Question 3 - Final Project

Vy Nguyen

2025-03-02

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(ggplot2)
library(car)
```

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

```
## Loading required package: carData
##
## Attaching package: 'car'
##
## The following object is masked from 'package:dplyr':
##
##     recode
##
## The following object is masked from 'package:purrr':
##
##     some
```

```r
library(corrplot)
```

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

```
## corrplot 0.95 loaded
```

```r
library(boot)
```

```
##
## Attaching package: 'boot'
##
## The following object is masked from 'package:car':
##
##     logit
```

```
red_wines <- read.csv2("Data/winequality-red.csv")
white_wines <- read.csv2("Data/winequality-white.csv")
```

First Look at the Data:

```
head(red_wines)
```

```
##   fixed.acidity volatile.acidity citric.acid residual.sugar chlorides
## 1          7.4              0.7            0            1.9     0.076
## 2          7.8             0.88            0            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              0.7            0            1.9     0.076
## 6          7.4             0.66            0            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  3.2      0.68     9.8
## 3                  15                   54   0.997 3.26      0.65     9.8
## 4                  17                   60   0.998 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
## 1       5
## 2       5
## 3       5
## 4       6
## 5       5
## 6       5
```

```
head(white_wines)
```

```
##   fixed.acidity volatile.acidity citric.acid residual.sugar chlorides
## 1             7             0.27        0.36           20.7     0.045
## 2           6.3              0.3        0.34            1.6     0.049
## 3           8.1             0.28         0.4            6.9      0.05
## 4           7.2             0.23        0.32            8.5     0.058
## 5           7.2             0.23        0.32            8.5     0.058
## 6           8.1             0.28         0.4            6.9      0.05
##   free.sulfur.dioxide total.sulfur.dioxide density   pH sulphates alcohol
## 1                  45                  170   1.001    3      0.45     8.8
## 2                  14                  132   0.994  3.3      0.49     9.5
## 3                  30                   97  0.9951 3.26      0.44    10.1
## 4                  47                  186  0.9956 3.19       0.4     9.9
## 5                  47                  186  0.9956 3.19       0.4     9.9
## 6                  30                   97  0.9951 3.26      0.44    10.1
```

```
##   quality
## 1       6
## 2       6
## 3       6
## 4       6
## 5       6
## 6       6
```

Casting Data into Appropriate Types:

```r
red_wines$quality <- as.numeric(red_wines$quality)
red_wines$fixed.acidity <- as.numeric(red_wines$fixed.acidity)
red_wines$volatile.acidity <- as.numeric(red_wines$volatile.acidity)
red_wines$citric.acid <- as.numeric(red_wines$citric.acid)
red_wines$residual.sugar <- as.numeric(red_wines$residual.sugar)
red_wines$chlorides <- as.numeric(red_wines$chlorides)
red_wines$free.sulfur.dioxide <- as.numeric(red_wines$free.sulfur.dioxide)
red_wines$total.sulfur.dioxide <- as.numeric(red_wines$total.sulfur.dioxide)
red_wines$density <- as.numeric(red_wines$density)
red_wines$pH <- as.numeric(red_wines$pH)
red_wines$sulphates <- as.numeric(red_wines$sulphates)
red_wines$alcohol <- as.numeric(red_wines$alcohol)

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

Second Look at Data:

```r
head(red_wines)
```

```
##   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
## 1       5
## 2       5
## 3       5
## 4       6
## 5       5
## 6       5
```

```r
head(white_wines)
```

```
##   fixed.acidity volatile.acidity citric.acid residual.sugar chlorides
## 1           7.0             0.27        0.36           20.7     0.045
## 2           6.3             0.30        0.34            1.6     0.049
## 3           8.1             0.28        0.40            6.9     0.050
## 4           7.2             0.23        0.32            8.5     0.058
## 5           7.2             0.23        0.32            8.5     0.058
## 6           8.1             0.28        0.40            6.9     0.050
##   free.sulfur.dioxide total.sulfur.dioxide density   pH sulphates alcohol
## 1                  45                  170  1.0010 3.00      0.45     8.8
## 2                  14                  132  0.9940 3.30      0.49     9.5
## 3                  30                   97  0.9951 3.26      0.44    10.1
## 4                  47                  186  0.9956 3.19      0.40     9.9
## 5                  47                  186  0.9956 3.19      0.40     9.9
## 6                  30                   97  0.9951 3.26      0.44    10.1
##   quality
## 1       6
## 2       6
## 3       6
## 4       6
## 5       6
## 6       6
```

