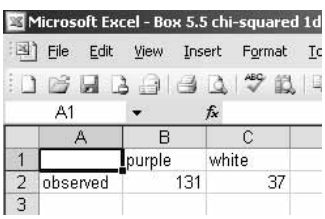
5.4.1 Chi-squared goodness of fit test when there is one degree of freedom

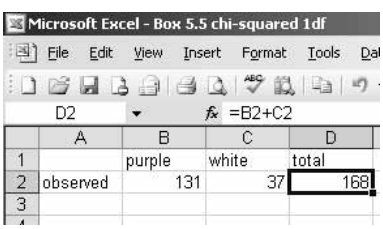
EXAMPLE 5.4. The genetics of tepal colour in *Allium schoenoprasum*

BOX 5.5. How to calculate a goodness of fit chi-squared test when there is one degree of freedom

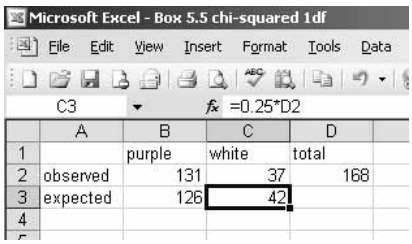
Step 1. Enter the data into the spreadsheet using appropriate row and column headings.



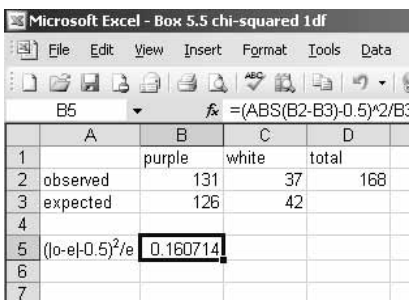
Calculate the total: use the formula '=b2+c2' in cell d2.



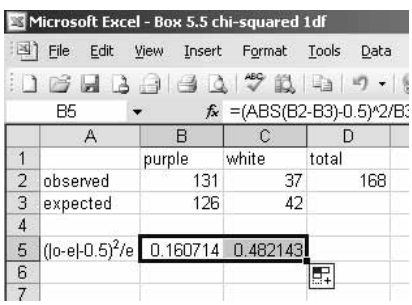
Step 2. Calculate the expected values.
In this case, we expect three-quarters (0.75) of the flowers to be purple, and one-quarter (0.25) to be white. In a new row (expected) use the formula '=0.75*d2' for the expected purple value and '=0.25*d2' for the expected white value.



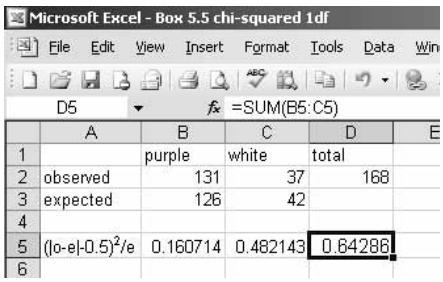
Step 3. calculate the values of $(\text{obs} - \text{exp})^2/\text{exp}$ – but remember to use the Yates' correction $(|\text{obs} - \text{exp}| - 0.5)^2/\text{exp}$, where the vertical bars mean 'take the **absolute** value of'. The 'absolute value' of a number (sometimes called its modulus) is the number, but ignoring any minus sign. In Excel, this is done using the 'abs' function. Use row 5, and in cell b5 type in the formula '=(abs(b2 – b3) – 0.5)^2/b3'. Click on the green tick, or press 'return'.



Drag this across into cell c5.



