

Project 1

MA8701

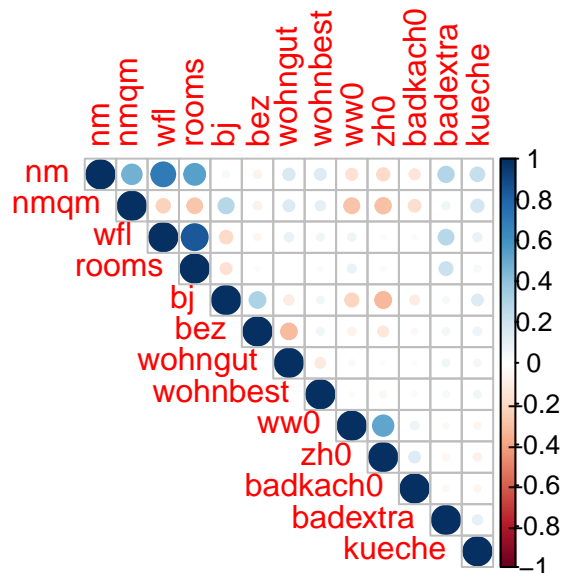
Group 5 : Yellow Submarine

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).



```
## 'data.frame': 2053 obs. of 13 variables:
## $ nm : num 741 716 528 554 698 ...
## $ nmqm : num 10.9 11.01 8.38 8.52 6.98 ...
## $ wfl : int 68 65 63 65 100 81 55 79 52 77 ...
## $ rooms : int 2 2 3 3 4 4 2 3 1 3 ...
## $ bj : Factor w/ 44 levels "1918","1924",...: 1 37 1 25 37 22 2 2 5 4 ...
## $ bez : Factor w/ 25 levels "1","2","3","4",...: 2 2 2 16 16 16 6 6 6 6 ...
## $ wohngut : int 1 1 1 0 1 0 0 0 0 0 ...
## $ wohnbest: int 0 0 0 0 0 0 0 0 0 0 ...
## $ ww0 : int 0 0 0 0 0 0 0 0 0 0 ...
## $ zh0 : int 0 0 0 0 0 0 0 0 0 0 ...
## $ badkach0: int 0 0 0 0 0 0 0 0 0 0 ...
```

```
## $ badextra: int 0 0 0 1 1 0 1 0 0 0 ...
## $ kueche : int 0 0 0 0 1 0 0 0 0 0 ...
```

