JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

PS 1 → next monday
Quiz 1 → solutn

PS2 is posted
Datasets → NMES

Lecture 7

Vector representation of MLR continued, assessing the impact of Gaussian residuals assumption

MLR model expressed in vector notation

We have for each subject i, $i = 1, \dots, n$

$$Y_i = \underbrace{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi}}_{\text{systematic component: } \mu_i} + \underbrace{\varepsilon_i}_{\text{error residual}}$$

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

$$\underset{\sim}{Y}_{n\times 1} = \underset{\sim}{1}_{n\times 1} \beta_0 + \beta_1 \underset{\sim}{X}_{1\, n\times 1} + \beta_2 \underset{\sim}{X}_{2\, n\times 1} + \dots \beta_p \underset{\sim}{X}_{n\, n\times 1} + \underset{\sim}{\varepsilon}_{n\times 1}$$

$$\underset{\sim}{Y}_{n\times 1} = \underbrace{(\underset{\sim}{1}, \underset{\sim}{X}_1, \underset{\sim}{X}_2, \dots, X_p)}_{X_{n\times(p+1)} \text{ design matrix}} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}_{p+1\times 1} + \underset{\sim}{\varepsilon}_{n\times 1}$$

MLR model expressed in vector notation

$$\Rightarrow \underset{\sim}{Y}_{n\times 1} = \underbrace{X_{n\times(p+1)}\,\underset{\sim}{\beta}_{(p+1)\times 1}}_{\underset{\sim}{\mu}} + \underset{\sim}{\varepsilon}_{n\times 1}$$

Say $n=3,\ p=2 \Rightarrow Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$

$$\underset{\sim}{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \quad X = \begin{bmatrix} 1 & X_{11} & X_{21} \\ 1 & X_{12} & X_{22} \\ 1 & X_{13} & X_{23} \end{bmatrix}_{3\times 3} \quad \underset{\sim}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}_{3\times 1} \quad \underset{\sim}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}_{3\times 1}$$

$$\underbrace{X_{3\times 3}\,\underset{\sim}{\beta}_{3\times 1}}_{} = \begin{pmatrix} \boxed{1\cdot\beta_0 + \beta_1 X_{11} + \beta_2 X_{21}} \\ 1\cdot\beta_0 + \beta_1 X_{12} + \beta_2 X_{22} \\ 1\cdot\beta_0 + \beta_1 X_{13} + \beta_2 X_{23} \end{pmatrix}_{3\times 1} = \underset{\sim}{\mu}_{3\times 1}$$

Defines the mean of Y_i for each i

MLR model expressed in vector notation

What about distribution of $\underset{\sim}{\varepsilon}$? and $\underset{\sim}{Y}$?

In general, we can define the multivariate normal distribution as: $\underset{\sim}{Y}_{n\times 1} \sim \text{MVN}(\underset{\sim}{\mu}_{n\times 1}, V_{n\times n})$

where $\underset{\sim}{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$ $\underset{\sim}{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}$

$V = \begin{bmatrix} V_{11} & & V_{ij} \\ & V_{22} & \\ V_{ji} & & \ddots & V_{nn} \end{bmatrix}$ $V_{ii} = \text{Var}(Y_i)$

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

If $\varepsilon_i \sim N(0, \sigma^2)$, independent

$\underset{\sim}{\varepsilon} \sim \text{MVN}(\underset{\sim}{0}, \sigma^2 I)$ $\quad I = \begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & \ddots & 1 \end{bmatrix}$ identity matrix

$\underset{\sim}{Y} \sim \text{MVN}(X\underset{\sim}{\beta}, \sigma^2 I)$ $\quad V = \begin{bmatrix} \sigma^2 & & 0 \\ & \sigma^2 & \\ 0 & & \ddots & \sigma^2 \end{bmatrix}$

$X\underset{\sim}{\beta}$

MLE or LS solution expressed in vector notation

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

MLE or least squares: Going to drop the "~"

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

X_i = ith row of X

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

$(y_1 - X_1\beta \quad y_2 - X_2\beta)\begin{pmatrix} y_1 - X_1\beta \\ y_2 - X_2\beta \end{pmatrix}$

$(y_1 - X_1\beta)^2 + (y_2 - X_2\beta)^2$

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

$$= X'(Y - X\beta)$$

SLR, $\hat{\beta}_1 = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$

Set to 0, solve for $\hat{\beta} = (X'X)^{-1}X'Y$

Predicted values and residuals in vector notation

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

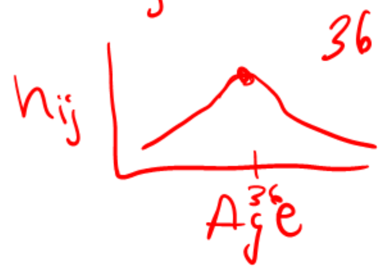
$$\hat{Y} = X\hat{\beta} = \underbrace{X(X'X)^{-1}X'}_{H_{n\times n} = \text{Hat matrix}}Y$$

