

Multiple Linear Regression

MLR Model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_{p-1} x_{(p-1)i} + \varepsilon_i$$

$i = 1, \dots, n$

p = # of parameters in the model
(betas)

$p-1$ = # of predictors

In matrix notation, we can write

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

$(n \times 1)$

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{(p-1)1} \\ 1 & x_{12} & x_{22} & \dots & x_{(p-1)2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{(p-1)n} \end{bmatrix}$$

$(n \times p)$

Annotations for X matrix:
- 1st col: 1st obs
- 2nd col: 2nd obs
- 1st pred: 1st predictor
- 2nd pred: 2nd predictor
- last pred: last predictor

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}$$

$(p \times 1)$

$$\epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Last time:

$$y = X\beta + \epsilon$$

$(n \times 1) \quad (n \times p)(p \times 1) \quad (n \times 1)$

Model Assumptions:

① the relationship b/w y & $x_j \quad j=1, \dots, p-1$ is linear

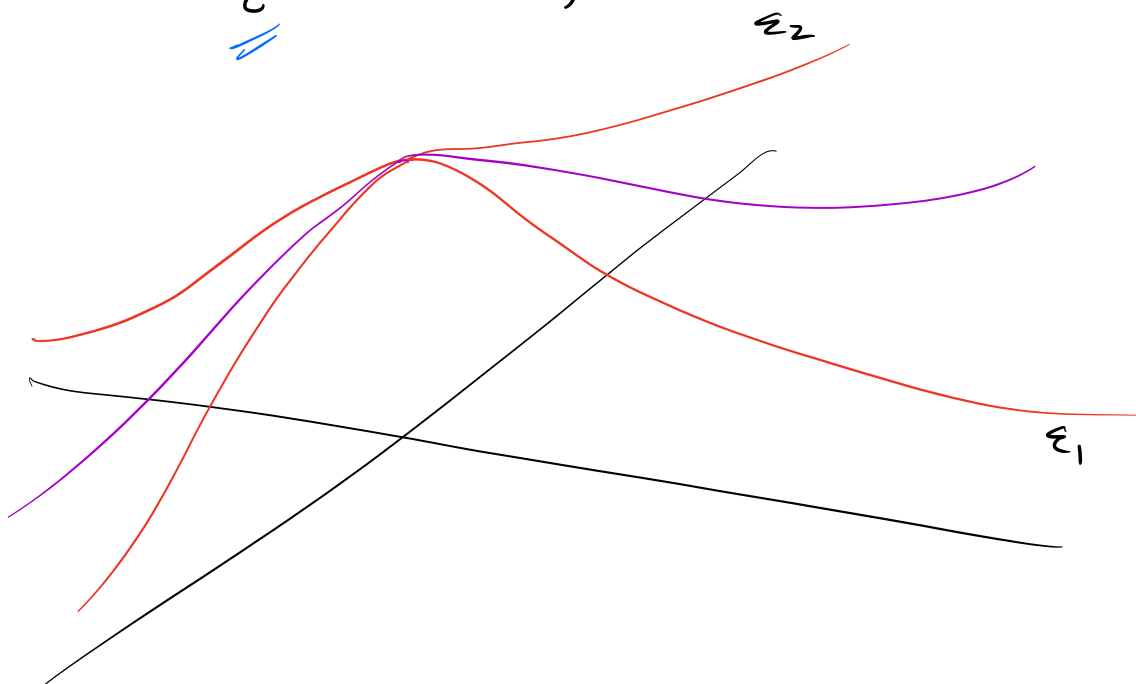
② X is fixed

③ $E(\epsilon) = 0$

$$\text{Var}(\epsilon_{ji}) = \sigma^2 \quad \forall i, j$$

$$\underline{\epsilon} \sim N(0, \sigma^2 I_n)$$

"multivariate
Normal
distribution"



Multivariate Normal dist has this pdf:

$$Z \sim N(\mu, \Sigma) \Rightarrow f(z_1, \dots, z_n) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{(z-\mu)^T \Sigma^{-1} (z-\mu)}{2} \right\}$$

What do I mean by covariance matrix?

Suppose I have a random vector

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

then the covariance matrix of X is

$$\text{Var}(X) = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_n) \\ & \text{Var}(X_2) & & \vdots \\ & & \ddots & \text{Cov}(X_{n-1}, X_n) \\ & & & \text{Var}(X_n) \end{bmatrix}$$

↑ always symmetric

$$= E((X-\mu)(X-\mu)^T)$$

Last Time we solved the matrix version of the least squares problem:

$$Q(\beta) = (Y - X\beta)^T (Y - X\beta)$$

$$\frac{\partial Q}{\partial \beta} \stackrel{!}{=} 0 \Rightarrow \hat{\beta} = (X^T X)^{-1} X^T Y$$

↑ LS estimator

Assume $X^T X$
has full
rank

What about fitted values?

$$\begin{aligned}\hat{y} &= X\hat{\beta} \\ &= X \underbrace{(X^T X)^{-1} X^T}_{H} Y \\ &= HY\end{aligned}$$

Define

$$H = X(X^T X)^{-1} X^T$$

as the "hat matrix"

How about residuals?

$$e = y - \hat{y} = y - HY = (I - H)y$$

How can I define SSE or $\hat{\sigma}^2$ in terms of the matrices?

