# Data Analysis Exercises on Gun Issues

# Edward Nelson

# California State University, Fresno

# Table of Contents

**Page**

**Preface 3**

**Chapter 1 – Research Design 4**

**Chapter 2 – Formulating Research Questions and Hypotheses 10**

**Chapter 3 – Causality 14**

**Chapter 4 – Describing the Data 19**

**Chapter 5 – Two Variable Tables – Exploring the Relationship Between Geography  
 and Views on Gun Ownership 30**

**Chapter 6 – Two Variable Tables – Exploring the Relationship Between Perceived Gun   
 Safety and Views on Gun Ownership 39**

**Chapter 7 – Two Variable Tables – Exploring the Relationship Between Politics  
 and Views on Gun Ownership 47**

**Chapter 8 – Two Variable Tables – Exploring the Relationship Between Socioeconomic  
 Status and Views on Gun Ownership 58**

**Chapter 9 – Two Variable Tables – Exploring the Relationship Between Age and   
 Cohorts and Views on Gun Ownership 68**

**Chapter 10 – Two Variable Tables – Exploring the Relationship Between Gender  
 and Views on Gun Ownership 77**

**Chapter 11 – Three Variable Tables – Exploring the Relationships of Politics,   
 Gender, and Views on Gun Ownership 85**

**Chapter 12 – Three Variable Tables – Exploring the Relationships of Own Guns,   
 Gender, and Views on Gun Ownership 94**

**Chapter 13 – Three Variable Tables – Exploring the Relationships of Gender,   
 How Often Follow What’s Going on in Government, and Views   
 on Gun Ownership 102**

**Chapter 14 – Writing Research Reports 108**

**Appendix A – Notes to the Instructor 111**

**Appendix B – Working with PSPP 118**

# Preface

The goal of these exercises is to introduce students to quantitative data analysis using a Pew Research Center survey that includes a series of questions on gun issues and many other topics. There are a series of 14 exercises that include an introduction to the scientific approach, two and three variable data analysis, and writing research reports. The data set we’ll be using is a subset of a 2016 Pew political survey conducted by the Pew Research Center which is one of the preeminent national research centers in the U.S. The data are already weighted to make the sample better represent the population from which the same was drawn.

Since these exercises were written so each exercise was independent of the other exercises, there is quite a bit of duplication from exercise to exercise, particularly in exercises five through ten.

Statistical analysis is limited to descriptive statistics, crosstabulation, Chi Square, and measures of association (primarily Cramer’s V and Somers’ D). Each of these statistics are explained in the exercises but there is no discussion of how to compute them.

There are many statistical packages available for you to use. This series of exercises uses PSPP which is a free alternative to SPSS. I plan to provide an alternative series that uses SPSS which should be available on this website sometime in 2020.

There are two appendices – Appendix A provides notes to the instructor and Appendix B provides information on working with PSPP.

## You have permission to use these exercises and the data set and to revise them to fit your needs.  Feel free to revise them in any way you want.  Just recognize the source of the original exercise.  I would like to hear from you about your experiences using the exercises. If you would like to contact me (Edward Nelson), please email me at [**ednelson@csufresno.edu**](mailto:ednelson@csufresno.edu).  I’m Professor Emeritus at California State University, Fresno in the Sociology department.  I taught research methods, statistics, and critical thinking before retiring and now teach a critical thinking course part time.  Please feel free to contact me about any questions or problems you may have when using the exercises.

##### Data Analysis Exercises on Gun IssuesEdward Nelson, California State University, Fresno

# Exercise 1 – Research Design

## Part 1—What is a Research Design?

A research design is your plan of action.  It lays out how you plan to go about answering your questions.  The research design includes how you plan to select the cases for analysis (sampling), how you will measure concepts, how you plan to collect your data, and how you will analyze the data.  In this exercise we’re going to discuss the data set that we’re going to use in these exercises and explore the research design for this study.

## Part 2 – The Research Study We’ll be Using

The research study we’ll be using in these exercises is an August 2016 political survey conducted by the Pew Research Center.[[1]](#footnote-1) The range of topics covered in this survey was quite broad covering topics such as immigration, religion, problems facing our country today, government regulation of businesses, and gun issues. In the rest of this exercise, we’ll describe how the survey was conducted.

The Pew Research Center is one of the leading survey research centers in the United States. It regularly conducts national surveys dealing with various topics.[[2]](#footnote-2) It conducted two major surveys focusing on religion in 2007 and 2014, national surveys of approximately 35,000 adults. It is well known for its political surveys which include the August 2016 survey we’re using in this exercise. This survey was conducted after the Democratic and Republican nominating conventions and about three months before the November election. While part of this survey focused on the upcoming election, many other questions focused on other issues. These questions form the basis for this set of exercises.

## Part 3 – Sampling

Populations are the complete set of individuals that we want to study.  For example, the population for the Pew survey we’re using in these exercises is all adults 18 years or age and older living in the United States at the time the survey was conducted.  Another example is the Census. The U.S. does a complete enumeration of all individuals living in the United States every ten years (i.e., each year ending in a zero).  We call this a census because it attempts to reach every person in the population.  Still another example of a population is all the students in a particular school or all college students in your state or the nation.  Populations are often large and it’s too costly and time consuming to carry out a complete enumeration.  So, what we do is select a sample from the population where a sample is a subset of the population and then use the sample data to make an inference about the population.[[3]](#footnote-3)

There are many different ways to select samples.  Probability samples are samples in which every individual in the population has a known, non-zero, chance of being in the sample (i.e., the probability of selection).  This isn’t the case for non-probability samples.  An example of a non-probability sample is an instant poll which you hear about on radio and television shows.  A show might invite you to go to a website and answer a question such as whether you favor or oppose gun controls.  This is a purely volunteer sample and we have no idea of the probability of selection.

No probability sample is ever a perfect representation of the population from which the sample is drawn.  This is because every sample contains some amount of sampling error.  Sampling error is inevitable.  There is always some amount of sampling error present in every sample.  The question then is how can we reduce sampling error?[[4]](#footnote-4)

* One way is to increase the sample size.  The larger the sample size, the less the sampling error.  A simple random sample[[5]](#footnote-5) of 400 will have half the sampling error that a simple random sample of 100 has.  To reduce the amount of sampling error by half, you have to quadruple the sample size.
* Another way you can reduce sampling error is to stratify the sample by some variable that is related to whatever you are studying.  For example, if you are trying to explain why some people favor stricter gun control and others oppose it, then you could stratify your sample by sex.  Assuming that sex is related to how people feel about gun control (and it is), this will reduce sampling error.

The August 2016 survey that we are using in these exercises was a telephone survey of adults in the United States conducted during the time period from August 9 through 16 (2016). It was a large survey of 2,010 adults. Another type of error that occurs in most surveys is coverage error which occurs when certain parts of our population are not included in our sample. For example, if we only called respondents on their landline, we would be excluding a big part of our population – those who only had a cell phone and those who did not have a landline. This is often referred to as coverage error because the sample would not adequately cover the entire population. For that reason, in the Pew survey about 25% of all calls were completed via a landline and 75% on a cell phone. Notice that this does not eliminate all coverage error since about 2% of the population does not have a telephone. Even though there is still some coverage error it would be quite small.

Another problem would occur if we only made one attempt at reaching someone. In order to avoid that problem, at least 7 attempts were made to reach respondents. Calls were made on different days of the week and different times of the day to increase the likelihood of reaching respondents.

Still another problem would occur if we only conducted surveys in English. This would exclude all non-English speaking respondents. For that reason, surveys were conducted in both English and Spanish. However, that still means that we are excluding respondents who don’t speak English or Spanish.

Another type of error is nonresponse error. We’re hardly ever able to reach all potential respondents in our sample because we’re unable to contact some and others refuse to be interviewed. Response rate is the percent of all eligible units in our sample that we’re able to reach and complete an interview. Response rates have been dropping significantly in telephone surveys over the last 20 to 30 years. In this Pew survey the response rate was 9 to 10 percent.

One thing that is always confusing is how we can make an inference from a sample that is only a small percent of the population. The U.S. Census reports that there were 308.7 million adults 18 years of age or older in 2010. Our sample would be only a tiny fraction of that population, well under 1%. It seems nonsensical that we would be able to make a valid inference from a sample that represented such a tiny proportion of the entire population. But we can. It turns out that what is important is the absolute size of the sample and not the proportion that it represents of the population and a sample of around 2,000 is a large sample. There are statistical procedures that we can use to estimate how much sampling error there will be in our sample. The Pew Research Center reports that we would expect our sample estimate to be within +/- 2.5 percentage points of the true value in the population with a sample of 2,010. Hard be believe, but true. You’ll have to take Pew’s word for this since the calculations require more statistical background than most of us possess.

Now it’s your turn. Write a paragraph addressing each of the questions below.

1. Is it possible to eliminate all sources of error in any survey?
2. Give another example of a population not discussed in this exercise.
3. The August 2016 Pew survey is an example of a probability sample. What did Pew do to make their study as systematic as possible?
4. What is an example of an unsystematic sample other than instant polls discussed in this exercise?
5. If their sample had only consisted of 500 respondents, would you expect the sampling error to be bigger or smaller than +/- 2.5 percentage points?
6. What are the possible sources of coverage error in the Pew survey described above?
7. What are the potential problems (i.e., nonresponse error) that a response rate of 9 to 10% might create?

## Part 4 – Measurement

The Pew survey asked a variety of questions about guns. One of the questions that we’re going to focus on in these exercises asked whether respondents thought it was more important to protect the right of Americans to own guns, or to control gun ownership. This question is at the center of much of the debate about guns in American society and is one of the variables that we’re going to focus on in these exercises.

Variables are measures of concepts where concepts are abstract ideas. Concepts can be defined as the abstract ideas that we want to use in our study.  Another way to think about concepts is to view them as the tools we’re going to use to try to answer our research questions. We might call this concept “opinions about gun ownership.”

Concepts have to be defined.  There are two different ways to define concepts. First, there is the theoretical definition.  This answers the question – what do we mean by these concepts? Second, there is the operational definition.  How do we measure these concepts?  What are the operations we go through to measure the concepts?

So, in this example the concept is “how people feel about controlling gun ownership” and the variable we’re using to measure this concept is the question from the Pew survey described above. Your research design should identify the concepts that you want to use in your study and both your theoretical and operational definitions of these concepts.

Here’s another example. Let’s say one of our concepts is religiosity which refers to how religious people are. Notice that this is different from another concept that we could call religious identification or preference which refers to the religion that a person identifies with. One of the common ways of measuring religiosity asks how often a person attends religious services. In the Pew survey, the question was “Aside from weddings and funerals, how often do you attend religious services... more than once a week, once a week, once or twice a month, a few times a year, seldom, or never?”

Now it’s your turn again. Write a paragraph addressing each of the questions below.

1. Another area that the Pew survey dealt with was government regulation of businesses, a much-debated topic in politics. The issue here is whether such regulation does more harm than good or whether more good comes from such regulation than harm. Write a question that you think gets at this issue that could be used in a survey.
2. Another topic is government aid to those who don’t have enough money to meet their basic needs. Write a question that you think gets at this issue.
3. Still another hotly contested issue involves immigration. Some feel that border security is more important while others feel that it’s more important to provide a path to citizenship for those who are here illegally. Write a question that you think gets at this issue.

## Part 5 – Data Collection

Science is an empirical enterprise.  That means that it is data based.  There are two ways that we collect data:

* we observe people and use our observations as data, and
* we ask people questions and use what they tell us as data.

We’re going to focus on survey research in these exercises. Sometimes surveys are referred to as sample surveys because we often select a sample of individuals from the population and ask them questions.  Then we use their answers to these questions as our data.  Surveys can take various forms:

* in-person interviews,
* mailed questionnaires,
* telephone surveys,
* web-based surveys, and
* surveys that combine two or more of these types of surveys (i.e., often called mixed-mode surveys).

Keep in mind that error is inevitable whenever we study something.  Since we can’t eliminate all error, our goal is to minimize error.  We have already talked about several different types of error that can occur in a survey – sampling error, coverage error, measurement error, and nonresponse error. If you want to read more about these types of error, take a look at *An Introduction to Survey Research*.[[6]](#footnote-6)

## Part 6 – Data Analysis

Once we have our data, then we want to analyze the data in such a way that we can begin to answer our research questions.  Other exercises in this series will explore data analysis and give you practice in analyzing survey data.  We’ll have much more to say about data analysis in these exercises.

Typically, we start by looking at variables one at a time (i.e., univariate analysis).  We use various statistical tools such as frequency distributions, measures of central tendency, measures of dispersion, measures of skewness, and different types of charts and graphs to help us describe variables.

Then we look at relationships between pairs of variables (i.e., bivariate analysis).  We use various statistical procedures such as crosstabulation, tests of significance, measures of association, and correlation and regression to help explore these relationships.

After looking at two-variable relationships we’ll focus on sets of three or more variables which is often referred to as multivariate analysis. The same statistical tools that we used in looking at bivariate relationships are also used in multivariate analysis.

One very important point to consider is the question of causality.  Survey design can never give us a complete picture of the causal patterns in our data, but it can help us begin to tease out what these causal patterns might look like.

We’ll come back to data analysis in other exercises in this series.

## Part 7 – Accessing the Data Set for the Pew Survey

The data for these exercises is the 2016 Pew Political Survey that was conducted from August 9 through 16 in 2016. The name of the data set is “pew\_2016\_political\_survey\_subset.” You can access it from the Social Science Research and Instructional Council’s website by clicking on [this link](https://www.ssric.org/files/2019-12/pew_2016_political_survey_subset.sav). If you are using this exercise as part of a class, your instructor will show you how to access the data.

This is not the complete data set. We selected the variables that we wanted to use in these exercises. Some of the variables were recoded and a few new variables were created from existing variables. Variable names were changed to make them easier to use.[[7]](#footnote-7)

In these exercises we’re going to use PSPP to analyze the data. PSPP is a free statistical package that can be used as an alternative to SPSS. We’ll show you how to use PSPP in later exercises.

## Part 8 – What’s Next?

In Exercise 2 we’re going to discuss research questions and formulating hypotheses. We’re also going to talk more about the research study that we’re using in this series of exercises.

##### Data Analysis Exercises on Gun IssuesEdward Nelson, California State University, Fresno

# Exercise 2 – Formulating Research Questions and Hypotheses

## Part 1—What are Research Questions?

All research starts with one or more research questions.  These are the questions that you want to answer in your research study.  For example, you might want to find out why some people vote Democrat and others vote Republican.  Or you might want to find out why some people think it is more important to protect the rights of Americans to own guns while others feel it is more important to control gun ownership.

There are lots of ways that we might go about trying to answer these questions.  Some might rely on what their friends or family tell them.  Others might rely on what people in authority like their religious leaders tell them.  Still others might use what is often called common sense to answer these questions.  But we’re going to use the scientific approach to try to answer these questions.  Thomas Sullivan defined science as a “method of obtaining knowledge about the world through systematic observations.”[[8]](#footnote-8)  That’s what we mean when we say that science is empirical; it’s based on observations.  Also, notice that we’re talking about a particular type of observations – systematic observations.

First, we have to learn how to formulate good research questions.  Let’s start by looking at some examples of poor questions and consider why they are poor questions?

* Women are more likely than men to vote Democrat in presidential elections.  This one is easy.  It’s not a question.  It’s actually a hypothesis which we will discuss later.
* Why are women more likely than men to vote Democrat in presidential elections?  This one is a little more difficult.  We want to start with the more general question such as why some people vote Democrat and others vote Republican?  Then we would consider possible answers to this question.  One of these answers might be that gender influences voting.  Since science is empirical, we would start by looking at data to see if, in fact, gender does influence voting and we would discover that in recent presidential elections women are more likely to vote Democrat.  This would lead us to ask why women are more likely than men to vote Democrat.  But we would start our study with the more general question.
* Why do dogs bark?  This is certainly a question and perhaps an interesting one.  But it’s not a question that social scientists would be interested in.  Social scientists focus on questions that involve behavior, attitudes, and opinions.

What are the characteristics of a good research question?

* We start by looking at general questions such as what influences voting or why do some people favor stricter gun control and others oppose it.  As our study progresses, we move to more focused questions such as why women are more likely to want to control gun ownership than men.
* We focus on questions that ask about behavior, attitudes, and opinions.
* Good questions are clearly stated.  Questions such as what about gun control aren’t clear and therefore aren’t useful.
* As with everything we write, we want to make sure that we use correct spelling and good grammar.  So proofread everything you write including your questions.

## Part 2 – Not It’s Your Turn

Write three research questions that you would be interested in trying to answer. Be sure to consider the discussion in Part 1 while formulating your research questions.

## Part 3 – The Research Study We’ll be Using

The research study we’ll be using in these exercises is an August 2016 political survey conducted by the Pew Research Center.[[9]](#footnote-9) The range of topics covered in this survey was quite broad covering topics such as immigration, religion, problems facing our country today, government regulation of businesses, and gun issues.

Let’s start by looking at the variables used in this Pew political survey. Some variables indicate how respondents feel about issues that involve guns.

* Whether respondents own guns
* Whether it is more important to protect the right to own guns or to control gun ownership
* Whether gun ownership does more to protect people from becoming crime victims or whether it puts people’s safety as risk
* Whether they favor or oppose several proposals about gun policy:
  + preventing those with mental illness from purchasing guns,
  + making gun sales subject to background checks,
  + banning assault weapons,
  + creating a federal data base to track guns sales,
  + banning high-capacity ammunition clips, and
  + barring gun purchases by people on no-fly or watch lists.

Other variables describe how big a problem respondents think certain things are in our country. These include:

* crime,
* terrorism,
* immigration,
* availability of good-paying jobs,
* race relations,
* gap between rich and poor, and
* environmental conditions.

Some variables indicate respondent’s opinion on various issues including:

* government regulation of businesses,
* helping the needy,
* solving the important problems facing our country, and
* immigration.

Other variables deal with politics including:

* political party identification,
* political ideology (i.e., conservative, moderate, liberal),
* whether they are registered to vote,
* how often they vote,
* how closely they follow political issues and public affairs, and
* how favorable they are toward Republicans, Democrats, and the Supreme Court.

Some variables describe how respondents feel about the way things are going today in our country including:

* satisfaction with the way things are going,
* whether things are better or worse or about the same as in the past, and
* whether the future will be better or worse or about the same as today.

Others deal with religion such as:

* the religion they identify with,
* attendance at religious services, and
* whether they consider themselves a born-again or Evangelical Christian.

Finally, other variables describe respondent’s background such as:

* where they live,
* age,
* education,
* income,
* gender,
* race, and
* number of children.

## Part 4 – Dependent and Independent Variable

Dependent variables are things that you are trying to explain. For example, one of the questions asks, “What do you think is more important – to protect the rights of Americans to own guns, or to control gun ownership?” This will be our dependent variable for many of these exercises. We want to explain why respondents hold different views on gun ownership.

Independent variables are variables that you think might help explain why people hold different views on gun issues. For example, men and women might hold different views on guns so gender could be one of our independent variables. Gun ownership is another possible independent variable.

We aren’t limited to just one independent variable. So, we might want to know how gun ownership and gender are jointly related to how people feel about guns.

## Part 5 – Hypotheses

Hypotheses specify the relationship that we expect to find between variables. Earlier in this exercise we suggested that gender might be related to how people answer the question in the Pew survey which asks, “What do you think is more important – to protect the rights of Americans to own guns, or to control gun ownership?” That means that how people feel about gun ownership will be our dependent variable and gender our independent variable. This is because we think that gender might explain why respondents hold different views on gun ownership. Our hypothesis could be that women are more likely to want to control gun ownership while men are more likely to want to protect the rights of Americans to own guns.

## Part 6 – Now It’s Your Turn Again

Let’s continue to use the Pew question on gun control as our dependent variable. In other words, we want to explain why some people feel it is more important to control gun ownership and others feel it is more important to protect the right to own guns. Look back at the list of variables in the Pew survey that is in part 2 of this exercise. Select three variables from this list that you think might explain why people hold different views on gun control. In other words, these would be your independent variables. Write three hypotheses linking each of your independent variables to how people feel about gun control. For each hypothesis briefly explain why you think your hypothesis might be true.

## What’s Next?

In Exercise 3 we’re going to consider causality and causal models and we’ll start using the data set that we’ll be using in these exercises.

##### Data Analysis Exercises on Gun IssuesEdward Nelson, California State University, Fresno

# Exercise 3 – Causality

## Part 1 – Scientific Approach

All research starts with a question. In this exercise the question will be why some people feel it is more important to protect the rights of Americans to own guns and others think it’s more important to control gun ownership.

The research study we’ll be using in these exercises is an August 2016 political survey conducted by the Pew Research Center.[[10]](#footnote-10) The name of the data set is “pew\_2016\_political\_survey\_subset.” You can access it from the Social Science Research and Instructional Council’s website by clicking on [this link](https://www.ssric.org/files/2019-12/pew_2016_political_survey_subset.sav). If you are using this exercise as part of a class, your instructor will show you how to access the data.

Assuming you have already installed PSPP on your computer, open the program by double clicking on the PSPP icon. Click on UTILTIES in the menu bar at the top of the screen and scroll down until you see the variable *guns1*. If you continue scrolling, you will see that there are nine variables that start with the letters “guns”. These are the questions about gun issues that Pew included in this survey. Scroll through the other variables so you have an idea of the different variables included in the data set. Use your mouse to click on a variable in the list and a box will open showing you the variable label and value labels for this variable. When you are done close this box by clicking on CANCEL.

There are a lot of ways that people try to answer questions. Some rely on what people in authority tell them. Other rely on their religious or cultural traditions. Still others rely on what some call common sense. In these exercises we’re going to use another approach – the scientific approach. Thomas Sullivan in his book *Applied Sociology* defines science as “*a method of obtaining knowledge about the world though systematic observations*.” Science isn’t the only way of obtaining knowledge but it’s the best way to answer questions such as the ones we’re going to explore in these exercises. Sullivan goes on to list five characteristics of science.[[11]](#footnote-11)

* “All scientific knowledge is based on *observations*.” That’s what we mean when we say science is empirical. We often refer to these observations as data and say that science is data based.
* “Scientific knowledge is based on *systematic observations*.” Some observations are unsystematic such as what our friends tell us. This is what we would call anecdotal data and is not scientific knowledge. Experimental design and surveys would be examples of approaches that are systematic.
* Science “focuses on *causation*.” Science strives to go beyond simple description and focus on understanding why things happen.
* “Scientific knowledge is provisional.” We never claim to have found the final answer to any question. We always allow for the possibility that our answer is wrong. In that case, we would reject our answer in favor of another approach to answering the question.
* “The final characteristic of scientific knowledge is that of *objectivity*.” We try to take our values and biases into account when we seek an answer to a question.

Let’s focus on the third of these characteristics. The scientific approach is most interested in causality. That is, it seeks to understand the factors that cause or lead to or influence some other factor. What are the criteria that we have to meet when we say that one variable causes or influences another variable?

* The two variables must be related to each other. In other words, there has to be a statistical relationship between the two variables.
* We must be able to show which variable is the cause and which is the effect. Often this is referred to as causal ordering.
* We must be able to show that the relationship between the two variables is not due to **any** other variable or combination of variables.

Exercises five through ten focus on the first two of these criteria and on the relationship between two variables. Often this is referred to as bivariate analysis. Exercises eleven through thirteen focus on the third criterion and on sets of three or more variables. This is typically referred to as multivariate analysis.

## Part 2 – Variables in the Pew Data Set

Here’s a partial list of variables in the Pew data set. The variable name is in parentheses. The letters “rec” indicate that this a recoded variable that has collapsed categories in the original variable.

* Geographic variables describing where the respondents live (variables start with “geog”)
  + Region of the country (*geog\_region*)
  + Population density (*geog\_density*)
  + Urban/suburban/rural nature of the community (*geog\_usr*)
* Background or demographic variables describing respondents (variables start with “demo”)
  + Age (*demo\_age*, *demo\_agerec1*, *demo\_agerec2*)
  + Generation (*demo\_gen*)
  + Education (*demo\_educ*, *demo\_educrec*)
  + Family income (*demo\_income, demo\_incomerec*)
  + Gender (*demo\_gender*)
  + Race (*demo\_race1*, *demo\_race2*)
  + Number of children in household (*demo\_child*)
  + Hispanic, Latino, Spanish origin (*demo\_hisp*)
* Political variables (variables start with “pol”)
  + Party identification (*pol1*, *pol2*)
  + Ideology – conservative, moderate, liberal (*pol3, pol3rec*)
  + Political ideology combined with party (*pol4*)
  + Registered to vote (*pol5*)
  + How often votes (*pol6*)
  + How often respondent follows what’s going on in government and public affairs? (*pol7, pol7rec*)
* Religious variables (variables start with “relig”)
  + Religious preference (*relig1, relig1rec*)
  + Considers self a born-again or evangelical Christian? (*relig2*)
  + How often attend religious services? (*relig3, relig3rec*)
* Gun variables (variables start with “guns”)
  + Does anyone in family own guns? (*guns1, guns1rec*)
  + Is it more important to protect rights to own guns or to control gun ownership? (*guns2, guns2rec*)
  + Does gun ownership do more to protect or to put safety at risk? (*guns3*)
  + Do respondents favor or oppose?
    - Preventing those with mental illness from purchasing guns (*guns4*)
    - Guns checks for private gun transactions and gun shows (*guns5*)
    - Ban on assault weapons (*guns6*)
    - Creating a federal data base containing data on gun sales (*guns7*)
    - Banning high capacity ammo clips (*guns8*)
    - Banning gun purchases for those on no-fly or watch list (*guns9*)
* How big a problem is? (variables start with “prob”)
  + Crime (*prob1*)
  + Terrorism (*prob2*)
  + Immigration (*prob3*)
  + Availability of good-paying Jobs (*prob4*)
  + Race relations (*prob5*)
  + Gap between rich and poor (*prob6*)
  + Environmental conditions (*prob7*)
* How favorable/unfavorable are respondents toward the following? (variables start with “favor”)
  + Republicans (*favor1*)
  + Democrats (*favor2*)
  + Supreme Court (*favor3*)
* Forced choice questions on:
  + Government regulation of businesses (*govreg*)
  + Help needy (*helpneedy*)
  + Can solve problems (*solveprob*)
  + Clear solution to big issues facing country today (*solution*)
* Does diversity make U.S. a better/worse place to live or does it make no difference? (diversity)
* Immigration (variables start with “immig”)
  + Which is respondents’ priority – border security or path to citizenship? (*immig1*)
  + Do respondents think that?
    - Immigrants take jobs from Americans or fill jobs Americans don’t want (*immig2*)
    - Immigrants are as honest and hard working as Americans or less honest and hardworking (*immig3*)
    - Immigrants are more likely to commit crime or no more likely to commit crime (*immig4*)
* Way things are going today (variables start with “sat”)
  + Satisfaction with present (*sat1*)
  + Better, worse, or same as 50 years ago (*sat2*)
  + Future better, worse, or about same (*sat3*)
* Weight variables
  + Original Pew weight variable (*weight*)
  + Adjusted weight variable (use this one) (*weightadj*)

## Part 3 – Dependent and Independent Variables

Dependent variables are what we are trying to explain. In these exercises we want to explain why some respondents want to protect gun ownership and other want to control it. Therefore, our dependent variable is the variable *guns2*.

Independent variables are the variables you think might help explain your dependent variable. There are many possible variables that might explain why some respondents want to protect gun ownership and other want to control it. For example, in Exercise 5 we’re going to focus on one possible set of independent variables – where people live (geography) and see if they account for some of the variation in our dependent variable.

Look at the list of variables in Part 2 and select three possible independent variables and write a paragraph for each variable explaining why you think that variable might influence our dependent variable – *guns2*.

## Part 4 – Causal Models

Causal models specify the set of independent variables that we think might influence our dependent variable. We often use diagrams to represent our causal model. We’re going to put the variable in a box and use arrows to represent possible causal relationships. Here’s one possible causal model.

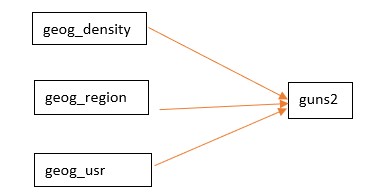


Figure 3-1

This model suggests that each of the geographical variables influences how people feel about gun ownership. We’ll explore this model in Exercise 5.

Let’s look at another model. One of the more important concepts in the social sciences is socioeconomic status which refers to the hierarchical structure of society in which some individuals have higher status and others have lower status. This is what we often refer to as a concept. Concepts can be thought of as abstract ideas. Concepts have to be measurable in order to be useful. We often use income and education as measures of socioeconomic status. We can represent this in the following diagram.

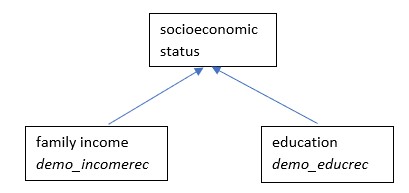


Figure-.3-2

Now let’s assume that we think that socioeconomic status influences how people feel about gun ownership. Again, we’re going to use *guns2* as our measure of how people feel about gun ownership. We could represent this model with the following diagram.

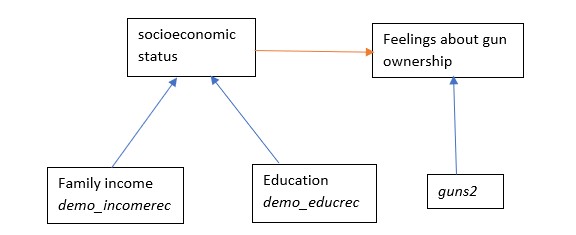


Figure 3-3

Draw a causal model representing a possible causal relationship between one of the independent variables you selected in Part 3 and your dependent variable – *guns2*.

Now draw another causal model using all three of the independent variables from Part 2 and *guns2*.

## What’s Next?

In Exercise 4 we’re going to discuss how we can describe the variables in our model. We’ll talk about frequency distributions, measures of central tendency, measures of dispersion or variability, charts and graphs, and measures of skewness.

##### Data Analysis Exercises on Gun IssuesEdward Nelson, California State University, Fresno

# Exercise 4 – Describing the Data

## Part 1 – Frequency Distributions

Typically, we start by looking at variables one at a time (i.e., univariate analysis).  We use various statistical tools such as frequency distributions, measures of central tendency, measures of dispersion, measures of skewness, and different types of charts and graphs to help us describe variables.

The research study we’ll be using in these exercises is an August 2016 political survey conducted by the Pew Research Center.[[12]](#footnote-12) The name of the data set is “pew\_2016\_political\_survey\_subset.” You can access it from the Social Science Research and Instructional Council’s website by clicking on [this link](https://www.ssric.org/files/2019-12/pew_2016_political_survey_subset.sav). If you are using this exercise as part of a class, your instructor will show you how to access the data.

Assuming you have already installed PSPP on your computer, open the program by double clicking on the PSPP icon. Click on UTILTIES in the menu bar at the top of the screen and scroll down until you see the variable *guns1*. If you continue scrolling, you will see that there are nine variables that start with the letters “guns”. These are the questions about gun issues that Pew included in this survey. Close this box by clicking on CANCEL.

We’re going to start by using two geographical variables – *geog\_density* and *geog\_region* which describe the population density and region of the country in which respondents live. Let’s start by getting frequency distributions for these two variables. A frequency distribution consists of the response categories and the count or number of respondents who gave each response. Various percents are also computed. To get a frequency distribution, click on ANALYZE in the menu bar, point at DESCRIPTIVE STATISTICS, and then click on FREQUENCIES. Scroll down until you see *geog\_density* and *geog\_region*. Your screen should look like Figure 4-1.

