



6. Inference for Two-Way Tables

This chapter introduces a method for analyzing the relationship between two categorical variables. The data are usually summarized with a two-way table (see Section 3.2). Inferential procedures are applied to test whether the two categorical variables are independent.

Example 3-2: Age and Political Interest, continued, from page 56. This data set was summarized in Table 3-6, page 59, and it appeared that there was an association between age and interest in politics, with interest appearing to increase with age. The 1265 respondents to this survey are a random sample (or close to it) of all American adults so we might ask if we can conclude that these data provide evidence of an association between age and interest for all American adults or whether the observed association in the sample could be due to sampling variability. A chi-square test of independence between two categorical variables can be used to assess this (“independence” means “no association”). The steps for carrying out a chi-square test of independence follow. The steps are very similar to those for creating a two-way table as described in **Creating a two-way table**, page 55. Remember that in this example, the data were entered as counts, so the **Weight Cases** option under **Data** was activated (see **Entering count data**, page 56).

1. Click **Analyze**, click **Descriptive Statistics**, and click **Crosstabs**. The “Crosstabs” window in Figure 6-1 appears.
2. Click on *age*, then click  to move *age* to the “Row(s)” box.
3. Click on *interest*, then click  to move *interest* to the “Column(s)” box.
4. Click the “Cells” button at the bottom of the “Crosstabs” window. The “Crosstabs: Cell Display” window in Figure 6-2 appears.
5. If you want the expected count in each cell, click **Expected** in the “Counts” box so that a check appears in the box in front of **Expected**. If you want the row percents, click **Row** in the “Percentages” box so that a check appears in the box in front of **Row**.
6. Click **Continue**.
7. Click the **Statistics** button at the bottom of the “Crosstabs” window (Figure 6-1). The “Crosstabs: Statistics” window in Figure 6-3 appears.
8. Click **Chi-square** so that a check appears in the box in front of it.
9. Click **Continue**.
10. Click **OK**.

The resulting SPSS output appears in Table 6-1 and Table 6-2. The row percentages were previously computed in Table 3-6, page 59, and show that the percentage of people who are “not much” interested in politics decreases with age while the percentage who are “very much” interested increases with age. Comparing the expected counts to the observed counts also shows this pattern. We can also see that the expected counts are all at least 5 so the expected cell frequency condition for using the chi-square test is met (page 547 of the text). The test statistic and p-value are obtained from the first line (Pearson Chi-square) of Table 6-2; ignore the rest of the table. This line gives us a test statistic of $X^2 = 41.446$ with $df = 4$ and p-value (“Asymp. Sig.”) of less than .0005 (it’s .000 when rounded to 3 decimal places so must be less than .0005; it’s not exactly 0). The p-value is very small indicating that the data provide very strong evidence that there is an association between age and political interest among all U.S. adults and that, in fact, older people tend to be more interested in politics than younger people. Remember that a significance test does not tell us how strong the association is, so we need to look at the row percents in Table 6-1 to see how big a difference between the age groups there is.

