Chapter 8
Crosstabulation and χ^2 Analyses

THE PURPOSE of crosstabulation is to show in tabular format the relationship between two or more categorical variables. Categorical variables include those in which distinct categories exist such as gender (female, male), ethnicity (Asian, White, Hispanic), place of residence (urban, suburban, rural), responses (yes, no), grades (A, B, C, D, F), and many more. Crosstabulation can be used with continuous data only if such data are divided into separate categories, such as age (0–19 years, 20–39 years, 40–59 years, 60–79 years, 80–99 years), total points (0–99, 100–149, 150–199, 200–250), and so on. While it is acceptable to perform crosstabulation with continuous data that has been categorized, it is rare to perform chi-square analyses with continuous data because a great deal of useful information about the distribution is lost by the process of categorization. For instance, in the total points distribution (above), two persons who scored 99 and 100 points, respectively, would be in the first and second categories and would be considered identical to two persons who scored 0 and 149 points, respectively. Nonetheless, crosstabulation with continuous data is often used for purposes of data description and display. The SPSS command **Crosstabs** and the subcommands **Cells** and **Statistics** are used to access all necessary information about comparisons of frequency data.

8.1 Crosstabulation

While the **Frequencies** command can tell us there are 5 Natives, 20 Asians, 24 Blacks, 45 Whites, and 11 Hispanics (and that there are 64 females and 41 males) in our **grades.sav** file, it cannot give us the number of female Asians or male Whites. This is the function of the **Crosstabs** command. It would be appropriate to "cross" two variables (**ethnic** by **gender**) to answer the questions posed above. This would produce a table of 10 different cells with associated frequencies inserted in each cell by crossing two (2) levels of **gender** with five (5) levels of **ethnic**. It is possible to cross three or more variables, although only with a very large data set would a researcher be likely to perform a crosstabulation with three variables because there would be many low-count and empty cells if the number of subjects was not sufficient. For the present sample, an **ethnic** by **gender** by **grade** crosstabulation would probably not be recommended. This procedure would create a 5 (**ethnic**) × 2 (**gender**) × 5 (**grade**) display of frequencies—a total of 50 cells to be filled with only 105 subjects. A large number of low-count or empty cells would be guaranteed. If such a crosstabulation were created with a larger *N*, SPSS would produce five different 5 × 2 tables to display these data.

8.2 Chi-Square (χ^2) Tests of Independence

In addition to frequencies (or the *observed values*) within each cell, SPSS can also compute the expected value for each cell. *Expected value* is based on the assumption that the two variables are independent of each other. A simple example demonstrates the derivation of expected value. Suppose there is a group of 100 persons in a room and that 30 are male and 70 are female. If there are 10 Asians in the group, it would be anticipated

(expected)—if the two variables are independent of each other—that among the 10 Asians there would be 3 males and 7 females (the same proportion as is observed in the entire group). However, with the same group of 100, if 10 of them were football players we would *not* expect 3 male football players and 7 female football players. In American society, most football players are male, and the two categories (gender and football players) are *not* independent of each other. If there were no additional information given, it would be expected that all 10 of the players would be male. The purpose of a chi-square test of independence is to determine whether the observed values for the cells deviate significantly from the corresponding expected values for those cells.

You may wish to test levels of a *single* variable. For instance, if you wish to determine whether actual values differ significantly from expected values for the five levels of ethnicity included in the **ethnic** variable you need to use a different chi-square test. The actual test you use is described on pages 225–227 of the Nonparametric Procedures chapter. However, with only one variable SPSS cannot compute expected values. You need to specify which expected values you wish. In the **ethnic** variable in the **grades.sav** file, among the 105 subjects there are actually 5 Natives, 20 Asians, 24 Blacks, 45 Caucasians, and 11 Hispanics. The default expected value (explained on page 226) is equal numbers in each group—that is 105/5, or 21 in each of the five groups. You may specify different expected values if you wish. For instance you may want to see if your sample differs significantly from the actual ethnic breakdown of your community. Then you enter numbers that represent the proportion of each ethnic group as your expected values.

