# Exercises on Exploring Confidence in Societal Institutions

# and Spending Priorities

# Table of Contents

Page

Exercise 1 – Exploring Confidence in Societal Institutions 4

Exercise 2 – Exploring Spending Priorities 7

Exercise 3 – Exploring the Relationship Between Confidence in Institutions and 11  
Spending Priorities

Exercise 4 – Exploring the Relationship Between Confidence in Institutions 15  
and Spending Priorities Controlling for Political Party Identification

Appendix – Codebook for the Data Set 25

## Preface

The data set used in this exercise is the 2018 General Social Survey. The name of the data set is GSS18CONFIDENCE.SAV which is a subset of the full data set. Some of the variables in the GSS have been recoded to make them easier to use and some new variables have been created.  The data have been weighted according to the instructions from the National Opinion Research Center.  These exercises use SPSS to analyze the data but you could also easily use PSPP.  A good reference for using SPSS is [*SPSS Statistics (Version 26): A Basic Tutorial*](http://www.ssric.org/node/696)by Edward Nelson and John Korey.

The focus of these exercises is on frequency distributions and two-variable and three-variable tables. The exercises revolve around respondents' confidence in societal institutions and spending priorities.

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

The exercises were written so that each exercise is independent of the other exercises. That means that there is some redundancy across the exercises. If you choose to use several exercises you may want to remove some of the redundant material.

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

# Exercise 1 – Exploring Confidence in Societal Institutions

## **Goal of Exercise**

## The goal of this exercise is to rank order societal institutions in terms of the amount of confidence that respondents have in them.  The exercise also gives you practice using FREQUENCIES in SPSS.

## **Part I—Getting the Frequency Distributions**

We’re going to use the General Social Survey (GSS) for this exercise.  The GSS is a national probability sample of adults in the United States conducted by the National Opinion Research Center (NORC).  The GSS started in 1972 and has been an annual or biannual survey ever since. For this exercise we’re going to use a subset of the 2018 GSS. Your instructor will tell you how to access this data set which is called GSS18CONFIDENCE.SAV. You can also download the data by clicking on this [link](http://ssric.org/node/552).

The GSS is an example of a social survey.  The investigators selected a sample from the population of all adults in the United States.  This particular survey was conducted in 2018 and is a relatively large sample of a little more than 2,300 adults.  In a survey we ask respondents questions and use their answers as data for our analysis.  The answers to these questions are used as measures of various concepts.  In the language of survey research these measures are typically referred to as variables.  Often, we want to describe respondents in terms of social characteristics such as marital status, education, and age.  Other times we want to describe respondents in terms of opinions about social issues such as abortion and gun control.  In this exercise we want to look at the confidence that respondents have in societal institutions.  These are all variables in the GSS.

The GSS uses the following question to measure confidence in institutions – “I am going to name some institutions in this country. As far as the people running these institutions are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them?”  This exercise focuses on the following nine societal institutions.  Respondents are asked whether they have a great deal, only some, or hardly any confidence in these institutions.  The names of the variables are in parentheses.[[1]](#footnote-1)

* military (CONARMY),
* major companies (CONBUS),
* organized religion (CONCLERG),
* education (CONEDUC),
* executive branch of the federal government (CONFED),
* banks and financial institutions (CONFINAN),
* U.S. Supreme Court (CONJUDGE),
* organized labor (CONLABOR), and
* congress (CONLEGIS).

Run FREQUENCIES in SPSS for all nine variables.  (See Frequencies in Chapter 4 of the SPSS [online SPSS tutorial](http://ssric.org/node/696).) There should be four columns of numbers.

* Just to the right of the value label you should see the FREQUENCY column.  This tells you the number of respondents that said a great deal, only some, or hardly any.
* To the right of that column there will be the PERCENT column.  This converts the frequencies to percents using all cases (2,348) as the denominator.  Notice that this includes those respondents who gave a valid response and those who have missing information (i.e., said they didn’t know (DK), refused to answer the question (NA), or were not asked the question (IAP)).
* To the right of the percent column there is the VALID PERCENT column.  This converts the frequencies to percents using only those cases with valid information (i.e., replied great deal, only some, or hardly any) in the denominator (i.e., 1,546 for CONBUS).
* Finally the far-right column is the CUMULATIVE PERCENT column.  This cumulates the valid percents.  Look at the table for CONBUS.  The first entry in the cumulative percents column is 19.7 because 19.7% of the cases responded a great deal.  The second entry is 85.7 because 85.7% said either a great deal or hardly any.  The third entry is 100.0 because all 100% of the cases gave one of these three responses.

It’s important to understand the difference between the percent column and the valid percent column.  Write out the arithmetic for computing both the percent and the valid percent for those who said they had a great deal of confidence in major companies (i.e., CONBUS).

## Part II – Rank Ordering the Institutions in Terms of the Percent who had a Great Deal of Confidence

In the chart below, write the name of the institution and the percent of respondents who had a great deal of confidence in each institution.  Arrange the institutions from rank 1 to rank 9 where 1 is the institution in which respondents had the most confidence and 9 is the institution in which respondents had the least confidence.

|  |  |  |
| --- | --- | --- |
| Percent of Respondents who had a Great Deal of Confidence in Each Institution | | |
| Rank | Institution | Percent Great Deal of Confidence |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |
| 6 |  |  |
| 7 |  |  |
| 8 |  |  |
| 9 |  |  |

## Part III – Rank Ordering the Institutions in Terms of the Percent who had at Least Some or a Great Deal of Confidence

In the chart below write the name of the institution and the percent of respondents who had at least some **or** a great deal of confidence in each institution.  Arrange the institutions from rank 1 to rank 9 where 1 is the institution in which respondents had the most confidence and 9 is the institution in which respondents had the least confidence.

|  |  |  |
| --- | --- | --- |
| Percent of Respondents who had at Least Some or a Great Deal of Confidence in Each Institution | | |
| Rank | Institution | Percent at Least Some or Great Deal of Confidence |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |
| 6 |  |  |
| 7 |  |  |
| 8 |  |  |
| 9 |  |  |

## Part IV – Summary

Did each method of ranking produce the same rank order of institutions?  What does that tell you about the use of statistics in analyzing data?

What conclusions can you draw about the level of confidence that adults in the U.S. have in our societal institutions?

# **Exercise 2 – Exploring Spending Priorities**

## **Goal of Exercise**

The goal of this exercise is to determine which areas or problems people think we ought to be spending more money on and which we ought to be spending less on.  We will also consider the way in which question wording affects what people tell us.  The exercise also gives you practice in using FREQUENCIES in SPSS.

