Introduction

1:1 Problem statement and research motivation

The ever-growing wine industry has seen a decline in both consumption and production recently. The growth of the industry relies on the satisfaction of consumers. Thus, the relationship between alcohol content and wine quality is a critical topic in the wine industry, as prior research suggests that alcohol content can significantly impact the sensory characteristics of a wine. Ensuring the wine industry relishes its glory days again with the help of our research gave us the motivation to figure out whether wine with higher alcohol content has a significantly higher quality rating than then the wines with lower alcohol content.

1:2 The data set

Our Dataset name is the wine dataset, and we took it from Kaggle. It contains 1599 samples and 12 features. Our research mainly focuses on alcohol content(percentage) and wine quality, rated on a scale of 1 to 10. Features include acidity, residual sugar, and pH levels, which give us extra information about wine characteristics. Our dataset is relevant for inquiring about the relationship between alcohol content and wine quality while considering other wine attributes.

1:3 Research question

Is there a difference in the average quality rating between wines with low alcohol content and those with high alcohol content?

We will investigate this through statistical hypothesis testing. We'll use histograms to assess normality, boxplots to illustrate differences, and t-tests or Wilcoxon tests to evaluate significance. This analysis will be performed in R using RStudio, utilizing the Kaggle Wine Dataset to explore the effect of alcohol content on wine quality.

1:4 Null hypothesis and alternative hypothesis (H0/H1)

In this analysis, we want to investigate the effect of alcohol content on the quality rating of white wine. To determine this, we approach it scientifically, establishing two competing hypotheses.

NULL HYPOTHESIS: There is no difference in the mean quality ratings between the wines with low alcohol content and wines with high alcohol content.

 ALTERNATIVE HYPOTHESIS: There is a difference in the mean quality ratings between wines with low alcohol content and wines with high alcohol content.

This will be tested using wilcox Test. The alcohol concentration will be separated into low and high categories. And lastely, the p value determines which to reject and accept.