Trouble with Turbulence

Connie Wu, Jason McEachin, Joe Choo, Scott Heng

10/10/2020

Introduction

Background Information

Turbulence is highly versatile motion that is often times difficult to predict and understanding fluid motion and its effect on natural problems poses a great challenge. Interpreting turbulence data is an incredibly important task in the engineering from understanding the cosmos to the cosmic cycle.

Parametric modeling is effective when we want to compactly represent features as model parameters. Unlike certain "black box" machine learning techniques, it offers a higher level of interpretability, which is especially useful for a practical setting. Rather than simply outputting a classification result, a parametric model allows us to investigate in more detail which aspects of turbulence differ between high and low particle cluster volumes.

In our project, we use linear regression and apply it to interpret the difference in model parameters in order to achieve the the following objectives:

1. Build a model that predict its particle cluster volume distribution in terms of the moments.
2. Investigate and interpret how model parameters affects the probability distribution for particle cluster volumes

Data

The data set procured for this case study consists of information about cluster volumes. In total, it contains 89 observations with 7 variables. Details of each variable is specified in Table 1.1. The original response variable, a probability distribution for particle cluster volumes, is difficult to interpret, and therefore is summarized into its first 4 raw moments.

Performing some exploratory data analysis, we can observe from Figure 1 that St, with the exception of the 0 values, follow a linear relationship with the first moment, while for Re, there is high variability for the first moment when $Re = 90$, and low variability otherwise. There is high variability across all Fr values with respect to the first moment. Some of these insights influenced our decisions for modelling approaches which will be further discussed in the Methods section.

Methodology

For the approach to modeling the first moment, we first conducted a simple linear regression with an inverse-logit transformation on Fr (this was used to address the Fr points labeled 'Inf') while keeping Re and St the same. However this model had a low R-squared value with a pattern found in the residuals. The next

Table 1: Table 1.1 Description Table of Data set variables

Metric	Value	Description
St	0 < St < 3	Particle property: effect on inertia (e.g. size, density)
Re	90, 224, 398	Reynolds Number: turbulent flow property
Fr	Infinity, 0, 3	Particle propety: gravitational acceleration
R moment 1	Continuous response variable	First raw moment of probability distribution
R moment 2	Continuous response variable	Second raw moment of volume probability distribution
R moment 3	Continuous response variable	Third raw moment of volume probability distribution
R moment 4	Continuous response variable	Fourth raw moment of volume probability distribution

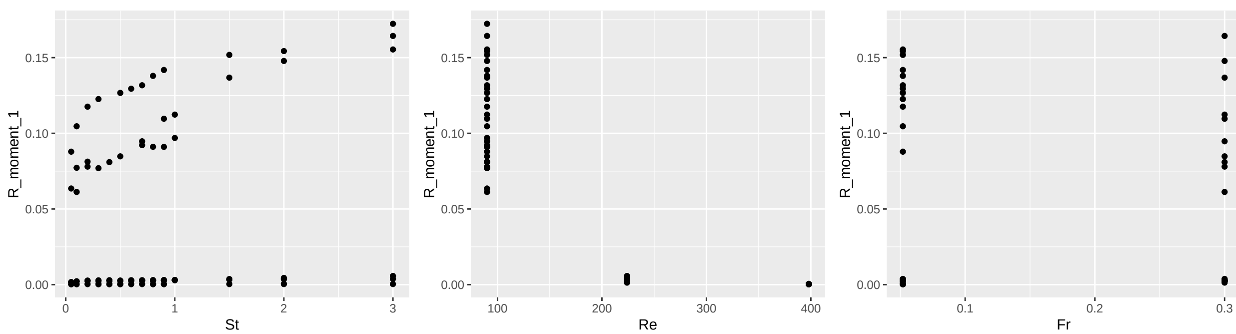


Figure 1: Exploratory data analysis plots on predictors vs first moment

approach was conducted with the addition of interaction terms, and the results were similar: low R-squared value and pattern in the residuals.

Referring back to the exploratory data analysis, we saw that the $Fr.logit$ and Re behaved more like categorical variables with only 3 unique data points for each. Therefore, for the next approach, we made Fr and Re as factors and created a model with both the simple linear regression and the addition of interaction terms. The factored model with interaction terms resulted in the best R-squared value (.982), a low RMSE (.0087), and a more reasonably random residuals plot. The improved summary, error, and model statistics while maintaining the interpretability of the model makes this factored linear regression with interaction terms model especially appealing.

To see if we can find a better model, we also conducted a Ridge Regression and Generalized Additive Model, but these resulted in very high MSE for higher moments.

For our chosen linear regression model, we can write this in statistical notation below:

$$Moment_x = \beta_0 + \beta_1 * St + \sum_{i=2}^{4} \beta_i * Re_i + \sum_{j=5}^{7} \beta_j * Fr.logit_j$$

$$+ \sum_{k=8}^{10} \beta_k * (\text{all two-way interactions between St, Re and Fr.logit})$$

$$\text{For } 1 \leq x \leq 4,$$

Considering the nature of our response variables, since each moment is derived from the same probability distribution, we decided to perform log transformations for the higher-order moments ($Moment_2$, $Moment_3$ and $Moment_4$). Furthermore, our design decision to include interaction effects stem from the exploratory data plot showing certain correlations between the predictor variables, so as to allow our model to be more interpretable to the collinearity between pair of predictors. It is important to note that the baseline level for the now categorical variable `Fr.logit` is 0.51, whereas the baseline level for the categorical variable `Re` is 90 in all of our models.

In order to validate our models and evaluate which performed the best out of all of the ones we tried in the Appendix, we performed 5-fold cross validation using the package `caret`. We interpreted a low RMSE and a high R-squared value to indicate that the model fit well to the data. We will talk more about these two values in the Results section for each model.

Results

The model's coefficients align with our prior beliefs. A larger particle size could increase the probability distribution for particle cluster volumes. Further, the coefficients of significance are `St`, `Re` and their interaction. For every one unit increase in `St`, we expect that the first moment will increase by 0.024. From this positive value we can conclude that the per probability distribution is increasing as we increase particle size. When `Fr`= 0.5744 and 1, there is a -0.035778 and -0.038679 increase in the first moment. When `Re` = 224 and 398, there is an 0.103 and -0.107 increase in the first moment. When `Fr`=0.5744, and `Re` = 224 there is an 0.029 increase in the first moment. When `Fr` = 1, and `Re` = 224 and 398, there is a 0.034 and 0.034 increase in the first moment. The R-squared value of 0.9892 shows the factors in the model successfully explain the most of the variance. The root mean square error (RMSE) is 0.0063 which is the lowest among the models, as can be expected since the others have been derived from this model.

It is important to note that there are NA's in the row representing the interaction between `Fr.logit` with a value of 0.574 and `Re` with a value of 398. This is because there are no combinations of these two categorical variables in the dataset, and as a result, there is no data to determine the coefficient for this interaction. However, we chose to keep this interaction because the other interactions between `Fr.logit` and `Re` are significant compared to the baseline.

We have plotted the model diagnostics for this model in the Appendix. Although the residuals seem to be clustered at the left end of the plot in the Residuals vs. Fitted plot, they do seem to be randomly distributed around the 0 line. Additionally, the QQ-plot seems to follow the diagonal line mostly well except at the ends, but this is not a huge concern. Finally, there seems to be 3 points of high leverage, which would be interesting to look into in future studies.

For the model implemented with the log of the second moment as the response variable, we see that all predictor variables and the interaction effects between `Fr` and `Re` are significant (p-values <0.05). A high adjusted R-squared value of 0.889 shows that the model explains a large amount of variability in its predictions considering multiple predictors and their collinearity. The RMSE calculated from 5-fold cross validation is 1.33 and the R-squared is 0.852, indicating that the model is fit well to the data. Holding all else constant, a unit increase in `St` results in a $e^{0.8586}$ multiplicative increase of the second moment. When `Fr` is 0.574 and 1, there is a $e^{-6.678}$ and $e^{-6.737}$ change of the second moment respectively. Similar when `Re` is 224 and 398, there is a $e^{-7.43451}$ and $e^{-10.7873}$ change of the second moment respectively. Additionally, when `Fr` is 0.574 and `Re` is 224 and 398, there is a $e^{4.477}$ change in the second moment. When `Fr` is 1, and `Re` is 224 and 398, there an additional $e^{4.694}$ and $e^{6.883}$ change in the second moment.

For the third moment, we can see that 4 of the interaction terms, the different combinations of `St:Fr.logit` and `St:Re`, are no longer significant compared to their baseline levels, although their main effects are still significant at an alpha level of 0.05. Fortunately, a high adjusted R-squared value of 0.875 shows that the model explains a large amount of variability in its predictions considering multiple predictors and their collinearity. Overall, we can see that individually, `St` has a positive correlation with the third moment, and `Fr.logit` and `Re` have negative correlations with the third moment as they increases from their baseline levels. More specifically, when holding all else constant and given `Fr.logit` and `Re` are at their respective baseline levels of 0.51 and 90, we can see that increasing `St` by one unit will increase the third moment by a multiplicative factor of $e^{0.022}$, or by 1.022 times. We can also see that when `Fr.logit` is at the baseline level of 0.51, `St` is 1, and `Re` increases from its baseline level of 90 to 224, the third moment increases by a multiplicative factor of $e^{-11.09-0.07}$, or by 0.000014 times. Finally, when `Re` is at the baseline level of 90, `St` is 1, and `Fr.logit` increases from its baseline level of 0.51 to 0.57, the third moment increases by a multiplicative factor of $e^{-12.81+0.299}$, or by 0.0000037 times. The RMSE of this model using 5-fold cross validation is 2.13, whereas the R-squared is 0.86, both of which indicate that we are fitting the training data relatively well.

For the fourth moment, we can similarly see that the interaction terms including `St` are no longer significant, while their main effects are still significant. The residual standard error is 2.693, which is higher than other models, but is expected since moment 4 is a higher-order derivation of the probability distribution compared to moments 1, 2 and 3. However, a high adjusted R-squared value of 0.8784 tells us that the model explains much of the variability in the predictive capabilities of the model, but is significant lower than the other models with lower-order moments. Holding all else constant, a unit increase in `St` causes a $e^{1.7147}$ increase in the 4th moment. When `Fr`$=$ 0.5744 and 1, there is an $e^{0.3206}$ and $e^{0.0065}$ increase in the 4th moment respectively. When `Re` $=$ 224 and 398, there is an $e^{0.1384}$ and $e^{-1.3817}$ increase in the 4th moment respectively. Additionally, when `Fr`=0.5744, and `Re` $=$ 224, there is an additional $e^{12.435}$ increase in the 4th moment. When `Fr` $=$ 1, and `Re` $=$ 224 and 398, there is an additional $e^{12.719}$ and $e^{19.134}$ increase in the 4th moment respectively. The RMSE of this model using 5-fold cross validation is 2.94, whereas the R-squared is 0.85, both of which indicate that we are fitting the training data relatively well.

Conclusion

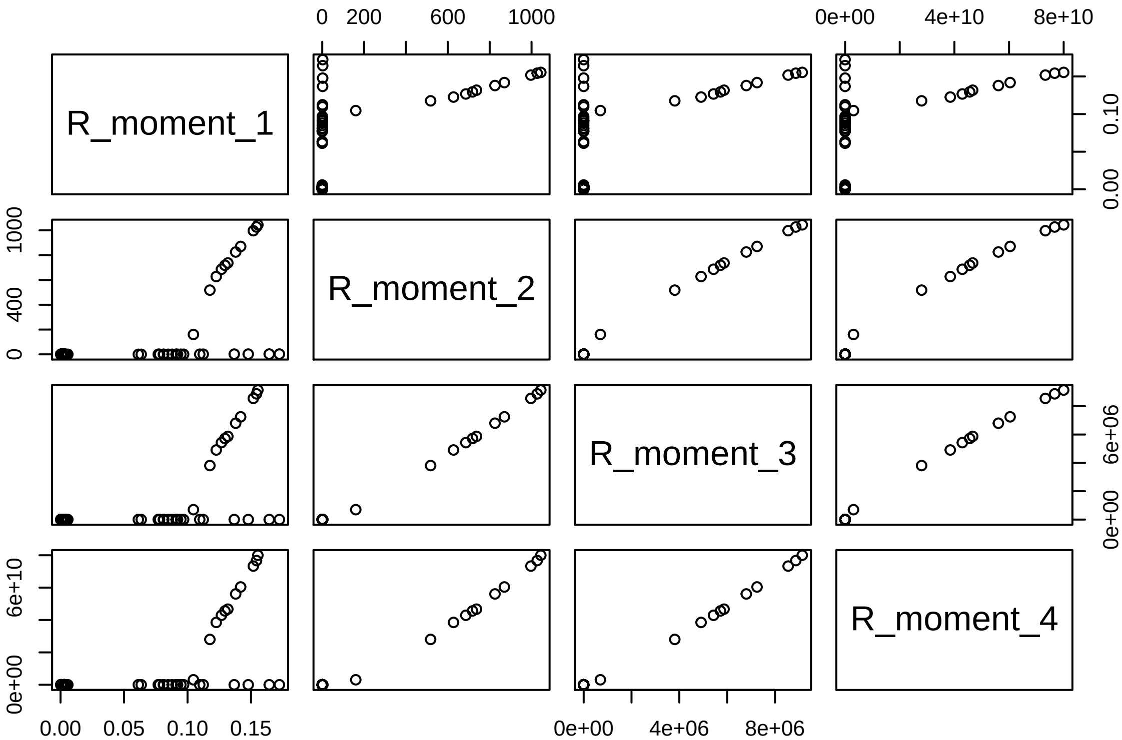
Using linear regression model, we see that there is a good fit with the data after treating `Fr` and `Re` as factors, as well as considering log transformations for higher-order moments. There is however, a decrease in the adjusted R-squared values as models are fit with higher-order moments, signifying that using a linear model to predict moments of a probability distribution decreases as the order increases. We can also conclude that `St`, `Re` and `Fr` are all significant in predicting moments up to the 4th derivative, along with their interaction effects with the exception of `St`.