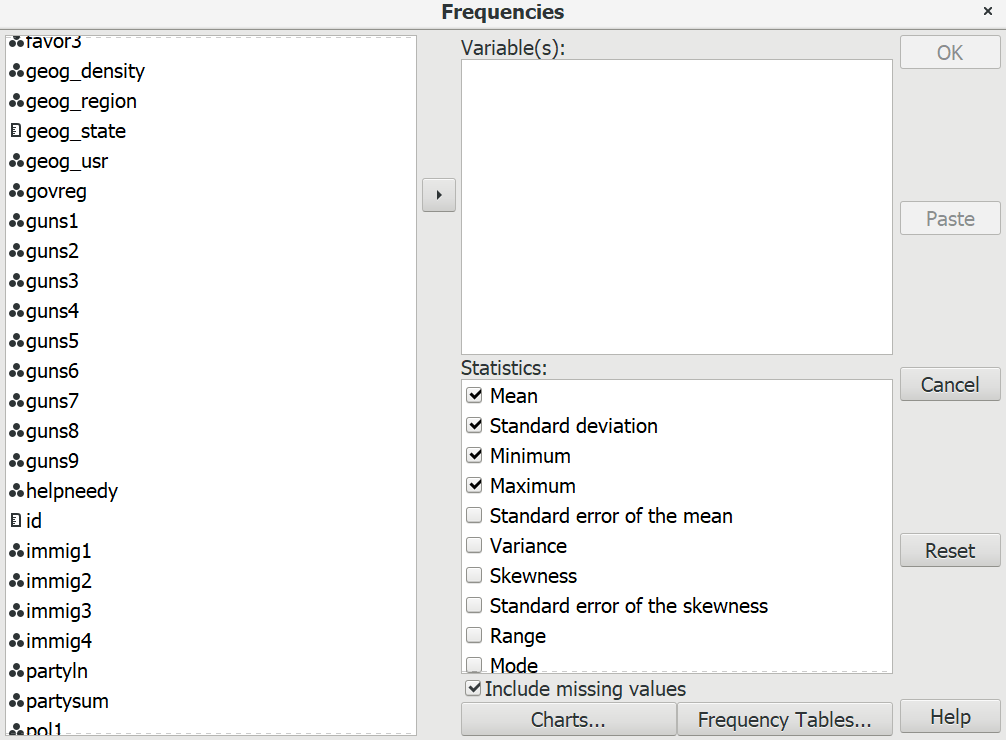


Figure 4-1

Click on *geog\_density* and *geog\_region* and then click on the right-pointing arrow to move them to the VARIABLES box on the right. Notice that in the STATISTICS box in the lower right there are four statistics that are selected. These are the default selections. For the moment, uncheck those boxes. We’ll talk more about statistics later in this chapter. Your screen should look like Figure 4-2.

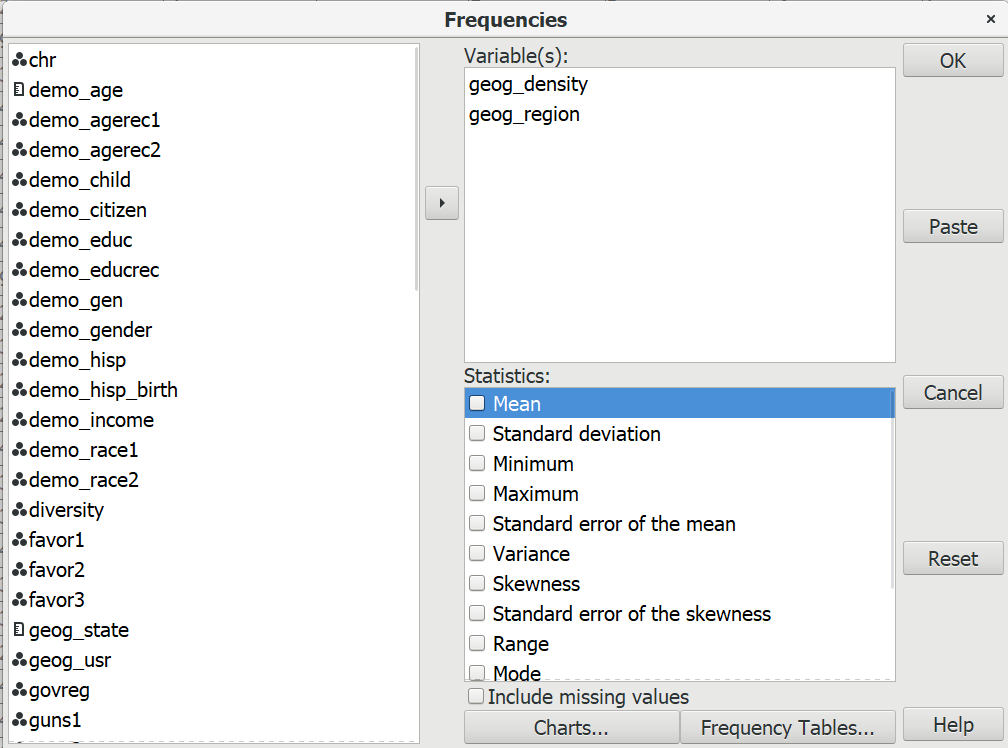


Figure 4-2

Now click on OK and you should see the frequency distributions for these two variables. Figure 4-3 shows what it will look like for *geog\_density* and *geog\_region*.

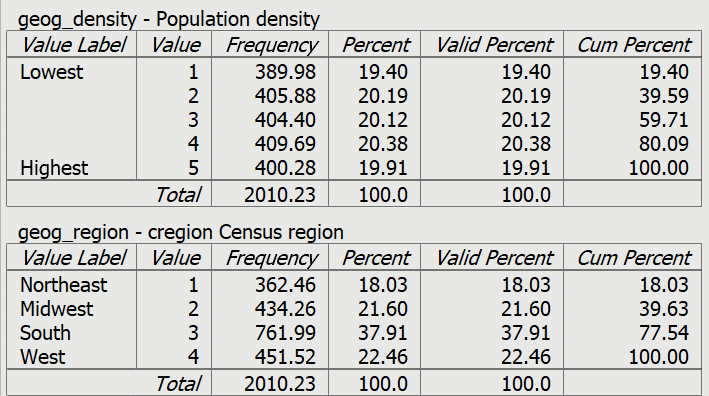


Figure 4-3

Let’s take a look at the top table and see what the numbers in the table tell us.

* The top line of the table shows the variable name and variable label.
* The first column of the table shows the response categories and a brief description of each category (i.e., the value label).
* The second column shows the values that are assigned to each response category.
* The third column shows the number of respondents that gave each response (i.e., frequency or count).
* The last three columns covert the frequencies into percents.
  + The column labeled “percent” divides the frequency by the total number of cases (2010)[[13]](#footnote-13) and multiples the resulting proportion by 100 to get the percent. Notice that percents are computed to the nearest hundredth.
  + The column labeled “valid percent” divides the frequency by the number of respondents that gave valid answers to the question. In this example, the percent column and the valid percent column are identical. This is because there are no cases with missing data in these variables. Where there are cases with missing information (e.g., don’t know or refused) these percents will probably not be the same.
  + The column “cum percent” or cumulative percent shows the number of respondents who gave a particular response and all lower responses. For example, the valid percent for the second lowest density category is 20.19% so the cumulative percent for that row is 20.19 + 19.40 or 39.59%. This tell us that 39.59% of the respondents live in areas that are low density defined as the lowest two density categories. Sometimes, as in this case, this is useful information. Other times it’s not very useful.

Look over the frequency distributions for *geog\_region* and answer the following questions.

* What are the four response categories?
* What do the various percent columns tell you?
* Why are the valid and cumulative percents identical?
* Write a paragraph summarizing what this table tells you about the distribution of respondents over these four regions.

## Part 2 – Levels of Measurement and Central Tendency

There are three commonly used measures of central tendency – the mode, median, and mean of a distribution.  The mode is the most common value or values in a distribution.[[14]](#footnote-14) The median is the middle value of a distribution.[[15]](#footnote-15) The mean is the sum of all the values divided by the number of values.

The choice of which measure of central tendency to use depends, in part, on the level of measurement of the variable. Measures are typically classified as nominal, ordinal, interval, and ratio.

* A **nominal measure** is one in which respondents are sorted into a set of categories which are qualitatively different from each other.  For example, we could classify individuals by their marital status.  Individuals could be married or widowed or divorced or separated or never married.  Our categories should be mutually exclusive and exhaustive.  Mutually exclusive means that every individual can be sorted into one and only one category.  Exhaustive means that every individual can be sorted into a category.  We wouldn’t want to use single as one of our categories because some people who are single can also be divorced and therefore could be sorted into more than one category.  We wouldn’t want to leave widowed off our list of categories because then we wouldn’t have any place to put these individuals. The categories in a nominal level measure have no inherent order to them.  This means that it wouldn’t matter how we ordered the categories.  They could be arranged in any number of different ways.
* An **ordinal measure** is a nominal measure in which the categories are ordered from low to high or from high to low.  We could classify individuals in terms of the highest educational degree they achieved as does the variable *demo\_educ* in our data set.  Some individuals did not attend high school (value 1); others attended high school but did not complete it (value 2); others graduated from high school but didn’t go on to college (value 3).  Others attended college but do not have a college degree (value 4). Still other individuals completed a two-year junior college degree but then left college (value 5).  Others completed their bachelor’s degree (value 6) and others went on to graduate work but did not complete a graduate degree (value 7) while others completed a graduate or professional degree (value 8).  These categories are ordered from low to high. But notice that while the categories are ordered they lack an equal unit of measurement.  That means that the differences between categories are not necessarily equal.  For example, the difference between values 1 and 2 is not the same as the difference between values 4 and 5 in terms of amount of education.
* An **interval measure** is an ordinal measure with equal units of measurement.  For example, consider temperature measured in degrees Fahrenheit.  Now we have equal units of measurement – degrees Fahrenheit.  The difference between 20 degrees and 40 degrees is the same as the difference between 70 degrees and 90 degrees.  Now the numbers have the properties of real numbers and we can add and subtract them.  But notice one thing about the Fahrenheit scale.  There is no absolute zero point. There can be both positive and negative temperatures.  That means that we can’t compare values by taking their ratios.  For example, we can’t divide 80 degrees Fahrenheit by 40 degrees and conclude that 80 is twice as hot at 40.  To do that we would need a measure with an absolute zero.[[16]](#footnote-16)
* A **ratio measure** is an interval measure with an absolute zero point.  The variable *demo\_age* in our data set records age in years and therefore has an absolute zero point.
* Dichotomies are variables that have only two categories. Most variables can be converted into dichotomies. For example, marital status could be expressed as married and not married. Race could be expressed as white and non-white. Dichotomies are always treated as ordinal.

Here’s how to decide which measure of central tendency to use.

* Nominal variables have only one choice. You must use the mode.
* Ordinal variables can use either the mode or the median.
* Interval and ratio variables can use the mode or median or the mean.

Run a frequency distribution for each of the following variables. Decide which measures of central tendency would be appropriate for each variable and tell PSPP to compute them. To get a particular statistic, click on the appropriate measure in the STATISTICS box in the lower right. Uncheck the box for any statistic you don’t want. Keep in mind that there could be more than one appropriate measure for a particular variable. Note that PSPP lists the median as the 50th percentile.[[17]](#footnote-17) Write a paragraph for each variable explaining what these measures tell you.

* *demo\_age*
* *demo\_agerec2*
* *demo\_child*
* *demo\_educ*
* *demo\_income*
* *pol3*

Now run a frequency distribution for favor1 and favor2. Ask for the appropriate measures of central tendency. Use these measures to compare the two variables to determine whether respondents feel more favorably toward the Republican or the Democratic parties.

## Part 3 – Measures of Dispersion or Variability

Dispersion or variability refers to the degree that values in a distribution are spread out or dispersed.  The measures of dispersion that we’re going to discuss are appropriate for interval and ratio level variables.[[18]](#footnote-18)  We’re going to discuss four such measures – the range, the interquartile range, the variance, and the standard deviation.

The range is the difference between the highest and lowest values in the distribution.  Run FREQUENCIES for *demo\_age* and compute the range by looking at the frequency distribution.  You can also ask PSPP to compute it for you.  Click on RANGE in the STATISTICS box.  You should get 78 which is 96 – 18. The range is not a very stable measure since it depends on the two most extreme values – the highest and lowest values.  These are the values most likely to change from sample to sample.

A more stable measure of dispersion is the interquartile range which is the difference between the third quartile (Q3) and the first quartile (Q1).  The third quartile is the same thing as the seventy-fifth percentile which is the value that has 25% of the cases above it and 75% of the cases below it.  The first quartile is the same as the twenty-fifth percentile which is the value that has 75% of the cases above it and 25% of the cases below it.  Look at the cumulative percent column in the frequency distribution for age.  The first quartile will be the category than contains the cumulative percent of 25.0 and the third quartile will be the category that contains the cumulative percent of 75.0.  Once you know Q3 and Q1 you can calculate the interquartile range by subtracting Q1 from Q3.  Since it’s not based on the most extreme values it will be more stable from sample to sample.  From the cumulative percent column, you should see that Q3 will equal 60 and Q1 will equal 32 and the interquartile range will equal 60 – 32 or 28.

The variance is the sum of the squared deviations from the mean divided by the number of cases minus 1 and the standard deviation is just the square root of the variance.  Your instructor may want to go into more detail on how to calculate the variance by hand.  PSPP will calculate them for you.  Select standard deviation and variance from the STATISTICS box.  Notice that the square root of the variance (318.63) equals the standard deviation (17.85).

The variance and the standard deviation can never be negative.  A value of 0 means that there is no variation or dispersion at all in the distribution.  All the values are the same.  The more variation there is, the larger the variance and standard deviation.

So, what does the variance and the standard deviation of the age distribution tell you?  That’s hard to answer because you don’t have anything to compare it to.  But if you knew the standard deviation for both men and women you would be able to determine whether men or women have more variation.  Instead of comparing the standard deviations for men and women you would compute a statistic called the Coefficient of Relative Variation (CRV).  CRV is equal to the standard deviation divided by the mean of the distribution.   A CRV of 2 means that the standard deviation is twice the mean and a CRV of 0.5 means that the standard deviation is one-half of the mean.  You would compare the CRV’s for men and women to see whether men or women have more variation relative to their respective means.

You might also have wondered why you need both the variance and the standard deviation when the standard deviation is just the square root of the variance.  You’ll have to take my word for it that you will need both as you go further in statistics.

Run a frequency distribution for *demo\_age* and *demo\_child*. Have PSPP compute the range, variance, and standard deviation for each variable. Compute the Interquartile Range yourself. Write a paragraph explaining what these measures tell you.

Sometimes we violate assumptions intentionally because the statistic can still provide useful information. The variables *favor1* and *favor2* describe respondent’s favorability toward the Republican and Democratic parties. Respondents were asked to choose one of four categories for each party – very favorable (value 1), mostly favorable (value 2), mostly unfavorable (value 3), and very unfavorable (value 4). This is clearly an ordinal variable. The difference between mostly favorable and mostly unfavorable is probably larger than the difference between very and mostly favorable.

But let’s go ahead and compute means and standard deviations anyway. Ask PSPP to compute them for you. Since high values represent unfavorable ratings, the Republican party is viewed more unfavorably (mean = 2.77) than is the Democratic party (mean = 2.56).

But is there more dispersion or variability in the ratings for one of the parties? Compute the Coefficient of Relative Variation (CRV) for each party following the instructions on the previous page. What did you discover? This is an example of deliberately violating the assumptions underlying the statistics. We gained some valuable information, but we always must keep in mind that violating the assumptions regarding level of measurement might create a problem.

## Part 4 – Charts

PSPP will construct pie charts, bar charts, and histograms.

A pie chart is a chart that shows the frequencies or percents of a variable with a small number of categories.  It is presented as a circle divided into a series of slices.  The area of each slice is proportional to the number of cases or the percent of cases in each category.  It is normally used with nominal or ordinal variables but can be used with interval or ratio variables which have a small number of categories. Figure 4-4 is a pie chart for *geog\_region*.

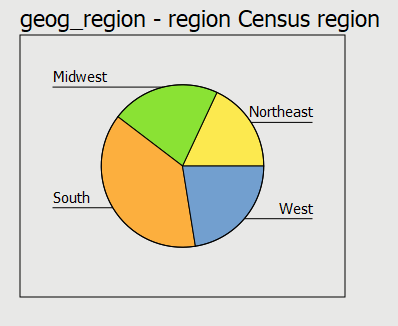


Figure 4-4

A bar chart is a chart that shows the frequencies or percents of a variable and is presented as a series of vertical bars that do not touch each other.  The height of each bar is proportional to the number of cases or the percent of cases in each category.  It is normally used with nominal or ordinal variables. Figure 4-5 is a bar chart of this same variable (*geog\_region*)*.*

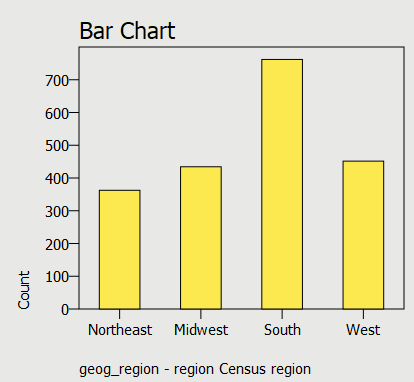


Figure 4-5

A histogram is a graph that shows the frequencies or percents of a variable with a larger number of categories. It is presented as a series of vertical bars that touch each other. The height of each bar is proportional to the number of cases or the percent of cases in each category. It is used with interval or ratio variables. Figure 4-6 is a histogram of *demo\_age*.

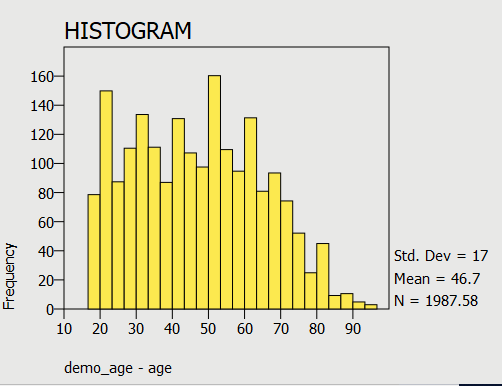


Figure 4-6

To get a chart from PSPP, click on the CHARTS button and check the box for the type of chart you want. If you don’t want to get the frequency distribution, click on FREQUENCIES TABLES and select NEVER under DISPLAY FREQUENCIES TABLES.

Decide which type of chart you want for the following variables and write a paragraph describing what you learned from the chart.

* *demo\_agerec1*
* *demo\_educ*
* *demo\_race1*
* *pol3*

## Part 5 - Skewness

A normal distribution is a unimodal (i.e., single peak) distribution that is perfectly symmetrical.  In a normal distribution the mean, median, and mode are all equal.  Figure 4-7 shows what a normal distribution looks like.

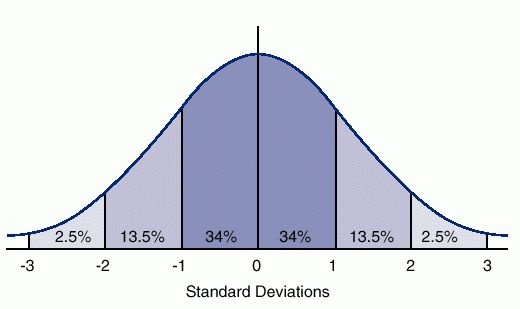


Figure 4-7

The horizontal axis is marked off in terms of standard scores where a standard score tells us how many standard deviations a value is from the mean of the normal distribution.  So, a standard score of +1 is one standard deviation above the mean and a standard score of -1 is one standard deviation below the mean.  The percents tell us the percent of cases that you would expect between the mean and a particular standard score **if** the distribution was perfectly normal.  You would expect to find approximately 34% of the cases between the mean and a standard score of +1 or -1.  In a normal distribution, the mean, median, and mode are all equal and are at the center of the distribution.  So, the mean always has a standard score of zero.

Skewness measures the deviation of a particular distribution from this symmetrical pattern.  In a skewed distribution one side has longer tails than the other side.  If the longer tail is to the left, then it is called a negatively skewed distribution.  If the longer tail is to the right, then it is called a positively skewed distribution.  One way to remember this is to recall that any value to the left of zero is negative and any value to the right of zero is positive.  Figure 4-8 shows you graphs of positively and negatively skewed distributions compared to a normal distribution.

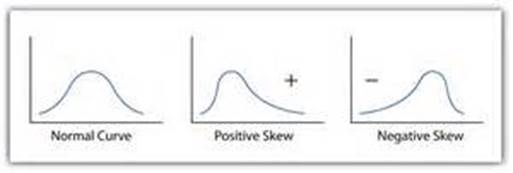


Figure 4-8

The best way to determine the skewness of a distribution is to tell PSPP to give you a histogram along with the mean, median, and the measure of skewness.  Figure 4-9 shows the histogram for *demo\_age* along with the mean, median, and the measure of skewness.

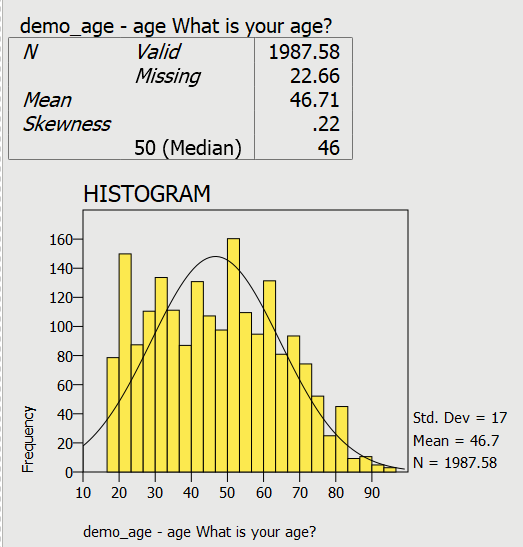


Figure 4-9

PSPP inserted the normal curve on the histogram. To get this, check the box for SUPERIMPOSE NORMAL CURVE when you ask for the histogram. This histogram is not perfectly normal in shape, but it’s not very far off. When the mean is larger than the median, the distribution is positively skewed. In this histogram the mean and the median are very similar telling you the distribution is not very skewed. Now look at the skewness value (0.22).  The larger the skewness value, the more skewed the distribution.  Positive skewness values indicate a positive skew and negative values indicate a negative skew.  There are various rules of thumb suggested for what constitutes a lot of skew but for our purposes we’ll just say that the larger the value, the more the skewness and the sign of the value indicates the direction of the skew. In this histogram, the skewness value is close to 0 meaning that the distribution is not very skewed.

Tell PSPP to give you the histogram for *demo\_child* and also ask for the mean, median, and measure of skewness. Don’t tell PSPP to superimpose the normal curve on this histogram. It won’t be very useful here. Because there are so few categories, the bar chart and the histogram will be identical. Write a paragraph explaining what the chart and these statistics tell you about skewness.

## What’s Next??

In Exercise 5 we’ll move from looking at variables one at a time (univariate analysis) to looking at pairs of variables (bivariate analysis). We’ll explore the relationship between where people live (geography) and how respondents feel about gun control.

##### Data Analysis Exercises on Gun IssuesEdward Nelson, California State University, Fresno

# Exercise 5 – Two-Variable Tables – Exploring the Relationship Between Geography and Views on Gun Ownership

## Part 1 – Research Questions

In Exercise 1 we discussed research design which is your plan of action. It lays out how you plan to go about answering your research questions. In Exercise 2 we focused on formulating these research questions and writing hypotheses.

The research study we’ll be using in these exercises is an August 2016 political survey conducted by the Pew Research Center.[[19]](#footnote-19) The name of the data set is “pew\_2016\_political\_survey\_subset.” You can access it from the Social Science Research and Instructional Council’s website by clicking on [this link](https://www.ssric.org/files/2019-12/pew_2016_political_survey_subset.sav). If you are using this exercise as part of a class, your instructor will show you how to access the data.

The research question we’ll be exploring in this exercise is why some respondents think it’s more important to protect the rights of Americans to own guns while others feel it is more important to control gun ownership. The name of this variable in the Pew data set is *guns2*.[[20]](#footnote-20) Let’s start by running the frequency distribution for *guns2*. To get a frequency distribution, click on ANALYZE in the menu bar, point at DESCRIPTIVE STATISTICS, and then click on FREQUENCIES. Scroll down until you see *guns2* and move it to the VARIABLE(S) box by clicking on the right-pointing arrow to the left of the box. Click on OK and you should see the frequency distribution. If you want more information about frequency distributions, refer to Exercise 4.

This variable is actually a combination of two other variables in the data – one that asks respondents to choose between these two options and another that asks how strongly they feel about it. So *guns2* has four response categories:

* feels strongly about protecting gun ownership,
* wants to protect gun ownership but doesn’t feel strongly about it,
* wants to control gun ownership but doesn’t feel strongly about it, and
* feels strongly about controlling gun ownership.

Notice that most respondents feel strongly about their opinion. Gun issues are something that Americans feel strongly about and is an important political issue in elections.

In Exercise 3 we discussed causal models which suggest some possible causal relationships that we might want to explore. Our dependent variable is what we are trying to explain. Since we want to explain why some respondents want to protect gun ownership and others want to control it, our dependent variable is *guns2*.

Independent variables are the variables you think might help explain your dependent variable. There are many possible variables that might explain why some respondents want to protect gun ownership and others want to control it. We discussed some of them in Exercise 3. In this exercise we’re going to focus on one possible set of independent variables – where people live (geography). There are three measures of geography in our data set that we’re using.

One variable (*geog\_density*) is population density which categorizes respondents by the number of people per square mile in the community in which they live. A second variable (*geog\_region*) is the region of the country in which they live. This variable uses the Census geographical classification of northeast, midwest, south, and west. A third variable (*geog\_usr*) classifies respondents by the type of community in which they live (i.e., urban, suburban, rural). Run frequency distributions for these three variables so you can see what they look like.

## Part 2 – Exploring the Relationship Between Population Density and

## Views on Gun Ownership

### Stating the Hypothesis.

Now we’re ready to develop our hypothesis. Let’s start with population density. High density areas like large urban centers are more likely to vote Democratic while low density areas like rural areas are more likely to vote Republican. We also know that Democrats are more likely to favor controlling gun ownership while Republicans are more likely to favor protecting it. **Therefore, we could hypothesize that** **higher density areas are more likely to want to control gun ownership and lower density areas are more likely to favor protecting gun ownership***.*

### Using Crosstabulation to Test the Hypothesis.

Crosstabulation can be used to explore relationships between categorical variables (i.e., one in which cases are coded into a relatively small number of categories). Let’s start by comparing higher and lower density areas in terms of how they feel about protecting and controlling gun ownership. Population density (*geog\_density*) will be our independent variable and how respondents feel about gun ownership (*guns2*) will be our dependent variable.

Assuming that you have already opened PSPP, click on ANALYZE in the menu bar at the top of your screen, use your mouse to point at DESCRIPTIVE STATISTICS, and then click on CROSSTABS. This will open the crosstabs dialog box. Your screen should look like Figure 5-1. If your screen displays the variable labels instead of the variable names, right click on the list of variable names and uncheck the box that says PREFER VARIABLE LABELS and now you should see the variable names.

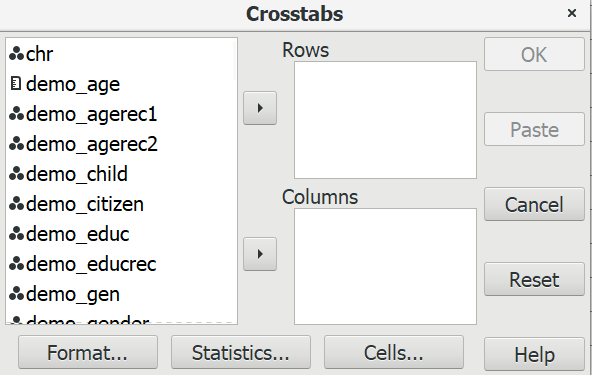


Figure 5-1

Scroll down the list of variables and find *geog\_density* and click on the right-pointing arrow to the left of columns. An easier way to do this is to enter the first letter of the variable name and PSPP will skip to the first variable that starts with a “g”. Now find *guns2* in the list of variables and click on the arrow to the left of rows. If you make a mistake, click on the variable name, make sure the arrow is pointing in the direction you want to move the variable, and click. Your screen should look like Figure 5-2.

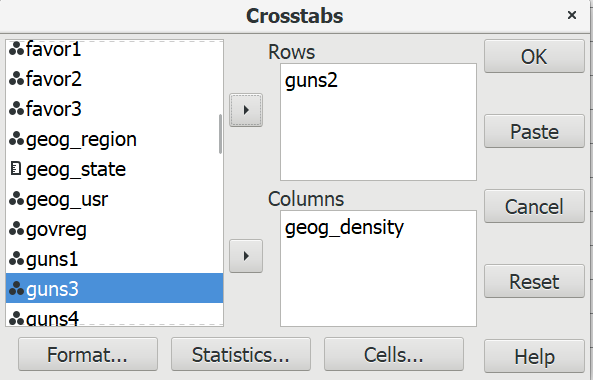


Figure 5-2

Use the following rule to decide which box to insert a variable in: “Put your independent variable in the columns and your dependent variable in the rows.”

Click on the cells button at the bottom of the dialog box. You’ll notice that the boxes for Count, Row, Column, and Total are already checked. These are the default options. Count is the number of cases in each cell of the table. Row, column, and total are three different ways of computing percents. Leave the count box checked. Now the question is which percents we want to use. Here’s a simple rule to follow: “If your independent is in the columns, choose the columns percents. If your independent variable is in the rows, choose the row percents.” Since we decided to put the independent variable in the columns, we’ll want the column percents. Leave the box for column percents checked and uncheck the boxes for the row and total percents. Your screen should look like Figure 5-3.

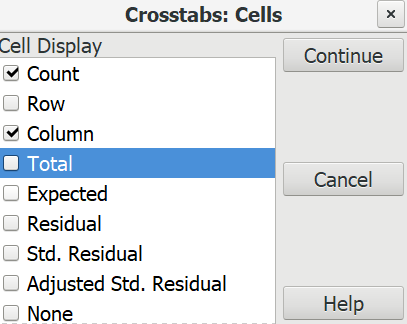


Figure 5-3

Click on the CONTINUE button to go back to Figure 5-2 and then click on OK. This will open the output window and your screen should look like Figure 5-4.

It’s easy to make sure that you have the correct percents. Your independent variable (*geog\_density*) should be in the columns and your dependent variable (*guns2*) should be in the rows. The column percents should sum down to 100%. Look at your table and make sure that you have made the correct selections.

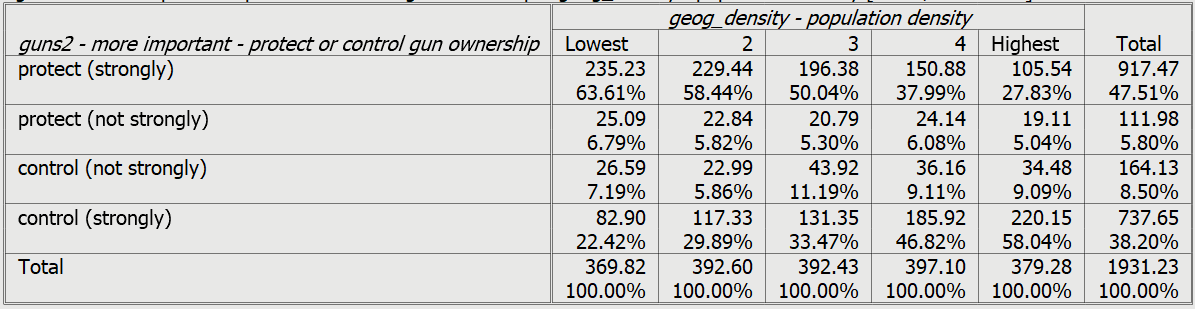


Figure 5-4

How are you going to interpret these percents?  Here’s a simple rule for interpreting percents.

