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

$$\underset{(n\times 1)}{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix},$$

$$\underset{(n\times p)}{X} = \begin{array}{rc} & \begin{array}{ccccc} & \text{1st pred} & \text{2nd pred} & \cdots & \text{last pred.} \end{array} \\ \begin{array}{r} \text{1st obs} \rightarrow \\ \text{2nd obs} \rightarrow \\ \\ \\ \end{array} & \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{(p-1)1} \\ 1 & X_{12} & X_{22} & & X_{(p-1)2} \\ \vdots & & \vdots & & \vdots \\ 1 & X_{1n} & X_{2n} & & X_{(p-1)n} \end{bmatrix} \end{array}$$

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

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

Last time:

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

Model Assumptions:

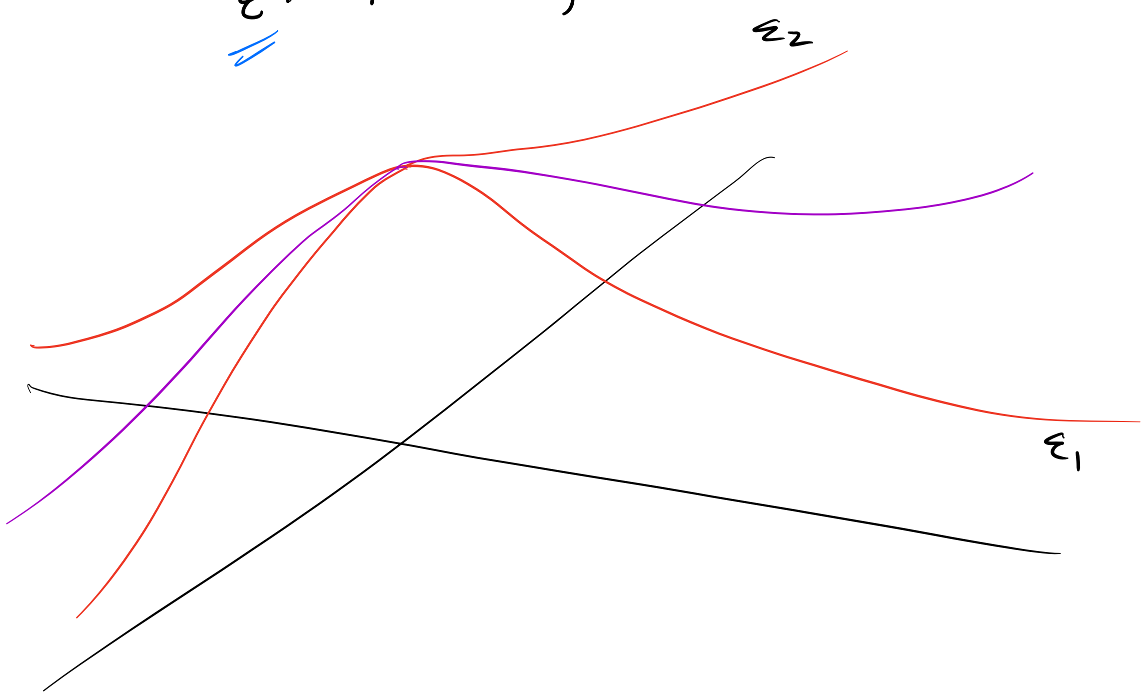
① the relationship b/w y & x_j $j=1,\dots,p-1$ is linear

② X is fixed

③ $E(\varepsilon)=0$

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

$\varepsilon \sim N(0, \sigma^2 I_n)$ "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

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

always symmetric

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

//

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

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

$$\Downarrow$$

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

LS estimator

Assume $X^T X$ has full rank

What about fitted values?

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

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

$$= HY$$

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

with vectors

$\|\cdot\|$

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

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

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

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

# of betas

(SLR: $p=2$)

What about the distributions of

$$Y = \overset{\text{fixed}}{X}\overset{\text{fixed}}{\beta} + \overset{\text{rand}}{\varepsilon}$$

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

SLR

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

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

ii. why?

