WFT Yield Linear Regression with Known Yield Trackers

Adam Yang

8/2/2018

Introduction

Over the course of 14 hp development, the main aim of the characterization team is to figure out how to maximize the core yield per chip on our product wafers. The team has mainly utilized Diagon (SRAM) yield and Sail (Logic) yield to predict the resulting product chip yield. In addition, we've identified Fin Residuals, Epi Etch Out, and faulty DC10 tool usage to be some major product yield detractors.

The goal of this excercise is to create a linear regression model based on each of the factors that we've discovered to have an impact on wafer final test (WFT) yield. The variables used in the regression are:

Variable	Label	Source
wft8to24	Wafer Final Test 8 to 24 Core Yield	Brian Walsh
wft22to24	Wafer Final Test 22 to 24 Core Yield	Brian Walsh
dgvmax	Diagon A102 1.25v Perfect Yield (SRAM)	Baldy
sailvmax	Sail 1.25v Perfect Yield (Logic)	Baldy
sailvmin_ly	Sail 0.5v Limited Yield (Epi Etch Out Tracker)	Baldy
fin_res	Y = Yes, M = Maybe, N = No for Fin Res	Adam Yang
fin_res_metric_2	Numeric Representation of Fin Res	Adam Yang
PC.DC10	Did the wafer go through DC10 tool in PC?	Baldy
CT.DC10	Did the wafer go through DC10 tool in CT?	Baldy
DC10	Did the wafer go through DC10 tool in PC or CT?	Baldy

Data Loading and Cleaning

```
setwd("~/Desktop/WFTAnalysis/")
getwd()
```

```
## [1] "/Users/adamyang/Desktop/WFTAnalysis"
```

```
# Pull csv file and create wft df
wft <- read.csv("WFT.csv")
# Change column names for easier calls
colnames(wft)[3:7] <- c("wft8to24", "wft22to24", "dgvmax", "sailvmax", "sailvmin_ly")
# Convert wft yields to percentage values
wft$wft8to24 <- wft$wft8to24*100
wft$wft22to24 <- wft$wft22to24*100
# Create USwft df for only US wafers
USwft <- wft[wft$FAMILY_CODE == "US",]
```

```
library(car)
library(reshape2)
library(ggplot2)
```

```
library(sandwich)
library(stargazer)
```

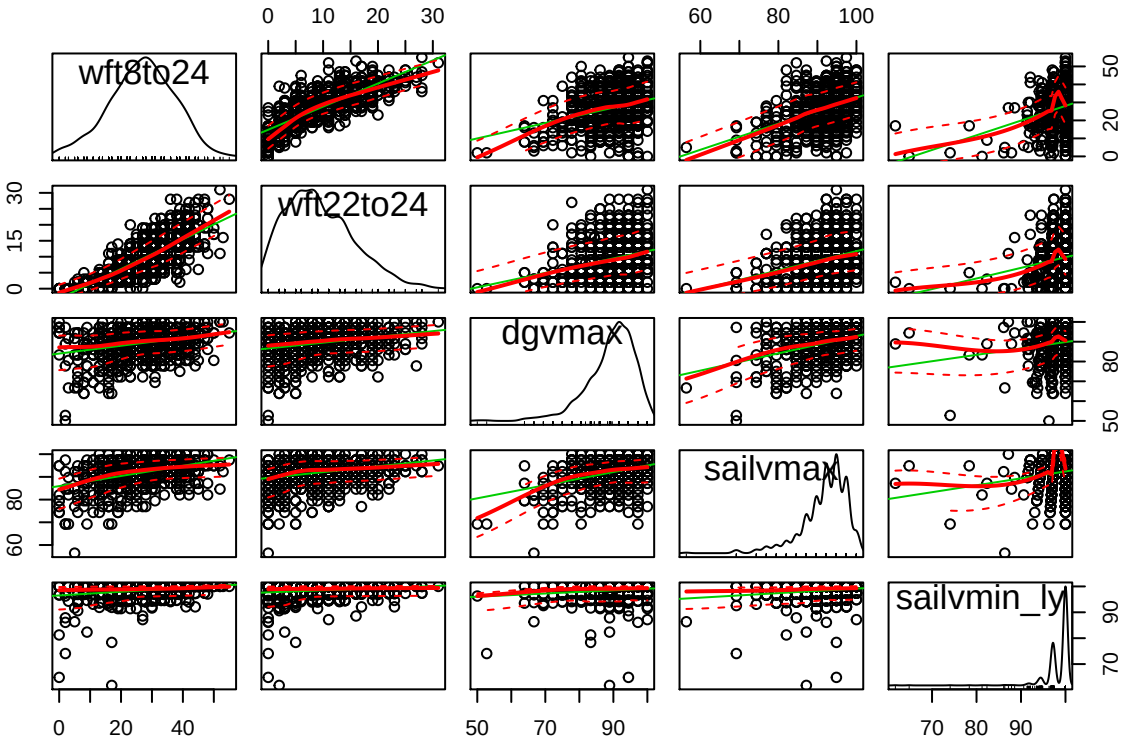
```
##
## Please cite as:
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
library(lmtest)
```

