Carleton University
School of Mathematics and Statistics
STAT 2509 A - Assignment #4

DUE: July 17th, 2025 (to be submitted on BrightSpace before 10:00am)

1. A department store conducted an experiment to investigate the effects of advertising expenditures on the weekly sales for its men's wear, children's wear and women's wear departments. Five weeks for observation were randomly selected from each department and an advertising budget x_1 (in \$ 100's) was assigned for each. The weekly sales y (in \$ 1000's) are shown in the table below for each of the 15 one-week sales periods.

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 $x_2 = \begin{cases} 1, & if\ children's\ wear \\ 0, & otherwise \end{cases}$, $x_3 = \begin{cases} 1, & if\ women's\ wear \\ 0, & otherwise \end{cases}$

and y = weekly sales (in \$ 1000)
x_1 = advertising expenditure (in \$ 100)

	Department		
Advertising Expenditure (x_1)	**Men's wear**	**Children's wear**	**Women's wear**
1	5.2	8.2	10.0
2	5.9	9.0	10.3
3	7.7	9.1	12.1
4	7.9	10.5	12.7
5	9.4	10.5	13.6

Suppose that all the necessary assumptions of multiple linear regression model hold.

(a) Write down/list all possible models (i.e. substitute the values of the dummy variable into the model).

Use SPSS to answer parts b), c) and d).

(b) Test at 5% level of significance, whether the full model can be used for the study.

(c) Test at 5% level of significance, whether the interaction terms are needed in the model.

(d) Based on the results in parts c), examine at 5% level of significance, whether the x_1 variable (i.e. advertising expenditure) is necessary for the prediction of the total weekly sales.

2. Refers to question 1.

Use SPSS to answer following questions and compare your results to results in Q1.

(a) Determine the subset of variables that is selected as best by the **Forward Selection Procedure** using *significance level (to-add-variable),* α = 0.05.

(b) Determine the subset of variables that is selected as best by the **Backward Elimination Procedure** using *significance level (to-delete-variable),* α = 0.10.

(c) Determine the subset of variables that is selected as best by the **Stepwise Regression Procedure** using *significance level (to-add-variable),* α = 0.05 and *significance level (to-delete-variable),* α = 0.10.

3. An ecological study was conducted to compare the rates of growth of vegetation at four swampy undeveloped sites and to determine the cause of any differences that might be observed. Six plants were randomly selected at each of the four sites and the leaf length of ten leaves of each plant were measured. The data in the table are the mean leaf length (in centimeters) per plant.

Site I	**Site II**	**Site III**	**Site IV**
5.7	6.2	5.4	3.7
6.3	5.3	5.0	3.2
6.1	5.7	6.0	3.9
6.0	6.0	5.6	4.0
5.8	5.2	4.9	3.5
6.2	5.5	5.2	3.6

$$\sum_{j=1}^{6} y_{1j} = 36.1, \quad \sum_{j=1}^{6} y_{2j} = 33.9, \quad \sum_{j=1}^{6} y_{3j} = 32.1, \quad \sum_{j=1}^{6} y_{4j} = 21.9$$

$$\sum_{j=1}^{6} y_{1j}^2 = 217.47, \quad \sum_{j=1}^{6} y_{2j}^2 = 192.31, \quad \sum_{j=1}^{6} y_{3j}^2 = 172.57, \quad \sum_{j=1}^{6} y_{4j}^2 = 80.35$$

Establish whether the mean leaf lengths are the same for the four swamp locations; if not, do a follow-up analysis using Tukey's h.s.d. to determine which are the same and which differ. (Use α =0.05). **List all necessary assumptions and test whether the treatment variances are equal.** Also perform a non-parametric analysis. Verify your results using SPSS.