STAT 2509 C
Assignment #3

DUE: March 3rd, 2015 (to be handed in during the class)

1. A study was conducted involving the relationship between the selling price (in thousands of dollars) of a house (y) and two independent variables, the number of rooms (x_1) and the number of square feet (x_2). The following data were collected on 22 properties sold in a particular residential area;

House	Selling price (y)	Rooms (x_1)	Sq. Ft (x_2)
1	25.75	5	986
2	37.95	5	998
3	46.45	7	1690
4	46.55	8	1829
5	47.95	6	1186
6	49.95	6	1734
7	52.45	7	1684
8	54.05	7	1846
9	54.85	7	1690
10	52.05	7	1910
11	54.39	7	1784
12	53.45	6	1690
13	59.51	7	1590
14	60.10	8	1855
15	63.85	8	2212
16	62.05	10	2784
17	69.45	7	2190
18	82.30	8	2259
19	81.85	7	1919
20	70.05	7	1685
21	112.45	10	2654
22	127.05	10	2756

Consider the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$.

(a) State the MLR model and all assumptions which are necessary for the statistical inference.

(b) Use matrices to compute the estimates of the population parameters β_0, β_1, β_2 and hence obtain the fitted least squares prediction line.

Hint: $\mathbf{X^T X} = \begin{bmatrix} 22 & 160 & 40931 \\ 160 & 1204 & 310676 \\ 40931 & 310676 & 81042365 \end{bmatrix}$, $\mathbf{X^T Y} = \begin{bmatrix} 1364.5 \\ 10430.2 \\ 2719760 \end{bmatrix}$, $\mathbf{Y^T Y} = \sum y_i^2 = 95\ 497.48$

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} 1.68378472 & -0.399634527 & 0.000681592 \\ -0.399634527 & 0.171636536 & -0.000456131 \\ 0.000681592 & -0.000456131 & 0.00000142 \end{bmatrix}, \quad \sum y_i = 1\,364.5$$

(c) Set up the ANOVA table and hence test for the significance of the model. Use $\alpha = 0.05$.

(d) Test whether x_1 term (i.e. whether the number of rooms) contributes to the given model. Use t-test with $\alpha = 0.05$.

(e) Find the values of the coefficient of determination, r^2, and the *adjusted* r^2 and interpret their meanings in this problem.

(f) Run SAS to verify your above results and also use the SAS output to answer part (d) using partial F-test with $\alpha = 0.05$.

2. Consider the following model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \varepsilon$$

where $x_2 = \begin{cases} 1, & if\ drug\ \ B \\ 0, & otherwise \end{cases}$ $x_3 = \begin{cases} 1, & if\ drug\ \ C \\ 0, & otherwise \end{cases}$

$x_1 = \ln(dose)$ y = potency of drug

Run **SAS** to test whether the 3 lines are parallel, i.e. test whether the slopes of these 3 lines are the same. Use $\alpha = 0.05$.

Use the following data:

	Drug Product		
Dose	**A**	**B**	**C**
0.2	2.0	1.8	1.3
0.4	4.3	4.1	2.0
0.8	6.5	4.9	2.8
1.6	8.9	5.7	3.4

Note: Remember to put a **FOOTNOTE** statement with your name and student number in the first line of your program.

Hand in both, your **program** and your **output**, with the appropriate measures **highlighted.**

Hand in only the **required** output when possible.