[Commentary by **John F Hall**] [Last updated: 10 August 2017]

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

Chapter 4: Getting Started with SPSS

Chapter 4 video tutorials (direct link to companion website)
[**NB:** All video tutorials for chapter 4 are on the same web page and cannot (yet) be disaggregated]

4.1.1: Overview of video tutorials 1 -6

Warning!

All the **video tutorials** for chapter 4 are on the same (very sensitive) webpage and can be accidentally triggered by stray mouse pointers. If you are not careful you can have two or more simultaneous commentaries playing and can't always tell which commentary relates to which video. It would be far better to split all these videos across separate pages, but this not yet possible.

All the **SPSS syntax** for all the analyses is in the same **Syntax Editor**, which can get very complicated and possibly confusing. The syntax needs to be split into separate SPSS *.sps files for each analysis.

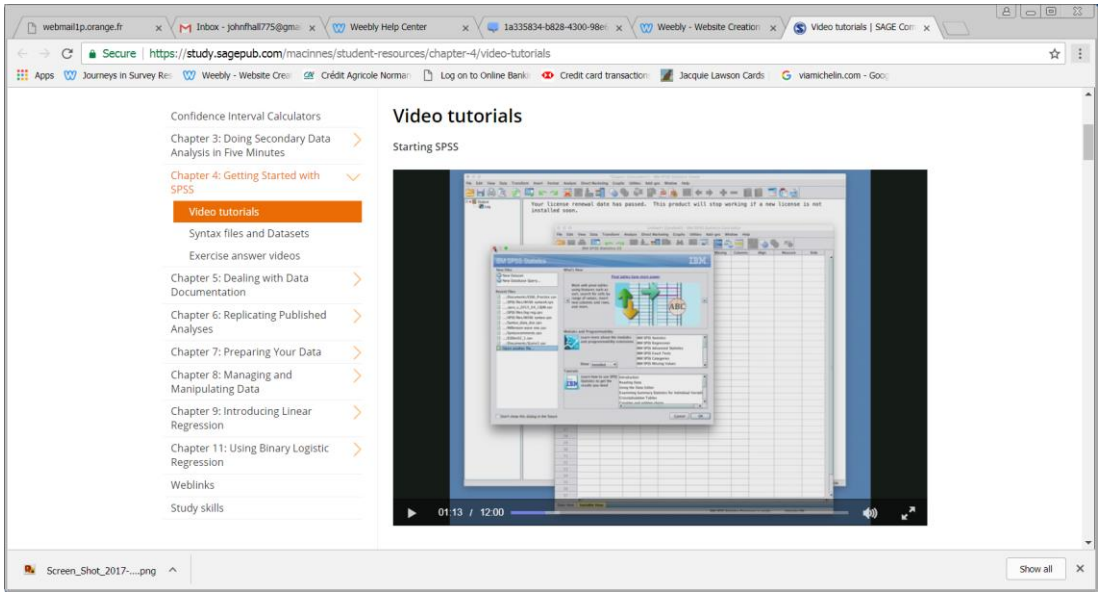
The following notes assume you are familiar with copy/paste, highlighting and dragging with left mouse down and that you have access to a licenced copy of SPSS. They are based on communications with John MacInnes and Sage when I first accessed the companion website and are offered not as criticism, but as constructive supplementary comments intended to help guide users through Chapter 4.

The rationale for the order of video topics is not particularly evident: in fact the chapters can be read in almost any order without losing pedagogic efficacy. Some topics could do with much more preliminary explanation of what is being done and why. However, at over 300 pages the book is already quite big and additional material could well make it unwieldy.

Contents:	**Variable(s) used in examples**
Video 1: Starting SPSS	[None: demonstration only]
Video 2: Frequencies	Country of survey
Video 3: Crosstabs	Country of survey by whether born there
Video 4: Histogram	Age of respondent
Video 5: Confidence Intervals	*"Gays and lesbians free to live life as they wish"*
Video 6: Crosstabs & chi-square	*"Gays and lesbians free to live life as they wish"* by **gender**, **age group** and **religion**

Contents (contd.)

Video 1: Starting SPSS



Quick tour through SPSS with gentle Scottish voice-over by MacInnes himself, demonstrating SPSS opening screen, **Data Editor** and **Output Viewer** windows, opening an SPSS saved file, different data types (**numeric** and **string**) using Edit >> Options to make your screen presentation the same as his, displaying the **Data Editor** in Variable View or Data View,[1] changing the displayed attributes of variables with View >> Customise Variable View, changing the font size, dragging column dividers (to reveal full contents of cells) and a look at the **Output Viewer**.

The author explains that each **row** of the matrix contains the data for all **variables** of a single **case**, each **column** contains all **values** for a single **variable** and each **cell** contains the **value** of a single **variable** for a single **case**.

