Linear Models: (PEP) Positive Expiratory Pressure

Rodulfo D. Alfafara II

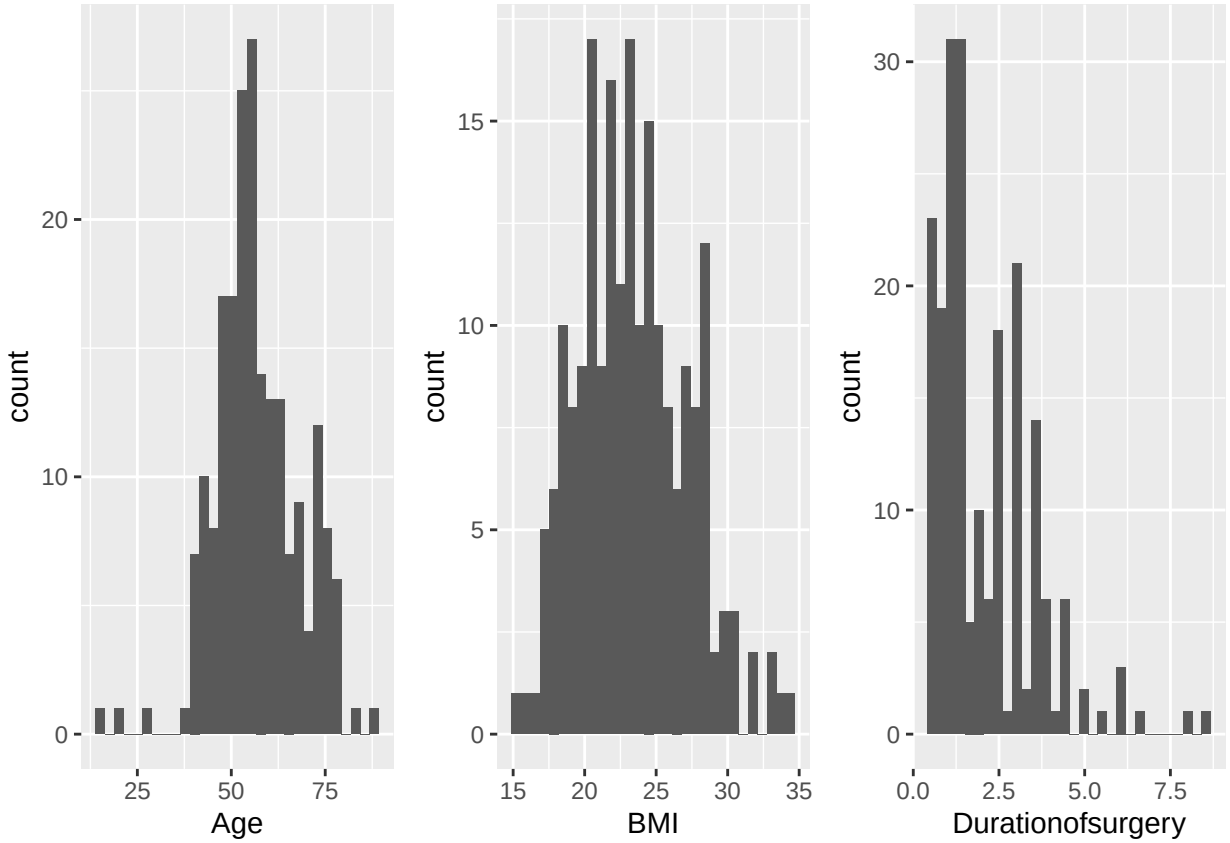
2023-09-18

In a randomized controlled trial by Zhang et al. (2015), the impact of an oscillating positive expiratory pressure (PEP) device on postoperative outcomes in patients who underwent thoracic or upper abdominal surgery was investigated. The study included 99 patients in the treatment group and 104 in the control group, focusing on the length of hospital stay. The analysis involved reading the data from PEP.csv, creating histograms to check for the need for transformations, and generating new variables new_stay (adjusted hospital stay) and log_surgery (log-transformed duration of surgery). Scatterplot matrices and loess curves were used to explore relationships between these variables and others, such as BMI and maximum temperature on the first day post-surgery. Regression analyses were conducted to evaluate the significance of predictors like log_surgery, group, and MaxtempDay1, and model assumptions were checked. The analysis also identified potential outliers based on standardized residuals, leverage, and Cook's distance. This comprehensive approach aimed to assess the PEP device's effectiveness in reducing hospital stays and to identify key predictors and outliers in the dataset.

```
#read the dataset
PEP <- read.csv("PEP.csv", header = TRUE)
```

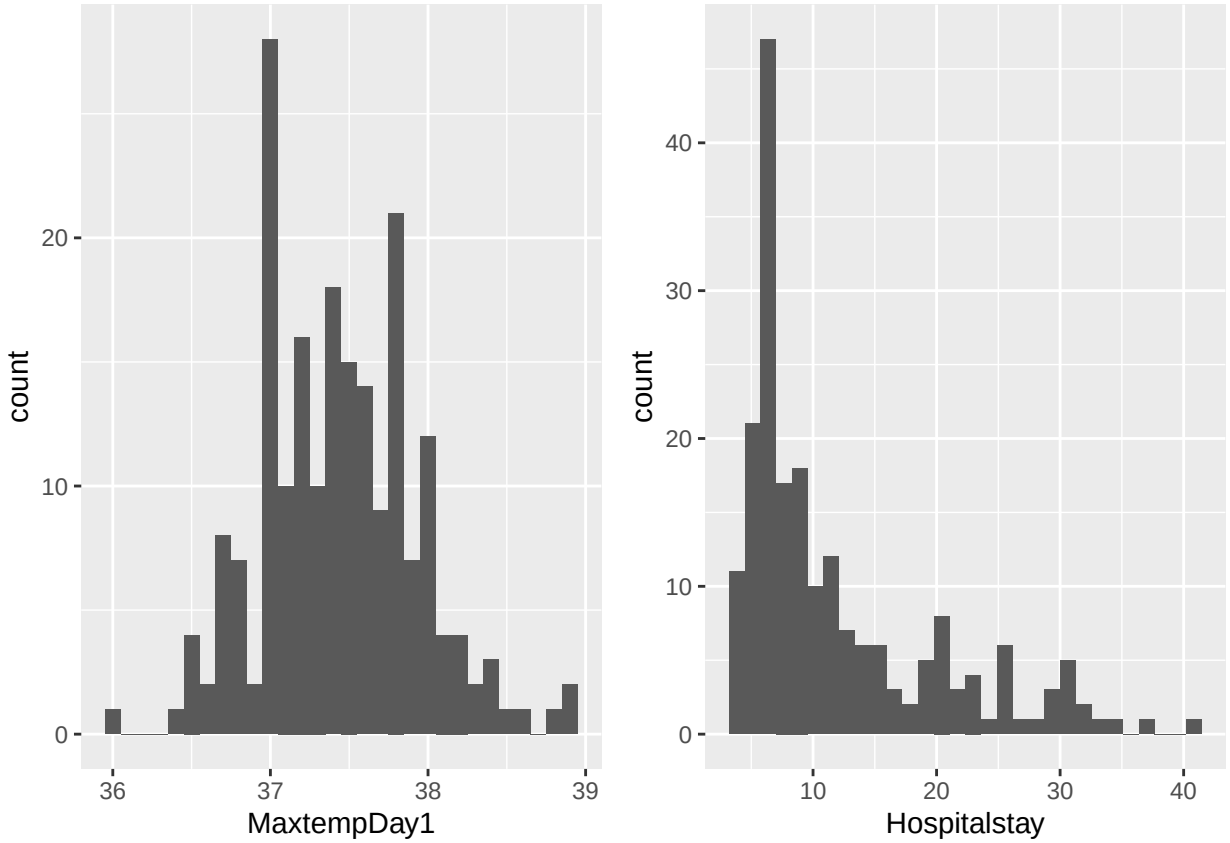
```
#Displayed histogram of each continuous variable
p1<- ggplot(PEP, aes(x=Age))+geom_histogram()
p2<- ggplot(PEP, aes(x=BMI))+geom_histogram()
p3<- ggplot(PEP, aes(x=Durationofsurgery))+geom_histogram()
grid.arrange(p1,p2,p3, ncol=3)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



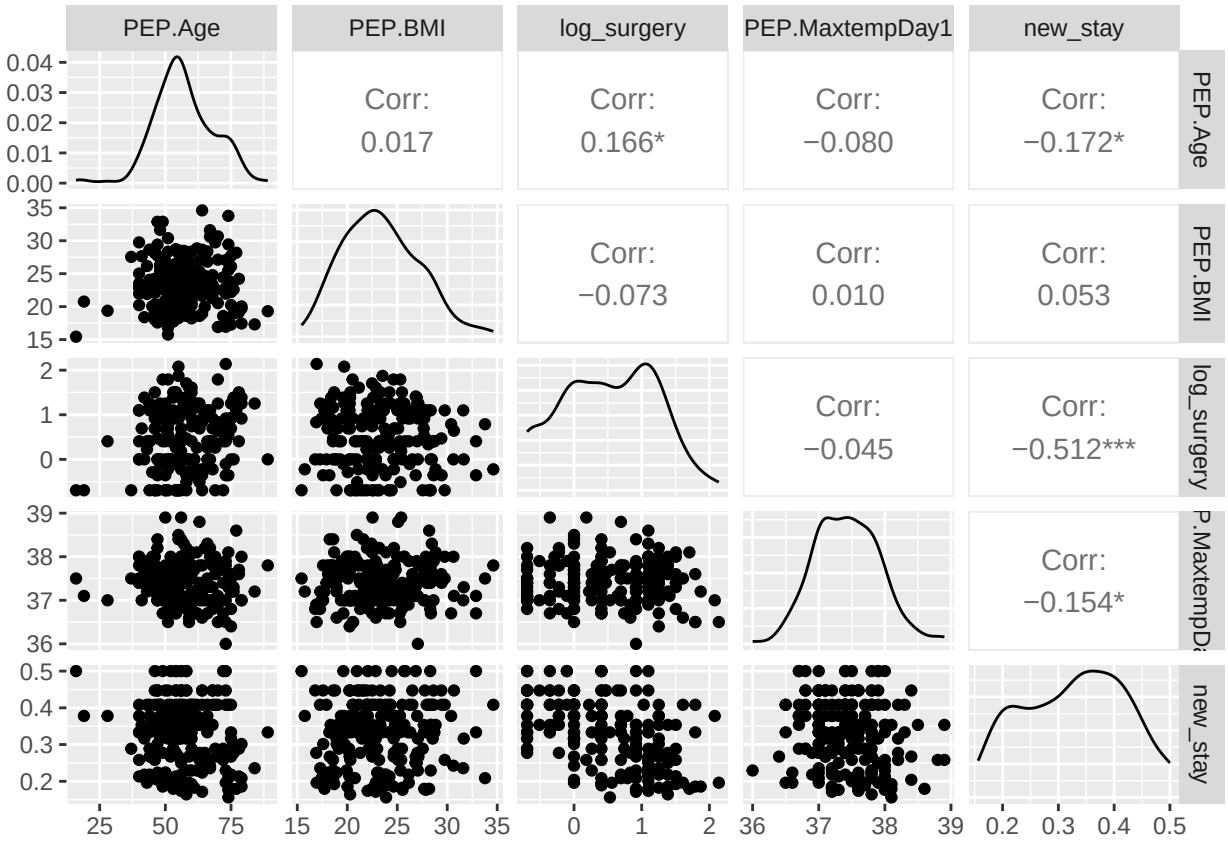

```
#Displayed histogram of each continuous variable
p4<- ggplot(PEP, aes(x=MaxtempDay1))+geom_histogram()
p5<- ggplot(PEP, aes(x=Hospitalstay))+geom_histogram()
grid.arrange(p4,p5, ncol=2)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Looking at the histograms, it appears that the continuous predictors relatively do not follow normal distribution, especially if we observe the duration of surgery and hospital stay, which are both evidently right-skewed. With this said, the continuous variables will most definitely benefit from transformations. Since these two show very strong concentration towards the left side of the histograms, then these would require strong transformation methods such as log transformation or square root transformation.

```
#Created new variables new_stay and log_surgery
#Obtained a scatterplot matrix of new_stay, log_surgery and all other numerical variables except Hospit
new_stay<-PEP$Hospitalstay^(-0.5)
log_surgery<-log(PEP$Durationofsurgery)
log_PEP<-data.frame(PEP$Age,PEP$BMI,log_surgery,PEP$MaxtempDay1,new_stay)
ggpairs(log_PEP)
```

