LungCapData

Amira Ibrahim

October 2, 2019

```
LungCapData <- read.delim(file.choose(),header = TRUE)
attach(LungCapData)
```

check names

```
names(LungCapData)
```

```
## [1] "LungCap"   "Age"       "Height"    "Smoke"     "Gender"    "Caesarean"
```

summary of data :

```
summary(LungCapData)
```

```
##     LungCap            Age            Height      Smoke        Gender
##  Min.   : 0.507   Min.   : 3.00   Min.   :45.30   no :648   female:358
##  1st Qu.: 6.150   1st Qu.: 9.00   1st Qu.:59.90   yes: 77   male  :367
##  Median : 8.000   Median :13.00   Median :65.40
##  Mean   : 7.863   Mean   :12.33   Mean   :64.84
##  3rd Qu.: 9.800   3rd Qu.:15.00   3rd Qu.:70.30
##  Max.   :14.675   Max.   :19.00   Max.   :81.80
##  Caesarean
##  no :561
##  yes:164
##
##
##
##
```

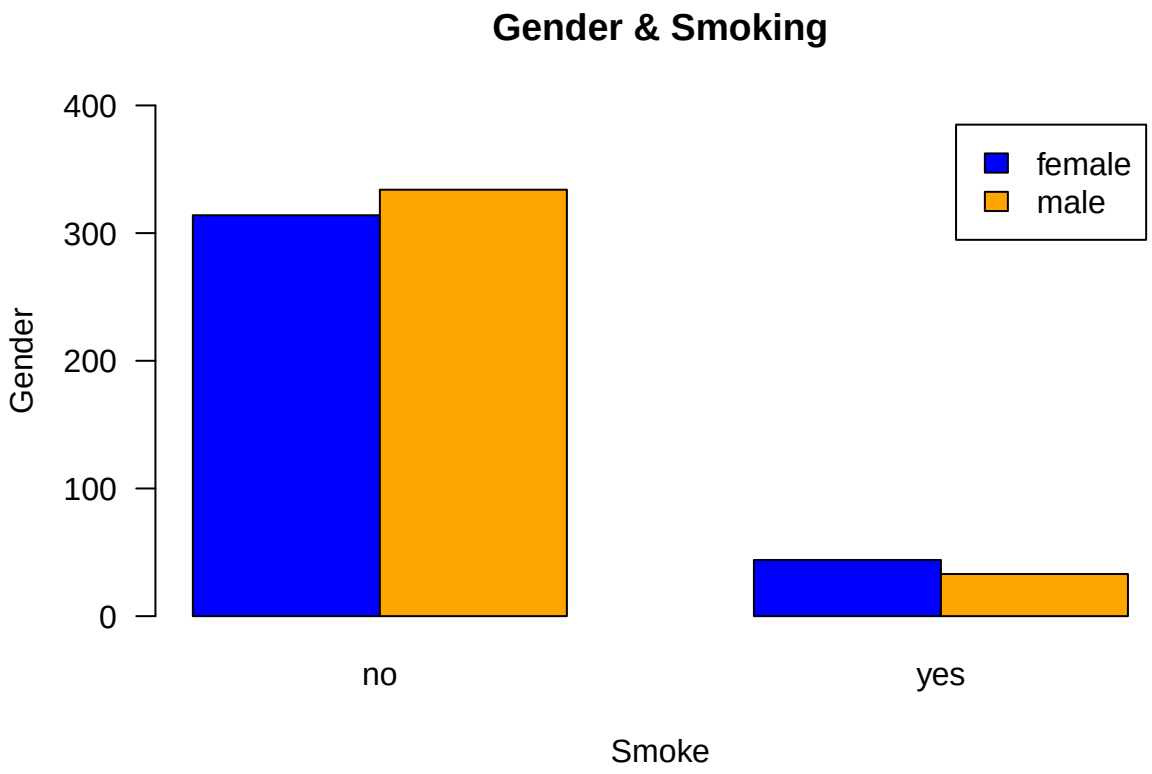
relation between Gender and Smoke :

```
Table1 <- table(Gender ,Smoke)
Table1
```

```
##         Smoke
## Gender    no yes
##   female 314  44
##   male   334  33
```

Table1 : Gender according to Smoking status

```
barplot(Table1 , beside = TRUE , legend=TRUE ,xlab = "Smoke" , ylab = "Gender" ,
        main = "Gender & Smoking" ,ylim = c(0,400),col = c("blue" , "orange"),las=1 )
```



categorical variables by chisq test :

H0 : No relation between smoking frequency and gender

```
chisq.test(Table1 , correct = TRUE)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  Table1
## X-squared = 1.7443, df = 1, p-value = 0.1866
```

p-value $>$ 0.05 , Fail to reject H0

calculate OR , RR :

```
library(epiR)
```

```
## Loading required package: survival
## Warning: package 'survival' was built under R version 3.6.1
## Package epiR 1.0-2 is loaded
## Type help(epi.about) for summary information
##
```

```
epi.2by2(Table1 , method = "cohort.count" , conf.level = 0.95)
```

```
##              Outcome +    Outcome -      Total        Inc risk *
## Exposed +          314           44        358              87.7
## Exposed -          334           33        367              91.0
## Total              648           77        725              89.4
##                   Odds
## Exposed +         7.14
## Exposed -        10.12
## Total             8.42
##
## Point estimates and 95% CIs:
## -------------------------------------------------------------------
## Inc risk ratio                                0.96 (0.92, 1.01)
## Odds ratio                                    0.71 (0.44, 1.14)
## Attrib risk *                                 -3.30 (-7.79, 1.19)
## Attrib risk in population *                   -1.63 (-5.32, 2.06)
## Attrib fraction in exposed (%)                -3.76 (-9.12, 1.34)
## Attrib fraction in population (%)             -1.82 (-4.34, 0.64)
## -------------------------------------------------------------------
##  Test that odds ratio = 1: chi2(1) = 2.077 Pr>chi2 = 0.15
##  Wald confidence limits
##  CI: confidence interval
##  * Outcomes per 100 population units
```

Odds of Females not smoking are 0.71 times odds of males not smoking

```
1/0.71
```

```
## [1] 1.408451
```

Odds of males not smoking are 1.4 times odds of Females not smoking

check normality

```
library(moments)
skewness(LungCap)
```

```
## [1] -0.2274017
```

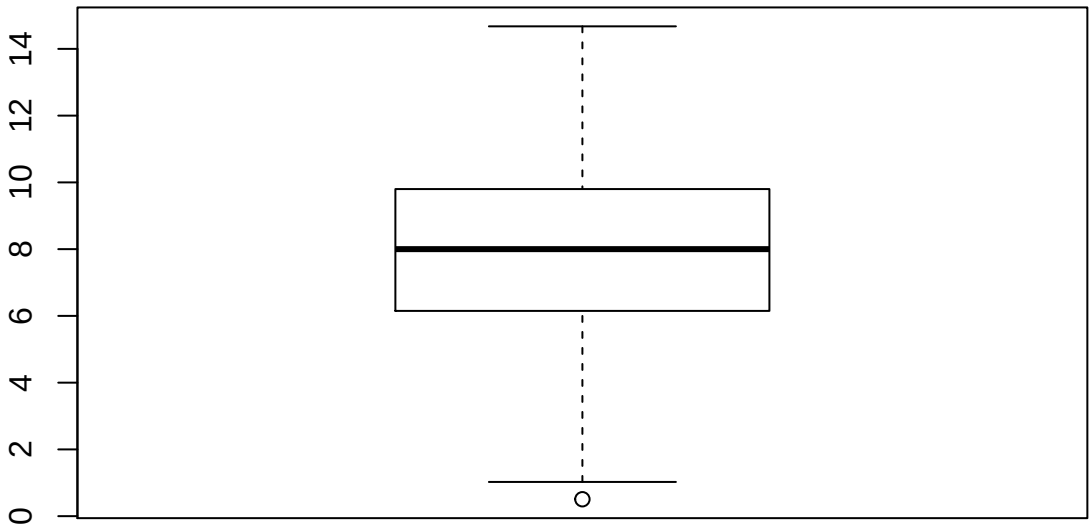
accepted level from -1 to +1

```
kurtosis(LungCap)
```

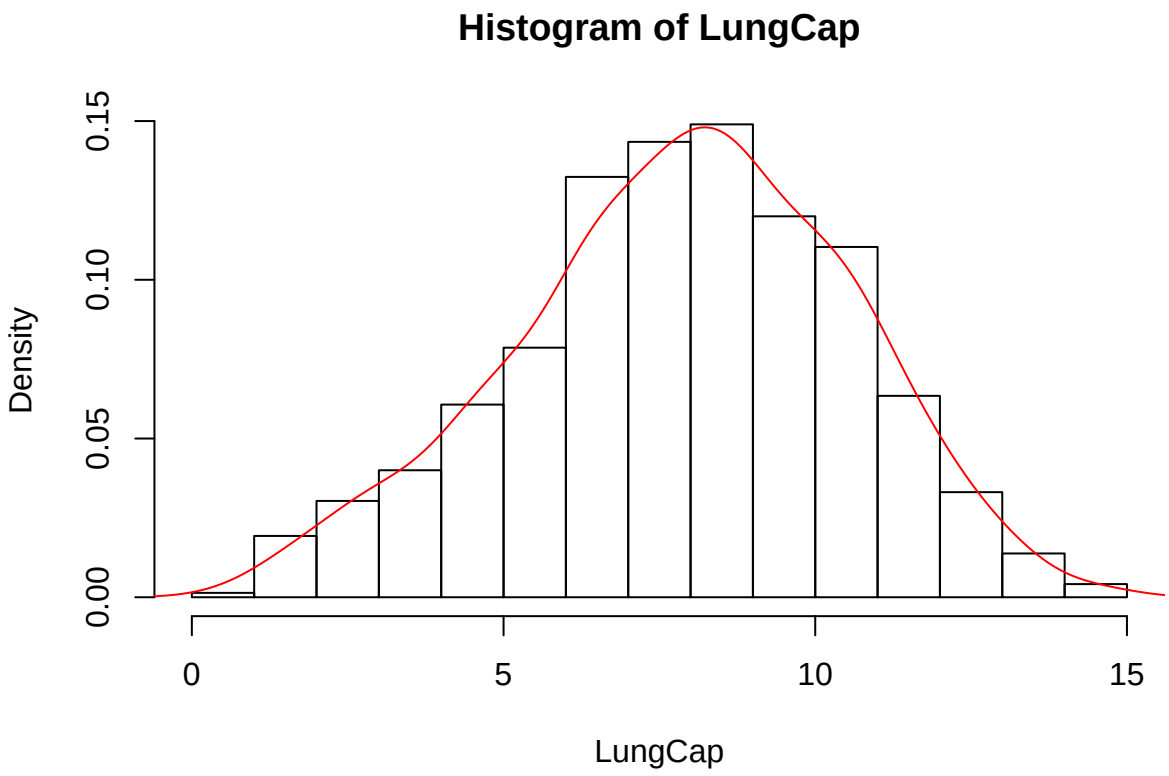
```
## [1] 2.68148
```

accepted level from -2 to +2 may to +3

```
boxplot(LungCap)
```




```
hist(LungCap,freq = FALSE)
lines(density(LungCap),col="red",lwd=1)
```



###visually ,data is normally distributed

One-sample t-test for lung Capacity :

Test H0 =8 , conf.interval = 0.95 :

```
t.test(LungCap , mu=8 , alternative = "two.sided" , conf.level = 0.95)
```