With scalars:

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

with vectors

$$SSE = e^T e = \underline{\|e\|^2}$$

$\|\cdot\|$

$$(e_1 \dots e_n) \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = e_1^2 + e_2^2 + \dots + e_n^2$$

$\hat{\sigma}^2$ is our unbiased est of σ^2

$$\hat{\sigma}^2 = \frac{SSE}{n-p}$$

↖ # of betas

(SLR: $p=2$)

What about the distributions of

$$Y = X\beta + \epsilon$$

fixed, rand
↓ ↓ ↓

i. $Y \sim N(\underset{(n \times 1)}{X\beta}, \underset{(n \times n)}{\sigma^2 I})$

ii. $\hat{\beta} \sim N(\underset{(p \times 1)}{\beta}, \underline{\underline{\sigma^2 (X^T X)^{-1}}})$

iii. $\hat{y} \sim N(X\beta, \sigma^2 H)$
 $\sigma^2 \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum x^2} \right)$

SQR

iv. $e \sim N(0, \sigma^2 (I - H))$

ii. why?

$$\begin{aligned} E(\hat{\beta}) &= E((X^T X)^{-1} X^T Y) = (X^T X)^{-1} X^T E(Y) \\ &= (X^T X)^{-1} \underbrace{X^T X}_{\substack{p \times n \quad n \times p \\ (p \times p)}} \beta = I_p \beta = \beta \end{aligned}$$

$$\begin{aligned} \text{Var}(\hat{\beta}) &= \text{Var}(\underbrace{(X^T X)^{-1}}_{p \times p} \underbrace{X^T}_{p \times n} Y) = \text{Var}(A Y) \\ &= A \text{Var}(Y) A^T \\ &= (X^T X)^{-1} X^T \sigma^2 I (X^T X)^{-1} \\ &\quad \vdots \\ &= \sigma^2 (X^T X)^{-1} \end{aligned}$$

For a constant matrix A & RV Y :

$$\text{Var}(AY) = \underset{p \times n \times n}{A} \underset{p \times n \quad n \times n}{\text{Var}(Y)} \underset{n \times p}{A^T}$$

$$\begin{aligned}
 E(\hat{y}) &= E(HY) = HE(Y) = HX\beta \\
 &= X(\cancel{X^T X})^{-1} \cancel{X^T} X\beta \\
 &= X\beta
 \end{aligned}$$

$$\begin{aligned}
 \text{Var}(\hat{y}) &= \text{Var}(HY) = H \text{Var}(Y) H^T \\
 &= H \cdot \sigma^2 I \cdot H^T \\
 &= \sigma^2 H H^T = \sigma^2 H^2 = \sigma^2 H
 \end{aligned}$$

①

This is actually symmetric

$$\begin{aligned}
 H^T &= (X(X^T X)^{-1} X^T)^T = (X^T)^T ((\underline{X^T X})^{-1})^T X^T \\
 &= X (\underline{(X^T X)^T})^{-1} X^T \\
 &= X (X^T X)^{-1} X^T = H
 \end{aligned}$$

$$\begin{aligned}
 H^2 &= H \cdot H = X(\cancel{X^T X})^{-1} \cancel{X^T} \underline{X(X^T X)^{-1} X^T} \\
 &= X (X^T X)^{-1} X^T = H
 \end{aligned}$$

