

# Model estimation

## Regression summary and plots for DV = price

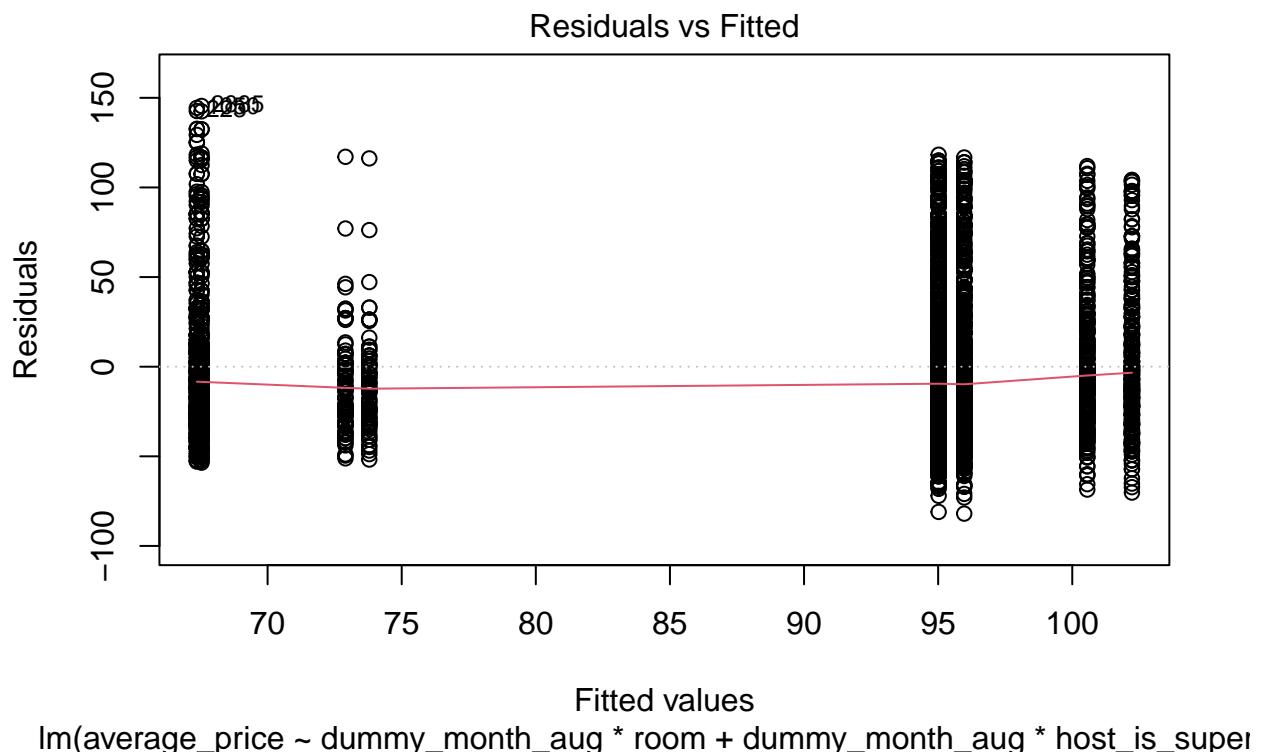
Antwerp

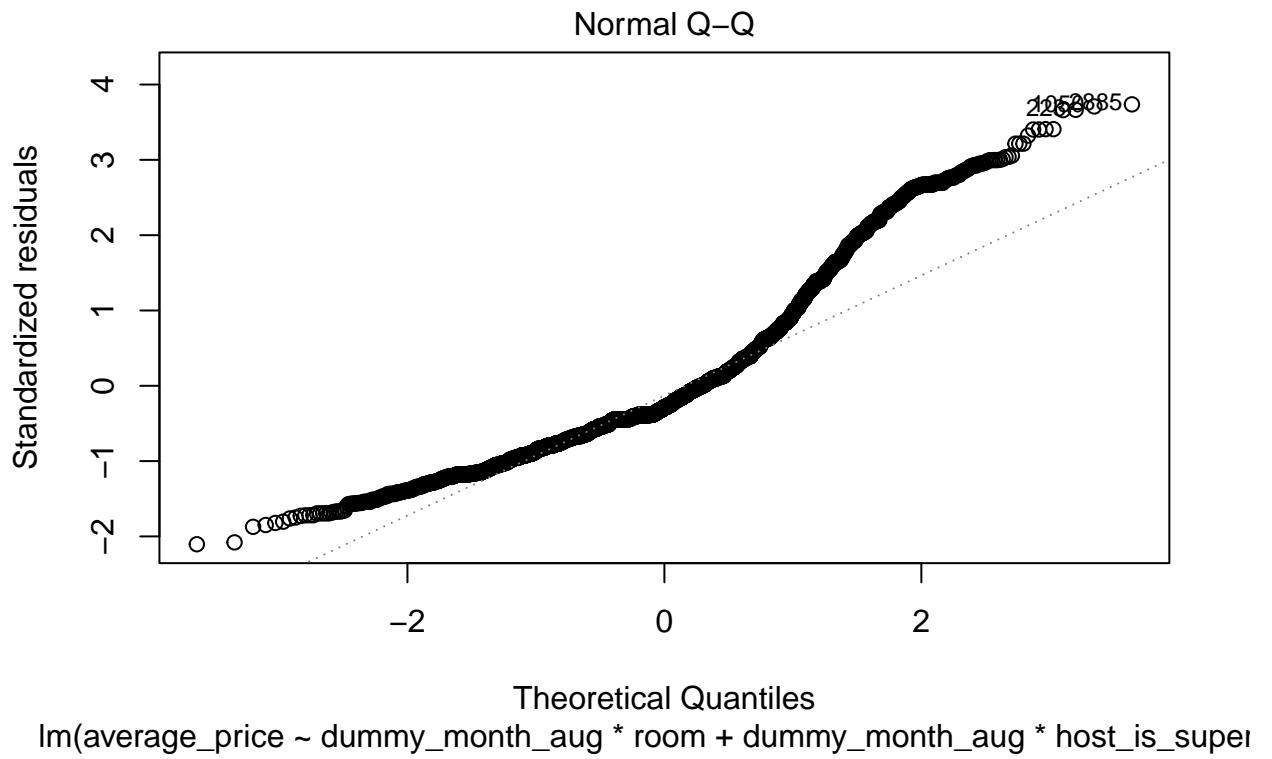
```
##  
## Call:  
## lm(formula = average_price ~ dummy_month_aug * room + dummy_month_aug *  
##      host_is_superhost, data = dataset)  
##  
## Residuals:  
##    Min      1Q Median      3Q     Max  
## -81.97 -25.97 -11.18  15.91 145.46  
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 67.5383   2.0701  32.626 < 2e-16 ***  
## dummy_month_aug -0.1763   2.9146  -0.060  0.95176  
## room          28.4338   2.2571  12.598 < 2e-16 ***  
## host_is_superhostTRUE 6.2533   2.2150   2.823  0.00478 **  
## dummy_month_aug:room -0.7772   3.1773  -0.245  0.80677  
## dummy_month_aug:host_is_superhostTRUE -0.7123   3.1184  -0.228  0.81933  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 38.98 on 3652 degrees of freedom  
## Multiple R-squared:  0.08178,   Adjusted R-squared:  0.08052  
## F-statistic: 65.05 on 5 and 3652 DF,  p-value: < 2.2e-16
```

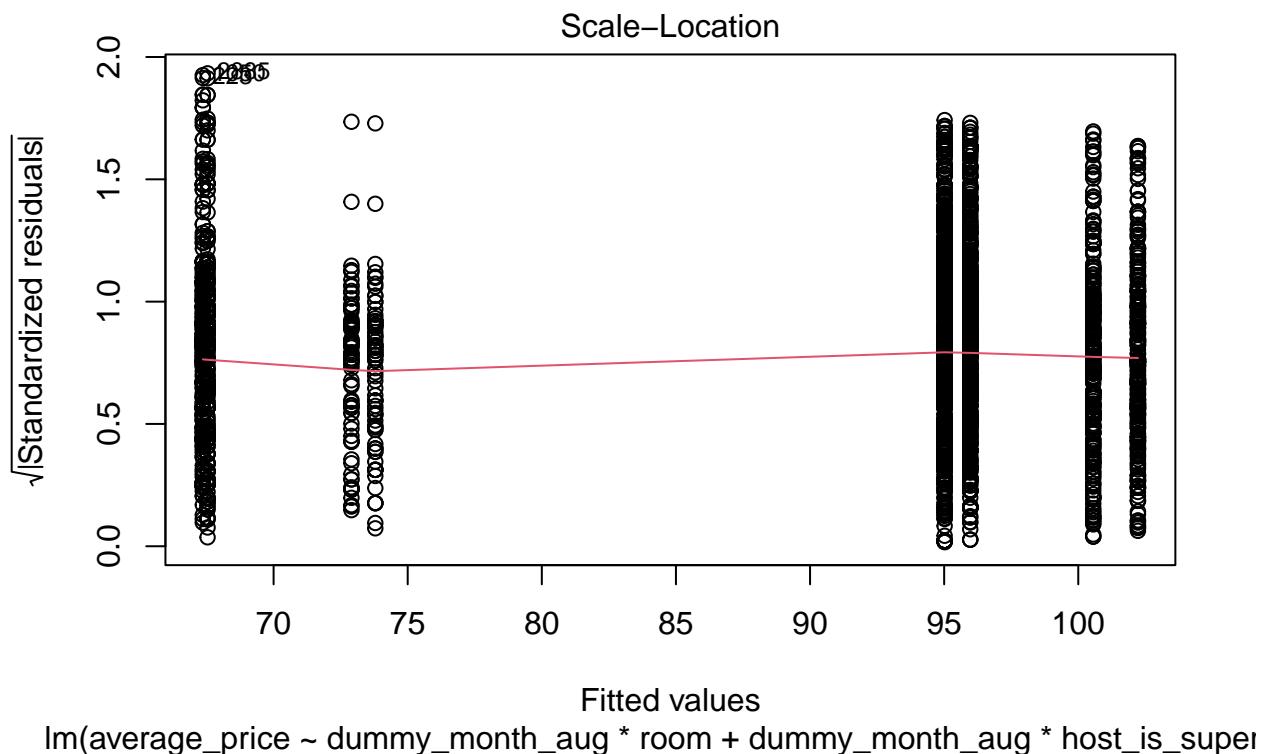
The negative median indicates that the residuals are skewed to the left.

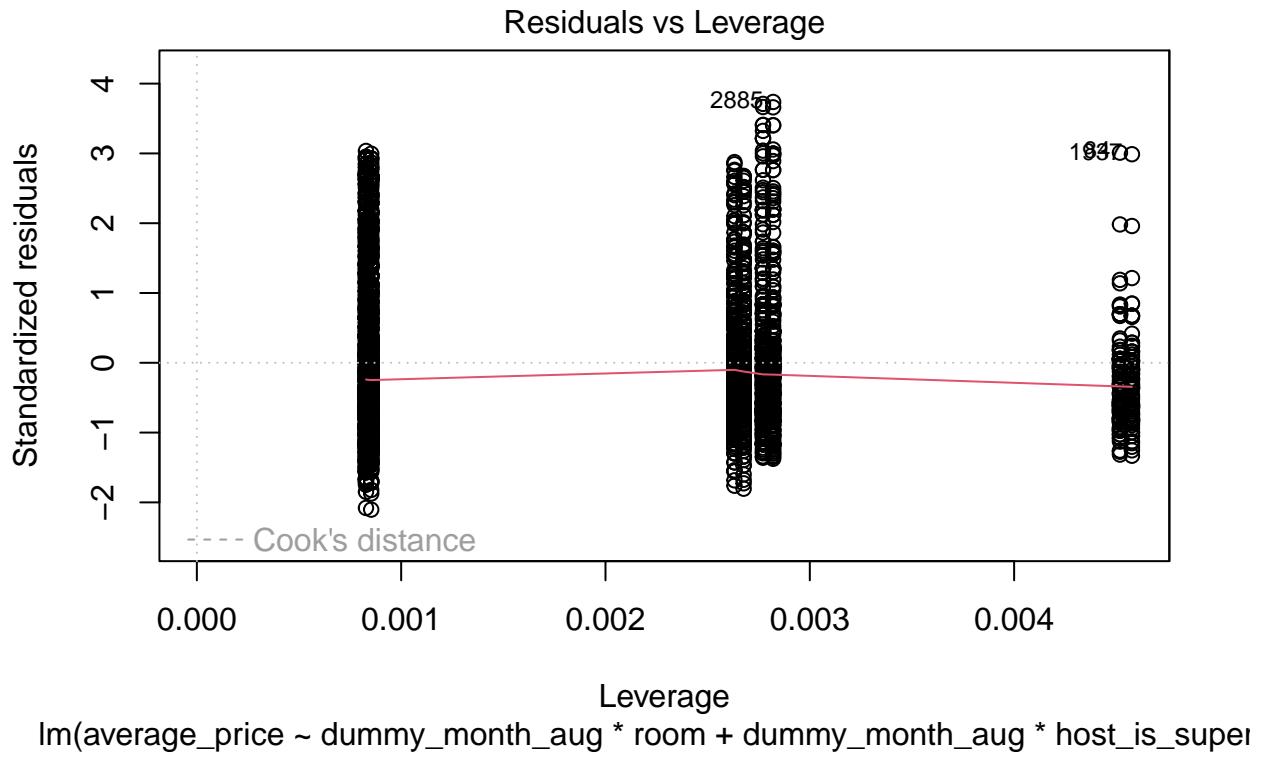
However, since the sample is large, we proceed with the model despite the fact that the normality assumption is violated.

The p-value of the F-statistic is below 0.05, which indicates that the model is significant. However, a very low percentage of the variance in the data is explained by the model (only ~8%).









