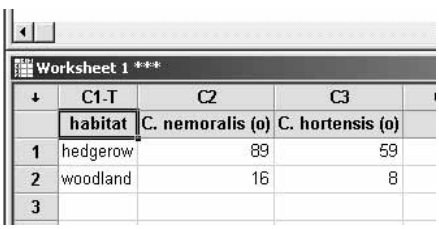
5.4.2. Chi-squared test for association when there is only one degree of freedom

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

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

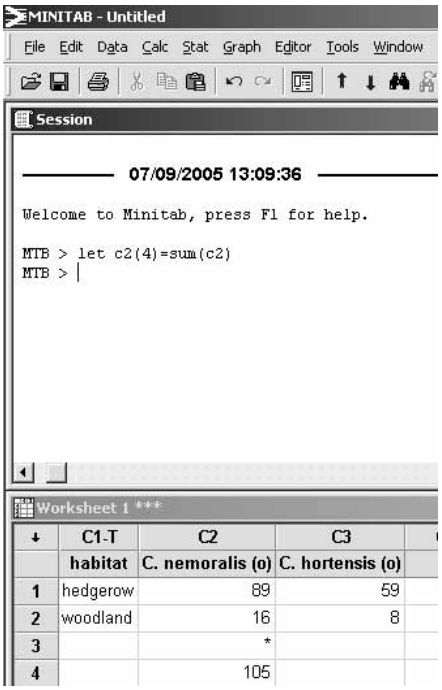
Step 1. Enter the data into the Worksheet window of Minitab.



	C1-T	C2	C3
	habitat	C. nemoralis (o)	C. hortensis (o)
1	hedgerow	89	59
2	woodland	16	8
3			

Step 2. Calculate the totals for the columns and rows.
Click in the session window, and go to 'Editor'. Select 'Enable Commands'.

At the 'MTB >' prompt, type 'let c2(4) = sum(c2)' and press the 'return' key. This will add all the entries in column 3, and place the result in cell 4 of that column.

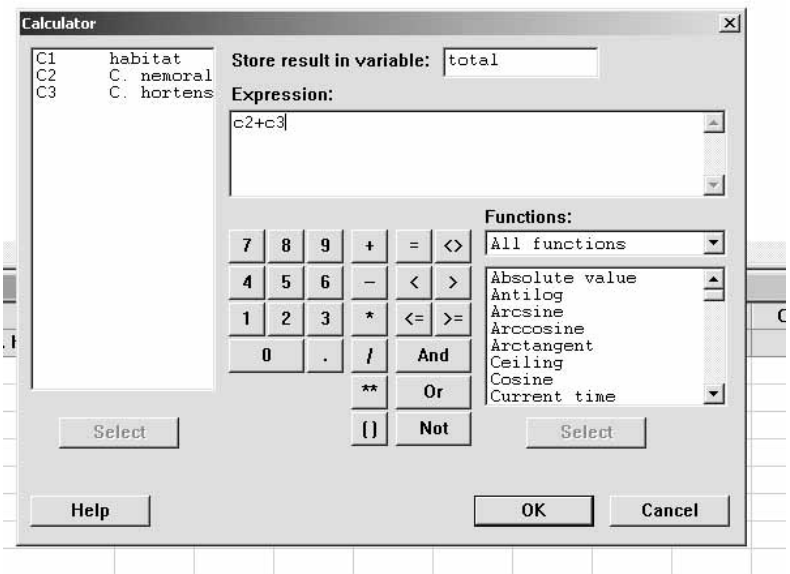


	C1-T	C2	C3
	habitat	C. nemoralis (o)	C. hortensis (o)
1	hedgerow	89	59
2	woodland	16	8
3		*	
4		105	

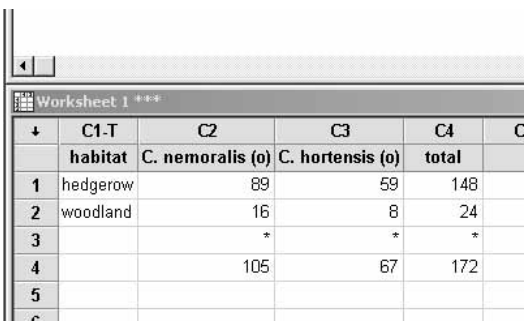
Repeat for column 3.

