SPSS Companion for Research Methods

Revised Third Edition

Melanie Arthur
Portland State University

Contents

Preface

PART 1: BASIC MOVES IN SPSS 1

PART 2: CHAPTER KEYED SPSS EXERCISES 31

1-Human Inquiry and Science 31
2-Paradigms, Theory, and Social Research 32
3-The Ethics and Politics of Social Research 32
4-Research Design 32
5-Conceptualization, Operationalization, and Measurement 34
6-Indexes, Scales, and Typologies 36
7-The Logic of Sampling 38
8-Experiments 39
9-Survey Research 41
10-Qualitative Field Research 41
11-Unobtrusive Research 41
12-Evaluation Research 42
13-Qualitative Data Analysis 43
14-Quantitative Data Analysis 43
15-The Elaboration Model 44
16-Social Statistics 45
17-Reading and Writing Social Research 47

Preface

This booklet contains a concise, user-friendly introduction to SPSS Student Version 11.0, with updates for SPSS Student Version 12.0, along with chapter-by-chapter SPSS exercises. The exercises have been organized to correlate with the chapters in Earl Babbie's The Practice of Social Research, 10th Edition, but can be used with any Wadsworth social research methods or statistics text. The data sets referenced in these exercises are available for bundling with Wadsworth's **SPSS Student Version 11.0 CD-ROM and/or 12.0 CD-ROM.** Instructions for accessing these data sets via SPSS can be found in the liner notes accompanying the SPSS Student Version CD-ROM.

SPSS FOR WINDOWS STUDENT VERSION

A GUIDE FOR STUDENTS

PART 1: BASIC MOVES IN SPSS 11.0 and SPSS 12.0

INTRODUCTION

For sociology and the other human sciences, personal computers have become a blessing accompanied by the obligatory curse. The blessing is this: Powerful, user-friendly programs like SPSS have made data analysis a lot easier. Not all that many years ago, the author of this guide was balancing Styrofoam cups of stale coffee and sleepy frustration at 3:00 AM, waiting for a de-bug run from a mainframe. And, yes, we were using early versions of SPSS back in those days. The guides were about the size of city telephone directories. Today, with SPSS versions 11.0 and 12.0, most of what you need to know is actually built into the software. Following the built-in tutorial will give you a rapid course in how to use SPSS. Think of this guide as a bridge. It's designed to match up the material in your text with things you can do in SPSS.

The obligatory curse is actually more like a caution. The ease of using a program like SPSS makes data analysis accessible to those who would as soon not when it comes to math and statistics. That's not such a bad thing, but you need to remember that just as there was adding-in-your-head before calculators, there was a whole lot of intensive calculation labor and concept-learning involved in data analysis back in the old days. Today, the tedious calculations are gone, but you still have to know the logic of sampling, how to collect data, and how to tell which statistical test is appropriate for a given level of measurement. And, if you're going to ask SPSS for a correlation coefficient, you still need to know where those things come from and what they tell you.

LEARNING OBJECTIVES:

What You're Supposed to Get Out of Part 1 of This Guide

→ When to use SPSS
→ Basic navigation in SPSS
  A. Getting in
  B. Understanding the SPSS Data Editor and dialog boxes
  C. Understanding the use of basic tools that accompany the Data Editor and selected dialog boxes
  D. Getting out without losing what you've done
→ How to enter your own data into SPSS
→ How to define variables for SPSS
→ How codebooks are used to convert survey data into numbers
→ How to create, save, and retrieve SPSS files
→ How to produce and interpret basic data summaries
→ How to transform variables to make them easier to analyze and compare using the
→ Understand what the General Social Survey (GSS) is and learn how to import a data file into your SPSS Data Editor.
→ How to produce a bar graph from one or more variables

NAVIGATING IN SPSS USING A PRACTICE DATA SET (PDS)

Let's say you've gone through the irritating stuff that we can't put in this sort of guide. You've got the student version of SPSS 11.0 or 12.0 installed on your computer. Maybe you used the installation program in your Windows 95, 98, or 2000. Or maybe you've got SPSS 11.0 or 12.0 installed on a bunch of lab computers. SPSS 11.0 and 12.0 are very similar versions of the software, though version 12.0 is somewhat newer. If you are not certain which version you are using, follow the directions in the textbox on page seven (7) of this manual to find out.

SPSS includes access to a set of files from the General Social Survey (GSS). You will also be using another GSS data set that's included with this manual. We'll tell you how to get at those files as we proceed. You'll use them to pursue certain prescribed exercises for this guide, for exploration purposes guided by your curiosity, and, no doubt, as required by your instructor.

Some Key Terms

Before we get started with a sample data set, you need to understand a couple of key terms. The first is a "variable". A variable is the basic unit of study in statistical analysis. It can be something as simple as height or weight, or as complex as life satisfaction. But all variables have two things in common. First, they can all be measured or quantified. Second, they all vary (what a surprise!).

In other words, they take on more than one value.

The second term you need to be familiar with is “level of measurement”. This refers to the measurement characteristics of the variables being studied. All variables fall into one of three levels of analysis. The first is a “nominal” level variable (sometimes called a *categorical* variable). Examples of this level of analysis would be hair color (blond, brunette, purple) religion (Catholic, Protestant, Jewish, Moslem) or favorite flavor of ice cream (vanilla, chocolate, coconut). These variables only vary qualitatively. No category is higher or lower than another. They’re just different. The second level of analysis is an “ordinal” level variable. These are rank ordered variables that can be arranged from high to low. So variables like job satisfaction (very satisfied, somewhat satisfied, miserable) or educational attainment (grade school, high school, college) would be examples of this level of measurement. The third level of measurement is an “interval-ratio” level variable. The values of these variables have real meaning. Height or weight would be good examples of interval ratio variables. So would income, or the score on your next test. It’s important to be able to identify the appropriate level of measurement for a given variable, because the kind of statistical test you apply to any kind of data depends on its level of measurement.

It’s important to understand that the same variable can be examined at different levels of measurement. Let’s take education as an example. It can be measured nominally (private or public school), ordinally (B.A., M.A., PhD) or as an interval ratio variable (years of education). It seems a little complicated at first, but you’ll get the hang of it as you start handling the data we’ll be working with. In your text, the way to deal with nominal, ordinal, and interval/ratio data is explained for you in terms of the kinds of statistical tests you can use for each type of data.

The last concept you will need to know is “measure of association.” Measures of association are statistics that tell us the extent to which two variables are related. Two of these statistics are used in this workbook; lambda and gamma. Lambda is used for nominal level variables while gamma is used for ordinal level variables. The values for lambda range from 0 to 1. The closer lambda is to 1, the more strongly the two variables are related. Gamma ranges from -1 to +1. Again, the closer gamma is to one, the stronger the relationship is between the two variables. The sign for gamma tells us whether the variables are positively or negatively related.

A Practice Data Set

To get you used to some basic moves with SPSS, we’ll start with having you enter a small data set. It’s *Display 1* on the next page. Take a look at it.

You’ll see that the Practice Data Set (PDS) includes 30 cases (n = 30) and four (4) variables. As you will know by now, a case is one of the individual "somethings" you're studying, depending on your *unit of measurement*. You could be studying PTA groups from a sample of schools, or newspaper editorials. In this case, our unit of measurement is people--30 college students--so each case

DISPLAY 1

PRACTICE DATA SET (PDS)

For Variables: SEX, ACADRANK, CONTROL, and LIFESAT
(n = 30)

Variable Name:

Case	SEX	ACADRANK	CONTROL	LIFESAT
1	2	4	3	4
2	1	2	4	5
3	1	1	3	3
4	1	1	2	3
5	2	2	3	4
6	2	3	4	5
7	2	3	5	5
8	1	1	3	3
9	2	1	2	2
10	2	1	3	3
11	1	2	3	3
12	1	3	5	4
13	1	2	3	3
14	2	3	5	4
15	1	3	4	5
16	2	4	3	4
17	1	2	1	2
18	2	1	2	3
19	2	1	3	2
20	1	3	4	4
21	1	2	3	4
22	1	3	5	5
23	2	3	4	4
24	1	2	2	3
25	1	1	2	2
26	2	1	3	4
27	2	4	4	5
28	2	2	3	4
29	1	1	2	3
30	1	2	4	5

is one person. The four variables are SEX, ACADRANK, CONTROL, and LIFESAT. LIFESAT stands for "Life Satisfaction," or how happy a person is with their life (5=strongly agree, 4=agree, 3= undecided, 2= disagree, 1=strongly disagree). CONTROL is presumed to be a measure of how much control a person feels they have over their life (5=strongly agree, 4= agree, 3= undecided, 2=disagree, 1=strongly disagree). ACADRANK stands for academic rank –freshman (1), sophomore (2), junior (3), or senior (4). You know what SEX stands for and that it varies between only two values or attributes--namely, male (1) or female (2).

Level of Measurement

In our practice data set, SEX is nominal, LIFESAT and CONTROL are ordinal level data, and ACADRANK is interval-ratio. (SPSS refers to interval-ratio data as "scale" data for technical reasons that aren't really important for you to understand right now.)

Variable Language in SPSS

You'll notice that each *variable name* in SPSS has eight--or fewer than eight--letters in it. Academic rank, for example, becomes ACADRANK. The eight-space variable is an old tradition required by computers to keep them happy. SPSS 12.0 has eliminated this limitation (allowing up to 64 characters), however, SPSS 11.0 still requires variable names to be a maximum of eight characters.

Being forced to condense variable names into eight letters may be a challenge--especially if you want to remember what they are. Fortunately, we can also tag a variable name with a *variable label* that can take up as many as 255 spaces--if we are really long-winded. SPSS makes that easy to do. And, once it's done, it's easy to access the variable label for a particular variable. You'll see how that works as you go through this guide and the SPSS Tutorial.

Getting a Grip on Coding Data

The Practice Data Set is pre-coded. That is, the information for each case has already been converted to numbers. Computers are number manipulators at heart, which is why they got the name computers. Powerful software--like SPSS--handles string (also called alphanumeric data) very comfortably. It's happy to read a datum like, "Irish Catholic." But you might want to keep in mind that it does so by making slick, fast conversions--assigning numbers to letters and back again--or associating numbers with strings of letters that we read as words.

In any case, in spite of programs like SPSS, we social science folks usually design our questionnaires or surveys with an eye to converting data into numbers. That's what we've done with your practice data set. Without looking at our data in number-form we cannot run statistical tests. If we can't run statistical tests, we can't test hypotheses. However, our numbers often represent qualities, not simply quantities. Sex, for example, is coded 1 for "male" and 2 for "female". The numbers 1 and 2 represent categories or qualities, rather than quantitative differences. This is also true for LIFESAT and CONTROL. As you go through this exercise, you'll see how SPSS makes deft, easy-to-access associations between numbers and words that signify ideas, opinions, or attitudes.

The Point of the Data Set

Why would anyone collect data like this? Here are some possible answers: (1) To see if level of education has an effect on life satisfaction; (2) to see if level of education affects the sense of control one has over one's life; (3) to see if the sense of control one reports having over one's life has an impact on life satisfaction. Maybe you can think of some other possibilities. If you were the one who made up the questionnaire, selected the sample, and collected the data, you'd probably have some clear objectives (research questions) in mind. Otherwise, you'd be doing a lot of hard work for nothing.

In fact, the Practice Data Set we are using here is contrived. It is not from an actual research project. However, it *is* based on various studies the author of this guide has conducted with college-student samples in the context of research methods classes.

Assumptions About the PDS

Let's make some assumptions about this made-up data set. We'll pretend that the PDS is taken from questionnaires administered to selected students over a period of three weeks. The original questionnaire included many more items than we are looking at here. To keep life simple, we will examine only four variables for 30 cases (n = 30). Think of these data as derived from a research project conducted at a state university population in an Eastern region of the United States. Assume as well, that the 30 cases represent the student population of State U. with respect to variables that interest us. So, the ratio of males to females in the sample, as well as the proportion of freshmen, sophomores, juniors and seniors, is roughly similar to what we find in the entire student population. We have chosen these four variables because we want to test hypotheses about the relationships among them. For example, we might hypothesize that life satisfaction is affected by a person's academic rank. In this case, life satisfaction would be considered the *dependent variable.* This is the variable we're interested in understanding. Academic rank is termed the *independent variable*, or the variable that has an effect on the dependent variable. We might note that on a blackboard or scratch pad as:

X acadrank → **Y** lifesat

Of course that's just a hopeful notation. As you have or will learn in your studies, the only way we can establish that X *causes* Y is in an experimental design. Among the other major kinds of social science research, such as survey research, the best we can hope for is a demonstration of the *probability* that X and Y are related.

Codebook Excerpt

In *Display 2*, you'll see a portion of the questionnaire codebook used to translate student responses into numeric data (numbers). You will want to give it some attention. Type in italics is from the actual questionnaire. Non-italicized type is part of the codebook.

