

Appendix C

Free and Open-Source Statistical Software

CHAPTER OBJECTIVES

- 1.** Explain why free and open-source software is helpful for statistical analysis.
- 2.** Compare and contrast PSPP and Statistical Package for the Social Sciences (SPSS) when approaching the main statistical tests discussed in the textbook.
- 3.** Compare and contrast R and SPSS when approaching the main statistical tests discussed in the textbook.
- 4.** Explore other free and open-source statistical software options.

Since publication of the first edition of *Quantitative Research Methods for Communication: A Hands-On Approach*, we have been searching for alternative statistical software packages that suit the needs and budgets of quantitative researchers. Although many free or open-source statistical software packages are available on the Internet, the two that we recommend are PSPP and R. Although we talked about these two software packages in various places throughout this book, we focused our attention on two more traditional statistical software packages: IBM's Statistical Package for the Social Sciences (SPSS) and Excel. In the first edition of this book, we also included Statistical Analysis System (SAS), but quickly realized that most people using this textbook were not actively using SAS for educational purposes.

We chose to include information on PSPP and R within this book because both of them are, first and foremost, free. Some of the more traditional statistical software packages are very expensive and are not cost-effective for some students and professors. Furthermore, some schools have even stopped providing undergraduate and graduate students access to some of the more expensive software packages, so PSPP and R are good alternatives.

PSPP and R are also good pedagogically speaking as well. PSPP, on the one hand, is written in a fashion that closely resembles SPSS, so students learning on PSPP can cross-transfer their computer skills to SPSS if they get the opportunity to further their statistical training. R, on the other hand, is a very powerful programming language and quickly becoming a software powerhouse internationally. In fact, R can do everything from simple descriptive statistics to advanced multivariate modeling, so students learning on R could use this program for all of their statistical needs from a beginning undergraduate course through a career in academia. The rest of this appendix discusses some of the nuances of both PSPP and R.

PSPP

PSPP is an open-source statistical package built on the same basic principles and functionality as SPSS. As an open-source alternative, the program is freely available for everyone under a GNU license through the Free Software Foundation. While you can burn copies of the software and have them available for purchase in your college/university bookstore (see the GNU policy on “free” - <https://www.gnu.org/philosophy/free-sw.en.html>), you can also make it freely available from your own website (like we’ve done) or direct students to download the software for free from the PSPP website. PSPP is available for both Windows and Apple operating systems, so everyone can easily download and install the software and be up and running in no time.

In these times when colleges and universities are drastically slashing budgets for software and other technical equipment, relying on open-source alternatives is very important. Thankfully, PSPP has filled a very important gap that is perfect as a substitute for SPSS. One note of caution, however, is needed. PSPP does not, nor will it ever, have the same complete functionality as SPSS. As a statistical tool, it is quite limited and does not have the same breadth of statistical features as proprietary software packages. But for a basic communication research course, it works very well as a substitute package.

In this section, we examine downloading and installing PSPP, the basic PSPP interface, and statistical analyses using PSPP.

Downloading and Installing PSPP

Downloading and installing PSPP is pretty easy whether you are installing the software on a PC or Mac computer. Start by going to the PSPP project’s main website (<http://www.gnu.org/software/pspp/pspp.html>). Once you are on the website (Figure C.1), you’ll find hyperlinks for a variety of different options. The one that we’re most concerned with at this point is getting and downloading PSPP, so select the one that is circled in Figure C.1, which is titled “Get PSPP.”

The screenshot shows the PSPP main website. At the top, there is a navigation bar with links for "Get PSPP", "FAQ", "Documentation", "Contribute", and "Quick Tour". Below this, a section titled "PSPP" describes the program as a free replacement for SPSS. A list of features follows, including support for over 1 billion variables, compatibility with SPSS syntax, and various output formats. A "USEFUL LINKS" section on the right provides links to news, mailing lists, developer resources, and other software. At the bottom, there is a footer with contact information and a copyright notice.

PSPP is a program for statistical analysis of sampled data. It is a Free replacement for the proprietary program SPSS, and appears very similar to it with a few exceptions.

The most important of these exceptions are, that there are no "time bombs"; your copy of PSPP will not "expire" or deliberately stop working in the future. Neither are there any on the number of cases or variables which you can use. There are no additional packages to purchase in order to get "advanced" functions; all functionality that PSPP currently has is included in the core package.

PSPP can perform descriptive statistics, T-tests, linear regression and non-parametric tests. Its backend is designed to perform its analyses as fast as possible, regardless of the input data. You can use PSPP with its graphical interface or the more traditional syntax commands.

A brief list of some of the features of PSPP follows:

- Supports over 1 billion cases.
- Supports over 1 billion variables.
- Syntax and data files are compatible with SPSS.
- Choice of terminal or graphical user interface.
- Choice of text, postscript or html output formats.
- Inter-operates with [Gnumeric](#), [OpenOffice.org](#) and other free software.
- Easy data import from spreadsheets, text files and database sources.
- Fast statistical procedures, even on very large data sets.
- No license fees.
- No expiration period.
- No unethical "end user license agreements".
- Fully indexed user manual.
- **Free Software**: licensed under [GPLv3](#) or later.
- Cross platform; Runs on many different computers and many different operating systems.

PSPP is particularly aimed at statisticians, social scientists and students requiring fast convenient analysis of sampled data.

Please send **FSF** & **GNU** inquiries to gnu@gnu.org. There are also other ways to contact the FSF. Please send broken links and other corrections (or suggestions) to bug-gnu-pspp@gnu.org. Copyright © 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007 Free Software Foundation, Inc., 51 Franklin St., Suite 330, Boston, MA 02110, USA - Verbatim copying and distribution of this entire article are permitted worldwide, without royalty, in any medium, provided this notice, and the copyright notice, are preserved.

Updated: \$Date: 2010/05/21 08:56:04 \$ \$Author: jmd \$

FIGURE C.1
PSPP Main Website

Once you're on the Get PSPP webpage (Figure C.2a), you'll need to scroll through the list of operating system options and find the one that corresponds to your computer's operating system. If you have a PC running Windows, then you'll want to select "Another Mingw-based version" (Figure C.2b). You then want to download the most recent build. When we wrote this appendix, the most recent build was on December 13, 2011, so we could have selected "PSPP-Master-2011-12-13" or the "PSPP-Master-single-user-2011-12-13" file (Figure C.2c). The single-user file is necessary only if you do not have administrative privileges on a given computer. Once you select the appropriate build, you will be taken to a downloading website called SourceForge (Figure C.2d). After five seconds (depending on your web browser), you will either receive a pop-up window (Figure C.2e) or a notification within the browser that a file is ready for download (Figure C.2f). Once you have downloaded the file, you can launch it and follow the onscreen instructions to install the program.

If you're running a Mac computer, you'll need to follow the instructions for the "Disk Image Installers" installation.

Both PC and Mac files are easily downloaded and installed, so this process should be fairly easy for you. If you run into trouble installing the program, we recommend talking to a computer specialist on your campus who is familiar with your computer's operating system. Because PSPP is an open-source software package, there is no support hotline. However, there is a PSPP wiki (<http://pspp.kiberpapa.org/wiki/doku.php>) and a pretty extensive user guide (<http://www.gnu.org/software/pspp/manual/pspp.pdf>).

