# Exercises on Data Analysis for Gender Differences

# Edward Nelson

# California State University, Fresno

# Table of Contents

# Page

# Preface 3

# Exercise 1 – Exploring Gender Differences in Political Attitudes and Behavior 4 Exercise 2 – Gender Differences in Voting Controlling for Family Income 8

Exercise 3 – Exploring Gender Differences on Social Issues 12

Exercise 4 -- Exploring Gender Differences on Gun Control by Adding Control Variables 15

Appendix – Codebook for Data Set 18

# Preface

The data set used in this exercise is the 2018 General Social Survey. The name of the data set is GSS18GENDER.SAV which is a subset of the full data set. 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.  These exercises use SPSS to analyze the data but you could also easily use PSPP.  A good reference for using SPSS is [*SPSS Statistics (Version 26): A Basic Tutorial*](http://www.ssric.org/node/696)by Edward Nelson and John Korey.

The focus of these exercises is on two-variable and three-variable data analysis. The exercises revolve around respondents' opinion about gun permits.

The exercises do not explain how to use SPSS. For that you could use the [SPSS tutorial](http://www.ssric.org/node/696) referred to throughout the exercises. I have added to the data set the variables that students are asked to create. The names of these variables all start with ZZ. Students are given the option of recoding political views in two different ways. Both options are included in the ZZ variables. That means you could skip some of the SPSS procedures by referring students to these variables in the data set. I have also included in the data set some variables not used in the exercises so you could develop your own exercises around these variables.

You have permission to use these exercises and to revise them to fit your needs which would include adding materials of your own or deleting parts of the exercises that you don't want to use.

# Exercise 1 – Exploring Gender Differences in Political Attitudes and Behavior

## Goal of Exercise

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

## Part I—Political Attitudes and Behavior

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 2018 GSS survey. Your instructor will tell you how to access this data set which is called GSS18GENDER.SAV. You can also download the data by clicking on this [link](http://ssric.org/node/529).

Political attitudes and behavior are some of the most widely studied areas in sociology, political science, psychology, economics, and history.  To get a sense of the research that has been done, go to [**Google Scholar**](https://scholar.google.com/) and enter “political attitudes and behavior” in the search box.

Some of the research has focused on differences by gender, income, education, race, age, and religion.  One of the reasons why social scientists are so interested in these individual characteristics is that they serve as markers of a person’s position in the social structure.  For example, men and women occupy different positions in the social structure that affect their life chances in such areas as access to quality health care, higher education, and legal representation.  This is also true for whites and blacks, older and younger, highly educated and less educated, and so on.

In this exercise we’re going to consider gender differences in political attitudes and behavior.  Our data set provides us with several different measures` of political attitudes and behavior including:

* political affiliation (PARTYID),
* political orientation (POLVIEWS), and
* voting (PRES12, PRES16, VOTE12, VOTE16).[[1]](#footnote-1)

## Part II – Gender Differences in Political Affiliation

Political affiliation or identification refers to the political party with which a person identifies.  The GSS uses a series of questions to determine a person’s political affiliation.

* “Generally speaking, do you usually think of yourself as a Republican, Democrat, Independent, or what?”
* Other questions ask how strong a Democrat or Republican a person is and whether independents lean toward the Democrats or toward the Republicans or do not lean either way.

The variable we’re going to use is named PARTYID.  Run FREQUENCIES in SPSS to get the frequency distribution for this variable.  (See Chapter 4, FREQUENCIES, in the [online SPSS tutorial](http://ssric.org/node/696).)  The categories are strong Democrat (0), not strong Democrat (1), independent, near Democrat (2), independent (3 –  i.e., does not lean either way), independent near Republican (4), not strong Republican (5), strong Republican (6), other party (7), don’t know (8), and no answer (9)[[2]](#footnote-2). The last three values (7, 8, and 9) have already been defined as missing values.[[3]](#footnote-3)

We can use this variable in various ways to describe the political affiliation of the respondents in our sample.