## **Part I—Getting the Frequency Distributions**

We’re going to use the General Social Survey (GSS) for this exercise.  The GSS is a national probability sample of adults in the United States conducted by the National Opinion Research Center (NORC).  The GSS started in 1972 and has been an annual or biannual survey ever since. For this exercise we’re going to use a subset of the 2018 GSS. Your instructor will tell you how to access this data set which is called GSS18CONFIDENCE.SAV. You can also download the data by clicking on this [link](http://ssric.org/node/552).

The GSS is an example of a social survey.  The investigators selected a sample from the population of all adults in the United States.  This particular survey was conducted in 2018 and is a relatively large sample of a little more than 2,300 adults.  In a survey we ask respondents questions and use their answers as data for our analysis.  The answers to these questions are used as measures of various concepts.  In the language of survey research these measures are typically referred to as variables.  Often, we want to describe respondents in terms of social characteristics such as marital status, education, and age.  Other times we want to describe respondents in terms of opinions about social issues such as abortion and gun control.  In this exercise we want to look at the areas or problems for which respondents think we ought to be spending more or less.  These are all variables in the GSS.

There are a number of different areas or problems on which the government spends money.  Some people think we should be spending more on some of these problems or areas while others think we should be spending less or about the same.  The GSS uses the following question to measure spending priorities – “First I would like to talk with you about some things people think about today. We are faced with many problems in this country, none of which can be solved easily or inexpensively. I'm going to name some of these problems, and for each one I'd like you to tell me whether you think we're spending too much money on it, too little money, or about the right amount.”  For this exercise we picked five areas or problems – foreign aid (NATAID), halting the rising crime rate (NATCRIME), dealing with drug addiction (NATDRUG), welfare (NATFARE), and improving the conditions of blacks (NATRACE).  The names of the variables are in parentheses.[[2]](#footnote-2)

Run FREQUENCIES in SPSS for all five variables.  (See Frequencies in Chapter 4 of the [online SPSS tutorial](http://ssric.org/node/696).) There should be four columns of numbers.

* Just to the right of the value label you should see the FREQUENCY column.  This tells you the number of respondents that said too little, about right, or too much.
* To the right of that column there will be the PERCENT column.  This converts the frequencies to percents using all cases (2,348) as the denominator.  Notice that this includes those respondents who gave a valid response and those who have missing information (i.e., said they didn’t know (DK), refused to answer the question (NA), or were not asked the question (IAP).
* To the right of the percent column there is the VALID PERCENT column.  This converts the frequencies to percents using only those cases with valid information (i.e., replied too little, about right, or too much) in the denominator (i.e., 1,125 for NATAID).
* Finally the far-right column is the CUMULATIVE PERCENT column.  This cumulates the valid percents.  Look at the table for NATAID.  The first entry in the cumulative percents column is 16.6 because 16.6% of the cases responded too little.  The second entry is 54.9 because 54.9% said either too little or about right.  The third entry is 100.0 because all 100% of the cases gave one of these three responses.

It’s important to understand the difference between the percent column and the valid percent column.  Write out the arithmetic for computing both the percent and the valid percent for those who said that spending on foreign aid was too little, about right, or too much (i.e., NATAID).

## Part II – Rank Ordering the Spending Priorities in Terms of the Percent Who Said Too Much

In the chart below, write the name of the problem or area and the percent of respondents who thought we were spending too much.  Arrange the areas or problems from rank 1 to rank 5 where 1 is the area or problem on which respondents were most likely to think we were spending too much and 5 is the area or problem on which respondents were least likely to think we were spending too much.

|  |  |  |
| --- | --- | --- |
| Percent of Respondents Who Said We Were Spending Too Much | | |
| Rank | Area or problem | Percent Saying We Were Spending Too Much |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## Part III – Alternative Ways of Asking the Question

There is usually more than one way to ask a question.  In this case the GSS built an experiment into the survey.  What they did was to vary the way the area or problem was phrased.  Half of the respondents were **randomly** assigned to one version of the question and the other half was **randomly** assigned to the other version.  Here’s the way the questions were asked.

* foreign aid (NATAID) vs. assistance to other countries (NATAIDY)
* halting the rising crime rate (CRIME) vs. law enforcement (CRIMY)
* dealing with drug addiction (NATDRUG) vs. drug rehabilitation (NATDRUGY)
* welfare (NATFARE) vs. assistance to the poor (NATFAREY)
* improving the conditions of blacks (NATRACE) vs. assistance to blacks (NATRACEY)

Let’s see if the way the question was worded affected how respondents answered the question.  Run FREQUENCIES in SPSS for all ten variables and enter the percent saying we’re spending too much in the appropriate cell for each variable.  Then compute the difference in the two percents for each area or problem.  In order to standardize our answers, subtract the second of the two variables from the first of the two variables.  The percents for the first pair of variables are filled in to show you what to do.

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Question | Percent Saying We're Spending Too Much | Difference Between the Two Questions |
| NATAID | Foreign aid | 45.1% | -13.0 |
| NATAIDY | Assistance to other countries | 58.1% |  |
| NATCRIME | Halting the rising crime rate |  |  |
| NATCRIMY | Law enforcement |  |  |
| NATDRUG | Drug addiction |  |  |
| NATDRUGY | Drug rehabilitation |  |  |
| NATFARE | Welfare |  |  |
| NATFAREY | Assistance to the poor |  |  |
| NATRACE | Improving the condition of blacks |  |  |
| NATRACEY | Assistance to blacks |  |  |

# Part IV – Summary

Which areas or problems were respondents most likely to think we’re spending too much and which were they least likely to think we’re spending too much?

Did the way the questions were asked affect how respondents answered the questions?  For which of the areas or problems was the difference the greatest? the least?

What do you think this means for the nation’s spending priorities?

# Exercise 3 – Exploring the Relationship Between Confidence in Institutions and Spending Priorities

## Goal of Exercise

The goal of this exercise is to determine whether confidence in the executive branch is related to whether respondents think we are spending too little, about right, or too much on various spending priorities such as the military, the problems of big cities, the environment, welfare, and assistance to blacks. The exercise also gives you practice in using FREQUENCIES and CROSSTABS in SPSS.

## Part I—Choosing our Dependent Variable

We’re going to use the General Social Survey (GSS) for this exercise. The GSS is a national probability sample of adults in the United States conducted by the National Opinion Research Center (NORC). The GSS started in 1972 and has been an annual or biannual survey ever since. For this exercise we’re going to use a subset of the 2018 GSS. Your instructor will tell you how to access this data set which is called GSS18CONFIDENCE.SAV. You can also download the data by clicking on this [link](http://ssric.org/node/552).