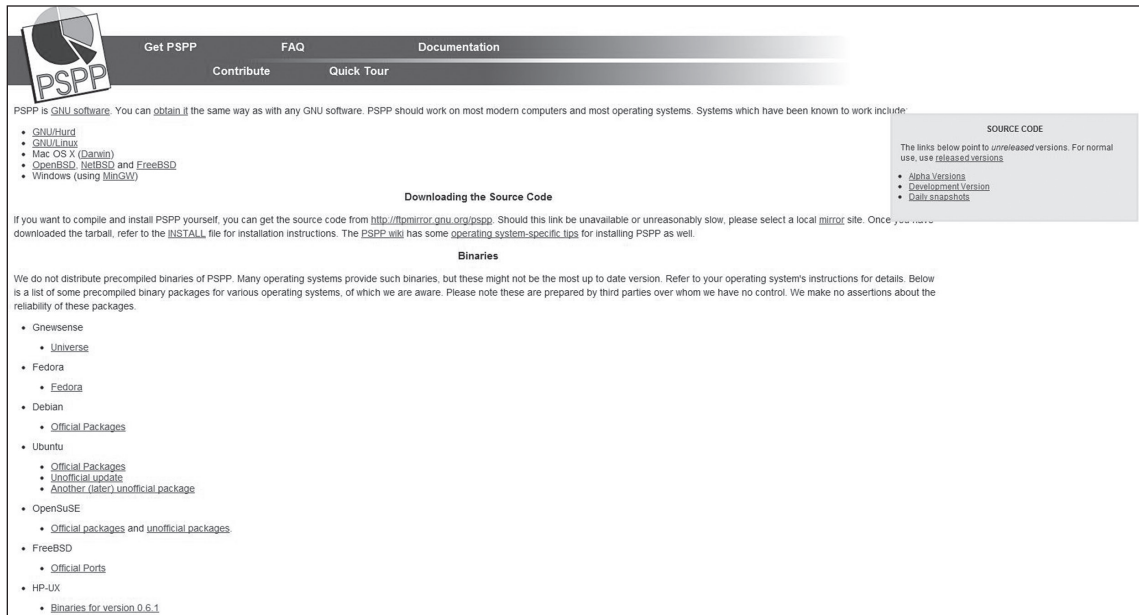


FIGURE C.2A
Get PSPP Main Page



FIGURE C.2B
Windows and Mac Options

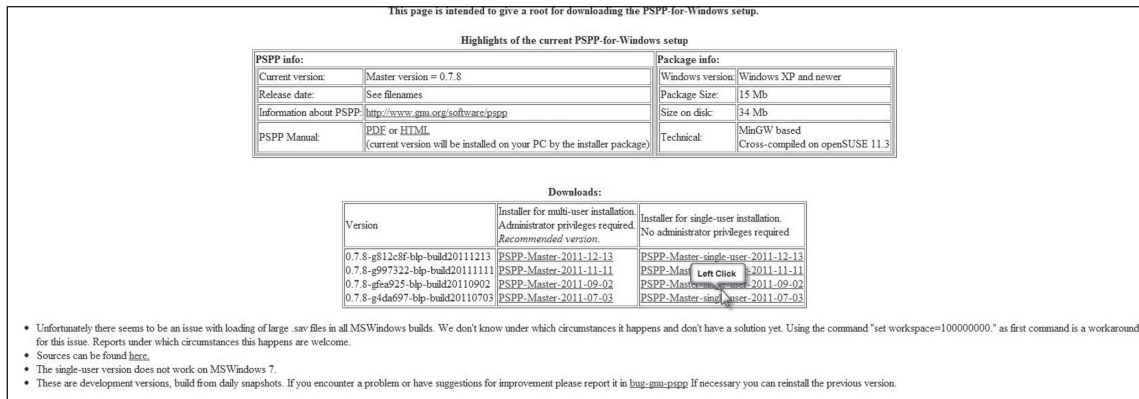


FIGURE C.2C
Select Version

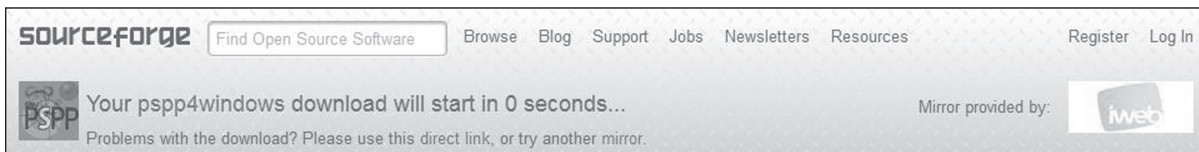


FIGURE C.2D
Redirect

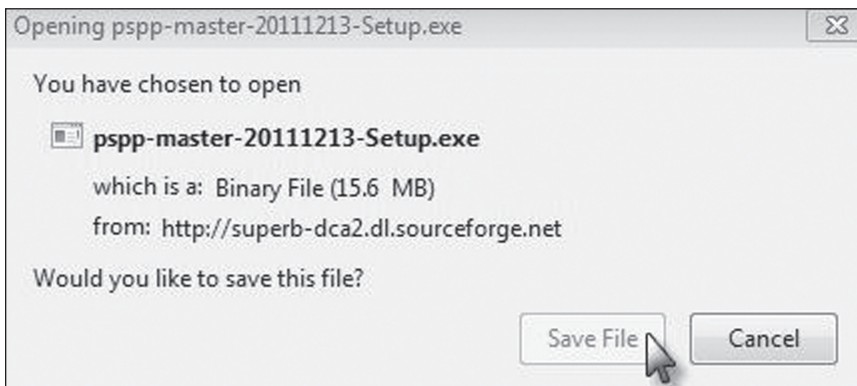


FIGURE C.2E
Save File

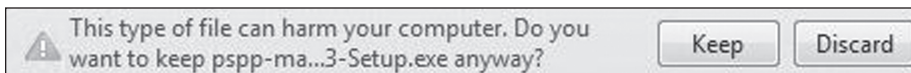


FIGURE C.2F
Download File

The Basic PSPP Interface

If you are familiar with Microsoft Excel or SPSS, then understanding the basic interface of PSPP is pretty easy. The program itself is designed to look and feel like any other spreadsheet program or statistical software package with a graphic user interface (GUI). Figure C.3 contains the main page of PSPP with the Recoded Data Set for the textbook open. You'll notice that if you look at the PSPP main page and the SPSS main page (Figure 8.4) the layout is identical.

Although the layout may be identical, we should note that the functionality between PSPP and SPSS is different. SPSS is a much more comprehensive statistical package and contains a considerable number of statistical options that PSPP does not. Remember, PSPP is an open-source software project with programmers from around the world participating in it as a hobby. Although these hobbyists have been improving the software since it originally launched in 1997, the program is still being developed and enhanced. As such, it is important to come back to the PSPP website periodically to ensure that you have the most recent build of the software itself. With each new build, you will need to download and reinstall the software.

	id	prca1	prca2	prca3	prca4	prca5	prca6	prca7	prca8	prca9	prca10	prca11	prca12	prca13	prca14	prca15	prca16
1	1.00	2.00	3.00	2.00	4.00	2.00	3.00	3.00	4.00	4.00	1.00	2.00	4.00	1.00	4.00	1.00	
2	2.00	1.00	1.00	3.00	2.00	4.00	5.00	5.00	1.00	1.00	1.00	1.00	2.00	3.00	1.00	1.00	
3	3.00	3.00	3.00	4.00	3.00	4.00	3.00	4.00	3.00	2.00	3.00	4.00	3.00	4.00	3.00	4.00	
4	4.00	1.00	2.00	4.00	2.00	4.00	2.00	1.00	2.00	2.00	2.00	2.00	2.00	3.00	2.00	4.00	2.00
5	5.00	2.00	2.00	3.00	2.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
6	6.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
7	7.00	1.00	1.00	2.00	2.00	2.00	1.00	2.00	2.00	1.00	1.00	1.00	2.00	2.00	1.00	2.00	
8	8.00	3.00	2.00	2.00	3.00	2.00	2.00	2.00	2.00	2.00	2.00	4.00	2.00	2.00	2.00	2.00	2.00
9	9.00	2.00	2.00	2.00	3.00	2.00	2.00	2.00	2.00	2.00	3.00	3.00	2.00	3.00	2.00	4.00	2.00
10	10.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	4.00	4.00	2.00	4.00	4.00	4.00	
11	11.00	1.00	4.00	1.00	3.00	2.00	3.00	2.00	4.00	3.00	1.00	1.00	3.00	4.00	1.00	1.00	
12	12.00	2.00	5.00	2.00	5.00	2.00	5.00	2.00	1.00	4.00	4.00	2.00	3.00	2.00	5.00	3.00	
13	13.00	3.00	2.00	4.00	4.00	4.00	2.00	4.00	2.00	3.00	1.00	4.00	4.00	5.00	1.00	4.00	
14	14.00	1.00	4.00	3.00	2.00	3.00	3.00	2.00	4.00	3.00	1.00	4.00	3.00	3.00	4.00	2.00	
15	15.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	2.00	2.00	2.00	1.00	1.00	
16	16.00	1.00	2.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	2.00	1.00	1.00	
17	17.00	1.00	1.00	1.00	1.00	2.00	2.00	3.00	3.00	4.00	2.00	2.00	3.00	3.00	3.00	2.00	
18	18.00	4.00	3.00	3.00	4.00	3.00	3.00	3.00	3.00	4.00	2.00	2.00	2.00	3.00	2.00	3.00	
19	19.00	4.00	2.00	2.00	3.00	3.00	2.00	3.00	2.00	2.00	2.00	3.00	2.00	3.00	2.00	2.00	
20	20.00	5.00	2.00	3.00	4.00	1.00	4.00	3.00	2.00	1.00	4.00	3.00	4.00	5.00	5.00	2.00	
21	21.00
22	22.00	3.00	3.00	3.00	3.00	4.00	3.00	4.00	4.00	2.00	4.00	4.00	4.00	3.00	5.00	3.00	
23	23.00	3.00	2.00	1.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	1.00	5.00	1.00	
24	24.00	2.00	2.00	3.00	3.00	4.00	3.00	3.00	2.00	3.00	2.00	2.00	3.00	3.00	4.00	3.00	
25	25.00	2.00	1.00	4.00	2.00	4.00	3.00	4.00	2.00	4.00	2.00	4.00	2.00	4.00	2.00	2.00	
26	26.00	1.00	1.00	1.00	1.00	2.00	2.00	1.00	2.00	1.00	1.00	1.00	2.00	2.00	3.00	1.00	

