Energy efficiency in buildings

Asiah Zibrila

Yao Zong

Deysy Londono

Introduction

Keeping a comfortable temperature in a building:

- Significant portion of the energy in the average home
- Contributes to global Energy consumption
- Concerns about exhaustions of energy resources
- Legal constraints in building energy performance

Goal

Model the relationship between the attributes of the building and the heating load.

Dataset

Name	Description	Units
X1	Relative Compactness	None
X2	Surface Area	m^2
X3	Wall Area	m^2
X4	Roof Area	m^2
X5	Overall Height	m
X6	Orientation (2, 3, 4, 5)	None
X7	Glazing Area (%)	None
X8	Glazing Area Distribution (1,2, 3, 4,5)	None
Y1	Heating Load (HL)	kWh/m^2
Y2	Cooling Load (CL)	kWh/m^2

Source: UC Irvine Machine learning repository

It contains energy performance data for 12 different building shapes

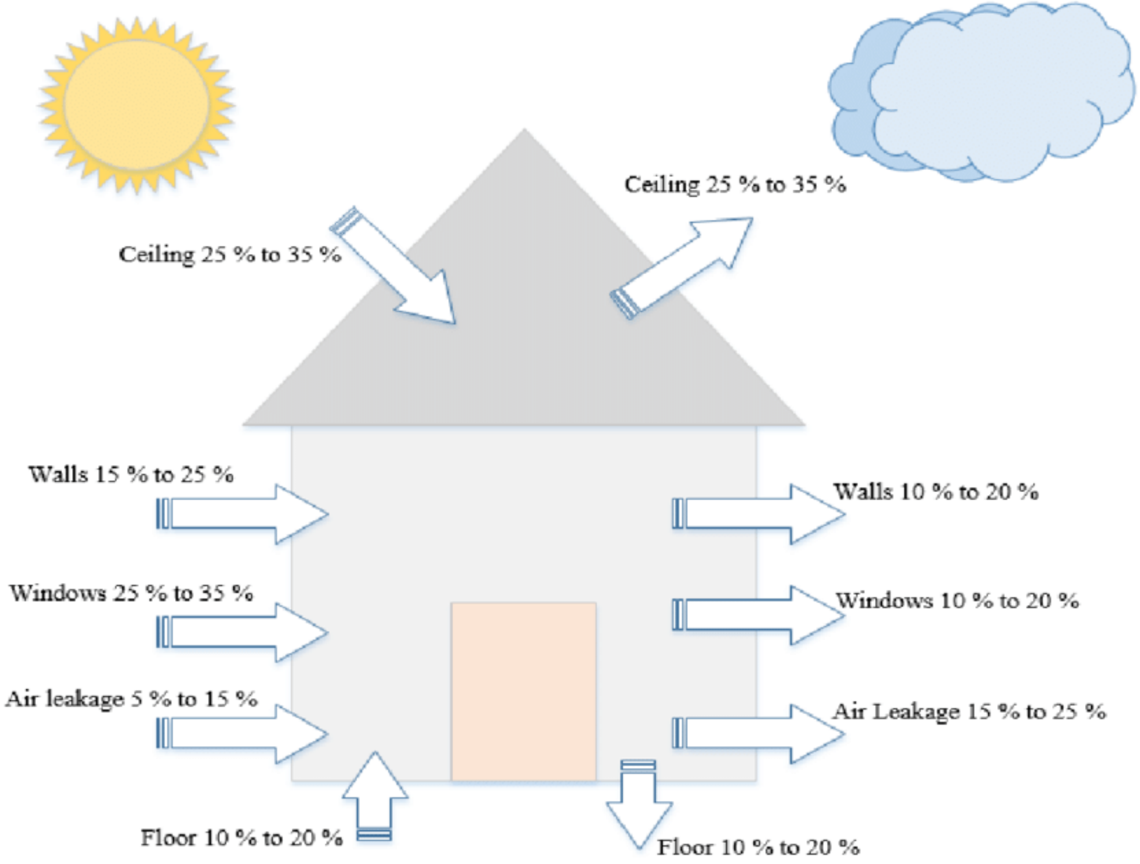
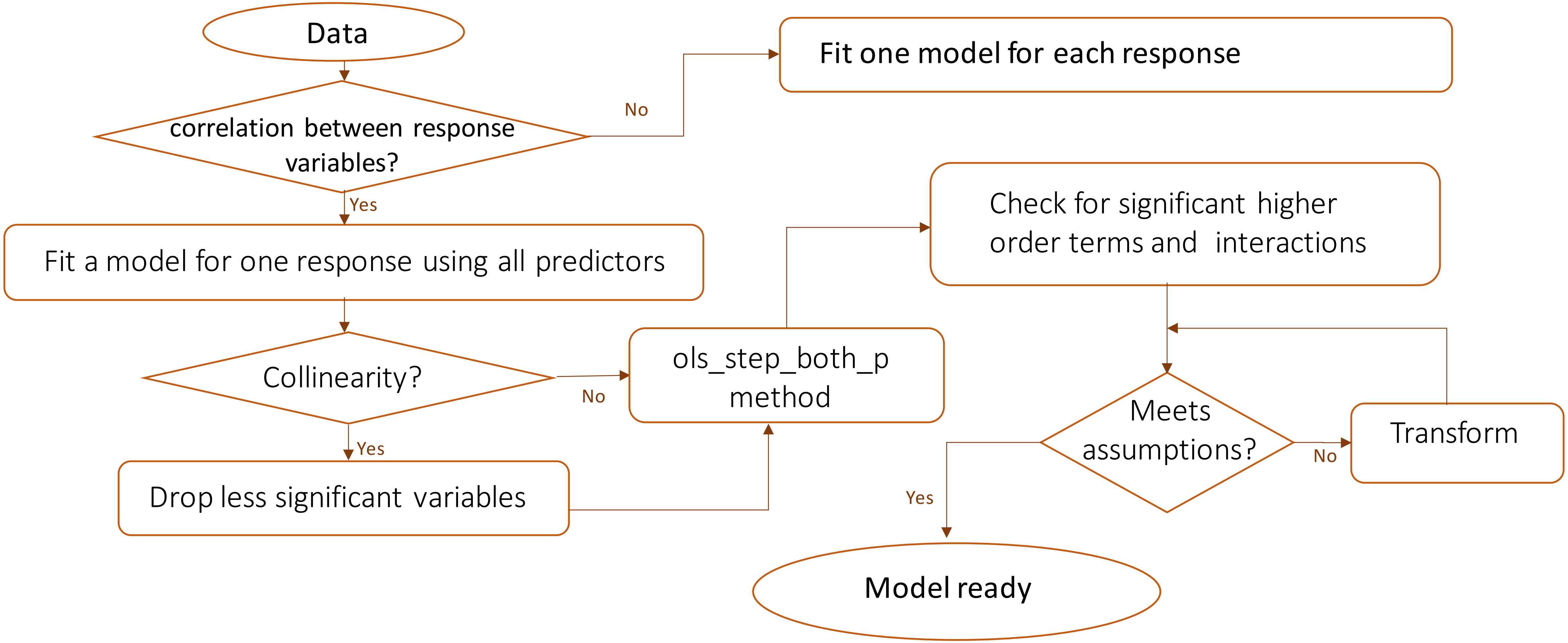


