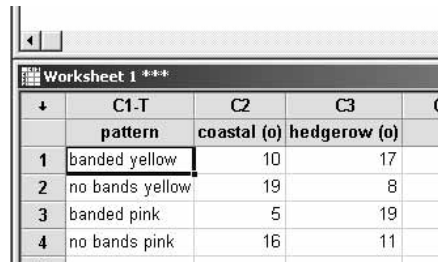


### 5.3. Chi-squared test for association

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

**BOX 5.4.** How to calculate an  $r \times c$  chi-squared test for association

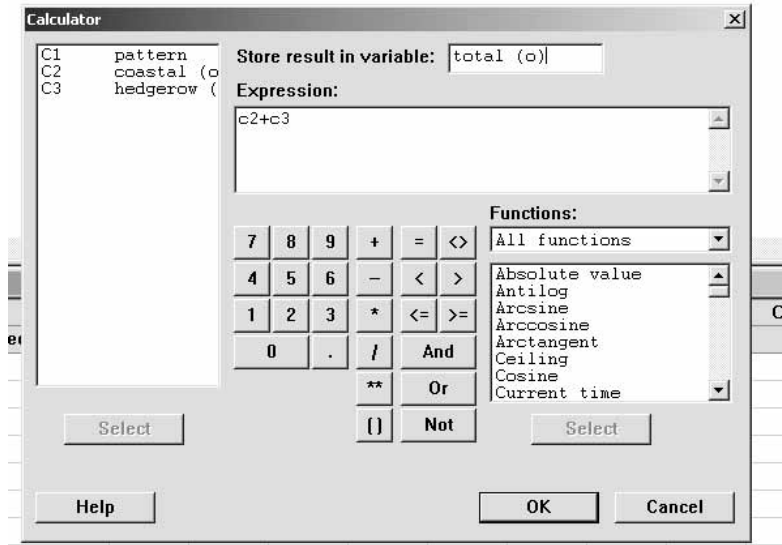
**Step 1.** Enter the data into the worksheet section of the Minitab screen. The letters 'o' in brackets at the end of the column names indicate observed values.



The screenshot shows a Minitab worksheet titled 'Worksheet 1 \*\*\*'. The data is organized into columns: C1-T (pattern), C2 (coastal (o)), C3 (hedgerow (o)), and C4 (total (o)). The rows represent different shell color patterns.

	C1-T pattern	C2 coastal (o)	C3 hedgerow (o)	C4 total (o)
1	banded yellow	10	17	
2	no bands yellow	19	8	
3	banded pink	5	19	
4	no bands pink	16	11	

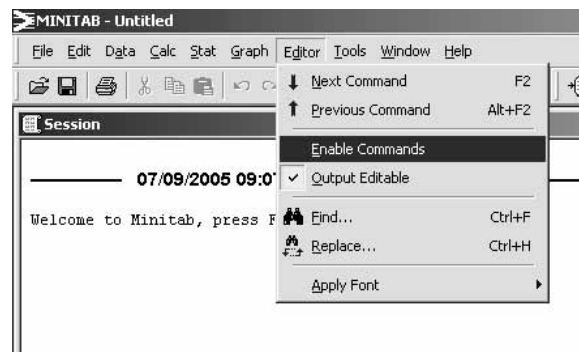
**Step 2.** Calculate the totals for the rows. Go to 'Calc', 'Calculator'; type 'c2 + c3' in the expression box, and type 'total (o)' in the 'Store result in variable' box.



