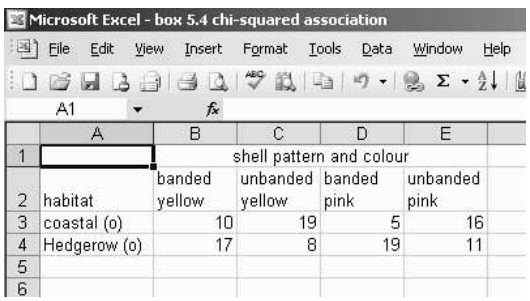
5.3. Chi-squared test for association

EXAMPLE 5.3. Shell colour in *Cepea nemoralis* in coastal and hedgerow habitats

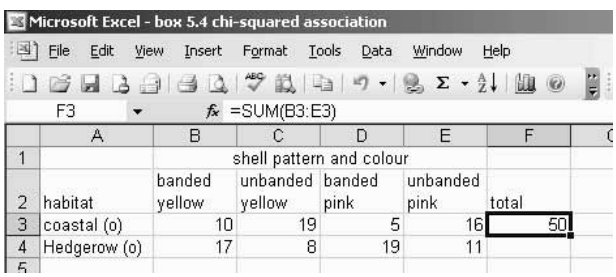
BOX 5.4. How to calculate an r × c chi-squared test for association

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

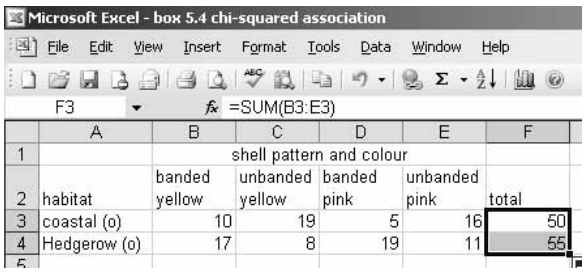


(The '(o)' mean 'observed', to distinguish them from the expected values to be calculated later.)

Calculate the totals for the rows and columns. Start with the rows: in cell f3, type the formula '=sum(b3:e3)', and click on the green tick or press 'return'.



Click on cell f3 to highlight it, then hover the cursor over the bottom-right corner of the cell. It should change from an open horizontal–vertical cross into an addition sign. When it does, hold down the left mouse button, move the cursor down into cellf4, and release the button. The formula will have been copied, and the total for the second row calculated.

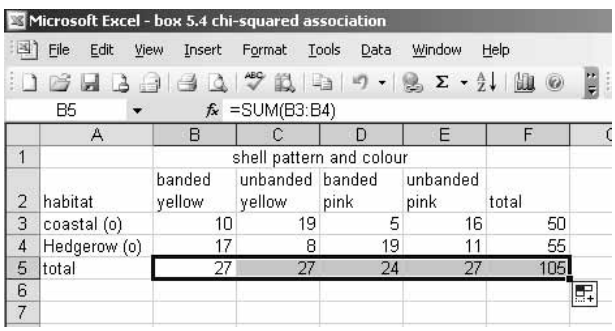

Microsoft Excel - box 5.4 chi-squared association

File Edit View Insert Format Tools Data Window Help

F3 =SUM(B3:E3)

	A	B	C	D	E	F
1		shell pattern and colour				
2	habitat	banded yellow	unbanded yellow	banded pink	unbanded pink	total
3	coastal (o)	10	19	5	16	50
4	Hedgerow (o)	17	8	19	11	55

Now for the columns: start in cell b5, and enter the formula '=sum(b3:b4)'. Click on the green tick or press 'return', then drag this across into cells c5 to f5.


Microsoft Excel - box 5.4 chi-squared association

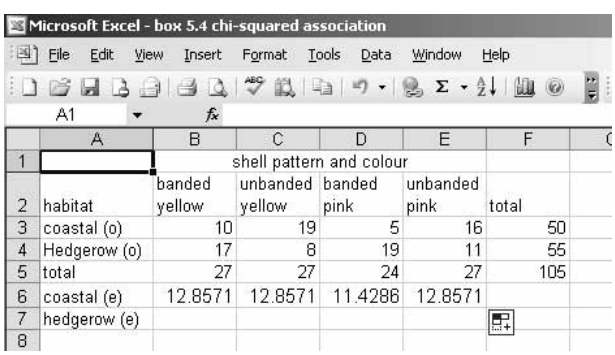
File Edit View Insert Format Tools Data Window Help

B5 =SUM(B3:B4)

	A	B	C	D	E	F
1		shell pattern and colour				
2	habitat	banded yellow	unbanded yellow	banded pink	unbanded pink	total
3	coastal (o)	10	19	5	16	50
4	Hedgerow (o)	17	8	19	11	55
5	total	27	27	24	27	105
6						
7						

Step 2. Calculate the expected values.
Create two new rows, coastal (e) and hedgerow (e), where the '(e)' indicate expected values. (In this example, rows 6 and 7 will be used.)