Image adapted from Scientific Figure on ResearchGate.. https://www.researchgate.net/figure342551422/

Methodology

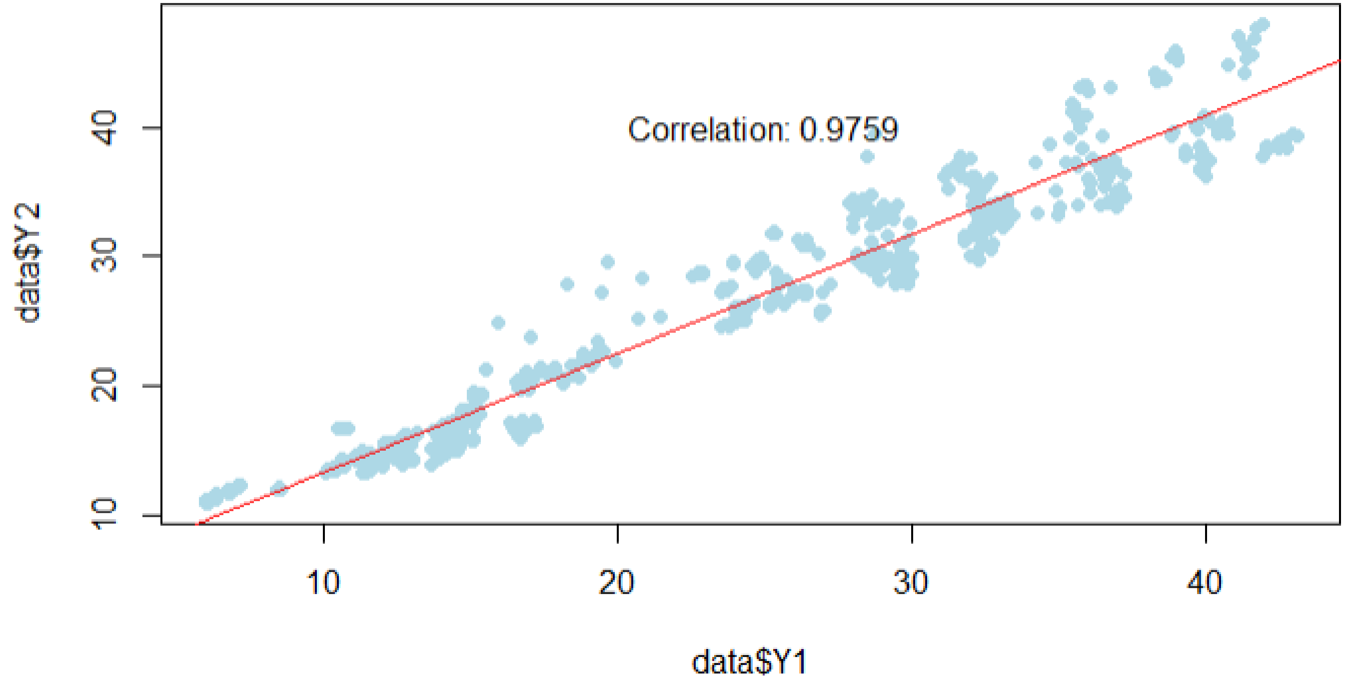


Fitting the Model

Correlation between dependent variables: 0.9759

Y1: Heating Load
Y2: Cooling Load

Only keeping **Y1**



Fitting the Model - Multicollinearity

VIF Multicollinearity Diagnostics

	VIF	detection
x1	105.5241	1
x2	Inf	1
x3	Inf	1
x4	Inf	1
x5	31.2055	1
factor(x6)3	1.5000	0
factor(x6)4	1.5000	0
factor(x6)5	1.5000	0
x7	1.2604	0
factor(x8)1	3.9271	0
factor(x8)2	3.9271	0
factor(x8)3	3.9271	0
factor(x8)4	3.9271	0
factor(x8)5	3.9271	0

Drop X2

VIF Multicollinearity Diagnostics

	VIF	detection
x1	105.5241	1
x3	27.6627	1
x4	211.9383	1
x5	31.2055	1
factor(x6)3	1.5000	0
factor(x6)4	1.5000	0
factor(x6)5	1.5000	0
x7	1.2604	0
factor(x8)1	3.9271	0
factor(x8)2	3.9271	0
factor(x8)3	3.9271	0
factor(x8)4	3.9271	0
factor(x8)5	3.9271	0

Drop X4

VIF Multicollinearity Diagnostics

	VIF	detection
x1	9.2503	0
x3	3.1619	0
x5	9.6261	0
factor(x6)3	1.5000	0
factor(x6)4	1.5000	0
factor(x6)5	1.5000	0
x7	1.2604	0
factor(x8)1	3.9271	0
factor(x8)2	3.9271	0
factor(x8)3	3.9271	0
factor(x8)4	3.9271	0
factor(x8)5	3.9271	0

Fitting the Model - Stepwise Selection

Independent variables: X1(Relative Compactness), X3(Wall Area), X5(Overall Height), X7(Glazing Area %), X8(Orientation, dummy)

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	-15.184	2.589		-5.864	0.000	-20.267	-10.100
x5	5.607	0.182	0.973	30.756	0.000	5.249	5.965
x7	16.848	0.867	0.222	19.431	0.000	15.146	18.551
x3	0.035	0.004	0.151	8.340	0.000	0.027	0.043
factor(x8)1	4.528	0.522	0.175	8.673	0.000	3.503	5.553
factor(x8)2	4.436	0.522	0.172	8.497	0.000	3.411	5.461
factor(x8)3	4.183	0.522	0.162	8.012	0.000	3.158	5.208
factor(x8)4	4.388	0.522	0.170	8.406	0.000	3.363	5.413
factor(x8)5	4.182	0.522	0.162	8.011	0.000	3.158	5.207
x1	-14.532	2.958	-0.152	-4.912	0.000	-20.340	-8.725

