(HRS DATA 'LSB' : Linear Regression Model for each of BMI2010 to BMI2018)

2023-06-20

INSTALL PACKAGES

```
#install.packages("car")
#install.packages("MASS")
#install.packages("lme4")
#install.packages("mlmRev")
#install.packages("agridat")
#install.packages("MCMCglmm")
#install.packages("ggplot2")
#install.packages("scapeMCMC")
library("car")
```

```
## Loading required package: carData
```

```
library("MASS")
library("lme4")
```

```
## Loading required package: Matrix
```

```
library("mlmRev")
library("agridat")
library("MCMCglmm")
```

```
## Loading required package: coda

## Loading required package: ape
```

```
library("ggplot2")
library(stats)
```

IMPORT THE LSB FILE

```
## IMPORT THE LBS FILE
LSB <- read.csv(file = "LSB-2010-2018.csv")
```

###Remove missing values #Replace 999 with NA

```r
LSB[LSB == 999] <- NA
```

#USING MICE

```r
library(mice)
```

```
##
## Attaching package: 'mice'
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
## The following objects are masked from 'package:base':
##
##     cbind, rbind
```

```r
lSb <- mice(LSB, method = "pmm", m = 5)
```

```
##
##  iter imp variable
##   1   1  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   1   2  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   1   3  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   1   4  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   1   5  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   2   1  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   2   2  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   2   3  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   2   4  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   2   5  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   3   1  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   3   2  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   3   3  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   3   4  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   3   5  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   4   1  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   4   2  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   4   3  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   4   4  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   4   5  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   5   1  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   5   2  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   5   3  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   5   4  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
##   5   5  NewID  Lsat2010  Lsat2012  Lsat2014  Lsat2016  Lsat2018  Srh2010  Srh2012  Srh2014  Srh2016
```

```
## Warning: Number of logged events: 4
```

```
lsb <- complete(lSb)
head(lsb)
```

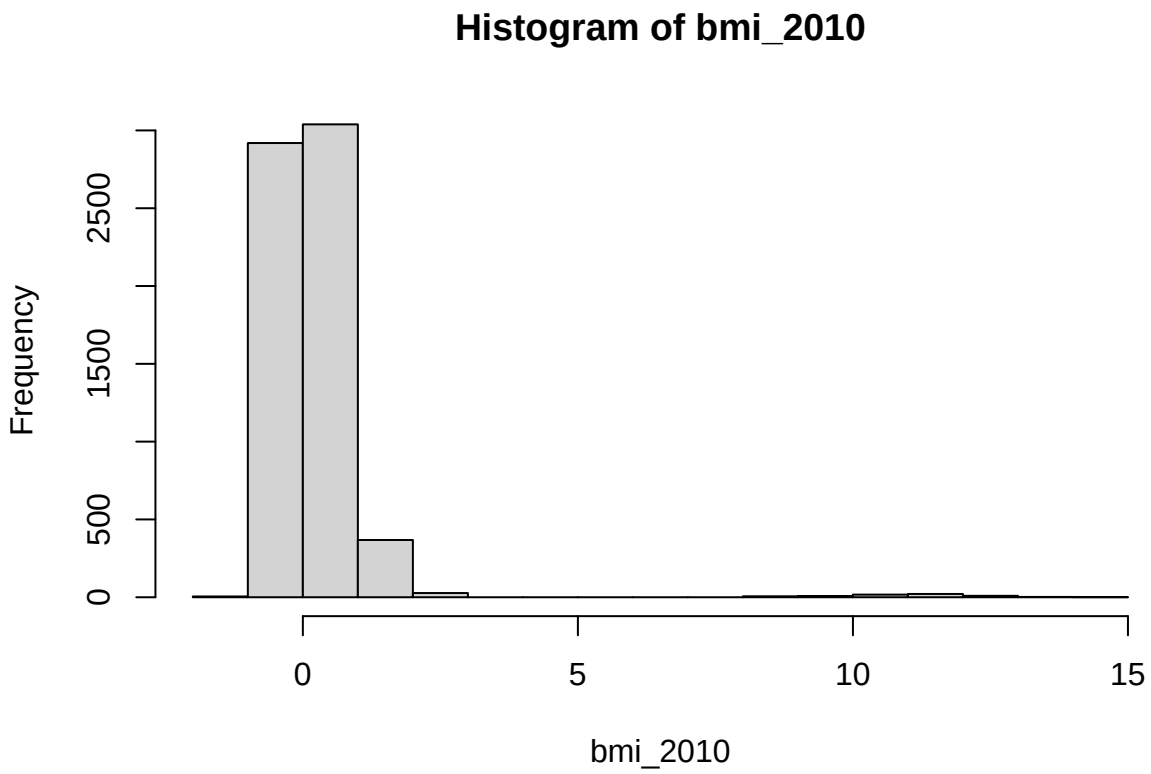
```
##   NewID Mari Gender dep smoke alc age2010 age2012 age2014 age2016 age2018
## 1     1    1      0   0     1   1      54      56      58      60      62
## 2     2    0      0   0     1   1      68      70      72      74      76
## 3     3    1      1   0     1   1      64      66      68      70      72
## 4     4    0      0   1     1   1      68      70      72      74      76
## 5     5    1      1   0     1   1      37      39      41      43      45
## 6     6    1      1   1     1   1      60      62      64      66      68
##   Lsat2010 Lsat2012 Lsat2014 Lsat2016 Lsat2018 Srh2010 Srh2012 Srh2014 Srh2016
## 1        1        0        0        1        1       1       0       0       0
## 2        0        0        1        1        2       0       0       0       0
## 3        0        0        0        0        1       0       0       0       0
## 4        1        1        2        1        1       0       0       0       0
## 5        0        0        0        0        0       1       2       0       0
## 6        0        0        0        1        0       0       0       0       0
##   Srh2018      bmi2010    bmi2012     bmi2014     bmi2016     bmi2018
## 1       0  0.098857035  0.2577475  0.08681049  0.19616334 -0.15715976
## 2       0  0.222581130 11.0221229 -0.20106838  1.00854701  0.91797906
## 3       0 -0.048775619 -0.2826723 -0.36005603 -0.04877562 -0.37152569
## 4       1  0.009230769  0.6119929  0.12442421 -0.01506160 -0.33191857
## 5       0 -0.265820274 -0.1585461  0.23769311 -0.22293313 -0.07180021
## 6       0  0.556625598  0.5280088  0.23049790  0.29764103 -0.11094017
```

Normality check for bmi_2010

```
bmi_2010 <- lsb$bmi2010
head(bmi_2010)
```

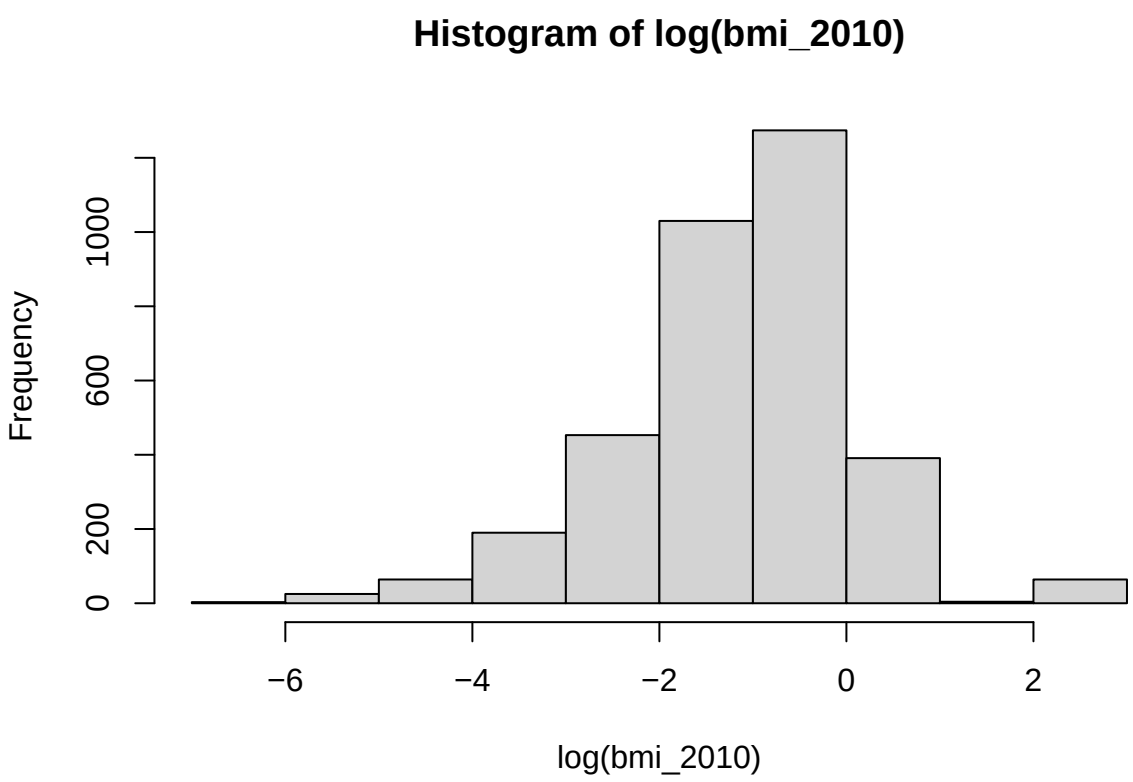
```
## [1]  0.098857035  0.222581130 -0.048775619  0.009230769 -0.265820274
## [6]  0.556625598
```

```
hist(bmi_2010)
```




```
hist(log(bmi_2010))
```

```
## Warning in log(bmi_2010): NaNs produced
```



###Looks fairly normal, however normality is not a core assumption to use this so we move futher to fitting the linear regression model and perform the model diagnostics afterwards.

