THM 1: $A, M \in S^{n\times n}$, $M \succ 0$, $\exists\, C$, $\det|C| \neq 0$ s.t.

- $C^T M C = I$
- $C^T A C = \Lambda$ (diagonal)

Pf: $M = M^{\frac{1}{2}} M^{\frac{1}{2}}$, & $B \triangleq (M^{-\frac{1}{2}})^T A (M^{-\frac{1}{2}})$

$\Rightarrow \exists\, P_{n\times n}$ $PP^T = I$ s.t.

$P^T B P = \Lambda$ or $(M^{-\frac{1}{2}}P)^T A (M^{-\frac{1}{2}}P) = \Lambda$

& $(M^{-\frac{1}{2}}P)^T M (M^{-\frac{1}{2}}P) = P^T P = PP^T = I$

$\Rightarrow C = M^{-\frac{1}{2}} P$. Also, $B \succeq 0$ if $A \succeq 0$.

THM 2: $\underset{n\times p}{X} = \underset{n\times r}{S}\, \underset{r\times r}{D}\, \underset{r\times p}{T}^T$ (g inverse)

Let $X^- = \underset{p\times r}{T}\, \underset{r\times r}{D_r^{-1}}\, \underset{r\times n}{S^T}$, then

$X X^- X = S D_r T^T T D_r^{-1} S^T S D_r T^T$

$= S D_r T^T = X$.

But there are infinitely many g-inverses:

$\tilde{X} = X^- + (I - X^- X) B$

Then $X \tilde{X} X = X X^- X + X(I - X^- X) B X$

$= X X^- X = X$.

THM 3: Application of X^- (sol. of linear system)

$$Ax = b$$

A sol. x^* exists iff it is a consistent set of eqs. i.e.,

$\text{rank}(A|b) = \text{rank}(A)$

& $x^* = A^- b$ is a sol.

Pf: $Ax^* = AA^- b = AA^- A x = Ax = b$ □

THM 4. Partitioned matrices

$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ $B = \begin{pmatrix} B_1 & B_2 \\ B_{21} & B_{22} \end{pmatrix}$

Then (i): $\left| \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix} \right| = |A_{11}||A_{22}|$

(verify $\left|\begin{pmatrix} I & 0 \\ 0 & A_{22} \end{pmatrix}\right| = |A_{22}|$)

(Then $\left|\begin{pmatrix} A_{11} & A_{12} \\ 0 & I \end{pmatrix}\right| = |A_{11}|$)

(ii): $|I + AB| = |I + BA|$

$\begin{pmatrix} I & A \\ -B & I \end{pmatrix}\begin{pmatrix} I & 0 \\ B & I \end{pmatrix} = \begin{pmatrix} I+AB & A \\ 0 & I \end{pmatrix}$; $\begin{pmatrix} I & 0 \\ B & I \end{pmatrix}\begin{pmatrix} I & A \\ -B & I \end{pmatrix} = \begin{pmatrix} I & A \\ 0 & I+BA \end{pmatrix}$

THM 5: $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ $A^{-1} = B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$

$\Rightarrow \begin{cases} B_{11} = A_{11\cdot 2}^{-1} = (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1} \\ B_{21} = -A_{22}^{-1} A_{21} A_{11\cdot 2}^{-1} \end{cases}$

$(I + BA)^{-1} = I - B(I + AB)^{-1} A$

($(A+BCD)^{-1}$ in HW3)

Multi-variate Normal

THM 6: MGF of $Y \sim \mathcal{N}_p(\mu, \Sigma)$ is

$$\psi_Y(t) = \mathbb{E} e^{t^T y} = e^{\mu^T t + \frac{1}{2} t^T \Sigma t}$$

Prop 7: $Y \sim \mathcal{N}_p(\mu, \Sigma)$ iff

$$a^T Y \sim \mathcal{N}_1(a^T \mu, a^T \Sigma a)$$

Pf: Cramér-wold's device.

Prop 8: $Y \sim \mathcal{N}_p(\mu, \Sigma)$, Then $AY \perp\!\!\!\perp BY$ iff $A \Sigma B^T = 0$.

