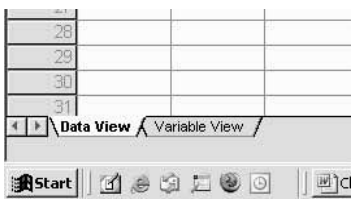
5.1. Chi-squared goodness of fit test

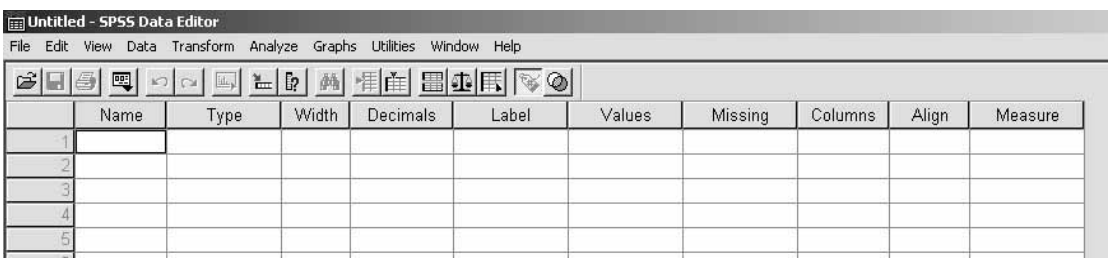
EXAMPLE 5.1. The distribution of holly leaf miners on *Ilex aquifolia*

BOX 5.1. How to calculate a goodness of fit chi-squared test

Step 1. Set up the variables.
(i) When SPSS starts, select the 'Type in data' option.
(ii) Then choose 'Variable View' from the tabs at the bottom left of the screen.

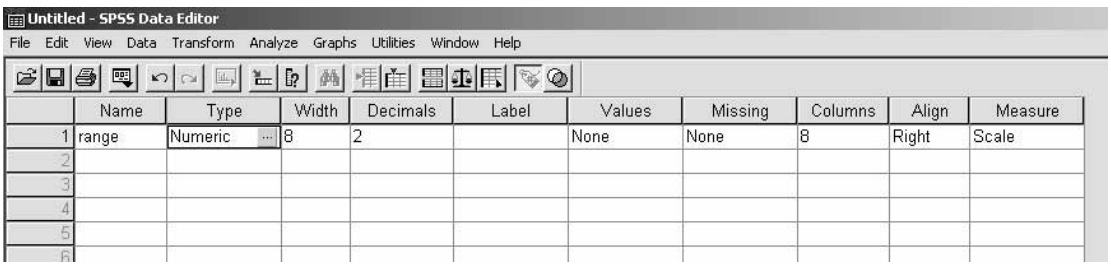


You will see a screen something like this:

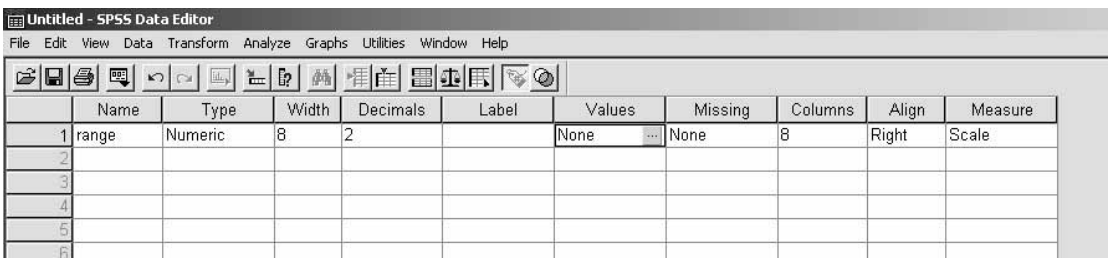


Each row represents a variable for the analysis.

