Chi-Square Using Weight Cases

David C. Howell

Lisa McGuire, at Virginia Tech, sent me a note asking if I would elaborate on the use of the **Weight Cases** command in SPSS to run a chi-square analysis on a contingency table. I was scheduled to do something boring, so I jumped at the chance. Although what I say here is specific to SPSS, similar procedures are also possible in SAS and Minitab.

I have adapted a study by Astin et al. (1995), which I took from an excellent nonparametric text by Hollander and Wolfe (1999). Astin et al. studied 87 women, of whom 50 were battered women and 37 were maritally distressed, but not battered. These women were further divided into those who fit a diagnosis of Post Traumatic Stress Disorder and those who did not. The data are shown below:

PTSD	Battered	Not Battered	Totals
Yes	29	7	36
No	21	30	51
Totals	50	37	87

Data Files with Individual Observations

One possible SPSS data file that could represent these data is shown below. I have shown only sections of the table to save space.

	ptsd	battered	var	var
1	Yes	Yes		
2	Yes	Yes		
3	Yes	Yes		
4	...	...		
5	Yes	Yes		
6	Yes	Yes		
7	Yes	No		
8	Yes	No		
9	...	...		
10	Yes	No		
11	No	Yes		
12	No	Yes		
13	No	Yes		
14	No	Yes		
15	No	Yes		
16	No	Yes		
17	...	...		
18	No	Yes		
19	No	Yes		
20	No	Yes		
21	No	No		
22	No	No		
23	No	No		
24	...	...		
25	No	No		
26	No	No		
27	No	No		
28	No	No		
29				
30				

temp.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities

9 : battered

Data View   Variable View

Notice that I have used ellipses (...) to show where I left out data to keep the graphic from being too long. {Those of you who use my book will note that I usually code data numerically (e.g. 1 = Yes, 2 = No), rather than alphanumerically. For chi-square it doesn't matter, and I coded with string variables (e. g. words) for clarity.}

With this file you can run a chi-square test on a contingency table. The complete data file would be 87 lines long, because we have 87 women in the study. I had to create this file with all 87 lines, and it was a pain in the neck typing a "Yes/Yes" pair 29 times, a "Yes/No" pair 7 times, a "No/Yes" pair 21 times, and a "No/No" pair 30 times. Imagine what I would have had to do if we had 500 subjects.

The resulting chi-square follows:

PTSD * Battered Crosstabulation

Count

		Battered		Total
		No	Yes	
PTSD	No	30	21	51
	Yes	7	29	36
Total		37	50	87

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	13.389[b]	1	.000		
Continuity Correction[a]	11.826	1	.001		
Likelihood Ratio	14.086	1	.000		
Fisher's Exact Test				.000	.000
N of Valid Cases	87				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 15.31.

You can see that the first table perfectly replicates the data table Astin et al. reported, though SPSS reorders rows and columns to keep "yes" and "No" in their alphabetical order. You can also see that chi-square = 13.389 on 1 df , which has a probability under H_0 = .000. We will reject the null

hypothesis and conclude that battered women are more likely to suffer from PTSD than non-battered women who also face marital distress. PTSD is apparently not just a result of distress; battering adds its own important contribution.

Shortening the Data Input Process

The data table that I created above was, as I said, a pain in the neck. I had to type in 87 lines of data without making any errors. But I know that many of those lines were copies of each other. For example, there were 29 lines with Yes in both columns, 7 with Yes in the first and No in the second column, and so on. That suggests that I might be able to shortcut all that work if I could find a way of telling SPSS how many lines look alike. For instance, I want to tell SPSS to take a line that has Yes in both columns, and (internally) repeat that line 29 times without my typing it.

To accomplish this, I start with the data table below:

This table is only four lines long, but it contains all of the information that the first data file did. The only trick is in getting SPSS to understand that.