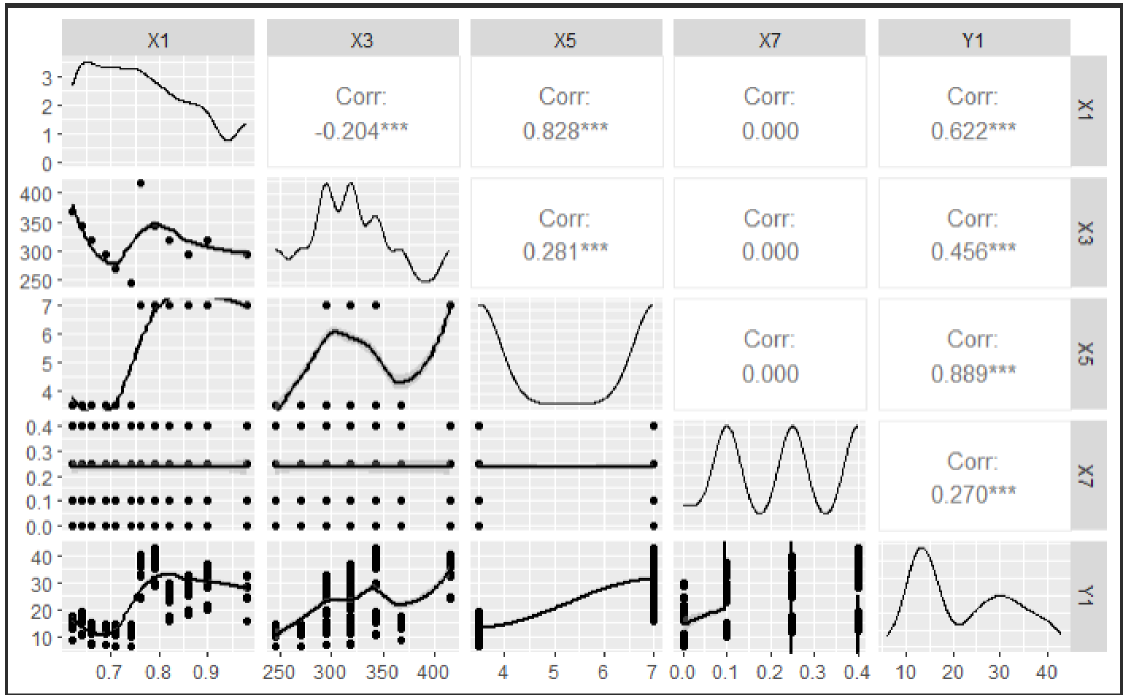
Stepwise Selection Summary

Step	Variable	Added/ Removed	R-Square	Adj. R-Square	C(p)	AIC	RMSE
1	x5	addition	0.791	0.791	1237.6840	4532.4914	4.6149
2	x7	addition	0.864	0.864	542.0180	4205.3741	3.7273
3	x3	addition	0.910	0.910	103.5860	3890.8720	3.0352
4	factor(x8)	addition	0.919	0.918	21.2450	3821.9456	2.8927
5	x1	addition	0.921	0.920	-0.7950	3799.8793	2.8496

```
Call:
lm(formula = Y1 ~ X1 + X3 + X5 + X7 + factor(X8), data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.3068 -1.5588  0.0232  1.4189  7.3450 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -15.183790   2.589482  -5.864 6.76e-09 ***
X1          -14.532402   2.958472  -4.912 1.10e-06 ***
X3            0.034976   0.004194   8.340 3.49e-16 ***
X5            5.606753   0.182300  30.756  < 2e-16 ***
X7           16.848333   0.867099  19.431  < 2e-16 ***
factor(X8)1   4.527653   0.522063   8.673  < 2e-16 ***
factor(X8)2   4.435986   0.522063   8.497  < 2e-16 ***
factor(X8)3   4.183000   0.522063   8.012 4.24e-15 ***
factor(X8)4   4.388208   0.522063   8.406  < 2e-16 ***
factor(X8)5   4.182444   0.522063   8.011 4.28e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.85 on 758 degrees of freedom
Multiple R-squared:  0.9212,	Adjusted R-squared:  0.9202 
F-statistic: 984.3 on 9 and 758 DF,  p-value: < 2.2e-16
```

Fitting the Model – Higher Order Terms

Pairwise Correlation Plot :



4th Order Model: Adjusted R-squared = 0.985

```
Call:
lm(formula = Y1 ~ X1 + X3 + X5 + X7 + factor(X8) + I(X1^2) +
    I(X1^3) + I(X3^2) + I(X3^3) + I(X1^4) + I(X3^4), data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-4.1708 -0.7470  0.0030  0.7912  4.0091

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.374e+04  7.989e+02  -42.23   <2e-16 ***
X1           1.447e+05  3.590e+03   40.32   <2e-16 ***
X3           7.225e+01  2.295e+00   31.48   <2e-16 ***
X5           1.697e+01  4.816e-01   35.23   <2e-16 ***
X7           1.685e+01  3.763e-01   44.77   <2e-16 ***
factor(X8)1  4.528e+00  2.266e-01   19.98   <2e-16 ***
factor(X8)2  4.436e+00  2.266e-01   19.58   <2e-16 ***
factor(X8)3  4.183e+00  2.266e-01   18.46   <2e-16 ***
factor(X8)4  4.388e+00  2.266e-01   19.37   <2e-16 ***
factor(X8)5  4.182e+00  2.266e-01   18.46   <2e-16 ***
I(X1^2)     -2.727e+05  6.824e+03  -39.96   <2e-16 ***
I(X1^3)      2.258e+05  5.692e+03   39.66   <2e-16 ***
I(X3^2)     -3.672e-01  1.108e-02  -33.15   <2e-16 ***
I(X3^3)      8.109e-04  2.344e-05   34.59   <2e-16 ***
I(X1^4)     -6.942e+04  1.761e+03  -39.42   <2e-16 ***
I(X3^4)     -6.580e-07  1.836e-08  -35.84   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.237 on 752 degrees of freedom
Multiple R-squared:  0.9853,    Adjusted R-squared:  0.985
F-statistic:  3354 on 15 and 752 DF,  p-value: < 2.2e-16
```

Fitting the Model – Interaction Terms

Significant Interaction Terms:

X1:X3

X1:X5

X1:X7

X3:X5

X3:X7

Adjusted R-squared: 0.9353

```
Call:
lm(formula = Y1 ~ X1 + X3 + X5 + X7 + factor(X8) + X1:X3 + X1:X5 +
    X1:X7 + X3:X5 + X3:X7, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-6.7845 -1.0662 -0.0875  0.7896  6.8684

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -236.96103   96.19470  -2.463 0.013987 *
X1           189.06644   94.48693   2.001 0.045753 *
X3             0.05769    0.08703   0.663 0.507595
X5            50.23088   15.12360   3.321 0.000939 ***
X7           -51.61230    8.04777  -6.413 2.51e-10 ***
factor(X8)1    4.52765    0.47007   9.632  < 2e-16 ***
factor(X8)2    4.43599    0.47007   9.437  < 2e-16 ***
factor(X8)3    4.18300    0.47007   8.899  < 2e-16 ***
factor(X8)4    4.38821    0.47007   9.335  < 2e-16 ***
factor(X8)5    4.18244    0.47007   8.897  < 2e-16 ***
X1:X3          0.39709    0.08367   4.746 2.48e-06 ***
X1:X5        -46.94903   15.14544  -3.100 0.002008 **
X1:X7         58.73302    6.71975   8.740  < 2e-16 ***
X3:X5         -0.04490    0.01618  -2.775 0.005653 **
X3:X7          0.07403    0.01629   4.544 6.43e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.566 on 753 degrees of freedom
Multiple R-squared:  0.9365,    Adjusted R-squared:  0.9353
F-statistic: 793.5 on 14 and 753 DF,  p-value: < 2.2e-16
```

