3.5.2.5: Data checks 2 – Sexism

[Screenshots: SPSS15, last updated: 24 Aug 2011, previous version 1 October 2009]

Exemplar: Fifth form survey fifthx.sav
Variable to be derived: Negative attitudes to women (Sexism)
Source variables: [9 items selected from] v248 to v261
Previous tutorial: 3.5.2.4 The COMPUTE command 1 - Attachment to status quo

Question Q.33 is slightly more complex. It consists of 14 statements measuring opinions about women, some negative, some neutral, some positive, with which pupils can agree or disagree on a 4-point scale. (Codes 1 to 4 on columns 48 to 61 of record 2, read into SPSS using positional variable naming convention as **v248 to v261**: SPSS automatically generates intermediate variable names)

Q33. Here are some statements made about women, We would like to know if you agree or disagree with them. (Please put a ring round the number which indicates your answer).

	Disagree Strongly	Disagree	Agree	Agree Strongly	
a) Careers are fine for women but real fulfilment is a home and kids.	1	2	3	4	(48)
b) Women should not expect men to pay for them when dating etc.	1	2	3	4	(49)
c) Half of all top jobs should be reserved for women.	1	2	3	4	(50)
d) It is a good thing that women can become airline pilots, plumbers etc.	1	2	3	4	(51)
e) Women are too emotional.	1	2	3	4	(52)
f) Women are not as ambitious as men.	1	2	3	4	(53)
g) Women are as intelligent as men.	1	2	3	4	(54)
h) Women do not need to be beautiful to be successful	1	2	3	4	(55)
j) Husbands rather than wives should have the final voice in family matters.	1	2	3	4	(56)
k) There is no difference in brain-power between men and women	1	2	3	4	(57)
l) If women are paid as much as men they should pay for themselves when dating etc.	1	2	3	4	(58)
m) Women should get equal pay for doing the same work as men.	1	2	3	4	(59)
n) Beauty contests are degrading to women and should stop.	1	2	3	4	(60)
o) Romantic love is dead	1	2	3	4	(61)

Nine[1] of these items, five of which are negative (a,e,f,j,o) and four positive (d,h,m,n) will be used to construct an index of "Sexism" (negative attitudes to women). Here is an extract from the user manual showing the original frequency counts for the responses to these questions.

Q.33 Here are some statements made about women. We would like to know if you agree or disagree with them. (Please put a ring round the number which indicates your answer).

Code:	Disagree Strongly 1	Disagree 2	Agree 3	Agree Strongly 4	Blank -1*	
a) Careers are fine for women but real fulfilment is a home and kids	39	54	27	12	10	(V248)
b) Women should not expect men to pay for them when dating etc.	16	35	56	25	10	(V249)
c) Half of all top jobs should be reserved for women	23	47	48	15	9	(V250)
d) It is a good thing that women can become airline pilots, plumbers etc.	17	20	71	23	11	(V251)
e) Women are too emotional	21	47	48	15	11	(V252)
f) Women are not as ambitious as men	37	62	23	6	14	(V253)
g) Women are as intelligent as men	8	9	68	47	10	(V254)
h) Women do not need to be beautiful to be successful	8	21	74	31	8	(V255)
j) Husbands rather than wives should have the final voice in family matters.	26	55	33	15	13	(V256)
k) There is no difference in brainpower between men and women	7	16	71	33	15	(V257)
l) If women are paid as much as men they should pay for themselves when dating etc.	7	41	52	26	15	(V258)
m) Women should get equal pay for doing the same work as men	4	6	67	59	6	(V259)
n) Beauty contests are degrading to women and should stop	28	74	20	8	12	(V260)
o) Romantic love is dead	44	62	12	13	11	(V261)

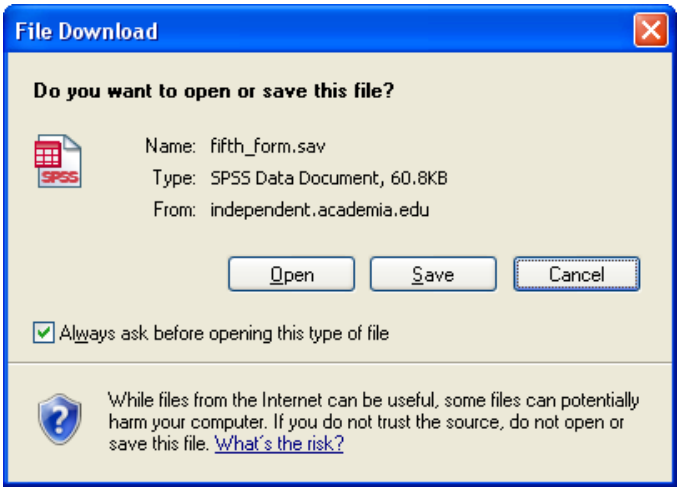
In the following exercise we are going to use the SPSS procedure **COUNT** to create a new, **derived variable** containing, for each pupil, a count of the **agree** or **agree strongly** responses (codes 3 or 4) to the negative items (a,e,f,j,o) together with the **disagree** or **disagree strongly** responses (codes 1 or 2) to the positive items (d,h,m,n). This will yield a score in the range 0 to 9 where 9 indicates high sexism.

[1] These nine items were chosen after multivariate analysis involving correlation to examine internal structure of the construct : we'll come to that later.

