Introduction

Wine Quality evaluation is a main part of the certification process of wines and can be used to improve wine making and to stratify wines such as premium brands or pricing. And this might be also important part of our better lifestyle to choose good quality wine based of facts.

Red Wine Quality Data Set

I found the available dataset about red wine quality from the UCI machine learning repository (https://archive.ics.uci.edu/ml/datasets/wine+quality). The dataset contains a total of 12 variables such as critic acid levels, pH, alcohol, and so on. There was also a quality measure between 0 and 10. These were recorded for 1,599 observations.

Hypothesis: What factors makes better quality of wine?

H0: None of the variables effects on the quality of wines.

H1: The variables effects on the quality of wines and make a wine 'low', 'mid' and 'high' quality of wines

Justification for the inclusion of the control

To understand, how much each attribute correlates with the quality score of wine, I resulted first of all the standard correlation coefficient between every variable. The correlation coefficient ranges from –1 to 1. When it is close to 1, it means that there is a strong positive correlation. In opposite way, if the coefficient is close to –1, it means that there is a strong negative correlation to each other. The coefficient 0 means, that there is no linear correlation.

Concludes

- "Quality" tends to go up when the "alcohol" goes up.
- "volatile acidity" and the "quality" results a slightly negative correlation.
- The correlation between attributes such as "fixed acidity" and "density" shows 0.67 as a slightly positive correlation (the correlation matrix is below)
- Tuning features like 'alcohol', 'fixed acidity and 'pH' may make wine scores higher or lower by having higher p values. **This shows that it would be profitable to change the physicochemical properties of wine for improving wine quality and human preference on wine.**

Results: Regression table (Correlation, Coefficients, Standard Errors, T values, P values)
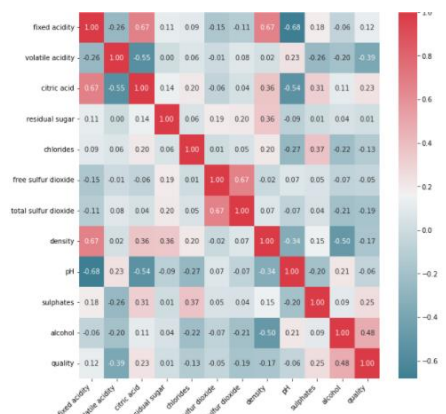
	Coefficients	Standard Errors	t values	p values
fixed acidity	7.663025	14.193401	0.539901	5.893414e-01
volatile acidity	0.018661	0.017377	1.073915	2.830240e-01
citric acid	-0.548181	0.081098	-6.759485	1.939293e-11
residual sugar	-0.154798	0.098560	-1.570597	1.164757e-01
chlorides	0.008009	0.010046	0.797214	4.254459e-01
free 4sulfur dioxide	-0.922925	0.280782	-3.286981	1.034799e-03
total sulfur dioxide	0.004456	0.001454	3.064570	2.216560e-03
density	-0.003246	0.000488	-6.652516	3.954126e-11
pH	-8.497988	14.487069	-0.586591	5.575617e-01
sulphates	-0.133681	0.128307	-1.041878	2.976269e-01
alcohol	0.553347	0.076569	7.226815	7.651657e-13