[Commentary by **John F Hall**] [**Draft only**: New page 4 July 2018]

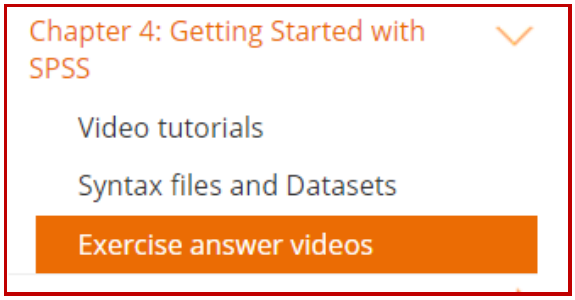
John MacInnes
An Introduction to Secondary Data Analysis with IBM SPSS Statistics
(Sage, Dec. 2017)

Chapter 4: Getting Started with SPSS **4.2.2: Exercise answer video 7 (4'54")**

Previous guides:

Macinnes 2017 Aide-mémoire for easier navigation of companion website
MacInnes 4.1.1 Overview of video tutorials 1 to 6
MacInnes 4.1.2 Downloading the European Social Survey Practice File
MacInnes 4.1.3 Downloading the SPSS syntax
MacInnes 4.1.4_Checking_the_SPSS_files
MacInnes 4.1.5 Guide to video tutorials 7 and 8
MacInnes 4.1.6 Guide to video tutorial 9
MacInnes 4.1.7 Guide to video tutorials 10 and 11
MacInnes 4.1.8 Guide to video tutorial 12 for Chapter 4
Macinnes_4.2.1_exercise_answer_videos_1-6

Chapter.4: Exercise answer videos (direct link to companion website)



Extract from page 104

exercises using the ESS6 practice dataset

1 Which countries have the (a) highest and (b) lowest proportion of their adult populations born in other countries?
2 Across all countries, what proportion of those not born in the country in which they are living are women? (Hint: use the correct weight variable!)
3 Across all countries, is the employment rate for those born in the country higher or lower than those not born there?
4 What is the average age of those not born in the country they live in, compared to those born there?
5 Are those born in the country or those not born in the country more likely to belong to a religious denomination?
6 In which countries do people report the highest and lowest level of trust in political parties?
7 What seems to have more impact on trust in political parties: age, gender or religion?
8 What proportion of men and women have married by the time they are 30 years old across all the countries in the survey?
9 In which country is the proportion of people who say they never watch television the highest?
10 Produce a chart of the mean size of households across the countries in the survey, ranking average household size from the smallest to the largest.

Answers, and explanations of how to arrive at these results, are available on the companion website.

Exercise answer video 7:

Exemplar: ESS6_Practice.sav
[Extracted direct from the ESS Round 6 2012 integrated mother file[1]]

Research question 7: (page 104)

What seems to have more effect on trust in political parties: age, gender or religion?

Actually, a prior question, to make students think, should perhaps be:

What sort of things could affect people's trust in political parties?

. . and then look for likely candidate variables in the data set. However, this is where we get down to some serious sociology, aided by a bit of statistics.

Dependent variable:	**[trstprt]**	"Trust in political parties"
Independent variables:	**[agea]**	Age of respondent
	[gndr]	Gender
	[rlgblg]	Belong to a particular religion or denomination
SPSS commands:	**RECODE**	
	CROSSTABS	

In direct syntax, JM recodes **[agea]** into a new variable **[agegroup]** before he explains recoding.
He then uses the GUI to check that variable **[pweight]** is included and that all cases will be used.
The syntax to ensure this would be:

```
weight by pweight.
use all.
```

Does **RECODE** in direct syntax:

```
135 recode agea (lo thru 29 = 1) (30 thru 49 = 2) (50 thru 64 = 3) (65 thru hi = 4) into agegroup.
136 value labels agegroup 1 "up to 29" 2 " 30 to 49" 3 " 50 to 64" 4 "65+".
137 freq agegroup.
```

```
recode agea
   (lo thru 29 =1)(30 thru 49 =2)(50 thru 65 = 3)(65 thru hi =4)
   into agegroup.
value labels agegroup 1 'up to 29' 2 '30 to 49' 3 '50 to 64' 4 '65+'.
```

[1] **ESS6e02_3.sav** ESS Round 6: European Social Survey Round 6 Data (2012). Data file edition 2.3. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC. See Appendix for how this was done.

