individual project

Huijun Hao

5/10/2020

R Markdown

```
#import data file
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 3.6.3
```

```
TCB<- read_excel("C:/Users/zhang/Desktop/individual project/183 individual.xlsx")
TCB
```

```
## # A tibble: 126 x 5
##     `MR No`   TCB   TSB     BW Gender
##       <dbl> <dbl> <dbl>  <dbl> <chr>
##  1 1492989   8.3   7.2   2730 f
##  2 1493428  10.1   9     3760 f
##  3 1493611  13.7  11.3   4130 f
##  4 1493706   8.9   6.8   2455 f
##  5 1493704   8.7   7.7   3320 f
##  6 1493869   6.8   6.3   3290 f
##  7 1494140   9.2   7     3245 f
##  8 1494738   8.8   6.7   3215 f
##  9 1495268   8.7   7.5   2735 f
## 10 1496182   8.7   6.5   4180 f
## # ... with 116 more rows
```

```
#Get the regression model
lmTCB=lm(TCB~TSB,data=TCB)
summary(lmTCB)
```

```
##
## Call:
## lm(formula = TCB ~ TSB, data = TCB)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1564 -0.4876 -0.0514  0.4794  3.8492
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.97118    0.44631   6.657 8.07e-10 ***
## TSB          0.80732    0.06128  13.175  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9853 on 124 degrees of freedom
## Multiple R-squared:  0.5833, Adjusted R-squared:  0.5799
## F-statistic: 173.6 on 1 and 124 DF,  p-value: < 2.2e-16
```

```
#Draw scatter plot
car::scatterplot(TCB~TSB,data=TCB)
#Normality test
shapiro.test(lmTCB$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  lmTCB$residuals
## W = 0.96395, p-value = 0.001945
```

```
#Independent test
TSA::runs(lmTCB$residuals)
```
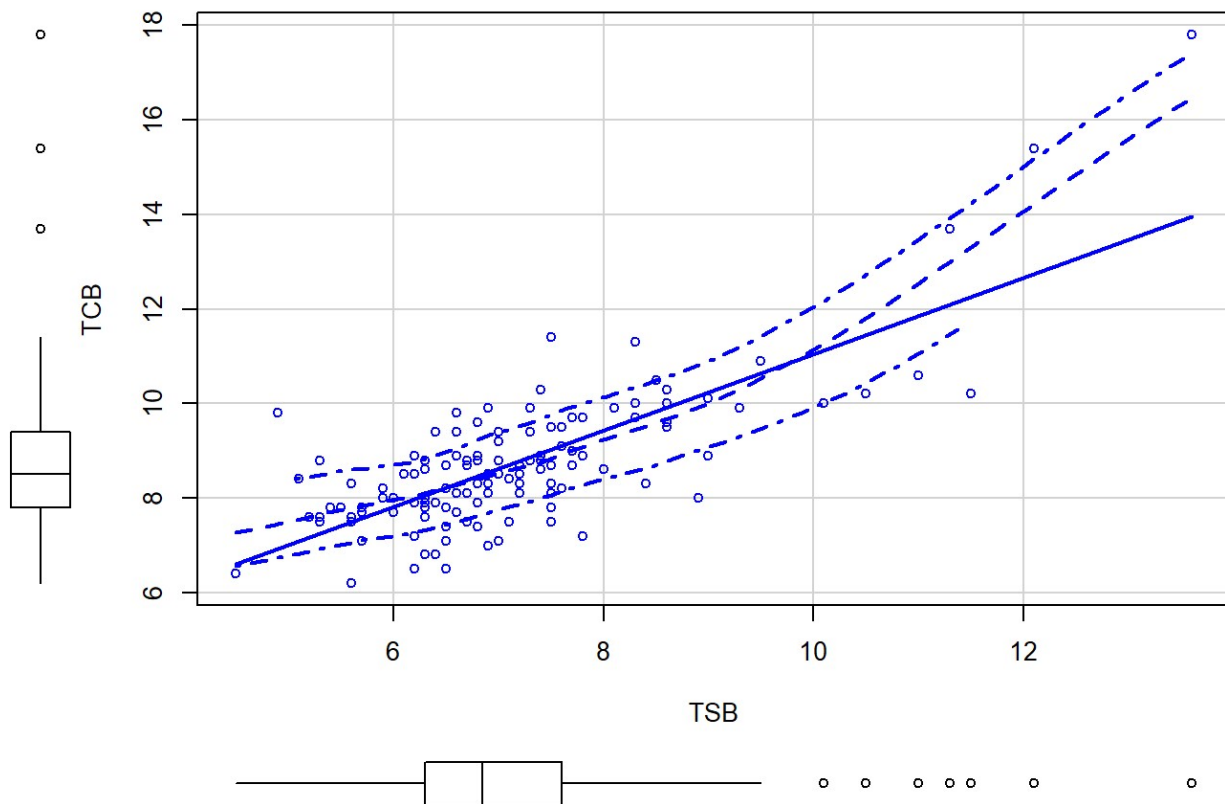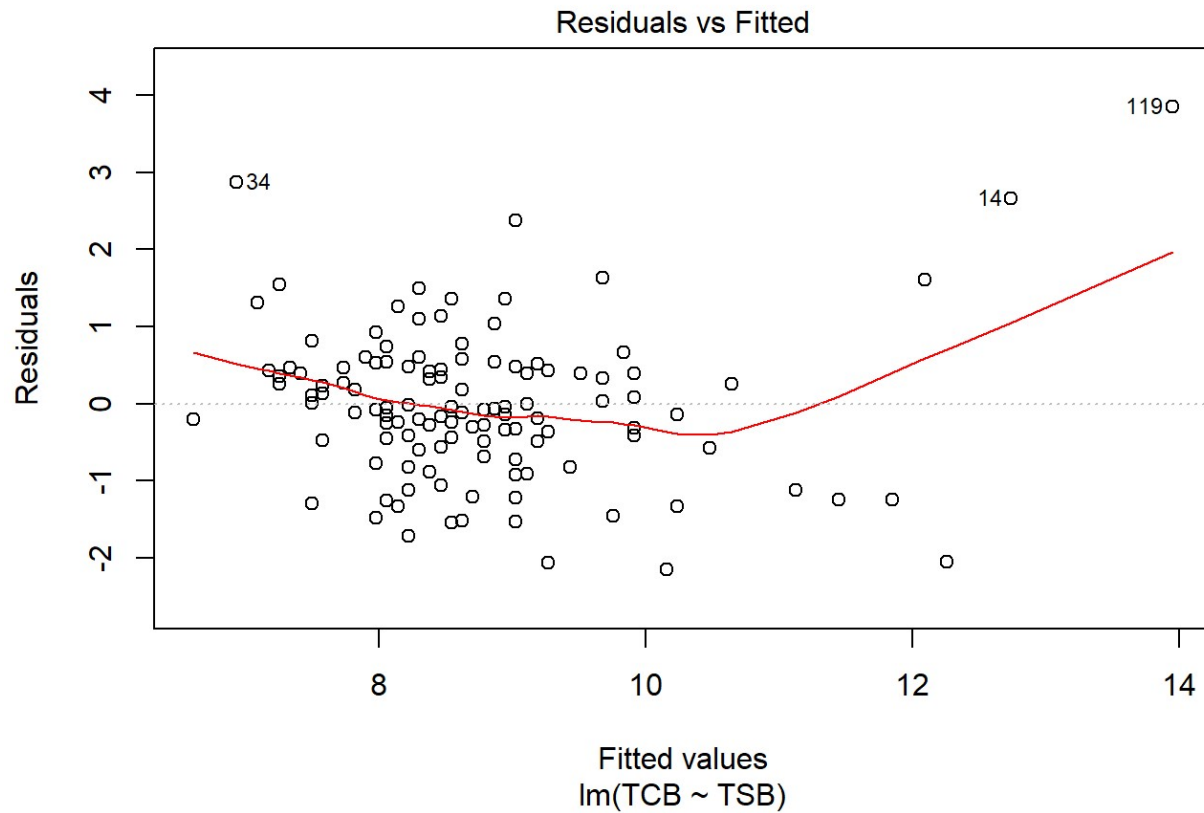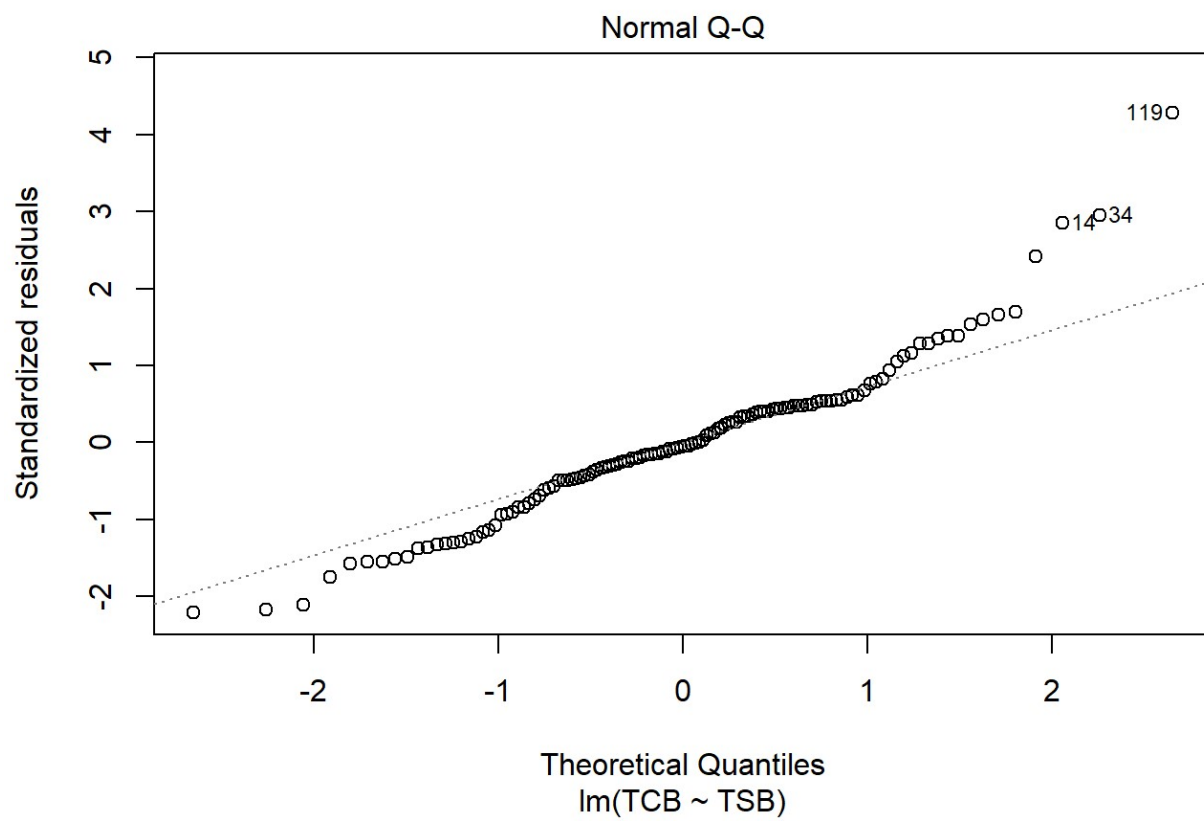
```
## $pvalue
## [1] 0.823
##
## $observed.runs
## [1] 62
##
## $expected.runs
## [1] 63.74603
##
## $n1
## [1] 67
##
## $n2
## [1] 59
##
## $k
## [1] 0
```
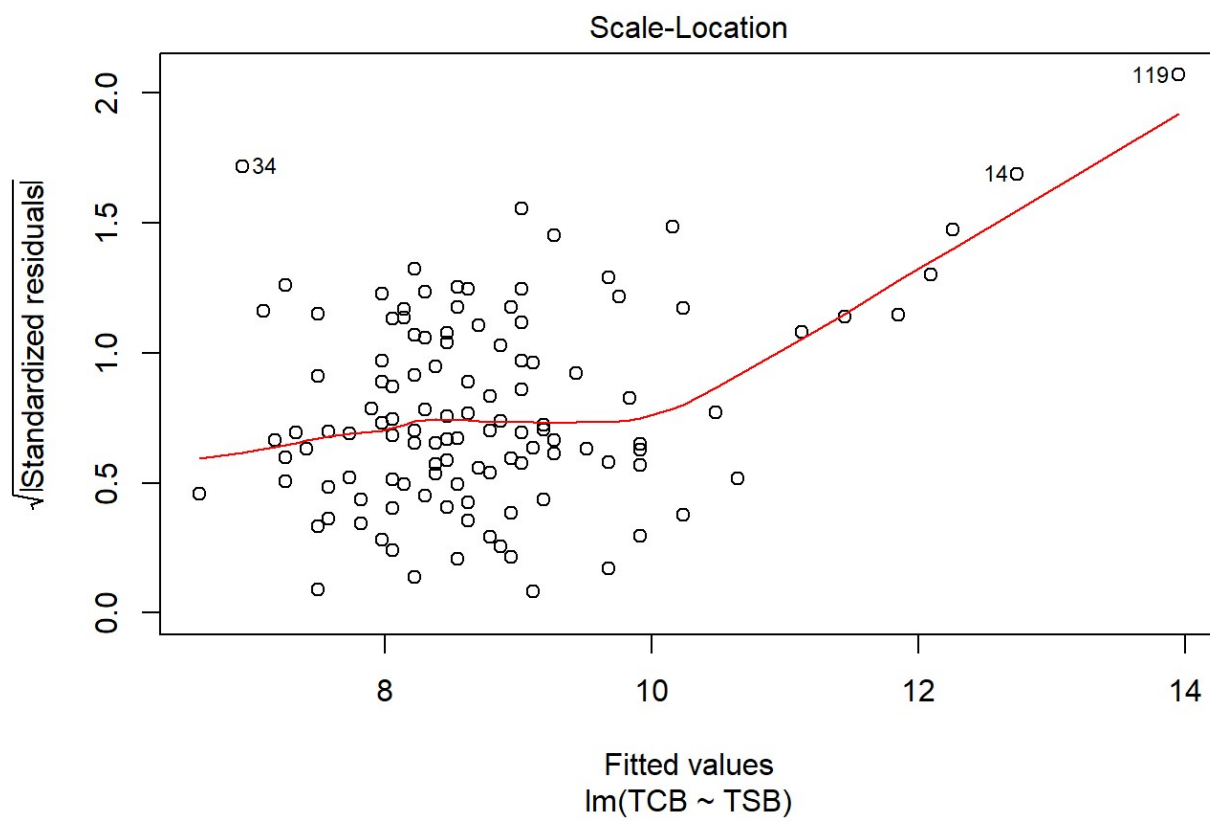
```
#Constant variance test
car::ncvTest(lmTCB)
```
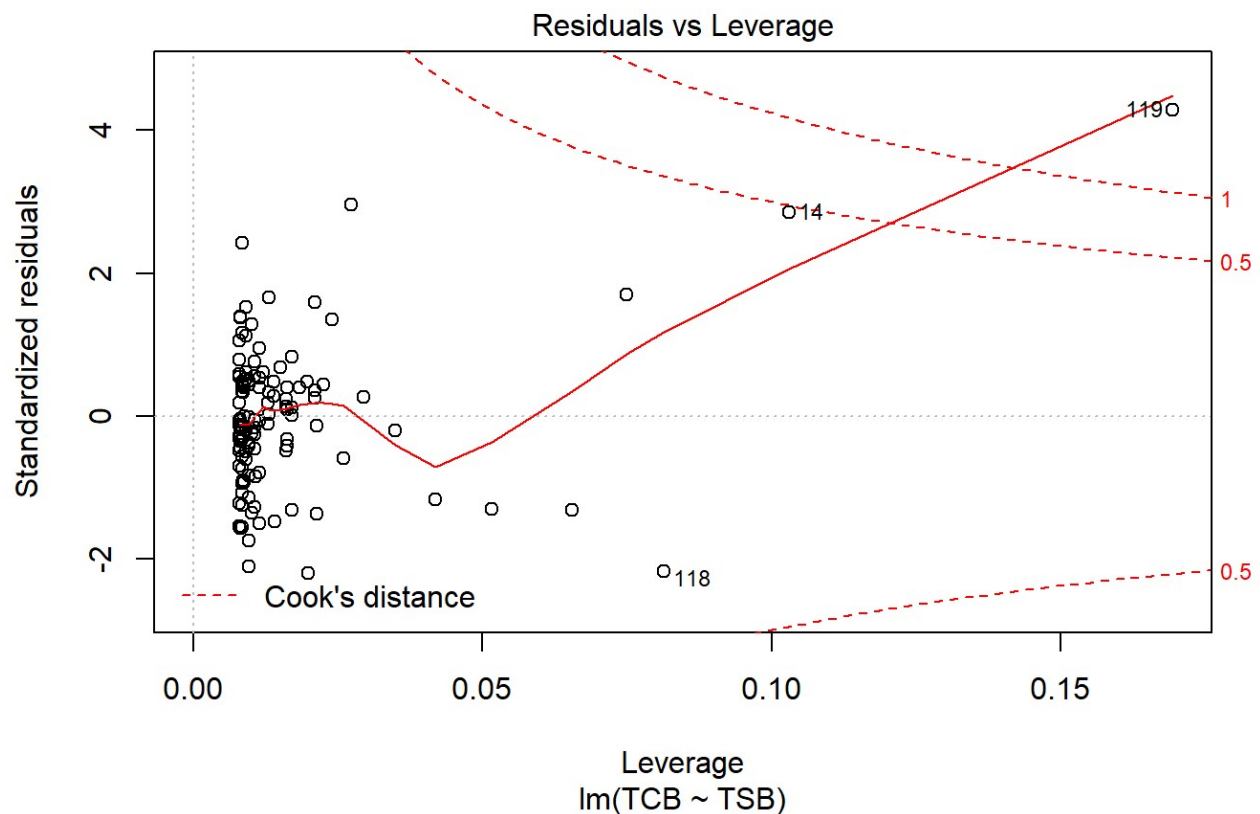
```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 48.88869, Df = 1, p = 2.7091e-12
```

```
#Draw the plot test for normality and constant variance
plot(lmTCB)
```

