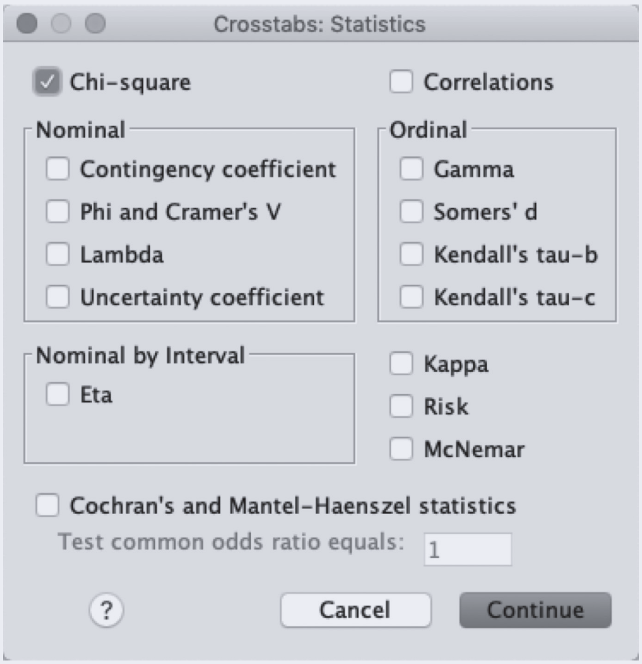
SPSS DEMONSTRATION [GSS18SSDS-A]

Demonstration 1: Producing the Chi-Square Statistic for Cross-Tabulations

The SPSS Crosstabs procedure was previously demonstrated in Chapter 9. This procedure can also be used to calculate a chi-square value for a bivariate table.

Click on *Analyze*, *Descriptive Statistics*, and *Crosstabs*, then on the *Statistics* button. You will see the Dialog box shown in Figure 8.3. To request the chi-square statistic, click on the Chi-square box in the upper left corner. You can also request expected frequencies via the Cells button.

Figure 8.3



Click on *Continue*. In this demonstration, we will look at the relationship between educational degree (DEGREE) and political views (POLVIEWS). Place POLVIEWS in the Row(s) box and DEGREE in the Column(s) box. Then, click on *OK* to run the procedure.

The resulting output includes the chi-square statistics as shown in Figure 8.4. SPSS produces quite a bit of output, perhaps more than what is expected. We will concentrate on the first row of information, the Pearson chi-square.

The Pearson chi-square has a value of 65.102 with 24 degrees of freedom. SPSS calculates the significance of this chi-square to be .000. Educational degree

Figure 8.4

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	65.102[a]	24	.000
Likelihood Ratio	64.671	24	.000
Linear-by-Linear Association	8.364	1	.004
N of Valid Cases	1437		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.25.

and political views are related. Specifically, as educational degree increases (graduate degree attainment), it appears that respondents are more likely to report being "extremely liberal."

The last portion of the output from SPSS allows us to check for the assumption that all expected values in each cell of the table are 5 or greater. The output indicates that zero of the cells or 0.0% have a value less than 5. This is lower than our threshold of 20%.

Demonstration 2: Producing Nominal and Ordinal Measures of Association for Bivariate Tables

The SPSS Crosstabs procedure can also be used to calculate measures of association. We'll begin by investigating the relationship between belief in the Bible (BIBLE) and support for legal abortions for women for any reason (ABANY).

Click on *Analyze*, *Descriptive Statistics*, then *Crosstabs* to get to the Crosstabs dialog box. Put ABANY in the Row(s) box and BIBLE in the Column(s) box. Then click on the Statistics button. The Statistics dialog box (Figure 8.3) has about a dozen statistics from which to choose. Note that four statistics are listed in separate categories for "Nominal" and "Ordinal" data. Lambda is listed in the former, and gamma and Kendall's tau-*b* in the latter. Cramer's *V* can be easily obtained by checking the Phi and Cramer's *V* box. The other measures of association, such as Somer's *d* and Phi, will not be discussed in this textbook.

Since both variables are nominal, check the box for lambda. It is critical that we, as users of statistical programs, understand which statistics to select in any procedure. SPSS, like most programs, can't help us select the appropriate statistic for an analysis. Now click on *Continue* and then *OK* to create the table.

The first table is Case Processing Summary, showing the number of valid and missing cases (not shown here). The second table should be a bivariate table of our two variables (not shown). Below is a table labeled "Directional Measures" (Figure 8.5). For now, we will only concern ourselves with the first two columns. Lambda is listed with three values. We've learned that the value of lambda depends on which variable is considered the dependent variable. In our example, attitude toward abortion for any reason is dependent, so lambda is .303. This indicates a moderate relationship between the two variables. We can conclude that knowing the respondent's belief about the Bible increases the ability to predict his or her abortion attitude by 30.3%.

