

Block 3: Analysing two variables (and sometimes three)

Derived variables

3.5.2.7: The **COMPUTE** command 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: V248 v252 v253 v256 V261 (negative: a,e,f,j,o)
 V251 V255 V259 V260 (positive: d,h,m,n)
Previous tutorials: [3.5.2.6 The COUNT command 2 - Sexism](#)

In the Fifth Form survey, we have already used **COUNT** to create **SEXISM** (as a crude measure of negative attitudes to women) but 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.

We can generate this much superior measure using the SPSS command **COMPUTE**.

The general format for **COMPUTE** is:

COMPUTE <computed variable> = <arithmetic expression>.

...in which **blue** indicates **SPSS keywords** etc., and **<plum>** things you have to **type in yourself**.

Question Q.33 is more complex than Q.34. It consists of 14 statements measuring opinions about women, some negative, some positive, with which pupils can agree or disagree on a 4-point scale. (Codes 1-4 on columns 48 to 61 of card 2, read into SPSS using the **positional** naming convention as **V248 to V261**) Only **nine** of these items will be used for this new measure of sexism.

Using **COMPUTE** to generate a measure of "Sexism" presents an additional problem. Four of the nine items are actually coded the opposite way round to the other five. If we add them all together they will simply cancel each other out. We want a high score to indicate high sexism, so we need to reverse (at least temporarily) the scores on the four positive items **V251 V255 V259 V261** (d,h,m and n) before adding the items together. The resultant variable will have values in the range 9 to 36 where 36 indicates high sexism,

The values of the source variables can be changed using a **RECODE** statement:

RECODE V251 V255 V259 V260 (4=1) (3=2) (2=3) (1=4).

Be very careful with recodes! The changes will be made permanent if you save the **Data Editor**. To avoid such problems, it's much safer to use **TEMPORARY**

TEMPORARY.
RECODE V251 V255 V259 V260 (4=1) (3=2) (2=3) (1=4).

... and even safer to create a new set of intermediate variables with **RECODE....INTO**.

RECODE V251 V255 V259 V260 (4=1) (3=2) (2=3) (1=4)(**ELSE = SYSMIS**)
INTO R251 R255 R259 R260.

...where **Rddd** indicates a recoded variable, but the positional info is retained.

You can then use these for calculating the sexism score. They can be discarded later, unless you want to keep them for another analysis.

Here's the facsimile of Q.33 again

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- <i>power</i> 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)

First we need a name for the new variable. We've already got **SEXISM** so let's call this one **SEXISM2**.

Going back to the general format for **COMPUTE** :

COMPUTE <computed variable> = <arithmetic expression>.

...in which **blue** indicates SPSS keywords etc., and **plum** indicates things **you have to write in**.

A small table helps

SPSS command	New variable	Source variables	
		Negative	Positive
COMPUTE	SEXISM2	V248 V252 V253 V256 V261	V251 V255 V259 V260 recoded into R251 R255 R259 R260

Our SPSS command will therefore be:

COMPUTE SEXISM2 =
V248 + V252 + V253 + V256 + V261 + R251 + R255 + R259 + R261 .

...or if you want to keep the variables in questionnaire order:

COMPUTE SEXISM2 =
V248 + R251 + V252 + V253 + R255 + V256 + R259 + R260 + V261 .

If you use the **SUM** facility, you have to use **SUM.n** where n is the number of variables to be summed and you **must use commas** to separate the variable names (unless they are adjacent in the file, in which case you can use **Vddd TO Vddd**)

COMPUTE SEXISM2 =
SUM.9 (V248, V252, V253, V256, V261, R251, R255, R259, R260) .

