

## Stochastik III - Übungsblatt 5

Abgabe: 19. 12. 2012 vor Beginn der Übung

### Aufgabe 1 (3 + 2 + 4 Punkte)

In zehn schweizer Skiorten wurde in der letzten Saison die Anzahl der Besucher registriert. Es wird angenommen, dass diese linear von der Gesamtlänge der Pisten und der Liftkapazität abhängt. Die Daten befinden sich auf der Vorlesungshomepage in der Datei `ski.txt`. Nehmen Sie an, dass die üblichen Normalverteilungsannahmen gelten.

- (a) Stellen Sie ein geeignetes multiples lineares Regressionsmodell auf und bestimmen Sie  $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)^\top$ . Testen Sie zum Niveau  $\alpha = 0.05$  die Hypothesen
- $H_0^{(1)}$  : Es besteht keine Abhängigkeit der Besucherzahl von den beiden Einflußgrößen.
  - $H_0^{(2)}$  : Es besteht keine Abhängigkeit der Besucherzahl von der Pistenlänge.
  - $H_0^{(3)}$  : Es besteht keine Abhängigkeit der Besucherzahl von der Liftkapazität.

Interpretieren Sie das Ergebnis.

- (b) Nehmen Sie nun an, dass die Besucherzahl nur von der Liftkapazität abhängt, d.h. wir betrachten ein einfaches lineares Regressionsmodell. Bestimmen Sie  $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)^\top$  für dieses Modell und testen Sie die Hypothese  $H_0^{(4)}$ , dass keine Abhängigkeit der Besucherzahl von der Liftkapazität besteht. Interpretieren Sie das Ergebnis.
- (c) Betrachten Sie wieder das Modell aus (b). In einem elften Skigebiet sei die Liftkapazität  $x_0 = 2500$ . Bestimmen Sie zum Niveau  $\gamma = 0.95$  ein Konfidenzintervall für den erwarteten Zielwert  $\beta_1 + \beta_2 x_0$ , ein Prognoseintervall für  $Y_0 = \beta_1 + \beta_2 x_0 + \varepsilon_0$  sowie ein Konfidenzband für die Regressionsgerade. Zeichnen Sie die Daten, die Regressionsgeraden, das Konfidenzintervall und das Prognoseintervall zusammen mit dem Konfidenzband in ein Schaubild.

### Aufgabe 2 (2 + 3 + 4 Punkte)

In dieser Aufgabe soll der Datensatz `wasser.txt` untersucht werden. Er enthält die experimentell bestimmten Siedepunkte von Wasser bei unterschiedlichem Luftdruck. Dabei bezeichnet die Spalte `Siedepunkt` den Siedepunkt in Grad Fahrenheit und die Spalte `Luftdruck` den Luftdruck in „inches of mercury“, einem amerikanischen Luftdruckmass.

Als Modell für diese Aufgabe nehmen wir an, dass der Siedepunkt linear in Abhängigkeit vom Luftdruck zunimmt. Sie können von identisch normalverteilten und unabhängigen Störgrößen ausgehen.

- (a) Führen Sie in R die lineare Regression durch ohne den Befehl `lm()` zu verwenden (d.h. mit Matrixoperationen). Geben Sie die Regressionsparameter aus und plotten Sie die Daten zusammen mit der Regressionsgerade in ein gemeinsames Schaubild.
- (b) Testen Sie mit R, ob der Siedepunkt doppelt so stark wächst wie der Luftdruck, und ob der Achsenabschnitt (d.h. der Funktionswert der Gerade bei Luftdruck 0) bei 150 liegt (in

einem gemeinsamen Test).

- (c) Leiten Sie eine Formel für die Schranken des Konfidenzbands für die Regressionsgerade her und plotten Sie sie zusätzlich in das Schaubild.

**Aufgabe 3** (1 + 2 + 2 + 2 + 2 Punkte)

Die Anzahl der Autos pro 100 Einwohner möge vom Pro-Kopf-Einkommen (in 10000 Euro) und dem Benzinpreis (in Euro) abhängen. Auf der Homepage der Vorlesung ist die Datei `autos.txt` verfügbar, die folgende Daten enthält:

Land	Autos	Einkommen	Benzinpreis
Belgien	30	29.4	136
Dänemark	28	33.0	129
Deutschland	35	31.2	113
Finnland	24	21.3	113
Frankreich	33	26.4	140
Griechenland	8	10.2	129
Irland	20	11.4	92
Island	34	29.4	131
Kanada	42	26.1	39
österreich	27	23.1	113

Bearbeiten Sie Teil a) und b) mit R und Teil c) - e) von Hand (d.h. R höchstens als Taschenrechner). Nehmen Sie für Teil c) - e) zusätzlich an, dass der Vektor der Störgrößen  $\mathcal{N}(0, \sigma^2 I)$ -verteilt ist. In Norwegen sei zusätzlich zu obigen Werten das Pro-Kopf-Einkommen  $x_{0,2} = 27.9$  und der Benzinpreis  $x_{0,3} = 138$  (für Teil d) und e)).

- (a) Legen Sie ein lineares Regressionsmodell zugrunde und schätzen Sie die Regressionsebene mit dem Befehl `lm()` in R.
- (b) Plotten Sie die Regressionsebene mit `persp()` in ein Schaubild und beschriften Sie die Achsen.
- (c) Bestimmen Sie zum Niveau  $\alpha = 0.05$  ein Konfidenzellipse für den Parametervektor  $\beta$ .
- (d) Bestimmen Sie zum Niveau  $\alpha = 0.05$  ein Konfidenzintervall für den erwarteten Zielwert  $\varphi(1, x_{0,2}, x_{0,3}) = \beta_1 + x_{0,2}\beta_2 + x_{0,3}\beta_3$ .
- (e) Bestimmen Sie zum Niveau  $\alpha = 0.05$  ein Prognoseintervall für die Zielvariable  $Y_0$ .