

# POL 199

## Basic Overview of Working with Data in SPSS

### FAMILIARIZING YOURSELF WITH SPSS

#### Types of Files:

Because SPSS is a multi-function program that allows you to enter data, play around with it, and view it in a variety of different contexts, there are three different types of files in SPSS (as opposed to a program like Microsoft Word, for instance, where all the files are basically Word documents). The different types of files are:

- **Data:** This is the actual database that stores your data (files are “.sav” files). It is similar to Excel. In the bottom left-hand corner of your screen, you will see two tabs—one for “Data View” and one for “Variable View”. You can click on either one to see your dataset in two different ways.
  - **Data View:** In this view, you see your dataset in the form of a spreadsheet, just like an Excel table. Here, you can see all the actual values of the variables you have.
  - **Variable View:** In this view, you see your dataset as a list of variables. In other words, instead of seeing the actual values of the variables you have, you see only what the variables are and how they are labeled.
- **Syntax:** This is a way for you to store your work (files are “.sys” files). It keeps a log of all the different commands you have used, and what you have done to the data. As you work with data, you will find that it is often hard to remember what you did yesterday, the week before, or even half an hour ago. A syntax file allows you to keep a record of this. It also allows you to repeat certain commands that you may use over and over again.
- **Output:** This allows you to see the results of your data manipulations (files are “.spo” files). If you ask SPSS to create a graph, the graph will appear here. If you ask it to create a table, the table will appear here.

To open a file of a particular type, go to:

**File → Open → [Data/Syntax/Output]**

Then an “Open File” dialog will pop-up on your screen. Navigate through it just as you would for any other type of program.

To create a new file of a particular type, go to:

**File → New → [Data/Syntax/Output]**

### Basic Organization of any data-file:

Your data is arranged like a spreadsheet, with rows and columns.

#### Rows = Cases

Each row is one case in your dataset. That means that each row should represent ONE of whatever your unit of analysis is. In other words, if your unit of analysis is individuals, each row should represent one person. Remember that sometimes your unit of analysis can be defined by two things—for instance, if your unit of analysis is one country in one year, each row should represent one country in one year.

#### Columns = Variables

Each column is one variable. So, by looking down one column, you can see all the different values that each variable takes.

To make sure you can keep track of your data, it is generally good practice to make the first few variables in your dataset be identifying variables. This means that these variables identify the cases in a unique way. So, if you have 57 survey respondents, it is generally a good idea to assign each respondent a unique ID # and make that the first variable in your dataset. If, on the other hand, you are looking at several countries over several years, then you should have a variable for the country, and another variable for the year. By looking at those two variables together, we can then identify the cases.

### **BASIC PRINCIPLES FOR WORKING WITH DATA**

*Principle #1: Get to know your data; including how it was collected.*

Whenever you start working with a new dataset, play around with it in the beginning to make sure you understand the basic structure of the data.

*Principle #2: Never lose data.*

Never, never, never lose data. If you are worried that you might save over something or ruin a dataset, then make sure you save a clean copy of the data for yourself. Then create a “working copy” where you can make changes, and manipulate the data. In case you ever make a mistake, you can always go back to your original copy.

*Principle #3: Create systems of dealing with data that you will remember.*

As you work with data, you will realize that it is hard to remember what everything is. The best way to handle this is to get into the habit of keeping careful records of your work. Syntax files (discussed above) can help with this. Another thing that is important is to create systems for yourself that make sense to you. This means systems of naming variables, labeling them, and saving files. In terms of naming variables, try to be as specific as you can in your variable names. One way to do this is to note the presence or absence of a quality/characteristic (e.g. don't use 'sex' use 'male' so that 1 = male & 0 = female).

### **ENTERING DATA IN SPSS (“LOGGING” DATA)**

There are two main steps to entering data.

1. First, create and properly label the new variable.
2. Second, enter the data.

### Creating New Variables

- First click on the Variable View tab on the lower left-hand corner of your screen. This view is where you enter in your new variables.
- Under Name enter in the name of your new variable. Be sure to use a naming convention that you will remember and that makes sense to you.
- *For the following columns, clicking on the box will cause a small gray box to appear on the right-hand side. You will need to click on the gray box to change the settings. After entering in the name and clicking on tab, the default settings for the other columns will pop up.*
- Under Type be sure that Numeric is selected if your data is numeric or ordinal, or if you will be using codes (e.g. 1 = male, 2 = female). If you want to be able to type in letters instead of numbers, you need to click on String, then click on Ok.
- Under Width be sure the column will be wide enough to enter in the given values (the default setting usually is wide enough so you can ignore this).
- Under Decimals be sure there is the right number of places past the decimal. If you want things to show up as whole numbers, change the Decimals to zero, if you want it to the one's digit, set it to one, etc.
- Under Labels enter in the entire name of the variable so that you know what it is. (e.g. type in "Type of schooling" instead of just "school"). Be descriptive and thorough in your labels so that you remember what the variable is.
- Values is where you enter in your codes. Click on the gray box. Next to Value type in the numerical code (e.g. 1). Next to Value Label type in what that number represents (e.g. Male). Then click Add. Once you have entered in all of the codes, click on Ok.
- You don't really need to do anything with the next three columns (Missing, Columns, and Align).
- For Measure make sure the correct level of measurement is selected (usually the default setting will make it correct).

### Entering the Data

- When all of the variables have been entered in Variable View, click on the Data View tab in the lower left-hand corner of your screen.
- Your new variables will appear horizontally, across the top of the data sheet. (Note that if you roll your mouse over the variable name at the top of the screen in Data View, then the Variable label should appear).
- Enter the numerical values in their appropriate places.

