TMA 4315 14. desember 2006

Oppgave 1

a) [Se boka s. 144]

Respons er ordinell, multinomisk.
Prop. odds modell antar en latent kontinuerlig variabel som måler CGI.

$$\ln \frac{\pi_0}{\pi_1+\pi_2} = \beta_{00} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

$$\ln \frac{\pi_0+\pi_1}{\pi_2} = \beta_{01} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Siden den ~~inverse~~ inverse av logit-funksjonen er den logistiske funksjon har vi

$$\pi_0 = \frac{e^{\beta_{00}+\underset{\sim}{x}^T \underset{\sim}{\beta}}}{1+e^{\beta_{00}+\underset{\sim}{x}^T \underset{\sim}{\beta}}}$$

$$\pi_0+\pi_1 = \frac{e^{\beta_{01}+\underset{\sim}{x}^T \underset{\sim}{\beta}}}{1+e^{\beta_{01}+\underset{\sim}{x}^T \underset{\sim}{\beta}}}$$

$$\Rightarrow \underline{\underline{\pi_1 = \text{— '' —} - \pi_0}}$$

og endelig

$$\underline{\underline{\pi_2 = 1 - \pi_0 - \pi_1}}$$

b) (*) $$\frac{P(CGI \le 0 \mid \underset{\sim}{x})}{P(CGI > 0 \mid \underset{\sim}{x})} = \frac{\pi_0}{\pi_1 + \pi_2} = e^{\beta_{00} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3}$$

Ser at hvis x_3 økes med 1 (og x_1, x_2 ikke endres) vil dette multipliseres med e^{β_3}. Tilsvarende for $P(CGI \le 1 \mid \underset{\sim}{x}) / P(CGI > 1 \mid \underset{\sim}{x})$

Fortolkning: Hvis $e^{\beta_3} < 1$ (dvs $\beta_3 < 0$) vil økning i initiell CGI gi en reduksjon i odds for å være i tilstand 0 eller ≤ 1 etter behandlingen.

Ser av (*) at økningen i x_1 og x_2 medfører på samme måte multiplikasjon med e^{β_1}, e^{β_2}, henholdsvis.

Fortolkningen: Hvis f.eks. $e^{\beta_1} < 1$ vil det å gå over fra hver annen uke til hver tredje uke gi en reduksjon i odds for de "gode tilstandene" 0 og ≤ 1.

Hvis $e^{\beta_2} < 1$ vil ~~kvinner~~ menn ha mindre odds enn kvinner for å gå til de "gode tilstandene" 0 og ≤ 1.

c) Saturert modell:

Hver rad i tabellen har sin egen π-vektor, $(\pi_{i0}, \pi_{i1}, \pi_{i2})$. Modellen vil dermed ha $12 \cdot 2 = 24$ frie parametre siden $\pi_{i0} + \pi_{i1} + \pi_{i2} = 1$.

Fra pensum er

$$D = 2\left[l_{\text{saturert modell}} - l_{\text{aktuell modell}} \right]$$

der l_m betyr maksimert log likelihood for modell m

Det er vist i et notat fra siste forelesning (se web-side) at

$$D = 2 \sum_{i=1}^{N} \sum_{j=0}^{2} y_{ij} \ln \frac{y_{ij}}{\hat{y}_{ij}}$$

der $\hat{y}_{ij} = n_i \hat{\pi}_{ij}$ og $\underset{\sim}{y_i} = (y_{i0}, y_{i1}, y_{i2})$

er responsen for i-te linje i tabellen, mens $\hat{\pi}_{ij}$ er estimerte π_{ij} fra den aktuelle modellen.

df = ant. frie parametre i saturert modell
− – " – aktuell modell
$= 24 - 5 = \underline{\underline{19}}$

d) $B*K + B*I$ har lineære prediktorer

$$\beta_{0j} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3$$

(sett inn på høyre side i modellen i (a)).

Den har 7 parametre, dvs $df = 24 - 7 = \underline{\underline{17}}$

Antall parametre og ~~devianse~~ frihetsgrader for modellene:

	$K+I$	$B+K+I$	$B*I+K$	$B*K+B*I$
Param.	4	5	6	7
df	20	19	18	17

Tester $H_0: K+I$ mot hver av de andre modellene.

Mot $B+K+I$: $\Delta D = 10.64 - 10.56 = 0.08$, $\approx \chi^2_1$ under H_0
Forkast ikke!

