1. A plant distills liquid air to produce oxygen. The following data were used by this plant to assess the relationship between the **percentage of impurity in the oxygen (PERC)** with the **amount (ppm) of impurity in the liquid air (AMT)**. Use the attached SAS output at the end of the exam and the scatterplot below to answer the following questions.

![Scatterplot with line](image)

(a) **(3 pts)** What is the response variable?

PERC

(b) **(6 pts)** Write down the least squares regression line and describe the relationship between PERC and AMT.

\[
\text{PERC} = 96.454 - 2.9 \text{ AMT}
\]

**Negative linear relationship with unit increase in AMT resulting in decrease of 2.9%**
(c) (5 pts) What is the estimated error variance? What does this variance represent in terms of the estimated regression model?

\[ \text{MSE} = .18297 \]

VARABILITY OF OBSERVATIONS ABOUT REG LINE

(d) (4 pts) Explain the difference between the residuals (RESIDUAL) and studentized residuals (STUDENT RESIDUAL) columns in the output.

\[ \text{STUDENT} = \frac{\text{RESIDUAL}}{s(\text{RESIDUAL})} \]

(e) (3 pts) If the liquid air had a pollution count of 1.00, what is the predicted percentage of impurity in the oxygen?

\[ 96.454 - 2.9(1.00) = 93.554 \]

(f) (4 pts) If the observed percentage of impurity at 1.00 were 93.0, what is the residual value?

\[ 93.0 - 93.554 = -.554 \]

(g) (3 pts) Suppose you plan to construct a 95% prediction interval for impurity amounts of 1.30 and 1.50. Given the sample mean of AMT is 1.354, which amount would have a narrower interval and why?

1.3 since it is closer to 1.354
(h) (5 pts) What is the P-value and degrees of freedom for the test $H_0 : \beta_1 = 0$? What is the conclusion?

\[ P\text{-value} < .0001 \quad df = 13 \]

\[ \text{REJECT } H_0 \text{ : DATA SUPPORT } \beta_1 \neq 0 \]

(i) (6 pts) Construct a 95% confidence interval for $\beta_1$.

\[ -2.9 \pm t(.975,13) \cdot 30557 \]

\[ t(.975,13) = 2.16 \]

(j) (4 pts) What is the estimated Pearson correlation between PERC and AMT?

\[-\sqrt{.8739}\]

(k) (4 pts) The plant’s execs are concerned that the estimated intercept is not 100%. Explain why this should not be that much of a concern given this study.

\[ \begin{align*}
\hat{\alpha} & = 96.45 \\
\text{AT } AMT = 0 \text{, PERC} & \geq 96.45 \\
\text{However, no data near } 0 \text{, AMT} = 0 \text{ is outside scope.}
\end{align*} \]
2. Short answer questions. Each part is unrelated.

(a) (3 pts) Suppose the estimated regression equation is \( \hat{Y} = 2 + 3X \). Give the estimated regression equation if the variable \( U = 3X + 6 \) were used in place of \( X \).

\[
X = \frac{U - 6}{3}
\]

\[
\hat{Y} = 2 + 3 \cdot \frac{U - 6}{3} = U - 4
\]

(b) (4 pts) Explain how a 95% confidence interval for the slope can be used to test \( H_0 : \beta_1 = 5 \) at the \( \alpha = .05 \) level?

LOOK TO SEE IF 5 IS IN INTERVAL
IF OUTSIDE REJECT \( H_0 \)

(c) (5 pts) Rob Poorman Auto Sales has decided to use \( R^2 \) to select the best model in predicting car demand based on several demographic variables. Explain when (and when not) this is a reasonable approach.

\( R^2 \) USED WHEN \( p \) IS FIXED

\( C_p, \) ADJUSTED \( R^2 \) USED WHEN \( p \) ARBITRARY
(d) (8 pts) For each of the following residual plots ($e_i$ vs $\hat{Y}$), state which assumptions of the linear regression model (if any) appear violated.

- **Nonlinear**
  - Residuals show a clear upward trend, indicating a nonlinear relationship.

- **Outlier**
  - Residuals contain a point that is far from the others, indicating an outlier.

- **Good**
  - Residuals are randomly scattered around the horizontal line, indicating a linear relationship.