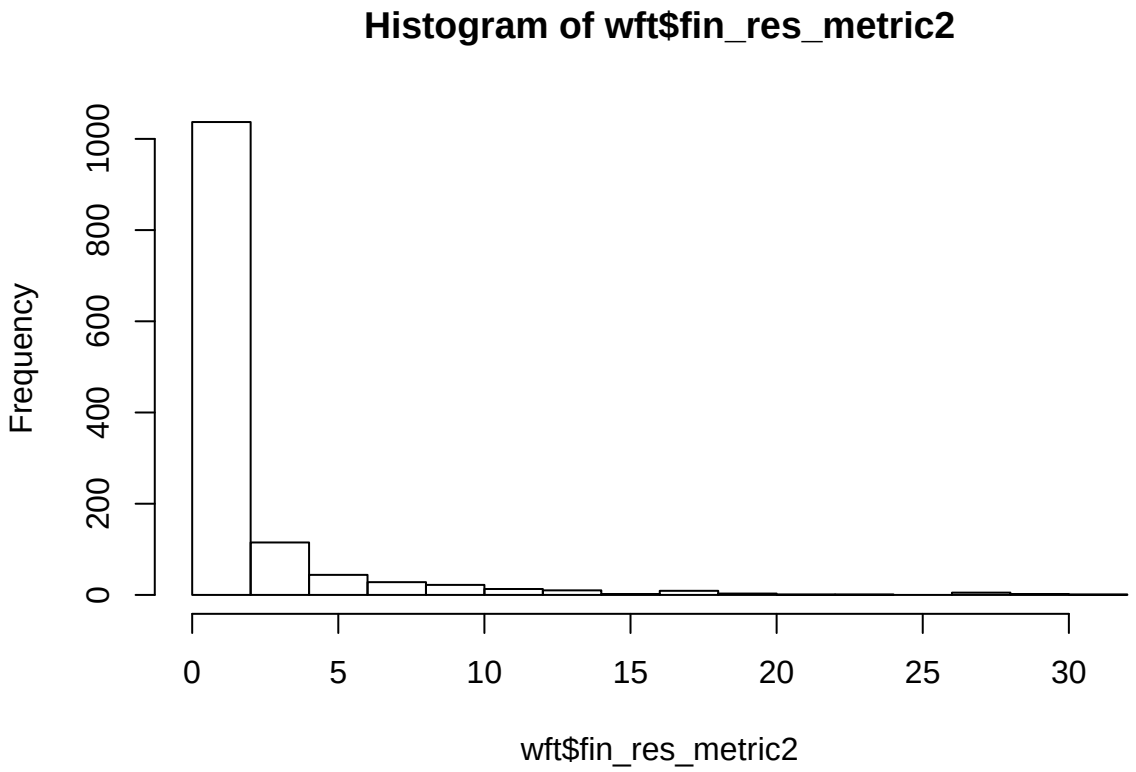
```
## Loading required package: zoo
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

Scatterplot Matrix and Check for Multicolinearity

```
#scatterplotMatrix(wft[3:7], c("wft8to24", "wft22to24", "dgvmax", "sailvmax", "sailvmin_ly"))
scatterplotMatrix(~wft8to24+wft22to24+dgvmax+sailvmax+sailvmin_ly, data = wft)
```

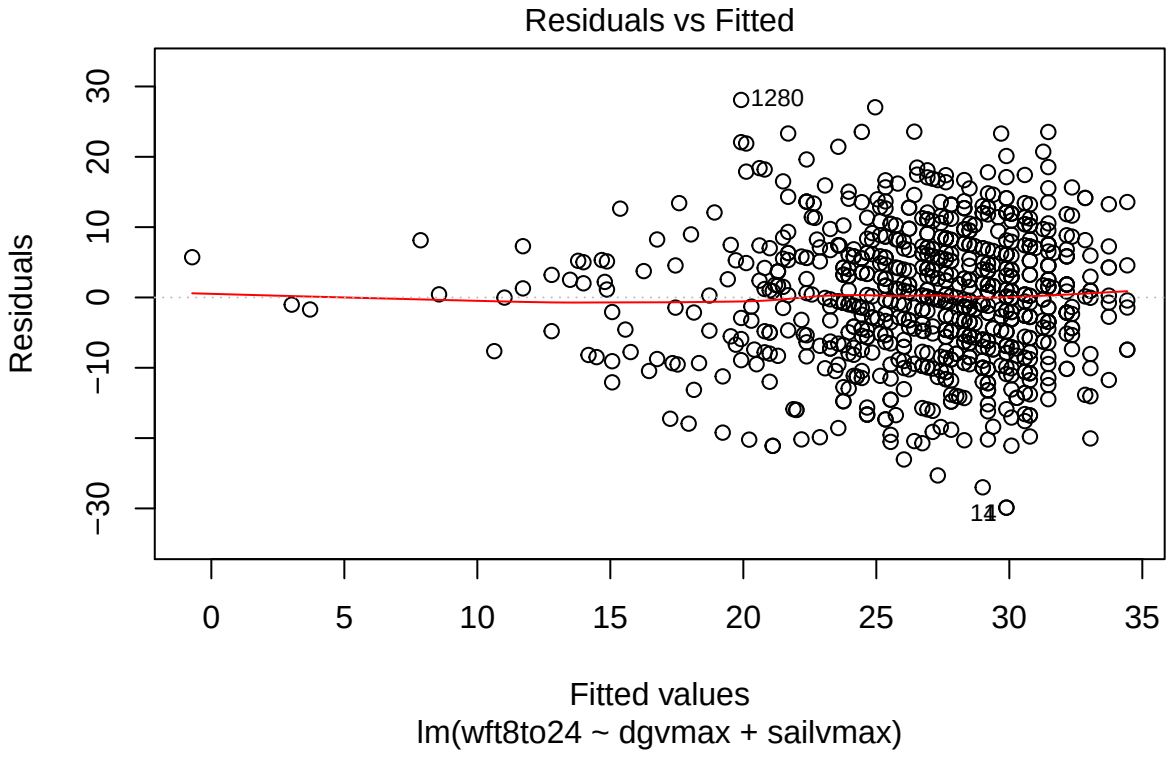



```
hist(wft$fin_res_metric2)
```

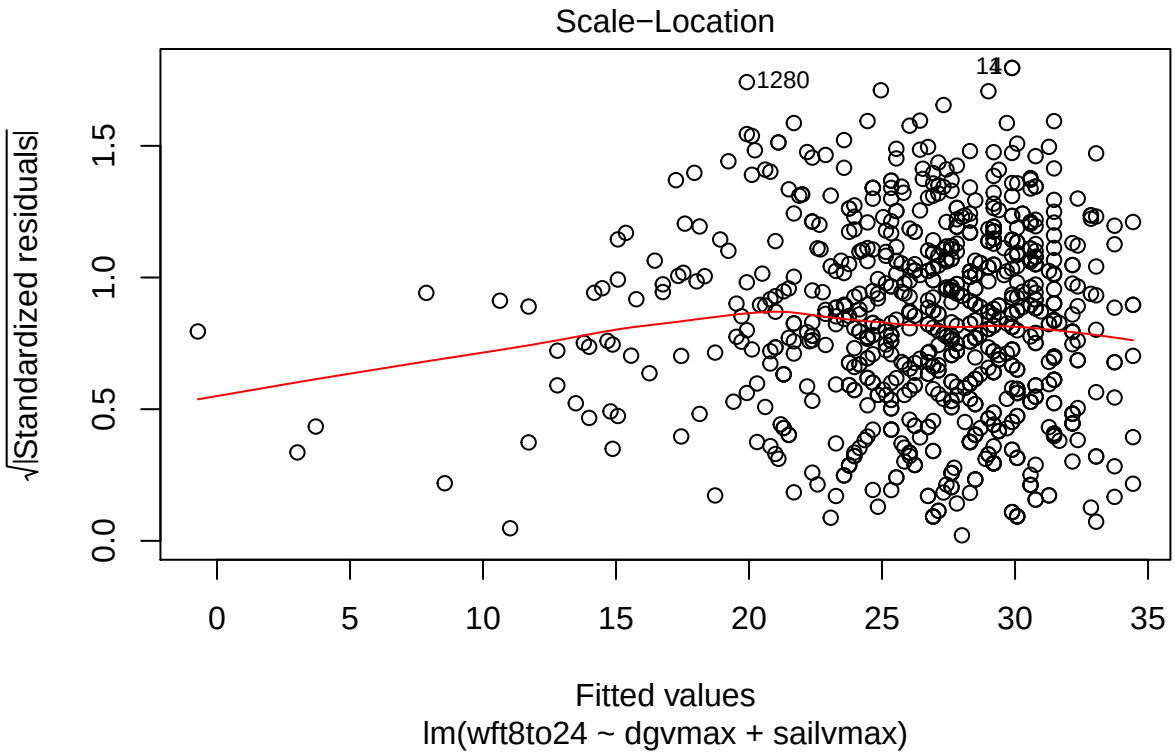


exists by lot so not a totally random sample.

