Lineaariset mallit, kl 2016, Harjoitus 3, viikko 15

1. Tarkastellaan (täysiasteista) lineaarista mallia $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon} \sim \mathsf{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_n\right)$, jossa matriisin $\mathbf{X}$ $(n \times p)$ ensimmäinen sarake on ykkösvektori $\mathbf{1}_n$. Tehtävässä 1.4 on todettu, että $\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n\bar{y}^2 = \mathbf{y}'\left(\mathbf{I}_n - \mathbf{J}\right)\mathbf{y}$, 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 $\mathrm{SST} = \mathbf{y}'\left(\mathbf{I}_n - \mathbf{J}\right)\mathbf{y}$. Osoita, että $\mathrm{SSR} = \mathbf{y}'\left(\mathbf{P} - \mathbf{J}\right)\mathbf{y}$ ja $\mathrm{SSE} = \mathbf{y}'\left(\mathbf{I}_n - \mathbf{P}\right)\mathbf{y}$, jossa $\mathbf{P} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$ (vrt. monisteen s. 10 esitetty tulos SST = SSR + SSE).

2. (Jatkoa tehtävälle 1.5) (i) Johda tehtävän 1.5 varianssianalyysimallissa parametrien $\mu_1, ..., \mu_p$ PNS-estimaattien lausekkeet.

(ii) Selvitä parametrivektorin $\boldsymbol{\mu} = \left(\mu_1, ..., \mu_p\right)$ PNS-estimaattorin $\hat{\boldsymbol{\mu}} = \left(\hat{\mu}_1, ..., \hat{\mu}_p\right)$ todennäköisyysjakauma. Ovatko estimaattorit $\hat{\mu}_1, ..., \hat{\mu}_p$ riippumattomia?

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

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

(i) Esitä tilanne tilanne lineaarisena mallina (mallin matriisiesitystä käyttäen) ja johda parametrien β_1 ja β_2 PNS-estimaattien lausekkeet.

(ii) Selvitä parametrivektorin $\boldsymbol{\beta} = [\beta_1 \ \ \beta_2]'$ PNS-estimaattorin $\hat{\boldsymbol{\beta}} = [\hat{\beta}_1 \ \ \hat{\beta}_2]'$ todennäköisyysjakauma. Ovatko estimaattorit $\hat{\beta}_1$ ja $\hat{\beta}_2$ riippumattomia?

4. Tarkastellaan yhden satunnaisen selittäjän lineaarista mallia

$$Y_i = \beta X_i + \varepsilon_i, \quad i = 1, ..., n, \quad \varepsilon_1, ..., \varepsilon_n \ \perp\!\!\!\perp, \ \varepsilon_i \sim \mathsf{N}\left(0, \sigma^2\right).$$

(i) Osoita, että parametrin β PNS-estimaattori on harhaton 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.

(ii) Perustele, miksei harhattomuus välttämättä päde ilman edellä mainittua riippumattomuutta.

Vihje: Voit olettaa, että kaikki laskelmissa tarvittavat odotusarvot ovat äärellisinä olemassa. Kohdassa (i) kannattaa käyttää yhtälöä $Y_i = \beta X_i + \varepsilon_i$. Kohdassa (ii) ei vaadita tarkkaa matemaattista perustelua.

Menemättä yksityiskohtiin todetaan, että edellisen tehtävän tuloksesta seuraa, että satunnaisten selittäjien tapauksessa monisteen s. 11 mainitun ehdon $\mathbf{X} \perp\!\!\!\perp \boldsymbol{\varepsilon}$ rikkoontuminen muuttaa PNS-estimaattorin ominaisuuksia ja tekee siitä käyttökelvottoman tilastollisessa päättelyssä. Seuraava tehtävä osoittaa, että samanlaisia ongelmia aiheutuu myös kiinteiden selittäjien tapauksessa, jos virhetermin kovarianssimatriisi ei ole oletetun kaltainen.

5. Tarkastellaan (täysiasteista kiinteiden selittäjien) lineaarista mallia $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, jossa aikaisemmasta poiketen $\boldsymbol{\varepsilon} \sim \mathsf{N}_n(\mathbf{0}, \boldsymbol{\Sigma})$ ja $\boldsymbol{\Sigma}$ on mikä tahansa positiivisesti definiitti matriisi. Johda PNS-estimaattorin $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ todennäköisyysjakauma tässä tilanteessa.

Vihje: Monisteen Liitteet A.2.4 ja A.1.