Pf: $W = \begin{pmatrix} A \\ B \end{pmatrix} Y \sim \mathcal{N}\left(\begin{pmatrix} A \\ B \end{pmatrix} \mu, \begin{pmatrix} A\Sigma A^T & A\Sigma B^T \\ B\Sigma A^T & B\Sigma B^T \end{pmatrix} \right)$

THM 9: $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathcal{N}_p\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$

Let $\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$. then

$$\mathrm{Cov}\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} I & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I \end{pmatrix}$$

$$= \begin{pmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}$$

$$\Rightarrow (Y_1 - \Sigma_{12}\Sigma_{22}^{-1} Y_2) \perp\!\!\!\perp Y_2$$

$$\Rightarrow Y_1 - \Sigma_{12}\Sigma_{22}^{-1} Y_2 \mid Y_2 = y_2 \sim \mathcal{N}(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

$$\Rightarrow Y_1 \mid Y_2 = y_2 \sim \mathcal{N}(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2), \Sigma_{11\cdot 2})$$

where $\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. $\square$.

THM 10: $X \sim (\mu, \Sigma)$

Then $\mathbb{E}\, x^T A x = \mu^T A \mu + \mathrm{Tr}(A\Sigma)$.

Sidenote 11: $Q = \sum_{i=1}^{n-1} (X_i - X_{i+1})^2$

Then $Q = x^T A x$

where $A = \begin{pmatrix} 1 & -1 & & 0 \\ -1 & 2 & \ddots & \\ & \ddots & 2 & -1 \\ 0 & & -1 & 1 \end{pmatrix}$ $\square$.

Quadratic forms

$Y \sim N_p(0, I)$

Let $Q = y^T A y$, $A^T = A$. What is the dist. of Q?

THM (fundamental)

$Q \overset{d}{\sim} \chi^2_r$ iff $A^2 = A$ & $\text{rank}(A) = r$

- $\psi_Q(t) = E e^{tQ} = \dots = \det|I - 2tA|^{-\frac{1}{2}}$

$= \det|I - 2tD|^{-\frac{1}{2}}$

$= \prod_{i=1}^{p} (1 - 2t\lambda_i)^{-\frac{1}{2}}$

$(A = TDT^T)$. If $A^2 = A \Rightarrow \lambda(A) = 0$ or 1

$\Rightarrow \psi_Q(t) = (1-2t)^{-\frac{r}{2}}$

$\Rightarrow Q \sim \chi^2_r$.

- $\psi_Q(t) = \prod_{i=1}^{p} (1 - 2\lambda_i t)^{-\frac{1}{2}} = (1-2t)^{-\frac{r}{2}}$

$\Rightarrow \prod_{i=1}^{p} (1 - 2\lambda_i t) = (1-2t)^r \quad \forall t$

$\Rightarrow \lambda_i = 0$ for $(p-r)$ i's & $\lambda_i = 1$ for r i's. $\square$.

Ex $Q = Y^T(I - \frac{11^T}{n})Y$, $A^2 = A$

$r(A) = Tr(A) = Tr(I - \frac{11^T}{n})$

$= p - 1$.

$\Rightarrow Y \sim N(0, \sigma^2 I)$ thus

$Q/\sigma^2 \sim \chi^2_{p-1}$.

Ex

If $Y \sim N_p(0, \Sigma)$, Let

$W = \Sigma^{-\frac{1}{2}} Y \sim N(0, I)$

$Q = Y^T A Y = W^T \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} W$

Thus $Q \sim \chi^2_r$ iff

$A \Sigma A = A$ & $r(A) = r$ $\square$

THM $Y \sim N(0, I)$ Then $A^2 = A$, $B^2 = B$,

$y^T A y \perp\!\!\!\perp y^T B y$ iff $AB = 0$

Pf: $AY \perp\!\!\!\perp BY$ if $AB^T = 0$. &

$(\Leftarrow)$: $y^T A^T A y = y^T A y \perp\!\!\!\perp y^T B y = y^T B^T B y$.

$(\Rightarrow)$: $y^T A y \perp\!\!\!\perp y^T B y$ & $\sim \chi^2 \Rightarrow$

$y^T(A+B)y \sim \chi^2 \Rightarrow (A+B)$ is idem

$\Rightarrow A^2 + BA + AB + B^2 = A + B$

$\Rightarrow BA + AB = 0$

$\Rightarrow ABA + AB = 0$ &

$BA + ABA = 0$

$\Rightarrow AB = BA = 0$

$\boxed{\text{THM}}$: $Y\sim N_p(\mu, I)$ & $Q_i = Y^TP_iY$

If $Q_i\sim\chi^2_{r_i}$ & $Q_1 - Q_2 \ge 0$, $r_1>r_2$

Then $Q_1-Q_2 \perp\!\!\!\perp Q_2$ & $Q_1-Q_2\sim\chi^2_{r_1-r_2}$

Pf: $0\le Q_1-Q_2 = y^T(P_1-P_2)y\ \forall y$.