Fitting the Model – Final Model

Interaction and Higher Order terms combined:

Adjust R-squared: 0.997

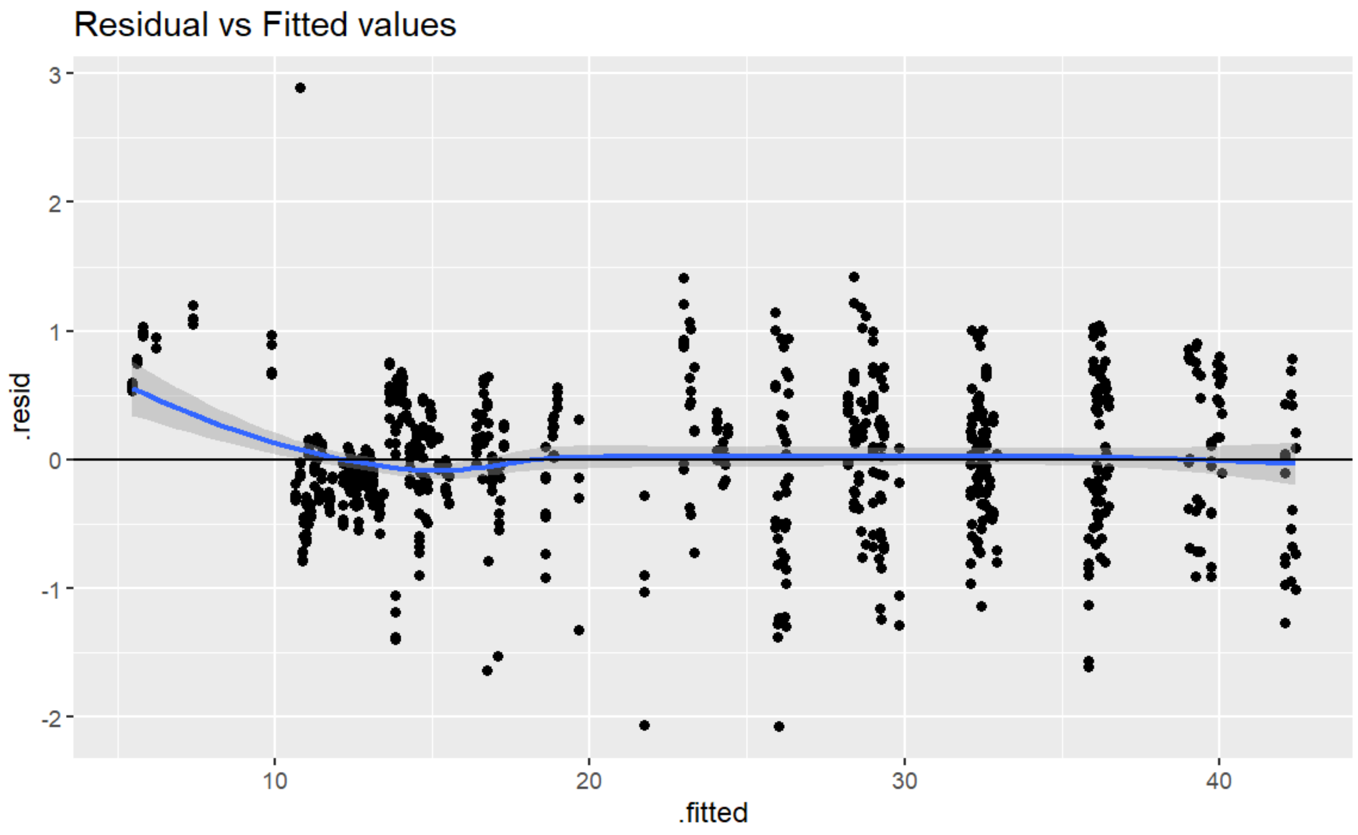
RMSE: 0.5542

$$\widehat{Y_1} = -4090 + 17520X1 + 76.82X3 + 81.93X5 - 51.61X7 + 4.52X8_1$$
$$+4.436X8_2 + 4.18X8_3 + 4.38X8_4 + 4.182X8_5 - 32730X1^2 + 27140X1^3$$
$$-0.373X_2^2 + 0.0081X3^3 - 83660X1^4 - 0.000006X3^4$$
$$-2.5(X1 \times X3) - 98.56(X1 \times X5) + 58.73(X1 \times X7) + 740.3(X3 \times X7)$$

```
Call:
lm(formula = Y1 ~ X1 + X3 + X5 + X7 + factor(X8) + X1:X3 + X1:X5 +
    X1:X7 + X3:X7 + I(X1^2) + I(X1^3) + I(X3^2) + I(X3^3) + I(X1^4) +
    I(X3^4), data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-2.07036 -0.30361 -0.02345  0.35724  2.88851

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.090e+04  4.619e+02 -88.547  < 2e-16 ***
X1           1.752e+05  2.346e+03  74.672  < 2e-16 ***
X3           7.682e+01  1.129e+00  68.065  < 2e-16 ***
X5           8.193e+01  2.123e+01   3.859 0.000124 ***
X7          -5.161e+01  1.738e+00 -29.692  < 2e-16 ***
factor(X8)1  4.528e+00  1.015e-01  44.593  < 2e-16 ***
factor(X8)2  4.436e+00  1.015e-01  43.690  < 2e-16 ***
factor(X8)3  4.183e+00  1.015e-01  41.198  < 2e-16 ***
factor(X8)4  4.388e+00  1.015e-01  43.220  < 2e-16 ***
factor(X8)5  4.182e+00  1.015e-01  41.193  < 2e-16 ***
I(X1^2)     -3.273e+05  5.090e+03 -64.311  < 2e-16 ***
I(X1^3)      2.714e+05  4.910e+03  55.276  < 2e-16 ***
I(X3^2)     -3.732e-01  6.017e-03 -62.034  < 2e-16 ***
I(X3^3)      8.161e-04  1.432e-05  56.986  < 2e-16 ***
I(X1^4)     -8.366e+04  1.694e+03 -49.380  < 2e-16 ***
I(X3^4)     -6.593e-07  1.238e-08 -53.244  < 2e-16 ***
X1:X3       -2.550e+00  8.220e-02 -31.025  < 2e-16 ***
X1:X5       -9.856e+01  2.782e+01  -3.543 0.000420 ***
X1:X7        5.873e+01  1.451e+00  40.466  < 2e-16 ***
X3:X7        7.403e-02  3.519e-03  21.037  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5542 on 748 degrees of freedom
Multiple R-squared:  0.9971,    Adjusted R-squared:  0.997
```

