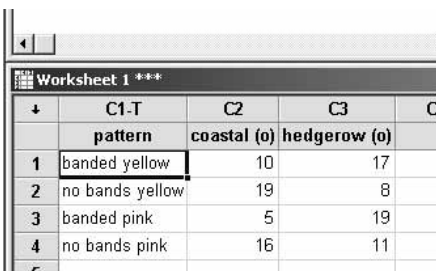
## 5.5.2. An r × c G test for association

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

**BOX 5.8. How to calculate an r × c *G* test for association**

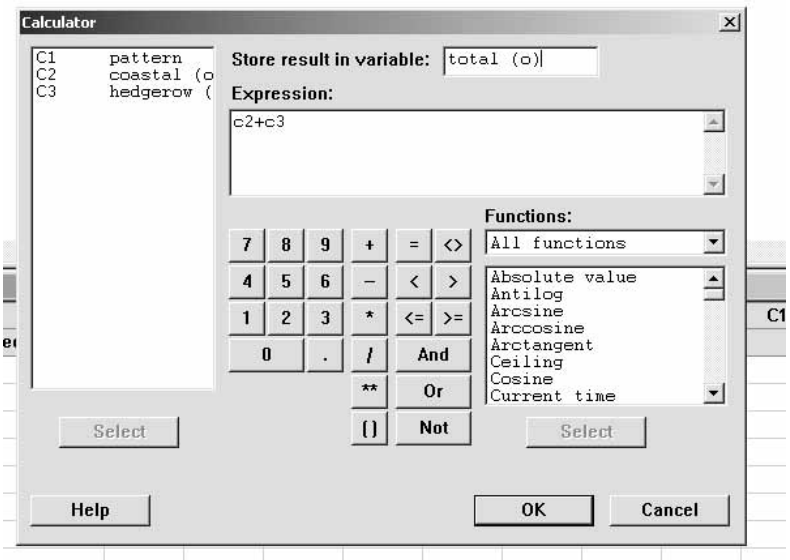
**Step 1.** Enter the data into the worksheet section of the Minitab screen. The letters 'o' in brackets at the end of the column names indicate observed values.



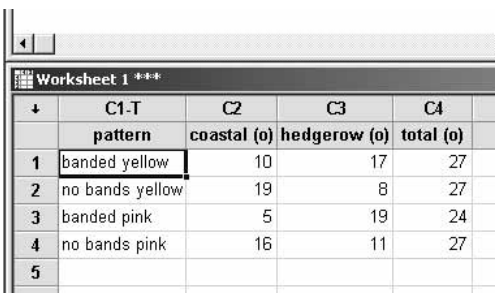
Worksheet 1 ***

| ↓ | C1-T | C2 | C3 | C |
|---|---|---|---|---|
| | pattern | coastal (o) | hedgerow (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.



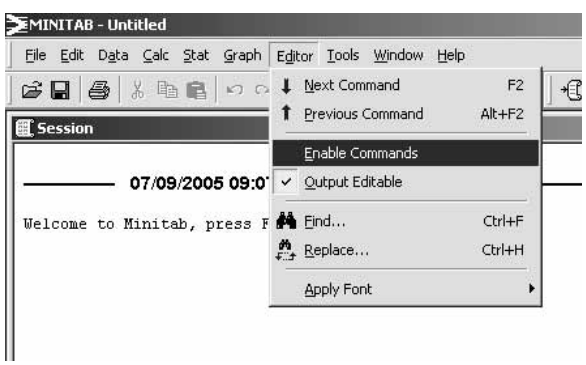

Now click on 'OK'.



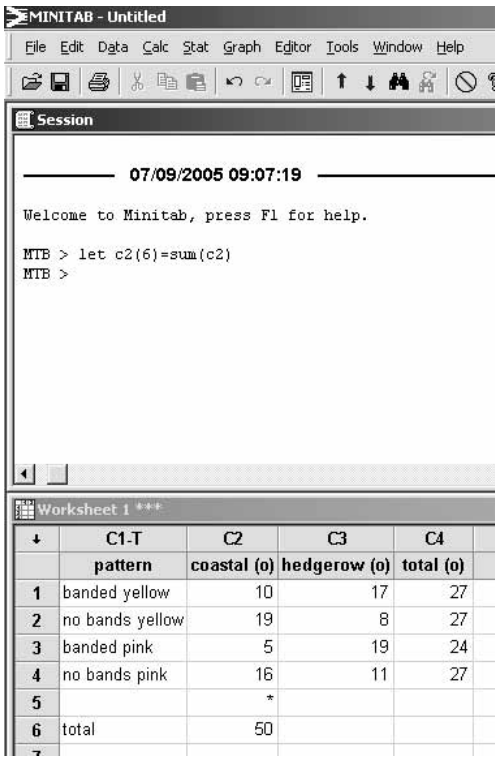
Worksheet 1 ***

| | 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'.

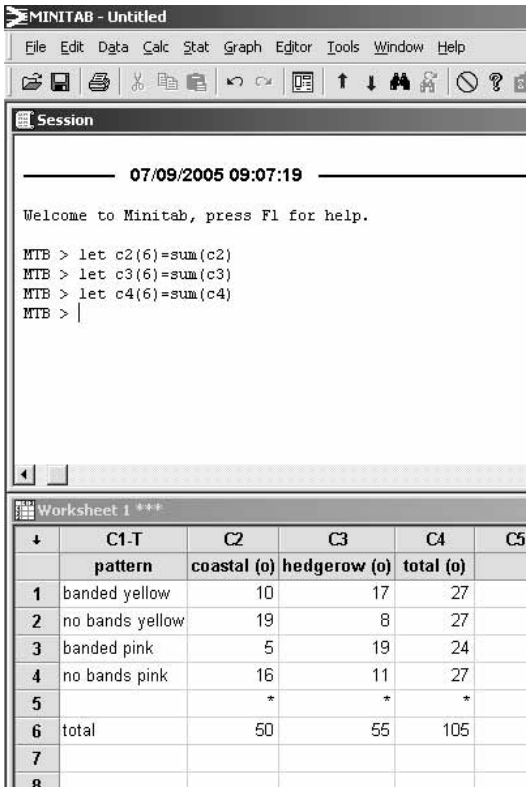


At the 'MTB >' prompt, type 'let c2(6) = sum(c2)'. This will add all the numbers in column 2, and place the result in cell 6 in column 2.



