



Practice Exercises: Lesson 6.1

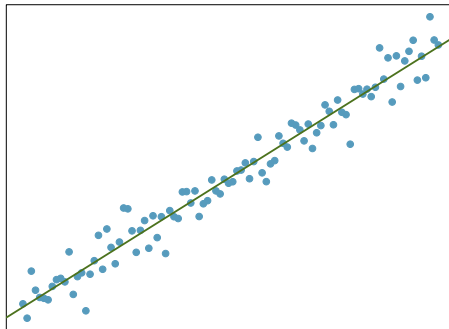
Diez, D. M., Çetinkaya-Rundel, M., Barr, C. D. (2019). OpenIntro Statistics (4th ed.). OpenIntro.
<https://www.openintro.org/book/os/> CC BY-SA 3.0

STAT 1201
Introduction to Probability and Statistics

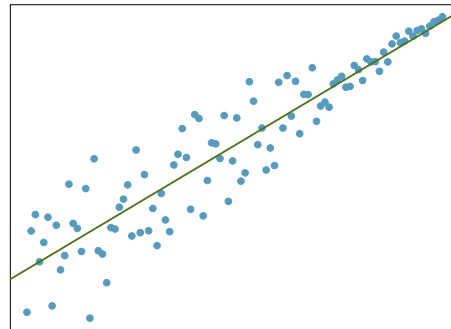
ONLINE AND DISTANCE EDUCATION

Exercises

8.1 Visualize the residuals. The scatterplots shown below each have a superimposed regression line. If we were to construct a residual plot (residuals versus x) for each, describe what those plots would look like.

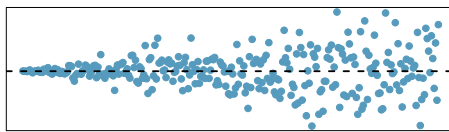


(a)

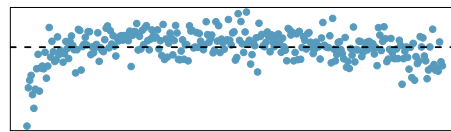


(b)

8.2 Trends in the residuals. Shown below are two plots of residuals remaining after fitting a linear model to two different sets of data. Describe important features and determine if a linear model would be appropriate for these data. Explain your reasoning.

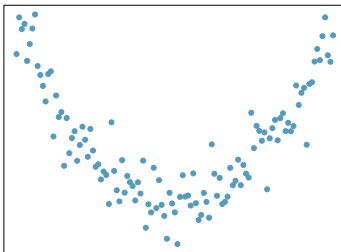


(a)

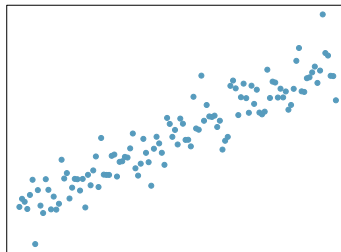


(b)

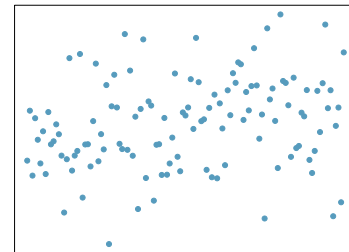
8.3 Identify relationships, Part I. For each of the six plots, identify the strength of the relationship (e.g. weak, moderate, or strong) in the data and whether fitting a linear model would be reasonable.



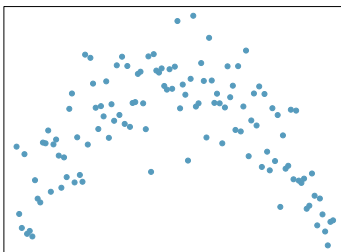
(a)



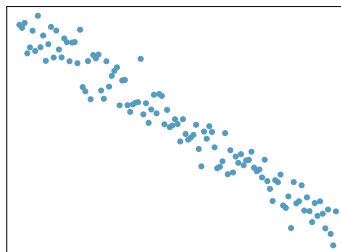
(b)



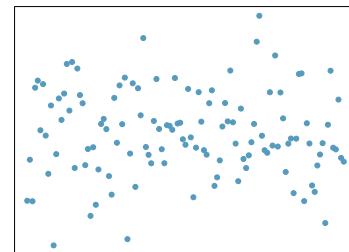
(c)



(d)