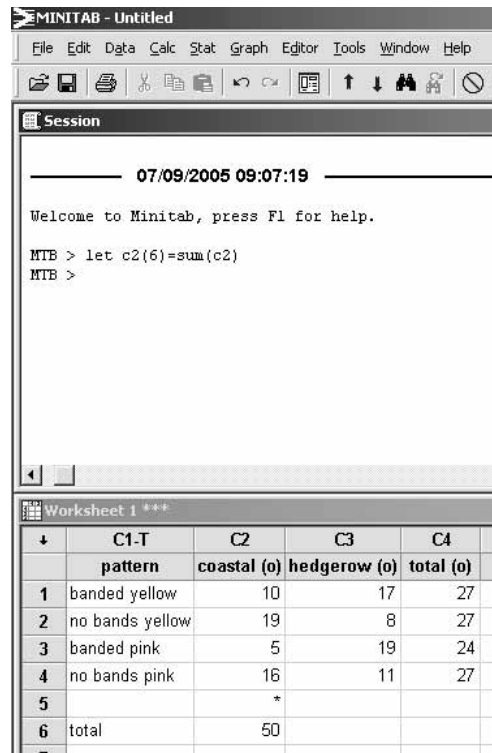
Now click on 'OK'.

	C1-T	C2	C3	C4
	pattern	coastal (o)	hedgerow (o)	total (o)
1	banded yellow	10	17	27
2	no bands yellow	19	8	27
3	banded pink	5	19	24
4	no bands pink	16	11	27
5				

**Step 3.** Calculate the totals for the columns. In c1(6), enter 'total' (leave row 5 blank to avoid confusion between data and totals). Click in the 'Session' (top) window, go to 'Editor' and select 'Enable Commands'.



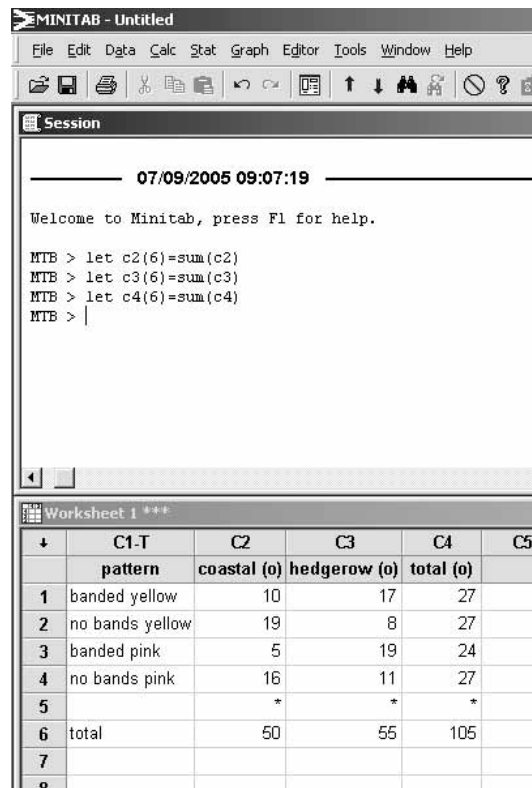
At the 'MTB >' prompt, type 'let c2(6)=sum(c2)'. This will add all the numbers in column 2, and place the result in cell 6 in column 2.



The screenshot shows the Minitab software interface. The top window is the 'Session' window, which displays the date and time '07/09/2005 09:07:19', a welcome message, and the command 'let c2(6)=sum(c2)' entered at the 'MTB >' prompt. The bottom window is the 'Worksheet 1 \*\*\*' window, which contains a data table with 7 rows and 5 columns. The columns are labeled 'C1-T', 'C2', 'C3', and 'C4'. The rows contain data for different patterns and their totals.

	C1-T	C2	C3	C4
	pattern	coastal (o)	hedgerow (o)	total (o)
1	banded yellow	10	17	27
2	no bands yellow	19	8	27
3	banded pink	5	19	24
4	no bands pink	16	11	27
5		*		
6	total	50		
7				

Repeat the process for columns 3 and 4.