White wine

```r
white_wines <- white_wines[-c(2782, 4746), ]

white_wines <- white_wines[, !names(white_wines) %in% "density"]

cor_matrix <- cor(white_wines)
print(cor_matrix)
```

```
##                      fixed.acidity volatile.acidity  citric.acid residual.sugar
## fixed.acidity           1.00000000      -0.02441003  0.288768348     0.08747237
## volatile.acidity       -0.02441003       1.00000000 -0.153366001     0.04888207
## citric.acid             0.28876835      -0.15336600  1.000000000     0.09018479
## residual.sugar          0.08747237       0.04888207  0.090184786     1.00000000
## chlorides               0.02280905       0.06905578  0.113874037     0.08684580
## free.sulfur.dioxide    -0.04741662      -0.09693490  0.099277105     0.31677768
## total.sulfur.dioxide    0.09274987       0.08970170  0.122612124     0.40909913
```

```
## pH                   -0.42614748        -0.03389030 -0.164276406      -0.20001998
## sulphates            -0.01731599        -0.03832149  0.061794881      -0.03113152
## alcohol              -0.12114054         0.06670014 -0.076211140      -0.45948671
## quality              -0.11443554        -0.19617797 -0.009748182      -0.09988258
##                       chlorides free.sulfur.dioxide total.sulfur.dioxide
## fixed.acidity         0.02280905        -0.047416620          0.0927498703
## volatile.acidity      0.06905578        -0.096934903          0.0897017038
## citric.acid           0.11387404         0.099277105          0.1226121243
## residual.sugar        0.08684580         0.316777677          0.4090991299
## chlorides             1.00000000         0.104082022          0.1997637779
## free.sulfur.dioxide   0.10408202         1.000000000          0.6113573538
## total.sulfur.dioxide  0.19976378         0.611357354          1.0000000000
## pH                   -0.09086912        -0.005387324         -0.0002365303
## sulphates             0.01629569         0.057139153          0.1332279425
## alcohol              -0.36053861        -0.255724146         -0.4513585896
## quality              -0.21019638         0.018554518         -0.1711118710
##                             pH    sulphates     alcohol     quality
## fixed.acidity        -0.4261474831 -0.01731599 -0.12114054 -0.114435535
## volatile.acidity     -0.0338902990 -0.03832149  0.06670014 -0.196177970
## citric.acid          -0.1642764064  0.06179488 -0.07621114 -0.009748182
## residual.sugar       -0.2000199805 -0.03113152 -0.45948671 -0.099882576
## chlorides            -0.0908691192  0.01629569 -0.36053861 -0.210196384
## free.sulfur.dioxide  -0.0053873238  0.05713915 -0.25572415  0.018554518
## total.sulfur.dioxide -0.0002365303  0.13322794 -0.45135859 -0.171111871
## pH                    1.0000000000  0.15517328  0.12124109  0.100651951
## sulphates             0.1551732786  1.00000000 -0.01778527  0.054587740
## alcohol               0.1212410887 -0.01778527  1.00000000  0.436052648
## quality               0.1006519508  0.05458774  0.43605265  1.000000000
```

```
cor_matrix <- cor(white_wines[sapply(white_wines, is.numeric)])

# Plot the correlation matrix as a circle plot
corrplot(cor_matrix, method = "circle")
```
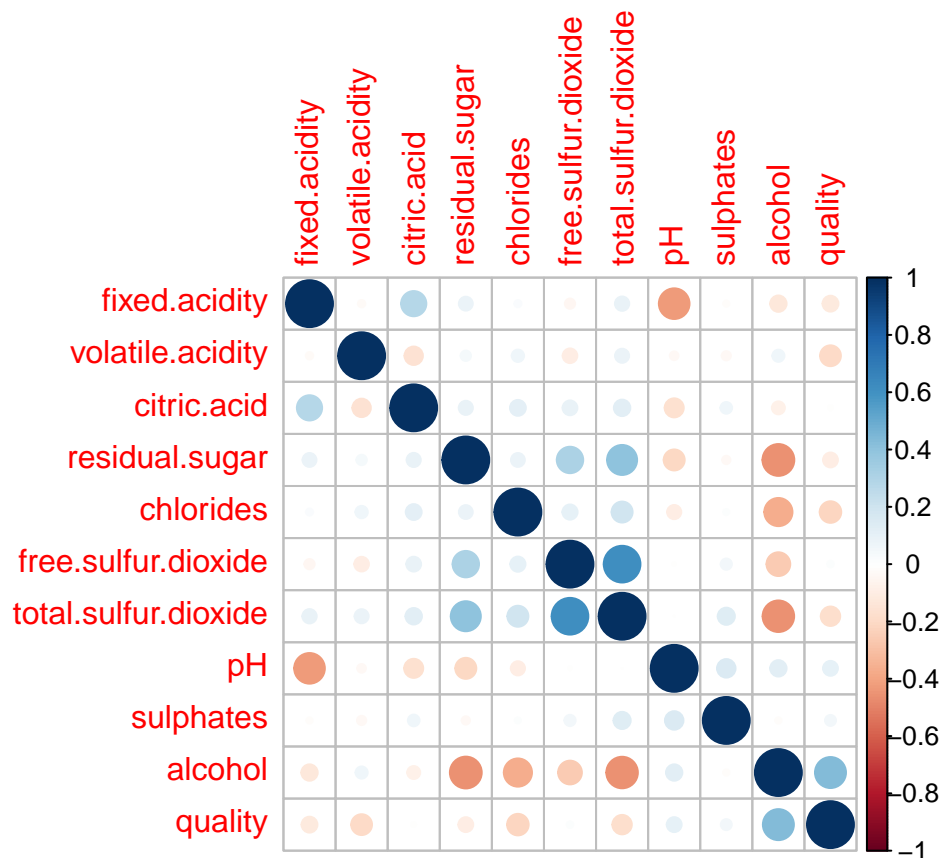
white wine Models

