Lineaariset mallit, kl 2011, Harjoitus 3, viikko 14

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

2. Jatkoa edelliselle tehtävälle. Osoita, että $\hat{\mathbf{y}}'\hat{\mathbf{y}} = \mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}}$ ja että $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2\bar{x}_2 - \cdots - \hat{\beta}_p\bar{x}_p$, kun mallissa on vakio eli $x_{i1} = 1$ kaikilla $i = 1, ..., n$. Tässä $\bar{x}_j = (x_{1j} + \cdots + x_{nj})/n$, jossa x_{ij} on matriisin $\mathbf{X}$ yleinen alkio ($i = 1, ..., n$, $j = 1, ..., p$), $\hat{\boldsymbol{\beta}} = [\hat{\beta}_1 \ \cdots \ \hat{\beta}_p]'$ ja muut merkinnät ovat kuten monisteessa (ks. s. 8-10).

3. Tehtävässä 1.4 on todettu, että $\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n\bar{y}^2 = \mathbf{y}'(\mathbf{I}_n - \mathbf{J})\,\mathbf{y}$, jossa $\mathbf{y} = [y_1 \ \cdots \ y_n]'$ ja $\mathbf{J} = \mathbf{1}_n\left(\mathbf{1}_n'\mathbf{1}_n\right)^{-1}\mathbf{1}_n'$ eli monisteen s. 10 merkinnöin SST $= \mathbf{y}'(\mathbf{I}_n - \mathbf{J})\,\mathbf{y}$. Osoita, että SSR $= \mathbf{y}'(\mathbf{P} - \mathbf{J})\,\mathbf{y}$ ja SSE $= \mathbf{y}'(\mathbf{I}_n - \mathbf{P})\,\mathbf{y}$, jossa $\mathbf{P} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$ ja kuten monisteen vastaavassa kohdassa on myös tässä matriisin $\mathbf{X}$ ensimmäinen sarake ykkösvektori $\mathbf{1}_n$. Tästä saat vaihtoehtoisen perustelun monisteessa s. 10 esitetylle tulokselle SST = SSR + SSE.

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 $Y_i = \beta X_i + \varepsilon_i$ ja tarkastella erotuksen $\hat{\beta} - \beta$ odotusarvoa. Kaikkien tarvittavien odotusarvojen oletetaan olevan äärellisinä olemassa. Viimeisessä kohdassa ei vaadita tarkkaa matemaattista perustelua.)

5. Tarkastellaan tehtävän 1 mallia yhden selittävän muuttujan lineaarisen regressiomallin tapauksessa, jolloin malliyhtälö on havaintoyksiköittäin ilmaistuna $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, $i = 1, ..., n$. Merkitään $\hat{\boldsymbol{\beta}} = [\hat{\beta}_1 \ \ \hat{\beta}_2]'$ ja $\bar{x} = (x_1 + \cdots + x_n)/n$. (i) Muodosta Lauseen 2.1(i) tulosta käyttäen $\mathsf{Cov}(\hat{\boldsymbol{\beta}})$ ja edelleen $\mathsf{Var}(\hat{\beta}_2)$ ja $\mathsf{Var}(\hat{\beta}_1 + \hat{\beta}_2\bar{x})$. (ii) Oletetaan, että selittävien muuttujien arvot $x_1, ..., x_n$ voidaan valita vapaasti väliltä $[c, d]$. Miten ne on valittava, jos $\mathsf{Var}(\hat{\beta}_2)$ halutaan minimoida? Onko tämä valinta muuten järkevä? (*Huom.*: Kohdassa (ii) ei vaadita yksityiskohtaisia matemaattisia todistuksia. Niissä voit myös olettaa, että n on parillinen.)