NATIONAL UNIVERSITY OF SINGAPORE

Department of Statistics and Applied Probability

ST3131 Regression Analysis

(Semester 2 : AY 2019/2020)

Due Date: Thursday, 9th April 2020, by 4pm

INSTRUCTIONS TO STUDENTS

1. Students are supposed to submit the assignment on time. Any submission after 4 pm of the due date are not accepted.
2. **Students are required to complete this assignment individually.**
3. Be sure to lay out systematically the various steps in your working.
4. Submission is done via Luminus ("Assignment" folder).

The quality of Pinot Noir wine is thought to be related to the properties of clarity, aroma, body, flavor, oakiness and the region (location) where the wine was produced. Data for 38 wines are given in the file `wine.csv` in Luminus. Data can be read in R by

```
> data<-read.table("wine.csv", sep = ",", header = TRUE)
```

Purpose: propose a model that you think it is the best among all the possible models. Your report can be organized in the following manner:

1. **Exploring Data**

   You need to describe the given data set by summary statistics and/or figures/plots, possible distribution of the response, possible relationship between variables.

2. **The Model**

   Fit the multiple linear regression model, present the coefficients table and Anova table. Describe the parameter estimation significance. Give steps of model adequacy checking and propose treatment if it is needed. Do a model selection to find the best model. You may provide some numerical summary/graphs drawn after fitting a model to see if the model fitted adequately. State your final model.

Format of your report:
Your report is limited to **no more than 6 printing pages** (R code is not counted in these 6 pages).
Tables and figures are numbered and referred in the text by number. Formulas are displayed in the center of a separate line.
Remarks: To complete the assignment, you may use almost all the knowledge that you have learnt from the course.