The chi-square statistic is computed by summing the squared deviations [observed value (f_o) minus expected value (f_e)] divided by the expected value for each cell:

$$\chi^2 = \Sigma[(f_o - f_e)^2/f_e]$$

For every *inferential* statistic (this is the first of many in this book!), there are three questions that you need to ask as a researcher. The first question is, "Are you fairly certain that the difference tested by this statistic is real instead of random?" In this case, if there is a large discrepancy between the observed values and the expected values, the χ^2 statistic would be large, suggesting a significant difference between observed and expected values. ("Significant difference" is statistics-speak for "fairly certain that this difference is real.") Along with this statistic, a probability value is computed. The value of p is the probability that the difference between observed and expected values tested by the χ^2 statistic is due to chance. With $p < .05$, it is commonly accepted that the observed values differ significantly from the expected values and that the two variables are *NOT* independent of each other. More complete descriptions and definitions are included in the Output section of this chapter.

The second question that you need to ask for every inferential statistic is, "How big is the difference tested by the statistic?" Depending on what your data look like, it is possible to have a small p value (and be very confident that the difference is real) even if the difference is very small, or to have a large p value (and not be confident that the difference is real) even if the difference in your data appears to be large. The χ^2 statistic itself isn't a good measure of how big the difference is between the observed and expected values because it is largely dependent on the number of dimensions and sample size. Thus comparisons of one chi-square value with another are often misleading.

To control for this difficulty, Pearson suggested the phi (φ) statistic, which divides the chi-square value by N and then takes the positive square root of the result. The purpose was to standardize a measure of association (also known as *effect size*) to values between 0 and 1 (with 0 indicating completely independent variables and a value close to 1 indicating a strong association between variables). However, if one of the dimensions of the crosstabulation is larger than 2, φ may attain a value larger than 1.0. To control for this, Cramér's V was introduced (the positive square root of $\chi^2/[N(k-1)]$, where k is the smaller of the number of rows and columns). This effect size measure *does* vary between 0 and 1.0 and is a commonly used measure of the strength of association between variables in a χ^2 analysis.

The final question that you need to ask for every inferential statistic is, "Is this finding important?" The answer to that question will depend on whether you're confident that the finding isn't the result of randomness, how big the effect is, and how much you care. Example: you may not care about a small decrease in the likelihood of blinking because of taking a certain medication, but you may care a lot about a small decrease in the likelihood of dying because of the same medication.

The file we use to illustrate **Crosstabs** is our example described in the first chapter. The data file is called **grades.sav** and has an $N = 105$. This analysis creates crosstabulations and calculates chi-square statistics for **gender** by **ethnic**.

8.3 Step by Step
8.3.1 Crosstabulation and Chi-Square Tests of Independence

To access the initial SPSS screen from the Windows display, perform the following sequence of steps:

In Screen	Do This	Step 1
2.1	✳ 🪟 → ✳ All Programs ▷ → ✳ 📁 IBM SPSS Statistics → ✳ IBM SPSS Statistics	Front1

Mac users: To access the initial SPSS screen, successively click the following icons:

In Screen	Do This	Step 1a
2.1	✳ 🚀 → ✳ ⬛ → ✳ Σα	Front1

After clicking the SPSS program icon, Screen 1 appears on the monitor.

	Step 2
Create and name a data file or edit (if necessary) an already existing file (see Chapter 3)	

*Screens 1 and 2 (displayed on the inside front cover) allow you to access the data file used in conducting the analysis of interest. The following sequence accesses the **grades.sav** file for further analyses:*

In Screen	Do This	Step 3
Front1	✳ **File** → **Open** → ✳ **Data** [*or* ✳ 📁]	
Front2	type grades.sav → ✳ **Open** [*or* ✳✳ grades.sav]	Data

Whether first entering SPSS or returning from earlier operations the standard menu of commands across the top is required. As long as it is visible you may perform any analyses. It is not necessary for the data window to be visible.