. . and as a check on the new variable:

freq agegroup. [Note the use of abbreviated syntax: **freq** for **frequencies**]

. . produces the following table:

agegroup

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	up to 29	12328	21.7	21.7	21.7
	30 to 49	17861	31.4	31.4	53.1
	50 to 64	15285	26.9	26.9	80.0
	65+	11361	20.0	20.0	100.0
	Total	56835	100.0	100.0	

[Weighted]

In file **ess6_practice.sav** the value **999** is declared as a **missing**,

14	agea	Age of respondent, calculated	{999, Not available}...	999

From the manual (p 1590)

[Help >> Command Syntax Reference Scroll down to > RECODE]

"LOWEST and HIGHEST (LO and HI) specify the lowest and highest values encountered in the data. ***LOWEST and HIGHEST include user-missing values*** *but not the system-missing value."*

Because JM used SPSS keyword **hi** in **(65 thru hi = 4)** value **999** has been recoded to valid value **4**. This means that 133 cases missing for **[agea]** have been wrongly assigned to the **65+** category.

The oldest person in the sample is 103, so he could have used **(65 thru 103 =4)**. Value 999 is then automatically assigned to **system missing**.

```
recode agea
  (lo thru 29 =1)(30 thru 49 =2)(50 thru 65 = 3)(65 thru 103 =4)
  into agegroup.
value labels agegroup 1 'up to 29' 2 '30 to 49' 3 '50 to 64' 4 '65+'.

frequencies agegroup .
```

agegroup

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	11034	19.4	19.5	19.5
	2	18528	32.6	32.7	52.2
	3	15592	27.4	27.5	79.7
	4	11468	20.2	20.3	100.0
	Total	56622	99.6	100.0	
Missing	**System**	**213**	0.4		
Total		56835	100.0		

[Weighted]

Alternative syntax to declare a user-missing value for age group would be:

```
if (missing (agea)) agegroup = 9.
formats agegroup (f2.0).
missing values agegroup (9)
```

. . or

```
recode agea
   (lo thru 29 =1)(30 thru 49 =2)(50 thru 65 = 3)(65 thru 103 =4)(999 = 9)
   into agegroup.
formats agegroup (f2.0).
missing values agegroup (9).
value labels agegroup 1 'up to 29' 2 '30 to 49' 3 '50 to 64' 4 '65+'.

frequencies agegroup .
```

agegroup

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	up to 29	11034	19.4	19.5	19.5
	30 to 49	18528	32.6	32.7	52.2
	50 to 64	15592	27.4	27.5	79.7
	65+	11468	20.2	20.3	100.0
	Total	56622	99.6	100.0	
Missing	**9**	**213**	0.4		
Total		56835	100.0		

[Weighted]

Zero order contingency tables (2'30")

Dependent (in column)	**Independent** (in rows)	**Label**
trstprt	rlgblg	Belonging to particular religion or denomination
	gndr	Gender
	agegroup	[No label]

JM uses the GUI to produce tables the sociological way round (independent variable(s) in rows, dependent in columns) with row %. With only 21 variables it is relatively easy to find the variables he needs via the GUI, but the full ESS6 file has 626 variables. Using the GUI in such circumstances is tortuous, tedious and time-consuming. Syntax is always quicker and easier.

He does not use **PASTE** to save the syntax, but it would be:

```
crosstabs rlgblg gndr agegroup by trstprt /cells row.
```

Belonging to particular religion or denomination * Trust in political parties Crosstabulation

% within Belonging to particular religion or denomination

		Trust in political parties											
		No trust at all	1	2	3	4	5	6	7	8	9	Complete trust	Total
Belonging to particular religion or denomination	Yes	21.6%	10.4%	13.4%	14.4%	10.9%	15.1%	6.7%	4.4%	1.9%	0.5%	0.7%	100.0%
	No	20.4%	10.5%	12.9%	13.9%	11.8%	15.6%	7.5%	4.6%	2.0%	0.4%	0.4%	100.0%
Total		21.2%	10.4%	13.2%	14.2%	11.2%	15.3%	7.0%	4.5%	2.0%	0.5%	0.6%	100.0%

[Weighted]

