Introduction to SPSS

Chapter 3

Agron Kaci

John Jay College
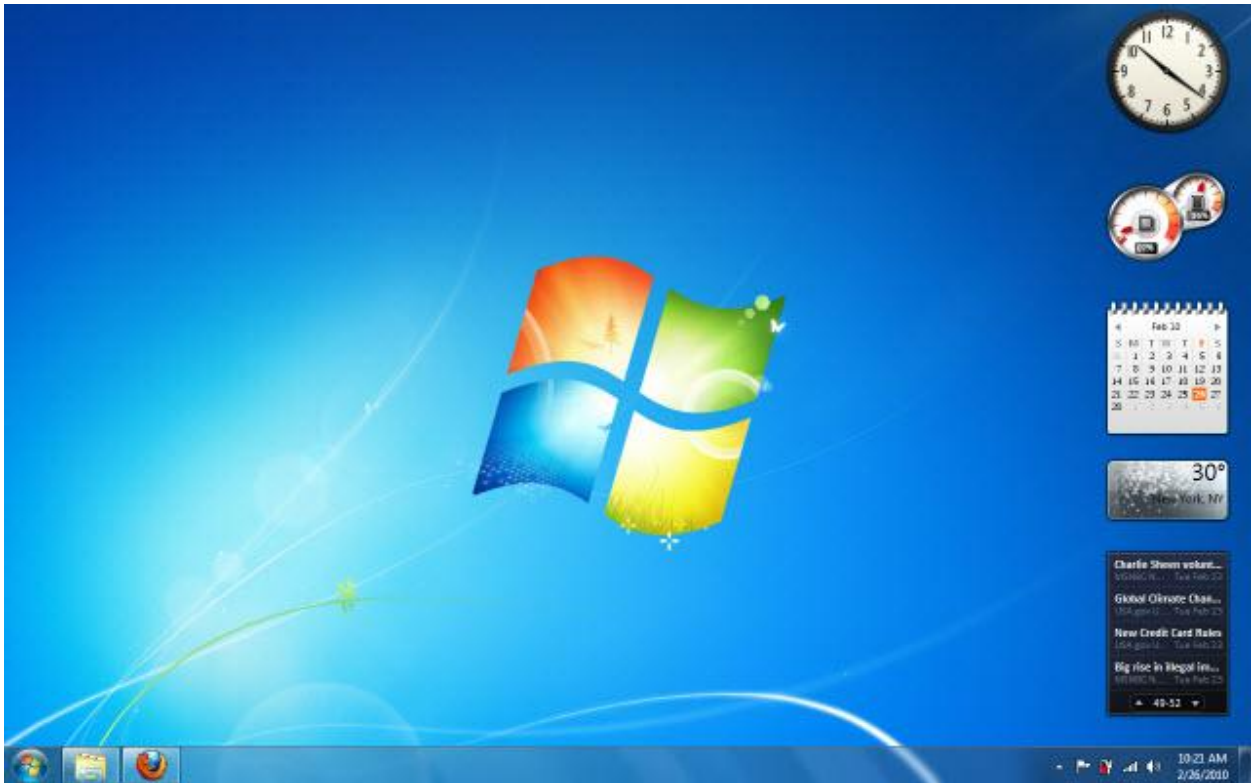
Introduction to SPSS

Using Windows

Figure 1-1

Figure 1-1 above shows the opening screen or the desktop of a computer with Windows 7 operating system.

To launch IBM SPSS software, let's start by clicking the "Start" button. The Start Menu will appear, and your screen should look something like Figure 1-2 below. This is the basic Start Menu, and it can also be customized by adding your most often used programs to it. If you are operating with Windows XP or Vista, there would not be many differences. In this menu, you will see All Programs, Documents, Control Panel, and Help and Support button. Notice that if a small triangle is present, that means there are more menus for that item. For example, the little black triangle on the right of the Microsoft Word menu will tell you that there are recent Word documents listed in the submenu, as soon as you move the mouse on that menu. This is a quick way to recall whatever you were working on recently, since one click will open that file and the applications program it belongs to (such as Word, PowerPoint, Excel, SPSS, etc.).
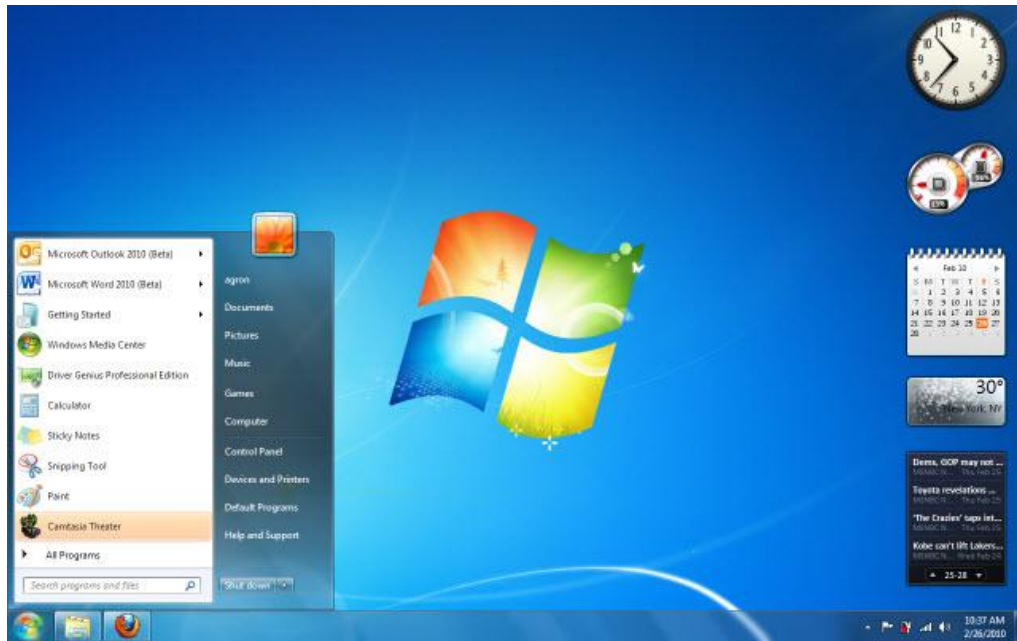
Figure 1-2

