SECTION 3.2 Exercises

35. **What's my line?** You use the same bar of soap to shower each morning. The bar weighs 80 grams when it is new. Its weight goes down by 6 grams per day on the average. What is the equation of the regression line for predicting weight from days of use?

36. **What's my line?** An eccentric professor believes that a child with IQ 100 should have a reading test score of 50, and that reading score should increase by 1 point for every additional point of IQ. What is the equation of the professor's regression line for predicting reading score from IQ?

37. **Gas mileage** We expect a car's highway gas mileage to be related to its city gas mileage. Data for all 1198 vehicles in the government's *2008 Fuel Economy Guide* give the regression line predicted highway mpg $= 4.62 + 1.109$ (city mpg).

(a) What's the slope of this line? Interpret this value in context.

(b) What's the intercept? Explain why the value of the intercept is not statistically meaningful.

(c) Find the predicted highway mileage for a car that gets 16 miles per gallon in the city. Do the same for a car with city mileage 28 mpg.

38. **IQ and reading scores** Data on the IQ test scores and reading test scores for a group of fifth-grade children give the following regression line: predicted reading score $= -33.4 + 0.882$(IQ score).

(a) What's the slope of this line? Interpret this value in context.

(b) What's the intercept? Explain why the value of the intercept is not statistically meaningful.

(c) Find the predicted reading scores for two children with IQ scores of 90 and 130, respectively.

39. **Acid rain** Researchers studying acid rain measured the acidity of precipitation in a Colorado wilderness area for 150 consecutive weeks. Acidity is measured by pH. Lower pH values show higher acidity. The researchers observed a linear pattern over time. They reported that the regression line $\widehat{\text{pH}} = 5.43 - 0.0053(\text{weeks})$ fit the data well.[16]

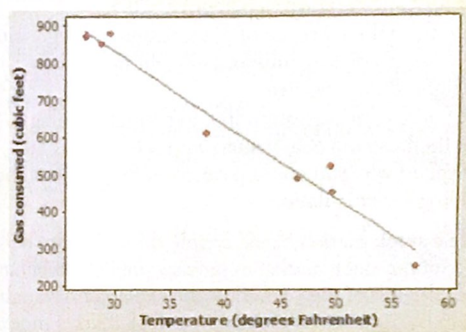
Pg 166

(a) Identify the slope of the line and explain what it means in this setting.

(b) Identify the y intercept of the line and explain what it means in this setting.

(c) According to the regression line, what was the pH at the end of this study?

40. **How much gas?** In Exercise 4 (page 158), we examined the relationship between the average monthly temperature and the amount of natural gas consumed in Joan's midwestern home. The figure below shows the original scatterplot with the least-squares line added. The equation of the least-squares line is $\hat{y} = 1425 - 19.87x$.



(a) Identify the slope of the line and explain what it means in this setting.

(b) Identify the y intercept of the line. Explain why it's risky to use this value as a prediction.

(c) Use the regression line to predict the amount of natural gas Joan will use in a month with an average temperature of 30°F.

41. **Acid rain** Refer to Exercise 39. Would it be appropriate to use the regression line to predict pH after 1000 months? Justify your answer.

42. **How much gas?** Refer to Exercise 40. Would it be appropriate to use the regression line to predict Joan's natural-gas consumption in a future month with an average temperature of 65°F? Justify your answer.

43. **Least-squares idea** The table below gives a small set of data. Which of the following two lines fits the data better: $\hat{y} = 1 - x$ or $\hat{y} = 3 - 2x$? Make a graph of the data and use it to help justify your answer. (*Note*: Neither of these two lines is the least-squares regression line for these data.)

x:	-1	1	1	3	5
y:	2	0	1	-1	-5

44. **Least-squares idea** Trace the graph from Exercise 40 on your paper. Show why the line drawn on the plot is called the least-squares line.

45. **Acid rain** In the acid rain study of Exercise 39, the actual pH measurement for Week 50 was 5.08. Find and interpret the residual for this week.

pg 168

46. **How much gas?** Refer to Exercise 40. During March, the average temperature was 46.4°F and Joan used 490 cubic feet of gas per day. Find and interpret the residual for this month.