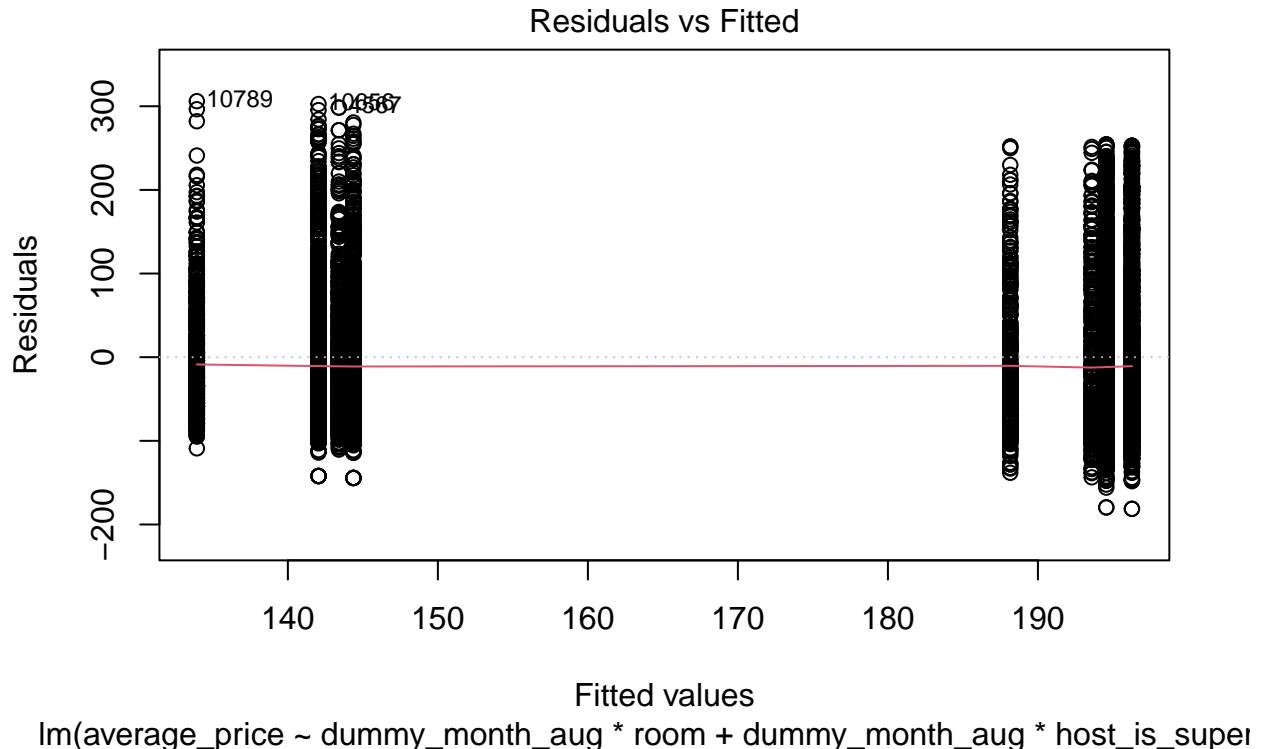
Amsterdam

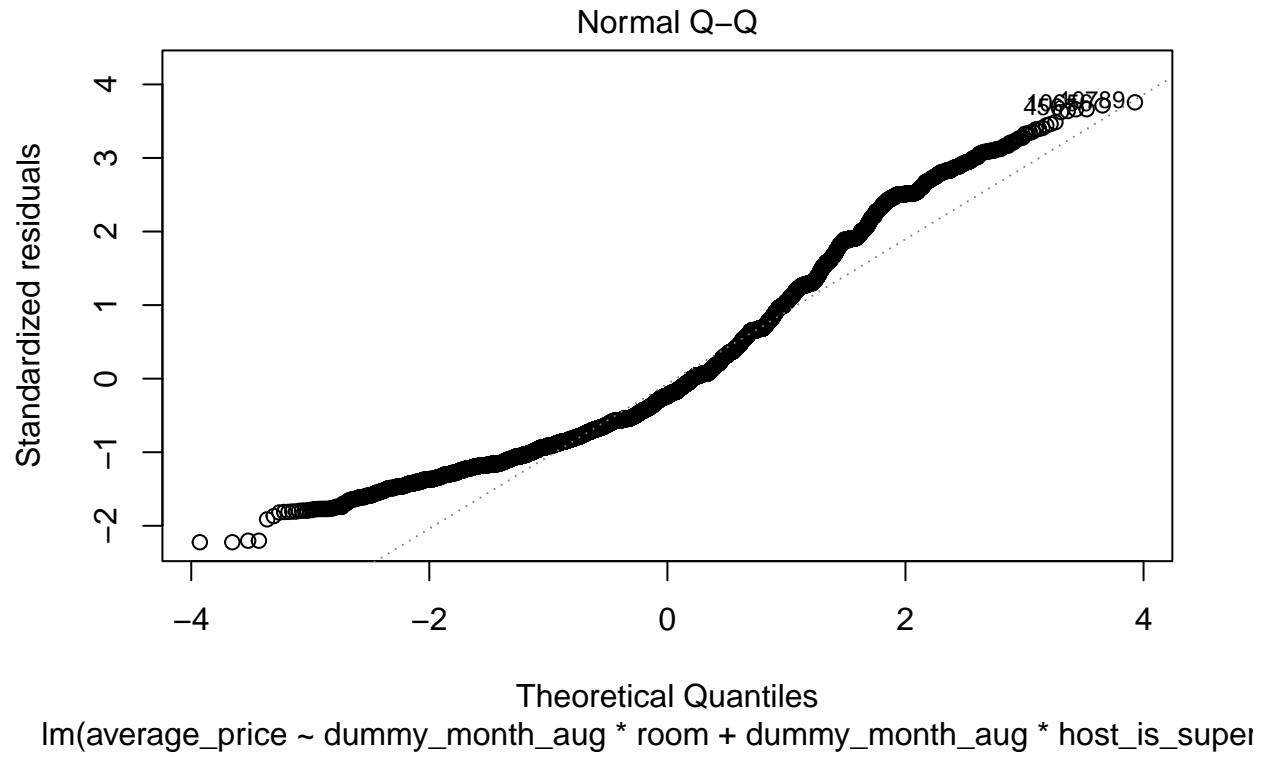
```
##
## Call:
## lm(formula = average_price ~ dummy_month_aug * room + dummy_month_aug *
##     host_is_superhost, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -181.27  -59.88  -19.27   48.26  306.07
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 142.045    2.169   65.478 <2e-16 ***
## dummy_month_aug                2.326    3.043    0.764  0.4447
## room                         54.224    2.426   22.350 <2e-16 ***
## host_is_superhostTRUE        -8.112    2.975   -2.727  0.0064 **
## dummy_month_aug:room          -4.036    3.410   -1.184  0.2366
## dummy_month_aug:host_is_superhostTRUE 7.142     4.190    1.705  0.0883 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 81.53 on 11652 degrees of freedom
## Multiple R-squared:  0.08604,    Adjusted R-squared:  0.08564
## F-statistic: 219.4 on 5 and 11652 DF,  p-value: < 2.2e-16
```

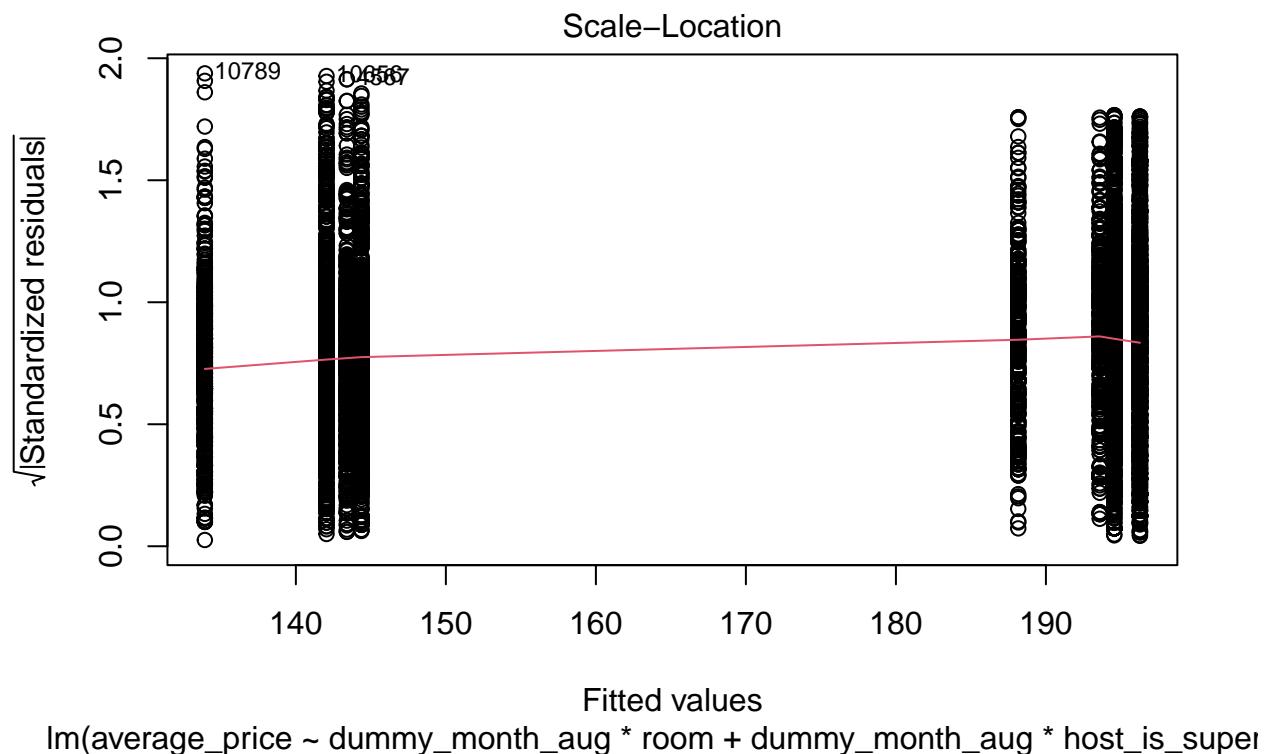
The negative median indicates that the residuals are skewed to the left.

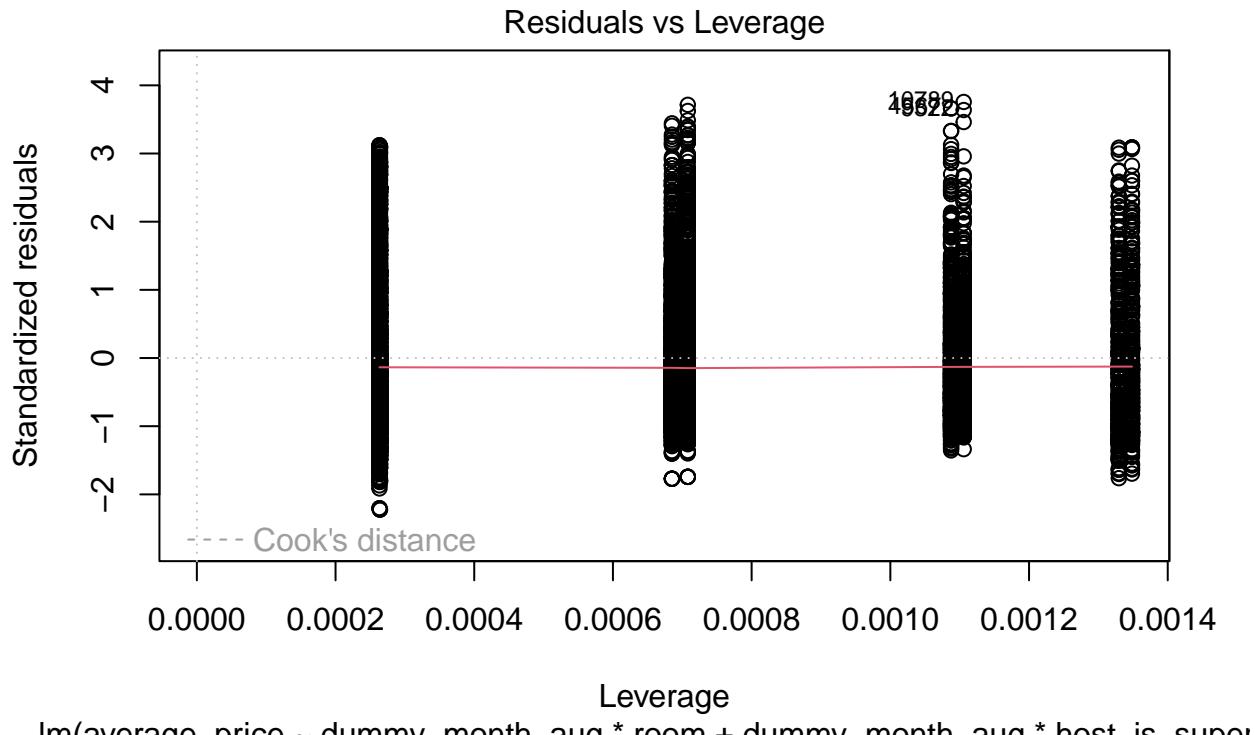
However, since the sample is large, we proceed with the model despite the fact that the normality assumption is violated.

The p-value of the F-statistic is below 0.05, which indicates that the model is significant. However, a very low percentage of the variance in the data is explained by the model (only ~9%).









Berlin

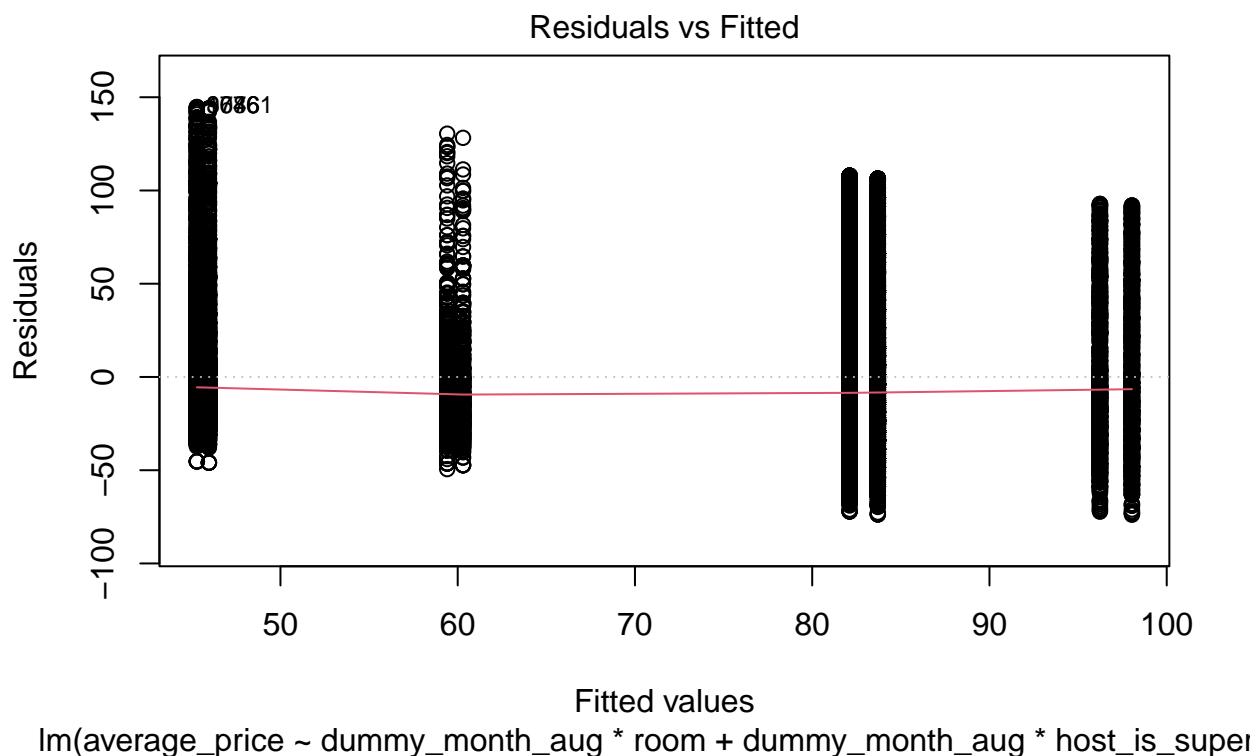
```
##
## Call:
## lm(formula = average_price ~ dummy_month_aug * room + dummy_month_aug *
##      host_is_superhost, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -73.879 -22.108 - 8.703 14.967 144.717 
## 
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                 45.9640    0.4296 107.000 <2e-16 ***
## dummy_month_aug            -0.6808    0.6064  -1.123  0.262    
## room                      37.7391    0.5600  67.396 <2e-16 ***
## host_is_superhostTRUE     14.3359    0.8629  16.613 <2e-16 ***
## dummy_month_aug:room      -0.9140    0.7919  -1.154  0.248    
## dummy_month_aug:host_is_superhostTRUE -0.2109    1.2293  -0.172  0.864  
## ---                        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 34.07 on 30397 degrees of freedom
##   (20 observations deleted due to missingness)
## Multiple R-squared:  0.2438, Adjusted R-squared:  0.2437
```

