NATIONAL UNIVERSITY OF SINGAPORE

Department of Statistics and Data Science

ST2137 Statistical Computing and Programming

(Semester 1 : AY 2021/2022)

Due Date: 5pm Friday, 19 November 2021

INSTRUCTIONS TO STUDENTS

1. Students are supposed to submit the assignment on time. Any submission after 5 pm of the due date are labelled as late and will get penalty (minus 2% of awarded mark for each hour late).
2. **Students are required to complete this assignment individually.**
3. Be sure to lay out systematically the various steps in your report.
4. Submission is done via Luminus ("Assignment 2 Submission" folder).

The price of a house depends on many factors. The data given in the file `house_selling_prices_OR.csv` (on LumiNUS) concern the selling price of a house relating to other variables given in the data. **Purpose of this assignment: Write a report to propose a linear model that you think it is the best among all the possible linear models and provide the interpretation for your final model.** Some instructions for the report:

1. **Exploring response variable and regressors**

   What you could do:

   - Summarize the response variable (house price in thousand dollars) using summary statistics, figures and/or plots. Comment if it is suitable to fit a linear regression model for this response.

   - Check the possible relationship of the response and the regressors by descriptive statistics, tests, figures or plots (histogram, scatter plot, boxplot, etc).

2. **The Model**

   What you should do:

   - Fit a multiple linear regression model for the response. Present the coefficients table and ANOVA table for the fitted model.

   - Describe the parameter estimation significance (by the test or confidence interval of the coefficients). Give some steps of model adequacy checking and propose a new model if the fitted model is not adequate.

- At first, you might start with a bulky model. However, with the check on the adequacy of the model and the significance test on the regressor(s), you may consider to fit a simpler model.

- If you have proposed a new model which you think it is better than the previous model, then you need to verify why the new model is better than the previous one.

- Investigate if a proposed model has outlier/influential point.

- State your final model and give detail interpretation for this final model (which should include the interpretation on the effect of each regressor to the response).

3. **Format of your report**:

   Your report is limited to **no more than 6 printing pages, font size 12**. The **SAS code must be attached at the end of the report**, as appendix (which is NOT COUNTED in the 6 pages of the report).

   Any table or figure in the report should be numbered clearly.

4. **Remarks**: **Use SAS** for this assignment. If you submit the assignment with the results/code from other software, you will get zero for this assignment.