47. **Husbands and wives** The mean height of American women in their early twenties is 64.5 inches and the standard deviation is 2.5 inches. The mean height of men the same age is 68.5 inches, with standard deviation 2.7 inches. The correlation between the heights of husbands and wives is about $r = 0.5$.

pg 173

(a) Find the equation of the least-squares regression line for predicting husband's height from wife's height. Show your work.
(b) Use your regression line to predict the height of the husband of a woman who is 67 inches tall. Explain why you could have given this result without doing the calculation.

48. **The stock market** Some people think that the behavior of the stock market in January predicts its behavior for the rest of the year. Take the explanatory variable x to be the percent change in a stock market index in January and the response variable y to be the change in the index for the entire year. We expect a positive correlation between x and y because the change during January contributes to the full year's change. Calculation from data for an 18-year period gives

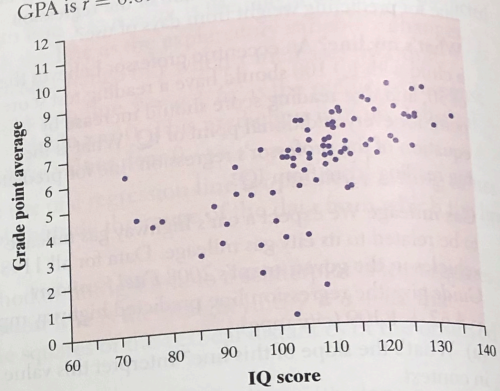
$$\bar{x} = 1.75\% \quad s_x = 5.36\% \quad \bar{y} = 9.07\%$$
$$s_y = 15.35\% \quad r = 0.596$$

(a) Find the equation of the least-squares line for predicting full-year change from January change. Show your work.
(b) The mean change in January is $\bar{x} = 1.75\%$. Use your regression line to predict the change in the index in a year in which the index rises 1.75% in January. Why could you have given this result (up to roundoff error) without doing the calculation?

49. **Husbands and wives** Refer to Exercise 47.
(a) Find r^2 and interpret this value in context.
(b) For these data, $s = 1.2$. Explain what this value means.

50. **The stock market** Refer to Exercise 48.
(a) What percent of the observed variation in yearly changes in the index is explained by a straight-line relationship with the change during January?
(b) For these data, $s = 8.3$. Explain what this value means.

51. **IQ and grades** Exercise 3 (page 158) included the plot shown below of school grade point average (GPA) against IQ test score for 78 seventh-grade students. (GPA was recorded on a 12-point scale with A+ = 12, A = 11, A− = 10, B+ = 9, . . . , D− = 1, and F = 0.) Calculation shows that the mean and standard deviation of the IQ scores are $\bar{x} = 108.9$ and $s_x = 13.17$. For the GPAs, these values are $\bar{y} = 7.447$ and $s_y = 2.10$. The correlation between IQ and GPA is $r = 0.6337$.



(a) Find the equation of the least-squares line for predicting GPA from IQ. Show your work.
(b) What percent of the observed variation in these students' GPAs can be explained by the linear relationship between GPA and IQ?
(c) One student has an IQ of 103 but a very low GPA of 0.53. Find and interpret the residual for this student.

52. **Will I bomb the final?** We expect that students who do well on the midterm exam in a course will usually also do well on the final exam. Gary Smith of Pomona College looked at the exam scores of all 346 students who took his statistics class over a 10-year period.[17] The least-squares line for predicting final-exam score from midterm-exam score was $\hat{y} = 46.6 + 0.41x$. Octavio scores 10 points above the class mean on the midterm. How many points above the class mean do you predict that he will score on the final? (This is an example of the phenomenon that gave "regression" its name: students who do well on the midterm will on the average do less well, but still above average, on the final.)

53. **Bird colonies** Exercise 6 (page 159) examined the relationship between the number of new birds y and percent of returning birds x for 13 sparrowhawk colonies. Here are the data once again.

