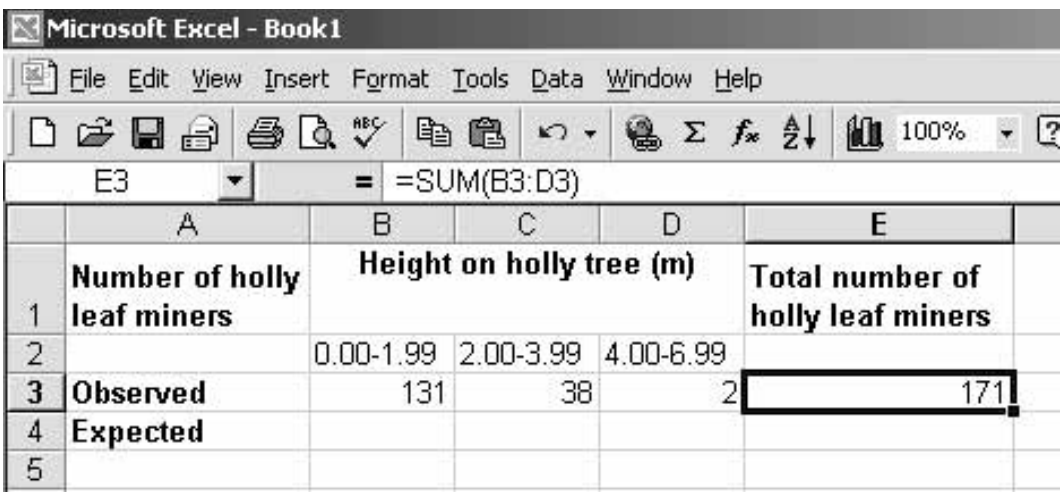
5.1. Chi-squared goodness of fit test

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

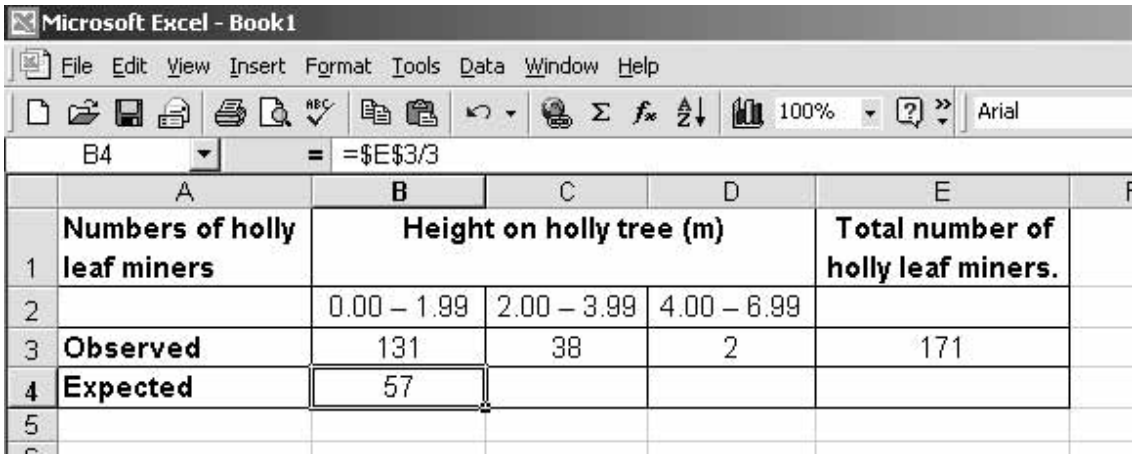
BOX 5.1. How to calculate a goodness of fit chi-squared test

Step 1. Enter the data into the spreadsheet using suitable headings. The format of a results table is usually OK, and we can get Excel to calculate the total number of observations for us.