These code books are given titles in SPSS to make them easy to handle. This PDS might be named dataset.rp4 (where .rp4 stands for research project 4). *Display 2* shows you what a portion of the codebook we're concerned with might look like. *You will use these codes for entering the PDS into your own SPSS file*. It's not here simply to annoy you.

Getting In to SPSS and Meeting the Data Editor

The directions for starting up SPSS differ only slightly depending whether you are using a student version of SPSS that you have installed on your computer, or whether you are using a full version of SPSS, as you might find in a computer lab. Either way, turn on your computer, access PROGRAMS from the START menu, and look for SPSS in the flip-out box. For students versions, look for SPSS 11.0 FOR WINDOWS STUDENT VERSION or SPSS 12.0 FOR WINDOWS STUDENT VERSION. For full versions of SPSS, you should find a folder entitled SPSS for Windows, which will open another flip-out box, allowing you to click on either SPSS 11.0 for Windows or SPSS 12.0 for Windows.

<u>Determining what version of SPSS you are using</u>
You may already know what version of SPSS came with your text because you installed it on your computer. However, if you find yourself in a computer lab it may not be so immediately obvious whether you are using version 11.0 or version 12.0. The first way you can find out, is by checking which version is listed on the start menu when you opened the program, following the directions above. If you have already opened the program, you can go to the Help menu, choose, About… and you will find the version listed on the resulting display box.

DISPLAY 2

FRAGMENT OF A RESEARCH QUESTIONNAIRE CONVERTED TO A CODEBOOK

STUDENT SURVEY

Please respond to all of the items. Do NOT write your name or student ID on the questionnaire. No effort will be made to identify individual respondents. Only summaries of the data will be made available to others and your individual responses are considered private and confidential.

Item 1. Please mark the appropriate space: Female ___ Male ___

(Col, 1, SEX) Male = 1; Female = 2

Item 2. Please mark your present academic rank:
Freshman ___
Sophomore ___
Junior___
Senior ___

(Col. 2, ACADRANK) 1 = Freshman; 2 = Sophomore; 3 = Junior; 4 = Senior

Part 2.

Please respond to the following statements by specifying the extent to which you agree or disagree with it. Let: SA = "strongly agree" D = "disagree"
A = "agree" SD = "strongly disagree"
U = "undecided"

Item 3. I control the circumstances of my life.
Please circle your response: SA A U D SD

Item 4. I am satisfied with my life.
Please circle your response: SA A U D SD

(Cols. 3 - 4. Likert format variables CONTROL thru LIFESAT) 1 = strongly disagree; 2 = disagree; 3 = undecided; 4 = agree; 5 = strongly agree.

The first display you see will be the Screen Display for SPSS (which will also provide the version number). It will linger briefly. Then you'll see a box entitled *SPSS for Windows Student Version* with a menu and a files window in it. That opening box is on top of a bigger window called the SPSS **Data Editor**. The Data Editor is your basic work place for entering data or doing things to data already entered. Just now, the Data Editor should be empty and waiting for you to do something with it. But, of course, you have to deal with the opening menu box that's in front of the Data Editor. It asks:

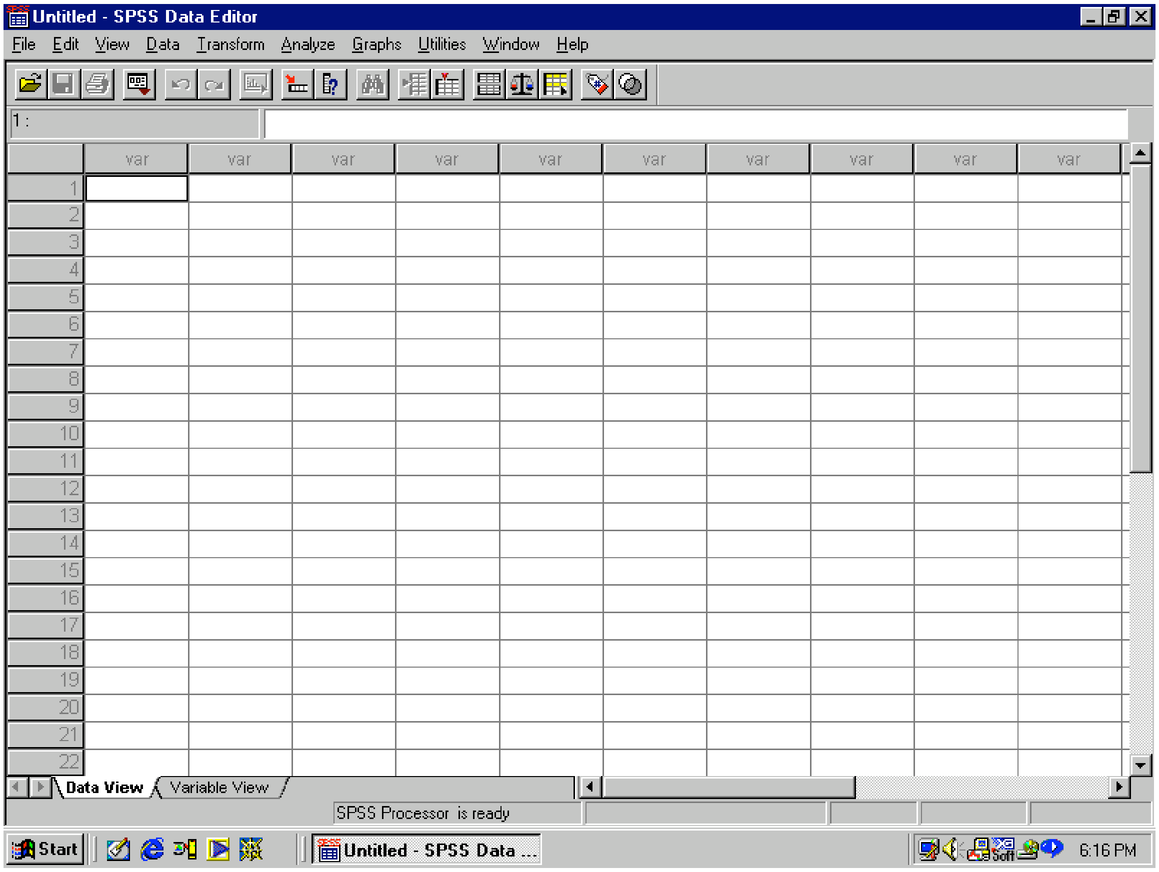
> What would you like to do?
>
> O Run the Tutorial
> O Type in Data
> O Run an Existing Query
> O Create a New Query Using Data Base Wizard
> O Open an Existing Data Source

In the opening menu box, click on *Type in Data*. Make sure you've left a fat dot in the little porthole next to *Type in Data*, then click *OK.*

The opening menu box also lets you delete that box when you next enter SPSS. DON'T DO THAT. You need that menu box until you become a data savant. Meanwhile, however, take a lesson from that menu box--as well as from the Data Editor when first you enter it. SPSS has lots of tools and options to choose from. That, in itself can be confusing. Remember: Most of the time, you only need a few tools and a few options. This fact requires no guilt.

You are now in the Data Editor. (Shown in *Figure 1*.) The page you are on is called the *Data View* page (it's labeled at the bottom left corner). If you click on *Variable View*, it will take you to a different page. Look at the main body of the display. SPSS organizes data with variables in the columns and cases in the rows. The top of each variable column is labeled *var* (for *variable)*. That's because none of the variables have been named or defined yet. The Data Editor allows you to enter large numbers of variables and large numbers of cases. In other words, SPSS handles pretty large samples with considerable ease. In the student Version, you can handle up to 1500 cases and 50 different variables in a single data set. Full versions are limited only by the processing capabilities of your computer.

Figure 1. SPSS Data Editor



To explore the pull-down tags lined up over the tool bar, like File or Data, click on any of them to see what's in the drop-down menu. To explore the tool bar icons at the top of the Data Editor, rest your mouse arrow on any icon to get a pop-up explanation in a little yellow box. Do that for a while before returning to this guide if you like. You'll notice that the tool bar items and the drop-down menus are often redundant--you can get at the same operation from more than one place on the screen.

Preparing to Enter the First Ten Cases

Look at the Practice Data Set (PDS). Prepare to enter the data for the first ten cases in the set. You'll be filling in ten rows and four columns for a total of 40 data boxes. Take the data from *Display 1* on page 3, directly to the Data Editor *in the same order*.

But first, try the short exercise found overleaf.

Data Boxes: Quick Exercise

If you are not sure what a data box is, move your mouse arrow around in the main body of the Data Editor. Click on any box. You'll get a heavy-line border around that box. You've designated a space you want to enter data in. If there's already data in that box, you've specified it in order to change that data--if you want to. To conclude this exercise, click on case-row number one under *var* column 1. You're at a good starting place for entering the first ten cases.

PDS Exercise 1: Entering Ten Cases

Click on the data box for case 1, under column 1. You can enter data directly into a data box. You can also click on the data entry panel at the top of the Data Editor and type in the data you want to appear in the data box you've marked. Try it both ways. Type the numeral 2 in the data entry panel. Press ENTER. 2.00 will appear in the marked data box and the case 2 box underneath it is now marked with a heavy border. Simultaneously, the dimmed *var* label at the top of the column is now darker and reads: *var00001*. Return to the entry panel and enter the numeral "1" in the entry panel. Press ENTER. 1.00 appears in the second case data box under *var00001*. Now go ahead and enter the data given in Display 1 directly into marked boxes, press ENTER and see how that works. If you make a mistake, click on the erroneous box, enter a correction in the entry panel or "write over" your error and press ENTER. Presto. It's done. Keep entering data until ten cases have been entered for all four variables in the Practice Set. When you finish, you should have data for ten cases entered under variables named *var00001* through *var00004*.

Check What You've Entered

Entering data inevitably leads to errors. Data checking--also called *data cleaning*--is an important part of data entry. You'll learn all about it in your text. Of course, you are not likely to have made errors so far. Even so, check what's in the Data Editor at this point to see if it matches the Practice Data Set for cases 1 through 10. Now we get to make sense of what we've just done and turn it into information that someone else can understand. But first, we'll save what you've already done into an SPSS system file. A *system file* is a data set that has been organized by SPSS so that the data can be analyzed.

PDS Exercise 2: Naming, Saving, and Retrieving Your File

As we go along, we'll look at the general business of getting files. Right now, here's how you name, save, and retrieve your current file.

STEP 1. Mouse-click on the *File* option at the top of the Data Editor. Select *Save* from the drop down menu. If you have not named a new file, a box appears asking you to name your file. That should occur in this case. Write in any name that appeals to you, though shorter file names are easier to use and remember. For example: MYFILE.sav. The four character file tag ".sav" will be added automatically by SPSS, allowing both you and SPSS to always know that this is an SPSS data file. When you are playing around with a data set, you might also want to save different versions of it--like myfile1.sav or myfile2.sav. You'll see why in more detail later on.

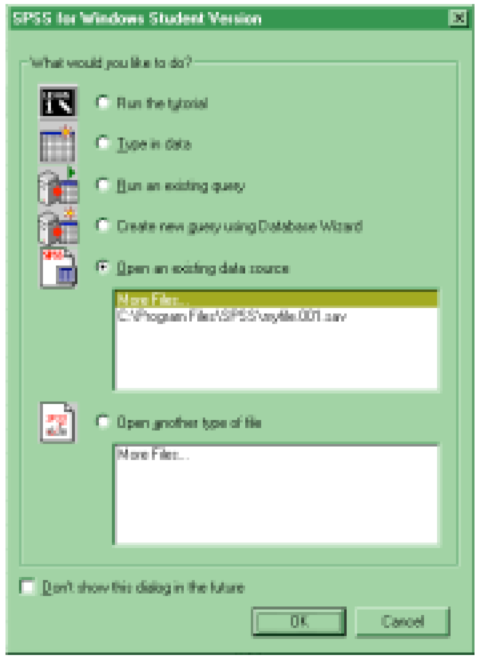
STEP 2. In the *Save File* box Click on the *OK* button. The file you've created with ten cases worth of data in it is now saved in the SPSS data file folder.

STEP 3. Exit SPSS by clicking on the window close-box or by selecting *Exit* from the *File* drop-down menu or closing the SPSS window. Take a break.

STEP 4. Re-enter SPSS. Look at the opening menu box.

STEP 5. The first time through, we had you mark *Type in Data*. This time you need to retrieve your file. Look in the first window at the bottom of the dialog box. Your file will be there. After you've done some work with SPSS, there will be lists of files in your opening menu boxes. Data files will be in the top panel and output files will be listed in the bottom panel. An output is the display you get in a viewing window after you've performed a data operation on a data set.

Figure 2. Opening Menu Box:
SPSS 11.0 for Windows Student Version



You select your file by double clicking it or by selecting it and clicking OK. The menu dialog box will go away and, after a brief pause, your file data will appear in the Data Editor. Notice that the name of your file will be shown at the top of the Data Editor. You're ready to go back to work.