In particular, if $y\in N(P_1)$,

$0\le y^T(-P_2)y\le 0$ & $y^TP_2^TP_2y=0 \Rightarrow$

$N(P_1)\subseteq N(P_2)$. $\Rightarrow$

$\forall y$, $y^TP_2(I-P_1)y=0$ Since $(I-P_1)y\in N(P_1)$. $\Rightarrow$

$(P_1-P_2)^2 = P_1^2 - P_1P_2 - P_2P_1 + P_2^2 = P_1-P_2$

& $\text{rank}(P_1-P_2) = \text{Tr}(P_1-P_2) = r_1-r_2$. □

WEEK 4 Lecture 12

Recall $y\sim N(0,\Sigma)$ then

$Q=y^TAy$ has mgf $\frac{1}{|I-2tA|^{\frac12}}$

- If $y\sim N(0,\Sigma)$, then

$$\frac{1}{|I-2t\Sigma A|^{\frac12}} \text{ or } \frac{1}{|I-2tA\Sigma|^{\frac12}} \text{ or}$$

$$\frac{1}{|I-2t\Sigma^{-\frac12}A\Sigma^{\frac12}|^{\frac12}}$$

$(|I-AB|=|I-BA|)$.

$\boxed{\text{THM}}$

Let $y\sim N(0,\Sigma)$, $Q_i = y^TA_iy$, $i=1,2$

Then $Q_1\perp\!\!\!\perp Q_2$ iff $A_1\Sigma A_2=0$

Pf: $(\Leftarrow)$: $\psi_{Q_1,Q_2}(t_1,t_2) = E\, e^{t_1Q_1+t_2Q_2}$

$$= \int e^{t_1Q_1+t_2Q_2}\frac{1}{2\pi|\Sigma|^{\frac12}} e^{-\frac12 y^T\Sigma^{-1}y}\,dy$$

$$= c\int e^{-\frac12 y^T[\Sigma^{-1}-2t_1A_1-2t_2A_2]y}\,dy$$

$$= \frac{|\Sigma^{-1}-2t_1A_1-2t_2A_2|^{-\frac12}}{|\Sigma|^{\frac12}} \quad (*)$$

$$(*) = \frac{1}{|I-2t_1A_1\Sigma-2t_2A_2\Sigma|^{\frac12}}$$

by cond. (Since $A_1\Sigma A_2=0$)

$$= \frac{1}{|I-2t_1A_1\Sigma|^{\frac12}\,|I-2t_2A_2\Sigma|^{\frac12}}$$

$$= \psi_{Q_1}(t_1)\,\psi_{Q_2}(t_2)$$

$\boxed{\text{refer to A.4.9}}$

Want to find dist. of $Q=y^TAy$

when $y\sim N_p(\mu,\Sigma)$

Non-central χ^2 dist.

Def: Let $X_i\sim N(\mu_i,1)$.

$X_i\perp\!\!\!\perp X_j$ for $i\neq j$, then

define $Y=\sum_{i=1}^n X_i^2$. Y has a

non-central χ^2 with df n &

non-centrality para. $\delta^2=\sum_{i=1}^n\mu_i^2 = \|\mu\|_2^2$

$$Y\sim\chi^2_n(\delta^2)$$

Let $a_1=\mu/\|\mu\|$, $\mu=\begin{pmatrix}\mu_1\\ \vdots\\ \mu_n\end{pmatrix}$, & $\{a_1,\dots,a_n\}$ be an orthonormal basis

for R^n. Let $A=\begin{pmatrix}a_1^T\\ \vdots\\ a_n^T\end{pmatrix}$, $A^TA=I$.

Let $W=AX\sim N(A\mu, I)$

$$A\mu = \begin{pmatrix}\|\mu\|\\ 0\\ \vdots\\ 0\end{pmatrix} \Rightarrow AX\sim N\left(\begin{pmatrix}\|\mu\|\\ 0\end{pmatrix}, I\right)$$

$$Y=\sum_i X_i^2 = X^TA^TAX$$

$$= W^TW = \sum_i W_i^2$$

$$= \underbrace{W_1^2}_{\chi_1^2(\|\mu\|^2)} + \underbrace{\sum_{i=2}^n W_i^2}_{\sim\chi^2_{n-1}}$$

