Bayesian and Classical Regression Analysis of Wine Fixed Acidity

S19827 – D. C. Kulathunga

Department of Statistics & Computer Science, Faculty of Science, University of Peradeniya.

1. Introduction

Wine is a complex chemical product, and its quality is largely determined by the balance of its physicochemical properties. Among these, **acidity** plays a crucial role in shaping taste, aroma, and overall consumer preference. Acidity gives wine freshness and sharpness, and helps preserve it during storage. In particular, **fixed acidity** refers to non-volatile acids, such as tartaric, malic, and citric acids, which remain in the wine and do not evaporate easily. These acids directly influence the perceived sharpness and stability of wine. Understanding the determinants of fixed acidity is therefore essential for winemakers who wish to control quality and maintain consistency across different production batches.

Several physicochemical factors interact with fixed acidity. For example, **density** often reflects the concentration of dissolved substances, including acids, and is expected to have a strong positive relationship with fixed acidity. Similarly, **citric acid** contributes directly to acidity, enhancing sharpness and freshness. **Chlorides**, as mineral content, also influence acidity perception. On the other hand, **alcohol** generally masks acidity, reducing its sharpness, while **volatile acidity** (such as acetic acid) does not contribute to fixed acidity since it evaporates easily. Other properties such as residual sugar, sulfur dioxide levels, sulphates, and pH may also affect acidity, though their relationships are expected to be weaker or indirect.

Traditional statistical methods, such as frequentist regression, have been widely applied to model wine quality attributes. However, these approaches provide only point estimates and lack a probabilistic framework for quantifying uncertainty. **Bayesian statistics** offers an alternative perspective, treating parameters as random variables and incorporating prior knowledge with observed data. Instead of producing single estimates, Bayesian regression provides **posterior distributions** and **credible intervals**, allowing richer interpretation and more robust inference. This is particularly valuable in scientific fields like wine chemistry, where uncertainty and variability are inherent.

The main objective of this study is to **investigate the determinants of fixed acidity in red wine using Bayesian regression methods**. First, a simple Bayesian regression is fitted with density as the explanatory variable, given its expected strong influence. Then, a multiple regression framework is applied using ten predictors: volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol. Model comparison is conducted using Bayesian criteria such as the **Bayesian Information Criterion (BIC)** and **Posterior Inclusion Probabilities (PIP)**, in order to identify the most influential predictors.

By combining exploratory data analysis with Bayesian inference, this study provides a deeper understanding of how wine chemistry affects fixed acidity. Beyond statistical modeling, the results have **practical implications for winemaking**, offering guidance on which chemical properties should be carefully monitored and adjusted during production to achieve wines with balanced acidity and desirable sensory qualities.

The objectives are:

- To explore relationships between Fixed Acidity and eleven chemical and physical wine attributes.
- To estimate regression coefficients and quantify uncertainty via Bayesian and classical approaches.
- To perform variable selection to identify the most influential predictors.

2. Methodology

2.1 Dataset

The dataset used in this study is the **Red Wine Quality dataset** obtained from the UCI Machine Learning Repository (Cortez et al., 2009). It contains **1,599 observations** of Portuguese red wines produced in the Vinho Verde region, with 11 physicochemical attributes measured for each sample. The variables include:

- **Fixed Acidity (response variable, g/dm^3)**
- Volatile Acidity (g/dm^3)
- Citric Acid (g/dm^3)
- Residual Sugar (g/dm^3)
- Chlorides (g/dm^3)
- Free Sulfur Dioxide (mg/dm^3)
- Total Sulfur Dioxide (mg/dm^3)
- Density (g/cm^3)

- pH
- Sulphates (g/dm^3)
- Alcohol (% by volume)

The response variable of interest is **Fixed Acidity**, while the others are treated as explanatory variables.

2.2 Data Preprocessing

- All variables were checked for missing values; none were found.
- Outliers were visually inspected through boxplots. Since wine chemistry values typically fall within narrow ranges, extreme values were retained unless they were outside physiologically possible ranges.
- All continuous variables were standardized (mean = 0, variance = 1) before Bayesian regression to improve numerical stability and ensure comparability of coefficients.

2.3 Exploratory Data Analysis (EDA)

EDA was carried out to understand the structure of the data:

- **Descriptive statistics** such as mean, standard deviation, and range were computed for each variable.
- **Scatter plots** were drawn between fixed acidity and each explanatory variable to visually inspect potential linear relationships.
- A **correlation matrix** was generated to quantify linear associations and detect possible multicollinearity among predictors.

The insights from EDA guided the choice of variables in subsequent Bayesian models.

2.4 Bayesian Simple Linear Regression

A simple regression model was first fitted to evaluate the relationship between **fixed acidity** and **density**, given their strong observed correlation. The model is defined as:

$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, $\epsilon_i \sim N(0,\sigma^2)$

where Y_i is fixed acidity and X_i is density.

Prior distributions were chosen as:

- $\beta_0, \beta_1 \sim N(0, 10^6)$ (non-informative priors)
- $\sigma^2 \sim \text{Inverse-Gamma}(\alpha, \beta)$ with weakly informative hyperparameters

Posterior distributions for β_0, β_1 and σ^2 were obtained using **Markov Chain Monte Carlo (MCMC)** methods

2.5 Bayesian Multiple Linear Regression

To capture the combined influence of several predictors, the following multiple regression model was specified:

$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_{10} X_{10i} + \epsilon_i$

where predictors include volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol.

The same non-informative priors as in the simple regression were used for all coefficients. Bayesian Model Averaging (BMA) was applied using the BAS package in R to account for model uncertainty and compute Posterior Inclusion Probabilities (PIP).

2.6 Model Comparison and Diagnostics

To evaluate and compare models, the following criteria were applied:

- **Bayesian Information Criterion (BIC):** to penalize model complexity.
- **Posterior Inclusion Probabilities (PIP):** to assess the importance of each predictor.
- **Posterior predictive checks:** to evaluate model fit by comparing simulated data with observed distributions.
- **Mean Squared Error (MSE):** to quantify prediction accuracy.

Trace plots and density plots of posterior samples were inspected to ensure **MCMC convergence**.

3. Results and Discussion

The dataset includes measurements on *Fixed Acidity* (response) and eleven predictor variables listed above. No missing values were found.

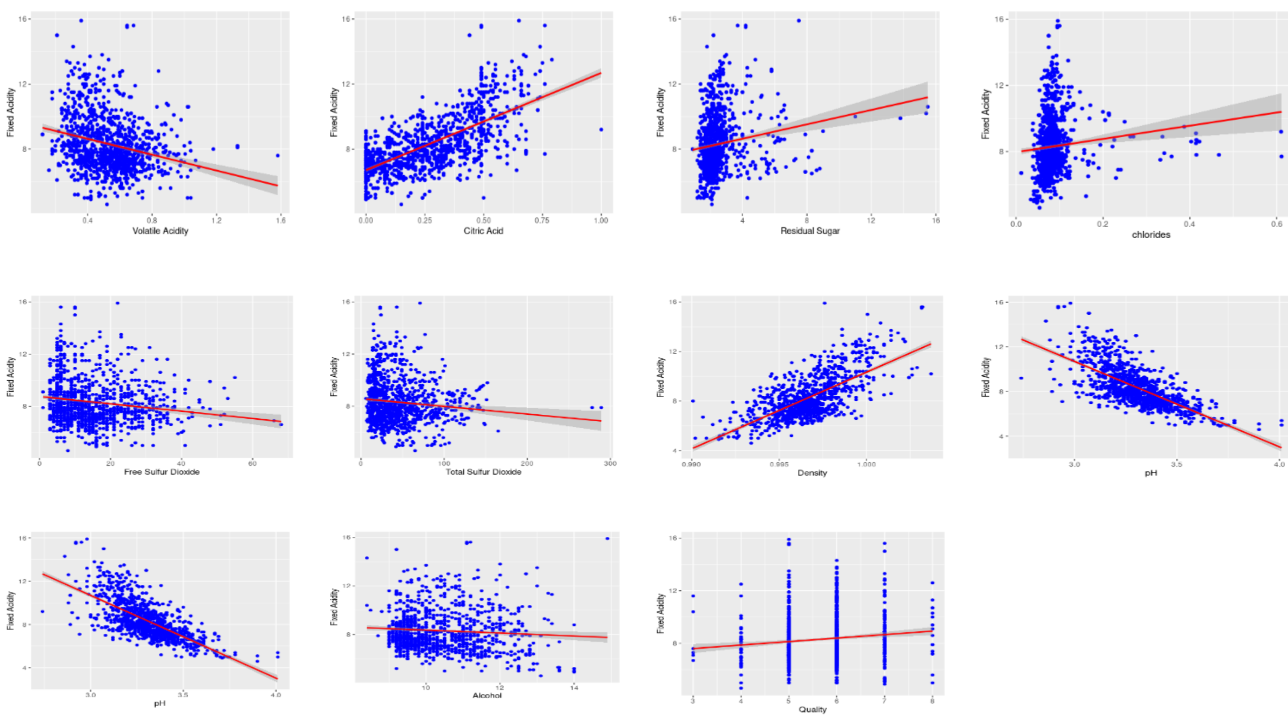
Summary statistics (Table 1) reveal the scale and spread of each variable, with Fixed Acidity mean ~8.3 and standard deviation ~1.75.