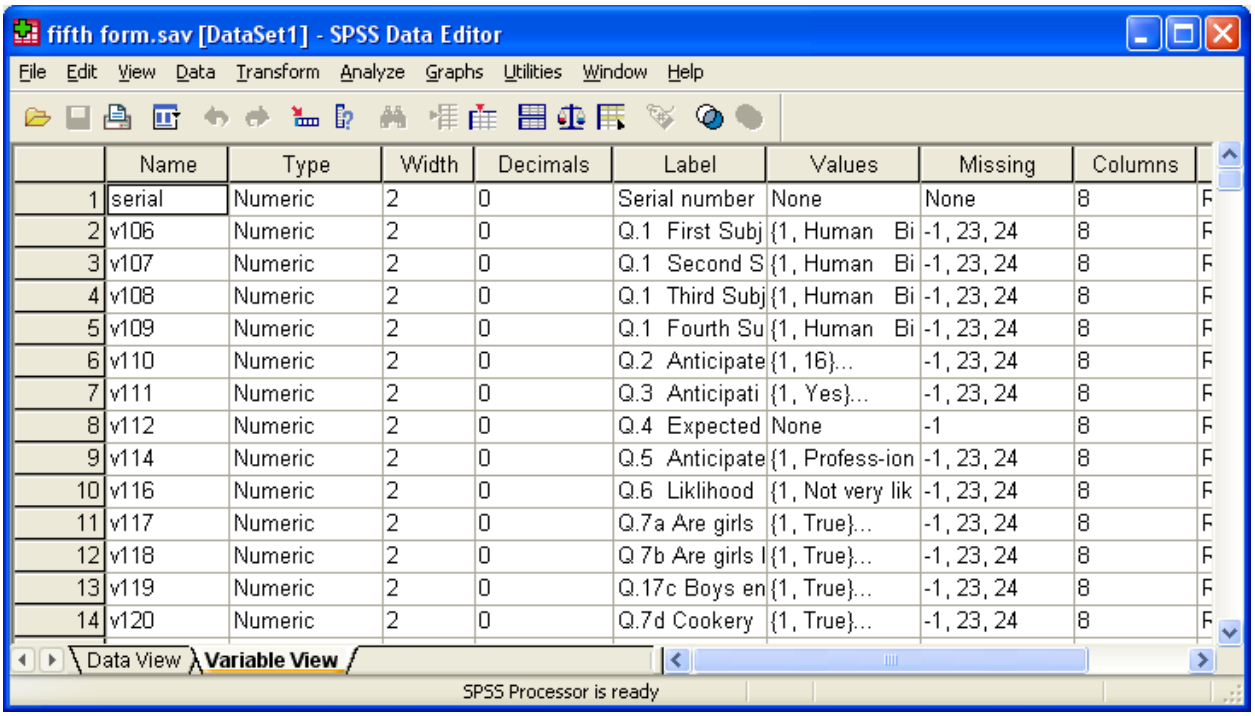
First, however, it is essential to check the original data, and a lot safer, especially in situations like this when working with batteries in which some items may need values to be reversed or recoded before generating scores.

[NB: You won't actually be able to do this next bit without immediate direct access to SPSS: don't worry just follow the tutorial anyway.]

Start by downloading the SPSS saved file Fifth Form Survey



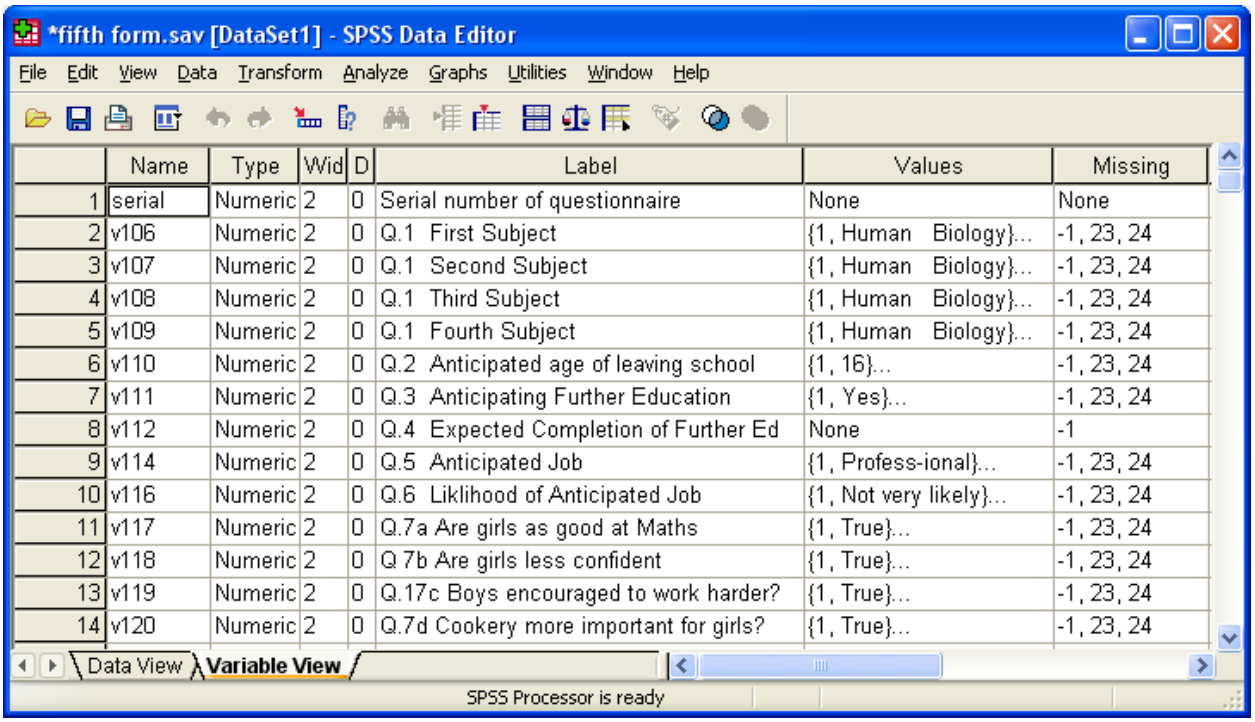
Click on **Open** to get[2]:



	Name	Type	Width	Decimals	Label	Values	Missing	Columns
1	serial	Numeric	2	0	Serial number	None	None	8
2	v106	Numeric	2	0	Q.1 First Subj	{1, Human Bi	-1, 23, 24	8
3	v107	Numeric	2	0	Q.1 Second S	{1, Human Bi	-1, 23, 24	8
4	v108	Numeric	2	0	Q.1 Third Subj	{1, Human Bi	-1, 23, 24	8
5	v109	Numeric	2	0	Q.1 Fourth Su	{1, Human Bi	-1, 23, 24	8
6	v110	Numeric	2	0	Q.2 Anticipate	{1, 16}...	-1, 23, 24	8
7	v111	Numeric	2	0	Q.3 Anticipati	{1, Yes}...	-1, 23, 24	8
8	v112	Numeric	2	0	Q.4 Expected	None	-1	8
9	v114	Numeric	2	0	Q.5 Anticipate	{1, Profess-ion	-1, 23, 24	8
10	v116	Numeric	2	0	Q.6 Liklihood	{1, Not very lik	-1, 23, 24	8
11	v117	Numeric	2	0	Q.7a Are girls	{1, True}...	-1, 23, 24	8
12	v118	Numeric	2	0	Q 7b Are girls l	{1, True}...	-1, 23, 24	8
13	v119	Numeric	2	0	Q.17c Boys en	{1, True}...	-1, 23, 24	8
14	v120	Numeric	2	0	Q.7d Cookery	{1, True}...	-1, 23, 24	8