His approach to this vocabulary is similar to the scheme I use for my teach-yourself course Survey Analysis Workshop (SPSS) but my course starts much further back in the data capture process, with a chalk-board exercise on the nature of survey data, listing a few simple questions, suggesting shorthand names to be used as column headers, entering rows of imaginary people and single word summaries of their responses, converting the answers from string to numeric, gradually building up a research vocabulary (**cases**, **variables**, **values**) and explaining that we have just created a flat **data matrix** of the type used by SPSS and Excel. (See 1.1.2 Introduction to survey data, p.2)

[1] In the video Data View displays **labels** instead of **values.**

During registration for the first class, students were given a short self-completion questionnaire[2] with some questions replicated from the British Social Attitudes Survey, some about themselves and their experience of surveys and computers. After a brief explanation of the Aladdin's cave of data stored on the Vax mainframe (no PCs or Macs in those days) and introduction to some simple commands for the Vax Management System (**VMS**: explained as having to learn the Turkish equivalent of Open Sesame") and the file editor **EDT** (explained as having to learn Greek to access the goodies inside) they got a coffee break before being introduced to the visual display units (VDUs) in the computer lab and obtain their individual passwords to access their personal disk area on the Vax mainframe. They were then asked to copy a short pre-pared SPSS syntax file from the Faculty disk area to their own disk area and run it on the Vax. Everyone got some output (though the few spares came in handy) and had something to discuss before we opened the Bulgarian wine.

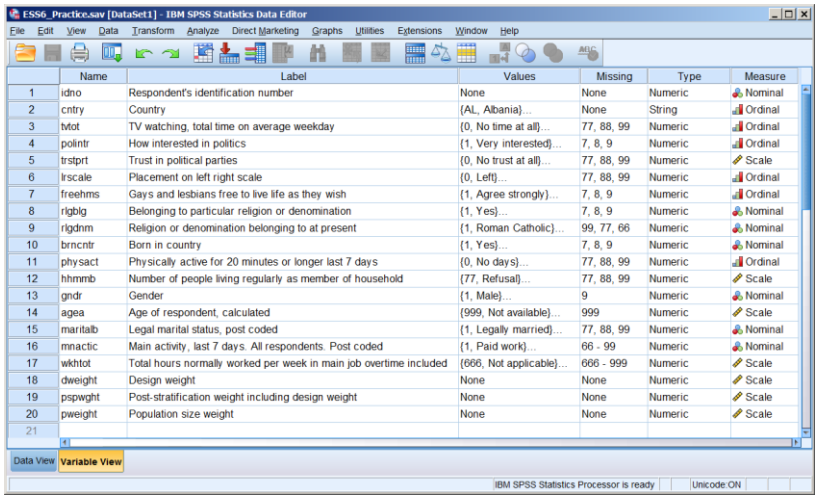
For the final three years of the course (1990-1992), two of my (1990) 2nd year undergraduates[3] on the BSocSci (Social Research) were employed as demonstrators for the lab sessions.

In the next session students were asked to code the data from 10 questionnaires and transfer them to a 10-line data sheet, one line per questionnaire. They then had to type these 10 lines into an 80-column ASCII file, prior to entering them into SPSS via **DATA LIST.**[4]

This was my introduction to them of “learning by doing”: it was very effective and gave students with little or no previous experience of surveys or computers their first appreciation of the nature of survey data.

Video 2: Frequency tables

With file **ESS6_Practice.sav** open:



	Name	Label	Values	Missing	Type	Measure
1	idno	Respondent's identification number	None	None	Numeric	Nominal
2	cntry	Country	{AL, Albania}...	None	String	Ordinal
3	tvtot	TV watching, total time on average weekday	{0, No time at all}...	77, 88, 99	Numeric	Ordinal
4	polintr	How interested in politics	{1, Very interested}...	7, 8, 9	Numeric	Ordinal
5	trstprt	Trust in political parties	{0, No trust at all}...	77, 88, 99	Numeric	Scale
6	lrscale	Placement on left right scale	{0, Left}...	77, 88, 99	Numeric	Ordinal
7	freehms	Gays and lesbians free to live life as they wish	{1, Agree strongly}...	7, 8, 9	Numeric	Ordinal
8	rlgblg	Belonging to particular religion or denomination	{1, Yes}...	7, 8, 9	Numeric	Nominal
9	rlgdnm	Religion or denomination belonging to at present	{1, Roman Catholic}...	99, 77, 66	Numeric	Nominal
10	brncntr	Born in country	{1, Yes}...	7, 8, 9	Numeric	Nominal
11	physact	Physically active for 20 minutes or longer last 7 days	{0, No days}...	77, 88, 99	Numeric	Ordinal
12	hhmmb	Number of people living regularly as member of household	{77, Refusal}...	77, 88, 99	Numeric	Scale
13	gndr	Gender	{1, Male}...	9	Numeric	Nominal
14	agea	Age of respondent, calculated	{999, Not available}...	999	Numeric	Scale
15	maritalb	Legal marital status, post coded	{1, Legally married}...	77, 88, 99	Numeric	Nominal
16	mnactic	Main activity, last 7 days. All respondents. Post coded	{1, Paid work}...	66 - 99	Numeric	Nominal
17	wkhtot	Total hours normally worked per week in main job overtime included	{666, Not applicable}...	666 - 999	Numeric	Scale
18	dweight	Design weight	None	None	Numeric	Scale
19	pspwght	Post-stratification weight including design weight	None	None	Numeric	Scale
20	pweight	Population size weight	None	None	Numeric	Scale
21						