* If your percents sum down to 100%, then compare the percents straight across.
* If your percents sum across to 100%, then compare the percents straight down.

Since the percents sum down to 100%, you want to compare straight across.

Look at the first row.  Approximately 64% of respondents living in the lowest density area strongly feel that it is more important to protect gun ownership. These percentages decrease as density increases. Only 28% of those who live in the highest density area strongly feel that protecting gun ownership is more important. That’s a difference of 36 percentage points.  When you look at the last row you see the opposite pattern. Respondents living in the more densely settled areas are more likely to feel strongly that controlling gun ownership is more important.

We never want to make too much of small differences.  Why not?  No sample is ever a perfect representation of the population from which the sample is drawn.  This is because every sample contains some amount of sampling error.  Sampling error is inevitable.  There is always some amount of sampling error present in every sample.  The larger the sample size, the less the sampling error and the smaller the sample size, the more the sampling error.  But in this case the differences are rather large so we would conclude that respondents living in less densely settled areas are more likely to think that protecting gun ownership is more important and those living in areas with greater density are more likely to want to control gun ownership.

### Chi Square.

In this example, we think that population density and how people feel about gun ownership are related to each other. We’ll call this our research hypothesis.  It’s what we expect to be true.  But there is no way to prove the research hypothesis directly.  So, we’re going to use a method of indirect proof.  We’re going to set up another hypothesis that says that the research hypothesis is not true and call this the null hypothesis.[[21]](#footnote-21)  In our case, the null hypothesis would be that the two variables are unrelated to each other.

Here are our two hypotheses:

* research hypothesis – population density and opinion about gun ownership are related to each other, and
* null hypothesis – population density and opinion about gun ownership are unrelated to each other. In other words, they are independent of each other.

It’s the null hypothesis that we are going to test.

When you ran the crosstab, PSPP computed Chi Square for you.  Click on the STATISTICS button and the box for Chi Square should be checked since it’s the default.

Now you will see another output box below the crosstabulation called “Chi-Square Tests.”  We want the test that is called “Pearson Chi-Square” which is in the first row of the box.  Ignore all the other rows. Your screen should look like Figure 5-5.

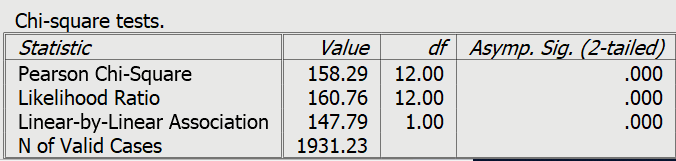


Figure 5-5

The value of Chi Square is 158.29.  Your instructor may or may not want to go into the computation of the Chi Square value but we’re not going to cover the computation in this exercise.

The degrees of freedom (df) is 12.  Degrees of freedom is the number of values that are free to vary.  This table has 4 rows and 5 columns which means there 4 x 5 or 20 cells in the table. Twelve of these cell values are free to vary. Once they are filled in, the values of the remaining 8 cells are fixed.

The two-tailed significance value is 0.000. Actually, this is a rounded value. PSPP rounds to the nearest third decimal place. So, it really means that the significance value is <.0005. This tells us there is only a tiny probability that we would be wrong if we rejected the null hypothesis.  In other words, we would be wrong less than 5 out of 10,000 times.  With odds like that we’re going to reject the null hypothesis.  A common rule is to reject the null hypothesis if the significance value is less than .05 or less than five out of one hundred.  Since, <.0005 is clearly less than .05, we reject the null hypothesis.  This means that we have support for our research hypothesis that the two variables are related.

### Measures of Association.

A measure of association is a numerical value that tells us how strongly related two variables are.[[22]](#footnote-22)  There are several characteristics of a good measure of association.

* They range from a value of 0 (i.e., no relationship) to 1 (i.e., the strongest possible relationship).
* For variables that have an underlying order from low to high they can be positive or negative.  Ordinal variables have an underlying order (i.e., from high to low or from low to high) while nominal variables don’t have an underlying order. A positive value indicates that as one variable increases, the other variable also increases.  A negative value indicates that as one variable increases, the other variable decreases.[[23]](#footnote-23)
* Some measures specify which variable is dependent and which is independent.  The independent variable is some variable that you think might help you explain the variation in the dependent variable.  For example, if your two variables were education and voting you might choose education as the independent variable and voting as your dependent variable because you think that education will help explain why some people vote Democrat and others vote Republican. Measures of association that specify which variable is dependent and which is independent are called asymmetric measures and measures that don’t specify which is dependent and which is independent are called symmetric measures.

There are many measures of association to choose from. When choosing a measure of association, we’ll start by considering the level of measurement of the two variables.

* If one or both of the variables is nominal, then choose one of these measures.[[24]](#footnote-24)
  + Contingency coefficient (CC)
  + Phi
  + Cramer’s V
* If both of the variables are ordinal, then choose from this list.
  + Gamma
  + Somers’ D
  + Kendall’s tau-b
  + Kendall’s tau-c
* Dichotomies should be treated as ordinal. Most variables can be recoded into dichotomies. For example, marital status can be recoded into married or not married. Race can be recoded as white or non-white. All dichotomies should be considered ordinal.

Population density and how respondents feel about gun ownership are both ordinal variables. Both variables have an inherent order to them. Additionally, population density might influence how people feel about gun ownership. This suggests that Somers’ D would be a good choice for a measure of association since it is the only one that allows us to specify one of the variables as dependent. To tell PSPP to compute D, click on STATISTICS and check the box for D. Your screen should look like Figure 5-6.

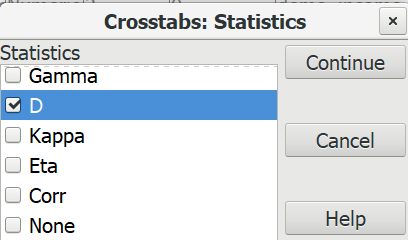


Figure 5-6

Now click on CONTINUE and then on OK. Your screen should look like Figure 5-7.

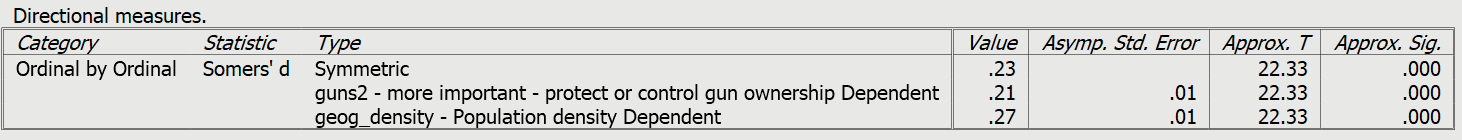


Figure 5-7

Since *guns2* is our dependent variable, the value of D that we want is .21, The significance value on the far right tells us that D is statistically significant. While .21 doesn’t seem very big, it’s at least a moderate relationship. Actually, the best use of measures of association is comparing the strength of relationships in several tables. We’ll come back to such measures in later chapters when we want to see which relationships are stronger and which are weaker.

## Part 3 – Exploring the Relationship Between Region of Country and

## Views on Gun Ownership

Now it’s your turn to interpret the tables. In this section we’re going to focus on the relationship between the region of the country in which respondents live and their views on gun ownership. Our independent variable will be *geog\_region* and the dependent variable will continue to be *guns2*. The south and the midwest are more politically conservative than other regions of the country and conservatives are more likely to want to protect gun ownership. **So, we could hypothesize that respondents in the south and the midwest are more likely to want to protect gun ownership while respondents in other regions are more likely to want to control gun ownership.**

Run the crosstab for *geog\_region* and *guns2*. Be sure to put the independent variable in the columns and the dependent variable in the rows and to get the column percents. Ask for Chi Square and use Cramer’s V as your measure of association. To get Cramer’s V check the box for Phi. We’ll explain why in a little bit. Your tables should look like Figure 5-8.

Crosstab of guns2 by geog_region - 1st part of output

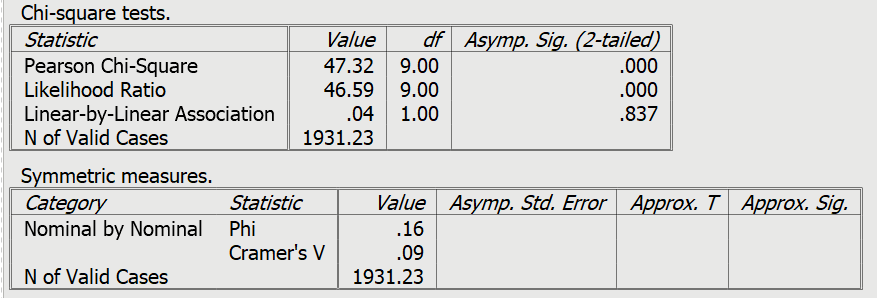



Figure 5-8

The variable *geog\_region* is a nominal variable unlike *geog\_density* which is ordinal. The reason is that there’s no inherent order to the four regions. They could be listed in any order. That means that we need to use one of the measures of association for nominal variables. Phi is a measure of association that is used for tables with two columns and two rows. Cramer’s V is used for tables of any size and Phi is a special case of Cramer’s V. For a table with two columns and two rows, V is equal to Phi. To get Cramer’s V in PSPP you should check the box for Phi.

Answer the following questions for this table.

* What do the column percents tell you about the relationship between these two variables?
* What does Chi Square tell you? Is the relationship statistically significant? How do you know? What does that mean?
* What does Cramer’s V tell you?
* How would you summarize the relationship between *geog\_region* and *guns2*?
* Do the data support your hypothesis?

## Part 4 – Exploring the Relationship Between the Urban, Suburban, Rural Character of Respondents’ Community and Views on Gun Ownership

Now it’s your turn to carry out the analysis and interpretation. Use *geog\_usr* as your independent variable and *gun2* as your dependent variable. Run the tables and answer the following questions.

* What is your hypothesis?
* Using the crosstab and the percentages, what is the relationship between these two variables?
* Describe the pattern that you see in the percents. Use the percents to illustrate the pattern.
* Using Chi Square, is the relationship statistically significant? How do you know? What does this mean?
* Using an appropriate measure of association, how strong is the relationship? Why did you use the measure you selected?
* How would you summarize the relationship between *geog\_usr* and *guns2*?
* Do the data support your hypothesis?

## What’s Next?

In Exercise 6 we’ll continue exploring the relationship between pairs of variables. We’ll look at the relationship between perceptions about gun safety and how people feel about gun ownership.

##### Data Analysis Exercises on Gun IssuesEdward Nelson, California State University, Fresno

# Exercise 6 – Two-Variable Tables – Exploring the Relationship Between Perceived Gun Safety and Views on Gun Ownership

## Part 1 – Research Questions

In Exercise 5 we explored the relationship between geography and how respondents felt about controlling and protecting gun ownership. We found that respondents in higher density areas such as larger metropolitan cities felt that controlling gun ownership was more important than protecting the right to own guns while those in smaller density areas felt just the opposite. We also discovered that respondents living in the south and the midwest were more likely to want to protect the right to own guns than those in other regions of the country.

The research study we’re using in these exercises is an August 2016 political survey conducted by the Pew Research Center.[[25]](#footnote-25) The name of the data set is “pew\_2016\_political\_survey\_subset.” You can access it from the Social Science Research and Instructional Council’s website by clicking on [this link](https://www.ssric.org/files/2019-12/pew_2016_political_survey_subset.sav). If you are using this exercise as part of a class, your instructor will show you how to access the data.

In this exercise we’re going to look at the relationship between perceptions of gun safety and feelings about controlling and protecting gun ownership. The question in the survey about gun safety was, “Do you think that gun ownership in this country does more to protect people from becoming victims of crime or does more to put people’s safety at risk?” The name of the variable in our data set is *guns3*. The question about gun ownership was, “What do you think is more important – to protect the right of Americans to own guns, or to control gun ownership?” and is named *guns2*.

Let’s start by running frequency distributions for *guns2* and *guns3*. To get a frequency distribution, click on ANALYZE in the menu bar, point at DESCRIPTIVE STATISTICS, and then click on FREQUENCIES. Scroll down until you see *guns2* and *guns3* and move them to the VARIABLE(S) box by clicking on the right-pointing arrow to the left of the box. Click on OK and you should see the frequency distributions. For more information on FREQUENCIES, see Exercise 4.

In Exercise 3 we discussed causal models which suggest some possible causal relationships that we want to explore. Our dependent variable is what we are trying to explain. Since we trying to explain why some respondents want to protect gun ownership and others want to control it, our dependent variable is *guns2*.

Independent variables are the variables you think might help explain your dependent variable. There are many possible variables that might explain why some respondents want to protect gun ownership and others want to control it. We discussed some of them in Exercise 3. In this exercise we’re going to consider whether perceptions about gun safety might influence their views on gun ownership.

Before we start, we’re going to look at a different question to demonstrate the basic principles of data analysis. We mentioned earlier that we discovered in Exercise 5 that respondents in higher density areas such as larger metropolitan cities felt that controlling gun ownership was more important than protecting the right to own guns while those in smaller density areas felt just the opposite. How did we come to that conclusion? How did we analyze our data? That’s what we’ll explore in the next section of this exercise. After we have demonstrated the basics of data analysis, we’ll return to the question about gun safety and gun ownership. If you have already worked through Exercise 5, skip to Part 3.

## Part 2 – Exploring the Relationship Between Population Density and

## Views on Gun Ownership

### Stating the Hypothesis.

High density areas like large urban centers are more likely to vote Democratic while low density areas like rural areas are more likely to vote Republican. We also know that Democrats are more likely to favor controlling gun ownership while Republicans are more likely to favor protecting it. **Therefore, we could hypothesize that** **higher density areas are more likely to want to control gun ownership and lower density areas are more likely to favor protecting gun ownership***.*

### Using Crosstabulation to Test the Hypothesis.

Crosstabulation can be used to explore relationships between categorical variables (i.e., one in which cases are coded into a relatively small number of categories). Let’s start by comparing higher and lower density areas in terms of how they feel about protecting and controlling gun ownership. Population density (*geog\_density*) will be our independent variable and how respondents feel about gun ownership (*guns2*) will be our dependent variable.

Assuming that you have already opened PSPP, click on ANALYZE in the menu bar at the top of your screen, use your mouse to point at DESCRIPTIVE STATISTICS, and then click on CROSSTABS. This will open the crosstabs dialog box. Your screen should look like Figure 6-1. If your screen displays the variable labels instead of the variable names, right click on the list of variable names and uncheck the box that says PREFER VARIABLE LABELS and now you should see the variable names.

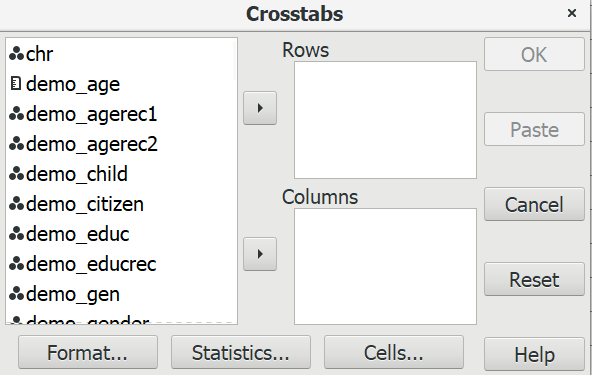


Figure 6-1

Scroll down the list of variables and find *geog\_density* and click on the right-pointing arrow to the left of columns. An easier way to do this is to enter the first letter of the variable name and PSPP will skip to the first variable that starts with a “g”. Now find *guns2* in the list of variables and click on the arrow to the left of rows. If you make a mistake, click on the variable name, make sure the arrow is pointing in the direction you want to move the variable, and click the arrow. Your screen should look like Figure 6-2.

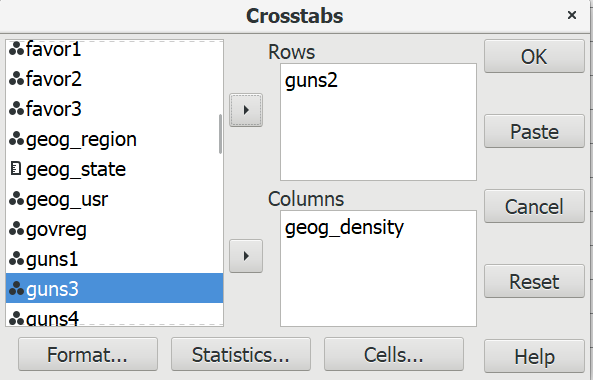


Figure 6-2

Use the following rule to decide which box to insert a variable in: “Put your independent variable in the columns and your dependent variable in the rows.”

Click on the cells button at the bottom of the dialog box. You’ll notice that the boxes for Count, Row, Column, and Total are already checked. These are the default options. Count is the number of cases in each cell of the table. Row, column, and total are three different ways of computing percents. Leave the count box checked. Now the question is which percents we want to use. Here’s a simple rule to follow: “If your independent is in the columns, choose the columns percents. If your independent variable is in the rows, choose the row percents.” Since we decided to put the independent variable in the columns, we’ll want the column percents. Leave the box for column percents checked and uncheck the boxes for the row and total percents. Your screen should look like Figure 6-3.

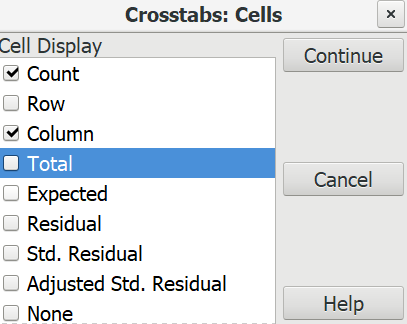


Figure 6-3

Click on the CONTINUE button to go back to Figure 6-2 and then click on OK. This will open the output window and your screen should look like Figure 6-4.

It’s easy to make sure that you have the correct percents. Your independent variable (*geog\_density*) should be in the columns and your dependent variable (*guns2*) should be in the rows. The column percents should sum down to 100%. Look at your table and make sure that you have made the correct selections.

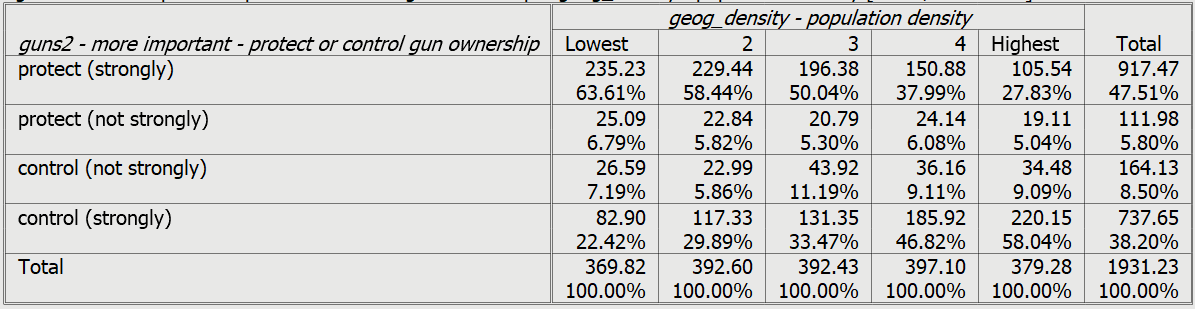


Figure 6-4

How are you going to interpret these percents?  Here’s a simple rule for interpreting percents.

* If your percents sum down to 100%, then compare the percents straight across.
* If your percents sum across to 100%, then compare the percents straight down.

Since the percents sum down to 100%, you want to compare straight across.

Look at the first row.  Approximately 64% of respondents living in the lowest density area strongly feel that it is more important to protect gun ownership. These percentages decrease as density increases. Only 28% of those who live in the highest density area strongly feel that protecting gun ownership is more important. That’s a difference of 36 percentage points.  When you look at the last row you see the opposite pattern. Respondents living in the more densely settled areas are more likely to feel strongly that controlling gun ownership is more important.

We never want to make too much of small differences.  Why not?  No sample is ever a perfect representation of the population from which the sample is drawn.  This is because every sample contains some amount of sampling error.  Sampling error is inevitable.  There is always some amount of sampling error present in every sample.  The larger the sample size, the less the sampling error and the smaller the sample size, the more the sampling error.  But in this case the differences are rather large so we would conclude that respondents living in less densely settle areas are more likely to think that protecting gun ownership is more important and those living in areas with greater density are more likely to want to control gun ownership.

### Chi Square.

In this example, we think that population density and how people feel about gun ownership are related to each other. We’ll call this our research hypothesis.  It’s what we expect to be true.  But there is no way to prove the research hypothesis directly.  So, we’re going to use a method of indirect proof.  We’re going to set up another hypothesis that says that the research hypothesis is not true and call this the null hypothesis.[[26]](#footnote-26)  In our case, the null hypothesis would be that the two variables are unrelated to each other.

Here are our two hypotheses:

* research hypothesis – population density and opinion about gun ownership are related to each other, and
* null hypothesis – population density and opinion about gun ownership are unrelated to each other. In other words, they are independent of each other.

It’s the null hypothesis that we are going to test.

When you ran the crosstab, PSPP computed Chi Square for you.  Click on the STATISTICS button and the box for Chi Square should be checked since it’s the default.

Now you will see another output box below the crosstabulation called “Chi-Square Tests.”  We want the test that is called “Pearson Chi-Square” in the first row of the box.  Ignore all the other rows in this box. Your screen should look like Figure 6-5.

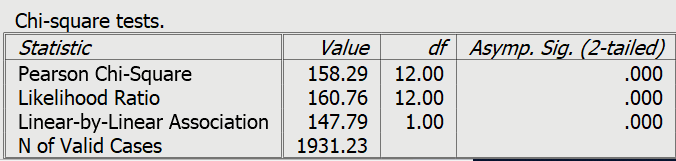


Figure 6-5

The value of Chi Square is 158.29.  Your instructor may or may not want to go into the computation of the Chi Square value but we’re not going to cover the computation in this exercise.

The degrees of freedom (df) is 12.  Degrees of freedom is the number of values that are free to vary.  This table has 4 rows and 5 columns which means there 4 x 5 or 20 cells in the table. Twelve of these cell values are free to vary and once they are filled in, the values of the remaining 8 cells are fixed.

The two-tailed significance value is 0.000. Actually, this is a rounded value. PSPP rounds to the nearest third decimal place. So, it really means that the significance value is <.0005. This tells us that there is only a tiny probability that we would be wrong if we rejected the null hypothesis.  In other words, we would be wrong less than 5 out of 10,000 times.  With odds like that we’re going to reject the null hypothesis.  A common rule is to reject the null hypothesis if the significance value is less than .05 or less than five out of one hundred.  Since, <.0005 is clearly less than .05, we reject the null hypothesis.  This means that we have support for our research hypothesis that the two variables are related.

### Measures of Association. A measure of association is a numerical value that tells us how strongly related two variables are.[[27]](#footnote-27)  There are several characteristics of a good measure of association.

* They range from a value of 0 (i.e., no relationship) to 1 (i.e., the strongest possible relationship).
* For variables that have an underlying order from low to high they can be positive or negative.  Ordinal variables have an underlying order (i.e., from high to low or from low to high) while nominal variables don’t have an underlying order. A positive value indicates that as one variable increases, the other variable also increases.  A negative value indicates that as one variable increases, the other variable decreases.[[28]](#footnote-28)
* Some measures specify which variable is dependent and which is independent.  The independent variable is some variable that you think might help you explain the variation in the dependent variable.  For example, if your two variables were education and voting you might choose education as the independent variable and voting as your dependent variable because you think that education will help explain why some people vote Democrat and others vote Republican. Measures of association that specify which variable is dependent and which is independent are called asymmetric measures and measures that don’t specify which is dependent and which is independent are called symmetric measures.

There are many measures of association to choose from. When choosing a measure of association, we’ll start by considering the level of measurement of the two variables.

* If one or both of the variables is nominal, then choose one of these measures.[[29]](#footnote-29)
  + Contingency coefficient (CC)
  + Phi
  + Cramer’s V
* If both of the variables are ordinal, then choose from this list.
  + Gamma
  + Somers’ D
  + Kendall’s tau-b
  + Kendall’s tau-c
* Dichotomies should be treated as ordinal. Most variables can be recoded into dichotomies. For example, marital status can be recoded into married or not married. Race can be recoded as white or non-white. All dichotomies should be considered ordinal.

Population density and how respondents feel about gun ownership are both ordinal variables. Both variables have an inherent order to them. Additionally, population density might influence how people feel about gun ownership. This suggests that Somers’ D would be a good choice for a measure of association since it is the only one that allows us to specify one of the variables as dependent. To tell PSPP to compute D, click on STATISTICS and check the box for D. Your screen should look like Figure 6-6.

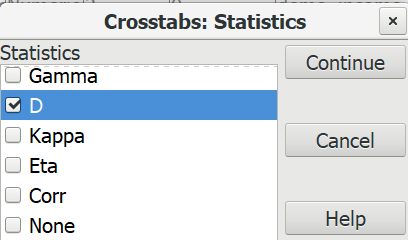


Figure 6-6

Now click on CONTINUE and then on OK. Your screen should look like Figure 6-7.

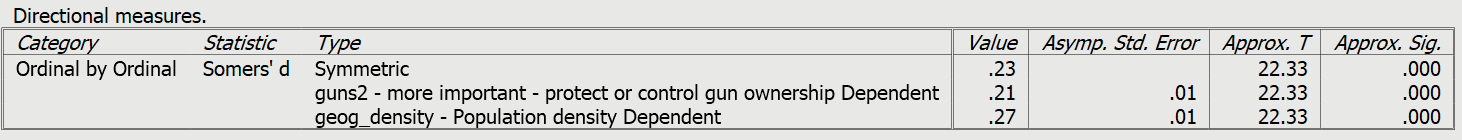


Figure 6-7

Since *guns2* is our dependent variable, the value of D that we want is .21, The significance value on the far right tells us that D is statistically significant. While .21 doesn’t seem very big, it’s at least a moderate relationship. Actually, the best use of measures of association is comparing the strength of relationships in several tables. We’ll come back to such measures in later chapters when we want to see which relationships are stronger and which are weaker.

## Part 3 – Exploring the Relationship Between Gun Safety and Views on Gun Ownership

Now it’s your turn to carry out the analysis and interpretation. Use *guns3* as your independent variable and *guns2* as your dependent variable. Run the tables and answer the following questions.

* What is your hypothesis?
* Using the crosstab and the percentages, what is the relationship between these two variables? In other words, describe the pattern that you see in the percents. Use the percents to illustrate the pattern.
* Using Chi Square, is the relationship statistically significant? How do you know? What does this mean?
* Using an appropriate measure of association, how strong is the relationship? Why did you use the measure you selected?
* How would you summarize the relationship between *guns3* and *guns2*?
* Do the data support your hypothesis?

## What’s Next?

In the next exercise we’ll explore the relationship between politics and how respondents view gun ownership. We’ll use crosstabulation, Chi Square, and a measure of association as we did in this exercise.

##### Data Analysis Exercises on Gun IssuesEdward Nelson, California State University, Fresno

# Exercise 7 – Two-Variable Tables – Exploring the Relationship Between Politics and Views on Gun Ownership

## Part 1 – Research Questions

In Exercise 5 we explored the relationship between geography and how respondents felt about controlling and protecting gun ownership. We found that respondents in higher density areas such as larger metropolitan cities felt that controlling gun ownership was more important than protecting the right to own guns while those in smaller density areas felt just the opposite. We also discovered that respondents living in the south and the midwest were more likely to want to protect the right to own guns than those in other regions of the country.

The research study we’re using in these exercises is an August 2016 political survey conducted by the Pew Research Center.[[30]](#footnote-30) The name of the data set is “pew\_2016\_political\_survey\_subset.” You can access it from the Social Science Research and Instructional Council’s website by clicking on [this link](https://www.ssric.org/files/2019-12/pew_2016_political_survey_subset.sav). If you are using this exercise as part of a class, your instructor will show you how to access the data.

In this exercise we’re going to look at the relationship between politics and feelings about controlling and protecting gun ownership. We’re going to look at two dimensions of politics – party identification and political ideology (i.e., conservative, moderate, and liberal). The Pew survey also asks respondents what they think is more important – “to protect the right of Americans to own guns, or to control gun ownership?” This variable is named *guns2*.

In Exercise 3 we discussed causal models which suggest some possible causal relationships that we want to explore. Our dependent variable is what we are trying to explain. Since we trying to explain why some respondents want to protect gun ownership and others want to control it, our dependent variable is *guns2*.

Independent variables are the variables you think might help explain your dependent variable. There are many possible variables that might explain why some respondents want to protect gun ownership and other want to control it. We discussed some of them in Exercise 3. In the next section of this exercise we’ll explore how the Pew survey measured political party identification and political ideology which will be our independent variables.

## Part 2 – Exploring Politics

Let’s start with party identification. The two major parties are the Democratic and the Republican parties. A large number of individuals choose not to identify with either party or say they are independents. Additionally, a much smaller number of people say they have no preference or identify with third parties. Pew asked the following question – “In politics today, do you consider yourself a Republican, Democrat, or independent.” This variable is named *pol1*. Run a frequency distribution for this variable. For more information on FREQUENCIES, see Exercise 4. PSPP will open an output window that contains your frequency distribution. To move back and forth between the data and the output windows, click on WINDOWS in the menu bar at the top of your screen and check the box for OUTPUT. Your screen should look like Figure 7-1.

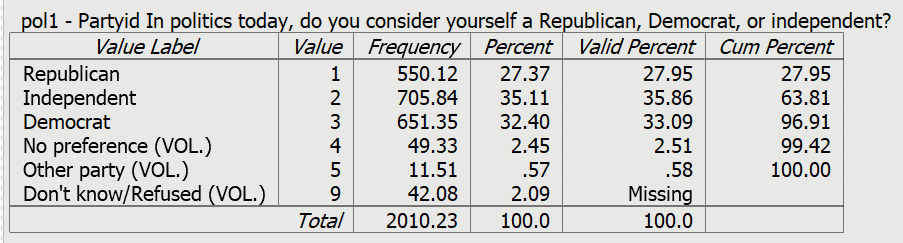


Figure 7-1

Most respondents choose one of the three categories offered them but a few volunteer that they have no preference or that they identify with some other party. Others say they don’t know or refused to answer the question. Don’t knows and refusals are treated as missing data since they didn’t answer the question.

Even though approximately 36% said they were independent, we know that some of these independents actually lean either Republican or Democrat. Pew asked those who didn’t choose Republican or Democrat which way they leaned. We combined *pol1* with this follow-up question and called this new variable *pol2*. Run a frequency distribution for *pol2*. Your screen should look like Figure 7-2.

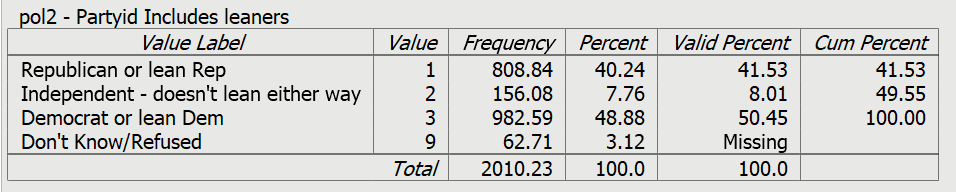


Figure 7-2

To measure political views or ideology, Pew asked respondents if they considered themselves to be very conservative, conservative, moderate, liberal, or very liberal. This variable is named *pol3*. Run a frequency distribution for *pol3* and your screen should look like Figure 7-3.