The screenshot shows the Minitab software interface. The 'Session' window displays the following text:

```

07/09/2005 09:07:19

Welcome to Minitab, press F1 for help.

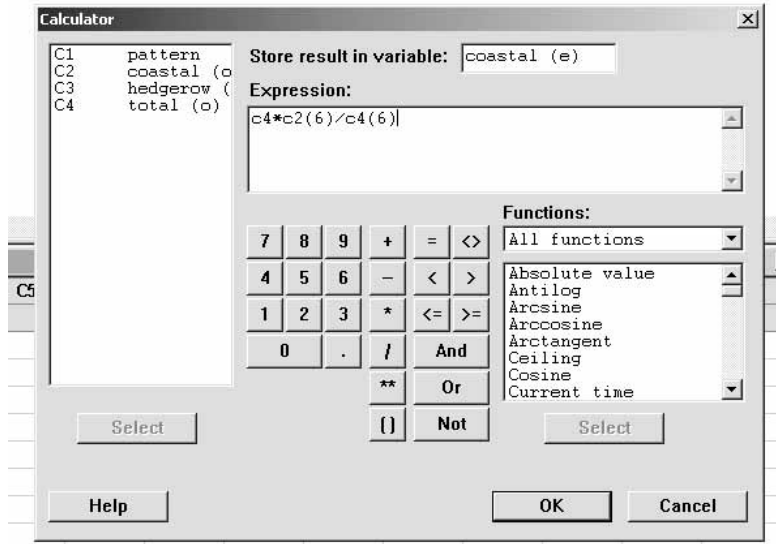
MTB > let c2(6)=sum(c2)
MTB > let c3(6)=sum(c3)
MTB > let c4(6)=sum(c4)
MTB > |

```

The 'Worksheet 1 \*\*\*' window shows a data table with the following structure:

	C1-T	C2	C3	C4	C5
	pattern	coastal (o)	hedgerow (o)	total (o)	
1	banded yellow	10	17	27	
2	no bands yellow	19	8	27	
3	banded pink	5	19	24	
4	no bands pink	16	11	27	
5		*	*	*	
6	total	50	55	105	
7					
8					

**Step 4.** Calculate the expected values. Go to 'Calc', 'Calculator' and enter 'coastal (e)' into the 'Store results in variable' box. Enter ' $c4 * c2(6) / c4(6)$ ' into the 'Expression' box.



Now click on 'OK'.

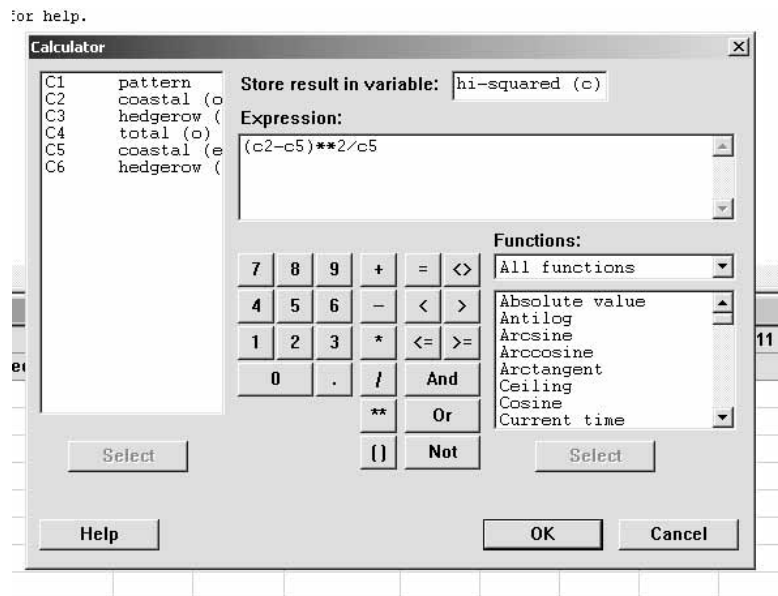
	C1-T	C2	C3	C4	C5
	pattern	coastal (o)	hedgerow (o)	total (o)	coastal (e)
1	banded yellow	10	17	27	12.8571
2	no bands yellow	19	8	27	12.8571
3	banded pink	5	19	24	11.4286
4	no bands pink	16	11	27	12.8571
5		*	*	*	*
6	total	50	55	105	50.0000
7					

Repeat the process for the expected values for the hedgerow snails, using the column heading 'hedgerow (e)' and the expression 'c4\*c3(6)/c4(6)'.