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

Regression

We start with a vanilla regression for reference.

```
##
## Call:
## lm(formula = nm ~ ., data = df_mod)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -616.33  -78.78   -1.42    82.19   705.60
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  162.3104    27.8995   5.818 6.94e-09 ***
## wfl           6.9216     0.2635  26.263 < 2e-16 ***
## rooms        -12.9199     6.4332  -2.008 0.044743 *
## bj1924       -100.1093    19.6709  -5.089 3.94e-07 ***
## bj1939        -51.0820    40.0059  -1.277 0.201801
## bj1948        -43.4699    17.1866  -2.529 0.011507 *
## bj1957        -24.2381    13.1852  -1.838 0.066170 .
## bj1957.5       18.7138    19.6667   0.952 0.341443
## bj1960         19.5617    15.5390   1.259 0.208223
## bj1966         5.9203    13.9939   0.423 0.672292
## bj1967         17.4326    27.0595   0.644 0.519499
## bj1968         6.1619    32.4242   0.190 0.849297
## bj1969        -35.1239    24.9312  -1.409 0.159042
## bj1970         8.1467    24.1903   0.337 0.736322
## bj1971        22.7388    27.0303   0.841 0.400318
## bj1972         3.4642    18.3819   0.188 0.850538
## bj1973        22.2193    22.3580   0.994 0.320445
## bj1974        43.7002    29.1988   1.497 0.134645
## bj1975        12.5650    38.5568   0.326 0.744548
## bj1976       -86.6050    57.0576  -1.518 0.129211
## bj1977        97.6443    60.8334   1.605 0.108629
## bj1978        44.0685    66.2852   0.665 0.506236
## bj1979        50.1127    61.2159   0.819 0.413101
## bj1980        49.9373    36.6220   1.364 0.172852
## bj1981        88.5097    39.0632   2.266 0.023571 *
## bj1982       -17.1652    52.8476  -0.325 0.745363
## bj1983        74.8158    25.1821   2.971 0.003004 **
## bj1984        80.9532    37.2329   2.174 0.029805 *
## bj1985       105.8678    34.4111   3.077 0.002123 **
## bj1986        59.2255    45.4133   1.304 0.192336
## bj1987        49.1158    34.2715   1.433 0.151977
## bj1988       147.9157    42.3852   3.490 0.000494 ***
## bj1989        77.6490    39.4970   1.966 0.049445 *
## bj1990       154.2909    47.4466   3.252 0.001166 **
## bj1991        71.3473    41.2410   1.730 0.083785 .
```

```

## bj1992      86.5411    31.9958    2.705 0.006894 **
## bj1993      90.3129    25.1462    3.592 0.000337 ***
## bj1994     239.5327    41.9447    5.711 1.30e-08 ***
## bj1995      90.1354    49.9745    1.804 0.071442 .
## bj1996     123.4211    34.5818    3.569 0.000367 ***
## bj1997      88.8192    43.9147    2.023 0.043255 *
## bj1998     177.0494    40.6209    4.359 1.38e-05 ***
## bj1998.5    119.0793    28.0125    4.251 2.23e-05 ***
## bj1999      47.0015    57.0361    0.824 0.410002
## bj2000     120.2847    35.8825    3.352 0.000817 ***
## bj2001     218.5516    67.0896    3.258 0.001143 **
## bez2       -35.9851    25.4861   -1.412 0.158122
## bez3       -16.2744    26.2572   -0.620 0.535455
## bez4       -34.4740    26.0011   -1.326 0.185036
## bez5       -38.4664    25.8800   -1.486 0.137350
## bez6       -59.2431    29.6860   -1.996 0.046109 *
## bez7      -101.9950    29.8102   -3.421 0.000636 ***
## bez8       -65.3975    30.1790   -2.167 0.030355 *
## bez9       -52.0535    25.3943   -2.050 0.040515 *
## bez10      -63.8332    30.9749   -2.061 0.039453 *
## bez11      -98.8313    29.9727   -3.297 0.000993 ***
## bez12      -32.0354    28.2229   -1.135 0.256478
## bez13      -41.7103    27.8436   -1.498 0.134287
## bez14     -115.8630    30.5571   -3.792 0.000154 ***
## bez15      -85.0417    33.4341   -2.544 0.011048 *
## bez16     -109.2551    27.5635   -3.964 7.64e-05 ***
## bez17      -76.9986    29.6753   -2.595 0.009537 **
## bez18      -39.0532    28.6017   -1.365 0.172278
## bez19      -67.3556    27.6692   -2.434 0.015008 *
## bez20      -82.5750    31.8512   -2.593 0.009598 **
## bez21      -73.1990    30.3489   -2.412 0.015960 *
## bez22     -102.4685    38.8056   -2.641 0.008342 **
## bez23     -116.8833    46.5163   -2.513 0.012059 *
## bez24     -114.4170    36.5471   -3.131 0.001770 **
## bez25      -83.9379    27.1701   -3.089 0.002034 **
## wohngut     24.9111     8.3923    2.968 0.003030 **
## wohnbest    123.2647    23.5906    5.225 1.92e-07 ***
## ww0        -173.0875    20.8217   -8.313 < 2e-16 ***
## zh0        -82.6242    14.3232   -5.769 9.26e-09 ***
## badkach0    -34.4896     8.6321   -3.996 6.69e-05 ***
## badextra    48.6276    11.9878    4.056 5.18e-05 ***
## kueche     101.8619    13.2662    7.678 2.52e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 146 on 1976 degrees of freedom
## Multiple R-squared:  0.6591, Adjusted R-squared:  0.646
## F-statistic: 50.28 on 76 and 1976 DF,  p-value: < 2.2e-16