	C1-T	C2	C3
	habitat	C. nemoralis (o)	C. hortensis (o)
1	hedgerow	89	59
2	woodland	16	8
3		*	*
4		105	67

Now add the rows. Go to 'Calc', 'Calculator', put 'total' in the 'Store result as variable' window, and type 'c2 + c3' in the 'Expression' window.

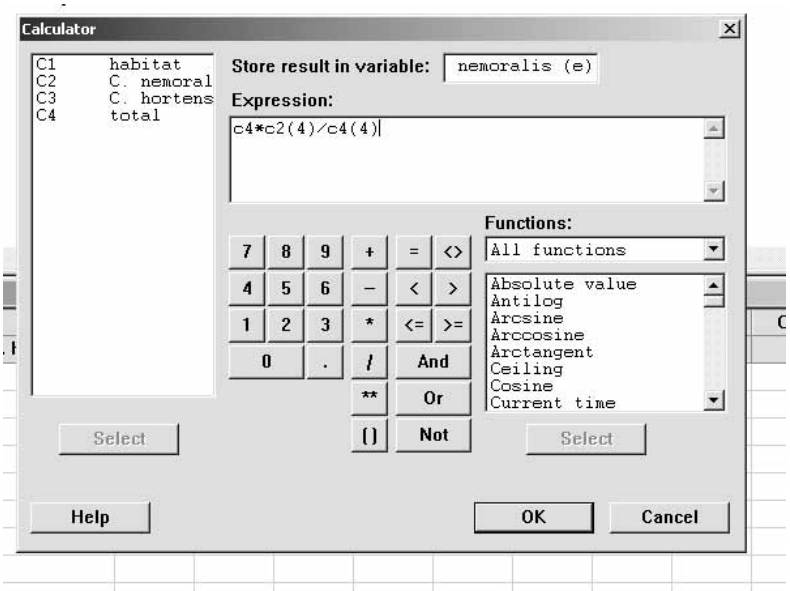


Click on 'OK'.

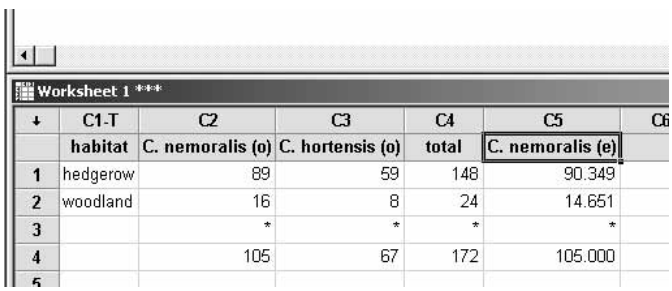


	C1-T	C2	C3	C4	C
	habitat	C. nemoralis (o)	C. hortensis (o)	total	
1	hedgerow	89	59	148	
2	woodland	16	8	24	
3		*	*	*	
4		105	67	172	
5					

Step 3. Calculate the **expected** values. Go to 'Calc', Calculator'. Enter '*C. nemoralis* (e)' in the 'Store result as variable' window; and type 'c4*c2(4)/c4(4)' in the 'Expression' window.



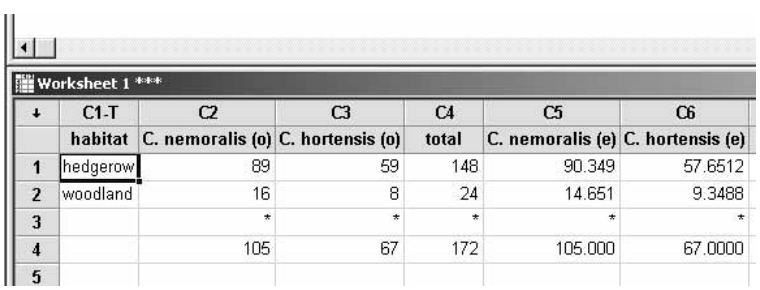
Click on 'OK'.



Worksheet 1 ***

	C1-T	C2	C3	C4	C5	C6
	habitat	C. nemoralis (o)	C. hortensis (o)	total	C. nemoralis (e)	
1	hedgerow	89	59	148	90.349	
2	woodland	16	8	24	14.651	
3		*	*	*	*	
4		105	67	172	105.000	
5						

Repeat the process for *C. hortensis*, remembering that the source data here are in column 3 instead of column 2.



