Logistinen regressio
Harjoitukset 15.3.2001

6.1. Tutkitaan `wisc` aineistossa lukiolaisten aikomuksia opiskella yliopistossa. Tutki hypoteesia, että vanhempien kannustus saattaa olla ns. väliintuleva muuttuja. Hypoteesin mukaan paremmassa sosiaalisessa ja taloudellisessa asemassa olevat vanhemmat pystyvät kannustamaan paremmin lapsiaan opiskeluun kuin huonommassa asemassa olevat. (Vihje: Vertaile muuttujan `parent` sisältämää mallia malliin, jossa `parent` ei ole mukana. Pidä `iq` kvantitatiivisena ja `ses` luokitteluasteikollisena selittäjänä.

6.2. Oletetaan, että toisistaan riippumattomat satunnaismuuttujat $Y_i,\ i = 1,2$ noudattavat binomijakaumaa $Y_i \sim b(n_i, \pi_i)$. Oletetaan, että erään logitmallin avulla saatiin parametrien π_1 ja π_2 suurimman uskottavuuden estimaateiksi $\hat{\pi}_1$ ja $\hat{\pi}_2$. Merkitään "onnistumisten" lukumääriä $y_{i1},\ i = 1,2$ ja vastaavasti $y_{i2} = n_i - y_{i1},\ i = 1,2$. Esimerkiksi π_1 voisi olla keuhkosyövän todennäköisyys tupakoivien joukossa ja π_2 vastaava todennäköisyys tupakoimattomien joukossa. Merkitään teoreettisia frekvenssejä μ_{ij}. Teoreettisten frekvenssien suurimman uskottavuuden estimaatit ovat $\hat{\mu}_{i1} = n_i\hat{\pi}_i$ ja $\hat{\mu}_{i2} = n_i(1 - \hat{\pi}_i),\ i = 1,2$. Totea laskemalla, että devianssi $2L_f - 2L_M$ on

$$2\sum_{i,j=1}^{2} y_{ij} \log(\frac{y_{ij}}{\hat{\mu}_{ij}}).$$

6.3. Selitetään opiskelijoiden valintoja (muuttuja `wallet`) muuttujilla `male, business, punish` ja `explain` (Aineisto `wallet`). Tutki, löytyykö jokin yhdysväikutuksia sisältävä malli, joka on tilastollisesti parempi kuin päävaikutusten malli. CATMODin MODEL lauseessa yhdysvaikutustermit tehdään samalla tavalla kuin GENMODissa.

6.4. Selitetään edelleen opiskelijoiden valintoja `wallet` aineistossa. Kolmeluokkaisen vastemuuttujan multinomisessa logitmallissa saadaan kaksi riippumatonta logityhtälöä. Tutki hypoteesia, että näiden yhtälöiden kertoimet ovat samat. Tulkitse tulokset.

6.5. Selitetään uskoa kuolemanjälkeiseen sukupuolella ja rodulla (`afterlif`). Estimoi multinominen logitmalli. Lausu kaikki kolme logityhtälöä saamiasi estimaatteja käyttäen. Mikä on todennäköisyys, että

(a) valkoinen nainen uskoo
(b) valkoinen mies uskoo
(c) musta nainen ei usko
(d) valkoinen mies on epävarma?