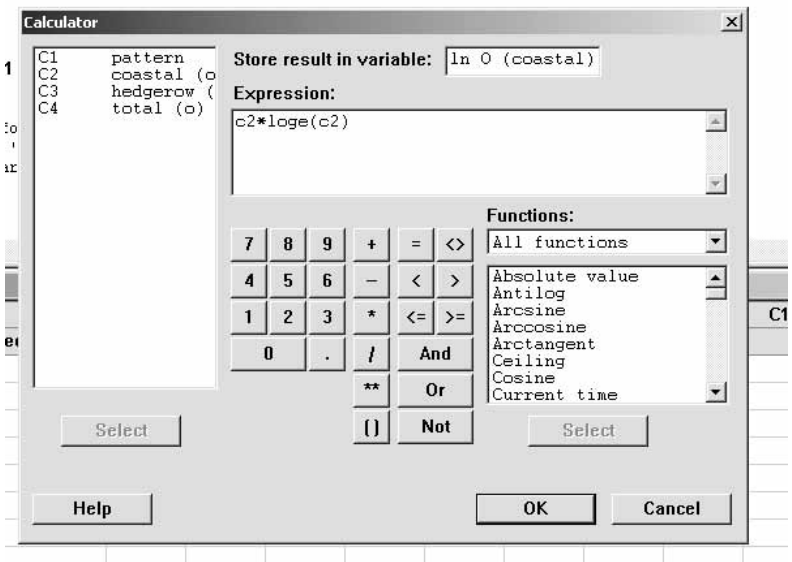

| | C1-T | C2 | C3 | C4 |
|---|---|---|---|---|
| | pattern | coastal (o) | hedgerow (o) | total (o) |
| 1 | banded yellow | 10 | 17 | 27 |
| 2 | no bands yellow | 19 | 8 | 27 |
| 3 | banded pink | 5 | 19 | 24 |
| 4 | no bands pink | 16 | 11 | 27 |
| 5 | | * | | |
| 6 | total | 50 | | |

Repeat the process for columns 3 and 4.

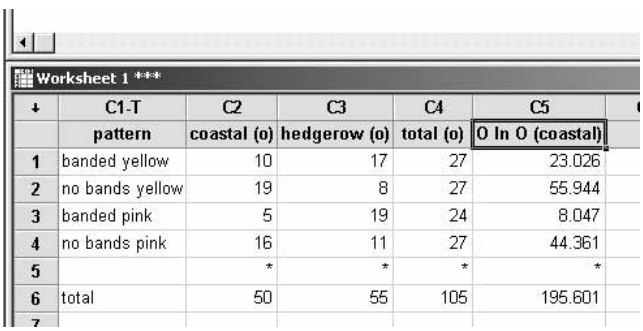



| ↓ | C1-T | C2 | C3 | C4 | C5 |
|---|---|---|---|---|---|
| | pattern | coastal (o) | hedgerow (o) | total (o) | |
| 1 | banded yellow | 10 | 17 | 27 | |
| 2 | no bands yellow | 19 | 8 | 27 | |
| 3 | banded pink | 5 | 19 | 24 | |
| 4 | no bands pink | 16 | 11 | 27 | |
| 5 | | * | * | * | |
| 6 | total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

**Step 4.** Find the values of O ln (O), where 'ln' means 'the natural logarithm of'. Go to 'Calc', 'Calculator', and enter 'O ln O (coastal)' in the 'Store result in variable' window. In the expression window, type 'c2*loge(c2)', where 'log e' means 'logarithm to base e', or natural logarithm.



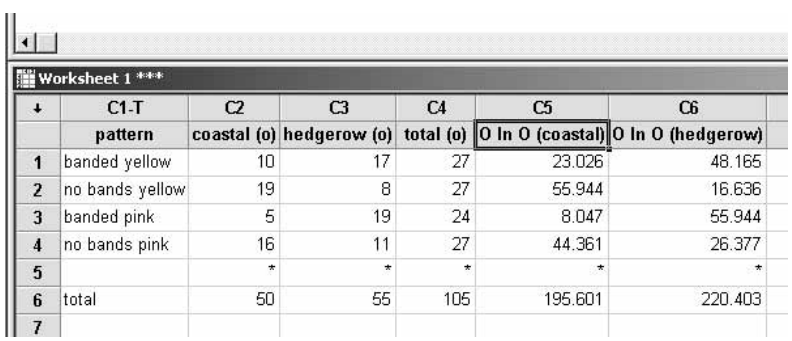

Click on ‘OK’.



Worksheet 1 ***

| | C1-T | C2 | C3 | C4 | C5 |
|---|---|---|---|---|---|
| | pattern | coastal (o) | hedgerow (o) | total (o) | O ln O (coastal) |
| 1 | banded yellow | 10 | 17 | 27 | 23.026 |
| 2 | no bands yellow | 19 | 8 | 27 | 55.944 |
| 3 | banded pink | 5 | 19 | 24 | 8.047 |
| 4 | no bands pink | 16 | 11 | 27 | 44.361 |
| 5 | | * | * | * | * |
| 6 | total | 50 | 55 | 105 | 195.601 |
| 7 | | | | | |

Repeat the process for the hedgerow snails, remembering that their data are in column 3.