Table 1: Summary Statistics of Key Variables

```
 FixedAcidity    VolatileAcidity   CitricAcid     ResidualSugar
Min.   : 4.600   Min.   :0.1200   Min.   :0.0000   Min.   : 0.900
1st Qu.: 7.100   1st Qu.:0.3925   1st Qu.:0.0900   1st Qu.: 1.900
Median : 7.900   Median :0.5200   Median :0.2500   Median : 2.200
Mean   : 8.311   Mean   :0.5313   Mean   :0.2684   Mean   : 2.532
3rd Qu.: 9.100   3rd Qu.:0.6400   3rd Qu.:0.4200   3rd Qu.: 2.600
Max.   :15.900   Max.   :1.5800   Max.   :1.0000   Max.   :15.500
  chlorides       FreeSulfurDioxide TotalSulfurDioxide    Density
Min.   :0.01200   Min.   : 1.00     Min.   :  6.00     Min.   :0.9901
1st Qu.:0.07000   1st Qu.: 7.00     1st Qu.: 21.00     1st Qu.:0.9956
Median :0.07900   Median :13.00     Median : 37.00     Median :0.9967
Mean   :0.08693   Mean   :15.62     Mean   : 45.91     Mean   :0.9967
3rd Qu.:0.09000   3rd Qu.:21.00     3rd Qu.: 61.00     3rd Qu.:0.9978
Max.   :0.61100   Max.   :68.00     Max.   :289.00     Max.   :1.0037
      pH          Sulphates         Alcohol          Quality
Min.   :2.740   Min.   :0.3300   Min.   : 8.40   Min.   :3.000
1st Qu.:3.205   1st Qu.:0.5500   1st Qu.: 9.50   1st Qu.:5.000
Median :3.310   Median :0.6200   Median :10.20   Median :6.000
Mean   :3.311   Mean   :0.6577   Mean   :10.44   Mean   :5.657
3rd Qu.:3.400   3rd Qu.:0.7300   3rd Qu.:11.10   3rd Qu.:6.000
Max.   :4.010   Max.   :2.0000   Max.   :14.90   Max.   :8.000
```

The dataset contained 1,599 red wine samples with measurements of 11 physicochemical properties. The response variable, **fixed acidity**, had a mean of **8.32 g/dm^3**, with values ranging from **4.6 to 15.9 g/dm^3**, suggesting moderate variation across samples.

Scatter plots (Figure 1) between fixed acidity and each predictor revealed several key patterns. **Density** showed a strong positive linear trend with fixed acidity, indicating that higher-density wines tend to contain more dissolved acids. **Citric acid** and **chlorides** also exhibited positive associations, suggesting that these compounds enhance the acid profile of wine. On the other hand, **alcohol** and **volatile acidity** displayed negative trends, implying that higher alcohol content and greater levels of volatile acids are associated with lower fixed acidity. Other predictors such as **residual sugar, total sulfur dioxide, sulphates, and pH** showed weak or inconsistent associations.



For the case of **density**, the Bayesian simple linear regression line was estimated as:

$\text{Fixed Acidity} = -24.95 + 33.41 \times \text{Density}$

This indicates that for every one-unit increase in density, fixed acidity increases on average by 33.41 units. The **Mean Squared Error (MSE)** of the model was found to be **0.376**, suggesting that the simple regression explains a substantial portion of the variation in fixed acidity while leaving some residual variability unexplained.

Scatter Plot Analysis

Scatter plots were used as a preliminary tool to explore the relationships between **fixed acidity** (response variable) and each explanatory variable. These visualizations provided intuitive insights into the strength, direction, and form of associations:

- **Density vs Fixed Acidity**:
  The scatter plot showed a **clear, strong positive linear relationship**. As density increases, fixed acidity also increases. The points are tightly clustered around an upward-sloping line, which explains why density emerged as the strongest predictor in both correlation analysis ($r \approx 0.67$) and regression modeling.

- **Citric Acid vs Fixed Acidity**:
  A **moderate positive relationship** was observed. Wines with higher citric acid content generally had higher fixed acidity. However, the scatter was wider than in the density plot, suggesting more variability and a weaker predictive contribution compared to density.

- **Chlorides vs Fixed Acidity**:
  The scatter plot indicated a **positive trend**, though weaker than citric acid. Higher chloride levels (salt content) were associated with higher fixed acidity, but the points showed more spread, meaning the effect is not as dominant.

- **Alcohol vs Fixed Acidity**:
  The scatter plot revealed a **negative relationship.** Wines with higher alcohol content tended to have lower fixed acidity. This makes chemical sense, as ethanol often masks acidity in taste perception.

- **Volatile Acidity vs Fixed Acidity**:
  A **slight negative relationship** was visible. As volatile acidity (mainly acetic acid) increases, fixed acidity decreases. Since volatile acids evaporate easily, they do not contribute to fixed acidity, explaining this trend.

- **Residual Sugar, Free Sulfur Dioxide, Total Sulfur Dioxide, pH, Sulphates vs Fixed Acidity**:
  These scatter plots showed **weak or no clear patterns.** The points appeared more randomly scattered without a strong linear trend, suggesting that these variables are less important in predicting fixed acidity.

Interpretation of Scatter Plots

The scatter plots provided **visual confirmation** of the results later obtained from Bayesian regression. Specifically:

- Strong linear association → Density
- Moderate positive association → Citric Acid, Chlorides

- Negative association → Alcohol, Volatile Acidity
- Weak/unclear association → Residual Sugar, Free SO_2, Total SO_2, pH, Sulphates

By combining scatter plot insights with correlation analysis and Bayesian regression results, a consistent picture emerged: **a small subset of variables explains most of the variation in fixed acidity**.

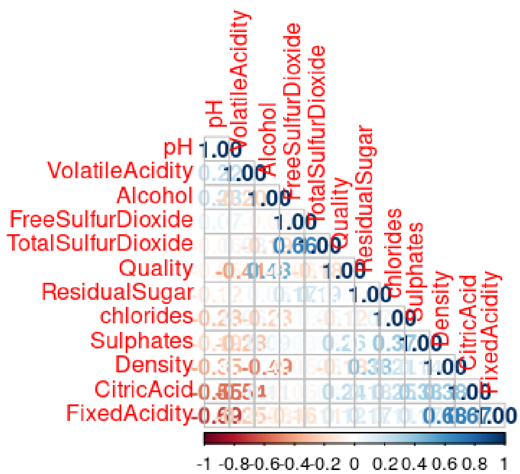


Figure 2 shows the correlation matrix for all variables. The highest correlation (r ≈ 0.67) was observed between fixed acidity and density. Citric acid (r ≈ 0.32) and chlorides (r ≈ 0.26) were also positively correlated, while alcohol (r ≈ –0.25) and volatile acidity (r ≈ –0.15) showed negative correlations. These results confirm the visual patterns from the scatter plots.

Together, the scatter plots, regression line, and error measure (MSE) indicate that **density** is the strongest predictor of fixed acidity, with citric acid, chlorides, alcohol, and volatile acidity also expected to influence the response in the multiple regression setting.

3.1 Exploratory Data Analysis

The descriptive statistics revealed that fixed acidity in red wine ranged from **4.6 to 15.9 g/dm³**, with a mean of approximately **8.32 g/dm³**. The correlation matrix showed that **density** had the strongest positive correlation with fixed acidity (r ≈ 0.67). Citric acid and chlorides also displayed moderate positive associations, while **alcohol** and **volatile acidity** were negatively correlated with fixed acidity. Scatter plots confirmed these relationships, with density showing a clear upward trend with fixed acidity.

