**RELG2R: Exercise Using SPSS to Explore Relationships Among Variables and Spuriousness**

Author:   Ed Nelson
Department of Sociology M/S SS97
California State University, Fresno
Fresno, CA 93740
Email:  **ednelson@csufresno.edu**

**Note to the Instructor:** The data set used in this exercise is gss14\_subset\_for\_classes\_RELG.sav which is a subset of the 2014 General Social Survey.  Some of the variables in the GSS have been recoded to make them easier to use and some new variables have been created.  The data have been weighted according to the instructions from the National Opinion Research Center.  This exercise uses RECODE, FREQUENCIES, and CROSSTABS in SPSS to explore relationships among variables.  In CROSSTABS, students are asked to use percentages, Chi Square, and an appropriate measure of association.  Two-variable and three-variable relationships will be explored, along with the concepts of explanation, spuriousness, and replication.  A good reference on using SPSS is *SPSS for Windows Version 23.0 A Basic Tutorial* by Linda Fiddler, John Korey, Edward Nelson (Editor), and Elizabeth Nelson.  The online version of the book is on the [**Social Science Research and Instructional Council's Website**](http://ssric.org/node/582). You have permission to use this exercise and to revise it to fit your needs.  Please send a copy of any revision to the author.  Included with this exercise (as separate files) are more detailed notes to the instructors, the SPSS syntax necessary to carry out the exercise (SPSS syntax file), and the SPSS output for the exercise (SPSS output file).  Please contact the author for additional information.

I’m attaching the following files.

* [**Data subset**](http://ssric.org/files/gss14_subset_for_classes_RELG.sav) (.sav format).
* [**Extended notes for instructors**](http://ssric.org/files/Extended_Notes_for_Instructors_for_RELG2R.docx). MS Word (.docx) format.
* [**SPSS syntax file**](http://ssric.org/files/SPSS_Syntax_for_RELG2R.sps) (.sps format).
* [**SPSS output file**](http://ssric.org/files/SPSS_Output_for_RELG2R.spv) (.spv format).
* [**This page**](http://ssric.org/files/RELG2R.docx) in MS Word (,docx) format.

**Goals of Exercise**

The goal of this exercise is to explore the relationship between religiosity and other variables using crosstabulation.  This exercise will focus on two-variable relationships and then on three-variable relationships.  The concepts of explanation, spuriousness, and replication will also be explored.  The exercise also provides practice in using several SPSS commands – RECODE, FREQUENCIES,  and CROSSTABS to explore relationships among variables.

**Part I--Recoding**

We’re going to use the General Social Survey (GSS) for this exercise.  The GSS is a national probability sample of adults in the United States conducted by the National Opinion Research Center.  For this exercise we’re going to use a subset of the 2014 GSS.  Your instructor will tell you how to access this data set which is called gss14\_subset\_for\_classes\_RELG.sav.

Religiosity is the strength of an individual’s attachment to his or her religious affiliation.  Several questions on the GSS are possible indicants of religiosity.  One of the questions asks respondents to estimate the strength of their religious affiliation.  That variable in the data set is named R8\_RELITEN.  Respondents were also asked how often they attend religious services (R6\_ATTEND) and how often they pray (R7\_PRAY).  These are all possible indicants of religiosity, but we’re going to use R6\_ATTEND in this exercise.

Before you start, run FREQUENCIES in SPSS to get the frequency distribution for R6\_ATTEND.  (See Chapter 4, Frequencies, in the online SPSS book cited on page 1 of this exercise.)

The variable R6\_ATTEND has nine categories.  Let’s start by reducing the number of categories. We’ll combine every week (7) and more than once a week (8) into one category and give this category a value of 1.  Combine once a month (4), two to three times a month (5), and nearly every week (6) into another category and give this a value of 2.  Finally, combine never (0), less than once a year (1), once a year (2), and several times a year (3) into another category and give this a value of 3.  Now we have three categories--often (1), sometimes (2), and infrequently (3).  When you use RECODE in SPSS, you can recode in two different ways—into the same variable or into different variables.  If you recode into the same variable, be careful.  It’s easier, but if you make a mistake, you will not be able to go back and recode it again.  You will have to close SPSS without saving the data set and then reopen the data set to get a fresh, clean copy of the data.  So for this exercise recode into different variables.  You’ll have to give your recoded variable a new name.  (See Chapter 3, Recode into Different Variables in the online SPSS book.)  Let’s call it R6\_ATTEND1.  To make your output more readable, add value labels for this variable.