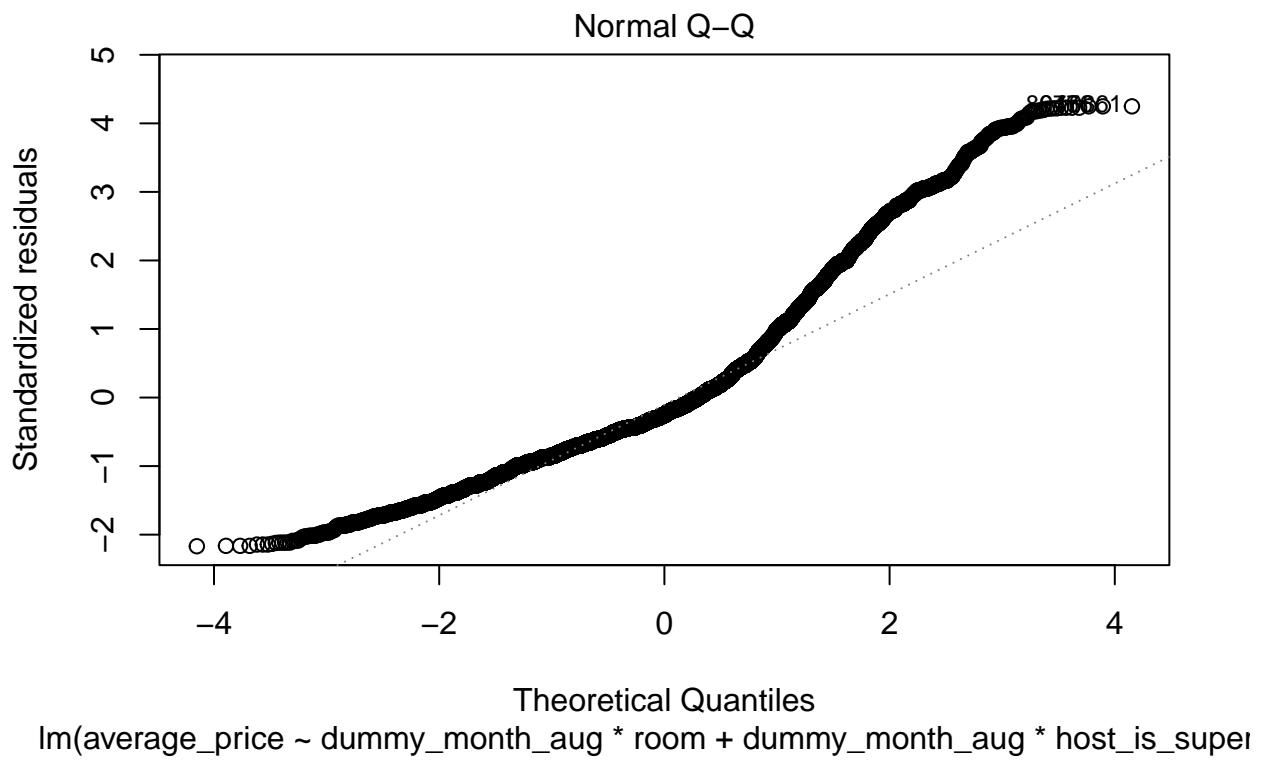
```
## F-statistic: 1960 on 5 and 30397 DF, p-value: < 2.2e-16
```

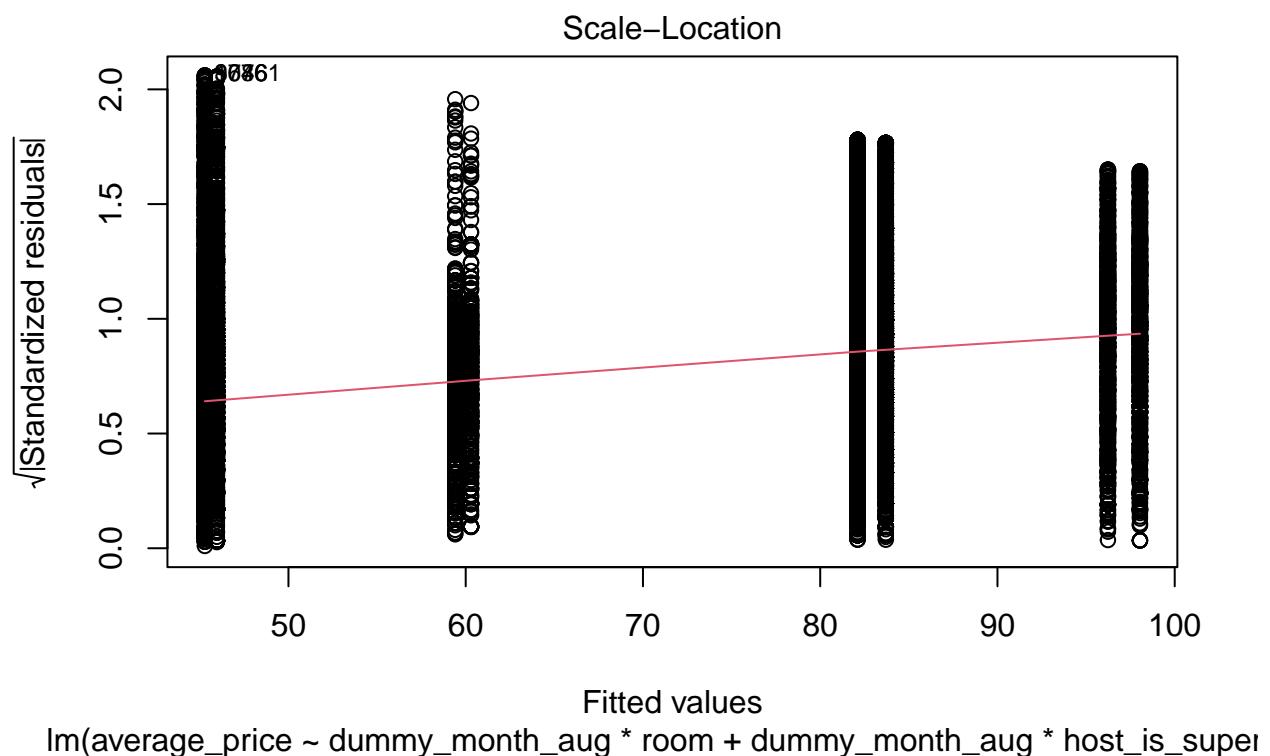
The negative median indicates that the residuals are skewed to the left.

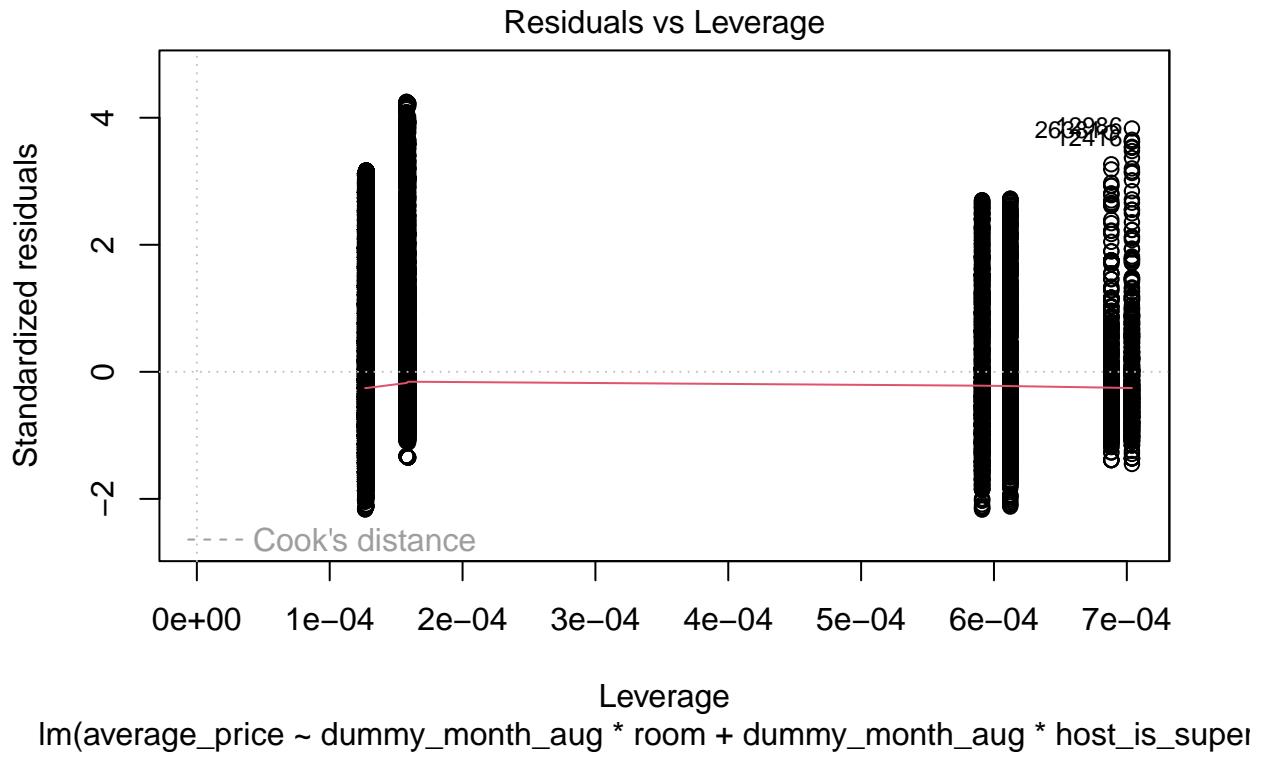
However, since the sample is large, we proceed with the model despite the fact that the normality assumption is violated.

The p-value of the F-statistic is below 0.05, which indicates that the model is significant. 24% of the variance in the DV is explained by the model.









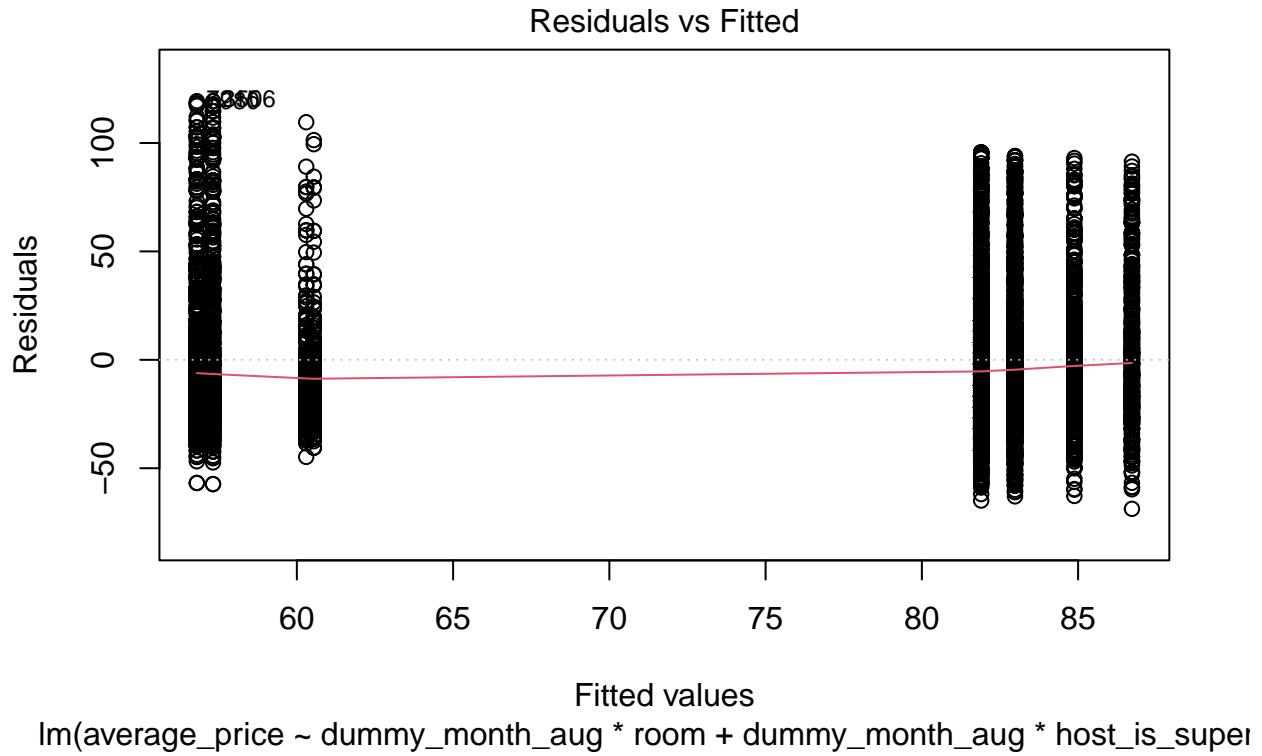
Brussels

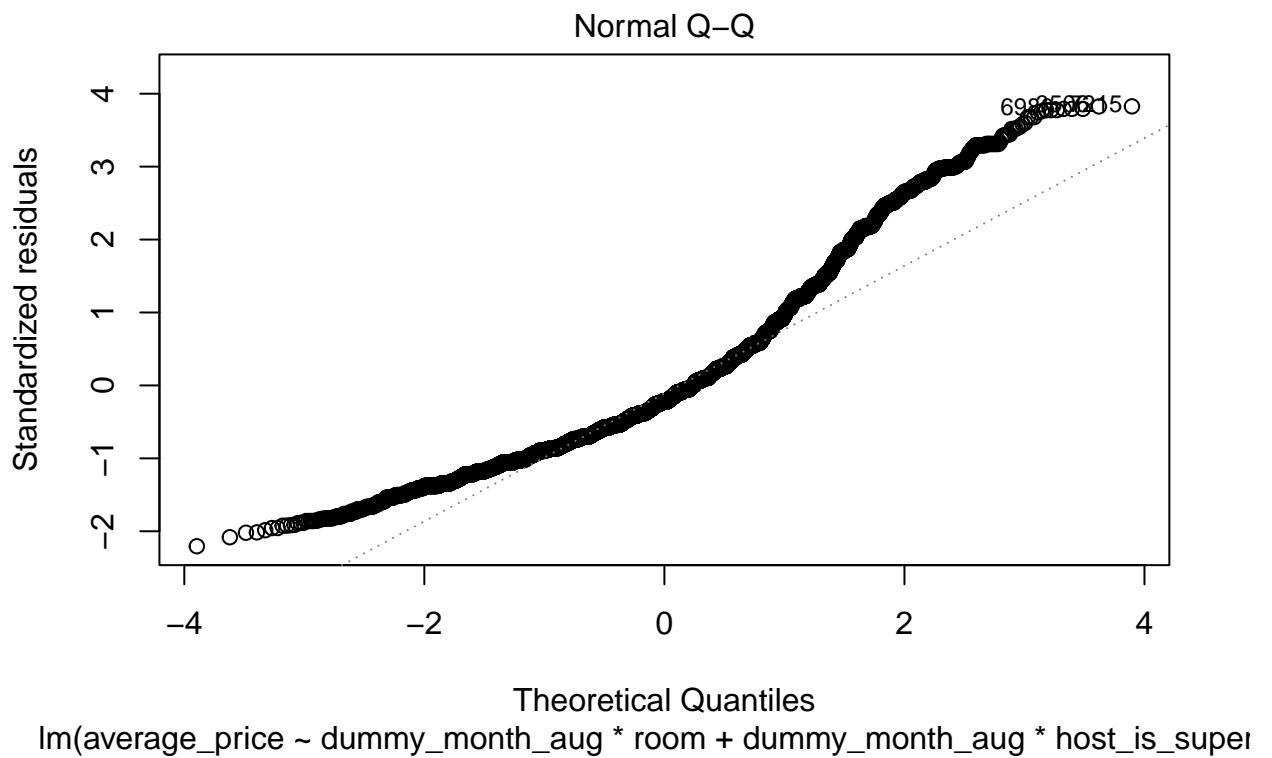
```
##
## Call:
## lm(formula = average_price ~ dummy_month_aug * room + dummy_month_aug *
##     host_is_superhost, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -68.723 -21.898 -6.898  14.926 119.290 
## 
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                 56.7996    0.8252  68.835 < 2e-16 ***
## dummy_month_aug              0.5173    1.1652   0.444  0.65709    
## room                         26.1800    0.9574  27.344 < 2e-16 ***
## host_is_superhostTRUE        3.7432    1.1934   3.137  0.00171 **  
## dummy_month_aug:room         -1.5994    1.3512  -1.184  0.23656    
## dummy_month_aug:host_is_superhostTRUE -0.7607    1.6848  -0.452  0.65161  
## ---                        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 31.18 on 10202 degrees of freedom
## Multiple R-squared:  0.1229, Adjusted R-squared:  0.1224 
## F-statistic: 285.8 on 5 and 10202 DF,  p-value: < 2.2e-16
```