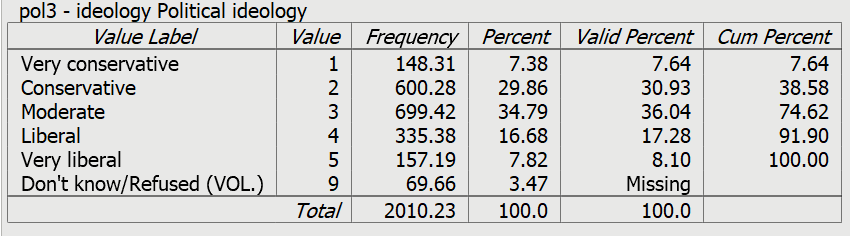


Figure 7-3

Party identification can be combined with political ideology thus producing still another variable which we called *pol4*. Run a frequency distribution for this new variable and your screen should look like Figure 7.4.

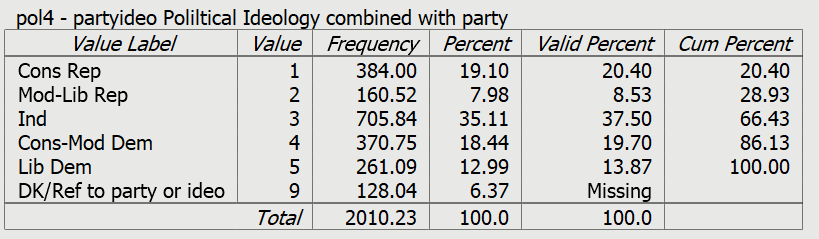


Figure 7-4

Write several paragraphs summarizing what we learned about political behavior in the United States from these four frequency distributions. Discuss each of the distributions separately and then summarize what they collectively tell us. Use the valid percents in your answers.

Now we’re ready to begin exploring the relationship of these political variables with how people feel about gun ownership. We’re going to use a statistical procedure called crosstabulation to do this. We’re also going to use several statistics to help us – Chi Square and Somers’ D which is a measure of association.

Before we start, we’re going to look at a different question to demonstrate the basic principles of data analysis. If you have already worked through Exercise 5 and feel comfortable with crosstabulation, Chi Square, and measures of association you can skip to Part 3.

We mentioned earlier that we discovered in Exercise 5 that respondents in higher density areas such as larger metropolitan cities felt that controlling gun ownership was more important than protecting the right to own guns while those in smaller density areas felt just the opposite. How did we come to that conclusion? How did we analyze our data? That’s what we’ll explore in the next section of this exercise. After we have demonstrated the basics of data analysis, we’ll return to the question about politics and gun ownership.

## Part 3 – Exploring the Relationship Between Population Density and

## Views on Gun Ownership

### Stating the Hypothesis.

High density areas like large urban centers are more likely to vote Democrat while low density areas like rural areas are more likely to vote Republican. We also know that Democrats are more likely to favor controlling gun ownership while Republicans are more likely to favor protecting it. **Therefore, we could hypothesize that** **higher density areas are more likely to want to control gun ownership and lower density areas are more likely to favor protecting gun ownership***.*

### Using Crosstabulation to Test the Hypothesis.

Crosstabulation can be used to explore relationships between categorical variables (i.e., one in which cases are coded into a relatively small number of categories). Let’s start by comparing higher and lower density areas in terms of how they feel about protecting and controlling gun ownership. Population density (*geog\_density*) will be our independent variable and how respondents feel about gun ownership (*guns2*) will be our dependent variable.

Assuming that you have already opened PSPP, click on ANALYZE in the menu bar at the top of your screen, use your mouse to point at DESCRIPTIVE STATISTICS, and then click on CROSSTABS. This will open the crosstabs dialog box. Your screen should look like Figure 7-5. If your screen displays the variable labels instead of the variable names, right click on the list of variable names and uncheck the box that says PREFER VARIABLE LABELS and now you should see the variable names.

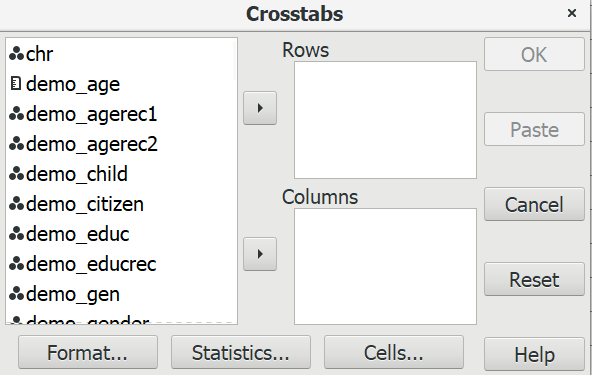


Figure 7-5

Scroll down the list of variables and find *geog\_density* and click on the right-pointing arrow to the left of columns. An easier way to do this is to enter the first letter of the variable name and PSPP will skip to the first variable that starts with a “g”. Now find *guns2* in the list of variables and click on the arrow to the left of rows. If you make a mistake, click on the variable name, make sure the arrow is pointing in the direction you want to move the variable, and click the arrow. Your screen should look like Figure 7-6.

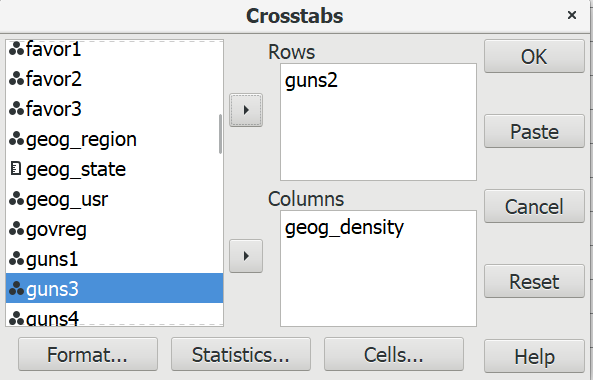


Figure 7-6

Use the following rule to decide which box to insert a variable in: “Put your independent variable in the columns and your dependent variable in the rows.”

Click on the cells button at the bottom of the dialog box. You’ll notice that the boxes for Count, Row, Column, and Total are already checked. These are the default options. Count is the number of cases in each cell of the table. Row, column, and total are three different ways of computing percents. Leave the count box checked. Now the question is which percents we want to use. Here’s a simple rule to follow: “If your independent is in the columns, choose the columns percents. If your independent variable is in the rows, choose the row percents.” Since we decided to put the independent variable in the columns, we’ll want the column percents. Leave the box for column percents checked and uncheck the boxes for the row and total percents. Your screen should look like Figure 7-7.

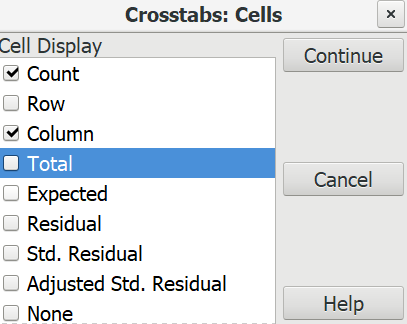


Figure 7-7

Click on the CONTINUE button to go back to Figure 7-6 and then click on OK. This will open the output window and your screen should look like Figure 7-8.

It’s easy to make sure that you have the correct percents. Your independent variable (*geog\_density*) should be in the columns and your dependent variable (*guns2*) should be in the rows. The column percents should sum down to 100%. Look at your table and make sure that you have made the correct selections.

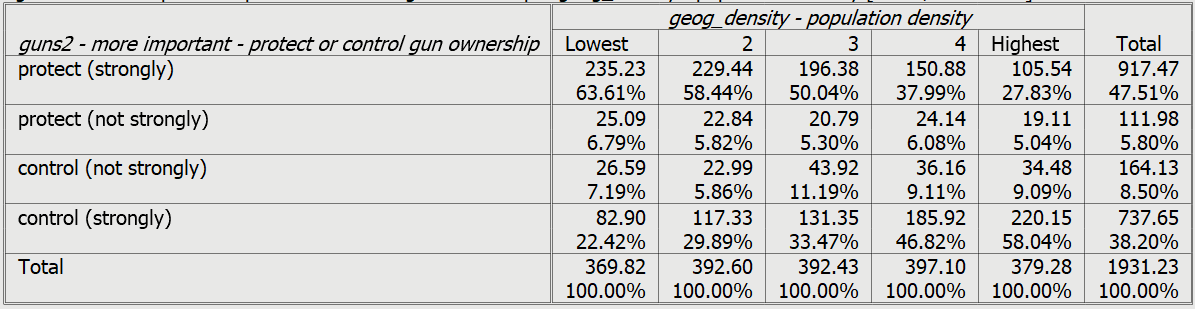


Figure 7-8

How are you going to interpret these percents?  Here’s a simple rule for interpreting percents.

* If your percents sum down to 100%, then compare the percents straight across.
* If your percents sum across to 100%, then compare the percents straight down.

Since the percents sum down to 100%, you want to compare straight across.

Look at the first row.  Approximately 64% of respondents living in the lowest density area feel strongly that it is more important to protect gun ownership. These percentages decrease as density increases. Only 28% of those who live in the highest density area strongly feel that protecting gun ownership is more important. That’s a difference of 36 percentage points.  When you look at the last row you see the opposite pattern. Respondents living in the more densely settled areas are more likely to strongly feel that controlling gun ownership is more important.

We never want to make too much of small differences.  Why not?  No sample is ever a perfect representation of the population from which the sample is drawn.  This is because every sample contains some amount of sampling error.  Sampling error is inevitable.  There is always some amount of sampling error present in every sample.  The larger the sample size, the less the sampling error and the smaller the sample size, the more the sampling error.  But in this case the differences are rather large so we would conclude that respondents living in less densely settled areas are more likely to think that protecting gun ownership is more important and those living in areas with greater density are more likely to want to control gun ownership.

### Chi Square.

In this example, we think that population density and how people feel about gun ownership are related to each other. We’ll call this our research hypothesis.  It’s what we expect to be true.  But there is no way to prove the research hypothesis directly.  So, we’re going to use a method of indirect proof.  We’re going to set up another hypothesis that says that the research hypothesis is not true and call this the null hypothesis.[[31]](#footnote-31)  In our case, the null hypothesis would be that the two variables are unrelated to each other.

Here are our two hypotheses:

* research hypothesis – population density and opinion about gun ownership are related to each other, and
* null hypothesis – population density and opinion about gun ownership are unrelated to each other. In other words, they are independent of each other.

It’s the null hypothesis that we are going to test.

When you ran the crosstab, PSPP computed Chi Square for you.  Click on the STATISTICS button and the box for Chi Square should be checked since it’s the default.

Now you will see another output box below the crosstabulation called “Chi-Square Tests.”  We want the test that is called “Pearson Chi-Square” in the first row of the box.  Ignore all the other rows in this box. Your screen should look like Figure 7-9.

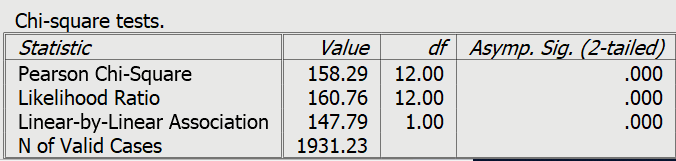


Figure 7-9

The value of Chi Square is 158.29.  Your instructor may or may not want to go into the computation of the Chi Square value but we’re not going to cover the computation in this exercise.

The degrees of freedom (df) is 12.  Degrees of freedom is the number of values that are free to vary.  This table has 4 rows and 5 columns which means there 4 x 5 or 20 cells in the table. Twelve of these cell values are free to vary and once they are filled in, the values of the remaining 8 cells are fixed.

The two-tailed significance value is 0.000. Actually, this is a rounded value. PSPP rounds to the nearest third decimal place. So, it really means that the significance value is <.0005. This tells us that there is only a tiny probability that we would be wrong if we rejected the null hypothesis.  In other words, we would be wrong less than 5 out of 10,000 times.  With odds like that we’re going to reject the null hypothesis.  A common rule is to reject the null hypothesis if the significance value is less than .05 or less than five out of one hundred.  Since, <.0005 is clearly less than .05, we reject the null hypothesis.  This means that we have support for our research hypothesis that the two variables are related.

### Measures of Association. A measure of association is a numerical value that tells us how strongly related two variables are.[[32]](#footnote-32)  There are several characteristics of a good measure of association.

* They range from a value of 0 (i.e., no relationship) to 1 (i.e., the strongest possible relationship).
* For variables that have an underlying order from low to high they can be positive or negative.  Ordinal variables have an underlying order (i.e., from high to low or from low to high) while nominal variables don’t have an underlying order. A positive value indicates that as one variable increases, the other variable also increases.  A negative value indicates that as one variable increases, the other variable decreases.[[33]](#footnote-33)
* Some measures specify which variable is dependent and which is independent.  The independent variable is some variable that you think might help you explain the variation in the dependent variable.  For example, if your two variables were education and voting you might choose education as the independent variable and voting as your dependent variable because you think that education will help explain why some people vote Democrat and others vote Republican. Measures of association that specify which variable is dependent and which is independent are called asymmetric measures and measures that don’t specify which is dependent and which is independent are called symmetric measures.

There are many measures of association to choose from. When choosing a measure of association, we’ll start by considering the level of measurement of the two variables.

* If one or both of the variables is nominal, then choose one of these measures.[[34]](#footnote-34)
  + Contingency coefficient (CC)
  + Phi
  + Cramer’s V
* If both of the variables are ordinal, then choose from this list.
  + Gamma
  + Somers’ D
  + Kendall’s tau-b
  + Kendall’s tau-c
* Dichotomies should be treated as ordinal. Most variables can be recoded into dichotomies. For example, marital status can be recoded into married or not married. Race can be recoded as white or non-white. All dichotomies should be considered ordinal.

Population density and how respondents feel about gun ownership are both ordinal variables. Both variables have an inherent order to them. Additionally, population density might influence how people feel about gun ownership. This suggests that Somers’ D would be a good choice for a measure of association since it is the only one that allows us to specify one of the variables as dependent. To tell PSPP to compute D, click on STATISTICS and check the box for D. Your screen should look like Figure 7-10.

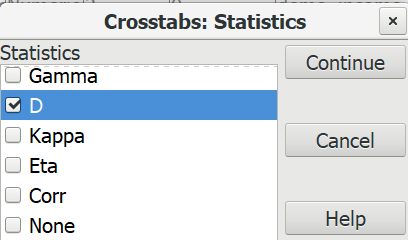


Figure 7-10

Now click on CONTINUE and then on OK. Your screen should look like Figure 7-11.

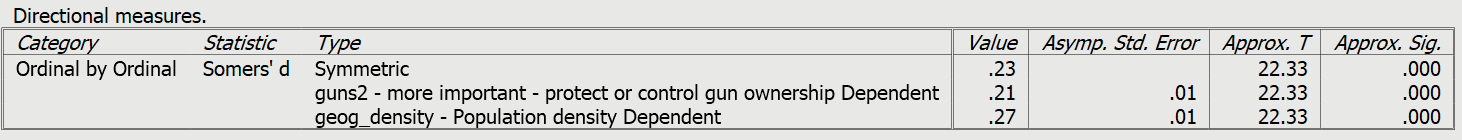


Figure 7-11

Since *guns2* is our dependent variable, the value of D that we want is .21, The significance value on the far right tells us that D is statistically significant. While .21 doesn’t seem very big, it’s at least a moderate relationship. Actually, the best use of measures of association is comparing the strength of relationships in several tables. We’ll come back to such measures in later chapters when we want to see which relationships are stronger and which are weaker.

## Part 4 – Exploring the Relationship Between Politics and Gun Ownership

In Part 2 we looked at four different measures of political behavior:

* political party identification (*pol1*),
* political identification taking into account those who lean one way or the other (*pol2*),
* political ideology (*pol3*), and
* political identification and political ideology combined (*pol4*).

In this section we’re going to explore the relationship between each of these variables and their views on gun ownership (*guns2*). Let’s start with *pol1* and *guns2*. Run the crosstabulation for these two variables. Our dependent variable is *guns2* because that’s what we’re trying to explain. Put the dependent variable, *guns2*, in the rows and the independent variable, *pol1*, in the columns. Ask for the column percents. Your screen should look like Figure 7-12.

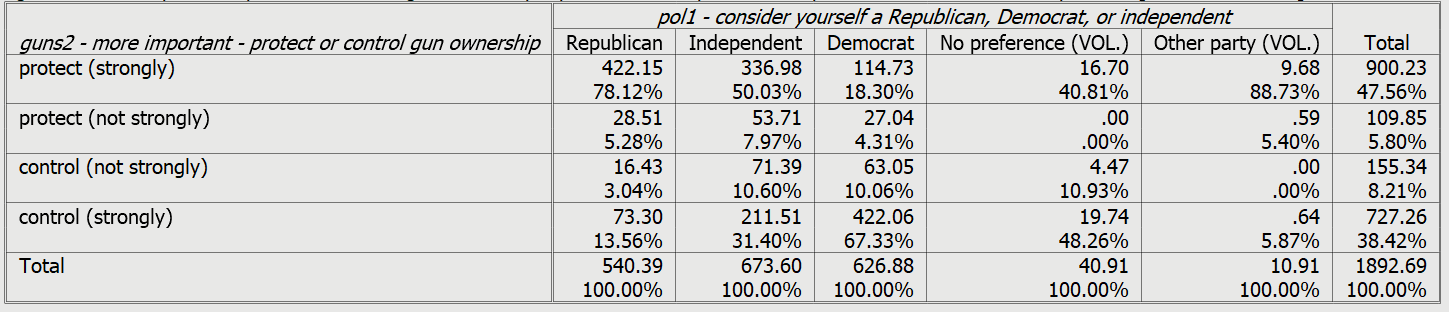


Figure 7-12

You can see from the table that relatively few respondents say, “no preference” (41) or “some other party” (11). Let’s drop them from the table. The easiest way to do that is to make them into missing values in which case they will automatically be eliminated from the table. Open the data window in PSPP and click on the VARIABLE VIEW tab at the bottom of the screen. Scroll down to the variable, *pol1*, and double-click in the MISSING VALUES cell. Enter the values 4 and 5 in the two empty cells. Look back at Figure 7-1 and you’ll see that 4 is the value for “no preference” and 5 is the value for “other party.” Click on OK in the lower left of the screen.

Now rerun the crosstabulation but this time request Chi Square and Somers’ D as your measure of association. This time the columns for “no preference” and “other party” should be gone from the table and your crosstab should look like Figure 7-13.

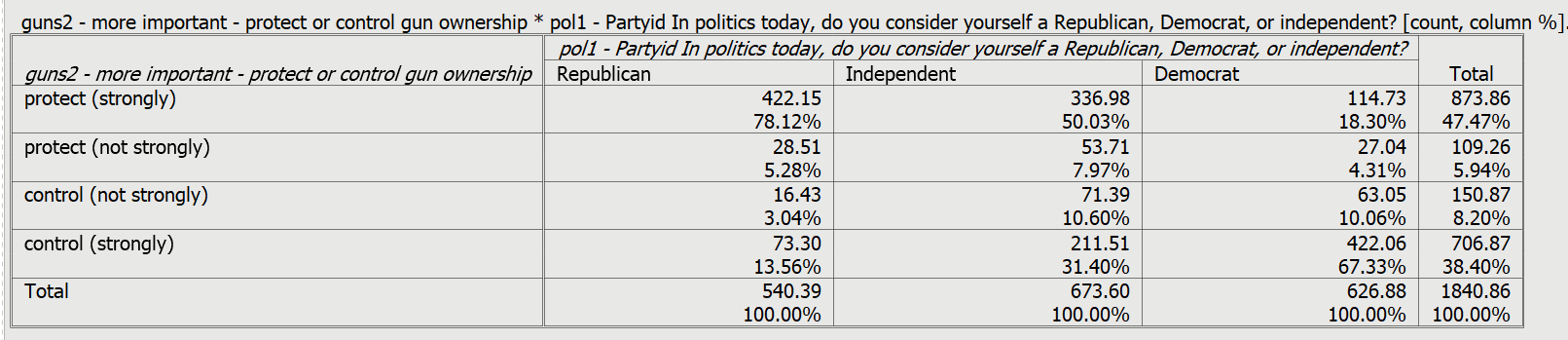


Figure 7-13

Run the crosstabs for *pol1*, *pol2*, *pol3*, and *pol4*. Remember to put the dependent variable in the rows and the independent variable in the columns. Be sure to ask for the column percents, Chi Square, and Somers’ D.

For **each** crosstabulation, answer the following questions.

* What is your hypothesis?
* Using the crosstab and the percentages, what is the relationship between the two variables? In other words, describe the pattern that you see in the percents. Use the percents to illustrate the pattern.
* Using Chi Square, is the relationship statistically significant? How do you know? What does this mean?
* Using Somers’ D as your measure of association, how strong is the relationship? Is the relationship stronger for some pairs of variables than for other pairs?
* Do the data support your hypothesis?

Now write a summary paragraph explaining what you have learned about the relationship between politics and how people view gun ownership.

## What’s Next?

In the next exercise we’ll explore the relationship between socioeconomic status and how respondents view gun ownership. We’ll use crosstabulation, Chi Square, and a measure of association as we did in this exercise.

##### Data Analysis Exercises on Gun IssuesEdward Nelson, California State University, Fresno

# Exercise 8 – Two-Variable Tables – Exploring the Relationship Between Socioeconomic Status and Views on Gun Ownership

## Part 1 – Research Questions

In Exercise 5 we explored the relationship between geography and how respondents felt about controlling and protecting gun ownership. We found that respondents in higher density areas such as larger metropolitan cities felt that controlling gun ownership was more important than protecting the right to own guns while those in smaller density areas felt just the opposite. We also discovered that respondents living in the south and the midwest were more likely to want to protect the right to own guns than those in other regions of the country.

The research study we’re using in these exercises is an August 2016 political survey conducted by the Pew Research Center.[[35]](#footnote-35) The name of the data set is “pew\_2016\_political\_survey\_subset.” You can access it from the Social Science Research and Instructional Council’s website by clicking on [this link](https://www.ssric.org/files/2019-12/pew_2016_political_survey_subset.sav). If you are using this exercise as part of a class, your instructor will show you how to access the data.

In this exercise we’re going to look at the relationship between socioeconomic status and how respondents feel about controlling and protecting gun ownership. We’re going to look at two measures of socioeconomic status – education and family income. The Pew surveys also asks respondents what they think is more important – “to protect the right of Americans to own guns, or to control gun ownership?” This variable is named *guns2*.

In Exercise 3 we discussed causal models which suggest some possible causal relationships that we want to explore. Our dependent variable is what we are trying to explain. Since we trying to explain why some respondents want to protect gun ownership and others want to control it, our dependent variable is *guns2*.

Independent variables are the variables you think might help explain your dependent variable. There are many possible variables that might explain why some respondents want to protect gun ownership and others want to control it. We discussed some of them in Exercise 3. In the next section of this exercise we’ll explore how the Pew survey measured socioeconomic status.

## Part 2 – Exploring Socioeconomic Status

Socioeconomic status (SES) is an important concept in the social sciences. A concept is an abstract idea. Power is another example of a concept. Since concepts are abstract ideas they must be defined. There are two types of definitions. One is the conceptual definition which explains what the concept means theoretically. Dictionary.com is a useful online dictionary. They define socioeconomic status as “an individuals or group’s position within a hierarchical social structure.” A second type of definition is the operational definition which specifies the way a concept is measured (i.e., the operations you carry out to measure the concept). Their operational definition says that “socioeconomic status depends on a combination of variables, including occupation, education, income, wealth, and place of residence.”[[36]](#footnote-36)

One way to measure SES is to use the respondent’s education. There are two variables that we can use in the Pew survey – *demo\_educ* and *demo\_educrec*. Run a frequency distribution for these two variables. Your screen should look like Figure 8-1 for these two variables.

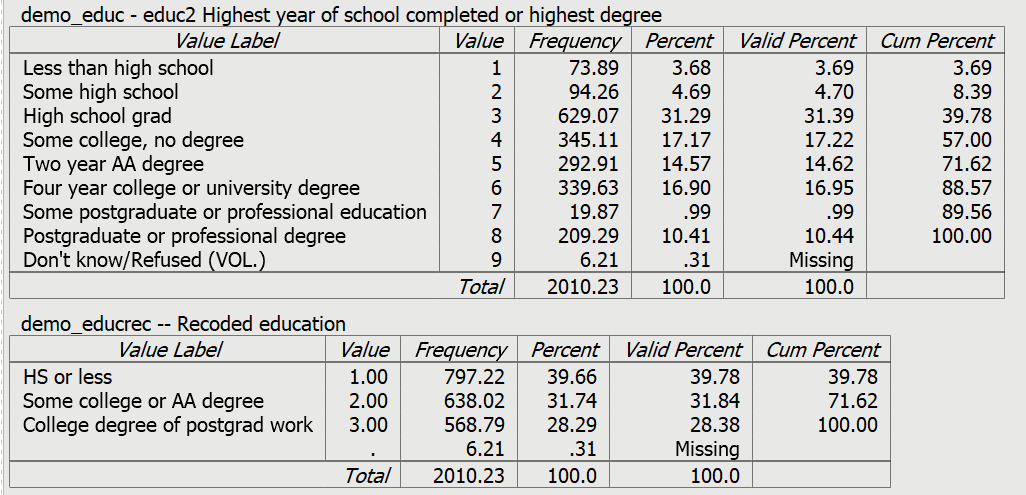


Figure 8\_1

There is more detail in the frequency distribution for demo\_*educ* but demo\_*educrec* will be easier to use. Value 1 of *demo\_educrec* combines the first three categories of *demo\_educ*; value 2 combines the next two categories; value 3 combines the last three categories.

The other measure of SES is total family income in the year previous to the survey (2015). Run the frequency distribution for *demo\_income* and *demo\_incomerec*. Your screen should look like Figure 8-2.

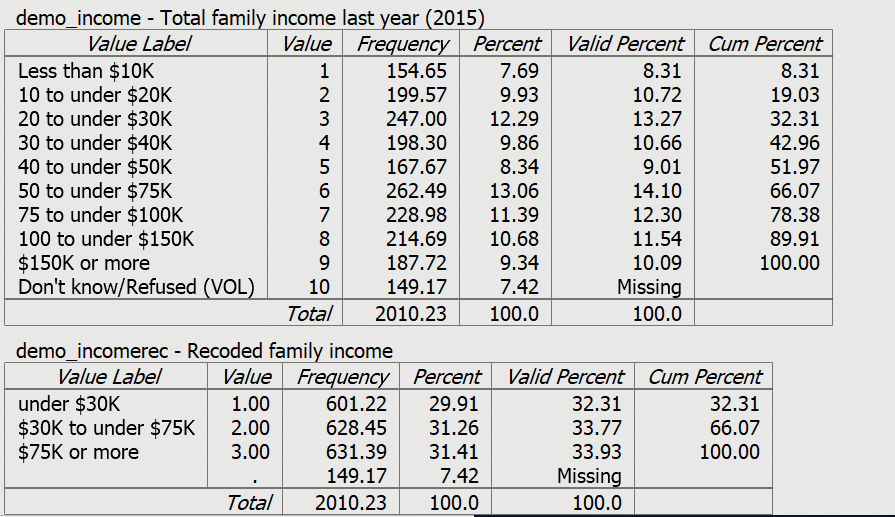


Figure 8.2

The recoded variable combines the nine categories of *demo\_income* into three categories. The first recoded category combine the first three categories; the second recoded category combines the middle three categories; the third recoded category combines the last three categories.

Write a paragraph explaining what these frequency distributions tell you about education and family income in the United States.

Now we’re ready to begin exploring the relationship between socioeconomic status and how people feel about gun ownership. We’re going to use a statistical procedure called crosstabulation to do this. We’re also going to use several statistics to help us – Chi Square and Somer’s D which is a measure of association.

Before we start, we’re going to look at a different question to demonstrate the basic principles of data analysis. If you have already worked through Exercise 5 and feel comfortable with crosstabulation, chi Square, and measures of association you can skip to Part 4.

We mentioned earlier that we discovered in Exercise 5 that respondents in higher density areas such as larger metropolitan cities felt that controlling gun ownership was more important than protecting the right to own guns while those in smaller density areas felt just the opposite. How did we come to that conclusion? How did we analyze our data? That’s what we’ll explore in the next section of this exercise. After we have demonstrated the basics of data analysis, we’ll return to the question about socioeconomic status and gun ownership.

## Part 3 – Exploring the Relationship Between Population Density and

## Views on Gun Ownership

### Stating the Hypothesis.

High density areas like large urban centers are more likely to vote Democrat while low density areas like rural areas are more likely to vote Republican. We also know that Democrats are more likely to favor controlling gun ownership while Republicans are more likely to favor protecting it. **Therefore, we could hypothesize that** **higher density areas are more likely to want to control gun ownership and lower density areas are more likely to favor protecting gun ownership***.*

### Using Crosstabulation to Test the Hypothesis.

Crosstabulation can be used to explore relationships between categorical variables (i.e., one in which cases are coded as having a relatively small number of categories). Let’s start by comparing higher and lower density areas in terms of how they feel about protecting and controlling gun ownership. Population density (*geog\_density*) will be our independent variable and how respondents feel about gun ownership (*guns2*) will be our dependent variable.

Assuming that you have already opened PSPP, click on ANALYZE in the menu bar at the top of your screen, use your mouse to point at DESCRIPTIVE STATISTICS, and then click on CROSSTABS. This will open the crosstabs dialog box. Your screen should look like Figure 8-3. If your screen displays the variable labels instead of the variable names, right click on the list of variable names and uncheck the box that says PREFER VARIABLE LABELS and now you should see the variable names.

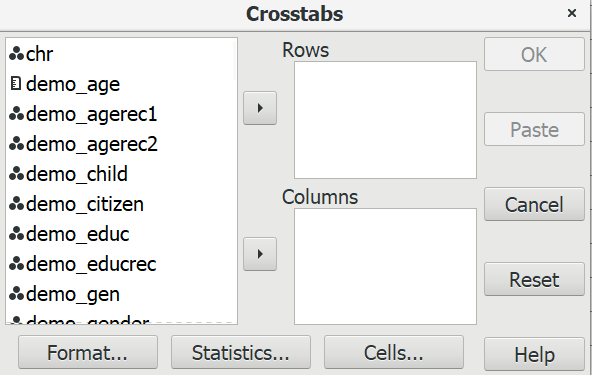


Figure 8-3

Scroll down the list of variables and find *geog\_density* and click on the right-pointing arrow to the left of columns. An easier way to do this is to enter the first letter of the variable name and PSPP will skip to the first variable that starts with a “g”. Now find *guns2* in the list of variables and click on the arrow to the left of rows. If you make a mistake, click on the variable name, make sure the arrow is pointing in the direction you want to move the variable, and click the arrow. Your screen should look like Figure 8-4.

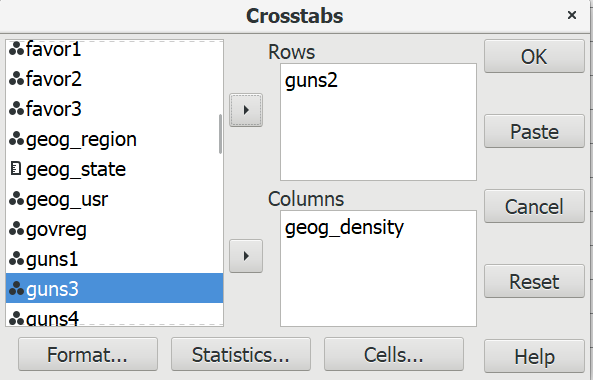


Figure 8-4

Use the following rule to decide which box to insert a variable in: “Put your independent variable in the columns and your dependent variable in the rows.”

Click on the cells button at the bottom of the dialog box. You’ll notice that the boxes for Count, Row, Column, and Total are already checked. These are the default options. Count is the number of cases in each cell of the table. Row, column, and total are three different ways of computing percents. Leave the count box checked. Now the question is which percents we want to use. Here’s a simple rule to follow: “If your independent is in the columns, choose the columns percents. If your independent variable is in the rows, choose the row percents.” Since we decided to put the independent variable in the columns, we’ll want the column percents. Leave the box for column percents checked and uncheck the boxes for the row and total percents. Your screen should look like Figure 8-5.

