**MT2002 Statistical Modeling**



|  |  |
| --- | --- |
| **Assignment No:** 03 | **CLO:** |
|  | **Semester:** Fall2023 |
| **Due date: As of GCR date** | **Marks:**  60 |

# Instructions

1. **Plagiarized work will result in zero marks.**
2. **No retake or late submission will be accepted.**
3. The submission should be a SINGLE UNZIPPED NOTEBOOK submitted on googles classroom.
4. This notebook should properly document what you did? How you did it? And the source-code for each part as well as the generated outputs.
5. Your submission file should be according to the following **format: id\_section\_A2** e.g., i22123456\_A\_A2. (Note: A2 in the end denotes Assignment 2).

**Marks distribution for this assignment.**

|  |  |
| --- | --- |
| Multiple Linear Model | 20 marks |
| Confounding variable | 40 marks |
| Multi-Collinearity |

Build a multiple linear regression model using the Iris dataset. Use **sepal length** and **petal length** as features. **Petal width** is the dependent variable (y), and **sepal length** and **petal length** are independent variables (x). Your task is to:

Estimate the slopes and intercept via Multiple linear models.

Confounding Variables:

* Identify confounding variables by utilizing the technique discussed in class.

Multicollinearity:

* Investigate multicollinearity among independent variables and apply effective strategies to deal with it in your statistical model. Support your strategy with proper arguments.