SPSS also calculates a symmetrical lambda for those tables, where there is no independent or dependent variable. This calculation goes beyond the scope of this book. In addition, SPSS provides the Goodman and Kruskal tau statistic, another nominal measure of association, even though it was not requested. These measures will always be produced when lambda is requested.

If we checked the box for the Phi and Cramer's *V* in the Statistics dialog box, the Symmetric Measures table should be included in the output (Figure 8.5). Cramer's *V* is .412, which indicates a moderate association between belief in the Bible and support for legal abortions for women for any reason.

We can also use the same procedures to calculate gamma for ordinal measures. For this demonstration, we'll examine the relationship between educational attainment (DEGREE) and attitudes toward premarital sex (PREMARSX). Respondents were asked about their views toward premarital sex (not wrong at all, sometimes wrong, almost always wrong, always wrong). Both variables are ordinal measurements.

Click on *Analyze*, *Descriptive Statistics*, then *Crosstabs* to get to the Crosstabs dialog box. Put PREMARSX in the Row(s) box and DEGREE in the Column(s)

Figure 8.5

Directional Measures

			Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Nominal by Nominal	Lambda	Symmetric	.146	.023	5.684	.000
		Abortion if woman wants for any reason Dependent	.303	.028	9.545	.000
		Feelings about the bible Dependent	.009	.039	.238	.812
	Goodman and Kruskal tau	Abortion if woman wants for any reason Dependent	.170	.022		.000[c]
		Feelings about the bible Dependent	.068	.010		.000[c]

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on chi-square approximation

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.412	.000
	Cramer's V	.412	.000
N of Valid Cases		957	

box. Then click on the *Statistics* button. The Statistics dialog box has about a dozen statistics from which to choose. Click on *gamma and Kendall's tau-b* listed in the ordinal box. SPSS produces two separate tables (other than Case Processing Summary); the first is the bivariate table between PREMARSX and DEGREE, and the second is the table of symmetric measures, gamma and Kendall's tau-*b*, which we requested (Figure 8.6).

The Kendall's tau-*b* statistic is in the first row under the column labeled "Value" followed by the gamma statistic. For this bivariate table, both the Kendall's tau-*b* statistic (.067) and the gamma statistic (.107) indicate a weak positive relationship between educational level and attitudes toward premarital sex. Using Kendall's tau-*b*, we can reduce about 7% of our error in predicting attitudes toward premarital sex by using information about respondent's education. Using gamma, about 11% of the error in predicting attitudes toward premarital sex would be reduced if we had information about respondent's educational attainment. Note that given how PREMARSX is coded, the positive gamma indicates that as DEGREE increases, respondents are more likely to view premarital sex as not wrong at all.

Figure 8.6

Symmetric Measures

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Ordinal by Ordinal	Kendall's tau-b	.067	.028	2.402	.016
	Gamma	.107	.045	2.402	.016
N of Valid Cases		995			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

SPSS PROBLEMS [GSS18SSDS-A]

1. Does one's preferred candidate in the 2016 U.S. presidential election vary by respondent's sex?
    a. Use SPSS to investigate the relationship between SEX and PRES16 (voted for Clinton or Trump in the 2016 election). Create a bivariate table with the appropriate percentages. Does the table have a large number of cells with expected values less than 5? Are there any surprises in the data?
    b. Have SPSS calculate chi-square for the table.
    c. Test the null hypothesis at the .05 significance level. What do you conclude?
    d. Select another demographic variable (DEGREE or CLASS) and investigate its relationship with PRES16.

2. Is there a relationship between sexual orientation (SEXORNT) and general happiness rating (HAPPY)? Have SPSS calculate the cross-tabulation treating SEXORNT as the independent variable and HAPPY as the dependent variable, along with chi-square (set alpha at .05). What can you conclude?
3. Examine the relationship between respondent's health (HEALTH) and educational degree (DEGREE). Define DEGREE as your independent variable.
    a. Request the appropriate measures of association to describe the relationship between these two variables. Interpret your measures of association.
    b. Have SPSS calculate the chi-square for the table. Test the null hypothesis at the .05 level. What do you conclude?
4. Examine the relationship between confidence in press (CONPRESS) and highest degree earned (DEGREE).
    a. Which variable is the dependent variable? Explain.
    b. Identify and calculate the appropriate measure of association to describe the relationship between the two variables.
    c. Add SEX as a control variable and calculate the gamma for each partial table. Is the relationship stronger for women or men? Can you think of reasons why this might be so?
5. Use the appropriate measure of association and investigate the relationship between the abortion attitudes in GSS18SSDS-A (e.g., ABANY, ABNOMORE, and ABPOOR) and various demographic variables (you might begin with SEX, CLASS, or DEGREE). For example, you might examine whether attitude toward each of the abortion items has a similar relationship to SEX. In other words, if females are supportive of abortion if one chooses an abortion because they are low income and can't afford more children (ABPOOR), are they also supportive of abortion in other circumstances (ABNOMORE)? Try exploring these relationships further by adding control variables. You might create tables of abortion attitude by SEX, CLASS, and DEGREE. When you have finished the analysis, write a short report summarizing the findings. Suggest possible causes for the relationships you found.

