Block 3: Analysing two variables (and sometimes three) **Derived variables**

3.5.2.4: The **COMPUTE** command 1 – Attachment to status quo

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

Exemplar:	Fifth form survey fifthx.sav
Variable to be derived:	Attachment to status quo
Source variables:	v275 to v278 (Q34 o, p, q ,r)
Previous tutorial:	3.5.2.3 The COUNT command 1 - Attachment to status quo

For the Fifth Form survey, we have already used the SPSS command **COUNT** to create a new variable **STATQUO** (a measure of teenagers' "Attachment to status quo") by counting the number of agreements with items o, p, q, and r in Q.34. A much more accurate and useful score can be obtained by combining **all** the responses for each item, so that **disagreement** can be taken into account as well as **agreement**, and also the **level** of agreement or disagreement. To generate this much superior measure, we shall use the SPSS command **COMPUTE**.

Remember, four items in question Q.34 (o,p,q,r) are replicated from a scale developed by Himmelweit to measure "attachment to status quo" among teenagers.

Q 34.	Do you agree or disagree with the following statements: (Ring the numbers)					
		Disagree Strongly	Tend to Disagree	Tend to Agree	Agree Strongly	
	o) It is best to be like the others and not to stand out from the rest.	1	2	3	4	(75)
	p) People who are content with what they have will have a better life than those who are always trying to improve their position.	1	2	3	4	(76)
	q) We are all born to our various social positions and it won't do to change them.	1	2	3	4	(77)
	r)The greatest source of happiness in life is to be satisfied with whatever you have.	1	2	3	4	(78)

On the questionnaire all items appear to be loaded in the same direction, so the values can be added together to yield a simple score, but **it's always best to check**. We did this in a previous tutorial **Data checks 1 – Status quo** so there's no need to do it again here.

In this edition of the file the values of all four variables tally exactly with those in the user manual and we can confidently proceed to create a new variable using the **COMPUTE** command.

COMPUTE creates new variables or new values for existing variables through an arithmetic expression or by combining existing variables according to one or more functions.   Only one transformation may be requested per **COMPUTE** command.

The general format is: **COMPUTE** **<computed variable>** **=** **<arithmetic expression> .**

in which **blue** indicates required **SPSS syntax** and **<plum>** things you have to **type in yourself**.

e.g. **COMPUTE** **STATQUO2** **=** **V275 + V276 + V27 7+ V278 .**

This command will create a new variable **STATQUO2** by adding together the values of the four variables **v275**, **v276**, **v277** and **v278**. If any of the four variables have missing values, then the value for the resultant variable **STATQUO2** will be set to system missing. As well as a full range of arithmetic operators **COMPUTE** also has available some special functions. There's a selection of these listed in the appendix (page 18). For a fuller explanation see the SPSS User Guide or use the **help** button.

A useful one is **SUM[.n]** which calculates the sum of all variables in the expression, but with optional **[.n]** requires a minimum number of valid values, otherwise the value for the new variable will be set to system missing.

e.g. **COMPUTE** **STATQUO2 = SUM.4 (V275 TO V278) .**

…which requires all four questions to be answered before calculating a score.

Another is the truncation function **TRUNC** which knocks off the figures after the decimal point after division. Thus, if you have a variable like actual income in ££, €€ or $$, or age in whole years you can generate quick groupings by dividing by ten or a thousand etc.

eg **COMPUTE** **AGEGROUP = TRUNC (AGE/10) .**
COMPUTE **INCGROUP = TRUNC (INCOME/1000) .**

These would not normally be used in reports, but can be helpful when exploring a new data set. The other functions are far too advanced for beginners and very rarely used anyway, except as intermediate stages.

Attachment to status quo

First we need a name for the new variable. We've already got **STATQUO** so let's call this one **STATQUO2**.

Going back to the general format:

COMPUTE   **<computed variable> = <arithmetic expression>.**

in which **blue** indicates required **SPSS syntax** and **<plum>** things you have to **type in yourself**.

A small table helps:

SPSS command	New variable	Source variables
COMPUTE	**STATQUO2**	**v275 v276 v277 v278**

We can therefore write:

COMPUTE   **STATQUO2   =   V275 + V276 + V277 + V278.**

Because the variables are adjacent in the file, you can also use

COMPUTE   STATQUO2 = SUM.4 (V275 TO V278).

Do not use   **COMPUTE   STATQUO2 = SUM (V275 TO V278).** It gives a different answer because it calculates a score even if one or more items are missing!

Start by downloading the saved file fifthx.sav [Don't worry if you haven't got immediate direct access to SPSS, just follow the tutorial anyway]

You should be able to do this yourself by now, so…

Scroll down to find variables **v275** to **v278** containing the data for items Q34o to Q34r, 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 and reduce the pane vertically to show only these variables:
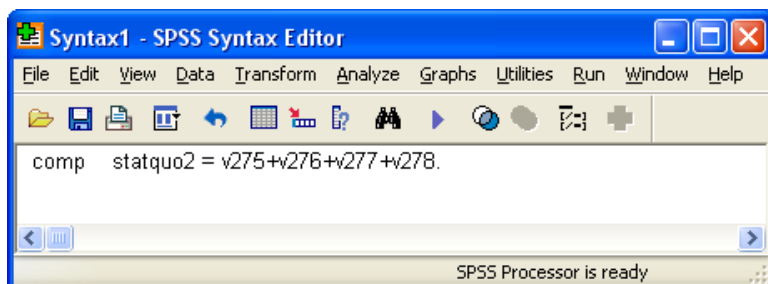

Now you know why positional variable names and informative variable labels are so useful! You can work straight from the questionnaire to the data set and *vice versa*.

The syntax way

If you want to do this using the drop-down menus, feel free, but it's cumbersome and takes forever. It's so much easier to open a new syntax file:
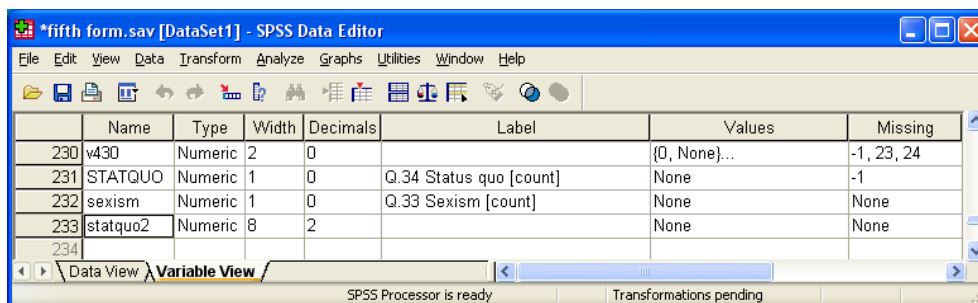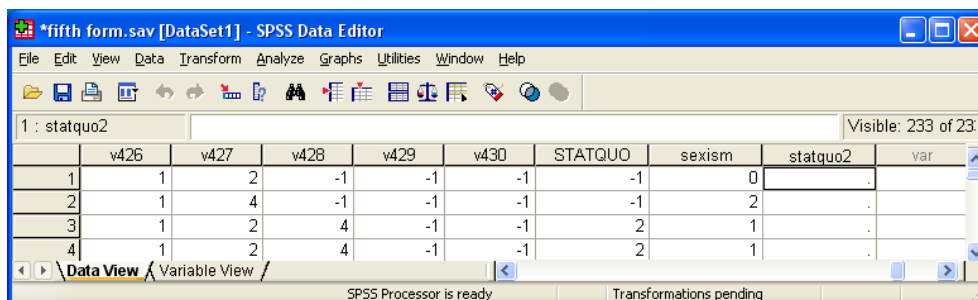

and type in your command line:

Leave the cursor in or on the new line and press **[CTRL]+R** or **Run…Current**.

This creates a new variable **STATQUO2** to measure attachment to status quo, but this time the resultant score will lie in the range 4 to 16 because we are adding together 4 items with values in the range 1 to 4. **COUNT** ignores any missing values and will produce a score for every case, whereas, with one important exception, **COMPUTE** will automatically assign the system missing value ' **.** ' (dot) if one or more items in the list is missing.
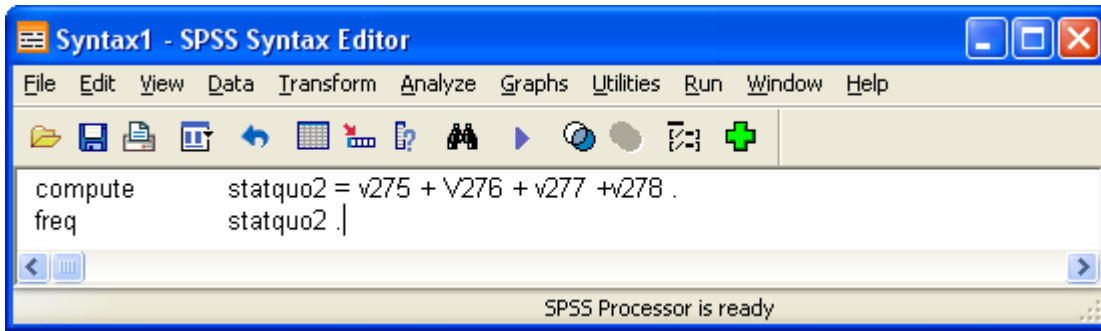
Nothing appears to happen, but, if you go back to the **Data Editor** and scroll down to the last row, you will see that **statquo2** has been appended.

However if you switch to **Data View** and press **[CTRL]►** to skip to the last column

… you will see that no scores have yet been entered for **STATQUO2**. Scroll up and down the column to confirm. This is because SPSS is waiting for an **EXECUTE** command or for a statistical procedure before it calculates any scores. We need a frequency count for **STATQUO2** anyway, so:

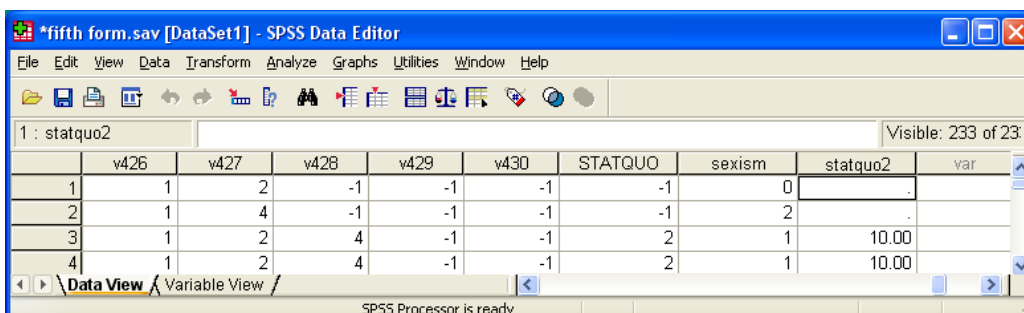
Run the command to get.
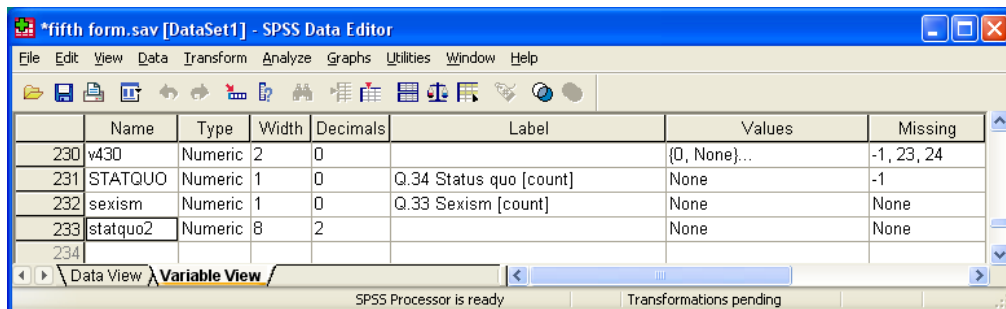
Statistics

statquo2

N	Valid	106
	Missing	36

statquo2

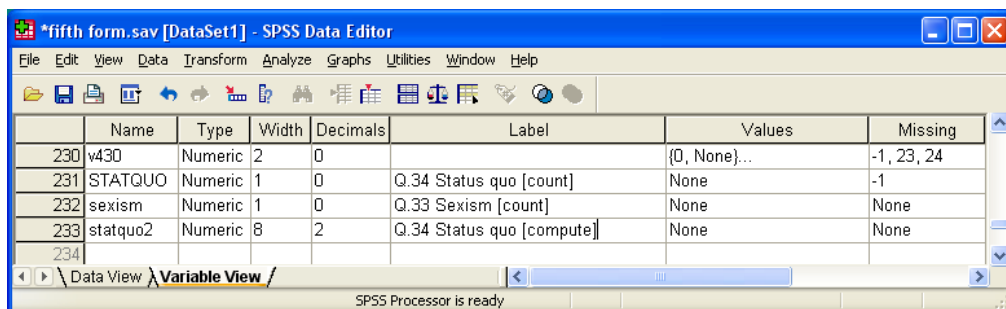|		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	4.00	1	.7	.9	.9
	5.00	3	2.1	2.8	3.8
	6.00	2	1.4	1.9	5.7
	7.00	6	4.2	5.7	11.3
	8.00	13	9.2	12.3	23.6
	9.00	18	12.7	17.0	40.6
	10.00	26	18.3	24.5	65.1
	11.00	17	12.0	16.0	81.1
	12.00	14	9.9	13.2	94.3
	13.00	4	2.8	3.8	98.1
	14.00	2	1.4	1.9	100.0
	Total	106	74.6	100.0	
Missing	System	36	25.4		
Total		142	100.0		