Create a linear regression of bmi_2010 (response variable) AND age_2010+Lsat_2010+Srh_2010

bmi_2010 = a + (age_2010)×b1 + (Lsat_2010)xb2 + (Srh_2010)xb3 + (Mari)xb4 + (Gender)xb5 + (dep)xb6 + (smoke)xb6 + (alc)xb7

```
lm_bmi_2010 = lm(lsb$bmi2010 ~ +lsb$age2010 + lsb$Lsat2010 + lsb$Srh2010 +
                   lsb$Mari + lsb$Gender + lsb$dep+ lsb$smoke + lsb$alc, data = lsb)
#Review the results
summary(lm_bmi_2010)
```

```
##
## Call:
## lm(formula = lsb$bmi2010 ~ +lsb$age2010 + lsb$Lsat2010 + lsb$Srh2010 +
##     lsb$Mari + lsb$Gender + lsb$dep + lsb$smoke + lsb$alc, data = lsb)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8574 -0.4397 -0.1573  0.1673 14.1002
##
```

```
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.618507   0.153074   4.041 5.39e-05 ***
## lsb$age2010  -0.001883   0.002716  -0.693 0.488182
## lsb$Lsat2010  0.015203   0.020734   0.733 0.463430
## lsb$Srh2010  -0.159550   0.022047  -7.237 5.13e-13 ***
## lsb$Mari     -0.060424   0.031417  -1.923 0.054486 .
## lsb$Gender   -0.159729   0.031084  -5.139 2.85e-07 ***
## lsb$dep      -0.020931   0.036307  -0.577 0.564291
## lsb$smoke    -0.116287   0.030839  -3.771 0.000164 ***
## lsb$alc      -0.041778   0.032689  -1.278 0.201286
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.183 on 6413 degrees of freedom
## Multiple R-squared:  0.01751,    Adjusted R-squared:  0.01629
## F-statistic: 14.29 on 8 and 6413 DF,  p-value: < 2.2e-16
```

###The results of the p-value for variables like Srh2010, Gender, and smoke indicates that they are probably an excellent addition to the model. However, there's are chances that the other predictors are not meaningful for the Linear regression model.

#Create a linear regression of bmi_2012 (response variable) AND age_2012+Lsat_2012+Srh_2012 ### bmi_2012 = a + (age_2012)×b1 + (Lsat_2012)xb2 + (Srh_2012)xb3 + (Mari)xb4 + (Gender)xb5 + (dep)xb6 + (smoke)xb6 + (alc)xb7

```
lm_bmi_2012 = lm(lsb$bmi2012~+lsb$age2012 + lsb$Lsat2012 + lsb$Srh2012
                 + lsb$Mari + lsb$Gender + lsb$dep+ lsb$smoke + lsb$alc, data = lsb)
#Review the results
summary(lm_bmi_2012)
```

```
##
## Call:
## lm(formula = lsb$bmi2012 ~ +lsb$age2012 + lsb$Lsat2012 + lsb$Srh2012 +
##     lsb$Mari + lsb$Gender + lsb$dep + lsb$smoke + lsb$alc, data = lsb)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5622 -0.3890 -0.1081  0.2203 12.2413
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.3711032  0.1181031   3.142  0.00168 **
## lsb$age2012   0.0005431  0.0020273   0.268  0.78881
## lsb$Lsat2012  0.0069885  0.0157972   0.442  0.65822
## lsb$Srh2012  -0.2226127  0.0166431 -13.376  < 2e-16 ***
## lsb$Mari      0.0068794  0.0234676   0.293  0.76942
## lsb$Gender   -0.1558004  0.0231757  -6.723 1.94e-11 ***
## lsb$dep       0.0272423  0.0269994   1.009  0.31301
## lsb$smoke    -0.0462675  0.0230044  -2.011  0.04434 *
## lsb$alc      -0.0250416  0.0243397  -1.029  0.30359
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.882 on 6413 degrees of freedom
## Multiple R-squared:  0.03959,    Adjusted R-squared:  0.03839
## F-statistic: 33.04 on 8 and 6413 DF,  p-value: < 2.2e-16
```

###The results of the p-value for variables like srh2012, gender and 'dep' indicates that they are probably an excellent addition to the model. However, there's are chances that the other predictors are not meaningful for the Linear regression model.

#Create a linear regression of bmi_2014 (response variable) AND age_2014+Lsat_2014+Srh_2014 ### bmi_2014 = a + (age_2014)×b1 + (Lsat_2014)xb2 + (Srh_2014)xb3 + (Mari)xb4 + (Gender)xb5 + (dep)xb6 + (smoke)xb6 + (alc)xb7

```
lm_bmi_2014 = lm(lsb$bmi2014~+lsb$age2014 + lsb$Lsat2014 + lsb$Srh2014 +
                   lsb$Mari + lsb$Gender + lsb$dep+ lsb$smoke + lsb$alc, data = lsb)
#Review the results
summary(lm_bmi_2014)
```

```
##
## Call:
## lm(formula = lsb$bmi2014 ~ +lsb$age2014 + lsb$Lsat2014 + lsb$Srh2014 +
##     lsb$Mari + lsb$Gender + lsb$dep + lsb$smoke + lsb$alc, data = lsb)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6196 -0.4306 -0.1441  0.1988 13.0802
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.632e-01  1.504e-01   4.411 1.05e-05 ***
## lsb$age2014  -6.310e-03  2.506e-03  -2.518 0.011831 *
## lsb$Lsat2014  2.892e-02  1.957e-02   1.478 0.139419
## lsb$Srh2014  -1.504e-01  2.163e-02  -6.953 3.94e-12 ***
## lsb$Mari      6.988e-02  2.893e-02   2.416 0.015731 *
## lsb$Gender   -1.115e-01  2.865e-02  -3.892 0.000101 ***
## lsb$dep       5.553e-02  3.332e-02   1.666 0.095669 .
## lsb$smoke     6.636e-05  2.846e-02   0.002 0.998140
## lsb$alc      -1.484e-02  3.011e-02  -0.493 0.622187
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.09 on 6413 degrees of freedom
## Multiple R-squared:  0.01331,    Adjusted R-squared:  0.01208
## F-statistic: 10.82 on 8 and 6413 DF,  p-value: 3.008e-15
```

###The results of the p-value for variables like age2014, Lsat2014, Srh2014, mari, and gender indicates that they are probably an excellent addition to the model. However, there's are chances that the other predictors are not meaningful for the Linear regression model.

#Create a linear regression of bmi_2016 (response variable) AND age_2016+Lsat_2016+Srh_2016 ### bmi_2016 = a + (age_2016)×b1 + (Lsat_2016)xb2 + (Srh_2016)xb3 + (Mari)xb4 + (Gender)xb5 + (dep)xb6 + (smoke)xb6 + (alc)xb7

```r
lm_bmi_2016 = lm(lsb$bmi2016~+lsb$age2016 + lsb$Lsat2016 + lsb$Srh2016
                 + lsb$Mari + lsb$Gender + lsb$dep+ lsb$smoke + lsb$alc, data = lsb)
#Review the results
summary(lm_bmi_2016)
```

```
##
## Call:
## lm(formula = lsb$bmi2016 ~ +lsb$age2016 + lsb$Lsat2016 + lsb$Srh2016 +
##     lsb$Mari + lsb$Gender + lsb$dep + lsb$smoke + lsb$alc, data = lsb)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5944 -0.4080 -0.1252  0.1953 13.1694
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.499529   0.151281   3.302 0.000965 ***
## lsb$age2016   0.001551   0.002447   0.634 0.526130
## lsb$Lsat2016  0.017214   0.018764   0.917 0.358960
## lsb$Srh2016  -0.173172   0.021343  -8.114 5.84e-16 ***
## lsb$Mari      0.025478   0.028264   0.901 0.367404
## lsb$Gender   -0.301832   0.027940 -10.803  < 2e-16 ***
## lsb$dep      -0.050996   0.032395  -1.574 0.115488
## lsb$smoke    -0.137226   0.027678  -4.958 7.31e-07 ***
## lsb$alc      -0.111534   0.029342  -3.801 0.000145 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.063 on 6413 degrees of freedom
## Multiple R-squared:  0.039,  Adjusted R-squared:  0.0378
## F-statistic: 32.53 on 8 and 6413 DF,  p-value: < 2.2e-16
```