EXCEL DEMONSTRATION [GSS18SSDS-E]

Producing a Chi-Square Statistic in Excel

In Chapter 9 Excel Demonstration 1, we investigated the relationship between condition of health (HEALTH) and general happiness (HAPPY). We treated

HEALTH as the independent variable and HAPPY as the dependent variable. In this demonstration, we will continue our investigation of the relationship between HEALTH and HAPPY by using Excel to run a chi-square test and determine whether or not there is significant relationship between HEALTH and HAPPY.

To begin our investigation, use Excel's PivotTable feature to create a bivariate table of HAPPY and HEALTH data (see Figure 8.7). For a review of how to do such, see Chapter 9 Excel Demonstration 1.

The chi-square test compares the expected frequencies (f_e) with the observed frequencies (f_o). In our table (see Figure 8.7), column percentages of HEALTH are displayed. We can easily change the column percentages to observed frequencies. In the PivotTable Fields window, under Σ Values, click on the "*i*" next to "Count of HEALTH." A window will appear (see Figure 8.8). Click on "Summarize by," and in the dropdown menu, choose "Count." Now click on "Show data as," and in the dropdown menu, choose "No Calculation."

The bivariate table of HEALTH and HAPPY data is now displaying observed frequencies. In cell D2, we've entered OBSERVED FREQUENCIES to recall what we've created (see Figure 8.9). It's very important to keep our Excel sheet organized in the event we save our work and come back to it later.

We will now use Excel to create expected frequencies of our table. To do this, we will first copy and paste the data table we've created. Select the entire data table (including OBSERVED FREQUENCIES) and then, in the main toolbar, select *Edit* and *Copy*.

We will paste our copied table below our original table. Navigate to any cell (we've chosen D14) and in the main toolbar, and select *Edit* and *Paste Special*. It's important to choose *Paste Special* (and not *Paste*). A *Paste Special* window will appear (see Figure 8.10). Select *Values* and then *OK*.

