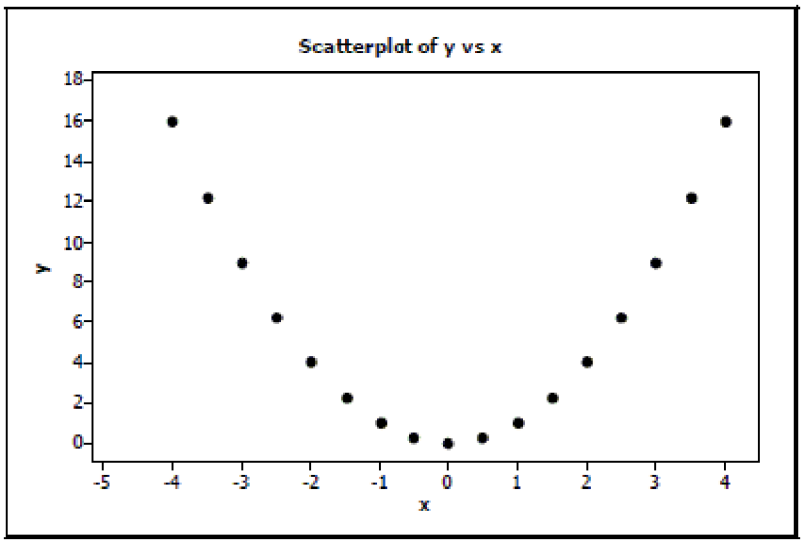
SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

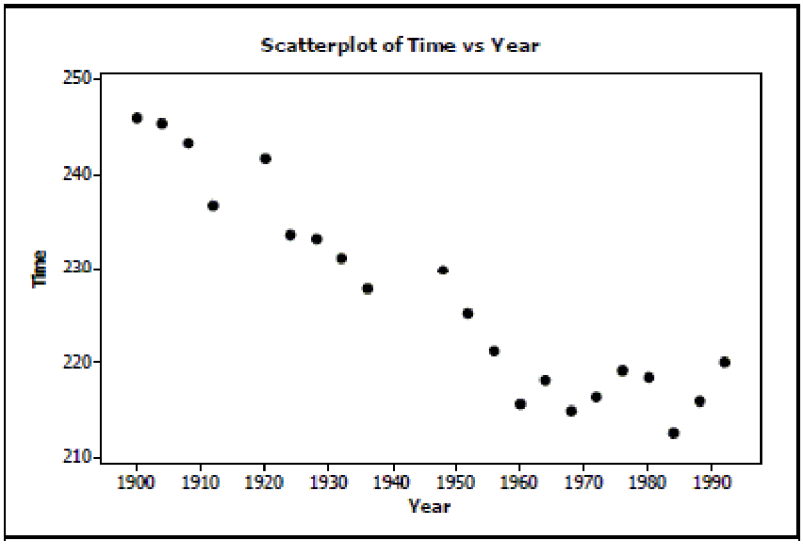
Solve the problem.

1) After conducting a survey of his students, a professor reported that "There appears to be a strong correlation between grade point average and whether or not a student works." Comment on this observation. 1) _____

2) The following scatterplot shows a relationship between x and y that results in a correlation coefficient of $r = 0$. Explain why $r = 0$ in this situation even though there appears to be a strong relationship between the x and y variables. 2) _____



3) The following scatterplot shows the relationship between the time (in seconds) it took men to run the 1500m race for the gold medal and the year of the Olympics that the race was run in: 3) _____



a. Write a few sentences describing the association.
b. Estimate the correlation. r = _____

4) Identify what is wrong with each of the following statements: 4) _____

a. The correlation between Olympic gold medal times for the 800m hurdles and year is -0.66 seconds per year.

b. The correlation between Olympic gold medal times for the 100m dash and year is -1.37.

c. Since the correlation between Olympic gold medal times for the 800m hurdles and 100m dash is -0. 41, the correlation between times for the 100m dash and the 800m hurdles is +0.41.

d. If we were to measure Olympic gold medal times for the 800m hurdles in minutes instead of seconds, the correlation would be -0.66/60 = -0.011.

5) After conducting a survey at a pet store to see what impact having a pet had on the condition of the yard, a news reporter stated "There appears to be a strong correlation between the owning a pet and the condition of the yard." Comment on this observation. 5) _____

6) On the axes below, sketch a scatterplot described: 6) _____

a. a strong positive association

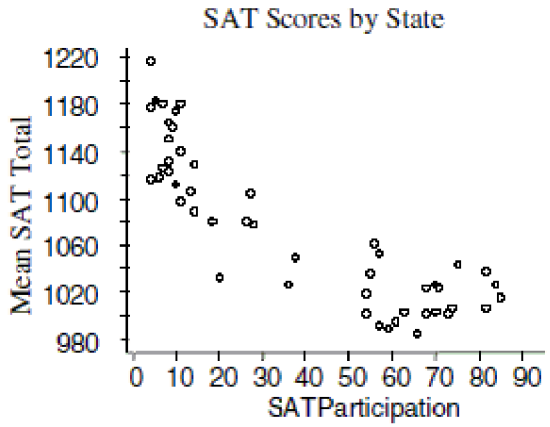
b. a weak negative association

7) A study by a prominent psychologist found a moderately strong positive association between the number of hours of sleep a person gets and the person's ability to memorize information. 7) _____

a. Explain in the context of this problem what "positive association" means.

b. Hoping to improve academic performance, the psychologist recommended the school board allow students to take a nap prior to any assessment. Discuss the psychologist's recommendations.

8) A common objective for many school administrators is to increase the number of students taking SAT and ACT tests from their school. The data from each state from 2003 are reflected in the scatterplot. 8) _____



a. Write a few sentences describing the association.
b. Estimate the correlation. r = _____
c. If the point in the top left corner (4, 1215) were removed, would the correlation become stronger, weaker, or remain about the same? Explain briefly.
d. If the point in the very middle (38, 1049) were removed, would the correlation become stronger, weaker, or remain about the same? Explain briefly.

9) After conducting a marketing study to see what consumers thought about a new tinted contact lens they were developing, an eyewear company reported, "Consumer satisfaction is strongly correlated with eye color." Comment on this observation. 9) _____

10) On the axes below, sketch a scatterplot described: 10) _____
a. a strong negative association

b. a strong association but r is near 0

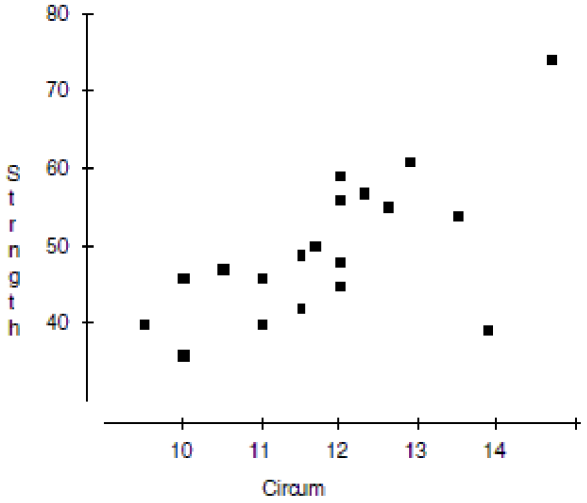
c. a weak but positive association

11) A school board study found a moderately strong negative association between the number of hours high school seniors worked at part-time jobs after school hours and the students' grade point averages. 11) _____

a. Explain in this context what "negative association" means.

b. Hoping to improve student performance, the school board passed a resolution urging parents to limit the number of hours students be allowed to work. Do you agree or disagree with the school board's reasoning. Explain.

12) Researchers investigating the association between the size and strength of muscles measured the forearm circumference (in inches) of 20 teenage boys. Then they measured the strength of the boys' grips (in pounds). Their data are plotted. 12) _____



a. Write a few sentences describing the association.

b. Estimate the correlation. r = _____

c. If the point in the lower right corner (at about 14" and 38 lbs.) were removed, how would the correlation become stronger, weaker, or remain about the same?

d. If the point in the upper right corner (at about 15" and 75 lbs.) were removed, would the correlation become stronger, weaker, or remain about the same?

13) One of your classmates is reading through the program for Friday night''s football game. Among other things, the program lists the players'' positions and their weights. Your classmate comments, "There is a strong correlation between a player's position and their weight." 13) _____

a. Explain why your classmate's statement is in error.

b. What other variable might be listed in the program that could be used to correctly identify a correlation with weight?

14) Match the following descriptions with the most likely correlation coefficient. 14) _____

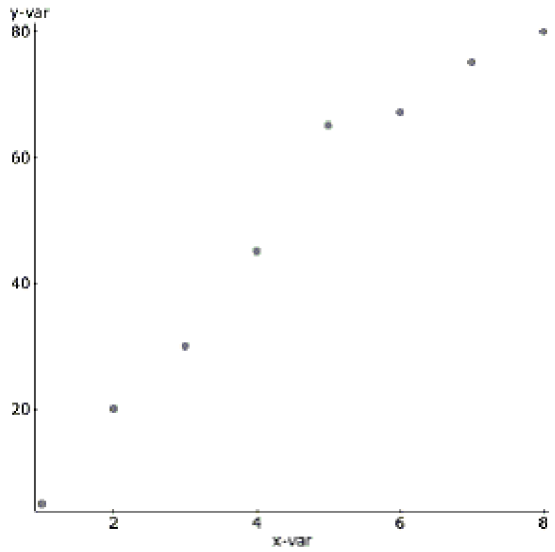
____ The number of hours you study and your exam score.
____ The number of siblings you have and your GPA.
____ The number of hours you practice a task and the number of minutes it takes you to complete it.
____ The number of hours you use a pencil and its length.

A. -0.78
B. 0.13
C. 0.46
D. 0.89

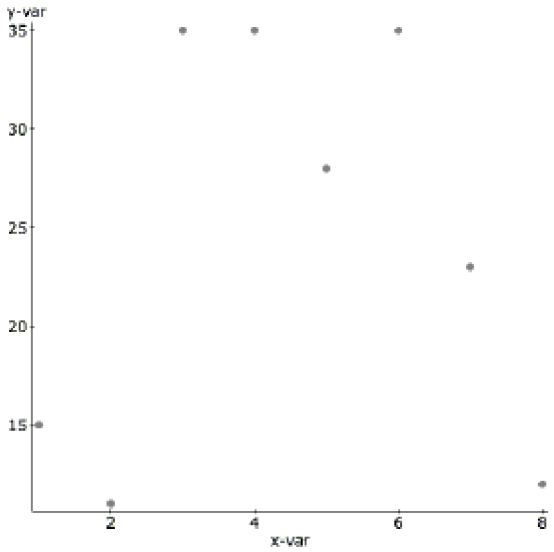
15) A researcher notes that there is a positive correlation between the temperature on a summer day and the number of bees that he can count in his garden over a 5-minute time span. 15) _____
a. Describe what the researcher means by a positive correlation.
b. If the researcher calculates the correlation coefficient using degrees Fahrenheit instead of Celsius, will the value be different?

16) Match each graph with the appropriate correlation coefficient. 16) _____
_____ 0.98 _____ 0.73 _____ 0.09 _____ -0.99

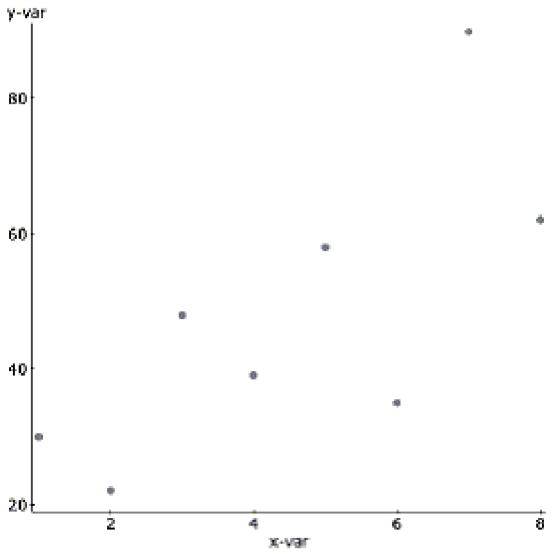
A.



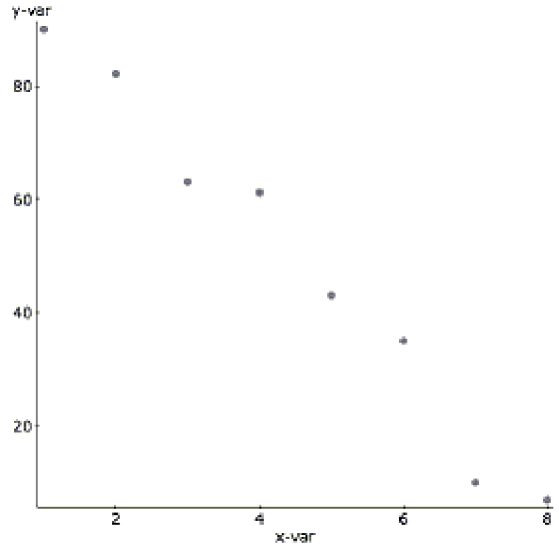
B.



C.



D.



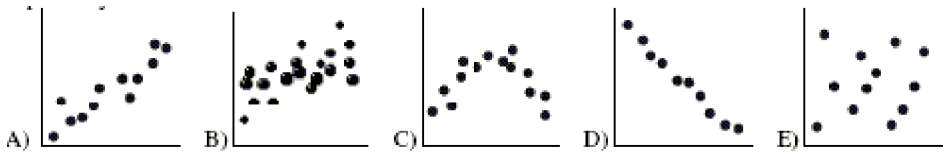
17) One your classmates is working on a science project for a unit on weather. She tracks the temperature one day, beginning at sunrise and finishing at sunset. Given that you are know for being the stats expert, she asks you about calculating the correlation for her data. What is the best advice you could give her? 17) _____

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

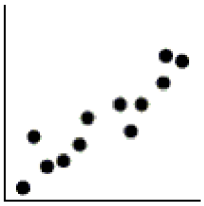
18) Researchers studying growth patterns of children collect data on the heights of fathers and sons. The correlation between the fathers' heights and the heights of their 16 year-old sons is most likely to be . . . 18) _____
A) near +1.0
B) near 0
C) near -1.0
D) near +0.7
E) somewhat greater than 1.0

19) The auto insurance industry crashed some test vehicles into a cement barrier at speeds of 5 to 25 mph to investigate the amount of damage to the cars. They found a correlation of $r = 0.60$ between speed (MPH) and damage ($). If the speed at which a car hit the barrier is 1.5 standard deviations above the mean speed, we expect the damage to be _?__ the mean damage. 19) _____
A) 0.90 SD above
B) 0.36 SD above
C) equal to
D) 1.5 SD above
E) 0.60 SD above

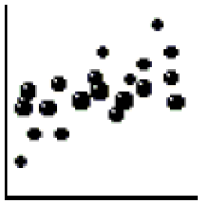
20) Which scatterplot shows a strong association between two variables even though the correlation is probably near zero? 20) _____



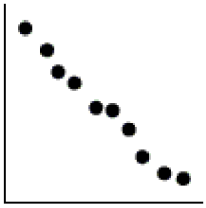
A)



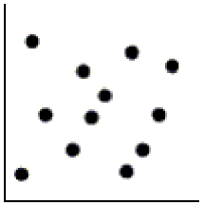
B)



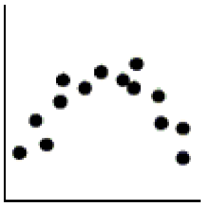
C)



D)



E)



21) The correlation between X and Y is $r = 0.35$. If we double each X value, decrease each Y by 0.20, and interchange the variables (put X on the Y-axis and vice versa), the new correlation 21) _____

A) is 0.70
B) cannot be determined.
C) is 0.35
D) is 0.50
E) is 0.90

22) A consumer group collected information on HDTVs . They created a linear model to estimate the cost of an HDTV (in $) based on the screen size (in inches). Which is the most likely value of the slope of the line of best fit? 22) _____

A) 700 B) 7 C) 0.70 D) 70 E) 7000

23) The correlation between a family's weekly income and the amount they spend on restaurant meals is found to be $r = 0.30$. Which must be true? 23) _____

I. Families tend to spend about 30% of their incomes in restaurants.
II. In general, the higher the income, the more the family spends in restaurants.
III. The line of best fit passes through 30% of the (*income*, *restaurant$*) data points.

