Lineaariset mallit, kl 2016, Harjoitus 6, viikko 18

Huom.: Helatorstain takia laskuharjoitukset pidetään vain perjantaina 10-12. Jos et pääse laskuharjoituksiin, voit toimittaa ratkaisut sähköpostitse laskuharjoitusten pitäjälle (petri.t.aaltonen@helsinki.fi).

1. Tarkastellaan mallia

$$Y_i = \beta_1 + \beta_2 x_{i1} + \beta_3 x_{i2} + \varepsilon_i, \quad i = 1, ..., n,$$

jossa $\varepsilon_1, ..., \varepsilon_n \sim \mathsf{N}\left(0, \sigma^2\right) \perp\!\!\!\perp$. Merkitsemällä $\bar{x}_j = n^{-1} \sum_{i=1}^{n} x_{ij}$, $j = 1, 2$, ja $\alpha = \beta_1 + \beta_2 \bar{x}_1 + \beta_3 \bar{x}_2$ voidaan malliyhtälö kirjoittaa muodossa

$$Y_i = \alpha + \beta_2(x_{i1} - \bar{x}_1) + \beta_3(x_{i2} - \bar{x}_2) + \varepsilon_i, \quad i = 1, ..., n.$$

(i) Osoita, että parametrin β_2 PNS-estimaattorin varianssi on

$$\mathsf{Var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1)^2 \left(1 - r_{12}^2\right)},$$

jossa r_{12} on selittäjien x_{i1} ja x_{i2}, $i = 1, ..., n$, välinen otoskorrelaatiokerroin (ks. HT 5.3 tai monisteen alaviite 10, s. 11).

(ii) Muodosta $100(1-\alpha)$%:n luottamusväli parametrille β_2. Mitä tälle luottamusvälille tapahtuu, kun selittäjien x_1 ja x_2 välinen korrelaatio kasvaa?

2. Jatkoa HT:lle 4.3, jossa tarkastellaan kahta riippumatonta lineaarista mallia

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\varepsilon}_i \sim \mathsf{N}_{n_i}\left(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i}\right), \quad \boldsymbol{\varepsilon}_1 \perp\!\!\!\perp \boldsymbol{\varepsilon}_2, \quad \boldsymbol{\beta}_i \in \mathbb{R}^p, \ \sigma^2 > \mathbf{0}, \ \mathsf{r}\left(\mathbf{X}_i\right) = p, \ i = 1, 2.$$

(i) Muodosta $100(1-\alpha)$%:n luottamusväli parametrivektorin $\boldsymbol{\beta}_1$ lineaarikombinaatiolle $\mathbf{a}'\boldsymbol{\beta}_1$ ($\mathbf{a} = [a_1 \ \cdots \ a_p]' \neq 0$) käyttäen HT:n 4.3 ratkaisussa esitettyä yhdistettyä mallia.

(ii) Luottamusväli lineaarikombinaatiolle $\mathbf{a}'\boldsymbol{\beta}_1$ voidaan muodostaa myös käyttäen ensimmäistä mallia $\mathbf{Y}_1 = \mathbf{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1$, $\boldsymbol{\varepsilon}_1 \sim \mathsf{N}_{n_i}\left(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i}\right)$, $\mathsf{r}\left(\mathbf{X}_i\right) = p$. Miten tämä luottamusväli eroaa edellisessä kohdassa saadusta luottamusvälistä vai eroaako lainkaan?

3. Jatkoa HT:lle 4.4, jossa tarkastellaan edellisen tehtävän mallia ehdolla $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. Muodosta $100(1-\alpha)$%:n luottamusväli parametrivektorin $\boldsymbol{\beta}_1$ lineaarikombinaatiolle $\mathbf{a}'\boldsymbol{\beta}_1$ ($\mathbf{a} = [a_1 \ \cdots \ a_p]' \neq 0$) käyttäen molempia malleja ja olettaen, että rajoite $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ on voimassa.

4. Oletetaan, että tavanomaisessa lineaarisessa mallissa $\mathbf{Y} \sim \mathsf{N}\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n\right)$ ($\boldsymbol{\beta} \in \mathbb{R}^p$, $\sigma^2 > \mathbf{0}$, $\mathsf{r}\left(\mathbf{X}\right) = p$) selittäjät ovat ortogonaalisia eli $\mathbf{X}'\mathbf{X}$ on diagonaalimatriisi. Johda parametrin β_j PNS-estimaatti, t-testisuure hypoteesille $H : \beta_j = 0$ ja $100(1-\alpha)$%:n luottamusväli parametrille β_j ($= \boldsymbol{\beta}$:n j. komponentti). Miten nämä muuttuvat, jos mallista poistetaan joku selittäjistä x_k ($k \neq j$)?

5. Oletetaan, että edellisessä tehtävässä $p = 3$, jolloin $\mathbf{X}$ on $n \times 3$ matriisi ja $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)$.

(i) Johda testi hypoteesille $H : \beta_1 = \beta_2 = \beta_3$.

(ii) Oletetaan, että hypoteesi H on tullut hylätyksi. Muodosta $100(1 - \alpha)\%$:n luottamusväli erotukselle $\beta_1 - \beta_2$. Miten tulkitset tilannetta, jossa luottamusväli on $(-0.8, 1.2)$?