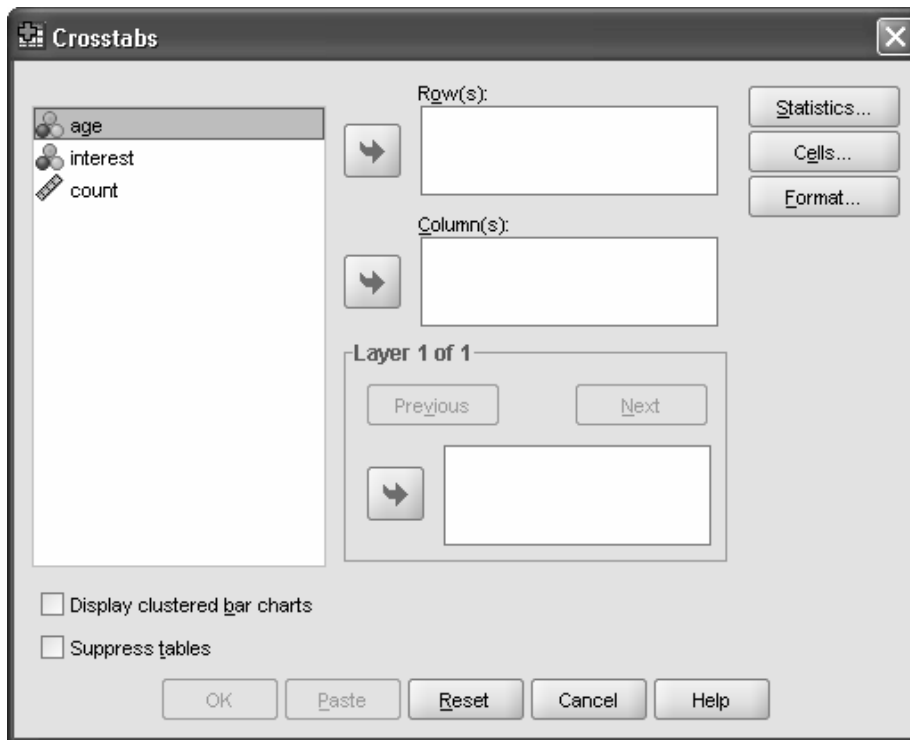


Figure 6-1

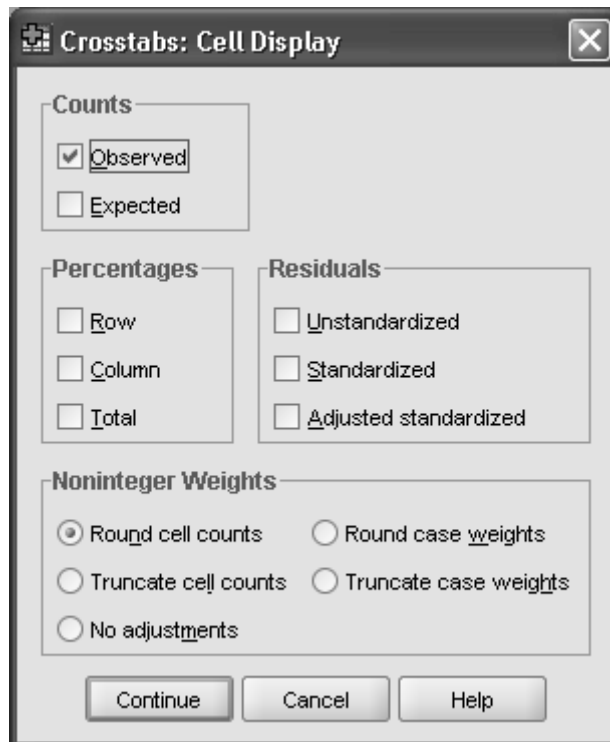


Figure 6-2

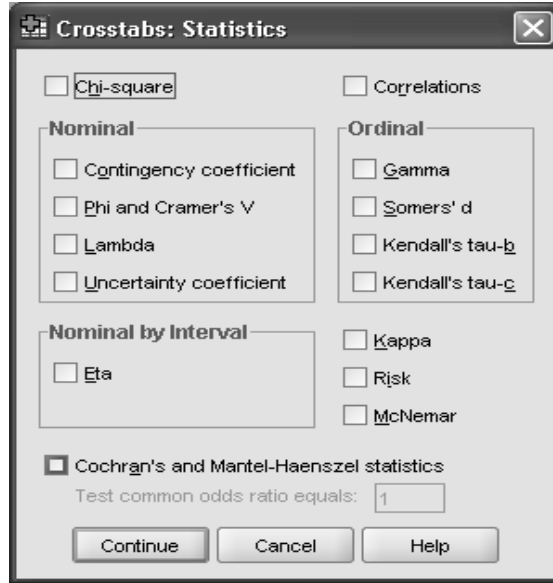


Figure 6-3

AGE * INTEREST Crosstabulation

			INTEREST			Total
			not much	somewhat	very much	
AGE	18-35	Count	146	192	47	385
		Expected Count	116.0	184.4	84.6	385.0
		% within AGE	37.9%	49.9%	12.2%	100.0%
	36-55	Count	146	260	125	531
		Expected Count	159.9	254.4	116.7	531.0
		% within AGE	27.5%	49.0%	23.5%	100.0%
	56-94	Count	89	154	106	349
		Expected Count	105.1	167.2	76.7	349.0
		% within AGE	25.5%	44.1%	30.4%	100.0%
Total	Count	381	606	278	1265	
	Expected Count	381.0	606.0	278.0	1265.0	
	% within AGE	30.1%	47.9%	22.0%	100.0%	

Table 6-1

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	41.446 ^a	4	.000
Likelihood Ratio	43.068	4	.000
Linear-by-Linear Association	33.886	1	.000
N of Valid Cases	1265		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 76.70.

Table 6-2