

anova_splittimes.R

poojasingh

Wed Nov 25 13:05:44 2015

```
#Read Marathon data dile
#fname = file.choose()
fname = "/Users/poojasingh/Documents/HStatE139/git/statse139-project2/statse139-project/Previous Boston
dfm = read.csv(fname, header=T,sep=" ")
names(dfm)

## [1] "BibNum"      "Year"        "Age"         "Gender1F2M"   "StartHr"
## [6] "StartMin"    "K0.5"        "K5.10"       "K10.15"      "K15.20"
## [11] "K20.25"      "K25.30"      "K30.35"      "K35.40"      "K40.Fin"
## [16] "HalfMar"     "X"           "Age2014"

n= length(dfm)
n

## [1] 18

#Add all times
times = as.matrix(dfm[,7:15], ncol=9)
times.20.40 = times = as.matrix(dfm[,11:15], ncol=9)
dfm$totaltime = rowSums(times)
dfm$cummttime.20.40 = rowSums(times)
attach(dfm)

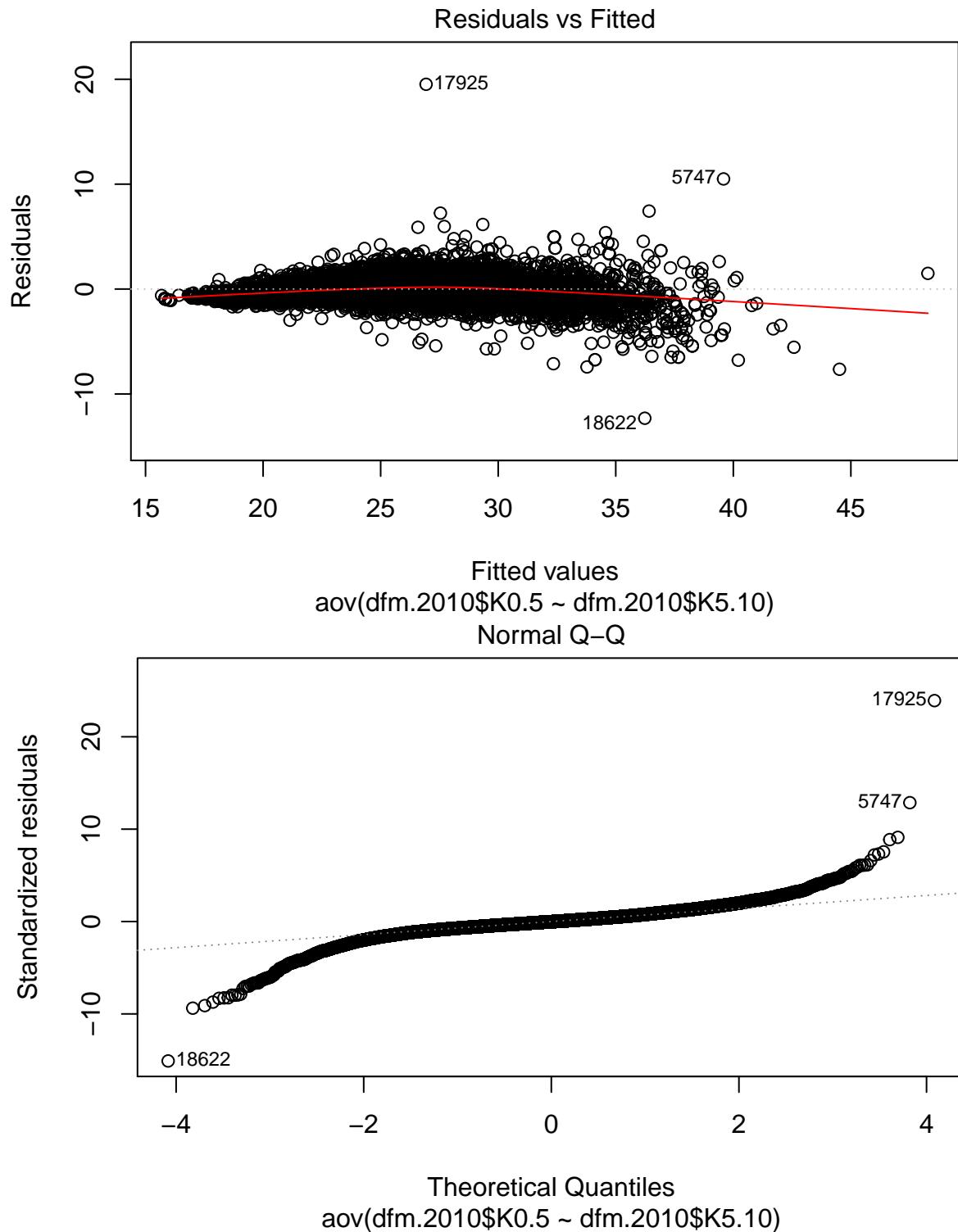
#Split data by year
dfm.2010 = subset(dfm, (dfm$Year==2010))
length(dfm.2010$Year) ## of rows for 2010

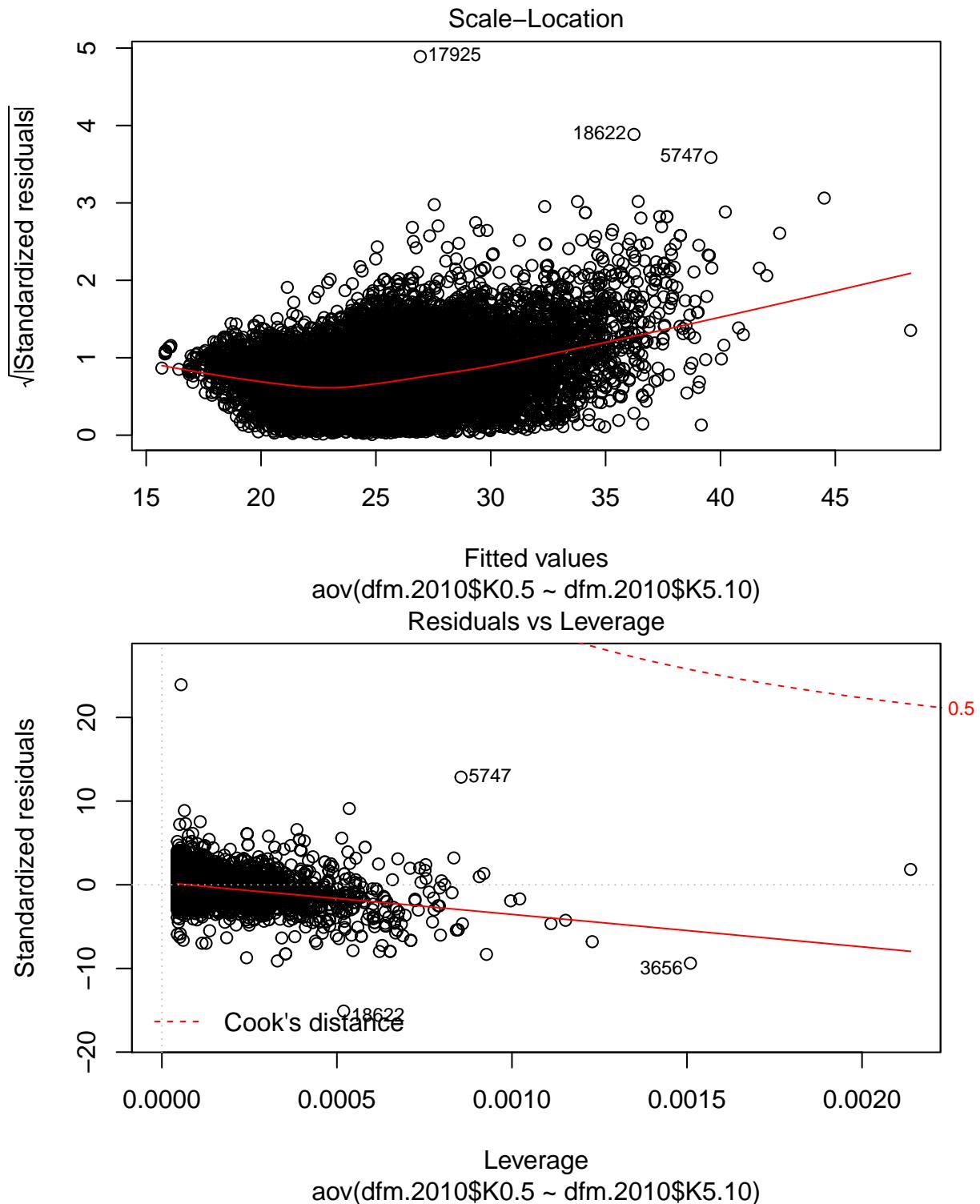
## [1] 22670

#Fit AOV model - varitions between two split times