After completion of Step 3 a screen with the desired menu bar appears. When you click a command (from the menu bar), a series of options will appear (usually) below the selected command. With each new set of options, click the desired item. The sequence to access chi-square statistics begins at any screen with the menu of commands visible:

In Screen	Do This	Step 4
Menu	✳ **Analyze** → ✳ **Descriptive Statistics** → ✳ **Crosstabs**	8.1

A new window now appears (Screen 8.1, below) that provides the framework for conducting a crosstabs analysis. The procedure is to click on the desired variable in the list to the left (**gender** in this example) and then click the top right-arrow (➡) to select gender as the row variable. Then click a second variable (**ethnic** in this case) and click the middle right-arrow (to select ethnicity as the column variable). That is all that is required to create a crosstabulation of two variables. This will create a 2 (**gender**) by 5 (**ethnic**) table that contains 10 cells. Each of the 10 cells will indicate the number of subjects in each category—4 Native females, 13 Asian females, and so forth for the other 8 cells in the table.

The lowest box in the window allows for crosstabulation of three or more variables. If, for instance, we wanted to find the gender by ethnicity breakdown for the three sections, you would click the **section** variable in the list of variables and then click the lowest of the three right arrows. This would result in three tables: a gender by ethnicity breakdown for the first section, a gender by ethnicity breakdown for the second section, and a gender by ethnicity breakdown for the third section. The **Previous** and **Next** to the left and right of **Layer 1 of 1** are used if you wanted a gender by ethnicity analysis for more than one variable. For instance, if you wanted this breakdown for both **section** and **year** (year in school), you would click **section**, click the lowest right arrow, click **Next**, click **year**, and then click the lowest right arrow again. This produces three 2 × 5 tables for **section** and four 2 × 5 tables for **year**. The sequence (below) shows how to create a **gender** by **ethnic** by **section** crosstabulation.

Screen 8.1 The Crosstabs Window
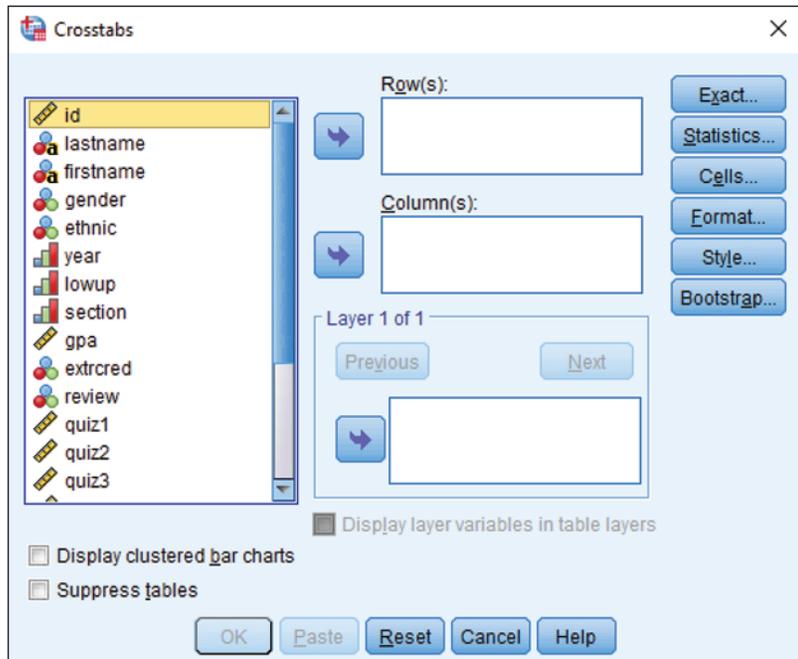
The following sequence begins from Screen 8.1. To arrive at this screen, perform whichever of Steps 1–4 (page 123) are necessary.

In Screen	Do This	Step 5
8.1	🔆 **gender** → 🔆 *upper* ➡ → 🔆 **ethnic** → 🔆 *middle* ➡	
	→ 🔆 **section** → 🔆 *lowest* ➡ → 🔆 **OK**	Back1

For a **gender** by **ethnic** crosstabulation, omit the two clicks that involve the **section** variable.

It is rare for a researcher to want to compute only cell frequencies. In addition to frequencies, it is possible to include within each cell a number of additional options. Those most frequently used are listed below with a brief definition of each. When you press the **Cells** button (Screen 8.1), a new screen appears (Screen 8.2, below) that allows you to select a number of options.