## **SORTING DATA**

Sometimes you will want to look at the data in the spreadsheet to understand it better. To do this, do:

## **Data → Sort cases...**

In the column on the left-hand side, click on the variable(s) you want to sort by, and use the arrow to move them to the right hand side. If you are sorting by more than one variable, you need to move them into the right-hand side in the order you want them to sort.

## **RECODING AN EXISTING VARIABLE**

You will sometimes want to recode a variable so that it is in the direction you want. Or maybe you'll want to take a nominal variable with multiple categories and make it a dummy variable (i.e., let's say you have race and there are 7 categories. You might just want to know whether someone is White or not). The steps for doing this are:

### **Transform → Recode → Into Different Variable**

- A dialog box will appear on your screen.
- In the box on the left, highlight the variable that you want to recode and use the arrow to place it in the "Input Variable" box on the top right.
- Move your cursor to the "name" box and give the new variable that you want to create a name (if you were creating a dummy variable for White, for example, the new name might be white). Then move your cursor to the "Label" box. This lets you describe what the variable is (just like the "label" box in the Variable View file). You might type "white dummy variable."
- Now click "old and new values." For every current value assigned to the variable, you need to assign a new value. Type the old value into the "Value: old value" box on the left hand side of the screen. Now click on "New value: value" and enter that you want the new value to be. (If the original value was 1, for example, and it was for Asians, then your new value would be 0, since everyone who is not white will score a 0 and everyone who is white will score a 1).
- Then click "Add."
- Do this for all of your values and then click "continue," then "change," then "OK."
- Your new recoded variable should now show up on your spreadsheet in the last column.
- To make sure that you keep good track of how you coded the new variable, click on "Variable View" at the bottom of your spreadsheet, scroll down to the last row (where you should find your variable), and type in the new values into the "values" column.

You can now use the new variable to analyze data.

## **CREATING A NEW VARIABLE FROM EXISTING VARIABLES**

Go to the menu bar at the top of the screen and choose "Transform," then "Compute." The new variable you want should be named in the "Target Variable" box. Then choose from your existing variables to create a new function that will be what the target variable is (for example, you might want to add several different variables to make an additive scale). Click OK.

**EXAMPLE**

ID#	Pres. Vote 2000	Pres. Vote 1996	Abortion views
1	Gore	Clinton	Never
2	Gore	Dole	OK except 3rd tri.
3	Bush	Clinton	Rape/Incest Only
4	Gore	Clinton	Always
5	Gore	Dole	Rape/Incest Only
6	Bush	Dole	Always
7	Gore	Clinton	Rape/Incest Only
8	Gore	Clinton	Never
9	Gore	Clinton	Rape/Incest Only
10	Gore	Dole	Always
11	Gore	Clinton	Rape/Incest Only
12	Gore	Clinton	OK except 3rd tri.
13	Bush	Clinton	Rape/Incest Only
14	Gore	Clinton	Always
15	Bush	Dole	Never
16	Bush	Dole	OK except 3rd tri.
17	Bush	Dole	Rape/Incest Only
18	Gore	Clinton	Always
19	Bush	Clinton	Rape/Incest Only
20	Gore	Dole	OK except 3rd tri.

+++++

Variable 69 Vote for President 2000? (presv00)

Who did you vote for?

- 1. AL GORE
- 3. GEORGE W. BUSH
- 5. PAT BUCHANAN
- 6. RALPH NADER
- 7. OTHER (SPECIFY)
- 8. DK
- 9. RF
- 0. NA; INAP, 5, 8, 9, 0 in C5

	0	1	3	5
	-----	-----	-----	-----
Count	629	590	530	3
	6	7	8	9
	-----	-----	-----	-----
Count	33	9	3	10

+++++

Variable 6 Presidential Choice in 1996(presv96)

IF R VOTED FOR PRESIDENT IN 1996:

Which one did you vote for?

[DO NOT PROBE 'DK' RESPONSE]

- 1. BILL CLINTON
- 3. BOB DOLE
- 5. ROSS PEROT
- 7. OTHER (SPECIFY)
- 8. DON'T KNOW
- 9. RF
- 0. NA; INAP, 0,5,8,9 in A3

	0	1	3	5
	-----	-----	-----	-----
Count	543	665	408	138
	7	8	9	
	-----	-----	-----	
Count	27	9	17	

+++++

Variable 55 Opinion About Abortion (abortion)

There has been some discussion about abortion during recent years. Which one of the opinions on this page best agrees with your view? You can just tell me the number of the opinion you choose.

- 1. BY LAW, ABORTION SHOULD NEVER BE PERMITTED.
- 2. THE LAW SHOULD PERMIT ABORTION ONLY IN CASE OF RAPE, INCEST, OR WHEN THE WOMAN'S LIFE IS IN DANGER.
- 3. THE LAW SHOULD PERMIT ABORTION FOR REASONS OTHER THAN RAPE, INCEST, OR DANGER TO THE WOMAN'S LIFE, BUT ONLY AFTER THE NEED FOR THE ABORTION HAS BEEN CLEARLY ESTABLISHED.
- 4. BY LAW, A WOMAN SHOULD ALWAYS BE ABLE TO OBTAIN AN ABORTION AS A MATTER OF PERSONAL CHOICE.
- 7. OTHER (SPECIFY) [Volunteered by respondent]
- 8. DK
- 9. RF
- 0. NA

	0	1	2	3
	-----	-----	-----	-----
Count	3	215	525	265
	4	7	8	9
	-----	-----	-----	-----
Count	753	28	11	7

+++++