Producing a Confidence Interval Around a Mean

SPSS calculates confidence intervals around a sample mean or proportion with the Explore procedure. Activate the Explore procedure by selecting the *Analyze* menu, *Descriptive Statistics*, then *Explore*. The opening dialog box has spaces for both dependent and independent variables.

Place HRSRELAX (after an average work day, about how many hours do you have to relax or pursue the activities you enjoy?) in the Dependent List box and SEX (respondent's sex) in the Factor List box, as shown in Figure 6.7. We expect a gender difference in average relaxation hours with women having less hours than men. In fact, social scientists have documented how women (not men) assume a second shift after work, managing family responsibilities at home and consequently having less time for themselves. We first analyzed this variable in Chapter 4, but here we calculate confidence intervals based on the same data.

Figure 6.7
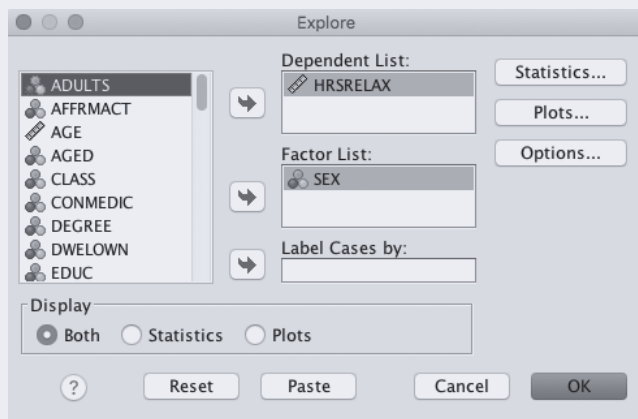
Explore Dialog Box for HRSRELAX and SEX

Click on the *Statistics* button. Note that the Descriptives choice also includes the confidence interval for the mean, which by default is calculated at the 95% confidence level. Let's change that to the 99% level by erasing the "95" and substituting "99."

Click on *Continue* to return to the main dialog box. Recall that Explore produces several statistics and plots by default. For this example, we don't need to view the graphics, so click on the *Statistics* button in the *Display* section. Click on *OK* to run the procedure.

The output from the Explore procedure (Figure 6.8) is divided into two parts, one for males and the other for females. The mean number of HRSRELAX for

Figure 6.8

Descriptives

Respondents sex				Statistic	Std. Error
Hours per day R have to relax	Male	Mean		3.95	.141
		99% Confidence Interval for Mean	Lower Bound	3.59	
			Upper Bound	4.31	
		5% Trimmed Mean		3.68	
		Median		4.00	
		Variance		8.647	
		Std. Deviation		2.941	
		Minimum		0	
		Maximum		24	
		Range		24	
		Interquartile Range		3	
		Skewness		2.396	.117
		Kurtosis		10.203	.233
	Female	Mean		3.52	.130
		99% Confidence Interval for Mean	Lower Bound	3.19	
			Upper Bound	3.86	
		5% Trimmed Mean		3.27	
		Median		3.00	
		Variance		7.680	
		Std. Deviation		2.771	
		Minimum		0	
		Maximum		20	
		Range		20	
		Interquartile Range		3	
		Skewness		2.201	.115
		Kurtosis		8.580	.229

males is 3.95; for females, it's 3.52. Our data indicate that, on average, men have slightly more time to relax in a day.

The 99% confidence interval for males runs from 3.59 to 4.31 hours. One way to interpret this result is to state that, in 100 samples each of size 438 males from the U.S. adult population, we would expect the confidence interval to include the true population value for the mean ideal number of children 99 times out of those 100. We can never be sure that in this particular sample the confidence interval includes the population mean. Still, our best estimate for the average relaxation hours per day falls within a narrow range of 0.72 (4.31 – 3.59). For females, the 99% confidence interval is 3.19 to 3.86 hours, with a range of 0.67 (3.86 – 3.19).

SPSS Output for HRSRELAX by Sex

We'll continue to explore the relationship between gender and relaxation hours in Chapter 8.

SPSS PROBLEMS [GSS18SSDS-B]

S1. Recall that the GSS sample includes men and women from 18 to 89 years of age. Does it matter that we have responses from men and women of diverse ages? Would our HRSRELAX results change if we selected a younger sample of men and women?

   a. To take the SPSS demonstration one step further, use the Select Cases procedure to select respondents based on the variable AGE who are less than or equal to 35 years old. Do this by selecting *Data* and then *Select Cases*. Next, select *If Condition is satisfied* and then click on *If*. Find and highlight the variable AGE in the scroll-down box on the left of your screen. Click the arrow next to the scroll-down box. AGE will now appear in the box on the right.

   Now, tell SPSS that you want to select respondents who are 35 years of age or less. The box on the right should now read AGE <= 35. Click *Continue* and then *OK*.

   b. Using this younger sample, repeat the Explore procedure (with the 99% confidence interval selected in the Statistics box) that we just completed in the demonstration. What differences exist between men and women in this younger sample on hours per day the respondent has to relax? How do these results compare with those based on the entire sample?