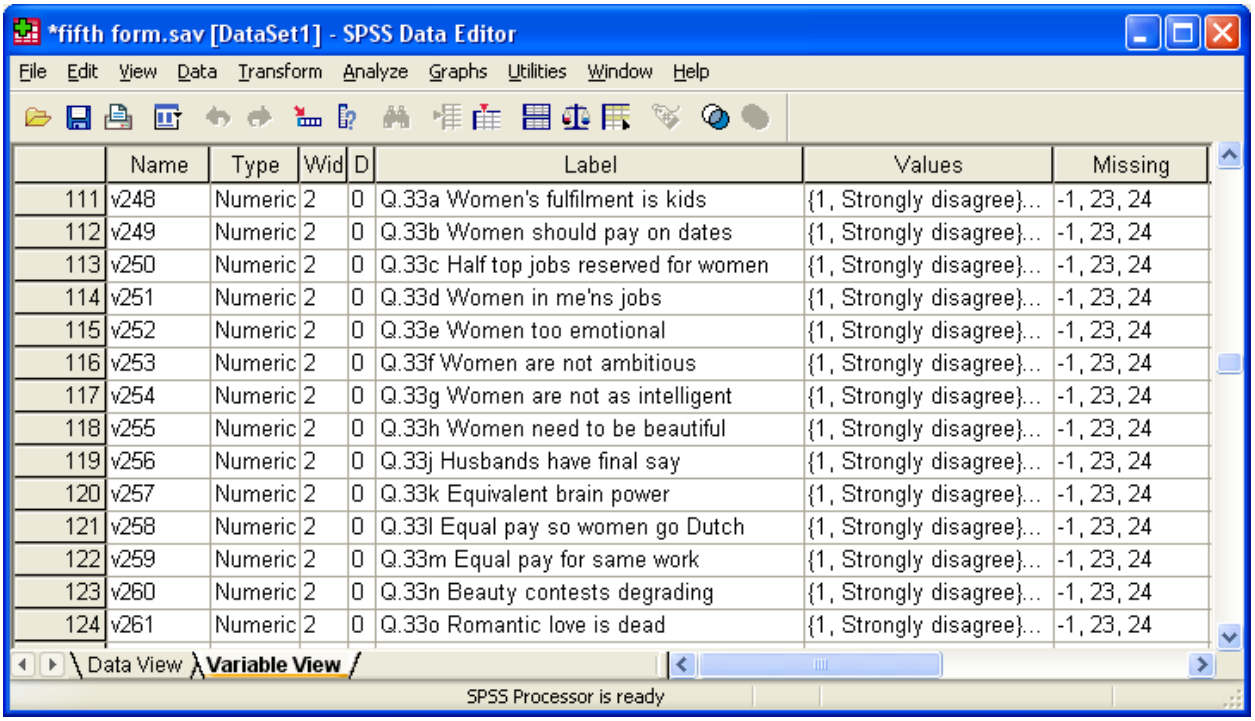
[2] The size of the Data Editor displayed will depend on how it was left when last used. You can drag the corners of the pane to reduce or enlarge it, and the edges up or down and left or right, using the left mouse button. This pane is adjusted to display only the first few variables in the file.

Use the mouse to drag the column separators around to make the display easier to read, then reduce the pane horizontally to include only essential information.



	Name	Type	Wid	D	Label	Values	Missing
1	serial	Numeric	2	0	Serial number of questionnaire	None	None
2	v106	Numeric	2	0	Q.1 First Subject	{1, Human Biology}...	-1, 23, 24
3	v107	Numeric	2	0	Q.1 Second Subject	{1, Human Biology}...	-1, 23, 24
4	v108	Numeric	2	0	Q.1 Third Subject	{1, Human Biology}...	-1, 23, 24
5	v109	Numeric	2	0	Q.1 Fourth Subject	{1, Human Biology}...	-1, 23, 24
6	v110	Numeric	2	0	Q.2 Anticipated age of leaving school	{1, 16}...	-1, 23, 24
7	v111	Numeric	2	0	Q.3 Anticipating Further Education	{1, Yes}...	-1, 23, 24
8	v112	Numeric	2	0	Q.4 Expected Completion of Further Ed	None	-1
9	v114	Numeric	2	0	Q.5 Anticipated Job	{1, Profess-ional}...	-1, 23, 24
10	v116	Numeric	2	0	Q.6 Liklihood of Anticipated Job	{1, Not very likely}...	-1, 23, 24
11	v117	Numeric	2	0	Q.7a Are girls as good at Maths	{1, True}...	-1, 23, 24
12	v118	Numeric	2	0	Q.7b Are girls less confident	{1, True}...	-1, 23, 24
13	v119	Numeric	2	0	Q.17c Boys encouraged to work harder?	{1, True}...	-1, 23, 24
14	v120	Numeric	2	0	Q.7d Cookery more important for girls?	{1, True}...	-1, 23, 24

Scroll down to find variables **v248** to **v261** containing the data for Q.33 (a) to (o) and reduce the pane vertically to show only these:



	Name	Type	Wid	D	Label	Values	Missing
111	v248	Numeric	2	0	Q.33a Women's fulfilment is kids	{1, Strongly disagree}...	-1, 23, 24
112	v249	Numeric	2	0	Q.33b Women should pay on dates	{1, Strongly disagree}...	-1, 23, 24
113	v250	Numeric	2	0	Q.33c Half top jobs reserved for women	{1, Strongly disagree}...	-1, 23, 24
114	v251	Numeric	2	0	Q.33d Women in me'ns jobs	{1, Strongly disagree}...	-1, 23, 24
115	v252	Numeric	2	0	Q.33e Women too emotional	{1, Strongly disagree}...	-1, 23, 24
116	v253	Numeric	2	0	Q.33f Women are not ambitious	{1, Strongly disagree}...	-1, 23, 24
117	v254	Numeric	2	0	Q.33g Women are not as intelligent	{1, Strongly disagree}...	-1, 23, 24
118	v255	Numeric	2	0	Q.33h Women need to be beautiful	{1, Strongly disagree}...	-1, 23, 24
119	v256	Numeric	2	0	Q.33j Husbands have final say	{1, Strongly disagree}...	-1, 23, 24
120	v257	Numeric	2	0	Q.33k Equivalent brain power	{1, Strongly disagree}...	-1, 23, 24
121	v258	Numeric	2	0	Q.33l Equal pay so women go Dutch	{1, Strongly disagree}...	-1, 23, 24
122	v259	Numeric	2	0	Q.33m Equal pay for same work	{1, Strongly disagree}...	-1, 23, 24
123	v260	Numeric	2	0	Q.33n Beauty contests degrading	{1, Strongly disagree}...	-1, 23, 24
124	v261	Numeric	2	0	Q.33o Romantic love is dead	{1, Strongly disagree}...	-1, 23, 24

Now you know why positional variable names and informative variable labels are so important! They are so much easier to find in the Data Editor than mnemonics and you (and, more important, others) can work straight from the original questionnaire to the data set and *vice versa*. If you insist on mnemonics, at least think about using question numbers as variable names, but don't leave the full stops in.