Models

```r
# CHeck IVF to decide which factor to use
model_full <- lm(quality ~ ., data = white_wines)
#vif(model)
```

```r
model_alcohol <- lm(quality ~ fixed.acidity + volatile.acidity + citric.acid + chlorides + pH + free.sul
vif(model_alcohol)
```

```
##        fixed.acidity     volatile.acidity          citric.acid
##             1.357263             1.125179             1.160494
##            chlorides                   pH   free.sulfur.dioxide
##             1.205178             1.335865             1.734020
##            sulphates total.sulfur.dioxide              alcohol
##             1.056650             2.133801             1.657651
##        residual.sugar
##             1.461398
```

```r
summary(model_alcohol)
```

```
##
```

```
## Call:
## lm(formula = quality ~ fixed.acidity + volatile.acidity + citric.acid +
##     chlorides + pH + free.sulfur.dioxide + sulphates + total.sulfur.dioxide +
##     alcohol + residual.sugar, data = white_wines)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4113 -0.4991 -0.0301  0.4656  3.1831
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           1.9863730  0.3493813   5.685 1.38e-08 ***
## fixed.acidity        -0.0484339  0.0148781  -3.255  0.00114 **
## volatile.acidity     -1.9394934  0.1139313 -17.023  < 2e-16 ***
## citric.acid          -0.0393944  0.0959621  -0.411  0.68144
## chlorides            -0.9603716  0.5414869  -1.774  0.07619 .
## pH                    0.1731651  0.0825096   2.099  0.03589 *
## free.sulfur.dioxide   0.0056530  0.0008541   6.619 4.02e-11 ***
## sulphates             0.4236637  0.0970944   4.363 1.31e-05 ***
## total.sulfur.dioxide -0.0009074  0.0003723  -2.437  0.01483 *
## alcohol               0.3639995  0.0112736  32.288  < 2e-16 ***
## residual.sugar        0.0250903  0.0026049   9.632  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.754 on 4885 degrees of freedom
## Multiple R-squared:  0.2754, Adjusted R-squared:  0.2739
## F-statistic: 185.7 on 10 and 4885 DF,  p-value: < 2.2e-16
```

```r
interaction_model <- lm(quality ~ fixed.acidity + volatile.acidity + citric.acid + chlorides + pH + fre
```

```r
AIC(interaction_model, model_alcohol)
```

```
##                   df      AIC
## interaction_model 13 11138.92
## model_alcohol     12 11142.18
```

```r
summary(interaction_model)
```

```
##
## Call:
## lm(formula = quality ~ fixed.acidity + volatile.acidity + citric.acid +
##     chlorides + pH + free.sulfur.dioxide + sulphates + total.sulfur.dioxide +
##     alcohol + residual.sugar + residual.sugar * alcohol, data = white_wines)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3836 -0.5026 -0.0302  0.4585  3.2011
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           1.6596071  0.3772324   4.399 1.11e-05 ***
```

```