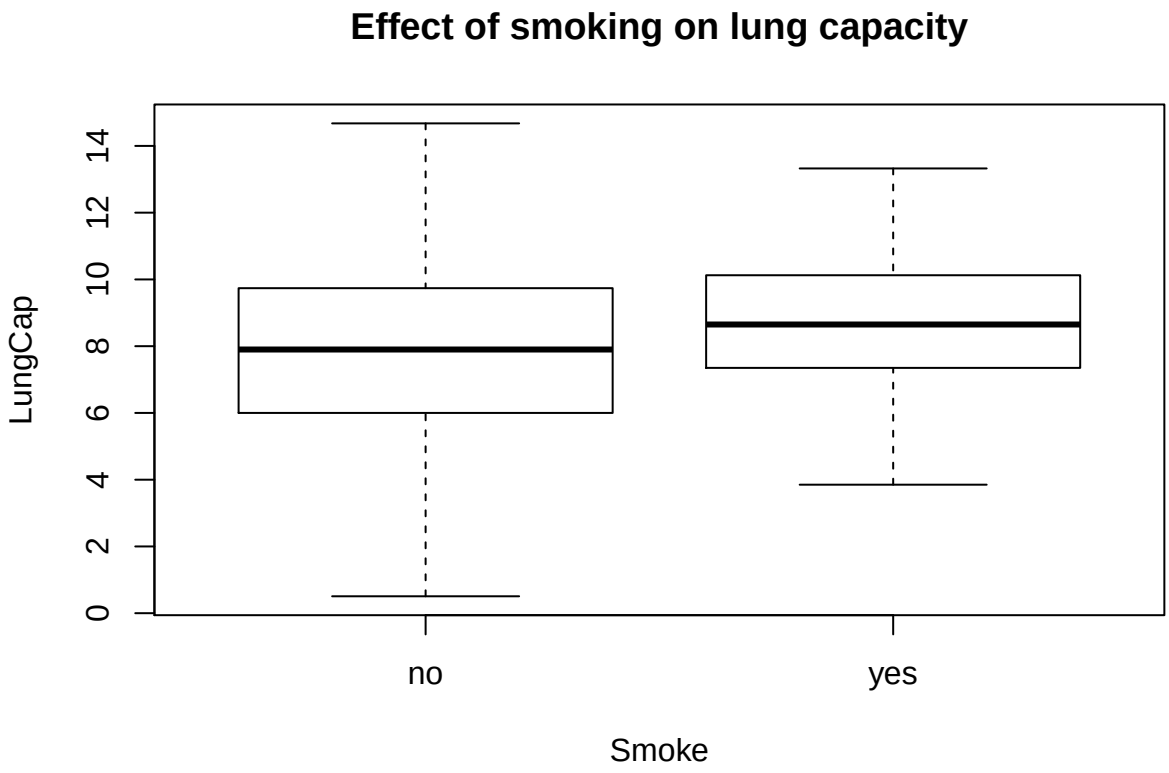
```
##
##  One Sample t-test
##
## data:  LungCap
## t = -1.3842, df = 724, p-value = 0.1667
## alternative hypothesis: true mean is not equal to 8
## 95 percent confidence interval:
##  7.669052 8.057243
## sample estimates:
## mean of x
##  7.863148
```

p-value >0.05 , fail to reject H0

Relation between Smoke & lung Capacity :

H0 : mean of smokers = mean of non smokers :

```
boxplot(LungCap~Smoke , main = "Effect of smoking on lung capacity")
```



check variance :

```
var(LungCap[Smoke == "yes"])
```

```
## [1] 3.545292
```

```
var(LungCap[Smoke == "no"])
```

```
## [1] 7.431694
```

so variance not equal

```
t.test(LungCap~Smoke , mu=0 , alternative = "two.sided" , var.eq = F, conf.level =  0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  LungCap by Smoke
```

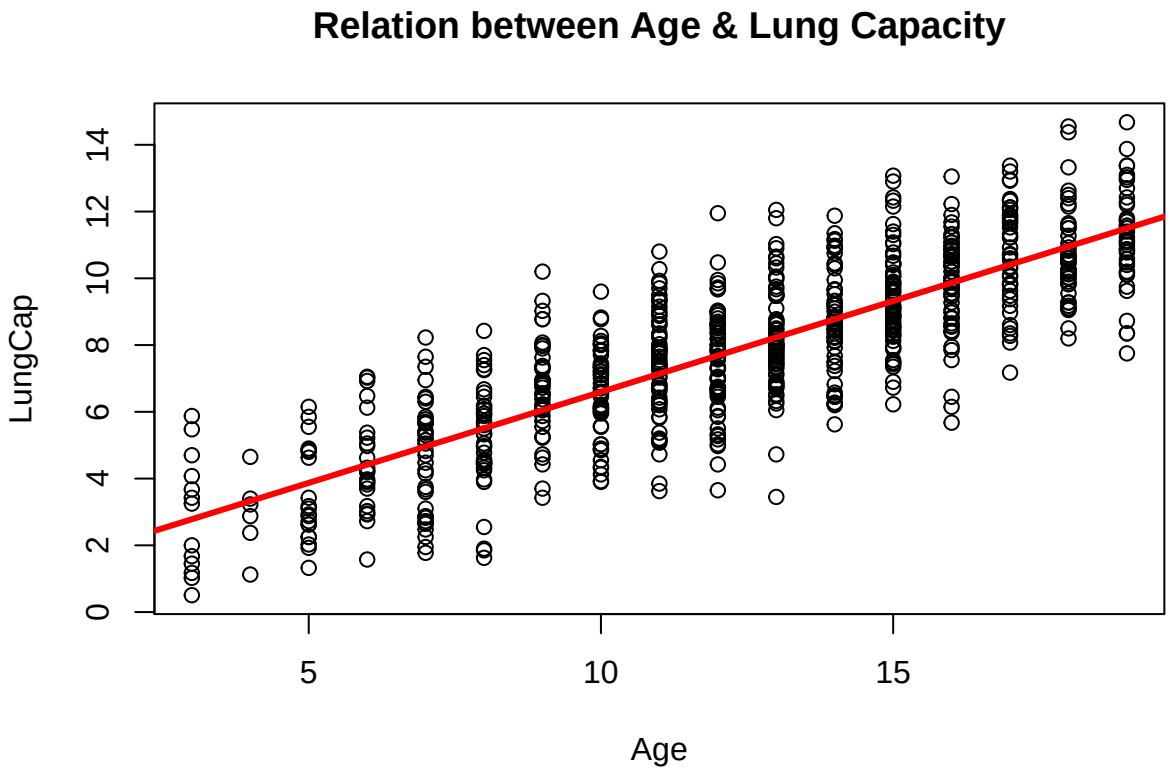
```
## t = -3.6498, df = 117.72, p-value = 0.0003927
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.3501778 -0.4003548
## sample estimates:
##  mean in group no mean in group yes
##          7.770188          8.645455
```

p-value < 0.05 , reject H0 , Smoking has a significant effect on lung capacity

fit a model of the relation between Age , LungCap :

use simple linear regression

```
model1 <- lm(LungCap~Age)
plot(Age,LungCap,main = "Relation between Age & Lung Capacity")
abline(model1 ,col=2 , lwd=3)
```



correlation between Lung capacity & Age

```
cor(Age,LungCap ,method="pearson")
```

```
## [1] 0.8196749
```

there is +ve strong correlation

Denisty plots : check if the response variable is close to normal :

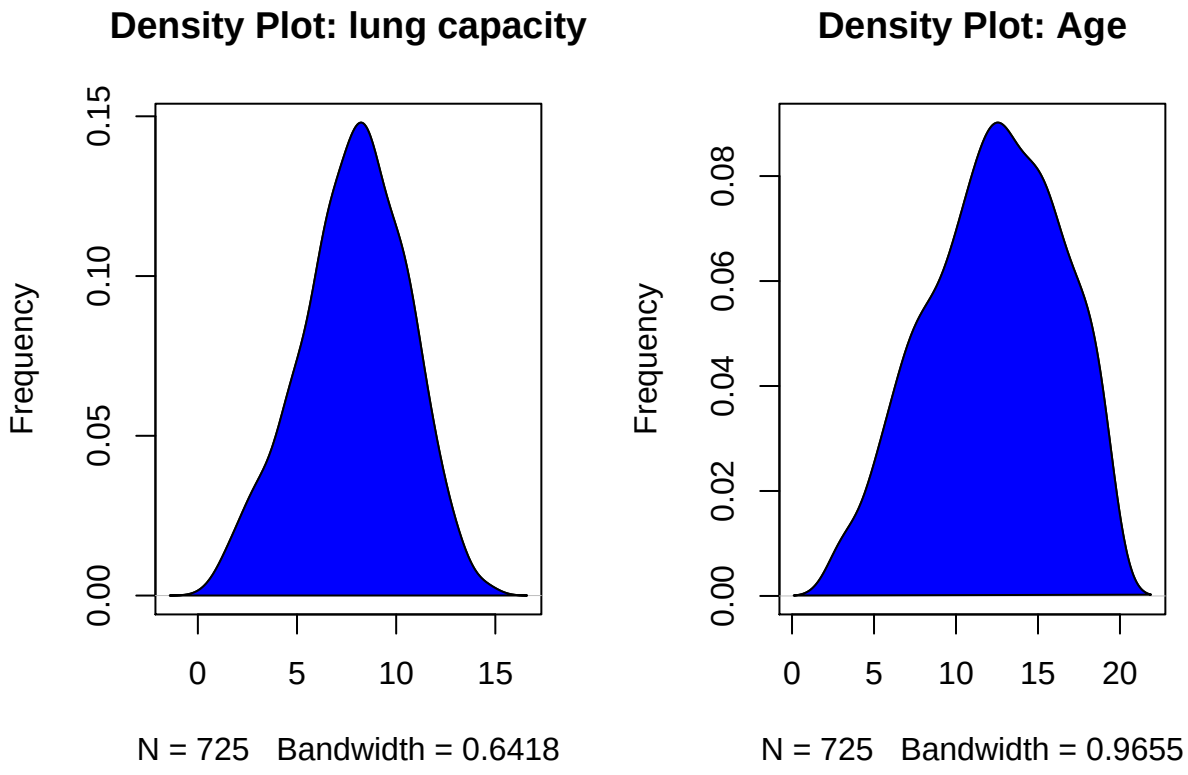
```
library(e1071)
```

```
## Warning: package 'e1071' was built under R version 3.6.1
```

```
##
## Attaching package: 'e1071'
```

```
## The following objects are masked from 'package:moments':
##
##     kurtosis, moment, skewness
```

```
par(mfrow=c(1, 2))  # divide graph area in 2 columns
plot(density(LungCap), main="Density Plot: lung capacity", ylab="Frequency")
# density plot for 'lung capacity'
polygon(density(LungCap), col="blue")
plot(density(Age), main="Density Plot: Age", ylab="Frequency")  # density plot for 'dist'
polygon(density(Age), col="blue")
```



built linear model equation :

```
model1 <- lm(LungCap~Age)
model1
```

```
##
## Call:
## lm(formula = LungCap ~ Age)
##
## Coefficients:
## (Intercept)          Age
##      1.1469       0.5448
```

Equation : lungCap = 1.1469 + 0.5448 * Age

increase in 1 year of Age associated with 0.5448 increase in lung Capacity

check the residuals and significance

H0 : slope = 0

```
summary(model1)
```

```
##
## Call:
## lm(formula = LungCap ~ Age)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.7799 -1.0203 -0.0005  0.9789  4.2650
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.14686    0.18353   6.249 7.06e-10 ***
## Age          0.54485    0.01416  38.476  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.526 on 723 degrees of freedom
## Multiple R-squared:  0.6719, Adjusted R-squared:  0.6714
## F-statistic:  1480 on 1 and 723 DF,  p-value: < 2.2e-16
```

p-value <0.05 , reject H0 ,there is significant diffrence

67% of the variation in Lung Capicity is explained by Age

test H0: variation mean squared regression = variation mean squared errors

```
anova(model1)
```

```
## Analysis of Variance Table
##
## Response: LungCap
##            Df Sum Sq Mean Sq F value    Pr(>F)
## Age         1 3447.0  3447.0  1480.4 < 2.2e-16 ***
## Residuals 723 1683.5     2.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
sqrt(2.3)
```

```
## [1] 1.516575
```

p-value < 0.05 , reject H0

Getting the coeffecient confedience interval :