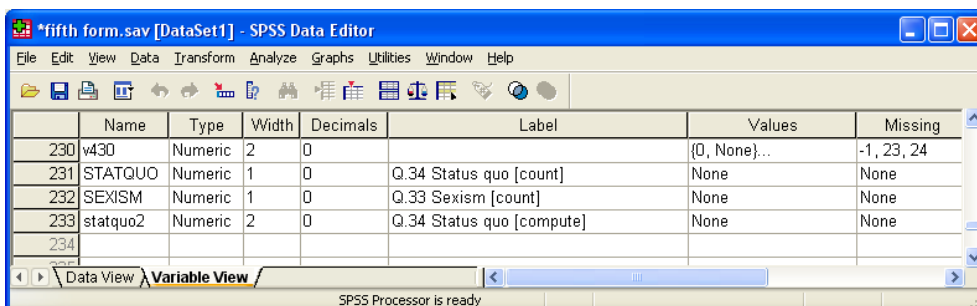
Do NOT use

COMPUTE SEXISM2 =
SUM (V248, V252, V253, V256, V261, R251, R255, R259, R260) .

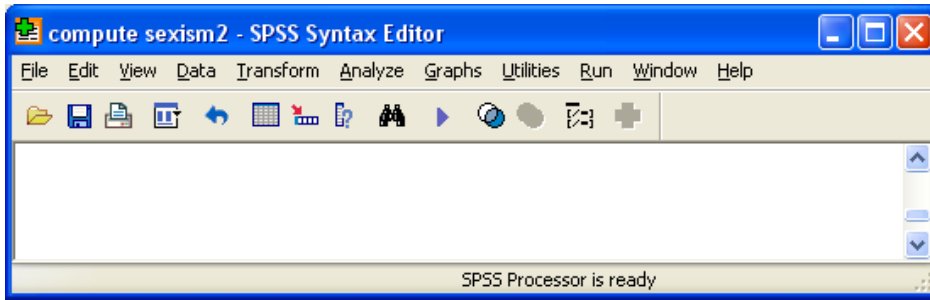
It gives a **misleading answer** because it calculates a score even if one or more items are missing!

Whenever items are being used in constructing new variables it's essential to **check the distributions** against those in the user manual. We did this in a previous tutorial **Data checks 2 – Sexism** so there's no need to do it again here. In this version of **Fifth Form Survey** the values of all nine variables tally exactly with those in the the user manual and we can confidently proceed to create a new variable using the **COMPUTE** command.

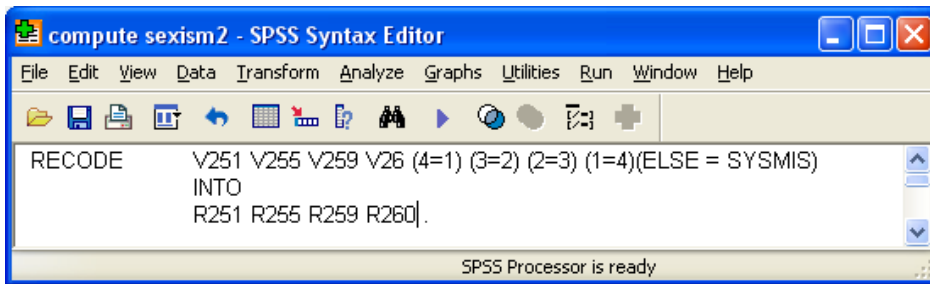
Go back to the Data Editor, scroll to the end: you should have **STATQUO**, **SEXISM** and **statquo2** if you did the previous tutorials. Adjust the column separators and adjust the pane to see only these variables.



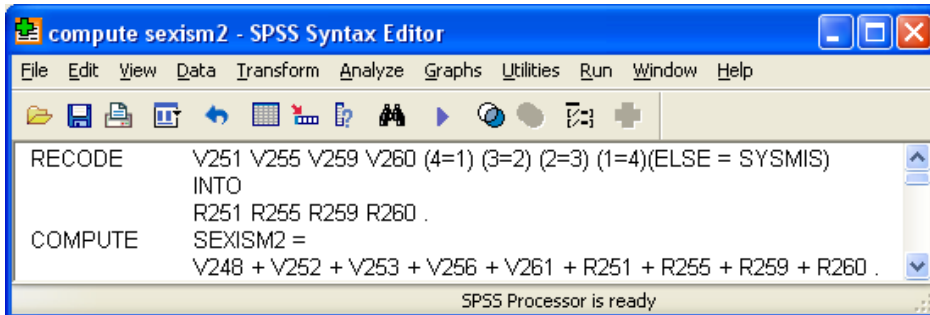
- File
- ...New
- ...Syntax



Write in the **RECODE** command :