## fixed.acidity           -0.0460924  0.0149067   -3.092 0.001999 **
## volatile.acidity        -1.9259270  0.1140357  -16.889  < 2e-16 ***
## citric.acid             -0.0515964  0.0960681   -0.537 0.591236
## chlorides               -0.8860237  0.5422236   -1.634 0.102312
## pH                       0.1853708  0.0826456    2.243 0.024944 *
## free.sulfur.dioxide      0.0055933  0.0008541    6.548 6.41e-11 ***
## sulphates                0.4068904  0.0973280    4.181 2.96e-05 ***
## total.sulfur.dioxide    -0.0009061  0.0003721   -2.435 0.014927 *
## alcohol                  0.3908219  0.0162497   24.051  < 2e-16 ***
## residual.sugar           0.0732171  0.0211675    3.459 0.000547 ***
## alcohol:residual.sugar  -0.0048282  0.0021074   -2.291 0.022005 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7537 on 4884 degrees of freedom
## Multiple R-squared:  0.2762, Adjusted R-squared:  0.2746
## F-statistic: 169.4 on 11 and 4884 DF,  p-value: < 2.2e-16
```

```
interactions_model <- lm(quality ~ fixed.acidity + volatile.acidity + citric.acid + chlorides + pH + fre

#confint(interactions_model, "alcohol")
```

```
summary(interactions_model)
```

```
##
## Call:
## lm(formula = quality ~ fixed.acidity + volatile.acidity + citric.acid +
##     chlorides + pH + free.sulfur.dioxide + sulphates + total.sulfur.dioxide +
##     alcohol + residual.sugar + residual.sugar * alcohol + total.sulfur.dioxide *
##     alcohol, data = white_wines)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4267 -0.4978 -0.0257  0.4670  3.1968
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                2.4808824  0.4982850   4.979 6.62e-07 ***
## fixed.acidity             -0.0483503  0.0149255  -3.239  0.00121 **
## volatile.acidity          -1.8886746  0.1149272 -16.434  < 2e-16 ***
## citric.acid               -0.0450678  0.0960504  -0.469  0.63894
## chlorides                 -0.8782893  0.5419353  -1.621  0.10516
## pH                         0.1652748  0.0829842   1.992  0.04647 *
## free.sulfur.dioxide        0.0056397  0.0008539   6.605 4.40e-11 ***
## sulphates                  0.4151429  0.0973297   4.265 2.03e-05 ***
## total.sulfur.dioxide      -0.0075835  0.0026748  -2.835  0.00460 **
## alcohol                    0.3178841  0.0331801   9.581  < 2e-16 ***
## residual.sugar             0.0914229  0.0223546   4.090 4.39e-05 ***
## alcohol:residual.sugar    -0.0065596  0.0022154  -2.961  0.00308 **
## total.sulfur.dioxide:alcohol 0.0006467 0.0002565   2.521  0.01174 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.7532 on 4883 degrees of freedom
## Multiple R-squared:  0.2771, Adjusted R-squared:  0.2754
## F-statistic:   156 on 12 and 4883 DF,  p-value: < 2.2e-16
```
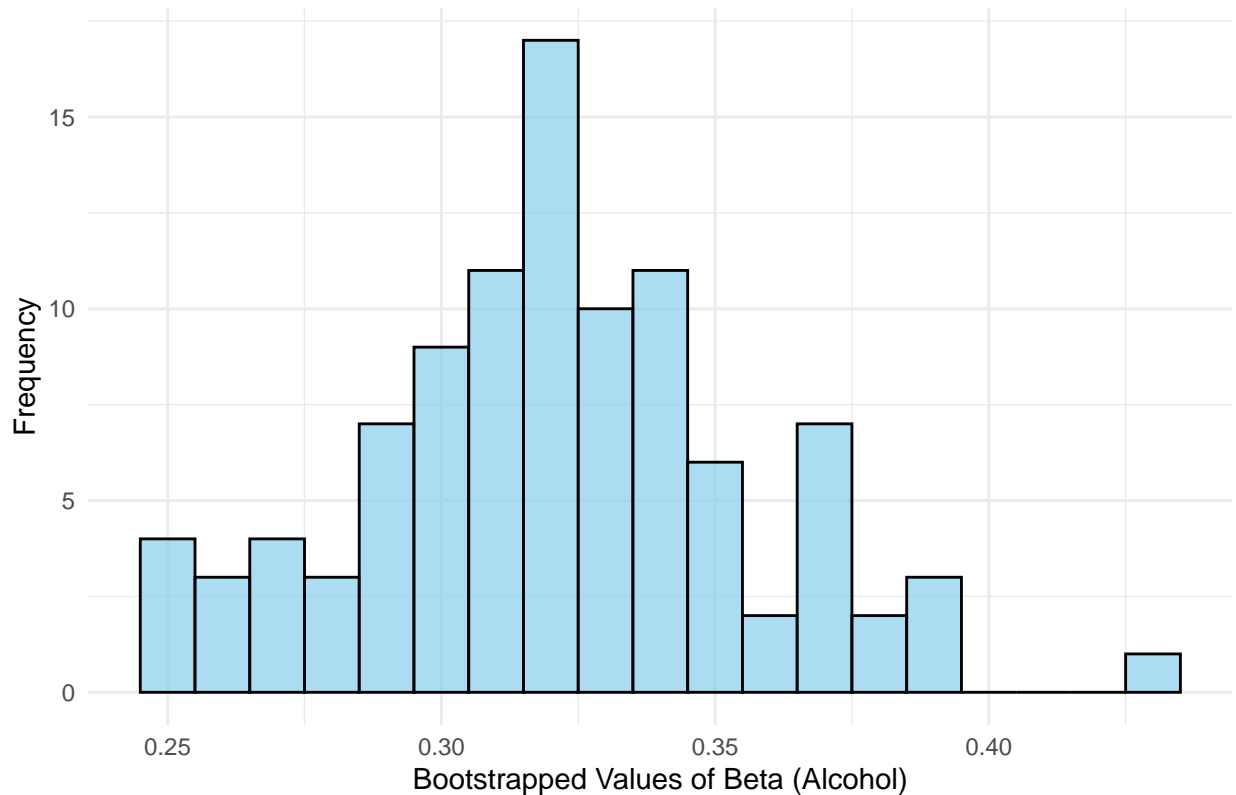
```r
set.seed(123)
boot_results <- boot(data = white_wines, statistic = function(data, indices) {
  model <- lm(quality ~ fixed.acidity + volatile.acidity + citric.acid + chlorides + pH + free.sulfur.d

  return(coef(model)["alcohol"])
}, R = 100)

boot.ci(boot_results, type = "perc", index = 1)
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 100 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = boot_results, type = "perc", index = 1)
##
## Intervals :
## Level      Percentile
## 95%    ( 0.2491,  0.3882 )
## Calculations and Intervals on Original Scale
## Some percentile intervals may be unstable
```

```r
boot_df <- data.frame(boot_values = boot_results$t)
ggplot(boot_df, aes(x = boot_values)) +
  geom_histogram(binwidth = 0.01, fill = "skyblue", color = "black", alpha = 0.7) +
  labs(title = "Bootstrap Distribution of Beta (Alcohol)", x = "Bootstrapped Values of Beta (Alcohol)",
  theme_minimal()
```

