

Opgave 3

Til sammenligning af indkomstfordelingen i to amerikanske stater, Kalifornien og Wisconsin, har man udtaget 250 tilfældige husholdninger fra Kalifornien og 150 fra Wisconsin. Resultatet, efter passende gruppering af indkomsten i tre grupper, så sådan ud:

Indkomst:	Høj	Mellem	Lav
Kalifornien	70	80	100
Wisconsin	34	40	76

(a) Opstil en statistisk model til analyse af disse antal. Estimer i denne model sandsynligheden for at en tilfældig husstand i Kalifornien ligger i højeste indkomstgruppe, og angiv approksimative 95% sikkerhedsgrænser for denne sandsynlighed.

Da antallene af husstande i de to stater er fastlagt på forhånd, virker det mest naturligt at antage, at de to talsæt (70, 80, 100) og (34, 40, 76) er uafhængige observationer af polynomialfordelte variable af orden 3, med antalsparametre henholdsvis 250 og 150, og (i første omgang) med hvert sit sæt af sandsynlighedsparametre.

Maksimaliseringsestimatorerne i denne model er sædvanlige relative hyppigheder, så vi får (for eksempel) at sandsynligheden for at en tilfældig kalifornisk husstand har indkomst "Høj" skal estimeres ved

$$\hat{p}_{11} = \frac{70}{250} = 0.28.$$

95% sikkerhedsgrænser fås på sædvanlig måde efter formlen

$$p_{11} \pm 1.96 \times \sqrt{\frac{\hat{p}_{11}(1 - \hat{p}_{11})}{250}} = 0.28 \pm 0.05566$$

svarende til sikkerhedsintervallet [0.224, 0.336].

(b) Opstil en reduceret model, som svarer til hypotesen om at indkomstfordelingen er den samme i de to stater, og udregn de fittede værdier (estimerede middelværdier for observationerne) i denne model.

Den reducerede model kan karakteriseres som ovenfor, med den ændring at de to polynomialfordelinger nu skal have samme sandsynlighedsparametre. De fittede værdier fås efter formlen

$$\hat{E}Y_{ij} = \frac{y_{i..}y_{.j.}}{y_{..}}$$

som fører til værdierne

Indkomst:	Høj	Mellem	Lav
Kalifornien	65.0	75.0	110.0
Wisconsin	39.0	45.0	66.0

(c) Foretag et test for denne hypotese.

Testet er ækvivalent med det sædvanlige test for uafhængighed. Vi får

$$-2 \log q = 2(70 \times \log 70 + \cdots + 76 \times \log 76$$

$$-104 \times \log 104 - \cdots - 150 \times \log 150 + 400 \times \log 400) = 4.3309$$

som skal vurderes i en χ^2 -fordeling med 2 frihedsgrader. Da 95% fraktilen i denne fordeling er 5.991, kan teststørrelsen ikke på nogen måde siges at være for stor. Vi godkender således hypotesen om samme indkomstfordeling i Kalifornien og Wisconsin.

(Pearsons teststørrelse bliver 4.3388.)