

# Project 1

MA8701

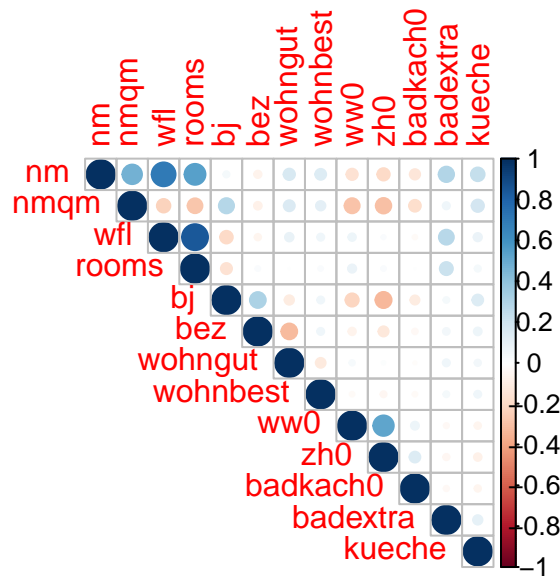
Group 5 : Yellow Submarine

07 February, 2021

## The Data Set

For our project work we use the Munich Rent 2003 data set as described in <https://rdrr.io/cran/LinRegInter/active/man/munichrent03.html>.

The data set has the covariates - **nmqm**: rent per square meter (numeric) - **wfl**: area in square meters (numeric) - **rooms**: number of rooms (numeric) - **bj**: year of construction (factor) - **bez**: district (factor) - **wohngut**: quality of location (factor) - **wohnbest**: high quality of location (factor) - **ww0**: hot water supply available (factor) - **zh0**: central heating (factor) - **badkach0**: tiled bathroom (factor) - **badextra**: high-quality bathroom (factor) - **kueche**: upscale kitchen equipment (factor) and the response - **nm**: rental price (numeric).



```
##      nm nmqm wfl rooms  bj bez wohngut wohnbest ww0 zh0 badkach0 badextra
## 1 741.39 10.90 68    2 1918  2      1      0  0  0      0      0
## 2 715.82 11.01 65    2 1995  2      1      0  0  0      0      0
## 3 528.25  8.38 63    3 1918  2      1      0  0  0      0      0
## 4 553.99  8.52 65    3 1983 16      0      0  0  0      0      1
## 5 698.21  6.98 100   4 1995 16      1      0  0  0      0      1
## 6 935.65 11.55 81    4 1980 16      0      0  0  0      0      0
##      kueche
## 1      0
## 2      0
## 3      0
## 4      0
```

```
## 5      1
## 6      0
```

We store the data set in an R data frame for all further computations.

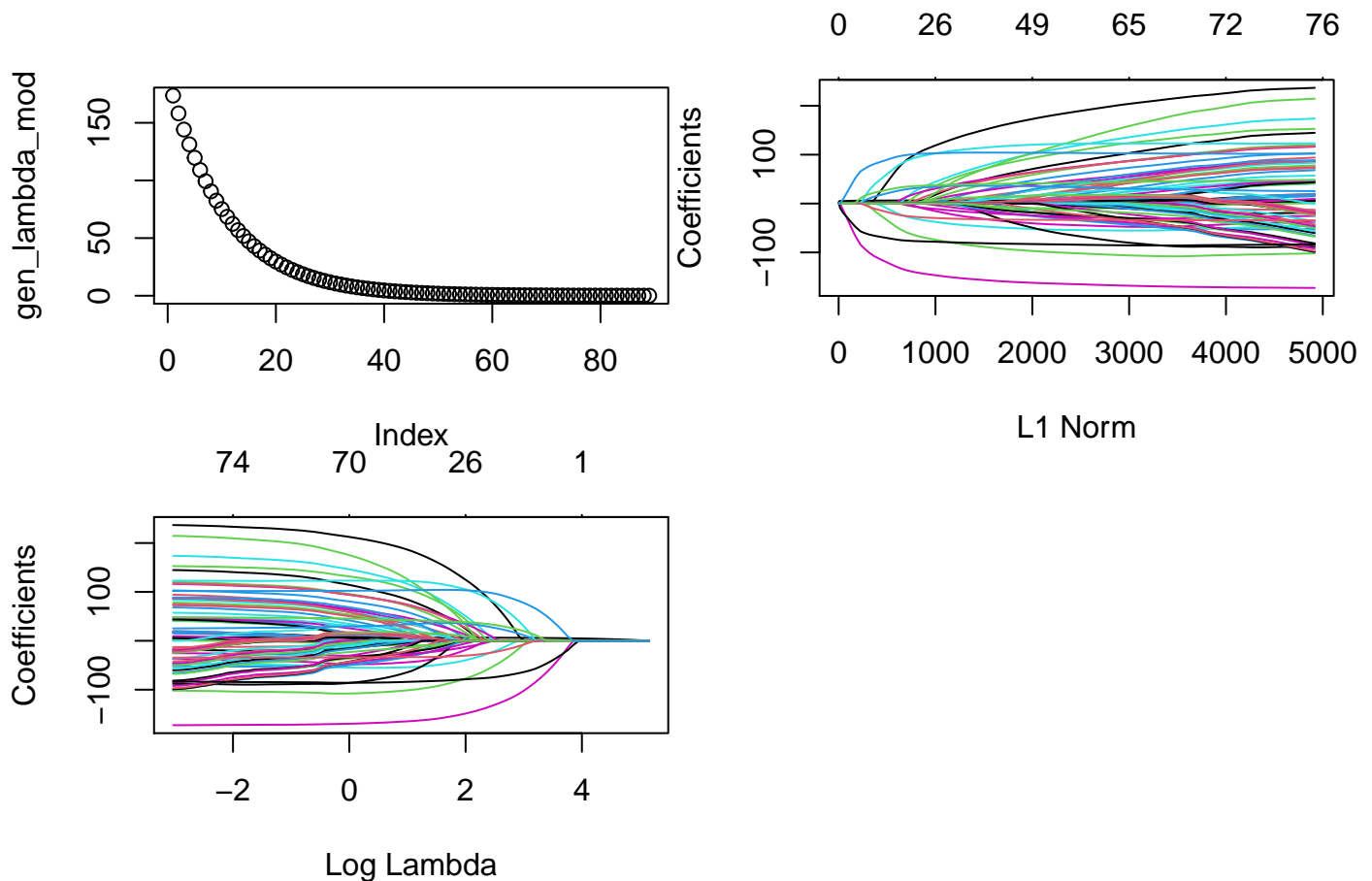
## Regression

We start with a vanilla regression for reference.

Remark: Interestingly in the regression, the significance of different **bjs** and **bezs** varies a lot.

## Shrinkage

After we saw the results for the linear regression, we continue with several methods including subset selection and shrinkage methods.



For the  $\lambda$  with one standard deviation, we observe that many of the **bjs** and **bezs** get shrinked, but not all of them - and the values differ from the linear regression. Whereas the other kept covariants roughly keep their parameter.

Above we considered a fixed  $\lambda$ , now we analyse which  $\lambda$  is optimal using cross validation.

## Group lasso

In the grouped lasso, the **bj** and **bez** are all shrinked or are all included, respectively. This coincides better with our intuition, that this criterion is considered or not considered. Whereas in the regression and lasso

before, just some years of construction and some areas where significant.

..... STILL TO DO BY FB!!!!

## [1] 1.35951