

3.2: Three (or more) variables**3.2.1 Elaboration****3.2.1.3 Elaboration 3 (Income differences 2009 – 2014: CROSSTABS)***(Zero order two-way contingency tables)***Previous tutorial:** [3.2.1.2 Elaboration 2 \(Income differences BSA 2009 - 2014\)](#)

This set of tutorials uses data from the [British Social Attitudes](#) series to explore the following research questions.

- 1: Is there a difference between the earnings (from paid work) of men and women?
- 2: What other variables might account for differences in earnings?
- 3: What effect do they have by themselves?
- 4: What happens to any differences in earnings between men and women when controlling for these other variables?

Data source: [British Social Attitudes 1983 to 2014: Cumulative SPSS file](#)**Dependent variable Y:** Earnings from paid work**Independent variable X:** Sex of respondent

Test variables T:

- Employee/Self-employed
- Public/private sector
- Hours worked
- Social class of job

- Age completed full time education
- Qualifications from school
- Qualifications post-school
- Highest qualification

- Age last birthday
- Age group

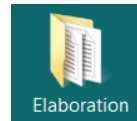
- Region
- Country

Tutorial [3.2.1.2 Elaboration 2 \(Income differences BSA 2009 - 2014\)](#) deals with:

- Selection and extraction of relevant variables
- Checking with SPSS command **FREQUENCIES**
- Editing of metadata where necessary or appropriate
- Restriction of the sample to include only respondents
 - a) from 2009 to 2014
 - b) for whom earnings (from paid work) are not missing
- Creation of two new derived variables for earnings and mode of work
- Saving the data to file **bsa89-14_elab4.sav** (19 variables, N = 8789),

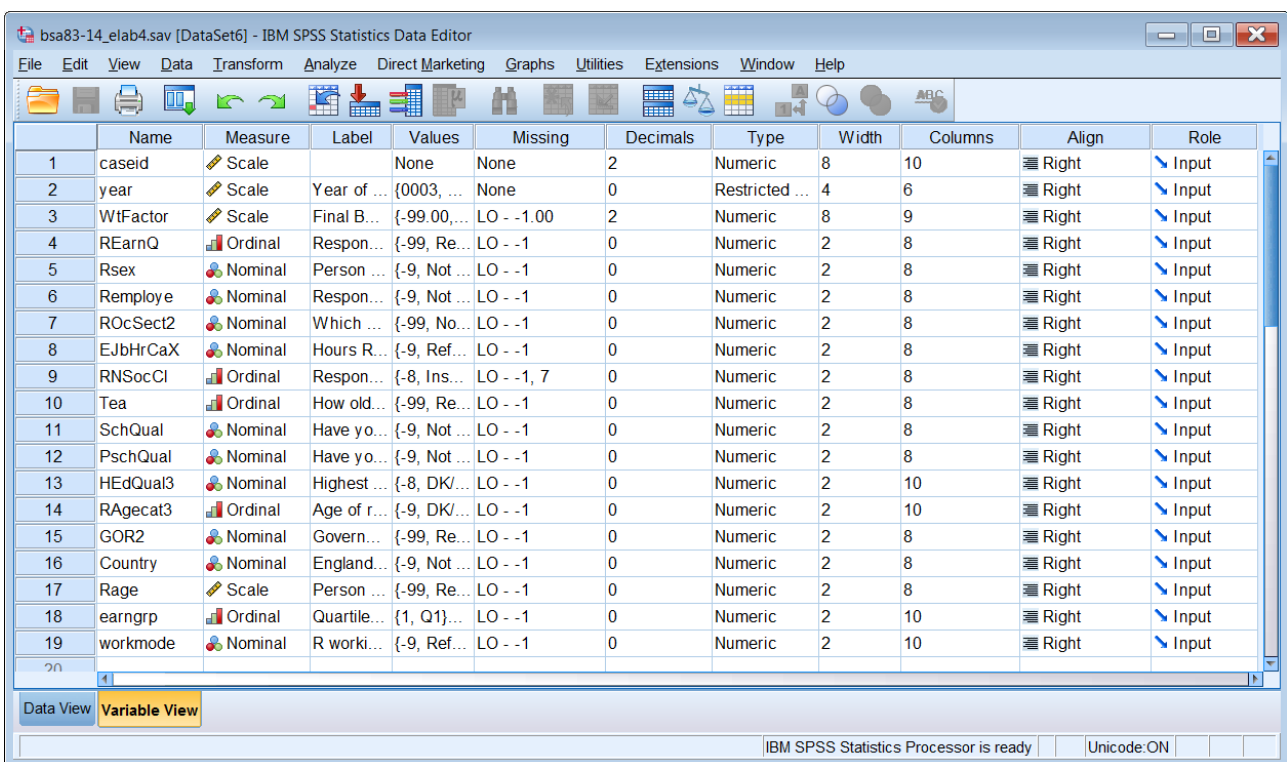
The sample in this file has been restricted to respondents from waves 2009 to 2014 only, for whom valid earnings data are available: those from waves 1983 to 2008 have been discarded, as have those for whom earnings information is missing or skipped as not applicable.

