SPSS DEMONSTRATIONS [GSS18SSDS-A]

Demonstration 1: Producing Frequency Distributions

In SPSS, you can review the frequency distribution for a single variable or for several variables at once. The frequency procedure is found in the *Descriptive Statistics* menu under *Analyze*. For this chapter, we will use the GSS18SSDS-A data set.

In the Frequencies dialog box, click on the variable name(s) in the left column and transfer the name(s) to the Variable(s) box. More than one variable can be selected at one time.

For our demonstration, let's select the variable POLVIEWS (respondent's political views). Click on *OK* to process the frequency. Respondents were asked to answer the question by indicating 1 = extremely liberal, 2 = liberal, 3 = slightly liberal, 4 = moderate, 5 = slightly conservative, 6 = conservative, and 7 = extremely conservative.

SPSS will produce two tables in a separate Output window, a small statistics table (not presented here), and a frequency table. Use the Window scroll keys to move up and down the window to find the statistics and frequency tables for POLVIEWS. What level of measurement is this variable? (Refer to Chapter 1 to review definitions.)

In the first table, Statistics, SPSS identifies all the valid and missing responses to this question. Responses are coded missing if no answer was given.

In the frequency table (see Figure 2.12), the variable label is reported. The first column lists the value and value label for each category of POLVIEWS. The next four columns contain important frequency information about the variable. The Frequency column shows the number of respondents who gave a particular response. Thus, we can see that 1,500 respondents are included in the data set, but only 1,437 provided a valid response, with 63 responses missing.

The Percent column calculates what percentage of the whole sample (1,500 cases) each of the responses represents. Thus, 10.7% of the total sample indicated

Figure 2.12

Think of self as liberal or conservative

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Extremely liberal	77	5.1	5.4	5.4
	Liberal	179	11.9	12.5	17.8
	Slightly liberal	161	10.7	11.2	29.0
	Moderate	537	35.8	37.4	66.4
	Slghtly conservative	183	12.2	12.7	79.1
	Conservative	235	15.7	16.4	95.5
	Extrmly conservative	65	4.3	4.5	100.0
	Total	1437	95.8	100.0	
Missing	Don't know	48	3.2		
	No answer	15	1.0		
	Total	63	4.2		
Total		1500	100.0		

slightly liberal political views. In most instances, percentages reported in the third column, Valid Percent, are more useful. This column removes all the cases defined as missing and recalculates percentages based only on the valid responses. Recalculated based only on valid cases (1,437), the percentage of those who answered slightly liberal is 11.2. The last column, Cumulative Percent, calculates cumulative percentages beginning with the first response. We know that 29% of the valid sample reported that they were liberal (extremely liberal, liberal, or slightly liberal).

Demonstration 2: Recoding Variables

Some variables may need to be recoded or reduced into a smaller number of categories or intervals to better present and understand the data. We could, for example, collapse POLVIEWS into a variable with three categories: (1) liberal, (2) moderate, and (3) conservative. To accomplish this, we could use the SPSS commands *Transform–Recode Into Different Variables*.

For more detailed instruction on recoding variables, please refer to the section on Recoding Variables in the SPSS Appendix on the text's study site, which explains how to recode the variable EDUC (respondent's years of education).

After reviewing the SPSS Appendix, recode POLVIEWS into a new variable called RPOLVIEWS. Frequencies for Recoded POLVIEWS, 3 Categories (RPOLVIEWS) should look like Figure 2.13.

Figure 2.13

Recoded POLVIEWS, 3 Categories

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Liberal	417	27.8	29.0	29.0
	Moderate	537	35.8	37.4	66.4
	Conservative	483	32.2	33.6	100.0
	Total	1437	95.8	100.0	
Missing	System	63	4.2		
Total		1500	100.0		

Demonstration 3: Producing a Bar Graph

SPSS greatly simplifies and improves the production of graphics. The program offers a separate choice from the main menu bar, *Graphs*, which lists more than a dozen types of graphs that SPSS can create.

Under the *Graphs* menu, select *Legacy Dialogs* and then *Bar*, which will produce various types of bar graphs. We will use bar graphs to display the distribution of the nominal variable MARITAL (marital status of respondent). After clicking on *Bar*, you will be presented with the initial dialog box, with the Simple bar graphs option already selected.

Almost all graphics procedures in SPSS begin with a dialog box that allows you to choose exactly the type of chart you want to construct. Many graph types can display more than one variable (the Clustered or Stacked choices). We will keep things simple here, so click on *Simple*, then on *Define*. When you do so, the main dialog box for simple bar graphs opens.

The variable MARITAL should be placed in the box labeled *Category Axis*. In the *Bars Represent* box, click on the *% of cases* radio button. This choice changes the default statistic from the number of cases to percentages, which are normally more useful for comparison purposes. Click on *OK* to submit your request. (SPSS automatically excludes missing values. You can change this by clicking on *Options*. Click in the box labeled *Display groups defined by missing values* to turn on this choice. Then, click on *Continue*, then on *OK* to submit your request to SPSS.)

The bar graph for MARITAL is presented in an output window labeled SPSS Viewer. You can see in Figure 2.14 that the bar graph for MARITAL has five bars because the only valid responses to this question are (1) "married," (2) "widowed," (3) "divorced," (4) "separated," and (5) "single, never been married."