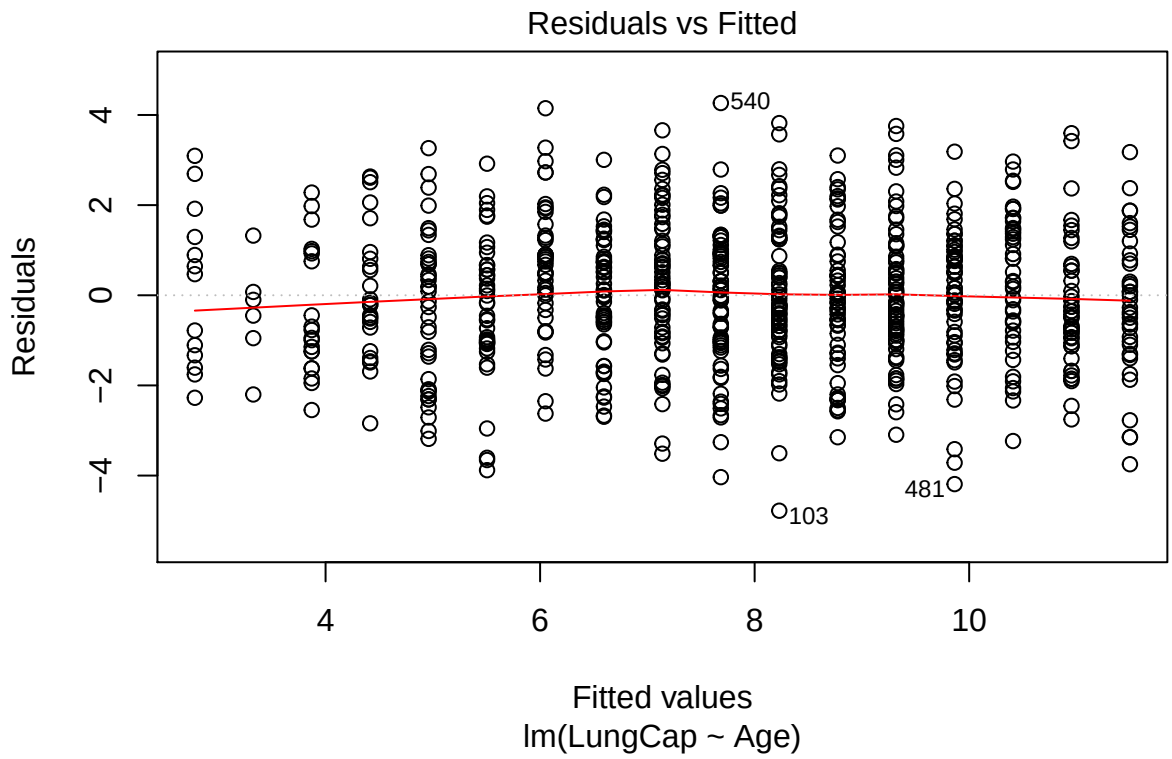
```
confint(model1)
```

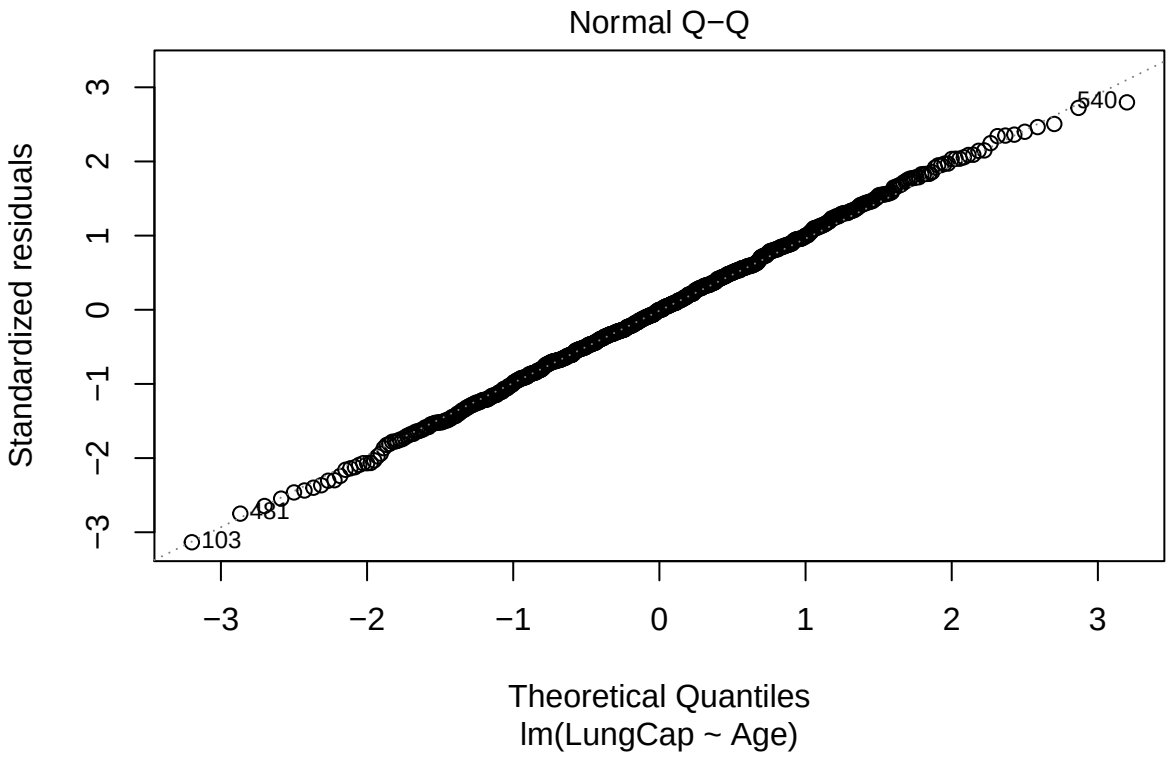
```
##                 2.5 %    97.5 %
## (Intercept) 0.7865454 1.5071702
## Age         0.5170471 0.5726497
```

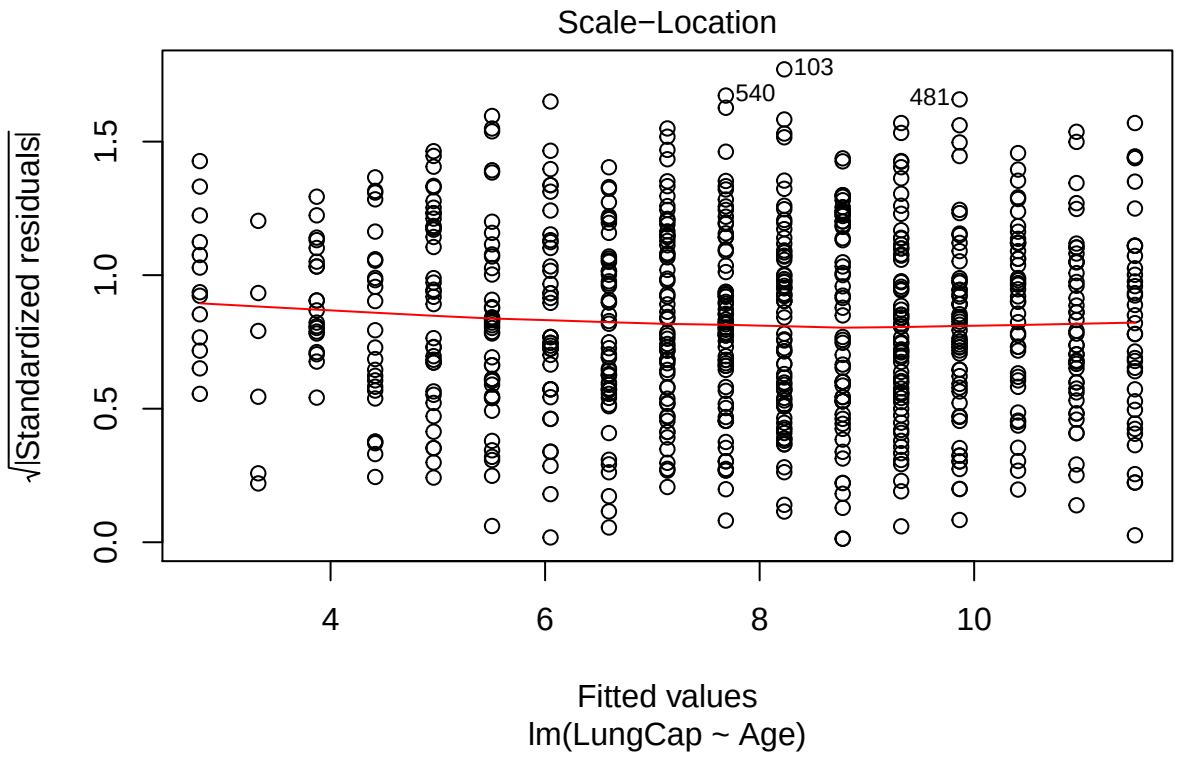
confedience interval not pass through zero , there is significant diffrence

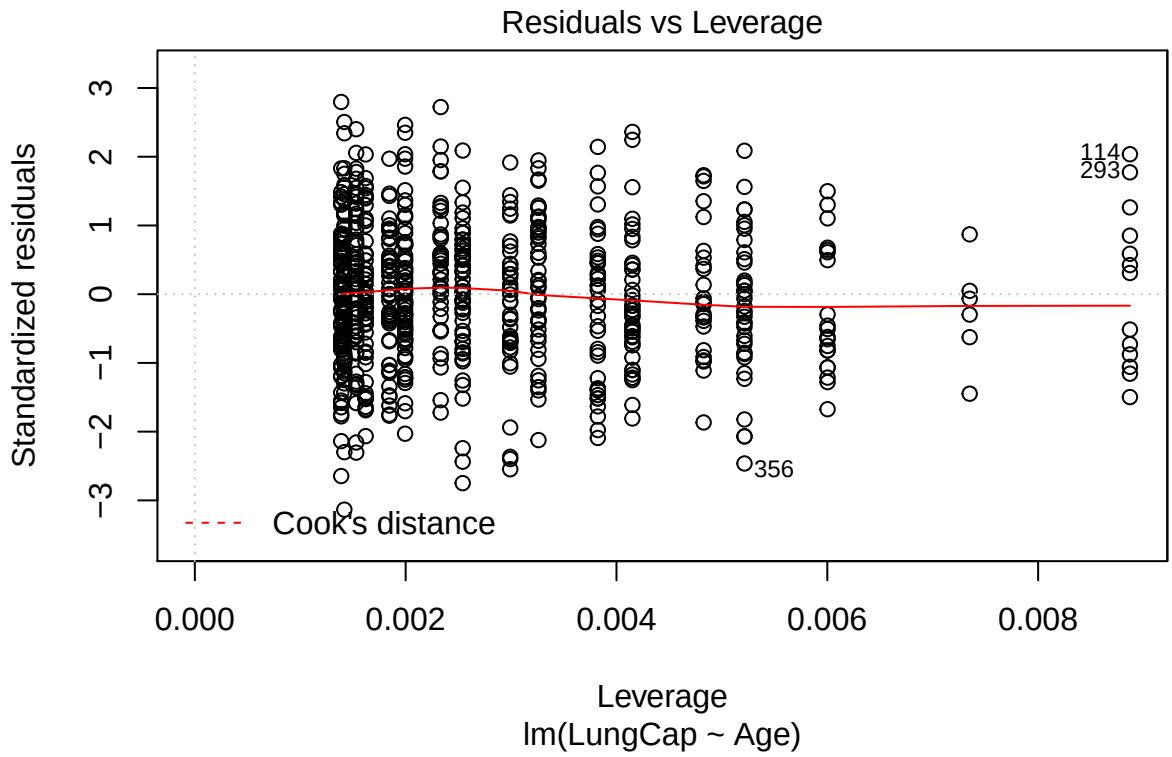
visualize the assumption

```
plot(model1)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(LungCap ~ Age)
103
481
540
−3
−2
−1
0
1
2
3


Scale−Location
103
540
481
√|Standardized residuals|
0.0
0.5
1.0
1.5
4
6
8
10
Fitted values
lm(LungCap ~ Age)



The diagnostic plots show residuals in four different ways:

Residuals vs Fitted. Used to check the linear relationship assumptions. A horizontal line, without distinct patterns is an indication for a linear relationship, what is good.

Normal Q-Q. Used to examine whether the residuals are normally distributed. It's good if residuals points follow the straight dashed line.

Scale-Location (or Spread-Location). Used to check the homogeneity of variance of the residuals (homoscedasticity). Horizontal line with equally spread points is a good indication of homoscedasticity.

Residuals vs Leverage. Used to identify influential cases, that is extreme values that might influence the regression results when included or excluded from the analysis.

fit a model using Age & Height as explanatory variables :

H0 B0=B1=B2= 0

```
mlr <- lm(LungCap~Age+Height , data = LungCapData)
summary(mlr)
```

```
##
## Call:
## lm(formula = LungCap ~ Age + Height, data = LungCapData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4080 -0.7097 -0.0078  0.7167  3.1679
```

```
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -11.747065   0.476899 -24.632  < 2e-16 ***
## Age           0.126368   0.017851   7.079 3.45e-12 ***
## Height        0.278432   0.009926  28.051  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.056 on 722 degrees of freedom
## Multiple R-squared:  0.843,  Adjusted R-squared:  0.8425
## F-statistic:  1938 on 2 and 722 DF,  p-value: < 2.2e-16
```

p-value $<$ 0.05 , reject H0

84.25% of variation in lung capacity is explained by Age and Height

Equation : lung capacity = -11.747 + (0.126*Age) + (0.278*Height)

increase in 1 year of Age with an increase in 0.126 of lung capacity adjusting for Height

pearson correlation between Age ,Height :