Residuals vs Fitted

Fitted values
lm(TCB ~ TSB)

Normal Q-Q

Standardized residuals

Theoretical Quantiles
lm(TCB ~ TSB)

Scale-Location

√|Standardized residuals|

Fitted values
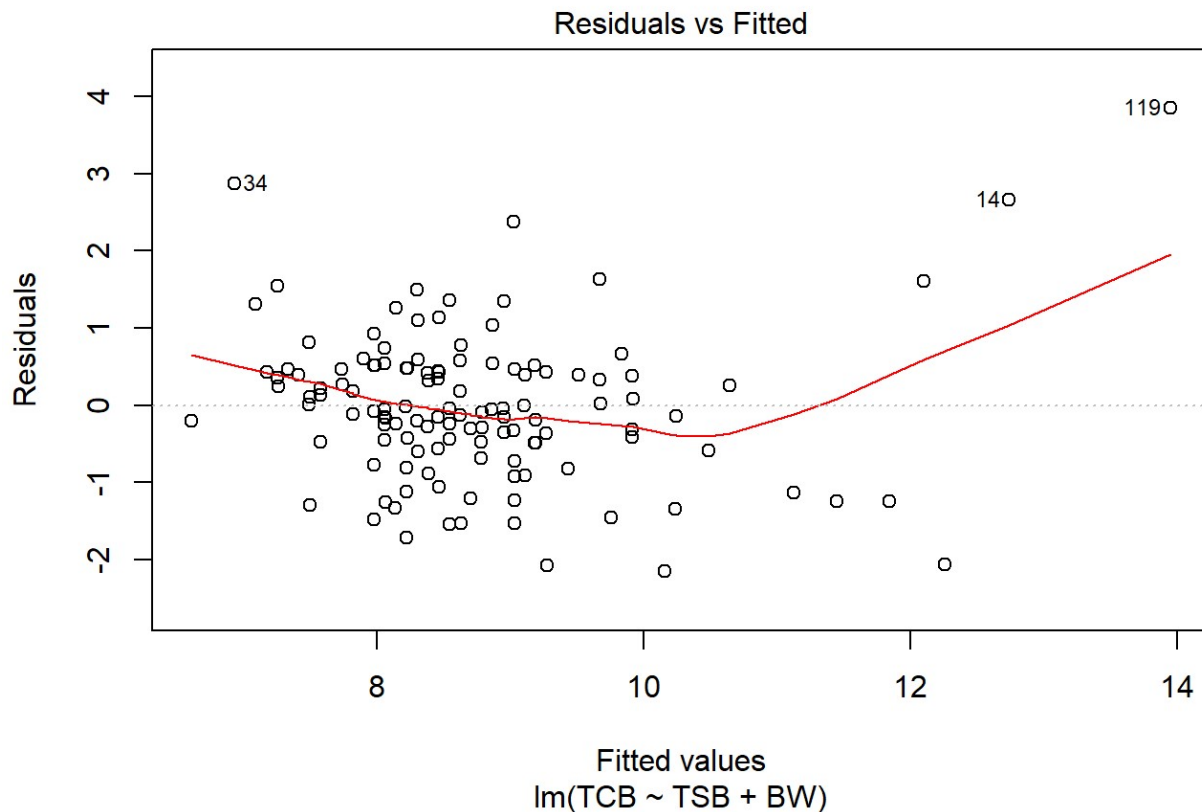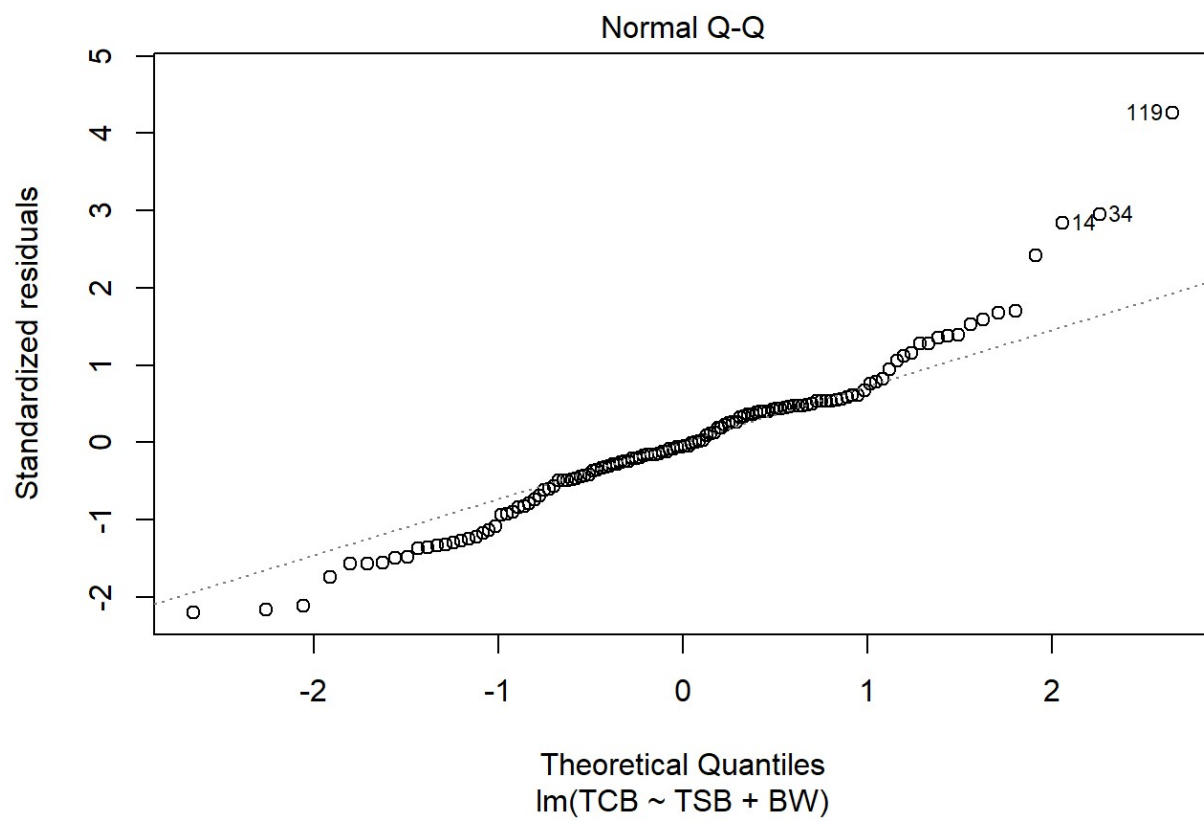lm(TCB ~ TSB)

Leverage
lm(TCB ~ TSB)