Later on, when you've got several files to look at, you can go from whatever file you are working on in the Data Editor to a *New* File or you can *Open* an existing file. For example: Click on the *File* option. Select *Open* from the drop-down menu. You'll get a pop-out menu offering you three options, namely: *Data*, *Output*, or *Other*…. (This is true in student versions. In Full versions of SPSS there are additional options for *Syntax* and *Scripts*, which will not be needed or used in this guide.) On selecting *Data*, you'll open a window with a set of data files that accompany SPSS. We'll be using several of these data sets as we proceed through this guide. You'll also see any data files you've prepared, such as myfile.sav. Selecting *Output* opens a window with all your output files. At present, of course, that window will be empty. If you're curious about the *Other…* option, pursue that through your SPSS Tutorial. It isn't needed for the exercises in this guide.

PDS Exercise 3: Taking a Variable View in the Data Editor: Defining Data

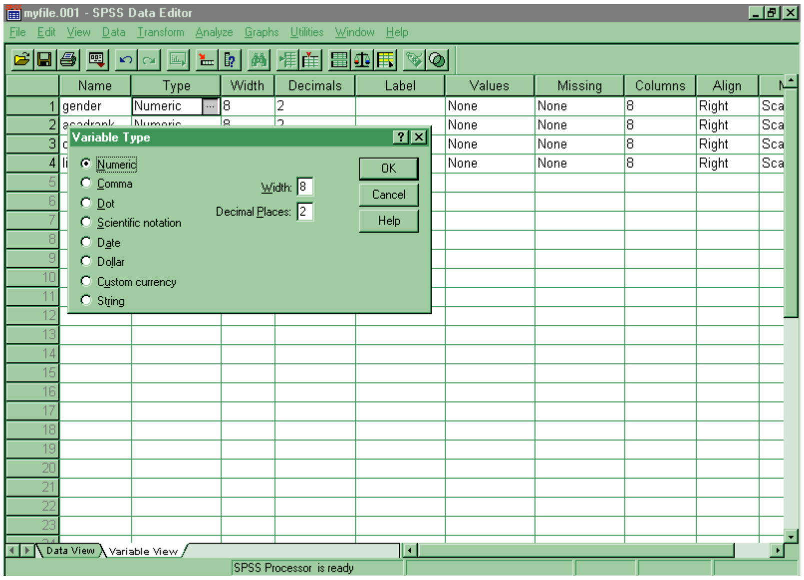
You will find that entering and working with numeric (number) data is pretty simple. But before you can do anything sensible with the data, you have to tell SPSS what you are up to. You also need to organize your data in the Data editor so you know what it means. As it turns out, SPSS allows you to define your data by simply shifting your point of view in the Data Editor and going through a few simple steps.

To see how this is done, look at the bottom of the bottom left base of the Data Editor and notice two tabs, one marked *Data View*, the other marked *Variable View*, which we already pointed out. (You might also have noticed these tabs in Figure 1.) With your first ten PDS cases in the Data Editor, amuse yourself by switching back and forth between the two tabs by mouse-clicking first one, then the other. See what happens. In the Data View, you'll see your four cases stacked on the left. In the Variable View, you'll see each of the four *variables* in a stack to the left of the Editor. While in the Variable View mode, note the column headings. These are, from left to right: Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, and Measure. That's a total of 10 column headings, which amounts to a total of 10 procedures you will need to follow to define your data.

1.) To assign each of your four variables a title with eight or fewer letters--called a **variable name**, simply mouse-tag the data block currently called var0001. Doing that will, of course, produce a bold border around the data box. Type the word "gender." (Just type it on the keyboard.) That name will magically appear in the data box, replacing the default variable name, var00001. Using the same procedure, enter "acadrank" in place of var00002, "control" in place of var00003, and "lifesat" in place of var00004. While SPSS 12.0 would allow you to use longer variable names, if you keep to the shorter names given here you can move back and forth between versions, if necessary.

2.) To define a variable as to its type, click on the top box under the column **Type** for your newly named variable, GENDER. This time, along with the bold border, you will also see a shaded button at the right of the data block. Click the button and a dialog box appears atop the Data Editor labeled *Variable Type*. It's represented in Figure 3, below. You may be surprised at the number of data types handled by SPSS, but, having stifled your amazement, simply mark the "numeric" option. Notice that you can also use this dialog box to set the width and decimal places for your variable. You should find this useful later on. For now, click OK, and proceed. Use the same procedure for your remaining three variables. Recall that all of them have been pre-coded into a numeric format. Mark them all numeric.

Figure 3: Type dialog box



3.) A data field is divided up into columns. The *width* of a data field is the number of columns required to fit the data. A numeric datum like one's age, for example, might be entered as 17, requiring a width of two columns. A datum like a mean word count for a sample of newspaper editorials might require a four-column width, and a ratio might require 1 column to the left of a decimal point and three to the right of it for a total column width of 4. In SPSS, the default setting for numeric data in the boxes stacked under any variable is 8 spaces for numbers followed by 2 decimal places. As it turns out all of the variables in our practice set are numeric data that require only 1 space (for a single number) and no decimal places. Click on the data block under **Width** for the PDS variable GENDER. This time, you'll get a button with "up" and "down" arrows. Use them to select the numeral "1." Use the same procedure to assign a one-column width to each of your PDS variables.

4.) Often numeric data includes decimals. When you click on a data box under the **Decimals** column for any variable, you'll get another "up-down" button. Use it to enter "0" for each of the four PDS variables. We require no decimal places for the number values assigned our variables.

5.) Because variable names are often too cryptic to be self-evident, you'll give each variable name a variable label. A variable label can be about as long as you want it to be—up to 225 characters—but the point of the variable label is to make clear what a variable is all about. To write in a variable label, select a data box under the column **Label** in the Data Editor. This time, you'll get a bold border around the data box that allows you to type in an elaboration or explanation. The procedure is the same as that used for entering variable names (Procedure 1.) For variable GENDER, type in "Gender." The term is self-evident, but the computer needs the variable label anyway to make you output displays easy to read. For variable ACADRANK, type in "Academic

Rank." For variable CONTROL, type in "Control Over One's Life." For variable LIFESAT, type in "Life Satisfaction."

6.) Especially when you want to read output, such as frequency tables or comparative analyses among variables, it's very useful to make your variable attributes evident. Instead of "2," for example in an output controlling for gender, it will be much easier to read "female." The **Values** column in the Variable View display allows you to write in your appropriate value labels. Remember that each number you entered in a data box is a code for something. For the variable GENDER, 1 stands for "male" and 2 stands for "female." For the variables CONTROL and LIFESAT, both ordinal variables dealing with levels of agreement to a statement, 1 stands for "strongly disagree," 2 stands for "disagree," and so on. To begin assigning value labels to the number-coded attributes for your PDS variables, click on Values in the GENDER row. You'll get a button to the right of the block. Click it to open the *Value Labels* dialog box shown below.

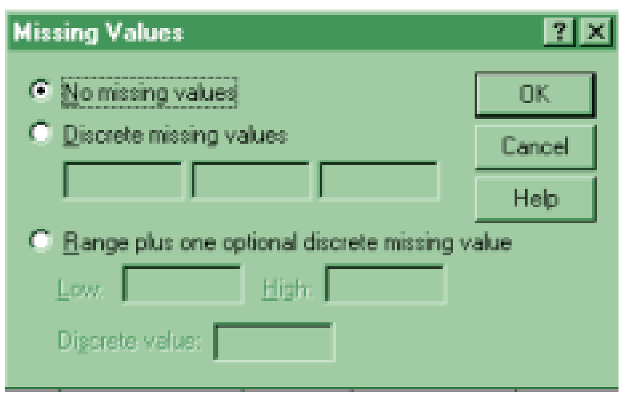
Figure 4. Value Labels Dialog Box

For the variable GENDER, write the numeral "1" in the upper left panel of the box labeled *Value*. In the *Value Label* panel directly below it, write "male." (Write the numeral and the word without quotation marks, however.) Now, press the ADD button on the dialog box and watch *1 = male* appear in the larger panel of the dialog box. See how it's done? Refer to your Codebook excerpt as needed and follow this same procedure to enter value labels for all your PDS variables.

For variable ACADRANK assign value labels as follows: 1 = freshman; 2 = sophomore; 3 = junior; 4 = senior. For CONTROL and LIFESAT assign: 1 = strongly disagree; 2 = disagree; 3 = undecided; 4 = agree; 5 = strongly agree.

7.) You won't have to deal with missing values in the PDS exercises, but you need to know how to provide that information to SPSS. Click on one of the boxes under the **Missing** column while in *Variable View* in the Data Editor and notice that a button is provided for you. Clicking the button opens a dialog box for missing values like the one illustrated in Figure 5.

Figure 5: Missing Values Dialog Box



The way you will go about dealing with missing values depends on the most reasonable strategy for your sample and the nature of your data. Sometimes it seems best to throw out the entire case, with all associated variables. In other instances it seems best to assign a likely value or a mean value, especially if the sample is fairly large. In any case, the art of handling missing values is discussed in your text and we will not give it much attention here. All you need to do for these PDS exercise variables is mark the porthole for "No missing values," press OK, and proceed to the next Column in Variable View, which, in fact, is the "Columns" column.

8.) The **Columns** options, provided by an "up – down" button, allows you to decide how wide your data blocks should be while in Data View. Eight is a good

standard decision. Select 8 for each variable. Or, if you like, select different numbers and flip back and for between the variable and data view modes to see what happens.

9.) Under the **Align** column while in *Variable View*, click on the first block, then on the button that appears, and notice your options. You can choose to align your datum to the left or to the right of a data block. Or, you can have it placed in the center of a data block. It's up to you. Experiment a bit, flipping back and forth between data and variable view modes to see what happens. Personally, I think centering the data is easiest to read.

10.) To assign a <u>level of measurement</u> (discussed on page 6) to each of your variables, click on any block in the **Measure** column while in Variable View. You are given three options when you select a block and tag the button. Choose the one that applies for that variable. In this case, GENDER is nominal and ACADRANK is scale, while CONTROL and LIFESAT are ordinal. ACADRANK could be viewed as ordinal or even nominal, depending one's frame of reference. In this case, let's assume that precise numbers of credit hours earned, taken as a ratio to total credit hours required for a college degree give us scale data.

PDS Exercise 4: Enter All 30 Cases from Display 1

In order to see how to explore your data, we first have to finish entering it. You know how it's done. Using a ruler and a sharp eye, enter all the data from the Practice Data Set found in *Display 1* on page 3.

Data Cleaning

There are a variety of ways to clean data. In the case of your PDS it's a pretty simple task. Simply go over the data you've entered in the Data editor and make sure you entered it correctly. If you want to review the principles of possible code and contingency data cleaning at this point, go to your text and re-visit the appropriate discussion.

DATA ANALYSIS EXERCISES

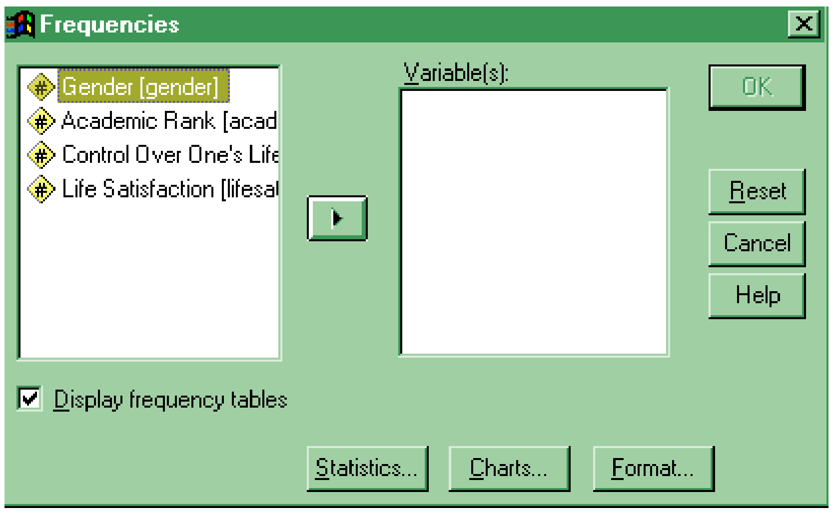
PDS Exercise 5: Frequencies

You can perform data analysis on any file you have opened in the Data Editor. One of the first and simplest things to look at for any variable is a summary count for each value of your variable. That means running *frequencies* for a particular variable or set of variables.

Here is how you can look at the frequencies for the variables in the Practice Data Set.

STEP 1. Select *Analyze* from the option strip at the top of the Data Editor. From the drop-down menu, select *Descriptive Statistics*. A pop-out menu will appear to the right of the drop-down menu. Click on Frequencies. A dialog box will appear that looks like what you see in *Figure 6*.