	A	B	C	D	E
1	**Number of holly leaf miners**	**Height on holly tree (m)**			**Total number of holly leaf miners**
2		0.00-1.99	2.00-3.99	4.00-6.99	
3	**Observed**	131	38	2	171
4	**Expected**				
5					

Step 2. Calculate the expected values. In this case, we are testing against a random distribution, so we expect the same number of holly leaf miners in each height range. We have three ranges, and a total of 171 holly leaf miners (recorded in cell e3 in the example above), so we expect 171/3 holly leaf miners in each range. Let Excel do the calculation for you by entering '=e3/3' into cell b4. (The '$' signs mean an absolute (rather than relative) reference to the contents of cell e3. We use this because we are going to drag the formula across into the other three cells.)

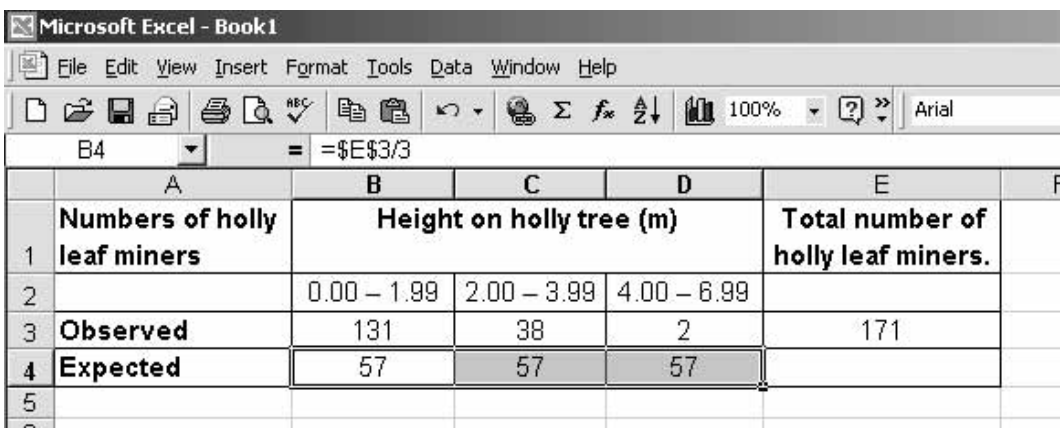

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help

B4 = =$E$3/3

	A	B	C	D	E
1	**Numbers of holly leaf miners**	**Height on holly tree (m)**			**Total number of holly leaf miners.**
2		0.00 – 1.99	2.00 – 3.99	4.00 – 6.99	
3	**Observed**	131	38	2	171
4	**Expected**	57			
5					

Place the cursor (usually a thick vertical–horizontal cross) on the bottom right-hand corner of cell b4. It should change to a thin horizontal–vertical cross. Hold down the left mouse button and drag the cursor across until cells c4 and d4 are highlighted, then release the mouse button.


Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help

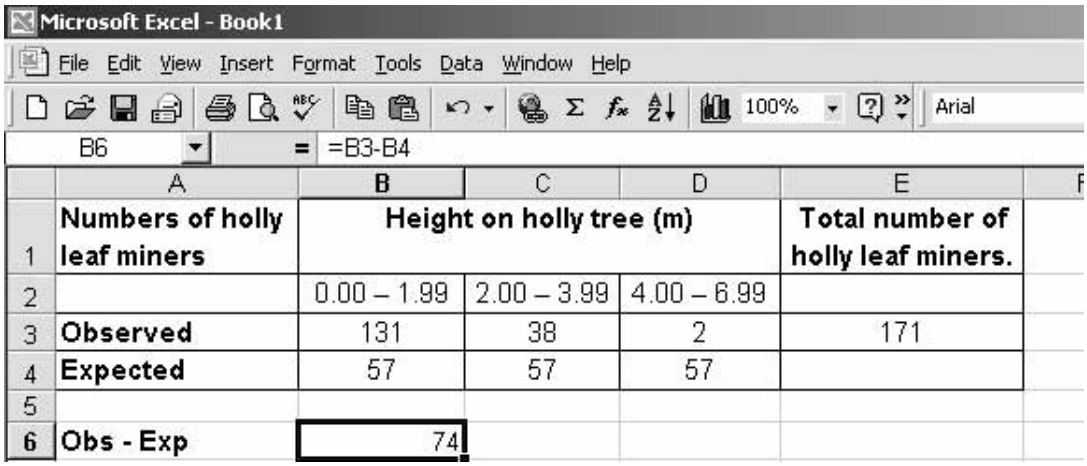
B4 = =$E$3/3

	A	B	C	D	E
1	**Numbers of holly leaf miners**	**Height on holly tree (m)**			**Total number of holly leaf miners.**
2		0.00 – 1.99	2.00 – 3.99	4.00 – 6.99	
3	**Observed**	131	38	2	171
4	**Expected**	57	57	57	
5					