Screen 8.2 The Crosstabs: Cell Display Window
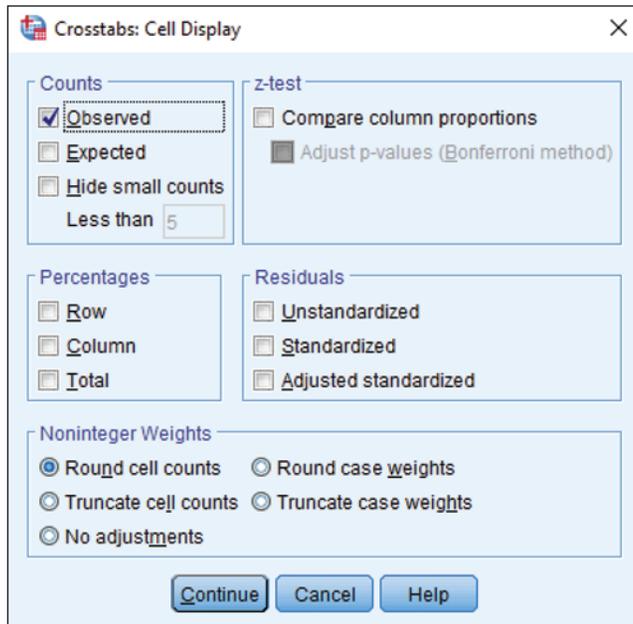
The **Observed** count is selected by default. The **Expected** count (more frequently referred to as the *expected value*) is in most cases also desired. Inclusion of other values depends on the preference of the researcher.

- **Observed Count:**   The actual number of subjects or cases within each cell.
- **Expected Count:**   The expected value for each cell (see page 121)
- **Row Percentages:**   The percent of values in each cell for that row.
- **Column Percentages:**   The percent of values in each cell for that column.
- **Total Percentages:**   The percent of values in each cell for the whole table.
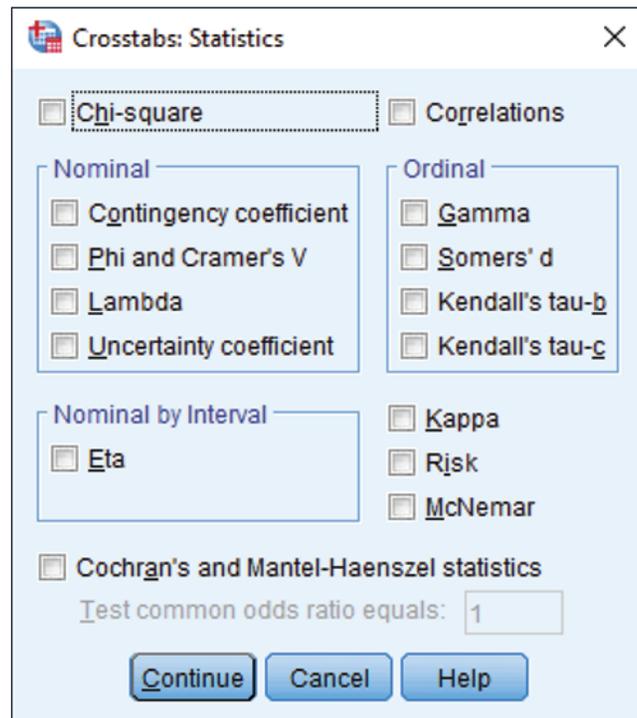- **Unstandardized Residuals:** Observed value minus expected value.

To create a **gender** *by* **ethnic** *crosstabulation that includes observed count, expected count, total percentages, and unstandardized residuals within each cell, perform the following sequence of steps. Begin at Screen 8.1; press* **Reset** *if variables remain from a prior analysis.*

In Screen	Do This	Step 5a
8.1	✳ **gender** → ✳ *upper* 🔷 → ✳ **ethnic** → ✳ *middle* 🔷 → ✳ **Cells**	
8.2	✳ **Expected** → ✳ **Total** → ✳ **Unstandardized** → ✳ **Continue**	
8.1	✳ **OK**	Back1

Note: We don't click on **Observed**, because this option is already selected by default.