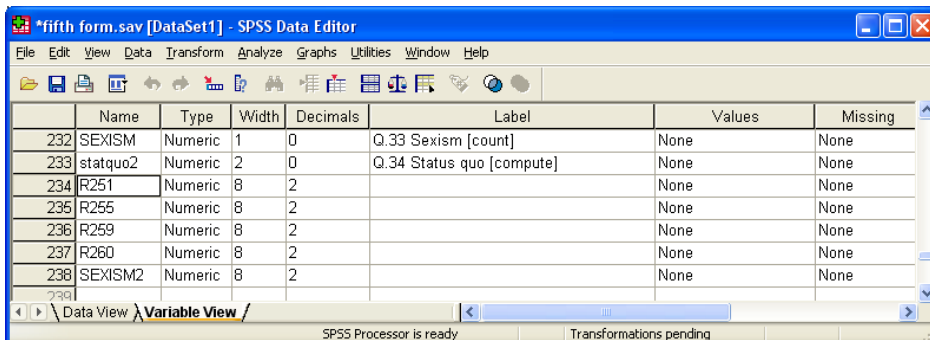


...followed by the **COMPUTE** command for **SEXISM2**.

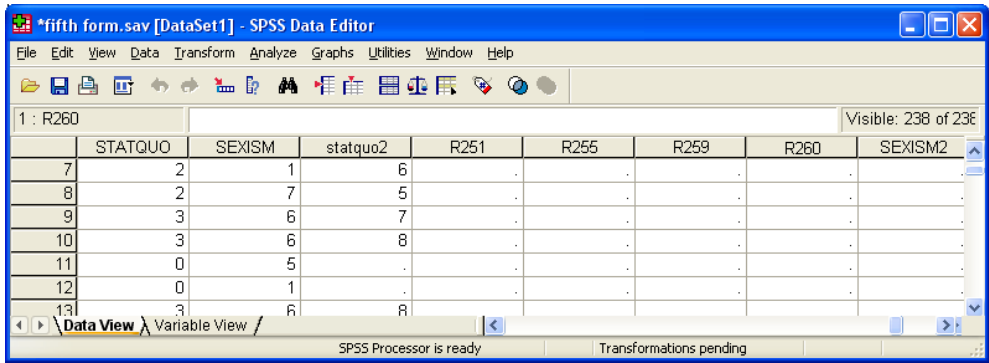


[NB: If you do this with long batteries of attitude scales containing missing values, you will rapidly reduce your effective sample size, so for some statistical purposes it may be better to **RECODE** the missing values to a central value such as mean, median or mode before calculating any scores.]

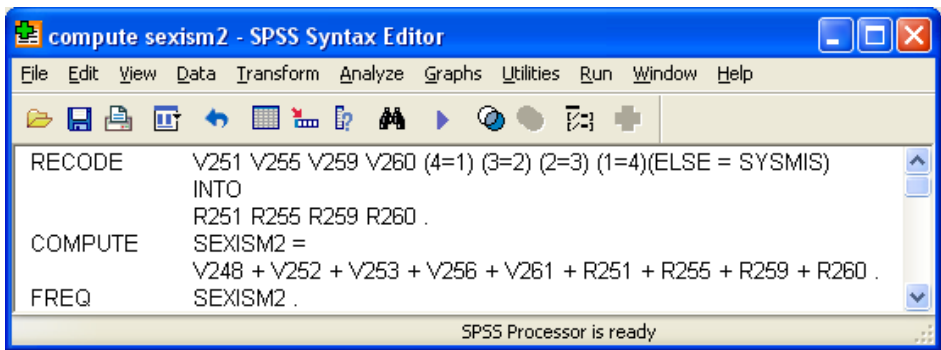
Run both commands. As before (with STATQUO2) nothing appears to happen, but if you go back to the **Data Editor** and scroll to the last row, you will see that **R251 R255 R259 R260** and **SEXISM2** have been appended at the end of the file.