FIGURE C.3
PSPP Main Page

So, beyond the basic interface looking similar, the statistical functionality is also very similar to that of SPSS. For example, you can manage your data in the same fashion that you can with SPSS by examining either “Data View” or “Variable View” at the bottom left hand of the screen. When you load a dataset, the data are always shown to you in Data View. If you want to easily manage your data, click on Variable View (Figure C.4). In Variable View, you can see a number of factors related to your dataset. We’ve specifically circled three column headings at the top that are especially important for data management purposes.

The first column circled is titled “Name.” The Name column is what you label each individual variable within the dataset. You’ll notice in this image that we start with the name “ID,” which corresponds with the ID number of each participant in the study, from 1 to 654. After the ID, we have the 24 items of the Personal Report of Communication Apprehension–24 and first two items showing of the Generalized Ethnocentrism Measure. Notice, when we enter the data, that we have to number each and every item on a separate measure, so the item names start at “prca1” and go through “prca24” and so on.

The second column circled is titled “Label.” The Label column allows the researcher to type in more descriptive terms for each variable. For the first variable,

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	id	Numeric	8	2	Identification #	None	None	8	Right	Scale
2	prca1	Numeric	8	2	Communication A	None	None	8	Right	Scale
3	prca2	Numeric	8	2		None	None	8	Right	Scale
4	prca3	Numeric	8	2		None	None	8	Right	Scale
5	prca4	Numeric	8	2		None	None	8	Right	Scale
6	prca5	Numeric	8	2		None	None	8	Right	Scale
7	prca6	Numeric	8	2		None	None	8	Right	Scale
8	prca7	Numeric	8	2		None	None	8	Right	Scale
9	prca8	Numeric	8	2		None	None	8	Right	Scale
10	prca9	Numeric	8	2		None	None	8	Right	Scale
11	prca10	Numeric	8	2		None	None	8	Right	Scale
12	prca11	Numeric	8	2		None	None	8	Right	Scale
13	prca12	Numeric	8	2		None	None	8	Right	Scale
14	prca13	Numeric	8	2		None	None	8	Right	Scale
15	prca14	Numeric	8	2		None	None	8	Right	Scale
16	prca15	Numeric	8	2		None	None	8	Right	Scale
17	prca16	Numeric	8	2		None	None	8	Right	Scale
18	prca17	Numeric	8	2		None	None	8	Right	Scale
19	prca18	Numeric	8	2		None	None	8	Right	Scale
20	prca19	Numeric	8	2		None	None	8	Right	Scale
21	prca20	Numeric	8	2		None	None	8	Right	Scale
22	prca21	Numeric	8	2		None	None	8	Right	Scale
23	prca22	Numeric	8	2		None	None	8	Right	Scale
24	prca23	Numeric	8	2		None	None	8	Right	Scale
25	prca24	Numeric	8	2		None	None	8	Right	Scale
26	ethno1	Numeric	8	2	Ethnocentrism	None	None	8	Right	Scale
27	ethno2	Numeric	8	2		None	None	8	Right	Scale

FIGURE C.4
PSPP Variable
View

we typed in the label “Identification #” to ensure that users would know what the less descriptive name “ID” stood for within the dataset. You’ll notice that we don’t label every item with a different name because this almost becomes overkill. However, if you scroll to the bottom of the dataset, you’ll find that we do provide comprehensive labels for each of the summed variable totals. You can see these labels in Figure C.5.

The third column circled in Figure C.3 is titled “Measure.” This column helps a researcher organize the various variables in the four levels of variables into three distinct groupings: nominal, ordinal, and scale (interval/ratio). If you scroll through the entire dataset, you’ll see that most of the variables are listed as scale data because they generally are part of an interval measure. However, if you look in

FIGURE C.5
PSPP Variable
View—Labels

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
141	sex	Numeric	8	2	Biological Sex	(1,00,'Male')_	None	8	Right	Scale
142	politics	Numeric	8	2	Political Affiliation	(1,00,'Democrat')_	None	8	Right	Nominal
143	class	Numeric	8	2	School Classification	(1,00,'First Year Stud')_	None	8	Right	Ordinal
144	online	Numeric	8	2	Time Spent Online	(1,00,'0-5 hour')_	None	8	Right	Ordinal
145	age	Numeric	8	2	Age	None	None	8	Right	Scale
146	edition	Numeric	8	2	Book Edition	(1,00,'1st Edition - 2')	None	8	Right	Nominal
147	Group_CA	Numeric	8	2	Group CA	None	None	10	Right	Scale
148	Meetings_CA	Numeric	8	2	Meetings CA	None	None	13	Right	Scale
149	Interpersonal_C	Numeric	8	2	Interpersonal CA	None	None	18	Right	Scale
150	PublicSpeaking	Numeric	8	2	Public Speaking CA	None	None	19	Right	Scale
151	Big_CA	Numeric	8	2	Communication Apprehension Total	None	None	10	Right	Scale
152	Ethnocentrism	Numeric	8	2	Ethnocentrism	None	None	15	Right	Scale
153	Humor_Assessm	Numeric	8	2	Humor Assessment	None	None	18	Right	Scale
154	Nonverbal_Imm	Numeric	8	2	Nonverbal Immediacy	None	None	21	Right	Scale
155	Assertiveness	Numeric	8	2	Assertiveness	None	None	15	Right	Scale
156	Responsiveness	Numeric	8	2	Responsiveness	None	None	16	Right	Scale
157	Group_WTC	Numeric	8	2	Group WTC	None	None	11	Right	Scale
158	Meetings_WTC	Numeric	8	2	Meetings WTC	None	None	14	Right	Scale
159	Interpersonal_W	Numeric	8	2	Interpersonal WTC	None	None	19	Right	Scale
160	PublicSpeaking	Numeric	8	2	Public Speaking WTC	None	None	20	Right	Scale
161	Stranger_WTC	Numeric	8	2	Stranger WTC	None	None	14	Right	Scale
162	Acquaintance_W	Numeric	8	2	Acquaintance WTC	None	None	18	Right	Scale
163	Friend_WTC	Numeric	8	2	Friend WTC	None	None	12	Right	Scale
164	Big_WTC	Numeric	8	2	Willingness to Communicate Total	None	None	10	Right	Scale
165	PublicSpeaking	Numeric	8	2	Everyone should be required to take public speaking	None	None	23	Right	Scale
166	Attitude_About	Numeric	8	2	Attitude about Higher Education	None	None	28	Right	Scale

Figure C.6a, you'll see that the variable political affiliation is defined as a nominal variable.