Bootstrap Distribution of Beta (Alcohol)


```r
# Load necessary packages
library(boot)
library(ggplot2)

set.seed(123)  # Set seed for reproducibility
boot_results <- boot(data = white_wines, statistic = function(data, indices) {
  model <- lm(quality ~ fixed.acidity + volatile.acidity + citric.acid + chlorides + pH + free.sulfur.di

  return(coef(model)["alcohol"])
}, R = 100)

ci_lower <- quantile(boot_results$t, 0.025)
ci_upper <- quantile(boot_results$t, 0.975)

boot_df <- data.frame(boot_values = boot_results$t)

ggplot(boot_df, aes(x = boot_values)) +
  geom_histogram(binwidth = 0.01, fill = "skyblue", color = "black", alpha = 0.7) +
  geom_vline(xintercept = ci_lower, color = "red", linetype = "dashed") +
  geom_vline(xintercept = ci_upper, color = "red", linetype = "dashed") +
  labs(title = "Bootstrap Distribution of Beta (Alcohol) with 95% CI",
       x = "Bootstrapped Values of Beta (Alcohol)",
       y = "Frequency") +
  theme_minimal() +
  annotate("text", x = ci_lower, y = 3, label = paste("CI Lower:", round(ci_lower, 3)), color = "red") +
  annotate("text", x = ci_upper, y = 3, label = paste("CI Upper:", round(ci_upper, 3)), color = "red")
```
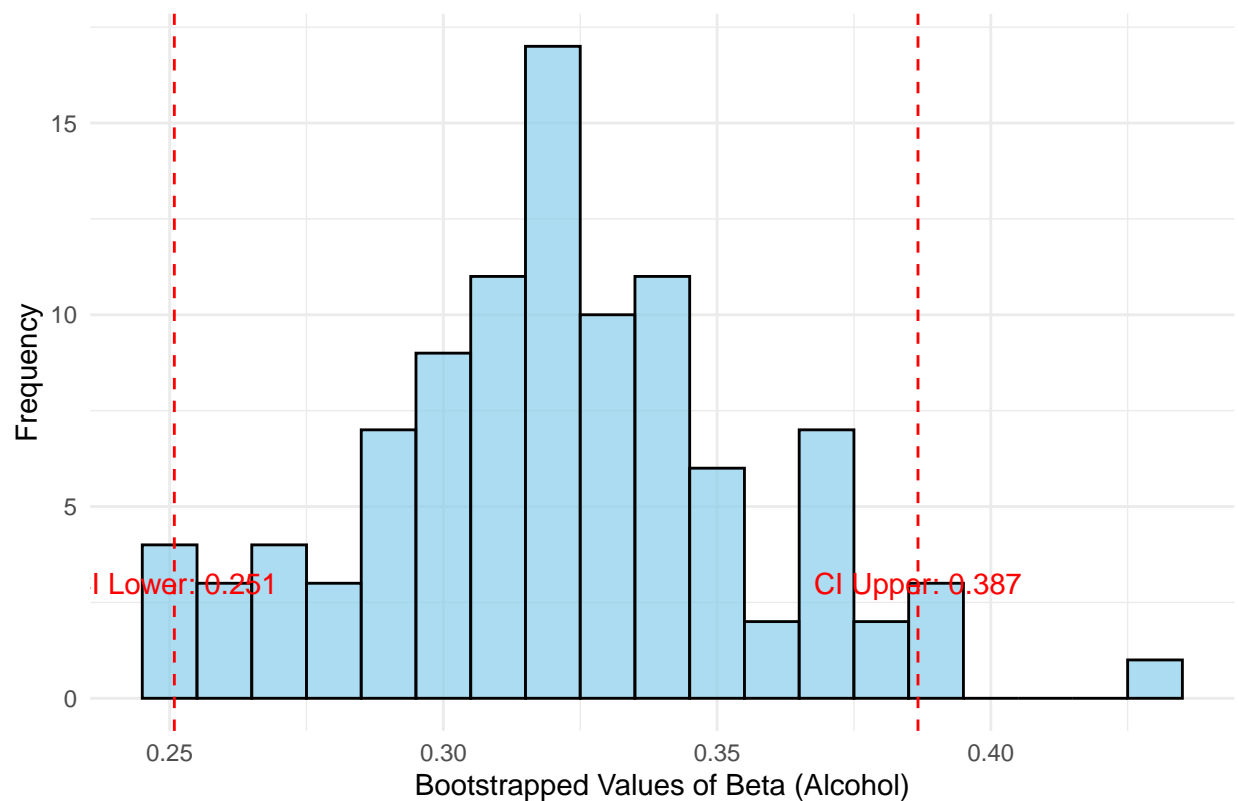
Bootstrap Distribution of Beta (Alcohol) with 95% CI


```r
model_no_alcohol <- lm(quality ~ fixed.acidity + volatile.acidity + citric.acid + chlorides + pH + free
```

```r
summary(model_no_alcohol)
```

```
##
## Call:
## lm(formula = quality ~ fixed.acidity + volatile.acidity + citric.acid +
##     chlorides + pH + free.sulfur.dioxide + sulphates + total.sulfur.dioxide +
##     residual.sugar, data = white_wines)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4363 -0.6215 -0.0083  0.4800  3.2187
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          6.4309501  0.3536935  18.182  < 2e-16 ***
## fixed.acidity       -0.0788144  0.0163545  -4.819 1.49e-06 ***
## volatile.acidity    -1.2731079  0.1234118 -10.316  < 2e-16 ***
## citric.acid          0.1819641  0.1054260   1.726   0.0844 .
## chlorides           -6.9010117  0.5609276 -12.303  < 2e-16 ***
## pH                   0.2248987  0.0908619   2.475   0.0134 *
## free.sulfur.dioxide  0.0079232  0.0009376   8.451  < 2e-16 ***
## sulphates            0.4773660  0.1069275   4.464 8.21e-06 ***
## total.sulfur.dioxide -0.0043689  0.0003927 -11.126  < 2e-16 ***
```

```
## residual.sugar       -0.0045684  0.0026848  -1.702    0.0889 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8305 on 4886 degrees of freedom
## Multiple R-squared:  0.1208, Adjusted R-squared:  0.1192
## F-statistic: 74.58 on 9 and 4886 DF,  p-value: < 2.2e-16
```