Working files were saved in folder **Elaboration**



bsa83-14_elab1	04/08/2016 18:30	SPSS Statistics Dat...	3,720 KB
bsa83-14_elab2	03/08/2016 17:41	SPSS Statistics Dat...	376 KB
bsa83-14_elab3	08/08/2016 09:25	SPSS Statistics Dat...	782 KB
bsa83-14_elab4	08/08/2016 09:26	SPSS Statistics Dat...	821 KB
bsa83-14_elab1	08/08/2016 09:12	SPSS Statistics Syn...	1 KB
bsa83-14_elab2	08/08/2016 09:18	SPSS Statistics Syn...	1 KB
bsa83-14_elab3	08/08/2016 09:20	SPSS Statistics Syn...	1 KB
bsa83-14_elab4	08/08/2016 09:21	SPSS Statistics Syn...	1 KB

Double click on **bsa83-14_elab4** 08/08/2016 09:26 SPSS Statistics Dat... 821 KB

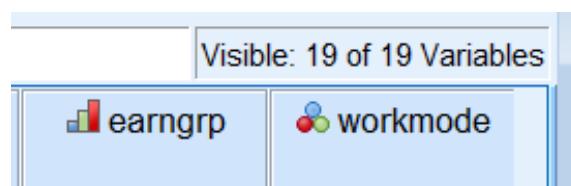
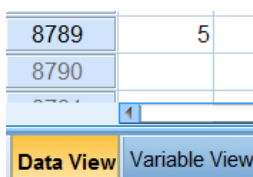


Switch to **Data View** and press **Ctrl End**: the **Data Editor** contains:

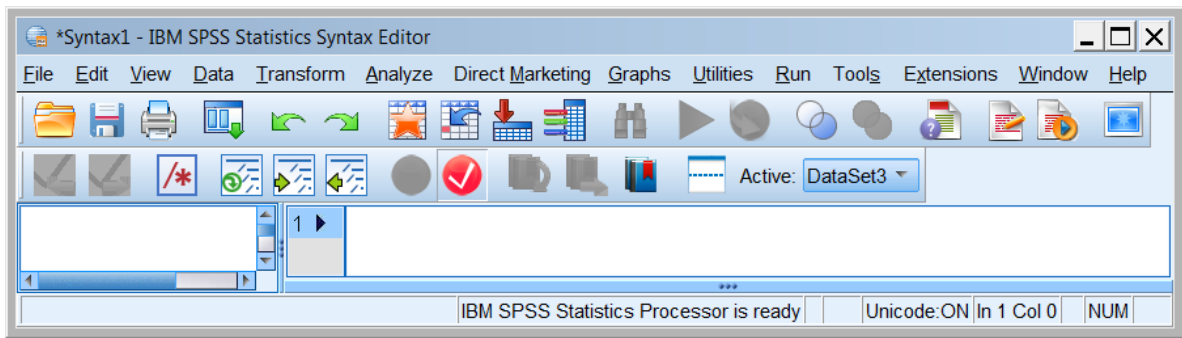
8.789 cases (bottom left)

and

19 variables (top right)



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



display labels.

[NB: display command not available from the GUI]

Variable Labels

Variable	Position	Label
caseid	1	<none>
year	2	Year of Interview
WtFactor	3	Final BSA weight
REarnQ	4	Respondent earnings quartiles (dv)
Rsex	5	Person 1 SEX
Remploye	6	Respondent currently employee or self-employed dv
ROcSect2	7	Which of the types of organisation on this card do/did you work for?
EJbHrCaX	8	Hours R works per week, excluding overtime [employee]. DV
RNSocCl	9	Respondent : social class [pre-SOC2000] best estimate dv
Tea	10	How old when completed your continuous full-time education?
SchQual	11	Have you passed any school examinations
PschQual	12	Have you achieved any post-school qualifications
HEdQual3	13	Highest educational qual obtained - dv
RAgecat3	14	Age of respondent(grouped)<6 category> dv
GOR2	15	Government office region 2003 version
Country	16	England, Scotland or Wales?
Rage	17	Person 1 age last birthday
earngrp	18	Quartile group of R's earnings from paid work
workmode	19	R working full- or part-time