---

2 See: 1.1.1 Pre-course questionnaire on interests and skills

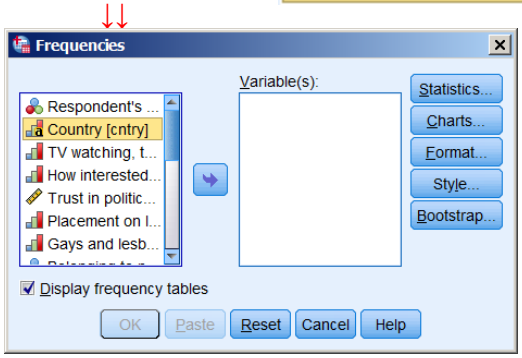
3 Deborah Youdell is now Professor of Sociology of Education, Birmingham University;
Katie Featherstone is now Director of Postgraduate Studies (Research) School of Nursing and Midwifery Studies Cardiff University

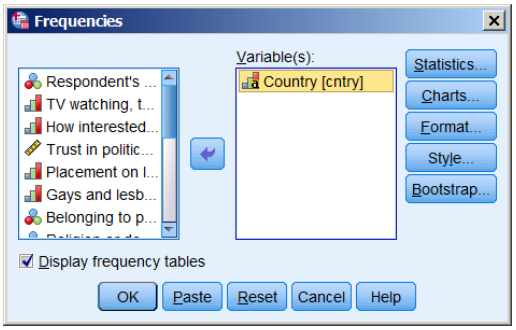
4 See 1.2.1 Data transfer sheet and 1.2.2 Preliminary data exercise: no SPSS for Windows or Excel in those days!

MacInnes produces a frequency count for country of survey **[cntry]** by using the GUI menu:

Analyze >> Descriptive Statistics >> Frequencies

In left pane highlight Country [cntry] and click to transfer to right pane





To generate syntax and copy it to your **Syntax Editor** press Paste

Syntax generated:

```
FREQUENCIES VARIABLES=cntry
  /ORDER=ANALYSIS.
```

To see the frequency table press OK

Frequency output generated:

Statistics

Country

N	Valid	54673
	Missing	0

There are 54673 countries and no missing cases for country.

Frequency table generated

Country

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Albania	1201	2.2	2.2	2.2
	Belgium	1869	3.4	3.4	5.6
	Bulgaria	2260	4.1	4.1	9.7
	Switzerland	1493	2.7	2.7	12.5
	Cyprus	1116	2.0	2.0	14.5
	Czech Republic	2009	3.7	3.7	18.2
	Germany	2958	5.4	5.4	23.6
	Denmark	1650	3.0	3.0	26.6
	Estonia	2380	4.4	4.4	31.0
	Spain	1889	3.5	3.5	34.4
	Finland	2197	4.0	4.0	38.5
	France	1968	3.6	3.6	42.1
	United Kingdom	2286	4.2	4.2	46.2
	Hungary	2014	3.7	3.7	49.9
	Ireland	2628	4.8	4.8	54.7
	Israel	2508	4.6	4.6	59.3
	Iceland	752	1.4	1.4	60.7
	Italy	960	1.8	1.8	62.4
	Lithuania	2109	3.9	3.9	66.3
	Netherlands	1845	3.4	3.4	69.7
	Norway	1624	3.0	3.0	72.6
	Poland	1898	3.5	3.5	76.1
	Portugal	2151	3.9	3.9	80.0
	Russian Federation	2484	4.5	4.5	84.6
	Sweden	1847	3.4	3.4	88.0
	Slovenia	1257	2.3	2.3	90.3
	Slovakia	1847	3.4	3.4	93.6
	Ukraine	2178	4.0	4.0	97.6
	Kosovo	1295	2.4	2.4	100.0
	Total	54673	100.0	100.0	

Again a matter of pedagogy perhaps, but the GUI route takes a lot longer than simply opening a new **Syntax Editor** and typing in:

freq cntry.

. . to get exactly the same table.

MacInnes does this in the video, but he also explains that the data need to be weighted, and why, then runs the following:

WEIGHT BY pweight.
FREQ cntry.

This video repays repeated watching.

With his emphasis in the book on using syntax I was rather hoping he would use:

File >> New >> Syntax

to create a new **Syntax Editor** and get users to type directly into it. (He actually does this much later in the book.)