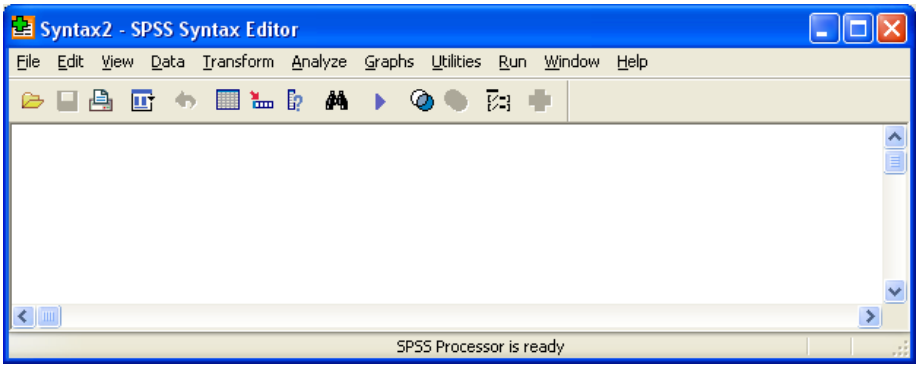
Nine of these items, five of which are negative (a,e,f,j,o) **V248 v252 v253 v256 v261** and four positive (d,h,m,n) **v251 v255 v259 v260** will be used to construct a crude index of "Sexism". Before that, however, we need to check these data against the original user manual, so we need to look at the distribution of values for these nine variables.

The syntax way

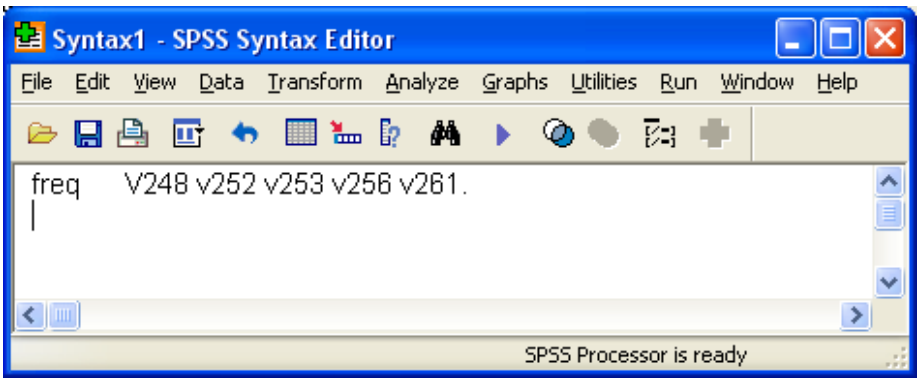
It's actually quicker to tabulate the whole set with **freq v248 to v261** and pick out the variables we want, but for now we'll do it using only the nine variables needed to construct the index. Also, to keep the output manageable (and fit on an A4 page) we'll tabulate the negative and positive items separately, starting with the negative items.

For a small job like this, open a new SPSS syntax file with:

File
…New
…Syntax



…and write **freq V248 v252 v253 v256 v261.** in the box. **Don't forget the full stop!**



Leave the cursor in or on the line and press **[CTRL]+R** to run the job.

SPSS will produce the following tables:

Statistics

		Q.33a Women's fulfilment is kids	Q.33e Women too emotional	Q.33f Women are not ambitious	Q.33j Husbands have final say	Q.33o Romantic love is dead
N	Valid	132	131	128	129	131
	Missing	10	11	14	13	11

This is a summary table for all five (negative) variables showing the number of valid and missing cases for each variable. It is followed by separate frequency tables, one for each variable:

Q.33a Women's fulfilment is kids

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	39	27.5	29.5	29.5
	Disagree	54	38.0	40.9	70.5
	Agree	27	19.0	20.5	90.9
	Agree strongly	12	8.5	9.1	100.0
	Total	132	93.0	100.0	
Missing	-1	10	7.0		
Total		142	100.0		

Q.33e Women too emotional

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	21	14.8	16.0	16.0
	Disagree	47	33.1	35.9	51.9
	Agree	48	33.8	36.6	88.5
	Agree strongly	15	10.6	11.5	100.0
	Total	131	92.3	100.0	
Missing	-1	11	7.7		
Total		142	100.0		

Q.33f Women are not ambitious

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	37	26.1	28.9	28.9
	Disagree	62	43.7	48.4	77.3
	Agree	23	16.2	18.0	95.3
	Agree strongly	6	4.2	4.7	100.0
	Total	128	90.1	100.0	
Missing	-1	14	9.9		
Total		142	100.0		

Q.33o Romantic love is dead

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	44	31.0	33.6	33.6
	Disagree	62	43.7	47.3	80.9
	Agree	12	8.5	9.2	90.1
	Agree strongly	13	9.2	9.9	100.0
	Total	131	92.3	100.0	
Missing	-1	11	7.7		
Total		142	100.0		

From these tables we can see that there are 142 cases in the file, that '**-1**' has been declared as

a missing value for all five variables and that there are no values outside the range of 1 to 4.

You can produce the same tables using the drop-down menus, but this takes a lot longer and is cumbersome, confusing and frustrating. You can also produce a summary table using SPSS syntax, but it looks very complicated for beginners (and also for me) so it's easier for now to do it with the menus. Here's the syntax generated by SPSS.

