# UNIVERSITETET I OSLO

## Det matematisk-naturvitenskapelige fakultet

| | |
|---|---|
| Eksamen i: | ST 202A — Generaliserte lineære modeller. |
| Eksamensdag: | Fredag 13. desember 1996. |
| Tid for eksamen: | 09.00 – 15.00. |
| Oppgavesettet er på 4 sider. | |
| Vedlegg: | Ingen. |
| Tillatte hjelpemidler: | Formelsamlinger for ST 101, ST 102, ST 103, lommeregner. |

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

## Oppgave 1.

I en undersøkelse blant 786 tilfeldig valgte yrkesaktive kvinner ble det spurt om fagforeningsmedlemskap og to forklaringsvariable som måler yrkeserfaring og klasse. De har hver tre kategorier; yrkeserfaring: Mindre enn tre år, tre til 10 år, mer enn ti år og klasse: Arbeidere, funksjonærer på lavere nivå og funksjonærer på høyere nivå.

La $\pi_{ij}$ være sannsynligheten for fagforeningsmedlemsskap for $i$-te nivå av faktoren yrkeserfaring og $j$-te nivå av faktoren klasse. Tilpasning av en model med logit link til datene gir følgende derivansanalysetabell

| | f.gr | Devians |
|---|---|---|
| 1 | 8 | 57.5 |
| erfaring | 6 | 28.5 |
| klasse | 6 | 35.8 |
| erfaring + klasse | 4 | 5.7 |

a) Vil du si det er noe signifikant samspill mellom erfaring og klasse?
Er hovedeffektene signifikante?
Du trenger ingen tabell for å svare på dette.

(Fortsettes side 2.)

b) For modellen med hovedeffekter gav tilpasningen følgende estimater for parametrene

| Parameter | Estimat | Standardfeil |
|---|---|---|
| 1 | 0.25 | 0.09 |
| ERF(1) | –0.61 | 0.12 |
| ERF(2) | 0.10 | 0.10 |
| KLASSE(1) | 0.03 | 0.13 |
| KLASSE(2) | –0.41 | 0.10 |

Her er modellen skrevet

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \mu + \mu_{1(i)} + \mu_{2(j)} \qquad i,j = 1,2,3$$

og det er valgt en parameterisering der $\sum_{i=1}^{3} \mu_{1(i)} = 0$ og $\sum_{j=1}^{3} \mu_{2(j)} = 0$.

En alternativ modelformulering er

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \delta + \alpha_i + \beta_j \qquad i = 1,2,3 \quad j = 1,2,3$$

der $\alpha_1 = \beta_1 = 0$.

Forklar sammenhengen mellom de to parameteriseringene, og angi hva estimatet for $\alpha_2$ og $\alpha_3$ vil bli.

Den nedre delen av korrelasjonsmatrisen mellom estimatorene er som følger, der variablene er ordnet som ovenfor

| | | | |
|---|---|---|---|
| 0.13 | | | |
| –0.32 | –0.41 | | |
| 0.35 | 0.04 | 0.00 | |
| –0.29 | 0.04 | –0.01 | -0.60 |

c) Hva er oddsforholdet for fagforeningsmedlemsskap mellom nyansatte og personer med lang yrkeserfaring? Beregn et 95% konfidensintervall.

d) Variabelen erfaring betegner en ordinal variable. Det er derfor rimelig å gi de tre kategoriene skårene 1, 2 og 3. Forklar hvordan en kan estimere en modell med bare hovedeffekter der denne informasjonen benyttes. Skriv ut eksplisitte uttrykk for de ni prediktorene som du vil benytte.

e) Forklar hvordan du tester modellen fra punkt d).

f) Uttrykk oddsforholdene for fagforeningsmedlemsskap mellom de tre gruppene arbeidserfaring i modellen fra punkt d). Kommenter resultatet.

