The following exercises were taken from Dr. D. Dickey WEB site for ST512

http://www.stat.ncsu.edu/people/dickey/courses/st512/

1) I ran a regression model with interaction obtaining

\[ Y = 100 + 2C - 3T + 0.2C*T \]

where C is the amount of some chemical, T is time in days, and Y is yield. My error mean square was 12 and my error sum of squares was 240.

a) How many data points did I have? \( n = \) 

b) For what amounts of chemical C, if any, will the predicted Y increase over time?

c) If I remove the interaction term, the error sum of squares will become 348. Find, if possible, the t test for the interaction coefficient.

\( t = \) 

d) How many degrees of freedom ______ does this interaction t test have?

2) I ran a regression of yield Y on PH (soil pH), M (soil moisture), and N (available soil nitrogen) getting these sequential (type I) and partial (type II) sums of squares from the commands PROC REG: MODEL Y = PH M N ;

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>TYPE I (sequential)</th>
<th>TYPE II (partial)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PH</td>
<td>500</td>
<td>140</td>
</tr>
<tr>
<td>M</td>
<td>80</td>
<td>60</td>
</tr>
<tr>
<td>N</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

Find the following if possible (if any is not possible, put “NP”)

a) The (corrected) model sum of squares _____ for the regression described above.

b) The error sum of squares _____ from the regression described above.

c) The (corrected) model sum of squares _____ in the regression of Y on PH and M

d) The (corrected) model sum of squares _____ in the regression of Y on PH and N

3) A multiple regression of Y on X1 and X2 includes an intercept. The resulting equation from the regression is

\[ \hat{Y} = 10 + 3X1 - 6X2 \]

The X'X matrix and its inverse are

\[ X'X = \begin{bmatrix} 30 & 5 & 5 \\ 5 & 10/3 & 10/3 \end{bmatrix}, \quad (X'X)^{-1} = \begin{bmatrix} 4/90 & -1/15 & 0 \\ -1/15 & C & -1 \end{bmatrix} \]

The error mean square is 25 and the (corrected) regression sum of squares is 382.5. Answer the following if possible from the given information

a) (*) Find the missing entry (C). \( C = \) 

b) Give the number of observations on Y. \( n = \)
c) Give the error degrees of freedom __________.

d) Give \( \sum(Y_i - \bar{Y})^2 \) = __________.

e) Give \( X'Y = \) __________.

f) Estimate the mean \( Y \) for \( X_1 = 3 \) and \( X_2 = 2 \). \( Y = \) __________.

g) We obtained \( b_2 = -6 \). Estimate the variance of \( b_2 \) __________.

h) We also obtained \( b_1 = 3 \) and so \( b_1 - b_2 = 9 \). Estimate the variance of \( b_1 - b_2 = \) __________.

i) Give the rank of \( X \) __________.

i) Give \( \bar{Y} = \) __________.

4) I fit a regression model to predict root weight \( R \) as a function of stem height \( S \) and leaf area \( L \). Here is my regression equation with standard errors in parentheses below the coefficients:

\[
\hat{R} = 10 + 3S + 2L \\
(3) \quad (1.2) \quad (1.2)
\]

I had 20 data points and my error mean square was 2. The corrected total sum of squares is 65. All questions refer to this regression.

a) Draw a picture (graph) to represent \( R = 10 + 3S + 2L \).

b) In your picture, what corresponds to the 10 in the equation above?

c) Explain what the coefficient 3 in the above equation tells you about your graph.

d) Another worker is surprised by the coefficient 3 in this study. He tells you that the coefficient should be 4 according to theory. Give a t statistic to test his hypothesis.

e) How many degrees of freedom does the above t statistic have?

f) In SAS we would get an overall F test for our model. Letting \( B_0 \), \( B_1 \) and \( B_2 \) be the theoretical coefficients being estimated in our model, write down the hypothesis being tested.

\[ H_0 : \]

and the calculated F statistic =

g) Suppose the F test from problem 6 has p value .035. I know from this that I should reject my null hypothesis. Explain what this p value represents (using a drawing of the F distribution if you like).
h) To see if I can omit S from the model, I can use either a t or F test. Find the calculated F test statistic.

i) Compute, if possible, the increase in error sum of squares that would result if S were omitted from the model.

5) A regression of Y on a column of 1's, X1, and X2 is run to estimate the parameters of the model

\[ Y = B_0 + B_1 X_1 + B_2 X_2 + e. \]

The estimated regression equation based on n=20 observations is

\[ \hat{Y} = 30 + 12 X_1 - 5 X_2, \]

that is, \( b_0=30, b_1=12, \) and \( b_2=-5. \) Suppose we also know the \( \text{inv}(X'X) \) matrix is

\[
\begin{pmatrix}
0.13333 & -0.13333 & 0.10000 \\
-0.13333 & 0.23333 & -0.20000 \\
0.10000 & -0.20000 & 0.20000
\end{pmatrix} = \frac{1}{30} \begin{pmatrix}
4 & -4 & 3 \\
-4 & 7 & -6 \\
3 & -6 & 6
\end{pmatrix}
\]

and the error sum of squares is 90.

a) Estimate sigma squared, the variance of e:

b) Estimate the covariance between \( b_1 \) and \( b_2: \)

c) I am thinking about what would happen if I ran this experiment over again. If my new sample has \( b_1 > B_1 \), should I expect \( b_2 > B_2 \) or \( b_2 < B_2 ? \) Explain your answer.

d) I want to compute a 95% prediction interval for an individual \( Y \) at \( X_1=3, X_2=3. \) Fill in the two missing entries below with numbers.

