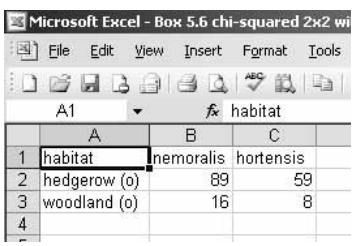
5.4.2. Chi-squared test for association when there is only one degree of freedom

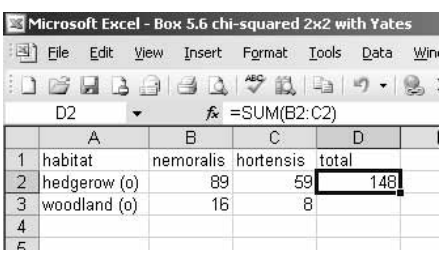
EXAMPLE 5.5. Frequency of *Cepea nemoralis* and *Cepea hortensis* in a hedge

BOX 5.6. How to calculate a 2 × 2 chi-squared test for association

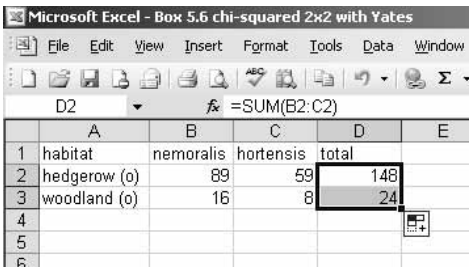
Step 1. Enter the data into the spreadsheet using appropriate row and column headings.



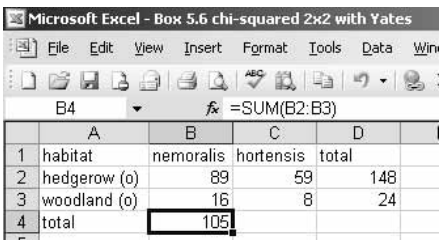
(The '(o)' indicate observed values, to distinguish them from the expected values to be calculated later.)
Calculate the row and column totals.
Create a new column (column d) called 'total'. In cell d2, type '=sum(b2:c2)'. Click on the green tick, or press 'return'.



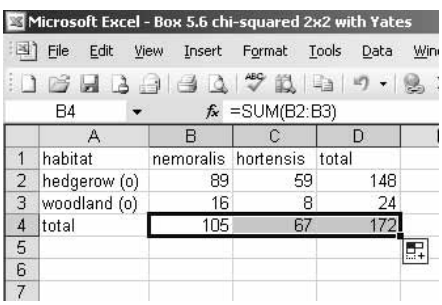
Highlight cell d2, and hover the cursor over the bottom right-hand corner of the cell. It should change from an open horizontal–vertical cross into an addition sign. When this happens, hold down the left mouse button, and move the cursor down into cell d3. Release the mouse button. The formula has been copied, and the total for the woodland snails has been calculated.



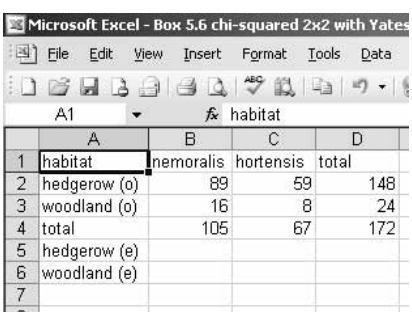
For the column totals, start at cell b4, and type in '=sum(b2:b3)'. Click on the green tick, or press 'return'.



Now drag this across into cells c4 and d4.



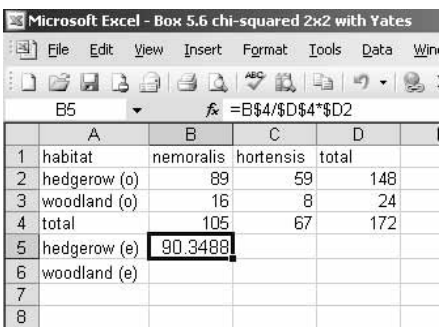
Step 2. Calculate the expected values.
Create two new rows for the expected values.



The simplest way to calculate the expected values is to enter a formula into each cell (for example, '=b4/d4*d2' in cell b5), but the following method is quicker (particularly when there are lots of cells to fill).