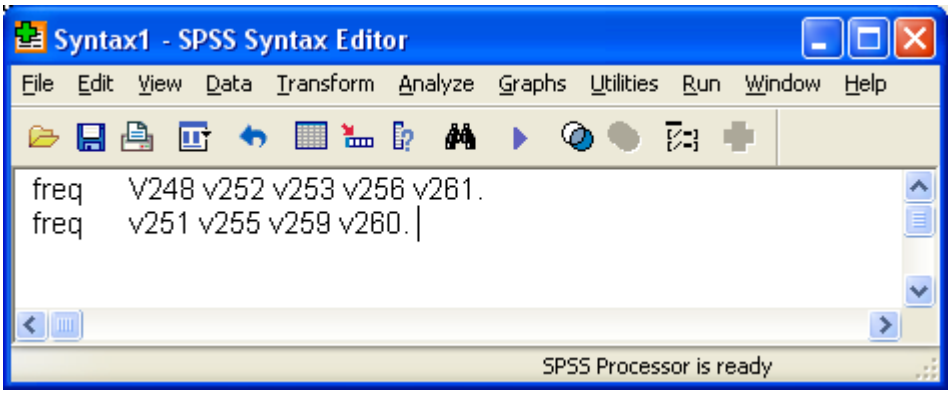
```
TABLES
  /FORMAT BLANK MISSING('.') /TABLES
  (LABELS)  BY
  ( v248 + v252 + v253 + v256 + v261 )
  /STATISTICS COUNT ((F5.0) 'Count' ) .
```

…which produces:

	Q.33a Women's fulfilment is kids	Q.33e Women too emotional	Q.33f Women are not ambitious	Q.33j Husbands have final say	Q.33o Romantic love is dead
	Count	Count	Count	Count	Count
Strongly disagree	39	21	37	26	44
Disagree	54	47	62	55	62
Agree	27	48	23	33	12
Agree strongly	12	15	6	15	13

It's easier with the menus, but could be very fiddly if the variables were spread all over the file. For analytical purposes, to compare distributions, we'd be better off with % rather than n in the cells, but this time we're checking frequencies against the user manual.

Now for the positive items: retrieve the syntax file and write **freq v251 v255 v259 v260.** on the next line:



to produce:

Statistics

		Q.33d Women in me'ns jobs	Q.33h Women need to be beautiful	Q.33m Equal pay for same work	Q.33n Beauty contests degrading
N	Valid	131	134	136	130
	Missing	11	8	6	12

As before, the summary table shows the number of valid and missing cases for each of the four variables.

It is followed by separate frequency tables, one for each variable.

Q.33d Women in men's jobs

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	17	12.0	13.0	13.0
	Disagree	20	14.1	15.3	28.2
	Agree	71	50.0	54.2	82.4
	Agree strongly	23	16.2	17.6	100.0
	Total	131	92.3	100.0	
Missing	-1	11	7.7		
Total		142	100.0		

Q.33h Women need to be beautiful

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	8	5.6	6.0	6.0
	Disagree	21	14.8	15.7	21.6
	Agree	74	52.1	55.2	76.9
	Agree strongly	31	21.8	23.1	100.0
	Total	134	94.4	100.0	
Missing	-1	8	5.6		
Total		142	100.0		

Q.33m Equal pay for same work

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	4	2.8	2.9	2.9
	Disagree	6	4.2	4.4	7.4
	Agree	67	47.2	49.3	56.6
	Agree strongly	59	41.5	43.4	100.0
	Total	136	95.8	100.0	
Missing	-1	6	4.2		
Total		142	100.0		

Q.33n Beauty contests degrading

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	28	19.7	21.5	21.5
	Disagree	74	52.1	56.9	78.5
	Agree	20	14.1	15.4	93.8
	Agree strongly	8	5.6	6.2	100.0
	Total	130	91.5	100.0	
Missing	-1	12	8.5		
Total		142	100.0		

You can produce the same tables using the drop-down menus, but this takes a lot longer and can be cumbersome, confusing and frustrating.

You can also produce a summary table using SPSS syntax, but it looks very complicated for beginners (and also for me) so it's easier for now to do it with the menus. Here's the syntax generated by SPSS.

```
TABLES
  /FORMAT BLANK MISSING('.') /TABLES
  (LABELS)  BY
  ( v251 + v255 + v259 + v260 )
  /STATISTICS COUNT ((F5.0) 'Count' ) .
```

	Q.33d Women in men's jobs	Q.33h Women need to be beautiful	Q.33m Equal pay for same work	Q.33n Beauty contests degrading
	Count	Count	Count	Count
Strongly disagree	17	8	4	28
Disagree	20	21	6	74
Agree	71	74	67	20
Agree strongly	23	31	59	8

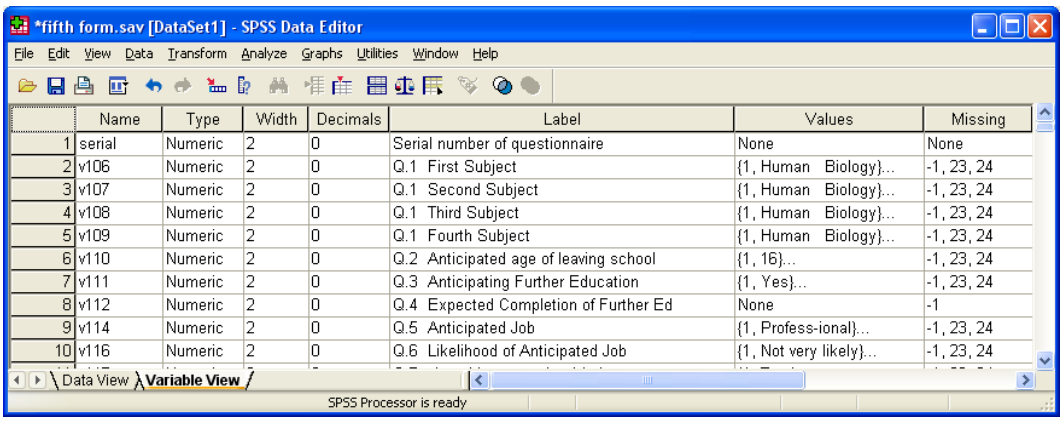
Again it's easier with the drop-down menus, but this time it was more fiddly as the variables were not adjacent in the file, but we've lost the information on missing values. Again for analytical purposes, to compare distributions, we'd be better off with % rather than n in the cells, but this time we're checking frequencies against the user manual.

In this particular data set the variables are all actually still in their original form, but it's important always to carry out such checks when creating derived variables as sometimes the original variables may have been changed or permanently recoded and any derived variables based on them may consequently be rendered meaningless.