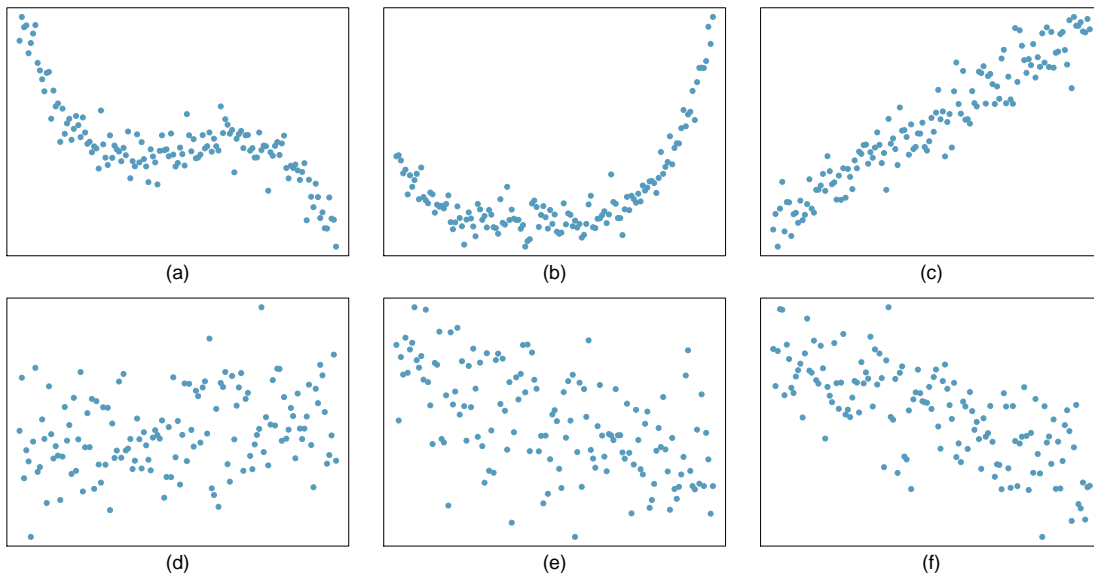


(e)



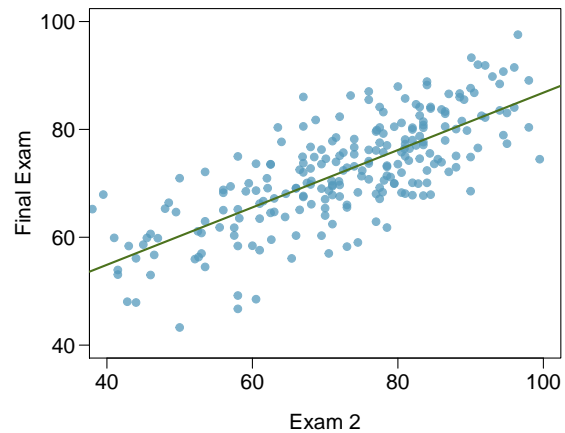
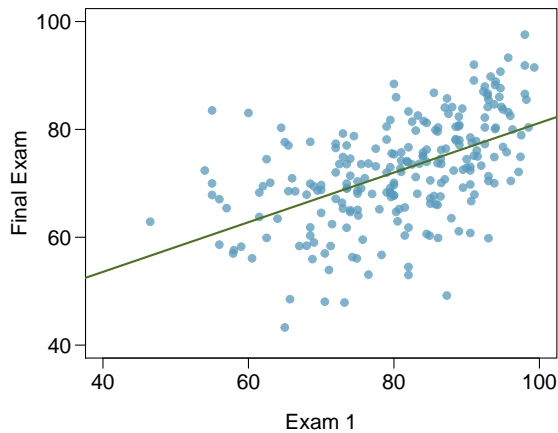
(f)

8.4 Identify relationships, Part II. For each of the six plots, identify the strength of the relationship (e.g. weak, moderate, or strong) in the data and whether fitting a linear model would be reasonable.

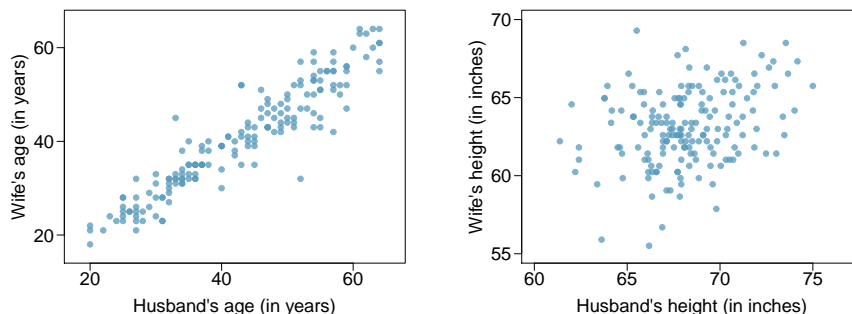


8.5 Exams and grades. The two scatterplots below show the relationship between final and mid-semester exam grades recorded during several years for a Statistics course at a university.

