2.5 Identifitseeritavus ja hinnatavus

Kas teatud liiki andmed põhimõtteliselt võiksid sisaldada mingit informatsiooni meid huvitava näitaja kohta? Sellele küsimusele aitavad vastata mõisted *identifitseeritavus* ja *hinnatavus*.

Definitsioon 2.1 *Juhusliku suuruse jaotuse* ***parameeter*** θ ***on identifitseeritav****, kui kahe parameetri* θ *võimaliku väärtuse* θ_1 *ja* θ_2 *korral järeldub jaotusfunktsioonide identsusest parameetri väärtuste võrdsus:*

$$F(x; \theta_1) = F(x; \theta_2) \ \forall x \Rightarrow \theta_1 = \theta_2.$$

Antud definitsioonis võib parameeter θ olla kas skalaar või vektor ja ka juhuslik suurus X mille jaotusfunktsiooni vaatame võib olla nii ühe- kui ka mitmemõõtmeline juhuslik suurus. Pidevate suuruste korral võib definitsioonis jaotusfunktsioonide identsuse asemel nõuda ka tihedusfunktsioonide identsust (tihedusfunktsioonide identsusest järeldub jaotusfunktsioonide identsus).

Millised parameetrid siis on identifitseeritavad, millised mitte? Toome paar näidet. Oletame et vaatlusandmeteks on 10 (ausa) mündiviske tulemused. Parameetriks mille vastu huvi tunneme on kojamehe vanaema sünniaasta. Kui kojamehe vanaema sünniaastaks on 1901, siis on kõigi võimalike katseseeriate nägemise tõenäosus $(1/2)^{10}$. Täpselt samasugune on kõigi katseseeriate tõenäosus ka siis, kui kojamehe vanaema sünniaastaks oleks 1932. Seega pole parameeter „kojamehe vanaema sünniaasta“ mündivisete põhjal identifitseeritav, me ei saa kasutada müdivisete tulemusi statistiliselt pädevate järelduste tegemiseks kojamehe vanaema vanuse kohta.

Teine näide. Viskame kaks korda münti, millel kulli saamise tõenäosus on p. Vaatlusvektori võimalikud väärtused ja nende esinemistõenäosused on järgmised:

$\mathbf{y}$	$P(\mathbf{y})$
{kull; kull}	p^2
{kull; kiri}	$p(1-p)$
{kiri; kull}	$(1-p)p$
{kiri; kiri}	$(1-p)^2$

Olgu p_1 ja p_2 kaks võimalikku parameetri p (kulli saamise tõenäosus) väärtust. Olgu mistahes vaatluste jada $\mathbf{y}$ korral antud vaatluste jada (vaatlusvektori) nägemise tõenäosused võrdsed. Siis ka vaatluste vektori {kull;

kull} nägemise tõenäosused mõlema parameetri väärtuse korral peavad olema võrdsed, $p_1^2 = p_2^2$. Kuna aga tõenäosused ei saa olla negatiivsed, siis järeldub sellest võrdusest, et $p_1 = p_2$ ja järelikult on parameeter p identifitseeritav — jaotusfunktsioonide samasusest saab järeldada parameetrite võrdust, $F(\mathbf{y}; p_1) = F(\mathbf{y}; p_2) \, \forall \mathbf{y} \Rightarrow p_1 = p_2$. Ehk kulli nägemise tõenäosuse p määramisel on kahe mündiviske tulemustest kasu (saadud hinnangu täpsus ei pruugi olla kuigivõrd hea, kuid mündivisete tulemused sisaldavad vähemalt mingitki informatsiooni kulli nägemise tõenäosuse kohta).

Kommentaar: parameetri/parameetervektori identifitseeritavus ei tähenda tingimata, et vastava parameetri väärtust oleks võimalik leida või hinnata ühe konkreetse valimi korral.

Mitte alati pole parameetrite vektor θ identifitseeritav, samas võib aga siiski olla võimalik vastata meid huvitavale küsimusele parameetrite vektori kohta (identifitseeritav on meid huvitav parameetrite vektori funktsioon).