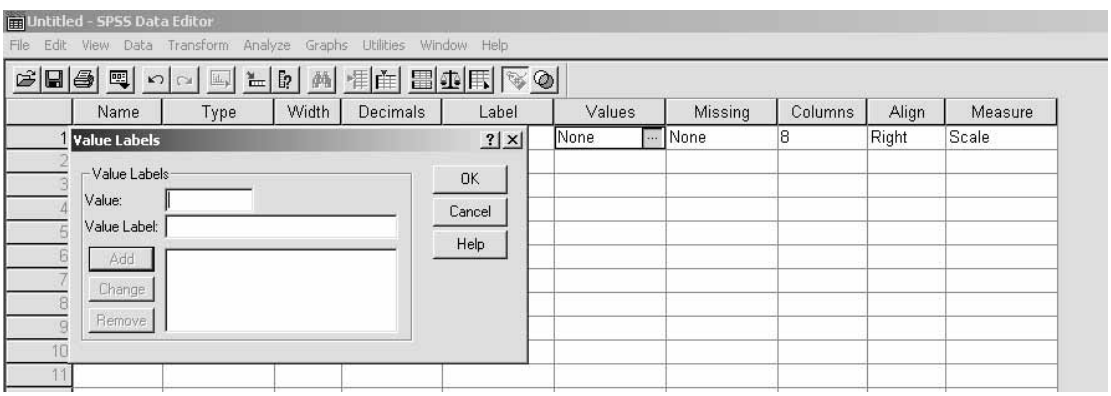
(iii) In the name for variable 1, type 'range' (SPSS won't accept capital letters as parts of a Name, and Names are limited to eight characters, so 'height range' is far too long). Most of the other characteristics of the variable will be give default values as below:



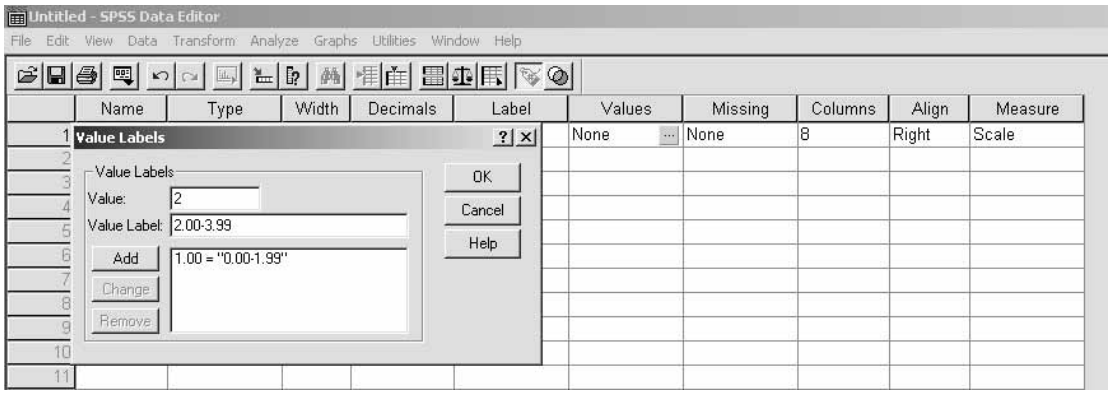
(iv) To input the values for the ranges, click in the 'Values' cell for 'Range'. This will produce a grey area at the right of the cell:



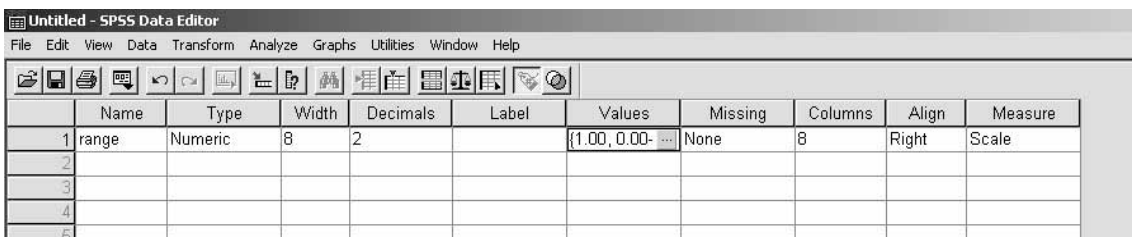
and clicking on this will give a dialogue box for inputting the values.



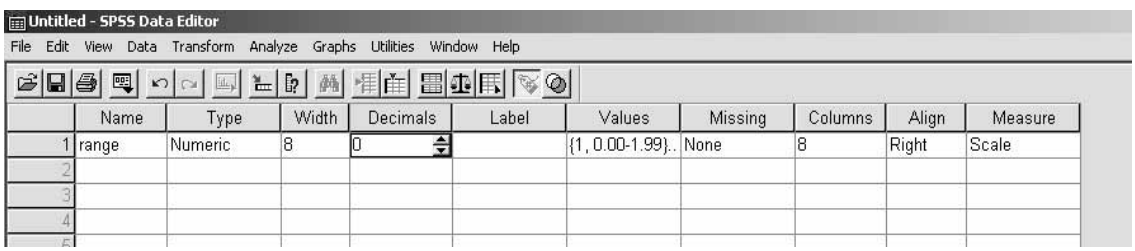
Type in a value of 1 and a value label of '0.00–1.99'. Then click on 'Add'. This will add your new value–label pair to the window at the bottom of the dialogue box.