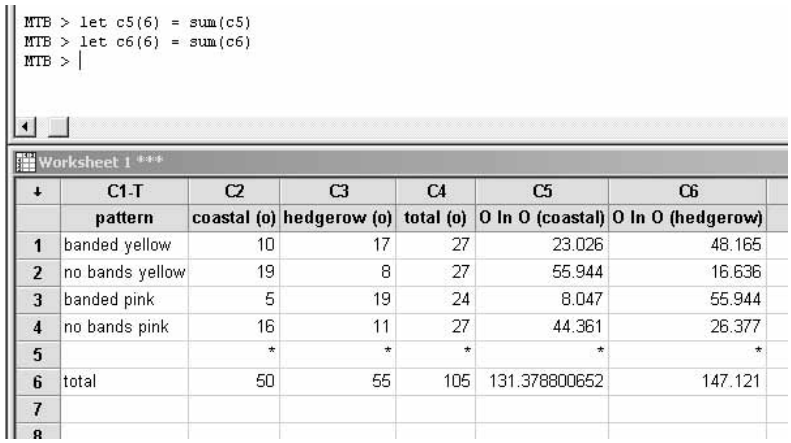


Worksheet 1 ***

| | C1-T | C2 | C3 | C4 | C5 | C6 |
|---|---|---|---|---|---|---|
| | pattern | coastal (o) | hedgerow (o) | total (o) | O ln O (coastal) | O ln O (hedgerow) |
| 1 | banded yellow | 10 | 17 | 27 | 23.026 | 48.165 |
| 2 | no bands yellow | 19 | 8 | 27 | 55.944 | 16.636 |
| 3 | banded pink | 5 | 19 | 24 | 8.047 | 55.944 |
| 4 | no bands pink | 16 | 11 | 27 | 44.361 | 26.377 |
| 5 | | * | * | * | * | * |
| 6 | total | 50 | 55 | 105 | 195.601 | 220.403 |
| 7 | | | | | | |

The entries in row 6 of columns 5 and 6 currently have no meaning. Convert them to column totals by deleting the numbers there at present

(highlight and press 'delete'), then after enabling commands, type in 'let c5(6) = sum(c5)' and 'let c6(6) = sum(c6)'.

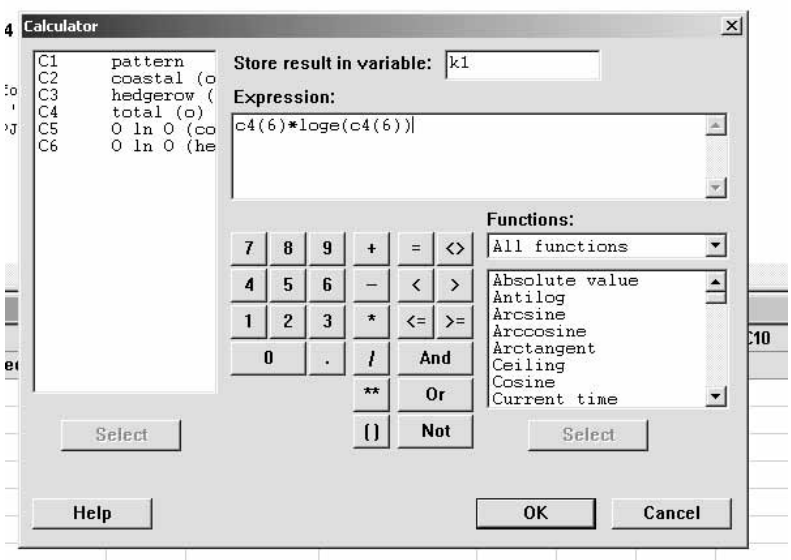


```
MTB > let c5(6) = sum(c5)
MTB > let c6(6) = sum(c6)
MTB > |
```

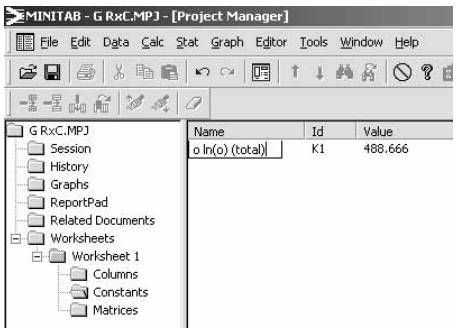
Worksheet 1 ***

| ↓ | C1-T | C2 | C3 | C4 | C5 | C6 |
|---|---|---|---|---|---|---|
| | pattern | coastal (o) | hedgerow (o) | total (o) | O ln O (coastal) | O ln O (hedgerow) |
| 1 | banded yellow | 10 | 17 | 27 | 23.026 | 48.165 |
| 2 | no bands yellow | 19 | 8 | 27 | 55.944 | 16.636 |
| 3 | banded pink | 5 | 19 | 24 | 8.047 | 55.944 |
| 4 | no bands pink | 16 | 11 | 27 | 44.361 | 26.377 |
| 5 | | * | * | * | * | * |
| 6 | total | 50 | 55 | 105 | 131.378800652 | 147.121 |
| 7 | | | | | | |
| 8 | | | | | | |

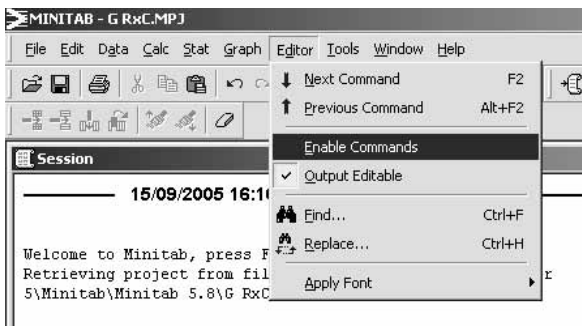
Next, calculate O ln O for the grand total (from cell 6 of column 4). Go to 'Calc', 'Calculator', enter 'k1' in the 'Store result as variable' window, and in the 'Expression' window type 'c4(6)*loge(c4(6))'.




Click on 'OK'. The main screen shouldn't change.
Open the project manager (at bottom left), open the 'Constants' folder, right-click on the word 'Unnamed' next to k1, and select 'Rename'. Give it the name 'o ln(o) (total)'.




Minimize project manager.
Now calculate o ln(o) for the column totals. Click in the upper ('Session') window, go to 'Editor' and select 'Enable Commands'.