Go back to the **Data editor:**

5

Scroll down and you will see that this time, although **STATQUO2** scores have now been generated, some cells are still empty. This is because some pupils were missing on one or more of the four items and SPSS has automatically set their score to system missing.
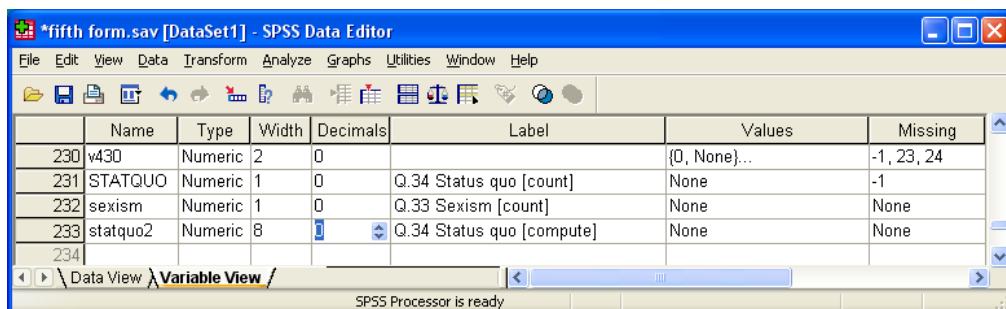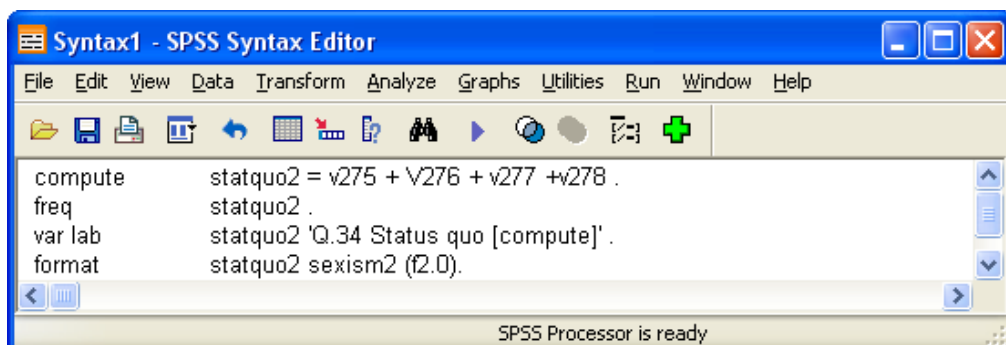
Go back to **Variable View**

We still need to give **STATQUO2** a label and and get rid of the decimal places. Both can be done directly in the **Data Editor**, the label by writing your own text in the **Label** column

…and the format by changing 2 to 0 in the **Decimals** column

Alternatively you can add **variable labels** and **format** commands to the syntax file

..and run the frequency count again

…which produces:

Q.34 Status quo [compute]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	4	1	.7	.9	.9
	5	3	2.1	2.8	3.8
	6	2	1.4	1.9	5.7
	7	6	4.2	5.7	11.3
	8	13	9.2	12.3	23.6
	9	18	12.7	17.0	40.6
	10	26	18.3	24.5	65.1
	11	17	12.0	16.0	81.1
	12	14	9.9	13.2	94.3
	13	4	2.8	3.8	98.1
	14	2	1.4	1.9	100.0
	Total	106	74.6	100.0	
Missing	System	36	25.4		
Total		142	100.0		

This table shows a range of 4 to 14 (out of a theoretical 16). We can improve it by adjusting the score so that it becomes a **ratio scale** with a true 0 point and a theoretical range of 0 – 12 instead of 4 – 16. To do this we subtract the number of items in the scale (4) using a further SPSS command **compute statquo2 = statquo2 – 4**. This is the equivalent of the original items having been coded 0 to 3 instead of 1 to 4, and avoids the error (common in many reports, especially when comparing group means) of discussing means as though they come from a ratio scale. Do **freq** again:

7

...to yield:

Q.34 Status quo [compute]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	1	.7	.9	.9
	1	3	2.1	2.8	3.8
	2	2	1.4	1.9	5.7
	3	6	4.2	5.7	11.3
	4	13	9.2	12.3	23.6
	5	18	12.7	17.0	40.6
	6	26	18.3	24.5	65.1
	7	17	12.0	16.0	81.1
	8	14	9.9	13.2	94.3
	9	4	2.8	3.8	98.1
	10	2	1.4	1.9	100.0
	Total	106	74.6	100.0	
Missing	System	36	25.4		
Total		142	100.0		

You could also use **compute statquo2 = sum.4 (v275 to v278)-4** but emphatically **do not use compute statquo2 = sum (v275 to v278) – 4** because it will calculate a score even if one or more constituent items are missing. Of the 36 missing cases in the above table, 18 are missing all four items: another 18 are missing at least one item, but would still be given a score.