- **Non-constant Variance**
  - Residuals show a funnel-like shape, indicating non-constant variance (heteroscedasticity).
3. An experiment was performed to best predict the burn time of a tobacco leaf using the leaf percentages of nitrogen ($X_1$), chlorine ($X_2$), and potassium ($X_3$). Use the attached SAS output of the top 10 “best” models to answer the following questions.

(a) (4 pts) The log(burn time) was chosen as the response variable. Describe a possible reason for using this transformation of burn time?

**NONLINEAR RELATIONSHIP BETWEEN $Y$ & $X'S$**

**PERHAPS TO STABILIZE VARIANCE**

(b) (4 pts) The full model contained all the **quadratic** terms and first-order interactions (e.g., $X_1^2$ and $X_1X_2$). Prior to analysis, the original variables were standardized. Why was this done?

**MULTICOLLINEARITY OF PRESENT WHEN LOOKING AT POLYNOMIAL MODELS**

(c) (4 pts) Write down the fitted regression line for the best model based on adjusted $R^2$.

$$Y = 0.46463 - 0.3009X_1 - 0.28385X_2 + 0.22094X_3 + 0.10558X_1^2$$

(d) (6 pts) Explain what this regression line tells you about the relationship between the percentages of the three elements and the burning time.

$\uparrow X_3 \rightarrow \uparrow Y \text{ by } 0.22094 \text{ units}$

$\uparrow X_1 \text{ 1 unit} \rightarrow \text{change } Y \text{ by } (-0.3009 + 0.10558X_2) \text{ units}$

**SIMILAR FOR $X_2$**

**CHANGE NOT CONSTANT BECAUSE OF INTERACTION**
(e) (4 pts) Are there any models in this list that are unacceptable based on $C_p$? Explain.

```
UNACCEPTABLE IF \[ C_p > p \]
```

YES 1 MODEL

\[
\frac{4.9242}{4} > 4
\]

(f) (6 pts) Describe a formal test (with degrees of freedom) that could be used to test whether this “best” model could be used in place of the full model with nine parameters.

Full model

- Model: 9
- Error: $n-10$
- Total: $n-1$

Reduced model

- Model: 4
- Error: $n-5$
- Total: $n-1$

\[
\frac{(SSE(R) - SSE(F))/5}{SSE(F)/(n-10)} = F^*
\]

Compare to $F(5, n-10)$
### Results for Problem #1

#### 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>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>16.49078</td>
<td>16.49078</td>
<td>90.13</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>13</td>
<td>2.37855</td>
<td>0.18297</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>14</td>
<td>18.86933</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 0.42774
Dependent Mean 92.52667

#### Parameter Estimates

| Parameter | DF | Estimate | Standard Error | t Value | Pr > |t| |
|-----------|----|----------|----------------|---------|-------|
| Intercept | 1  | 96.45456 | 0.42822        | 225.24  | <.0001|
| AMT       | 1  | -2.90096 | 0.30557        | -9.49   | <.0001|

#### Output Statistics

<table>
<thead>
<tr>
<th>Obs</th>
<th>AMT</th>
<th>PERC</th>
<th>Value</th>
<th>Mean</th>
<th>Predict</th>
<th>Std Error</th>
<th>Residual</th>
<th>Std Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.1</td>
<td>93.3000</td>
<td>93.2635</td>
<td>0.1350</td>
<td>0.0365</td>
<td>0.406</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.45</td>
<td>92.0000</td>
<td>92.2482</td>
<td>0.1143</td>
<td>-0.2482</td>
<td>0.412</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.36</td>
<td>92.4000</td>
<td>92.5093</td>
<td>0.1105</td>
<td>-0.1093</td>
<td>0.413</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.59</td>
<td>91.7000</td>
<td>91.8420</td>
<td>0.1319</td>
<td>-0.1420</td>
<td>0.407</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.08</td>
<td>94.0000</td>
<td>93.3215</td>
<td>0.1386</td>
<td>0.6785</td>
<td>0.406</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.75</td>
<td>94.6000</td>
<td>94.2788</td>
<td>0.2151</td>
<td>0.3212</td>
<td>0.370</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.2</td>
<td>93.6000</td>
<td>92.9734</td>
<td>0.1201</td>
<td>0.6266</td>
<td>0.411</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.99</td>
<td>93.1000</td>
<td>93.5826</td>
<td>0.1567</td>
<td>-0.4826</td>
<td>0.398</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.83</td>
<td>93.2000</td>
<td>94.0468</td>
<td>0.1945</td>
<td>-0.8468</td>
<td>0.381</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1.22</td>
<td>92.9000</td>
<td>92.9154</td>
<td>0.1178</td>
<td>-0.0154</td>
<td>0.411</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1.47</td>
<td>92.2000</td>
<td>92.1902</td>
<td>0.1160</td>
<td>0.009844</td>
<td>0.412</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1.81</td>
<td>91.3000</td>
<td>91.2038</td>
<td>0.1778</td>
<td>0.0962</td>
<td>0.389</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>2.03</td>
<td>90.1000</td>
<td>90.5656</td>
<td>0.2342</td>
<td>-0.4656</td>
<td>0.358</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1.75</td>
<td>91.6000</td>
<td>91.3779</td>
<td>0.1638</td>
<td>0.2221</td>
<td>0.395</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>1.68</td>
<td>91.9000</td>
<td>91.5810</td>
<td>0.1487</td>
<td>0.3190</td>
<td>0.401</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Studentized Residuals and Cook's Distance