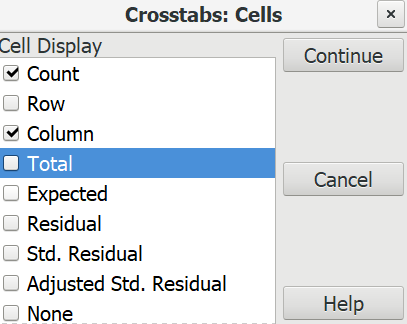


Figure 8-5

Click on the CONTINUE button to go back to Figure 8-4 and then click on OK. This will open the output window and your screen should look like Figure 8-6.

It’s easy to make sure that you have the correct percents. Your independent variable (*geog\_density*) should be in the columns and your dependent variable (*guns2*) should be in the rows. The column percents should sum down to 100%. Look at your table and make sure that you have made the correct selections.

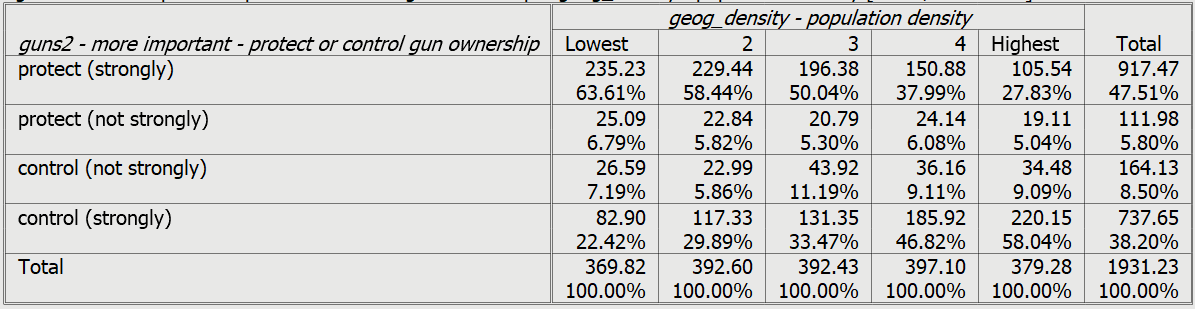


Figure 8-6

How are you going to interpret these percents?  Here’s a simple rule for interpreting percents.

* If your percents sum down to 100%, then compare the percents straight across.
* If your percents sum across to 100%, then compare the percents straight down.

Since the percents sum down to 100%, you want to compare straight across.

Look at the first row.  Approximately 64% of respondents living in the lowest density area feel strongly that it is more important to protect gun ownership. These percentages decrease as density increases. Only 28% of those who live in the highest density area strongly feel that protecting gun ownership is more important. That’s a difference of 36 percentage points.  When you look at the last row you see the opposite pattern. Respondents living in the more densely settled areas are more likely to strongly feel that controlling gun ownership is more important.

We never want to make too much of small differences.  Why not?  No sample is ever a perfect representation of the population from which the sample is drawn.  This is because every sample contains some amount of sampling error.  Sampling error is inevitable.  There is always some amount of sampling error present in every sample.  The larger the sample size, the less the sampling error and the smaller the sample size, the more the sampling error.  But in this case the differences are rather large so we would conclude that respondents living in less densely settled areas are more likely to think that protecting gun ownership is more important and those living in areas with greater density are more likely to want to control gun ownership.

### Chi Square.

In this example, we think that population density and how people feel about gun ownership are related to each other. We’ll call this our research hypothesis.  It’s what we expect to be true.  But there is no way to prove the research hypothesis directly.  So, we’re going to use a method of indirect proof.  We’re going to set up another hypothesis that says that the research hypothesis is not true and call this the null hypothesis.[[37]](#footnote-37)  In our case, the null hypothesis would be that the two variables are unrelated to each other.

Here are our two hypotheses:

* research hypothesis – population density and opinion about gun ownership are related to each other, and
* null hypothesis – population density and opinion about gun ownership are unrelated to each other. In other words, they are independent of each other.

It’s the null hypothesis that we are going to test.

When you ran the crosstab, PSPP computed Chi Square for you.  Click on the STATISTICS button and the box for Chi Square should be checked since it’s the default.

Now you will see another output box below the crosstabulation called “Chi-Square Tests.”  We want the test that is called “Pearson Chi-Square” in the first row of the box.  Ignore all the other rows in this box. Your screen should look like Figure 8-7.

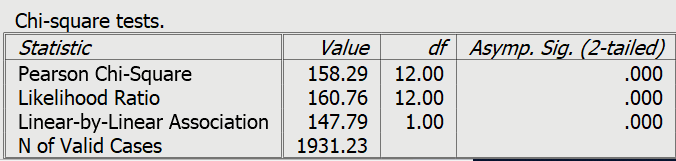


Figure 8.7

The value of Chi Square is 158.29.  Your instructor may or may not want to go into the computation of the Chi Square value but we’re not going to cover the computation in this exercise.

The degrees of freedom (df) is 12.  Degrees of freedom is the number of values that are free to vary.  This table has 4 rows and 5 columns which means there 4 x 5 or 20 cells in the table. Twelve of these cell values are free to vary and once they are filled in, the values of the remaining 8 cells are fixed.

The two-tailed significance value is 0.000. Actually, this is a rounded value. PSPP rounds to the nearest third decimal place. So, it really means that the significance value is <.0005. This tells us that there is only a tiny probability that we would be wrong if we rejected the null hypothesis.  In other words, we would be wrong less than 5 out of 10,000 times.  With odds like that we’re going to reject the null hypothesis.  A common rule is to reject the null hypothesis if the significance value is less than .05 or less than five out of one hundred.  Since, <.0005 is clearly less than .05, we reject the null hypothesis.  This means that we have support for our research hypothesis that the two variables are related.

### Measures of Association. A measure of association is a numerical value that tells us how strongly related two variables are.[[38]](#footnote-38)  There are several characteristics of a good measure of association.

* They range from a value of 0 (i.e., no relationship) to 1 (i.e., the strongest possible relationship).
* For variables that have an underlying order from low to high they can be positive or negative.  Ordinal variables have an underlying order (i.e., from high to low or from low to high) while nominal variables don’t have an underlying order. A positive value indicates that as one variable increases, the other variable also increases.  A negative value indicates that as one variable increases, the other variable decreases.[[39]](#footnote-39)
* Some measures specify which variable is dependent and which is independent.  The independent variable is some variable that you think might help you explain the variation in the dependent variable.  For example, if your two variables were education and voting you might choose education as the independent variable and voting as your dependent variable because you think that education will help explain why some people vote Democrat and others vote Republican. Measures of association that specify which variable is dependent and which is independent are called asymmetric measures and measures that don’t specify which is dependent and which is independent are called symmetric measures.

There are many measures of association to choose from. When choosing a measure of association, we’ll start by considering the level of measurement of the two variables.

* If one or both of the variables is nominal, then choose one of these measures.[[40]](#footnote-40)
  + Contingency coefficient (CC)
  + Phi
  + Cramer’s V
* If both of the variables are ordinal, then choose from this list.
  + Gamma
  + Somers’ D
  + Kendall’s tau-b
  + Kendall’s tau-c
* Dichotomies should be treated as ordinal. Most variables can be recoded into dichotomies. For example, marital status can be recoded into married or not married. Race can be recoded as white or non-white. All dichotomies should be considered ordinal.

Population density and how respondents feel about gun ownership are both ordinal variables. Both variables have an inherent order to them. Additionally, population density might influence how people feel about gun ownership. This suggests that Somers’ D would be a good choice for a measure of association since it is the only one that allows us to specify one of the variables as dependent. To tell PSPP to compute D, click on STATISTICS and check the box for D. Your screen should look like Figure 8-8.

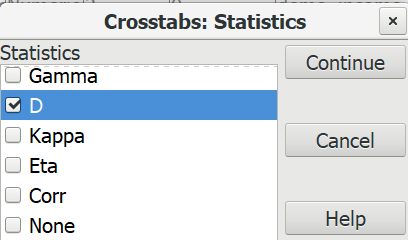


Figure 8.8

Now click on CONTINUE and then on OK. Your screen should look like Figure 8-9.

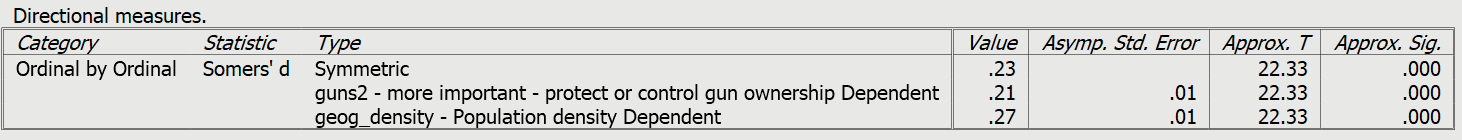


Figure 8-9

Since *guns2* is our dependent variable, the value of D that we want is .21, The significance value on the far right tells us that D is statistically significant. While .21 doesn’t seem very big, it’s at least a moderate relationship. Actually, the best use of measures of association is comparing the strength of relationships in several tables. We’ll come back to such measures in later chapters when we want to see which relationships are stronger and which are weaker.

## Part 4 – Exploring the Relationship Between Socioeconomic Status and Gun Ownership

In Part 2 we looked at two different measures of socioeconomic status:

* education (*demo\_educrec)* and
* family income (*demo\_incomerec*).

Notice that we’re using the recoded variables for both education and family income.

In this section we’re going to explore the relationship between each of these variables and their views on gun ownership (*guns2*). Let’s start with *demo\_educrec*. Run the crosstabulation for education and how people feel about gun ownership. Our dependent variable is *guns2* because that’s what we’re trying to explain. Remember to put the dependent variable in the rows and the independent variable in the columns. Be sure to ask for the column percents, Chi Square, and Somers’ D.

Next run the crosstabulation for *demo\_incomerec* and *guns2*.

For **each** crosstabulation, answer the following questions.

* What is your hypothesis?
* Using the crosstab and the percentages, what is the relationship between the two variables? In other words, describe the pattern that you see in the percents. Use the percents to illustrate the pattern.
* Using Chi Square, is the relationship statistically significant? How do you know? What does this mean?
* Using Somers’ D as your measure of association, how strong is the relationship? Is the relationship stronger for some pairs of variables than for other pairs?
* Do the data support your hypothesis?

Now write a summary paragraph explaining what you have learned about the relationship between politics and how people view gun ownership.

## What’s Next?

In the next exercise we’ll explore the relationship between age and how respondents view gun ownership. We’ll also introduce the concept of cohorts (or generations). We’ll use crosstabulation, Chi Square, and a measure of association as we did in this exercise.

##### Data Analysis Exercises on Gun IssuesEdward Nelson, California State University, Fresno

# Exercise 9 – Two-Variable Tables – Exploring the Relationship Between Age and Cohorts on Views about Gun Ownership

## Part 1 – Research Questions

In Exercise 5 we explored the relationship between geography and how respondents felt about controlling and protecting gun ownership. We found that respondents in higher density areas such as larger metropolitan cities felt that controlling gun ownership was more important than protecting the right to own guns while those in smaller density areas felt just the opposite. We also discovered that respondents living in the south and the midwest were more likely to want to protect the right to own guns than those in other regions of the country.

The research study we’re using in these exercises is an August 2016 political survey conducted by the Pew Research Center.[[41]](#footnote-41) The name of the data set is “pew\_2016\_political\_survey\_subset.” You can access it from the Social Science Research and Instructional Council’s website by clicking on [this link](https://www.ssric.org/files/2019-12/pew_2016_political_survey_subset.sav). If you are using this exercise as part of a class, your instructor will show you how to access the data.

In this exercise we’re going to look at the relationship between age and how respondents feel about controlling and protecting gun ownership. We’re also going to explore what we mean by cohorts (or generations) and see how cohorts are related to gun ownership. Our measure of how respondents feel about gun ownership will be a question from the Pew survey that asks respondents what they think is more important – “to protect the right of Americans to own guns, or to control gun ownership?” This variable is named *guns2*.

In Exercise 3 we discussed causal models which suggest some possible causal relationships that we want to explore. Our dependent variable in these models was what we are trying to explain. Since we trying to explain why some respondents want to protect gun ownership and others want to control it, our dependent variable is *guns2*.

Independent variables are the variables you think might help explain your dependent variable. There are many possible variables that might explain why some respondents want to protect gun ownership and others want to control it. We discussed some of them in Exercise 3. In the next section of this exercise we’ll discuss age and cohorts or generations.

## Part 2 – Exploring Age and Cohorts

Age and cohort are important concepts in the social sciences. A concept is an abstract idea. Since concepts are abstract ideas they must be defined. There are two types of definitions. One is the conceptual definition which explains what the concept means theoretically. Dictionary.com[[42]](#footnote-42) is a useful online dictionary which will give you a short conceptual definition of most concepts. A second type of definition is the operational definition which specifies the way a concept is measured (i.e., the operations you carry out to measure the concept).

Age is typically measured by asking respondents how old they are in years or by asking them what year they were born in. In the Pew data set, this variable is named *demo\_age*.

The definition of cohorts is less self-evident. Here’s a fairly standard definition. Remember that the Pew survey was conducted in 2016 so the ages refer to their age in that year.

* Silent generation – born 1928 to 1945 – ages 71 to 88
* Boomer generation – born 1946 to 1964 – ages 52 to 70
* Generation X – born 1965 to 1980 – ages 36 to 53
* Millennial generation – born after 1980 – ages 35 or younger

Recently researchers have referred to those born from 1997 to 2012 as generation Z. However, the Pew survey only interviewed individuals who were 18 and over. In 2016 there would be relatively few respondents falling in this cohort so we’re just going to define millennials as those born after 1980.

Cohorts or generations consist of individuals who share similar experiences in their formative years. In the Pew data set generations or cohorts is named *demo\_gen*.

* The silent generation experienced the depression and World War II.
* Boomers experienced the post-World War II baby boom and economic expansion.
* Generation X experienced the Vietnam War and various social movements including the Civil Rights Movement and the Women’s Movement
* Millennials are generally the children of boomers who became adults in the 21st century.

Run frequency distributions for *demo\_age* and *demo\_gen*. One of the things you immediately notice for age is that there are a lot of categories. In fact, too many for out purposes. So, we created two new variables – demo\_agerec1 which collapsed age into six categories and demo\_agerec2 which collapsed it into four categories. Run frequency distribution for these two variables. “Rec” means that these are recoded variables.

When you look at the frequency distribution for *demo\_gen* you’ll notice that there are a few respondents who were born before 1928. These are typically referred to as the Greatest generation but there are so few of them that we have assigned them a missing value which means they will not appear in any tables we run for this variable.

Now we’re ready to begin exploring the relationship between age and cohorts and how people feel about gun ownership. We’re going to use a statistical procedure called crosstabulation to do this. We’re also going to use several statistics to help us – Chi Square and Somer’s D which is a measure of association.

Before we start, we’re going to look at a different question to demonstrate the basic principles of data analysis. If you have already worked through Exercise 5 and feel comfortable with crosstabulation, chi Square, and measures of association you can skip to Part 3.

We mentioned earlier that we discovered in Exercise 5 that respondents in higher density areas such as larger metropolitan cities felt that controlling gun ownership was more important than protecting the right to own guns while those in smaller density areas felt just the opposite. How did we come to that conclusion? How did we analyze our data? That’s what we’ll explore in the next section of this exercise. After we have demonstrated the basics of data analysis, we’ll return to the question about age and cohorts and gun ownership.

## Part 2 – Exploring the Relationship Between Population Density and

## Views on Gun Ownership

### Stating the Hypothesis.

High density areas like large urban centers are more likely to vote Democrat while low density areas like rural areas are more likely to vote Republican. We also know that Democrats are more likely to favor controlling gun ownership while Republicans are more likely to favor protecting it. **Therefore, we could hypothesize that** **higher density areas are more likely to want to control gun ownership and lower density areas are more likely to favor protecting gun ownership***.*

### Using Crosstabulation to Test the Hypothesis.

Crosstabulation can be used to explore relationships between categorical variables (i.e., one in which cases are coded as having a relatively small number of categories). Let’s start by comparing higher and lower density areas in terms of how they feel about protecting and controlling gun ownership. Population density (*geog\_density*) will be our independent variable and how respondents feel about gun ownership (*guns2*) will be our dependent variable.

Assuming that you have already opened PSPP, click on ANALYZE in the menu bar at the top of your screen, use your mouse to point at DESCRIPTIVE STATISTICS, and then click on CROSSTABS. This will open the crosstabs dialog box. Your screen should look like Figure 9-1. If your screen displays the variable labels instead of the variable names, right click on the list of variable names and uncheck the box that says PREFER VARIABLE LABELS and now you should see the variable names.

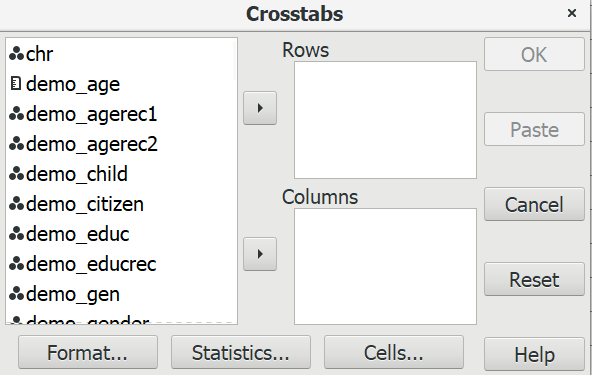


Figure 9.1

Scroll down the list of variables and find *geog\_density* and click on the right-pointing arrow to the left of columns. An easier way to do this is to enter the first letter of the variable name and PSPP will skip to the first variable that starts with a “g”. Now find *guns2* in the list of variables and click on the arrow to the left of rows. If you make a mistake, click on the variable name, make sure the arrow is pointing in the direction you want to move the variable, and click the arrow. Your screen should look like Figure 9-2.

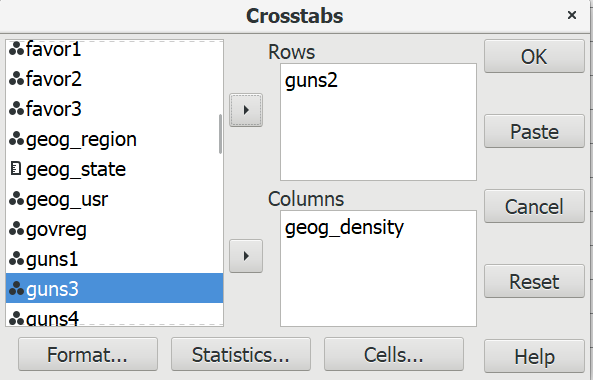


Figure 9-2

Use the following rule to decide which box to insert a variable in: “Put your independent variable in the columns and your dependent variable in the rows.”

Click on the cells button at the bottom of the dialog box. You’ll notice that the boxes for Count, Row, Column, and Total are already checked. These are the default options. Count is the number of cases in each cell of the table. Row, column, and total are three different ways of computing percents. Leave the count box checked. Now the question is which percents we want to use. Here’s a simple rule to follow: “If your independent is in the columns, choose the columns percents. If your independent variable is in the rows, choose the row percents.” Since we decided to put the independent variable in the columns, we’ll want the column percents. Leave the box for column percents checked and uncheck the boxes for the row and total percents. Your screen should look like Figure 9-3.

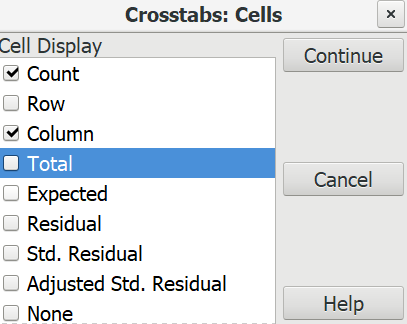


Figure 9-3

Click on the CONTINUE button to go back to Figure 9-2 and then click on OK. This will open the output window and your screen should look like Figure 9-4.

It’s easy to make sure that you have the correct percents. Your independent variable (*geog\_density*) should be in the columns and your dependent variable (*guns2*) should be in the rows. The column percents should sum down to 100%. Look at your table and make sure that you have made the correct selections.

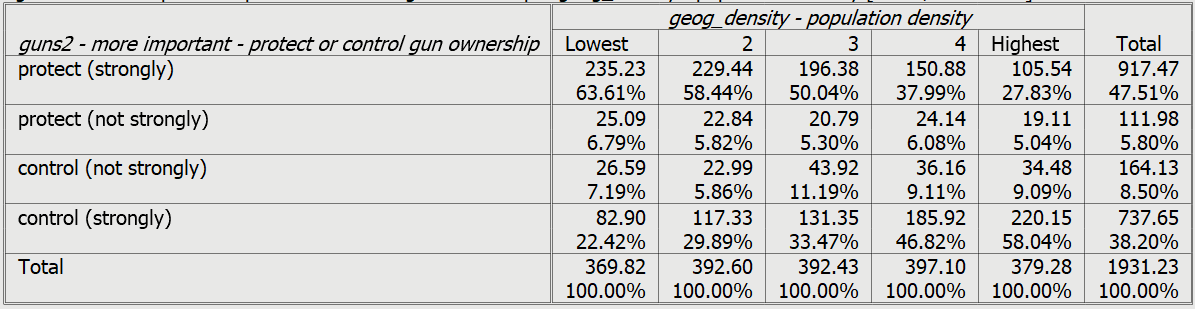


Figure 9-4

How are you going to interpret these percents?  Here’s a simple rule for interpreting percents.

* If your percents sum down to 100%, then compare the percents straight across.
* If your percents sum across to 100%, then compare the percents straight down.

Since the percents sum down to 100%, you want to compare straight across.

Look at the first row.  Approximately 64% of respondents living in the lowest density area feel strongly that it is more important to protect gun ownership. These percentages decrease as density increases. Only 28% of those who live in the highest density area strongly feel that protecting gun ownership is more important. That’s a difference of 36 percentage points.  When you look at the last row you see the opposite pattern. Respondents living in the more densely settled areas are more likely to strongly feel that controlling gun ownership is more important.

We never want to make too much of small differences.  Why not?  No sample is ever a perfect representation of the population from which the sample is drawn.  This is because every sample contains some amount of sampling error.  Sampling error is inevitable.  There is always some amount of sampling error present in every sample.  The larger the sample size, the less the sampling error and the smaller the sample size, the more the sampling error.  But in this case the differences are rather large so we would conclude that respondents living in less densely settled areas are more likely to think that protecting gun ownership is more important and those living in areas with greater density are more likely to want to control gun ownership.

### Chi Square.

In this example, we think that population density and how people feel about gun ownership are related to each other. We’ll call this our research hypothesis.  It’s what we expect to be true.  But there is no way to prove the research hypothesis directly.  So, we’re going to use a method of indirect proof.  We’re going to set up another hypothesis that says that the research hypothesis is not true and call this the null hypothesis.[[43]](#footnote-43)  In our case, the null hypothesis would be that the two variables are unrelated to each other.

Here are our two hypotheses:

* research hypothesis – population density and opinion about gun ownership are related to each other, and
* null hypothesis – population density and opinion about gun ownership are unrelated to each other. In other words, they are independent of each other.

It’s the null hypothesis that we are going to test.

When you ran the crosstab, PSPP computed Chi Square for you.  Click on the STATISTICS button and the box for Chi Square should be checked since it’s the default.

Now you will see another output box below the crosstabulation called “Chi-Square Tests.”  We want the test that is called “Pearson Chi-Square” in the first row of the box.  Ignore all the other rows in this box. Your screen should look like Figure 9-5.

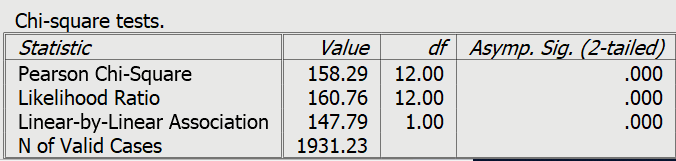


Figure 9-5

The value of Chi Square is 158.29.  Your instructor may or may not want to go into the computation of the Chi Square value but we’re not going to cover the computation in this exercise.

The degrees of freedom (df) is 12.  Degrees of freedom is the number of values that are free to vary.  This table has 4 rows and 5 columns which means there 4 x 5 or 20 cells in the table. Twelve of these cell values are free to vary and once they are filled in, the values of the remaining 8 cells are fixed.

The two-tailed significance value is 0.000. Actually, this is a rounded value. PSPP rounds to the nearest third decimal place. So, it really means that the significance value is <.0005. This tells us that there is only a tiny probability that we would be wrong if we rejected the null hypothesis.  In other words, we would be wrong less than 5 out of 10,000 times.  With odds like that we’re going to reject the null hypothesis.  A common rule is to reject the null hypothesis if the significance value is less than .05 or less than five out of one hundred.  Since, <.0005 is clearly less than .05, we reject the null hypothesis.  This means that we have support for our research hypothesis that the two variables are related.

### Measures of Association. A measure of association is a numerical value that tells us how strongly related two variables are.[[44]](#footnote-44)  There are several characteristics of a good measure of association.

* They range from a value of 0 (i.e., no relationship) to 1 (i.e., the strongest possible relationship).
* For variables that have an underlying order from low to high they can be positive or negative.  Ordinal variables have an underlying order (i.e., from high to low or from low to high) while nominal variables don’t have an underlying order. A positive value indicates that as one variable increases, the other variable also increases.  A negative value indicates that as one variable increases, the other variable decreases.[[45]](#footnote-45)
* Some measures specify which variable is dependent and which is independent.  The independent variable is some variable that you think might help you explain the variation in the dependent variable.  For example, if your two variables were education and voting you might choose education as the independent variable and voting as your dependent variable because you think that education will help explain why some people vote Democrat and others vote Republican. Measures of association that specify which variable is dependent and which is independent are called asymmetric measures and measures that don’t specify which is dependent and which is independent are called symmetric measures.

There are many measures of association to choose from. When choosing a measure of association, we’ll start by considering the level of measurement of the two variables.

* If one or both of the variables is nominal, then choose one of these measures.[[46]](#footnote-46)
  + Contingency coefficient (CC)
  + Phi
  + Cramer’s V
* If both of the variables are ordinal, then choose from this list.
  + Gamma
  + Somers’ D
  + Kendall’s tau-b
  + Kendall’s tau-c
* Dichotomies should be treated as ordinal. Most variables can be recoded into dichotomies. For example, marital status can be recoded into married or not married. Race can be recoded as white or non-white. All dichotomies should be considered ordinal.

Population density and how respondents feel about gun ownership are both ordinal variables. Both variables have an inherent order to them. Additionally, population density might influence how people feel about gun ownership. This suggests that Somers’ D would be a good choice for a measure of association since it is the only one that allows us to specify one of the variables as dependent. To tell PSPP to compute D, click on STATISTICS and check the box for D. Your screen should look like Figure 9-6.

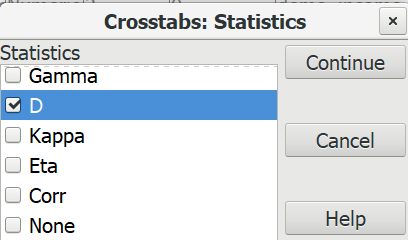


Figure 9-6

Now click on CONTINUE and then on OK. Your screen should look like Figure 9-7.

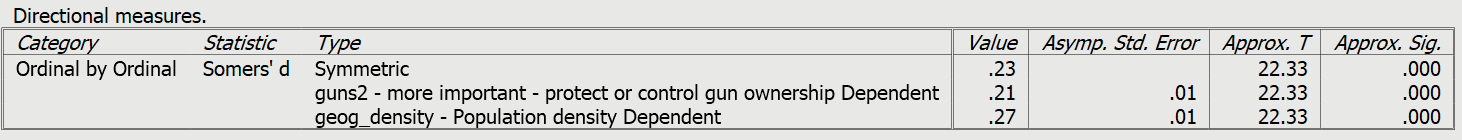


Figure 9-7

Since *guns2* is our dependent variable, the value of D that we want is .21, The significance value on the far right tells us that D is statistically significant. While .21 doesn’t seem very big, it’s at least a moderate relationship. Actually, the best use of measures of association is comparing the strength of relationships in several tables. We’ll come back to such measures in later chapters when we want to see which relationships are stronger and which are weaker.

## Part 3 – Exploring the Relationship Between Age and Cohorts and Gun Ownership

In this section we’re going to explore the relationship between respondents’ age and cohort and their views on gun ownership (*guns2*). We’re not going to use *demo\_age* as our measure of age. Rather we’re going to use *demo\_agerec1* which recodes age into six categories.

Let’s start with *demo\_agerec1*. Run the crosstabulation for age and how people feel about gun ownership. Our dependent variable is *guns2* because that’s what we’re trying to explain. Remember to put the dependent variable in the rows and the independent variable in the columns. Be sure to ask for the column percents, Chi Square, and Somers’ D.

Next run the crosstabulation for *demo\_gen* and *guns2*.

For **each** crosstabulation, answer the following questions.

* What is your hypothesis?
* Using the crosstab and the percentages, what is the relationship between the two variables? In other words, describe the pattern that you see in the percents. Use the percents to illustrate the pattern.
* Using Chi Square, is the relationship statistically significant? How do you know? What does this mean?
* Using Somers’ D as your measure of association, how strong is the relationship? Is the relationship stronger for one pair of variables than for other pair?
* Do the data support your hypothesis?

Now write a summary paragraph explaining what you have learned about the relationship between age and cohorts and how people view gun ownership.

## What’s Next?

In the next exercise we’ll explore the relationship between gender and how respondents view gun ownership. We’ll use crosstabulation, Chi Square, and a measure of association as we did in this exercise.

##### Data Analysis Exercises on Gun IssuesEdward Nelson, California State University, Fresno

# Exercise 10 – Two-Variable Tables – Exploring the Relationship Between Gender and Views on Gun Ownership

## Part 1 – Research Questions

In Exercise 5 we explored the relationship between geography and how respondents felt about controlling and protecting gun ownership. We found that respondents in higher density areas such as larger metropolitan cities felt that controlling gun ownership was more important than protecting the right to own guns while those in smaller density areas felt just the opposite. We also discovered that respondents living in the south and the midwest were more likely to want to protect the right to own guns than those in other regions of the country.

The research study we’re using in these exercises is an August 2016 political survey conducted by the Pew Research Center.[[47]](#footnote-47) The name of the data set is “pew\_2016\_political\_survey\_subset.” You can access it from the Social Science Research and Instructional Council’s website by clicking on [this link](https://www.ssric.org/files/2019-12/pew_2016_political_survey_subset.sav). If you are using this exercise as part of a class, your instructor will show you how to access the data.

In this exercise we’re going to look at the relationship between gender and how respondents feel about controlling and protecting gun ownership. Our measure of how respondents feel about gun ownership will be a question from the Pew survey that asks respondents what they think is more important – “to protect the right of Americans to own guns, or to control gun ownership?” This variable is named *guns2*.

In Exercise 3 we discussed causal models which suggest some possible causal relationships that we want to explore. Our dependent variable in these models was what we are trying to explain. Since we trying to explain why some respondents want to protect gun ownership and others want to control it, our dependent variable is *guns2*.

Independent variables are the variables you think might help explain your dependent variable. There are many possible variables that might explain why some respondents want to protect gun ownership and others want to control it. We discussed some of them in Exercise 3. In this exercise we’re going to focus on gender as a possible independent variable.

## Part 2 – Measuring Gender

The data for the Pew survey we’re using was a telephone survey of adults in the U.S. At the conclusion of the survey interviewers were instructed to record the respondents’ sex. Recent surveys are beginning to ask respondents whether they identify as male, female, or non-binary. Some surveys ask more detailed questions about transgender individuals. No single way of asking this question has emerged as of yet. In this 2016 Pew survey we will be using the variable *demo\_gender* which is based on the interviewer’s perception of the respondent’s sex.

