of Democratic U.S. House members and U.S. Senators (circle one)

increases. decreases.

D. Which two of the following statements describe the relationship between the percentage of unionized workers and the percentage of state legislators who are Democrats? (check two)

- [x] The relationship is negative.
- [ ] The relationship is positive.
- [ ] The relationship is stronger than the relationship between the percentage of unionized workers and the percentage of Democratic U.S. House representatives and U.S. senators.
- [ ] The relationship is weaker than the relationship between the percentage of unionized workers and the percentage of Democratic U.S. House representatives and U.S. senators.

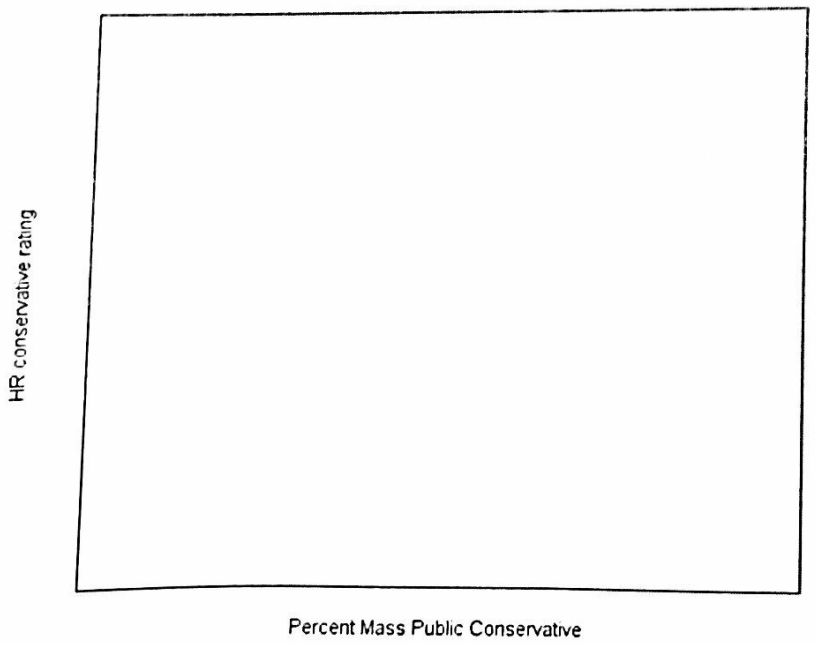
2. (Dataset: States. Variables: HR_conserv11, Conserv_public.) Two congressional scholars are discussing extent to which members of the U.S. House of Representatives stay in touch with the voters in their stat

Scholar 1: "When members of congress vote on important public policies, they are closely attuned to the ideological make-ups of their states. Members from states having many liberals will tend to cast votes in the liberal direction. Representatives from states with mostly conservative constituencies, by contrast, will take conservative positions on important policies."

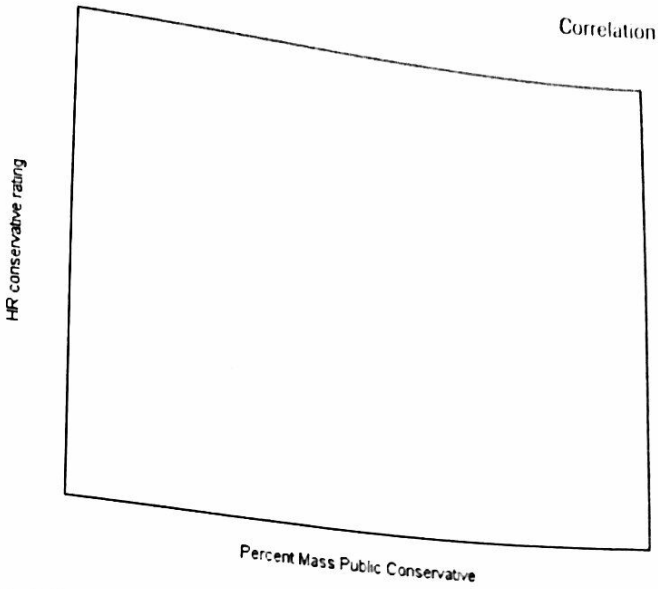
Scholar 2: "You certainly have a naïve view of congressional behavior. Once they get elected, members of congress adopt a 'Washington, D.C. state of mind,' perhaps voting in the liberal direction on one policy and the conservative direction on another. One thing is certain: The way members vote has little to do with the ideological composition of their states."

Think about an independent variable that measures the percentage of self-described "conservatives" amo the mass public in a state, with low values denoting low percentages of conservatives and high values denoti high percentages of conservatives. And consider a dependent variable that gauges the degree to which the state's House delegation votes in a conservative direction on public policies. Low scores on this dependent variable tell you that the delegation tends to vote in a liberal direction and high scores say that the delegatio votes in a conservative direction.