Figure 6: Dialog Box: Frequencies



In the window to the left of the box is a list of your variables. If the list is long, you can scan down the window to see what's there. If you rest your mouse arrow on a variable you'll get a small description box giving the full variable label for that variable. In this case, you will only see the four variables from our PDS.

STEP 2. You begin by selecting variables. Select GENDER by mouse-tagging it. A dark blue strip illuminates it. Press the arrow button to move GENDER into the open window at the right of the box. If you make a mistake, you can click on the variable name you've just moved into the selection window and it will return to the variable list window. Or you can press the *Reset* button and all the variables in the selection window will go back into the variable list window. This reset feature applies to many of your SPSS dialog boxes.

STEP 3. Do the same thing for the three remaining variables. Select the variable and press the arrow button until all four PDS variables are in the box to the top right of the Frequencies dialog box.

STEP 4. Press *OK.* After a brief interlude, an output display will open up in a viewing window. You can scan through it to see what you have. Notice that most of the tools available in the Data Editor are also available in a viewing window.

Display 3: Frequencies Output

Frequency Table

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	16	48.5	53.3	53.3
	female	14	42.4	46.7	100.0
	Total	30	90.9	100.0	
Missing	System	3	9.1		
Total		33	100.0		

Academic Rank

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	freshman	10	30.3	33.3	33.3
	sophomore	9	27.3	30.0	63.3
	junior	8	24.2	26.7	90.0
	senior	3	9.1	10.0	100.0
	Total	30	90.9	100.0	
Missing	System	3	9.1		
Total		33	100.0		

Control Over One's Life

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly disagree	1	3.0	3.3	3.3
	disagree	6	18.2	20.0	23.3
	undecided	12	36.4	40.0	63.3
	agree	7	21.2	23.3	86.7
	strongly agree	4	12.1	13.3	100.0
	Total	30	90.9	100.0	
Missing	System	3	9.1		
Total		33	100.0		

Life Satisfaction

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	disagree	4	12.1	13.3	13.3
	undecided	9	27.3	30.0	43.3
	agree	10	30.3	33.3	76.7
	strongly agree	7	21.2	23.3	100.0
	Total	30	90.9	100.0	
Missing	System	3	9.1		
Total		33	100.0		

You can print your output by selecting the print icon on the tool bar in the viewing window or by selecting *Print* from the drop-down menu after you select *File*. From most laser printers, the output should look like what you see in *Display 3*. Look through the display to see how it summarizes responses to the Practice Data Set questionnaire.

Closing a Viewing Window

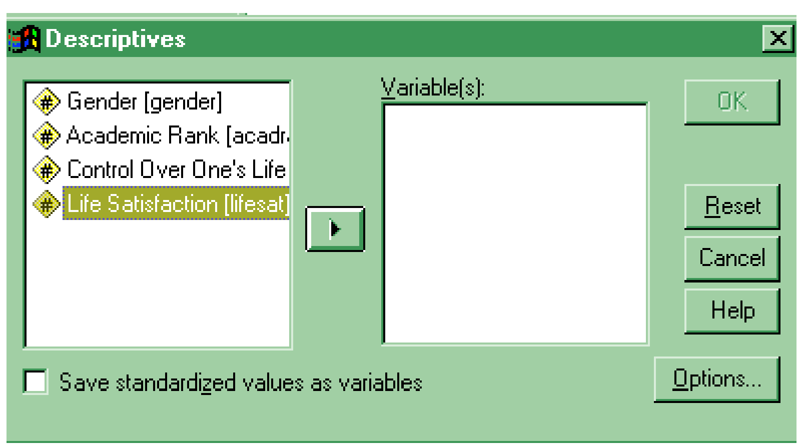
When you decide you've had enough of the display in a viewing window, you can save it or let it go. To save your file, select *Save* on the viewing window tool bar. A window will appear that is the sort one sees in Windows 95 or 98 for saving or getting at files. In the narrow panel at the bottom of the window, write in the name of the file you want.

If you press *Close* before saving, you'll get a warning message asking you if you want to save your file as, for example, *Output 1*. If you press *NO*, all the data in the viewing window goes to cyber-heaven and is lost. If you want to save it, use the default file name (*Output 1*, or 2, or whatever). Or, write in any file name you like (with eight characters in it). SPSS will automatically add a file tag of ".spo" , which always denotes an SPSS output file. Press *OK*. The output file will be saved in the SPSS data file folder and you will be returned to the Data Editor. Remember that the output and data files are saved separately in SPSS, so saving your output window will NOT ensure that your data is also saved.

PDS Exercise 6: Looking at Means

It's often useful to look at the mean responses to your variables in order to see how you might want to proceed in analyzing your data. You can do that in several ways, but for this exercise we'll simply choose the Descriptives alternative provided by SPSS. To look at the means for your PDS variables, here's how you proceed:

Figure 7: Dialog Box: Descriptives



STEP 1. For Version 11.0: From the Data Editor (with the PDS file in place, of course), in either the data or the variable view mode, select *Analyze.* From the drop-down menu, select, *Descriptive Statistics.* From the pop-out menu, select, *Descriptives*. A dialog box appears that looks like Figure *7.* **For Version 12:** Proceed as for Version 11.0, but it is not necessary to be in the Data Editor window to access the *Analyze* menu.

STEP 2. Move the variables you select from the panel to the left of the box over to the panel at the upper right of the box. You're using basically the same procedure you used before in the *Frequencies* box. While in this dialog box, after you've moved all four variables to the right-hand box, press *Options* to open a second dialog box. This one gives you the different statistical measures you can select. Check the boxes for minimum, maximum, mean, and standard deviation. Press *Continue*, then press *OK*. The viewing window that opens will give you a display that looks about like what you see in *Display 4*.

Display 4
DESCRIPTIVES SUMMARY FOR THE FOUR PDS VARIABLES

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Gender	30	1	2	1.47	.51
Academic Rank	30	1	4	2.13	1.01
Control Over One's Life	30	1	5	3.23	1.04
Life Satisfaction	30	2	5	3.67	.99
Valid N (listwise)	30				

It's important to note that the computer will calculate any statistic you want, even if it doesn't make any sense. This is referred to as "garbage in – garbage out". And here's a good example. The mean for SEX is 1.47. The mean for ACADRANK is 2.13. Does it make sense to talk about an "average" sex? Not likely! If you realized these means would be silly, useless, or both, and you rebelled against the instructions in STEP 2, give yourself a small gold star. Otherwise, receive the cautionary lesson. The computer will do what you tell it to do. If you ask silly questions, you'll get silly answers. The means that actually make sense here are those for CONTROL and LIFESAT.

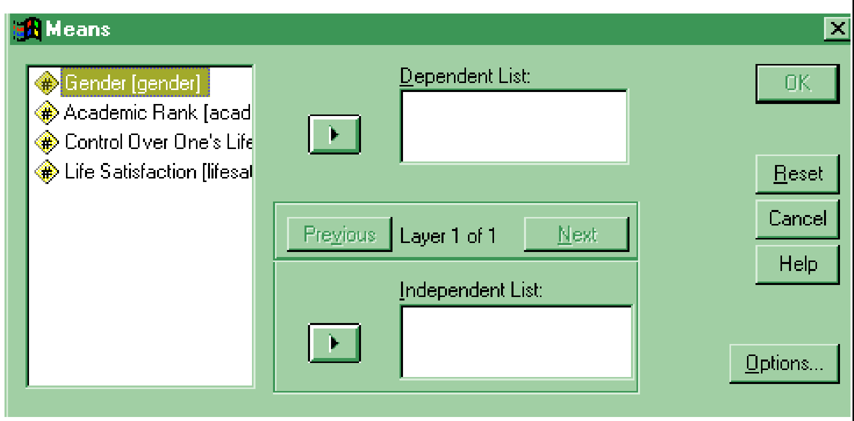
PDS Exercise 7: Comparing Means

You often want to see if mean responses on a variable vary with some selected independent variable. In the case of the PDS, GENDER and ACADRANK are most likely to be selected as independent variables. Why? (If you're still unsure about independent and dependent variables, look at Chapter 8 of *The Practice of Social Research*, 10th Edition). Because we might hypothesize that the sense of control one has over one's life and reported life satisfaction may be dependent on gender or academic rank--or both.

Let's see how academic rank is ordered with respect to the mean responses to CONTROL and LIFESAT. Here, we'll think of these subjective *attitude measures* as dependent variables.

STEP 1. With you PDS data set in place in the Data Editor, select *Analyze*. From the drop-down menu, select *Compare Means*. From the pop-out menu, select *Means* and you'll get the dialog box shown in *Figure 8*. You'll use it to produce what you see in Display 5.

Figure 8: Dialog Box: Means



STEP 2. Using the procedure you are now familiar with, select ACADRANK and move it to the panel for independent variables. Next, select and move CONTROL and LIFESAT to the panel at the top of the box for dependent variables. Press *OK*. The output you get in the viewing window will look like *Display 5*.

Looking at the means, does it seem that juniors and seniors have, on average, higher LIFESAT and CONTROL scores than freshmen and sophomores?

Display 5
MEANS: CONTROL AND LIFESAT BY ACADRANK*
(*Case Processing Summary is excluded.)

Report

Academic Rank		Control Over One's Life	Life Satisfaction
freshman	Mean	2.50	2.80
	N	10	10
	Std. Deviation	.53	.63
sophomore	Mean	2.89	3.67
	N	9	9
	Std. Deviation	.93	1.00
junior	Mean	4.50	4.50
	N	8	8
	Std. Deviation	.53	.53
senior	Mean	3.33	4.33
	N	3	3
	Std. Deviation	.58	.58
Total	Mean	3.23	3.67
	N	30	30
	Std. Deviation	1.04	.99

PDS Exercise 8: Crosstabs

To explore this issue further, let's look at a crosstabulation of academic rank and perceived control over one's life.

STEP 1. From the Data Editor, select *Analyze* (in Version 12.0, you need not be in the Data Editor). From the drop-down menu, select *Descriptive Statistics*. From the pop-out menu select *Crosstabs.* Look at the dialog box that appears. It's very similar to others we've looked at. It has the same basic format and you perform operations in it in much the same way as you have in other dialog boxes.

STEP 2. Select ACADRANK and move it to the *Columns* window inside the box. Select CONTROL and move it to the *Rows* window (usually the independent variable is put in the columns and the dependent variable is put in the rows). Press *OK*. Almost immediately, SPSS will organize the data into a display that will appear in a viewing window. The right side of the viewer has the information you're looking for. However, depending on the formatting parameters of your computer and monitor, you will probably have to scroll the window down or up and back and forth to see all of the display. You can select the print icon from the tool bar or from the File pull-down menu to print your display. It will look like what you see in Display 6. How does it look to you? Do you see a pattern that suggests that either upper or lower class students report having greater control over their life? Are the numbers spread out evenly in the table, or do they tend to "bunch" in certain areas? It's hard to tell with these data because our sample is so small. The numbers are scattered all over the box and it's hard to see much of a pattern.

Display 6

CROSSTABS OUTPUT: ACADRANK BY CONTROL*

(*Case Processing Summary is excluded.)

Acadrank*Control Crosstabulation

Count

		Academic Rank				Total
		Freshman	Sophomore	Junior	Senior	
Control	strongly disagree	0	1	0	0	1
	disagree	5	1	0	0	6
	undecided	5	5	0	2	12
	agree	0	2	4	1	7
	strongly agree	0	0	4	0	4
Total		10	9	8	3	30

To see if we have a pattern, one that suggests a relationship between the two variables, we need to do some data compressing. We can do that by *recoding* the values of our variables. We can collapse them into more general categories that will squeeze the frequencies together and allow us to see a pattern--if there is one. If you are not quite sure what we are getting at, follow the next exercise and see what happens.

PDS Exercise 9: Recoding Variables

What we are going to do is collapse the values for both academic rank and "Control over one's life." In effect, we are going to convert both variables into nominal data with only two values for each. (We refer to variables with only two values as "dichotomous" variables.) In the case of academic rank, we'll compress freshman and sophomores into one group and we'll compress juniors and seniors into a second group.

The variable CONTROL will also be compressed into only two values which will be, in effect, "low" and "high." To do that, we have to make a research decision. There are five values for CONTROL. The middle value is "undecided." Our judgment call here will be to lump "undecideds" with those who responded with "disagree" of "strongly disagree." We'll assume that people who are "undecided" don't have a positive sense of having control over their life circumstances.

In any case, here's what we will do in this exercise. Following the simplest, clearest route for recoding variables in SPSS, we're going to convert ACADRANK into a new variable named NRANK. It will have two values, namely:

1 = “freshmen and sophomores” and 2 = “juniors and seniors”.