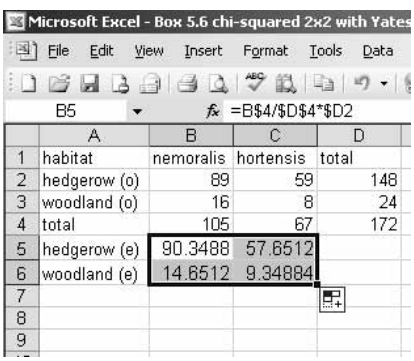
In cell b5, enter the formula '=b$4/$d$4*$d2'. The d4 is an absolute reference that doesn't change wherever we drag the formula. The $d2 is a fixed column but variable row reference: it will always refer to column d, but if the formula is in row 5, it will refer to row2, and if the formula is in row 6, it will refer to row 3. The b$4 is a fixed row but variable column reference: it will always refer to row 5, but the column will be that where the formula is located. This is so that we can drag the formula all the way across the expected array.

Click on the green tick, or press 'return'.



	A	B	C	D
1	habitat	nemoralis	hortensis	total
2	hedgerow (o)	89	59	148
3	woodland (o)	16	8	24
4	total	105	67	172
5	hedgerow (e)	90.3488		
6	woodland (e)			
7				
8				

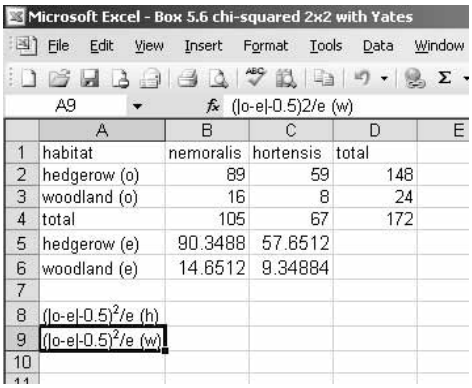
Drag the formula across and down to cell c6 (you may have to first drag across, then drag the row down.)



	A	B	C	D
1	habitat	nemoralis	hortensis	total
2	hedgerow (o)	89	59	148
3	woodland (o)	16	8	24
4	total	105	67	172
5	hedgerow (e)	90.3488	57.6512	
6	woodland (e)	14.6512	9.34884	
7				
8				
9				

Step 3. Perform the 2 × 2 chi-squared test for **association**, and decide what the result means.

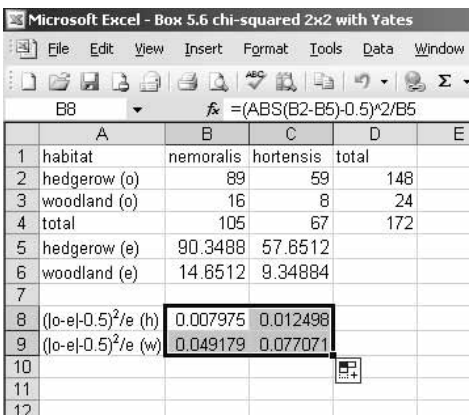
To calculate chi-squared, we need to find $(\text{obs} - \text{exp})^2/\text{exp}$ (including the Yates' correction) for all combinations of habitat and species, then add them all together. Start by creating two rows in which to store the results.



	A	B	C	D	E
1	habitat	nemoralis	hortensis	total	
2	hedgerow (o)	89	59	148	
3	woodland (o)	16	8	24	
4	total	105	67	172	
5	hedgerow (e)	90.3488	57.6512		
6	woodland (e)	14.6512	9.34884		
7					
8	$(\|o-e\|-0.5)^2/e$ (h)				
9	$(\|o-e\|-0.5)^2/e$ (w)				
10					

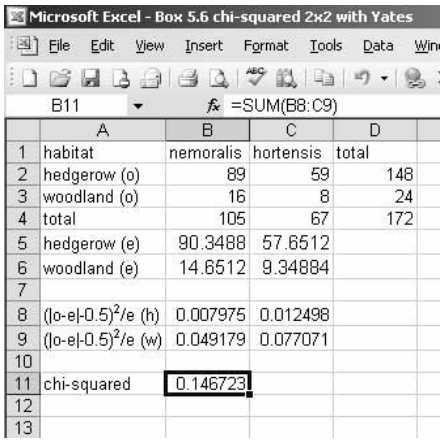
(In this example, the '(h)' and '(w)' denote hedgerow and woodland respectively.)

