



Statistik II

Übungsblatt 7

Abgabe: Do. 7.12.2006, 10.00 Uhr, Briefkasten: Statistik II.

Bei jeder Aufgabe können maximal 5 Punkte erreicht werden.

1. Folgender Datensatz über 1.544 Mordfälle im Bundesstaat North Carolina gibt Informationen über die Hautfarbe des Täters, die Hautfarbe des Opfers, sowie ob das Todesurteil gesprochen wurde, oder nicht:

Hautfarbe		Todesurteil	
Opfer	Täter	Ja	Nein
nicht weiß	nicht weiß	29	587
	weiß	4	76
weiß	nicht weiß	33	251
	weiß	33	541

Diese Daten sollen verwendet werden um zu prüfen, ob Todesurteile "fair" vergeben werden. Dies ergibt drei Fragen

- (a) Ist der Urteilsspruch unabhängig von der Hautfarbe des Täters?
- (b) Ist der Urteilsspruch unabhängig von der Hautfarbe des Opfers?
- (c) Ist der Urteilsspruch unabhängig von der Hautfarbe des Täters und des Opfers?

Formulieren Sie die obigen Fragen als loglineare Modelle, und überprüfen Sie welche Unabhängigkeiten gegeben sind.

2. Erstellen Sie die jeweiligen Mosaic Plots für die Fragen (a) bis (c) in Aufgabe 1. Plotten Sie jeweils die Graphiken für die Rohdaten und die Modelldaten. Vergleichen Sie die Unterschiede. Ergeben sich aus den Mosaic Plots andere Interpretationen, als aus den loglinearen Modellen in Aufgabe 1?
3. Berechnen Sie für die Daten aus Aufgabe 1 eine logistische Regression für Todesurteil. Interpretieren Sie das Ergebnis — auch im Unterschied zu den Ergebnissen aus Aufgabe 1.
4. Cressie und Read haben in ihrem 1989 erschienenen Buch "Goodness-of-Fit-Statistics for Discrete Multivariate Data" eine Familie von Gütekriterien vorgeschlagen, die sogenannte *Power-Divergence-Statistic*, definiert als

$$PD(\lambda) = \frac{2}{\lambda(\lambda + 1)} \sum_{i=1}^k X_i \left[\left(\frac{X_i}{\hat{m}_i} \right)^\lambda - 1 \right],$$

wobei \hat{m}_i die erwartete Zellhäufigkeit in Zelle i bezeichnet und λ ein zu wählender Parameter mit $-\infty < \lambda < \infty$ ist.

Zeigen Sie, daß die Power-Divergence-Statistic für $\lambda = 1$ der χ^2 -Statistik entspricht.

5. In einer amerikanischen Studie über homosexuelle Männer sollte untersucht werden, wie der Zusammenhang zwischen dem Verlust der Krankenversicherung, der Erkrankung an AIDS und der Ausbildungssituation ist. Folgende Daten wurden erfasst:

		Versicherungsverlust	
Ausbildung	AIDS	Ja	Nein
≤Schulabschluss	ja	1	23
	nein	8	116
Universität	ja	3	41
	nein	24	699
≥Promotion	ja	3	25
	nein	6	591

Bestimmen Sie mittels loglinearer Modellen:

- Gibt es einen Hinweis darauf, dass die AIDS Diagnose in Verbindung mit dem Versicherungsverlust steht?
- Besteht eine Verbindung zwischen dem Ausbildungsgrad und der Assoziation zwischen Diagnose und Versicherungsverlust?