We now have our 'expected' values in the table.

Step 3. Work out chi-squared. Unless you are a maths whiz, it is best to do this in small steps.

First, find the difference between the observed and expected values. Enter '=B3-B4' into cell b6, say, then drag the result across into cells c6 and d6.

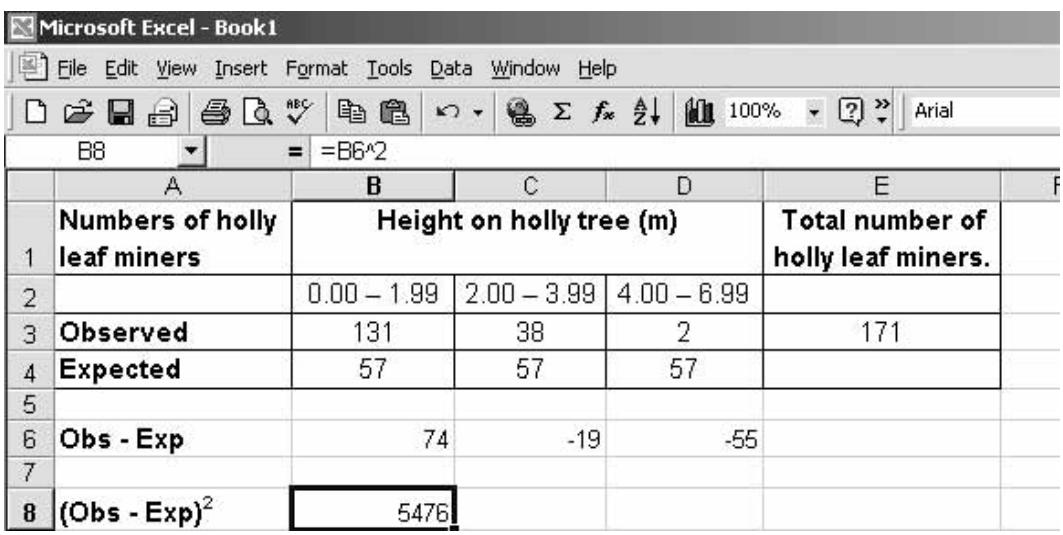

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help

B6 = =B3-B4

	A	B	C	D	E
1	**Numbers of holly leaf miners**	**Height on holly tree (m)**			**Total number of holly leaf miners.**
2		0.00 – 1.99	2.00 – 3.99	4.00 – 6.99	
3	**Observed**	131	38	2	171
4	**Expected**	57	57	57	
5					
6	**Obs - Exp**	74			

Next, square these values by entering '=B6^2' into cell b8, say, and drag the result across into cells c8 and d8.

