1 The Scientific Study of Society

1.1 OVERVIEW

In this chapter we introduce you to some of the important building blocks of a scientific approach to studying the social world. As you can already tell from reading the first chapter of *The Fundamentals of Social Research* – which we will refer to as "*FSR*" or the "main text" from here on – data are an important part of what we do both to explore the world and to test hypotheses based on causal theories. An important part of working with data is learning how to use a statistical software package. In the sections that follow, we introduce you to the Stata program and some basics that you will need to get up and running. In doing this, we also introduce some general principles of good computing practices for effectively working with data.

1.2 "A WORKBOOK? WHY IS THERE A WORKBOOK?"

You might be asking yourself this question, and it's perfectly fair to do so. Allow us to try to explain how this workbook fits in with the main *FSR* text.

As you will see in the weeks and months to follow in your class, the main textbook will teach you about the use of statistics in the social sciences, mostly by using equations and examples. So yes, in some ways, it will feel rather math-y. (And we think that's cool, though we realize that it's not everyone's cup of tea.) One of the ways that people learn about the practice of statistics is to use computer software to calculate statistics directly. To that end, many instructors want students to learn to use a particular computer software package so they can begin to conduct statistical analyses themselves.[1] We have discovered through years

[1] This particular software companion book teaches students to use Stata, but we have also produced parallel books for instructors who wish to have their students learn SPSS or R.

of teaching that this transition between equations in a book and software output on a computer screen is a very difficult one. The goal of this software companion book is to make this connection stronger, even seamless.

If we are successful, this book will do two things. First, it will teach the nuts and bolts about how to use Stata. Though many (perhaps most) students today are quite computer-literate, we believe that having a reference guide for students to learn the techniques, or for them to teach themselves out of class time, will be helpful. Second, and more importantly, this software guide will provide explicit hand-holding to you as you learn to connect the key principles from the main text to the practical issues of producing and interpreting statistical results.

Each chapter of this software guide works in parallel with that of the main *FSR* text. So when you learn the equations of two-variable regression analysis in Chapter 10 of the main text, you will learn the details about using Stata to estimate two-variable regression models in Chapter 10 of this companion book. And so on. In the end, we hope that the very important (but perhaps rather abstract) equations in the text become more meaningful to you as you learn to estimate the statistics yourself, and then to learn to interpret them meaningfully and clearly. Those three things – formulae, software, and interpretation – together provide a very solid foundation and basic understanding of social science.

Let's start.

1.3 GETTING STARTED WITH STATA

To get started with Stata, we recommend that you set yourself up in front of a computer that has the program installed with a copy of *FSR* close by. You should also have the set of computer files that accompany this text (which you can download from the text's website, www.cambridge.org/fsr) in a directory on the computer on which you are working. You will get the most out of this workbook by working in Stata as you read this workbook.

The instructions in this book can help you learn Stata whether you use a Windows-based PC or a Mac. Once the program is launched, Stata works identically, no matter which platform you use. Mac users should be aware, though, that our screenshots will come from a Windows-based PC. Some of those screenshots that involve finding and opening files on your computer, therefore, will look somewhat unfamiliar to Mac users, but we assume that Mac users are at least somewhat used to this. Overall, the differences between running Stata on Windows compared to a Mac are minimal. That said, we have created a help guide on the differences between working with Stata on a Windows-based PC and a Mac operating system, which can be found online at www.cambridge.org/fsr.

Finally, we wrote this book while using versions 16 and 17 of Stata. Particularly for the statistical fundamentals you will learn in this book, the differences between versions – at least as old as Stata 12 – are not severe. In fact, if you use any version of Stata between 12 and 17, you might not notice the difference between what appears on your screen and what appears in the screenshots contained in this book.

1.3.1 Launching Stata

When you are sitting in front of a computer on which Stata has been properly installed, you can launch the program by double-clicking on the Stata icon or by finding the Stata program on your start menu. Once you have successfully launched the Stata program, you will sometimes be prompted with a small box of options for updating the program like what we see in Figure 1.1. If this box does pop up when you launch the program, then we recommend that, for now, you click the option "Check next time Stata is launched" and then click "OK."