###The results of the p-value for variables like, age2016, Lsat2016, srh2016, gender and alcohol indicates that it is probably an excellent addition to the model. However, there's are chances that the other predictors are not meaningful for the Linear regression model. #Create a linear regression of bmi_2018 (response variable) AND age_2018+Lsat_2018+Srh_2018 ### bmi_2018 = a + (age_2018)×b1 + (Lsat_2018)xb2 + (Srh_2018)xb3 + (Mari)xb4 + (Gender)xb5 + (dep)xb6 + (smoke)xb6 + (alc)xb7

```r
lm_bmi_2018 = lm(lsb$bmi2018~+lsb$age2018 + lsb$Lsat2018 + lsb$Srh2018
                 + lsb$Mari + lsb$Gender + lsb$dep+ lsb$smoke + lsb$alc, data = lsb)
#Review the results
summary(lm_bmi_2018)
```

```
##
## Call:
## lm(formula = lsb$bmi2018 ~ +lsb$age2018 + lsb$Lsat2018 + lsb$Srh2018 +
##     lsb$Mari + lsb$Gender + lsb$dep + lsb$smoke + lsb$alc, data = lsb)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5298 -0.4424 -0.1507  0.1955 14.9362
##
```

```
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.070877   0.174643   6.132 9.20e-10 ***
## lsb$age2018  -0.007976   0.002734  -2.918  0.00354 **
## lsb$Lsat2018 -0.064922   0.021100  -3.077  0.00210 **
## lsb$Srh2018  -0.241540   0.024580  -9.827  < 2e-16 ***
## lsb$Mari     -0.029750   0.031480  -0.945  0.34467
## lsb$Gender   -0.101960   0.031289  -3.259  0.00113 **
## lsb$dep       0.081068   0.036350   2.230  0.02577 *
## lsb$smoke    -0.069346   0.031052  -2.233  0.02557 *
## lsb$alc      -0.142156   0.032876  -4.324 1.56e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.19 on 6413 degrees of freedom
## Multiple R-squared:  0.03114,    Adjusted R-squared:  0.02993
## F-statistic: 25.77 on 8 and 6413 DF,  p-value: < 2.2e-16
```

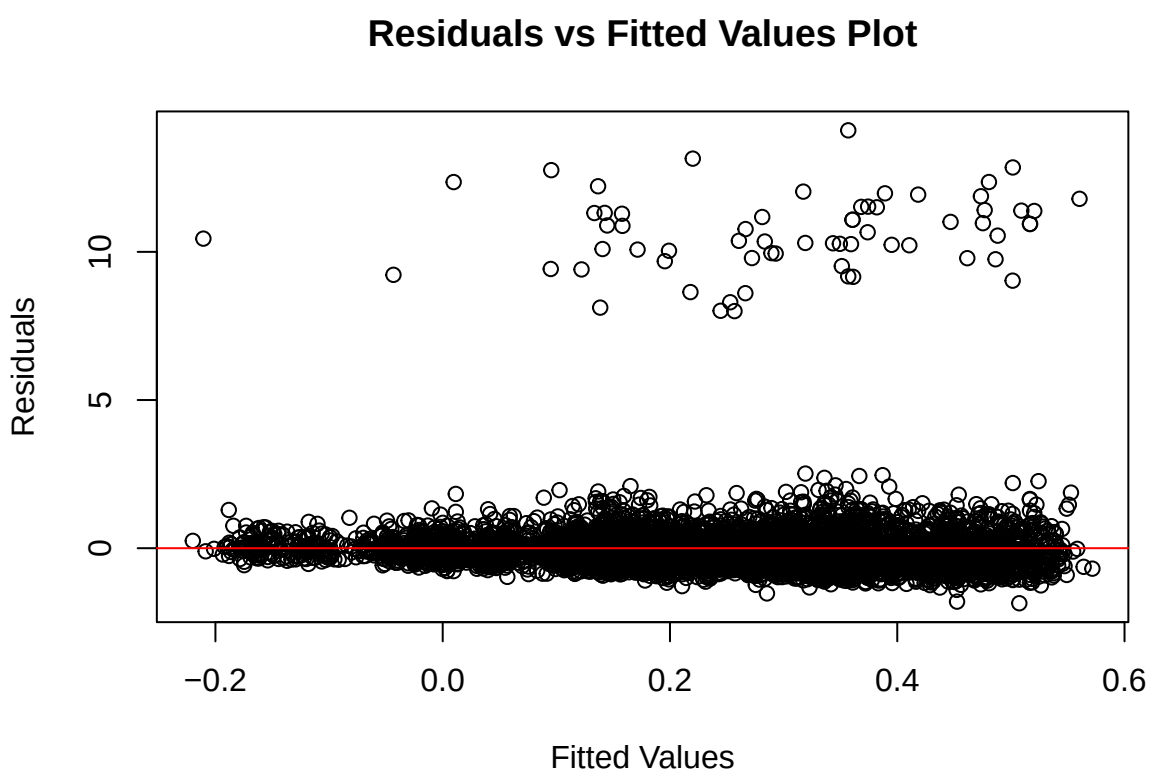
###The results of the p-value for variables like, Lsat2018, age2018, srh2018, gender and alcohol indicates that they are probably an excellent addition to the model. However, there's are chances that the other predictors are not meaningful for the Linear regression model.

#MODEL DIAGNOSTICS

####model diagnostics lm_bmi_2010

Residuals vs Fitted Values Plot with adjusted margins

```r
par(mar = c(5, 5, 4, 2) + 0.1)  # Adjusting the plot margins
plot(lm_bmi_2010$fitted.values, lm_bmi_2010$residuals,
     xlab = "Fitted Values", ylab = "Residuals",
     main = "Residuals vs Fitted Values Plot")
abline(h = 0, col = "red")
```



Check for homoscedasticity

```
ncvTest(lm_bmi_2010)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 617.8867, Df = 1, p = < 2.22e-16
```

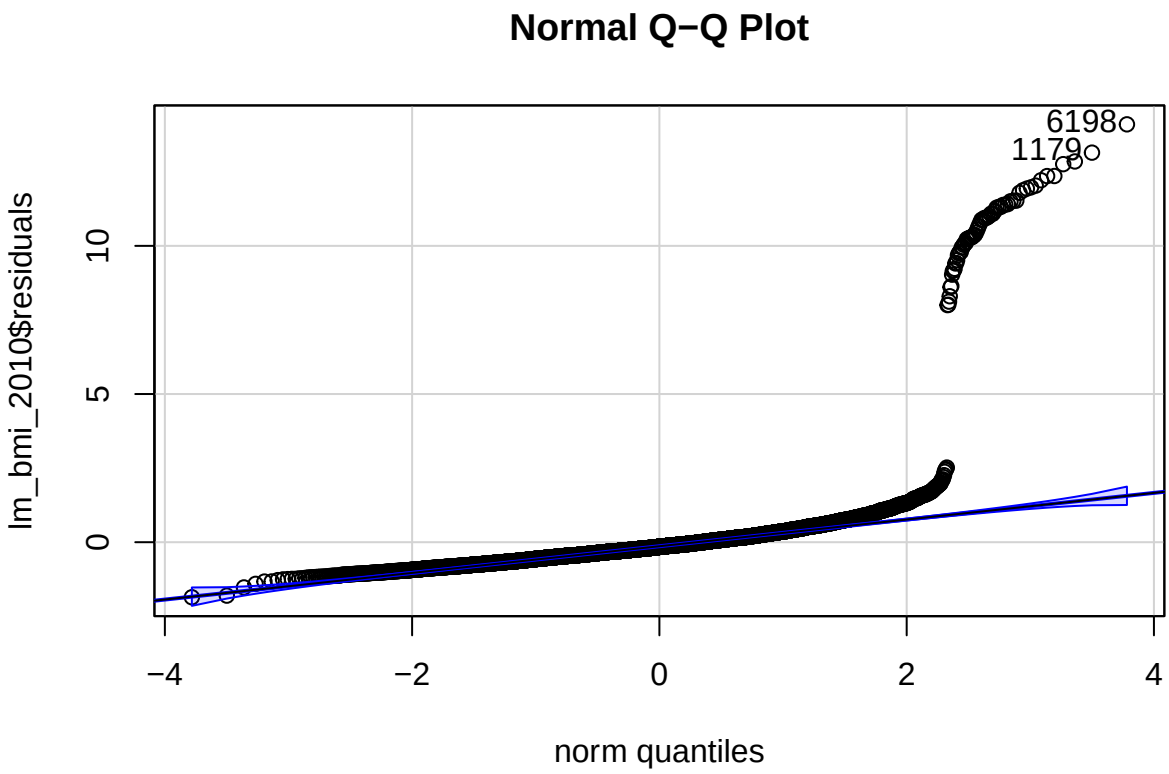
Check for normality of residuals

Q-Q Plot for residuals

```
qqPlot(lm_bmi_2010$residuals, main = "Normal Q-Q Plot")
```