$$\hat{R} = Y - \hat{Y} = Y - X\hat{\beta}$$

$$= Y - HY$$

$$= (I-H)Y$$

$$\hat{Y}_i = \sum_{j=1}^{n} h_{ij} Y_j$$



Distribution of $\hat{\beta}$

Note that if $\underset{\sim}{Y} \sim MVN(\underset{\sim}{\mu}, V)$, then

$\Rightarrow A\underset{\sim}{Y} \sim MVN(A\underset{\sim}{\mu}, \underset{\sim}{AVA'})$

$$\hat{\beta} = \underbrace{(X'X)^{-1}X'}_{A}Y$$

$$E(\hat{\beta}) = E(AY) = AE(Y) = AX\beta = \underbrace{(X'X)^{-1}X'X}\beta = \beta$$

$$Var(\hat{\beta}) = Var(AY) = (X'X)^{-1}X' Var(Y) X(X'X)^{-1}$$

SLR $Var(\hat{\beta}_1) = \sigma^2/SSX$

$$(X'X)^{-1}X'\sigma^2 I X(X'X)^{-1}$$

$$\sigma^2 (X'X)^{-1}\underbrace{X'X(X'X)^{-1}}$$

$$\leftarrow \sigma^2 (X'X)^{-1}$$

Distribution of $\hat{Y}$

$$\hat{Y} = X\hat{\beta} = \underbrace{X(X'X)^{-1}X'}Y = \underbrace{H}Y$$

$$E(HY) = HE(Y) = HX\beta = X\underbrace{(X'X)^{-1}}\underbrace{X'X}\beta$$

$$= X\beta$$

$$Var(\hat{y}) = H\,Var(Y)\,H'$$

$$= X(X'X)^{-1}X'\sigma^2 I\,X(X'X)^{-1}X'$$

$$= \sigma^2 X(X'X)^{-1}\underbrace{X'X(X'X)^{-1}}_{I}X'$$

$$= \sigma^2 H$$

Properties of the Hat matrix

1) hat matrix is symmetric

$$H' = \left[X\underbrace{(X'X)^{-1}}X'\right]' = X(X'X)^{-1}X' = H$$

2) hat matrix is idempotent: $H \cdot H = H$

See Lecture 7 Handout

Distribution of $\hat{R}$

$$\hat{R} = Y - \hat{Y} = Y - HY = \underbrace{(I - H)}Y$$

$$\hat{R} \sim MVN$$

$$E(\hat{R}) = E(Y - \hat{Y}) = X\beta - X\beta = 0$$

$$Var(\hat{R}) = \sigma^2 (I - H)$$

Relationship between $\hat{Y}$ and $\hat{R}$

$$Cov(\underbrace{\hat{Y}}_{HY}, \underbrace{\hat{R}}_{(I-H)Y}) = E[\underline{HY}\{(I-H)Y\}']$$

$$= E[HYY'(I-H)]$$

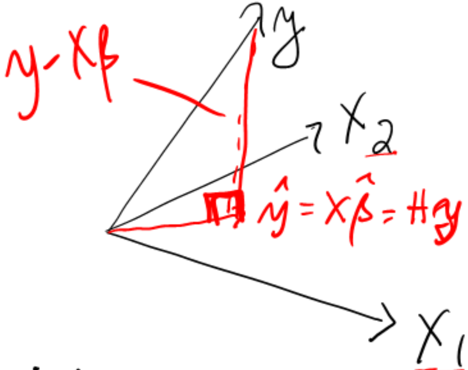
$$= H\underbrace{E[YY']}(I-H)$$

$$= H\sigma^2 I(I-H)$$

$$= \sigma^2 \underbrace{H(I-H)}_{H-HH = H-H = 0}$$

Geometry of least squares

Consider $\underset{\sim}{Y}_{n\times 1}$, $\underset{\sim}{X_1}_{n\times 1}$, $\underset{\sim}{X_2}_{n\times 1}$



H projects y onto the plane spanned by X_1, X_2

$\Rightarrow \hat{y} = Hy = X\hat{\beta}$

1) minimize the distance between y and $\hat{y} = X\hat{\beta}$

2) Shortest distance is the one that has a right angle between the predicted value and residual

3) residual is orthogonal to the plane spanned by X

4) Score equations:

$\Rightarrow X'(Y - X\beta) = 0$

Simulation study

- We derived the distribution of the estimated regression coefficients assuming the residuals were Gaussian.

- Does approximate normality of the estimated regression coefficients hold even when the residuals are non-Gaussian?

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}, \quad \underset{\sim}{\varepsilon} \sim \underbrace{MVN}(\underset{\sim}{0}, \sigma^2 I)$$

$$\underbrace{\hat{\beta}} \sim \underbrace{MVN}(\beta, \sigma^2 (X'X)^{-1})$$

Yes $\Rightarrow$ $\underbrace{(X'X)^{-1}X'}\underline{Y}$ = weighted average of Y

Next time....

- Deriving the distribution of linear combinations of regression coefficients
- Deriving the distribution of non-linear combinations of regression coefficients using the Delta method
- LAB: You will generate the distribution of combinations of regression coefficients using bootstrap!