The Hat Matrix

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

$$\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \underbrace{\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top}_{\boldsymbol{H}} \boldsymbol{y} = \boldsymbol{H}\boldsymbol{y}$$

$$\widehat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \widehat{\boldsymbol{y}} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y}$$

$\boldsymbol{H}$: projects onto $\text{span}(\boldsymbol{X})$

$\boldsymbol{I} - \boldsymbol{H}$: projects onto orthogonal complement $[\text{span}(\boldsymbol{X})]^\perp$

Properties

(a) symmetric: $\boldsymbol{H}^\top = \boldsymbol{H}, (\boldsymbol{I}-\boldsymbol{H})^\top = \boldsymbol{I}-\boldsymbol{H}$

(b) idempotent: $\boldsymbol{H}\boldsymbol{H} = \boldsymbol{H}$, $(\boldsymbol{I}-\boldsymbol{H})(\boldsymbol{I}-\boldsymbol{H}) = \boldsymbol{I}-\boldsymbol{H}$

(c) positive semi-definite: $\boldsymbol{z}^\top \boldsymbol{H}\boldsymbol{z} \geq 0, \boldsymbol{z}^\top(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{z} \geq 0 \ \forall \ \boldsymbol{z} \in \mathbb{R}^n$

(d) eigenvalues:

$$\boldsymbol{H} : \{\underbrace{1, ..., 1}_{p}, \underbrace{0, ..., 0}_{n-p}\}; \qquad \boldsymbol{I}-\boldsymbol{H} : \{\underbrace{1, ..., 1}_{n-p}, \underbrace{0, ..., 0}_{p}\}$$

(e) trace: $\mathrm{tr}(\boldsymbol{H}) = p, \mathrm{tr}(\boldsymbol{I}-\boldsymbol{H}) = n-p$

<u>Proof</u>: $\mathrm{tr}[\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top] = \mathrm{tr}[\underbrace{(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}}_{\boldsymbol{I}_{p\times p}}]$.

Exercise Prove (a)-(c).

Gauss-Markov Theorem

Proof

$$\underbrace{\mathbb{E}(\boldsymbol{Ay}) = \boldsymbol{\beta} \quad \Rightarrow \quad \boldsymbol{AX\beta} = \boldsymbol{\beta}}_{\text{for } \underline{\text{any}}\ \boldsymbol{\beta}} \quad \Rightarrow \quad \boldsymbol{AX} = \boldsymbol{I}$$

$$\begin{aligned}
\mathbb{V}\mathrm{ar}(\boldsymbol{\ell}^\top \widetilde{\boldsymbol{\beta}}) - \mathbb{V}\mathrm{ar}(\boldsymbol{\ell}^\top \widehat{\boldsymbol{\beta}}) &= \sigma^2 \boldsymbol{\ell}^\top [\boldsymbol{AA}^\top - (\boldsymbol{X}^\top \boldsymbol{X})^{-1}]\boldsymbol{\ell} \\
&= \sigma^2 \boldsymbol{\ell}^\top [\boldsymbol{AA}^\top - \boldsymbol{A} \underbrace{\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top}_{\boldsymbol{H}} \boldsymbol{A}^\top]\boldsymbol{\ell} \\
&= \sigma^2 \boldsymbol{\ell}^\top \boldsymbol{A}(\boldsymbol{I} - \boldsymbol{H}) \underbrace{\boldsymbol{A}^\top \boldsymbol{\ell}}_{\boldsymbol{z}} = \sigma^2 \boldsymbol{z}^\top (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{z} \geq 0
\end{aligned}$$

Unbiased Estimation of σ^2

$$\boxed{\widehat{\sigma}^2 = \frac{1}{n-p}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}$$

Step (1) Can write

$$\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 = \|\boldsymbol{y} - \widehat{\boldsymbol{y}}\|^2 = \boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y},$$

and show (below) it can be further reduced to $\boldsymbol{\varepsilon}^\top(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{\varepsilon}$.

Exercise Show that $\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y} = \boldsymbol{\varepsilon}^\top(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{\varepsilon}$.

Unbiased Estimation of σ^2

Step (2) For any squared matrix $\boldsymbol{A}$,

$$
\begin{aligned}
\mathbb{E}(\boldsymbol{\varepsilon}^\top \boldsymbol{A}\boldsymbol{\varepsilon}) &= \mathbb{E}\Big[\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\varepsilon_i\varepsilon_j\Big] \\
&\overset{(\dagger)}{=} \sum_{i\neq j} a_{ij}\,\underbrace{\mathbb{E}(\varepsilon_i)}_{0}\,\underbrace{\mathbb{E}(\varepsilon_j)}_{0} + \sum_{i=1}^{n} a_{ii}\,\underbrace{\mathbb{E}(\varepsilon_i^2)}_{\sigma^2} \\
&= \sigma^2 \mathrm{tr}(\boldsymbol{A}),
\end{aligned}
$$

where $(\dagger)$ is due to $\varepsilon_i \perp\!\!\!\perp \varepsilon_j$ for $i \neq j$. The result then follows from the fact that $\mathrm{tr}(\boldsymbol{I} - \boldsymbol{H}) = n - p$.

Review of One-Sample T-Test

Given $X_1, ..., X_n \overset{iid}{\sim} \mathrm{N}(\mu, \sigma^2)$, want to test $H_0 : \mu =$ (something).

target	μ
estimator	$\bar{X}$
variance (of estimator)	σ^2/n

Under H_0,

$$\frac{\bar{X} - (\text{something})}{\sqrt{\sigma^2/n}} \sim \mathrm{N}(0,1) \quad \Rightarrow \quad \frac{\bar{X} - (\text{something})}{\sqrt{S^2/n}} \sim t_{(n-1)}$$

where

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

is unbiased estimator of σ^2. [*<u>Note</u>: Requires normality of all* X_i.]

