6.3 Confidence interval for the model based mean, prediction interval

Normal linear model $\mathbb{Y} | \mathbb{X} \sim \mathcal{N}_n(\mathbb{X}\beta, \sigma^2 I_n)$

$$\Rightarrow \forall i \quad Y_i | X_i \sim \mathcal{N}(X_i^T\beta, \sigma^2)$$

$$\varepsilon_i = Y_i - X_i^T\beta \overset{iid}{\sim} \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$\mathcal{X} \subseteq \mathbb{R}^k$: sample space of $X_1, \dots, X_n$

Take $x_{new} \in \mathcal{X}$, $\varepsilon_{new} \sim \mathcal{N}(0, \sigma^2)$

$$\varepsilon_{new} \perp\!\!\!\perp \boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$$

$$Y_{new} := x_{new}^T\beta + \varepsilon_{new}$$

→ "new" observation being sampled / obtained) independently of the "old" observations $\mathbb{Y} = (Y_1, \dots, Y_n)^T$

$$Y_{new} | X_{new} = x_{new} \sim \mathcal{N}(x_{new}^T\beta, \sigma^2)$$

Problems: Interval estimation of

(i) $\mu_{new} = \mathbb{E}(Y_{new} | X_{new} = x_{new}) \overset{\text{being assumed}}{=} x_{new}^T\beta$

(ii) Y_{new} itself (prediction)

Theorem 6.3 Confidence interval for the model based mean, prediction interval

16

Let $Y|X \sim N_n(X\beta, \sigma^2 I_n)$, rank $(X_{n\times k}) = k$, $\hat{\beta} = (X^TX)^{-1}X^TY$ is the LSE of the regression parameters β. Let $x_{new} \in \mathcal{X}$, $x_{new} \neq 0$. Let $\varepsilon_{new} \sim N(0, \sigma^2)$ is independent of $\varepsilon = Y - X\beta$. Finally, let $Y_{new} = x_{new}^T \beta + \varepsilon_{new}$. The following then holds:

(i) The quantity $\hat{\mu}_{new} := x_{new}^T \hat{\beta}$ is the BLUE of $\mu_{new} = x_{new}^T \beta$. The standard error of $\hat{\mu}_{new}$ is $S.E.(\hat{\mu}_{new}) = \sqrt{MS_e\, x_{new}^T (X^TX)^{-1} x_{new}}$ and the lower and the upper bound of the $(1-\alpha)100\%$ confidence interval for μ_{new} are

$$(\hat{\mu}^L_{new}, \hat{\mu}^U_{new}) = \hat{\mu}_{new} \pm S.E.(\hat{\mu}_{new})\, t_{n-k}\left(1-\frac{\alpha}{2}\right).$$

PROOF: ~~Point (i) of~~

- BLUE property: Theorem 2.5 (Gauss-Markov for linear comb.)
- Point (viii), Theorem 6.2

$$\forall \beta^0 \quad P\left(\left(\hat{\mu}_{new} \pm \sqrt{MS_e\, x_{new}^T (X^TX)^{-1} x_{new}}\; t_{n-k}\left(1-\frac{\alpha}{2}\right)\right) \ni x_{new}^T \beta^0 \;;\; \beta = \beta^0\right) = 1-\alpha$$

MM

(iii) a (random) interval with the bounds

$$(Y_{new}^L, Y_{new}^U) = \hat{\mu}_{new} \pm S.E.P.(X_{new})\, t_{n-k}(1-\tfrac{\alpha}{2}),$$

where $S.E.P.(X_{new}) = \sqrt{MS_e(1 + X_{new}^T (X^T X)^{-1} X_{new})}$,

covers with the probability of $(1-\alpha)$ the value of Y_{new}.

That is $\forall \beta^0 \in \mathbb{R}^k\ \forall \sigma_0^2 > 0$

$$P((Y_{new}^L, Y_{new}^U) \ni Y_{new}\ ;\ \beta = \beta^0, \sigma^2 = \sigma_0^2) = 1-\alpha.$$

Quick but <u>INCORRECT</u> "proof":

$$Y_{new} = X_{new}^T \beta + \varepsilon_{new}$$

$$\hat{Y}_{new} := X_{new}^T \hat{\beta} + \underbrace{\hat{\varepsilon}_{new}}_{:=0}$$

indep. $:= 0$ since $\mathbb{E}(\varepsilon_{new} / X_{new} = x_{new}) = 0$

$$var(\hat{Y}_{new} / \mathbb{X}) = var(X_{new}^T \hat{\beta} / \mathbb{X}) + var(\hat{\varepsilon}_{new} / \mathbb{X}) =$$

$$= X_{new}^T \underbrace{var(\hat{\beta} / \mathbb{X})}_{\sigma^2 (X^T X)^{-1}} X_{new} + \sigma^2$$