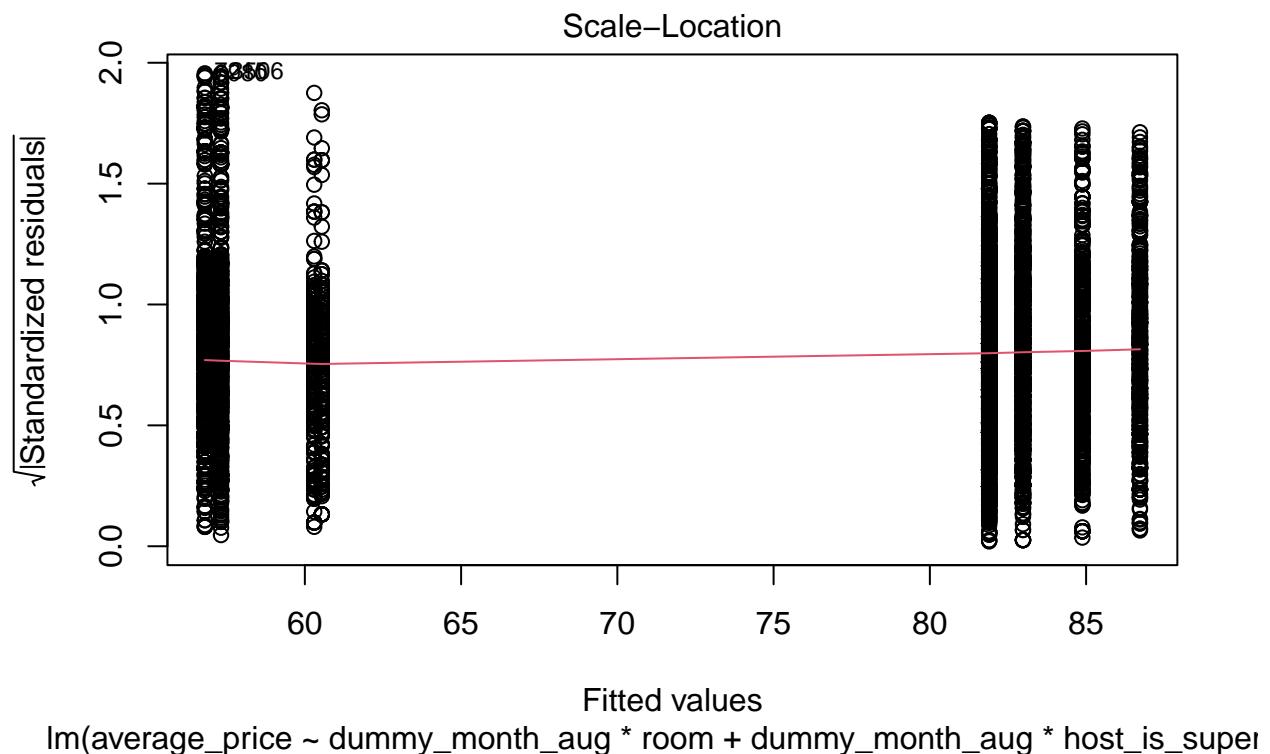
The negative median indicates that the residuals are skewed to the left.

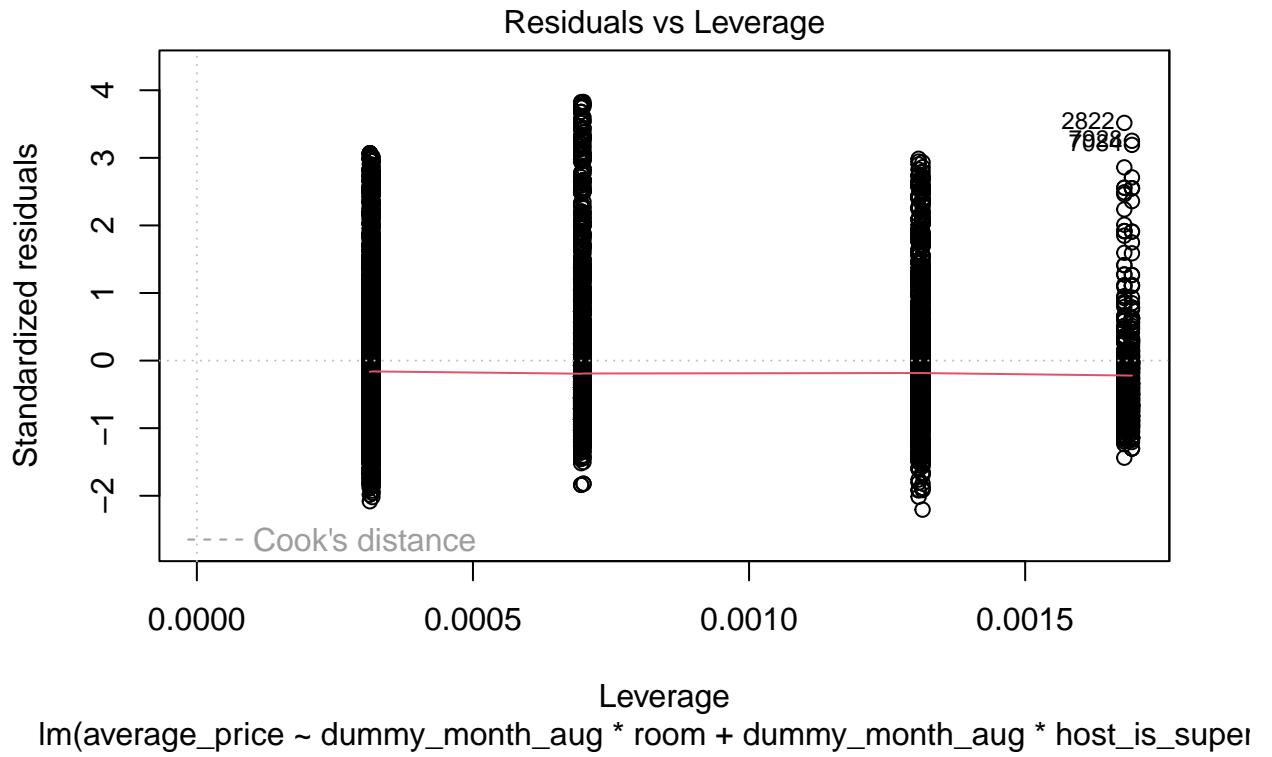
However, since the sample is large, we proceed with the model despite the fact that the normality assumption is violated.

The p-value of the F-statistic is below 0.05, which indicates that the model is significant. However, a very low percentage of the variance in the data is explained by the model (only ~12%).









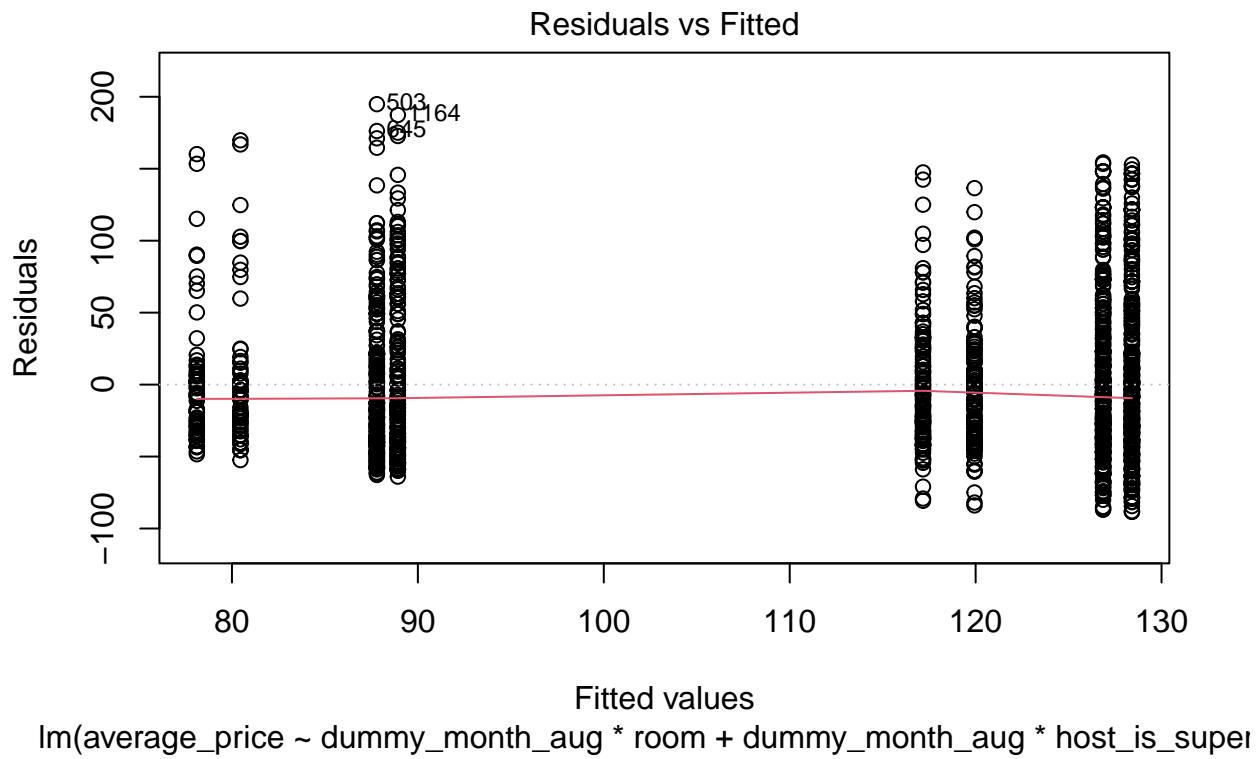
Rotterdam

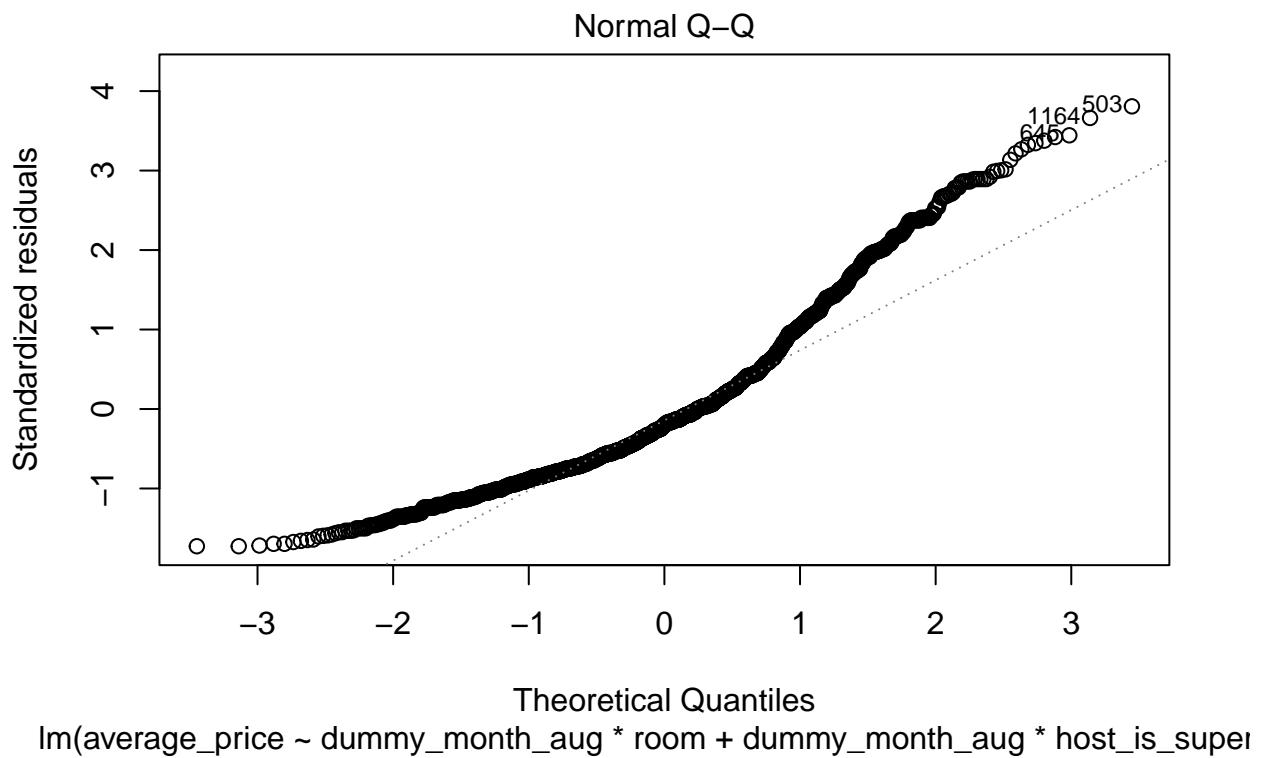
```
##
## Call:
## lm(formula = average_price ~ dummy_month_aug * room + dummy_month_aug *
##     host_is_superhost, data = dataset)
##
## Residuals:
##    Min      1Q Median      3Q     Max
## -88.41 -37.77 -10.26  23.14 194.86
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 88.9234    3.0963  28.719 <2e-16 ***
## dummy_month_aug            -1.1271    4.3143  -0.261  0.7939
## room                      39.4846    3.6391  10.850 <2e-16 ***
## host_is_superhostTRUE     -8.4677    4.1213  -2.055  0.0401 *
## dummy_month_aug:room       -0.4258    5.0993  -0.083  0.9335
## dummy_month_aug:host_is_superhostTRUE -1.2172    5.8183  -0.209  0.8343
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 51.26 on 1760 degrees of freedom
## Multiple R-squared:  0.1238, Adjusted R-squared:  0.1213
## F-statistic: 49.74 on 5 and 1760 DF,  p-value: < 2.2e-16
```

