**Lineaaristen mallien kurssi - yleistentti 24.1.2013**

**1.** Määrittele yleinen (täysiasteinen) lineaarinen malli kaikkine oletuksineen ja siihen liittyvät käsitteet (pienimmän neliösumman) sovite ja residuaali. Selvitä muutamalla virkkeellä mallin sekä sovitteen ja residuaalin tulkinta. Osoita lisäksi oikeaksi yhtälöt
(a) $\hat{\mathbf{y}}'\hat{\boldsymbol{\varepsilon}} = 0$ ja
(b) $(\hat{\varepsilon}_1 + \cdots + \hat{\varepsilon}_n)/n = 0$, kun mallissa on vakiotermi.
Edellä $\hat{\mathbf{y}}$ $(n \times 1)$ on aineistosta laskettu sovitevektori, $\hat{\boldsymbol{\varepsilon}} = [\hat{\varepsilon}_1 \ \cdots \ \hat{\varepsilon}_n]'$ on aineistosta laskettu residuaalivektori ja $n$ on havaintojen lukumäärä.

**2.** Olkoon $Y_1, ..., Y_6$ riippumattomia normaalisti jakautuneita satunnaismuuttujia, joille pätee $Y_i \sim \mathsf{N}(\mu_i, \sigma^2)$ ja

$$\mu_i = \begin{cases} \beta_1 - \beta_2, & \text{kun } i = 1,2,3 \\ \beta_1 + \beta_2, & \text{kun } i = 4,5,6. \end{cases}$$

Muotoile tilanne lineaarisena mallina ja estimoi parametrit $\beta_1$ ja $\beta_2$ soveltamalla lineaarisen mallin estimointiteoriaa. Kuvaa lyhyesti käyttämäsi estimointiperiaate ja selvitä parametrivektorin $\boldsymbol{\beta} = [\beta_1 \ \ \beta_2]'$ estimaattorin $\hat{\boldsymbol{\beta}} = [\hat{\beta}_1 \ \ \hat{\beta}_2]'$ jakauma. Ovatko estimaattorit $\hat{\beta}_1$ ja $\hat{\beta}_2$ riippumattomia? (Esitä myös perustelu estimaattorin $\hat{\boldsymbol{\beta}}$ jakaumalle ja sen komponenttien riippumattomuudelle tai riippuvuudelle.)

**3.** Oletetaan, että $n$ riippumatonta havaintoa $Y_1, ..., Y_n$ $(n > 2)$ on poimittu normaalijakaumasta, jonka varianssi on tuntematon $\sigma^2$. Ensimmäisen $n-1$ havainnon tiedetään olevan peräisin samasta normaalijakaumasta, mutta ennen viimeisen havainnon poimimista epäillään, että havaintojen odotusarvo on muuttunut äkillisesti ja saattaa poiketa ensimmäisen $n-1$ havainnon odotusarvosta. Muotoile tilanne lineaarisen mallin avulla ja johda testi hypoteesille, jonka mukaan $n$. havainto on peräisin samasta normaalijakaumasta kuin $n-1$ ensimmäistä havaintoa.

**4.** Yksisuuntainen varianssianalyysi: Ongelmanasettelu, tilastollinen malli, tavallisimmin testattava hypoteesi ja sen testaaminen.

## Muistin tueksi

- Satunnaisvektorin $\mathbf{X} \sim \mathsf{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ tiheysfunktio on

$$f(x) = (2\pi)^{-k/2} \det(\boldsymbol{\Sigma})^{-1/2} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\},$$

  jossa $\det(\boldsymbol{\Sigma})$ on kovarianssimatriisin $\boldsymbol{\Sigma}$ determinantti ja yksiulotteisessa tapauksessa $k = 1$.

- $F_{k,m}$-jakauman määrittelee satunnaismuuttuja $m\chi^2_k / k\chi^2_m$, jossa $\chi^2_k \perp\!\!\!\perp \chi^2_m$.
  Lisäksi $\mathsf{E}(\chi^2_k) = k$, $\mathsf{Var}(\chi^2_k) = 2k$.

- $t_k$-jakauman määrittelee satunnaismuuttuja $Z/\sqrt{\frac{1}{k}\chi^2_k}$, josssa $Z \sim \mathsf{N}(0,1)$ ja $Z \perp\!\!\!\perp \chi^2_k$.

- Jos $\mathbf{X} \sim \mathsf{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, niin $(\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi^2_k$.

- Jos $\mathbf{X} \sim \mathsf{N}_k(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_k)$ ja matriisi $\mathbf{P}$ $(k \times k)$ astetta $r$ oleva ortogonaalinen projektio, niin $(\mathbf{X} - \boldsymbol{\mu})' \mathbf{P} (\mathbf{X} - \boldsymbol{\mu}) / \sigma^2 \sim \chi^2_r$.