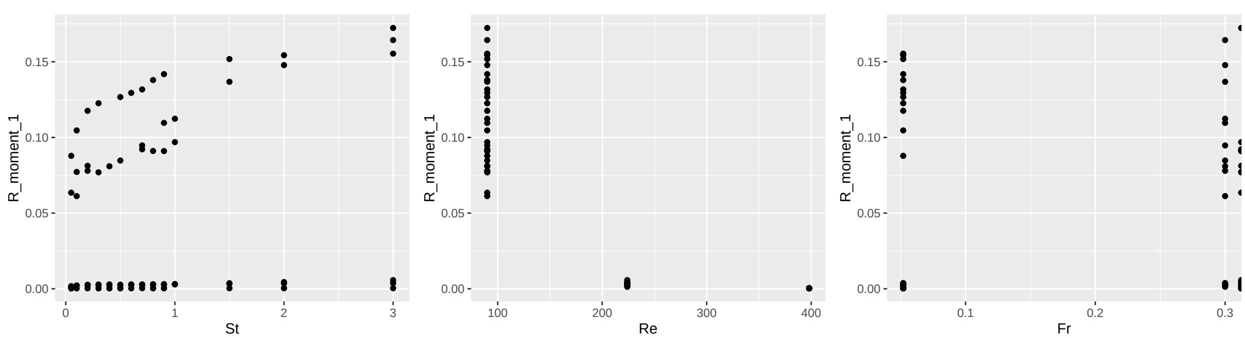
There are high levels of uncertainty with our models, as seen from the confidence intervals (seen in the appendix). For the majority of the predictors and their interaction effects, we see a wide range in their 95% confidence intervals, which we can interpret as the model having high degrees of predictive uncertainty. This can be attributed firstly to the limited number of observations in the data set, as well as the derivative nature of the response variables.

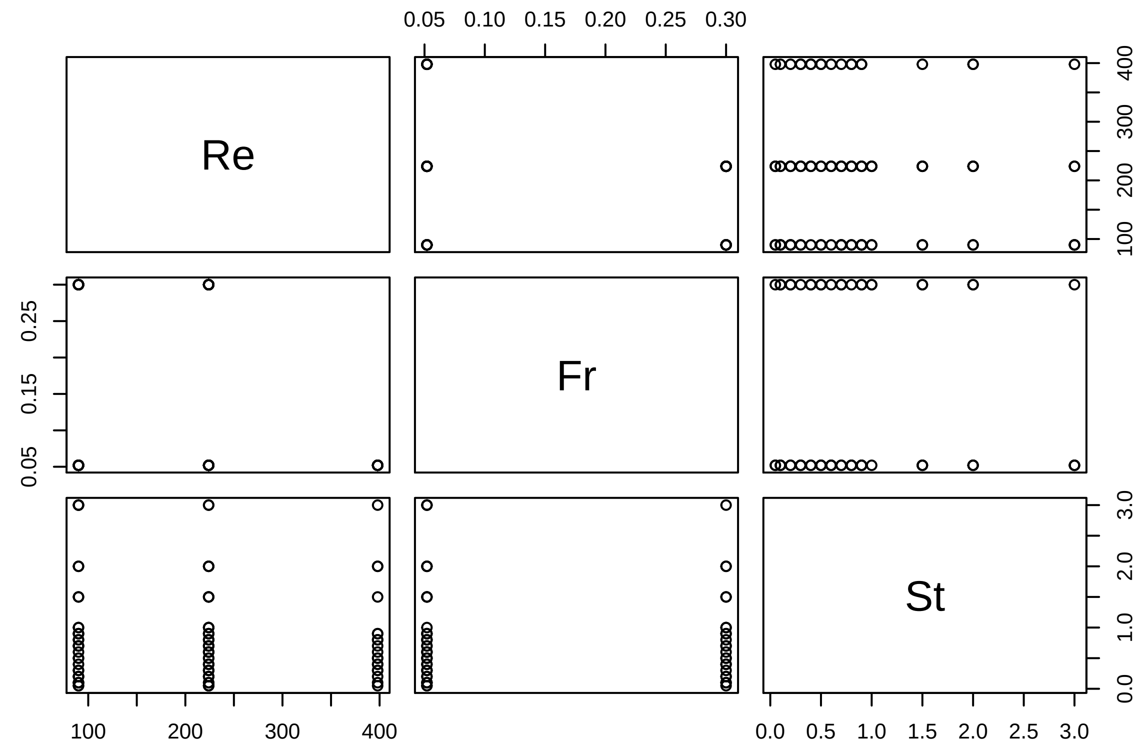
We attempted other models (ridge regression and regressions with higher-polynomials), but concluded that they did not fit as well as the linear regression with interactions and log transformations of the response variables. Ultimately, considering the nature of the study, the types of predictors and the response variables, along with the limited number of observations in the data set, using a linear model with acute modeling decisions was the best fit for the study and provided the best predictive accuracies overall. It is important to note that this study does not include interpretations on particle clustering probability distributions nor turbulence, but the moments that are derived from the distribution. Further manipulations are required to obtain interpretative results for these features.