```

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

Shrinkage

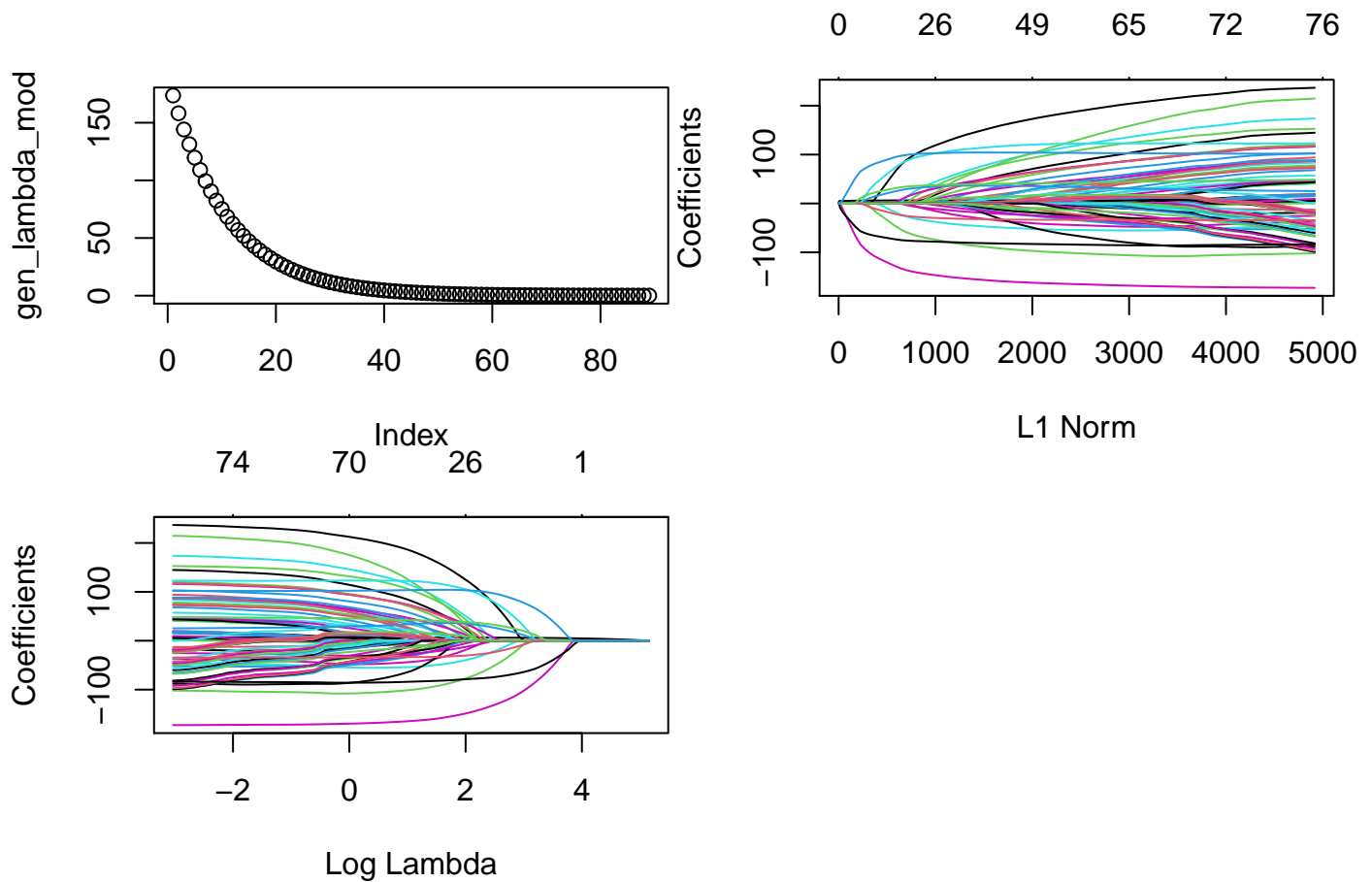
After we saw the results for the linear regression, we continue with several methods including subset selection and shrinkage methods.