* We could combine the strong and not strong categories and then combine the three values for independents giving us the following categories:
  + Democrat (values 0 and 1),
  + Independent (values 2, 3, and 4),
  + Republican (values 5 and 6).
* We could also combine all those who say they are Democrats and the independents that lean toward the Democrats into one category and then do the same for Republicans.  This would leave those who are independent and don’t lean either direction as independents and would give us the following categories:
  + Democrat (values 0, 1, and 2),
  + Independent (value 3),
  + Republican (values 4, 5, and 6).

There are good arguments that can be made for either way of combining categories.  We’re going to use the second approach in this exercise.  Use RECODE in SPSS to reduce the number of categories using this second approach.  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.  Let's name this variable PARTYID1.  (See Chapter 3, Recoding into Different Variables in the online SPSS tutorial.)  To make your output more readable, add value labels for this variable.

Compare the unrecoded frequency distribution (PARTYID) with the recoded frequency distribution (PARTYID1) to make sure you recoded correctly.  If you made a mistake, redo this part of the exercise.

Write a paragraph using the recoded frequency distribution to describe how Americans identify with the two major political parties in our country.  Use the percents to indicate whether more respondents see themselves as Democrats or as Republicans.  How large is the independent category?   Keep in mind that these are the respondents who are “true” independents.  In other words, they don’t lean either way.  How does this category compare in size to that of Democrats and Republicans?  What are the implications of these findings for national political elections?

Now we want to compare men and women in terms of their political affiliation so we can look at gender differences.  Run CROSSTABS in SPSS to produce a crosstabulation of two variables – SEX and PARTYID1.  (See Chapter 5, CROSSTABS, in the online SPSS tutorial.)  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 the respondent’s gender and political affiliation.  Were males more or less likely than females to see themselves as Republican or Democrat or independent?  How big was the gender difference for party affiliation where gender difference is the percent of males who are Democrats minus the percents of females who are Democrats?  Use the percents, Chi Square, and the measure of association to help you describe this relationship.

## Part III – Gender Differences in Political Orientation

Political orientation refers to whether people see themselves as politically liberal, moderate (i.e., middle-of-the-road), or politically conservative.  The question asked in the GSS is as follows:  “We hear a lot of talk these days about liberals and conservatives. I'm going to show you a seven-point scale on which the political views that people might hold are arranged from extremely liberal--point 1--to extremely conservative-- point 7. Where would you place yourself on this scale?”  The categories are extremely liberal (1), liberal (2), slightly liberal (3), moderate (4), slightly conservative (5), conservative (6), extremely conservative (7), not applicable (0 – i.e., was not asked), don’t know (8), and no answer (9).  The last three values (0, 8, and 9) have already been defined as missing values.

Run FREQUENCIES in SPSS to get a frequency distribution for the variable POLVIEWS.  What are the different ways we could recode this variable to reduce the categories to liberal (1), moderate (2), and conservative (3)?  Think about whether you want to include those who are slightly liberal and slightly conservative in the moderate category or whether you want to consider them liberal and conservative.   How does your decision affect your interpretation of the frequency distribution?

Now use RECODE in SPSS to carry out the recoding that you decided to use.  Call the recoded variable POLVIEWS1.  Compare the frequency distributions for the unrecoded and recoded variables to make sure you recoded correctly.  If you made a mistake, redo this part of the exercise.  To make your output more readable, add value labels for this variable.

Now we want to compare men and women in terms of their political orientation so we can look at gender differences.  Run CROSSTABS in SPSS to produce a crosstabulation of the two variables – SEX and POLVIEWS1.  Think carefully about which variable is your independent variable and which is your dependent variable.  Put the independent variable in the column and the dependent variable in the row of your table.  Tell SPSS to compute the appropriate percents, Chi Square, and an appropriate measure of association.

Write a paragraph describing the relationship between the respondent’s gender and political orientation.  Were males more or less likely than females to see themselves as liberal or conservative or moderate?  How big was the gender difference for political orientation?  Use the percents, Chi Square, and the measure of association to help you describe this relationship.