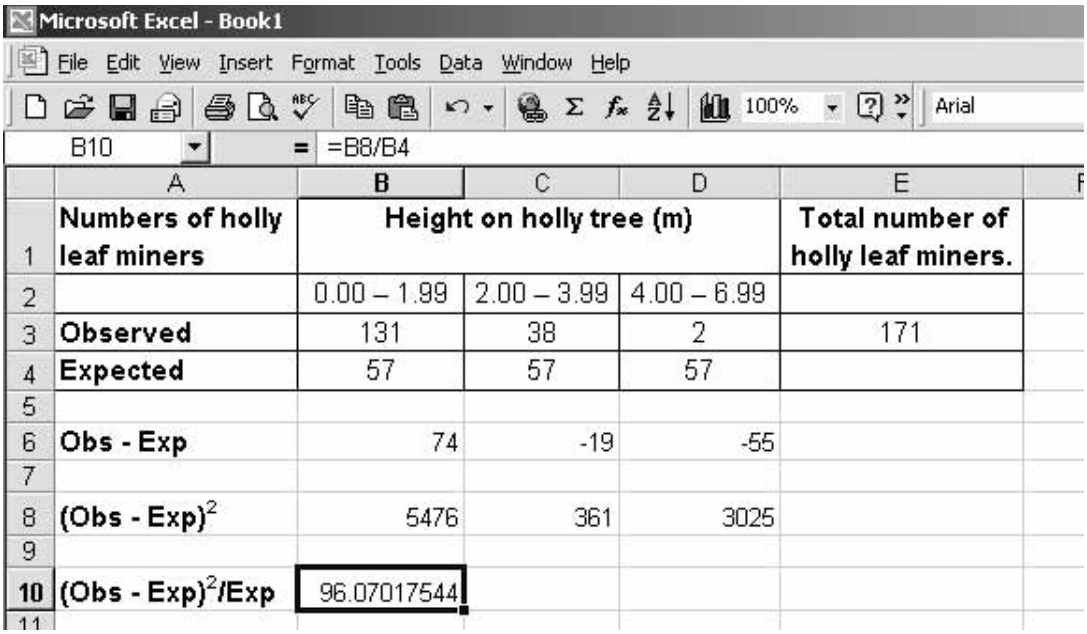

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help

B8 = =B6^2

	A	B	C	D	E
1	**Numbers of holly leaf miners**	**Height on holly tree (m)**			**Total number of holly leaf miners.**
2		0.00 – 1.99	2.00 – 3.99	4.00 – 6.99	
3	**Observed**	131	38	2	171
4	**Expected**	57	57	57	
5					
6	**Obs - Exp**	74	-19	-55	
7					
8	**(Obs - Exp)^2**	5476			

Then divide this by the expected number of holly leaf miners by entering '=b8/b4' into cell b10, say, and dragging across into cells c10 and d10.


Microsoft Excel - Book1

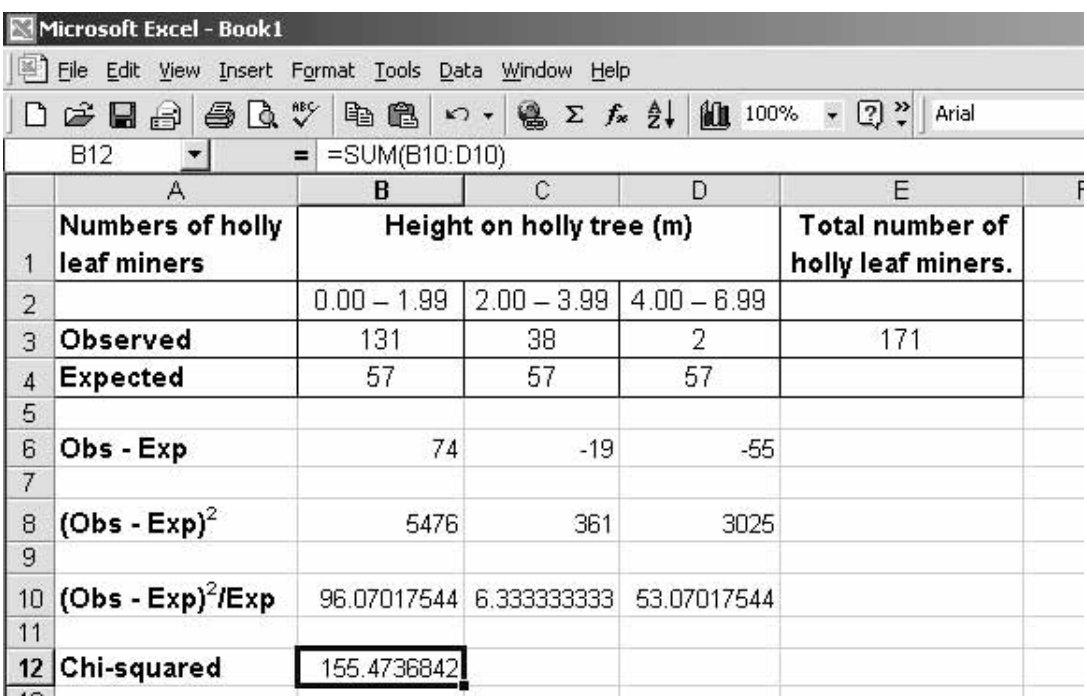
File Edit View Insert Format Tools Data Window Help