SPSS graphs can be edited by selecting *Edit*, then *Edit Content*, *In Separate Window*, which moves the graph to its own window (*Chart Editor*) and displays various editing tools and choices.

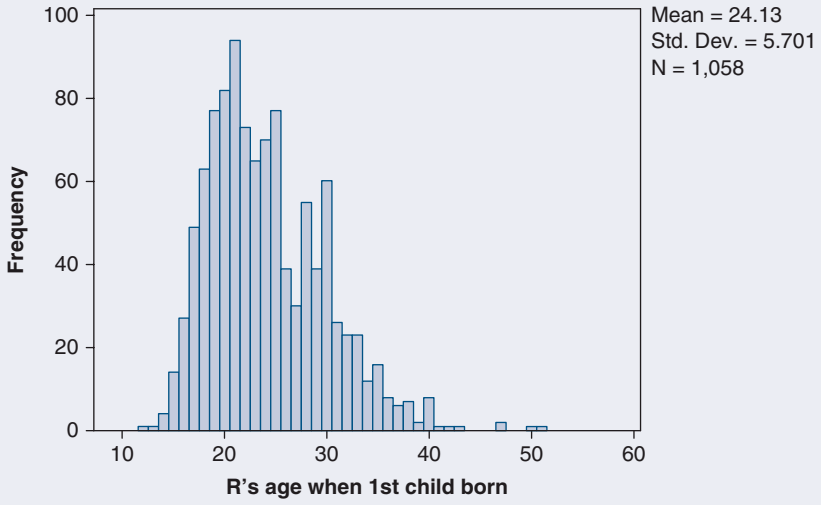
Figure 2.14

Demonstration 4: Producing a Histogram

Histograms are used to display interval or ratio variables. We'll use the variable AGEKDBRN—respondent's age when first child was born. Under the *Graphs–Legacy Dialogs* menu in SPSS, select the *Histogram*.

Histograms are created for one variable at a time (that's why there was no opening dialog box as for bar graphs). You simply insert (drag) the variable you want to display in the first empty box. You don't need to worry about missing values in histograms; unlike the bar graph default, SPSS automatically deletes them from the display. Notice that SPSS includes icons to indicate the level of measurement for each variable. Interval-ratio variables (or scale variables as SPSS refers to them) is matched with a ruler icon. Click on the *OK* button (on the bottom right-hand corner) to process this request. The resulting histogram is shown in Figure 2.15.

Figure 2.15



SPSS automatically decided the appropriate width for each interval based on the range of the variable and the optimal number of bars to be displayed on a screen. The histogram also includes the calculation for mean and standard deviation, which will be discussed in Chapters 3 and 4.

SPSS PROBLEMS [GSS18SSDS-A]

S1. Use the SPSS Frequencies command to produce a frequency table for the variable CONFED (confidence in the executive branch of federal government).

   a. What is the number of valid respondents?

   b. Based on the valid number of respondents, what *percentage* of the sample reported a great deal of confidence in the executive branch of federal government? What *proportion* of the sample reported hardly any confidence?

   c. What is the best way to graphically display these data? Explain.

S2. The GSS included a series of questions on respondents' attitudes about abortion if the woman wants an abortion for any reason (ABANY), is married and doesn't want any more children (ABNOMORE), and is low income and can't afford any more children (ABPOOR).

   a. Run frequencies for all the three variables.

   b. Prepare a general statement summarizing your results from the three frequency tables. Identify the level of measurement for each variable. How would you describe respondents' attitudes about abortion?

S3. Produce the frequency table for HAPMAR (happiness of marriage).

   a. What is the level of measurement for this variable?
   b. Identify two independent variables that may affect how someone responded to HAPMAR. Explain the relationship between these variables and HAPMAR.
   c. What is the best way to graphically display these data? Explain.

S4. The GSS2018 respondents were asked to report their highest year of school (EDUC). Using GSS18SSDS-A, run the frequency table for this variable. Collapse this interval-ratio variable into an ordinal measure (omitting those who did not respond to the question). How many categories do you have?

   a. Prepare a frequency and cumulative percentage table of your recoded EDUC variable.
   b. Prepare a graphic presentation of your recoded EDUC variable.

S5. Determine how best to represent the following variables graphically:

   a. PARTYID (respondent's political party identification)
   b. SEXORNT (sexual orientation)
   c. HEALTH (condition of health)
   d. EMAILHR (e-mail hours per week)

Note: Before selecting/constructing your graph, you may want to review the variable by first using the *Frequencies* or *Utilities—Variables* command. The levels of measurement for several variables might be mislabeled in SPSS. If you are using the *Utilities—Variables* option to review each variable and its level of measurement, you should confirm the level of measurement by reviewing the variable's frequency table (*Analyze—Descriptive—Frequencies*).