Starting IBM SPSS for Windows

The IBM SPSS (depicted hereof) for Windows icon should be found in *All Programs* menu. Double-click IBM SPSS Statistics folder and then click on the IBM SPSS Statistics 19 link. There is also a change in color from version 18, which was in red. Version 19 is shown in blue. Wait while IBM SPSS loads.

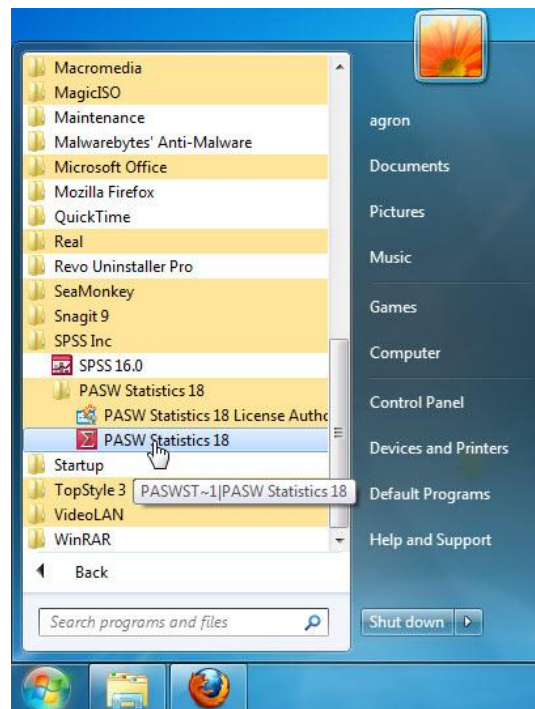*Note: In the figure below (and in all tutorials images), SPSS version 18 is depicted.*

Figure 1-3

Before IBM SPSS launches, it is customary to see a small dialog box that has the options to open an existing dataset or other types of files. Also, it allows you to type in your own data. For now, click "Cancel" to exit this box.

Figure 1-4

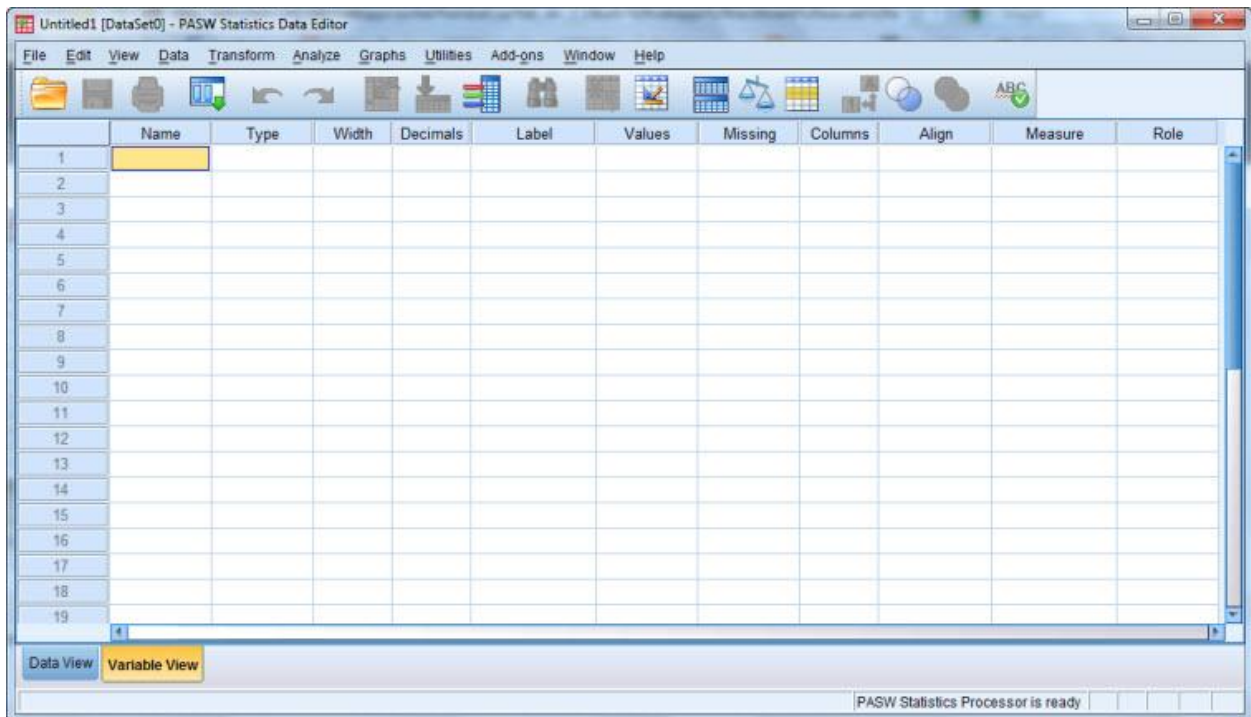Now the SPSS window opens as seen below:

Figure 1-5

The Taskbar shows an "IBM SPSS" button or icon .

Repeating basic computer literacy topics may be in order here. On the top-right corner, observe three small squares [   ]. The red cross, X, is used to close the window you are using. This will not close or exit the whole application. If more SPSS windows are open with different data sets, they will remain open.

The rectangle makes the window fill up the whole screen. Depending on your software options, SPSS may have been already full screen. In this case you will see two cascading rectangles [   ] in this box. You may want to play with the size of the window, so that it can make room for multi-tasking.

To the left of the other two boxes, you have the "Minimize" button, that looks like an underscore sign. This button will minimize a window into the taskbar. To retrieve this window again, just click on the taskbar button again.

Difference between Close and Exit

Close and Exit have a different meaning. Close will only close the window that is active, Exit will close the whole application. Let's look at the ways to close and/or exit SPSS.

1. Pull down the File menu and click "Exit". "Exit" will close the whole application, no matter how many data sets you have open.
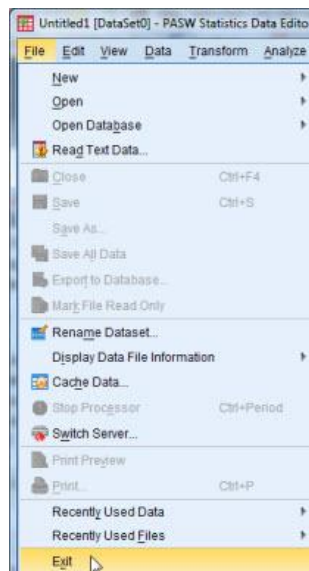

Figure 1-6

2. "Click and drag" is another technique that resembles the one described above. Click the "File" menu, *hold* the mouse button down and simultaneously *move* it down up to "Exit" and then release. IBM SPSS will now close.

3. By clicking the IBM SPSS icon in the top-left corner of the screen, to the left of the words "**Untitled: [Dataset0] - IBM SPSS Statistics Data Editor**", you'll see a short menu. Clicking Close will close the whole application. You can get the same result by double-clicking this icon. Of course, you have to click OK in the subsequent dialog box, where SPSS asks if you really want to close the application.
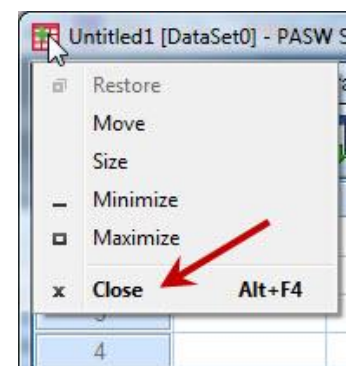
4. By clicking the X in the top-right corner of the screen, IBM SPSS application—not only the window you have open, but all—will close.

You can learn more about Windows 7, there are online resources that can be found **here>>>**.

Opening Data Files (datasets)

In class, we will use the 2004GSS.sav data file. You can download this data file from our course site link Datasets SPSS on the left navigation panel on Blackboard. Please refer to the podcast created earlier on how to open IBM SPSS files from Blackboard. This podcast is found in the Announcements area.

Let's launch IBM SPSS. Click on the "Start" button, go to the "IBM SPSS Statistics" folder, click on "IBM SPSS Statistics 19" and IBM SPSS will be launched.

Figure 1-7

You already saw how the screen looks like in Figure 1-5 above. On the title bar, you'll see the title "Untitled [Dataset0] - IBM SPSS Statistics Data Editor", because we haven't opened a dataset yet. You also will see the menu bar: File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, and Help.

