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

**Note to the Instructor:** The data set used in this exercise is Field\_2013\_subset\_for\_classes\_GUN\_CONTROL.sav which is a subset of a Field Poll conducted in February, 2013.  Some of the variables in this Field Poll have been recoded to make them easier to use and some new variables have been created.  The data have been weighted according to instructions from the Field Research Corporation.  This exercise uses FREQUENCIES to get frequency distributions and CROSSTABS to explore relationships between variables.  In CROSSTABS students are asked to use percentages.  Chi Square and measures of association will be considered in a later exercise.  A good reference on using SPSS is SPSS for Windows Version 23.0 A Basic Tutorial by Linda Fiddler, John Korey, Edward Nelson (Editor), and Elizabeth Nelson.  The online version of the book is on [the Social Science Research and Instructional Center's website.](http://ssric.org/node/582)  You have permission to use this exercise and to revise it to fit your needs.  Please send a copy of any revision to the author. Included with this exercise (as separate files) are more detailed notes to the instructors, the SPSS syntax necessary to carry out the exercise (SPSS syntax file), and the SPSS output for the exercise (SPSS output file). These, of course, will need to be removed as you prepare the exercise for your students.  Please contact the author for additional information.

I’m attaching the following files.

* [**Data subset which is a SPSS data file**](http://ssric.org/files/field_2013_subset_for_classes_GUN_CONTROL.sav) (.sav format).
* [**Extended notes for instructors**](http://ssric.org/files/Extended_Notes_for_GUN_CONTROL2g.docx). MS Word (.docx) format.
* [**SPSS syntax file**](http://ssric.org/files/SPSS_Syntax_for_GUN_CONTROL2G.sps) (.sps format).
* [**SPSS output file**](http://ssric.org/files/SPSS_Output_for_GUN_CONTROL2G.spv) (.spv format).
* [**This page**](http://ssric.org/files/GUN_CONTROL.docx) in MS Word (.docx) format.

## Goals of Exercise

The goal of this exercise is to explore the relationship between socioeconomic status and opinion on gun control.  The exercise also gives you practice in using several SPSS commands –FREQUENCIES and CROSSTABS.

### Part I—Socioeconomic Status

We’re going to use a Field Poll conducted in 2013 for this exercise.  The Field Poll is a statewide poll of registered voters in California conducted by the Field Research Corporation.  For this exercise we’re going to use a subset of this Field Poll. Your instructor will tell you how to access this data set which is called Field\_2013\_subset\_for\_classes\_GUN\_CONTROL.sav.

One of the most important and frequently used concepts in the social sciences is socioeconomic status which can be defined as “an individual's or group's position within a hierarchical social structure”[[1]](" \l "_edn1" \o ") or as “the social standing or class of an individual or group.”[[2]](" \l "_edn2" \o ")  Concepts are abstract ideas which need to be measured before they can be used in research.  Often researchers use individuals’ income, education, or occupation to measure their socioeconomic status.  In this exercise we’re going to use education and household income as our measures.

Respondent’s education is measured by the variable D\_EDUC1\_q103.  Run FREQUENCIES in SPSS to get the frequency distribution for this variable.  (See Chapter 4, FREQUENCIES, in the online SPSS book mentioned on page 1 of this exercise.) There are ten categories for this variable plus the missing value code (i.e., 11) which is the value for those who didn’t answer this question.

There are five columns in the output that SPSS gives you:

* The first column is the value label for the response category.
* The second column is the number of cases or frequency for each response.
* The third column is the percent.  The denominator for the percent is the total number of cases in the sample (834).
* The fourth column is the valid percent.  Here the denominator is the number of valid cases (830).  This is the total number of respondents (834) minus the number of cases with missing data (4).
* The fifth column is the cumulative percent.  Notice that these percents cumulate and eventually equal 100.0 for the last of the response categories.  Look at the cumulative percent for the third response category (i.e., high school graduate).  The cumulative percent is 20.7 which means that 20.7% of the respondents who answered this question (i.e., had valid responses) were high school graduates or had less education.

