

BRCA 2 Genomic Data Set

BRCA.sav [DataSet2] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

5:

	base	count	var																		
1	A	38514																			
2	C	24631																			
3	G	25685																			
4	T	38249																			
5																					
6																					
7																					
8																					
9																					
10																					
11																					
12																					
13																					
14																					
15																					
16																					
17																					
18																					
19																					
20																					
21																					
22																					
23																					
24																					
25																					
26																					
27																					
28																					
29																					
30																					
31																					
32																					
33																					
34																					
35																					
36																					
37																					
38																					
39																					
40																					
41																					
42																					
43																					
44																					

Data View Variable View /

SPSS Processor is ready Weight On

The dialog gives you the option to define the counts

The screenshot shows the SPSS Data Editor window for a file named 'BRCA.sav [DataSet2]'. The main window displays a data grid with the following data:

	base	count	var																			
1	A	38514																				
2	C	24631																				
3	G	25685																				
4	T	38249																				
5																						
6																						
7																						
8																						
9																						
10																						
11																						
12																						
13																						
14																						
15																						
16																						
17																						
18																						
19																						
20																						
21																						
22																						
23																						
24																						
25																						
26																						
27																						
28																						
29																						
30																						
31																						
32																						
33																						
34																						
35																						
36																						
37																						
38																						
39																						
40																						
41																						
42																						
43																						
44																						

The 'Weight Cases' dialog box is open, showing the following options:

- Do not weight cases
- Weight cases by

The 'Weight cases by' option is currently unselected. The 'Frequency Variable' field is empty. The 'Current Status' is 'Weight cases by base'. Buttons for 'OK', 'Reset', 'Cancel', and 'Help' are visible.

At the bottom of the window, the status bar indicates 'SPSS Processor is ready' and 'Weight On'.

Select the *Weight cases by* toggle, and pass the **count** variable into the *Frequency Variable* box

The screenshot shows the SPSS Data Editor window for a file named 'BRCA.sav [DataSet2]'. The main data grid contains the following data:

	base	count	var																		
1	A	38514																			
2	C	24631																			
3	G	25685																			
4	T	38249																			
5																					
6																					
7																					
8																					
9																					
10																					
11																					
12																					
13																					
14																					
15																					
16																					
17																					
18																					
19																					
20																					
21																					
22																					
23																					
24																					
25																					
26																					
27																					
28																					
29																					
30																					
31																					
32																					
33																					
34																					
35																					
36																					
37																					
38																					
39																					
40																					
41																					
42																					
43																					
44																					

The 'Weight Cases' dialog box is open, showing the following options:

- Do not weight cases
- Weight cases by

The 'Frequency Variable' field contains 'Count (count)'. The 'Current Status' is 'Weight cases by base'. Buttons for 'OK', 'Reset', 'Cancel', and 'Help' are visible.

The status bar at the bottom of the window shows 'Data View | Variable View |' and 'SPSS Processor is ready'.

Use *Analyze*, *Nonparametric Tests*, *Chi-Square* pulldown selections

The screenshot shows the SPSS Data Editor window for a file named 'BRCA.sav [DataSet2]'. The 'Analyze' menu is open, and the path 'Analyze > Nonparametric Tests > Chi-Square...' is highlighted. The data grid shows a table with columns 'base' and 'count'. The first four rows contain data for categories A, C, G, and T. The status bar at the bottom indicates 'Chi-Square' and 'SPSS Processor is ready'.

33 :	base	count																		
1	A	38514																		
2	C	24631																		
3	G	25685																		
4	T	38249																		
5																				
6																				
7																				
8																				
9																				
10																				
11																				
12																				
13																				
14																				
15																				
16																				
17																				
18																				
19																				
20																				
21																				
22																				
23																				
24																				
25																				
26																				
27																				
28																				
29																				
30																				
31																				
32																				
33																				
34																				
35																				
36																				
37																				
38																				
39																				
40																				
41																				
42																				
43																				
44																				

Pass the **base** variable into the *Test Variable List*, and click OK. here we test against the probability distribution with all probabilities equal.