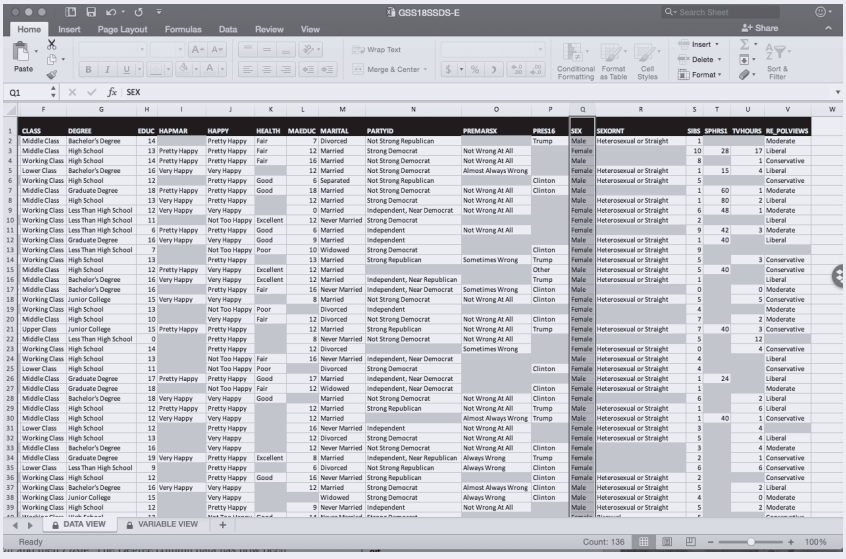
EXCEL DEMONSTRATIONS [GSS18SSDS-E]

Before we begin this chapter's Excel demonstrations, it is important to know that you can paste any tables, graphs, and charts you produce in Excel into a Word document by selecting the object in Excel you've created and then clicking on *Edit* and then *Copy*. You will be able to *Paste* (*Edit* and then *Paste*) your object into a document, e-mail, and so on.

Demonstration 1: Producing a Frequency Distribution

In Excel, you can produce a frequency distribution for one variable at a time. Open the GSS18SSDS-E data set. In the data view sheet of Excel, use your mouse to highlight the column "SEX," including all of the respondents (up to and including row 136). Your screen should look like Figure 2.16.

