STAT 2509 B
Assignment 3

Due time: Noon (12:00 pm) February 28, 2020
How to hand in: You can either hand in during the class on February 27, or put your papers in my mailbox or slide them under my door by the due time. If (and only if) you are in Tutorial Section B2, you can gave your papers to your TA.

1. Indicate whether or not each of the following models can be treated as an multiple linear regression (MLR) model:

   (i) $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$

   (ii) $y = \left(e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2^2}\right)\varepsilon$

   (iii) $y = \beta_0 + \beta_1 x_1 + \beta_2 e^{x_1} + \varepsilon$

   (iv) $y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_1^3 + \beta_4 x_2 + \varepsilon$

   (v) $y = \beta_0 e^{\beta_1 x_1 + \beta_2 x_2} + \varepsilon$

2. A medical study was conducted to study the relationship between infants' systolic blood pressure and two explanatory variables, age (days) and weight (kg). The data for 25 infants are given below.

Age (x_1)	Weight (x_2)	Systolic BP (y)
3	2.61	80
4	2.67	90
5	2.98	96
6	3.98	102
3	2.87	81
4	3.41	96
5	3.49	99
6	4.03	110
3	3.41	88
4	2.81	90
5	3.24	100
6	3.75	102
3	3.18	86
4	3.13	93
5	3.98	101
6	4.55	103
3	3.41	86
4	3.35	91
5	3.75	100
6	3.83	105
3	3.18	84
4	3.52	91
5	3.49	95
6	3.81	104
6	4.03	107

Consider the MLR model

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

(a) State all the assumptions that are necessary for the statistical inference under the MLR model.

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

Hint: $\boldsymbol{X}^\top \boldsymbol{X} = \begin{bmatrix} 25.00 & 114.00 & 86.46 \\ 114.00 & 552.00 & 404.07 \\ 86.46 & 404.07 & 304.5062 \end{bmatrix}$, $\boldsymbol{X}^\top \boldsymbol{Y} = \begin{bmatrix} 2380.00 \\ 11072.00 \\ 8306.16 \end{bmatrix}$,

$\left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \approx \begin{bmatrix} 2.3963567 & 0.11058177 & -0.8271483 \\ 0.1105818 & 0.06834592 & -0.1220909 \\ -0.8271483 & -0.12209090 & 0.4001512 \end{bmatrix}$,

$\boldsymbol{Y}^\top \boldsymbol{Y} = \sum_{i=1}^{n} y_i^2 = 228230$, and $\sum_{i=1}^{n} y_i = 2380$.

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

(d) Test whether age (x_1) contributes to explaining (or predicting) the systolic blood pressure (y) under the MLR model. Use t-test with $\alpha = 0.05$.

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

(f) Run SAS to verify your answers to the above questions. In addition, use the SAS output to answer subquestion (d) using the partial F-test with $\alpha = 0.05$.

3. An experimenter wished to compare the potencies of three different drug products. To do this, 12 test tubes were inoculated with a culture of the virus under study and incubated for 2 days at 35°C. Four dosage levels (0.2, 0.4, 0.8, and 1.6 mg per tube) were to be used from each of the three drug products (A, B and C), with only one dose-drug product combination for each of the 12 test-tube cultures. The data are shown in the following table:

Dose	Drug potency (y)		
	Drug A	Drug B	Drug 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

Let

$$x_1 = \ln(\text{dose}), \quad x_2 = \begin{cases} 1, & \text{if drug B} \\ 0, & \text{otherwise} \end{cases}, \quad x_3 = \begin{cases} 1, & \text{if drug C} \\ 0, & \text{otherwise} \end{cases},$$

and $y =$ drug potency. Consider the following MLR model

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

Run SAS to test whether the 3 lines corresponding to the effects of the 3 drugs are parallel (i.e. whether these 3 lines have the same slope). Use $\alpha = 0.05$.

Notes:

- Remember to put a `FOOTNOTE` statement with your name and student number in your SAS program.
- **In addition to the answers to the questions**, hand in both your **code** and your **output**, with the appropriate measures **highlighted**.
- Hand in **only** the **required output** when possible.