## Part IV – Gender Differences in Voting

Now we’re going to consider voting.  Keep in mind that we’re only looking at respondents’ reports of whether they voted and whether they voted for a particular candidate.  There is lots of evidence that indicates that people over report whether they voted and voting for the winning candidate.  Voting is clearly the socially desirable response and social desirability is a powerful influence on how people answer survey questions.

The GSS asks respondents who they voted for in the 2012 (Obama vs. Romney) and the 2016 (Trump vs. Clinton) presidential elections.  These variables are named PRES12 and PRES16.  They also asked respondents whether they voted in these elections. These variables are named VOTE12 and VOTE16.  Run FREQUENCIES in SPSS to get the frequency distributions.  A good source of information for the actual percent of people who voted in these elections can be found on [**the United States Elections Project's website.**](http://www.electproject.org/home/voter-turnout/voter-turnout-data)  Compare the percent who said they voted in the GSS with the actual voter turnout.  The Encyclopedia Britannica reports the percent of votes cast for Obama in 2012 and Trump in 2016 on their [**website.**](https://www.britannica.com/topic/United-States-Presidential-Election-Results-1788863)  Compare the actual percent that voted for Obama and Trump with the percent who said they voted for these candidates in the GSS.  Write a paragraph describing your findings and what might account for these differences.

Now we want to compare men and women in terms of voting so we can look at gender differences.  Run CROSSTABS in SPSS to produce four crosstabulations of gender with each of these four variables (PRES12, PRES16, VOTE12, VOTE16).  Think carefully about which variable is your independent variable and which is your dependent variable.  Put the independent variable in the column and the dependent variable in the row of your table.  Tell SPSS to compute the appropriate percents, Chi Square, and an appropriate measure of association.

Write a paragraph describing the relationship between the respondent’s gender and voting.  Were males more or less likely than females to vote?  Were males more or less likely to vote for the Democratic candidate for president?  How big were the gender differences for voting?  Use the percents, Chi Square, and the measure of association to help you describe this relationship.[[4]](#footnote-4)

## Part V – Summary

Write a paragraph summarizing what you learned about gender differences in political attitudes and behavior.  What are the implications of these findings for national political elections?

# Exercise 2 – Gender Differences in Voting Controlling for Family Income

## Goal 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 Family 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 2018 GSS survey. Your instructor will tell you how to access this data set which is called GSS18GENDER.SAV. You can also download the data by clicking on this [link](http://ssric.org/node/529).

In Exercise 1 we looked at gender differences in a number of areas including political affiliation, political orientation, and voting.  We made some interesting discoveries.

* Females were more likely than males to say they were Democrats while males were more likely to be Republican.
* 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 2012 but males were a little more likely to say they voted in 2016. However, the Chi Square tests were not significant in either election.
* The biggest difference was in whom respondents said they voted for in 2012 and 2016.  Females were 6 percentage points more likely to vote for Obama in 2012 and 15 percentage points more likely to say they voted for Clinton in 2016.

This is an interesting beginning but now we want to add other variables into the analysis.  Family income would be a good variable to consider since income is probably also related to voting.  But before we do that, let’s look at our measure of family income (INCOME16).[[5]](#footnote-5)

* OM06 which is a measure of respondent’s inco

By the way, the “16” at the end of the 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 26 categories.  Use FREQUENCIES to run a frequency distribution for INCOME16 so you can see these categories. (See Frequencies, in Chapter 4 in the [online SPSS tutorial](http://ssric.org/node/696).) 

Twenty-six categories are too many for crosstabulation so let’s reduce the number of categories by recoding income.  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 are as close as we can get to an equal distribution. Here’s how we are going to define the categories.

* Category 1 – under $40,000 which would be values 1 through 17
* Category 2 -- $40,000 to under $90,000 which would be values 18 through 21
* Category 3 -- $90,000 or more which would be values 22 through 26

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 variable INCOME161 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 tutorial.)  To make your output more readable, add value labels for these variables.