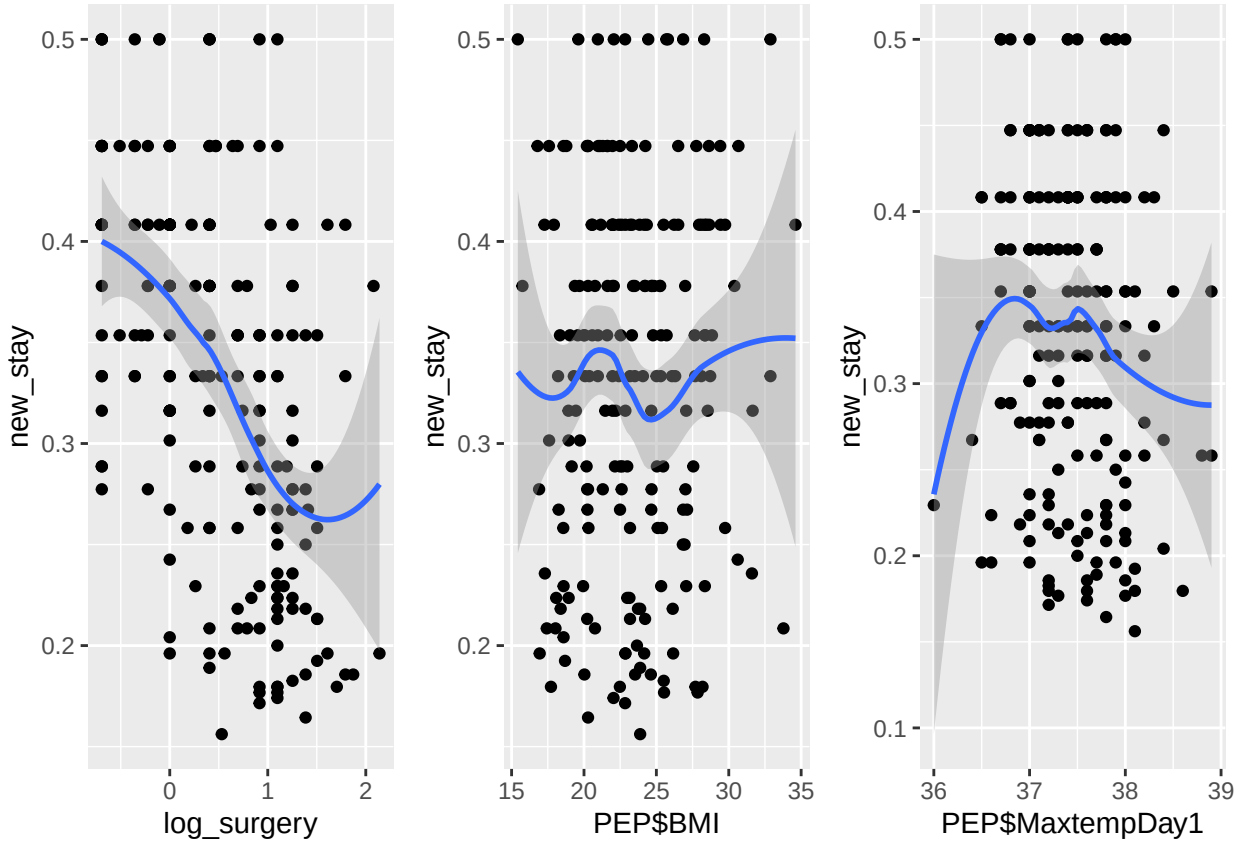


Referring to the obtained scatterplots, it seems that new_stay and log_surgery have somewhat of a negative correlation. Other than these two, it seems that there are no other correlated variables.

It seems that there is no symmetric pattern around a horizontal line at zero for all variable relationships. It can be seen that points are clustered, and some even show slight curvature. These deviations indicate non-normality. Moreover, the residuals are not evenly spread across all levels of the predicted values, which also suggests heteroscedasticity.

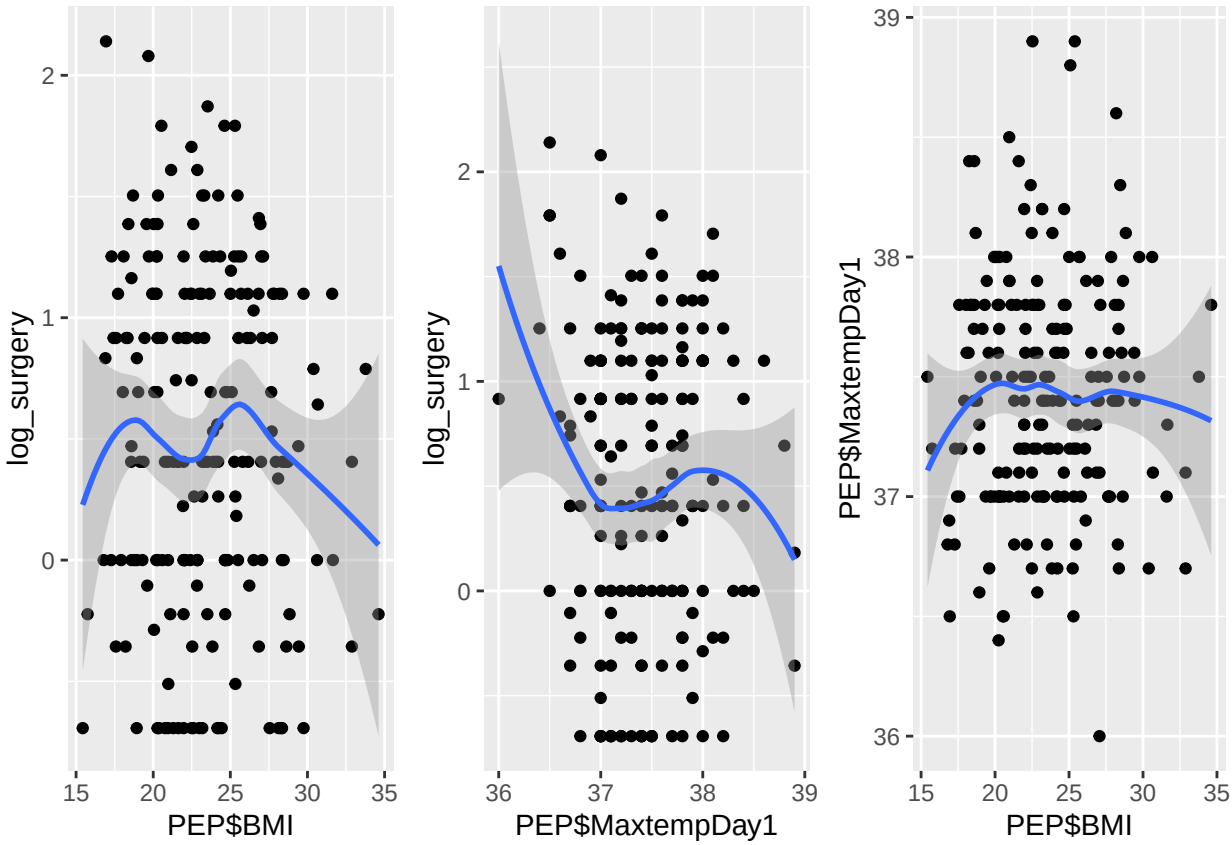
```
#Obtained a scatter plot matrix of new_stay, log_surgery, BMI and MaxtempDay1
#Added a loess curve to each sub-plot
log_PEP2<-data.frame(new_stay,log_surgery, PEP$BMI,PEP$MaxtempDay1)
p1 <- ggplot(data=log_PEP2,aes(x=log_surgery,y=new_stay))+geom_point()+geom_smooth(method="loess")
p2 <- ggplot(data=log_PEP2,aes(x=PEP$BMI,y=new_stay))+geom_point()+geom_smooth(method="loess")
p3 <- ggplot(data=log_PEP2,aes(x=PEP$MaxtempDay1,y=new_stay))+geom_point()+geom_smooth(method="loess")
grid.arrange(p1,p2,p3,ncol=3)
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```




```
#Obtained a scatter plot matrix of new_stay, log_surgery, BMI and MaxtempDay1
#Added a loess curve to each sub-plot
p4 <- ggplot(data=log_PEP2,aes(x=PEP$BMI,y=log_surgery))+geom_point()+geom_smooth(method="loess")
p5 <- ggplot(data=log_PEP2,aes(x=PEP$MaxtempDay1,y=log_surgery))+geom_point()+geom_smooth(method="loess
p6 <- ggplot(data=log_PEP2,aes(x=PEP$BMI,y=PEP$MaxtempDay1))+geom_point()+geom_smooth(method="loess")
grid.arrange(p4,p5,p6,ncol=3)
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```

