Serie 10

1. Herzinfarkt und Verhütungsmittel (Agresti, 1990, 2.1.3): 58 verheiratete Patientinnen unter 45 Jahren, die wegen Herzinfarkt ('myocardial ..') im Spital waren und dreimal mehr "ähnliche" Patientinnen, wurden befragt, ob sie Verhütungsmittel ('Contraceptives') verwendet hätten. Die Antworten sind

		Herzinfarkt (Y)	
		ja	nein
Verhütungs-	ja	23	34
mittel (X)	nein	35	132

Sie können diese Aufgabe gut von Hand (mit Taschenrechner) oder auch mit S-plus (Datensatz `myocard.dat`) lösen.

a) Berechnen Sie die prozentualen Anteile. Benutzen Sie die Ihnen am sinnvollsten erscheinende Möglichkeit.

b) Berechnen Sie die gemeinsamen (π_{ij}) und bedingten ($\pi_{j|i} := P[Y = j|X = i]$, und $\pi_{i|j}$) Verteilungen sowie die Randverteilungen ($\pi_{i.}, \pi.j$). (Maximum Likelihood Schätzungen, d.h. relative Häufigkeiten!).

c) Berechnen Sie *relative Risikos*, $P[Y = j|X = 1] \,/\, P[Y = j|X = 0]$, das Assoziationsmass (*odds ratio*), $(\pi_{j|i}/\pi_{j'|i})/(\pi_{j|i'}/\pi_{j'|i'}) \stackrel{!}{=} \pi_{11}\pi_{22}/\pi_{12}\pi_{21}$, und auch dessen Logarithmus. Dabei sind j und j' die beiden möglichen Werte von Y (analog für i, i' und X). — Interpretation ?

d) Welches sind die Wahrscheinlichkeitswerte unter der Unabhängigkeitshypothese in jeder Zelle? Berechnen Sie die *Deviance*– und *Pearson* χ^2–Statistik sowie deren P–Werte.

Deviance $D(\theta_0) := 2\left(\log L(\hat\theta_{ML}) - \log L(\theta_0)\right)$, wobei $L(\theta)$ die *Likelihood* als Funktion aller Parameter θ bezeichnet. Für die Multinomialverteilung, $\log L(\theta) = \text{const.} + \sum_k n_k \log \pi_k$ (und $\sum_k \pi_k \equiv 1$).

χ^2-Statistik $X^2 := \sum_k (beob_k - erw_k)^2/erw_k$, wobei $beob_k$ die beobachtete und erw_k die erwartete *Anzahl* für die k-te Gruppe bezeichnen. Für die Multinomialverteilung, $X^2 = \sum_k \frac{(n_k - n.\pi_k)^2}{n.\pi_k}$.

2. Wir wollen nun die *'myocardial infarction'* Daten (Datensatz `myocard.dat`) als *log-lineares Modell* analysieren mit S-plus und `glm(..)` (s. Anleitung letztes Mal). Das Modell lautet hier

$$\log E\langle n_{ij}\rangle = \mu + \alpha_i + \beta_j + \gamma_{i,j},$$

wobei unter der Unabhängigkeitannahme $\gamma_{i,j} \equiv 0$ ist und z.B. α_i $\log \pi_{i\cdot}$ entspricht. und ist.

a) Erstellen sie zuerst die Modell-Variabeln aus der Kontingenztafel, z.B. mit den S-plus-Befehlen

```
d.myoc _ read.table(".....datasets/myocard.dat", header=T)
Myo _ rep(c("Y","no"), 2);       Myo _ C(as.factor(Myo), treatment)
Con _ rep(c("Y","no"), c(2,2));  Con _ C(as.factor(Con), treatment)
d.myoc2 _ data.frame(Myo, Con, freq= unname(unlist(d.myoc)))
```

Schauen Sie `Myo` und `Con` an! Die Zusätze mit `C(..)` definieren sogenannte „Kontraste", d.h., wir legen hier fest, dass im Modell die Nebenbedingungen $\alpha_1 = \beta_1 = \gamma_{1,j} = \gamma_{i,1} = 0$ genommen werden. Sonst lauteten die Bedingungen $\sum_i \alpha_i = \sum_j \beta_j = \ldots = 0$.

b) Verwenden Sie nun `glm(.., family = poisson)` für die Überprüfung der Unabhängikeit von Krankheit und der Empfängnisverhütung. Die Abhängige Variable ist also `freq`. Berechnen Sie beide Modelle (mit und ohne Interaktions–Term).

***c)** Leiten Sie die Formeln für die (oben von S-plus verwendeten) Schätzungen (für α_2, β_2 und $\gamma_{2,2}$) im *„saturierten Modell"*,

$$\log n_{ij} = \mu + \alpha_i + \beta_j + \gamma_{i,j}, \quad \text{mit } \alpha_1 = \beta_1 = \gamma_{1,j} = \gamma_{i,1} = 0$$

her, und zeigen Sie, dass gerade $\underline{\hat\gamma_{22} = \log \textit{odds ratio}}$ gilt.
Hinweis: Schreiben Sie die vier Gleichungen des Modells für i und $j = 1, 2$ auf!

Abgabe: Mo, 27. Jan. 97.
Präsenz, Hilfe: Im SOL F5 (Sonneggstr.33), gemäss Vereinbarung: Tel. (63) 2-3408 oder via e-mail To: `maechler@stat.math.ethz.ch`.

Übung via WWW in `ftp://ftp.stat.math.ethz.ch/Teaching/maechler/Stat-Meth/96-10.ps`