# Review

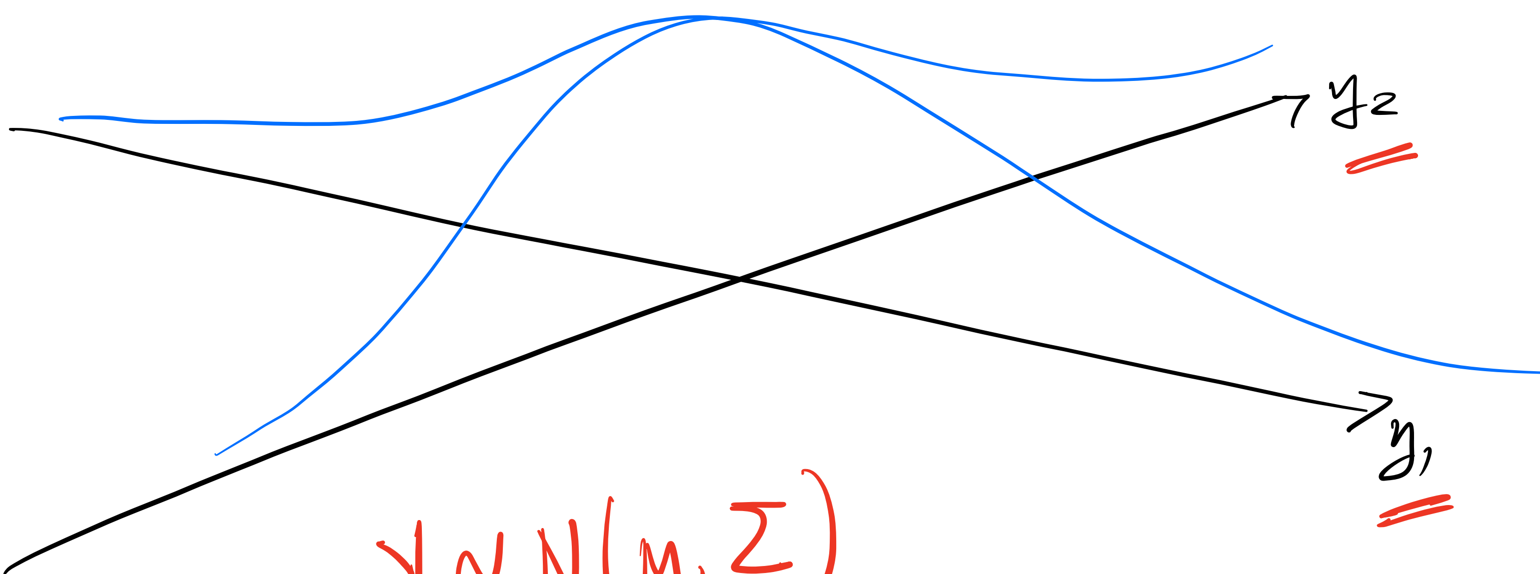
## Multivariate Normal Distribution

Def: A random $n\times 1$ vector $Y$ is said to have a multivariate Normal dist w/ mean $\underset{(n\times1)}{\mu}$ & covariance matrix $\underset{(n\times n)}{\Sigma}$ if

assume full rank

it has the following pdf:

$$f(y) = \frac{1}{(2\pi)^{n/2}\,|\Sigma|^{1/2}}\, e^{-\underset{1\times n}{(y-\mu)^T}\,\underset{n\times n}{\Sigma^{-1}}\,\underset{n\times 1}{(y-\mu)}/2}.$$




$$Y \sim N(\mu, \Sigma)$$

What is a covariance matrix?

$$Var(Y) = \Sigma = \begin{bmatrix} Var(Y_1) & Cov(Y_1,Y_2) & \cdots & Cov(Y_1,Y_n) \\ Cov(Y_1,Y_2) & Var(Y_2) & & \vdots \\ \vdots & & \ddots & Cov(Y_{n-1},Y_n) \\ Cov(Y_1,Y_n) & \cdots & & Var(Y_n) \end{bmatrix}$$

Notice $\Sigma$ is symmetric

Linear combo:

$$a_1Y_1 + a_2Y_2 + \cdots + a_nY_n$$

Let:

$$\underset{1\times n}{A} = [a_1 \ a_2 \ \cdots \ a_n]$$

$$\Rightarrow AY = a_1Y_1 + a_2Y_2 + \cdots + a_nY_n \sim N(\overset{1\times n\, n\times 1}{A\mu}, \overset{1\times n\, n\times n\, n\times 1}{A\Sigma A^T})$$

Properties:

① $\forall$ constant matrix $A$ we have
$(m \times n)$

$$\underbrace{A}_{m\times n}\underbrace{y}_{n\times 1} \sim N(\underbrace{\underset{m\times n}{A}\mu}_{\text{new mean}}, \underbrace{\underset{m\times n}{A}\underset{n\times n}{\Sigma}\underset{n\times m}{A^T}}_{\text{new variance.}})$$

② Any linear combination of the elements of $y$ are also Normally distributed.