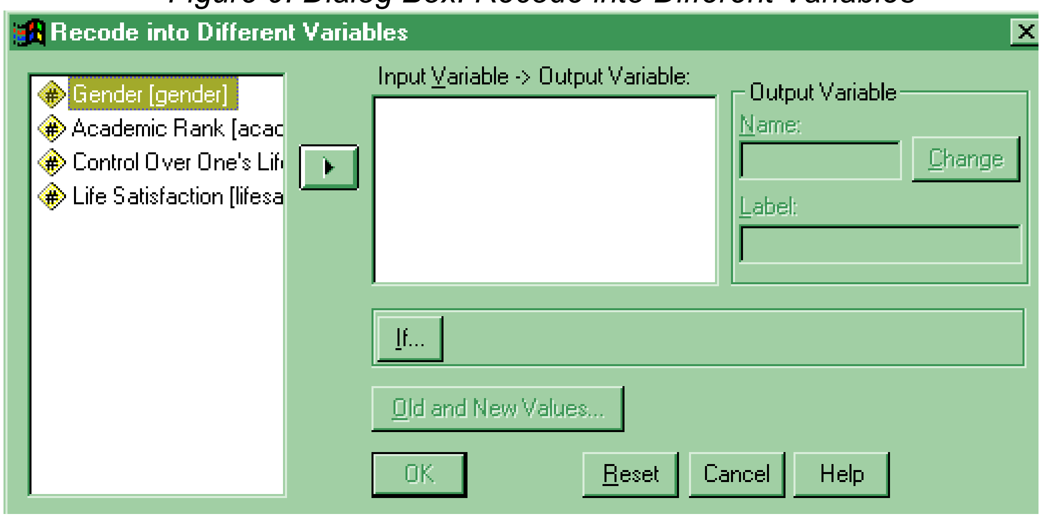
Next, we’ll convert CONTROL into a new variable called NCONTROL. It will have these labels: 1 = “low control,” and 2 = “high control.”

After we’ve done our recoding, we’ll look at a crosstabs between the two recoded variables--NRANK and NCONTROL to see if a pattern suggesting a relationship has appeared.

STEP 1. From the Data Editor, with the PDS file in place, select *Transform* from the options strip at the top of the screen (in Version 12.0 you must have the PDS data file in place, but need not be in the Data Editor). From the drop-down menu, select *Recode.* You’ll get a pop-out menu with two options. Select *Into Different Variable.* A dialog box will appear that is titled *Recode into Different Variables*. It looks like what you see in *Figure 9.*

STEP 2. We’ll start with ACADRANK. Looking at the Recode Into Different Variable box, notice your variable list is in the vertical window on the left. Select ACADRANK and move it over into the open window named *Input Variable ◊ Output Variable*. Use the arrow button between the windows and follow the same procedure you’ve used before.

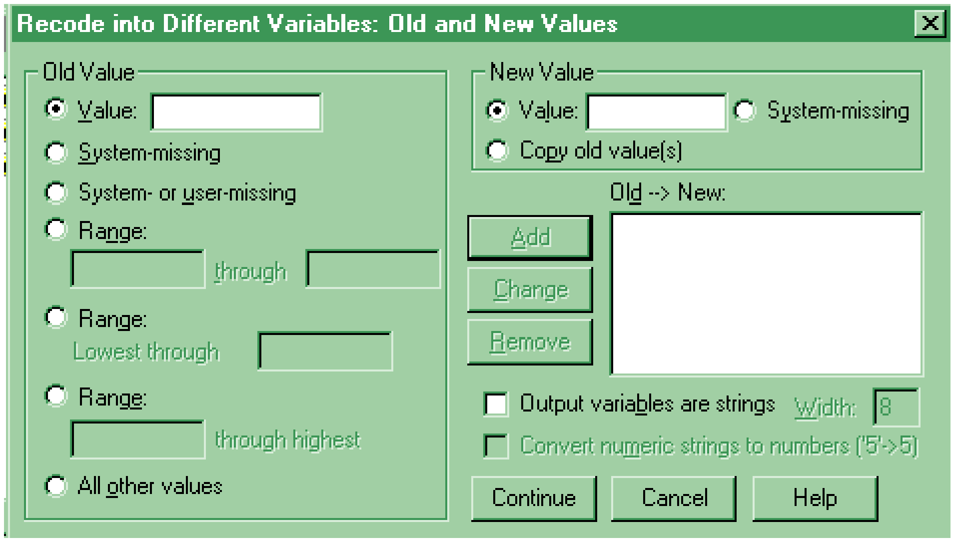
Figure 9. Dialog Box: Recode into Different Variables



STEP 3. To the right of the box, is a section marked *Output Variable*. Within that region of the box is a narrow, horizontal panel marked *Name*. Write the name of the new variable, NRANK, into the panel. Next, in the same section of the box, lower down, find the panel marked *Label*. Write: *Recoded Academic Rank* into that panel. If you are satisfied with what you’ve written, press the *CHANGE* button near the bottom of the box. ACADRANK ◊ NRANK will appear in the *Numeric Variable ◊ Output Variable* window.

STEP 4. You’re still in the *Recode Into Different Variable* box. Press the *Old* *and New Values* button. A box appears like that you see in *Figure 10*.

Figure 10. Dialog Box: Recode into Different Variables: Old and New Values



STEP 5. In the *Old Value* section to the left of the box, write the numeral "1" into the *Value* panel. Under the *New Value* section to the right of the box write the numeral 1 into that *Value* panel. The *ADD* button illuminates (darkens). Press it. You'll see "1 ◊ 1" in the large window called *Old ◊ New*:

What you've done, of course, is tell SPSS that the old value for Freshmen--which is 1--is now 1 in the new variable, NRANK.

STEP 6. Repeat Step 5 in this exercise for values 2 (sophomores) = 1; 3 (juniors) = 2, and 4 (seniors) = 2. When you finish your value recoding, the window in the *Old ◊ New* box should look like this:

1 ◊ 1
2 ◊ 1
3 ◊ 2
4 ◊ 2

You have successfully compressed the four values of the old variable into only two values for the new variable.

STEP 7. Press *Continue,* then *OK*. The dialog box goes away and, after a brief pause, the variable NRANK appears in your Data editor in a new, fifth column. However, you'll also note that the values are written into the data boxes with two decimal places. For example: 2.00 instead of 2. We'll attend to this matter shortly, in Step 9.

STEP 8. Repeat steps 1 through 7 for the variable CONTROL. You'll need to clear your input for NRANK by pressing *Reset*. Remember that values 1, 2, and 3 = 1, and values 4, 5 = 2.

STEP 9. This is a multi-part review step. Switch to the *Variable View* mode and go through the procedures you learned in PDS Exercise 3 to define your new variables, NRANK and NCONTROL in helpful ways. Be sure to label the values of your new variables by clicking on the shaded area of the values column in Variable View mode. These new value labels will show up on your crosstab tables.

PDS Exercise 10: Crosstabs Using Recoded Variables

STEP 1. Return to PDS Exercise 8: *Crosstabs*. Review it or consult it as needed for this exercise. Prepare a crosstabs output for NRANK by NCONTROL, using NRANK as your *Column* variable and NCONTROL as the *Row* variable.

STEP 2. Study the results in the viewing window and in *Display 7*.

Looking at *Display 7*, do you see a pattern? How would you define the pattern? How would you attempt to explain it?

Display 7
CROSSTABS: NCONTROL BY NRANK*
(*Case Processing Summary is Omitted.)

recoded academic rank * recoded control Crosstabulation

		Recoded Academic Rank		Total
		1	2	
Recoded Control	1	17	2	19
Over one's life	2	2	9	11
Total		19	11	30

Importing and Exploring Data Sets

The rest of this guide will depend on subsets of the GSS survey data as well as several other data sets provided for exercises and for exploration. To get at the data sets accessible through **SPSS, Student Version**, select *File* while you are in the Data Editor. Select *Open*. A pop-out menu offers you three options: *Data*, *Output*, and *Other*. (Alternatively, you can click on the file folder icon on the menu bar). In **SPSS, Full Version**, the procedure is the same, but there are additional options here that you need not worry about. For the exercises in the guide, you will want to access either *Data* or *Output*. The first option opens a window containing your PDS data set along with the selection of data sets offered in SPSS 11.0 Student Version. The *Output* option opens a window containing all of the output files you've saved.

Select *Open* and look over the interesting variety of data sets. Select any of them and, after a moment, the file will appear in your Data Editor. Take some time to explore different data sets to see what you find. Let your mouse arrow rest on the different variables to see what they are. To get a better idea about what each variable is, select *Utilities* from the options strip at the top of the Data Editor. From the drop-down menu select *Variables.* A dialog box called *Variables* appears. The window to the left of the box has the variable list. The larger window to the right of the box has a display that gives you information about each variable you select from the variable list. (F1 in the display stands for the data field, namely a numeric display one character wide. NAP stands for "Not at Phone;" DK stands for "Don't Know," and NA is for "Not Answered." Explore the variable lists for different data sets.

The most comprehensive and complex data sets you will find, like GSS93 subset, are from the General Social Survey. The GSS has been ongoing since 1972. Year by year, careful multi-stage sampling has produced a representative sample of 1,500 English-speaking people over the age of 17 within the continental United States. That means more than 30,000 people have yielded information to carefully trained interviewers on a wide range of information.

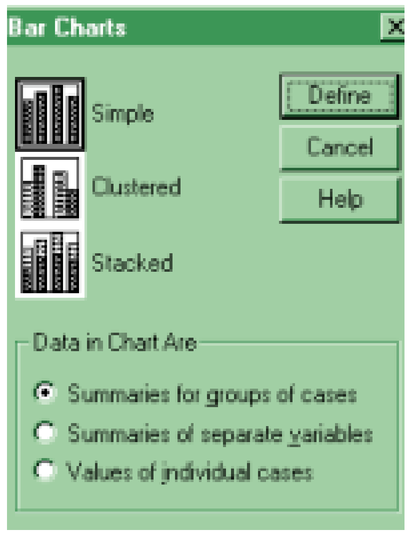
Open GSS93 subset and explore the different variables. Look at them in the data mode and in the variable view mode. Check out the variables list through the *Utilities* option. You'll notice that the variables fall into rough categories. Some, for example, like *age* and *educ* are scale variables that give us background or demographic measures of the sample population. Other variables, usually ordinal in nature, give us measures of social attitudes. Examples include *news*, *sexeduc*, and *cappun*. Check out those variables and see what they appear to measure. As you look at the different variables, try to decide which ones might be taken as dependent variables and which ones might serve best as independent variables. Imagine some relationships among variables that you might like to explore. Open the data set related to home sales. Ask yourself what the *unit of analysis* is here. Again, think about which variables might precede and tend to predict other variables.

To close part 1, let's see how to prepare a simple bar graph. Then you can browse around in data sets and prepare bar graphs (or pie graphs). Feel free, too, to look at various descriptives for any variables that interest you.

Data Set Exercise 1: Creating a Simple Bar Graph

STEP 1. While in the Data Editor, open GSS93 subset. Select *Graphs* from the tool bar. From the drop-down menu, select *Bar*. A dialog box opens called *Bar Charts*. It looks like *Figure 11*.

Figure 11. Dialog Box: Bar Charts

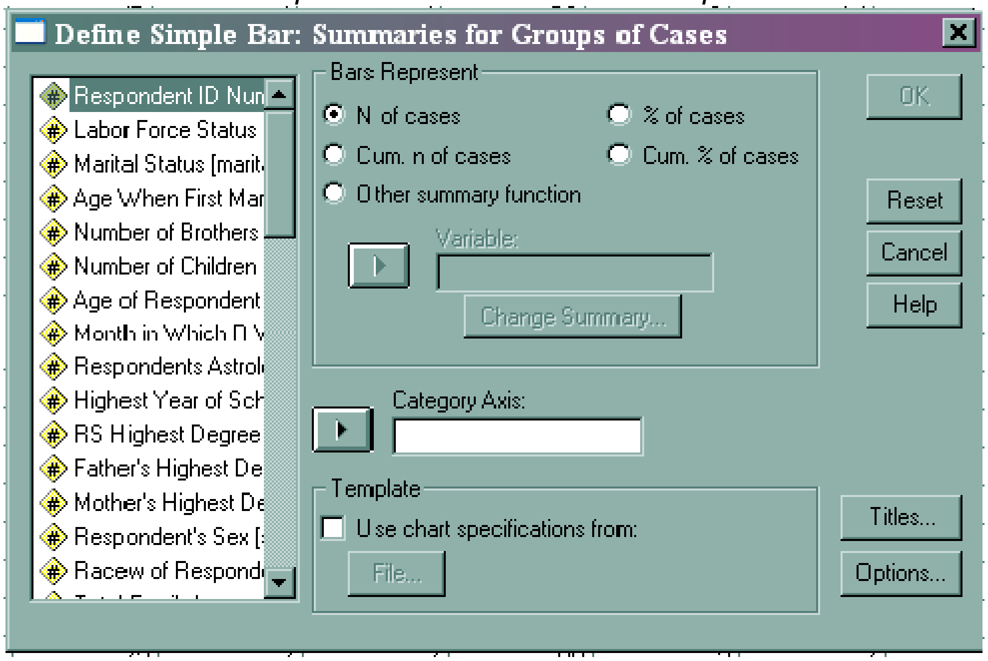


STEP 2. In the *Bar Charts* dialog box, press the *Simple* button. Put a dot in the porthole next to *Summaries for groups of cases.* Finally, press the *Define* button and observe the dialog box you see in *Figure 12*, shown below.