Gender * Trust in political parties Crosstabulation

% within Gender

		Trust in political parties											Total
		No trust at all	1	2	3	4	5	6	7	8	9	Complete trust	
Gender	Male	21.7%	10.7%	13.0%	13.8%	11.7%	14.4%	7.2%	4.6%	1.9%	0.5%	0.5%	100.0%
	Female	20.8%	10.2%	13.4%	14.5%	10.9%	16.0%	6.8%	4.4%	2.0%	0.5%	0.6%	100.0%
Total		21.2%	10.4%	13.2%	14.2%	11.2%	15.2%	7.0%	4.5%	2.0%	0.5%	0.6%	100.0%

[Weighted]

agegroup * Trust in political parties Crosstabulation

% within agegroup

		Trust in political parties											Total
		No trust at all	1	2	3	4	5	6	7	8	9	Complete trust	
agegroup	up to 29	19.4%	11.0%	13.0%	12.7%	11.7%	15.4%	8.3%	5.4%	2.0%	0.5%	0.6%	100.0%
	30 to 49	21.9%	11.2%	13.4%	14.6%	11.7%	14.4%	6.2%	4.1%	1.4%	0.5%	0.4%	100.0%
	50 to 64	22.5%	10.2%	13.3%	15.0%	10.8%	14.6%	6.4%	4.2%	2.0%	0.4%	0.5%	100.0%
	65+	20.1%	8.9%	13.0%	13.6%	10.6%	17.4%	7.7%	4.7%	2.7%	0.7%	0.7%	100.0%
Total		21.2%	10.4%	13.2%	14.2%	11.2%	15.2%	7.0%	4.5%	2.0%	0.5%	0.5%	100.0%

[Weighted]

JM comments "Because there are so many values for **[trstprt]** it's not very easy to tell what's going on." In fact, the tables are so wide they won't even fit in his viewer and he can only see the full tables by dragging the right edge of the window out.

The tables are cluttered and difficult to interpret so he says, "Let's use the **MEANS** procedure" (3'17")

```
MEANS trstprt by rlgblg  gndr,. agegroup
    /cells mean count.
```

Trust in political parties * Belonging to particular religion or denomination

Trust in political parties

Belonging to particular religion or denomination	Mean	N
Yes	2.97	32900
No	3.04	21510
Total	3.00	54410

Trust in political parties * Gender

Trust in political parties

Gender	Mean	N
Male	2.97	24674
Female	3.01	30366
Total	2.99	55041

Trust in political parties * agegroup

Trust in political parties

agegroup	Mean	N
up to 29	3.13	10582
30 to 49	2.87	18114
50 to 64	2.91	15195
65+	3.18	11011
Total	2.99	54903

JM concludes that none of the three variables have much impact on **[trstprt]**

Religion ". . very little difference"
Gender ". . slight difference, but again not much happening here."
Age group ". . seems to be highest amongst younger people and also amongst older people."

[**NB:** In JM's version of **ESS6-Practice.sav**, the variable **[trstprt]** is declared as **Scale**, but technically it is **Ordinal**. In the survey trade it is common practice to treat Likert-type ordinal variables as scale, but usually only if there at least 7 points. In the full ESS6 file and in file ess6_practice.sav on this site, it is correctly declared as **Ordinal**.]

None of the syntax for video 7 is either on the site or saved with **PASTE**. There is little if any correspondence between the contents of Chapter 4 and video 7.

He does not consider recoding **[trstprt]** into fewer categories or suggest other variables which might affect trust in political parties. For example, much more interesting is:

```
means trstprt by cntry
 /cells mean count.
```

Report

Trust in political parties

Country	Mean	N
Denmark	5.31	450
Norway	5.15	401
Netherlands	5.02	1371
Switzerland	4.99	643
Finland	4.89	447
Sweden	4.86	773
Belgium	4.23	917
Germany	3.68	7017
United Kingdom	3.61	5046
Iceland	3.58	25
Hungary	3.24	822
Estonia	3.20	110
Israel	3.14	538
France	3.13	5278
Russian Federation	3.01	11381
Ireland	2.94	348
Lithuania	2.82	247
Slovakia	2.74	450
Czech Republic	2.69	878
Cyprus	2.46	70
Albania	2.32	219
Slovenia	2.27	172
Poland	2.21	3174
Kosovo	2.01	131
Italy	2.00	5131
Ukraine	1.90	3699
Spain	1.88	3873
Portugal	1.87	889
Bulgaria	1.80	607
Total	2.99	55107