Compare the unrecoded frequency distribution with the recoded frequency distribution to make sure you recoded correctly.  If you made a mistake, redo this part of the exercise.

Now that you have recoded income, let’s see if both income and gender are related to voting.  Run CROSSTABS in SPSS to produce two crosstabulations – one for INCOME161 and PRES16 and a second crosstab for SEX and PRES16.  (See Chapter 5, CROSSTABS, in the online SPSS tutorial.)  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 family income and voting and the relationship between sex and voting.  Were males more or less likely than females to vote for Trump and were lower income groups more or less likely to vote for Trump?   Use the percents, Chi Square, and the measure of association to help you describe this relationship.

## Part II – Controlling for Family Income

Now that we have recoded family income and have verified that both gender and income are related to 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 PRES16 because we’re trying to explain why some people voted for Trump and others voted for Clinton.  Our independent variable is SEX because we think that gender might help 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 SEX and PRES16 separately for each of the three income categories.  That means we’ll have three crosstabs – one for each of the income categories. 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 – INCOME161 – in the third box down on the CROSSTABS dialog box and click on OK.

The output will have the independent variable (SEX) in the column, the dependent variable (PRES16) in the row, and the control variable (INCOME161) 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.[[6]](#footnote-6)

## Part III – Interpreting the Partial Tables

Let’s define the gender gap for voting as the percent of males who voted for Trump minus the percent of females who voted for Trump.  Since we have three partial tables (i.e., one for each income category), we will have three gender gaps.  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 the measure of association in the table.

Write a paragraph summarizing what you discovered when you looked at the gender gaps.  Use the Chi Squares and the measure of association that SPSS calculated to help you. 

## Part IV – Repeating the Analysis for the 2012 Election

Let’s repeat the analysis we did in Parts 2 and 3 but this time we’ll use the 2012 election (PRES12).  Have SPSS create the tables for you and then fill in the table at the end of this exercise. This time we'll define the gender gap as the percent of males who voted for Obama minus the percent of females who voted for Romney.

When you look at the gender gaps what you should see is two different patterns for the two presidential elections.  Write a paragraph explaining how your results differ for the 2012 and 2016 presidential elections. Then write a second paragraph suggesting why you think you found different gender gaps for the two elections.

**Gender Gaps for 2016 Election**

|  |  |  |  |
| --- | --- | --- | --- |
| Family Income | Gender Gap | Chi Square | Measure of Association |
| Lowest income category |  |  |  |
| Middle income category |  |  |  |
| Highest income category |  |  |  |
| All incomes together |  |  |  |

**Gender Gaps for 2012 Election**

|  |  |  |  |
| --- | --- | --- | --- |
| Family Income | Gender Gap | Chi Square | Measure of Association |
| Lowest income category |  |  |  |
| Middle income category |  |  |  |
| Highest income category |  |  |  |
| All incomes together |  |  |  |

## Part V – Summary

What did you discover about the effect of family income on the gender gap in the 2012 and 2016 presidential elections? What does this suggest about the importance of repeating our analysis in different time periods?

# Exercise 3 – Exploring Gender Differences on Social Issues

## Goal of Exercise

The goal of this exercise is to explore differences between men and women on the social issues of abortion, capital punishment, and gun control.  The exercise also gives you practice in using two SPSS commands – FREQUENCIES and CROSSTABS.

## Part I—Social Issues

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 2018 GSS survey. Your instructor will tell you how to access this data set which is called GSS18GENDER. You can also download the data by clicking on this [link](http://ssric.org/node/529).

In Exercises 1 and 2 we looked at gender differences in voting and discovered that females were more likely than males to vote for the Democrat candidate for president.  The Center for the American Woman and Politics at Rutgers – The State University of New Jersey has [data on the gender gap in presidential voting from 1980 through 2016](https://cawp.rutgers.edu/sites/default/files/resources/ggpresvote.pdf) that show that females were more likely than males to vote for the Democrat candidate in all elections from 1980 to the present.