Percent return:	74	66	81	52	73	62	52	45	62	46	60	46	38
New adults:	5	6	8	11	12	15	16	17	18	18	19	20	20

(a) Enter the data into your calculator and make a scatterplot.

(b) Use your calculator's regression function to find the equation of the least-squares regression line. Add this line to your scatterplot from (a).

(c) Explain in words what the slope and y intercept of the regression line tell us.

(d) An ecologist uses the line to predict how many birds will join another colony of sparrowhawks, to which 60% of the adults from the previous year return. What's the prediction?

54. **Do heavier people burn more energy?** Exercise 10 (page 159) presented data on the lean body mass and resting metabolic rate for 12 women who were subjects in a study of dieting. Lean body mass, given in kilograms, is a person's weight leaving out all fat. Metabolic rate, in calories burned per 24 hours, is the rate at which the body consumes energy. Here are the data again.

Mass:	36.1	54.6	48.5	42.0	50.6	42.0	40.3	33.1	42.4	34.5	51.1	41.2
Rate:	995	1425	1396	1418	1502	1256	1189	913	1124	1052	1347	1204

(a) Enter the data into your calculator and make a scatterplot.

(b) Use your calculator's regression function to find the equation of the least-squares regression line. Add this line to your scatterplot from (a).

(c) Explain in words what the slope of the regression line tells us.

(d) Another woman has a lean body mass of 45 kilograms. What is her predicted metabolic rate?

55. **Bird colonies** Refer to Exercise 53.

(a) Use your calculator to make a residual plot. Describe what this graph tells you about how well the line fits the data.

(b) Which point has the largest size residual? Explain what this residual means in context.

56. **Do heavier people burn more energy?** Refer to Exercise 54.

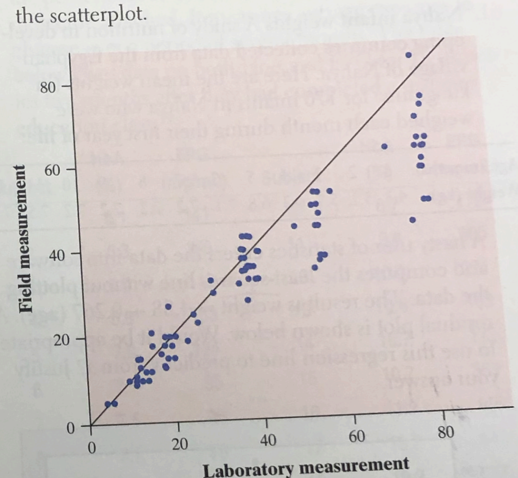
(a) Use your calculator to make a residual plot. Describe what this graph tells you about how well the line fits the data.

(b) Which point has the largest residual? Explain what the value of that residual means in context.

57. **Bird colonies** Refer to Exercises 53 and 55. For the regression you performed earlier, $r^2 = 0.56$ and $s = 3.67$. Explain what each of these values means in this setting.

58. **Do heavier people burn more energy?** Refer to Exercises 54 and 56. For the regression you performed earlier, $r^2 = 0.768$ and $s = 95.08$. Explain what each of these values means in this setting.

59. **Oil and residuals** The Trans-Alaska Oil Pipeline is a tube that is formed from 1/2-inch-thick steel and that carries oil across 800 miles of sensitive arctic and subarctic terrain. The pipe segments and the welds that join them were carefully examined before installation. How accurate are field measurements of the depth of small defects? The figure below compares the results of measurements on 100 defects made in the field with measurements of the same defects made in the laboratory.[18] The line $y = x$ is drawn on the scatterplot.