T-Test for Linear Regression

For any $\boldsymbol{c} \in \mathbb{R}^p$, want to test $H_0 : \boldsymbol{c}^\top \boldsymbol{\beta} = \text{(something)}$.

target	$\boldsymbol{c}^\top \boldsymbol{\beta}$
estimator	$\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}}$
variance (of estimator)	$\boldsymbol{c}^\top \mathbb{V}\text{ar}(\widehat{\boldsymbol{\beta}})\boldsymbol{c} = \sigma^2[\boldsymbol{c}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{c}]$

Under H_0,

$$\frac{\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}} - \text{(something)}}{\sqrt{\sigma^2[\boldsymbol{c}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{c}]}} \sim \text{N}(0,1) \;\Rightarrow\; \frac{\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}} - \text{(something)}}{\sqrt{\widehat{\sigma}^2[\boldsymbol{c}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{c}]}} \sim t_{(n-p)},$$

where

$$\widehat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2$$

is unbiased estimator of σ^2. [<u>*Note*</u>*: Requires normality of all* y_i.]

Meaning of $\boldsymbol{c}^\top \boldsymbol{\beta}$

Merely a general notation. For example, can test

$$
\begin{aligned}
&H_0 : \beta_1 = 0 && \text{by letting} \quad \boldsymbol{c} = (1, 0, ..., 0)^\top, \\
&H_0 : \beta_1 - \beta_2 = 0 && \text{by letting} \quad \boldsymbol{c} = (1, -1, 0, ..., 0)^\top,
\end{aligned}
$$

and so on.

Exercise For $\widehat{\boldsymbol{\beta}} \in \mathbb{R}^2$, verify

$$
\begin{bmatrix} 1 & -1 \end{bmatrix}
\begin{bmatrix} \mathbb{V}\mathrm{ar}(\widehat{\beta}_1) & \mathbb{C}\mathrm{ov}(\widehat{\beta}_1, \widehat{\beta}_2) \\ \mathbb{C}\mathrm{ov}(\widehat{\beta}_2, \widehat{\beta}_1) & \mathbb{V}\mathrm{ar}(\widehat{\beta}_2) \end{bmatrix}
\begin{bmatrix} 1 \\ -1 \end{bmatrix}
= \mathbb{V}\mathrm{ar}(\widehat{\beta}_1 - \widehat{\beta}_2).
$$

Interpretation of Regression Estimates

Exercise Isolate variable j by re-arranging $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_0 & \boldsymbol{x}_j \end{bmatrix}$. Compute from $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$ that

$$\widehat{\beta}_j = [\boldsymbol{x}_j^\top (\boldsymbol{I} - \boldsymbol{H}_0) \boldsymbol{x}_j]^{-1} [\boldsymbol{x}_j^\top (\boldsymbol{I} - \boldsymbol{H}_0) \boldsymbol{y}]$$

where $\boldsymbol{H}_0 = \boldsymbol{X}_0 (\boldsymbol{X}_0^\top \boldsymbol{X}_0)^{-1} \boldsymbol{X}_0^\top$.

Interpretation That is,

$$\widehat{\beta}_j = (\boldsymbol{u}_j^\top \boldsymbol{u}_j)^{-1} \boldsymbol{u}_j^\top \boldsymbol{v},$$

where

$$\boldsymbol{u}_j = (\boldsymbol{I} - \boldsymbol{H}_0) \boldsymbol{x}_j \quad \text{and} \quad \boldsymbol{v} = (\boldsymbol{I} - \boldsymbol{H}_0) \boldsymbol{y}.$$

Thus, $\widehat{\beta}_j$ is effect of $\boldsymbol{x}_j$ on $\boldsymbol{y}$ after having accounted for the effects of all other variables in the model.

Tools

For

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}, \quad \text{where} \quad \boldsymbol{A}_{21} = \boldsymbol{A}_{12}^\top,$$

if

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{bmatrix} = \boldsymbol{A}^{-1},$$

then

$$\begin{aligned} \boldsymbol{B}_{11} &= (\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}, \\ \boldsymbol{B}_{22} &= (\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1}, \\ \boldsymbol{B}_{12} &= -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{B}_{22}, \qquad \boldsymbol{B}_{21} = \boldsymbol{B}_{12}^\top. \end{aligned}$$

Appendix

Functional Estimation In General

Model
$$\mathbb{E}(y|\boldsymbol{x}) = f(\boldsymbol{x}) = \sum_{j=1}^{q} \beta_j \varphi_j(\boldsymbol{x}),$$
as a more flexible (e.g., nonlinear, smooth, ...) function but approximate/represent it with basis expansion and turn "exotic" problem of functional estimation into a "basic" one of estimating the coefficients $\beta_1, ..., \beta_q$, i.e., linear regression.

$$\begin{bmatrix} f(\boldsymbol{x}_1) \\ \vdots \\ f(\boldsymbol{x}_n) \end{bmatrix} = \begin{bmatrix} \varphi_1(\boldsymbol{x}_1) & \cdots & \varphi_q(\boldsymbol{x}_1) \\ \vdots & \ddots & \vdots \\ \varphi_1(\boldsymbol{x}_n) & \cdots & \varphi_q(\boldsymbol{x}_n) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_q \end{bmatrix}$$

For example, $\{\varphi_j(\boldsymbol{x})\}_{j=1}^{q} \sim$ Fourier basis; wavelets; B-splines. Here, q often big, so need extra constraints to estimate $\boldsymbol{\beta}$. (Why?)