While not circled in Figure C.4, there is another very important column for you to understand the Variable View in PSPP. In Figure C.6a, you'll see that we have circled the "Values" column. If you remember from your results looking at SPSS earlier in the textbook, instead of seeing biological sex represented as "1" or "2" in various printouts, you actually see the labels "female" and "male." To get these labels to exist in PSPP or SPSS printouts, we need to provide those labels here. Figure C.6b shows you the labels dialogue box for biological sex. In this dialogue box, you can see that when we originally coded the data, males in the sample were coded as "1" and females in the sample were coded as "2". Now if we run a frequency table of the female and male breakdown of the sample, we'll see the words "female" and "male" appear in the frequency table instead of the numerical values (Figure C.6c).

Now that we've explained the basic user interface of PSPP, let's examine some of the statistical analyses possible in PSPP.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
sex	Numeric	8	2	Biological Sex	1.00, 'Male'	None	8	Right	Scale
politics	Numeric	8	2	Political Affiliation	(1.00, 'Democrat')_	None	8	Right	Nominal
class	Numeric	8	2	School Classification	(1.00, 'First Year Student')_	None	8	Right	Ordinal
online	Numeric	8	2	Time Spent Online	(1.00, '0-.5 hour')_	None	8	Right	Ordinal
age	Numeric	8	2	Age	None	None	8	Right	Scale
edition	Numeric	8	2	Book Edition	(1.00, '1st Edition - 2008')_	None	8	Right	Nominal

FIGURE C.6A
Values Column

The dialog box titled "Value Labels" contains the following elements:

- Input fields for "Value:" and "Value Label:".
- Buttons: "Add", "Apply", "Remove", "OK", "Cancel", and "Help".
- A list box containing:
 - 1.00 = 'Male'
 - 2.00 = 'Female'

FIGURE C.6B
Values Dialogue Box

```

FREQUENCIES
FREQUENCIES
/VARIABLES= sex
/FORMAT=AVALUE TABLE
/STATISTICS=NONE.
    
```

Biological Sex

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Male	1.00	320	48.93	49.92	49.92
Female	2.00	321	49.08	50.08	100.00
.	.	13	1.99	Missing	
Total		554	100.0	100.0	

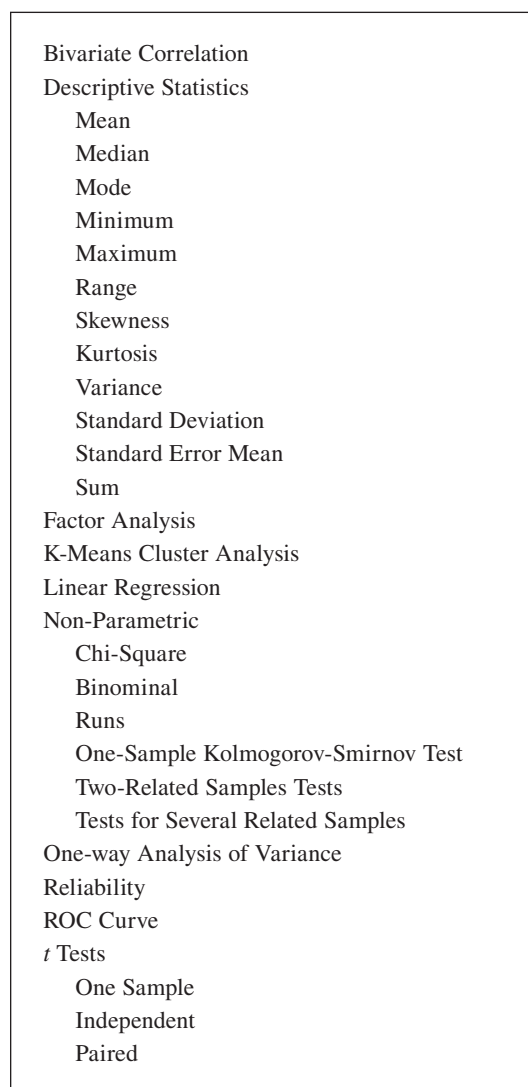
FIGURE C.6C
Frequency Table

Statistical Analyses Using PSPP

PSPP is a pretty comprehensive statistical program. All of the statistical tests that we demonstrate in the book and the online materials (with the exception of those in Appendix E) can be conducted using PSPP. Figure C.7 shows a comprehensive list of the different tests that can be conducted in PSPP.

Running statistics in PSPP is very similar to running statistics in SPSS. In fact, the basic process starts the same. In SPSS, all of the tests we discussed were under the “Analyze” drop-down menu, which is also true in PSPP. Figure C.8 shows you the “Compare

FIGURE C.7
Statistical Tests
in PSPP



Bivariate Correlation
Descriptive Statistics
Mean
Median
Mode
Minimum
Maximum
Range
Skewness
Kurtosis
Variance
Standard Deviation
Standard Error Mean
Sum
Factor Analysis
K-Means Cluster Analysis
Linear Regression
Non-Parametric
Chi-Square
Binominal
Runs
One-Sample Kolmogorov-Smirnov Test
Two-Related Samples Tests
Tests for Several Related Samples
One-way Analysis of Variance
Reliability
ROC Curve
<i>t</i> Tests
One Sample
Independent
Paired

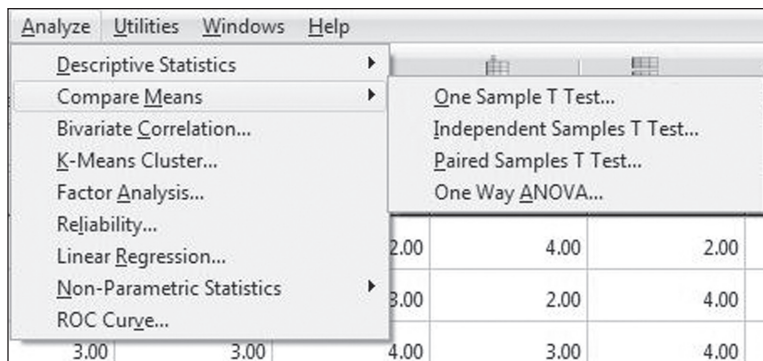


FIGURE C.8
PSPP Analyze Menu

Means” menu in the Analyze drop-down menu. You’ll notice that the PSPP menu is considerably shorter in terms of the number of statistical categories when you compare it to the Analyze menu seen in SPSS.

Not only is PSPP’s Analyze drop-down menu designed to be used in a fashion similar to SPSS, the processes you use to run the different statistical tests and then the printouts of the results also looks the same. Again, you can use all of the information within the chapters of this book related to SPSS as a general guide for running and interpreting results in PSPP.

R

R is one of the most powerful free open-source statistical software packages anywhere. In fact, R is quickly becoming a force to be reckoned with. To learn more about the prominence of R in various research capacities, check out Robert A. Muenchen’s (2010a) article “The Popularity of Data Analysis Software.” But we’re betting you’re not here to learn about R’s popularity, but to learn how to use R. Well, we are not R experts. In fact, we have a steep learning curve (just like you do) ahead of us if we’re going to become proficient at using R.

According to Kurt Hornik (2011), R, first and foremost, is a computer language that stems from a language originally developed at Bell Labs by Becker, Chambers, and Wilks called S in combination with Sussman’s Scheme programming language. R itself was developed by a working team at the University of Auckland, New Zealand. Since the mid-1990s, a group of computer scientists and statisticians called the “R Core Team” have continuously updated the source code database making R a growing statistical software that can be utilized by anyone for free. Furthermore, the international nature of R has caused it to become a widely used statistical software language similar in size and scope to SAS. Thankfully, we will not need to learn the software language to utilize the basic statistical properties of R itself.

In this section, we examine downloading and installing R and R-Commander, the basic R-Commander interface, and statistical analyses using R-Commander.