Switch to **Data View** and skip to the last column



No scores have yet been calculated. SPSS is waiting for an **EXECUTE** command or a request for statistical analysis. We need a frequency count anyway to check the distribution of scores, so add a **FREQUENCIES** command:



..and run the job to get:

SEXISM2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	10.00	4	2.8	3.5	3.5
	12.00	3	2.1	2.6	6.1
	13.00	3	2.1	2.6	8.8
	14.00	3	2.1	2.6	11.4
	15.00	8	5.6	7.0	18.4
	16.00	9	6.3	7.9	26.3
	17.00	5	3.5	4.4	30.7
	18.00	9	6.3	7.9	38.6
	19.00	17	12.0	14.9	53.5
	20.00	12	8.5	10.5	64.0
	21.00	11	7.7	9.6	73.7
	22.00	3	2.1	2.6	76.3
	23.00	5	3.5	4.4	80.7
	24.00	9	6.3	7.9	88.6
	25.00	2	1.4	1.8	90.4
	26.00	2	1.4	1.8	92.1
	27.00	6	4.2	5.3	97.4
28.00	2	1.4	1.8	99.1	
29.00	1	.7	.9	100.0	
	Total	114	80.3	100.0	
Missing	System	28	19.7		
Total		142	100.0		

Go back to the **Data Editor**

	STATQUO	SEXISM	statquo2	R251	R255	R259	R260	SEXISM2
29	0	0	.	2.00	2.00	1.00	.	.
30	2	1	6	2.00	2.00	2.00	3.00	19.00
31	0	3	3	2.00	2.00	2.00	3.00	19.00
32	0	0
33	2	1	6	2.00	1.00	2.00	3.00	16.00
34	3	3	7	2.00	2.00	2.00	3.00	21.00
35	1	5	.	3.00	2.00	4.00	2.00	24.00

Scroll down and you will see that this time, although **SEXISM2** scores have now been generated, some cells are still empty. This is because 28 pupils were missing on one or more of the nine items and SPSS has automatically set their score to system missing.

Go back to **Variable View**

	Name	Type	Width	Decimals	Label	Values	Missing
232	SEXISM	Numeric	1	0	Q.33 Sexism [count]	None	None
233	statquo2	Numeric	2	0	Q.34 Status quo [compute]	None	None
234	R251	Numeric	8	2		None	None
235	R255	Numeric	8	2		None	None
236	R259	Numeric	8	2		None	None
237	R261	Numeric	8	2		None	None
238	SEXISM2	Numeric	8	2		None	None

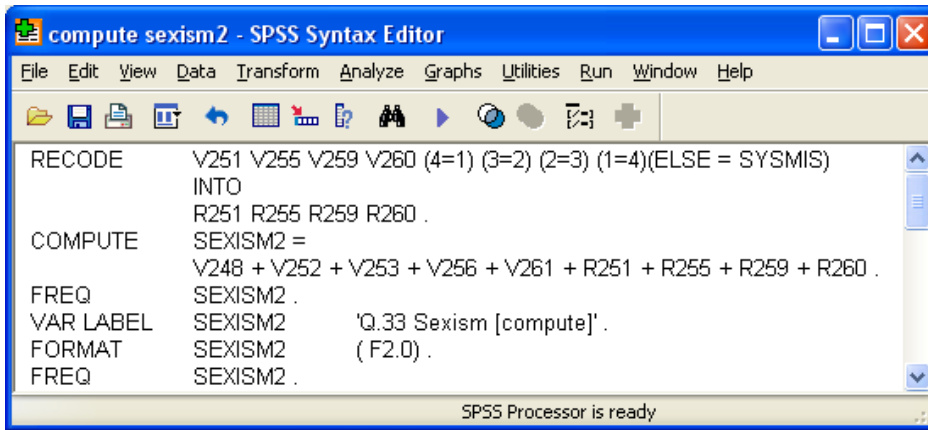
We still need to give **SEXISM2** a label and get rid of the decimal places. Both can be done directly in the **Data Editor**, the label by writing your own text in the **Label** column:

	Name	Type	Width	Decimals	Label	Values	Missing
232	SEXISM	Numeric	1	0	Q.33 Sexism [count]	None	None
233	statquo2	Numeric	2	0	Q.34 Status quo [compute]	None	None
234	R251	Numeric	8	2		None	None
235	R255	Numeric	8	2		None	None
236	R259	Numeric	8	2		None	None
237	R260	Numeric	8	2		None	None
238	SEXISM2	Numeric	8	2	Q.33 Sexism [compute]	None	None

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

	Name	Type	Width	Decimals	Label	Values	Missing
232	SEXISM	Numeric	1	0	Q.33 Sexism [count]	None	None
233	statquo2	Numeric	2	0	Q.34 Status quo [compute]	None	None
234	R251	Numeric	8	2		None	None
235	R255	Numeric	8	2		None	None
236	R259	Numeric	8	2		None	None
237	R260	Numeric	8	2		None	None
238	SEXISM2	Numeric	8	0	Q.33 Sexism [compute]	None	None

Alternatively you can add **variable labels** and **format** commands to the syntax file and run the frequency count again to force SPSS to execute the changes.



Q.33 Sexism [compute]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	10	4	2.8	3.5	3.5
	12	3	2.1	2.6	6.1
	13	3	2.1	2.6	8.8
	14	3	2.1	2.6	11.4
	15	8	5.6	7.0	18.4
	16	9	6.3	7.9	26.3
	17	5	3.5	4.4	30.7
	18	9	6.3	7.9	38.6
	19	17	12.0	14.9	53.5
	20	12	8.5	10.5	64.0
	21	11	7.7	9.6	73.7
	22	3	2.1	2.6	76.3
	23	5	3.5	4.4	80.7
	24	9	6.3	7.9	88.6
	25	2	1.4	1.8	90.4
	26	2	1.4	1.8	92.1
	27	6	4.2	5.3	97.4
	28	2	1.4	1.8	99.1
	29	1	.7	.9	100.0
		Total	114	80.3	100.0
Missing	System	28	19.7		
Total		142	100.0		

This table shows a range of 10 to 29 (out of a theoretical 36). 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 – 27 instead of 9 – 36. To do this we subtract the number of items in the scale (9) using another SPSS command **compute sexism2 = sexism2 – 9**. 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.

We're going to need another frequency count as well.

```

compute sexism2 - SPSS Syntax Editor
File Edit View Data Transform Analyze Graphs Utilities Run Window Help
RECODE V251 V255 V259 V260 (4=1) (3=2) (2=3) (1=4)(ELSE = SYSMIS)
INTO
R251 R255 R259 R260 .
COMPUTE SEXISM2 =
V248 + V252 + V253 + V256 + V261 + R251 + R255 + R259 + R260 .
FREQ SEXISM2 .
VAR LABEL SEXISM2 'Q.33 Sexism [compute]' .
FORMAT SEXISM2 (F2.0) .
FREQ SEXISM2 .
COMPUTE SEXISM2 = SEXISM2 - 9 .
FREQ SEXISM2 .
SPSS Processor is ready
  
```

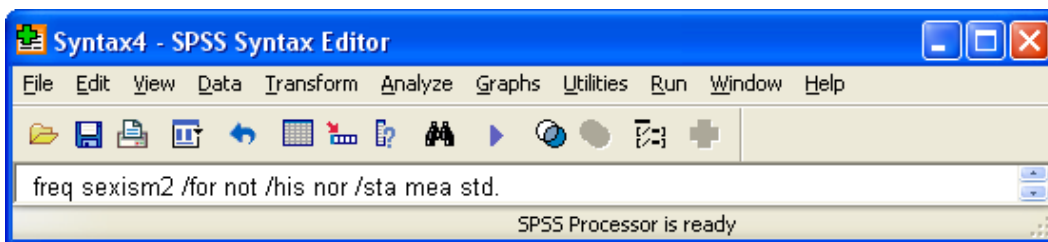
...to get:

Q.33 Sexism [compute]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	4	2.8	3.5	3.5
	3	3	2.1	2.6	6.1
	4	3	2.1	2.6	8.8
	5	3	2.1	2.6	11.4
	6	8	5.6	7.0	18.4
	7	9	6.3	7.9	26.3
	8	5	3.5	4.4	30.7
	9	9	6.3	7.9	38.6
	10	17	12.0	14.9	53.5
	11	12	8.5	10.5	64.0
	12	11	7.7	9.6	73.7
	13	3	2.1	2.6	76.3
	14	5	3.5	4.4	80.7
	15	9	6.3	7.9	88.6
	16	2	1.4	1.8	90.4
	17	2	1.4	1.8	92.1
	18	6	4.2	5.3	97.4
	19	2	1.4	1.8	99.1
	20	1	.7	.9	100.0
		Total	114	80.3	100.0
Missing	System	28	19.7		
Total		142	100.0		

Because **SEXISM2** is a genuine interval scale (also now a ratio scale) it is legitimate to ask for a mean and standard deviation (e.g. **/statistics mean stdev**) plus a histogram overlaid with a normal curve (e.g. **.../histogram normal**)

You can also suppress the frequency table (e.g. ...**/format notable**) and just get the statistics and the chart.:



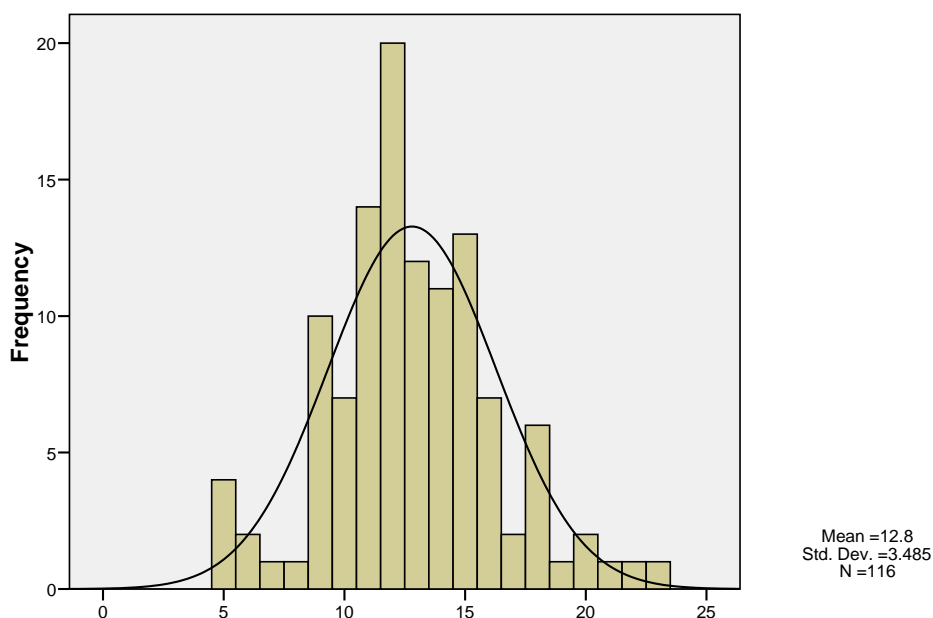
...which produces:

Statistics

Q.33 Sexism [compute]

N	Valid	116
	Missing	26
Mean		12.80
Std. Deviation		3.485

Histogram



[**NB:** SPSS automatically prints the mean, std and N. The histogram has a label on the output, but it's not transferred when copied. You'll have to write your own underneath the chart.]

The point-and click way

If you prefer, you can produce all the above tables, statistics and histogram using the drop-down menus, but it will take you a very long time. I'm not even going to bother. You have been warned!

End of section 3.5: Derived variables [**COMPUTE** and **COUNT**]

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