**GENDER2G: Exercise Using SPSS to Explore Gender Differences in Voting Controlling for Income**

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\_GENDER\_DIFFERENCES.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 in SPSS to recode some of the variables, FREQUENCIES to get frequency distributions, and CROSSTABS to explore the relationships among variables.  In CROSSTABS students are asked to use percentages, Chi Square, and an appropriate measure of association.  You could skip the part of the exercise that involves recoding since those variables are included in the data set.  Then you could go directly to the parts of the exercise that deal with 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 Center'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). These, of course, will need to be removed as you prepare the exercise for your students.  Please contact the author for additional information.

I’m attaching the following files.

* [**Data subset**](http://ssric.org/files/gss14_subset_for_classes_GENDER_DIFFERENCES.sav) (.sav format).
* [**Extended notes for instructors**](http://ssric.org/files/Extended_Notes_for_Instructors_for_GENDER2G.docx). MS Word (.docx) format.
* [**SPSS syntax file**](http://ssric.org/files/SPSS_Syntax_for_GENDER2G.sps) (.sps format).
* [**SPSS output file**](http://ssric.org/files/SPSS_Output_for_GENDER2G.spv) (.spv format).
* [**This page**](http://ssric.org/files/GENDER2G.docx) in MS Word (.docx) format.

**Goals of Exercise**

The goal of this exercise is to explore differences between men and women in voting controlling for income.  The exercise also gives you practice in using several SPSS commands – RECODE, FREQUENCIES, and CROSSTABS.

**Part I—Adding Income into the Analysis**

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\_GENDER\_DIFFERENCES.sav.

In Exercise Gender1G we looked at gender differences in a number of areas including political affiliation, political orientation, political interest, and voting.  We made some interesting discoveries.

* Females were more likely than males to say they were Democrats and independents while males were more likely to be Republican.
* Males were more likely than females to be very interested in politics while females were more likely to be not very or not at all interested.
* Males and females were very similar in their political outlook (i.e., liberal, moderate, conservative).
* Females were a little more likely to say they voted in 2008 but males and females were about equally likely to say they voted in 2012.
* The biggest difference was in whom respondents said they voted for in 2008 and 2012.  Females were about 9 percentage points more likely to vote for Obama in 2008 and 6 percentage points more likely in 2012.

This is an interesting beginning but now we want to add other variables into the analysis.  Income would be an interesting variable to consider since income might be related to both gender and voting.  But before we do that let’s look at our two possible measures of income:

* F1\_INCOME06 which is a measure of family income and
* F3\_RINCOM06 which is a measure of respondent’s income**[[1]](http://ssric.org/node/531/edit%22%20%5Cl%20%22_ftn1%22%20%5Co%20%22)**.

By the way, the “06” at the end of each variable’s name doesn’t refer to the year but rather to the coding scheme for income.  Just think of this as the reference number for the particular coding scheme used in the GSS.  Another important point is that income is not measured in dollars but has been classified into 25 categories.  Take a look at the categories by clicking on “Utilities” in the SPSS menu bar and then clicking on “Variables” in the drop-down menu.  Scroll down until you see F1\_INCOME06 and click on it.**[[2]](http://ssric.org/node/531/edit%22%20%5Cl%20%22_ftn2%22%20%5Co%20%22)**  Now you will see the list of 25 categories.  Then do the same for F3\_RINCOM06.

Twenty-five categories are too many for crosstabulation so let’s reduce the number of categories by recoding both income variables.  Run FREQUENCIES in SPSS for the two income variables. (See Frequencies, in Chapter 4 in the online SPSS book mentioned on page 1.)  Let’s reduce the number of categories to three and try to choose categories that divide our distribution into three categories that have approximately the same number of cases or as close as we can get to an equal distribution.  Here’s how we are going to define the categories.

* F1\_INCOME06 (family income)
	+ Category 1 – under $35,000 which would be values 1 through 16
	+ Category 2 -- $35,000 to under $75,000 which would be values 17 through 20
	+ Category 3 -- $75,000 or more which would be values 21 through 25
* F3\_RINCOM06 (respondent’s income)
	+ Category 1 – under $22,500 which would be values 1 though 13
	+ Category 2 -- $22,500 through under $50,000 which would be values 14 through 18
	+ Category 3 -- $50,000 or more which would be values 19 through 25

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 these variables F1\_INCOME061 and F3\_RINCOM061 where the 1 at the end of the variable name indicates this is a recoded variable.  (See Chapter 3, Recoding into Different Variables in the online SPSS book.)  To make your output more readable, add value labels for these variables.

Compare the unrecoded frequency distributions with the recoded frequency distributions to make sure you recoded correctly.  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\_GENDER\_DIFFERENCES.sav file again.  The reason for this is that you have altered the coding of these 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.

Now that you have recoded both measures of income let’s see if we were right that they are related to gender and voting.  Run CROSSTABS in SPSS to produce four crosstabulations – one for F1\_INCOME061 and D5\_SEX, a second for F3\_RINCOM061 and D5\_SEX, a third for F1\_INCOME061 and P6\_PRES12, and a fourth for F3\_RINCOM061 and P6\_PRES12.  (See Chapter 5, CROSSTABS, in the online SPSS book.)  You’ll need to decide which of these variables you want to use as your independent variable and which you want to use as your dependent variable.  The dependent variable is what you are trying to explain and the independent variable is the variable that you think will help you explain the variation in your dependent variable.  Put the independent variable in the column and the dependent variable in the row of your table.  If you do this, you will always want to tell SPSS to compute the column percents.  Also tell SPSS to compute Chi Square, and an appropriate measure of association.

