Using SPSS with *The Fundamentals of Political Science Research*

Paul M. Kellstedt and Guy D. Whitten
Department of Political Science
Texas A&M University

Contents

1 Overview

In this document we provide a guide for how to produce the statistics and graphs for our book using the program SPSS. This guide is not intended as a comprehensive introduction to SPSS. There are numerous books intended to serve this purpose.[1]

The ordering of topics in this appendix is intended to follow closely the ordering of topics in the book. After a couple of brief discussions of housekeeping issues (keeping track of your work and getting data into the program), we present the statistical and graphical tools in roughly the order in which they occur in the text. Throughout this guide we refer to specific sections of our book *The Fundamentals of Political Science Research (FPSR).*

1.1 Getting started—launching SPSS

SPSS is a menu-driven statistical software package that is suitable for most common forms of data analysis in the social sciences.

[1]For an up-to-date listing of such books we recommend going to SPSS's website at www.spss.com

To open the program you can click on the SPSS icon if it appears on your computer's desktop. Otherwise, you will need to locate SPSS using the "Start" menu in the way you would any other software.

Copies of SPSS on different computers may not have identical default settings. One thing that will be helpful in navigating different data sets, especially new data sets, or data sets with many variables, will be to do the following when launching SPSS Under the "Edit" menu, select "Options," which will open a radio box. On the "General" tab, be sure that the radio buttons for "Display labels" and "Alphabetical" are filled in. In general, this will make it easier to navigate around a data set with which you are unfamiliar.

1.2 Which version?

This text has been written using version 17.0 of SPSS on a Windows-based computer. Most of the syntax herein will work with earlier versions of SPSS and on different platforms.

1.3 The `data set window` and the `output window`

Although you have only launched a single program, in the Windows program tray, there are actually two separate windows open. The first is the `data set window`, and the second is the `output window`. The `data set window` contains the data set and variable definitions—more on this in a minute—and the `output window` is where the results of your statistical analysis will appear. So running a command on a data set in the `data set window` will produce results in the `output window` (which we will get to shortly). As you work through SPSS you will swap back and forth between these windows rather frequently.

Within the `data set window`, there are two important views, denoted by tabs in the lower-left-hand corner. One tab says `Data View` and the adjacent tab says `Variable View`. The `Data View` tab looks exactly like a spreadsheet, and it represents the way in which the data are stored. Each variable is in a separate column, with the name of the variable in the first row across the top. Likewise, each case is in a separate row. For example, in a data set collected as part of an election study, each column would represent a variable. One such column in such a data set would probably represent the variable for whether or not a respondent to the survey voted; another column would represent the variable for the level of education attained by the respondent; and so on. Each row represents the responses of a separate respondent. So a particular cell is where information is stored for a unique respondent to a particular survey question.

In different types of data sets, of course, the columns and rows will contain different types of information. In a time-series observational data set, the columns would again contain in-

formation about a particular variable—such as the unemployment rate—and the rows would represent the particular time unit at which the data were aggregated—such as months, quarters, or years.

The `Variable View` tab is accessed by clicking on that tab with the mouse. It contains information about what the variables actually mean—among other things, the variable label, the value labels, and any declared missing values. When you click in the `Variable View` tab, you will notice that the variable names that were previously in the first row of each column are now in the first column of each row. Each subsequent column, then, contains information about the variable in that particular row. The column `Label` gives a fuller description of what the variable actually is, helping to elucidate the meaning of the often arcane variable names in data sets; it is sometimes helpful to expand the column width by hovering the mouse over the word `Label` in the first row. The column `Values` gives the value labels so that you know what the numbers in the cells of the data set stand for. For any particular variable, you can click on the cell under `Values` for the particular variable you want to learn about, then click the small gray box at the right-hand-side of the cell; a dialog box will open telling you what the numbers for the variable stand for.[2] Finally, if any values for a given variable are missing—for example, if a survey respondent did not answer a particular question, or if there is missing data for a particular time period—the column `Missing` contains information indicating what values signal that a case has "missing" information for that variable. When there are "missing" values declared, those observations with the designated missing values will be dropped from any analysis that uses that variable.

