Yleistetyt lineaariset mallit
6. harjoitukset, 9. vko 2002

6.1. Tarkastellaan `budworm`-aineistoa (vrt. 5. harjoitukset). Selitä hyönteisten kuolleisuutta sukupuolella ja käytetyllä myrkkyannoksella (`ldose`) ja ota malliin myös annoksen ja sukupuolen yhdysvaikutus. Onko sukupuoli merkitsevä selittäjä mallissa? Ota sitten selittäjäksi muuttujan `ldose` sijaan `ldose-3`. Onko sukupuoli nyt merkitsevä? Tulkitse tulos.

6.2. Tarkastellaan aineistoa `Cambridge`, jossa on annettu tieliikenneonnettomuudet Cambridgessa 1978-81 kahdelta tieosuudelta `TRoad` ja `MRoad` kolmena eri vuorokauden aikana (ks. Altham s. 39). Aineisto on myös tiedostossa `http://mtl.uta.fi/tilasto/GLIM/R_luennot/Cambridge.txt`. Oletetaan, että onnettomuuksien lukumäärät ovat toisistaan riippumattomia satunnaismuuttujia $Y_{ij} \sim Po(\mu_{ij}),\ i = 1,2; j = 1,2,3$, missä i on tieindeksi ja j aikaindeksi.

(a) Selitä onnettomuuksien määrää ensin pelkästään tiellä ja vuorokauden ajalla. Oleta tien ja vuorokauden ajan vaikutukset toisistaan riippumattomiksi. Tulkitse malli. Yksinkertaista mallia, jos se on perusteltua. Laske mallin antamat sovitteet (estimoidut onnettomuuksien määrät) ja residuaalit.

(b) Ota malliin mukaan myös liikennemäärät ja tee samat tarkastelut kuin edellisessä kohdassa.

6.3. Tehtävässä 5.7 selitettiin uusien AIDS-tapausten määrää UK:ssa (`AIDS`-aineisto) logaritmisesti lineaarisella Poissonin mallilla

$$\log(\mu_i) = \alpha + \beta i,$$

missä i on kuukauden järjestysnumero.

(a) Esitä Apulausetta 2.2 (Altham s. 17-18, Lemma 2) soveltaen mallin kertoimien suurimman uskottavuuden estimaattorin asymptoottinen kovarianssimatriisi.

(b) Laske numeerisesti kovarianssimatriisin estimaatti.

6.4. Oletetaan, että satunnaismuuttujat $Y_1, Y_2, \ldots, Y_n$ ovat riippumattomia ja noudattavat samaa jakaumaa (otos jakaumasta). Kirjoita jakauman tiheysfunktio (tf) ja logaritmoitu uskottavuusfunktio (uf) eksponentiaalisen jakuman standardimuotoa käyttäen (ks. Altham, s. 15), kun

(a) $Y_i \sim N(\mu, \sigma^2)$

(b) $Y_i \sim Po(\mu)$

(c) $Y_i \sim Gam(\mu, \nu),\ i = 1, 2, \ldots, n$. Gammajakauman tiheysfunktio on muotoa

$$f(y; \mu, \nu) = \frac{1}{\Gamma(\nu)} (\frac{\nu}{\mu})^{\nu} y^{\nu - 1} \exp(-\frac{\nu}{\mu} y),\ y \geq 0.$$

(Vihje: Gammajakauman skaalausparametri $\varphi = \frac{1}{\nu}$)

6.5. Sama kuin tehtävä 4 , mutta nyt oletetaan, että $g(\mu_i) = \alpha + \beta x_i,\ i = 1, 2, \ldots, n$. Kunkin jakauman linkkifunktio on luonnollinen ja $\mathrm{E}(Y_i) = \mu_i$. Jakaumat ovat

(a) $Y_i \sim N(\mu_i, \sigma^2)$

(b) $Y_i \sim Po(\mu_i)$

(c) $Y_i \sim Gam(\mu_i, \nu),\ i = 1, 2, \ldots, n$.

6.6. (a) Määritä tehtävässä 4 annettujen jakaumien odotusarvot ja varianssit sekä

(b) tehtävässä 5 annettujen jakaumien odotusarvot ja varianssit Apulauseen 2.1 avulla (Altham, s. 15, Lemma 1).

6.7. The Daily Telegraph (28.10.1988) julkaisi artikkelin "Johtajat alttiita ajamaan humalassa" (Altham, s. 42). Artikkelissa esitettiin Royal Ascotin ja Henley Regatan yhteydessä järjestettyjen puhallustestien tuloksia (Royal Ascot ja Henley Regatta ovat tunnetusti "kosteita" englantilaisia urheilutapahtumia niinkuin hevos- ja soutukilpailutkin).

(a) Estimoi täysi malli ja riippumattomuusmalli (kiinnijääminen ja paikka riippumattomia).

(b) Vertaile mallien antamia tuloksia ja tulkitse.

(c) Jääkö RA:ssa todennäköisemmin kiinni kuin HR:ssä (jos puhallutetaan)?

RA:ssa puhallutettiin kaikkiaan 24+2210 ja HR:ssä 5+680 henkilä. RA:ssa jäi siis kiinni 24 maistanutta ja HR:ssa 5(ks. Altham s. 42). Katso http://mtl.uta.fi/tilasto/GLIM/R_luennot/Regatta.txt