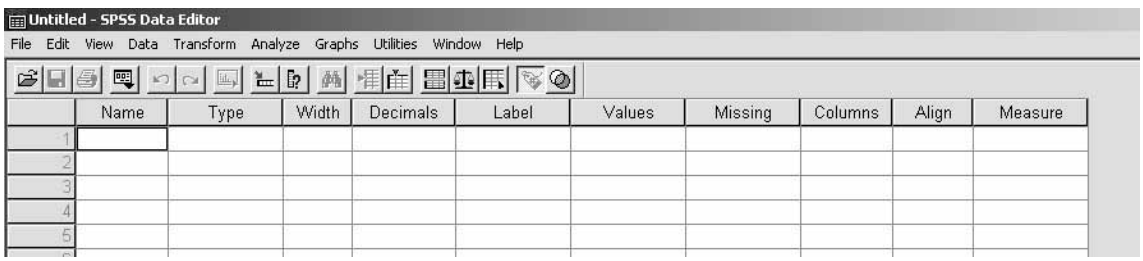
5.3. Chi-squared test for association

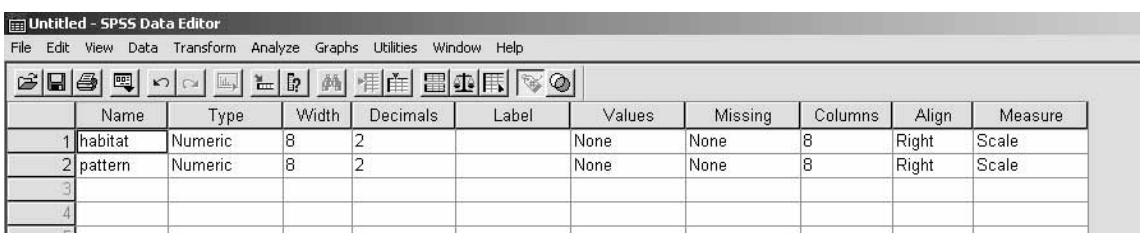
EXAMPLE 5.3. Shell colour in *Cepea nemoralis* in coastal and hedgerow habitats

BOX 5.4. How to calculate an r × c chi-squared test for association

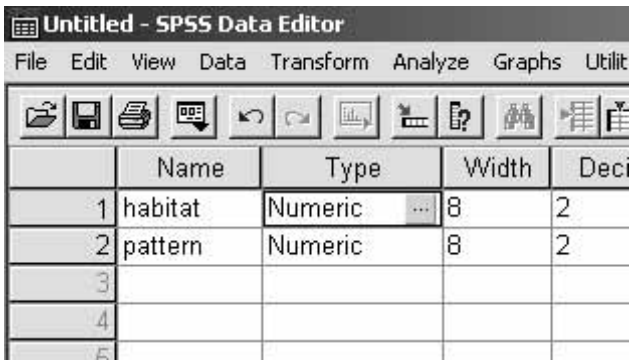
Step 1. Set up the variables.
When SPSS starts up, select 'Variable view' using the tabs at the bottom left. You should get something like this:



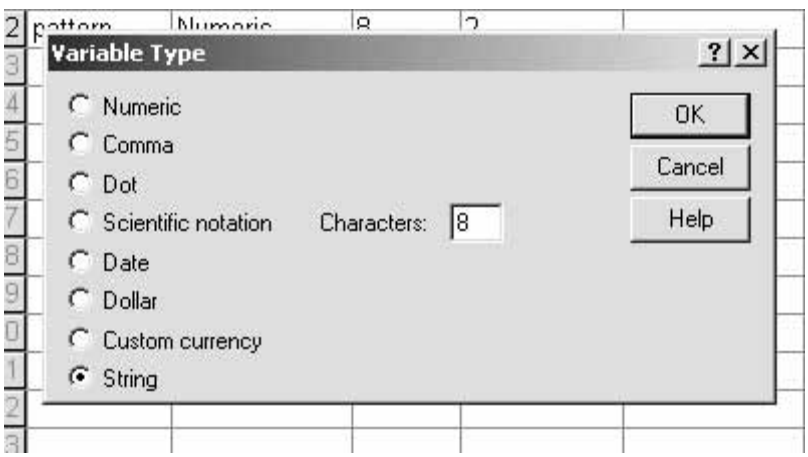
For the first variable name, type in 'habitat', and for the second 'pattern'. Default properties are set for each variable.