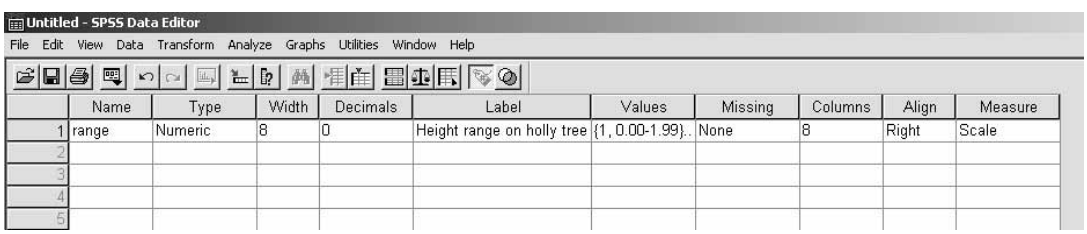
Repeat for the other two value–label pairs (we have three height ranges), then click on 'OK'. You should finish up with something like this:



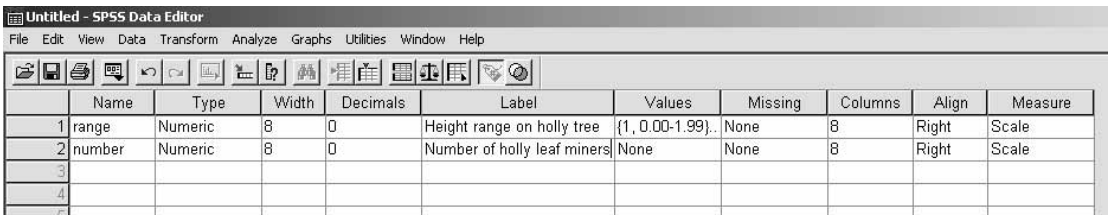
(If you click on the 'Decimals' cell and change the number of decimal places to zero using the arrows, the '1.00' at the start of the Values becomes simply '1'.)



(v) In the 'Label' column, enter a description of the ranges, for example 'Height range on holly tree'. Note that the cell width expands to fit the text.

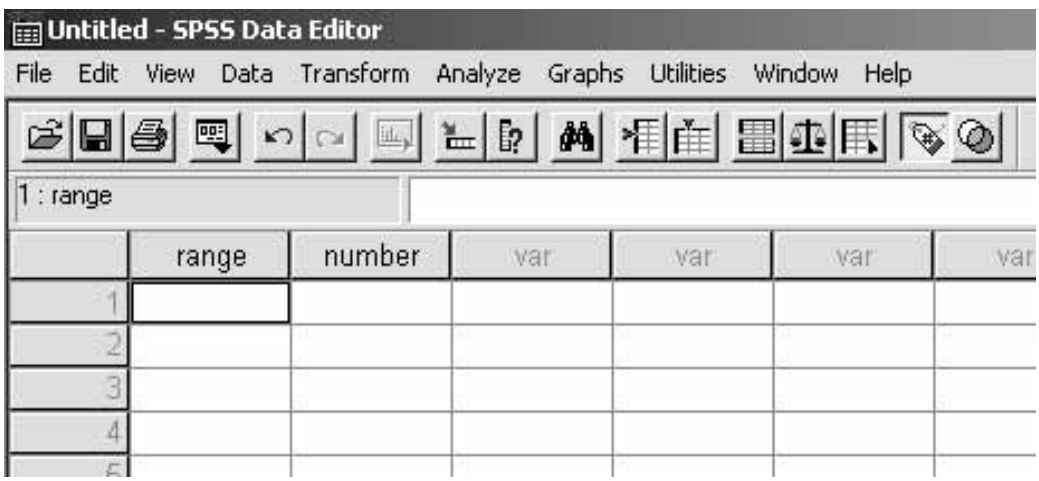


(vi) Next set up the variable to contain the actual counts of holly leaf miners. Give it the name 'number', and set 'Decimals' to zero (we can't have fractions of a holly leaf miner). Put in the label 'Number of holly leaf miners'.

