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

Provide an appropriate response.

1) In 2007, the number of wins had a mean of 81.79 with a standard deviation of 10.89 for the teams of baseball’s American league. The equation that predicts the number of wins (y) using the number of runs allowed (x) is \( \hat{y} = 59.62 - 0.10x \). What is the predicted number of wins for a team that allowed 800 runs? Round your answer to the nearest integer.

A) 160  
B) 80  
C) 82  
D) 168

2) Based on findings from the Health and Nutrition Examination Survey conducted by the National Center for Health Statistics from April 1971 to June 1974, the regression equation predicting the average weight of a male aged 18-24 (y) based on his height (x) is given by \( \hat{y} = -172.63 + 4.842x \). (www.cdc.gov/nchs/data/ad/ad014acc.pdf) Interpret the slope of the regression line.

A) for every unit increase in height, the predicted weight increases by 4.842 pounds  
B) for every unit increase in weight, the predicted height decreases by 4.842 pounds  
C) for every unit increase in height, the predicted weight decreases by 4.842 pounds  
D) for every unit increase in weight, the predicted height increases by 4.842 pounds

3) The regression equation relating dexterity scores (x) and productivity scores (y) for the employees of a company is \( \hat{y} = 5.50 + 1.91x \). Ten pairs of data were used to obtain the equation. The same data yield \( r = 0.986 \) and \( \bar{y} = 56.3 \). What is the best predicted productivity score for a person whose dexterity score is 20?

A) 43.7  
B) 38.20  
C) 111.91  
D) 58.20  
E) 56.30

4) Nine data points of data yield \( r = 0.867 \) and the regression equation \( \hat{y} = 19.4 + 0.93x \). Also, \( \bar{y} = 64.7 \). What is the best predicted value of y for \( x = 40 \)?

A) 56.6  
B) 37.2  
C) 64.7  
D) 79.6  
E) 57.8

Select the most appropriate answer.

5) The y-intercept is the

A) predicted value of y.  
B) point where the regression line crosses the x-axis.  
C) smallest value for the residual sum of squares.  
D) change in the predicted value of y per unit increase in x.  
E) predicted value of y when x = 0.

6) The slope is the

A) smallest value for the residual sum of squares.  
B) predicted value of y when x = 0.  
C) change in the predicted value of y per unit increase in x.  
D) point where the regression line crosses the y-axis.  
E) predicted value of y.
Provide an appropriate response.

7) Which statement is true about residuals?
   A) Not all observations have residuals.
   B) The larger the absolute value of a residual, the closer the predicted value is to the actual value.
   C) In a scatterplot, the residual for an observation is the horizontal distance between the point and the regression line.
   D) Residuals measure the size of prediction errors.
   E) None of these

8) A regression line for predicting the selling prices of homes in Chicago is \( y = 168 + 102x \), where \( x \) is the square footage of the house. A house with 1500 square feet recently sold for $140,000. What is the residual for this observation?
   A) 13,000   B) -13,000   C) 13,168   D) 1316.80   E) -13,168

9) A regression line for predicting Internet usage (%) for 39 countries is \( \hat{y} = -3.61 + 1.55x \), where \( x \) is the per capita GDP, in thousands of dollars, and \( y \) is Internet usage. What is the residual for a country with a per capita GDP of $28,000 and actual Internet use of 38 percent?
   A) -5.4   B) -1.79   C) 5.4   D) -4.5   E) 1.79

10) A regression line for predicting Internet usage (%) for 39 countries is \( \hat{y} = -3.61 + 1.55x \), where \( x \) is the per capita GDP, in thousands of dollars, and \( y \) is Internet usage. Interpret the residual for one of the 39 countries with per capita GDP of $15,000 and actual Internet use of 20 percent.
    A) The actual Internet usage for this country is 3.25% lower than expected from the regression equation.
    B) The actual Internet usage for this country is 3.25% higher than expected from the regression equation.
    C) The actual Internet usage for this country is 0.36% lower than expected from the regression equation.
    D) The actual Internet usage for this country is 0.36% higher than expected from the regression equation.
    E) The actual Internet usage for this country is 3.6% higher than expected from the regression equation.