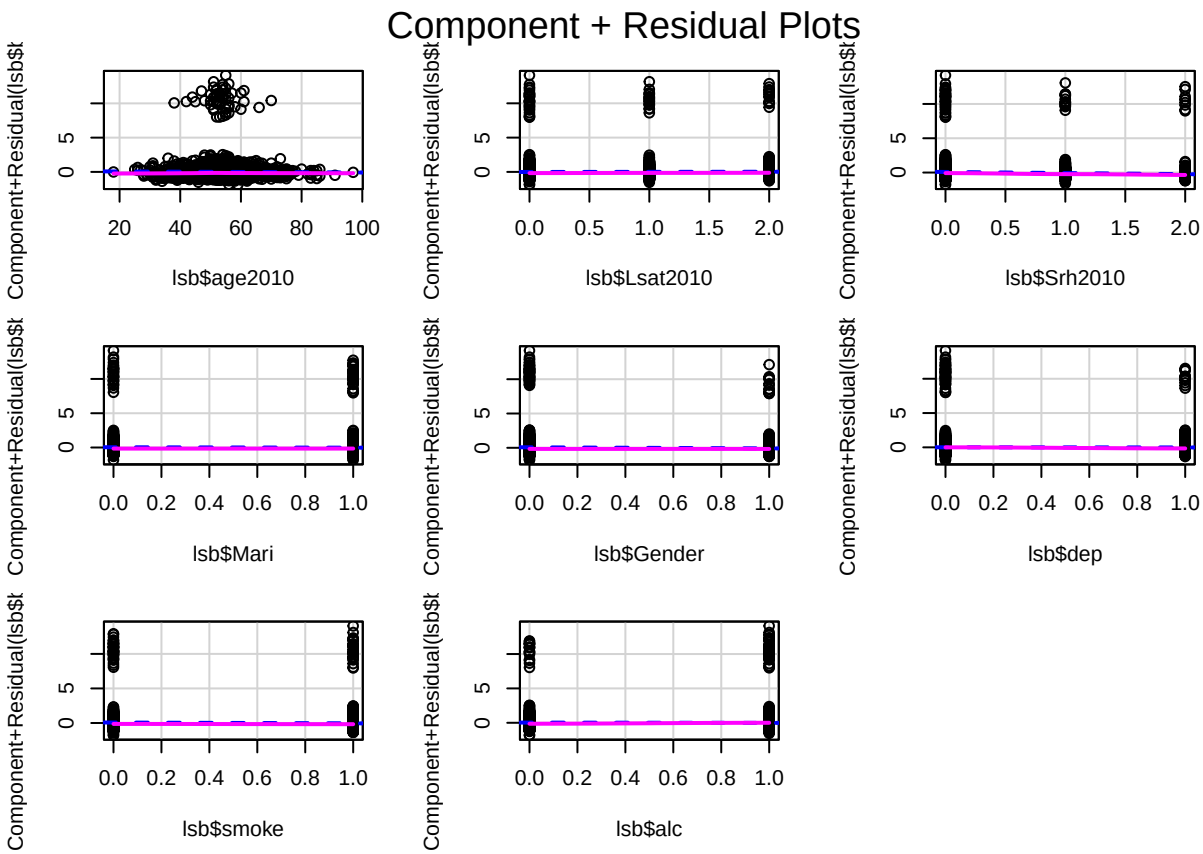
```
## [1] 6198 1179
```

```
qqline(lm_bmi_2010$residuals)
```



Check for linearity

```
crPlots(lm_bmi_2010)
```



check for independence

Durbin-Watson Test

```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

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

```
dwtest(lm_bmi_2010)
```

```
##
##  Durbin-Watson test
##
```

```
## data:  lm_bmi_2010
## DW = 2.0352, p-value = 0.9199
## alternative hypothesis: true autocorrelation is greater than 0
```

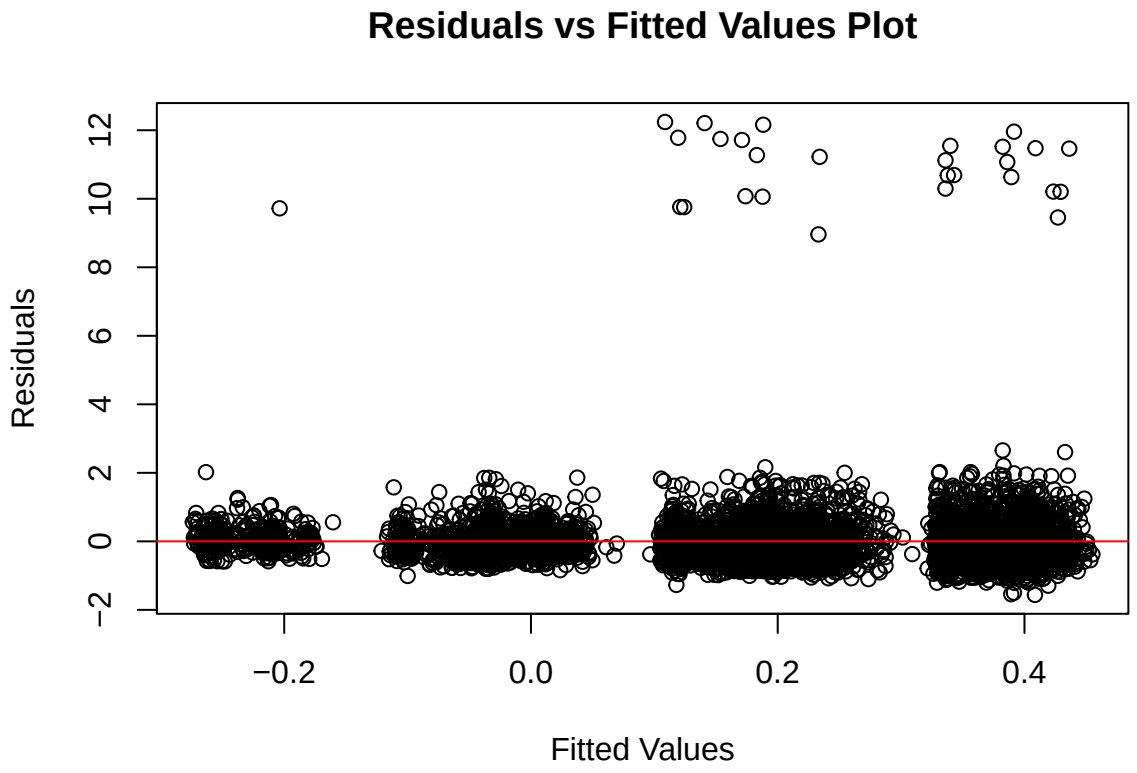
###INTERPRETATION: The plot of the Residuals vs Fitted Values, displaiys fair homoscedasricity, however the p-value for the ncvTest further confirms lack of homoscedasticity. The qq-plot of the residuals, violates the normality assumptions. The Component+Residual plots affirms Linearity and the Durbin-Watson (DW) statistic close to 2 shows evidence of Independence . However, since some of the assuptions seem to be violated, this linear model might not be a good fit for predicting Bmi2010

####model diagnostics lm_bmi_2012

Residuals vs Fitted Values Plot with adjusted margins

```r
par(mar = c(5, 5, 4, 2) + 0.1)  # Adjusting the plot margins
plot(lm_bmi_2012$fitted.values, lm_bmi_2012$residuals,
     xlab = "Fitted Values", ylab = "Residuals",
     main = "Residuals vs Fitted Values Plot")

abline(h = 0, col = "red")
```



Check for heteroscedasticity

```
ncvTest(lm_bmi_2012)
```

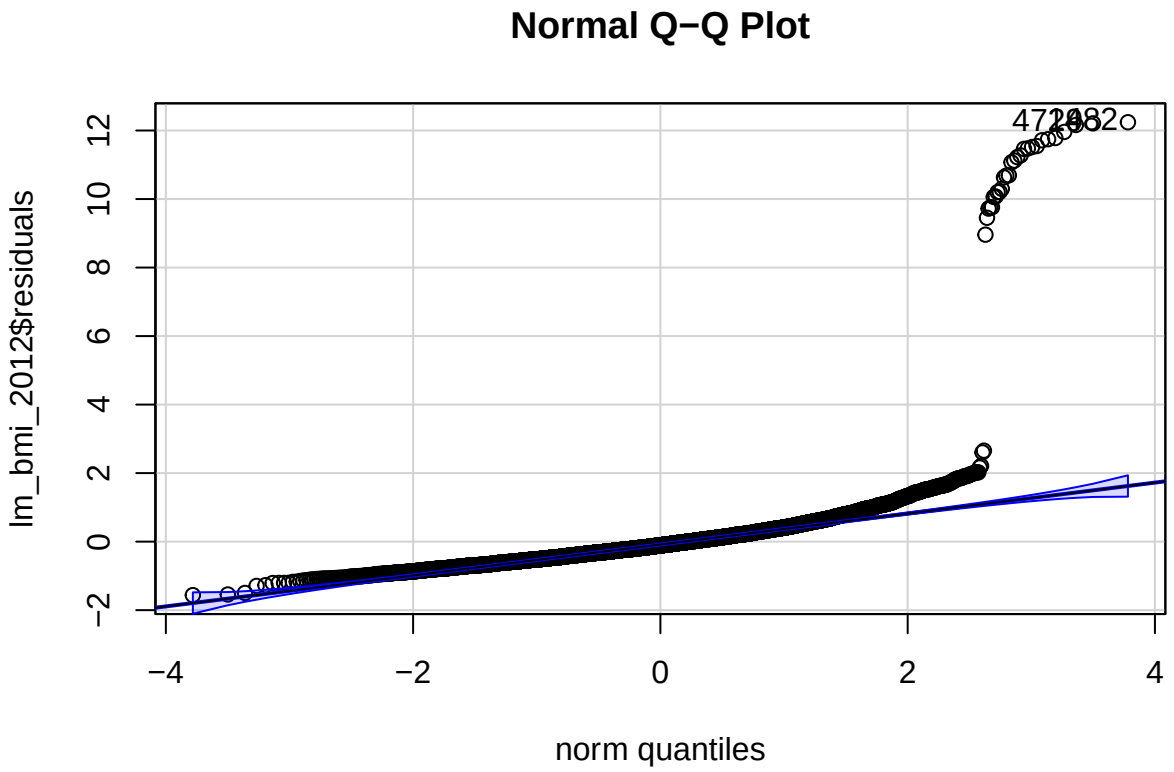
```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 474.1475, Df = 1, p = < 2.22e-16
```

Check for normality of residuals

