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

Harjoitukset 6 (to 23.4.)

1. Tarkastellaan yhden selittäjän lineaarista regressiomallia $Y_1, ..., Y_n \ \perp\!\!\!\perp\ , Y_i \sim \mathsf{N}\left(\beta_1 + \beta_2 x_i, \sigma^2\right)$. Johda F-testisuure nollahypoteesille $\beta_2 = 0$ ja osoita, että testisuure voidaan lausua (tavanomaisin merkinnöin) muodossa

$$F = \frac{\hat{\beta}_2^2 \sum_{i=1}^n (x_i - \bar{x})^2}{S^2}.$$

(*Vihje*: Mallin matriisiesitys ja F-testisuureen yleinen lauseke (ks. monisteen s. 18). Testisuureen haetun lausekkeen johtamisessa tarvitset myös 2×2 -matriisin käänteismatriisin laskukaavaa.)

2. Jatkoa edelliselle ja HT 3.4:lle. (i) Osoita, että edellisen tehtävän mallissa

$$\mathrm{SSE} = \sum_{i=1}^n (y_i - \bar{y})^2 - \hat{\beta}_2^2 \sum_{i=1}^n (x_i - \bar{x})^2 = \left(1 - r_{xy}^2\right)(n-1)\, s_y^2.$$

Yllä r_{xy} on havainnoista (y_i, x_i), $i = 1, ..., n$, laskettu korrelaatiokerroin (ks. monisteen alaviite s. 11), $s_y^2 = (n-1)^{-1} \sum_{i=1}^n (y_i - \bar{y})^2$ ja residuaalineliösumma SSE on määritelty monisteen sivulla 10. (ii) Osoita tämän avulla, että edellisen tehtävän F-testi on yhtäpitävä t-testin

$$\sqrt{n-2}\, r_{xy} \Big/ \sqrt{1 - r_{xy}^2} \overset{H}{\sim} t_{n-2}$$

kanssa. Minkä hypoteesin testiksi tämä voidaan myös tulkita? (*Vihje*: Monisteen s. 10 oleva yhtälö SST = SSR + SSE, jossa SSR:ää voidaan muokata HT 3.4:n tuloksen avulla. Muista myös, että nyt $S^2 = \mathrm{SSE}/(n-2)$).

3. Tarkastellaan jakson 3.2 alun tilannetta (s. 19), jossa malliyhtälö on $Y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$, $i = 1, ..., n$, ja testattava hypoteesi $H : \beta_2 = \cdots = \beta_p = 0$. Osoita ensin, että $\mathrm{SSE} = \left(1 - \mathsf{R}^2\right) \sum_{i=1}^n (y_i - \bar{y})^2$, jossa R^2 on selitysaste ja SSE on residuaalineliösumma (ks. s. 10), ja tämän perusteella edelleen, että F-testisuure edellä mainitulle hypoteesille voidaan kirjoittaa

$$F = (n-p)\, \mathsf{R}^2 / (p-1)\left(1 - \mathsf{R}^2\right).$$

Tulkitse kaava sanoin.

4. Jaetaan lineaarisen mallin $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ selittäjät kahteen ryhmään:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}. \tag{1}$$

Yllä $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^{'} \quad \boldsymbol{\beta}_2^{'}]'$, $\boldsymbol{\beta}_1$ ja $\boldsymbol{\beta}_2$ ovat $p_1 \times 1$ ja $p_2 \times 1$ -vektoreita, $\mathbf{X}_1$ ja $\mathbf{X}_2$ ovat täysiasteisia $n \times p_1$ ja -$n \times p_2$-matriiseja (p_1,$p_2 \geq 1$) ja $\mathbf{X}$ on täysiasteinen $n \times p$ -matriisi. Merkitään (1*):llä typistettyä mallia

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}^*. \tag{1*}$$

a) Olkoot ryhmien selittäjät keskenään ortogonaalisia eli $\mathbf{X}_1^{'}\mathbf{X}_2 = \mathbf{0}$ ($p_1 \times p_2$ -nollamatriisi). Osoita, että malleista (1) ja (1*) lasketut PNS-estimaattorit $\hat{\boldsymbol{\beta}}_1$:t ovat samoja.

b) Osoita, että ylipäänsä malleista (1) ja (1*) lasketut PNS-estimaattorit $\hat{\boldsymbol{\beta}}_1$:t ovat eri suuria.

