

Reeksamen august 2009, Opgave 3

(a)

Procenttabellen ser sådan ud:

KØN	M	K
AARSAG	-----	-----
Hjerte	41.77	37.31
Kræft	35.49	36.17
Hjerne	11.77	20.10
Ulykke	10.97	6.42
	-----	-----
SUM	100.00	100.00

Disse procenter kan fortolkes som estimatorer for sandsynlighedsparametrene i en model, hvor søjlesummerne 3153 og 2289 opfattes som givne, og indholdet af de to søjler opfattes som udfaldene af to uafhængige polynomialfordelte variable med hvert sit sæt af sandsynlighedsparametre. Den mest markante forskel mellem kønnene er, at kvinderne har næsten dobbelt så stor sandsynlighed som mændene for at dø af en hjernerelateret kredsløbslidelse (hjerneblødning, slagtilfælde). For mændene kompenseres dette af større sandsynligheder for at dø af hjertelaterede kredsløbslidelser eller omkomme i ulykker.

(b)

Vi antager, at de otte tal er fremkommet som observationer af uafhængige Poissonfordelte variable med parametre λ_{ak} , og tester multiplikativitetshypotesen $\lambda_{ak} = \alpha_a \beta_k$, som svarer til hypotesen om at de to polynomialfordelinger i spørgsmål (a) har de samme sandsynlighedsparametre. Altså at fordelingen af dødsårsag er den samme for mænd og kvinder. Vi får

$$-2 \log q = 2(1317 \log 1317 + \dots + 147 \log 147$$

$$-3153 \log 3153 - \dots - 493 \log 493 + 5442 \log 5442) = 97.62$$

(tilsvarende Pearson størrelse: 97.38). Denne teststørrelse skal vurderes i en χ^2 -fordeling med 3 frihedsgrader hvor den er ekstremt signifikant (99.99%-fraktilen er 21.1). Så vi forkaster hypotesen, og konkluderer at der, løst sagt, er forskel på hvad mænd og kvinder dør af. Dvs. de forskelle vi fandt i opgave (a) er faktisk signifikante.

(c)

De betingede sandsynligheder der tales om her kan vi estimere på sædvanlig måde ved at se på deltabellen bestående af række 1 og 3, og her betragte modellen hvor søjlesummerne opfattes som givne ("to binomialfordelinger"). Vi får estimatorerne

	KØN	M	K
AARSAG	-----		
Hjerte		78.02	64.99
Hjerne		21.98	35.01

SUM		100.00	100.00

hvor f.eks. de to tal i første række er estimerede sandsynligheder (i procent) for at dødsårsagen er en hjerterelateret kredsløbslidelse, givet at det er en kredsløbslidelse, for hver af de to køn. Fra spørgsmål (a) og (b) ved vi på det nærmeste, at disse to sandsynligheder er forskellige. Men det helt korrekte test for dette får vi ved at teste for uafhængighed i den tilsvarende 2×2 antalstabel. Vi får teststørrelsen

$$-2 \log q = 2(1317 \log 1317 + \dots + 460 \log 460$$

$$-1688 \log 1688 - \dots - 831 \log 831 + 3002 \log 3002) = 62.34$$

(tilsvarende Pearson størrelse: 62.65). Den skal vurderes i en χ^2 -fordeling med 1 frihedsgrad, hvor den er ekstremt signifikant (99.99%-fraktilen er 15.2). Så der er ingen tvivl om, at blandt dem der dør af kredsløbssygdomme er det kvinderne, der oftest dør af lidelser som er relateret til hjernen.