

78185 Yleistetyt lineaariset mallit

Harjoitus 1, syksy 2014

1. Bernoulli-jakauman pistetodennäköisyysfunktio on

$$f(y; p) = p^y(1 - p)^{1-y}, \quad y \in \{0, 1\}, \quad p \in (0, 1).$$

- a) Osoita, että Bernoulli-jakauma kuuluu eksponentiaaliseen jakaumaperheeseen

$$f(y; \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\},$$

- b) Laske jakauman odotusarvo ja varianssi funktioiden $a(\phi)$ ja $b(\theta)$ avulla.

2. Eksponenttijakauman tiheysfunktio on

$$f(y; \lambda) = \lambda e^{-\lambda y}, \quad y > 0, \quad \lambda > 0.$$

- a) Osoita, että eksponenttijakauma kuuluu eksponentiaaliseen jakaumaperheeseen.

- b) Laske jakauman odotusarvo ja varianssi funktioiden $a(\phi)$ ja $b(\theta)$ avulla.

3. Gammajakauman tiheysfunktio on muotoa

$$f(y; \mu, \nu) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\mu}\right)^\nu y^{\nu-1} \exp\left(-\frac{\nu}{\mu}y\right),$$

jossa $y > 0$, $\mu > 0$, $\nu > 0$.

- a) Osoita, että gammajakauma kuuluu eksponentiaaliseen jakaumaperheeseen.

- b) Laske jakauman odotusarvo ja varianssi funktioiden $a(\phi)$ ja $b(\theta)$ avulla.

4. Aineistossa `lapset85.dat` on tietoja Oulun läänissä 1985-86 syntyneistä lapsista ja heidän vanhemmistaan. Muuttujankuvaukset ovat tiedostossa `muuttujat.pdf`.

- a) Lue aineisto R-ohjelmistoon (voit käyttää myös muita tilasto-ohjelmistoja, esim. SPSS, SAS,...) ja kuvaile muuttujia sopivien tunnuslukujen ja graafisten esitysten avulla.

- b) Halutaan tutkia äidin raskaudenaikaisen tupakoinnin vaikutusta lapsen syntymäpainoon, kun mahdolliset analyysiä sekoittavat tekijät aineistossa ovat äidin ikä, äidin sosiaaliluokka, pariteetti, lapsen gestaatioikä ja sukupuoli. Sovita aineistoon erilaisia lineaarisia regressiomalleja ja mieti näiden mallien regressiokertoimien tulkintaa.

Ohjeita R-ohjelmiston käyttöön

Luetaan havaintoaineisto ASCII-tiedostosta `tiedosto.dat`. Tiedosto on levyn `e` hakemistossa `aineistot` ja tiedoston ensimmäisellä rivillä on muuttujien nimet.

```
> tiedosto <-read.table("e:/aineistot/tiedosto.dat",header=TRUE)

> summary(tiedosto)    #yksinkertaisia tunnuslukuja muuttujista
> names(tiedosto)     #muuttujien nimet
> dim(tiedosto)       #tiedoston dimensiot (rivit, sarakkeet)
```

Lineaarisisessa regressiomallissa selitettävänä muuttujana on y ja selittävinä muuttujina x_2 , x_3 ja x_4 . Vakiokerroin on oletusarvoisesti mukana mallissa.

```
> mod1<-lm(y~x2+x3+x4,data=tiedosto)
```

Lineaarisisessa regressiomallissa selitettävänä muuttujana on y ja selittävinä muuttujina x_2 , x_3^2 ja x_4 . Vakiokerroin ei ole mukana mallissa (-1).

```
> mod2<-lm(y~-1+x2+I(x3^2)+x4,data=tiedosto)

> summary(mod2)       #lineaarisen regressioanalyysin tulokset
> coef(mod2)         #regressiokertoimet

> v<-c(2,1,3,1,2,3,3)
> u<-factor(v)       #muodostetaan faktorimuuttuja
> u
[1] 2 1 3 1 2 3 3
Levels: 1 2 3
```

Huom! Kun k -luokkainen faktorimuuttuja lisätään malliin, niin funktio `lm` muodostaa automaattisesti $k - 1$ kpl 0/1-muuttujia.