Variables in the working file

The dependent variable **earngrp** has four categories with an approximately equal number of cases in each:

earngrp Quartile group of R's earnings from paid work

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Q1	2157	24.5	24.5	24.5
	2 Q2	2430	27.6	27.6	52.2
	3 Q3	2196	25.0	25.0	77.2
	4 Q4	2006	22.8	22.8	100.0
	Total	8789	100.0	100.0	

We are now going to examine the distribution of these earnings groups within categories of the other variables.

Elaboration: Zero order tables

Research question 3: What effect do they have by themselves?

To produce two-way tables for the independent and test variables use the SPSS command **CROSSTABS**

[General format: **CROSSTABS** <row variable> **by** <column variable(s)> /cells <cell contents> /statistics <stats list>.

My preference is to have the dependent variable in the rows: the default output has counts only.

crosstabs rsex **by** earngrp.

Rsex Person 1 SEX * earngrp Quartile group of R's earnings from paid work Crosstabulation

Count		earngrp Quartile group of R's earnings from paid work				Total
		1 Q1	2 Q2	3 Q3	4 Q4	
Rsex Person 1 SEX	1 Male	559	1113	1245	1382	4299
	2 Female	1598	1317	951	624	4490
Total		2157	2430	2196	2006	8789

Difference (Men-Women) +1039 +204 -294 -758 +191

There are approximately equal numbers of men (4299) and women (4490) in the sample, so even with these absolute numbers we can see that men have higher earnings than women. There are almost three times fewer men than women (559 :1598) in the lowest earnings group Q1 and more than twice as many men than women (1382 : 624) in the highest earnings group Q4.

To get a more accurate picture we need to standardise the figures in the table by converting the absolute figures to percentages:

crosstabs rsex **by** earngrp /cells count row.

Rsex Person 1 SEX * earngrp Quartile group of R's earnings from paid work Crosstabulation

		earngrp Quartile group of R's earnings from paid work				Total	
		1 Q1	2 Q2	3 Q3	4 Q4		
Rsex Person 1 SEX	1 Male	Count	559	1113	1245	1382	4299
		% within Rsex Person 1 SEX	13.0%	25.9%	29.0%	32.1%	100.0%
2 Female	Count	1598	1317	951	624	4490	
	% within Rsex Person 1 SEX	35.6%	29.3%	21.2%	13.9%	100.0%	
Total	Count	2157	2430	2196	2006	8789	
	% within Rsex Person 1 SEX	24.5%	27.6%	25.0%	22.8%	100.0%	

SPSS command **CROSSTABS** produces such cluttered output, especially if you request both counts and row percentages, that you have to edit the output manually. All we really need is this:

**Rsex Person 1 SEX * earngrp Quartile group of R's earnings from paid work
Crosstabulation**

% within Rsex Person 1 SEX

	earngrp Quartile group of R's earnings from paid work				n = 100%
	1 Q1	2 Q2	3 Q3	4 Q4	
1 Male	13.0%	25.9%	29.0%	32.1%	4299
2 Female	35.6%	29.3%	21.2%	13.9%	4490
All	24.5%	27.6%	25.0%	22.8%	8789

From this table we can calculate the **epsilon** statistic (percentage point difference) between men and women for each earnings group:

Epsilon¹ **-22.6** **-3.4** **+7.8** **+18.2**

Q.1: Is there a difference between the earnings (from paid work) of men and women?

A: Yes, for the total sample (2009 to 2014 only) for whom we have data.

If there were no difference in earnings between men and women, the percentages for men and women would be the same as the percentages for the whole sample, but there are marked differences. The percentages for the whole sample have been partitioned into conditional percentages for men and women: there is a clear and steep gradient in favour of men from **-22.6** points in the lowest earnings group Q1 to **+18.2** in the highest earnings group Q4.

We shall see later what happens when we introduce one or more test variables, but first let us answer the next question.

Q.2: What other variables might account for differences in earnings?

Let's have a look.

Work-related

crosstabs remploye to rnsoccl workmode **by** earngrp /cells count row.