In column 1 cell 8, type 'o ln(o)'.
At the 'MTB >' prompt, type 'let c2(8) = c2(6)*loge(c2(6))', and press 'return'.

```
MTB > let c2(8) = c2(6)*loge(c2(6))
MTB >
```

Worksheet 1 ***

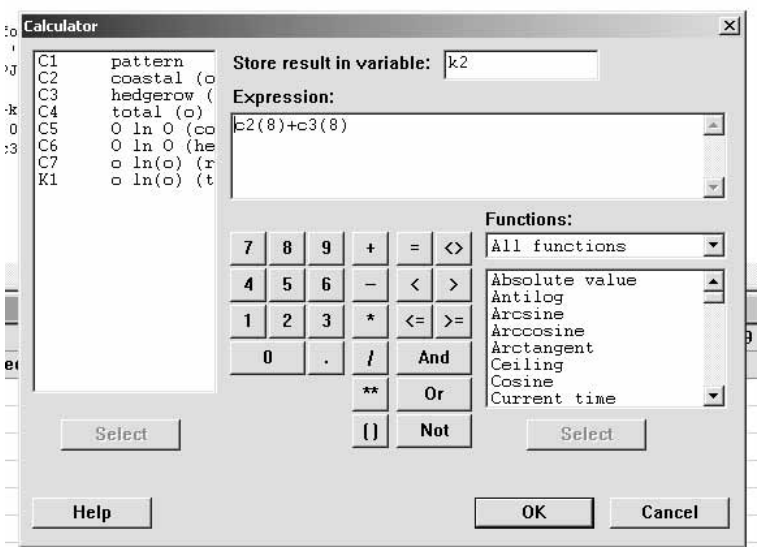
| ↓ | C1-T | C2 | C3 | C4 | C5 | C6 | C7 |
|---|---|---|---|---|---|---|---|
| | pattern | coastal (o) | hedgerow (o) | total (o) | O ln O (coastal) | O ln O (hedgerow) | |
| 1 | banded yellow | 10.000 | 17 | 27 | 23.026 | 48.165 | |
| 2 | no bands yellow | 19.000 | 8 | 27 | 55.944 | 16.636 | |
| 3 | banded pink | 5.000 | 19 | 24 | 8.047 | 55.944 | |
| 4 | no bands pink | 16.000 | 11 | 27 | 44.361 | 26.377 | |
| 5 | | * | * | * | * | * | |
| 6 | total | 50.000 | 55 | 105 | 131.379 | 147.121 | |
| 7 | | * | | | | | |
| 8 | o ln(o) | 195.60115 | | | | | |
| 9 | | | | | | | |

Repeat for column 3 using the command 'let c3(8) = c3(6)*loge(c3(6))'.

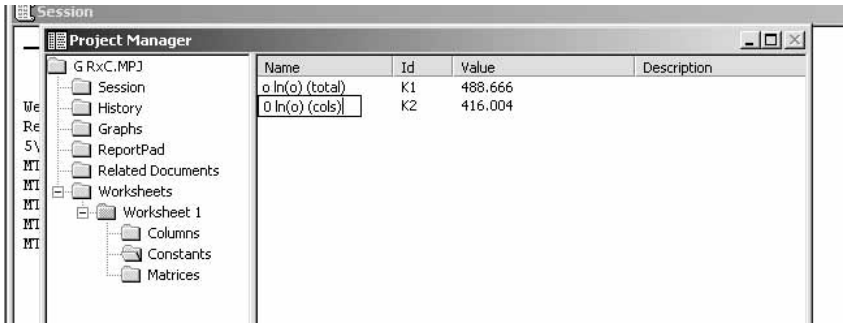
Worksheet 1 ***

| ↓ | C1-T | C2 | C3 | C4 | C5 | C6 | C7 | C8 |
|---|---|---|---|---|---|---|---|---|
| | pattern | coastal (o) | hedgerow (o) | total (o) | O ln O (coastal) | O ln O (hedgerow) | o ln(o) (row total) | |
| 1 | banded yellow | 10.000 | 17.000 | 27 | 23.026 | 48.165 | 88.988 | |
| 2 | no bands yellow | 19.000 | 8.000 | 27 | 55.944 | 16.636 | 88.988 | |
| 3 | banded pink | 5.000 | 19.000 | 24 | 8.047 | 55.944 | 76.273 | |
| 4 | no bands pink | 16.000 | 11.000 | 27 | 44.361 | 26.377 | 88.988 | |
| 5 | | * | * | * | * | * | * | |
| 6 | total | 50.000 | 55.000 | 105 | 131.379 | 147.121 | 343.236 | |
| 7 | | * | * | | | | | |
| 8 | o ln(o) | 195.601 | 220.40332518 | | | | | |
| 9 | | | | | | | | |

Add up the o ln(o) for the columns, and store it in a variable. Go to 'Calc', 'Calculator', type 'k2' in the 'Store results in variable' window, and put 'c2(8) + c3(8)' in the expression window.

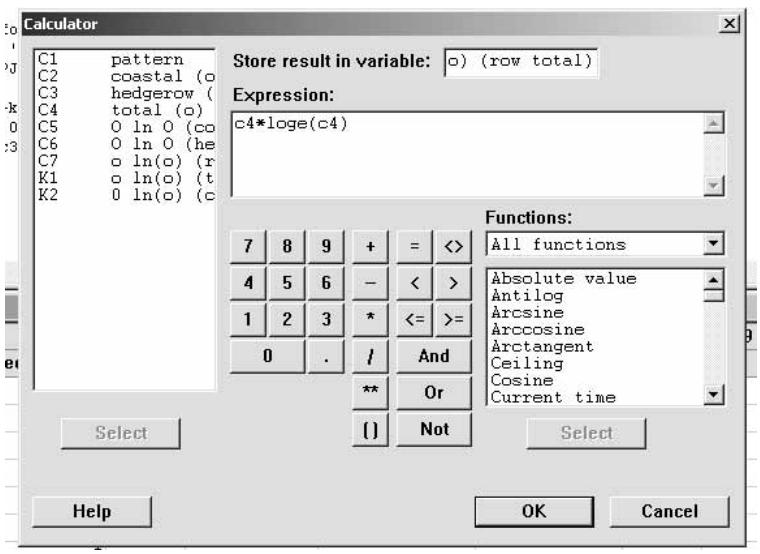



Click on 'OK', and open the project manager. Rename K2 as 'o ln(o) (cols)'.