Definitsioon 2.2 ***Parameetri*** *θ* ***funktsioon*** *$g(\theta)$ on* ***identifitseeritav,*** *kui kahe parameetri θ võimaliku väärtuse θ_1 ja θ_2 korral järeldub jaotusfunktsioonide identsusest funktsiooni väärtuste võrdsus:*

$$F(x; \theta_1) = F(x; \theta_2) \ \forall x \Rightarrow g(\theta_1) = g(\theta_2).$$

Selgitame parameetrite funktsiooni identifitseeritavust samuti näite abil. Vaatame andmestikku, kus on 2 mehe ja 2 naise andmed. Meestel tehtud mõõtmiste jaotuseks olgu:

$$y \sim N(\mu + \alpha_{mees}; \sigma^2)$$

ja naistel tehtud mõõtmiste jaotuseks olgu

$$y \sim N(\mu + \alpha_{naine}; \sigma^2)$$

ehk tegemist on tüüpilise dispersioonanalüüsi mudeliga.

Selgub, et parameetervektor $\boldsymbol{\beta} = (\mu; \alpha_{mees}; \alpha_{naine})^T$ pole identifitseeritav: kahe erineva parameetervektori väärtuse $\boldsymbol{\beta}_1 = (1; 2; 3)^T$ ja $\boldsymbol{\beta}_2 = (0; 3; 4)^T$ korral on vaatluste vektori $\mathbf{y}$ tihedusfunktsioon täpselt samasugune ja järelikult pole parameetrite vektor $\boldsymbol{\beta}$ identifitseeritav — sest mõlema parameetervektori korral oleks uuritava tunnuse jaotuseks meestel $N(3; \sigma^2)$ ja naistel $N(4; \sigma^2)$.

Küll aga osutub identifitseeritavaks parameetervektori $\boldsymbol{\beta}$ funktsioon

$$g(\boldsymbol{\beta}) = (0; 1; -1) \cdot \boldsymbol{\beta} = \alpha_{mees} - \alpha_{naine}$$

ehk meeste ja naiste keskväärtuste erinevus.

Olgu vaatlusvektor $\mathbf{y}$ on normaaljaotusega, $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}; \sigma^2\mathbf{I})$. Parameetrite $\boldsymbol{\beta}$ funktsioon $h(\boldsymbol{\beta})$ identifitseeritav siis ja ainult siis, kui tegemist on $\mathbf{X}\boldsymbol{\beta}$ funktsiooniga, st kui leidub funktsioon g, nii et $h(\boldsymbol{\beta}) = g(\mathbf{X}\boldsymbol{\beta})$.

Näitame esmalt, et kui $h(\boldsymbol{\beta})$ on kirja pandav kui $\mathbf{X}\boldsymbol{\beta}$ funktsioon, $h(\boldsymbol{\beta}) = g(\mathbf{X}\boldsymbol{\beta})$, siis on tegemist parameetrite $\boldsymbol{\beta}$ identifitseeritava funktsiooniga.

Võrdusest $h(\boldsymbol{\beta}) = g(\mathbf{X}\boldsymbol{\beta})$ järeldub, et kui $\mathbf{X}\boldsymbol{\beta}_1 = \mathbf{X}\boldsymbol{\beta}_2$ siis $h(\boldsymbol{\beta}_1) = h(\boldsymbol{\beta}_2)$, sest $h(\boldsymbol{\beta}_1) = g(\mathbf{X}\boldsymbol{\beta}_1) = g(\mathbf{X}\boldsymbol{\beta}_2) = h(\boldsymbol{\beta}_2)$.

Vaatame, kas sellisel juhul on $h(\boldsymbol{\beta})$ identifitseeritav?

Selles veendumiseks vaatame, kas valimi jaotusfunktsioonide võrdsusest saab järeldada, et $\mathbf{X}\boldsymbol{\beta}_1 = \mathbf{X}\boldsymbol{\beta}_2$.

$$
\begin{aligned}
f_{\mathbf{y}}(\boldsymbol{\beta}_1) &= f_{\mathbf{y}}(\boldsymbol{\beta}_2)\ \forall \mathbf{y} \\
|2\pi\sigma_1^2|^{-n/2} \exp\left(-\frac{1}{2\sigma_1^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_1)^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_1)\right) &= \\
&= |2\pi\sigma_2^2|^{-n/2} \exp\left(-\frac{1}{2\sigma_2^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_2)^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_2)\right) \forall \mathbf{y}
\end{aligned}
$$