The Weight Cases Command

We will begin with the **weighted cases** command. This command tells SPSS to take each case (line of data) and weight it by some variable--in our case, "freq." In other words, "Take the first line (Yes/Yes) and count it 29 times. Then take the next line (Yes/No), and count it 7 times, etc." To do this we use the **Data/Weight Cases** command from the main menu. See below:
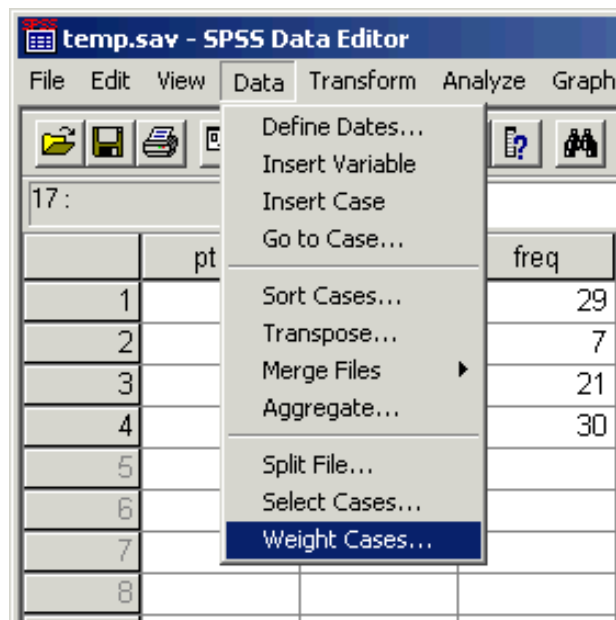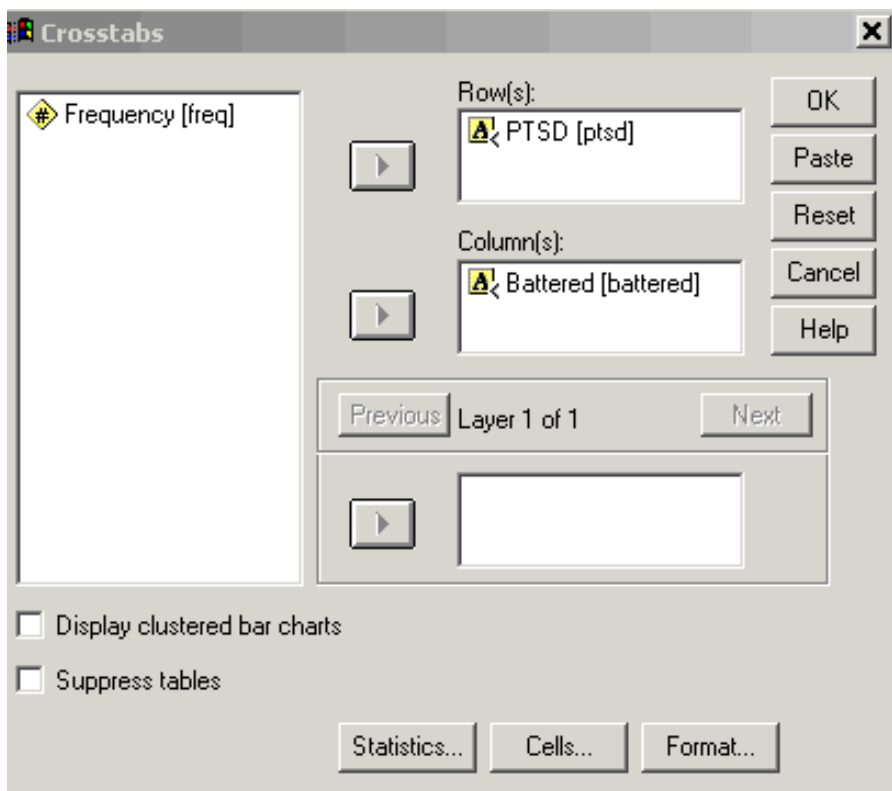