Minimize project manager.

Calculate o ln(o) for the total of each row. Go to 'Calc', 'Calculator', enter 'o ln(o) (row total)' in the 'Store result in variable' window, and type 'c4*loge(c4)' in the 'Expression' window.




Click on 'OK'.

Worksheet 1 ***

| ↓ | C1-T | C2 | C3 | C4 | C5 | C6 | C7 | C8 |
|---|---|---|---|---|---|---|---|---|
| | pattern | coastal (o) | hedgerow (o) | total (o) | O ln O (coastal) | O ln O (hedgerow) | o ln(o) (row total) | |
| 1 | banded yellow | 10.000 | 17.000 | 27 | 23.026 | 48.165 | 88.988 | |
| 2 | no bands yellow | 19.000 | 8.000 | 27 | 55.944 | 16.636 | 88.988 | |
| 3 | banded pink | 5.000 | 19.000 | 24 | 8.047 | 55.944 | 76.273 | |
| 4 | no bands pink | 16.000 | 11.000 | 27 | 44.361 | 26.377 | 88.988 | |
| 5 | | * | * | * | * | * | * | |
| 6 | total | 50.000 | 55.000 | 105 | 131.379 | 147.121 | 488.666 | |
| 7 | | * | * | | | | | |
| 8 | o ln(o) | 195.601 | 220.40332518 | | | | | |
| 9 | | | | | | | | |

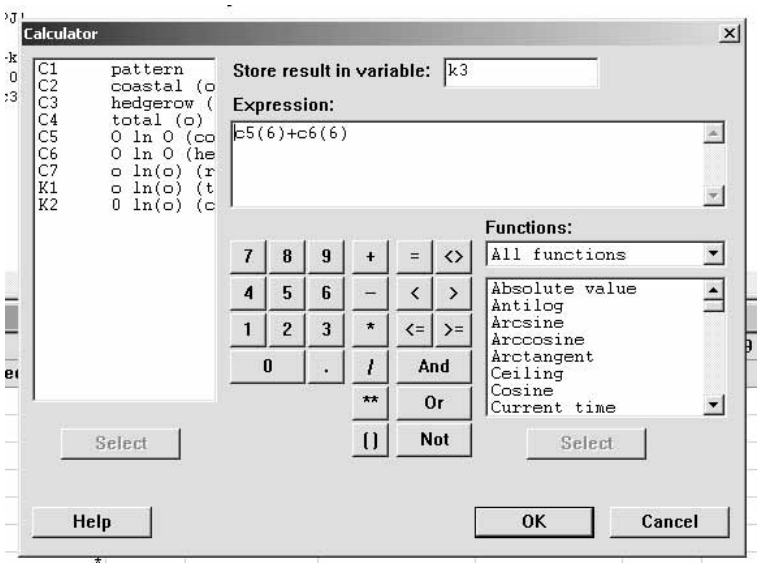
Note that the number in cell c7(6) isn't actually the total. To change it to the total, we add the row totals together and store the result in cell c7(6). First delete the contents of cell c7(6), then (after ensuring commands are enabled) type in the command 'let c7(6) = sum(c7)'.

```
MTB > let c7(6) = sum(c7)
MTB >
```

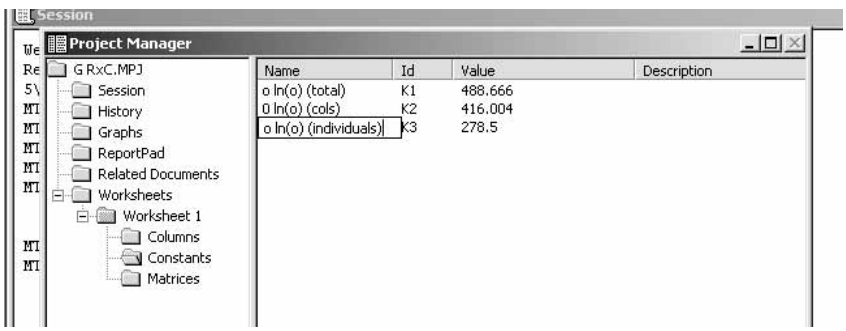
Worksheet 1 ***

| ↓ | C1-T | C2 | C3 | C4 | C5 | C6 | C7 |
|---|---|---|---|---|---|---|---|
| | pattern | coastal (o) | hedgerow (o) | total (o) | O ln O (coastal) | O ln O (hedgerow) | o ln(o) (row total) |
| 1 | banded yellow | 10.000 | 17.000 | 27 | 23.026 | 48.165 | 88.988 |
| 2 | no bands yellow | 19.000 | 8.000 | 27 | 55.944 | 16.636 | 88.988 |
| 3 | banded pink | 5.000 | 19.000 | 24 | 8.047 | 55.944 | 76.273 |
| 4 | no bands pink | 16.000 | 11.000 | 27 | 44.361 | 26.377 | 88.988 |
| 5 | | * | * | * | * | * | * |
| 6 | total | 50.000 | 55.000 | 105 | 131.379 | 147.121 | 343.236078075 |
| 7 | | * | * | | | | |
| 8 | o ln(o) | 195.601 | 220.403 | | | | |
| 9 | | | | | | | |

Finally, we find the sum of all the o ln(o) for the individual measurements, and store that in another constant. This is just the sum of c5(6) and c6(6). Go to 'Calc', 'Calculator', put k3 in the 'Store result in variable' window, and type 'c5(6) + c6(6)' in the 'Expression' window.