A) II only
B) II and III only
C) III only
D) I, II, and III
E) I only

24) A medical researcher finds that the more overweight a person is, the higher his pulse rate tends to be. In fact, the model suggests that 12-pound differences in weight are associated with differences in pulse rate of 4 beats per minute. Which is true? 24) _____

I. The correlation between pulse rate and weight is 0.33
II. If you lose 6 pounds, your pulse rate will slow down 2 beats per minute.
III. A positive residual means a person's pulse rate is higher than the model predicts.

A) II only
B) I only
C) II and III only
D) none
E) III only

25) Education research consistently shows that students from wealthier families tend to have higher SAT scores. The slope of the line that predicts *SAT score from family income* is 6.25 points per $1000, and the correlation between the variables is 0.48. Then the slope of the line that predicts *family income from SAT score* (in $1000 per point) … 25) _____

A) is 6.25 B) is 0.037 C) is 3.00 D) is 13.02 E) is 0.16

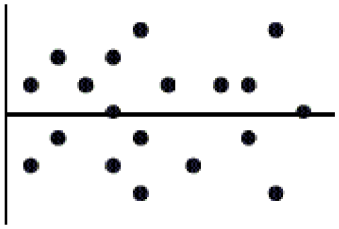
26) A regression analysis of company profits and the amount of money the company spent on advertising found $r^2 = 0.72$. Which of these is true? 26) _____

I. This model can correctly predict the profit for 72% of companies.
II. On average, about 72% of a company's profit results from advertising.
III. On average, companies spend about 72% of their profits on advertising.

A) none of these
B) II only
C) I and III
D) III only
E) I only

27) A least squares line of regression has been fitted to a scatterplot; the model's residuals plot is shown. 27) _____

Which is true?



A) The linear model is poor because the correlation is near 0.
B) The linear model is appropriate.
C) none of these
D) The linear model is poor because some residuals are large.
E) A curved model would be better.

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

28) **Earning power** A college's job placement office collected data about students' GPAs and the salaries they earned in their first jobs after graduation. The mean GPA was 2.9 with a standard deviation of 0.4. Starting salaries had a mean of $47,200 with a SD of $8500. The correlation between the two variables was $r = 0.72$. The association appeared to be linear in the scatterplot. (*Show work*) 28) _____

a. Write an equation of the model that can predict salary based on GPA.
b. Do you think these predictions will be reliable? Explain.
c. Your brother just graduated from that college with a GPA of 3.30. He tells you that based on this model the residual for his pay is -$1880. What salary is he earning?

29) **Assembly line** Your new job at *Panasony* is to do the final assembly of camcorders. As you learn how, you get faster. The company tells you that you will qualify for a raise if after 13 weeks your assembly time averages under 20 minutes. The data shows your average assembly time during each of your first 10 weeks. 29) _____

Week	Time(min)
1	43
2	39
3	35
4	33
5	32
6	30
7	30
8	28
9	26
10	25

a. Which is the explanatory variable?
b. What is the correlation between these variables?
c. You want to predict whether or not you will qualify for that raise. Would it be appropriate to use a linear model? Explain.

30) **Associations** For each pair of variables, indicate what association you expect: positive(+), negative(-), curved(C), or none(N). 30) _____

a. power level setting of a microwave; number of minutes it takes to boil water
b. number of days it rained in a month (during the summer); number of times you mowed your lawn that month
c. number of hours a person has been up past a normal bedtime; number of minutes it takes the person to do a crossword puzzle
d. number of hockey games played in Minnesota during a week; sales of suntan lotion in Minnesota during that week
e. length of a student's hair; number of credits the student earned last year

31) **Music and grades** (True Story) A couple of years ago, a local newspaper published research results claiming a positive association between the number of years high school children had taken instrumental music lessons and their performances in school (GPA). 31) _____

a. What does "positive association" mean in this context?
b. A group of parents then went to the School Board demanding more funding for music programs as a way to improve student chances for academic success in high school. As a statistician, do you agree or disagree with their reasoning? Explain briefly.

32) **Gas mileage again** In the *Data Desk* lab last week you analyzed the association between a car's fuel economy and its weight. Another important factor in the amount of gasoline a car uses is the size of the engine. Called "displacement", engine size measures the volume of the cylinders in cubic inches. The regression analysis is shown. 32) _____

Dependent variable is: **MPG**
89 total cases of which 0 are missing
R squared = 60.9% R squared (adjusted) = 60.0%
s = 3.056 with 89 - 2 = 87 degrees of freedom

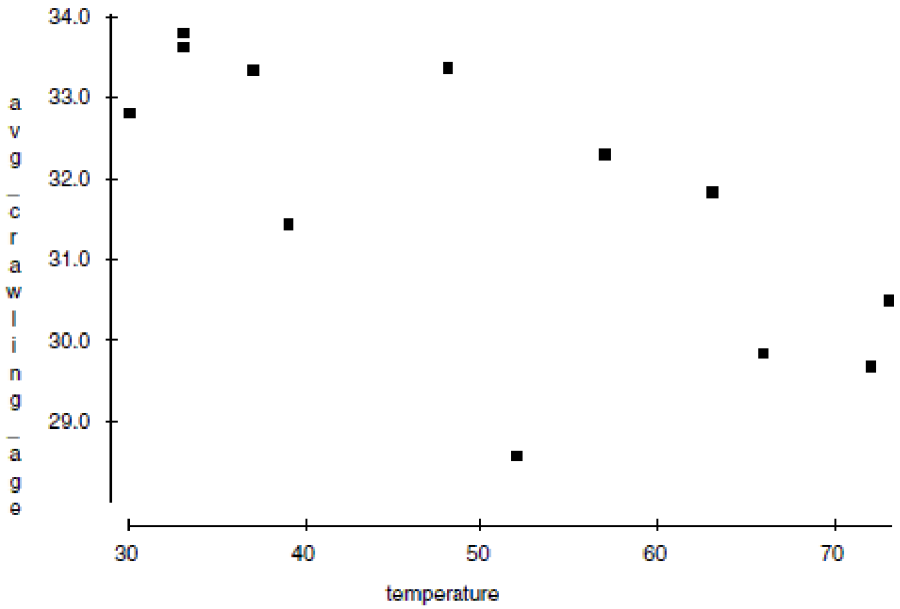
Source	Sum of Squares	df	Mean Square	F-ratio
Regression	696.744	1	696.744	74.6
Residual	448.236	48	9.33826	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	34.9799	1.231	28.4	Š 0.0001
Eng. Displcmt	-0.066196	0.0077	-8.64	Š 0.0001

a. How many cars were included in this analysis?
b. What is the correlation between engine size and fuel economy?
c. A car you are thinking of buying is available with two different size engines, 190 cubic inches or 240 cubic inches. How much difference might this make in your gas mileage? (*Show work*)

33) **Crawling** Researchers at the University of Denver Infant Study Center 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. The study sought an association between babies' first crawling age (in weeks) and the average temperature during the month they first try to crawl (about 6 months after birth). Between 1988 and 1991 parents reported the birth month and age at which their child was first able to creep or crawl a distance of four feet in one minute. Data were collected on 208 boys and 206 girls. The graph below plots average crawling ages (in weeks) against the mean temperatures when the babies were 6 months old. The researchers found a correlation of $r = -0.70$ and their line of best fit was 33) _____

$$\widehat{AvAge} = 36 - 0.08 AvTemp\ .$$



a. Draw the line of best fit on the graph. (Show your method clearly.)
b. Describe the association in context.
c. Explain (in context) what the slope of the line means.
d. Explain (in context) what the y-intercept of the line means.
e. Explain (in context) what R^2 means.
f. In this context, what does a negative residual indicate?

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

34) It takes a while for new factory workers to master a complex assembly process. During the first month new employees work, the company tracks the number of days they have been on the job and the length of time it takes them to complete an assembly. The correlation is most likely to be 34) _____

A) exactly -1.0
B) near +0.6
C) exactly +1.0
D) near -0.6
E) near 0

35) A lakeside restaurant found the correlation between the daily temperature and the number of meals they served to be 0.40. On a day when the temperature is two standard deviations above the mean, the number of meals they should plan on serving is _?_ the mean. 35) _____

A) equal to
B) 0.16 SD above
C) 0.4 SD above
D) 2.0 SD above
E) 0.8 SD above

36) For families who live in apartments the correlation between the family's income and the amount of rent they pay is $r = 0.60$. Which is true? 36) _____

I. In general, families with higher incomes pay more in rent.
II. On average, families spend 60% of their income on rent.
III. The regression line passes through 60% of the (*income$*, *rent$*) data points.

A) I and II only
B) I, II, and III
C) II only
D) I only
E) I and III only

37) A regression analysis of students' AP* Statistics test scores and the number of hours they spent doing homework found $r^2 = 0.32$. Which of these is true? 37) _____

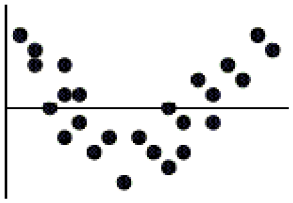
I. 32% of student test scores can be correctly predicted with this model.
II. Homework accounts for 32% of your grade in AP* Stats.
III. There's a 32% chance that you'll get the score this model predicts for you.

A) I only
B) III only
C) I and II
D) II only
E) none of these

38) Variables *X* and *Y* have $r = 0.40$. If we decrease each *X* value by 0.1, double each *Y* value, and then interchange them (put *X* on the *Y*-axis and vice versa) the new correlation will be 38) _____

A) -0.40 B) 0.15 C) 0.80 D) 0.40 E) 0.60

39) The residuals plot for a linear model is shown. Which is true? 39) _____



A) The linear model is okay because approximately the same number of points are above the line as below it.
B) The linear model is no good since the correlation is near 0.
C) The linear model is no good since some residuals are large.
D) The linear model is okay because the association between the two variables is fairly strong.
E) The linear model is no good because of the curve in the residuals.

40) A regression model examining the amount of weight a football player can bench press found that 10 cm differences in chest size are associated with 8 kg differences in weight pressed. Which is true? 40) _____

I. The correlation between chest size and weight pressed is $r = 0.80$
II. As a player gets stronger and presses more weight his chest will get bigger.
III. A positive residual means that the player pressed more than predicted.

A) none B) I and II C) III only D) I only E) I and III

41) Suppose we collect data hoping to be able to estimate the prices of commonly owned new cars (in $) from their lengths (in feet). Of these possibilities, the slope of the line of best fit is most likely to be 41) _____

A) 3 B) 300 C) 3000 D) 30 E) 30,000

42) Medical records indicate that people with more education tend to live longer; the correlation is 0.48. The slope of the linear model that predicts *lifespan* from *years of education* suggests that on average people tend to live 0.8 extra years for each additional year of education they have. The slope of the line that would predict *years of education* from *lifespan* is 42) _____

A) 0.288 B) 1.25 C) 1.67 D) 0.384 E) 0.8

43) This regression analysis examines the relationship between the number of years of formal education a person has and their annual income. According to this model, about how much more money do people who finish a 4-year college program earn each year, on average, than those with only a 2-year degree? 43) _____

Dependent variable is **Income**
R-squared = 25.8%
s = 3888 with 57 degrees of freedom

Variable	Coefficient	s.e. of Coeff
Constant	3984.45	6600
Education	2668.45	600.1

A) $2710 B) $7968 C) $9321 D) $2006 E) $5337

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

44) **Associations** For each pair of variables, indicate what association you expect: positive linear(**+**), negative linear(**-**), curved(**C**), or none(**N**). 44) _____

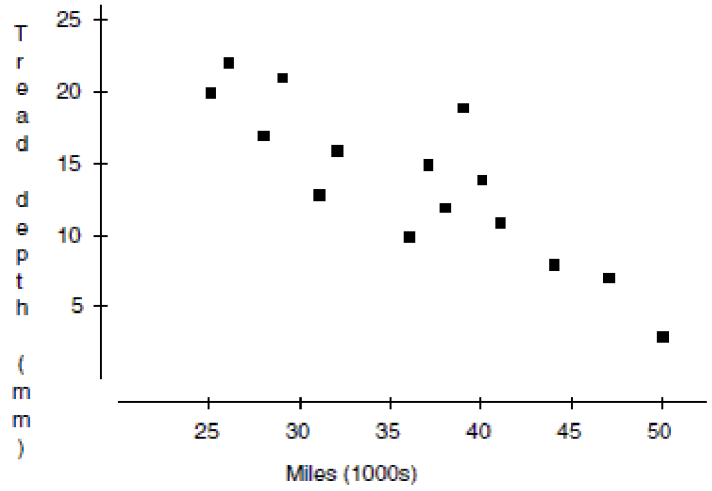
a. the number of miles a student lives from school; the student's GPA
b. a person's blood alcohol level; time it takes the person to solve a maze
c. weekly sales of hot chocolate at a Montana diner; the number of auto accidents that week in that town
d. the price charged for fund-raising candy bars; number of candy bars sold
e. the amount of rainfall during growing season; the crop yield (bushels per acre)

45) **Email** At CPU every student gets a college email address. Data collected by the college showed a negative association between student grades and the number of emails the student sent during the semester. 45) _____

a. Briefly explain what "negative association" means in this context.
b. After seeing this study the college proposes trying to improve academic performance by limiting the amount of email students can send through the college address. As a statistician, what do you think of this plan? Explain briefly.

46) **Car commercials** A car dealer investigated the association between the number of TV commercials he ran each week and the number of cars he sold the following weekend. He found the correlation to be $r = 0.56$. During the time he collected the data he ran an average of 12.4 commercials a week with a standard deviation of 1.8, and sold an average of 30.5 cars with a standard deviation of 4.2. Next weekend he is planning a sale, hoping to sell 40 cars. Create a linear model to estimate the number of commercials he should run this week. Write a sentence explaining your recommendation. 46) _____

47) **Taxi tires** A taxi company monitoring the safety of its cabs kept track of the number of miles tires had been driven (in thousands) and the depth of the tread remaining (in mm). Their data are displayed in the scatterplot. They found the equation of the least squares regression line to be $\hat{tread} = 36 - 0.6miles$, with $r^2 = 0.74$. 47) _____



a. Draw the line of best fit on the graph. (Show your method clearly.)
b. What is the explanatory variable?
c. The correlation $r =$ ____
d. Describe the association in context.
e. Explain (in context) what the slope of the line means.
f. Explain (in context) what the y - intercept of the line means.
g. Explain (in context) what R^2 means.
h. In this context, what does a negative residual mean?

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

48) A silly psychology student gathers data on the shoe size of 30 of his classmates and their GPA's. The correlation coefficient between these two variables is most likely to be 48) _____

A) exactly –1.0
B) near +0.6
C) exactly +1.0
D) near 0
E) near -0.6

49) researcher studied the relationship between family income and amount of money spent on an automobile. She calculated that R^2 = 45%. Which is the correct interpretaion? 49) _____

A) The car price fluctuates 45% more than income.
B) None of these
C) The probability of predicting the correct price of a car is 45%.
D) 45% of the variability in car price can be explained by using income.
E) 45% of the price of the car can be predicted by using income.

50) If r = -0.4 for the relationship between the time of day and amount of coffee in an office worker's mug, which are true? 50) _____

I. r^2 = -16%
II. There is a linear relationship between time and amount of coffee.
III. 16% of the variability is correctly predicted by time of day.

A) III
B) II and III only
C) I
D) II
E) none of these

51) The relationship between the longevity of an animal's life and its gestation time is 0.70. If an animal is one standard deviation below average in life expectancy, the gestation time is predicted to be __?__ below average. 51) _____

A) 1 SD
B) 0.49 SD
C) none of these
D) 0.7 SD
E) 1.4 SD

52) We can use the length of a man's hand span to predict his height, with a correlation coefficient of r = 0.60. If change our measurements from cm to m, the new correlation will be 52) _____

A) none of these
B) 0.006
C) 0.06
D) 6
E) 0.60

53) If a data set has a relationship that is best described by a linear model, than the residual plot will 53) _____

A) have no pattern with a correlation near 0.
B) none of these
C) also have a linear pattern with a similar correlation.
D) be an unknown shape.
E) have a curved pattern, like a parabola.