[Weighted: rows sorted in descending order]

The effect of religion on trust in political parties is more marked if **trstprt** is tabulated by **rlgdnm** (denomination/faith):

```
means trstprt by rlgdnm
 /cells mean count.
```

Report

Trust in political parties

Religion or denomination belonging to at present	Mean	N
Protestant	4.12	5155
Eastern religions	3.74	277
Islamic	3.50	1866
Other non-Christian religions	3.32	136
Jewish	3.08	544
Other Christian denomination	2.92	604
Eastern Orthodox	2.76	8305
Roman Catholic	2.63	15859
Total	2.97	32744

[Weighted: rows sorted in descending order]

[**NB:** Mean **trstprt** for those not belonging to a particular religion (**rlgblg** code 2) is 3.04 (n=21510)]

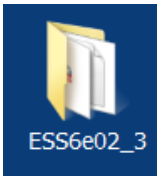
End of: **4.2.2: Exercise answer video 7**
Forward to: **4.2.3: Exercise answer video 8**
Back to: **Macinnes (2017)**

Appendix

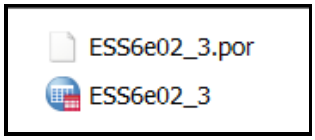
Some variable attributes in the file on the companion website are slightly different to those in the main file for ESS6. An extract can be obtained from the European Social Survey site.

On page ESS6 - 2012 Data Download, in the right hand margin under **Integrated File Download**, click on Download ESS Round 6 (2012).

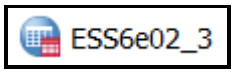
It will arrive in a *.zip file **ESS6e02_3.spss.zip** from which folder **ESS6e02_3** can be extracted to your specified location (in this case my desktop). The file you need is **ESS6_e2_3.sav** (51.2 mb)

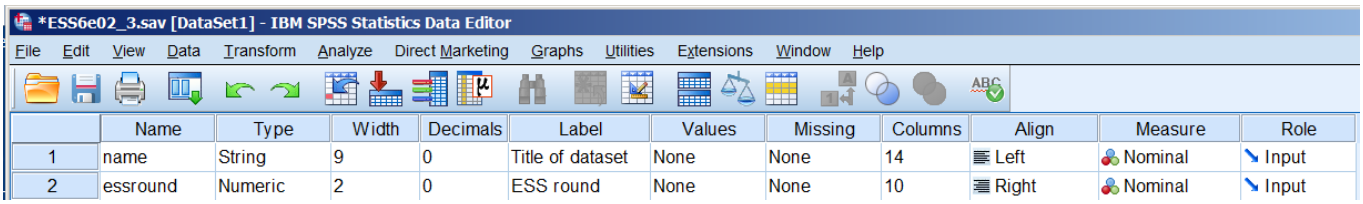


contains two files



Double click on





	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	name	String	9	0	Title of dataset	None	None	14	Left	Nominal	Input
2	essround	Numeric	2	0	ESS round	None	None	10	Right	Nominal	Input

File **ESS6e02_3.sav** contains 626 variables:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
625	pspwght	Numeric	4	2	Post-stratificati...	None	None	10	Right	Scale	Input
626	pweight	Numeric	8	2	Population siz...	None	None	10	Right	Scale	Input
627											

. . and 54673 cases:

	inwyye	inwehh	inwemm	inwtm	spltadme	supqad1	supqad2	supqdd	supqmm	supqyr	dweight	pspwght	pweight
54672	2013	14	42	72	1	6	1	28	2	2013	0.03	0.04	0.10
54673	2013	16	32	82	2	6	1	28	2	2013	0.17	0.22	0.10
54674													

From the SPSS **Data Editor**, the appropriate extract can be obtained with:

```
title 'Extract variables for ESS6-Practice.sav'.
save outfile "C:\ <filepath> \ <filename>"
 /keep
   idno cntry tvtot polintr trstprt lrscale freehms rlgblg rlgdnm brncntr physact
   hhmmb gndr agea maritalb mnactic wkhtot dweight pspwght pweight.
```

In my case the filepath is specified as:

'C:\Users\John Hall\Desktop\Research 2018\MacInnes 2017\JFH SPSS files\Ch4_Replications \ess6 _practice.sav'