```
# Create model for 8-24c yield vs. only dg and sail yield.
model1 <- lm(wft8to24 ~ dgvmax+sailvmax, data = wft)
# Residuals-Fitted Values plot (Zero Conditional Mean)
plot(model1, 1)
```



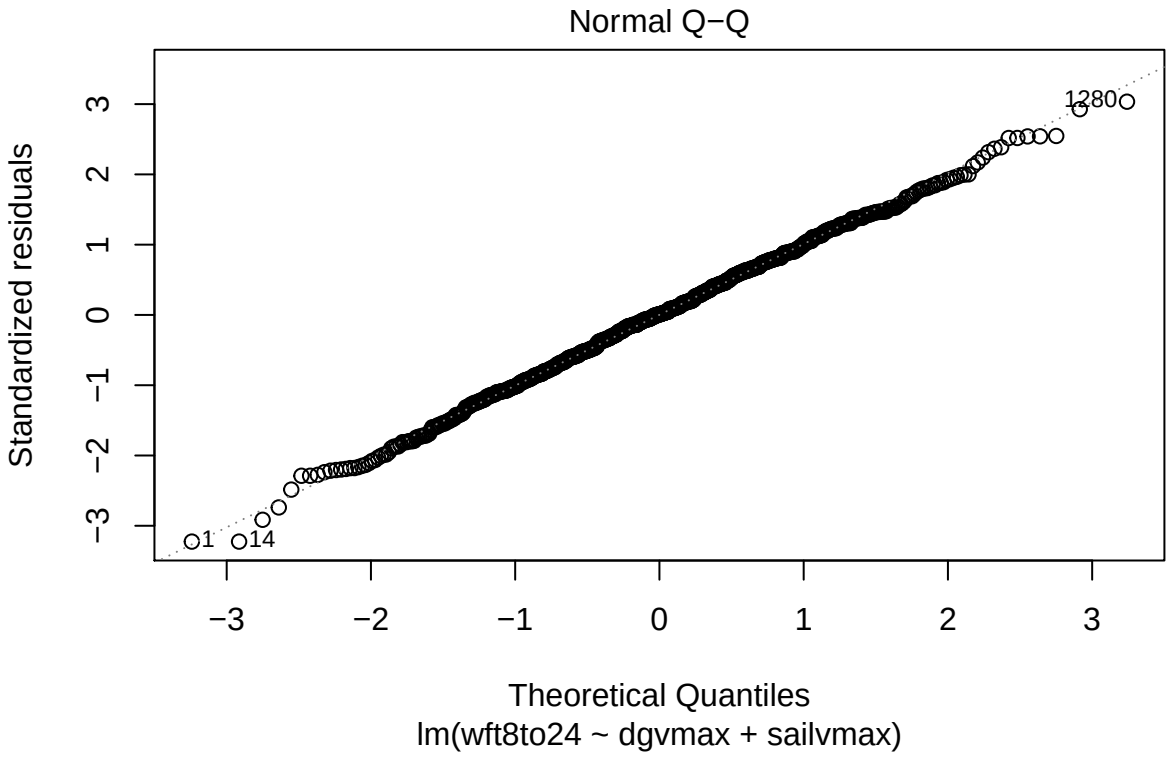
```r
# Scale-Location plot (Homoskedasticity)
plot(model1, 3)
```




```r
# Breusch-Pagan Test for Homoskedasticity
bptest(model1)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model1
## BP = 6.8171, df = 2, p-value = 0.03309
```

```r
# Q-Q Plot (Normality of Residuals)
plot(model1, 2)
```

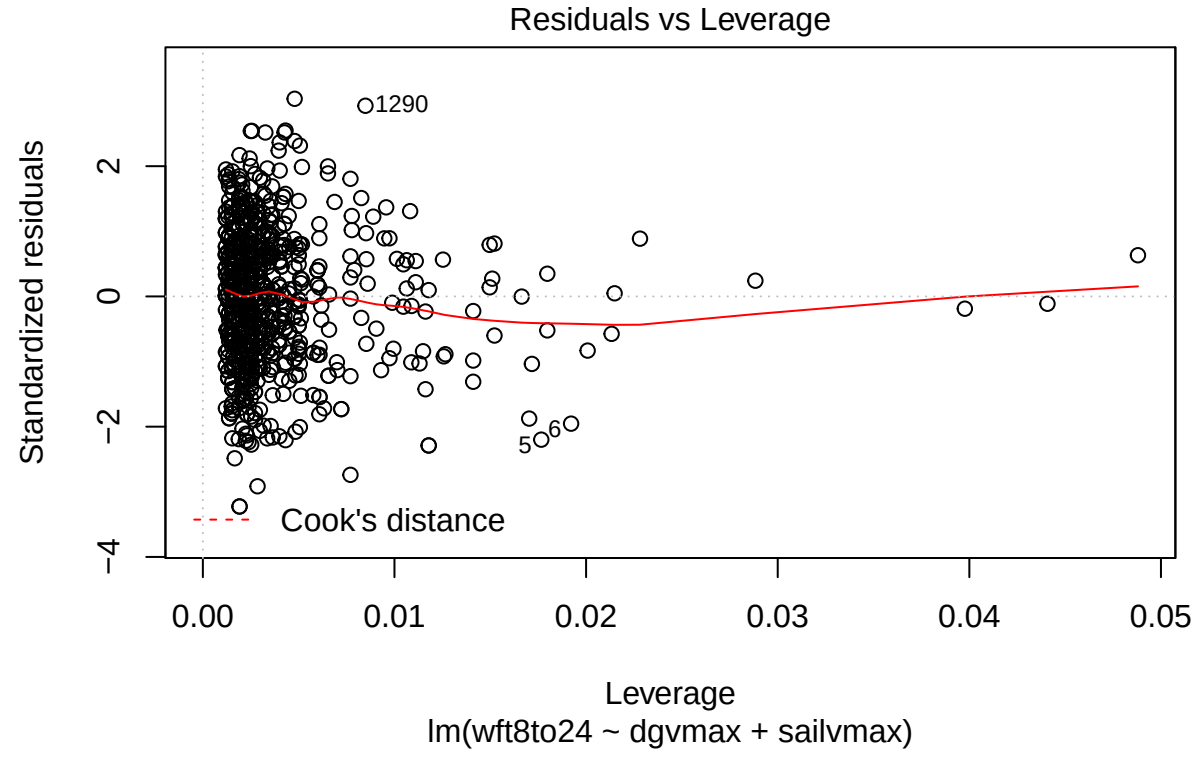



