dataset1

ning

07/12/2021

```
yacht<-read.table("yacht_hydrodynamics.data")
head(yacht)
```

```
##     V1    V2   V3   V4   V5    V6   V7
## 1 -2.3 0.568 4.78 3.99 3.17 0.125 0.11
## 2 -2.3 0.568 4.78 3.99 3.17 0.150 0.27
## 3 -2.3 0.568 4.78 3.99 3.17 0.175 0.47
## 4 -2.3 0.568 4.78 3.99 3.17 0.200 0.78
## 5 -2.3 0.568 4.78 3.99 3.17 0.225 1.18
## 6 -2.3 0.568 4.78 3.99 3.17 0.250 1.82
```

```
library(randomForest)
```

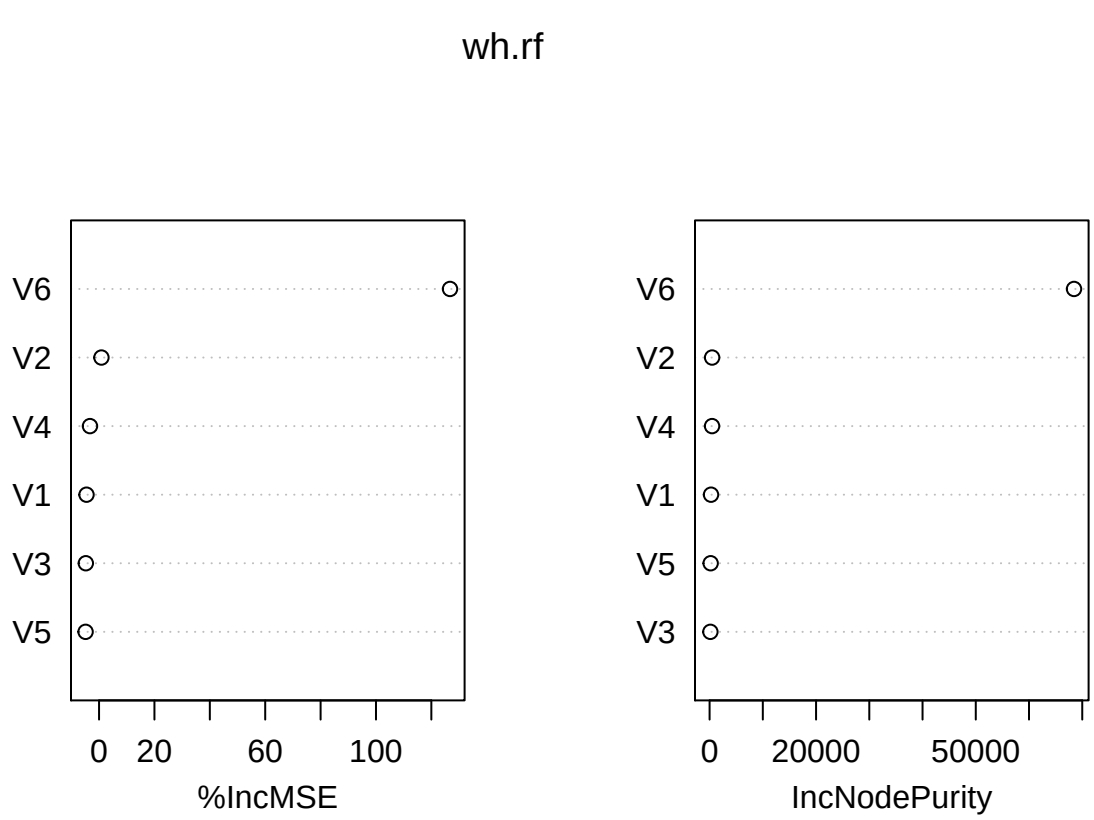
```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

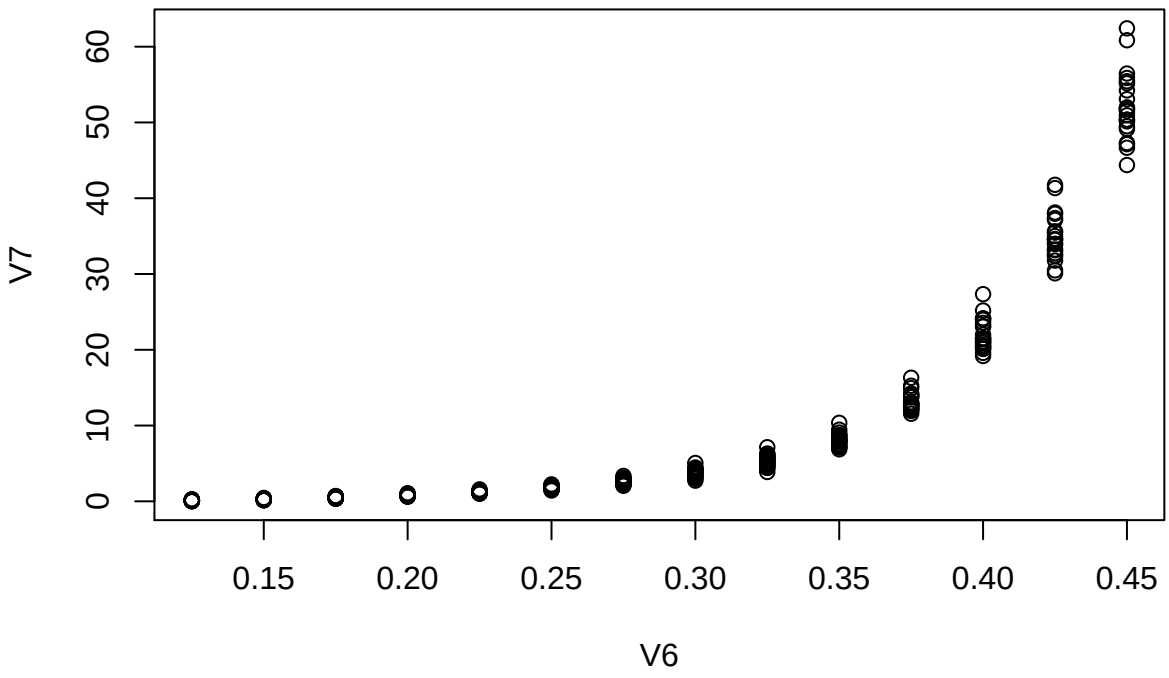
```
wh.rf <- randomForest(data=yacht, V7~ . , mtry=4, nodesize=7,
                      importance=TRUE, keep.forest=TRUE)
round(importance(wh.rf),3)
```