Thus far we have only created tabulations of numbers within cells. Usually, along with crosstabulation, a chi-square analysis is conducted. This requires a click on the **Statistics** pushbutton (see Screen 8.1). When this button is clicked, a new window opens (Screen 8.3, below). Many different tests of independence or association are listed here. Only **Chi-square**, and **Phi and Cramer's V** (see page 122) and **Correlations** (explained in Chapter 10) will be considered here. As in the **Cells** window, the procedure is to click in the small box to the left of the desired statistic before returning to the previous screen to conduct the analysis. See the SPSS help files for technical descriptions of the other statistics included in this chart.

Screen 8.3 The Crosstabs: Statistics Window

In the sequence below, we create a **gender** *by* **ethnic** *crosstabulation, request* **Observed** *count,* **Expected count**, *and* **Unstandardized Residuals** *within each cell, and select the* **Chi-square** *and* **Phi and Cramer's V** *as forms of analysis. We don't include correlations because they have no meaning when there is no intrinsic order to the associated variables. The starting point is Screen 8.1. Complete results from this analysis are included in the Output section.*

In Screen	Do This	Step 5b
8.1	✳ **gender** ➝ ✳ *upper* ➡ ➝ ✳ **ethnic** ➝ ✳ *middle* ➡ ➝ ✳ **Cells**	
8.2	✳ **Expected** ➝ ✳ **Unstandardized** ➝ ✳ **Continue**	
8.1	✳ **Statistics**	
8.3	✳ **Chi-square** ➝ ✳ **Phi and Cramer's V** ➝ ✳ **Continue**	
8.1	✳ **OK**	**Back1**

Often we may wish to conduct a crosstabulation and chi-square analysis on a *subset* of a certain variable. For instance, in the **gender** by **ethnic** crosstabulation described earlier, we may wish to delete the "Native" category from the analysis since there are only five of them and earlier analyses have indicated that there is a problem with low count cells. This then creates a 2 (levels of gender) × 4 (levels of ethnicity after excluding the first level) analysis. After you have selected the variables for crosstabulation, have chosen cell values and desired statistics, click on the **Data** command in the Menu Bar at the top of the screen. In the menu of options that opens below, click on **Select Cases**. At this point, a new window will open (Screen 4.8 from Chapter 4). In this window, click the small circle to the right of **If condition is satisfied** (so a black dot fills it), then click the **If** button just below.

A new dialog box again opens (see Screen 4.3, also from Chapter 4) titled **Select Cases: If**. This window provides access to a wide variety of operations described in Chapter 4. For the present we are only concerned with how to select levels 2, 3, 4, and 5 of the **ethnic** variable. First step is to select **ethnic** from the variable list to the left, then click the ⬛ to paste the variable into the "active" box, then click the ⬛ (on the small keyboard below the active box), and then click the 2. You have now indicated that you wish to select all levels of **ethnic** greater or equal to 2. Then click **Continue**. Screen 4.8 will reappear; click **OK**, the original screen will appear (Screen 8.1) and then click **OK**. Your analysis will now be completed with only four levels of ethnicity. The step-by-step sequence follows.

The starting point for this sequence is Screen 8.1. It may be necessary to press **Reset**.

In Screen	Do This	Step 5c
8.1	✳ **gender** ⟶ ✳ *upper* ➡ ⟶ ✳ **ethnic** ⟶ ✳ *middle* ➡ ⟶ ✳ **Cells**	
8.2	✳ **Expected** ⟶ ✳ **Unstandardized** ⟶ ✳ **Continue**	
8.1	✳ **Statistics**	
8.3	✳ **Chi-square** ⟶ ✳ **Phi and Cramer's V** ⟶ ✳ **Continue**	
8.1	✳ **Data** *(at the top of the screen in the menu bar)* ⟶ ✳ **Select Cases**	
4.8	✳ *the* **O** *to the left of* **If condition is satisfied** ⟶ ✳ **If** *bar just beneath*	
4.3	✳ **ethnic** ✳ ➡ ⟶ ✳ **>=** ⟶ ✳ 2 ⟶ ✳ **Continue**	
4.8	✳ **OK**	
8.1	✳ **OK**	Back1

8.4 Weight Cases Procedure: Simplified Data Setup

All of the chi-square analyses assume that the data is set up as described in Chapter 4: Each person has one line, and every line is a case. But sometimes this is inconvenient: If you have thousands of people in your study, and you already know how many people are in each cell, you can no doubt make better use of your time than entering the same information in thousands of rows. In that case, you can use the **Weight Cases** procedure with a simplified data file.

