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

Answer the question appropriately.

1) A random sample of records of electricity usage of homes gives the amount of electricity used in July and size (in square feet) of 135 homes. A regression was done to predict the amount of electricity used (in kilowatt-hours) from size. The residuals plot indicated that a linear model is appropriate. The model is

\[ \hat{\text{usage}} = 1258 + 0.8 \times \text{size} \]

Explain what the slope of the line says about the electricity usage and home size.

A) On average, the amount of electricity used is 0.8 kilowatt hours less than the size of the house.
B) On average, the amount of electricity used increases by 1258 kilowatt-hours when the size of the house is increased by a square foot.
C) On average, the amount of electricity used increases by 0.8 kilowatt-hours when the size of the house is increased by a square foot.
D) On average, the size of the house increases by 0.8 feet for every kilowatt-hour used.
E) On average, the size of the house increases by 1258 feet for every kilowatt-hour used.

Use the model to make the appropriate prediction.

2) A random sample of records of electricity usage of homes in the month of July gives the amount of electricity used and size (in square feet) of 135 homes. A regression was done to predict the amount of electricity used (in kilowatt-hours) from size. The residuals plot indicated that a linear model is appropriate. The model is

\[ \hat{\text{usage}} = 1297 + 0.2 \times \text{size} \]

How much electricity would you predict would be used in a house that is 2361 square feet?

A) 3658.2 kilowatt-hours
B) 824.8 kilowatt-hours
C) 1769.2 kilowatt-hours
D) 472.2 kilowatt-hours
E) 5320.00 kilowatt-hours

Tell what the residual plot indicates about the appropriateness of the linear model that was fit to the data.

3) A) Model is not appropriate. The relationship is nonlinear.
B) Model may not be appropriate. The spread is changing.
C) Model is appropriate.
Answer the question appropriately.

5) The relationship between the number of games won by a minor league baseball team and the average attendance at their home games is analyzed. A regression to predict the average attendance from the number of games won has an $R^2 = 27.8\%$. The residuals plot indicated that a linear model is appropriate. Write a sentence summarizing what $R^2$ says about this regression.

A) Differences in average attendance explain 72.2% of the variation in the number of games won.
B) The number of games won explains 72.2% of the variation in average attendance.
C) In 27.8% of games won the attendance was at least as large as the average attendance.
D) Differences in average attendance explain 27.8% of the variation in the number of games won.
E) The number of games won explains 27.8% of the variation in average attendance.

Use the given data to find the equation of the regression line. Round to 3 significant digits, if necessary.

6) Ten Ford Escort classified ads were selected. The age and prices of several used Ford Escorts are given in the table.

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$10,000</td>
</tr>
<tr>
<td>2</td>
<td>$8500</td>
</tr>
<tr>
<td>2</td>
<td>$8000</td>
</tr>
<tr>
<td>3</td>
<td>$6000</td>
</tr>
<tr>
<td>3</td>
<td>$5900</td>
</tr>
<tr>
<td>4</td>
<td>$5800</td>
</tr>
<tr>
<td>4</td>
<td>$5000</td>
</tr>
<tr>
<td>5</td>
<td>$3000</td>
</tr>
<tr>
<td>6</td>
<td>$2000</td>
</tr>
<tr>
<td>6</td>
<td>$1900</td>
</tr>
</tbody>
</table>

A) $\hat{\text{price}} = 7.05 - 0.000616 \text{ age}$
B) $\hat{\text{price}} = 10000 - 1600 \text{ age}$
C) $\hat{\text{price}} = 7200 - 692 \text{ age}$
D) $\hat{\text{price}} = -1580 + 11300 \text{ age}$
E) $\hat{\text{price}} = 11300 - 1580 \text{ age}$
7) A random sample of records of electricity usage of homes in the month of July gives the amount of electricity used and size (in square feet) of 135 homes. A regression was done to predict the amount of electricity used (in kilowatt-hours) from size. The residuals plot indicated that a linear model is appropriate. The model is 
\[ \hat{y} = 1244 + 0.6x \] 
The people in a house that is 2314 square feet used 500 kilowatt-hours less than expected. How much did they use? 
A) 2132.4 kilowatt-hours 
B) 644.4 kilowatt-hours 
C) 1783.33 kilowatt-hours 
D) 3058.6 kilowatt-hours 
E) 1888.4 kilowatt-hours

8) A random sample of records of electricity usage of homes gives the amount of electricity used in July and size (in square feet) of 135 homes. A regression was done to predict the amount of electricity used (in kilowatt-hours) from size. The residuals plot indicated that a linear model is appropriate. Do you think the slope is positive or negative? Why? 
A) Positive. More square feet indicates more houses. 
B) Negative. Smaller homes should use less electricity. 
C) Negative. Larger homes should use less electricity. 
D) Positive. Larger homes should use more electricity. 
E) Positive. The larger the number of houses the more electricity used.

Fill in the missing information.

<table>
<thead>
<tr>
<th>9)</th>
<th>( x )</th>
<th>( s_x )</th>
<th>( y )</th>
<th>( s_y )</th>
<th>( r )</th>
<th>( \hat{y} = b_0 + b_1x )</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>8</td>
<td>?</td>
<td>?</td>
<td>-0.6</td>
<td>( \hat{y} = 190 - 2x )</td>
<td></td>
</tr>
</tbody>
</table>

A) \( \bar{y} = 174; s_y = 0.15 \) 
B) \( \bar{y} = 158; s_y = 26.67 \) 
C) \( \bar{y} = 32; s_y = 4.00 \) 
D) \( \bar{y} = 222; s_y = 26.67 \) 
E) \( \bar{y} = 206; s_y = 6.80000019 \)

Answer the question appropriately.

10) 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. 
A) 0.56 SD below 
B) 2.5 SD above 
C) equal to 
D) 0.56 SD above 
E) 1.4 SD above
Fill in the missing information.

<table>
<thead>
<tr>
<th></th>
<th>( \bar{x} )</th>
<th>( s_x )</th>
<th>( \bar{y} )</th>
<th>( s_y )</th>
<th>( r )</th>
<th>( \hat{y} = b_0 + b_1x )</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>?</td>
<td>1.2</td>
<td>?</td>
<td>120</td>
<td>?</td>
<td>( \hat{y} = -110 +50x )</td>
</tr>
</tbody>
</table>

A) \( \bar{y} = -245; r =0.42 \)
B) \( \bar{y} = 135; r = -0.50 \)
C) \( \bar{y} = -107.3; r =0.42 \)
D) \( \bar{y} = -50; r =0.03 \)
E) \( \bar{y} =25; r =0.50 \)

Answer the question appropriately.

