I. Multiple choice. Answer Questions 1 and 2 below with the letters (a) through (d) from the following choices: (9 points)

(a) near zero
(b) near 5
(c) a large positive value
(d) a large negative value

1. Suppose the test statistic \( t = \frac{\bar{X} - 5}{s/\sqrt{n}} \) is used to test the hypotheses \( H_0 : \mu = 5 \) versus \( H_1 : \mu < 5 \). If the null hypothesis is true, then the test statistic would be expected to be _______?  If \( \mu = 5 \) then \( \bar{X} - 5 \) should be close to zero.

2. Consider use of the test statistic \( \frac{SSR}{s^2} \) for testing the null hypothesis \( H_0 \) : there is no linear relationship between \( X \) and \( Y \) versus \( H_1 \) : there is a linear relationship between \( X \) and \( Y \). If the alternative hypothesis is true, then the test statistic would be expected to be _______?  The more definitive the linear relationship, the larger SSR is with respect to SSE.

3. The quantity \( \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \) computes the sum of the squared vertical distances

(a) from the regression line to the sample average (SSR)
(b) from the data values to the sample average
(c) from the data values to the regression line

II. The burning rate of a solid propellant should be at least 50 cm/sec. Suppose that a sample of size \( n = 25 \) is used to test the hypotheses \( H_0 : \mu = 50 \) vs \( H_1 : \mu < 50 \) at the \( \alpha = .05 \) level.

Suppose it is known that the distribution of burning rates is normally distributed with \( \sigma = 2 \) cm/sec.

(1) What is the probability of a Type I error? (Don’t make this difficult -- it’s easy. I want a number - no work is necessary.) (3 points)

\[ \text{Type I error} = \text{significance level} = .05 \]
(2) What is the probability of making a Type II error when the actual mean burning rate is 49 cm/sec.

\[ H_0: \mu = 50 \quad \text{Reject } H_0 \text{ if } \bar{X} - 50 \leq \frac{X - 50}{2.05} \leq -1.645 \]

\[ \beta = P \left( \frac{\bar{X} - 50}{2.05} \geq 1.645 \right) \quad \text{if } \mu = 49 \]

\[ = P \left( \frac{\bar{X} - 50}{2.05} \geq -1.658 \right) \]

\[ = P \left( \bar{X} - 50 \geq 49.342 \right) \]

\[ = P \left( \bar{X} \geq 49.342 \right) \leq 0.023 \]

\[ \leq 0.077 \]

III. The K and L procedures are two methods for predicting the shear strength of steel plate girders. These two methods were applied to each of seven specific girders, and the data are shown below:

<table>
<thead>
<tr>
<th>Girder</th>
<th>K</th>
<th>L</th>
<th>( \bar{d} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.15</td>
<td>.99</td>
<td>.16</td>
</tr>
<tr>
<td>2</td>
<td>1.21</td>
<td>1.06</td>
<td>.15</td>
</tr>
<tr>
<td>3</td>
<td>1.08</td>
<td>1.07</td>
<td>.01</td>
</tr>
<tr>
<td>4</td>
<td>1.30</td>
<td>1.25</td>
<td>.05</td>
</tr>
<tr>
<td>5</td>
<td>1.20</td>
<td>1.15</td>
<td>.05</td>
</tr>
</tbody>
</table>

Paired data. For example, readings 1.15 and .99 were both made on girder 1.

\[ \frac{1.15 - .99}{5} = 0.084 \]

\[ S_d^2 = (0.76^2 - 0.66^2 - 0.64^2 - 0.54)^2 \]

\[ = 0.0448 \]

State and test the appropriate hypotheses for testing whether there is a difference in the mean strength readings of the two techniques at the \( \alpha = 0.01 \) level of significance. Also, find a P-value for this test. You may assume that the strengths are normally distributed using each of the techniques.

1. \( H_0: \mu_d = 0 \)
   \( H_1: \mu_d \neq 0 \)

2. \( \alpha = 0.01 \)

3. \( t = \frac{\bar{d}}{S_d/\sqrt{n}} \approx t(4, df) \)

4. Reject \( H_0 \) if \( |t| > t_{0.05}(4) \)
   \( 3.61 > 4.604 \)

5. \( t = \frac{0.084}{0.066/\sqrt{5}} = 2.80 \)

6. Since \( 2.80 \neq 4.604 \)
   we do not reject \( H_0 \)

7. We cannot conclude there is a difference between the two methods.
IV. The data below are the estimated miles per gallon (MPG) and weight (in tons) for 8 American-made automobiles. (31 points)

<table>
<thead>
<tr>
<th>Weight (X)</th>
<th>MPG (Y)</th>
<th>ΣX = 11.3</th>
<th>ΣY = 181</th>
<th>ΣXY = 244.8</th>
<th>ΣX² = 16.69</th>
<th>ΣY² = 4,347</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.3</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.8</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.9</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(1) Find the equation of the regression line for predicting MPG from the weight of the car.

\[ b = \frac{\sum xy - \left( \frac{\sum x \sum y}{n} \right)}{\sum x^2 - \left( \frac{\sum x^2}{n} \right)} = \frac{244.8 - (1.3)(181)/8}{16.69 - (1.3)^2/8} = \frac{-10.8625}{-1.72875} = -14.906 \]

\[ a = \bar{y} - b \bar{x} = 22.625 + 14.906(1.125) = 43.68 \]

\[ \hat{y} = 43.68 - 14.906X \]

I will give the results using the incorrect value of \( \Sigma xy = 244.8 \) at the end of this key.