Write a paragraph describing the relationship between both measures of income and sex and between the two measures of income and voting.  Were males more or less likely than females to have higher or lower income and were lower income groups more or less likely to vote for Obama?   Use the percents, Chi Square, and the measure of association to help you describe this relationship.

**Part II – Controlling for Income**

Now that we have recoded our two income variables and have verified that both measures of income are related to gender and voting we’re ready to bring income our analysis.  Up until now we have only looked at variables two at a time.  Now we’re going to consider three variables simultaneously.  Our dependent variable is P6\_PRES12 because we’re trying to explain why some people voted for Obama and others voted for Romney.  Our independent variable is D5\_SEX because we think that gender might help us explain voting behavior.  Our third variable, income, will become our control variable.  We’re going to control for income by holding it constant.  That means we’re going to divide our sample into three groups – those low in income, those with middle income, and those high in income.  Then we’re going to look at the crosstabulation of D5\_SEX and P6\_PRES12 separately for each of the three income categories.  That means we’ll have three crosstabs – one for each of the income categories – and we’ll call these partial tables since each table contains part of the data.

Telling SPSS to run a three-variable table is very similar to running a two-variable table but with one important difference.  Put your dependent variable in the row and your independent variable in the column and tell SPSS to compute the column percents, Chi Square, and an appropriate measure of association.  This is exactly what you did in Part 1.  Now put your control variable – F1\_INCOME061 – in the third box down on the CROSSTABS dialog box and click on “OK.”  Then repeat this process but this time control for F3\_RINCOM061.

The output will have the independent variable (D5\_SEX) in the column, the dependent variable (P6\_PRES12) in the row, and the control variable (i.e., one of your two recoded income variables) along the left margin.  You’ll have four tables stacked on top of each other.  The top table will contain only the lowest income category, the second table down will contain only the middle income group and the third table down will have only the highest income category.  The bottom table will be your two-variable table which includes all the respondents regardless of income.**[[3]](http://ssric.org/node/531/edit%22%20%5Cl%20%22_ftn3%22%20%5Co%20%22)**  You’ll also have the Chi Square tables and tables with your Gamma values.   Notice that there is a Chi Square and a Gamma for each of your partial tables and for your two-variable table.

**Part III – Interpreting the Partial Tables**

Let’s define the gender gap for voting as the percent of males who voted for Obama minus the percent of females who voted for Obama.  Since we have three partial tables (i.e., one for each income category), we will have three gender gaps.  Calculate the gender gap for both income variables (i.e., family income and respondent’s income).  To help you there is a table you can fill in at the end of this exercise.  Enter the gender gap, the significance value for Chi Square, and the value of Gamma in the table.

Write a paragraph summarizing what you discovered when you looked at the gender gaps.  Use the Chi Squares and the Gammas that SPSS calculated to help you.  Did you find a pattern for the gender gaps for each income measure?  What do the Chi Squares and Gammas tell you?

**Part IV – Repeating the Analysis for the 2008 Election**

You should have found a different pattern to the gender gaps for the two income variables.  It’s hard to come up with a good explanation for this finding.  One possibility is that this was just a random occurrence.  However, we have a way to check on that.  Let’s repeat the analysis we did in Parts 2 and 3 but this time we’ll use the 2008 election (P5\_PRES08).  Have SPSS create the tables for you and then fill in the table which you will find at the end of this exercise.

When you look at the gender gaps what you should see is two different patterns for the two income measures.  For respondent’s income you should see a large and statistically significant difference between men and women for the high income category and a smaller and non-significant difference for the lower and middle income categories.  However, for family income there is a large and significant difference for the low income category and a smaller and non-significant difference for the middle and high income categories.  Moreover, these patterns occur for both the 2008 and 2012 presidential elections suggesting that they aren’t just random occurrences.

**Gender Gaps for 2012 Election**

|  |  |  |  |
| --- | --- | --- | --- |
| Respondent’s Income | Gender Gap | Chi Square | Gamma |
| Low income category |  |  |  |
| Middle income category |  |  |  |
| High income category |  |  |  |
| All incomes together |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Family Income | Gender Gap | Chi Square | Gamma |
| Low income category |  |  |  |
| Middle income category |  |  |  |
| High income category |  |  |  |
| All incomes together |  |  |  |

**Gender Gaps for 2008 Election**

|  |  |  |  |
| --- | --- | --- | --- |
| Respondent’s Income | Gender Gap | Chi Square | Gamma |
| Low income category |  |  |  |
| Middle income category |  |  |  |
| High income category |  |  |  |
| All incomes together |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Family Income | Gender Gap | Chi Square | Gamma |
| Low income category |  |  |  |
| Middle income category |  |  |  |
| High income category |  |  |  |
| All incomes together |  |  |  |

**[[1]](http://ssric.org/node/531/edit%22%20%5Cl%20%22_ftnref1%22%20%5Co%20%22)** Variable names are in all capitals.

**[[2]](http://ssric.org/node/531/edit%22%20%5Cl%20%22_ftnref2%22%20%5Co%20%22)** If you don’t see the variable names but instead sees the variable labels, close the list of variables and click on “Edit” in the menu bar and then click on “Options.”  In the “General” tab click on “Display names” and “Alphabetical” in the upper-left and then click on “OK” once or twice depending on what was previously selected.

**[[3]](http://ssric.org/node/531/edit%22%20%5Cl%20%22_ftnref3%22%20%5Co%20%22)** If you compute the gender differences for the two-variable tables in Part 1 and compare them to the gender differences for this two-variable table they will be close but not identical.  This might be confusing.  The reason is that CROSSTABS includes only those cases with valid information (i.e., not missing data) for **all** variables in the table.  When you add the control variable into the analysis, then CROSSTABS will also eliminate those cases with missing information on this variable.  That means that you are using a different subset of cases for each of these two-variable tables.