# Peatükk 3

# Parameeterfunktsiooni $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ hinnangust

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

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

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

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

ja sellisel juhul võime kirjutada:

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

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

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

Juhul kui $\boldsymbol{\beta}$ ise on hinnatav, siis

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

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

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

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

Tõestus.

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

Varustatud ülaltoodud abitulemusega, on edasine muidugi lihtne:

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

Seega võiksime kasutada $\sigma^2$ nihketa hinnanguna suurust:

$$\hat{\sigma}^2 = \mathbf{y}^T(\mathbf{I}-\mathbf{P_X})\mathbf{y}/\mathrm{rank}(\mathbf{I}-\mathbf{P_X})$$

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* ($\chi^2$) *jaotusega juhuslik suurus vabadusastmete arvuga* $p$ *ja mittetsentraalsuse parameetriga* $ncp = \boldsymbol{\mu}^T\boldsymbol{\mu}$,

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

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

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

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

$$\mathbf{O}^T\mathbf{y} \sim N(\mathbf{O}^T\boldsymbol{\mu}; \mathbf{O}^T\mathbf{O}(= \mathbf{I}_{\mathrm{rank}(\mathbf{A})}))$$

ja järelikult (definitsiooni kohaselt)

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

□

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

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

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

$$\begin{aligned} \mathbf{y}^T(\mathbf{I} - \mathbf{P_X})\mathbf{y}/\sigma^2 &\sim \chi^2_{\mathrm{rank}(\mathbf{I}-\mathbf{P_X})} \\ &\sim \chi^2_{n-\mathrm{rank}(\mathbf{X})} \end{aligned} \tag{3.1}$$

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

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

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

$$\mathbf{y}^T(\mathbf{I}-\mathbf{P_X})\mathbf{y}/\chi_{1-\alpha/2;n-p} \ldots \mathbf{y}^T(\mathbf{I}-\mathbf{P_X})\mathbf{y}/\chi_{\alpha/2;n-p}.$$

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

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

Aga kui $Y_1$ ja $Y_2$ on sõltumatud juhuslikud suurused $Y_1 \perp Y_2$, siis on sõltumatud ka nende juhuslike suuruste mistahes funktsioonid, $f_1(Y_1) \perp f_2(Y_2)$.

Hinnatava parameeterfunktsiooni $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ hinnangu jaotus oli normaaljaotus:

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

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

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

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

Milleks saame seda teadmist kasutada? Muidugi hüpoteeside testimiseks:

$$t := \frac{\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}-\boldsymbol{\lambda}^T\boldsymbol{\beta}_0}{\sqrt{MSE\cdot\boldsymbol{\lambda}^T(\mathbf{X}^T\mathbf{X})^-\boldsymbol{\lambda}}} \overset{H_0}{\sim} t_{n-p}.$$

Märkus:

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

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

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

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

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

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

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

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

## Ülesanded

1. Vaata lihtsat dispersioonanalüüsi mudelit

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

Defineerime $\boldsymbol{\beta}^T = (\mu, \alpha_1, \alpha_2)$. Antud mudeli parameetreid hinnati muiste 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

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

   - 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.