Figure 2.16



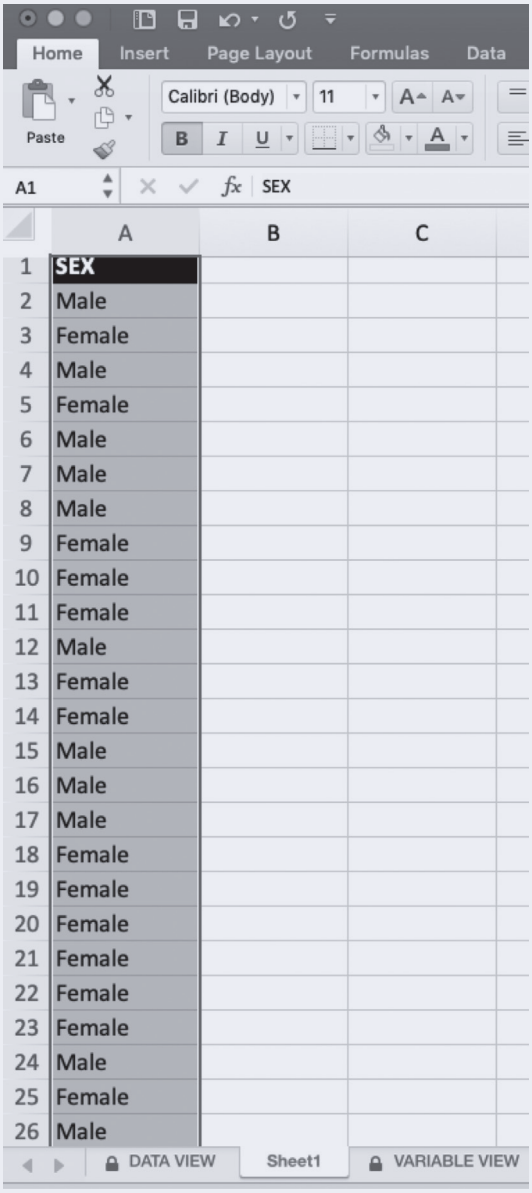
	CLASS	DEGREE	EDUC	HAPMAR	HAPPY	HEALTH	MAEDUC	MARITAL	PARTYID	PREMARSX	PRES16	SEX	SEXORNT	SIBS	SPHRS1	TVHOURS	RE_POLVIEWS
2	Middle Class	Bachelor's Degree	14		Pretty Happy	Fair	7	Divorced	Not Strong Republican		Trump	Male	Heterosexual or Straight	1			Moderate
3	Middle Class	High School	13	Pretty Happy	Pretty Happy	Fair	12	Married	Strong Democrat	Not Wrong At All		Female		10	28	17	Liberal
4	Working Class	High School	14	Pretty Happy	Pretty Happy	Fair	16	Married	Not Strong Democrat	Not Wrong At All		Male		8		1	Conservative
5	Lower Class	Bachelor's Degree	16	Very Happy	Very Happy		12	Married	Not Strong Democrat	Almost Always Wrong		Female	Heterosexual or Straight	1	15	4	Liberal
6	Working Class	High School	12		Pretty Happy	Good	6	Separated	Not Strong Republican		Clinton	Male	Heterosexual or Straight	5			Conservative
7	Middle Class	Graduate Degree	18	Pretty Happy	Pretty Happy	Good	18	Married	Not Strong Democrat	Not Wrong At All	Clinton	Male		1	60	1	Moderate
8	Middle Class	High School	13	Very Happy	Pretty Happy		12	Married	Strong Democrat	Not Wrong At All		Male	Heterosexual or Straight	1	80	2	Liberal
9	Working Class	Less Than High School	12	Very Happy	Very Happy		0	Married	Independent, Near Democrat	Not Wrong At All		Female	Heterosexual or Straight	6	48	1	Moderate
10	Working Class	Less Than High School	11		Not Too Happy	Excellent	12	Never Married	Strong Democrat			Female	Heterosexual or Straight	2			Liberal
11	Working Class	Less Than High School	6	Pretty Happy	Pretty Happy	Good	6	Married	Independent	Not Wrong At All		Female		9	42	3	Moderate
12	Working Class	Graduate Degree	16	Very Happy	Very Happy	Good	9	Married	Independent			Male	Heterosexual or Straight	1	40		Liberal
13	Working Class	Less Than High School	7		Not Too Happy	Poor	10	Widowed	Strong Democrat		Clinton	Female	Heterosexual or Straight	9			
14	Working Class	High School	13		Pretty Happy		13	Married	Strong Republican	Sometimes Wrong	Trump	Female	Heterosexual or Straight	5		3	Conservative
15	Middle Class	High School	12	Pretty Happy	Very Happy	Excellent	12	Married			Other	Male	Heterosexual or Straight	5	40		Conservative
16	Middle Class	Bachelor's Degree	16	Very Happy	Very Happy	Excellent	12	Married	Independent, Near Republican		Trump	Male	Heterosexual or Straight	1			Liberal
17	Middle Class	Bachelor's Degree	16		Pretty Happy	Fair	16	Never Married	Independent, Near Democrat	Sometimes Wrong	Clinton	Male		0		0	Moderate
18	Working Class	Junior College	15	Very Happy	Very Happy		8	Married	Not Strong Democrat	Not Wrong At All	Clinton	Female	Heterosexual or Straight	5		5	Conservative
19	Working Class	High School	13		Not Too Happy	Poor		Divorced	Independent			Female		4			Moderate
20	Middle Class	High School	10		Very Happy	Fair	12	Divorced	Not Strong Democrat	Not Wrong At All	Clinton	Female		7		2	Moderate
21	Upper Class	Junior College	15	Pretty Happy	Pretty Happy		12	Married	Strong Republican	Not Wrong At All	Trump	Female	Heterosexual or Straight	7	40	3	Conservative
22	Middle Class	Less Than High School	0		Pretty Happy		8	Never Married	Not Strong Democrat	Not Wrong At All		Female		5		12	
23	Working Class	High School	14		Pretty Happy		12	Divorced		Sometimes Wrong		Female	Heterosexual or Straight	0		4	Conservative
24	Working Class	High School	13		Not Too Happy	Fair	16	Never Married	Independent, Near Democrat			Male	Heterosexual or Straight	4			Liberal
25	Lower Class	High School	11		Not Too Happy	Poor		Divorced	Strong Democrat		Clinton	Female	Heterosexual or Straight	4			Conservative
26	Middle Class	Graduate Degree	17	Pretty Happy	Pretty Happy	Good	17	Married	Independent, Near Democrat			Male	Heterosexual or Straight	1	24		Liberal
27	Middle Class	Graduate Degree	18		Not Too Happy	Fair	12	Widowed	Independent, Near Democrat		Clinton	Female	Heterosexual or Straight	1			Moderate
28	Middle Class	Bachelor's Degree	18	Very Happy	Very Happy	Good		Married	Not Strong Democrat	Not Wrong At All	Clinton	Female		6		2	Liberal
29	Middle Class	High School	12	Pretty Happy	Pretty Happy		12	Married	Strong Republican	Not Wrong At All	Trump	Male	Heterosexual or Straight	1		6	Liberal
30	Middle Class	High School	12	Very Happy	Very Happy		12	Married		Almost Always Wrong	Trump	Male	Heterosexual or Straight	1	40	1	Conservative
31	Lower Class	High School	12		Pretty Happy		16	Never Married	Independent	Not Wrong At All		Female	Heterosexual or Straight	3		4	
32	Working Class	High School	13		Very Happy		12	Divorced	Strong Democrat	Not Wrong At All		Female	Heterosexual or Straight	5		4	Liberal
33	Middle Class	Bachelor's Degree	16		Very Happy		12	Never Married	Not Strong Democrat	Not Wrong At All	Clinton	Female		3		4	Moderate
34	Middle Class	Graduate Degree	19	Very Happy	Pretty Happy	Excellent	8	Married	Independent, Near Republican	Always Wrong	Trump	Female		2		1	Conservative
35	Lower Class	Less Than High School	9		Pretty Happy		6	Divorced	Not Strong Republican	Always Wrong	Clinton	Female		6		6	Conservative
36	Working Class	High School	12		Pretty Happy	Good	16	Never Married	Strong Republican		Clinton	Female	Heterosexual or Straight	2			Conservative
37	Working Class	Bachelor's Degree	16	Very Happy	Very Happy		12	Married	Strong Democrat	Almost Always Wrong	Clinton	Male	Heterosexual or Straight	5		2	Liberal
38	Working Class	Junior College	15		Very Happy			Widowed	Strong Democrat	Always Wrong	Clinton	Male	Heterosexual or Straight	4		0	Moderate
39	Working Class	High School	12		Pretty Happy		9	Never Married	Independent	Not Wrong At All		Male	Heterosexual or Straight	5		2	Moderate

