Linear Regression

Professor Jie Peng, PhD

Department of Statistics

University of California, Davis

Geometric Interpretation

Column Space of the Design Matrix

- $\mathbf{1}_n$: $n \times 1$ vector of ones; $\mathbf{x} = (X_1, \cdots, X_n)^T$: $n \times 1$ vector of X values.
- Design matrix: $\mathbf{X} = (\mathbf{1}_n, \mathbf{x})$.
- $\mathrm{col}(\mathrm{X})$: linear subspace of $\mathbb{R}^n$ generated by the columns of $\mathbf{X}$

$$\mathrm{col}(\mathrm{X}) = \{\mathbf{v} \in \mathbb{R}^{\mathrm{n}} : \exists \mathrm{c}_0, \mathrm{c}_1 \in \mathrm{R}, \mathrm{s.t.}, \mathbf{v} = \mathrm{c}_0 \mathbf{1}_{\mathrm{n}} + \mathrm{c}_1 \mathbf{x}\}.$$

Projection to $\mathrm{col}(\mathrm{X})$

$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ projects a vector in $\mathbb{R}^n$ to $\mathrm{col}(\mathrm{X})$: For any $\mathbf{w} \in \mathbb{R}^n$

- $\mathbf{Hw} \in \mathrm{col}(\mathrm{X})$, i.e., there exists $c_0, c_1 \in \mathbf{R}$ such that

  $\mathbf{Hw} = c_0 \mathbf{1}_n + c_1 \mathbf{x}$.

- $\mathbf{w} - \mathbf{Hw} \perp \mathrm{col}(\mathrm{X})$, i.e., for any $\mathbf{v} \in \mathrm{col}(\mathrm{X})$, the inner product

  $< \mathbf{w} - \mathbf{Hw}, \mathbf{v} >= (\mathbf{w} - \mathbf{Hw})^T \mathbf{v} = 0$.

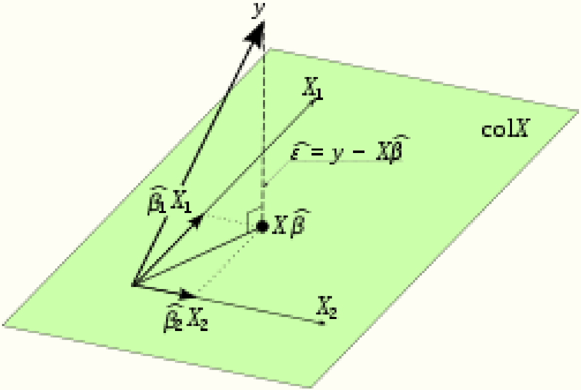
Fitted Values and Residuals

- $\widehat{\mathbf{Y}} = \mathbf{HY} = \hat{\beta}_0 \mathbf{1}_n + \hat{\beta}_1 \mathbf{x} \in \mathrm{col}(\mathrm{X})$
- $\mathbf{e} = \mathbf{Y} - \mathbf{HY} \perp \mathrm{col}(\mathrm{X})$
- Since $\mathbf{1}_n, \mathbf{x}, \widehat{\mathbf{Y}} \in \mathrm{col}(\mathrm{X})$, so

$$
\begin{aligned}
< \mathbf{e}, \mathbf{1}_n > &= \sum_{i=1}^{n} e_i = 0 \\
< \mathbf{e}, \mathbf{x} > &= \sum_{i=1}^{n} X_i e_i = 0 \\
< \mathbf{e}, \widehat{\mathbf{Y}} > &= \sum_{i=1}^{n} \hat{Y}_i e_i = 0
\end{aligned}
$$

Geometric Interpretation

Figure: Orthogonal projection of response vector **Y** onto the linear subspace of $\mathbb{R}^n$ generated by the columns of the design matrix **X**



Sums of Squares: Matrix Form

Error Sum of Squares

$$SSE = \sum_{i=1}^{n} e_i^2$$

can be expressed in matrix form:

$$SSE = \mathbf{e}'\mathbf{e} = \mathbf{Y}'(\mathbf{I}_n - \mathbf{H})'(\mathbf{I}_n - \mathbf{H})\mathbf{Y} = \mathbf{Y}'(\mathbf{I}_n - \mathbf{H})\mathbf{Y}.$$

- $\mathbf{I}_n - \mathbf{H}$ is a projection matrix.
- $df(SSE) = rank(\mathbf{I}_n - \mathbf{H}) = n - 2.$

Total Sum of Squares

$$SSTO = \sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n(\overline{Y})^2$$

can be expressed in matrix form:

$$SSTO = \mathbf{Y}'\mathbf{Y} - \frac{1}{n}\mathbf{Y}'\mathbf{J}_n\mathbf{Y} = \mathbf{Y}'\left(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n\right)\mathbf{Y}.$$

- $\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n$ is a projection matrix.

$$\mathbf{J}_n = \mathbf{1}_n\mathbf{1}_n' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

- $df(SSTO) = rank(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n) = n - 1.$

Regression Sum of Squares

$$SSR = \sum_{i=1}^{n} (\widehat{Y}_i - \overline{Y})^2$$

can be expressed in matrix form:

$$\begin{aligned} SSR &= \left(\widehat{\mathbf{Y}} - \overline{\mathbf{Y}}\right)' \left(\widehat{\mathbf{Y}} - \overline{\mathbf{Y}}\right), & \overline{\mathbf{Y}} = \frac{1}{n}\mathbf{J}_n\mathbf{Y} \\ &= \mathbf{Y}'\left(\mathbf{H} - \frac{1}{n}\mathbf{J}_n\right)'\left(\mathbf{H} - \frac{1}{n}\mathbf{J}_n\right)\mathbf{Y} \\ &= \mathbf{Y}'(\mathbf{H} - \frac{1}{n}\mathbf{J}_n)\mathbf{Y} \end{aligned}$$

- $\mathbf{H} - \frac{1}{n}\mathbf{J}_n$ is a projection matrix.
- $df(SSR) = rank(\mathbf{H} - \frac{1}{n}\mathbf{J}_n) = 1$.