Screen 8.4 The Weight Cases Window
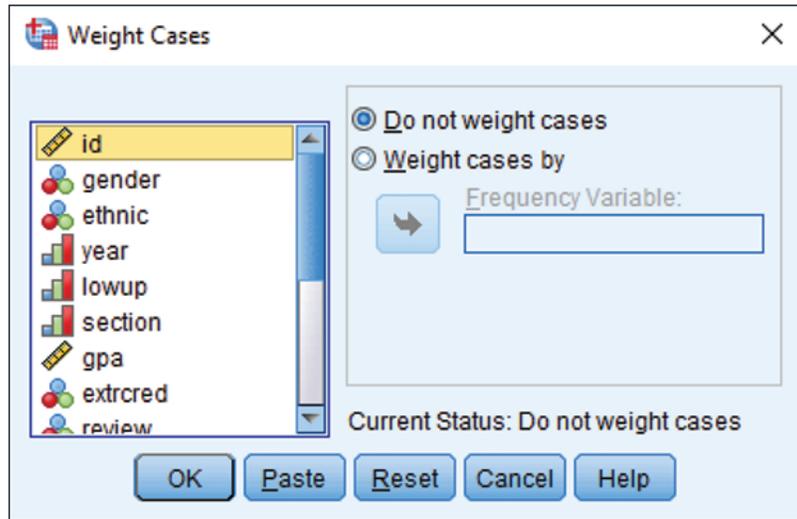
	Gender	Ethnic	Weight
1	Female	Native	4
2	Female	Asian	13
3	Female	Black	14
4	Female	White	26
5	Female	Hispanic	7
6	Male	Native	1
7	Male	Asian	7
8	Male	Black	10
9	Male	White	19
10	Male	Hispanic	4

To do this, you first have to set up your categorical variables (the same way you would in a regular data file). For the example in this chapter, you would set up a variable to code for ethnicity and a variable to code for gender. Then you make an additional variable—you can name it whatever you like, but we'd suggest something like **"Weight"** or **"Case-Count"**—that includes how many cases have that particular combination of categorical variables. For example, in the example used in this chapter, there are 4 Native American females, 13 Asian females, 14 Black females, and so on (see the output below on page 129 for all of the counts). To enter this data, you could set the data file up like this (left):

So instead of one row per case, there is one row per combination of categorical variables. The next step is then to tell SPSS that you want it to pretend that each row is actually there many times—as many times as the weighting variables (the variable **weight** in the chart) indicate. Here is the step-by-step sequence that will accomplish that

In Screen	Do This	Step 5d
8.4	✳ **Data** → ✳ **Weight Cases** → ✳ **Weight cases by** → ✳ **weight** → ✳ ➜ → ✳ **OK**	

Then you are ready to run the Crosstabs analysis as if you had actually entered every category multiple times. If, for instance, you completed the steps shown in Step 5a, despite the tiny file (shown above) the results would be identical to the results if you completed the sequences in Step 5a with the far larger **grades.sav** file.

Upon completion of Step 5, 5a, 5b, or 5c, the output screen will appear (Screen 1, inside back cover). All results from the just-completed analysis are included in the Output Navigator. Make use of the scroll bar arrows (▲▼▶◀) to view the results. Even when viewing output, the standard menu of commands is still listed across the top of the window. Further analyses may be conducted without returning to the data screen.

8.5 Printing Results

Results of the analysis (or analyses) that have just been conducted require a window that displays the standard commands (**File** **Edit** **Data** **Transform** **Analyze . . .**) across the top. A typical print procedure is shown below beginning with the standard output screen (Screen 1, inside back cover).

To print results, from the Output screen perform the following sequence of steps:

In Screen	Do This	Step 6
Back1	*Select desired output or edit (see pages 16–22)* ➝ ✳ **File** ➝ ✳ **Print**	
2.9	*Consider and select desired print options offered, then* ➝ ✳ **OK**	

*To exit you may begin from any screen that shows the **File** command at the top.*

In Screen	Do This	Step 7
Menu	✳ **File** ➝ ✳ **Exit**	2.1

Note: After clicking **Exit**, there will frequently be small windows that appear asking if you wish to save or change anything. Simply click each appropriate response.

