

## **Unit 7**

### **Relationships in Data**

## Topic 26

### Graphical Displays of Association

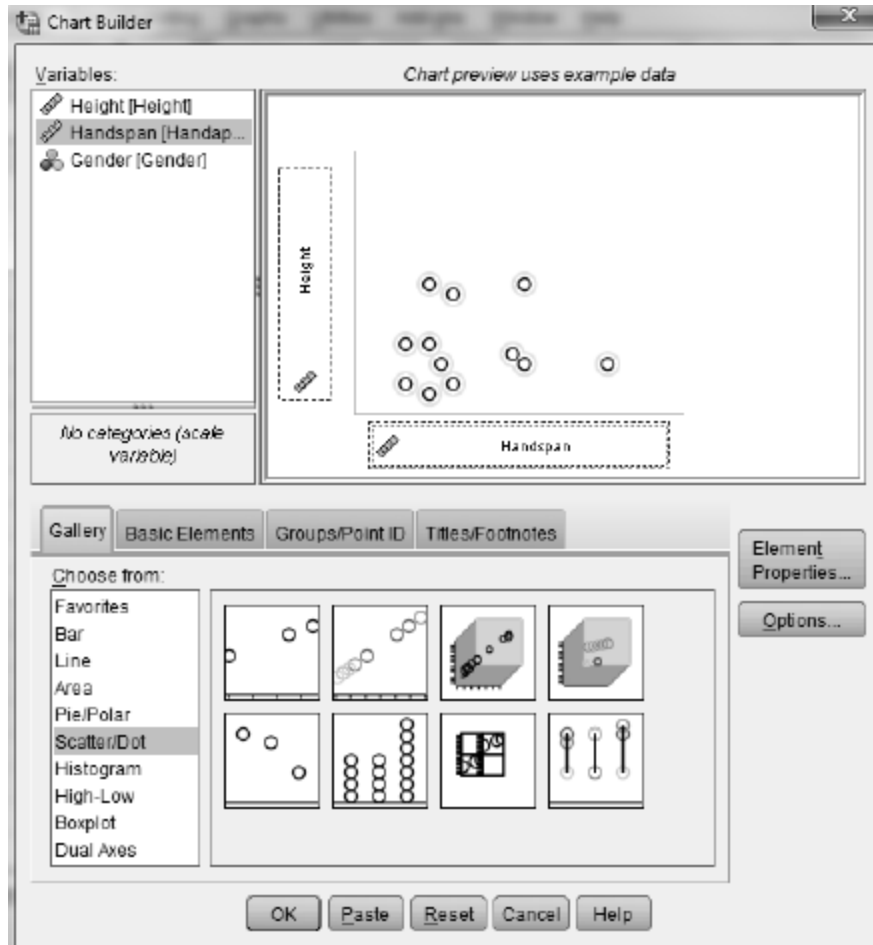
#### In-Class Activities

#### Activity 26-5: Heights, Handspans, and Foot Lengths

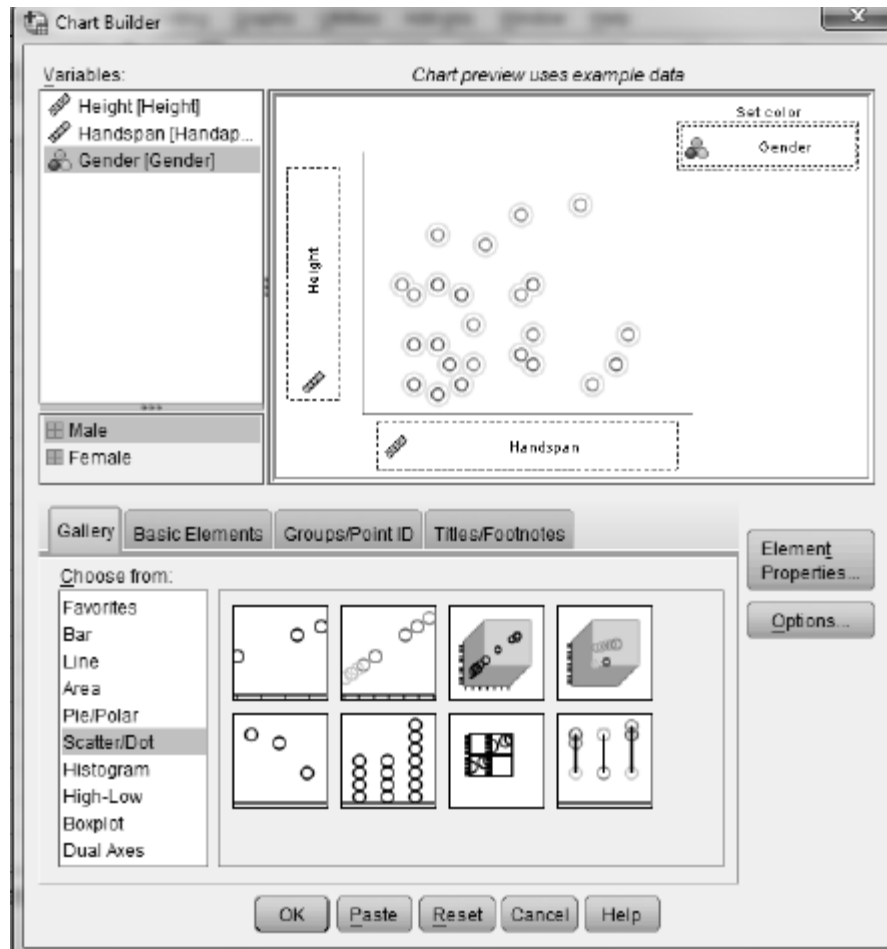
26-5, 28-1, 29-12, 29-13, 29-14

You will use SPSS in parts a, b, and c of this activity.

- a. Use SPSS to create a scatterplot of *height* vs. *handspan*. Do these variables appear to be associated? If so, describe the association in your main textbook?
  1. Enter the data in SPSS. See Activity 2-1 for instructions on entering new data. Create one variable for the *height* data, another for the *handspan* data, and a third for the *gender* data. Be sure that *gender* is a nominal variable, and create value labels. To check or change the data type, use the *Measure* column in the **Variable View** tab.
  2. Select **Graphs > Chart Builder** to bring up the dialog box shown below.
  3. From the **Gallery** select **Scatter/Dot**, and drag the picture of the simple scatterplot (the one in the upper left corner) to the **Chart preview** box.
  4. **Drag the response variable (*Height*) to the **Y-Axis?** box and drag the explanatory variable (*Handspan*) to the **X-Axis?** box**
  5. **Click **OK** to create the plot.**



- b. Use SPSS to create a labeled scatterplot of *height* vs. *handspan*, using different labels for men and women. Use the graph to comment on the three questions posed in your main textbook.
1. Select **Graphs > Chart Builder** to bring up the dialog box shown below.
  2. From the **Gallery** select **Scatter/Dot**, and drag the picture of the scatterplot with different colored circles to the **Chart preview** box.
  3. Drag the response variable (*Height*) to the **Y-Axis?** box and drag the explanatory variable (*Handspan*) to the **X-Axis?** box. **Drag the group-identifying variable (*Gender*) to the Set color box.**
  4. **Click OK** to create the graph.



