R Notebook

Code

Dataset z https://pubs.com/joelrudinas03/WQD

Obsahuje menení rusnych parametru vin a jejich hodnocení (kvalitu).

Hide

```
winequality <- read.csv("http://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality-white.csv", sep = ";")
```

```
Warning message:
R graphics engine version 14 is not supported by this version of RStudio. The Plots tab will be disabled until a newer version of RStudio is installed.
```

Hide

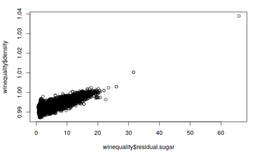
```
summary(winequality)
```

```
 fixed.acidity    volatile.acidity  citric.acid     residual.sugar    chlorides       free.sulfur.dioxide total.sulfur.dioxide   density             pH          sulphates     
 Min.   : 3.800   Min.   :0.0800   Min.   :0.0000   Min.   : 0.600   Min.   :0.00900   Min.   :  2.00     Min.   :  9.0      Min.   :0.9871   Min.   :2.720   Min.   :0.2200  
 1st Qu.: 6.300   1st Qu.:0.2100   1st Qu.:0.2700   1st Qu.: 1.700   1st Qu.:0.03600   1st Qu.: 23.00     1st Qu.: 108.0     1st Qu.:0.9917   1st Qu.:3.090   1st Qu.:0.4100  
 Median : 6.800   Median :0.2600   Median :0.3200   Median : 5.200   Median :0.04300   Median : 34.00     Median : 134.0     Median :0.9937   Median :3.180   Median :0.4700  
 Mean   : 6.855   Mean   :0.2782   Mean   :0.3342   Mean   : 6.391   Mean   :0.04577   Mean   : 35.31     Mean   : 138.4     Mean   :0.9940   Mean   :3.188   Mean   :0.4898  
 3rd Qu.: 7.300   3rd Qu.:0.3200   3rd Qu.:0.3900   3rd Qu.: 9.900   3rd Qu.:0.05000   3rd Qu.: 46.00     3rd Qu.: 167.0     3rd Qu.:0.9961   3rd Qu.:3.280   3rd Qu.:0.5500  
 Max.   :14.200   Max.   :1.1000   Max.   :1.6600   Max.   :65.800   Max.   :0.34600   Max.   :289.00     Max.   :440.0      Max.   :1.0390   Max.   :3.820   Max.   :1.0800  
    alcohol         quality     
 Min.   : 8.00   Min.   :3.000  
 1st Qu.: 9.50   1st Qu.:5.000  
 Median :10.40   Median :6.000  
 Mean   :10.51   Mean   :5.878  
 3rd Qu.:11.40   3rd Qu.:6.000  
 Max.   :14.20   Max.   :9.000  
```

Obsah cuknu vs. hustota vina

Hide

```
plot(winequality$residual.sugar, winequality$density)
```



Jednoduchy linearni model

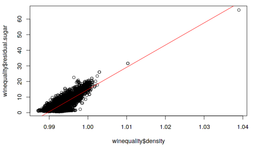
Hide

```
lm.1 = lm(winequality$residual.sugar ~ winequality$density)
summary(lm.1)
```

```

Call:
lm(formula = winequality$residual.sugar ~ winequality$density)

Residuals:
    Min      1Q  Median      3Q     Max 
-9.8434 -1.9079 -0.0151  1.9298 12.8099 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)          -1497.85      13.11  -107.4   <2e-16 ***
winequality$density   1422.74      13.19   107.9   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.76 on 4896 degrees of freedom
Multiple R-squared:  0.7039,    Adjusted R-squared:  0.7038 
F-statistic: 1.164e+04 on 1 and 4896 DF,  p-value: < 2.2e-16
```

Hide

```
plot(winequality$residual.sugar ~ winequality$density)
abline(lm.1,col='red')
```



Jedno menení je hodne vzdalene od ostatnich. nuzhodla jsem se jej vyloucit.

Hide

```
max_density = max(winequality$density)
winequality2 = winequality[!winequality$density==max_density,]
```

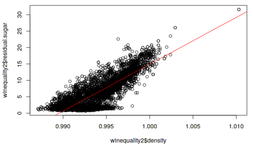
Hide

```
lm.2 = lm(winequality2$residual.sugar ~ winequality2$density)
summary(lm.2)
```

```

Call:
lm(formula = winequality2$residual.sugar ~ winequality2$density)

Residuals:
    Min      1Q  Median      3Q     Max 
-9.9638 -1.8980 -0.0128  1.9299 12.8775 

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)          -1412.72      13.42  -105.3   <2e-16 ***
winequality2$density  1427.63      13.50   105.7   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.76 on 4895 degrees of freedom
Multiple R-squared:  0.6955,    Adjusted R-squared:  0.6954 
F-statistic: 1.118e+04 on 1 and 4895 DF,  p-value: < 2.2e-16
```

Hide

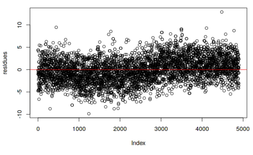
```
plot(winequality2$residual.sugar ~ winequality2$density)
abline(lm.2, col='red')
```



Residua

Hide

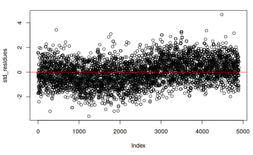
```
residuals = resid(lm.2)
plot(residuals)
abline(h=0, col='red')
```



Standardizovana residua

Hide

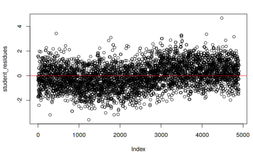
```
std_residuals = rstandard(lm.2)
plot(std_residuals)
abline(h=0, col='red')
```



Studentizovana residua

Hide

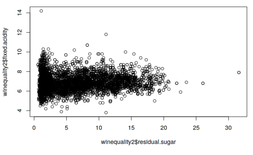
```
student_residues = rstudent(lm.2)
plot(student_residues)
abline(h=0, col='red')
```



Jiny model

Hide

```
plot(winequality2$residual.sugar, winequality2$fixed.acidity)
```



Hide

```
lm.3 = lm(winequality2$residual.sugar ~ winequality2$density + winequality2$fixed.acidity)
summary(lm.3)
```

```

Call:
lm(formula = winequality2$residual.sugar ~ winequality2$density + 
    winequality2$fixed.acidity)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.4500 -1.8006  0.0539  1.8469 11.8904 

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)             -1.474e+03  1.339e+01 -110.09   <2e-16 ***
winequality2$density     1.495e+03  1.355e+01  110.33   <2e-16 ***
winequality2$fixed.acidity -8.689e-01  4.691e-02  -18.52   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.668 on 4894 degrees of freedom
Multiple R-squared:  0.7155,    Adjusted R-squared:  0.7153 
F-statistic:  6153 on 2 and 4894 DF,  p-value: < 2.2e-16
```