2 Keeping track of your work

Any time that you work with data, you should have the following questions in mind:

1. If I put these data and my work with them aside for a month, would I be able to go back and *exactly* recreate my work?
2. If I put these data and my work with them aside, would *someone else* be able to go back and *exactly* recreate my work?

If the answer to both questions is "yes," then you have done a good job keeping track of your work.

[2] You can edit the value labels by using this procedure.

2.1 Replication is high flattery

All of this emphasis on keeping track of your work may seem unecessary or tedious, especially when you consider the second question. But keep in mind the crucial role that our statistical hypothesis testing plays on the road to scientific knowledge (see Chapter 1). When a researcher comes up with a hypothesis test that supports a new theory or challenges an established theory, the first thing that other researchers will want to do is to **replicate** their work. Initial replication of another researcher's work may involve either recreating the same results with the original data or the collection of new data and attempting to recreate the same results. Either way, it is crucial that the researcher who is replicating someone else's work be able to repeat what they did.

Replication may seem like a distrustful exercise but, keep in mind that one of our rules of the road (again, from Chapter 1) is to "consider only empirical evidence." Replication is a very serious consideration of another researcher's empirical evidence. Boring results seldom get replicated. Thus, rather than being distrustful or insulting, replication is actually high flattery.

2.2 Using the `output window`

An excellent way to keep track of your work in SPSS is to properly utilize the `output window`. Recall that the `output window` is opened automatically when you launch SPSS. The results of all statistical analyses that you perform in SPSS will appear in the `output window`.

Within the `output window`, several features are notable. First among these is that the file can be annotated by clicking on "Insert" in the menu bar and clicking "New Text." This can be helpful if you need to make notes to yourself about the analysis you've just performed. Second, note that the output can be edited, or entirely deleted, by clicking on the left pane of the output window, which will highlight the analysis you've just done. It can, for example, be deleted entirely if you make a mistake. Finally, note how to save your output in the expected way, by clicking "File" and "Save," or by clicking the disk icon on the menu bar. As always, be careful about the drive and directory where you save your output. All of this will help you to know that you (and someone else) will be able to re-create the results that you've just produced.

2.3 Keeping notes

One of the best ways to keep track of what you have done is to write about it in a separate file or in a notebook. An obvious worry about writing about what you are doing while you are doing it is that you might work more slowly. However, what you lose in working speed is likely

to be more than made up for by avoiding the mistakes that writing allows you to avoid. We agree wholeheartedly with the words of Diedre McCloskey that "writing is thinking." And, in this case, slowing down and thinking about what you are doing will often save you time in the long run.

3 Getting data into the program

Upon launching SPSS a window will open, and a dialog box will appear atop it, asking "What would you like to do?" If you are working on an existing data set, click the radio button that says "Open an existing data source" and click "OK." That will open a new dialog box asking you to specify the drive and directory where the data set resides. Navigate to the location of the data set that you plan to use, click on it, and click "Open" and then "Data." Note that SPSS assumes that you will be using data sets saved in SPSS format, which have ".sav" as their file extension. In the dialog box that appears when clicking "Open" and "Data," however, SPSS allows you to import files from other programs, such as Stata or a Microsoft Excel file.

Before analyzing the data set you've just opened, it is helpful to keep in mind a few tricks for transforming your data set so that you can perform the analysis that you have in mind.

3.1 Making your own data set

If you make your own data set, we recommend that you enter your data into a spreadsheet like Excel and then convert your spreadsheet into SPSS format.

3.1.1 Naming your variables

Variable names should convey as much information about a variable as possible. They should be designed for use in such a way that they prompt your memory of what you did when you created the variable. When creating variable names in SPSS you should keep in mind the following list of concerns:

- Variable names may not normally begin with a $ or a # symbol.
- Variable names may not contain spaces. For instance, "presidential_vote" is allowed, but "presidential vote" is not a valid name for a single variable.
- Variable names in SPSS are not case sensitive. This means that "vote1984" and "Vote1984" refer to the same variable (and duplicate variable names are not allowed).

- Variable names in SPSS can be up to 64 characters long; in practice, however, you will find that variable names longer than 10 characters are pretty unwieldy.