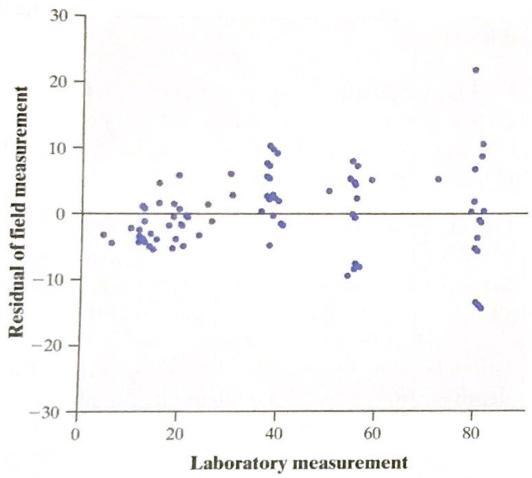
(a) Describe the overall pattern you see in the scatterplot, as well as any deviations from that pattern.

(b) If field and laboratory measurements all agree, then the points should fall on the $y = x$ line drawn on the plot, except for small variations in the measurements. Is this the case? Explain.

(c) The line drawn on the scatterplot ($y = x$) is *not* the least-squares regression line. How would the slope and y intercept of the least-squares line compare? Justify your answer.

60. **Oil and residuals** Refer to Exercise 59. The following figure shows a residual plot for the least-squares regression line. Discuss what the residual plot tells

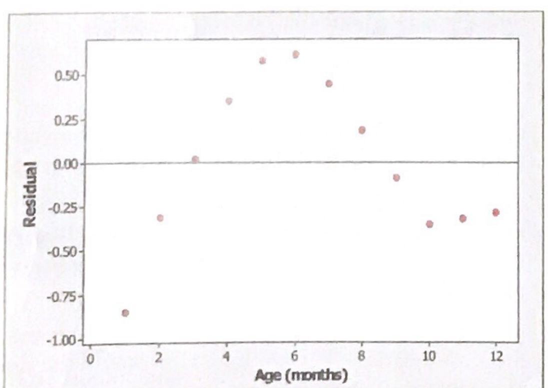
you about how well the least-squares regression line fits the data.



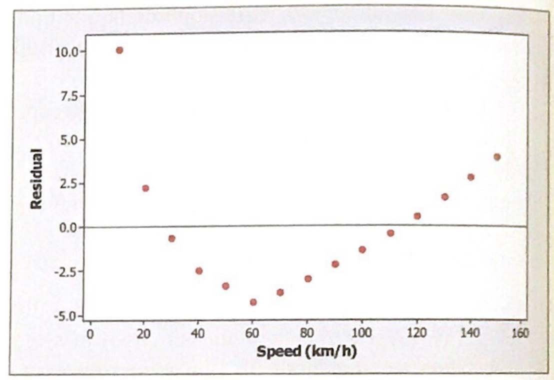
61. **Nahya infant weights** A study of nutrition in developing countries collected data from the Egyptian village of Nahya. Here are the mean weights (in kilograms) for 170 infants in Nahya who were weighed each month during their first year of life:

Age (months):	1	2	3	4	5	6	7	8	9	10	11	12
Weight (kg):	4.3	5.1	5.7	6.3	6.8	7.1	7.2	7.2	7.2	7.2	7.5	7.8

A hasty user of statistics enters the data into software and computes the least-squares line without plotting the data. The result is $\widehat{\text{weight}} = 4.88 + 0.267$ (age). A residual plot is shown below. Would it be appropriate to use this regression line to predict y from x? Justify your answer.

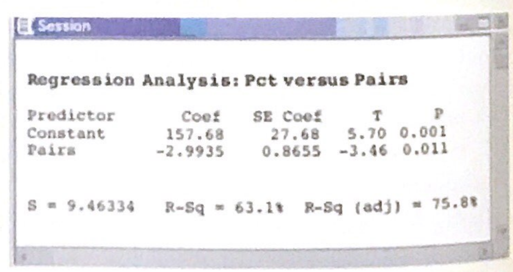


62. **Driving speed and fuel consumption** Exercise 9 (page 159) gives data on the fuel consumption y of a car at various speeds x. Fuel consumption is measured in liters of gasoline per 100 kilometers driven and speed is measured in kilometers per hour. A statistical software package gives the least-squares regression line and the residual plot shown below. The regression line is $\hat{y} = 11.058 - 0.01466x$. Would it be appropriate to use the regression line to predict y from x? Justify your answer.