<table>
<thead>
<tr>
<th>Obs</th>
<th>AMT</th>
<th>Residual</th>
<th>-2</th>
<th>1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.1</td>
<td>0.0899</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>1.45</td>
<td>-0.602</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.014</td>
</tr>
<tr>
<td>3</td>
<td>1.36</td>
<td>-0.264</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.002</td>
</tr>
<tr>
<td>4</td>
<td>1.59</td>
<td>-0.349</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.006</td>
</tr>
<tr>
<td>5</td>
<td>1.08</td>
<td>1.677</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.165</td>
</tr>
<tr>
<td>6</td>
<td>0.75</td>
<td>0.869</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.128</td>
</tr>
<tr>
<td>7</td>
<td>1.2</td>
<td>1.526</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.100</td>
</tr>
<tr>
<td>8</td>
<td>0.99</td>
<td>-1.213</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.114</td>
</tr>
<tr>
<td>9</td>
<td>0.83</td>
<td>-2.223</td>
<td>****</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.644</td>
</tr>
<tr>
<td>10</td>
<td>1.22</td>
<td>-0.0374</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.000</td>
</tr>
<tr>
<td>11</td>
<td>1.47</td>
<td>0.0239</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.000</td>
</tr>
<tr>
<td>12</td>
<td>1.81</td>
<td>0.247</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.006</td>
</tr>
<tr>
<td>13</td>
<td>2.03</td>
<td>-1.301</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.362</td>
</tr>
<tr>
<td>14</td>
<td>1.75</td>
<td>0.562</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.027</td>
</tr>
<tr>
<td>15</td>
<td>1.68</td>
<td>0.796</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.044</td>
</tr>
</tbody>
</table>
### RESULTS FOR PROBLEM #3

<table>
<thead>
<tr>
<th>Model</th>
<th>C(p)</th>
<th>R-Square</th>
<th>R-Square</th>
<th>Adjusted R-Square</th>
<th>AIC</th>
<th>SBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1.6559</td>
<td>0.8575</td>
<td>0.8348</td>
<td>-93.4815</td>
<td>-86.47553</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.2586</td>
<td>0.8602</td>
<td>0.8310</td>
<td>-92.0370</td>
<td>-83.62982</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.2996</td>
<td>0.8599</td>
<td>0.8307</td>
<td>-91.9792</td>
<td>-83.57202</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.3625</td>
<td>0.8595</td>
<td>0.8302</td>
<td>-91.8907</td>
<td>-83.48356</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.6017</td>
<td>0.8579</td>
<td>0.8283</td>
<td>-91.5567</td>
<td>-83.14963</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.6392</td>
<td>0.8577</td>
<td>0.8280</td>
<td>-91.5047</td>
<td>-83.09753</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4.0757</td>
<td>0.8679</td>
<td>0.8335</td>
<td>-91.7545</td>
<td>-81.94613</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4.9242</td>
<td>0.8229</td>
<td>0.8025</td>
<td>-88.9491</td>
<td>-83.34427</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5.0568</td>
<td>0.8615</td>
<td>0.8254</td>
<td>-90.3231</td>
<td>-80.51470</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5.0991</td>
<td>0.8612</td>
<td>0.8250</td>
<td>-90.2629</td>
<td>-80.45449</td>
<td></td>
</tr>
</tbody>
</table>