(Why does this follow immediately from ①?)

③ If we have 2 MVN vectors $\underset{m\times 1}{y_1}$ & $\underset{n\times 1}{y_2}$ such that,

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$$

where $E(y_1)=\mu_1, E(y_2)=\mu_2$ $\qquad \Sigma_{12}=\Sigma_{21}^T$

$Var(y_1)=\Sigma_{11}, Var(y_2)=\Sigma_{22}$

& $Cov(y_1, y_2)=\Sigma_{12}$,

then the conditional distribution of

$y_1$ given $y_2$ is:

$$y_1|y_2 \sim N\left(\mu_1+\Sigma_{12}\Sigma_{22}^{-1}(y_2-\mu_2), \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

④ If $Y \sim N(\underline{\underline{0}}, \underline{\underline{I_n}})$ then

$$\|Y\|^2=\sum_{i=1}^{n} y_i^2 \sim \chi^2_n.$$

If $Y \sim N(\underline{0}, \underline{I_n})$ &

A is an $\underline{n\times n}$ matrix that is:

① $\underline{\text{symmetric}}$ & $\qquad A=A^T$

② $\underline{\text{idempotent}}$, then $\qquad A^2=A$

projection matrix

$Y^TAY \sim \chi^2_{df}$ where $\underline{df=rank(A)}$.

# Recap MLR model:

Responses: $Y_{(n\times 1)}$

Predictors: $X_{(n\times p)}$ $= \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{(p-1)1} \\ \vdots & \vdots & & & \vdots \\ 1 & x_{1n} & \cdots & \cdots & x_{(p-1)n} \end{bmatrix}$

(*) For now, assume $X$ has full rank & $n>p \Rightarrow \operatorname{rank}(X)=p$.

Coefficients: $\beta_{(p\times 1)}$

Errors: $\varepsilon_{(n\times 1)}$

Model:

$$y = \underbrace{X\beta}_{\text{const.}} + \underbrace{\varepsilon}_{\text{random}} \quad \text{with} \quad \varepsilon \sim N(0, \sigma^2 I_n)$$

$X^TY \approx \operatorname{Cov}(X,Y)$

$X^TX \approx \operatorname{Var}(X)$

## Least Squares Solution:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}}\, Q(\beta) = \underset{\beta}{\operatorname{argmin}}\, (Y-X\beta)^T(Y-X\beta)$$

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

(if $X^TX$ has full rank)

# Other Regression Quantities

hat matrix

Fitted Values:

$$\hat{Y} = X\hat{\beta} = \overset{H_n}{\underbrace{X(X^TX)^{-1}X^T}}\,Y = HY$$

Residuals:

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

Sum of Squared Error:

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

---

## Useful Properties

1. $H$ is symmetric
2. $H$ is idempotent.

$\therefore$ 3. $H$ is a projection matrix

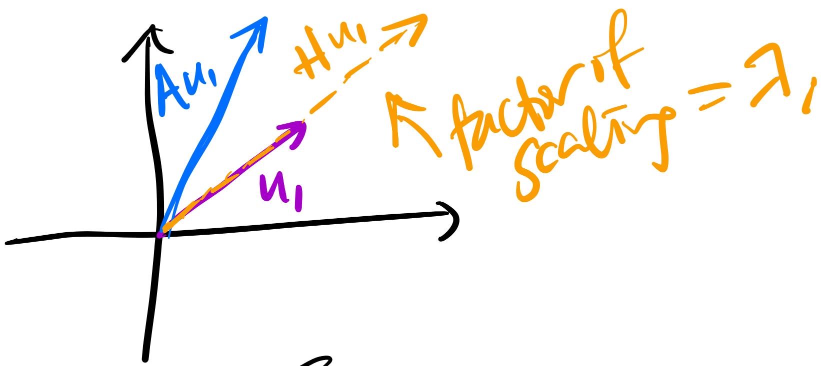
(1-3) also hold for $I-H$.