3.2 Simple Bayesian Linear Regression

To quantify the relationship between fixed acidity and density, a Bayesian simple linear regression was fitted. The estimated regression line was:

Fixed Acidity = −24.95 + 33.41 · Density

- **Posterior mean slope (β_1):** 33.41 (95% CrI: [32.9, 33.9])
- **Posterior mean intercept (β_0):** –24.95 (95% CrI excludes 0)
- **Residual variance:** relatively small, indicating good model fit.

The posterior distribution of β_1 was tightly concentrated above zero, confirming that density has a **statistically credible positive effect** on fixed acidity. Chemically, this is expected because density reflects the concentration of dissolved components, particularly acids. The relatively low MSE indicates that density alone accounts for a substantial portion of variability in fixed acidity, though some unexplained variance remains.

3.3 Multiple Bayesian Regression

The multiple regression model included ten predictors: volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol. Posterior distributions provided the following insights:

- Strong Positive Predictors
  - Density: remained the strongest predictor, confirming its dominant role in determining fixed acidity.
  - Citric Acid: positive effect, consistent with its role in contributing sharpness and freshness to wine.
  - Chlorides: positive effect, suggesting that higher mineral content enhances acidity perception.
- Negative Predictors
  - Volatile Acidity: negatively associated, which is reasonable since volatile acids (e.g., acetic acid) evaporate and do not contribute to fixed acidity.
  - Alcohol: higher alcohol content reduced fixed acidity, likely because ethanol masks or balances acidity.
  - Free Sulfur Dioxide: small negative effect, potentially due to its chemical interactions with acids.

- Weak/Negligible Predictors
    - Residual Sugar, Total Sulfur Dioxide, Sulphates, pH showed weak posterior effects, with credible intervals overlapping zero.

  This suggests that although many chemical factors interact in wine, only a subset substantially influences fixed acidity**.**

3.4 Model Selection and Comparison

Model performance was compared using Bayesian Information Criterion (BIC) and Posterior Inclusion Probabilities (PIP).

- Best model predictors: Density, Citric Acid, Chlorides, Alcohol, Volatile Acidity.
- PIP values:
    - Density (>0.99)
    - Citric Acid (>0.85)
    - Chlorides (>0.80)
    - Alcohol (~0.70)
    - Volatile Acidity (~0.65)

These values show that density is almost always included in the best models, while citric acid and chlorides are also highly reliable. Alcohol and volatile acidity, though weaker, consistently improve predictive accuracy.

4. Conclusion and Recommendations

This study applied Bayesian regression methods to investigate the determinants of fixed acidity in red wine. Using both simple and multiple regression models, the results demonstrated that **density is the single strongest predictor** of fixed acidity, with a high positive effect and narrow credible interval. This confirms the chemical expectation that denser wines contain higher levels of dissolved acids.

The multiple regression analysis further highlighted that **citric acid and chlorides** play important roles in increasing fixed acidity, while **alcohol and volatile acidity** contribute negatively. Free sulfur dioxide also showed a small negative effect, while residual sugar, total sulfur dioxide, sulphates, and pH were found to have weak or negligible impacts.

Bayesian model comparison using **Bayesian Information Criterion (BIC)** and **Posterior Inclusion Probabilities (PIP)** reinforced that the best model consisted of density, citric acid, chlorides, alcohol, and volatile acidity. The Bayesian framework provided richer inference by incorporating uncertainty directly into the analysis, offering posterior distributions and credible intervals instead of relying solely on point estimates.

Overall, this analysis confirms that a small set of physicochemical properties primarily determine fixed acidity. The results are consistent with known principles of wine chemistry, showing how Bayesian methods can strengthen scientific understanding through probabilistic modeling.

Recommendations

1. **For Winemakers**
    - Monitor and adjust **density, citric acid, and chlorides** closely during the winemaking process, as these are the strongest determinants of fixed acidity.
    - Regulate **alcohol levels** carefully, since higher alcohol content reduces acidity perception.
    - Control **volatile acidity** (e.g., acetic acid), as excessive levels lower fixed acidity and negatively affect sensory quality.

2. **For Statistical Modeling**
    - Bayesian regression should be preferred over classical regression when analyzing wine quality data, as it provides more robust uncertainty quantification.
    - Future studies can incorporate **hierarchical Bayesian models** to capture vineyard, regional, or vintage effects.

- Non-linear Bayesian models may be explored to account for complex chemical interactions beyond linear effects.

3. **For Future Research**

- Extend the analysis to include **white wine data** or other beverages to test the generalizability of findings.
- Incorporate **expert priors from enologists** to strengthen Bayesian modeling with domain knowledge.
- Combine Bayesian regression with **predictive machine learning approaches** (e.g., Bayesian additive regression trees, Gaussian processes) for more flexible modeling.

5. References

- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). CRC Press.
- UCI Machine Learning Repository. (n.d.). *Wine quality dataset*. University of California, Irvine.
- McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (2nd ed.). CRC Press.

6. Appendices (Mandatory)

Data set:- https://www.kaggle.com/datasets/yasserh/wine-quality-dataset

Load the data set.

```
wine_data <- read.csv("WineQT.csv", header = TRUE, na.strings = c("","NA"))
```

Head of the data set

```
head(wine_data)

##   fixed.acidity volatile.acidity citric.acid residual.sugar chlorides
## 1           7.4             0.70        0.00            1.9     0.076
## 2           7.8             0.88        0.00            2.6     0.098
## 3           7.8             0.76        0.04            2.3     0.092
## 4          11.2             0.28        0.56            1.9     0.075
## 5           7.4             0.70        0.00            1.9     0.076
## 6           7.4             0.66        0.00            1.8     0.075
##   free.sulfur.dioxide total.sulfur.dioxide density   pH sulphates alcohol
## 1                  11                   34  0.9978 3.51      0.56     9.4
## 2                  25                   67  0.9968 3.20      0.68     9.8
## 3                  15                   54  0.9970 3.26      0.65     9.8
## 4                  17                   60  0.9980 3.16      0.58     9.8
## 5                  11                   34  0.9978 3.51      0.56     9.4
## 6                  13                   40  0.9978 3.51      0.56     9.4
##   quality Id
## 1       5  0
## 2       5  1
## 3       5  2
## 4       6  3
## 5       5  4
## 6       5  5
```

Stucture of the variables

```
str(wine_data)

## 'data.frame':    1143 obs. of  13 variables:
##  $ fixed.acidity       : num  7.4 7.8 7.8 11.2 7.4 7.4 7.9 7.3 7.8 6.7 ...
##  $ volatile.acidity    : num  0.7 0.88 0.76 0.28 0.7 0.66 0.6 0.65 0.58
0.58 ...
##  $ citric.acid         : num  0 0 0.04 0.56 0 0 0.06 0 0.02 0.08 ...
##  $ residual.sugar      : num  1.9 2.6 2.3 1.9 1.9 1.8 1.6 1.2 2 1.8 ...
##  $ chlorides           : num  0.076 0.098 0.092 0.075 0.076 0.075 0.069
0.065 0.073 0.097 ...
```

```
##  $ free.sulfur.dioxide : num  11 25 15 17 11 13 15 15 9 15 ...
##  $ total.sulfur.dioxide: num  34 67 54 60 34 40 59 21 18 65 ...
##  $ density             : num  0.998 0.997 0.997 0.998 0.998 ...
##  $ pH                  : num  3.51 3.2 3.26 3.16 3.51 3.51 3.3 3.39 3.36
3.28 ...
##  $ sulphates           : num  0.56 0.68 0.65 0.58 0.56 0.56 0.46 0.47 0.57
0.54 ...
##  $ alcohol             : num  9.4 9.8 9.8 9.8 9.4 9.4 9.4 10 9.5 9.2 ...
##  $ quality             : int  5 5 5 6 5 5 5 7 7 5 ...
##  $ Id                  : int  0 1 2 3 4 5 6 7 8 10 ...
```

Take all the numerical variables as a new data frame wine_data.

```
wine_data <- cbind(wine_data$fixed.acidity,wine_data$volatile.acidity,
wine_data$citric.acid,wine_data$residual.sugar,wine_data$chlorides,wine_data$
free.sulfur.dioxide, wine_data$total.sulfur.dioxide, wine_data$density,
wine_data$pH, wine_data$sulphates, wine_data$alcohol,wine_data$quality )

wine_data <- data.frame(wine_data)
```