Now that you have recoded this variable, run FREQUENCIES in SPSS to get a frequency distribution for R6\_ATTEND1.  Compare this distribution to the distribution you ran before you recoded to see if you made any mistakes.  If you made a mistake, redo this part of the exercise. If you recoded into the same variable, you will have to exit SPSS (or close your file) being sure **NOT** to save it.  Then get back into SPSS and open the gss14\_subset\_for\_classes\_RELG.sav file again.  The reason for this is that you have altered the coding of the variable and will have to get another copy of the data file to start over.  If you saved the data file, then you would have written over the original copy.  So be careful.  That’s why we said to recode into different variables in this exercise.

**Part II—Analysis of Two Variable Relationships**

Let’s start by exploring the relationship between our measure of religiosity and whether or not respondents think pornography ought to be illegal for everyone or only illegal for those under the age of 18. The variable PORN1\_PORNLAW includes the respondents’ answers to the question “Which of these statements comes closest to your feelings about pornography laws?  There should be laws against the distribution of pornography whatever the age.  There should be laws against the distribution of pornography to persons under 18.  There should be no laws against the distribution of pornography.”

Use CROSSTABS in SPSS to get the crosstabulation of R6\_ATTEND1 and PORN1\_PORNLAW. (See Chapter 5, Crosstabulation in the online SPSS book.)  Be careful when you select the independent and dependent variables.  Be sure to select the correct percentages, Chi Square, and an appropriate measure of association.  Write a paragraph or two describing the relationship between these variables using all this information.

**Part III—Analysis of Two Variable Relationships Continued**

We know that there are other variables related to R6\_ATTEND1 and PORN1\_PORNLAW.  Most research has shown that women are more likely than men to attend church.  There are good reasons to think that women are also more likely than men to feel that pornography ought to be illegal for everyone.  Women are typically the objects of pornography and are demeaned by pornography.  Men are more likely to view pornography than women.  Let’s see if we find these relationships in our data.

Use CROSSTABS to get the crosstabulation of D5\_SEX and R6\_ATTEND1 and the relationship of D5\_SEX and PORN1\_PORNLAW.  Be careful to select the proper independent and dependent variables and to ask for the correct percentages, Chi Square, and an appropriate measure of association.

Write a paragraph or two describing the relationships you find.  Were they what you expected to find?

**Part IV—Analysis of Multivariate Relationships**

Perhaps the reason that more religious people are more likely to feel that pornography ought to be illegal for everyone regardless of age is that women are more religious than men and women are also more likely to feel that pornography ought to be illegal for everyone.  If this was true and we were to take the effect of gender out of the relationship, then we would expect the relationship between R6\_ATTEND1 and PORN1\_PORNLAW to disappear (or to be reduced).  This would mean that the relationship between R6\_ATTEND1 and PORN1\_PORNLAW is a spurious relationship.  A spurious relationship is one in which there is a statistical relationship between two variables but that relationship is not causal.  The relationship can be explained away by some other variable (or variables).

To check on this, we would divide our sample into two groups – all men and all women.  In other words, sex would be our control variable.  If the relationship between R6\_ATTEND1 and PORN1\_PORNLAW goes away for both men and women (or decreases for both), then we would say the relationship was spurious and that we have explained away the relationship between religiosity and feelings about pornography laws.  This is often referred to as explanation.  In SPSS we would run a three-variable crosstab using R6\_ATTEND1 as our independent variable, PORN1\_PORNLAW as our dependent variable and D5\_SEX as our control variable.

If the relationship between R6\_ATTEND1 and PORN1\_PORNLAW does not change when we control for sex, then we would say that we have replicated the relationship.  The control variable has not affected the relationship between the independent and dependent variables.  We call this replication because the relationship between R6\_ATTEND1 and PORN1\_PORNLAW has been replicated (or repeated) for both men and women.

Run the three-variable crosstab using R6\_ATTEND1 as our independent variable, PORN1\_PORNLAW as our dependent variable and D5\_SEX as our control variable. (See Chapter 8, Crosstabulation Revisited in the online SPSS book.)  Be sure to get the correct percentages, Chi Square, and an appropriate measure of association.

Write a paragraph or two describing what you found when you controlled for sex.  Use the percentages, Chi Square, and an appropriate measure of association to help interpret your findings.

**Part V – Conclusions**

Now let’s pull all of this together.  Discuss what you learned about the relationship between R6\_ATTEND1 and PORN1\_PORNLAW.  Was it spurious due to sex?  How do you know?  If a relationship is not spurious due to one variable, can we conclude that it is not spurious at all?  Be sure to make it clear that you understand what spurious means.