WEEK 4 Lec 13

Let $X_i \sim \mathcal{N}(\mu_i, 1)$, $X_i \perp X_j$, $1 \le i \ne j \le n$

$Y = \sum_{i=1}^n X_i^2$ is non-central $\chi_n^2(\delta^2)$

where $\delta^2 = \|\mu\|_2^2$, $\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$

$$Y \sim \chi_n^2(\delta^2)$$

$a_1 = \mu / \|\mu\|$, $\|a_1\| = 1$, $A = \begin{pmatrix} a_1^T \\ \vdots \\ a_n^T \end{pmatrix} \Rightarrow$

$$W = AX \sim \mathcal{N}\left(\begin{pmatrix} \|\mu\| \\ 0 \end{pmatrix}, I\right) \Rightarrow$$

$$Y = \|AX\|_2^2 = \sum_i W_i^2 = W_1^2 + \sum_{i=2}^n W_i^2.$$

$$\mathbb{E}Y = (n-1) + \|\mu\|_2^2 + 1 = \delta^2 + n$$

$$\text{Var}\,Y = \text{Var}\left(\sum_i X_i^2\right) = \sum_i \text{Var}(X_i^2)$$

$$= \sum_i \text{Var}\left((X_i - \mu_i)^2 + \mu_i(X_i - \mu_i)\right)$$

$$= 2n + 4\|\mu\|_2^2 = 2n + 4\delta^2$$

Can you approx the dist. of Y by something simpler?

$$\frac{Y - \mathbb{E}Y}{\sqrt{\text{Var}Y}} \dot\sim \mathcal{N}(0,1) \Rightarrow$$

$$\frac{Y - n - \delta^2}{\sqrt{2n + 4\delta^2}} \dot\sim \mathcal{N}(0,1)$$

Recall reproductive property:

$$X_i \sim \chi_{r_i}^2, \; i = 1, 2, \; X_1 \perp\!\!\!\perp X_2 \Rightarrow$$

$$X_1 + X_2 \sim \chi_{r_1 + r_2}^2$$

$Y_1 = \sum_{j=1}^{r_1} X_j^2$, $\quad X_i \sim \mathcal{N}(\mu_i, 1)$

$Y_2 = \sum_{j=r_1+1}^{n} X_j^2$, $\quad 1 \le i \le n$

$$Y_1 + Y_2 \overset{d}{\sim} \chi_{r_1 + r_2}^2(\delta_1^2 + \delta_2^2)$$

Verify mgf of $Y_1 + Y_2$: $(\lambda = \delta_1^2 + \delta_2^2)$

$$\psi_Y(t) = \frac{e^{\lambda t/(1-2t)}}{(1-2t)^{\frac{n}{2}}} \qquad 2t < 1$$

General Result for dist. of quadratic form

Let $Y \sim \mathcal{N}_p(\mu, \Sigma)$

What is the dist. of $Q = Y^T A Y$

Answer: $\text{rank}(A) = r$, $\Sigma \succ 0$

write $Q = Y^T \Sigma^{-\frac{1}{2}} T T^T \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} T T^T \Sigma^{-\frac{1}{2}} Y$

where $TT^T = T^T T = I$ & $T^T \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} T = \Lambda$

$$Q = W^T \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{pmatrix} W = \sum_{i=1}^p W_i^2 \lambda_i$$

λ_i: eigen-vals of $\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$ &

$$W \sim \mathcal{N}_p(T^T \Sigma^{-\frac{1}{2}} \mu, I)$$

$\Rightarrow Q = \sum_{i=1}^r \lambda_i W_i^2$ is a weigted sum of non-central $\chi_1^2(\delta_i^2)$

where $\delta_i^2 = (t_i^T \Sigma^{-\frac{1}{2}} \mu)^2$

- λ_i's are non-zero eigen-vals of $\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$ (or $A\Sigma$, ΣA).

(Ex): If $\mu=0$, $\Sigma=I$ & $A^2=A$

$$Q \overset{d}{=} \Sigma_i \lambda_i \chi_1^2(\delta_i^2) \text{ where } \delta_i=0$$

λ_i are non-zero eV's of $A\Sigma=A$

$$\Rightarrow Q \overset{d}{=} \sum_{i=1}^{r} \chi_1^2(0) \sim \chi_r^2$$

(Ex) If $\mu=0$, $A=\Sigma^{-1}$.