In this exercise we’re going to look at three social issues – abortion, capital punishment, and gun control.

* The GSS asked respondents whether they thought that it should be legal for a woman to obtain an abortion in seven scenarios:[[7]](#footnote-7)
  + ABANY – for any reason,
  + ABDEFECT – if there was a strong chance of a birth defect,
  + ABHLTH – if the woman’s health was seriously endangered,
  + ABNOMORE – if the woman was married and wanted no more children,
  + ABPOOR – if the woman was low income and couldn’t afford more children,
  + ABSINGLE – if the woman was not married, and
  + ABRAPE – if the woman was pregnant as a result of rape.
* Respondents were asked if they favored or opposed the death penalty for murder – CAPPUN.
* The GSS also asked if they favored or opposed gun permits – GUNLAW.

Run FREQUENCIES in SPSS to get frequency distributions for all nine of these variables.  (See Frequencies, in Chapter 4 in the [online SPSS tutorial](http://ssric.org/node/696).)  Write a paragraph summarizing what you discovered about attitudes towards these issues.

## Part II – Gender Differences

Before we look at gender differences for these social issues let’s think about what we would expect to find.  A hypothesis states the relationship you expect to find between two variables.  For example, for abortion we might hypothesize that women would be more likely than men to think that abortion should be legal **or** we could hypothesize that men would be more likely than women to feel this way **or** we could suggest that there are no differences between men and women.  We would state our hypothesis before we looked at the data.

We would also want to explain why we think our hypothesis will be true.  In other words, we would write an argument for which the hypothesis is the conclusion to the argument.  If we thought that women were more likely than men to think abortion should be legal, we might argue that women would want to have control over their bodies and therefore would be more likely to think that abortion should be legal.  But if we thought that women were less likely than men to think that abortion should be legal we might point to the fact that by every measure of religiosity women are more religious than men and those who are more religious tend to be more opposed to abortion.  Therefore women would be less likely to feel that abortion should be legal.

Now that you understand what a hypothesis and an argument is, write three paragraphs that state your hypothesis and argument for the issues of abortion, capital punishment, and gun control.  Be sure to indicate whether you think that women are more or less likely than men to think that abortion should be legal and to favor or oppose capital punishment and gun control and why you think that is the case.

## Part III – Gender Differences for Abortion, Capital Punishment, and Gun Control

Run CROSSTABS in SPSS to determine how men and women (SEX) feel about these social issues.  (See Chapter 5, CROSSTABS, in the online SPSS book.)  Use all seven variables for abortion (i.e., ABANY, ABDEFECT, ABHLTH, ABNOMORE, ABPOOR, ABSINGLE, ABRAPE), capital punishment (CAPPUN), and gun control (GUNLAW).  You’ll need to decide what you want to use as your independent variable and what 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 gender and these nine variables that are our measures of how people feel about these three social issues.  Were males more or less likely than females to think that abortion should be legal, to favor capital punishment, and to favor gun permits?   Use the percents, Chi Square, and the measure of association to help you describe these relationships.  For abortion be aware that the gender differences might be different for the various measures of how people feel about abortion.  To help you organize your findings, there is a table at the end of this exercise where you can fill in the gender differences, the significance of Chi Square, and the value of the measure of association.  We’ll define the gender gap as the percent of males minus the percent of females that think abortion should be legal and that favor capital punishment and gun permits.

## Part IV – Summary

Look back at the hypotheses and arguments you wrote in Part 2.  Do the gender differences support your hypotheses?  In the case of abortion, keep in mind that the results for some of the seven measures might support the hypothesis while the others might not support it.  What do the data suggest about your arguments?

