5.1. Chi-squared goodness of fit test

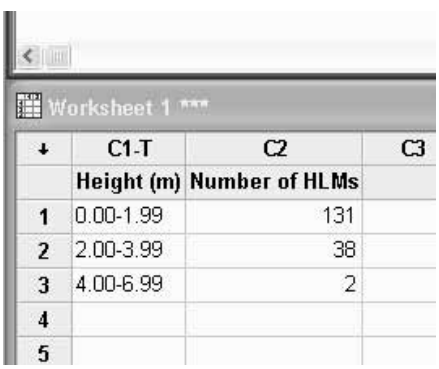
EXAMPLE 5.1. The distribution of holly leaf miners on *Ilex aquifolia*

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

Step 1. Enter the data into the spreadsheet section of the Minitab window. For the holly leaf miner data, the height ranges (classes) will need to be entered as text, so change the first column to text format by clicking on the column heading, then using 'Editor', 'Format column', 'Text'. The column heading should change from plain 'C1' to 'C1-T', which indicates a column of text.

Next, enter the column title ('Height (m)') in the space just below the column heading, and enter the height ranges.

In the second column ('Number of HLMs') enter the recorded numbers of holly leaf miners. (Note that the widths of the columns adjust to accommodate the labels.)

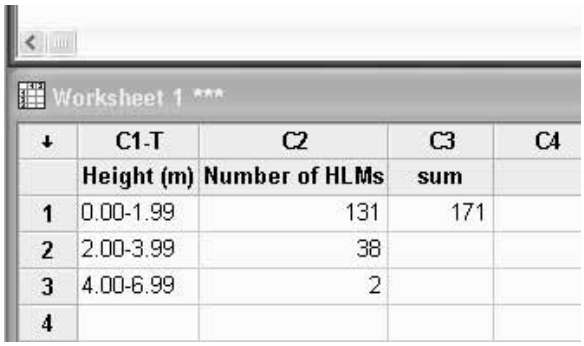


↓	C1-T	C2	C3
	Height (m)	Number of HLMs	
1	0.00-1.99	131	
2	2.00-3.99	38	
3	4.00-6.99	2	
4			
5			

Step 2. Now we need to calculate the expected number of holly leaf miners for each height range, assuming that they are all equal.

Go to 'Calc', 'Calculator', and enter 'Sum' in the 'store result in variable' window.

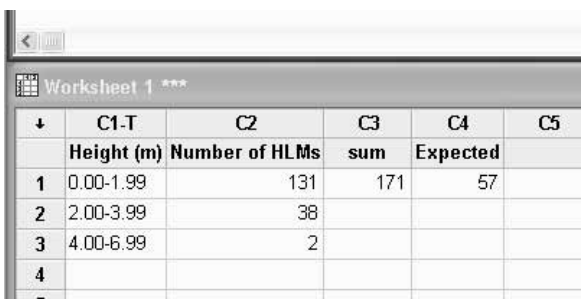
Next, go to the 'Expression' window and enter 'Sum(c2)'. (Alternatively, scroll down the 'function' window and click on 'sum', then 'select; then go to the left-hand window, click on 'C2 Number of HLMs' and 'Select'.) Now click on 'OK'. The sum should appear in the spreadsheet window.



Worksheet 1 ***

↓	C1-T	C2	C3	C4
	Height (m)	Number of HLMs	sum	
1	0.00-1.99	131	171	
2	2.00-3.99	38		
3	4.00-6.99	2		
4				

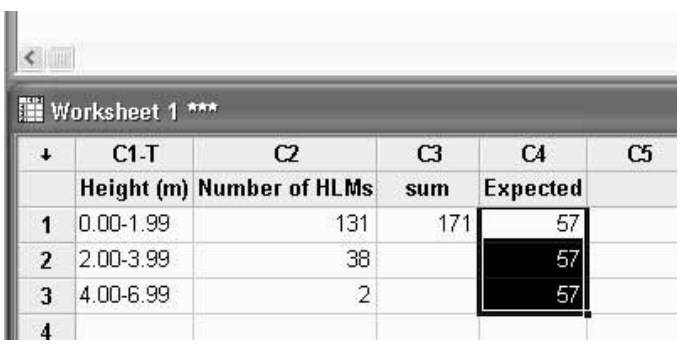
Go to 'Calc', 'Calculator' again, and this time put 'Expected' in the 'Store Results in Variable' window. In the 'Expression' window, enter 'sum/3' (where 3 is the number of categories of height) and press 'OK'.



Worksheet 1 ***

↓	C1-T	C2	C3	C4	C5
	Height (m)	Number of HLMs	sum	Expected	
1	0.00-1.99	131	171	57	
2	2.00-3.99	38			
3	4.00-6.99	2			
4					

Now click on the cell containing the number 57. There will be a small handle at the bottom right. Hover the cursor over this, and hold down the left mouse button. Now drag the handle down until two more cells are highlighted, and release the mouse button.