Subset selection

```
## Subset selection object
## 11 Variables (and intercept)
##           Forced in Forced out
## wfl          FALSE      FALSE
## rooms        FALSE      FALSE
## bj           FALSE      FALSE
## bez          FALSE      FALSE
## wohngut      FALSE      FALSE
## wohnbest     FALSE      FALSE
## ww0          FALSE      FALSE
## zh0          FALSE      FALSE
## badkach0     FALSE      FALSE
## badextra     FALSE      FALSE
## kueche       FALSE      FALSE
## 1 subsets of each size up to 8
## Selection Algorithm: exhaustive
##           wfl rooms bj  bez wohngut wohnbest ww0 zh0 badkach0 badextra kueche
## 1 ( 1 ) "*" " " " " " " " " " " " " " " " " " " " " " " " " " " " "
## 2 ( 1 ) "*" " " " " " " " " " " " " "*" " " " " " " " " " " " "
## 3 ( 1 ) "*" " " " " " " " " " " " " "*" " " " " " " " " " " " "
## 4 ( 1 ) "*" " " "*" " " " " " " " " "*" " " " " " " " " " " " "
## 5 ( 1 ) "*" " " "*" " " " "*" " " " " "*" " " " " " " " " " " " "
## 6 ( 1 ) "*" " " "*" " " " "*" "*" " " " " " " " " " " " " " " " "
## 7 ( 1 ) "*" " " "*" " " " "*" "*" "*" " " " " " " " " " " " " " "
## 8 ( 1 ) "*" " " "*" "*" "*" "*" "*" " " " " " " " " " " " " " " "
## [1] 8
```