Figure 8.7

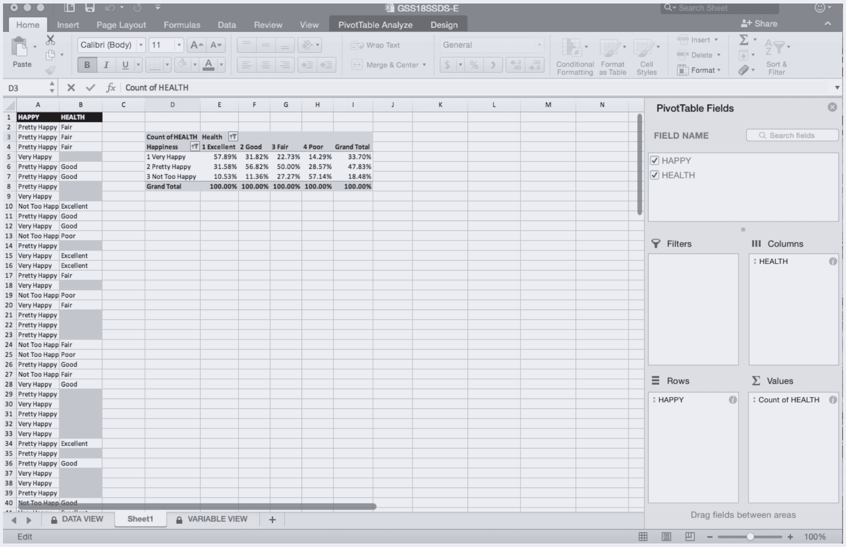


Figure 8.8

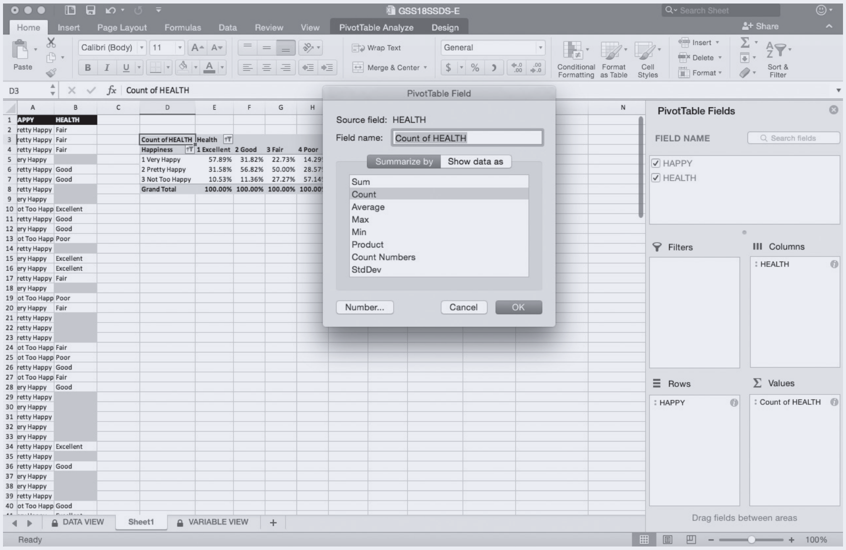
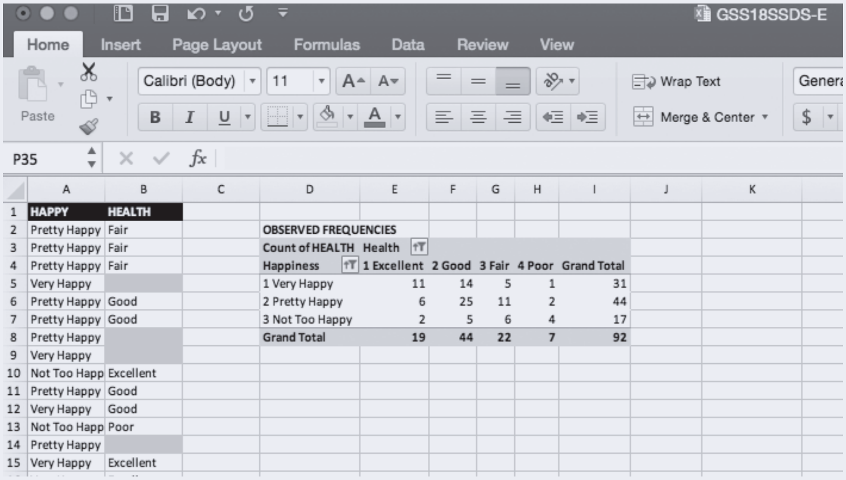


Figure 8.9



Click on cell D14 and replace "OBSERVED FREQUENCIES" with "EXPECTED FREQUENCIES." Then, with the exception of the column and row marginals, select all of the data cells. Data cells E17–H17, E18–H18, and E19–H19 should all be selected. Click on *Edit* → *Clear* → *All*. We are now ready to use Excel to calculate expected frequencies (see Figure 8.11).