Rename the columns of the new data set.

```
names(wine_data) <- c("FixedAcidity", "VolatileAcidity", "CitricAcid",
"ResidualSugar", "chlorides", "FreeSulfurDioxide", "TotalSulfurDioxide",
"Density", "pH","Sulphates", "Alcohol", "Quality")

head(wine_data)

##   FixedAcidity VolatileAcidity CitricAcid ResidualSugar chlorides
## 1          7.4            0.70       0.00           1.9     0.076
## 2          7.8            0.88       0.00           2.6     0.098
## 3          7.8            0.76       0.04           2.3     0.092
## 4         11.2            0.28       0.56           1.9     0.075
## 5          7.4            0.70       0.00           1.9     0.076
## 6          7.4            0.66       0.00           1.8     0.075
##   FreeSulfurDioxide TotalSulfurDioxide Density   pH Sulphates Alcohol
Quality
## 1                11                 34  0.9978 3.51      0.56     9.4
5
## 2                25                 67  0.9968 3.20      0.68     9.8
5
## 3                15                 54  0.9970 3.26      0.65     9.8
5
## 4                17                 60  0.9980 3.16      0.58     9.8
6
## 5                11                 34  0.9978 3.51      0.56     9.4
5
```

```
## 6                13                   40   0.9978 3.51      0.56      9.4
5
```

Check the missing values

```
sum(is.na(wine_data) == TRUE)

## [1] 0
```

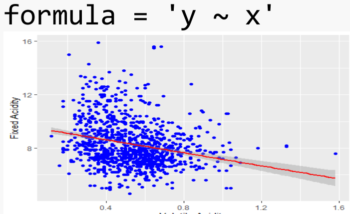
Summary output of the variables.

```
summary(wine_data)

##   FixedAcidity    VolatileAcidity    CitricAcid     ResidualSugar
##  Min.   : 4.600   Min.   :0.1200   Min.   :0.0000   Min.   : 0.900
##  1st Qu.: 7.100   1st Qu.:0.3925   1st Qu.:0.0900   1st Qu.: 1.900
##  Median : 7.900   Median :0.5200   Median :0.2500   Median : 2.200
##  Mean   : 8.311   Mean   :0.5313   Mean   :0.2684   Mean   : 2.532
##  3rd Qu.: 9.100   3rd Qu.:0.6400   3rd Qu.:0.4200   3rd Qu.: 2.600
##  Max.   :15.900   Max.   :1.5800   Max.   :1.0000   Max.   :15.500
##    chlorides         FreeSulfurDioxide TotalSulfurDioxide    Density
##  Min.   :0.01200   Min.   : 1.00     Min.   :  6.00     Min.   :0.9901
##  1st Qu.:0.07000   1st Qu.: 7.00     1st Qu.: 21.00     1st Qu.:0.9956
##  Median :0.07900   Median :13.00     Median : 37.00     Median :0.9967
##  Mean   :0.08693   Mean   :15.62     Mean   : 45.91     Mean   :0.9967
##  3rd Qu.:0.09000   3rd Qu.:21.00     3rd Qu.: 61.00     3rd Qu.:0.9978
##  Max.   :0.61100   Max.   :68.00     Max.   :289.00     Max.   :1.0037
##        pH          Sulphates         Alcohol          Quality
##  Min.   :2.740   Min.   :0.3300   Min.   : 8.40   Min.   :3.000
##  1st Qu.:3.205   1st Qu.:0.5500   1st Qu.: 9.50   1st Qu.:5.000
##  Median :3.310   Median :0.6200   Median :10.20   Median :6.000
##  Mean   :3.311   Mean   :0.6577   Mean   :10.44   Mean   :5.657
##  3rd Qu.:3.400   3rd Qu.:0.7300   3rd Qu.:11.10   3rd Qu.:6.000
##  Max.   :4.010   Max.   :2.0000   Max.   :14.90   Max.   :8.000
```

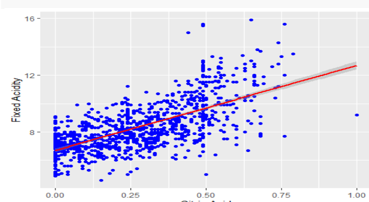
Standard deviations of each variable

```
st_devs <- c(sd(wine_data$FixedAcidity), sd(wine_data$VolatileAcidity),
sd(wine_data$CitricAcid),sd(wine_data$ResidualSugar),
sd(wine_data$chlorides), sd(wine_data$FreeSulfurDioxide),
sd(wine_data$TotalSulfurDioxide), sd(wine_data$Density), sd(wine_data$pH),
sd(wine_data$Sulphates), sd(wine_data$Alcohol), sd(wine_data$Quality))

st_devs

##  [1]  1.747595017  0.179633193  0.196685852  1.355917467  0.047267338
##  [6] 10.250486123 32.782130307  0.001925067  0.156664060  0.170398714
## [11]  1.082195610  0.805824248
```

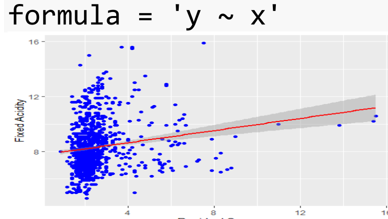
```
library(ggplot2)
scplot1 <- ggplot(data =
wine_data, aes(x =
VolatileAcidity, y =
FixedAcidity)) +
  geom_point(color =
"blue") +
  xlab("Volatile
Acidity") +
  ylab("Fixed Acidity") +
  geom_smooth(method =
"lm", color = "red")

scplot1
## `geom_smooth()` using
formula = 'y ~ x'
```




```
scplot2 <- ggplot(data =
wine_data, aes(x =
CitricAcid, y =
FixedAcidity)) +
  geom_point(color =
"blue") +
  xlab("Citric Acid") +
  ylab("Fixed Acidity")
+
  geom_smooth(method =
"lm", color = "red")

scplot2
## `geom_smooth()` using
formula = 'y ~ x'
```



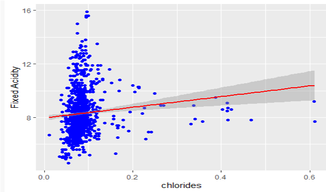

```
scplot3 <- ggplot(data =
wine_data, aes(x =
ResidualSugar, y =
FixedAcidity)) +
  geom_point(color =
"blue") +
  xlab("Residual Sugar")
+
  ylab("Fixed Acidity")
+
  geom_smooth(method =
"lm", color = "red")

scplot3
## `geom_smooth()` using
formula = 'y ~ x'
```




```
scplot4 <- ggplot(data =
wine_data, aes(x =
chlorides, y =
FixedAcidity)) +
  geom_point(color =
"blue") +
  xlab("chlorides") +
  ylab("Fixed Acidity") +
  geom_smooth(method =
"lm", color = "red")

scplot4
## `geom_smooth()` using
formula = 'y ~ x'
```

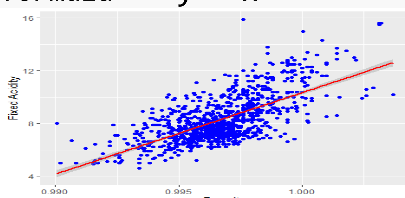
```
scplot5 <- ggplot(data =
wine_data, aes(x =
FreeSulfurDioxide, y =
FixedAcidity)) +
  geom_point(color =
"blue") +
  xlab("Free Sulfur
Dioxide") +
  ylab("Fixed Acidity")
+
  geom_smooth(method =
"lm", color = "red")

scplot5
## `geom_smooth()` using
```

```
scplot6 <- ggplot(data =
wine_data, aes(x =
TotalSulfurDioxide, y =
FixedAcidity)) +
  geom_point(color =
"blue") +
  xlab("Total Sulfur
Dioxide") +
  ylab("Fixed Acidity")
+
  geom_smooth(method =
"lm", color = "red")

scplot6
## `geom_smooth()` using
```



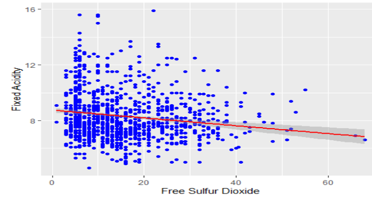
scplot7 code:

```
scplot7 <- ggplot(data =
wine_data, aes(x =
Density, y =
FixedAcidity)) +
  geom_point(color =
"blue") +
  xlab("Density") +
  ylab("Fixed Acidity") +
  geom_smooth(method =
"lm", color = "red")

scplot7
## `geom_smooth()` using
formula = 'y ~ x'
```