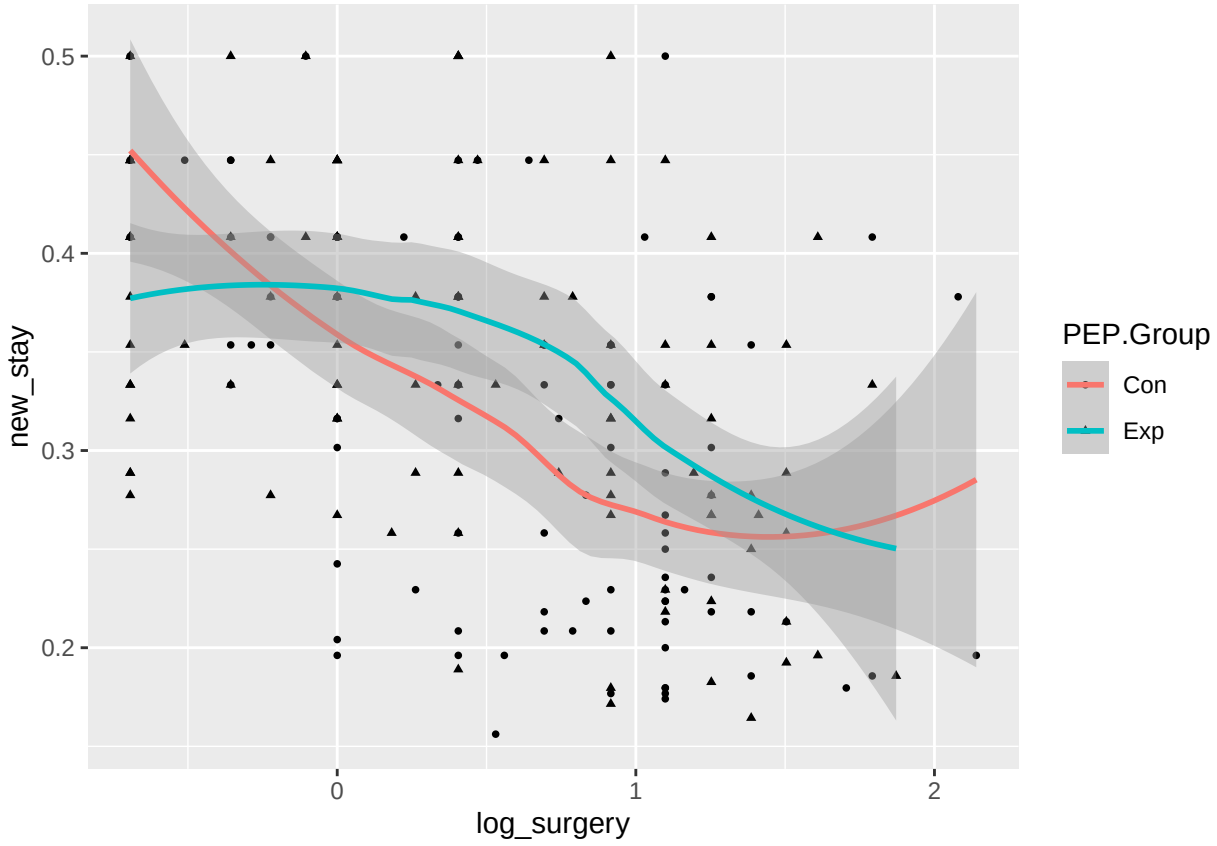


Based on the generated scatterplots with loess curve, log_surgery has the best linear relationship with new_stay. Moreover, BMI and MaxtempDay1 both do not have a good linear relationship with new_stay. Additionally, if we look at the relationships of the other variables with one another, it appears that they do not have a strong linear relationship as well, which means that there are no multicollinearity among the variables.

```
#created a data frame which also includes the non-numerical variables
log_PEP3<-data.frame(PEP$Group, PEP$Gender, new_stay,log_surgery, PEP$BMI,PEP$MaxtempDay1)

#Obtained a graph of new_stay versus log_surgery using different symbols for the different levels of Gr
#Added loess curves for the different levels of Group
log_PEP3$PEP.Group <- as.factor(log_PEP3$PEP.Group)
ggplot(log_PEP3, aes(x= log_surgery, y=new_stay, color=PEP.Group))+geom_point(aes(shape=PEP.Group),colo
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

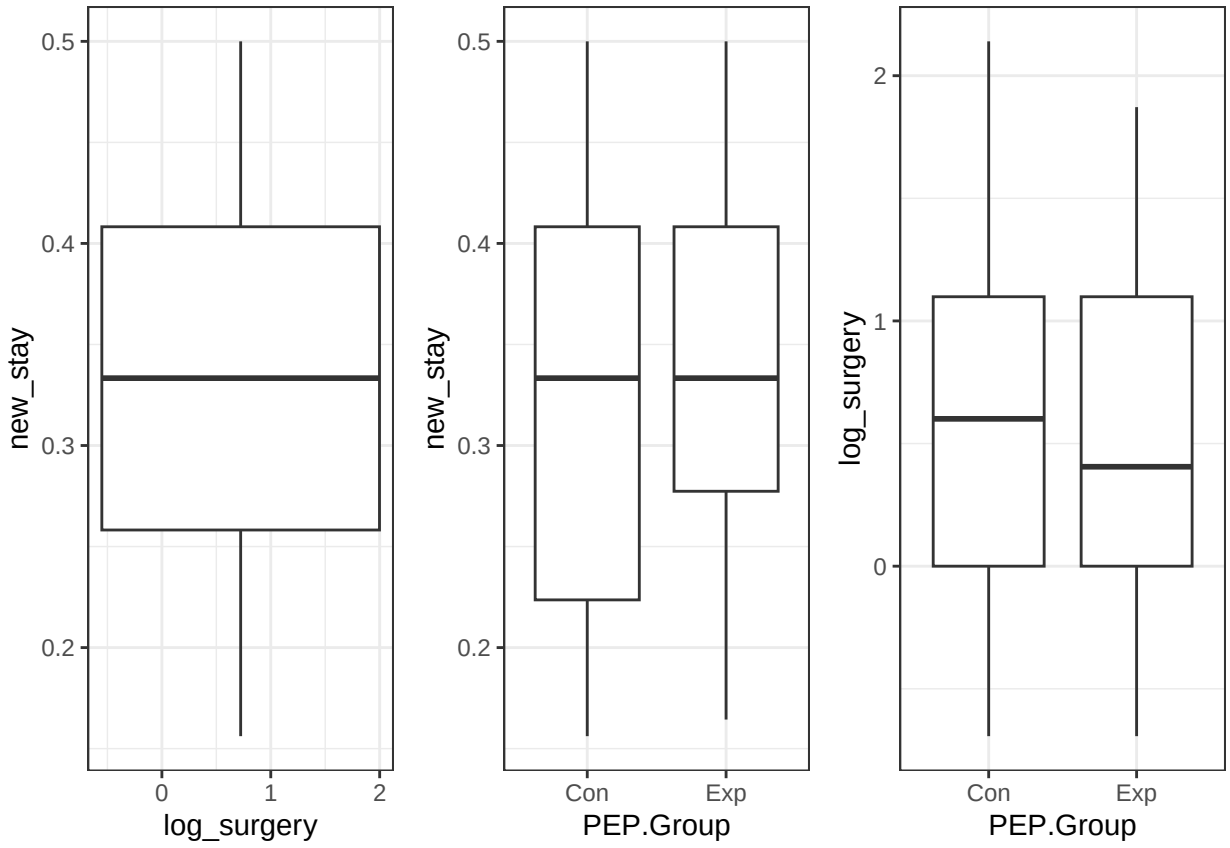


From the plot, it can be seen that the change in response to the changing duration of surgery becomes different as the Group factor levels change, especially for the intersection points. Based on the figure, for subjects that belong to the controlled group, there appears to be somewhat of a negative relationship where the length of stay decreases with the increase of surgery duration. However, when duration of surgery goes beyond the 1.5 mark it seems that the length of stay is increasing once again. On the contrary, for the experimental group, the loess curve seems to continue decreasing in terms of length of stay as duration surgery increases, showing no signs of increasing again even at the longer surgery duration points.Despite these reasons, we have to observe how for majority of the data points(log_surgery: range[0-1.5]); the two curves are approximately parallel from one another, which means that the rate of change for both levels are roughly the same for this range. This means that although at the intersection points there are changes in response to the changing duration of surgery as the Group factor levels change;for majority of the data points(within the interval of the two points of intersection), there is no evident change. For this reason, it appears that log_surgery and Group cannot be promising predictors.

To confirm our observation, we will be plotting a boxplot.

```
#plotted boxplot
p1 <- ggplot(log_PEP3, aes(x = log_surgery, y = new_stay)) + geom_boxplot() + theme_bw()
p2 <- ggplot(log_PEP3, aes(x = PEP.Group, y = new_stay)) + geom_boxplot() + theme_bw()
p3 <- ggplot(log_PEP3, aes(x = PEP.Group, y = log_surgery)) + geom_boxplot() + theme_bw()
grid.arrange(p1,p2,p3,ncol=3)
```

```
## Warning: Continuous x aesthetic
## i did you forget `aes(group = ...)`?
```



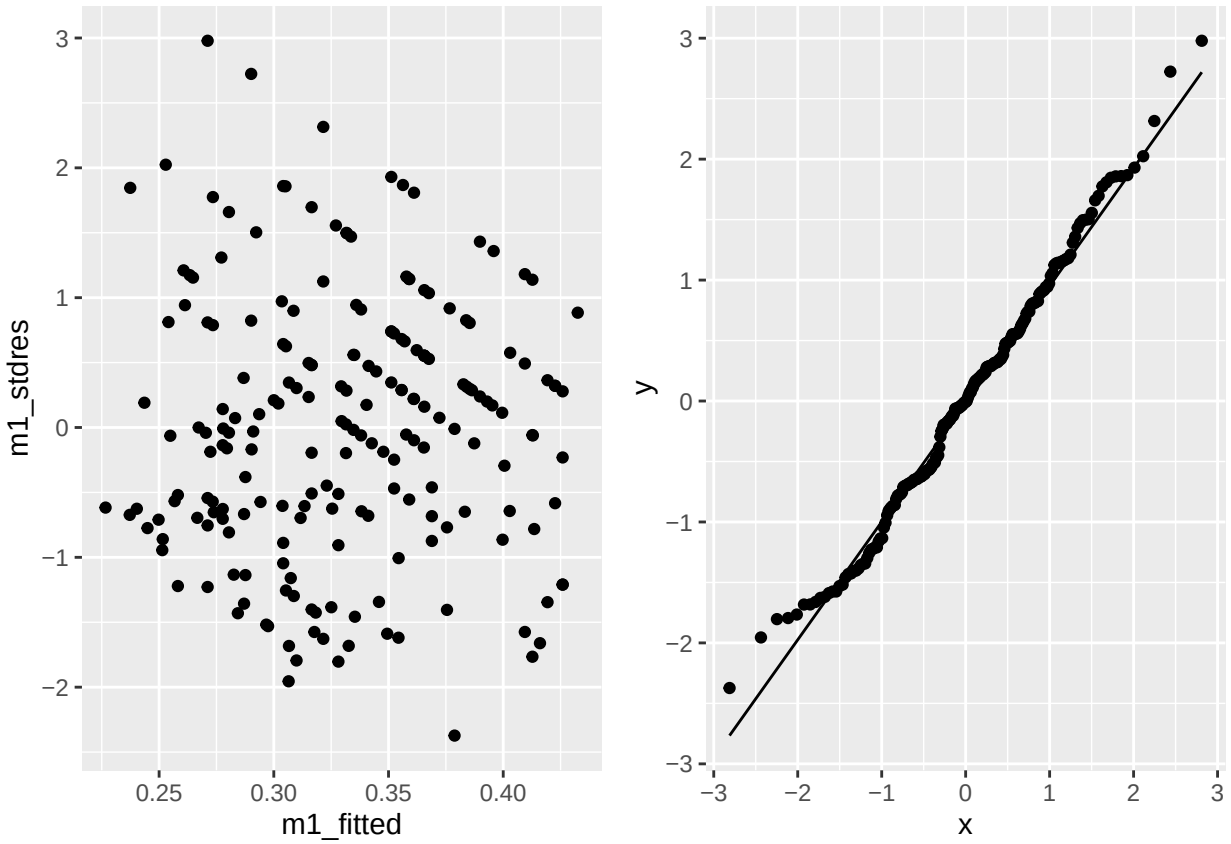
From the boxplot, we can observe that as the Group levels change from Con to Exp, there is not much change in the range of values for both new_stay and log_surgery, which means that changing levels have little to no effect, thus, confirming our previous observation that log_surgery and Group do not seem to be promising predictors.

```
m1<-lm(data=log_PEP3,new_stay ~ log_surgery + PEP.MaxtempDay1 )
m1_stdres <- rstandard(m1)
m1_fitted<-fitted(m1)
summary(m1)
```

```
##
## Call:
## lm(formula = new_stay ~ log_surgery + PEP.MaxtempDay1, data = log_PEP3)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.182670 -0.052590 -0.000605  0.048849  0.228846
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.597823   0.412566   3.873 0.000146 ***
## log_surgery     -0.067981   0.007768  -8.752 8.77e-16 ***
## PEP.MaxtempDay1 -0.032947   0.011015  -2.991 0.003130 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.07743 on 200 degrees of freedom
## Multiple R-squared:  0.2941, Adjusted R-squared:  0.2871
## F-statistic: 41.67 on 2 and 200 DF,  p-value: 7.447e-16
```

```
mydata1<-tibble(m1_stdres,m1_fitted)
p1<-ggplot(mydata1,aes(y=m1_stdres,x=m1_fitted))+geom_point()
p2<-ggplot(mydata1,aes(sample=m1_stdres))+stat_qq()+stat_qq_line()
grid.arrange(p1,p2,ncol=2)
```



Based on the residuals vs fitted plot, the residuals seem to be somewhat randomly scattered with no discernible pattern especially for residuals below the horizontal line at 0. However, above the horizontal line at 0, it seems that there is a bit of downward pattern for some points. This suggests that the model may somewhat be appropriate. On the other hand, the Normal probability plot of residuals appear to be generally linear, with only slight curvatures towards the beginning and the end of the line; hinting that perhaps, normality can actually be assumed and that the model may be deemed appropriate.

The multiple linear regression model can be written as:

$$newstay = \beta 0 + \beta 1 logsurgery + \beta 2 MaxtempDay1 + \epsilon$$

We have the fitted equation as:

$$newstay = 1.597823\text{-}0.067981 * logsurgery\text{-}0.032947 * MaxtempDay1 + \epsilon$$

With a p-value of 7.447e-16, the model is significant. Moreover, both log_surgery and MaxtempDay1 are significant predictors of new_stay as their p-values are both very small values close to 0. Despite this, it is

important to note that their R-squared value of 0.2941 suggests that the model is explaining only a relatively small proportion of the variability in the response variable, new_stay.

Based on the scatter plots generated earlier, it is clear that Age is not a confounding variable because it does not have a strong relationship with any of the other predictors. To further confirm this, we will be fitting a model with Age as a predictor.

```
#created a data frame which also includes age
log_PEP4<-data.frame(PEP$Age,PEP$Group, PEP$Gender, new_stay,log_surgery, PEP$BMI,PEP$MaxtempDay1)

m2<-lm(data=log_PEP4,new_stay ~ log_surgery + PEP.MaxtempDay1 + PEP.Age )
summary(m2)
```

```
##
## Call:
## lm(formula = new_stay ~ log_surgery + PEP.MaxtempDay1 + PEP.Age,
##     data = log_PEP4)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.186443 -0.053716  0.000507  0.049654  0.227453
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.6973307  0.4146459   4.093 6.18e-05 ***
## log_surgery     -0.0657841  0.0078353  -8.396 8.59e-15 ***
## PEP.MaxtempDay1 -0.0343394  0.0109920  -3.124  0.00205 **
## PEP.Age         -0.0008503  0.0004953  -1.717  0.08763 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07705 on 199 degrees of freedom
## Multiple R-squared:  0.3044, Adjusted R-squared:  0.294
## F-statistic: 29.03 on 3 and 199 DF,  p-value: 1.295e-15
```

Looking at the summary, the coefficients of log_surgery and MaxtempDay1 did not drastically change. The change was very small for both. Moreover, both are still significant with very small p-values still close to 0 and close to the previous p-values both had for the first fitted model. Additionally, it is important to note that Age is insignificant with a p-value of 0.08763. With these said, we can now confirm that Age is definitely not a confounding factor.

Given that Group is a categorical predictor,we will be interested to observe its different levels' interaction with the other predictors of the model. For the next part, we will fit a multiple regression model including interaction terms for Group with the other predictors from the model.