Now we’re ready to begin exploring the relationship between gender how people feel about gun ownership. We’re going to use a statistical procedure called crosstabulation to do this. We’re also going to use several statistics to help us – Chi Square and Somer’s D which is a measure of association.

Before we start, we’re going to look at a different question to demonstrate the basic principles of data analysis. If you have already worked through Exercise 5 and feel comfortable with crosstabulation, chi Square, and measures of association you can skip to Part 3.

We mentioned earlier that we discovered in Exercise 5 that respondents in higher density areas such as larger metropolitan cities felt that controlling gun ownership was more important than protecting the right to own guns while those in smaller density areas felt just the opposite. How did we come to that conclusion? How did we analyze our data? That’s what we’ll explore in the next section of this exercise. After we have demonstrated the basics of data analysis, we’ll return to the question about gender and gun ownership.

## Part 2 – Exploring the Relationship Between Population Density and

## Views on Gun Ownership

### Stating the Hypothesis.

High density areas like large urban centers are more likely to vote Democrat while low density areas like rural areas are more likely to vote Republican. We also know that Democrats are more likely to favor controlling gun ownership while Republicans are more likely to favor protecting it. **Therefore, we could hypothesize that** **higher density areas are more likely to want to control gun ownership and lower density areas are more likely to favor protecting gun ownership***.*

### Using Crosstabulation to Test the Hypothesis.

Crosstabulation can be used to explore relationships between categorical variables (i.e., one in which cases are coded as having a relatively small number of categories). Let’s start by comparing higher and lower density areas in terms of how they feel about protecting and controlling gun ownership. Population density (*geog\_density*) will be our independent variable and how respondents feel about gun ownership (*guns2*) will be our dependent variable.

Assuming that you have already opened PSPP, click on ANALYZE in the menu bar at the top of your screen, use your mouse to point at DESCRIPTIVE STATISTICS, and then click on CROSSTABS. This will open the crosstabs dialog box. Your screen should look like Figure 10-1. If your screen displays the variable labels instead of the variable names, right click on the list of variable names and uncheck the box that says PREFER VARIABLE LABELS and now you should see the variable names.

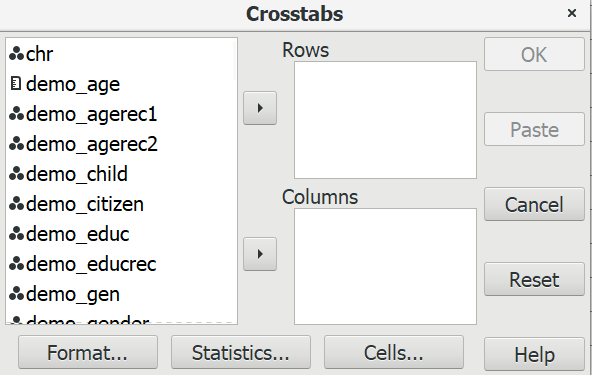


Figure 10-1

Scroll down the list of variables and find *geog\_density* and click on the right-pointing arrow to the left of columns. An easier way to do this is to enter the first letter of the variable name and PSPP will skip to the first variable that starts with a “g”. Now find *guns2* in the list of variables and click on the arrow to the left of rows. If you make a mistake, click on the variable name, make sure the arrow is pointing in the direction you want to move the variable, and click the arrow. Your screen should look like Figure 10-2.

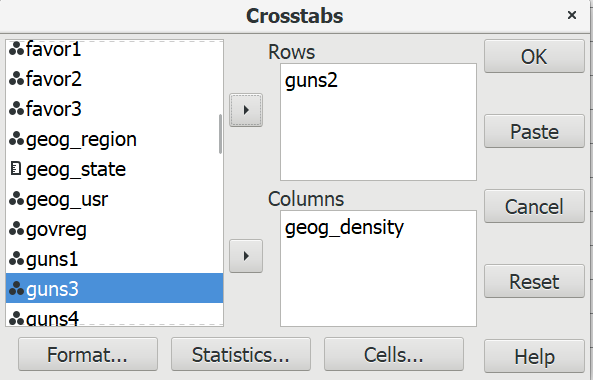


Figure 10-2

Use the following rule to decide which box to insert a variable in: “Put your independent variable in the columns and your dependent variable in the rows.”

Click on the cells button at the bottom of the dialog box. You’ll notice that the boxes for Count, Row, Column, and Total are already checked. These are the default options. Count is the number of cases in each cell of the table. Row, column, and total are three different ways of computing percents. Leave the count box checked. Now the question is which percents we want to use. Here’s a simple rule to follow: “If your independent is in the columns, choose the columns percents. If your independent variable is in the rows, choose the row percents.” Since we decided to put the independent variable in the columns, we’ll want the column percents. Leave the box for column percents checked and uncheck the boxes for the row and total percents. Your screen should look like Figure 10-3.

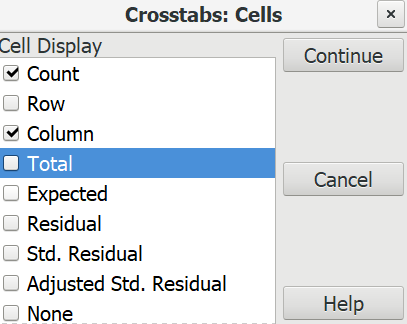


Figure 10-3

Click on the CONTINUE button to go back to Figure 9-2 and then click on OK. This will open the output window and your screen should look like Figure 10-4.

It’s easy to make sure that you have the correct percents. Your independent variable (*geog\_density*) should be in the columns and your dependent variable (*guns2*) should be in the rows. The column percents should sum down to 100%. Look at your table and make sure that you have made the correct selections.

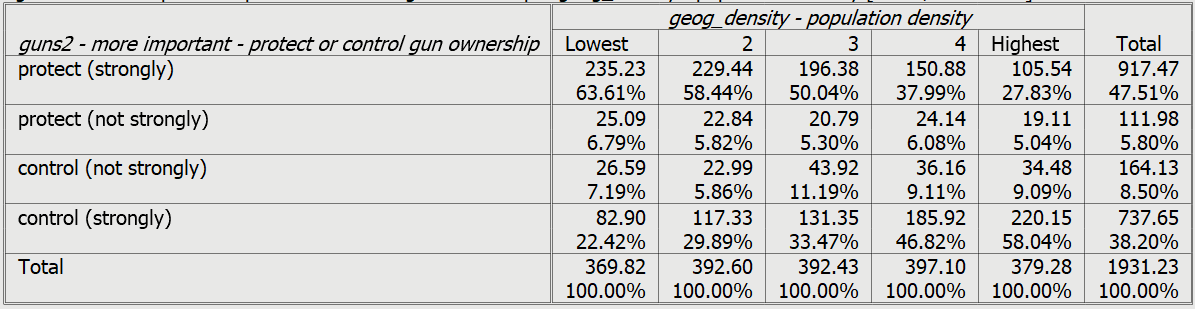


Figure 10-4

How are you going to interpret these percents?  Here’s a simple rule for interpreting percents.

* If your percents sum down to 100%, then compare the percents straight across.
* If your percents sum across to 100%, then compare the percents straight down.

Since the percents sum down to 100%, you want to compare straight across.

Look at the first row.  Approximately 64% of respondents living in the lowest density area feel strongly that it is more important to protect gun ownership. These percentages decrease as density increases. Only 28% of those who live in the highest density area strongly feel that protecting gun ownership is more important. That’s a difference of 36 percentage points.  When you look at the last row you see the opposite pattern. Respondents living in the more densely settled areas are more likely to strongly feel that controlling gun ownership is more important.

We never want to make too much of small differences.  Why not?  No sample is ever a perfect representation of the population from which the sample is drawn.  This is because every sample contains some amount of sampling error.  Sampling error is inevitable.  There is always some amount of sampling error present in every sample.  The larger the sample size, the less the sampling error and the smaller the sample size, the more the sampling error.  But in this case the differences are rather large so we would conclude that respondents living in less densely settled areas are more likely to think that protecting gun ownership is more important and those living in areas with greater density are more likely to want to control gun ownership.

### Chi Square.

In this example, we think that population density and how people feel about gun ownership are related to each other. We’ll call this our research hypothesis.  It’s what we expect to be true.  But there is no way to prove the research hypothesis directly.  So, we’re going to use a method of indirect proof.  We’re going to set up another hypothesis that says that the research hypothesis is not true and call this the null hypothesis.[[48]](#footnote-48)  In our case, the null hypothesis would be that the two variables are unrelated to each other.

Here are our two hypotheses:

* research hypothesis – population density and opinion about gun ownership are related to each other, and
* null hypothesis – population density and opinion about gun ownership are unrelated to each other. In other words, they are independent of each other.

It’s the null hypothesis that we are going to test.

When you ran the crosstab, PSPP computed Chi Square for you.  Click on the STATISTICS button and the box for Chi Square should be checked since it’s the default

Now you will see another output box below the crosstabulation called “Chi-Square Tests.”  We want the test that is called “Pearson Chi-Square” in the first row of the box.  Ignore all the other rows in this box. Your screen should look like Figure 10-5.

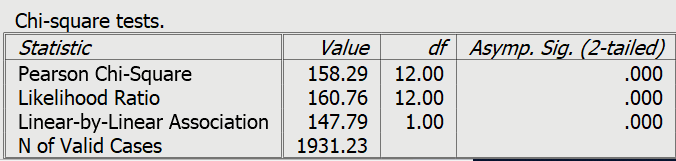


Figure 10-5

The value of Chi Square is 158.29.  Your instructor may or may not want to go into the computation of the Chi Square value but we’re not going to cover the computation in this exercise.

The degrees of freedom (df) is 12.  Degrees of freedom is the number of values that are free to vary.  This table has 4 rows and 5 columns which means there 4 x 5 or 20 cells in the table. Twelve of these cell values are free to vary and once they are filled in, the values of the remaining 8 cells are fixed.

The two-tailed significance value is 0.000. Actually, this is a rounded value. PSPP rounds to the nearest third decimal place. So, it really means that the significance value is <.0005. This tells us that there is only a tiny probability that we would be wrong if we rejected the null hypothesis.  In other words, we would be wrong less than 5 out of 10,000 times.  With odds like that we’re going to reject the null hypothesis.  A common rule is to reject the null hypothesis if the significance value is less than .05 or less than five out of one hundred.  Since, <.0005 is clearly less than .05, we reject the null hypothesis.  This means that we have support for our research hypothesis that the two variables are related.

### Measures of Association. A measure of association is a numerical value that tells us how strongly related two variables are.[[49]](#footnote-49)  There are several characteristics of a good measure of association.

* They range from a value of 0 (i.e., no relationship) to 1 (i.e., the strongest possible relationship).
* For variables that have an underlying order from low to high they can be positive or negative.  Ordinal variables have an underlying order (i.e., from high to low or from low to high) while nominal variables don’t have an underlying order. A positive value indicates that as one variable increases, the other variable also increases.  A negative value indicates that as one variable increases, the other variable decreases.[[50]](#footnote-50)
* Some measures specify which variable is dependent and which is independent.  The independent variable is some variable that you think might help you explain the variation in the dependent variable.  For example, if your two variables were education and voting you might choose education as the independent variable and voting as your dependent variable because you think that education will help explain why some people vote Democrat and others vote Republican. Measures of association that specify which variable is dependent and which is independent are called asymmetric measures and measures that don’t specify which is dependent and which is independent are called symmetric measures.

There are many measures of association to choose from. When choosing a measure of association, we’ll start by considering the level of measurement of the two variables.

* If one or both of the variables is nominal, then choose one of these measures.[[51]](#footnote-51)
  + Contingency coefficient (CC)
  + Phi
  + Cramer’s V
* If both of the variables are ordinal, then choose from this list.
  + Gamma
  + Somers’ D
  + Kendall’s tau-b
  + Kendall’s tau-c
* Dichotomies should be treated as ordinal. Most variables can be recoded into dichotomies. For example, marital status can be recoded into married or not married. Race can be recoded as white or non-white. All dichotomies should be considered ordinal.

Population density and how respondents feel about gun ownership are both ordinal variables. Both variables have an inherent order to them. Additionally, population density might influence how people feel about gun ownership. This suggests that Somers’ D would be a good choice for a measure of association since it is the only one that allows us to specify one of the variables as dependent. To tell PSPP to compute D, click on STATISTICS and check the box for D. Your screen should look like Figure 10-6.

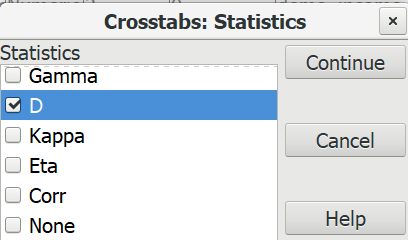


Figure 10-6

Now click on CONTINUE and then on OK. Your screen should look like Figure 10-7.

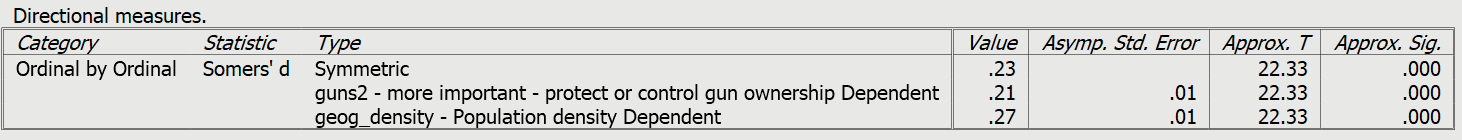


Figure 10-7

Since *guns2* is our dependent variable, the value of D that we want is .21, The significance value on the far right tells us that D is statistically significant. While .21 doesn’t seem very big, it’s at least a moderate relationship. Actually, the best use of measures of association is comparing the strength of relationships in several tables. We’ll come back to such measures in later chapters when we want to see which relationships are stronger and which are weaker.

## Part 3 – Exploring the Relationship Between Gender and Views on Gun Ownership

In this section we’re going to explore the relationship between gender and respondents’ views on gun ownership (*guns2*). So, our independent variable will be *demo\_gender*.

Run the crosstabulation for gender and how people feel about gun ownership. Our dependent variable is *guns2* because that’s what we’re trying to explain. Remember to put the dependent variable in the rows and the independent variable in the columns. Be sure to ask for the column percents, Chi Square, and Somers’ D.

Answer the following questions for this table.

* What is your hypothesis?
* Using the crosstab and the percentages, what is the relationship between the two variables? In other words, describe the pattern that you see in the percents. Use the percents to illustrate the pattern.
* Using Chi Square, is the relationship statistically significant? How do you know? What does this mean?
* Using Somers’ D as your measure of association, how strong is the relationship? Is the relationship stronger for one pair of variables than for the other pair?
* Do the data support your hypothesis?

Now write a summary paragraph explaining what you have learned about the relationship between gender and how people view gun ownership.

## What’s Next?

In the next exercise we’ll move from looking at relationships between pairs of variables (bivariate analysis) to exploring relationships among sets of three variables (multivariate analysis). We’ll consider which is more important – politics or gender and we’ll introduce the concept of spuriousness. We’ll use crosstabulation, Chi Square, and a measure of association as we did in this exercise.

##### Data Analysis Exercises on Gun IssuesEdward Nelson, California State University, Fresno

# Exercise 11 – Three Variable Tables – Exploring the Relationships of Politics, Gender, and Views on Gun Ownership

## Part 1 – Research Questions

In Exercises 5 through 10 we explored relationships between pairs of variables which is often referred to as bivariate analysis. Exercises 11 through 13 focus on sets of three variables which is often referred to as multivariate analysis. Multivariate analysis isn’t limited to just three variables but that’s what we’re going to focus on in these exercises.

The research study we’re using in these exercises is an August 2016 political survey conducted by the Pew Research Center.[[52]](#footnote-52) The name of the data set is “pew\_2016\_political\_survey\_subset.” You can access it from the Social Science Research and Instructional Council’s website by clicking on [this link](https://www.ssric.org/files/2019-12/pew_2016_political_survey_subset.sav). If you are using this exercise as part of a class, your instructor will show you how to access the data.

In this exercise we’re going to look at politics, gender, and how respondents feel about controlling and protecting gun ownership. Our measure of how respondents feel about gun ownership will be a question from the Pew survey that asks respondents what they think is more important – “to protect the right of Americans to own guns, or to control gun ownership?” This variable is named *guns2*. We’re going to use a recode of this variable which is named *guns2rec*.

Our dependent variable is what we’re trying to explain. Since we trying to explain why some respondents want to protect gun ownership and others want to control it, our dependent variable is *guns2rec*.

Independent variables are the variables you think might help explain your dependent variable. There are many possible variables that might explain why some respondents want to protect gun ownership and others want to control it. We discussed some of them in Exercise 3. In this exercise we’re going to focus on politics and gender as possible independent variables.

Often, we identify some variable that we think might be affecting the relationship between our dependent and independent variables and call this the control variable. There are several reasons we might want to introduce a control variable into the analysis. We’ll talk about control variables in this exercise.

## Part 2 – Measuring Gender

The Pew survey we’re using was a telephone survey of adults in the U.S. At the conclusion of the survey interviewers recorded the respondent’s sex. Recent surveys are beginning to ask respondents whether they identify as male, female, or non-binary. Some surveys ask more detailed questions about transgender individuals. No single way of asking this question has emerged as of yet. In the 2016 Pew survey we’ll be using the variable *demo\_gender* which is based on the interviewer’s perception of the respondent’s sex.

## Part 3 – Exploring the Two-Variable Relationships for Gender, Politics, and How Respondents Feel about Gun Ownership

Let’s start by looking at the two-variable relationship between gender and both politics and how respondents’ feel about gun ownership. As in the previous exercises, we’re going to use a statistical procedure called crosstabulation. We’re also going to use several statistics to help us – Chi Square and Somers’ D which is a measure of association.

## How is *demo\_gender* related to both politics and how people feel about gun ownership? There are several variables we could use to operationalize politics but we’ll focus on *pol3rec* and *pol4*. *Pol3rec* is a recode of *pol3* and describes respondents’ political ideology as conservative, moderate, or liberal. *Pol4* is a combination of political ideology and political party and classifies respondents as conservative Republican, moderate or liberal Republican, independent, conservative or moderate Democrat, or liberal Democrat.

## Run the crosstab for gender and each of the political variables. *Demo\_gender* should be your independent variable. Be sure to put your independent variable in the columns and get the column percents, Chi Square, and Somers’D. If you need a review of two-variable tables, refer to Exercise 5.

## Write a paragraph describing the relationship between gender and each of the two political variables. Be sure to use the statistics you told PSPP to compute in your paragraph.

Now let’s turn our attention to the relationship between each of the political variables and how respondents feel about gun ownership using *guns2rec* as your dependent variable. *Guns2rec* is a recode of *guns2* and dichotomizes this variable into two categories – supports protecting gun ownership and supports controlling gun ownership.  
  
Write a paragraph describing the relationships between each of the two political variables and how people feel about gun ownership. Be sure to use the statistics you ran in your paragraph.

## Part 4 – Exploring the Three-Variable Relationships for Gender, Political Ideology, and How Respondents Feel about Gun Ownership

## Now let’s turn our attention to three-variable relationships. Unfortunately, PSPP doesn’t let you easily run three-variable tables from the graphical interface. But you can run them from the syntax window. Click on FILE in the menu bar at the top of your screen and then on NEW and finally on SYNTAX. We’re going to enter the following crosstabs command in the syntax window. Here’s the command. You can copy and paste the command into the syntax window or you can type it in yourself.

## CROSSTABS

## /TABLES = GUNS2REC BY POL3REC BY DEMO\_GENDER

## /STATISTICS = CHISQ D

/CELLS = COUNT COLUMN.

Here’s a brief explanation for each of the parts of this command.

* CROSSTABS is the name of the command.
* TABLES is the subcommand tells PSPP that GUNS2REC is the dependent variable, POL3REC is the independent variable, and DEMO\_GENDER is the control variable.
* STATISTICS is the subcommand that tells PSPP which statistics you want to run (Chi Square and Somers’ D).
* CELLS tells PSPP which statistics you want in each of the cells of the table (frequencies or counts and the column percents).
* The forward slash (/) is the separator that separates the different parts of the command.
* The period (.) at the end of the entire command tells PSPP that this ends the command.
* It doesn’t matter whether you use capitals or lower-case letters.
* To run the command, click on RUN in the menu bar in the syntax window. Then click on ALL in the drop-down menu which means to run all the commands.

## Now click on WINDOW in the menu bar and then check the box for the output window. Your screen should look like Figure 11-1.

## Crosstabs output

Figure 11-1

Notice that you have two tables – one for males and the other for females. Often these are referred to as partial tables since each of the tables contains part, but not all, of the cases in your sample. In each cell of the table, there are two numbers – the count or number of cases in each cell and the column percents. Notice that the percents sum down to 100 telling you that these are column percents.

While the percents are different in the partial tables for males and females, the pattern in the same. Conservatives are more likely to want to protect gun ownership while liberals are more likely to want to control gun ownership. What do the Chi Square tests and the measures of association tell you? Your screen should look like Figure 11-2.

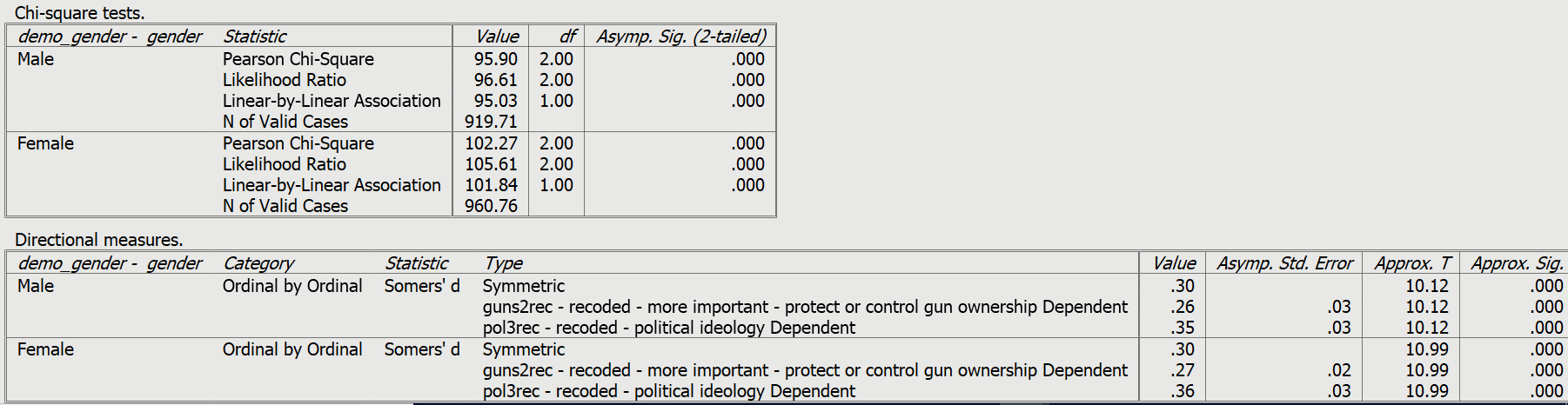


Figure 11-2

Both of the Chi Square tests are statistically significant and the Somers’ D values are both moderately large and statistically significant. This tells us that these relationships are unlikely to have occurred by chance and are moderately strong. In other words, political ideology is related to how respondents feel about gun control. Conservatives are more likely to want to protect gun ownership and liberals are more likely to want to control gun ownership. And this is true for both males and females.

Now let’s change the TABLES subcommand and rerun this three-variable table. Now your syntax should look like this.

## CROSSTABS

## /TABLES = GUNS2REC BY DEMO\_GENDER BY POL3REC

## /STATISTICS = CHISQ D

/CELLS = COUNT COLUMN.

Run this command. Your screen should look like Figure 11-3.

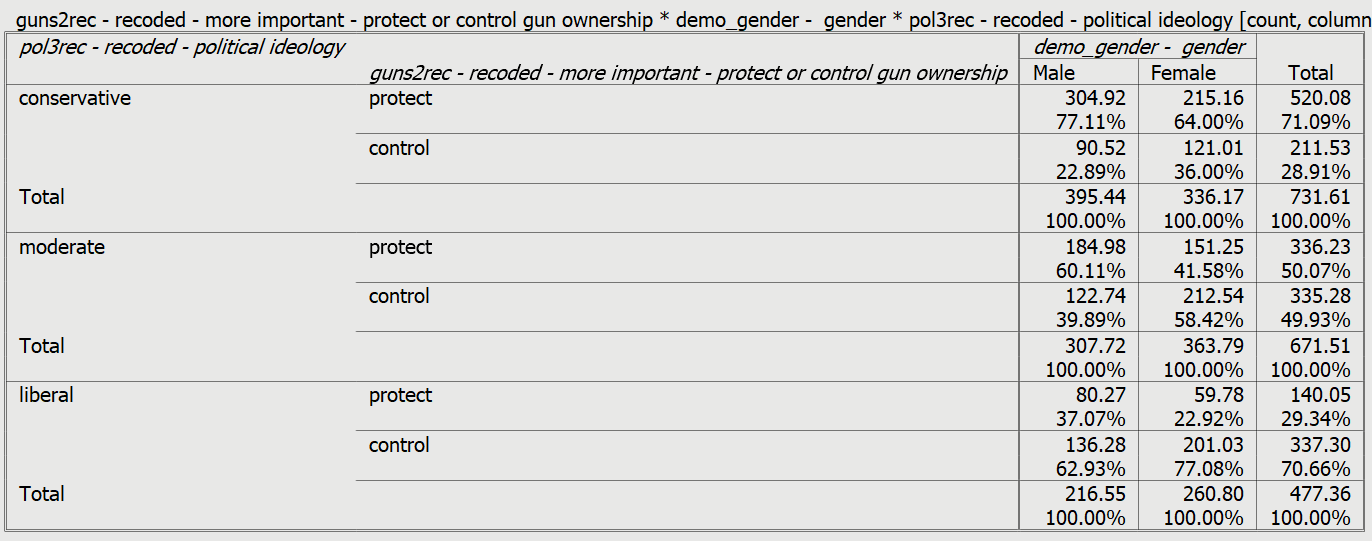


Figure 11-3

Notice what happened. Now the table shows the relationship between gender and respondents’ views on gun ownership holding political ideology constant. Now there are three partial tables. The percents have been rearranged. The Chi Square tests are statistically significant and the Somers’ D values are smaller but still statistically significant.

Since our dependent variable is a dichotomy (i.e., has only two values) we can rearrange the values into a table that is easier to read. Here’s what that would look like.

Percent that Favor Controlling Gun Ownership   
by Gender and Political Ideology

|  |  |  |  |
| --- | --- | --- | --- |
|  | Male | Female | Parentage Point Diff |
| Conservative | 22.9 | 36.0 | -13.1 |
| Moderate | 39.9 | 58.4 | -18.5 |
| Liberal | 62.9 | 77.1 | -14.2 |
| Percentage Point Diff | -40.0 | -41.1 |  |

Figure 11-4

Notice that we have entered the percent that favored controlling gun ownership. We could have entered the percent that favored protecting gun ownership. The numbers would have been different but the pattern wouldn’t have changed.

We computed the percentage point difference by subtracting the percent for females from the percent for males. We could have subtracted the percent for males from the percent for females. That would have changed the sign from negative to positive but it wouldn’t have changed the numerical values.

And we could have used the crosstab in either Figure 11-1 or 11-3 to fill in the percents in the table. We would have gotten the same results.

What does this table tell us?

* Females are more likely to want to control gun ownership than males for all three political ideologies – conservative, moderates, and liberal. The percentage differences vary but not very much meaning that the relationship between gender and how respondents feel about gun ownership is about the same for all three ideologies.
* Liberals are much more likely to want to control gun ownership than conservatives. Moderates fall between conservatives and liberals. The percentage point differences are almost identical for males and females. Here we subtracted the percent for liberals from the percent for conservatives.

In this example, our dependent variable (*guns2rec*) was a recode of *guns2* and is a dichotomy. Any variable can be made into a dichotomy by recoding it. However, the disadvantage of doing this is that it often results in the loss of information. In this case, we really haven’t lost much information and the resulting table is much easier to read and interpret.

## Part 5 – Causality and Spuriousness

One of the reasons that we control for other variables is to determine if the relationship between two variables might be due to some other variable. Here’s a simple illustration. Children take standardized achievement tests every year in school. Did you know that there is a relationship between foot size and test scores? Kids with big feet have higher test scores and kids with small feet have lower tests scores. But no one thinks that foot size causes or influences test scores. No parent says I hope my children have big feet so they will be smarter and do better in school.

But that raises the question – why is there a relationship between foot size and test scores. If it’s not a causal relationship, then it must be due to some other variable or combination of variables. In that case we would say it is a spurious relationship. It’s easy to think of a possible third variable in this example – grade in school. Kids in higher grades typically have bigger feet and higher test scores. How would we test this explanation? We would control for grade in school. In other words, we would hold grade constant. We would do this by sorting our sample into groups by grade – all those in grade 6 and all those in grade 7 and so on up to all those in grade 12 which would be a total of seven groups. If the relationship is due to grade in school, then the relationship between foot size and test scores would go away (or get much weaker) in each of these groups. If it did, then we would know that the relationship between foot size and test scores was due to this third variable and we would say that it is a spurious relationship.

How could we extend this example to our earlier analysis of gender, political ideology, and whether respondents favored protecting gun ownership or controlling gun ownership? Let’s take another look at Figure 11-5.

Percent that Favor Controlling Gun Ownership by  
 Gender and Political Ideology

|  |  |  |  |
| --- | --- | --- | --- |
|  | Male | Female | Parentage Point Diff |
| Conservative | 22.9 | 36.0 | -13.1 |
| Moderate | 39.9 | 58.4 | -18.5 |
| Liberal | 62.9 | 77.1 | -14.2 |
| Percentage Point Diff | -40.0 | -41.1 |  |

Figure 11-5

Notice that when we hold gender constant the relationship between political ideology and how respondents feel about gun ownership really doesn’t change. For males, the percentage point difference between conservatives and liberals is 40.0 percentage points. For females, it is 41.1 percentage points. That tells us that the relationship between political ideology and opinions on gun ownership is not spurious due to gender. It might be spurious due to some other third variable, but it’s not spurious due to gender. We usually call this **replication** because the two-variable relationship is replicated or repeated in each of the partial tables (i.e., males and females). Non-spuriousness means that the relationship between two variables is not due to that particular variable.

## Part 6 – Criteria for Establishing Causality

What are the criteria that we must meet to establish a causal relationship?

* We must show that there is a statistical relationship between our independent and dependent variables.
* We must show that one of the variables is the possible causal variable and the other is the possible effect. This is often referred to as causal ordering.
* We must show that the two-variable relationship is not due to **any** other variable or combination of variables. In other words, we must show that it is not spurious.

Clearly, we can’t show that a two-variable relationship is not due to **all** other variables. We can’t even know what all the other variables might be. This means we can never conclusively prove causality using survey analysis. We can only present enough evidence to make it likely that it is a causal relationship. That’s what has been done with smoking and lung cancer. We haven’t proven conclusively that smoking causes lung cancer but we have ruled out so many possible third variables that now it’s highly credible (or believable) that it is a causal relationship.

So, in the example we looked at in Part 4 we would say that there is a statistical relationship between political ideology and how people feel about gun ownership. We could argue that ideology might influence how people feel about gun ownership but that it is unlikely that how they feel about gun ownership would influence their ideology. And we would say that it’s not a spurious relationship when we hold gender constant. But we could not say that we have proven that it is a causal relationship.

