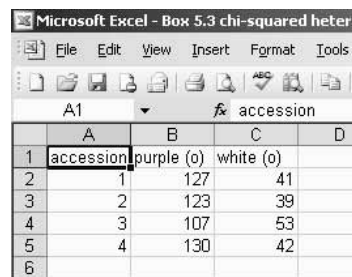


5.2. Heterogeneity in a goodness of fit test

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 data into the spreadsheet using suitable row and column headings.



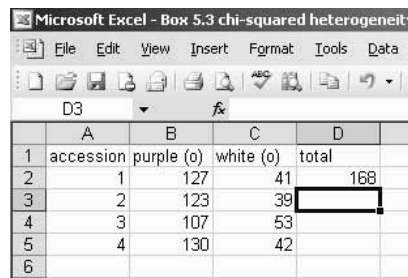
	A	B	C	D
1	accession	purple (o)	white (o)	
2	1	127	41	
3	2	123	39	
4	3	107	53	
5	4	130	42	
6				

(The '(o)' means 'observed', to distinguish these values from the **expected** values that we are going to calculate later.)

Step 2. Calculate the expected values.

First calculate the totals for the rows and columns.

Into cell d2, say, type the formula ' $=b2 + c2$ ', then either click on the green tick or press 'return'.



	A	B	C	D
1	accession	purple (o)	white (o)	total
2	1	127	41	168
3	2	123	39	
4	3	107	53	
5	4	130	42	
6				

Click on cell d2 again, and hover the cursor over the bottom-right corner of the cell. It should change from an open horizontal-vertical cross into an addition sign. Hold down the left mouse button, and drag the cursor down: release it when you reach cell d5.

Microsoft Excel - Box 5.3 chi-squared heterogeneity

	A	B	C	D	E
1	accession	purple (o)	white (o)	total	
2	1	127	41	168	
3	2	123	39	162	
4	3	107	53	160	
5	4	130	42	172	
6					
7					
8					

For the column totals, first find the sum of the first column by typing ‘=sum(b2:b5)’ into cell b6, then clicking on the green tick or pressing ‘return’.

Microsoft Excel - Box 5.3 chi-squared heterogeneity

	A	B	C	D	E
1	accession	purple (o)	white (o)	total	
2	1	127	41	168	
3	2	123	39	162	
4	3	107	53	160	
5	4	130	42	172	
6	total	487			
7					
8					

Now drag the contents of cell b6 across into cells c6 and d6.

Microsoft Excel - Box 5.3 chi-squared heterogeneity

	A	B	C	D	E
1	accession	purple (o)	white (o)	total	
2	1	127	41	168	
3	2	123	39	162	
4	3	107	53	160	
5	4	130	42	172	
6	total	487	175	662	
7					
8					
9					

We expect three-quarters (0.75) of the flowers to be purple. In cell e2, type the formula ‘=0.75*d2’, then click on the green tick or press return.

	A	B	C	D	E
1	accession	purple (o)	white (o)	total	purple (e)
2	1	127	41	168	126
3	2	123	39	162	
4	3	107	53	160	
5	4	130	42	172	
6	total	487	175	662	

Now drag this down into cells e3 to e6.

	A	B	C	D	E	F
1	accession	purple (o)	white (o)	total	purple (e)	
2	1	127	41	168	126	
3	2	123	39	162	121.5	
4	3	107	53	160	120	
5	4	130	42	172	129	
6	total	487	175	662	496.5	

We expect one-quarter (0.25) of the flowers to be white, so use the formula ' $=0.25*d2$ ' in cell f2, and drag this down into cells f3 to f6.

	A	B	C	D	E	F	G
1	accession	purple (o)	white (o)	total	purple (e)	white (e)	
2	1	127	41	168	126	42	
3	2	123	39	162	121.5	40.5	
4	3	107	53	160	120	40	
5	4	130	42	172	129	43	
6	total	487	175	662	496.5	165.5	

Step 3. Calculate chi-squared. Unless you are a maths whiz, it is best to do this in several stages.