12) 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?
   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 and II  B) I only  C) III only  D) II only  E) None

13) A golf ball is dropped from 15 different heights (in inches) and the height of the bounce is recorded (in inches.)
   The regression analysis gives the model \( \hat{y} =0.5 +0.72 \) drop. A golf ball dropped from 63 inches bounced 1 inch less than expected. How high did it bounce?
   A) 46.86 inches  B) 65.22 inches  C) 44.36 inches  D) 85.42 inches  E) 44.86 inches

14) A golf ball is dropped from 15 different heights (in inches) and the height of the bounce is recorded (in inches.)
   The regression analysis gives the model \( \hat{y} =-0.3 +0.69 \) drop. Explain what the slope of the line says about the bounce height and the drop height of the ball.
   A) On average, the bounce height will be 0.69 inches less than the drop height.
   B) On average, the bounce height increases by -0.3 inches for every extra inch of drop height.
   C) On average, the drop height increases by -0.3 inches for every extra inch of bounce height.
   D) On average, the drop height increases by 0.69 inches for every extra inch of bounce height.
   E) On average, the bounce height increases by 0.69 inches for every extra inch of drop height.

15) A correlation of zero between two quantitative variables means that
   A) we have done something wrong in our calculation of \( r \).
   B) None of the above.
   C) there is no linear association between the two variables.
   D) there is no association between the two variables.
   E) re-expressing the data will guarantee a linear association between the two variables.

16) Consider the four points (20, 30), (30, 70), (40, 40), and (50, 80). The least squares line is \( \hat{y} =13 +70x \). Explain what "least squares" means using these data as a specific example.
   A) The line \( \hat{y} =13 +70x \) minimizes the sum of the squared horizontal distances from the points to the line.
   B) The line \( \hat{y} =13 +70x \) minimizes the square of the standard deviation.
   C) The line \( \hat{y} =13 +70x \) minimizes the sum of the squared vertical distances from the points to the line.
   D) The line \( \hat{y} =13 +70x \) minimizes the sum of the vertical distances from the points to the line.
   E) The line \( \hat{y} =13 +70x \) minimizes the sum of the squared difference between the x and y values.
A random sample of 150 yachts sold in the United States last year was taken. A regression to predict the price (in thousands of dollars) from length (in feet) has an $R^2 = 15.8\%$. What is correlation between length and price?

A) 0.918  
B) 0.025  
C) 0.709  
D) 0.397  
E) 0.158

Explain what is wrong with each interpretation. Assume calculations are done correctly.

A sociology student does a study to determine whether people who exercise live longer. He claims that someone who exercises 7 days a week will live 15 years longer than someone who doesn’t exercise at all.

A) Predictions based on a regression line are for average values of x and y. The actual average life expectancy changes every year so an accurate prediction is impossible.

B) There is nothing wrong with the interpretation.

C) A linear model is inappropriate for sociology studies.

D) Predictions based on a regression line are for average values of y for a given x. The actual life expectancy will vary around the prediction.

E) The $R^2$ has to be greater than 90% to make a statement like this.

Answer the question appropriately.

A random sample of records of electricity usage of homes gives the amount of electricity used in July and size (in square feet) of 135 homes. A regression was done to predict the amount of electricity used (in kilowatt-hours) from size. The residuals plot indicated that a linear model is appropriate. What units does the slope have?

A) Slope is kilowatt-hours per house.

B) Slope is kilowatt-hours per square foot.

C) Slope is square feet per house.

D) Slope is square feet per kilowatt-hour.

E) Slope is houses per kilowatt-hour.

The relationship between the number of games won by a minor league baseball team and the average attendance at their home games is analyzed. A regression analysis to predict the average attendance from the number of games won gives the model attendance $= -2800 + 222$ wins. The Hackenburg Monkeys averaged 14,169 fans at each game. They won 43 times. Calculate the residual and explain what it means.

A) 26,515 people. The Hackenburg Monkeys averaged 26,515 more fans than would be predicted for a team with 43 wins.

B) 7423 people. The Hackenburg Monkeys averaged 7423 more fans than would be predicted for a team with 43 wins.

C) 14,156 people. On average the Hackenburg Monkeys will have 14,156 extra people.

D) -7423 people. The Hackenburg Monkeys averaged 7423 less fans than would be predicted for a team with 43 wins.

E) 6746 people. The Hackenburg Monkeys were expected to average 6746 people for each game.

A random sample of records of electricity usage of homes gives the amount of electricity used and size (in square feet) of 135 homes. A regression to predict the amount of electricity used (in kilowatt-hours) from size has an $R^2$ squared of 71.6%. The residuals plot indicated that a linear model is appropriate. Write a sentence summarizing what $R^2$ says about this regression.

A) Size differences explain 71.6% of the variation in electricity usage.

B) Differences in electricity usage explain 71.6% of the variation in the size of house.

C) Size differences explain 71.6% of the variation in the number of homes.

D) Size differences explain 28.4% of the variation in electricity usage.

E) Differences in electricity usage explain 28.4% of the variation in the number of house.
22) Using advertised prices for used Ford Escorts a linear model for the relationship between a car’s age and its price is found. The regression has an \( R^2 = 87.1\% \). Describe the relationship

A) Positive, weak linear relationship. As the age increases the price goes down.
B) Positive, strong linear relationship. As the age increases the price goes up.
C) Negative, strong linear relationship. As the age increases the price goes down.
D) Negative, weak linear relationship. As the age decreases the price goes down.
E) Negative, strong linear relationship. As the age decreases the price stays the same.

23) Which statement about residuals plot is true?

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

Fill in the missing information.

24) \[
\begin{array}{c|c|c|c|c}
\bar{x} & s_x & \bar{y} & s_y & r \\
30 & ? & 6 & 9 & 6 \\
\end{array}
\]

\[
y = b_0 + b_1 x
\]

A) 4  B) 3  C) 7  D) 2  E) 9

Answer the question appropriately.

25) A golf ball is dropped from 15 different heights (in inches) and the height of the bounce is recorded (in inches.)

The regression analysis gives the model \( \hat{y} = -84 + 18x \). A golf ball company is trying to show that its new ball will increase your driving distance. If the new ball is dropped from several heights would the company rather see positive or negative residuals. Explain.

A) Neither. The ball should bounce the same as expected otherwise it wasn’t manufactured properly.
B) Negative. This would mean the ball is bouncing more than expected and you would more likely be able to hit it longer.
C) Positive. This would mean the ball is bouncing more than expected and you would more likely be able to hit it longer.
D) Negative. The ball isn’t bouncing as high as expected so you would more likely be able to hit it longer.
E) Positive. This would mean the ball is being dropped from higher distances so you would more likely be able to hit it longer.

26) Residuals are . . .