Checking assumptions

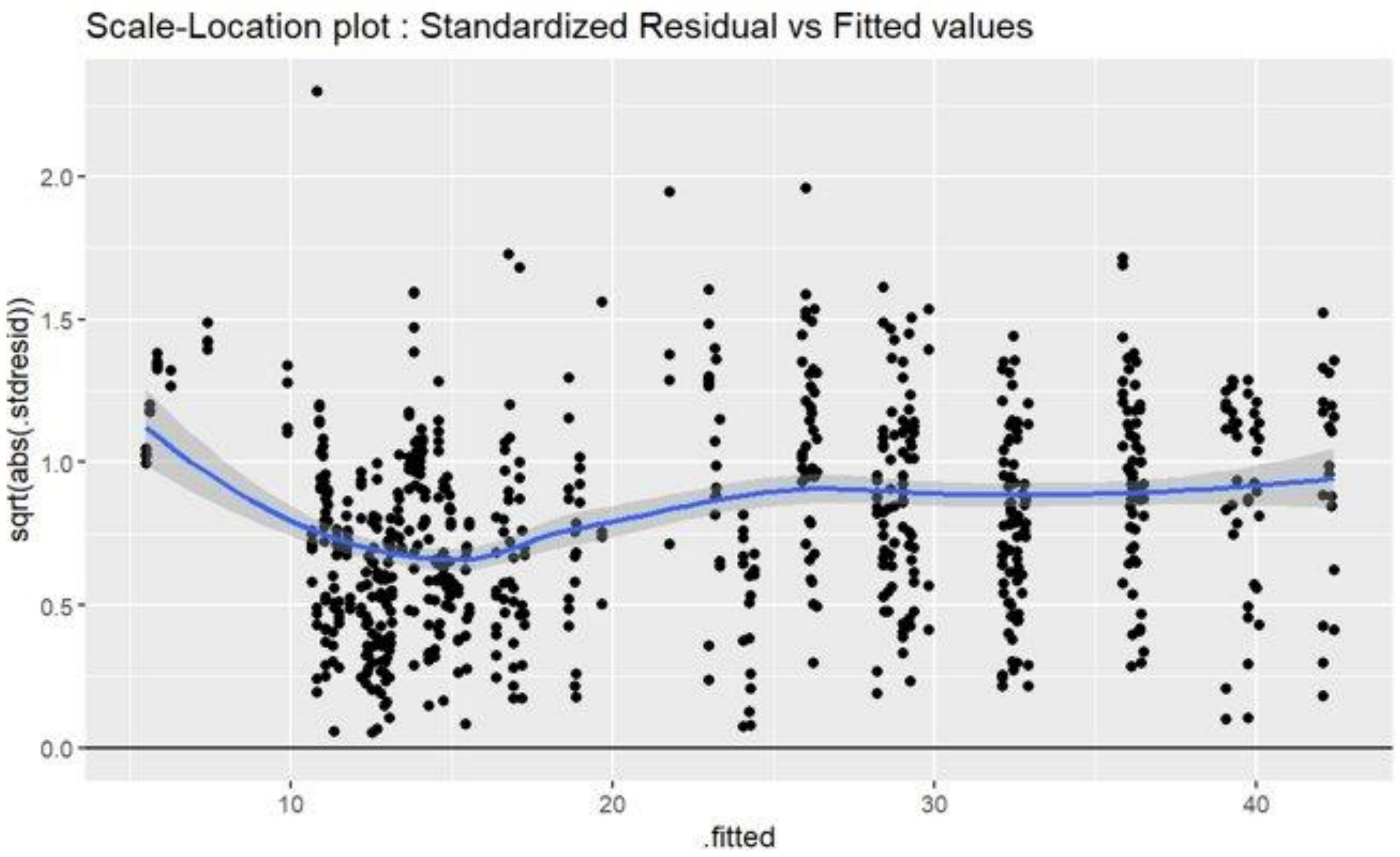
1. Linearity



Checking assumptions

2. Independence

3. Equal Variance



Test for heteroscedasticity

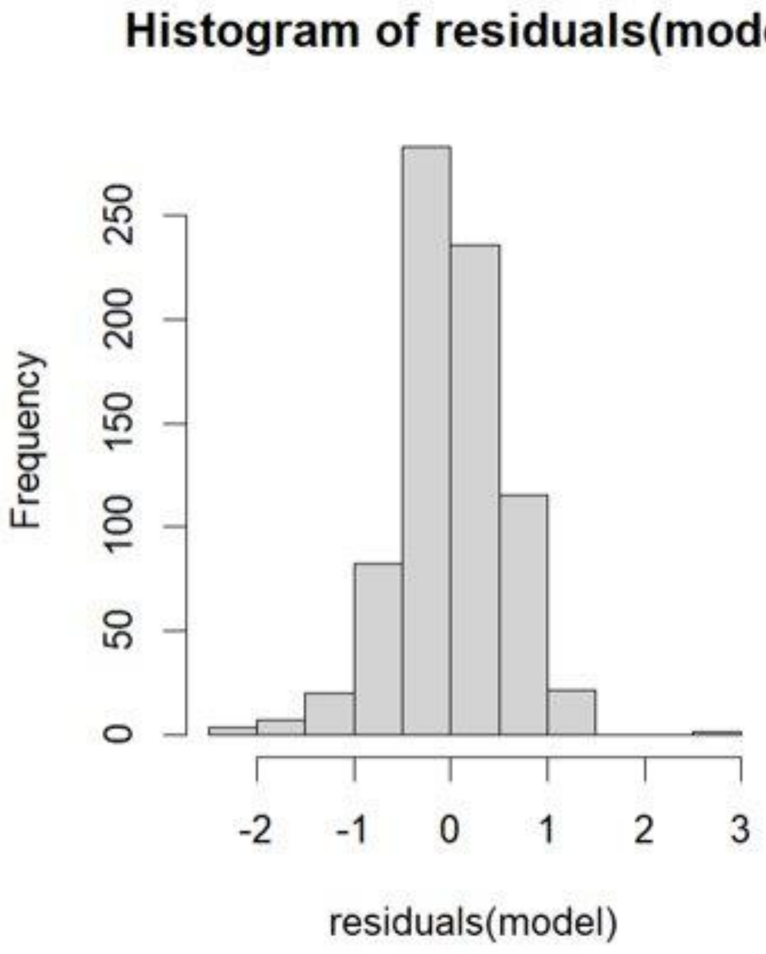
- H_0: heteroscedasticity is not present (homoscedasticity)
- H_A : heteroscedasticity is present