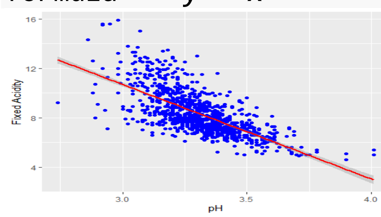



```
scplot10 <- ggplot(data =
wine_data, aes(x =
Alcohol, y =
FixedAcidity)) +
  geom_point(color =
"blue") +
  xlab("Alcohol") +
  ylab("Fixed Acidity") +
  geom_smooth(method =
"lm", color = "red")

scplot10
## `geom_smooth()` using
formula = 'y ~ x'
```

```
formula = 'y ~ x'
```

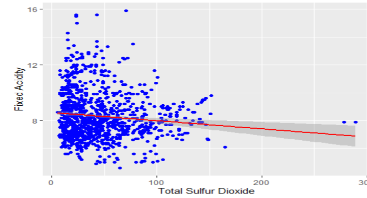



```
scplot8 <- ggplot(data =
wine_data, aes(x = pH, y
= FixedAcidity)) +
  geom_point(color =
"blue") +
  xlab("pH") +
  ylab("Fixed Acidity")
+
  geom_smooth(method =
"lm", color = "red")

scplot8
## `geom_smooth()` using
formula = 'y ~ x'
```

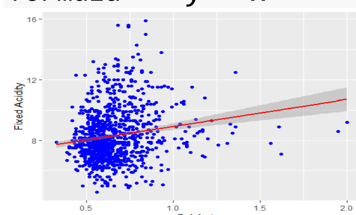


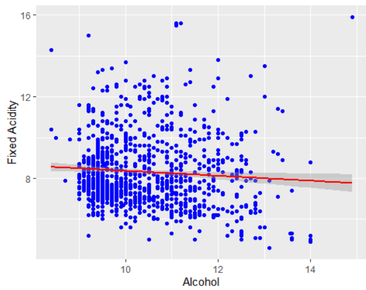

```
scplot11 <- ggplot(data
= wine_data, aes(x =
Quality, y =
FixedAcidity)) +
  geom_point(color =
"blue") +
  xlab("Quality") +
  ylab("Fixed Acidity")
+
  geom_smooth(method =
"lm", color = "red")

scplot11
## `geom_smooth()` using
formula = 'y ~ x'
```

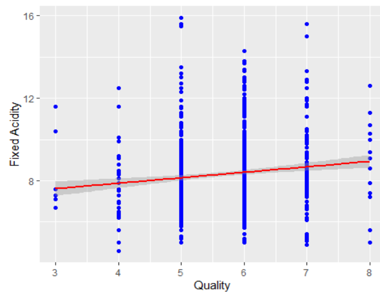
```
formula = 'y ~ x'
```




```
scplot9 <- ggplot(data =
wine_data, aes(x =
Sulphates, y =
FixedAcidity)) +
  geom_point(color =
"blue") +
  xlab("Sulphates") +
  ylab("Fixed Acidity")
+
  geom_smooth(method =
"lm", color = "red")

scplot9
## `geom_smooth()` using
formula = 'y ~ x'
```



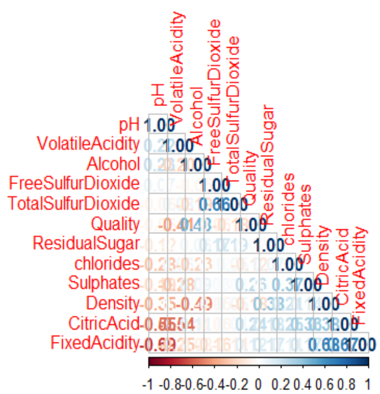





```
library(corrplot)

## Warning: package 'corrplot' was built under R version 4.3.3

## corrplot 0.92 loaded
```

The correlation coefficients among each and every pair of variables.

```
corrplot(corr = cor(wine_data), method = "number" , order = 'FPC', type =
'lower')
```




```
cor(wine_data)

##                      FixedAcidity VolatileAcidity  CitricAcid ResidualSugar
## FixedAcidity           1.00000000    -0.250728322  0.67315725   0.171830535
## VolatileAcidity       -0.25072832     1.000000000 -0.54418694  -0.005751097
## CitricAcid             0.67315725    -0.544186937  1.00000000   0.175814854
## ResidualSugar          0.17183054    -0.005751097  0.17581485   1.000000000
## chlorides              0.10788857     0.056336259  0.24531249   0.070863112
## FreeSulfurDioxide     -0.16483079    -0.001962479 -0.05758910   0.165338797
## TotalSulfurDioxide    -0.11062837     0.077747722  0.03687111   0.190790035
## Density                0.68150088     0.016511520  0.37524326   0.380146952
## pH                    -0.68516260     0.221491518 -0.54633914  -0.116958936
## Sulphates              0.17459183    -0.276078597  0.33123176   0.017474504
## Alcohol               -0.07505485    -0.203909273  0.10625034   0.058420606
```

```
## Quality               0.12197010    -0.407393513  0.24082084   0.022001931
##                       chlorides FreeSulfurDioxide TotalSulfurDioxide
Density
## FixedAcidity         0.10788857      -0.164830793        -0.11062837
0.68150088
## VolatileAcidity      0.05633626      -0.001962479         0.07774772
0.01651152
## CitricAcid           0.24531249      -0.057589104         0.03687111
0.37524326
## ResidualSugar        0.07086311       0.165338797         0.19079003
0.38014695
## chlorides            1.00000000       0.015280458         0.04816316
0.20890071
## FreeSulfurDioxide    0.01528046       1.000000000         0.66109287 -
0.05415032
## TotalSulfurDioxide   0.04816316       0.661092872         1.00000000
0.05017483
## Density              0.20890071      -0.054150318         0.05017483
1.00000000
## pH                  -0.27775907       0.072803706        -0.05912572 -
0.35277462
## Sulphates            0.37478389       0.034445122         0.02689368
0.14313929
## Alcohol             -0.22991709      -0.047094832        -0.18816480 -
0.49472690
## Quality             -0.12408453      -0.063259641        -0.18333915 -
0.17520792
##                             pH   Sulphates     Alcohol     Quality
## FixedAcidity        -0.68516260  0.17459183 -0.07505485  0.12197010
## VolatileAcidity      0.22149152 -0.27607860 -0.20390927 -0.40739351
## CitricAcid          -0.54633914  0.33123176  0.10625034  0.24082084
## ResidualSugar       -0.11695894  0.01747450  0.05842061  0.02200193
## chlorides           -0.27775907  0.37478389 -0.22991709 -0.12408453
## FreeSulfurDioxide    0.07280371  0.03444512 -0.04709483 -0.06325964
## TotalSulfurDioxide -0.05912572  0.02689368 -0.18816480 -0.18333915
## Density             -0.35277462  0.14313929 -0.49472690 -0.17520792
## pH                   1.00000000 -0.18549903  0.22532220 -0.05245303
## Sulphates           -0.18549903  1.00000000  0.09442113  0.25771026
## Alcohol              0.22532220  0.09442113  1.00000000  0.48486621
## Quality             -0.05245303  0.25771026  0.48486621  1.00000000
```

```
lmTemp1 <- lm(formula = FixedAcidity ~ . , data = wine_data)
summary(lmTemp1)
```

```
##
## Call:
## lm(formula = FixedAcidity ~ ., data = wine_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -2.6456 -0.3589 -0.0060  0.3479  3.6466
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -6.316e+02  1.550e+01 -40.749  < 2e-16 ***
## VolatileAcidity      1.972e-01  1.420e-01   1.389  0.16523
## CitricAcid           1.815e+00  1.612e-01  11.256  < 2e-16 ***
## ResidualSugar       -2.485e-01  1.657e-02 -14.996  < 2e-16 ***
## chlorides           -3.718e+00  4.787e-01  -7.768 1.78e-14 ***
## FreeSulfurDioxide    6.643e-03  2.502e-03   2.655  0.00804 **
## TotalSulfurDioxide -5.308e-03  8.129e-04  -6.529 9.95e-11 ***
## Density              6.546e+02  1.545e+01  42.362  < 2e-16 ***
## pH                  -5.276e+00  1.533e-01 -34.426  < 2e-16 ***
## Sulphates           -7.293e-01  1.319e-01  -5.529 3.99e-08 ***
## Alcohol              5.550e-01  2.717e-02  20.429  < 2e-16 ***
## Quality              2.218e-02  2.921e-02   0.759  0.44777
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6294 on 1131 degrees of freedom
## Multiple R-squared:  0.8715, Adjusted R-squared:  0.8703
## F-statistic: 697.6 on 11 and 1131 DF,  p-value: < 2.2e-16
```

```r
library(BAS)
```

```
## Warning: package 'BAS' was built under R version 4.3.3
```