63. **Merlins breeding** Exercise 13 (page 160) gives data on the number of breeding pairs of merlins in an isolated area in each of nine years and the percent of males who returned the next year. The data show that the percent returning is lower after successful breeding seasons and that the relationship is roughly linear. The figure below shows Minitab regression output for these data.

pg 182



(a) What is the equation of the least-squares regression line for predicting the percent of males that return from the number of breeding pairs? Use the equation to predict the percent of returning males after a season with 30 breeding pairs.

(b) What percent of the year-to-year variation in percent of returning males is explained by the straight-line relationship with number of breeding pairs the previous year?

(c) Use the information in the figure to find the correlation r between percent of males that return and number of breeding pairs. How do you know whether the sign of r is $+$ or $-$?

(d) Interpret the value of s in this setting.

64. **Does social rejection hurt?** Exercise 14 (page 160) gives data from a study that shows that social exclusion causes "real pain." That is, activity in an area of the brain that responds to physical pain goes up as distress from social exclusion goes up. A scatterplot shows a moderately strong, linear relationship. The figure below shows Minitab regression output for these data.

```
Session

Regression Analysis: Brain versus Distress

Predictor       Coef    SE Coef      T      P
Constant    -0.12608    0.02465  -5.12  0.000
distress    0.060782   0.009979   6.09  0.000

S = 0.0250896   R-Sq = 77.1%   R-Sq (adj) = 75.1%
```

(a) What is the equation of the least-squares regression line for predicting brain activity from social distress score? Use the equation to predict brain activity for social distress score 2.0.

(b) What percent of the variation in brain activity among these subjects is explained by the straight-line relationship with social distress score?

(c) Use the information in the figure to find the correlation r between social distress score and brain activity. How do you know whether the sign of r is + or −?

(d) Interpret the value of s in this setting.

65. **Outsourcing by airlines** Exercise 5 (page 158) gives data for 14 airlines on the percent of major maintenance outsourced and the percent of flight delays blamed on the airline.

Airline	Outsource percent	Delay percent
AirTran	66	14
Alaska	92	42
American	46	26
America West	76	39
ATA	18	19
Continental	69	20
Delta	48	26
Frontier	65	31
Hawaiian	80	70
JetBlue	68	18
Northwest	76	43
Southwest	68	20
United	63	27
US Airways	77	24

(a) Make a scatterplot with outsourcing percent as x and delay percent as y. Hawaiian Airlines is a high outlier in the y direction. Because several other airlines have similar values of x, the influence of this outlier is unclear without actual calculation.

(b) Find the correlation r with and without Hawaiian Airlines. How influential is the outlier for correlation?

(c) Find the least-squares line for predicting y from x with and without Hawaiian Airlines. Draw both lines on your scatterplot. Use both lines to predict the percent of delays blamed on an airline that has outsourced 76% of its major maintenance. How influential is the outlier for the least-squares line?

66. **Managing diabetes** People with diabetes measure their fasting plasma glucose (FPG; measured in units of milligrams per milliliter) after fasting for at least 8 hours. Another measurement, made at regular medical checkups, is called HbA. This is roughly the percent of red blood cells that have a glucose molecule attached. It measures average exposure to glucose over a period of several months. The table below gives data on both HbA and FPG for 18 diabetics five months after they had completed a diabetes education class.[19]

Subject	HbA (%)	FPG (mg/mL)	Subject	HbA (%)	FPG (mg/mL)
1	6.1	141	10	8.7	172
2	6.3	158	11	9.4	200
3	6.4	112	12	10.4	271
4	6.8	153	13	10.6	103
5	7.0	134	14	10.7	172
6	7.1	95	15	10.7	359
7	7.5	96	16	11.2	145
8	7.7	78	17	13.7	147
9	7.9	148	18	19.3	255

(a) Make a scatterplot with HbA as the explanatory variable. There is a positive linear relationship, but it is surprisingly weak.