100% Arial

B10 = =B8/B4

	A	B	C	D	E
1	**Numbers of holly leaf miners**	**Height on holly tree (m)**			**Total number of holly leaf miners.**
2		0.00 – 1.99	2.00 – 3.99	4.00 – 6.99	
3	**Observed**	131	38	2	171
4	**Expected**	57	57	57	
5					
6	**Obs - Exp**	74	-19	-55	
7					
8	**(Obs - Exp)^2**	5476	361	3025	
9					
10	**(Obs - Exp)^2/Exp**	96.07017544			
11					

Finally, find chi-squared by adding together all the values of $(\text{Obs} - \text{Exp})^2/\text{Exp}$ by entering '=SUM(B10:D10)' into cell b12, say.


Microsoft Excel - Book1

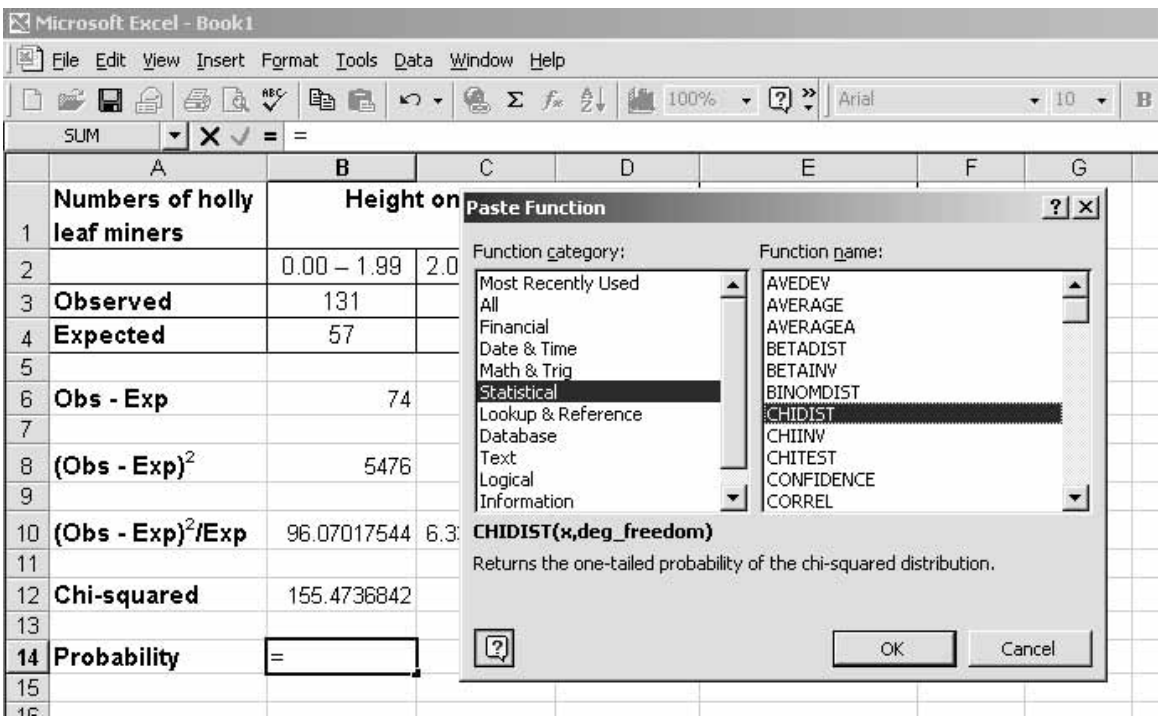
File Edit View Insert Format Tools Data Window Help

