5.5.1. G goodness of fit test

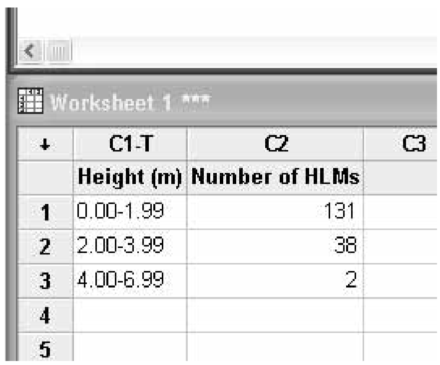
EXAMPLE 5.1. The distribution of holly leaf miners on *Ilex aquifolia*

BOX 5.7. How to calculate a goodness of fit *G* test

Step 1. Enter the data into the spreadsheet section of the Minitab window. For the holly leaf miner data, the height ranges 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.

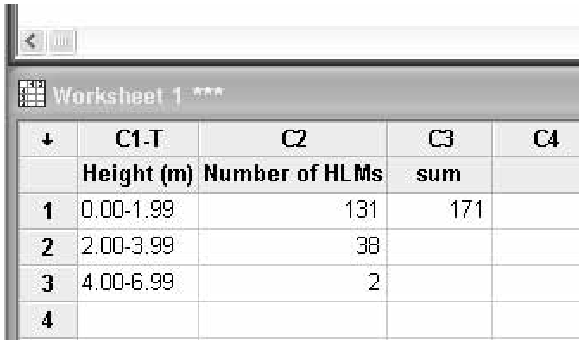
In the second column ('Number of HLMs') enter the recorded numbers of holly leaf miners. (Note that the widths of the columns adjust to accommodate the labels.)



↓	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 (class), assuming that they are all equal.

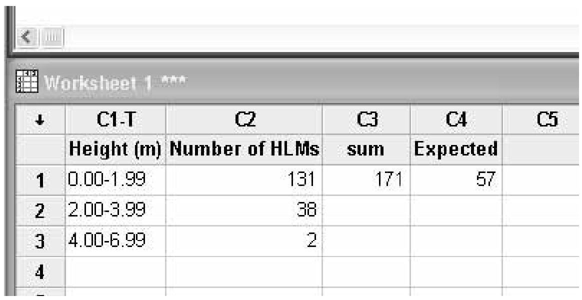
Go to 'Calc', 'Calculator', and enter 'Sum' in the 'Store result in variable' window.
Next, go to the 'Expression' window and enter 'Sum(c2)'. (Alternatively, scroll down the 'Function' window and click on 'Sum', then 'Select'; then go to the left-hand window, click on 'C2 Number of HLMs' and 'Select'.)
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				

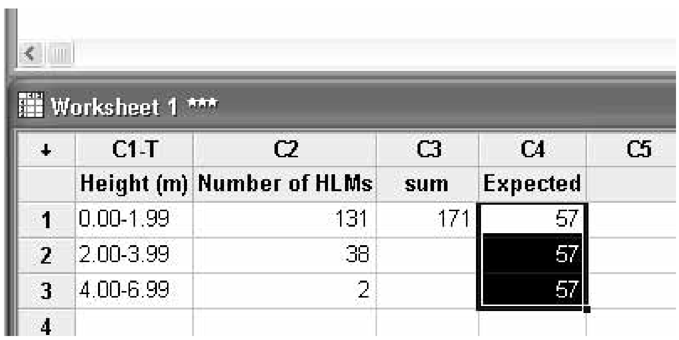
Go to ‘Calc’, ‘Calculator’ again, and this time put ‘Expected’ in the ‘Store results in variable’ window. In the ‘Expression’ window, enter ‘sum/3’ (where 3 is the number of categories of height) and press ‘OK’.



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.