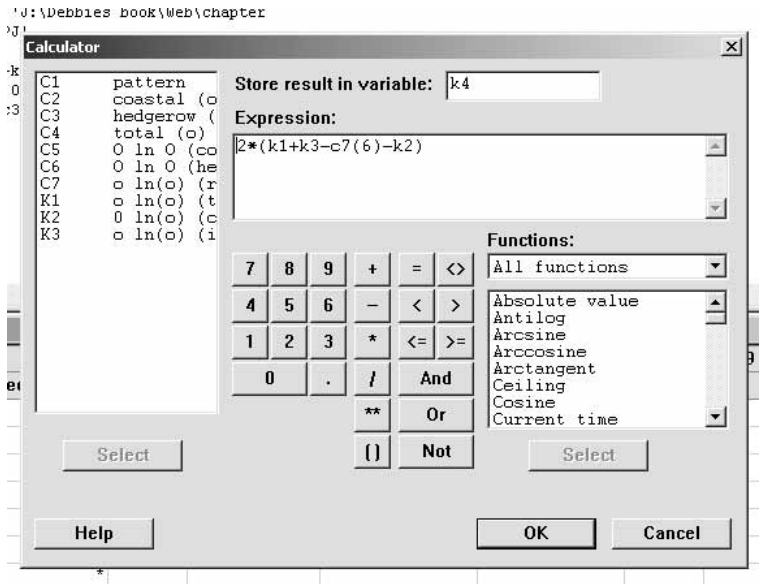



Click on 'OK'.
Open the project manager, and rename k3 as 'o ln(o) (individuals)'.

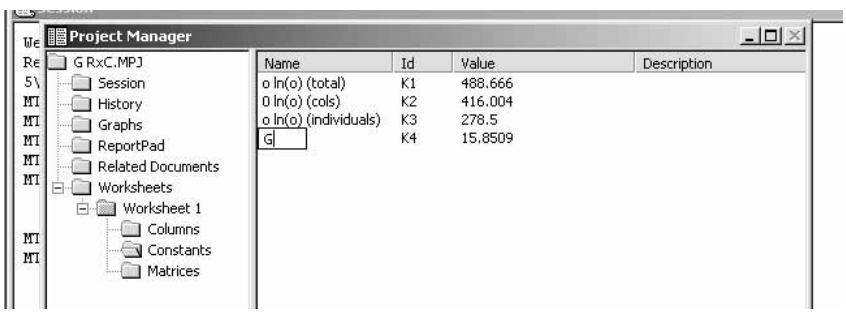



**Step 5**. Calculate $G_{\text{calculated}}$.

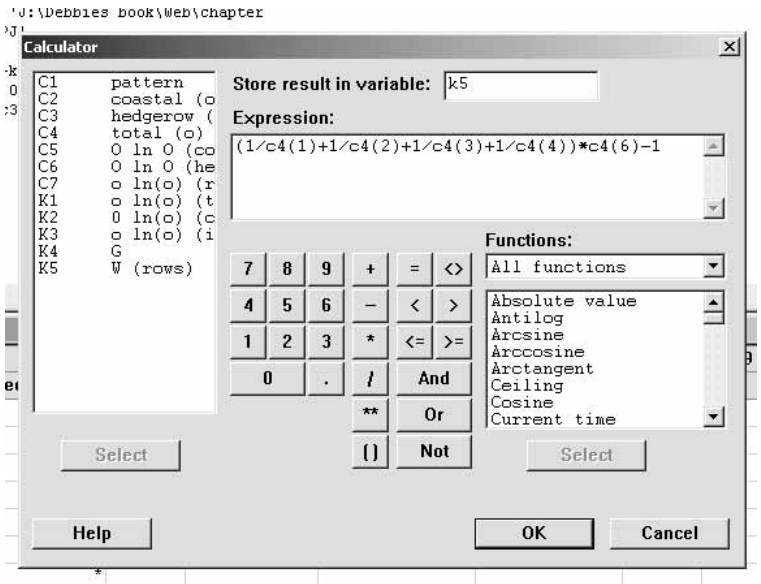
$G = 2 \times$ [o ln(o) (total) + o ln(o) (individuals) − o ln(o) (rows) − o ln(o) (columns)] Go to 'Calc', 'Calculator', put k4 in the 'Store result in variable' window, and type '2*(k1 + k3 − c7(6) − k2)' in the expression window.



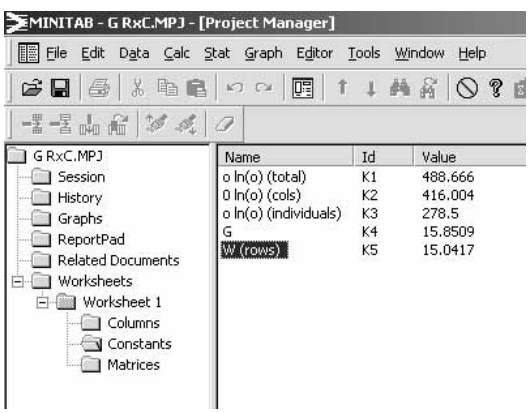

Click on 'OK'.
Go to project manager, and rename k4 as 'G'.



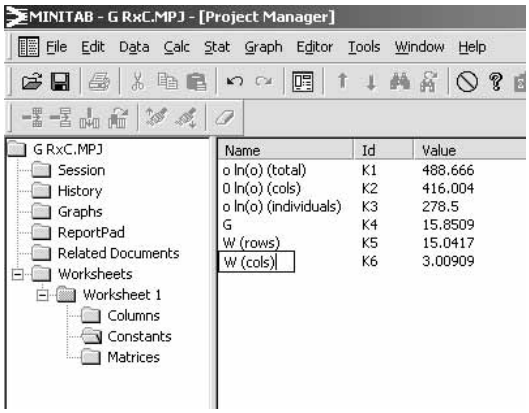

Minimize project manager.
Now we need to apply the Williams' correction.
First, we find 1/row total for each row, add them all together, multiply by the grand total, and subtract 1. This is slightly messy, but go to 'Calc', 'Calculator', put k5 in the 'Store result in variable' window, and type '(1/c4(1) + 1/c4(2) + 1/c4(3) + 1/c4(4))*c4(6) – 1' in the 'Expression' window.