Exercise:
Show 1-3 for
$H$ & $I-H$.

Eigenvectors & Eigenvalues:

If $u_1$ is an eigenvector of $H$ w/ eigenvalue $\lambda_1$: then

$$Hu_1 = \lambda_1 u_1$$




$$H = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

$$U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$H = U \Lambda U^T$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^T$$

$$H = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad (H - \lambda I)x = 0$$

$$Hx = \lambda x = \lambda I x$$

# Decomposing H:

Let H have the eigendecomposition:

$\boxed{\underset{n\times n}{H} = U \Lambda U^T}$ where $\underset{n\times n}{U}$ is the matrix of eigenvectors of H, $U = \begin{bmatrix} - u_1 - \\ - u_2 - \\ \vdots \\ - u_n - \end{bmatrix}$,

and $\Lambda$ is a diagonal matrix with H's eigenvalues on the diagonal, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$.

The matrix $U$ is "orthogonal", i.e.

$\boxed{U^T U = I}$ ($\therefore$ $U^{-1} = U^T$).

$\boxed{\text{Use } H^2 = H}$

**Claims** for eigendecomp of H:

① $\lambda_i = 0$ or $1$ always

② $\text{rank}(H) = \text{trace}(H) = p$

Why are these true for H?

Pf:

① $H^2 = (U\Lambda U^T)^2$

$$= (U\Lambda \underbrace{U^T\ U}\Lambda U^T)$$

$$= U\Lambda I \Lambda U^T$$

$$= U\Lambda^2 U^T$$

Since $H^2 = H$

$$= U\Lambda U^T$$

$$U\Lambda^2 U^T = U\Lambda U^T$$

$$\cancel{U^T U}\Lambda^2 \cancel{U^T U} = \cancel{U^T U}\Lambda \cancel{U^T U}$$

$$\Lambda^2 = \Lambda$$

$$\begin{bmatrix} \lambda_1^2 & \cdots & & 0 \\ 0 & \lambda_2^2 & & \vdots \\ 0 & & \ddots & 0 \\ \vdots & & & \\ 0 & & & \lambda_p^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ \vdots & \lambda_2 & & \vdots \\ & & \ddots & 0 \\ 0 & \cdots & & \lambda_n \end{bmatrix}$$

$\boxed{\lambda_i^2 = \lambda_i}$

$\lambda_i^2 - \lambda_i = 0$

$\lambda_i(1-\lambda_i) = 0$

$\lambda_i = 0$ or

$1-\lambda_i = 0$

②

$$H = U \Lambda U^T$$

$$= \begin{bmatrix} \text{---}\, u_1 \,\text{---} \\ \text{---}\, u_2 \,\text{---} \\ \vdots \\ \text{---}\, u_n \,\text{---} \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{bmatrix}$$

<u>Defs</u>:

<u>rank</u> = # of nonzero eigenvalues

<u>trace</u> = sum of diagonal elements

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & \\ \vdots & & \ddots & \\ a_{n1} & & & a_{nn} \end{bmatrix} \Rightarrow tr(A) = \sum_{i=1}^{n} a_{ii}$$

trace is a cyclic operator:

$$tr(ABC) = tr(BCA) = tr(CAB)$$

$$tr(H) = tr(U\Lambda U^T) = tr(\Lambda U^T U)$$

$$= tr(\Lambda)$$

$$= \underbrace{1+1+\cdots+1}_{\#\text{ of nonzeros}} + 0 \cdots + 0. = rank(H)$$

Why is $tr(H) = p$?

$$tr(H) = tr(X(X^TX)^{-1}X^T) = tr(\underbrace{X^T X}_{p\times n\ n\times p}(X^TX)^{-1})$$

$$= tr(I_p) = p$$

Pf (concise version written before class)

① b/c $H$ is idempotent,

$$\Rightarrow \quad H^2 = U\Lambda U^T U \Lambda U^T$$

$$\Rightarrow \quad = U \Lambda I \Lambda U^T$$

$$\Rightarrow \quad = U \Lambda^2 U^T = U \Lambda U^T = H$$

which implies $\Lambda^2 = \Lambda$.

So $\lambda_i^2 = \lambda_i \;\; \forall \; i = 1, \dots, n. \rightarrow \lambda_i = 0 \text{ or } 1.$

② $tr(H) = tr(X(X^TX)^{-1}X^T)$

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

$= tr(I_p) = p$

Why is $tr(H) = rank(H)$?

rank is # of nonzero eigenvalues

same as trace if they're all 0 or 1!

