Lineaariset mallit, kl 2015, Harjoitus 3, viikko 15

1. Tarkastellaan lineaarista mallia $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon} \sim \mathsf{N}\left(\mathbf{0}, \sigma^2\mathbf{I}_n\right)$ ($\boldsymbol{\beta} \in \mathbb{R}^p$, $\sigma^2 > \mathbf{0}$, $\mathsf{r}\left(\mathbf{X}\right) = p$). Osoita, että residuaalivektorille $\hat{\boldsymbol{\varepsilon}}$ pätee $\hat{\boldsymbol{\varepsilon}} = \left(\mathbf{I}_n - \mathbf{P}\right)\boldsymbol{\varepsilon}$, jossa $\mathbf{P} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$ (ks. monisteen s. 9) ja laske tämän ja liitteen A.1 tulosten avulla $\mathsf{E}\left(\hat{\boldsymbol{\varepsilon}}\right)$, $\mathsf{Cov}\left(\hat{\boldsymbol{\varepsilon}}\right)$ ja $\mathsf{Cov}(\hat{\boldsymbol{\varepsilon}}, \hat{\boldsymbol{\beta}})$.

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

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

Muotoile tilanne lineaarisena mallina käyttäen matriiseja ja muodosta parametrien β_1 ja β_2 SU-estimaattorit. Selvitä lisäksi Lauseen 2.1(i) tulosta käyttäen 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?

3. Olkoon oikea (täysiasteinen) malli $\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$ ($\boldsymbol{\varepsilon} \sim \mathsf{N}\left(\mathbf{0}, \sigma^2\mathbf{I}_n\right)$, $\boldsymbol{\beta}_1 \in \mathbb{R}^{p_1}$, $\boldsymbol{\beta}_2 \in \mathbb{R}^{p_2}$, $\sigma^2 > \mathbf{0}$). Oletetaan, että $\boldsymbol{\beta}_1$ estimoidaan kuitenkin käyttäen mallia, josta $\mathbf{X}_2$ on jätetty pois (eli malliyhtälö on $\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_*$). Laske näin saadun $\boldsymbol{\beta}_1$:n PNS-estimaattorin odotusarvo ja selvitä milloin tämä estimaattori on harhaton?

4. Tarkastellaan tehtävän 1 mallia erikoistapauksessa $p = 1$, jolloin malliyhtälö on havaintoyksiköittäin ilmaistuna $Y_i = \beta x_i + \varepsilon_i$, $i = 1, ..., n$, ja $\mathbf{X}$ on vektori $\mathbf{x} = [x_1 \;\; \cdots \;\; x_n]'$. (i) Osoita, että PNS-estimaattori parametrille β on

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2}.$$

(ii) Oletetaan nyt, että selittävän muuttujan x_i paikalla mallissa ja PNS-estimaattorissa $\hat{\beta}$ on satunnainen X_i. Osoita PNS-estimaattorin harhattomuus eli $\mathsf{E}(\hat{\beta}) = \beta$, kun monisteen s. 11 mainittu ehto (a) eli $(X_1, ..., X_n) \perp\!\!\!\perp (\varepsilon_1, ..., \varepsilon_n)$ pätee. Perustele lisäksi, miksei harhattomuus välttämättä päde ilman tätä riippumattomuutta. (*Vihje*: Kohdassa (ii) kannattaa sijoittaa $Y_i = \beta X_i + \varepsilon_i$. Viimeisessä kohdassa ei vaadita tarkkaa matemaattista perustelua.)