Model Comparison

```
AIC(interactions_model, interaction_model)
```

```
##                    df      AIC
## interactions_model 14 11134.55
## interaction_model  13 11138.92
```

```
AIC(interactions_model, model_no_alcohol)
```

```
##                    df      AIC
## interactions_model 14 11134.55
## model_no_alcohol   11 12087.23
```

We favor model that has the interaction term

Calulate F-test:

```
pred_interaction <- predict(interactions_model)
pred_alcohol <- predict(model_alcohol)
pred_no_alcohol <- predict(model_no_alcohol)

# Plot Actual vs Predicted for Interaction Model
plot(white_wines$quality, pred_interaction,
     main = "Predicted vs Actual with Interaction and No Alcohol Models",
     xlab = "Actual Quality", ylab = "Predicted Quality",
     pch = 19, col = "black",
     xlim = c(min(white_wines$quality), max(white_wines$quality)),
     ylim = c(min(white_wines$quality), max(white_wines$quality)))


# Add a line for the Alcohol Model (no interaction)
abline(lm(pred_alcohol ~ white_wines$quality), col = "blue", lwd = 2)

# Add a line for the No Alcohol Model (same slope, different intercept)
abline(lm(pred_no_alcohol ~ white_wines$quality), col = "green", lwd = 2)

# Add a line for the Interaction Model
lines(lowess(white_wines$quality, pred_interaction), col = "red", lwd = 2)
```
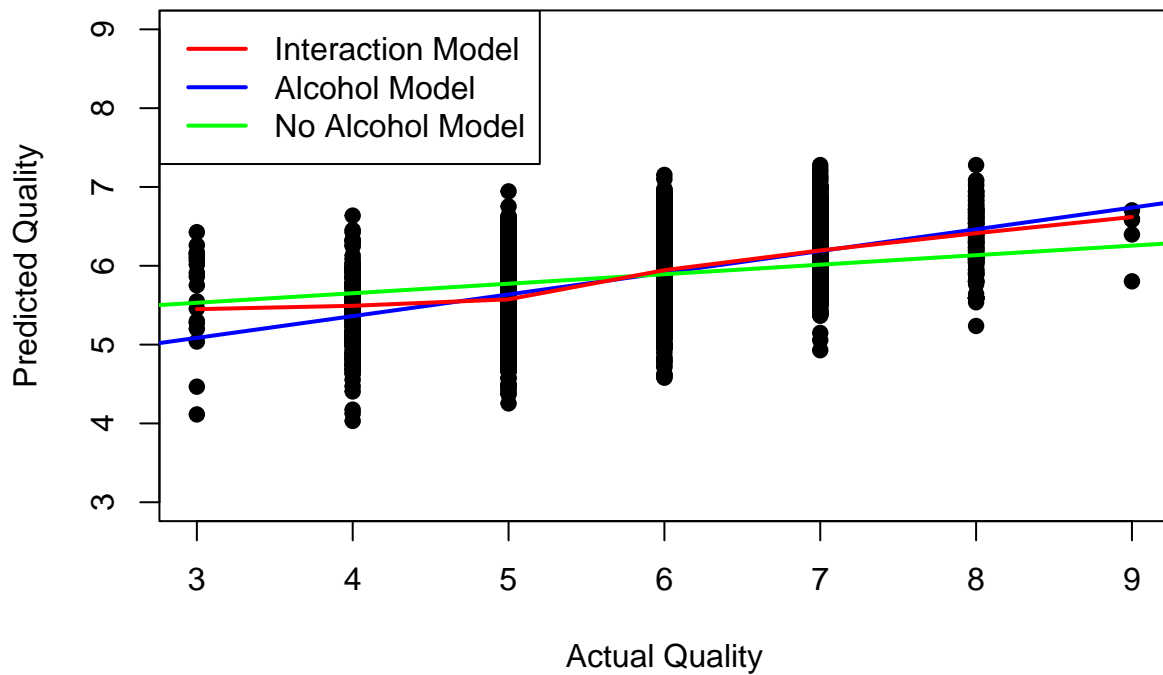
```
# Add a legend to explain the lines
legend("topleft", legend = c("Interaction Model", "Alcohol Model", "No Alcohol Model"),
       col = c("red", "blue", "green"), lwd = 2)
```