A. Below is an empty graphic shell showing the relationship between the independent variable and the dependent variable. Draw a regression line inside the shell that depicts what the relationship should look like if Scholar 1 is correct.



B. Below is another graphic shell showing the relationship between the independent variable and the dependent variable. Draw a regression line inside the shell that depicts what the relationship should look like if Scholar 2 is correct.



C. The States dataset contains the variable Conserv_public, the percentage of the mass public calling themselves conservative. This is the independent variable. The dataset also contains HR_conserv11, a measure of conservative votes by states' House members. Scores on this variable can range from 0 (low conservatism) to 100 (high conservatism). This is the dependent variable. Perform a regression analysis.

According to the regression equation, a 1-percentage-point increase in conservatives in the mass public is associated with (check one)

- ❑ about a 27-point decrease in House conservatism scores.
- ❑ about a 2-point increase in House conservatism scores.
- ❑ about a 8-point increase in House conservatism scores.

D. If you were to use this regression to estimate the mean House conservatism score for states having 30 percent conservatives, your estimate would be (circle one)

a score of about 35. a score of about 45. a score of about 55.

E. The adjusted *R*-squared for this relationship is equal to _____. This tells you that about _____ percent of the variation in HR_conserv11 is explained by Conserv_public.

F. Use Graphs → Legacy Dialogs → Scatter/Dot to create a scatterplot of the relationship between Conserv_public (X-axis) and HR_conserv11 (Y-axis). In the Chart Editor, add a linear regression line to the scatterplot. Enhance the graph's data-ink ratio by following the procedures described in this chapter for creating an erased graph. Print the graph.

G. Based on your inspection of the regression results, the scatterplot and linear prediction line, and adjusted *R*-squared, which congressional scholar is more correct? (check one)

- ❑ Scholar 1 is more correct.
- ❑ Scholar 2 is more correct.

H. Explain your answer in G, making specific reference to the statistical and graphic evidence.

_____

_____

_____

_____

_____

_____

3. (Dataset: States. Variables: TO_0812, Obama2012.) An article of faith among Democratic Party strategists (and a source of apprehension among Republican strategists) is that high voter turnouts help Democratic candidates. Why should this be the case? According to the conventional wisdom, Democratic electorates are less likely to vote than are Republican voters. Thus, low turnouts naturally favor Republican candidates. As turnouts push higher, the reasoning goes, a larger number of potential Democratic voters will go to the polls, creating a better opportunity for Democratic candidates. Therefore, as turnouts go up, so should the Democratic percentage of the vote.[8]

   A. Use regression to test this conventional wisdom. The States dataset contains TO_0812, the percentage-point change in presidential election turnout between 2008 and 2012. States in which turnout declined between 2008 and 2012 have negative values on TO_0812, whereas states in which turnout increased have positive values on TO_0812. (For example, Utah's turnout increased by a bit more than 2 percentage points between 2008 and 2012, giving Utah a score of 2.1 on TO_0812. Florida's turnout dropped by 4 points, giving a value of −4 on TO_0812.) TO_0812 is the independent variable. Another variable, Obama2012, the percentage of the vote cast for Democratic candidate Barack Obama, is the dependent variable.

   Based on your results, the regression equation for estimating the percentage voting for Obama is: (put the constant in the first blank)

   _____ + _____ * TO_0812.

   B. The *P*-value for the regression coefficient on TO_0812 is _____, and the adjusted *R*-squared is _____.

   C. Consider your findings in A and B. One may conclude that (check one)

   - ☐ the conventional wisdom is correct
   - ☐ the conventional wisdom is incorrect

   D. Explain your answer in C, making specific reference to the regression results.

4. (Dataset: States. Variables: abortlaw10, ProChoice.) As you are no doubt aware, in its momentous decision in *Roe v. Wade* (1973), the U.S. Supreme Court declared that states may not outlaw abortion. Even so, many state legislatures have enacted restrictions and regulations that, while not banning abortion, make an abortion more difficult to obtain. Other states, however, have few or no restrictions. What factors might explain these differences in abortion laws among the states? We know that the mass public remains divided on this issue. Public opinion in some states is more favorable toward permitting abortion, whereas in other states public opinion is less favorable. Does public opinion guide state policy on this issue?

   The States dataset contains abortlaw10, which measures the number of abortion restrictions a state has enacted into law. Values on abortlaw10 range from 0 (least restrictive) to 10 (most restrictive). This is the dependent variable. The dataset also has the variable ProChoice, the percentage of the mass public that is pro-choice. This is the independent variable.