The percents and valid percents are very close for this variable.  That’s because there were only four respondents who didn’t answer the question.  When there are more cases with missing information, these percents can be quite different.

Ten categories is a lot.  Some of these response categories have relatively few cases in them. For example, there were only 15 respondents who answered “trade or vocational school.” Often we combine categories to reduce the number of categories.  This is called recoding.  There is another variable called D\_EDUC2\_recoded\_EDUC1 which is a recode of D\_EDUC1\_q103 and reduces the number of categories to four.  Run FREQUENCIES in SPSS to get the frequency distribution for this variable.  Notice that all the categories have relatively large number of cases in them.

Let’s make sure you know how to interpret the output that SPSS gives you for D\_EDUC2\_recoded\_EDUC1.  Answer the following questions.

* How many respondents were college grads?
* What percent of all the cases (834) were college grads?  How was this percent computed?  Show the arithmetic for the calculation.
* What percent of those respondents who actually answered the question (830) were college grads?  How was this percent computed?  Show the arithmetic for the calculation.
* What percent were college grads or had less education?
* What percent had less education than college grads?

### Part II – Opinion on Gun Control

The second concept we will be using in this exercise is the respondent’s opinion on gun control.  This refers to whether or not they are in favor of controls on gun ownership.  The question used to measure this concept is the following” “What do you think is more important – to protect the right of Americans to own guns, or to impose greater controls on gun ownership?”  This variable is named G1\_q13.  Run FREQUENCIES in SPSS to get the frequency distribution for this variable.

Write a paragraph describing what the frequency distribution tells you about respondents’ opinions on gun control.  Be sure to use the valid percents in your answer and indicate what percent of respondents answered this question and what percent had no opinion.

### Part III – Relationship between Respondents’ Education and Opinion on Gun Control

Now we want to explore the relationship between socioeconomic status using education as our measure of socioeconomic status and how respondents feel about gun control.  To do this we’re going to run CROSSTABS in SPSS to produce a crosstabulation of our two measures – D\_EDUC2\_recoded\_EDUC1 and G1\_q13.  (See Chapter 5, CROSSTABS, in the online SPSS book.)

It’s important to distinguish between our dependent variable and our independent variable.  The dependent variable is what you are trying to explain and the independent variable is the variable that you think will help you explain the variation in your dependent variable. We want to explain why some people favor increased controls on guns and why others oppose it.  Our hypothesis is that socioeconomic status will help us answer this question.  In other words, it will help explain the variation in people’s opinion about gun control.  More specifically, our hypothesis is that those with higher socioeconomic status will be more in favor of increased gun control while those lower in socioeconomic status will be more opposed to gun control.  A hypothesis specifies the relationship you expect to find between two variables.

When you run the crosstab for your two variables, put the independent variable in the column and the dependent variable in the row of your table.  If you do this, you will always want to tell SPSS to compute the column percents.  Remember that you want to compare the percents straight across, not down or on the diagonal.

Write a paragraph describing the relationship between respondents’ education and their opinion on gun control.  Were those with more education more or less likely than those with less education to favor gun control? Use the percents to help you describe this relationship.

### Part IV – Relationship between Respondents’ Household Income and Gun Control

Now let’s explore the relationship between socioeconomic status using household income as our measure of socioeconomic status and how respondents feel about gun control.  We’re going to run a CROSSTABS in SPSS to produce a crosstabulation of our two measures – D\_INCOME1\_q109 and G1\_q13.   Write a hypothesis indicating what you expect the relationship between these two variables to look like.

Run CROSSTABS to produce a crosstabulation of your two variables.  Make sure you put the independent variable in the column and the dependent variable in the row and that you ask for the correct percents.

Write a paragraph describing the relationship between respondents’ household income and their opinion on gun control.  Were those with higher income more or less likely than those with lower income to favor gun control? Use the percents to help you describe this relationship.

### Part V – Relationship between Respondents’ Assessment of their Own Financial Situation and Gun Control

There are two questions in the Field Poll that ask respondents to assess their own financial situation.