Predicted vs Actual with Interaction and No Alcohol Models



```
par(mfrow=c(1,2))  # Split the plot into two panels

# Residuals for interaction model
plot(white_wines$quality, pred_interaction - white_wines$quality,
     main = "Residuals: Interaction Model",
     xlab = "Actual Quality", ylab = "Residuals",
     pch = 19, col = "pink")
abline(h = 0, col = "black", lwd = 2)

# Residuals for no alcohol model
plot(white_wines$quality, pred_no_alcohol - white_wines$quality,
     main = "Residuals: No Alcohol Model",
     xlab = "Actual Quality", ylab = "Residuals",
     pch = 19, col = "green")
abline(h = 0, col = "black", lwd = 2)
```
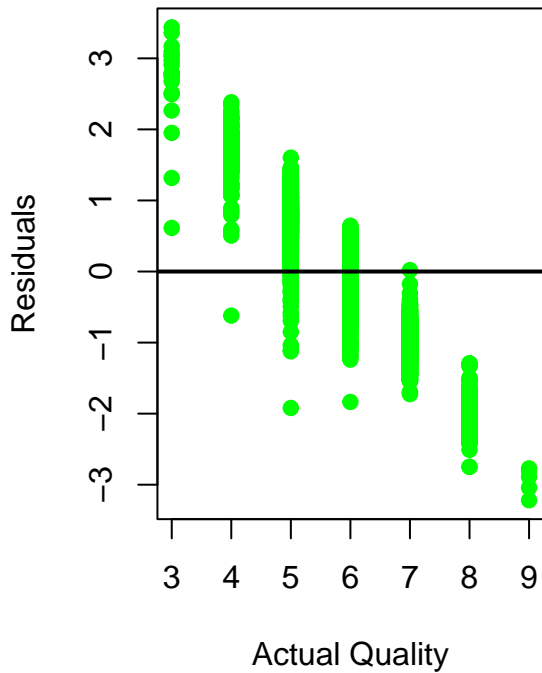
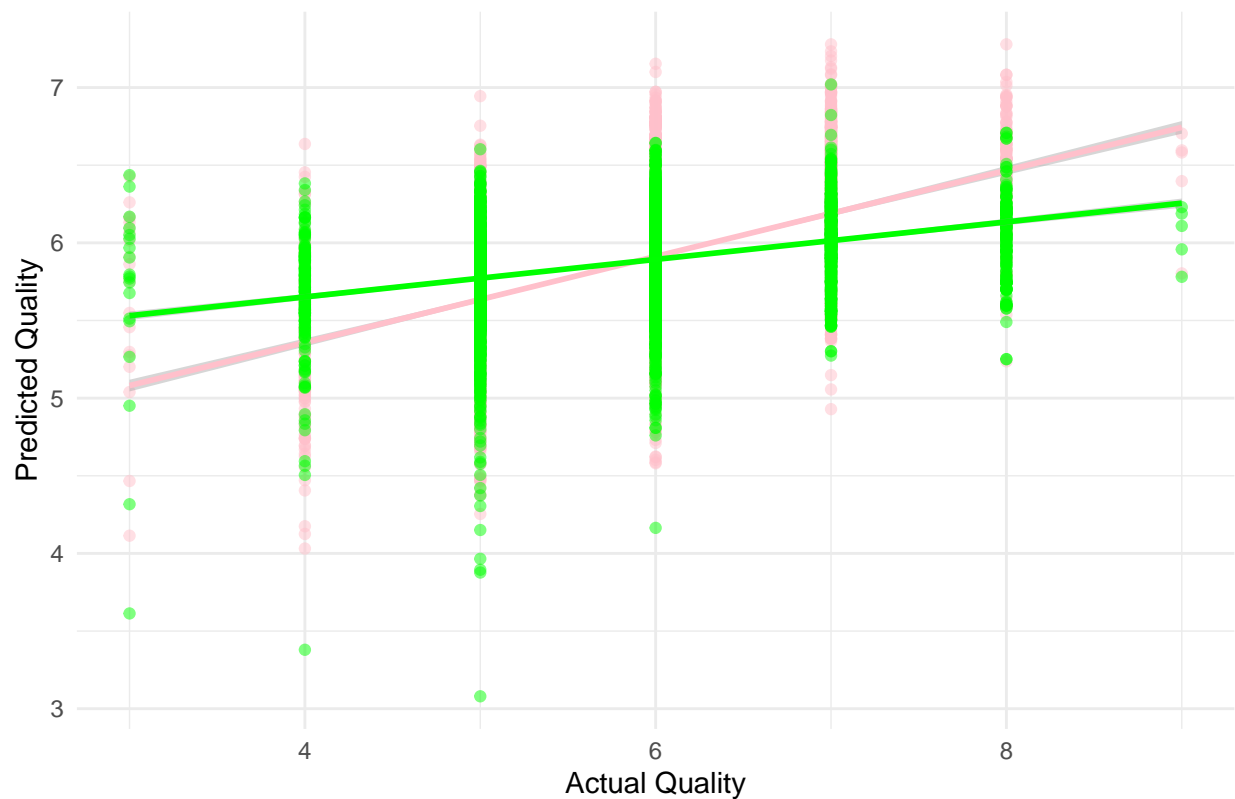## Residuals: Interaction Model
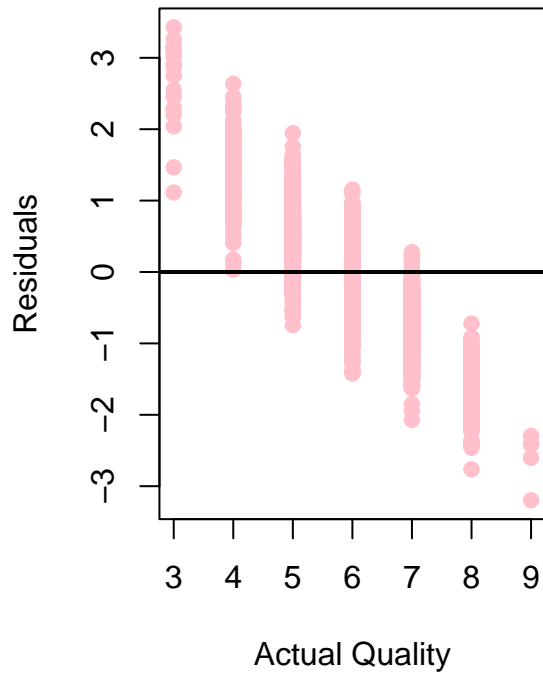
Residuals: No Alcohol Model


```
ggplot(data = white_wines, aes(x = quality)) +
  geom_point(aes(y = pred_interaction), color = "pink", alpha = 0.5) +
  geom_smooth(aes(y = pred_interaction), color = "pink", method = "lm") +
  geom_point(aes(y = pred_no_alcohol), color = "green", alpha = 0.5) +
  geom_smooth(aes(y = pred_no_alcohol), color = "green", method = "lm") +
  labs(title = "Predicted vs Actual: Interaction vs No Alcohol Model",
       x = "Actual Quality", y = "Predicted Quality") +
  theme_minimal()
```