Based on the plot we can see both normality and variance test are satisfied. In the QQ plot most of the points are in the linear. There are few outliers. In the residual plot most of points have constant variance.
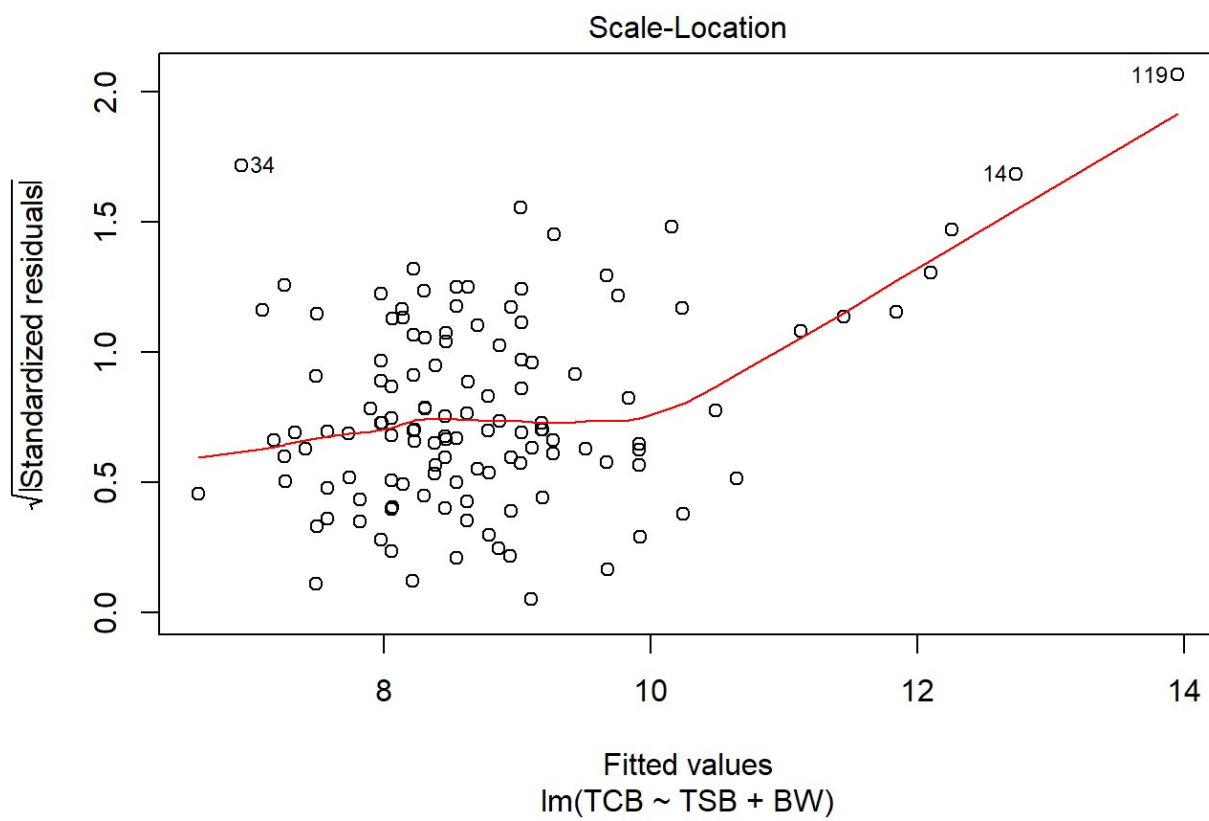
```
#Test whether Birthweight affect correlation
lmBW=lm(TCB~TSB+BW,data=TCB)
summary(lmBW)
```
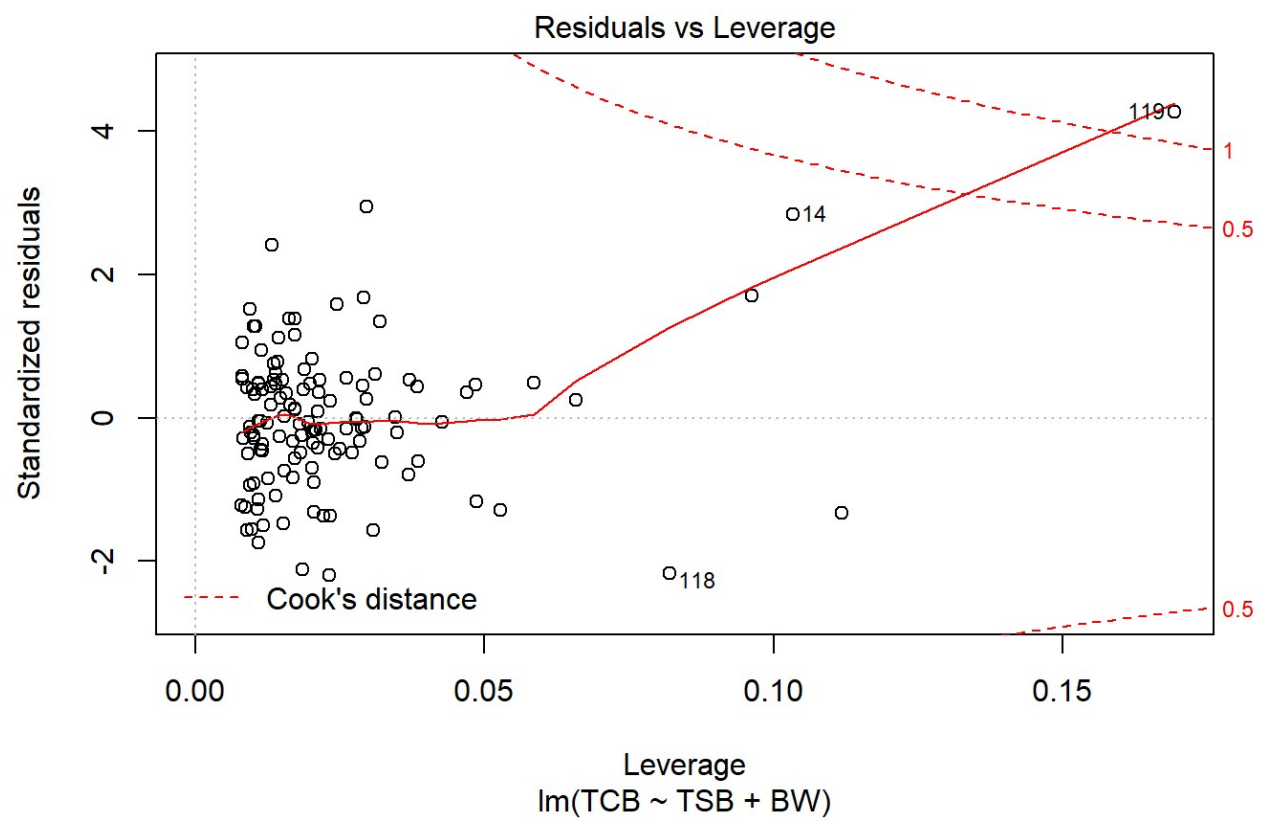
```
##
## Call:
## lm(formula = TCB ~ TSB + BW, data = TCB)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1546 -0.4866 -0.0507  0.4738  3.8493
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.949e+00  8.291e-01   3.557 0.000533 ***
## TSB         8.069e-01  6.287e-02  12.835  < 2e-16 ***
## BW          7.514e-06  2.393e-04   0.031 0.975006
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9893 on 123 degrees of freedom
## Multiple R-squared:  0.5833, Adjusted R-squared:  0.5765
## F-statistic: 86.09 on 2 and 123 DF,  p-value: < 2.2e-16
```

```
#Draw the plot of the new regression model include weight
plot(lmBW)
```


Residuals vs Fitted

lm(TCB ~ TSB + BW)

Normal Q-Q

Standardized residuals

Theoretical Quantiles
lm(TCB ~ TSB + BW)

Scale-Location

√|Standardized residuals|

Fitted values
lm(TCB ~ TSB + BW)