Then $Q=y^TAy=y^T\Sigma^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}\Sigma^{-\frac{1}{2}}y$

$$=\tilde{y}^T\tilde{y} \sim \chi_r^2(0)$$

If $\mu\neq 0$, then $\delta_i^2=\mu^T\Sigma^{-\frac{1}{2}}t_it_i^T\Sigma^{-\frac{1}{2}}\mu$

& $\sum_{i=1}^{r}\delta_i^2=\mu^T\Sigma^{-\frac{1}{2}}(\Sigma_i t_it_i^T)\Sigma^{-\frac{1}{2}}\mu$

But $T^T\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}T=I$

$$\Rightarrow Q \sim \chi_r^2(\mu^T\Sigma^{-1}\mu) \text{ or } \mu^TA\mu.$$

WEEK 5 Lecture 14

non-central t variate

Recall:

$$t_n(0) \overset{d}{=} \frac{\mathcal{N}(0,1)}{\sqrt{\chi_n^2/n}} \quad \& \quad \mathcal{N}(0,1) \perp\!\!\!\perp \chi_n^2$$

$$F_{n,m} \overset{d}{=} \frac{\chi_n^2/n}{\chi_m^2/m} \quad \& \quad \chi_n^2 \perp\!\!\!\perp \chi_m^2$$

non-central t & F:

$$t_n(\theta) \overset{d}{=} \frac{\mathcal{N}(\theta,1)}{\sqrt{\chi_n^2(0)/n}} \quad \& \quad \perp\!\!\!\perp$$

$$F_{n,m}(\delta^2) \overset{d}{=} \frac{\chi_n^2(\delta^2)/n}{\chi_m^2(0)/m} \quad \& \quad \perp\!\!\!\perp$$

(Ex) Let $X_i \sim \mathcal{N}(\mu,\sigma^2)$, $1\le i\le n$ & $X_i \perp\!\!\!\perp X_j$, $i\neq j$. Let

$$S^2=\frac{1}{n-1}\Sigma_i(X_i-\bar{X})^2$$

where $\bar{X}=\Sigma_iX_i/n$. What is the dist. of $\boxed{T=\frac{\bar{X}-a}{S/\sqrt{n}}}$?

We know $\bar{X} \perp\!\!\!\perp S^2$ &

$\bar{X}\sim\mathcal{N}(\mu,\frac{\sigma^2}{n})$, $\frac{(n-1)S^2}{\sigma^2}\sim\chi_{n-1}^2$

& $\frac{\bar{X}-a}{\sigma/\sqrt{n}} \sim \mathcal{N}(\frac{\mu-a}{\sigma/\sqrt{n}},1)$

Thus,
$$T=\frac{\left(\frac{\bar{X}-a}{\sigma/\sqrt{n}}\right)\frac{\sigma}{\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}}\frac{\sigma}{\sqrt{n-1}}} \overset{d}{=} \frac{\mathcal{N}(\theta,1)}{\sqrt{\chi_{n-1}^2/n-1}}$$

$$\Rightarrow T \sim t_{n-1}^2(\theta)$$

where $\theta=\mu-a$.

Fisher-Cochran's Theorem

If $Y \sim N_p(0, I)$ & $Y^TY = \sum_{i=1}^{k} Q_i$

where $Q_i = Y^TA_iY$, $\text{rank}(A) = r_i$.

Then T.F.A.E.:

1. $Q_i \perp\!\!\!\perp Q_j \quad i \neq j$
2. $Q_i \sim \chi^2_{r_i}$, $i = 1, \dots, k$
3. $\sum_{i=1}^{k} r_i = p$

Loynes' Lemma: If $M^2 = M = M^T$ & $P^T = P \geq 0$ & $I - M - P \geq 0$. Then

$$PM = MP = 0$$

Pf: Let $Y \in \mathbb{R}^n$ & $Z = MY$, $Z^TZ = Y^TM^TMY$

or $Z^TZ = Y^TMMY = Z^TMZ \Rightarrow$

$Z^T(I-M)Z = 0$

- $0 \leq Z^T(I-M-P)Z = -Z^TPZ \leq 0$

$\Rightarrow Z^TPZ = 0 \Rightarrow PZ = 0$ or

$PMY = 0 \quad \forall Y$

$\Rightarrow PM = MP = 0$

Marsaglia & Goraybill's lemma

Suppose $D_i = D_i^T$ & $I - D_i - D_j \geq 0$ $1 \leq i \leq k$. Then any 2 of the following statements imply the third:

1. $D_i^2 = D_i$, $i = 1, \dots, k$
2. $D_iD_j = 0 \quad \forall i \neq j$
3. $D = \sum_i D_i$ is idempotent

Pf: (1)(2) $\Rightarrow$ (3): Trivial

(3)(1) $\Rightarrow$ (2): $D - D_i - D_j = \sum_{k \neq i,j} D_k \geq 0$

$I - D \geq 0$ by (3). (& idempotent)