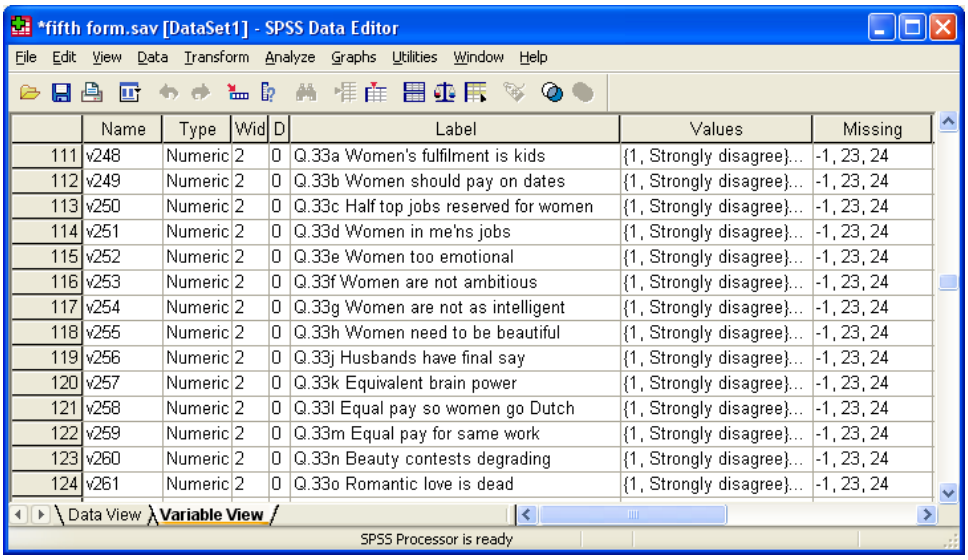
For an instance of this actually happening, see the slide show for my repetition of the exercises from Julie Pallant's **SPSS Survival Manual** (2nd edition 2005) Old Dog, Old Tricks 5: Exercises from SPSS Survival Manual or the full paper Old Dog, Old Tricks: Using SPSS Syntax to Avoid the Mouse Trap (pages 45-76)

The point-and-click way

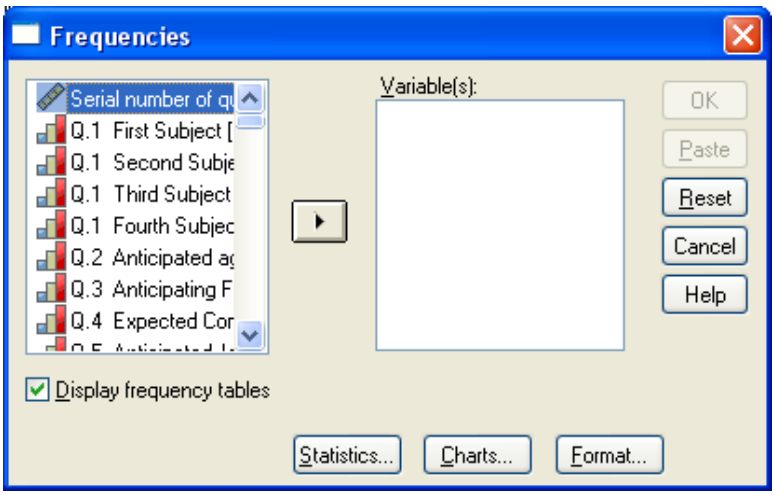
Here's how to get the same tables using the drop-down menus. Go back to the Data Editor:



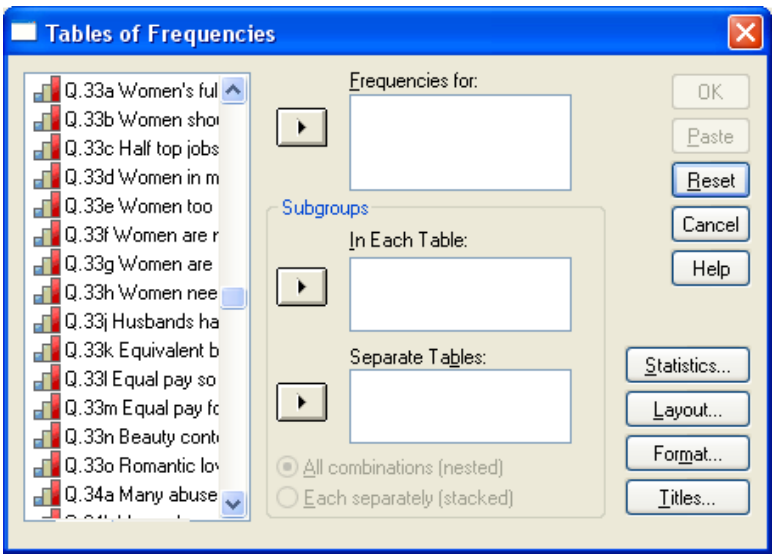
…and scroll to find **V248…V261**:



Analyze
…Descriptive statistics
…Frequencies

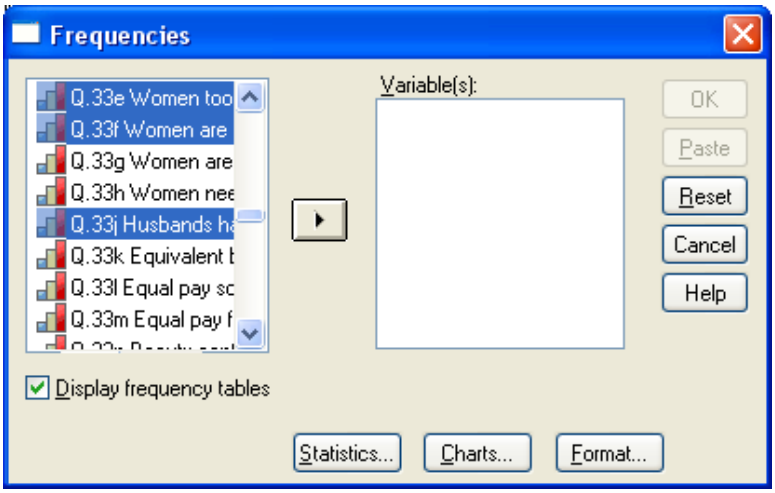


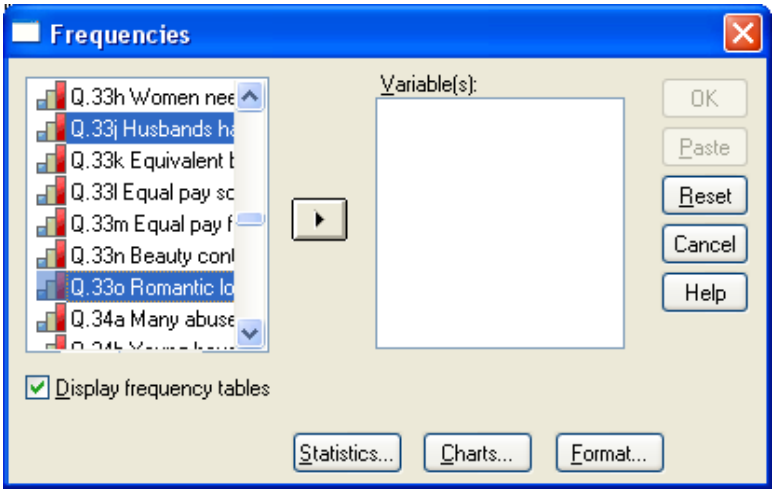
Scroll to find the variables we need:



