

11. Chi Square

Objectives

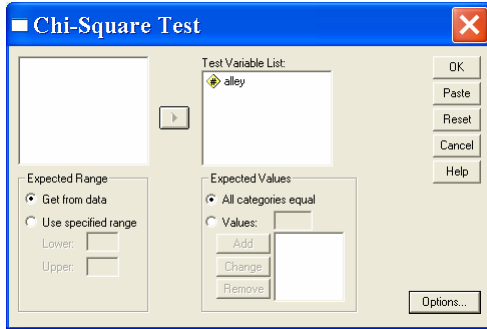
- ◆ Calculate goodness of fit Chi Square
- ◆ Calculate Chi Square for contingency tables
- ◆ Calculate effect size
- ◆ Save data entry time by weighting cases

A Chi Square is used to analyze categorical data. It compares observed frequencies to expected or predicted frequencies. We will examine simple goodness of fit Chi Squares that involved only one variable and more complicated contingency tables that include 2 or more variables. Each type is programmed through different menu options. Let's start with goodness of fit.

Goodness of Fit Chi Square All Categories Equal

Let's begin by using the example in Section 19.1 of the text. We want to test the null hypothesis that rats are equally likely to choose Alley A and Alley B when trying to escape.

- ✓ **Open** *alley chosen.sav*.
- ✓ Select **Analyze/Nonparametric Tests/Chi Square**.



✓ Select alley as the **Test Variable**.

Under **Expected Values**, All categories equal is the default. This is what we want since our null hypothesis is that each alley is equally

Chi-Square Test

Frequencies

ALLEY

	Observed N	Expected N	Residual
A	9	16.0	-7.0
B	23	16.0	7.0
Total	32		

Test Statistics

	ALLEY
Chi-Square ^a	6.125
df	1
Asymp. Sig.	.013

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 16.0.

As you can see, the expected values were 16 each, just as we expected. Now, compare this Chi Square to the value computed in Section 19.1 of the textbook. Once again, they are in agreement.

Goodness of Fit Chi Square Categories Unequal

Now, let's try an example where the expected values are not equal across categories. The difference is we have to specify the expected proportions. This example is based on Exercise 19.3 in the textbook, but the numbers in the data set are slightly different. In the exercise, Howell discusses his theory that when asked to sort one-sentence characteristics like "I eat too fast" into piles ranging from "not at all like me" to

“very much like me,” the percentage of items placed in each pile will be approximately 10%, 20%, 40%, 20%, and 10%. In our data set, the frequencies are 7, 11, 21, 7, and 4 respectively.

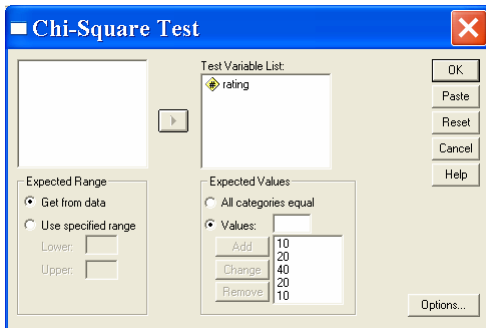
✓ **Open** *unequal categories.sav*. There is no need to save *alley chosen.sav* since we did not change the data file in anyway.

✓ Select **Analyze/Nonparametric Statistics/Chi Square**.

✓ Select rating as the **Test Variable**.

Under **Expected Values**, select

Values. Now, we have to type in the expected *proportion* of cases that should fit each category. These must be specified in order to match the ascending numeric order of the



Chi-Square Test

Frequencies

RATING

	Observed N	Expected N	Residual
not at all like me	7	5.0	2.0
somewhat unlike me	11	10.0	1.0
neither like me or unlike me	21	20.0	1.0
somewhat like me	7	10.0	-3.0
very much like me	4	5.0	-1.0
Total	50		

Test Statistics

	RATING
Chi-Square ^a	2.050
df	4
Asymp. Sig.	.727

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 5.0.