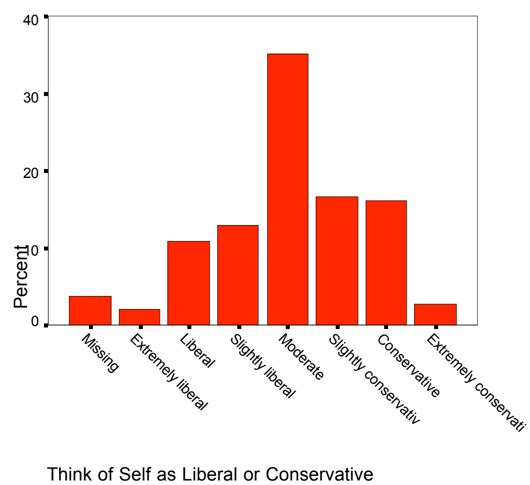
STEP 3. In the *Define Simple Bar* dialog box, mark the porthole next to *% of cases*. Next, select the variable polviews from the panel of variables on the left of the box and move it to the panel window marked *Category Axis*, using the arrow button. *Press OK.* Observe the viewing window. It should look like what you see in *Display 8*.

To extend this exercise, select *Analyze* from the tool bar at the top of the Data Editor. Choose *Descriptive Statistics*, then *Frequencies*. Pull the variable, polyview into the main panel. This time, while in the Frequencies dialog box, press, *Statistics*. Select *mean*, *median*, and *mode* by checking those options in the dialog box that opens when you do this. Press *Continue* in the Statistics box to return to the Frequencies dialog box. Now, press the *Charts* button and consider the dialog box you get this time. Notice that you can simply choose a bar chart. However, this time, mark the porthole opposite *N of Cases*. Press *OK* and observe what shows up in the viewing window.

Figure 12
Define Simple Bar: Summaries for Groups of Cases.



Display 8. Simple Bar Graph:
Summary of Political Views from Extremely Liberal to Extremely Conservative



Compare the bar charts you get for looking at the number of cases per variable attribute as opposed to looking at the percentage of cases. Notice the differences and the similarities. You'll see that they are very similar because the distribution of political views around the value "moderate" is rather close to a rough normal curve.

Part 2

CHAPTER KEYED SPSS EXERCISES
INTRODUCTION

The exercises included in Part 2 are designed to supplement important topics in your text, chapter by chapter. All of them will utilize the data sets that accompany SPSS (all versions) and this manual.

You were introduced to GSS93 subset in Part 1. Take some time now to explore the other data sets available to you. While in the Data Editor, select *File* ◊ *Open* ◊ *Data*, and observe the data sets. Select GSS93 subset*, World95, Employee Data,* and any other data sets that arouse your curiosity. You'll see that data sets like GSS93 subset include a wide variety of demographic or background variables, many of which can be taken as independent variables, along with a fair number of social attitude measures, many of which can be looked at as dependent variables. While in GSS93 subset, make some notes to yourself about relationships it might be interesting to explore. For example, among social attitudes measured in GSS93 subset is one's reported views on capital punishment (cappun). One might wonder, among other things, whether education, age, or gender might be a stronger predictor of one's position on capital punishment. Or, perhaps, we might wonder how these variables might interact with respect to *cappun.*

Your background in statistics may not be advanced. In fact, you may not have taken a basic statistics course when you take this course in research methods. For that reason, exercises involving statistical procedures may be thought of as highly recommended options.

PART 1: AN INTRODUCTION TO INQUIRY

Chapter 1 ***Human Inquiry and Science***

Exploring Variable Language

Open GSS93 subset. Explore the variables *sex, degree* (highest degree received) and *relig* (religious preference) by preparing a frequency output for each variable. If you're unsure how to do this, go back to exercise 5 in this manual. You should also prepare bar charts for each variable. You can do this by first clicking on the "charts" box at the bottom of the frequencies window, and then clicking on the "bar charts" option. The bar charts will give you a visual sense of the attributes for each of the three variables. Meanwhile, notice that in SPSS language we are inclined to speak of attributes as *values* assigned to a variable. Meanwhile, think about how the ratio of *females to males* might affect data summaries. Consider the numerical preponderance of people in the samples who have either a high school degree or less than a high school degree with respect to social attitudes respecting abortion or electing a black president. Ask yourself if the high percentage of Protestants in the sample will have much effect on general social attitudes. Finally, go back to the "charts" window and change the chart values from "frequencies" to "percentages" and note how this changes the bar charts that the computer generates. Does this make the bar charts easier to interpret?

Chapter 2 *Paradigms, Theory, and Social Research*

One of the most important things for you to learn is that computers can't do the thinking for you. This means that you can't develop theories from data analysis. It's the other way around. Theories are constructed by thinking about the issues and reading what others have written on the subject. That's the hard work. Once you've developed a theory, you can test it with software programs like SPSS. As you'll see, that's the easy part.

Chapter 3 *The Ethics and Politics of Social Research*

Exploring your GSS93 subset data set, briefly explain your understanding of whether the research data was (1) gathered through voluntarily participation (2) without doing harm to the subjects, (3) while protecting participants' anonymity, and (4) not deceiving them.

PART 2: THE STRUCTURING OF INQUIRY

Chapter 4 *Research Design*

Criteria for Nomothetic Causality:

1. Open the GSS93 subset. Run a Crosstabs procedure between the variables *spanking* and *degree*. The first variable regards approval of using spanking to discipline a child and the second is the respondent's level of education, in terms of highest degree earned. Put spanking into the rows window and degree into the columns window. Select *Statistics* and then, in the statistics dialog box, select *gamma* and *Chi Square*. Press *Continue.* Press *Cells* and select percentages by *Columns*, then press *OK*. You need not have much knowledge of the statistical tests you've applied to grasp what you'll find in your viewing window. Study the percentages and the frequencies in the cells to see if you can detect any pattern in the data. Note that the Chi-square and gamma tests find the relationship between these two variables significant at a .004 and .000 level, respectively. Those numbers tell us that the differences in percentages we see across the rows are very unlikely to have appeared due to

a random distribution of responses.

2. Be sure to save your output into an SPSS file with some name that helps you
identify it for review or for later use.

3. Explain how the criteria of causality may be met in this exploration. (1) Which of the two variables can be said to precede the other in time? (2) If neither can be said to meet that criterion, what can be said about either variable being a "cause" of the other? (3) Are the two variables empirically related? (4) Is it possible or even likely that the relationship between *spanking* and *degree* is caused by—or more accurately--predicted by some antecedent variable? Some conceivable antecedent variables might turn out to include measures of socio-economic class. Can you suggest why this might be the case?

Units of Analysis:

Open the data set World 95 and explore it briefly. Do the same for data sets called Coronary Artery and Home Sales. What units of analysis are used in each of these data sets?

The Time Dimension:

The time dimension is the anvil on which we hammer out our understandings of *changes* in data. Public attitudes, incomes, and demographic measures, like the proportions of people in various age categories do not remain constant. So, often, we are interested in patterns or trends in our data. If we have the funding and the resources, we may want to pursue a *longitudinal study*. That is, we may want to take related sets of observations over specified periods of time. Your SPSS 11.0 data sets don't include specific longitudinal measures—even though they are definitely included in actual GSS data sets. Still, you can get some sense of what's involved in a *trend study* by manipulating the *age* variable. Try this exercise:

1. Open GSS93 subset. Recode the *age* variable such that ages 18-30 = 1; 31-44 = 2; 45-65 = 3, and 66-89 = 4. Call your new variable *age2*.

2. Proceed to the variable view in your Data Editor to define your new variable, *age2*. Make it decimal free with a width of 1. Write in the value labels for use in output displays and record the variable as ordinal.

3. With the GSS93 subset data set in place in your Data Editor, run a crosstabs procedure for polviews (political views) and *age 2*. In the Crosstabs dialog box, place your four age groups variable (*age2*) in the columns panel, and *polviews* in the rows panel. Press *Statistics* and select *Chi-square* ◊ *Continue*.

Back in the original dialog box, press *Cells*. In that dialog box, mark percentages by columns and press *Continue*. Press *OK* and consider the output you get in your viewing window.

> ***SPSS Procedures***:
>
> Follow the recode procedures you were introduced to in Part 1. However, when it comes to assigning *Old and New Values*, mark the *Range* porthole on the left side of the box and fill in a range, such as 18 through 30 across the two panels. Having written in the *Value* as 1 in the upper left panel, press *Add* beside the *Old◊ New* panel on the right side of the box. You'll get 1 = "18-30." Proceed to add your other three values in the same way, press *Continue*. Press *OK*.
>
> Make sure you save your Data Editor display so that the revision to your data set is retained for future use.

You will find that the Chi-square test reveals a definite, statistically significant difference in the attitude of the four age groups toward their political views. Think about why these different age groups vary in their political views. Is it because age has an effect on how a person votes, or because their views concerning politics are partially determined by the time period in which they grew up?

Chapter 5 *Conceptualization, Operationalization, and Measurement*

Operationalizing Concepts

1. Operationalization is the process of finding variables that reflect the underlying concept that you're interested in studying. Let's say you want to study the relationship between the concepts of education and attitudes towards corporal punishment. The researcher's question is how to measure (or operationalize) these two concepts. The variable DEGREE (highest degree obtained) is one operationalization of education. How well does this operationalization reflect the underlying concept of education? What does it miss? Likewise, SPANKING is one measure of attitudes towards corporal punishment. How good a measure is it? Now think about another concept – attitudes towards legalization of drugs. Can you operationalize this concept with a variable in the GSS data set?

2. Run a Frequencies procedure for the GSS93 subset variable *agewed*. That's the year one was first married. Print the contents of your output window. Examining the hardcopy, figure out how to reduce the number of variable attributes (reported ages) that are spread out from age 13 to age 58. Consider

a recode with 3 variable attributes. Explain how you would justify your breaks. There's no need to do a recode, simply work out how you *would* recode *agewed* if you wanted to. In recoding a variable to make it easier to work with, will you be drawing away from the conceptualization of your variable?

Exploring Reliability and Validity

Reliability is getting the same kinds of results when a particular measure is applied repeatedly. For example, a set of survey items measuring attitudes that are intended to reveal racial prejudice might yield similar results in similar sample populations over time. The measures would exhibit reliability. On the other hand, reliability does not mean we are necessarily measuring what we think we are measuring; a reliable measuring technique may not be valid. I might reliably run my right rear tire up over the curb every time I try to parallel park. That doesn't mean my parallel parking method is valid.

While we may agree on a valid method for parallel parking, we will generally have a harder time agreeing on a whether or not an operationalized concept is measuring what we think it should be measuring. That's partly because our concepts are always constructs based on reaching agreements about things—like what constitutes prejudice, or what income level should be considered "middle class."

1. To explore dimensions of validity, open your PDS data set. Think about *lifesat* and *control*. In terms of *face validity* do these measures seem to make sense? Does simply asking people to evaluate how satisfied they are with their lives, for example, give us a valid measure of that concept?

2. Open GSS93 subset. Consider what may be involved in the concept of conservatism. Choose three variables that you think are good indicators of this concept. Next, run a crosstabs procedure for each of these three variables with *polviews.* Select the chi-square and gamma statistics to get a rough sense of how strongly related each of the three variables appears to be to a self-assessment of one's self as either liberal or conservative.

3. Attempt to understand how the exercise you've just gone through relates to *criterion-related validity* on the one hand or *construct validity* on the other. Review your text treatment as seems helpful, recalling that criterion-related validity is also called *predictive* validity and that construct validity has to do with the *logical relationships* among variables.

4. Reconsider the PDS variable, *lifesat*. Think about the notion of *content validity*. Then, formulate an argument that the variable, *lifesat*, is too limited; it fails to cover the range of possible meanings in the concept we're calling "life satisfaction." Explain why you've taken this position.

Chapter 6 ***Indexes, Scales, and Typologies***

To supplement your understanding of the discussion of index construction in your text, create a variable called *conserv* using the C*ompute* procedure. In effect, you will be using the summation of several attitude variables to create a simple index.