There are a number of different areas or problems on which the government spends money. Some people think we should be spending more on some of these problems or areas while others think we should be spending less or about the same. The GSS uses the following question to measure spending priorities – “First I would like to talk with you about some things people think about today. We are faced with many problems in this country, none of which can be solved easily or inexpensively. I'm going to name some of these problems, and for each one I'd like you to tell me whether you think we're spending too much money on it, too little money, or about the right amount.” For this exercise were going to focus on whether respondents think we are spending too much, about right, or too little on some of these issues.[[3]](#footnote-3)

* the military
  + the military, assistance, and defense (NATARMS)
  + national defense (NATARMSY)
* big cities
  + solving the problems of big cities (NATCITY)
  + assistance to big cities (NATCITYY)
* the environment
  + improving and protecting the environment (NATENVIR)
  + the environment (NATENVIY)
* welfare
  + welfare (NATFARE)
  + assistance to the poor (NATFAREY)​
* assistance to blacks
  + improving the conditions of blacks (NATRACE)
  + assistance to blacks (NATRACEY)

Notice that there are two versions of each question. Respondents were randomly assigned to each version. This allows us to see whether question wording affects what people tell us.

These variables will be our dependent variables. Remember that the dependent variable is what we’re trying to explain. We want to explain why some people think we are spending too little on these issues while others think we are spending too much and still others feel we are spending about the right amount of money.

Let's start with spending on the military. Run FREQUENCIES in SPSS for NATARMS and NATARMSY. (See Frequencies in Chapter 4 of the [online SPSS tutorial](http://ssric.org/node/696).) Notice that each version of the question produces about the same result. The responses are pretty evenly split. About 31% of respondents think we are spending too little while about 41% think we are spending too much and about 28% think we are spending about the right amount of money. Clearly there is disagreement on how much we should be spending on the military but the two versions of the question produce similar results.

## Part II – Choosing Our Independent Variable

Independent variables are variables that we think might explain why some people feel we are spending too little on a particular issue or area while others think we are spending too much or about the right amount. There are many possible independent variables. We’re going to focus on the amount of confidence that respondents have in the executive branch of the federal government (CONFED).

We want to write a hypothesis that specifies the relationship that you expect to find between confidence in the executive branch and how respondents feel about spending on the environment. Our hypothesis might be that the less confidence people have in the executive branch of the federal government, the more likely they are to think we are spending too little on the environment and the less likely they are to feel we are spending too much.

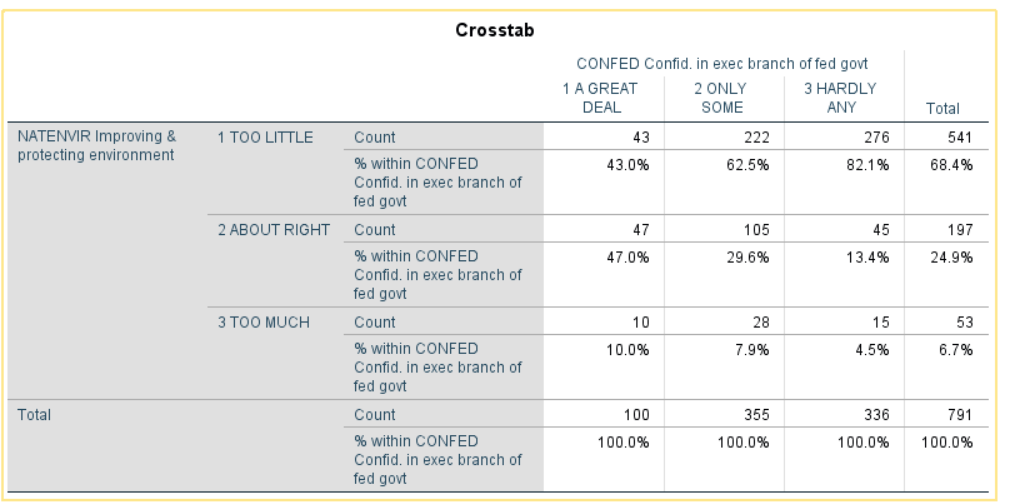
Now we have a clear hypothesis that specifies the relationship we expect to find between confidence in the executive branch and spending priorities on the environment. But why do we expect to find this relationship? We need to provide a clear and convincing argument to support our hypothesis. If we’re asked the why question, how will we respond? We might point out that Republicans generally want less environmental regulations and the President in 2018 (i.e., when the survey was conducted) was a Republican. For this reason those who have less confidence in the executive branch will feel that we ought to be spending more on the environment.

Now it’s your turn. Another issue that was covered in the survey was spending on education (NATEDUC). Write a hypothesis that specifies the relationship that you expect to find between confidence in the executive branch (CONFED) and spending on education (NATEDUC). Then write a rationale explaining why you expect to find this relationship.

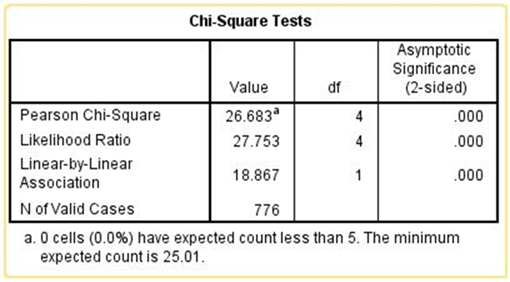
## Part III – Let’s Look at the Data

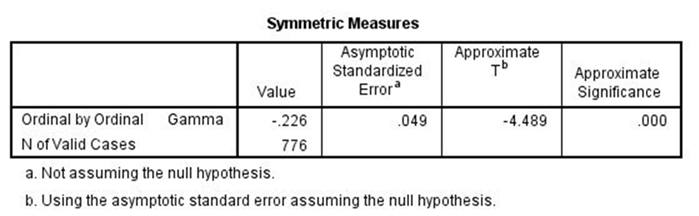
Now that we have a hypothesis and a rationale for our hypothesis, it’s time to look at the data. First, you need to be clear which is the dependent and independent variable. The dependent variable is what you are trying to explain which is why some people feel we are spending too much on the environment and others think we are spending too much or about the right amount. Use NATENVIR as your dependent variable. The independent variable is the variable you think might help you explain differences of opinion on spending. Use CONFED as your independent variable. In this case our hypothesis suggests that confidence in the executive branch influences spending priorities for the environment. Run CROSSTABS in SPSS to get the table that shows the relationship between these two variables. (See Crosstabulation in Chapter 5 of the online SPSS tutorial.) Put the independent variable in the column and the dependent variables in the rows of your table. If you do this, you will always want to tell SPSS to compute the column percents. Also tell SPSS to compute Chi Square and an appropriate measure of association.

