STA303H1-S: Methods of Data Analysis II
Assignment 1 - Question 3 Due data February 3, 2023

Student Name.....ID number.....

Instructions: *Show your answers in details.*

Q3 (12 points): In this question, I simulated 60 observations from the following polynomial regression

$$Y = 0.5X + X^2 + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, 1).$$

The **R** code that I used is given below

```
set.seed(7)
x <- rnorm(60)
y <- 0.5*x + x^2 + rnorm(60)
```

1. Use `ggplot()` function to plot X versus Y and argue that the simple linear regression might be a candidate model (reduced model) that can fit the data. Add the fitted line to this plot.
2. Estimate the coefficients of the correct polynomial model (full model).
3. Use `ggplot()` to plot X versus Y and add the smooth polynomial curve to this plot.
4. Use F-test from ANOVA to compare between the two models and draw your final conclusion.
5. Calculate and interpret the coefficient of determination of your selected final model.
6. Determine which (if any) of the observation points are leverage and/or outliers?