54) A regression model examining the amount of distance a long distance runner runs (in miles) to predict the amount of fluid the runner drinks (ounces) has a slope of 4.6. Which interpretation is appropriate? 54) _____

A) We predict 4.6 miles for every ounce that is drunk.
B) The correlation is needed to interpret this value.
C) Each mile adds 4.6 more ounces.
D) We predict for every mile run, the runner drinks 4.6 more ounces.
E) A runner drinks a minimum of 4.6 oz.

55) A regression equation is found that predicts the increased cost of a home owner's electricity bill given the number of holiday lights they put on the outside of their house. The equation is $\hat{dollars} = 2.5 + 0.02(light)$. If a house has 400 lights and a $15 increase in their electricity cost, find their residual. 55) _____

A) -$15 B) $5 C) $15 D) -$5 E) $20

56) Computer output in the scenario described in problem #8 reports that $s = 2.3$. Which is the correct interpretation of this value? 56) _____

A) The slope of the regression line is 2.3 lights per dollar.
B) The correlation is 2.3.
C) The average prediction error of the regression line is $2.30.
D) The initial cost, even with no lights is $2.30
E) The slope of the regression line is $2.30 per light.

57) Using the equation in number #8 again, if a homeowner doubles the number of lights he uses from 500 to 1000, how much do we predict he will increase his electric bill by? 57) _____

A) $2 B) $35 C) $12.50 D) $22.50 E) $10

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

58) **Associations** For each pair of variables, indicate what association you expect: positive linear(**+**), negative linear(**-**), curved(**C**), or none(**N**). 58) _____

a. the number of hours in the sun; the number of mold cultures on a piece of bread
b. the number of hours a store is open; the number of sales the store has
c. the number of hours you practice golf; your golf score
d. the price of gasoline; the number of families that take summer road trips
e. the size of a front lawn; the number of children who live in the house

59) **Put to Work** Some students have to work part time jobs to pay for college expenses. A researcher examined the academic performance of students with jobs versus those without. He found a positive association between the number of hours worked and GPA. Explain what "positive association" means in this context. 59) _____

60) **High Score** The longer you play a video game, the higher score you can usually achieve. An analysis of a popular game found the following relationship between the hours a player has played a game and their corresponding high score on that game. 60) _____

Dependent variable is **High Score**
R-squared = 76.5%
s = 383.3 with 89 degrees of freedom

Variable	Coefficient	s.e. of Coeff
Constant	524.8	145.3
Hours	2498.8	324.5

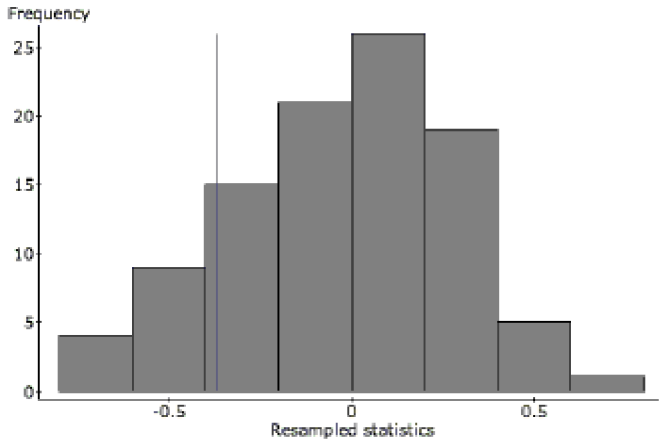
a. Write the regression equation and define the variables of your equation in context.
b. Interpret the slope in context.
c. Interpret the y-intercept in context.
d. Interpret *s* in context.
e. What is the correlation coefficient? Interpret this value in context.

61) **Time Wasted** A group of students decide to see if there is link between wasting time on the internet and GPA. They don't expect to find an extremely strong association, but they 're hoping for at least a weak relationship. Here are the findings. 61) _____

linear regression results: Dependent Variable: GPA Sample size: 10 R (correlation coefficient) = -0.37199274 R-sq = 0.1383786 s = 0.85365134		
Parameter	Estimate	Std. Err.
Intercept	4.06191	0.74405
Hours/week	-0.0297	0.02616

a. How strong is the relationship the students found? Describe in context with statistical justification.

One student is concerned that the relationship is so weak, there may not actually be any relationship at all. To test this concern, he runs a simulation where the 10 GPA's are randomly matched with the 10 hours/week. After each random assignment, the correlation is calculated. This process is repeated 100 times. Here is a histogram of the 100 correlations. The correlation coefficient of -0.371 is indicated with a vertical line.



b. Do the results of this simulation confirm the suspicion that there may not be any relationship? Refer specifically to the graph in your explanation.

An article in the *Journal of Statistics Education* reported the price of diamonds of different sizes in Singapore dollars (SGD). The following table contains a data set that is consistent with this data, adjusted to US dollars in 2004:

2004 US $	Carat	2004 US $	Carat	2004 US $	Carat
494.82	0.12	688.24	0.15	748.10	0.16
768.03	0.17	944.90	0.18	1076.18	0.19
1105.03	0.20	1071.75	0.21	1289.20	0.23
1508.88	0.25	1504.44	0.26	1597.63	0.27
1826.18	0.28	1908.28	0.29	2038.09	0.32
2096.89	0.33	2409.76	0.35		

62) Make a scatterplot and describe the association between the size of the diamond (carat) and the cost (in US dollars). 62) _____

63) Create a model to predict diamond costs from the size of the diamond. 63) _____

64) Do you think a linear model is appropriate here? Explain. 64) _____

65) Interpret the slope of your model in context. 65) _____

66) Interpret the intercept of your model in context. 66) _____

67) What is the correlation between cost and size? 67) _____

68) Explain the meaning of R^2 in the context of this problem. 68) _____

69) Would it be better for a customer buying a diamond to have a negative residual or a positive residual from this model? Explain. 69) _____

In an effort to decide if there is an association between the year of a postal increase and the new postal rate for first class mail, the data were gathered from the United States Postal Service. In 1981, the United States Postal Service changed their rates on March 22 and November 1. This information is shown in the table.

Year	Rate
1971	0.08
1974	0.10
1975	0.13
1978	0.15
1981	0.18
1981	0.20
1985	0.22
1988	0.25
1991	0.29
1995	0.32

70) Make a scatterplot and describe the association between the year and the first class postal rate. 70) _____

71) Create a model to predict postal rates from the year. 71) _____

72) Do you think a linear model is appropriate here? Explain. 72) _____

73) Interpret the slope of your model in context. 73) _____

74) Interpret the intercept of your model in context. 74) _____

75) What is the correlation between year and postal rate? 75) _____

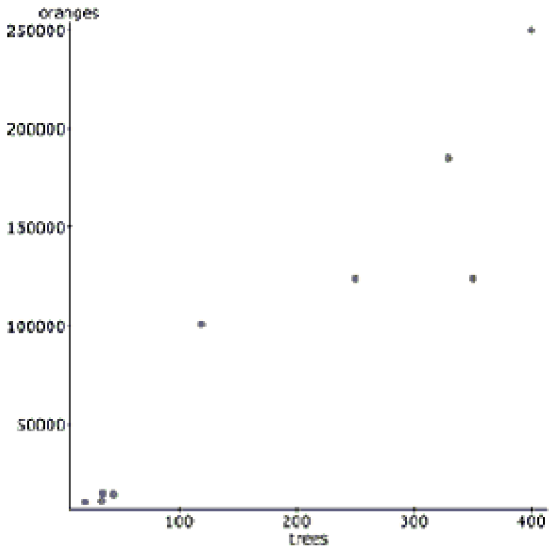
76) Explain the meaning of R^2 in the context of this problem. 76) _____

77) Would it be better for customers for a year to have a negative residual or a positive residual from this model? Explain. 77) _____

A study examined the number of trees in a variety of orange groves and the corresponding number of oranges that each grove produces in a given harvest year. Linear regression was calculated and the results are below.

linear regression results:
Dependent Variable: oranges
Independent Variable: trees
Sample size: 9
R-sq = 0.886
s = 31394.7

Parameter	Estimate	Std. Err.
Constant	390.59	16328.8
Trees	525.84	71.22



78) Write the regression equation. Define all variables used in your equation. 78) _____

79) Interpret the slope in context. 79) _____

80) Interpret *s* in context. 80) _____

81) Does the value of *s* concern you? How might you deal with this data differently to address this problem? 81) _____

82) Since r^2 is not 100%, there must be other factors in influencing the number of oranges harvested. What percentage is that and what is another factor you think might be involved? 82) _____

83) The farmer with 35 had 15,400 oranges; find the value of his residual. Show your work. 83) _____

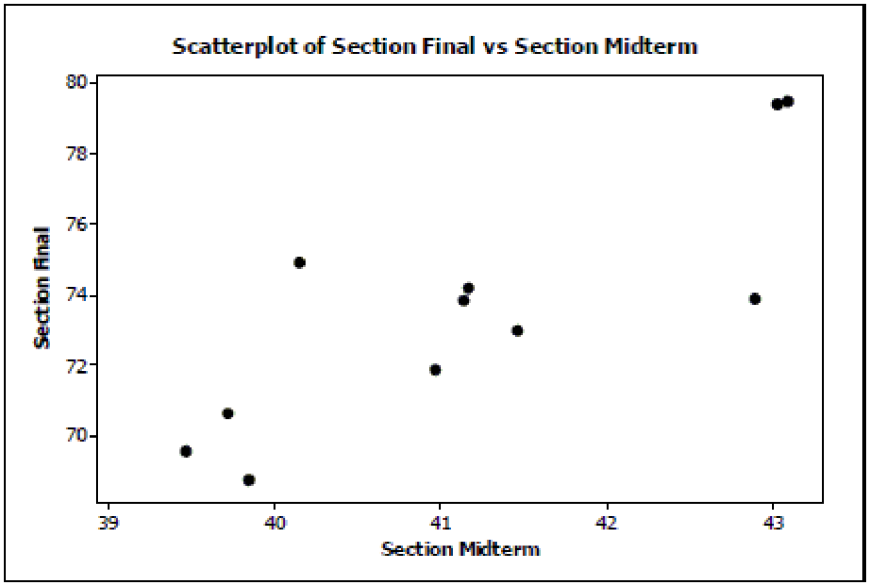
84) Is the farmer in problem #5 pleased or displeased with the value of his residual? Why? 84) _____

85) Find the value of the correlation coefficient and interpret this value in context. 85) _____

86) If these data were collected in California, would you feel confident in using this equation to make predictions about Florida orange groves also? Explain. 86) _____

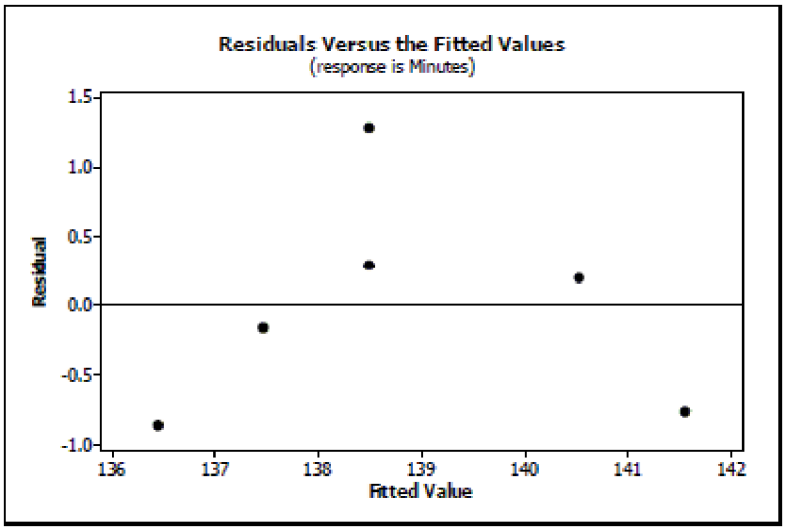
Solve the problem.

87) The following is a scatterplot of the average final exam score versus midterm score for 11 sections of an introductory statistics class: 87) _____



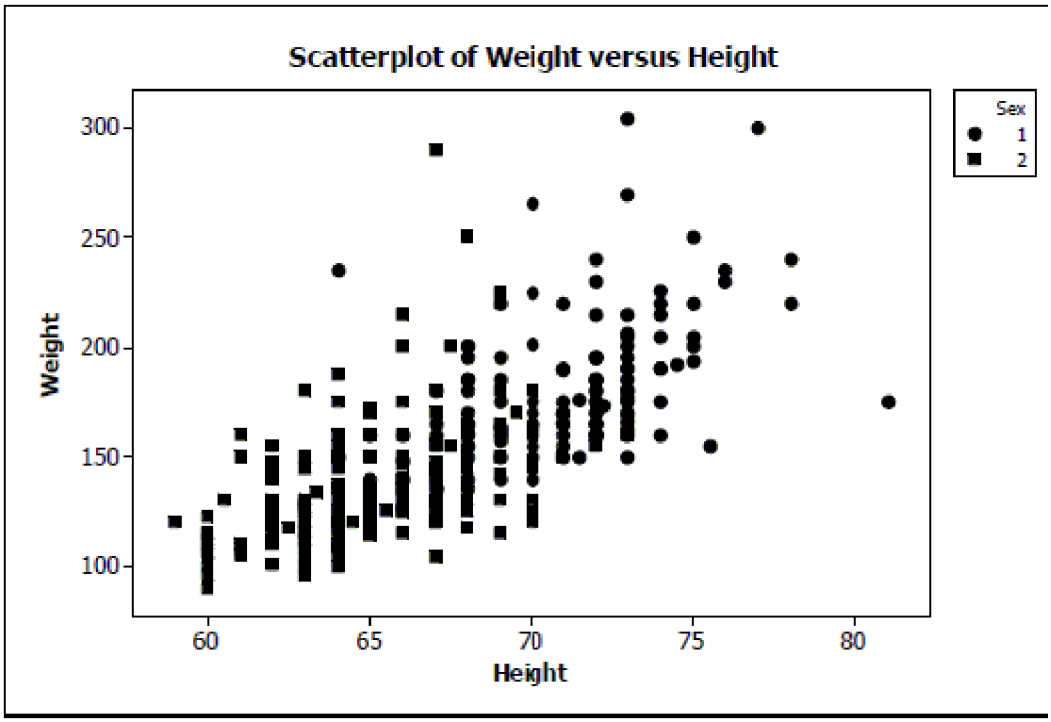
The correlation coefficient for these data is $r = 0.829$. If you had a scatterplot of the final exam score versus midterm score for all individual students in this introductory statistics course, would the correlation coefficient be weaker, stronger, or about the same? Explain.

88) A plot of the residuals versus the fitted values for record-breaking times of female marathon runners for the years 1998 - 2003 is: 88) _____



Based on this residuals plot, does it seem reasonable to use linear regression for this model? Explain.

89) Here is a scatterplot of weight versus height for students in an introductory statistics class. The men are coded as "1" and appear as circles in the scatterplot; the women are coded as "2" and appear as squares in the scatterplot. 89) _____

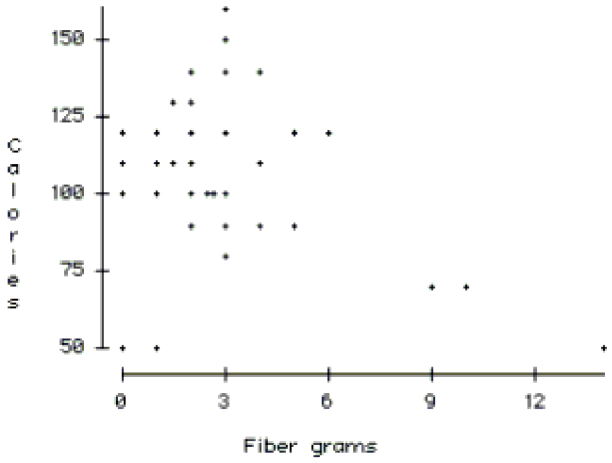


a. Do you think there is a clear pattern? Describe the association between weight and height.

b. Comment on any differences you see between men and women in the plot.

c. Do you think a linear model from the set of all data could accurately predict the weight of a student with height 70 inches? Explain.

