Multiple Linear Regression

MLR Model:

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

$$i = 1 \ldots n$$

p = # of parameters in the model
(betas)

$p-1$ = # of predictors

In matrix notation, we can write

$$\underset{(n \times 1)}{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{bmatrix} ,$$

$$\underset{(n \times p)}{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{(p-1)1} \\ 1 & X_{12} & X_{22} & & X_{(p-1)2} \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{1n} & X_{2n} & & X_{(p-1)n} \end{bmatrix}$$

1st pred, 2nd pred, last pred.

1st obs, 2nd obs

$$\underset{(p \times 1)}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}$$

$$\mathcal{E} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$
$(n \times 1)$

Last time:

$$Y = X\beta + \mathcal{E}$$

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

Model Assumptions:

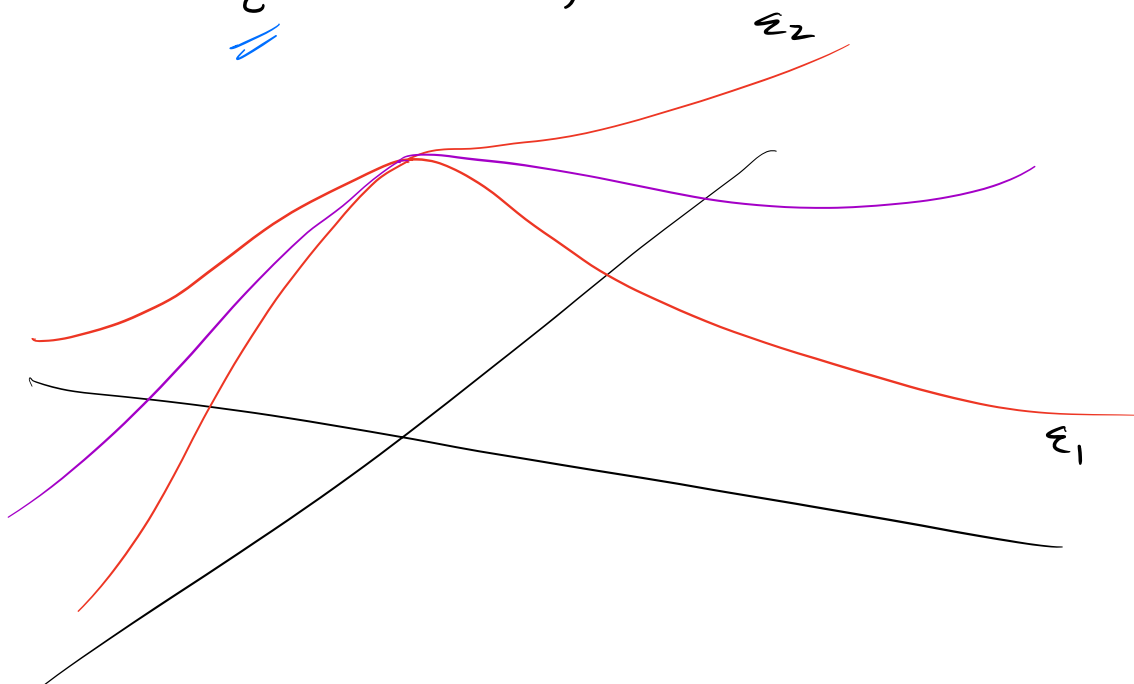① the relationship b/w Y & X_j $j = 1, \dots, p+1$ is linear

② X is fixed

③ $E(\mathcal{E}) = 0$

$Var(\varepsilon_i) = \sigma^2 \quad \forall i$

$\mathcal{E} \sim N(0, \sigma^2 I_n) \quad \leftarrow$ "multivariate Normal distribution"

Multivariate Normal dist has this pdf:

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

⌐ 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

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

↑ always symmetric

$$= E\left((X-\mu)(X-\mu)^T \right) \qquad //$$

Last Time we solved the matrix version of

the least squares problem:

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

⇓

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

↑ LS estimator   Assume $X^T X$ has full rank

What about fitted values?

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

$$= X\,\underbrace{(X^TX)^{-1}X^T}\,Y$$

Define
$$H = X(X^TX)^{-1}X^T$$
as the "hat matrix"

$$= HY$$

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:

$$SSE = \sum_{i=1}^{n} e_i^2 = e^Te = \underline{\|e\|^2}$$

with vectors

$$\|\cdot\|$$

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

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

$$\hat{\sigma}^2 = \frac{SSE}{n-p} \quad \nwarrow \text{# of betas}$$

$$(SLR: p=2)$$

What about the distributions of

fixed, rand
$Y = X\beta + \varepsilon$

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

$\hat{Y} = HY$

ii. $\hat{\beta} \sim N(\beta^{\checkmark}, \sigma^2 (X^TX)^{-1})$

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

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

ii. why?

$E(\hat{\beta}) = E\left((X^TX)^{-1} X^T Y \right) = (X^TX)^{-1} X^T E(Y)$

$\qquad = (X^TX)^{-1} \underset{\underset{(p \times p)}{\underbrace{\underset{p \times n \;\; n \times p}{X^T X}}}}{X^T X} \beta = I_p \beta = \beta$

$Var(\hat{\beta}) = Var(\overset{A}{\overbrace{(X^TX)^{-1}}} \underset{p \times n}{X^T} Y) = Var(AY)$

$\qquad \qquad \underset{p \times p}{} \qquad = A\, Var(Y) A^T$

$\qquad \qquad = (X^TX)^{-1} X^T \sigma^2 I \left((X^TX)^{-1} X^T \right)^T$

$\qquad \qquad \vdots$

$\qquad \qquad = \sigma^2 (X^TX)^{-1}$

$\lceil$ For a constant matrix A & RV Y:

$\qquad Var(AY) = \underset{p \times n \;\; n \times 1}{A} \underset{p \times n \;\; n \times n \;\; n \times p}{Var(Y)\, A^T} \qquad \rfloor$

$$E(\hat{y}) = E(Hy) = H E(y) = H X \beta$$

$$= X \cancel{(X^TX)^{-1} X^T X} \beta$$

$$= X\beta$$

$$Var(\hat{y}) = Var(Hy) = H \, Var(y) \, H^T$$

$$= H \cdot \sigma^2 I \cdot H^T$$

$$= \sigma^2 H H^T = \sigma^2 H^2 = \sigma^2 H$$

① this actually symmetric

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

$$= X\left((X^TX)^T\right)^{-1} X^T$$

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

$$H^2 = H \cdot H = X\cancel{(X^TX)^{-1}X^T} \, \underline{\underline{X}}(X^TX)^{-1}X^T$$

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

② H is an $\underline{\underline{idempotent}}$ matrix $H \cdot H = I \cdot H$
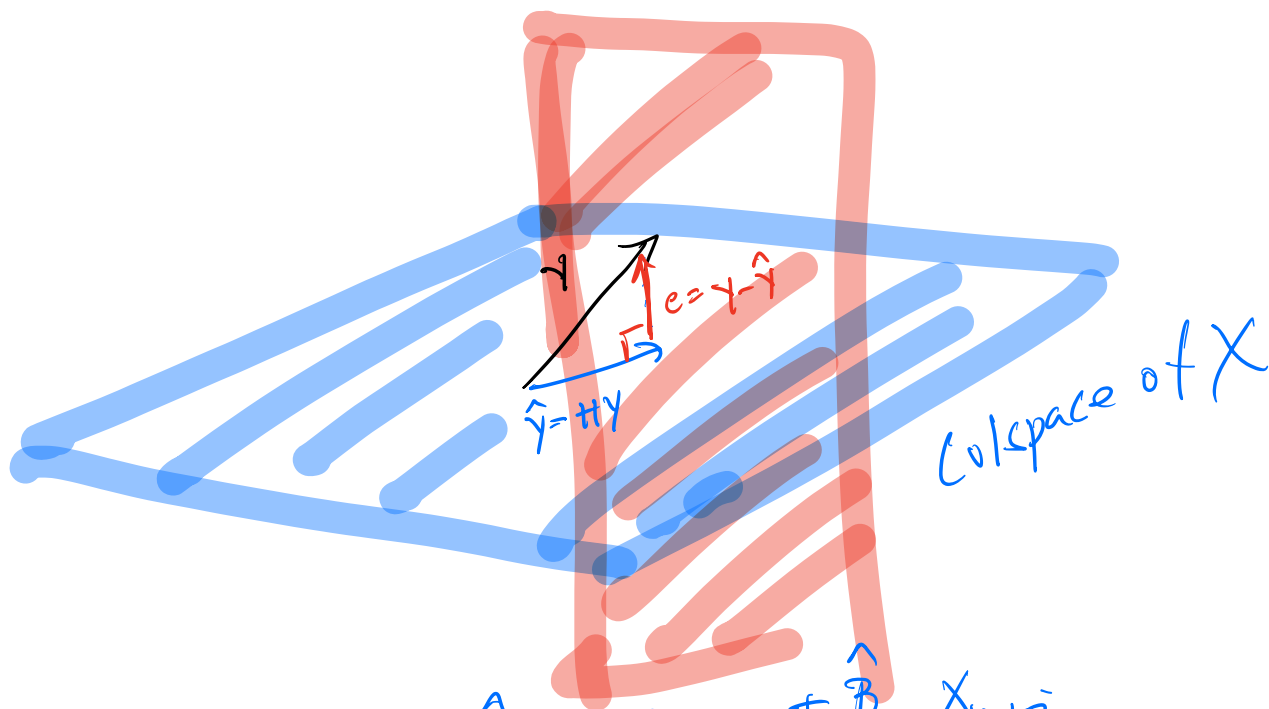   powers

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

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

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

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

H is symmetric [1] & idempotent [2]

$$\Leftrightarrow H \text{ is a } \underline{\text{projection}} \text{ matrix}$$

Colspace of X

$\hat{y} = HY$

γ

$e = y - \hat{y}$

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

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

$$Var(e) = Var(y - \hat{y}) = Var(y - Hy)$$
$$= Var((I-H)y)$$
$$= (I-H) Var(y) (I-H)^T$$
$$= \sigma^2 (I-H)(I-H)^T$$
$$= \sigma^2 (I \cdot I^T - HI^T - IH^T + HH^T)$$
$$= \sigma^2 (I - H - \cancel{H + H})$$
$$= \sigma^2 (I - H)$$

$I - H$ is also ① symmetric $\Rightarrow$ $I-H$ is also a projection matrix
② idempotent

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

Is V a proj. matrix?

What is the distribution of SSE?

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

$$e = (I-H)Y \sim N(\underbrace{(I-H)X\beta}_{0}, \underbrace{(I-H)(\sigma^2 I)(I-H)^T}_{\sigma^2(I-H)})$$

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

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

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

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

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

For projection matrices, the rank = the trace.

$$rank(I-H) = tr(I-H) = tr(I) - tr(H)$$

$$= n - tr(\underline{X}(X^TX)^{-1}X^T)$$

$$= n - tr((X^TX)^{-1}X^TX)$$

$$= n - tr(I_p) = n - p$$

"Trace is cyclic" $\iff$ $tr(ABC) = tr(BCA)$
$$= tr(CAB)$$
$$= tr(ABC)$$

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

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

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

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