

Yleistetyt lineaariset mallit

8. harjoitukset, 11. vko 2002

- 8.1. Tarkastellaan aineistoa `kyphosis`(ks. . . . `/glim/datat/kyphosis.txt`). Selitetään kyttyräselkäisyyden (`Kyphosis`) todennäköisyyttä selkäleikkauksen jälkeen, kun selittäjinä ovat muuttujat `Age`, `Number` ja `Start` sekä niiden kaikki kahdenväliset yhdysvaikutukset. Mikä on paras malli? (Muuttujien kuvaus tiedoston . . . `/glim/R_luennot/Glim_in_R` alussa. Tarkastele tilannetta AIC:n, devianssin, mallin kertoimien merkitsevyyden ja uskottavuussuhdetestin (khin neliö) kannalta.
- 8.2. Vertaile parasta malliasi malliin `kyph.glm4 <- glm(Kyphosis ~ poly(Age,2) + I((Start>12)*(Start-12))`. Tarkastele vaikuttavia havaintoja Cookin etäisyyden ja `dfbetas`-mitan avulla. Poista vaikuttavia havaintoja yksitellen ja tee aina uusi analyysi. Miten tulokset muuttuvat? Mikä on poiston jälkeen vaikuttavin?
- 8.3. Tarkastellaan englantilaisten lääkäreiden tupakointia ja sepelvaltimotautiin liittyvää kuolleisuutta koskevaa tutkimusta, joka aloitettiin vuonna 1951 (Dobson 2001, s. 154, kopio jaettu luennolla ja on liitteenä tehtävien paperiversiossa). (aineisto tiedostossa . . . `/glim/datat/smoking.cdeath.txt`). Selitä kuolleisuutta iällä ja tupakoinnilla.
- (a) Estimoi malli selittämällä "raakaa" kuolleisuuslukua suoraan iällä ja tupakoinnilla.
 - (b) Suhteuta kuolleisuus kyseiseen riskiryhmään kuuluvien lääkäreiden määrällä.
- Vertaile malleja ja niiden antamia ennusteita.
- 8.4. Ota edellisen tehtävän mallissa selittäjiksi myös iän neliö ja tupakointikä (`smokeage`). Vertaile tätä mallia edellisessä tehtävässä saatuun. Tulkitse kertoimet, laske mallin antamat ennusteet ja residuaalit. Etsi vaikuttavin havainto. Tulkitse.
- 8.5. (Ks. tehtävä 5.7) Tarkastellaan uusien kuukausittain (36 kk) rekisteröityjen AIDS-tapausten määriä Englannissa marraskuuhun 1985 asti (Altham s. 36, aineisto on annettu tiedostossa http://mt1.uta.fi/tilasto/GLIM/R_luennot/poireg). Oletetaan, että uusien tapausten lukumäärät $Y_i \sim Po(\mu_i)$, $\mu_i \geq 0, i = 1, \dots, 36$ ovat riippumattomia ja noudattavat Poissonin jakaumaa. Estimoi malli,
- (a) kun selittäjänä on kuukauden järjestysluku ja käytetään neliöjuurilinkkiä.

(b) Vertaa tulosta logaritmilinkillä saatuun tulokseen. Piirrä estimoitujen mallien kuvaajat.

(c) Kumpi malli on parempi?

8.6. Tarkastellaan iteratiivista painotettua pienimmän neliösumman menetelmään

$$\mathbf{X}^T \mathbf{W} \mathbf{X} \beta_{q+1} = \mathbf{X}^T \mathbf{W} \mathbf{z} \quad (0.0.1)$$

yleistetyn lineaarisen mallin parametrivektorin β estimoimiseksi. Esitä edellisen tehtävän logaritmisessä Poissonin mallin tilanteessa

(a) "working" residuaalivektorin r^W ja

(b) \mathbf{z} :n lauseke.

8.7. Laske R:n avulla tehtävän 5 Poissonin logaritmisesta mallista (viimeisen iteraatiokierroksen) painomatriisi \mathbf{W} (painot w_i) sekä \mathbf{z} -vektori. Totea, että näillä arvoilla yhtälö 0.0.1 toteutuu.