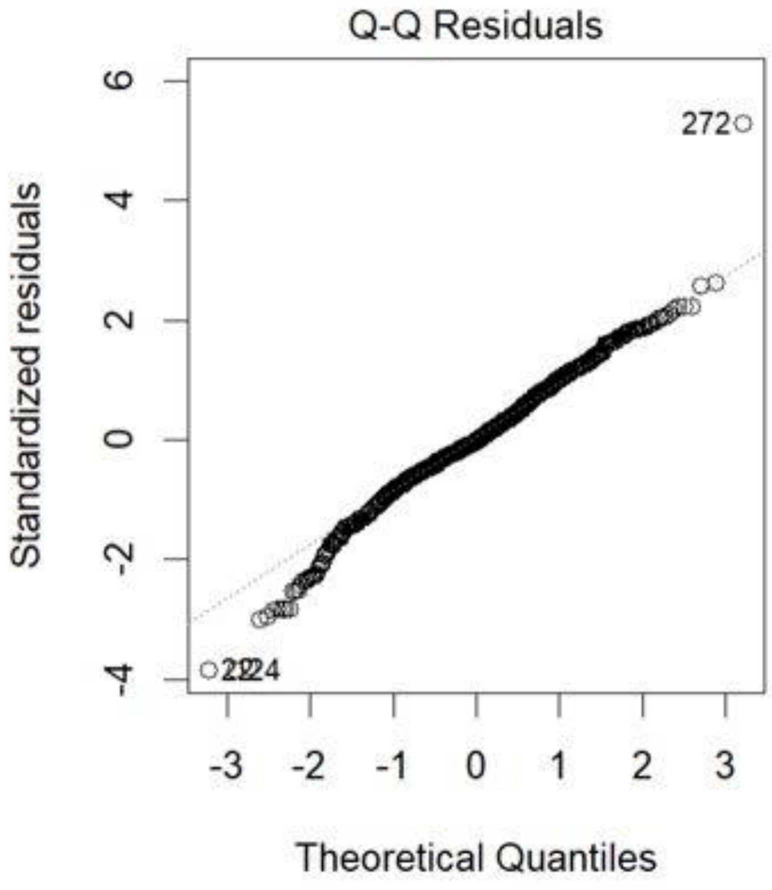
- The Breusch-Pagan test.
- BP = 169.24, df = 19, p-value $< 2.2e\text{-}16 < \alpha = 0.05$

- Null hypothesis rejected

Checking assumptions

4. Normality





Checking assumptions

4. Normality (Statistical Test)

- Ho : the sample data are significantly normally distributed
- Ha : the sample data are not significantly normally distributed
- Shapiro-Wilk test

```
        Shapiro-Wilk normality test

data:  residuals(model)
W = 0.98323, p-value = 1.061e-07
```

- Not normally distributed.