	C1-T	C2	C3	C4	C5	C6	C7
	pattern	coastal (o)	hedgerow (o)	total (o)	coastal (e)	hedgerow (e)	
1	banded yellow	10	17	27	12.8571	14.1429	
2	no bands yellow	19	8	27	12.8571	14.1429	
3	banded pink	5	19	24	11.4286	12.5714	
4	no bands pink	16	11	27	12.8571	14.1429	
5		*	*	*	*	*	
6	total	50	55	105	50.0000	55.0000	
7							

**Step 5.** Calculate the test statistic for this  $r \times c$  chi-squared test for association. This is the sum of the terms such as  $(\text{observed} - \text{expected})^2 / (\text{expected})$ , and in this case we have eight of them. The easiest way to do this is to calculate the eight parts separately, in two columns, and then add them all up.

Go to 'Calc', 'Calculator' and enter 'chi-squared (c)' (c for coastal) into the 'Store results in variable' box, and type  $(c2 - c5)**2/c5$  into the 'Expression' window.



Now click on 'OK'.

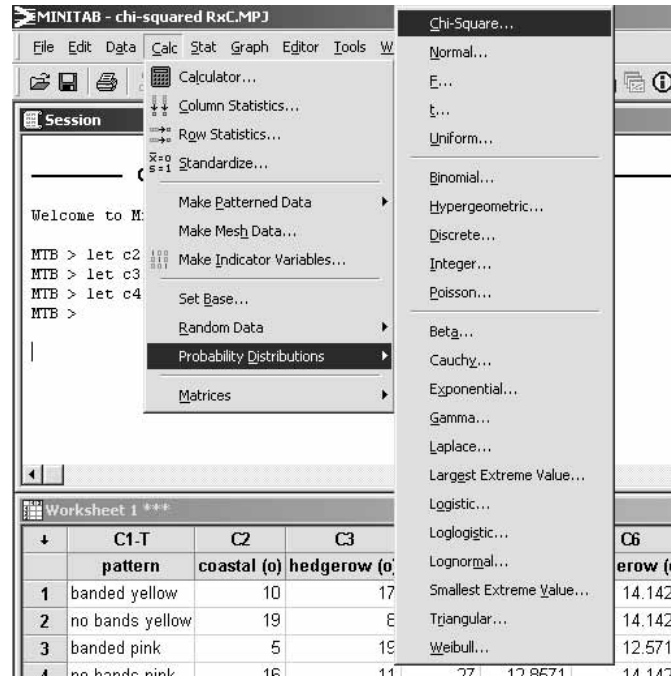
	C1-T	C2	C3	C4	C5	C6	C7
	pattern	coastal (o)	hedgerow (o)	total (o)	coastal (e)	hedgerow (e)	chi-squared (c)
1	banded yellow	10	17	27	12.8571	14.1429	0.63492
2	no bands yellow	19	8	27	12.8571	14.1429	2.93492
3	banded pink	5	19	24	11.4286	12.5714	3.61607
4	no bands pink	16	11	27	12.8571	14.1429	0.76825
5		*	*	*	*	*	*
6	total	50	55	105	50.0000	55.0000	0.00000
7							
8							

Repeat the process for the hedgerow data, using the column heading 'chi-squared (h)' and the expression  $(c3 - c6)**2/c6$ .

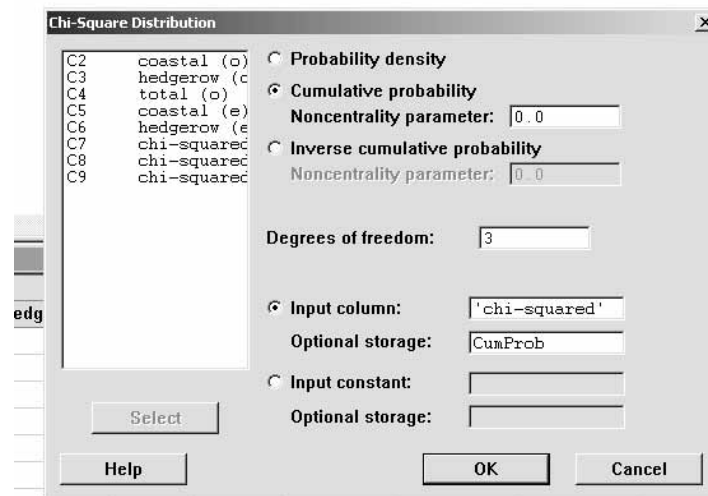


**Step 6.** Perform the  $r \times c$  chi-squared test for association, and decide what the result means.

Go to 'Calc', 'Probability Distributions', 'Chi-squared'.