Current research states that a good diet should contain 20-35 grams of dietary fiber. Research also states that each day should start with a healthy breakfast. The nutritional information for 77 breakfast cereals was reviewed to find the grams of fiber and the number of calories per serving. The scatterplot below shows the relationship between fiber and calories for the cereals.

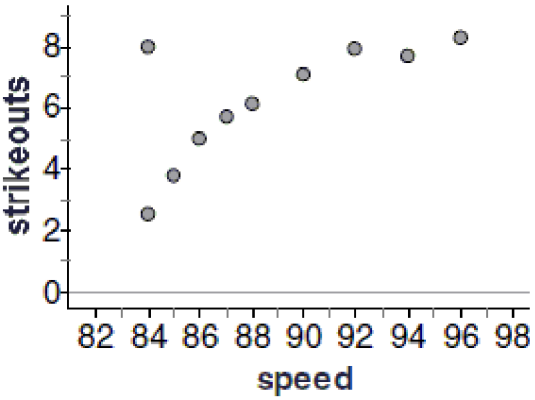


90) Do you think there is a clear pattern? Describe the association between fiber and calories. 90) _____

91) Comment on any unusual data point or points in the data set. Explain. 91) _____

92) Do you think a model could accurately predict the number of calories in a serving of cereal that has 22 grams of fiber? Explain. 92) _____

Baseball coaches use a radar gun to measure the speed of pitcher's fastball. They also record outcomes such as hits and strikeouts. The scatterplot below shows the relationship between the average speed of a fastball and the average number of strikeouts per nine innings for each pitcher on the Bulldogs, based on the past season.



93) Do you think there is a pattern? Describe the association between speed and the number of strikeouts. 93) _____

94) Comment on any unusual data point or points in the data set. Explain. 94) _____

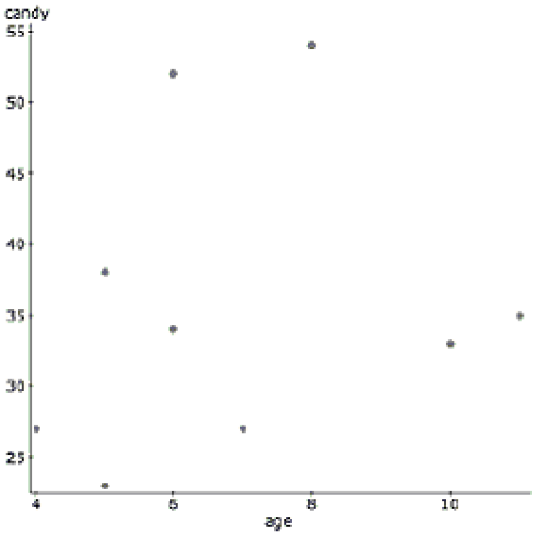
95) Do you think the association would be stronger or weaker if we used data from one month of the season? 95) _____

96) Do you think a model based on these data could accurately predict the average number of strikeouts for a pitcher with an average fastball speed of 70 mph.? Explain. 96) _____

Halloween is a fun night. It seems that older children might get more candy because they can travel further while trick-or-treating. But perhaps the youngest kids get extra candy because they are so cute. Here are some data that examine this question, along with the regression output.

Dependent Variable: candy
Sample size: 9
R (correlation coefficient) = 0.19534425
R-sq = 0.038159375
s = 11.297554

Parameter	Estimate	Std. Err.
Intercept	13.569231	9.0783516
Age	3.4038462	1.0175376

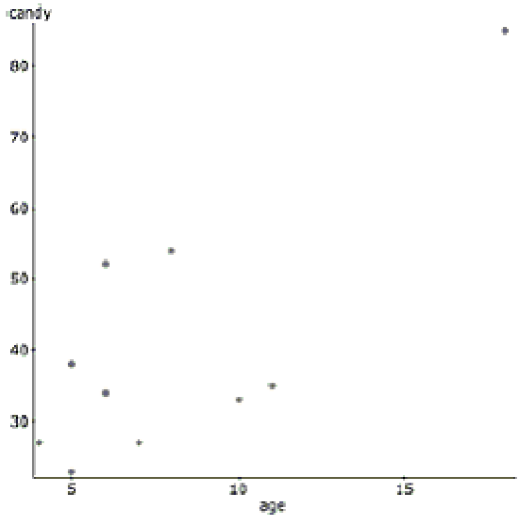


97) Based on the graph and the regression output, what conclusions do you draw regarding the relationship between age and the number of pieces of candy a trick-or-treater collects? 97) _____

98) The next day, a young girl reveals that her older brother also went trick-or-treating, but didn't want to admit that he participated. He was added to the data set and these are the results. 98) _____

Dependent Variable: candy
Sample size: 10
R (correlation coefficient) = 0.76362369
R-sq = 0.58312115
s = 12.709041

Parameter	Estimate	Std. Err.
Intercept	13.569231	9.0783516
Age	3.4038462	1.0175376



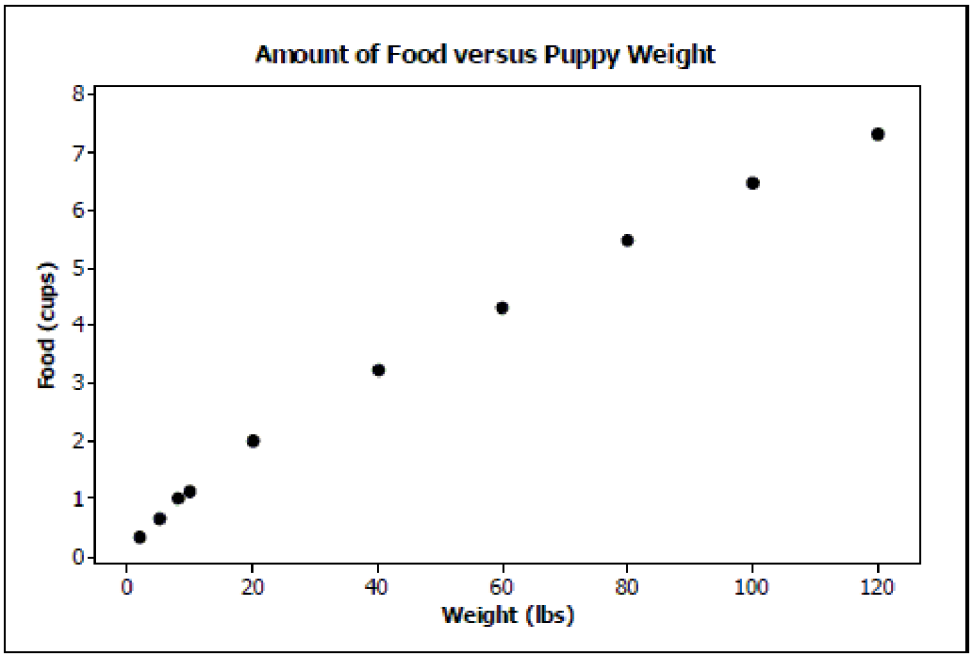
Describe the effect of this new candy collector on the regression model.

Solve the problem.

99) If you have a puppy who you are feeding Hill's Science Diet® original puppy food, the feeding guidelines for puppies who are 10 to 12 months old are as follows (Source: Hill's feeding guidelines): 99) _____

Weight (lbs)	2	5	8	10	20	40	60	80	100	120
Food (cups)	1/3	2/3	2	1-1/8	2	3-1/4	4-1/3	5.5	6.5	7-1/3

A scatterplot of the data is:



Does it seem reasonable to perform a linear regression to predict amount of food from the puppy's weight based on this data set? Explain.

100) You are given the following costs to build a square deck for your house: 100) _____

Width (ft)	4	5	6	7	8	9	10	11	12	13	14	15
Cost ($)	150	255	350	500	650	800	1000	1200	1450	1700	1950	2250

a. Use re-expressed data to create a model that predicts the cost of the deck based on the width.
b. Why do you think that your model is appropriate?
c. Find the predicted cost of a square deck that is 10.5 feet wide.
d. Is it reasonable to use this model to predict the cost of a square deck that is 20 feet wide? Explain.

101) The average movie ticket prices in selected years since 1948 are listed in the table below. 101) _____

Year	Movie Ticket Price
1948	$0.36
1954	$0.49
1958	$0.68
1963	$0.86
1967	$1.22
1971	$1.65
1974	$1.89
1974	$2.03
1976	$2.13
1977	$2.23
1978	$2.34
1979	$2.47
1980	$2.69
1981	$2.78
1982	$2.94
1983	$3.15
1984	$3.36
1985	$3.55
1986	$3.71
1987	$3.91
1988	$4.11

a. Use re-expressed data to create a model that predicts ticket prices. (Hint: scale the year)
b. Find the movie ticket price this model predicts for 2004.

102) During a chemistry lab, students were asked to study a radioactive element which decays over time. The results are in the table. 102) _____

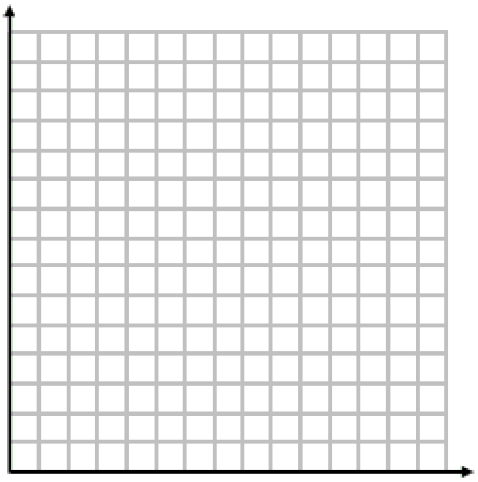
Time (in days)	0	2	4	6	8	10
Element (in grams)	320	226	160	115	80	57

a. Model the remaining mass of the element.
b. Find the predicted amount of the element remaining after thirty minutes.

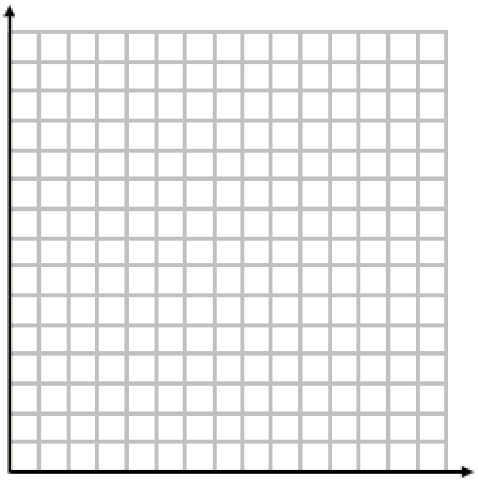
During a science lab, students heated water, allowed it to cool, and recorded the temperature over time. They computed the difference between the water temperature and the room temperature. The results are in the table.

Time (in minutes)	10	20	30	40	50	60
Difference in temp. (degrees F)	68	36	20	10	6	4

103) Sketch a scatterplot. 103) _____



104) Newton's Law of Cooling suggests an exponential function is appropriate. Re-express the data using logarithms and sketch a new scatterplot. 104) _____

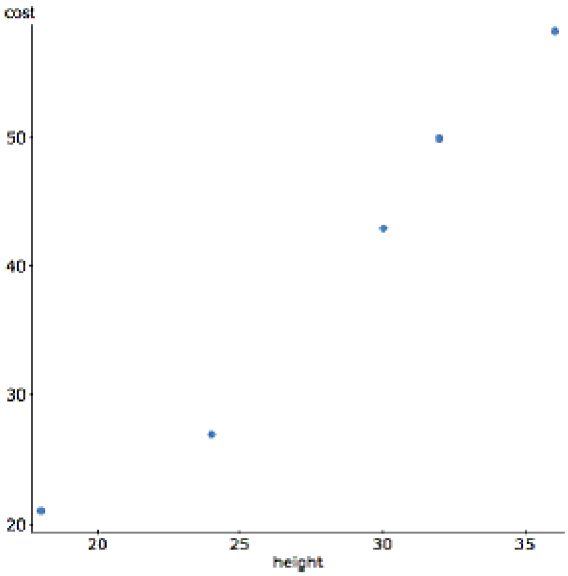


105) Write the equation of the least-squares regression line for the transformed data. Draw the regression line on the scatterplot in question 2. 105) _____

106) Use the equation $\log(\widehat{difftemp}) = 2.057 - 0.025 time$ to predict the difference in temperature after 45 minutes. 106) _____

107) Use the equation $\log(\widehat{difftemp}) = 2.057 - 0.025 time$ to predict the difference in temperature at time 0 minutes. What does this value represent? 107) _____

The bigger the stop sign, the more expensive it is. Here is a graph of the height of a sign in inches versus its cost in dollars.

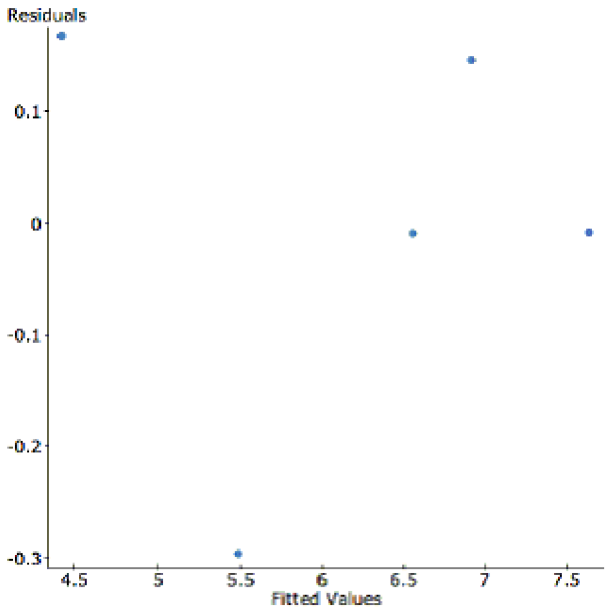


108) Describe why performing linear regression with these data is not a good decision. 108) _____

To achieve linearity, the data was transformed using a square root function of cost. Here are the results and a residual plot.

Dependent Variable: sqrt(cost)
R (correlation coefficient) = 0.98946627
R-sq = 0.97904349
s: 0.2141

Parameter	coeff	se
Intercept	1.1857	0.4346
height	0.1792	0.0151



109) Do you think this transformation for linearity was successful? Why? 109) _____

110) Write the transformed regression equation. Make sure to define any variables used in your equation. 110) _____

111) Interpret R-sq in the context of this problem. 111) _____

112) Use your equation to predict the cost of a 48" stop sign. 112) _____

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

113) All but one of these statements contain a mistake. Which could be true? 113) _____

A) The correlation between the amount of fertilizer used and the yield of beans is 0.42.
B) The correlation between a football player's weight and the position he plays is 0.54.
C) There is a correlation of 0.63 between gender and political party.
D) The correlation between a car's length and its fuel efficiency is 0.71 miles per gallon.
E) There is a high correlation (1.09) between height of a corn stalk and its age in weeks.

114) Residuals are . . . 114) _____

A) the difference between observed responses and values predicted by the model.
B) data collected from individuals that is not consistent with the rest of the group.
C) none of these
D) variation in the data that is explained by the model.
E) possible models not explored by the researcher.

115) Which statement about influential points is true? 115) _____

I. Removal of an influential point changes the regression line.
II. Data points that are outliers in the horizontal direction are more likely to be influential than points that are outliers in the vertical direction.
III. Influential points have large residuals.

A) II and III
B) I and III
C) I, II, and III
D) I and II
E) I only

116) Which is true? 116) _____

I. Random scatter in the residuals indicates a model with high predictive power.
II. If two variables are very strongly associated, then the correlation between them will be near +1.0 or -1.0.
III. The higher the correlation between two variables the more likely the association is based in cause and effect.