From the main Excel toolbar, click on *Edit* and then *Copy*. This will copy the SEX column data. On the bottom of the Excel window, next to the sheets *DATA VIEW* and *VARIABLE VIEW*, click on the "+" sign. Doing this will open a blank sheet that's capable of editing. Click on the empty A1 cell, and then select *Edit* and then *Paste*. The SEX column data have been reproduced in your new sheet (see Figure 2.17). You are now ready to produce a frequency distribution and any other graphs or charts you may desire. (*Note:* This copy-and-paste procedure of the data is an extra step because, as explained in the Chapter 1 Excel Demonstration, both the *DATA VIEW* and *VARIABLE VIEW* sheets are in "Protect" mode to avoid any accidental changes to the data set. If you prefer, you can turn off the "Protect" mode and work directly from the *DATA VIEW* sheet. To do so, on the Excel Home Tab, click on *Format* and under *Protection* select *Unprotect Sheet*.)

Click on cell A1 (labeled "SEX"). On the "Insert" Excel tab, select "PivotTable." A Create PivotTable window will open. You will be able to select the range of data you want to analyze. Your copied and pasted data should automatically be selected if you clicked on cell A1 as directed before selecting the "PivotTable" option. Under "Choose where to place the PivotTable," select "Existing worksheet," and then click on an empty cell in this sheet. We will click on cell C2 (see Figure 2.18).

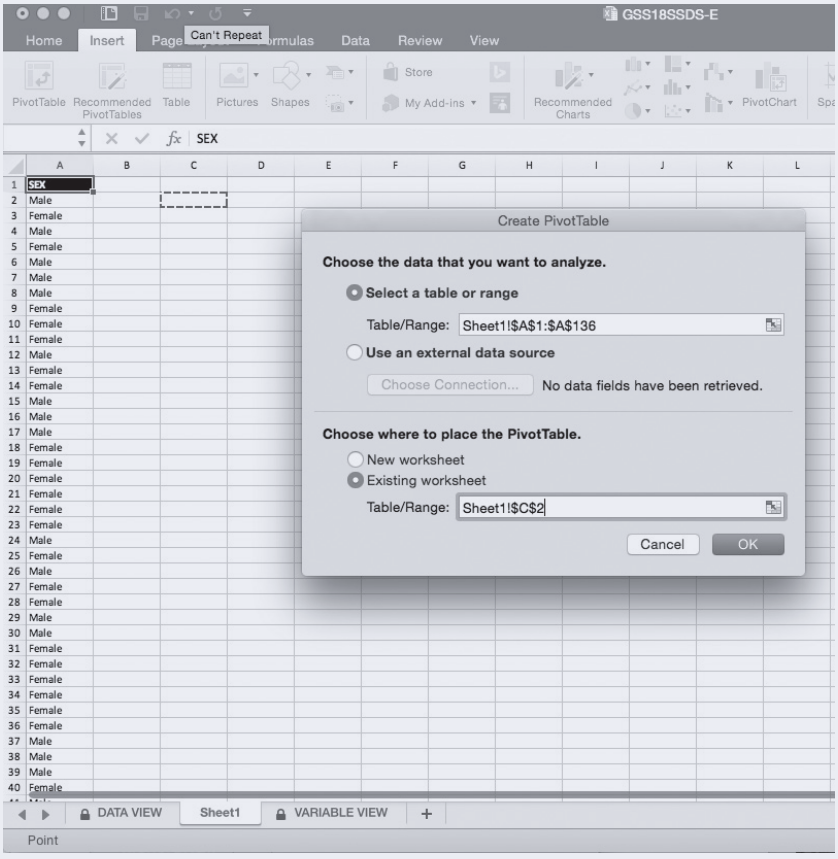
A "PivotTable Fields" window will open on the right side of the Excel sheet. In the "Field Name" box, you should see "SEX." Click on "SEX" and drag it to "Rows." Then, drag "SEX" to the "Values" box. Repeat this step three times. Doing so will easily allow you to produce frequencies, percentages, and cumulative percentages. As you follow these steps, you will see Excel begin to construct a frequency distribution that starts in cell C2 and builds out. Click on "Row

Figure 2.17



Labels" in C2. Rename "Row Labels" to "SEX." Doing this will make our frequency distribution more presentable. Click on D2. A window will open. In the "Field name" box, replace "Count of SEX" with "Frequency." Make sure in the "Summarize by" option that "Count" is highlighted. Select *OK.* Click on E2. In the "Field name" box, replace "Count of SEX2" with "%." Make sure the option