Residuals vs Leverage

lm(TCB ~ TSB + BW)

```
#Make birthweight as a coefficient of TSB
lmbw=lm(TCB~TSB+TSB*BW,data=TCB)
summary(lmbw)
```

```
##
## Call:
## lm(formula = TCB ~ TSB + TSB * BW, data = TCB)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3456 -0.4706 -0.0159  0.5373  3.3805
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.7492926  4.0153990   3.424 0.000841 ***
## TSB         -0.7052367  0.5541127  -1.273 0.205533
## BW          -0.0031329  0.0011673  -2.684 0.008286 **
## TSB:BW       0.0004376  0.0001594   2.746 0.006950 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.964 on 122 degrees of freedom
## Multiple R-squared:  0.6076, Adjusted R-squared:  0.5979
## F-statistic: 62.96 on 3 and 122 DF,  p-value: < 2.2e-16
```

```
#Normality test
shapiro.test(lmBW$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  lmBW$residuals
## W = 0.96388, p-value = 0.001919
```

```
#Independent test
TSA::runs(lmBW$residuals)
```

```
## $pvalue
## [1] 0.823
##
## $observed.runs
## [1] 62
##
## $expected.runs
## [1] 63.74603
##
## $n1
## [1] 67
##
## $n2
## [1] 59
##
## $k
## [1] 0
```

```
#Constant variance test
car::ncvTest(lmBW)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 48.83095, Df = 1, p = 2.79e-12
```
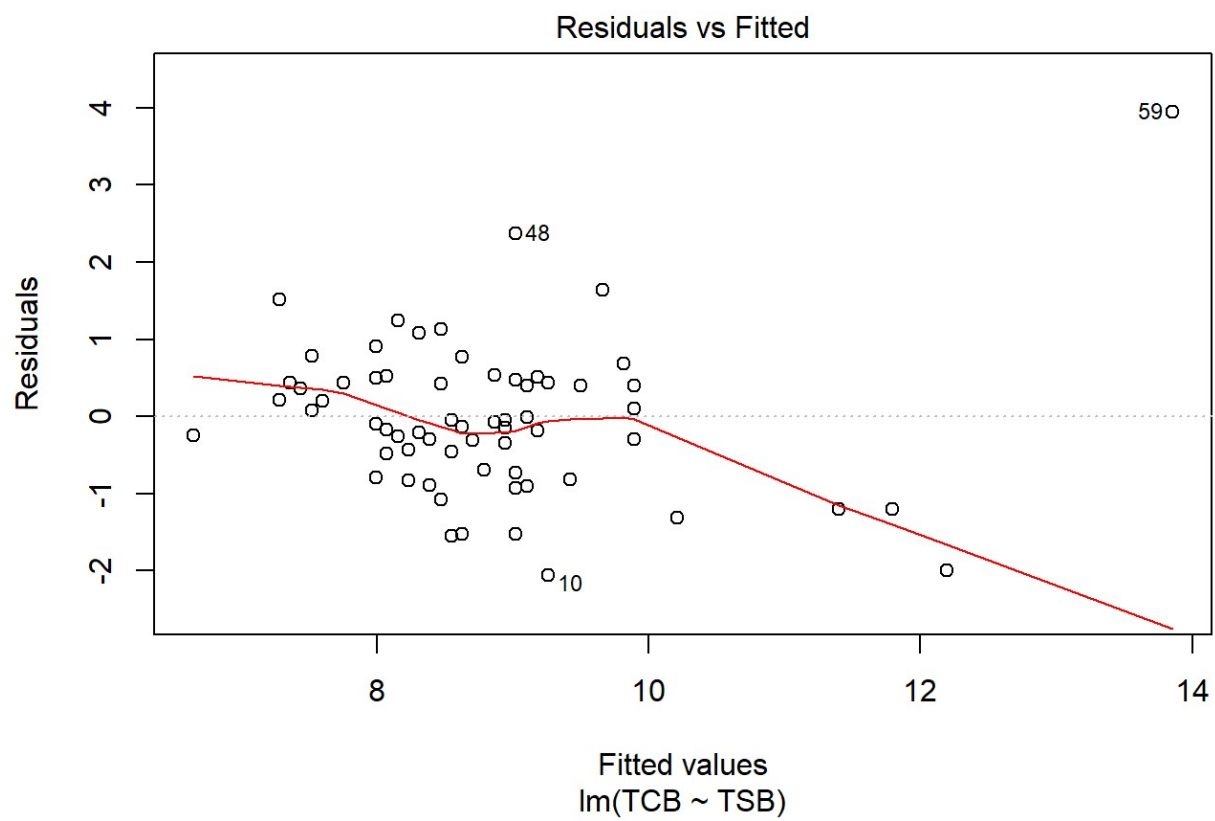
```
#Male
Male <- read_excel("C:/Users/zhang/Desktop/individual project/Male.xlsx")
Male
```

```
## # A tibble: 66 x 5
##     `MR No`   TCB   TSB Gender    BW
##       <dbl> <dbl> <dbl> <chr>  <dbl>
##  1 1493066   8.1   7.5 m       3455
##  2 1493254   9.4   6.6 m       3575
##  3 1493728  10.5   8.5 m       3630
##  4 1493668   8.3   7.5 m       3665
##  5 1494636  11.3   8.3 m       2840
##  6 1494779  10.3   8.6 m       3584
##  7 1494914   9.7   7.8 m       3095
##  8 1495271   7.5   5.3 m       4075
##  9 1495618   8.9   9   m       3590
## 10 1496357   7.2   7.8 m       3725
## # ... with 56 more rows
```
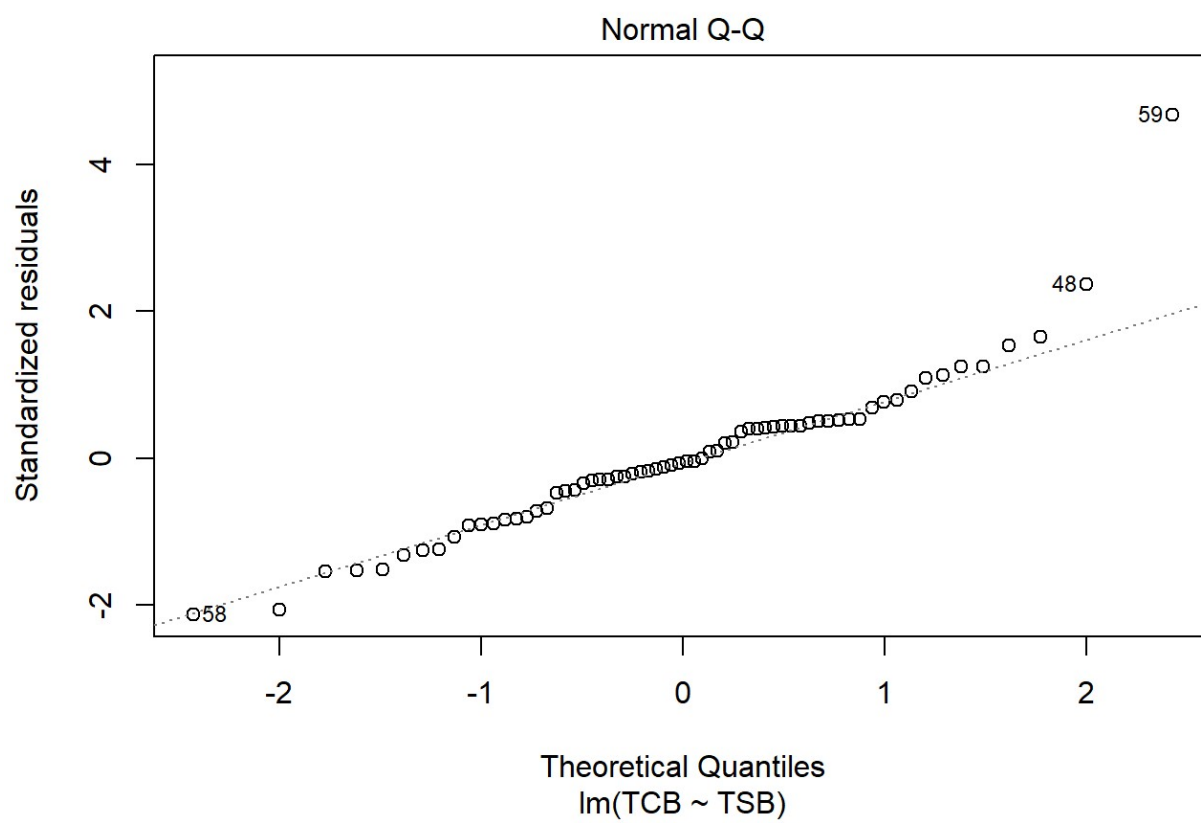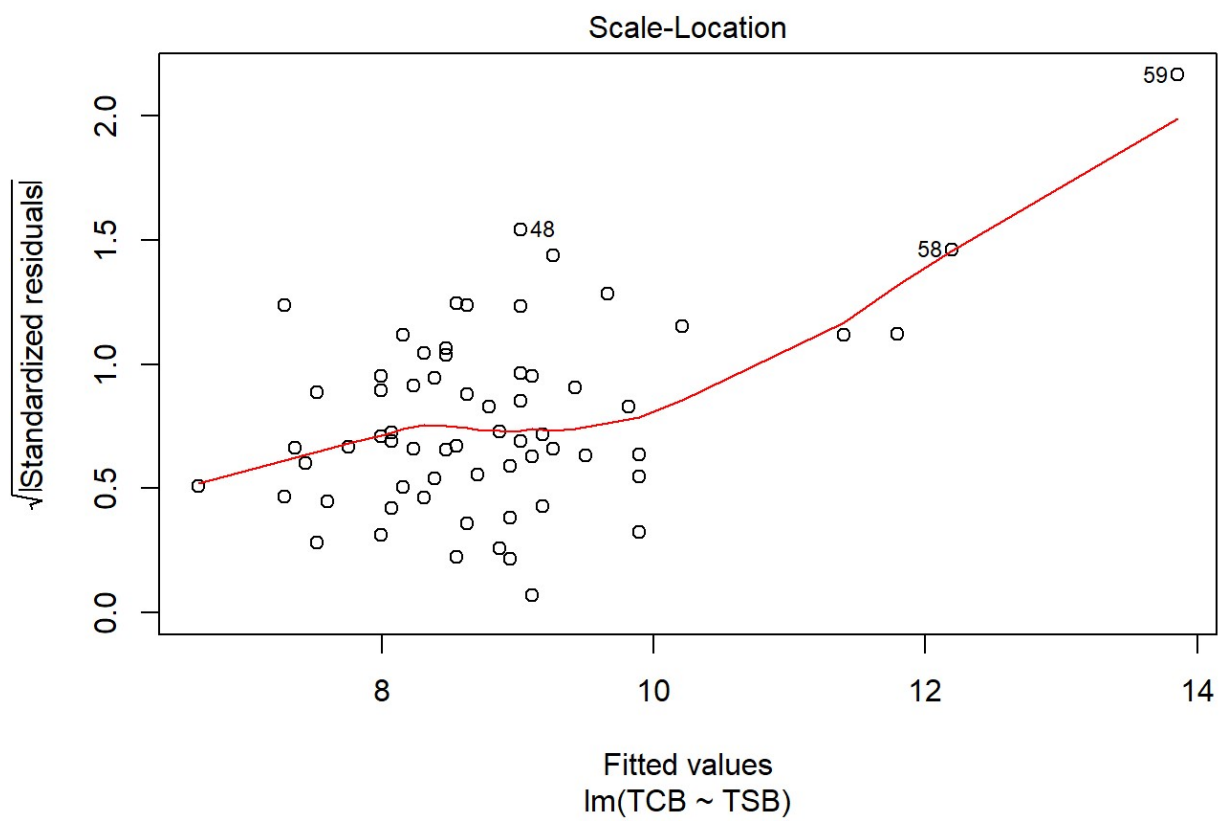
```
#Make a regression model include only male
lmMale=lm(TCB~TSB,data=Male)
summary(lmMale)
```
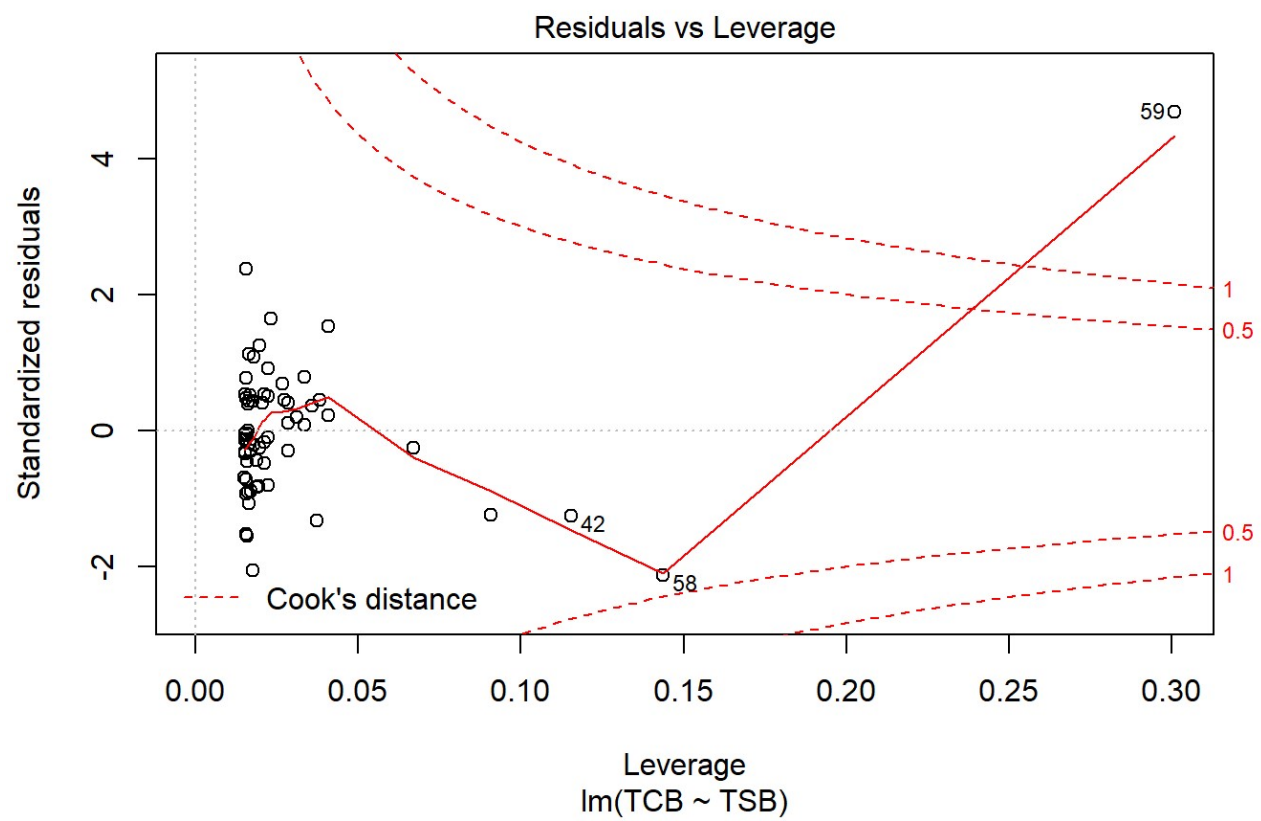
```
##
## Call:
## lm(formula = TCB ~ TSB, data = Male)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0632 -0.6353 -0.0592  0.4959  3.9465
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.09005    0.62181   4.969 5.30e-06 ***
## TSB          0.79143    0.08443   9.374 1.26e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.008 on 64 degrees of freedom
## Multiple R-squared:  0.5786, Adjusted R-squared:  0.572
## F-statistic: 87.87 on 1 and 64 DF,  p-value: 1.265e-13
```

```
#draw the plot test the correlation base on male.
plot(lmMale)
```

Residuals vs Fitted

Residuals

59

48

10

Fitted values
lm(TCB ~ TSB)

Normal Q-Q

lm(TCB ~ TSB)

