

Peatükk 3

Parameeterfunktsiooni $\lambda^T \beta$ hinnangust

Vaatame hinnatavat parameeterfunktsiooni $\lambda^T \beta = \mathbf{v}^T \mathbf{X} \beta$. Nagu juba teame, on hinnang $\lambda^T \hat{\beta}$ nihketa, $E(\lambda^T \hat{\beta}) = \lambda^T \beta$. Ka hinnangu dispersioon on lihtsalt leitav:

$$\begin{aligned} D(\lambda^T \hat{\beta}) &= D(\mathbf{v}^T \mathbf{P}_X \mathbf{y}) \\ &= \mathbf{v}^T \mathbf{P}_X D(\mathbf{y}) \mathbf{P}_X \mathbf{v} \\ &= \sigma^2 \mathbf{v}^T \mathbf{P}_X \mathbf{v} \\ &= \sigma^2 \lambda^T (\mathbf{X}^T \mathbf{X})^{-1} \lambda. \end{aligned}$$

Paneme tähele, et juhul kui β on ise hinnatav, siis on ka hinnangu $\hat{\beta}$ dispersioon üheselt määratud:

$$D(\hat{\beta}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$$

ja sellisel juhul võime kirjutada:

$$D(\lambda^T \hat{\beta}) = \lambda^T D(\hat{\beta}) \lambda.$$

Juhul, kui vaatlusvektor on normaaljaotusega, $\mathbf{y} \sim N(\mathbf{X}\beta; \sigma^2 \mathbf{I})$, siis on ka hinnangu $\lambda^T \hat{\beta}$ jaotuseks normaaljaotus (sest lineaarteisendus ei hävita normaaljaotust):

$$\lambda^T \hat{\beta} \sim N(\lambda^T \beta; \lambda^T (\mathbf{X}^T \mathbf{X})^{-1} \lambda \sigma^2).$$

Juhul kui β ise on hinnatav, siis

$$\hat{\beta} \sim N(\beta; (\mathbf{X}^T \mathbf{X})^{-1} \sigma^2).$$

Et saada reaalselt arvutatavat $\hat{\beta}$ dispersiooni hinnangut, peaksime leidma viisi hindamaks parameetrit σ^2 . Soovitav oleks nihketa hinnang — poleks ju ilus oma töö täpsuse iseloomustamisel tehtavat viga pisemaks muuta. Alustuseks uurime, milline tuleb vigade ruutude summa $\mathbf{y}^T(\mathbf{I} - \mathbf{P}_X)\mathbf{y}$ keskväärus. Esmalt aga tõestame ühe abitulemuse:

Lemma 3.1 Olgu $\mathbf{A}$ suvaline $n \times n$ matriks, $E\mathbf{y} = \boldsymbol{\mu}$ ja $D\mathbf{y} = \mathbf{V}$. Siis

$$E(\mathbf{y}^T \mathbf{A} \mathbf{y}) = \text{tr}(\mathbf{A} \mathbf{V}) + \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu}.$$

Tõestus.

$$\begin{aligned} E(\mathbf{y}^T \mathbf{A} \mathbf{y}) &= E(\text{tr}(\mathbf{A} \mathbf{y} \mathbf{y}^T)) \\ &= \text{tr}(\mathbf{A} E(\mathbf{y} \mathbf{y}^T)) \\ &= \text{tr}(\mathbf{A}(\mathbf{V} + \boldsymbol{\mu} \boldsymbol{\mu}^T)) \\ &= \text{tr}(\mathbf{A} \mathbf{V}) + \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu}. \end{aligned}$$

Varustatud ülaltoodud abitulemusega, on edasine muidugi lihtne:

$$\begin{aligned} E[\mathbf{y}^T(\mathbf{I} - \mathbf{P}_X)\mathbf{y}] &= \text{tr}((\mathbf{I} - \mathbf{P}_X)\sigma^2) + \boldsymbol{\beta}^T \mathbf{X}^T(\mathbf{I} - \mathbf{P}_X)\mathbf{X}\boldsymbol{\beta} \\ &= \text{tr}((\mathbf{I} - \mathbf{P}_X)\sigma^2) \\ &= \sigma^2 \text{tr}(\mathbf{I} - \mathbf{P}_X) \\ &= \sigma^2(n - \text{rank}(\mathbf{X})). \end{aligned}$$

Seega võiksime kasutada σ^2 nihketa hinnanguna suurust:

$$\hat{\sigma}^2 = \mathbf{y}^T(\mathbf{I} - \mathbf{P}_X)\mathbf{y} / \text{rank}(\mathbf{I} - \mathbf{P}_X)$$

antud suurust tuntakse ka nime keskmine ruutviga ehk MSE (Mean Square Error) all.