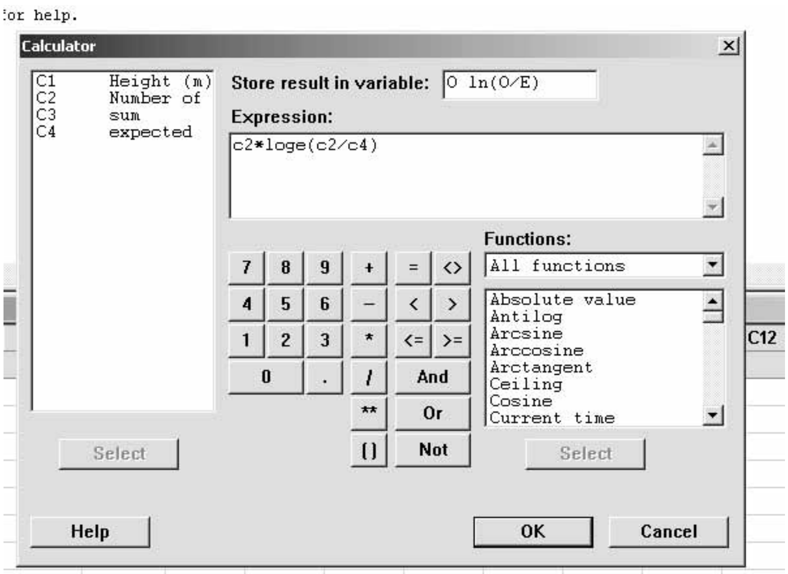
Worksheet 1 ***

↓	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					

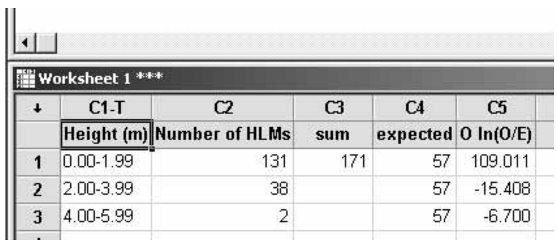
We now have a column of expected values to go with our column of actual values.

Step 3. Perform the test.

First, calculate the values of 'O ln (O/E)' for each height range, where O is the observed value and E is the expected value. Go to 'Calc', 'Calculator', put 'O ln(O/E)' in the 'Store result as variable', and in the 'Expression' window type 'c2*loge(c2/c4)'. 'loge' is a function that calculates natural logarithms (logarithms to base e, where e = 2.718 . . .).



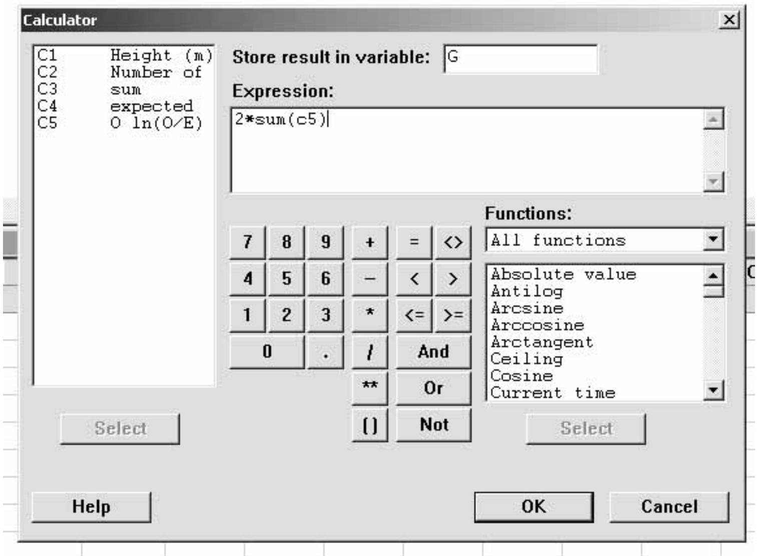
Click on 'OK'.



Worksheet 1 ***

↓	C1-T	C2	C3	C4	C5
	Height (m)	Number of HLMs	sum	expected	O ln(O/E)
1	0.00-1.99	131	171	57	109.011
2	2.00-3.99	38		57	-15.408
3	4.00-5.99	2		57	-6.700

G is twice the sum of the values of O ln(O/E). Go to 'Calc', Calculator' and enter 'G' in the 'Store result as variable' window. Type '2*sum(c5)' in the 'Expression' window.