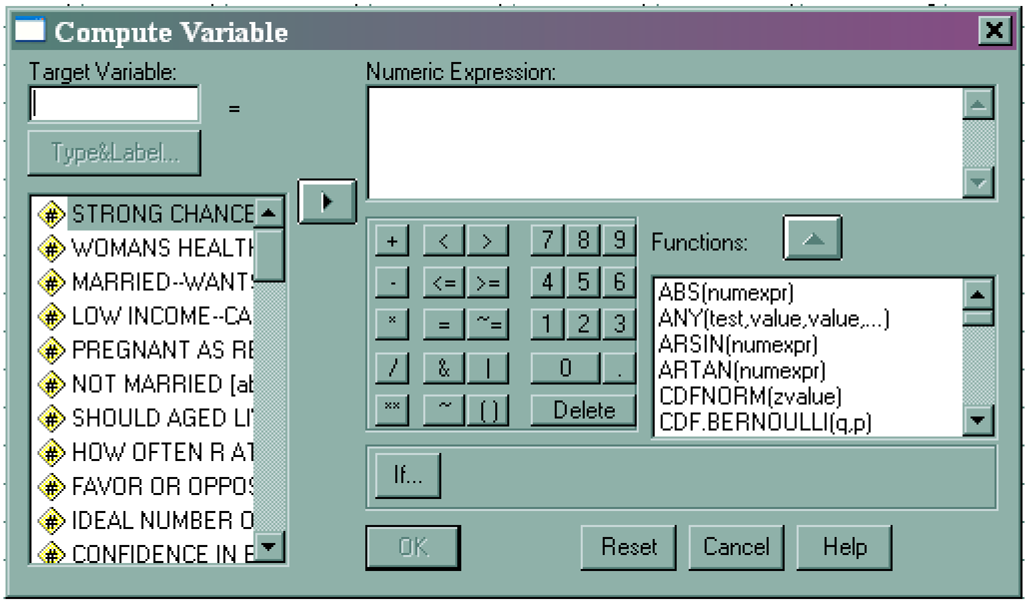
1. For this exercise, you will need to open a data set named GSS7500s.sav included on the GSS data disk that may have come bundled with your SPSS CD-ROM. If you're unsure how to open this file, just ask your instructor. Search through the variables in this data set for attitude measures that, in your view, would suggest a conservative political attitude. See if you agree that approval of a law prohibiting interracial marriage (*racmar*), the favoring of capital punishment (*cappun*), attendance at religious services (attend), and *rejection* of a woman's right to have a legal abortion in the case of rape (*abrape*) would seem to make a cluster of variables suggesting a conservative political attitude.

2. *Recodes*. To make our analysis and make it easier to handle, we'll convert all of our potential index items into nominal variables. However, when you examine the attributes of these variables, you'll find that they are "going in different directions" when it comes to suggesting conservatism. For example*, racmar* and *cappun* code the favorable response as "1" and the opposition response as "2." Meanwhile, favoring the right of a woman to an abortion in the case of rape (*abrape*) makes 1 the "agree" response. Also, *attend* turns out to be an ordinal variable with four response options from low to high attendance.

To develop an index by summing four variables which seem, *on their face*, to represent conservatism, recode all the variables as follows: Recode *cappun* to *cappun2* and *racmar* to *racmar2* by reversing the values such that 1 = 2 and 2 =1.Recode *attend* to *attend2* such that the old value 1 and 2, both equal 2, while the old values 3 and 4 both equal 1. Leave *abrape* alone since the conservative response is already "2" for opposition. Define your new variables in the Data Editor and make sure to save the new version of GSS7500s.sav so you won't lose all your new variables.

3. Sum the values of the four variables to create an index value we will call *conserv*. Select *Transform ◊ Compute*. Observe the dialog box called *Compute Variable* on your screen or in *Figure 2-1*.

Figure 2-1: *Compute Dialog Box*



SPSS Compute Procedure: In the Compute Variable dialog box find the *Target Variable* panel and write "*conserv*." In the window called *Numeric Expression* write: *racmar2 + cappun2 + abrape + attend* Press *OK*. Scan the Data Editor to locate your new variable.

4. Use the variable view mode to define your new variable as seems useful. The level of measurement now will be scale, even though all of the composite variables have been reduced to a nominal level. Summing the variables as we've done here creates definite mathematical intervals along the range from the lowest possible score, which is 4, to the highest possible score, which is 8.

5. Produce a frequencies output accompanied by a simple bar chart for the values of *conserv*. Explore the output window on your screen or print it for easier analysis

Testing Your Index

6. You've already established that you index items show face validity. Now, consider the pages in your text that deal with index validation. You'll see that you can produce an *item analysis* of your index items by running a crosstabs procedure, comparing *conserv* with *racmar2*, *cappun2*, *abrape*, and *attend2.*

Notice that you'll want to place *conserv* in the columns position, with index items in the rows position. Be sure to specify percentages by columns and see what you discover in your viewing window.

7. To check on the external validity of your conservatism index, run a crosstabs of *conserv* by *polviews*.

8. A good index should have an ordinal range among items. For example, in our rough conservatism index, *racmar2* may seem a more extreme measure of conservatism than *cappun2*. Does your data analysis show this to be the case? Explore your data outputs, especially comparing cell frequencies in the 2 x 2 contingency tables, to see if a rough ordinal range is present among the items.

Chapter 7 *The Logic of Sampling*

1. Open GSS93 subset and scroll down the Data Editor to get a sense of the number of respondents in the data set, paying special attention to the variable *educ*, which is the respondent's level of education. Let's explore the nature and consistency of probability sampling by exploring this variable in the GSS93 subset sample.

For the variable *educ*. Select *Analyze* from the tool bar and proceed to use the frequencies procedure, selecting *Mean* from the *Statistics* dialog box. Save the contents of your output window for future reference.

> ***SPSS Procedure*:**
>
> When you take a random sample from your total sample of 1500, you are treating N =1500 as a study population. To draw each random sample of n = 150, select: *Data ◊ Select Cases.* In the *Select Cases* dialog box, mark *Random sample of cases*. Press the *Sample* button. A dialog box appears called *Select Cases: Random Sample*. Mark *Approximately* and write the numeral 10 in the small open window which is followed by the expression *% of all cases*. Press *Continue* to return to the previous dialogue box. Mark the porthole for "filtered." Press *OK*. Scan your Data Editor to see that the random sample has been selected. You can now use the Frequencies procedure to get the mean of *educ* for this sample.
>
> To get rid of a filtered Data Editor display, press *Data ◊ Select Cases*. In the *Select Cases* dialogue box, mark *All Cases*. Press *OK*. Notice that no cases are now filtered in the Data Editor, and you can proceed.

2. Take five separate 10% random samples from your GSS93 subset population by following the above procedure. Each random sample, of course, will include about 150 cases. Establish the mean response for each of your

five samples by using the frequencies procedure and selecting *Mean* from the *Statistics* option. Save each output for future reference or simply write down the valid n, the number missing, and the mean for each sample as you proceed.

3. Compare the means for each of these five 10% samples with the mean for the complete sample. On a piece of paper, create a simple summary table, arranging the sample means in order from the greatest minus difference to the greatest plus difference above and below the mean of the complete sample.

4. Select: *Analyze* ◊ *Descriptive Statistics* ◊ *Frequencies*. Press the *Charts* button and select *Histograms*. Also, mark the porthole for superimposing the normal curve on your output display. Press *Continue.* Select *educ* while in the *Frequencies* dialog box., then press*OK*.

Study your output. Print it for convenience. Notice that the spread of your sample means very roughly follows the distribution for years of education in the histogram. As it turns out, if you run enough of these 10% sample means they will tend to spread around 13.04 in a way that follows the normal curve. Study this phenomenon in relationship to your text discussion of sampling distributions and the calculation of standard errors.

PART 3: MODES OF OBSERVATION

Chapter 8 *Experiments*

1. In this exercise you will evaluate an experiment on the effectiveness of fertilizing plants. In order to do these analyses you'll need to enter the following data into SPSS, just like you did for the PDS file. Here are the data:

	Treatment	Starting Height	Ending Height
Plant 1	1	4	53
Plant 2	1	5	54
Plant 3	2	2	63
Plant 4	2	4	67
Plant 5	3	3	76
Plant 6	3	7	78

Enter these data into the SPSS data editor. You should have 6 cases and three variables. Name the variables TREAT, START and END, respectively.
Here are the details about this experiment. Namely, 6 corn plants were grown over a time, *t*. Two of them, in treatment 1, got no fertilizer. Two plants, in treatment 2, got a little bit of fertilizer. Two more plants, those in treatment 3, got

all the fertilizer they might have hoped for. An initial observation for each plant was its height, presumably not long after germination. Let's say the height was taken in centimeters. After a period of time in which all 6 plants were treated the same way, got the same light, were grown in the same kind of soil, and so on, a final observation—height of plant in centimeters—was taken. Simply by looking at the raw data, you can see that there are differences among conditions 1, 2, and 3.

2. Let's figure the differences. Select *Transform* ◊ *Compute* from the Data Editor tool bar. In the *Compute Variable* dialog box, write "growth" in the *Target Variable* panel. From the variables you see in the panel to the left of the box, select *end* and press the arrow to move it into the *Numeric Expressions* panel. Next, press "–"(minus) on the keypad. Then select the variable *start* and move that into the *Numeric Expressions* panel. That window display should now read:

end – start

Press, *OK* and observe your new variable, *growth*, in the Data Editor. In the variable view mode, define it with 0 decimal places and a width of 2.

3. Now let's get a better sense of the differences between treatments1, 2, and 3. From the Data Editor tool bar, select *Analyze* ◊ *Compare Means* ◊ *Means*. In the *Means* dialog box, move *treat* into the independent list panel and move *growth* to the dependent list panel. Press *OK*. Clearly, as you will see in the viewing window, there are differences among the means. Save you output for future reference.

4. To get a visual sense of the differences between means in treatments 1, 2, and 3, select *Graphs* ◊ *Bar*. In the *Bar Charts* dialog box, press *Simple*. Leave the porthole for *Summaries for groups of cases* marked. Press *Define*. In the dialog box move treatment (*treat*) to the *Category Axis*. Under *Bars Represent*, mark *Other Summary functions*. That opens the Variable panel. Select *growth* and move it into that panel using the arrow button. Notice the way the variable is now expressed in that panel. Press *OK* and study what you get in the viewing window. The differences among means in the three treatments are now visually evident.

5. What hypothesis might you have written down to specify the objective of your data analysis? (Write a null hypothesis if you haven't already.) Have we met the terms of a classical experimental design, assuming our assumptions about this data are valid? How might an experiment of this sort be made stronger?

Chapter 9 *Survey Research*

1. Pursuing *secondary analysis,* open GSS93 subset. Run a crosstabs between age2, the four-age-groups variable you created earlier, and *life*. The *life* variable asks respondents if their life is exciting or dull. While in the Crosstabs dialog box, select *Statistics* and check gamma. Gamma is a handy measure of association between ordinal variables. (You should recall that *age2* became an ordinal variable when it was recoded from *age*.)

2. Study the contingency table in you output viewing window. Notice, simply from studying the cell frequencies, that there is a pattern. That pattern is also suggested by a modest, but mildly significant gamma measure. In short, not too surprisingly, we find a relationship between getting older and seeing one's life as less exciting.

Chapter 10 *Qualitative Field Research*

SPSS isn't really designed to handle qualitative research. There are, however, several software packages available for this purpose. Ask your teacher about them.

Chapter 11 *Unobtrusive Research*

Research in the social sciences has the potential drawback of altering the phenomenon being studied. When an investigator interviews subjects or observes social interactions, their presence may affect the processes under examination. People tend to behave differently when they are being studied. *Unobtrusive research* attempts to avoid this problem by collecting or using data in a way that doesn't alter the relationships being studied. Think about the topic of suicide. One way to study suicide would be to interview people who have attempted suicide. The problem with this is that their responses may be influenced by what the person thinks s/he should say, or by the characteristics of the interviewer (sex, race, age, etc.) An alternative method would be to use data that don't affect the results. For example, Emile Durkheim's groundbreaking 1897 study, *Suicide*, used public records to explore the relationships between demographic data and personal acts of self-destruction in a way that could have no imaginable effect on either the suicides or the demographic data associated with suicide rates. Other examples of unobtrusive measures include content analysis and historical/comparative analysis.

Chapter 12 *Evaluation Research*

Since evaluation research is concerned with examining the outcome of public policies or programs of social intervention, it is not really feasible to use your SPSS data sets to that end in a clear and unambiguous way. You might want to write out a brief explanation to yourself as to why that would be the case.

On the other hand, it is possible to do a little "back engineering" to see the line of reasoning behind evaluation research by looking at relationships among variables in data sets like GSS93 subset, GSS7500s, or even World95.

1. Open World 95. Explore the data set until you locate *lit_fema* and *babymort*. The first variable is the percentage of women who can read, the second is the infant mortality rate operationalized as the number of infant deaths per 1000 population per year. Select *Analyze* ◊ *Correlation* ◊ *bivariate*. The use of a Pearson correlation measure (coefficient) is quite appropriate here, since we are dealing with two scale (ratio level) variables. While in the correlation dialog box, select *Options* and note the default value is marked for "Exclude missing cases pairwise." Under *Statistics*, check the box next to *Means and Standard Deviations*. Press *Continue*. Press *OK* and examine your output.

