

78185 Yleistetyt lineaariset mallit

Harjoitus 2, syksy 2014

1. Seuraavassa taulukossa on ristiintaulukoituna syntymäpaino (**matala** syntymäpaino, **normaali** syntymäpaino) ja vauvojen kuolleisuus (**kuollut** vuoden sisällä syntymästä, **elossa** vuoden kuluttua syntymästä). Estimoi riskisuhde ja ristitulosuhde (OR) kuolemislle vuoden sisällä syntymästä. Laske myös likimääräiset 95% luottamusvälit.

	Kuollut	Elossa	Yhteensä
Matala	618	4597	5215
Normaali	422	67093	67515
Yhteensä	1040	71690	72730

2. Prostatasyöpöpotilaan hoitomenetelmä riippuu suuresti siitä onko syöpä levinnyt ympäröiviin imusolmukkeisiin. Havaintoaineisto koostuu 53 prostata-syöpöpotilaasta, joilta on mitattu muun muassa seuraavien muuttujien arvot (muuttuja *nodalinv* on saatu leikkauksen jälkeen, muiden muuttujien arvot saatu ennen leikkausta):

nodalinv: Syövän levinneisyys imusolmukkeisiin. 1="kyllä", 0="ei".

grad: Neulalla otetun solunäytteen perusteella tehty arvio syövän asteesta. 1="vakava", 0="vähemmän vakava".

xray: Röntgenkuvan perusteella tehty arvio syövän asteesta. 1 = "vakava", 0 = "vähemmän vakava".

Seuraavassa taulukossa on ristiintaulukoituna muuttujat *nodalinv*, *grad* ja *xray*:

		xray	
		0	1
nodalinv	grad		
0	0	21	3
	1	8	1
1	0	5	4
	1	4	7

- (a) Laske vaaran $p = P(\text{nodalinv} = 1)$ estimaatti syövän levinneisyydelle ja vaaran likim. 95% luottamusväli, kun oletetaan, että populaatio on vaaran suhteen homogeeninen.
- (b) Laske vaarasuhteen ja ristitulosuhteen estimaatti syövän levinneisyydelle ja vastaavat 95% luottamusvälit, kun verrataan ryhmiä $\text{grad} = 0$ ja $\text{grad} = 1$.

- (c) Laske vaarasuhteen ja ristitulosuhteen estimaatti syövän levinneisyydelle ja vastaavat 95% luottamusvälit, kun verrataan ryhmiä $\{grad = 0, xray = 0\}$ ja $\{grad = 1, xray = 0\}$. Huom! Muuttuja $xray$ on vakio 0.
- (d) Laske vaarasuhteen ja ristitulosuhteen estimaatti syövän levinneisyydelle ja vastaavat 95% luottamusvälit, kun verrataan ryhmiä $\{grad = 0, xray = 1\}$ ja $\{grad = 1, xray = 1\}$. Huom! Muuttuja $xray$ on vakio 1.
3. Olkoot $\{y_1, \dots, y_{n_1}\}$, $\{y_{n_1+1}, \dots, y_{n_1+n_2}\}$ ja $\{y_{n_1+n_2+1}, \dots, y_{n_1+n_2+n_3}\}$ riippumattomat satunnaisotokset Bernoullijakaumista parametreilla p_1 , p_2 ja p_3 . Siis p_1 , p_2 ja p_3 ovat populaatioihin liittyvät vaarat. Tilannetta vastaa logistinen regressiomalli

$$\text{logit}(p_i) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3}, \quad i = 1, 2, 3,$$

jossa

$$x_{i2} = \begin{cases} 1 & \text{jos } i = 2, \\ 0 & \text{jos } i \neq 2. \end{cases} \quad \text{ja} \quad x_{i3} = \begin{cases} 1 & \text{jos } i = 3, \\ 0 & \text{jos } i \neq 3. \end{cases}$$

Johda parametrin β suurimman uskottavuuden estimaatti.

Vihje: 1) Satunnaisotokset ovat riippumattomia, 2) Laske parametrien p_1 , p_2 ja p_3 su-estimaatit ja käytä su-estimaatin invarianssiominaisuutta (Tilastollisen päättelyn kurssilta tuttu asia).