5

Cross-Tabulation and Measures of Association for Nominal and Ordinal Variables

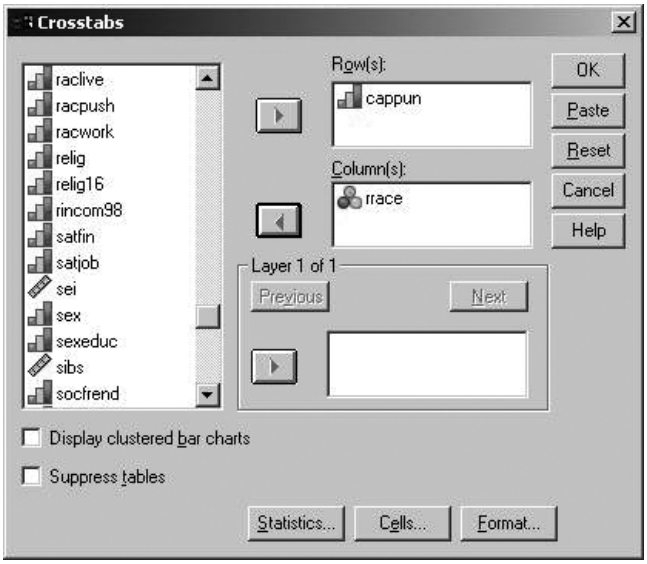
The most basic type of cross-tabulation (crosstabs) is used to analyze relationships between two variables. This allows a researcher to explore the relationship between variables by examining the intersections of categories of each of the variables involved. The simplest type of cross-tabulation is bivariate analysis, an analysis of two variables. However, the analysis can be expanded beyond that.

Bivariate Analysis

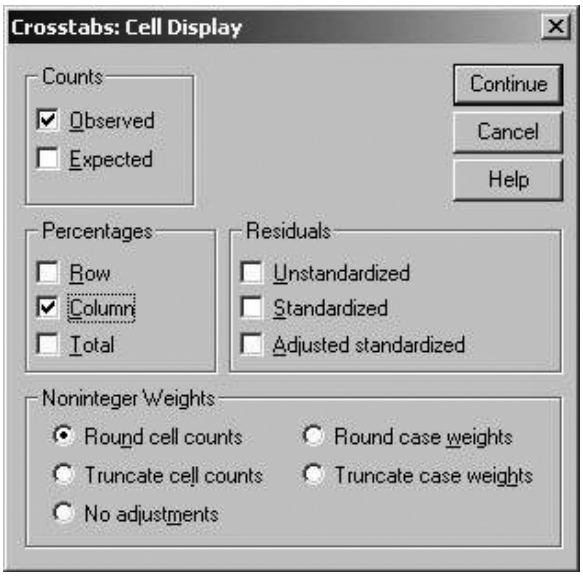
By example, follow these menus to conduct a cross-tabulation of two variables:

ANALYZE → DESCRIPTIVES → CROSSTABS . . .

After selecting those menus, you will be presented with a dialog box like the one above. Here, you will have the opportunity to select the row and column variables for the bivariate table. As is customary, it is recommended that you choose the independent variable as the column variable. Above, "cappun" (view on capital punishment/the death penalty) has been selected for the row



variable, and "rrace" (recoded version of respondent's race) has been chosen as the column variable. We use "rrace" because the number of categories has been collapsed to four. This makes it easier to interpret data from cross-tabulations when the number of categories is kept smaller. Next, click on the "Cells" button to choose options about what information will be given in the output table.



Be sure that the "Column" box is checked under "Percentages." This will ensure that you have information from the appropriate perspective to analyze your variables based on which is the predictor. Click "Continue" in the "Cell Display" dialog box, then "OK" in the original "Crosstabs" dialog box. The table that follows comes from the output produced by following the aforementioned steps.

FAVOR OR OPPOSE DEATH PENALTY FOR MURDER * Race Cross-Tabulation

			Race				Total
			other	white	black	Hispanic	
FAVOR OR OPPOSE DEATH PENALTY FOR MURDER	FAVOR	Count	14	412	45	14	485
		% within race recode	58.3%	71.4%	48.4%	82.4%	68.2%
	OPPOSE	Count	10	165	48	3	226
		% within race recode	41.7%	28.6%	51.6%	17.6%	31.8%
Total		Count	24	577	93	17	711
		% within race recode	100.0%	100.0%	100.0%	100.0%	100.0%