2. Notice in the Descriptives summary box that the mean for *lit_fema* is 67.26% with a very broad spread among countries; the standard deviation is 28.61%. Notice too, that we only have data on *lit_fema* for 85 out of the 109 cases (countries) in the sample.

3. Explore the correlations matrix box and consider the meaning of a negative .843 correlation coefficient. As you'll see, it's considered significant at the 0.01 level. As correlation coefficients go, that's a strong one. (For those who have had some background in statistics, you'll understand that r =.843 accounts for $.843^{2} =$ 71% of the variance between the correlated variables.)

4. To visualize what you have here, while still in the output viewing window, select *Graphs* ◊ *Scatter*. In the Scatterplot dialog box, press Simple, then press Define. In the Simple Scatterplot dialog box that opens for you, move *lit_fema* to the *Y Axis* panel and *babymort* to the *X Axis* panel, using the arrow buttons you are now familiar with. Should you wish to check your *Options* while in this dialog box, you'll note that "Exclude cases listwise" is checked—which is what you want. Press *Continue*, then press *OK* and be amazed at your output. You will see how obvious it is that female literacy and lower infant mortality rates are strongly related.

5. Now consider the whole point of evaluation research and think "backwards." If you had the political clout to mount progressive social programs to reduce infant mortality in your country, would you not consider programs that increased female literacy? And, if you did successfully establish such programs, might not your *evaluation research* include analysis of the sort you've just explored?

PART 4: ANALYSIS OF DATA

Chapter 13 *Qualitative Data Analysis*

As we pointed out earlier, SPSS isn't really suitable for analyzing qualitative data. Ask your teacher about the various software packages available for these types of studies.

Chapter 14 *Quantitative Data Analysis*

You've had a mini-introduction to data coding in part 1. To supplement the much more elaborate treatment of coding in your text, try the following:

1. Scan the Data Editor with your GSS93 subset data set in place to see if you can detect any cases of possible code errors. You may not find any, but you will discover variables that were not responded to and missing value codes.

2. Explore the variables *polviews*, *pillok*, *letdie1*, and *tvhours*, looking at frequencies for each of them. Try writing out the appearance of your codebook for these items--assuming you were responsible for constructing a codebook. Be sure to include your idea of the best wording for the survey question for each variable.

Univariate Analysis

1. Open GSS93 subset. Explore the variable *news*. It summarizes how often respondents report reading a newspaper. Switch between the data and the variable view modes to check out the variable. Also, pull down the *Variables* list by accessing the *Utilities* option in the Data Editor tool bar.

2. Examine the variable *news* using the frequencies procedure. Press the *Statistics* button while you are in the *Frequencies* dialog box. Select *mean, median*, and *mode*. Select the bar chart option after pressing *Charts*. Press *Continue*, then *OK*.

3. Explore (or print the contents of) your output window to see how you can collapse extreme categories. Specifically, after examining the valid percent column, your bar chart and your measures of central tendency, figure out how you might merge the 5 attributes of news into only 2.

4. Recode *news* into a new variable, *news2.* To make life easier for yourself, recode *news* attributes 1 and 2 into a value "2" for *news2*. Recode the old values 3, 4, and 5 into the value "1." By recoding in this way, of course, frequent newspaper readers have a higher value. Why do we want to set up the vales in this way? Well, let's assume that we are going to propose a working

hypothesis: Namely higher education tends to be associated with more frequent newspaper reading. In fact, that is just the sort of hypothesis we will want to consider.

5. Run a frequency on *news2*, using the same procedure you employed in step 2, above. You'll notice that frequent newspaper readers still tend to be in a fairly large majority.

6. Recode the GSS93 subset variable *degree* into a new variable called *hiloeduc*. Let 0, 1, and 2, in the variable *degree* equal 1 in the new variable. Let the values 3 and 4 in *degree* equal 2. To assure yourself that this recode makes sense, Run a frequencies procedure on the variable *degree* and look at the value labels, frequencies, and percentages for the values 0 through 4.

7. While in variable view mode, define *hiloeduc* as a numeric variable with a width of 1 space and 0 decimal places. Give it the variable label "Level of education." Assign "1" the value label "high school or less" and assign "2" the value label "more than high school." Designate the variable as *nominal*.

Bivariate Analysis

8. Run a crosstabs procedure on *hiloeduc* by *news2*. While in the Crosstabs dialog box, place *hiloeduc* in the columns panel and *news2* in the rows panel. In the *Cells* dialog box, select, *Percentage by columns*. Before leaving the Crosstabs dialog box, leave a check mark beside *Display Clustered Bar charts*. Optionally, select the *Statistics* option for lambda. You'll see that these are statistical measures of association associated with nominal by nominal comparisons. Press *OK*.

9. Study your output window. See what you can make of the lambda results by looking carefully at the actual frequencies you see in the contingency table created by your output. Reviewing your text discussion as you need to, explain the conclusions you may draw from the table.

Chapter 15 *The Elaboration Model*

1. Open GSS93 subset and run a crosstabs procedure on *cappun* by *hiloeduc*. Move *hiloeduc* to the columns panel and *cappun* to the rows panel. Check the option for *Display Clustered Bar Charts*. Press *Continue*, then *OK* and examine your output in the viewing window.

2. *While still in the viewing window*, select: *Analyze* ◊ *Descriptive Statistics* ◊ *Crosstabs*. *hiloeduc* and *cappun* should still be in the dialog box unless you have reset it. From the variable list on the left of the dialog box, select *sex* and move it into the bottom window. Once again, select *Display Clustered Bar*

Charts. Press *Continue*, then *OK*.
3. Looking at your output, you will see that your partial tables (for the male and female parts—partials--of your sample) are stacked. Look at the cell frequencies and at your clustered bar charts to explore the relationships of level of education to attitudes about the death penalty. In terms of the elaboration model, what has occurred when we applied the *test variable sex*? Have we *specified* or *replicated* the relationship between level of education and the attitude towards the death penalty?

4. What test variable (other than sex) can you imagine that might conceivably produce either an *explanation* or an *interpretation* of the relationship between *hiloeduc* and *cappun*?

Chapter 16 *Social Statistics*

Descriptive Statistics.

1. Open GSS93 subset. Look at the frequencies for *polviews*. Then, run a crosstabs of *polviews* by *sex*. Save your output in a file for further use. Now, create *your own* table for these variables, using the kind of format used in your text. Do a thorough and neat job; pretend it will be placed in a research report.

Nominal Measures of Association

2. Examine a crosstabs between *sex* and *grass*, selecting the statistics Phi and Cramer's V and the Contingency coefficient. You will recall that *grass* asks whether or not marijuana should be legalized. Noting in your output viewing window that all three selected statistics are significant at about the .001 level, study the distributions of frequencies in the contingency table to see why. If you failed to ask for percentages by rows or columns, percentage the table to get a better sense of the pattern. Discuss the nature of these nominal measures with your instructor or consult a basic statistics text. You'll find that the measures are easy to calculate by hand in 2 x 2 tables and that the Contingency coefficient is the equivalent of a correlation for nominal variables. Save your output for future reference.

Ordinal Measures of Association

3. Prepare a crosstabulation between *polviews* and *cappun*. Before doing that, check out the two variables to satisfy yourself that they are both at the ordinal level of measurement. Put *polviews* in the columns panel and *cappun* in the rows panel. Check the clustered bar chart option and, from the *Statistics* dialog box, select the ordinal measure of association *gamma*. Press *Continue* and *OK*. Study your output. You will note that the gamma measure is significant at a .000 level. That means that there is less than 1 in 1000 chances that these

variables are unrelated in the general population.

Chi-Square

4. Throughout this guide, we've asked you to select and appraise the results of Chi- Square tests. As an exercise, after first reviewing your text discussion of Chi Square on pages 464-468, look over the Chi-Square outputs you will have accumulated in your SPSS output files. Attempt to make sense of how useful they were.

Interval/Ratio Measures of Association

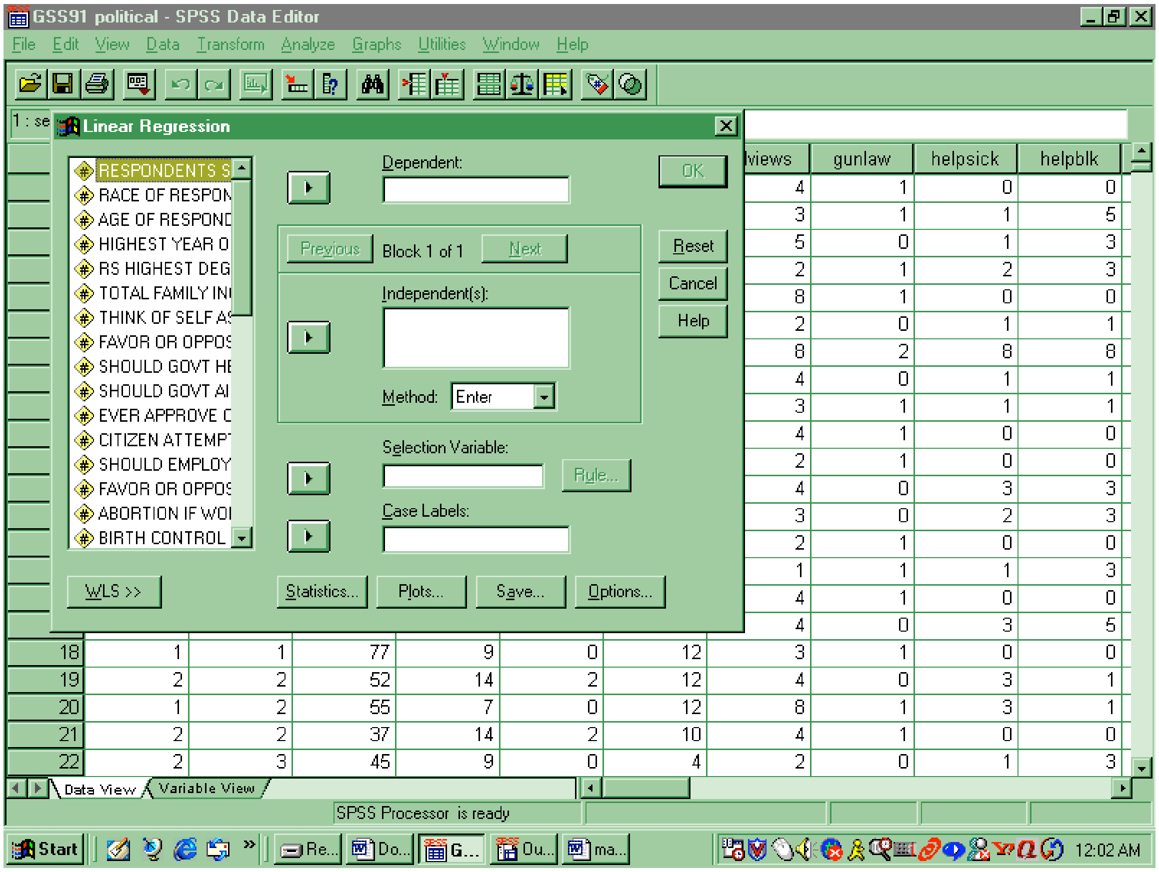
We've explored the way you can use SPSS to explore simple bivariate correlations in our World 95 exercise for chapter 12. You also got some sense of what linear regression is about by simply thinking about the approximate regression line in the scatterplot we derived for female literacy by infant mortality.

6. Open the GSS93 subset. To extend your understanding of linear regression, select *Analyze* ◊ *Regression* ◊ *Linear*. Study the *Linear Regression* dialog box in Figure 2-2. Enter *rincome91* in the dependent variable panel at the top of the box and enter *educ* (the highest year of school completed) in the independents panel. Press the *Plots* button. In the *Linear Regression: Plots* dialog box, under *Standardized Residual Plots*, check *Normal probability plot*. Press *Continue*, then *OK* and observe your output.

7. Study your output in the viewing window, then print what's there for easier analysis. Study your results in relationship to your text discussion of regression analysis. You will see that your output gives you a great deal of information. Consult your instructor or a statistics text to see how you might want to use the analysis of variance (ANOVA) summary and how to interpret the unstandardized (B) and standardized (Beta) coefficients.

Simply following the text discussion and looking at the Normal P-P Plot of Regression in your output, see if you can follow how an estimate of the value of Y (the dependent variable) would plug in to the standard regression equation: **Y'** = **a** (the intercept value on the y axis) **+ bX** (the change on Y for 1 unit of X) where X which is years of education completed for a given case.

Figure 2-2: Linear Regression Dialog Box



Chapter 17 ***Reading and Writing Social Research***

Think up an interesting research question using variables found in one of the data sets you have been working with. Analyze your data with the appropriate SPSS procedures. You might use simple descriptive statistics, crosstabs, correlations and/or bivariate regression, depending on the type of data you're using and the question you're asking. Write up your results.