S2. Calculate the 90% confidence interval for the following variables, comparing lower, working, middle, and upper classes (CLASS) in the GSS sample. Use the Explore procedure using CLASS as your factor variable (*Analyze*, *Descriptive Statistics*, *Explore*). Make a summary statement of your findings. (Reset Select Cases to *All cases* before completing this exercise.)

   a. EDUC (Respondent's highest year of school completed)
   b. HOMPOP (Number of persons in household)

EXCEL DEMONSTRATION [GSS18SSDS-E]

Producing a Confidence Interval Around a Mean

In this demonstration, we will use Excel to produce a confidence interval around the mean number of hours spent watching television per day (TVHOURS). We will then produce a separate confidence interval around the mean TVHOURS for respondents who are conservative, liberal, and moderate using the recoded political views variable (RE_POLVIEWS) that is included in GSS18SSDS-E.
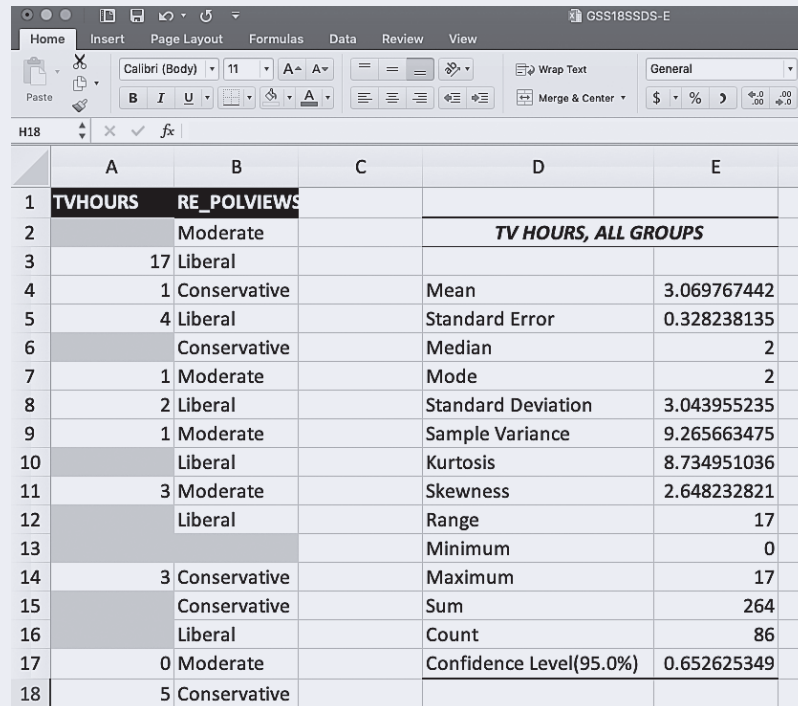
   Copy TVHOURS and RE_POLVIEWS data from the protected Data View sheet and paste it into a new Excel sheet. Next navigate to Excel's Data tab and select *Data Analysis*. A window of Analysis Tools will appear. Select

Descriptive Statistics and then *OK*. Click in the empty box next to "Input Range" and highlight the column of TVHOURS data from A2 to A136. Do not select A1 for it contains the variable name, TVHOURS. Under "Output options," select *Summary statistics* and *Confidence Level for Mean*. Notice how next to Summary Statistics, 95% is listed. This indicates that Excel will produce the 95% confidence level. If you'd like, you can enter any other confidence level here (for example, 90% or 99%). Select *Output Range*. Click in the empty box next to "Output Range," and then select any cell in the current sheet you are working in. This will tell Excel where to place the Descriptive Statistics box it will generate. In our example, we've chosen for the output box to begin in cell D2. Click *OK*.

A Descriptive Statistics box will appear (see Figure 6.9). We will label this box TVHOURS, ALL GROUPS. The 95% confidence level for the mean is .65 (rounded). If we add .65 to the mean TVHOURS (3.07), we will get the upper limit of 3.72. If we subtract .65 from the mean TVHOURS (3.07), we will get the lower limit of 2.42.

One way to interpret our confidence interval is to state that, in 100 samples each of size 135 participants from the U.S. adult population, we would expect the confidence interval to include the true population value for TVHOURS 95 times out of

Figure 6.9

	A	B	C	D	E
1	TVHOURS	RE_POLVIEWS			
2		Moderate		*TV HOURS, ALL GROUPS*	
3	17	Liberal			
4	1	Conservative		Mean	3.069767442
5	4	Liberal		Standard Error	0.328238135
6		Conservative		Median	2
7	1	Moderate		Mode	2
8	2	Liberal		Standard Deviation	3.043955235
9	1	Moderate		Sample Variance	9.265663475
10		Liberal		Kurtosis	8.734951036
11	3	Moderate		Skewness	2.648232821
12		Liberal		Range	17
13				Minimum	0
14	3	Conservative		Maximum	17
15		Conservative		Sum	264
16		Liberal		Count	86
17	0	Moderate		Confidence Level(95.0%)	0.652625349
18	5	Conservative			

those 100. We can never be sure that in this particular sample, the confidence interval includes the population mean. Still, our best estimate for the average number of hours spent watching television per day falls within a narrow range of 0.65 (3.07–2.42).

