

JYVÄSKYLÄN YLIOPISTO  
Matematiikan ja tilastotieteen laitos  
Johdatus tilastolliseen mallintamiseen

Harjoitustyö 2  
Viimeinen palautuspäivä 5.12.2011

**Kirjoita siisti raportti jollakin tekstinkäsittelyllä. Esitä raporttissasi vain mielestäsi tarpeelliset taulukot ja kuvat, joista helposti käy ilmi mitä ne sisältävät. Viiden sivun raportti on riittävä. Liitä käyttämäsi koodi raporttisi loppuun. Koodi mukaan luettuna sivumäärä voi olla enemmän kuin viisi.**

Jokainen kirjoittaa raporttinsa itsenäisesti.

**Palautus paperisena työhuoneeni (MaD321) vieressä olevaan lokerikkoon. Nido paperit yhteen!**

1. Aineisto `plans` sisältää tietoja erään Wisconsinin lukion (high school) pojista:

- `ses` = 1,2,3,4; sosioekonominen asema: 1 = alin taso, 2 = alempi keskitaso, 3 = ylempi keskitaso ja 4 = ylin taso;
- `iq` = 1,2,3,4; älykkyys: 1 = alin taso, 2 = alempi keskitaso, 3 = ylempi keskitaso ja 4 = ylin taso;
- `encourage` = 1,2; vanhempien rohkaisu jatko-opintoihin: 1 = vähäinen, 2 = runsas;
- muuttajat `yes` ja `no` ilmoittavat jatko-opintosuunnitelmista collegessa, `yes` tarkoittaa, että jatkosuunnitelma on, ja `no` tarkoittaa, ettei jatkosuunnitelmaa ole. Ota `yes` vasteksi.
- `freq` on summa `yes` + `no` eli se ilmoittaa kussakin ryhmässä olevien poikien lukumäärän.

Halutaan tutkia, mitkä taustamuuttujat ovat yhteydessä jatko-opintosuunnitelmiin. Rakenna sellainen logistinen regressiomalli, jonka avulla voit vastata seuraaviin kysymyksiin:

- a) Mikä on todennäköisyys, että pojalla on jatko-opintosuunnitelma, kun hänen älykyytensä ja sosioekonominen tasonsa ovat alimmalla tasolla eikä hän saa vanhemmiltaan rohkaisua?
- b) Mikä on tämä todennäköisyys siinä tapauksessa, että vanhemmat rohkaisevat häntä?
- c) Laske kohtien a) ja b) todennäköisyyksien osamäärät ja vertaa niitä `encourage`-muuttujaan liittyvään ristitulosuhteeseen (OR:ään).
- d) Tee malliisi liittyvät diagnostiset tarkastelut graafisesti. Esitä kommenttisi.

2. Aineisto `allshoots`; ks. esimerkki 5.1 ja harj. 8 teht. 1. Sovelletaan aineistoon Poisson-regressiota. Kertolaskulla `n*ybar` (pyöristä lähimpään kokonaislukuun `round()`-funktioilla) saat otokseen tulleissa versoissa olevan runkoyksiköiden kokonaismäärän `y`. Oleta, että

$y_i \sim \text{Po}(n_i\theta_i)$ . Rakenna Poisson-regressiomalli, jossa selittäjinä ovat Day ja Type. Tutki tarvitaanko mallissa interaktiota ja/tai Day:n polynomia. Perusta ehdotuksesi sekä merkitsevyyystesteihin että jäännösten diagnostisiin tarkasteluihin. Tulkitse mallisi kertoimet.