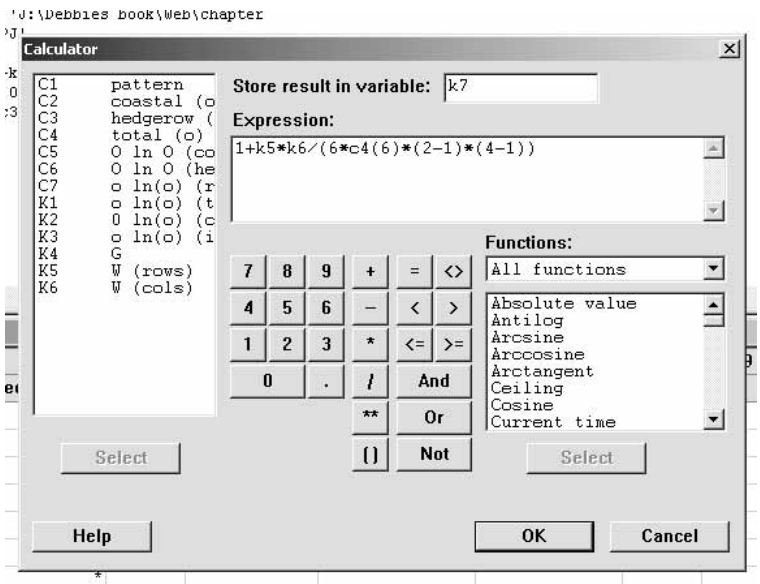

Click on 'OK'. Go to project manager, and rename k5 as 'W (rows)'.



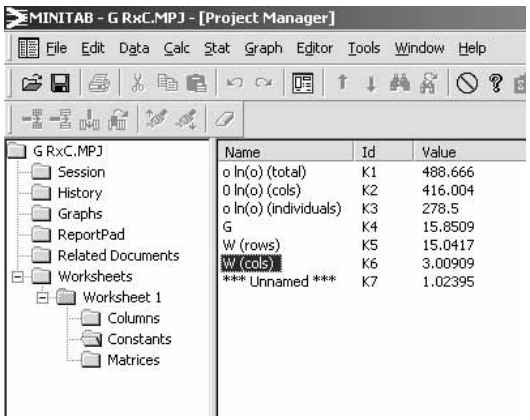

Repeat for the column totals, using k6 and the formula '(1/c2(6) + 1/c3(6))*c4(6) – 1'. Open project manager, and rename k6 as 'W (cols)'.



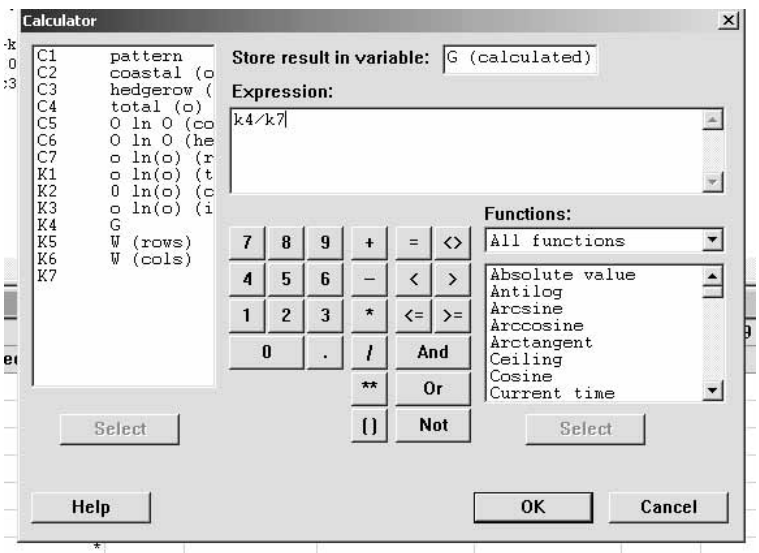

$W$ is $1 + (W\ (\text{rows}) \times W\ (\text{cols}))/(6 \times \text{total} \times (\text{cols} - 1) \times (\text{rows} - 1)$. We have a $2 \times 4$ results table, an the total number of observations is in c4(6). Go to 'Calc', 'Calculator', put k7 in the 'Store result in variable' window, and type '1 + k5*k6/(6*c4(6)*(2 – 1)*(4 – 1))' in the 'Expression' window.




Click on 'OK'. Go to project manager, and rename k7 as 'W'.




$G_{calculated} = G/W$. This is k4/k7: put the result in G(calculated).



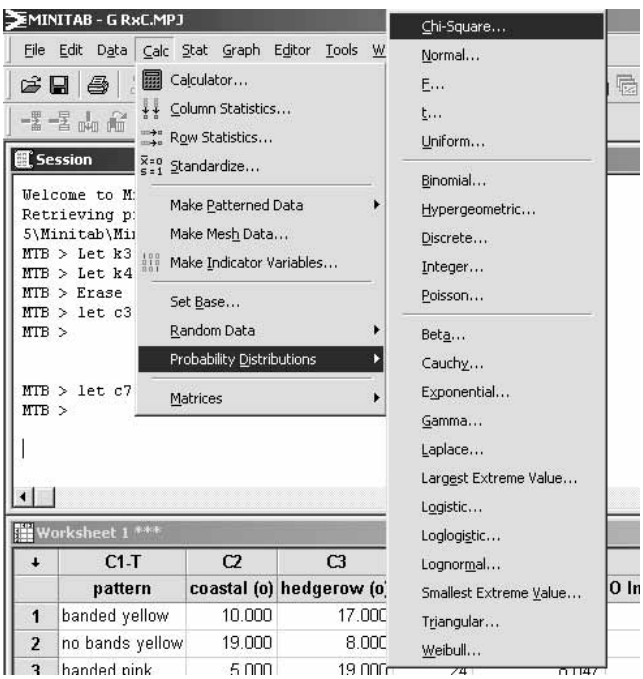

Click on ‘OK’.