Bayesian Simple Linear Regression

Fit a simple linear regression model of Density versus Fixed Acidity.

```r
apTemp.lm1 <- lm(formula = FixedAcidity ~ Density, data = wine_data)
summary(lmTemp1)
```

```
##
## Call:
## lm(formula = FixedAcidity ~ ., data = wine_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6456 -0.3589 -0.0060  0.3479  3.6466
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -6.316e+02  1.550e+01 -40.749  < 2e-16 ***
## VolatileAcidity      1.972e-01  1.420e-01   1.389  0.16523
## CitricAcid           1.815e+00  1.612e-01  11.256  < 2e-16 ***
## ResidualSugar       -2.485e-01  1.657e-02 -14.996  < 2e-16 ***
## chlorides           -3.718e+00  4.787e-01  -7.768 1.78e-14 ***
```

```
## FreeSulfurDioxide   6.643e-03  2.502e-03   2.655  0.00804 **
## TotalSulfurDioxide -5.308e-03  8.129e-04  -6.529 9.95e-11 ***
## Density             6.546e+02  1.545e+01  42.362  < 2e-16 ***
## pH                 -5.276e+00  1.533e-01 -34.426  < 2e-16 ***
## Sulphates          -7.293e-01  1.319e-01  -5.529 3.99e-08 ***
## Alcohol             5.550e-01  2.717e-02  20.429  < 2e-16 ***
## Quality             2.218e-02  2.921e-02   0.759  0.44777
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6294 on 1131 degrees of freedom
## Multiple R-squared:  0.8715, Adjusted R-squared:  0.8703
## F-statistic: 697.6 on 11 and 1131 DF,  p-value: < 2.2e-16
```

Obtain residuals and n.

```
resid <- residuals(apTemp.lm1)
n <- length(resid)
```

Calculate MSE

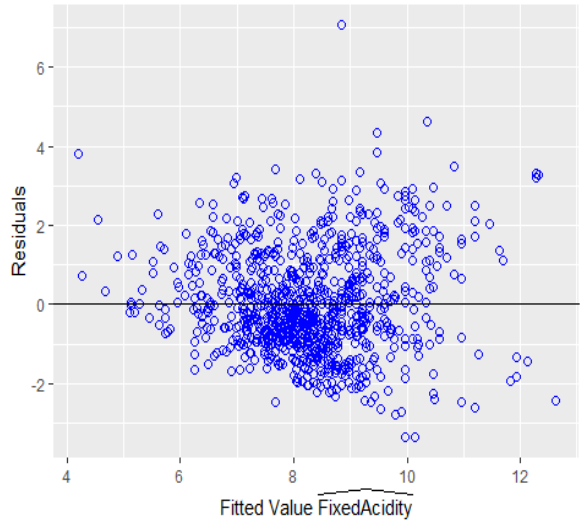
```
MSE <- 1/ (n-2) * sum((resid ^ 2))
MSE

## [1] 1.637071
```

Combine residuals and fitted values into a data frame.

```
result <- data.frame(fitted_values = fitted.values(apTemp.lm1) , residuals =
residuals(apTemp.lm1))
```

Load library and plot residuals versus fitted values.

```
library(ggplot2)
ggplot(data = result, aes(x = fitted_values, y = residuals)) +
  geom_point(color = "blue", pch = 1, size = 2) +
  geom_abline(intercept = 0, slope = 0) +
  xlab(expression(paste("Fitted Value ", widehat(FixedAcidity)))) +
  ylab("Residuals")
```



The observation with the largest fitted value.

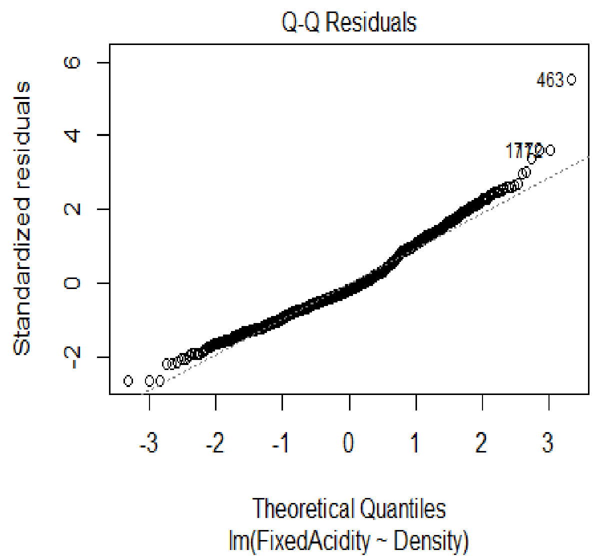
```
which.max(as.vector(fitted.values(apTemp.lm1)))

## [1] 1023
```

Shows this observation has the largest Density

```
which.max(wine_data$Density)

## [1] 1023
```

Normal probability plot of the residuals.

```r
plot(apTemp.lm1, which = 2)
```



Credible Intervals for Slope Beta and y-Intercept alpha.

```r
output <- summary(apTemp.lm1)$coef[, 1:2]
output

##              Estimate Std. Error
## (Intercept) -608.3393   19.60348
## Density      618.6733   19.66775

out <- cbind(output, confint(apTemp.lm1))
colnames(out) <- c("Posterior Mean", "Posterior Std", "2.5", "97.5")
round(out, 3)

##             Posterior Mean Posterior Std      2.5     97.5
## (Intercept)       -608.339        19.603 -646.802 -569.876
## Density            618.673        19.668  580.084  657.262
```

Construct current prediction.

```r
alpha <- apTemp.lm1$coefficients[1]
beta <- apTemp.lm1$coefficients[2]
new_x <- seq(min(wine_data$Density) , max(wine_data$Density) , length.out =
100)
y_hat <- alpha + beta*new_x
```

Get lower and upper bounds for mean.

```
ymean <- data.frame(predict(apTemp.lm1 , newdata = data.frame(Density =
new_x) , interval = "confidence" , level = 0.95))
```

Get lower and upper bounds for prediction.

```
ypred <- data.frame(predict(apTemp.lm1 , newdata = data.frame(Density =
new_x) , interval = "prediction" , level = 0.95))

output <- data.frame(x = new_x , y_hat = y_hat , ymean_lwr = ymean$lwr ,
ymean_upr = ymean$upr , ypred_lwr = ypred$lwr , ypred_upr = ypred$upr)
```

Extract potential outlier data point.

```
outlier <- data.frame(x = wine_data$Density[3241] , y =
wine_data$Density[3241])
```

Scatter plot of original.

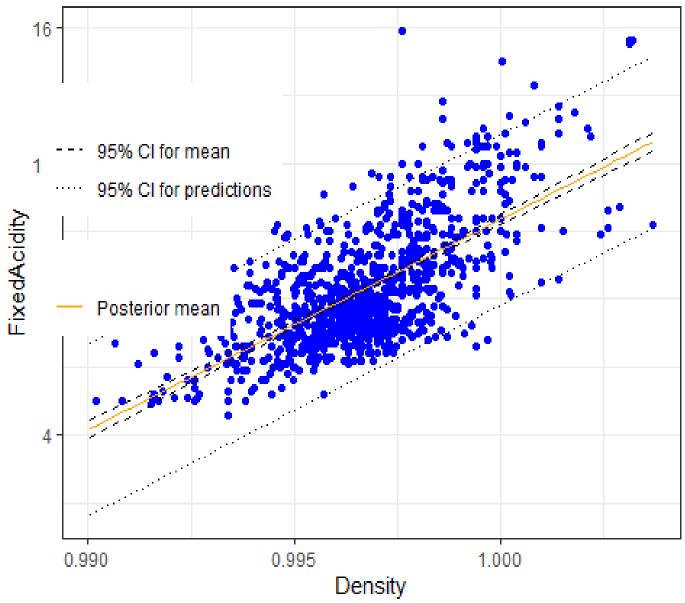
```
library(ggplot2)
plot1 <- ggplot(data = wine_data , aes(x = Density , y = FixedAcidity)) +
geom_point(color = "blue")
```

Add bounds of mean and prediction.

```
plot2 <- plot1 + geom_line(data = output , aes(x = new_x , y = y_hat , color=
"first") , lty = 1) + geom_line(data = output , aes(x = new_x , y = ymean_lwr
, lty = "second")) +  geom_line(data = output , aes(x = new_x , y = ymean_upr
, lty = "second")) + geom_line(data = output , aes(x = new_x , y = ypred_upr
, lty="third")) + geom_line(data = output , aes(x = new_x , y = ypred_lwr ,
lty = "third")) + scale_colour_manual(values = c("orange") , labels =
"Posterior mean" , name = "") + scale_linetype_manual(values = c(2,3) ,
labels = c("95% CI for mean" , "95% CI for predictions") , name = "") +
theme_bw() + theme(legend.position = c(1,0) , legend.justification = c(3.-
0.35,-0.8))
```

Identify potential outlier.

```
plot2 + geom_point(data = outlier , aes(x = x , y = y) , color = "orange" ,
pch = 1 , cex = 6)

## Warning: Removed 1 rows containing missing values (`geom_point()`).
```



Bayesian Multiple Linear Regression Use BAS.LM to run regression model.

