JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Lecture 6

The classical linear regression model: Maximum likelihood estimates = least square solution, Distributions of key results, vector notation

$\hat{\beta}$

Maximum likelihood estimation under gaussian residuals

- Model: $Y_i = \underbrace{\mu_i(\beta, X_i)} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$ independent $i = 1, \dots, n$

  $\mu_i(\beta, X_i) = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi} \qquad Y_i \sim N(\mu_i(\beta, X_i), \sigma^2)$

- Likelihood function: a mathematical function of $\mu_i(\underbrace{\beta}, X_i)$ and σ^2 for fixed y

  $$L(\beta, \sigma^2 \mid y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-1}{2\sigma^2}(y_i - \underbrace{\beta_0 - \beta_1 X_{1i} - \dots - \beta_p X_{pi}}_{\mu_i(\beta, X_i)})^2\right)$$

- Log Likelihood function:

  $$\ell(\beta, \sigma^2 \mid y) = \log L(\beta, \sigma^2 \mid y)$$
  $$= \sum_{i=1}^{n} \left[-\frac{1}{2}\log(2\pi) - \log(\sigma) - \frac{1}{2\sigma^2} \underbrace{(y_i - \mu_i(\beta, X_i))^2}_{*} \right]$$

log likelihood will be maximized when we can minimize *

Maximum likelihood estimation under gaussian residuals

- Solution for β_j ⇒ derivative of log-likelihood wrt to β_j
  ⇒ set $=0$ and solve for $\beta_j = \hat{\beta}_{j,mle}$

$\ell(\beta, \sigma^2 | y)$ as a function of β is proportional to $\sum_{i=1}^{n} (y_i - \mu_i(\beta, X_i))^2$

$$U_{\beta_j}(\beta | \sigma^2) = \frac{d}{d\beta_j} \ell(\beta, \sigma^2 | y)$$

score equation

$$= \frac{d}{d\beta_j} \sum_{i=1}^{n} (y_i - \mu_i(\beta, X_i))^2$$

$$= \sum_{i=1}^{n} 2 (y_i - \underbrace{\mu_i(\beta, X_i)}_{\beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi}})(-X_{ji})$$

set $=0$ and solve for β_j

Maximum likelihood estimation under gaussian residuals

- Solution for β_j

$$U_{\underset{\sim}{\beta}} = \sum_{i=1}^{n} \left(y_i - \mu_i(\underset{\sim}{\beta}, X_i) \right) \begin{bmatrix} 1 \\ X_{1i} \\ X_{2i} \\ \vdots \\ X_{pi} \end{bmatrix}_{(p+1)\times 1} = 0$$

This does not involve $\underset{\sim}{\sigma}^2$

* to confirm the maximum
$\rightarrow \frac{d}{d\beta_j} U_{\underset{\sim}{\beta}}$ evaluated at $\hat{\beta}$
to show these are negative

$p+1$ equations
$p+1$ unknowns: $\underset{\sim}{\beta}$

Maximum likelihood estimation under gaussian residuals

- Solution for σ^2

Assume the MLEs for $\beta \Rightarrow \hat{\beta}$

$$U_{\sigma^2}(\hat{\beta}) = \frac{d}{d\sigma^2} \sum_{i=1}^{n} \left(-\log(\sigma) - \frac{1}{2\sigma^2}\left(y_i - \mu_i(\hat{\beta}, X_i)\right)^2 \right)$$

$$= \sum_{i=1}^{n} \left(\frac{1}{-2\sigma^2} + \frac{1}{2(\sigma^2)^2}\left(y_i - \mu_i(\hat{\beta}, X_i)\right)^2 \right)$$

↳ set equal to 0 and solve for σ^2

$$E(\hat{\sigma}^2_{MLE}) = \frac{n-(p+1)}{n}\sigma^2$$

$$\hat{\sigma}^2_{MLE} = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \mu_i(\hat{\beta}, X_i)\right)^2$$

$$\tilde{\sigma}^2 = \frac{1}{n-(p+1)} \sum_{i=1}^{n} \left(y_i - \mu_i(\hat{\beta}, X_i)\right)^2$$

MLEs for simple linear regression $p=1$

Solve for $\beta_0, \beta_1 \Rightarrow$ 2 score equations

$$\underbrace{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 X_i) = 0}$$ and $$\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 X_i)X_i = 0$$

$$\sum_{i=1}^{n} y_i - n\beta_0 - \beta_1 \sum_{i=1}^{n} X_i = 0$$

$$\sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} X_i = n\beta_0$$

$$\frac{1}{n}\sum_{i=1}^{n} y_i - \beta_1 \frac{1}{n}\sum_{i=1}^{n} X_i = \beta_0$$

$$\overline{y} - \beta_1 \overline{X} = \hat{\beta}_0$$

MLEs for simple linear regression

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) x_i = 0$$

$$\sum_{i=1}^{n} (y_i - (\bar{y} - \beta_1 \bar{x}) - \beta_1 x_i) x_i = 0$$

$$\sum_{i=1}^{n} (\underbrace{y_i x_i - \bar{y} x_i} - \beta_1 (x_i - \bar{x}) x_i) = 0$$

$$\sum_{i=1}^{n} (y_i - \bar{y}) x_i - \beta_1 \sum_{i=1}^{n} (x_i - \bar{x}) x_i = 0$$

$$\frac{\sum_{i=1}^{n} (y_i - \bar{y}) x_i}{\sum_{i=1}^{n} (x_i - \bar{x}) x_i} = \hat{\beta}_1 \overset{?}{=} \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

