

## Blatt 6

### Residuenanalyse und Modellwahl

#### Aufgabe 6.1 Forbes' Datensatz

1857 veröffentlichte der schottische Physiker James D. Forbes Messungen des Siedepunkts von Wasser bei verschiedenen Luftdrücken, `data(forbes, package = "MASS")`, die er zuvor in Schottland und den Alpen bestimmt hatte. Ziel war es, den Luftdruck aus dem Siedepunkt des Wassers vorherzusagen, da ersterer auf die Höhe über dem Meeresspiegel zu schließen erlaubt, Barometer zur damaligen Zeit aber keine robusten Instrumente und daher schlecht mitzuführen waren.

Passen Sie ein einfaches lineares Modell an die Daten an und überprüfen Sie die Modellannahmen (Homoskedastizität, Normalverteilung, etc.; vgl. Kap. 6). Korrigieren Sie ggf. das Modell. Zeichnen Sie zusätzlich zur Regressionsgeraden punktweise Konfidenzintervalle sowie simultane Konfidenzbänder für die Regressionsgerade sowie punktweise Vorhersageintervalle, jeweils zum Niveau 95%.

**Hinweis:** `plot.lm` stellt einige diagnostische Standardgrafiken zur Verfügung; weitere nützliche Funktionen sind `summary.lm`, `fitted`, `resid`, `rstandard`, `rstudent`, `hatvalues`.

#### Aufgabe 6.2 Occam's razor

Laden Sie den Datensatz<sup>1,2</sup> `statex.x77` aus `datasets`. Er enthält Informationen von 50 Bundesstaaten der USA hinsichtlich Einwohnerzahl (1975), Einkommen pro Kopf (1974), Analphabetenquote (1970), Lebenserwartung (1969–1971), Mordrate (1976), Anteil an Personen mit Abitur (1970), mittlerer Anzahl an Tagen mit Temperaturen unter Null (1931–1960) und Fläche (1970) der jeweiligen Bundesstaaten. Untersuchen Sie, durch welche Größen die Lebenserwartung geeignet erklären lässt. Nutzen Sie dazu die Methode der Kreuzvalidierung sowie das Akaike und das Bayessche Informationskriterium. Für die beiden letztgenannten stehen unter R die Funktionen `AIC` und `BIC` aus dem Paket `stats` zur Verfügung. Als kleine Hilfe zur Generierung aller linearer Modelle kann das Code-Segment unter <https://www.tu-ilmeneau.de/analysis/team/philipp-schmitz/> verwendet werden.

Nächste Übung: Donnerstag 29. Juni 2017 um 13:00 Uhr im Sr C 115

---

<sup>1</sup>U.S. Department of Commerce, Bureau of the Census, Statistical Abstract of the United States (1977).

<sup>2</sup>U.S. Department of Commerce, Bureau of the Census, County and City Data Book (1977).