Let's click on the "File" menu. In the File menu that pulls down, hover your mouse at "Open" link and then click "Data" in the submenu that appears on its right.

This opens the Open Data dialog box. This method of opening data sets from the Open→Data menu, assumes that you have saved the file locally in your computer. Now, while in the "Open data" dialog box, navigate to find the folder that contains the data file and double-click on the file name, 2004GSS. The data set now appears to have filled the columns and rows that were empty. The method that we

will use in class will be to open the file directly from Blackboard. When we start modifying the data files we would need to save them and then open them according the method described now: File-->Open... There is a keyboard shortcut for opening data files faster, and that is Ctrl + O.

Use any method to open the file. Let's talk about the new table of data in SPSS. The columns contain the variables (that used to be questions or concepts in a survey) and the rows contain the cases (or observations, which are the answers given by participants). The variable name appears on the header of each column. The variable name in earlier versions was limited to 8 characters; so many times they do not tell you much about the variable. To find out more about variable names click the "Utilities" menu in the menu bar. Select "Variables" and click on any variable on the left box. This procedure is depicted in the Figures below.
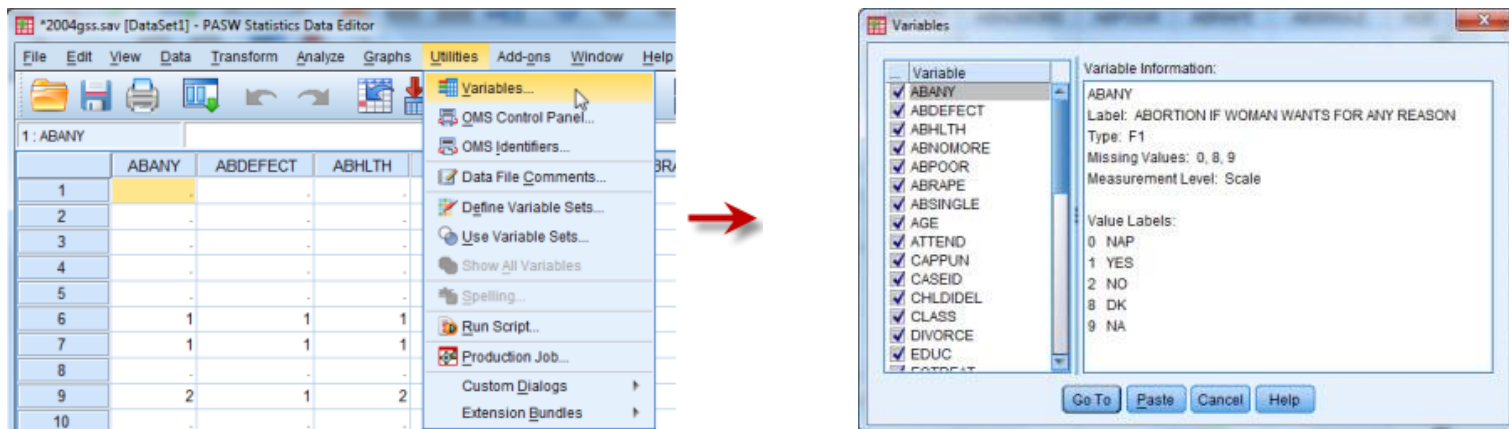

Figure 1-8

When you click on any of variables at left a short description of this variable, called label is shown in the right box. In addition to the label of the variable (long name of the variable), you will find the value labels, which are the attributes that make this variable. These are the possible answers to the questions in the survey. As you will see later, these values are very important!

A Brief Analysis of data with IBM SPSS

OK, so we know how to launch SPSS and how to open a data set in it. What else does SPSS serve?

In the Figure 1-8 above, we selected the variable name ABANY, which was labeled as "ABORTION IF WOMAN WANTS FOR ANY REASON". The possible answers are *YES*, *NO*, DK or "*Don't know*", NA or "*No answer*", and NAP or "*Not applicable*". (*Not applicable* is used for respondents who were randomly not asked the question). These value labels have values 1, 2, 8, 9, and 0 respectively.

If you recall from the first lecture, the most basic type of question in a research project should be to describe what is going on in the sample. Since you read the label and value labels of the variable ABANY, let's discover the percent of people that thought it should be legal for a woman to have an abortion for any reason.

In SPSS, click on the "Analyze" button in the menu bar. As the menu that pulls down, you see many statistical procedures collected in groups. If you hold the mouse over any collection, a submenu appears at the right side.