- c. Use SPSS to create scatterplots and conduct a similar analysis of *height* vs. *foot length*, describing the relationship between them and comparing that relationship between the two sexes.

### Activity 26-6: Televisions and Life Expectancy

#### 26-6, 27-3, 28-19

You will use SPSS to create the requested scatterplot in part b of this activity. Complete all other parts as directed in your main textbook.

- b. Use SPSS to produce a scatterplot of *life expectancy* vs. *televisions per thousand people*. The data are stored in the SPSS file `TVlife06.SAV`. Does there appear to be an association between the two variables? If so, describe its direction, strength, and form. See Activity 26-5 for instructions on creating scatterplots.

## **Activity 26-7: Kentucky Derby Times**

You will use SPSS in parts b and c of this activity. Complete all other parts as directed in your main textbook.

- b. Use SPSS to create the requested histogram. The data are stored in the SPSS file `KYDerby.SAV`.
- c. Use SPSS to create the requested scatterplot.

## **Exercises**

### **Exercise 26-9: Challenger Disaster**

**26-9, 27-17**

Use SPSS to create the scatterplots requested in parts a and d of this exercise. The data are stored in the SPSS file `Challenger.SAV`.

### **Exercise 26-11: Broadway Shows**

**2-14, 26-10, 26-11, 27-15**

Use SPSS to produce the labeled scatterplot requested in part a. The data are stored in the SPSS file `Broadway06.SAV`.

### **Exercise 26-14: Fast Food Sandwiches**

**26-14, 26-15, 28-16**

Use SPSS to create the scatterplot requested in part a and to create the labeled scatterplot requested in part b. The data are stored in the SPSS file `Arbys06.SAV`.

### **Exercise 26-15: Fast Food Sandwiches**

**26-14, 26-15, 28-16**

Use SPSS to produce the requested scatterplots. The data are stored in the SPSS file `Arbys06.SAV`.

**Exercise 26-16: College Alumni Donations**  
**26-16, 27-16**

Use SPSS to analyze the distribution donor percentages requested in part a. Use SPSS to produce the scatterplots requested in parts c and d. The data are stored in the SPSS file `HMCDonors.SAV`.

**Exercise 26-17: Peanut Butter**

Use SPSS to produce the scatterplot requested in part c and to produce the labeled scatterplot request in part d. The data are stored in the SPSS file `PeanutButter.SAV`.

**Exercise 26-21: Comparison Shopping**  
**23-22, 23-23, Lab 6, 26-21**

Use SPSS to create the scatterplot requested in part b. The data are stored in the SPSS file `Shopping.SAV`.

**Exercise 26-24: Word Twist**  
**23-24, 23-25, 26-24, 28-27**

Use SPSS to produce the scatterplot requested in part b. The data are stored in the SPSS file `WordTwist.SAV`.

**Exercise 26-26: Cat Jumping**  
**26-26, 26-27, 28-29, 28-30, 29-20**

Use SPSS to produce the scatterplots requested in part a. The data are stored in the SPSS file `CatJumping.SAV`.

**Exercise 26-27: Cat Jumping**  
**26-26, 26-27, 28-29, 28-30, 29-20**

Use SPSS to produce the labeled scatterplot requested in part a and to produce the comparative dotplots requested in part b. The data are stored in the SPSS file `CatJumping.SAV`.

**Exercise 26-30: Life Expectancy**  
**26-29, 26-30, 28-34, 28-35**

Use SPSS to produce the scatterplots requested in part a. The data are stored in the SPSS file `LifeExpectancy2008.SAV`.

**Exercise 26-32: Scrabble Names**  
**26-32, 29-23**

Use SPSS to produce the scatterplot requested in part a, to create the new variable and dotplot requested in part g, and to produce the scatterplot requested in part h. The data are stored in the SPSS file `ScrabbleNames.SAV`.

**Exercise 26-33: Airline Maintenance**

Enter the data into SPSS and use SPSS to produce the scatterplot requested in part b.

## Topic 27

### Correlation Coefficient

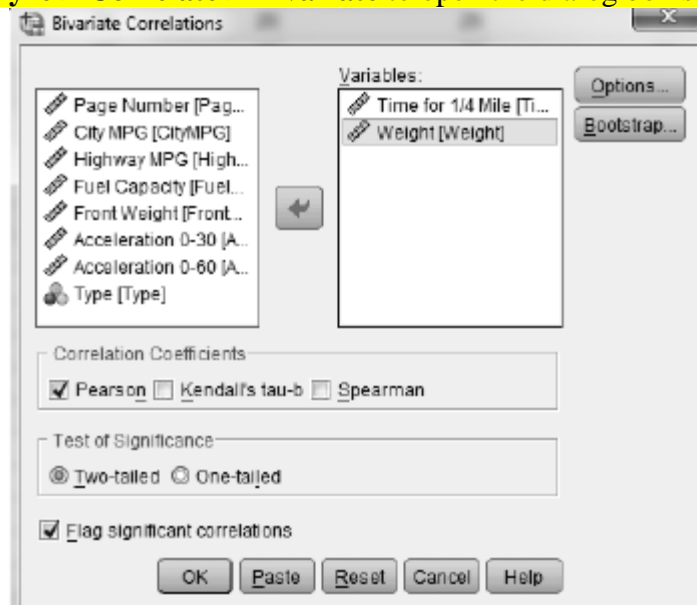
#### In-Class Activities

##### Activity 27-1: Car Data

26-3, 27-1, 28-10

You will use SPSS in parts a and b of this activity. Complete parts c—f as directed in your main textbook.

- a. Use SPSS to calculate the value of the correlation coefficient between *time to travel 1/4 mile* and *weight*. Record this value in the table on page 594 of your main textbook in the column corresponding to scatterplot A.
  1. Open the SPSS data file *Cars99.SAV*.
  2. Select **Analyze > Correlate > Bivariate** to open the dialog box shown below.



3. Enter the two variables whose correlation coefficient you want in the **Variables** box.
4. Make sure that **Pearson** is checked in the **Correlation Coefficients** area
5. Click **OK**. A table like the one shown below will be produced in an Output window. From it you will see that the desired correlation coefficient is  $r = -0.450$ . Record this value in the table on page 594 of your main textbook under the letter A shown there.