   A. If you were to use regression analysis to test the idea that public opinion on abortion affects state abortion policy, you would expect to find (check one)

   - ☐ a negative sign on ProChoice's regression coefficient
   - ☐ a positive sign on ProChoice's regression coefficient

B. Using regression, analyze the abortlaw10–ProChoice relationship. According to the results, the regression equation for estimating the number of abortion restrictions is (fill in the blanks)

_____ (constant) _____ (regression coefficient) *ProChoice.

C. The *P*-value for the regression coefficient is _____. The value of adjusted *R*-squared is _____.

D. According to States, about 60 percent of Colorado residents are pro-choice. In Arkansas, by contrast, only about 40 percent of the public holds this view. Based on the regression equation: (fill in the blanks)

- ❑ You would estimate that Colorado would have about _____ abortion restrictions.
- ❑ You would estimate that Arkansas would have about _____ abortion restrictions.

E. Adjusted *R*-squared is equal to _____. This means that (complete the sentence)

_____

_____

F. Use Graphs → Legacy Dialogs → Scatter/Dot to create a scatterplot of the relationship between ProChoice (X-axis) and abortlaw10 (Y-axis). In the Chart Editor, add a linear regression line to the scatterplot. Enhance the graph's data/ink ratio by following the procedures described in this chapter for creating an erased graph. Print the graph.

5. Suppose that a critic, upon examining the variables in States, and viewing your results in Exercise 4, expresses skepticism about the relationship between mass-level abortion attitudes and the number of state-level restrictions on abortion:

"There is a key aspect of state governance that you have not taken into account: the percentage of state legislators who are women (womleg_2015). If you were to examine the correlation coefficients among abortlaw10, ProChoice, and womleg_2015, you will find two things. First, the womleg_2015–abortlaw10 correlation will be negative and pretty strong . . . say, at least –.50. Second, the womleg_2015–ProChoice correlation will be positive and fairly strong—at least +.50. Third, when you perform a multiple regression analysis of abortlaw10, using ProChoice and womleg_2015 as independent variables, you will find that womleg_2015 is statistically significant, while ProChoice will fade to statistical insignificance."

A. Obtain the correlation matrix that will allow you to test the critic's claim. Record the correlations in the following table:

		No. of abortion restrictions (abortlaw10)	Percent mass public pro-choice (ProChoice)	Percent female legislators (womleg_2015)
No. of abortion restrictions (abortlaw10)	Pearson Correlation	1		
Percent mass public pro-choice (ProChoice)	Pearson Correlation	?	1	
Percent female legislators (womleg_2015)	Pearson Correlation	?	?	1

B. Consider the skeptical critic's first claim regarding the relationship between womleg_2015 and abortlaw10. According to the correlation coefficient, this claim is:

❑ Correct because _____

_____

❑ Incorrect because _____

C. Consider the skeptical critic's second claim regarding the relationship between womleg_2015 and ProChoice. According to the correlation coefficient, this claim is:

❑ Correct because _____

❑ Incorrect because _____

D. Run the multiple regression suggested by the critic. Write the correct values next to the question marks in the following table:

	Coefficient	*t*-statistic	*P*-value
Number of restrictions (abortlaw10)	?	?	?
Percent mass public pro-choice (ProChoice)	?	?	?
Percent female legislators (womleg_2015)	?		
Constant			

E. Based on the evidence in part D, is the critic's third claim regarding the multiple regression analysis correct? This claim is

❑ Correct because _____

❑ Incorrect because _____

6. (Dataset: GSS2012. Variables: tolerance, educ, age, polviews.) What factors affect a person's level of tolerance of unpopular groups? Consider three hypotheses:

Hypothesis 1: In a comparison of individuals, older people will be less tolerant than younger people.

Hypothesis 2: In a comparison of individuals, those with higher levels of education will have higher levels of tolerance than those with lower levels of education.

Hypothesis 3: In a comparison of individuals, conservatives will be less tolerant than liberals.

GSS2012 includes the following variables, as described in the following table:

GSS2012 variable	Label	Coding	Status in this exercise
Tolerance	Tolerance	0 (low) to 15 (high)	Dependent variable
Age (age)	R's age (years)	18 to 89	Independent variable
Education (educ)	Highest year of school	0 to 20	Independent variable
Political views (polviews)	Ideological self-placement	1 (extremely liberal) to 7 (extremely conservative)	Independent variable

A. Run Correlate. Focus on the correlations between the dependent variable and each of the independent variables. Write the correlations in the following table:

	Tolerance
Age (age)	?
Education (educ)	?
Political views (polviews)	?