```
# Q-Q Plot for residuals
qqPlot(lm_bmi_2012$residuals, main = "Normal Q-Q Plot")
```

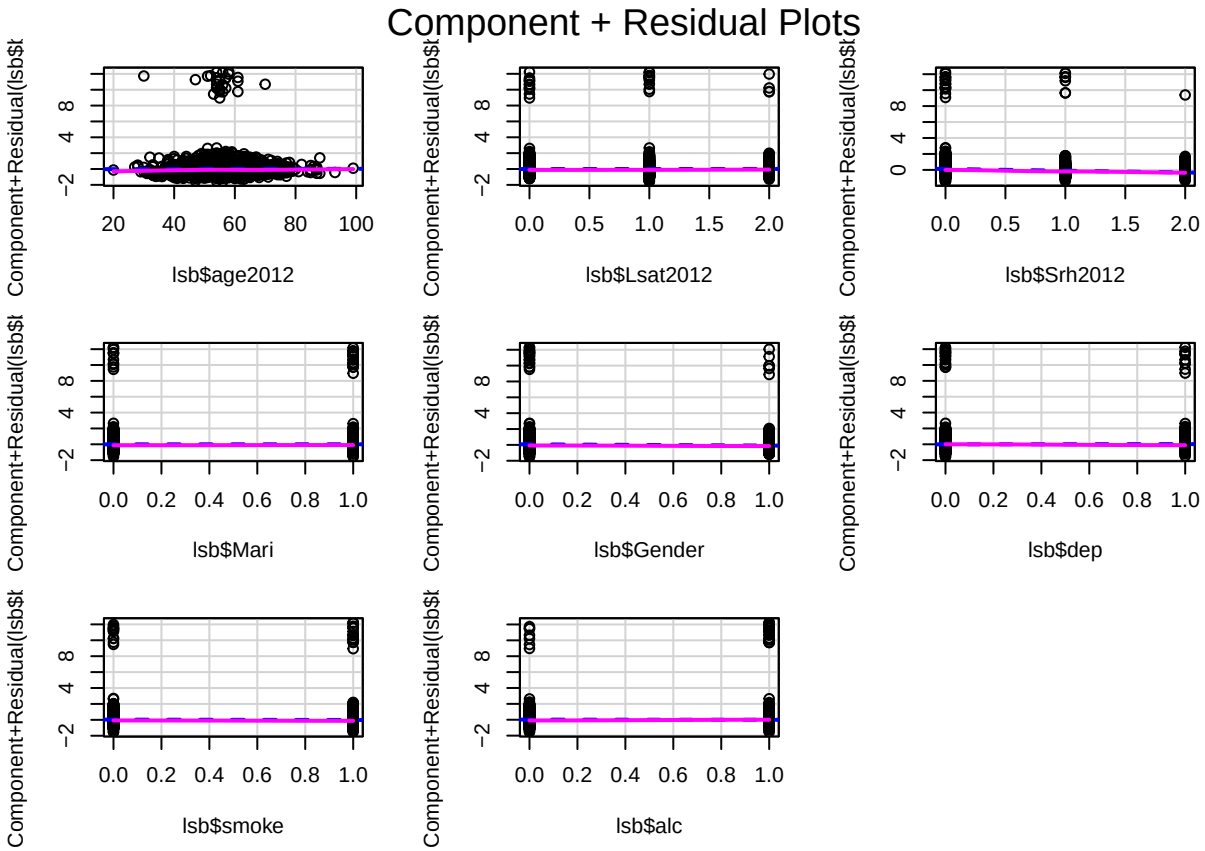
```
## [1] 1482 4729
```

```
qqline(lm_bmi_2012$residuals)
```



Check for linearity

```
crPlots(lm_bmi_2012)
```



check for independence

Durbin-Watson Test

```
dwtest(lm_bmi_2012)
```

```
##
##  Durbin-Watson test
##
## data:  lm_bmi_2012
## DW = 1.9738, p-value = 0.1454
## alternative hypothesis: true autocorrelation is greater than 0
```

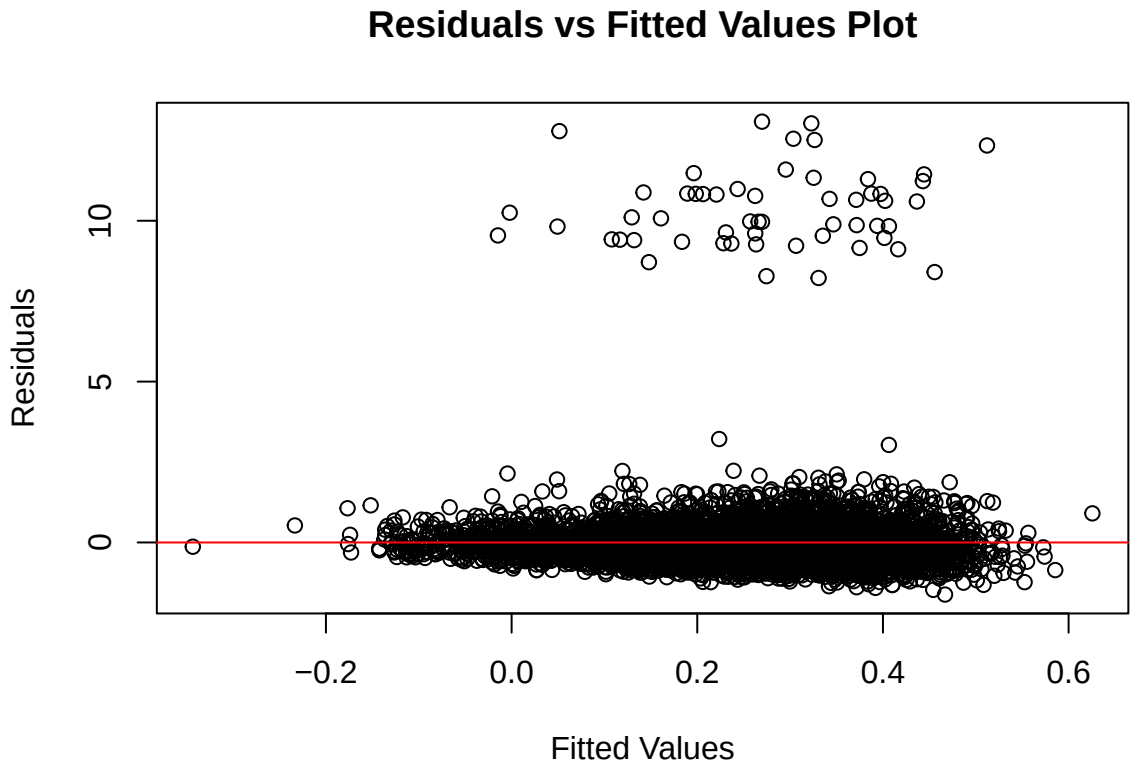
###INTERPRETATION: The plot of the Residuals vs Fitted Values, displaiys fair homoscedasricity, however the p-value for the ncvTest further confirms lack of homoscedasticity. The qq-plot of the residuals,

violates the normality assumptions. The Component+Residual plots affirms Linearity and the Durbin-Watson (DW) statistic close to 2 shows evidence of Independence . However, since some of the assuptions seem to be violated, this linear model might not be a good fit for predicting Bmi2012

####model diagnostics lm_bmi_2014

Residuals vs Fitted Values Plot with adjusted margins

```
par(mar = c(5, 5, 4, 2) + 0.1)  # Adjusting the plot margins
plot(lm_bmi_2014$fitted.values, lm_bmi_2014$residuals,
     xlab = "Fitted Values", ylab = "Residuals",
     main = "Residuals vs Fitted Values Plot")
abline(h = 0, col = "red")
```



Check for heteroscedasticity

