Chapter 8 – Cross-Tabulation, Chi-Square, and Non Parametric Measures of Association

Using college student data from file, do the following problems. Print your outputs after typing your interpretations on them. Please circle the key parts of the output that you discuss.

8.1. Run crosstabs and interpret the results of chi-square and phi (or Cramer's *V*), as discussed in Chapter 6 and in the interpretation of Output 8.1, for: (a) gender and marital status and (b) age group and marital status.

Selection of Statistic

Problem 8.1 is designed to analyze two categorical variables (i.e., ones that are nominal or have only a few ordered levels). Remember, nominal variables are variables that have distinct unordered levels; each subject is in only one level (you can only be male *or* female). Chi-square (χ^2) or phi/Cramer's *V* are good choices for statistics when comparing two nominal variables. They are less appropriate if either variable has three or more *ordered* levels because these statistics do not take in account the order and, thus, sacrifice power.

Chi-square requires a relatively large sample size and/or a relatively even split of the subjects among the levels because the expected counts in 80% of the cells should be greater than 5. Fisher's exact test for 2x2 (only) crosstabs can be reported instead of chi-square for small samples. Chi-square and the Fisher's exact test provide similar information about differences between groups (and relationships among variables); however, they only tell us whether the difference (or relationship) is statistically significant (not likely to be due to chance). They do *not* tell the effect size (the strength of the relationship).

Phi and Cramer's *V* provide information about the *strength* of the association between two categorical variables and can be used as a measure of the effect size similar to *r*. If one has a 2x2 cross tabulation, phi is the appropriate statistic. For larger crosstabs, Cramer's *V* is used. The numbers in the crosstabs description refer to the number of

levels in each of the variables. Thus, for gender and marital, the crosstab is 2x3 because gender has two levels and marital status has three levels.

Assumptions of Chi-square

- If the expected frequencies, are less than 5, the test of significance is too liberal. At least 80% of the expected frequencies should be 5 or larger. All should be at least 5 if you have a 2x2 chi-square.
- **Data are treated as nominal, even if ordered.**

Assumptions of Phi/Cramer's V **(measure of association for nominal variables)**

How to Produce the Selected SPSS Output

To answer Problem 8.1a with Windows:

- Click on Analyze ⇒ Desc Stat ⇒ Crosstabs. This will open the Crosstabs window
- Highlight gender
- Click on the arrow to move gender over to the Row(s) box
- Highlight marital
- Click on the arrow to move it over to the Column(s) box
- Click on Statistics. This will open the Crosstabs: Statistics window
- Click on Phi/Cramer's *V* and chi-square
- Click on Continue
- Click on cells. This will open the Crosstabs: Cell Display
- Click on Expected, Total and Standardized
- Click on Continue and O.K.

To answer Problem 8.1a with syntax:

```
CROSSTABS
  /TABLES=gender  BY marital
  /FORMAT= AVALUE TABLES
  /STATISTIC=CHISQ PHI
  /CELLS= COUNT EXPECTED TOTAL SRESID .
```

SPSS Output for Problem 8.1a

Crosstabs

One observation is missing.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
gender of student * marital status	49	98.0%	1	2.0%	50	100.0%

Count is the number of subjects who responded in that cell.

Variable name. The levels are: single, married, and divorced.

gender of student * marital status Crosstabulation

			marital status			
			single	married	divorced	Total
gender of student	males	Count	14	7	5	26
		Expected Count	10.6	9.6	5.8	26.0
		% of Total	28.6%	14.3%	10.2%	53.1%
		Std. Residual	1.0	-.8	-.3	
	females	Count	6	11	6	23
		Expected Count	9.4	8.4	5.2	23.0
		% of Total	12.2%	22.4%	12.2%	46.9%
		Std. Residual	-1.1	.9	.4	
Total		Count	20	18	11	49
		Expected Count	20.0	18.0	11.0	49.0
		% of Total	40.8%	36.7%	22.4%	100.0%

Number of subjects expected by chance to be in that cell

The differences between count and expected count, then standardized. Residuals greater than 2 indicates that the difference is a major contributor to a significant chi-square. None of the residuals are greater than 2 in this output.