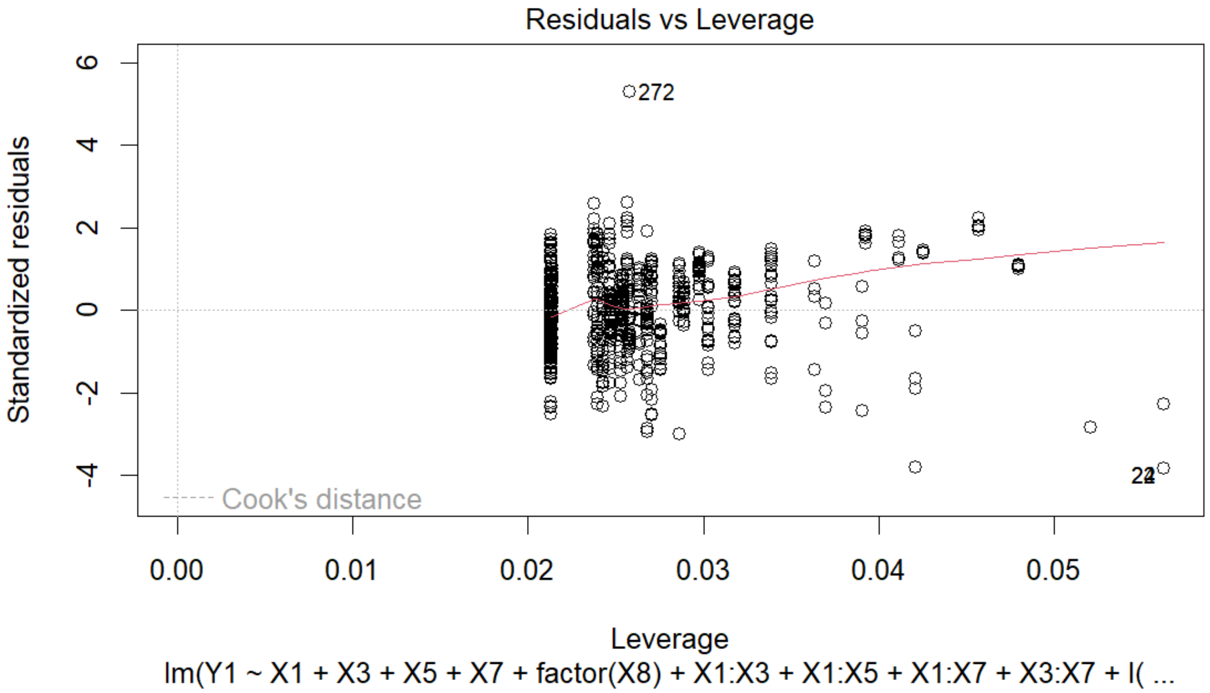
Checking assumptions

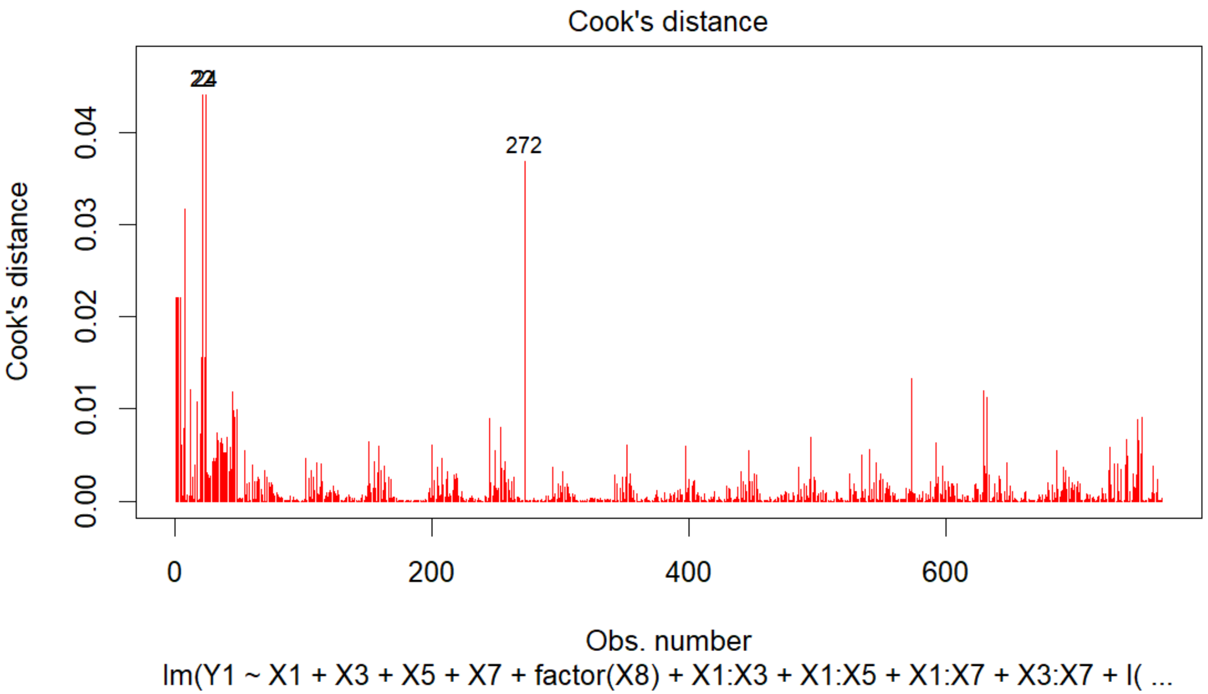
5. Multicollinearity

```
VIF Multicollinearity Diagnostics

                  VIF detection
X1             9.2503         0
X3             3.1619         0
X5             9.6261         0
factor(X6)3 1.5000         0
factor(X6)4 1.5000         0
factor(X6)5 1.5000         0
X7             1.2604         0
factor(X8)1 3.9271         0
factor(X8)2 3.9271         0
factor(X8)3 3.9271         0
factor(X8)4 3.9271         0
factor(X8)5 3.9271         0

NOTE:  VIF Method Failed to detect multicollinearity
```