3.2 Getting to know someone else's data

When working someone else's data, it is important to get to know each variable with which you intend to work. To get to know each variable, we recommend reading any documentation that comes with their data set. Many data sets have accompanying text files, often called "codebooks" that explain how the cases were selected and the various variables measured. A careful read through of this material will help you to know a data set and to avoid making mistakes. We also recommend that you produce descriptive statistics and/or descriptive graphs for each variable before you run any hypothesis tests with more than one variable.

It is particularly important to know what values represent missing values when you are working with someone else's data. In SPSS missing values can either be represented as a period (.), or by a particular numeric value, or, in some data sets, in both ways. If you are not careful about this, you may mistake a missing value for an actual value and come to faulty conclusions as a result.

4 Making changes to a data set

When you are working with a data set, you will often want to make changes before you conduct your analyses. We have already stressed the importance of keeping track of your work. This is especially crucial when you are making changes to a data set. Another important part of changing a data set is to always be able to go back to the data set with which you started working. We highly recommend that you save the original data under a different name after you make changes to it and that you preserve the original data set.

4.1 Recoding values of a variable

It is often the case that we want to recode our variables before we use them to test hypotheses. As an example, let's say that we want to test a theory about what makes someone more or less likely to vote for an incumbent presidential candidate. If we are using survey data from the 2004 National Election Study, the variable V045026, with a variable label "Voter: R's vote for President" provides a useful measure of our dependent variable. We can see from the NES codebook that the possible values for this variable are 1. John Kerry, 3. George W. Bush, 5. Ralph Nader, 7. Other, and 9. Refused. For our purposes, the we want to distinguish between those voters who voted for the incumbent (the 3s), and those who voted

for someone else (the 1s, 5s, and 7s). Survey respondents who refused to say who they voted for present a bit of a dilemma but, since we do not know whether they voted for or against the incumbent, the safe thing to do is to treat them as missing for our variable. One of the best ways to conduct a recode is to create a new variable equal to the old variable and then recode the new variable. This is particularly helpful in this example because the old variable name "V045026" did not convey very much information to us.

In such cases, you need to recode that variable. To do this, we use the menu "Transform" and then "Recode" "into Different Variables," which will open a dialog box.[3] From the left-hand scroll box, select the old variable and click the $\hookrightarrow$ button to select that variable. On the right-hand side, you will need to provide a new variable name and variable label. It's a good idea to try to be as descriptive as possible in these situations; you might be surprised how easy it is to forget your own data manipulations weeks or days after you've initially made the changes. Then click on the button for "Old and New Values," which will open still another dialog box. On the left-hand side of that box, you can input the "Old Values" of the old variable—and you can use the self-explanatory "range" boxes if appropriate—and then, on the right-hand side, decide on a "New Value" for how you want the old values to be recoded into new values. When you've made your first recoding, click the "Add" button, and then repeat the process until you've finished, being careful to recode any missing values in either the old variable, the new variable, or both. When you've finished with all possible values, click on the "Continue" and then the "OK" buttons.

In the example for variable V045025 in the 2004 NES, using the above procedures, we would categorize a vote for the incumbent as a 1, and all others as 0. So under "Old Values," type 1 (for the initial value for John Kerry votes), and under "New Values," type 0, and then click "Add." Repeat this for 5 and 7, recoding both those old values to 0. Then, for Bush voters, put 3 in "Old Values" and 1 in "New Values" and click "Add" as well. Finally, be sure to deal with missing cases. A 9 in the old variable would need to be declared "Missing" under the "New Value" using the radio button there. When you're done, click on the "Continue" and then the "OK" buttons.

When you've done this, you'll want to be sure to save changes to your data set by clicking the familiar disk icon in the tray, or following the "File" and "Save" menu commands.

4.1.1 Checking your recodes

Any time that you recode a variable, it is important to produce a table to check that you have done what you set out to do with your recode command. We provide an example of this below in the section on tables.

[3] In general, it is unwise to choose the option to "Recode into Same Variable" option, as you will lose information that might be useful to you or someone else using this data set in the future.