**Gender Gaps for Social Issues**

|  |  |  |  |
| --- | --- | --- | --- |
| Abortion | Gender Gap | Chi Square | Measure of Association |
| For any reason |  |  |  |
| Strong chance of birth defect |  |  |  |
| Woman's health seriously endangered |  |  |  |
| Not married and wants no more children |  |  |  |
| Low income and can't afford more children |  |  |  |
| Not married |  |  |  |
| Pregnant as a result of rape |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Capital Punishment | Gender Gap | Chi Square | Measure of Association |
| Favor death penalty for murder |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Gun Control | Gender Gap | Chi Square | Measure of Association |
| Favor gun permits |  |  |  |

# Exercise 4 -- Exploring Gender Differences on Gun Control by Adding Control Variables

## Goal of Exercise

The goal of this exercise is to explore differences between men and women in their feelings about gun control by adding other variables into the analysis.  The exercise also gives you practice using CROSSTABS in SPSS to explore relationships among variables.

## Part I—Gun Control

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 2018 GSS survey. Your instructor will tell you how to access this data set which is called GSS18GENDER.SAV. You can also download the data by clicking on this [link](http://ssric.org/node/529).

In Exercise 3 we looked three social issues – abortion, capital punishment, and gun control – and compared how men and women felt about these issues.  In this exercise we’re going to focus on gun control and expand our analysis by adding in control variables.  This process of adding other variables into the analysis is often referred to as elaboration.

Let’s start by running CROSSTABS in SPSS to determine how men and women (SEX) feel about gun control (GUNLAW)[[8]](#footnote-8).  (See Chapter 5, CROSSTABS, in the [online SPSS tutorial](http://ssric.org/node/696).)  You’ll need to decide which variable 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 gender and how respondents feel about gun control.  Were males more or less likely than females to favor gun permits and by how much?  Use the percents, Chi Square, and the measure of association to help you describe these relationships.

## Part II – Elaborating a Relationship by Adding Other Variables

In Part 1 you discovered that women were more likely than men to favor gun permits.  The gender gap was the percent of men who favored gun permits minus the percent of women who favored permits or 66.0% - 77.0% or -11.0 percentage points.

But why were men more likely to oppose gun permits?  One possible explanation is that men are more likely to own guns and those who own guns are more likely to oppose gun permits.  If this explanation is correct then we would expect the gender gap between men and women in opposing gun permits to disappear or at least decrease considerably when we hold gun ownership constant.  In other words, if the gender gap is because men are more likely to own guns then when we compared men and women who both owned guns, the gap ought to decrease.  And the same result ought to hold when we compared men and women who both didn’t own guns.  This is what we mean by holding gun ownership constant.

But we’re getting ahead of ourselves.  Let’s check to make sure that men are more likely to own guns by running CROSSTABs in SPSS to compare men and women (SEX) in terms of gun ownership (OWNGUN).  Make sure you properly select your independent and dependent variables.  If you put your independent variable in the column and your dependent variable in the row you will want to get the column percents.  Also tell SPSS to compute Chi Square and an appropriate measure of association.  Write a paragraph interpreting this table.  Be sure to use the percents, Chi Square, and the measure of association in your interpretation.

Now let’s check to make sure that those who own guns are more opposed to gun permits by running CROSSTABS to compare those who own guns and those who don’t own guns (OWNGUN) in terms of how they feel about gun permits (GUNLAW).  Again, put your independent variable in the column and ask for the column percents, Chi Square, and an appropriate measure of association.  Write a paragraph using all this information to interpret the table.

## Part III – Controlling for Gun Ownership

Recall that in Part 2 we asked why were men more likely to oppose gun permits?  We suggested that one possible interpretation is that men are more likely to own guns and those who own guns are more likely to oppose gun permits.  Now that we have shown sex to be related to gun ownership and gun ownership to be related to how people feel about gun permits, we’re ready to see if our interpretation of this relationship is correct.

If this explanation is correct then we would expect that the gender gap between men and women in opposing gun permits ought to disappear or at least decrease considerably when we hold gun ownership constant.  As we said in part 2, if the gender gap is because men are more likely to own guns then if we compared men and women who both owned guns the gender gap ought to decrease.  And the same result ought to hold when we compared men and women who both didn’t own guns.