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

<table>
<thead>
<tr>
<th>x</th>
<th>s_x</th>
<th>y</th>
<th>s_y</th>
<th>r</th>
<th>( y = b_0 + b_1 x )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.08</td>
<td>1.6</td>
<td>-0.2</td>
<td>( \hat{y} = )</td>
<td></td>
</tr>
</tbody>
</table>

A) \( y = 7.93 - 0.01x \)  
B) \( y = -52.10 + 20.00x \)  
C) \( y = 19.90 - 4.00x \)  
D) \( y = -4.00 + 19.90x \)  
E) \( y = 20.00 + 2.63x \)

Answer the question appropriately.

28) Using advertised prices for used Ford Escorts a linear model for the relationship between a car’s age and its price is found. The regression has an \( R^2 = 87.7\% \). Write a sentence summarizing what \( R^2 \) says about this regression.

A) The price of the car explains 87.7\% of the variation in age.  
B) The price of the car explains 9.36\% of the variation in price.  
C) The price of the car explains 12.3\% of the variation in age.  
D) The age of the car explains 12.3\% of the variation in price.  
E) The age of the car explains 87.7\% of the variation in price.

29) A residuals plot is useful because

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

30) A golf ball is dropped from 15 different heights (in inches) and the height of the bounce is recorded (in inches.) The regression analysis gives the model bounce = 0.3 + 0.72 drop. A golf ball dropped from 70 inches bounced 51.7 inches. What is the residual for this bounce height?

A) 51.4 inches  
B) 1 inches  
C) -1 inches  
D) 0.72 inches  
E) 2 inches

31) A golf ball is dropped from 15 different heights (in inches) and the height of the bounce is recorded (in inches.) The regression analysis gives the model bounce = -0.5 + 0.69 drop. A golf ball dropped from 66 inches bounced a height whose residual is -0.5 inches. What is the bounce height?

A) 65.155 inches  
B) 0.5 inches  
C) 65.5 inches  
D) 45.54 inches  
E) 44.54 inches
Tell what the residual plot indicates about the appropriateness of the linear model that was fit to the data.

32) A) Model is appropriate.
   B) Model is not appropriate. The relationship is nonlinear.
   C) Model may not be appropriate. The spread is changing.

Use the given data to find the equation of the regression line. Round to 3 significant digits, if necessary.

33) Ten students in a tutor program were randomly selected. Their grade point averages (GPAs) when they entered the program were less than 2.5. The following data were obtained regarding their GPAs on entering the program versus their current GPAs.

<table>
<thead>
<tr>
<th>Entering GPA (E)</th>
<th>Current GPA (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5</td>
<td>2.6</td>
</tr>
<tr>
<td>1.8</td>
<td>2.1</td>
</tr>
<tr>
<td>2.3</td>
<td>2.5</td>
</tr>
<tr>
<td>1.6</td>
<td>1.6</td>
</tr>
<tr>
<td>1.5</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0</td>
<td>2.4</td>
</tr>
<tr>
<td>2.1</td>
<td>2.2</td>
</tr>
<tr>
<td>2.4</td>
<td>2.7</td>
</tr>
<tr>
<td>1.9</td>
<td>2.3</td>
</tr>
<tr>
<td>2.1</td>
<td>2.3</td>
</tr>
</tbody>
</table>

A) $\hat{C} = 0.527 + 0.873E$
B) $\hat{C} = 2.51 + 0.529E$
C) $\hat{C} = 0.711 + 0.346E$
D) $\hat{C} = 0.873 + 0.627E$
E) $\hat{C} = 0.0065 + 0.879E$

34) Managers rate employees according to job performance and attitude. The results for several randomly selected employees are given below.

<table>
<thead>
<tr>
<th>Attitude</th>
<th>59</th>
<th>63</th>
<th>65</th>
<th>69</th>
<th>58</th>
<th>77</th>
<th>76</th>
<th>69</th>
<th>70</th>
<th>64</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance</td>
<td>72</td>
<td>67</td>
<td>78</td>
<td>82</td>
<td>75</td>
<td>87</td>
<td>92</td>
<td>83</td>
<td>87</td>
<td>78</td>
</tr>
</tbody>
</table>

A) Performance = 11.7 + 1.02 Attitude
B) Performance = 92.3 - 0.669 Attitude
C) Performance = 100.3 - 0.453 Attitude
D) Performance = 2.81 + 1.35 Attitude
E) Performance = - 47.3 + 2.02 Attitude
Explain what is wrong with each interpretation. Assume calculations are done correctly.
35) A biology student does a study to investigate the association between the amount of sunlight and the number of roses on a rosebush in one summer. (The R^2 value is 58%) He claims that the amount of sunlight determines 58% of the number of roses on a rosebush in one summer.
A) The R^2 has to be greater than 90% to make a statement like this.
B) The amount of sunlight accounts for 58% of the variation in the number of roses. It does not determine the number of roses.
C) The amount of sunlight will increase the number of roses 58% of the time.
D) There is nothing wrong with the interpretation.
E) The amount of variation in sunlight changes 58% of the time. This tells us nothing about the number of roses.

Use the given data to find the equation of the regression line. Round to 3 significant digits, if necessary.
36) The relationship between the number of games won by a minor league baseball team (x) and the average attendance at their home games (y) is analyzed. The mean number of games won was 69 with a standard deviation of 16. The mean attendance was 6452 with a standard deviation of 1626. The correlation between the games won and attendance was 0.52.
A) attendance =1080 +77.8 wins
B) attendance =2810 +52.8 wins
C) attendance = -4090 +153 wins
D) attendance =6450 +0.00512 wins
E) attendance = -560 +102 wins

37) The relationship between the price of yachts (y) and their length (x) is analyzed. The mean length was 46 feet with a standard deviation of 13. The mean price was $86,000 with a standard deviation of 17,000. The correlation between the price and the length was 0.37.
A) price =73,000 +0.000283 length
B) price =63,700 +484 length
C) price = -4,540,000 +584 length
D) price = -1,090,000 +509 length
E) price =25,800 +1310 length

Answer the question appropriately.
38) The relationship between the number of games won by a minor league baseball team and the average attendance at their home games is analyzed. A regression analysis to predict the average attendance from the number of games won gives the model attendance = -2800 +195 wins. The Buckley Bobcats averaged 4489 fans at each game. They won 41 times. Calculate the residual and explain what it means.
A) 706 people. The Buckley Bobcats averaged 706 more fans than would be predicted for a team with 41 wins.
B) 4474 people. On average the Buckley Bobcats will have 4474 extra people.
C) 5195 people. The Buckley Bobcats were expected to average 5195 people for each game.
D) 15,284 people. The Buckley Bobcats averaged 15,284 less fans than would be predicted for a team with 41 wins.
E) -706 people. The Buckley Bobcats averaged 706 less fans than would be predicted for a team with 41 wins.
39) Using advertised prices for used Ford Escorts a linear model for the relationship between a car’s age and its price is found. The regression has an $R^2 = 87.9\%$. Why doesn't the model explain 100% of the variation in the price of an Escort?