To create a formula to be dragged along a row, go to cell b6 and enter the formula '=b5/f5*f3'. The 'b5' is a relative cell reference which will change as the formula is dragged across, but the 'f5' and 'f3' are absolute, and remain constant as the formula is dragged. Click on the green tick or press 'return', the drag the formula across into cells c6 to e6.

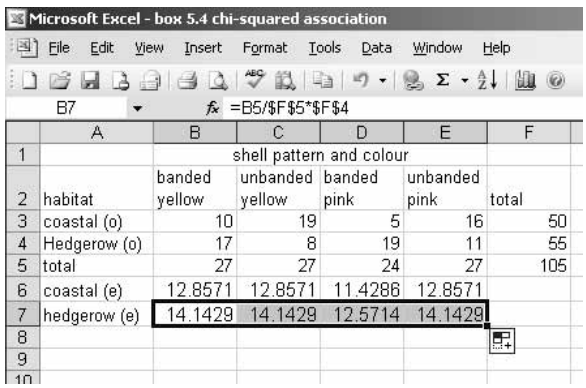

Microsoft Excel - box 5.4 chi-squared association

File Edit View Insert Format Tools Data Window Help

A1

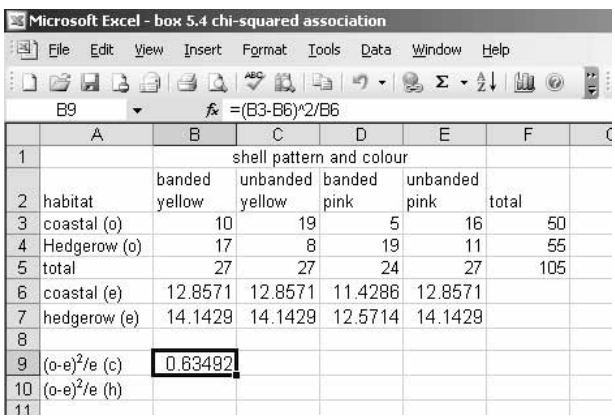
	A	B	C	D	E	F
1		shell pattern and colour				
2	habitat	banded yellow	unbanded yellow	banded pink	unbanded pink	total
3	coastal (o)	10	19	5	16	50
4	Hedgerow (o)	17	8	19	11	55
5	total	27	27	24	27	105
6	coastal (e)	12.8571	12.8571	11.4286	12.8571	
7	hedgerow (e)					
8						

Repeat for the hedgerow snails: use the formula '=b5/f5*f4' in cell b7, then drag across into cells c7 to e7.



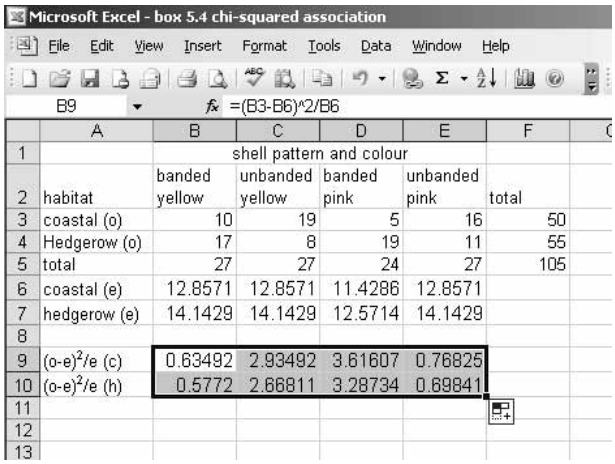
	A	B	C	D	E	F
1		shell pattern and colour				
2	habitat	banded yellow	unbanded yellow	banded pink	unbanded pink	total
3	coastal (o)	10	19	5	16	50
4	Hedgerow (o)	17	8	19	11	55
5	total	27	27	24	27	105
6	coastal (e)	12.8571	12.8571	11.4286	12.8571	
7	hedgerow (e)	14.1429	14.1429	12.5714	14.1429	
8						
9						
10						

Step 3. Calculate chi-squared test for association. First, calculate the values of $(\text{obs} - \text{exp})^2/\text{exp}$ for all possible combinations of habitat and pattern/colour. Use rows 9 and 10, and enter the formula '=(b3-b6)^2/b6' into cell b9. Click on the green tick, or press 'return'.



