

Tentamen, Generaliserade linjära modeller (GLM) 2025-01-09

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. Frågan 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 att 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 godkänt krävs minst 50 utav de totalt 100 poängen. För detaljerade betygskriterier, se kursbeskrivningen.

Good luck!

Del A: besvara samtliga frågor

Problem 1. (14 poäng)

För den binära variabeln Y_i , antag följande model

$$\eta_i = \beta_0 + \beta_1 x_i$$

och en logistisk länkfunktion, där x_i är en dummy som indikerar om i tillhör grupp 2 ($x_i = 1$), eller grupp 1 ($x_i = 0$).

Vi har följande skattningar från ett stickprov

$$\hat{\beta}_0 = -0.25, \text{ och, } \hat{\beta}_1 = 0.25.$$

- (a) Givet en predicerad sannolikheten på 0.5, vilken grupp tillhör observationen?
- (b) För en observation i grupp 1, vad är det förväntade värdet på Y_i ?
- (c) Om vi slumpmässigt och oberoende drar 5 variabler från modellen, och antar att all tillhör grupp 2, vad är

$$\Pr \left(\sum_{i=1}^5 Y_i \leq 3 \right)$$

Problem 2. (14 poäng)

Vi har $n = 9$ observationer på en binär variable Y_i , med ett kovariat x_i som är en dummy som indikerar om i tillhör grupp 2 ($x_i = 1$), eller grupp 1 ($x_i = 0$). En skattning med *Probit* länk ger

$$\hat{\beta}_0 = -0.8416, \text{ och, } \hat{\beta}_1 = 0.8416.$$

- (a) Vad är den predicerade sannolikheten för ett lyckat utfall om $x_i = 0$?
- (b) En observation har predicerad sannolikhet $\frac{1}{2}$, hör den till grupp 1 eller 2?
- (c) Det predicerade, förväntade värdet på $\sum_{i=1}^9 Y_i$, för de givna observationerna, är 3 - hur många observationer kommer från grupp 1?

Problem 3. (14 poäng)

Antalet ordrar (köp eller försäljningar), Y_i , som en elektronisk aktiehandelsportal motar varje minut i , antas vara $Po(\lambda_i)$. För att ta hänsyn till tid på dygnet, delas dygnet in i 24-timmars intervall. För en observation, definieras

$$x_i = \exp \left\{ -\frac{(t_i - 12)^2}{20} \right\}$$

där t_i är timdelen för observationen. T.ex., för klockan 19.03 är $t_i = 19$.

För kanonisk länkfunktion antas den linjära prediktorn

$$\eta_i = \beta_0 + \beta_1 x_i.$$

(a) Antag att $\beta_0 = -.15$ och $\beta_1 = .5$, hur många orders kan du förvänta dig per minut mellan 15.00 och 16.00?

(b) Givet att det är morgon och att $E(Y_i) = 0.9348674$, hur många timmar är det kvar till lunch (12.00)? Antag att $\beta_0 = -.15$ och $\beta_1 = .5$

(c) Vad är sannolikheten att det kommer in mer än 4 order under de första 5-minuters intervallet efter klockan 14.00? Antag att $\beta_0 = -.175$ och $\beta_1 = .65$

Problem 4. (14 poäng)

För en delmängd av endast manliga repondenter i ett stort, representativt sample, undersöker vi utfaller Y_i

”Now thinking about the time since your 18th birthday (including the past 12 months) how many female partners have you had sex with?”

