

Willis

respondents to rate something on a scale from 0 to 100, with 0 being very cold, negative feelings, 50 representing neutral feelings, and 100 being very warm positive feelings.

Let's look at the correlations among several feeling thermometers to get an idea of how people view different subjects. If people think subjects are similar, they should express similar sentiments about them. Perhaps knowing how someone feels about one subject can help us predict how they might feel about a variety of subjects.

1. Use the Analyze > Correlate > Bivariate procedure to generate a correlation matrix for the following feeling thermometers in the NES dataset:
 fit_Tump_pre, fit_Police, fit_Rep, fit_Obama, and fit_HClinton_pre. Be sure to weight observations using new, so your results are nationally representative. Fill in the missing Pearson correlation coefficients in the following table.

	fit_Tump_pre	fit_Police	fit_Rep	fit_Obama	fit_HClinton_pre
fit_Tump_pre	1.00				
fit_Police	.33	1.00			
fit_Rep	-.72	-.29	1.00		
fit_Obama	-.105	-.44	.81	1.00	
fit_HClinton_pre					1.00

2. According to the correlation coefficient, as feeling thermometer scores for the police increase, feeling thermometer scores for Republicans (circle one)

increase. decrease.

3. According to the correlation coefficient matrix, which of the following variables has the strongest correlation (either negatively or positively) with feeling thermometer scores for Donald Trump? (circle one)

Police Republicans Obama H. Clinton

4. According to the correlation coefficient matrix, the following two feeling

thermometer scores have the strongest correlation (either negatively or positively): Obama and HClinton.

2. (Derasen States, Variables: vep6_turnout, clinton6.) 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.

1. Use the Analyze > Regression > Linear procedure to test this conventional wisdom. The States dataset contains vep6_turnout, the percentage of the state voting-eligible population that turned out to vote in the 2006 presidential election. This is the independent variable. Another variable, clinton6, the percentage of the vote cast for Democratic candidate Hillary Clinton, is the dependent variable.

• Use linear regression analysis to examine the relationship between voter turnout and Hillary Clinton's vote share. Complete the following table.

	turnout	clinton6
Unadjusted	17.12	13.99
Adjusted	17.12	13.99
R-squared		.07
Adjusted R-squared		.05

2. Based on your results, the regression equation for estimating the percentage voting for Hillary Clinton is (fill in the blanks, put the constant _cons in the last blank)

Clinton Voter Percentage = 17.12
Turnout -

Percepsio

option

- The conventional wisdom is correct because the independent variable
- has a positive relationship w/ vote share for Clinton.
- The conventional wisdom is incorrect because

3. (Dataset States: Variables: `abortionor7`, `prochoice_percent`.) 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 public remains divided on this issue. Public opinion in some states is more favorable toward permitting abortion, whereas public opinion is less favorable in other states. Does public opinion guide state policy on this issue?

The `Survey` dataset contains `abortion2007`, which measures the number of abortion restrictions a state has enacted into law. Values on `abortion2007` range from 0 (least restrictive) to 13 (most restrictive). This is the dependent variable. The dataset also has the variable, `prochoice`, percent the percentage of the mass public that is pro-choice (thus opposed to putting restrictions on abortion access). This is the independent variable.

1. 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 probchoice_percent's regression coefficient.

- \square a positive sign on ρ_{choice} means a positive regression coefficient

2. Using SPSS's Analyze ► Regression ► Linear procedure, analyze the relationship between `abortlaw2017` and `prochoice_percent`. Complete the following table:

Parameter	Estimate	Standard Error	t-Statistic	Pr > t
α_1	14.9	1.15	12.9	<.001
α_2	3.45	0.448	7.7	<.001
α_3	0.001	0.001	0.4	0.69
α_4	0.001	0.001	0.4	0.69
α_5	0.001	0.001	0.4	0.69
α_6	0.001	0.001	0.4	0.69
α_7	0.001	0.001	0.4	0.69
α_8	0.001	0.001	0.4	0.69
α_9	0.001	0.001	0.4	0.69
α_{10}	0.001	0.001	0.4	0.69
α_{11}	0.001	0.001	0.4	0.69
α_{12}	0.001	0.001	0.4	0.69
α_{13}	0.001	0.001	0.4	0.69
α_{14}	0.001	0.001	0.4	0.69
α_{15}	0.001	0.001	0.4	0.69
α_{16}	0.001	0.001	0.4	0.69
α_{17}	0.001	0.001	0.4	0.69
α_{18}	0.001	0.001	0.4	0.69
α_{19}	0.001	0.001	0.4	0.69
α_{20}	0.001	0.001	0.4	0.69
α_{21}	0.001	0.001	0.4	0.69
α_{22}	0.001	0.001	0.4	0.69
α_{23}	0.001	0.001	0.4	0.69
α_{24}	0.001	0.001	0.4	0.69
α_{25}	0.001	0.001	0.4	0.69
α_{26}	0.001	0.001	0.4	0.69
α_{27}	0.001	0.001	0.4	0.69
α_{28}	0.001	0.001	0.4	0.69
α_{29}	0.001	0.001	0.4	0.69
α_{30}	0.001	0.001	0.4	0.69
α_{31}	0.001	0.001	0.4	0.69
α_{32}	0.001	0.001	0.4	0.69
α_{33}	0.001	0.001	0.4	0.69
α_{34}	0.001	0.001	0.4	0.69
α_{35}	0.001	0.001	0.4	0.69
α_{36}	0.001	0.001	0.4	0.69
α_{37}	0.001	0.001	0.4	0.69
α_{38}	0.001	0.001	0.4	0.69
α_{39}	0.001	0.001	0.4	0.69
α_{40}	0.001	0.001	0.4	0.69
α_{41}	0.001	0.001	0.4	0.69
α_{42}	0.001	0.001	0.4	0.69
α_{43}	0.001	0.001	0.4	0.69
α_{44}	0.001	0.001	0.4	0.69
α_{45}	0.001	0.001	0.4	0.69
α_{46}	0.001	0.001	0.4	0.69
α_{47}	0.001	0.001	0.4	0.69
α_{48}	0.001	0.001	0.4	0.69
α_{49}	0.001	0.001	0.4	0.69
α_{50}	0.001	0.001	0.4	0.69
α_{51}	0.001	0.001	0.4	0.69
α_{52}	0.001	0.001	0.4	0.69
α_{53}	0.001	0.001	0.4	0.69
α_{54}	0.001	0.001	0.4	0.69
α_{55}	0.001	0.001	0.4	0.69
α_{56}	0.001	0.001	0.4	0.69
α_{57}	0.001	0.001	0.4	0.69
α_{58}	0.001	0.001	0.4	0.69
α_{59}	0.001	0.001	0.4	0.69
α_{60}	0.001	0.001	0.4	0.69
α_{61}	0.001	0.001	0.4	0.69
α_{62}	0.001	0.001	0.4	0.69
α_{63}	0.001	0.001	0.4	0.69
α_{64}	0.001	0.001	0.4	0.69
α_{65}	0.001	0.001	0.4	0.69
α				

3. According to the results, the regression equation for estimating the number of abortion restrictions is (fill in the blanks, put the constant ____cons in the last blank)

Number of Restrictions = 15 Percentage Pro-
choice = 14.9

4. According to the data, 70 percent of Virginia residents are pro-choice. In Tennessee, by contrast, only 40 percent of the public hold this view. Based on the regression equation (fill in the blank).

- You would estimate that Virginia would have 4.7 abortion restrictions.
- You would estimate that Tennessee would have 8.9 abortion restrictions.

3. Adjusted R -squared is equal to .18. This means that

The 186 variation in abortion restrictions
is explained by the % of pro choice
public opinion.

6. Run the Graphs ► Legacy Dialogs ► Scatter. Dot procedure to obtain a

scatterplot with a linear prediction overlay. Make sure the x-axis and y-axis are appropriately labeled. Also, change the color and pattern of the linear prediction line. If you prefer, make other enhancements to the graph's appearance. Print the graph.