A) The model was calculated incorrectly. It should explain all the variation in price.
B) The model is only right 87.9\% of the time.
C) 12.1\% of the time the buyer is getting ripped off by an unscrupulous seller.
D) The prices of all used Ford Escorts were not used.
E) There are other factors besides age that affect the price. These include things such as mileage, options, and condition of the car.

Use the model to make the appropriate prediction.

40) The relationship between the number of games won during one season by a minor league baseball team and the average attendance at their home games is analyzed. A regression analysis to predict the average attendance from the number of games won gives the model $\hat{\text{attendance}} = -2000 + 207$ wins. Predict the average attendance of a team with 460 wins. Explain any possible problems with this prediction.

A) 97,220 people. A team doesn't play that many games and their stadiums probably can't hold that many people.
B) 12 people. There are other factors besides number of games won.
C) 95,220 people. It is only an estimate.
D) 93,220 people. A team doesn't play that many games and their stadiums probably can't hold that many people.
E) 7522 people. There is no problem with this prediction.

41) The relationship between the selling price (in dollars) of used Ford Escorts and their age (in years) is analyzed. A regression analysis to predict the price from the age gives the model $\hat{\text{price}} = 13,996 - 1479$ age. You want to sell a 17 year old Escort. Use the model to determine an appropriate price. Explain any problems.

A) -$11,147 You won't sell a car for a negative amount. The model doesn't give meaningful prices for Escorts this old.
B) -$25,143 You won't sell a car for a negative amount. The model doesn't give meaningful prices for Escorts this old.
C) $9 The car should be worth more than this.
D) $25,143 There is no way the car is worth this much.
E) -$39,139 There are no problems with this prediction.

Answer the question appropriately.

42) A random sample of 150 yachts sold in the United States last year was taken. A regression to predict the price (in thousands of dollars) from length (in feet) has an $R^2 = 16.20\%$. What would you predict about the price of the yacht whose length was one standard deviation above the mean?

A) The price should be 1 SD above the mean in price.
B) The price should be 0.402 SDs above the mean in price.
C) The price should be 0.915 SDs above the mean in price.
D) The price should be 1 SD below the mean in price.
E) The price should be 0.805 SDs above the mean in price.
Use the given data to find the equation of the regression line. Round to 3 significant digits, if necessary.

43) Ten Jeep Cherokee classified ads were selected. The age and prices of several used Ford Escorts are given in the table.

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$19,000</td>
</tr>
<tr>
<td>1</td>
<td>$18,500</td>
</tr>
<tr>
<td>2</td>
<td>$16,000</td>
</tr>
<tr>
<td>3</td>
<td>$13,000</td>
</tr>
<tr>
<td>3</td>
<td>$12,600</td>
</tr>
<tr>
<td>4</td>
<td>$10,000</td>
</tr>
<tr>
<td>4</td>
<td>$9,000</td>
</tr>
<tr>
<td>5</td>
<td>$6,000</td>
</tr>
<tr>
<td>6</td>
<td>$4,000</td>
</tr>
<tr>
<td>6</td>
<td>$2,900</td>
</tr>
</tbody>
</table>

A) \( \hat{\text{price}} = 22000 - 3110 \text{ age} \)
B) \( \hat{\text{price}} = -3110 + 22000 \text{ age} \)
C) \( \hat{\text{price}} = 19000 - 3000 \text{ age} \)
D) \( \hat{\text{price}} = 7.05 - 0.000319 \text{ age} \)
E) \( \hat{\text{price}} = 17200 - 891 \text{ age} \)

44) A golf ball was dropped from 8 different heights. The drop height and the bounce height were recorded.

<table>
<thead>
<tr>
<th>Drop Height (in.)</th>
<th>Bounce Height (in.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>96</td>
<td>73</td>
</tr>
<tr>
<td>84</td>
<td>65</td>
</tr>
<tr>
<td>72</td>
<td>55</td>
</tr>
<tr>
<td>60</td>
<td>46</td>
</tr>
<tr>
<td>48</td>
<td>38</td>
</tr>
<tr>
<td>36</td>
<td>29</td>
</tr>
<tr>
<td>24</td>
<td>19</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
</tr>
</tbody>
</table>

A) \( \hat{\text{bounce}} = -0.335 + 1.305 \text{ drop} \)
B) \( \hat{\text{bounce}} = 0.321 + 0.765 \text{ drop} \)
C) \( \hat{\text{bounce}} = 0.95 - 9.1 \text{ drop} \)
D) \( \hat{\text{bounce}} = 73 - 0.765 \text{ drop} \)
E) \( \hat{\text{bounce}} = 0.215 + 0.866 \text{ drop} \)
45) Ten students in a graduate program were randomly selected. Their grade point averages (GPAs) when they entered the program were between 3.5 and 4.0. The following data were obtained regarding their GPAs on entering the program versus their current GPAs.

<table>
<thead>
<tr>
<th>Entering GPA (E)</th>
<th>Current GPA (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5</td>
<td>3.6</td>
</tr>
<tr>
<td>3.8</td>
<td>3.7</td>
</tr>
<tr>
<td>3.6</td>
<td>3.9</td>
</tr>
<tr>
<td>3.6</td>
<td>3.6</td>
</tr>
<tr>
<td>3.5</td>
<td>3.9</td>
</tr>
<tr>
<td>3.9</td>
<td>3.8</td>
</tr>
<tr>
<td>4.0</td>
<td>3.7</td>
</tr>
<tr>
<td>3.9</td>
<td>3.9</td>
</tr>
<tr>
<td>3.5</td>
<td>3.8</td>
</tr>
<tr>
<td>3.7</td>
<td>4.0</td>
</tr>
</tbody>
</table>

A) \( \hat{C} = 4.91 + 0.0212E \)
B) \( \hat{C} = 3.67 + 0.0313E \)
C) \( \hat{C} = 5.81 + 0.497E \)
D) \( \hat{C} = 4.23 + 0.746E \)
E) \( \hat{C} = 2.51 + 0.329E \)

**Answer the question appropriately.**

46) If you create a regression model for predicting the weight of a motorcycle (in pounds) from its length (in feet), is the slope most likely to be 0.8, 8, 80, 800, or 8000?

A) 8000  B) 8  C) 80  D) 0.8  E) 800

**Tell what the residual plot indicates about the appropriateness of the linear model that was fit to the data.**

47)

A) Model may not be appropriate. The spread is changing.
B) Model is appropriate.
C) Model is not appropriate. The relationship is nonlinear.
Explain what is wrong with each interpretation. Assume calculations are done correctly.

