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

Determine the regression equation for the data. Round the final values to three significant digits, if necessary.

1) \[ \begin{array}{c|cccc} x & 2 & 4 & 5 & 6 \\ \hline y & 7 & 11 & 13 & 20 \end{array} \]

A) \( \hat{y} = 2.8x \)  
B) \( \hat{y} = 0.15 + 2.8x \)  
C) \( \hat{y} = 3x \)  
D) \( \hat{y} = 0.15 + 3x \)

2) \[ \begin{array}{c|cccc} x & 1 & 3 & 5 & 7 & 9 \\ \hline y & 143 & 116 & 100 & 98 & 90 \end{array} \]

A) \( \hat{y} = 151 - 6.8x \)  
B) \( \hat{y} = 140 - 6.2x \)  
C) \( \hat{y} = -151 + 6.8x \)  
D) \( \hat{y} = -140 + 6.2x \)

3) Ten students in a graduate program were randomly selected. The following data represent their grade point averages (GPAs) at the beginning of the year (x) versus their GPAs at the end of the year (y).

\[ \begin{array}{cc} x & y \\ 3.5 & 3.6 \\ 3.8 & 3.7 \\ 3.6 & 3.9 \\ 3.6 & 3.6 \\ 3.9 & 3.9 \\ 4.0 & 3.7 \\ 3.9 & 3.9 \\ 3.5 & 3.8 \\ 3.7 & 4.0 \end{array} \]

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

Use the regression equation to predict the y-value corresponding to the given x-value. Round your answer to the nearest tenth.

4) Eight pairs of data yield the regression equation \( \hat{y} = 55.8 + 2.79x \). Predict \( y \) for \( x = 3.1 \).

A) 175.8  
B) 64.4  
C) 71.1  
D) 57.8

5) The regression equation relating attitude rating (x) and job performance rating (y) for ten randomly selected employees of a company is \( \hat{y} = 11.7 + 1.02x \). Predict the job performance rating for an employee whose attitude rating is 63.

A) 12.6  
B) 74.9  
C) 76.0  
D) 80.1
The regression equation for the given data points is provided. Graph the regression equation and the data points.

6) \[
\begin{array}{c|cccccc}
 x & 1 & 3 & 5 & 7 & 9 \\
 y & 73 & 46 & 30 & 28 & 20 \\
\end{array}
\]

\[^{\wedge} y = 70.4 - 6.2x\]

- A)
- B)
- C)
- D)
7) Is the data point, P, an outlier, a potential influential observation, both, or neither?

A) Both  B) Outlier  C) Neither  D) Potential influential observation

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

8) Describe what scatter diagrams are, and discuss their importance.

9) Define the terms "predictor variable" and "response variable." Give an example of each.

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

10) True or false? Any point that is an influential observation is also an outlier, while an outlier may or may not be an influential observation.

A) True  B) False

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

11) Sketch a scatter diagram for a set of data points for which it would be appropriate to fit a regression line. Then, sketch a second scatter diagram for a set of data points for which it would not be appropriate to fit a regression line. Explain why it would not be reasonable to fit a regression line for the second scatter diagram.

12) A regression equation is obtained for the following set of data.

<table>
<thead>
<tr>
<th>x</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>9</th>
<th>10</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>28</td>
<td>33</td>
<td>39</td>
<td>45</td>
<td>47</td>
<td>52</td>
</tr>
</tbody>
</table>

For what range of x-values would it be reasonable to use the regression equation to predict the y-value corresponding to a given x-value? Why?
MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Obtain the linear correlation coefficient for the data. Round your answer to three decimal places.

13) \begin{tabular}{|c|c|c|c|c|}
\hline
x & 24.3 & 46.0 & 39.3 & 40.1 \\
\hline
y & 4 & 2 & 5 & 3 \\
\hline
\end{tabular}

A) 0  
B) 0.655  
C) 0.736  
D) -0.736

14) A study was conducted to compare the number of hours spent in the computer lab on an assignment (x) and the grade on the assignment (y), for each of eight randomly selected students in a computer class. The results are recorded in the table below.