First, calculate ' $(\text{obs}-\text{exp})^2/\text{exp}$ ' for the purple flowers. Into cell g2, type the formula ' $=(b2-e2)^2/e2$ ', click on the green tick or press 'return', and drag this down into cells g3 to g6.

	A	B	C	D	E	F	G	
1	accession	purple (o)	white (o)	total	purple (e)	white (e)	(obs-exp) ² /exp (p)	
2		1	127	41	168	126	42	0.007936508
3		2	123	39	162	121.5	40.5	0.018518519
4		3	107	53	160	120	40	1.408333333
5		4	130	42	172	129	43	0.007751938
6	total		487	175	662	496.5	165.5	0.181772407

(The ‘p’) at the end of the heading indicates that the column refers to the purple flowers.)

Repeat the process for the white flowers using the formula ‘=(c2-f2)²/f2’ in cell h2.

	A	B	C	D	E	F	G	H	
1	accession	purple (o)	white (o)	total	purple (e)	white (e)	(obs-exp) ² /exp (p)	(obs-exp) ² /exp (w)	
2		1	127	41	168	126	42	0.007936508	0.023809524
3		2	123	39	162	121.5	40.5	0.018518519	0.055555556
4		3	107	53	160	120	40	1.408333333	4.225
5		4	130	42	172	129	43	0.007751938	0.023255814
6	total		487	175	662	496.5	165.5	0.181772407	0.545317221

The ‘summed’ value of chi-squared is found by adding together all the values of ‘(obs-exp)²/exp’ for the actual data (i.e. NOT the totals). These are conveniently in a rectangular block, and so Excel will do this using the ‘sum’ function.

Find a convenient cell (b8, say), and label it. Type in the formula ‘=sum(g2:h5)’, and click on the green tick or press ‘return’.

	A	B	C	D	E	F	G	H	
1	accession	purple (o)	white (o)	total	purple (e)	white (e)	(obs-exp) ² /exp (p)	(obs-exp) ² /exp (w)	
2		1	127	41	168	126	42	0.007936508	0.023809524
3		2	123	39	162	121.5	40.5	0.018518519	0.055555556
4		3	107	53	160	120	40	1.408333333	4.225
5		4	130	42	172	129	43	0.007751938	0.023255814
6	total		487	175	662	496.5	165.5	0.181772407	0.545317221
7									
8	chi-sq. sum							5.770161	

Find the ‘deviation’ chi-squared by adding the values of ‘(obs-exp)²/exp’

for the totalled data (these values are in cells g6 and h6). This example uses cell b9.

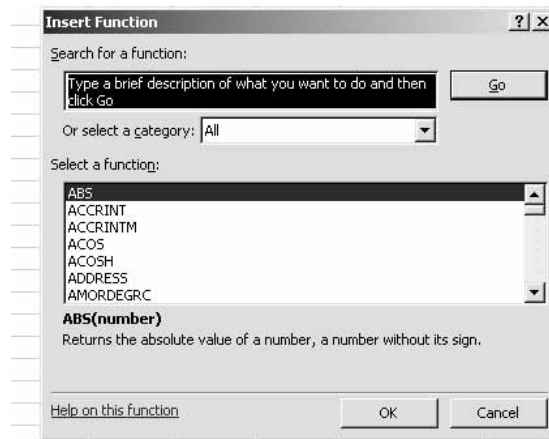
	A	B	C	D	E	F	G	H
1	accession	purple (o)	white (o)	total	purple (e)	white (e)	(obs-exp) ² /exp (p)	(obs-exp) ² /exp (w)
2	1	127	41	168	126	42	0.007936508	0.023809524
3	2	123	39	162	121.5	40.5	0.018518519	0.055555556
4	3	107	53	160	120	40	1.408333333	4.225
5	4	130	42	172	129	43	0.007751938	0.023255814
6	total	487	175	662	496.5	165.5	0.181772407	0.545317221
7								
8	chi-sq. sum	5.770161						
9	chi-sq. dev	0.72709						
10								
11								

The heterogeneity chi-squared is the summed chi-squared minus the deviation chi-squared. In this case, the formula is ' $=b8-b9$ ', and we put the value in cell b10.