100% Arial

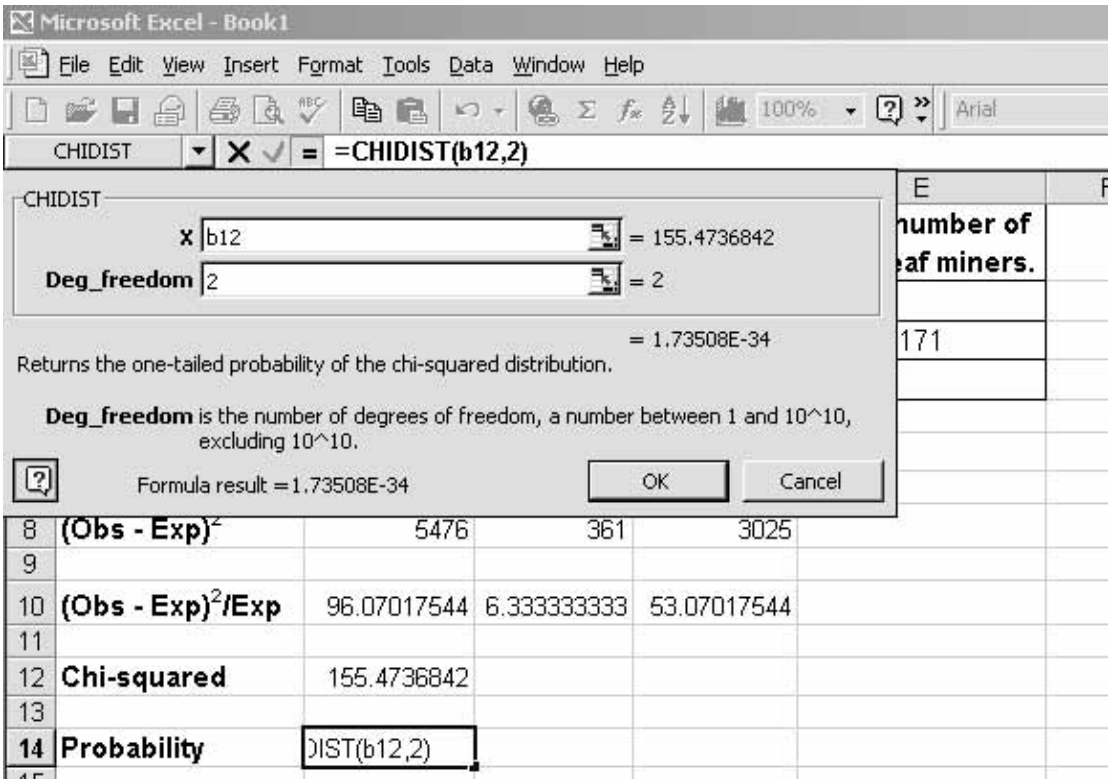
B12 = =SUM(B10:D10)

	A	B	C	D	E
1	**Numbers of holly leaf miners**	**Height on holly tree (m)**			**Total number of holly leaf miners.**
2		0.00 – 1.99	2.00 – 3.99	4.00 – 6.99	
3	**Observed**	131	38	2	171
4	**Expected**	57	57	57	
5					
6	**Obs - Exp**	74	-19	-55	
7					
8	**(Obs - Exp)^2**	5476	361	3025	
9					
10	**(Obs - Exp)^2/Exp**	96.07017544	6.333333333	53.07017544	
11					
12	**Chi-squared**	155.4736842			