Because **STATQUO2** is a genuine interval scale (also now a ratio scale) it is legitimate to produce descriptive statistics and a histogram overlaid with a normal curve. As an option, we can also suppress the frequency table (.. **/format notable**) and just ask for selected statistics (e.g.**/statistics mean stdev**) and/or graphics (e.g**.…/histogram normal**).
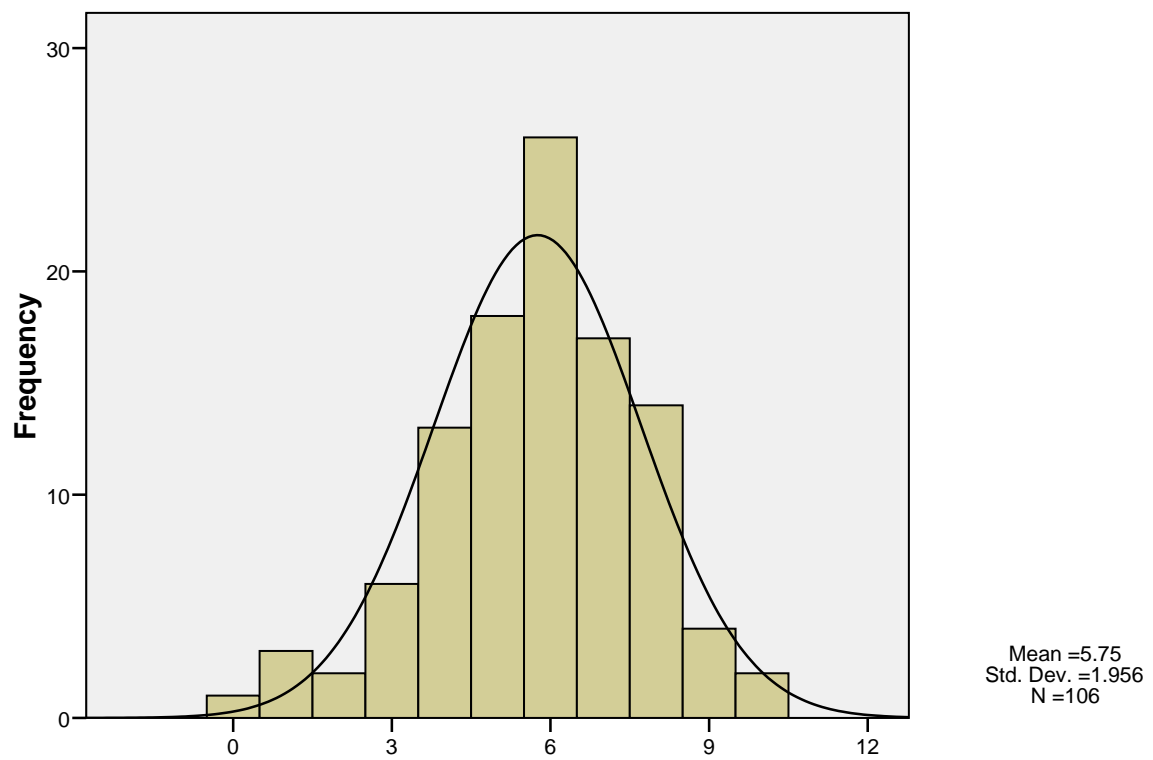


```
compute      statquo2 = v275 + V276 + v277 +v278 .
freq         statquo2 .
var lab      statquo2 'Q.34 Status quo [compute]' .
format       statquo2  (f2.0).
freq         statquo2 .
compute      statquo2 = statquo2 - 4 .
freq         statquo2 .
freq         statquo2 / for not /sta mean stdev /his nor .
```

Run to get:

Statistics

Q.34 Status quo [compute]

N	Valid	106
	Missing	36
Mean		5.75
Std. Deviation		1.956

Histogram
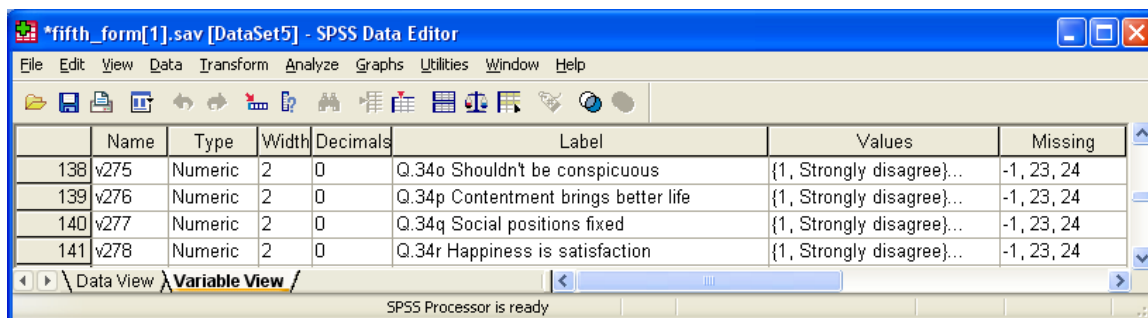

Mean =5.75
Std. Dev. =1.956
N =106

Nice one to show your statistics tutors: they love normal distributions!

[**NB:** There's a label on the output, but it has not copied across.]

The point-and click way

You can use the drop-down menus if you prefer, but it takes a lot longer.

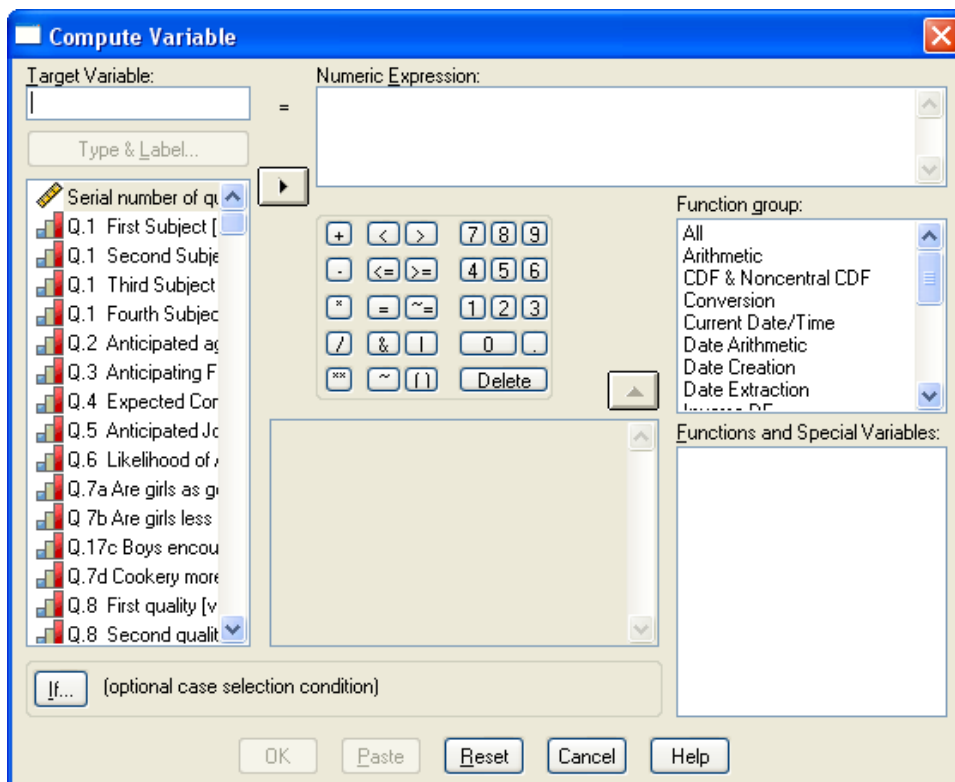Go back to the **Data Editor** and find the four items we need:
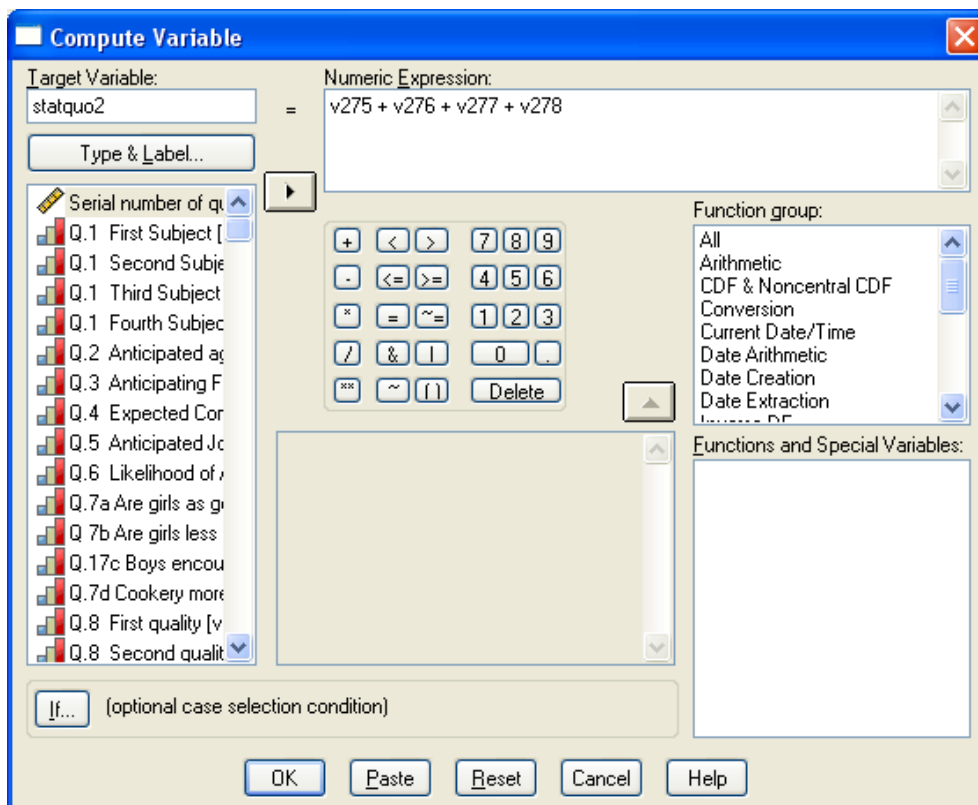