After selecting **Weight Cases**, we get
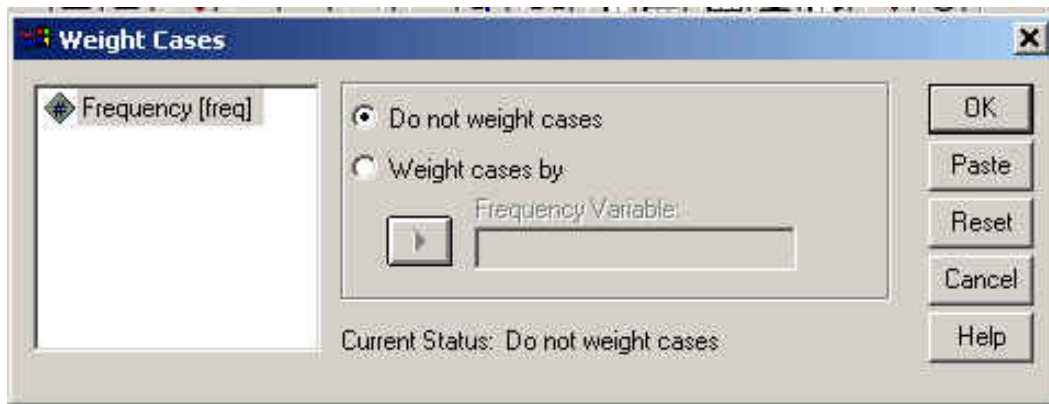
Select the "Weight cases by" button and indicate that Freq is the variable by which we will weight cases. This gives:
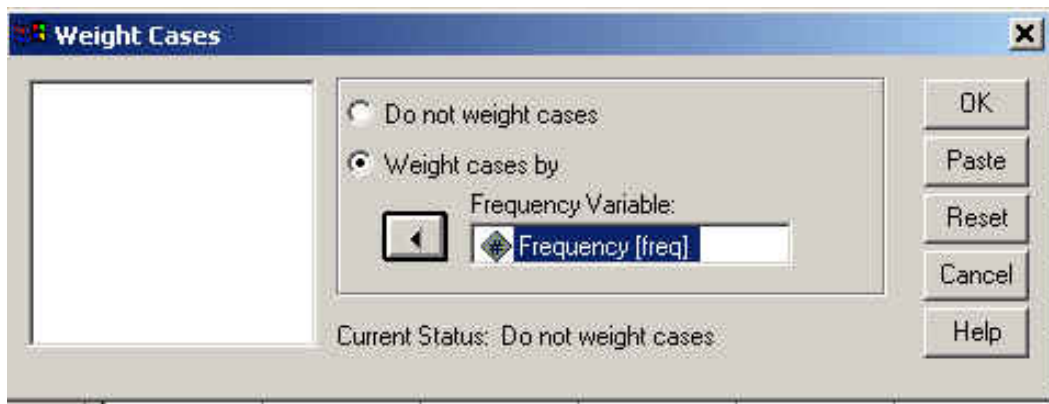
Now we are ready to run our Chi-square. Just use the standard **Analyze/Descriptive/Crosstabs** command, and be sure to click on the Statistics button and tell it to compute chi-square. The dialog box will look like this:

Notice that this looks just like any other chi-square for a contingency table. There is no indication that we have weighted cases. We just have to remember that we did that.

Supposed we made a mistake.

Suppose that we had good intentions, and were going to use the Weight Cases command, but somehow we were distracted. Our data file only has 4 lines of data, and without the weighted cases command, we would get:

PTSD * Battered Crosstabulation

Count

		Battered		Total
		No	Yes	
PTSD	No	1	1	2
	Yes	1	1	2
Total		2	2	4

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.000b	1	1.000		
Continuity Correction a	.000	1	1.000		
Likelihood Ratio	.000	1	1.000		
Fisher's Exact Test				1.000	.833
N of Valid Cases	4				

a. Computed only for a 2x2 table

b. 4 cells (100.0%) have expected count less than 5. The minimum expected count is 1.00.

Ooops!! I hope that anyone looking at such output would realize that there was a serious problem, even if they couldn't identify what it is. The problem is that each cell was treated equally (and counted once), rather than being weighted by Freq. We just need to go back to the main menu and tell it to weight cases.

A Word of Caution

After using the Weight Cases command to run any analysis, be sure that you turn if off before you go on to other analyses. You only want to do weighting when weighting is important, but it is very easy to forget that you have turned it on, and just fail to turn of off when you no longer need it. This is equally true when you use other commands in that menu, such as Select Cases or Split File.

I hope that this clarifies how to use the weighted cases option to run chi-square on a contingency table. Let me know if there is confusion.

[Return to Dave Howell's Statistical Home Page](#)

Send mail to: [David.Howell@uvm.edu)](#)

Last revised: 11/14/2001