A) II only
B) I, II, and III
C) none
D) I only
E) I and II

117) A company's sales increase by the same amount each year. This growth is . . . 117) _____

A) linear
B) quadratic
C) power
D) logarithmic
E) exponential

118) Another company's sales increase by the same percent each year. This growth is . . . 118) _____

A) logarithmic
B) quadratic
C) exponential
D) power
E) linear

119) A scatterplot of $\frac{1}{\sqrt{y}}$ vs. x shows a strong positive linear pattern. It is probably true that 119) _____

A) large values of X are associated with large values of Y.
B) the residuals plot for regression of Y on X shows a curved pattern.
C) the correlation between X and Y is near +1.0.
D) the scatterplot of Y vs X also shows a linear pattern.
E) accurate predictions can be made for Y even if extrapolation is involved.

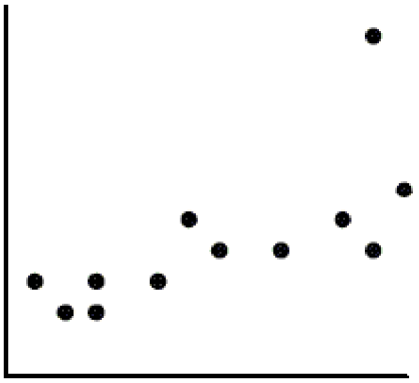
120) It's easy to measure the circumference of a tree's trunk, but not so easy to measure its height. Foresters developed a model for ponderosa pines that they use to predict the tree's height (in feet) from the circumference of its trunk (in inches): $\ln \hat{h} = -1.2 + 1.4(\ln C)$. A lumberjack finds a tree with a circumference of 60"; how tall does this model estimate the tree to be? 120) _____

A) 83' B) 11' C) 93' D) 5' E) 19'

121) Two variables that are actually not related to each other may nonetheless have a very high correlation because they both result from some other, possibly hidden, factor. This is an example of 121) _____

A) regression.
B) an outlier.
C) a lurking variable.
D) extrapolation.
E) leverage.

122) If the point in the upper right corner of this scatterplot is removed from the data set, then what will happen to the slope of the line of best fit (*b*) and to the correlation (*r*)? 122) _____



A) both will increase.
B) both will decrease.
C) *b* will decrease, and *r* will increase.
D) both will remain the same.
E) *b* will increase, and *r* will decrease.

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

123) **Breaking strength** A company manufactures polypropylene rope in six different sizes. To assess the strength of the ropes they test two samples of each size to see how much force (in kilograms) the ropes will hold without breaking. The table shows the results of the tests. We want to create a model for predicting the breaking strength from the diameter of the rope. 123) _____

Diameter (mm)	Strength (kg)	
4	60	76
7	157	153
10	254	262
12	334	388
15	551	529
20	938	893

a. Find a model that uses re-expressed data to straighten the scatterplot.
b. The company is thinking of introducing a new 25mm rope. How strong should it be? (Write a sentence in context based on one of your models.)

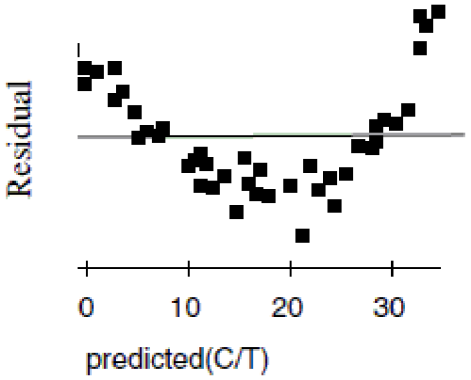
124) **Math and Verbal** Suppose the correlation between SAT Verbal scores and Math scores is 0.57 and that these scores are normally distributed. If a student's Verbal score places her at the 90th percentile, at what percentile would you predict her Math score to be? (SHOW WORK) 124) _____

125) **Penicillin** Doctors studying how the human body assimilates medication inject some patients with penicillin, and then monitor the concentration of the drug (in units/cc) in the patients' blood for seven hours. The data are shown in the scatterplot. First they tried to fit a linear model. The regression analysis and residuals plot are shown. 125) _____

Dependent variable is: **Concentration**
No Selector
R squared = 90.8% R squared (adjusted) = 90.6%
s = 3.472 with 43 - 2 = 41 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	4900.55	1	4900.55	407
Residual	494.199	41	12.0536	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	40.3266	1.295	31.1	≤ 0.0001
Time	-5.95956	0.2956	-20.2	≤ 0.0001



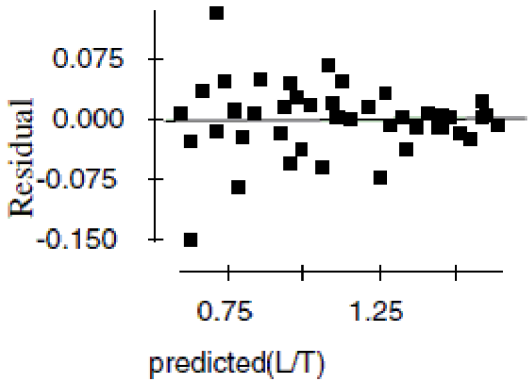
a. Find the correlation between time and concentration.
b. Using this model, estimate what the concentration of penicillin will be after 4 hours.
c. Is that estimate likely to be accurate, too low, or too high? Explain.

Now the researchers try a new model, using the re-expression log(*Concentration*). Examine the regression analysis and the residuals plot below.

Dependent variable is: **LogCnn**
No Selector
R squared = 98.0% R squared (adjusted) = 98.0%
s = 0.0451 with 43 - 2 = 41 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	4.11395	1	4.11395	2022
Residual	0.083412	41	0.002034	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	1.80184	0.0168	107	≤ 0.0001
Time	-0.172672	0.0038	-45.0	≤ 0.0001



d. Explain why you think this model is better than the original linear model.

e. Using this new model, estimate the concentration of penicillin after 4 hours.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

126) All but one of the statements below contain a mistake. Which one could be true? 126) _____

A) The correlation between weight and length of foot is 0.488.
B) The correlation between the breed of a dog and its weight is 0.435.
C) The correlation between gender and age is -0.171.
D) If the correlation between blood alcohol level and reaction time is 0.73, then the correlation between reaction time and blood alcohol level is -0.73.
E) The correlation between height and weight is 0.568 inches per pound.

127) A correlation of zero between two quantitative variables means that 127) _____

A) there is no linear association between the two variables.
B) we have done something wrong in our calculation of *r*.
C) there is no association between the two variables.
D) re-expressing the data will guarantee a linear association between the two variables.
E) none of these

128) A residuals plot is useful because 128) _____

I. it will help us to see whether our model is appropriate.
II. it might show a pattern in the data that was hard to see in the original scatterplot.
III. it will clearly identify influential points.

A) I, II, and III
B) I only
C) I and II only
D) II only
E) I and III only

129) Which of the following is not a goal of re-expressing data? 129) _____

A) Make the scatter in a scatterplot spread out evenly rather than following a fan shape.
B) Make the spread of several groups more alike.
C) Make the form of a scatterplot more nearly linear.
D) Make the distribution of a variable more symmetric.
E) All of these are goals of re-expressing data.

130) The correlation coefficient between the hours that a person is awake during a 24-hour period and the hours that same person is asleep during a 24-hour period is most likely to be 130) _____
A) exactly +1.0
B) near -0.8
C) near 0
D) exactly -1.0
E) near +0.8

131) The correlation coefficient between high school grade point average (GPA) and college GPA is 0.560. For a student with a high school GPA that is 2.5 standard deviations above the mean, we would expect that student to have a college GPA that is _____ the mean. 131) _____
A) 0.56 SD above B) 2.5 SD above C) equal to D) 1.4 SD above

132) A regression analysis of students' college grade point averages (GPAs) and their high school GPAs found $R^2 = 0.311$. Which of these is true? 132) _____
I. High school GPA accounts for 31.1% of college GPA.
II. 31.1% of college GPAs can be correctly predicted with this model.
III. 31.1% of the variance in college GPA can be accounted for by the model
A) I only
B) I and II
C) none of these
D) II only
E) III only

133) Although there are annual ups and downs, over the long run, growth in the stock market averages about 9% per year. A model that best describes the value of a stock portfolio is probably: 133) _____
A) linear
B) power
C) logarithmic
D) exponential
E) quadratic

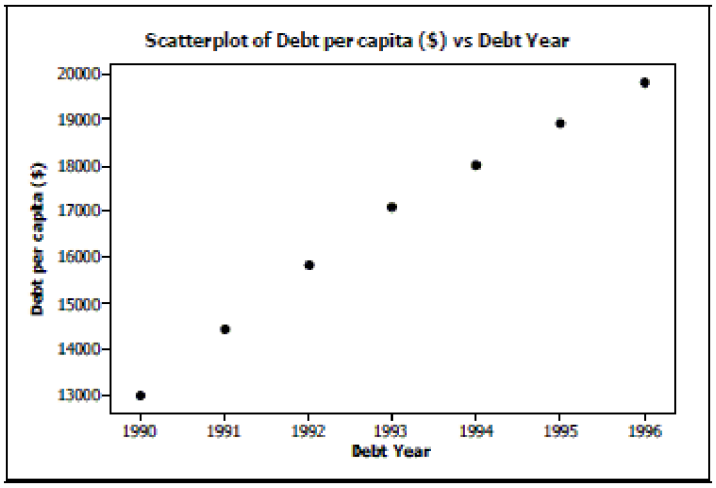
134) When using midterm exam scores to predict a student's final grade in a class, the student would prefer to have a 134) _____
A) positive residual, because that means the student's final grade is higher than we would predict with the model.
B) residual equal to zero, because that means the student's final grade is exactly what we would predict with the model.
C) positive residual, because that means the student's final grade is lower than we would predict with the model.
D) negative residual, because that means the student's final grade is lower than we would predict with the model.
E) negative residual, because that means the students final grade is higher than we would predict with the model.

135) The model $\sqrt{\hat{distance}} = 3.30 + 0.235(speed)$ can be used to predict the stopping distance (in feet) for a car traveling at a specific speed (in mph). According to this model, about how much distance will a car going 65 mph need to stop? 135) _____
A) 18.6 feet B) 4.3 feet C) 729.0 feet D) 345.0 feet E) 27.0 feet

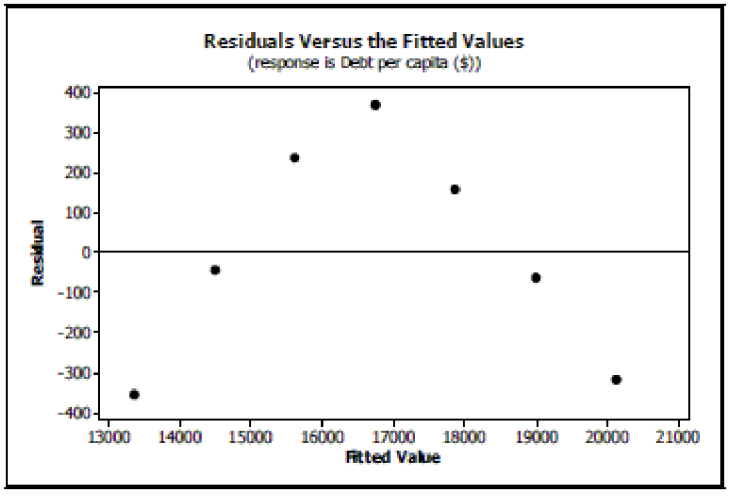
SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

136) **Storks** Data show that there is a positive association between the population of 17 European countries and the number of stork pairs in those countries. 136) _____
a. Briefly explain what "positive association" means in this context.
b. Wildlife advocates want the stork population to grow, and jokingly suggest that citizens should be encouraged to have children. As a statistician, what do you think of this suggestion? Explain briefly.

137) **Personal debt** According to *The World Almanac and Book of Facts 2004*, the debt per capita for the years 1990-2001 gives the following scatterplot: 137) _____



Regression output gives the equation of the regression line as
$\hat{Debt} = -2{,}231{,}226 + 1128(Year)$ with $R^2 = 98.8\%$.
a. What is the response variable?
b. What is the correlation coefficient r?
c. Explain in context what the slope of the line means.
d. Explain in context what $R^2 = 98.8\%$ means.
e. You decide to take a look at a residuals plot before making any predictions. Based on the following residuals plot, does linear regression seem appropriate for these data? Explain.



138) **Studying for exams** A philosophy professor has found a correlation of 0.80 between the number of hours students study for his exams and their exam performance. During the time he collected the data, students studied an average of 10 hours with a standard deviation of 2.5 hours, and scored an average of 80 points with a standard deviation of 7.5 points. 138) _____

a. Create a linear model to estimate the number of points a student will score on the next exam from the number of hours the student studies.

b. If a student studies for 15 hours, what score should the student expect on the next exam? Show your work.

139) **Height and weight** Suppose that both height and weight of adult men can be described with Normal models, and that the correlation between these variables is 0.65. If a man's height places him at the 60th percentile, at what percentile would you expect his weight to be? 139) _____

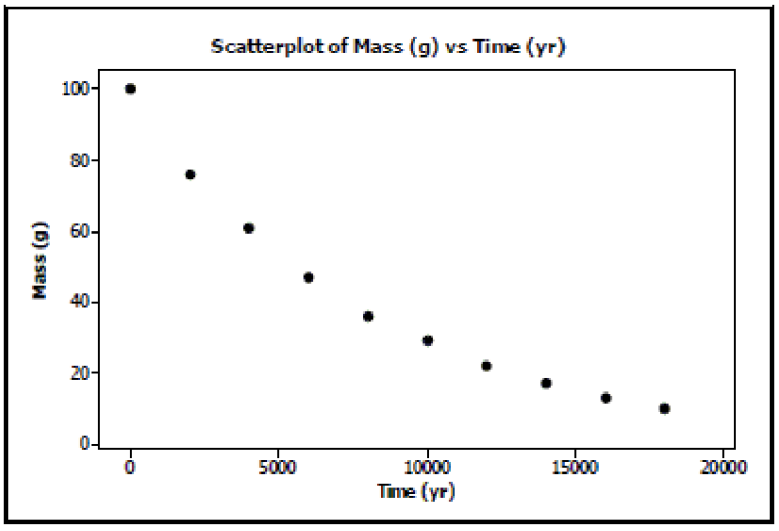
140) **Carbon dating** QuarkNet, a project funded by the National Science Foundation and the U.S. Department of Energy, poses the following problem on its website: 140) _____

"Last year, deep within the Soudan mine, QuarkNet teachers began a long-term experiment to measure the amount of carbon-14 remaining in an initial 100-gram sample at 2000-year intervals. The experiment will be complete in the year 32001. Fortunately, a method for sending information backwards in time will be discovered in the year 29998, so, although the experiment is far from over, the results are in."

Here is a portion of the data:

Time (yr)	0	2000	4000	6000	8000	10,000	12,000	14,000	16,000	18,(
Mass (g)	100	76	61	47	36	29	22	17	13	1(

A scatterplot of these data looks like:



a. Straighten the scatterplot by re-expressing these data and create an appropriate model for predicting the mass from the year.

b. Use your model to estimate what the mass will be after 7500 years.

c. Can you use your model to predict when 50 g of the sample will be left? Explain.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

141) All but one of the statements below contain a mistake. Which one could be true? 141) _____
A) The correlation between the species of tree and its height is $r = 0.56$.
B) The correlation between age and weight of a newborn baby is $r = 0.83$ oz per day.
C) The correlation between blood alcohol level and reaction time is $r = 0.73$.
D) There is a high correlation between cigarette smoking and gender.
E) The correlation between a person's age and vision (20/20?) is $r = -1.04$.