As above, we are going to use a single formula, and drag it across all cells. In cell b8, type '=(abs(b2 – b5) – 0.5)^2/b5', then drag it into the other cells.



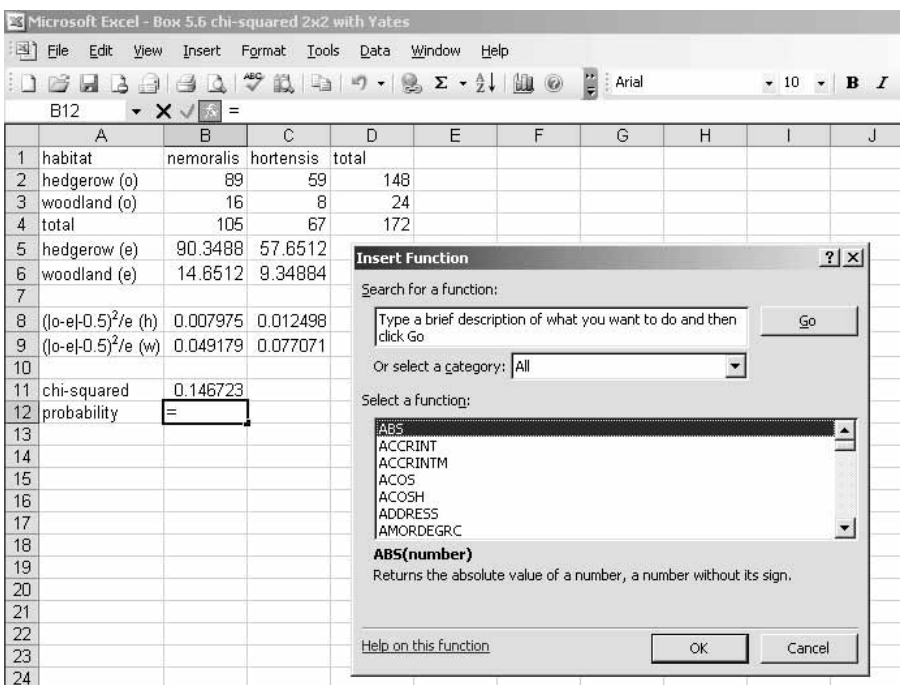
	A	B	C	D	E
1	habitat	nemoralis	hortensis	total	
2	hedgerow (o)	89	59	148	
3	woodland (o)	16	8	24	
4	total	105	67	172	
5	hedgerow (e)	90.3488	57.6512		
6	woodland (e)	14.6512	9.34884		
7					
8	$(\|o-e\|-0.5)^2/e$ (h)	0.007975	0.012498		
9	$(\|o-e\|-0.5)^2/e$ (w)	0.049179	0.077071		
10					
11					

Now add them all together to find chi-squared. In a suitable cell (with a label nearby) type '=sum(b8:c9)', and either click on the green tick or press 'return'.



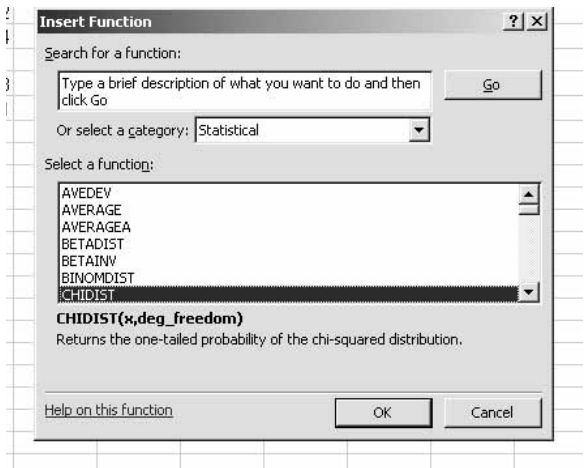
	A	B	C	D
1	habitat	nemoralis	hortensis	total
2	hedgerow (o)	89	59	148
3	woodland (o)	16	8	24
4	total	105	67	172
5	hedgerow (e)	90.3488	57.6512	
6	woodland (e)	14.6512	9.34884	
7				
8	$(\|o-e\|-0.5)^2/e$ (h)	0.007975	0.012498	
9	$(\|o-e\|-0.5)^2/e$ (w)	0.049179	0.077071	
10				
11	chi-squared	0.146723		
12				
13				

Now choose another appropriate cell for the result of the chi-squared test, and label it. Go to 'Insert', 'Function'.