```
#fitted multiple regression model including interaction terms for Group with the other predictors from
m3<-lm(data = log_PEP4, new_stay ~ log_surgery * PEP.Group + PEP.MaxtempDay1* PEP.Group)

m3_stdres <- rstandard(m3)
m3_fitted<-fitted(m3)
summary(m3)
```

```
##
## Call:
```

```
## lm(formula = new_stay ~ log_surgery * PEP.Group + PEP.MaxtempDay1 *
##     PEP.Group, data = log_PEP4)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.190415 -0.051952  0.000405  0.044421  0.251454
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    2.04621    0.57186   3.578 0.000436 ***
## log_surgery                   -0.08477    0.01153  -7.354 5.05e-12 ***
## PEP.GroupExp                  -0.91440    0.82706  -1.106 0.270244
## PEP.MaxtempDay1               -0.04486    0.01523  -2.946 0.003611 **
## log_surgery:PEP.GroupExp       0.03118    0.01557   2.002 0.046621 *
## PEP.GroupExp:PEP.MaxtempDay1   0.02441    0.02208   1.106 0.270179
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07667 on 197 degrees of freedom
## Multiple R-squared:  0.3183, Adjusted R-squared:  0.301
## F-statistic:  18.4 on 5 and 197 DF,  p-value: 5.624e-15
```

Given that this is an unbalanced study, and given that we are interested to include the interaction term (if there are any significant) between the two levels; the model for this situation is:

$$Yi = \mu + \alpha i + \beta 1 + \beta 2 + \gamma i1 + \gamma i2 + \epsilon i12k$$

where:

Yi is the Hospitalstay response;

μ is the overall mean or intercept;

αi is the main Group effect for i = 1(Con), 2(Exp);

$\beta 1$ is the main effect for Durationofsurgery;

$\beta 2$ is the main effect for MaxtempDay1;

γ i1 is the interaction effect between group and Durationofsurgery;

γ i2 is the interaction effect between group and MaxtempDay1;

ϵ i12k is the random error term or residual term, which represents the unexplained variation in the response variable

We have the fitted equation as:

$$newstay = 2.04621\text{-}0.91440 * Group\text{-}0.08477 * logsurgery$$
$$\text{-}0.04486 * MaxtempDay1 + 0.03118 * logsurgery : Group$$
$$+0.02441 * MaxtempDay1 : Group + \epsilon i12k$$

```r
#Got the 95% confidence intervals for the regression coefficients
confint(m3)
```

```
##                                     2.5 %      97.5 %
## (Intercept)                    0.9184563395  3.17397111
## log_surgery                   -0.1075023007 -0.06203731
## PEP.GroupExp                  -2.5454190501  0.71661904
## PEP.MaxtempDay1               -0.0748874301 -0.01482515
## log_surgery:PEP.GroupExp       0.0004710881  0.06189508
## PEP.GroupExp:PEP.MaxtempDay1 -0.0191275855  0.06795607
```

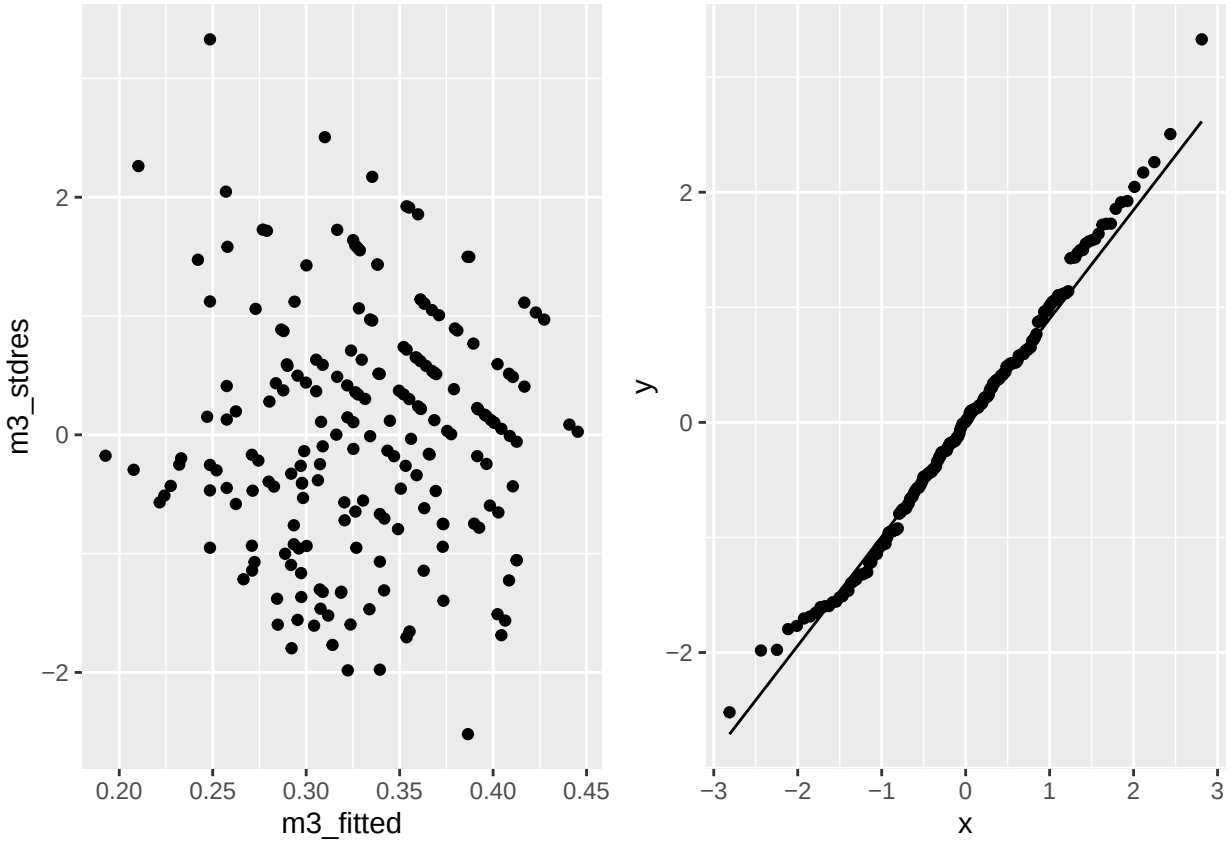
```r
#Got the table to see number of counts
table(log_PEP4$PEP.Group)
```

```
##
## Con Exp
## 104  99
```

Notice how by default, R already selected the first level (Con) as the reference category. Referring to the table above, we can see that Con is the level with the highest number of counts which makes sense for us to stick with using Con as the reference category. Therefore, as also instructed, I shall proceed with Con as the reference category by not changing the reference category.

Overall regression is significant, with p-value: 5.624e-15 which is very close to 0. However, only two main effects appear to be significant with p-values of 5.05e-12(log_surgery) and 0.003611(MaxtempDay1).Group seems to be insignificant with a p-value of 0.270244.For the interaction terms, only the interaction term between Group and log_surgery is significant with a p-value of 0.046621. The interaction term between Group and MaxtempDay1 is insignificant with a p-value of 0.270179. Therefore, since the interaction term between Group and log_surgery is significant; even if Group's main effect is insignificant, it is recommended to retain Group as it has a significant interaction with log_surgery.Next, we can also see that all terms including interaction terms have coefficients within their respective confidence intervals. However, it is important to note that all terms have a wide interval where we have to be extra careful of our predictions. Moreover,the R2 = 31.83% This means that the model only explains about 31.83% of the variation.

```r
#checked the assumptions
mydata3<-tibble(m3_stdres,m3_fitted)
p1<-ggplot(mydata3,aes(y=m3_stdres,x=m3_fitted))+geom_point()
p2<-ggplot(mydata3,aes(sample=m3_stdres))+stat_qq()+stat_qq_line()
grid.arrange(p1,p2,ncol=2)
```



Based on the residuals vs fitted plot, the residuals seem to be somewhat randomly scattered with no discernible pattern especially for residuals below the horizontal line at 0. However, above the horizontal line at 0, it seems that there is a bit of downward pattern for some points. This suggests that the model may somewhat be appropriate. On the other hand, the Normal probability plot of residuals appear to be generally linear, with only slight curvatures towards the beginning and the end of the line; hinting that perhaps, normality can actually be assumed and that the model may be deemed appropriate. Therefore, I can say that the assumptions have somewhat been met. However, the model can further be improved by removing the insignificant predictors such as the interaction term between Group and MaxtempDay1.

```
#checked the assumptions
m3_stdres<-rstandard(m3)
leverage_m3<-hatvalues(m3)
CooksD_m3<-cooks.distance(m3)
print(data.frame(log_PEP4,m3_stdres,leverage_m3,CooksD_m3))
```

```
##     PEP.Age PEP.Group PEP.Gender  new_stay log_surgery PEP.BMI PEP.MaxtempDay1
## 1        48       Con          M 0.4082483   0.0000000   25.50            37.4
## 2        44       Con          M 0.2672612   1.0986123   22.49            37.8
## 3        53       Con          M 0.1796053   1.0986123   17.72            37.2
## 4        56       Con          F 0.4082483   1.0296194   26.51            37.5
## 5        68       Con          M 0.4472136   0.4054651   21.30            36.8
## 6        46       Con          M 0.2886751   1.0986123   20.18            37.8
## 7        74       Con          M 0.2085144   0.7884574   33.78            37.5
## 8        45       Con          M 0.3535534  -0.2876821   20.05            38.0
## 9        63       Con          F 0.2773501   1.2527630   27.01            37.4
## 10       73       Con          M 0.5000000  -0.1053605   22.84            37.9
```

```
## 11       70       Con          M 0.2773501   0.8329091   16.90            36.9
## 12       58       Con          M 0.4472136   0.4700036   18.59            37.4
## 13       43       Con          F 0.4082483   0.2231436   21.93            37.2
## 14       51       Con          F 0.1961161   0.5596158   24.16            37.7
## 15       56       Con          M 0.5000000   1.0986123   25.69            38.0
## 16       58       Con          F 0.1796053   1.7047481   22.48            38.1
## 17       51       Con          M 0.2000000   1.0986123   23.66            37.5
## 18       73       Con          F 0.2294157   0.9162907   27.06            36.0
## 19       67       Con          M 0.2581989   0.4054651   23.18            38.2
## 20       53       Con          M 0.3535534   0.9162907   18.34            37.8
## 21       62       Con          F 0.2294157   0.2623643   25.33            37.2
## 22       48       Con          M 0.3333333   0.9162907   26.30            37.3
## 23       40       Con          M 0.3333333   0.4054651   23.51            37.4
## 24       51       Con          M 0.4082483   0.0000000   28.34            37.4
## 25       40       Con          M 0.2581989   1.0986123   29.75            37.5
## 26       60       Con          M 0.3162278   0.7419373   21.47            37.8
## 27       84       Con          M 0.2357023   1.2527630   17.30            37.2
## 28       79       Con          M 0.3015113   1.2527630   19.72            37.0
## 29       52       Con          M 0.2085144   0.4054651   20.76            38.0
## 30       73       Con          F 0.4472136  -0.3566749   17.58            37.8
## 31       58       Con          M 0.4472136  -0.5108256   20.98            37.9
## 32       16       Con          F 0.5000000  -0.6931472   15.43            37.5
## 33       60       Con          M 0.3535534   0.9162907   20.57            37.0
## 34       52       Con          F 0.1961161   0.0000000   22.86            37.0
## 35       55       Con          M 0.3779645   2.0794415   19.69            37.0
## 36       78       Con          M 0.2132007   1.5040774   24.22            37.6
## 37       66       Con          M 0.3535534   0.4054651   28.37            36.7
## 38       40       Con          M 0.2132007   1.0986123   20.22            38.0
## 39       40       Con          M 0.3333333   1.0986123   25.01            38.0
## 40       79       Con          F 0.1856953   1.3862944   20.03            38.0
## 41       44       Con          M 0.2182179   1.2527630   23.88            37.4
## 42       69       Con          M 0.4082483  -0.6931472   29.75            38.0
## 43       67       Con          M 0.2425356   0.0000000   30.62            38.0
## 44       48       Con          F 0.3779645   0.0000000   20.28            37.7
## 45       40       Con          M 0.3779645   0.0000000   21.97            37.4
## 46       64       Con          F 0.4472136  -0.3566749   28.63            37.0
## 47       66       Con          M 0.2500000   1.0986123   26.83            37.3
## 48       48       Con          F 0.3162278   0.0000000   31.65            37.3
## 49       47       Con          M 0.4082483  -0.6931472   23.18            38.2
## 50       63       Con          M 0.2581989   0.6931472   25.09            38.8
## 51       44       Con          M 0.3333333   0.4054651   24.07            37.6
## 52       52       Con          F 0.3162278   0.0000000   24.65            37.8
## 53       67       Con          F 0.2357023   1.0986123   31.60            37.0
## 54       77       Con          F 0.1796053   1.0986123   28.19            38.6
## 55       28       Con          M 0.3779645   0.4054651   19.37            37.0
## 56       44       Con          M 0.3779645   1.2527630   23.39            37.5
## 57       49       Con          M 0.1856953   1.7917595   24.62            37.6
## 58       74       Con          F 0.1767767   0.9162907   25.53            37.3
## 59       50       Con          M 0.2236068   1.0986123   22.99            37.8
## 60       76       Con          M 0.3015113   0.0000000   18.94            37.3
## 61       42       Con          F 0.2182179   1.3862944   18.39            37.8
## 62       59       Con          F 0.4082483   0.0000000   20.59            36.5
## 63       47       Con          F 0.2236068   1.0986123   23.14            37.0
## 64       62       Con          F 0.3779645   1.2527630   25.25            36.7
```