Appendix

EDA



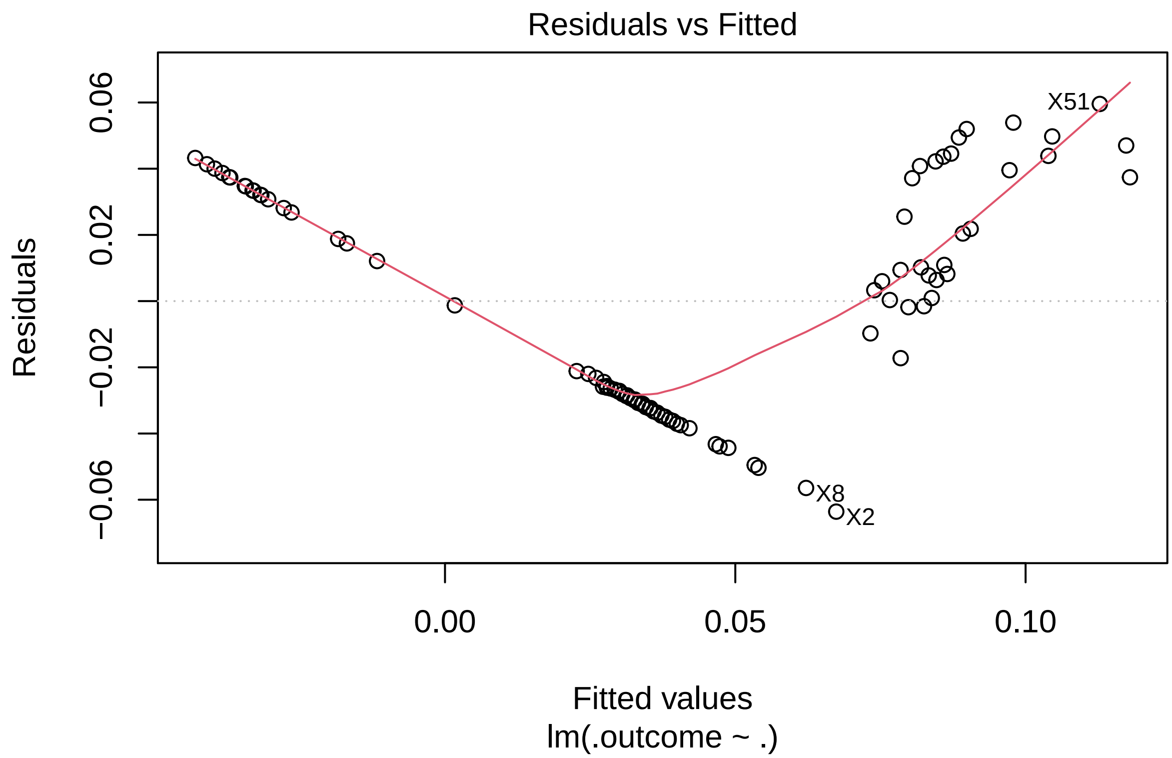


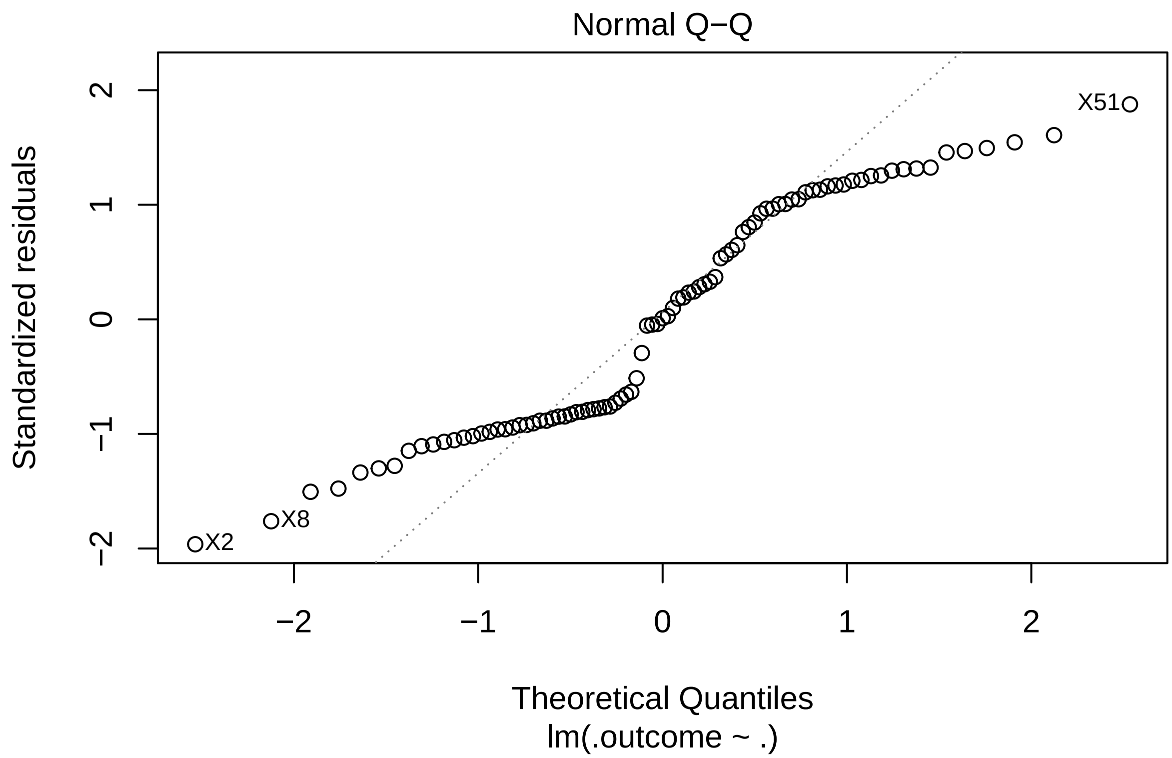


Modeling

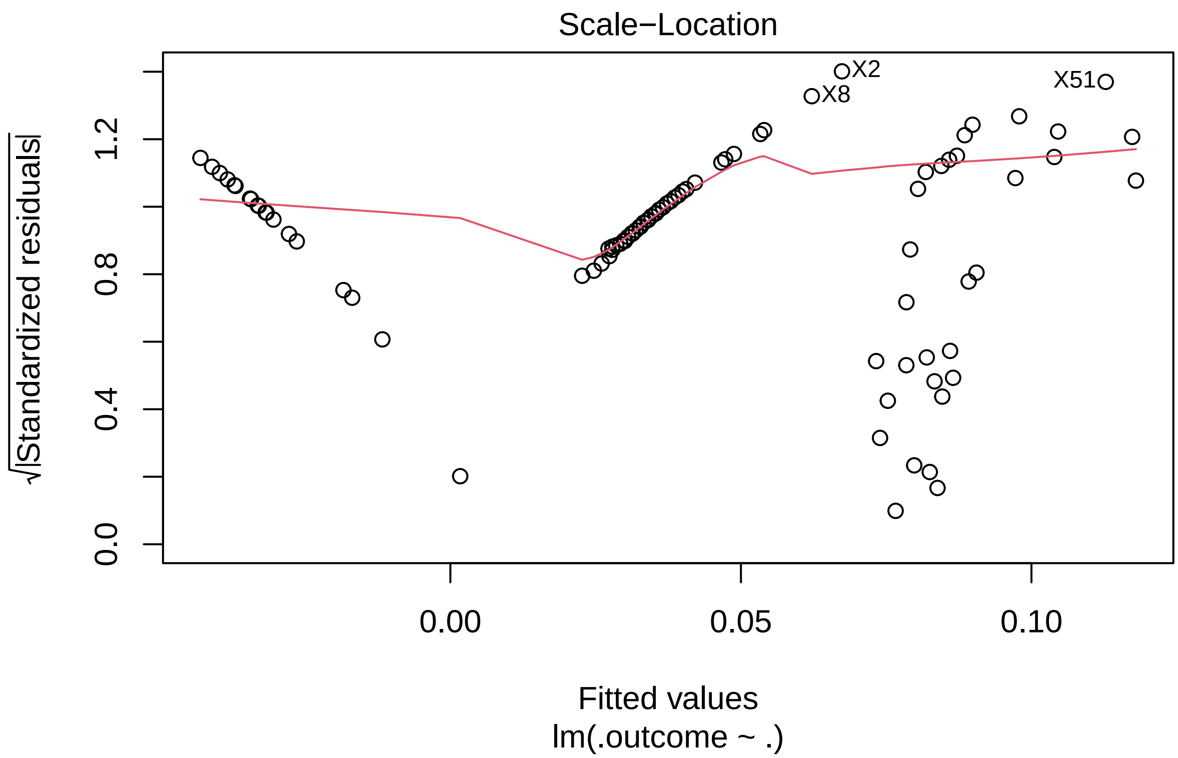
Simple linear modeling

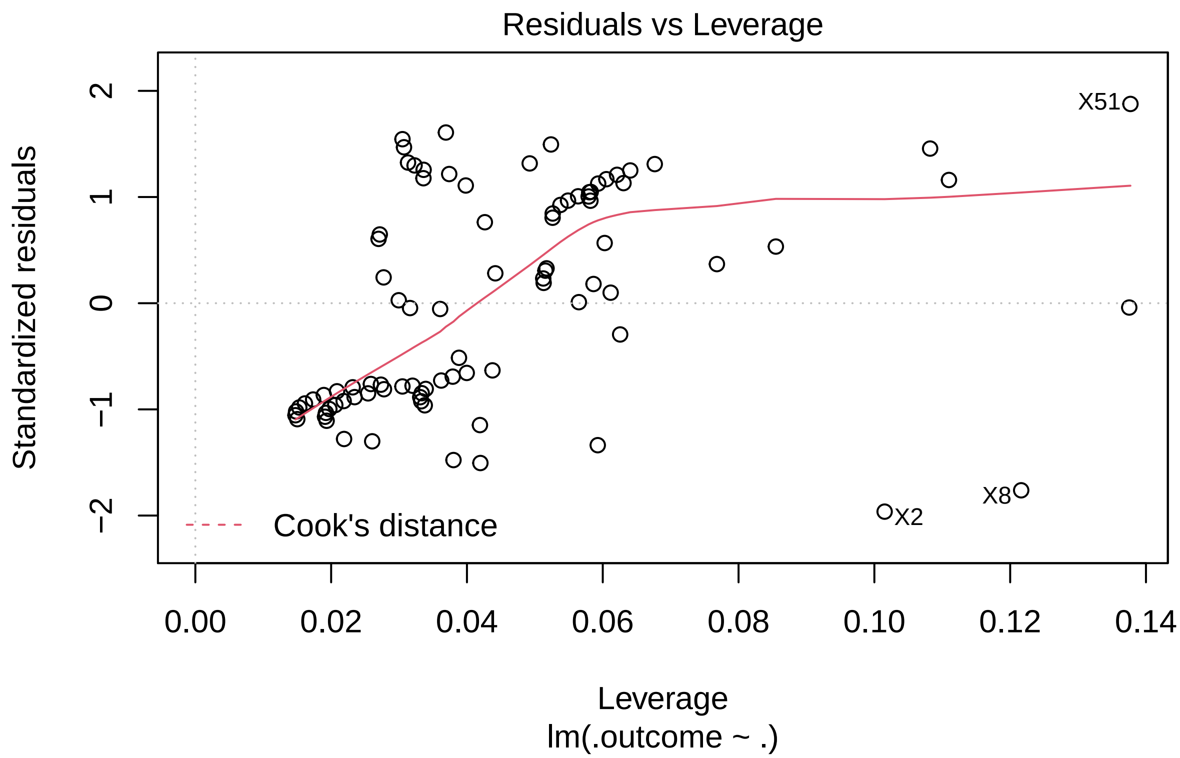
```
## Linear Regression
##
## 89 samples
##  3 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 72, 73, 70, 70, 71
## Resampling results:
##
##   RMSE        Rsquared   MAE
##   0.03413316  0.6495622  0.03077955
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```


Residuals vs Fitted
Residuals
0.06
0.02
−0.02
−0.06
X51
X8
X2
0.00
0.05
0.10
Fitted values
lm(.outcome ~ .)



Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(.outcome ~ .)
X51
X8
X2
−2
−1
0
1
2


Scale−Location
√|Standardized residuals|
0.0
0.4
0.8
1.2
0.00
0.05
0.10
X2
X8
X51
Fitted values
lm(.outcome ~ .)



Interaction effects

```
## [1] 0.0009855535
```

```
##
## Call:
## lm(formula = .outcome ~ ., data = dat)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.064305 -0.028656  0.006188  0.026969  0.056868
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.346e-01  2.947e-02   4.568 1.72e-05 ***
## St               4.392e-02  1.656e-02   2.653  0.00958 **
## Fr.logit        -6.140e-02  3.881e-02  -1.582  0.11750
## Re              -4.801e-04  1.115e-04  -4.305 4.60e-05 ***
## `St:Fr.logit` -1.434e-02  1.997e-02  -0.718  0.47469
## `St:Re`       -1.008e-04  3.913e-05  -2.577  0.01177 *
## `Fr.logit:Re`  2.660e-04  1.351e-04   1.968  0.05242 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0323 on 82 degrees of freedom
```

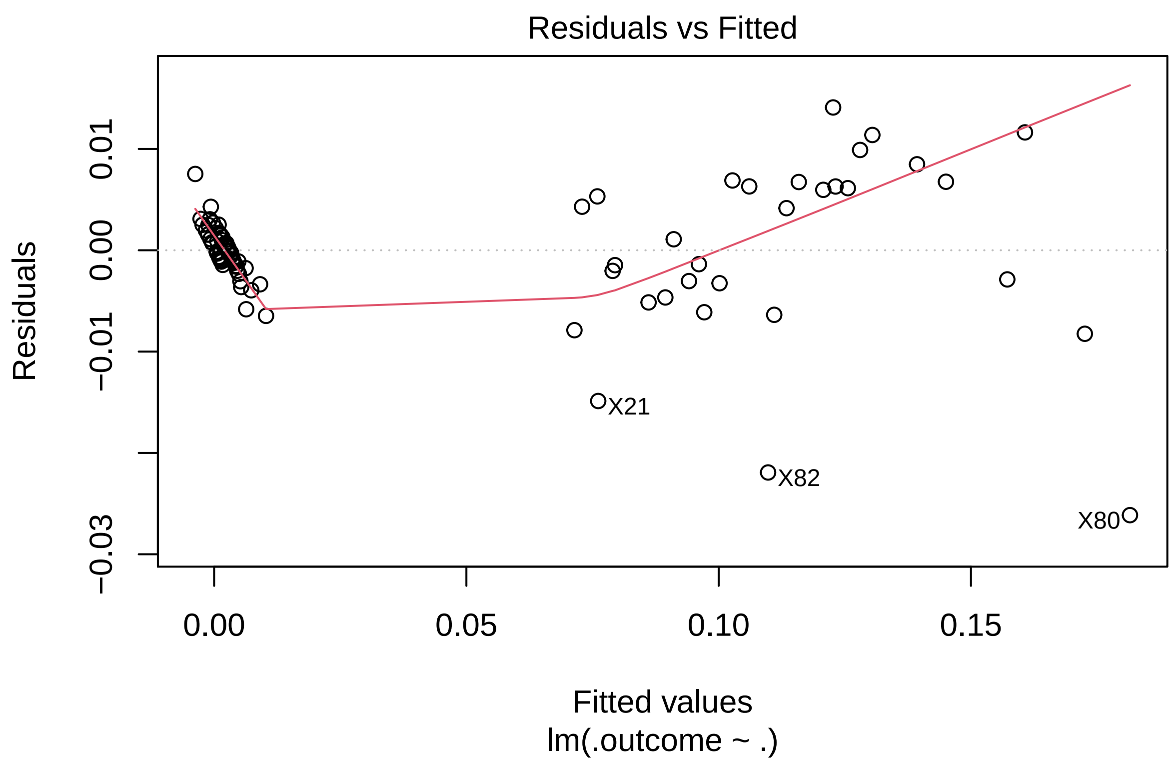
```
## Multiple R-squared:  0.6883, Adjusted R-squared:  0.6655
## F-statistic: 30.18 on 6 and 82 DF,  p-value: < 2.2e-16
```

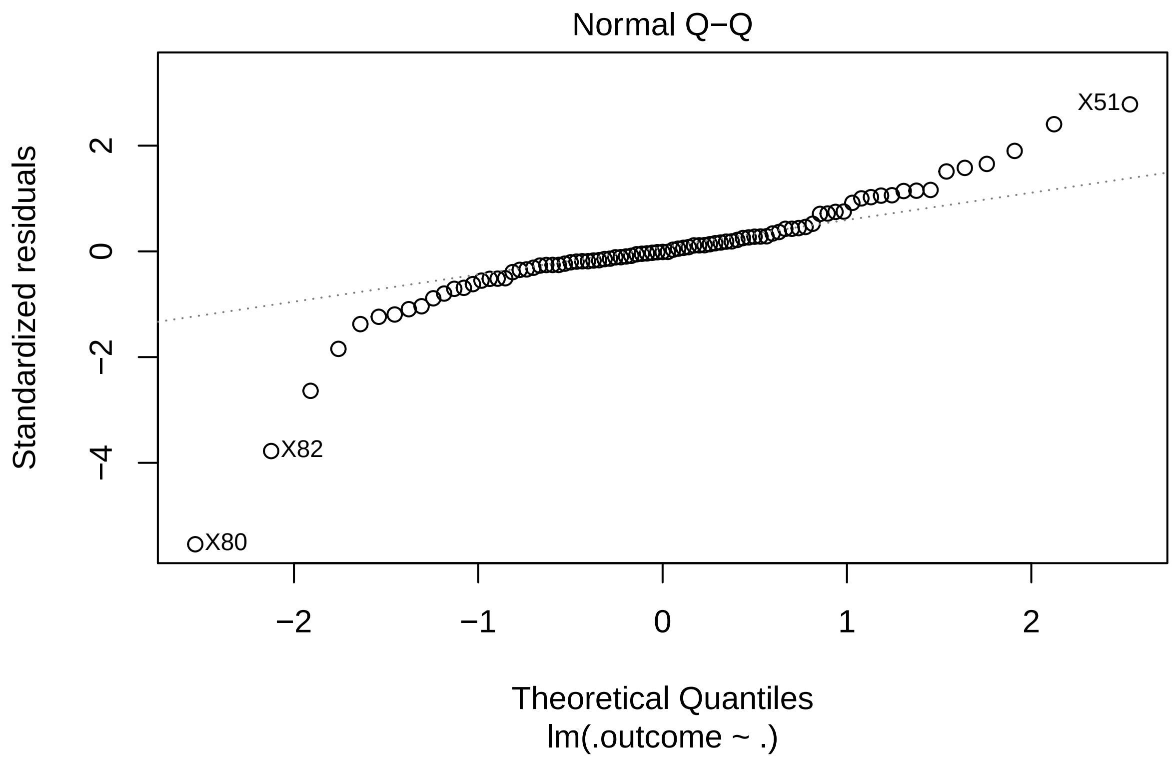
Predictors as factors

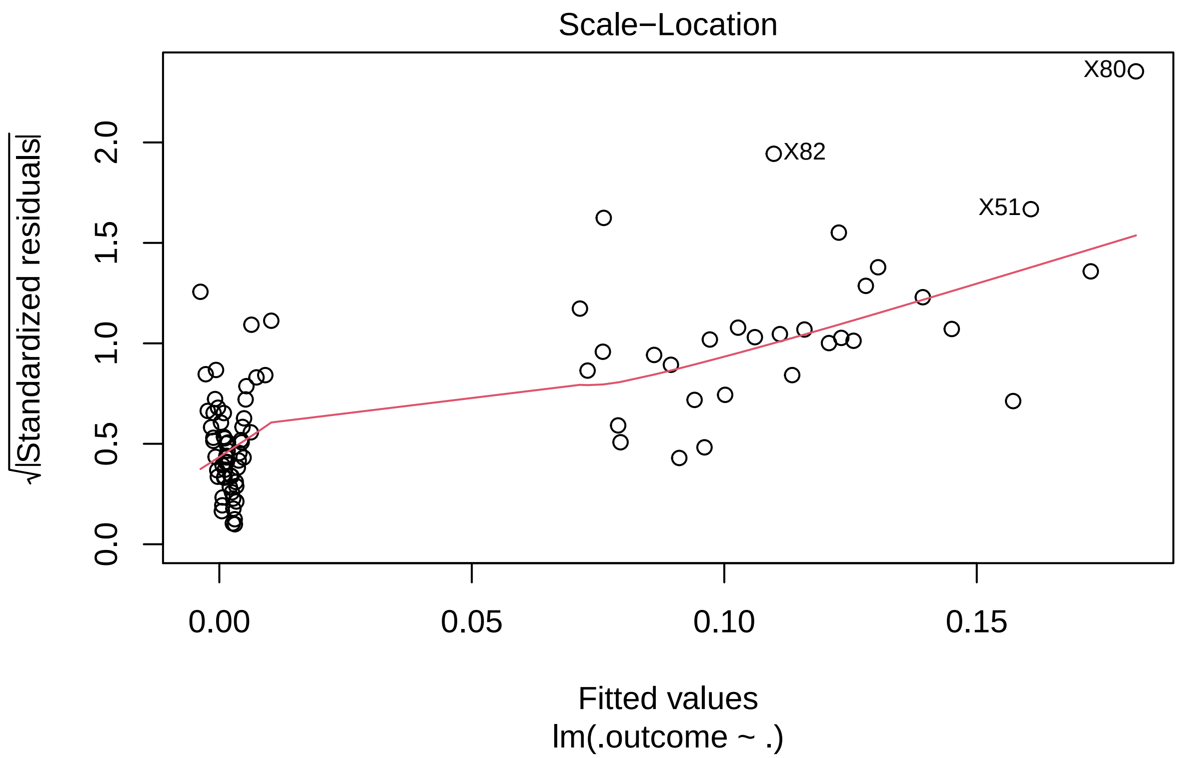
```
## [1] 0.0001623039
```

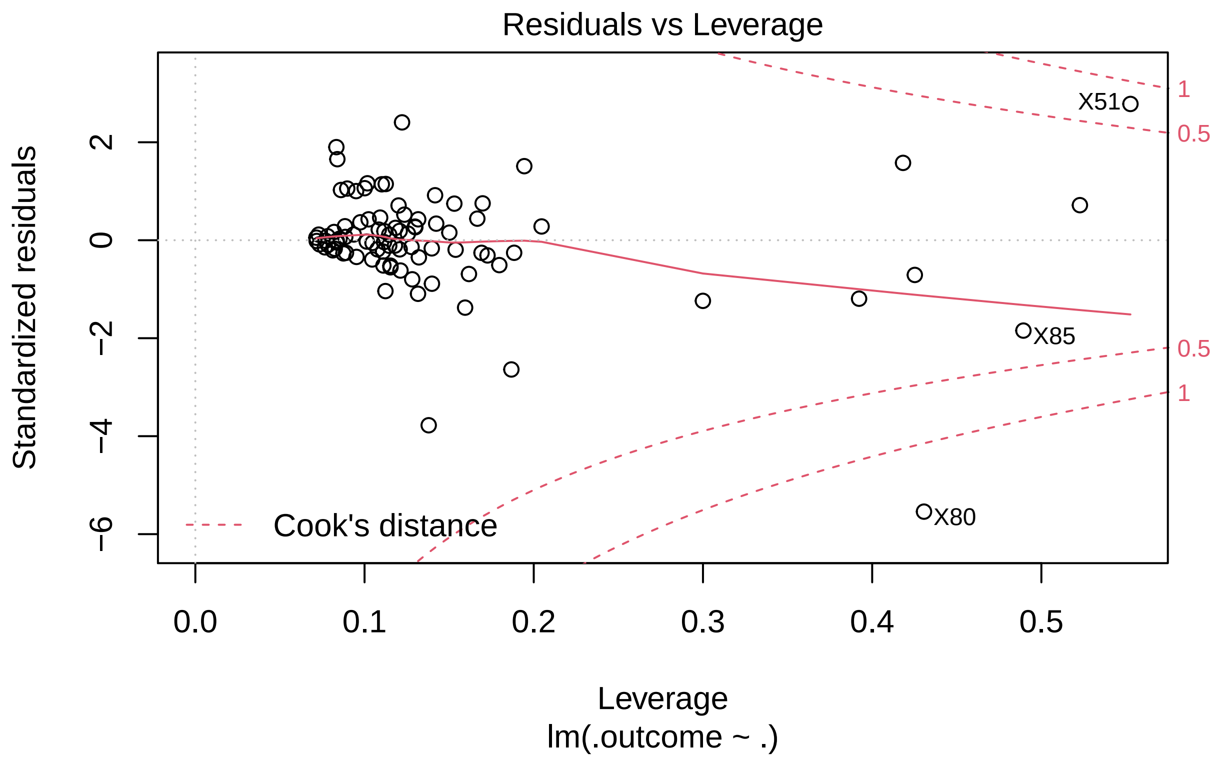
Final Models