At this point, you should see one large window like that in Figure 1.2. Within this main Stata window, you will see four other windows labeled "History" (on the left side), "Variables" (on the top right side), "Properties" (on the bottom right side) and "Command" (across the bottom). The remaining area in the middle, known as the "Results" window, is not labeled. If you are seeing all of this, you are ready to go.

1.3.2 Getting Stata to Do Things

In almost any mainstream statistical program today, there are multiple ways to accomplish the same tasks. In Stata, almost any command can be executed using pull-down menus, typed commands in the command window, or typed commands in a do-file window. The choice of which

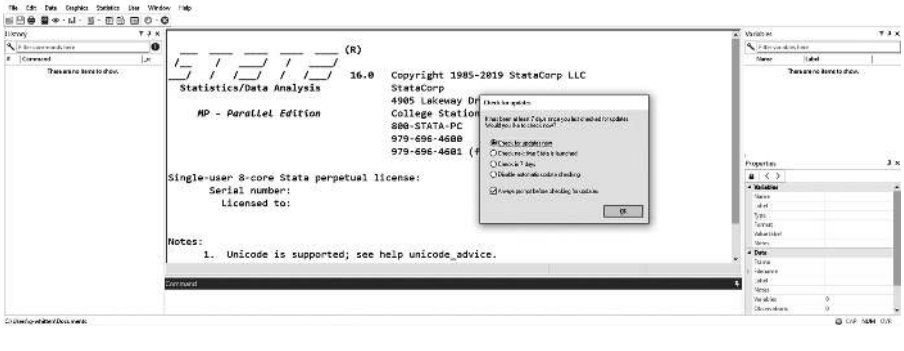

Figure 1.1 Stata after initial launch with update options box

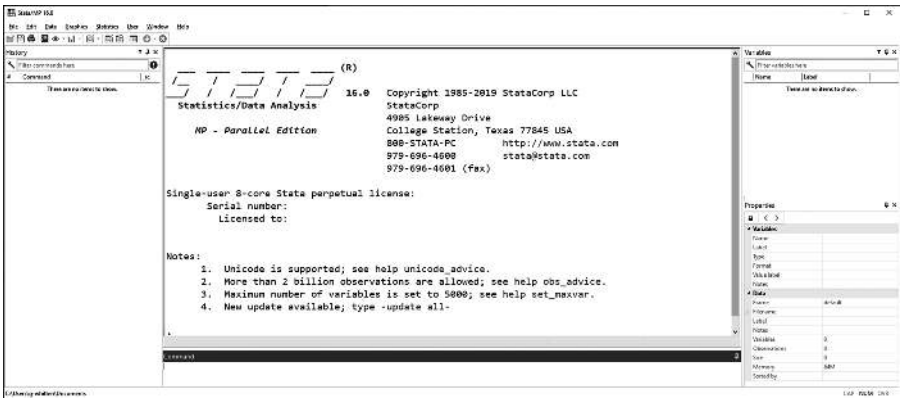

Figure 1.2 Stata initial launch

of these options to use is a matter of personal comfort. But, as we discuss below, no matter which way you choose to get Stata to do things, you need to keep track of what you are doing. We now discuss the three ways to get Stata to do things by showing an example of opening a data set. We recommend that you try all three, but especially the example of using a do-file in the final subsection below.

Using Pull-Down Menus

If you prefer to use pull-down menus, you need to start with either the row of textual headings across the top left of the program (starting with "File," then "Edit," etc.) or, immediately under that, the row of icons (a picture of a folder opening, a picture of a floppy disk, etc.). In our initial example, we are going to open a data set, so we need to start either with the textual heading "File" or with the icon that looks like a folder being opened. In Figure 1.3, we show what this will look like if we click on "File."