B. Based on the *direction* of each correlation coefficient, does it appear that each hypothesis has merit? Answer yes or no and explain:

C. Run multiple regression analysis. Fill in the following table:

Tolerance	Coefficient	*t*-statistic	*P*-value
Age (age)	?	?	?
Education (educ)	?	?	?
Political views (polviews)	?	?	?
Constant	?		
Adjusted *R*-squared	?		

D. Consider whether each hypothesis—Hypothesis 1, Hypothesis 2, and Hypothesis 3—is supported by your analysis. For each hypothesis, check the correct box and explain your answer.

❑ Hypothesis 1 is supported because

❑ Hypothesis 1 is not supported because

❑ Hypothesis 2 is supported because

❑ Hypothesis 2 is not supported because

❑ Hypothesis 3 is supported because

❑ Hypothesis 3 is not supported because

E. Use the regression equation to estimate the tolerance score for the typical respondent, which we will define as a person having the mean values of all the independent variables. Run Descriptives to obtain mean values for each independent variable. Write the means in the table that follows:

	Age (age)	Education (educ)	Political views (polviews)
Mean	?	?	?

F. When you use the mean values to estimate the tolerance score for the typical person, you obtain an estimate equal to (fill in the blank) _____.

That concludes the exercises for this chapter.

NOTES

1. Regression analysis on variables measured by percentages can be confusing. Always stay focused on the exact units of measurement. One percentage point would be 1.00. So if attend_pct increases by 1.00, then womleg_2015 decreases, on average, by .535, or about one-half of a percentage point.
2. The *t*-ratio for the Y-intercept permits you to test the null hypothesis that, in the population, the Y-intercept is 0. In this case, we have no interest in testing the hypothesis that states having 0 frequent attenders have 0 percent women in their state legislatures.
3. Most data analysis programs, SPSS included, provide two values of *R*-square—a plain version, which SPSS labels "R Square," and an adjusted version, "Adjusted R Square." Adjusted *R*-square is often about the same as (but is always less than) plain *R*-square. What is the difference? Just like a sample mean, which provides an estimate of the unseen population mean, a sample *R*-square provides an estimate of the true value of *R*-square in the population. And just like a sample mean, the sample *R*-square is equal to the population *R*-square, give or take random sampling error. However, unlike the random error associated with a sample mean, *R*-square's errors can assume only positive values—squaring any negative error, after all, produces a positive number—introducing upward bias into the estimated value of *R*-square. This problem, which is more troublesome for small samples and for models with many independent variables, can be corrected by adjusting plain *R*-square "downward." For a sample of size *N* and a regression model with *k* predictors, adjusted *R*-square is equal to: 1 – (1 – *R*-square)[(*N* – 1)/(*N* – *k* – 1)]. See Barbara G. Tabachnick and Linda S. Fidell, *Using Multivariate Statistics,* 3rd ed. (New York: HarperCollins, 1996), 164–165.
4. Of course, the smallest value of BA_or_more in the actual data is substantially higher than 0. If you do a quick Descriptives run, you will find that the lowest value of BA_or_more is 17.3 percent.
5. Edward R. Tufte, *The Visual Display of Quantitative Information,* 2nd ed. (Cheshire, Conn.: Graphics Press, 2001). Tufte's work has inspired other excellent treatments of visual communication. For example, see Stephen Few, *Show Me the Numbers: Designing Tables and Graphs to Enlighten* (Oakland, Calif.: Analytics Press, 2004); and Howard Wainer, *Graphic Discovery: A Trout in the Milk and Other Visual Adventures* (Princeton: Princeton University Press, 2005).
6. You will want to keep the Properties window open for your entire excursion into the Chart Editor. Each time you select a different part of the graph for editing, SPSS automatically adjusts the Properties window to reflect the editable features of the graphic element you have selected. Naturally, you can open the Properties window upon entering the Chart Editor by clicking the Properties button.
7. With the Chart Editor still open, click File → Save Chart Template. In the Save Chart Template window, click in the All Settings box, which selects all chart features. Now uncheck the box next to Text Content. (You don't want SPSS to apply the same axis titles to all of your scatterplots.) Click Continue. Find a good place to save the template (and concoct a descriptive name for the file), which SPSS saves with the .sgt extension. To apply the template to future editing projects: In Chart Editor, click File → Apply Chart Template, find the .sgt file, and click Open. Experience teaches that SPSS will apply most of the template's features to the new graphic, although some minor editing may still be required.
8. Michael D. Martinez and Jeff Gill, "The Effects of Turnout on Partisan Outcomes in U.S. Presidential Elections 1960–2000," *Journal of Politics* 67, no. 4 (November 2005): 1248–1274. Martinez and Gill find that the Democratic advantage in higher turnouts has declined over time.