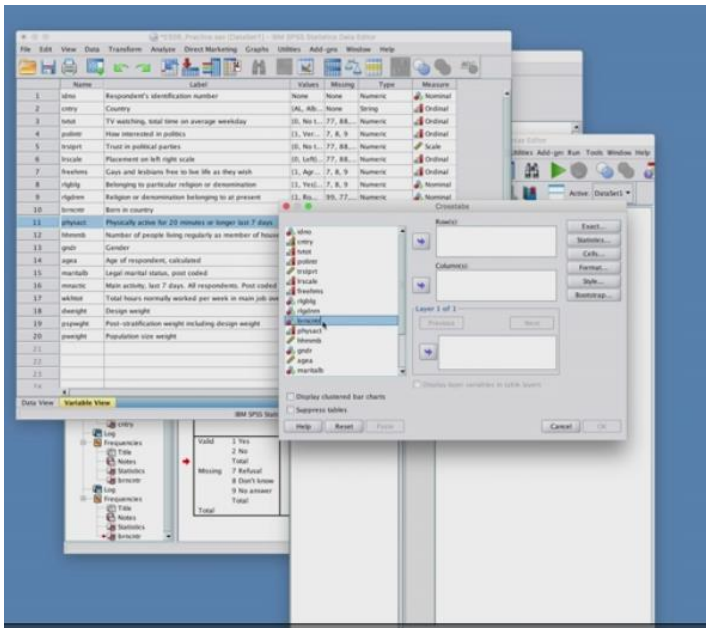
One thing that really bugs me about the **FREQUENCIES** command is that it always displays **Cumulative Percent**, even for categorical variables, when all I really need is **Frequency** and **Valid percent**, often only the latter. It would be nice if there were options to choose which columns to display, but for now I can edit the **Pivot Table** to get rid of the **Percent** and **Cumulative Percent** columns, leaving only **Frequency** and **Valid Percent**.

		Frequency	Valid Percent
Valid	AL Albania	1201	2.2
	BE Belgium	1869	3.4
	BG Bulgaria	2260	4.1
	CH Switzerland	1493	2.7
	CY Cyprus	1116	2.0
	CZ Czech Republic	2009	3.7
	DE Germany	2958	5.4
	DK Denmark	1650	3.0
	EE Estonia	2380	4.4
	ES Spain	1889	3.5
	FI Finland	2197	4.0
	FR France	1968	3.6
	GB United Kingdom	2286	4.2
	HU Hungary	2014	3.7
	IE Ireland	2628	4.8
	IL Israel	2508	4.6
	IS Iceland	752	1.4
	IT Italy	960	1.8
	LT Lithuania	2109	3.9
	NL Netherlands	1845	3.4
	NO Norway	1624	3.0
	PL Poland	1898	3.5
	PT Portugal	2151	3.9
	RU Russian Federation	2484	4.5
	SE Sweden	1847	3.4
	SI Slovenia	1257	2.3
	SK Slovakia	1847	3.4
	UA Ukraine	2178	4.0
	XK Kosovo	1295	2.4
	Total	54673	100.0

. . but it's quicker and easier with **CTABLES**.[5]

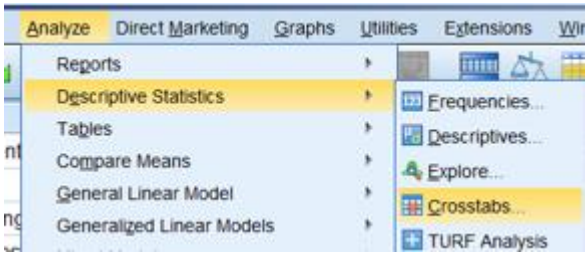
[5] See **Appendix 1: Example of CTABLES**

Video 3: Crosstab brncntr x cntry



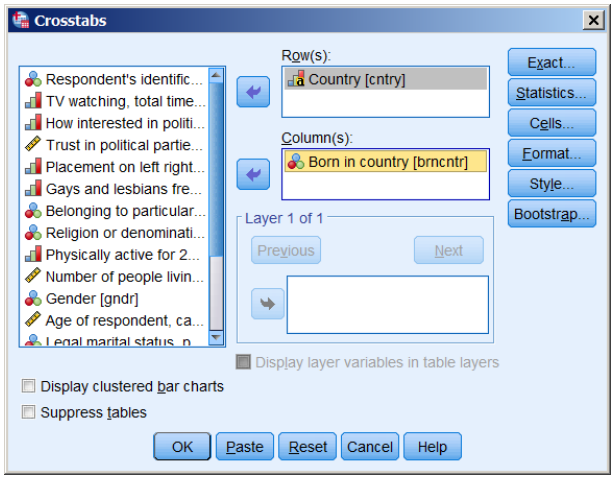
Again MacInnes uses the GUI to produce a two-way contingency table of country of survey **[cntry]** by whether respondents were born there **[brncntr]** this time with:

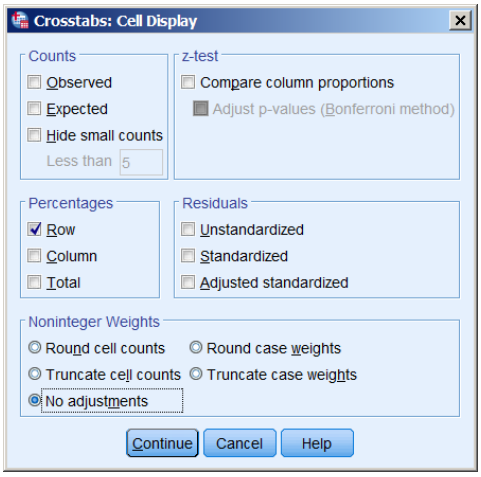
Analyze >> Descriptive Statistics >> Crosstabs