We’re going to divide our sample into two groups – those who own guns and those who don’t own guns.  Then we’re going to look at the crosstabulation of SEX and GUNLAW separately for both of these two groups.  That means we’ll have two crosstabs – one for those who own guns and one for those who don’t own guns.  We’ll call these partial tables since each table contains part of the data.

Run CROSSTABS in SPSS to determine how men and women (SEX) feel about gun control (GUNLAW).  Put the independent variable in the column and the dependent variable in the row of your table.  Then put your control variable (OWNGUN) in the bottom box of the CROSSTABS dialog box.  You will also want to tell SPSS to compute the column percents, Chi Square, and an appropriate measure of association.

Your three-variable table will have three tables stacked on top of each other.  The top table will contain all those who said they had a gun in their home.  Look at the percents, Chi Square, and measure of association for this table and write a paragraph describing your findings.  What was the gender gap for those who owned guns?  Was this difference statistically significant?  Use your measure of association to estimate the strength of the relationship.

The middle table (i.e., below the top table) will contain all those who said they didn’t have a gun in their home.   Write a paragraph describing your findings following the instructions above.

The bottom table will include all respondents including both those who owned a gun and those who didn’t.  By the way, it will look a little different than the two-variable table you ran in Part 1 because SPSS omits all cases with missing values on **any** of the three variables in your table. This means that your table from Part 1 and the three-variable table in Part 3 will be based on a slightly different subset of cases.

## Part IV – Interpreting the Three-Variable Table

We suggested that one possible explanation for men being more opposed to gun permits is that men are more likely to own guns and those who own guns are more likely to oppose gun permits.  If this explanation is correct then we would expect the gender gap between men and women in opposing gun permits ought to disappear or at least decrease considerably when we hold gun ownership constant.

But did that happen?  It did decrease for one of the partial tables but in the other partial table it increased.  So what does that mean?  Our explanation turned out to be incorrect but in the process of analyzing the data we discovered something important.  The gender gap decreased for one of the partial tables and increased for the other partial table.  This is referred to as specification.  We have specified the condition under which the gender gap got smaller and the condition under which it got bigger.

Study the three-variable table and write a paragraph explaining precisely what happened and see if you can suggest why this happened.  It might help you to list the percent who favored guns permits for the four subgroups – men who owned guns, women who owned guns, men who didn’t own guns, and women who didn’t own guns.  Try listing these four groups in ascending order of the percent who favored gun permits.  How does this help you understand what is going on in the three-variable table?

## Part 5 – Summary

Write a paragraph summarizing what you learned about elaborating a two-variable table by adding a third variable into the analysis. How is data analysis an ongoing process in which we discover new things about relationships and then explore these new findings?

# **Appendix**

# **Codebook for the Subset of the 2018 General Social Survey**

The General Social Survey (GSS) is a large, national probability sample of adults in the United States. It began in 1972 and continued on an almost yearly basis until 1996. In 1996, the GSS became a biannual survey and the sample size increased. Many questions are asked on each survey, while other questions are rotated from survey to survey. This subset from the 2018 GSS includes all the cases (2,348) and 76 variables. This data set has already been weighted using the weight variable supplied by the GSS (WTSS). Some of the original GSS variables were recoded and a few new variables created. Some of the new variables have names similar to those in the original GSS data set. The data set was created to accompany these exercises and is named GSS18GENDER.SAV.

**Variable** **Description of Variable**

ABANY Abortion if woman wants for any reason

ABDEFECT Abortion if strong chance of serious defect

ABHLTH Abortion if woman's health seriously endangered

ABNOMORE Abortion if married and wants no more children

ABPOOR Abortion if low income and can't afford more children

ABRAPE Abortion if pregnant as result of rape

ABSINGLE Abortion if not married

AGE Age of respondent

AGED Should aged live with their children?

