TMA4315 Generaliserte lineære modeller
Eksamen 29. november 2007
Løsningsskisse

Oppgave 1: Innfører notasjon

$a \in \{1, 2, \ldots, 9\}$

$a = 1$: alder 40-44, $a = 2$: alder 45-49, osv.

$s \in \{1, 2, 3, 4\}$

$s = 1$: kun sigaretter
$s = 2$: både sigaretter og sigarer/pipe
$s = 3$: ikke-røyk
$s = 4$: kun sigarer/pipe

a) For $x_1 = 3$ og $x_2 = 1$ er den estimerte modellen

$$\ln \hat{\mu} = -3.6322 + 0.98039 - 0.04781 = -2.69962$$

$$\underline{\underline{\hat{\mu}}} = e^{-2.69962} = \underline{\underline{0.06723}}$$

frihetsgrader = # observasjoner − # estimerte parametre

$$= 36 - 12 = \underline{\underline{24}}$$

Finner fra tabell kritisk verdi i en χ^2_{24}-fordeling

$$z_{0.05, 24} = 36.42$$

Siden $D = 21.487 < 36.42$ tilpasser modellen de observerte dataene bra.

b) Hypotesetest. H_0: modell 2 mot H_1: modell 1

$$\Delta D = D_2 - D_1 = 3910.7 - 24.487 = 3886.213$$

$$df = df_2 - df_1 = 32 - 24 = 8$$

Ser at vi får en p-verdi ≈ 0, og forkaster H_0. Dvs. foretrekker modell 1 fremfor modell 2

Hypotesetest. H_0: modell 3 mot H_1: modell 1

$$\Delta D = D_3 - D_1 = 75.734 - 21.487 = 54.247$$

$$df = df_3 - df_1 = 31 - 24 = 7$$

Igjen blir p-verdien ≈ 0 og vi forkaster H_0

⊃: Modell 1 er „den beste" modellen

c) $$\ln \mu(a,s) = \beta_1 + \sum_{i=2}^{9} \beta_i I(a=i) + \sum_{i=1}^{3} \gamma_i I(s=i)$$

$$\ln r(a,s_1,s_2) = \ln \mu(a,s_1) - \ln \mu(a,s_2)$$

$$= \sum_{i=1}^{3} \gamma_i \left[I(s_1=i) - I(s_2=i) \right]$$

$$r(a,s_1,s_2) = \exp\left\{ \sum_{i=1}^{3} \gamma_i \left[I(s_1=i) - I(s_2-i) \right] \right\}$$

Vi ser at $r(a,s_1,s_2)$ ikke avhenger av a. Dette skylles at modellen ikke har med interaksjonsledd mellom røykestatus og alder.

$$r(a,4,1) = e^{\gamma_1(0-1)} = e^{-\gamma_1}$$

$$\hat{r}(a,4,1) = e^{-0.36915} = \underline{\underline{0.6913}}$$

90%-konfidensintervall for γ_1 er

$$[+0.36915 \pm 1.64 \cdot 0.03791]$$

90%-konfidensintervall for $r(a,4,1) = e^{-\gamma_1}$ blir

dermed

$$e^{-0.36915 \pm 1.64 \cdot 0.03791}$$

$$= \underline{\underline{[0.6496,\ 0.7357]}}$$

Da intervallet ikke inneholder 1 er det signifikant forskjell. Det er en fordel å røyke sigar/pipe fremfor sigaretter.

Oppgave 2

$$f(y;\theta) = \frac{\theta^2}{2} y\, e^{-\theta y} = \exp\{-\theta y + 2\ln\theta + \ln y - \ln 2\}$$

$$= \exp\{a(y)b(\theta) + c(\theta) + d(y)\}$$

med

$$a(y) = y$$
$$b(\theta) = -\theta$$
$$c(\theta) = 2\ln\theta$$
$$d(y) = \ln y - \ln 2$$

$\supset$: Y_i-ene har fordelinger fra samme eksponensielle familie.