Downloading and Installing R and R-Commander

The first step in downloading and installing R is to visit the R website (<http://www.r-project.org/>). When you arrive at the website, you'll see a variety of hyperlinks (Figure C.9). Find the hyperlink (circled in Figure C.9) that says "download R." Once you're on the download page, you'll see a variety of CRAN (Comprehensive R Archive Network) Mirror sites that you can choose from (Figure C.10a). Depending on where you are in the world, you'll want to select a CRAN Mirror site that is closer to you geographically. The closer a mirror site is to you, the faster the download time will be (generally speaking). Another factor that affects download time is the amount of server allotment the CRAN Mirror is given at the hosting institution. If you ever find one CRAN Mirror site taking forever, try a different one. Obviously, the speed of your own Internet connection will also affect the download times. In Figure C.10b, the cursor is pointing to the CRAN Mirror hosted by Carnegie Mellon University, which is one of the oldest in the United States and has pretty fast download speeds for those of us in America.

We should also mention a proprietary version of R, which comes with software support, called Revolution. Revolution (<http://www.revolutionanalytics.com/>) is freely downloadable for students and faculty, and you can then purchase a technical support plan as necessary. Revolution is the same basic R engine; it's just designed to be a little more accessible for those in business and education who need technical assistance.

When you select your CRAN Mirror, you will find a webpage that looks similar to what you see in Figure C.10c. Depending on whether you are installing R on a PC, Mac, or Linux machine, click the appropriate hyperlink. For our purposes, we're going

The R Project for Statistical Computing

PCA 5 vars

```
prcomp(x = data[, 1:5])
```

Clustering 4 groups

Factor 1 [41%]

Factor 3 [19%]

Getting Started:

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOs. **To download R, please choose your preferred CRAN mirror.**
- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News:

- R version 2.14.1 (December Snowflakes) has been released on 2011-12-22.
- The R Journal Vol.3/2 is available.
- useR! 2012. will take place at Vanderbilt University, Nashville Tennessee, USA, June 12-15, 2012.
- R version 2.13.2 has been released on 2011-09-30.

This server is hosted by the Institute for Statistics and Mathematics of the WU Wien.

FIGURE C.9
R Home Page



The Comprehensive R Archive Network is available at the following URLs, please choose a location close to you. Some statistics on the status of the mirrors can be found here: [main page](#), [windows release](#), [windows old release](#).

Country	URL	Organization
Argentina	http://mirror.fcaglp.unlp.edu.ar/CRAN/	Universidad Nacional de La Plata
	http://r.mirror.mendoza-conicet.gob.ar/	CONICET Mendoza
Australia	http://cran.csiro.au/	CSIRO
	http://cran.ms.unimelb.edu.au/	University of Melbourne
Austria	http://cran.atr-project.org/	Wirtschaftsuniversitaet Wien
Belgium	http://www.freeststatistics.org/cran/	K.U.Leuven Association
Brazil	http://cran-r.c3d.ufpr.br/	Universidade Federal do Parana
	http://cran.focruz.br/	Oswaldo Cruz Foundation, Rio de Janeiro
	http://www.vps.fuvz.usp.br/CRAN/	University of Sao Paulo, Sao Paulo
	http://brieger.esalq.usp.br/CRAN/	University of Sao Paulo, Piracicaba
Canada	http://cran.stat.sfu.ca/	Simon Fraser University, Burnaby
	http://mirror.its.dal.ca/cran/	Dalhousie University, Halifax
	http://probability.ca/cran/	University of Toronto
	http://cran.skazkaforvov.com/	iWeb, Montreal
	http://cran.parentingamerica.com/	iWeb, Montreal
Chile	http://dirichlet.mat.puc.cl/	Pontificia Universidad Catolica de Chile, Santiago
China	http://ftp.ctex.org/mirrors/CRAN/	CTEX.ORG
	http://cran.csdl.cn/	Computer Network Information Center, CAS, Beijing
	http://mirror.bnu.edu.cn/cran/	Beijing Jiaotong University, Beijing
	http://mirrors.usc.edu.cn/CRAN/	University of Science and Technology of China
	http://mirrors.xmu.edu.cn/CRAN/	Xiamen University

Navigation Links:

- About R
- What is R?
- Contributors
- Screenshots
- What's new?
- Download, Packages
- CRAN
- R Project
- Foundation
- Members & Donors
- Mailing Lists
- Bug Tracking
- Developer Page
- Conferences
- Search
- Documentation
- Manuals
- FAQs
- The R Journal
- Wiki
- Books
- Certification
- Other
- Misc
- Bioconductor
- Related Projects
- User Groups
- Links

FIGURE C.10A
CRAN Mirrors

USA

http://cran.opensourceresources.org/	opensourceresources.org
http://cran.cnr.Berkeley.edu	University of California, Berkeley, CA
http://cran.stat.ucla.edu/	University of California, Los Angeles, CA
http://streaming.stat.iastate.edu/CRAN/	Iowa State University, Ames, IA
http://rweb.quant.ku.edu/cran/	University of Kansas, Lawrence, KS
http://software.rc.fas.harvard.edu/mirrors/R/	Harvard University, Cambridge, MA
http://watson.nci.nih.gov/cran_mirror/	National Cancer Institute, Bethesda, MD
http://cran.mtu.edu/	Michigan Technological University, Houghton, MI
http://cran.wustl.edu/	Washington University, St. Louis, MO
http://cran.case.edu/	Case Western Reserve University, Cleveland, OH
http://ftp.osuosl.org/pub/cran/	Oregon State University
http://lib.stat.cmu.edu/R/CRAN/	Statlib, Carnegie Mellon University, Pittsburgh, PA
http://cran.mirrors.hoobly.com	Hoobly Classifieds, Pittsburgh, PA
http://mirrors.nics.utk.edu/cran/	National Institute for Computational Sciences, Oak Ridge, TN
http://cran.revolutionanalytics.com	Revolution Analytics, Dallas, TX
http://cran.sixsigmaonline.org/	sixsigmaonline.org, Houston, TX
http://cran.fhcr.c.org/	Fred Hutchinson Cancer Research Center, Seattle, WA
http://cran.cs.wvu.edu/	Western Washington University, Bellingham, WA

FIGURE C.10B
Select a CRAN Mirror Site

to select PC (Figure C.10d). You'll want to follow the hyperlink titled "install R for the first time." If you are a programmer and want the specific build specifications, then the other links could be useful, but everyone else will want to install R for the first time.

As of January 1, 2012, when this appendix was written, the last complete build of R was version 2.14.1 for Windows. Obviously, this number will change depending on when you purchase this book, so do not be surprised if this number has gone up by

the time you are reading this appendix. At this point, just click the hyperlink “Download R 2.14.1 for Windows” (Figure C.10e), and then click the “Save File” function (Figure C.10f). Once the download is complete, just launch the executable file and follow the onscreen instructions for installing R.

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for MacOS X](#)
- [Download R for Windows](#)

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2011-12-22, December Snowflakes): [R-2.14.1.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features](#) and [bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

FIGURE C.10C
Download Instructions and Links

R for Windows

Subdirectories:

base	Binaries for base distribution (managed by Duncan Murdoch). This is what you want if you install R for the first time .
contrib	Binaries of contributed packages (managed by Uwe Ligges)

Please do not submit binaries to CRAN. Package developers might want to contact Duncan Murdoch or Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