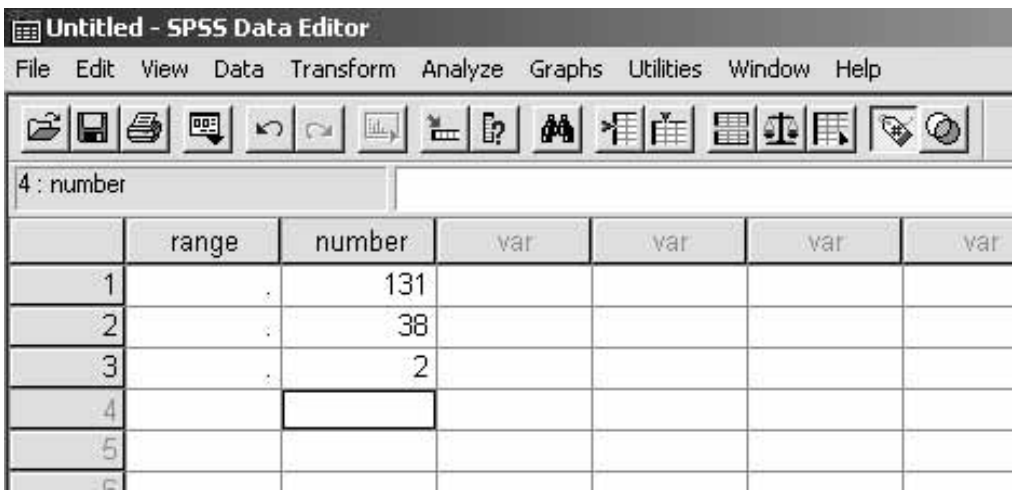


	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	range	Numeric	8	0	Height range on holly tree	{1, 0.00-1.99}..	None	8	Right	Scale
2	number	Numeric	8	0	Number of holly leaf miners	None	None	8	Right	Scale
3										
4										

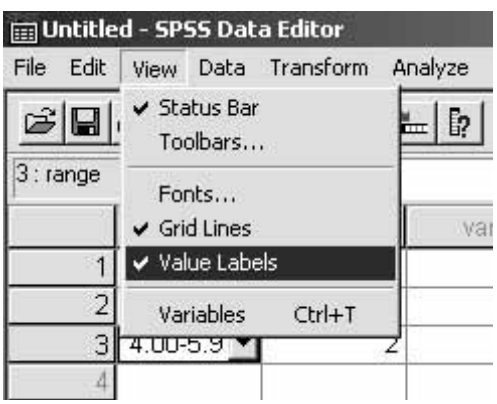
Step 2. Enter the data
(i) Transfer to 'Data View' using the tab at the bottom left of the screen. You should get something like this:



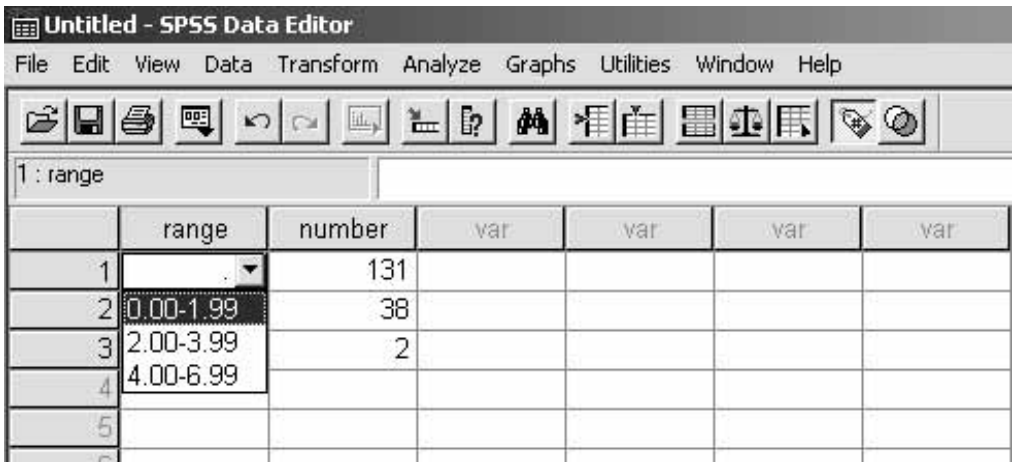
(ii) Another quirk of SPSS is that you have to input the numbers before it lets you put in the labels. Put the numbers of holly leaf miners into the second column:



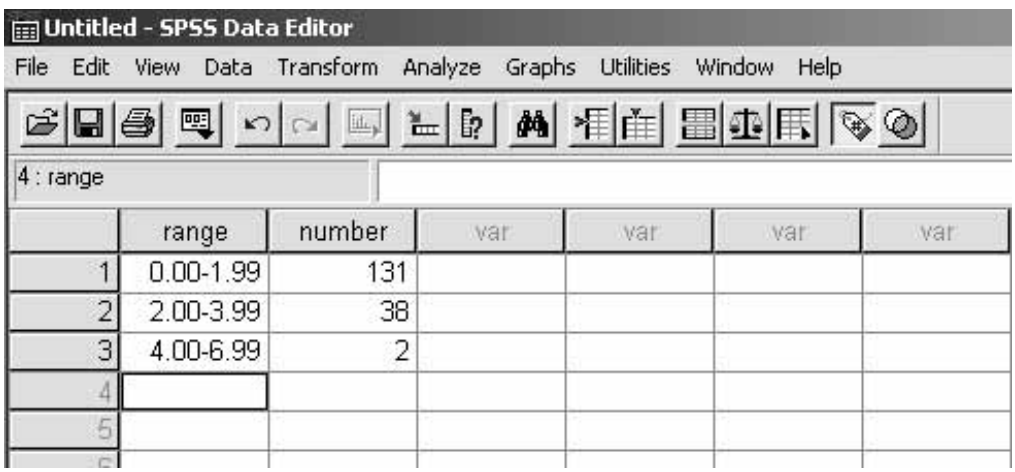
(iii) Check that value labels are enabled by going to 'View' and ensuring that 'Value Labels' is selected.



Now click in the first 'range' cell. You will get a drop-down menu of the values you put in while in variable view.



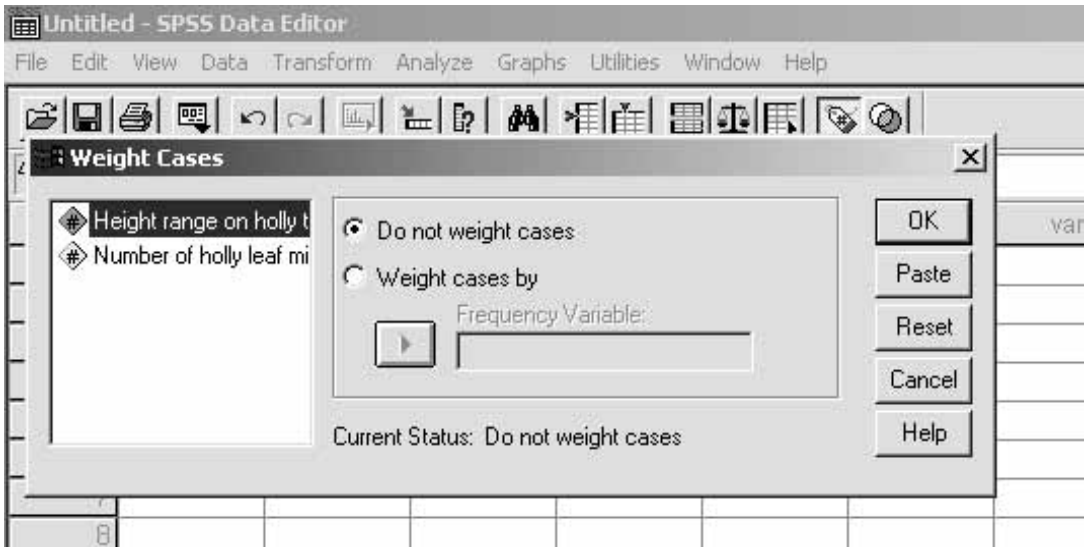
Select the first one (0.00–1.99). Repeat for the other two ranges.