Follow the procedure as shown below:

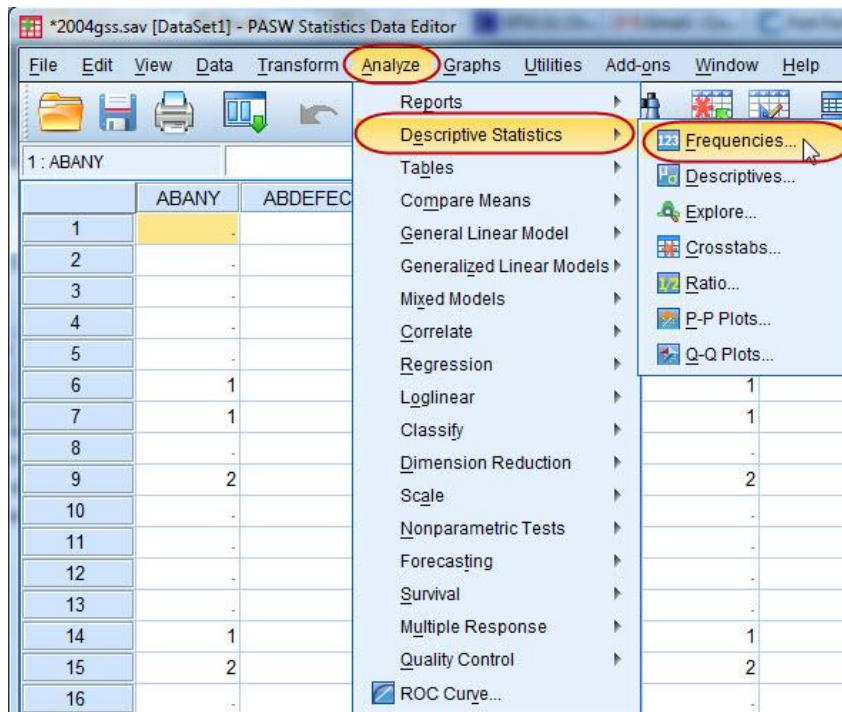**Analyze** ➡ **Descriptive Statistics** ➡ **Frequencies**

Figure 1-9

In the Frequencies dialog box click on ABANY, and move it into the Variable(s) box by clicking the arrow at the right side of the box.
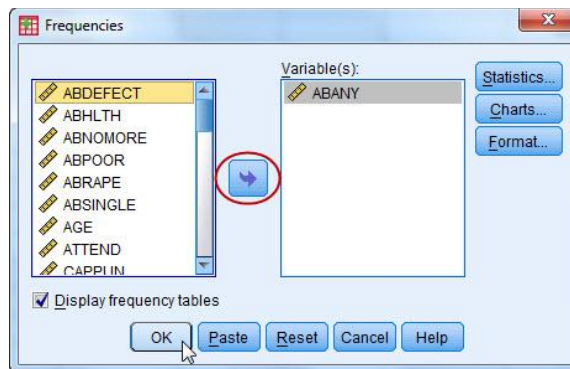

Figure 1-10

For now, just click OK. This will cause SPSS to process our request and display it in the Output window, which looks like Figure 1-11 below.
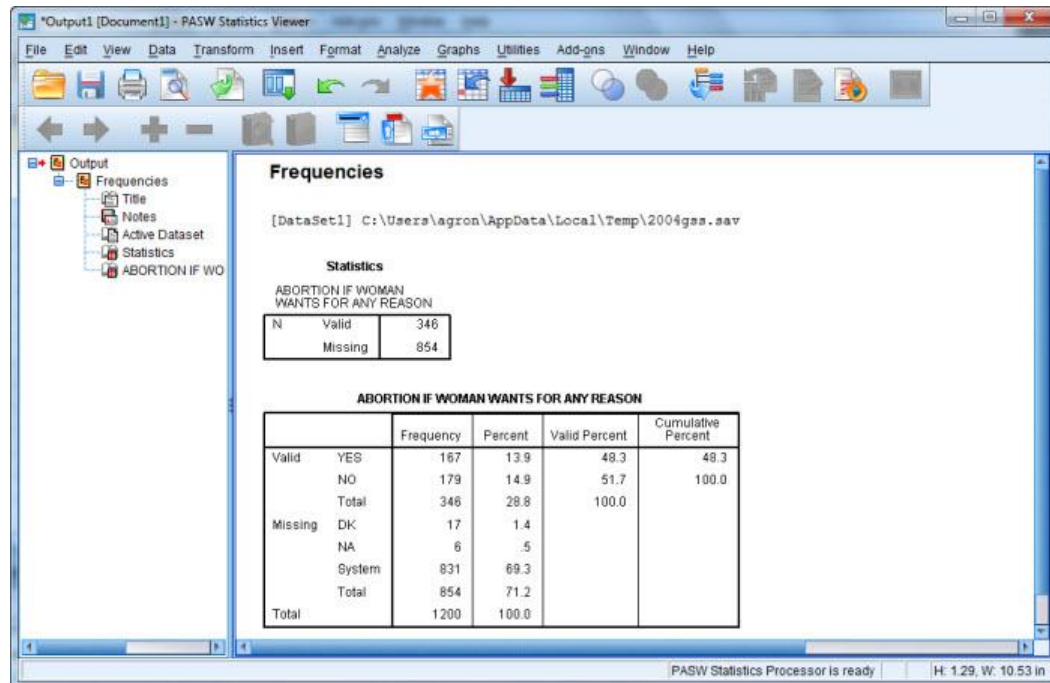
Figure 1-11

The Output Window is made of two smaller windows. The left hand panel contains the outline of the output that IBM SPSS gives you and can be used to quickly navigate to what you want to view. If the material or content is too long in the main window to the right, you can click on the title of the procedure in the window on the left and that content of that procedure will show in the window at the right side. The plus sign allows you to expand the outline of a single SPSS procedure.