8.6 Output

8.6.1 Crosstabulation and Chi-Square (χ^2) Analyses

What follows is partial output from sequence Step 5b, page 126.

			Ethnicity					
			Native	Asian	Black	White	Hispanic	Total
Gender	Female	Count	4	13	14	26	7	64
		Expected Count	3.0	12.2	14.6	27.4	6.7	64.0
		Residual	1.0	.8	−.6	−1.4	.3	
	Male	Count	1	7	10	19	4	41
		Expected Count	2.0	7.8	9.4	17.6	4.3	41.0
		Residual	−1.0	−.8	.6	1.4	−.3	
Total		Count	5	20	24	45	11	105
		Expected Count	5.0	20.0	24.0	45.0	11.0	105.0

	Value	df	Significance
Pearson Chi-Square	1.193[a]	4	.879
Likelihood Ratio	1.268	4	.867
Linear-by-Linear Association	.453	1	.501
Phi	.107	--	.879
Cramér's V	.107	--	.879
N of valid cases	105		

[a]3 cells (30%) have expected count less than 5. The minimum expected count is 1.95.

The first step in interpreting a crosstabulation or chi-square analysis is to observe the actual values and the expected values within each cell. Preliminary observation indicates that observed values and expected values are quite similar. The greatest

discrepancy is for Whites (females: actual count = 26, expected = 27.4; and males: actual count = 19, expected = 17.6). Note also that the residual value (the number beneath the other two) is simply the observed value minus the expected value. Even without looking at the chi-square statistics you would anticipate that observed values and expected values do not differ significantly (that is, gender and ethnicity in this sample are independent of each other). The results support this observation with a low chi-square value (1.193) and a significance greater than .8 (.879). Notice that measures of association are also small and do not approach significance. As suggested in the Step by Step section, low-count cells are a problem; 3 of 10 have an expected value less than 5. The usual response would be to redo the analysis after deleting the "Native" category. To further assist understanding, definitions of output terms follow:

Term	Definition/Description
Count	The top number in each of the 10 cells (4, 13, 14, . . .), indicates the number of subjects in each category.
Expected Count	The second number in each of the 10 cells (3.0, 12.2, 14.6, . . .), indicates the number that would appear there if the two variables were perfectly independent of each other.
Residual	The observed value minus the expected value.
Row Total	The total number of subjects for each row (64 females, 41 males).
Column Total	The total number of subjects in each category for each column (5 Natives, 20 Asians, 24 Blacks, 45 Whites, 11 Hispanics).
Chi Square: Pearson and Likelihood Ratio	Two different methods for computing chi-square statistics. With a large N, these two values will be close to equal. The formula for the Pearson chi-square is $$\chi^2 = \sum [(f_0 - f_e)^2 / f_e]$$
Value	For **PEARSON** and **MAXIMUM LIKELIHOOD** methods, as the test-statistic value gets larger the likelihood that the two variables are *not* independent (e.g., *are* dependent) also increases. The values close to 1 (1.193, 1.268) suggest that gender balance is not dependent on which ethnic groups are involved.
Degrees of Freedom (df)	Degrees of freedom is the number of levels in the first variable minus 1 (2 − 1 = 1) times the number of levels in the second variable minus 1 (5 − 1 = 4); 1 × 4 = 4.
Significance	The likelihood that these results could happen by chance. The large p value here indicates that observed values do *not* differ significantly from expected values.
Linear-by-Linear Association	This statistic tests whether the two variables correlate with each other. This measure is often meaningless because there is no logical or numeric relation to the order of the variables. For instance, there is no logical order (from a low value to a high value) of ethnicity. Therefore, the correlation between gender and ethnicity is meaningless. If, however, the second variable were income, ordered from low to high, a valid correlation results.
Minimum Expected Count	The minimum expected count is for the first cell of the second row (male Natives). The expected value there is rounded off to the nearest 10th (2.0). The value accurate to two decimals is 1.95.
Phi	A measure of the strength of association or effect size between two categorical variables. A value of .107 represents a very weak association between gender and ethnicity. The equation: $$\varphi = \sqrt{\chi^2 / N}$$
Cells with Expected Count < 5	Three of the ten cells have an expected frequency less than 5. If you have many low-count cells (more than 25% is one accepted criteria), the overall chi-square value is less likely to be valid.
Cramér's V	A measure of the strength of association or effect size between two categorical variables. It differs from phi in that Cramér's V varies strictly between 0 and 1, while in certain cases phi may be greater than 1. The equation follows (Note: k is the smaller of the number of rows and columns): $$V = \sqrt{\chi^2 / [N(k-1)]}$$
Approximate Significance	This is the same as the significance for the Pearson chi-square. The high value (.879) indicates very weak association.