MLEs for simple linear regression

$$\frac{\sum x_i}{n} = \bar{x}$$

$$\sum x_i = n\bar{x}$$

$$\underbrace{\sum (y_i - \bar{y}) x_i}_{} = \sum y_i x_i - \bar{y} \sum x_i$$

$$= \sum y_i x_i - n\bar{y}\bar{x}$$

$$= \sum y_i x_i - n\bar{y}\bar{x} - n\bar{y}\bar{x} + n\bar{y}\bar{x}$$

$$= \sum \left[y_i x_i - \bar{y} x_i - \bar{x} y_i + \bar{y}\bar{x} \right]$$

$$= \underbrace{\sum (y_i - \bar{y})(x_i - \bar{x})}_{}$$

Take away messages

1. for ε_i assumed to be iid $N(0,\sigma^2)$,
the least squares solution $\Leftrightarrow$ ML solution

2. Each β_j, $j=1,\ldots,p$ is a linear function of y

$$\hat{\beta}_j = \sum_{i=1}^{n} w_{ij}(X_i)\, y_i \qquad p=1 \quad w_{ij} = \frac{(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

3. $\hat{\beta}$ are not robust to outliers $\Rightarrow (y_i - \bar{y})$

4. Observations with large $(X_i - \bar{X})$ have greater weights $\Leftrightarrow$ leverage

Properties of sums of independent Gaussian random variables

Understand the distribution of $\hat{\beta}$ $\Rightarrow$ generating CIs

$Y_1, \ldots, Y_n$ are independent $N(\mu_i, \sigma_i^2)$ $\Rightarrow$ conducting hypothesis test

Define $d = \sum_{i=1}^{n} a_i Y_i$ $=$ d as a linear combination of $\underset{\sim}{Y}$ with weights a_i

$$d \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

Distribution of $\hat{\beta}_1$ in SLR assuming Gaussian residuals

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{\underbrace{\sum_{i=1}^{n} (x_i - \bar{x})^2}_{SSX}} = \sum a_i y_i$$

$$a_i = \frac{(x_i - \bar{x})}{SSX}$$

$$\hat{\beta}_1 \sim N\left(\sum_{i=1}^{n} a_i E(y_i), \sum_{i=1}^{n} a_i^2 \underbrace{Var(y_i)}_{\sigma^2} \right)$$

$$\sim N\left(\sum_{i=1}^{n} a_i (\beta_0 + \beta_1 x_i), \sigma^2 \sum_{i=1}^{n} a_i^2 \right)$$

Distribution of $\hat{\beta}_1$ in SLR assuming Gaussian residuals

$$E(\hat{\beta}_1) = \sum_{i=1}^{n} a_i (\beta_0 + \beta_1 X_i)$$

$$= \frac{\sum_{i=1}^{n} (X_i - \bar{X})(\beta_0 + \beta_1 X_i)}{SSX} = \frac{\beta_0 \cancel{\sum (X_i - \bar{X})}^{\,0}}{SSX} + \frac{\beta_1 \sum (X_i - \bar{X}) X_i}{SSX}$$

$$= \beta_1 \frac{SSX}{SSX} = \beta_1 \quad \text{unbiased}$$

$$Var(\hat{\beta}_1) = \sigma^2 \sum_{i=1}^{n} a_i^2 = \sigma^2 \frac{\sum (X_i - \bar{X})^2}{SSX^2} = \sigma^2 / SSX$$

$$\hat{\beta}_0 \sim N\left(\beta_0, \frac{\sigma^2}{n} + ?\right) \Rightarrow \text{see Handout}$$

