STOCKHOLM UNIVERSITY
Department of Statistics
Johan Koskinen

Omtentamen, Generaliserade linjära modeller (GLM)
2024-01-04

Tid för examination: 14.00-18.00

Tillåtna hjälpmedel: Miniräknare, tilldelad tabellsamling, och medtaget formelblad (1 dubbel-sidigt A4)

Observera att inget formelblad kommer utdelas

Tentamen består av två delar. Del A: frågor 1-5; Del B: 6a och 6b. Frågorna i del A är tillsammans värda SJUTTIO (70) poäng. Fågan i del B är värd TRETTIO (30) poäng. Besvara **alla** frågor i del A. Besvara **endast en** av frågorna i del B. Du väljer vilken fråga i del B som du vill besvara och lämna in. Endast **ett** (1) svar kommer at rättas! Om du lämnar in svaret på två B-frågor förbehåller examinatorn sig rätten att inte rätta någon och du får noll (0) poäng på del B.

För godkännt krävs minst 50 utav det totalt 100 poängen. För detaljerade betygskriterier, se kursbeskrivningen.

Good luck!

Del A: besvara samtliga frågor

Problem 1. (14 poäng)
Antag att vi har utfall $i = 1, \ldots, n$, som oberoende följer $Y_i \sim Bern(p_i)$. Antag vidare den kanoniska länkfunktionen med linjär prediktor

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}$$

För $n = 4$, ges värden på x_i i Tabellen:

		x_{i1}	x_{i2}	x_{i3}
i	1	-2	1	1
	2	-1	1	0
	3	0	0	1
	4	1	0	0

Vidare, antag att $\beta_0 = -\frac{1}{2}$, $\beta_1 = 1$, $\beta_2 = -\frac{1}{2}$, samt $\beta_3 = \frac{1}{2}$.

(a) För $i = 1, \ldots, n$, beräkna η_i

(b) Beräkna p_i för $i = 1, \ldots, n$

(c) Utvärdera likelihoodfunktionen för de givna β_j samt värden

i	1	2	3	4
y_i	0	0	1	1

Problem 2. (14 poäng)
Antag förutsättningar som i (a), men med *probit* länkfunktionen.

(a) Vilken av observationerna $i = 3$ och $i = 4$, har större predicerad sannolikhet $\Pr(Y_i = 1 \mid x_{i1}, x_{i2}, x_{i3})$?

(b) För observation $i = 2$, antag att du kunde ändra värdet på x_{23} genom att lägga till ett nummer c. För vilket värde c är $\Pr(Y_2 = 1 \mid x_{21}, x_{22}, x_{23})$ lika med $\frac{1}{2}$?

(c) Utvärdera loglikelihood funktionen för y_i, $i = 1, \ldots, 4$

Problem 3. (14 poäng)
Pete har några favoritplatser i Albany, Western Australia, där han tycker om att fiska, såsom the Rocks, hans uppblåsbara båt, och Calm Bay. Antag att antalet fiskar Y_{ij} han fångar på dag i på plats j är $Po(\lambda_{ij})$. Antag vidare att Pete använder en kanonisk länk men linjär prediktor

$$\eta_{ij} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \gamma_1 x_{i2} x_{j2} + \gamma_2 x_{i3} x_{j2} + \gamma_3 x_{i2} x_{j3} + \gamma_4 x_{i3} x_{j3},$$

där x_{i1} är vattentemperaturen (Celcius) dag i, x_{i2} är cosinus av vindriktiningen (i radianer) dag i, x_{i3} är sinus av vindriktiningen (i radianer) dag i, och x_{jk} är dummy variabler som indikerar vilken plats han fiskar på, $k = 2, 3$.

Peter har skattat parametrarna

β_0	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4
5.00	-0.25	0.57	-0.28	-0.51	0.30	-1.50	1.75

(a) Antag att det är en dag då det är 16 grader varmt, och att vindriktningen är 317 grader (approximativt NV), vilket motsvarar 5.53 radianer, med $\cos(5.53) = .73$ och $\sin(5.53) = -.68$. Om Pete vill maximera det förväntade antalet fiskar han fångar (enligt modellen), vilken av platserna $j = 1, 2, 3$ skall han fiska på?