4.2 Case selection

In certain instances, you may wish to perform analyses on only a subset of the cases in your data set. For example, in a data set of survey respondents, you may be interested only in examining the data from female respondents. Or, in a time-series data set, you may be interested particularly in the dynamics of government approval only for a particular administration. In such situations, the SPSS menu can specify which cases you want to analyze, and which cases you wish omitted from analysis. From the main toolbar, select "Data" and then "Select Cases." This will open a dialog box for you to specify the criteria of your choice for including or excluding cases for subsequent analysis. You will note that the radio button "All cases" is the default, meaning that SPSS will use all cases in the data set for analysis unless told otherwise. The subsequent radio buttons provide options should you wish to analyze only a subset of cases. Typically, the radio button "If condition is satisfied" is the most simple way to select a subset of cases. Selecting this will open yet another dialog box, and you can fill in the condition using the mathematical symbols (such as the '=' sign) and the relevant variable name(s) (which you can see from the scroll box on the left-hand side) in the box at the top, and clicking "Continue" and then "OK," and the condition specified will be applied to the active data set. Note that, in the `Data View` window, the de-selected cases will have a slash mark through their case numbers on the left-hand side to indicate that those cases will not be used in subsequent analysis.

For example, if you are using a time-series data set and wish only to examine cases during the year 2006, you would need to find the variable that contains information for the year (in many cases, the variable would be named "year"). Next, you would need to find how that variable is coded. For example, do the values of the variable year read "2006" or simply "06"? (If you don't know, look under the `Data View` window at some values for the variable of interest, or, in other circumstances, look at how non-obvious variables like a respondent's gender are coded in the `Variable View` window under `Labels`.)

When you wish to return to analyzing all cases, of course, you will need to repeat this process but select the "All cases" option.

5 Producing descriptive statistics and graphs

Once you have your data in SPSS it is important to get to know your data. In Chapter 6 we discussed a variety of tools that can be used to get to know your data one variable at a time. In this section we discuss how to produce and present such information. An important first step to getting to know your data is to figure out the what is the measurement metric for each variable. For categorical and ordinal variables, we suggest producing frequency tables. For continuous variables, there are a wide range of descriptive statistics.

5.1 Describing categorical and ordinal variables

As we discussed in Chapter 6, a frequency table is the best way to numerically examine and present the distribution of values for a categorical or ordinal variable. In SPSS this is done through the "Analyze" menu. Under "Descriptive Statistics," click "Frequencies...", which will open a dialog box. At that point, you will be able to choose the variable(s) for which you want frequency distributions by clicking on the variable name in the left-hand window, followed by clicking the $\hookrightarrow$ symbol.

This command produces a five-column table in which the first column contains the variable values (or value labels if there are value labels for this variable); the second column is the number of cases in the data set that take on each value (labeled "Frequency"); the third column (labeled "Percent") is the percentage of cases that take on each value, *including any cases with missing values*; the fourth column (labeled "Valid Percent") contains the same calculation but excludes any cases with missing values; and the fifth column (labeled "Cumulative Percent") is the cumulative percentage of cases from top to bottom.

Pie graphs or bar charts, such as Figures 6.1 and 6.2 in *FPSR*, are a graphical way to get to know categorical and ordinal variables. From the same dialog box that you used to generate the frequency distribution, by clicking on the "Charts..." option, you can produce either a bar chart or a pie chart by selecting the appropriate radio button. These charts then appear in the output window.

5.2 Describing continuous variables

A full battery of descriptive statistics for a continuous variable can be found in one of two ways. First, you can produce these statistics by clicking "Analyze" "Descriptive Statistics" and "Frequencies," as above, and then clicking on the "Statistics..." option to choose which particular statistics you would like to see. (Avoid the temptation to mindlessly click them all.) Note also that this option allows for you to compute the standard error of the mean (from Chapter 7 of *FPSR* in addition to the standard deviation (from Chapter 6). A second way to get statistics, but without the frequency distributions, is to click on "Analyze" "Descriptive Statistics" and "Descriptives..." and then clicking on "Options..." to choose which statistics you would like to have calculated.

To produce a box-whisker plot or a histogram, select "Graphs," then "Chart Builder..." Doing this opens a dialog box, and in the lower-left corner of that dialog box, you have several options from which to choose. "Box plot" and "Histogram" are two of those options. You begin the process by dragging the type of graph you want to produce up into the chart preview area in the upper-right corner of the dialog box. Next, you click on the "Element Properties..." box, which opens yet another dialog box. It is here that you choose the variable

to plot, and there are various options you can select to add a title to your chart, for example. Clicking "OK" will produce the chart in the `output window`.