```
cor(Age , Height, method = "pearson")
```

```
## [1] 0.8357368
```

there is +ve strong correlation

Getting the coeffecient confedience interval :

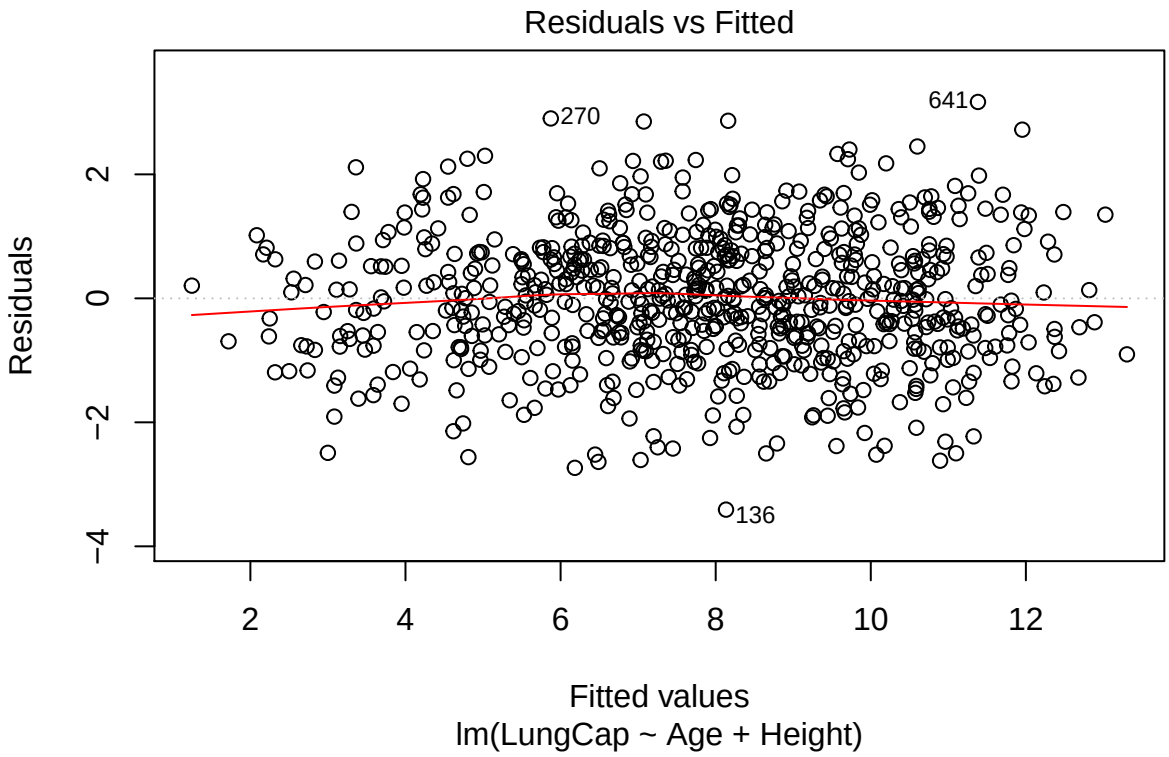
```
confint(mlr)
```

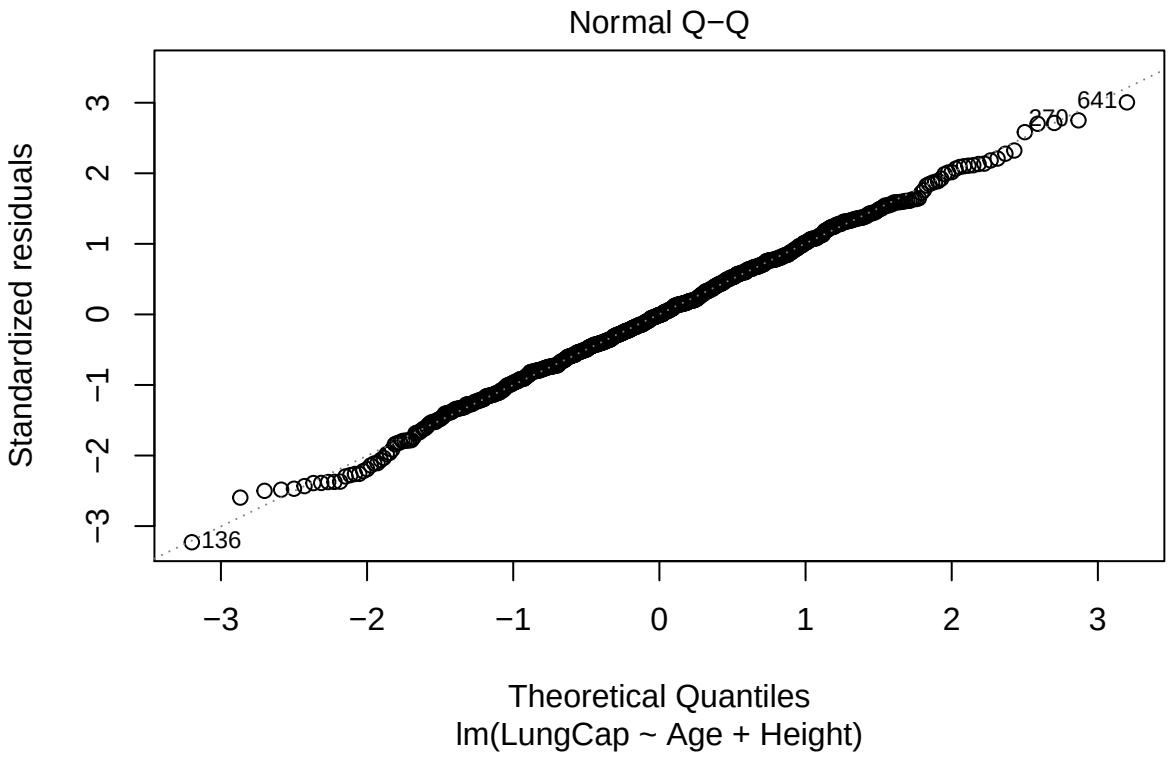
```
##                    2.5 %      97.5 %
## (Intercept) -12.68333877 -10.8107918
## Age           0.09132215   0.1614142
## Height        0.25894454   0.2979192
```

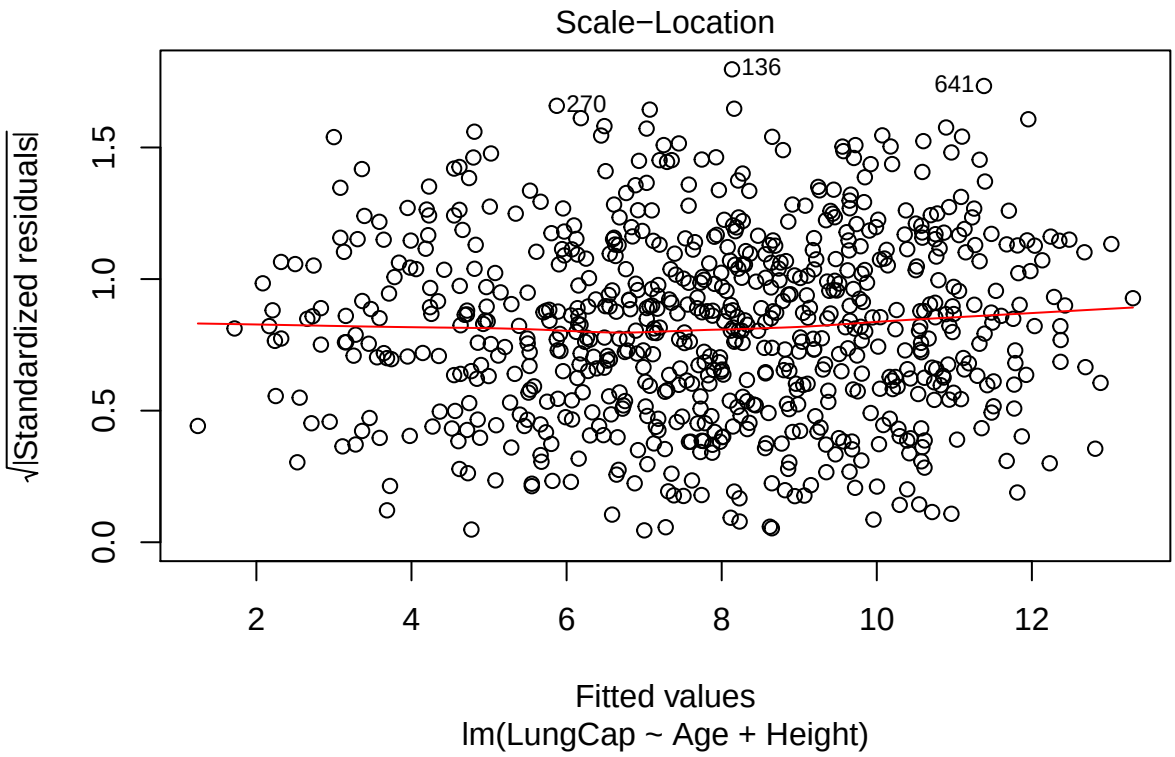
confedience interval not pass through zero , there is significant diffrence

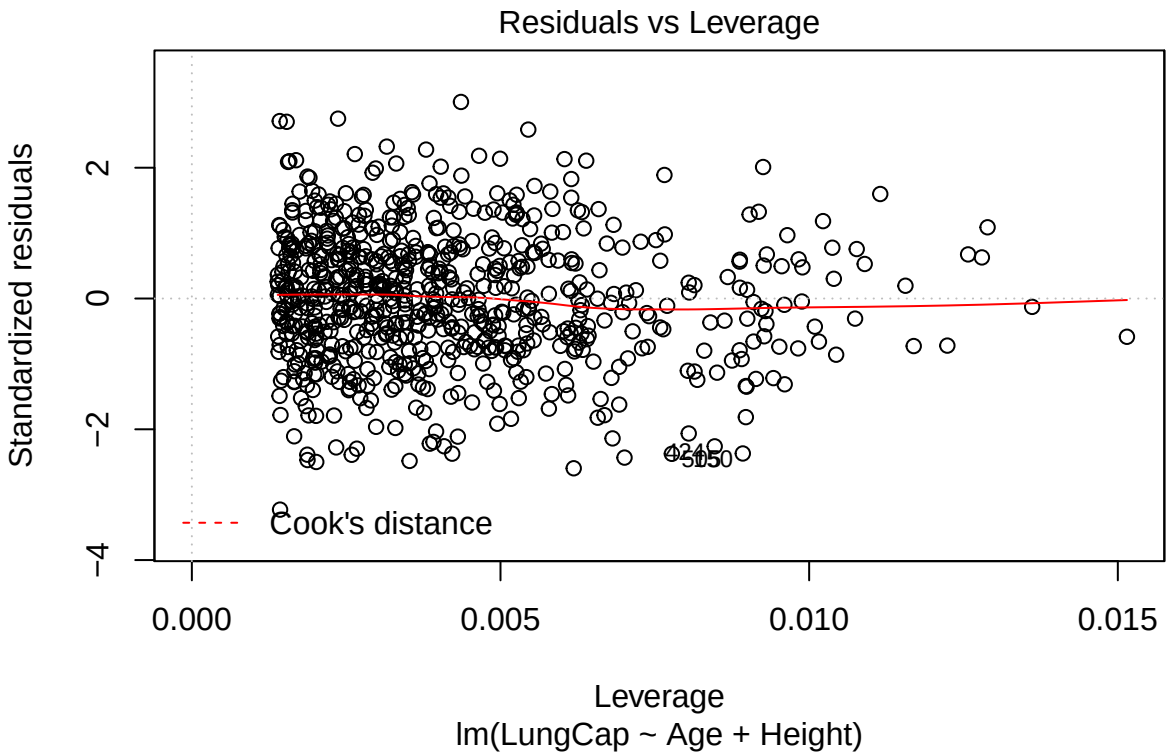
visualize the assumption

```
plot(mlr)
```


Residuals vs Fitted
270
641
136
Residuals
2
0
−2
−4
2
4
6
8
10
12
Fitted values
lm(LungCap ~ Age + Height)


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(LungCap ~ Age + Height)
136
370
641
−3
−2
−1
0
1
2
3


Scale−Location
136
641
270
√|Standardized residuals|
0.0
0.5
1.0
1.5
2
4
6
8
10
12
Fitted values
lm(LungCap ~ Age + Height)



The diagnostic plots show residuals in four different ways:

Residuals vs Fitted. Used to check the linear relationship assumptions. A horizontal line, without distinct patterns is an indication for a linear relationship, what is good.

Normal Q-Q. Used to examine whether the residuals are normally distributed. It's good if residuals points follow the straight dashed line.

Scale-Location (or Spread-Location). Used to check the homogeneity of variance of the residuals (homoscedasticity). Horizontal line with equally spread points is a good indication of homoscedasticity.

Residuals vs Leverage. Used to identify influential cases, that is extreme values that might influence the regression results when included or excluded from the analysis.

If we convert Height into categorical variable:

creat Height categorical A<50 , B=50-55 , c=55-60 , D=60-65 , E=65-70 , F>70

