

STAT 511: Project #2

Spring 2021

Due on Friday, April 9

Instructions:

- This is a **group** assignment. Each group will submit **one report** with all names.
- Reports will include any relevant R output, summaries, interpretations and conclusions.
- Upload R code as **one separate** file to CANVAS. In total, submit **two files** each group.

Dataset: APPENC07.txt (Real Estate Sales)

The city tax assessor was interested in predicting residential home sales prices in a midwestern city as a function of various characteristics of the home and surrounding property. Data on 522 transactions were obtained for home sales during the year 2002.

Variables of Analysis:

- Response variable (Y): Sales price of a residence (in dollars), Column 2 in the dataset.
- Explanatory variable (X_1): Number of bedrooms, Column 4.
- Explanatory variable (X_2): Number of bathrooms, Column 5.
- Explanatory variable (X_3): Garage size (number of cars that garage can hold), Column 7.

What you will complete:

- Part 1: Model Estimation and Interpretation
 - Fit a multiple linear regression to estimate sales price using all three predictors.

- Interpret **each of the partial slope** coefficients.
 - Report the **adjusted R-squared** of your MLR model and interpret this number.
- Part 2: Prediction
- Predict the sales price for a house with 3 bedrooms, 3 bathrooms, and a 2-car garage in that city.
 - Calculate the 95% **confidence interval** for the sales price in previous question. Interpret this interval.
 - Calculate the 95% **prediction interval** for the sales price in previous question. Interpret this interval.
- Part 3: Hypothesis Testing
- Test the significance of each **partial slope** and conclude. Choose your own α .
 - Conduct an F-test for the **overall model significance** and conclude. Choose your own α .
 - Are the number of bathrooms (X_2) and garage size (X_3) **jointly significant** in the MLR model? Conduct a partial F-test and conclude. Choose your own α .
- Part 4: Multicollinearity
- Obtain the scatterplot matrix and correlation matrix of the three predictors. Does there possibly exist any multicollinearity issue? If so, among which predictors?

- Remove one of two predictors which are strongly correlated from the original MLR model. Run a new MLR model only with the rest two predictors. Does each of the two partial slope coefficients appear to be significant?

Requirements:

Project reports need to be well organized and written. Your report should include

- Report title (a concrete title)
- Author's names and emails
- A table of contents
- Clear labeling of each section and subsection
- Any relevant output
- Writings including summaries, interpretations and conclusions
- Appropriate length (Preferably less than 15 pages).

Reports are graded based on completion, output correctness and report writing. See the evaluation form. There is no unique solution to this project. Therefore, everyone's report should look different.