Ridge

Lasso



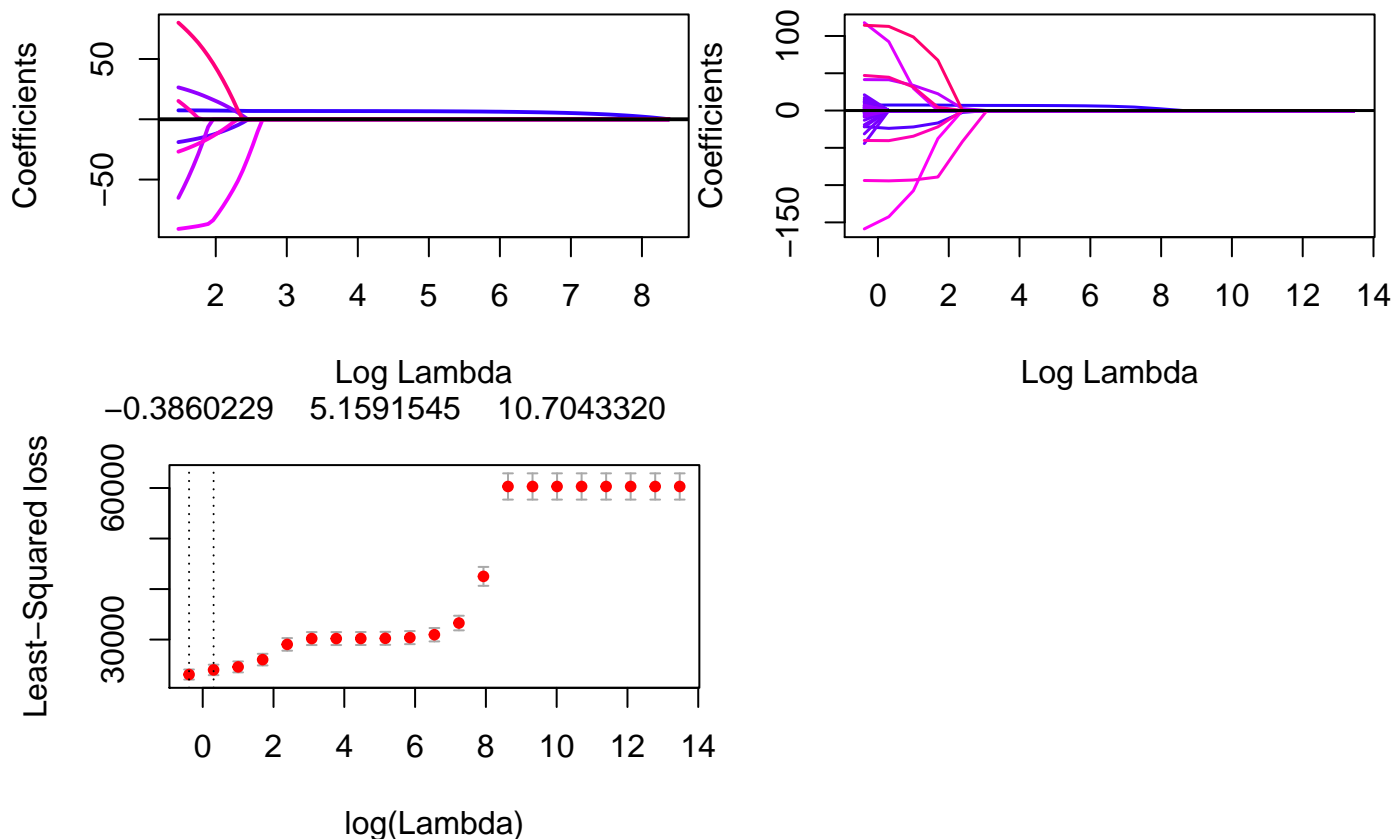
For the λ with one standard deviation, we observe that many of the **bjs** and **bez**s get shrunk, but not all of them - and the values differ from the linear regression. Whereas the other kept covariants roughly keep their parameter.

Above we considered a fixed λ , now we analyse which λ is optimal using cross validation.

Group lasso

In the grouped lasso, the **bj** and **bez** are all shrunk 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!!!!



```
## List of 9
## $ lambda      : num [1:21] 712775 356387 178194 89097 44548 ...
## $ cvm         : num [1:21] 60304 60304 60304 60304 60304 ...
## $ cvsd        : num [1:21] 2599 2599 2599 2599 2599 ...
## $ cvupper     : num [1:21] 62902 62902 62902 62902 62902 ...
## $ cvlo        : num [1:21] 57705 57705 57705 57705 57705 ...
## $ name        : Named chr "Least-Squared loss"
## ..- attr(*, "names")= chr "L2"
## $ gglasso.fit:List of 9
## ..$ b0         : Named num [1:21] 570 570 570 570 570 ...
## .. ..- attr(*, "names")= chr [1:21] "s0" "s1" "s2" "s3" ...
## ..$ beta       : num [1:76, 1:21] 0 0 0 0 0 0 0 0 0 0 0 ...
## .. ..- attr(*, "dimnames")=List of 2
## .. .. ..$ : chr [1:76] "wfl" "rooms" "bj1924" "bj1939" ...
## .. .. ..$ : chr [1:21] "s0" "s1" "s2" "s3" ...
## ..$ df         : Named int [1:21] 0 0 0 0 0 0 0 0 0 1 1 ...
## .. ..- attr(*, "names")= chr [1:21] "s0" "s1" "s2" "s3" ...
## ..$ dim        : int [1:2] 76 21
## ..$ lambda     : num [1:21] 712775 356387 178194 89097 44548 ...
## ..$ npasses    : int 2657
## ..$ jerr       : int 0
## ..$ group      : int [1:76] 1 2 3 3 3 3 3 3 3 3 ...
## ..$ call       : language gglasso(x = x, y = y, group = group, loss = "ls", lambda = lambda, delta = d
## ..- attr(*, "class")= chr [1:2] "gglasso" "ls"
## $ lambda.min   : num 0.68
## $ lambda.1se   : num 1.36
## - attr(*, "class")= chr "cv.gglasso"
```

```

## [1] 1.35951

##              s0
## (Intercept) 106.107086
## wfl         7.413067
## rooms      -23.871982
## bj1924      0.000000
## bj1939      0.000000
## bj1948      0.000000
## bj1957      0.000000
## bj1957.5    0.000000
## bj1960      0.000000
## bj1966      0.000000
## bj1967      0.000000
## bj1968      0.000000
## bj1969      0.000000
## bj1970      0.000000
## bj1971      0.000000
## bj1972      0.000000
## bj1973      0.000000
## bj1974      0.000000
## bj1975      0.000000
## bj1976      0.000000
## bj1977      0.000000
## bj1978      0.000000
## bj1979      0.000000
## bj1980      0.000000
## bj1981      0.000000
## bj1982      0.000000
## bj1983      0.000000
## bj1984      0.000000
## bj1985      0.000000
## bj1986      0.000000
## bj1987      0.000000
## bj1988      0.000000
## bj1989      0.000000
## bj1990      0.000000
## bj1991      0.000000
## bj1992      0.000000
## bj1993      0.000000
## bj1994      0.000000
## bj1995      0.000000
## bj1996      0.000000
## bj1997      0.000000
## bj1998      0.000000
## bj1998.5    0.000000
## bj1999      0.000000
## bj2000      0.000000
## bj2001      0.000000
## bez2        0.000000
## bez3        0.000000
## bez4        0.000000
## bez5        0.000000
## bez6        0.000000