Select 'Cumulative Probability', and enter '3' (for this example) in the 'degrees of freedom' window. For the 'Input Column', enter 'Chi-squared', and for 'Optional Storage' enter 'CumProb'.

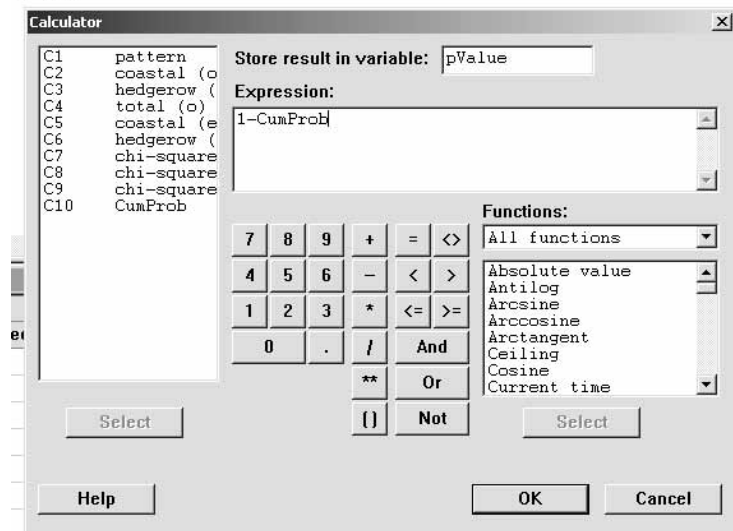


Click on 'OK'.

	C1-T	C2	C3	C4	C5	C6	C7	C8	C9	C10	C
	pattern	coastal (o)	hedgerow (o)	total (o)	coastal (e)	hedgerow (e)	chi-squared (c)	chi-squared (h)	chi-squared	CumProb	
1	banded yellow	10	17	27	12.8571	14.1429	0.63492	0.57720	15.1852	0.998335	
2	no bands yellow	19	8	27	12.8571	14.1429	2.93492	2.66811			
3	banded pink	5	19	24	11.4286	12.5714	3.61607	3.28734			
4	no bands pink	16	11	27	12.8571	14.1429	0.76825	0.69841			
5		*	*	*	*	*	*	*			
6	total	50	55	105	50.0000	55.0000	0.00000	0.00000			
7											

The final step is to find the probability that the null hypothesis (there is no association between the distribution of shell patterns observed and the habitat (coastal and hedgerow) of *Cepea nemoralis*) can be rejected. Go to 'Calc', 'Calculator'. Enter 'pValue' in the 'Store results in variable' window, and type '1-CumProb' in the 'Expression' window.

for help.



Click on 'OK'.

	C1-T	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
	pattern	coastal (o)	hedgerow (o)	total (o)	coastal (e)	hedgerow (e)	chi-squared (c)	chi-squared (h)	chi-squared	CumProb	pValue
1	banded yellow	10	17	27	12.8571	14.1429	0.63492	0.57720	15.1852	0.998335	0.0016650
2	no bands yellow	19	8	27	12.8571	14.1429	2.93492	2.66811			
3	banded pink	5	19	24	11.4286	12.5714	3.61607	3.28734			
4	no bands pink	16	11	27	12.8571	14.1429	0.76825	0.69841			
5		*	*	*	*	*	*	*			
6	total	50	55	105	50.0000	55.0000	0.00000	0.00000			

In this example the  $p$  value ( $p = 0.002$ ) is below the threshold level of  $p = 0.05$ . Therefore the null hypothesis is rejected. There is a highly significant association ( $\chi^2_{\text{calculated}} = 15.18, p = 0.002$ ) between the distribution of shell patterns and habitat (coastal and hedgerow) of *Cepea nemoralis*.