**RELG1R: Exercise Using SPSS to Explore Measurement, Validity, and Relationships Among Variables**

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 COMPUTE, RECODE and IF in SPSS to create new variables, FREQUENCIES, and CROSSTABS to explore the relationships among variables.  In CROSSTABS students are asked to use percentages, Chi Square, and an appropriate measure of association.  The exercise is moderately difficult because it requires students to carefully deal with 18 different combinations of three variables in order to create the new measure of religiosity.  However, it is a good exercise to test one’s ability to think through a problem and then write the appropriate SPSS commands.  You could skip the part of the exercise that involves the creation of the new measure of religiosity since that variable (RELIGOS) is included in the data set.  Then you could go directly to Parts III and IV which deal with validity and relationships between variables.  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).  Validity is also discussed and students are asked to use the idea of construct validity to validate the measure they created.  A good reference on validity is *Reliability and Validity Assessment* by Edward G. Carmines and Richard A. Zeller (Sage, 1979).  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_RELG1R.docx). MS Word (.docx) format.
* [**SPSS syntax file**](http://ssric.org/files/SPSS_Syntax_for_RELG1R.sps) (.sps format).
* [**SPSS output file**](http://ssric.org/files/SPSS_Output_for_RELG1R.spv) (.spv format).
* [**This page**](http://ssric.org/files/RELG1R.docx) in MS Word (,docx) format.

**Goals of Exercise**

The goal of this exercise is to create a measure of religiosity.  We will also validate our measure. Validity refers to whether we are measuring what we think we are measuring.  If we can show that we are measuring what we say we are measuring, that we have validated the measure.  Once we have validated the measure, we’ll see how it is related to other variables.  The exercise also gives you practice in using several SPSS commands – COMPUTE, RECODE and IF to create new variables, FREQUENCIES, and CROSSTABS to explore the 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 survey. 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 in the GSS are possible indicants of religiosity.  One of the questions asks respondents to estimate the strength of their religious affiliation.  That variable 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.  Instead of choosing one, let’s combine all three variables into one composite variable.

Before you start, run FREQUENCIES in SPSS to get the frequency distributions for the following three variables: R8\_RELITEN, R6\_ATTEND, R7\_PRAY.  (See Chapter 4, Frequencies in the online SPSS book cited on page 1 of this exercise.)

Let’s start by reducing the number of categories for each variable by using RECODE in SPSS. The variable R8\_RELITEN records the respondent’s self-reported strength of affiliation.  The categories are strong (1), somewhat strong (2), not very strong (3), and no religion (4).  Let’s combine somewhat strong, not very strong, and no religion into one category and give that category a value of 2.  Now we have two categories--strong (1) and not strong (2).  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.  Call this one R8\_RELITEN1.  (See Chapter 3, Recoding into Different Variables in the online SPSS book.)  To make your output more readable, add value labels for this variable.

Now let’s recode R6\_ATTEND and call the recoded variable R6\_ATTEND1.  Let’s 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).  To make your output more readable, add value labels for this variable.

Finally, let’s recode R7\_PRAY and call the recoded variable R7\_ PRAY1.  Combine several times a day (1) and once a day (2) into one category and give that a value of 1.  Combine several times a week (3) and once a week (4) into another category and give that a value of 2.  Combine less than once a week (5) and never (6) into another category and give that a value of 3.  Now we have three categories--often (1), sometimes (2), and infrequently (3).  To make your output more readable, add value labels for this variable.

Now that you have recoded these variables, run FREQUENCIES in SPSS to get a frequency distribution for these three variables.  Compare these distributions to the distributions you ran before you started 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 these three variables 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—Creating a Measure of Religiosity**

Now that we have reduced the number of categories into a more manageable number, let’s create a new variable, which will be a combination of these three variables.  We’ll call this new variable REL.  To do this we’ll use the IF command in SPSS.

If individuals say they have a strong attachment to their religious affiliation (recoded value of 1 on R8\_RELITEN1), attend church often (recoded value of 1 on R6\_ATTEND1), and pray often (recoded value of 1 on R7\_PRAY1), then they are highly religious.  Let’s give these individuals a value of 1 on our new variable REL.

If individuals say they don’t have a strong attachment to their religious affiliation (recoded value of 2 on R8\_RELITEN1), attend church infrequently (recoded value of 3 on R6\_ATTEND1), and pray infrequently (recoded value of 3 on R7\_PRAY1), then they are not religious.  Let’s give these individuals a value of 3 on REL.

Everyone else will be somewhere between highly religious and not very religious.  Let’s give these individuals a value of 2 on REL.

Our new variable, REL should have three categories--1 represents those who are highly religious, 2 those who are medium in religiosity, and 3 those who are low in religiosity.  If a respondent has a missing value for any of the three variables (R8\_RELITEN1, R6\_ATTEND1, R7\_PRAY1), then he or she will automatically be assigned a system missing value for REL.

To use the IF command, click on TRANSFORM and then on COMPUTE.  (See Chapter 3, Creating New Variables Using Compute and If in the online SPSS book.)  Enter the name of the new variable (REL) in the Target Variable box.  Then click on the If button.  Select the option that says “Include if case satisfies condition” by clicking on the circle to the left of it.  Now enter your IF statement in the large box.  Think of all the possibilities.  R6\_ATTEND1 and R7\_ PRAY1 can have three values (1 or 2 or 3).  R8\_ RELITEN1 can only have two values (1 or 2).  That means there are 18 different possible combinations of these three variables (3 times 3 times 2).  Write one IF statement for each of these 18 different combinations.  That’s 18 different combinations.