```
## Linear Regression
##
## 89 samples
##  3 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 71, 72, 71, 70, 72
## Resampling results:
##
##   RMSE         Rsquared   MAE
##   0.008673306  0.9820262  0.004945823
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(.outcome ~ .)
X51
X82
X80
2
0
−2
−4
−2
−1
0
1
2

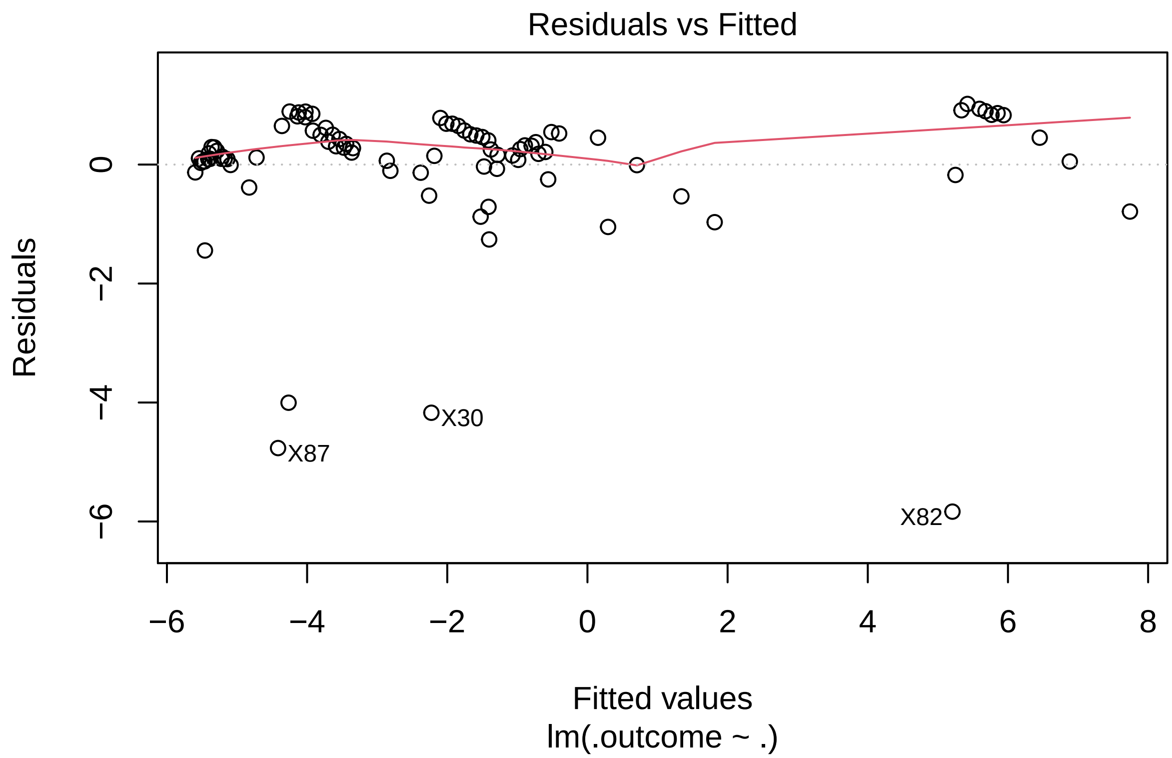

Scale−Location
X80
X82
X51
2.0
1.5
1.0
0.5
0.0
√|Standardized residuals|
0.00
0.05
0.10
0.15
Fitted values
lm(.outcome ~ .)



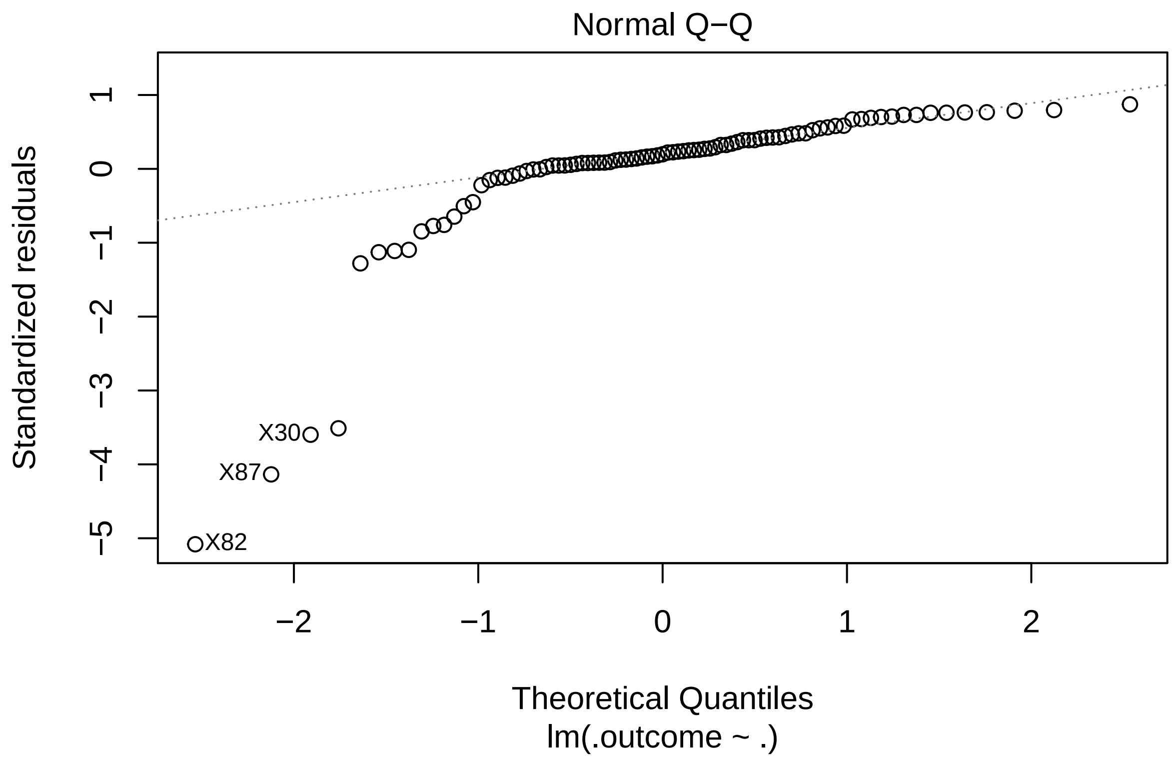
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.1085805	0.0023779	45.662236	0.0000000
St	0.0243128	0.0017443	13.938697	0.0000000
Fr.logit0.574442516811659	-0.0357783	0.0035405	-10.105471	0.0000000
Fr.logit1	-0.0386790	0.0032193	-12.014556	0.0000000
Re224	-0.1030601	0.0029874	-34.498397	0.0000000
Re398	-0.1067095	0.0036031	-29.615831	0.0000000
‘St:Fr.logit0.574442516811659‘	0.0089440	0.0023797	3.758539	0.0003334
‘St:Fr.logit1‘	0.0059558	0.0019952	2.985061	0.0038125
‘St:Re224‘	-0.0273996	0.0019383	-14.135990	0.0000000
‘St:Re398‘	-0.0258336	0.0025060	-10.308724	0.0000000
‘Fr.logit0.574442516811659:Re224‘	0.0288311	0.0036570	7.883737	0.0000000
‘Fr.logit1:Re224‘	0.0336571	0.0037050	9.084172	0.0000000
‘Fr.logit1:Re398‘	0.0342939	0.0040512	8.465154	0.0000000