- Based on these graphs, which of the two exams has the strongest correlation with the final exam grade? Explain.
- Can you think of a reason why the correlation between the exam you chose in part (a) and the final exam is higher?



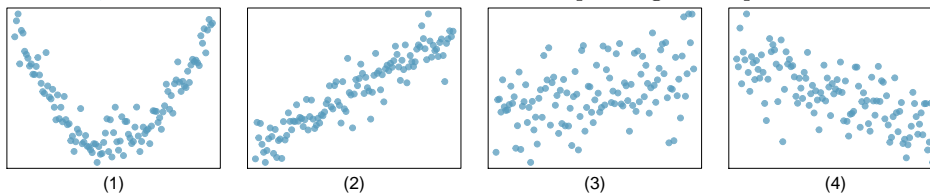
8.6 Husbands and wives, Part I. The Great Britain Office of Population Census and Surveys once collected data on a random sample of 170 married couples in Britain, recording the age (in years) and heights (converted here to inches) of the husbands and wives.⁵ The scatterplot on the left shows the wife's age plotted against her husband's age, and the plot on the right shows wife's height plotted against husband's height.



- Describe the relationship between husbands' and wives' ages.
- Describe the relationship between husbands' and wives' heights.
- Which plot shows a stronger correlation? Explain your reasoning.
- Data on heights were originally collected in centimeters, and then converted to inches. Does this conversion affect the correlation between husbands' and wives' heights?

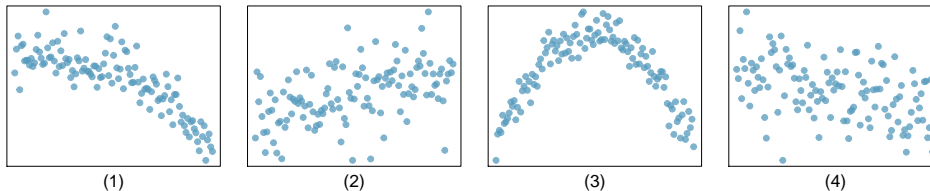
8.7 Match the correlation, Part I. Match each correlation to the corresponding scatterplot.

- $R = -0.7$
- $R = 0.45$
- $R = 0.06$
- $R = 0.92$

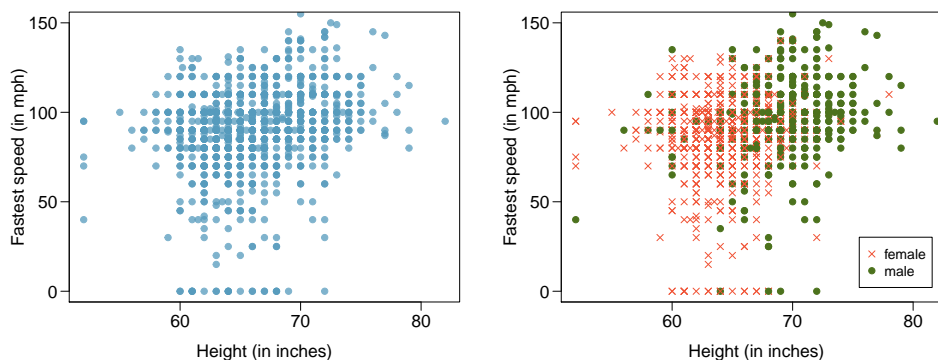


8.8 Match the correlation, Part II. Match each correlation to the corresponding scatterplot.

- $R = 0.49$
- $R = -0.48$
- $R = -0.03$
- $R = -0.85$



8.9 Speed and height. 1,302 UCLA students were asked to fill out a survey where they were asked about their height, fastest speed they have ever driven, and gender. The scatterplot on the left displays the relationship between height and fastest speed, and the scatterplot on the right displays the breakdown by gender in this relationship.



