

# Module 3 Assignment — Multiple Linear Regression Modeling [due Feb 1]



New Attempt

- Due Feb 1, 2022 by 11:59pm
- Points 100
- Submitting a text entry box or a file upload

[Update 1/28: Title of the assignment was changed]

**Deadline: Feb 01st midnight**

## Data & Documentation Files:


- [AmesHousingDataDocumentation.txt](https://northeastern.instructure.com/courses/97835/files/13264858?wrap=1)  
(<https://northeastern.instructure.com/courses/97835/files/13264858?wrap=1>)   
([https://northeastern.instructure.com/courses/97835/files/13264858/download?download\\_frd=1](https://northeastern.instructure.com/courses/97835/files/13264858/download?download_frd=1))
- [AmesHousing.csv](https://northeastern.instructure.com/courses/97835/files/13264859?wrap=1) (<https://northeastern.instructure.com/courses/97835/files/13264859?wrap=1>)   
([https://northeastern.instructure.com/courses/97835/files/13264859/download?download\\_frd=1](https://northeastern.instructure.com/courses/97835/files/13264859/download?download_frd=1))

NOTE: Read the documentation carefully. Perform necessary pre-processing of data such as converting qualitative (text) data into categorical variables or their respective indicator columns.

## Assignment Problem Statement:

As an analytics expert, we should be able to build meaningful models, evaluate them, and interpret them for decision makers. In this assignment, for the housing market data of Ames, Iowa, we perform estimate a linear regression model, interpret results, and evaluate the fitted model.

The **response variable** for analysis will be **Sale Price** of house.

1. Estimate a linear regression model to predict “**Sale Price**” of house in Ames, IA as a function of **relevant explanatory variables** that you think might influence sale price. Write a brief report on the findings in terms of regression coefficient estimates (including intercept) and their statistical significance. Additionally, write remarks on the goodness of fitness statistics (such as  $R^2$ , Adjusted  $R^2$ , and F-statistics reported by the post estimation regression command). 
2. Evaluate the fitted model for its adequacy using following testing strategies.
  - A. Generate two residual plots. First, plot residuals against the statistically significant predictors with highest magnitude of coefficient, and highest t-statistic respectively. Next, plot residuals against

predictors with lowest magnitude of coefficient, and lowest t-statistic respectively. Compare these four plots and interpret your findings in terms of how they differ.

- B. Generate the Q-Q plot and write your findings. Ensure to identify top 5 outlier observations by specifying the necessary option in the R-command used.
- C. Evaluate the fitted model for presence of multicollinearity problem and write your findings.