```
##    %IncMSE IncNodePurity
## V1  -4.531       270.824
## V2   0.854       475.063
## V3  -4.783       154.207
## V4  -3.281       466.482
## V5  -4.871       202.534
## V6 126.740     68454.096
```

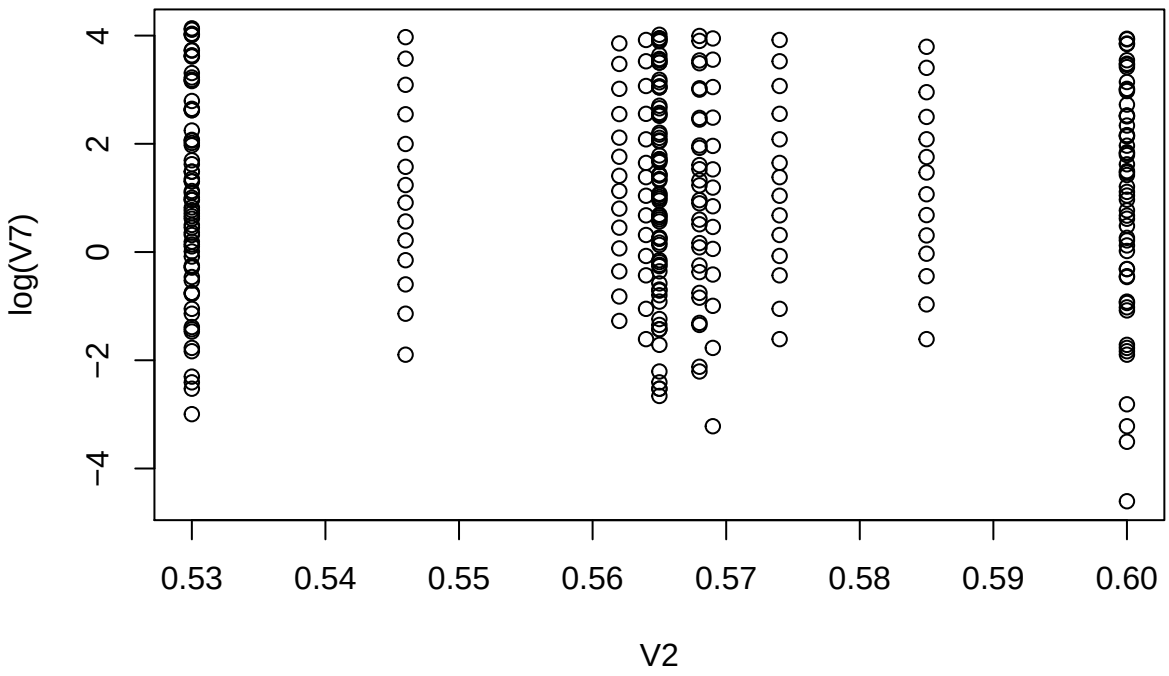
```
varImpPlot(wh.rf)
```



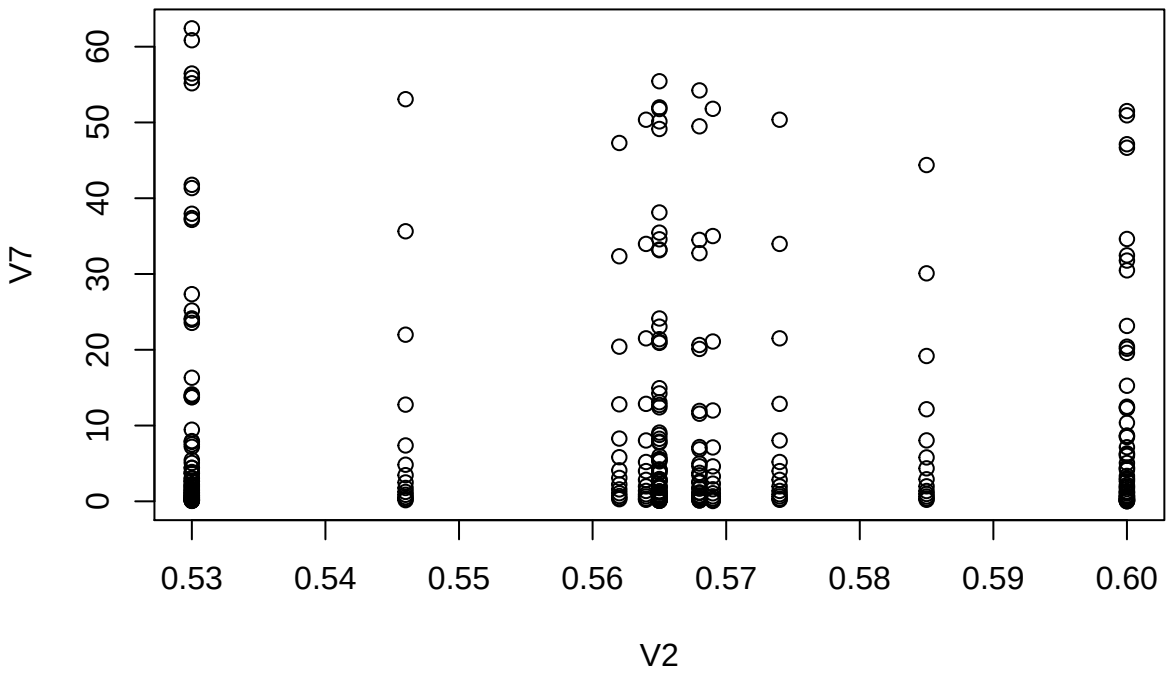