Once you have clicked on "File," you will want to direct Stata to the location on your computer where you have placed the *FSR* Stata companion files (as we noted above, these can be downloaded from www.cambridge.org/fsr). In our running example, these files are located in the directory "C:\MyFSRStataFiles." So, to find our initial data set, named "African American Home Ownership 1985-2014", we would point Stata to this directory and then click on the file "African American Home Ownership 1985-2014" as shown in Figure 1.4.

Once you have done this correctly, your Stata screen should look like Figure 1.5. A few things have changed:

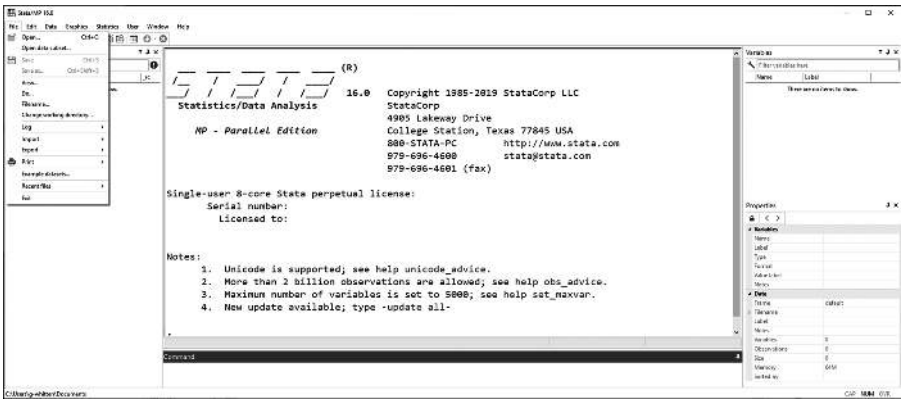

Figure 1.3 Stata with pull-down menu for "File" selected

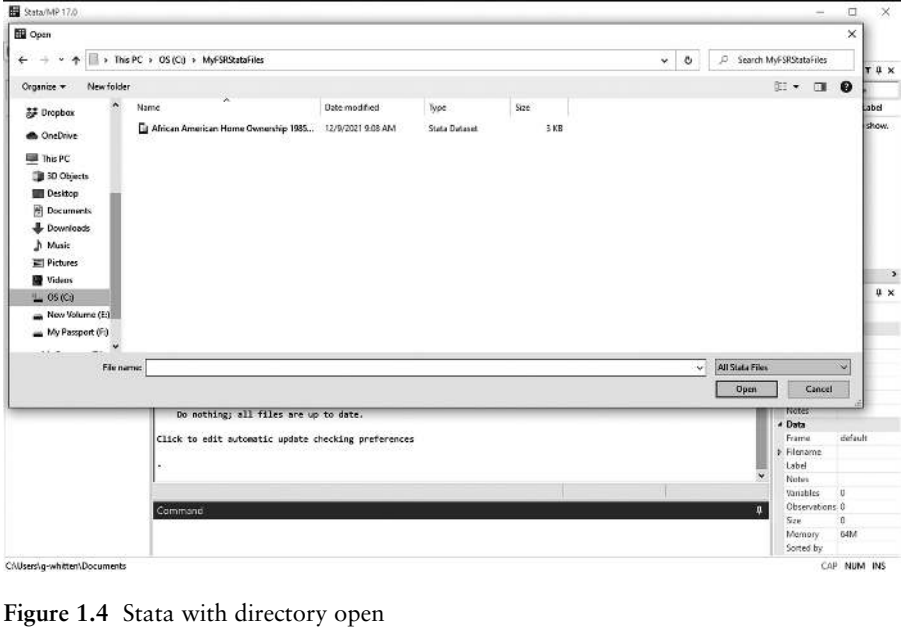

Figure 1.4 Stata with directory open

- In the "Results" window, we can see the text
  `. use "C:\MyFSRStataFiles\African American Home Ownership 1985-2014.dta"`
  where
  - the "." in front of this line indicates that this is a command that Stata has executed,
  - the name of the command is "use" which is the main Stata command for accessing a data set,