**Remploye Respondent currently employee or self-employed dv * earngrp Quartile group of R's earnings from paid work
Crosstabulation**

			earngrp Quartile group of R's earnings from paid work				Total
			1 Q1	2 Q2	3 Q3	4 Q4	
Remploye Respondent currently employee or self-employed dv	1 Employee	Count	1787	2130	1934	1711	7562
		% within Remploye Respondent currently employee or self-employed dv	23.6%	28.2%	25.6%	22.6%	100.0%
	2 self-employed	Count	370	300	259	295	1224
		% within Remploye Respondent currently employee or self-employed dv	30.2%	24.5%	21.2%	24.1%	100.0%
Total		Count	2157	2430	2193	2006	8786
		% within Remploye Respondent currently employee or self-employed dv	24.6%	27.7%	25.0%	22.8%	100.0%

¹ Epsilon cannot be calculated by CROSSTABS: these epsilons were calculated by copying the body of the table into Excel, performing the calculations and copying the results back into Word.

[NB: Although this table is relatively clear, default **CROSSTABS** output for larger tables is generally far too cluttered and difficult to work with: the following tables have been manually edited to reduce the clutter a bit and shorten or delete inordinately long variable and value labels.]

Employe Respondent currently employee or self-employed dv * earngrp Quartile group of R's earnings from paid work Crosstabulation

		earngrp Quartile group of R's earnings from paid work				Total
		1 Q1	2 Q2	3 Q3	4 Q4	
1 Employee	Count	1787	2130	1934	1711	7562
	%	23.6%	28.2%	25.6%	22.6%	100.0%
2 self-employed	Count	370	300	259	295	1224
	%	30.2%	24.5%	21.2%	24.1%	100.0%
Total	Count	2157	2430	2193	2006	8786
	%	24.6%	27.7%	25.0%	22.8%	100.0%

Even this table is too cluttered, but if we ask for row percent only, we lose the base n for the percentages:

Employe Respondent currently employee or self-employed dv * earngrp Quartile group of R's earnings from paid work Crosstabulation

% within Employe Respondent currently employee or self-employed dv

		earngrp Quartile group of R's earnings from paid work				n = 100%
		1 Q1	2 Q2	3 Q3	4 Q4	
1 Employee		23.6%	28.2%	25.6%	22.6%	100.0%
2 self-employed		30.2%	24.5%	21.2%	24.1%	100.0%
Total		24.6%	27.7%	25.0%	22.8%	100.0%

The following tables have been manually edited to replace 100% with **base n**. This is extremely fiddly and time-consuming, but we'll see later that there is a better way to do it

Employe Respondent currently employee or self-employed dv * earngrp Quartile group of R's earnings from paid work Crosstabulation

% within Employe Respondent currently employee or self-employed dv

		earngrp Quartile group of R's earnings from paid work				n = 100%
		1 Q1	2 Q2	3 Q3	4 Q4	
1 Employee		23.6%	28.2%	25.6%	22.6%	7562
2 self-employed		30.2%	24.5%	21.2%	24.1%	1224
Total		24.6%	27.7%	25.0%	22.8%	8786

		earngrp Quartile group of R's earnings from paid work				n = 100%
		1 Q1	2 Q2	3 Q3	4 Q4	
1 PRIVATE SECTOR		25.9%	27.8%	23.0%	23.4%	4648
2 NATIONALISED OR PUBLIC		14.7%	42.2%	30.3%	12.8%	109
3 OTHER PUBLIC SECTOR		19.4%	28.0%	30.0%	22.6%	2455
4 CHARITY/ VOLUNTARY		25.2%	30.6%	29.3%	14.8%	317
Total		23.6%	28.2%	25.6%	22.6%	7529

[Numbers are a bit low in nationalised/public and in charity/voluntary]

Work-related (contd.)

EJbHrCaX Hours R works per week, excluding overtime [employee]. DV * earngrp Quartile group of R's earnings from paid work Crosstabulation

% within EJbHrCaX Hours R works per week, excluding overtime [employee]. DV

	earngrp Quartile group of R's earnings from paid work				n = 100%
	1 Q1	2 Q2	3 Q3	4 Q4	
0 less than 10 hours a week	55.4%	20.0%	16.2%	8.5%	130
1 10-15 hours a week	73.8%	17.4%	6.7%	2.0%	344
2 16-23 hours a week	63.8%	23.7%	8.4%	4.1%	898
3 24-29 hours a week	50.3%	29.2%	13.6%	7.0%	487
4 30 or more hours a week	10.5%	29.9%	31.2%	28.3%	5459
5 varies too much to say	29.3%	24.3%	17.9%	28.6%	140
Total	23.6%	28.3%	25.7%	22.5%	7458

workmode R working full- or part-time * earngrp Quartile group of R's earnings from paid work Crosstabulation

