Lineaariset mallit, kl 2016, Harjoitus 4, viikko 16

1. Tarkastellaan (täysiasteista) lineaarista mallia

$$\begin{aligned} &Y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, ..., n, \\ &\varepsilon_1, ..., \varepsilon_n \ \perp\!\!\!\perp\ , \ \varepsilon_i \sim \mathsf{N}\left(0, w_i \sigma^2\right), \end{aligned} \tag{1}$$

jossa $w_1, ..., w_n$ ovat tunnettuja positiivisia reaalilukuja. Määrittelemällä diagonaalimatriisi $\mathbf{W} = \mathsf{diag}\,[w_1 \ \cdots \ w_n]$ (ks. monisteen s. 47) nähdään, että kysymyksessä on tehtävän 3.5 mallin erikoistapaus, jossa kovarianssimatriisi $\mathbf{\Sigma} = \sigma^2\mathbf{W}$.

(i) Muunna malliyhtälöä (1) niin, että päädyt tavanomaiset oletukset toteuttavaan lineaariseen malliin, jonka parametrit ovat $\boldsymbol{\beta} = \left(\beta_1, ..., \beta_p\right)$ ja σ^2.

(ii) Esitä kohdassa (i) saamasi muunnettu malli matriisimuodossa ja päättele kurssilla esitettyjä tuloksia soveltaen mitkä ovat parametrien $\boldsymbol{\beta}$ ja σ^2 SU-estimaatit.

(iii) Käyttäen kohdassa (i) saamaasi muunnettua mallia ja Lausetta 2.1 selvitä parametrien $\boldsymbol{\beta}$ ja σ^2 SU-estimaattorien todennäköisyysjakaumat.

Huom.: Yksi (joskaan ei ainoa) tapa esittää parametrien $\boldsymbol{\beta}$ ja σ^2 SU-estimaatit on

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{-1}\mathbf{y} \ \text{ ja } \ \hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{W}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}),$$

jossa $\mathbf{y} = [y_1 \ \cdots \ y_n]'$ ja $\mathbf{X} = [x_{ij}]$ $(i = 1, ..., n,\ j = 1, ..., p)$. Näin saatu $\boldsymbol{\beta}$:n estimaatti tunnetaan painotettuna PNS-estimaattina. Luvut w_i tunnetaan esimerkiksi, kun selitettävän muuttujan havainnot ovat keskiarvotietoja eli y_i on keskiarvo n_i:stä riippumattomasta havainnosta, joiden vaihtelu on saman kokoista. Tällöin $w_i = 1/n_i$ (vrt. otoskeskiarvon varianssi riippumattomien samoin jakautuneiden havaintojen tapauksessa). Joissakin tapauksissa virhevarianssin voidaan olettaa olevan verrannollinen (positiivisarvoisen) selittäjän x_j kanssa, jolloin $w_i = x_{ij}$ (tämän oletuksen realistisuutta voidaan tutkia sen jälkeen, kun mallin parametrit on estimoitu).

2. Tarkastellaan kahden riippumattoman normaalisen otoksen mallia

$$Y_1, ..., Y_n \ \perp\!\!\!\perp\ , \ Y_i \sim \begin{cases} \mathsf{N}\left(\mu_1, \sigma^2\right), & \text{kun } i = 1, ..., n_1 \\ \mathsf{N}\left(\mu_2, \sigma^2\right), & \text{kun } i = n_1 + 1, ..., n_1 + n_2 = n \end{cases}$$

(μ_1, $\mu_2 \in \mathbb{R}$, $\sigma^2 > 0$, n_1, $n_2 > 1$).

(i) Harjoitustehtävän 3.2 ratkaisusta saadaan parametrin $\boldsymbol{\mu} = [\mu_1 \ \ \mu_2]'$ PNS-estimaattori $\hat{\boldsymbol{\mu}}$ ja sen todennäköisyysjakauma. Mikä on parametrin σ^2 harhaton estimaattori ja sen jakauma?

(ii) Estimoi parametrit μ_1 ja μ_2 ehdolla $\mu_1 = \mu_2$ ja selvitä μ_1:n (ja μ_2:n) rajoitetun PNS-estimaattorin jakauma. Mikä on tässä tapauksessa parametrin σ^2 harhaton estimaattori ja sen jakauma? (Monisteen s. 16-17 esitetyistä estimointitavoista kannattaa etsiä yksinkertaisinta.)

3. Tarkastellaan kahta riippumatonta lineaarista mallia

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

Muodosta näistä matriiseja käyttäen yksi malli ja esitä parametrin $\boldsymbol{\beta} = \left[\boldsymbol{\beta}_1' \ \ \boldsymbol{\beta}_2'\right]'$ PNS-estimaattorin $\hat{\boldsymbol{\beta}} = [\hat{\boldsymbol{\beta}}_1' \ \ \hat{\boldsymbol{\beta}}_2']'$ jakauma. Ovatko PNS-estimaattorit $\hat{\boldsymbol{\beta}}_1$ ja $\hat{\boldsymbol{\beta}}_2$ riippumattomia?

Aputulos: Jos $\mathbf{A}$ ja $\mathbf{B}$ ovat epäsingulaarisia neliömatriiseja, niin

$$\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^{-1} \end{bmatrix}.$$

4. Jatkoa edelliselle. Estimoi parametri $\boldsymbol{\beta} = \left[\boldsymbol{\beta}_1' \ \ \boldsymbol{\beta}_2'\right]'$ ehdolla $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. Mikä on $\boldsymbol{\beta}_1$:n (ja $\boldsymbol{\beta}_2$:n) rajoitetun PNS-estimaattorin jakauma? (Tässä on käytännön kannalta oletuksena, että rajoite $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ on muuttujien luonne huomioon ottaen mielekäs, mikä on mahdollisesti todettu myös formaalissa testauksessa. Tässäkin tapauksessa monisteen s. 16-17 esitetyistä estimointitavoista kannattaa etsiä yksinkertaisinta.)

5. Tarkastellaan tehtävän 3 mallin erikoistapauksena kahta riippumatonta yhden selittäjän lineaarista regressiomallia

$$Y_1, ..., Y_n \ \perp\!\!\!\perp, \ Y_i \sim \begin{cases} \mathsf{N}\left(\beta_1 + \beta_2 x_i, \sigma^2\right), & \text{kun } i = 1, ..., n_1 \\ \mathsf{N}\left(\beta_3 + \beta_4 x_i, \sigma^2\right), & \text{kun } i = n_1 + 1, ..., n_1 + n_2 = n. \end{cases}$$

(i) Muotoile tilanne lineaarisena mallina käyttäen matriiseja.

(ii) Asetetaan parametreja β_2 ja β_4 sitova rajoite $\beta_2 = \beta_4$. Ilmaise tämä rajoite matriisimuodossa käyttäen monisteen yhtälöitä (2.7) ($\mathbf{A}\boldsymbol{\beta} = \mathbf{c}$) ja (2.9) ($\boldsymbol{\beta} = \mathbf{C}\boldsymbol{\phi} + \mathbf{d}$).

(iii) Esitä rajoitetun PNS-estimaatin lauseke, kun rajoite $\beta_2 = \beta_4$ muotoillaan yhtälön (2.9) mukaisesti.

(iv) Pohdi rajoitteen $\beta_2 = \beta_4$ tulkintaa, kun selitettävä muuttuja on lääkärien kuukausipalkka, selittävä muuttuja on työvuosien määrä ja sukupuoli määrää ryhmäjaon.