```
plot(V7~V6,data=yacht)
```



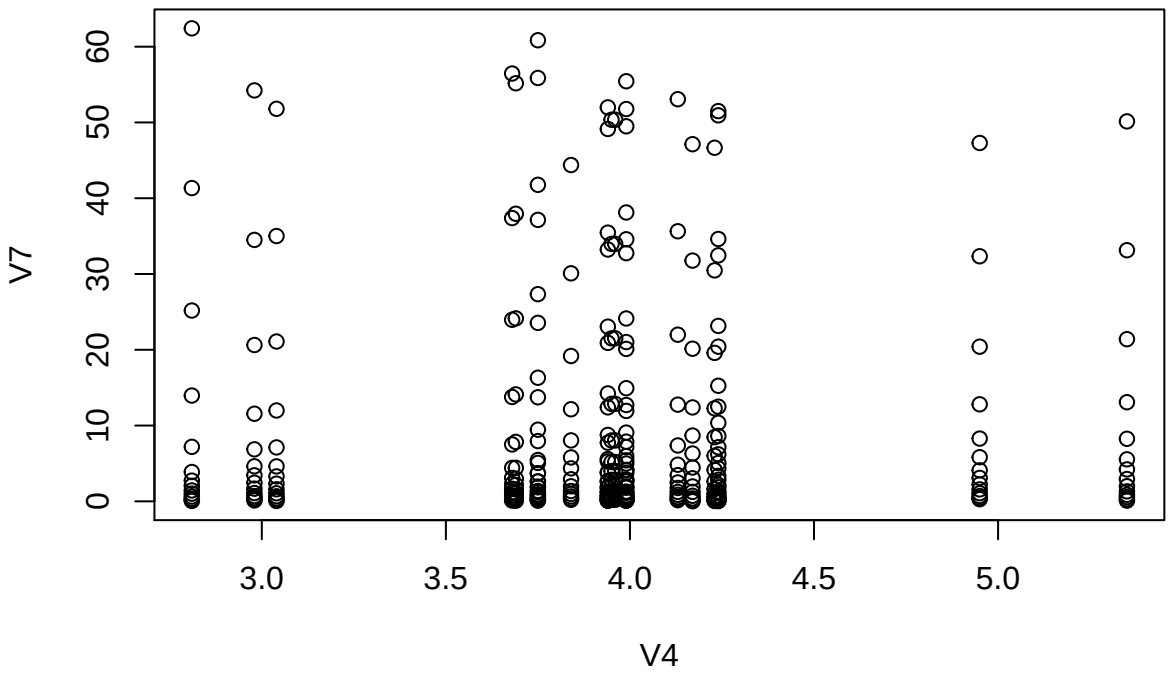

```
plot(log(V7)~V2,data=yacht)
```



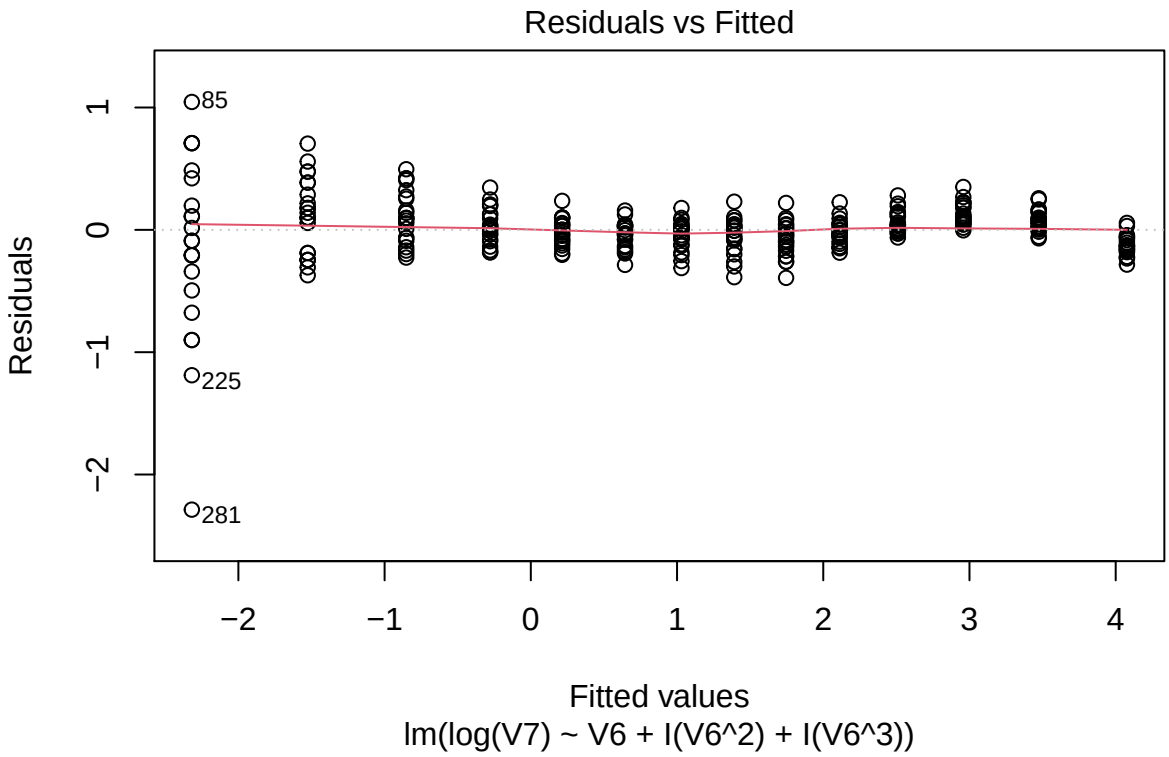

```
plot(V7~V2,data=yacht)
```

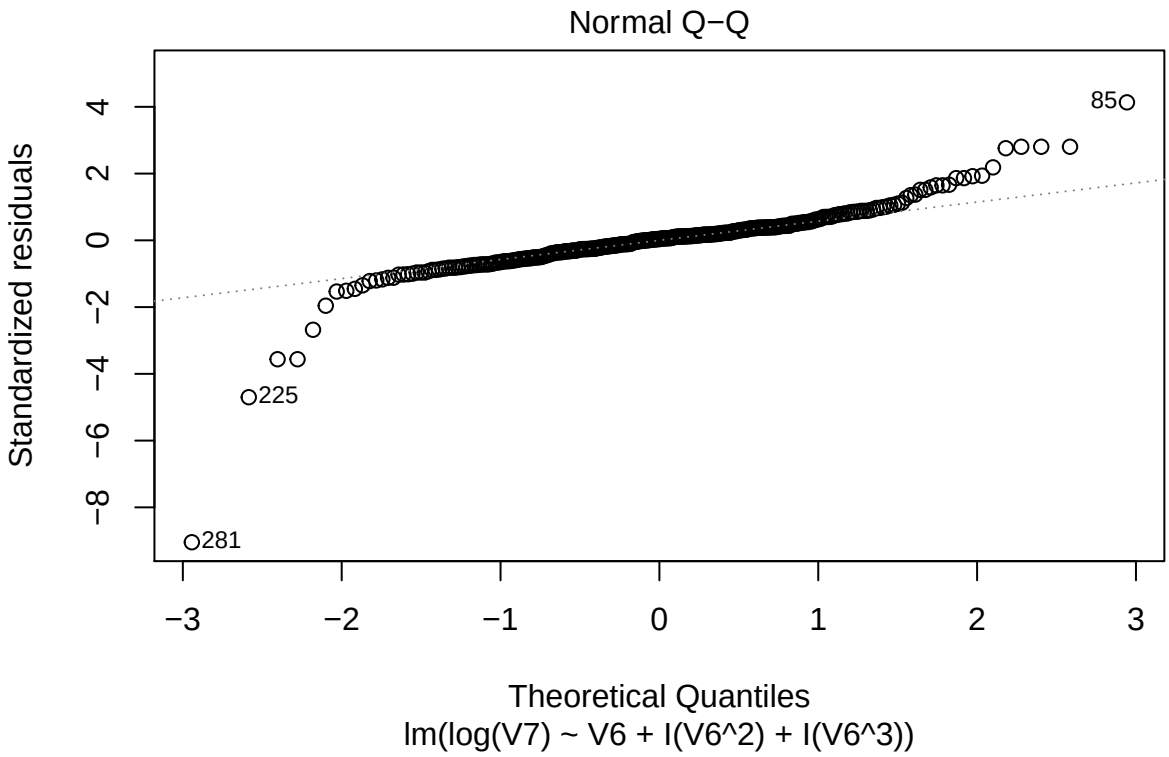


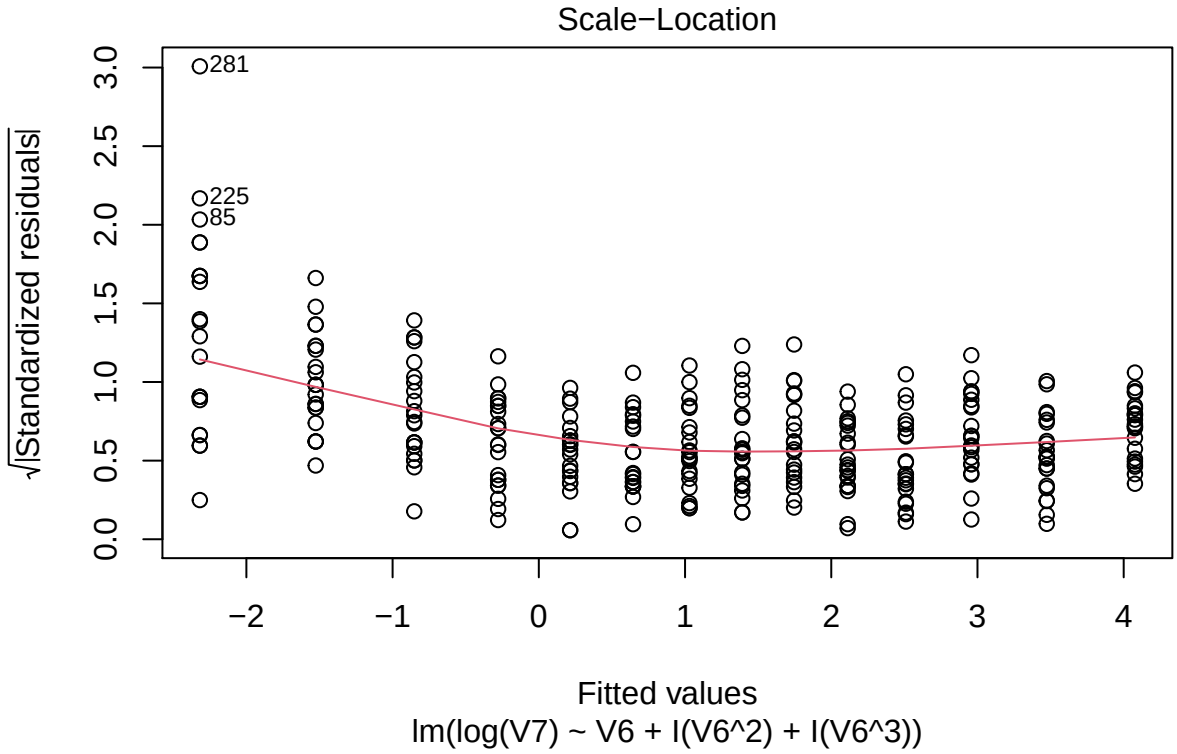