% within workmode R working full- or part-time

	earngrp Quartile group of R's earnings from paid work				n = 100%
	1 Q1	2 Q2	3 Q3	4 Q4	
1 Full time	10.5%	29.9%	31.2%	28.3%	7562
2 Part time	59.3%	23.8%	10.5%	6.5%	1224
Total	23.6%	28.3%	25.7%	22.5%	8786

[Analysis will eventually be restricted to those working full time only (30+ hours per week)].

RNSocCI Respondent : social class [pre-SOC2000] best estimate dv * earngrp Quartile group of R's earnings from paid work Crosstabulation

% within RNSocCI Respondent : social class [pre-SOC2000] best estimate dv

	earngrp Quartile group of R's earnings from paid work				n = 100%
	1 Q1	2 Q2	3 Q3	4 Q4	
1 I (SC=1)	3.8%	9.8%	28.7%	57.8%	666
2 II (SC=2)	12.1%	21.8%	30.7%	35.4%	3394
3 III (non-manual) (SC=3)	36.8%	34.1%	19.9%	9.2%	1710
4 III (manual) (SC=4)	23.3%	35.5%	28.9%	12.3%	1569
5 IV (SC=5)	49.6%	35.0%	12.3%	3.1%	1107
6 V (SC=6)	61.9%	30.2%	4.2%	3.8%	265
Total	24.6%	27.7%	25.0%	22.7%	8711

For the purposes of elaboration **rnsoccl** can be grouped into two categories (1,2 : 3 to 6) or manual-/non-manual (1 to 3: 4 to 6) possibly three, but this will depend on the number of cases in each, and on the proportion of men to women.

Education

crosstabs tea to hedqual3 by earngrp /cells row.

Tea How old when completed your continuous full-time education?[compressed] dv * earngrp Quartile group of R's earnings from paid work Crosstabulation

% within Tea How old when completed your continuous full-time education?[compressed] dv

		earngrp Quartile group of R's earnings from paid work				n = 100%
		1 Q1	2 Q2	3 Q3	4 Q4	
Tea	1 15 or under	36.0%	32.4%	21.3%	10.3%	1013
	2 16	28.9%	31.7%	24.2%	15.3%	2668
	3 17	25.9%	29.8%	25.8%	18.5%	784
	4 18	26.9%	29.4%	24.3%	19.4%	1301
	5 19 or over	14.5%	21.3%	27.4%	36.8%	2964
	6 Still at school	75.0%		25.0%		4
	7 Still at college or university	70.5%	15.9%	6.8%	6.8%	44
Total		24.5%	27.7%	25.0%	22.8%	8778

Those still at school, college or university can be discarded.

SchQual Have you passed any school examinations * earngrp Quartile group of R's earnings from paid work Crosstabulation

% within SchQual Have you passed any school examinations

		earngrp Quartile group of R's earnings from paid work				n = 100%
		1 Q1	2 Q2	3 Q3	4 Q4	
Passed any school examinations	1 Yes	22.4%	26.5%	26.0%	25.1%	7518
	2 No	37.5%	34.0%	19.3%	9.2%	1265
Total		24.5%	27.6%	25.0%	22.8%	8783

PschQual Have you achieved any post-school qualifications * earngrp Quartile group of R's earnings from paid work Crosstabulation

% within PschQual Have you achieved any post-school qualifications

		earngrp Quartile group of R's earnings from paid work				n = 100%
		1 Q1	2 Q2	3 Q3	4 Q4	
Achieved any post-school qualifications	1 Yes	20.4%	26.3%	26.7%	26.6%	6624
	2 No	37.2%	31.7%	19.7%	11.3%	2156
Total		24.5%	27.6%	25.0%	22.8%	8783

HEdQual3 Highest educational qual obtained - dv * earngrp Quartile group of R's earnings from paid work Crosstabulation

% within HEdQual3 Highest educational qual obtained - dv

		earngrp Quartile group of R's earnings from paid work				n = 100%
		1 Q1	2 Q2	3 Q3	4 Q4	
	1 Degree	10.2%	17.0%	28.8%	44.0%	1185
	2 Higher educ below degree/A level	26.1%	29.4%	24.9%	19.6%	1389
	3 O level or equiv/CSE	31.8%	33.5%	21.3%	13.5%	1039
	4 No qualification	48.6%	33.2%	11.3%	6.9%	319
Total		24.6%	27.1%	24.0%	24.3%	3932

[Numbers have reduced significantly, but check sample]

Age

crosstabs ragecat3 by earngrp /cells row.

