LINEAARISET MALLIT, 5 OP (aineopinnot). 12.3.–30.4.2009. Kirjallisuus: Pentti Saikkonen: Lineaarinen malli. Luennoi: yliopistonlehtori Pekka Pere.

Harjoitukset 3 (to 2.4.)

1. Monisteen sivulla 7 esitetään yhtäsuuruus

$$\sum_{i=1}^{n} \mathbf{x}_i(y_i - \mathbf{x}_i^{'}\boldsymbol{\beta}) = \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\boldsymbol{\beta}.$$

Perustele huolellisesti suureiden dimensiot ja yhtälön paikkansapitävyys komponenteittain. (Huomaa, että merkinnät yllä ovat monisteen sivun 7 ja HT 1.1:n mukaiset. Monisteen liitteessä B s:lla 51 $\mathbf{x}_i$:t ovat sivun 7 merkintöjen kanssa ristiriitaisesti n-vektoreita. Jos korvaat sivulla 51 $\mathbf{x}_i$:t $\mathbf{x}_{(i)}$:llä, niin kaikki merkinnät ovat yhdenmukaisia.)

2. Olkoon neliömatriisi $\mathbf{A}$ $(n \times n)$ (ortogonaalinen) projektio eli symmetrinen ja idempotentti. Osoita, että $\mathbf{A}$ on positiivisesti semidefiniitti eli $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ $\forall\ \mathbf{x} \in \mathbb{R}^n$ $(\mathbf{A} \geq \mathbf{0})$ ja että $\mathbf{A}$:n aste on $\mathbf{A}$:n jälki eli $\mathsf{r}(\mathbf{A}) = \mathsf{tr}\,(\mathbf{A})$. (*Vihje*: Pääakselihajotelma ja HT 1.5.)

3. Matriisi $\mathbf{I}_n - \mathbf{P} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ $(n \times n)$ projisoi $\mathbb{R}^n$:n vektorit avaruuteen $\mathcal{R}(\mathbf{X})^{\perp}$. Osoita, että matriisin $\mathbf{I}_n - \mathbf{P}$

a) aste on $n - p$ (*vihje*: HT 3.2 ja monisteen s. 15)

b) ominaisarvojen summa on $n - p$ (*vihje*: $\mathsf{tr}\,(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i$, jossa λ_i:t ovat matriisin $\mathbf{A}$ ominaisarvot — monisteen s. 50).

c) ominaisarvoista $n - p$ kappaletta on ykkösiä ja p kappaletta nollia (*vihje*: HT 1.5).

4. Esitä yhden selittäjän lineaarisen regressiomallin $Y_1, ..., Y_n$ $\perp\!\!\!\perp$, $Y_i \sim \mathsf{N}\left(\beta_1 + \beta_2 x_i, \sigma^2\right)$ normaaliyhtälöt komponenttimuodossa (ilman matriiseja) ja osoita, että niiden ratkaisuna saatavat PNS-estimaatit voidaan lausua muodossa

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad \text{ja} \qquad \hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x},$$

jossa esimerkiksi $\bar{y} = (y_1 + \cdots + y_n)/n$. Esitä $\hat{\beta}_2$ käyttäen havainnoista (y_i, x_i), $i = 1, ..., n$, laskettuja keskihajontoja ja otoskorrelaatiokerrointa. (*Huom.*: Otoskorrelaatiokertoimen määritelmä löytyy monisteen s. 11 alaviitteestä ja esimerkiksi y-havaintojen keskihajonta on $s_y = \sqrt{(n-1)^{-1} \sum_{i=1}^{n} (y_i - \bar{y})^2}$).

5. Tarkastellaan yhden selittäjän lineaarista regressiomallia kahden ryhmän tapauksessa:

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

jossa $n_1, n_2 > 2$.

a) Osoita, että kysymyksessä on lineaarinen malli käyttäen lineaarisen mallin matriisiesitystä (eli $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$).

b) Osoita, että parametrien β_1 ja β_2 (vastaavasti β_3 ja β_4) PNS-estimaatit saadaan kuten edellisessä tehtävässä käyttäen vain havaintoyksiköiden $i = 1, \ldots, n_1$ (vastaavasti $i = n_1 + 1, \ldots, n$) havaintoja.

c) Miten tulkitset ehtoa $\beta_2 = \beta_4$, kun y on lääkärien kuukausipalkka, x on työvuosien määrä ja sukupuoli määrää ryhmäjaon?

6. Tutkitaan yksinkertaista yhden selittäjän lineaarista regressiomallia $Y_i \sim \mathsf{N}\left(\beta x_i, \sigma^2\right)$, $i = 1, \ldots, n$. (Ainoa selittävä muuttuja ei ole tässä välttämättä vakio.) Osoita, että PNS-estimaattori parametrille β on

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

7. Osoitetaan, että jos lineaarisessa regressiomallissa $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ ($\boldsymbol{\varepsilon} \sim \mathsf{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$) selittäjät ovat stokastisia, niin kerroinparametrien PNS-estimaattori $\hat{\boldsymbol{\beta}}$ $(p \times 1)$ ei ole välttämättä normaalijakautunut (vaikka kiinteiden selittäjien tilanteessa on lauseen 2.1 mukaan). Oletetaan, että $p = n = 1$ eli että mallissa on vain yksi selittäjä ja että havaintoja on vain yksi! Oletetaan lisäksi, että $\sigma^2 = 1$, että selittäjä $X \sim \mathsf{N}(0, 1)$ ja että mallin jäännös ε ja selittäjä X ovat toisistaan riippumattomia. Todista, että tällöin $\hat{\beta} - \beta$ on Cauchy-jakautunut. (*Vihje1:* Sijoita $\hat{\beta}$:n kaavaan $Y_1 = \beta X_1 + \varepsilon_1$. *Vihje2:* Satunnaismuuttuja, joka on kahden riippumattoman standardinormaalin satunnaismuuttujan osamäärä, noudattaa Cauchy-jakaumaa.)