```
## 65       50       Con          M 0.3535534  -0.3566749   22.53           38.9
## 66       72       Con          F 0.5000000  -0.6931472   24.44           37.4
## 67       68       Con          F 0.4082483  -0.2231436   21.97           37.3
## 68       50       Con          M 0.3779645   0.4054651   26.99           37.5
## 69       50       Con          M 0.3333333   0.6931472   20.07           37.5
## 70       57       Con          M 0.4082483   0.0000000   28.46           38.3
## 71       70       Con          F 0.4472136   0.6418539   30.67           37.1
## 72       56       Con          F 0.4472136  -0.6931472   20.28           37.0
## 73       53       Con          M 0.3779645   0.4054651   25.26           37.3
## 74       74       Con          F 0.4082483  -0.3566749   29.44           37.4
## 75       64       Con          M 0.4082483  -0.2231436   34.61           37.8
## 76       51       Con          F 0.3779645  -0.2231436   15.75           37.2
## 77       74       Con          M 0.1561738   0.5306283   23.88           38.1
## 78       56       Con          M 0.4082483   0.4054651   27.95           37.4
## 79       64       Con          M 0.2236068   0.8329091   18.94           36.6
## 80       55       Con          M 0.3333333   0.3364722   28.10           37.8
## 81       61       Con          F 0.3779645   0.0000000   22.06           37.2
## 82       53       Con          M 0.3535534   1.3862944   19.57           37.6
## 83       63       Con          M 0.2294157   1.1631508   18.59           37.8
## 84       60       Con          F 0.4082483   0.4054651   23.32           37.0
## 85       47       Con          M 0.2041241   0.0000000   18.59           38.4
## 86       62       Con          M 0.3535534  -0.2231436   28.84           38.1
## 87       75       Con          M 0.4082483   0.0000000   17.27           36.8
## 88       46       Con          M 0.2085144   0.6931472   18.03           37.8
## 89       46       Con          M 0.3535534   0.6931472   19.03           37.5
## 90       43       Con          M 0.1961161   0.4054651   26.16           37.9
## 91       66       Con          M 0.4082483   0.0000000   22.41           37.5
## 92       59       Con          F 0.2294157   1.0986123   19.94           38.0
## 93       56       Con          F 0.1740777   1.0986123   22.04           37.6
## 94       56       Con          F 0.1767767   1.0986123   27.85           38.0
## 95       41       Con          F 0.3535534   0.6931472   27.65           37.0
## 96       70       Con          M 0.4082483   1.7917595   20.53           36.5
## 97       73       Con          F 0.1961161   2.1400662   16.94           36.5
## 98       47       Con          F 0.3015113   0.9162907   17.59           37.0
## 99       55       Con          F 0.3162278   0.4054651   28.55           37.6
## 100      75       Con          F 0.3333333   0.9162907   26.08           37.2
## 101      75       Con          F 0.2182179   0.6931472   23.73           37.2
## 102      60       Con          F 0.3333333  -0.3566749   18.21           37.6
## 103      79       Con          F 0.2085144   0.9162907   17.43           37.0
## 104      50       Con          M 0.4472136  -0.6931472   22.49           37.1
## 105      60       Exp          M 0.4082483   1.6094379   21.16           37.5
## 106      56       Exp          M 0.2500000   1.3862944   26.95           37.9
## 107      55       Exp          M 0.3333333  -0.6931472   20.37           37.0
## 108      56       Exp          F 0.2182179   1.0986123   26.13           36.9
## 109      57       Exp          M 0.3535534   1.2527630   25.71           38.0
## 110      61       Exp          M 0.1825742   1.2527630   25.51           37.2
## 111      56       Exp          F 0.2581989   0.1823216   25.39           38.9
## 112      37       Exp          M 0.2886751  -0.6931472   27.55           37.5
## 113      56       Exp          F 0.3162278   0.0000000   27.06           37.1
## 114      59       Exp          M 0.2672612   1.4109870   26.85           37.1
## 115      49       Exp          M 0.2773501   0.9162907   20.28           37.0
## 116      44       Exp          M 0.2886751  -0.6931472   22.99           37.3
## 117      72       Exp          M 0.2581989   0.4054651   18.57           37.7
## 118      60       Exp          M 0.2773501   1.2527630   20.22           37.1
```

```
## 119      46       Exp          F 0.2581989   1.5040774   20.31            38.0
## 120      73       Exp          M 0.4082483   0.4054651   21.97            37.0
## 121      55       Exp          M 0.3162278   0.9162907   19.44            37.9
## 122      76       Exp          M 0.2773501   1.3862944   22.60            37.2
## 123      56       Exp          M 0.3333333   0.5306283   27.68            37.0
## 124      78       Exp          M 0.2886751   0.4054651   19.15            37.7
## 125      42       Exp          M 0.3162278   0.0000000   22.86            37.6
## 126      52       Exp          F 0.4082483   0.0000000   17.92            37.4
## 127      47       Exp          M 0.4472136   0.0000000   18.78            37.4
## 128      60       Exp          M 0.2886751   1.1939225   25.04            37.2
## 129      54       Exp          M 0.3779645   0.4054651   21.63            37.2
## 130      51       Exp          F 0.3162278  -0.6931472   18.93            37.2
## 131      89       Exp          M 0.3333333   0.0000000   19.30            37.8
## 132      72       Exp          M 0.2886751   0.7419373   22.49            36.7
## 133      55       Exp          F 0.2672612   0.9162907   18.25            38.4
## 134      66       Exp          M 0.3779645   0.2623643   24.22            37.0
## 135      55       Exp          M 0.4082483  -0.6931472   22.60            37.8
## 136      66       Exp          M 0.3333333   1.0986123   22.41            38.3
## 137      54       Exp          F 0.2773501  -0.6931472   21.30            37.4
## 138      55       Exp          M 0.3333333   0.0000000   21.97            37.3
## 139      58       Exp          M 0.1643990   1.3862944   20.28            37.8
## 140      56       Exp          M 0.1889822   0.4054651   23.88            37.7
## 141      54       Exp          M 0.3333333   0.2623643   23.18            37.4
## 142      46       Exp          F 0.4082483  -0.6931472   28.13            37.4
## 143      19       Exp          F 0.3779645  -0.6931472   20.76            37.1
## 144      55       Exp          F 0.3535534  -0.6931472   28.35            37.7
## 145      54       Exp          M 0.4082483  -0.6931472   24.16            37.0
## 146      54       Exp          M 0.4472136  -0.6931472   21.97            37.5
## 147      40       Exp          M 0.4082483   0.0000000   22.86            37.8
## 148      70       Exp          M 0.2672612   1.2527630   27.12            37.8
## 149      54       Exp          M 0.4472136   0.0000000   26.51            37.6
## 150      50       Exp          F 0.2236068   1.2527630   18.08            37.6
## 151      56       Exp          F 0.4472136   0.0000000   20.20            37.6
## 152      58       Exp          M 0.3779645   0.0000000   24.68            37.0
## 153      50       Exp          F 0.3333333  -0.6931472   20.96            37.0
## 154      55       Exp          F 0.3535534   0.0000000   20.96            38.5
## 155      65       Exp          F 0.2773501  -0.2231436   24.67            38.2
## 156      54       Exp          M 0.3333333   0.4054651   28.72            37.5
## 157      52       Exp          F 0.2672612   0.0000000   24.65            37.1
## 158      75       Exp          F 0.1796053   0.9162907   27.70            37.6
## 159      44       Exp          M 0.4082483   0.4054651   28.65            37.9
## 160      54       Exp          M 0.3779645   0.4054651   24.21            36.7
## 161      58       Exp          F 0.3162278   1.2527630   21.97            38.2
## 162      60       Exp          M 0.2294157   1.0986123   28.35            37.8
## 163      68       Exp          F 0.3535534   0.6931472   24.77            37.8
## 164      55       Exp          M 0.1856953   1.8718022   23.53            37.2
## 165      46       Exp          F 0.4472136  -0.6931472   24.24            37.2
## 166      51       Exp          M 0.4472136   0.0000000   16.79            36.8
## 167      58       Exp          M 0.5000000  -0.3566749   26.85            37.4
## 168      47       Exp          M 0.3333333  -0.3566749   32.88            37.1
## 169      50       Exp          F 0.3779645   0.0000000   24.84            37.7
## 170      63       Exp          F 0.3535534  -0.5108256   25.33            37.0
## 171      51       Exp          F 0.3779645   0.6931472   24.68            37.7
## 172      61       Exp          F 0.4082483   0.0000000   22.06            37.7
```

```
## 173      60       Exp          M 0.3779645  -0.2231436   23.51            36.8
## 174      48       Exp          M 0.4082483  -0.1053605   26.23            37.1
## 175      47       Exp          M 0.4472136   0.4700036   29.41            37.6
## 176      62       Exp          M 0.4082483  -0.3566749   23.84            36.7
## 177      51       Exp          F 0.3779645   0.7884574   30.39            36.7
## 178      54       Exp          F 0.5000000  -0.1053605   19.61            36.7
## 179      56       Exp          M 0.4472136  -0.2231436   21.11            37.8
## 180      60       Exp          M 0.1924501   1.5040774   18.69            38.1
## 181      47       Exp          M 0.4472136   1.0986123   27.76            37.0
## 182      69       Exp          F 0.2886751   0.2623643   22.65            37.6
## 183      48       Exp          F 0.4082483   1.2527630   24.34            37.4
## 184      75       Exp          M 0.3162278   0.9162907   22.20            37.5
## 185      54       Exp          M 0.2132007   1.5040774   23.18            37.3
## 186      54       Exp          M 0.3535534   1.5040774   23.31            37.4
## 187      56       Exp          F 0.4472136   0.6931472   20.27            37.0
## 188      58       Exp          M 0.5000000  -0.6931472   28.30            36.8
## 189      64       Exp          F 0.4472136  -0.6931472   21.63            37.1
## 190      53       Exp          M 0.2886751   1.5040774   25.47            36.8
## 191      51       Exp          F 0.3333333   1.7917595   25.30            36.5
## 192      49       Exp          F 0.5000000   0.4054651   32.88            36.7
## 193      52       Exp          M 0.5000000   0.4054651   25.80            37.0
## 194      54       Exp          F 0.5000000   0.4054651   20.96            37.9
## 195      68       Exp          F 0.4472136   0.9162907   23.31            37.0
## 196      53       Exp          F 0.2886751   0.9162907   22.14            36.8
## 197      56       Exp          M 0.4472136   0.4054651   21.60            38.4
## 198      46       Exp          F 0.5000000   0.9162907   22.06            37.8
## 199      59       Exp          M 0.1961161   1.6094379   22.86            36.6
## 200      65       Exp          F 0.3535534   1.0986123   19.98            37.0
## 201      75       Exp          M 0.2672612   1.2527630   20.24            36.4
## 202      53       Exp          M 0.3535534   0.9162907   21.60            37.0
## 203      64       Exp          M 0.1714986   0.9162907   22.84            37.2
##        m3_stdres leverage_m3    CooksD_m3
## 1    0.521823661  0.01727999 7.980129e-04
## 2    0.128501633  0.02170399 6.105699e-05
## 3   -1.379851707  0.01815282 5.866968e-03
## 4    1.727043082  0.01478027 7.457689e-03
## 5    1.139775961  0.02956277 6.595769e-03
## 6    0.410891245  0.02170399 6.242691e-04
## 7   -1.163962514  0.01088735 2.485439e-03
## 8   -0.165846678  0.03229105 1.529677e-04
## 9    0.197123161  0.02039160 1.348102e-04
## 10   1.913024132  0.02383815 1.489499e-02
## 11  -0.568365101  0.02346033 1.293445e-03
## 12   1.553093048  0.01013067 4.114377e-03
## 13   0.652295226  0.01615215 1.164232e-03
## 14  -1.463730009  0.01143888 4.131909e-03
## 15   3.328761294  0.02920007 5.554798e-02
## 16  -0.175925612  0.06061383 3.328392e-04
## 17  -0.933408206  0.01637770 2.417779e-03
## 18  -1.704148511  0.09491712 5.075979e-02
## 19  -0.531319165  0.02931806 1.421078e-03
## 20   1.060348866  0.01746884 3.331685e-03
## 21  -1.655076599  0.01550034 7.188044e-03
## 22   0.498105366  0.01333830 5.590150e-04
```