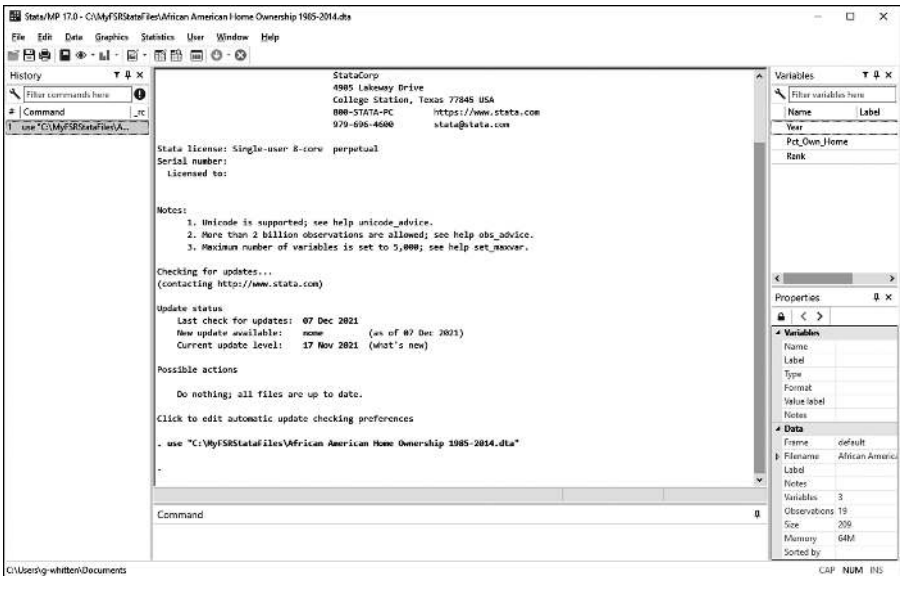

Figure 1.5 Stata with data loaded in

  - the text in quotes tells us the name of the file and location where the file was obtained.

- In the upper left corner, at the top of the "History" window, we see the number 1 in the "#" column followed by the beginning of the text of the command. This is where Stata keeps track of each command that it has executed.
- On the right side, we can see that there is new information in the "Variables" and "Properties" windows.
  - In the "Variables" window, we can see the names of the three variables that are contained in this data set.
  - In the "Properties" window, we can see some information about each variable and some information about the data set. In particular, we can tell that the data set contains three variables and nineteen observations.

Using the Command Window

You can type commands directly into the command window that you see across the bottom of the initial window that opens when you launch the program. These commands are typed in one at a time, and are executed by the program when you hit the "Enter" button on your keyboard.

So, if we want to load the data set "African American Home Ownership 1985-2014," which is a Stata data set (with the ".dta" suffix), you

would type the following command into the "Command" window and hit the "Enter" key on your computer:

```
use "C:\MyFSRStataFiles\African American Home Ownership
1985-2014.dta", clear
```

If you have done this correctly, your Stata will look like Figure 1.5, except that it will have the "`, clear`" at the end.[2] The "`, clear`" tells us that Stata cleared out any data that we had sitting in the program's memory before it opened our data set.

Using a Do-File

A third way to issue commands in Stata is to use a do-file. While this method of working will seem a little bit clumsy at first, it is our preferred method of working in Stata for reasons that we will explain below. To work with a do-file, you need to open a new window called a "Do-file Editor." To do this, go to the pull-down menus on the top left of the program and select "Window," "Do-file Editor," and finally left-click on "New Do-file Editor," as shown in Figure 1.6. We will eventually cover a lot of different things that one can do with a do-file, but, for now, all that we want you to do is to type the following command into the new do-file:

```
use "C:\MyFSRStataFiles\African American Home Ownership
1985-2014.dta", clear
```

Once you have typed this command into the do-file editor, you will then want to select the entire line – you can do this by left-clicking at the beginning of the line and then moving to the end and releasing the left mouse button – and then click on the icon at the right side of the top of the do-file window that looks like a piece of paper with writing on it with an arrow pointing to the right in the lower right corner of the icon. Clicking on this icon, named "Execute Selection (do)," will tell the program to execute the selected line of code. In Figure 1.7, we show what this will look like right before you click on "Execute Selection (do)." Once you have done this correctly, you will see output in the main window that looks like Figure 1.5.