Figure 2.18

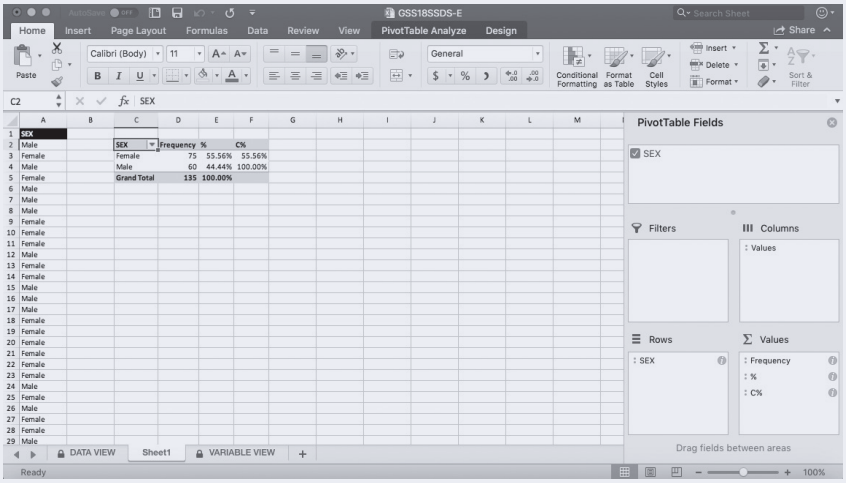


"Count" is highlighted in "Summarize by." In "Show data as," select "% of Column Total" in the dropdown box. Select *OK*. Click on F3. In the "Field name" box, replace "Count of SEX3" with "C%" for cumulative percentage. Make sure the option "Count" is highlighted in "Summarize by." In "Show data as," select "% of Running Total In" in the dropdown box. Select *OK*. Your frequency distribution is ready for interpretation (see Figure 2.19).

The first column of the frequency table contains the categories (Female/Male) of your variable (SEX). The second column contains the frequency for each category. Notice how there are 135 respondents (75 females and 60 males) in this data set. We don't have missing data in this variable, so we can easily interpret the percentage as presented. For example, we could say that our data set is made up of almost 56% females and only 44% males. Given our variable is nominal, it wouldn't be useful to interpret the cumulative percentage distribution we created.

If you are working with an ordinal or interval-ratio variable, interpreting the cumulative percentage would be beneficial. But, you would need to make sure the

Figure 2.19



categories of the variable are placed in order from low to high or high to low. You can easily order categories in any frequency distribution created in Excel by adding a "1," "2," or "3" in front of each category. Then reorder the categories by clicking on the little arrow in the variable (in our example, this is cell C2). A window will appear. Click on *Sort*, then choose either *Ascending* or *Descending*. Your categories should be ordered from either low to high (ascending) or high to low (descending).

Demonstration 2: Producing a Bar Graph

We will now produce a bar graph using the SEX variable we were working with in Demonstration 1. On the "Insert" Excel tab, next to "Recommend Charts," click on the dropdown image of the bar graph. There are many bar graphs to choose from. Let's choose the first one under "2-D Column" (see Figure 2.20).

In the "PivotChart Fields" window on the right, under "Values," select "Frequencies" and drag it to the "SEX" window at the top. This will remove "Frequencies" from the frequency distribution as well as the bar graph we created. Do the same for "C%." You can delete the legend in the bar graph by selecting it and clicking delete on your keyboard. At the top of the bar graph, you should see "Total" as a tentative title. We've clicked on the tentative title and retitled it to read "Respondent's Sex, GSS18SSDS-E" (see Figure 2.21). Excel graphs can be edited by clicking on the bar graph you created and the selecting the Excel "Design" tab. From this menu, you can change the colors of the graph you created and add a chart element (a title, a legend, and more). We suggest you experiment with the many different design options Excel offers its users.

Figure 2.20

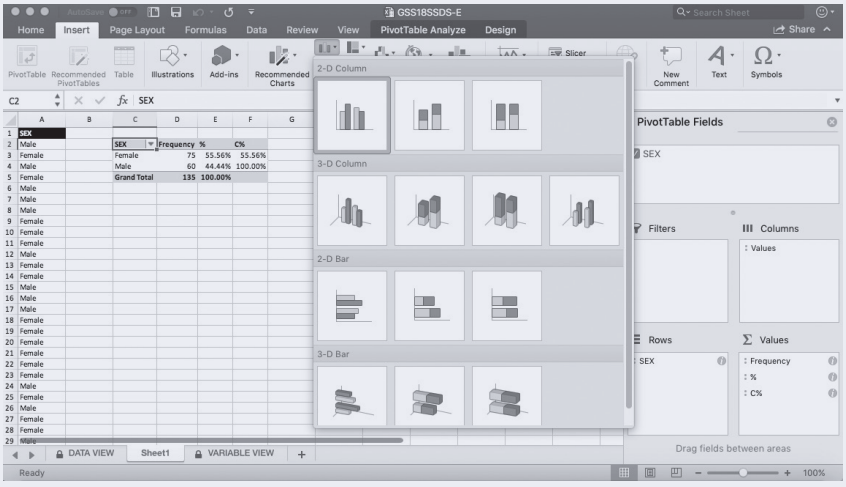
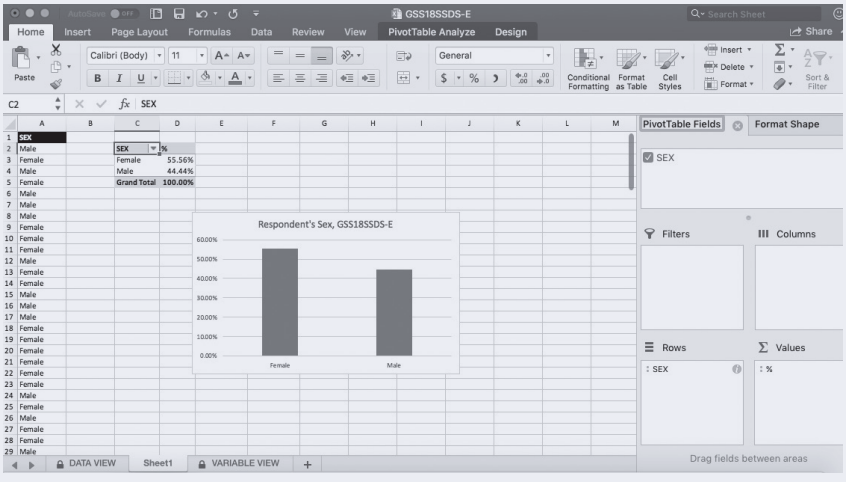