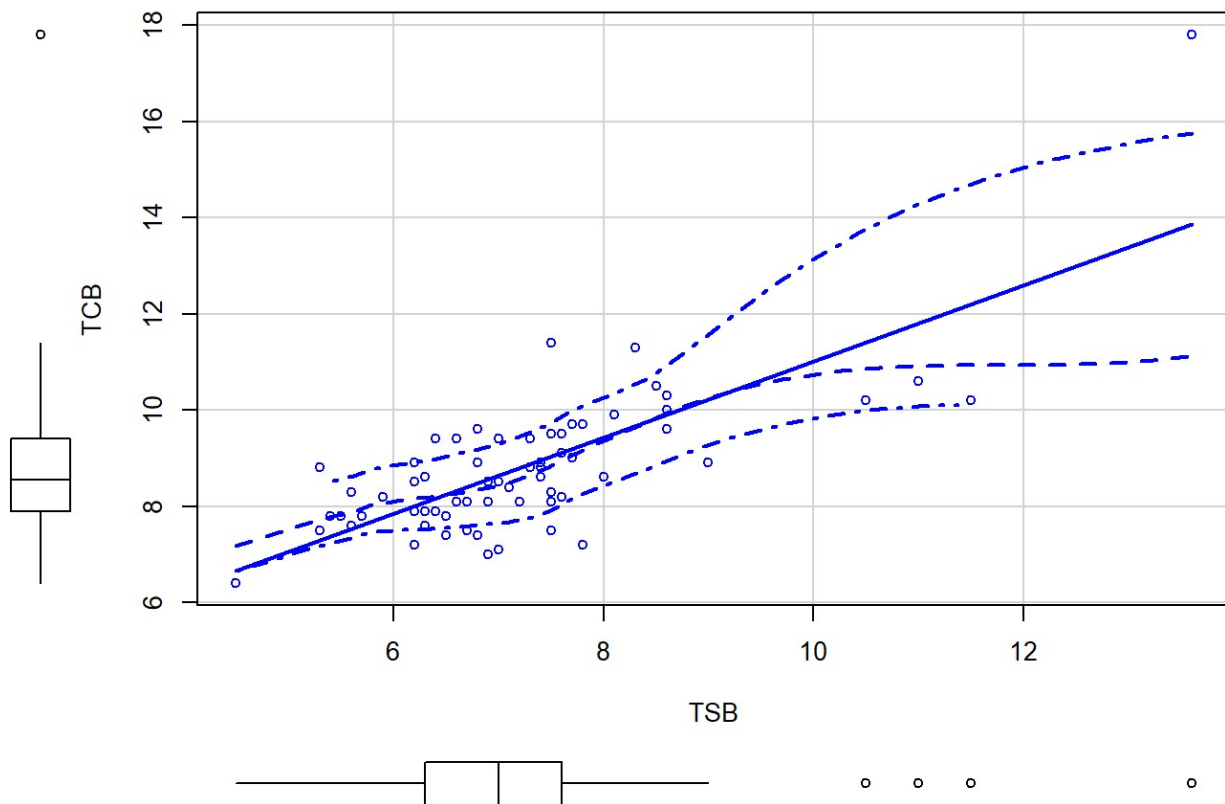**Residuals vs Leverage**

lm(TCB ~ TSB)

```
#Draw scatter plot
car::scatterplot(TCB~TSB,data=Male)
```

```
#Normality test
shapiro.test(lmMale$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  lmMale$residuals
## W = 0.94981, p-value = 0.009607
```

```
#Independent test
TSA::runs(lmMale$residuals)
```

```
## $pvalue
## [1] 0.759
##
## $observed.runs
## [1] 32
##
## $expected.runs
## [1] 33.72727
##
## $n1
## [1] 36
##
## $n2
## [1] 30
##
## $k
## [1] 0
```
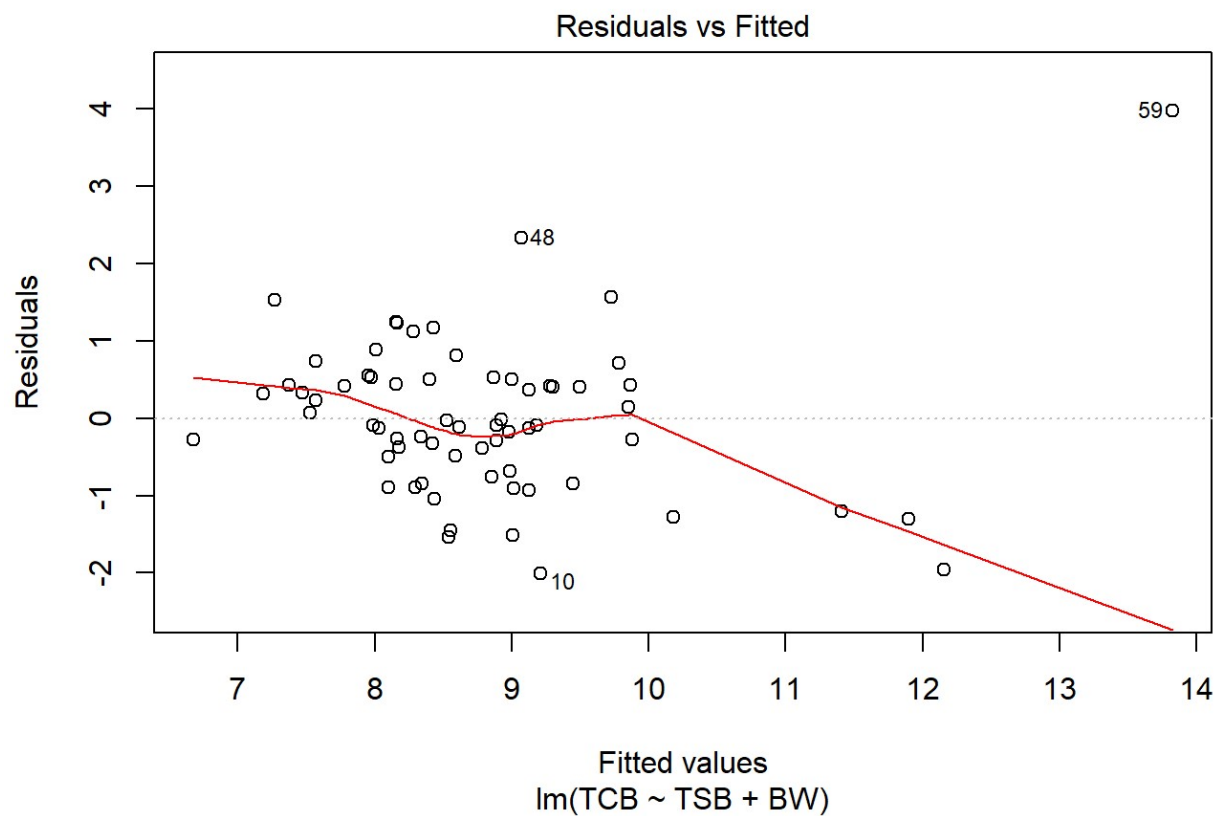
```
#Constant variance test
car::ncvTest(lmMale)
```
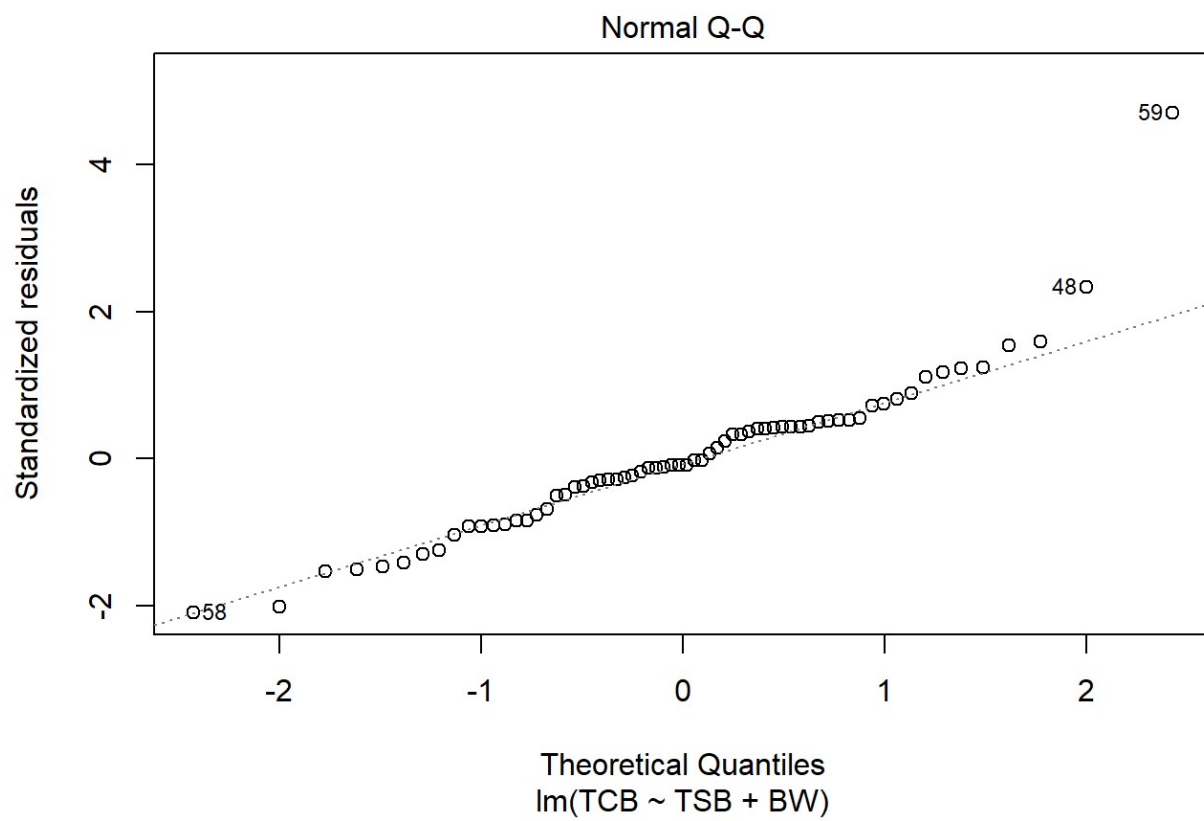
```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 55.45267, Df = 1, p = 9.5737e-14
```
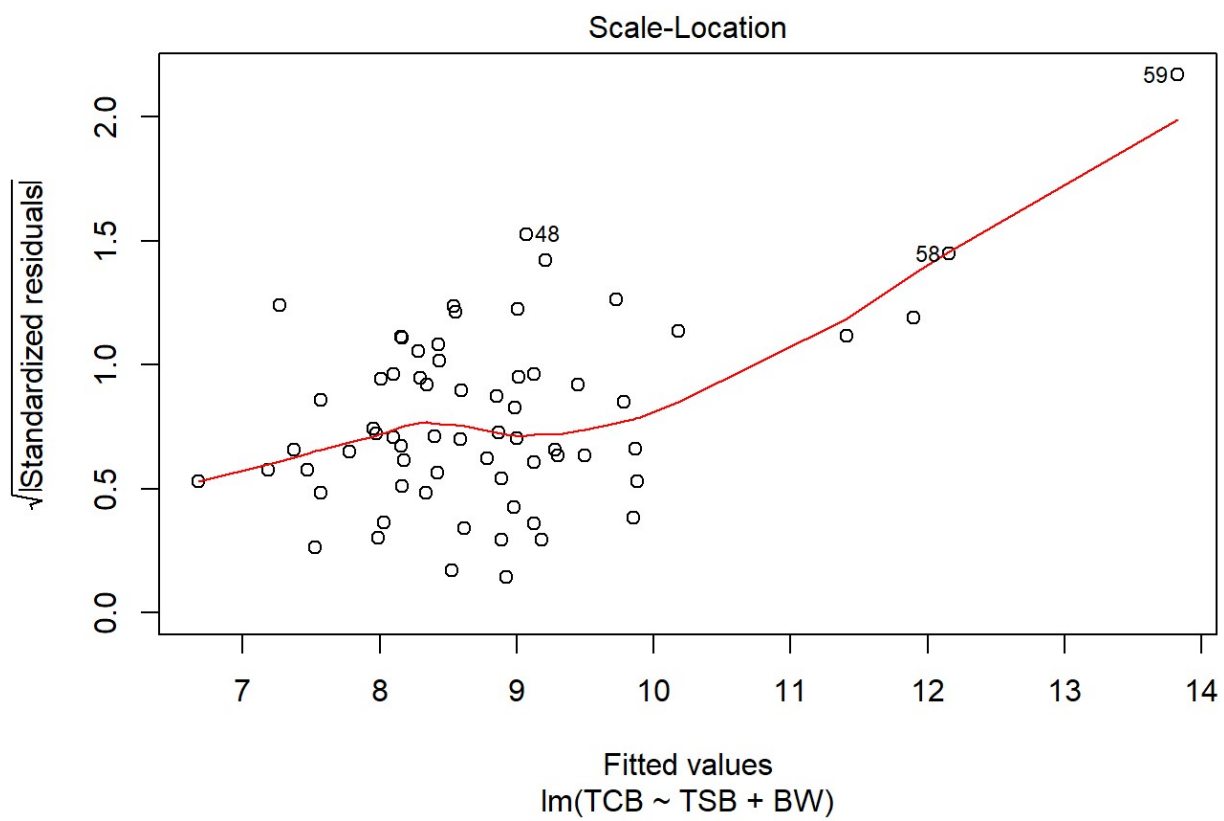
```
#Based on male,one more factor add in how the model will change.
lmmW=lm(TCB~TSB+BW,data=Male)
summary(lmmW)
```
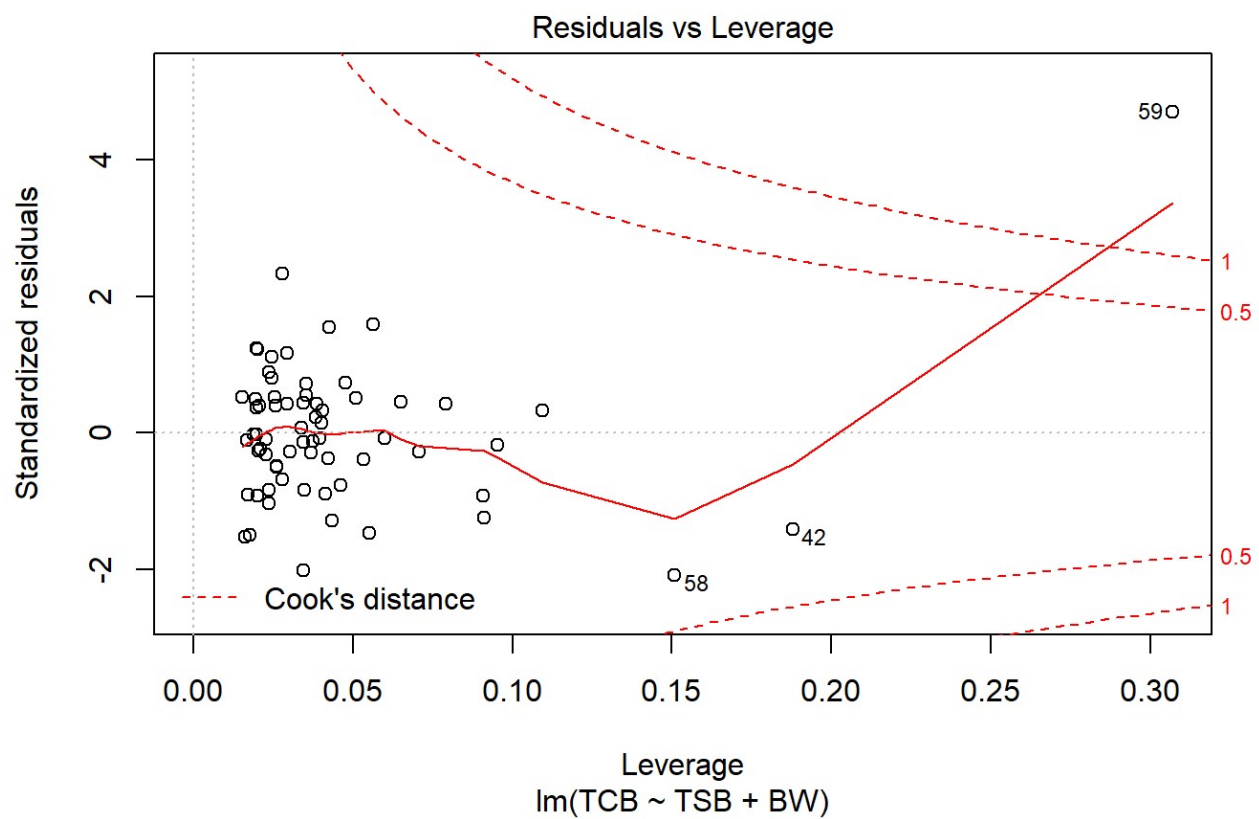
```
## 
## Call:
## lm(formula = TCB ~ TSB + BW, data = Male)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0133 -0.6385 -0.0860  0.4841  3.9766
## 
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.5301254  1.3203740   2.674  0.00955 **
## TSB          0.7928643  0.0850848   9.319 1.83e-13 ***
## BW          -0.0001345  0.0003554  -0.379  0.70630
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.015 on 63 degrees of freedom
## Multiple R-squared:  0.5795, Adjusted R-squared:  0.5662
## F-statistic: 43.42 on 2 and 63 DF,  p-value: 1.404e-12
```

```
plot(lmmW)
```