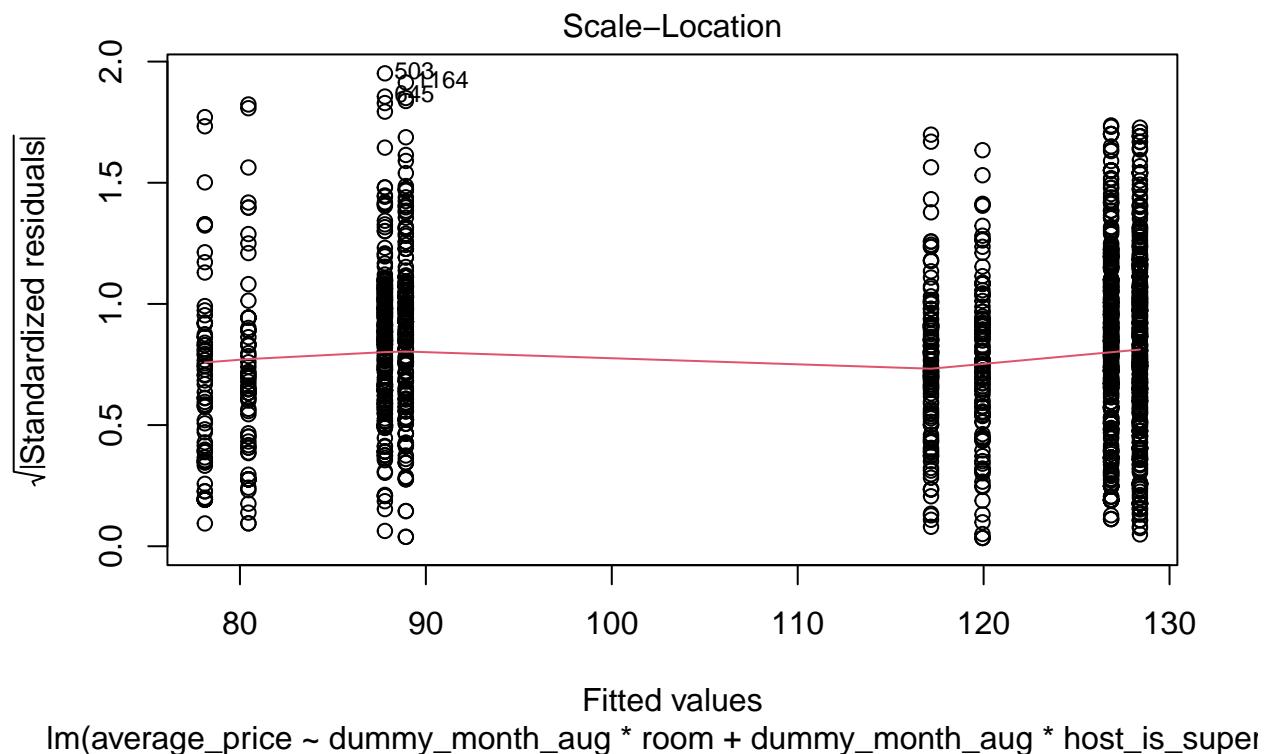
The negative median indicates that the residuals are skewed to the left.

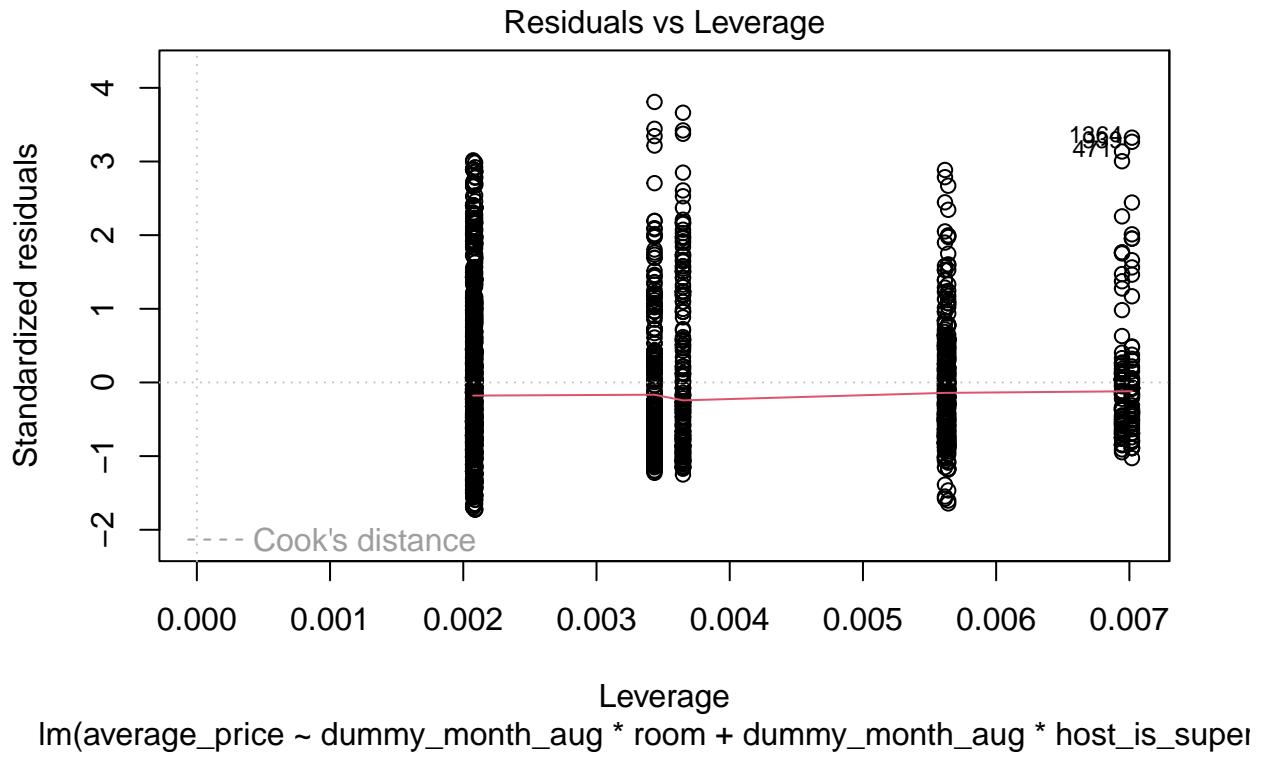
However, since the sample is large, we proceed with the model despite the fact that the normality assumption is violated.

The p-value of the F-statistic is below 0.05, which indicates that the model is significant. However, a very low percentage of the variance in the data is explained by the model (only ~12%).









## Regression summary for DV = average nights

## Antwerp

```

## 
## Call:
## lm(formula = average_nights ~ dummy_month_aug * room + dummy_month_aug *
##      host_is_superhost, data = dataset)
##
## Residuals:
##    Min      1Q  Median      3Q     Max
## -884.8 -510.8   254.4   380.3   385.6
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                758.65767  25.33976  29.939 < 2e-16 ***
## dummy_month_aug             -14.35731   35.67758  -0.402   0.687  
## room                      -19.24248   27.62867  -0.696   0.486  
## host_is_superhostTRUE      131.14186   27.11429   4.837  1.38e-06 ***
## dummy_month_aug:room        19.60416   38.89282   0.504   0.614  
## dummy_month_aug:host_is_superhostTRUE 0.03212   38.17235   0.001   0.999  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 477.1 on 3652 degrees of freedom

```

```

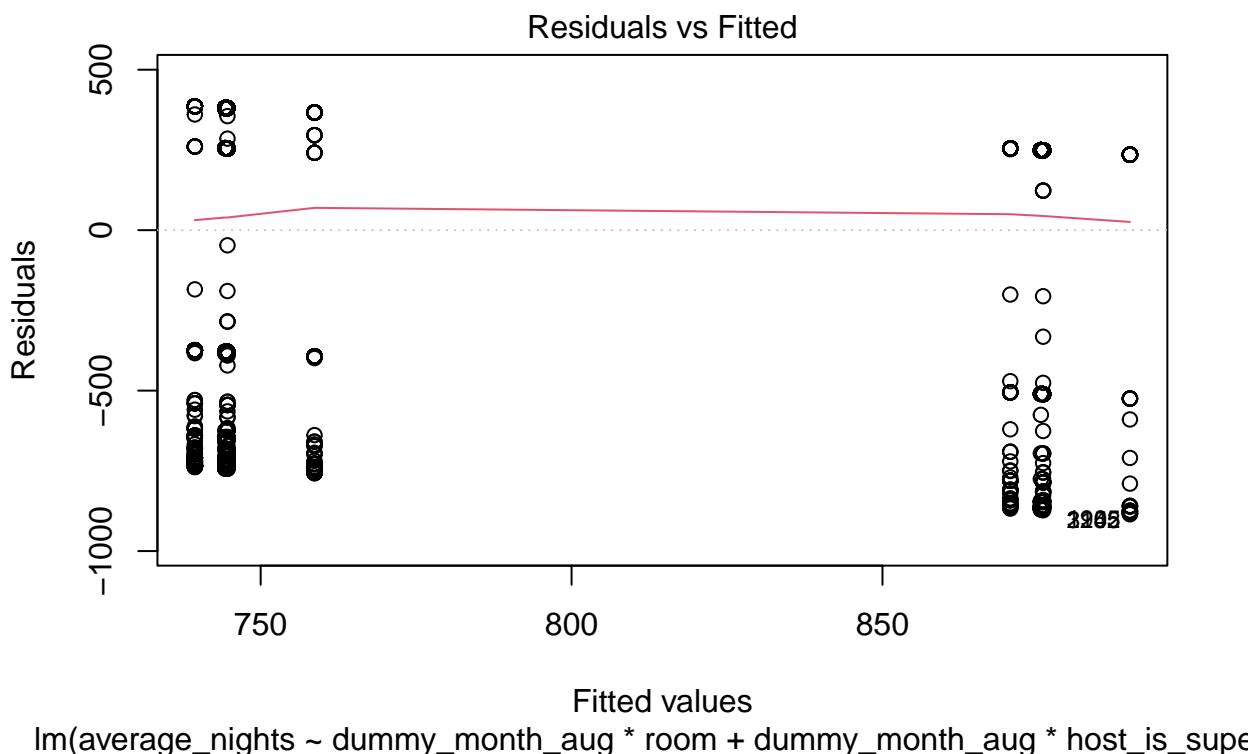
## Multiple R-squared:  0.01292,    Adjusted R-squared:  0.01156
## F-statistic: 9.557 on 5 and 3652 DF,  p-value: 4.501e-09

```

The positive median indicates that the residuals are skewed to the right.

However, since the sample is large, we proceed with the model despite the fact that the normality assumption is violated.

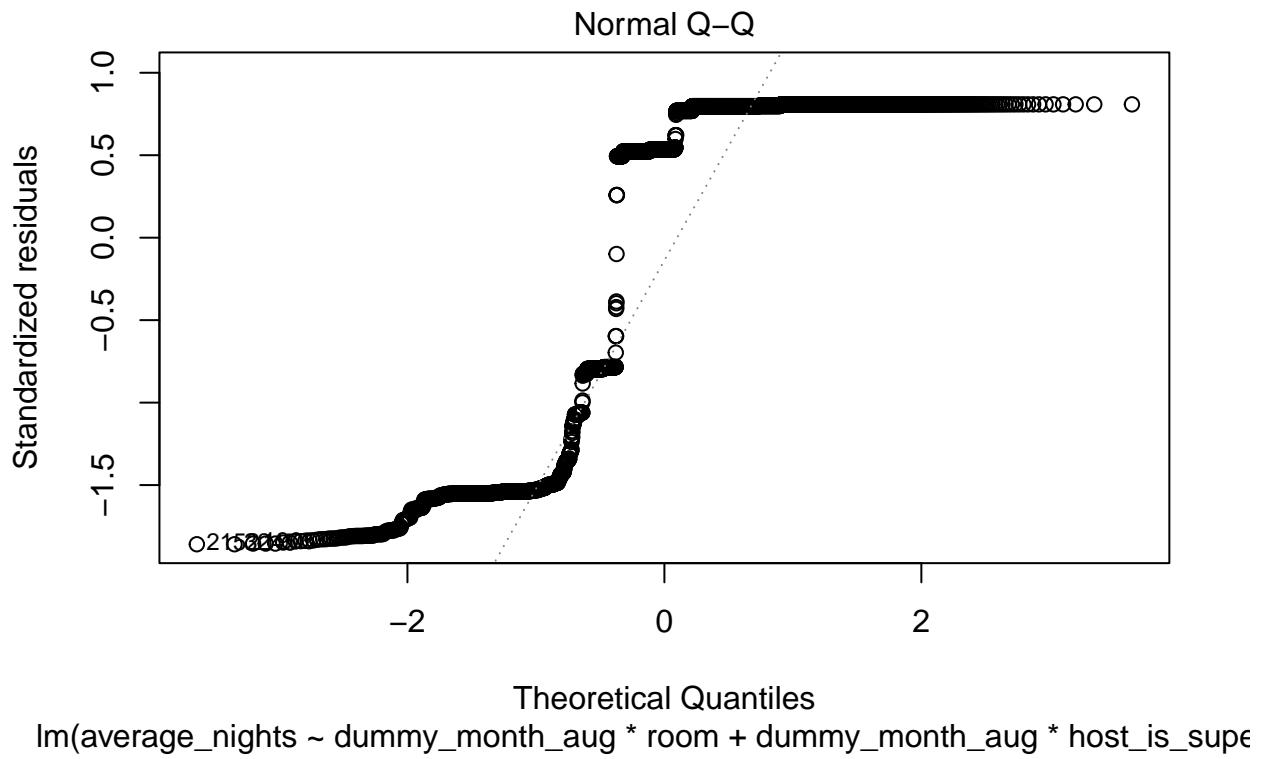
The p-value of the F-statistic is below 0.05, which indicates that the model is significant. However, a very low percentage of the variance in the data is explained by the model (only ~1%).