$$E[Y_i] = -\frac{c'(\theta)}{b'(\theta)} = -\frac{2\cdot\frac{1}{\theta}}{-1} = \underline{\underline{\frac{2}{\theta}}}$$

$$\mathrm{Var}[Y_i] = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{(b'(\theta))^3}$$

$$= \frac{0 - 2\left(-\frac{1}{\theta^2}\right)(-1)}{(-1)^3} = \underline{\underline{\frac{2}{\theta^2}}}$$

b) $$U_j(\beta) = \sum_{i=1}^{N}\left[\frac{y_i - \mu_i}{Var(Y_i)}\, x_{ij}\, \frac{\partial \mu_i}{\partial \eta_i}\right]$$

Vi har $\frac{\partial \eta}{\partial \mu} = \frac{1}{\mu} \Rightarrow \frac{\partial \mu}{\partial \eta} = \mu = e^{x^T\beta}$

og

$$\left.\begin{array}{l} \ln\mu = x^T\beta \\ \mu = \frac{2}{\theta} \end{array}\right\} \Rightarrow \mu = \frac{2}{\theta} = e^{x^T\beta} \Rightarrow \theta = 2e^{-x^T\beta}$$

slik at vi får:

$$U_j(\beta) = \sum_{i=1}^{N}\left[\frac{y_i - \frac{2}{\theta_i}}{\frac{2}{\theta_i^2}}\, x_{ij}\, e^{x_i^T\beta}\right]$$

$$= \sum_{i=1}^{N}\left[\frac{y_i - e^{x_i^T\beta}}{\frac{1}{2}e^{2x_i^T\beta}}\, x_{ij} \cdot e^{x_i^T\beta}\right]$$

$$= 2\sum_{i=1}^{N}\left[\frac{y - e^{x_i^T\beta}}{e^{x_i^T\beta}} \cdot x_{ij}\right]$$

$$\mathcal{J}_{jk}(\beta) = \sum_{i=1}^{N}\left[\frac{x_{ij}\, x_{ik}}{Var(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2\right]$$

$$= \sum_{i=1}^{N}\left[\frac{x_{ij}\, x_{ik}}{\frac{2}{\theta_i^2}} \cdot \left(e^{x_i^T\beta}\right)^2\right]$$

$$\mathcal{J}_{jk}(\beta) = \sum_{i=1}^{N} \frac{x_{ij}\, x_{ik}}{\frac{1}{2} e^{2x_i^T\beta}} \cdot e^{2x_i^T\beta} = 2\sum_{i=1}^{N} x_{ij}\, x_{ik}$$

$\supset$: $\underline{\underline{\mathcal{J}(\beta) = 2X^TX}}$ (merk: ikke funksjon av β).

Rekursjonsligningen blir

$$\mathcal{J}\beta^{(m)} = \mathcal{J}\beta^{(m-1)} + U(\beta^{(m-1)})$$

$$\underline{\underline{\beta^{(m)} = \beta^{(m-1)} + \mathcal{J}^{-1}U(\beta^{(m-1)})}}$$

c) $$L = \prod_{i=1}^{N} \frac{\theta_i^2}{2} y_i\, e^{-\theta_i y_i}$$

$$l = \sum_{i=1}^{N}\left[2\ln\theta_i + \ln y_i - \theta_i y_i - \ln 2\right]$$

For vår modell får man da

$$l(\hat{\beta}) = \sum_{i=1}^{N}\left[2\ln\hat{\theta}_i + \ln y_i - \hat{\theta}_i y_i - \ln 2\right]$$

Benytter $\hat{\mu}_i = \frac{2}{\hat{\theta}_i} \Leftrightarrow \hat{\theta}_i = \frac{2}{\hat{\mu}_i} = \frac{2}{\hat{y}_i}$

og får

$$l(\hat{\beta}) = \sum_{i=1}^{N} \left[2\ln 2 - 2\ln \hat{y}_i + \ln y_i - 2\frac{y_i}{\hat{y}_i} - \ln 2 \right]$$