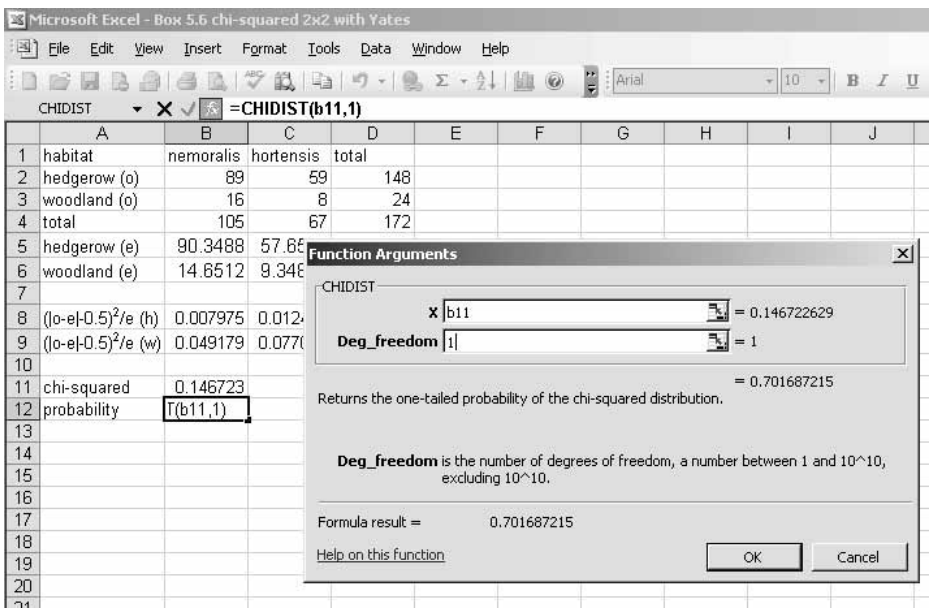


	A	B	C	D	E	F	G	H	I	J
1	habitat	nemoralis	hortensis	total						
2	hedgerow (o)	89	59	148						
3	woodland (o)	16	8	24						
4	total	105	67	172						
5	hedgerow (e)	90.3488	57.6512							
6	woodland (e)	14.6512	9.34884							
7										
8	$(\|o-e\|-0.5)^2/e$ (h)	0.007975	0.012498							
9	$(\|o-e\|-0.5)^2/e$ (w)	0.049179	0.077071							
10										
11	chi-squared	0.146723								
12	probability	=								
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										

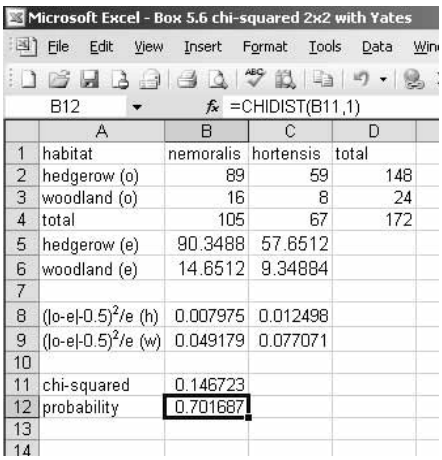
Select the category 'Statistical', and click on 'CHIDIST'.



Click on 'OK'. In the next dialogue box, enter the cell reference where chi-squared is stored (b11 in this case) and the number of degrees of freedom (1 in this case).



Click on 'OK'.



	A	B	C	D
1	habitat	nemoralis	hortensis	total
2	hedgerow (o)	89	59	148
3	woodland (o)	16	8	24
4	total	105	67	172
5	hedgerow (e)	90.3488	57.6512	
6	woodland (e)	14.6512	9.34884	
7				
8	$(\|o-e\|-0.5)^2/e$ (h)	0.007975	0.012498	
9	$(\|o-e\|-0.5)^2/e$ (w)	0.049179	0.077071	
10				
11	chi-squared	0.146723		
12	probability	0.701687		
13				
14				

The probability of accepting the null hypothesis is 0.70, which is very high and above the threshold of $p = 0.05$. Therefore we conclude that there is no significant association ($\chi^2_{\text{calculated}}$ 0.15, $p = 0.70$) between the distribution of snail species and the two habitats.