Here is the crosstabulation for the question that asks about spending on the environment (NATENVIR).



To interpret the crosstabulation always compare the percents in the direction opposite to the way in which they sum to 100%. Since you asked for the column percents, the percents sum down to 100. That means that you want to compare the percents straight across. Look at the first row (i.e., too little). The more confidence people have in the executive branch, the less likely they are to feel that we are spending too little. On the other hand, the more confidence they have in the executive branch, the more likely they are to feel we’re spending too much. For example, 43.0% of those who have a great deal of confidence in the executive branch think we are spending too little on the environment compared to 82.1% of those who have hardly any confidence. On the other hand, 4.5% of those who have hardly any confidence in the executive branch feel we are spending too much on the environment compared to 10.0% of those who have a great deal of confidence.





Chi Square is statistically significant and tau-c (an appropriate measure of association) is -0.226 suggests a moderately strong relationship. Clearly the data support our hypothesis.

Run the crosstabulation for the other question that asks about the environment (NATENVIY)? Did you find the same relationship for both versions of the question on the environment?

## Part IV – More Analysis

Now repeat this analysis for the hypothesis that you wrote in Part 2. Tell SPSS to get the crosstabulation of the two variables (CONFED and NATEDUC) and write a paragraph describing the relationship between your two variables and whether the table supports your hypothesis. Be sure to use both versions of the question and to tell SPSS to compute Chi Square and the appropriate measure of association. Use all this information to interpret the relationship between your two variables.

## Part V – Summary

Write a paragraph indicating what you learned about the relationship between confidence in the executive branch and spending priorities for education. Be as specific as possible.

## Exercise 4 – Exploring the Relationship Between Confidence in Institutions and Spending Priorities Controlling for Political Party Identification

## **Goal of Exercise**

The goal of this exercise is to determine whether confidence in the executive branch is related to whether respondents think we are spending too little, about right, or too much on the environment and education with controls for political party identification. The exercise also gives you practice in using RECODE, FREQUENCIES and CROSSTABS in SPSS.

## **Part I—Choosing our Dependent Variable**

We’re going to use the General Social Survey (GSS) for this exercise. The GSS is a national probability sample of adults in the United States conducted by the National Opinion Research Center (NORC). The GSS started in 1972 and has been an annual or biannual survey ever since. For this exercise we’re going to use a subset of the 2018 GSS. Your instructor will tell you how to access this data set which is called GSS18CONFIDENCE.SAV. You can also download the data by clicking on this [link](http://ssric.org/node/552).

There are a number of different areas or problems on which the government spends money. Some people think we should be spending more on some of these problems or areas while others think we should be spending less or about the same. The GSS uses the following question to measure spending priorities – “First I would like to talk with you about some things people think about today. We are faced with many problems in this country, none of which can be solved easily or inexpensively. I'm going to name some of these problems, and for each one I'd like you to tell me whether you think we're spending too much money on it, too little money, or about the right amount.” For this exercise were going to focus on whether respondents think we are spending too much, about right, or too little on the environment (NATENVIR) and education (NATEDUC).

These variables will be our dependent variables. Remember that the dependent variable is what we’re trying to explain. We want to explain why some people think we are spending too little on these issues while others think we are spending too much and still others feel we are spending about the right amount of money.

## **Part II – Choosing Our Independent Variable**

Independent variables are variables that we think might explain why some people feel we are spending too little on the environment and education while others think we are spending too much or about the right amount. There are many possible independent variables. We’re going to focus on the amount of confidence that respondents have in the executive branch of the federal government (CONFED).

We want to write a hypothesis that specifies the relationship that you expect to find between confidence in the executive branch and how respondents feel about spending on the environment. Our hypothesis might be that the less confidence people have in the executive branch of the federal government, the more likely they are to think we are spending too little on the environment and the less likely they are to feel we are spending too much.

Now we have a clear hypothesis that specifies the relationship we expect to find between confidence in the executive branch and spending priorities on the environment. But why do we expect to find this relationship? We need to provide a clear and convincing argument to support our hypothesis. If we’re asked the why question, how will we respond? We might point out that Republicans generally want less environmental regulations and the President in 2018 (i.e., when the survey was conducted) was a Republican. For this reason those who have less confidence in the executive branch will feel that we ought to be spending more on the environment.

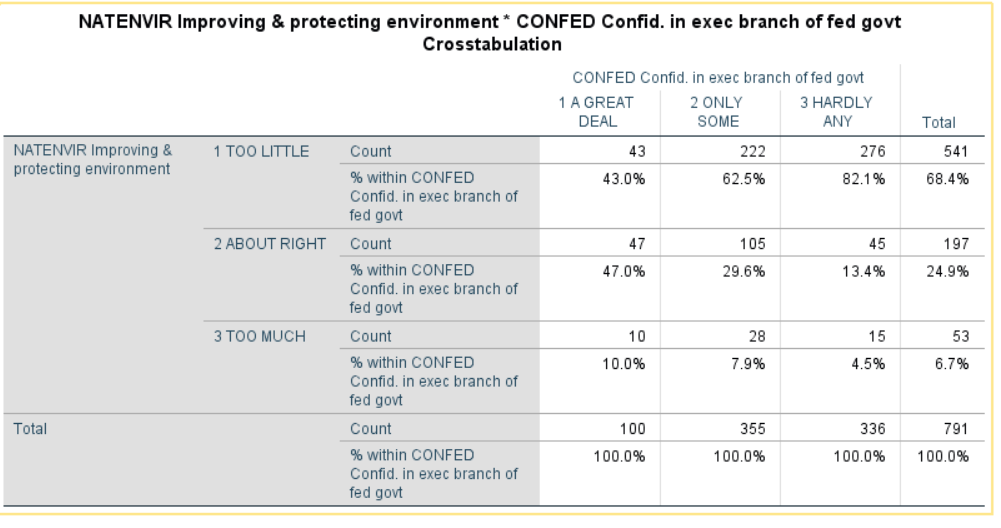
Now it’s your turn. The other spending priority that we’re going to consider in this exercise is spending on education. Write a hypothesis that specifies the relationship that you expect to find between confidence in the executive branch (CONFED) and spending on education (NATEDUC). Then write an argument explaining why you expect to find this relationship.

