Excersice Sheet 6

A broker wants to use linear regression to find out which factors have a large influence on the price of a property. For this purpose, the variables described in Table 1 are given for the last 88 sales in the broker's region.

Table 1 House price record

Variabel	Description
`price`	house price ($\times$ 1,000 EUR)
`bdrms`	number bedrooms
`lotsize`	parking area (m^2)
`sqrm`	house area (m^2)
`country`	$==1$ when in country house style
`lprice`	`log(price)`
`llotsize`	`log(lotsize)`
`lsqrm`	`log(sqrm)`

1. Create a linear regression model with `price` as dependent variable and `bdrms`, `lotsize`, `sqrm` und `country` as independent variables.
   a) Determine the regression coefficients and p-values of the dependent variable and compare their influence within the model on the predicted value for `price`.
   b) Determine how much variance of the dependent variable is explained.
   c) Check the residuals (graphically) for normal distribution and homoskedasticity.

```
# Solution for Task 1...
```

2. Given be the linear regression model from task 1.
   a) Create a scatterplot to display the relationship between the predicted value for `price` and the residual size.
   b) For some houses, the price forecast of the broker model is more than EUR 100,000 off. Highlight houses with a residual size of more than 100 or less than 100. What could be the reasons for high model inaccuracies?
   c) Can the R^2-value be increased by using a linear transformation of one of the independent variables?

```
# Solution for Task 2...
```

3. Graphically display the relationship between `bdrms` and `price`. Check whether this relationship is also reflected in the regression model from Task 1. Create a regression model with `bdrms` as the only independent variable. Compare the regression coefficients with those of the model from Task 1 and interpret the differences.

```
# Solution for Task 3...
```

Dataset:

- http://isgwww.cs.uni-magdeburg.de/cv/lehre/VisAnalytics/material/exercise/datasets/hprice.csv