```
plot(V7~V4,data=yacht)
```

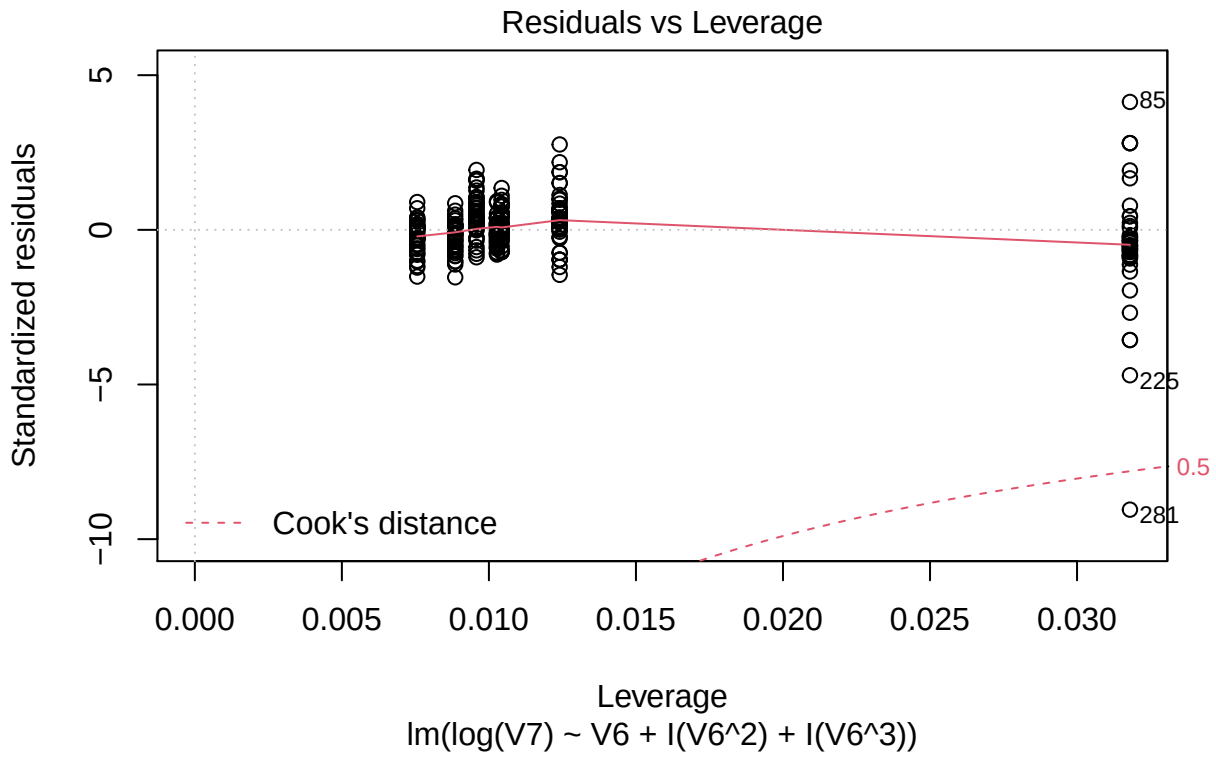



```
plot(lm(log(V7)~V6+I(V6^2)+I(V6^3),data=yacht))
```

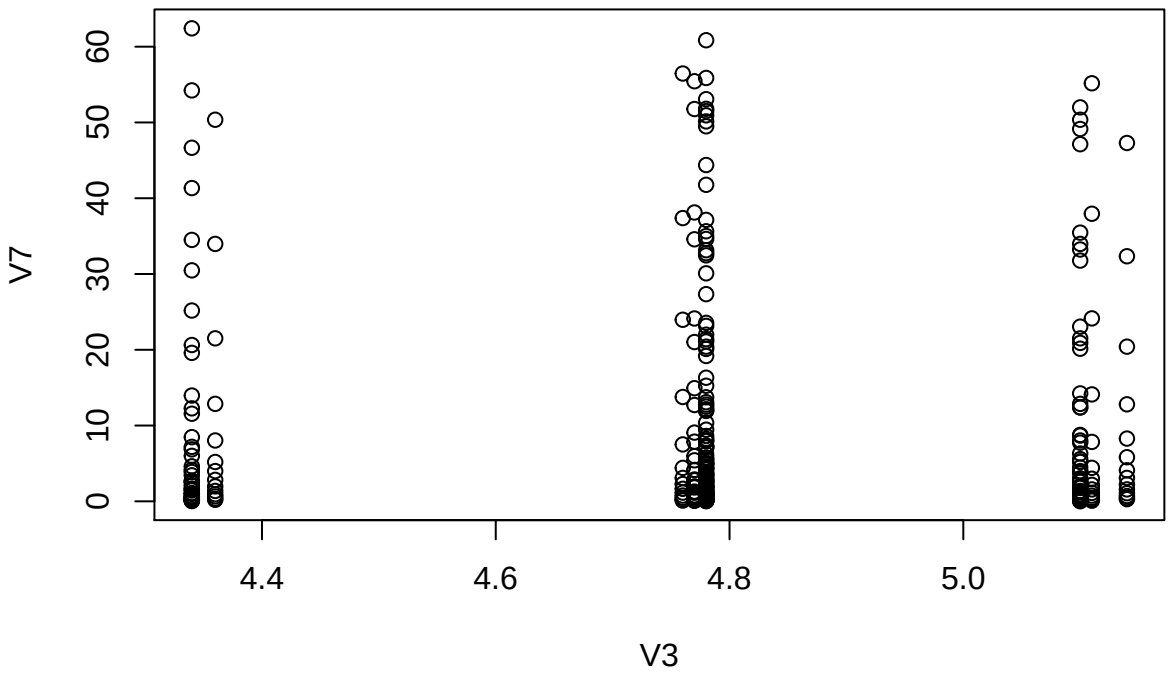

Residuals vs Fitted
Residuals
Fitted values
lm(log(V7) ~ V6 + I(V6^2) + I(V6^3))
85
225
281
−2
−1
0
1
2
3
4


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(log(V7) ~ V6 + I(V6^2) + I(V6^3))
85
225
281
−8
−6
−4
−2
0
2
4
−3
−2
−1
0
1
2
3


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
281
225
85
−2
−1
0
1
2
3
4
Fitted values
lm(log(V7) ~ V6 + I(V6^2) + I(V6^3))




```
plot(V7~V3,data =yacht)
```




```
#RMSE
set.seed(569)
train_ind<-sample.int(nrow(yacht),0.7*nrow(yacht))
train <- yacht[train_ind , ]
test <- yacht[-train_ind , ]
reg_train<-lm(log(V7)~V6+I(V6^2)+I(V6^3),data=train)
reg2_train<-lm(V7~V6+I(V6^2)+I(V6^3),data=train)
reg3_train<-lm(log(V7)~V4+V6+I(V6^2)+I(V6^3),data=train) #best 4 from forward selection and exhaustive
reg4_train<-lm(log(V7)~V3+V6+I(V6^2)+I(V6^3),data=train)#best 4 from backward selection
reg5_train<-lm(log(V7) ~V1+V2+V4+V6+I(V6^2)+I(V6^3), data=train) # 6 from min cp
reg6_train<-lm(log(V7) ~V1+V2+V3+V5+V6+I(V6^2)+I(V6^3), data=train)
reg7_train<-lm(log(V7)~V6+I(V6^2)+I(V6^3)+V3*V4-V3-V4, data=train)
sqrt(mean((test$V7-exp(predict(reg_train,newdata = test)))^2))
```