```
# Shapiro-Wilk test of normality
shapiro.test(model1$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  model1$residuals
## W = 0.99855, p-value = 0.7293
```

```
# Histogram of Residuals
hist(model1$residuals)
```


```
# Residuals-Leverage plot
plot(model1, 5) # No problem with Cook's distance
```




```
# a general rule is that if 1 % (or more) data points have standardized residuals > 2.5, the model cont
a <- rstudent(model1)
```

```r
b <- a[a < -3]
```

```r
#linearHypothesis(model1, c("dgvmax = 0", " sailvmax = 0"), vcov = vcovHC)
linearHypothesis(model1, c("dgvmax = 0", " sailvmax = 0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## dgvmax = 0
## sailvmax = 0
##
## Model 1: restricted model
## Model 2: wft8to24 ~ dgvmax + sailvmax
##
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1    840 88680
## 2    838 72106  2     16573 96.305 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

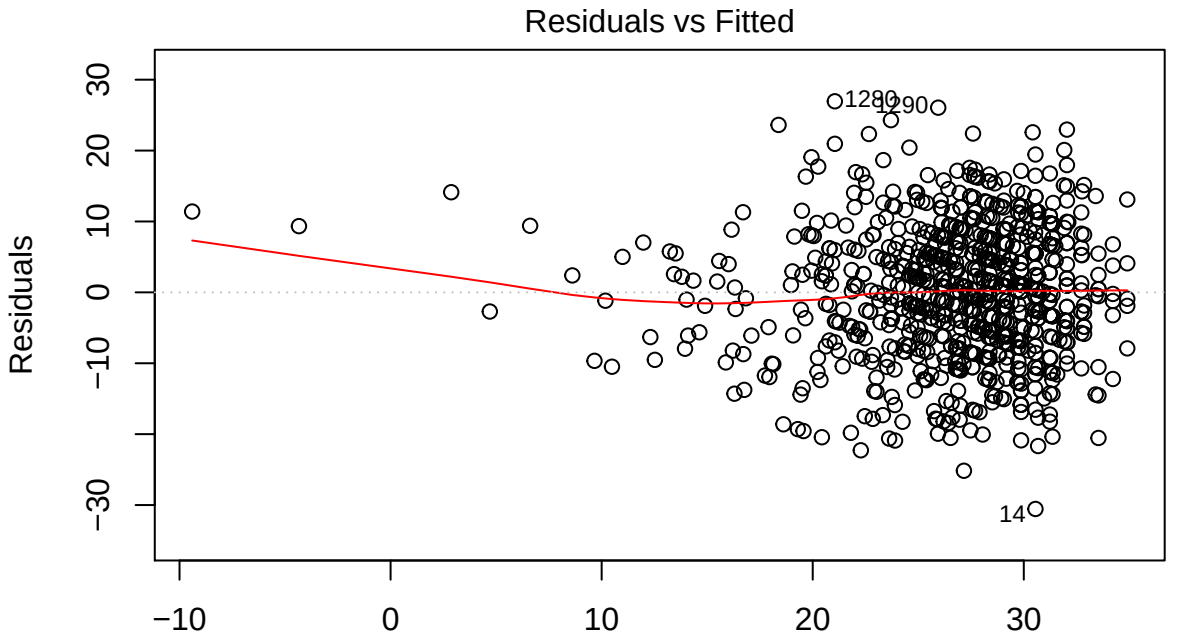
```r
paste("adj.r.square:", summary(model1)$adj.r.squared)
```

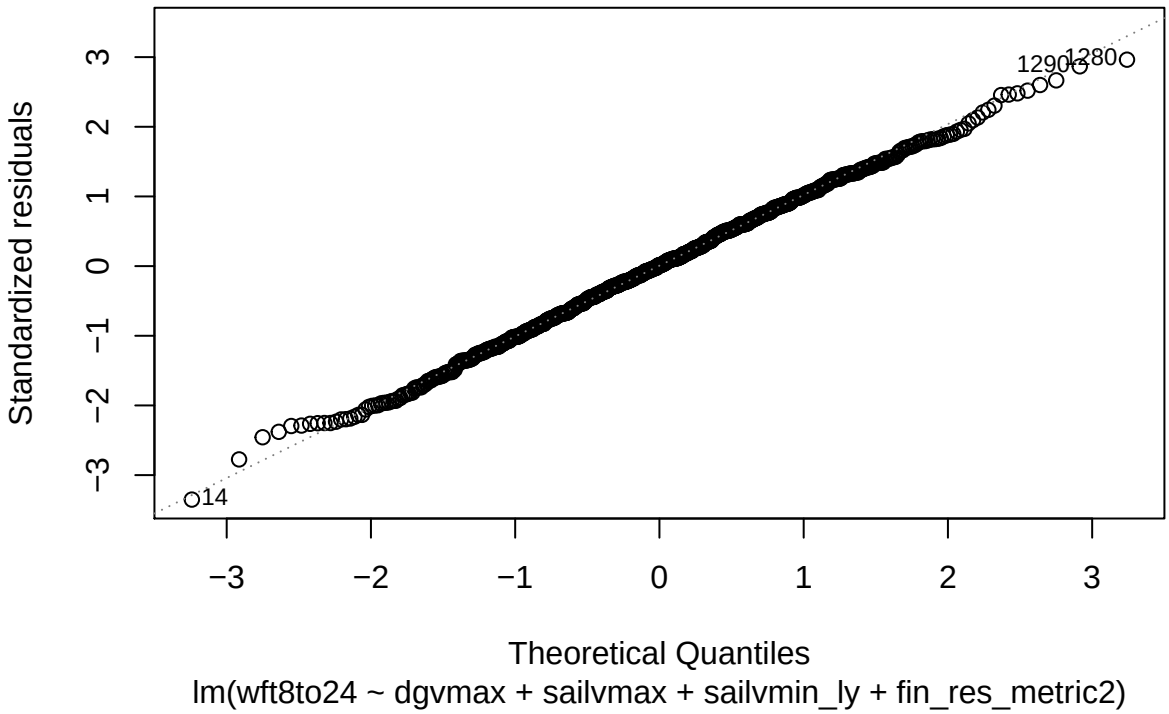
```
## [1] "adj.r.square: 0.18494932526888"
```