Correlations

		Time for 1/4 Mile	Weight
Time for 1/4 Mile	Pearson Correlation	1	-.450*
	Sig. (2-tailed)		.000
	N	73	73
Weight	Pearson Correlation	-.450**	1
	Sig. (2-tailed)	.000	
	N	73	109

\*\* . Correlation is significant at the 0.01 level (2-tailed).

- b. Now use SPSS to calculate the value of the correlation coefficient for the other eight scatterplots. Record these values in the table on page 594 of your main textbook below the appropriate letters.

### Activity 27-2: Governors' Salaries

You will use SPSS in parts b, d, and f—h of this activity. Complete all other parts as directed in your main textbook.

- b. Use SPSS to produce a scatterplot of *governor's salary* vs. *median housing price*. The data are stored in the SPSS file **Governors08.SAV**. Describe the association (direction, strength, and form) between these two variables. See Activity 26-5 for instructions for producing scatterplots.
- d. Use SPSS to calculate the value of this correlation. Record this value in your main textbook and comment on the accuracy of your guess. See Activity 27-1 for instructions on calculating correlations.
- f. Suppose Hawaii gives its governor a \$100,000 raise. Make this change in the data. Then reproduce the scatterplot and recalculate the value of the correlation coefficient. Has the correlation coefficient changed much?
- g. Repeat part f, after giving the governor of Hawaii an additional \$100,000.
- h. Return Hawaii to its original value. Now change California's governor's salary to \$0. (Former governor Schwarzenegger waived his salary as governor.) Then reproduce the scatterplot and recalculate the correlation coefficient. Has the correlation coefficient changed much?

### Activity 27-3: Televisions and Life Expectancy 26-6, 27-3, 28-19

You will use SPSS in part c of this activity.

- c. Use SPSS to calculate this correlation coefficient. The data are stored in the SPSS file **TVlife06.SAV**. How accurate was your guess. See Activity 27-1 for instruction on calculating correlations. Record you value in your main textbook.

### Activity 27-5: Housing Prices

19-22, 26-1, **27-5**, 27-25, 28-2, 28-12, 28-13, 29-3

You will use SPSS in part d of this activity.

- d. Confirm your calculation in part b by using SPSS to calculate the value of the correlation coefficient between *house price* and *size*. The data are stored in the SPSS file `HousePricesAG.SAV`. See Activity 27-1 for instructions on calculating correlations.

### Activity 27-6: Exam Score Improvements

23-19, 23-20, 23-21, **27-6**

You will use SPSS in parts a, b, d, f, h, and j of this activity.

- a. Use SPSS to produce a scatterplot of *exam 2 score* vs. *exam 1 score*. The data are stored in the SPSS file `ExamScores.SAV`. Comment on the direction, strength, and form of the association. See Activity 26-5 for instructions on creating scatterplots.
- b. Use SPSS to calculate the correlation coefficient between *exam 2 score* and *exam 1 score*. See Activity 27-1 for instructions on calculating correlations.
- d. Use SPSS to make this change (subtract 10 points from every score on exam 1). Use **Transform > Compute Variable**. Reproduce the scatterplot of *exam 2 score* vs. *new exam 1 score* and recalculate the correlation coefficient. How did the correlation value change?
- f. Use SPSS to make this change: double everyone's score on exam 2. Reproduce the scatterplot of *new exam 2 score* vs. *exam 1 score* and recalculate the correlation. How did the correlation value change?
- h. Make up some hypothetical bivariate data with the property described in part g. *Hint*: Choose any values for the exam 1 scores, and then make sure each exam 2 score is 10 points higher. Do this for at least 5 hypothetical students. Then use SPSS to produce a scatterplot and calculate the correlation coefficient. Does this confirm the value you expected in part g or do you need to revise your thinking?
- j. Make up some hypothetical bivariate data with the property described in part i. Then use SPSS to produce a scatterplot and calculate the correlation coefficient. Does this confirm the value you expected in part I or do you need to revise your thinking?

## **Activity 27-7: Draft Lottery**

### **10-18, 27-7, 29-9**

You will use SPSS in parts c—e of this activity.

- c. Use SPSS to produce a scatterplot of *draft number* vs. *sequential data of the birthday*. The data are stored in the SPSS file `DraftLottery.SAV`. Based on the scatterplot, guess the value of the correlation coefficient. Explain the reasoning behind your guess. See Activity 26-5 for instructions on creating scatterplots.
- d. Use SPSS to calculate the value of the correlation coefficient. Does the value surprise you? If so, look back at the scatterplot to see whether, in hindsight, its value makes sense. Summarize what the value of this correlation coefficient reveals about how the draft numbers were distributed across birthdays throughout the year. See Activity 27-1 for instructions on calculating correlations.
- e. Data for 1971 are also stored in the SPSS file `DraftLottery.SAV`. Examine a scatterplot and calculate the correlation coefficient between *draft number* and *sequential date* for that year's lottery. Comment on your findings in your main textbook.

## **Exercises**

### **Exercise 27-8: Hypothetical Exam Scores**

#### **7-12, 8-22, 8-23, 9-22, 10-22, 27-8**

Use SPSS to calculate the correlation coefficients requested in parts c and f. The data are stored in the SPSS file `HypoExams.SAV`.

### **Exercise 27-10: Monopoly**

#### **27-10, 27-11**

Use SPSS to produce the scatterplots and calculate the correlation coefficients requested in parts a, b, and c. The data are stored in the SPSS file `Monopoly.SAV`.

### **Exercise 27-10: Monopoly**

#### **27-10, 27-11**

Use SPSS to calculate the correlation coefficients requested in parts d and f. The data are stored in the SPSS file `Monopoly.SAV`.

**Exercise 27-12: Monthly Temperatures**  
**9-26, 26-12, 27-12**

Enter the data into SPSS and use SPSS to calculate the correlation coefficient requested in part b.

**Exercise 27-15: Broadway Shows**  
**2-14, 26-10, 26-22, 27-15**

Use SPSS to calculate the correlation coefficients requested in part a. The data are stored in the SPSS file `Broadway06.SAV`.

**Exercise 27-16: College Alumni Donations**  
**26-16, 27-16**

Use SPSS to produce the scatterplots requested in part a and to calculate the correlation coefficients requested in parts b and e.

**Exercise 27-17: Challenger**  
**26-9, 27-17**

Use SPSS to calculate the correlation coefficients that are requested. The data are stored in the SPSS file `Challenger.SAV`.

**Exercise 27-18: Solitaire**  
**11-22, 11-23, 13-12, 15-3, 21-20, 27-18**

Use SPSS to produce the requested scatterplots and calculate the requested correlation coefficients. The data are stored in the SPSS file `Solitaire.SAV`.