Mot $B*I+K$: $\Delta D = 10.64 - 8.52 = 2.12$, $\approx \chi^2_2$ under H_0
Forkast ikke!

Mot $B*K+B*I$: $\Delta D = 10.64 - 8.33 = 2.31$, $\approx \chi^2_3$ under H_0
Forkast ikke!

Velger derfor modellen $K+I$ (dvs. behandl. intervall har ikke betydning).

e) $\widehat{e^{\beta_1}} = e^{\hat{\beta}_1} = e^{-0.21992} = \underline{0.8026}$

Konf int:

$\approx 95\%$ for β_1: $\hat{\beta}_1 \pm 1.96 \cdot \widehat{SD(\hat{\beta}_1)}$

$\approx 95\%$ for e^{β_1}: $e^{\hat{\beta}_1 \pm 1.96 \cdot \widehat{SD(\hat{\beta}_1)}}$

dvs $e^{-0.21992 \pm 1.96 \cdot 0.75607}$

dvs $[0.1824, 3.5324]$

Dette inneholder 1, så vi kan ikke forkaste $H_0: e^{\beta_1} = 1$ (dvs $\beta_1 = 0$) som betyr at behandlingsintervall ikke har signifikant betydning.
Dette stemmer med konklusjonen i (d)

$\widehat{e^{\beta_2}} = e^{\hat{\beta}_2} = e^{-2.15758} = \underline{0.1156}$

$\widehat{e^{\beta_3}} = e^{\hat{\beta}_3} = e^{-2.27249} = \underline{0.1031}$

Ser på en kvinne, init. COI = 5, injeksj. hver annen uke:
dvs $x_1 = 0$, $x_2 = 0$, $x_3 = 5$.

Fra (a) $$\hat{\pi}_0 = \frac{e^{\hat{\beta}_{00} + 5\hat{\beta}_3}}{1+e^{\hat{\beta}_{00}+5\hat{\beta}_3}}$$

$$= \frac{e^{8.47538 + 5\cdot(-2.27249)}}{1+e^{8.47538+5\cdot(-2.27249)}}$$

$$= \underline{\underline{0,0528}}$$

Standardavvik beregnes ved å Taylor-utvikle

$\hat{\pi}_0$ omkring (β_{00}, β_3)

La $$f(x,y) = \frac{e^{x+5y}}{1+e^{x+5y}}$$

Da er $$\frac{\partial f}{\partial x} = \frac{e^{x+5y}}{(1+e^{x+5y})^2} = f(1-f)$$

$$\frac{\partial f}{\partial y} = \frac{5e^{x+5y}}{(1+e^{x+5y})^2} = 5f(1-f)$$

Dermed

$$\hat{\pi}_0 \approx \pi_0 + \pi_0(1-\pi_0)(\hat{\beta}_{00} - \beta_{00})$$
$$+ 5\pi_0(1-\pi_0)(\hat{\beta}_3 - \beta_3)$$

slik at

$$\mathrm{var}(\hat{\pi}_0) \simeq \pi_0^2 (1-\pi_0)^2 \left[\mathrm{Var}(\hat{\beta}_{00}) + 25\,\mathrm{Var}(\hat{\beta}_3) + 10\,\mathrm{Cov}(\hat{\beta}_{00}, \hat{\beta}_3) \right]$$

Estimerer dette ved å sette $\hat{\pi}_0$ for π_0, og lese av $\mathrm{Var}(\cdot)$, $\mathrm{Cov}(\cdot,\cdot)$ fra R-utskrift. Dette gir

$$\widehat{SD(\hat{\pi}_0)} = 0.0528 \cdot 0.9472 \cdot \sqrt{11.1599 + 25 \cdot 0.4879 - 10 \cdot 1.8574}$$

$$= \underline{\underline{0.1094}}$$

Oppgave 2.

a) $f(y_i;\theta_i) = \theta_i e^{-\theta_i y_i} = e^{-\theta_i y_i + \ln \theta_i}$

som kommer fra familien $f(y;\theta) = e^{-\theta y + \ln \theta}$
Med notasjon fra boka:

$a(y) = y$, $b(\theta) = -\theta$, $c(\theta) = \ln\theta$, $d(y) = 0$.

