Prof. Dr. E. Spodarev
D. Meschenmoser

Wintersemester 2009 / 2010
21. Januar 2010

Statistik II – Übungsblatt 12

Abgabe: 28. Januar 2010, vor den Übungen

Aufgabe 1
Auf der Homepage der Vorlesung befindet sich die Datei `schweiz.txt`, die die Ergebnisse einer Befragung von 872 Haushalten in der Schweiz enthält. Dabei wurden folgende Daten erhoben.

- Teilnahme: Ist die Person erwerbstätig?
- Alter: Alter in Jahrzehnten
- Bildung: Dauer der Berufsausbildung in Jahren
- JKinder: Anzahl der jungen Kinder (jünger als 7)
- AKinder: Anzahl der älteren Kinder (älter als 7)
- Herkunft: Ist die Person Ausländer?

Teilnahme	**Alter**	**Bildung**	**JKinder**	**AKinder**	**Herkunft**
no	3	8	1	1	no
yes	4.5	8	0	1	no
⋮	⋮	⋮	⋮	⋮	⋮

a) Betrachte das Logit-Modell und regressiere die binäre Variable „Teilnahme“ (Zielvariable) auf alle übrigen (erklärenden) Variablen. Interpretiere den Output von `summary()`. (3)

b) Erweitere das Logit-Modell aus Aufgabe a), indem als weitere erklärende Variable das Quadrat des Alters mit aufgenommen wird und interpretiere wie in a) den R-Output. (3)

Hinweis: Als erster Parameter in `glm()` muss hier
`Teilnahme~1+Alter+Bildung+JKinder+AKinder+Herkunft+I(Alter^2)`
angegeben werden.

c) Welches der beiden Modelle aus a) und b) ist im Sinne des AIC-Kriteriums besser? (1)

d) Teste mit Hilfe des Likelihood-Quotiententests aus der Vorlesung, ob die Regressionskoeffizienten β_3 (Bildung), β_4 (JKinder), β_5 (AKinder) und β_6 (Herkunft) im Logit-Modell aus Aufgabe a) gleich 0 sind (Nullhypothese: $H_0 : \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$). Interpretiere das Testergebnis. (3)

Vorgehensweise in R:
`schweiza` bezeichne das Logit-Modell aus Aufgabe a) und `schweizd` das Logit-Modell unter H_0.
Anwendung des Likelihood-Quotiententests:

```
> anova(schweiza, schweizd, test = "Chisq")
```

e) Teste analog zu d) die Hypothese $H_0 : \beta_6 = 0$ vs. $H_1 : \beta_6 \neq 0$ im Logitmodell und interpretiere das Ergebnis. (3)

Aufgabe 2 (5)

Beweise Satz 5.2.1 aus der Vorlesung in allgemeiner Form, d.h. für beliebiges i.

Aufgabe 3

Auf der Vorlesungshomepage ist die Datei `pca-example.txt` verfügbar. Sie enthält 100 dreidimensionale Vektoren (x_1, x_2, x_3) von Pseudo-Zufallszahlen.

a) Führe Schritt für Schritt eine Hauptkomponentenanalyse auf den Daten durch. Du kannst R benutzen, um z.B. Eigenwerte zu berechnen, nicht aber die Methoden `prcomp()` und `princomp()`. Gib die Hauptkomponenten in Abhängigkeit von x_1, x_2, x_3 an. (3)

b) Berechne den Anteil der Gesamtvarianz, der durch die i-te Hauptkomponente erklärt wird, d.h. berechne Var (i-te Hauptkomponente) / $\sum_{j=1}^{3}$ Var (j-te Hauptkomponente). (2)

c) Projiziere die Daten auf die Ebene, die durch die ersten beiden Hauptkomponenten aufgespannt wird, und plotte diese Punkte zusammen mit den beiden Hauptkomponenten. (3)