Kuna antud võrdus peab kehtima kõigi vaatlusvektorite $\mathbf{y}$ korral siis peab võrdus keshtima ka valikute $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_1$ ja $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_2$ korral, kust saame võrdused:

$$
\begin{cases}
|2\pi\sigma_1^2|^{-n/2} \cdot 1 = |2\pi\sigma_2^2|^{-n/2} \exp\left(-\frac{1}{2\sigma_2^2}(\mathbf{X}\boldsymbol{\beta}_1-\mathbf{X}\boldsymbol{\beta}_2)^T(\mathbf{X}\boldsymbol{\beta}_1-\mathbf{X}\boldsymbol{\beta}_2)\right) \\
|2\pi\sigma_1^2|^{-n/2} \exp\left(-\frac{1}{2\sigma_1^2}(\mathbf{X}\boldsymbol{\beta}_1-\mathbf{X}\boldsymbol{\beta}_2)^T(\mathbf{X}\boldsymbol{\beta}_1-\mathbf{X}\boldsymbol{\beta}_2)\right) = |2\pi\sigma_2^2|^{-n/2} \cdot 1,
\end{cases}
$$

sest $(\mathbf{X}\boldsymbol{\beta}_2 - \mathbf{X}\boldsymbol{\beta}_1)^T(\mathbf{X}\boldsymbol{\beta}_2 - \mathbf{X}\boldsymbol{\beta}_1) = (-1)^2(\mathbf{X}\boldsymbol{\beta}_1 - \mathbf{X}\boldsymbol{\beta}_2)^T(\mathbf{X}\boldsymbol{\beta}_1 - \mathbf{X}\boldsymbol{\beta}_2)$. Asendades esimesest võrrandist saadud $|2\pi\sigma_1^2|^{-n/2}$ väärtuse teise võrrandisse saame

$$
\begin{aligned}
|2\pi\sigma_2^2|^{-n/2}\exp\left(\left(-\frac{1}{2\sigma_2^2}-\frac{1}{2\sigma_1^2}\right)(\mathbf{X}\boldsymbol{\beta}_1-\mathbf{X}\boldsymbol{\beta}_2)^T(\mathbf{X}\boldsymbol{\beta}_1-\mathbf{X}\boldsymbol{\beta}_2)\right) &= |2\pi\sigma_2^2|^{-n/2} \\
\exp\left(\left(-\frac{1}{2\sigma_2^2}-\frac{1}{2\sigma_1^2}\right)(\mathbf{X}\boldsymbol{\beta}_1-\mathbf{X}\boldsymbol{\beta}_2)^T(\mathbf{X}\boldsymbol{\beta}_1-\mathbf{X}\boldsymbol{\beta}_2)\right) &= 1 \\
\left(\left(-\frac{1}{2\sigma_2^2}-\frac{1}{2\sigma_1^2}\right)(\mathbf{X}\boldsymbol{\beta}_1-\mathbf{X}\boldsymbol{\beta}_2)^T(\mathbf{X}\boldsymbol{\beta}_1-\mathbf{X}\boldsymbol{\beta}_2)\right) &= 0 \\
(\mathbf{X}\boldsymbol{\beta}_1-\mathbf{X}\boldsymbol{\beta}_2)^T(\mathbf{X}\boldsymbol{\beta}_1-\mathbf{X}\boldsymbol{\beta}_2) &= 0 \\
\mathbf{X}\boldsymbol{\beta}_1-\mathbf{X}\boldsymbol{\beta}_2 &= 0 \\
\mathbf{X}\boldsymbol{\beta}_1 &= \mathbf{X}\boldsymbol{\beta}_2
\end{aligned}
$$

Seega oleme näidanud, et lineaarse mudeli $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}; \sigma^2\mathbf{I})$ korral on identifitseeritavad kõik parameetrite funktsioonid mis on kirja pandavad kui $\mathbf{X}\boldsymbol{\beta}$-funktsioonid.