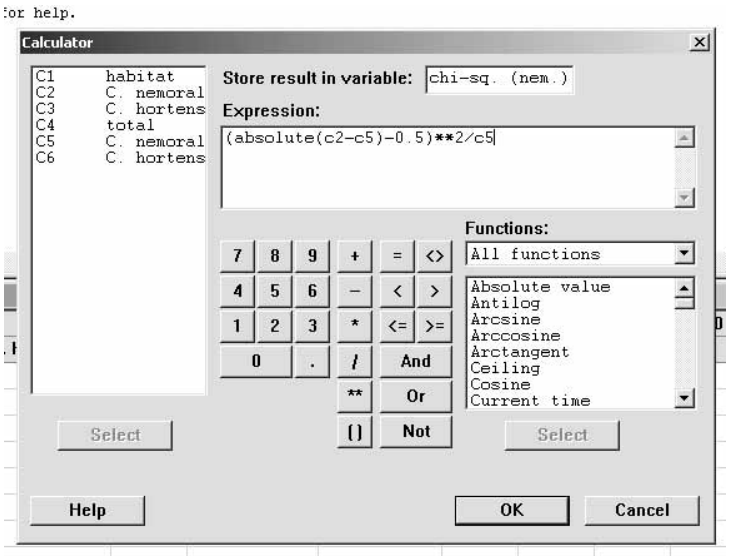
Worksheet 1 ***

	C1-T	C2	C3	C4	C5	C6
	habitat	C. nemoralis (o)	C. hortensis (o)	total	C. nemoralis (e)	C. hortensis (e)
1	hedgerow	89	59	148	90.349	57.6512
2	woodland	16	8	24	14.651	9.3488
3		*	*	*	*	*
4		105	67	172	105.000	67.0000
5						

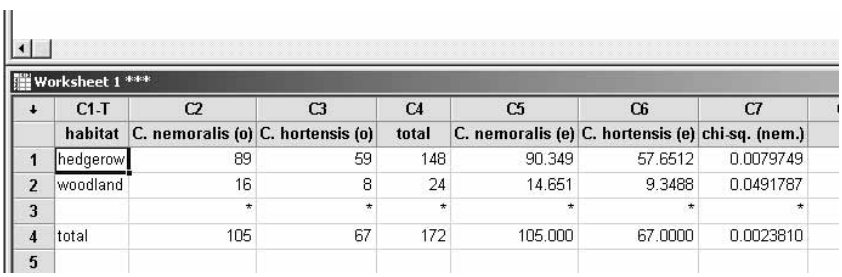
Step 4. Calculate the individual values of chi-squared. Because we have a 2×2 table, we will only have one degree of freedom, so we need to incorporate Yates' correction. This involves the use of the 'Absolute'

operator, which calculates the absolute value of a number (the value ignoring any minus signs).

Go to 'Calc', 'Calculator', and enter 'chi-sq. (nem.)' in the 'Store results as variable' window. Type '(absolute(c2–c5)–0.5)**2/c5' in the expression window.



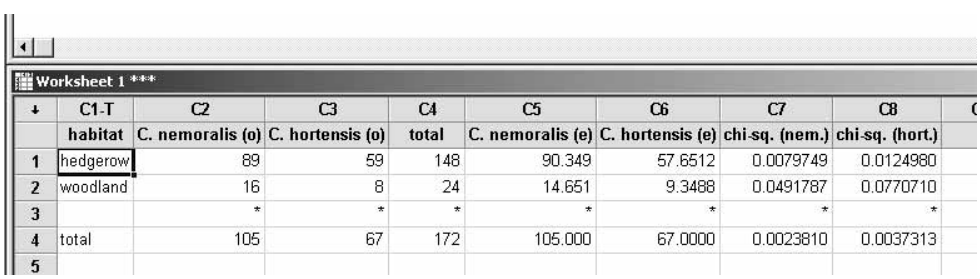
Click on 'OK'.



Worksheet 1 ***

	C1-T	C2	C3	C4	C5	C6	C7
	habitat	C. nemoralis (o)	C. hortensis (o)	total	C. nemoralis (e)	C. hortensis (e)	chi-sq. (nem.)
1	hedgerow	89	59	148	90.349	57.6512	0.0079749
2	woodland	16	8	24	14.651	9.3488	0.0491787
3		*	*	*	*	*	*
4	total	105	67	172	105.000	67.0000	0.0023810
5							

Repeat for *C. hortensis*, remembering that the observed and expected values are now in columns 3 and 6 respectively.