RAgecat3 Age of respondent(grouped)<6 category> dv * earngrp Quartile group of R's earnings from paid work Crosstabulation

% within RAgecat3 Age of respondent(grouped)<6 category> dv

	earngrp Quartile group of R's earnings from paid work				n = 100%
	1 Q1	2 Q2	3 Q3	4 Q4	
1 18-25	45.0%	33.5%	17.5%	4.0%	424
2 26-35	22.5%	28.8%	27.5%	21.1%	1193
3 36-45	22.0%	26.4%	24.0%	27.5%	1551
4 46-55	23.9%	25.4%	23.7%	27.0%	1433
5 56-65	23.5%	31.0%	24.3%	21.3%	874
6 >66	36.0%	24.7%	19.3%	20.0%	150
Total	24.9%	27.9%	24.1%	23.1%	5625

Location

GOR2 Government office region 2003 version * earngrp Quartile group of R's earnings from paid work Crosstabulation

% within GOR2 Government office region 2003 version

	earngrp Quartile group of R's earnings from paid work				n = 100%
	1 Q1	2 Q2	3 Q3	4 Q4	
1 North East	26.8%	32.4%	24.5%	16.2%	444
2 North West	27.5%	28.7%	24.2%	19.5%	1024
3 Yorkshire and Humberside	30.2%	27.6%	22.6%	19.6%	725
4 East Midlands	22.2%	29.0%	27.3%	21.5%	721
5 West Midlands	25.3%	33.1%	23.1%	18.5%	801
6 SW	27.7%	31.7%	22.4%	18.1%	772
7 Eastern	22.8%	22.7%	24.8%	29.7%	958
8 Inner London	19.7%	18.2%	29.8%	32.4%	346
9 Outer London	15.6%	24.2%	27.0%	33.2%	563
10 South East	22.4%	24.4%	24.8%	28.3%	1253
11 Wales	28.1%	32.2%	25.2%	14.4%	416
12 Scotland	24.5%	28.3%	27.5%	19.6%	766
Total	24.5%	27.6%	25.0%	22.8%	8789

Country England, Scotland or Wales? * earngrp Quartile group of R's earnings from paid work Crosstabulation

% within Country England, Scotland or Wales?

	earngrp Quartile group of R's earnings from paid work				n = 100%
	1 Q1	2 Q2	3 Q3	4 Q4	
1 England	24.3%	27.3%	24.7%	23.6%	7607
2 Scotland	24.5%	28.3%	27.5%	19.6%	766
3 Wales	28.1%	32.2%	25.2%	14.4%	416
Total	24.5%	27.6%	25.0%	22.8%	8789

Year

`crosstabs year by earngrp /cells row.`

year Year of Interview * earngrp Quartile group of R's earnings from paid work Crosstabulation

% within year Year of Interview

		earngrp Quartile group of R's earnings from paid work				n = 100%
		1 Q1	2 Q2	3 Q3	4 Q4	
year Year of Interview	2009	23.4%	24.5%	27.6%	24.5%	1689
	2010	24.4%	30.3%	25.4%	19.9%	1474
	2011	25.0%	28.6%	24.9%	21.5%	1484
	2012	26.7%	26.6%	24.5%	22.2%	1385
	2013	25.3%	28.4%	22.4%	23.9%	1459
	2014	22.6%	27.8%	24.7%	25.0%	1298
Total		24.5%	27.6%	25.0%	22.8%	8789

Except in 2010, the distribution is reasonably even.

Because SPSS command **CROSSTABS** produces such cluttered output, especially if you request both counts and row percents, it's better from now on to use **CTABLES**. The syntax is quite complex, but the tables are much clearer. However, even **CTABLES** doesn't calculate the epsilons, but I'm working on it.

End of tutorial: **3.2.1.3 Elaboration 3 (Income differences 2009 – 2014: CROSSTABS)**
[Last updated: 11 August 2016]

Next tutorial: **3.2.1.4 Elaboration 4 (Income differences 2009 – 2014_CTABLES)**

Back to: [3.2.1.2 Elaboration 2 \(Income differences BSA 2009 - 2014\)](#)

Back to: [3.2 Three \(or more\) variables](#)