FIGURE C.10D
R for Windows

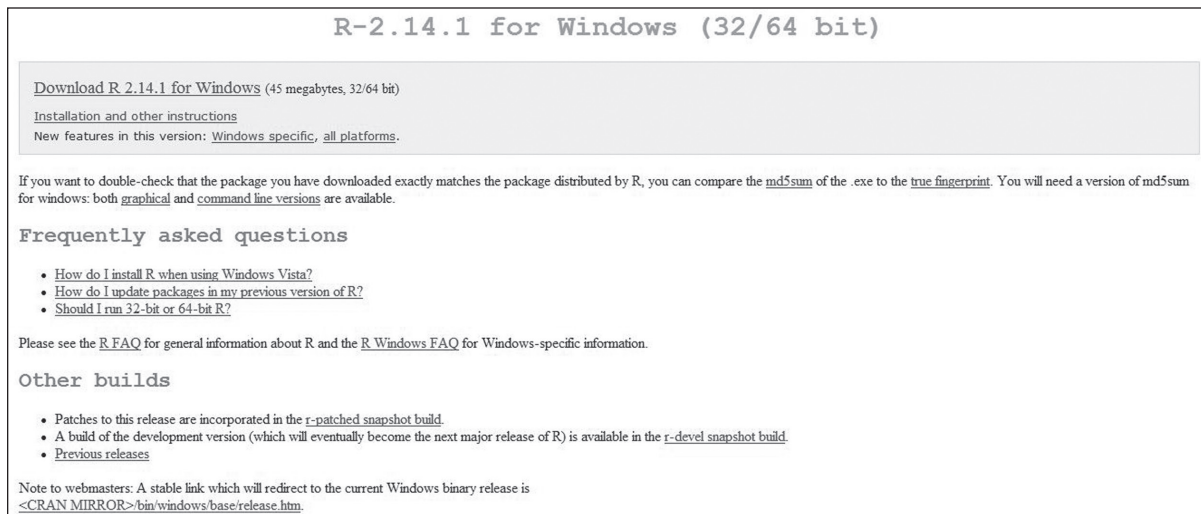


FIGURE C.10E
R-2.14.1 for Windows

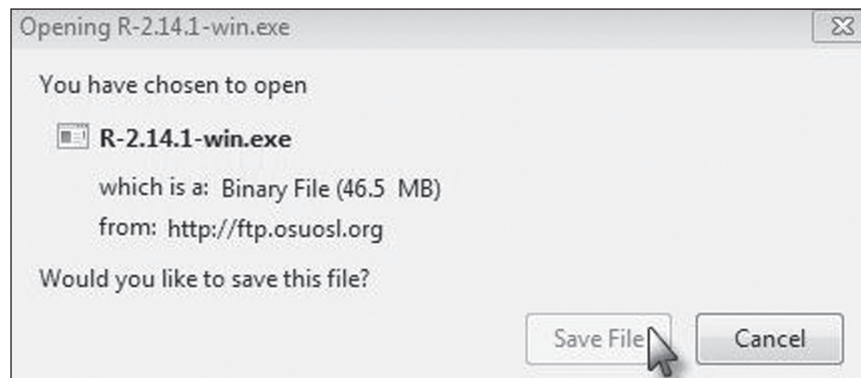


FIGURE C.10F
Save File

The Basic R-Commander Interface

Unlike PSPP, when R is finished downloading, you are not really done installing R. The first time you launch R, you'll be presented with the screen you see in Figure C.11. At this point, you need to do a couple of other installs before we can actually start utilizing R.

First, you need to go to the drop-down menu titled "Packages" and then select "Install package(s)" (Figure C.12a). You will then be asked to once again select a CRAN Mirror site for the rest of your installation period (Figure C.12b). Once you have selected the appropriate CRAN site, you will then be presented with a long list of possible packages

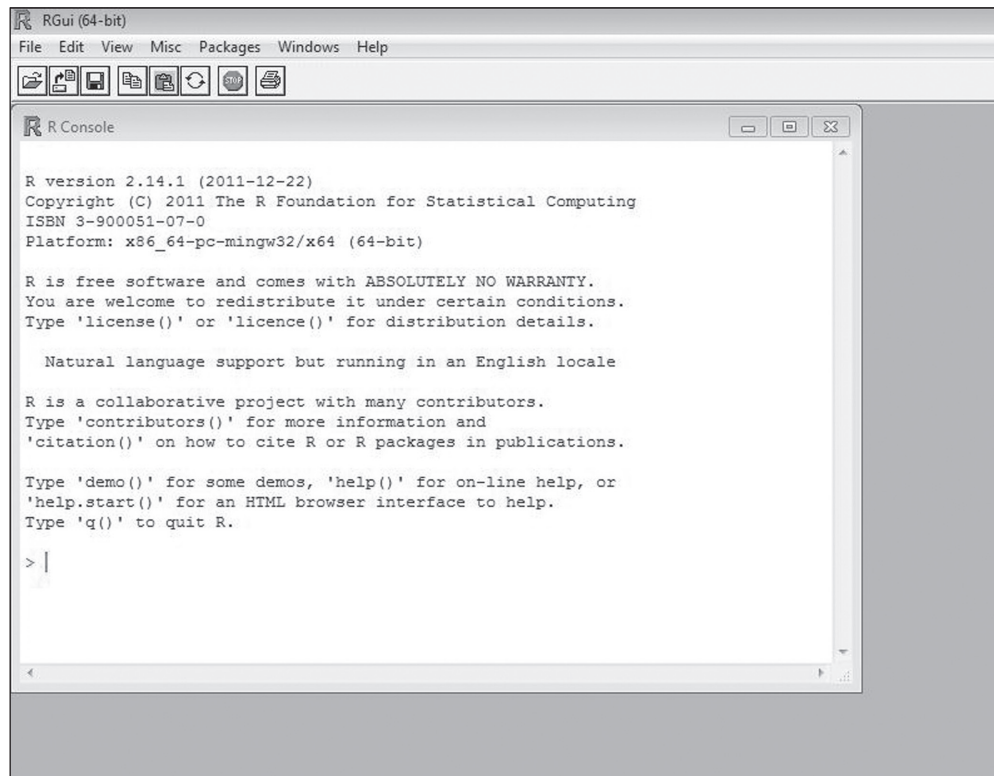
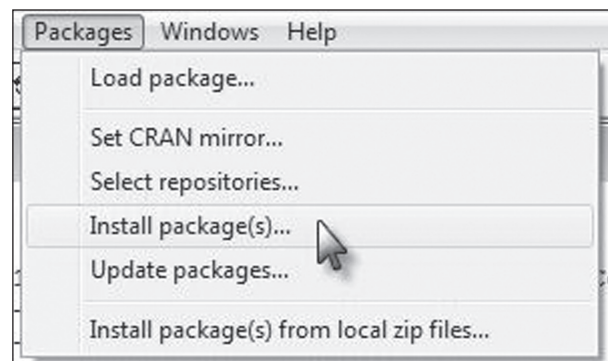


FIGURE C.11
R Main Page

that you can install. Scroll down the list until you find the one called “Rcmdr,” or R-Commander (Figure C.12c). Select Rcmdr, and click OK. Immediately, the program will start downloading Rcmdr for you into R.

Once Rcmdr has finished downloading, you can now load the package. Once again, go up to the “Packages” drop-down menu, and this time, select “Load Package” (Figure C.12d). You will then be presented with a pop-up menu that contains