**Exercise 27-19: Climatic Conditions**

Use SPSS to calculate the requested correlation coefficients. The data are stored in the SPSS file `Climate.SAV`.

**Exercise 27-20: Muscle Fatigue**  
**23-2, 23-27, 23-28, 26-22, 27-20**

Enter the data into SPSS and use SPSS to calculate the correlation coefficient in part a.

**Exercise 27-21: Digital Cameras**  
**10-5, 26-18, 27-21**

Use SPSS to calculate the correlation coefficients requested in parts a and b. The data are stored in the SPSS file `DigitalCameras.SAV`.

**Exercise 27-24: Utility Usage**  
**26-28, 27-24, 28-28, 29-21**

Use SPSS to calculate the correlation coefficients requested in part b. The data are stored in the SPSS file `Utilities.SAV`.

**Exercise 27-25: House Prices**  
**19-22, 26-1, 27-5, 27-25, 28-2, 28-12, 28-13, 29-3**

Use SPSS to calculate the correlation coefficients requested in parts a and b. The data are stored in the SPSS file `HousePricesAG.SAV`.

**Exercise 27-26: Kentucky Derby**  
**26-7, 27-26, 27-27, 27-28, 28-32, 28-33**

Use SPSS to calculate the correlation coefficient requested in part a. Use SPSS to create the new variable *Winning Speed* and to produce the scatterplot and calculate the correlation between it and *year* as requested in part c. The data are stored in the SPSS file `KYDerby.SAV`.

**Exercise 27-27: Kentucky Derby**  
**26-7, 27-26, 27-27, 27-28, 28-32, 28-33**

Use SPSS to create the new variable *Winning Speed* and to produce the histogram requested in part a. Use SPSS to produce the scatterplot and calculate the correlation coefficient requested in part b. The data are stored in the SPSS file `KYDerby.SAV`.

**Exercise 27-28: Kentucky Derby**  
**26-7, 27-26, 27-27, 27-28, 28-32, 28-33**

Use SPSS to calculate the correlation coefficients requested in parts a and b. The data are stored in the SPSS file `KYDerby.SAV`.

## Topic 28

### Least Squares Regression

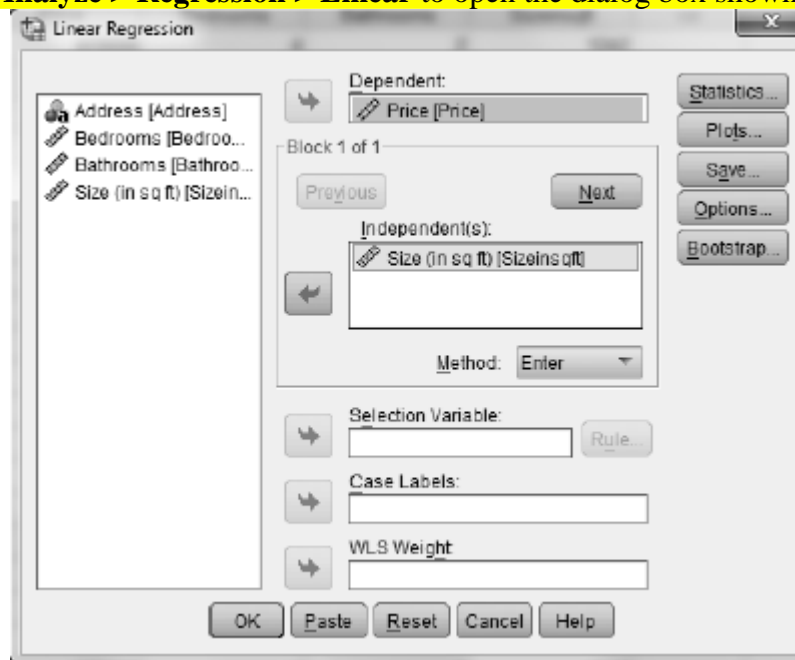
#### In-Class Activities

##### Activity 28-2: House Prices

19-22, 26-1, 27-5, 27-25, **28-2**, 28-12, 28-13, 29-3

You will use SPSS in part c of this activity. Complete all other parts as directed in your main textbook.

- c. Use SPSS to confirm your calculation of the least squares line. *Hint: Do not be surprised if the coefficients differ a bit based on rounding discrepancies.*
  1. Open the SPSS data file `HousePricesAG.SAV`.
  2. Select **Analyze > Regression > Linear** to open the dialog box shown below.



3. Enter the explanatory variable (*Size*) in the **Independent(s)** box, and enter the response variable (*Price*) in the **Dependent** box.
4. Click **OK**.

The output contains, among other things, The *Coefficients* and *Model Summary* tables shown below. The value in the *B* column of the *Coefficients* table in the (*Constant*) row is the value of the intercept, *a*. The value in the *B* column in the row with the

explanatory variable is the value of the slope,  $b$ . The value in the *R Square* column of the *Model Summary* table is the coefficient of determination,  $r^2$ .

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.780 <sup>a</sup>	.608	.587	51309.577

a. Predictors: (Constant), Size (in sq ft)

**Coefficients<sup>a</sup>**

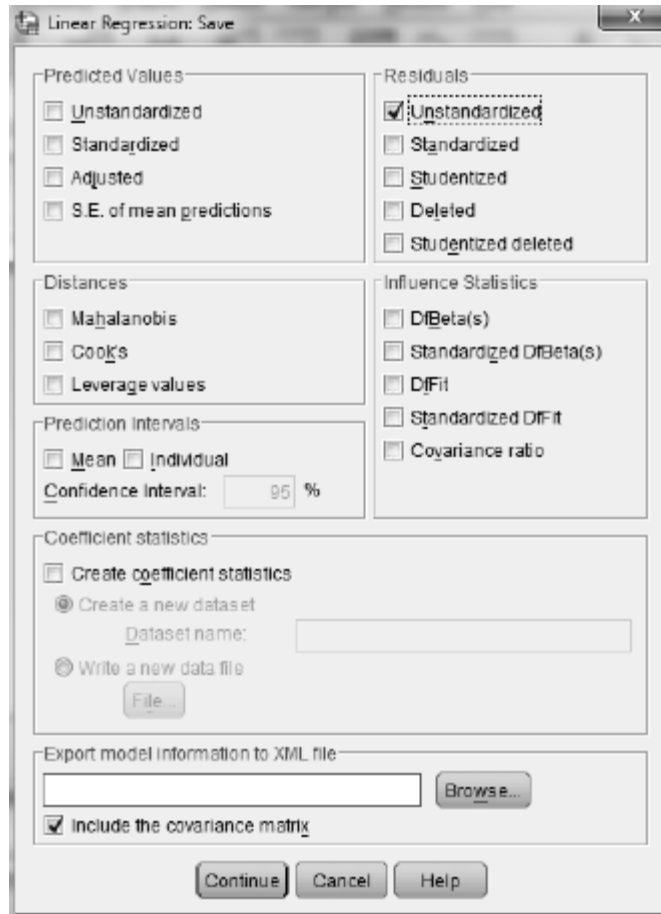
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	265222.120	42642.017		6.220	.000
	Size (in sq ft)	168.592	31.884	.780	5.288	.000

