1. **Cereals.** For many people, breakfast cereal is an important source of fiber in their diets. Cereals also contain potassium, a mineral shown to be associated with maintaining a healthy blood pressure. An analysis of the amount of fiber (in grams) and the potassium content (in milligrams) in servings of 27 breakfast cereals produced the regression model \( \text{Potassium} = 38 + 27 \text{Fiber} \). If your cereal provides 9 grams of fiber per serving, how much potassium does the model estimate you will get?

2. **Horsepower.** In Chapter 7's Exercise 33 we examined the relationship between the fuel economy (mpg) and horsepower for 15 models of cars. Further analysis produces the regression model \( \text{mpg} = 46.87 - 0.084 \text{HP} \). If the car you are thinking of buying has a 200-horsepower engine, what does this model suggest your gas mileage would be?

3. **More cereal.** Exercise 1 describes a regression model that estimates a cereal’s potassium content from the amount of fiber it contains. In this context, what does it mean to say that a cereal has a negative residual?

5. **Another bowl.** In Exercise 1, the regression model \( \text{Potassium} = 38 + 27 \text{Fiber} \) relates fiber (in grams) and potassium content (in milligrams) in servings of breakfast cereals. Explain what the slope means.

7. **Cereal again.** The correlation between a cereal’s fiber and potassium contents is \( r = 0.903 \). What fraction of the variability in potassium is accounted for by the amount of fiber that servings contain?

13. **What slope?** If you create a regression model for predicting the Weight of a car (in pounds) from its Length (in feet), is the slope most likely to be 3, 30, 300, or 3000? Explain.

15. **Real estate.** A random sample of records of sales of homes from Feb. 15 to Apr. 30, 1993, from the files maintained by the Albuquerque Board of Realtors gives the Price and Size (in square feet) of 117 homes. A regression to predict Price (in thousands of dollars) from Size has an R-squared of 71.4%. The residuals plot indicated that a linear model is appropriate.
   a) What are the variables and units in this regression?
   b) What units does the slope have?
   c) Do you think the slope is positive or negative? Explain.

21. **More real estate.** Consider the Albuquerque home sales from Exercise 15 again. The regression analysis gives the model \( \text{Price} = 47.82 + 0.061 \text{Size} \).
   a) Explain what the slope of the line says about housing prices and house size.
   b) What price would you predict for a 3000-square-foot house in this market?
   c) A real estate agent shows a potential buyer a 1200-square-foot home, saying that the asking price is $6000 less than what one would expect to pay for a house of this size. What is the asking price, and what is the $6000 called?

27. **Cigarettes.** Is the nicotine content of a cigarette related to the “tars”? A collection of data (in milligrams) on 29 cigarettes produced the scatterplot, residuals plot, and regression analysis shown:

![Scatterplot and residuals plot for cigarette data]

Dependent variable is: nicotine
R squared = 92.4%

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.154030</td>
</tr>
<tr>
<td>Tar</td>
<td>0.065052</td>
</tr>
</tbody>
</table>

a) Do you think a linear model is appropriate here? Explain.
b) Explain the meaning of \( R^2 \) in this context.
28. **Attendance 2006.** In the previous chapter you looked at the relationship between the number of wins by American League baseball teams and the average attendance at their home games for the 2006 season. Here are the scatterplot, the residuals plot, and part of the regression analysis:

![Graph showing relationship between home attendance and wins.]

- **Dependent variable is: Home Attendance**
- **R squared = 48.5%**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-14364.5</td>
</tr>
<tr>
<td>Wins</td>
<td>538.915</td>
</tr>
</tbody>
</table>

a) Do you think a linear model is appropriate here? Explain.
b) Interpret the meaning of $R^2$ in this context.
c) Do the residuals show any pattern worth remarking on?
d) The point in the upper right of the plots is the New York Yankees. What can you say about the residual for the Yankees?

29. **Another cigarette.** Consider again the regression of Nicotine content on Tar (both in milligrams) for the cigarettes examined in Exercise 27.

a) What is the correlation between Tar and Nicotine?
b) What would you predict about the average Nicotine content of cigarettes that are 2 standard deviations below average in Tar content?
c) If a cigarette is 1 standard deviation above average in Nicotine content, what do you suspect is true about its Tar content?

32. **Last inning 2006.** Refer again to the regression analysis for average attendance and games won by American League baseball teams, seen in Exercise 28.

a) Write the equation of the regression line.
b) Estimate the Average Attendance for a team with 50 Wins.
c) Interpret the meaning of the slope of the regression line in this context.
d) In general, what would a negative residual mean in this context?
e) The St. Louis Cardinals, the 2006 World Champions, are not included in these data because they are a National League team. During the 2006 regular season, the Cardinals won 83 games and averaged 42,588 fans at their home games. Calculate the residual for this team, and explain what it means.
Online clothes. An online clothing retailer keeps track of its customers’ purchases. For those customers who signed up for the company’s credit card, the company also has information on the customer’s Age and Income. A random sample of 500 of these customers shows the following scatterplot of Total Yearly Purchases by Age:

The correlation between Total Yearly Purchases and Age is $r = 0.037$. Summary statistics for the two variables are:

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>29.67 yrs</td>
<td>8.51 yrs</td>
</tr>
<tr>
<td>Total Yearly Purchase</td>
<td>$572.52</td>
<td>$253.62</td>
</tr>
</tbody>
</table>

a) What is the linear regression equation for predicting Total Yearly Purchase from Age?
b) Do the assumptions and conditions for regression appear to be met?
c) What is the predicted average Total Yearly Purchase for an 18-year-old? For a 50-year-old?
d) What percent of the variability in Total Yearly Purchases is accounted for by this model?
e) Do you think the regression might be a useful one for the company? Explain.

SAT scores. The SAT is a test often used as part of an application to college. SAT scores are between 200 and 800, but have no units. Tests are given in both Math and Verbal areas. Doing the SAT-Math problems also involves the ability to read and understand the questions, but can a person’s verbal score be used to predict the math score? Verbal and math SAT scores of a high school graduating class are displayed in the scatterplot, with the regression line added.

a) Describe the relationship.
b) Are there any students whose scores do not seem to fit the overall pattern?
c) For these data, $r = 0.685$. Interpret this statistic.
d) These verbal scores averaged 596.3, with a standard deviation of 99.5, and the math scores averaged 6122, with a standard deviation of 96.1. Write the equation of the regression line.
e) Interpret the slope of this line.
f) Predict the math score of a student with a verbal score of 500.
g) Every year some student scores a perfect 1600. Based on this model, what would be that student’s Math score residual?