142) Which statement about correlation is true? 142) _____
I. Regression based on data that are summary statistics tends to result in a higher correlation.
II. If $r^2 = 0.95$, the response variable increases as the explanatory variable increases.
III. An outlier always decreases the correlation.
A) I only
B) none of these
C) I, II, and III
D) II only
E) III only

143) Which statement about residuals plots is true? 143) _____
I. A curved pattern indicates nonlinear association between the variables.
II. A pattern of increasing spread indicates the predicted values become less reliable as the explanatory variable increases.
III. Randomness in the residuals indicates the model will predict accurately.
A) I, II, and III
B) I only
C) II only
D) I and II only
E) I and III only

144) Which of the following is <u>not</u> a source of caution in regression between two variables? 144) _____
A) an outlier
B) All of these are potential problems.
C) subgroups with differences
D) a lurking variable
E) extrapolation

145) Which statement about re-expressing data is <u>not</u> true? 145) _____
I. Unimodal distributions that are skewed to the left will be made more symmetric by taking the square root of the variable.
II. A curve in which the direction of the association changes from negative to positive will not benefit from re-expression.
III. One goal of re-expression may be to make the variability of the response variable more uniform.
A) II and III
B) I, II, and III
C) III only
D) II only
E) I only

146) Over the past decade a farmer has been able to increase his wheat production by about the same number of bushels each year. His most useful predictive model is probably… 146) _____

A) exponential
B) power
C) linear
D) logarithmic
E) quadratic

147) Another farmer has increased his wheat production by about the same percentage each year. His most useful predictive model is probably… 147) _____

A) linear
B) power
C) logarithmic
D) exponential
E) quadratic

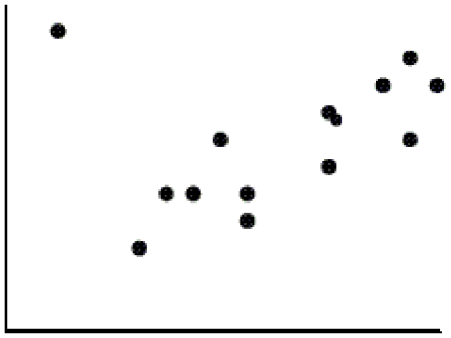
148) The model $\sqrt{\hat{str}} = 12 + 20dia$ can be used to predict the breaking strength of a rope (in pounds) from its diameter (in inches). According to this model, how much force should a rope one-half inch in diameter be able to withstand? 148) _____

A) 4.7 lbs B) 484 lbs C) 22 lbs D) 256 lbs E) 16 lbs

149) A scatterplot of log(Y) vs. log(X) reveals a linear pattern with very little scatter. It is probably true that … 149) _____

A) the correlation between *X* and *Y* is near +1.
B) the residuals plot for regression of *Y* on *X* shows a curved pattern.
C) the scatterplot of *Y* vs *X* shows a linear association.
D) the calculator's LnReg function will model the association between *X* and *Y*.
E) the correlation between *X* and *Y* is near 0.

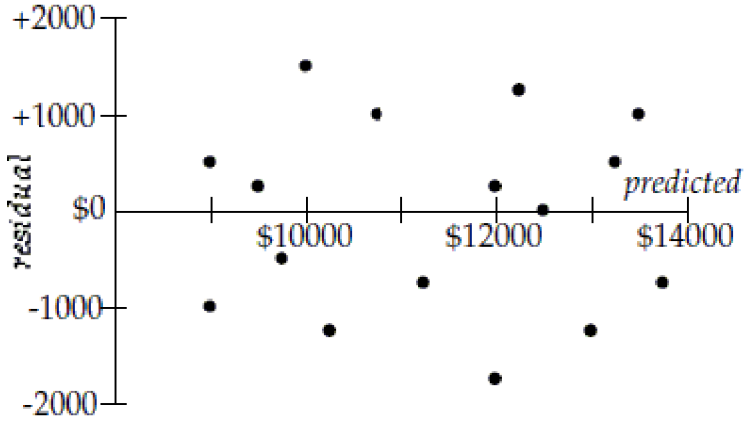
150) If the point in the upper left corner of the scatterplot is removed, what will happen to the correlation (*r*) and the slope of the line of best fit (*b*)? 150) _____



A) *r* will increase and *b* will decrease.
B) Both will decrease.
C) *r* will decrease and *b* will increase.
D) Both will increase.
E) They will not change.

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

151) **Subaru costs** Data collected from internet ads for 1999 Subarus were used to create a model to estimate the asking price of the car based on the number of miles it had been driven. The model has $r^2 = 0.47$ and equation $\hat{Price} = 15{,}327 - 0.11(Miles)$. The plot of residuals versus the predicted price is shown. 151) _____



a. Do you think you could make accurate estimates of Subaru prices with this model? Explain.
b. Interpret the slope of the line.
c. One of the cars in the data set had been driven 42,000 miles. How much was the owner asking for it? (Show work.)

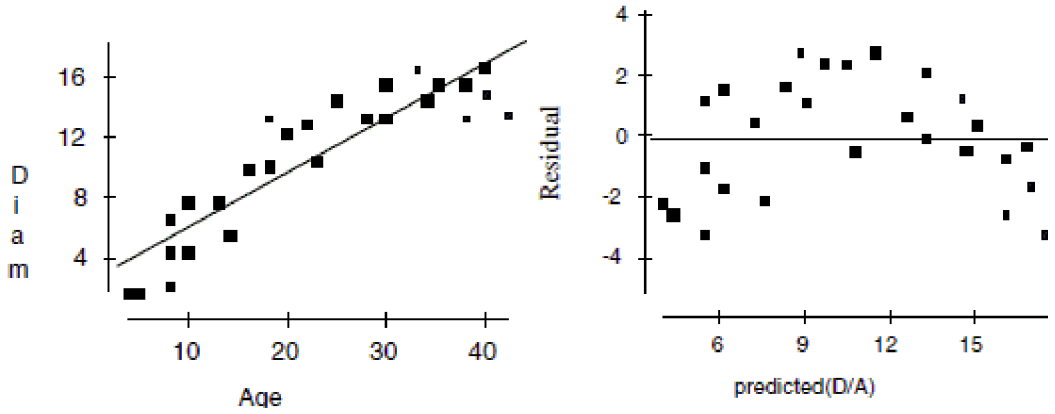
152) **Penicillin assimilation** Doctors studying how the human body assimilates medication inject a patient with penicillin, and then monitor the concentration of the drug in the patient's blood for several hours. The data are shown in the table. 152) _____

Time elapsed (Hours)	Concentration (Units/cc)
1	42
2	28
3	19
4	13
5	9
6	6
7	4

a. Straighten the scatterplot by re-expressing these data and create an appropriate model for predicting the concentration of penicillin.
b. Use your model to estimate what the concentration of penicillin will be after 8 hours.

153) **Blood pressure and cholesterol** Suppose that both blood pressure and cholesterol levels of adult women can be described with Normal models, and that the correlation between these variables is 0.60. If a woman's blood pressure places her at the 88th percentile, at what percentile would you predict her cholesterol level to be? 153) _____

154) **Maple trees** A forester would like to know how big a maple tree might be at age 50 years. She gathers data from some trees that have been cut down, and plots the diameters (in inches) of the trees against their ages (in years). First she makes a linear model. The scatterplot and residuals plot are shown. 154) _____

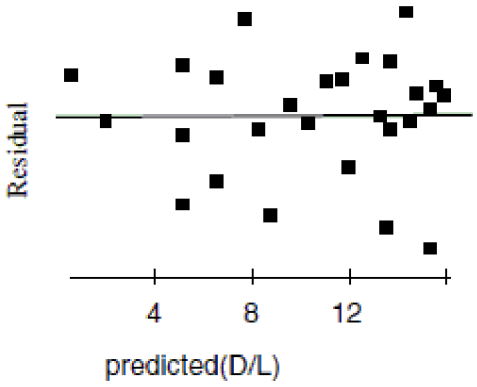


a. Describe the association shown in the scatterplot.
b. Do you think the linear model is appropriate? Explain.
c. If she uses this model to try to predict the diameter of a 50-year old maple tree, would you expect that estimate to be fairly accurate, too low, or too high? Explain.

Now she re-expresses the data, using the logarithm of age to try to predict the diameter of the tree. Here are the regression analysis and the residuals plot.

Dependent variable is: **Diam**
R squared = 84.3%

Variable	*Coefficient*	*s.e. of Coeff*
Constant	*- 8.60770*	*1.681*
Log(Age)	*15.0701*	*1.299*



d. Explain why you think this is a better model.
e. Using this model, predict the diameter of a maple tree at age 50 years.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

155) All but one of the statements below contain a mistake. Which one could be true? 155) _____
A) The number of apricots on a tree and the amount of fertilizer have a 1.12 correlation.
B) There is a strong correlation between type of preferred pet and income level.
C) The correlation between the height of a bean plant and the day is 0.78 in/day.
D) The correlation between the time it takes to get ready in the morning and gender is 0.78.
E) The correlation between your golf score and the number of hours you practice is -0.36.

156) R-sq is a measure of ... 156) _____
A) the change in the y-variable that corresponds with the change in the x-variable.
B) the probability that the regression line makes a correct prediction.
C) the initial predicted starting point of the response variable when x is zero.
D) the percentage of the accuracy of the regression equation.
E) the proportion of the variability in the response variable that is explained by the explanatory variable.

157) If a data point is influential it... 157) _____
A) is guaranteed to be extreme in the vertical direction.
B) will change the slope of the regression equation.
C) is guaranteed to be extreme in the horizontal direction.
D) has a small residual.
E) none of these

158) The relationship between the number of hours a person practices a task and the time it takes them to complete the task is calculated to have R-sq = 56.7%. The value of the correlation coefficient is 158) _____
A) -0.753 B) -0.238 C) 0.238 D) 2.38 E) 0.753

159) A residual plot that has no pattern is a sign that... 159) _____
A) the model is not a good one, because there is no pattern.
B) the original data is curved and the regression line is a good model.
C) the original data is curved and the regression line is not a good model.
D) the original data is straight and the regression line is a good model.
E) the original data is straight and the regression line is not a good model.

160) The price of first class stamp has followed inflation over time and has increased at a constant percentage over time. The most useful predictive model is probably... 160) _____
A) exponential
B) quadratic
C) linear
D) power
E) logarithmic

161) A business owner notes that for every extra hour his store is open, his total sales increase by a fixed amount. His most useful predictive model is probably... 161) _____
A) linear
B) quadratic
C) power
D) exponential
E) logarithmic

162) In predicting the growth of the volume of a small bay by measuring the height of the water at a dock, a researcher is using a model of $\sqrt[3]{\widehat{volume}} = 2.34 + 4.56(height)$, where height is measured in m and volume cubic miles. If the height rises to 3.45 m, what is the predicted volume? 162) _____

A) 7×10^7 m^3
B) 2.62 m^3
C) 18.1 m^3
D) 5902 m^3
E) 1.2×10^{12} m^3

An 8th grade class develops a linear model that predicts the number of cheerios (a small round cereal) that fit on the circumference of a plate by using the diameter in inches. Their model is $\widehat{cheerios} = 0.56 + 5.11(diameter)$.

163) The slope of this model is best interpreted in context as… 163) _____

A) For every 1 inch of diameter, the circumference holds about 0.56 more cheerios.
B) It takes 5.11 cheerios to fill a plate's circumference.
C) A mistake, because π is about 3.14 and that should be the slope.
D) For every 5.11 inches of diameter, the circumference is about 1 cheerio bigger.
E) For every 1 inch of diameter, the circumference holds about 5.11 more cheerios.

164) If the diameter is increased from 4 inches to 14 inches, the predicted number of cheerios will **increase** by about… 164) _____

A) 10
B) 51
C) 72
D) none of these
E) 21

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Solve the problem.

165) **Mistakes.** Describe the mistake made in the following analyses: 165) _____

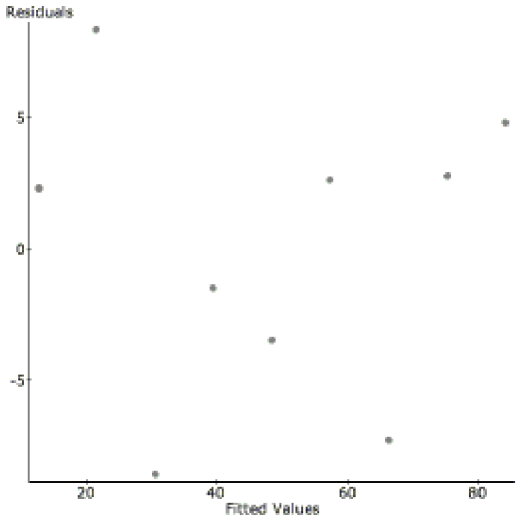
a. Ten teachers compute their average test scores for all their students. Then the superintendent collects their data and finds the school average. He repeats this process for eight different schools and finds a positive correlation between the age of the school average age of the teachers at a school and their average score.

b. The mayor of a city is concerned that the population of the city is growing faster than revenue. He calculates that over the last 5 years, the year and the size of the city have a R-sq of 95.7%. With such a high value, the mayor confidently predicts the population for the next three years of fiscal planning.

166) **Shrimp** From 1982 to 1990, there was a decrease in the number of white shrimp harvested from the Galveston Bay. Here is the regression analysis and a residual plot. The year has been shortened to two digits (82, 83…) and the dependent variable is the number of shrimp collected per hour. 166) _____

Dependent Variable: Shrimp/hour
R-sq = 0.9496342
s: 6.0232354

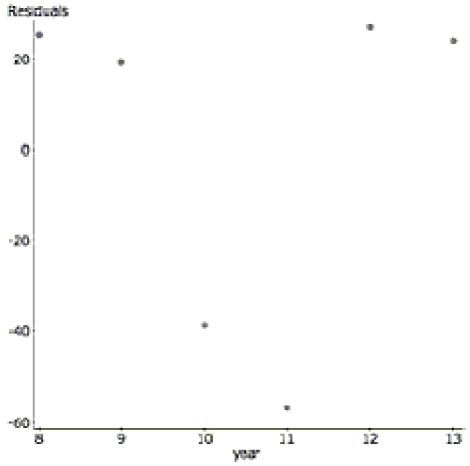
Parameter	**Estimate**	**Std. Err.**
constant	816.71111	66.903419
year	-8.9333333	0.77759635



a. Write the regression equation and define your variables.
b. Find the correlation coefficient and interpret it in context.
c. Interpret the value of the slope in context.
d. In 1991, the shrimp production rebounded (in part due to the effects of El Nino) to 81 shrimp/hour. Find the value of this residual.
e. The prediction for 1991 was very inaccurate. What name do statisticians give to this kind of prediction error?

167) **Students** A growing school district tracks the student population growth over the years from 2008 to 2013. Here are the regression results and a residual plot. 167) _____

students = 119.53 + 172.03 year
Sample size: 6
R-sq = 0.987

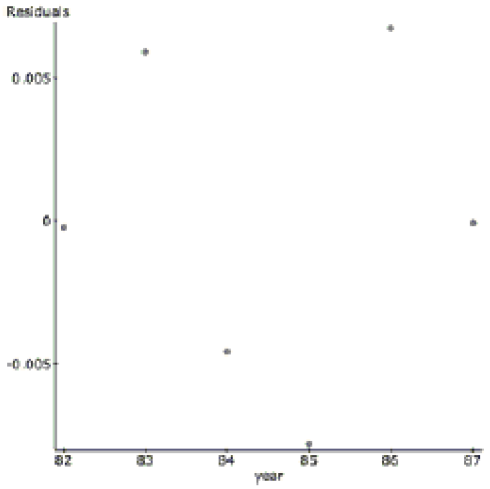


a. Explain why despite a high R-sq, this regression is not a successful model.

To linearize the data, the log (base 10) was taken of the student population. Here are the results.

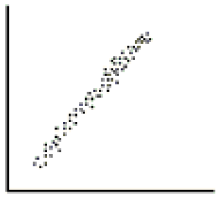
Dependent Variable: log(students)
Sample size: 6
R-sq = 0.994

Parameter	Estimate	Std. Err.
constant	2.871	0.0162
year	0.0389	0.00152

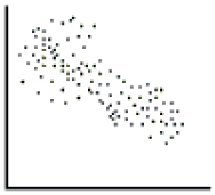


b. Describe the success of the linearization.
c. Interpret R-sq in the context of this problem.
d. Predict the student population in 2014.