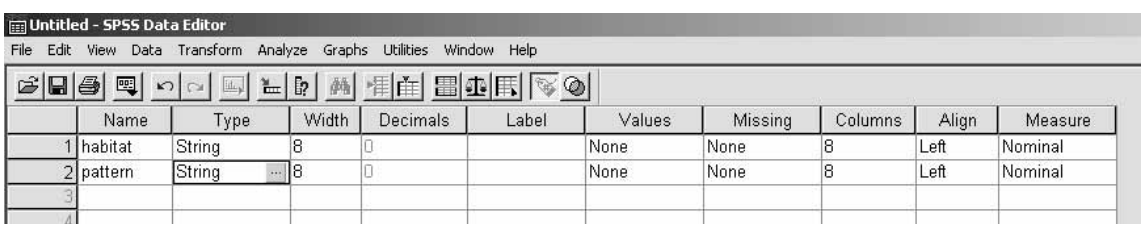
We could use numbers to represent our data, but that would introduce a layer of possible confusion. It is probably best to convert both our variables to text ('string') type by clicking in the 'Type' cell. This should produce a grey area at the right of the cell.



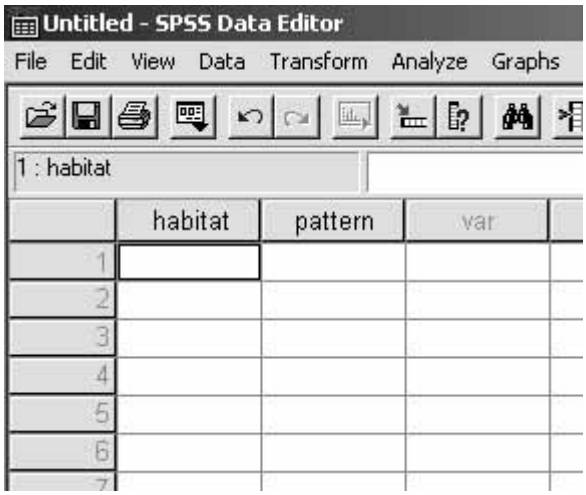
Click on this grey area to get a menu of possible types, and select the radio button for 'String'.



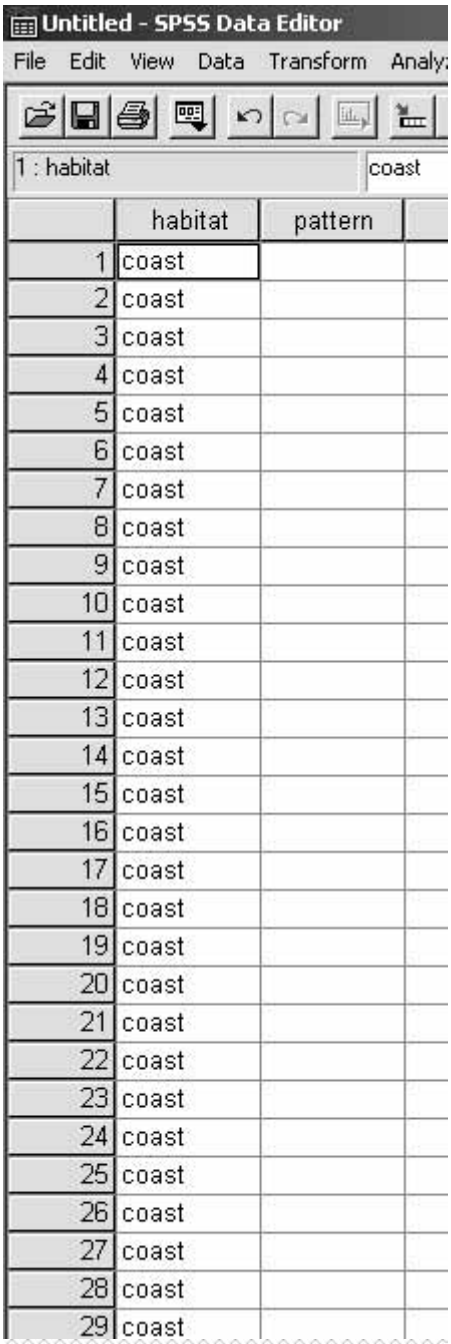
Click on 'OK' to close the dialogue box. Repeat the process for the other variable.



Step 2. Enter the data.
Change to 'Data View' using the tabs at the bottom left of the screen.



We have two habitats: coastal and hedgerow. Each row represents a separate snail, so we need 50 rows with 'coastal' in the habitat column, and 55 rows with 'hedgerow' in the habitat column. To avoid misspellings (which SPSS would interpret as new categories), it is probably best to do some copying and pasting.



