

## Tutorial for SPSS for Windows

Jean McSween

Geospatial and Statistical Data Center

University of Virginia Library

This document provides a tutorial for beginners to SPSS for Windows using the basic SPSS skills discussed in “Getting Started with SPSS for Windows.” This tutorial will use the data from the General Social Survey to explore an example of a research hypothesis.

### Section 1: Organizing and Using Data in SPSS

1. Data set
2. Research Hypothesis
3. Opening Data in SPSS
4. Coding of variables
5. Creating New Variables from Old Variables
6. Recoding Same Variable
7. Recoding into Different Variable
8. Missing Values
9. Sorting Cases
10. Selecting Cases
11. Splitting Data Files

### Section 2: Summary Statistics

1. Descriptive Statistics
2. Frequencies
3. Crosstabulations 2

### Section 1: Organizing and Using Data in SPSS

#### 1. Data set

This tutorial uses a limited data set from the General Social Survey (GSS). The data set and its codebook are available through the homepage of the Geostat Center under Interactive Data Resources. The GSS is one of thousands of quantitative studies accessible through the University of Virginia’s membership in the Inter-university Consortium for Political and Social Research (ICPSR) at <http://fisher.lib.virginia.edu/icpsr> .

The GSS is a cross-national survey conducted by the National Opinion Research Center (NORC) since 1972. The cumulative data set includes over 35,000 respondents with approximately 2,500 questions used in their surveys. In addition to a substantial core of questions, it includes modules for more detailed study into areas such as social networks, sociopolitical participation, mental health, and gender.

#### 2. Research Hypothesis

For this tutorial, we will use a hypothesis regarding gender to explore the different tools of SPSS. Our hypothesis states that gender affects one’s political party affiliation. Men are more likely to identify with the Republican Party and women are more likely to identify with the Democratic Party in the 1996.

#### 3. Opening the Data in SPSS

To start SPSS in the Geostat Lab, go to the Statistics Utilities folder and double click on the SPSS icon. You can also start SPSS by double-clicking on any SPSS file.

To open data file:

File -> Open -> Data...

The file name for this tutorial is found on J:\SPSS tutorial\Tutorial for SPSS.doc. This will bring your data into the Data Editor screen.

#### 4. Learning More about Your Data

It is always a good idea to begin by familiarizing yourself with your data set. As you can see, the data set for this tutorial contains only a limited GSS file, which includes 10 variables plus respondent identification numbers and the year of the study. For this section, we will use the variable for political ideology (POLVIEWS) for practice. 3

a. At the bottom left-hand side of your screen, you will see Data View and Variable View. Switch from Data View with the spreadsheet of the data to Variable View by clicking Variable View using your mouse. You will see brief descriptions of the variables in the study.

b. Or click View -> Variables using the top menu. SPSS will provide you with very brief descriptions of the variables.

c. For more detailed information about each variable's values, click on the box under the column for Values that corresponds with the variable of interest. In our case, find the column Values and the row for POLVIEWS. Click on the box. This will highlight the box, which includes variable label {0,NAP}.

d. Click the small gray box for the information. A screen labeled Value Labels will appear. For POLVIEWS, we find that 0="NAP," 1="Extremely Liberal," etc.

e. You should also refer to the GSS codebook at <http://www.icpsr.umich.edu/GSS99/codebook.htm> or through the Geostat Center homepage for greater detail about the survey question and additional data available through the GSS.

Practice:

Using the information above, examine the coding for the variables: SEX, MARITAL, EDUC, INCOME91, POLVIEWS and PARTYID.

a. What do 1 and 2 represent for the variable SEX?

b. What categories does the GSS offer for their question about marital status (MARITAL)?

c. How many categories does GSS offer for PARTYID?

d. Which variables contain missing data in the form of "Don't Know," "NAP" or "Not Applicable," "NA" or "No Answer," "Refused," "Not Sure," or "Inappropriate"?

#### 5. Creating New Variable From Old Variable

As you may have noticed for the PARTYID variable, party identification includes intensity of support for the respondent's party affiliation ("strong" or "not very strong"). Our hypothesis examines simple party identification. Therefore, we want to create a new variable called PID, which we will later recode to remove intensity of party affiliation while preserving our original variable PARTYID. The first step in the process is the creation of a new variable from the old variable.