1) Correlation measures the strength of a linear association between two quantitative variables. Whether or not a student works is a categorical variable, so correlation cannot be calculated between GPA and whether or not a student works.
2) The correlation coefficient only measures the strength of linear associations. The relationship between x and y that we see here is far from linear (in fact, it is a parabolic relationship).
3) a. There is a fairly strong, negative, linear relationship between the time (in seconds) it took men to run the 1500m race for the gold medal and the year of the Olympics that the race was run in. It appears that the gold medal times for the 1500m race have decreased over time.
   b. $r = -0.94$ (answers between -0.7 and -0.98 are acceptable)
4) a. Correlation has no units.
   b. Correlation has to be between -1 and +1.
   c. Correlation does not change if we reverse the role of the x and y variables.
   d. Correlation does not change when we change units.
5) The variables - owning a pet and condition of the yard - are both categorical variables. Correlation cannot be calculated with categorical variables.
6) a.

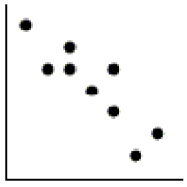


   b.

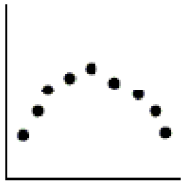


7) a. A positive association means in general people who had more sleep were able to memorize more information.
   b. The child psychologist is attributing association to cause and effect. There is an implication that more sleep will cause better memorization, therefore causing an increase in assessments scores. Perhaps people who had memorized more were able to sleep more restfully, or perhaps differences
   in brain chemistry allowed some people to memorize more and to sleep more easily.
8) a. There is a moderate, negative, linear association between the percent of students taking the SAT test and the total SAT score. It appears that the states with a larger percentage of students taking the SAT test have lower average total scores.
   b. $r = -0.76$ (answers between -0.6 and -0.9 are acceptable)
   c. If the point in the top left corner (4, 1215) were removed, the correlation would become stronger because the remaining points show a pattern with slightly less scatter.
   d. If the point in the very middle (38, 1049) were removed, the correlation would remain about the same; this point does not contribute much to the scatter.
9) There may be an association between customer satisfaction and eye color, but these are both categorical variables so they cannot be "correlated."

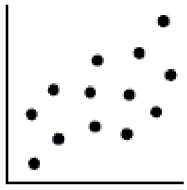
10) a.



b.



c.



11) a. Students who worked more hours tended to have lower grades.
b. They are mistakenly attributing the association to cause and effect. Maybe students with low grades are more likely to seek jobs, or maybe there is some other factor in their home life that leads both to lower grades and to the desire or need to work.

12) a. There is a moderate, positive, linear association between forearm circumference and grip strength among these boys. In general, the larger their forearms, the stronger their grip. One boy in particular had very large forearms and a very strong grip. There was one outlier - the boy with the second largest forearms had one of the weakest grips.
b. Actually $r = 0.652$ - any guess between 0.5 and 0.8 is pretty good.
c. The correlation would become stronger.
d. The correlation would become weaker.

13) a. Player position is a categorical variable. You can use correlation for categorical variables.
b. If the players' heights were listed, you could find correlation, to give one example.

14) __C__ The number of hours you study and your exam score.
__B__ The number of siblings you have and your GPA.
__A__ The number of hours you practice a task and the number of minutes it takes you to complete it.
__D__ The number of hours you use a pencil and its length.

15) a. As the number of degrees increases, the number of bees increases in a linear manner.
b. No. A change of units will not change the correlation.

16) _A_ 0.98 _C_ 0.73 _B_ 0.09 _D_ -0.99

17) She needs to graph her data first and see if it is a linear pattern. Which it almost certainly is not. It will go up and then come back down, probably in a curve. So correlation is not going to be appropriate.

18) D

19) A

20) E

21) C

22) D

23) A

24) E

25) B

26) A
27) B
28) a) $\hat{\$} = 2830 + 15{,}300gpa$
b) Somewhat reliable; based on this model, differences in GPA explain only 52% of the variability in salaries.
c) $51,440
29) a) Weeks worked
b) $r = -0.97$
c) No. The residuals plot shows a distinct curve, and predictions about what will happen three weeks in the future are likely to be unreliable.
30) a) -
b) C
c) +
d) -
e) N
31) a) In general, kids who studied music longer had higher GPA's.
b) Disagree; association does not mean cause and effect. Perhaps the greater parental commitment that supports music lessons also encourages higher grades. (or higher SES enhances both, or people who are better students anyway take music, etc.)
32) a) 89
b) -0.78
c) 3.3 mpg
33) a) Plot 2 points; for example (30,33.6) and (70,30.4)
b) The association is linear, moderately strong, and negative, with one outlier. Children seem to crawl earlier when the temperature is higher, though there was an unusually early age observed for a temperature just above 50°.
c) The model suggests that, on average, babies crawl 0.8 weeks earlier for every 10° higher the temperature is.
d) The model predicts that at a temperature of 0° babies would crawl at an average of 36 weeks old (though this may not mean much as no data were collected at such cold temperatures.)
e) 49% of the variability in crawling age can be explained by variations in temperature.
f) A negative residual would indicate that babies crawled at a younger age than the model predicted.
34) D
35) E
36) D
37) E
38) D
39) E
40) C
41) C
42) A
43) E
44) a) none
b) positive(+)
c) positive(+)
d) negative(-)
e) curved
45) a. Negative association implies that students who sent out more emails during the semester tended to have lower grades.
b. This plan assumes that association means cause and effect. The college incorrectly proposes to limit emails through the college address as a way of increasing student grades. Perhaps students with bad grades console themselves by emailing friends.

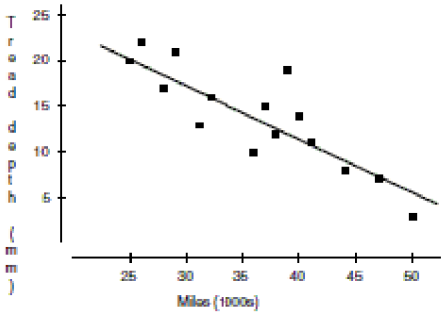
46) Explanatory variable (x): the number of cars the car dealer sold the following weekend
Response variable (y): the number of TV commercials the car dealer ran each week

$b_1 = \frac{rs_y}{s_x} = \frac{(0.56)(1.8)}{4.2} = 0.24$

Substituting $(\bar{x}, \bar{y})$, $12.4 = b_0 + (0.24)(30.5)$, so $b_0 = 5.08$.

Model: $\hat{comm} = 5.08 + 0.24cars$
Car dealer hopes to sell 40 cars: predicted number of commercials needed = 5.08 + 0.24(40) = 14.68, or 15 commercials this week.

47) a.



With 25000 miles: 36 - 0.6(25) = 21; (25, 21)
With 45,000 miles: 36 - 0.6(45) = 9; (45, 9)
Model goes through points: (25, 21) and (45, 9).
b. The explanatory variable is the number of miles tires had been driven (in thousands).
c. The correlation must have the same sign as the slope. $r = \sqrt{R^2} = \sqrt{0.74} = -0.86$
d. The association between the number of miles tires have been driven (in thousands) and the tire tread depth (in mm) is a moderately strong negative linear relationship. Tires with more miles tend to have lower tread depth. (In this model, the tire tread is expected to be an additional 0.6 mm lower for every additional 1000 miles the tires have been driven.) One tire had unusually deep tread for the number of miles driven.
e. This model suggests that for every additional 1000 miles the tires are driven, the depth of the tire tread will decrease by 0.6 mm, on average.
f. The model predicts that brand new tires (number of miles equals zero) have tread averaging 36 mm deep.
g. R^2 means that 74% of the variability in tread depth is explained by the variations in the number of miles the tires have been driven.
h. Residual equals the observed tread depth minus the predicted tread depth. A negative residual means that the observed amount of tread depth is less than the predicted amount of tread depth, using this model. This means that the tire tread is actually wearing out faster than the model predicts.

48) D
49) D
50) E
51) D
52) E
53) A
54) D
55) D
56) C
57) E

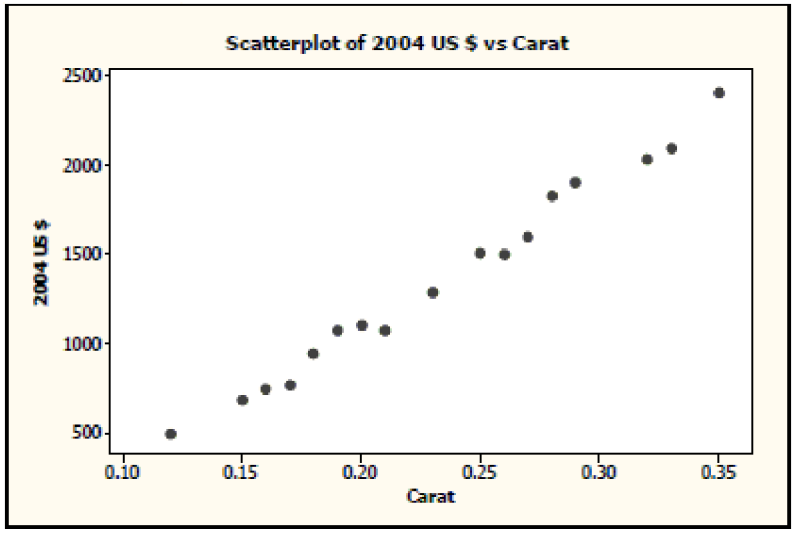
58) A—curve
B—positive(+)
C—negative(-)
D—negative(-)
E—none

59) Positive association implies that as students work more hours their GPA's tend to be higher.

60) a. $\hat{highscore}$ = 524.8 + 2498.8(*hours*)
b. For every one more hour of time played, the high score is predicted to increase by 2498.8
c. A beginning player is predicted to score 524.8
d. The typical miss of the predictions on the regression is 383.3 points
e. $\sqrt{0.765}$ = 0.875 , There is a strong, positive, linear relationship between hours played and high score points.

61) a. There is only a weak relationship. R^2 = 13.8% and r = -0.372. The relationship seen on the scatterplot is very weak.
b. It appears there may be no relationship at all. The value -0.372 does not appear to be unusual. 15 out of 100 times the correlation was even closer to negative one. So the association we are observing could be due to random variation.

62)



There is a strong, positive, linear association between the size of the diamond and its cost. The cost of a diamond increases with size.

63)

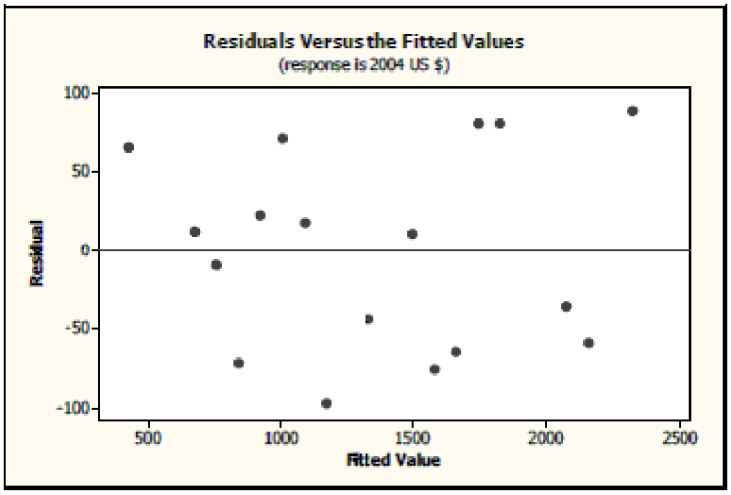
The regression equation is
2004 US $ = - 559 + 8225 Carat

Predictor	Coef	SE Coef	T	P
Constant	-558.52	57.88	-9.65	0.000
Carat	8225.1	239.1	34.40	0.000

S = 64.9355 R-Sq = 98.7% R-Sq(adj) = 98.7%

Predicted cost = -558.52 + 8225.1(carat)

64)



A linear model is appropriate for this problem. The residual plot shows no obvious pattern.

65) The slope of the model is 8225.1. The model predicts that for each additional carat, the cost of the diamond will increase by $8225.10, on average. This can also be interpreted as for each additional 0.01 carat, the cost of the diamond will increase by $82.251, on average.

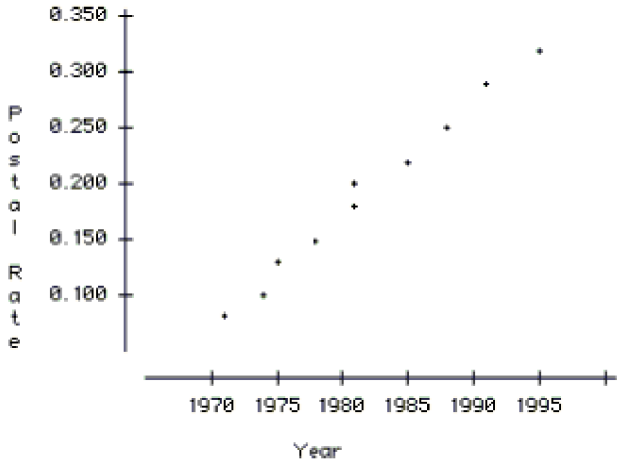
66) The intercept of the model is -558.52. The model predicts that a diamond of 0 carats costs -$558.52. This is not realistic.

67) The correlation, r, is $r = \sqrt{0.987} = 0.993$. Since the scatterplot shows a positive relationship, the positive value must be used.

68) $R^2 = 0.987$. So 98.7% of the variation in diamond prices can be accounted for by the variation in the size of the diamond.

69) It would be better for customers to have a negative residual from this model, since a negative residual would indicate that the actual cost of the diamond was less than the model predicted it to be.

70)



There is a strong, positive, linear association between the year and the first class postal rate. Postal rates have increased over time.

71)

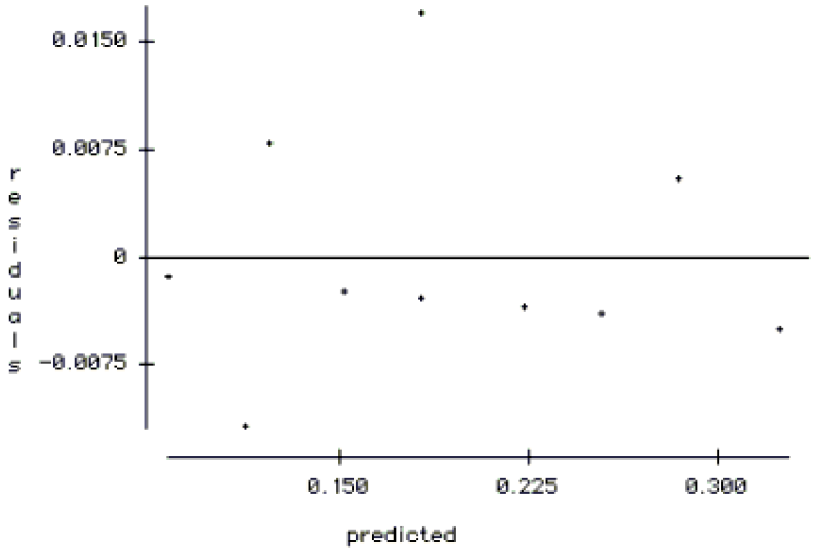
Dependent variable is: **Postal Rate**
No Selector
R squared = 99.0% R squared (adjusted) = 98.8%
s = 0.008641 with 10 - 2 = 8 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	0.0563626	1	0.0563626	755
Residual	0.000597404	8	0.0000746754	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	-19.9278	0.7324	-27.2	≤ 0.0001
Year	0.0101518	0.0003695	27.5	≤ 0.0001

$\hat{Rate}$ = -19.93 + 0.01015(*year*)

72)



Yes, a linear model is appropriate for this problem. A review of the residual plots shows no obvious pattern.

73) Slope of model is 0.0101518. The model predicts that for every additional year the first class postal rate will increase by $0.01, on average.

74) Intercept of the model is -19.93. The model predicts that at Year = 0, the first class postal rate was -$19.93. This is not realistic.

75) The correlation, *r*, is $r = \sqrt{0.990} = 0.9950$. Since the scatterplot shows a positive relationship, the positive value must be used.

