1. In simple linear regression, both the t and F tests can be used as model utility tests.
   - True
   - False

2. The sample correlation coefficient is a measure of the strength of a linear relationship between two continuous variables.
   - True
   - False

3. Many factors affect the length of a professional football game. A study was conducted to determine the relationship between the total number of penalty yards (x) and the time required to complete a game (y, in hours). The following is the summary data:
   \[ n = 9, S_{XX} = 26,256, S_{YY} = 3,956, S_{XY} = 244.8, \text{MSE} = 0.2390 \]
   The expected value of the slope is ______. (6 decimal places)

   Many factors affect the length of a professional football game. A study was conducted to determine the relationship between the total number of penalty yards (x) and the time required to complete a game (y, in hours). The following is the summary data:
   \[ n = 9, S_{XX} = 26,256, S_{YY} = 3,956, S_{XY} = 244.8, \text{MSE} = 0.2390 \]
   The 95% confidence interval for the slope is (_____, _____) (6 decimal places)

   Many factors affect the length of a professional football game. A study was conducted to determine the relationship between the total number of penalty yards (x) and the time required to complete a game (y, in hours). The following is the summary data:
   \[ n = 9, S_{XX} = 26,256, S_{YY} = 3,956, S_{XY} = 244.8, \text{MSE} = 0.2390 \]
   For a t-test for the association of time vs penalty yards with a significance level of 0.05, Which of the following are the correct hypotheses:
   - H0: \( \beta_0 = 0 \), Ha: \( \beta_0 \neq 0 \)
   - H0: \( \beta_1 = 0 \), \( \beta_1 > 0 \)
   - H0: \( \beta_0 = 0 \), \( \beta_0 > 0 \)
   - H0: \( \beta_1 = 0 \), Ha: \( \beta_1 \neq 0 \).
   The test statistic is _______ (3 decimal places).
The conclusion is:

☐ There is an association between length of the game and the number of penalty yards.

☐ There is no association between length of the game and the number of penalty yards.

Many factors affect the length of a professional football game. A study was conducted to determine the relationship between the total number of penalty yards (x) and the time required to complete a game (y, in hours). The following is the summary data:

\[ n = 9, S_{XX} = 26,256, S_{YY} = 3.956, S_{XY} = 244.8, \text{MSE} = 0.2390. \]

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of freedom</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>2.2824</td>
<td>1</td>
<td>2.2824</td>
</tr>
<tr>
<td>Error</td>
<td>1.6731</td>
<td>7</td>
<td>0.2390</td>
</tr>
<tr>
<td>Total</td>
<td>3.9556</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

For a F-test for the association of time vs penalty yards with a significance level of 0.05, which of the following are the correct hypotheses:

☐ H0: \( \beta_1 = 0 \), Ha: \( \beta_1 \neq 0 \).

☐ H0: There is an association between time and penalty yards.

☐ Ha: There is no association between time and penalty yards.

☐ H0: There is no association between time and penalty yards.

☐ Ha: There is an association between time and penalty yards.

☐ H0: \( \beta_0 = 0 \), Ha: \( \beta_0 \neq 0 \).

The value of the test statistic is _____ (2 decimal places)
4. Crimini mushrooms are more common than white mushrooms, and they contain a high amount of copper, which is an essential element according to the U.S. Food and Drug Administration. A study was conducted to determine whether the weight of a mushroom is linearly related to the amount of copper it contains. A random sample of crimini mushrooms was obtained, and the weight (in grams) and the total copper content (in mg) was measured for each. The scatterplot is show below:

![Scatterplot of mushroom weight vs copper content]

The summary statistics are: $S_{xx}=137.48, S_{yy}=5.7787, S_{xy}=21.275$

The sample correlation coefficient is _____ . (4 decimal places).

