Yleistetyt lineaariset mallit
5. harjoitukset, 8. vko 2002

5.1. Tarkastellaan luennolla esitettyä aineistoa `budworm`(Collet 1999, s. 75), jossa on tutkittu kokeellisesti cypermethrin-nimisen myrkyn vaikutusta tiettyjen tuhohyönteisten kuolleisuuteen (Huom! Aineisto ja mm. luennolla käsitellyn asian R-koodi ovat hakemistossa `http://mtl.uta.fi/tilasto/GLIM/R_luennot/binreg`). Selitä kuolleisuutta (kuoleman todennäköisyyttä) myrkkyannoksen logaritmilla (2-kantainen) logistista regressiota

$$\text{logit}(\frac{\pi_i}{1-\pi_i}) = \alpha + \beta x_i, \; i = 1, \ldots, 20 \tag{0.0.1}$$

käyttäen (logit-linkki).

(a) Estimoi parametrien suurimman uskottavuuden estimaatit

(b) Laske mallin uskottavuusfunktion ja logaritmoidun uskottavuusfunktion maksimit

(c) Laske täyden mallin uskottavuusfunktion ja logaritmoidun uskottavuusfunktion maksimit

5.2. Jatketaan edellisen tehtävän aineiston tarkastelua.

(a) Laske mallin devianssi edellisessä tehtävässä laskettujen logaritmodun uskottavuusfunktion maksimien avulla.

(b) Estimoi myös nollamalli ja testaa selittäjän merkitsevyys devianssien erotuksen avulla.

5.3. Tarkastellaan edelleen `budworm`-aineistoa. Olkoot y_i, $i = 1, \ldots, 12$ kuolleiden lukumäärät. Merkitään $y_{i1} = y_i$ ja $y_{i2} = 20 - y_i$. Osoita, että mallin 0.0.1 devianssi on muotoa

$$2 \sum_{i=1}^{12} \sum_{j=1}^{2} y_{ij} \log \frac{y_{ij}}{\hat{y}_{ij}}, \tag{0.0.2}$$

missä $\hat{y}_{ij} = 20\hat{\pi}_{ij}, \hat{\pi}_{i1} = \hat{\pi}_i$ ja $\hat{\pi}_{i2} = 1 - \hat{\pi}_i$. Laske devianssi kaavalla 0.0.2

5.4. Ota `budworm`-aineistossa myös sukupuoli selittäjäksi. Ilmeisesti koiraiden ja naaraiden kuolleisuuskäyrät poikkeavat toisistaan. Onko kyse vain tasoerosta vai ovatko käyrät erimuotoiset? Yritä löytää paras malli.

5.5. Olkoot $X \sim Po(\lambda)$ ja $Y \sim Po(\mu)$ toisistaan riippumattomia satunnaismuuttujia. Silloin $X + Y \sim Po(\lambda + \mu)$. Osoita, että X :n ehdollinen jakauma ehdolla $X + Y = N$ noudattaa binomijakaumaa $Bi(N, \pi)$, missä $\pi = \lambda/(\lambda + \mu)$.

5.6. Olkoot $X_i \sim Po(\lambda_i)$, $i = 1, \ldots, k$ toisistaan riippumattomia satunnaismuuttujia. Silloin $X_1 + X_2 + \ldots + X_k \sim Po(\lambda_1 + \lambda_2 + \ldots + \lambda_k)$. Osoita, että satunnaismuuttujien $X_1, X_2, \ldots, X_k$ ehdollinen yhteisjakauma ehdolla $X_1 + X_2 + \ldots + X_k = N$ noudattaa multinomijakaumaa parametrein N ja $\pi_1, \pi_2, \ldots, \pi_k$, missä $\pi_i = \lambda_i/(\lambda_1 + \lambda_2 + \ldots + \lambda_k)$, $i = 1, 2, \ldots, k$.

5.7. Tarkastellaan uusien kuukausittain (36 kk) rekisteröityjen AIDS-tapausten määriä Englannissa marraskuuhun 1985 asti (Altham s. 36, aineisto on annettu tiedostossa `http://mtl.uta.fi/tilasto/GLIM/R_luennot/poireg`). Oletetaan, että uusien tapausten lukumäärät $Y_i \sim Po(\mu_i)$, $\mu_i \geq 0, i = 1, \ldots, 36$ ovat riippumattomia ja noudattavat Poissonin jakaumaa.

(a) Estimoi malli, jossa odotusarvojen logaritmilla on lineaarinen trendi (selittäjänä kuukauden järjestysluku).

(b) Laske mallin antamat odotusarvojen estimaatit (sovitteet) ja piirrä odotusarvokäyrä pisteparveen.

(c) Sopiiko malli hyvin aineistoo?

(d) Yhdistä ainakin joitain kuukausia pidemmiksi aikaväleiksi, jotta saadaan suurempia sovitteita. Miten mallin sopivuus (devianssi) muuttuu?