\[ \text{____________________} (+/-) t*\sqrt{\text{____________________}}. \]

6) I am interested in the relationship between \( \text{YIELD} \) in potted plants as a function of soil \( \text{Ph} \), temperature (\( \text{TEMP} \)) and moisture (\( \text{WATER} \)). I observe yield and soil conditions for several plants and run this regression:

(a) PROC REG; MODEL YIELD = PH TEMP WATER / SS1;

The results are at the end of the quiz and all questions refer to this printout. As usual, I assume a model

\[ \text{Yield} = \text{Beta0} + \text{Beta1*PH} + \text{Beta2*TEMP} + \text{Beta3*WATER} + e \]

and I use X and Y to denote the usual regression matrices.

a) How many rows ____ and columns ____ does the X matrix have?
b) Give an estimate of Beta1 ______ and calculate the t statistic for testing $H_0: \beta_1 = 18$ ($t =$______). How many degrees of freedom does $t$ have? ____

c) Calculate, if possible, an $F$ for testing the null hypothesis that the coefficients of TEMP and WATER can both simultaneously be set to 0.

$$F = _____$$

d) The P-value 0.3380 for WATER is somehow related to the statistic $t = 0.988$. Explain this relationship, using a picture if you like.

-1

e) Calculate (or just find) $b'X'Y$ where $b = (X'X)^{-1} (X'Y)$.

$$b'X'Y = _______$$

---

Model: MODEL1
Dependent Variable: YIELD

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>Prob&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3</td>
<td>2006.05659</td>
<td>668.68553</td>
<td>8.743</td>
<td>0.0012</td>
</tr>
<tr>
<td>Error</td>
<td>16</td>
<td>1223.74341</td>
<td>76.48396</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C Total</td>
<td>19</td>
<td>3229.80000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 8.74551  R-square 0.6211  Dep Mean 114.10000  Adj R-sq 0.5501  C.V. 7.66478

Parameter Estimates

| Variable     | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob > |T| |
|--------------|----|--------------------|----------------|-----------------------|--------|---|
| INTERCEPT    | 1  | -18.656243         | 26.26313587    | -0.710                | 0.4988 | 0.3380 |
| PH           | 1  | 25.866095          | 5.56684983     | 4.646                 | 0.0003 |
| TEMP         | 1  | -0.549719          | 0.19337876     | -2.843                | 0.0118 |
| WATER        | 1  | 0.491309           | 0.49743915     | 0.988                 | 0.3380 |

Variable DF Type I SS

| INTERCEPT | 1 | 260376 |
| PH | 1 | 1289.049922 |
| TEMP | 1 | 642.396128 |
| WATER | 1 | 74.610536 |

7) I regressed Y on X1 and X2 getting these Type I and Type II sums of squares from PROC REG. What would have happened if I had run the regression in the opposite order (PROC REG; MODEL Y = X2 X1; )
8) Here is the result of running PROC REG on some data using the /I SS1 SS2 options to deliver sums of squares and the inverse of X'X (I have deleted the extraneous data that usually borders this inverse). Note that the inverse is in two parts.

a) Fill in the missing numbers below:

Model: MODEL1

X'X Inverse

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Type I SS</th>
<th>Type II SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEP</td>
<td>1</td>
<td>90000</td>
<td>18000</td>
</tr>
<tr>
<td>X1</td>
<td>1</td>
<td>800</td>
<td>900</td>
</tr>
<tr>
<td>X2</td>
<td>1</td>
<td>700</td>
<td>700</td>
</tr>
</tbody>
</table>

Answer:

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Type I SS</th>
<th>Type II SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEP</td>
<td>1</td>
<td>90000</td>
<td>18000</td>
</tr>
<tr>
<td>X2</td>
<td>1</td>
<td>(______)</td>
<td>(______)</td>
</tr>
<tr>
<td>X1</td>
<td>1</td>
<td>(______)</td>
<td>(______)</td>
</tr>
</tbody>
</table>

Dependent Variable: Y

Analysis of Variance
### Source DF Squares     Square    F Value    Prob>F
| Model   | 4       | 34603.17258 | 8650.79315 | (_______) | 0.0001 |
| Error   | (___)   | (___________) | 10.47076   |           |        |
| C Total | (___)   | 34844.00000 |

- Root MSE: 3.23585
- R-square: (_______)
- Dep Mean: 186.00000

### Parameter Estimates

| Variable | DF | Parameter | Standard Error | T for H0: Parameter=0 | Prob > |T| |
|----------|----|-----------|----------------|------------------------|--------|---|
| INTERCEP | 1  | 109.396250 | 4.98172872     | 21.959                 | 0.0001 |
| X1       | 1  | 6.434904   | (___________)  | (________)             | 0.0001 |
| X2       | 1  | 2.684596   | 0.17803728     | 15.079                 | 0.0001 |
| X3       | 1  | -5.410582  | 0.27433882     | -19.722                | 0.0001 |
| X4       | 1  | 1.937675   | 0.08422635     | 23.006                 | 0.0001 |

(continued on next page)

(Warning: The values provided are approximate and may not match the exact values in the document.)

### Variable DF Type I SS Type II SS
| INTERCEP | 1  | 968688     | 5049.202823 |
| X1       | 1  | 2362.216057| 3439.005860 |
| X2       | 1  | 21940      | 2380.802    |
| X3       | 1  | (________) | 4072.785802 |
| X4       | 1  | (________) | 5541.709506 |

\[23.006^2*10.47076 > 5541.922\]

a) Compute the rank of X'X. Rank = __________