```
## [1] 3.065069
```

```
sqrt(mean((test$V7-predict(reg2_train,newdata = test))^2))
```

```
## [1] 1.677241
```

```
sqrt(mean((test$V7-exp(predict(reg3_train,newdata = test)))^2))
```

```
## [1] 3.442238
```

```r
sqrt(mean((test$V7-exp(predict(reg4_train,newdata = test)))^2))
```

```
## [1] 3.283484
```

```r
sqrt(mean((test$V7-exp(predict(reg5_train,newdata = test)))^2))
```

```
## [1] 3.271653
```

```r
sqrt(mean((test$V7-exp(predict(reg6_train,newdata = test)))^2))
```

```
## [1] 3.242125
```

```r
sqrt(mean((test$V7-exp(predict(reg7_train,newdata = test)))^2))
```

```
## [1] 3.527683
```

```r
#max adjr2 and min cp
library(leaps)
s <- regsubsets(log(V7) ~V1+V2+V3+V4+V5+V6+I(V6^2)+I(V6^3), data=yacht, method="exhaustive")
ss <- summary(s)

s2 <- regsubsets(log(V7) ~V1+V2+V3+V4+V5+V6+I(V6^2)+I(V6^3), data=yacht, method="forward")
ss2 <- summary(s2)

s3 <- regsubsets(log(V7) ~V1+V2+V3+V4+V5+V6+I(V6^2)+I(V6^3), data=yacht, method="backward")
ss3 <- summary(s3)

s4 <- regsubsets(log(V7) ~V1+V2+V3+V4+V5+V6+I(V6^2)+I(V6^3)+V1*V2+V1*V3+V1*V4+V1*V5+V1*V6+V2*V3+V2*V4*V
ss4 <- summary(s4)
ss4
```

```
## Subset selection object
## Call: regsubsets.formula(log(V7) ~ V1 + V2 + V3 + V4 + V5 + V6 + I(V6^2) +
##     I(V6^3) + V1 * V2 + V1 * V3 + V1 * V4 + V1 * V5 + V1 * V6 +
##     V2 * V3 + V2 * V4 * V2 * V5 + V2 * V6 + V3 * V4 + V3 * V5 +
##     V3 * V6 + V4 * V5 + V4 * V6 + V5 * V6, data = yacht, method = "exhaustive")
## 24 Variables  (and intercept)
##           Forced in Forced out
## V1            FALSE      FALSE
## V2            FALSE      FALSE
## V3            FALSE      FALSE
## V4            FALSE      FALSE
## V5            FALSE      FALSE
## V6            FALSE      FALSE
## I(V6^2)       FALSE      FALSE
## I(V6^3)       FALSE      FALSE
## V1:V2         FALSE      FALSE
## V1:V3         FALSE      FALSE
## V1:V4         FALSE      FALSE
## V1:V5         FALSE      FALSE
```

```
## V1:V6        FALSE      FALSE
## V2:V3        FALSE      FALSE
## V2:V4        FALSE      FALSE
## V2:V5        FALSE      FALSE
## V4:V5        FALSE      FALSE
## V2:V6        FALSE      FALSE
## V3:V4        FALSE      FALSE
## V3:V5        FALSE      FALSE
## V3:V6        FALSE      FALSE
## V4:V6        FALSE      FALSE
## V5:V6        FALSE      FALSE
## V2:V4:V5     FALSE      FALSE
## 1 subsets of each size up to 8
## Selection Algorithm: exhaustive
##          V1  V2  V3  V4  V5  V6  I(V6^2) I(V6^3) V1:V2 V1:V3 V1:V4 V1:V5 V1:V6
## 1  ( 1 ) " " " " " " " " " " "*" " "     " "     " "   " "   " "   " "   " "
## 2  ( 1 ) " " " " " " " " " " "*" "*"     " "     " "   " "   " "   " "   " "
## 3  ( 1 ) " " " " " " " " " " "*" "*"     "*"     " "   " "   " "   " "   " "
## 4  ( 1 ) " " " " " " " " " " "*" "*"     "*"     " "   " "   " "   " "   " "
## 5  ( 1 ) " " " " " " " " " " "*" "*"     "*"     "*"   " "   "*"   " "   " "
## 6  ( 1 ) " " " " " " " " " " "*" "*"     "*"     "*"   " "   "*"   " "   " "
## 7  ( 1 ) " " " " " " "*" " " "*" "*"     "*"     " "   " "   " "   " "   " "
## 8  ( 1 ) " " " " " " "*" " " "*" "*"     "*"     " "   " "   " "   "*"   " "
##          V2:V3 V2:V4 V2:V5 V4:V5 V2:V6 V3:V4 V3:V5 V3:V6 V4:V6 V5:V6 V2:V4:V5
## 1  ( 1 ) " "   " "   " "   " "   " "   " "   " "   " "   " "   " "   " "
## 2  ( 1 ) " "   " "   " "   " "   " "   " "   " "   " "   " "   " "   " "
## 3  ( 1 ) " "   " "   " "   " "   " "   " "   " "   " "   " "   " "   " "
## 4  ( 1 ) " "   " "   " "   " "   " "   "*"   " "   " "   " "   " "   " "
## 5  ( 1 ) " "   " "   " "   " "   " "   " "   " "   " "   " "   " "   " "
## 6  ( 1 ) " "   " "   " "   " "   " "   " "   " "   " "   "*"   " "   " "
## 7  ( 1 ) " "   "*"   " "   " "   "*"   " "   " "   " "   "*"   " "   " "
## 8  ( 1 ) " "   "*"   " "   " "   "*"   " "   " "   " "   "*"   " "   " "
```

```
ss4$adjr2
```

```
## [1] 0.9698118 0.9735425 0.9806705 0.9812965 0.9818202 0.9823177 0.9830878
## [8] 0.9833463
```

```
ss4$cp
```

```
## [1] 257.24244 188.27725  57.01563  46.31700  37.57114  29.36874  16.25690
## [8]  12.53342
```

```
summary(regsubsets(V7 ~V1+V2+V3+V4+V5+V6+I(V6^2)+I(V6^3), data=yacht, method="exhaustive"))$adjr2
```

```
## [1] 0.8623568 0.9600416 0.9834056 0.9841778 0.9844947 0.9844615 0.9845048
## [8] 0.9844777
```

```
summary(regsubsets(V7 ~V1+V2+V3+V4+V5+V6+I(V6^2)+I(V6^3), data=yacht, method="exhaustive"))$cp
```

```
## [1] 2409.433966  483.146928   24.996555   10.853893    5.668224    7.313239
## [7]    7.475490    9.000000
```

```
ss
```

```
## Subset selection object
## Call: regsubsets.formula(log(V7) ~ V1 + V2 + V3 + V4 + V5 + V6 + I(V6^2) +
##     I(V6^3), data = yacht, method = "exhaustive")
## 8 Variables  (and intercept)
##         Forced in Forced out
## V1          FALSE      FALSE
## V2          FALSE      FALSE
## V3          FALSE      FALSE
## V4          FALSE      FALSE
## V5          FALSE      FALSE
## V6          FALSE      FALSE
## I(V6^2)     FALSE      FALSE
## I(V6^3)     FALSE      FALSE
## 1 subsets of each size up to 8
## Selection Algorithm: exhaustive
##          V1  V2  V3  V4  V5  V6  I(V6^2) I(V6^3)
## 1  ( 1 ) " " " " " " " " " " "*" " "     " "
## 2  ( 1 ) " " " " " " " " " " "*" "*"     " "
## 3  ( 1 ) " " " " " " " " " " "*" "*"     "*"
## 4  ( 1 ) " " " " " " "*" " " "*" "*"     "*"
## 5  ( 1 ) " " "*" " " "*" " " "*" "*"     "*"
## 6  ( 1 ) "*" "*" " " "*" " " "*" "*"     "*"
## 7  ( 1 ) "*" "*" "*" " " "*" "*" "*"     "*"
## 8  ( 1 ) "*" "*" "*" "*" "*" "*" "*"     "*"
```