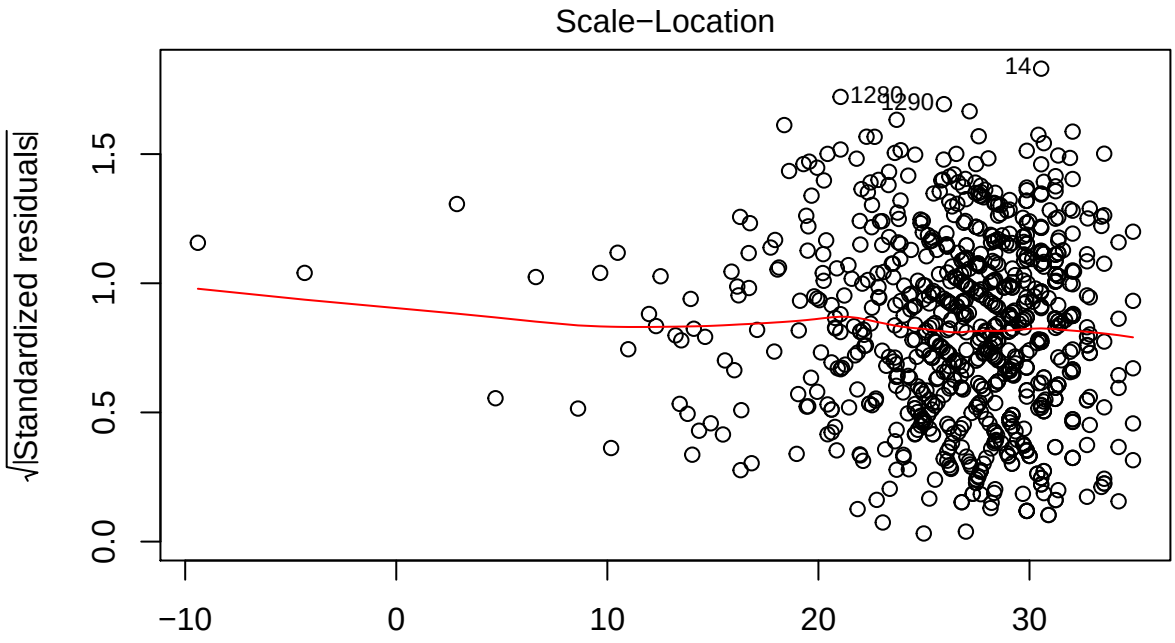
```r
coeftest(model1)
```

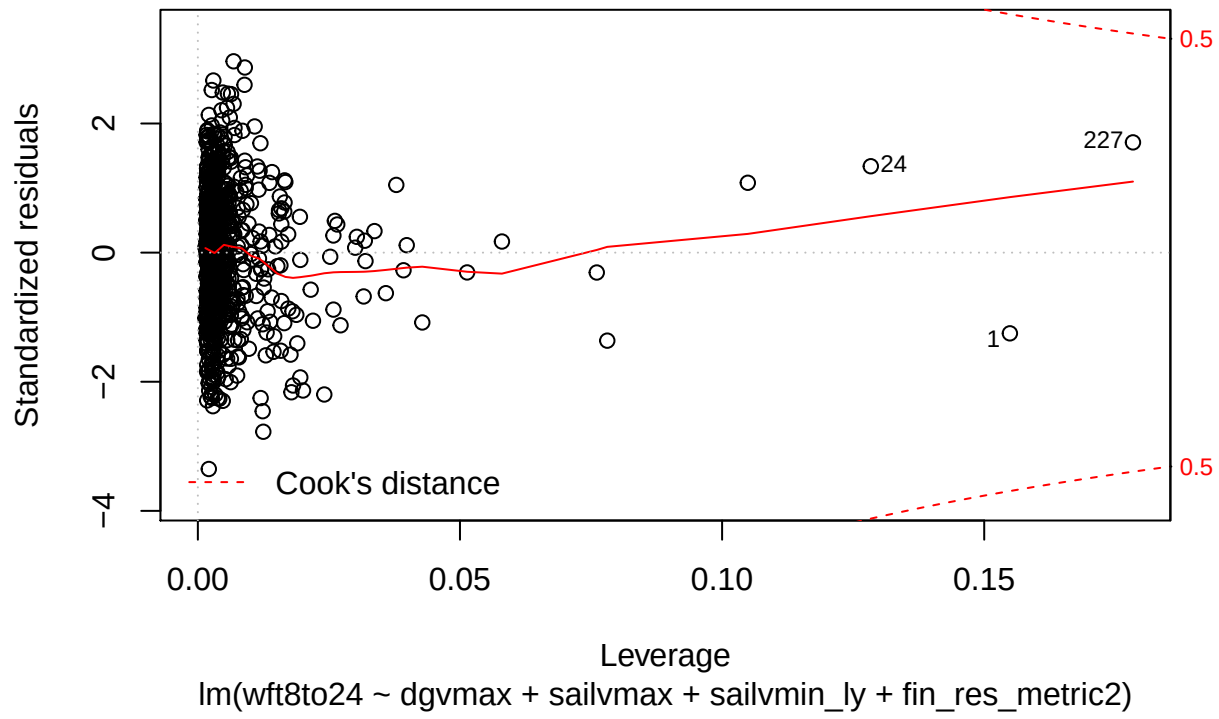
```
##
## t test of coefficients:
##
##               Estimate Std. Error t value  Pr(>|t|)
## (Intercept) -52.075764   5.702401 -9.1323 < 2.2e-16 ***
## dgvmax        0.249257   0.048869  5.1005 4.191e-07 ***
## sailvmax      0.615849   0.060044 10.2567 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
model <- lm(wft8to24 ~ dgvmax+sailvmax+sailvmin_ly+fin_res_metric2, data = wft)
plot(model)
```


