

## SPSS PC Version 10: Crosstabs with Chi-Square Test<sup>1</sup>

The following uses a set of variables from the "1995 National Survey of Family Growth" to demonstrate how to use some procedures available in SPSS PC Version 10.

Chi-square (pronounced 'ky,' not like the trendy drink) is used to test for the significance of relationships between variables cross-classified in a bivariate table. The chi-square test results in a chi-square statistic that tells us the degree to which the conditional distributions (the distribution of the dependent variable across different values of the independent variable) differ from what we would expect under the assumption of "statistical independence." In other words, as in other hypothesis tests, we are setting up a null hypothesis and trying to reject it. In this case, the null hypothesis says that there is no relationship between the variables in our bivariate table (i.e., that they are statistically independent) and that any difference between the conditional distributions that we see is actually just due to random sampling error. If we reject the null hypothesis, we will lend support to the research hypothesis that there is a real relationship between the variables in the population from which the sample is drawn.

### **Using SPSS to conduct a Chi-Square Test of Significance ( $X^2$ ) :**

- To conduct a chi-square test, you need two variables that can reasonably be handled in a bivariate table (i.e., with a limited number of categories). In SPSS choose *Analyze, Descriptive Statistics..., Crosstabs*. This will open the dialog box in which you choose the two variables to be included in the bivariate table and chi-square test. Remember that the independent variable should occupy the columns and the dependent variable the rows of your table. Choose these by highlighting variables from your list and clicking the appropriate right-pointing arrows next to the "Column(s)" and "Row(s)" boxes. Once you've selected your variables, request a chi-square statistic for the table(s) by clicking on the "Statistics" box and then marking the selection box next to "Chi-square" in the upper left hand corner of the resulting dialog box. Hit "Continue" to return to the original dialog box.
- Now you may want to request column percentages so you can get an idea of the strength of the differences between the conditional distributions (remember that the chi-square test just tells us whether or not the relationship is present; a large chi-square does not indicate a stronger relationship because it is greatly affected by the size of the sample). You can do this by clicking on "Cells" and choosing "Column" under the "Percentages" heading. While you are here, you can also ask SPSS to report the expected frequencies for each cell by clicking on "Expected" under the "Counts" heading. These will help you see exactly where the obtained value of the chi-square statistic comes from. Return to the main dialog box by hitting the "Continue" button. After you've made all your selections, hit "OK."

The syntax for this would look like:

```
CROSSTABS
  /TABLES=cohever BY region
  /FORMAT= AVALUE TABLES
  /STATISTIC=CHISQ
  /CELLS= COUNT EXPECTED COLUMN TOTAL .
```

- Here's an example of the output you will receive, examining the relationship between region of residence (the independent variable) and our indicator of whether the NSFG respondent has ever cohabited (the

---

<sup>1</sup>Prepared by Kyle Crowder of the Sociology Department of Western Washington University, and modified by Patty Glynn, University of Washington. 12/12/2000 C:\all\help\helpnew\chissps.wpd

dependent variable):

**COHEVER Whether R has Ever Cohabited \* REGION Region Where R Lives Crosstabulation**

			REGION Region Where R Lives				Total
			1.00 Northeast	2.00 Midwest	3.00 South	4.00 West	
COHEVER Whether R has Ever Cohabited	1.00	Count	853	1090	1496	1183	4622
		Expected Count	864.6	1104.9	1598.3	1054.2	4622.0
		% within REGION Region Where R Lives	42.0%	42.0%	39.9%	47.8%	42.6%
	2.00	Count	1176	1503	2255	1291	6225
		Expected Count	1164.4	1488.1	2152.7	1419.8	6225.0
		% within REGION Region Where R Lives	58.0%	58.0%	60.1%	52.2%	57.4%
Total	Count	2029	2593	3751	2474	10847	
	Expected Count	2029.0	2593.0	3751.0	2474.0	10847.0	
	% within REGION	100.0%	100.0%	100.0%	100.0%	100.0%	
	Region Where R Lives						

**Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	39.461 <sup>a</sup>	3	.000
Likelihood Ratio	39.297	3	.000
Linear-by-Linear Association	9.835	1	.002
N of Valid Cases	10847		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 864.57.

- Notice that SPSS produces the familiar bivariate table, including the expected cell frequencies and column percentages (if you requested them). The next box contains the chi-square score for the table (labeled Pearson chi-square), the table's degrees of freedom, and the p-value associated with the obtained chi-square score. Also note that SPSS reports how many of the cells in the table have an expected frequency below 5. Remember, this is an important diagnostic tool because chi-square becomes unreliable when your table has cells with expected frequencies below 5.
- As always, you can reach a conclusion about your hypothesis by comparing the obtained Pearson chi-square to the critical value of chi-square, or by comparing the reported p-value to your chosen alpha level. In our example, either method would lead us to reject the null hypothesis at all conventional levels of confidence. Right?