Click on:

Transform
  …Compute variable

Scroll down to find the items o, p, q and r in Q.33 (**v275** to **v278**), type **statquo2** in the **Target Variable box** and then highlight and transfer the variables one at a time with ▶ into the **Numeric Expression** box: click on the + button in between each one.
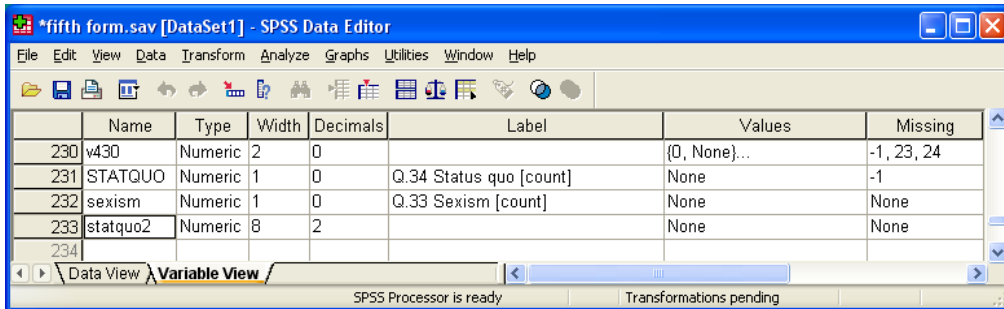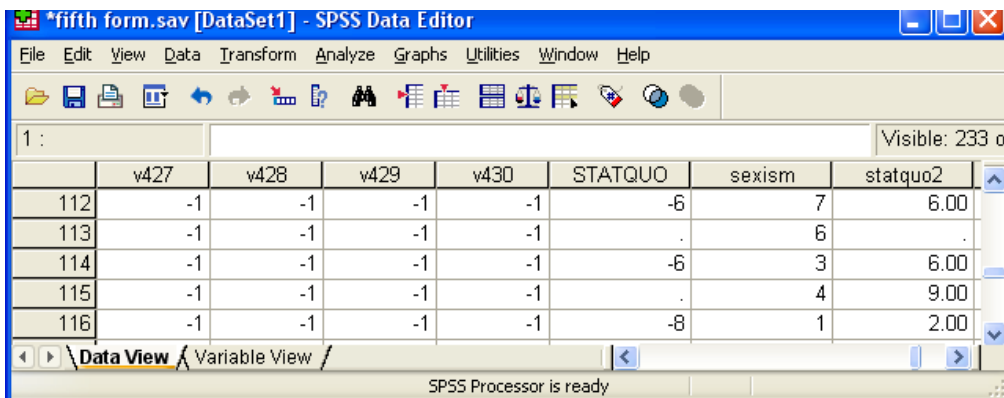
You have to do this **four times** (!!!) to get to

And click on OK to get:

```
COMPUTE statquo2 = v275 + v276 + v277 + v278 .
EXECUTE .
```

As before, **statquo2** has been appended in the last row of the **Data Editor**

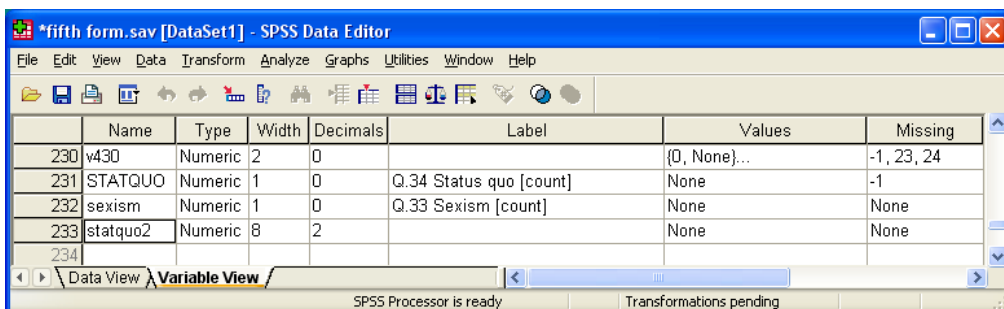|	Name	Type	Width	Decimals	Label	Values	Missing
230	v430	Numeric	2	0		{0, None}...	-1, 23, 24
231	STATQUO	Numeric	1	0	Q.34 Status quo [count]	None	-1
232	sexism	Numeric	1	0	Q.33 Sexism [count]	None	None
233	statquo2	Numeric	8	2		None	None
234							