Figure 8.10

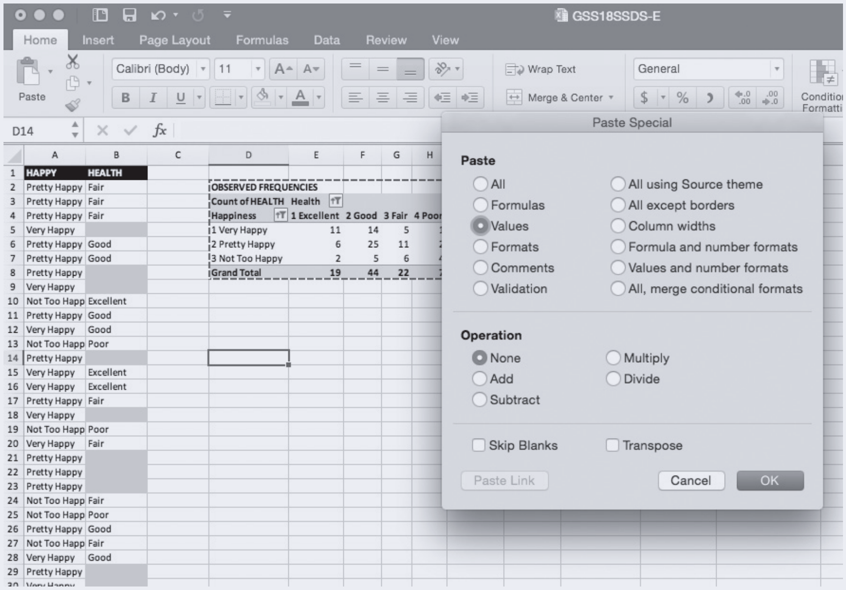
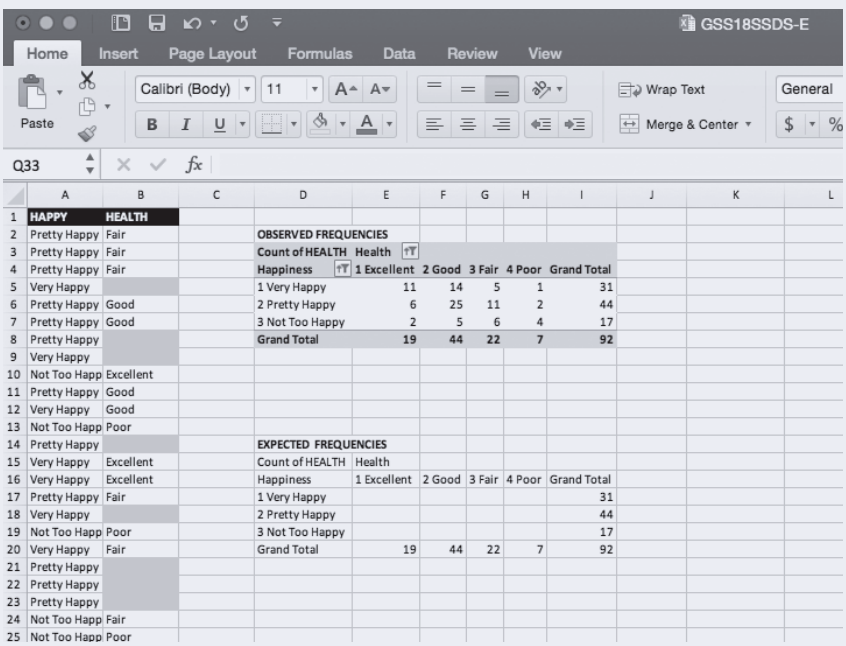


Figure 8.11



We will begin with cell E17 (Excellent Health and Very Happy). In cell E17, type =E20*I17/I20 and then *Enter*. Notice how this formula is telling Excel to multiply the column marginal by the row marginal and then dividing the result by the total (see Figure 8.12). The expected frequency (6.40) for those who are in excellent health and very happy will appear (see Figure 8.13).

We can now move forward calculating the rest of the expected frequencies for the remaining cells. You should write the following formula in each cell:

- Cell E18 (Excellent Health and Pretty Happy): =E20*I18/I20
- Cell E19 (Excellent Health and Not Too Happy): =E20*I19/I20
- Cell F17 (Good Health and Very Happy): =F20*I17/I20
- Cell F18 (Good Health and Pretty Happy): =F20*I18/I20
- Cell F19 (Good Health and Not Too Happy): =F20*I19/I20

Figure 8.12

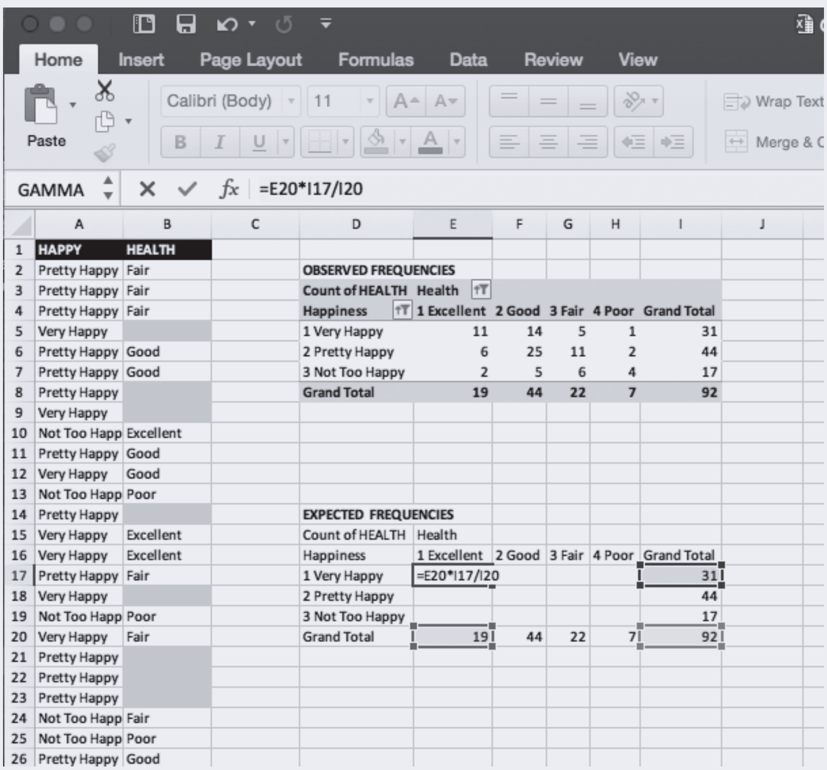
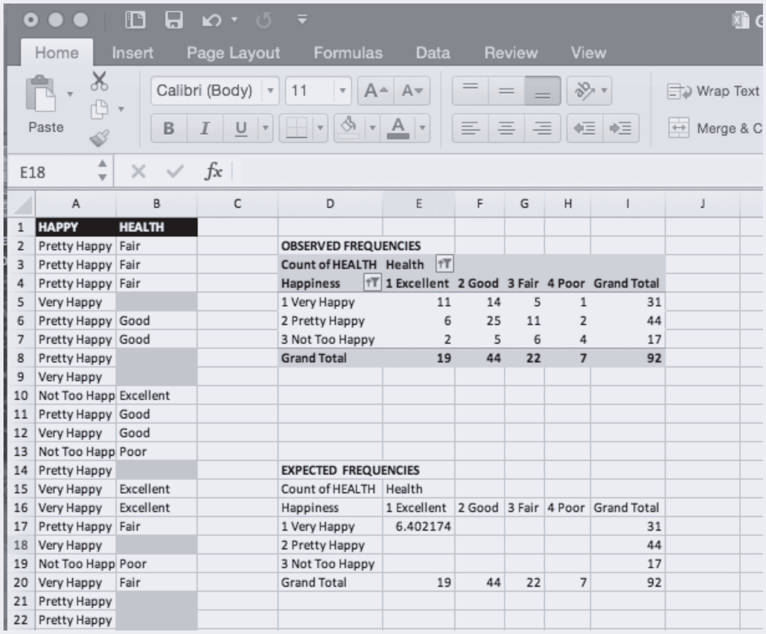