5. The temperature of the upper layer of ocean water is affected by sunlight and wind. There is often a very sharp difference in temperature between the surface zone and the more stationary deep zone. The thermocline layer marks the abrupt drop-off in temperature. The following data were obtained in a study of temperature (x, measured in °C) versus depth (y, measured in meters) above the thermocline layer in the Mediterranean Sea.

The ANOVA table from the data is:

<table>
<thead>
<tr>
<th>source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>108.54</td>
<td>1</td>
<td>108.54</td>
</tr>
<tr>
<td>Error</td>
<td>78.06</td>
<td>6</td>
<td>13.01</td>
</tr>
<tr>
<td>Total</td>
<td>186.6</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

The equation of the line is: $y^\hat{} = 23.091 - 0.084 x$

The correlation coefficient is ____ (4 decimal places).
6. The following are four QQ-plots of the residuals from different data sets. For which of these plots is the normality assumption valid?

A. 

B. 

C. 

D. 

☐ a. A
☐ b. B
☐ c. C
☐ d. D
7. The following are residual plots (residuals versus predictor variable) for different data sets. Please match them with the options available.

   ___ 1. [Image of residual plot]
   ___ 2. [Image of residual plot]
   ___ 3. [Image of residual plot]
   ___ 4. [Image of residual plot]

   a. The data violates the linear assumption.
   b. The data violates both the linear assumption and constant variance.
   c. The data violates neither the linear assumption nor the constant variance assumption.
   d. The data violates the constant variance assumption.
   e. The data violates only the linear assumption.

8. For \( x = x^* \) and a fixed confidence level, a prediction interval for an observed value \( Y \) is wider than a confidence interval for the mean value of \( Y \).
   - True
   - False
9. For a fixed confidence level, the width of a confidence interval for the mean value of \( Y \) is the same for any value of \( x^* \).
   - True
   - False

10. For \( x = x^* \), a confidence interval for the mean value of \( Y \) and a prediction interval for an observed value of \( Y \) are centered at the same value.
   - True
   - False

11. A new solar collector is being tested for use in charging batteries that can provide electricity for an entire home. A random sample of days was selected and the amount of solar radiation was measured (\( x \), in langleys) for each. The total battery charge was measured as a proportion (\( y \), between 0 and 1). The summary statistics are given.

\[
\hat{\beta}_0 = 0.2007 \quad n = 21 \quad MSE = 0.06135
\]
\[
\hat{\beta}_1 = 0.00446 \quad \bar{x} = 103.095 \quad s_{xx} = 12335.8
\]

Fill in the blanks. (Give your answer to five decimal places.)

The 95% confidence interval for the slope is (____ , ____ ).

A new solar collector is being tested for use in charging batteries that can provide electricity for an entire home. A random sample of days was selected and the amount of solar radiation was measured (\( x \), in langleys) for each. The total battery charge was measured as a proportion (\( y \), between 0 and 1). The summary statistics are given.

\[
\hat{\beta}_0 = 0.2007 \quad n = 21 \quad MSE = 0.06135
\]
\[
\hat{\beta}_1 = 0.00446 \quad \bar{x} = 103.095 \quad s_{xx} = 12335.8
\]

Fill in the blanks. (Give your answer to four decimal places.)

The 95% confidence interval for the mean value at the amount of solar radiation of 130 langleys is (____ , ____ ).
A new solar collector is being tested for use in charging batteries that can provide electricity for an entire home. A random sample of days was selected and the amount of solar radiation was measured (\(x\), in langleys) for each. The total battery charge was measured as a proportion (\(y\), between 0 and 1). The summary statistics are given.

\[
\hat{\beta}_0 = 0.2007 \quad n = 21 \quad \text{MSE} = 0.06135
\]

\[
\hat{\beta}_1 = 0.00446 \quad \bar{x} = 103.095 \quad S_{xx} = 12335.8
\]

Fill in the blanks. (Give your answer to four decimal places.)
The 95\% confidence interval for the observed value at the amount of solar radiation of 130 langleys is (______, ______)