a. Dependent Variable: Price

### Activity 28-3: Animal Trotting Speeds

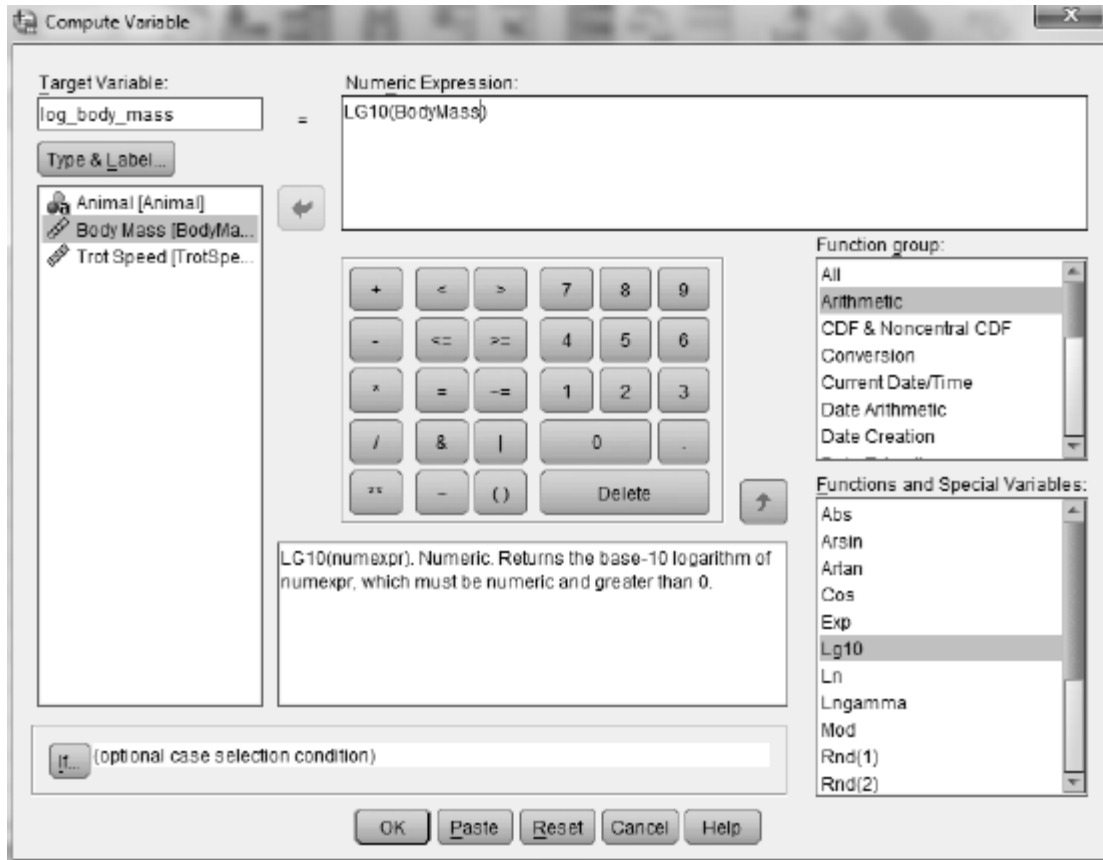
You will use SPSS in parts d—g of this activity.

- d. Use SPSS to produce a residual plot of *residuals* vs. *body mass* for these data.
  1. Open the SPSS data file `TrotSpeeds.SAV`.
  2. Select **Analyze > Regression > Linear** and make *body mass* the independent variable and *trot speed* the dependent variable. Click **Save** to open the dialog box shown below.
  3. In the **Residuals** area check **Unstandardized**. Click **Continue** followed by **OK**. This will run the regression. A new variable, *Res\_1*, will be created in the Data view that contains the residuals.



4. Produce a scatterplot with the residuals on the vertical axis and *body mass* on the horizontal axis. See Activity 26-5 for instructions on producing scatterplots.
- e. Use SPSS to take the logarithm (base 10) of the *body mass* variable.
  1. Select **Transform > Compute Variable** to open the dialog box shown below.
  2. Make the target variable *log\_body\_mass*.
  3. Select **Arithmetic** from the **Function group** menu. **Lg10** computes the base 10 logarithm. Double click it. *Lg10(?)* will appear in the **Numeric Expression** box. Replace the question mark with the *body mass* variable. The dialog box should look like that shown below.
  4. Click **OK** to create the new variable.