$I - D_i - D_j = (I - D) + (D - D_i - D_j) \geq 0$

Then by Loynes's lemma where

$M = D_i$; $P = D_j$

$\Rightarrow D_iD_j = 0$

(2)(3) $\Rightarrow$ (1): $D_ix = \lambda x \Rightarrow$

$DD_ix = \lambda Dx$ or

$D_i^2x = \lambda Dx$ or

$\lambda^2 x = \lambda Dx \Rightarrow$

$\lambda = 0$ or 1 since $D^2 = D$

$\Rightarrow D_i^2 = D_i$. $\square$

WEEK5 Lecture 15

Recall $Y \sim \mathcal{N}_p(\mu, \Sigma)$.

$$Q = y^T A y \overset{d}{=} \sum_{i=1}^{p} \lambda_i \chi_1^2(\delta_i^2)$$

where λ_i's are e.v.s of $A\Sigma$ or ΣA or $\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$

$$Q = y^T \Sigma^{-\frac{1}{2}} T T^T \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} T \underbrace{T^T \Sigma^{-\frac{1}{2}} y}_{w}$$

$$w \sim \mathcal{N}(T^T \Sigma^{-\frac{1}{2}} \mu, I)$$

$$\Rightarrow Q = w^T D w = (w_1 \cdots w_p) \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_p \end{pmatrix}$$

$$= \sum_{i=1}^{p} \lambda_i w_i^2 = \sum_{i=1}^{r} \lambda_i w_i^2$$

$$w_i^2 \sim \chi_1^2\left((t_i^T \Sigma^{-\frac{1}{2}} \mu)^2\right), \quad \lambda_1 \geq \cdots \geq \lambda_r > 0.$$

Ex $\mu \neq 0$. $A = \Sigma^{-1}$, $Q = Y^T A Y$?

$$Q = Y^T \Sigma^{-\frac{1}{2}} \Sigma^{-\frac{1}{2}} Y = \tilde{Y}^T \tilde{Y},$$

$$\tilde{Y} \sim \mathcal{N}(\Sigma^{-\frac{1}{2}} \mu, I) \Rightarrow$$

$$Q \sim \chi_p^2(\mu^T \Sigma^{-1} \mu)$$

Ex $A^2 = A$ & $\Sigma = I$ & $\mu \neq 0$?

$$A\Sigma = A \Rightarrow Q = \sum_{i=1}^{r} \chi_1^2(\delta_i^2)$$

$$\delta_i^2 = (t_i^T \Sigma^{-\frac{1}{2}} \mu)^2 = (t_i^T \mu)^2$$

$$\Rightarrow \sum_{i=1}^{r} \delta_i^2 = \mu^T \left(\sum_{i=1}^{r} t_i t_i^T\right) \mu$$

But $T^T A T = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \Rightarrow$

$$A = (T_1 \ T_2) \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} T_1^T \\ T_2^T \end{pmatrix}$$

$$= T_1 T_1^T$$

$$\Rightarrow \sum_{i=1}^{r} \delta_i^2 = \mu^T A \mu$$

Proof of Fisher-Cochran Theorem cont'd

If $Y \sim \mathcal{N}_p(0, I)$ & $Y^T Y = \sum_i^k Q_i$,

$Q_i = Y^T A_i Y$. $\text{rank}(A_i) = r_i$.

T.F.A.E.

(i) $Q_i \perp\!\!\!\perp Q_j \quad 1 \leq i \neq j \leq k$

(ii) $Q_i \sim \chi_{r_i}^2$

(iii) $\sum_{i=1}^{k} r_i = p$

Strategy: Show (i) ⇒ (ii) ⇒ (iii) ⇒ (i)

Pf: (i) ⇒ (ii): $Q_i \perp\!\!\!\perp Q_j \Rightarrow Q_1 \perp\!\!\!\perp \sum_{i=2}^{k} Q_i$

or $Y^T A_1 Y \perp\!\!\!\perp Y^T \left(\sum_{i=2}^{k} A_i\right) Y$

$\overset{\text{Craig's lemma}}{\Longrightarrow} A_1 \left(\sum_{i=2}^{k} A_i\right) = 0$ (Lec. 12)

$\Rightarrow A_1 (I - A_1) = 0 \Rightarrow A_1^2 = A_1$

$\Rightarrow A_i^2 = A_i, \quad i = 1, \ldots, k$

$\Rightarrow Q_i \sim \chi_{r_i}^2$ Fund. Thm.