Figure 8.13



	A	B	C	D	E	F	G	H	I	J
1	HAPPY	HEALTH								
2	Pretty Happy	Fair		OBSERVED FREQUENCIES						
3	Pretty Happy	Fair		Count of HEALTH	Health					
4	Pretty Happy	Fair		Happiness	1 Excellent	2 Good	3 Fair	4 Poor	Grand Total	
5	Very Happy			1 Very Happy	11	14	5	1	31	
6	Pretty Happy	Good		2 Pretty Happy	6	25	11	2	44	
7	Pretty Happy	Good		3 Not Too Happy	2	5	6	4	17	
8	Pretty Happy			Grand Total	19	44	22	7	92	
9	Very Happy									
10	Not Too Happ	Excellent								
11	Pretty Happy	Good								
12	Very Happy	Good								
13	Not Too Happ	Poor								
14	Pretty Happy			EXPECTED FREQUENCIES						
15	Very Happy	Excellent		Count of HEALTH	Health					
16	Very Happy	Excellent		Happiness	1 Excellent	2 Good	3 Fair	4 Poor	Grand Total	
17	Pretty Happy	Fair		1 Very Happy	6.402174				31	
18	Very Happy			2 Pretty Happy					44	
19	Not Too Happ	Poor		3 Not Too Happy					17	
20	Very Happy	Fair		Grand Total	19	44	22	7	92	
21	Pretty Happy									
22	Pretty Happy									

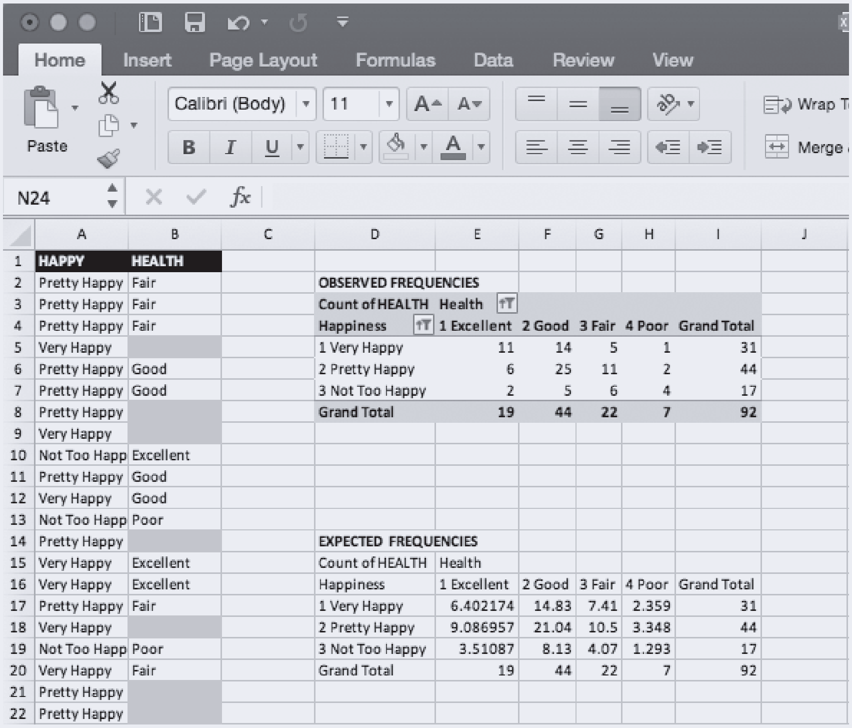
- Cell G17 (Fair Health and Very Happy): =G20*I17/I20
- Cell G18 (Fair Health and Pretty Happy): =G20*I18/I20
- Cell G19 (Fair Health and Not Too Happy): =G20*I19/I20
- Cell H17 (Poor Health and Very Happy): =H20*I17/I20
- Cell H18 (Poor Health and Pretty Happy): =H20*I18/I20
- Cell H19 (Poor Health and Not Too Happy): =H20*I19/I20