## **Part III – Let’s Look at the Data**

Now that we have a hypothesis and a rationale for our hypothesis, it’s time to look at the data. First, you need to be clear which is the dependent and independent variable. The dependent variable is what you are trying to explain which is why some people feel we are spending too much on the environment and others think we are spending too little or about the right amount. The independent variable is the variable you think might help you explain differences of opinion on spending. In this case our hypothesis suggests that confidence in the executive branch influences spending priorities for the environment.

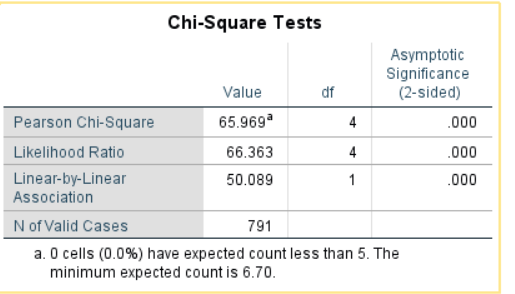
Run CROSSTABS in SPSS to get the table that shows the relationship between these two variables (NATENVIR and CONFED). (See Crosstabulation in Chapter 5 of the  [online SPSS tutorial](http://ssric.org/node/696).) Put the independent variable in the column and the dependent variables in the rows of your table. If you do this, you will always want to tell SPSS to compute the column percents. Also tell SPSS to compute Chi Square and an appropriate measure of association.

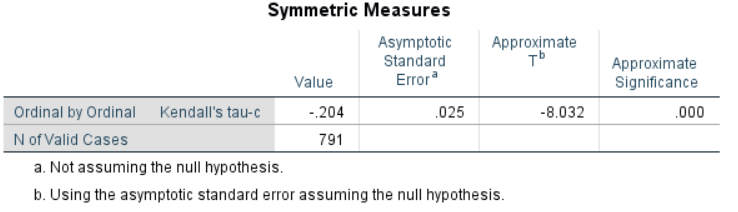
Here is the crosstabulation for the question that asks about spending on the environment (NATENVIR).



To interpret the crosstabulation always compare the percents in the direction opposite to the way in which they sum to 100%. Since you asked for the column percents, the percents sum down to 100. That means that you want to compare the percents straight across. Look at the first row (i.e., too little). The more confidence people have in the executive branch, the less likely they are to feel that we are spending too little. On the other hand, the more confidence they have in the executive branch, the more likely they are to feel we’re spending too much. For example, 43.0% of those who have a great deal of confidence in the executive branch think we are spending too little on the military compared to 82.1% of those who have hardly any confidence. But 10.0% of those who have a great deal of confidence think we are spending too much on the military compared to 4.8% of those who had hardly any confidence.

Here are the Chi Square test and a measure of association. We used tau-c as our measure of association in this example.





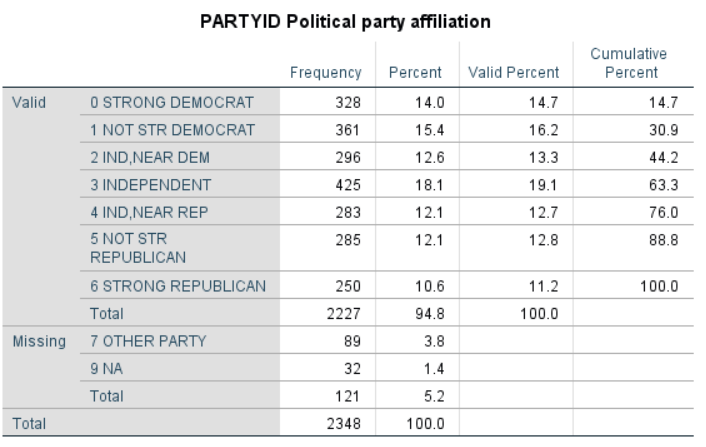
Chi Square is statistically significant and the tau-c (i.e., an appropriate measure of association) of -0.204 suggests a somewhat strong relationship. Clearly the data support our hypothesis.

## **Part IV – Now It’s Your Turn**

In Part 2, you wrote a hypothesis that specified the relationship that you expect to find between confidence in the executive branch (CONFED) and spending on education (NATEDUC). Now repeat the analysis that we carried out in Part 3 by telling SPSS to get the crosstabulation of CONFED and NATEDUC and write a paragraph describing the relationship between your two variables and whether the table supports your hypothesis. Be sure to tell SPSS to compute Chi Square and the appropriate measure of association. Use all this information to interpret the relationship between your two variables.

## **Part V – Bringing another Variable into the Analysis**

There are other variables that are related to both confidence in the executive branch and spending on the environment. One of those variables is political party identification (PARTYID). Run a frequency distribution for this variable to see how it is coded.

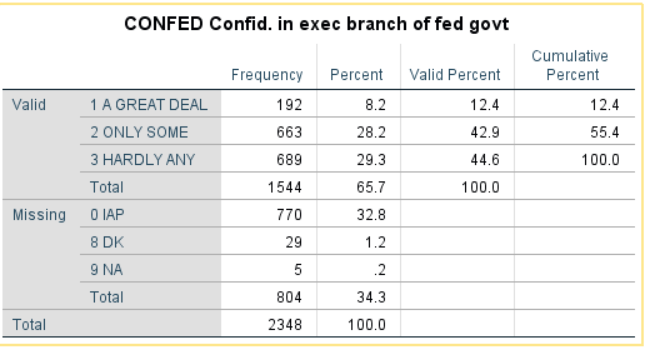


We don’t want to use all seven categories for this variable so let’s recode PARTYID by combining categories in the following way.

* Combine strong Democrat (value 0), not strong Democrat (value 1), and Independents, near Democrat (value 2) into once category. Give this category a value of 1 and call it “Democrat.”
* Leave Independent (value 3) as a separate category. Give this category a value of 2 and call it “Independent.”
* Combine Independent, near Republican (value 4), not strong Republican (value 5) and strong Republican (value 6) into another category. Give this category a value of 3 and call it “Republican.”

There are other ways we could combine the categories but let’s use this method. We’re going to use recoding into a different variable in SPSS. Give this new variable the name of PARTYID1. To make your output easier to read, add variable and value labels for your new recoded variable. If you don’t know how to recode into another variable or how to add variable and value labels, your instructor will show you how. (See also Chapter 3 in the online SPSS tutorial.)

There is another variable that we also want to recode. Run a frequency distribution for CONFED.

This sh

Notice that there aren’t many respondents who have a great deal of confidence in the executive branch (i.e. only 12.4%). This is going to cause a problem later on. We’ll explain why later in this exercise. So let’s recode it now in the following way.

