Q2.1 [15 pts] I have data on yield using three fertilizers A, B, and C and want to run a multiple regression that will regress a column Y on a matrix X. My data are shown below. I am especially interested in comparing fertilizer A to the average of B and C. Find a second comparison so that associated X matrix will be full rank and I automatically get a t-test for the comparison I am interested in. Fill in a full rank X matrix (including an intercept column).

DATA:

<table>
<thead>
<tr>
<th>Y</th>
<th>FERTILIZER</th>
<th>X MATRIX</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>A</td>
<td>1 2 0</td>
</tr>
<tr>
<td>18</td>
<td>A</td>
<td>1 2 0</td>
</tr>
<tr>
<td>22</td>
<td>A</td>
<td>1 2 0</td>
</tr>
<tr>
<td>14</td>
<td>A</td>
<td>1 2 0</td>
</tr>
<tr>
<td>22</td>
<td>B</td>
<td>1 -1 1</td>
</tr>
<tr>
<td>28</td>
<td>B</td>
<td>1 -1 1</td>
</tr>
<tr>
<td>32</td>
<td>B</td>
<td>1 -1 1</td>
</tr>
<tr>
<td>26</td>
<td>B</td>
<td>1 -1 1</td>
</tr>
<tr>
<td>25</td>
<td>C</td>
<td>1 -1 -1</td>
</tr>
<tr>
<td>34</td>
<td>C</td>
<td>1 -1 -1</td>
</tr>
<tr>
<td>30</td>
<td>C</td>
<td>1 -1 -1</td>
</tr>
<tr>
<td>29</td>
<td>C</td>
<td>1 -1 -1</td>
</tr>
</tbody>
</table>

1. Fert A vs Average(Fert B, Fert C)

2. Fert B vs Fert C

<table>
<thead>
<tr>
<th>Contrast</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>A vs (B, C)</td>
<td>2</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>B vs C</td>
<td>0</td>
<td>1</td>
<td>-1</td>
</tr>
</tbody>
</table>
Q2.2 [40pts] I measure yield \( Y \) as a function of \( P \) and \( W \), where \( P \) is the deviation of soil pH from 5 and \( W \) is the deviation of the amount of water applied to the soil from 8 liters. A soil pH of 5 and 8 liters of water might represent standard growing conditions.

Here are the data long:

\[
\begin{array}{c|cccc}
Y & I & W & P \\
10 & 1 & -4 & 0 \\
18 & 1 & -2 & 2 \\
23 & 1 & -1 & 0 \\
13 & 1 & 0 & -4 \\
25 & 1 & 4 & 1 \\
32 & 1 & 3 & 0 \\
26 & 1 & 0 & 1 \\
\end{array}
\]

Let \( X \) represent the matrix with columns \( I \), \( W \) and \( P \) then.

\[
X = \begin{bmatrix}
1 & -4 & 0 \\
1 & -2 & 2 \\
1 & -1 & 0 \\
1 & 0 & -4 \\
1 & 4 & 1 \\
1 & 3 & 0 \\
1 & 0 & 1 \\
\end{bmatrix}
\]

\[
X'X = \begin{bmatrix}
7 & 0 & 0 \\
0 & 46 & 0 \\
0 & 0 & 22 \\
\end{bmatrix}
\]

a. Matrix \( X'X \) has dimensions 3 rows and 3 columns

b. Compute \( X'X \) and \((X'X)^{-1}\)

\[
X'X = \begin{bmatrix}
7 & 0 & 0 \\
0 & 46 & 0 \\
0 & 0 & 22 \\
\end{bmatrix} \quad (X'X)^{-1} = \begin{bmatrix}
\frac{1}{7} & 0 & 0 \\
0 & \frac{1}{46} & 0 \\
0 & 0 & \frac{1}{22} \\
\end{bmatrix}
\]

c. Vector \( X'Y \) is given

\[
X'Y = \begin{bmatrix}
147 \\
72 \\
35 \\
\end{bmatrix}
\]

And from the PROC REG output we have Error Mean Squares is 24.94368 with Error df = 7-2-1 = 4 degrees of freedom.
d. Estimate the parameters in the regression of $Y$ on $W$ and $P$ (with an intercept) and give the standard error for regression coefficient of $W$. 