Figure 8.14 displays a table of observed frequencies for HEALTH and HAPPY followed by a table of expected frequencies for the same two variables.

Now that we have established the observed and expected frequencies, we can ask Excel to calculate the chi-square and report the associated *p* value for our data. In Excel, this function is called CHITEST.

In cell D24, we will type "p value for chi-square." We will do this simply to keep our work organized. In cell F24, we will type the formula =CHITEST(E5:H7,E17:H19) and then *Enter*. The *p* value (.015) will appear

Figure 10.14

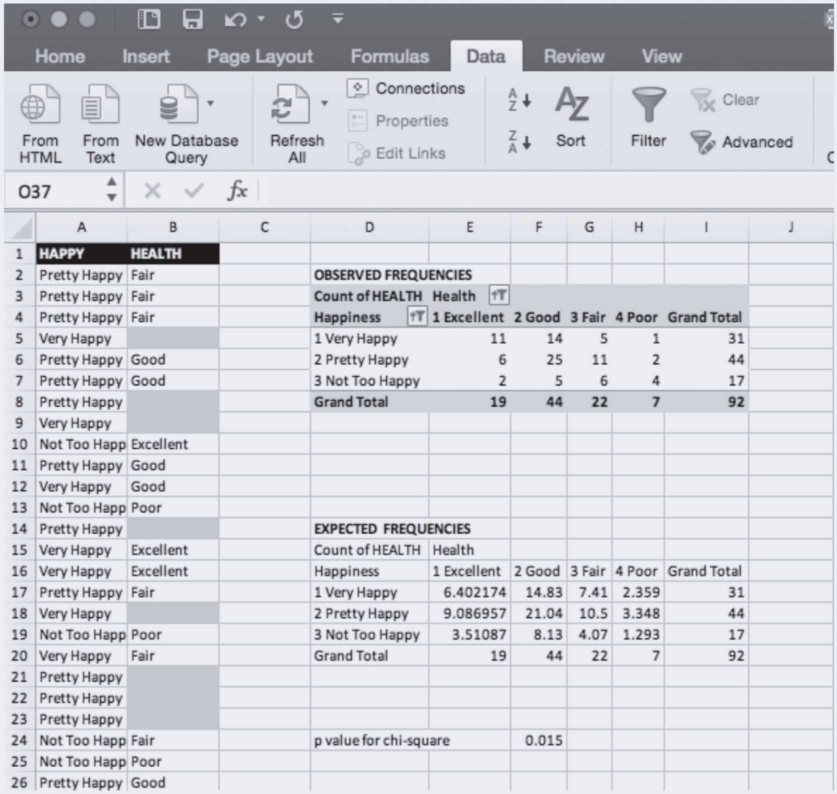


	A	B	C	D	E	F	G	H	I	J
1	HAPPY	HEALTH								
2	Pretty Happy	Fair		OBSERVED FREQUENCIES						
3	Pretty Happy	Fair		Count of HEALTH	Health					
4	Pretty Happy	Fair		Happiness	1 Excellent	2 Good	3 Fair	4 Poor	Grand Total	
5	Very Happy			1 Very Happy	11	14	5	1	31	
6	Pretty Happy	Good		2 Pretty Happy	6	25	11	2	44	
7	Pretty Happy	Good		3 Not Too Happy	2	5	6	4	17	
8	Pretty Happy			Grand Total	19	44	22	7	92	
9	Very Happy									
10	Not Too Happ	Excellent								
11	Pretty Happy	Good								
12	Very Happy	Good								
13	Not Too Happ	Poor								
14	Pretty Happy			EXPECTED FREQUENCIES						
15	Very Happy	Excellent		Count of HEALTH	Health					
16	Very Happy	Excellent		Happiness	1 Excellent	2 Good	3 Fair	4 Poor	Grand Total	
17	Pretty Happy	Fair		1 Very Happy	6.402174	14.83	7.41	2.359	31	
18	Very Happy			2 Pretty Happy	9.086957	21.04	10.5	3.348	44	
19	Not Too Happ	Poor		3 Not Too Happy	3.51087	8.13	4.07	1.293	17	
20	Very Happy	Fair		Grand Total	19	44	22	7	92	
21	Pretty Happy									
22	Pretty Happy									