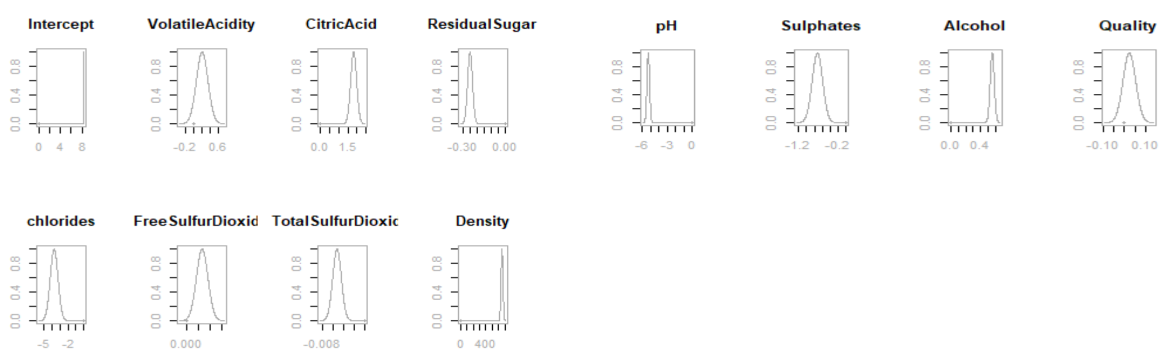
```
library(BAS)
wine.bas = bas.lm(FixedAcidity ~ . , data =  wine_data , prior = "BIC" ,
modelprior = Bernoulli(1) ,  include.always = ~ . , n.models = 1)
wine.bas

##
## Call:
## bas.lm(formula = FixedAcidity ~ ., data = wine_data, n.models = 1,
##     prior = "BIC", modelprior = Bernoulli(1), include.always = ~.)
##
##
##  Marginal Posterior Inclusion Probabilities:
##          Intercept     VolatileAcidity          CitricAcid
ResidualSugar
##                  1                   1                   1
1
##          chlorides   FreeSulfurDioxide  TotalSulfurDioxide
Density
##                  1                   1                   1
1
##                 pH           Sulphates             Alcohol
Quality
##                  1                   1                   1
1
```

Posterior Means and Posterior Standard Deviations.

```
wine.coef = coef(wine.bas)
wine.coef

##
##  Marginal Posterior Summaries of Coefficients:
##
##  Using  BMA
##
##  Based on the top  1 models
##                      post mean   post SD     post p(B != 0)
## Intercept             8.311e+00   1.862e-02   1.000e+00
## VolatileAcidity       1.972e-01   1.420e-01   1.000e+00
## CitricAcid            1.815e+00   1.612e-01   1.000e+00
## ResidualSugar        -2.485e-01   1.657e-02   1.000e+00
## chlorides            -3.718e+00   4.787e-01   1.000e+00
## FreeSulfurDioxide     6.643e-03   2.502e-03   1.000e+00
## TotalSulfurDioxide   -5.308e-03   8.129e-04   1.000e+00
## Density               6.546e+02   1.545e+01   1.000e+00
## pH                   -5.276e+00   1.533e-01   1.000e+00
## Sulphates            -7.293e-01   1.319e-01   1.000e+00
## Alcohol               5.550e-01   2.717e-02   1.000e+00
## Quality               2.218e-02   2.921e-02   1.000e+00
```

visualization of the coefficients.

```
par(mfrow = c(2, 4) , col.lab = "darkgrey" , col.axis = "darkgrey" , col =
"darkgrey")
plot(wine.coef , ask = F)
```




```
out <- confint(wine.coef)[, 1:2]
```

Extract the upper and lower bounds of the credible intervals.

```
names <- c("posterior mean", "posterior std", colnames(out))
out <- cbind(wine.coef$postmean , wine.coef$postsd , out)
colnames(out) <- names
round(out , 4)

##                    posterior mean posterior std     2.5%    97.5%
## Intercept                  8.3111        0.0186   8.2746   8.3476
## VolatileAcidity            0.1972        0.1420  -0.0815   0.4759
## CitricAcid                 1.8150        0.1612   1.4986   2.1314
## ResidualSugar             -0.2485        0.0166  -0.2810  -0.2159
## chlorides                 -3.7184        0.4787  -4.6576  -2.7792
## FreeSulfurDioxide          0.0066        0.0025   0.0017   0.0116
## TotalSulfurDioxide        -0.0053        0.0008  -0.0069  -0.0037
## Density                  654.5712       15.4517 624.2540 684.8885
## pH                        -5.2761        0.1533  -5.5768  -4.9754
## Sulphates                 -0.7293        0.1319  -0.9880  -0.4705
## Alcohol                    0.5550        0.0272   0.5017   0.6083
## Quality                    0.0222        0.0292  -0.0351   0.0795
```

Use BAS.LM to run regression model.

```
wine.bas2 <- bas.lm(FixedAcidity ~ . , data = wine_data , prior = "BIC" ,
modelprior = Bernoulli(1) ,
n.models = 1)
wine.bas2

##
## Call:
## bas.lm(formula = FixedAcidity ~ ., data = wine_data, n.models = 1,
##     prior = "BIC", modelprior = Bernoulli(1))
##
##
##  Marginal Posterior Inclusion Probabilities:
##          Intercept     VolatileAcidity          CitricAcid
ResidualSugar
##                NaN                   0                   0
0
##          chlorides   FreeSulfurDioxide  TotalSulfurDioxide
Density
##                  0                   0                   0
0
##                 pH           Sulphates             Alcohol
Quality
##                  0                   0                   0
0
```