model2 = aov(dfm.2010$K0.5~dfm.2010$K5.10)
model2

## Call:
## aov(formula = dfm.2010$K0.5 ~ dfm.2010$K5.10)
##
## Terms:
##          dfm.2010$K5.10 Residuals
## Sum of Squares    252466.92 15100.48
## Deg. of Freedom          1      22668
##
## Residual standard error: 0.8161854
## Estimated effects may be unbalanced
```

```
plot(model2)
```





```
#Define matrix with column1 representing different groups of split times and column 2 representing dist
m = matrix(nrow = 5, ncol = 2)
c1 = c("K0.5", "K5.10", "K10.15", "K15.20", "K20.25", "K25.30", "K30.35", "K35.40", "K40.Fin")
c2 = c(dfm.2010\$K0.5, dfm.2010\$K5.10, dfm.2010\$K10.15, dfm.2010\$K15.20, dfm.2010\$K20.25, dfm.2010\$K25.30)
m=cbind(c1,c2)
head(m)
```

```

##      c1      c2
## [1,] "K0.5"  "25.4"
## [2,] "K5.10" "26.32"
## [3,] "K10.15" "28.93"
## [4,] "K15.20" "26.72"
## [5,] "K20.25" "25.2"
## [6,] "K25.30" "25.92"

#Fit AOV model - variation across all split times
model1 = aov(m[,2]~m[,1])
model1

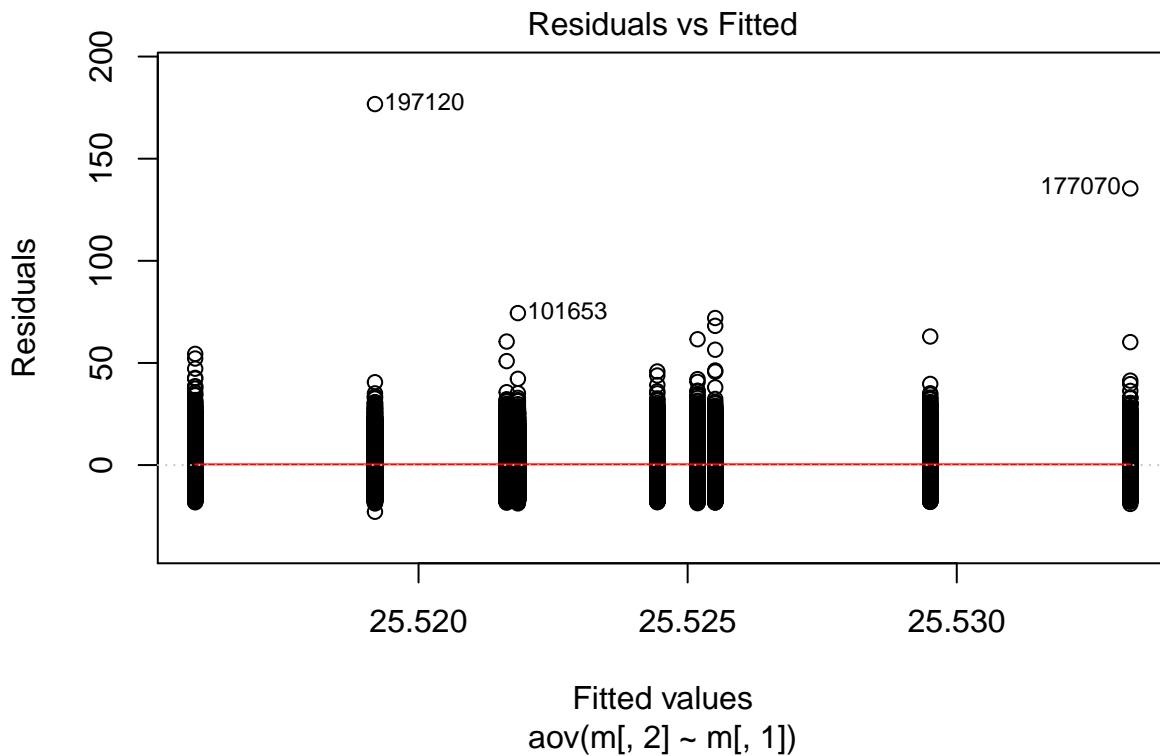
```