We could have changed the number of characters in the string when we set up the variables, but it is just as easy to use codes for the colours. I shall use 'by', 'py', 'bp', and 'pp' for 'banded yellow', 'plain yellow', 'banded pink',

and 'plain pink' respectively. More copying and pasting needed in column 2 . . .

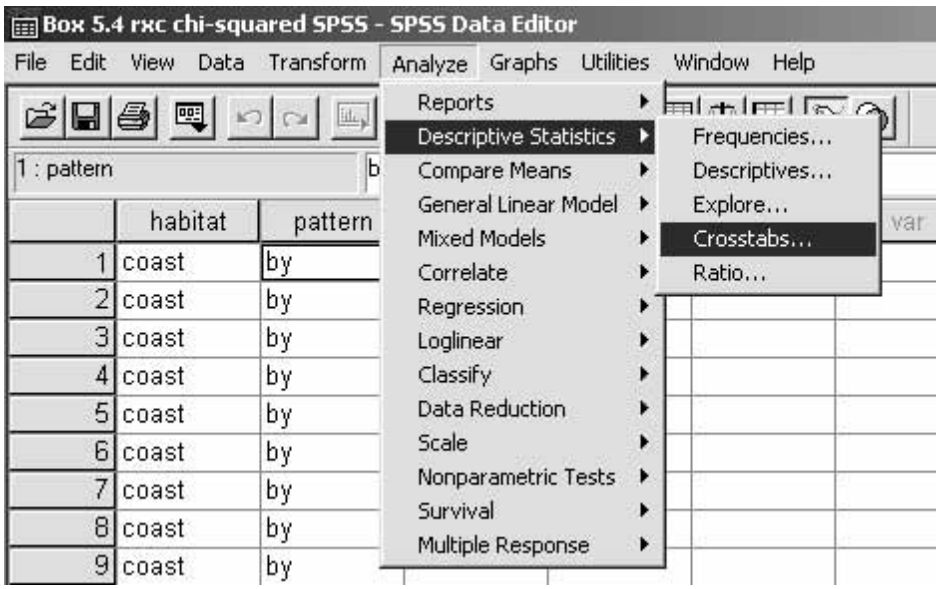
Untitled - SPSS Data Editor

File Edit View Data Transform Analy:

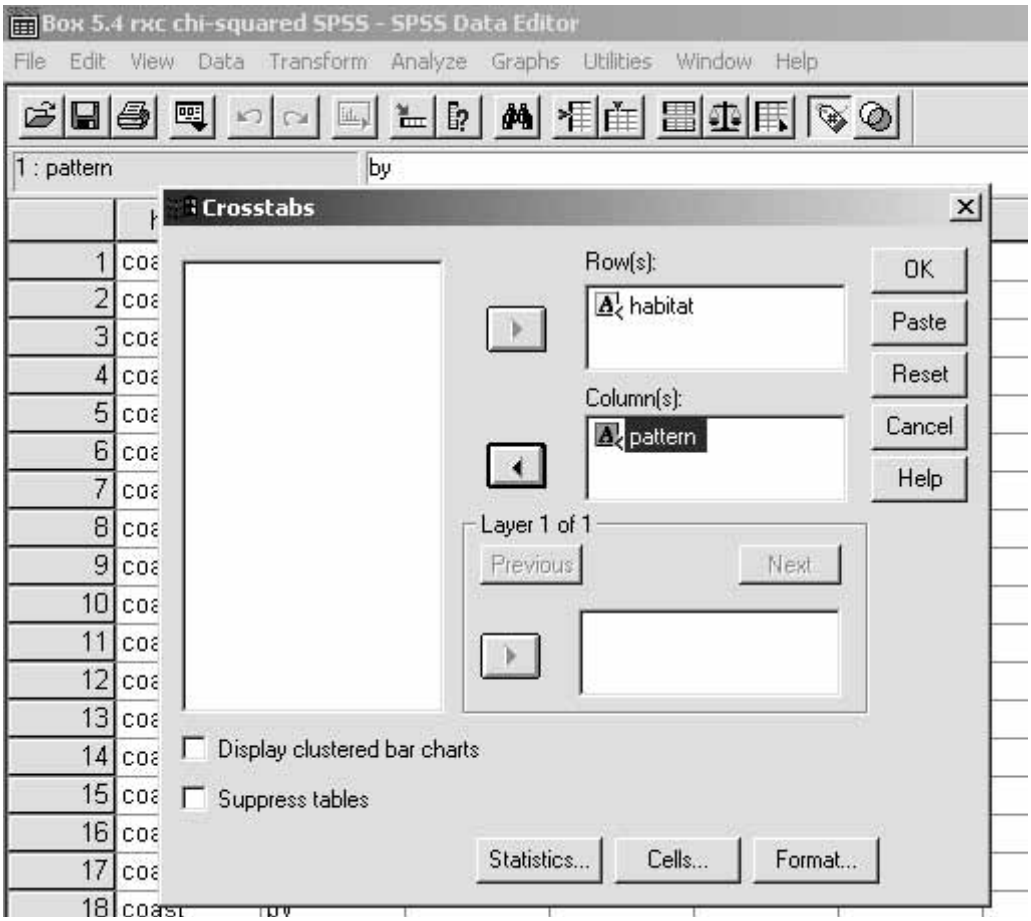
1 : pattern | by

	habitat	pattern
1	coast	by
2	coast	by
3	coast	by
4	coast	by
5	coast	by
6	coast	by
7	coast	by
8	coast	by
9	coast	by
10	coast	by
11	coast	py
12	coast	py
13	coast	py
14	coast	py
15	coast	py
16	coast	py
17	coast	py
18	coast	py
19	coast	py
20	coast	py
21	coast	py
22	coast	py
23	coast	py
24	coast	py
25	coast	py
26	coast	py
27	coast	py
28	coast	py
29	coast	py
30	coast	bp

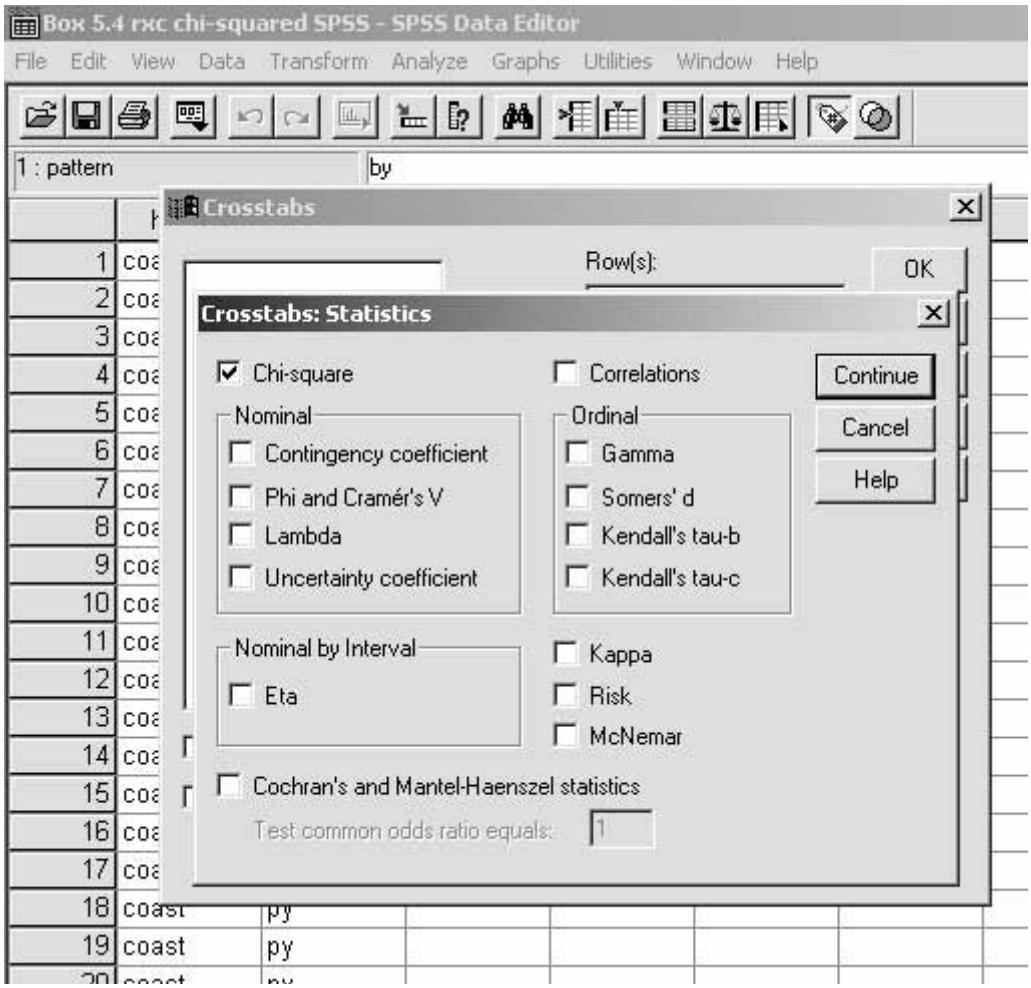
Step 3. Perform the test
Go to 'Analyze', 'Descriptive Statistics', 'Crosstabs'.