```

```

## bez7          0.000000
## bez8          0.000000
## bez9          0.000000
## bez10         0.000000
## bez11         0.000000
## bez12         0.000000
## bez13         0.000000
## bez14         0.000000
## bez15         0.000000
## bez16         0.000000
## bez17         0.000000
## bez18         0.000000
## bez19         0.000000
## bez20         0.000000
## bez21         0.000000
## bez22         0.000000
## bez23         0.000000
## bez24         0.000000
## bez25         0.000000
## wohngut       41.318858
## wohnbest      92.318048
## ww0           -142.415965
## zh0           -94.286712
## badkach0      -40.350166
## badextra      44.631976
## kueche        112.865650

##              s0
## (Intercept) 100.907935
## wfl          7.371921
## rooms       -17.563424
## bj1924       0.000000
## bj1939       0.000000
## bj1948       0.000000
## bj1957       0.000000
## bj1957.5     0.000000
## bj1960       0.000000
## bj1966       0.000000
## bj1967       0.000000
## bj1968       0.000000
## bj1969       0.000000
## bj1970       0.000000
## bj1971       0.000000
## bj1972       0.000000
## bj1973       0.000000
## bj1974       0.000000
## bj1975       0.000000
## bj1976       0.000000
## bj1977       0.000000
## bj1978       0.000000
## bj1979       0.000000
## bj1980       0.000000
## bj1981       0.000000
## bj1982       0.000000

```



```

## bj1983      0.000000
## bj1984      0.000000
## bj1985      0.000000
## bj1986      0.000000
## bj1987      0.000000
## bj1988      0.000000
## bj1989      0.000000
## bj1990      0.000000
## bj1991      0.000000
## bj1992      0.000000
## bj1993      0.000000
## bj1994      0.000000
## bj1995      0.000000
## bj1996      0.000000
## bj1997      0.000000
## bj1998      0.000000
## bj1998.5    0.000000
## bj1999      0.000000
## bj2000      0.000000
## bj2001      0.000000
## bez2        0.000000
## bez3        0.000000
## bez4        0.000000
## bez5        0.000000
## bez6        0.000000
## bez7        0.000000
## bez8        0.000000
## bez9        0.000000
## bez10       0.000000
## bez11       0.000000
## bez12       0.000000
## bez13       0.000000
## bez14       0.000000
## bez15       0.000000
## bez16       0.000000
## bez17       0.000000
## bez18       0.000000
## bez19       0.000000
## bez20       0.000000
## bez21       0.000000
## bez22       0.000000
## bez23       0.000000
## bez24       0.000000
## bez25       0.000000
## wohngut     24.120239
## wohnbest    0.000000
## ww0         -48.870443
## zh0         -89.859070
## badkach0    -23.908694
## badextra     8.772584
## kueche      72.633830

## 77 x 3 sparse Matrix of class "dgCMatrix"
##           group lasso general lasso vanilla LS