Kui vaatleme funktsiooni $h(\boldsymbol{\beta})$ mis pole vaadeldav $\mathbf{X}\boldsymbol{\beta}$ funktsioonina, st. kui leiduvad $\boldsymbol{\beta}_1$ ja $\boldsymbol{\beta}_2$, nii et $\mathbf{X}\boldsymbol{\beta}_1 = \mathbf{X}\boldsymbol{\beta}_2$ aga $h(\boldsymbol{\beta}_1) \neq h(\boldsymbol{\beta}_2)$, siis pole vastav funktsioon ka identifitseeritav:

$$\mathbf{X}\boldsymbol{\beta}_1 = \mathbf{X}\boldsymbol{\beta}_2 \Rightarrow f(\mathbf{y}; \boldsymbol{\beta}_1) = f(\mathbf{y}; \boldsymbol{\beta}_2),$$

aga eelduse järgi $h(\boldsymbol{\beta}_1) \neq h(\boldsymbol{\beta}_2)$, seega $h(\boldsymbol{\beta})$ pole identifitseeritav.

Järeldus: parameetervektor $\boldsymbol{\beta}$ on identifitseeritav, kui $\mathbf{X}^T\mathbf{X}$ on pööratav, sest siis on parameetervektor $\boldsymbol{\beta}$ ise esitatav kui $\mathbf{X}\boldsymbol{\beta}$ funktsioon: $\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = g(\mathbf{X}\boldsymbol{\beta})$.

Identifitseeritavuse mõiste kasutamiseks peame teadma uuritava tunnuse $\mathbf{y}$ jaotust. Lineaarseid mudeleid soovime aga kasutada ka siis, kui me uuritava tunnuse jaotust ei tea. Seega oleks hea, kui oskaksime öelda, kas antud andmed võiksid sisaldada informatsiooni meid huvitavate parameetrite või näitajate (parameetrite funktsioonide) kohta ka siis, kui me ei tea millise jaotusega on vaatlusvektor $\mathbf{y}$. Selles aitab meid *hinnatavuse* mõiste.

Definitsioon 2.3 *Parameetrite lineaarne funktsioon* $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ *on* ***hinnatav****, kui leidub lineaarne nihketa hinnang meid huvitavale suurusele* $\boldsymbol{\lambda}^T\boldsymbol{\beta}$*:*
$\exists \mathbf{v}$*, nii et* $E(\mathbf{v}^T\mathbf{y}) = \lambda^T\boldsymbol{\beta}$ *(mistahes* $\boldsymbol{\beta}$ *väärtuse korral).*

On lihtne näha, et parameetrite lineaarne funktsioon on hinnatav normaaljaotuse eelduse kehtides siis ja ainult siis kui parameetrite lineaarne funktsioon on identifitseeritav. Kui parameetervektor $\lambda^T\boldsymbol{\beta}$ on hinnatav, siis

$$\begin{aligned} E(\mathbf{v}^T\mathbf{y}) &= \boldsymbol{\lambda}^T\boldsymbol{\beta}\ \forall\boldsymbol{\beta} \\ \mathbf{v}^T\mathbf{X}\boldsymbol{\beta} &= \boldsymbol{\lambda}^T\boldsymbol{\beta}\ \forall\boldsymbol{\beta} \\ \mathbf{v}^T\mathbf{X} &= \boldsymbol{\lambda}^T \end{aligned}$$

ja järelikult on $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ esitatav kui $\mathbf{X}\boldsymbol{\beta}$ (lineaarne) funktsioon (mis on identifitseeritav, kui vaatlusvektor on normaaljaotusega).

Kui aga $\mathbf{y}$ on normaaljaotusega ja parameetrite lineaarne funktsioon $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ on identifitseeritav, siis $\boldsymbol{\lambda}^T\boldsymbol{\beta} = \mathbf{v}^T\mathbf{X}\boldsymbol{\beta} = E(\mathbf{v}^T\mathbf{y})$ (sest $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ peab olema $\mathbf{X}\boldsymbol{\beta}$ funktsioon) ja seega on ta ka hinnatav.

Hinnatavuse (*estimability*) eeliseks on see, et tema kontrollimiseks pole tarvis teha eelduseid uuritava tunnuse jaotuse kohta. Puudus: hinnatavuse mõiste katab ainult parameetrite lineaarseid funktsioone (ja teda ei saa kasutada parameetrite mittelineaarsete funktsioonide korral).

Meenutuseks:

- $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ on hinnatav $\Leftrightarrow \exists\mathbf{v}$, nii et $\mathbf{v}^T\mathbf{X} = \boldsymbol{\lambda}^T$.

- Kui maatriks $\mathbf{X}^T\mathbf{X}$ on pööratav, siis on parameetervektor $\boldsymbol{\beta}$ ise hinnatav ning hinnatavad on ka kõik võimalikud parameetrite lineaarkombinatsioonid.