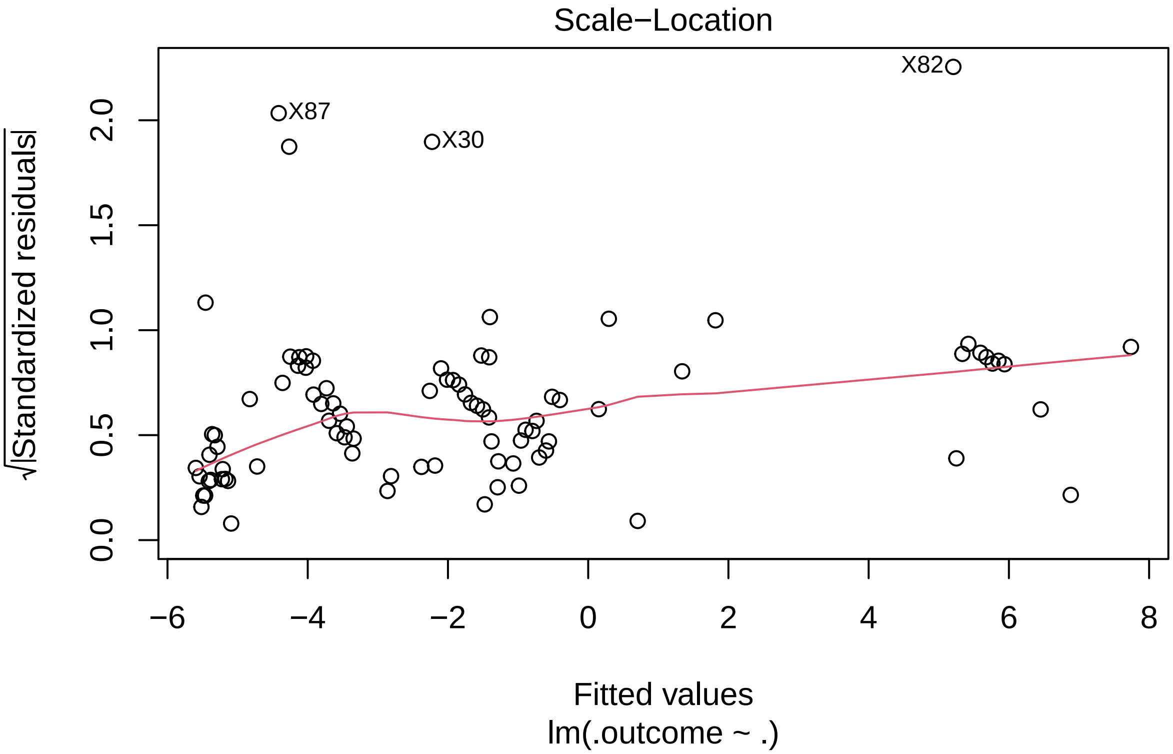
	2.5 %	97.5 %
(Intercept)	0.1038445	0.1133165
St	0.0208388	0.0277868
Fr.logit0.574442516811659	-0.0428298	-0.0287268
Fr.logit1	-0.0450908	-0.0322671
Re224	-0.1090100	-0.0971102
Re398	-0.1138858	-0.0995333
‘St:Fr.logit0.574442516811659‘	0.0042045	0.0136835
‘St:Fr.logit1‘	0.0019820	0.0099296
‘St:Re224‘	-0.0312600	-0.0235391
‘St:Re398‘	-0.0308247	-0.0208425
‘Fr.logit0.574442516811659:Re224‘	0.0215475	0.0361147
‘Fr.logit1:Re224‘	0.0262779	0.0410364
‘Fr.logit0.574442516811659:Re398‘	NA	NA
‘Fr.logit1:Re398‘	0.0262253	0.0423625