```
## 23  -0.011590995  0.01052083 2.380857e-07
## 24   0.521823661  0.01727999 7.980129e-04
## 25  -0.168007685  0.01637770 7.833065e-05
## 26   0.374201373  0.01482458 3.511788e-04
## 27  -0.470363103  0.02209302 8.330534e-04
## 28   0.280014952  0.02695063 3.619472e-04
## 29  -1.301543493  0.01988448 5.728002e-03
## 30   0.878370742  0.02977720 3.946543e-03
## 31   0.768065234  0.03817946 3.902839e-03
## 32   1.029406411  0.04467147 8.258477e-03
## 33   0.589004015  0.02032851 1.199804e-03
## 34  -2.519477872  0.02824421 3.074979e-02
## 35   2.262166197  0.06495725 5.925076e-02
## 36  -0.250704531  0.03145737 3.402333e-04
## 37  -0.160215026  0.03549808 1.574552e-04
## 38  -0.467896294  0.02920007 1.097494e-03
## 39   1.122422922  0.02920007 6.315620e-03
## 40  -0.511888915  0.03944504 1.793372e-03
## 41  -0.582143470  0.02039160 1.175730e-03
## 42   0.104568830  0.04961820 9.514723e-05
## 43  -1.309244267  0.02450499 7.176614e-03
## 44   0.300432790  0.01734179 2.654818e-04
## 45   0.123364372  0.01727999 4.460072e-05
## 46   0.405659287  0.04158934 1.190151e-03
## 47  -0.393899210  0.01677207 4.411145e-04
## 48  -0.748549341  0.01883748 1.792963e-03
## 49   0.225491322  0.05712021 5.133834e-04
## 50   0.151841829  0.07980850 3.332748e-04
## 51   0.106042590  0.01048587 1.986057e-05
## 52  -0.453240721  0.01894048 6.610011e-04
## 53  -0.761511283  0.02328145 2.303785e-03
## 54  -0.568612022  0.07062538 4.094971e-03
## 55   0.340005882  0.02005928 3.944002e-04
## 56   1.582439682  0.02072445 8.832449e-03
## 57  -0.294249045  0.04596447 6.952429e-04
## 58  -1.557669669  0.01333830 5.466782e-03
## 59  -0.447178919  0.02170399 7.394022e-04
## 60  -0.942334260  0.01883748 2.841453e-03
## 61  -0.197636453  0.03144315 2.113410e-04
## 62  -0.009561918  0.05970299 9.675419e-07
## 63  -0.921145506  0.02328145 3.370899e-03
## 64   1.120626003  0.04015487 8.756028e-03
## 65   0.301996514  0.09599066 1.614019e-03
## 66   0.970246522  0.04604925 7.573726e-03
## 67   0.214866035  0.02592441 2.047861e-04
## 68   0.632320122  0.01010882 6.805122e-04
## 69   0.366914725  0.01007378 2.283329e-04
## 70   1.064706137  0.03876959 7.620299e-03
## 71   1.571810766  0.01535697 6.422083e-03
## 72   0.025884781  0.05945084 7.058530e-06
## 73   0.515129025  0.01172187 5.245641e-04
## 74   0.124773140  0.02937092 7.851545e-05
## 75   0.511040631  0.02504659 1.118212e-03
## 76  -0.244958297  0.02846710 2.930345e-04
```

```
## 77  -1.796380155  0.02439185 1.344666e-02
## 78   0.970730036  0.01052083 1.669894e-03
## 79  -1.467934653  0.04013898 1.501828e-02
## 80   0.147249755  0.01398982 5.127300e-05
## 81   0.005335667  0.02118401 1.026912e-07
## 82   1.473431922  0.02659745 9.886816e-03
## 83  -0.298713054  0.02356332 3.588802e-04
## 84   0.739029822  0.02005928 1.863323e-03
## 85  -1.596505169  0.04510255 2.006475e-02
## 86  -0.033860987  0.03398843 6.723520e-06
## 87   0.169542751  0.03846059 1.916267e-04
## 88  -1.095345855  0.01433074 2.907292e-03
## 89   0.631989989  0.01007378 6.774210e-04
## 90  -1.521248946  0.01635126 6.411513e-03
## 91   0.580616354  0.01651154 9.432909e-04
## 92  -0.253241401  0.02920007 3.214936e-04
## 93  -1.215942194  0.01736409 4.354457e-03
## 94  -0.950078585  0.02920007 4.525036e-03
## 95   0.339460308  0.01875973 3.671791e-04
## 96   2.046694696  0.07180892 5.401270e-02
## 97  -0.429985800  0.09101267 3.085326e-03
## 98  -0.096803697  0.02032851 3.240842e-05
## 99  -0.118249893  0.01048587 2.469634e-05
## 100  0.439547107  0.01487933 4.863551e-04
## 101 -1.320505451  0.01291822 3.803457e-03
## 102 -0.747738949  0.02799597 2.683961e-03
## 103 -1.322313012  0.02032851 6.047035e-03
## 104  0.086005997  0.05491688 7.163777e-05
## 105  1.717141711  0.03602656 1.836617e-02
## 106 -0.435758636  0.03768089 1.239204e-03
## 107 -1.054901242  0.03938516 7.604238e-03
## 108 -1.329014408  0.02882260 8.736605e-03
## 109  0.873402678  0.03823659 5.054623e-03
## 110 -1.606411619  0.02414372 1.064096e-02
## 111 -0.950593651  0.11272874 1.913449e-02
## 112 -1.509681622  0.03491264 1.374155e-02
## 113 -0.752250648  0.01682746 1.614220e-03
## 114 -0.404832167  0.03165632 8.929551e-04
## 115 -0.646015962  0.02095554 1.488784e-03
## 116 -1.563300065  0.03409265 1.437668e-02
## 117 -1.067240469  0.01462940 2.818384e-03
## 118 -0.382432833  0.02629562 6.582870e-04
## 119 -0.216665331  0.04673701 3.835976e-04
## 120  0.716975209  0.01633554 1.422799e-03
## 121  0.109263335  0.02567289 5.242851e-05
## 122 -0.261134964  0.02857971 3.343716e-04
## 123 -0.180060062  0.01656703 9.102975e-05
## 124 -0.666789002  0.01462940 1.100151e-03
## 125 -0.617542666  0.01601273 1.034328e-03
## 126  0.538054905  0.01372961 6.716829e-04
## 127  1.049817241  0.01372961 2.557046e-03
## 128 -0.246899315  0.02240017 2.327977e-04
## 129  0.371562465  0.01149973 2.676846e-04
## 130 -1.225332108  0.03498715 9.072602e-03
```

```
## 131 -0.339859058  0.02177451 4.285050e-04
## 132 -0.704811374  0.03226158 2.760084e-03
## 133 -0.409732655  0.05873201 1.745868e-03
## 134  0.217909996  0.01678705 1.351232e-04
## 135  0.159976824  0.04266515 1.900957e-04
## 136  0.580928853  0.05391514 3.205349e-03
## 137 -1.686449078  0.03406781 1.671833e-02
## 138 -0.472750021  0.01389256 5.247711e-04
## 139 -1.597867868  0.03377172 1.487318e-02
## 140 -1.976732779  0.01462940 9.668775e-03
## 141 -0.260812402  0.01070815 1.227143e-04
## 142  0.050745903  0.03406781 1.513730e-05
## 143 -0.433151207  0.03675132 1.193060e-03
## 144 -0.595328053  0.03921131 2.410710e-03
## 145 -0.057931159  0.03938516 2.293277e-05
## 146  0.595257601  0.03491264 2.136362e-03
## 147  0.648096215  0.02177451 1.558250e-03
## 148 -0.327182113  0.02949537 5.422306e-04
## 149  1.104793059  0.01601273 3.310445e-03
## 150 -0.956709234  0.02423281 3.788491e-03
## 151  1.104793059  0.01601273 3.310445e-03
## 152  0.033019178  0.01959941 3.632626e-06
## 153 -1.054901242  0.03938516 7.604238e-03
## 154  0.118419815  0.06933529 1.741240e-04
## 155 -1.143847996  0.04868067 1.115875e-02
## 156 -0.133449298  0.01076853 3.231021e-05
## 157 -1.396379336  0.01682746 5.562171e-03
## 158 -1.768598487  0.01627342 8.624061e-03
## 159  0.961686300  0.02196894 3.462358e-03
## 160  0.239740832  0.03011177 2.974044e-04
## 161  0.434111876  0.05045649 1.668992e-03
## 162 -0.935647962  0.02538599 3.800451e-03
## 163  0.416003749  0.01881003 5.529421e-04
## 164 -1.142630721  0.05031542 1.152875e-02
## 165  0.513853780  0.03498715 1.595524e-03
## 166  0.895114334  0.02775231 3.811782e-03
## 167  1.498338301  0.02195624 8.399785e-03
## 168 -0.781850699  0.02484088 2.595302e-03
## 169  0.221389374  0.01845878 1.536232e-04
## 170 -0.653435875  0.03244336 2.386185e-03
## 171  0.709377535  0.01563242 1.331902e-03
## 172  0.620087859  0.01845878 1.205174e-03
## 173 -0.179500732  0.03207642 1.779615e-04
## 174  0.384327747  0.01870062 4.691446e-04
## 175  1.433213998  0.01223334 4.239958e-03
## 176  0.100482750  0.04086241 7.169269e-05
## 177  0.512291133  0.03289119 1.487602e-03
## 178  1.497186486  0.03492246 1.351896e-02
## 179  1.006723192  0.02654324 4.605811e-03
## 180 -1.070874737  0.05236205 1.056089e-02
## 181  1.725478400  0.02496208 1.270361e-02
## 182 -0.793707803  0.01288671 1.370708e-03
## 183  1.426063497  0.02244893 7.783639e-03
## 184  0.001207786  0.01487961 3.672236e-09
```

```
## 185 -1.001111591  0.03171243 5.470650e-03
## 186  0.886174860  0.03124980 4.222046e-03
## 187  1.433176149  0.01773981 6.182589e-03
## 188  1.113267764  0.04726185 1.024673e-02
## 189  0.487157010  0.03675132 1.509111e-03
## 190 -0.137258149  0.04707063 1.551009e-04
## 191  0.593902480  0.08020460 5.126100e-03
## 192  1.856009941  0.03011177 1.782477e-02
## 193  1.923618525  0.01633554 1.024173e-02
## 194  2.171799717  0.02196894 1.765813e-02
## 195  1.593154156  0.02095554 9.054422e-03
## 196 -0.553032543  0.02947359 1.548017e-03
## 197  1.638773448  0.05553698 2.631988e-02
## 198  2.505484663  0.02167006 2.317433e-02
## 199 -1.365262638  0.06386025 2.119196e-02
## 200  0.488299265  0.02496208 1.017373e-03
## 201 -0.719651689  0.06570959 6.070722e-03
## 202  0.358509071  0.02095554 4.585064e-04
## 203 -1.982379516  0.01591616 1.059323e-02
```

```
#checked the assumptions
max(abs(m3_stdres))
```

```
## [1] 3.328761
```

```
max(leverage_m3)
```

```
## [1] 0.1127287
```

```
max(CooksD_m3)
```

```
## [1] 0.05925076
```

The highest residual is observation 15 with 3.328761. It is a 56 Male from the control group

The highest leverage is observation 111 with 0.1127287. It is a 56 Female from the experimental group.

The highest CooksD is observation 35 with 0.05925076. It is a 55 Male from the control group.

