



## Statistik II

### Übungsblatt 4

**R-Übung:** Dieses Übungsblatt soll die R Kenntnisse vertiefen.  
Dabei sollen bereits bekannte Datensätze mit linearen Modellen in R analysiert werden

1. "Trees"-Datensatz:

Erstellen Sie für den Tree-Datensatz folgende Regressions Modelle.

(a) Führen Sie eine lineare Regression der Form

$$y = \beta_0 + \beta_1 \text{Höhe} + \beta_2 \text{Durchmesser}$$

für das Volumen in Abhängigkeit von Höhe und Durchmesser durch.

(b) Führen Sie eine polynomiale Regression der Form

$$y = \beta_0 + \beta_1 \text{Höhe} + \beta_2 \text{Durchmesser} + \beta_3 \text{Durchmesser}^2$$

für das Volumen in Abhängigkeit von Höhe und Durchmesser durch.

(c) Logarithmieren Sie die Variablen Volumen, Höhe und Durchmesser und führen Sie dann eine lineare Regression wie in (a) durch, also

$$\ln y = \beta_0 + \beta_1 \ln(\text{Höhe}) + \beta_2 \ln(\text{Durchmesser}).$$

2. "Swiss-Datensatz": Laden Sie den Datensatz "Swiss" aus der R-Datensatz Bibliothek. Informieren Sie sich über die einzelnen Variablen und die Herkunft der Daten. Betrachtet werden soll die Variable "Infant.Mortality". Erstellen Sie ein Regressionsmodell mittels:

- Schrittweiser Vorwärts-Regression (via AIC).
- Schrittweiser Rückwärts-Regression (via AIC).
- Schrittweiser Vorwärts-Regression mittels Korrelation.
- Schrittweiser Regression mittels des `step` Kommandos.
- Vergleichen Sie die Ergebnisse aus (a)-(d)!

3. "SMSA-Datensatz":

(a) Führen Sie eine lineare Regression für die Mortalitätsrate durch.

- Starten Sie mit dem saturierten Modell.
- Eliminieren Sie alle nicht signifikanten Variablen.
- Nehmen Sie nun alle Ausreißer heraus, sie Ihrer Meinung nach das Modell zu stark beeinflussen.
- Vergleichen Sie Ihr bestes Modell mit dem aus (a). Was hat sich geändert?

4. Folgende Daten stammen von einem bei der Papierherstellung durchgeführten Experiment. Dabei wurde in der Phase des Blattpressens mit fünf verschiedenen Druckstärken gearbeitet. Ziel des Experiments war es, den Einfluß der Druckstärke auf die Reißfestigkeit des Papiers zu untersuchen. Für jede der fünf Druckstärken  $A, B, C, D$  und  $E$  wurden vier Blatt Papier dem Reißfest

unterzogen. Die angegebenen Werte geben den sogenannten Reißfaktor an, eine Maßzahl, die den Prozentsatz einer Standardkraft widerspiegelt, der benötigt wird, das Blatt zu zerreißen.

Druckstärke beim Blattpressen	Reißfaktor			
A	112	119	117	113
B	108	99	112	118
C	120	106	102	109
D	110	101	99	104
E	100	102	96	101

- (a) Vergleichen Sie die Gruppen graphisch!
- (b) Führen Sie eine Varianzanalyse für dieses Experiment durch!
- (c) Angenommen die Stufen  $A$  bis  $E$  stehen für konkrete Werte:  $A = 35.0$ ,  $B = 49.5$ ,  $C = 70.0$ ,  $D = 99.0$  und  $E = 140.0$ . Analysieren Sie nun diese Daten mit einer linearen Regression. Wie unterscheidet sich dieses Modell von dem aus (b)?

5. "Barley"-Datensatz:

Analysieren Sie den Barley-Datensatz

- (a) mit graphischen Mitteln.
- (b) mit Hilfe der Varianzanalyse.

Ziehen Sie dabei alle Variablen in Betracht und möglicherweise auch Interaktionen zwischen den Variablen. Sind die Ergebnisse aus (a) und (b) konsistent?

- (c) Führen Sie die Tukey Post-Hoc Tests durch.