

Regression Assignment (15 marks)

PART 1 (10 marks):

Download Housing.csv. Based on these data:

- Visualize some correlations between variables in the data set (2 marks)
- Pick 2 linear regression models to predict median house value
- Check for collinearity using VIF to remove highly correlated variables from the models (1 mark)
- Plot the distribution of the residuals against the fitted values to check for heteroscedasticity (1 mark)
- Use ncvTest or equivalent to test for heteroscedasticity (1 mark)
(<https://www.rdocumentation.org/packages/car/versions/3.0-12/topics/ncvTest>)
- Test for normality of the residuals (use at least one of Wald test, Q-Q plots, etc). 1 mark
- Compare the 2 models using AIC and pick the best model. 1 mark
- Report the coefficients of the winning model and their statistics (including confidence intervals) and interpret the resulting model coefficients. 3 marks

PART 2 (5 marks)

- Predict admission using GRE, GPA, and undergrad institution ranks
- Admission = 1 or 0. Hence use logistic regression (GLM)
- Report the statistics, confidence intervals, etc for the logistic regression and interpret the results (what are the most significant variables that predict whether someone will get admitted? Explain in terms odds by exponentiating the coefficients) - 3 marks
- Can you test an interaction effect? Let's say GPA matters even more if you are from a lower ranked institution (lower GPAs may be tolerated if you are from a higher ranked institution). So include a GPA*rank term in the model and try to interpret the resulting coefficient. - 2 marks