```
#got the probability of highest residual statistic
pt(3.328761, df = 197,lower.tail=FALSE )
```

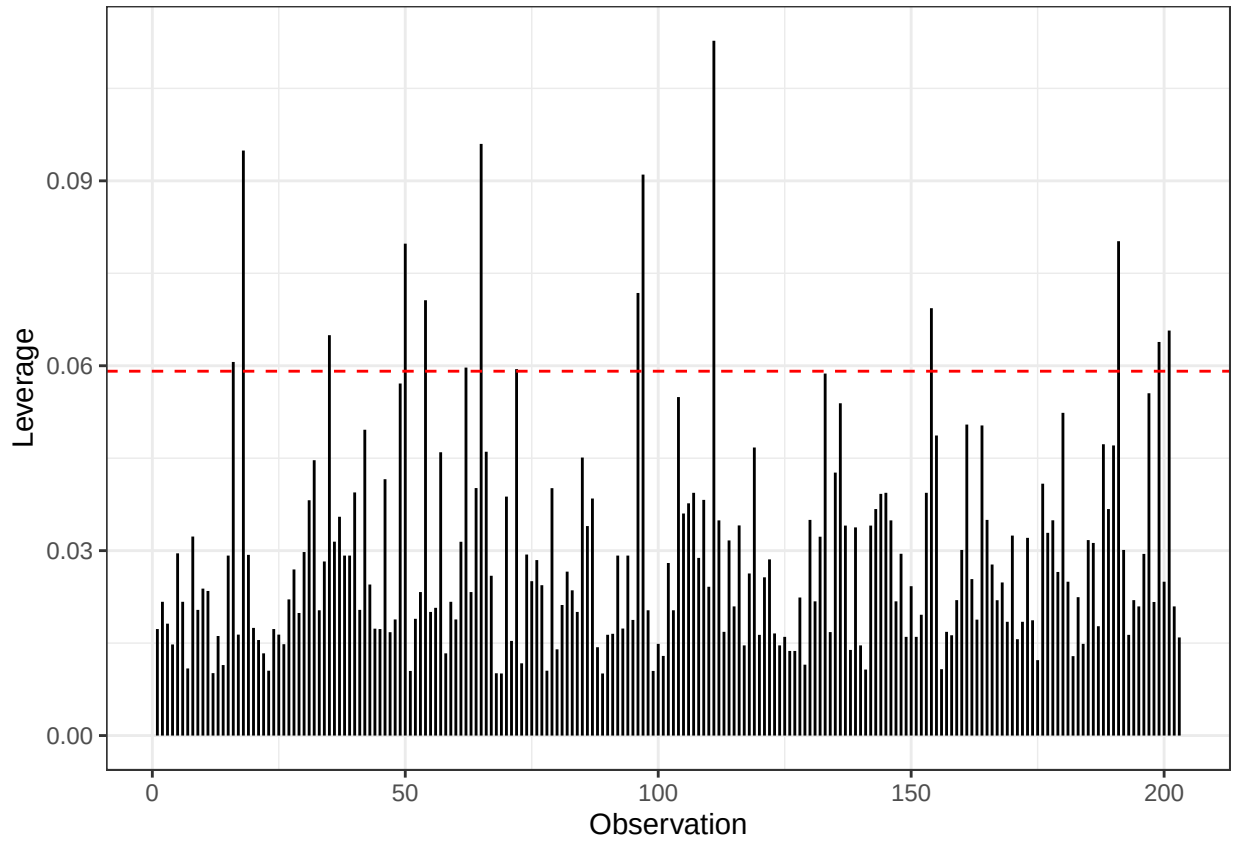
```
## [1] 0.0005205421
```

The probability of highest residual statistic is 0.0005205421, which means that observation 15 is an outlier because this value is significant.

```
#prepared computations for the plots
n <- nrow(log_PEP4)
index<-as_tibble(c(1:n))
colnames(index)<-c('Observation')
```

```
lev <- as_tibble(hatvalues(m3))
colnames(lev) <- c('Leverage')
dfbetas <- as_tibble(dfbetas(m3))
colnames(dfbetas) <- c('DFBeta0', 'DFBeta1', 'DFBeta2','DFBeta3','DFBeta4','DFBeta5')
cooksD <- as_tibble(cooks.distance(m3))
colnames(cooksD)<-c('CooksD')
diagnostics<-tibble(index,lev,dfbetas,cooksD)
```

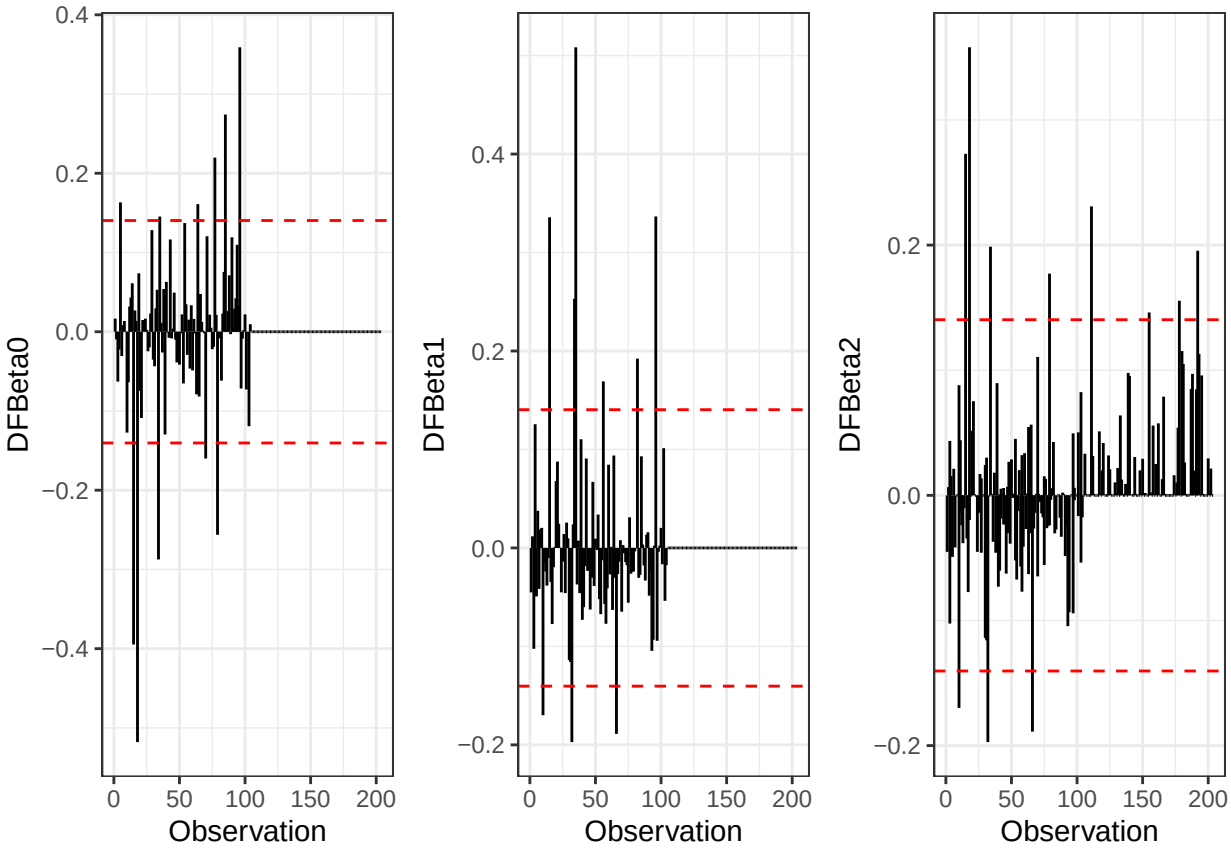
```
#plotted to see which observations have high influence
p_plus_1<-ncol(dfbetas)
thresh_leverage<-2*p_plus_1/n#applied rule of thumb 2
p1<-ggplot(diagnostics, aes(x = Observation, ymax=pmax(0, Leverage), ymin = pmin(0, Leverage))) +
geom_linerange()+theme_bw() + geom_hline(yintercept=thresh_leverage, col="red", linetype="dashed") +
labs(y="Leverage")
p1
```



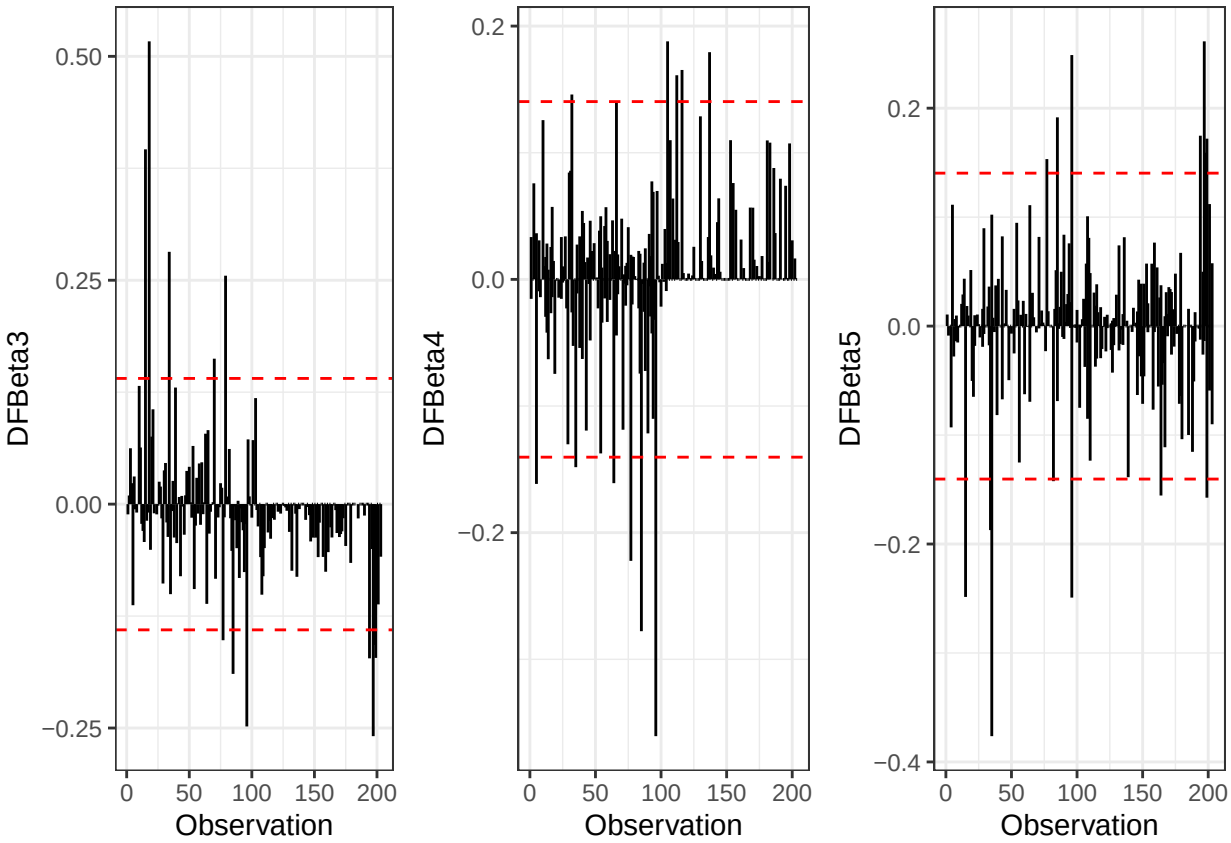
From the plot above, it seems that Observation 111 with leverage of 0.1127287 is highly influential along with 12 other observations that also have relatively very high leverages.

```
#plotted DFBETAs
thresh <- 2/sqrt(n) #applied rule of thumb
p2<-ggplot(diagnostics, aes(x = Observation, ymax=pmax(0, DFBeta0), ymin = pmin(0, DFBeta0))) +
geom_linerange()+theme_bw() + geom_hline(yintercept=thresh, col="red", linetype="dashed") +
geom_hline(yintercept=-thresh,col="red",linetype="dashed") +
labs(y="DFBeta0")
p3<- ggplot(diagnostics, aes(x = Observation, ymax=pmax(0, DFBeta1), ymin = pmin(0, DFBeta1))) +
```