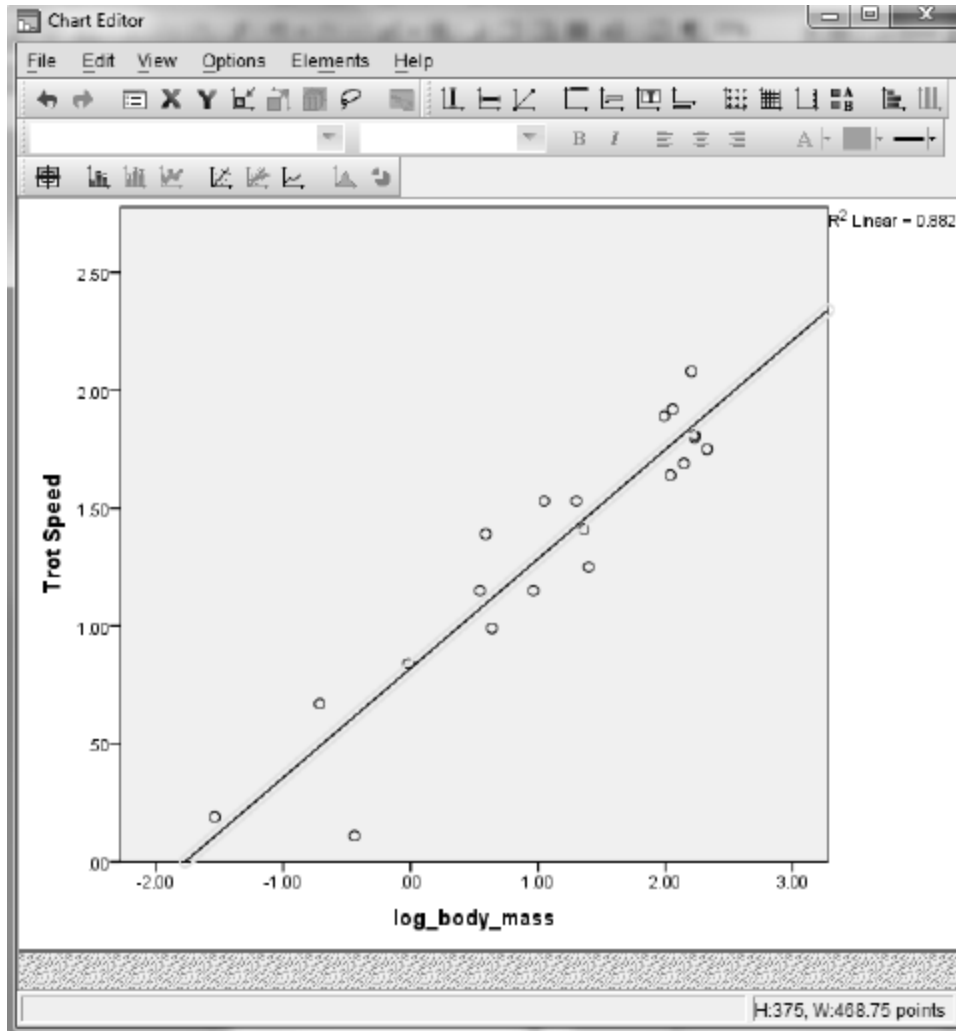




f. Use SPSS to determine the least squares line for predicting *trot speed* from  $\log(\text{body mass})$ . Also produce a scatterplot with this line drawn on it. Record the equation of this line and the value of  $r^2$  in your main textbook. Proceed as directed below to produce this plot. Such a plot is known as a *fitted line plot*. You will need to select

**Analyze > Regression > Linear** to get the equation of the least squares line.

1. Select **Graphs > Chart Builder** and create a scatterplot of *trot speed* vs.  $\log(\text{body mass})$ . See Activity 26-5 for instructions.
2. Double click the resulting graph to open the *Chart Editor* shown below.
3. Click the **Add Fit Line at Total** icon. It is the fifth icon from the left in the row immediately above the graph window.
4. Close the Chart Editor and the scatterplot with the regression line drawn in it will replace the original scatterplot in the Output window.



- g. Use SPSS to produce a residual plot for the least squares line in part f. Does this plot reveal a random scattering of points, indicating that the linear is reasonable for the transformed data?

#### Activity 28-4: Textbook Prices

7-29, 28-4, 28-5, 29-5

You will use SPSS in parts b, d, and h—j of this self-check activity. The data are stored in the SPSS file `TextbookPrices.SAV`.

- b. Use SPSS to examine scatterplots of *price* vs. *pages* and *price* vs. *year*. Comment on what these graphs reveal about the association between *price* and these other variables. See Activity 26-5 for instructions on scatterplots.
- d. Use SPSS to determine the least squares line for predicting a textbook's price based on its number of pages. Report the equation of this regression line, and draw the line on the scatterplot. *Hint:* Be sure to use good statistical notation in report this equation. See Activity 28-2 for instructions.
- h. Use SPSS to create and comment on a residual plot for this least squares line.

- i. Use SPSS to determine the least squares line for predicting a textbook's price based on its year of publication. Report the equation of this regression line and the value of  $r^2$ .
- j. Create a residual plot for this least squares line. Comment on what it reveals about how well the linear model summarizes the relationship between *price* and *year*. See Activity 28-3 for instructions.

## Exercises

### Exercise 28-6: Airfares

**28-6, 28-7, 28-8, 28-9, 29-16**

Use SPSS to create the scatterplot requested in part b. The data are stored in the SPSS file `Airfares.SAV`.

### Exercise 28-7: Airfares

**28-6, 28-7, 28-8, 28-9, 29-16**

Use SPSS to create the residual plot requested in part c. The data are stored in the SPSS file `Airfares.SAV`.

### Exercise 28-8: Airfares

**28-6, 28-7, 28-8, 28-9, 29-16**

Use SPSS to calculate the least squares lines and values of  $r^2$  requested in parts b, d, and e. The data are stored in the SPSS file `Airfares.SAV`.

### Exercise 28-11: Electricity Bills

Use SPSS to create the dotplot requested in part a and produce the scatterplot requested in part b. The data are stored in the SPSS file `ElectricBill.SAV`.

### Exercise 28-12: House Prices

**19-22, 26-1, 27-5, 27-25, 28-2, 28-12, 28-13, 29-3**

Use SPSS to calculate the regression line and the value of  $r^2$  requested in part b. The data are stored in the SPSS file `HousePricesAG.SAV`.

**Exercise 28-13: House Prices****19-22, 26-1, 27-5, 27-25, 28-2, 28-12, 28-13, 29-3**

Use SPSS to create the graphical and numerical summaries requested in part a, produce the scatterplot requested in part b, calculate the regression line and the value of  $r^2$  requested in part c, and produce the residual plot requested in part d. Also use SPSS to perform the analyses requested in part e. The data are stored in the SPSS file `HousePricesAG.SAV`.

**Exercise 28-14: Honda Prices****7-10, 10-19, 12-21, 28-14, 28-15, 29-10, 29-11**

Use SPSS to produce the scatterplot requested in part a, calculate the regression line requested in part b, calculate the value of  $r^2$  requested in part d, and perform the analyses requested in part f. The data are stored in the SPSS file `HondaPrices.SAV`.

**Exercise 28-15: Honda Prices****7-10, 10-19, 12-21, 28-14, 28-15, 29-10, 29-11**

Redo the previous exercise with *manufacture* as the explanatory variable. Use SPSS to produce the scatterplot requested in part a, calculate the regression line requested in part b, calculate the value of  $r^2$  requested in part d, and perform the analyses requested in part f. The data are stored in the SPSS file `HondaPrices.SAV`.

**Exercise 28-16: Fast Food Sandwiches****26-14, 26-15, 28-16**

Use SPSS to produce the scatterplot requested in part a, calculate the equation of the regression line and the value of  $r^2$  requested in part b, and produce the residual plot requested in part c. The data are stored in the SPSS file `Arbys06.SAV`.

**Exercise 28-17: Box Office Blockbusters****28-17, 28-18**

Use SPSS to conduct the requested full regression analysis. The data are stored in the SPSS file `BoxOffice05.SAV`.

**Exercise 28-18: Box Office Blockbusters**  
**28-17, 28-18**

Use SPSS to conduct the requested full regression analysis. The data are stored in the SPSS file `BoxOffice05.SAV`.

**Exercise 28-19: Televisions and Life Expectancy**  
**26-6, 27-3, 28-19**

Use SPSS to produce the scatterplot requested in part a and to compute the transformations and create the scatterplots requested in parts b and c. The data are stored in the SPSS file `TVlife05.SAV`.