Create new variable without conditional statement:

The new variable will contain exactly the same information as the old variable.

a. Return to Data Editor screen in Data View.

b. Using the top menu bar, click Transform -> Compute. This will take you to a screen labeled Compute Variable.

c. In the box for Target Variable, type PID1 for the name of the new variable.

d. Using the box below with the list of variables in the data set, highlight the variable for party identification then click on the run arrow (>). PARTYID will appear in the box for Numeric Expression.

e. Click OK at the bottom of the Compute Variable screen.

Create a new variable with conditional statement:

The new variable will contain the same information as the old variable if it was specially selected using the conditional statements.

a. Create a new variable named MWDS from MARITAL using steps b-d. Click on If instead of OK. This will bring up a screen for Compute Variable: If Cases.

Make sure you delete the previous information from PID1.

b. Click on Include if Cases Satisfies Condition. 5

c. Highlight the variable for marital status then click on the run arrow. MARITAL will appear in the box. Using the symbols and numbers below, click <=5. This will create a new variable without the category for "9=Don't Know."

d. Click Continue.

[Option: If you would like to save your recoding into syntax, click Paste. The syntax function allows you to save your work with coding and manipulating data as a reminder of your coding scheme and a time saver for later work on the data.

To execute command in syntax, highlight commands that you wish to run. Then click the run command button on the top menu (>).]

e. OK on the next screen.

Practice:

Create a new variable (INCOME) from the old variable (INCOME91) excluding all categories that might be considered missing values and skew our results.

[Hint: Use conditional statement in recreating new variable.]

## 6. Recoding Same Variable

Now that you have mastered the art of creating new variables, you need to recode them to address our research question about gender and party identification. For party identification, we will make the assumptions that independents who lean toward a political party are similar to weak party identifiers, and people who support other parties are independents.

a. Click Transform -> Recode -> Into Same Variables...

b. Highlight the new party identification variable (PID1) then click on the run arrow. This will move PID1 into the box for Numeric Variables.

c. Click Old and New Values. This will take you into the screen for Recode into Same Variables: Old and New Values.

d. On the left side of the screen, the box is labeled Old Value. In the box labeled Value, type the number of the value you wish to change. In the box labeled New Value, type the new value you want to assign it. Click Add after each old value has been assigned a new value. The change will appear in the box Old->New.

Make the following changes:

0 -> 1

1 -> 1

2 -> 1

3 -> 2

7 -> 2

4 -> 3  
5 -> 3 6  
6 -> 3

e. Once you have recoded the variable, click Continue. You will return to the previous screen. Then click OK.

Practice:

Recode EDUC into fewer categories.

Recode the categories for “No formal schooling” (0) through “8th grade” (8) into 1.

Recode “9th

grade” (9) through “11th

grade” (11) into 2.

Recode “12th

grade” (12) into 3.

Recode “1 year of college” (13) through “3 years” (15) into 4.

Recode “4 years” (16) into 5.

Recode “5 years” (17) through “8 years” (20) into 6.

[Hint: In screen for Recode into Same Variables: Old and New Values, highlight values in box for Old -> New then use the Remove to delete the values assigned for PID1.]

#### 7. Recoding into Different Variable

There is one very useful shortcut to creating a new variable from an existing variable through recoding. We could simply omit the separate step of creating a variable. Let’s use the example of party identification again to create PID2, which should be identical to PID1.

a. Click Transform -> Recode -> Into Different Variables...

b. Highlight the party identification variable (PARTYID) then click on the run arrow (>). This will move PARTYID into the box for Input Variable ! Output Variable.

c. Under Output Variable, type PID2 in the box for name. For label, type Party Identification 2.

d. Click Change. 7

e. Click Old and New Values. This will take you into the screen for Recode into Same Variables: Old and New Values.

f. On the left side of the screen, the box is labeled Old Value. In the box labeled Value, type the number of the value you wish to change. In the box labeled New Value, type the number of the value you want to be assigned for the response.

Click Add after each old value has been assigned a new value.

Make the following changes:

0 -> 1

1 -> 1

2 -> 1

3 -> 2

7 -> 2

4 -> 3

5 -> 3

6 -> 3

g. Once you have recoded the variable, click Continue. You will return to the previous screen. Then click OK. PID1 and PID2 should have exactly the same values.

