Chapter Five: Cross Tabulations

In this chapter, we'll look at how SPSS for Windows can be used to create contingency tables, sometimes called cross tabulations (or crosstabs), bivariate, or two-variable tables. A contingency table helps us look at whether the value of one variable is associated with, or "contingent" upon, that of another. It is most useful when each variable contains only a few categories. Usually, though not always, such variables will be nominal or ordinal.

Some techniques for examining relationships among interval or ratio variables are presented in later chapters.

To make it easier to follow the instructions in this chapter, we recommend that you set certain options in SPSS in the same way that we have. First, click on Edit in the menu bar, then on Options, and General. Under Variable Lists, click on Display names, and Alphabetical. These choices will ensure that the variables in dialog boxes will look like they do in our examples (see Figure 5-1).
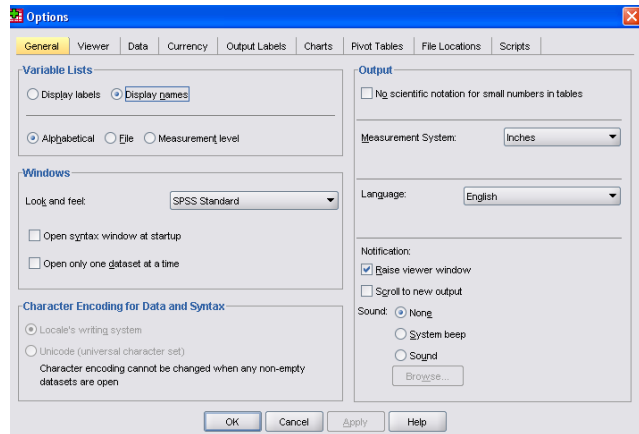
Figure 5-1

Now click on Pivot Tables in the tabs. SPSS offers a number of different "looks" for contingency tables. You might want to experiment with the different choices. For now, however, click on Academic under TableLook (see Figure 5-2). Then click on OK.
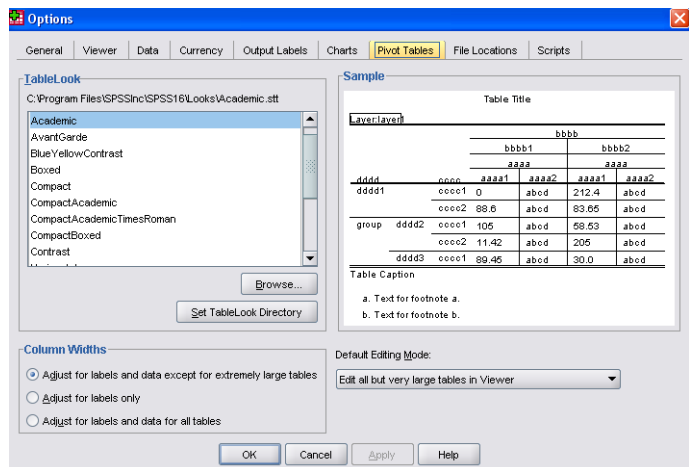
Figure 5-2

To illustrate the Crosstabs technique, we'll use the General Social Survey subset (GSS06A). Open this file following the instructions in Chapter 1 under "Getting a Data File."

Crosstabs are particularly useful in hypothesis testing such as the question, "Let's see if there is any difference between men and women in their attitudes towards abortion." To create a contingency table (crosstabs), from the menu, click on Analyze, Descriptive Statistics, and Crosstabs. This will open the dialog box shown in Figure 5-3.

You then choose the row (usually the dependent) variable and column (usually the independent) variable.[9] In Appendix A, you will see that there are seven variables that deal with opinions about abortion. Let's choose *abhlth* (abortion if the woman health is endangered) for our row variable and *sex* (respondent's sex) for column variable. To do this, select the variable you want from the list and click on it to highlight it, then use the arrow keys to the right of the List box to move the variable into either the Row or the Column box (for now, ignore the Bottom box – more about it in Chapter 8). If you've done everything correctly, your screen will look like Figure 5-4, but don't click OK yet!