5. Johda F-testi nollahypoteesille $\mu = \mu_0$, kun mallina on $Y_1, ..., Y_n \sim \mathsf{N}\left(\mu, \sigma^2\right)$ $\perp\!\!\!\perp$ ja totea, että se on yhtäpitävä tilastollisen päättelyn kurssilla (monisteen kohdassa 5.4.2) esitetyn t-testin kanssa. (*Vihje*: Mallin matriisiesitys (ks. monisteen s. 4), F-testisuureen yleinen lauseke (ks. monisteen s. 18) ja tulos $F_{1,k} \doteq t_k^2$ ("ovat samoja jakaumia").)

6. Jatkoa HT 5.5:lle. Johda F-testi nollahypoteesille $\mu_1 = \mu_2$ ja osoita, että yhtäpitävä testi voidaan perustaa t_{n-2}-jakaumaa noudattavaan t-testisuureeseen ($n = n_1 + n_2$). (*Vihje*: Mallin matriisiesitys (ks. monisteen s. 5), F-testisuureen yleinen lauseke (ks. monisteen s. 18) ja tulos $F_{1,k} \doteq t_k^2$.)

7. Jatkoa edelliselle tehtävälle. Oletetaan, että hypoteesi $\mu_1 = \mu_2$ on tullut hylätyksi. Muodosta $100 \times (1 - \alpha)$ prosentin luottamusväli erotukselle $\mu_1 - \mu_2$.

8. On estimoitu PNS:llä malli $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon} \sim \mathsf{N}\left(\mathbf{0}, \sigma^2\mathbf{I}_n\right)$ ($\boldsymbol{\beta} \in \mathbb{R}^p$, $\sigma^2 \in \mathbb{R}^+$, $\mathsf{r}\left(\mathbf{X}\right) = p$) ja saatu estimaatti $\hat{\boldsymbol{\beta}}$. Ennustetaan Y_{n+1}:n arvoa luontevalla kaavalla $\mathbf{x}_{n+1}^{'}\hat{\boldsymbol{\beta}}$ (halutaan laskea ennuste selitettävälle muuttujalle, kun selittävät muuttujat ovat $\mathbf{x}_{n+1}$:n mukaiset).

a) Osoita, että $\mathsf{E}(\mathbf{x}_{n+1}^{'}\hat{\boldsymbol{\beta}}) = \mathbf{x}_{n+1}^{'}\boldsymbol{\beta} = \mathsf{E}(Y_{n+1})$.

b) Osoita, että $\mathsf{Var}(\mathbf{x}_{n+1}^{'}\hat{\boldsymbol{\beta}}) = \sigma^2\mathbf{x}_{n+1}^{'}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{n+1}$.

c) Osoita, että $\mathsf{Var}(Y_{n+1} - \mathbf{x}_{n+1}^{'}\hat{\boldsymbol{\beta}}) = \sigma^2[1 + \mathbf{x}_{n+1}^{'}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{n+1}]$. Onko varianssi $\mathsf{Var}(Y_{n+1} - \mathbf{x}_{n+1}^{'}\hat{\boldsymbol{\beta}})$ suurempi vai pienempi kuin σ^2 (perustele)?

d) Olkoon mallin ainoa selittäjä vakio. Mitä ovat tällöin $\mathbf{x}_{n+1}^{'}\hat{\boldsymbol{\beta}}$, $\mathsf{Var}(\mathbf{x}_{n+1}^{'}\hat{\boldsymbol{\beta}})$ ja $\mathsf{Var}(Y_{n+1} - \mathbf{x}_{n+1}^{'}\hat{\boldsymbol{\beta}})$ ja mitä tällöin varianssille $\mathsf{Var}(Y_{n+1} - \mathbf{x}_{n+1}^{'}\hat{\boldsymbol{\beta}})$ tapahtuu, kun havaintojen lukumäärä n kasvaa kohti ääretöntä?

Tulkitse tulokset a)–d) sanoin.