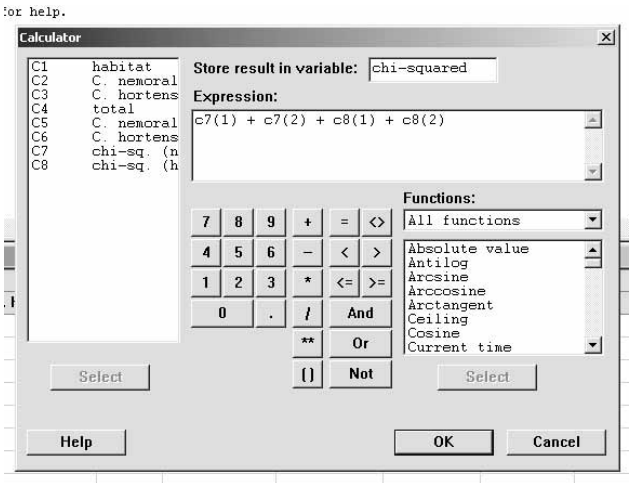
Worksheet 1 ***

	C1-T	C2	C3	C4	C5	C6	C7	C8
	habitat	C. nemoralis (o)	C. hortensis (o)	total	C. nemoralis (e)	C. hortensis (e)	chi-sq. (nem.)	chi-sq. (hort.)
1	hedgerow	89	59	148	90.349	57.6512	0.0079749	0.0124980
2	woodland	16	8	24	14.651	9.3488	0.0491787	0.0770710
3		*	*	*	*	*	*	*
4	total	105	67	172	105.000	67.0000	0.0023810	0.0037313
5								

(The values in row 4 of columns 7 and 8 have no meaning.)

Step 5. Calculate the overall value of chi-squared, and work out what it means in this context.

The overall value of chi-squared is the sum of the four individual values of chi-squared. Go to 'Calc', 'Calculator', and enter 'chi-squared' in the 'Store result as variable' window. In the 'Expression' window, enter 'c7(1) + c7(2) + c8(1) + c8(2)'.

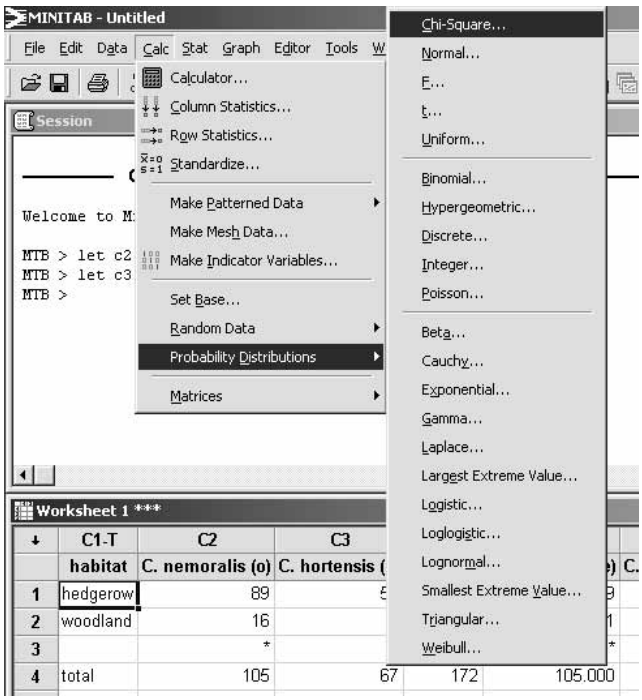


Click on 'OK'.