(ii) ⇒ (iii): $Q_i \sim \chi_{r_i}^2 \Rightarrow A_i^2 = A_i$

$$\Rightarrow \sum_{i=1}^{k} r_i \overset{\text{idem.}}{=} \sum_{i=1}^{k} Tr(A_i) = Tr\left(\sum_i A_i\right)$$

$$= Tr(I) = p.$$

(iii) ⇒ (i): $A_1 + A = I$, $A = \sum_{i=2}^{k} A_i$

Let T be s.t. $T^T T = T T^T = I$ &

$$T^T A_1 T = \Lambda \Rightarrow$$

$$T^T A_1 T + T^T A T = T^T T = I \Rightarrow$$

$$\underset{\begin{pmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_{r_1} & & \\ & & & 0 & \\ & & & & \ddots \ 0 \end{pmatrix}}{\Lambda} + \underset{\begin{pmatrix} 1-\lambda_1 & & & & \\ & \ddots & & & \\ & & 1-\lambda_{r_1} & & \\ & & & 1 & \\ & & & & \ddots \ 1 \end{pmatrix}}{T^T A T} = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix} \Rightarrow$$

(key step)

But $r(T^TAT) = r(A) = r(\sum_{i=2}^{p} A_i)$

$\leq \sum_{i=2}^{p} r(A_i) = p - r_1$ (by (iii))

$\Rightarrow 1 - \lambda_i = 0 \quad \forall i = 1, \cdots, r$

$\Rightarrow \lambda_i = 1$

$\Rightarrow A_1^2 = A_1 \Rightarrow A_i^2 = A_i, \ i = 1, \cdots, k$

Next. $I - A_i - A_j \geq 0, \ A_i^T = A_i, \ A_j^T = A_j$

$\Rightarrow A_i A_j = 0 \quad \forall \quad 1 \leq i \neq j \leq k$

(Laynes Lemma)

But $Q_i \perp\!\!\!\perp Q_j$ iff $A_i \Sigma A_j = 0$

(craig's Lemma)

$\Rightarrow Q_i \perp\!\!\!\perp Q_j \quad \forall \quad 1 \leq i \neq j \leq k.$ $\square$.

$$SSTr = \sum_{j}^{n} \sum_{i}^{k} (\bar{Y}_i - \bar{Y})^2 \frac{\sigma^2}{n} \Big/ \frac{\sigma^2}{n}$$

$\Rightarrow \frac{SSTr}{\sigma^2} \sim \chi^2_{k-1}$ (under H_0)

$\Rightarrow \mathbb{E} \frac{SSTr}{(k-1)} = \sigma^2 + n \frac{\sum_i (\tau_i - \bar{\tau})^2}{k-1}$

WEEK 5 Lecture 16

Applications of Cochran's thm.

Ex $y \sim \mathcal{N}(0, I)$ &

$y^Ty = y^T \frac{11^T}{n} y + y^T (I - \frac{11^T}{n}) y \Rightarrow$

$y^T (I - \frac{11^T}{n}) y \sim \chi^2_r(0), \ r = Tr(I - \frac{11^T}{n}) = n-1$

$y^T \frac{11^T}{n} y \sim \chi^2_1$

Ex In reg. Problem:

$$Y = X\beta + \epsilon, \quad \beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_{p-1} \end{pmatrix}$$

WT Test $\beta_1 = \cdots = \beta_{p-1} = 0$ (Omibus test)

& $\hat{\beta} = (X^TX)^{-1} X^T y, \quad r(\underset{n \times p}{X}) = p.$

$$Y^TY = \underbrace{Y^T(I-P)Y}_{SSE} + \underbrace{Y^T(P - \frac{11^T}{n})Y}_{SS Reg} + Y^T \frac{11^T}{n} Y$$

Test H_0: $\frac{SS Reg}{SSE} \frac{n-p}{p-1} \sim F_{p-1, n-p}$

Ex ANOVA (one-way)

$Y_{ij} = \mu + \tau_i + \epsilon_{ij}, \ j = 1, \cdots, n, \ i = 1, \cdots, k$

$\epsilon_{ij} \sim NID(0, \sigma^2)$

Q: Test $\tau_1 = \cdots = \tau_k$. ?