FIGURE C.12A
Install Packages



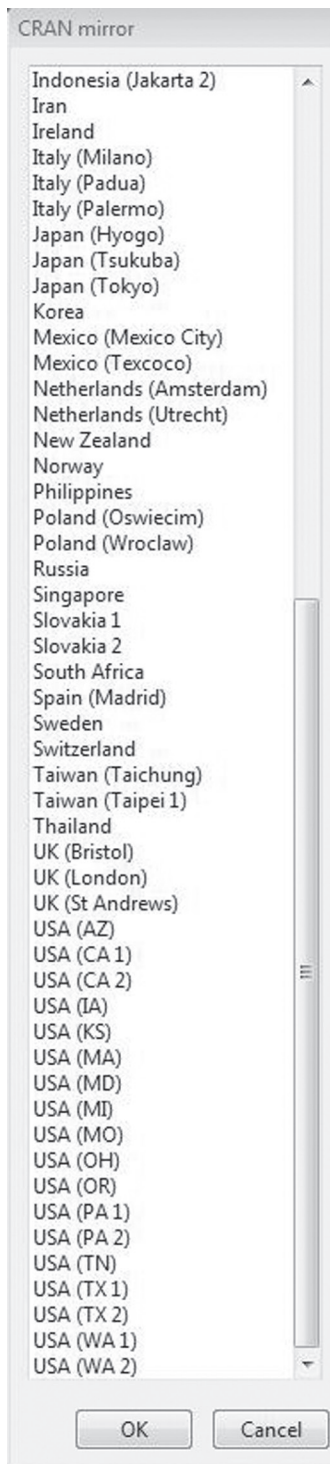


FIGURE C.12B
Select CRAN Mirror

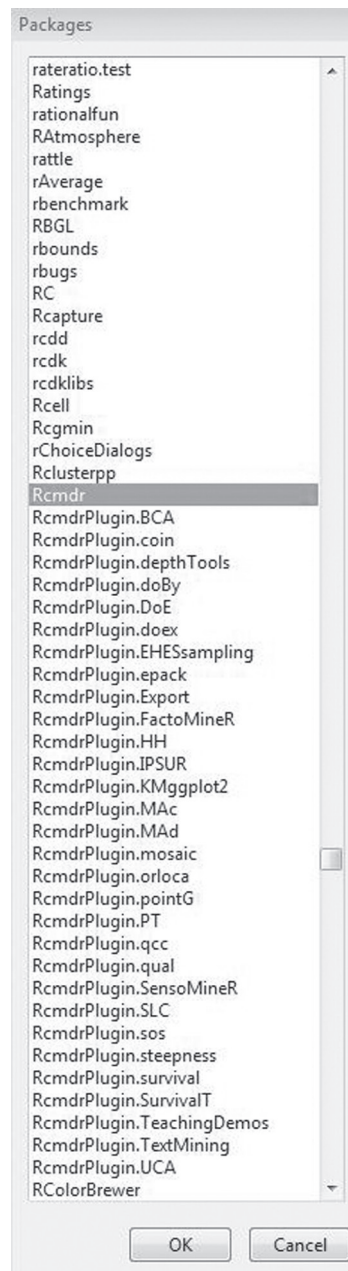


FIGURE C.12C
Rcmdr

FIGURE C.12D
Load Package

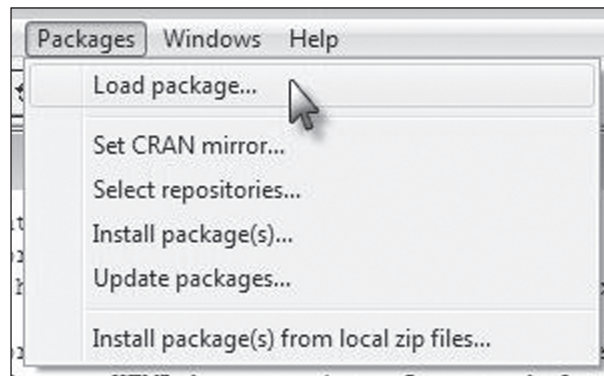
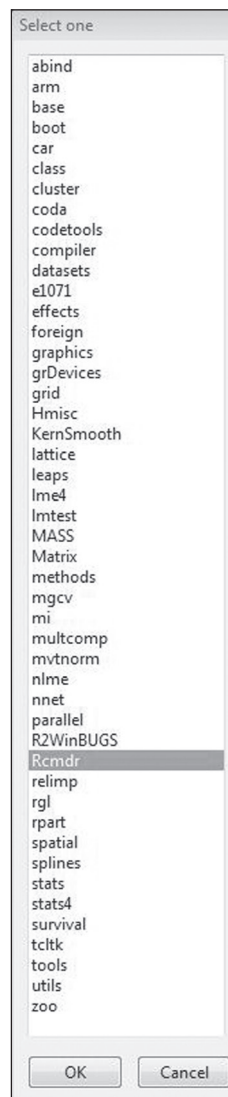


FIGURE C.12E
Installed Packages



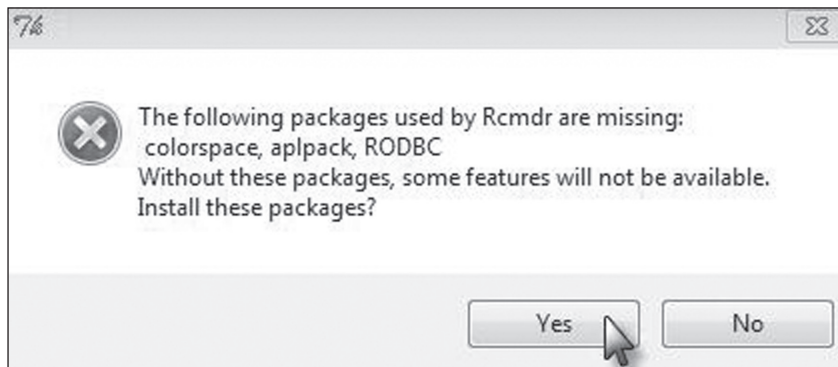


FIGURE C.12F
Rcmdr Warning Screen

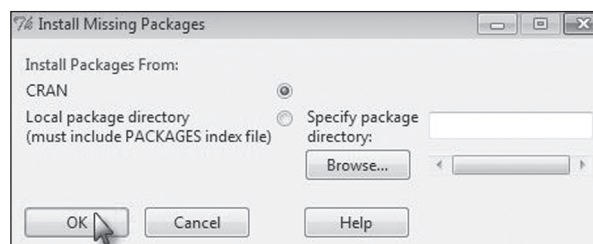


FIGURE C.12G
Install Missing Packages Dialogue Box

all of the preloaded R packages (now including Rcmdr). Find Rcmdr, and click OK (Figure C.12e). Because this is the first time you are loading Rcmdr, there are a number of other packages that Rcmdr needs to load for you to use the program. You will be shown a warning message (Figure C.12f) telling you that other packages are needed; just click OK. You will then be presented with another dialogue box asking you where you want to load those other packages from (Figure C.12g). If you have just downloaded Rcmdr from a CRAN Mirror site, click OK, and the program will immediately start downloading the necessary packages.

Statistical Analyses Using R-Commander

Figure C.13 is what the loaded version of Rcmdr looks like. You'll notice that the GUI is quite different from what you are used to seeing in SPSS, PSPP, or Excel. As such, it does take a little bit of an adjustment. So, let's start by loading the R dataset that goes along with the textbook.

To load a dataset, go to the drop-down menu in Rcmdr titled "Data," and then select "Load Data Set" (Figure C.14). You will then need to search your hard drive for the dataset and click OK. At this point, the dataset will be loaded; however, you will not see the dataset in the main screen like you do in other software packages. If you want to see the dataset, there are two buttons at the top of Rcmdr to help you view your data. The first button is titled "Edit data set," and allows you to actually change the numbers

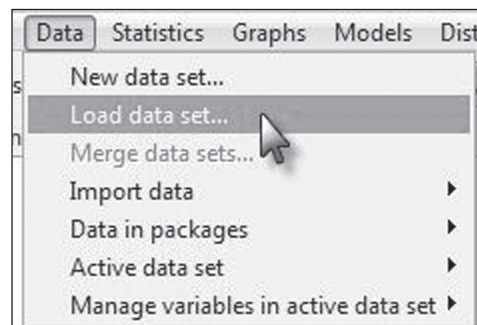
FIGURE C.13
R-Commander
(Rcmdr) Main Screen



within your dataset (Figure C.15). When you click on this button, an editable spreadsheet with all of your variables pops-up (Figure C.16). The View data set button will present you with a similar pop-up window, but you will not be able to edit the information within the window.

As for calculating statistics within R, instead of going to the “Analyze” button, you will click on the “Statistics” button. For example, if you want to calculate the frequencies of females and males within the dataset, you’ll go to the Statistics drop-down menu, choose “Summaries,” and then click on “Frequency distributions”

FIGURE C.14
Load Dataset



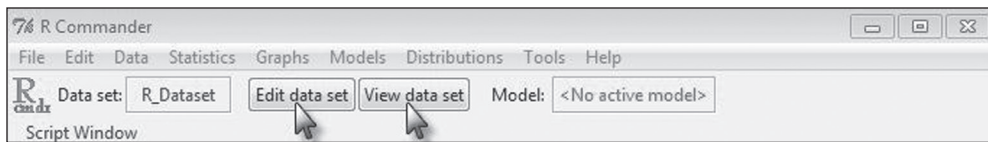


FIGURE C.15
Editing and View Data Menu

 The screenshot shows the R Data Editor window. The title bar reads 'R Data Editor'. The window contains a table with 11 columns and 19 rows. The columns are labeled 'id', 'prca1', 'prca2', 'prca3', 'prca4', 'prca5', 'prca6', 'prca7', 'prca8', and 'prca9'. The rows are numbered 1 through 19. The data values are integers ranging from 1 to 5.