Definitsioon 3.1 Kui $p \times 1$ vektor $\mathbf{y}$ on normaaljaotusega, $\mathbf{y} \sim N(\boldsymbol{\mu}; \mathbf{I})$, siis $\mathbf{y}^T \mathbf{y}$ on mittetsentraalse hii-ruut (χ^2) jaotusega juhuslik suurus vabadusastmete arvuga p ja mittetsentraalsuse parameetriga $n\boldsymbol{\mu}^T \boldsymbol{\mu}$,

$$\mathbf{y}^T \mathbf{y} \sim \chi_{df=p; ncp=\boldsymbol{\mu}^T \boldsymbol{\mu}}^2$$

Teoreem 3.1 Olgu $\mathbf{y} \sim N(\boldsymbol{\mu}; \mathbf{I})$, ja olgu maatriks $\mathbf{A}$ ortogonaalne projektor (sümmeetriline ja idempotentne). Siis

$$\mathbf{y}^T \mathbf{A} \mathbf{y} \sim \chi_{df=rank(\mathbf{A}); ncp=\boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu}}^2.$$

Tõestus.

Maatriksi $\mathbf{A}$ saame esitada kahe ortogonaalse maatriksi korrutisena (maatriksi spektraallahutus, idempotentse maatriksi omaväärtused saavad olla vaid kas 0 või 1): $\mathbf{A} = \mathbf{O} \mathbf{O}^T$, kusjuures $\mathbf{O}^T \mathbf{O} = \mathbf{I}_r$. Seega

$$\mathbf{O}^T \mathbf{y} \sim N(\mathbf{O} \boldsymbol{\mu}; \mathbf{O} \mathbf{O}^T (= \mathbf{I}_r))$$

ja järelikult (definitsiooni kohaselt)

$$\begin{aligned} (\mathbf{O} \mathbf{y})^T (\mathbf{O} \mathbf{y}) &\sim \chi_{rank(\mathbf{A}), \boldsymbol{\mu}^T \mathbf{O}^T \mathbf{O} \boldsymbol{\mu}}^2 \\ \mathbf{y}^T \mathbf{A} \mathbf{y} &\sim \chi_{rank(\mathbf{A}), \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu}}^2. \end{aligned}$$

□

Märkus 1: analoogne tulemus on tõestatud ka veidi üldisemal kujul:

Teoreem 3.2 Kui $\mathbf{y} \sim N(\boldsymbol{\mu}; \mathbf{V})$, siis $\mathbf{y}^T \mathbf{A} \mathbf{y} \sim \chi_{r, ncp}^2$, $r := rank(\mathbf{A})$, $ncp := \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu}$ siis ja ainult siis kui maatriks $\mathbf{A} \mathbf{V}$ on idempotentne.

Märkus 2: Kuna maatriks $\mathbf{I} - \mathbf{P}_{\mathbf{X}}$ on ortogonaalne projektor (sümmeetriline ja idempotentne), siis prognoosijääkide ruutude summa jagatud σ^2 -ga on hii-ruut jaotusega:

$$\begin{aligned} \mathbf{y}^T (\mathbf{I} - \mathbf{P}_{\mathbf{X}}) \mathbf{y} / \sigma^2 &\sim \chi_{rank(\mathbf{I} - \mathbf{P}_{\mathbf{X}})}^2 \\ &\sim \chi_{n - rank(\mathbf{X})}^2 \end{aligned} \quad (3.1)$$

