

# STAT 3690 Lecture 16

zhiyanggeezhou.github.io

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca)

Mar 07, 2022

## What is a linear model?

- Responses are linear functions with respect to unknown parameters.

## Univariate/multiple linear regression (J&W Sec. 7.2–7.5)

- Interested in the relationship between random scalar  $Y$  and random  $q$ -vector  $[X_1, \dots, X_q]^\top$
- Model
  - Population version:  $Y \mid X_1, \dots, X_q \sim ([1, X_1, \dots, X_q]\boldsymbol{\beta}, \sigma^2)$ , where  $\boldsymbol{\beta} = [\beta_0, \dots, \beta_q]^\top$ , i.e.,
    - \*  $E(Y \mid X_1, \dots, X_q) = [1, X_1, \dots, X_q]\boldsymbol{\beta} = \beta_0 + \sum_{j=1}^q X_j\beta_j$
    - \*  $\text{var}(Y \mid X_1, \dots, X_q) = \sigma^2$
  - Sample version  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ 
    - \*  $\mathbf{Y} = [Y_1, \dots, Y_n]^\top$  and design matrix

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{q1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{nq} \end{bmatrix}_{n \times (q+1)}$$

- Independent realizations  $[Y_i, X_{i1}, \dots, X_{iq}]^\top \sim [Y, X_1, \dots, X_q]^\top, i = 1, \dots, n$
- $\text{rk}(\mathbf{X}) = q + 1 < p + q + 1 \leq n$
- \*  $\boldsymbol{\varepsilon} = [\varepsilon_1, \dots, \varepsilon_n]^\top \sim (\mathbf{0}_n, \sigma^2 \mathbf{I}_n)$

- 
- Least squares (LS) estimation (no need of normality)
    - $\hat{\boldsymbol{\beta}}_{\text{LS}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$
    - $\hat{\sigma}_{\text{LS}}^2 = (n - q - 1)^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{LS}})^\top (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{LS}}) = (n - q - 1)^{-1} \mathbf{Y}^\top (\mathbf{I} - \mathbf{H}) \mathbf{Y}$ 
      - \* Hat matrix  $\mathbf{H} = [h_{ij}]_{n \times n} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ 
        - Symmetric
        - Idempotent:  $\mathbf{H}^2 = \mathbf{H}\mathbf{H} = \mathbf{H}$
        - $\text{rk}(\mathbf{H}) = \text{rk}(\mathbf{X})$
        - Each eigenvalue of  $\mathbf{H}$  is either zero or one
      - \*  $E(\hat{\sigma}_{\text{LS}}^2) = \sigma^2$

a. Prove that  $\hat{\beta}_{LS} = \arg \min_{\beta} Q(\beta) = (Y - X\beta)^T (Y - X\beta) = \arg \min_{\beta} Y^T Y - 2\beta^T X^T Y + \beta^T X^T X \beta$

Known:  $\frac{\partial Q(\beta)}{\partial \beta} = -2X^T Y + 2X^T X \beta$  (Matrix calculus)

Let  $\partial Q(\beta) / \partial \beta = 0$ , then  $X^T Y = X^T X \beta$ .

So,  $\hat{\beta}_{LS} = (X^T X)^{-1} X^T Y$  is a stationary point, i.e., a candidate for the minimizer

Actually, for any  $\beta$ ,

$$\begin{aligned} Q(\beta) &= (Y - X\hat{\beta}_{LS} + X\hat{\beta}_{LS} - X\beta)^T (Y - X\hat{\beta}_{LS} + X\hat{\beta}_{LS} - X\beta) \\ &= \{Y - X\hat{\beta}_{LS} + X(\hat{\beta}_{LS} - \beta)\}^T \{Y - X\hat{\beta}_{LS} + X(\hat{\beta}_{LS} - \beta)\} \\ &= (Y - X\hat{\beta}_{LS})^T (Y - X\hat{\beta}_{LS}) + \underbrace{(\hat{\beta}_{LS} - \beta)^T X^T X (\hat{\beta}_{LS} - \beta)}_{\geq 0} + \underbrace{2(Y - X\hat{\beta}_{LS})^T X (\hat{\beta}_{LS} - \beta)}_{=0} \end{aligned}$$

$$\therefore Q(\beta) = (Y - X\hat{\beta}_{LS})^T (Y - X\hat{\beta}_{LS}) + \underbrace{(\hat{\beta}_{LS} - \beta)^T X^T X (\hat{\beta}_{LS} - \beta)}_{\geq 0} = Q(\hat{\beta}_{LS})$$

That is,  $Q(\beta) \geq (Y - X\hat{\beta}_{LS})^T (Y - X\hat{\beta}_{LS}) = Q(\hat{\beta}_{LS})$  for all  $\beta$ .

