

Aufgabe 1:

Unter <http://www.stat.uni-muenchen.de/service/datenarchiv/miete/miete03.html> finden Sie Daten zum Münchner Mietspiegel 2003. Schätzen Sie den Einfluss der Wohnfläche (`wf1`) und der Anzahl der Räume (`rooms`) auf die Nettomiete (`nm`) und untersuchen Sie, ob es für dieses Beispiel Probleme mit Kollinearität gibt.

Aufgabe 2:

Sei $Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{ip-1} + \beta_p x_{ip} + \varepsilon_i$, $i = 1, \dots, n$ ein volles Modell und $Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{ip-1} + \varepsilon_i$, $i = 1, \dots, n$ ein Submodell.

- (a) Betrachten Sie das adjustierte Bestimmtheitsmaß

$$R_{adj}^2 = 1 - \frac{MSE}{MST} = 1 - \frac{\hat{\sigma}^2}{SST/(n-1)} = 1 - \frac{n-1}{n-p'}(1-R^2).$$

Zeigen Sie, dass das Bestimmtheitsmaß R^2 des vollen Modells immer größer gleich dem R^2 des Submodells ist. Wann gilt Gleichheit?

Zeigen Sie, dass R_{adj}^2 diese Monotonieeigenschaft nicht besitzt und im Gegensatz zu R^2 auch negativ werden kann.

- (b) Es sei die übliche Annahme (2.5) gegeben. Zeigen Sie, dass die Anwendung des AIC im vorliegenden Fall äquivalent ist zu einem F-Test mit kritischem Wert

$$F^{AIC} = (n-p')(\exp(2/n) - 1).$$

Aufgabe 3:

Betrachten Sie wieder den Datensatz `teengamb` (Blatt 3/Aufgabe 1).

- (a) Zunächst betrachten wir wieder das Modell, das nur die Haupteffekte aller erhobenen Variablen enthält. Gibt es Hinweise auf Heteroskedastie? Schätzen Sie ein alternatives Modell und geben Sie die zugehörige Varianz–Kovarianz–Matrix der Störterme an.
- (b) Passen Sie nun ein Modell mit allen zweifach-Interaktionen an:

```
library(faraway)
teengamb$sexF <- factor(teengamb$sex, levels = c(0,1), labels = c("m", "w"))
lm.gamble <- lm(gamble ~ (income + sexF + status + verbal)^2, data = teengamb)
```

Ausgehend von diesem Modell möchten Sie nun eine Variablenselektion anhand des AIC-Kriteriums durchführen. Geben Sie zunächst die Definition des AIC-Kriteriums an und beschreiben Sie die einzelnen Komponenten. Welchem Trade-Off versucht das Kriterium gerecht zu werden?

- (c) Führen Sie nun eine schrittweise Variablenselektion gemäß des AIC durch. Verwenden Sie dabei die Funktion `stepAIC` aus dem package `MASS` mit Option `direction = "both"`. Erklären und interpretieren Sie den Output, achten Sie dabei insbesondere auf die Behandlung der Interaktionsterme und der dazugehörigen Haupteffekte. Welche Terme enthält das zuletzt ausgewählte Modell?
- (d) Vergleichen Sie die schrittweise Variablenselektion mit einer Selektion basierend auf allen möglichen Modellen mit Haupteffekten und Zweifach–Interaktionen. Benutzen Sie dazu die Funktion `regsubsets` aus dem Package `leaps`.
- (e) Betrachtet man die Diagnostik des Modells aus der Angabe vor Variablenselektion, fällt z.B. die Beobachtung 24 auf. Passen Sie das Modell ohne diese Beobachtungen an und wiederholen Sie die schrittweise Variablenselektion aus (b). Welches Modell wird nun ausgewählt? Sollte die Analyse von Ausreißern und einflussreichen Punkten vor oder nach der Variablenselektion erfolgen?