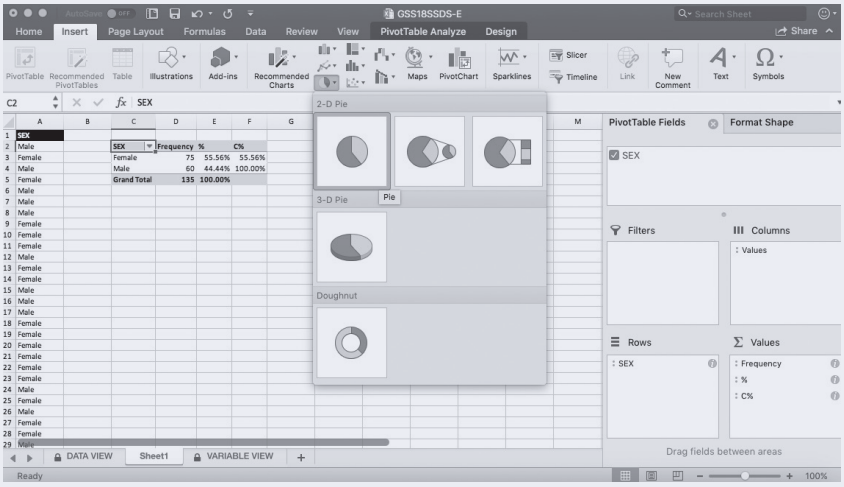
Figure 2.21



Demonstration 3: Producing a Pie Chart

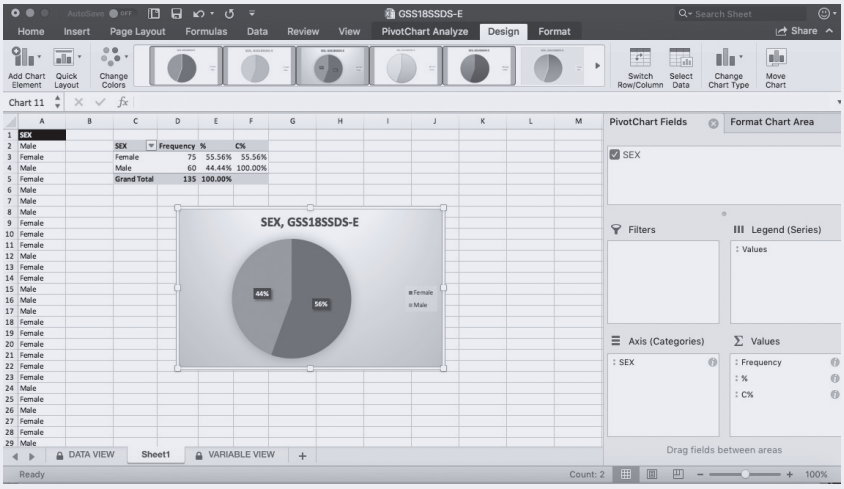
Let's produce a pie chart with Excel, again using our SEX variable. On the "Insert" Excel tab, next to "Recommend Charts," click on the dropdown image of the pie chart. There are many pie charts to choose from. Let's choose the first one under "2-D Pie" (see Figure 2.22).

Figure 2.22



Excel will create a pie chart for the variable SEX. You can retitle the tentative title as we've done to read "SEX, GSS18SSDS-E." You can further edit the pie chart by navigating to the "Design" tab. We've chosen one of the designs Excel offers that includes percentages for the pie slices (see Figure 2.23). As with the bar graph, we suggest you experiment with the many different design options Excel offers its users.