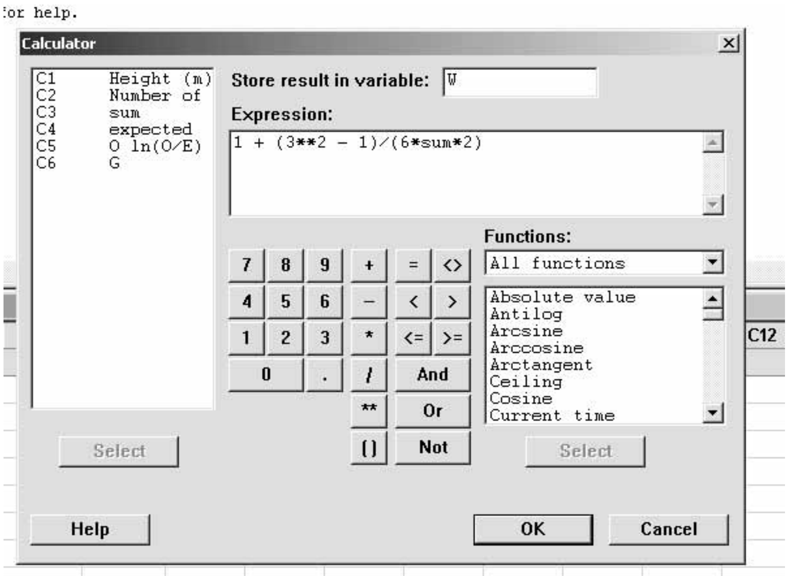


Click on 'OK'.

Worksheet 1 ***

↓	C1-T	C2	C3	C4	C5	C6
	Height (m)	Number of HLMs	sum	expected	O ln(O/E)	G
1	0.00-1.99	131	171	57	109.011	173.807
2	2.00-3.99	38		57	-15.408	
3	4.00-5.99	2		57	-6.700	
4						

Now we do the Williams' correction. $W = 1 + (a^2 - 1)/6nv$, and in this case, a = number of categories = 3, n = the total number of observations = 171, and v = number of degrees of freedom = $a - 1 = 2$ in this case. Go to 'Calc', Calculator', and enter 'W' in the 'Store results as variable' window. In the 'Expression window, type '1 + (3**2 – 1)/(6*sum*2)'.

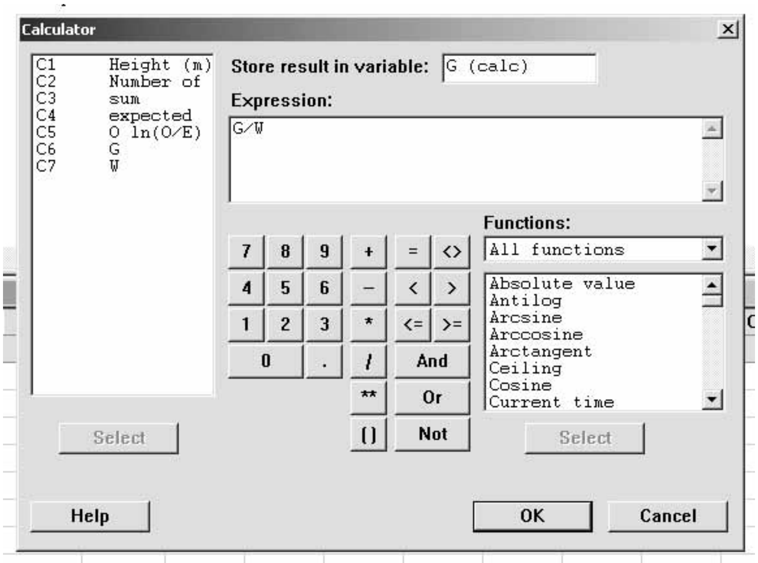


Click on 'OK'.

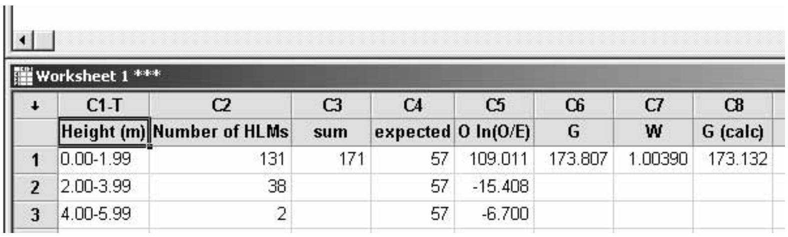
Worksheet 1 ***

↓	C1-T	C2	C3	C4	C5	C6	C7
	Height (m)	Number of HLMs	sum	expected	O ln(O/E)	G	W
1	0.00-1.99	131	171	57	109.011	173.807	1.00390
2	2.00-3.99	38		57	-15.408		
3	4.00-5.99	2		57	-6.700		
4							

Find $G_{calculated}$ by dividing the original value of G by W. Go to 'Calc', 'Calculator', enter 'G (calc)' in the 'Store results as variable' window, and type 'G/W' in the 'Expression' window.