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

Predicted vs Actual: Interaction vs No Alcohol Model

Not include interaction :

```
anova(model_alcohol, model_no_alcohol)
```

```
## Analysis of Variance Table
##
## Model 1: quality ~ fixed.acidity + volatile.acidity + citric.acid + chlorides +
##     pH + free.sulfur.dioxide + sulphates + total.sulfur.dioxide +
##     alcohol + residual.sugar
## Model 2: quality ~ fixed.acidity + volatile.acidity + citric.acid + chlorides +
##     pH + free.sulfur.dioxide + sulphates + total.sulfur.dioxide +
##     residual.sugar
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1   4885 2777.1
## 2   4886 3369.8 -1   -592.66 1042.5 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

AIC method:

```
AIC(model_alcohol, model_no_alcohol)
```

```
##                df      AIC
## model_alcohol     12 11142.18
## model_no_alcohol  11 12087.23
```

Interaction vs No Alcohol Model:

```
AIC(interaction_model, model_no_alcohol)
```

```
##                   df      AIC
## interaction_model 13 11138.92
## model_no_alcohol  11 12087.23
```

P-value

H_0: beta_alcohol = 0

F statistic

```
anova(interaction_model, model_no_alcohol)
```

```
## Analysis of Variance Table
##
## Model 1: quality ~ fixed.acidity + volatile.acidity + citric.acid + chlorides +
##     pH + free.sulfur.dioxide + sulphates + total.sulfur.dioxide +
##     alcohol + residual.sugar + residual.sugar * alcohol
## Model 2: quality ~ fixed.acidity + volatile.acidity + citric.acid + chlorides +
##     pH + free.sulfur.dioxide + sulphates + total.sulfur.dioxide +
##     residual.sugar
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1   4884 2774.1
## 2   4886 3369.8 -2   -595.64 524.33 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-value is extremely small 2.2*e-16, which is well below the 0.01 significance level, we can reject the null hypothesis that the coefficient of alcohol is zero. This provides strong evidence that alcohol has a significant effect on white wine quality.

Furthermore, combining the F-test results with the AIC comparison, we find strong support for alcohol being an important factor in influencing white wine quality.

Thus, higher alcohol content will result in higher quality of wine.

Red wine

```
red_correlation <- cor(red_wines)
```

```r
# CHeck IVF to decide which factor to use
red_model_full <- lm(quality ~ ., data = red_wines)
vif(red_model_full)
```

```
##        fixed.acidity      volatile.acidity           citric.acid
##             7.767512              1.789390              3.128022
##        residual.sugar             chlorides   free.sulfur.dioxide
##             1.702588              1.481932              1.963019
## total.sulfur.dioxide               density                    pH
##             2.186813              6.343760              3.329732
##             sulphates               alcohol
##             1.429434              3.031160
```

```r
red_reduced_model <- lm(quality ~ fixed.acidity + volatile.acidity + citric.acid + chlorides + pH + fre
```

```r
summary(red_model_full)
```

```
##
## Call:
## lm(formula = quality ~ ., data = red_wines)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.68911 -0.36652 -0.04699  0.45202  2.02498
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          2.197e+01  2.119e+01   1.036   0.3002
## fixed.acidity        2.499e-02  2.595e-02   0.963   0.3357
## volatile.acidity    -1.084e+00  1.211e-01  -8.948  < 2e-16 ***
## citric.acid         -1.826e-01  1.472e-01  -1.240   0.2150
## residual.sugar       1.633e-02  1.500e-02   1.089   0.2765
## chlorides           -1.874e+00  4.193e-01  -4.470 8.37e-06 ***
## free.sulfur.dioxide  4.361e-03  2.171e-03   2.009   0.0447 *
## total.sulfur.dioxide -3.265e-03  7.287e-04  -4.480 8.00e-06 ***
## density             -1.788e+01  2.163e+01  -0.827   0.4086
## pH                  -4.137e-01  1.916e-01  -2.159   0.0310 *
## sulphates            9.163e-01  1.143e-01   8.014 2.13e-15 ***
## alcohol              2.762e-01  2.648e-02  10.429  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.648 on 1587 degrees of freedom
## Multiple R-squared:  0.3606, Adjusted R-squared:  0.3561
## F-statistic: 81.35 on 11 and 1587 DF,  p-value: < 2.2e-16
```

```r
AIC(red_model_full, red_reduced_model)
```

```
##                   df      AIC
## red_model_full    13 3164.277
## red_reduced_model 12 3268.271
```

```
anova(red_model_full, red_reduced_model)
```

```
## Analysis of Variance Table
##
## Model 1: quality ~ fixed.acidity + volatile.acidity + citric.acid + residual.sugar +
##     chlorides + free.sulfur.dioxide + total.sulfur.dioxide +
##     density + pH + sulphates + alcohol
## Model 2: quality ~ fixed.acidity + volatile.acidity + citric.acid + chlorides +
##     pH + free.sulfur.dioxide + sulphates + total.sulfur.dioxide +
##     residual.sugar + density
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1   1587 666.41
## 2   1588 712.08 -1   -45.672 108.76 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```