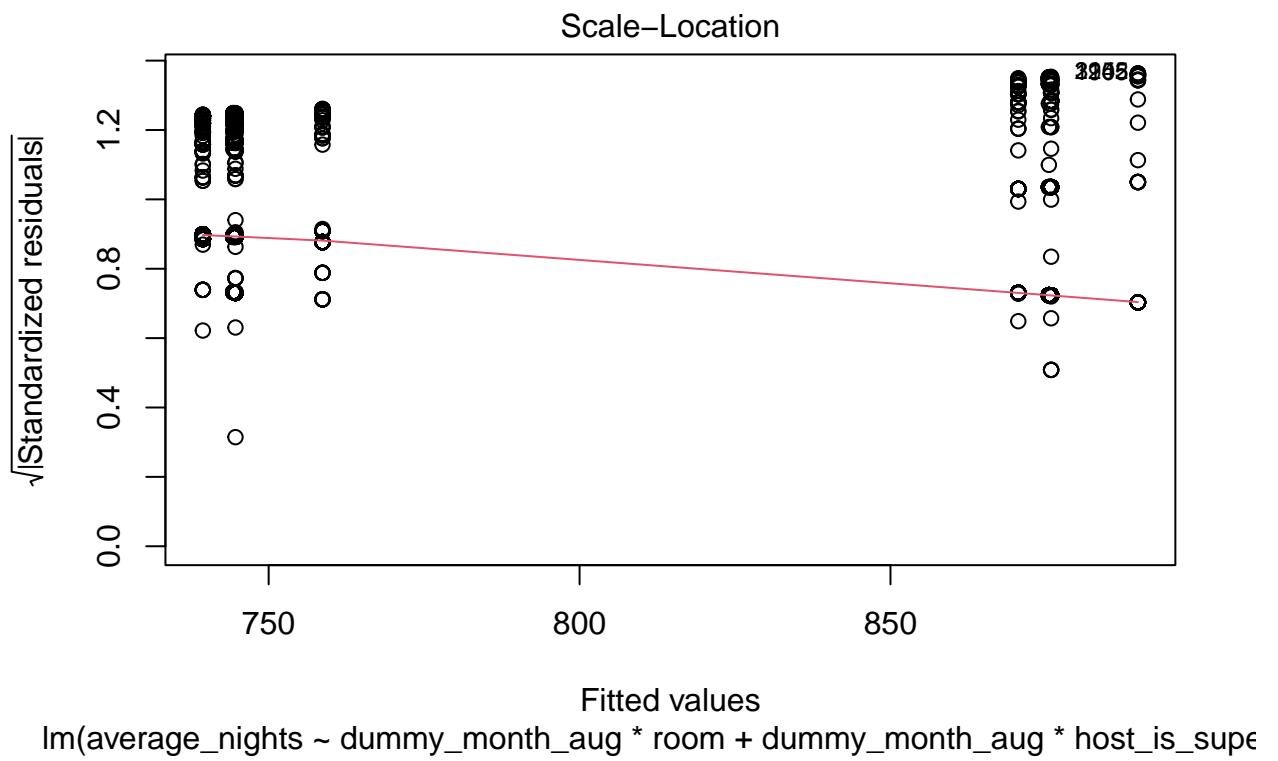


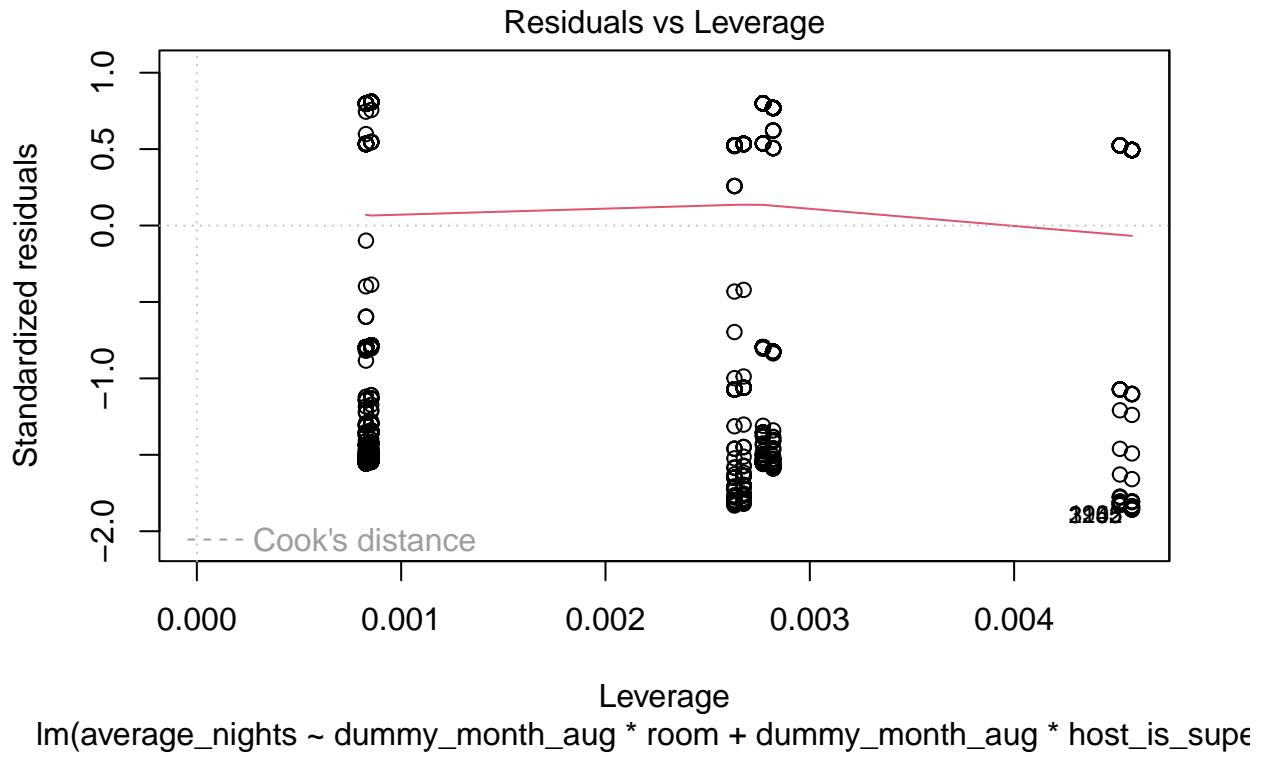
```

lm(average_nights ~ dummy_month_aug * room + dummy_month_aug * host_is_supervisor)

```







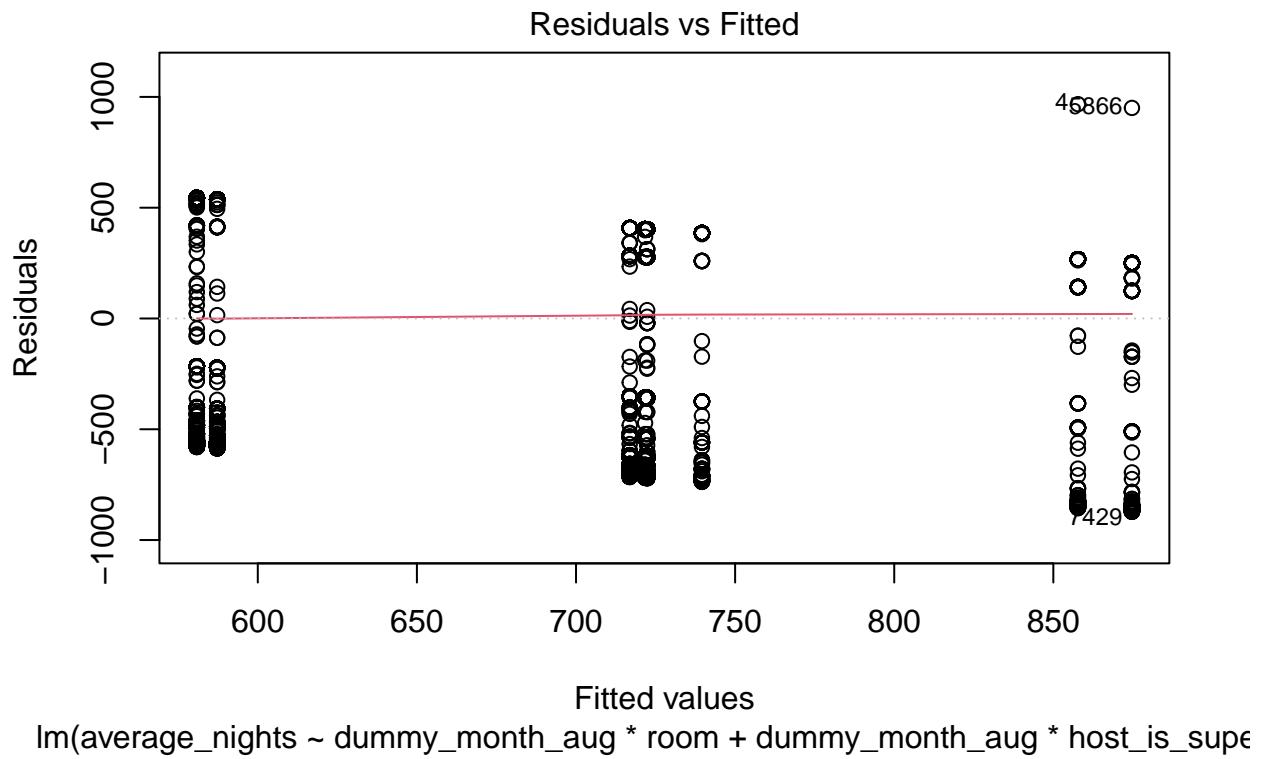
Amsterdam

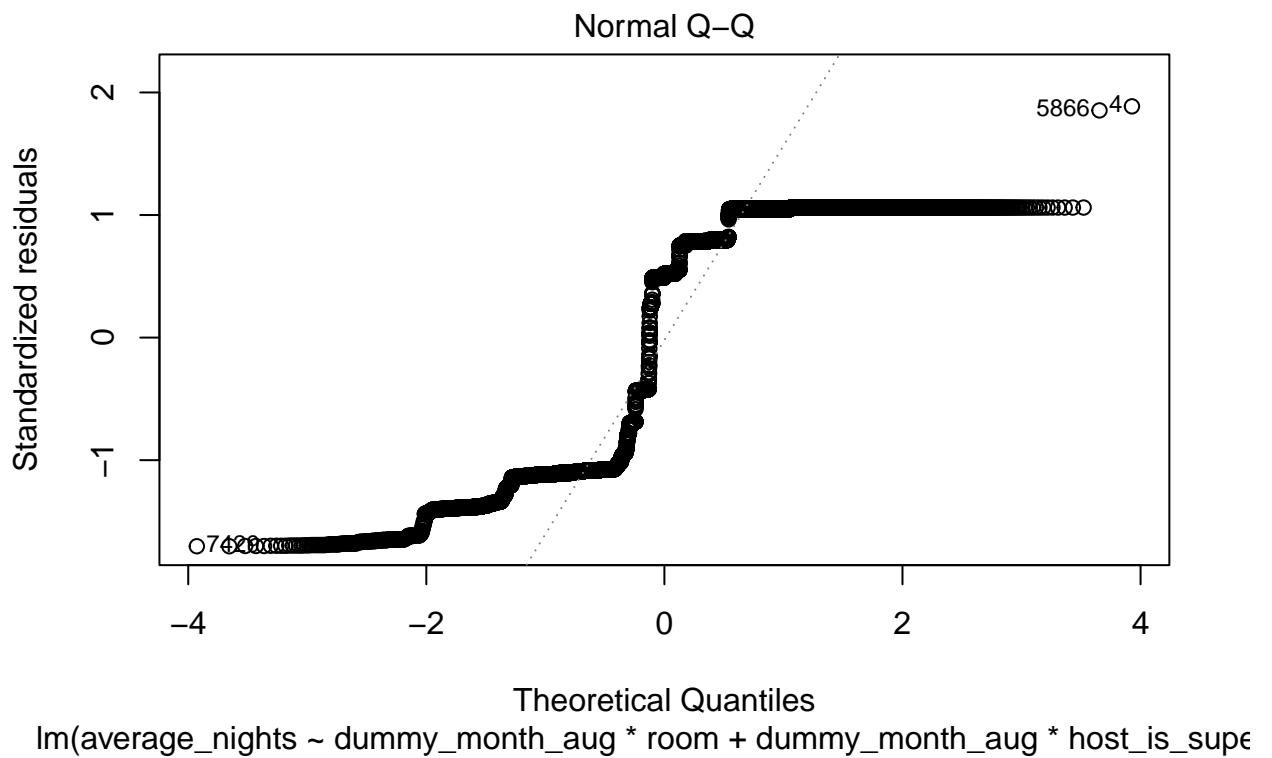
```
##
## Call:
## lm(formula = average_nights ~ dummy_month_aug * room + dummy_month_aug *
##     host_is_superhost, data = dataset)
##
## Residuals:
##    Min      1Q Median      3Q     Max
## -872.7 -558.2  267.3  537.8  967.3
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                722.3608   13.6481  52.928 < 2e-16 ***
## dummy_month_aug            -5.4805   19.1449  -0.286  0.775
## room                      -135.1272   15.2635  -8.853 < 2e-16 ***
## host_is_superhostTRUE     152.3419   18.7147   8.140 4.35e-16 ***
## dummy_month_aug:room       -0.9058   21.4550  -0.042  0.966
## dummy_month_aug:host_is_superhostTRUE -11.4909   26.3601  -0.436  0.663
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 512.9 on 11652 degrees of freedom
## Multiple R-squared:  0.03424,    Adjusted R-squared:  0.03383
## F-statistic: 82.63 on 5 and 11652 DF,  p-value: < 2.2e-16
```

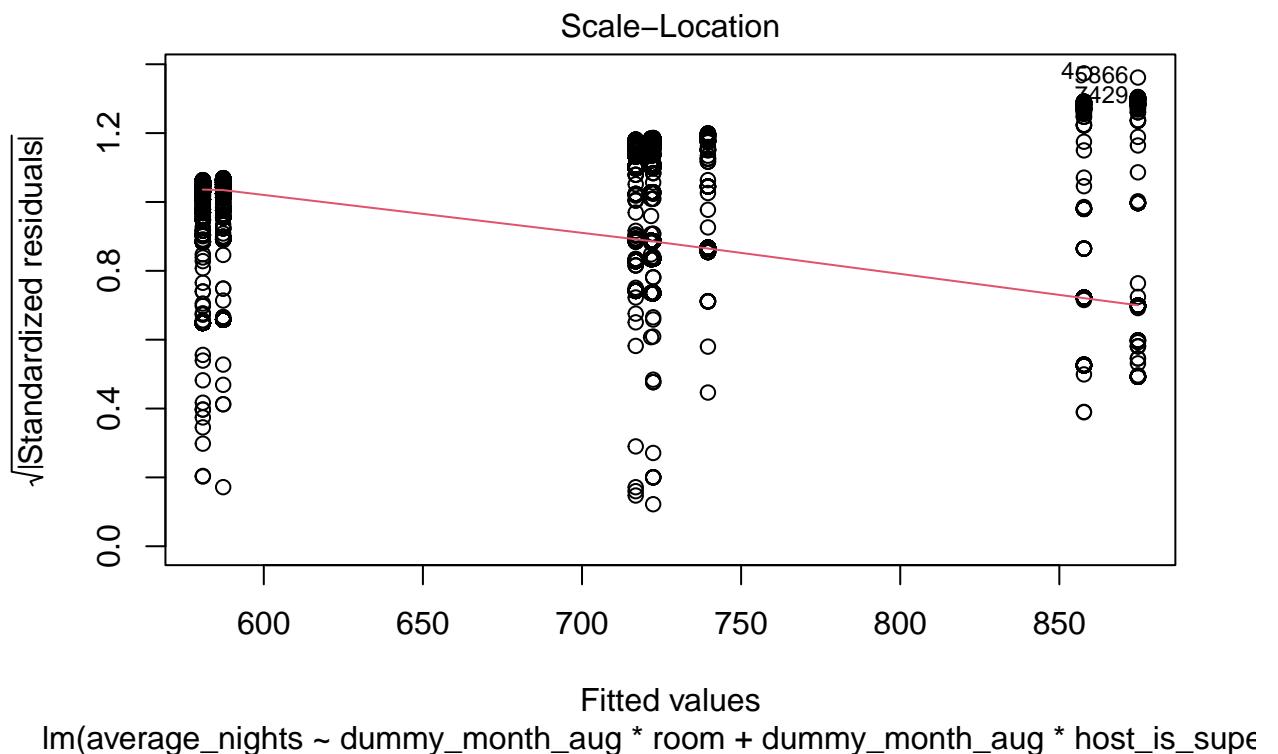
The positive median indicates that the residuals are skewed to the right.

