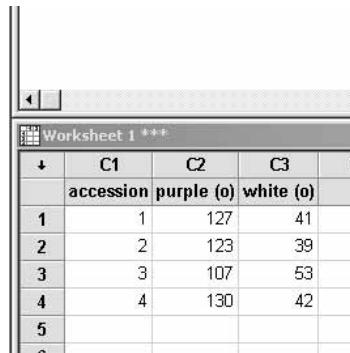


5.2. Heterogeneity in a goodness of fit test

EXAMPLE 5.2. The genetics of tepal colour in *Allium schoenoprasum*

BOX 5.3. How to calculate a chi-squared test for heterogeneity

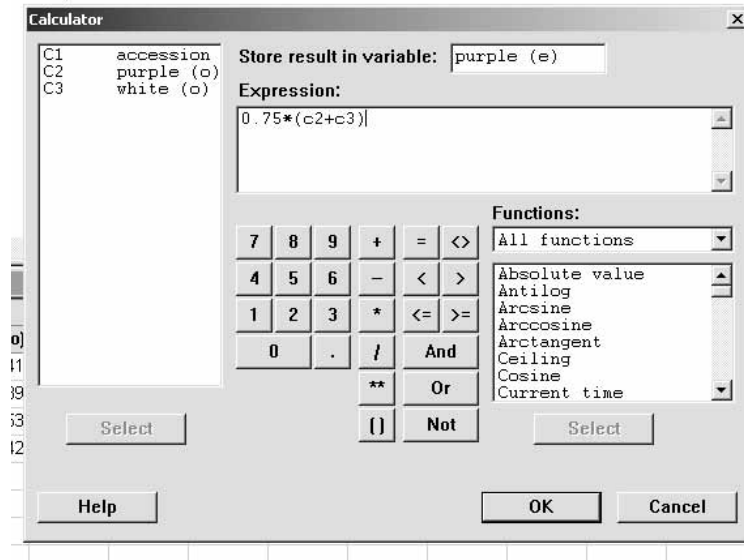
Step 1. Enter the accession data into the worksheet section of the Minitab screen. The letters ‘o’ in brackets at the end of the column names indicate observed values.



	C1	C2	C3	C4
	accession	purple (o)	white (o)	
1	1	127	41	
2	2	123	39	
3	3	107	53	
4	4	130	42	
5				

Step 2. Calculate the **expected** values. Go to ‘Calc’, ‘Calculator’. We expect three-quarters (0.75) of the flowers to be purple, so we calculate the total number of flowers in each accession and multiply by 0.75 to give the expected number of purple flowers. Store the results in a variable called purple (e), where the ‘e’ means ‘expected’.

for help.



Click on 'OK' to perform the calculation.

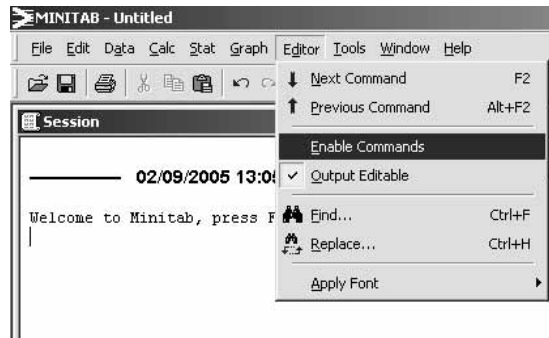
	C1	C2	C3	C4	C5
	accession	purple (o)	white (o)	purple (e)	
1	1	127	41	126.0	
2	2	123	39	121.5	
3	3	107	53	120.0	
4	4	130	42	129.0	

Repeat the process to find the expected number of white flowers, remembering that we only expect one-quarter (0.25) of the flowers to be white.

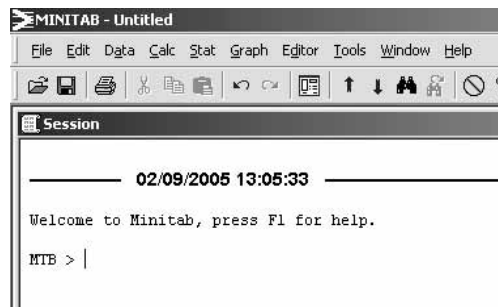
	C1	C2	C3	C4	C5	C6
	accession	purple (o)	white (o)	purple (e)	white (e)	
1	1	127	41	126.0	42.0	
2	2	123	39	121.5	40.5	
3	3	107	53	120.0	40.0	
4	4	130	42	129.0	43.0	
5						

Step 3. Calculate the totals and put them in row 6 (leaving row 5 empty means that we won't confuse them with our data). This needs to be done

using a session command, and to do this, you need to enable session commands. Go to 'Editor' and select 'Enable Commands'.



You should get a 'MTB >' prompt in the session window.



The important session command is 'let'. To calculate the sum of the values in column 2, type in 'let c2(6) = sum(c2)' and press 'return'.