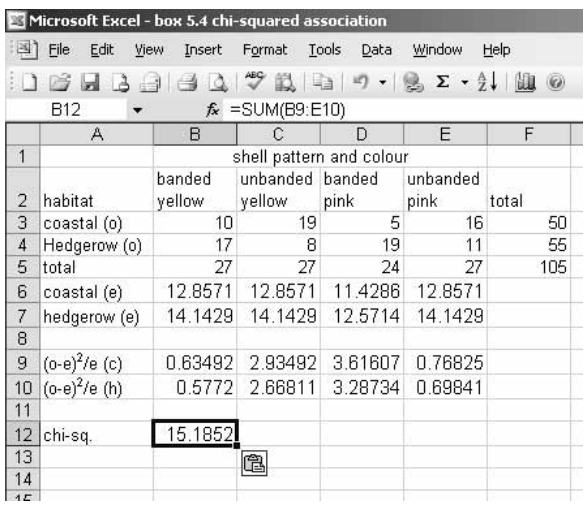
	A	B	C	D	E	F
1		shell pattern and colour				
2	habitat	banded yellow	unbanded yellow	banded pink	unbanded pink	total
3	coastal (o)	10	19	5	16	50
4	Hedgerow (o)	17	8	19	11	55
5	total	27	27	24	27	105
6	coastal (e)	12.8571	12.8571	11.4286	12.8571	
7	hedgerow (e)	14.1429	14.1429	12.5714	14.1429	
8						
9	$(o-e)^2/e$ (c)	0.63492				
10	$(o-e)^2/e$ (h)					
11						

Because the observed and expected values are in identically arranged arrays, this formula can be dragged across and down to cell e10 to calculate all the other values of $(\text{obs} - \text{exp})^2/\text{exp}$. (You will have to do the drag across first, then drag the whole row down.)



	A	B	C	D	E	F
1		shell pattern and colour				
2	habitat	banded yellow	unbanded yellow	banded pink	unbanded pink	total
3	coastal (o)	10	19	5	16	50
4	Hedgerow (o)	17	8	19	11	55
5	total	27	27	24	27	105
6	coastal (e)	12.8571	12.8571	11.4286	12.8571	
7	hedgerow (e)	14.1429	14.1429	12.5714	14.1429	
8						
9	$(o-e)^2/e$ (c)	0.63492	2.93492	3.61607	0.76825	
10	$(o-e)^2/e$ (h)	0.5772	2.66811	3.28734	0.69841	
11						
12						
13						

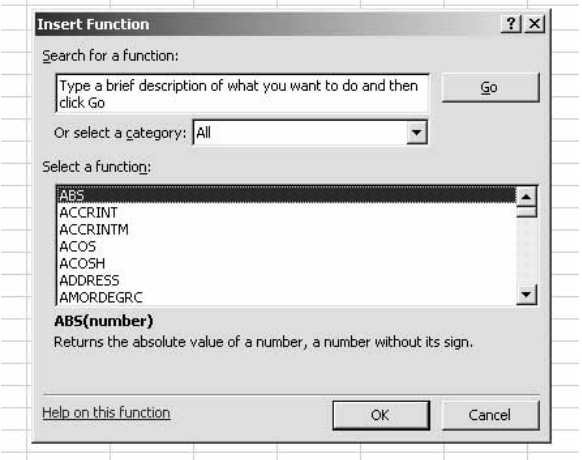
Next, add them all up to find the value of chi-squared. They are in a nice rectangular array, so a single 'sum' function will do the job. We shall use cell b12, and the formula '=sum(b9:e10)'.