The screenshot shows the SPSS Data Editor interface with the 'Chi-Square Test' dialog box open. The dialog box is titled 'Chi-Square Test' and has a blue header. It contains the following elements:

- Count [count]:** A text box with a pencil icon and a dropdown arrow.
- Test Variable List:** A list box containing 'Nucleotide [base]'. Buttons 'OK', 'Reset', 'Cancel', and 'Help' are to its right.
- Expected Range:** Two radio buttons: 'Get from data' (selected) and 'Use specified range'. Below 'Use specified range' are 'Lower:' and 'Upper:' text boxes.
- Expected Values:** Two radio buttons: 'All categories equal' (selected) and 'Values:'. Below 'Values:' are 'Add', 'Change', and 'Remove' buttons.
- Options...:** A button at the bottom right of the dialog box.

The background shows a grid of data with columns labeled 'var' and rows numbered from 16 to 44. The status bar at the bottom indicates 'Data View Variable View' and 'SPSS Processor is ready'.

The results are computed.

Nucleotide

	Observed N	Expected N	Residual
A	38514	31769.8	6744.3
C	24631	31769.8	-7138.8
G	25685	31769.8	-6084.8
T	38249	31769.8	6479.3
Total	127079		

Test Statistics

	Nucleotide
Chi-Square ^a	5522.597
df	3
Asymp. Sig.	.000

a. 0 cells (.0%) have expected frequencies less than 5.
The minimum expected cell frequency is 31769.8.

To test against a different set of probabilities, use the *Expected Values* toggle

The screenshot shows the SPSS Data Editor window with a Chi-Square Test dialog box open. The dialog box is titled "Chi-Square Test" and has a blue header. It contains several sections:

- Count [count]:** A text box with a pencil icon and a button.
- Expected Range:** Two radio buttons: "Get from data" (selected) and "Use specified range". Below "Use specified range" are "Lower:" and "Upper:" text boxes.
- Test Variable List:** A list box containing "Nucleotide [base]" and an "OK" button.
- Expected Values:** Two radio buttons: "All categories equal" and "Values: 0.3" (selected). Below "Values: 0.3" are "Add", "Change", and "Remove" buttons.
- Buttons:** "Reset", "Cancel", "Help", and "Options..." buttons.

The background shows a grid of data with columns labeled "var" and rows numbered from 16 to 44. The status bar at the bottom indicates "Data View | Variable View" and "SPSS Processor is ready".

We test against the probability distribution we probabilities (0.3 ,0.2,0.2,0.3).
We type these values successively into the *Values* box, and click *Add*

The screenshot shows the SPSS Data Editor interface with a 'Chi-Square Test' dialog box open. The dialog box is titled 'Chi-Square Test' and has a blue header. It contains the following elements:

- Count [count]:** A text box with a pencil icon and a right-pointing arrow.
- Test Variable List:** A list box containing 'Nucleotide [base]' with a right-pointing arrow.
- Expected Range:** Two radio buttons: 'Get from data' (selected) and 'Use specified range'. Below 'Use specified range' are 'Lower:' and 'Upper:' text boxes.
- Expected Values:** Two radio buttons: 'All categories equal' and 'Values' (selected). Below 'Values' is a list box containing '0.3', '0.2', '0.2', and '0.3'. To the left of this list are 'Add', 'Change', and 'Remove' buttons. The 'Add' button is highlighted.
- Buttons:** 'OK', 'Reset', 'Cancel', 'Help', and 'Options...' are located on the right side of the dialog box.

The background shows a grid of data with columns labeled 'var' and rows numbered from 16 to 44. The status bar at the bottom indicates 'Data View Variable View' and 'SPSS Processor is ready'.

The new results are produced.

Nucleotide

	Observed N	Expected N	Residual
A	38514	38123.7	390.3
C	24631	25415.8	-784.8
G	25685	25415.8	269.2
T	38249	38123.7	125.3
Total	127079		

Test Statistics

	Nucleotide
Chi-Square ^a	31.492
df	3
Asymp. Sig.	.000

a. 0 cells (.0%) have expected frequencies less than 5.
The minimum expected cell frequency is 25415.8.

