Data Analysis Project 1

MA8701

Group 5 : Yellow Submarine

15 February, 2021

Note on Open Science

To pursue the idea of reproducible research, the chosen dataset as well as the code for our analysis are publicly accessible:

- dataset: https://data.ub.uni-muenchen.de/2/1/miete03.asc
- code: https://github.com/FlorianBeiser/MA8701

The Data Set

In this project, we analyse a real dataset using shrinkage methods. For our project work we use the Munich Rent 2003 data set as described in https://rdrr.io/cran/LinRegInteractive/man/munichrent03.html. The data set has 12 covariates, of which many are suffering multicollinearity, a brief introduction to these parameters are listed below:

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

For the data anaysis, the aim is to perform regression. Our data set is suited for that, since it suffers from multicolinearity as we see in Figure 1. For further data analysis, we store the data set in an R data frame.

Regression

We start with a vanilla LM regression for reference.

```
##
## Call:
## lm(formula = nm ~ ., data = df_mod)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

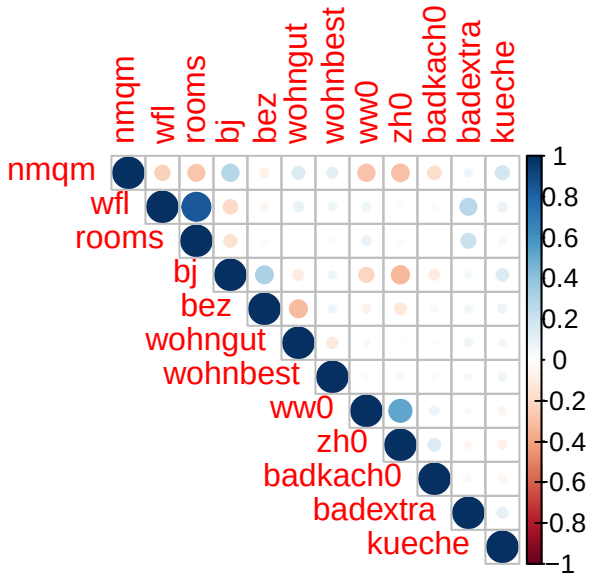

Figure 1: Correlation between the covariates

```
## -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
```

```
-> ->
```

```
-> -> -> -> -> -> -> -> -> -> -> ->
```

```
->
```

```
->
```