Enter **[cntry]** in the Row(s) pane,
[brncntr] in the Column(s) pane

. . and ask for **Row %**





Click on Continue then OK to produce the following table:

Country * Born in country Crosstabulation

% within Country

		Born in country		
		Yes	No	Total
Country	Albania	99.3%	0.7%	100.0%
	Belgium	86.0%	14.0%	100.0%
	Bulgaria	99.4%	0.6%	100.0%
	Switzerland	77.5%	22.5%	100.0%
	Cyprus	88.8%	11.2%	100.0%
	Czech Republic	97.4%	2.6%	100.0%
	Germany	89.9%	10.1%	100.0%
	Denmark	93.1%	6.9%	100.0%
	Estonia	83.7%	16.3%	100.0%
	Spain	88.5%	11.5%	100.0%
	Finland	95.7%	4.3%	100.0%
	France	89.4%	10.6%	100.0%
	United Kingdom	88.4%	11.6%	100.0%
	Hungary	98.8%	1.2%	100.0%
	Ireland	85.4%	14.6%	100.0%
	Israel	68.8%	31.2%	100.0%
	Iceland	94.4%	5.6%	100.0%
	Italy	92.7%	7.3%	100.0%
	Lithuania	97.1%	2.9%	100.0%
	Netherlands	90.9%	9.1%	100.0%
	Norway	87.5%	12.5%	100.0%
	Poland	98.8%	1.2%	100.0%
	Portugal	93.9%	6.1%	100.0%
	Russian Federation	94.0%	6.0%	100.0%
	Sweden	87.3%	12.7%	100.0%
	Slovenia	91.0%	9.0%	100.0%
	Slovakia	98.3%	1.7%	100.0%
	Ukraine	92.1%	7.9%	100.0%
	Kosovo	94.4%	5.6%	100.0%
Total		90.8%	9.2%	100.0%

Macinnes says the countries are listed in alphabetical order: it's not immediately obvious how, but it's because the international abbreviations used as values are not shown. (See table on page 5)

Clicking on Paste produces the following syntax:

```
CROSSTABS
  /TABLES=cntry BY brncntr
  /FORMAT=AVALUE TABLES
  /CELLS=ROW
  /COUNT ROUND CELL.
```

. . which can be done more simply by typing:

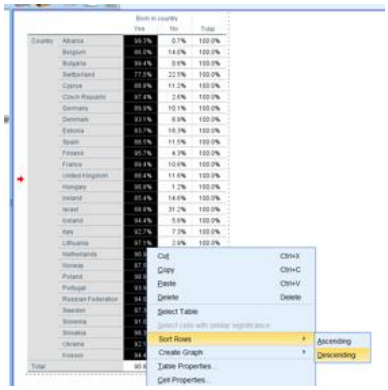
crosstabs cntry by brncntr /cells row .

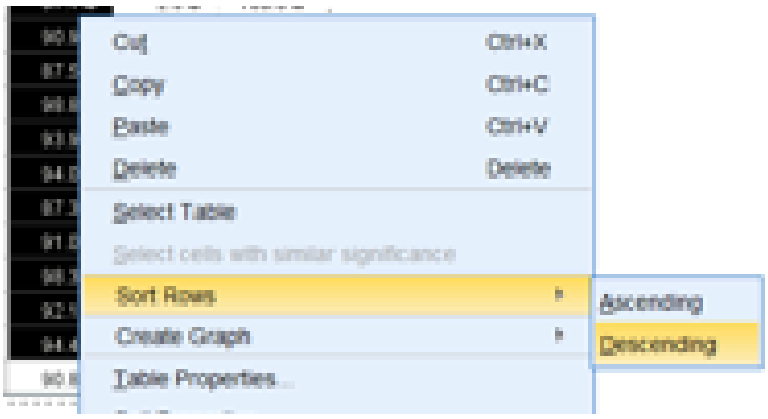
. . in the **Syntax Editor**.

MacInnes then demonstrates how the countries can be re-organised in descending order of the percentage born in the country where they now live.

Double click on the table to enter **Pivot** mode: highlight the cells in the **Yes** column (but **not** the total cell)

Sort Rows >> Descending





Country * Born in country Crosstabulation

% within Country

		Born in country		
		Yes	No	Total
Country	Bulgaria	99.4%	0.6%	100.0%
	Albania	99.3%	0.7%	100.0%
	Poland	98.8%	1.2%	100.0%
	Hungary	98.8%	1.2%	100.0%
	Slovakia	98.3%	1.7%	100.0%
	Czech Republic	97.4%	2.6%	100.0%
	Lithuania	97.1%	2.9%	100.0%
	Finland	95.7%	4.3%	100.0%
	Iceland	94.4%	5.6%	100.0%
	Kosovo	94.4%	5.6%	100.0%
	Russian Federation	94.0%	6.0%	100.0%
	Portugal	93.9%	6.1%	100.0%
	Denmark	93.1%	6.9%	100.0%
	Italy	92.7%	7.3%	100.0%
	Ukraine	92.1%	7.9%	100.0%
	Slovenia	91.0%	9.0%	100.0%
	Netherlands	90.9%	9.1%	100.0%
	Germany	89.9%	10.1%	100.0%
	France	89.4%	10.6%	100.0%
	Cyprus	88.8%	11.2%	100.0%
	Spain	88.5%	11.5%	100.0%
	United Kingdom	88.4%	11.6%	100.0%
	Norway	87.5%	12.5%	100.0%
	Sweden	87.3%	12.7%	100.0%
	Belgium	86.0%	14.0%	100.0%
	Ireland	85.4%	14.6%	100.0%
	Estonia	83.7%	16.3%	100.0%
	Switzerland	77.5%	22.5%	100.0%
	Israel	68.8%	31.2%	100.0%
Total		90.8%	9.2%	100.0%