**Exercise 28-20: Planetary Measurements**  
**8-12, 10-20, 19-20, 27-13, 28-20, 28-21**

Use SPSS produce the scatterplot requested in part a, calculate the regression equation and produce the residual plot requested in part b, calculate the transformations and produce the scatterplots requested in part c, calculate the regression equation and the value of  $r^2$  requested in part d, and produce the residual plot requested in part f. The data are stored in the SPSS file `Planets.SAV`.

**Exercise 28-21: Planetary Measurements**  
**8-12, 10-20, 19-20, 27-13, 28-20, 28-21**

Use SPSS produce the scatterplot requested in part a, compute the transformation requested in part b, calculate the transformations and produce the scatterplots requested in part c, calculate the regression equation requested in part d, and produce the residual plot requested in part e. The data are stored in the SPSS file `Planets.SAV`.

**Exercise 28-22: Gestation and Longevity**  
**28-22, 28-23**

Use SPSS to produce the scatterplot requested in part a, calculate the regression equation and the value of  $r^2$  requested in part b, produce the residual plot requested in part d, and produce the scatterplots and calculate the regression equations and the values of  $r^2$  and produce the residual plots requested in parts g and h. The data are stored in the SPSS file `Gestation06.SAV`.

**Exercise 28-23: Gestation and Longevity**  
**28-22, 28-23**

Use SPSS to reproduce the regression equation to be used in part a. The data are stored in the SPSS file `Gestation06.SAV`.

**Exercise 28-27: Word Twist**  
**23-24, 23-25, 26-24, 28-27**

Use SPSS to produce the fitted line plot, calculate the regression line, and store the residuals requested in part a. The data are stored in the SPSS file `WordTwist.SAV`.

**Exercise 28-28: Utility Usage**  
**26-28, 27-24, 28-28, 29-21**

Use SPSS to calculate the equation of the regression line and the value of  $r^2$  requested in part a, and produce the residual plot requested in part d. The data are stored in the SPSS file `Utilities.SAV`.

**Exercise 28-29: Cat Jumping**  
**26-26, 28-29, 28-30, 29-20**

Use SPSS to calculate the equation of the regression and the value of  $r^2$  requested in part a, and produce the residual plot requested in part d. The data are stored in the SPSS file `CatJumping.SAV`.

**Exercise 28-30: Cat Jumping**  
**26-26, 28-29, 28-30, 29-20**

Use SPSS to compute the requested new variable and perform the requested full regression analysis. The data are stored in the SPSS file `CatJumping.SAV`.

**Exercise 28-33: Kentucky Derby**  
**26-7, 27-26, 27-27, 27-28, 28-32, 28-33**

Use SPSS to split the file and calculate the equation of the regression lines requested in part a, and produce the residual plots requested in part b. The data are stored in the SPSS file `KYDerby.SAV`.

**Exercise 28-34: Life Expectancy**  
**26-29, 26-30, 28-34, 28-35**

Use SPSS to produce the scatterplots requested in part a, and calculate the equation of the regression line and the value of  $r^2$  requested in part b. The data are stored in the SPSS file `LifeExpectancy2008.SAV`.

**Exercise 28-35: Life Expectancy**  
**26-29, 26-30, 28-34, 28-35**

Use SPSS to perform the requested log transformations and produce the scatterplots and calculate the equation of the regression lines. The data are stored in the SPSS file `LifeExpectancy2008.SAV`.

**Exercise 28-36: Governors' Salaries**  
**27-2, 28-36**

Use SPSS to produce the fitted line plot and calculate the equation of the regression line requested in part a, and calculate the equation of the regression line requested in part c. The data are stored in the SPSS file `Governors08.SAV`.

# Topic 29

## Inference for Correlation and Regression

### In-Class Activities

#### Activity 29-1: Studying and Grades

#### 29-1, 29-2, 29-7, 29-8

You will use SPSS in parts b, c, n, o, p, and r of this activity. Complete all other parts as directed.

- b. Open the SPSS file `UOPgpa.SAV` and use SPSS to construct a scatterplot of *GPA* vs. *study hours*. Describe any association revealed by the scatterplot. Is the association what you expected? Explain. See Activity 26-5 for instructions on generating scatterplots.
- c. Use SPSS to determine the equation of the least squares line for predicting *GPA* from *study hours*, being sure to write it in the context of these two variables. See Activity 28-2 for instructions on least squares.
- n. The table shown below is part of the standard output from SPSS for this regression. You have already seen in Topic 28 how to get the sample slope and intercept. In the *B* column of the *Unstandardized Coefficients* column, you see that  $a = 2.886$  and  $b = .089$ . This table also gives the standard errors for both  $a$  and  $b$ . They are in the *Std. Error* column in the *Unstandardized Coefficients* column. Thus,  $SE(a) = .120$  and  $SE(b) = .028$ . Is this reasonably close to the standard deviation of your 100 simulated slope coefficients?

Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.886	.120		24.034	.000
	Study Hours per Week	.089	.028	.343	3.226	.002

a. Dependent Variable: GPA

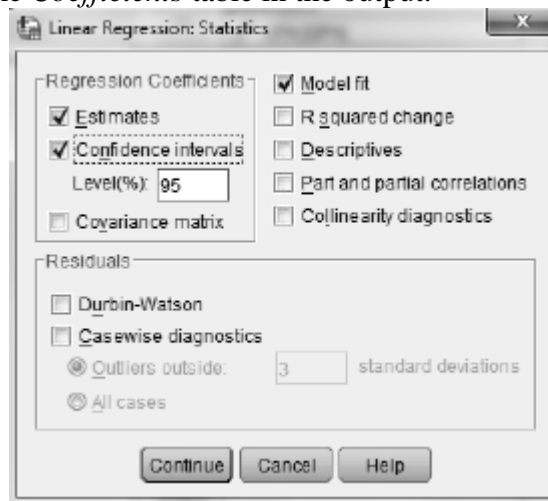
- o. Use the standard error from part n to calculate (by hand) the test statistic for testing whether the population slope differs from zero. Does this value agree with the value shown in the *t* column of the table shown above in the *Study Hours per Week* row?
- p. Use the *t*-Distribution Critical Values table (Table III) and your test statistic in part o to determine (as best you can) the *p*-value of the test. Is this *p*-value consistent with the value shown in the *Sig.* column of the table shown above in the *Study Hours per Week* row? The value in the table is the exact *p*-value for testing  $H_0 : \beta = 0$  vs.  $H_a : \beta \neq 0$ . Is this *p*-value consistent with your simulation results? Explain. If you want to run a one-sided test using SPSS results, you need to adjust the *Sig.* value as shown in the table below.