Worksheet 1 ***

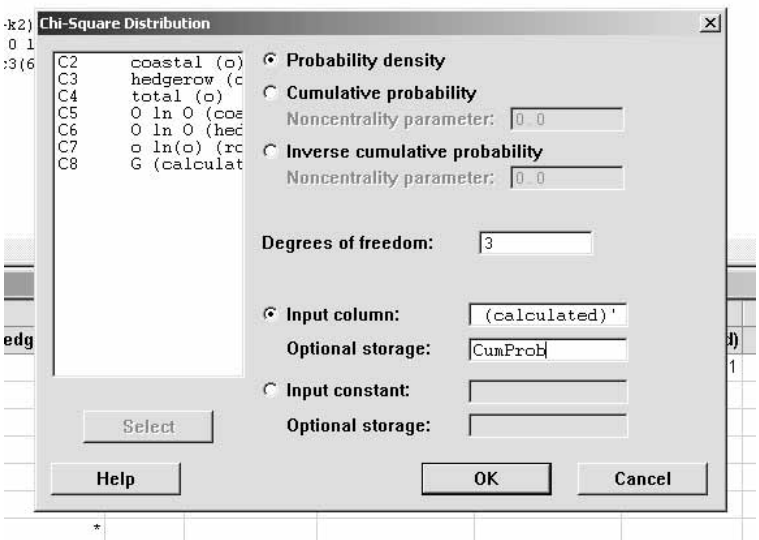
| | C1-T | C2 | C3 | C4 | C5 | C6 | C7 | C8 |
|---|---|---|---|---|---|---|---|---|
| | pattern | coastal (o) | hedgerow (o) | total (o) | O ln O (coastal) | O ln O (hedgerow) | o ln(o) (row total) | G (calculated) |
| 1 | banded yellow | 10.000 | 17.000 | 27 | 23.026 | 48.165 | 88.988 | 15.4801 |
| 2 | no bands yellow | 19.000 | 8.000 | 27 | 55.944 | 16.636 | 88.988 | |
| 3 | banded pink | 5.000 | 19.000 | 24 | 8.047 | 55.944 | 76.273 | |
| 4 | no bands pink | 16.000 | 11.000 | 27 | 44.361 | 26.377 | 88.988 | |
| 5 | | * | * | * | * | * | * | |
| 6 | total | 50.000 | 55.000 | 105 | 131.379 | 147.121 | 343.236078075 | |
| 7 | | * | * | | | | | |
| 8 | o ln(o) | 195.601 | 220.403 | | | | | |
| 9 | | | | | | | | |

The value of $G_{calculated}$ is therefore 15.4801.

**Step 6.** Perform the *G* test, and decide what the result means.
Go to 'Calc', 'Probability Distributions', 'Chi-squared'.



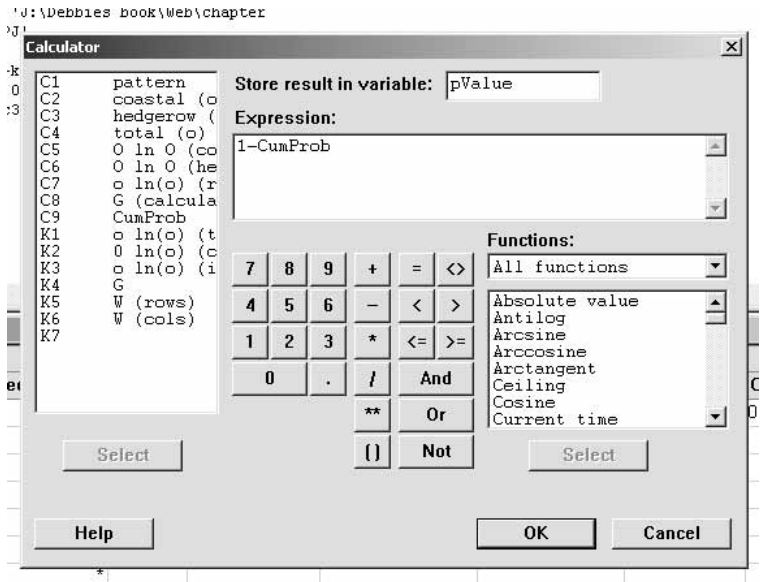

Select 'Cumulative Probability', and enter '3' (for this example) in the 'Degrees of Freedom' window. For the 'input column', enter 'chi-squared', and for 'optional storage' enter 'CumProb'.




Click on 'OK'.

| C6 | C7 | C8 | C9 | C1 |
|---|---|---|---|---|
| O ln O (hedgerow) | o ln(o) (row total) | G (calculated) | CumProb | |
| 48.165 | 88.988 | 15.4801 | 0.0006829 | |
| 16.636 | 88.988 | | | |
| 55.944 | 76.273 | | | |
| 26.377 | 88.988 | | | |
| * | * | | | |
| 147.121 | 343.236078075 | | | |

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*) is not rejected. Go to 'Calc', 'Calculator'. Enter '*p* Value' in the 'Store results in variable' window, and type '1-CumProb' in the 'Expression' window.




Click on 'OK'.

| C7 | C8 | C9 | C10 | C11 |
|---|---|---|---|---|
| chi-squared (c) | chi-squared (h) | chi-squared | CumProb | pValue |
| 0.63492 | 0.57720 | 15.1852 | 0.998335 | 0.0016650 |
| 2.93492 | 2.66811 | | | |
| 3.61607 | 3.28734 | | | |
| 0.76825 | 0.69841 | | | |
| * | * | | | |
| 0.00000 | 0.00000 | | | |

The very low value ($p = 0.0016650$) is below the threshold of $p = 0.05$ which means that the null hypothesis can be rejected. There is a highly significant association ($G = 15.18$, $p = 0.0017$) between the distribution of shell patterns and habitat (coastal and hedgerow) of *Cepea nemoralis*.