but this time, because of the automatic **EXECUTE** command, the scores have also been calculated and entered. Switch to **Data View** and press **[CTRL]►** to skip to the last column, then scroll up and down to see them.
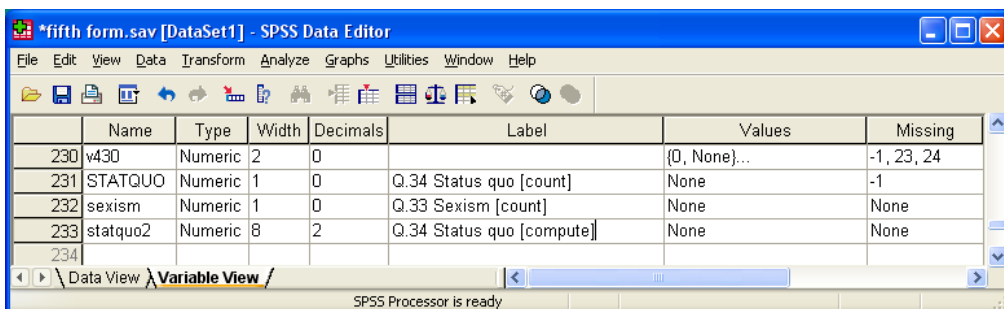
	v427	v428	v429	v430	STATQUO	sexism	statquo2
112	-1	-1	-1	-1	-6	7	6.00
113	-1	-1	-1	-1	.	6	.
114	-1	-1	-1	-1	-6	3	6.00
115	-1	-1	-1	-1	.	4	9.00
116	-1	-1	-1	-1	-8	1	2.00

We still need to add a variable label and change the format to get rid of the decimal places.

Go back to **Variable View**

	Name	Type	Width	Decimals	Label	Values	Missing
230	v430	Numeric	2	0		{0, None}...	-1, 23, 24
231	STATQUO	Numeric	1	0	Q.34 Status quo [count]	None	-1
232	sexism	Numeric	1	0	Q.33 Sexism [count]	None	None
233	statquo2	Numeric	8	2		None	None
234							

…and write your own label in the **Label** column

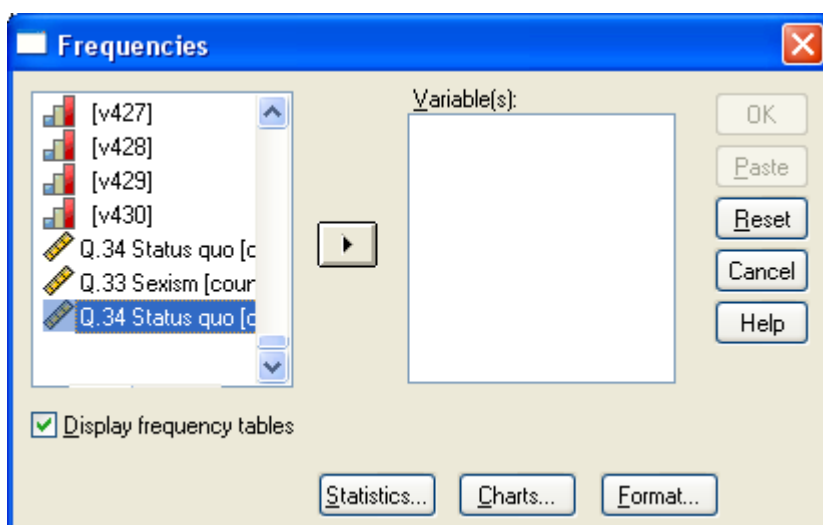|	Name	Type	Width	Decimals	Label	Values	Missing
230	v430	Numeric	2	0		{0, None}...	-1, 23, 24
231	STATQUO	Numeric	1	0	Q.34 Status quo [count]	None	-1
232	sexism	Numeric	1	0	Q.33 Sexism [count]	None	None
233	statquo2	Numeric	8	2	Q.34 Status quo [compute]	None	None
234							

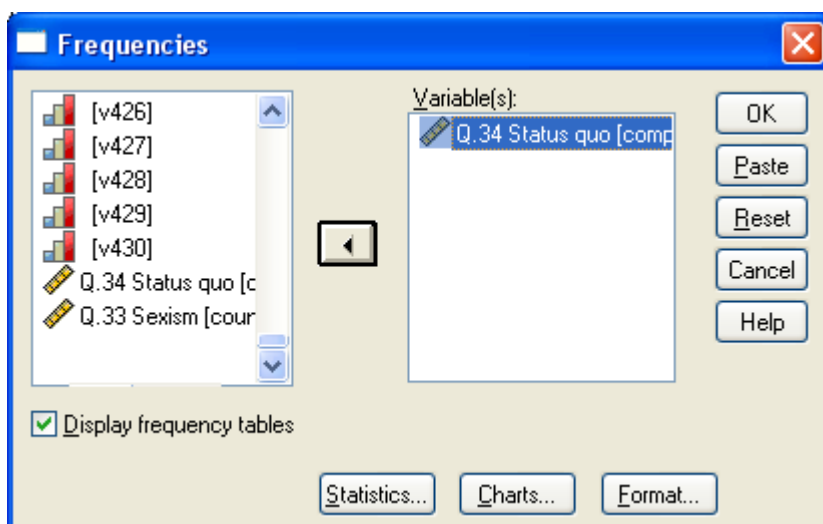and amend the format by changing **2** to **0** in the **Decimals** column

Now for the frequency count. You should be able to do this by yourself by now, so go back to the **Data Editor**:

Analyze
   Descriptive statistics
……Frequencies

Highlight **statquo2**
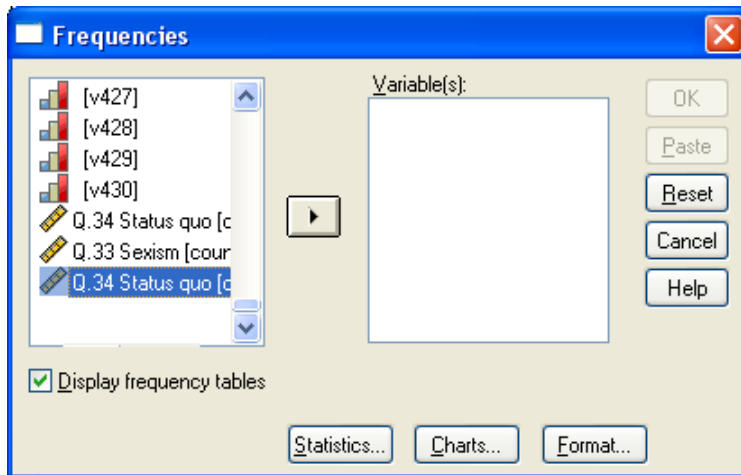

…transfer it with ▶ into the **Variable(s)** box.