Residuals vs Fitted

lm(TCB ~ TSB + BW)

Normal Q-Q

Standardized residuals

Theoretical Quantiles
lm(TCB ~ TSB + BW)

Scale-Location

√|Standardized residuals|

Fitted values
lm(TCB ~ TSB + BW)

Residuals vs Leverage

Leverage
lm(TCB ~ TSB + BW)

```
#Normality test
shapiro.test(lmmW$residuals)
```

```
##
##   Shapiro-Wilk normality test
##
## data:  lmmW$residuals
## W = 0.94826, p-value = 0.008038
```

```
#Independent test
TSA::runs(lmmW$residuals)
```

```
## $pvalue
## [1] 0.759
##
## $observed.runs
## [1] 32
##
## $expected.runs
## [1] 33.72727
##
## $n1
## [1] 36
##
## $n2
## [1] 30
##
## $k
## [1] 0
```

```
#Constant variance test
car::ncvTest(lmmW)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 56.12118, Df = 1, p = 6.8139e-14
```

```
#Make birthweight as a coefficient of TSB
lmMW=lm(TCB~TSB+(TSB*BW),data=Male)
summary(lmMW)
```

```
##
## Call:
## lm(formula = TCB ~ TSB + (TSB * BW), data = Male)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.2606 -0.5664 -0.0274  0.5195  3.5090
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.3760680  6.0761955   1.872   0.0659 .
## TSB         -0.2800494  0.8156834  -0.343   0.7325
## BW          -0.0024451  0.0017825  -1.372   0.1751
## TSB:BW       0.0003156  0.0002386   1.322   0.1909
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.009 on 62 degrees of freedom
## Multiple R-squared:  0.5911, Adjusted R-squared:  0.5713
## F-statistic: 29.87 on 3 and 62 DF,  p-value: 4.515e-12
```
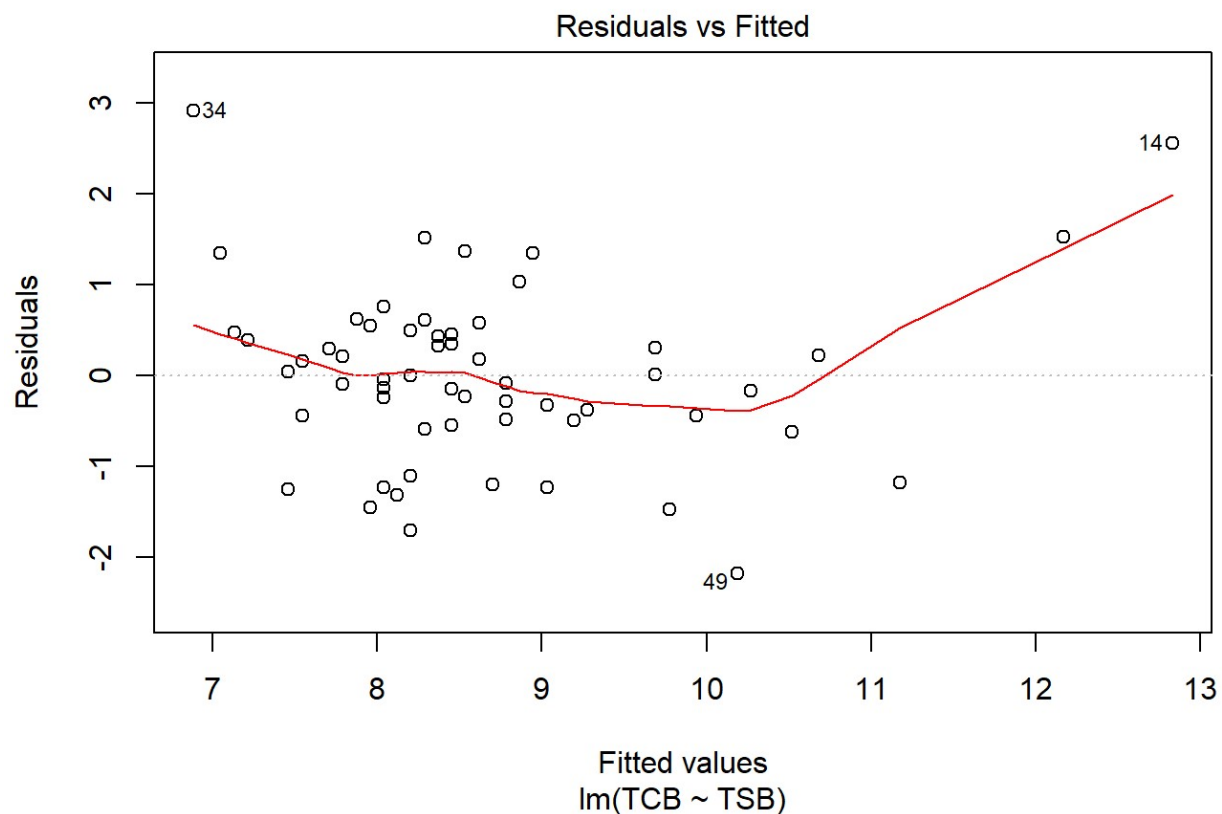
```
#Female
#Import the data only relate to female
Female<-read_excel("C:/Users/zhang/Desktop/individual project/Female.xlsx")
Female
```