As you can see, SPSS calculated the expected values based on the proportions that we indicated-check the math if you would like. In this case, the fact that the Chi Square is not significant supports the hypothesis. The observed frequencies of ratings fit with the predicted frequencies.

Chi Square for Contingency Tables

Let's use the example illustrated in Section 19.2 of the textbook. We want to examine the hypothesis that perceived fault of a rape victim is related to whether or not the defendant is found guilty.

As you can see the data are nicely displayed in a table in the textbook. This is a great opportunity to show you a nifty trick called weighting cases. Essentially, this will allow us to get SPSS to analyze the data presented in the textbook, without having to enter 358 pieces of data!

✓ Select **File/New/Data**.

✓ In Variable View, create two variables. **Name** one fault and specify the **Values** such that 1 = low fault and 2 = high fault. **Name** the other variable verdict and specify the **Values** such that 1= guilty and 2 = not guilty. Then return to the Data View.

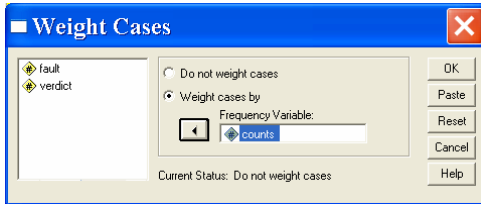
✓ There are four possible combinations of the two variables, as illustrated in the textbook. They are low fault/guilty, low fault/not guilty, high fault/guilty, and high fault/not guilty. So, enter 1,1,2, 2 under fault and 1, 2, 1, 2 under verdict, in the first four rows. A sample follows.

	fault	verdict
1	1.00	1.00
2	1.00	2.00
3	2.00	1.00
4	2.00	2.00

✓ Now, we need to indicate the frequencies for each combination. Create a new variable and **Name** it counts. Then enter the appropriate frequencies for each combination. A sample follows-this time, I will display it with the variable labels on so it is similar to the table in the textbook.

	fault	verdict	counts
1	low fault	guilty	153.00
2	low fault	not guilty	24.00
3	high fault	guilty	105.00
4	high fault	not guilty	76.00

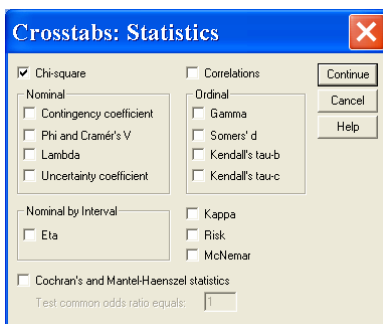
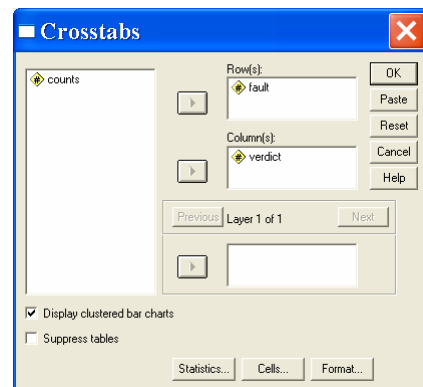
- ✓ Select **Data/Weight Cases**.



- ✓ Select **Weight cases by** and select counts as the **Frequency Variable**. Click **Ok**. Until we

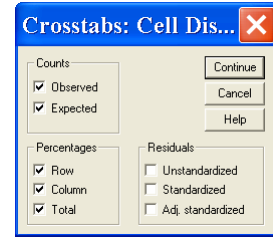
- ✓ Select **Analyze/Descriptive Statistics/Crosstabs**.

- ✓ To be consistent with the presentation in the textbook, select fault for **Rows** and verdict for **Columns**. Select **Display clustered bar charts** to help us visualize the data. Click on



- ✓ Select **Chi-square**, and then click **Continue**. Under **Nominal**, select **Phi and Cramer's V** as well so we can get a measure of effect

✓ Under **Count**, select **Observed** and **Expected**. Under **Percentages**, select **Row**, **Column**, and **Total**. Then click **Continue**. In the main dialog box, click **Ok**. The output follows.



Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
FAULT * VERDICT	358	100.0%	0	.0%	358	100.0%

FAULT * VERDICT Crosstabulation

			VERDICT		Total
			guilty	not guilty	
FAULT	low fault	Count	153	24	177
		Expected Count	127.6	49.4	177.0
		% within FAULT	86.4%	13.6%	100.0%
		% within VERDICT	59.3%	24.0%	49.4%
		% of Total	42.7%	6.7%	49.4%
	high fault	Count	105	76	181
		Expected Count	130.4	50.6	181.0
		% within FAULT	58.0%	42.0%	100.0%
		% within VERDICT	40.7%	76.0%	50.6%
		% of Total	29.3%	21.2%	50.6%
Total	Count	258	100	358	
	Expected Count	258.0	100.0	358.0	
	% within FAULT	72.1%	27.9%	100.0%	
	% within VERDICT	100.0%	100.0%	100.0%	
	% of Total	72.1%	27.9%	100.0%	

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	35.930 ^a	1	.000		
Continuity Correction ^a	34.532	1	.000		
Likelihood Ratio	37.351	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	35.830	1	.000		
N of Valid Cases	358				

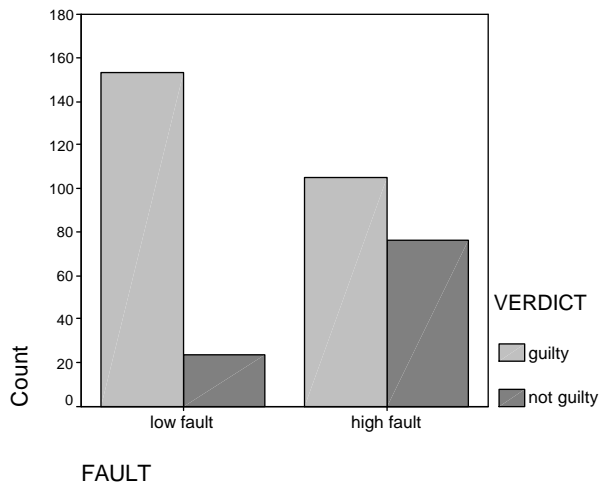
a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 49.44.

Symmetric Measures

		Value	Approx. Sig.
Nominal by	Phi	.317	.000
Nominal	Cramer's V	.317	.000
N of Valid Cases		358	

- a. Not assuming the null hypothesis.
- b. Using the asymptotic standard error assuming the null hypothesis.



The first thing to notice is that in the table named **Case Processing Summary**, the N is 358. This demonstrates that weighting cases worked as we intended! Now compare the Expected Counts to the values in the text. Finally, compare the Chi Square values. We are interested in the Pearson Chi Square because it was calculated the same way as the one in the textbook. Once again, the results are consistent with the textbook.

In this chapter you learned to use SPSS to calculate Goodness of Fit tests with and without equal frequencies. You also learned to calculate Chi Square for contingency

tables, and learned a trick to reduce data entry by weighting cases. Complete the following exercises to help you become familiar with these commands.

Exercises

1. Using *alley chosen.sav*, use a Goodness of Fit Chi Square to test the hypothesis that rats are more likely than chance to choose Alley D. Compare your results to Section 19.1, Extension of the Multicategory Case, in the textbook.
2. Solve Exercise 19.3 from the textbook using SPSS. Create the data file yourself.
3. Create your own data file to represent the observed data presented in the textbook in Table 19.2 using Weight Cases.
4. Using the data file you created in Exercise 3, calculate a Chi Square using crosstabs to examine the hypothesis that the number of bystanders is related to seeking assistance. Be sure to calculate Cramer's Phi. Compare your results to the textbook.