(? since $var(\varepsilon_{new} / X) = \sigma^2$)

$$= \sigma^2 (1 + X_{new}^T (X^T X)^{-1} X_{new})$$

$\longrightarrow S.E.P.(X_{new})$: replace unknown σ^2 by MS_e

$\Rightarrow$ „prediction": $\hat{Y}_{new} \pm S.E.P.(X_{new}) \cdot$ quantile

Proof: All statements given $X_1, \dots, X_n, X_{new}$:

$$Y_{new} \sim \mathcal{N}(\mu_{new}, \sigma^2)$$

$$\hat{\mu}_{new} \sim \mathcal{N}(\mu_{new}, \sigma^2 v), \quad v = \underbrace{x_{new}^T (X^T X)^{-1} x_{new}}_{>0 \text{ (WHY?)}}$$

$$\parallel$$

LSE of μ_{new} based on $(\mathbb{Y}, \mathbb{X})$

$$\mathbb{Y} \perp\!\!\!\perp Y_{new} \implies Y_{new} \perp\!\!\!\perp \hat{\mu}_{new}$$

$$\implies Y_{new} - \hat{\mu}_{new} \sim \mathcal{N}(0, \sigma^2(1+v))$$

Further, SSe is a function of $\mathbb{Y}$ (and $\mathbb{X}$)

$$\mathbb{Y} \perp\!\!\!\perp Y_{new} \implies Y_{new} \perp\!\!\!\perp SSe$$

We also know (theorem 6.2(v)): $\hat{\mu}_{new} \perp\!\!\!\perp SSe$

$$\implies Y_{new} - \hat{\mu}_{new} \perp\!\!\!\perp SSe$$

Finally, $\frac{SSe}{\sigma^2} \sim \chi^2_{n-k}$

Hence

$$\underbrace{\frac{\dfrac{Y_{new} - \hat{\mu}_{new}}{\sqrt{\sigma^2(1+v)}} \sim \mathcal{N}(0,1)}{\sqrt{\dfrac{SSe}{(n-k)\sigma^2}} \sim \chi^2_{n-k}}}_{\displaystyle \frac{Y_{new} - \hat{\mu}_{new}}{\sqrt{MSe(1+v)}}} \perp\!\!\!\perp \sim t_{n-k}$$

$$\implies \forall \beta^0 \in \mathbb{R}^k \ \forall \sigma_0^2 > 0, \quad Y_{new}^0 = x_{new}^T \beta^0 + \varepsilon_{new}^0, \ \varepsilon_{new}^0 \sim \mathcal{N}(0, \sigma_0^2)$$

$$1 - \alpha = P\left(\left| \frac{Y_{new}^0 - \hat{\mu}_{new}}{\sqrt{MSe(1+v)}} \right| < t_{n-k}\left(1 - \tfrac{\alpha}{2}\right); \beta = \beta^0, \sigma^2 = \sigma_0^2 \right)$$

$$= P\left(\left(\hat{\mu}_{new} \pm \underbrace{\sqrt{MSe(1+v)}}_{S.E.P.(x_{new})} \, t_{n-k}\left(1 - \tfrac{\alpha}{2}\right) \right) \ni Y_{new}^0; \beta = \beta^0, \sigma^2 = \sigma_0^2 \right)$$

Confidence interval for the model based mean: 17

$$CIM(x_{new}) := \hat{\mu}_{new} \pm \sqrt{MSe \cdot v}\; t_{n-k}\left(1-\frac{\alpha}{2}\right), \quad v = x_{new}^T (X^T X)^{-1} x_{new}$$

Prediction interval:

$$PI(x_{new}) := \hat{\mu}_{new} \pm \underbrace{\sqrt{MSe(1+v)}}_{S.E.P.(x_{new})}\; t_{n-k}\left(1-\frac{\alpha}{2}\right)$$

Sequence of CIM's and PI's for a grid of values $x_{new} \in \mathcal{X}$ 18

→ confidence band AROUND the regression function

→ prediction band

Illustration: Kojeni

Hosi∅ 19

- width of the band

CIM

$\underbrace{MSe \cdot v}_{\text{precis. of LSE}}$

PI

$\underbrace{MSe \cdot v}_{\text{precis. of LSE}} + \underbrace{MSe}_{\equiv \widehat{var}(Y|X)}$

→ what is going to happen if $n \to \infty$?

Importance of normality assumption for a valid prediction (also homosced.)