```
ss2
```

```
## Subset selection object
## Call: regsubsets.formula(log(V7) ~ V1 + V2 + V3 + V4 + V5 + V6 + I(V6^2) +
##     I(V6^3), data = yacht, method = "forward")
## 8 Variables  (and intercept)
##         Forced in Forced out
## V1          FALSE      FALSE
## V2          FALSE      FALSE
## V3          FALSE      FALSE
## V4          FALSE      FALSE
## V5          FALSE      FALSE
## V6          FALSE      FALSE
## I(V6^2)     FALSE      FALSE
## I(V6^3)     FALSE      FALSE
## 1 subsets of each size up to 8
## Selection Algorithm: forward
##          V1  V2  V3  V4  V5  V6  I(V6^2) I(V6^3)
## 1  ( 1 ) " " " " " " " " " " "*" " "     " "
## 2  ( 1 ) " " " " " " " " " " "*" "*"     " "
## 3  ( 1 ) " " " " " " " " " " "*" "*"     "*"
## 4  ( 1 ) " " " " " " "*" " " "*" "*"     "*"
## 5  ( 1 ) " " "*" " " "*" " " "*" "*"     "*"
## 6  ( 1 ) "*" "*" " " "*" " " "*" "*"     "*"
## 7  ( 1 ) "*" "*" "*" "*" " " "*" "*"     "*"
## 8  ( 1 ) "*" "*" "*" "*" "*" "*" "*"     "*"
```

```
ss3
```

```
## Subset selection object
## Call: regsubsets.formula(log(V7) ~ V1 + V2 + V3 + V4 + V5 + V6 + I(V6^2) +
##     I(V6^3), data = yacht, method = "backward")
## 8 Variables  (and intercept)
##         Forced in Forced out
## V1          FALSE      FALSE
## V2          FALSE      FALSE
## V3          FALSE      FALSE
## V4          FALSE      FALSE
## V5          FALSE      FALSE
## V6          FALSE      FALSE
## I(V6^2)     FALSE      FALSE
## I(V6^3)     FALSE      FALSE
## 1 subsets of each size up to 8
## Selection Algorithm: backward
##          V1  V2  V3  V4  V5  V6  I(V6^2) I(V6^3)
## 1  ( 1 ) " " " " " " " " " " "*" " "     " "
## 2  ( 1 ) " " " " " " " " " " "*" "*"     " "
## 3  ( 1 ) " " " " " " " " " " "*" "*"     "*"
## 4  ( 1 ) " " " " "*" " " " " "*" "*"     "*"
## 5  ( 1 ) " " " " "*" " " "*" "*" "*"     "*"
## 6  ( 1 ) "*" " " "*" " " "*" "*" "*"     "*"
## 7  ( 1 ) "*" "*" "*" " " "*" "*" "*"     "*"
## 8  ( 1 ) "*" "*" "*" "*" "*" "*" "*"     "*"
```

```
ss$adjr2
```

```
## [1] 0.9698118 0.9735425 0.9806705 0.9812589 0.9816951 0.9819435 0.9819566
## [8] 0.9819076
```

```
ss2$adjr2
```

```
## [1] 0.9698118 0.9735425 0.9806705 0.9812589 0.9816951 0.9819435 0.9818843
## [8] 0.9819076
```

```
ss3$adjr2
```

```
## [1] 0.9698118 0.9735425 0.9806705 0.9808176 0.9815708 0.9818288 0.9819566
## [8] 0.9819076
```

```
ss$cp
```

```
## [1] 206.578918 144.019780  24.787724  15.864224   9.547201   6.402748   7.187637
## [8]   9.000000
```

```
ss2$cp
```

```
## [1] 206.578918 144.019780  24.787724  15.864224   9.547201   6.402748   8.387351
## [8]   9.000000
```

```r
ss3$cp
```

```
## [1] 206.578918 144.019780  24.787724  23.255666  11.622507   8.312461   7.187637
## [8]   9.000000
```

```r
#AIC
AIC(lm(log(V7) ~V1+V2+V4+V6+I(V6^2)+I(V6^3), data=yacht)) #6 selected min cp
```

```
## [1] 24.9129
```

```r
AIC(lm(log(V7) ~V1+V2+V3+V5+V6+I(V6^2)+I(V6^3), data=yacht)) #7 max adjr2
```

```
## [1] 25.66454
```

```r
AIC(lm(log(V7)~V6+I(V6^2)+I(V6^3),data=yacht)) #best 3
```

```
## [1] 42.95161
```

```r
AIC(lm(V7~V6+I(V6^2)+I(V6^3),data=yacht))#best 3 without log
```

```
## [1] 1292.359
```

```r
AIC(lm(log(V7)~V4+V6+I(V6^2)+I(V6^3),data=yacht))#best 4 from forward selection and exhaustive selectio
```

```
## [1] 34.41452
```

```r
AIC(lm(log(V7)~V3+V6+I(V6^2)+I(V6^3),data=yacht))#best 4 from backward selection
```

```
## [1] 41.58377
```

```r
AIC(lm(log(V7)~V6+I(V6^2)+I(V6^3)+V3*V4-V3-V4,data=yacht))#with interaction from exhaustive
```

```
## [1] 33.79683
```

```r
#summary
summary(lm(log(V7) ~V1+V2+V4+V6+I(V6^2)+I(V6^3), data=yacht)) #6 min cp
```

```
##
## Call:
## lm(formula = log(V7) ~ V1 + V2 + V4 + V6 + I(V6^2) + I(V6^3),
##     data = yacht)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.24833 -0.10301 -0.00105  0.10969  0.92605
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.062e+00  4.960e-01 -16.252  < 2e-16 ***
## V1           2.127e-02  9.367e-03   2.270  0.02389 *
## V2          -1.853e+00  6.470e-01  -2.863  0.00448 **
## V4           1.121e-01  2.749e-02   4.078 5.83e-05 ***
## V6           7.116e+01  4.261e+00  16.700  < 2e-16 ***
## I(V6^2)     -1.848e+02  1.570e+01 -11.770  < 2e-16 ***
## I(V6^3)      1.996e+02  1.811e+01  11.021  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2483 on 301 degrees of freedom
## Multiple R-squared:  0.9823, Adjusted R-squared:  0.9819
## F-statistic:  2784 on 6 and 301 DF,  p-value: < 2.2e-16
```

```
summary(lm(log(V7) ~V1+V2+V3+V5+V6+I(V6^2)+I(V6^3), data=yacht)) #7 selected by by adjr2
```

```
##
## Call:
## lm(formula = log(V7) ~ V1 + V2 + V3 + V5 + V6 + I(V6^2) + I(V6^3),
##     data = yacht)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.25263 -0.09748 -0.00243  0.10401  0.92568
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.530e+00  5.721e-01 -14.910  < 2e-16 ***
## V1           2.146e-02  9.363e-03   2.292 0.022622 *
## V2          -1.081e+00  6.107e-01  -1.770 0.077724 .
## V3           2.978e-01  7.602e-02   3.917 0.000111 ***
## V5          -2.968e-01  7.778e-02  -3.815 0.000165 ***
## V6           7.116e+01  4.260e+00  16.707  < 2e-16 ***
## I(V6^2)     -1.848e+02  1.569e+01 -11.775  < 2e-16 ***
## I(V6^3)      1.996e+02  1.810e+01  11.025  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2482 on 300 degrees of freedom
## Multiple R-squared:  0.9824, Adjusted R-squared:  0.982
## F-statistic:  2388 on 7 and 300 DF,  p-value: < 2.2e-16
```

```
summary(lm(log(V7)~V6+I(V6^2)+I(V6^3),data=yacht)) #best 3
```

```
##
## Call:
## lm(formula = log(V7) ~ V6 + I(V6^2) + I(V6^3), data = yacht)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.28690 -0.09902  0.01018  0.09868  1.04531
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -8.7162     0.3695  -23.59   <2e-16 ***
## V6            71.1627     4.4088   16.14   <2e-16 ***
## I(V6^2)     -184.7861    16.2431  -11.38   <2e-16 ***
## I(V6^3)      199.6060    18.7382   10.65   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2569 on 304 degrees of freedom
## Multiple R-squared:  0.9809, Adjusted R-squared:  0.9807
## F-statistic:  5193 on 3 and 304 DF,  p-value: < 2.2e-16
```

