ST102 Våren 1995 – Oppgave 2 a–d

X_{1j} = antall eneggede tvillingpar med mor i alderskategori j $(j=1,2,3)$

X_{2j} = antall toeggede tvillingpar med mor i alderskategori j $(j=1,2,3)$

p_{1j} = sannsynligheten for at tilfeldig valgt tvillingpar er enegget og har mor i alderskategori j $(j=1,2,3)$

p_{2j} = sannsynligheten for at tilfeldig valgt tvillingpar er toegget og har mor i alderskategori j $(j=1,2,3)$

a) Hvert tvillingpar faller i én av de $k=6$ kategoriene $(j=1,2,3)$

- Enegget og mor i alderskategori j
- Toegget og mor i alderskategori j

Sannsynlighetene for disse er

henholdsvis p_{1j} $(j=1,2,3)$ og p_{2j} $(j=1,2,3)$.

Forudsat at vi har uafhængighed, vil da $(X_{11}, X_{12}, X_{13}, X_{21}, X_{22}, X_{23})$ være multinomisk fordelt

Vi har at $0 \le p_{ij} \le 1$ for $i=1,2$ og $j=1,2,3$ og at

$$\sum_{i=1}^{2} \sum_{j=1}^{3} p_{ij} = 1$$

Modellen har derfor $k-1 = 6-1 = 5$ frit varierende parametre.

ML-estimatorerne (SME) er givet ved

$$\hat{p}_{ij} = \frac{X_{ij}}{200} \qquad i=1,2 \ ; \ j=1,2,3$$

Vi har at $X_{ij} \sim$ binomisk$(200, p_{ij})$

Av dette følger det at

$$E(X_{ij}) = 200 \cdot p_{ij} \quad \text{og} \quad V(X_{ij}) = 200\, p_{ij}(1-p_{ij})$$

Dermed har vi at

$$E(\hat{p}_{ij}) = \frac{E(X_{ij})}{200} = \frac{200 \cdot p_{ij}}{200} = p_{ij}$$

og

$$V(\hat{p}_{ij}) = \frac{V(X_{ij})}{200^2} = \frac{200\, p_{ij}(1-p_{ij})}{200^2} = \frac{p_{ij}(1-p_{ij})}{200}$$

(Du skal ikke finne kovariansene.)

Vi setter nå

$$p_{i\cdot} = \sum_{j=1}^{3} p_{ij} = \text{Sannsynlighet for}$$

eneggete (i=1) og

toeggete (i=2) tvillingpar

$$p_{\cdot j} = \sum_{i=1}^{2} p_{ij} = \text{Sannsynlighet for at}$$

man er i alderskategori j

(j=1,2,3)

b) Hvis mors alder er uavhengig av om tvillingparet er enegget eller toegget, er (tvillingtype 1 er enegget, 2 er toegget)

$$p_{ij} = P(\text{tvillingtype } i \cap \text{alderskategori } j)$$

$$= P(\text{tvillingtype } i)\,P(\text{alderskategori } j)$$

$$= p_{i\circ}\, p_{\circ j}$$

Antall parametere blir nå

$$k-1-((2-1)+(3-1)) = 6-1-3 = 2$$

Nå er ML estimatene

$$p_{i\circ}^{*} = \frac{X_{i1}+X_{i2}+X_{i3}}{200} = \frac{X_{i\circ}}{200}$$

$$p_{\circ j}^{*} = \frac{X_{1j}+X_{2j}}{200} = \frac{X_{\circ j}}{200}$$

og dermed under uavhengighets-antagelsen

$$p_{ij}^{*} = p_{i\circ}^{*}\, p_{\circ j}^{*} = \frac{X_{i\circ}\, X_{\circ j}}{200^2}$$

c) Kji-kvadrat observatoren blir

$$X^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(X_{ij} - E_{ij})^2}{E_{ij}}$$

der

$$E_{ij} = 200 \cdot p_{ij}^{\circ} = \frac{X_{i\cdot} X_{\cdot j}}{200}$$

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

Vi vil forkaste H_0 hvis X^2 er tilstrekkelig stor, siden vi vil vente at det er forskjell på X_{ij}-ene og E_{ij}-ene hvis H_0 er gal

Ut fra tabellen i oppgaven finner vi

$$E_{11} = \frac{75 \cdot 60}{200} = 22.5$$

$$E_{12} = \frac{75 \cdot 40}{200} = 45.0$$

$$E_{13} = \frac{75 \cdot 20}{200} = 7.5$$

$$E_{21} = \frac{125 \cdot 60}{200} = 37.5$$

$$E_{22} = \frac{125 \cdot 120}{200} = 75.0$$

$$E_{23} = \frac{125 \cdot 20}{200} = 12.5$$

Kji-kvadratobservatoren blir

$$X^2 = 6.22$$

Av kji-kvadrattabellen finner vi at $X^2_{0.05,2} = 5.99$ og $X^2_{0.025,5} = 7.38$

Sidan X^2 får en verdi mellom disse, er P-verdien mellom 2.5% og 5%. Vi forkaster derfor H_0, så dataene tyder på at det er avhengighet mellom mors alder og enegget/toegget.

d) Likelihooden er generelt gitt som

$$L(p_{11}, p_{12}, \dots, p_{23}) = K \cdot \prod_{i=1}^{2} \prod_{j=1}^{3} p_{ij}^{x_{ij}}$$

Sannsynlighetskvoten er generelt gitt som

$$\Lambda = \frac{\max_{H_0} L}{\max L}$$

og vi forkaster når Λ er tilstrekkelig liten.

Her har vi at

$$\Lambda = \frac{L(p_{11}^*, p_{12}^*, \dots, p_{23}^*)}{L(\hat{p}_{11}, \hat{p}_{12}, \dots, \hat{p}_{23})}$$

$$= \prod_{i=1}^{2} \prod_{j=1}^{3} \left(\frac{p_{ij}^*}{\hat{p}_{ij}} \right)^{x_{ij}}$$

Vi ser nå på $-2 \log \Lambda$ som er tilnærmet kji-kvadrat fordelt med 2 frihetsgrader under H_0

Vi finner

$$-2 \log \Lambda = 2 \sum_{i=1}^{2} \sum_{j=1}^{3} X_{ij} \log \left(\frac{\hat{p}_{ij}}{p_{ij}^{*}} \right)$$

$$= 2 \sum_{i=1}^{2} \sum_{j=1}^{3} X_{ij} \log \left(\frac{X_{ij}}{200\, p_{ij}^{*}} \right)$$

$$= 2 \sum_{i=1}^{2} \sum_{j=1}^{3} X_{ij} \log \left(\frac{X_{ij}}{E_{ij}} \right)$$

For tallene i oppgaven får testobservatoren verdien

$$-2 \log \Lambda = 6.19$$

Dette er nesten det samme som kji-kvadrat observatoren

(I ekstraoppgave 11 fra Storviks hefte er det forklart hvorfor de to testobservatorene blir tilnærmet like.)