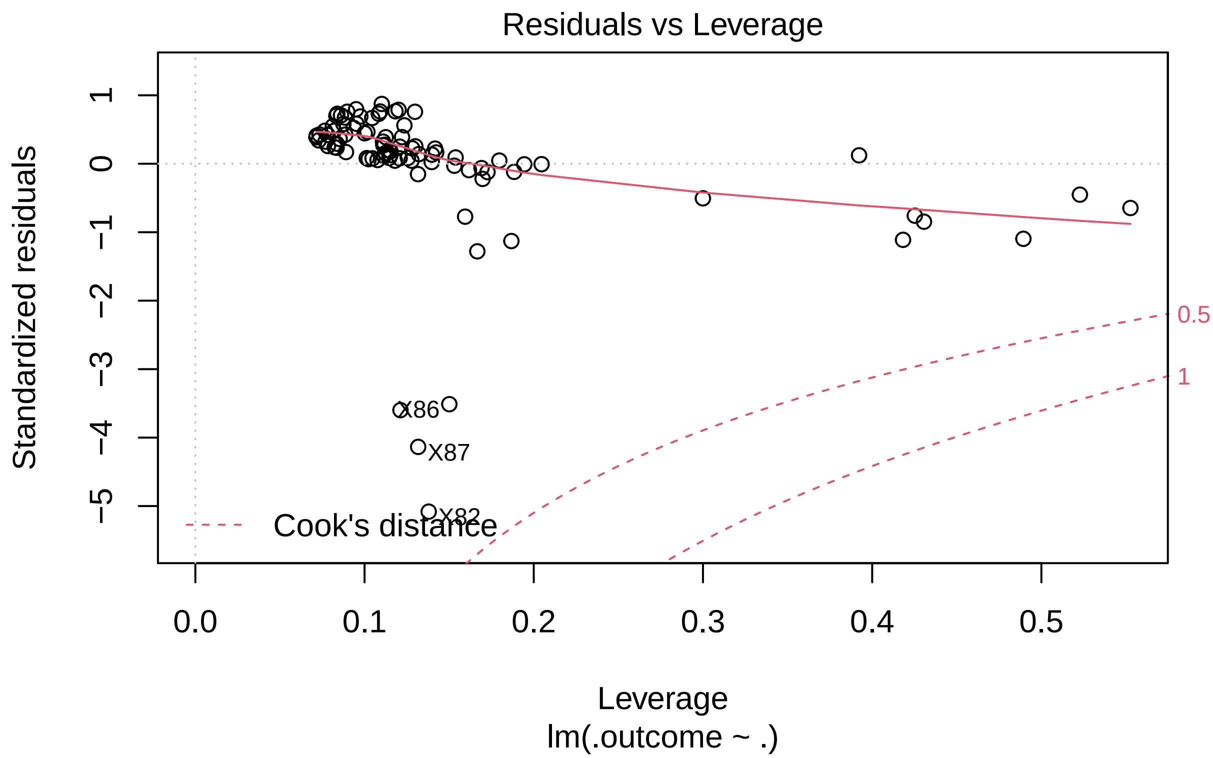
```
## Linear Regression
##
## 89 samples
##  3 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 69, 72, 72, 71, 72
## Resampling results:
##
##   RMSE      Rsquared   MAE
##   1.236948  0.8834658  0.7745262
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```



Residuals vs Fitted
Residuals
0
−2
−4
−6
X87
X30
X82
−6
−4
−2
0
2
4
6
8
Fitted values
lm(.outcome ~ .)


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(.outcome ~ .)
X30
X87
X82
1
0
−1
−2
−3
−4
−5
−2
−1
0
1
2

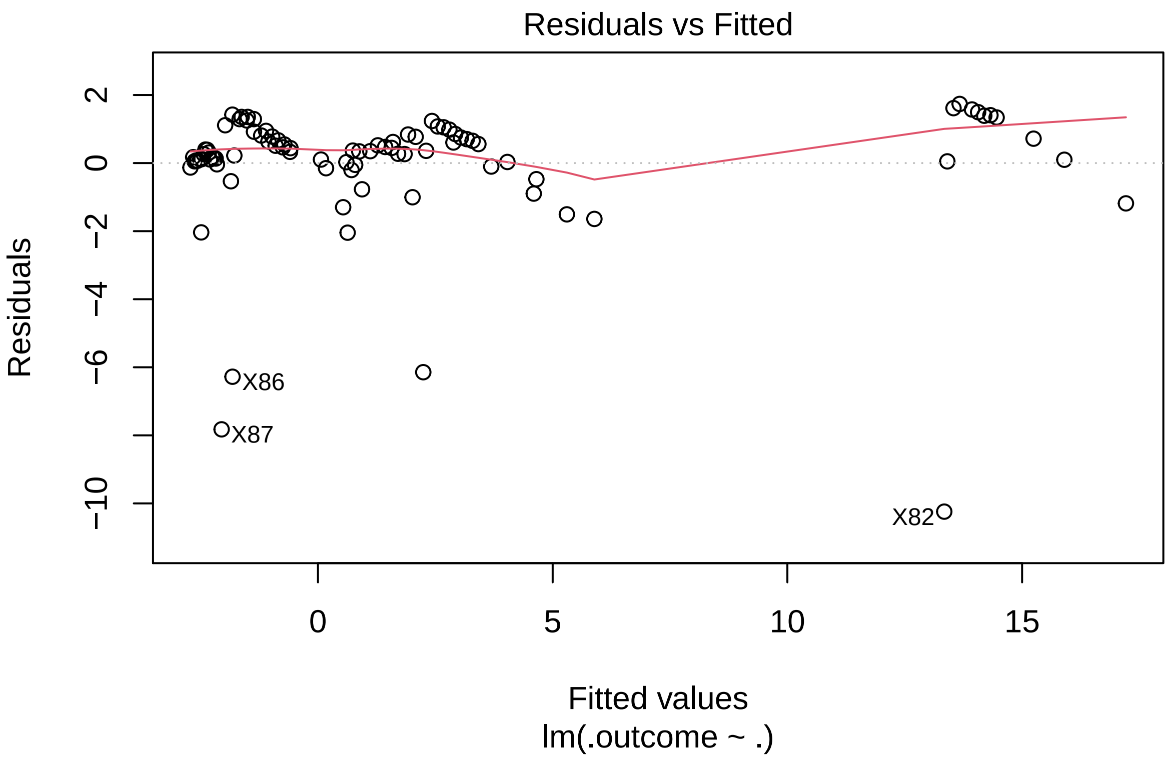

Scale−Location
X87
X30
X82
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
−6
−4
−2
0
2
4
6
8
Fitted values
lm(.outcome ~ .)

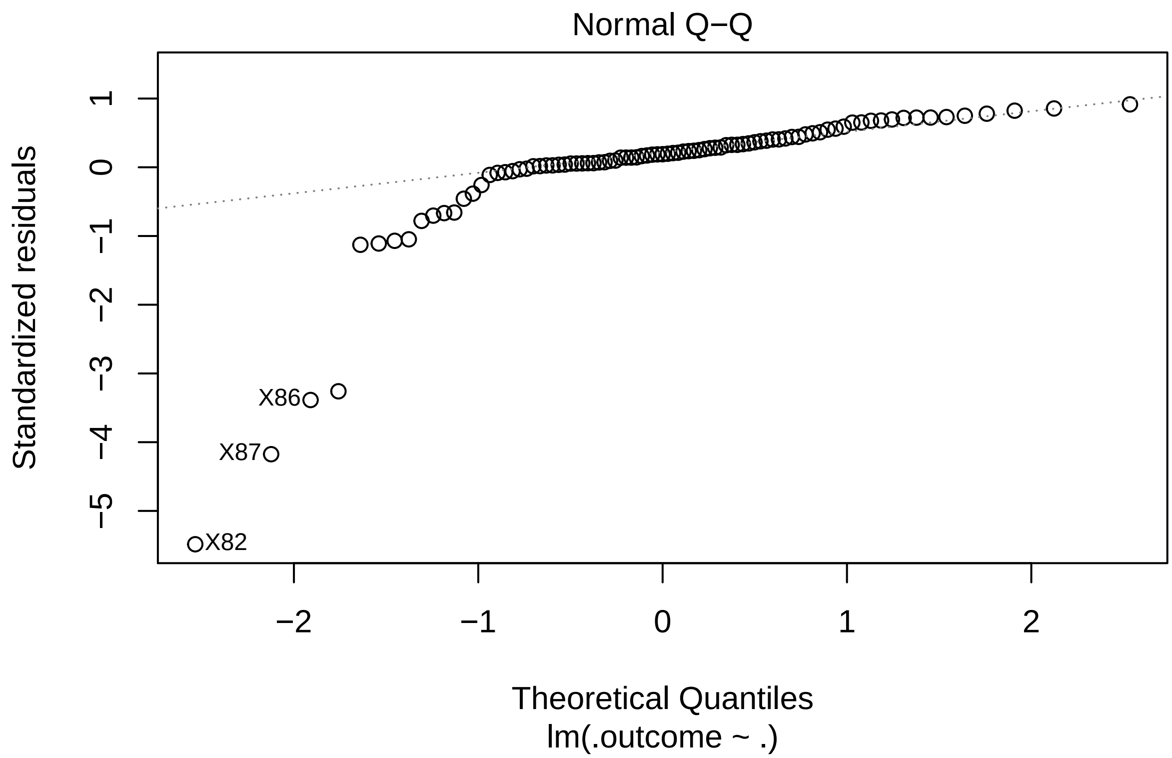


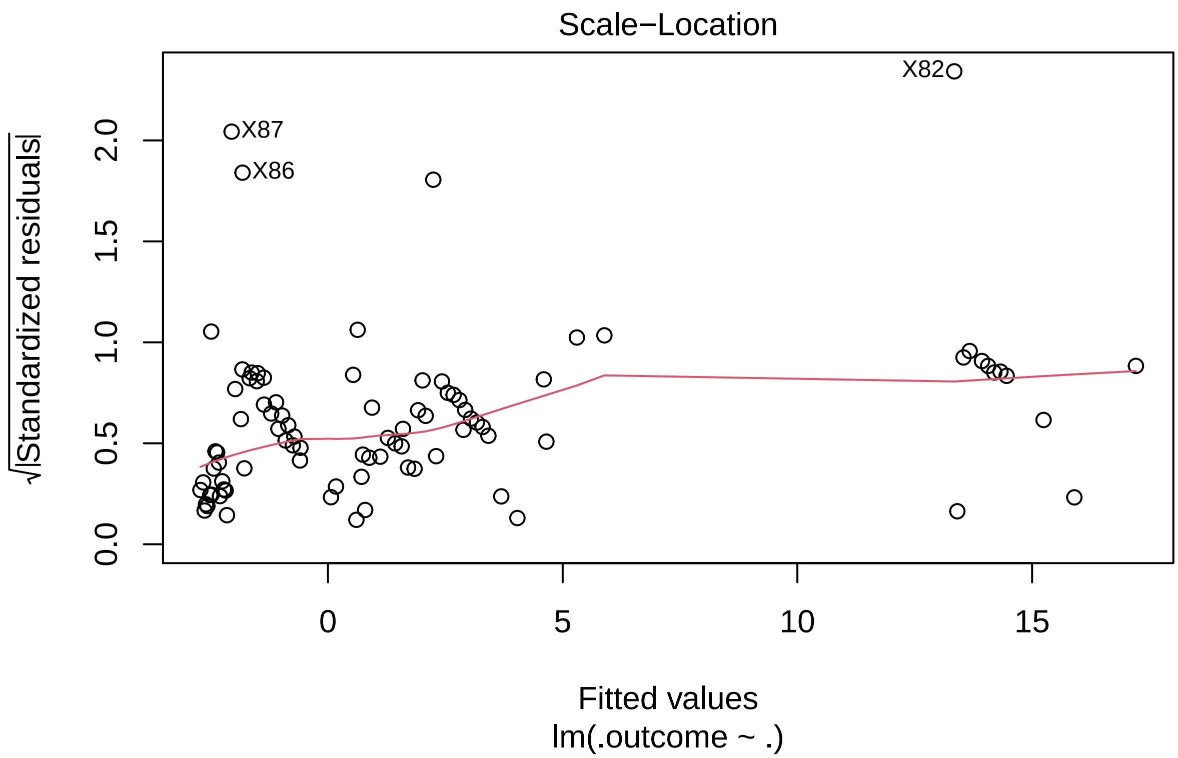
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.1649892	0.4703068	10.9821705	0.0000000
St	0.8586953	0.3449847	2.4890826	0.0149943
Fr.logit0.574442516811659	-6.6781466	0.7002444	-9.5368803	0.0000000
Fr.logit1	-6.7377944	0.6367273	-10.5819149	0.0000000
Re224	-7.4345116	0.5908513	-12.5827109	0.0000000
Re398	-10.7873786	0.7126326	-15.1373643	0.0000000
‘St:Fr.logit0.574442516811659‘	0.2507829	0.4706531	0.5328402	0.5956988
‘St:Fr.logit1‘	0.1123923	0.3946148	0.2848152	0.7765604
‘St:Re224‘	-0.0040912	0.3833572	-0.0106721	0.9915130
‘St:Re398‘	-0.5934658	0.4956405	-1.1973715	0.2348831
‘Fr.logit0.574442516811659:Re224‘	4.4777949	0.7232946	6.1908315	0.0000000
‘Fr.logit1:Re224‘	4.6944332	0.7327880	6.4062638	0.0000000
‘Fr.logit1:Re398‘	6.8834362	0.8012510	8.5908610	0.0000000

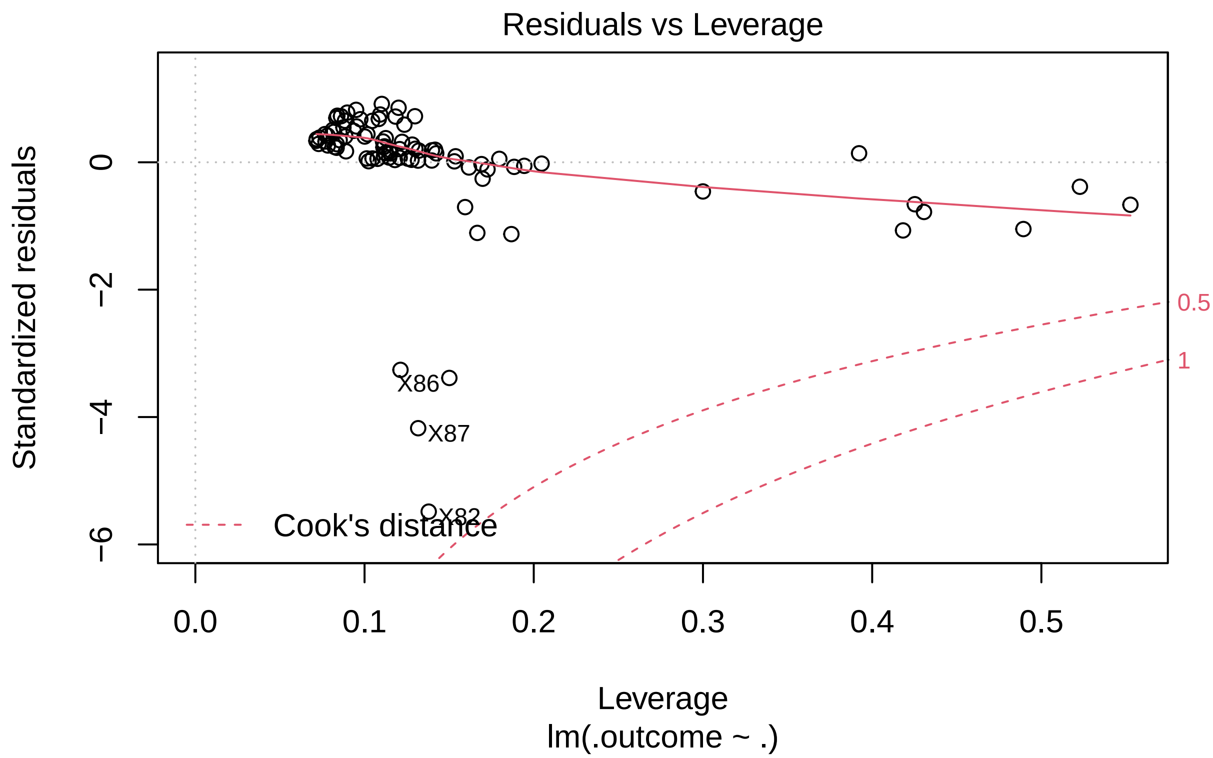
	2.5 %	97.5 %
(Intercept)	4.2282921	6.1016863
St	0.1715988	1.5457918
Fr.logit0.574442516811659	-8.0728041	-5.2834891
Fr.logit1	-8.0059468	-5.4696420
Re224	-8.6112940	-6.2577291
Re398	-12.2067093	-9.3680478
‘St:Fr.logit0.574442516811659‘	-0.6866040	1.1881697
‘St:Fr.logit1‘	-0.6735511	0.8983357
‘St:Re224‘	-0.7676133	0.7594309
‘St:Re398‘	-1.5806194	0.3936878
‘Fr.logit0.574442516811659:Re224‘	3.0372289	5.9183609
‘Fr.logit1:Re224‘	3.2349594	6.1539070
‘Fr.logit0.574442516811659:Re398‘	NA	NA
‘Fr.logit1:Re398‘	5.2876064	8.4792659

```
## Linear Regression
##
## 89 samples
##  3 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 70, 71, 72, 72, 71
## Resampling results:
##
##   RMSE      Rsquared   MAE
##   2.130962  0.8143657  1.281979
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```


Residuals vs Fitted
Residuals
2
0
−2
−4
−6
−10
X86
X87
X82
0
5
10
15
Fitted values
lm(.outcome ~ .)


Normal Q−Q
Standardized residuals
1
0
−1
−2
−3
−4
−5
X86
X87
X82
−2
−1
0
1
2
Theoretical Quantiles
lm(.outcome ~ .)

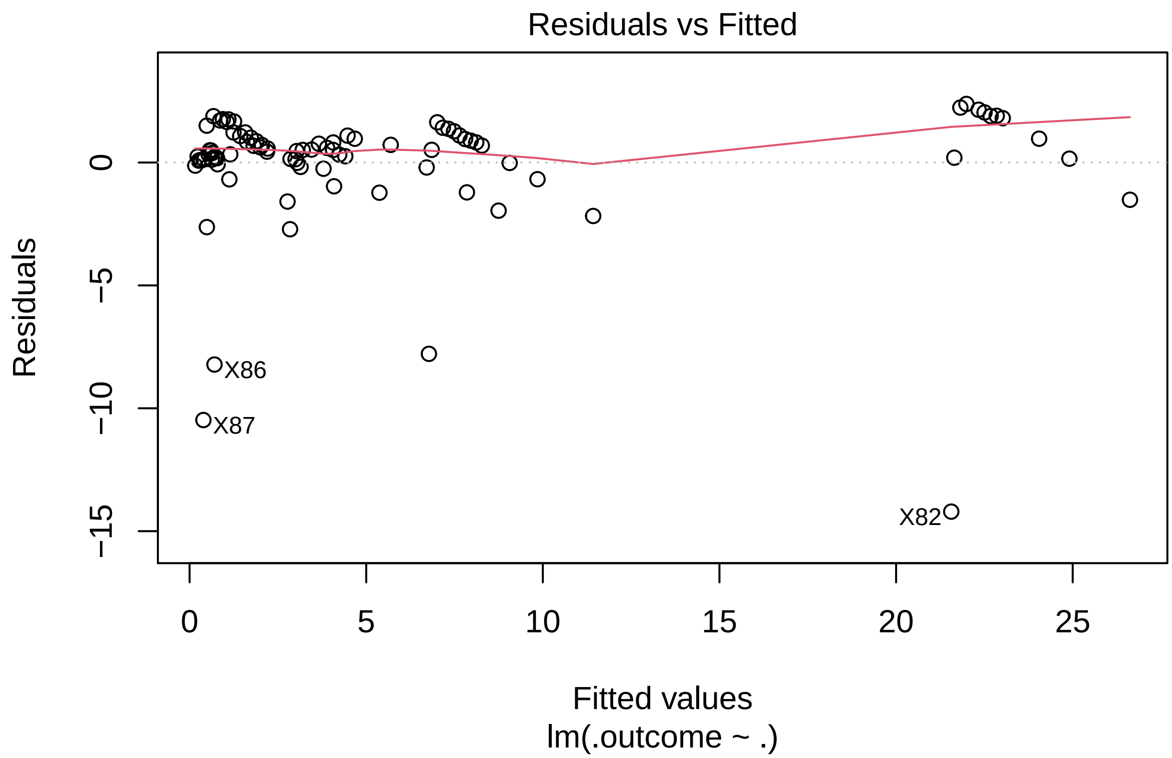

Scale−Location
X82
X87
X86
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
0
5
10
15
Fitted values
lm(.outcome ~ .)



