Frekvenssiaineistojen analyysi
Harjoitukset 25.10.2001

6.1. Tarkastellaan taulukkoa 5.1. Sovita taulukon aineistoon logistinen regressiomalli, kun kuorsauksen määrä pisteytetään luvuilla

1 (0, 2, 4, 6),
2 (0, 1, 2, 3),
3 (1, 2, 3, 4).

(a) Vertaile näiden kolmen eri pisteyksen vaikutusta parametrien estimaatteihin.
(b) Vertaile muunnoksen vaikutusta todennäköisyyksien $\hat{\pi}(x)$ estimaatteihin.
(c) Pistemäärät saadaan toisistaan lineaarisella muunnoksella $y = a + bx$. Mikä on siis lineaarisen muunnoksen vaikutus tuloksiin?

6.2. Tarkastellaan taulukkoa 5.5. Selitetään epämuodostumien esiintymistä alkoholin käytöllä.

(a) Laske epämuodostumien suhteelliset osuudet alkoholin kulutuksen eri tasoilla. Sovita aineistoon logitmalli, kun kulutuksen määrinä käytetään lukuja (0, 0.5, 1.5, 4, 7). Vertaa aineistosta laskettuja suhteellisia osuuksia mallin antamiin sovitteisiin.
(b) Sovita aineistoon myös probitmalli. Vertaa tulosta logitmallin antamaan tulokseen.

Taulukko 5.5 Äidin alkoholin käyttö ja lapsen epämuodostumat.

	Epämuodostuma		
Alkoholin kulutus	Ei	On	Yhteensä
0	17066	48	17114
$<$1	14464	38	14502
1-2	788	5	793
3-5	126	1	127
$>=$6	37	1	38

6.3. Vertaillaan virheellisten tuotteiden lukumääriä rinnakkaisissa valmistusprosesseissa A ja B. Kumpaakin prosessia tarkkaillaan 10 viikkoa. Viallisten tuotteiden lukumäärät tuote-erissä menetelmällä A olivat 8, 7, 6, 6, 3, 4, 7, 2, 3, 4 ja menetelmällä B vastaavasti 9, 9, 8, 14, 8, 13, 11, 5, 7, 6. Tarkastele viallisten lukumääriä toisistaan riippumattomina Poissonin satunnaismuuttujina, joiden odotusarvot ovat μ_A ja μ_B. Vertaile prosesseja A ja B sovittamalla aineistoon logaritmisesti lineaarinen Poissonin regressiomalli $log\ \mu = \alpha + \beta x$, missä $x = 1$ tarkoittaa menetelmää B ja $x = 0$ menetelmää A.

6.4. Taulukossa 5.6 on esitetty autovakuutuksen perusteella suoritettujen korvausten lukumääriä koskeva hypoteettinen aineisto. Autot on luokiteltu iän (2 tasoa) ja koon (3 tasoa) mukaan. Oleta, että korvausten lukumäärät noudattavat toisistaan riippumatta Poissonin jakaumaa. Selitä korvausten lukumääriä auton iällä ja koolla. Suhteuta korvausten lukumäärät autojen lukumääriin eli käytä autojen määrää `offset`-muuttujana.

Taulukko 5.6 Vakuutuskorvaukset auton iän ja koon mukaan.

Autoja	Korvauksia	Koko	Ikä
500	42	pieni	1
1200	37	keskikokoinen	1
100	1	suuri	1
400	101	pieni	2
500	73	keskikokoinen	2
300	14	suuri	2

6.5. Kun kuorsauksen määrä pisteytetään $\{0, 2, 4, 5\}$ on Taulukkon 5.1 aineistosta estimoitu probitmalli

$$\text{probit}[\hat{\pi}(x)] = -2.061 + 0.188x.$$

(a) Millä kuorsauksen tasolla sydäntaudin todennäköisyys olisi $\frac{1}{2}$?

(b) Tulkitse, mitä tarkoittaa $\text{probit}[\hat{\pi}(0)] = -2.061$.