```
ncvTest(lm_bmi_2014)
```

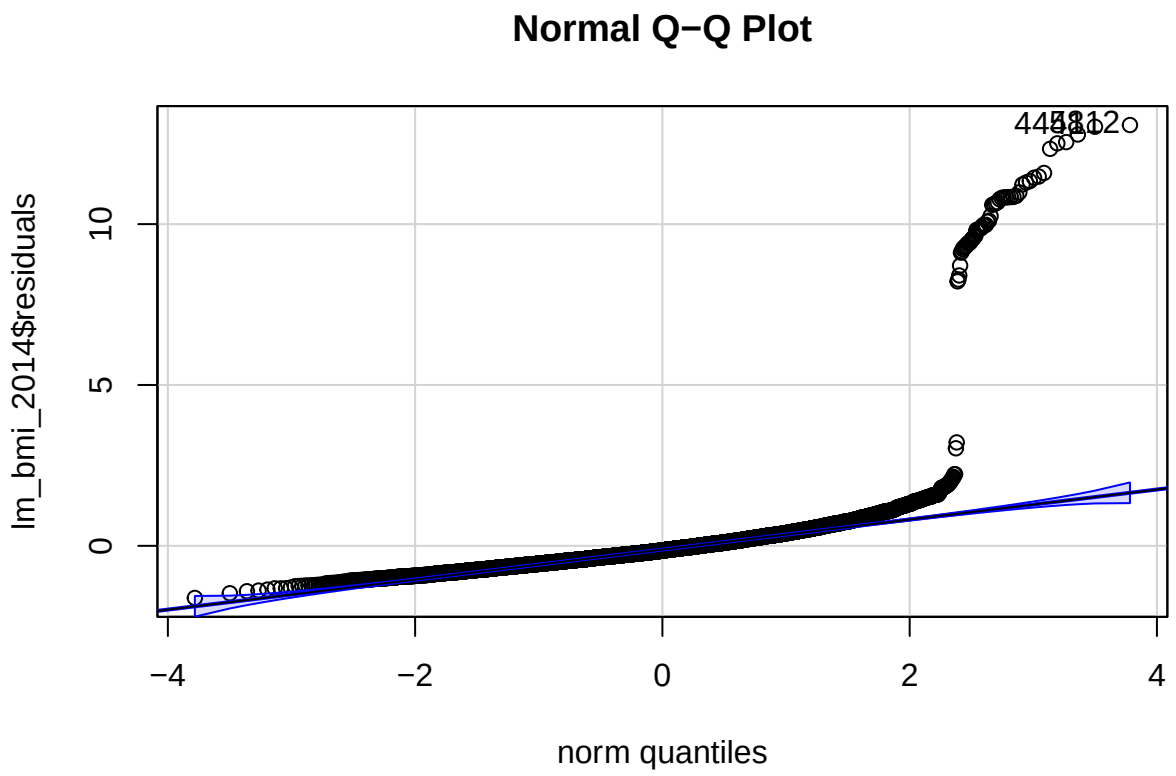
```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 192.6234, Df = 1, p = < 2.22e-16
```

Check for normality of residuals

```
# Q-Q Plot for residuals
qqPlot(lm_bmi_2014$residuals, main = "Normal Q-Q Plot")
```

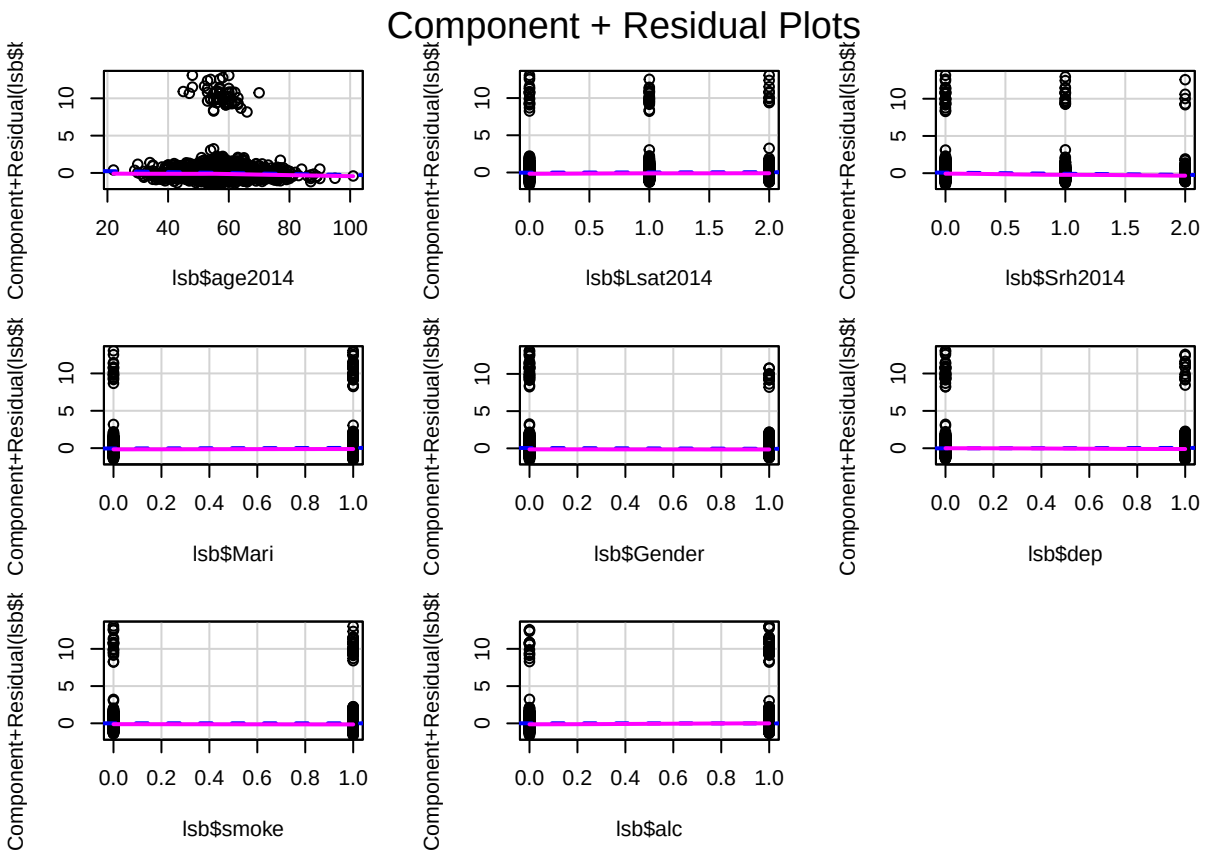
```
## [1] 5812 4441
```

```
qqline(lm_bmi_2014$residuals)
```



Check for linearity

```
crPlots(lm_bmi_2014)
```



check for independence

Durbin-Watson Test

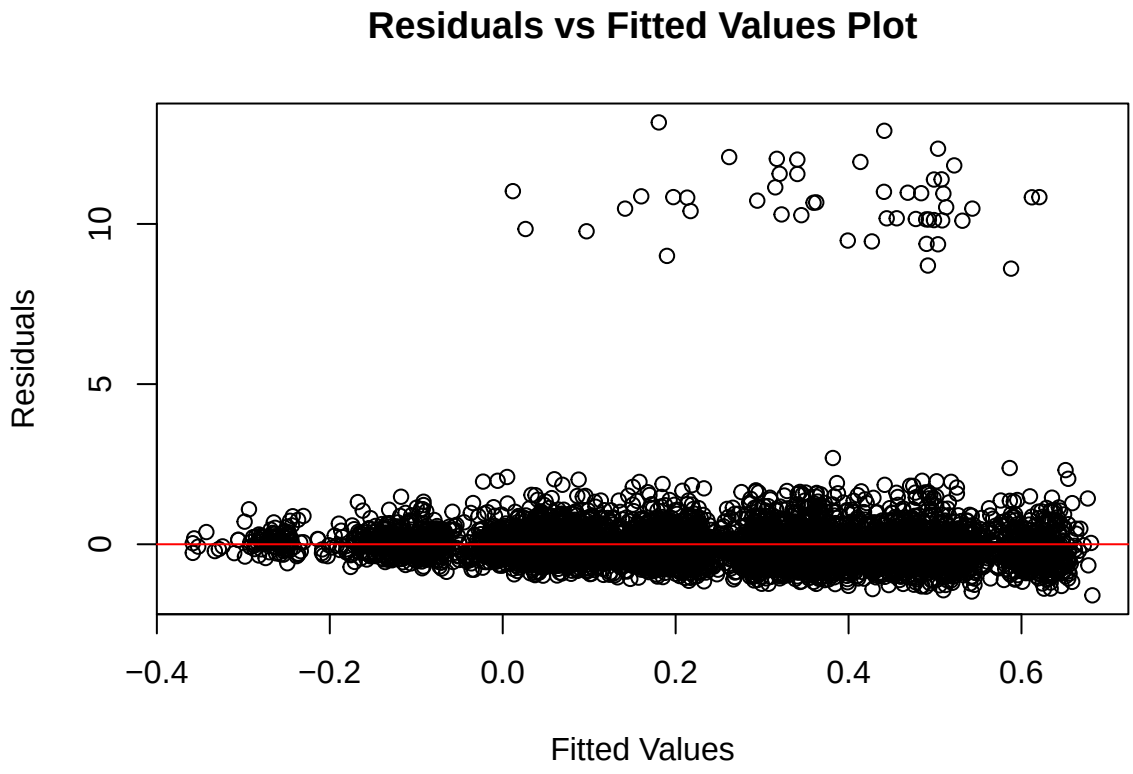
```
dwtest(lm_bmi_2014)
```

```
##
##  Durbin-Watson test
##
## data:  lm_bmi_2014
## DW = 1.99, p-value = 0.342
## alternative hypothesis: true autocorrelation is greater than 0
```

###INTERPRETATION: The plot of the Residuals vs Fitted Values, displaiys fair homoscedasricity, however the p-value for the ncvTest further confirms lack of homoscedasticity. The qq-plot of the residuals, violates the normality assumptions. The Component+Residual plots affirms Linearity and the Durbin-Watson (DW) statistic close to 2 shows evidence of Independence . However, since some of the assuptions seem to be violated, this linear model might not be a good fit for predicting Bmi2014

####model diagnostics lm_bmi_2016 # Residuals vs Fitted Values Plot with adjusted margins