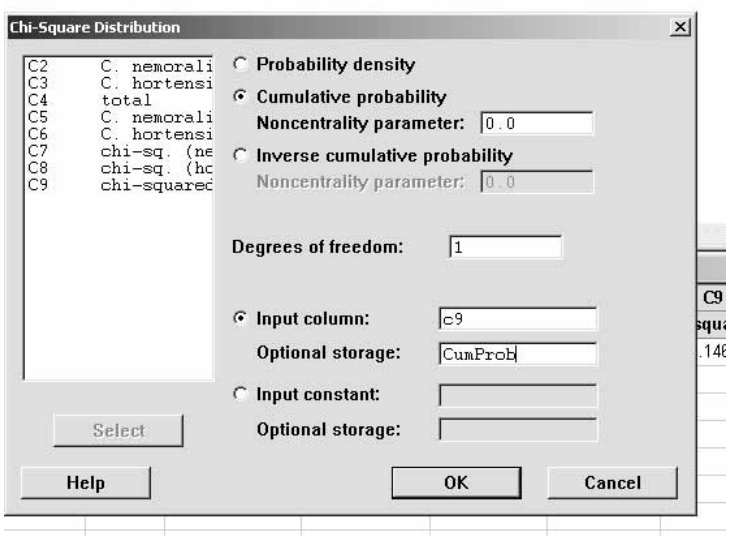
Worksheet 1 ***

	C1-T	C2	C3	C4	C5	C6	C7	C8	C9
	habitat	C. nemoralis (o)	C. hortensis (o)	total	C. nemoralis (e)	C. hortensis (e)	chi-sq. (nem.)	chi-sq. (hort.)	chi-squared
1	hedgerow	89	59	148	90.349	57.6512	0.0079749	0.0124980	0.146723
2	woodland	16	8	24	14.651	9.3488	0.0491787	0.0770710	
3		*	*	*	*	*	*	*	
4	total	105	67	172	105.000	67.0000	0.0023810	0.0037313	
5									

Go to 'Calc', 'Probability Distributions' and select 'Chi-square'.



Select 'Cumulative Probability', enter '1' in the **'degrees of freedom'** window; put 'c9' (or 'chi-squared') in the 'Input Column' window, and 'CumProb' in the 'Optional storage'.

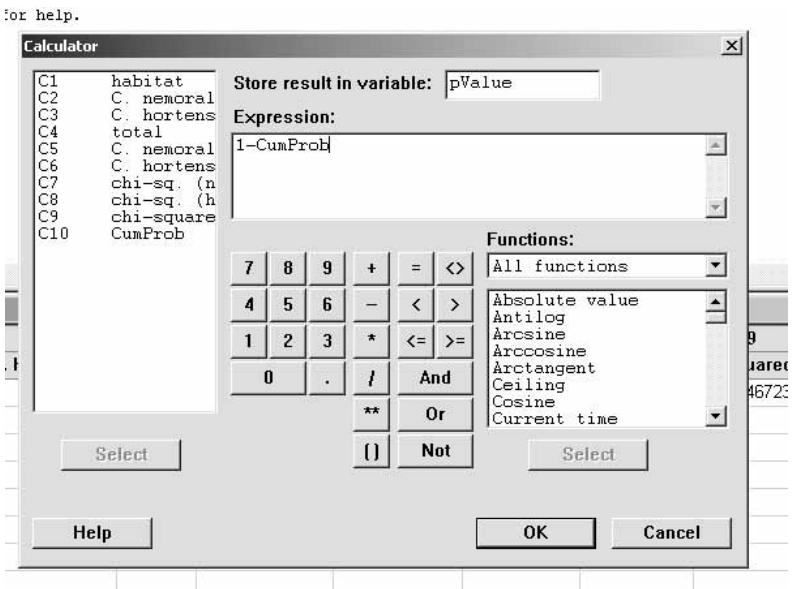


Click on 'OK'.

Worksheet 1 ***

↓	C1-T	C2	C3	C4	C5	C6	C7	C8	C9	C10
	habitat	C. nemoralis (o)	C. hortensis (o)	total	C. nemoralis (e)	C. hortensis (e)	chi-sq. (nem.)	chi-sq. (hort.)	chi-squared	CumProb
1	hedgerow	89	59	148	90.349	57.6512	0.0079749	0.0124980	0.146723	0.298313
2	woodland	16	8	24	14.651	9.3488	0.0491787	0.0770710		
3		*	*	*	*	*	*	*		
4	total	105	67	172	105.000	67.0000	0.0023810	0.0037313		
5										

Next, Calculate the p value. Go to 'Calc', 'Calculator', and enter 'pValue' in the 'Store result as variable' window. Type '1-CumProb' in the 'Expression' window.



Click on 'OK'.

Worksheet 1 ***

↓	C1-T	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
	habitat	C. nemoralis (o)	C. hortensis (o)	total	C. nemoralis (e)	C. hortensis (e)	chi-sq. (nem.)	chi-sq. (hort.)	chi-squared	CumProb	pValue
1	hedgerow	89	59	148	90.349	57.6512	0.0079749	0.0124980	0.146723	0.298313	0.701687
2	woodland	16	8	24	14.651	9.3488	0.0491787	0.0770710			
3		*	*	*	*	*	*	*			
4	total	105	67	172	105.000	67.0000	0.0023810	0.0037313			

This p value ($p = 0.7$) is large and exceeds the threshold of $p = 0.05$. We do not reject the null hypothesis. There is no significant association ($\chi^2_{\text{calculated}} = 0.15$, $p = 0.7$) between the distribution of snail species and the two habitats.