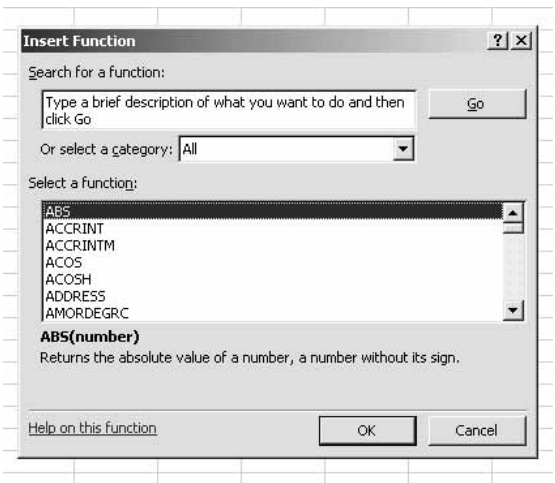
Add these together by typing the formula '=sum(b5:c5)' into cell d5. Click on the green tick, or press 'return'.



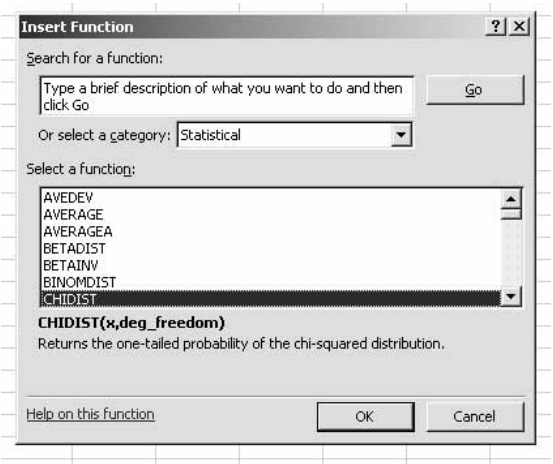
	A	B	C	D	E
1		purple	white	total	
2	observed	131	37	168	
3	expected	126	42		
4					
5	$(\|o-e\|-0.5)^2/e$	0.160714	0.482143	0.64286	
6					

This is our value of chi-squared.

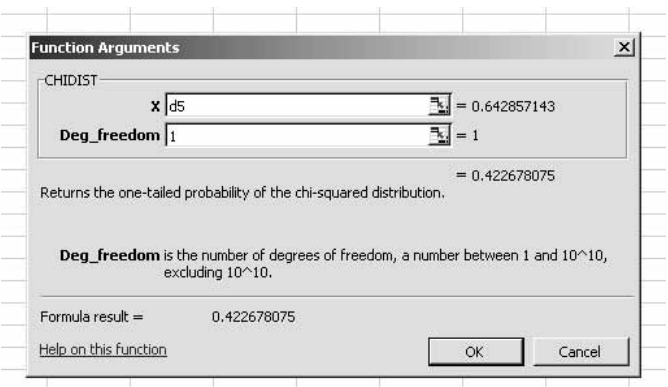
Step 4. Perform the test, and decide what the result means.
Select a suitable cell (we shall use b7), go to 'Insert', 'Function'.



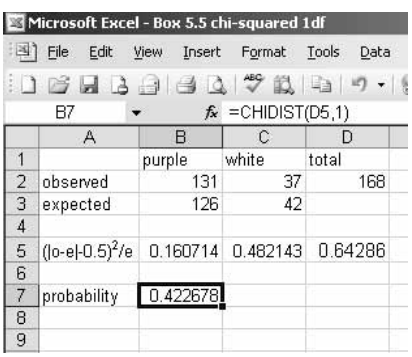
Select the 'Statistical' category, and choose 'CHIDIST' from the list provided.



Click on 'OK'. Insert the location of the cell where the value of chi-squared is stored (d5 in this case) and the number of degrees of freedom (1 in this case).



Click on 'OK'.



	A	B	C	D
1		purple	white	total
2	observed	131	37	168
3	expected	126	42	
4				
5	$(\|o-e\|-0.5)^2/e$	0.160714	0.482143	0.64286
6				
7	probability	0.422678		
8				
9				

The probability (p) = 0.42. This value is greater than $p = 0.05$, so we do not reject the null hypothesis. The flower colours in the F2 generation of *Allium schoenoprasum* plants do not differ significantly ($\chi^2_{\text{calculated}}$ 0.64, $p = 0.42$) from the predicted ratio of three purple: one white. This indicates that the genetic model is probably correct.