4. Dataset States: Variables: `abrtlaw2017`, `prochoice_percent`, `womleg_2017` Suppose that a critic, upon examining the variables in the States dataset and viewing your results in Exercise 3, 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_2017`). If you were to examine the correlation coefficients among `abrtlaw2017`, `prochoice_percent`, and `womleg_2017`, you will find two things. First, the correlation between `abrtlaw2017` and `womleg_2017` will be negative and pretty strong... say, at least -0.50 . Second, the correlation between `prochoice_percent` and `womleg_2017` will be positive and fairly strong—at least $+0.50$. Third, when you perform a multiple regression analysis of `abrtlaw_2017` using `prochoice_percent` and `womleg_2017` as independent variables, you will find that `womleg_2017` is statistically significant, while `prochoice_percent` will fade to statistical insignificance.

1. Use the Analyze > Correlate > Bivariate procedure to obtain a correlation matrix for `abrtlaw_2017`, `prochoice_percent`, and `womleg_2017`. Write the correlation coefficients next to the question marks in the following table:

	1.00		
	-.48	1.00	
	-.54	.25	1.00

2. Consider the skeptical critic's first claim regarding the relationship between `womleg_2017` and `abrtlaw2017`. According to the correlation coefficient, this claim is (circle one and explain your answer):

correct incorrect

5. Consider the skeptical critic's second claim regarding the relationship between `womleg_2017` and `prochoice_percent`. According to the correlation coefficient, this claim is (circle one and explain your answer):

- because
- both `womleg` and `abrtlaw`
- are negatively correlated

correct incorrect

- because
- `prochoice + womleg` are
- positively correlated

4. Use the Analyze > Regression > Linear procedure to estimate the multiple regression model suggested by the critic. Write the correct values next to the question marks in the following table:

	-.11	.04	-2.03
	-.2	.05	-3.45
	17.76		<.001
	50		
	.4		
	.37		

5. Based on the evidence in part D, is the critic's third claim regarding the multiple regression analysis correct? (circle one and explain your answer):

Correct

~~Incorrect~~

- because
- the multiple regression result shows the partial regression coefficient for abortion is still negative & statistically significant

6. Create a bubble plot that depicts the relationship between `abortion2017` (y-axis) and `prochoice_percent` (x-axis), weighted by `womleg_2017`. This is a special graphing procedure demonstrated in a screencast video. Print the graph.
5. (Dataset: GSS, Variables: `tolerance`, `educ`, `age`, `polviews`, `wtss`.) 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.

The GSS dataset includes the following variables, as described in the table below.

Variable	Measure	Scale	Excluded cases
<code>age</code>	Range: 18-99	18 to 99	Ind. excluded: none
<code>educ</code>	Range: 1-17	1 to 17	Ind. excluded: none
<code>polviews</code>	Range: 1-7	1 to 7	Ind. excluded: none

1. Use the **Analyze** > **Regression** > **Linear** procedure to run a multiple regression analysis with the dependent variable and independent variables specified above. (Don't forget to weight observations using `wtss`.) After you run the model, run the script to obtain adjusted R-squared. Fill in the following table.

Variable	Unadjusted R-squared	Adjusted R-squared	Significance
<code>age</code>	-.04	.002	3.04
<code>educ</code>	.17	.02	11.07
<code>polviews</code>	.03	.03	4.11
<code>Constant</code>	2.84	2.8	8.71
<code>Intercept</code>	16.85	16.85	2.1

2. Based on the evidence in part A, does it appear that Hypothesis 1 has merit? (circle one and explain your answer)

☒ Yes ☐ No

- because
- partial regression coefficient on the variable for age is negative & statistically significant
- the older the people, the less tolerant

3. Based on the evidence in part A, does it appear that Hypothesis 2 has merit? (circle one and explain your answer)

☒ Yes ☐ No

- because
- partial regression coefficient for the variable for education is positive & statistically significant
- the more educated, the more tolerant