Comment: I've been using SPSS since 1972 and this was the first time I'd used this facility, Never too late to learn something new.

Video 4: Histogram agea

Again uses the GUI to produce a frequency count of **[agea]** "*Age of respondent, calculated*".

This is one occasion where it is legitimate to produce **Cumulative percent** as it can be used to mark suitable cutting points for creating age groups. JM also produces quite a dense histogram and shows how it can be simplified by widening the intervals from single years to 15 years.

Video 5: Freehms & CIs

Demonstrates a frequency count for **[freehms]** "*Gays and lesbians free to live life as they wish*" and explains that the %% in the table are only an **estimate,** based on a **sample**, of the **parameter** for the **population** from which the sample is drawn. He then shows an Excel file in which **confidence intervals** for the estimates are calculated.

Video 6: Crosstabs & chi-square

Uses the GUI to produce two-way contingency tables of **[freehms]** "*Gays and lesbians free to live life as they wish*" tabulated by **[rlgblg]** "*Belonging to particular religion or denomination*" **[gndr**] "*Gender*" and **[lrscale]** "*Placement on left right scale*".

Shows the syntax generated by SPSS from the GUI using Analyze >> Descriptive Statistics >> Crosstabs

```
CROSSTABS
  /TABLES=freehms BY rlgblg gndr rlgdnm lrscale
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI GAMMA
  /CELLS=COUNT COLUMN
  /COUNT ROUND CELL.
```

but then shows how the same analysis can be obtained by directly written (abbreviated) syntax:

cross freehms by rlgblg gndr rlgdnm lrscale /cel row /sta chi phi gamma .

To be honest, I think it's far too early to introduce statistics like chi-square and the phi and gamma coefficients: students need time to assimilate the idea of contingency tables by experimenting with different independent variables to compare percentages and see what effect (if any) they have on attitudes.

Assuming X = <independent variable> and Y = <dependent variable> MacInnes follows what I call the "accountancy" convention, putting the dependent variable in the rows, the independent variable in the columns, and asking for row percent so that the **columns** add up to 100% (as preferred by accountants, statisticians and psychologists)

crosstabs <dependent variable> by <independent variable> /cells colpct.

Following a common "sociological" convention, I would normally put the dependent variable in the columns, the independent variable in the rows and ask for row% so that the **rows** add up to 100%.

crosstabs <independent variable> by <independent variable> /cells rowpct.

That makes it much easier to compare percentages within categories of the independent variable, but this is clearly a matter of taste.

When X has only 2 categories, as in

```
crosstabs freehms BY rlgblg gndr /cells colpct .
```

. . comparison of %% is relatively easy across rows

Crosstab

			rlgblg Belonging to particular religion or denomination 1 Yes	2 No	Total
Freehms Gays and lesbians free to live life as they wish	1 Agree strongly	Count	7572	7875	15447
		% within rlgblg	23.7%	37.9%	29.3%
	2 Agree	Count	11026	6532	17558
		% within rlgblg	34.5%	31.5%	33.3%
	3 Neither agree nor disagree	Count	5186	2582	7768
		% within rlgblg	16.2%	12.4%	14.7%
	4 Disagree	Count	3824	1816	5640
		% within rlgblg	12.0%	8.7%	10.7%
	5 Disagree strongly	Count	4339	1956	6295
		% within rlgblg	13.6%	9.4%	11.9%
Total		Count	31947	20761	52708
		% within	100.0%	100.0%	100.0%

. . but when X has many categories, it becomes much more difficult as in

```
crosstabs freehms BY rlgdnm
        /cells colpct .
```

Crosstab

			rlgdnm Religion or denomination belonging to at present								
			1 Roman Catholic	2 Protestant	3 Eastern Orthodox	4 Other Christian denomination	5 Jewish	6 Islamic	7 Eastern religions	8 Other non-Christian religions	Total
freehms Gays and lesbians free to live life as they wish	1 Agree strongly	Count	4560	1514	638	169	232	255	116	53	7537
		% within rlgdnm Religion or denomination belonging to at present	28.6%	29.6%	9.0%	29.0%	45.0%	12.3%	36.4%	36.6%	23.7%
	2 Agree	Count	6392	2441	1231	154	151	445	123	34	10971
		% within rlgdnm Religion or denomination belonging to at present	40.0%	47.8%	17.4%	26.5%	29.3%	21.5%	38.6%	23.4%	34.5%
	3 Neither agree nor disagree	Count	2406	579	1518	105	47	428	44	24	5151
		% within rlgdnm Religion or denomination belonging to at present	15.1%	11.3%	21.4%	18.0%	9.1%	20.7%	13.8%	16.6%	16.2%
	4 Disagree	Count	1393	349	1541	62	33	387	27	24	3816
		% within rlgdnm Religion or denomination belonging to at present	8.7%	6.8%	21.7%	10.7%	6.4%	18.7%	8.5%	16.6%	12.0%
	5 Disagree strongly	Count	1210	225	2160	92	52	555	9	10	4313
		% within rlgdnm Religion or denomination belonging to at present	7.6%	4.4%	30.5%	15.8%	10.1%	26.8%	2.8%	6.9%	13.6%
Total		Count	15961	5108	7088	582	515	2070	319	145	31788
		% within rlgdnm Religion or denomination belonging to at present	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