It should be noted that the ways in which to customize your figures are numerous, and, importantly, that they vary significantly from version to version of SPSS.

5.3 Putting descriptive statistics into tables

When we produce descriptive statistics in SPSS we get a lot of output. Usually this output is much more than what we need to present in a paper that describes our variables one at a time. We therefore suggest making your own tables in whatever word processing program you are working with.

5.4 Putting SPSS graphs into documents

Once you have produced an SPSS graph that you want to include in a document, one of the easiest ways to do so is to right-click on the graph in the SPSS output window and select "copy" and then right-click on the location where you want to place the graph in your word processing program and select "paste."

6 Bivariate hypothesis tests in SPSS

In this section we go through the four different types of bivariate hypothesis tests discussed in Chapters 8 and 9 and discuss how to conduct these analyses in SPSS.

6.1 Tabular analysis

In tabular analysis, we are testing the null hypothesis that the column variable and row variable are unrelated to each other. We will review the basics of producing a table in which the rows and columns are defined by the values of two different variables, generating hypothesis-testing statistics, and then presenting what you have found.

To produce a two-variable table in SPSS select "Analyze," then "Descriptive Statistics" then "Crosstabs..." which will open a dialog box. You should put your dependent variable in the box labeled "Row(s)" and your independent variable in the box labeled "Column(s)." You do this by finding each variable from the scroll-able list on the left-hand side of the dialog

box, and then clicking the ↪ symbol. This will create tables that conform to conventional expectations.

By clicking on the box for "Cells..." you can choose what information you want displayed in the cells. When the dialog box opens, under "Percentages" be sure to click "Column." *In this case, the default for* SPSS *is to include only the observed cell counts.* This is another instance where clicking checkboxes haphazardly will end up creating problems instead of solving them.

As detailed in Chapter 8 of *FPSR*, these column percentages allow for the comparison of interest—they tell us how the independent variable values differ in terms of their distribution across values of the dependent variable. It is crucial, when working with tables of this nature, to put the appropriate variables across the rows and columns of the table and then to present the column frequencies.

6.1.1 Generating test statistics

In Chapter 8 of *FPSR*, we discuss in detail the logic of Pearson's chi-squared test statistic which we use to test the null hypothesis that the row and column variables are not related. This is produced in SPSS in the "Crosstabs" dialog box by clicking on the "Statistics..." button, and checking the box "Chi-square" in the top left corner, then clicking "Continue" to get back to the "Crosstabs" box. When you click "OK" to run the analysis, the results of the Chi square test will be in a separate box immediately beneath the table. The calculated value of the Chi square statistic, based on the difference between observed and expected cell counts, appears in the row labeled "Pearson Chi-Square," and in the column "Value." The degrees of freedom appear in the next column, labeled "df," and the p-value appears on the column labeled "Asymp. Sig. (2-sided)." When that value is less than .05, we reject the null hypothesis.

6.1.2 Putting tabular results into papers

We recommend that you make your own tables in whatever word processing program you choose to use instead of copying and pasting the tables that you make in SPSS . The first reason for doing so is that you will think about your results more closely when you are producing your own tables. This will help you to catch any mistakes that you might have made and to write more effectively about what you have found. Another reason for doing so is that tables constructed by you will tend to look better. By controlling how the tables are constructed, you will be able to communicate with maximum clarity.

As a part of making your own tables, you should have the goal in mind that your table communicates something on its own. In other words if someone only looked at your table,

would they be able to figure out what was going on? If the answer is "yes," then you have constructed an effective table. We offer the following advice ideas for making useful tables:

- Give your tables a title that conveys the essential result in your table
- Make your column and row headings as clear as possible
- Put notes at the bottom of your tables to explain the table's contents

6.2 Difference of means

Difference of means tests are conducted when we have a continuous dependent variable and a limited independent variable.