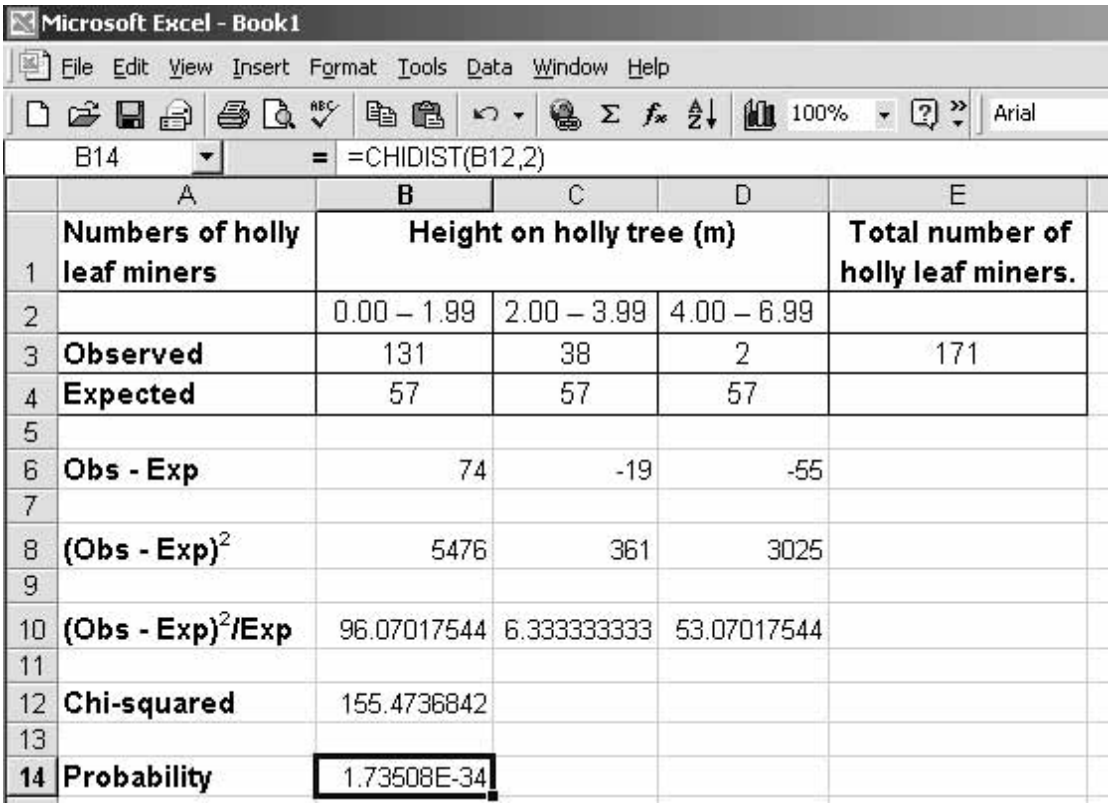
Step 4. Perform the test. We are going to work out the probability that indicates whether the null hypothesis can be rejected or not rejected. To do this, select another cell, b14, say, and start the function wizard ('Insert', 'Function'). From the function category, select 'Statistical', and from the function name list select 'CHIDIST'.



Click on 'OK'. In the following dialogue box, in the first box enter the cell location where the value of chi-squared is held (you can enter the value of chi-squared manually, but entering the cell number eliminates the possibility of typing mistakes). In the second box, enter the number of degrees of freedom (2 in this case).



Now hit 'OK'. The number that appears in the cell is the probability at which the null hypothesis is not rejected, and in this case it is very small (1.73508E-34 means 1.73508×10^{-34}, which is a very small number indeed).


Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help

B14 = =CHIDIST(B12,2)

	A	B	C	D	E
1	**Numbers of holly leaf miners**	**Height on holly tree (m)**			**Total number of holly leaf miners.**
2		0.00 – 1.99	2.00 – 3.99	4.00 – 6.99	
3	**Observed**	131	38	2	171
4	**Expected**	57	57	57	
5					
6	**Obs - Exp**	74	-19	-55	
7					
8	**(Obs - Exp)^2**	5476	361	3025	
9					
10	**(Obs - Exp)^2/Exp**	96.07017544	6.333333333	53.07017544	
11					
12	**Chi-squared**	155.4736842			
13					
14	**Probability**	1.73508E-34			

The probability ($p \approx 0.0$) is below the threshold of $p = 0.05$, therefore we reject the null hypothesis. There is a 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.