

Components That Influence the Rating of Wine

Zhongjie Wu

Department of Economics, Lehigh University

BIS 044: Business Analytics I

Dr. Troy Adair

December 1, 2021

Components That Influence the Rating of Wine

1. Introduction and statement of hypothesis

Research Purpose and Motivation

Humans have a rich history of brewing and tasting wine.

Statement of Hypothesis

- Sulphates has positive effect on the quality rating of wine.
- Chlorides has negative effect on the quality rating of wine.

2. Description of data sources and methodology

Description of Data

I obtained the dataset from Kaggle, one machine learning website with a lot of useful datasets. The original dataset is uploaded by Kaggle user “ruthgn”, while I subtracted some irrelevant variables and only kept one dependent variable: Quality, and two independent variables: Sulphates and Chlorides.

- Quality: This variable is measured by ratings provided by customers and wine experts. The data range from 0 to 10. 0 refers to “very bad” while 10 refers to “very good”. Data type of the variable is float.
- Sulphates: Sulphates usually exist as sulfur dioxide molecules and sulfite ions in wine. The amount of Sulphates is measured as viscosity (g/dm³). Data type of the variable is float.

- Chlorides: We could consider chlorides as salt exists in wine. The amount of Chlorides is measured as viscosity (g/dm^3). Data type of the variable is float.

Methodology

Considering that this is the assignment that measures our ability of data analysis by visualizations. I would use histogram to check if the independent variables are normally distributed so that they don't disturb the rules of simple regression. I would use two scatter plots with regression lines to check my two hypotheses.

However, examine hypotheses by eyes is not serious enough. I would use simple regression method as well to see if my hypotheses are correct.

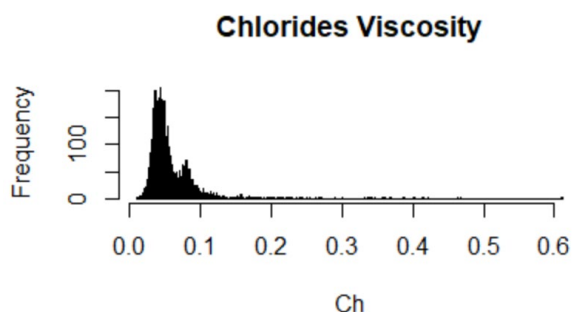
3. Analysis/visualizations (with explanations)

Histogram of Chlorides

```
wine <- read.csv(file = 'C:\\Users\\wuzho\\Downloads\\Wine.csv')
```

```
Ch <- wine$chlorides
```

```
hist(Ch, breaks=1000, main="Chlorides  
Viscosity", col="darkmagenta")
```

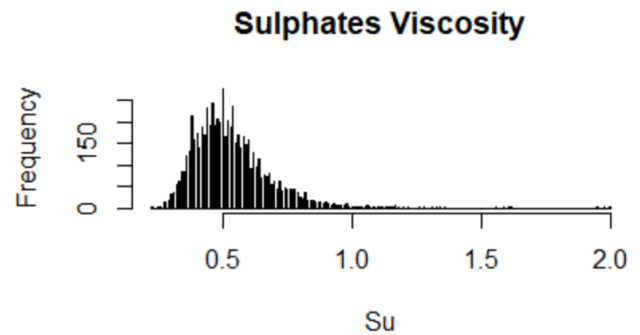


We could tell from the diagram that the distribution of Chlorides is skewed. We may need to remove observants with Chlorides larger than 0.15.

Histogram of Sulphates

```
Su <- wine$sulphates
```

```
hist(Su, breaks=1000, main="Sulphates  
Viscosity", col="darkmagenta")
```



We could tell from the histogram that the distribution of Sulphates is also skewed. We need to remove observants with value larger than 1.0.

Data Cleaning

From the data cleaning suggestions above, we removed observants with unsatisfactory value.

```
df<-wine[!(wine$chlorides > 0.15 | wine$sulphates > 1),]
```

After data cleaning, we still have 6337 observants.