For mettet modell har vi en parameter θ_i for hver obs. y_i. Finner dermed $\hat{\theta}_i^{max}$ ved

$$\frac{\partial}{\partial \theta_i}\left(2\ln\theta_i + \ln y_i - \theta_i y_i - \ln 2 \right) = 0$$

$$\frac{2}{\theta_i} - y_i = 0 \Rightarrow \hat{\theta}_i^{max} = \frac{2}{y_i}$$

slik at

$$l(\hat{\beta}_{max}) = \sum_{i=1}^{N} \left[2\ln\frac{2}{y_i} + \ln y_i - \frac{2}{y_i}\cdot y_i - \ln 2 \right]$$

$$= \sum_{i=1}^{N} \left[2\ln 2 - 2\ln y_i + \ln y_i - 2 - \ln 2 \right]$$

$$= \sum_{i=1}^{N} \left[\ln 2 - \ln y_i - 2 \right]$$

$$D = 2\left[l(\hat{\beta}_{max}) - l(\hat{\beta}) \right]$$

$$= 2\sum_{i=1}^{N} \Big[\cancel{\ln 2} - \ln y_i - 2 - 2\cancel{\ln 2} + 2\ln\hat{y}_i - \ln y_i + 2\frac{y_i}{\hat{y}_i} + \cancel{\ln 2} \Big]$$

$$= 2\sum_{i=1}^{N} \left[2\ln\frac{\hat{y}_i}{y_i} - 2\left(1 - \frac{y_i}{\hat{y}_i}\right) \right]$$

$$= 4\sum_{i=1}^{N} \left[\ln\frac{\hat{y}_i}{y_i} + \frac{y_i - \hat{y}_i}{\hat{y}_i} \right]$$

Devians residualene blir

$$d_i = 2\, \text{sign}(y_i - \hat{y}_i) \sqrt{\ln \frac{\hat{y}_i}{y_i} + \frac{y_i - \hat{y}_i}{\hat{y}_i}}$$

d)

$$r_i = \frac{o_i - e_i}{\sqrt{\sigma_i^2}} = \frac{y_i - \frac{2}{\hat{\theta}_i}}{\sqrt{\frac{2}{\hat{\theta}_i^2}}}$$

$$= \frac{y_i - 2/\frac{2}{\hat{y}_i}}{\sqrt{2/\left(\frac{2}{\hat{y}_i}\right)^2}} = \frac{y_i - \hat{y}_i}{\sqrt{2\hat{y}_i^2}} = \sqrt{2}\, \frac{y_i - \hat{y}_i}{\hat{y}_i}$$

$$X^2 = 2\sum_{i=1}^{N} \left(\frac{y_i - \hat{y}_i}{\hat{y}_i}\right)^2$$

Rekkeutvikler $f(y) = \ln \frac{\hat{y}}{y} + \frac{y - \hat{y}}{\hat{y}}$

$$= \ln \hat{y} - \ln y + \frac{y - \hat{y}}{\hat{y}}$$

omkring $\hat{y}$. Får $f(\hat{y}) = 0$

$$f'(y) = -\frac{1}{y} + \frac{1}{\hat{y}} \quad , \quad f'(\hat{y}) = -\frac{1}{\hat{y}} + \frac{1}{\hat{y}} = 0$$

$$f''(y) = +\frac{1}{y^2} \quad , \quad f''(\hat{y}) = \frac{1}{(\hat{y})^2}$$

Dvs.

$$f(y) \approx f(\hat{y}) + f'(\hat{y})(y - \hat{y}) + \frac{1}{2} f''(\hat{y})(y - \hat{y})^2$$

$$= 0 + 0 + \frac{1}{2} \frac{1}{(\hat{y})^2} (y - \hat{y})^2$$

slik at

$$D \approx 4 \cdot \sum_{i=1}^{N} \frac{1}{2} \frac{1}{(\hat{y}_i)^2} \cdot (y_i - \hat{y}_i)^2 = \underline{\underline{2 \sum_{i=1}^{N} \left(\frac{y_i - \hat{y}_i}{\hat{y}_i} \right)^2}}$$