The screenshot shows the Minitab interface. The top window is the 'Session' window, which displays the following text:

```

02/09/2005 13:05:33

Welcome to Minitab, press F1 for help.

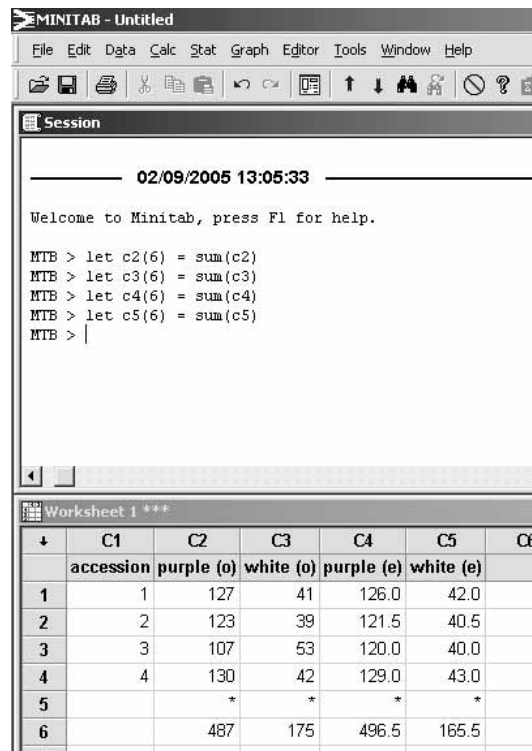
MTB > let c2(6) = sum(c2)
MTB > |

```

The bottom window is the 'Worksheet 1 ***' window, which contains a data table with the following structure:

	C1	C2	C3	C4	C5	C6
	accession	purple (o)	white (o)	purple (e)	white (e)	
1	1	127	41	126.0	42.0	
2	2	123	39	121.5	40.5	
3	3	107	53	120.0	40.0	
4	4	130	42	129.0	43.0	
5		*				
6		487				
7						
8						

(The * in column 2 cell 5 means that there is nothing there.)
Repeat the process for columns 3, 4, and 5.



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

```

02/09/2005 13:05:33

Welcome to Minitab, press F1 for help.

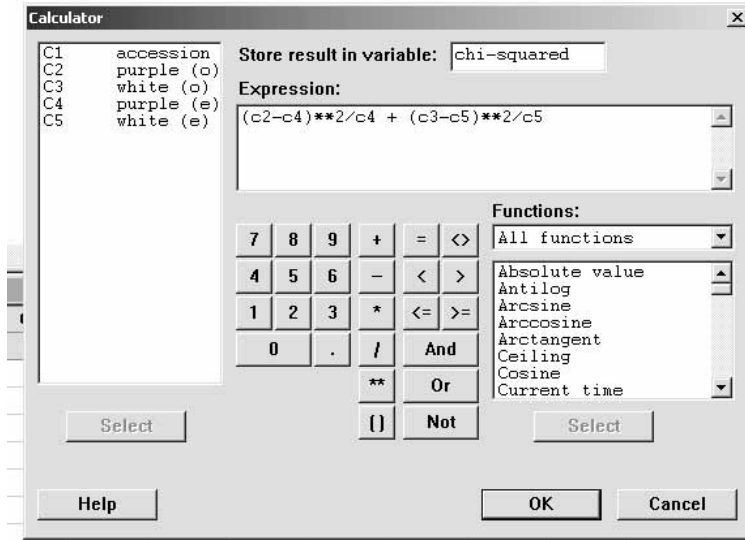
MTB > let c2(6) = sum(c2)
MTB > let c3(6) = sum(c3)
MTB > let c4(6) = sum(c4)
MTB > let c5(6) = sum(c5)
MTB > |

```

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

	C1	C2	C3	C4	C5	C6
	accession	purple (o)	white (o)	purple (e)	white (e)	
1	1	127	41	126.0	42.0	
2	2	123	39	121.5	40.5	
3	3	107	53	120.0	40.0	
4	4	130	42	129.0	43.0	
5		*	*	*	*	
6		487	175	496.5	165.5	

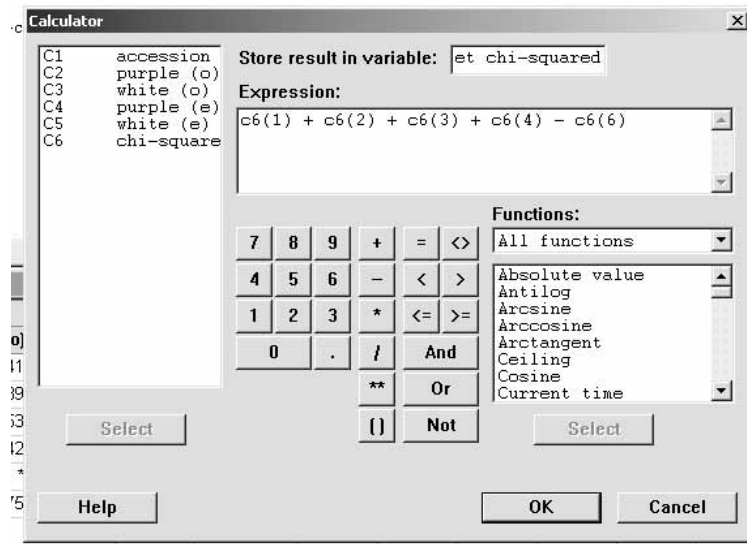
Step 4. Work out chi-squared for each accession and for the totals. Go to 'Calc', 'Calculator' and enter $(c2 - c4)^2/c4 + (c3 - c5)^2/c5$, putting the results in a column called 'Chi-squared'.