I also tend to leave out the counts as they just clutter up the tables, but this means there is no N shown as a base for %.

In fact **CTABLES** can produce much clearer and less cluttered output with row% in the cells and base Ns in the Total column:

		freehms Gays and lesbians free to live life as they wish					
		1 Agree strongly Row N %	2 Agree Row N %	3 Neither agree nor disagree Row N %	4 Disagree Row N %	5 Disagree strongly Row N %	Total Count
gndr Gender	1 Male	26.1%	34.4%	14.9%	11.8%	12.8%	25274
	2 Female	31.8%	32.0%	14.8%	9.9%	11.5%	27991

		freehms Gays and lesbians free to live life as they wish					
		1 Agree strongly Row N %	2 Agree Row N %	3 Neither agree nor disagree Row N %	4 Disagree Row N %	5 Disagree strongly Row N %	Total Count
rlgblg Belonging to particular religion or denomination	1 Yes	23.7%	34.5%	16.2%	12.0%	13.6%	31947
	2 No	37.9%	31.5%	12.4%	8.7%	9.4%	20762

This orientation makes it easier to interpret and to compute epsilons (percentage point differences).

The syntax for **CTABLES** can become quite complicated for newbies, but see Appendix 1 for a worked example.

When the independent variable is dichotomous (has only two categories) the epsilons can be used for a special analysis technique known as **Elaboration** (see 3.2 Three (or more) variables and the worked example in Appendix 2). .

End of: **4.1.1: Overview of video tutorials 1 – 6**

Back to: MacInnes (2017)

Appendix 1: Example of CTABLES

For counts only:

ctables /tab cntry.

To get **valid percent** and **totals** it's a bit more complicated: I had to use the GUI, but at least I didn't get the pesky **cumulative percent**

Syntax generated by Paste

```
CTABLES
  /VLABELS VARIABLES=cntry DISPLAY=DEFAULT
  /TABLE cntry [COUNT 40.0, COLPCT.COUNT F40.1]
  /CATEGORIES VARIABLES=cntry ORDER=A KEY=VALUE EMPTY=INCLUDE  TOTAL = YES
```

		Count
Country	Albania	1201
	Belgium	1869
	Bulgaria	2260
	Switzerland	1493
	Cyprus	1116
	Czech Republic	2009
	Germany	2958
	Denmark	1650
	Estonia	2380
	Spain	1889
	Finland	2197
	France	1968
	United Kingdom	2286
	Hungary	2014
	Ireland	2628
	Israel	2508
	Iceland	752
	Italy	960
	Lithuania	2109
	Netherlands	1845
	Norway	1624
	Poland	1898
	Portugal	2151
	Russian Federation	2484
	Sweden	1847
	Slovenia	1257
	Slovakia	1847
	Ukraine	2178
	Kosovo	1295

		Count	Column N %
Country	Albania	1201	2.2
	Belgium	1869	3.4
	Bulgaria	2260	4.1
	Switzerland	1493	2.7
	Cyprus	1116	2.0
	Czech Republic	2009	3.7
	Germany	2958	5.4
	Denmark	1650	3.0
	Estonia	2380	4.4
	Spain	1889	3.5
	Finland	2197	4.0
	France	1968	3.6
	United Kingdom	2286	4.2
	Hungary	2014	3.7
	Ireland	2628	4.8
	Israel	2508	4.6
	Iceland	752	1.4
	Italy	960	1.8
	Lithuania	2109	3.9
	Netherlands	1845	3.4
	Norway	1624	3.0
	Poland	1898	3.5
	Portugal	2151	3.9
	Russian Federation	2484	4.5
	Sweden	1847	3.4
	Slovenia	1257	2.3
	Slovakia	1847	3.4
	Ukraine	2178	4.0
	Kosovo	1295	2.4
	Total	54673	100.0

My tutorial 3.2.1.4 Elaboration 4 (Income differences 2009 – 2014 CTABLES) has fully worked examples from the British Social Attitudes Survey (cumulative waves 2009 – 2014) using **CTABLES** to examine the model:

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

Y = Dependent variable
X = Independent variable
T = Test variable(s)