```
CatHeight <- cut(Height,breaks = c(0,50,55,60,65,70,100), labels = c("A","B","C","D","E","F"))
```

fit model using Age , Height (as categorical variable) as explanatory variables :

```
m2 <- lm(LungCap~Age+CatHeight)
summary(m2)
```

```
##
```

```
## Call:
## lm(formula = LungCap ~ Age + CatHeight)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8719 -0.7751  0.0281  0.7521  3.4160
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.97553    0.29394   3.319  0.00095 ***
## Age          0.20110    0.01859  10.816  < 2e-16 ***
## CatHeightB   1.48361    0.31780   4.668 3.62e-06 ***
## CatHeightC   2.68562    0.29818   9.007  < 2e-16 ***
## CatHeightD   3.93857    0.30623  12.862  < 2e-16 ***
## CatHeightE   5.00703    0.32105  15.596  < 2e-16 ***
## CatHeightF   6.53873    0.34635  18.879  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.159 on 718 degrees of freedom
## Multiple R-squared:  0.812,  Adjusted R-squared:  0.8104
## F-statistic: 516.8 on 6 and 718 DF,  p-value: < 2.2e-16
```

Lung capacity $= 0.976 + 0.201Age + 1.484\text{Xb} + 2.686Xc + 3.939\text{Xd} + 5.007Xe + 6.539\text{Xf}$

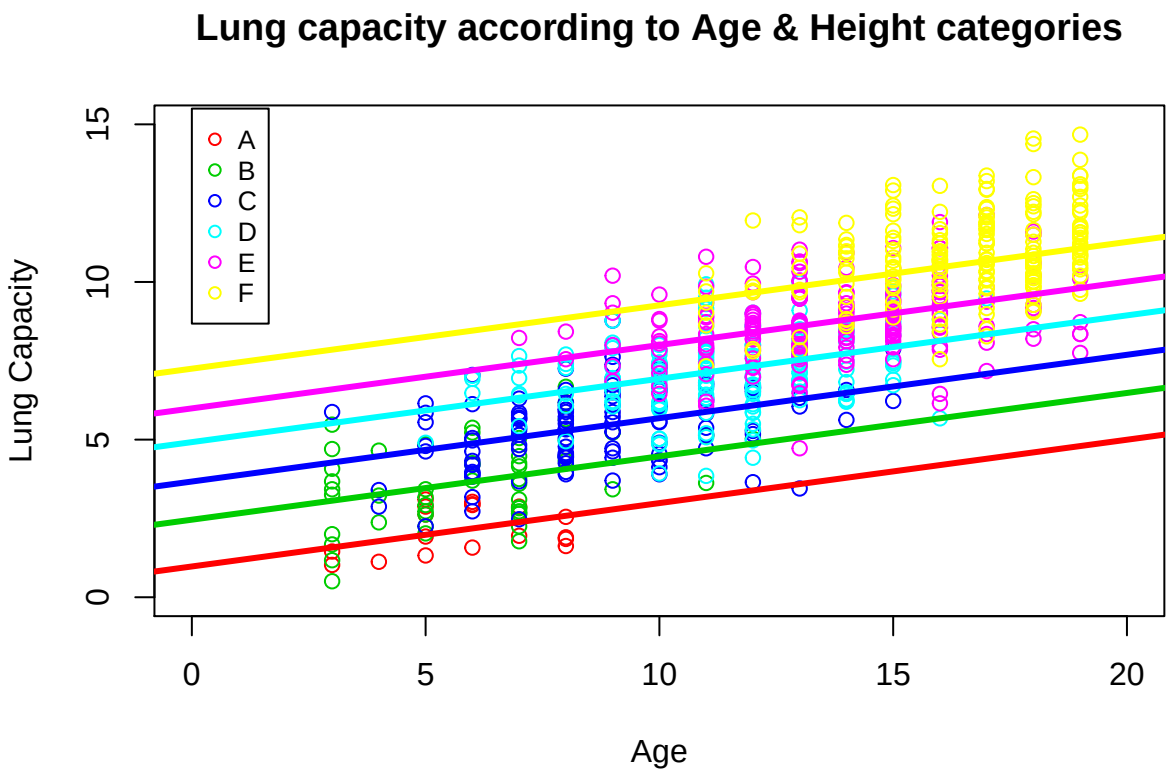
lung capacity for category A $= 0.976 + 0.201Age$ *###lung capacity for category B* $= 2.46 + 0.201$ **Age**

lung capacity for category c $= 3.67 + 0.201Age$ *###lung capacity for category D* $= 4.92 + 0.201$ **Age**

lung capacity for category E $= 5.99 + 0.201Age$ *###lung capacity for category F* $= 7.25 + 0.201$ **Age**

plot Data with different colors For Height categories:

```
plot(Age[CatHeight=="A"],LungCap[CatHeight=="A"] , col=2,xlim = c(0,20), ylim = c(0,15),xlab = "Age" , y
points(Age[CatHeight=="B"], LungCap[CatHeight=="B"],col=3)
points(Age[CatHeight=="C"], LungCap[CatHeight=="C"],col=4)
points(Age[CatHeight=="D"], LungCap[CatHeight=="D"],col=5)
points(Age[CatHeight=="E"], LungCap[CatHeight=="E"],col=6)
points(Age[CatHeight=="F"], LungCap[CatHeight=="F"],col=7)
legend(0,15.5,legend = c("A","B","C","D","E","F"),col = 2:7,pch = 1,cex = 0.8)
abline(a=0.976,b=0.201,col=2,lwd=3)
abline(a=2.46,b=0.201,col=3,lwd=3)
abline(a=3.67,b=0.201,col=4,lwd=3)
abline(a=4.92,b=0.201,col=5,lwd=3)
abline(a=5.99,b=0.201,col=6,lwd=3)
abline(a=7.25,b=0.201,col=7,lwd=3)
```

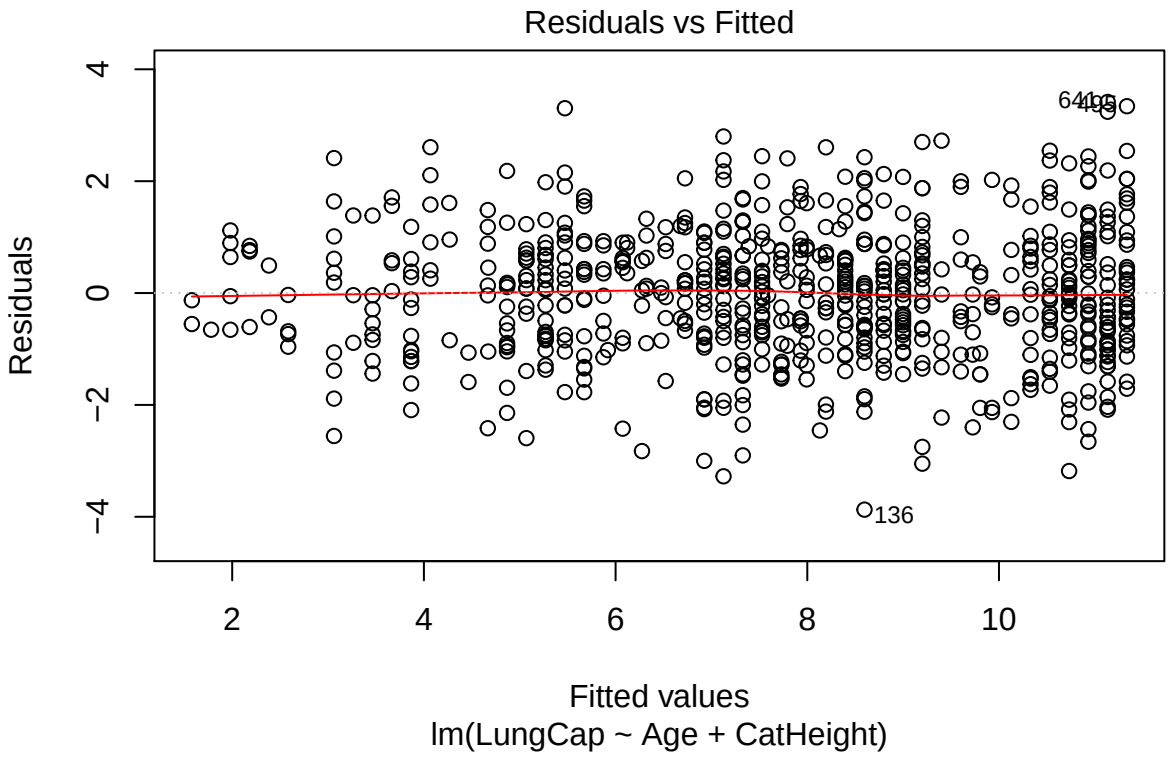


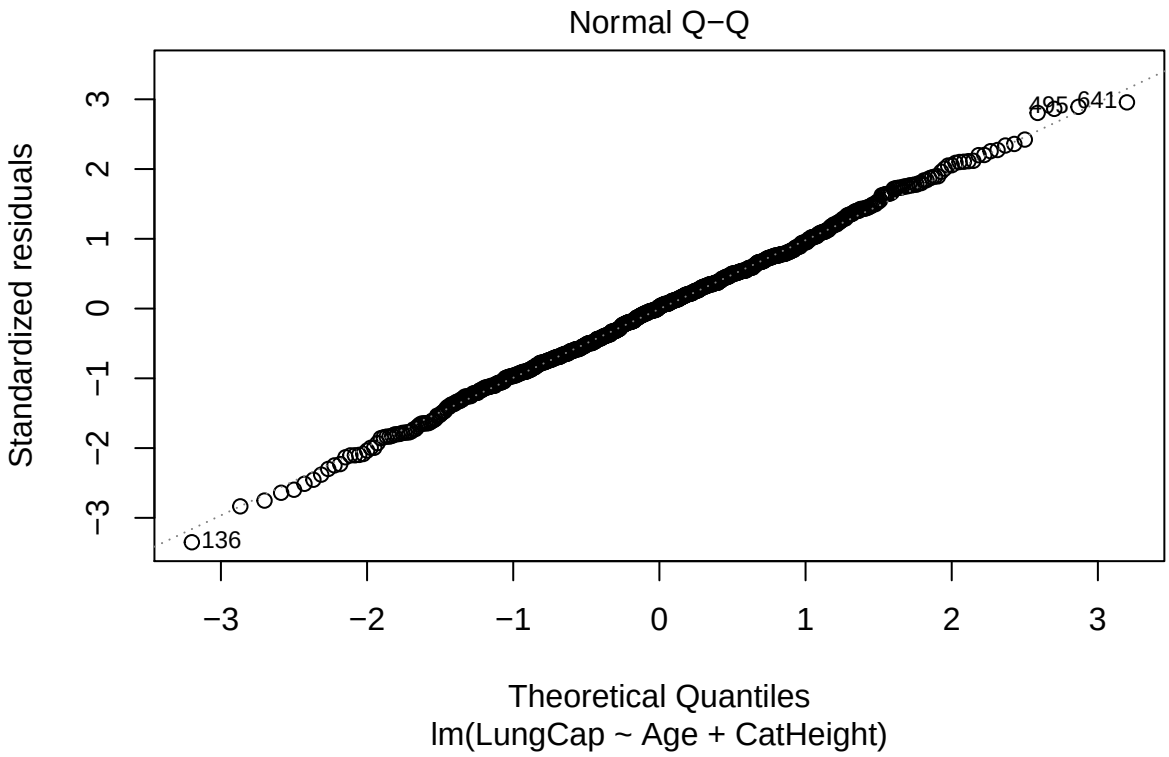
###increase in 1 year associate with 0.201 change in lung capacity independent on Height categories

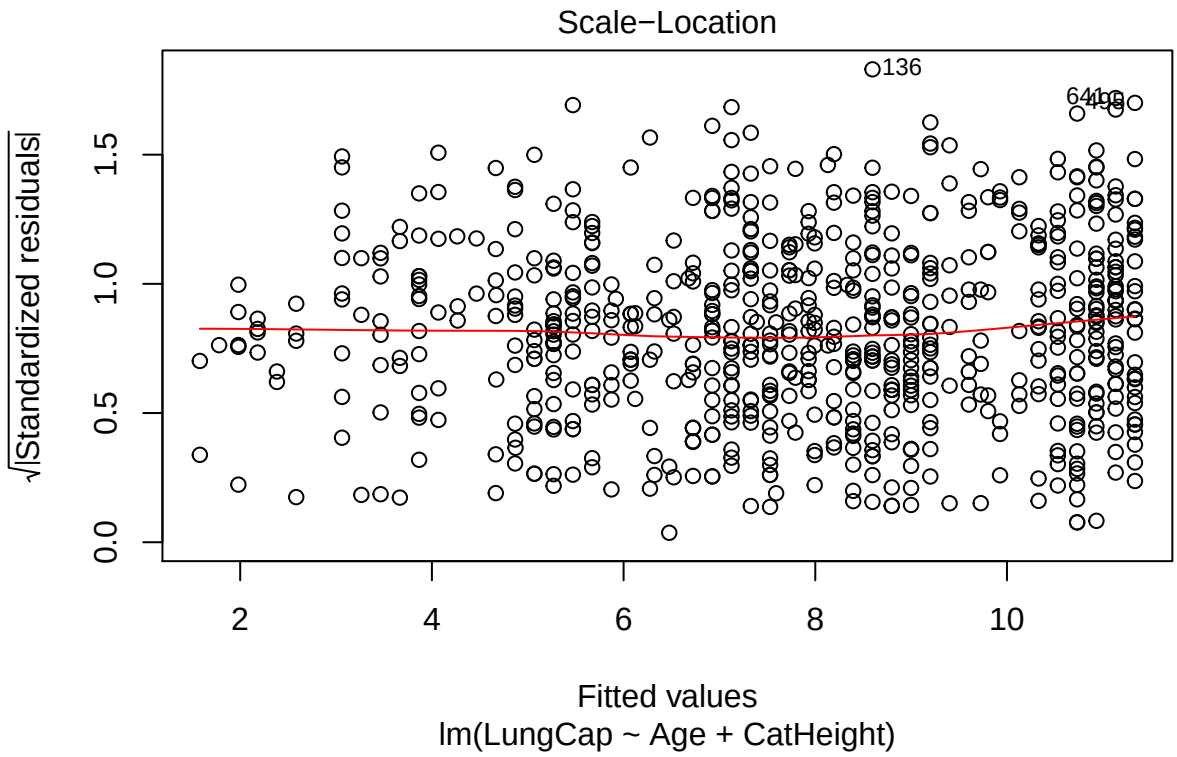
Age effect is the same for all Height categories

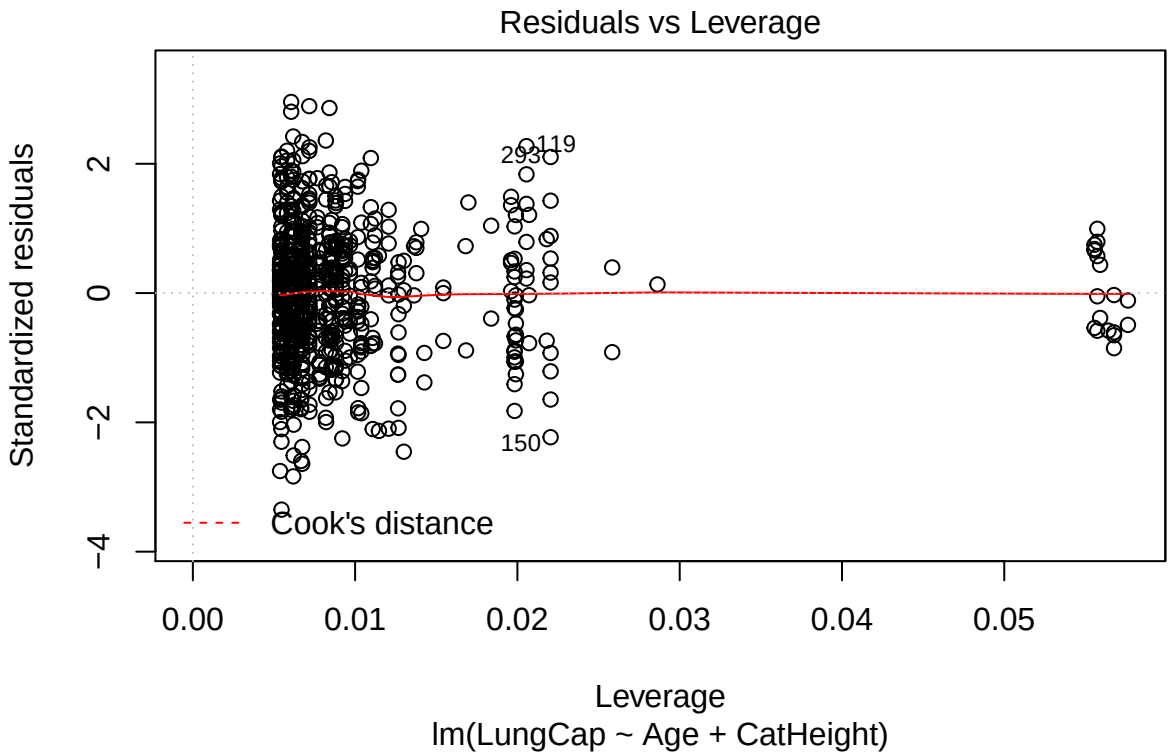
visualize the assumption