\[
\begin{bmatrix}
1/7 & 0 & 0 \\
0 & 1/46 & 0 \\
0 & 0 & 1/22
\end{bmatrix}
\begin{bmatrix}
147 \\
72 \\
35
\end{bmatrix}
= 
\begin{bmatrix}
21 \\
1.5652 \\
1.5909
\end{bmatrix}
\]

\[
\text{Var}(b_0) = (1/7) \times \text{MSE} = 0.14286 \times 24.94368 = 3.5634
\]

\[
\begin{align*}
\sigma_{b_0} &= \sqrt{3.5634} = 1.8877 \\
\text{Var}(b_1) &= (1/46) \times \text{MSE} = 0.02174 \times 24.94368 = 0.5422539 \\
\sigma_{b_1} &= \sqrt{0.5422539} = 0.7364 \\
\text{Var}(b_2) &= (1/46) \times \text{MSE} = 0.04545 \times 24.94368 = 1.1338 \\
\sigma_{b_2} &= \sqrt{1.1338} = 1.0648
\end{align*}
\]

e. Write the prediction equation,

\[
\hat{Y} = 21 + 1.5652W + 1.5909P
\]

std errors : \( (\______) \ (0.7364) \ (\______) \)

f. Compute the sample-to-sample correlation between the $W$ coefficient and the $P$ coefficient.

Correlation between $W$ and $P$ is 0.
Q2.3 [40pts] A multiple regression is used to fit the model

\[ Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + e_i \]

to some observational data. The error mean square is \(\text{MSE}=50\) with 26 degrees of freedom. We also have these corrected sums of squares from PROC REG:

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Type I [sequential] Sum of Squares</th>
<th>Type II [partial] Sum of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>1</td>
<td>380</td>
<td>100</td>
</tr>
<tr>
<td>X2</td>
<td>1</td>
<td>20</td>
<td>130</td>
</tr>
<tr>
<td>X3</td>
<td>1</td>
<td>100</td>
<td>NP</td>
</tr>
<tr>
<td>X4</td>
<td>1</td>
<td>80</td>
<td>80</td>
</tr>
</tbody>
</table>

\( t_{26,0.975} = 2.06 \)

[A] [10 points] Fill in the missing partial sums of squares where possible. If not possible, write NP.

[B] [25 points] With reference to the 4 variable model above, compute the following if possible. If not possible, put NP.

[i] The model F test [corrected] \( F = 2.9 \)

Model SS = 380 + 20 + 100 + 80 = 580
Model df = 4
Model MS = 580/4 = 145
Model F = Model MS / Error MS = 145 / 50 = 2.9

(ii) \( R^2 = \frac{580}{1880} = 0.3085, 30.85\% \)

\( R^2 \) squared = Model SS / total SS
Error SS = Error MS * Error df = 50 * 26 = 1300
Total SS = Model SS + Error SS = 580 + 1300 = 1880

[iii] The F test for testing \( H_0: \beta_2 = \beta_3 = \beta_4 = 0 \) \( F = 1.33 \)

RSS(\(X_2, X_3, X_4/int, X_1\)) = 20 + 100 + 80 = 200
Hypothesis df = 3
Hypothesis MS = 200/3 = 66.67
Hypothesis F = Hypothesis MS/ Error MS = 66.67/50 = 1.33

(iv) The F test for testing \( H_0: \beta_1 = 0 \) \( F = 2.00 \)
Hypothesis MS = 100/1 = 100
Hypothesis F = Hypothesis MS/ Error MS = 100/50 = 2.00

[v] The F test for testing \( H_0: \beta_2 = \beta_3 = 0 \) \( F = \text{NP} \)
[C] [5 points] Suppose my estimate of $\beta_2$ is 7. Compute a 95% confidence interval for $\beta_2$.

1. Calculate $F$ to test $H_0: \beta_2 = 0$
   $H_1: \beta_2 \neq 0$

2. $F_{\beta_2} = \frac{130/1}{50} = 2.6$

3. Take square root of calculated $F$ value, result is the value for the $t$-statistic = estimate for $b_2/se_b^2$
   $F_{\beta_2} = 2.6 = t^2$

4. $t = \sqrt{F_{\beta_2}} = \sqrt{2.6} = 1.612452$

5. Get $se_b^2$
   $t = 1.612452 = \frac{b_2}{se_b} = \frac{7}{s_{b_2}}$
   $s_{b_2} = \frac{7}{1.612452} = 4.34$

6. Calculate the confidence interval

$$\hat{\beta}_2 \pm t_{26,0.975} \times s_{b_2}$$

95% CI : $= 7 \pm 2.06 \times 4.34$

$conf(-1.94 \leq \beta_2 \leq 15.94) = 95\%$