The window on the left has a tree-like menu which is made of five headers:

1. The Title of the SPSS procedure (Frequencies),
2. Notes,
3. Active Dataset (which shows the location of the data set used from the local computer),
4. Statistics (a summary of the number of valid and missing cases), and
5. The name of the actual table containing the frequency distribution numbers and percentages.

If you go straight to "Statistics" header, you will see in the table that there were 346 valid cases (those who answered) and 854 missing cases. If you click on the actual procedure header – ABORTION IF WOMAN WANTS FOR ANY REASON – you observe that out of 1200 cases, 167 participants said **yes** (making about 14% of the whole sample), 179 answered **no** (making about 15% of the whole sample), 17 said they **didn't know** (making about 1.4% of the whole sample), 6 **didn't answer** (making about half percent of the whole sample), and 831 were entered in the dataset as **not applicable**. As we've said above, the "not applicable" 831 respondents were not asked this question (making more than 69% of the whole sample). It is a common practice to not ask all respondents a particular question in a survey research. In this case, only about 370 (exactly 369) of the approximately 1200 respondents were asked this question. The percentages we mentioned before were taken from the Percent column. You may wonder what is the Valid Percent column… Well, usually, we are interested only in those who were asked the questions and answered it. Thus, we will leave out those respondents who were not asked the question, those who did not answer when asked, and those who did not know the answer. Those who answered constitute those who have an opinion about the research topic. So, now, we need to know what percent of the respondents said **yes** or **no** of those who had an opinion. In this case, we

look at the Valid Percents column in the table. We, then, observe that 48.3% of "valid" respondents (those who had an opinion) answered that abortion when woman wants it for any reason should be legal, while 51.7% thought it should be not be legal.

In addition to descriptive statistics, research projects need to address questions of relationships between two or more variables. We call this *inferential statistics*. Let's enter another variable in the arena. Who would you think would favor allowing a legal abortion when a woman wants it for any reason, men or women? Maybe you are thinking "women", since they are the ones who are more attached to the issue (this could be your future hypothesis). To find out the answer to this question we will perform a *crosstabulation* procedure.
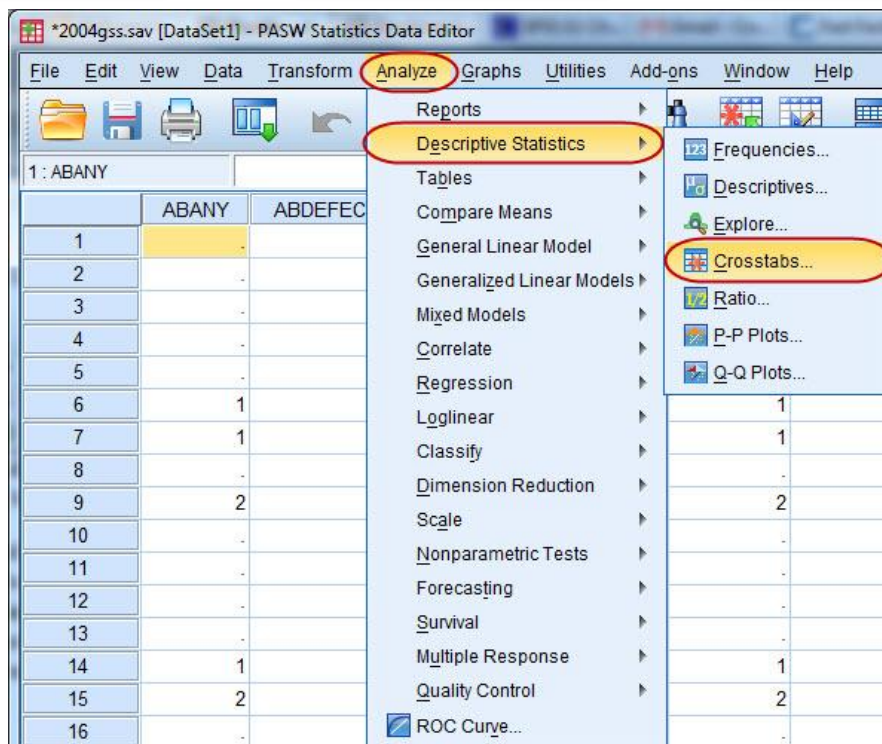
Analyze ➡ **Descriptive Statistics** ➡ **Crosstabs**