```
plot(m2)
```


Residuals vs Fitted
Residuals
Fitted values
lm(LungCap ~ Age + CatHeight)
136


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(LungCap ~ Age + CatHeight)
136
495
641
−3
−2
−1
0
1
2
3


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2
4
6
8
10
136
Fitted values
lm(LungCap ~ Age + CatHeight)



The diagnostic plots show residuals in four different ways:

Residuals vs Fitted. Used to check the linear relationship assumptions. A horizontal line, without distinct patterns is an indication for a linear relationship, what is good.

Normal Q-Q. Used to examine whether the residuals are normally distributed. It's good if residuals points follow the straight dashed line.

Scale-Location (or Spread-Location). Used to check the homogeneity of variance of the residuals (homoscedasticity). Horizontal line with equally spread points is a good indication of homoscedasticity.

Residuals vs Leverage. Used to identify influential cases, that is extreme values that might influence the regression results when included or excluded from the analysis.

fit model using Age , Smoking as explanatory variables :

```
mlr1 <- lm(LungCap~Age+Smoke)
summary(mlr1)
```

```
##
## Call:
## lm(formula = LungCap ~ Age + Smoke)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.8559 -1.0289 -0.0363  1.0083  4.1995
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.08572    0.18299   5.933 4.61e-09 ***
## Age          0.55540    0.01438  38.628  < 2e-16 ***
## Smokeyes    -0.64859    0.18676  -3.473 0.000546 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.514 on 722 degrees of freedom
## Multiple R-squared:  0.6773, Adjusted R-squared:  0.6764
## F-statistic: 757.5 on 2 and 722 DF,  p-value: < 2.2e-16
```

67% of variation in lung capacity is explained by Age & Smoke

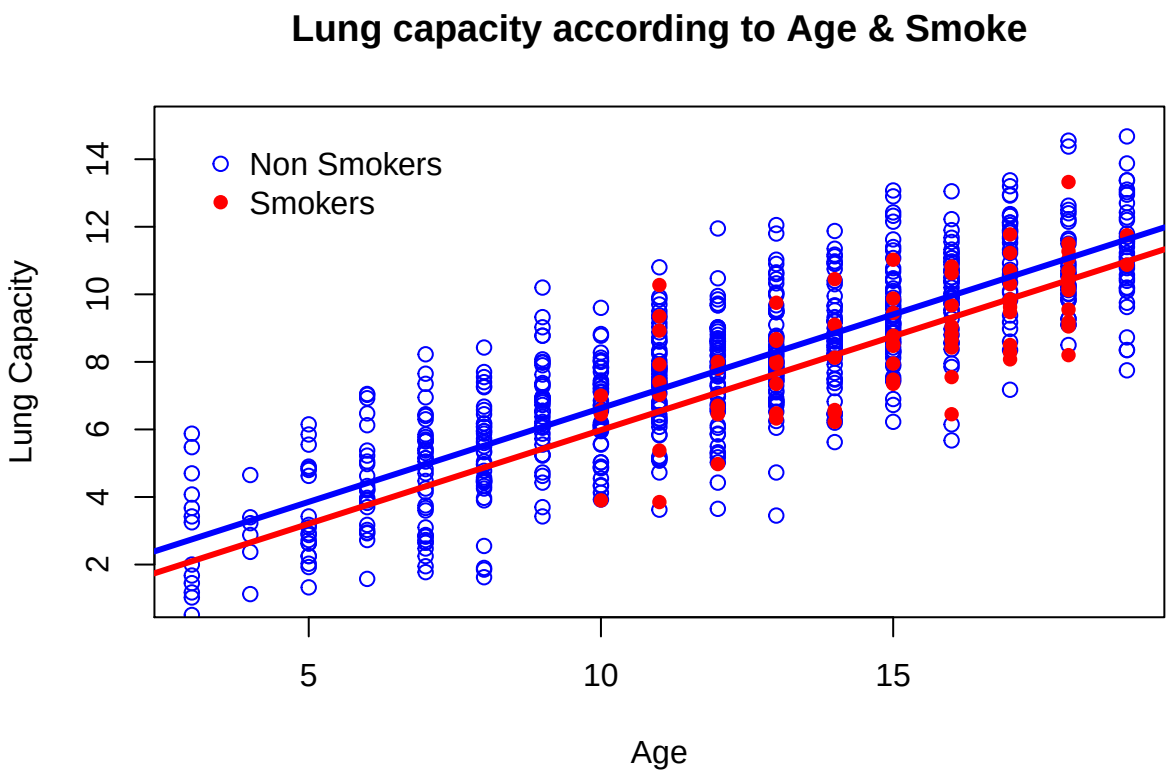
Equation : LungCap = 1.086 + (0.555*Age) +(-0.649*Smoke yes)

lung capacity in non smokers = 1.086 + (0.555*Age)

lung capacity in smokers = 0.437 + (0.555*Age)

Plot the data to differ between Smokers & non smokers :