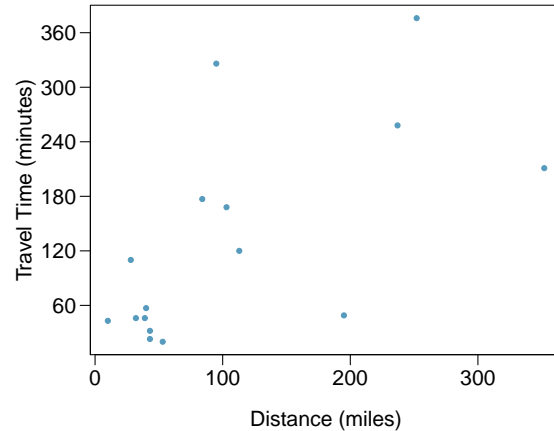
- Describe the relationship between height and fastest speed.
- Why do you think these variables are positively associated?
- What role does gender play in the relationship between height and fastest driving speed?

⁵D.J. Hand. *A handbook of small data sets*. Chapman & Hall/CRC, 1994.

8.10 Guess the correlation. Eduardo and Rosie are both collecting data on number of rainy days in a year and the total rainfall for the year. Eduardo records rainfall in inches and Rosie in centimeters. How will their correlation coefficients compare?

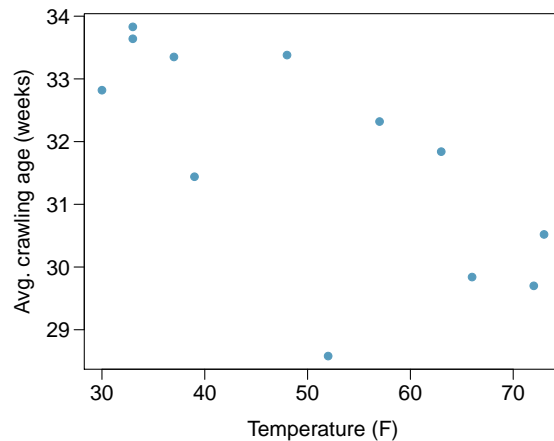
8.11 The Coast Starlight, Part I. The Coast Starlight Amtrak train runs from Seattle to Los Angeles. The scatterplot below displays the distance between each stop (in miles) and the amount of time it takes to travel from one stop to another (in minutes).

- Describe the relationship between distance and travel time.
- How would the relationship change if travel time was instead measured in hours, and distance was instead measured in kilometers?
- Correlation between travel time (in miles) and distance (in minutes) is $r = 0.636$. What is the correlation between travel time (in kilometers) and distance (in hours)?



8.12 Crawling babies, Part I. A study conducted at the University of Denver investigated whether babies take longer to learn to crawl in cold months, when they are often bundled in clothes that restrict their movement, than in warmer months.⁶ Infants born during the study year were split into twelve groups, one for each birth month. We consider the average crawling age of babies in each group against the average temperature when the babies are six months old (that's when babies often begin trying to crawl). Temperature is measured in degrees Fahrenheit ($^{\circ}\text{F}$) and age is measured in weeks.

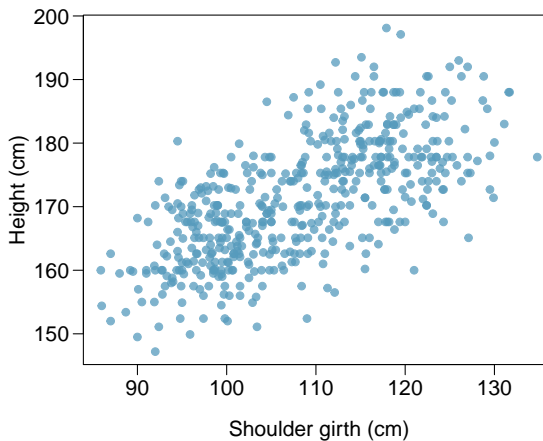
- Describe the relationship between temperature and crawling age.
- How would the relationship change if temperature was measured in degrees Celsius ($^{\circ}\text{C}$) and age was measured in months?
- The correlation between temperature in $^{\circ}\text{F}$ and age in weeks was $r = -0.70$. If we converted the temperature to $^{\circ}\text{C}$ and age to months, what would the correlation be?



⁶J.B. Benson. "Season of birth and onset of locomotion: Theoretical and methodological implications". In: *Infant behavior and development* 16.1 (1993), pp. 69–81. ISSN: 0163-6383.