This is tedious, but it’s the best way to think the problem through logically and make sure you don’t miss any possibility.   To help us do this, before we start let’s run a crosstabulation with R6\_ATTEND1 as the column variable, R7\_PRAY1 as the row variable, and R8\_RELITEN1 as the control variable.  Don’t ask for any percents since you only want to know how many cases are in each combination.  Each cell in the table represents one of the 18 possible combinations.  Print out the table and write the combination in each cell.  For example, for the cell that represents R6\_ATTEND1 = 1 and R7\_PRAY1 = 1 and R8\_RELITEN1 = 1, write 111.  For the cell that represents R6\_ATTEND1 = 2 and R7\_PRAY1 = 1 and R8\_RELITEN1 = 1, write 211.  Do this for all 18 combinations.  Now number these combinations from 1 to 18 using 1 for cell 111 and 18 for cell 332.  Use 2 through 17 for the other cells.

Now you’re ready to write the IF statements.  After you have entered your IF statement, click on continue and enter the numeric value you want to assign to the REL variable in the Numeric Expression box and click on OK.  The numeric value is the number you assigned to each category (i.e., 1 through 18).  Do this for each of the 18 possible combinations.  After each of the combinations (except the first time), SPSS will ask you if it is OK to change the existing variable.  Click on Yes.  Once you have done this the first time, it will go faster since SPSS will remember what you entered before and you can modify what you entered previously. To make your output more readable, add value labels for this variable.

Now we need to recode the REL variable we just created.  Recode 1 as 1 and 18 as 3.  Recode 2 through 17 as 2.  Let’s call this new variable REL1.  Assign value labels to these recoded categories (i.e., 1 is high in religiosity, 2 is medium in religiosity, 3 is low in religiosity).

You’re might have some problems doing this part of the exercise.  Your instructor will help you if you are having problems.

Run FREQUENCIES in SPSS to get a frequency distribution for your new variable, REL1.  There is another variable in the data set, RELIGOSR, which should be identical to your variable (REL1).  Run FREQUENCIES for RELIGOSR and compare the two distributions.  If they are not the same, you made a mistake and will have to start over.  See your instructor if you can’t figure out your mistake.

**Part III--Validity**

We created a variable, REL1, which we claim is a measure of religiosity.  But how do we know it measures religiosity?  This is a question of validity.  Are we measuring what we say we are measuring?

What we can do is look for variables that are likely to be closely related to religiosity and see if they are strongly related.  For example, if our measure is a valid measure of religiosity, then we would expect highly religious individuals to be more likely to believe in life after death than less religious individuals.  The variable R22\_POSTLIFE tells us whether respondents say they believe in life after death.  We would also expect highly religious respondents to be less likely to have seen an X-rated movie in the last year (variable is S11\_XMOVIE).

If our new variable (REL1) behaves as we expect it to, then we can claim that we have demonstrated its validity.  This is called construct validity.  If it does not behave as we expect it to, then it’s a little more complicated.  It may be that our measure is not valid.  Or it may be that our expectations are wrong.  Or it may be there is something else wrong with our survey.  But the important point is that if REL1 behaves as we expect it to, then we have evidence of the construct validity of our new measure.

To check on the validity of your new measure (REL1), run two crosstabulations—one for REL1 and R22\_POSTLIFE and another for REL1 and S11\_XMOVIE.  (See Chapter 5, Crosstabulations in the online SPSS book.)  Think carefully about which should be the independent variable and which should be the dependent variable.  Be sure to get the appropriate percents, Chi Square, and an appropriate measure of association.  Write a paragraph indicating whether you think your measure of religiosity, REL1, is a valid measure.  Indicate your reasoning.

**Part IV--Analysis**

Now that we have created a measure of religiosity (REL1) and have some evidence that it is valid, we can explore its relationship with other variables.  Let’s look in the data set for some measures of the respondent’s opinion on social issues.  You can click on Utilities in the menu bar of SPSS and then on Variables in the Utilities’ menu to see a list of the variables in the data set.  There are questions on the legalization of marijuana (M1\_GRASS), homosexual relations (S7\_HOMOSEX), suicide (SUI1\_SUICIDE1), allowing incurable patients to die (SUI5\_LETDIE1), pornography laws (PORN1\_PORNLAW), sex before marriage (S9\_PREMARSX) and others.  Select one other variable that you think ought to be related to religiosity and complete the following steps:

1.         Write a hypothesis stating how you expect religiosity (REL1) to be related to this variable.

2.         Write a paragraph or two that indicates why you think your hypothesis is true.  In other words, write an argument in which your hypothesis is the conclusion.

3.         Use SPSS to run the crosstabulation of REL1 and your variable.  Think about which is the independent and dependent variable.  Remember to get the correct percentages.  Use Chi Square and an appropriate measure of association.

4.         Write a paragraph interpreting the table that SPSS gave you and indicate whether the data support your hypothesis.  Use the percentages, Chi Square and the measure of association to help you interpret the table.