Multiple Linear Regression

p = # of parameters
# of betas

MLR Model:

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

for $i = 1, \ldots, n$

Assumptions

1. Y has a linear relationship w/ $\{X_1, X_2, \ldots, X_{p-1}\}$
2. The X_j's are fixed, $j = 1, \ldots, p-1$
3. $E(\varepsilon_i) = 0$
   $var(\varepsilon_i) = \sigma^2$
   $Cor(\varepsilon_i, \varepsilon_j) = 0$ if $i \neq j$ $\longrightarrow$ $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

The system of equations defined above can be compactly expressed as:

$$Y = X\beta + \varepsilon$$ where

cols = variables

$\downarrow$

$$\underset{(n\times 1)}{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \underset{n\times p}{X} = \begin{pmatrix} 1 & X_{11} & X_{21} & \cdots & X_{(p-1)1} \\ 1 & X_{12} & X_{22} & & X_{(p-1)2} \\ 1 & \vdots & \vdots & & \vdots \\ \vdots & & & & \\ 1 & X_{1n} & X_{2n} & & X_{(p-1)n} \end{pmatrix} \xleftarrow{} \text{obs}$$

$$\underset{(p\times 1)}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}, \quad \text{and} \quad \underset{(n\times 1)}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Model Assumptions

① No change ↑ same as above

② X is fixed

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \Rightarrow \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

③ $E(\underset{(n\times 1)}{\varepsilon}) = \underset{(n\times 1)}{0}$

$var(\underset{(n\times 1)}{\varepsilon}) = \sigma^2 \underset{(n\times n)}{I_n}$

Classical

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

What is a covariance matrix?

For example, if I have a random vector $\underline{\underline{X}}$ with mean vector μ, we say its covariance matrix is defined as:

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

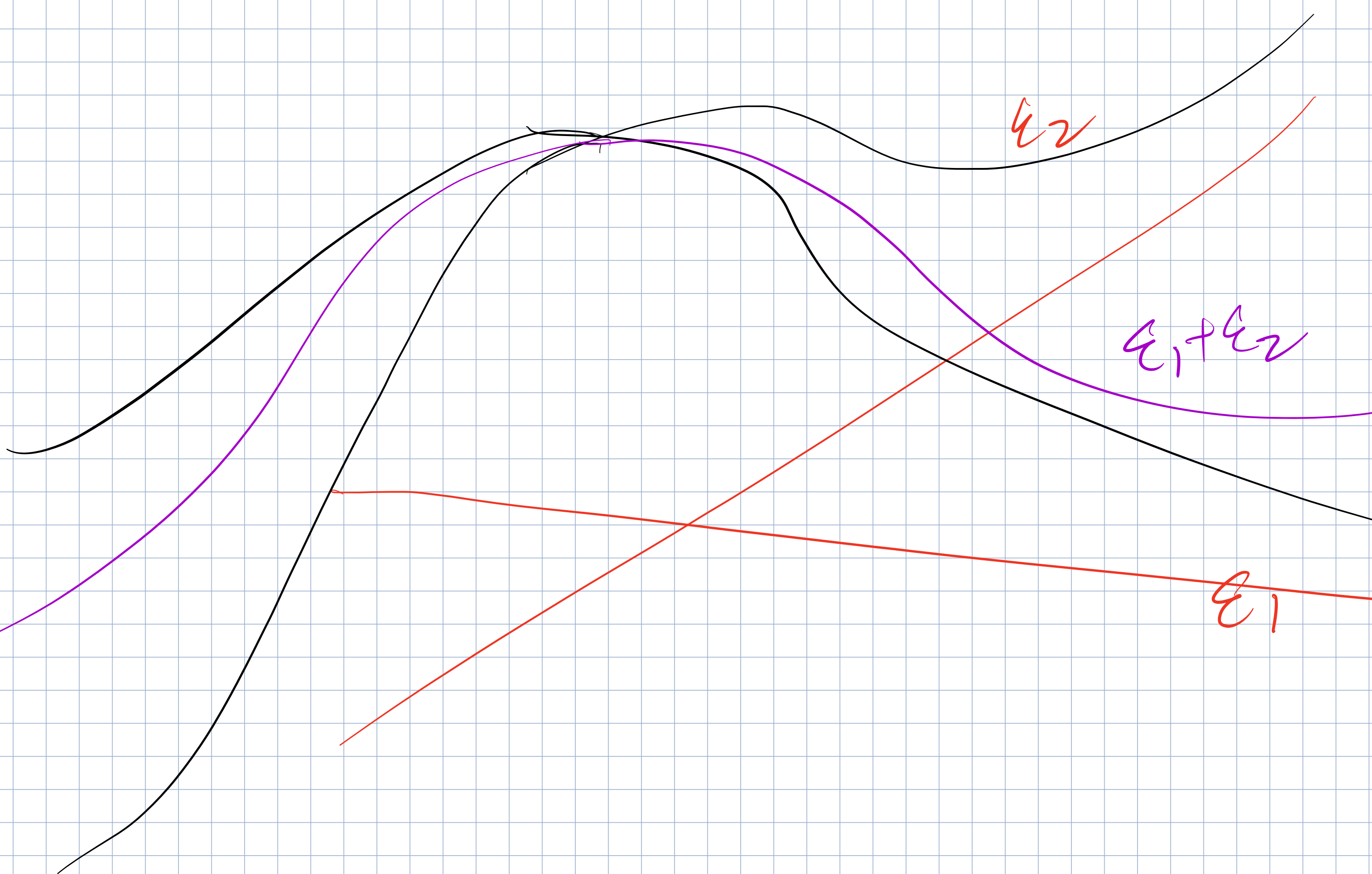
$$= \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) & Cov(X_1, X_3) & \cdots & Cov(X_1, X_n) \\ & Var(X_2) & & & \vdots \\ & & \ddots & & \vdots \\ & & & \ddots & Cov(X_{n-1}, X_n) \\ & & & & Var(X_n) \end{bmatrix}$$