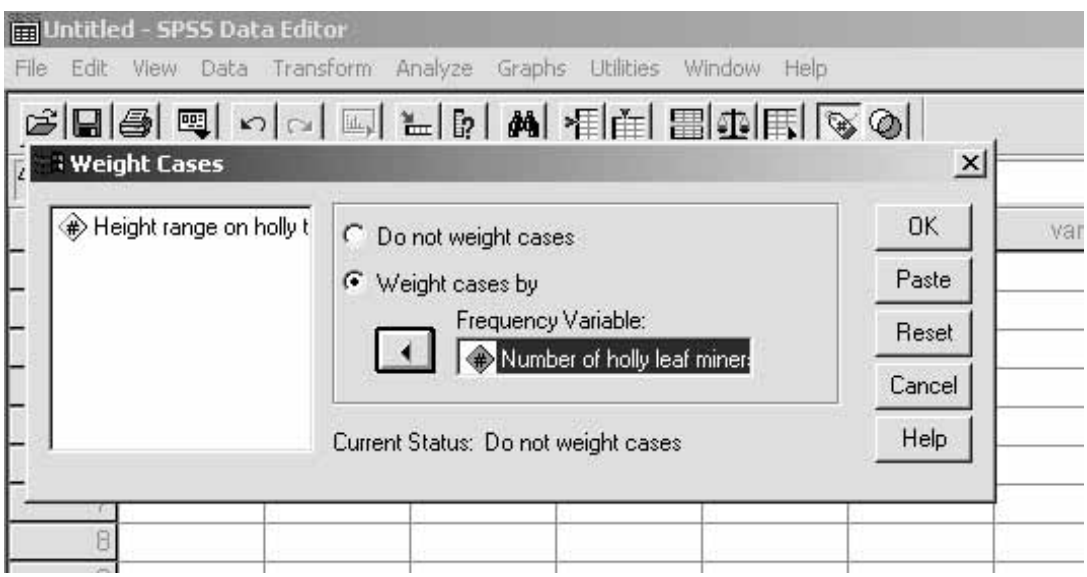
Step 3. Perform the test.

(i) The ranges are our test **variables**, and the weightings for each variable are the number of holly leaf miners. The first thing to do is assign the weightings.

Go to 'Data' and select 'Weight Cases'. This brings up a dialogue box.

