

Statistical Analysis – 2

Batch 19 Term 2

Group Assignment

Total Points: 40

This deliverable has 40% weightage in the Consolidated Score Sheet.

Due Date: 20th May 2023, 11:55 PM

General Instructions:

1. This is a **group assignment**. Please adhere to the group details posted on LMS.
2. **Do NOT submit .zip files.**
3. Please note that both **the report (pdf file) and the code file (Excel / R / Python file) are mandatory.**
4. Any late submission will attract a penalty as mentioned in the course outline.
5. **Please adhere to the given instructions, otherwise, your submission will not be accepted, or a severe penalty will be applied.**
6. The Honor Code for this submission is **1N**.
7. Upload your submission to the '**SA2 Group Assignment**' submission folder on LMS.
8. **Email submissions are NOT allowed.** All the submissions must be made on LMS.
9. Handwritten content will not be considered for evaluation.
10. There is no penalty for early submissions!

Assignment Deliverables:

1. A **.pdf document** with relevant answers (add Excel / R / Python code outputs for the necessary questions) and explanation
2. **Excel workbook / R / Python code files** used to solve the assignment
3. **Assignment Submission Form** must be attached **separately** with your name and PGID

Assignment Details

Please refer to the Excel sheet for the dataset on House Prices and the variable descriptions.

Sample Selection

- Each group will have a unique sample drawn from the population. This population has 21613 observations. The sample that you are using is considered as sampling with replacement. The sample size for your data is 5000.
- When you open the worksheet titled "House Prices", you will notice that the numbers in Column A (ID) will be unique for your sample. Press function key F9 once and all the numbers will change.
- Copy the range B1:W5001 (5001 rows and 22 columns) and paste them in a separate workbook as values. If you want, you can save this new workbook as a csv file.
- These 5000 observations consisting of 22 variables will be your sample for further analysis.

Assignment Questions

Objective is to identify variables that impact the House Price

1. Carry out EDA to identify if there is any relationship between House Prices and independent variables (4 marks)
2. Use EDA to identify the need for any interaction variables (4 marks)
3. Create additional variables as required (including dummy variables) (4 marks)
4. Fit a regression equation with House Prices as the Dependent variable (4 marks)
5. What is the maximum R^2 and adjusted R^2 possible will all variables in the equation significant at $\alpha = 0.10$? (4 marks)
6. What is the impact of Water Front and View on House Prices? (4 marks)
7. Is there any Multicollinearity? (4 marks)
8. Carry out a test for Heteroskedasticity (4 marks)
9. Carry out a test for Normality assumption about errors (4 marks)
10. Interpret the regression coefficients (4 marks)