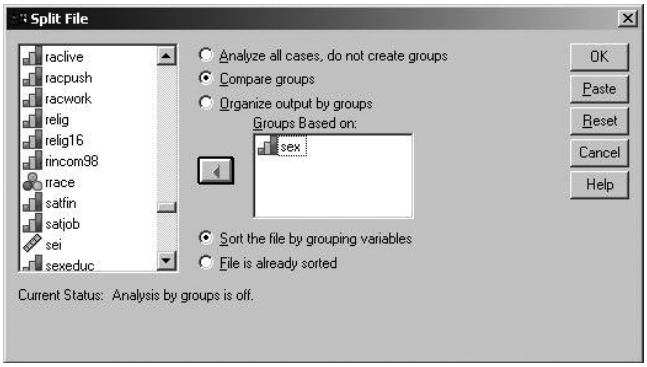
Based on the information in the table, it is easy to see that there is some sort of relationship between the variables of interest in this case. Note that by looking at the percentages across the columns (categories of the independent variable), one can see that there are differences in opinion by race about the death penalty. According to these GSS (General Social Survey) data, whites and Hispanics are more likely to favor the death penalty than blacks or others.

Adding Another Variable or Dimension to the Analysis

Suppose we want to further explore the bivariate relationship that we briefly examined in the preceding section. By adding another variable, such as respondent's sex, we can further explore how opinions about capital punishment are held in the United States. One way that we can perform this type of analysis is to split our data file by respondent's sex. At that point, any analysis that we do will be performed across the categories of the variable with which we have split the data set.

In order to split the data file by respondent's sex, use these menus:

DATA → SPLIT FILE . . .



You will be given a "Split File" dialog box like the one above. Here, choose the "Compare groups" radio button. This will brighten the "Groups Based on" box and allow you to now move variables into that box, which will then be used to split the data file. Find "sex" from the variable bank on the left and move it over into the "Groups Based on" box. It is often a good idea to make sure that the file is sorted by grouping variables, although this is not necessary. Click "OK."

SPSS will now perform the "Split File" function. You will know that the data file has been split by the indicator in the lower right-hand window of the data editor. It will say "Split File On."

Now go back to the "Crosstabs" menu and perform the same operations that were done in the previous section. (The variables and setting should remain the same from before, so unless you've restarted SPSS in between, it will just be a matter of choosing "OK" once the dialog box appears.)

ANALYZE → DECRIPTIVES → CROSSTABS . . .

FAVOR OR OPPOSE DEATH PENALTY FOR MURDER * Race Cross-Tabulation

				Race				
R's SEX				other	white	black	Hispanic	Total
MALE	FAVOR OR OPPOSE DEATH PENALTY FOR MURDER	FAVOR	Count	9	225	19	8	261
			% race	64.3%	78.1%	59.4%	80.0%	75.9%
		OPPOSE	Count	5	63	13	2	83
			% race	35.7%	21.9%	40.6%	20.0%	24.1%
	Total		Count	14	288	32	10	344
			% race	100.0%	100.0%	100.0%	100.0%	100.0%
FEMALE	FAVOR OR OPPOSE DEATH PENALTY FOR MURDER	FAVOR	Count	5	187	26	6	224
			% race	50.0%	64.7%	42.6%	85.7%	61.0%
		OPPOSE	Count	5	102	35	1	143
			% race	50.0%	35.3%	57.4%	14.3%	39.0%
	Total		Count	10	289	61	7	367
			% race	100.0%	100.0%	100.0%	100.0%	100.0%

The preceding table is presented as part of the output that SPSS returns. Although it is similar to the table given in the prior section, note that it has twice as many cells. It has been split into two tables, one for males and one for females. In this instance, among other things, it can be seen from the table that Hispanic females have the highest percentage of all categories of men and women who "favor the death penalty for murder." In all other racial categories, men tend to be more likely to favor the death penalty. By adding this new dimension, we are able to obtain some additional insight into public opinion on this matter. See your statistics or research methods book(s) for more details.

Measures of Association for Nominal and Ordinal Variables

PRE statistics allow us to determine the proportional reduction of error achieved by adding one or more variables to an analysis (even if just one independent variable). "PRE measures are derived by comparing the errors made in predicting the dependent variable while ignoring the independent variable with errors made when making predictions that use information about the independent variable" (Frankfort-Nachmias & Leon-Guerrero, 2006). For nominal variables, we utilize lambda. For details on how lambda is calculated, see Chapter 7 of *Social Statistics for a Diverse Society* (2006).

