%	%	%	n= 100%
4	RSex Sex of respondent	Male	12.3	22.4	29.7	35.6	831
5		Female	34.1	26.6	25.6	13.6	858
6							
7		Epsilon	=c4-c5				

Enter

		rearngrp Quartile earnings group				
		Q1	Q2	Q3	Q4	Total
		%	%	%	%	n= 100%
RSex Sex of respondent	Male	12.3	22.4	29.7	35.6	831
	Female	34.1	26.6	25.6	13.6	858

Epsilon -21.9

Highlight cell C7

Ctrl C or Right click >> Copy

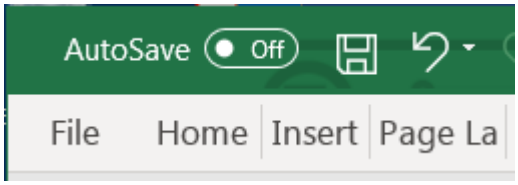
Highlight cells D7 to F7

	A	B	C	D	E	F	G
1			rearngrp Quartile earnings group				
2			Q1	Q2	Q3	Q4	Total
3			%	%	%	%	n= 100%
4	RSex Sex of respondent	Male	12.3	22.4	29.7	35.6	831
5		Female	34.1	26.6	25.6	13.6	858
6							
7		Epsilon	-21.9				

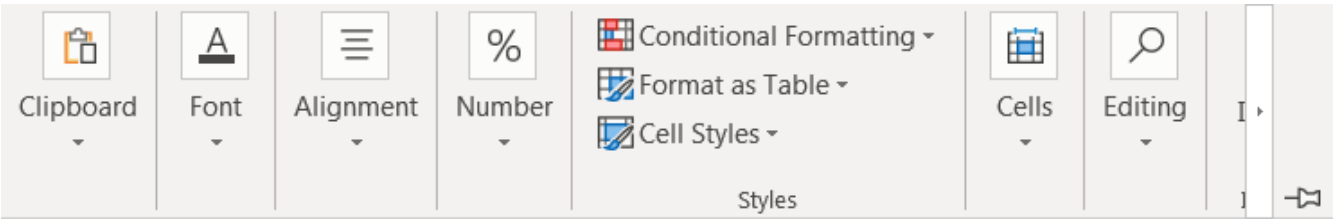
Ctrl V or Right click > Paste

		rearngrp Quartile earnings group				
		Q1	Q2	Q3	Q4	Total
		%	%	%	%	n= 100%
RSex Sex of respondent	Male	12.3	22.4	29.7	35.6	831
	Female	34.1	26.6	25.6	13.6	858
	Epsilon	-21.9	-4.2	4.1	22.0	

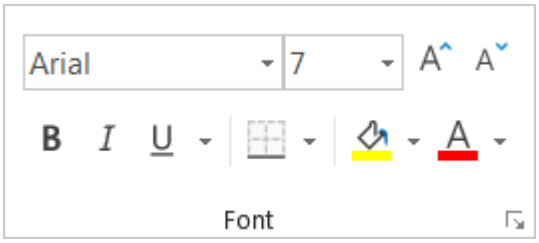
You can change the font, size and colour to taste:



Click **Home**



Click **Font** and change size to 10.



		rearngrp Quartile earnings group				
		Q1	Q2	Q3	Q4	Total
		%	%	%	%	n= 100%
RSex Sex of respondent	Male	12.3	22.4	29.7	35.6	831
	Female	34.1	26.6	25.6	13.6	858
	Epsilon	-21.9	-4.2	4.1	22.0	

Excel doesn't seem to have a facility for leading + signs, but you can change the table in **Word**

		rearngrp Quartile earnings group				
		Q1	Q2	Q3	Q4	Total
		%	%	%	%	n= 100%
RSex Sex of respondent	Male	12.3	22.4	29.7	35.6	831
	Female	34.1	26.6	25.6	13.6	858
	Epsilon	-21.9	-4.2	+4.1	+22.0	

Bolden the epsilons:

		rearngrp Quartile earnings group				
		Q1	Q2	Q3	Q4	Total
		%	%	%	%	n= 100%
RSex Sex of respondent	Male	12.3	22.4	29.7	35.6	831
	Female	34.1	26.6	25.6	13.6	858
	Epsilon	**-21.9**	**-4.2**	**+4.1**	**+22.0**	

Change the colour of the positive epsilons:

		rearngrp Quartile earnings group				
		Q1	Q2	Q3	Q4	Total
		%	%	%	%	n= 100%
RSex Sex of respondent	Male	12.3	22.4	29.7	35.6	831
	Female	34.1	26.6	25.6	13.6	858
	Epsilon	**-21.9**	**-4.2**	**+4.1**	**+22.0**	