48) A psychologist does an experiment to determine whether an outgoing person can be identified by his or her handwriting. She claims that the $R^2$ of 89% shows that this linear model is appropriate.

   A) $R^2$ does not tell whether the model is appropriate, but measures the strength of the linear relationship. High $R^2$ could also be due to an outlier.
   B) $R^2$ does not tell whether the model is appropriate, but gives the percentage of data points that are close to the model. You can sometimes have a high $R^2$ with a nonlinear relationship.
   C) This $R^2$ means that 89% of the dependent values will fall within one standard deviation of the mean and tells nothing about the appropriateness of the model.
   D) An $R^2$ this high means there is a very weak linear association and the model is probably inappropriate.
   E) There is nothing wrong with the interpretation.

Use the model to make the appropriate prediction.

49) The relationship between the number of games won by a minor league baseball team and the average attendance at their home games is analyzed. A regression analysis to predict the average attendance from the number of games won gives the model attendance $\hat{=} -2400 +199$ wins. Predict the average attendance of a team with 59 wins.

   A) 12 people
   B) 11,741 people
   C) 9341 people
   D) -2142 people
   E) 14,141 people

Answer the question appropriately.

50) A random sample of records of electricity usage of homes in the month of July gives the amount of electricity used and size (in square feet) of 135 homes. A regression was done to predict the amount of electricity used (in kilowatt-hours) from size. The residuals plot indicated that a linear model is appropriate. The model is $\hat{y} = 1244 +0.8 \text{size}$. What would a negative residual mean for people living in a house that is 2245 square feet?

   A) Their house is bigger than expected.
   B) They are using less electricity than expected.
   C) They are using the least amount of electricity of all of the houses sampled.
   D) They are using more electricity than expected.
   E) Their house is smaller than expected.

Answer the question.

51) A teacher studied students' grades and established a strong positive association between SAT scores and college grades. Describe three different possible cause-and-effect relationships that might be present.

   A) Perhaps higher SAT scores cause higher college grades, higher college grades cause higher SAT scores, or both could be caused by a lurking variable such as the students' IQ.
   B) There is only one cause-and-effect relationship: higher SAT scores cause higher college grades.
   C) There are only two cause-and-effect relationships: higher SAT scores cause higher college grades, or higher college grades cause higher SAT scores.
   D) Perhaps higher SAT scores cause higher college grades, higher college grades cause higher SAT scores, or both could be caused by a lurking variable such as the students' high school.
   E) There is only one cause-and-effect relationship: higher college grades cause higher SAT scores.
52) The scatterplot below displays the average longevity (in years) plotted against gestation (in days) for a number of different mammals. For what range of gestation lengths is a linear model appropriate?

![Scatterplot of longevity vs. gestation]

A) A single linear model is appropriate for the entire data set.
B) A linear model should be used for each pair of adjacent data points.
C) One linear model for 0 through 100 days and another linear model for 150 through 350 days.
D) A linear model should not be used for any part of the data.
E) One linear model for 0 through 200 days and another linear model for 200 through 350 days.

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

54) A reporter studied the causes of a fire to a house, and established a strong positive association between the damages (in dollars) and the number of firefighters at the scene. Describe three different possible cause-and-effect relationships that might be present.
A) There are only two cause-and-effect relationships: damages cause more firefighters at the scene, or firefighters cause more damages at the scene.
B) There is only one cause-and-effect relationship: the damages cause more firefighters at the scene.
C) Perhaps the damages cause more firefighters at the scene, firefighters cause more damages at the scene, or both could be caused by a lurking variable such as the size of the blaze.
D) Perhaps the damages cause more firefighters at the scene, firefighters cause more damages at the scene, or both could be caused by a lurking variable such as the type of the building.
E) There is only one cause-and-effect relationship: firefighters cause more damages at the scene.
55) The scatterplot below displays the total home attendance (in millions) for major league baseball’s National League for the years 1960 through 2002. This total home attendance is the grand total of all attendees at all National League games during the season. For what range of years is a linear model appropriate?

A) A linear model should be used for each pair of adjacent data points.
B) A single linear model is appropriate for the entire data set.
C) A linear model should not be used for any part of the data set.
E) None of these

56) Which of the following scatterplots of residuals suggests that a linear model may not be applicable?

A) II, III
B) I
C) IV
D) I, IV
E) None of the above
Solve the problem.

57) The figure below shows the recent trend in first-birth rate for American women between the ages of 18 and 19. (The first-birth rate is the number of 18 to 19 year-olds per 1000 who give birth to their first child).

The regression analysis of this data yields the following values:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1509.5</td>
</tr>
<tr>
<td>Year</td>
<td>-0.725</td>
</tr>
</tbody>
</table>

\( R^2 = 0.6174 \)

Use this model to predict the first-birth rate for 18 to 19 year-olds in 2007.

A) 50 per 1000  
B) 60 per 1000  
C) 54 per 1000  
D) 57 per 1000  
E) 52 per 1000

Answer the question.

58) The scatterplot below displays the yearly production in millions of pounds of flue-cured tobacco in the U.S. For what range of years is a linear model appropriate?

A) One linear model for 1919 through about 1960 and another linear model for about 1960 through 2000.  
B) A single linear model is appropriate for the entire data set.  
C) A linear model should be used for each pair of adjacent data points.  
D) A linear model should not be used for any part of the data.  
E) None of these
59) Which of the labeled points below are outliers?

A) Points A, B, C, and D  
B) Points B and D  
C) Point D  
D) Points A, B, and D  
E) Points A and B

Solve the problem.
60) The figure below shows the life expectancy for persons living in the United States.

The regression analysis of the data yields the following values:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-353.87</td>
</tr>
<tr>
<td>Year</td>
<td>0.2157</td>
</tr>
</tbody>
</table>

\[ R^2 = 0.9539 \]

Use the regression model to predict the life expectancy (to the nearest year) in 2010.

A) 82 years  
B) 83 years  
C) 79 years  
D) 78 years  
E) 80 years
Answer the question.

61) 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)?

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

62) A study of consumer behavior finds a strong positive association between sales of ice cream and sales of soda. Describe three different possible cause-and-effect relationships that might be present.

A) Perhaps sales of ice cream cause higher sales of soda, sales of soda cause higher sales of ice cream, or both could be caused by a lurking variable such as the outdoor temperature.
B) There are only two cause-and-effect relationships: sales of ice cream cause higher sales of soda, or sales of soda cause higher sales of ice cream.
C) There are no possible cause-and-effect relationship, because there should be an arithmetic mistake.
D) There is only one cause-and-effect relationship: sales of soda cause higher sales of ice cream.
E) There is only one cause-and-effect relationship: sales of ice cream cause higher sales of soda.
63) A college admissions officer, defending the college's use of SAT scores in the admissions process, produced the graph below. It represents the mean GPAs for last year's freshmen, grouped by SAT scores. It shows that increased SAT score is associated with increased GPA. What concerns you about the graph, the statistical methodology, or the conclusion reached?