## Part 7 – Exploring the Three-Variable Relationship for Gender, Political Party Combined with Ideology, and How Respondents Feel about Gun Ownership

Now it’s your turn. We’re going to rerun the three variable table but this time we’re going to use political party combined with ideology (*pol4*) as our political variable. Redo the analysis from Part 4 using *pol4* instead of *pol3rec*. This means that the command you should paste into the syntax window will be the following.

## CROSSTABS

## /TABLES = GUNS2REC BY POL4 BY DEMO\_GENDER

## /STATISTICS = CHISQ D

/CELLS = COUNT COLUMN.

Fill in the percents for the following table.

Percent that Favor Controlling Gun Ownership   
by Gender and Political Party Combined with Ideology

|  |  |  |  |
| --- | --- | --- | --- |
|  | Males | Females | Percentage Point Diff |
| Conservative Republicans |  |  |  |
| Moderate-Liberal Republicans |  |  |  |
| Independents |  |  |  |
| Conservative-Moderate Democrats |  |  |  |
| Liberal Democrats |  |  |  |
| Percentage Point Diff |  |  |  |

Figure 11-6

Now answer the following questions.

* Describe the relationship between political party/ideology and how respondents felt about gun ownership when you held gender constant (i.e., controlled for gender). Did controlling for gender change the relationship?
* What do the Chi Square and Somers’ D values tell you?
* Was the relationship between political party/ideology and how respondents felt about gun ownership a spurious relationship when you held gender constant (i.e., controlled for gender)? How do you know?

## What’s Next?

In the next exercise we’re going to explore the three-variable relationships for the presence of guns in the household, gender, and how respondents feel about controlling and protecting gun ownership.

##### Data Analysis Exercises on Gun IssuesEdward Nelson, California State University, Fresno

# Exercise 12 – Three Variable Tables – Exploring the Relationships of Own Guns, Gender, and Views on Gun Ownership

## Part 1 – Research Questions

In Exercises 5 through 10 we explored relationships between pairs of variables which is often referred to as bivariate analysis. Exercises 11 through 13 focus on sets of three variables which is often referred to as multivariate analysis. Multivariate analysis isn’t limited to just three variables but that’s what we’re going to focus on in these exercises.

The research study we’re using in these exercises is an August 2016 political survey conducted by the Pew Research Center.[[53]](#footnote-53) The name of the data set is “pew\_2016\_political\_survey\_subset.” You can access it from the Social Science Research and Instructional Council’s website by clicking on [this link](https://www.ssric.org/files/2019-12/pew_2016_political_survey_subset.sav). If you are using this exercise as part of a class, your instructor will show you how to access the data.

In this exercise we’re going to look at the presence of guns in the household, gender, and how respondents feel about controlling and protecting gun ownership. Our measure of how respondents feel about gun ownership will be a question from the Pew survey that asks respondents what they think is more important – “to protect the right of Americans to own guns, or to control gun ownership?” This variable is named *guns2*. We’re going to use a recode of this variable which is named *guns2rec*. *Guns2rec* is a recode of *guns2* and dichotomizes this variable into two categories – supports protecting gun ownership and supports controlling gun ownership.

Our dependent variable is what we’re trying to explain. Since we trying to explain why some respondents want to protect gun ownership and others want to control it, our dependent variable is *guns2rec*.

Independent variables are the variables you think might help explain your dependent variable. There are many possible variables that might explain why some respondents want to protect gun ownership and others want to control it. We discussed some of them in Exercise 3. In this exercise we’re going to focus on the presence of guns in the household and gender as possible independent variables.

Often, we identify some variable that we think might be affecting the relationship between our dependent and independent variables and call this the control variable. There are several reasons we might want to introduce a control variable into the analysis. We’ll talk about control variables in this exercise.

## Part 2 – Measuring Gender

The Pew survey we’re using was a telephone survey of adults in the U.S. At the conclusion of the survey interviewers recorded the respondent’s sex. Recent surveys are beginning to ask respondents whether they identify as male, female, or non-binary. Some surveys ask more detailed questions about transgender individuals. No single way of asking this question has emerged as of yet. In the 2016 Pew survey we’ll be using the variable *demo\_gender* which is based on the interviewer’s perception of the respondent’s sex.

## Part 3 – Measuring Whether There Are Guns Present in Household

The presence of guns in the household is another important variable to consider. The Pew survey asked the following question in their survey: “Do you, or does anyone in your household, own a gun, rifle or pistol? [IF YES: Is that you or someone else in your household?]**”** This was coded into the following categories:

* Yes, respondent
* Yes, someone else
* Yes, both/multiple(VOL.)
* No, nobody in household owns a gun
* Don't know/Refused (VOL.)

The name of this variable in our data set is *guns1*. We recoded it into the following categories:

* Yes, respondent
* Yes, someone else, but not respondent
* No, nobody in household owns a gun
* Don’t know/refused (VOL.)

This variable was named *guns1rec*. We did this by combining the first and third categories of the original variable.

## Part 4 – Exploring the Two-Variable Relationships for Gender, the Presence of Guns in Household, and How Respondents Feel about Gun Ownership

Let’s start by looking at the two-variable relationship between gender and whether there are guns present in household, and how respondents feel about gun ownership. As in the previous exercises, we’re going to use a statistical procedure called crosstabulation. We’re also going to use several statistics to help us – Chi Square and Somers’ D which is a measure of association.

How is *demo\_gender* related to both the presence of guns in household and how people feel about gun ownership? Run the crosstab for *demo\_gender* and *guns1rec* and *guns2rec*. “Rec” means that these are recoded variables. Be sure to put your independent variable in the columns and get the column percents, Chi Square, and Somers’ D. If you need a review of two-variable tables, refer to Exercise 5.  
  
Write a paragraph describing the relationship between gender and each of the other two variables. Be sure to use the statistics you told PSPP to compute in your paragraph.

Now run the crosstab for *guns1rec* and *guns2rec*. Be sure to put your independent variable in the columns and get the column percents, Chi Square, and Somers’ D.

Write a paragraph describing the relationship between *guns1rec* and *guns2rec*. Be sure to use the statistics you told PSPP to compute in your paragraph.

## Part 5 – Exploring the Three-Variable Relationships for Gender, Presence of Guns in Household, and How Respondents Feel about Gun Ownership

## Now let’s turn our attention to three-variable relationships. Unfortunately, PSPP doesn’t let you easily run three-variable tables from the graphical interface. But you can run them from the syntax window. Click on FILE in the menu bar at the top of your screen and then on NEW and finally on SYNTAX. We’re going to enter the following crosstabs command in the syntax window. Here’s the command. You can copy and paste the command into the syntax window or you can type it in yourself.

## CROSSTABS

## /TABLES = GUNS2REC BY GUNS1REC BY DEMO\_GENDER

## /STATISTICS = CHISQ D

/CELLS = COUNT COLUMN.

Here’s a brief explanation for each of the parts of this command.

* CROSSTABS is the name of the command.
* TABLES is the subcommand tells PSPP that GUNS2REC is the dependent variable, GUNS1REC is the independent variable, and DEMO\_GENDER is the control variable.
* STATISTICS is the subcommand that tells PSPP which statistics you want to run (Chi Square and Somers’ D).
* CELLS tells PSPP which statistics you want in each of the cells of the table (frequencies or counts and the column percents).
* The forward slash (/) is the separator that separates the different parts of the command.
* The period (.) at the end of the entire command tells PSPP that this ends the command.
* It doesn’t matter whether you use capitals or lower-case letters.
* To run the command, click on RUN in the menu bar in the syntax window. Then click on ALL in the drop-down menu which means to run all the commands.

Now click on WINDOW in the menu bar and then check the box for the output window. Your screen should look like Figure 12-1.

## Crosstab Output

Figure 12-1

Notice that you have two tables – one for males and the other for females. Often these are referred to as partial tables since each of the tables contains part, but not all, of the cases in your sample. In each cell of the table, there are two numbers – the count or number of cases in each cell and the column percents. Notice that the percents sum down to 100 telling you that these are column percents.

Since our dependent variable is a dichotomy, we can rearrange the three-variable table into another table that is simpler and easier to read. Fill in the percents for Figure 12-2 below.

Percent that Favor Controlling Gun Ownership   
by Gender and Presence of Guns in Household

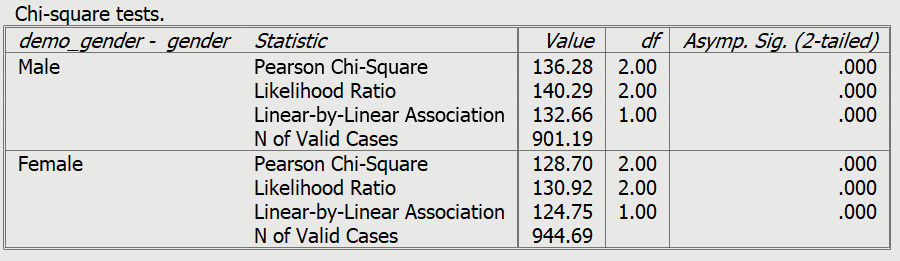
|  |  |  |  |
| --- | --- | --- | --- |
|  | Male | Female | Percentage Point Diff |
| Respondent owns gun | 19.9 | 29.1 | -9.2 |
| Some else in household owns guns | 26.7 | 42.2 | -15.5 |
| No guns in household | 58.5 | 71.6 | -13.1 |
| Percentage Point Diff | -38.6 | -42.5 |  |

Figure 12-2

Notice that we have entered the percent that favored controlling gun ownership. We could have entered the percent that favored protecting gun ownership. The numbers would have been different but the pattern wouldn’t have changed.

We computed the percentage point difference by subtracting the percent for females from the percent for males. We could have subtracted the percent for males from the percent for females. That would have changed the sign from negative to positive but it wouldn’t have changed the numerical values.

What do the Chi Square tests and the measure of association tell you? Your screen should look like Figure 12-3.



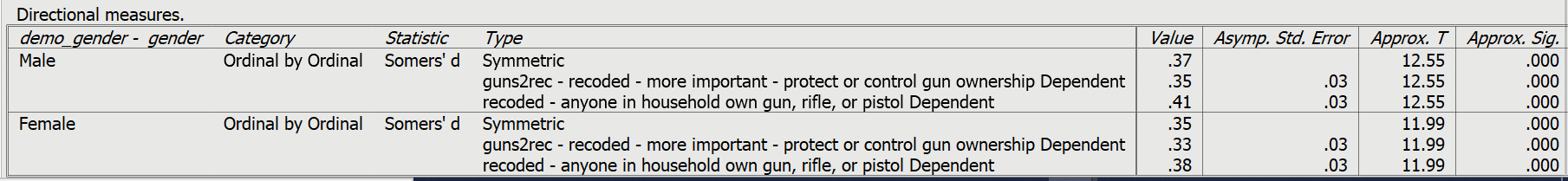


Figure 12-3

Using the statistics in Figure 12-1, 12-2, and 12-3, answer the following questions.

* What is the relationship between presence of guns in household and how respondents feel about protecting or controlling gun ownership **for males**?
* What is the relationship between presence of guns in household and how respondents feel about protecting or controlling gun ownership **for females**?
* Is the relationship similar or different for males and for females?
* Is the relationship similar or different to the two-variable table that you ran in Part 3?

When you compare the two partial tables (i.e., one for males and another for females) with the two-variable table from Part 3 there are three possibilities.

* The partial tables are similar to the two-variable table. This is referred to as **replication** since the partial tables replicate the original two-variable table. This means that the control variable (i.e., gender) is not affecting the relationship between presence of guns in household and how respondents feel about protecting or controlling gun ownership.
* The relationship between presence of guns in household and how respondents feel about protecting or controlling gun ownership disappears or is considerably weakened in **all** the partial tables. Further, this assumes that the control variable is **antecedent**. Antecedent means that the control variable is causally prior to both the independent and dependent variables. This is referred to as **explanation** because the control variable has explained away the relationship between the independent and dependent variables. Often, we say that the two-variable relationship is **spurious**. In other words, there is a statistical relationship but it’s not a causal relationship.
* The relationship between presence of guns in household and how respondents feel about protecting or controlling gun ownership varies in the partial tables. This means that the relationship is stronger for some of the partial tables and weaker for other partial tables. This is referred to as **specification** because the control variable specifies the conditions under which the two-variable relationship is stronger and weaker.

Let’s talk about causal ordering. Our independent variable is presence of guns in household (*guns1rec*) and our dependent variable is how respondents feel about protecting or controlling gun ownership (*guns2rec*). Our control variable is *demo\_gender*. Gender is an antecedent variable since it is a possible cause of both the independent and dependent variables. Here’s how we would diagram that.

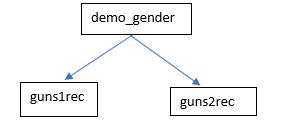


Figure 12-4

In other words, *demo\_gender* could possibly cause or influence the other two variables but those variables could not cause or influence *demo\_gender.*

How would you decide if the relationship in each of the partial tables is similar to the original two-variable table or whether it is different? How would you tell if the relationship varies in the partial tables? Here are three things to look at.

* For each partial table compute the percent difference and then compare these differences for the two partial tables.
* Look to see if Chi Square is statistically significant in the two partial tables.
* Compare the Somers’ D values for the two partial tables.

Now write a paragraph discussing which of the three outcomes – replication, explanation, or specification – best describes your three-variable table. Be sure to use the percent differences, Chi Square, and Somers’ D in your paragraph. Don’t make too much of small differences.

## Part 6 – What if the Control Variable Isn’t an Antecedent Variable?

Let’s think about another example. In Part 4 we saw that gender was related to how respondents felt

about protecting or controlling gun ownership. Females were more likely to want to control gun

ownership while males were more likely to want to protect gun ownership. But why? Here’s one

possible reason. Females might be more likely to think that gun ownership puts people’s safety at risk

and those who think that guns put people’s safety at risk are more likely to want to control gun

ownership. Pew asked a question about gun safety – “Do you think that gun ownership in this country

does more to protect people from becoming victims of crime or puts people’s safety at risk?” This

variable is named *guns3*. So perhaps the possible causal relationship looks this.

Digram of hypothesized impact of gender on guns3, and of guns3 on guns2rec

Figure 12-5

In this example *guns3* is an intervening variable. Demo\_gender is a possible cause of how respondents feel about gun safety and gun safety is a possible cause of how respondents feel about protecting or controlling gun ownership.

How would we determine whether *guns3* helps answer the question – why are females more likely to want to control gun ownership? Does *guns3* help interpret the relationship between *demo\_gender* and *guns2rec*? What we can do is run the three-variable table with *demo\_gender* as the independent variable, *guns2rec* as the dependent variable, and *guns3* as the control variable. If our argument is correct, then this relationship ought to be weaker in all the partial tables.

Run the three-variable table described above by entering the following command into a syntax window.

## CROSSTABS

## /TABLES = GUNS2REC BY DEMO\_GENDER BY GUNS3

## /STATISTICS = CHISQ D

/CELLS = COUNT COLUMN.

In Part 4 we said there were three possible outcomes when we controlled for a third variable. We called these:

* replication,
* explanation, and
* specification.

Now we can add a fourth possible outcome which is typically referred to as **interpretation**. The difference between explanation and interpretation is that in explanation the control variable is an antecedent variable and in interpretation the control variable in an intervening variable. But in both explanation and interpretation the relationship between the independent and dependent variables either goes away or is considerably weaker in **all** the partial tables.

Answer the following questions.

* What happened to the relationship between *demo\_gender* and *guns2rec* when you held *guns3* constant (i.e., controlled for guns3)?
* Did this turn out to be an example of interpretation? How did you decide?
* Which of the four outcomes described above is this an example of? How do you know? What does this mean?

## Part 7 – What if the Causal Order is Unclear?

Sometimes the causal ordering of the variables is unclear. This could be for two reasons.

* There is no causal relationship between the two variables. An example of this would be the relationship between gender and race.
* We have no way of knowing the causal ordering. For example, think of two opinion variables such as protecting or controlling gun ownership and how respondents feel about immigration.

In these cases, we could still determine which variable has the stronger relationship with the dependent variable controlling for the other variable but we can’t say whether it is explanation or interpretation.

## What’s Next?

In the next exercise we’ll explore the three-variable relationships for gender, how often respondents follow what’s going on in government and public affairs, and how they feel about gun ownership.

##### Data Analysis Exercises on Gun IssuesEdward Nelson, California State University, Fresno

# Exercise 13 – Three Variable Tables – Exploring the Relationships of Gender, How Often Respondents Follow What’s Going on in Government and Public Affairs, and Views on Gun Ownership

## Part 1 – Research Questions

In Exercises 5 through 10 we explored relationships between pairs of variables which is often referred to as bivariate analysis. Exercises 11 through 13 focus on sets of three variables which is often referred to as multivariate analysis. Multivariate analysis isn’t limited to just three variables but that’s what we’re going to focus on in these exercises.

The research study we’re using in these exercises is an August 2016 political survey conducted by the Pew Research Center.[[54]](#footnote-54) The name of the data set is “pew\_2016\_political\_survey\_subset.” You can access it from the Social Science Research and Instructional Council’s website by clicking on [this link](https://www.ssric.org/files/2019-12/pew_2016_political_survey_subset.sav). If you are using this exercise as part of a class, your instructor will show you how to access the data.

In this exercise we’re going to look at gender, how often respondents follow what’s going on in government and public affairs, and how they feel about gun ownership. Our measure of how respondents feel about gun ownership will be a question from the Pew survey that asks respondents what they think is more important – “to protect the right of Americans to own guns, or to control gun ownership?” This variable is named *guns2*. We’re going to use a recode of this variable which is named *guns2rec*. *Guns2rec* is a recode of *guns2* and dichotomizes this variable into two categories – supports protecting gun ownership and supports controlling gun ownership.

Our dependent variable is what we’re trying to explain. Since we trying to explain why some respondents want to protect gun ownership and others want to control it, our dependent variable is *guns2rec*.

Independent variables are the variables you think might help explain your dependent variable. There are many possible variables that might explain why some respondents want to protect gun ownership and others want to control it. We discussed some of them in Exercise 3. In this exercise we’re going to focus on gender and how often respondents follow what’s going on in government and public affairs as possible independent variables.

Often, we identify some variable that we think might be affecting the relationship between our dependent and independent variables and call this the control variable. There are several reasons we might want to introduce a control variable into the analysis. We’ll talk about control variables in this exercise.

## Part 2 – Measuring Gender

The Pew survey we’re using was a telephone survey of adults in the U.S. At the conclusion of the survey interviewers recorded the respondent’s sex. Recent surveys are beginning to ask respondents whether they identify as male, female, or non-binary. Some surveys ask more detailed questions about transgender individuals. No single way of asking this question has emerged as of yet. In the 2016 Pew survey we’ll be using the variable *demo\_gender* which is based on the interviewer’s perception of the respondent’s sex.

## Part 3 – Measuring How Often Respondents Follow What’s Going On

## in Government and Public Affairs

The Pew survey asked the following question to measure how often respondents followed governmental and public affairs – “Would you say you follow what’s going on in government and public affairs: most of the time, some of the time, only now and then, or hardly at all?” The name of this variable in the data set is *pol7*. We’re using a recoded version that dichotomizes the response categories into most of the time or less than most of the time. The recoded version is *pol7rec*. For convenience we’ll refer to this variable as “how often follows public affairs” in the rest of this exercise.

## Part 4 – Exploring the Two-Variable Relationships Between Gender and How Often Respondents Follow Public Affairs and How Respondents Feel about Gun Ownership

Let’s start by looking at the two-variable relationships between gender and how often respondents follow public affairs and how respondents feel about gun ownership. As in the previous exercises, we’re going to use a statistical procedure called crosstabulation. We’re also going to use several statistics to help us – Chi Square and Somers’ D which is a measure of association.

How is *demo\_gender* related to how often respondents follow public affairs and how people feel about gun ownership? Run the crosstab for gender and each of these variables. Be sure to put *demo\_gender* in the columns and get the column percents, Chi Square, and Somers’ D – and be sure to use *pol7rec*. If you need a review of two-variable tables, refer to Exercise 5.  
  
Write a paragraph describing the relationship between gender and each of these variables. Be sure to use the statistics you told PSPP to compute in your paragraph.

## Now run the crosstab to find out how *pol7rec* is related to how people feel about gun ownership. Be sure to put *demo\_gender* in the columns and get the column percents, Chi Square, and Somers’ D. Write a paragraph describing the relationship between these two variables.

## Part 5 – Exploring the Three-Variable Relationships for Gender, How Often Respondents Follow Public Affairs, and How Respondents Feel about Gun Ownership

Now let’s turn our attention to three-variable relationships. Unfortunately, PSPP doesn’t let you easily run three-variable tables from the graphical interface. But you can run them from the syntax window. Click on FILE in the menu bar at the top of your screen and then on NEW and finally on SYNTAX. We’re going to enter the crosstabs command in the syntax window. Here’s the command. You can copy and paste the command into the syntax window or you can type it in yourself.

## CROSSTABS

## /TABLES = GUNS2REC BY POL7REC BY DEMO\_GENDER

## /STATISTICS = CHISQ D

/CELLS = COUNT COLUMN.

Here’s a brief explanation for each of the parts of this command.

* CROSSTABS is the name of the command.
* TABLES is the subcommand tells PSPP that GUNS2REC is the dependent variable, POL7REC is the independent variable, and DEMO\_GENDER is the control variable.
* STATISTICS is the subcommand that tells PSPP which statistics you want to run (Chi Square and Somers’ D).
* CELLS tells PSPP which statistics you want in each of the cells of the table (frequencies or counts and the column percents).
* The forward slash (/) is the separator that separates the different parts of the command.
* The period (.) at the end of the entire command tells PSPP that this ends the command.
* It doesn’t matter whether you use capitals or lower-case letters.
* To run the command, click on RUN in the menu bar in the syntax window. Then click on ALL which means to run all the commands.

## Now click on WINDOW in the menu bar and then check the box for the output window. Your screen should look like Figure 13-1.

## Crosstab output

Figure 13-1

Notice that you have two tables – one for males and the other for females. Often these are referred to as partial tables since each of the tables contains part, but not all, of the cases in your sample. In each cell of the table, there are two numbers – the count or number of cases in each cell and the column percents. Notice that the percents sum down to 100 telling you that these are column percents.

While the percents are different in the partial tables for males and females, the pattern in the same. Those who are more likely to follow public affairs are more likely to want to protect the right to own guns. But notice that the percent differences show that relationship is stronger for males than for females. Let’s look at what the Chi Square tests and the measure of association tell us. Your screen should look like Figure 13-2.

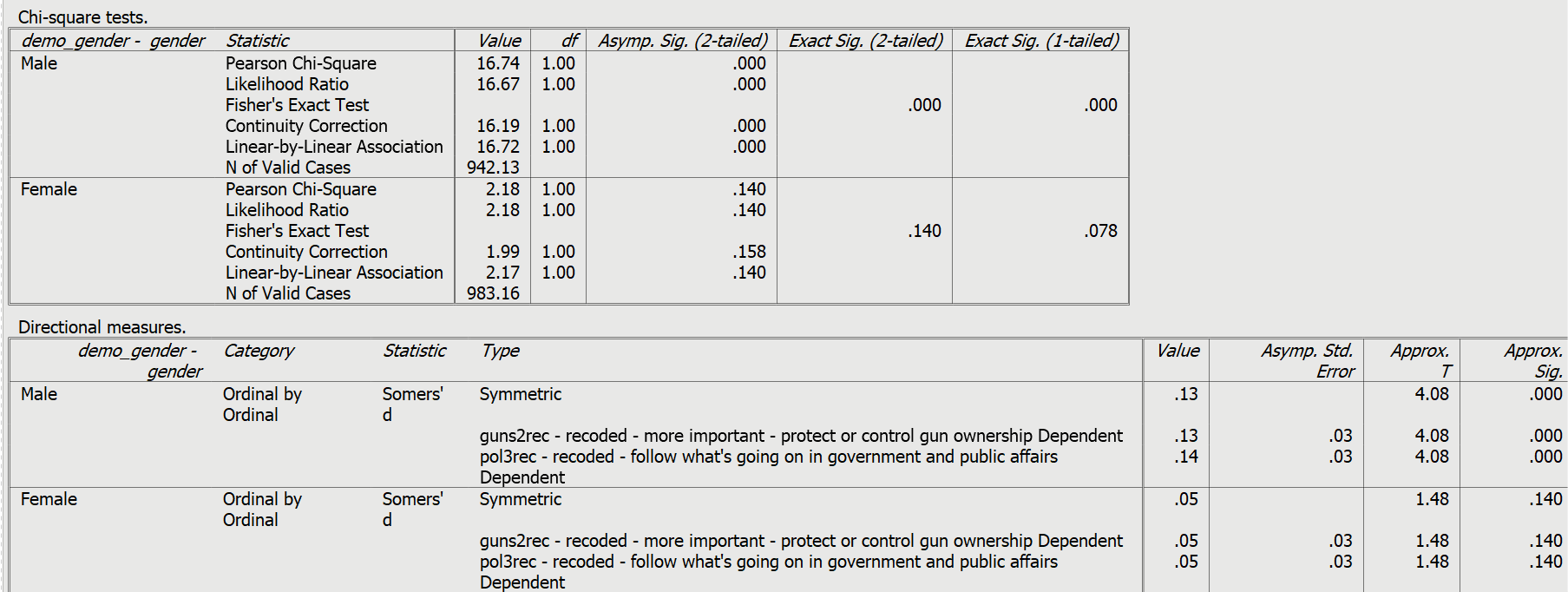


Figure 13-2

Notice that Chi Square is significant for males but not for females and that the Somers’ D values are larger for males. In other words, the relationship between how often respondents follow public affairs and views on gun ownership has virtually disappeared for females but not for males.

Since our dependent variable is a dichotomy (i.e., has only two values) we can rearrange the values into a table that is easier to read. Here’s what that would look like.

Percent that Favor Controlling Gun Ownership  
 by Gender and How Often Follow Public Affairs

|  |  |  |  |
| --- | --- | --- | --- |
|  | Male | Female | Parentage Point Diff |
| Follow most of time | 32.2 | 53.2 | -21.0 |
| Follow less | 45.3 | 57.9 | -12.6 |
| Percentage Point Diff | -13.1 | -4.7 |  |

Figure 13-3

Notice that we entered the percent that favored controlling gun ownership. We could have entered the percent that favored protecting gun ownership. The numbers would have been different but the pattern wouldn’t have changed.

We computed the percentage point difference by subtracting the percent for females from the percent for males. We could have subtracted the percent for males from the percent for females. That would have changed the sign from negative to positive but it wouldn’t have changed the numerical values.

What does this table tell us?

* Females are more likely to want to control gun ownership than males **both** for those who follow public affairs most of the time **and** for those who follow it less. But the percentage difference is much larger for those that follow most of the time.
* Those who follow public affairs less are more likely to want to control gun ownership than those who follow it more and this is true for both males and females. But the percentage differences are greater for males than for females.

In this example, our dependent variable (*guns2rec*) was a recode of *guns2* and is a dichotomy. Most variables can be made into a dichotomy by recoding. However, the disadvantage of doing this is that it often results in the loss of information. In this case, we really haven’t loss much information and the resulting table is much easier to read and interpret.

## Part 6 – What Do We Make of All This?

So, what’s going on here? The relationship between how much people follow public affairs and how they feel about controlling and protecting gun ownership depends on gender. There’s a stronger relationship for males than for females. And when we look at Figure 13-3 three of the percentages are fairly close to each other – 57.9%, 53.2%, 45.3% -- and one of the percents – 32.2% -- is considerably smaller. Those that are least likely to want to control gun ownership are males who follow public affairs more closely. That’s an example of what we call **specification**. The interaction of our two independent variables – gender and how closely respondents follow public affairs – specified the conditions under which respondents want to either protect or control gun ownership.

Exposure to information about public affairs means that people will also be exposed to more information about mass shootings and might, therefore, be more fearful of becoming a victim themselves. This might lead them to want to arm themselves with guns. However, notice that there are problems with this argument. Exposure to information about mass shootings might also lead them to want to control gun ownership. Both arguments seem equally likely.

## Part 7 – Exploring the Three-Variable Relationship Between Gender, How Much Respondents Follow Public Affairs, and Creating a Federal Data Base to Track Gun Sales

Now it’s your turn. We’re going to rerun the three variable table but this time we’re going to use a

different dependent variable – *guns7*. Here’s the question from the Pew survey – “Please tell me if you

would favor or oppose the following proposals about gun policy. What aboutcreating a federal

government database to track all gun sales?”

You want the three-variable table with *guns7* as your dependent variable, *pol7rec* as your independent variable, and *demo\_gender* as your control variable. Be sure to ask for the column percents, Chi Square, and Somers’ D. Here’s the command. You can copy and paste the command into the syntax window or you can type it in yourself.

## CROSSTABS

## /TABLES = GUNS7 BY POL7REC BY DEMO\_GENDER

## /STATISTICS = CHISQ D

/CELLS = COUNT COLUMN.

Fill in the percents for the following table.

Percent that Favor Creating a Federal Data Base to Track Gun Sales   
by Gender and How Often Follow Public Affairs

|  |  |  |  |
| --- | --- | --- | --- |
|  | Males | Females | Percentage Point Diff |
| Follow most of time |  |  |  |
| Follow less of time |  |  |  |
| Percentage Point Diff |  |  |  |

Figure 13-4

Now answer the following questions.

* Describe the relationship between how often respondents follow public affairs and how respondents feel about gun ownership when you hold gender constant (i.e., control for gender). Did controlling for gender change the relationship?
* Is this an example of specification? How did you decide?
* What do you think these findings mean? In other words, how would you explain the findings?

## What’s Next?

In the final exercise in this series we’ll discuss how to write research reports.

##### Data Analysis Exercises on Gun IssuesEdward Nelson, California State University, Fresno

# Exercise 14 -- Writing Research Reports

This chapter will focus on how to write research reports including:

* how to organize your report,
* creating tables,
* whether to use footnotes or endnotes,
* citing articles, papers and other materials,
* plagiarism,
* proofreading, and
* other guides to writing reports.

## An Outline of your Research Report

In the previous chapters we discussed how to use PSPP to analyze your data. We talked about using PSPP to describe your data, analyzing the relationship between pairs of variables, and extending our analysis to include sets of three or more variables. Now we need to think about how to write a research report so that others may read it and learn from our analysis.  This report might be for a class you are taking, or it might be a report that you are submitting to a research conference.  If you are going to submit your report to a journal for possible publication, you need to look carefully at the instructions that all journals provide on preparing a manuscript for publication.

Here's an outline for your report.  Don't think that this is the only way you can organize your report, but this is one way to do it.

* Title page including your name, date, and class or institutional affiliation.
* Abstract – An abstract is a short summary of what you did in the paper and the major findings of your analysis.  Abstracts are really short, so you need to be succinct. It should be less than 200 words or even shorter depending on the requirements of your professor or the research conference to which you are submitting your paper.
* Table of contents (optional).
* Body of the paper.
  + An introduction to the paper which explains why you wrote the report and provides an introduction to the topic of the paper.
  + Your review of the literature that summarizes what others discovered about this topic.  Virtually everything you might do has been written about by others.  You should review the relevant literature and summarize what others have found.  You don't want to simply list the relevant literature and consider the articles and books one by one.  Rather you want to summarize what others have done and look for themes around which you can organize your literature review.  If you are having trouble finding relevant literature, go to the library at your university or a nearby university and talk with a reference librarian.  They are trained in searching for relevant literature and will be able to help you.
  + The methodology of your study.
    - If you collected your own data, discuss how you chose your sample, how you measured the concepts, and how you collected your data.
    - If you used an existing data set, discuss the sampling, measurement, and data collection used in that study.  Studies that are part of data archives such as the Inter-university for Political and Social Research at the University of Michigan and the Roper Center for Public Opinion Research at Cornell University provide good summaries for all data sets that are housed at their archive.
  + Theory and Hypotheses – If you are using a theoretical perspective and/or testing hypotheses, describe the theory and state the hypotheses you plan to test.  Be sure to cite supporting literature that form the basis for your theory and hypotheses.
  + Empirical findings and interpretation – What are the empirical findings that came out of your data analysis and what did they tell you?  If you are testing hypotheses, did your analysis support your hypotheses?  Remember that you are telling a story.  Start simple and build up.  That means starting with looking at variables one at a time (i.e., univariate analysis), then proceeding to relationships between pairs of variables (i.e., bivariate analysis), and then looking at sets of three or more variables (i.e., multivariate analysis) to consider such things as spuriousness.
  + Conclusions and summary. This is a little like your abstract but not as short.  What did you do, what did you find in your study and what does it mean?
* Tables.  You may choose to put your tables in the body of your paper, or you may decide to put them all at the end of your paper.
* References.  For every article or book that you cite, you need to provide a full bibliographic reference at the end of the report.

