ST 102 våren 1997 – Oppgave 4

Vi har $n = 58626$ firebarnsmødre

La N_j = antall firebarnsmødre som har j jenter ; $j = 0, 1, 2, 3, 4$

Da er $(N_0, N_1, N_2, N_3, N_4)$ multinomisk fordelt.
Den simultane punktsannsynligheten er

$$P(N_0 = n_0, N_1 = n_1, \ldots, N_4 = n_4)$$

$$= \frac{n!}{n_0!\, n_1!\, n_2!\, n_3!\, n_4!} p_0^{n_0} p_1^{n_1} p_2^{n_2} p_3^{n_3} p_4^{n_4}$$

Her er p_j sannsynligheten for at en firebarnsmor har j jenter.

a) Hvis antall jenter i en firebarnsfamilie er binomisk fordelt, har vi at

$$p_j = \binom{4}{j} p^j (1-p)^{4-j} \qquad \text{for } j = 0, 1, \ldots, 4$$

Likelihooden blir da (da K og K' er konstanter)

$$L(p) = K \prod_{j=0}^{4} \left\{ \binom{4}{j} p^j (1-p)^{4-j} \right\}^{n_j}$$

$$= K' \prod_{j=0}^{4} \left\{ p^{j n_j} (1-p)^{4n_j - j n_j} \right\}$$

$$= K' \, p^{\sum_{j=0}^{4} j n_j} (1-p)^{4\sum_{j=0}^{4} n_j - \sum_{j=0}^{4} j n_j}$$

$$= K' \, p^y (1-p)^{N-y}$$

der $y = \sum_{j=0}^{4} j n_j$ = antall jenter

og $N = 4\sum_{j=0}^{n_j} n_j = 4n$ = antall barn

Log-likelihooden blir

$$l(p) = \log L(p) = \log K' + y \log p + (N-y)\log(1-p)$$

Vi deriverer og får

$$l'(p) = \frac{y}{p} - \frac{N-y}{1-p}$$

Vi setter $l'(p) = 0$ og løser likningen.
Det gir ML estimatet

$$\hat{p} = y/N$$

Ut fra tallene i oppgaven finner vi

$$y = 0 \cdot 4699 + 1 \cdot 15251 + 2 \cdot 20806$$
$$+ 3 \cdot 13901 + 4 \cdot 3969 = 114442$$

$$N = 4 \cdot n = 4 \cdot 58626 = 234\,504$$

Maksimum likelihood estimatet for p blir dermed

$$\hat{p} = \frac{114\,442}{234\,504} = 0.488$$

b) Vi vil teste nullhypotesen

$$H_0: \; p_j = \binom{4}{j} p^j (1-p)^{4-j} \qquad j = 0, 1, \ldots, 4$$

mot alternativet at likheten ikke holder.

Til det bruker vi kji-kvadrat-observatoren

$$X^2 = \sum_{j=0}^{4} \frac{(N_j - E_j)^2}{E_j}$$

der N_j-ene er som gitt i punkt a

og

$$E_j = n \binom{4}{j} \hat{p}^j (1-\hat{p})^{4-j}$$

Hvis H_0 er sann, er X^2 kji-kvadrat fordelt med $5-1-1 = 3$ frihetsgrader

Ut fra tallene i oppgaven, finner vi

$$E_0 = 58626 \cdot (1-0.488)^4 = 4028.75$$

$$E_1 = 58626 \cdot 4 \cdot 0.488 \cdot (1-0.488)^3 = 15359.60$$

$$E_2 = 58626 \cdot 6 \cdot 0{,}488^2(1-0{,}488)^2 = 21959{,}43$$

$$E_3 = 58626 \cdot 4 \cdot 0{,}488^3(1-0{,}488) = 13953.39$$

$$E_4 = 58626 \cdot 0{,}488^4 = 3324.83$$

Kji-kvadratobservatoren får verdien

$$X^2 = 297.86$$

Dette er mye større enn $X^2_{0.005,3} = 12.84$

Vi forkaster H_0, og konkluderer at antall jenter i en firebarnsfamilie ikke er binomisk fordelt.

Ved å sammenligne de observerte og forventede antallene, ser vi at den binomiske fordelingen predikerer for få familier med bare gutter eller bare jenter og for mange familier med to av hvert kjønn.