11) A regression line for predicting the selling prices of homes in Chicago is \( \hat{y} = 168 + 102x \), where \( x \) is the square footage of the house. Interpret the residual for a house with 1800 square feet that recently sold for $200,000.
    A) The house sold for $16,232 less than was to be expected from the regression equation.
    B) The house sold for $16,064 more than was to be expected from the regression equation.
    C) The house sold for $16,400 less than was to be expected from the regression equation.
    D) The house sold for $16,232 more than was to be expected from the regression equation.
    E) The house sold for $16,400 more than was to be expected from the regression equation.
12) 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 simple linear regression to predict the amount of electricity used (in kilowatt-hours) based on size has an $r^2 = 0.71$. Assume that a linear model is appropriate. Interpret $r^2$.

A) The prediction error for predicting electricity use is about the same when using the regression line and $\bar{y}$.
B) The prediction error using the regression line to predict electricity use is 71% smaller than the prediction error using $\bar{y}$ to predict it.
C) The prediction error using the regression line to predict electricity use is 71% larger than the prediction error using $\bar{y}$ to predict it.
D) The prediction error using the regression line to predict electricity use is 29% larger than the prediction error using $\bar{y}$ to predict it.
E) The prediction error using the regression line to predict electricity use is 29% smaller than the prediction error using $\bar{y}$ to predict it.

13) 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 $r^2 = 0.255$. The residuals plot indicated that a linear model is appropriate. Interpret $r^2$.

A) The prediction error using the regression line to predict attendance is 74.5% larger than the prediction error using $\bar{y}$ to predict it.
B) The prediction error using the regression line to predict attendance is 25.5% larger than the prediction error using $\bar{y}$ to predict it.
C) The prediction error for predicting attendance is about the same when using the regression line and $\bar{y}$.
D) The prediction error using the regression line to predict attendance is 25.5% smaller than the prediction error using $\bar{y}$ to predict it.
E) The prediction error using the regression line to predict attendance is 74.5% smaller than the prediction error using $\bar{y}$ to predict it.

14) 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 $r^2 = 0.256$. Assume that a linear model is appropriate. What is the correlation between the average attendance and the number of games won?

A) 0.863  B) 0.07  C) 0.256  D) 0.744  E) 0.506

Select the most appropriate answer.

15) Among the possible lines that can go through data points in a scatterplot, the regression line results from the least squares method and has the smallest value for the ____________.

A) residual sum  B) intercept  C) residual sum of squares  D) correlation  E) slope
The prediction error for an observation, which is the difference between the actual value and the predicted value of the response variable, is called ____________________.

A) a correlation
B) an extrapolation
C) an intercept
D) a residual
E) an outlier

Fill in the missing information.

17) $\overline{x} \quad s_x \quad \overline{y} \quad s_y \quad r \quad y = a + bx$

17) ______

A) -168 + 4.4x
B) 175 + 2.2x
C) -168 + 2.2x
D) -154 + 4.4x
E) -38 + 4.4x

18) $\overline{x} \quad s_x \quad \overline{y} \quad s_y \quad r \quad y = a + bx$

18) ______

A) $\overline{y} = -70; r = 0.02$
B) $\overline{y} = 104; r = -0.36$
C) $\overline{y} = 214; r = 0.36$
D) $\overline{y} = 6; r = 0.02$
E) $\overline{y} = 6; r = 0.36$

19) $\overline{x} \quad s_x \quad \overline{y} \quad s_y \quad r \quad y = a + bx$

19) ______

A) $\overline{x} = 3; s_x = 1.00$
B) $\overline{x} = 12; s_x = 1.00$
C) $\overline{x} = 3; s_x = 0.50$
D) $\overline{x} = 48; s_x = -18.00$
E) $\overline{x} = 12; s_x = 2.00$

Use a graphing calculator to find the regression line for the given data.

20) $\overline{x} \quad 2 \quad 4 \quad 5 \quad 6$

20) ______

A) $\overline{y} = 0.15 + 2.8x$
B) $\overline{y} = 3.0x$
C) $\overline{y} = 2.8x$
D) $\overline{y} = 3.0x - 0.15$
E) $\overline{y} = 0.15 + 3.0x$
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>x (entering GPA)</th>
<th>y (current GPA)</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) $\hat{y} = 4.91 + 0.0212x$
B) $\hat{y} = 3.67 + 0.0313x$
C) $\hat{y} = 5.81 + 0.497x$
D) $\hat{y} = 3.67 - 0.0313x$
E) $\hat{y} = 2.51 + 0.329x$