

## Einführung in R

### 13. Aufgabenblatt

#### Präsenzaufgabe 1

Der Datensatz *ipo1.txt* enthält Daten aus einer Untersuchung des Einflusses der Beweidung durch Hasen auf die Samenproduktion einer zweijährigen Pflanze. 99 Pflanzen wurden auf zwei Beweidungstypen (**grazed** = mit Beweidung, **ungrazed** = ohne Beweidung) aufgeteilt. (grazed: Beweidungsdruck wurde durch Kaninchen für zwei Wochen implementiert). Am Ende der Wachstumsperiode wurden die Anzahlen der Früchte jeder Pflanze ausgezählt. Weil die Größe der Pflanzen zu Beginn des Experiments auch einen Einfluss haben könnte, wurde die Größe der Wurzel vor dem Einpflanzen gemessen. Der Datensatz enthält die folgenden Variablen:

**Wurzel:** Wurzelgröße zu Beginn des Versuches

**Beweidung :** Beweidungsdruck durch Kaninchen

**Fruechte:** Anzahl der Früchte pro Pflanze am Ende der Wachstumsperiode

- Untersuchen Sie den Einfluss der Wurzelgröße auf die Anzahl der Früchte mittels einer linearen Regressionsanalyse. Plotten Sie die Daten gemeinsam mit der Regressionsgeraden des angepassten Modells.
- Untersuchen Sie den Einfluss der Wurzelgröße und des Beweidungsdruckes auf die Anzahl der Früchte und ob die Beweidung die Beziehung zwischen Wurzelgröße und Anzahl der Früchte beeinflusst (*mit einem ANCOVA-Modell mit Wechselwirkung*). Verwenden Sie  $\alpha = 0.05$  für Ihre Analyse. Vereinfachen Sie das Modell nach Signifikanzen und geben Sie das Endmodell an.

#### Präsenzaufgabe 2

Der Datensatz *birthwt.txt* wurde 1986 vom medizinischen Zentrum Baystate in Springfield erhoben. Untersucht wurden Parameter, die das Geburtsgewicht (**bwt**) von Neugeborenen beeinflussen. Wir wollen untersuchen, wie das Geburtsgewicht vom Alter der Mutter (**age**) und ihrem Rauchverhalten während der Schwangerschaft (**smoke**: Nichtraucher = 0, Raucher = 1) abhängt. Als Ausgangsmodell benutzen Sie das Modell mit Wechselwirkung. Optimieren Sie das Modell nach Signifikanzen. Stellen Sie das geschätzte Modell grafisch dar.

*Hinweis:* die Ausprägungen des Merkmals **smoke** sind mit Zahlen codiert (0 und 1). Weil dieses Merkmal nominal-skaliert ist (kategorial), soll die Variable **smoke** zunächst als Faktor undefiniert werden.

### Präsenzaufgabe 3

Man möchte untersuchen ob das Wachstum von Hamstern von ihrem Futter und von ihrem Typ abhängt. Der Datensatz *hamster1.txt* enthält Daten aus einer fiktiven Studie. Vier Wochen lang wurden 60 Hamster von zwei Hamstertypen (Typ: light, dark) mit drei verschiedenen Diäten (Futter: food1, food2, food3) gefüttert und ihre Gewichtszunahme in *g* (Wachstum) nach vier Wochen ausgewertet.

- Verwenden Sie ein lineares Modell mit Interaktionsterm als Ausgangsmodell. Optimieren Sie das Modell nach Signifikanzen.
- Optimieren Sie das Modell nach Informationsgehalt (AIC). Geben Sie das geschätzte Modell bzw. die geschätzten Mittelwerte an.

### Hausaufgabe 1 (8 Punkte)

Beantworten Sie die Fragestellungen a. und b. aus Präsenzaufgabe 1 dieses Mal anhand des Datensatzes *ipo2.txt*.

### Hausaufgabe 2 (12 Punkte)

In einer Studie wurde untersucht, ob das Wachstum von Hamstern (Wachstum) vom Futter (Futter: diet1, diet2, diet3), vom Hamstertyp (Typ: dunkel, hell) und von der Region (Region: wald, wueste) abhängt. Die erfassten Daten sind im Datensatz *hamster3.txt* enthalten. Verwenden Sie das maximale lineare Modell (3-fache Varianzanalyse mit allen Interaktionstermen) als Ausgangsmodell und optimieren Sie das Modell nach Signifikanzen und nach AIC. Geben Sie das geschätzte Modell bzw. die geschätzten Mittelwerte an.

### Zusatzaufgabe (8 Punkte)

Betrachten Sie im Datensatz *survey* aus dem Paket *MASS* die Variablen *Pulse* (Pulsrate, Schläge pro Minute), *Smoke* (Rauchverhalten, verschiedene Kategorien), *Sex* und *Age*. Wir möchten den Zusammenhang der Variable *Pulse* mit den Kovariablen *Smoke*, *Sex*, *Age* mittels einer linearen Regressionsanalyse untersuchen. Verwenden Sie das maximale lineare Modell (3-fache Varianzanalyse mit allen Interaktionstermen) als Ausgangsmodell:

**Pulse ~ Smoke \* Sex \* Age**

Optimieren Sie das Modell nach Signifikanz und AIC. Geben Sie das geschätzte Modell an. Verwenden Sie  $\alpha = 0.05$  für Ihre Tests.

Abgabe der Lösungen: bis **Montag 27.01.2020**,

*maendle@uni-bremen.de*