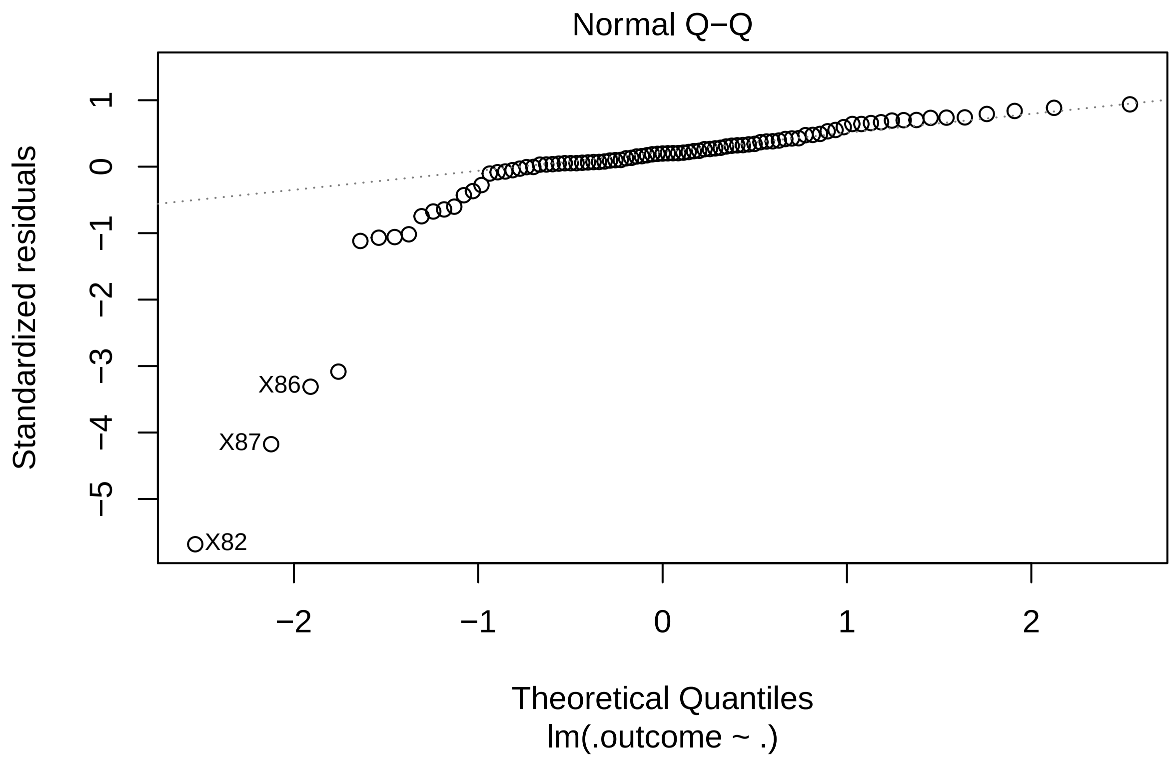
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	13.2762648	0.7647925	17.3593020	0.0000000
St	1.3118787	0.5609991	2.3384683	0.0219944
Fr.logit0.574442516811659	-12.8053561	1.1387071	-11.2455224	0.0000000
Fr.logit1	-12.8079364	1.0354185	-12.3698164	0.0000000
Re224	-11.0943403	0.9608169	-11.5467788	0.0000000
Re398	-16.0225724	1.1588523	-13.8262426	0.0000000
‘St:Fr.logit0.574442516811659‘	0.2985980	0.7653557	0.3901428	0.6975223
‘St:Fr.logit1‘	0.0643493	0.6417055	0.1002786	0.9203872
‘St:Re224‘	-0.0761679	0.6233990	-0.1221816	0.9030777
‘St:Re398‘	-1.0143783	0.8059891	-1.2585508	0.2120451
‘Fr.logit0.574442516811659:Re224‘	8.4942612	1.1761904	7.2218420	0.0000000
‘Fr.logit1:Re224‘	8.7407086	1.1916282	7.3350973	0.0000000
‘Fr.logit1:Re398‘	13.0493354	1.3029598	10.0151483	0.0000000

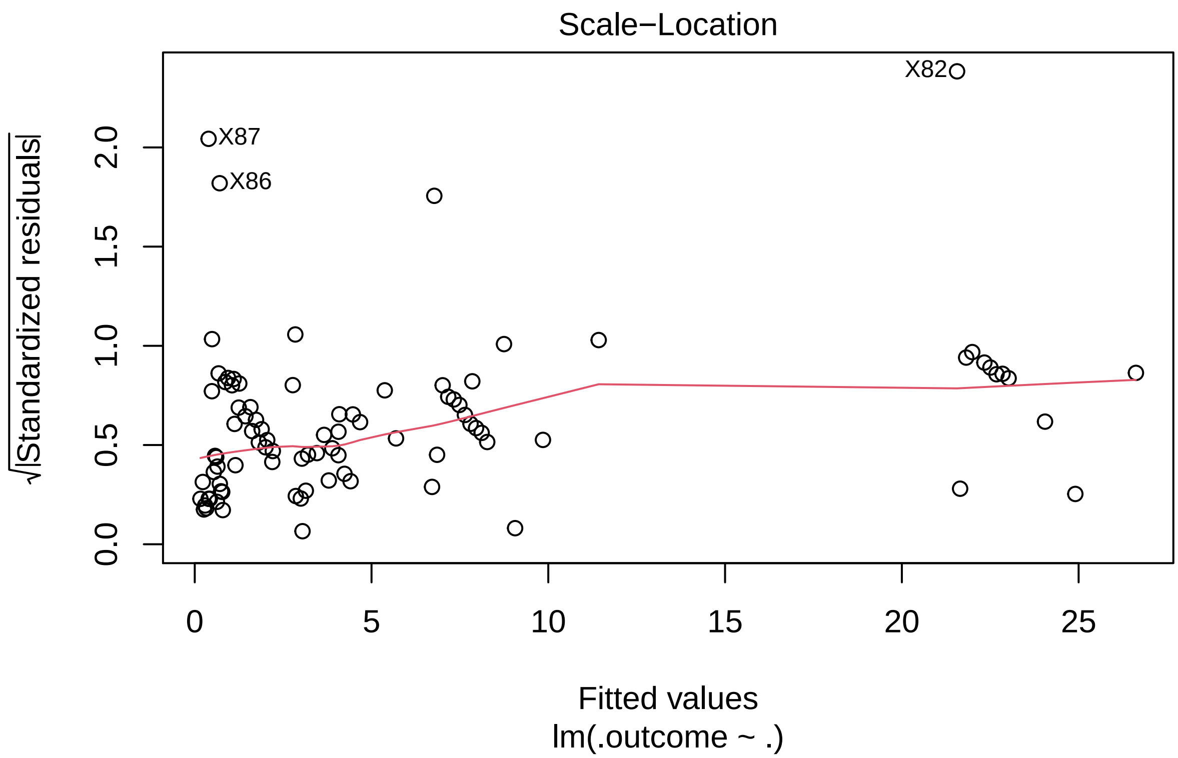
	2.5 %	97.5 %
(Intercept)	11.7530484	14.7994811
St	0.1945521	2.4292053
Fr.logit0.574442516811659	-15.0732878	-10.5374244
Fr.logit1	-14.8701510	-10.7457218
Re224	-13.0079730	-9.1807075
Re398	-18.3306267	-13.7145181
‘St:Fr.logit0.574442516811659‘	-1.2257400	1.8229360
‘St:Fr.logit1‘	-1.2137179	1.3424165
‘St:Re224‘	-1.3177745	1.1654388
‘St:Re398‘	-2.6196448	0.5908882
‘Fr.logit0.574442516811659:Re224‘	6.1516750	10.8368474
‘Fr.logit1:Re224‘	6.3673754	11.1140418
‘Fr.logit0.574442516811659:Re398‘	NA	NA
‘Fr.logit1:Re398‘	10.4542661	15.6444047

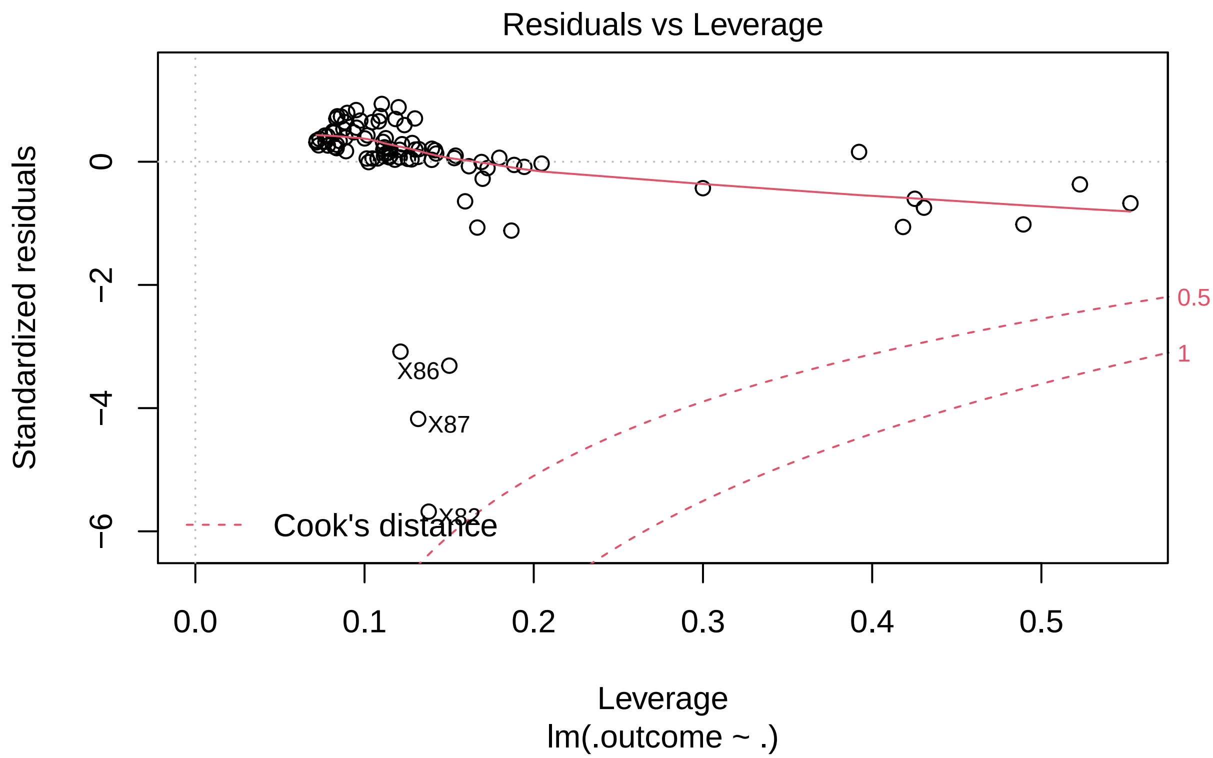
```
## Linear Regression
##
## 89 samples
##  3 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 71, 71, 71, 72, 71
## Resampling results:
##
##   RMSE      Rsquared   MAE
##   2.661394  0.8527879  1.570022
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```



Residuals vs Fitted
Residuals
0
−5
−10
−15
X86
X87
X82
0
5
10
15
20
25
Fitted values
lm(.outcome ~ .)


Normal Q−Q
Standardized residuals
1
0
−1
−2
−3
−4
−5
X86
X87
X82
−2
−1
0
1
2
Theoretical Quantiles
lm(.outcome ~ .)


Scale−Location
X82
X87
X86
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
0
5
10
15
20
25
Fitted values
lm(.outcome ~ .)

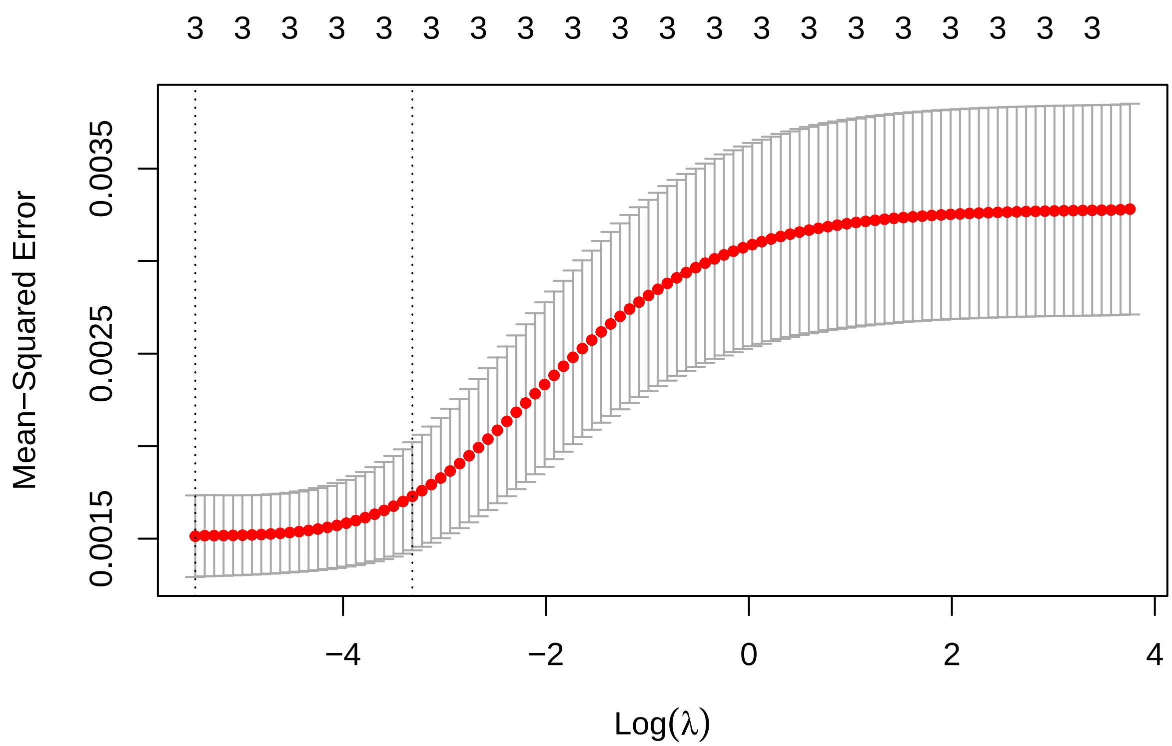


	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	21.4768836	1.0242751	20.9678854	0.0000000
St	1.7147879	0.7513377	2.2823132	0.0252691
Fr.logit0.574442516811659	-18.8360802	1.5250533	-12.3510967	0.0000000
Fr.logit1	-18.7901080	1.3867204	-13.5500340	0.0000000
Re224	-14.7802916	1.2868076	-11.4860151	0.0000000
Re398	-21.3493429	1.5520334	-13.7557242	0.0000000
‘St:Fr.logit0.574442516811659‘	0.3206283	1.0250294	0.3127991	0.7552900
‘St:Fr.logit1‘	0.0065094	0.8594265	0.0075742	0.9939766
‘St:Re224‘	-0.1384372	0.8349088	-0.1658112	0.8687459
‘St:Re398‘	-1.3817884	1.0794491	-1.2800867	0.2044082
‘Fr.logit0.574442516811659:Re224‘	12.4355392	1.5752541	7.8943069	0.0000000
‘Fr.logit1:Re224‘	12.7191484	1.5959297	7.9697425	0.0000000
‘Fr.logit1:Re398‘	19.1345746	1.7450344	10.9651561	0.0000000

	2.5 %	97.5 %
(Intercept)	19.4368629	23.5169043
St	0.2183692	3.2112066
Fr.logit0.574442516811659	-21.8734870	-15.7986734
Fr.logit1	-21.5520009	-16.0282151
Re224	-17.3431911	-12.2173922
Re398	-24.4404852	-18.2582005
‘St:Fr.logit0.574442516811659‘	-1.7208946	2.3621512
‘St:Fr.logit1‘	-1.7051867	1.7182056
‘St:Re224‘	-1.8013023	1.5244279
‘St:Re398‘	-3.5316977	0.7681208
‘Fr.logit0.574442516811659:Re224‘	9.2981488	15.5729296
‘Fr.logit1:Re224‘	9.5405790	15.8977178
‘Fr.logit0.574442516811659:Re398‘	NA	NA
‘Fr.logit1:Re398‘	15.6590374	22.6101118

Ridge Regression