ATTEND How often respondent attends religious services

CAPPUN Favor or oppose death penalty for murder

CHILDS Number of children

CLASS Subjective class identification

COLATH Allow anti‑religionist to teach

COLCOM Allow communist to teach

COLHOMO Allow homosexual to teach

COLMIL Allow militarist to teach

COLMSLM Allow anti-American Muslim Clergyman to teach in college

COLRAC Allow racist to teach

DEGREE Respondent's highest degree

EDUC Highest year of school completed

ETHNICITY Respondent’s race/ethnicity[[9]](#footnote-9)

FAIR People fair or try to take advantage

FEAR Afraid to walk at night in neighborhood

FINRELA Opinion of family income

GRASS Should marijuana be made legal?

GUNLAW Favor or oppose gun permits

HAPMAR Happiness of marriage

HAPPY General happiness

HEALTH Condition of health

ID Respondent’s identification (id) number

INCOME16 Total family income (2017)

LIBATH Allow anti‑religious book in library

LIBCOM Allow communist's book in library

LIBHOMO Allow homosexual's book in library

LIBMIL Allow militarist's book in library

LIBMSLM Allow anti-American Muslim clergyman's book in library

LIBRAC Allow racist's book in library

MADEG Mother's highest degree

MAEDUC Highest year school completed, mother

MARITAL Marital status

OWNGUN Have gun in home?

PADEG Father's highest degree

PAEDUC Highest year school completed, father

PARTYID Political Party Affiliation

POLVIEWS Think of self as liberal or conservative

PORNLAW Feelings about pornography laws

PRAY How often does respondent pray?

PRAYER Support Supreme Court Decision on prayer in public schools

PRES12 Vote for Romney or Obama in 2012

PRES16 Vote for Clinton or Trump in 2016

RACE Race of respondent

REGION Region of interview

RELIG Respondent's religious preference

RELITEN Strength of religious affiliation

RELPERSN Respondent considers self a religious person

SATFIN Satisfaction with financial situation

SEX Respondent's sex

SIZE Size of place respondent lives in thousands

SPDEG Spouse's highest degree

SPEDUC Highest year school completed, spouse

SPKATH Allow anti‑religionist to speak

SPKCOM Allow communist to speak

SPKHOMO Allow homosexual to speak

SPKMIL Allow militarist to speak

SPKMSLM Allow anti-American Muslim clergyman to speak

SPKRAC Allow racist to speak

TRUST Can people be trusted?

VOTE12 Did respondent vote in 2012 election?

VOTE16 Did respondent vote in 2016 election?

WTSS Weight variable for GSS18 (data subset already weighted by the variable WTSS)

YEAR Year of survey (2018 for all respondents)

ZODIAC Respondent's astrological sign

ZZINCOME161 Recoded family income  
ZZPARTYID1 Recoded political party affiliation  
ZZPOLVIEWS1 Recoded think of self as liberal or conservative – version 1  
ZZPOLVIEWS2 Recoded think of self as liberal or conservative – version 2

1. The variable names are in all capitals. [↑](#footnote-ref-1)
2. The values are in parentheses. [↑](#footnote-ref-2)
3. Missing values indicate answers that do not provide valid information. [↑](#footnote-ref-3)
4. The Center for the American Woman and Politics at Rutgers – The State University of New Jersey has [data on the gender gap in presidential voting from 1980 through 2016](https://cawp.rutgers.edu/sites/default/files/resources/ggpresvote.pdf).  [↑](#footnote-ref-4)
5. Variable names are in all capitals. [↑](#footnote-ref-5)
6. 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. [↑](#footnote-ref-6)
7. Variable names are in all capitals. [↑](#footnote-ref-7)
8. Variable names are in all capitals. [↑](#footnote-ref-8)
9. This variable was created by combining responses to a question asking the respondent’s race (coded as White, Black, and Other), and another question asking whether the respondent is Hispanic. Any respondent identifying as Hispanic was so classified, regardless of race. [↑](#footnote-ref-9)