```
geom_linerange()+theme_bw() + geom_hline(yintercept=thresh, col="red", linetype="dashed") +
geom_hline(yintercept=-thresh,col="red",linetype="dashed") +
labs(y="DFBeta1")
p4<- ggplot(diagnostics, aes(x = Observation, ymax=pmax(0, DFBeta2), ymin = pmin(0, DFBeta1))) +
geom_linerange()+theme_bw() + geom_hline(yintercept=thresh, col="red", linetype="dashed") +
geom_hline(yintercept=-thresh,col="red",linetype="dashed") +
labs(y="DFBeta2")
grid.arrange(p2,p3,p4,ncol=3)
```

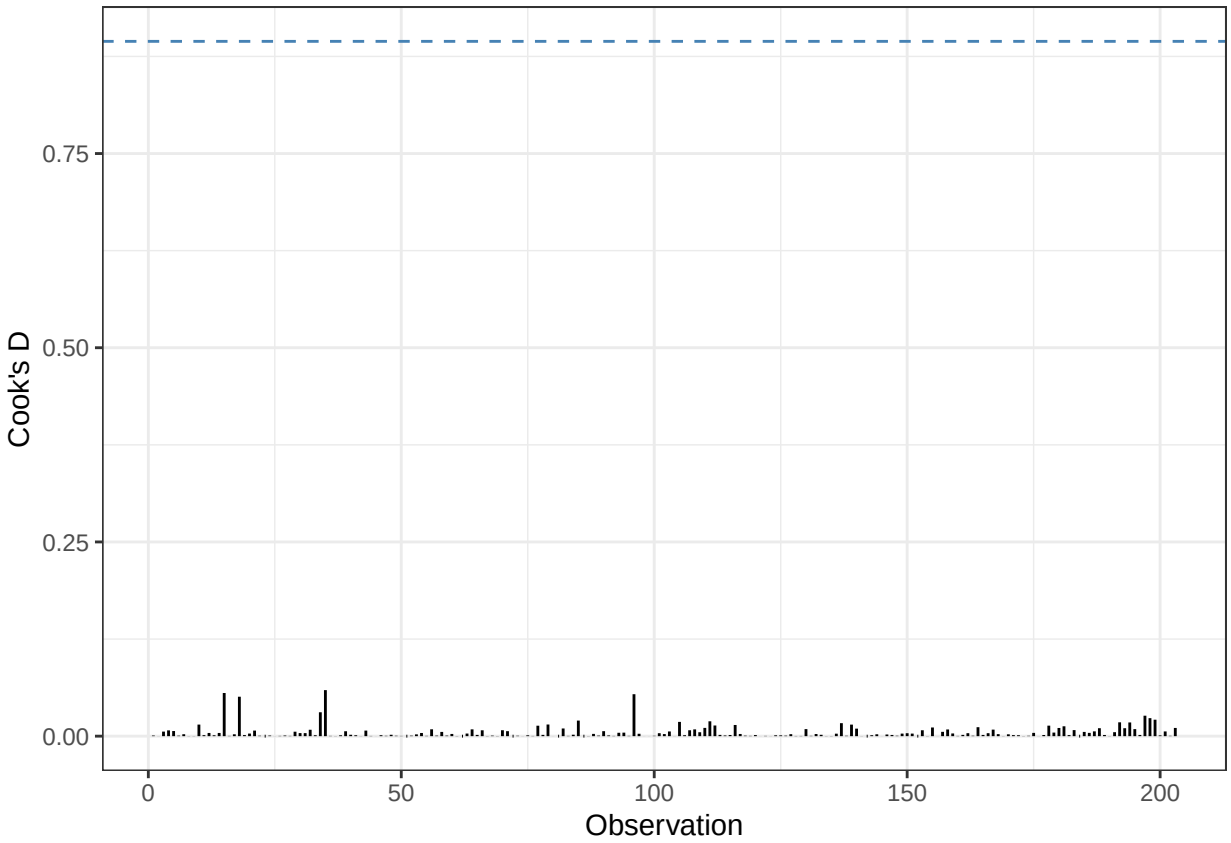



```
#plotted DFBETAs
p5<- ggplot(diagnostics, aes(x = Observation, ymax=pmax(0, DFBeta3), ymin = pmin(0, DFBeta2))) +
geom_linerange()+theme_bw() + geom_hline(yintercept=thresh, col="red", linetype="dashed") +
geom_hline(yintercept=-thresh,col="red",linetype="dashed") +
labs(y="DFBeta3")
p6<- ggplot(diagnostics, aes(x = Observation, ymax=pmax(0, DFBeta4), ymin = pmin(0, DFBeta3))) +
geom_linerange()+theme_bw() + geom_hline(yintercept=thresh, col="red", linetype="dashed") +
geom_hline(yintercept=-thresh,col="red",linetype="dashed") +
labs(y="DFBeta4")
p7<- ggplot(diagnostics, aes(x = Observation, ymax=pmax(0, DFBeta5), ymin = pmin(0, DFBeta4))) +
geom_linerange()+theme_bw() + geom_hline(yintercept=thresh, col="red", linetype="dashed") +
geom_hline(yintercept=-thresh,col="red",linetype="dashed") +
labs(y="DFBeta5")
grid.arrange(p5,p6,p7,ncol=3)
```



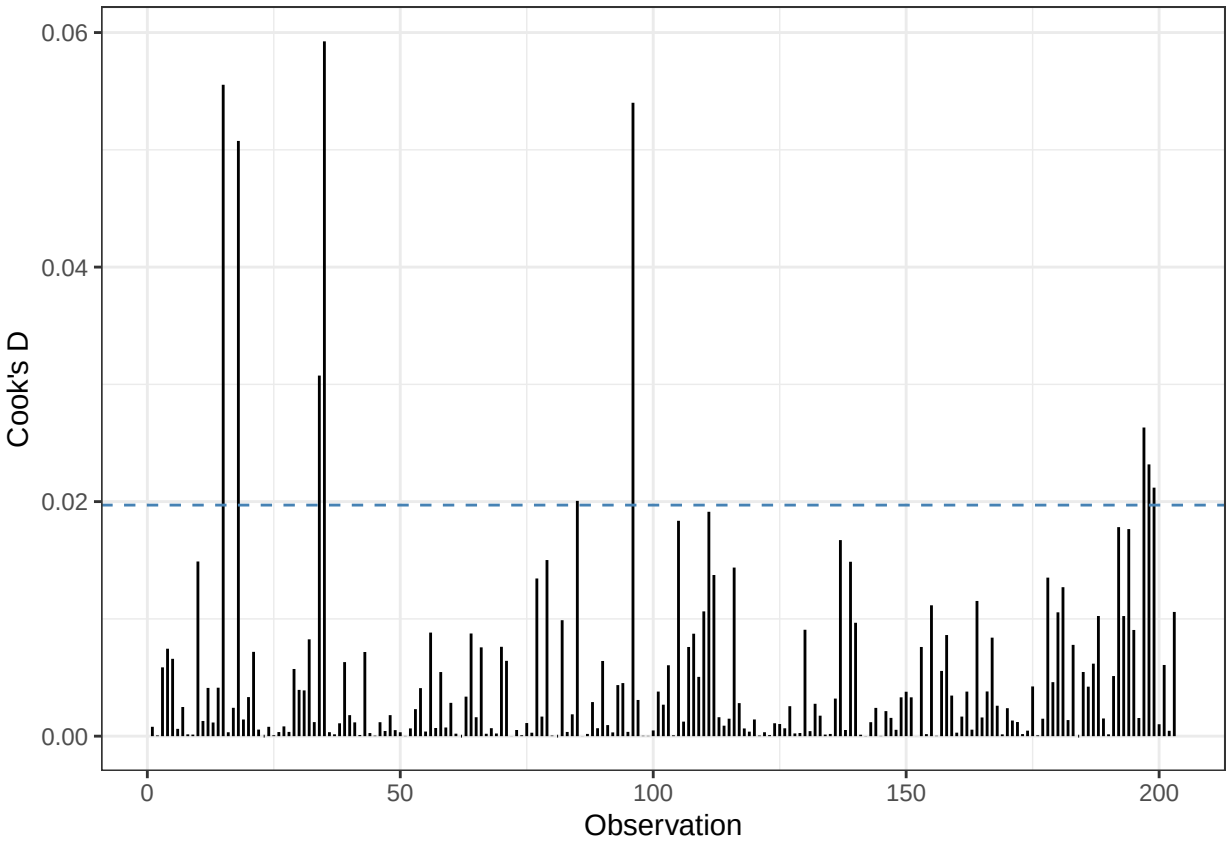
Only few observations influence the calculation of the coefficients. For Beta2, it has relatively more observations than the others, and among those that have influenced Beta2 is observation 111(highest leverage), 35(highest Cooks D) and 15 (highest residual). Observations 35 and 15 have also been influencing Beta1 and Beta3.

```
#plotted CooksD
p8<-ggplot(diagnostics, aes(x = Observation, ymax=pmax(0, CooksD), ymin = pmin(0, CooksD))) +
geom_linerange()+theme_bw() +
geom_hline(yintercept=qf(0.5,p_plus_1,n-p_plus_1), col="steelblue", linetype="dashed") + labs(y="Cook's
p8
```



Cook's D was not able to flag any observation as highly influential. But we can see that there are some observations which are relatively higher than most. Because of this, we will be investigating if the same can be said for the 4/n cutoff.

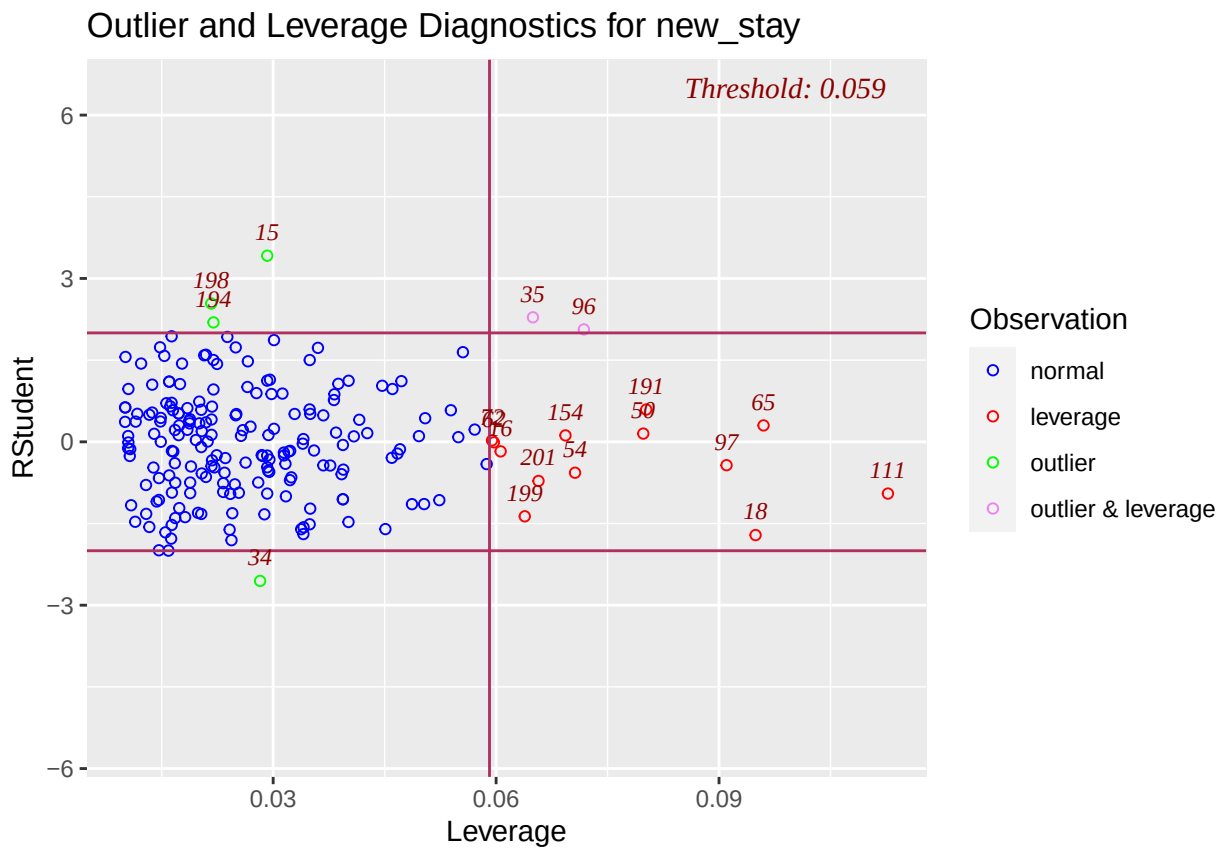
```
p9<-ggplot(diagnostics, aes(x = Observation, ymax=pmax(0, CooksD), ymin = pmin(0, CooksD))) +
geom_linerange()+theme_bw() + geom_hline(yintercept=4/n,col="steelblue", linetype="dashed") +
labs(y="Cook's D")#applied rule of thumb
p9
```



From the graph above, we can now see that observations 15 and 35 have been flagged by Cook's distance as highly influential along with 7 other observations.

With these results, I can say that observations 15,35 and 111 can be considered as outliers simply because 15 and 35 have been influential for the DFBetas, influencing more than one coefficient, and both have been flagged as highly influential by Cook's distance. Moreover, observation 111 has also been able to influential for some coefficients for the DFBetas, not to mention that its leverage value of 0.1127287 is far apart from the other high leverages. For these reasons, based on my analysis observations 15,35 and 111 are outliers.Lastly, to confirm our conclusions, we will be using the olsrr package to have a clearer view of the observations that are considered as outliers.

```
#visualized outlier observations
ols_plot_resid_lev(m3)
```



From the graph above, 15 has been flagged as an outlier, 35 has been flagged as an outlier & high leveraged. While 111 has not been flagged as an outlier, we can observe that it is so far apart from the other predictors; and we can recall that extreme values on the x-scale can severely be influential. With these reasons, we can confirm our earlier analysis and conclude that observations 15, 35 and 111 are outliers.