## 8. Missing Values

There remains another step for getting our data ready for analysis. As we saw in part 4 of the tutorial, some variables have missing data. To prevent these value labels from skewing our data analysis, it is necessary to purge the missing data. Let's use the variable for females staying home (FEHOME). This variable has five values assigned to responses:

0 = "NAP"

1 = "Agree"

2 = "Disagree"

8 = "Not Sure"

9 = "NA."

As you will see below, GSS has already coded 0 and 9 as missing values. However, the value for "Not Sure" (8) remains in your data. If we do not assign "Not Sure" as a missing value, our mean for FEHOME will be 2.05 with a standard deviation of 1.19. After assigning it as a missing value, the mean will be 1.84 with a standard deviation of .37. This represents a sizable difference, particularly for our standard deviations.

a. At the bottom left-hand side of the page, click on Variable View.

b. Find the column for Missing. Move down the column until you find the box for FEHOME. Click on the box. 8

c. Click on the small gray box to the right in the box. This will bring up the Missing Values screen where you will see two numbers in the boxes for Discrete missing values. You will the numbers 0 and 9 already appear in the boxes for Discrete missing values.

d. In the empty box, type 8 to add the value as a missing value.

e. Click OK to save the change.

Practice:

a. Assign the same change in missing values to the variable FEPOL.

b. Assign 98 as a missing value for EDUC.

c. Assign 8 as a missing value for POLVIEWS.

## 9. Sorting Cases

Once we have manipulated your data for analysis, it is sometimes necessary to organize and sort it. Sorting cases allows us to organize our data in ascending or descending order on the basis of one or more variables. For our research, we want to sort SEX in ascending order with males (1) first.

a. Return to the Data View screen for the Data Editor.

b. Click Data -> Sort Cases...

c. Highlight SEX then click on run arrow (>) to move variable into the box for Sort by... Below you will notice that it is already set for ascending order in the box for

Sort Order.

d. Click OK. 9

If you examine your data now, you will find that the ID order is no longer perfect ascending order. Rather the data is organized by SEX with all the 1's listed first.

Practice:

Change the sorting for SEX to descending order with the female (2) respondents appearing first.

[Hint: You will need to click Reset in the screen for Sort Cases.]

## 10. Selecting Cases

Selecting cases is a useful feature if we are interested in a specific subset of your data. In exploring our hypothesis about gender and party identification, we may want to more closely examine respondents who agreed with the statement, "Most men are better suited emotionally for politics than are most women."

a. Click Data -> Select Cases.

b. In the box labeled Select, click If condition is satisfied then If. This will bring you to another screen for Select Cases: If.

c. Highlight the variable FEPOL, then click on the run arrow (>) to move FEPOL into the empty box.

d. Type or click =1 after FEPOL to indicate that you wish to select all respondents who agreed that women were not emotionally suited from politics compared to men.

e. Click Continue to return to the previous screen. 10

f. Notice that there are two options under Unselected Cases Are – Filtered and Deleted. Make sure that Filtered is marked. The delete function will remove all cases that do not satisfy your conditions. The filter function will mark all nonselected cases without deleting them.

g. Click OK. Notice how the filtered cases are now marked.

h. To restore data, click Data -> Select Cases. Click All Cases under Select then click OK.

Practice:

Filter out (not delete) all respondents who disagreed (2) with the statement: "Women should take care of running their homes and leave running the country up to men" (FEHOME). Then restore filtered data.

## 11. Splitting Data Files

Using the commands for selecting cases, we were able to examine particular subsets. Splitting data files permits us to retain distinct subsets while examining each separately. In the case of our research design, we want to examine the genders separately for the same types of information.

a. Click Data -> Split File...

b. A screen will appear for Split File. The default setting for this screen is Analyze all cases, do not create separate groups. Either of the two options below – Compare groups and Organize output by groups – will split the data into groups. The difference between the functions is limited to presentation in data output.

c. Click Compare groups.

d. Highlight the variable SEX then click on the run arrow to move the variable into the box for Groups Based on.

e. Click OK. 11

If we were to look at the descriptive information for income, the data output would split the data for men and women as shown below. The mean income for men is more than the mean income for women.

Restore default:

- a. Click Data -> Split File...
- b. Click Analyze all cases, do not create groups then click OK.

## Section 2: Summary Statistics

### 1. Descriptives

Great job manipulating and organizing your data! Now we come to data analysis where all the previous work begins to pay off. The descriptive function will provide us with an overview of particular variables in our data set such as sample size, mean, and standard deviation.