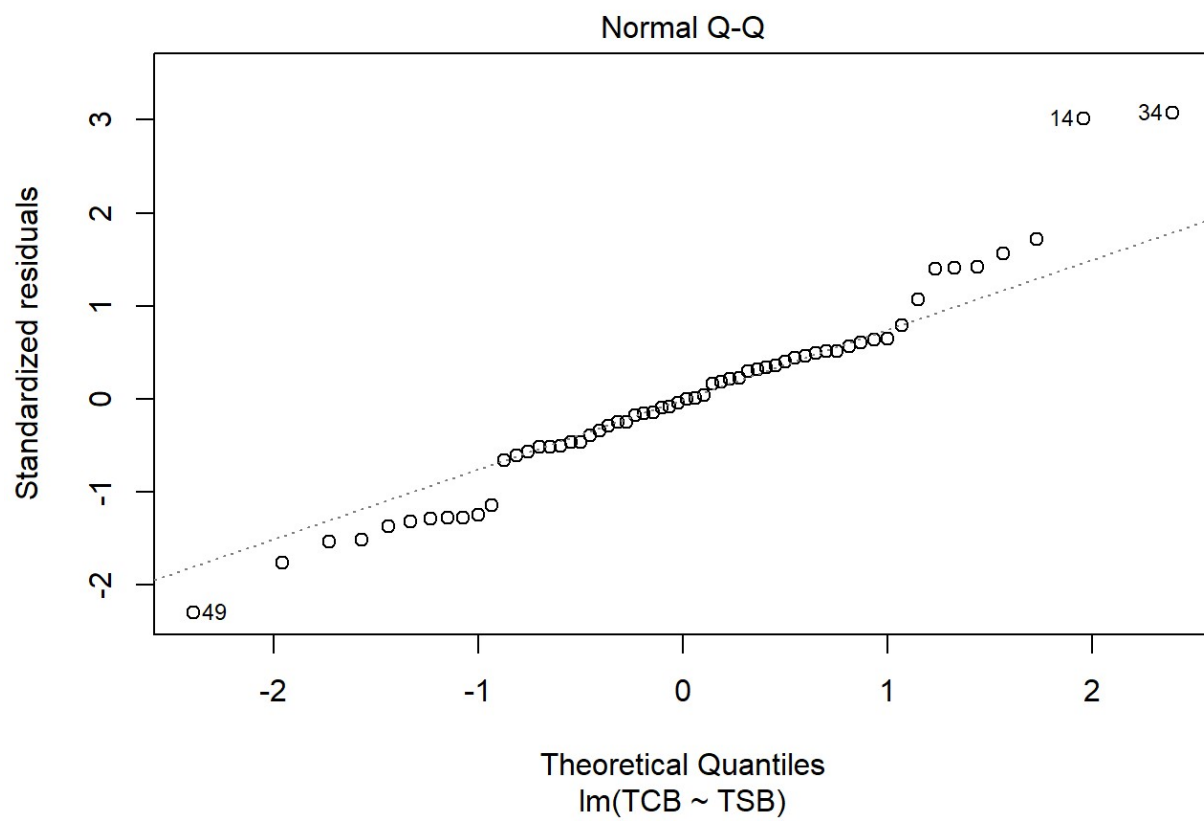```
A tibble: 60 x 5
`MR No`   TCB   TSB   BW Gender
<dbl> <dbl> <dbl> <dbl> <chr>
1 1492989   8.3   7.2 2730 f
2 1493428 10.1   9   3760 f
3 1493611 13.7 11.3 4130 f
4 1493706   8.9   6.8 2455 f
5 1493704   8.7   7.7 3320 f
6 1493869   6.8   6.3 3290 f
7 1494140   9.2   7   3245 f
8 1494738   8.8   6.7 3215 f
9 1495268   8.7   7.5 2735 f
10 1496182   8.7   6.5 4180 f
... with 50 more rows
```
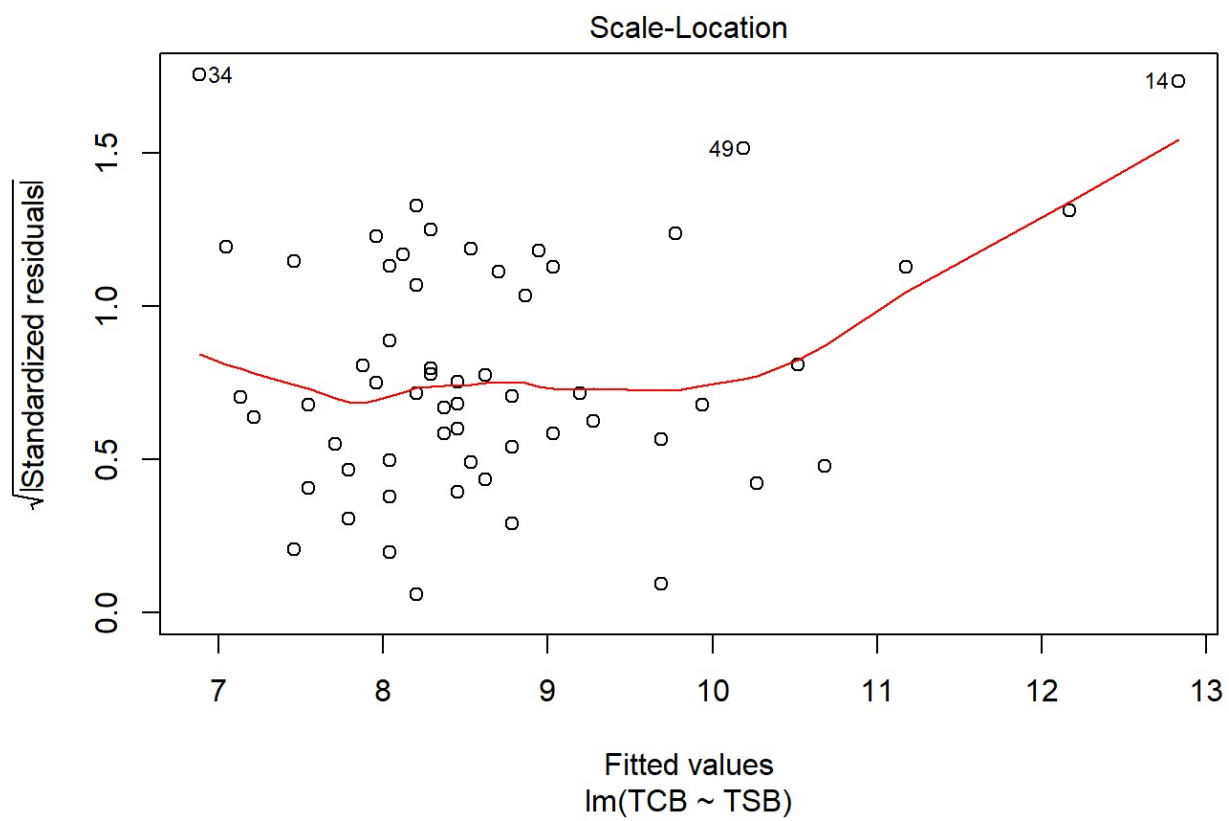
```
#Regression model only include female
lmFemale=lm(TCB~TSB,data=Female)
summary(lmFemale)
```
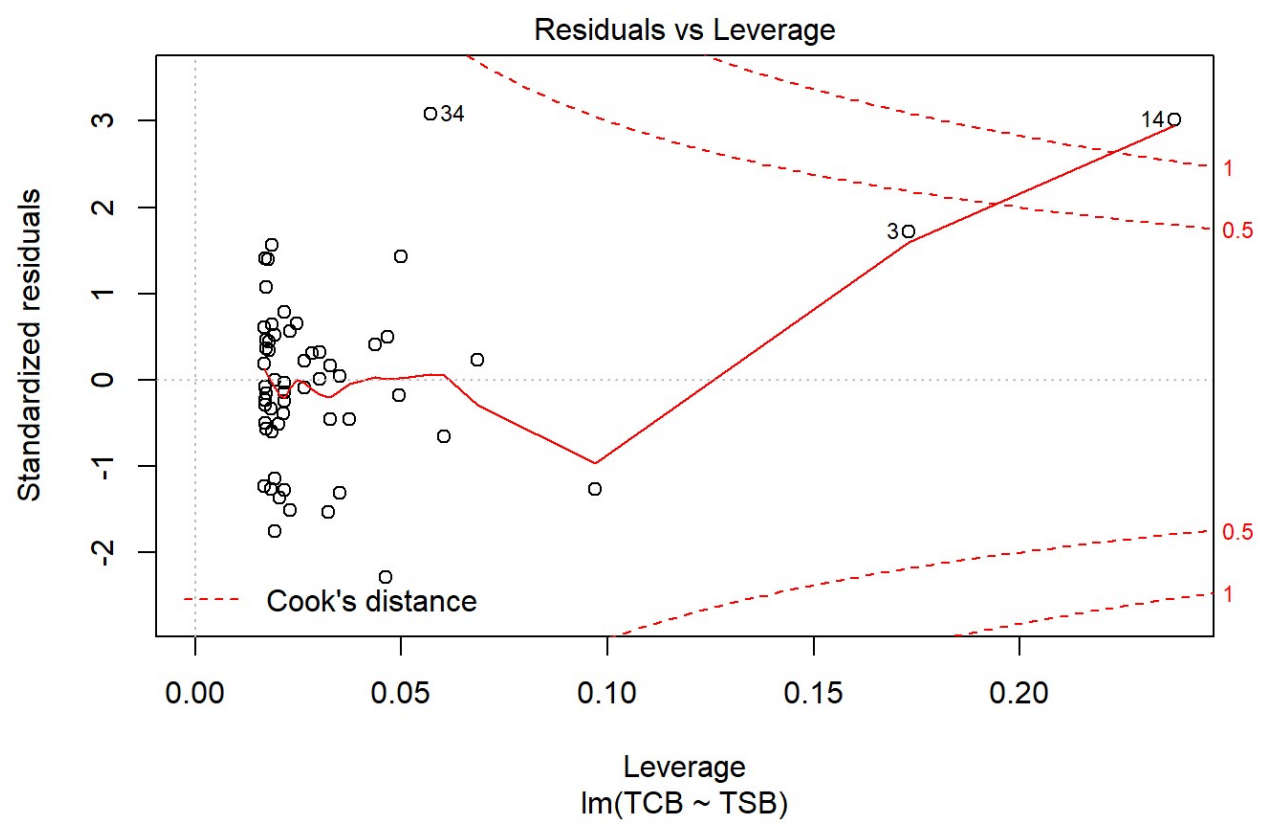
```
##
Call:
lm(formula = TCB ~ TSB, data = Female)
##
Residuals:
Min   1Q   Median   3Q   Max
-2.18742 -0.49544 -0.02079 0.47753 2.91919
##
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.83021   0.65416   4.326 6.05e-05 ***
TSB   0.82665   0.09095   9.090 9.44e-13 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
Residual standard error: 0.9757 on 58 degrees of freedom
Multiple R-squared: 0.5875, Adjusted R-squared: 0.5804
F-statistic: 82.62 on 1 and 58 DF, p-value: 9.439e-13
```

```
#Draw the plot test for normality and constant variance
plot(lmFemale)
```


Residuals vs Fitted

lm(TCB ~ TSB)

Normal Q-Q

Standardized residuals

Theoretical Quantiles
lm(TCB ~ TSB)

Scale-Location

√|Standardized residuals|

Fitted values
lm(TCB ~ TSB)

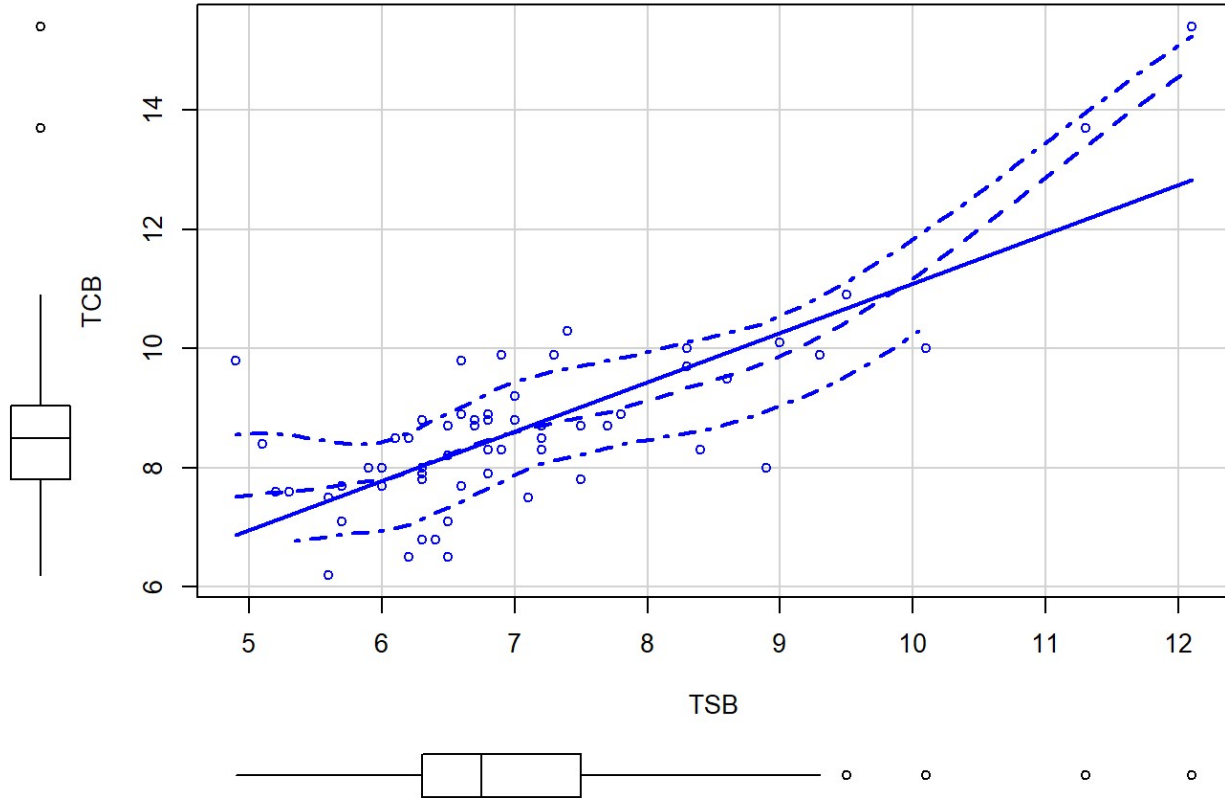Residuals vs Leverage

lm(TCB ~ TSB)

```
#Draw scatter plot
car::scatterplot(TCB~TSB,data=Female)
```

```
#Normality test
shapiro.test(lmFemale$residuals)
```

```
##
##   Shapiro-Wilk normality test
##
## data:  lmFemale$residuals
## W = 0.96834, p-value = 0.1209
```

```
#Independent test
TSA::runs(lmFemale$residuals)
```

```
## $pvalue
## [1] 0.905
##
## $observed.runs
## [1] 31
##
## $expected.runs
## [1] 30.96667
##
## $n1
## [1] 31
##
## $n2
## [1] 29
##
## $k
## [1] 0
```