A) The statistical methodology is a concern, because there may be lurking variables.
B) The statistical methodology is a concern, because SAT is not an adequate predictor of GPA on average.
C) The statistical methodology is a concern, because the GPA data is based on mean GPAs, not individual data. We also don't know the number of students in each SAT category.
D) The conclusion reached is a concern, because it should be increased GPA is associated with increased SAT score.
E) The conclusion reached is a concern, because there are no relationships between SAT scores and GPA.

64) The scatterplot below shows the percentage of the US population that is foreign born for the years 1910 - 1990. For what range of years is a linear model appropriate?

A) One linear model is appropriate for the years 1910 through 1950 and another linear model for the years 1950 through 1990.
B) One linear model is appropriate for the years 1910 through 1970 and another linear model for the years 1970 through 1990.
C) A linear model should be used for each pair of adjacent data points.
D) A linear model should not be used for any part of the data.
E) A single linear model is appropriate for the entire data set.
65) The figure below shows the association between female life expectancy and the average number of children women give birth to for several different countries. Also shown is the equation and correlation from a regression analysis. What is the correct conclusion to draw from the figure?

A) High average number of children women give birth to is causing reduced life expectancy, probably because of the increased fatigue and emotional stress exerted on mothers.

B) While there appears to be a very strong association, there is probably not a cause-and-effect relationship between female life expectancy and the average number of children women give birth to. Access to basic health care is probably a lurking variable that drives both female life expectancy and the average number of children women give birth to.

C) Those countries with low life expectancies clearly have no regard for children or expectant mothers.

D) Countries that have low life expectancies and high average number of children women give birth to seem to have less regard for the sanctity of human life.

E) The association must be coincidental. I would expect the association to have a positive slope, not the negative one illustrated above.
66) Which of the following scatterplots of residuals suggests that a linear model may not be applicable?

A) III
B) IV
C) II
D) I
E) None of the above

Solve the problem.

67) The table below displays the latitude (degrees north) and average daily minimum temperature in January (in degrees Fahrenheit) for some cities located in the northern hemisphere.

<table>
<thead>
<tr>
<th>Latitude (Deg North)</th>
<th>Average Daily Min Temp in Jan (°F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Athens, Greece</td>
<td>38</td>
</tr>
<tr>
<td>Bombay, India</td>
<td>19</td>
</tr>
<tr>
<td>Cairo, Egypt</td>
<td>30</td>
</tr>
<tr>
<td>London, England</td>
<td>51</td>
</tr>
<tr>
<td>Mexico City, Mexico</td>
<td>19</td>
</tr>
<tr>
<td>Moscow, Russia</td>
<td>55</td>
</tr>
<tr>
<td>Paris, France</td>
<td>49</td>
</tr>
<tr>
<td>Bangkok, Thailand</td>
<td>13</td>
</tr>
<tr>
<td>Tokyo, Japan</td>
<td>35</td>
</tr>
<tr>
<td>Manilla, Phillipines</td>
<td>14</td>
</tr>
</tbody>
</table>

The scatter plot and regression equation are shown below:
The regression analysis of this data yields the following values:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>76.94</td>
</tr>
<tr>
<td>Latitude</td>
<td>-1.038</td>
</tr>
</tbody>
</table>

R^2 = 0.7660

Use this model to predict the average daily minimum temperature (to the nearest tenth of a degree) in January for Panama City whose latitude is 9 degrees north.

A) 86.3°F  B) 67.6°F  C) 61.5°F  D) 54.0°F  E) 9.3°F

Answer the question.

68) A study finds a strong positive association between sizes of children feet and results in spelling tests. Describe three different possible cause-and-effect relationships that might be present.

A) Perhaps older children cause better spelling tests, spelling tests cause older children, or both could be caused by a lurking variable such as the school.
B) There are only two cause-and-effect relationships: older children cause better spelling tests, spelling tests cause older children.
C) Perhaps older children cause better spelling tests, spelling tests cause older children, or both could be caused by a lurking variable such as the age.
D) There are only one cause-and-effect relationship: older children cause better spelling tests, because older children do better on spelling tests than younger children, on average.
E) There are no possible cause-and-effect relationships.
69) Which of the following scatterplots of residuals suggests that a linear model may not be applicable?

A) I
B) III
C) IV
D) II
E) None of the above

70) Which of the labeled points below are outliers?

A) Points A, B, C, and D
B) Points C and D
C) Points A and C
D) Point A
E) Points A, C, and D
71) Which of the labeled points below are influential points?

A) Point C
B) Point A
C) Points B and D
D) Points A, B, C, and D
E) Points A, B, and D
Solve the problem.

72) The table below shows the gestation (in days) and average longevity (in years) for a number of different mammals:

<table>
<thead>
<tr>
<th>Mammal</th>
<th>Gestation (days)</th>
<th>Average Longevity (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black Bear</td>
<td>219</td>
<td>18</td>
</tr>
<tr>
<td>Cat (domestic)</td>
<td>63</td>
<td>12</td>
</tr>
<tr>
<td>Monkey (Rhesus)</td>
<td>166</td>
<td>15</td>
</tr>
<tr>
<td>Lion</td>
<td>100</td>
<td>15</td>
</tr>
<tr>
<td>Horse</td>
<td>330</td>
<td>20</td>
</tr>
<tr>
<td>Gorilla</td>
<td>258</td>
<td>20</td>
</tr>
<tr>
<td>Gray Squirrel</td>
<td>44</td>
<td>10</td>
</tr>
</tbody>
</table>

The scatter plot and regression equation are shown below:

The regression analysis of this data yields the following values:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>9.90</td>
</tr>
<tr>
<td>Gestation</td>
<td>0.0345</td>
</tr>
</tbody>
</table>

\[R^2 = 0.9048\]

Use this model to predict the average longevity (to the nearest tenth of a year) of an African elephant whose gestation is 660 days.

A) 38.7 years    B) 32.7 years    C) 41.9 years    D) 51.2 years    E) 22.8 years
73) 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)?

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

74) The scatterplot below displays world population (in millions) for the years 0 - 1800. Where the population is an estimate, the lower estimate is given. For what range of years is a linear model appropriate?

A) A single linear model is appropriate for the entire data set.
B) One linear model is appropriate for the years 0 through 600 and another linear model for the years 600 through 1800.
C) A linear model should be used for each pair of adjacent data points.
D) A linear model should not be used for any part of the data.
E) One linear model is appropriate for the years 0 through 1000 and another linear model for the years 1400 through 1800.
75) Over a period of years, a certain town observed the association between the number of people attending churches and the number of people in the city jail. The results are shown on the figure below. Also shown are the equation and $R^2$ value from a linear regression analysis. What is the best conclusion to draw from the figure?