(see Figure 8.15). There are ways to ask Excel to report the obtained chi-square value; however, such requires a number of additional steps that are beyond the scope of this demonstration. The *p* value associated with the chi-square value offers us enough information to either confirm or reject our null hypothesis. Also, it is important to cautiously interpret any chi-square results when you are working with a small sample size—especially when cells have small expected count values. For the sake of this emonstration, we will move forward with our interpretation of the results.

Excel has reported a *p* value for the chi-square test of the relationship between HEALTH and HAPPY to be .015. If we set our alpha value at .05, we can conclude that condition of health and general happiness are significantly related because our *p* value (.015) is less than our alpha (.05). We can reject the null hypothesis that there is no relationship between HEALTH and HAPPY.

Figure 8.15



	A	B	C	D	E	F	G	H	I	J
1	HAPPY	HEALTH								
2	Pretty Happy	Fair		OBSERVED FREQUENCIES						
3	Pretty Happy	Fair		Count of HEALTH	Health					
4	Pretty Happy	Fair		Happiness	1 Excellent	2 Good	3 Fair	4 Poor	Grand Total	
5	Very Happy			1 Very Happy	11	14	5	1	31	
6	Pretty Happy	Good		2 Pretty Happy	6	25	11	2	44	
7	Pretty Happy	Good		3 Not Too Happy	2	5	6	4	17	
8	Pretty Happy			Grand Total	19	44	22	7	92	
9	Very Happy									
10	Not Too Happ	Excellent								
11	Pretty Happy	Good								
12	Very Happy	Good								
13	Not Too Happ	Poor								
14	Pretty Happy			EXPECTED FREQUENCIES						
15	Very Happy	Excellent		Count of HEALTH	Health					
16	Very Happy	Excellent		Happiness	1 Excellent	2 Good	3 Fair	4 Poor	Grand Total	
17	Pretty Happy	Fair		1 Very Happy	6.402174	14.83	7.41	2.359	31	
18	Very Happy			2 Pretty Happy	9.086957	21.04	10.5	3.348	44	
19	Not Too Happ	Poor		3 Not Too Happy	3.51087	8.13	4.07	1.293	17	
20	Very Happy	Fair		Grand Total	19	44	22	7	92	
21	Pretty Happy									
22	Pretty Happy									
23	Pretty Happy									
24	Not Too Happ	Fair		p value for chi-square		0.015				
25	Not Too Happ	Poor								
26	Pretty Happy	Good								

EXCEL PROBLEMS [GSS18SSDS-E]

E1. Do attitudes toward premarital sex (PREMARSX) vary by respondent's sex (SEX)? Treat SEX as the independent variable and PREMARSX as the dependent variable.

a. Create a bivariate table of observed frequencies for SEX and PREMARSX.

b. Create a bivariate table of expected frequencies for SEX and PREMARSX.

c. Using Excel's CHITEST command, what is the *p* value for the chi-square test?

d. Test the null hypothesis at the .05 significance level. What do you conclude?

E2. Do political views (RE_POLVIEWS) vary by marital status (MARITAL)? Treat MARITAL as the independent variable and RE_POLVIEWS as the dependent variable.

a. Create a bivariate table of observed frequencies for MARITAL and RE_POLVIEWS.

b. Create a bivariate table of expected frequencies for MARITAL and RE_POLVIEWS.

c. Using Excel's CHITEST command, what is the *p* value for the chi-square test?

d. Test the null hypothesis at the .01 significance level. What do you conclude? Would your results change if you test the null hypothesis at the .05 significance level?

E3. Does a respondent's preferred candidate in the 2016 presidential election (PRES16) vary by political views (RE_POLVIEWS)? Treat RE_POLVIEWS as the independent variable and PRES16 as the dependent variable.

a. Create a bivariate table of observed frequencies for RE_POLVIEWS and PRES16.

b. Create a bivariate table of expected frequencies for RE_POLVIEWS and PRES16.

c. Using Excel's CHITEST command, what is the *p* value for the chi-square test?

d. Test the null hypothesis at the .05 significance level. What do you conclude? Would your results change if you test the null hypothesis at the .01 significance level?