Saadud tulemust saab kasutada näiteks σ^2 -le usaldusintervalli leidmiseks. Kui tähistame vabadusastmete arvuga df hii-ruut jaotuse α kvantiili sümbooliga $\chi_{\alpha; df}$ siis võime tulemust 3.1 kasutades kirjutada

$$\begin{aligned} P(\chi_{\alpha/2; n-p} \leq \mathbf{y}^T (\mathbf{I} - \mathbf{P}_{\mathbf{X}}) \mathbf{y} / \sigma^2 \leq \chi_{1-\alpha/2; n-p}) &= 1 - \alpha \\ P\left(1/\chi_{\alpha/2; n-p} \geq \frac{\sigma^2}{\mathbf{y}^T (\mathbf{I} - \mathbf{P}_{\mathbf{X}}) \mathbf{y}} \geq 1/\chi_{1-\alpha/2; n-p}\right) &= 1 - \alpha \\ P(\mathbf{y}^T (\mathbf{I} - \mathbf{P}_{\mathbf{X}}) \mathbf{y} / \chi_{\alpha/2; n-p} \geq \sigma^2 \geq \mathbf{y}^T (\mathbf{I} - \mathbf{P}_{\mathbf{X}}) \mathbf{y} / \chi_{1-\alpha/2; n-p}) &= 1 - \alpha. \end{aligned}$$

Ehk siis $1 - \alpha$ usaldusintervalli jääkide dispersioonile avaldub kujul:

$$\mathbf{y}^T (\mathbf{I} - \mathbf{P}_{\mathbf{X}}) \mathbf{y} / \chi_{1-\alpha/2; n-p} \cdots \mathbf{y}^T (\mathbf{I} - \mathbf{P}_{\mathbf{X}}) \mathbf{y} / \chi_{\alpha/2; n-p}.$$

Üritame nüüd kokku panna teadmised hinnatava parameeterfunktsiooni hinnangu jaotuse ja prognoosijääkide ruutude summa kohta. Esmalt tähele-
dame, et juhuslikud suurused $\mathbf{P}_X \mathbf{y}$ ja $(\mathbf{I} - \mathbf{P}_X) \mathbf{y}$ on sõltumatud (juhul kui
 $\mathbf{y}$ on normaaljaotusega piisab, kui näitame, et kovariatsioon on 0):

$$\begin{aligned} \text{cov}(\mathbf{P}_X \mathbf{y}, (\mathbf{I} - \mathbf{P}_X) \mathbf{y}) &= \mathbf{P}_X \text{cov}(\mathbf{y}, \mathbf{y}) (\mathbf{I} - \mathbf{P}_X)^T \\ &= \sigma^2 \mathbf{P}_X (\mathbf{I} - \mathbf{P}_X)^T \\ &= 0 \end{aligned}$$

Aga kui Y_1 ja Y_2 on sõltumatud juhuslikud suurused $Y_1 \perp Y_2$, siis on sõltu-
matud ka nende juhuslike suuruste mistahes funktsioonid, $f(Y_1) \perp f(Y_2)$.

Hinnatava parameeterfunktsiooni $\lambda^T \beta$ hinnangu jaotus oli normaaljao-
tus:

$$\begin{aligned} \lambda^T \hat{\beta} &\sim N(\lambda^T \beta, \lambda^T (\mathbf{X}^T \mathbf{X})^{-1} \lambda \sigma^2) \\ \frac{\lambda^T \hat{\beta} - \lambda^T \beta}{\sqrt{\lambda^T (\mathbf{X}^T \mathbf{X})^{-1} \lambda \sigma^2}} &\sim N(0; 1) \end{aligned}$$

ja jääkide ruutude summa jaotus oli hii-ruut jaotus. Kui $Y_1 \sim N(0; 1)$ ja
 $Y_2 \sim \chi_k^2$, ja kui $Y_1 \perp Y_2$, siis $t = Y_1 / \sqrt{Y_2/k} \sim t_k$. Seega

$$\begin{aligned} t &:= \frac{\lambda^T \hat{\beta} - \lambda^T \beta}{\sqrt{\lambda^T (\mathbf{X}^T \mathbf{X})^{-1} \lambda \sigma^2}} / \sqrt{\frac{\mathbf{y}^T (\mathbf{I} - \mathbf{P}_X) \mathbf{y}}{(n-p)\sigma^2}} \\ &= \frac{\lambda^T \hat{\beta} - \lambda^T \beta}{\sqrt{MSE \cdot \lambda^T (\mathbf{X}^T \mathbf{X})^{-1} \lambda}} \end{aligned}$$

on t -jaotusega juhuslik suurus, $t \sim t_{n-p}$, kus $p := \text{rank}(\mathbf{X})$.