Implications for data analysis

① MLE $\Longleftrightarrow$ LS when $\varepsilon_i \sim$ Normal

② $\hat{\beta}$ are linear combinations of $Y_1, \ldots, Y_n$

$\Rightarrow$ $\hat{\beta}$ are also Gaussian

$\Rightarrow$ even if ε_i are not normal

$\hat{\beta}$ $\Leftrightarrow$ weighted averages of $Y_1, \ldots, Y_n$

for sufficiently large n

CLT $\Rightarrow$ $\hat{\beta} \sim$ Normal

③ $\hat{\beta}$ $\Rightarrow$ sensitive to extreme observations

Y space X space

MLE or LS solution expressed in vector notation

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \varepsilon_i \qquad i = 1, \dots, n$$

$$\begin{array}{c|l}
i & \\
1 & Y_1 = 1 \cdot \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \dots + \beta_p X_{p1} + \varepsilon_1 \\
2 & Y_2 = 1 \cdot \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} + \dots + \beta_p X_{p2} + \varepsilon_2 \\
\vdots & \quad \vdots \\
n & Y_n = 1 \cdot \beta_0 + \beta_1 X_{1n} + \beta_2 X_{2n} + \dots + \beta_p X_{pn} + \varepsilon_n
\end{array}$$

$$\underset{n\times 1}{\underset{\sim}{Y}} = \underset{n\times 1}{\underset{\sim}{1}}\,\beta_0 + \underset{n\times 1}{\underset{\sim}{X_1}}\,\beta_1 + \underset{n\times 1}{\underset{\sim}{X_2}}\,\beta_2 + \dots + \underset{n\times 1}{\underset{\sim}{X_p}}\,\beta_p + \underset{n\times 1}{\underset{\sim}{\varepsilon}}$$

$$X = \underset{n \times (p+1)}{\left[\underset{\sim}{1}, \underset{\sim}{X_1}, \underset{\sim}{X_2}, \dots, \underset{\sim}{X_p}\right]} \quad \underset{p+1}{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}}$$

MLE or LS solution expressed in vector notation

$$\underset{\sim}{Y}_{n\times 1} = \underbrace{\begin{bmatrix} 1 & X_{11} & \cdots & X_{p1} \\ 1 & X_{12} & & X_{p2} \\ \vdots & \vdots & & \vdots \\ 1 & X_{1n} & & X_{pn} \end{bmatrix}_{n\times(p+1)} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}_{(p+1)\times 1}}_{n\times 1} + \underset{\sim}{\varepsilon}_{n\times 1}$$

cross product for 1st row: $1\cdot\beta_0 + \beta_1 X_{11} + \cdots + \beta_p X_{p1}$

$$\underset{\sim}{Y}_{n\times 1} = \underbrace{X \underset{\sim}{\beta}}_{\substack{n\times 1 \\ n\times(p+1)\ (p+1)\times 1}} + \underset{\sim}{\varepsilon}_{n\times 1}$$

X = design matrix

MLR model expressed in vector notation

What is the distribution for $\underset{\sim}{\varepsilon}_{n\times 1}$

$\underset{\sim}{Y} \sim MVN(\underset{\sim}{\mu}, V)$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \qquad \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} \qquad V = \begin{bmatrix} V_{11} & V_{12} & \cdots & V_{1n} \\ V_{21} & V_{22} & \cdots & V_{2n} \\ \vdots & & & \\ V_{n1} & V_{n2} & \cdots & V_{nn} \end{bmatrix}$$

$V_{ii} = Var(Y_i)$

$V_{ij} = Cov(Y_i, Y_j)$

$\varepsilon_i \sim N(0, \sigma^2)$, $Cov(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$

$$\underset{\sim}{\varepsilon} \sim MVN\left(\underset{\sim}{0}, \begin{bmatrix} \sigma^2 & & 0 \\ & \sigma^2 & \\ 0 & & \ddots \ \sigma^2 \end{bmatrix}\right)$$

$\sigma^2 I_{n\times n}$

$$\begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & \ddots \end{bmatrix}$$

MLR model expressed in vector notation

MLR model: $\underset{n\times1}{Y} = \underset{n\times1}{X\beta} + \underset{n\times1}{\varepsilon}$, $\underset{n\times1}{\varepsilon} \sim MVN(\underset{n\times1}{0}, \sigma^2 \underset{n\times n}{I})$

Choose $\hat{\beta}$ and $\hat{\sigma}^2$ to minimize $\sum_{i=1}^{n} r_i(\beta)^2 = \sum_{i=1}^{n} (y_i - X_i\beta)^2$

$$\sum_{i=1}^{n} (y_i - X_i\beta)^2 = \underset{1\times n}{(Y-X\beta)'}\underset{n\times1}{(Y-X\beta)}$$

$$U_\beta(\beta) = \frac{d}{d\beta}(Y-X\beta)'(Y-X\beta) = X'(Y-X\beta)$$

$\Rightarrow 0$ solve for β

$$X'Y - X'X\beta = 0$$

$$X'X\beta = X'Y$$

$$\hat{\beta} = (X'X)^{-1}X'Y \Rightarrow \text{solution}$$

$X'Y \Rightarrow \sum (X_i - \bar{X})(Y_i - \bar{Y})$

$X'X \Rightarrow \sum (X_i - \bar{X})^2$

Next time….

- We will use vector notation to derive the distribution of key results including the estimated regression coefficient vector, predicted values and residuals
- Geometry of least squares
- What happens to our inferences when the Gaussian assumption is violated? We will explore this via simulation study