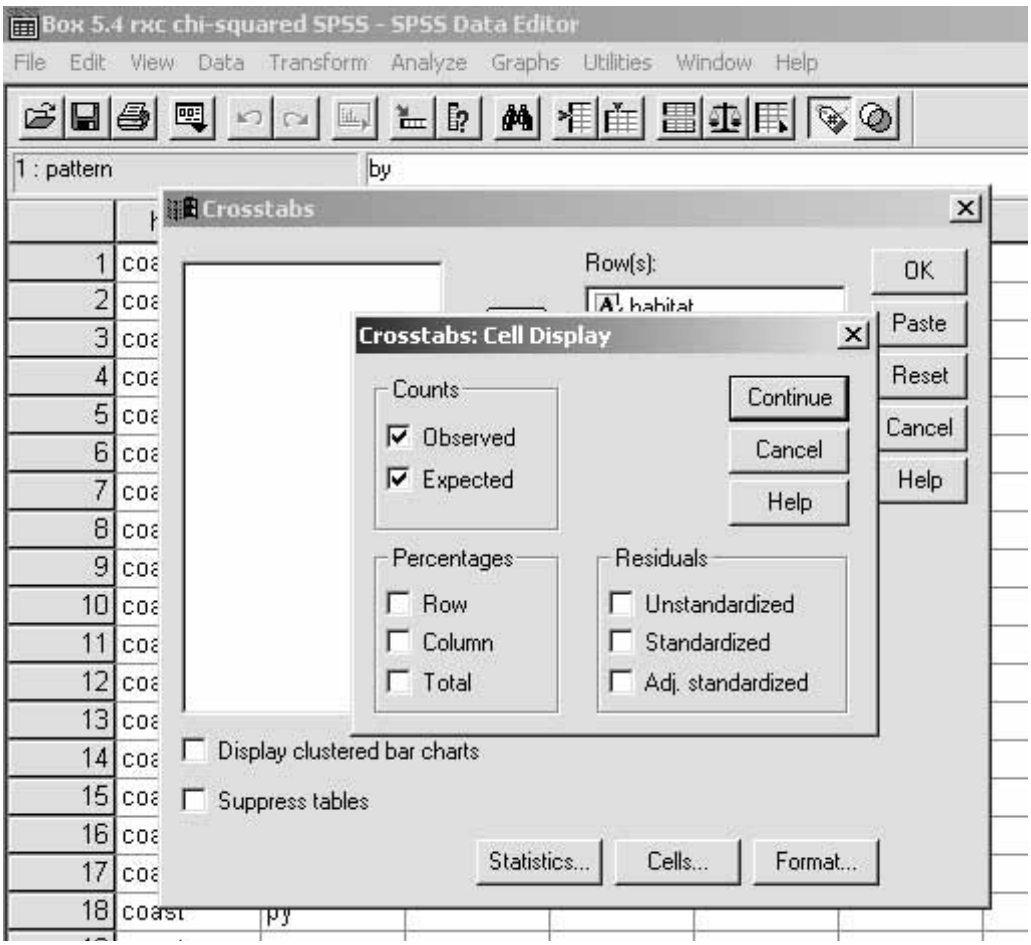
In our original table, the rows were the habitats and the columns were the shell patterns. Transfer the variables to the appropriate windows by clicking on the variable in the left-hand window to highlight them, then clicking on the appropriate arrow to transfer them.



Click on the 'Statistics' button. In the dialogue box, select 'Chi-squared', and click on 'Continue'.



Click on the 'Cells' button, and make sure that both 'Observed' and **'Expected'** are selected.



Click on 'Continue', and then on 'OK'. The output will appear in a separate window.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
HABITAT * PATTERN	105	100.0%	0	.0%	105	100.0%

HABITAT * PATTERN Crosstabulation

			PATTERN				Total
			bp	by	pp	py	
HABITAT	coast	Count	5	10	16	19	50
		Expected Count	11.4	12.9	12.9	12.9	50.0
	hedgerow	Count	19	17	11	8	55
		Expected Count	12.6	14.1	14.1	14.1	55.0
Total		Count	24	27	27	27	105
		Expected Count	24.0	27.0	27.0	27.0	105.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-siced)
Pearson Chi-Square	15.185[a]	3	.002
Likelihood Ratio	15.851	3	.001
N of Valid Cases	105		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 11.43.

Step 4. Decide what the results mean

The value of chi-squared is 15.18, and the Asymp. Sig. is 0.002. We can reject the null hypothesis, and conclude that there is a significant association (better than $p = 0.01$) between habitat and shell pattern.