* EC2\_q25 – “Would you say that you and your family are financially better off or worse off today than you were a year ago?” (Response categories are better off, no change, worse off, no opinion.)
* EC3\_q26 – “Looking ahead, do you think that a year from now you will be better off financially, worse off, or just about the same as now?”  (Response categories are better off in a year, no change, worse off in a year, no opinion.)

Notice that people in any income category from low to high could think of themselves as better or worse off.  But what we would expect the relationship to look like? Write two hypotheses that indicate what you expect the relationship will look like for each of these two variables and opinion on gun control.

Now run the two crosstabs and see what the relationship actually looks like.  Use the percents to describe the relationship being sure to discuss whether the data support your hypothesis. Note that one of the columns in each of these tables (i.e., the no opinion column) has very few cases. You should ignore the percents in these columns because of the small bases.

### Part VI – Conclusions about Socioeconomic Status and Gun Control

Look at the crosstabs you ran.  What do they tell you about the relationship between socioeconomic status and opinion on gun control?  Do you find basically the same relationship in all the tables?  What does this tell you about the relationship of socioeconomic status and opinion on gun control?

### Part VII – Missing Data

Let’s take a closer look at the amount of missing data for the five variables we used in this exercise.  Run FREQUENCIES again for these five variables.

* D\_EDUC2\_recoded\_EDUC1
* D\_INCOME1\_q109
* EC2\_q25
* EC3\_q26
* G1\_q13

Let’s look at the missing data for each of these variables?

* D\_EDUC2\_recoded\_EDUC1 is a recode of D\_EDUC1\_q103 so the missing data for this variable is referred to as system-missing data.  There are very few cases (4) with missing data here.
* G1\_q13 doesn’t have any missing data.  Remember that “no opinion” is not treated as missing data.
* D\_INCOME1\_q109 has quite a bit of missing data.  There were 122 respondents who said they either didn’t know their income or refused to tell the interviewer.  That’s 14.7% of the entire sample.  Income often produces the most missing data in many polls.
* EC2\_q25 and EC3\_q26 have something that we haven’t seen before.  There is a missing value called “question not asked.”  What does that mean?  It’s standard practice in survey research to ask some questions of a random subsample of the entire sample.  Notice that there are 415 cases where the question was not asked.  That’s about half of the entire sample.  These questions were asked of a random 50% of the entire sample.  So it’s still a random sample but it has fewer cases in it.  This procedure allows survey researchers to include more questions in a survey without increasing the length of the survey and thereby increasing interviewee fatigue.  There are two forms to this interview (A and B).  Some questions are in Form A only, some are in Form B only, and some are in both forms.

So what do we make of all this?

* If there is no missing data for a variable, we don’t have anything to worry about.
* If there is missing data created by a question being asked of a random subsample of the entire sample, we have two problems.  One is that we’re working with a smaller sample.  In this case, a sample of 415 rather than a sample of 834.  Smaller samples have certain limitations that we’ll discuss in later exercises.  The second problem is that variables from Form A cannot be crosstabulated with variables from Form B.  See if you can figure why not?  By the way, the form of the survey is in the variable label that appears in your output.  If the form is not mentioned, then the question was included in both forms.  Note that G1\_q13 was included in both Form A and Form B which means it was asked of the entire 834 respondents.
* One of the variables, D\_INCOME1\_q109, had quite a bit of missing data.  There were 122 respondents (14.7% of the entire sample) that didn’t answer the income questions.  So what’s the problem here?  If lower income or higher income respondents were more likely to not answer the question, then we could have a seriously biased estimate of what the income distribution for this population looks like.  So the moral of this story is that you should be suspicious of variables with a lot of missing data unless this is a result of not asking the question for that segment of the sample.

We’ll come back to the issue of missing data in later exercises.

[[1]](" \l "_ednref1" \o ") See [Dictionary.com](http://www.dictionary.com/browse/socioeconomic-status) for a definition of socioeconomic status.

[[2]](" \l "_ednref2" \o ") See the American Psychological Association’s [website](http://www.apa.org/topics/socioeconomic-status/) for another definition of socioeconomic status.