76) $R^2 = 0.990$. So 99.0% of the variation in first class postal rates can be accounted for by the variation in year.

77) It would be better for customers to have a negative residual from this model. A negative residual would indicate that the actual first class postal rate is lower than the model predicted it would be.

78) $\hat{oranges}$ = 390.59 + 525.84(*trees*)

79) We predict roughly 525 oranges for every tree in the grove.

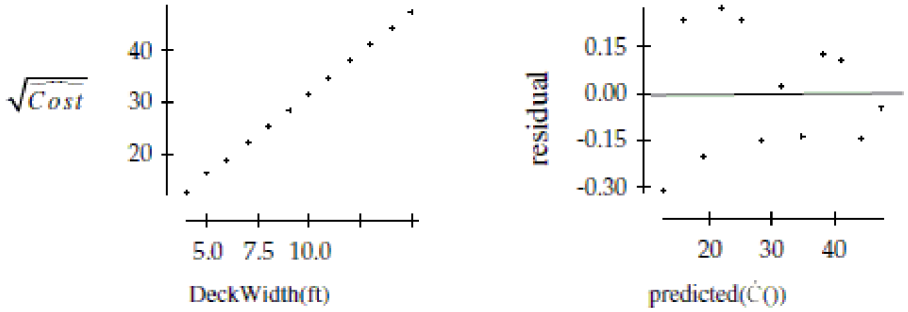
80) Our regression equation makes predictions that miss the data by about 31,395 oranges, on average.

81) Yes. For the four smaller groves, this error is about as big as the entire harvest. Since there are such extreme differences between the small orchards and the big orchards, it might be better to divide the data set into two separate groups. (Note: this observation is a bit of a stretch for some students in chapter 7. A small investigative task, as it were!)

82) 11.4% of the variation is due to other factors. It could be soil quality, water supply, weather, type of orange, etc…
83) The predicted number of oranges is 18795. 15400 – 18795 = -3395 oranges
84) He would be displeased because his grove produced 3395 fewer oranges than predicted for a grove of his size.
85) $\sqrt{0.886} = 0.941$, which indicates a strong, linear positive relationship between the number of oranges and the number of trees.
86) No. It would be a dangerous assumption to think that Florida oranges are similar to California oranges in production.
87) Relationships based on averages have higher correlation coefficients than relationships based on individual data. Therefore, a scatterplot of the final exam score versus midterm score for individual students would show much more scatter and a weaker correlation coefficient.
88) Since we see a clear pattern in the residuals plot, it does not seem reasonable to use linear regression for this model.
89) a. There is a moderately strong, positive association between weight and height. The variation in weight is larger for larger values of height.
    b. Men, on average, appear to be taller and heavier than women. We can clearly see two groups (with some overlap) in this scatterplot.
    c. Since there appears to be a difference between men and women in the plot, it is not correct to use a single model obtained by these data to make a prediction. Furthermore, there appears to be a great deal of scatter at 70", with weights varying by over 50 pounds for women and well over 100 pounds for men.
90) There is no clear pattern. At first glance, there appears to be a weak, negative association between grams of fiber and the number of calories in the cereals. Yet, the five points at the bottom of the graph are outside the pattern, with extremely low numbers of calories. Additionally, the three points on the right of the scatterplot have an unusually high amount of fiber, making them outliers and influential points.
91) The points in the bottom left corner seem to have extremely low calorie content for cereals between zero and six grams of fiber. The points with 9, 10 and 14 grams of fiber appear to have an unusually high amount of fiber for their calorie content, making them outliers and influential points. These three points would also be leverage points, creating the impression that there is a negative association between grams of fiber and the number of calories in the cereals.
92) This data contains information about cereals with fiber content between 0 and 14 grams. It would be extrapolation to try to use this data to predict the calorie content of cereals with 22 grams of fiber.
93) There appears to be a moderately strong, positive, but nonlinear (curved) association between speed and number of strikeouts. Pitchers with higher speeds tend to have more strikeouts. There is one point that doesn't fit the pattern. One pitcher had more strikeouts on average than his average speed would typically indicate.
94) There is one pitcher that deviates from the pattern. The pitcher has a slow fastball, about 84 mph, but a high number of strikeouts, about 8. Perhaps this pitcher has another pitch, like a knuckleball, that makes it difficult for opposing hitters.
95) The association would probably be weaker. This plot uses averages from an entire season. Data from a single month would have more variability.
96) 70 mph is lower than any of the speeds in these data, and extrapolation is risky business. Also, these data are for only pitchers on the Bulldogs, and may not be representative of others.
97) There is a very weak linear relationship. The scatterplot shows only the slightest of positive associations. r = 0.195 and r^2 = 3.8%. There is very little linearity at all.
98) This new point is very influential. It raised the slope from 0.888 candies/year of age to 3.4! r increased to 0.76 and r^2 increased to 58%. This one data point makes a very weak association look very strong.
99) We can see a bend in the scatterplot, so a linear regression is not appropriate. We must re-express the data in order to use linear regression.

100) a. Using the square root of the costs, we get the model: $\sqrt{\hat{Cost}} = -0.135 + 3.17(DeckWidth)$

b. The scatterplot of $\sqrt{\hat{Cost}}$ vs. *DeckWidth* is much straighter than the original scatterplot, and the residuals plot is scattered.



c. $\sqrt{\hat{Cost}} = -0.135 + 3.17(DeckWidth) = -0.135 + 3.17(10.5) \approx 33.15$

$\hat{Cost} = 33.15^2 \approx \$1098.92 \approx \$1100$

d. It is not reasonable to make a prediction for the cost of a square deck that is 20 feet wide, since prediction for a width of 20 feet would be extrapolation.

101) a. Let explanatory variable be Year - 1900; so, 1948 is input as 48.

Let response variable be log(Ticket Price)

Exponential model: $\log(\hat{Ticket}) = -1.73 + 0.0269(Year - 1900)$

b. $\log(\hat{Ticket}) = -1.73 + 0.0269(104) = 1.0676$

$\hat{Ticket} = 10^{1.0676} = \$11.68$

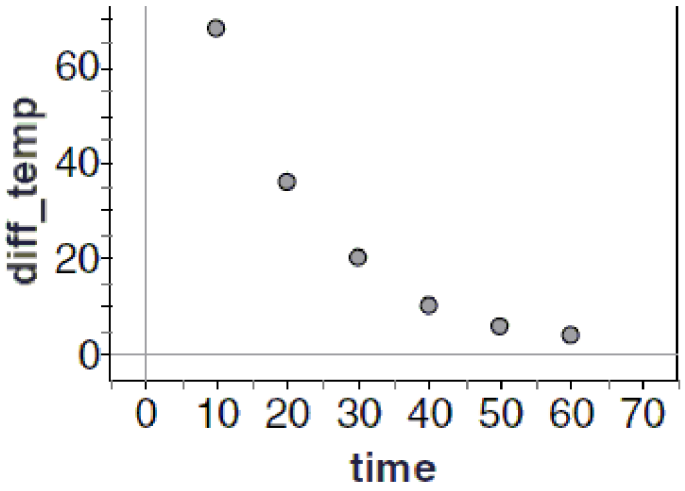
102) a. $\log \hat{Element} = 2.505 - 0.0749(time)$

b. Time is measured in days, so 30 minutes, or half an hour, is $\frac{1}{48}$ days.

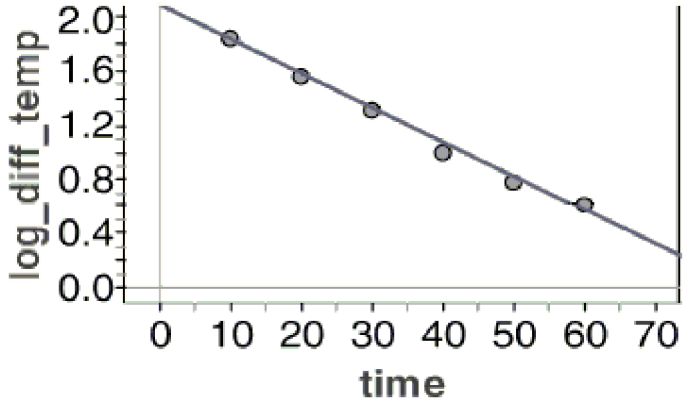
$\log \hat{Element} = 2.505 - 0.0749\left(\frac{1}{48}\right) \approx 2.5034$

$10^{2.5034} \approx 318.74$ grams

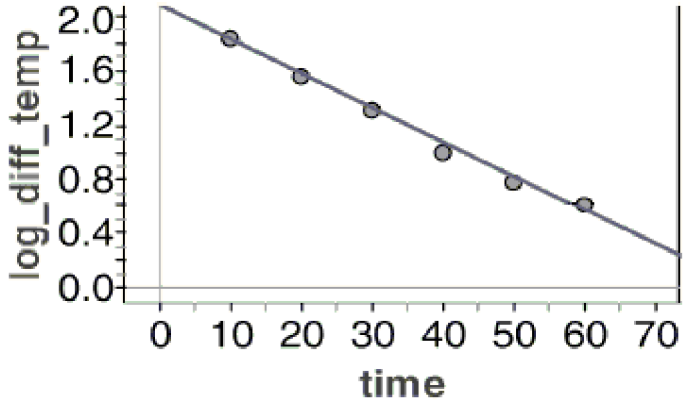
103)



104)



105) $\log(\widehat{difftemp}) = 2.057 - 0.025 time$



106) $\log(\widehat{difftemp}) = 2.057 - 0.025(45) = 0.932$

$\widehat{difftemp} = 10^{0.932} = 8.551$

107) $\log(\widehat{difftemp}) = 2.057$

$\widehat{difftemp} = 10^{2.057} = 114.025$

This represents the model's prediction of the difference in the temperature at the beginning of the experiment.

108) The data is clearly curved, so using a linear model would not be an appropriate choice.

109) The residual plot has no pattern, so it does appear that square root of cost and height have a linear relationship.

110) $sqrt(\widehat{cost}) = 1.1857 + 0.1792(height)$

111) 97.9% of the change in square root of cost is explained by the regression on height.

112) $1.1857 + 0.1792(48) = 9.7873; 9.7873^2 = \$95.79$

113) A

114) A

115) D

116) C

117) A

118) C

119) B

120) C

121) C

122) C

123) a. $\sqrt{\hat{str}} = 2.64 + 1.37dia$ (among other possibilities)
b. The model estimates this rope will have an approximate breaking strength of 1367 kg, but this extrapolation should be viewed with caution.
124) $z_M = 0.57z_V = 0.57(1.28) = 0.73$, corresponding to the 76th percentile
125) a. -0.953
b. 16.5 units/cc
c. Too high; the residuals are generally negative for times between 2 and 5 hours.
d. The residuals show a random pattern with no curvature.
e. 12.9 units/cc
126) A
127) A
128) C
129) E
130) D
131) D
132) E
133) D
134) A
135) D
136) a. Positive association implies that countries with larger populations tend to have more stork pairs and countries with smaller populations tend to have fewer stork pairs.
b. This suggestion assumes that association means causation. The wildlife advocates incorrectly propose human population growth as a way to increase the number of stork pairs. Perhaps there is a lurking variable, like land mass, that accounts for the positive association between the two variables.
137) a. The response variables is "debt per capita."
b. $r = \sqrt{0.988} = 0.994$ We know that the correlation coefficient will be positive, since there is a positive association between the two variables.
c. On average, debt per capita increases $1128 per year.
d. About 98.8% of the variability in debt per capita is explained by the model.
e. There is a definite curve in the residuals plot, which was not obvious in the original scatterplot. Thus, linear regression is <u>not</u> appropriate for these data.
138) a. Explanatory variable: number of hours spent studying
Response variable: score on exam
slope: 2.40; intercept: 56; Model: $\hat{Score} = 56 + 2.4(Hours)$
b. $\hat{Score} = 56 + 2.4(15) = 92$; A student who studies for 15 hours should expect to score 92 points on the exam, based on this model.
139) $r = 0.65$
height at 60th percentile: $z_{ht} = 0.25$
Regression to the mean predicts that z_{wt} will be r times as far from 0 as z_{ht} was, so
$z_{wt} = r(z_{ht}) = 0.65(0.25) \approx 0.16$; The man's weight will be approximately the 56th percentile.
140) a. Re-express the data using *Time* as the explanatory variable and log(*Mass*) as the response variable. The model is $\log(\hat{Mass}) = 2.00143 - 0.000055(Time)$.
b. $\log(\hat{Mass}) = 2.00143 - 0.000055(7500) = 1.58893$, so $\hat{Mass} = 10^{1.58893} = 38.8g$ remaining.
c. No. This model is to be used to predict *Mass* from *Time*, not *Time* from *Mass*. We would need to develop a new model using *Mass* as the explanatory variable and *Time* as the response variable to make this prediction.
141) C
142) A

143) D
144) B
145) B
146) C
147) D
148) B
149) B
150) D
151) a. Using the model, mileage explains only 47% of the variability in price and some of the residuals are nearly $2000. Estimates of price will be only moderately accurate.
b. Slope = -0.11; The model predicts that for every additional mile the car had been driven the price of the car would decrease $0.11, on average.
c. $\hat{price} = 15{,}327 - 0.11(42{,}000) = 15{,}327 - 4620$ = $10,707
Residual at $10,707 from residual plot: $1000
Asking price = predicted price plus residual = $10,707 + $1000 = $11,707
152) a. Re-express the data using *Time* as explanatory variable and log(*Concentration*) as response variable.
Model: $\log(\hat{Concentration}) = 1.789 - 0.169(Time)$
b. When Time = 8, $\log(\hat{Conc}) = 1.789 - 0.169(8) = 0.437$; $\hat{Conc} = 10^{0.437} = 2.74$ units/cc
153) Correlation: $r = 0.60$; Blood pressure at 88th percentile: $z_{BP} = 1.175$
Regression to the mean predicts that z_C will be r times as far from 0 as z_{BP} was.
$z_C = 0.60 z_{BP} = 0.60(1.175) = 0.705$, so cholesterol will be approximately the 76th percentile.
154) a. The association between age of a maple tree and its diameter is moderately strong, positive, and curved, not linear.
b. No, the plot of residuals shows an obvious pattern. Trees with diameters less than 6 inches have negative residuals, trees with diameters between 9 and 14 inches have positive residuals, and trees with diameters larger than 15 inches have negative residuals.
c. Using this model to predict the diameter of a 50-year old maple tree would be too high. The model in the original scatterplot is above the data points in the region of 50 years and the residuals above 15 inches are negative indicating that the model would overestimate the diameter of the tree.
d. There is no obvious pattern to the residual plot.
e. Model: $\hat{Diam} = -8.6077 + 15.0701[\log(Age)]$
At 50 years, $\hat{Diam} = -8.6077 + 15.0701[\log(50)] \approx 17.0$
Prediction for the diameter of a maple tree at age 50 years is 17.0 inches.
155) E
156) E
157) B
158) A
159) C
160) A
161) A
162) D
163) E
164) B
165) a. In using data that has been averaged over so many variables, he is likely to cloud actual associations that are of interest.
b. The data is probably not linear and should be linearized before regression is done. Also, 3 future years is in danger of extrapolation.

166) a. $\widehat{shrimp/hour}$ = 816.7 - 8.933(*year*)
b. -0.974; This tells us there is a strong, negative correlation between year and number of shrimp collected per hour.
c. For every year that goes by, the number of shrimp collected per hour is decreasing by about 8.9.
d. Prediction = 3.797 shrimp/hour; 81 – 3.797 = 77.2 shrimp/hour
e. This is extrapolation. We assume that the trend will continue, but it did not.

167) a. Even though R-sq = 98.7%, the residual plot has a curved pattern. Also, we believe that populations grow exponentially, so a linear model is probably not appropriate.
b. The residual plot has a slight curve, but this seems to be an improvement on our first model. Also, taking the log of a population growth is a correct model choice.
c. 99.4% of the variability in the **log** of student population is successfully explained by the regression on year.
d. 2.871 + 0.0389*14 = 3.4156; $10^{3.4156}$= 2604 students