```r
par(mar = c(5, 5, 4, 2) + 0.1)  # Adjusting the plot margins
plot(lm_bmi_2016$fitted.values, lm_bmi_2016$residuals,
     xlab = "Fitted Values", ylab = "Residuals",
     main = "Residuals vs Fitted Values Plot")
abline(h = 0, col = "red")
```



Check for heteroscedasticity

```r
ncvTest(lm_bmi_2016)
```

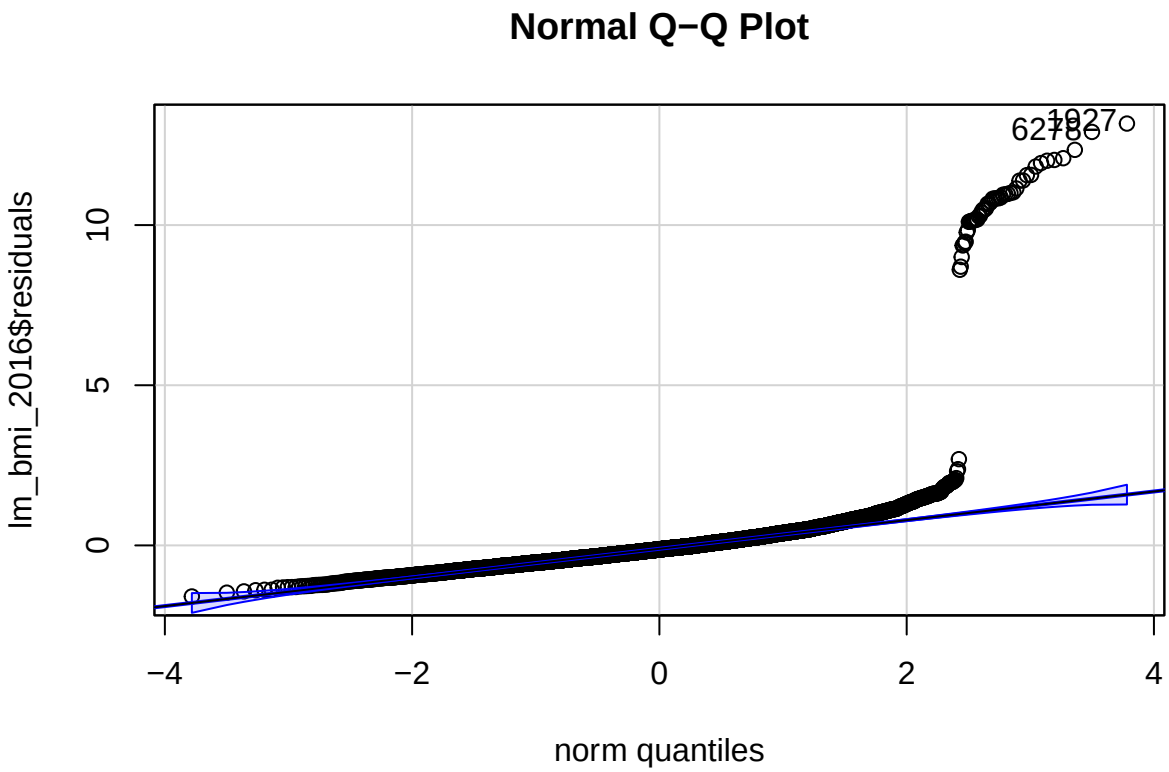
```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 1064.976, Df = 1, p = < 2.22e-16
```

Check for normality of residuals

```r
# Q-Q Plot for residuals
qqPlot(lm_bmi_2016$residuals, main = "Normal Q-Q Plot")
```

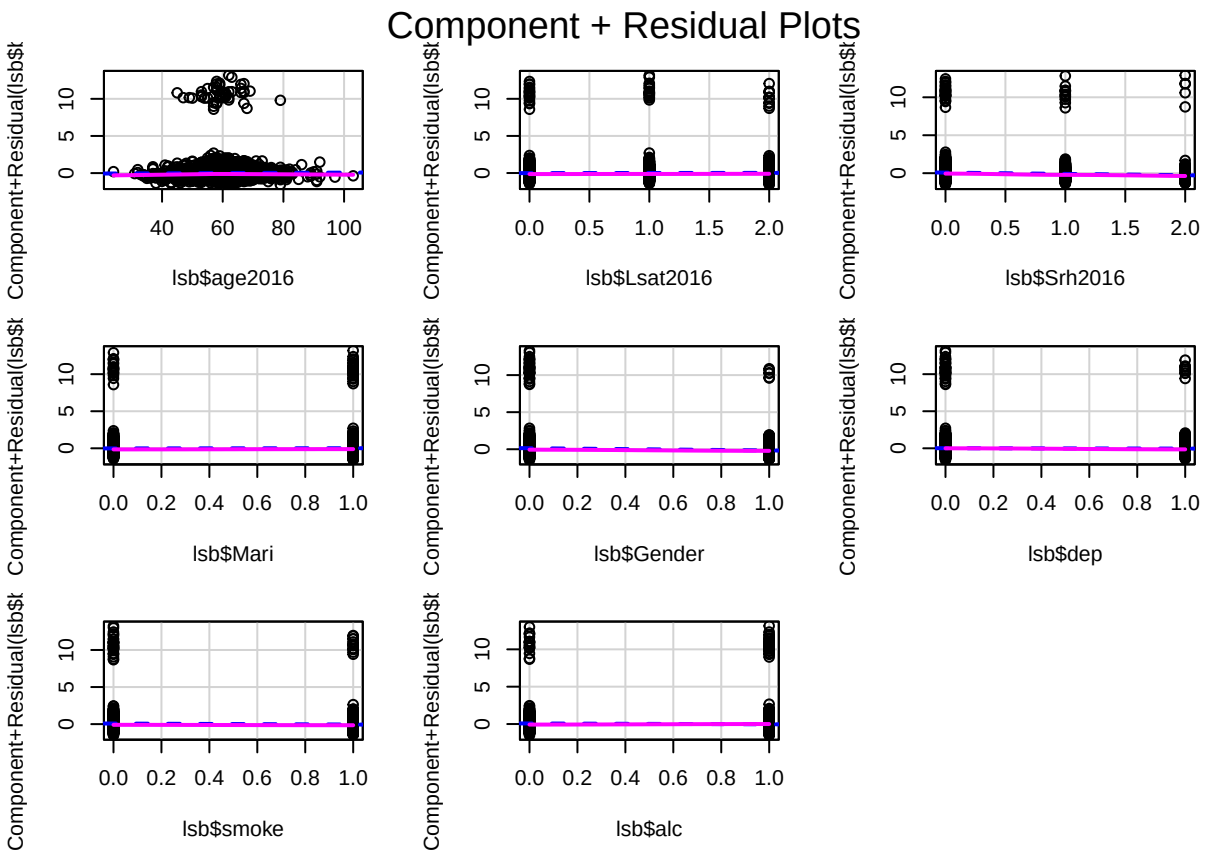
```
## [1] 1927 6278
```

```
qqline(lm_bmi_2016$residuals)
```



Check for linearity

```
crPlots(lm_bmi_2016)
```



check for independence

Durbin-Watson Test

```
dwtest(lm_bmi_2016)
```

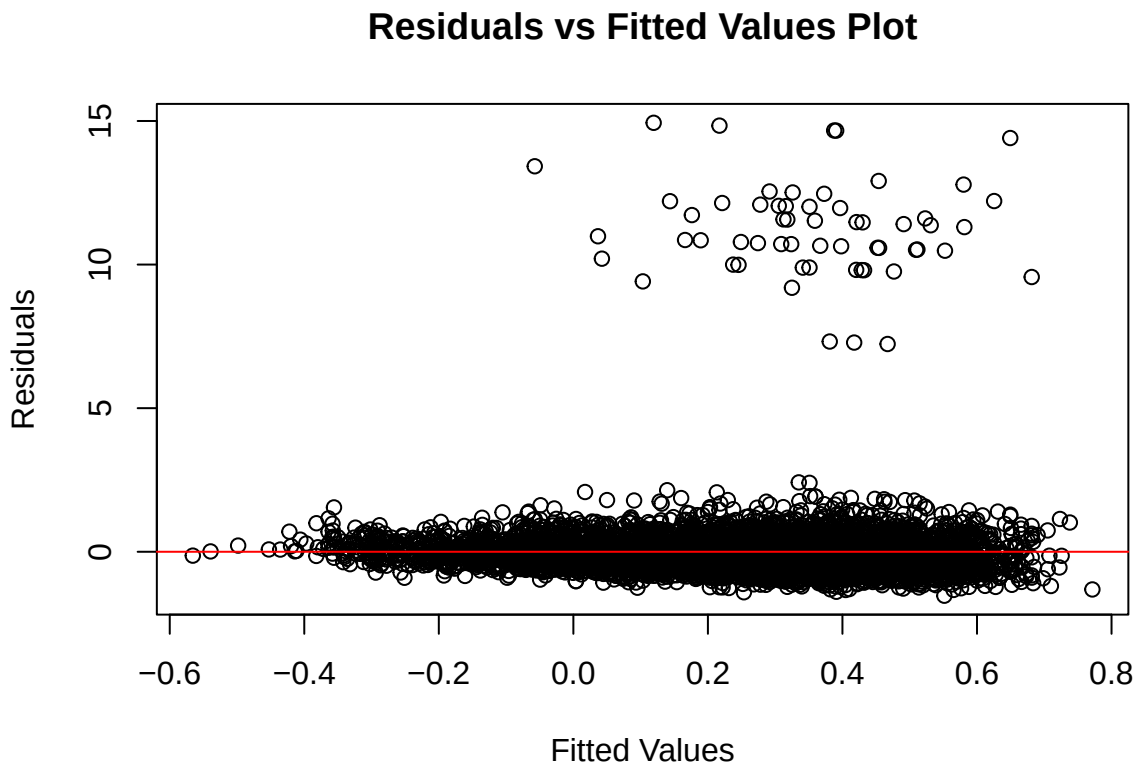
```
##
##  Durbin-Watson test
##
## data:  lm_bmi_2016
## DW = 1.9958, p-value = 0.4304
## alternative hypothesis: true autocorrelation is greater than 0
```

###INTERPRETATION: The plot of the Residuals vs Fitted Values, displaiys fair homoscedasricity, however the p-value for the ncvTest further confirms lack of homoscedasticity. The qq-plot of the residuals, violates the normality assumptions. The Component+Residual plots affirms Linearity and the Durbin-Watson (DW) statistic close to 2 shows evidence of Independence . However, since some of the assuptions seem to be violated, this linear model might not be a good fit for predicting Bmi2016

####model diagnostics lm_bmi_2018

Residuals vs Fitted Values Plot with adjusted margins