```

## (Intercept)	106.107086	121.53253762	162.310441
## wfl	7.413067	6.35790480	6.921638
## rooms	-23.871982	.	-12.919931
## bj1924	.	-77.41285147	-100.109344
## bj1939	.	.	-51.082040
## bj1948	.	-39.87004946	-43.469920
## bj1957	.	-12.77226096	-24.238117
## bj1957.5	.	.	18.713838
## bj1960	.	.	19.561674
## bj1966	.	-1.78110003	5.920349
## bj1967	.	.	17.432638
## bj1968	.	.	6.161898
## bj1969	.	-19.07548904	-35.123926
## bj1970	.	.	8.146714
## bj1971	.	.	22.738843
## bj1972	.	.	3.464200
## bj1973	.	.	22.219275
## bj1974	.	.	43.700203
## bj1975	.	.	12.564953
## bj1976	.	.	-86.605034
## bj1977	.	.	97.644285
## bj1978	.	.	44.068520
## bj1979	.	.	50.112745
## bj1980	.	.	49.937326
## bj1981	.	.	88.509713
## bj1982	.	.	-17.165153
## bj1983	.	0.03551727	74.815843
## bj1984	.	.	80.953167
## bj1985	.	5.50490755	105.867818
## bj1986	.	.	59.225499
## bj1987	.	.	49.115827
## bj1988	.	25.05760499	147.915666
## bj1989	.	.	77.648956
## bj1990	.	43.94662766	154.290945
## bj1991	.	.	71.347309
## bj1992	.	.	86.541067
## bj1993	.	12.74643665	90.312924
## bj1994	.	125.34863498	239.532748
## bj1995	.	.	90.135389
## bj1996	.	17.20771477	123.421116
## bj1997	.	.	88.819228
## bj1998	.	51.01474230	177.049378
## bj1998.5	.	34.40462157	119.079298
## bj1999	.	.	47.001514
## bj2000	.	26.95792991	120.284699
## bj2001	.	33.21189163	218.551590
## bez2	.	.	-35.985131
## bez3	.	.	-16.274425
## bez4	.	.	-34.474015
## bez5	.	.	-38.466358
## bez6	.	.	-59.243092
## bez7	.	.	-101.994969
## bez8	.	.	-65.397522
## bez9	.	.	-52.053469

## bez10	.	.	-63.833161
## bez11	.	.	-98.831306
## bez12	.	.	-32.035394
## bez13	.	8.87285870	-41.710326
## bez14	.	.	-115.863027
## bez15	.	.	-85.041679
## bez16	.	-5.64277000	-109.255107
## bez17	.	.	-76.998642
## bez18	.	.	-39.053201
## bez19	.	.	-67.355571
## bez20	.	.	-82.574987
## bez21	.	.	-73.198994
## bez22	.	.	-102.468535
## bez23	.	.	-116.883323
## bez24	.	.	-114.417039
## bez25	.	.	-83.937882
## wohngut	41.318858	35.58541690	24.911148
## wohnbest	92.318048	105.40971577	123.264686
## ww0	-142.415965	-148.97130968	-173.087458
## zh0	-94.286712	-78.75042575	-82.624164
## badkach0	-40.350166	-30.32923639	-34.489575
## badextra	44.631976	38.13432987	48.627634
## kueche	112.865650	103.33736726	101.861941