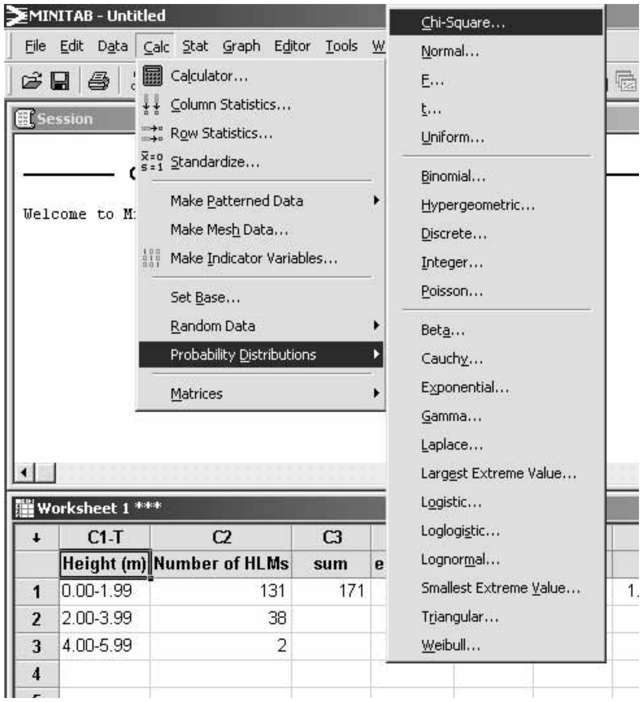
Click on 'OK'.



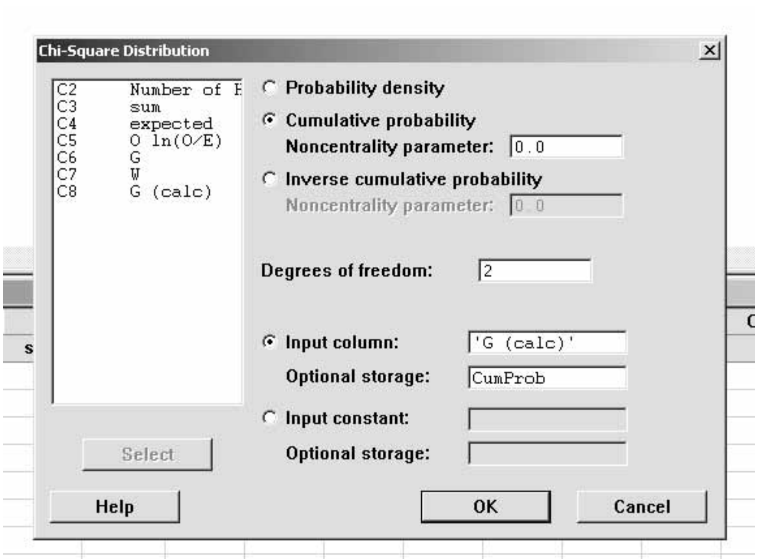
Worksheet 1 ***

↓	C1-T	C2	C3	C4	C5	C6	C7	C8
	Height (m)	Number of HLMs	sum	expected	O ln(O/E)	G	W	G (calc)
1	0.00-1.99	131	171	57	109.011	173.807	1.00390	173.132
2	2.00-3.99	38		57	-15.408			
3	4.00-5.99	2		57	-6.700			

Treat this as a chi-squared value to find the probability. Go to 'Calc', 'Probability Distributions', and select 'chi-square'.



Select ‘Cumulative probability’, enter ‘2’ for the degrees of freedom, enter ‘G (calc)’ or ‘c9’ for the input column, and ‘CumProb’ for the optional storage.



Click on 'OK'.

Worksheet 1 ***

↓	C1-T	C2	C3	C4	C5	C6	C7	C8	C9
	Height (m)	Number of HLMs	sum	expected	O ln(O/E)	G	W	G (calc)	CumProb
1	0.00-1.99	131	171	57	109.011	173.807	1.00390	173.132	1
2	2.00-3.99	38		57	-15.408				
3	4.00-5.99	2		57	-6.700				

The p value is found by subtracting the 'CumProb' value from 1, and this clearly gives zero within the limits of accuracy of the software. Therefore the probability of rejection of the null hypothesis (that the observed values come from a uniform distribution) is so close to 1 that the computer can't tell the difference. Since $p = 0.0$ is less than the threshold of $p = 0.05$ we reject the null hypothesis. There is a significant difference ($G = 173.13$, $p < 0.001$) 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.