## **Tables**

There are advantages and disadvantages to putting your tables in the body of the report or at the end of the report.  Putting them in the body of the report keeps them front and center for the reader but they often are bulky and get in the way of reading the report.  Putting them at the end of the report gets them out of the way and allows the reader to spread them out and look at them as he or she is reading the paper.  Your instructor or the research conference will usually tell you where to put your tables.

If they are placed at the end of the paper, put a note in the body of the report that says something like "Table 1 about here."  That will let the reader know where the table fits into your report.

Constructing a good table is important.  Sometimes your instructor will tell you to copy tables from the program you are using for statistical analysis (e.g., PSPP, SPSS, and others) into your paper.  Other times you will construct the tables yourself.  A good reference on creating tables is The Chicago Guide to Writing About Numbers by Jane E. Miller.[[55]](#footnote-55)  Your word processing program (e.g., Word in Microsoft Office) will provide you with templates that you can choose for your tables.

### Footnotes or Endnotes

Often you want the reader to be aware of something, but you don't want to put it in the body of the paper.  It may be a technical issue such as how you recoded a variable or why you chose a particular statistic.  Or you may want to tell the reader that you will discuss something later in the paper.  You can put comments like these in either a footnote or an endnote.  A footnote goes at the bottom of the page and an endnote goes at the end of the paper.  Your word processing program will allow you to enter either footnotes or endnotes in your paper.  Which you use is up to you unless your instructor or the research conference tells you that one or the other is required.

### Citing Articles, Papers, and Other Materials

There are many styles such as American Psychological Association (APA) or Modern Language Association (MLA) that you could use to cite materials that you refer to in your paper.  Remember that anytime you refer to someone else's work, you must acknowledge the source.  Your instructor or research conference will often specify which style you should use.

### Plagiarism

Plagiarism is using someone else's words or ideas without acknowledging the source.  If you are quoting from a document, you must cite the source.  Even if you are paraphrasing, you must acknowledge the source.  If you are using someone else's ideas, you must also acknowledge the source.  There is a good review of plagiarism written by Earl Babbie that can be found on the Internet by clicking [here](http://www1.chapman.edu/~babbie/plag00.html).  Click on the red arrows at the top to go forward or backward in this review of plagiarism.

Proofreading

Be sure to proofread your paper several times before submitting it. Use the spell and grammar checker in your word processing program. You could also ask a friend to read it and tell you about any errors or parts that are confusing.

### **Other Guides to Writing Reports**

There are many other guides to writing research reports.  One that is commonly used in Sociology is the *Guide to Writing Sociology Papers*.[[56]](#footnote-56)  You can find others on the internet by entering "writing research reports" in the search box.

##### Data Analysis Exercises on Gun IssuesEdward Nelson, California State University, Fresno

# Appendix A – Notes to the Instructor

The goal of these exercises is to introduce students to quantitative data analysis using a Pew 2016 political survey that includes a series of questions on gun issues and many other topics. The statistical tools used are the following:

* frequency distributions,
* measures of central tendency,
* measures of dispersion or variability,
* skewness,
* two- and three-variable crosstabulation,
* Chi Square, and
* measures of association (primarily Cramer’s V and Somers’ D).

## Overview of the Exercises

The first three exercises focus on the scientific approach, research design, hypotheses, and causality. They also include a brief introduction to the variables in the data set.

The remaining exercises provide an introduction to data analysis focusing on opinions about gun issues. Describing the data or univariate analysis is covered in the fourth exercise. Two-variable or bivariate analysis is covered in exercises five through ten, and three-variable or multivariate analysis in exercises eleven through thirteen. Exercise fourteen is a brief introduction to writing research reports. The dependent variable in the exercises is respondent’s views on gun ownership – whether they think it’s more important to protect the rights to own guns or whether it’s more important to control gun ownership. Here’s a breakdown of the fourteen exercises.

* Exercise 1 – research design (i.e., sampling, measurement, data collection, data analysis) and a brief introduction to the data set
* Exercise 2 – formulating research questions, developing hypotheses, and independent and dependent variables along with a more detailed introduction to the data set
* Exercise 3 – scientific approach, criteria for establishing cause and effect, independent and dependent variables, and causal models along with a more detailed introduction to the data set
* Exercise 4 – describing the data (frequency distributions, levels of measurement, measures of central tendency, measures of dispersion or variability, skewness, charts)
* Exercise 5 – two-variable crosstabulations – geography and views on gun ownership
* Exercise 6 – two-variable crosstabulations – views on gun safety and gun ownership
* Exercise 7 – two-variable crosstabulations – politics and views on gun ownership
* Exercise 8 – two-variable crosstabulations – socioeconomic status and views on gun ownership
* Exercise 9 – two-variable crosstabulations – age, cohorts (or generations), and views on gun ownership
* Exercise 10 – two-variable crosstabulations – gender and views on gun ownership
* Exercise 11 – three-variable crosstabulations – politics, gender, and views on gun ownership
* Exercise 12 – three-variable crosstabulations – guns in household, gender, and views on gun ownership
* Exercise 13 – three-variable crosstabulations – how often follow what’s going on in government and public affairs, gender, and views on gun ownership
* Exercise 14 – writing research reports

## Data Set

The data set we’ll be using is a subset of a 2016 Pew political survey conducted by the Pew Research Center which is one of the preeminent national research centers in the U.S. The original data set contains many variables from which I selected about 60 variables for these exercises. I renamed the variables to make them easier to use. Some variables were recoded and a few new variables were created out of existing variables. The name of the data set is “pew\_2016\_political\_survey\_subset.”

Note that the variable names start with letters to indicate the type of variable. This is followed by an underscore and letters that describe the variable. Here are a list of the first part of the names describing the type of variable.

* geog – geographic variables
* demo – background variables
* pol – political variables
* relig – religious variables
* guns – variables about guns
* prob – variables about social problems
* favor – variables describing favorability toward political parties and the Supreme Court
* gov – variables about the government
* diversity – variable about societal diversity
* immig – variables about immigration
* sat – variables about respondent’s satisfaction with way things are going today
* weight – weight variables

Here is a list of the variables in this subset along with their variable names and the names of the variables in the original Pew data set. The wording of the questions in the Pew survey are also included. The names of the variables in this subset are in italics. This variable information is included here for instructors only. In the exercises I only include the variable names (in italics) and a brief description of the variable.

* Geographic variables describing where respondents live (variables start with “geog”) (These Pew variables are based on respondent’s zip code.)
  + Region of the country (*geog\_region*) (Pew variable is region.)
  + Population density (*geog\_density*) (Pew variable is density.)
  + Rural/suburban/urban nature of the community (*geog\_usr*) (Pew variable is usr.)
  + State of residence (*geog\_state*) (Pew variable is state. Some of the states have relatively few cases.)
* Background or demographic variables describing respondents (variables start with “demo”)
  + Age (*demo\_age*, *demo\_agerec1*, *demo\_agerec2*) (Pew variable is age. Recoded variables have rec at the end of the variable name. If there is more than one recode, the number after rec indicates the number of the recode.)
  + Generation (*demo\_gen*) (Pew variable is gen. Generation is sometimes referred to as cohort.)
  + Education (*demo\_educ*, *demo\_educrec*) (Pew variable is educ. Wording of question: What is the highest level of school you have completed or the highest degree you have received?)
  + Family income (*demo\_income, demo\_incomerec*) (Pew variable is income. Question wording: Last year, that is in 2015, what was your total family income from all sources, before taxes?)
  + Gender (*demo\_gender*) (Pew variable is sex. Interviewers did not ask this question. The respondent’s gender was recorded by the interviewer.)
  + Race (*demo\_race1*, *demo\_race2*) (Pew variable is race. Question wording: Which of the following describes your race? You can select as many as apply. White, Black or African American, Asian or Asian American or some other race. *Demo\_race1* codes race into white non-Hispanic, black non-Hispanic, Hispanic, other. *Demo\_race2* breaks Hispanics into those born in the U.S. and those born outside the U.S.)
  + Hispanic, Latino, Spanish origin (*demo\_hisp*) (Pew variable is hisp. Question wording: Are you of Hispanic, Latino, or Spanish origin, such as Mexican, Puerto Rican or Cuban?)
  + Number of children in household (*demo\_child*) (Pew variable is child.)
* Political variables (variables start with “pol”)
  + Party identification (*pol1*) (Pew variable is party. Question wording: In politics today, do you consider yourself a Republican, Democrat, or independent?)
  + Party identification combined with leaners. (*pol2*) (This is a variable I created. Pew asked this follow up question: As of today do you lean more to the Republican Party or more to the Democratic Party? They named this variable partyln. This variable is combined with *pol1* to produce *pol2*. Partyln is not included in this subset.)
  + Ideology – conservative, moderate, liberal (*pol3, pol3rec*) (Pew variable is ideo. Question wording: In general, would you consider your political views as very conservative, conservative, moderate, liberal, very liberal.)
  + Political ideology combined with party (*pol4*) (Pew variable is partyideo. This a variable that Pew created that codes respondents as conservative Republican, moderate-liberal Republican, Independent, conservative-moderate Democrat, liberal Democrat.)
  + Registered to vote? (*pol5*) (Pew variable is reg. Asks whether respondent is absolutely certain they are registered, probably registered, or not registered at their current address.)
  + How often respondent votes (*pol6*) (Pew variable is oftvote. Question wording: How often would you say you vote – always, nearly always, part of the time, seldom?)
  + How often respondent follows what’s going on in government and public affairs? (*pol7, pol7rec*) (Pew variable is Q. 12. Question wording: Would you say you follow what’s going on in government and public affairs most of the time, some of the time, only now and then, or hardly at all?)
* Religious variables (variables start with “relig”)
  + Religious preference (*relig1, relig1rec*) (Pew variable is relig. Question wording: What is your present religion, if any? Are you Protestant, Roman Catholic, Mormon, Orthodox such as Greek or Russian Orthodox, Jewish, Muslim, Buddhist, Hindu, atheist, agnostic, something else, or nothing in particular? *Relig1rec* recodes *relig1* into five categories – Protestant, Roman Catholic, other Christian, Non-Christian, and unaffiliated.)
  + Considers self a born-again or evangelical Christian? (*relig2*) (Pew variable is born. Question wording: Would you describe yourself as a "born again" or evangelical Christian, or not?)
  + How often attend religious services? (*relig3, relig3rec*) (Pew variable in attend. Question wording: Aside from weddings and funerals, how often do you attend religious services... more than once a week, once a week, once or twice a month, a few times a year, seldom, or never? *Relig3rec* recodes *relig3* into three categories – once a week or more, sometimes which combines once or twice a month and a few times a year, and seldom or never.)
* Gun variables (variables start with “guns”)
  + Does anyone in family own guns? (*guns1, guns1rec*) (Pew variable is q90. Question wording: Do you, or does anyone in your household, own a gun, rifle or pistol? [IF YES: Is that you or someone else in your household?])
  + Is it more important to protect rights to own guns or to control gun ownership? (*guns2, guns2rec*) (Pew variable is q83. Question wording: What do you think is more important – to protect the right of Americans to own guns, OR to control gun ownership?)
  + Does gun ownership do more to protect or to put safety at risk? (*guns3*) (Pew variable is q85. Question wording: Do you think that gun ownership in this country does more to protect people from becoming victims of crime or does more to put people’s safety at risk?)
  + Do respondents favor or oppose?
    - Laws to prevent people with mental illness from purchasing guns (*guns4*) (Pew variable is q86a.)
    - Making private gun sales and sales at gun shows subject to background checks (*guns5*) (Pew variable is q86b)
    - A ban on assault style weapons (*guns6*) (Pew variable is q86c.)
    - Creating a federal government database to track all gun sales (*guns7*) (Pew variable is q86d.)
    - A ban on high-capacity ammunition clips that hold more than 10 bullets (*guns8*) (Pew variable is q86e.)
    - Barring gun purchases by people on the federal no-fly or watch lists (*guns9*) (Pew variable is q86f.)
* How big a problem is? (variables start with “prob”) (Pew question wording: I’m going to read from a list of things that may be problems in our country. First, how big a problem is **\_\_\_** in our country – a very big problem, a moderately big problem, a small problem, or not a problem at all?”
  + Crime (*prob1*) (Pew variable is q87a.)
  + Terrorism (*prob2*) (Pew variable is q87b.)
  + Immigration (*prob3*) (Pew variable is q87c.)
  + Availability of good-paying Jobs (*prob4*) (Pew variable is q87d.)
  + Relations between racial and ethnic groups (*prob5*) (Pew variable is q87e.)
  + Gap between rich and poor (*prob6*) (Pew variable is q87f.)
  + Condition of the environment (*prob7*) (Pew variable is q87g.)
* How favorable/unfavorable are respondents toward? (variables start with “favor”) (Pew question wording: Is your overall opinion of **\_\_\_** very favorable, mostly favorable, mostly unfavorable, or very unfavorable?)
  + Republicans (*favor1*) (Pew variable is q28a.)
  + Democrats (*favor2*) (Pew variable is q28b.)
  + Supreme Court (*favor3*) (Pew variable is q28c.)
* Forced choice questions (Pew question wording: Which comes closer to your view — even if neither is exactly right?)
  + Government regulation of business is necessary to protect the public interest or

government regulation of business usually does more harm than good. (*govreg*) (Pew

variable is q66a.)

* + The government should do more to help needy Americans, even if it means going deeper into debt or the government today can't afford to do much more to help the needy (*helpneedy*). (Pew variable is q66b.)
  + As Americans, we can always find ways to solve our problems and get what we want or the country can’t solve many of its important problems. (*solveprob*) (Pew variable is q66c.)
  + There are clear solutions to most big issues facing the country today or most big issues facing the country today don’t have clear solutions. (*solution*) (Pew variable is q66g.)
* Does diversity make U.S. a better/worse place to live or does it make no difference? (diversity) (Pew variable is q61. Question wording: On balance, do you think having an increasing number of people of many different races, ethnic groups and nationalities in the United States makes this country a better place to live, a worse place to live, or doesn’t make much difference either way?)
* Immigration (variables start with “immig”)
  + Which is respondents’ priority – border security or path to citizenship? (*immig1*) (Pew variable is q71. Question wording: What should be the priority for dealing with illegal immigration in the U.S.? better border security and stronger enforcement of our immigration laws; or creating a way for immigrants already here illegally to become citizens if they meet certain requirementsor should both be given equal priority? If respondents answered both, they were asked which comes closer to their views.)
  + When it comes to undocumented immigrants currently in the U.S which comes closer to your view — even if neither is exactly right. Undocumented immigrants mostly fill jobs that American citizens would like to have or undocumented immigrants mostly fill jobs American citizens don’t want. (immig2) (Pew variable is q76a.)
  + When it comes to undocumented immigrants currently in the U.S which comes closer to your view — even if neither is exactly right? Immigrants are as honest and hard working as Americans or less honest and hardworking (*immig3*) (Pew variable is q76b)**.**
  + When it comes to undocumented immigrants currently in the U.S which comes closer to your view — even if neither is exactly right? Undocumented immigrants living in the U.S. are more likely than American citizens to commit serious crimes or undocumented immigrants living in the U.S. are no more likely than American citizens to commit serious crimes. (*immig4*) (Pew variable is q76c.)
* Way things are going today (variables start with “sat”)
  + Satisfaction with present (*sat1*) (Pew variable is q2. Question wording: All in all, are you satisfied or dissatisfied with the way things are going in this country today?)
  + Better, worse, or same as 50 years ago (*sat2*) (Pew variable is q7. Question wording: Do you think the future of the next generation of Americans will be better, worse, or about the same as life today?)
  + Future better, worse, or about same (*sat3*) (Pew variable is q8. Question wording: Do you think the future of the next generation of Americans will be better, worse, or about the same as life today?)
* Weight variables
  + *weight* (This is the weight variable from the Pew Center. Use variable *weightadj* instead.)
  + *weightadj* (This is the adjusted weight variable that I created. Use this weight variable. See the next section on weighting.)

## Weighting the Data

Data are weighted so they better represent the population from which the sample was selected. The sample consists of 2,010 adults (18 years and older) living in the U.S. The weight variable in the Pew data set is named *weight*. When this weight variable is applied to the data it expands the sample to 6,080. I created an adjusted weight variable so the original sample size of 2,010 would be maintained. You should use this adjusted weight variable called *weightadj* in all your analysis. The data set that you can download from this website has already been weighted by *weightadj*. Note that weighting produces frequencies with decimals. Students should be told that they can ignore the decimals.

## Suggestions for Using the Exercises

Since these exercises were written so each exercise was independent of the other exercises, there is quite a bit of duplication from exercise to exercise.  If you are using several exercises, you will want to remove some of that duplication.

Exercises 5 through 10 focus on two-variable crosstabulations. Exercise 5 focuses on the relationship between different ways of describing the area where the respondent lives (i.e., geography) and views on gun ownership. One of the geography variables describes the population density (*geog\_density*) of the respondent’s location and crosstabs it with the variable *guns2* which measures whether the respondent thinks it is more important to protect the right to own guns or whether it is more important to control gun ownership. This analysis is repeated as an example in each of exercises 6 through 10. That makes for a lot of repetition so if you are using two or more of these six exercises you will want to take out the repetitive material which is used in each of these exercises.

One way to use these exercises is to use the first two exercises, one of the two-variable exercises, and one of the three-variable exercises. You can pick and choose whichever exercises you want to use in your class. Or you may want to choose just one of the exercises and use it.

Another possibility is to develop new exercises based on the data set provided. There are variables about gun issues not used in these exercises and there are many other variables besides those that focus on gun issues that could be the basis for new exercises.

## Statistical Analysis

Statistical analysis is limited to descriptive statistics, crosstabulation, Chi Square, and measures of association (primarily Cramer’s V and Somers’ D). Each of these statistics are explained in the exercises but there is no discussion of how to compute them. You may want to add material on computation.

There are many statistical packages available for you to use. This series of exercises uses PSPP which is a free alternative to SPSS. I plan to provide an alternative series that uses SPSS which should be available on this website sometime in 2020.

For those who are not familiar with PSPP, Appendix B contains notes that explain how to download PSPP and some other important points about PSPP.

## Permission to Modify

## You have permission to use these exercises and the data set and to revise them to fit your needs.  Feel free to revise them in any way you want.  Just recognize the source of the original exercise.  I would like to hear from you about your experiences using the exercises.

## Contact Information

If you would like to contact me (Edward Nelson), please email me at [**ednelson@csufresno.edu**](mailto:ednelson@csufresno.edu).  I’m Professor Emeritus at California State University, Fresno in the Sociology department.  I taught research methods, statistics, and critical thinking before retiring and now teach a critical thinking course part time.  Please feel free to contact me about any questions or problems you may have when using the exercises.

##### Data Analysis Exercises on Gun IssuesEdward Nelson, California State University, Fresno

# Appendix B – Working with PSPP

SPSS is the most widely used statistical package in the California State University System and many other colleges and universities. However, some colleges, including many community colleges, do not have a site license for SPSS. The cost of SPSS is often prohibitive, so these exercises will use PSPP. For more information on PSPP, click [here](http://www.gnu.org/software/pspp/). Their website says that “GNU PSPP is a program for statistical analysis of sampled data. It is a Free replacement for the proprietary program SPSS and appears very similar to it with a few exceptions.”

The easiest way to download PSPP is to click [here](http://pspp.awardspace.info/) and look for the “Downloads” box. Then download the latest version in either 32-bit or 64-bit format. If you’re not sure which version to download, go to the control panel and click on “System” and look for your system type. Then follow the instructions to download. You can open SPSS data files (both .sav and .por) in PSPP.

You can run PSPP from either the interface or in syntax mode. In these exercises, with one exception described below we’re going to focus on using the interface by pointing and clicking at appropriate points.

Some PSPP commands do not have the full capabilities of their corresponding SPSS commands. Other PSPP commands work somewhat differently from their SPSS counterparts. We’ll encounter this when we want to run three-variable tables in Exercises 11 through 13. To solve this problem, we’re going to run the commands for three-variable tables in what is called syntax mode. We’ll explain how to do that in Exercises 11 through 13.

PSPP will list the variables in your data file and you can select those variables you want to use. It’s easier to find the variables if they are listed by variable names. You can change the way PSPP lists the variables by right clicking anywhere on the list of variables and checking or unchecking the box for PREFER VARIABLE LABELS.

You can sort the variables alphabetically by name by clicking on EDIT in PSPP and then on OPTIONS. Click on DISPLAY NAMES and on SORT BY NAME. SORT BY NAME means that all your variables will be arranged alphabetically which will make it easier to find the variables you want to use.

1. Pew Research Center. (2016). August 2016 Political Survey. Retrieved February 16, 2019, 2019, from <http://www.people-press.org/dataset/august-2016-political-survey/> [↑](#footnote-ref-1)
2. If you would like to learn more about the Pew Research Center, click [here](http://www.pewresearch.org/). [↑](#footnote-ref-2)
3. These examples focus on populations made up of individuals. Populations can also consist of other units such as businesses or colleges and universities. All colleges and universities in your state or in the nation would be another example of a population. [↑](#footnote-ref-3)
4. There are many different types of error that can occur in a study. Sampling error is only one. We’ll discuss others in this exercise. We can’t ever eliminate all error. What we can do is to try to minimize error. [↑](#footnote-ref-4)
5. Simple random samples are one type of probability sample where every individual in the population has the same chance of being selected in the sample. [↑](#footnote-ref-5)
6. Ernest L. Cowles and Edward Nelson. (2019). *An Introduction to Survey Research Volume 1—The Basics of Survey Research* and *Volume 2—Carrying Out the Survey*, New York: Business Expert Press. [↑](#footnote-ref-6)
7. You can download the full data set by going to <http://www.people-press.org/dataset/august-2016-political-survey/> and clicking on Download Dataset. [↑](#footnote-ref-7)
8. Sullivan, Thomas J. (1992). *Applied Sociology – Research and Critical Thinking*, New York: Macmillan. [↑](#footnote-ref-8)
9. Pew Research Center. (2016). August 2016 Political Survey. Retrieved February 16, 2019, 2019, from <http://www.people-press.org/dataset/august-2016-political-survey/> [↑](#footnote-ref-9)
10. Pew Research Center. (2016). August 2016 Political Survey. Retrieved February 16, 2019, 2019, from <http://www.people-press.org/dataset/august-2016-political-survey/> [↑](#footnote-ref-10)
11. Sullivan, Thomas J. 1992. *Applied Sociology – Research and Critical Thinking*. Macmillan: New York, pp. 12-16. [↑](#footnote-ref-11)
12. Pew Research Center. (2016). August 2016 Political Survey. Retrieved February 16, 2019, 2019, from <http://www.people-press.org/dataset/august-2016-political-survey/> [↑](#footnote-ref-12)
13. You’re probably wondering why the frequencies are decimals. That’s a result of weighting the data. Surveys typically weight the data so the sample better represents the population from which the sample was selected. This results in frequencies that have decimals. When you refer to the frequencies round the decimal values off to the nearest whole number. So, 2,010.23 rounds to 2,010. [↑](#footnote-ref-13)
14. Frequency distributions can be grouped or ungrouped.  Think of age.  We could have a distribution that lists all the ages in years for the respondents in our survey.  One of the variables (*demo\_age*) in our data set does this.  But we could also divide age into a series of categories such as 18 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, and 65 and older. This is how the variable *demo\_agerec1* classifies respondents into a set of categories. In a grouped frequency distribution, the mode would be the most common category or categories.  [↑](#footnote-ref-14)
15. In a grouped frequency distribution, the median is the category that contains the middle value. [↑](#footnote-ref-15)
16. You might wonder why we didn’t use an example from the Pew Survey.  There isn’t one.  There are examples in other data sets. Think about profit for businesses over a fiscal year.  There is no absolute zero.  Profit could be positive or negative. [↑](#footnote-ref-16)
17. The 50th percentile is the value that has 50% of the cases above it and 50% below it. Thus, it’s the same as the median. They are just two equivalent ways of saying the same thing. [↑](#footnote-ref-17)
18. The Index of Qualitative Variation can be used to measure dispersion for a nominal variable. PSPP does not compute this statistic so we’re not going to discuss it here. [↑](#footnote-ref-18)
19. Pew Research Center. (2016). August 2016 Political Survey. Retrieved February 16, 2019, 2019, from <http://www.people-press.org/dataset/august-2016-political-survey/> [↑](#footnote-ref-19)
20. The wording of the question in the survey was, “What do you think is more important – to protect the right of Americans to own guns, or to control gun ownership?” [↑](#footnote-ref-20)
21. The null hypothesis is often called the hypothesis of no difference. We’re saying that there is no relationship between these two variables. In other words, there’s nothing there. [↑](#footnote-ref-21)
22. Chi Square tells us nothing about the strength of the relationship. Don’t make the mistake of assuming that the larger the value of Chi Square, the stronger the relationship. To measure the strength of a relationship, you need a measure of association. [↑](#footnote-ref-22)
23. Measures of association for nominal variables range from 0 to 1 while measures for ordinal variables range from   
    -1 to +1. [↑](#footnote-ref-23)
24. PSPP will also compute the Uncertainty Coefficient (UC) and Lambda but we’re not going to consider these measures. [↑](#footnote-ref-24)
25. Pew Research Center. (2016). August 2016 Political Survey. Retrieved February 16, 2019, 2019, from <http://www.people-press.org/dataset/august-2016-political-survey/> [↑](#footnote-ref-25)
26. The null hypothesis is often called the hypothesis of no difference. We’re saying that there is no relationship between these two variables. In other words, there’s nothing there. [↑](#footnote-ref-26)
27. Chi Square tells us nothing about the strength of the relationship. Don’t make the mistake of assuming that the larger the value of Chi Square, the stronger the relationship. To measure the strength of a relationship, you need a measure of association. [↑](#footnote-ref-27)
28. Measures of association for nominal variables range from 0 to 1 while measures for ordinal variables range from   
    -1 to +1. [↑](#footnote-ref-28)
29. PSPP will also compute the Uncertainty Coefficient (UC) and Lambda but we’re not going to consider these measures. [↑](#footnote-ref-29)
30. Pew Research Center. (2016). August 2016 Political Survey. Retrieved February 16, 2019, 2019, from <http://www.people-press.org/dataset/august-2016-political-survey/> [↑](#footnote-ref-30)
31. The null hypothesis is often called the hypothesis of no difference. We’re saying that there is no relationship between these two variables. In other words, there’s nothing there. [↑](#footnote-ref-31)
32. Chi Square tells us nothing about the strength of the relationship. Don’t make the mistake of assuming that the larger the value of Chi Square, the stronger the relationship. To measure the strength of a relationship, you need a measure of association. [↑](#footnote-ref-32)
33. Measures of association for nominal variables range from 0 to 1 while measures for ordinal variables range from   
    -1 to +1. [↑](#footnote-ref-33)
34. PSPP will also compute the Uncertainty Coefficient (UC) and Lambda but we’re not going to consider these measures. [↑](#footnote-ref-34)
35. Pew Research Center. (2016). August 2016 Political Survey. Retrieved February 16, 2019, 2019, from <http://www.people-press.org/dataset/august-2016-political-survey/> [↑](#footnote-ref-35)
36. Dictionary.com. (2019). Retrieved June 6, 2019, from <https://www.dictionary.com/browse/socioeconomic-status?s=t> [↑](#footnote-ref-36)
37. The null hypothesis is often called the hypothesis of no difference. We’re saying that there is no relationship between these two variables. In other words, there’s nothing there. [↑](#footnote-ref-37)
38. Chi Square tells us nothing about the strength of the relationship. Don’t make the mistake of assuming that the larger the value of Chi Square, the stronger the relationship. To measure the strength of a relationship, you need a measure of association. [↑](#footnote-ref-38)
39. Measures of association for nominal variables range from 0 to 1 while measures for ordinal variables range from   
    -1 to +1. [↑](#footnote-ref-39)
40. PSPP will also compute the Uncertainty Coefficient (UC) and Lambda but we’re not going to consider these measures. [↑](#footnote-ref-40)
41. Pew Research Center. (2016). August 2016 Political Survey. Retrieved February 16, 2019, 2019, from <http://www.people-press.org/dataset/august-2016-political-survey/> [↑](#footnote-ref-41)
42. See Dictionary.com for an example. (2019). Retrieved June 6, 2019, from <https://www.dictionary.com/browse/socioeconomic-status?s=t> [↑](#footnote-ref-42)
43. The null hypothesis is often called the hypothesis of no difference. We’re saying that there is no relationship between these two variables. In other words, there’s nothing there. [↑](#footnote-ref-43)
44. Chi Square tells us nothing about the strength of the relationship. Don’t make the mistake of assuming that the larger the value of Chi Square, the stronger the relationship. To measure the strength of a relationship, you need a measure of association. [↑](#footnote-ref-44)
45. Measures of association for nominal variables range from 0 to 1 while measures for ordinal variables range from   
    -1 to +1. [↑](#footnote-ref-45)
46. PSPP will also compute the Uncertainty Coefficient (UC) and Lambda but we’re not going to consider these measures. [↑](#footnote-ref-46)
47. Pew Research Center. (2016). August 2016 Political Survey. Retrieved February 16, 2019, 2019, from <http://www.people-press.org/dataset/august-2016-political-survey/> [↑](#footnote-ref-47)
48. The null hypothesis is often called the hypothesis of no difference. We’re saying that there is no relationship between these two variables. In other words, there’s nothing there. [↑](#footnote-ref-48)
49. Chi Square tells us nothing about the strength of the relationship. Don’t make the mistake of assuming that the larger the value of Chi Square, the stronger the relationship. To measure the strength of a relationship, you need a measure of association. [↑](#footnote-ref-49)
50. Measures of association for nominal variables range from 0 to 1 while measures for ordinal variables range from   
    -1 to +1. [↑](#footnote-ref-50)
51. PSPP will also compute the Uncertainty Coefficient (UC) and Lambda but we’re not going to consider these measures. [↑](#footnote-ref-51)
52. Pew Research Center. (2016). August 2016 Political Survey. Retrieved February 16, 2019, 2019, from <http://www.people-press.org/dataset/august-2016-political-survey/> [↑](#footnote-ref-52)
53. Pew Research Center. (2016). August 2016 Political Survey. Retrieved February 16, 2019, 2019, from <http://www.people-press.org/dataset/august-2016-political-survey/> [↑](#footnote-ref-53)
54. Pew Research Center. (2016). August 2016 Political Survey. Retrieved February 16, 2019, 2019, from <http://www.people-press.org/dataset/august-2016-political-survey/> [↑](#footnote-ref-54)
55. Jane E. Miller. *The Chicago Guide to Writing About Numbers*. 2015 (2nd edition). Chicago: University of Chicago Press. [↑](#footnote-ref-55)
56. Sociology Writing Group. *A Guide to Writing Sociology Papers*. 2013 (7th edition). New York: Worth Publishers. [↑](#footnote-ref-56)