![Graph showing the relationship between people attending churches and people in jail. The equation is $y = 0.0216x + 62.672$ with $R^2 = 0.7954$.]

- A) Although the association is negatively strong, going to church does not cause people to go to jail. Instead, size of the population in the town is probably a lurking variable that drives both the number of people in jail and the number of people attending church to increase together.
- B) Clearly, there must be some as-yet unknown problems associated with going to church.
- C) More people go to church not to go to jail.
- D) More detainees should go to church to improve their behavior.
- E) Although the association is positively strong, going to church does not cause people to go to jail. Instead, size of the population of the town is probably a lurking variable that drives both the number of people in jail and the number of people attending church to increase together.

Use the given data to find the equation of the regression line. Round to 3 significant digits, if necessary.

76) The relationship between the cost of a taxi ride ($y$) and the length of the ride ($x$) is analyzed. The mean length was 4.8 miles with a standard deviation of 1.3. The mean cost was $8.19 with a standard deviation of 1.7. The correlation between the cost and the length was 0.81.

- A) $\hat{\text{cost}} = 5.22 + 0.619 \text{ length}$
- B) $\hat{\text{cost}} = -477 + 101 \text{ length}$
- C) $\hat{\text{cost}} = 3.11 + 1.06 \text{ length}$
- D) $\hat{\text{cost}} = 1.91 + 1.31 \text{ length}$
- E) $\hat{\text{cost}} = -117 + 26.1 \text{ length}$

Fill in the missing information.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$s_x$</th>
<th>$y$</th>
<th>$s_y$</th>
<th>$r$</th>
<th>$\hat{y} = b_0 + b_1x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>19</td>
<td>-3</td>
<td>0.6</td>
<td>-</td>
<td>$\hat{y} = 30 - 4x$</td>
</tr>
</tbody>
</table>

- A) $\bar{x} = 12.25; s_x = 0.90$
- B) $\bar{x} = 49; s_x = -14.25$
- C) $\bar{x} = 11; s_x = 1.80$
- D) $\bar{x} = -46; s_x = 11.50$
- E) $\bar{x} = 2.75; s_x = 0.45$
Answer the question appropriately.

78) A random sample of 150 yachts sold in the United States last year was taken. A regression to predict the price (in thousands of dollars) from length (in feet) has an $R^2 = 15.8\%$. What would you predict about the price of the yacht whose length was two standard deviations below the mean?

A) The price should be 0.795 SDs below the mean in price.
B) The price should be 1 SD below the mean in price.
C) The price should be 1 SD above the mean in price.
D) The price should be 1.835 SDs below the mean in price.
E) The price should be 0.397 SDs below the mean in price.

79) A random sample of records of electricity usage of homes gives the amount of electricity used in July and size (in square feet) of 135 homes. A regression to predict the amount of electricity used (in kilowatt-hours) from size was completed. The residuals plot indicated that a linear model is appropriate. What are the variables and units in this regression?

A) Amount of electricity used (in kilowatt-hours) is $y$ and number of homes is $x$.
B) Number of homes is $y$ and amount of electricity used (in kilowatt-hours) is $x$.
C) Size (in square feet) is $y$ and number of homes is $x$.
D) Size (in square feet) is $y$ and amount of electricity used (in kilowatt-hours) is $x$.
E) Amount of electricity used (in kilowatt-hours) is $y$ and size (in square feet) is $x$.

Use the given data to find the equation of the regression line. Round to 3 significant digits, if necessary.

80) Two different tests are designed to measure employee productivity and dexterity. Several employees are randomly selected and tested with these results.

<table>
<thead>
<tr>
<th>Dexterity</th>
<th>23</th>
<th>25</th>
<th>28</th>
<th>21</th>
<th>21</th>
<th>25</th>
<th>26</th>
<th>30</th>
<th>34</th>
<th>36</th>
</tr>
</thead>
<tbody>
<tr>
<td>Productivity</td>
<td>49</td>
<td>53</td>
<td>59</td>
<td>42</td>
<td>47</td>
<td>53</td>
<td>55</td>
<td>63</td>
<td>67</td>
<td>75</td>
</tr>
</tbody>
</table>

A) Productivity = 6.08 + 1.56 Dexterity
B) Productivity = 10.7 + 1.53 Dexterity
C) Productivity = 5.05 + 1.91 Dexterity
D) Productivity = 2.36 + 2.03 Dexterity
E) Productivity = 75.3 - 0.329 Dexterity

Answer the question appropriately.

81) If you create a regression model for estimating the price of a salad based upon its weight (oz), is the slope most likely to be 0.2, 2, 20, 200, or 2000?

A) 0.2  B) 2000  C) 200  D) 20  E) 2

82) A golf ball is dropped from 15 different heights (in inches) and the height of the bounce is recorded (in inches.)

The regression analysis gives the model bounce = -0.4 + 0.75 drop. Interpret the meaning of the $y$-intercept.

A) According to the model, a ball dropped from 0 inches high will bounce - 0.4 inches. (This may not actually happen.)
B) According to the model, a ball dropped from 0.75 inches high will bounce 0 inches. (This may not actually happen.)
C) According to the model, a ball dropped from - 0.4 inches high will bounce 0 inches. (This may not actually happen.)
D) According to the model, a ball dropped from 0.75 inches high will bounce - 0.4 inches. (This may not actually happen.)
E) According to the model, a ball dropped from 0 inches high will bounce 0.75 inches. (This may not actually happen.)
83) The relationship between the number of games won by a minor league baseball team and the average attendance at their home games is analyzed. A regression to predict the average attendance from the number of games won has an $R^2 = 31.1\%$. The residuals plot indicated that a linear model is appropriate. What is the correlation between the average attendance and the number of games won.

A) 0.311  B) 0.097  C) 0.830  D) 0.689  E) 0.558

Use the model to make the appropriate prediction.

84) A golf ball is dropped from 15 different heights (in inches) and the height of the bounce is recorded (in inches.) The regression analysis gives the model $\hat{y} = -0.4 + 0.72 \cdot \text{drop}$. Predict the height of the bounce if dropped from 65 inches.

A) 46.4 inches  B) 47.2 inches  C) 46.8 inches  D) 65.32 inches  E) 90.83 inches

Fill in the missing information.

85) $\begin{array}{cccc} x & s_x & y & s_y & r & \hat{y} = b_0 + b_1 x \\ 13 & 1 & 25 & 3 & 0.4 & \hat{y} = ? \end{array}$

A) $\hat{y} = 3 + 1.92x$  B) $\hat{y} = 1.2 + 9.4x$  C) $\hat{y} = -14 + 3x$  D) $\hat{y} = 23.27 + 0.13x$  E) $\hat{y} = 9.4 + 1.2x$

Tell what the residual plot indicates about the appropriateness of the linear model that was fit to the data.