Figure 1-12

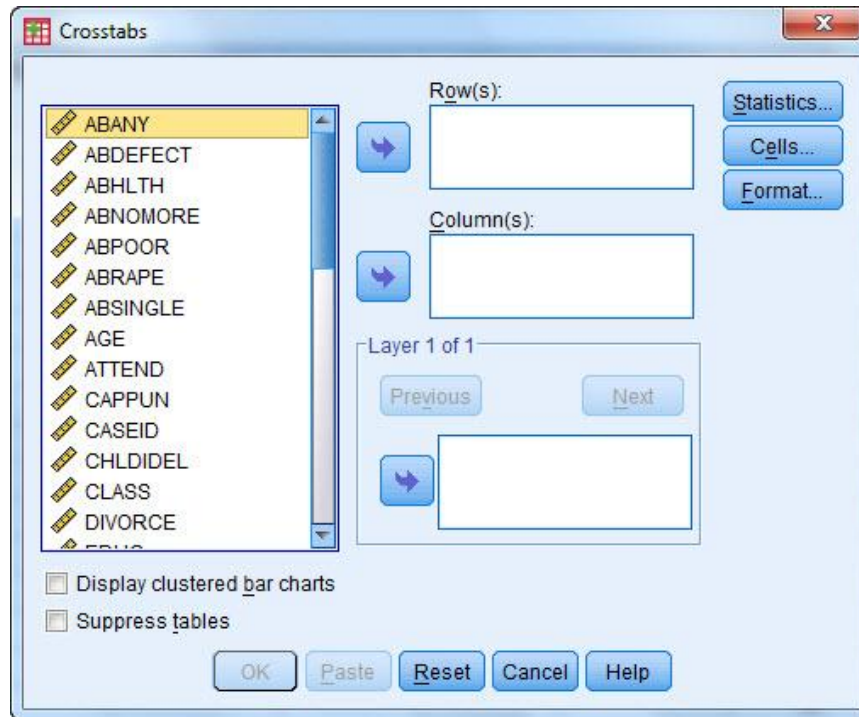The Crosstabs dialog box opens (Figure 1-13).

Figure 1-13

In this dialog box, you see three small fields in the right side. One says "Row(s)", another says "Column(s)", and the other says "Layer 1of 1". We need to fill the "Row(s) and "Column(s)" fields. Usually the independent variable goes on the Column(s) box and the dependent variable goes on the Row(s) box.

Click on the variable "ABANY" and move it to the Rows box. Sometimes, gender as a variable can be named as "gender" or as "sex". In this dataset it is SEX. Scroll down on the left box to find the variable SEX. Alternatively, you can press letter "S" in the keyboard to go straight to variables that start with that letter. Move it into the Columns box (column box generally is the place for independent variables). Make sure that your screen looks like the figure below, before moving to some options in the dialog box.
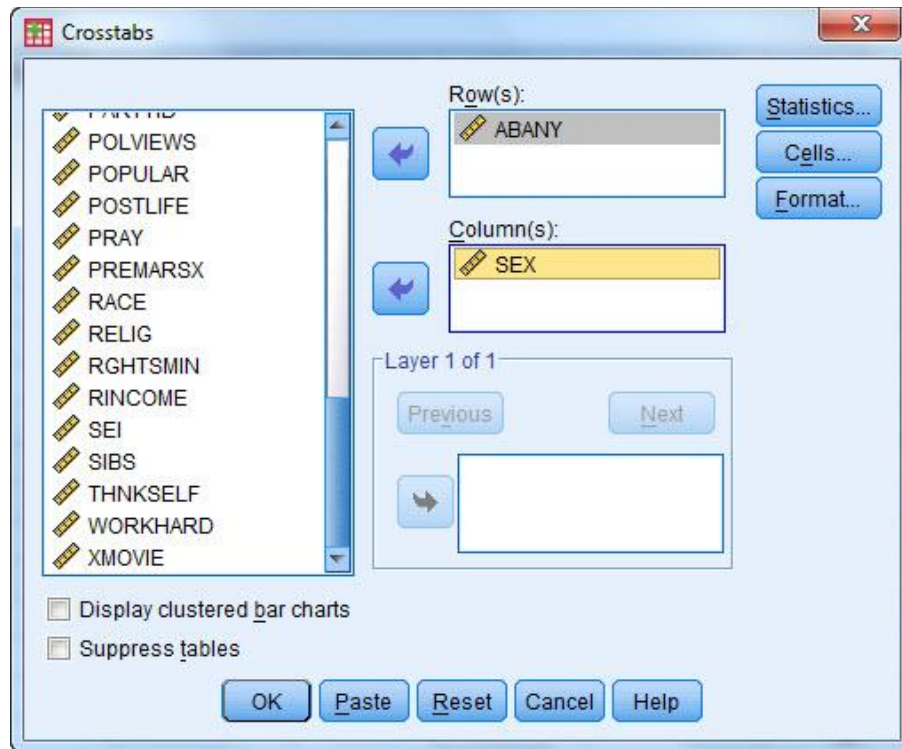
Figure 1-14