Dependent variable Y:	Earnings from paid work (Quartiles)
Independent variable X:	Sex of respondent
Test variable T:	Hours worked (Under 30, 30+)

```
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 .
```

		earngrp Quartile group of R's earnings from paid work				
		1 Q1	2 Q2	3 Q3	4 Q4	Total
		%	%	%	%	n= 100%
Rsex Person 1 SEX	1 Male	13.0	25.9	29.0	32.1	4299
	2 Female	35.6	29.3	21.2	13.9	4490
	Epsilon	**-22.6**	**-3.4**	**+7.8**	**+18.2**	

		earngrp Quartile group of R's earnings from paid work				
		1 Q1	2 Q2	3 Q3	4 Q4	Total
		%	%	%	%	n= 100%
workmode R working full- or part-time	1 Full time	10.5	29.9	31.2	28.3	5459
	2 Part time	59.3	23.8	10.5	6.5	1999
	Epsilon	**-48.8**	**+6.1**	**+20.7**	**+21.8**	

		workmode R working full- or part-time									
		1 Full time					2 Part time				
		earngrp Quartile group of R's earnings from paid work					earngrp Quartile group of R's earnings from paid work				
		1 Q1	2 Q2	3 Q3	4 Q4	Total	1 Q1	2 Q2	3 Q3	4 Q4	Total
		%	%	%	%	n= 100%	%	%	%	%	n= 100%
Rsex Person 1 SEX	1 Male	6.1	26.8	32.3	34.9	3028	45.0	24.0	15.6	15.4	416
	2 Female	16.0	33.9	29.9	20.2	2431	63.0	23.7	9.2	4.1	1583
	Epsilon	**-9.9**	**-7.1**	**+2.3**	**+14.7**		**-18.1**	**+0.3**	**+6.5**	**+11.3**	

Pencil and paper exercise: complete the blank table:

Difference in earnings of men and women controlling for hours worked (group Q4 only)

	All	Full time	Part time	Epsilon
All	% ( )	% ( )	% ( )	
Men	% ( )	% ( )	% ( )	
Women	% ( )	% ( )	% ( )	
Epsilon				

Step by step to arrive at:

	All	Full time	Part time	Epsilon
All	22.5% (7458)	28.3% (5459)	6.52% (1999)	**+13.8**
Men	32.5% (3444)	34.9% (3028)	15.4% (416)	**+19.5**
Women	13.8% (4014)	24.2% (2431)	4.1% (1583)	**+20.1**
Epsilon	**+18.7**	**+10.7**	**+11.3**	

Students find %% easier to understand, but the underlying logic is a platform for more advanced modelling.

The syntax for the ESS6 sample would be:

```
ctables /TABLE gndr BY freehms [ROWPCT.COUNT TOTALS [COUNT]]
 /CATEGORIES VARIABLES= freehms TOTAL=YES
 /TABLE rlgblg BY freehms [ROWPCT.COUNT TOTALS [COUNT]]
 /CATEGORIES VARIABLES= freehms TOTAL=YES .
```

Appendix 2: Elaboration

SPSS syntax suggested by Bruce Weaver for a new macro: **ELABORATE**

I've modified it for Y = freehms, X = gndr

```
* Encoding: UTF-8.
DEFINE elaborate
( Y = !CHAREND('/') /
 RowVar = !CHAREND('/') /
 ColVar = !CMDEND ).
CTABLES
 /TABLE !RowVar [C] > !ColVar [C] by !Y [c][ROWPCT.totaln, COUNT]
 /SLABELS POSITION=ROW VISIBLE=NO
 /CATEGORIES VARIABLES= !y TOTAL=YES
 /TITLES CAPTION ='NOTE: Cells show ROWPCT and COUNT'.
!ENDDEFINE.

* Read in some data to illustrate.
* Modify path on the FILE HANDLE command as needed.

FILE HANDLE TheDataFile
 /NAME='C:\Users\John Hall\Desktop\ch4jfh.sav' .

NEW FILE.
GET FILE = "TheDataFile".
* Call the macro.
elaborate Y = freehms / RowVar = gndr/ ColVar = rlgblg.
```

Example using data from ESS_Practice.sav

Model: Dependent Y = freehms Independent X = gndr Control T = rlgblg

Had to fiddle with the table to get this and have yet to try losing the % sign in each cell, and to put brackets round the n. Cells show % of base n.

				freehms Gays and lesbians free to live life as they wish					
				1 Agree strongly	2 Agree	3 Neither agree nor disagree	4 Disagree	5 Disagree strongly	Total
gndr Gender	1 Male	rlgblg Belonging to particular religion or denomination	1 Yes	19.5%	34.3%	15.6%	12.3%	12.9%	100.0%
				2856	5036	2288	1806	1890	13876
			2 No	31.9%	30.9%	12.3%	9.5%	10.6%	100.0%
				3707	3593	1426	1105	1234	11064
	2 Female	rlgblg Belonging to particular religion or denomination	1 Yes	24.1%	30.6%	14.8%	10.3%	12.6%	100.0%
				4705	5972	2893	2008	2449	18027
			2 No	40.8%	28.9%	11.4%	7.0%	7.1%	100.0%
				4153	2939	1156	711	723	9682

NOTE: Cells show ROWPCT and COUNT (base n for %)