- Mitte kõik parameetrite identifitseeritavad funktsioonid (näiteks $\boldsymbol{\beta}_1/\boldsymbol{\beta}_2$) ei pruugi olla hinnatavad (sest $\boldsymbol{\beta}_1/\boldsymbol{\beta}_2$ pole parameetrite lineaarne funktsioon). Kõik hinnatavad funktsioonid on aga (vähemalt normaaljaotuse eelduse kehtides) identifitseeritavad.

- Kõik hinnatavad parameeterfunktsioonid on kirja pandavad kui suuruse $\mathbf{X}\boldsymbol{\beta}$ (=E$\mathbf{y}$) funktsioonid.

  Seega mistahes olemasolevate vaatluste lineaarkombinatsiooni keskväärtus on hinnatav. Näiteks kujuta endale ette lihtsat ühefaktorilist dispersioonanalüüsi, faktortunnuseks olgu sugu. Kui valimis on vähemalt 1 mees, on meeste keskväärtus hinnatav. Kui valimis on vähemalt 1 naine, on naiste keskväärtus hinnatav. Kui valimis on vähemalt 1 mees ja 1 naine, siis on meeste ja naiste keskväärtuste erinevus hinnatav.

Ülesanded

Oletame, et 5-l objektil on mõõdetud kolme tunnuse (Y, X_1, X_2 väärtused). Saadud andmeid kasutatakse hindamaks mitmest regressioonimudelit ($Y = c_0 + c_1X_1 + c_2X_2 + \varepsilon$). Kasutatav andmestik on järgmine:

Y	1	1	1	2	2
X_1	1	2	2	4	5
X_2	0	0	0	0	0

Tähistame $\boldsymbol{\beta}^T := (c_0, c_1, c_2)$. Kas parameeterfunktsioon $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ on hinnatav, kui:
a) $\boldsymbol{\lambda}^T = (1,1,0)$
b) $\boldsymbol{\lambda}^T = (1,1,1)$?
Põhjenda oma otsust!

Vali ülaltoodutest välja hinnatav parameeterfunktsioon ja leia käsitsi (näita arvutusi!) $\widehat{\boldsymbol{\lambda}^T\boldsymbol{\beta}}$

2.6 Parim lineaarne nihketa hinnang - Best Linear Unbiased Estimator (BLUE)

Soovime hinnata hinnatavat parameeterfunktsiooni $\boldsymbol{\lambda}^T\boldsymbol{\beta}$. Soovime, et hinnang oleks nihketa hinnangute seas väiksema võimaliku varieeruvusega (veaga). Millist hinnangut peaksime kasutama?

Definitsioon 2.4 *Hinnangut $\widehat{\boldsymbol{\lambda}^T\boldsymbol{\beta}}$ kutsutakse parimaks lineaarseks nihketa hinnanguks (BLUE), kui ta rahuldab järgmiseid nõudeid:*

1. *Lineaarsus*
$$\widehat{\boldsymbol{\lambda}^T\boldsymbol{\beta}} = \mathbf{a}^T\mathbf{y}$$
2. *Nihketus*
$$\forall\boldsymbol{\beta}\ \ E(\widehat{\boldsymbol{\lambda}^T\boldsymbol{\beta}}) = \boldsymbol{\lambda}^T\boldsymbol{\beta}$$
3. *Minimaalne dispersioon (minimaalne keskmine ruutviga):*
$$D(\widehat{\boldsymbol{\lambda}^T\boldsymbol{\beta}}) \leq D(\widetilde{\boldsymbol{\lambda}^T\boldsymbol{\beta}})$$
$\forall\widetilde{\boldsymbol{\lambda}^T\boldsymbol{\beta}}$ *korral mis rahuldab nõudeid (1) ja (2). Teisisõnu:*
$$D(\mathbf{a}^T\mathbf{y}) \leq D(\mathbf{b}^T\mathbf{y})$$
$\forall\mathbf{b}$ *korral, mille puhul* $E\mathbf{b}^T\mathbf{y} = \boldsymbol{\lambda}^T\boldsymbol{\beta}$ $(\forall\boldsymbol{\beta})$.