Residuals vs Fitted
Residuals
Fitted values
lm(wft8to24 ~ dgvmax + sailvmax + sailvmin_ly + fin_res_metric2)
1280
1290
14


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(wft8to24 ~ dgvmax + sailvmax + sailvmin_ly + fin_res_metric2)
1290
1280
14






```
model$coefficients
```

```
##     (Intercept)          dgvmax        sailvmax     sailvmin_ly
##   -104.93457613      0.24539425      0.58201035      0.57097667
## fin_res_metric2
##      0.07480625
```

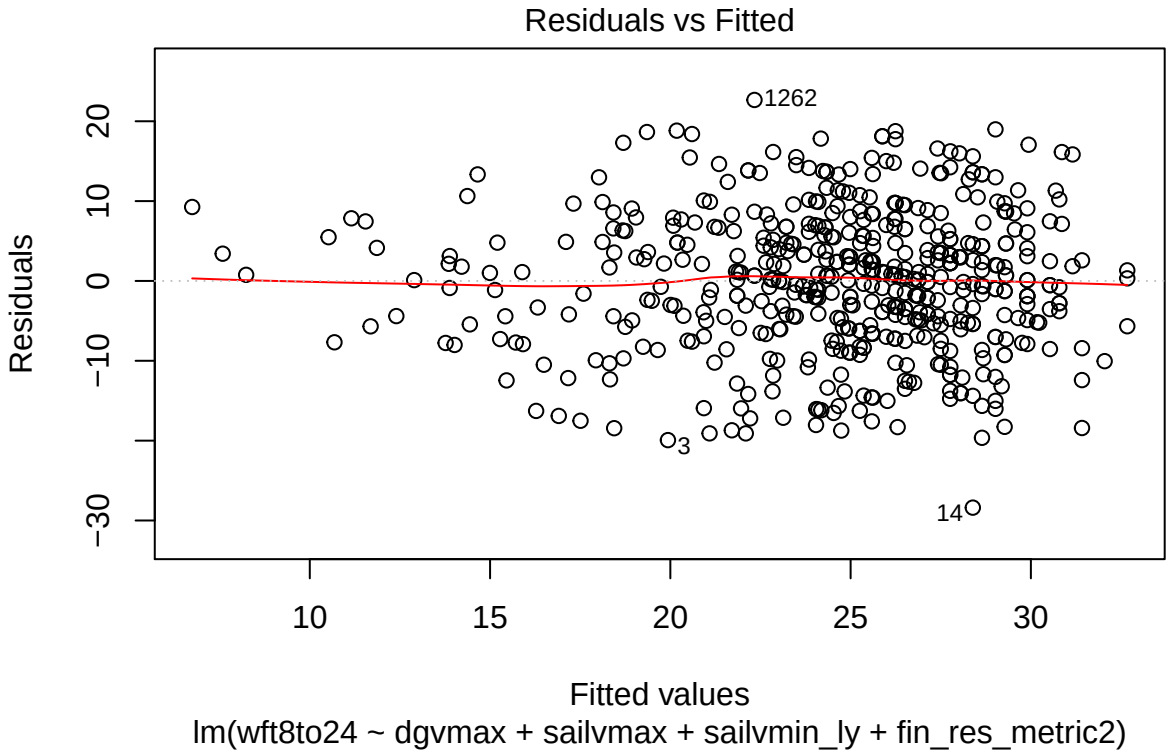
```r
sd(wft$wft8to24)
```

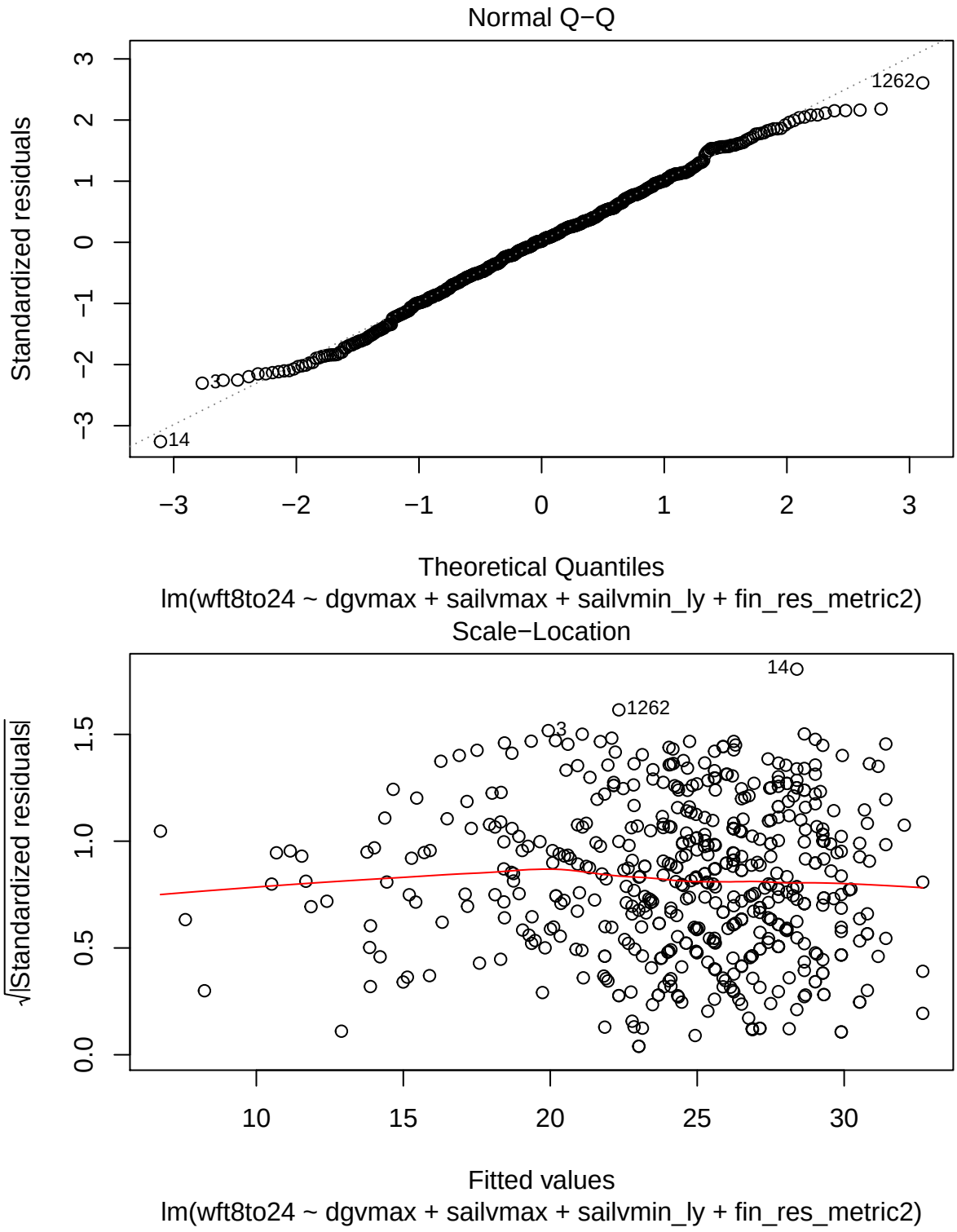
```
## [1] 10.42721
```

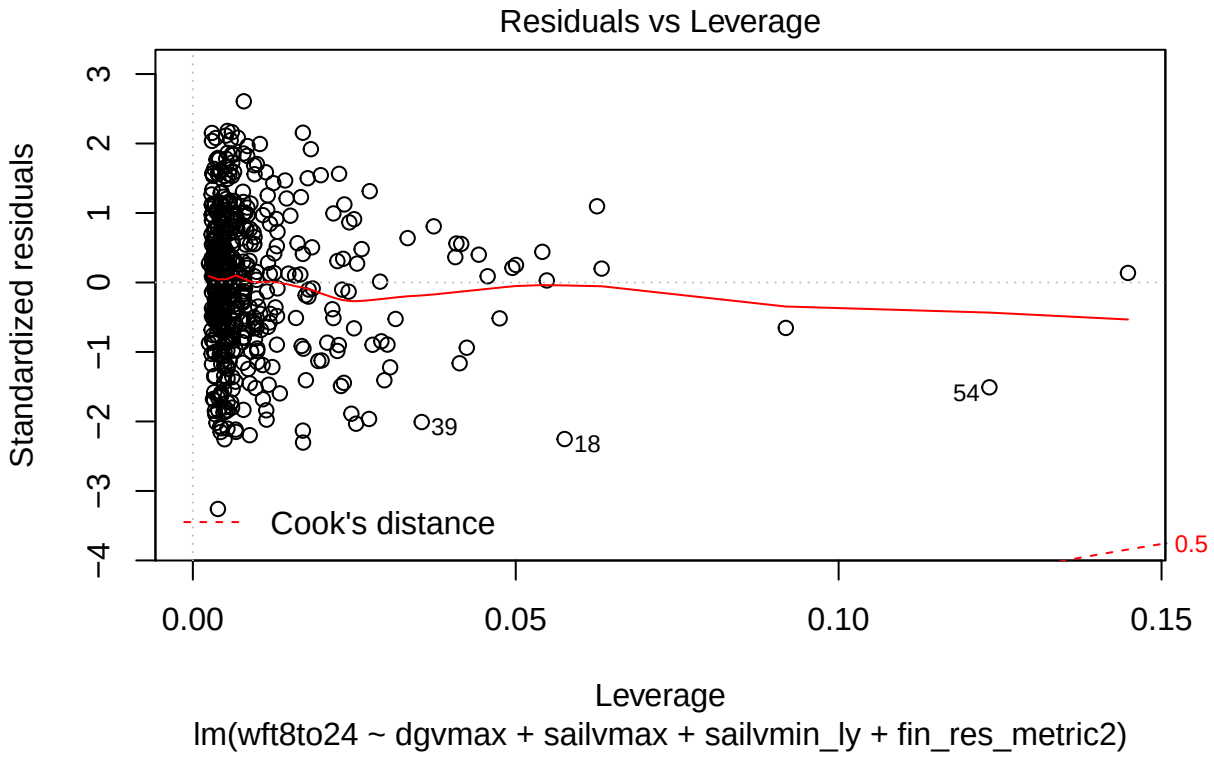
```r
summary(model)$adj.r.square
```

```
## [1] 0.2110533
```

```r
USmodel <- lm(wft8to24 ~ dgvmax+sailvmax+sailvmin_ly+fin_res_metric2, data = USwft)
plot(USmodel)
```