[2] The location of files is often a stumbling block for beginning users of a statistical software package. To keep things simple, we recommend that you create a folder on your computer's C drive named "MyFSRStataFiles" and put all of the files that you have downloaded from www.cambridge.org/fsr into that folder. If you are unable to do this, then on a computer using a Windows operating system you can find the exact name of the location of a file by right-clicking on that file, left-clicking "Properties," and then looking at the entry to the right of "Location." This filepath, or location, can be selected, copied, and pasted directly into your command window (or do-file) to ensure that it is exactly right. As discussed earlier, a help guide on the differences between working on a Windows-based PC and a Mac operating system can be found online at www.cambridge.org/fsr.

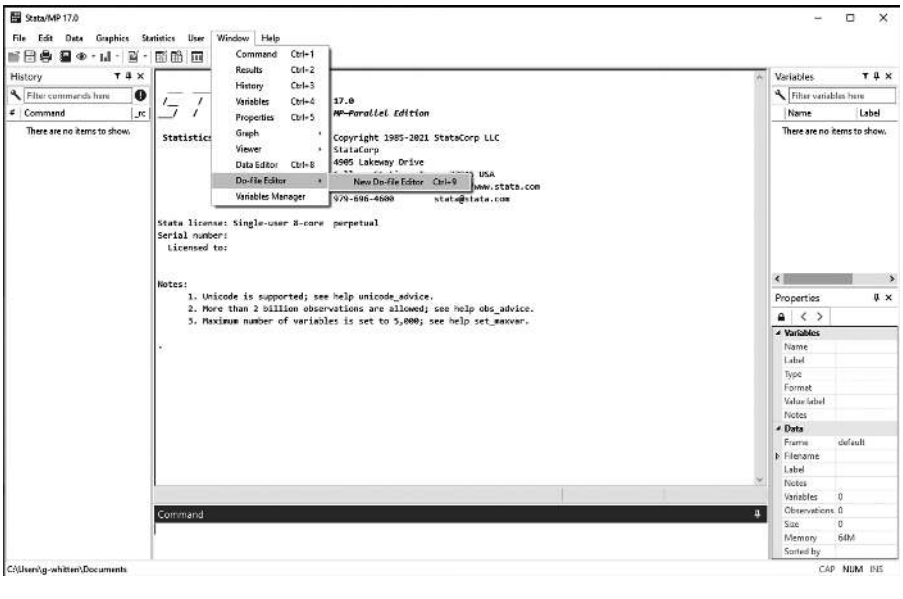

Figure 1.6 Opening a new do-file window

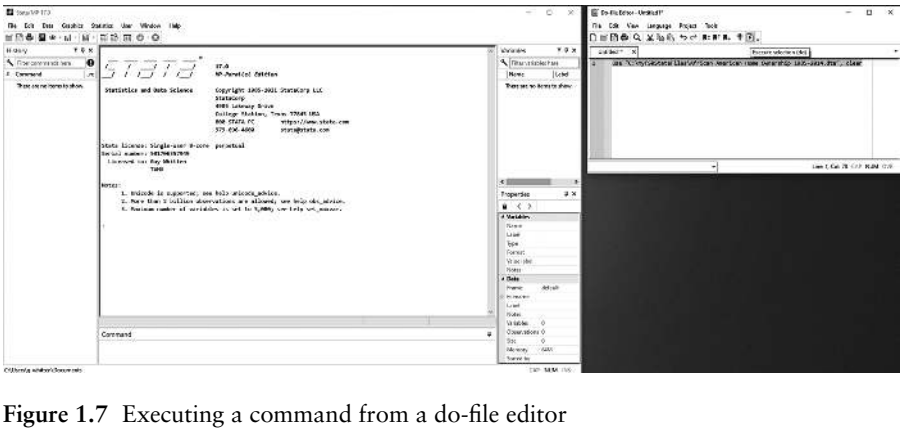

Figure 1.7 Executing a command from a do-file editor

1.3.3 Initially Examining Data in Stata

Now that we have shown you three different ways to get a data set into Stata, we want you to take a look at the data that you have loaded into the program.[3] These data are from the 1985–2014 General Social Surveys (GSSs), a data set that you will become quite familiar with in your statistics course because we make extensive use of it in this workbook as well as in the main text. The particular survey question that we examine in this

[3] We discuss how to manually enter your own data into a Stata file in an online appendix, available at www.cambridge.org/fsr.