Märkus:
Nõude (3) võib kirja panna ka kujul

$$\mathrm{E}(\mathbf{a}^T\mathbf{y} - \boldsymbol{\lambda}^T\boldsymbol{\beta})^2 \leq \mathrm{E}(\mathbf{b}^T\mathbf{y} - \boldsymbol{\lambda}^T\boldsymbol{\beta})^2.$$

Seega otsime lineaarset nihketa hinnangut mille korral hinnangu oodatav ruutviga oleks minimaalne.

Teoreem 2.6 ***Gauss-Markovi teoreem***
Vaatame mudelit
$$E\mathbf{y} = \mathbf{X}\boldsymbol{\beta},\ \ D(\mathbf{y}) = \sigma^2\mathbf{I}.$$
Kui $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ on hinnatav, siis
$$\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}} = \boldsymbol{\lambda}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y} \tag{2.17}$$
on parim lineaarne nihketa hinnang $\boldsymbol{\lambda}^T\boldsymbol{\beta}$-le.

Tõestus
Esmalt paneme tähele, et nihketuse nõudest (2) järeldub:

$$\begin{aligned} \forall\boldsymbol{\beta}\ \mathrm{E}(\widehat{\boldsymbol{\lambda}^T\boldsymbol{\beta}}) &= \boldsymbol{\lambda}^T\boldsymbol{\beta} \\ \forall\boldsymbol{\beta}\ \mathrm{E}(\mathbf{a}^T\mathbf{y}) &= \boldsymbol{\lambda}^T\boldsymbol{\beta} \\ \forall\boldsymbol{\beta}\ \mathbf{a}^T\mathbf{X}\beta &= \boldsymbol{\lambda}^T\boldsymbol{\beta} \\ \mathbf{a}^T\mathbf{X} &= \boldsymbol{\lambda}^T \end{aligned}$$

Kuna sama nõue peab kehtima ka mingi teise lineaarse hinnangu $\mathbf{b}^T\mathbf{y}$ jaoks, saame:

$$(\mathbf{a}^T - \mathbf{b}^T)\mathbf{X} = 0.$$

Vaatame selle teise lineaarse nihketa hinnangu dispersiooni:

$$\begin{aligned} \mathrm{D}(\mathbf{b}^T\mathbf{y}) &= \mathrm{D}(\mathbf{b}^T\mathbf{y} - \mathbf{a}^T\mathbf{y} + \mathbf{a}^T\mathbf{y}) \\ &= \mathrm{D}(\mathbf{b}^T\mathbf{y} - \mathbf{a}^T\mathbf{y}) + \mathrm{D}(\mathbf{a}^T\mathbf{y}) + 2\mathrm{cov}(\mathbf{b}^T\mathbf{y} - \mathbf{a}^T\mathbf{y}, \mathbf{a}^T\mathbf{y}) \\ &= \mathrm{D}(\mathbf{b}^T\mathbf{y} - \mathbf{a}^T\mathbf{y}) + \mathrm{D}(\mathbf{a}^T\mathbf{y}) + 2(\mathbf{b}^T - \mathbf{a}^T)\mathrm{cov}(\mathbf{y}, \mathbf{y})\mathbf{a} \end{aligned}$$

Juhul, kui $\mathbf{a}^T\mathbf{y} = \boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}} = \mathbf{v}^T\mathbf{P}_\mathbf{X}\mathbf{y} = \mathbf{v}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y}$, siis $\sigma^2(\mathbf{b}^T - \mathbf{a}^T)\mathbf{a} = 0$ (sest $(\mathbf{b}^T - \mathbf{a}^T)\mathbf{X} = 0$) ja saame tulemuseks:

$$\mathrm{D}(\mathbf{b}^T\mathbf{y}) = \mathrm{D}(\mathbf{b}^T\mathbf{y} - \mathbf{a}^T\mathbf{y}) + \mathrm{D}(\mathbf{a}^T\mathbf{y}) \quad (\geq \mathrm{D}(\mathbf{a}^T\mathbf{y})).$$

□

Muuseas, saadud BLUE-hinnang on üheselt määratud:

$$\begin{aligned} \mathrm{D}(\mathbf{b}^T\mathbf{y} - \mathbf{a}^T\mathbf{y}) &= 0 \\ \sigma^2(\mathbf{b} - \mathbf{a})^T(\mathbf{b} - \mathbf{a}) &= 0 \\ \mathbf{b} &= \mathbf{a}. \end{aligned}$$

2.7 Parim nihketa hinnang (BUE)

H. Cramer ja C.R.Rao näitasid, et mistahes nihketa hinnangu dispersioon ei saa olla väiksem teatavast piirist (mida tuntakse kui Cramer-Rao alampiiri).