Antag att $Y_i \sim NB(k = 3, p_i)$ och att vi vill använda variablerna

j	beskrivning
1	ålder
2	är din religion protestantism (ja, $x_{i2} = 1$)
3	är din religion katolicism (ja, $x_{i3} = 1$)
4	är din religion judendom (ja, $x_{i4} = 1$)
5	är du gift ($x_{i5} = 1$) eller ej ($x_{i5} = 0$)
6	är du för ($x_{i6} = 1$) sexualundervisning i skolan eller emot ($x_{i6} = 0$)

Du använder inte den kanoniska länkfunktionen, utan länkfunktionen

$$\log(E(Y_i | x_{i1}, \dots, x_{i6})) = \eta_i.$$

Med den linjära prediktorn

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i1}^2 + \beta_3 x_{i2} + \beta_4 x_{i3} + \beta_5 x_{i4} + \beta_6 x_{i5} + \beta_7 x_{i6}$$

får du skattningarna

j	$\hat{\beta}_j$
0	0.105
1	0.119
2	-0.001
3	-0.981
4	-0.474
5	-0.844
6	-0.313
7	0.651

(a) Vad är standardavvikelsen för antal sexpartners i gruppen trettiåriga, ogifta, protestantiska män, som är emot sexualundervisning i skolorna, enligt modellen?

(b) För en trettiårig, ogift, protestantisk man, som är emot sexualundervisning i skolorna, vad är sannolikheten att han haft högst en partner?

Problem 5. (14 poäng)

Antag att en person kan välja mellan spel A, B, och C. För individer $i = 1, \dots, n$, har vi mått x_{i1} och x_{i2} , som speglar riskbenägenhet och emotionell mognad. För $j = A, B, C$ antar vi en länk funktion

$$g(\Pr(Y_{ij} = 1 \mid x_{i1}, x_{i2})) = \log \left(\frac{\Pr(Y_{ij} = 1 \mid x_{i1}, x_{i2})}{\Pr(Y_{iA} = 1 \mid x_{i1}, x_{i2})} \right) = \eta_{ij},$$

där

$$\eta_{ij} = \beta_{0j} + \beta_{1j}x_{i1} + \beta_{2j}x_{i2}.$$

Betrakta följande parameterar som givna

	β_{0j}	β_{1j}	β_{2j}
B	-0.45	0.27	0.09
C	-1.25	0.49	1.85

(a) Vad är sannolikheten, enligt modellen, att någon med $x_{i1} = 1.25$ och $x_{i2} = 0.5$ väljer spel C framför A och B?

(b) En person i med $x_{i1} = 1.25$, har predicerade sannolikheter

$$\Pr(Y_{iB}) = 0.1831, \text{ och } \Pr(Y_{iC}) = 0.6296,$$

vad är x_{i2} ?

Del B: besvara endast en av frågorna 6a och 6b

En fråga är värd 30 poäng.

Problem 6.a

För exemplet i Fråga 5, betrakta den betingade modellen för

$$Y_i = \begin{cases} 1, & \text{om } i \text{ väljer spel C givet att } i \text{ valt C eller A} \\ 0, & \text{om } i \text{ väljer spel A givet att } i \text{ valt C eller A} \end{cases}$$

definierat endast på den delmängd av personer som valt antingen spel A eller C.

(i) Tag fram

$$\text{logit}(\Pr(Y_i = 1 \mid x_{i1}, x_{i2})),$$

och uttryck detta i termer av β_{kj} och x_{ij} från fråga 5, $k = 0, 1, 2$

(ii) Vad är

$$\Pr(Y_i = 1 \mid x_{i1}, x_{i2})$$

för personen i 5(a)?

Problem 6.b

Betrakta följande data

i	y	x_1	x_2
1	0	0	8
2	2	0	6
3	0	0	3
4	6	1	7
5	2	1	2
6	4	1	5
7	30	3	1
8	35	3	0
9	22	3	3

Antag $Y_i \mid x_i \sim NB(k = 4, p_i)$ och länkfunktionen $\log(E(Y_i)) = \eta_i$, med

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

(i) Härled scorefunktionen.

(ii) Bestäm om

$$\hat{\beta}_0 = -0.55926, \hat{\beta}_1 = 1.28761, \text{ och } \hat{\beta}_2 = 0.09217$$

eller

$$\tilde{\beta}_0 = -0.75926, \tilde{\beta}_1 = 1.59761, \text{ och } \tilde{\beta}_2 = 0.09217$$

maximerar likelihood-funktionen.