MULTIPLE LM: ESTIMATION

parameters to estimate : $\left.\begin{matrix}\beta_1,\dots,\beta_p\\ \sigma^2\end{matrix}\right\} \Rightarrow \Theta = \mathbb{R}^p \times \mathbb{R}^+$ parameter space

Similarly to the case of a simple linear model, the ML estimators are the same that we obtain using the OLS (minimize the sum of squares of the residuals).

$Y_i \sim N(\mu_i, \sigma^2)$ independent for $i=1,\dots,n$

$\mu_i = X\underline{\beta} = \beta_1 x_{i1} + \dots + \beta_p x_{ip}$

likelihood:

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i-\mu_i)^2\right\}$$

$$= (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu_i)^2\right\}$$

$$\ell(\theta) = \cancel{-\frac{n}{2}\log 2\pi} - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu_i)^2$$

as before, $\mu_i = \underset{\sim}{x}_i^T \underline{\beta}$

$$\sum_i (y_i-\mu_i)^2 = \sum_{i=1}^{n}(y_i - \underset{\sim}{x}_i^T\underline{\beta})^2 = (\underline{y}-X\underline{\beta})^T(\underline{y}-X\underline{\beta}) = \underbrace{S(\underline{\beta})}_{\text{sum of squares}}$$

for fixed $\sigma^2$, maximizing the likelihood is equivalent to minimizing $S(\underline{\beta})$, independently of the value of $\sigma^2$

$\Rightarrow \hat{\underline{\beta}} = \arg\min_{\underline{\beta}} S(\underline{\beta})$

notice that $S(\underline{\beta}) = (\underline{y}-X\underline{\beta})^T(\underline{y}-X\underline{\beta}) = \underline{y}^T\underline{y} - \overbrace{\underline{y}^T X\underline{\beta}}^{1\times1} - \overbrace{\underline{\beta}^T X^T\underline{y}}^{1\times1} + \underline{\beta}^T X^T X\underline{\beta} = \underline{y}^T\underline{y} - 2\underline{y}^T X\underline{\beta} + \underline{\beta}^T X^T X\underline{\beta}$

Recall:
- $\frac{\partial}{\partial\underline{\beta}} \underset{1\times p}{\underline{a}^T} \underset{p\times1}{\underline{\beta}} = \underset{p\times1}{\underline{a}}$
- $\frac{\partial}{\partial\underline{\beta}} \underset{1\times p}{\underline{\beta}^T} \underset{p\times p}{A} \underset{p\times1}{\underline{\beta}} = \underset{p\times1}{\underset{(p\times p)(p\times1)}{2A\underline{\beta}}}$

to find $\hat{\underline{\beta}}$ we need to solve $\frac{\partial}{\partial\underline{\beta}} S(\underline{\beta}) = 0$

$\frac{\partial}{\partial\underline{\beta}} S(\underline{\beta}) = \frac{\partial}{\partial\underline{\beta}}(\underline{y}^T\underline{y} - 2\underline{y}^T X\underline{\beta} + \underline{\beta}^T X^T X\underline{\beta}) = 0 - 2X^T\underline{y} + 2X^TX\underline{\beta}$

$-2X^T(\underline{y}-X\underline{\beta}) = 0 \rightarrow \begin{cases} \underline{x}_1^T(\underline{y}-X\underline{\beta}) = 0 \\ \vdots \\ \underline{x}_p^T(\underline{y}-X\underline{\beta}) = 0 \end{cases}$ "normal equations"

$\Rightarrow \frac{\partial}{\partial\underline{\beta}} S(\underline{\beta}) = 0 \Rightarrow X^TX\underline{\beta} = X^T\underline{y}$ To solve the equation I need $X^TX$ to be nonsingular (invertible) $\rightarrow$ ok since we required $X$ with full rank

$\Rightarrow \hat{\underline{\beta}} = (X^TX)^{-1}X^T\underline{y}$ critical point $\rightarrow$ is it a minimum?

Hessian: $\frac{\partial^2}{\partial\underline{\beta}\partial\underline{\beta}^T} S(\underline{\beta}) = 2X^TX \Big|_{\underline{\beta}=\hat{\underline{\beta}}} = 2X^TX$ has to be positive definite

Recall: $\Sigma$ is positive definite if $\forall \underline{a} \neq \underline{0}$, $\underline{a}^T\Sigma\underline{a} > 0$

does it hold for $X^TX$?

$\underline{a}^TX^TX\underline{a} = (X\underline{a})^T(X\underline{a}) \geq 0$ and it is $=0 \iff X\underline{a} = 0$

since we required $X$ to have full rank $\Rightarrow X\underline{a} = 0 \iff \underline{a} = \underline{0}$

$\Rightarrow \underline{a}^TX^TX\underline{a} > 0 \Rightarrow 2X^TX$ is positive definite

$\Rightarrow \hat{\underline{\beta}} = (X^TX)^{-1}X^T\underline{y}$ is the minimum of $S(\underline{\beta})$ and the ML estimate

ESTIMATE of $\sigma^2$

$\ell(\theta) = \ell(\underline{\beta}, \sigma^2) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(\underline{y}-X\underline{\beta})^T(\underline{y}-X\underline{\beta})$

$\ell(\hat{\underline{\beta}}, \sigma^2) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(\underline{y}-X\hat{\underline{\beta}})^T(\underline{y}-X\hat{\underline{\beta}})$

$\frac{\partial}{\partial\sigma^2}\ell(\hat{\underline{\beta}}, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(\underline{y}-X\hat{\underline{\beta}})^T(\underline{y}-X\hat{\underline{\beta}}) = 0 \Rightarrow \hat{\sigma}^2 = \frac{(\underline{y}-X\hat{\underline{\beta}})^T(\underline{y}-X\hat{\underline{\beta}})}{n} = \frac{\underline{e}^T\underline{e}}{n}$

the estimator (r.v.) $\hat{\sigma}^2(\underline{Y}) = \frac{(\underline{Y}-X\hat{\underline{\beta}})^T(\underline{Y}-X\hat{\underline{\beta}})}{n} = \frac{\underline{E}^T\underline{E}}{n}$

is biased: $\mathbb{E}[\hat{\sigma}^2(\underline{Y})] = \frac{n-p}{n}\sigma^2$ $\rightarrow$ the unbiased estimator is $\hat{S}^2 = \frac{(\underline{Y}-X\hat{\underline{\beta}})^T(\underline{Y}-X\hat{\underline{\beta}})}{n-p} = \frac{n}{n-p}\hat{\sigma}^2(\underline{Y})$

(n − # columns of X)

## Remarks

- the normal equations are $\begin{cases} (\underline{y}-X\hat{\underline{\beta}})^T\underline{x}_1 = 0 & \rightarrow \underline{e}^T\underline{x}_1 = 0 \\ \vdots \\ (\underline{y}-X\hat{\underline{\beta}})^T\underline{x}_p = 0 & \rightarrow \underline{e}^T\underline{x}_p = 0 \end{cases} \Rightarrow \underline{e}^TX = 0$

- if we include the intercept $\underline{x}_1 = \underline{1}$

  $\underline{e}^T\underline{x}_1 = 0 \Rightarrow \underline{e}^T\underline{1} = 0 \Rightarrow \sum_{i=1}^{n} e_i = 0 \Rightarrow \bar{e} = 0$