

# STAT 5214G: Advanced Methods of Regression

## Homework 3

**Electronically submit your solutions to Canvas by Sunday June 13th**

If you are a **DAAS student** you are required to submit your solutions using R Markdown.  
All students are required to submit their code.

---

The winequalityred.csv dataset on Canvas contains information about the red wine variants of the Portuguese “Vinho Verde” wine. We are interested in determining the best model for alcohol content ( $y$ ) from the available predictors: fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, and sulphates.

1. **[5 pts]** Fit the winning model from Homework 1 and 2.
2. **[20 pts]** Create a scatterplot matrix with correlations. Comment on what you observe.
3. **[25 pts]** Use tolerance/VIF's to assess whether there are near linear relationships of three or more variables. Comment on what you observe.
4. **[25 pts]** Use condition indices to evaluate whether there is multicollinearity in this model.
5. **[25 pts]** Based on all the information you have (homework 1 - homework 3), what can you conclude about this model?