### Parameter Estimates

<table>
<thead>
<tr>
<th>Model</th>
<th>C(p)</th>
<th>Intercept</th>
<th>x1</th>
<th>x2</th>
<th>x3</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1.6559</td>
<td>0.66463</td>
<td>-0.3009</td>
<td>-0.28385</td>
<td>0.22094</td>
</tr>
<tr>
<td>5</td>
<td>3.2586</td>
<td>0.67306</td>
<td>-0.29773</td>
<td>-0.28079</td>
<td>0.22434</td>
</tr>
<tr>
<td>5</td>
<td>3.2996</td>
<td>0.68651</td>
<td>-0.29838</td>
<td>-0.28029</td>
<td>0.22090</td>
</tr>
<tr>
<td>5</td>
<td>3.3625</td>
<td>0.64495</td>
<td>-0.29958</td>
<td>-0.30527</td>
<td>0.22461</td>
</tr>
<tr>
<td>5</td>
<td>3.6017</td>
<td>0.65608</td>
<td>-0.30359</td>
<td>-0.28188</td>
<td>0.21940</td>
</tr>
<tr>
<td>5</td>
<td>3.6392</td>
<td>0.66464</td>
<td>-0.30140</td>
<td>-0.28349</td>
<td>0.22062</td>
</tr>
<tr>
<td>6</td>
<td>4.0757</td>
<td>0.63515</td>
<td>-0.29371</td>
<td>-0.32928</td>
<td>0.23731</td>
</tr>
<tr>
<td>3</td>
<td>4.9242</td>
<td>0.68600</td>
<td>-0.30966</td>
<td>-0.26606</td>
<td>0.22320</td>
</tr>
<tr>
<td>6</td>
<td>5.0568</td>
<td>0.66701</td>
<td>-0.29804</td>
<td>-0.30010</td>
<td>0.22425</td>
</tr>
<tr>
<td>6</td>
<td>5.0991</td>
<td>0.67530</td>
<td>-0.30148</td>
<td>-0.27879</td>
<td>0.22384</td>
</tr>
</tbody>
</table>

### Parameter Estimates

<table>
<thead>
<tr>
<th>Model</th>
<th>C(p)</th>
<th>x11</th>
<th>x22</th>
<th>x33</th>
<th>x12</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1.6559</td>
<td></td>
<td></td>
<td></td>
<td>0.10558</td>
</tr>
<tr>
<td>5</td>
<td>3.2586</td>
<td></td>
<td></td>
<td></td>
<td>0.12202</td>
</tr>
<tr>
<td>5</td>
<td>3.2996</td>
<td></td>
<td></td>
<td>-0.02412</td>
<td>0.11266</td>
</tr>
<tr>
<td>5</td>
<td>3.3625</td>
<td></td>
<td>0.02137</td>
<td></td>
<td>0.10073</td>
</tr>
<tr>
<td>5</td>
<td>3.6017</td>
<td>0.00844</td>
<td></td>
<td></td>
<td>0.10455</td>
</tr>
<tr>
<td>5</td>
<td>3.6392</td>
<td></td>
<td></td>
<td></td>
<td>0.10200</td>
</tr>
<tr>
<td>6</td>
<td>4.0757</td>
<td></td>
<td>0.05199</td>
<td></td>
<td>0.12965</td>
</tr>
<tr>
<td>3</td>
<td>4.9242</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5.0568</td>
<td></td>
<td>0.01951</td>
<td>-0.02242</td>
<td>0.10774</td>
</tr>
<tr>
<td>6</td>
<td>5.0991</td>
<td></td>
<td></td>
<td></td>
<td>0.11437</td>
</tr>
</tbody>
</table>

### Parameter Estimates

<table>
<thead>
<tr>
<th>Model</th>
<th>C(p)</th>
<th>x13</th>
<th>x23</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1.6559</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.2586</td>
<td></td>
<td>-0.02987</td>
</tr>
<tr>
<td>5</td>
<td>3.2996</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.3625</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.6017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.6392</td>
<td>0.00797</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4.0757</td>
<td></td>
<td>-0.06514</td>
</tr>
<tr>
<td>3</td>
<td>4.9242</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5.0568</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5.0991</td>
<td>0.02660</td>
<td>-0.03766</td>
</tr>
</tbody>
</table>