(2) Find a 97% (no this isn’t a typo) confidence interval for \( \mu_{x_0} \) where \( x_0 = 1.15 \).

\[ \hat{y}_0 = 43.68 - 14.906(1.15) = 26.53 \]

\[ t_{0.05}(6) = 2.829 \]

\[ 26.53 \pm 2.829 \left( \frac{3.87}{\sqrt{8}} \right) + \left( \frac{1.15 - 1.4125}{1.72875} \right) \]

\[ \text{see below} \]

\[ 26.53 \pm 5.13 \]

i.e., 21.4 \( \rightarrow \) 31.66

(3) Find SST, SSR, and SSE.

\[ \text{SST} = \sum y^2 - \left( \frac{\sum y}{n} \right)^2/n \]

\[ = 4347 - (181/8)^2 \]

\[ = 251.875 \]

\[ \text{SSR} = b \sum xy = (14.906)(-10.8625) \]

\[ = 161.9 \]

\[ \text{SSE} = \text{SST} - \text{SSR} = 251.875 - 161.9 = 89.975 \]
V. Below is part of the SAS output from PROC REG. The dependent variable is the purity of oxygen produced in a chemical distillation process, and the dependent variable is the percentage hydrocarbons that are present in the main condenser of the distillation unit.

(21 points)

The REG Procedure
Model: MODEL1
Dependent Variable: purity

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>85.8566</td>
<td>85.8566</td>
<td>14.0261</td>
</tr>
<tr>
<td>Error</td>
<td>8</td>
<td>48.7923</td>
<td>6.1212</td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>9</td>
<td>134.8259</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(1) Fill in the blanks in the Analysis of Variance table shown above.

(2) What is the sample size?

\[ n = 10 \quad (n - 1 = 9) \]

(3) State the null and alternative hypotheses that are to be tested using the F-value you obtained in the table above.

\[ H_0: \text{there is no linear association between the two variables} \]
\[ H_1: \text{there is a linear association between the two variables} \]

(4) What is the critical region for the test referred to in (c) for an \( \alpha = .05 \) level test?

\[ F > F_{0.05}(1, 8) = 5.32 \]

(5) What is the P-value for this test? (Discuss how you found your answer.)

\[ P < .01 \quad \text{since} \quad 14.0261 > 11.26 = F_{0.01}(1, 8) \]

(6) Based on these data, what is the estimate of \( \text{Var}(\varepsilon) \) in the model \( Y = \alpha + \beta x + \varepsilon \) where \( Y \) indicates purity and \( x \) indicates hydrocarbon level?

\[ s^2 = 6.1212 \]

(7) Note that the SAS output for testing the significance of the slope and y-intercept is not given here. By using the information in the above table, what is the absolute value of the t-value used to test \( H_0 : \beta = 0 \)?

\[ t = \sqrt{F} \quad \text{from the analysis of variance table} \]
\[ \Rightarrow \quad t = \sqrt{14.0261} = 3.745 \]
\( IV. \) Using \( \Sigma XY = 224.8 \)

1. \( b = \frac{\Sigma X_iY_i - \Sigma X_i \Sigma Y_i / n}{\Sigma X_i^2 - (\Sigma X_i)^2 / n} \)
   
   \[ b = \frac{224.8 - (11.3)(181)/8}{16.69 - (11.3)^2/8} \]
   
   \[ b = -30.8625 \]
   
   \[ b = -42.35 \]

2. \( a = \overline{Y} - b \overline{X} = 22.625 + 42.35(1.4125) \)
   
   \[ a = 82.44 \]

\[ \hat{Y} = 82.44 - 42.35X \]

2. \( \hat{Y}_0 = 82.44 - 42.35(1.15) = 33.74 \)

\[ 33.74 \pm 2.829 \sqrt{\frac{1}{8} + \frac{(1.15-1.4125)^2}{1728.75}} \]

\[ 33.74 \pm 17.58 \]

\( n \rightarrow 16.16 \rightarrow 51.32 \)

3. \( \text{SST} = \Sigma Y_i^2 - (\Sigma Y_i)^2 / n = 4347 - (181)^2 / 8 \)

\[ = 251.875 \]

\( \text{SSR} = bS_{XY} = (42.35)(-30.8625) \)

\[ = 1307.1 \]

\( \text{SSE} = \text{SST} - \text{SSR} = 251.875 - 1307 \)

\[ = -1055 \]

\( \text{SSE can't be negative and this is what led us to realize there was an error. I told you to assume SSE is positive, i.e., use SSE = 1055 and proceed.} \)
\[ S^2 = \frac{SSE}{6} = \frac{1055}{6} = 175.83 \]

so \( s = \sqrt{175.83} = 13.26 \) which was used in the confidence interval in (2).