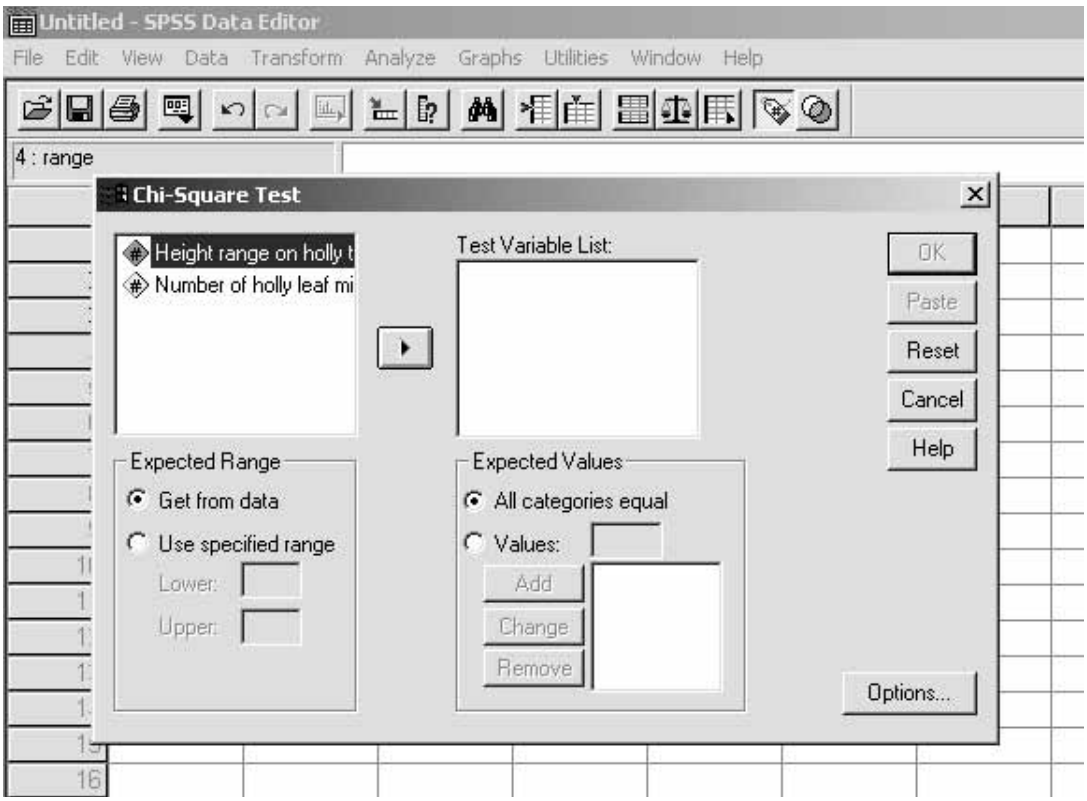


Make sure that the 'Weight cases by' radio button is clicked. In the left-hand window, there will be a list of variables. Select 'Number of holly leaf miners' from the list, and click on the arrow to transfer it to the box labelled 'Frequency Variable'.

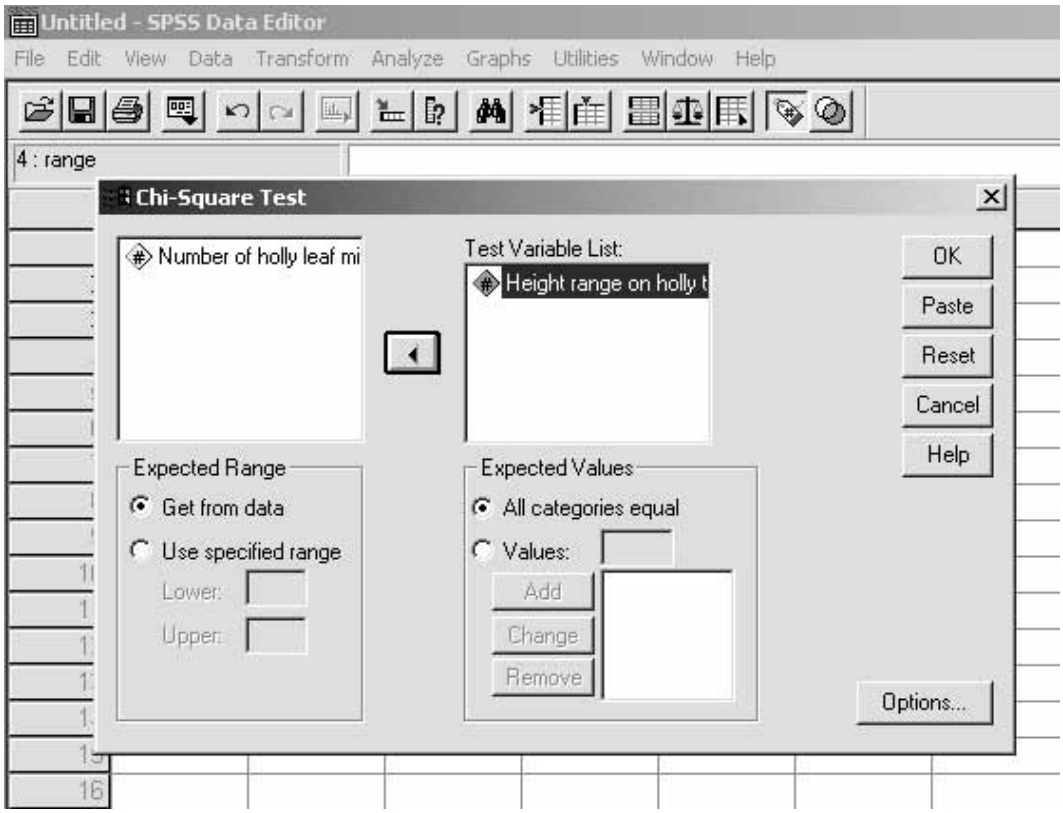


Click on 'OK'.

