Assignment 2

IE 208: Applied Statistics

Due date: **December 9, 2021 at 4:30pm**

Please note:

- You will receive no points for answers without explanation.
- Late submission is not allowed.
- You must submit your work through E-campus: no email submission.
- Prepare your answers in an Excel file and then submit the Excel file.
- 25 points for each question.

In the assignment, use $\alpha = 0.05$.

[Table 1] Sample data

Treatment 1	Treatment 2	Treatment 3	Treatment 4
21	15	18	34
18	27	23	18
12	28	18	29
18	20	13	41
24	29	20	25
13	15	20	

1. (25 pts) There are four treatments and we have sample data of the treatments as shown in Table 1.

(a) Set a proper hypothesis test to check if at least one of the four population means is different. Please complete the following table (show all calculations), and state the conclusion of the hypothesis test by analyzing the *p*-value.

Source	**Degrees of freedom**	**Sum of squares**	**Mean squares**	***F*-statistic**
Treatments	**?**	**?**	**?**	**?**
Error	**?**	**?**	**?**	
Total	**?**	**?**		

(b) Which treatment or treatments do you think have the highest population means? Please provide a detailed statistical support by performing pairwise comparisons.

2. (25 pts) Let's assume that the sample data for Treatments 1, 2, and 3 from Table 1 are paired. We want to check if the population mean of the three treatments are the same. Must use the paired sample assumption in the below questions.

(a) If we want to perform a hypothesis test, what is the null hypothesis? Please explain the null hypothesis in detail.

(b) Perform the hypothesis test. (Is the null hypothesis accepted or rejected?)

[Table 2] Sample data (each row is paired data)

Weight	Score
165	96
172	92
160	88
186	74
185	90
157	65
155	78
177	85
171	82
164	80
182	74

3. (25 pts) Sample data for weight and score are shown in Table 2. Please solve the below questions and show all calculations.

(a) Find the line that best fits the data when setting Score as the dependent variable. Also find the error variance and the R^2 value.

(b) Find the line that best fits the data when setting Weight as the dependent variable. Also find the error variance and the R^2 value.

[Table 3] Sample data (each row is paired data)

	y	x
1	570	22
2	355	20
3	235	15
4	160	14
5	140	10
6	130	12
7	311	18
8	404	21
9	650	25
10	200	15

4. (25 pts) Suppose we want to compare the two models using data from Table 3. Which model better describes y? Statistically show why.

① $y = \beta_0 + \beta_1 x$ ② $y = \beta_2 + \beta_3 x^2$