Lineare Modelle **2. Block**
Grün **WS 2017/18**

2 Das klassische lineare Modell: Modellannahmen

Aufgabe 1:

Im Folgenden bestimmen wir für den Münchner Mietspiegel Datensatz folgende zwei Regressionen:

$$\begin{aligned} \textit{mieteqm} &= \beta_0 + \beta_1 \cdot \textit{flaeche} + \epsilon \\ \textit{mieteqm} &= \beta_0 + \beta_1 \cdot \frac{1}{\textit{flaeche}} + \epsilon \end{aligned}$$

- Schätzen Sie die jeweiligen Koeffizienten mithilfe der KQ-Methode und interpretieren Sie sie.
- Erstellen Sie einen Scatterplot für die Beobachtungen *flaeche* und *mieteqm* und zeichnen Sie die beiden Prognoselinien ein.
- Erstellen Sie eine Graphik der Residuen und analysieren Sie, ob die Modellannahmen des klassischen linearen Modell jeweils erfüllt sind.
- Welches Modell passt besser?

Aufgabe 2:

Im Folgenden betrachten wir nur Wohnungen, die 1966 oder später gebaut wurden.

Bestimmen Sie separat für die Wohnungen in guter Lage sowie die Wohnungen in normaler Lage die Regressionsmodelle:

$$\textit{miete} = \beta_0 + \beta_1 \cdot \textit{flaeche} + \epsilon.$$

- Erstellen Sie ein Streudiagramm der Daten und zeichnen Sie die beiden Geraden ein.

Als nächstes erstellen Sie folgendes Modell für Wohnungen in guter und normaler Lage gemeinsam:

$$\textit{miete} = \beta_0 + \beta_1 \cdot \textit{flaeche} + \beta_2 \cdot \textit{lage} + \epsilon.$$

- Erstellen Sie ein Streudiagramm der Daten und zeichnen Sie die beiden Geraden, die sich aus dem Modell für die Wohnungen in guter Lage bzw. in normaler Lage ergeben, ein.
- Interpretieren Sie das Modell und vergleichen Sie es zu den beiden separaten Modellen.

Aufgabe 3:

Passen Sie die folgenden 2 Regressionsmodelle an den Mietspiegeldatensatz an:

$$\begin{aligned} \textit{mieteqm} &= \beta_0 + \beta_1 \cdot \textit{flaeche} + \beta_2 \cdot \textit{flaeche}^2 + \epsilon \\ \textit{mieteqm} &= \beta_0 + \beta_1 \cdot \textit{flaeche} + \beta_2 \cdot \textit{flaeche}^2 + \beta_3 \cdot \textit{flaeche}^3 + \epsilon \end{aligned}$$

- Visualisieren Sie die vorhergesagten Werte, indem Sie sie in das Streudiagramm der Daten jeweils einzeichnen.
- Visualisieren Sie auch die Residuen.

Vgl. Abb. 3.9.

Hinweis: Das Modell $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$ kann in R mithilfe folgendem Aufrufs angepasst werden:

```
> lm(y ~ x + I(x^2))
```

Nachdem mathematische Operatoren wie z.B. `+` im Formelinterface eine andere Bedeutung haben, muss man die Formel-Interpretation vermeiden und bekommt die ursprüngliche arithmetische Bedeutung mithilfe von `I()`. Siehe auch `?I`.

Aufgabe 4:

Passen Sie das folgende Regressionsmodell an den Mietspiegeldatensatz an:

$$mieteqm = \beta_0 + \beta_1 \cdot glage + \beta_2 \cdot blage + \epsilon,$$

wobei *glage* die Dummy-Variable für gute Lage und *blage* die Dummy-Variable für beste Lage ist. D.h. es wird Dummy-Kodierung mit normaler Lage als Referenzkategorie verwendet.

- Interpretieren Sie die geschätzten Regressionskoeffizienten.
- Bestimmen Sie die für die 3 Lagen vorhergesagten Werten.

Hinweis:

```
> mietspiegel$lage <- factor(mietspiegel$lage, 1:3, c("normal", "gut", "beste"))
> model1 <- lm(mieteqm ~ lage, data = mietspiegel)
```