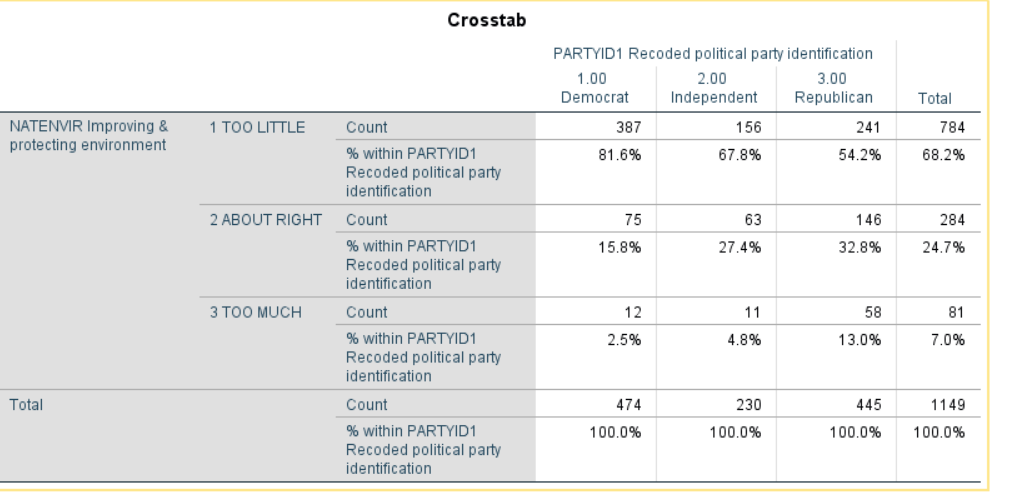
* Combine a great deal (value 1) and only some (value 2) into one category. Give this category a value of 1 and call it “some or more.”
* Leave hardly any (value 3) as a separate category. Give this category a value of 2 and call it “hardly any.”

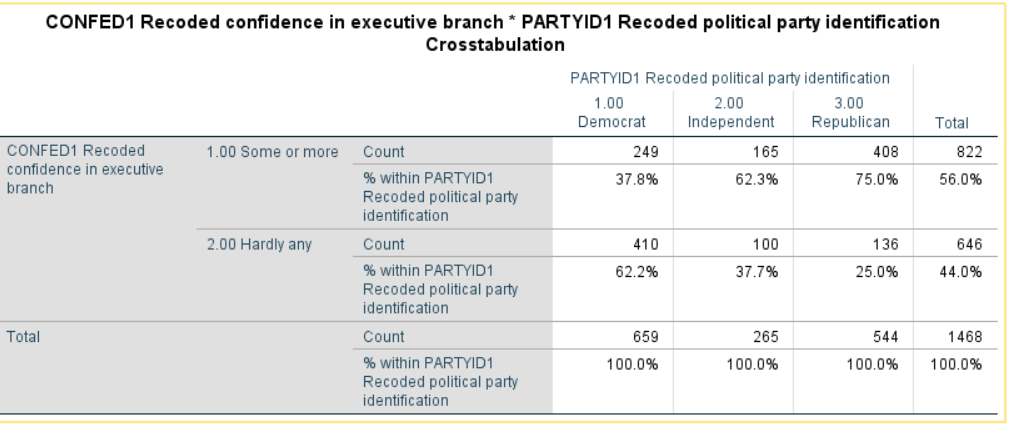
Give this new variable the name of CONFED1. To make your output easier to read, add variable and value labels for your new recoded variable. If you don’t know how to recode into another variable or how to add variable and value labels, your instructor will show you how. (See also Chapter 3 in the online SPSS tutorial.)

Now let’s make sure that party identification is related to both CONFED1 and NATENVIR. Tell SPSS to get the crosstabulation of PARTYID1 with each of these variables. This will require that you run two crosstabulations.  Use PARTYID1 as your independent variable.

* PARTYID1 and CONFED1
* PARTYID1 and NATENVIR

Write a paragraph describing the relationship between PARTYID1 and these two variables. Be sure to tell SPSS to compute Chi Square and the appropriate measure of association. Use all this information to interpret these two crosstabulations. Here are the tables that you should get from SPSS.[[4]](#footnote-4)





It’s clear from these tables that political party identification is related to both confidence and spending priorities for the environment.

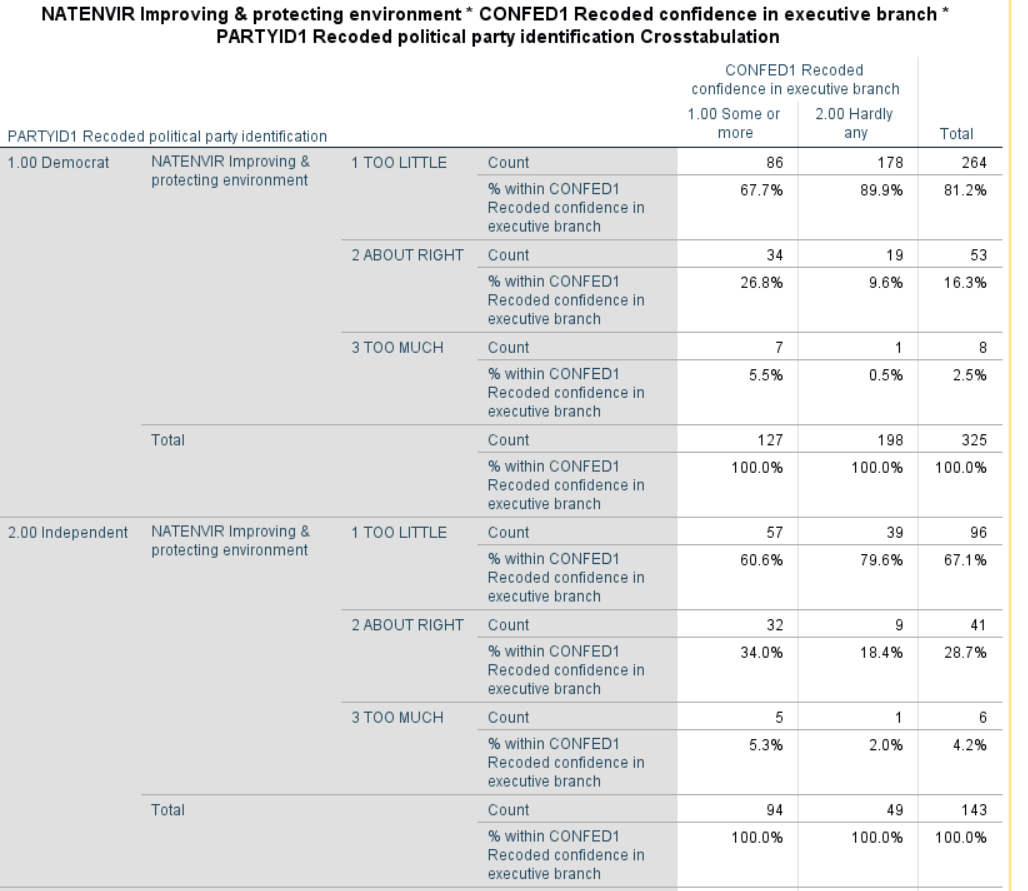
## **Part VI – Now It’s Your Turn Again**