Exercises

Answers to selected exercises are downloadable at **www.spss-step-by-step.net**.

For questions 1-5 below, compute a chi-square and:

1. Circle the observed (actual) values.
2. Box the expected values.
3. Put an * next to the unstandarized residuals.
4. Underline the significance value that shows whether observed and expected values differ significantly.
5. Make a statement about independence of the variables involved.
6. State the nature of the relationship. #5 identifies whether there is a relationship, now you need to indicate what that relationship is. Example: Men tend to help more with goal-disruptive problems, whereas women tend to help more with relational problems.
7. Is there a significant linear association?
8. Does linear association make sense for these variables?
9. Is there a problem with low-count cells?
10. If there is a problem, what would you do about it?

1. File: **grades.sav**. Variables: **gender** by **ethnic**. Select: **observed count**, **expected count**, **unstandardized residuals**. Compute: **Chi-square**, **Phi and Cramer's V**. Edit to fit on one page, print out, and then perform the 10 operations listed above.

2. File: **grades.sav**. Variables: **gender** by **ethnic**. Prior to analysis, complete the procedure shown in Step 5c (page 127) to eliminate the "Native" category (low-count cells). Select: **observed count**, **expected count**, **unstandardized residuals**. Compute: **Chi-square**, **Phi and Cramer's V**. Edit to fit on one page, print out, and then perform the 10 operations listed above.

3. File: **helping3.sav**. Variables: **gender** by **problem**. Select: **observed count, expected count, unstandarized residuals**. Compute: **Chi-square**, **Phi and Cramer's V**. Edit to fit on one page, print out, and then perform the 10 operations listed above.

4. File: **helping3.sav**. Variables: **school** by **occupat**. Prior to analysis, select cases: "**school** > 2 & **occupat** < 6". Select: **observed count, expected count, unstandarized residuals**. Compute: **Chi-square**, **Phi and Cramer's V**. Edit to fit on one page, print out, and then perform the 10 operations listed above.

5. File: **helping3.sav**. Variables: **marital** by **problem**. Prior to analysis, eliminate the "DTS" category (marital < 3). Select: **observed count**, **expected count**, **unstandarized residuals**. Compute: **Chi-square**, **Phi and Cramer's V**. Edit to fit on one page, print out, and then perform the 10 operations listed to the left.

6. Using the **grades.sav** file, run a crosstabulation on whether people attended the review session, and whether they did the extra credit. How many people both attended the review session and did extra credit, attended the review session but did not do extra credit, did not attend the review session but did extra credit, or neither attended the review session nor did extra credit? Can you confidently say that there is a relationship between attending the review session and getting extra credit?

7. The following table shows how many people in the town of Grover's Corners are happy or unhappy, along with whether their socioeconomic status is lower, middle, or upper class. Is there a relationship between social class and happiness, or are they independent? If there is a relationship, how large is this effect and what does it look like?

Happy?	Socioeconomic Status	Number of People
Yes	Lower	3630
Yes	Middle	4992
Yes	Upper	5196
No	Lower	5567
No	Middle	5105
No	Upper	5190

8. Professor Rteneggg predicts that timid creativity is associated with cerebrosolution levels. He finds that, among people with high levels of cerebrosolution, 25 have timid creativity and 55 do not have timid creativity. Among those with low levels of cerebrosolution, 46 have timid creativity and 44 do not have timid creativity. Is Rteneggg's prediction supported by your data? How do your results support this conclusion?