6.2.1 Examining differences graphically

When we use graphs to assess a difference of means, we are graphing the distribution of the continuous dependent variable for two or more values of the limited independent variable. Figure 8.1 shows how this is done with a box-whisker plot. In a situation with a discrete independent (X-axis) variable and a continuous dependent (Y-axis) variable, you need to drag the "Simple Boxplot" icon into the chart preview section in the upper-right portion of the dialog box. The "Element Properties..." button which opens another dialog box will enable you to specify the variables for each axis.

6.2.2 Generating test statistics

To conduct a difference of means t-test such as the one discussed on pages 146-150 of *FPSR*, select "Analyze," then "Compare means," then "One-Way ANOVA..." From the variable list on the left, click and drag the continuous dependent variable into the "Dependent List" box, and click and drag the discrete independent variable into the "Factor" box. Then click "OK" to run the analysis. The output in the output window provides the p-value in the far-right column labeled "Sig." When that value is less than .05, we reject the null hypothesis.

6.3 Correlation coefficients

Correlation coefficients summarize the relationship between two continuous variables.

6.3.1 Producing scatter plots

We can examine the relationship between two continuous variables in a scatter plot such as Figure 8.3 in *FPSR*. To produce a scatterplot, select "Graphs," then "Chart Builder..." Doing this opens a dialog box, and in the lower-left corner of that dialog box, you have several options from which to choose. "Scatter/Dot" is the option you should choose. You begin the process by dragging the type of graph you want to produce—there are many from which to select, but "Simple Scatter" is a good place to start—up into the chart preview area in the upper-right corner of the dialog box. Next, you click on the "Element Properties..." box, which opens yet another dialog box. It is here that you choose the variable to plot, and there are various options you can select to add a title to your chart, for example. Clicking "OK" will produce the chart in the `output window`.

6.3.2 Generating test statistics

To generate a correlation coefficient with the associated p-value in SPSS click "Analyze," "Correlate," and "Bivariate..." which will open up a dialog box. Click and drag the continuous variables that you want to correlate into the "Variables" box. In most cases, the default selections work well. To run your analysis, click "OK." The `output window` contains a matrix of the correlations—you can include more than two if you like—that contain the correlation coefficient as well as the p-value and the size of the sample used in the calculation.

6.4 Bivariate regression

The estimation of a bivariate regression model, as discussed in Chapter 9 of *FPSR*, is fairly straightforward. From the menu, select "Analyze," then "Regression," then "Linear..." which will open a dialog box. Click and drag a variable into each of the "Dependent" and "Independent(s)" boxes. Then click "OK."

The output produced in the `output window` contains more detail than you might expect. The R^2 statistic is found in the "Model Summary" portion of the output. One piece of information, however, is not as obvious: the sample size (n). The sample size is simply equal to the total number of degrees of freedom plus one.

In the "Coefficients" portion of the output, you will find estimated coefficients for both the y-intercept (or α), which is labeled "(Constant)," and for the slope of the regression line (or β), both of which are in the columns labeled "B." In adjacent columns are the standard errors for each coefficient, the standardized coefficient, the calculated t-statistic, and the p-value (in the "Sig." column).

7 Multiple regression

The estimation of a multiple regression model, as discussed in Chapters 10 and 11 of *FPSR*, is just an extension of the commands used for estimating a bivariate regression.

Using the drop-down menu, select "Analyze," then "Regression," then "Linear..." which will open a dialog box. Click and drag a variable into the "Dependent" box, but in the "Independent(s)" box, drag and click more than one independent variable from the list at the left into the box. Then click "OK."

7.1 Standardized coefficients

Standardized coefficients appear by default in the output, under the column "Standardized Coefficients: Beta."

7.2 Post-estimation diagnostics in SPSS for OLS

In Chapter 11 we discussed a number of diagnostic procedures that can be carried out once an OLS model has been estimated. These procedures are all available in SPSS . It is important to keep in mind that when asked to conduct such an analysis, SPSS will always do so using information from the last regression model that was estimated.

