



Bokmål

Faglig kontakt under eksamen
John Tyssedal tlf. 73593534

EKSAMEN I TMA4315 GENERALISERTE LINEÆRE MODELLER

Fredag 9. desember 2005

Tid: 09.00 – 13.00

Tillatte hjelpemidler: Alle trykte og håndskrevne hjelpemidler. Alle kalkulatorer er tillatt. Sensur: 9. januar 2005.

Oppgave 1

I en studie av hesteskokrabber fant en at hver hunnkrabbe alltid bodde sammen med en hannkrabbe. I tillegg hadde flere hunnkrabber en eller flere hannkrabber like i nærheten, kalt satellitter. Det var av interesse å finne ut hvilke egenskaper ved hunnkrabbene som særlig tiltrakk seg hannkrabber. For hver hunnkrabbe ble det registrert hvor mange satellitter hun hadde, bredden på skallet og fargen på skallet delt inn i kategoriene lyst, middels lyst, middels mørkt og mørkt. Vi innfører nå følgende variabler:

Y = antall satellitter for hver hunnkrabbe

$$Y_1 = \begin{cases} 1 & \text{dersom en hunnkrabbe har satellitter} \\ 0 & \text{elles} \end{cases}$$

X_1 = bredde av skallet

$$X_2 = \begin{cases} 1 & \text{dersom skallet er lyst} \\ 0 & \text{elles} \end{cases}$$

$$X_3 = \begin{cases} 1 & \text{dersom skallet er middels lyst} \\ 0 & \text{elles} \end{cases}$$

$$X_4 = \begin{cases} 1 & \text{dersom skallet er middels mørkt} \\ 0 & \text{elles} \end{cases}$$

Innsamlet datamateriale for 20 hunnkrabber er gitt nedenfor:

Krabbenr.	y_i	y_{1i}	x_{1i}	x_{2i}	x_{3i}	x_{4i}
1	8	1	28.3	0	1	0
2	4	1	26.0	0	0	1
3	0	0	21.0	0	0	0
4	3	1	25.0	1	0	0
5	8	1	25.7	0	1	0
6	6	1	27.5	0	1	0
7	5	1	26.1	1	0	0
8	4	1	28.9	0	0	1
9	4	1	22.9	0	1	0
10	3	1	26.2	0	0	1
11	8	1	30.0	0	1	0
12	3	1	26.2	0	1	0
13	4	1	25.4	0	1	0
14	0	0	27.5	0	0	0
15	3	1	27.0	0	0	0
16	1	1	24.5	0	1	0
17	1	1	27.3	0	0	1
18	0	0	22.0	0	0	1
19	2	1	30.2	1	0	0
20	3	1	26.0	0	1	0

For å finne ut om bredden på skallet påvirket sannsynligheten for at en hunnkrabbe hadde satellitter ble det utført en logistisk regresjon med Y_1 som responsvariabel. En utskrift med programpakken R er gitt nedenfor. I spørsmålene a), b) og c) vil du ha bruk for denne utskriften.

```
glm(formula = y1 ~ x1, family = binomial)
```

```
Deviance Residuals:
```

```
    Min       1Q   Median       3Q      Max
-2.5764  0.1595  0.3649  0.4483  1.0678
```

```
Coefficients:
```

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -14.7637      8.6268  -1.711  0.0870 .
x1           0.6562      0.3518   1.865  0.0621 .
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 16.908  on 19  degrees of freedom
Residual deviance: 12.099  on 18  degrees of freedom
AIC: 16.099
```

- a) Sett opp en modell for sannsynligheten for at en hunnkrabbe har satellitter. Vurder ut i fra utskriften om modellen har tilfredstillende tilpasning og om parametrene er signifikante. Bruk 10% signifikansnivå i vurderingene. Hva blir estimert sannsynlighet for at en hunnkrabbe med skallbredde 25 cm har satellitter?
- b) Hva blir estimatet for oddsen når skallbredden er 25 cm? Finn estimert multiplikativ forandring i oddsen når skallbredden forandrer seg med en enhet. Uttled et tilnærmet standardavvik for estimatoren for denne forandringen. Estimer hvilken skallbredde som gir en estimert sannsynlighet på 0.9 for å ha satellitter.

Definer $\sigma_i^2 = \pi_i(1 - \pi_i)$ der $\pi_i = P(Y_{1i} = 1)$, $i = 1, 2, \dots, 20$.

- c) Vis ut fra kjente resultat at teoretisk kovariansmatrise for parametrene er gitt

$$\text{ved} \begin{bmatrix} \sum_{i=1}^{20} \sigma_i^2 & \sum_{i=1}^{20} \sigma_i^2 x_{1i} \\ \sum_{i=1}^{20} \sigma_i^2 x_{1i} & \sum_{i=1}^{20} \sigma_i^4 x_{1i}^2 \end{bmatrix}^{-1}$$

Estimert kovariansmatrise for de to koeffisientene er gitt ved: $\begin{bmatrix} 74.42 & -3.02 \\ -3.02 & 0.124 \end{bmatrix}$

Finn et tilnærmet 95 % konfidensintervall for oddsen når skallbredden er 25 cm.

En fant det rimelig at antall satellitter for hver hunnkrabbe var Poisson fordelt og en ønsket difor å utføre en Poissonregresjon med log link for å finne ut om det var noen sammenheng mellom forventet antall satellitter og bredden og fargen på skallet til hunnkrabben. En utskrift med programpakken R er gitt nedenfor. I spørsmålene d), e) og f) vil du ha bruk for denne utskriften.

```
glm(formula = y ~ x1 + x2 + x3 + x4, family = poisson)
```

```
Deviance Residuals:
```

Min	1Q	Median	3Q	Max
-1.71150	-1.07231	0.07435	0.54191	1.50426

```
Coefficients:
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.44922	1.60729	-2.146	0.0319 *
x1	0.13426	0.05737	2.340	0.0193 *
x2	0.96755	0.66514	1.455	0.1458
x3	1.49468	0.59747	2.502	0.0124 *
x4	0.77971	0.64623	1.207	0.2276

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

```
' ' 1
```

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 42.114 on 19 degrees of freedom
Residual deviance: 21.990 on 15 degrees of freedom
AIC: 85.579

- d) Skriv opp den estimerte modellen og vurder signifikansen av de estimerte koeffisientene. Utfør også en test for å vurdere tilpasningen av modellen. Bruk 5% signifikansnivå i vurderingene. Hva blir konklusjonen?
- e) Lag en skisse som illustrerer sammenhengen mellom forventet antall satellitter og skallbredde for hver av de 4 fargene. Ta hensyn til de vurderinger av signifikans som ble utført i Id). Estimatet av multiplikativ forandring i forventet antall satellitter når skallbredden forandrer seg med 1 cm blir ifølge denne modellen uavhengig av farge. Forklar hvorfor. Finn også et 95% konfidensintervall for denne forandringen.
- f) Finn et estimat for hvilken skallbredde som gjør forventet antall satellitter lik 5 for krabber med middels lyst skall. Forklar også (uten tallregning) hvordan en kan finne et tilnærmet 95 % konfidensintervall for denne skallbredden.

Oppgave 2

- a) Utled uttrykket for skalert devians for Binomisk fordeling.

Kovariansmatrisen for estimatoren av koeffisientene er gitt ved $(XWX)^{-1}$ der X er den vanlige designmatrisen i lineær regresjon.

- b) Finn forskjellen i W matrisa med log link og med identitets link for Poisson fordelte variable. Finn og forskjellen i W matrisa med logit link og identitetslink for binomisk fordelte variable.