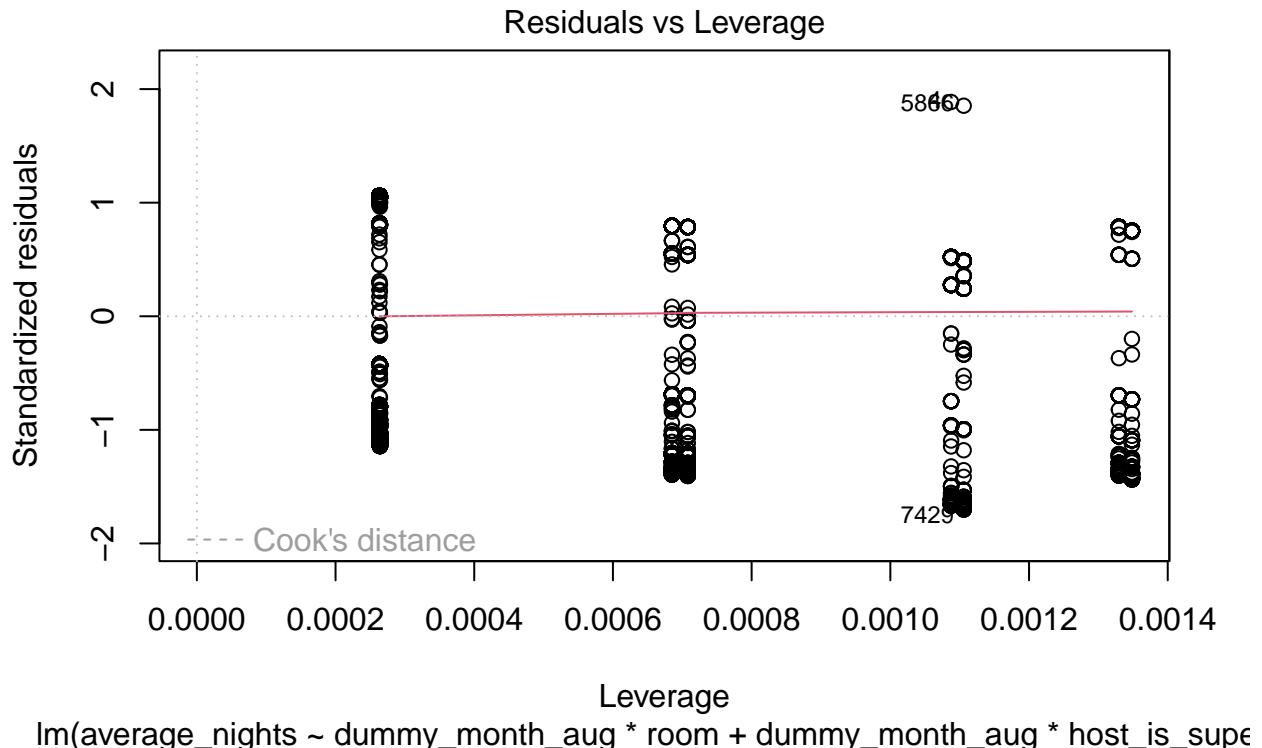
However, since the sample is large, we proceed with the model despite the fact that the normality assumption is violated.

The p-value of the F-statistic is below 0.05, which indicates that the model is significant. However, a very low percentage of the variance in the data is explained by the model (only ~3%).









Berlin

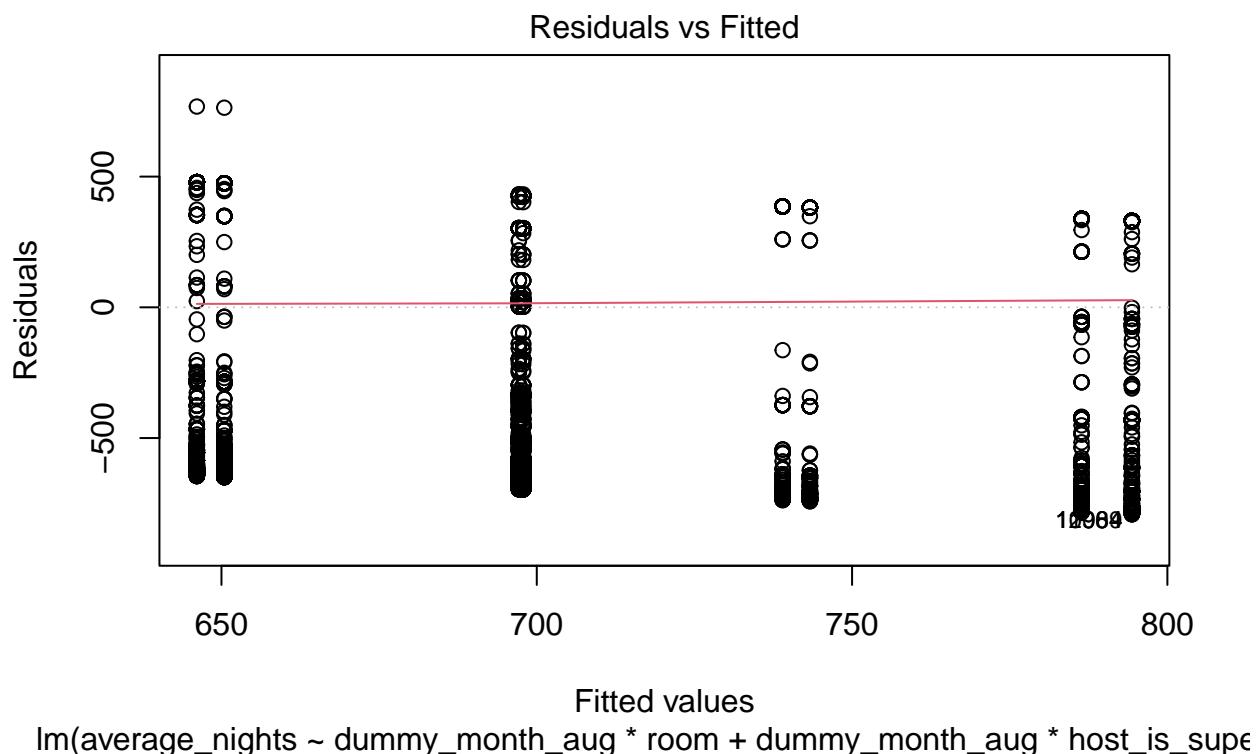
```
##
## Call:
## lm(formula = average_nights ~ dummy_month_aug * room + dummy_month_aug *
##      host_is_superhost, data = dataset)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -791.4 -631.1  386.0  427.8  767.9
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 650.452    6.488 100.255 < 2e-16 ***
## dummy_month_aug              -4.354    9.158 -0.475  0.634
## room                         47.402   8.457  5.605 2.10e-08 ***
## host_is_superhostTRUE        88.544   13.033  6.794 1.11e-11 ***
## dummy_month_aug:room          3.663   11.960  0.306  0.759
## dummy_month_aug:host_is_superhostTRUE  8.677   18.567  0.467  0.640
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 514.5 on 30397 degrees of freedom
## (20 observations deleted due to missingness)
## Multiple R-squared:  0.005938,   Adjusted R-squared:  0.005775
```

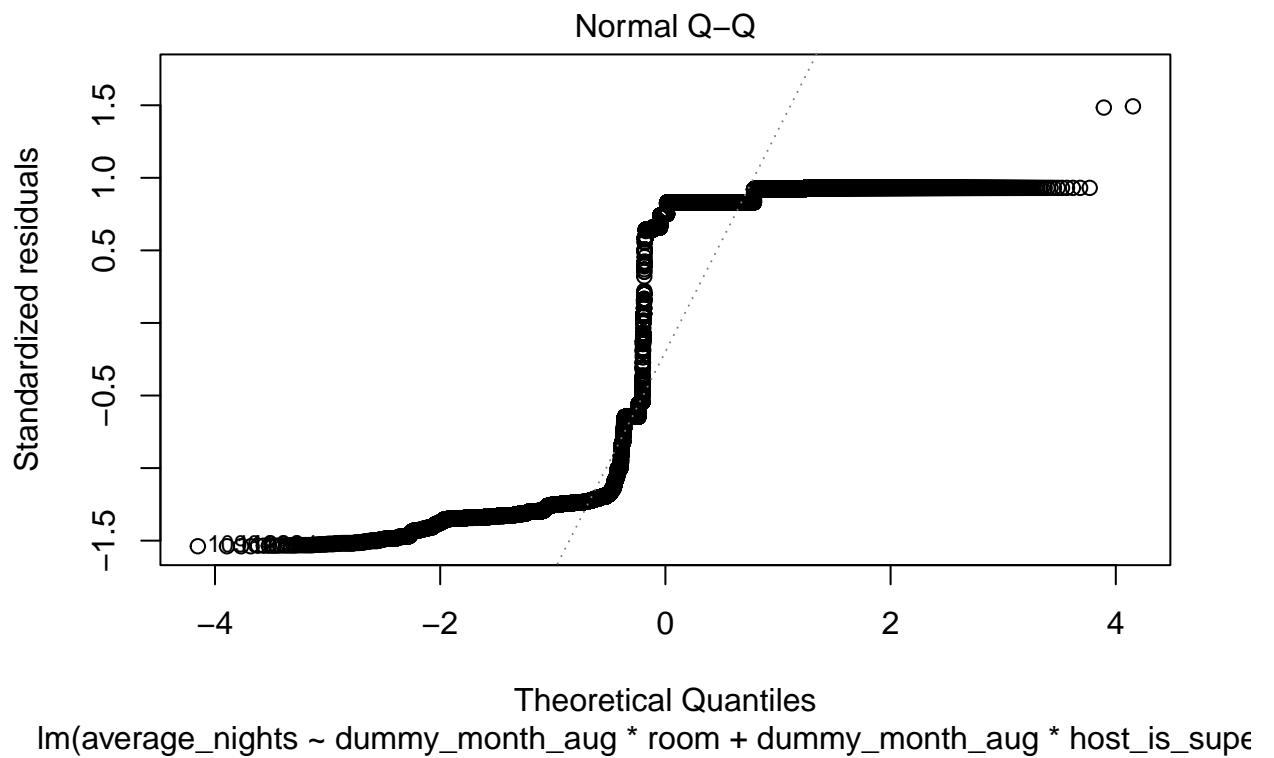
```
## F-statistic: 36.32 on 5 and 30397 DF, p-value: < 2.2e-16
```

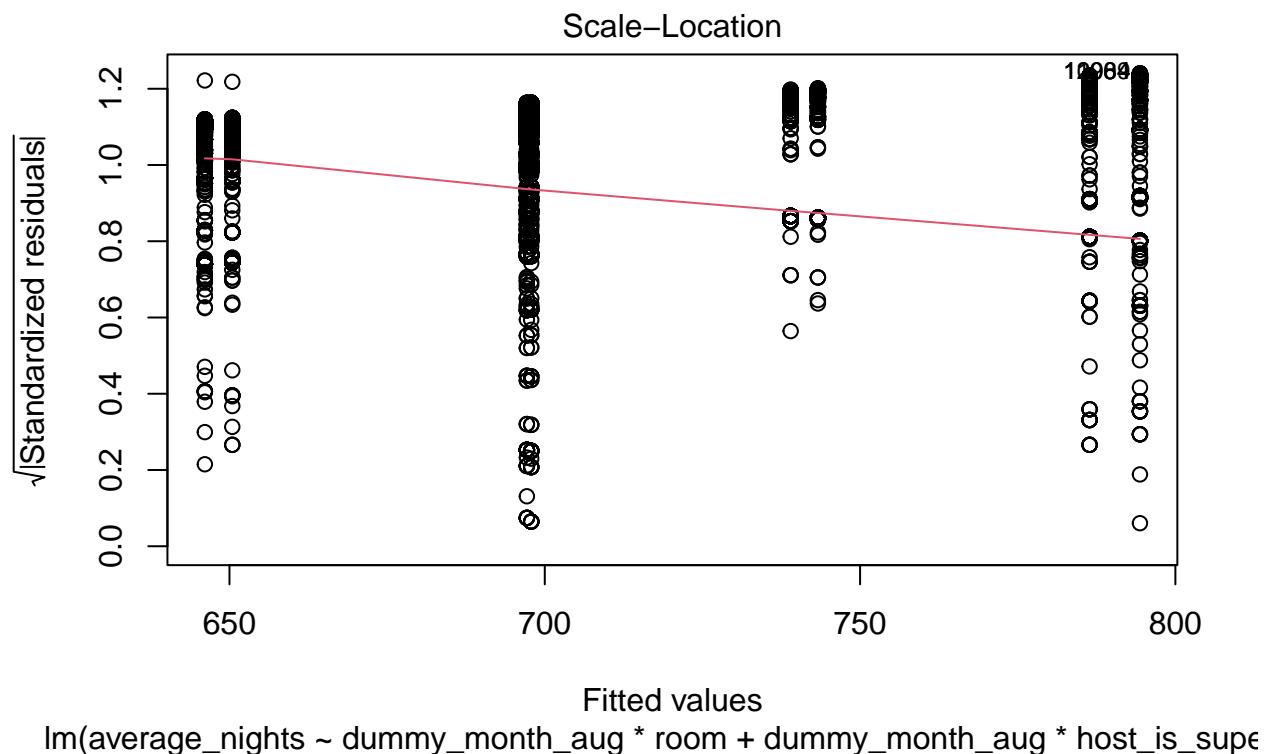
The positive median indicates that the residuals are skewed to the right.

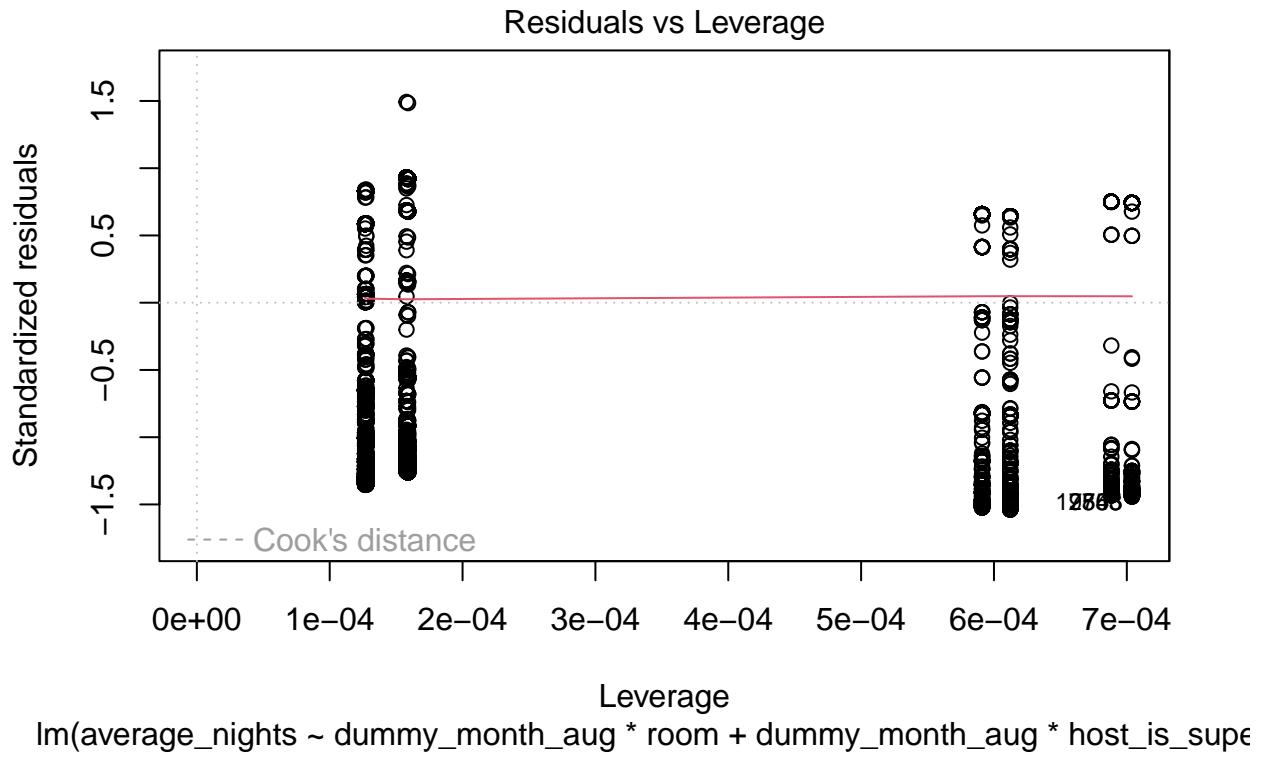
However, since the sample is large, we proceed with the model despite the fact that the normality assumption is violated.