Lambda

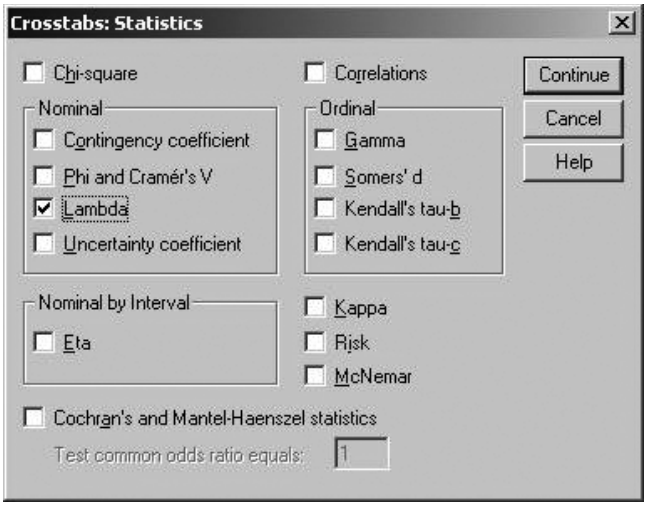
To compute lambda for the relationship between race and view on capital punishment, being again by selecting the cross-tabulation menu:

ANALYZE → DESCRIPTIVES → CROSSTABS . . .

Now, when presented the "Crosstabs" dialog box, and after entering the variables of interest, select the "Statistics" button. You will be given the dialog box on the following page.

Under the "Nominal" heading, select "Lambda." This will instruct SPSS to add lambda to the things it will present in the output. Now click "Continue" in the "Statistics" dialog box, then "OK" in the prior dialog box. On the following page you will find an image from the output SPSS would produce.

Typically, lambda is presented as an asymmetrical measure of association as is the case in *Social Statistics for a Diverse Society* (2006). Given that, the value of lambda to be used can be found in the "value" column in the row indicating the correct dependent variable. In this case, "cappun" (Favor or Oppose



Directional Measures

			Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Nominal by Nominal	Lambda	Symmetric	.008	.027	.311	.756
		FAVOR OR OPPOSE DEATH PENALTY FOR MURDER Dependent	.013	.042	.311	.756
		Race Dependent	.000	.000	.[c]	.[c]
	Goodman and Kruskal tau	FAVOR OR OPPOSE DEATH PENALTY FOR MURDER Dependent	.031	.014		.000[d]
		Race Dependent	.020	.010		.000[d]

a. Not assuming the null hypothesis.

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

c. Cannot be computed because the asymptotic standard error equals zero.

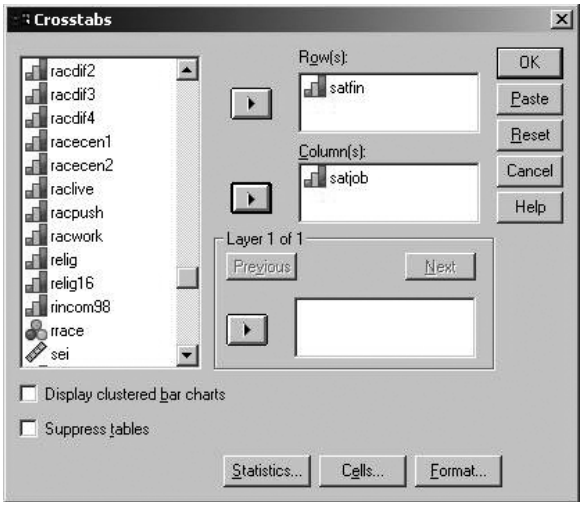
d. Based on chi-square approximation

Death Penalty for Murder) is the appropriate dependent variable. We see that lambda is 0.013 and that it is not statistically significant ($p = 0.756$).

Gamma and Somers' *d*

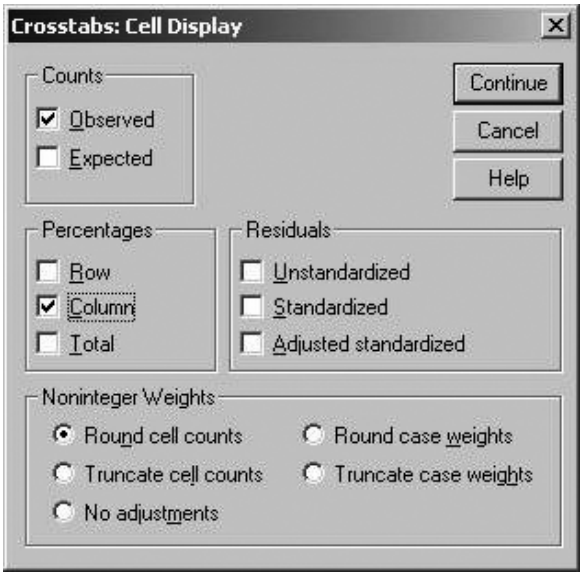
To compute other measures of association, like gamma and Somers' *d*, use the following guidelines. Gamma and Somers' *d* are both measures of association for ordinal variables. Gamma is symmetrical; Somers' *d* is asymmetrical.