	id	prca1	prca2	prca3	prca4	prca5	prca6	prca7	prca8	prca9
1	1	2	3	2	4	2	3	3	4	4
2	2	1	1	3	2	4	5	5	1	1
3	3	3	3	4	3	4	3	4	3	2
4	4	1	2	4	2	4	2	1	2	2
5	5	2	2	3	2	3	3	3	3	3
6	6	5	5	5	5	5	5	5	5	5
7	7	1	1	2	2	2	1	2	2	1
8	8	3	2	2	3	2	2	2	2	2
9	9	2	2	2	3	2	2	2	2	3
10	10	5	5	5	5	5	5	5	5	5
11	11	1	4	1	3	2	3	2	4	3
12	12	2	5	2	5	2	5	2	1	4
13	13	3	2	4	4	4	2	4	2	3
14	14	1	4	3	2	3	3	2	4	3
15	15	1	1	1	1	1	1	1	1	1
16	16	1	2	1	1	1	1	1	1	1
17	17	1	1	1	1	2	2	3	3	4
18	18	4	3	3	4	3	3	3	3	4
19	19	4	2	2	3	3	2	3	2	2

FIGURE C.16
Viewing and Editing
Dataset

(Figure C.17a). In the Frequency Distributions dialogue box, you'll see all of the variables within the dataset listed alphabetically (Figure C.17b). Scroll down until you find the variable titled "sex," highlight it, and then click OK (Figure C.17c). Once you've done that, you'll see that the output is now available in the R-Commander main screen in the Output Window. The first thing you'll notice about this output is that it is not presented in user-friendly tables like you see in SPSS, PSPP, or Excel. Instead, the information is presented to you in writing. Next to the two arrows in Figure C.17d, you'll see the results first for the frequency counts and then for the percentages of females and males in the sample. Notice that it does not provide you with information on missing data.

A Note On Other R Packages

Now that we've looked at the basic functionality of R, we highly encourage you to play around with Rcmdr and become familiar with where all the drop-down menus are in this program. Again, it's distinctly different from the other three software programs discussed in the book, so it does take a little more learning than we can present here.

FIGURE C.17A
Statistics Menu

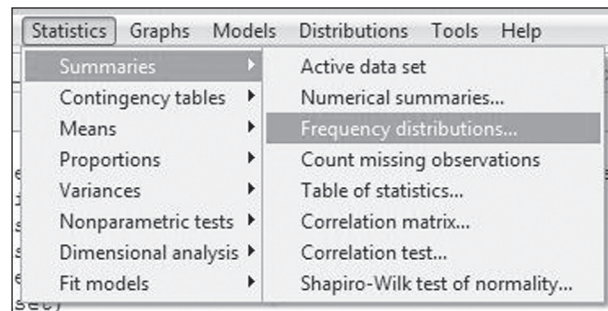


FIGURE C.17B
Frequency Distributions
Dialogue Box

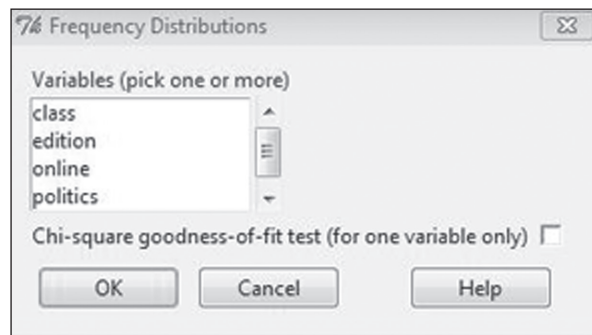
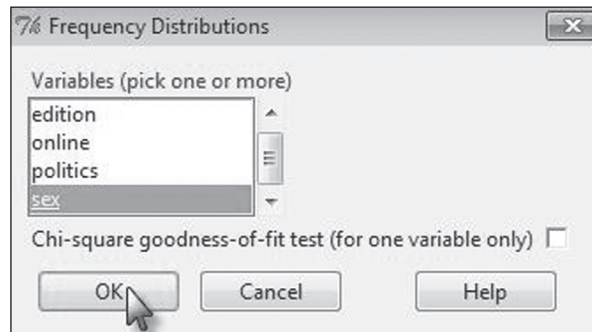


FIGURE C.17C
Frequency Distributions
Dialogue Box—
Sex Chosen



To help you learn more about R, there are a number of great books and articles on the subject that can be purchased or downloaded (Fox & Weisberg, 2011; Hogan, 2010; Muenchen, 2009, 2010a, 2010b).

One of the nice aspects of R is that it is very versatile, and there are hundreds of different add-on statistical packages that you could theoretically use. As someone who is just learning quantitative research methods, probably all that you need at this point is Rcmdr. In fact, Rcmdr has many advanced features we do not even come close to exploring in this text. However, R can be expanded as your knowledge of quantitative

```

R Commander
File Edit Data Statistics Graphs Models Distributions Tools Help
Data set: R_Dataset Edit data set View data set Model: <No active model>

Script Window
View(R_Dataset)
View(R_Dataset)
fix(R_Dataset)
View(R_Dataset)
fix(R_Dataset)
fix(R_Dataset)
.Table <- table(R_Dataset$sex)
.Table # counts for sex
round(100*.Table/sum(.Table), 2) # percentages for sex
remove(.Table)

Output Window
> fix(R_Dataset)
> View(R_Dataset)
> fix(R_Dataset)
> fix(R_Dataset)
> .Table <- table(R_Dataset$sex)
> .Table # counts for sex
Male Female
320 321
> round(100*.Table/sum(.Table), 2) # percentages for sex
Male Female
49.92 50.08
> remove(.Table)

Messages
[5] NOTE: The dataset R_Dataset has 654 rows and 166 columns.
[6] NOTE: The dataset R_Dataset has 654 rows and 166 columns.

```

FIGURE C.17D
Frequency Results

research methods expands. All of the advanced statistics discussed in Appendix E have packages that can be installed into R to help you run those tests. However, once you start delving into statistics outside of Rcmdr, you'll also need to become more versed in R as a programming language as well.

Because R is a constantly growing and evolving software package, we recommend reading *The R Journal* to learn more about the basic and advanced statistical computing packages available to users (<http://journal.r-project.org/>). One of the interesting and problematic aspects of R is that anyone can create and upload a new statistical package. To maintain some level of quality of control, we recommend only using those statistical packages that have articles within *The R Journal* discussing those packages and their statistical attributes.

CONCLUSION

This appendix has shown you two options for free and open-source statistical software. Although both PSPP and R have their strengths and weaknesses, either can be used for learning basic quantitative communication research. Whatever

statistical package you use, you'll need to spend time becoming comfortable with the software.

Beyond SPSS, Excel, PSPP, and R, however, are a range of other statistical software packages that are equally viable. For example, the very first edition of the textbook included how to use SAS

instead of Excel. One of our coauthors primarily uses STATA in his research because it contains a very robust statistical software package for a low cost (when compared to other software packages). Figure C.18 contains a list of other free and open-source statistical software packages.

FIGURE C.18
Free and Open-Source
Software

Software Name	Software Website
Gnumeric Spreadsheet	http://www.gnumeric.org/
jamovi	https://www.jamovi.org/
JASP*	https://jasp-stats.org/
MicrOsiris	http://www.microsiris.com/
OpenStat	http://openstat.info/OpenStat_Features.htm
Salstat	https://www.salstat.com/
SAS University Edition*	https://www.sas.com/en_us/software/university-edition.html
SOFA Statistics	http://www.sofastatistics.com/home.php
Statcato	http://www.statcato.org/statcato/index.php
Statistical Lab	http://www.statistiklabor.de/en/

* Our top picks.

REFERENCES

- Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (2nd ed.). Los Angeles, CA: Sage.
- Hogan, T. P. (2010). *Bare-bones R: A brief introductory guide*. Los Angeles, CA: Sage.
- Hornik, K. (2011). *The R FAQ*. Retrieved from <http://CRAN.R-project.org/doc/FAQ/>
- Muenchen, R. A. (2009). *R for SAS and SPSS users*. New York: Springer.
- Muenchen, R. A. (2010a). *The popularity of data analysis software*. Retrieved from <http://r4stats.com/popularity>
- Muenchen, R. A. (2010b). *R for SAS and SPSS users: The free version*. Retrieved from <http://sites.google.com/site/r4statistics/books/free-version/RforSASSPSSusers.pdf>