```
## [1] 0.001078549
```

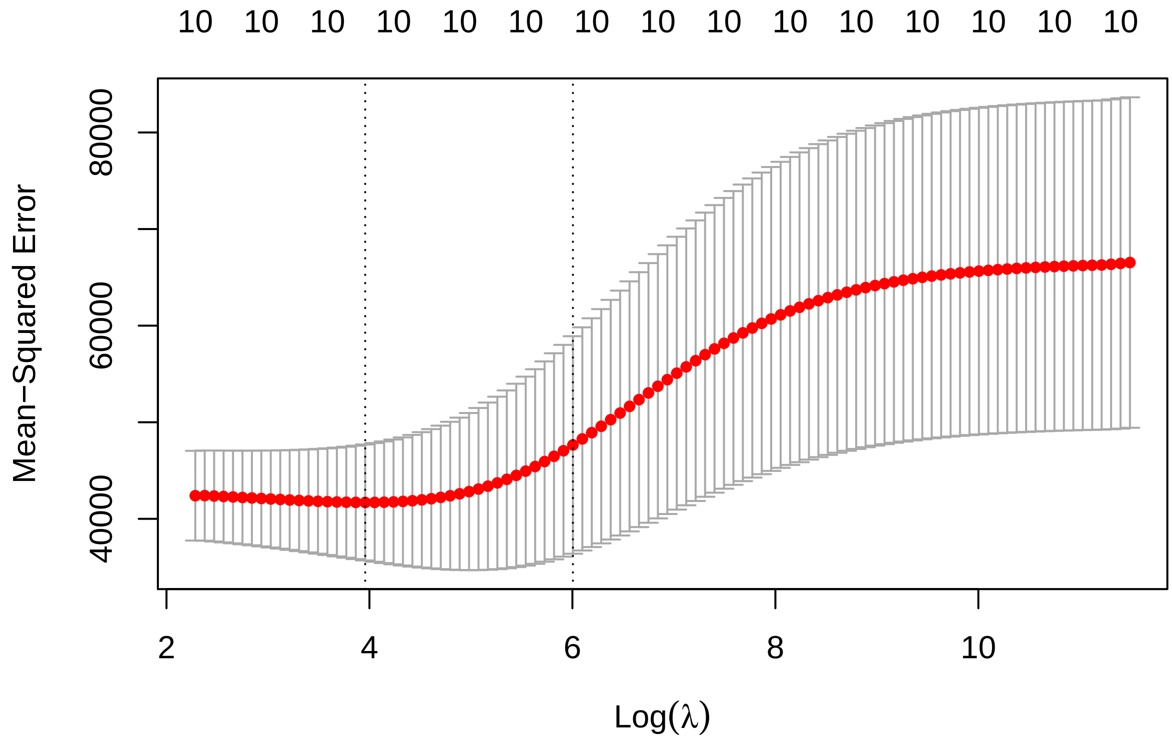



```
## [1] 0.004272015
```

```
## [1] 0.001091239
```

MSE stays basically the same

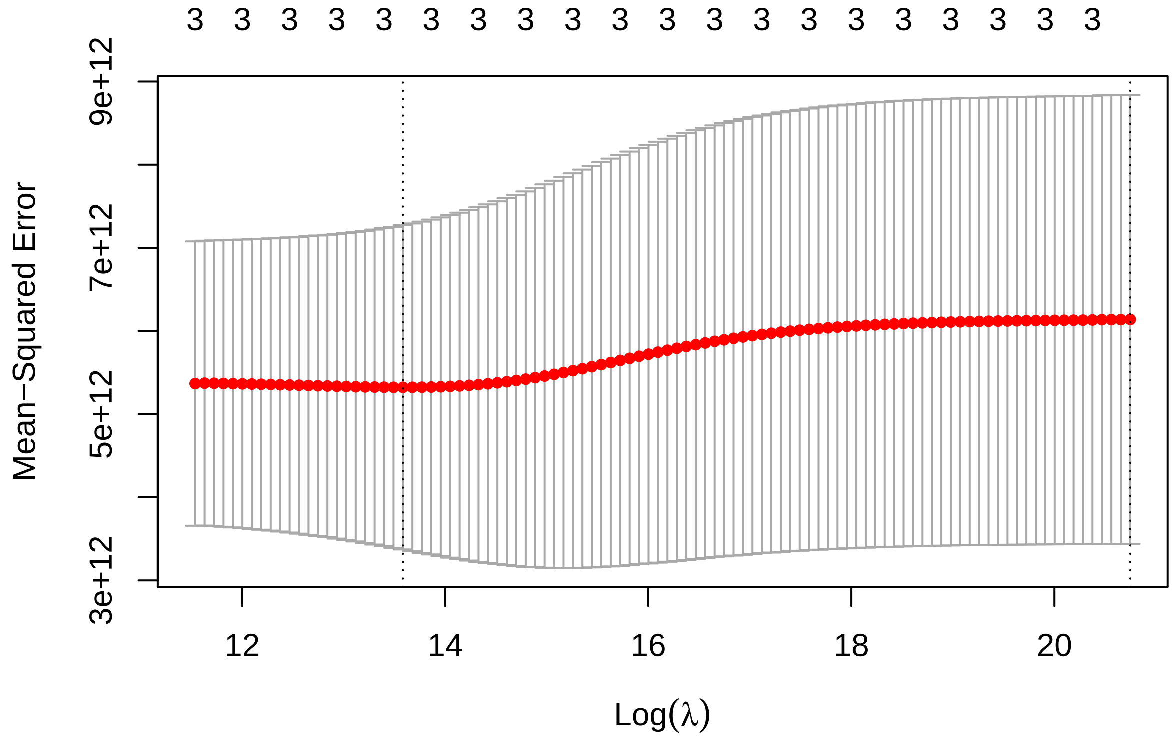
```
## [1] 43396.09
```




```
## [1] 52.36596
```

```
## [1] 46196.89
```

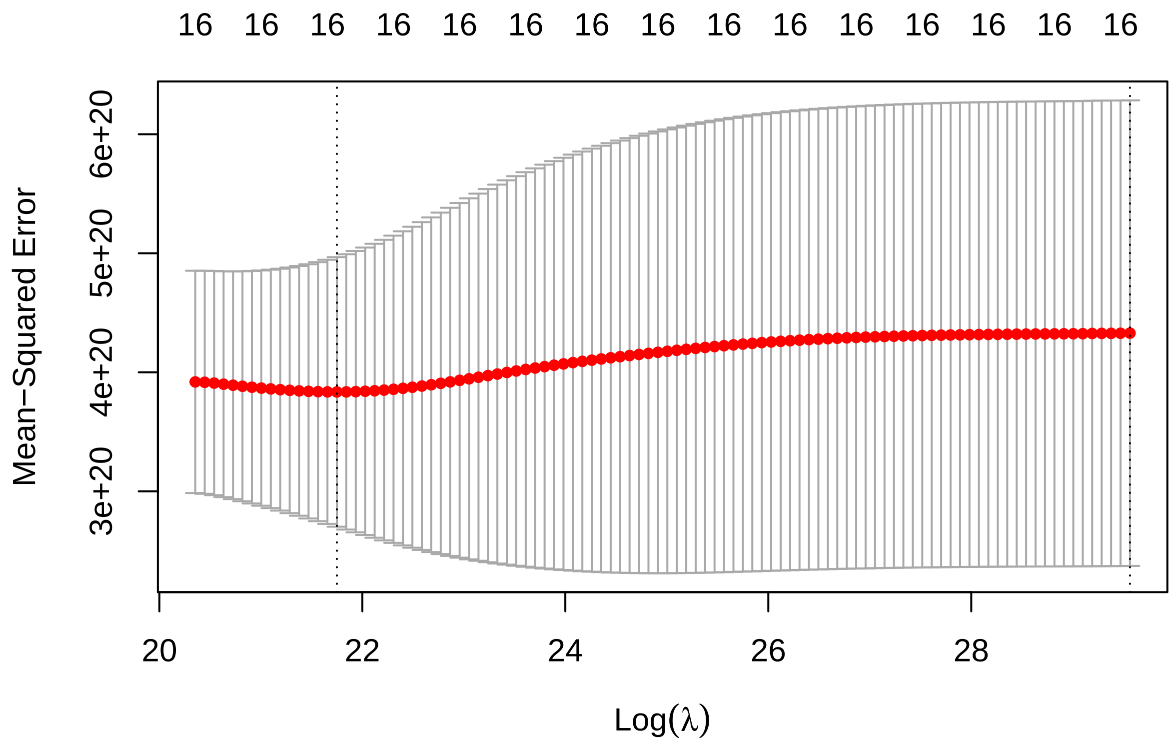
```
## [1] 3.166732e+12
```




```
## [1] 792973.2
```

```
## [1] 3.007604e+12
```

```
## [1] 3.509278e+20
```




```
## [1] 2789083228
```

```
## [1] 2.039584e+20
```

improvement? still large

GAMS

```
## [1] 0.001226049
```

```
## [1] 48573.71
```

```
## [1] 2.819773e+20
```