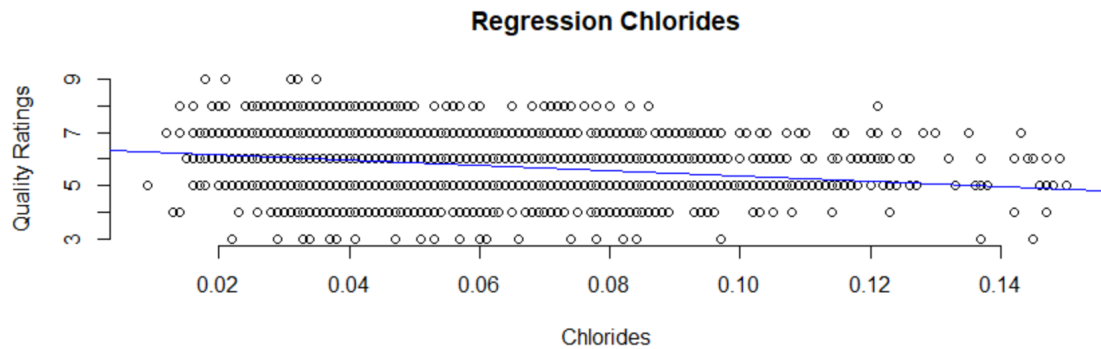
Scatter plot of Chlorides

```
x <- df$chlorides
```

```
y <- df$quality
```

```
plot(x, y, main = "Regression Chlorides", xlab = "Chlorides", ylab = "Quality  
Ratings", pch = 1, frame = FALSE)
```

```
abline(lm(y ~ x, data = df), col = "blue")
```



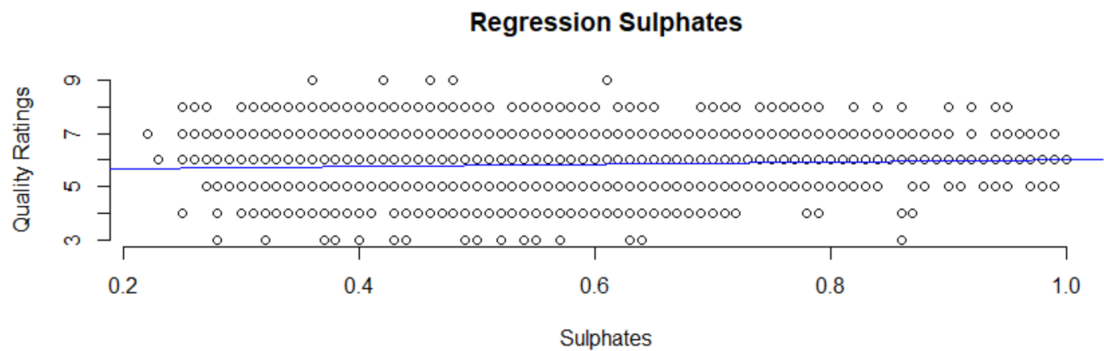
Scatter Plot of Sulphates

```
x <- df$sulphates
```

```
y <- df$quality
```

```
plot(x, y, main = "Regression Sulphates", xlab = "Sulphates", ylab = "Quality  
Ratings", pch = 1, frame = FALSE)
```

```
abline(lm(y ~ x, data = df), col = "blue")
```



Simple Linear Regression

1. SLR on quality and chlorides

```
res <- lm (quality ~ chlorides, data = df)
```

```
res
```

```
Call:
lm(formula = quality ~ chlorides, data = df)

Coefficients:
(Intercept)    chlorides
      6.354       -10.060
```

2. SLR on quality and sulphates

```
res <- lm (quality ~ sulphates, data = df)
```

```
res
```

```
Call:
lm(formula = quality ~ sulphates, data = df)

Coefficients:
(Intercept)    sulphates
      5.6069       0.4193
```

4. Summary

From the previous image, Sulphates has positive effect on the quality rating of wine, and Chlorides has negative effect on the quality rating of wine. In summary, the data shows that my conjecture is correct.