Expectation of SSE

$$
\begin{aligned}
E(SSE) &= E\left(\mathbf{Y}'(\mathbf{I}_n - \mathbf{H})\mathbf{Y}\right) = E\left(Tr((\mathbf{I}_n - \mathbf{H})\mathbf{Y}\mathbf{Y}')\right) \\
&= Tr\left((\mathbf{I}_n - \mathbf{H})E(\mathbf{Y}\mathbf{Y}')\right) \\
&= Tr\Big((\mathbf{I}_n - \mathbf{H})(\sigma^2\mathbf{I}_n + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}')\Big) \\
&= \sigma^2 Tr\left(\mathbf{I}_n - \mathbf{H}\right) + Tr\left((\mathbf{I}_n - \mathbf{H})\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'\right) \\
&= (n-2)\sigma^2
\end{aligned}
$$

The last equality is because $Tr\left(\mathbf{I}_n - \mathbf{H}\right) = n - 2$ and $(\mathbf{I}_n - \mathbf{H})\mathbf{X} = \mathbf{0}$.

Confidence Interval for σ^2

Under the Normal error model:

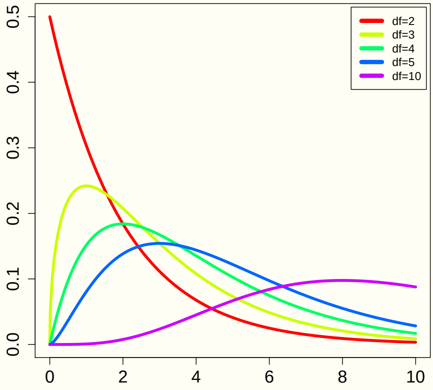
- $SSE \sim \sigma^2 \chi^2_{(n-2)}$
- Pivotal quantity:

$$\frac{SSE}{\sigma^2} \sim \chi^2_{(n-2)}$$

- $(1 - \alpha)100\%$-confidence interval for σ^2:

$$\left[\frac{SSE}{\chi^2(1-\alpha/2; n-2)}, \frac{SSE}{\chi^2(\alpha/2; n-2)}\right]$$

Probability Density Curves of χ^2 Distributions



Properties of Projection Matrices

Optional material.

- Eigen-decomposition: $Q\Lambda Q^T$, where Q is an orthogonal matrix of eigenvectors and Λ is a diagonal matrix of eigenvalues.
- Eigenvalues are either 1 or 0.
- The number of nonzero eigenvalues equals trace of the matrix equals the rank.
- For example. In simple linear regression:

$$rank(\mathbf{H}) = 2, \quad rank(\mathbf{I}_n - \mathbf{H}) = n - 2$$

Sampling Distribution of *SSE* under Normal Error Model

Optional material (Cont'd).

- $\mathbf{I}_n - \mathbf{H}$ is a projection matrix with rank $n - 2 \Longrightarrow$

$$\mathbf{I}_n - \mathbf{H} = \mathbf{Q}^T \Lambda \mathbf{Q},$$

  where $\Lambda = \mathrm{diag}\{1, \cdots, 1, 0, 0\}$ and $\mathbf{Q}$ is an orthogonal matrix.

- $(\mathbf{I}_n - \mathbf{H})\mathbf{X} = \mathbf{0} \Longrightarrow$

$$\mathbf{e} = (\mathbf{I}_n - \mathbf{H})\mathbf{Y} = (\mathbf{I}_n - \mathbf{H})(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) = (\mathbf{I}_n - \mathbf{H})\boldsymbol{\epsilon}$$

Optional material (Cont'd).

- $SSE = \mathbf{e}^T\mathbf{e} = \boldsymbol{\epsilon}^T(\mathbf{I}_n - \mathbf{H})\boldsymbol{\epsilon} = (\mathbf{Q}\boldsymbol{\epsilon})^T\Lambda(\mathbf{Q}\boldsymbol{\epsilon})$
- Let $\mathbf{z} = \mathbf{Q}\boldsymbol{\epsilon}$, then
$$SSE = \sum_{i=1}^{n-2} z_i^2$$
- Moreover
$$\mathbf{E}(\mathbf{z}) = \mathbf{Q}\mathbf{E}\{\boldsymbol{\epsilon}\} = \mathbf{0}, \quad \boldsymbol{\sigma}^2\{\mathbf{z}\} = \mathbf{Q}\boldsymbol{\sigma}^2\{\boldsymbol{\epsilon}\}\mathbf{Q}^T = \sigma^2\mathbf{Q}\mathbf{Q}^T = \sigma^2\mathbf{I}_n$$
  So under Normal error model, z_is are i.i.d. $N(0, \sigma^2)$.
- Thus $SSE \sim \sigma^2\chi^2_{(n-2)}$.