

1 Analyse von Kontingenztafeln: Das loglineare Modell

Aufgabe 10

Basierend auf der Kriminalitätsstatistik der USA des Jahres 1987 erschien folgende AP-Nachricht: Die Wahrscheinlichkeit, irgendwann einmal Opfer eines Mordes zu werden, beträgt für ein männliches neugeborenes Kind mit weißer Hautfarbe 0.0049 und 0.0263 bei nicht-weißer Hautfarbe; für Mädchen beträgt sie 0.0023 und 0.0072 bei weißer bzw. nicht-weißer Hautfarbe.

- (a) Bestimmen Sie die bedingten odds ratios zwischen *Hautfarbe* und *Mordopfer* (ja/nein), gegeben *Geschlecht*. Interpretieren Sie das Ergebnis.
- (b) Ist es notwendig in ein, hierfür angenommenes, loglineares Modell eine 3-Faktor Interaktion aufzunehmen? Geben Sie eine geeignete Begründung an.

Aufgabe 11

Folgende Tabelle enthält die erwarteten gemeinsamen Häufigkeiten, wenn bei einer Reihe von Personen die drei Variablen *Pfadfinder* (*P*), *Kriminell* (*K*) und *Sozioökonomischer Status* (*S*) betrachtet werden.

Sozioökonomischer Status	Pfadfinder	Kriminell	
		ja	nein
niedrig	ja	10	40
	nein	40	160
mittel	ja	18	132
	nein	18	132
hoch	ja	8	192
	nein	2	48

- (a) Welches loglineare Modell beschreibt diese erwarteten Häufigkeiten? Tipp: Betrachten Sie hierzu die bedingten odds ratios von *P* und *K* gegeben *S*!
- (b) Bestimmen Sie den (marginalen) odds ratio von *P* und *K*! Was fällt auf?

Aufgabe 12

In R steht der Datensatz `HairEyeColor` zur Verfügung.

- (a) Verschaffen Sie sich einen Überblick über die Datenstruktur und visualisieren Sie die Daten. Machen Sie sich hierfür mit dem Paket `vcd`, zur Visualisierung kategorialer Daten, vertraut.
- (b) Aggregieren Sie zunächst bezüglich der Variablen *Sex*. Analysieren Sie die sich ergebende Tafel mittels loglinearer Modelle unter Verwendung der R-Funktion `loglin()`! Sind die beiden Variablen *Hair* und *Eye* unabhängig? Welche Methoden gibt es, um die Unabhängigkeit graphisch zu untersuchen?
- (c) Analysieren Sie nun den ursprünglichen Datensatz `HairEyeColor` mittels loglinearer Modelle unter Verwendung der R-Funktion `loglin()`! Für welches Modell entscheiden Sie sich?

Vorgehen:

Gehen Sie vom saturierten Modell aus und testen Sie anhand des Devianzunterschiedes Schritt für Schritt, ob es gerechtfertigt ist, ein restriktiveres Modell anzunehmen.

Aufgabe 13

In folgender Kontingenztafel sind Unfalldaten für Florida aus dem Jahr 1988 zusammengefasst. Die einzelnen Unfallbeteiligten sind gruppiert bezüglich der Variablen *Sicherheitsgurt* (angelegt / nicht angelegt), *Herausgeschleudert* (ja / nein) und *Verletzungen* (tödlich / nicht tödlich).

Sicherheitsgurt	Herausgeschleudert	Verletzungen	
		nicht tödlich	tödlich
angelegt	ja	1105	14
	nein	411 111	483
nicht angelegt	ja	4626	497
	nein	157 342	1008

- (a) Speichern Sie obige Daten in einem R-Objekt vom Typ `data.frame`!
- (b) Berechnen Sie zu diesem Datensatz das saturierte Modell und das Modell ohne 3-Faktor-Interaktionen unter Verwendung der R-Funktion `loglm()` aus dem Paket `MASS`! Ist es gerechtfertigt auf die 3-Faktor-Interaktionen zu verzichten? Betrachten Sie hierzu die Devianz und die bedingten odds ratios!
- (c) Verwenden Sie nun zur Berechnung des Modells ohne 3-Faktor-Interaktionen die R-Funktion `glm()` und vergleichen Sie die Ergebnisse mit denen aus Teilaufgabe (b)! Ergeben sich dieselben Schätzungen? Falls nein, warum?