Milleks saame seda teadmist kasutada? Muidugi hüpoteeside testimiseks:

$$t := \frac{\lambda^T \hat{\beta} - \lambda^T \beta_0}{\sqrt{MSE \cdot \lambda^T (\mathbf{X}^T \mathbf{X})^{-1} \lambda}} \stackrel{H_0}{\sim} t_{n-p}.$$

Märkus:

Kuna $D(\lambda^T \hat{\beta}) = \lambda^T (\mathbf{X}^T \mathbf{X})^{-1} \lambda \sigma^2$ ja $E(MSE) = \sigma^2$ siis nihketa hinnanguks
hinnangu $\lambda^T \hat{\beta}$ dispersioonile on

$$\widehat{D(\lambda^T \hat{\beta})} = MSE \cdot \lambda^T (\mathbf{X}^T \mathbf{X})^{-1} \lambda$$

Seega võime t -statistiku ülestähendamisel kasutada ka ehk tuttavamana näivat kirjpilti:

$$t := \frac{\boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}} - \boldsymbol{\lambda}^T \boldsymbol{\beta}_0}{\sqrt{\widehat{\mathbf{D}(\boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}})}}.$$

Samuti saame välja kirjutada $(1 - \alpha)$ -usaldusintervalli $\boldsymbol{\lambda}^T \boldsymbol{\beta}$ -le:

$$P \left(t_{\alpha/2} \leq \frac{\boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}} - \boldsymbol{\lambda}^T \boldsymbol{\beta}}{\sqrt{\widehat{\mathbf{D}(\boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}})}} \leq t_{1-\alpha/2} \right) = 1 - \alpha$$

$$P \left(\boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}} + t_{\alpha/2} \sqrt{\widehat{\mathbf{D}(\boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}})}} \leq \boldsymbol{\lambda}^T \boldsymbol{\beta} \leq \boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}} + t_{1-\alpha/2} \sqrt{\widehat{\mathbf{D}(\boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}})}} \right) = 1 - \alpha,$$

kus $t_{1-\alpha/2}$ tähistab vabadusastmete arvuga $n - p$ t -jaotuse $1 - \alpha/2$ -kvantiili.

Ülesanded*

1. Vaata lihtsat dispersioonanalüüsi mudelit

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1, 2; j = 1, \dots, n_i.$$

Defineerime $\boldsymbol{\beta}^T = (\mu, \alpha_1, \alpha_2)$. Antud mudeli parameetreid hinnati muis- te kahe erineva teadlase poolt (kahes sõltumatus uuringus). Esimeses uuringus $n_1 = 2; n_2 = 5$; teises uuringus $n_1 = 10; n_2 = 30$. Esimesest uuringust on säilinud suuruse $\boldsymbol{\lambda}^T \boldsymbol{\beta}$ hinnang $\boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}}$, kus $\boldsymbol{\lambda}^T = (0; 1; -1)$. Paraku pole aga esimese uuringu autor kirja pannud jääkide dispersiooni hinnangut. Teisest uuringust on säilinud jääkide dispersiooni hinnang MSE_2 , kuid pole paraku säilinud parameetervektori $\boldsymbol{\beta}$ hinnangut. Tahame testida hüpoteesi $H_0 : \boldsymbol{\lambda}^T \boldsymbol{\beta} = 0$ ja kasutame selleks teststatistikut

$$t := \frac{\boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}}}{\sqrt{MSE_2 \cdot \boldsymbol{\lambda}^T (\mathbf{X}^T \mathbf{X})^{-1} \boldsymbol{\lambda}}}.$$

- Kas t -statistiku valemis peaks mudelimaatriks $\mathbf{X}$ olema 1. uuringu mudelimaatriks või 2. uuringu mudelimaatriks?
- Näita, et ülaltoodud t -statistik on nullhüpoteesi kehtides t -jaotusega. Leia antud t -statistiku vabadusastmete arv.