Click on 'OK'. The results should appear in a separate column.

	C1	C2	C3	C4	C5	C6
	accession	purple (o)	white (o)	purple (e)	white (e)	chi-squared
1	1	127	41	126.0	42.0	0.03175
2	2	123	39	121.5	40.5	0.07407
3	3	107	53	120.0	40.0	5.63333
4	4	130	42	129.0	43.0	0.03101
5		*	*	*	*	*
6		487	175	496.5	165.5	0.72709
7						

Step 5. Calculate the test value of chi-squared, called the heterogeneity chi-squared. This is the sum of the individual chi-squareds for each accession, minus the chi-squared value for the total. Use the 'Calc', 'Calculator' facility, and type in ' $c6(1) + c6(2) + c6(3) + c6(4) - c6(6)$ '. Store this in a variable called 'het chi-squared'.

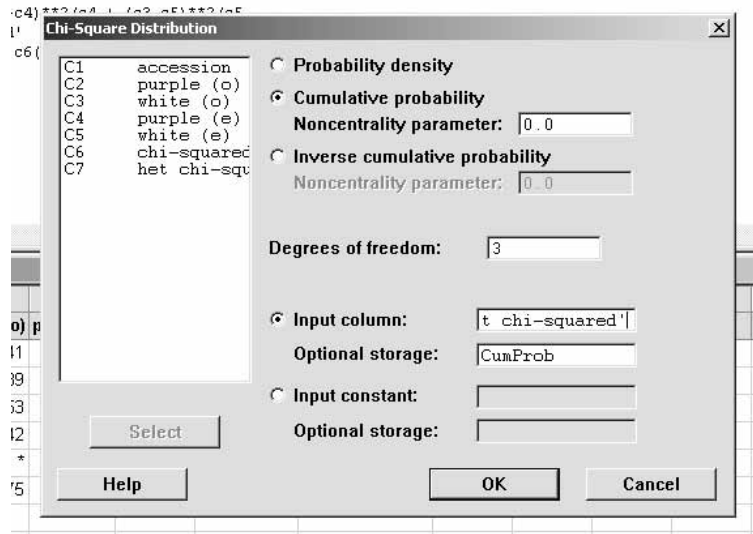


Click on 'OK'.

	C1	C2	C3	C4	C5	C6	C7	C8
	accession	purple (o)	white (o)	purple (e)	white (e)	chi-squared	het chi-squared	
1	1	127	41	126.0	42.0	0.03175	5.04307	
2	2	123	39	121.5	40.5	0.07407		
3	3	107	53	120.0	40.0	5.63333		
4	4	130	42	129.0	43.0	0.03101		
5		*	*	*	*	*		
6		487	175	496.5	165.5	0.72709		
7								

Step 6. Sort out the probability.

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'.



Then click on 'OK'.

	C1	C2	C3	C4	C5	C6	C7	C8	C9
	accession	purple (o)	white (o)	purple (e)	white (e)	chi-squared	het chi-squared	CumProb	
1	1	127	41	126.0	42.0	0.03175	5.04307	0.831330	
2	2	123	39	121.5	40.5	0.07407			
3	3	107	53	120.0	40.0	5.63333			
4	4	130	42	129.0	43.0	0.03101			
5		*	*	*	*	*			
6		487	175	496.5	165.5	0.72709			
7									

The final step is to find the probability that the null hypothesis (there is no heterogeneity between the accessions) can be rejected. Go to 'Calc', 'Calculator'. Enter 'pValue' in the 'Store results in variable' window, and type '1-CumProb' in the 'Expression' window. Click on 'OK'.

	C1	C2	C3	C4	C5	C6	C7	C8	C9
	accession	purple (o)	white (o)	purple (e)	white (e)	chi-squared	het chi-squared	CumProb	pValue
1	1	127	41	126.0	42.0	0.03175	5.04307	0.831330	0.168670
2	2	123	39	121.5	40.5	0.07407			
3	3	107	53	120.0	40.0	5.63333			
4	4	130	42	129.0	43.0	0.03101			
5		*	*	*	*	*			
6		487	175	496.5	165.5	0.72709			
7									

In this example the p value ($p = 0.17$) is greater than the threshold level of $p = 0.05$. Therefore the null hypothesis is not rejected. There is no statistically significant heterogeneity ($p = 0.17$) between the accessions of *Allium schoenoprasum*, and it is therefore reasonable to sum the data across all accessions and use a goodness of fit chi-squared test on the totals.