Reeksamen august 2008, Opgave 3

(a)

Hypotesen $\lambda_{ak} = \alpha_a\beta_k$ om multiplikativitet svarer til hypotesen om uafhængighed i en polynomialfordelingsmodel hvor totalsummen er fast, eller til hypotesen om at de fem sandsynlighedsparametre er ens i en binomialmodel hvor rækkesummerne er faste. Udtrykt i almindelige vendinger betyder den her, at forholdet mellem de to køns risiko for at komme til skade på knallert i trafikken er den samme i de fem aldersgrupper. Eller at de relative forhold mellem aldersgruppernes risiko er de samme for de to køn.

Selve testet foretages præcis som testet for uafhængighed:

$$-2\log q = 2(373\log 373 + \cdots + 6\log 6$$
$$-930\log 930 - \cdots - 46\log 46 + 1085\log 1085)$$

$=$ **3.2903**. Denne størrelse skal vurderes i en χ^2–fordeling med (5–1)(2–1)=4 frihedsgrader, hvor den er klart insignifikant. Vi godkender altså hypotesen om multiplikativitet (Pearsons teststørrelse bliver 3.3663, med samme konklusion).

(b)

At de forventede antal $\alpha_a\beta_k$ er større for mænd end for kvinder betyder naturligvis at $\beta_M > \beta_K$. Testet for $\beta_M = \beta_K$ (søjlehomogenitet) kan udføres som et test for $p = 0.5$ i en binomialfordeling med antalsparameter 1085, hvor 930 successer er observeret. Det giver selvfølgelig forkastelse. Teststørrelsen

$$-2\log q = 2\left(930\log\frac{930}{1085} + 155\log\frac{155}{1085} - 1085\log 0.5\right) = 614.18$$

skal vurderes i en χ^2–fordeling med 1 frihedsgrad, hvor den er alt, alt for stor.

(c)

Dette er modellen som i noterne kaldes "proportionalitet med en given baggrundsvariabel". Estimatet for β er $\hat\beta = \frac{1085}{5368354} = 0.0002021$, og de fittede værdier $\hat\beta n_a$ bliver

```
237.9    86.3   317.8   282.5   160.6
```

Alene forskellen mellem antallet 438 af tilskadekomster i den yngste aldersgruppe og den fittede værdi 237.9 er så ekstrem, at det er klart at modellen må forkastes. Afvigelsen $438 - 237.9 \approx 200$ burde jo ikke være meget større end den dobbelte standardafvigelse $2\sqrt{237.9} \approx 31$! Og dette til trods for, at befolkningstallet for aldersgruppen –17 tilsyneladende også omfatter børn i en alder, hvor man normalt ikke kører meget på knallert ...

Et egentligt test imod den fulde model (Pearsons version) ser sådan ud:

$$\frac{(438-237.9)^2}{237.9}+\cdots+\frac{(46-160.6)^2}{160.6}=392.87$$

(tilsvarende $-2\log q = 381.2$), som skal vurderes i en χ^2–fordeling med 4 frihedsgrader, hvor den er alt for stor. Tendensen er hele vejen igennem, at risikoen for tilskadekomst falder med alderen. Men ved formuleringen af konklusionen skal man være opmærksom på, at dette ikke nødvendigvis betyder, at de helt unge er elendige til at køre på knallert og de gamle meget dygtige. Det har nemlig lige så meget at gøre med *hvor meget* der køres på knallert i de forskellige aldersgrupper.