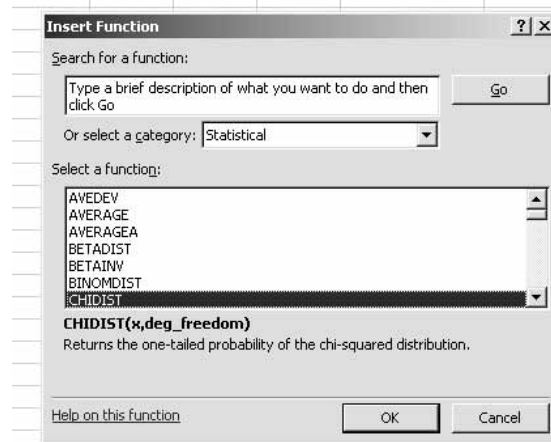
	A	B	C	D	E	F	G	H
1	accession	purple (o)	white (o)	total	purple (e)	white (e)	(obs-exp) ² /exp (p)	(obs-exp) ² /exp (w)
2	1	127	41	168	126	42	0.007936508	0.023809524
3	2	123	39	162	121.5	40.5	0.018518519	0.055555556
4	3	107	53	160	120	40	1.408333333	4.225
5	4	130	42	172	129	43	0.007751938	0.023255814
6	total	487	175	662	496.5	165.5	0.181772407	0.545317221
7								
8	chi-sq. sum	5.770161						
9	chi-sq. dev	0.72709						
10	chi-sq. het	5.043072						
11								
12								

Step 4. Perform the chi-squared test for heterogeneity, and decide what the result means.

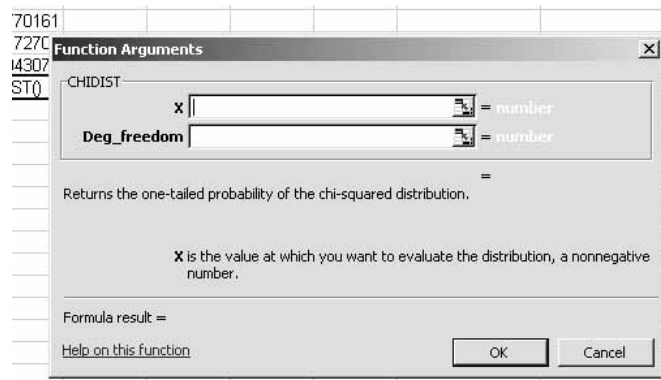
Click on a suitable free cell (b11 in this example), and go to 'Insert', 'Function'.



Select the 'Statistical' category, and choose 'CHIDIST'.



Click on 'OK'.



Insert the location of the cell containing the value of chi-squared (b10 in this case) and the number of **degrees of freedom** (3 in this case). Click on 'OK'.

	A	B	C	D	E	F	G	H
1	accession	purple (o)	white (o)	total	purple (e)	white (e)	(obs-exp) ² /exp (p)	(obs-exp) ² /exp (w)
2	1	127	41	168	126	42	0.007936508	0.023809524
3	2	123	39	162	121.5	40.5	0.018518519	0.055555556
4	3	107	53	160	120	40	1.408333333	4.225
5	4	130	42	172	129	43	0.007751938	0.023255814
6	total	487	175	662	496.5	165.5	0.181772407	0.545317221
7								
8	chi-sq. sum	5.770161						
9	chi-sq. dev	0.72709						
10	chi-sq. het	5.043072						
11	probability	0.16867						
12								

The probability of rejecting the null hypothesis is 0.17, which is above the threshold of $p = 0.05$, so we do not reject the null hypothesis. Therefore there is no statistically significant heterogeneity ($\chi^2_{\text{calculated}} = 5.04$, $p = 0.17$) 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.