(b) Subject 15 is an outlier in the y direction. Subject 18 is an outlier in the x direction. Find the correlation for all 18 subjects, for all except Subject 15, and for all except Subject 18. Are either or both of these subjects influential for the correlation? Explain in simple language why r changes in opposite directions when we remove each of these points.

(c) Add three regression lines for predicting FPG from HbA to your scatterplot: for all 18 subjects, for all except Subject 15, and for all except Subject 18. Is either Subject 15 or Subject 18 strongly influential for the least-squares line? Explain in simple language what features of the scatterplot explain the degree of influence.

67. **Bird colonies** Return to the data of Exercise 53 on sparrowhawk colonies. We'll use these data to illustrate influence.

(a) Make a scatterplot of the data suitable for predicting new adults from percent of returning adults. Then add two new points. Point A: 10% return, 15 new adults. Point B: 60% return, 28 new adults. In which direction is each new point an outlier?

(b) Add three least-squares regression lines to your plot: for the original 13 colonies, for the original colonies plus Point A, and for the original colonies plus Point B. Which new point is more influential for the regression line? Explain in simple language why each new point moves the line in the way your graph shows.

68. **Beer and blood alcohol** The example on page 182 describes a study in which adults drank different amounts of beer. The response variable was their blood alcohol content (BAC). BAC for the same amount of beer might depend on other facts about the subjects. Name two other variables that could account for the fact that $r^2 = 0.80$.

pg 185

STEP 4

69. **Predicting tropical storms** William Gray heads the Tropical Meteorology Project at Colorado State University. His forecasts before each year's hurricane season attract lots of attention. Here are data on the number of named Atlantic tropical storms predicted by Dr. Gray and the actual number of storms for the years 1984 to 2008:[20]

Year	Forecast	Actual	Year	Forecast	Actual
1984	10	12	1997	11	7
1985	11	11	1998	10	14
1986	8	6	1999	14	12
1987	8	7	2000	12	14
1988	11	12	2001	12	15
1989	7	11	2002	11	12
1990	11	14	2003	14	16
1991	8	8	2004	14	14
1992	8	6	2005	15	27
1993	11	8	2006	17	9
1994	9	7	2007	17	14
1995	12	19	2008	15	16
1996	10	13			

Analyze these data. How accurate are Dr. Gray's forecasts? How many tropical storms would you expect in a year when his preseason forecast calls for 16 storms? What is the effect of the disastrous 2005 season on your answers? Follow the four-step process.

STEP 4

70. **Beavers and beetles** Do beavers benefit beetles? Researchers laid out 23 circular plots, each 4 meters in diameter, in an area where beavers were cutting down cottonwood trees. In each plot, they counted the number of stumps from trees cut by beavers and the number of clusters of beetle larvae. Ecologists think that the new sprouts from stumps are more tender than other cottonwood growth, so that beetles prefer them. If so, more stumps should produce more beetle larvae. Here are the data:[21]

Stumps:	2	2	1	3	3	4	3	1	2	5	1	3
Beetle larvae:	10	30	12	24	36	40	43	11	27	56	18	40

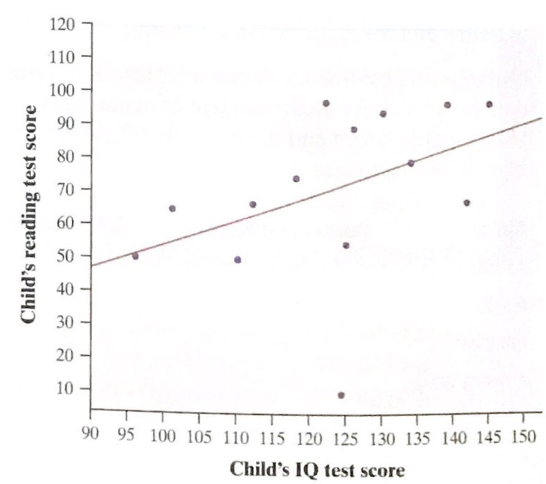
Stumps:	2	1	2	2	1	1	4	1	2	1	4
Beetle larvae:	25	8	21	14	16	6	54	9	13	14	50

Analyze these data to see if they support the "beavers benefit beetles" idea. Follow the four-step process.