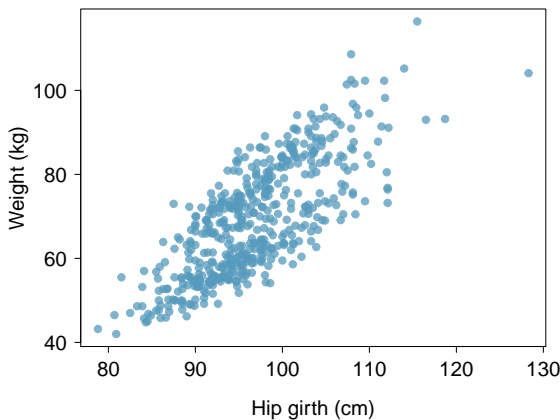
8.13 Body measurements, Part I. Researchers studying anthropometry collected body girth measurements and skeletal diameter measurements, as well as age, weight, height and gender for 507 physically active individuals.⁷ The scatterplot below shows the relationship between height and shoulder girth (over deltoid muscles), both measured in centimeters.

- Describe the relationship between shoulder girth and height.
- How would the relationship change if shoulder girth was measured in inches while the units of height remained in centimeters?



8.14 Body measurements, Part II. The scatterplot below shows the relationship between weight measured in kilograms and hip girth measured in centimeters from the data described in Exercise 8.13.

- Describe the relationship between hip girth and weight.
- How would the relationship change if weight was measured in pounds while the units for hip girth remained in centimeters?



8.15 Correlation, Part I. What would be the correlation between the ages of husbands and wives if men always married woman who were

- 3 years younger than themselves?
- 2 years older than themselves?
- half as old as themselves?

8.16 Correlation, Part II. What would be the correlation between the annual salaries of males and females at a company if for a certain type of position men always made

- \$5,000 more than women?
- 25% more than women?
- 15% less than women?

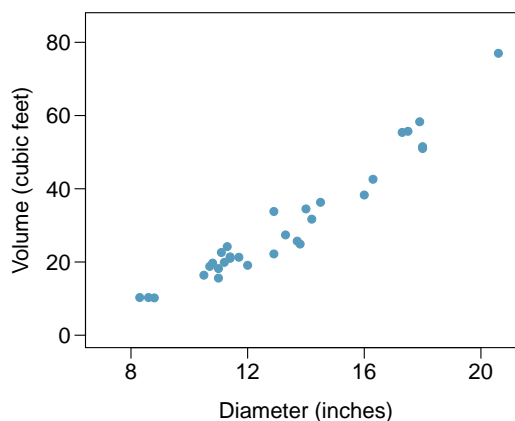
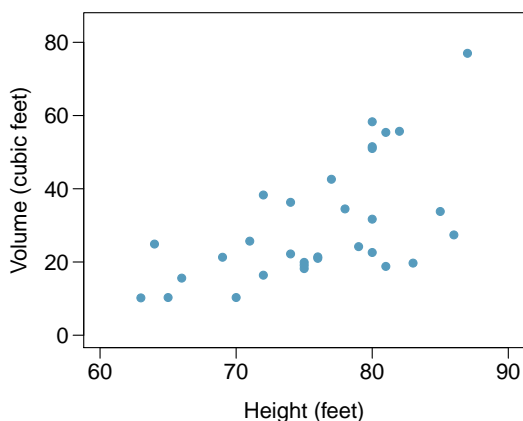
⁷G. Heinz et al. "Exploring relationships in body dimensions". In: *Journal of Statistics Education* 11.2 (2003).

Chapter exercises

8.37 True / False. Determine if the following statements are true or false. If false, explain why.

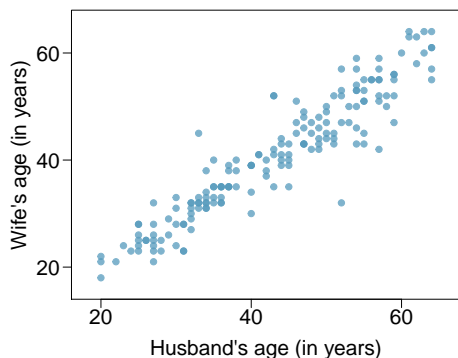
- A correlation coefficient of -0.90 indicates a stronger linear relationship than a correlation of 0.5 .
- Correlation is a measure of the association between any two variables.

8.38 Trees. The scatterplots below show the relationship between height, diameter, and volume of timber in 31 felled black cherry trees. The diameter of the tree is measured 4.5 feet above the ground.²⁰



- Describe the relationship between volume and height of these trees.
- Describe the relationship between volume and diameter of these trees.
- Suppose you have height and diameter measurements for another black cherry tree. Which of these variables would be preferable to use to predict the volume of timber in this tree using a simple linear regression model? Explain your reasoning.