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

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

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

$$= A\,Var(Y)A^T$$

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

$$\vdots$$

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

For a constant matrix A & RV Y:

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

$$E(\hat{Y}) = E(HY) = HE(Y) = HX\beta$$
$$= X\cancel{(X^TX)^{-1}X^TX}\beta$$
$$= X\beta$$

$$Var(\hat{Y}) = Var(HY) = H\,Var(Y)H^T$$
$$= H \cdot \sigma^2 I \cdot H^T$$
$$= \sigma^2 H H^T = \sigma^2 H^2 = \sigma^2 H$$

① H is actually symmetric

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

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

② H is an idempotent matrix

powers

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

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

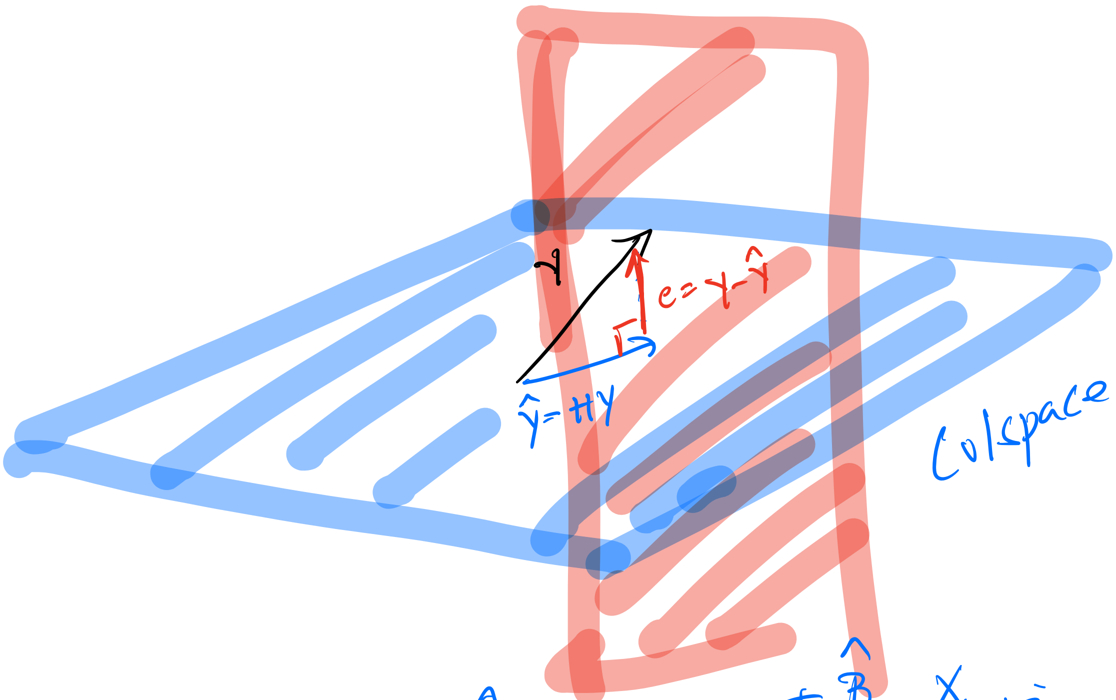
$$(AA^{-1})^T = I$$

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

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

H is ① symmetric & ② idempotent

$\iff$ H is a <u>projection</u> matrix



Colspace of X

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

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

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

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

$$\boxed{\text{If } V = aH + b(I-H) \quad \text{is } V \text{ a proj. matrix?}}$$

What is the distribution of SSE?

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

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

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

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

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

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

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

For projection matrices, the rank = the trace.

$$\text{rank}(I-H) = tr(I-H) = tr(I) - tr(H)$$

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

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

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

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

$$= tr(CAB)$$

$$= tr(ABC)$$

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

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

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

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