

House Rental Price Analysis

Data Preparation

I have introduced the dataset `municrent03` which is integrated in the R package `LinRegInteractive`, available at `README.md` file. Therefore, I can simply load this dataset to begin the subsequent steps of the analysis.

```
library(LinRegInteractive)
data(municrent03)
data <- municrent03
```

I began by examining the variable types to understand the structure of the dataset.

```
> str(data)
'data.frame': 2053 obs. of 12 variables:
 $ rent      : num  741 716 528 554 698 ...
 $ rentsqm   : num  10.9 11.01 8.38 8.52 6.98 ...
 $ area      : int   68 65 63 65 100 81 55 79 52 77 ...
 $ rooms     : int    2 2 3 3 4 4 2 3 1 3 ...
 $ yearc     : num  1918 1995 1918 1983 1995 ...
 $ bathextra : Factor w/ 2 levels "no","yes": 1 1 1 2 2 1 2 1 1 1 ...
 $ bathtile  : Factor w/ 2 levels "yes","no": 1 1 1 1 1 1 1 1 1 1 ...
 $ cheating  : Factor w/ 2 levels "yes","no": 1 1 1 1 1 1 1 1 1 1 ...
 $ district  : Factor w/ 25 levels "All-Umenz","Alt-Le",...: 10 10 10 17 17 17 21 21 21 21 ...
 $ location  : Ord.factor w/ 3 levels "normal"<"good"<...: 2 2 2 1 2 1 1 1 1 1 ...
 $ upkitchen : Factor w/ 2 levels "no","yes": 1 1 1 1 2 1 1 1 1 1 ...
 $ wwater    : Factor w/ 2 levels "yes","no": 1 1 1 1 1 1 1 1 1 1 ...
> names(data)
'rent' 'rentsqm' 'area' 'rooms' 'yearc' 'bathextra' 'bathtile' 'cheating' 'district' 'location' 'upkitchen' 'wwater'
```

After that, I reviewed the distributions, value ranges, and identified any potential missing values.

```
> summary(data)

      rent      rentsqm      area      rooms
Min.   : 77.31   Min.   : 1.470   Min.   : 17.0   Min.   :1.000
1st Qu.: 389.95   1st Qu.: 6.800   1st Qu.: 53.0   1st Qu.:2.000
Median : 534.30   Median : 8.470   Median : 67.0   Median :3.000
Mean   : 570.09   Mean   : 8.394   Mean   : 69.6   Mean   :2.598
3rd Qu.: 700.48   3rd Qu.:10.090   3rd Qu.: 83.0   3rd Qu.:3.000
Max.   :1789.55   Max.   :20.090   Max.   :185.0   Max.   :6.000

yearc      bathextra bathtile  cheating      district      location
Min.   :1918   no :1862   yes:1673   yes:1878   Neuh-Nymp: 177   normal:1205
1st Qu.:1948   yes: 191   no : 380   no : 175   Lud-Isar : 161   good : 803
Median :1960                                     Au-Haid  : 139   top  : 45
Mean    :1958                                     SchwWest : 137
3rd Qu.:1973                                     Maxvor  : 132
Max.    :2001                                     Laim    : 117
(Other) :1190

upkitchen  wwater
no :1903   yes:1981
yes: 150   no : 72
```

Note that the variables `rentsqm`, `rent` and `area` are related by the equation: $\text{rent} = \text{rentsqm} \times \text{area}$. For example, at row 100, we have:

```
> round(data$rent[100]/data$area[100],2) #compute rentsqm
11.3
> data$rentsqm[100]
11.3
```

Since `rentsqm` is a derived variable, I chose to exclude it and instead focus on total `rent`, which may better capture the underlying relationships with other features.

```
> data$rentsqm <- NULL
```

Exploratory Data Analysis

Graphical Model Learning

Inference and Querying

$$\mathbb{P}(X < 1, Y > 1) = \int_{-\infty}^1 \int_1^{+\infty} f(x, y) dx dy$$

Averaged Model - hc