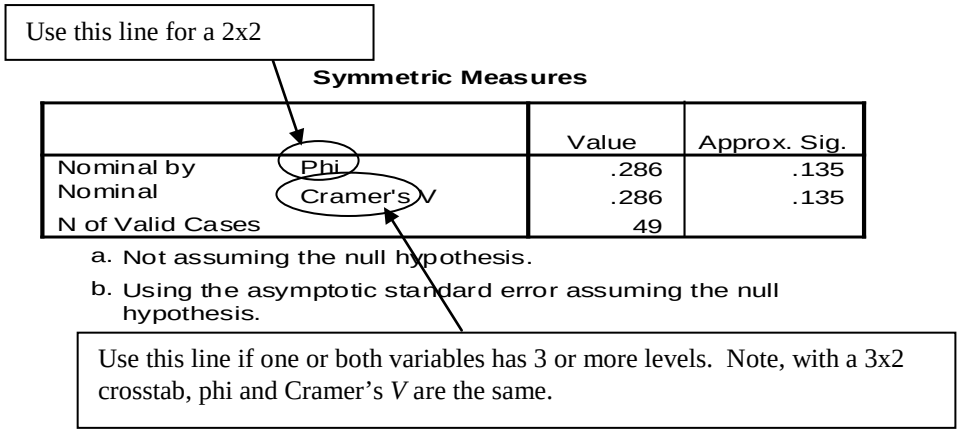
Significance level. This chi-square is not significant.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	4.011[a]	2	.135
Likelihood Ratio	4.095	2	.129
Linear-by-Linear Association	2.392	1	.122
N of Valid Cases	49		

This line indicates the results of the chi square statistic.

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 5.16.



How to Produce the Selected SPSS Output

To answer Problem 8.1b with Windows:

- Click on Analyze ⇒ Desc Stat ⇒ Crosstabs. This will open the Crosstabs window
- Highlight age group and move it into the Row(s) box
- Highlight marital and move it into the Column(s) box
- Click on Statistics. This will open the Crosstabs: Statistics window
- Click on Phi, Cramer's *V* and chi-square
- Click on Continue
- Click on Cells. This will open the Crosstabs: Cell Display window
- Click on Expected, Total and Standardized
- Click on Continue and O.K.

To answer Problem 8.1b with syntax:

```
CROSSTABS
  /TABLES=age  BY marital
  /FORMAT= AVALUE TABLES
  /STATISTIC=CHISQ PHI
  /CELLS= COUNT EXPECTED TOTAL SRESID .
```

SPSS Output for Problem 8.1b

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
age group * marital status	49	98.0%	1	2.0%	50	100.0%

age group * marital status Crosstabulation

			marital status			Total
			single	married	divorced	
age group	less than 22	Count	13	1	2	16
		Expected Count	6.5	5.9	3.6	16.0
		% of Total	26.5%	2.0%	4.1%	32.7%
		Std. Residual	2.5	-2.0	-.8	
	22-29	Count	7	6	5	18
		Expected Count	7.3	6.6	4.0	18.0
		% of Total	14.3%	12.2%	10.2%	36.7%
		Std. Residual	-.1	-.2	.5	
	30 or more	Count	0	11	4	15
		Expected Count	6.1	5.5	3.4	15.0
		% of Total	.0%	22.4%	8.2%	30.6%
		Std. Residual	-2.5	2.3	.3	
Total		Count	20	18	11	49
		Expected Count	20.0	18.0	11.0	49.0
		% of Total	40.8%	36.7%	22.4%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	23.173[a]	4	.000
Likelihood Ratio	28.888	4	.000
Linear-by-Linear Association	11.590	1	.001
N of Valid Cases	49		

a. 3 cells (33.3%) have expected count less than 5. The minimum expected count is 3.37.

This is not good because for chi-square no more than 20% of the cells should have expected counts less than 5.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.688	.000
	Cramer's V	.486	.000
N of Valid Cases		49	

a. Not assuming the null hypothesis.

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

For 3x3 and larger cross tabulations, phi and *V* are different; use Cramer's *V*.

Description of Output 8.1a and b

The Case Processing Summary table for Output 8.1a indicates that there is one person missing data about either gender or marital status, or both. There is also one person missing in Output 8.1b.

The 8.1a cross tabulation table shows the number of individuals (counts) with each marital status that are males or females. The expected count is the number one would expect to have in each category by chance based on the column and row totals. Each cell also shows a percent of the grand total, which indicates the percentages of persons of each gender occupying each marital status.

Look at the difference between the observed count and the expected count. In tables larger than 2x2, the differences are of different sizes. The chi-square will assist in determining whether these differences could have occurred by chance. *If the statistical test is significant*, then it is important to identify and discuss the effect size by examining phi or Cramer's *V*. It is also critical to determine which cells are contributing to the significant results if we want to discuss the findings fully; to do this, we examine the standardized residuals to see which one(s) are more than 2.0. Note that in Problem 8.1b, age is an ordered variable (going from low to high) so chi-square and Cramer's *V* are not as powerful as would be an ordinal statistic (e.g., Kruskal-Wallis).

You can see from footnote [a] for the chi-square tests table in Output 8.1a, that the assumption about cell sizes has not been violated because no cells have an expected count of less than 5. On the other hand, three of the cells for the 8.1b chi-square have an expected count of less than 5 so an assumption has been violated and we should be cautious about the interpretation of chi-square. However, the chi-square Sig. is so much less (.000) than .05 that we probably don't need to worry about it in this case.