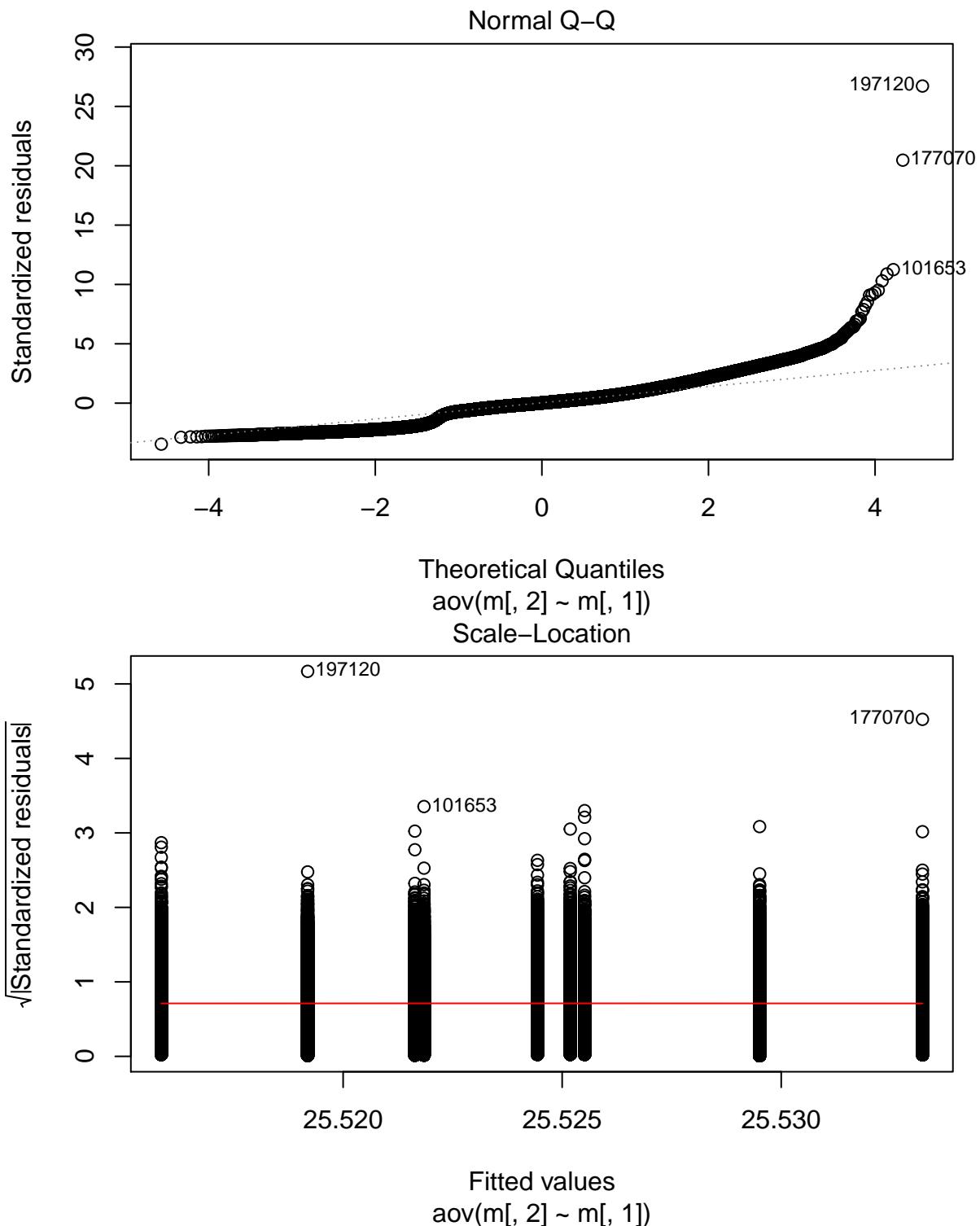
```

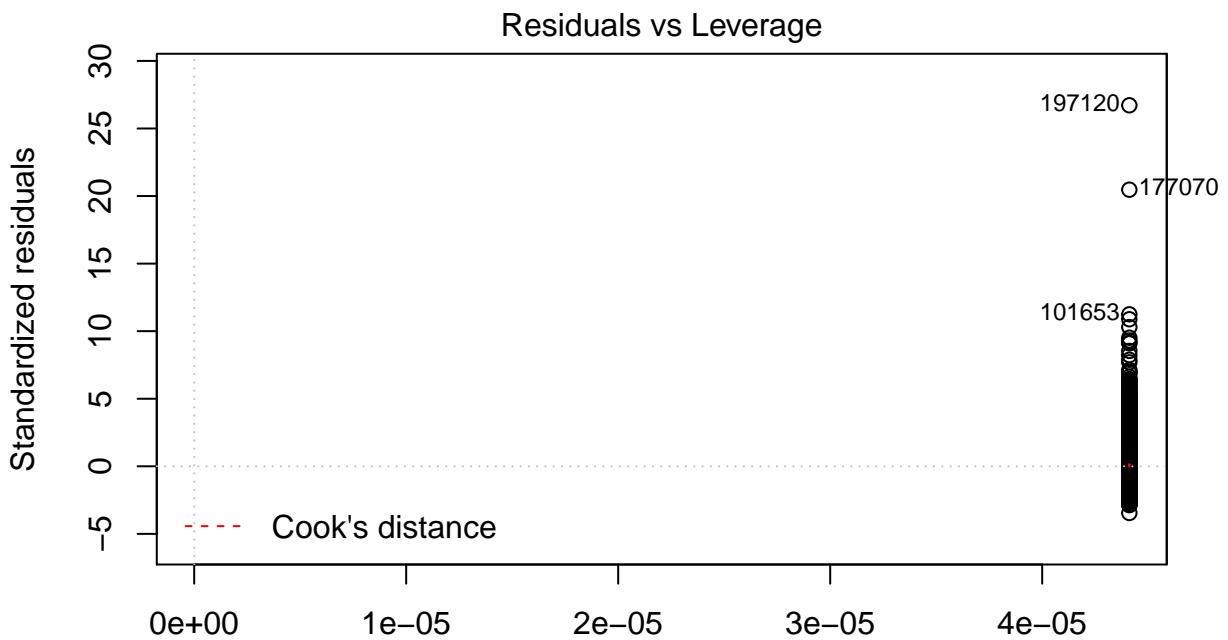
## Call:
##   aov(formula = m[, 2] ~ m[, 1])
##
## Terms:
##           m[, 1] Residuals
## Sum of Squares    5  8928288
## Deg. of Freedom     8   204021
##
## Residual standard error: 6.615256
## Estimated effects may be unbalanced

```

```
plot(model1)
```







```
#Fit ANOVA model - using the recommendation in the article
#TotalTime ~ mean of response + half.mar.time + scaled-time(20-40/half.mar.time)