b. Prove that  $E(\hat{\sigma}_{LS}^2) = \sigma^2$

$$\begin{aligned} Y^T (I - H) Y &= (Y - X\beta + X\beta)^T (I - H) (Y - X\beta + X\beta) \\ &= (Y - X\beta)^T (I - H) (Y - X\beta) + \underbrace{2\beta^T X^T (I - H) (Y - X\beta)}_{=0} + \underbrace{\beta^T X^T (I - H) X \beta}_{=0} \\ &= \text{tr}\{(I - H)(Y - X\beta)(Y - X\beta)^T\} \end{aligned}$$

$$\begin{aligned} \therefore E\{Y^T (I - H) Y | X\} &= E\{\text{tr}\{(I - H)(Y - X\beta)(Y - X\beta)^T\} | X\} \\ &= \text{tr}\{(I - H) E\{(Y - X\beta)(Y - X\beta)^T | X\}\} \\ &= \text{tr}\{(I - H) \text{Cov}(Y - X\beta | X)\} \\ &= \text{tr}\{(I - H) \text{Cov}(Y | X)\} \\ &= \text{tr}\{\sigma^2 (I - H)\} = \sigma^2 (n - 1) \end{aligned}$$

$$\therefore E(\hat{\sigma}_{LS}^2 | X) = \sigma^2 \text{ for all } X$$

$$\therefore E(\hat{\sigma}_{LS}^2) = \sigma^2$$

- Maximum likelihood (ML) estimation (in need of (conditional) normality)

$$\begin{aligned} -\hat{\beta}_{ML} &= (X^T X)^{-1} X^T Y = \hat{\beta}_{LS} \\ * \text{ Given } X, \hat{\beta}_{ML} &\sim MVN_{q+1}(\beta, \sigma^2 (X^T X)^{-1}) \\ -\hat{\sigma}_{ML}^2 &= n^{-1} Y^T (I - H) Y = n^{-1} (n - q - 1) \hat{\sigma}_{LS}^2 \\ * \text{ Given } X, n\hat{\sigma}_{ML}^2 / \sigma^2 &= (n - q - 1) \hat{\sigma}_{LS}^2 / \sigma^2 \sim \chi^2(n - q - 1) \end{aligned}$$

$$\therefore \ln L(X\beta, \sigma) = \text{const} - n \ln \sigma - \frac{1}{2\sigma^2} (Y - X\beta)^T (Y - X\beta)$$

$$\therefore \text{for each } \sigma, \text{ maximize } \ln L(X\beta, \sigma) \Leftrightarrow \text{minimize } (Y - X\beta)^T (Y - X\beta)$$

$$\therefore \hat{\beta}_{ML} = \hat{\beta}_{LS}$$

$$\therefore \hat{\sigma}_{ML} = \arg \max_{\sigma} \ln L(X\hat{\beta}_{LS}, \sigma)$$

$$\text{Let } \partial \ln L(X\hat{\beta}_{LS}, \sigma) / \partial \sigma = -n/\sigma + \sigma^{-3} (Y - X\hat{\beta}_{LS})^T (Y - X\hat{\beta}_{LS}) = 0$$

Then  $\hat{\sigma} = \{(Y - X\hat{\beta}_{LS})^T (Y - X\hat{\beta}_{LS}) / n\}^{1/2}$  is a stationary point.

$$\begin{aligned} \therefore \partial^2 \ln L(X\hat{\beta}_{LS}, \sigma) / \partial \sigma^2 \big|_{\sigma=\hat{\sigma}} &= n\hat{\sigma}^{-2} - 3\hat{\sigma}^{-4} (Y - X\hat{\beta}_{LS})^T (Y - X\hat{\beta}_{LS}) \\ &= \frac{n^2 - 3n^2}{(Y - X\hat{\beta}_{LS})^T (Y - X\hat{\beta}_{LS})} < 0 \end{aligned}$$

$$\therefore \hat{\sigma} = \arg \max_{\sigma} \ln L(X\hat{\beta}_{LS}, \sigma) = \hat{\sigma}_{ML}$$

- Inference (in need of (conditional) normality)
  - Inference on  $\mathbf{a}^\top \boldsymbol{\beta}$ , given  $\mathbf{a} \in \mathbb{R}^{q+1}$ 
    - \* Estimator  $\mathbf{a}^\top \hat{\boldsymbol{\beta}}_{\text{ML}}$
    - \*  $100(1 - \alpha)\%$  confidence interval for  $\mathbf{a}^\top \boldsymbol{\beta}$ :

$$\mathbf{a}^\top \hat{\boldsymbol{\beta}}_{\text{ML}} \pm t_{1-\alpha/2, n-q-1} \hat{\sigma}_{\text{LS}} [\mathbf{a}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{a}]^{1/2}$$

- Inference on  $Y_0 = \mathbf{X}_0^\top \boldsymbol{\beta} + \varepsilon_0$  with a new observation vector given  $\mathbf{X}_0 = [1, X_{01}, \dots, X_{0q}]^\top \in \mathbb{R}^{q+1}$ 
  - \* Prediction  $\hat{Y}_0 = \mathbf{X}_0^\top \hat{\boldsymbol{\beta}}_{\text{ML}}$
  - \*  $100(1 - \alpha)\%$  prediction interval for  $Y_0$

$$\mathbf{X}_0^\top \hat{\boldsymbol{\beta}}_{\text{ML}} \pm t_{1-\alpha/2, n-q-1} \hat{\sigma}_{\text{LS}} [1 + \mathbf{X}_0^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}_0]^{1/2}$$