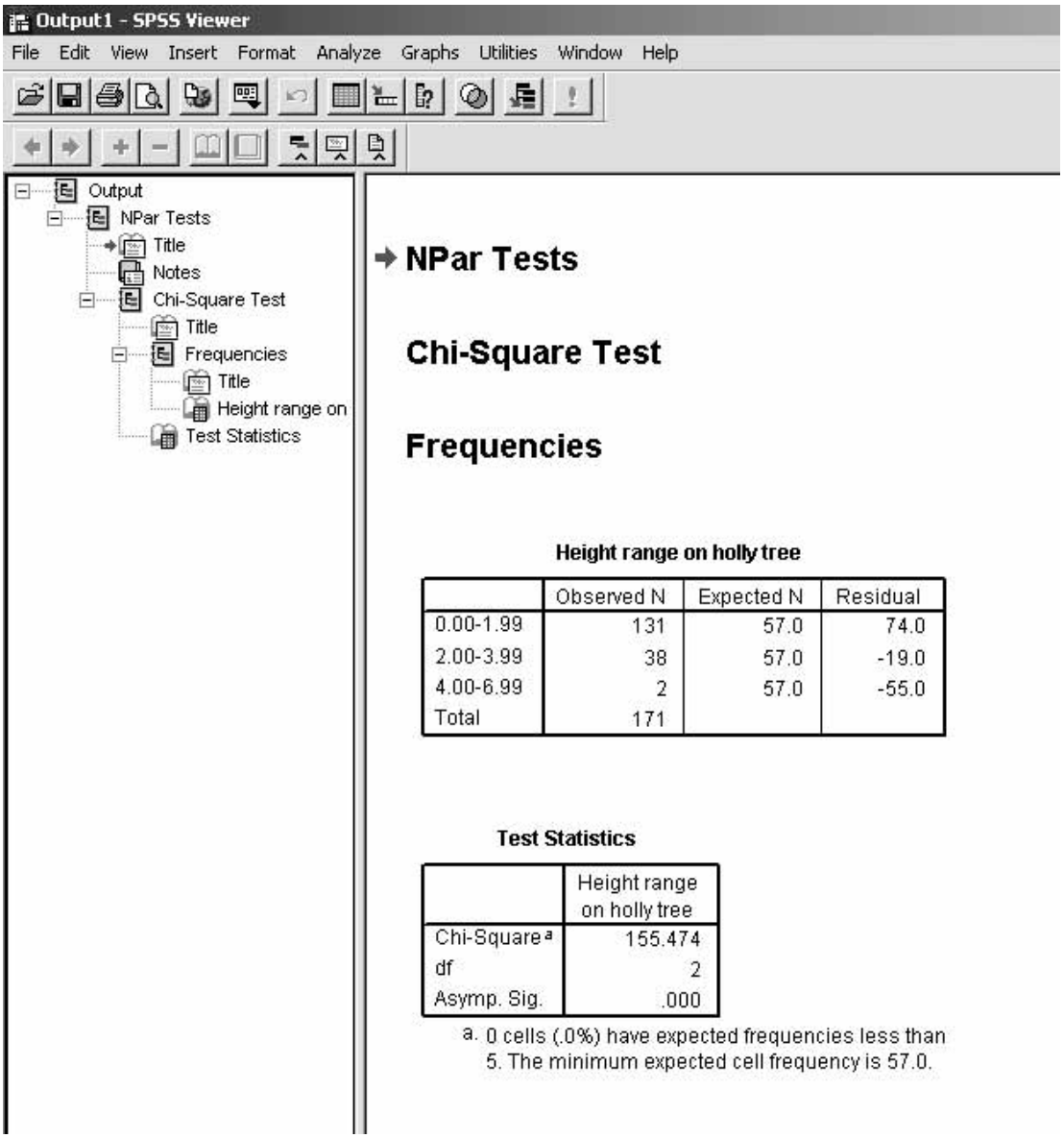
(ii) Now we tell SPSS which test to perform. Go to 'Analyze' [sic], 'Non-Parametric Tests' and 'Chi-Square'.



Select 'Height range on holly tree' and transfer it to the 'Test Variable List' window by clicking on the arrow. In 'Expected Range', select the 'Get From Data' radio button (SPSS will now use the data to calculate the expected values), and in 'Expected Values' select the 'All categories equal' radio button (SPSS will now calculate the expected values by assuming that all categories are equally likely).



(iii) Click on 'OK'. After thinking for a moment, the computer will produce an output window. This will contain two tables. The first table contains a summary of the numbers – the height ranges, the actual and expected numbers, and the residuals (differences between the observed and expected values). The second table contains the results of the test: the value of chi-squared (155.474), the number of **degrees of freedom** (df) (2 in this case), and the 'Asymp. Sig.' (asymptotic significance), the probability of getting this value of chi-squared if the data are evenly distributed – in this case, so small that it doesn't show up in the first three decimal places.



This means that we can reject the null hypothesis, and conclude that there is a difference between the numbers of holly leaf miners found at various heights (m) on the tree compared with those expected.