#Calculate mean of the totaltime
dfm.2010.totaltime.mean = mean(dfm.2010$totaltime)
dfm.2010.totaltime.mean

## [1] 126.8356

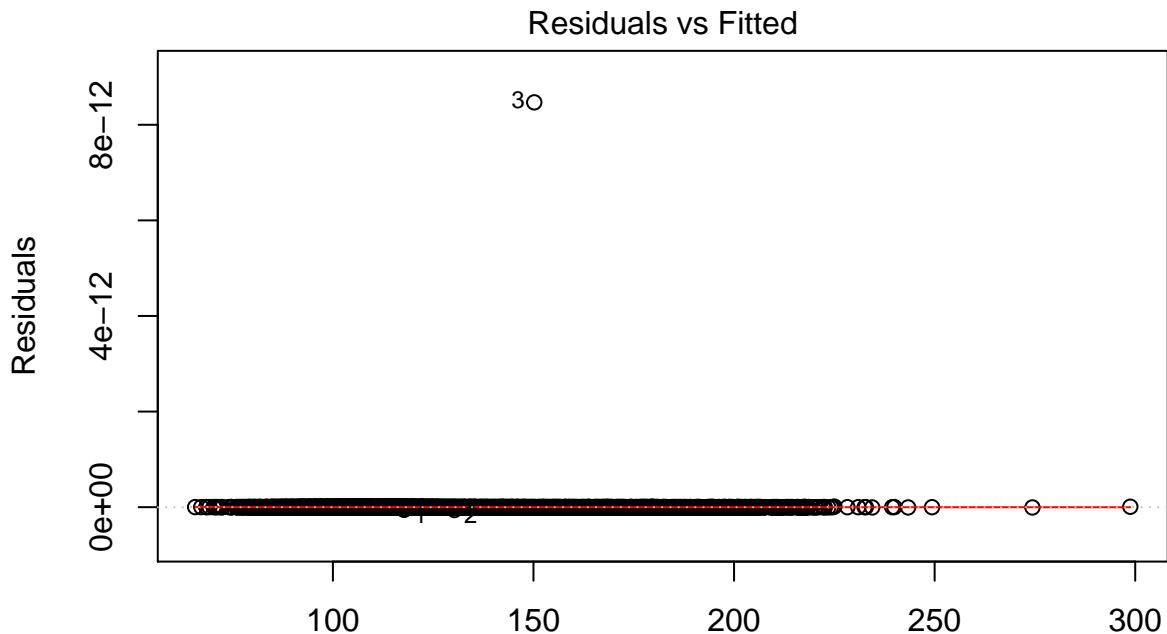
#Add the mean to half marathon
dfm.mean.plus.halfmar = dfm.2010.totaltime.mean + dfm.2010$HalfMar

#AOV model
modelr = aov(dfm.2010$totaltime ~ dfm.mean.plus.halfmar + (dfm.2010$cummttime.20.40/dfm.2010$HalfMar))
modelr

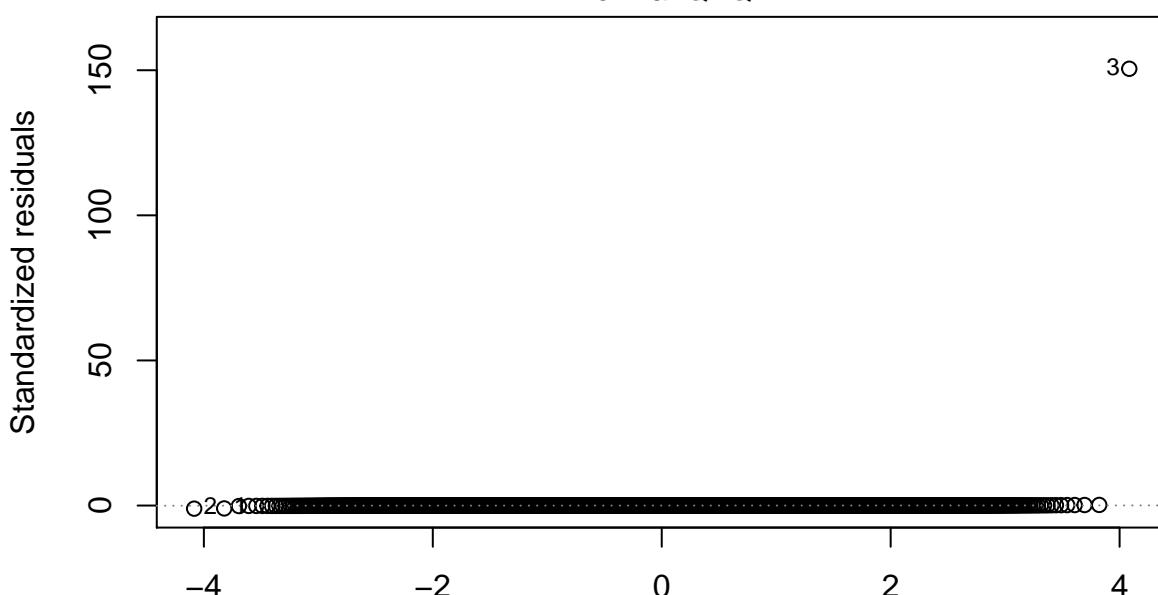
## Call:
##   aov(formula = dfm.2010$totaltime ~ dfm.mean.plus.halfmar + (dfm.2010$cummttime.20.40/dfm.2010$HalfMar))
## 
## Terms:
##   dfm.mean.plus.halfmar dfm.2010$cummttime.20.40
##   Sum of Squares      9606472          2868009
##   Deg. of Freedom       1                  1
##   dfm.2010$cummttime.20.40:dfm.2010$HalfMar Residuals
##   Sum of Squares          0                  0
##   Deg. of Freedom         1                22659
## 
## Residual standard error: 5.62714e-14
## Estimated effects may be unbalanced
```

```
## 7 observations deleted due to missingness
```

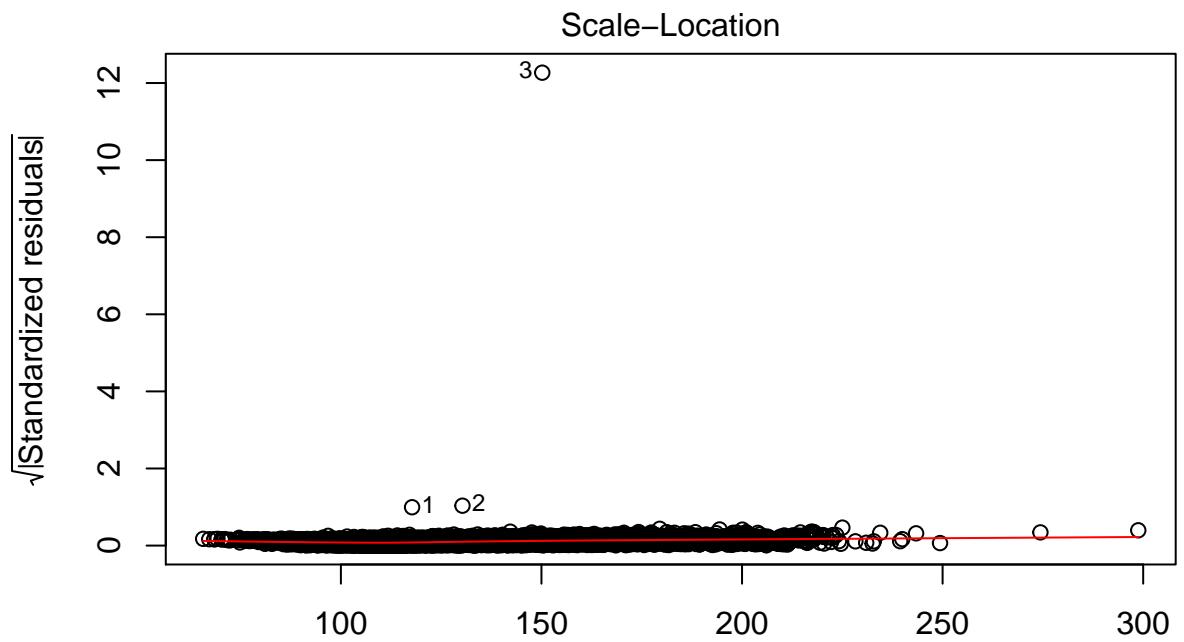
```
plot(modelr)
```



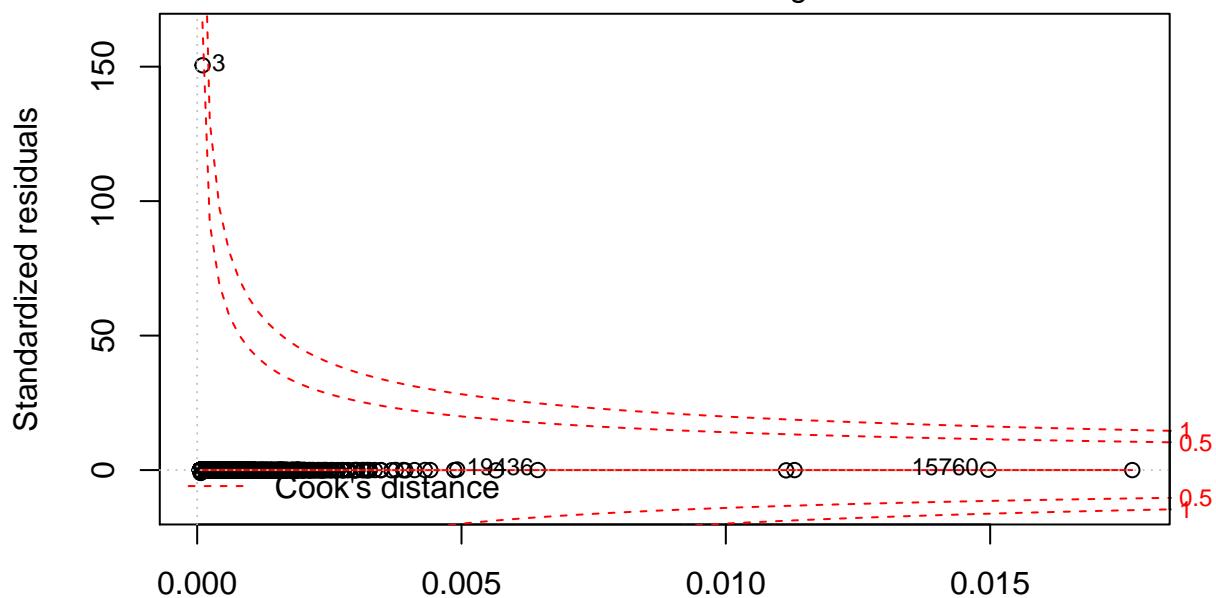
Fitted values
aov(dfm.2010\$totaltime ~ dfm.mean.plus.halfmar + (dfm.2010\$cummttime.20.40/d ..
Normal Q–Q



Theoretical Quantiles
aov(dfm.2010\$totaltime ~ dfm.mean.plus.halfmar + (dfm.2010\$cummttime.20.40/d ..



```
aov(dfm.2010$totaltime ~ dfm.mean.plus.halfmar + (dfm.2010$cummtime.20.40/d ..  
Residuals vs Leverage
```



```
aov(dfm.2010$totaltime ~ dfm.mean.plus.halfmar + (dfm.2010$cummtime.20.40/d ..
```