Checking assumptions

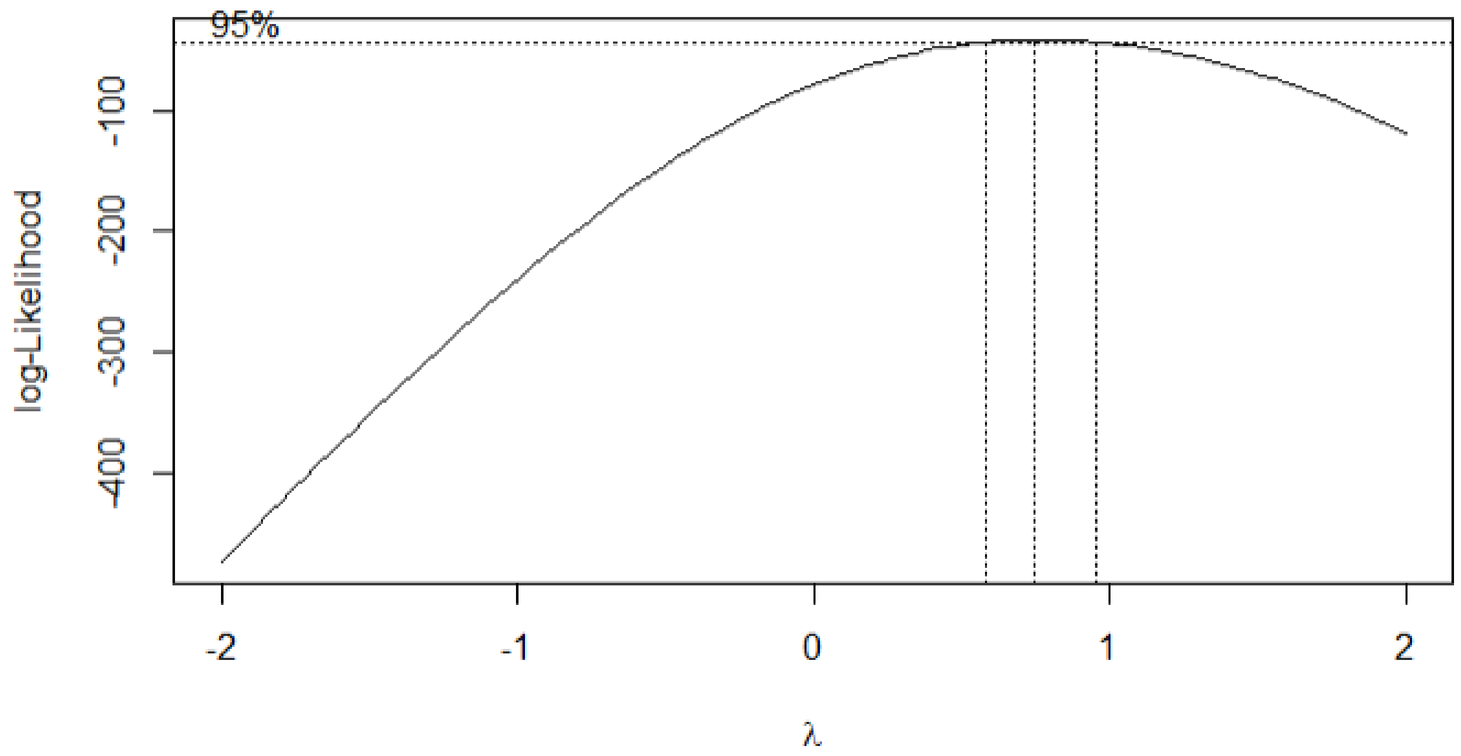
6. Outliers

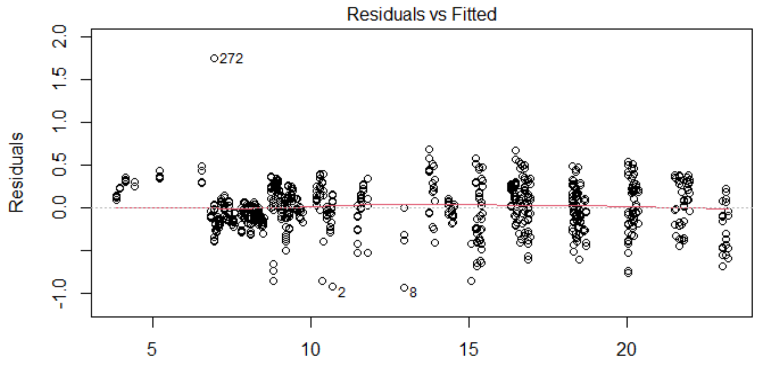


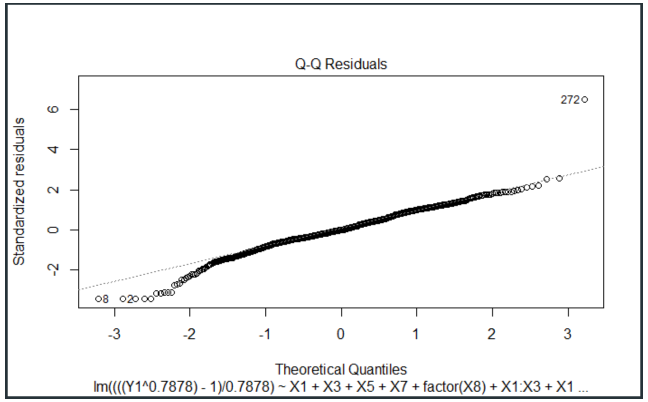


Box-cox Transformation

Search for best lamda: 0.7879



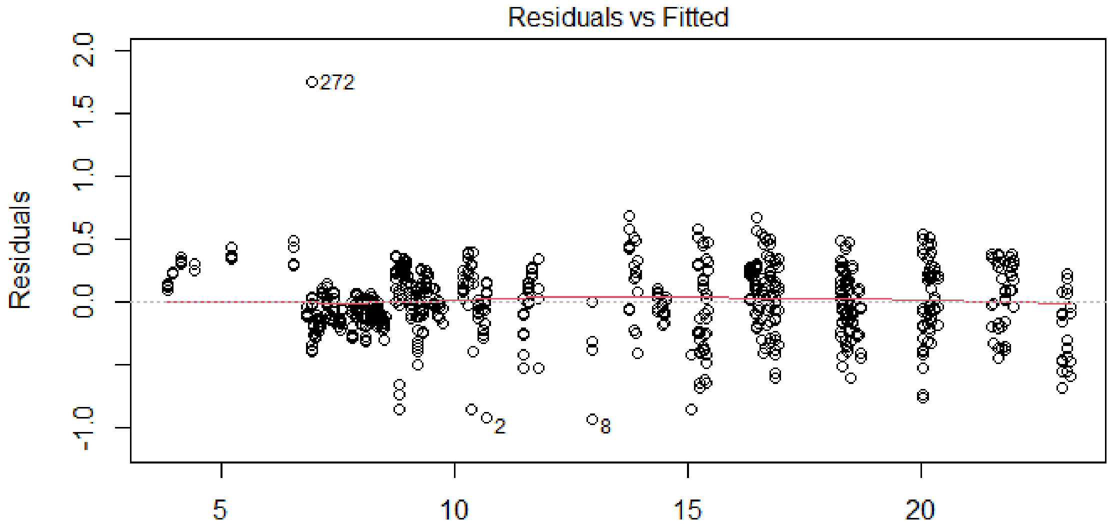


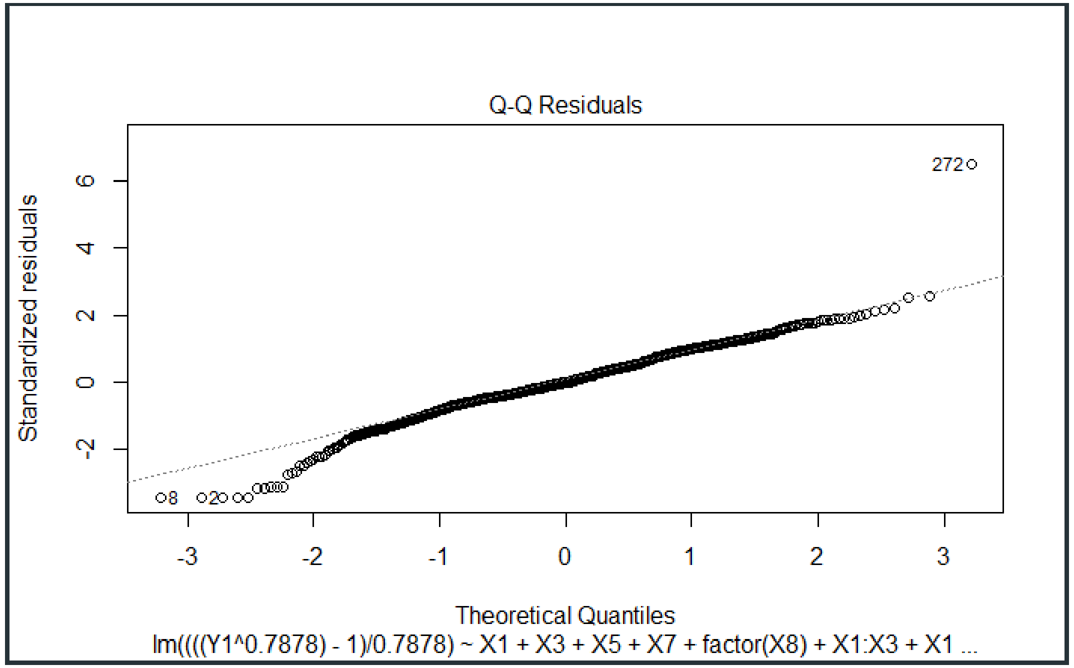


Box-cox Transformation

```
        studentized Breusch-Pagan test

data:  bcmodel
BP = 97.787, df = 19, p-value = 1.345e-12
```

```
        Shapiro-Wilk normality test

data:  residuals(bcmodel)
W = 0.9687, p-value = 9.264e-12
```





Conclusions and Recommendations

- The glazing area(X7) and glazing area distribution (X8) are slightly correlated to each other, however uncorrelated with the first six input variables, and the wall area (X3) is not strongly correlated to any other variable, which made them good candidates for the model .
- Our results indicate that the significant variables for modelling the heating load of a building are relative compactness (x1), wall area(x3), overall height(x5), orientation (x6), glazing area(x7) and glazing area distribution (x8) .
- A Box-Cox transformation on the data was performed and the MLR model still did not meet the assumptions, therefore the forecast accuracy may be distorted.
- Exploring alternative regression models such as robust regression or PLS might improve the model.

References

- L. Pérez-Lombard, J. Ortiz, C. Pout, A review on buildings energy consumption information, Energy Build. 40 (3) (2008) 394–398.
- K. Kavyalola, Robust modeling of heating and cooling loads using partial least squares towards efficient residential building design. Journal of Building Engineering 18 (2018) 467–475.
- UC Irvine Machine learning repository. Energy Efficiency Dataset. https://archive.ics.uci.edu/dataset/242/energy+efficiency
- A. Tsana's, A. Xiara, Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools, Energy Build. 49 (2012) 560–567.