$H_a : \beta > 0$	If $t > 0$ : $p\text{-value} = \text{Sig.}/2$ If $t \leq 0$ : $p\text{-value} = 1 - \text{Sig.}/2$
$H_a : \beta < 0$	If $t \geq 0$ : $p\text{-value} = 1 - \text{Sig.}/2$ If $t < 0$ : $p\text{-value} = \text{Sig.}/2$

Use the sample slope and its standard error to form a 95% confidence interval for the population slope coefficient. SPSS also calculated confidence intervals for the slope coefficient and intercept.

1. Select **Analyze > Regression > Linear**, set up the regression, and click **Statistics** to bring up the dialog box shown below.
2. Check **Confidence intervals** in the **Regression Coefficient** area. Enter the desired confidence level, as a percent, in the **Level (%)** box.
3. Click **Continue** and then **OK** to run the regression. The confidence intervals will be appended to the *Coefficients* table in the output.



## Activity 29-2: Studying and Grades 29-1, 29-2, 29-7, 29-8

You will use SPSS in parts a and b of this activity.

- a. Open the SPSS file `UOPgpa.SAV` and set up SPSS to run the regression. Click **Plots** to open the dialog box shown below. Check **Histogram** and **Normal probability plot** in the **Standardized Residual Plots** area. Move *\*ZRESID* to the **Y** box and move *\*ZPRED* to the **X** box. Click **Ccontinue** followed by **OK** to produce a residual plot, a histogram of the residuals, and a normal plot of the residuals. Do these plots reveal any marked features suggesting non-normality (condition 3)?



- b. You produced a residual plot in part a. Describe that plot. Does it reveal a strong pattern (such as curvature), suggesting a linear model is not appropriate (condition 2)? Does the residual plot suggest that the variable of the residuals differs substantially at various  $x$ -values (condition 4)?

### Activity 29-3: House Prices

19-22, 26-1, 27-5, 27-25, 28-2, 28-12, 28-13, 29-3

You will use SPSS in parts a, b, and c of this activity.

- Open the SPSS file `HousePricesAG.SAV` and use SPSS to produce the regression output, including 90% confidence intervals, for conducting inference about the population slope using *size* to predict *price*. Report the equation of the sample least squares line, and also report the standard error of the slope coefficient.
- Use SPSS output to conduct a test of whether the sample data provide evidence that *price* is positively associated with *size* in the population. Report all aspects of the test, including an examination of residual plots to check the technical conditions, and summarize your conclusions.
- Report the 90% confidence interval for the population slope coefficient. Also state whether 0 is in this interval, and comment on how this relates to the test in part b.

### Activity 29-5: Textbook Prices

7-29, 28-4, 28-5, 29-5

You will use SPSS in parts a, b, d, e, and g of this self-check activity. Consult your main textbook for instructions on all other parts.

- Open the SPSS file `TextbookPrices.SAV` and reproduce a scatterplot of *price* vs. *pages*, including the least squares line and 90% confidence intervals. Also report the equation of the line and the value of  $r^2$ , along with an interpretation of  $r^2$ .
- Use SPSS output to report the value of the slope coefficient, along with the symbol to represent it and its standard error. Provide an interpretation of both numbers.

- d. Use the SPSS output to conduct a significance test of whether the sample data provide strong evidence (at the  $\alpha = .01$  significance level) that there is a positive relationship in the population between a textbook's price and its number of pages. Report all aspects of the test, and summarize your conclusion.
- e. Report the 90% confidence interval for the slope coefficient and interpret your results.
- g. Comment on whether the technical conditions for these inference procedures appear to be satisfied. Include an analysis of appropriate residual plots.

## Exercises

### Exercise 29-10: Honda Prices

7-10, 10-19, 12-21, 28-14, 28-15, **29-10**, 20-11

Use SPSS to conduct the regression requested in part a, construct the confidence interval requested in part b, produce the residual plots requested in part d, and perform the analyses requested in part e. The data are stored in the SPSS file `HondaPrices.SAV`.

### Exercise 29-11: Honda Prices

7-10, 10-19, 12-21, 28-14, 28-15, **29-10**, **29-11**

Use SPSS to produce the requested analyzes in parts a-d. The data are stored in the SPSS file `HondaPrices.SAV`.

### Exercise 29-12: Heights, Handspans, and Foot Lengths

26-5, 28-1, **29-12**, **29-13**, 29-14

Enter the data into SPSS and use SPSS to produce the scatterplot requested in part a, calculate the regression equation requested in part b, calculate the value of  $r^2$  requested in part c, produce the residual plots requested in part d, conduct the test requested in part f, and construct the confidence interval requested in part g.

### Exercise 29-12: Heights, Handspans, and Foot Lengths

26-5, 28-1, **29-12**, **29-13**, 29-14

Enter the data into SPSS and use SPSS to conduct the requested analyses.

### Exercise 29-12: Heights, Handspans, and Foot Lengths

26-5, 28-1, **29-12**, **29-13**, 29-14

Enter the data into SPSS and use SPSS to conduct the requested analyses.

**Exercise 29-16: Airfares**  
**28-6, 28-7, 28-8, 28-9, 29-16**

Use SPSS to conduct the test requested in part a, construct the confidence interval requested in part b, and produce the residual plots requested in part d. The data are stored in the SPSS file `Airfares.SAV`.

**Exercise 29-17: Marriage Ages**  
**8-17, 9-6, 16-24, 17-22, 23-1, 23-12, 23-13, 26-4, 29-17, 29-18**

Use SPSS to calculate the regression equation requested in part a, conduct the test requested in part b, construct the confidence interval requested in part c, and conduct the test requested in part e. The data are stored in the SPSS file `MarriageAges.SAV`.

**Exercise 29-18: Marriage Ages**  
**8-17, 9-6, 16-24, 17-22, 23-12, 23-13, 26-4, 29-17, 29-18**

Use SPSS to determine the influential point requested in part a, and conduct the analysis requested in part b. The data are stored in the SPSS file `MarriageAges.SAV`.

**Exercise 29-19: Cricket Thermometers**  
**28-26, 29-19**

Use SPSS to conduct the test requested in part a, and construct the confidence interval requested in part b. The data are stored in the SPSS file `Cricketts.SAV`.

**Exercise 29-20: Cat Jumping**  
**26-26, 26-27, 28-29, 28-30, 29-20**

Use SPSS to calculate the requested  $p$ -values. The data are stored in the SPSS file `CatJumping.SAV`.

**Exercise 29-24: Birth Weights**  
**19-28, 19-29, 20-23, 20-24, 20-25, 29-24**

Use SPSS to perform the requested analyses. The data are stored in the SPSS file `NCBirths.SAV`.