```
plot(Age[Smoke=="no"],LungCap[Smoke=="no"] , col="blue" , ylim = c(1,15),xlab = "Age" , ylab = "Lung Ca
points(Age[Smoke=="yes"], LungCap[Smoke=="yes"],col="red",pch=16)
legend(3,15,legend = c("Non Smokers","Smokers"),col = c("blue","red"),pch = c(1,16),bty = "n")
abline(a=1.08,b=0.555,col="blue",lwd=3)
abline(a=0.431,b=0.555,col="red",lwd=3)
```



###increase in 1 year associate with 0.555 change in mean lung capacity , this increase is the same in

Smokers & non Smokers

For Smokers mean lung capacity decreased by 0.649 ,this decrease is the same in All ages

Effect of Age is independent on Smoking & vise versa , So no interaction between Age and Smoke

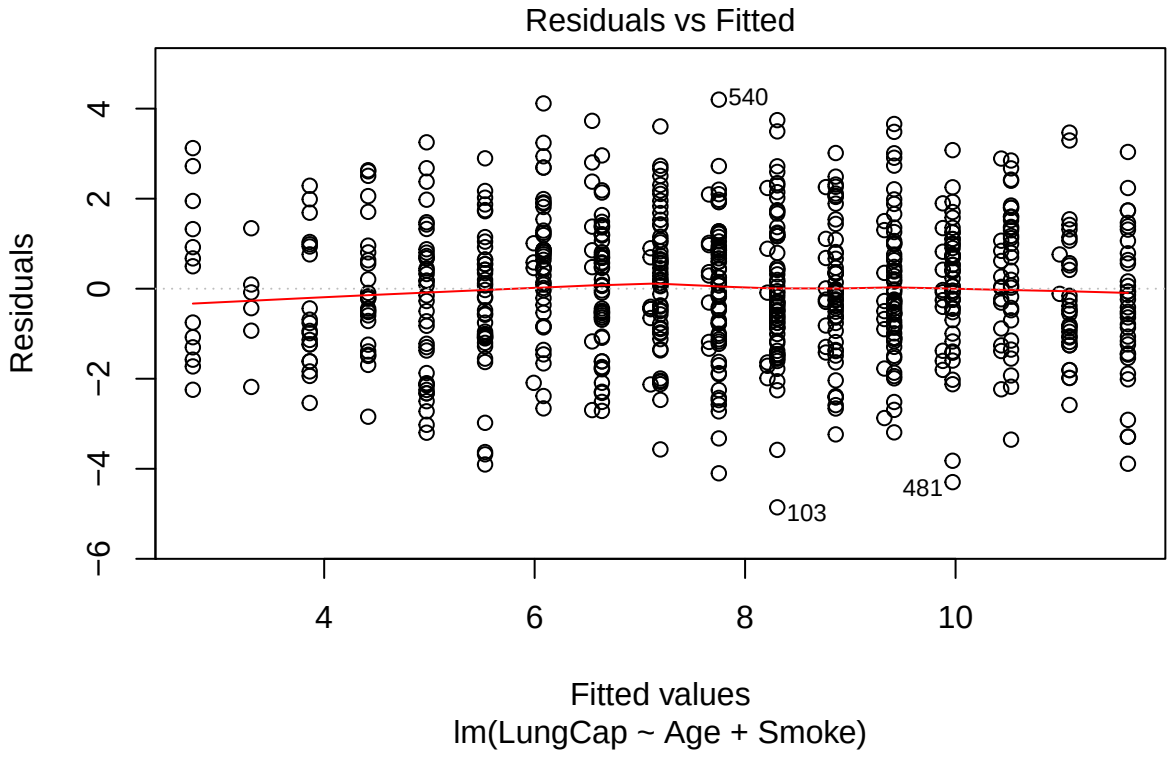
Getting the coeffecient confedience interval :

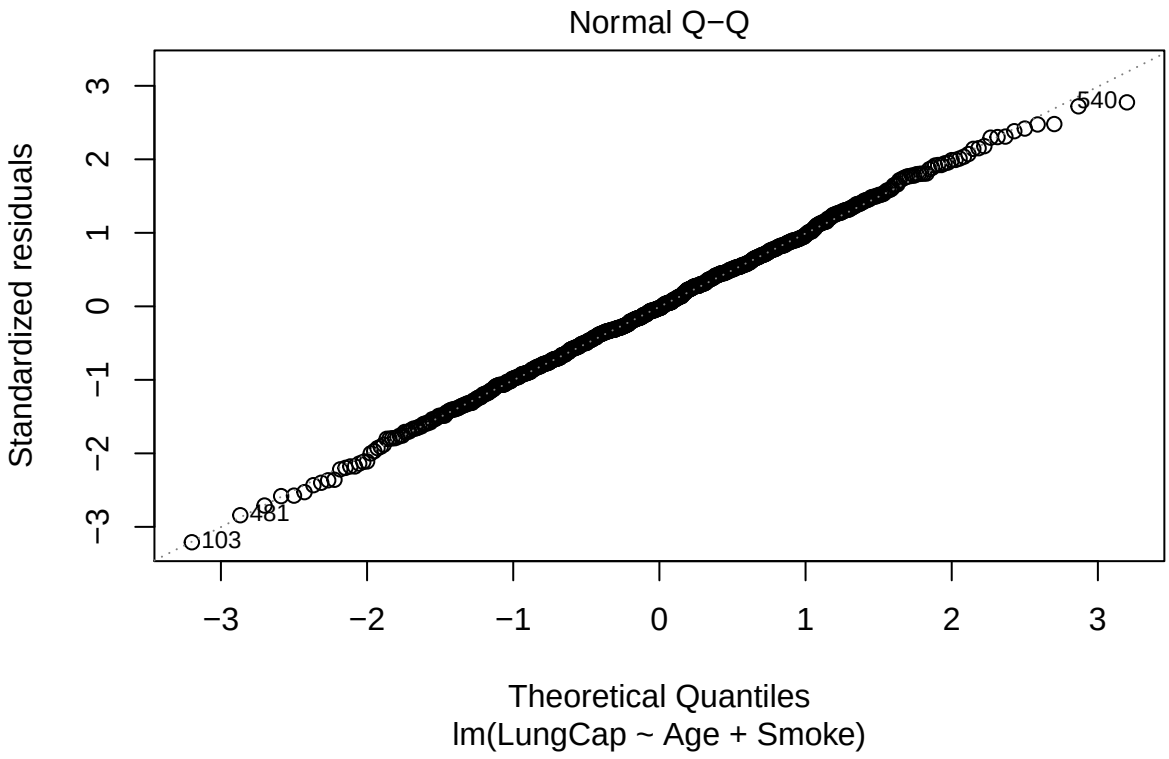
```
confint(mlr1)
```

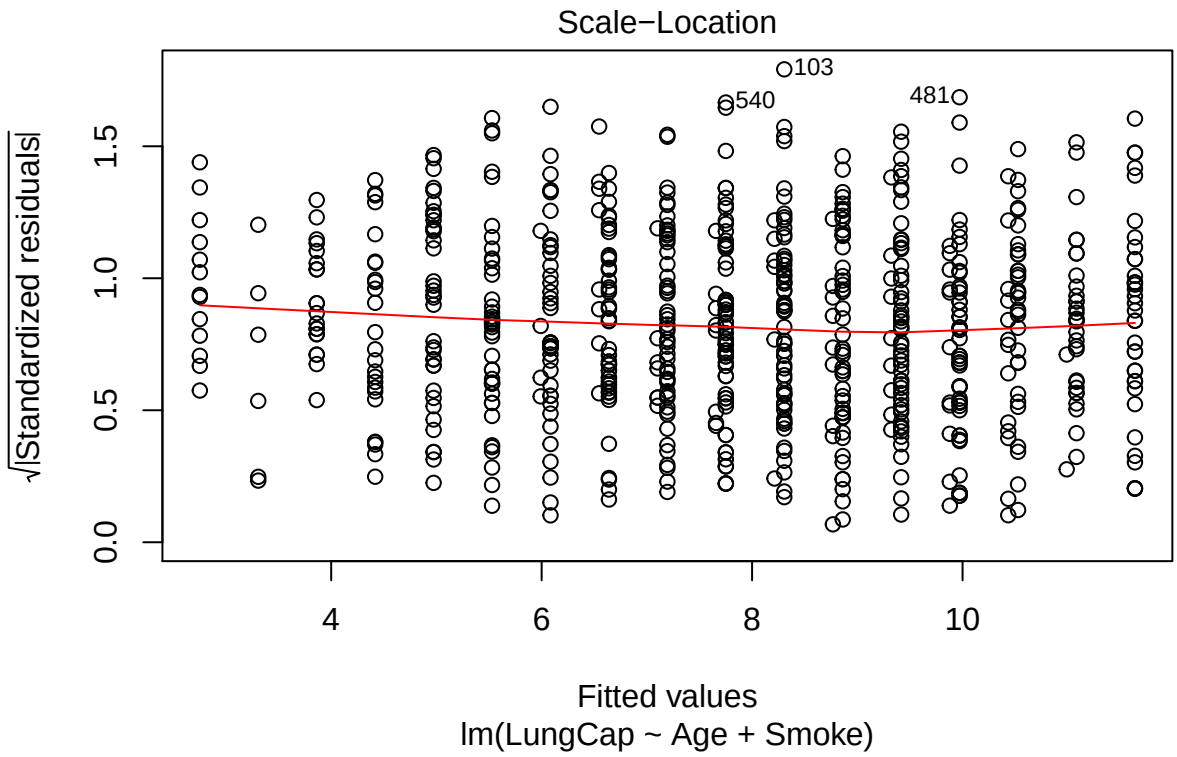
```
##                  2.5 %     97.5 %
## (Intercept)  0.7264702  1.4449793
## Age          0.5271678  0.5836240
## Smokeyes    -1.0152473 -0.2819294
```

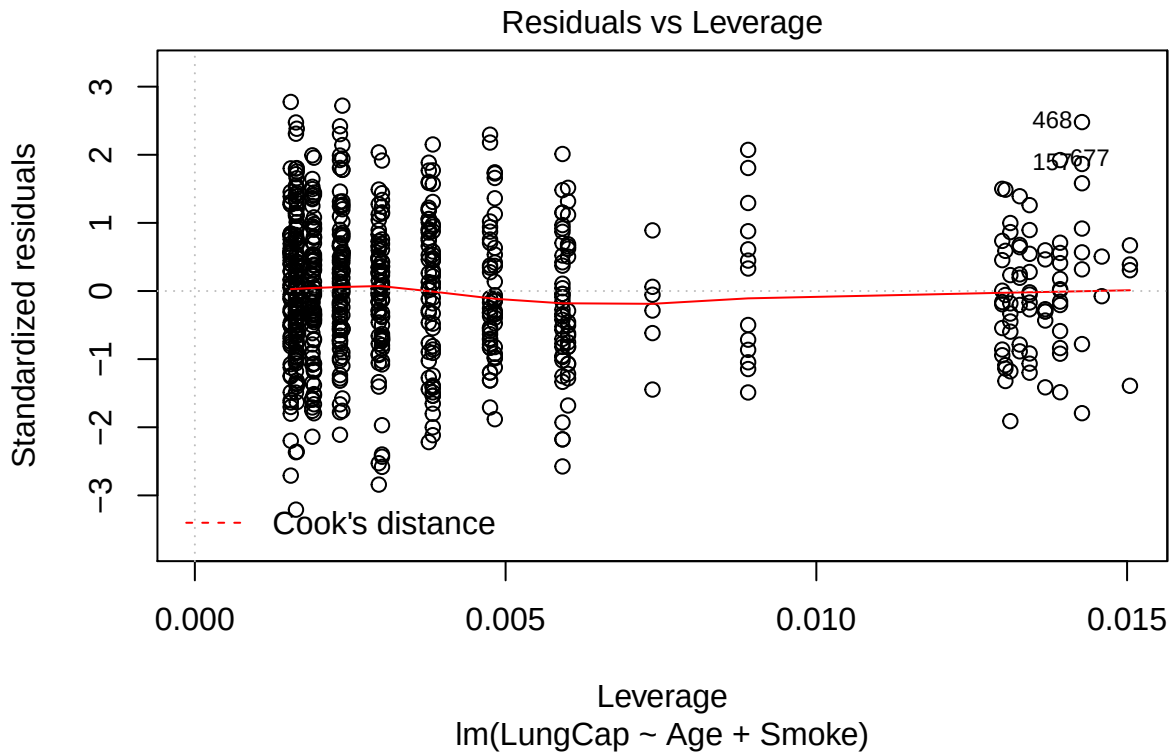
visualize the assumption

```
plot(mlr1)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(LungCap ~ Age + Smoke)
540
481
103
−3
−2
−1
0
1
2
3


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
4
6
8
10
103
540
481
Fitted values
lm(LungCap ~ Age + Smoke)



The diagnostic plots show residuals in four different ways:

Residuals vs Fitted. Used to check the linear relationship assumptions. A horizontal line, without distinct patterns is an indication for a linear relationship, what is good.

Normal Q-Q. Used to examine whether the residuals are normally distributed. It's good if residuals points follow the straight dashed line.

Scale-Location (or Spread-Location). Used to check the homogeneity of variance of the residuals (homoscedasticity). Horizontal line with equally spread points is a good indication of homoscedasticity.

Residuals vs Leverage. Used to identify influential cases, that is extreme values that might influence the regression results when included or excluded from the analysis.

fit model for all variables

```
mlr2 <- lm(LungCap~ Age+Height+Smoke+Gender+Caesarean,data = LungCapData)
summary(mlr2)
```

```
##
## Call:
## lm(formula = LungCap ~ Age + Height + Smoke + Gender + Caesarean,
##     data = LungCapData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3388 -0.7200  0.0444  0.7093  3.0172
##
```

```
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -11.32249    0.47097 -24.041  < 2e-16 ***
## Age            0.16053    0.01801   8.915  < 2e-16 ***
## Height         0.26411    0.01006  26.248  < 2e-16 ***
## Smokeyes      -0.60956    0.12598  -4.839 1.60e-06 ***
## Gendermale     0.38701    0.07966   4.858 1.45e-06 ***
## Caesareanyes  -0.21422    0.09074  -2.361   0.0185 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.02 on 719 degrees of freedom
## Multiple R-squared:  0.8542, Adjusted R-squared:  0.8532
## F-statistic: 842.8 on 5 and 719 DF,  p-value: < 2.2e-16
```

85.32% of variation in Lung capacity is explained by other variables

equation : Lung capacity = -11.32 + (0.16*Age) + (0.26*Height) + (-0.061*Smoke yes) + (0.38*Gender male) +

```
      (-0.21*caesarean yes)
```

Lung Capacity of non Smokers males = -10.94 +(0.16*Age) + (0.26*Height)

Lung Capacity of Smokers males = -11.001 +(0.16*Age) + (0.26*Height)

Lung Capacity of non Smokers females (caesarean) = -11.53 + (0.16*Age) + (0.26*Height)

Lung Capacity of non Smokers females (non caesarean) = -11.32 + (0.16*Age) + (0.26*Height)

Lung Capacity of Smokers females (caesarean) = -11.591 + (0.16*Age) + (0.26*Height)

Lung Capacity of Smokers females (non caesarean) = -11.381 + (0.16*Age) + (0.26*Height)

Getting the coeffecient confedience interval :