Figure 2.23



Demonstration 4: Producing a Histogram

We will create a histogram in Excel using the variable SPHRS1 (number of hours spouse worked last week). In the data view sheet of Excel, use your mouse to highlight the column "SPHRS1," including all of the respondents (up to and including row 136). From the main Excel toolbar, click on *Edit* and then *Copy*. This will copy the SPHRS1 column data. On the bottom of the Excel window, next to the sheets *DATA VIEW* and *VARIABLE VIEW*, click on the "+" sign. Doing this will open a blank sheet that's capable of editing. Click on the empty A1 cell, and then select *Edit* and then *Paste*. The SPHRS1 column data have been reproduced in your new sheet. You are now ready to ask Excel to create a histogram. On the "Insert" Excel tab, next to "Recommend Charts," click on the dropdown image of the histogram. There are three histograms to choose from. Let's choose the first one under "Histogram" (see Figure 2.24).

Excel will create a histogram for SPHRS1. You can retitle the tentative title as we've done to read "Spouse's Number of Hours Worked, GSS18SSDS-E." You can further edit the histogram by navigating to the "Design" tab (see Figure 2.25).

Figure 2.24

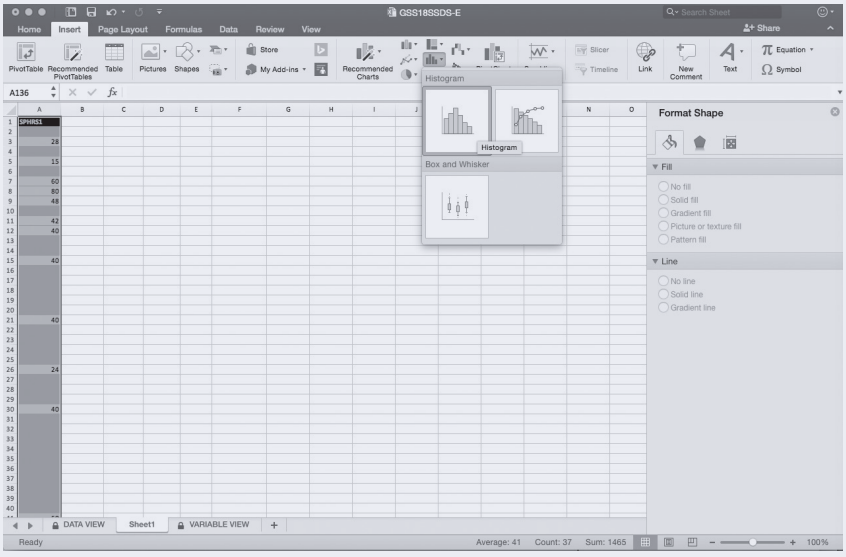
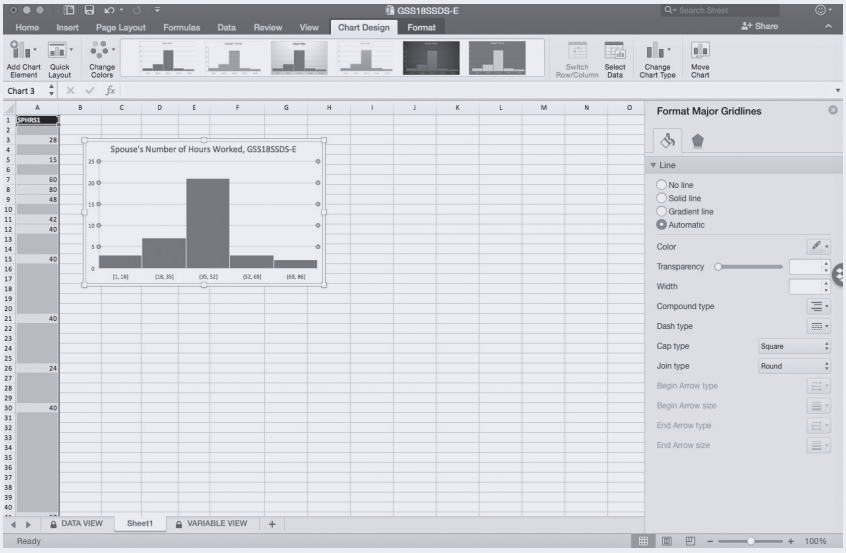


Figure 2.25



EXCEL PROBLEMS [GSS18SSDS-E]

E1. We will use Excel to examine the highest degree our respondents have earned.

a. Produce a frequency distribution for the variable DEGREE. Be sure to order the categories in the frequency distribution from lowest to highest.

b. What percentage of respondents reported their highest degree earned was high school?

c. How many respondents reported their highest degree earned was a bachelor's degree?

d. Considering the level of measurement for DEGREE, have Excel create an appropriate graph or chart. (*Hint*: consider level of measurement.) Be sure to name the variable in the title of the graph/chart, and also identify the data source.

E2. Use Excel to create a histogram of any interval-ratio variable in GSS18SSDS-E. Do not use SPHRS1 since we used that variable in Demonstration 4 to demonstrate how to produce a histogram in Excel.

   a. How many respondents are represented in your chosen variable?

   b. In two to three sentences, what does the histogram you've created show us about your chosen variable?

E3. For the following variables in GSS18SSDS-E, identify whether a bar graph, pie chart, and/or histogram is the most appropriate tool to visually represent each variable.

   a. ABANY (abortion if woman wants for any reason)

   b. HAPMAR (happiness of marriage)

   c. SEXORNT (sexual orientation)

   d. TVHOURS (hours per day watching TV)