*(Fortsettes side 3.)*

# Oppgave 2.

La $Y$ være en Poissonfordelt variabel med forventning $\mu$.

a) Bestem den momentgenererende funksjonen $M_Y(t) = E[\exp(tY)]$ og den kumulantgenererende funksjonen $K_Y(t) = \log(M_Y(t))$. Finn variansen $\sigma^2$ og skjevheten $E[(Y-\mu)/\sigma]^3$.

b) Beregn den kumulantgenererende funksjonen til $(Y-\mu)/\sigma$ og bestem grensen når $\mu \to \infty$. Kommenter resultatet.

c) Sett $Z = \sqrt{Y}$. Finn den tilnærmede fordelingen til $Z - \sqrt{\mu}$ når $\mu \to \infty$ og kommenter resultatet.

La i resten av oppgaven $Y_1, \ldots, Y_n$ være uavhengige Poissonfordelte variable med forventning $\mu_1, \ldots, \mu_n$, der $\mu_1, \ldots, \mu_n$ avhenger av observerte kovariater $x_1, \ldots, x_n$.

d) Anta at $\alpha + \beta x_i = \sqrt{\mu_i}\ \ i = 1, \ldots, n$ slik at linkfunksjonen er kvadratroten, og prediktoren $\eta_i = \alpha + \beta x_i\ \ i = 1, \ldots, n$.
Utled ligningene til bestemmelse av sannsynlighetsmaksimeringsestimatorene (SME) $(\widehat{\alpha}, \widehat{\beta})$ i dette tilfellet.

e) Hva er den tilnærmede fordelingen til $(\widehat{\alpha}, \widehat{\beta})$? Angi spesielt kovariansmatrisen i grensefordelingen. Under hvilke betingelser gjelder tilnærmelsen?

f) Forklar hvordan skåringsalgoritmen for å bestemme SME ser ut i dette tilfellet.

g) Bestem deviansen til denne modellen, og bruk ligningene til bestemmelse av SME til å forenkle uttrykket.

# Oppgave 3.

Vi skal i denne oppgaven se på en $3 \times 3$ tabell, der vi oppfatter de observerte verdiene $y_{11}, \ldots, y_{33}$ som realisasjoner av en eller flere multinomisk fordelte tilfeldige variable, eller som realisasjoner av uavhengige Poissonfordelte tilfeldige variable.

a) Vi ser først på tilfellet der hver rad oppfattes som en realisasjon av multinomiske variable, og radsummene derfor er størrelser som er fastlagt ved opplegget av undersøkelsen.

*(Fortsettes side 4.)*

Anta at sannsynlighetene har formen

$$\pi_{ij} = \frac{\exp(\beta_{ij})}{\sum_{j=1}^{3} \exp(\beta_{ij})} \qquad j = 1,2,3 \quad i = 1,2,3$$

der $\beta_{11} = \beta_{21} = \beta_{31} = 0$.

Forklar hvordan sannsynlighetsmaksimeringsestimatorene for $\beta$'ene bestemmes ved å tilpasse en passende Poissonmodell.

b) Anta nå at bare totalsummen i tabellen kan betraktes som fastlagt. Hvordan må du modifisere framgangsmåten fra punkt a)?

c) Anta nå at det er tilpasset en Poissonmodell til $y_{11}, \ldots, y_{33}$ der $\log(\mu_{ij}) = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{12(ij)}$ med bibetingelser $\sum_{i=1}^{3} \mu_{1(i)} = \sum_{j=1}^{3} \mu_{2(j)} = 0$ og $\sum_{j=1}^{3} \mu_{12(ij)} = 0$, $\sum_{i=1}^{3} \mu_{12(ij)} = 0 \quad i,j = 1,2,3.$

Forklar hvilke av størrelsene $\mu, \mu_{1(1)}, \ldots, \mu_{12(33)}$ som er fastlagt ved opplegget av undersøkelsen hvis

(i) totalsummen er gitt.

(ii) radsummene er gitt.

SLUTT