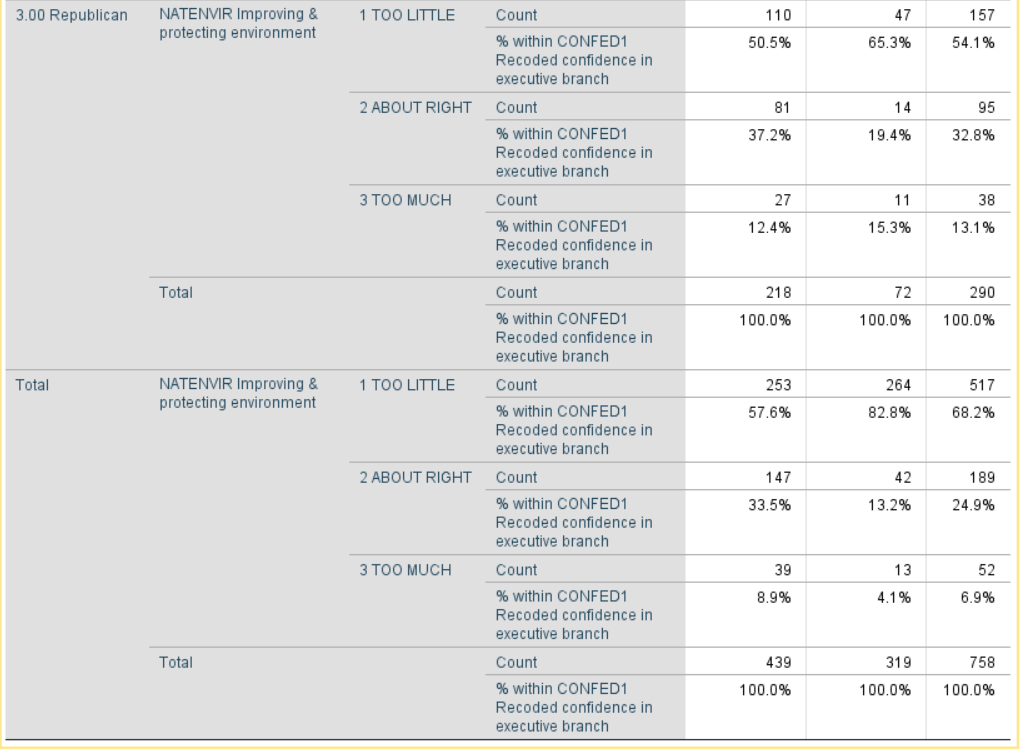
Now repeat the analysis that we carried out in Part V. Tell SPSS to get the crosstabulation of PARTYID1 and NATEDUC and write a paragraph describing the relationship between your two variables. Be sure to tell SPSS to compute Chi Square and an appropriate measure of association. Use all this information to interpret the relationship between your two variables.

## **Part VII – Bringing Party Identification into the Analysis**

Now that we know that political party identification is related to both confidence in the executive branch and spending priorities that raises another possibility. Perhaps the reason that those who have more confidence in the executive branch are less likely to feel that we’re spending too little on the environment is that Republicans have more confidence in the executive branch and are less inclined to think that we’re spending too little on the environment than do Democrats.

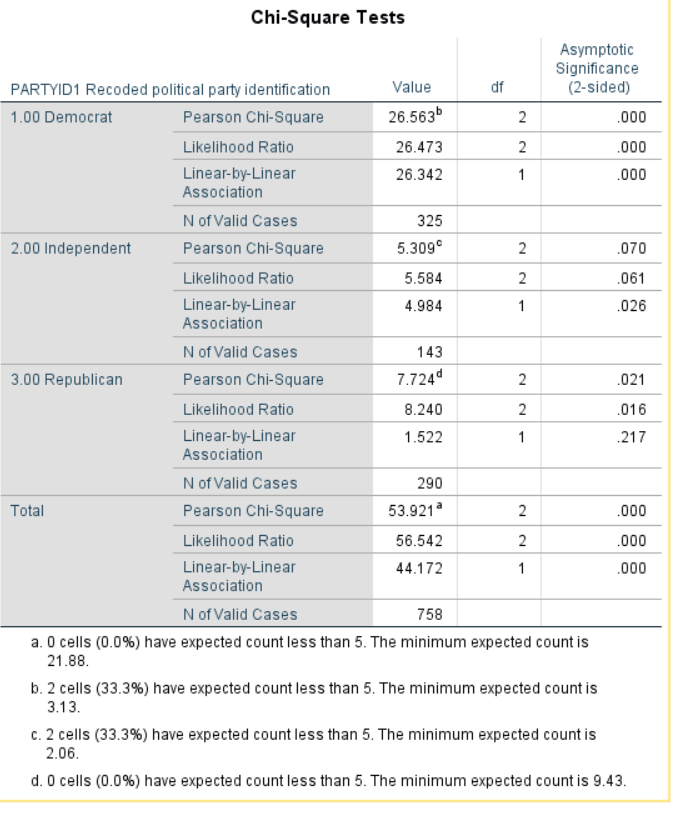
We can check on this by rerunning the crosstabulation of confidence in the executive branch and spending on the environment controlling for party identification. In other words, we’re going to separate our sample into three groups – Democrats, Independents, and Republicans – and look at the relationship between CONFED1 and NATENVIR separately for each of these three groups. Here’s what our tables will look like.

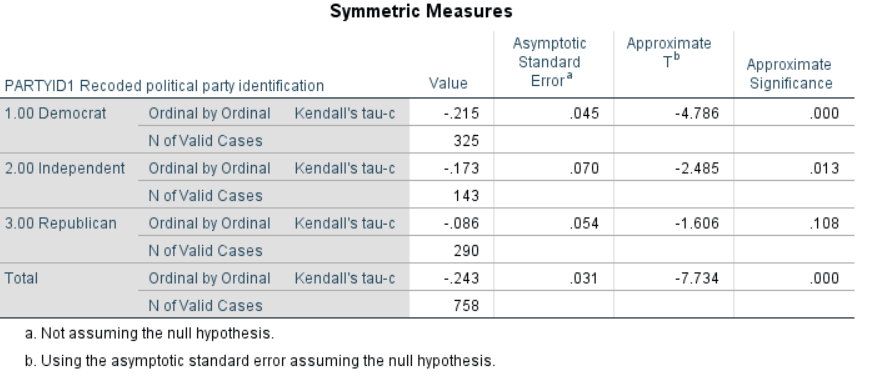




Note that we have four tables – one each for Democrats, Independents, and Republicans – and one for all three combined which SPSS calls total.