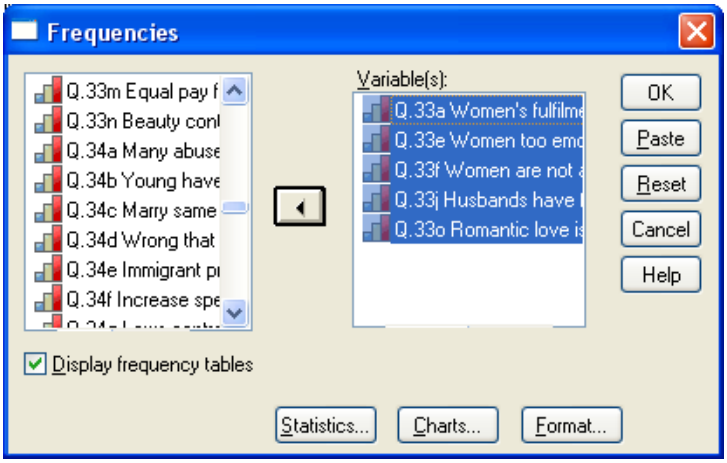
Negative items

The variables are not adjacent, so you need to drag them into the **Frequencies for** box one at a time or you can hold down **[CTRL**] and click on each of them to highlight them:





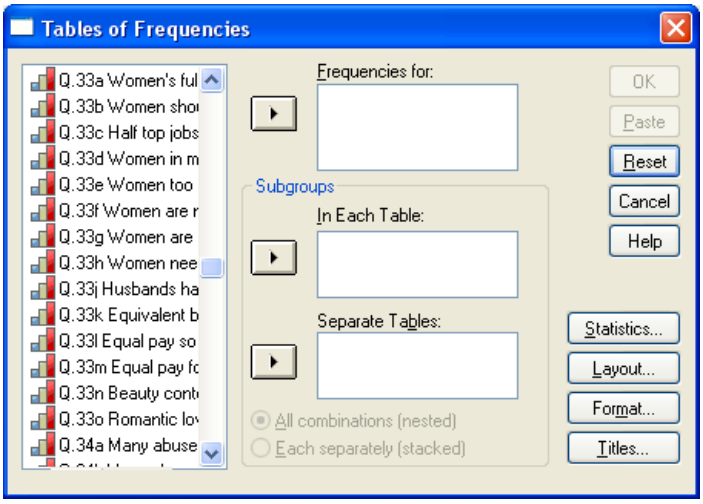
and drag all five across in one go:



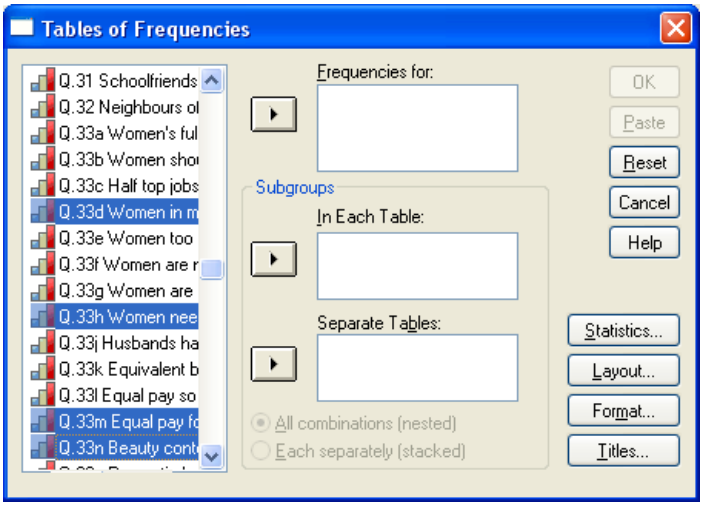
Press **OK** to run and get tables for the negative items.

Positive items

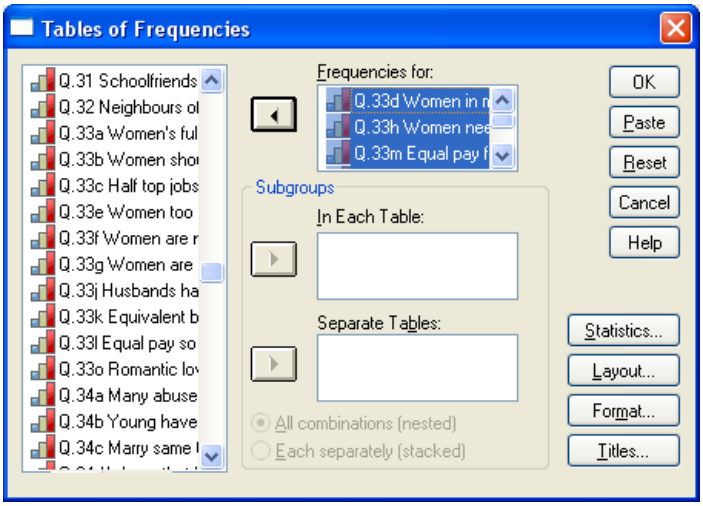
Analyze
…Descriptive statistics
…Frequencies



The variables are not adjacent, so you need to drag them into the **Frequencies for** box one at a time or you can hold down **[CTRL**] and click on each of them to highlight them:



and drag all four across in one go:

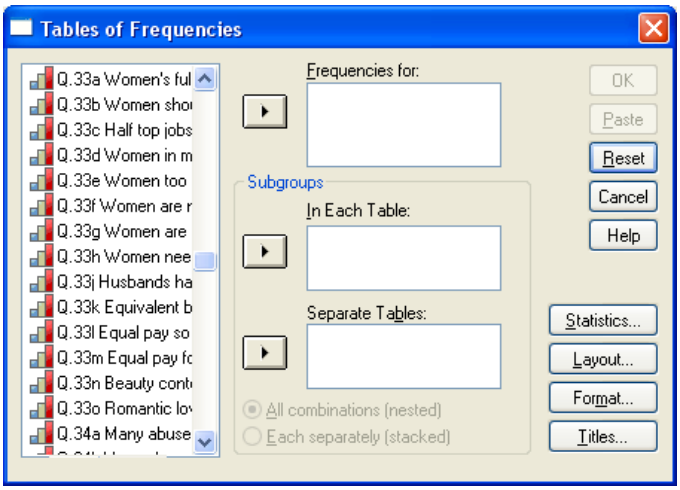


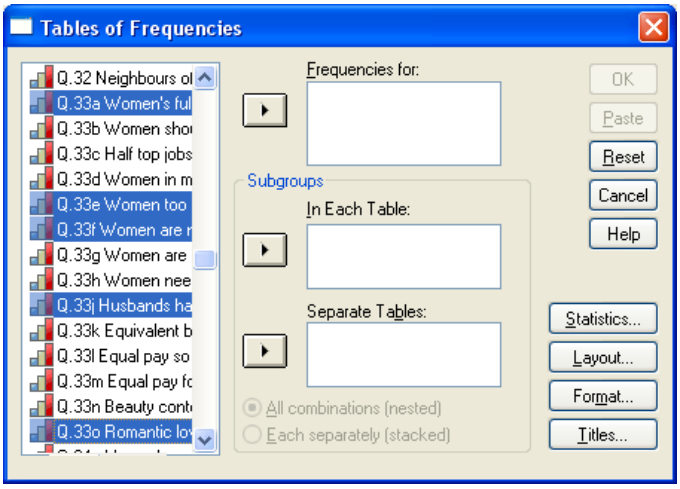
Press **OK** to get tables.

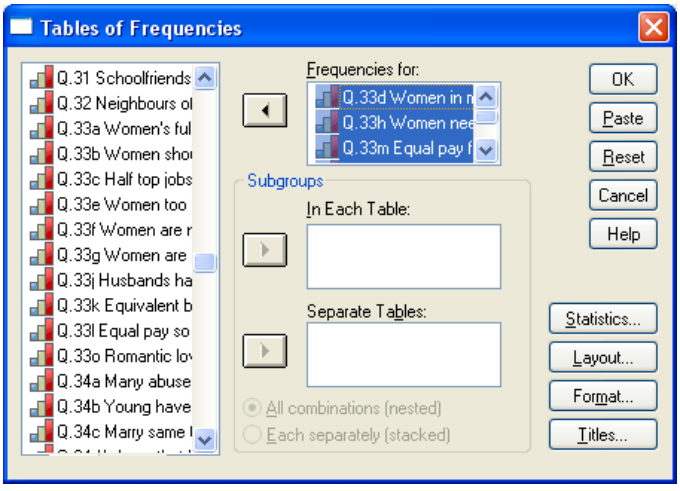
For the summary table, it's better to do the negative and positive items separately: that way the tables fit in the Viewer (and on an A4 Word page like this one!)

Negative items

Analyze
…Tables
…Tables of Frequencies



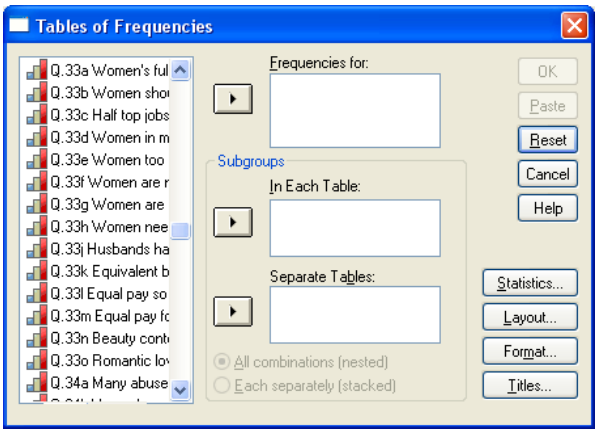


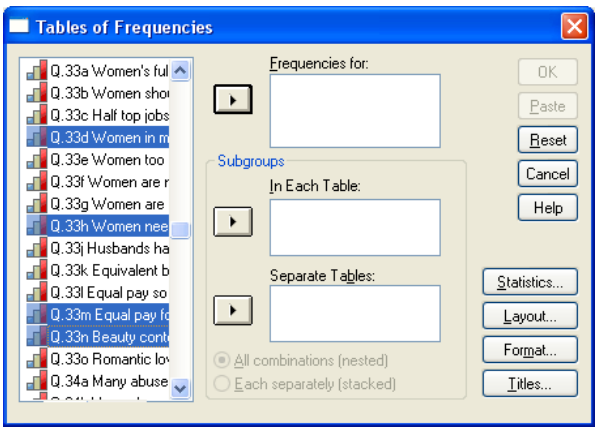


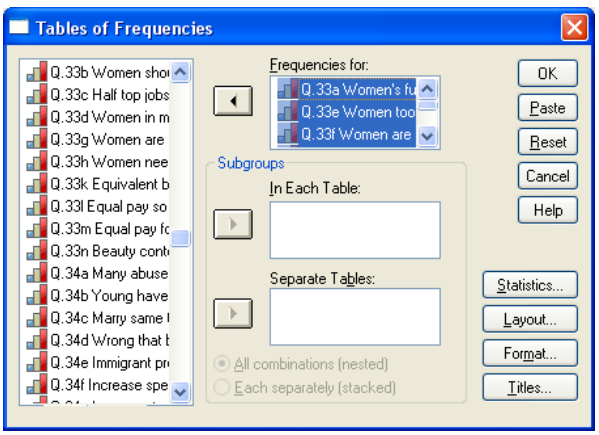
Press **OK** to get the table.

Positive items

Analyze
…Tables
…Tables of Frequencies







Press **OK** to get the table.

Again it's easier to produce summary tables with the drop-down menus, but this time it is more fiddly as the variables are not adjacent in the file. For analytical purposes, especially with large samples, we'd be better off with % rather than n in the cells, as % are easier to compare. A useful tool, but you do lose information on missing values.

Next tutorial(s): 3.5.2.6 The COUNT command 2 - Sexism

Feedback on ease of understanding and use of tutorial, please, to: johnfhall@orange.fr