86) [Residual plot image]

A) Model is not appropriate. The relationship is nonlinear.  
B) Model is appropriate.  
C) Model may not be appropriate. The spread is changing.

Use the model to make the appropriate prediction.

87) The relationship between the selling price (in dollars) of used Ford Escorts and their age (in years) is analyzed. A regression analysis to predict the price from the age gives the model $\hat{y} = 14,489 - 1337 \cdot \text{age}$. Predict the price of an Escort that is 2 years old.

A) $11,815$  B) $11$  C) $2674$  D) $17,163$  E) $13,154$
Answer the question.

88) 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
   A) an outlier.
   B) a lurking variable.
   C) regression.
   D) leverage.
   E) extrapolation.

89) Which of the labeled points below will exert the largest leverage on a linear model of the data?

90) The figure below examines the association between life expectancy and computer ownership for several countries. Also shown are the equation and \( R^2 \) value from a linear regression analysis. What is the best conclusion to draw from the figure?

   A) Persons who live longer buy more computers over the course of their longer lifetimes.
   B) Exposure to the radiation from computer monitors is causing a clear decline in life expectancy.
   C) Computer ownership promotes health and long life, probably due to the greater access that computer owners have to health information on the world-wide web.
   D) Although the association is strong, computer ownership probably does not promote longevity. Instead, national per capita wealth is probably a lurking variable that drives both life expectancy and computer ownership.
   E) Clearly, there must be some as-yet unknown health benefit associated with using computers.
91) Which of the labeled points below are influential points?

A) Points A and C
B) Points A and D
C) Points A, C, and D
D) Points A, B, C, and D
E) Point D

92) Which of the following scatterplots of residuals suggests that a linear model may not be applicable?

A) III
B) II
C) IV
D) I
E) None of the above

Solve the problem.
93) The table below shows the age and annual income of 12 randomly selected college graduates all living in the city of Seattle.

<table>
<thead>
<tr>
<th>Age</th>
<th>Annual Income (dollars)</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>28,520</td>
</tr>
<tr>
<td>31</td>
<td>36,750</td>
</tr>
<tr>
<td>55</td>
<td>72,155</td>
</tr>
<tr>
<td>47</td>
<td>43,225</td>
</tr>
<tr>
<td>38</td>
<td>34,197</td>
</tr>
<tr>
<td>50</td>
<td>60,030</td>
</tr>
<tr>
<td>29</td>
<td>25,005</td>
</tr>
<tr>
<td>23</td>
<td>31,625</td>
</tr>
<tr>
<td>33</td>
<td>55,975</td>
</tr>
<tr>
<td>40</td>
<td>37,064</td>
</tr>
<tr>
<td>52</td>
<td>75,082</td>
</tr>
<tr>
<td>25</td>
<td>19,055</td>
</tr>
</tbody>
</table>

The scatter plot and regression equation are shown below:

![Scatter plot with regression line]

The regression analysis of this data yields the following values:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-8447.05</td>
</tr>
<tr>
<td>Age</td>
<td>1380.95</td>
</tr>
</tbody>
</table>

\[ R^2 = 0.7182 \]

Use this model to predict the annual income (to the nearest dollar) of a 56 year old college graduate living in Seattle.

A) $61,367  B) $77,333  C) $70,230  D) $85,780  E) $68,886
Answer the question appropriately.

94) When using midterm exam scores to predict a student's final grade in a class, the student would prefer to have a
   
   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 higher than we would predict with the model.
   E) negative residual, because that means the student's final grade is lower than we would predict with the model.

95) A random sample of 150 yachts sold in the United States last year was taken. A regression analysis to predict the price (in thousands of dollars) from length (in feet) was completed. A linear model is appropriate. What are the units of the slope?
   
   A) Slope is dollars per foot.
   B) Slope is feet per thousand dollars.
   C) Slope is yachts per dollar.
   D) Slope is thousands of dollars per foot.
   E) Slope is feet per dollar.

Tell what the residual plot indicates about the appropriateness of the linear model that was fit to the data.

96) A) Model is appropriate.
   B) Model is not appropriate. The relationship is nonlinear.
   C) Model may not be appropriate. The spread is changing.
97) The figure below shows the association between life expectancy and infant mortality for several different countries. Also shown is the equation and correlation from a regression analysis. What is the correct conclusion to draw from the figure?

A) High infant mortality is causing reduced life expectancy, probably because of the increased emotional stress exerted on parents who have lost a child at birth.
B) Those countries with low life expectancies clearly have no regard for children or expectant mothers.
C) The association must be coincidental. I would expect the association to have a positive slope, not the negative one illustrated above.
D) Countries that have low life expectancies and high infant mortality rates seem to have less regard for the sanctity of human life.
E) While there appears to be a very strong association, there is probably not a cause-and-effect relationship between infant mortality and life expectancy. Access to basic health care is probably a lurking variable that drives both life expectancy and infant mortality.

98) Which of the labeled points below will exert the largest leverage on a linear model of the data?

A) Point A  B) Point D  C) Point C  D) Point E  E) Point B
The relationship between the number of games won by a minor league baseball teams and the average attendance at their home games is analyzed. A regression to predict the average attendance from the number of games won has an $r = 0.75$. Interpret this statistic.

A) Positive, fairly strong linear relationship. 56.25% of the variation in average attendance is explained by the number of games won.
B) Negative, fairly strong linear relationship. 56.25% of the variation in average attendance is explained by the number of games won.
C) Positive, fairly strong linear relationship. 75% of the variation in average attendance is explained by the number of games won.
D) Positive, weak linear relationship. 6.25% of the variation in average attendance is explained by the number of games won.
E) Negative, weak linear relationship. 6.25% of the variation in average attendance is explained by the number of games won.

An economist noticed that nations with more TV sets have higher life expectancy. He established a strong positive association between length of life and number of TV sets. Describe three different possible cause-and-effect relationships that might be present.

A) Perhaps more TV sets cause higher life expectancy, higher life expectancy cause more TV sets, or both could be caused by a lurking variable such as the wealth of the nation.
B) Perhaps more TV sets cause higher life expectancy, higher life expectancy cause more TV sets, or both could be caused by a lurking variable such as the TV sets brand.
C) There are only two cause-and-effect relationships: more TV sets cause higher life expectancy, or higher life expectancy cause more TV sets.
D) There is only one cause-and-effect relationship: more TV sets cause higher life expectancy.
E) There is only one cause-and-effect relationship: higher life expectancy cause more TV sets.