Nå er $b'(\theta) = -1$, $b''(\theta) = 0$

$c'(\theta) = \frac{1}{\theta}$, $c''(\theta) = -\frac{1}{\theta^2}$

Dermed (fra boka)

$$\underline{\underline{\mu_i}} = -\frac{c'(\theta_i)}{b'(\theta_i)} = \underline{\underline{\frac{1}{\theta_i}}}$$

$$\sigma_i^2 = \frac{b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)}{(b'(\theta_i))^3} = \frac{1}{\theta_i^2}$$

b) $$l \underset{def}{=} \sum_{i=1}^{N} \ln f(y_i; \theta_i)$$

$$= \sum_{i=1}^{N} (-\theta_i y_i + \ln \theta_i) \quad (*)$$

Nå er

$$D = 2[l_{saturert} - l_{aktuell\ modell}]$$

Saturert modell: alle θ_i ulike

$$\frac{\partial l}{\partial \theta_i} = -y_i + \frac{1}{\theta_i}$$

Settes dette lik 0 fås $\hat{\theta}_i = \frac{1}{y_i}$

Dermed fra $(*)$: $l_{saturert} = \sum_{i=1}^{N} \left(-1 + \ln \frac{1}{y_i}\right)$

I GLM-modellen estimeres μ_i ved en $\hat{\mu}_i$ (som vi her ikke trenger noe uttrykk for), som her kalles $\hat{y}_i$. Dermed blir $l_{aktuell}$ gitt ved $\hat{\theta}_i = \frac{1}{\hat{y}_i}$ i $(*)$.

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

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

Dermed er

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

$$= 2\sum_{i=1}^{N} \left[\frac{y_i - \hat{y}_i}{\hat{y}_i} - \ln \frac{y_i}{\hat{y}_i} \right] \quad \text{Q.E.D.}$$

c) Pearson-res: $r_i \underset{def}{=} \dfrac{y_i - \widehat{E(y_i)}}{\sqrt{\widehat{Var(y_i)}}}$

$$= \frac{y_i - \hat{y}_i}{\hat{y}_i}$$

siden $Var\, y_i = E(y_i)^2$ for eksponensiell fordeling.

For å vise at $X^2 \approx D$ Taylor-utvikler vi

$$f(x) = \frac{x-a}{a} - \ln \frac{x}{a} \quad \text{omkring } x = a.$$

Vi har $f(a) = 0$, $f'(x) = \frac{1}{a} - \frac{1}{x}$ så $f'(a) = 0$,

$$f''(x) = \frac{1}{x^2} \text{ så } f''(a) = \frac{1}{a^2}$$

Dermed: $f(x) \approx \frac{1}{2} \cdot \frac{1}{a^2}(x-a)^2$

og $D \approx \sum_{i=1}^{N} \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i^{\,2}} = \sum_{i=1}^{N} r_i^2 = X^2$ Q.E.D.

ved å
sette inn i
oppgitt D i
oppgavetekst

d) H_0 svarer til en GLM der
alle θ_i er like, lik en θ.

Da er $l = -\theta \sum_{i=1}^{N} y_i + N \ln \theta$

som ved å derivere og sette lik 0 gir $\cancel{\theta = \frac{1}{N}\sum_{i=1}^{N} y_i}$

$$\hat{\theta} = \frac{1}{\frac{1}{N}\sum_{i=1}^{N} y_i} = \frac{1}{\bar{y}}$$

dvs $\hat{\theta}_i = \frac{1}{\bar{y}}$ for hver i, dvs $\hat{\mu}_i = \hat{y}_i = \bar{y}$

Da blir fra (b)

$$D = 2 \sum_{i=1}^{N} \left(\frac{y_i - \bar{y}}{\bar{y}} - \ln \frac{y_i}{\bar{y}} \right)$$

som er $\approx \chi^2_{N-1}$ under H_0

ant. param. i saturert modell

ant. param. i modell under H_0

(Vi forkaster H_0 hvis D er stor).

Tilsvarende blir $\chi^2 = \sum_{i=1}^{N} \left(\frac{y_i - \bar{y}}{\bar{y}}\right)^2$

en testobservator som er $\approx \chi^2_{N-1}$ under H_0.

Merk at vi kan skrive

$$\chi^2 = \frac{\sum_{i=1}^{N} (y_i - \bar{y})^2}{\bar{y}^2} = \frac{(N-1)s^2}{\bar{y}^2}$$

Under H_0 er både s^2 og $\bar{y}^2$ estimatorer for μ^2 (siden $\mathrm{Var}(Y) = E(Y)^2$ i eksponensial-fordelingen)

„Forkastning vil da skje hvis s^2 blir "for stor" i forhold til $\bar{y}^2$ som er en <u>intuitiv</u> testmetode.