UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i:	STK3100 — Innføring i generaliserte lineære modeller.
Eksamensdag:	Tirsdag 16. desember 2008.
Tid for eksamen:	09.00 – 12.00.
Oppgavesettet er på 3 sider.	
Vedlegg:	Tabell over normalfordelingen.
Tillatte hjelpemidler:	Godkjent kalkulator og formelsamling for STK1100/STK1110 og for STK1120.

Kontroller at oppgavesettet er komplett
før du begynner å besvare spørsmålene.

NB. Oppgavesettet består av en oppgave med 8 punkter.

Oppgave 1.

(a) Vis at punktsannsynlighetene i Poissonfordelingen kan skrives på formen $f(y;\theta) = c(y)\exp(\theta y - a(\theta))$.

Finn sammenhengen mellom forventningen μ i Poissonfordelingen og parameteren θ.

Gi eksplisitte uttrykk for funksjonene $a(\theta)$ og $c(y)$.

(b) Anta at $Y_1, \ldots, Y_n$ er uavhengige og Poissonfordelte med forventning $\mu_i = \exp(\alpha + \beta x_i)$ der α og β er regresjonsparametre og x_i kjente forklaringsvariable. Påvis at denne modellen kommer inn under rammen for generaliserte lineære modeller.

(c) Sett opp log-likelihood for dataene i punkt (b) og vis at scorefunksjonen kan skrives

$$U(\alpha,\beta) = \begin{pmatrix} U_1(\alpha,\beta) \\ U_2(\alpha,\beta) \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} 1 \\ x_i \end{pmatrix} (Y_i - \mu_i)$$

Finn også et uttrykk for forventet informasjonsmatrise.

(Fortsettes side 2.)

(d) Forklar hva en mettet modell er. Vis at maximum likelihood estimatene for $\mu_i, i = 1, \ldots, n$, i den mettede modellen blir $\tilde{\mu}_i = Y_i$.

(e) Finn på denne bakgrunn et uttrykk for deviansen ved en generalisert lineær modell med responser fra Poissonfordelingen.

Grei ut om tester som kan gjøres ved hjelp av devianser.

(f) Man velger å se bort fra antallet hendelser Y_i for hvert individ i og vil heller gjøre analysen basert på indikatorvariablen for at det var minst en skade, dvs. $Y_i' = I(Y_i > 0)$. La $\pi_i = \mathrm{P}(Y_i' = 1)$. Påvis at dette blir en generalisert lineær modell for binære data med samme lineære prediktor som i punkt (b) og linkfunksjon

$$g(\pi_i) = \log(-\log(1 - \pi_i)).$$

Hva kalles denne linkfunksjonen?

(g) Under er det gjengitt resultater fra en analyse av bilforsikringsdata der responsen er om forsikringstaker har rapportert en eller flere skader siste år. Modellen for dataene er en Poissonregresjonsmodell tilsvarende punkt (b), men dataene er analysert med binære responser tilsvarende den utledede modellen i punkt (f). Variablene som er brukt er bilførerens alder, kategorisert til 6 grupper (`agecat`), og bilens verdi i 10.000 dollar (`veh_value`) (brukt som en numerisk variabel slik at en bil som f.eks. er verdt 25.000 dollar er kodet som 2.5). Kun biler med verdi under 40.000 dollar er tatt med i analysen.

Beregn på basis av `R`-utskriften estimater for relativ forskjell i raten for skader mellom

(i) alderskategori 2 og alderskategori 1

(ii) alderskategori 5 og alderskategori 2

(iii) to biler verdt henholdsvis 25.000 og 5.000 dollar

(iv) to bilførere der den ene føreren er i alderskategori 2 og har en bil verdt 20.000 dollar, mens den andre føreren er i alderskategori 1 og har en bil verdt 10.000 dollar

```
Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)       -1.79067    0.05636 -31.773  < 2e-16 ***
factor(agecat)2 -0.21697    0.05846  -3.711 0.000206 ***
factor(agecat)3 -0.25327    0.05674  -4.464 8.06e-06 ***
factor(agecat)4 -0.27294    0.05663  -4.820 1.44e-06 ***
factor(agecat)5 -0.51762    0.06396  -8.093 5.83e-16 ***
factor(agecat)6 -0.47167    0.07183  -6.567 5.14e-11 ***
veh_value         0.11656    0.01874   6.220 4.96e-10 ***
```

(Fortsettes side 3.)

(h) Finn dessuten 95% konfidensintervall for relativ endring i raten for skader mellom

(i) alderskategori 2 og alderskategori 1

(iv) to bilførere der den ene føreren er i alderskategori 2 og har en bil verdt 20.000 dollar, mens den andre føreren er i alderskategori 1 og har en bil verdt 10.000 dollar

For (iv) trenger du å vite at korrelasjonskoeffisienten mellom estimerte regresjonskoeffisienter svarende til aldersgruppe 2 og bilens verdi ble estimert til -0.0259.

SLUTT