

What Makes a Good Glass of Wine?

Research Question

Our project seeks to answer the question: what makes a good wine? More specifically, what chemical compositions of wine affect the variation of wine quality? To answer this question, we will analyze the wine quality dataset from Vinho Verde vineyard in Portugal, and more importantly try to narrow down our question to make it possible for it to be supported by evidence.

Data

The Wine Quality dataset was collected from Vinho Verde wine Samples, from northern Portugal and the dataset is divided into two: Red wine and White wine. Red Wine has 1599 observations, and white wine has 4898 observations (each observation being a specific wine). In our analysis, we split **quality** into “good wine” and “bad wine”, defining “good wine” as wine that receives a quality score higher than 7. The variables in the dataset:

- **quality**: the quality of the wine
- **good_wine**: 1 if **quality** > 7, 0 otherwise
- **fixed.acidity**: the amount of acid in wine that's not volatile (do not evaporate fast)
- **volatile.acidity**: the amount of acetic acid in wine
- **citric.acid**: found in small quantities and can add freshness and flavor to wines
- **residual.sugar**: amount of sugar left after fermentation
- **chlorides**: amount of salt in wine
- **free.sulfur.dioxide**: free amount of SO_2 exists in equilibrium between molecular SO_2 and bisulfite ion
- **total.sulfur.dioxide**: total amount of SO_2 , including
- **density**: density of wine measured in g/ml
- **pH**: acidity of wine and hydrogen ion concentration
- **sulphates**: produced by yeast, protecting wine against oxidation
- **alcohol**: percent alcohol content of the wine
- **color**: color of wine, red or white

The data dictionary can be found [here](#).

Methods and Results

We fit two different models to the wine quality dataset. The first model was obtained through a step-wise AIC algorithm, and produced a reduced model with the best AIC. The second model was a logistic regression model with interactive terms, which we considered to be the full model. After cross validation and more steps of model evaluation, we determined that the best AIC model is the final model.