4. Based on the evidence in part A, does it appear that Hypothesis 3 has merit? (circle one and explain your answer)

Yes

No

• because
• Partial regression coefficient for political ideology variable
• is negative statistically significant.
• The more conservative the less liberal

5. The adjusted R-squared statistic for the multiple regression model you estimated in part A equals 0.1.

6. Use the regression equation to estimate the tolerance score for the typical respondent, which we will define as a person having the median values of all the independent variables. Run Analyze → Descriptive Statistics → Frequencies with the statistics option (for medians) to obtain the median values for each independent variable. Write the medians in the following table (the median of polviews already appears in the table).

Median	tolc	score
47	13	4.15

7. When you use the median values to estimate the tolerance score for the typical person, you obtain an estimate equal to (fill in the blank) 4.15.
6. (Dataset: States; Variables: HR_conserv, Conserv_public) Two congressional scholars are discussing the extent to which members of the U.S. House of Representatives stay in touch with the voters in their states.

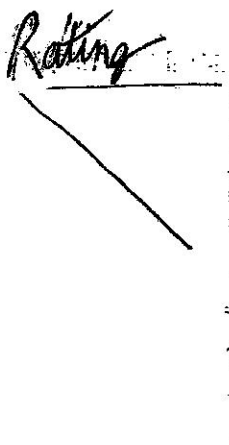
Scholar 1: When members of Congress vote on important public policies, they are closely attuned to the ideological makeup of their states. Members from states having lots of liberals will tend to cast votes in the liberal direction. Representatives from states with mostly conservative constituencies,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

Scholar 2: You certainly have a naive 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 in 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 among the mass public in a state, with low values denoting low percentages of conservatives and high values denoting 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 delegation votes in a conservative direction.

1. 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.



2. 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.

0% mass public, conservative

Rating 60%

010 Miss Public conservative

3. 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. Use the Analyze ▶ Regression ▶ Linear procedure to estimate the relationship between the variables.
 - According to the regression equation, a 1-percentage-point increase in conservatism in the mass public is associated with (check one)
 - ☐ about a 2-point decrease in House conservatism scores.
 - ☒ about a 2-point increase in House conservatism scores.
 - ☐ about an 8-point increase in House conservatism scores.
4. 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 the closest estimate)

30
35
40
45
50

5. The adjusted R^2 squared for this relationship is equal to 58. This tells you that about 58 percent of the variation in `HR_conserv11` is explained by `Conserv_public`.
6. Use the Graphs ▶ Legacy Dialogs ▶ Scatter Dot procedure to obtain a scatterplot

with a linear prediction overlay. Remember that `HR_conserv11` is the y-axis variable, and `Conserv_public` is the x-axis variable. Make sure the y-axis and x-axis are appropriately labeled and change the pattern of the linear prediction line. If you prefer, make other enhancements to the graph's appearance. Print the graph.

Based on your inspection of the regression results, the scatterplot and linear prediction line, and adjusted R^2 -squared, which congressional scholar is more correct?

- ☒ Scholar 1 is more correct because do of public that is conservative is positive
- related to conservatism scores of the delegation
- more conservative the higher the more
- conservative this members are
- ☐ Scholar 2 is more correct because

- That concludes the exercises for this chapter.
- 1 The `democ_fh_polity` variable's high correlations make sense because it is an average of democracy ratings by two other organizations: Freedom House and Polity. We evaluated missing observations using the Analyze ▶ Descriptive Statistics ▶ Frequencies procedure discussed in chapter 1.
 - 2 If you're interested in seeing which countries are rated the most and least violent in the world, refer to the procedure discussed in the section on "Obtaining Case-level Information with Case Summaries" in chapter 1.
 - 3 For this analysis, we are treating World dataset observations as if they were a sample. See the "A Closer Look: Treating Census as a Sample" discussion in chapter 1.
 - 4 Regression analysis on variables measured on constructed indexes like these can be