Here are the Chi Square and tau-c tables.[[5]](#footnote-5)





Note that again we have four tables -- one each for Democrats, Independents, and Republicans – and another for all three combined which SPSS calls total.

If the relationship between confidence and spending priorities had been due to party identification, then the relationship would have disappeared or decreased markedly in each of the first three tables.[[6]](#footnote-6) But what happened was that it decreased for Republicans but not for Independents and Democrats. In other words, we discovered that the relationship exists primarily for Democrats and Independents but not for Republicans. This is often referred to as specification.

## **Part VIII – Now It’s Your Turn Again**

Now repeat this analysis that we carried out in Part VII but this time use NATEDUC as your dependent variable. Tell SPSS to get the crosstabulation of CONFED1 and NATEDUC controlling for PARTYID1 and write a paragraph describing the relationship between these two variables when you control for political party identification. Be sure to tell SPSS to compute Chi Square and an appropriate measure of association. Use all this information to interpret the relationship between your two variables paying particular attention to what happened when you controlled for political party identification.

## **Part IX – Summary**

Write a paragraph indicating what you learned about the relationship between confidence in the executive branch and spending priorities for the environment and education. Be as specific as possible.

# **Appendix**

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

The General Social Survey (GSS) is a large, national probability sample of adults in the United States. It began in 1972 and continued on an almost yearly basis until 1996. In 1996, the GSS became a biannual survey and the sample size increased. Many questions are asked on each survey, while other questions are rotated from survey to survey. This subset from the 2018 GSS includes all the cases (2,348) and 88 variables. This data set has already been weighted using the weight variable supplied by the GSS (WTSS). Some of the original GSS variables were recoded and a few new variables created. Some of the new variables have names similar to those in the original GSS data set. The data set was created to accompany these exercises and is named GSS18CONFIDENCE.SAV.

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

ABANY Abortion if woman wants for any reason

ABDEFECT Abortion if strong chance of serious defect

ABHLTH Abortion if woman's health seriously endangered

ABNOMORE Abortion if married and wants no more children

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

ABRAPE Abortion if pregnant as result of rape

ABSINGLE Abortion if not married

AGE Age of respondent

ATTEND How often respondent attends religious services

CAPPUN Favor or oppose death penalty for murder

CHILDS Number of children

CLASS Subjective class identification

CONARMY Confidence in the military

CONBUS Confidence in major companies

CONCLERG Confidence in organized religion

CONEDUC Confidence in education

CONFED Confidence in executive branch of federal government

CONFINAN Confidence in banks and financial institutions

CONJUDGE Confidence in United States Supreme Court

CONLABOR Confidence in organized labor

CONLEGIS Confidence in Congress

CONMEDIC Confidence in medicine

CONPRESS Confidence in the press

CONSCI Confidence in the scientific community

CONTV Confidence in television

DEGREE Respondent's highest degree

EDUC Highest year of school completed

FAIR People fair or try to take advantage

FEAR Afraid to walk at night in neighborhood

FINRELA Opinion of family income

GRASS Should marijuana be made legal?

GUNLAW Favor or oppose gun permits

ID Respondent’s identification (id) number

INCOME16 Total family income (2017)

MADEG Mother's highest degree

MAEDUC Highest year school completed, mother

MARITAL Marital status

NATAID Spending – foreign aid

NATAIDY Spending – assistance to other countries

NATARMS Spending – military, armaments, defense

NATARMSY Spending – national defense

NATCITY Spending – solving problems of big cities

NATCITYY Spending – assistance to big cities

NATCRIME Spending – halting rising crime rate

NATCRIMY Spending – law enforcement

NATDRUG Spending – dealing with drug addiction

NATDRUGY Spending – drug rehabilitation

NATEDUC Spending – improving the nation's educational system

NATEDUCY Spending – education

NATENVIR Spending – improving and protecting the environment

NATENVIY Spending – the environment

NATFARE Spending – welfare

NATFAREY Spending – assistance to the poor

NATHEAL Spending – improving and protecting nation's health

NATHEALY Spending – health

NATRACE Spending – improving the conditions of blacks

NATRACEY Spending – assistance to blacks

NATSPAC Spending – space exploration program

NATSPACY Spending – space exploration

OWNGUN Have gun in home?

PADEG Father's highest degree

PAEDUC Highest year school completed, father

PARTYID Political Party Affiliation

POLVIEWS Think of self as liberal or conservative

PORNLAW Feelings about pornography laws

PRAY How often does respondent pray?

PRAYER Support Supreme Court Decision on prayer in public schools

PRES12 Vote for Romney or Obama in 2012

PRES16 Vote for Clinton or Trump in 2016

RACE Race of respondent

REGION Region of interview

RELIG Respondent's religious preference

RELITEN Strength of religious affiliation

RELPERSN Respondent considers self a religious person

SATFIN Satisfaction with financial situation

SEX Respondent's sex

SIZE Size of place respondent lives in thousands

SPDEG Spouse's highest degree

SPEDUC Highest year school completed, spouse

TRUST Can people be trusted?

VOTE12 Did respondent vote in 2012 election?

VOTE16 Did respondent vote in 2016 election?

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

YEAR Year of survey (2018 for all respondents)

ZZPARTYID1 Recoded political party affiliation  
ZZCONFED1 Recoded confidence in executive branch of federal government

1. Variable names are in all capitals. [↑](#footnote-ref-1)
2. Variable names are in all capitals. [↑](#footnote-ref-2)
3. The variable names are in all capitals. [↑](#footnote-ref-3)
4. Note that I have omitted the Chi Square and tau-c tables for sake of brevity. [↑](#footnote-ref-4)
5. Now we can explain why we recoded CONFED.  Note that at the bottom of the Chi Square table, SPSS tells you the minimum expected frequency for each table.  Chi Square assumes that this minimum expected frequency is at least 5 for each table.  If we had not recoded CONFED, we would have violated this assumption. [↑](#footnote-ref-5)
6. This is often referred to as spuriousness. [↑](#footnote-ref-6)