ANALYZE → DESCRIPTIVES → CROSSTABS

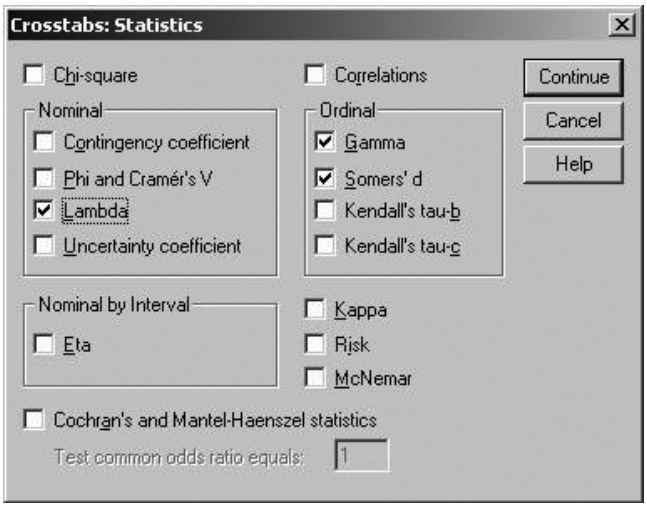


You will be given a "Crosstabs" dialog box. For this example, select "satfin" as the row variable and "satjob" as the column variable. "Satfin" is the variable representing how satisfied the respondent is with her/his financial situation. "Satjob" reveals the level of satisfaction that the respondent feels about her/his job or housework.

Now click the "Cells" button.



In the "Cells" dialog box, make sure that "Observed" counts are selected and that "Column" percentages have been requested. Now click "Continue." You will be returned to the "Crosstabs" dialog box. Here, click the "Statistics" button. You will be given the following dialog box:



In this box, select "Gamma" and "Somers' d." Click "Continue" and then click "OK" once you are returned to the original "Crosstabs" dialog box. The tables below come from the output that SPSS will create:

SATISFACTION WITH FINANCIAL SITUATION * JOB OR HOUSEWORK Cross-Tabulation

			JOB OR HOUSEWORK				
			VERY SATISFIED	MOD. SATISFIED	A LITTLE DISSAT	VERY DISSATISFIED	Total
SATISFACTION WITH FINANCIAL SITUATION	SATISFIED	Count	104	53	7	1	165
		% within JOB OR HOUSEWORK	36.6%	26.5%	12.1%	4.0%	29.1%
	MORE OR LESS	Count	117	82	22	9	230
		% within JOB OR HOUSEWORK	41.2%	41.0%	37.9%	36.0%	40.6%
	NOT AT ALL SAT	Count	63	65	29	15	172
		% within JOB OR HOUSEWORK	22.2%	32.5%	50.0%	60.0%	30.3%
Total		Count	284	200	58	25	567
		% within JOB OR HOUSEWORK	100.0%	100.0%	100.0%	100.0%	100.0%

Note that the standard cross-tabulation is produced above and gives an overview by column percents of the relationship between the two variables.

Directional Measures

			Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Nominal by Nominal	Lambda	Symmetric	.024	.023	1.054	.292
		SATISFACTION WITH FINANCIAL SITUATION Dependent	.039	.025	1.504	.133
		JOB OR HOUSEWORK Dependent	.007	.040	.177	.860
	Goodman and Kruskal tau	SATISFACTION WITH FINANCIAL SITUATION Dependent	.031	.009		.000[c]
		JOB OR HOUSEWORK Dependent	.023	.009		.000[c]
Ordinal by Ordinal	Somers' d	Symmetric	.218	.036	6.067	.000
		SATISFACTION WITH FINANCIAL SITUATION Dependent	.227	.037	6.067	.000
		JOB OR HOUSEWORK Dependent	.211	.035	6.067	.000

a. Not assuming the null hypothesis.

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

c. Based on chi-square approximation

The value for Somers' *d* is located in the value column in the row with the appropriate variable listed as the dependent variable. (Note that since Somers' *d* is asymmetrical, the two values given, where the dependent variables are different, turn out to be different.) Somers' *d* is statistically significant in this case ($p = 0.000$).

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal	Gamma	.341	.053	6.067	.000
N of Valid Cases		567			

a. Not assuming the null hypothesis.

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

Above, note the value for gamma: 0.341. It is also statistically significant ($p = 0.000$).