We will now use Excel to produce a separate confidence interval around the mean TVHOURS for respondents who are conservative, liberal, and moderate using the recoded political views variable (RE_POLVIEWS). This process begins by sorting the RE_POLVIEWS. Select all of the data in the TVHOURS and RE_POLVIEWS columns (include the first row, which is the variable label for each column). Now navigate to Excel's Data tab and select *Sort*. A Sort window will appear (see Figure 6.10). Under "Column," we will instruct Excel to sort our data by RE_POLVIEWS. Under "Sort On," we will instruct Excel to sort by "Values." And, under "Order," we will instruct Excel to sort from A to Z. Select *OK*. The data will be sorted by RE_POLVIEWS. This will make it very easy for us to produce a confidence interval around the mean TVHOURS for those who are conservative, moderate, and liberal.

We are ready once again to navigate to Excel's Data tab and select *Data Analysis*. A window of Analysis Tools will appear. Select *Descriptive Statistics* and then *OK*. Click in the empty box next to "Input Range" and highlight

Figure 6.10
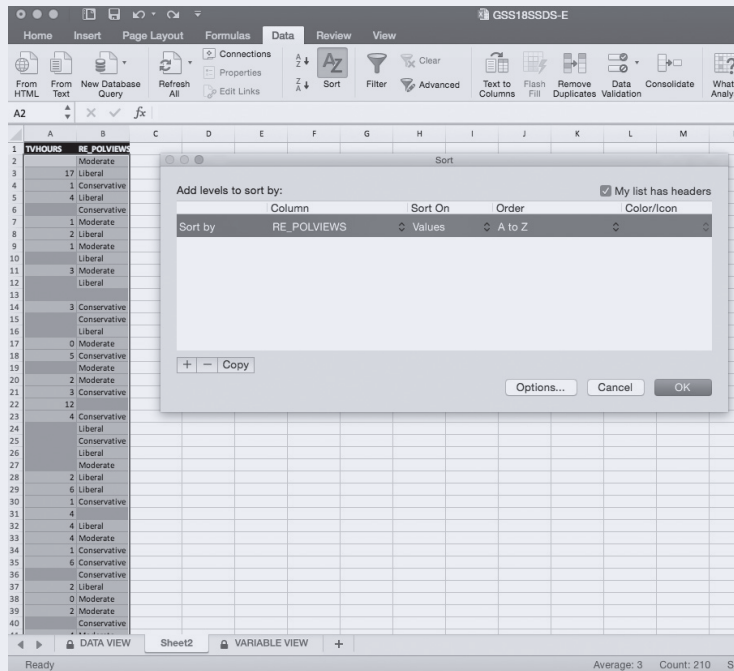
the column of TVHOURS data from A2 to A38. As we can see by looking at the RE_POLVIEWS column, A2 to A38 is TVHOURS data for our conservative respondents. We will not select A1 for it contains the variable name, TVHOURS. Under "Output options," select *Summary statistics* and *Confidence Level for Mean*. The 95% confidence level is the default option. Click in the empty box next to "Output Range," and then select any cell in the current sheet you are working in. This will tell Excel where to place the Descriptive Statistics box it will generate. In our example, we've chosen for the output box to begin in cell G2. See Figure 6.11. Click *OK*.

A Descriptive Statistics box will appear. We will label this box TVHOURS, Conservatives. The 95% confidence level for the mean TVHOURS for conservatives is 1.21 (rounded). If we add 1.21 to their mean TVHOURS (2.75), we will get the upper limit of 3.96. If we subtract 1.21 from the mean TVHOURS (2.75), we will get the lower limit of 1.54. One way to interpret this confidence interval is to state that, in 100 samples each of size 37 conservative participants from the U.S. adult population, we would expect the confidence interval to include the true

Figure 6.11

population value for TVHOURS 95 times out of those 100. Our best estimate for the average number of hours conservatives spend watching television per day falls within a range of 2.42 (3.96 – 1.54).

You can repeat the above steps (of course, you will need to modify the Output Range values) to produce a confidence interval around the mean TVHOURS for liberals and moderates. Figure 6.12 displays four Descriptive Statistics tables for our four groups: all respondents, conservatives, liberals, and moderates.

Figure 6.12

EXCEL PROBLEMS [GSS18SSDS-E]

E1. Use Excel to produce a 95% confidence level around the mean number of siblings (SIBS).

    a. What is the mean for SIBS?

    b. What is the 95% confidence level value?

    c. What is the lower limit?

    d. What is the upper limit?

    e. How would you interpret the confidence interval for SIBS?

E2. Repeat the Excel demonstration for TVHOURS and RE_POLVIEWS, but instead of working with a 95% confidence level, use a 90% confidence level.

    a. What is the 90% confidence interval of TVHOURS for conservatives?

    b. What is the 90% confidence interval of TVHOURS for liberals?

    c. What is the 90% confidence interval of TVHOURS for moderates?