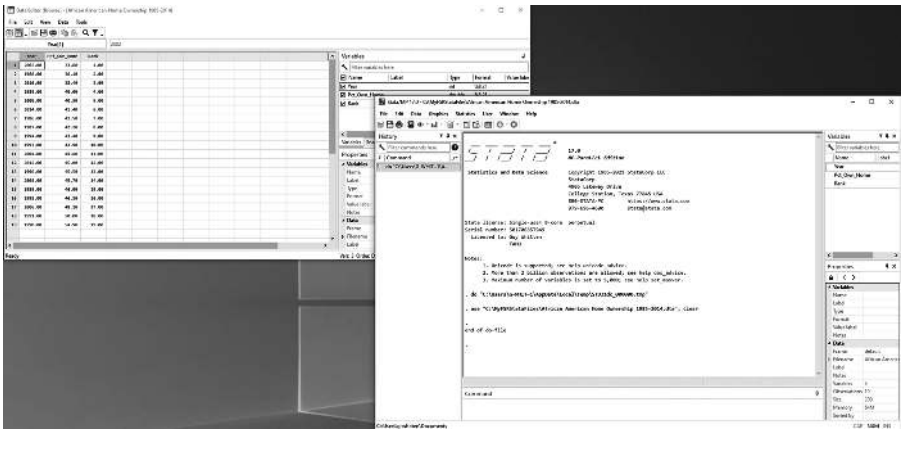

Figure 1.8 Initially examining data in Stata

illustration asks respondents whether they own or rent the dwelling in which they currently live. To get an initial look at these data, click on the "Data Editor (Browse)" icon which can be found in the top left of the main Stata window – it is the icon that looks like a spreadsheet with a magnifying glass over it. Once you have done this, your computer should look something like Figure 1.8. Each column in this spreadsheet contains values for a single variable and each row contains data from a single election.

In Figure 1.8, you will see data in three columns, labeled from left to right as "Year," "Pct_Own_Home," and "Rank." Those labels correspond to the names of the three variables in our data set. Down the rows you will see the cases of the three variables. For example, in row 10 of the data window, which is the tenth case in the data set, we see the following values for the three variables. "Year" takes on the value of 1993, "Pct_Own_Home" has the value of 43.90, and "Rank" has the value 10.00. These variable names make clear what each variable represents and what the corresponding values mean.

You can use the bars at the right and at the bottom of the data set to scroll through it. As you already know, this data set has only three variables, so you need to scroll to the right to see any other variables. However, you can use the bar at the right of the screen to scroll down when you have larger data sets. For now, you'll notice that the last of our cases has a value for "Year" of 1990 and a value for "Pct_Own_Home" of 54.90. You can also see that the data set has a total of nineteen cases.

You are now ready to proceed to the end-of-chapter exercises.

1.4 EXERCISES

1. Go through all of the steps described above. Once you have the "Data View" open (so that your computer looks like Figure 1.8), do the following:

   (**a**) Look at the values in the column labeled "Pct_Own_Home." This is our measure of homeownership. Do the following:

   i. Identify the year with the highest value for this variable.

   ii. Identify the year with the lowest value for this variable.

   iii. What does it mean if this variable goes up by one?

   (**b**) Look at the values in the column labeled "Rank." This is our measure of the rank, from lowest to highest, of the proportion of African American homeownership in the years from 1985 to 2014. Now do the following:

   i. Identify the year with the highest value for this variable.

   ii. Identify the year with the lowest value for this variable.

   iii. What does it mean if this variable goes up by one?