Multiple choice: Select the best answer for Exercises 71 to 78.

71. The figure below is a scatterplot of reading test scores against IQ test scores for 14 fifth-grade children. The line is the least-squares regression line for predicting reading score from IQ score. If another child in this class has IQ score 110, you predict the reading score to be close to

(a) 50. (b) 60. (c) 70. (d) 80. (e) 90.



72. The slope of the line in the figure above is closest to

(a) −1. (b) 0. (c) 1. (d) 2. (e) 46.

73. Smokers don't live as long (on average) as nonsmokers, and heavy smokers don't live as long as light smokers. You perform least-squares regression on the age at death of a large group of male smokers y and the number of packs per day they smoked x. The slope of your regression line

(a) will be greater than 0.
(b) will be less than 0.
(c) will be equal to 0.
(d) You can't perform regression on these data.
(e) You can't tell without seeing the data.

Exercises 74 to 78 refer to the following setting. Measurements on young children in Mumbai, India, found this least-squares line for predicting height y from arm span x:[22]

$$\hat{y} = 6.4 + 0.93x$$

Measurements are in centimeters (cm).

74. How much does height increase on average for each additional centimeter of arm span?
(a) 0.93 cm (c) 5.81 cm (e) 7.33 cm
(b) 1.08 cm (d) 6.4 cm

75. According to the regression line, the predicted height of a child with an arm span of 100 cm is about
(a) 106.4 cm. (c) 93 cm. (e) 7.33 cm.
(b) 99.4 cm. (d) 15.7 cm.

76. By looking at the equation of the least-squares regression line, you can see that the correlation between height and arm span is
(a) greater than zero.
(b) less than zero.
(c) 0.93.
(d) 6.4.
(e) Can't tell without seeing the data.

77. In addition to the regression line, the report on the Mumbai measurements says that $r^2 = 0.95$. This suggests that
(a) although arm span and height are correlated, arm span does not predict height very accurately.
(b) height increases by $\sqrt{0.95} = 0.97$ cm for each additional centimeter of arm span.
(c) 95% of the relationship between height and arm span is accounted for by the regression line.
(d) 95% of the variation in height is accounted for by the regression line.
(e) 95% of the height measurements are accounted for by the regression line.

78. One child in the Mumbai study had height 59 cm and arm span 60 cm. This child's residual is
(a) −3.2 cm. (c) −1.3 cm. (e) 62.2 cm.
(b) −2.2 cm. (d) 3.2 cm.

Exercises 79 and 80 refer to the following setting. In its *Fuel Economy Guide* for 2008 model vehicles, the Environmental Protection Agency gives data on 1152 vehicles. There are a number of outliers, mainly vehicles with very poor gas mileage. If we ignore the outliers, however, the combined city and highway gas mileage of the other 1120 or so vehicles is approximately Normal with mean 18.7 miles per gallon (mpg) and standard deviation 4.3 mpg.

79. **In my Chevrolet** (2.2) The 2008 Chevrolet Malibu with a four-cylinder engine has a combined gas mileage of 25 mpg. What percent of all vehicles have worse gas mileage than the Malibu?

80. **The top 10%** (2.2) How high must a 2008 vehicle's gas mileage be in order to fall in the top 10% of all vehicles? (The distribution omits a few high outliers, mainly hybrid gas-electric vehicles.)

81. **Marijuana and traffic accidents** (1.1) Researchers in New Zealand interviewed 907 drivers at age 21. They had data on traffic accidents and they asked the drivers about marijuana use. Here are data on the numbers of accidents caused by these drivers at age 19, broken down by marijuana use at the same age:[23]

	Marijuana Use per Year			
	Never	1–10 times	11–50 times	51 + times
Drivers	452	229	70	156
Accidents caused	59	36	15	50

(a) Make a graph that displays the accident rate for each class. Is there evidence of an association between marijuana use and traffic accidents?
(b) Explain why we can't conclude that marijuana use *causes* accidents.