```
summary(lm(log(V7)~V4+V6+I(V6^2)+I(V6^3),data=yacht))#best 4 from bs
```

```
##
## Call:
## lm(formula = log(V7) ~ V4 + V6 + I(V6^2) + I(V6^3), data = yacht)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.30684 -0.08750 -0.00826  0.09956  0.95864
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -9.05292    0.37837 -23.926   <2e-16 ***
## V4             0.08554    0.02634   3.247   0.0013 **
## V6            71.16266    4.34115  16.393   <2e-16 ***
## I(V6^2)     -184.78610   15.99395 -11.553   <2e-16 ***
## I(V6^3)      199.60604   18.45077  10.818   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.253 on 303 degrees of freedom
## Multiple R-squared:  0.9815, Adjusted R-squared:  0.9813
## F-statistic:  4020 on 4 and 303 DF,  p-value: < 2.2e-16
```

```
summary(lm(V7~V6+I(V6^2)+I(V6^3),data=yacht))#best 3 without log
```

```
##
## Call:
## lm(formula = V7 ~ V6 + I(V6^2) + I(V6^3), data = yacht)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.308 -1.210 -0.306  1.296 11.732
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -48.029      2.809  -17.10   <2e-16 ***
## V6            695.227     33.510   20.75   <2e-16 ***
## I(V6^2)     -3161.265    123.461  -25.61   <2e-16 ***
## I(V6^3)      4675.126    142.426   32.83   <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.953 on 304 degrees of freedom
## Multiple R-squared:  0.9836, Adjusted R-squared:  0.9834
## F-statistic:  6065 on 3 and 304 DF,  p-value: < 2.2e-16
```

```
summary(lm(log(V7)~V3+V6+I(V6^2)+I(V6^3),data=yacht)) #best 4 from fs and es
```

```
##
## Call:
## lm(formula = log(V7) ~ V3 + V6 + I(V6^2) + I(V6^3), data = yacht)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3197 -0.1014  0.0019  0.0997  1.0083
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -9.22073    0.46037 -20.029   <2e-16 ***
## V3             0.10536    0.05773   1.825    0.069 .
## V6            71.16266    4.39197  16.203   <2e-16 ***
## I(V6^2)     -184.78610   16.18118 -11.420   <2e-16 ***
## I(V6^3)      199.60604   18.66676  10.693   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.256 on 303 degrees of freedom
## Multiple R-squared:  0.9811, Adjusted R-squared:  0.9808
## F-statistic:  3925 on 4 and 303 DF,  p-value: < 2.2e-16
```

```
#prediction error

#best 3 without log
train <- 1:as.integer(dim(yacht)[1]/2)
reg1 <- lm(V7~V6+I(V6^2)+I(V6^3), data=yacht[train,])
error1 <- sum((yacht$V7[-train] - predict(reg1, yacht[-train,]))^2)
reg2 <- lm(V7~V6+I(V6^2)+I(V6^3), data=yacht[-train,])
error2 <- sum((yacht$V7[train] - predict(reg2, yacht[train,]))^2)
error <- (error1 + error2)/dim(yacht)[1]
error
```

```
## [1] 4.546575
```

```
#best 3
train <- 1:as.integer(dim(yacht)[1]/2)
reg1 <- lm(log(V7)~V6+I(V6^2)+I(V6^3), data=yacht[train,])
error1 <- sum((yacht$V7[-train] - exp(predict(reg1, yacht[-train,])))^2)
reg2 <- lm(log(V7)~V6+I(V6^2)+I(V6^3), data=yacht[-train,])
error2 <- sum((yacht$V7[train] - exp(predict(reg2, yacht[train,])))^2)
error <- (error1 + error2)/dim(yacht)[1]
error
```

```
## [1] 8.293667
```

```
#best4 from exhaustive and forward
train <- 1:as.integer(dim(yacht)[1]/2)
reg1 <- lm(log(V7)~V4+V6+I(V6^2)+I(V6^3), data=yacht[train,])
error1 <- sum((yacht$V7[-train] - exp(predict(reg1, yacht[-train,])))^2)
reg2 <- lm(log(V7)~V4+V6+I(V6^2)+I(V6^3), data=yacht[-train,])
error2 <- sum((yacht$V7[train] - exp(predict(reg2, yacht[train,])))^2)
error <- (error1 + error2)/dim(yacht)[1]
error
```

```
## [1] 9.690707
```

```
#best4 from backward
train <- 1:as.integer(dim(yacht)[1]/2)
reg1 <- lm(log(V7)~V3+V6+I(V6^2)+I(V6^3), data=yacht[train,])
error1 <- sum((yacht$V7[-train] - exp(predict(reg1, yacht[-train,])))^2)
reg2 <- lm(log(V7)~V3+V6+I(V6^2)+I(V6^3), data=yacht[-train,])
error2 <- sum((yacht$V7[train] - exp(predict(reg2, yacht[train,])))^2)
error <- (error1 + error2)/dim(yacht)[1]
error
```

```
## [1] 8.879072
```

```
#6 min cp
train <- 1:as.integer(dim(yacht)[1]/2)
reg1 <- lm(log(V7) ~V1+V2+V4+V6+I(V6^2)+I(V6^3), data=yacht[train,])
error1 <- sum((yacht$V7[-train] - exp(predict(reg1, yacht[-train,])))^2)
reg2 <- lm(log(V7) ~V1+V2+V4+V6+I(V6^2)+I(V6^3), data=yacht[-train,])
error2 <- sum((yacht$V7[train] - exp(predict(reg2, yacht[train,])))^2)
error <- (error1 + error2)/dim(yacht)[1]
error
```

```
## [1] 22.41788
```

```
#7 max adjr2
train <- 1:as.integer(dim(yacht)[1]/2)
reg1 <- lm(log(V7) ~V1+V2+V3+V5+V6+I(V6^2)+I(V6^3), data=yacht[train,])
error1 <- sum((yacht$V7[-train] - exp(predict(reg1, yacht[-train,])))^2)
reg2 <- lm(log(V7) ~V1+V2+V3+V5+V6+I(V6^2)+I(V6^3), data=yacht[-train,])
error2 <- sum((yacht$V7[train] - exp(predict(reg2, yacht[train,])))^2)
error <- (error1 + error2)/dim(yacht)[1]
error
```