Use the *Data*, *Weight Cases* pulldown selections to define the table

The screenshot shows the SPSS Data Editor interface for a file named 'HairEyes.sav'. A 'Weight Cases' dialog box is open, showing the 'Weight cases by' option selected with 'Count (count)' as the frequency variable. The main data grid is visible, showing a list of cases (rows) and variables (columns). The first few rows of data are as follows:

Case #	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6	Var 7	Var 8	Var 9	Var 10	Var 11	Var 12	Var 13	Var 14	Var 15	Var 16	Var 17	Var 18
8	4	2	2	29														
9	1	3	3	26														
10	2	3	3	17														
11	3	3	3	14														
12	4	3	3	14														
13	1	4	4	7														
14	2	4	4	94														
15	3	4	4	10														
16	4	4	4	16														

The dialog box also includes 'OK', 'Reset', 'Cancel', and 'Help' buttons. The status bar at the bottom indicates 'SPSS Processor is ready' and 'Weight On'.

Use the *Analyze, Descriptive Statistics, Crosstabs* pulldown selections

The screenshot shows the SPSS Data Editor interface with the 'Analyze' menu open. The 'Descriptive Statistics' sub-menu is also open, and the 'Crosstabs...' option is selected. The main data grid shows a dataset with columns 'eye', 'hair', and 'c' and rows numbered 1 through 54. The 'c' column contains numerical values for rows 9 through 16.

	eye	hair	c
1	1	1	
2	2	1	
3	3	1	
4	4	1	
5	1	2	
6	2	2	
7	3	2	
8	4	2	
9	1	3	26
10	2	3	17
11	3	3	14
12	4	3	14
13	1	4	7
14	2	4	94
15	3	4	10
16	4	4	16
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			
51			
52			
53			
54			

Pass **eye** into the *Row(s)* list, and **hair** into the *Column(s)* list

The image shows the SPSS Data Editor interface with the Crosstabs dialog box open. The dialog box is titled "Crosstabs" and has a blue header. It contains the following elements:

- Count [count]:** A list of variables to be included in the crosstabs.
- Row(s):** A list containing "Eye Colour [eye]".
- Column(s):** A list containing "Hair Colour [hair]".
- Layer 1 of 1:** A section for defining layers, with "Previous" and "Next" buttons.
- Display clustered bar charts:** A checkbox that is currently unchecked.
- Suppress tables:** A checkbox that is currently unchecked.
- Buttons:** "OK", "Reset", "Cancel", "Help", "Statistics...", "Cells...", and "Format...".

The background is a grid of variables, with 54 rows and 17 columns, all labeled "var". The status bar at the bottom indicates "Data View" and "SPSS Processor is ready".

Click the *Statistics* button on the *Crosstabs* dialog, and select *Chi-square*.
Click *Continue*, then *OK*.

The screenshot shows the SPSS Data Editor window titled "HairEyes.sav [DataSet1] - SPSS Data Editor". The main window displays a grid of data with columns labeled "var" and rows numbered 1 through 54. A "Crosstabs" dialog box is open, and its "Statistics" sub-dialog is also open. In the "Statistics" sub-dialog, the "Chi-square" checkbox is checked. Other options include "Nominal" (with "Contingency coefficient", "Phi and Cramér's V", "Lambda", and "Uncertainty coefficient"), "Ordinal" (with "Gamma", "Somers' d", and "Kendall's tau-b/c"), "Nominal by Interval" (with "Eta"), "Cochran's and Mantel-Haenszel statistics" (with a "Test common odds ratio equals" field set to 1), "Correlations", "Kappa", "Risk", and "McNemar". Buttons for "Continue", "Cancel", and "Help" are visible in the sub-dialog. The status bar at the bottom indicates "SPSS Processor is ready" and "Weight On".

The results are produced.

Eye Colour * Hair Colour Crosstabulation

Count		Hair Colour				Total
		Black	Brown	Red	Blond	
Eye Colour	Brown	68	119	26	7	220
	Blue	20	84	17	94	215
	Hazel	15	54	14	10	93
	Green	5	29	14	16	64
Total		108	286	71	127	592

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	138.290 ^a	9	.000
Likelihood Ratio	146.444	9	.000
Linear-by-Linear Association	28.292	1	.000
N of Valid Cases	592		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 7.68.