These settings will produce a table with cell frequencies, but raw numbers or frequencies by themselves are seldom useful. Percentages, however, make for a better and more useful comparison between different categories (or values) of variables. Luckily, we can "order" SPSS to compute percentages. Click on the "Cells" button. In the *Crosstabs: Cell Display* dialog box, check the box next to "Column" percentages (see figure 1-15 below).
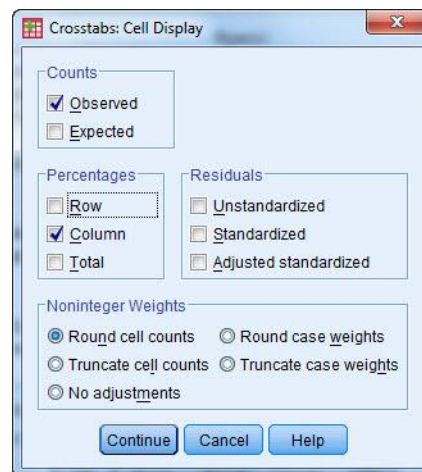

Figure 1-15

Click "Continue" and then click "OK". SPSS gives you the Output Window (figure 1-16 below).
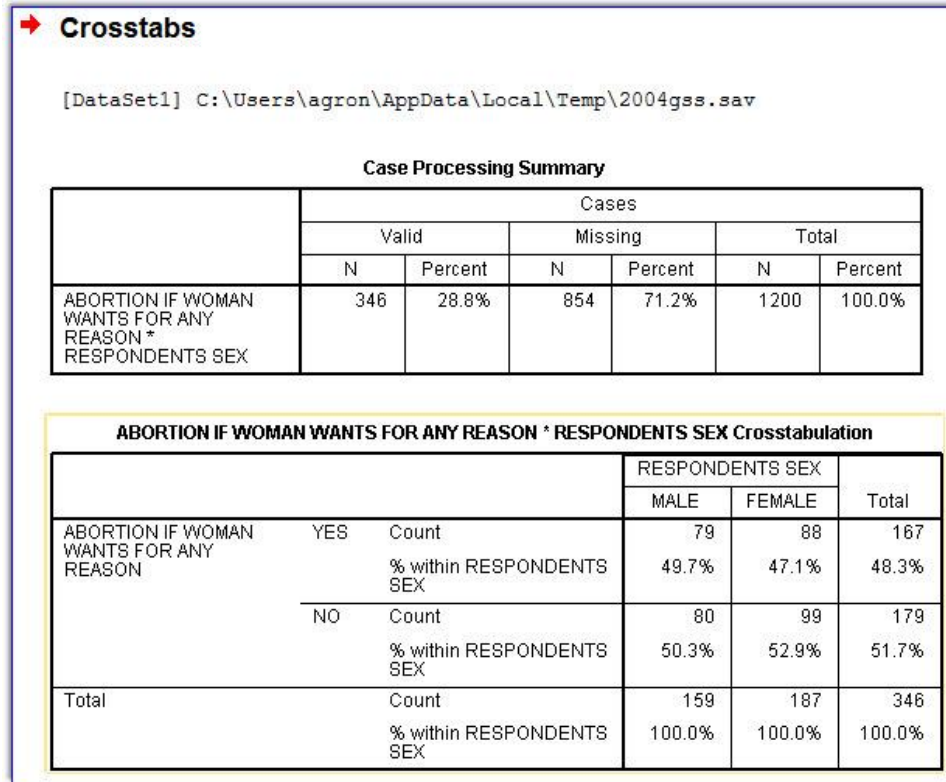
Crosstabs

[DataSet1] C:\Users\agron\AppData\Local\Temp\2004gss.sav

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
ABORTION IF WOMAN WANTS FOR ANY REASON * RESPONDENTS SEX	346	28.8%	854	71.2%	1200	100.0%

ABORTION IF WOMAN WANTS FOR ANY REASON * RESPONDENTS SEX Crosstabulation

			RESPONDENTS SEX		
			MALE	FEMALE	Total
ABORTION IF WOMAN WANTS FOR ANY REASON	YES	Count	79	88	167
		% within RESPONDENTS SEX	49.7%	47.1%	48.3%
	NO	Count	80	99	179
		% within RESPONDENTS SEX	50.3%	52.9%	51.7%
Total		Count	159	187	346
		% within RESPONDENTS SEX	100.0%	100.0%	100.0%

Figure 1-16

Look at the crosstabulation (otherwise called a contingency table, because one variable is contingent—depends on—the other). Let's compare men and women among those who are in favor of a woman right to obtain a legal abortion for any reason she chooses. We observe that 49.7% percent of men said yes to an abortion for any reason compared to 47.1% percent of women. We can see that, contrary to what we thought in the beginning, men approve more than women, although the difference is very little – 2.6%.