It's usually not necessary to have a table to report one chi-square. Thus, we have combined Output 8.1a and 8.1b into one table, Table 8.1.

Example of How to Write About Problem 8.1a and b

Results

Table 8.1 shows the Pearson chi-square results and indicates that males and females are not significantly different on whether they are single, married, or divorced ($\chi^2 = 4.01$, $df = 2$, $N = 49$, $p = .135$).

Table 8.1 also shows that marital status is different for each of the age groups ($\chi^2 = 23.17$, $df = 4$, $N = 49$, $p < .001$). Cramer's V indicates that the strength of the association between the two variables is .49 and, thus, the effect size is relatively large. Students under 22 are more likely than expected under the null hypothesis to be single and older student (over 30) are more likely to be married. However, 33% of the cells had expected frequencies less than 5 so we need to be cautious about this conclusion.

Table 8.1

Chi-square Analyses of Prevalence of Gender and Age Among Single, Married, and Divorced Students

Variable	*n*	Marital Status			χ^2	*p*
		Single	Married	Divorced		
Gender					4.01	.135
Males	26	14	7	5		
Females	23	6	11	6		
Age Group					23.17	<.001
Less than 22	16	13	1	2		
22-29	18	7	6	5		
30 or more	15	0	11	4		
Totals	49	20	18	11		

8.2. Select two other appropriate variables, run and interpret the output as we did in Output 8.1.

Selection of the Statistic

(Same as for Problem 8.1.)

How to Produce the Selected SPSS Output

(Same as 8.1, but the variables are different)

SPSS Output for Problem 8.2a

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
gender of student * television shows-sports	50	100.0%	0	.0%	50	100.0%

television shows-sports * gender of student Crosstabulation

			gender of student		
			males	females	Total
television shows-sports	no	Count	2	22	24
		Expected Count	12.5	11.5	24.0
		% of Total	4.0%	44.0%	48.0%
		Std. Residual	-3.0	3.1	
	yes	Count	24	2	26
		Expected Count	13.5	12.5	26.0
		% of Total	48.0%	4.0%	52.0%
		Std. Residual	2.9	-3.0	
Total		Count	26	24	50
		Expected Count	26.0	24.0	50.0
		% of Total	52.0%	48.0%	100.0%

All cells have standardized residuals greater than 2 so all are contributing to the significant chi-square. Males say yes more and no less than females who say the reverse.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	35.258[b]	1	.000		
Continuity Correction[a]	31.974	1	.000		
Likelihood Ratio	41.365	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	34.553	1	.000		
N of Valid Cases	50				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 11.52.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	-.840	.000
	Cramer's V	.840	.000
N of Valid Cases		50	

a. Not assuming the null hypothesis.

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

Description of Output 8.2a

(Same as for Problem 8.1.)

Example of How to Write About Problem 8.2a

Results

Chi-square analysis indicated a significant difference between males (24 out of 26) and females (2 out of 24) in the proportions who watch sports programs on television (χ^2= 35.26, *df* =1, *N* = 50, $p < .001$). As hypothesized, male students were more likely to watch sports on TV than are females. The phi value (-.84) indicates that the effect size was very large (Cohen, 1988).

Discussion

The finding that males tend to watch sport programs more than females is similar to findings in other studies (Smith, 1999; Jones, 1998). In today's society many boys feel pressured to excel at sports. This might be explained partially from the modeling of adult males who watch sports programs on television.

Table 8.2a[1]

Chi-square Analysis of Responses to Watching Sports and Gender

Gender	Yes	No	χ^2	*p*
Males	24	2	35.26	<.001
Females	2	22		

Example of How to Write About Problem 8.2b

Results

Chi-square analyses in Table 8.2b indicated significant differences between males and females in the proportions who watch sitcoms, movies, and sports, but not news. Females were more likely than males to watch sitcoms and movies. Males were more likely than females to watch sports. The effect sizes were very large for movies and sports.

Table 8.2b

Chi-square Analyses of "Yes" Responses to Type of Television Viewed and Gender

	Gender			
Type of show	Males (*n*=26)	Female (*n*=24)	χ^2	*p*
Sitcoms	11	21	11.06	.001
Movies	0	18	30.47	<.001
Sports	24	2	35.26	<.001
News	14	9	1.34	.247

[1] One would not usually make a table for one chi-square. See table 8.2b for an example of a table with several related chi-squares. If you have both text and a table do not repeat the statistics (χ^2 = 35.26, etc.) in the text.

8.3. Is there an association between having children or not and watching TV sitcoms?

8.4. Is there a difference between students who have children and those who do not in regard to their age group?

8.5. Compute an appropriate statistic and effect size measure for the relationship between gender and evaluation of social life.