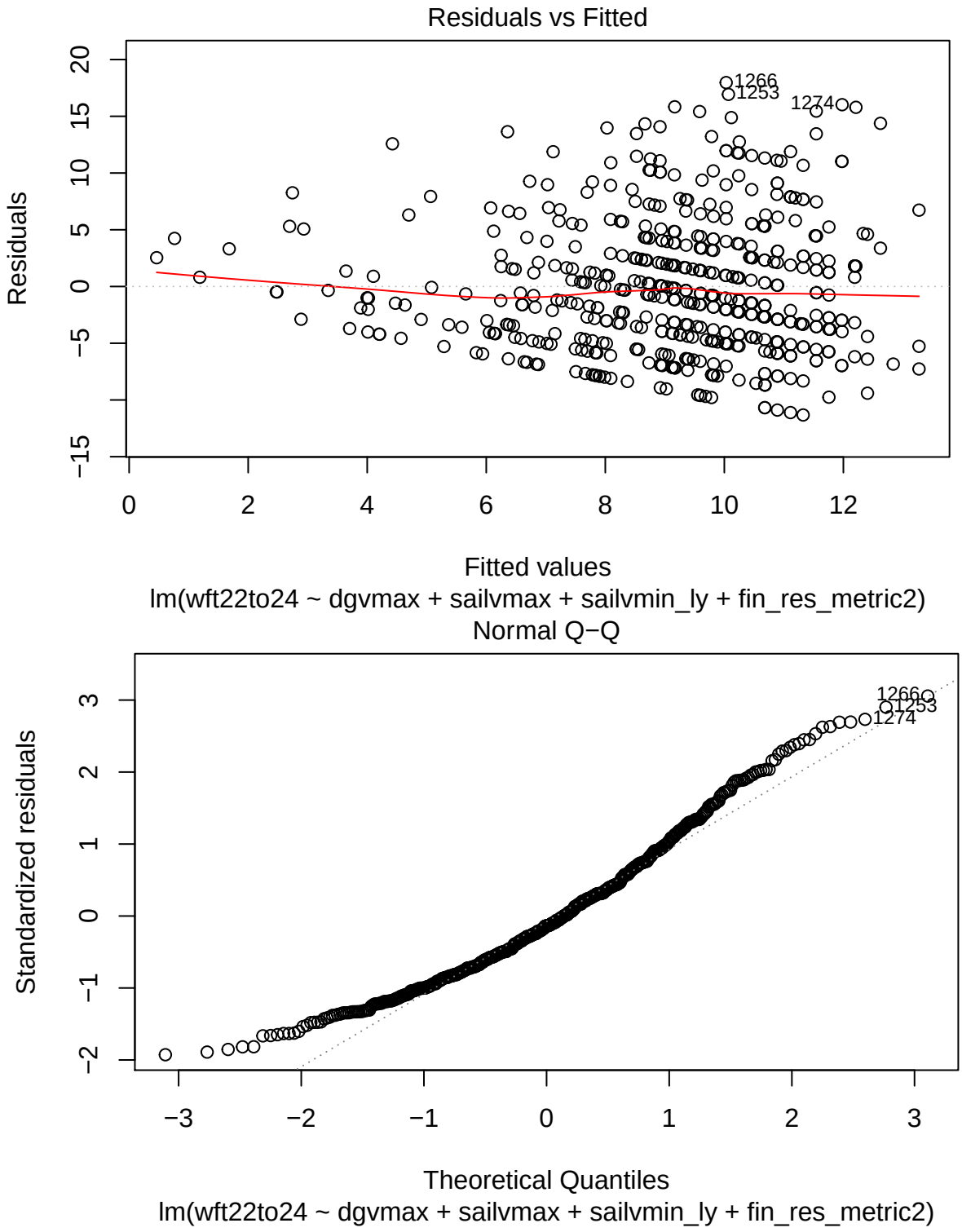

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(wft8to24 ~ dgvmax + sailvmax + sailvmin_ly + fin_res_metric2)
Scale−Location
√|Standardized residuals|
Fitted values
lm(wft8to24 ~ dgvmax + sailvmax + sailvmin_ly + fin_res_metric2)

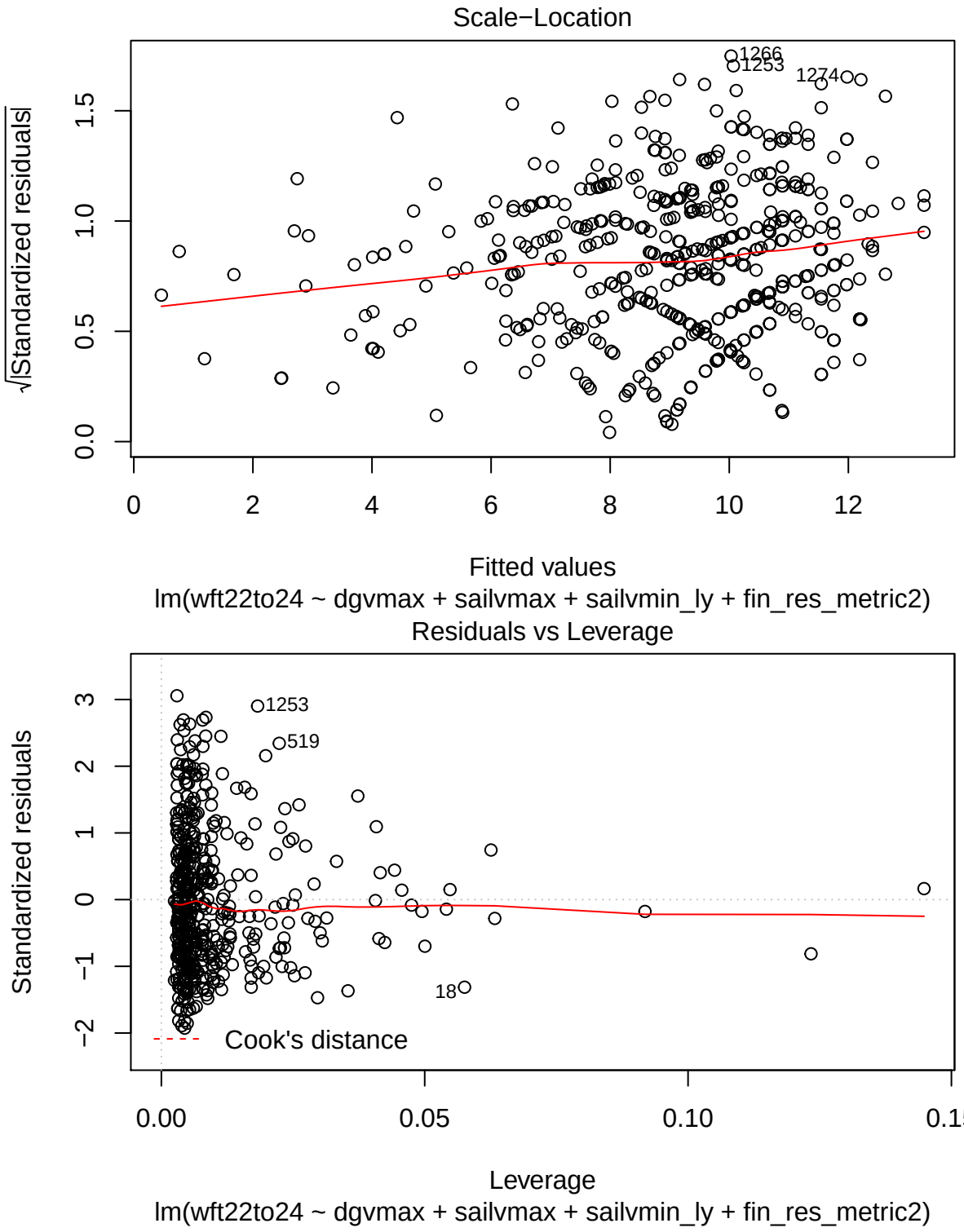



```
USmodel$coefficients
```

```
##     (Intercept)          dgvmax        sailvmax     sailvmin_ly
##     -95.1175539       0.2259661       0.5902504       0.4616980
## fin_res_metric2
##       0.1418049
```

```
USmodel2 <- lm(wft22to24 ~ dgvmax+sailvmax+sailvmin_ly+fin_res_metric2, data = USwft)
plot(USmodel2)
```


Residuals vs Fitted
Residuals
Fitted values
lm(wft22to24 ~ dgvmax + sailvmax + sailvmin_ly + fin_res_metric2)
1266
1253
1274
Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(wft22to24 ~ dgvmax + sailvmax + sailvmin_ly + fin_res_metric2)




```
USmodel2$coefficients
```

```
##     (Intercept)          dgvmax        sailvmax     sailvmin_ly
##     -51.9119588       0.1557406       0.2520808       0.2440057
## fin_res_metric2
##       0.1008990
```