The p-value of the F-statistic is below 0.05, which indicates that the model is significant. However, an extremely low percentage of the variance in the data is explained by the model (<1%).









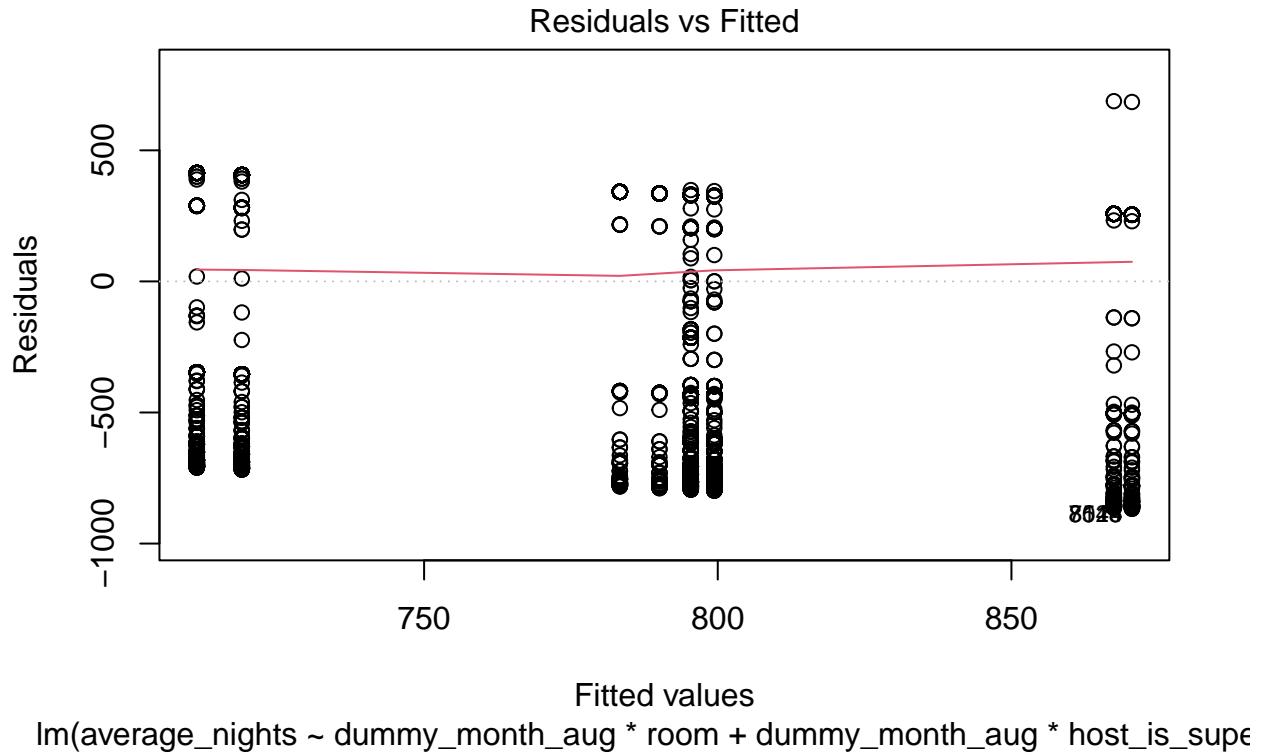
Brussels

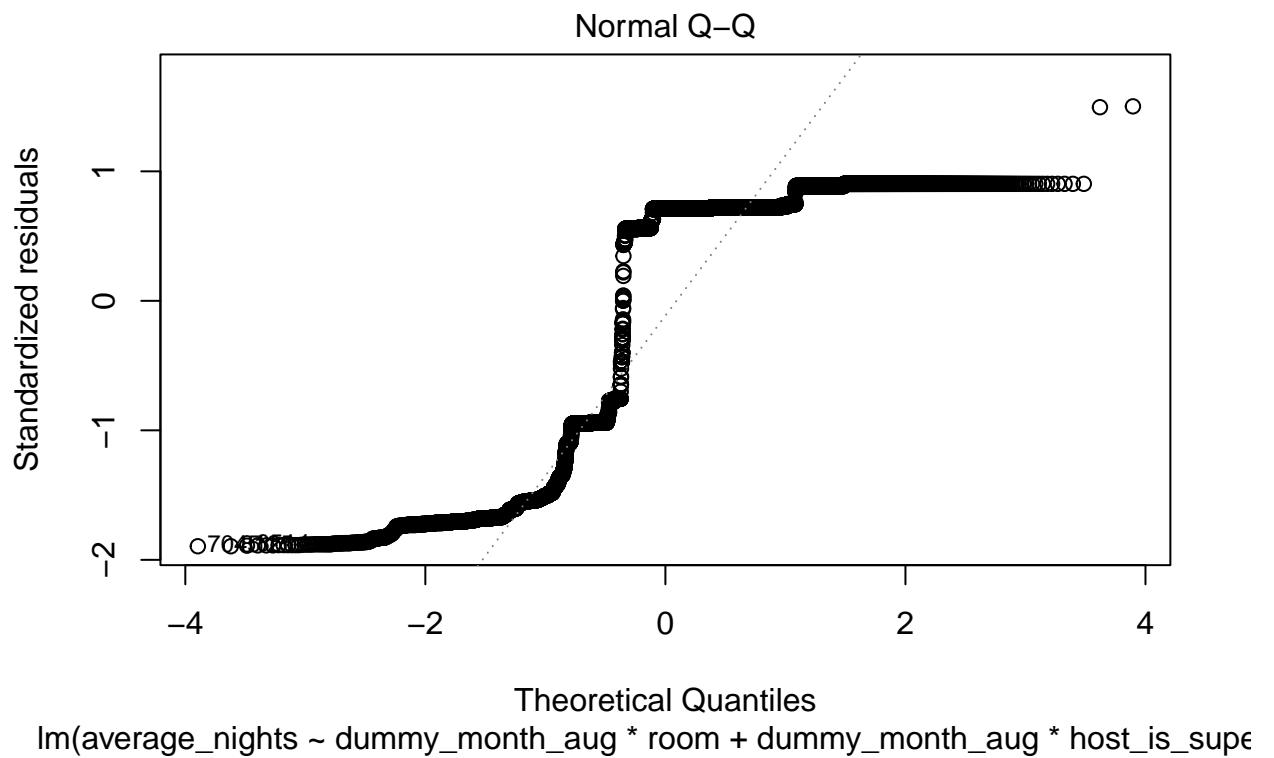
```
##
## Call:
## lm(formula = average_nights ~ dummy_month_aug * room + dummy_month_aug *
##     host_is_superhost, data = dataset)
##
## Residuals:
##    Min      1Q Median      3Q     Max
## -867.5 -434.4  325.6  329.6  687.6
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                718.9665   12.1219  59.312 < 2e-16 ***
## dummy_month_aug            -7.6618   17.1172 -0.448   0.654
## room                      80.4451   14.0652  5.719 1.10e-08 ***
## host_is_superhostTRUE     71.1089   17.5316  4.056 5.03e-05 ***
## dummy_month_aug:room       3.6626   19.8494  0.185   0.854
## dummy_month_aug:host_is_superhostTRUE  0.9224   24.7500  0.037   0.970
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 458.1 on 10202 degrees of freedom
## Multiple R-squared:  0.009922,   Adjusted R-squared:  0.009437
## F-statistic: 20.45 on 5 and 10202 DF,   p-value: < 2.2e-16
```

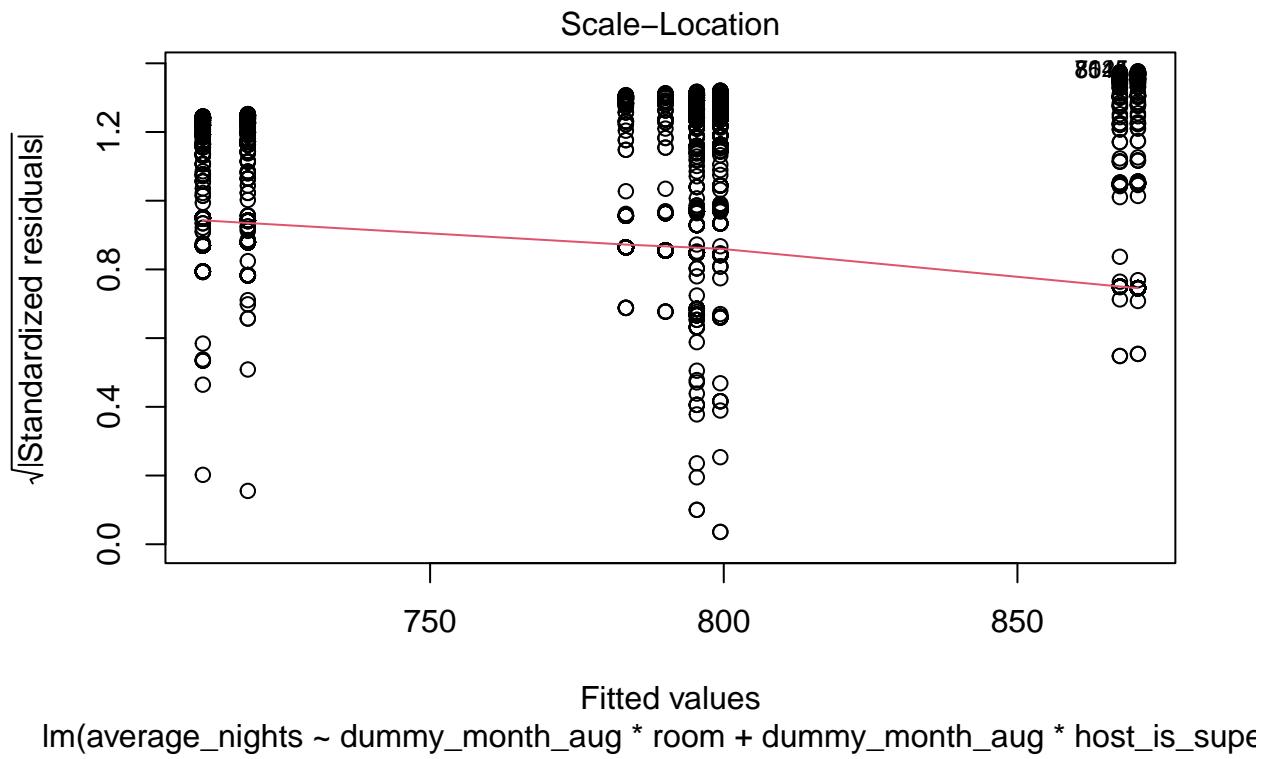
The positive median indicates that the residuals are skewed to the right.

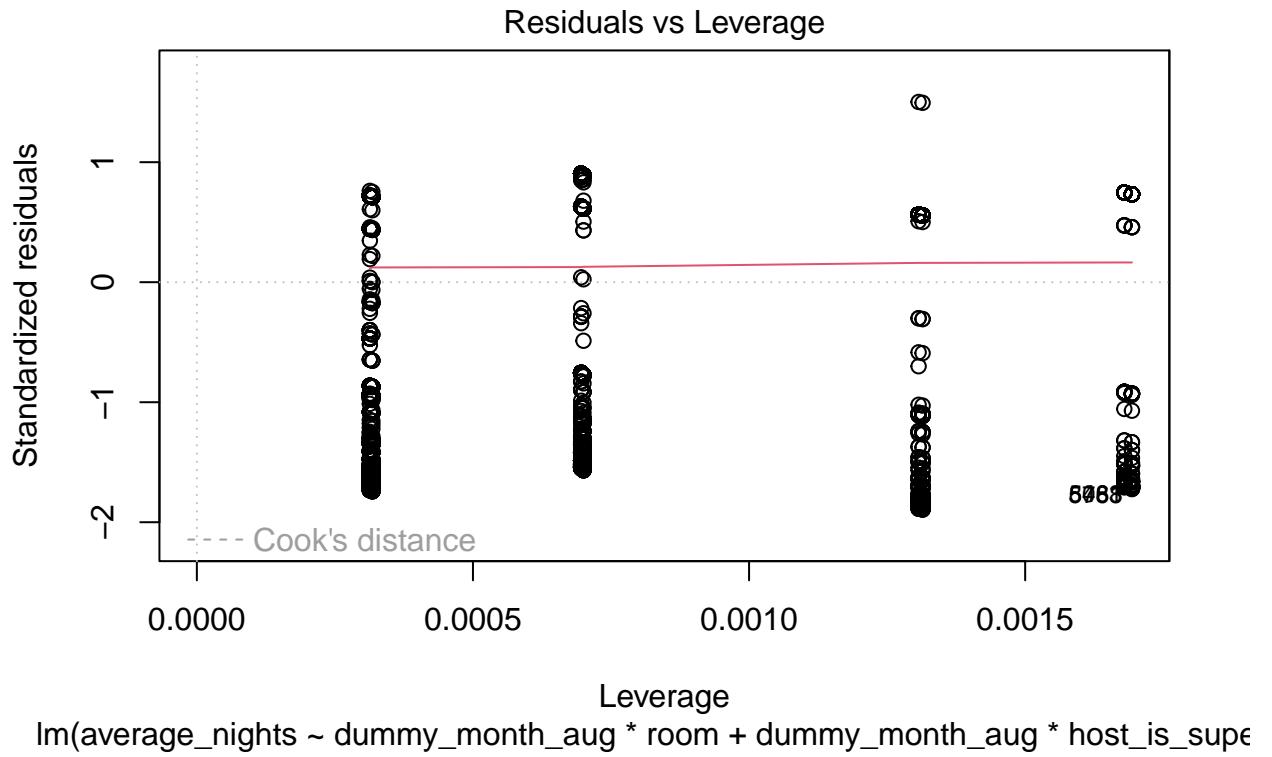
However, since the sample is large, we proceed with the model despite the fact that the normality assumption is violated.

The p-value of the F-statistic is below 0.05, which indicates that the model is significant. However, an extremely low percentage of the variance in the data is explained by the model (~1%).









### Rotterdam

```
##
## Call:
## lm(formula = average_nights ~ dummy_month_aug * room + dummy_month_aug *
##      host_is_superhost, data = dataset)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -814.4 -595.4  336.4  395.8  399.0
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 756.83     29.12  25.992 <2e-16 ***
## dummy_month_aug              -16.92     40.57  -0.417  0.677
## room                         -30.79     34.22  -0.900  0.368
## host_is_superhostTRUE        62.61     38.76   1.616  0.106
## dummy_month_aug:room          20.12     47.95   0.420  0.675
## dummy_month_aug:host_is_superhostTRUE -48.51     54.72  -0.887  0.375
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 482 on 1760 degrees of freedom
## Multiple R-squared:  0.002339, Adjusted R-squared:  -0.0004956
## F-statistic: 0.8251 on 5 and 1760 DF,  p-value: 0.5317
```

The positive median indicates that the residuals are skewed to the right.

However, since the sample is large, we proceed with the model despite the fact that the normality assumption is violated.

The p-value of the F-statistic is below 0.05, which indicates that the model is significant. However, an extremely low percentage of the variance in the data is explained by the model (<1%).