Bayesian Model Selection

```
modelSelection <- step(lmTemp1, k = log(nrow(wine_data)))

## Start:  AIC=-985.96
## FixedAcidity ~ VolatileAcidity + CitricAcid + ResidualSugar +
##     chlorides + FreeSulfurDioxide + TotalSulfurDioxide + Density +
##     pH + Sulphates + Alcohol + Quality
##
##                      Df Sum of Sq     RSS     AIC
## - Quality             1      0.23  448.27 -992.42
## - VolatileAcidity     1      0.76  448.80 -991.05
## <none>                             448.04 -985.96
## - FreeSulfurDioxide   1      2.79  450.83 -985.89
## - Sulphates           1     12.11  460.15 -962.51
## - TotalSulfurDioxide  1     16.89  464.93 -950.70
## - chlorides           1     23.90  471.94 -933.59
## - CitricAcid          1     50.19  498.23 -871.64
## - ResidualSugar       1     89.09  537.13 -785.71
## - Alcohol             1    165.33  613.37 -633.99
## - pH                  1    469.49  917.53 -173.69
## - Density             1    710.91 1158.95   93.29
##
## Step:  AIC=-992.42
## FixedAcidity ~ VolatileAcidity + CitricAcid + ResidualSugar +
##     chlorides + FreeSulfurDioxide + TotalSulfurDioxide + Density +
##     pH + Sulphates + Alcohol
##
##                      Df Sum of Sq     RSS     AIC
## - VolatileAcidity     1      0.62  448.88 -997.89
## <none>                             448.27 -992.42
## - FreeSulfurDioxide   1      2.84  451.11 -992.23
## - Sulphates           1     11.92  460.19 -969.47
## - TotalSulfurDioxide  1     17.49  465.76 -955.70
## - chlorides           1     24.72  472.98 -938.11
## - CitricAcid          1     50.09  498.36 -878.38
## - ResidualSugar       1     88.98  537.25 -792.50
## - Alcohol             1    187.88  636.15 -599.36
## - pH                  1    476.44  924.71 -171.83
## - Density             1    710.81 1159.08   86.38
##
## Step:  AIC=-997.89
## FixedAcidity ~ CitricAcid + ResidualSugar + chlorides + FreeSulfurDioxide
+
##     TotalSulfurDioxide + Density + pH + Sulphates + Alcohol
##
##                      Df Sum of Sq     RSS     AIC
## - FreeSulfurDioxide   1      2.53  451.42 -998.50
```

```
## <none>                              448.88 -997.89
## - Sulphates          1     13.97   462.86 -969.90
## - TotalSulfurDioxide  1     16.88   465.76 -962.74
## - chlorides           1     24.55   473.43 -944.06
## - CitricAcid          1     63.04   511.93 -854.72
## - ResidualSugar       1     89.98   538.86 -796.11
## - Alcohol             1    194.12   643.00 -594.15
## - pH                  1    478.02   926.90 -176.16
## - Density             1    775.71 1224.59  142.18
##
## Step:  AIC=-998.5
## FixedAcidity ~ CitricAcid + ResidualSugar + chlorides + TotalSulfurDioxide
+
##     Density + pH + Sulphates + Alcohol
##
##                      Df Sum of Sq     RSS     AIC
## <none>                              451.42 -998.50
## - Sulphates           1     13.34   464.75 -972.26
## - TotalSulfurDioxide  1     16.22   467.64 -965.18
## - chlorides           1     24.31   475.73 -945.58
## - CitricAcid          1     61.99   513.40 -858.47
## - ResidualSugar       1     87.97   539.39 -802.04
## - Alcohol             1    195.17   646.59 -594.83
## - pH                  1    475.52   926.93 -183.16
## - Density             1    773.19 1224.61  135.16
```

```
basModel1 <- bas.lm(formula = FixedAcidity ~ . , data = wine_data ,
prior = "BIC" , modelprior = uniform())
basCoeff <- coef(basModel1)
basCoeff
```

```
##
##  Marginal Posterior Summaries of Coefficients:
##
##  Using  BMA
##
##  Based on the top  2048 models
##                     post mean   post SD     post p(B != 0)
## Intercept            8.311e+00   1.864e-02   1.000e+00
## VolatileAcidity      7.468e-03   4.521e-02   5.028e-02
## CitricAcid           1.698e+00   1.385e-01   1.000e+00
## ResidualSugar       -2.471e-01   1.664e-02   1.000e+00
## chlorides           -3.583e+00   4.590e-01   1.000e+00
## FreeSulfurDioxide    2.696e-03   3.509e-03   4.289e-01
## TotalSulfurDioxide  -4.442e-03   9.463e-04   1.000e+00
## Density              6.580e+02   1.501e+01   1.000e+00
## pH                  -5.275e+00   1.532e-01   1.000e+00
## Sulphates           -7.362e-01   1.268e-01   1.000e+00
## Alcohol              5.662e-01   2.567e-02   1.000e+00
## Quality              5.318e-04   5.968e-03   3.352e-02
```

Find the index of the model with the largest logmarg.

```
best <- which.max(basModel1$logmarg)
```

Retreat the index of variables in the best model, with 0 as the index of the intercept.

```
bestModel <- basModel1$which[[best]]
bestModel

## [1]  0  2  3  4  6  7  8  9 10
```

Create an indicator vector indicating which variables are used in the best model.

```
bestGamma <- rep(0,basModel1$n.vars)

bestGamma[bestModel + 1] = 1
bestGamma

##  [1] 1 0 1 1 1 0 1 1 1 1 1 0
```

Coefficient Estimates Under Reference Prior for Best BIC Model Fit the best BIC model by imposing which variables to be used using the indicators.

```
wine.bestBIC <- bas.lm(FixedAcidity ~ . , data = wine_data, prior = "BIC" ,
n.models = 1 , bestmodel = bestGamma , modelprior = uniform())

wine.coeff <- coef(wine.bestBIC)

out <- confint(wine.coeff)[,1:2]
```

Combine results and construct summary table.

```
coef.BIC <- cbind(wine.coeff$postmean , wine.coeff$postsd , out)
names <- c("post mean" , "post sd" , colnames(out))
colnames(coef.BIC) <- names
coef.BIC

##                        post mean      post sd         2.5%        97.5%
## Intercept            8.311111111 1.866206e-02   8.274495064   8.34772716
## VolatileAcidity      0.000000000 0.000000e+00   0.000000000   0.00000000
## CitricAcid           1.686826597 1.351769e-01   1.421601577   1.95205162
## ResidualSugar       -0.245499225 1.651460e-02  -0.277901837  -0.21309661
## chlorides           -3.570017428 4.568016e-01  -4.466288722  -2.67374613
## FreeSulfurDioxide    0.000000000 0.000000e+00   0.000000000   0.00000000
## TotalSulfurDioxide  -0.003884293 6.084643e-04  -0.005078135  -0.00269045
## Density            657.246670764 1.491309e+01 627.986325974 686.50701555
## pH                  -5.262546266 1.522634e-01  -5.561295978  -4.96379655
## Sulphates           -0.729371714 1.260183e-01  -0.976626958  -0.48211647
```

```
## Alcohol                0.567111234 2.561184e-02   0.516859324   0.61736314
## Quality                0.000000000 0.000000e+00   0.000000000   0.00000000
```

Calculating Posterior Probability

```
wine_bas <- bas.lm(FixedAcidity ~ VolatileAcidity + CitricAcid +
ResidualSugar + chlorides + FreeSulfurDioxide + TotalSulfurDioxide + Density
+ pH + Sulphates + Alcohol + Quality , data = wine_data , prior = "BIC" ,
modelprior = uniform())

round(summary(wine_bas) , 3)

##                      P(B != 0 | Y)   model 1   model 2   model 3   model 4
## Intercept                    1.000     1.000     1.000     1.000     1.000
## VolatileAcidity              0.050     0.000     0.000     1.000     1.000
## CitricAcid                   1.000     1.000     1.000     1.000     1.000
## ResidualSugar                1.000     1.000     1.000     1.000     1.000
## chlorides                    1.000     1.000     1.000     1.000     1.000
## FreeSulfurDioxide            0.429     0.000     1.000     1.000     0.000
## TotalSulfurDioxide           1.000     1.000     1.000     1.000     1.000
## Density                      1.000     1.000     1.000     1.000     1.000
## pH                           1.000     1.000     1.000     1.000     1.000
## Sulphates                    1.000     1.000     1.000     1.000     1.000
## Alcohol                      1.000     1.000     1.000     1.000     1.000
## Quality                      0.034     0.000     0.000     0.000     0.000
## BF                              NA     1.000     0.738     0.048     0.044
## PostProbs                       NA     0.528     0.390     0.025     0.023
## R2                              NA     0.871     0.871     0.872     0.871
## dim                             NA     9.000    10.000    11.000    10.000
## logmarg                         NA -3524.919 -3525.223 -3527.959 -3528.051
##                        model 5
## Intercept                1.000
## VolatileAcidity          0.000
## CitricAcid               1.000
## ResidualSugar            1.000
## chlorides                1.000
## FreeSulfurDioxide        0.000
## TotalSulfurDioxide       1.000
## Density                  1.000
## pH                       1.000
## Sulphates                1.000
## Alcohol                  1.000
## Quality                  1.000
## BF                       0.036
## PostProbs                0.019
## R2                       0.871
## dim                     10.000
## logmarg              -3528.256
```

The marginal posterior inclusion probability (pip)

```
print(wine_bas)

##
## Call:
## bas.lm(formula = FixedAcidity ~ VolatileAcidity + CitricAcid +
##     ResidualSugar + chlorides + FreeSulfurDioxide + TotalSulfurDioxide +
##     Density + pH + Sulphates + Alcohol + Quality, data = wine_data,
##     prior = "BIC", modelprior = uniform())
##
##
##  Marginal Posterior Inclusion Probabilities:
##          Intercept     VolatileAcidity          CitricAcid
ResidualSugar
##            1.00000             0.05028             1.00000
1.00000
##          chlorides   FreeSulfurDioxide  TotalSulfurDioxide
Density
##            1.00000             0.42885             1.00000
1.00000
##                 pH           Sulphates             Alcohol
Quality
##            1.00000             1.00000             1.00000
0.03352
```