```
#Constant variance test
car::ncvTest(lmFemale)
```
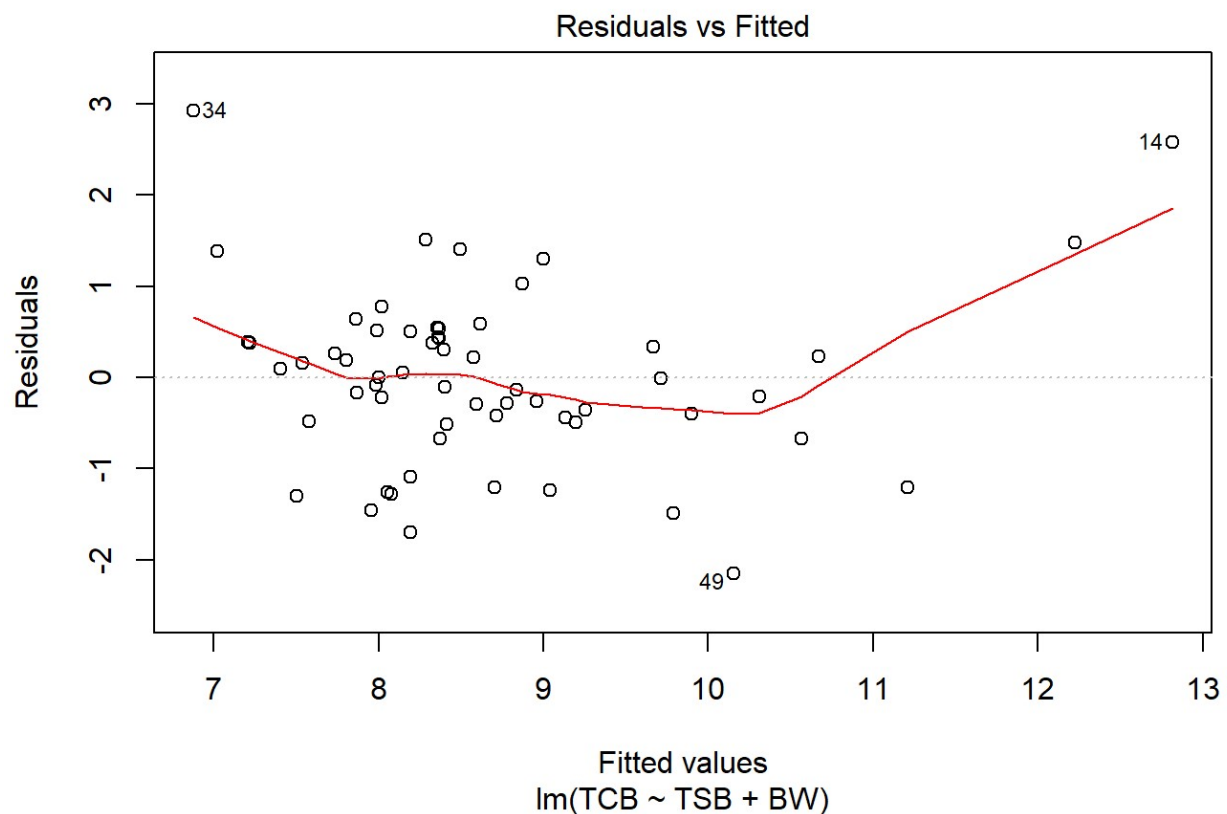
```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 3.857279, Df = 1, p = 0.049531
```
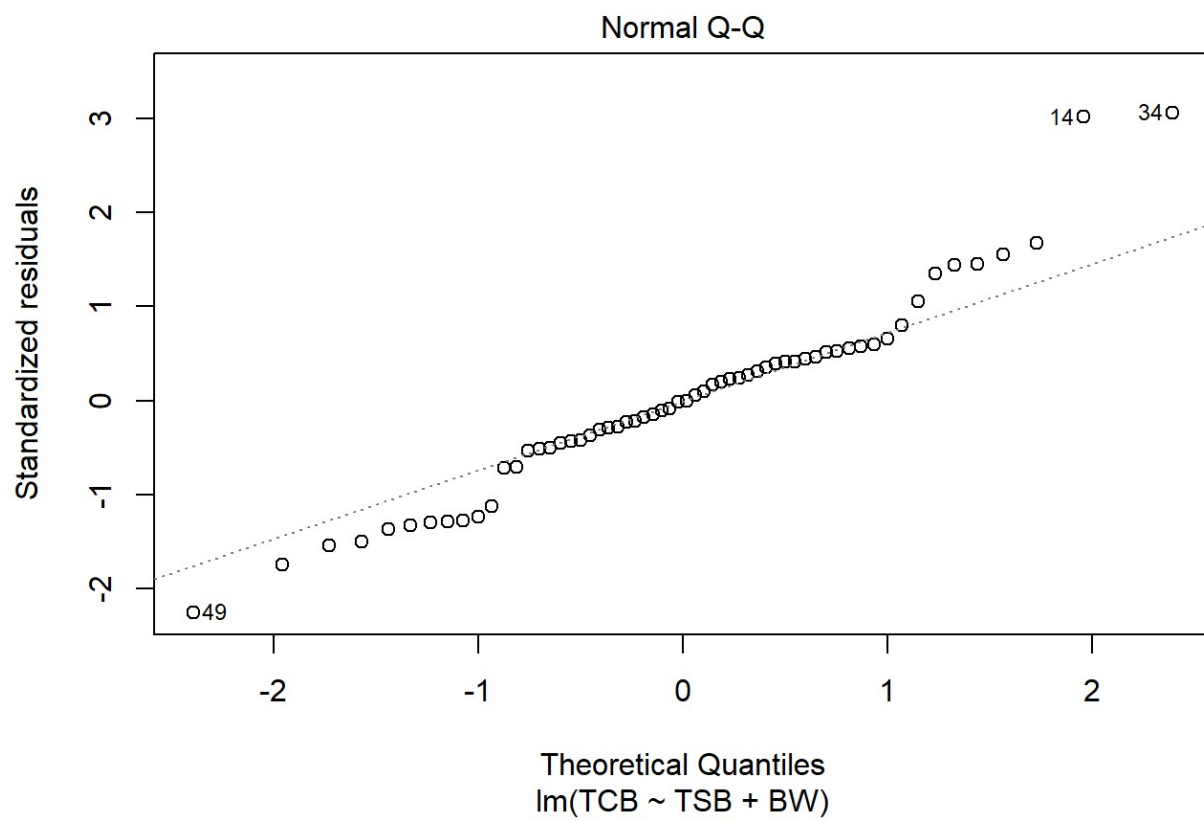
```
#Based on female,one more factor add in how the model will change.
lmFW=lm(TCB~TSB+BW,data=Female)
summary(lmFW)
```
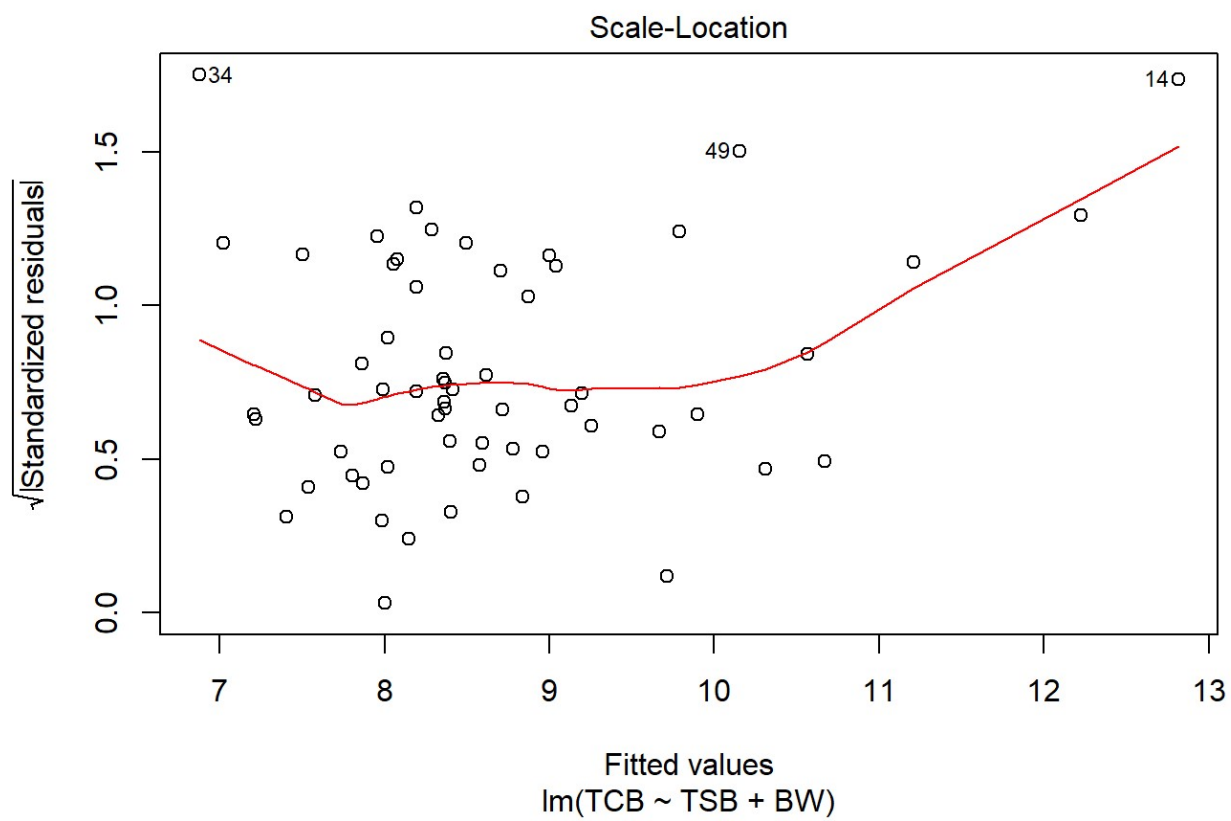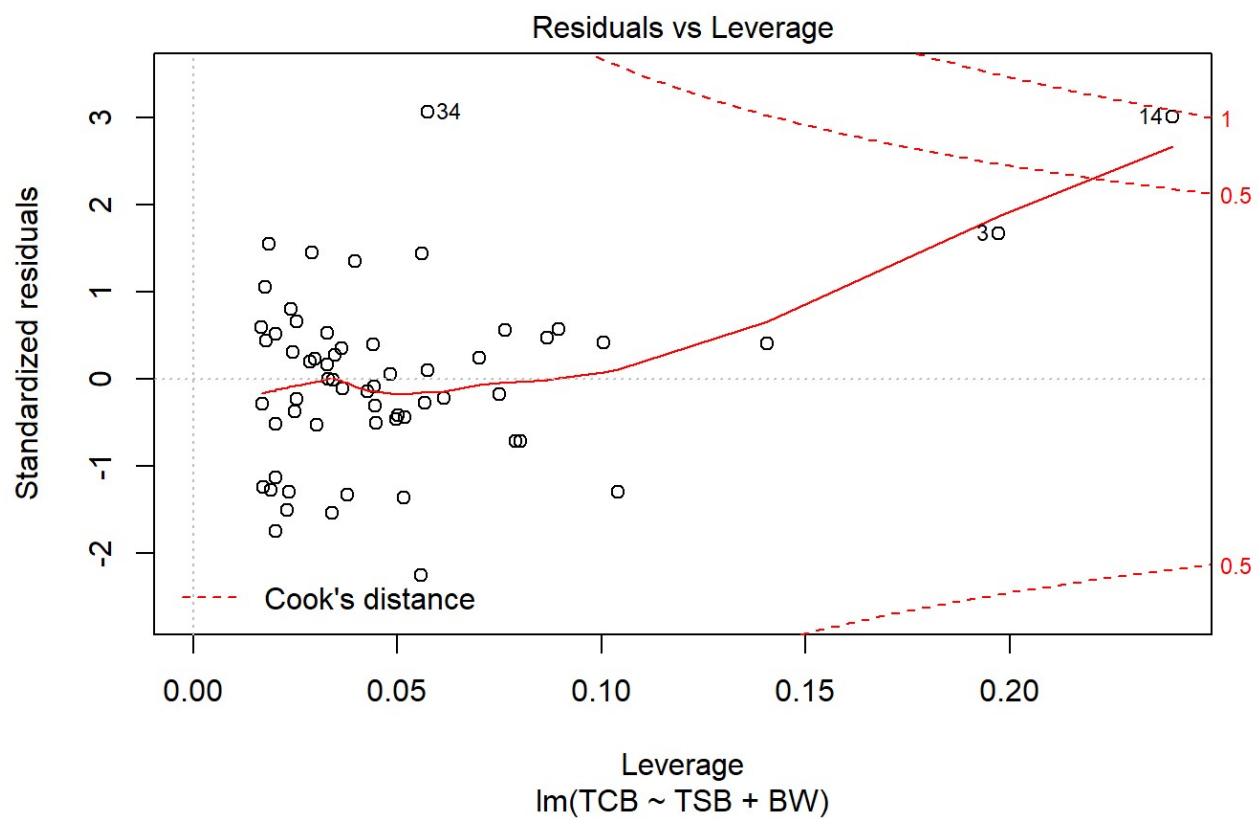
```
##
## Call:
## lm(formula = TCB ~ TSB + BW, data = Female)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -2.15337 -0.48445 -0.00746  0.45673  2.92389
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.5222158  1.0905341   2.313   0.0244 *
## TSB         0.8139588  0.0983851   8.273 2.42e-11 ***
## BW          0.0001224  0.0003454   0.355   0.7243
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9832 on 57 degrees of freedom
## Multiple R-squared:  0.5884, Adjusted R-squared:  0.574
## F-statistic: 40.75 on 2 and 57 DF,  p-value: 1.026e-11
```

```
plot(lmFW)
```


Residuals vs Fitted

Fitted values
lm(TCB ~ TSB + BW)

Normal Q-Q

lm(TCB ~ TSB + BW)

Scale-Location

√|Standardized residuals|

Fitted values
lm(TCB ~ TSB + BW)

Residuals vs Leverage

lm(TCB ~ TSB + BW)

```
#Normality test
shapiro.test(lmFW$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  lmFW$residuals
## W = 0.96655, p-value = 0.09879
```

```
#Independent test
TSA::runs(lmFW$residuals)
```

```
## $pvalue
## [1] 0.905
##
## $observed.runs
## [1] 31
##
## $expected.runs
## [1] 30.96667
##
## $n1
## [1] 31
##
## $n2
## [1] 29
##
## $k
## [1] 0
```

```
#Constant variance test
car::ncvTest(lmFW)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 3.721108, Df = 1, p = 0.053729
```

```
#Make birthweight as a coefficient of TSB
lmFWN=lm(TCB~TSB+(TSB*BW),data=Female)
summary(lmFWN)
```

```
##
## Call:
## lm(formula = TCB ~ TSB + (TSB * BW), data = Female)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.71837 -0.46015 -0.00857  0.52843  2.43242
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 18.7552421  5.8363678   3.214  0.00218 **
## TSB         -1.5196465  0.8310700  -1.829  0.07280 .
## BW          -0.0044862  0.0016633  -2.697  0.00922 **
## TSB:BW       0.0006576  0.0002327   2.826  0.00653 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.928 on 56 degrees of freedom
## Multiple R-squared:  0.6398, Adjusted R-squared:  0.6205
## F-statistic: 33.16 on 3 and 56 DF,  p-value: 1.871e-12
```