```
par(mar = c(5, 5, 4, 2) + 0.1)  # Adjusting the plot margins
plot(lm_bmi_2018$fitted.values, lm_bmi_2018$residuals,
     xlab = "Fitted Values", ylab = "Residuals",
     main = "Residuals vs Fitted Values Plot")
abline(h = 0, col = "red")
```



Check for heteroscedasticity

```
ncvTest(lm_bmi_2018)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 724.5497, Df = 1, p = < 2.22e-16
```

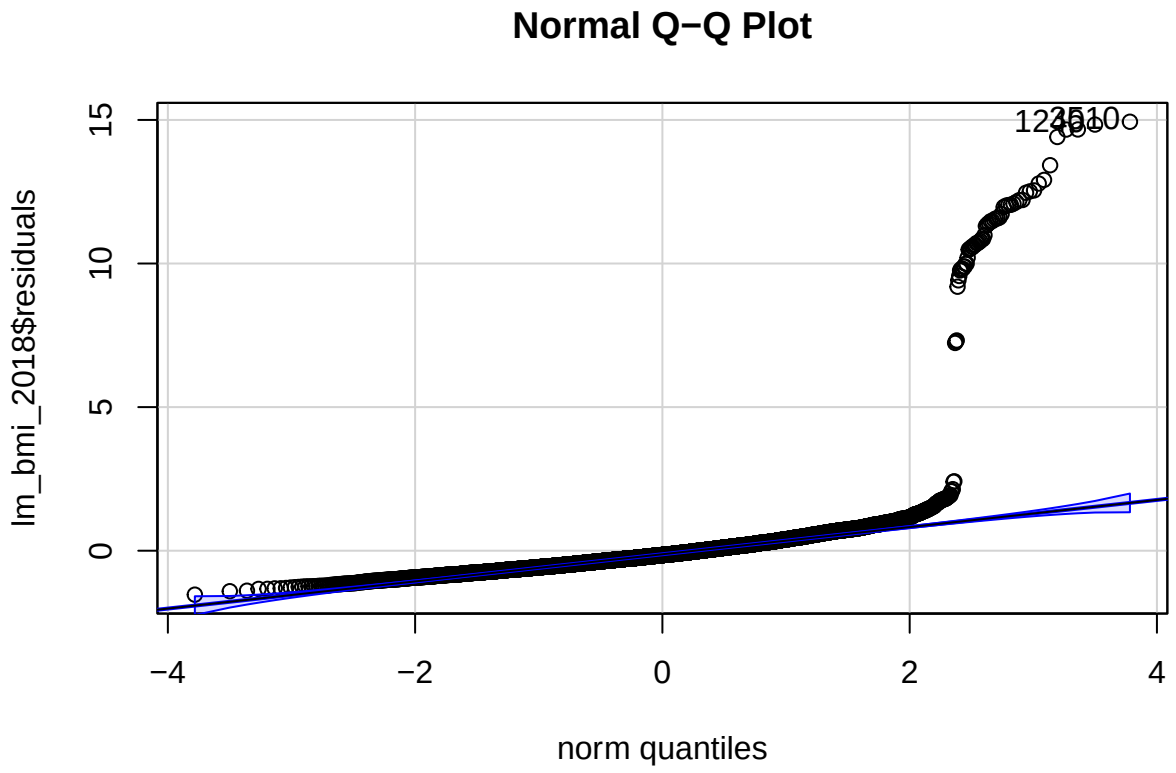
Check for normality of residuals

Q-Q Plot for residuals

```
qqPlot(lm_bmi_2018$residuals, main = "Normal Q-Q Plot")
```

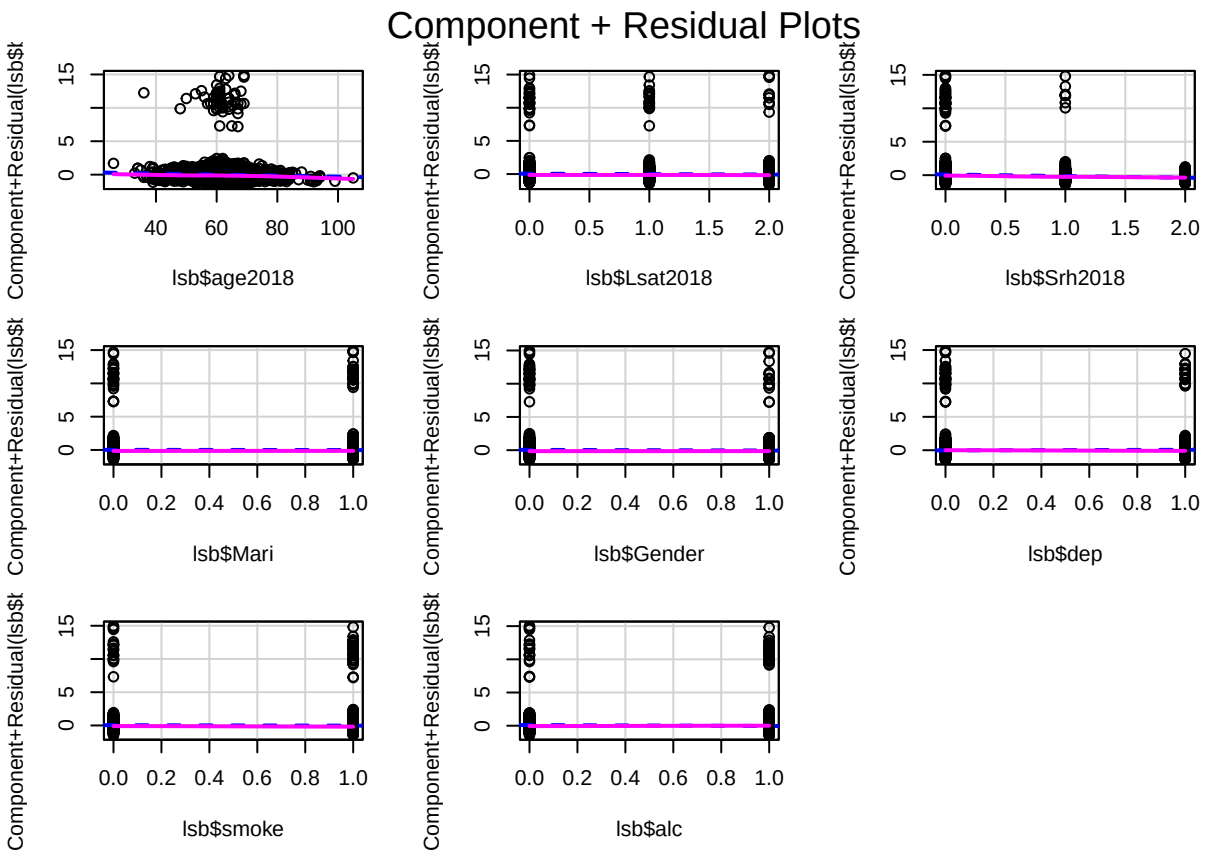
```
## [1] 3510 1240
```

```
qqline(lm_bmi_2018$residuals)
```



Check for linearity

```
crPlots(lm_bmi_2018)
```



check for independence

Durbin-Watson Test

```
dwtest(lm_bmi_2018)
```

```
##
##  Durbin-Watson test
##
## data:  lm_bmi_2018
## DW = 1.9607, p-value = 0.05681
## alternative hypothesis: true autocorrelation is greater than 0
```

###INTERPRETATION: The plot of the Residuals vs Fitted Values, displaiys fair homoscedasricity, however the p-value for the ncvTest further confirms lack of homoscedasticity. The qq-plot of the residuals, violates the normality assumptions. The Component+Residual plots affirms Linearity and the Durbin-Watson (DW) statistic close to 2 shows evidence of Independence . However, since some of the assuptions seem to be violated, this linear model might not be a good fit for predicting Bmi2018