Yleistetyt lineaariset mallit
9. harjoitukset, 12. vko 2002

9.1. Olkoot $Y_1, \ldots, Y_N$ toisistaan riippumattomia Poissonin jakaumaa $Y_i \sim Po(\mu_i)$ noudattavia satunnaismuuttujia. Tarkastellaan mallia

$$g(\mu_i) = \mathbf{x}_i^T \beta = \beta_1 + \sum_{j=2}^{k} x_{ij}\beta_j,\ i = 1, \ldots, N,$$

missä g on linkkifunktio ja $k < N$. Merkitään mallista saatuja odotusarvojen suurimman uskottavuuden estimaatteja $\hat{\mu}_i$, $i = 1, \ldots, n$. Osoita, että mallin devianssi on

$$D = 2[\sum_{i=1}^{N} y_i \log(\frac{y_i}{\hat{\mu}_i}) - \sum_{i=1}^{N}(y_i - \hat{\mu}_i)].$$

9.2. Olkoot $Y_1, \ldots, Y_N$ toisistaan riippumattomia Poissonin satunnaismuuttujia $Y_i \sim Po(\mu_i)$ ja

$$\log(\mu_i) = \beta_1 + \sum_{j=2}^{k} x_{ij}\beta_j,\ i = 1, \ldots, N.$$

(a) Osoita, että $\frac{\partial l(\beta)}{\partial \beta_1} = \sum_{i=1}^{N}(y_i - \mu_i)$, missä $l(\beta)$ on logaritmoitu uskottavuusfunktio.

(b) Osoita, että odotusarvojen suurimman uskottavuuden estimaatit toteuttavat yhtälön $\sum_{i=1}^{N} y_i = \sum_{i=1}^{N} \hat{\mu}_i$.

(c) Miten tehtävässä 1 esitetty devianssin lauseke sievenee edellisen kohdan tuloksen perusteella?

9.3. Tarkastellaan influenssarokotteen tehoa testanneen kokeen tuloksia (Dobson 2001, s. 159; Aineisto tiedostossa `... /datat/Influenza`).

(a) Testaa tavanomaisella χ^2-yhteensopivuustestillä ja sopivalla logaritmisesti lineaarisella mallilla nollahypoteesi, että koeryhmässä ja lumeryhmässä on sama vasta-ainejakauma.

(b) Laske nollahypoteesiin liittyvän mallin antamat sovitteen, devianssiresiduaalit ja Pearsonin residuaalit. Mitkä havainnot kasvattavat devianssia eniten?

9.4. 2×2-kontingenssitaulukon täysi malli voidaan kirjoittaa muodossa

$$\eta_{ij} = \lambda + (-1)^{i-1}\lambda^X + (-1)^{j-1}\lambda^Y + (-1)^{i+j}\lambda^{XY}, i = 1, 2;\ j = 1, 2,$$

missä $\eta_{ij} = \log(\mathrm{E}(Y_{ij}) = \log(n\pi_{ij})$, $n = \sum\sum Y_{ij}$ ja $\sum\sum \pi_{ij} = 1$. Osoita, että yhdysvaikutustermi

$$\lambda^{XY} = \frac{1}{4}\log\phi,$$

missä $\phi = (\pi_{11}\pi_{22})/(\pi_{12}\pi_{21})$. Riippumattomuusmalli liittyy siis tilanteseen $\phi = 1$.

9.5. Influenssarokotteen tehoa testanneessa kokeessa (ks. tehtävä 3) lumeryhmään oli valittu 38 henkilöä ja koeryhmään 35 henkilöä. Henkilöiden lukumäärä kummassakin ryhmässä jakautuu vasta-aineluokkiin (pieni/kohtalaine/suuri) multinomijakauman mukaan (kaksi riippumatonta multinomista otosta). Esitä, miten testataan multinomijakauman avulla hypoteesi, että henkilöt jakautuvat vasta-aineluokkiin samalla tavalla koe- ja lumeryhmässä?

9.6. Tarkastellaan 2×2-taulukon täyttä logaritmisesti lineaarista mallia

$$\log(\mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Z + \lambda_{ij}^{XZ}.$$

Olkoot havainnot $y_{11}, y_{12}, y_{21}, y_{22}$. Mitä on parametrin λ_{22}^{XZ} suurimman uskottavuuden estimaatti, kun identifiointirajoitteet ovat $\lambda_1^X = \lambda_1^Z = \lambda_{11}^{XZ} = \lambda_{12}^{XZ} = \lambda_{21}^{XZ} = 0$?

9.7. Esittele harjoitustyösi aihe (2p).