In the buttons within the Crosstabs dialog box, click on Cells. Here you have a number of choices for the information you would like to have in each cell of your table. The Observed box should already be selected—it shows the actual number of cases in each cell. You will also want to see percentages as well as raw numbers so that you can easily compare groupings of different sizes. You should always make sure that each category of the independent variable totals 100%; our general rule is to have the dependent variables be the rows and the independent variables the columns. So choose Columns for the percentages as in Figure 5-5.

Now click on Continue to get back to the Crosstabs dialog box. Once you are back there, click OK. SPSS will now open the Output Viewer window, which will show you your table (see Figure 5-6).
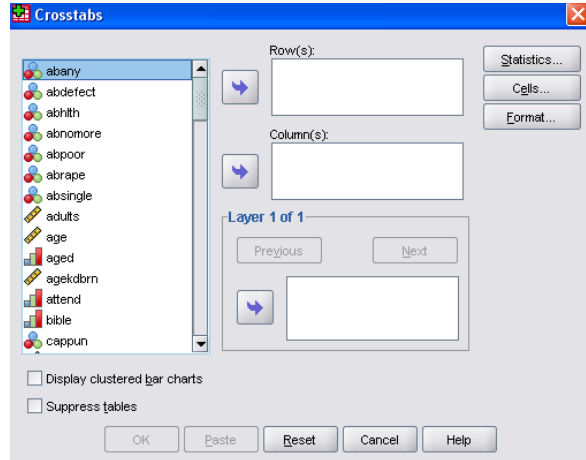

Figure 5-3
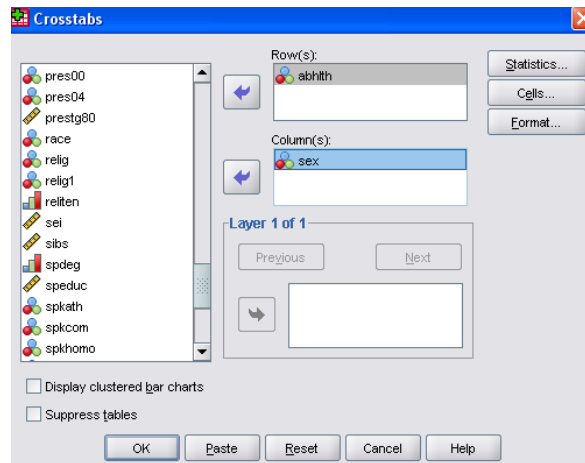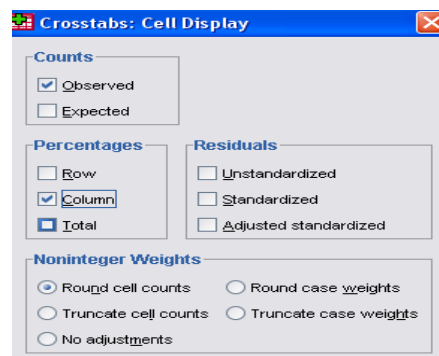

Figure 5-4

Figure 5-5

---

[9] The independent variable is the causal variable.

The Case Processing Summary shows the Valid, Missing, and Total cases. The high percent of missing cases here reflects the people who were not asked this particular question in the survey. The valid number of cases is used in the table.

Figure 5-6

The Crosstabs shows the 1916 valid cases arranged in a table that shows what percent of men and women said either Yes or No to the *abhlth* question. Note that 87.4% of the men and 85.9% of the women said Yes, a percentage point difference of only 1.5.

Figure 5-7

Your initial conclusion here might be that on abortion issues, there's virtually no difference between men and women in their responses. Is this correct or did you stop your analysis a little too soon? Let's look at a different abortion question. Repeat the steps above, but use *abnomore* as your dependent variable this time. Your results should look like Figure 5-7.