(b) En annan dag är det 12 grader varmt, men en vindriktningen på 225 grader (approximativt SV), vilket motsvarar 3.93 radianer, med $\cos(3.93) = -0.705$ och $\sin(3.93) = -0.71$. Vad är sannolikheten att Pete inte fångar någon fisk alls om han fiskar på plats 3?

(c) Med vindriktning som i (a), finn den temperatur när sannolikheten att fånga fisk är lika stor som att inte fånga fisk på plats 3?

Problem 4. (14 poäng)
För exemplet i problem 3 (b), anta istället att Y_i följer en negativ Binomialfördelning, med $k = 3$. Du använder inte den kanoniska länfunktionen, utan

$$\log \{E(Y_{ij} \mid x_{i1}, x_{i2}, x_{i3}, x_{j2}, x_{j3})\} = \eta_{ij}$$

Antag samma parameterskattningar som in problem 3.

(a) Vad är sannolikheten att Pete inte fångar någon fisk alls när han går till plats 3?

(b) Med antaganden som i (a), vad är sannolikheten att han fångar minst två fiskar plats 1?

Problem 5. (14 poäng)
Vi har observationer på vilka transportmedel som personer använder för att ta sig till Arlanda flygplats. Alternativen är

1. Taxi
2. Flygbussarna
3. Snälltåg/SJ
4. Arlanda express
5. Privat bil

För varje person och tillfälle har vi kovariaten

1. Inkomst (standardiserad)
2. Tidig resa
3. Ålder (standardiserad)
4. Avstånd till flygplats i log-kilometer
5. Reser med mer än en väska
6. Tjänsteresa

Vi får följande skattningar baserat på ett stickprov

	Flygbuss	Tåg	Arlanda express	Bil
0	0.50	0.10	0.45	0.10
1	-0.67	-0.10	-0.01	-0.07
2	-0.25	0.01	0.01	0.55
3	0.15	0.70	0.10	0.01
4	0.00	0.15	0.05	0.20
5	-0.17	-0.01	0.01	0.09
6	-0.85	-0.55	-0.05	-0.01

För en person med följande kovariat

x_0	x_1	x_2	x_3	x_4	x_5	x_6
1	.1	0	.1	3.4	0	0

vad är den predicerade sannolikheten att de kommer att ta taxi till flygplatsen?

Del B: besvara endast <u>en</u> av frågorna 6a och 6b

En fråga är värd 30 poäng

Problem 6.a
För $Y_i \sim Bern(p_i)$, med två prediktorer x_{i1} och x_{i2}, har vi följande data

i	y_i	x_{i1}	x_{i2}
1	1	3	1
2	1	3	0
3	0	2	0
4	1	0	1
5	0	1	1
6	0	0	0
7	1	1	1
8	1	3	1
9	1	2	0
10	0	3	1

Om vi antar logit länk, vilka av $\tilde{\beta}_j$ eller $\hat{\beta}_j$ maximerar (log-) likelihood funktionen? De givna värdena är

$$\tilde{\beta}_0 = -0.6627 \; , \tilde{\beta}_1 = 0.3766 \; , \tilde{\beta}_2 = 0.6991$$

samt

$$\hat{\beta}_0 = 0.5515 \; , \hat{\beta}_1 = -0.3751 \; , \hat{\beta}_2 = 0.6991$$

Problem 6.b
Antag att $Y_i \sim Po(\lambda_i)$ för n oberoende observationer med två prediktorer. Följande observationer är givna:

i	y	x_{i1}	x_{i2}
1	0	0	8
2	0	0	6
3	1	1	7
4	2	1	5
5	51	3	0

Med kanonisk länkfunktion fås följande skattningarna för två olika modeller

$$\tilde{\beta}_0 = -1.6718 \; , \tilde{\beta}_1 = 1.8691$$

samt

$$\hat{\beta}_0 = -1.13997 \; , \hat{\beta}_1 = 1.69194 \; , \hat{\beta}_2 = -0.06318$$

Testa

$$H_0 : \beta_2 = 0 \; , \; \text{mot } H_1 : \beta_1 \neq 0$$

på lämplig signifikansnivå