the covariance matrix is always symmetric

Multivariate Normal Dist

We say a variable Z has a multivariate Normal dist w/ mean μ & variance Σ if its pdf:

$$f(z_1, \ldots, z_n) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left\{ \frac{-(x-\mu)^T \Sigma^{-1} (x-\mu)}{2} \right\}$$



Last time we solved the least squares problem in matrix form:

$$Q(\beta) = \|Y - X\beta\|^2 = (Y - X\beta)^T (Y - X\beta)$$

has the minimizer

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

if $X^T X$ is invertible
$\Updownarrow$
X has full rank

Fitted Values

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

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

$$= HY$$

where $H = X(X^T X)^{-1} X^T$ is the "hat matrix"

Residuals

$e_i = y_i - \hat{y}_i$

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

How can I express SSE in terms of the matrix form?

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = (Y - \hat{Y})^T (Y - \hat{Y}) = e^T e$$
$$= \|Y - \hat{Y}\|^2 = \|e\|^2$$
$$= \|(I - H)Y\|^2$$

$$\sum_{i=1}^{n} e_i^2 = e_1^2 + e_2^2 + e_3^2 + \cdots + e_n^2 =$$
$$(e_1\ e_2 \cdots e_n)\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} = e^T e = \|e\|^2$$

What are the distributions of:

i. $Y \sim N(X\beta, \sigma^2 I_n)$

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

iii. $\hat{Y} \sim N(X\beta, \sigma^2 H)$

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

i. $Y = X\beta + \varepsilon$

$$E(Y) = E(X\beta + \varepsilon) = X\beta + E(\varepsilon) = X\beta$$

$$Var(Y) = Var(X\beta + \varepsilon) = Var(\varepsilon) = \sigma^2 I_n$$

ii. $\hat{\beta} = (X^TX)^{-1}X^TY$

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

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

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

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

$\vdots$

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

iii. $\hat{Y}$

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

$$E(\hat{Y}) = E(HY) = HE(Y) = HX\beta =$$

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

$$= X\beta$$

$$Var(\hat{Y}) = Var(HY) = H\, Var(Y) H^T$$

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

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

What is H^T? $\quad (ABC)^T = C^T B^T A^T$

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

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

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

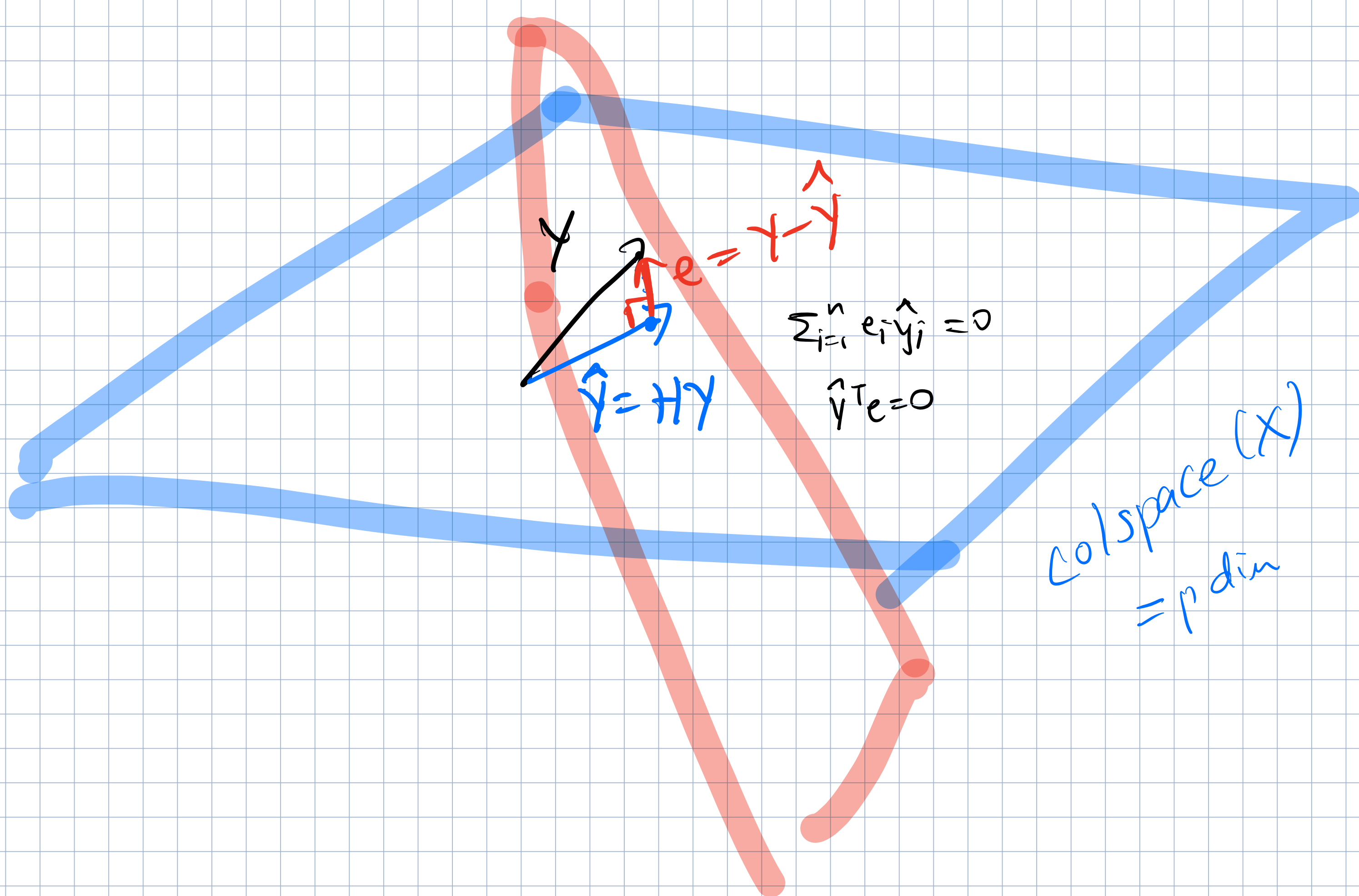
what's H^2?

$$\underbrace{X(X^TX)^{-1}X^T}_{H} \cdot \underbrace{X(X^TX)^{-1}X^T}_{H} = X(X^TX)^{-1}X^T = H$$

If $H^2 = H$ then H is called "idempotent"

If a matrix is ① symmetric &
② idempotent then

H is a "projection matrix".



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

$\sum_i x_i e_i = 0$

$e^T x_1 = 0$

$$
\begin{aligned}
Var(e) = Var(Y - \hat{Y}) &= Var((I-H)Y) \\
&= (I-H)\,Var(Y)\,(I-H)^T \\
&= (I-H)\,\sigma^2 I\,(I-H) \\
&= \sigma^2 (I-H)(I-H) \\
&= \sigma^2 (I - HI - IH + H^2) \\
&= \sigma^2 (I - H - \cancel{H + H}) \\
&= \sigma^2 (I-H)
\end{aligned}
$$

Notice $I-H$ is ① sym.

② idempotent

& so it's also a projection matrix

What's the dist of SSE?

$X\beta - HX\beta = X\beta - X\beta = 0$

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

$$e = (I-H)Y \sim N\left((I-H)X\beta, (I-H)\sigma^2 I (I-H)^T\right)$$

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

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

SSE
$\shortparallel$
e^Te
$\shortparallel$

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

[You can use the fact that if

Y is Normal then

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

For projection matrices, the rank = the trace.

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

$tr(ABC) =$
$tr(BCA) =$
$tr(CAB) =$
~~$tr(ABC)$~~
"cyclic"

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

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

$$SSE/\sigma^2$$

$$\|$$

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

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

$\Rightarrow$ An unbiased est for σ^2 is:

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

$$E\left(\frac{SSE}{n-p}\right) = \frac{1}{n-p} \sigma^2 E\left(\frac{SSE}{\sigma^2}\right)$$

$$= \frac{\sigma^2}{\cancel{n-p}} \cancel{(n-p)} = \sigma^2$$