- a. Click Analyze -> Descriptive Statistics -> Descriptives...
- b. Highlight the following variables: SEX, EDUC, FEHOME, FEPOL, PID1, and INCOME. This will move the variables into the box labeled Variables.
- c. Click Options to view all potential statistical information available through Descriptives. Notice that Mean, Std. Deviation, Minimum, and Maximum are all marked. 12
- d. Click Continue on the screen for Descriptives Options, then click OK.
- e. The descriptive analysis will appear on the Output screen.
- f. Minimize or click the Data Editor screen to return to your data.

Practice:

In our descriptive analysis, we did not examine political ideologies (POLVIEWS). Analyze the descriptive information for POLVIEWS alone.

[Hint: Remove the previous variables for analysis from the box labeled Variables by highlighting the variables then click the run arrow (<) for this assignment.]

### 2. Frequencies

Whereas the descriptive data provides us with an excellent overview of particular variables, frequencies provide us with detailed information about each value assigned to particular variables. This function allows us to get a better idea about the distribution of the values for each variable. For example, your descriptive analysis provided us with a mean for political ideology (4.19). Frequencies will provide us with the percentage of respondents who describe themselves as extremely liberal or extremely conservative.

Without charts:

- a. Click Analyze -> Descriptive Statistics -> Frequencies...
- b. Once in the screen for Frequencies, highlight the variable POLVIEWS then click the run arrow.
- c. If we click OK, our frequencies will appear for this variable. 13

With charts:

- a. Click Analyze -> Descriptive Statistics -> Frequencies...
- b. POLVIEWS already appears in the box for Variables.
- c. Click Charts, a button at the bottom of the screen for Frequencies: Charts.
- d. In the box for Chart Type, click Histograms then click on the box by With normal curves.
- e. Click Continue to return to the previous screen.

f. Click OK. In addition to the frequency information seen above, your output will include a histogram with a normal curve for the data. 14

Practice:

Analyze the frequency for SEX. Include a pie chart.

#### 4. Crosstabulation

Crosstabulation takes us to the next step in our analysis. It permits us to compare frequency information for the intersection between two variables. Thus, we can now compare the frequencies of men's party identification to women's party identification.

a. Click Analyze -> Descriptive Statistics -> Crosstabs...

b. A screen for Crosstabs will appear. Highlight the independent variable (SEX) then click on the run arrow for Row(s). Highlight the dependent variable (PID1)

then click on the run arrow for Column(s).

c. Click Cells... at the bottom center of the screen. This will take you to the screen Crosstabs: Cell Display.

d. Under Percentages, click on Row then click Continue to return to the previous screen.

e. Click OK. 15

Our hypothesis stated that gender affects one's political party affiliation. Men are more likely to identify with the Republican Party and women are more likely to identify with the Democratic Party in the 1996. Our crosstabulation results suggest that our hypothesis may be correct. We find that the percent of women identifying with the Democratic Party (50.9 %) is over ten percentage points higher than the percentage of men (39.8%) who identified with the Democratic Party. For Republican Party identification, we find that men identify with the party (42.9%) more than women (31.0%). Men and women identify with independent parties to virtually the same degree.

Congratulations! You have completed the tutorial. 16

Tutorial for SPSS for Windows - Answers

Jean McSween

Geospatial and Statistical Data Center

University of Virginia Library

Section 1: Organizing and Using Data in SPSS

#### 4. Learning More about Your Data

a. 1=Male and 2=Female

b. 6 categories: 1=Married, 2=Widowed, 3=Divorced, 4=Separated, 5=Never married, and 9=No answer

c. 10 categories: 0=Strong Democrat, 1=Not very strong Democrat, 2=Independent close to Democrat, 3=Independent (Neither, No response), 4=Independent close to Republican, 5=Not very strong Republican, 6=Strong Republican, 7=Other party refused to say, 8=Don't know, 9=No answer

d. All of the variables except ID, YEAR, and SEX.

5.

The dot (.) in place of the data indicates missing values. 17

6. Your data should now look like this:

8. The box for missing values should now read 0,9,8. 18

9. All the female respondents should now appear first.

10. Your data should now have slash marks on the far left-hand side indicating filter

cases. 19

## Section 2: Summary Statistics

1. This table should appear in your output file.
2. Frequencies for SEX with pie chart.