and click **OK** to run the job and get the same frequency table as on page 8.
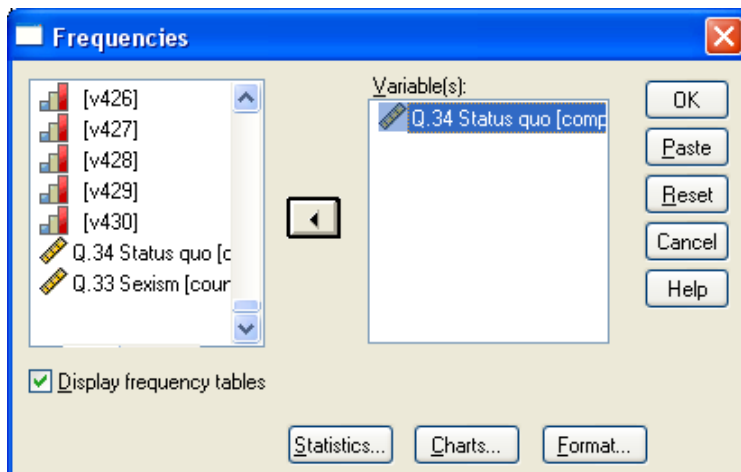
To suppress the frequency table and get only selected statistics and a histogram:

Analyze
  Descriptive statistics
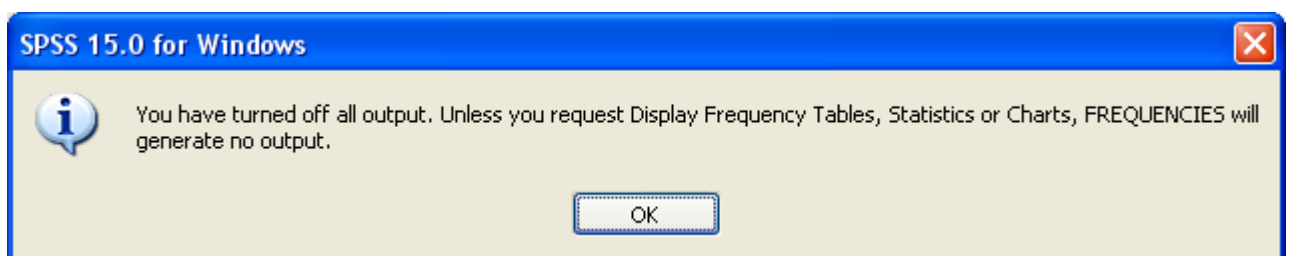**……Frequencies**
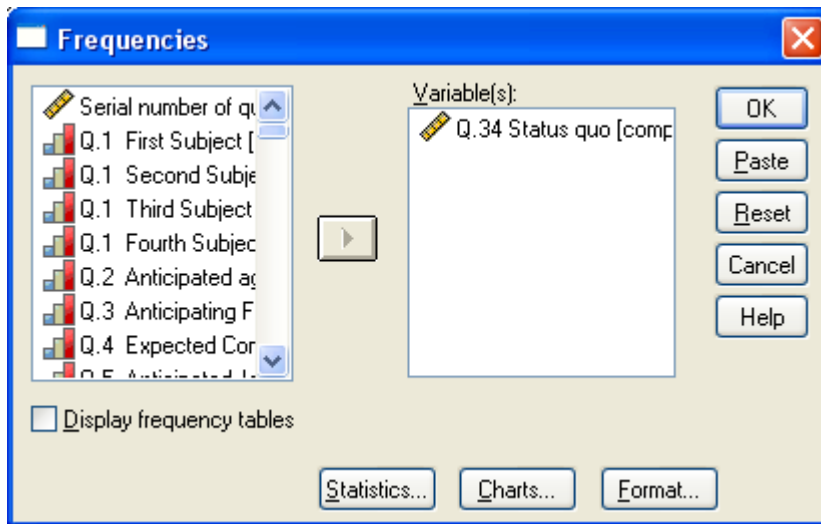
Highlight **statquo2**

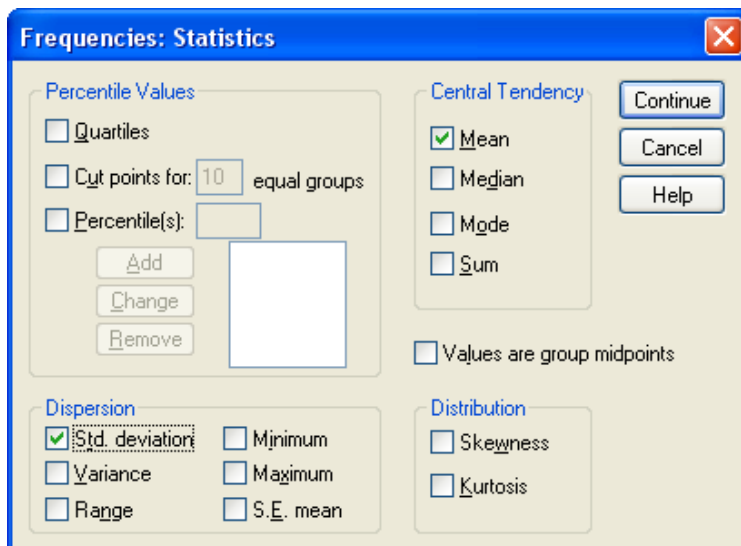and transfer it with ▶ into the **Variable(s)** box.

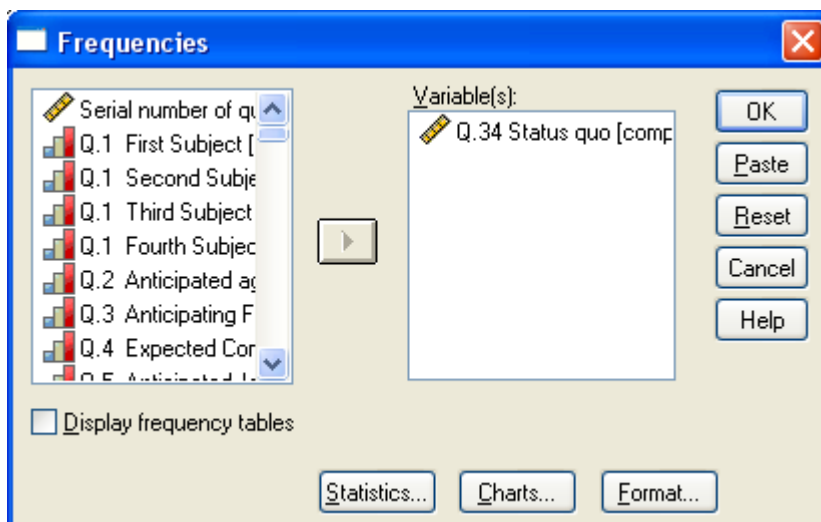Uncheck **Display frequency tables** and you'll get the following warning.