Now we see that 43.9% of the men and 39.3% of the women said Yes to "Abortion if a woman is married and wants no more children." When we compare Figure 5-6 with Figure 5-7, we see there is a large difference between total Yes answers (87% compared with 41%), which indicates that abortion as an issue needs to be broken down into specific conditions if you want to study it in depth. We also see that there is now a larger difference between men and women on this particular question. But is it a significant difference? To answer this we will need to do some statistical analysis.

For our next cross tabulation, again go to the menu and choose Analyze, Descriptive Statistics, and Crosstabs. In the Crosstabs dialog box place *abnomore* as the row variable and *sex* as the column variable. Now click on the Statistics button, then Chi-Square to obtain a measure of statistical significance, and on Phi and Cramer's V, which are measures of the strength of association between two variables when one or both are at the nominal level of measurement. Phi is appropriate for tables with two rows and two columns, while Cramer's V is appropriate otherwise. Your dialog box should look like Figure 5-8.
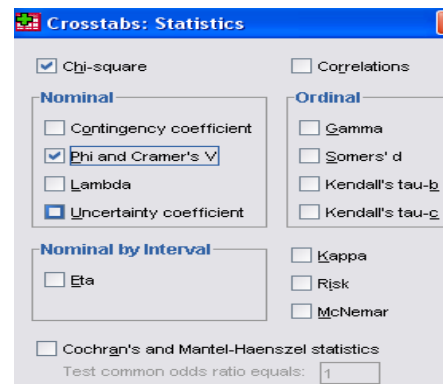

Figure 5-8

Click on **Continue**, then **OK**. The table in Figure 5-7 reappears, but with some additional information (you might have to scroll down to see it)—look for "Chi-Square Tests" (Figure 5-9).

The Pearson Chi Square test indicates that the relationship is statistically significant. It would occur by chance approximately 43 times out of 1000. The Cramer's V of .046 in Figure 5-10 (Symmetric Measures) indicates that there is virtually no relationship.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4.089ª	1	.043		
Continuity Correction[b]	3.903	1	.048		
Likelihood Ratio	4.086	1	.043		
Fisher's Exact Test				.045	.024
Linear-by-Linear Association	4.087	1	.043		
N of Valid Cases	1925				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 357.27.

b. Computed only for a 2x2 table

Figure 5-9

Let's look at a somewhat different table. For many years, scholars have observed that in the U.S., compared to other industrialized countries, social class has relatively little impact on political attitudes and behavior. To find this out, click on **Analyze**, **Descriptive Statistics**, and **Crosstabs**. If the variables you used before are still there, click on the **Reset** button, then move *polviews* to the Row box and *inco6rec* to the Column box. Since both of these variables are ordinal, we'll want to obtain different statistics to measure their relationship. Click on **Statistics** and then on **Kendall's tau c**. (Tau c is a measure of association that is appropriate when both variables are ordinal and do not have the same number of categories.) Then click on **Continue** and then on **Cells** and then on **Column percents**. Now click on **Continue** and then click on **OK**. What do the results show?

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.046	.043
	Cramer's V	.046	.043
	N of Valid Cases	1925	

Figure 5-10

Chapter Five Exercises

1. Suppose we measure class, not by income, but by what people perceive their social class to be (using the variable named *class*). How closely is this measure related to a person's self identified political views (*polviews*)? Note: before running this crosstab, look at the frequency distribution for *class*. (See Chapter 4 on univariate statistics.) You may want to recode this variable before proceeding. (See Chapter 3 on transforming data.)

2. Consult the codebook in Appendix A describing this dataset. Other than income and self perceived class, what background variables (such as region of country, age, marital status, religion, sex, race, or education) might help explain a person's political views? (Here as well, you may need to recode some variables before proceeding.)

3. Is trust related to race? Run crosstabs for *trust* (Can people be trusted?) with *race* and see what you find.

4. Is ideology a general characteristic or is it issue specific? That is, are people who are liberal (or conservative) on one issue (such as capital punishment) also liberal (or conservative) on other issues (such as gun control or legalizing marijuana)?