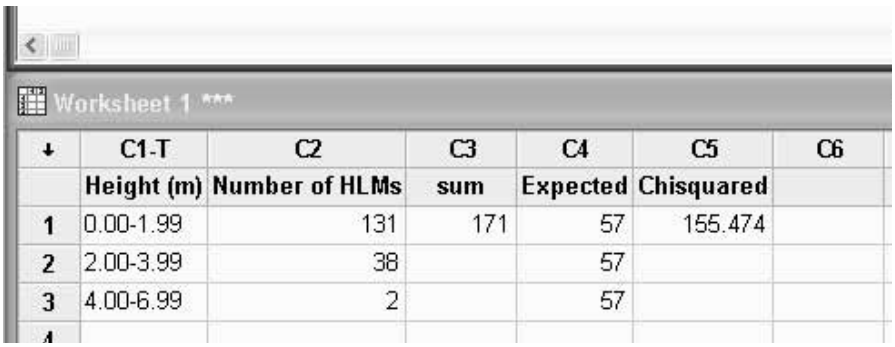
Worksheet 1 ***

↓	C1-T	C2	C3	C4	C5
	Height (m)	Number of HLMs	sum	Expected	
1	0.00-1.99	131	171	57	
2	2.00-3.99	38		57	
3	4.00-6.99	2		57	
4					

We now have a column of expected values to go with our column of actual values.

Step 3. Perform the chi-squared goodness of fit test.

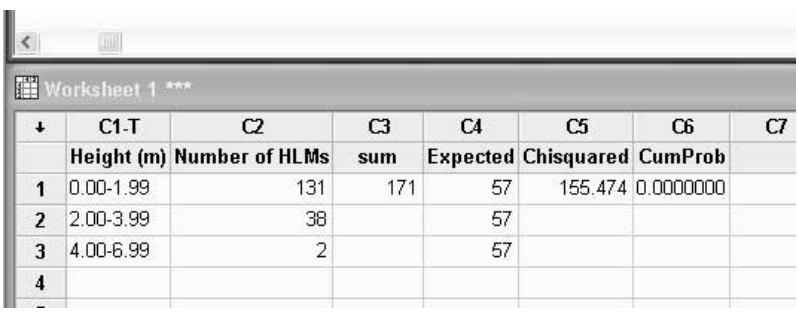
First, calculate chi-squared. This is done by going to 'Calc', 'Calculator', typing 'Chi-squared' in the 'Store results in variable' window, and 'sum((c2-c4)**2/c4)' in the 'Expression' window. Then hit 'OK'.



Worksheet 1 ***

	C1-T	C2	C3	C4	C5	C6
	Height (m)	Number of HLMs	sum	Expected	Chisquared	
1	0.00-1.99	131	171	57	155.474	
2	2.00-3.99	38		57		
3	4.00-6.99	2		57		
4						

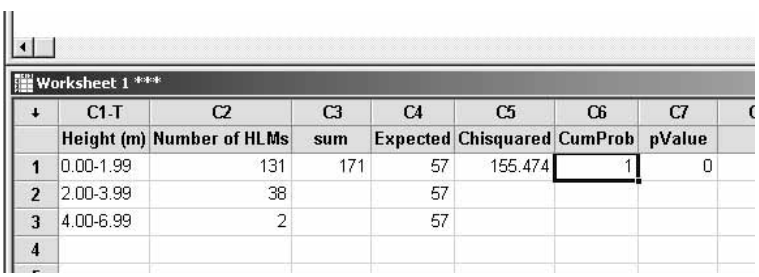
Go to 'Calc', 'Probability Distributions', 'Chi-Squared'. Select 'Cumulative Probability', and enter '2' (for this example) in the 'Degrees of Freedom' window. For the 'Input column', enter 'Chisquared', and for 'Optional storage' enter 'CumProb'. Then click on 'OK'.



Worksheet 1 ***

	C1-T	C2	C3	C4	C5	C6	C7
	Height (m)	Number of HLMs	sum	Expected	Chisquared	CumProb	
1	0.00-1.99	131	171	57	155.474	0.0000000	
2	2.00-3.99	38		57			
3	4.00-6.99	2		57			
4							

The final step is to find the probability that the null hypothesis can be rejected. Go to 'Calc', 'Calculator'. Enter 'pValue' in the 'Store results in variable' window, and type '1-CumProb' in the 'Expression' window. Click on 'OK'.



Worksheet 1 ***

	C1-T	C2	C3	C4	C5	C6	C7
	Height (m)	Number of HLMs	sum	Expected	Chisquared	CumProb	pValue
1	0.00-1.99	131	171	57	155.474	1	0
2	2.00-3.99	38		57			
3	4.00-6.99	2		57			
4							

In this example the *p* value is very small (the computer gives it a value of 0.0). Therefore the null hypothesis can be rejected. (The very high value of

chi-squared should have been a strong hint.) We therefore conclude that there is a very highly significant difference ($\chi^2_{\text{calculated}} = 155.47, p < 0.001$) between the numbers of holly leaf miners found at the various levels on the tree compared with those expected, such that the holly leaf miners are not found in equal numbers at all heights.