Ühe parameetri korral ütleb Cramer-Rao alampiir, et parameetri θ mistahes nihketa hinnangu $\hat{\theta}$ dispersioon ei saa olla suurem kui pöördväärtus Fisheri informatsioonist:

$$\mathrm{D}(\hat{\theta}) \geq \mathcal{I}(\theta)^{-1}.$$

Sama tulemuse üldistus mitmemõõtmelisele juhule (ja sobib kasutamiseks ka nihkega hinnangute korral) on kirja pandav järgneval kujul.

Kui parameetervektori $\boldsymbol{\theta}$ hinnangu $\hat{\boldsymbol{\theta}}$ keskväärtus on kirja pandav kui $\boldsymbol{\theta}$ mingi funktsioon, $\mathrm{E}\hat{\boldsymbol{\theta}} = \mathbf{g}(\boldsymbol{\theta})$, siis

$$\mathrm{D}(\hat{\boldsymbol{\theta}}) \geq_L \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}} [\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})]^{-1} \left[\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}} \right]^T,$$

Kus $\geq_L$ tähistab Löwneri osalist järjestust (kui $\mathbf{A} \geq_L \mathbf{B}$, siis $\mathbf{A} - \mathbf{B}$ on mittenegatiivselt määratud maatriks, st. $\mathbf{x}^T(\mathbf{A} - \mathbf{B})\mathbf{x} \geq 0 \ \forall \mathbf{x}$).

Kui tegemist on nihketa hinnanguga, $\mathbf{g}(\boldsymbol{\theta}) = \boldsymbol{\theta}$, siis $\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}} = \mathbf{I}$ ja

$$\mathrm{D}(\hat{\boldsymbol{\theta}}) \geq_L [\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})]^{-1},$$

kus

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) = \mathrm{E}\left\{ \left(\frac{\partial l}{\partial \boldsymbol{\theta}} \right)^T \frac{\partial l}{\partial \boldsymbol{\theta}} \right\}.$$

Normaaljaotuse eeldusel, lineaarsete mudelite kontekstis (vaata valemit 2.14):

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{X} / \sigma^2$$

ja seega on Fisheri informatsioon (juhul kui parameetervektor $\boldsymbol{\beta}$ on hinnatav) kirja pandav kui

$$\begin{aligned}
\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) &= \frac{1}{\sigma^2}\frac{1}{\sigma^2}\mathrm{E}\left\{[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{X}]^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{X}\right\} \\
&= \frac{1}{\sigma^4}\mathbf{X}^T\mathrm{E}\left[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\right]\mathbf{X} \\
&= \frac{1}{\sigma^4}\mathbf{X}^T\mathbf{I}\sigma^2\mathbf{X} \\
&= \mathbf{X}^T\mathbf{X}/\sigma^2
\end{aligned}$$

Seega ei saa parameetervektori $\boldsymbol{\beta}$ ükski nihketa hinnang olla väiksema dispersiooniga kui $\boldsymbol{\mathcal{I}}^{-1} = (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2$. Kuna aga lineaarse mudeli korral

$$\begin{aligned}
\mathrm{D}(\hat{\boldsymbol{\beta}}) =& \mathrm{D}\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}\right) \\
=& (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathrm{D}\left(\mathbf{y}\right)[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T]^T \\
=& (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\sigma^2\mathbf{I}X(\mathbf{X}^T\mathbf{X})^{-1} \\
=& (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2,
\end{aligned}$$

siis järelikult on meie poolt kasutatav hinnang parimaks nihketa hinnanguks parameetervektorile $\boldsymbol{\beta}$.

Kokkuvõtteks: Kui andmed on normaaljaotusega, $\boldsymbol{\varepsilon} \sim N$, siis on vähimruutude/suurima tõepära hinnang $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ parimaks nihketa hinnanguks.

Kui andmed pole normaaljaotusega, on sama hinnang parimaks lineaarseks nihketa hinnanguks (st. sellisel juhul võib olla võimalik leida hinnangut, mis on parem kui meie poolt pakutud hinnang, kuid seda hinnangut ei tasu otsida lineaarsete hinnangute klassist).