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>96</td>
</tr>
<tr>
<td>11</td>
<td>51</td>
</tr>
<tr>
<td>16</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>58</td>
</tr>
<tr>
<td>7</td>
<td>89</td>
</tr>
<tr>
<td>15</td>
<td>81</td>
</tr>
<tr>
<td>16</td>
<td>46</td>
</tr>
<tr>
<td>10</td>
<td>51</td>
</tr>
</tbody>
</table>

A) 0.462  
B) 0.017  
C) -0.335  
D) -0.284

15) Two separate tests, x and y, are designed to measure a student’s ability to solve problems. Several students are randomly selected to take both tests and their results are shown below.

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>48</td>
<td>73</td>
</tr>
<tr>
<td>52</td>
<td>67</td>
</tr>
<tr>
<td>58</td>
<td>73</td>
</tr>
<tr>
<td>44</td>
<td>59</td>
</tr>
<tr>
<td>43</td>
<td>58</td>
</tr>
<tr>
<td>43</td>
<td>56</td>
</tr>
<tr>
<td>40</td>
<td>58</td>
</tr>
<tr>
<td>51</td>
<td>64</td>
</tr>
<tr>
<td>59</td>
<td>74</td>
</tr>
</tbody>
</table>

A) 0.548  
B) 0.109  
C) 0.714  
D) 0.867

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

Provide an appropriate response.

16) What is the relationship between the linear correlation coefficient and the usefulness of the regression equation for making predictions?

17) Suppose data are collected for each of several randomly selected adults for height, in inches, and number of calories burned in 30 minutes of walking on a treadmill at 3.5 mph. How would the value of the linear correlation coefficient, r, change if all of the heights were converted to meters?

18) For each of 200 randomly selected cities, Pete compared data for the number of churches in the city (x) and the number of homicides in the past decade (y). He calculated the linear correlation coefficient and was surprised to find a strong positive linear correlation for the two variables. Does this suggest that when a city builds new churches this will tend to cause an increase in the number of homicides? Why do you think that a strong positive linear correlation coefficient was obtained?
MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

19) Determine which plot shows the strongest linear correlation.
   A) \[\text{Plot A}\]
   B) \[\text{Plot B}\]
   C) \[\text{Plot C}\]
   D) \[\text{Plot D}\]

19) ______
1) C
2) B
3) D
4) B
5) C
6) B
7) A
8) Answers will vary. One possible answer: Scatter diagrams (or scatterplots) are graphs of all data points, with the predictor variable on the horizontal axis and the response variable on the vertical axis. The purpose of a scatter diagram is to help visualize any apparent relationship between the two variables. Before finding a regression line for a set of data points, draw a scatter diagram. If the data points do not appear to be scattered about a straight line, do not determine a regression line.
9) For a linear equation \( y = b_0 + b_1 x \), \( y \) is the dependent variable and \( x \) is the independent variable. However, in the context of regression analysis, we more customarily call \( y \) the response variable and \( x \) the predictor variable or explanatory variable (because it is used to predict or explain the values of the response variable). Examples will vary.
10) B
11) Examples will vary. The points in the first scatter diagram should appear to be scattered about a straight line. The points in the second scatter diagram should not appear to be scattered about a straight line.
12) It would be reasonable to use the regression equation to predict the \( y \)-value corresponding to a given \( x \)-value for \( x \)-values in the range from 2 to 12. We can reasonably use the regression equation to make predictions for values of the predictor variable \( (x) \) within the range of the observed values of the predictor value, in this case from 2 to 12. However, to do so for values of the predictor variable outside that range may not be reasonable, because the linear relationship between the variables may not hold there. Using the regression equation to make predictions for values of the predictor variable outside the range of the observed values of the predictor variable is called extrapolation. Grossly incorrect predictions can result from extrapolation. (Explanations will vary.)
13) D
14) C
15) D
16) The strength of the linear relationship (measured by the linear correlation coefficient) indicates the usefulness of the regression equation for making predictions. The coefficient of determination, \( r^2 \), is a descriptive measure of the utility of the regression equation for making predictions. The coefficient of determination is the square of the linear correlation coefficient \( (r) \).
17) The value of \( r \) would remain the same, as \( r \) is independent of the choice of units.
18) The positive linear correlation coefficient suggests that cities with a lot of churches also tend to have a high number of homicides. However, the fact that there is a correlation does not imply that there is causation. It is unlikely that building new churches would lead to an increase in the number of homicides. It is more likely that the correlation between the two variables is explained by their association with another variable (called a lurking variable), population. Larger cities tend to have both more churches and more homicides than small cities.
19) B