```
## [1] 20.04742
```

```
#with interaction from exhaustive selection
train <- 1:as.integer(dim(yacht)[1]/2)
reg1 <- lm(log(V7) ~V6+I(V6^2)+I(V6^3)+V3*V4-V3-V4, data=yacht[train,])
error1 <- sum((yacht$V7[-train] - exp(predict(reg1, yacht[-train,])))^2)
reg2 <- lm(log(V7) ~V6+I(V6^2)+I(V6^3)+V3*V4-V3-V4, data=yacht[-train,])
error2 <- sum((yacht$V7[train] - exp(predict(reg2, yacht[train,])))^2)
error <- (error1 + error2)/dim(yacht)[1]
error
```

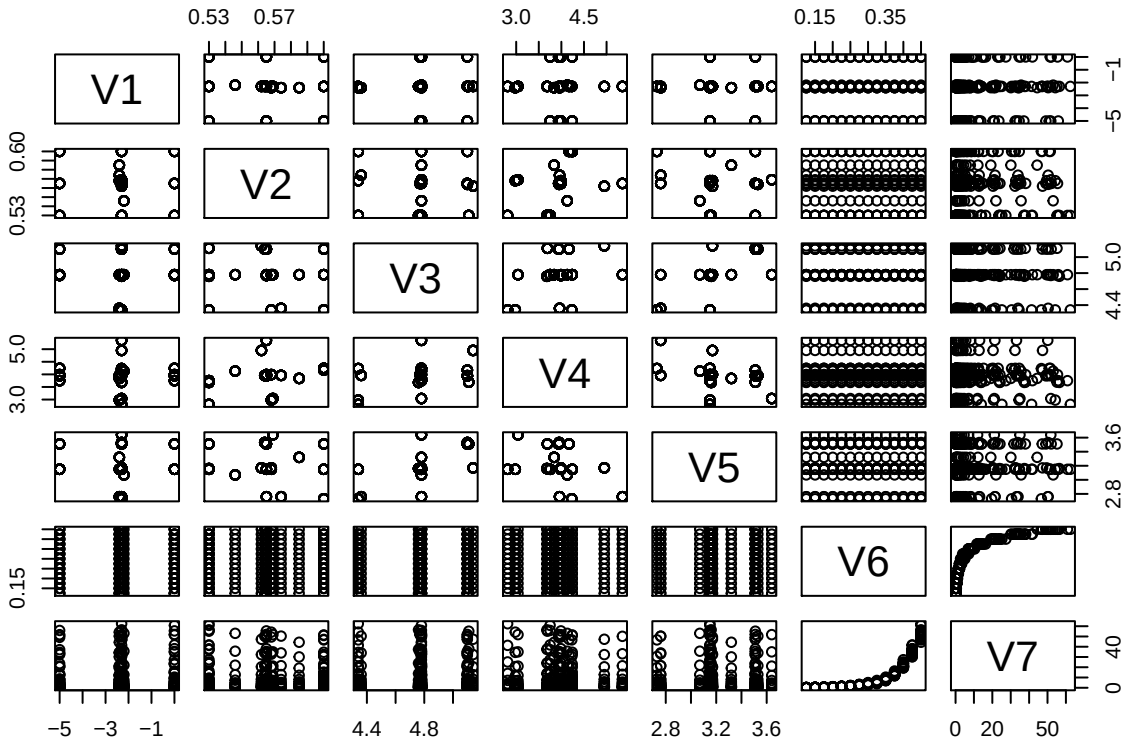
```
## [1] 9.895813
```

```
adjr2<-c(0.9806705,0.9834056,0.9812589,0.9808176,0.9819435,0.9819566,0.9812965)
cp<-c(24.787724,24.996555,9.547201,11.622507,6.402748,7.187637,46.31700)
AIC_value<-c(42.95161,1292.359,34.41452,41.58377,24.9129,25.66454,33.79683)
pred_error <- c(8.293667,4.546575,9.690707,8.879072,22.41788,20.04742,9.895813)
RMSE_value<-c(3.065069,1.677241,3.442238,3.283484,3.271653,3.242125,3.527683)
models<-c("log(V7)~V6+I(V6^2)+I(V6^3)","V7~V6+I(V6^2)+I(V6^3)","log(V7)~V4+V6+I(V6^2)+I(V6^3)","log(V7)
df <- data.frame(adjr2,cp,AIC_value,pred_error,RMSE_value,models)
df2<-data.frame(df, row.names = c("best3","best3_no_log","best4_forwad_exhaustive","best4_backward","wi
write.csv(df2,"/Users/wangning/Desktop/table.csv")
library(psych)
```

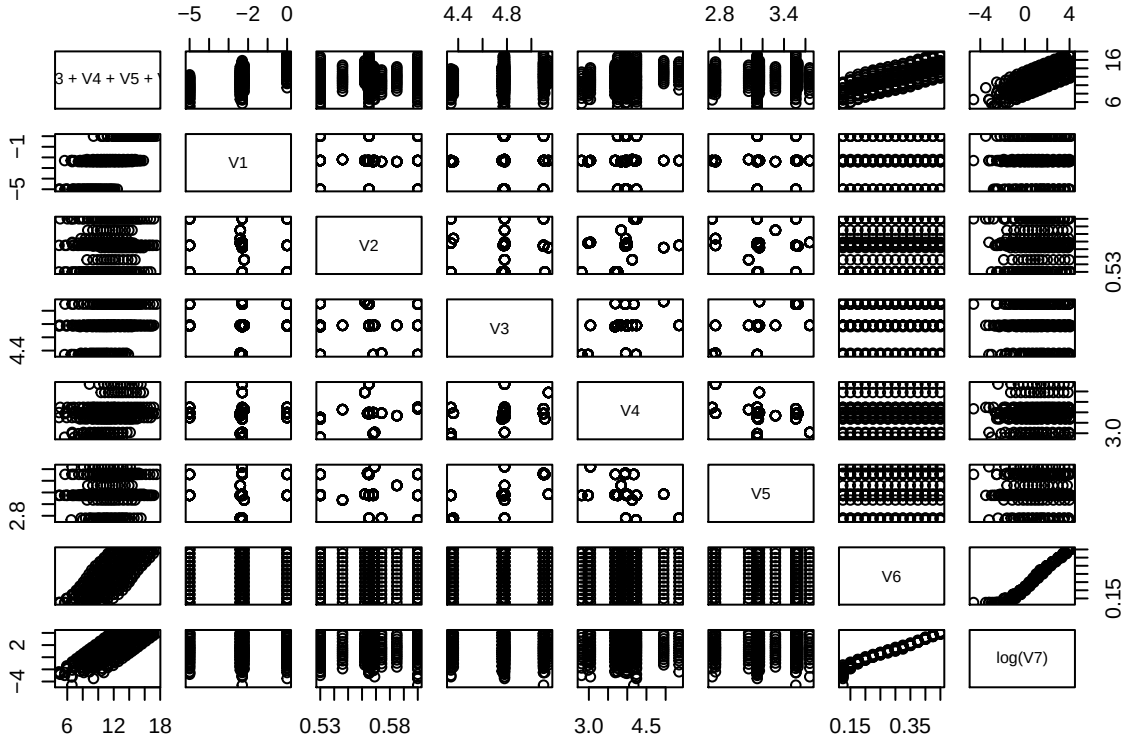
```
##
## Attaching package: 'psych'
```

```
## The following object is masked from 'package:randomForest':
##
##     outlier
```

```
pairs(~.,data=yacht)
```



```
pairs(V1+V2+V3+V4+V5+V6+log(V7)~V1+V2+V3+V4+V5+V6+log(V7),data=yacht)
```




```
summary(lm(log(V7)~V6+I(V6^2)+I(V6^3),data=yacht[-281,]))
```

```
##
## Call:
## lm(formula = log(V7) ~ V6 + I(V6^2) + I(V6^3), data = yacht[-281,
##     ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26339 -0.10120  0.01010  0.09245  0.97020
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -8.2497     0.3195  -25.82   <2e-16 ***
## V6            66.2877     3.8034   17.43   <2e-16 ***
## I(V6^2)     -168.7266    13.9920  -12.06   <2e-16 ***
## I(V6^3)      182.7646    16.1247   11.33   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.22 on 303 degrees of freedom
## Multiple R-squared:  0.9856, Adjusted R-squared:  0.9854
## F-statistic:  6895 on 3 and 303 DF,  p-value: < 2.2e-16
```

```r
summary(lm(log(V7)~V6+I(V6^2)+I(V6^3),data=yacht))
```

```
##
## Call:
## lm(formula = log(V7) ~ V6 + I(V6^2) + I(V6^3), data = yacht)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.28690 -0.09902  0.01018  0.09868  1.04531
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -8.7162     0.3695  -23.59   <2e-16 ***
## V6            71.1627     4.4088   16.14   <2e-16 ***
## I(V6^2)     -184.7861    16.2431  -11.38   <2e-16 ***
## I(V6^3)      199.6060    18.7382   10.65   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2569 on 304 degrees of freedom
## Multiple R-squared:  0.9809, Adjusted R-squared:  0.9807
## F-statistic:  5193 on 3 and 304 DF,  p-value: < 2.2e-16
```