7.2.1 Identifying outliers and influential cases in OLS

Table 11.9 shows the five largest DFBETA scores from a regression model. In SPSS you estimate DFBETA scores in the following way. In the regression dialog box, click on "Save..." and then find the section on the right-hand side for "Influence Statistics." Checking the box for "DfBeta(s)" will create a new variable with DFBETA values for each case in the data set. After clicking "Continue" and "OK" and estimating your model, click back into the `data set window` and then click on the `Variable View` tab, scrolling down to the bottom, where you will see that SPSS has created new variables with variable labels that inform you which variable is associated with which particular independent variable (and the constant) in your model.[4] You can then click back into the `Data View` window, scroll to that variable, and find the largest DFBETA scores, or, of course, use the commands to provide descriptive statistics for your new variable.

[4] A note of caution: SPSS names these variables differently depending on how many models and independent variables you have estimated.

7.2.2 Detecting multicollinearity in OLS

As we discussed on pages 227 and 228 of *FPSR*, one way to detect multicollinearity is to estimate a Variance Inflation Factor (or "VIF") for each independent variable after you have estimated your regression model. From the regression dialog box, click on "Statistics..." and then check the box for "Collinearity diagnostics." When you click "Continue" and "OK," the model will be estimated. In the output, this will produce the VIF statistic for each independent variable in the model in a column helpfully labeled "VIF."

8 Dealing with time series data before estimating OLS

As we discussed on pages 233-244 of *FPSR*, time series data require careful treatment. In SPSS it is quite helpful to identify a data set as a time series (unless told otherwise, SPSS assumes that the data it is working with are from a cross-sectional data set). From the pull-down menu, select "Data" and then "Define Dates..." which will open a dialog box. On the left-hand side, a scroll-able list labeled "Cases Are" gives you choices (such as yearly, quarterly, monthly) to select the aggregation interval of your data set. Next, on the right-hand side, enter the value for the earliest observation in your data set. Finally, click "OK" and SPSS will assume your data are a time series.

Once you have identified a data set as a time series data set, you can then create new variables to represent lagged or differenced variables. In order to do this, under the menu, select "Transform" and then "Compute Variable..." which will open a dialog box. The key features of this function are the "Target Variable" window in the upper-left corner, where you name your new variable; the "Type & Label..." button, where you should clearly label the new variable; and most importantly, the "Numeric Expression" window, where you instruct SPSS how to compute your new variable. To lag an existing variable called *Approve* by one period, in the "Numeric Expression" window you would type:

```
LAG(Approve,1)
```

Where the '1' indicates you wish to lag the variable approve by just one period.[5] Then click "OK" and SPSS will create your lagged variable.

To create a first-differenced variable for an existing variable named *Approve*, you will follow the same procedure, except in the "Numeric Expression" window, type:

```
Approve - LAG(Approve,1)
```

[5]Obviously, if you wish to create multiple variables for multiple lags of a single variable, you'll need to repeat this procedure for additional lag numbers.

This instructs SPSS to take the current value of the variable *Approve* and subtract it from its lagged value to produce the new "Target Variable." Again, you would click "OK" to compute the variable.

9 Binomial logit and binomial probit

The commands for estimating binomial logit and binomial probit models in SPSS are fairly similar to the commands for estimating OLS models. For binomial logit, from the drop-down menu, select "Analyze," "Regression," and then "Binary Logistic..." which opens a dialog box. Use the arrows to select your dependent variable in the "Dependent" box, and your list of independent variables in the box labeled "Covariates." Clicking the "OK" button estimates the model.

For a binomial probit model, "Analyze," "Regression," and then "Probit..." which opens a dialog box. Place your dependent variable in the usual way in the "Response Frequency" box, and your independent variable(s) in the box labeled "Covariate(s)." Then click "OK."

9.1 Obtaining predicted probabilities for binomial logit and binomial probit models

As discussed in Chapter 11 of *FPSR*, the results from models with dichotomous dependent variables can be interpreted in terms of predicted probabilities. In SPSS this is most easily done for logit models. In the logit menu, click on the "Save..." button, which opens another dialog box. In the "Predicted Values" portion, check the "Probabilities" box, then click "Continue" and "OK" to estimate the model. A new variable with predicted probabilities will be generated, which you can view by clicking back into the `data set window` and then clicking on the `Variable View` tab, and finally scrolling down to the bottom. There you will see that SPSS has created new variables with variable labels that inform you which variable is associated with your logit model.