H is an idempotent matrix
powers

$$H \cdot H = I \cdot H$$

$$AA^{-1} = I$$

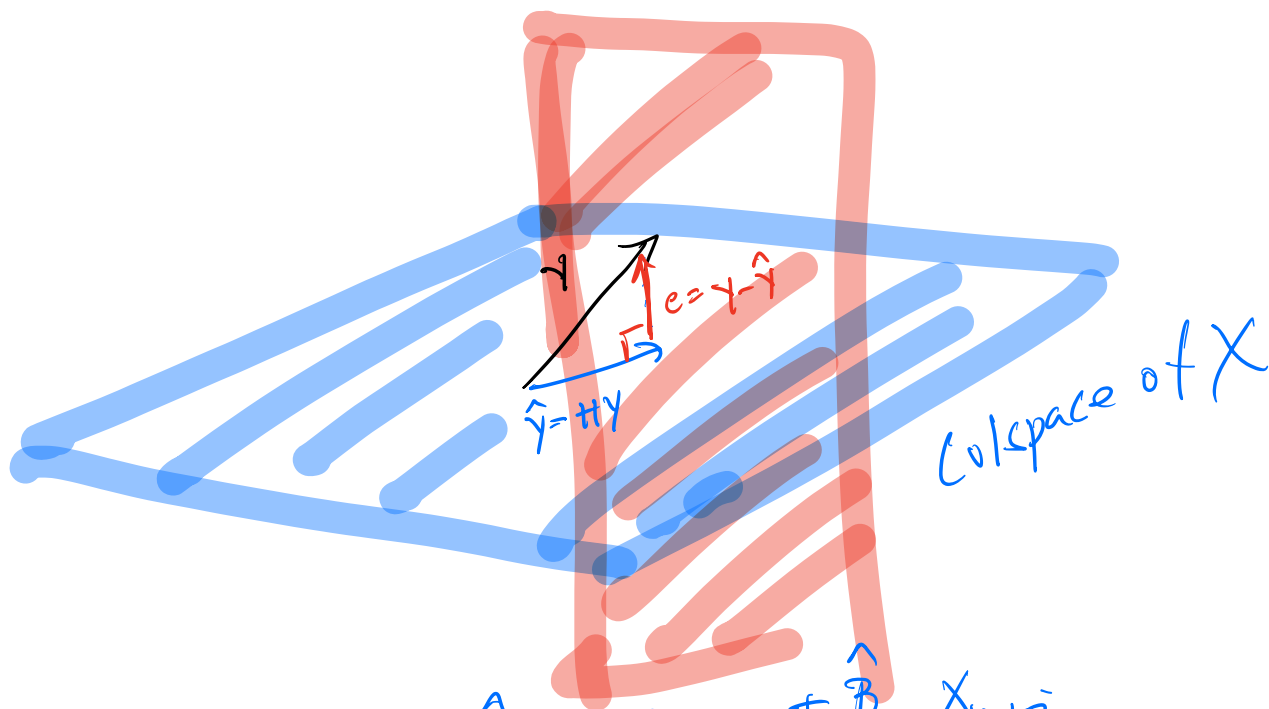
$$(AA^{-1})^T = I$$

$$\Rightarrow (A^{-1})^T A^T = I$$

$$(A^{-1})^T = (A^T)^{-1}$$

H is ^① symmetric & ^② idempotent

$\Leftrightarrow H$ is a projection matrix



$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots + \hat{\beta}_{p-1} x_{p-1i}$$

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

$$\text{Var}(e) = \text{Var}(Y - \hat{Y}) = \text{Var}(Y - HY) \\ = \text{Var}((I - H)Y) \\ = (I - H) \text{Var}(Y) (I - H)^T \\ = \sigma^2 (I - H)(I - H)^T$$

$$= \sigma^2 (I \cdot I^T - HI^T - IH^T + HH^T)$$

$$= \sigma^2 (I - H - \cancel{HI^T} + \cancel{IH^T})$$

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

$I - H$ is also

- ① symmetric
- ② idempotent

$\Rightarrow I - H$ is also a projection matrix

If $V = aH + b(I - H)$

Is V a
proj. matrix?

What is the distribution of SSE?

$$Y \sim N(X\beta, \sigma^2 I)$$

$$e = (I-H)Y \sim N(\underbrace{(I-H)X\beta}_0, \underbrace{(I-H)(\sigma^2 I)(I-H)^T}_{\sigma^2(I-H)})$$

$$(I-H)Y \sim N(0, \sigma^2(I-H))$$

$$\frac{(I-H)Y}{\sigma} \sim N(0, (I-H))$$

$$Y^T(I-H)Y \sim \chi^2_{df} \quad df = \text{rank}(I-H)$$

$$\left\| \frac{(I-H)Y}{\sigma} \right\|^2 = \frac{(Y^T(I-H)Y)}{\sigma^2} = \frac{e^T e}{\sigma^2} = \frac{SSE}{\sigma^2}$$

$$\left\| \frac{(I-H)Y}{\sigma} \right\|^2 = \frac{SSE}{\sigma^2} \sim \chi^2_{\text{rank}(I-H) = n-p}$$

For projection matrices, the rank = the trace.

$$\begin{aligned} \text{rank}(I-H) &= \text{tr}(I-H) = \text{tr}(I) - \text{tr}(H) \\ &= n - \text{tr}(X(X^T X)^{-1} X^T Y) \\ &= n - \text{tr}(X^T X)^{-1} X^T X \end{aligned}$$

$$= n - \text{tr}(I_p) = n - p$$

"Trace is cyclic" \Leftrightarrow $\text{tr}(ABC) = \text{tr}(BCA)$
 $= \text{tr}(CAB)$
 $= \text{tr}(ABC)$

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

\Rightarrow

$$E\left(\frac{SSE}{\sigma^2}\right) = n-p$$

$$\Rightarrow E\left(\frac{SSE}{n-p}\right) = \sigma^2$$

Take MSE as $\frac{1}{\sigma^2} = \frac{SSE}{n-p}$