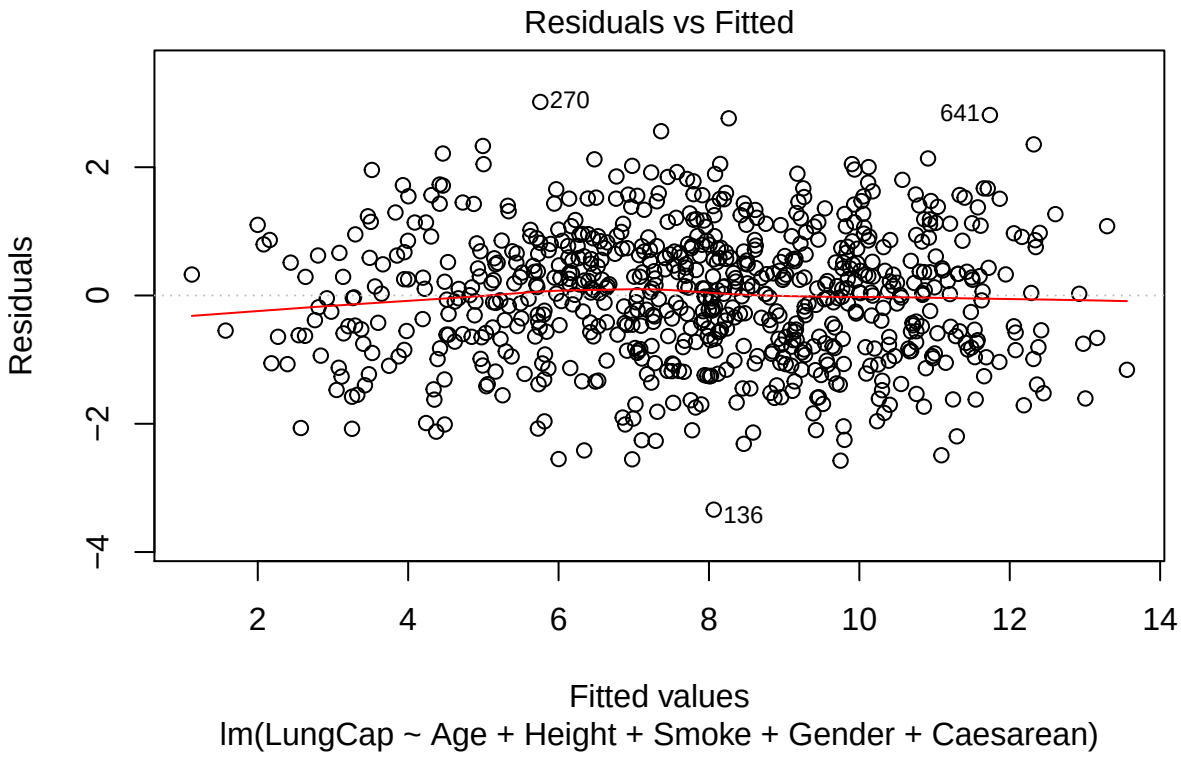
```
confint(mlr2)
```

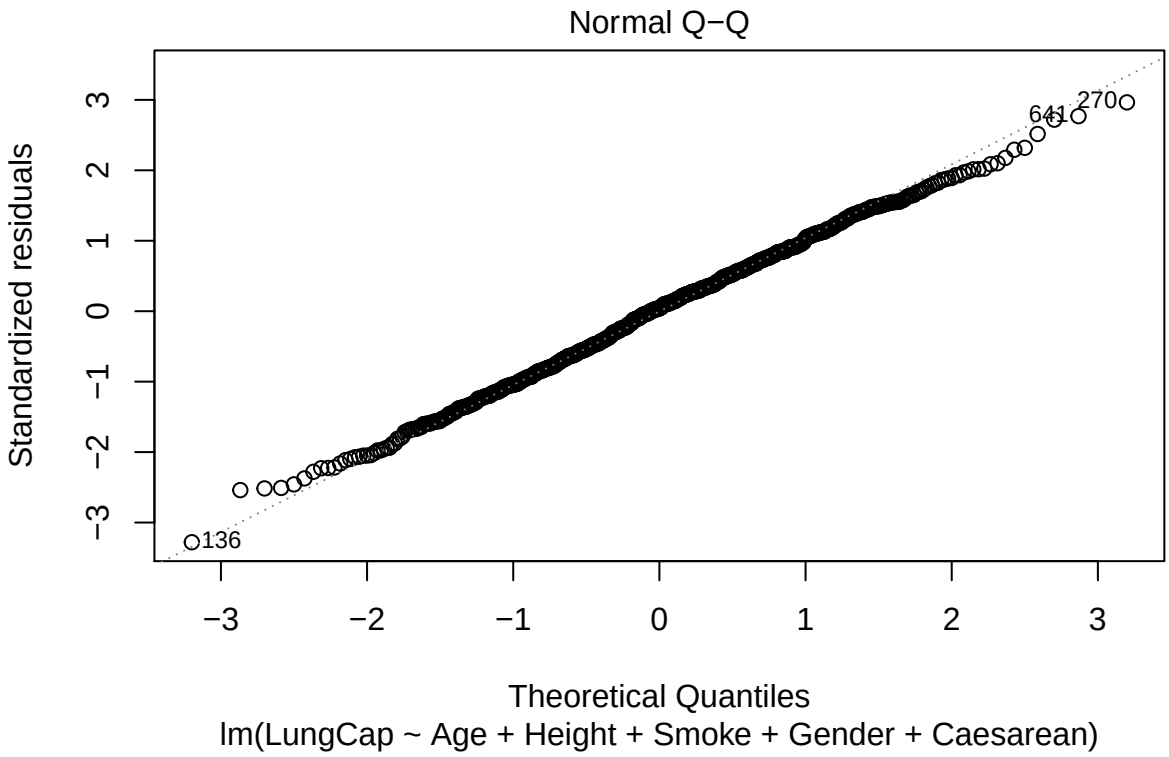
```
##                   2.5 %       97.5 %
## (Intercept)  -12.2471338 -10.39783728
## Age            0.1251765   0.19588271
## Height         0.2443581   0.28386751
## Smokeyes      -0.8568861  -0.36223237
## Gendermale     0.2306230   0.54340035
## Caesareanyes  -0.3923590  -0.03607738
```

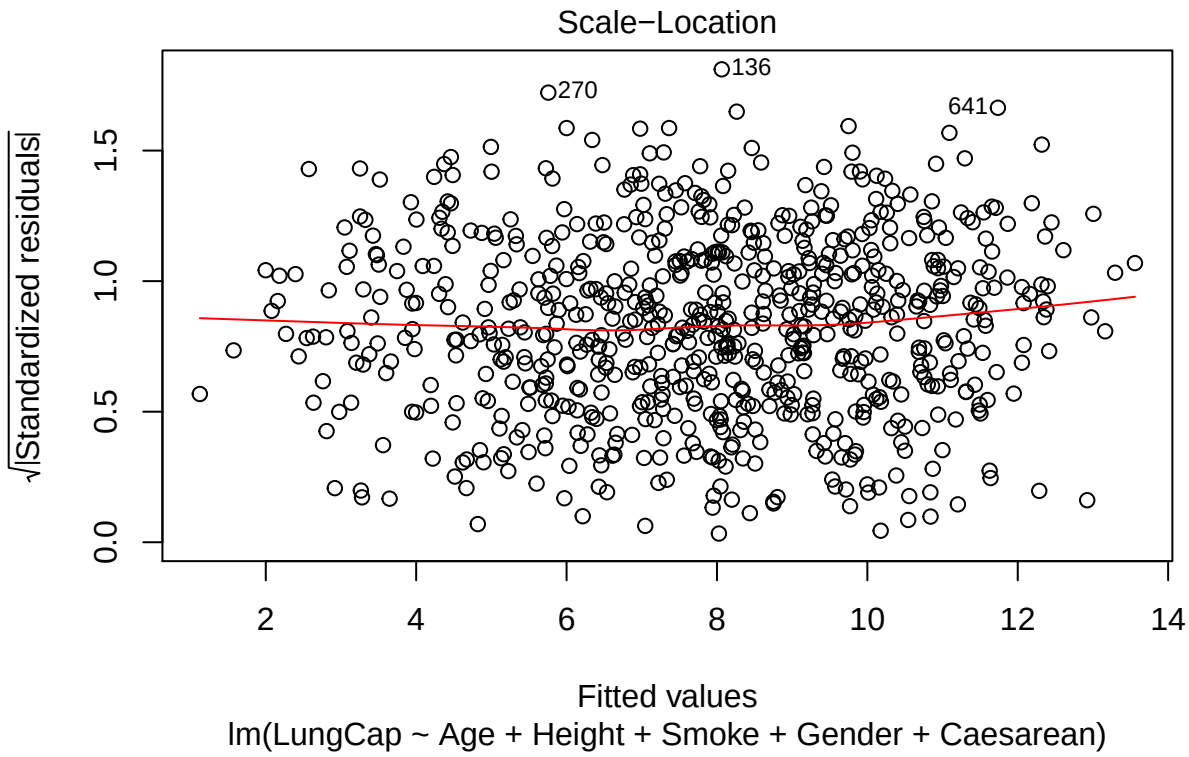
confedience interval not pass through zero , there is significant diffrence

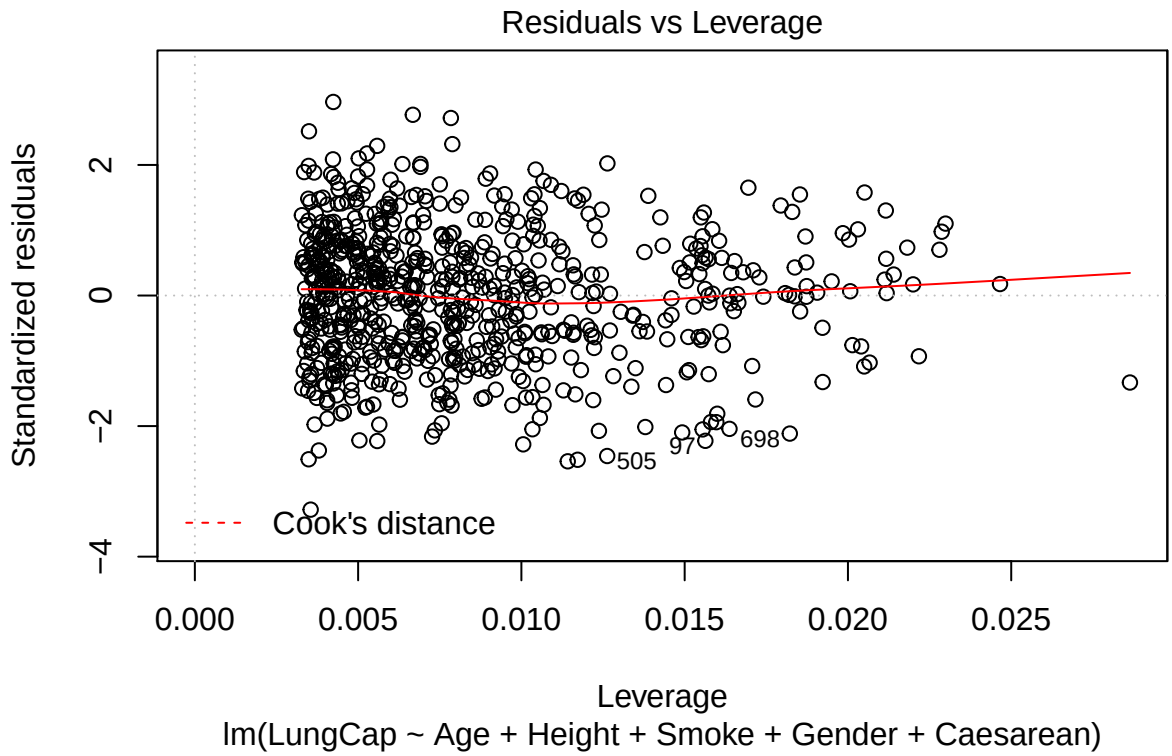
visualize the assumption

```
plot(mlr2)
```


Residuals vs Fitted
270
641
136
Residuals
2
0
−2
−4
2
4
6
8
10
12
14
Fitted values
lm(LungCap ~ Age + Height + Smoke + Gender + Caesarean)


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(LungCap ~ Age + Height + Smoke + Gender + Caesarean)
136
641
270


Scale−Location
136
270
641
√|Standardized residuals|
0.0
0.5
1.0
1.5
2
4
6
8
10
12
14
Fitted values
lm(LungCap ~ Age + Height + Smoke + Gender + Caesarean)



The diagnostic plots show residuals in four different ways:

Residuals vs Fitted. Used to check the linear relationship assumptions. A horizontal line, without distinct patterns is an indication for a linear relationship, what is good.

Normal Q-Q. Used to examine whether the residuals are normally distributed. It's good if residuals points follow the straight dashed line.

Scale-Location (or Spread-Location). Used to check the homogeneity of variance of the residuals (homoscedasticity). Horizontal line with equally spread points is a good indication of homoscedasticity.

Residuals vs Leverage. Used to identify influential cases, that is extreme values that might influence the regression results when included or excluded from the analysis.