Click **OK** but you must now ask for something to display. This warning does not appear if you select Statistics or Charts before unchecking **Display frequency tables**.

click on Statistics and check the boxes for **Std deviation** under **Dispersion** and **Mean** under **Central Tendency**
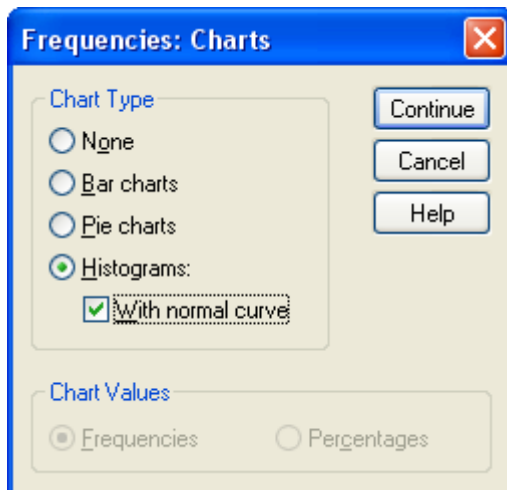

then click on Continue
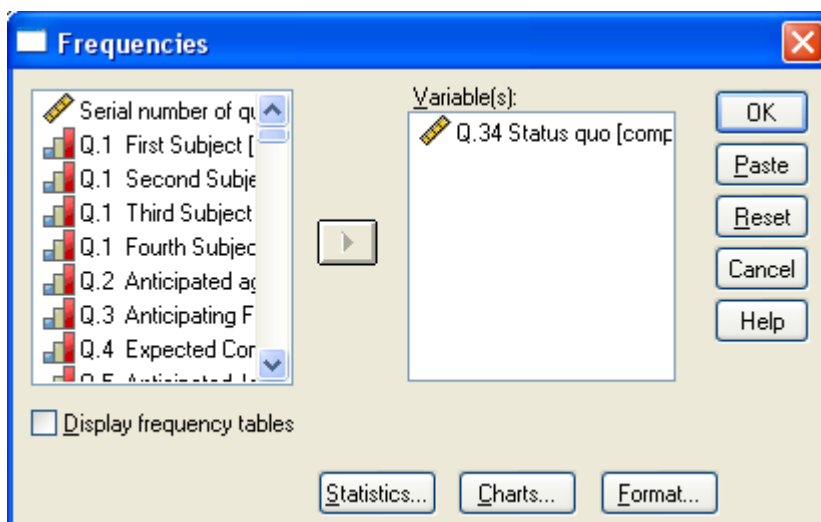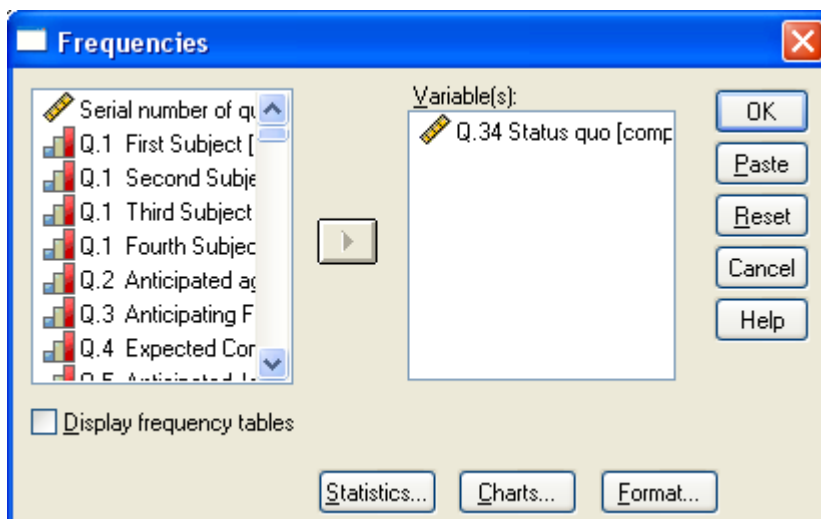

click on Charts

then check the button for **Histograms** under **Chart Type** and the box **With normal curve**

Continue

Continue

press **OK**

SPSS generates the following syntax:

```
FREQUENCIES
  VARIABLES=statquo2
  /FORMAT=LIMIT(5)
  /STATISTICS=MEAN MEDIAN
  /HISTOGRAM  NORMAL
  /ORDER=  ANALYSIS .
```

…and the same table and histogram as before

Statistics

Q.34 Status quo [compute]

N	Valid	106
	Missing	36
Mean		5.75
Std. Deviation		1.956

Histogram

Mean =5.75
Std. Dev. =1.956
N =106

I told you it was quicker in syntax!

Next tutorial(s):   3.5.2.5 Data checks 2 - Sexism

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

Appendix:

The COMPUTE command

COMPUTE creates new variables or new values for existing variables through an arithmetic expression or by combining existing variables according to one or more functions. Only one transformation may be requested per **COMPUTE** command.

The general format is:

COMPUTE <computed variable> = <arithmetic expression> .

The arithmetic operators available are:

+	addition		/	division
-	subtraction		**	exponentiation (ie 'to the power of')
*	multiplication			…together with brackets.

e.g.   **COMPUTE   STATQUO = V275+V276+V277+V278.**

If any of the variables included in the expression have missing values, then the value for the resultant variable is also declared missing (**except** for statistical functions as shown below)

As well as a full range of arithmetic operators **COMPUTE** has available some special functions. There are seven classes of functions within the SPSS **COMPUTE** command:

Arithmetic	Logical
Statistical	Random
Missing Value	Date/Time functions
Cross-case	

For a fuller explanation of these facilities see the SPSS User Guide or use the **help** button.

Some of the most common functions you are likely to use in survey analysis are:

TRUNC	truncate	SD[.n]	std deviation of values across list
SUM[.n]	sum of values across variable list	MIN[.n]	minimum of values across list
MEAN[.n]	mean of values across variable list	MAX[.n]	maximum of values across list

Where the optional [.n] indicates, for statistical functions, the minimum number of arguments required to be valid before the expression is declared missing.

A useful one is **SUM[.n]** which calculates the sum of all variables in the expression, but [with optional [.n] requires a minimum number of valid values, otherwise the value for the new variable will be set to system missing.

e.g.   **COMPUTE   STATQUO = SUM.4 (V275 TO V278).**

…which requires all four questions to be answered before calculating a score.

Another is the truncation function **TRUNC** which knocks off the figures after the decimal point after division. Thus, if you have a variable like actual income in ££, €€ or $$, or age in whole years you can generate quick groupings by dividing by ten or a thousand etc

eg   **COMPUTE   AGEGROUP=TRUNC (AGE/10).**
      COMPUTE   INCGROUP=TRUNC (INCOME/1000).