1. In a diabetes study, 1123 subjects were recruited, and a number of clinical traits and information were collected, including (see data in the attached file “HW1\_data.txt”):

Sex: male/female Age: age of the study subject

bmi: body mass index fbg: fasting blood glucose

fins: fasting insulin hba1c: hemoglobin A1c

tg: total glyceride tcho: total cholesterol

hdl: high density lipoprotein ldl: low density lipoprotein

Particularly, the investigators are interested in the effects of fbg (X1) and tg (X2) on hba1c (Y). In addition, hba1c > 6.5 is considered to be diabetic. So another question is whether the same model can be used to characterize the relationship between predictors and hba1c for all individuals, or whether two different models are needed: one for people with diabetes, and the other for those without diabetes. Now, you are assigned to analyze the data, and are asked to complete the following tasks.

1. Fit a simple linear regression model between hba1c and fbg for diabetic patients.
2. Obtain the parameter estimations and the corresponding standard errors.
3. Check on the validity of the assumptions including normality and homoscedasticity, using the approaches you think appropriate.

***Note*** that in order to check whether variances are constant, you may need to put fbg values into several intervals, and consider all fbg value in a particular interval being the same. So use a software to bin the data with the following range: <4; (4, 5); (5, 6.5); (6.5, 8); (8, 10); >10. Then set the original X values to the mid-point of the corresponding interval.

For example, we have five data points with X={4.6, 5.4, 6.1, 7, 7.8}. After the process, we will change the X’s to 4.5, 5.75, 5.75, 7.25, and 7.25, as they fall into intervals (4,5), (5, 6.5), (5, 6.5), (6.5, 8), and (6.5, 8).

1. Complete the ANOVA table, and determine whether the F-statistic will result in a significant result. Interpret your result.
2. Fit a simple linear regression model between hba1c and fbg for all subjects
3. Obtain the parameter estimations.
4. Check on the lack-of-fit. Again, we need to “create” pure errors before we could conduct the test. Use similar process described in A.2 above, and use intervals <4; (4, 4.5); (4.5, 5); …; (9, 9.5); (9.5, 10); >10.
5. Do you think this model similar to the model you obtain in A)? (No formal test needed here. We will study how to conduct a formal statistic test later.)
6. Fit a multiple linear regression model between hba1c and fbg + tg.
7. Obtain the parameter estimations and complete the ANOVA table.
8. Conduct a test for the whole model and interpret your result.
9. Manually calculate the corresponding and adjusted based on the quantities in the ANOVA table. You need to write out the formula and include at least one or two intermediate steps showing the values of the quantities.