8.39 Husbands and wives, Part III. Exercise 8.33 presents a scatterplot displaying the relationship between husbands' and wives' ages in a random sample of 170 married couples in Britain, where both partners' ages are below 65 years. Given below is summary output of the least squares fit for predicting wife's age from husband's age.



	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.5740	1.1501	1.37	0.1730
age_husband	0.9112	0.0259	35.25	0.0000

$df = 168$

- We might wonder, is the age difference between husbands and wives consistent across ages? If this were the case, then the slope parameter would be $\beta_1 = 1$. Use the information above to evaluate if there is strong evidence that the difference in husband and wife ages differs for different ages.
- Write the equation of the regression line for predicting wife's age from husband's age.
- Interpret the slope and intercept in context.
- Given that $R^2 = 0.88$, what is the correlation of ages in this data set?
- You meet a married man from Britain who is 55 years old. What would you predict his wife's age to be? How reliable is this prediction?
- You meet another married man from Britain who is 85 years old. Would it be wise to use the same linear model to predict his wife's age? Explain.

²⁰Source: R Dataset, stat.ethz.ch/R-manual/R-patched/library/datasets/html/trees.html.

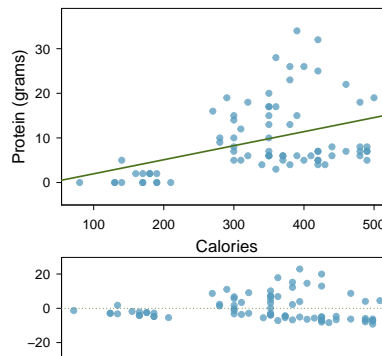
8.40 Cats, Part II. Exercise 8.26 presents regression output from a model for predicting the heart weight (in g) of cats from their body weight (in kg). The coefficients are estimated using a dataset of 144 domestic cat. The model output is also provided below.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.357	0.692	-0.515	0.607
body wt	4.034	0.250	16.119	0.000

$$s = 1.452 \quad R^2 = 64.66\% \quad R_{adj}^2 = 64.41\%$$

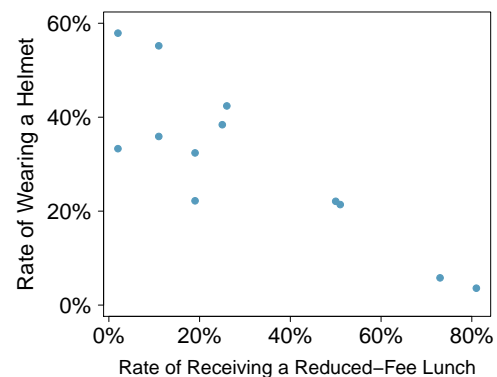
- We see that the point estimate for the slope is positive. What are the hypotheses for evaluating whether body weight is positively associated with heart weight in cats?
- State the conclusion of the hypothesis test from part (a) in context of the data.
- Calculate a 95% confidence interval for the slope of body weight, and interpret it in context of the data.
- Do your results from the hypothesis test and the confidence interval agree? Explain.

8.41 Nutrition at Starbucks, Part II. Exercise 8.22 introduced a data set on nutrition information on Starbucks food menu items. Based on the scatterplot and the residual plot provided, describe the relationship between the protein content and calories of these menu items, and determine if a simple linear model is appropriate to predict amount of protein from the number of calories.



8.42 Helmets and lunches. The scatterplot shows the relationship between socioeconomic status measured as the percentage of children in a neighborhood receiving reduced-fee lunches at school (**lunch**) and the percentage of bike riders in the neighborhood wearing helmets (**helmet**). The average percentage of children receiving reduced-fee lunches is 30.8% with a standard deviation of 26.7% and the average percentage of bike riders wearing helmets is 38.8% with a standard deviation of 16.9%.

- If the R^2 for the least-squares regression line for these data is 72%, what is the correlation between **lunch** and **helmet**?
- Calculate the slope and intercept for the least-squares regression line for these data.
- Interpret the intercept of the least-squares regression line in the context of the application.
- Interpret the slope of the least-squares regression line in the context of the application.
- What would the value of the residual be for a neighborhood where 40% of the children receive reduced-fee lunches and 40% of the bike riders wear helmets? Interpret the meaning of this residual in the context of the application.



8.43 Match the correlation, Part III. Match each correlation to the corresponding scatterplot.

- $r = -0.72$
- $r = 0.07$
- $r = 0.86$
- $r = 0.99$