How is this useful?

Consider estimating $\sigma^2$...

We know
$$Y \sim N(X\beta, \sigma^2 I_n)$$

$$\Rightarrow Y - X\beta \sim N(0, \sigma^2 I_n)$$

$(I-H)X\beta = \ldots$?

$$\Rightarrow \frac{Y-X\beta}{\sigma} \sim N(0, I_n)$$

why? consider

$$\Rightarrow \frac{(Y-X\beta)^T(I-H)(Y-X\beta)}{\sigma^2} = \frac{Y^T(I-H)Y}{\sigma^2} = \frac{SSE}{\sigma^2}$$

by ④

$$\Rightarrow \frac{SSE}{\sigma^2} = \frac{Y^T(I-H)Y}{\sigma^2} \sim \chi^2_{\text{rank}(I-H)}$$

$$\begin{aligned}\text{rank}(I-H) &= \text{tr}(I-H) \\ &= \text{tr}(U(I-\Lambda)U^T) \\ &= \text{tr}((I-\Lambda)U^TU) \\ &= \text{tr}\left(\begin{bmatrix} 0 & 0 \\ 0 & I_{n-p}\end{bmatrix}\right) = n-p.\end{aligned}$$

Why does this rank of a projection matrix idea work???

Notice: If $H = U \Lambda U^T$, then

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

$$U^T(I-H)U = U^T U - U^T H U$$

$$= U^T(I-\Lambda)U$$

$$= U^T \begin{bmatrix} 0 & 0 \\ 0 & I_{n-p} \end{bmatrix} U$$

So $\frac{U^T e}{\sigma} \sim N\left(0, U^T \begin{bmatrix} 0 & 0 \\ 0 & I_{n-p} \end{bmatrix} U\right)$

The first $p$ entries are $N(0,0)$...

deterministically 0!

$$\left\|\frac{U^T e}{\sigma}\right\|^2 = \overbrace{\frac{0^2}{\sigma^2} + \frac{0^2}{\sigma^2} + \cdots + \frac{0^2}{\sigma^2}}^{p \text{ of these}} + \overbrace{\frac{(U^T e)_{p+1}^2}{\sigma^2} + \cdots + \frac{(U^T e)_n^2}{\sigma^2}}^{(n-p) \text{ std normals!}}$$

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

Finally notice $\frac{SSE}{\sigma^2} = \frac{e^T e}{\sigma^2} = \frac{e^T U U^T e}{\sigma^2} = \left\|\frac{U^T e}{\sigma}\right\|^2$

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

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

$\hat{\sigma}^2 = MSE$!

---

<u>Claim</u>: The fitted values & the residuals are independent.

i.e. $\hat{Y} = X\hat{\beta} \perp\!\!\!\perp Y - X\hat{\beta} = (I-H)Y$

Pf:

$$H = U \Lambda U^T = U \begin{bmatrix} I_p & 0 \\ 0 & 0 \end{bmatrix} U^T$$

Let $U = [U_1 \; U_2]$ so then
$\quad\quad (n \times p) \quad n \times (n-p)$

$$H = [U_1, U_2] \begin{bmatrix} I_p & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} = U_1 U_1^T$$

Similarly,

$$I - H = [U_1, U_2] \begin{bmatrix} 0 & 0 \\ 0 & I_{n-p} \end{bmatrix} \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} = U_2 U_2^T$$

$\therefore$

$$U^T Y \sim N(U^T X \beta, \sigma^2 I_n)$$

$U^T = \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix}$ so

$$U^T Y = \begin{bmatrix} U_1^T Y \\ U_2^T Y \end{bmatrix} \sim N\left( \begin{pmatrix} U_1^T X \beta \\ U_2^T X \beta \end{pmatrix}, \sigma^2 \begin{bmatrix} I_p & 0 \\ 0 & I_{n-p} \end{bmatrix} \right)$$

$\Rightarrow U_1^T Y \perp\!\!\!\perp U_2^T Y$ b/c joint normal & cov 0.

$$U_1^T Y \perp\!\!\!\perp U_2^T Y$$

$$\Rightarrow U_1 U_1^T Y \perp\!\!\!\perp U_2 U_2^T Y$$

$$\Rightarrow HY \perp\!\!\!\perp (I-H)Y$$

$$\Rightarrow \hat{Y} \perp\!\!\!\perp e$$

This allows us to do inference on $\beta$!