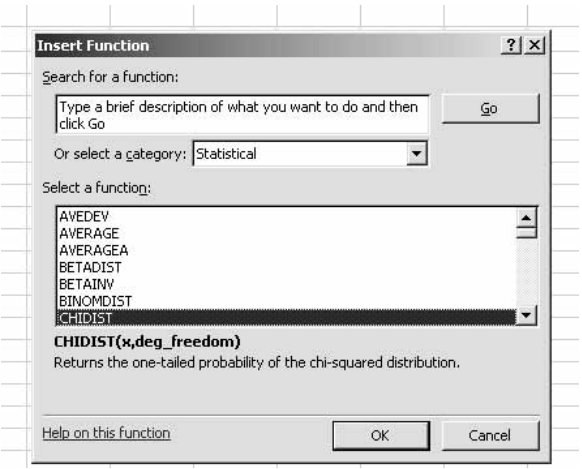
	A	B	C	D	E	F
1		shell pattern and colour				
2	habitat	banded yellow	unbanded yellow	banded pink	unbanded pink	total
3	coastal (o)	10	19	5	16	50
4	Hedgerow (o)	17	8	19	11	55
5	total	27	27	24	27	105
6	coastal (e)	12.8571	12.8571	11.4286	12.8571	
7	hedgerow (e)	14.1429	14.1429	12.5714	14.1429	
8						
9	$(o-e)^2/e$ (c)	0.63492	2.93492	3.61607	0.76825	
10	$(o-e)^2/e$ (h)	0.5772	2.66811	3.28734	0.69841	
11						
12	chi-sq.	15.1852				
13						
14						

Step 4. Perform the chi-squared test for association, and decide what the result means.

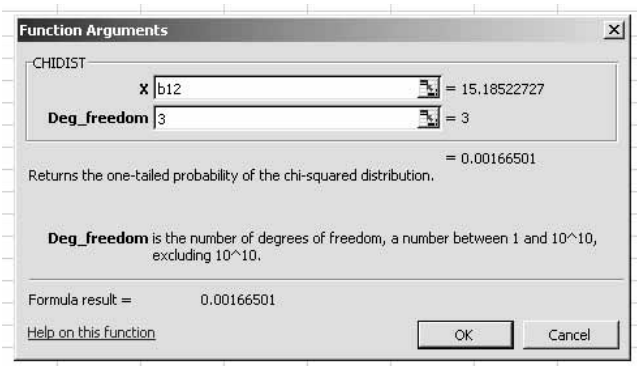
Select a suitable cell (we shall use b13), got to 'Insert' and select 'Function'.



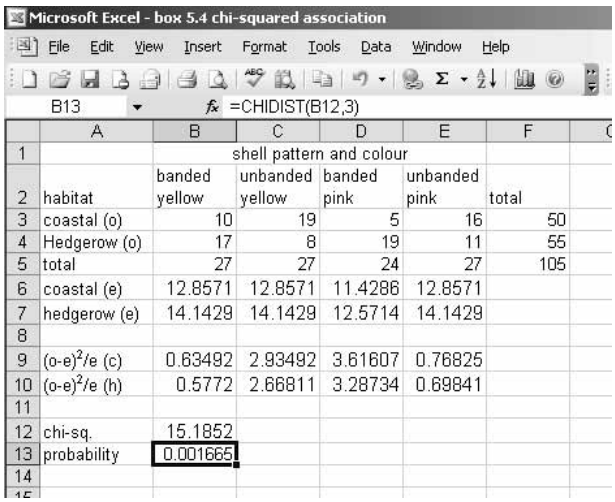
Go to the 'Statistical' category, and select 'CHIDIST'.



Press 'OK', then enter the cell where the value of chi-squared is stored (b12 in this case) and the number of **degrees of freedom** (3 in this case).



Click on 'OK'.



	A	B	C	D	E	F
1		shell pattern and colour				
2	habitat	banded yellow	unbanded yellow	banded pink	unbanded pink	total
3	coastal (o)	10	19	5	16	50
4	Hedgerow (o)	17	8	19	11	55
5	total	27	27	24	27	105
6	coastal (e)	12.8571	12.8571	11.4286	12.8571	
7	hedgerow (e)	14.1429	14.1429	12.5714	14.1429	
8						
9	$(o-e)^2/e$ (c)	0.63492	2.93492	3.61607	0.76825	
10	$(o-e)^2/e$ (h)	0.5772	2.66811	3.28734	0.69841	
11						
12	chi-sq.	15.1852				
13	probability	0.001665				
14						

In this example p is very low (0.0017) and below the threshold of $p = 0.05$, so we reject the null hypothesis. There is a highly significant association ($\chi^2_{\text{calculated}} = 15.19$, $p = 0.0017$) between the distribution of shell patterns and habitat (coastal and hedgerow) of *Cepea nemoralis*.