**Problem I. (**Each question = 9 points . Show your work for partial credit)

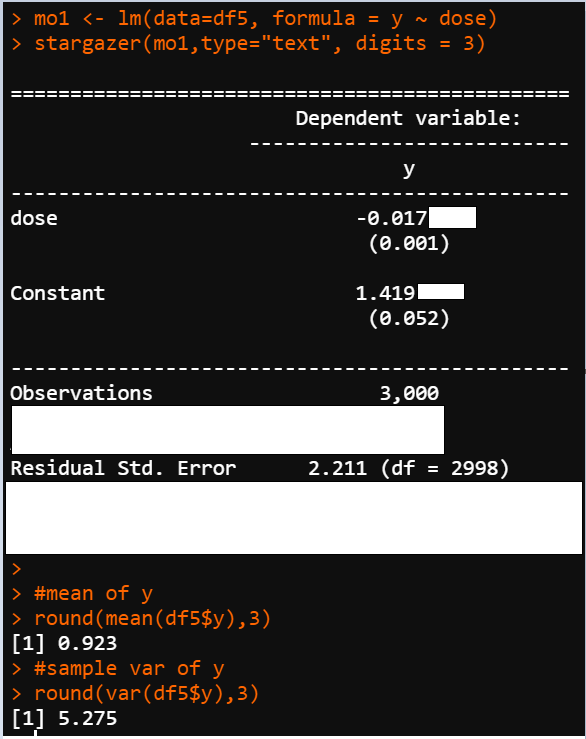
In a phase 3 trail, Moderna is conducting a double blind COVID-19 vaccine trail in which neither the volunteers nor the doctors treating them know what they are getting. Half the volunteers receive Covid-19 vaccine of various doses and the other half receive a placebo shot that simply includes salt water. To analyze the data, the researcher uses the following regression specification:

Where is the number of days that a person stays in hospital due to Covid-19 within 6 months after the trail and is the dose of the vaccine that range between 0 µg – 100 µg and can take values 0 µg (placebo), 25 µg (low dose), 50 µg (mid dose), 100 µg (high dose). The population coefficients of interest are and whereas denotes the error term.

To conduct the study, Moderna recruited 3,000 volunteers in the U.S. and estimated the above equation using an Ordinary Least Squares estimation process. Using a computer-generated algorithm to randomize the participants, half of the sample (n=1,500) received a placebo shot, and the other half received 25 µg, 50 µg, and 100 µg dose shots with, randomly allocated.

The regression output and the descriptive statistics for the outcome is given in **Table 1**.

**Table 1**



Using the results in **Table 1:**

1. Interpret the estimated slope and the intercept.
2. Conduct a statistical test to conclude whether the vaccine is effective. State the null and the alternative hypothesis, calculate the test statistic, p-value and conclude.
3. Using the R output (regression results and descriptive statistics) calculate the average dose of the vaccine received by (i) the full sample, and (ii) excluding those who received the placebo shot.
4. Calculate the R-squared of the regression and interpret the value.
5. Predict the outcome for a patient who received the placebo shot, and patients who received and 50 µg shot.
6. How effective is the vaccine? Interpret the magnitude of the estimated coefficients based on your previous results.

To decide on the right dose of the vaccine, Moderna conducted additional regression analysis using the following three models:

(1)

(2)

(3)

The first regression includes only the control group (n=1,500) and those who received the 25 µg vaccine (n=500) and denoted with a dummy variable , which equals one if the patient is in the treatment group and zero otherwise.

The second regression includes only the control group (n=1500) and those who received the 50 µg vaccine (n=500) and denoted with a dummy variable , which equals one if the patient is in the treatment group and zero otherwise.

The third regression includes only the control group (n=1500) and those who received the 100 µg vaccine (n=500) and denoted with a dummy variable , which equals one if the patient is in the treatment group and zero otherwise.

The regression results of these analyses are given in **Table 2.**

Using the results in **Table 2:**

1. Interpret the estimated coefficients (both Constant and treat) column (1)
2. Compare the results from all the regression models (columns 1,2,3) and recommend a vaccine dose based on your findings.

**Table 2**