$\sum_{i=1}^{k} \sum_{j=1}^{n} (Y_{ij} - \bar{Y})^2 = \sum_{i=1}^{k} \sum_{j=1}^{n} (Y_{ij} - \bar{Y}_{i\cdot} + \bar{Y}_{i\cdot} - \bar{Y})^2$

$= \sum_i \sum_j (Y_{ij} - \bar{Y}_{i\cdot})^2 + \sum_i \sum_j (\bar{Y}_{i\cdot} - \bar{Y})^2$

$= \sum_i \sum_j (Y_{ij} - \bar{Y}_{i\cdot})^2 + n \sum_i (\bar{Y}_{i\cdot} - \bar{Y})^2$

$= \sum_i \sum_j (\mu + \tau_i + \epsilon_{ij} - \mu - \tau_i - \bar{\epsilon}_{i\cdot})^2 +$

$n \sum_i (\mu + \tau_i + \bar{\epsilon}_{i\cdot} - \mu - \bar{\tau} - \bar{\epsilon})^2$

$SSE = \sum_j \sum_i (\epsilon_{ij} - \bar{\epsilon}_{i\cdot})^2 \frac{\sigma^2}{\sigma^2} \Rightarrow \frac{SSE}{\sigma^2} \sim \sum_{i=1}^{k} \chi^2_{n-1}$

& $\mathbb{E} \frac{SSE}{(n-1)k} = \sigma^2.$

Lack of fit test

H_0: $\mathbb{E}Y = \beta_0 + \beta_1 X + \beta_2 X^2$?

VS H_a: $\mathbb{E}Y \neq \beta_0 + \beta_1 X + \beta_2 X^2$.

$\sum_i \sum_j (Y_{ij} - \bar{Y}_i)^2 = SSE = \underset{\text{pure}}{SSPE} + \underset{\text{lack of fit}}{SSLof}$

LEC17 Linear Models

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

$$\mathbb{E}\varepsilon = 0, \quad Cov\,\varepsilon = \sigma^2 I$$

Goal: Estimate β.

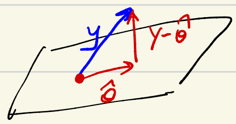
Write $Y = X\beta + \varepsilon = \theta + \varepsilon$,

$$\theta = X\beta \in C(X) = \Omega$$

Least-Square Criterion: find $\hat{\theta}$ s.t.

$$\|Y - \hat{\theta}\|^2 \leq \|Y - \theta\|^2 \quad \forall \theta \in \Omega.$$

How to find $\hat{\theta}$?



So $\hat{\theta}$ must be s.t.

$$(Y - \hat{\theta})^T \theta = 0, \quad \theta \in \Omega$$

$$(Y - \hat{\theta})^T X\beta = 0 \quad \forall \beta$$

$$\Rightarrow (Y - \hat{\theta})^T X = 0 \text{ or}$$

$$X^T \hat{\theta} = X^T Y \text{ or}$$

$$X^T X \hat{\beta} = X^T Y \quad \text{[Normal Eq.]}$$

There is always a sol. to the normal eqs.

Recall $Ax = b$ iff linear system consistent.

i.e., $\text{rank}(A) = \text{rank}(A|b)$

This is true for normal eq. because

$$r(X^TX | X^TY) = r(X^T(X|Y)) \leq r(X^T) = p$$

But LHS $\geq r(X^TX) = r(X^T) = p$

If $Ax = b$ has a sol, then $\tilde{x} = A^- b$ is a sol.:

$$A\tilde{x} = AA^-b = AA^-Ax = b$$

This means

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

is a least sq. estimate of β

In ANOVA, this issue arises.

$$\mathbb{E}Y = \begin{pmatrix} 1 & 1_k & & 0 \\ & & \ddots & \\ & 0 & & 1_k \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \vdots \\ \tau_m \end{pmatrix}$$

WTK: $\tau_i = \tau_j$? (μ is a nuisance para.)

So $\hat{\beta}$ is not unique. However,

$$\hat{Y} = \hat{\theta} = X\hat{\beta} \text{ is unique}$$

$$= X(X^TX)^- X^T Y$$

$$= P_{C(X)} Y$$

$e = Y - \hat{Y}$ residual vector.

So e^Te is unique.

Another way to derive
l.s.e. of β is as follows
WT find $\hat{\beta}$ s.t.
$$\|Y - X\hat{\beta}\|^2 \le \|Y - X\beta\|^2$$
let $\gamma = \|Y - X\beta\|^2$
Then $\frac{\partial \gamma}{\partial \beta} = 2X^TX\beta - 2X^TY = 0$
$$\frac{\partial^2 \gamma}{\partial \beta \partial \beta^T} = 2X^TX \succeq 0$$
$$\Rightarrow \quad \hat{\beta} = (X^TX)^- X^TY$$

Rogers: Matrix derivatives.

Note: $\hat{\beta}$ is MLE for $Y \sim N_p(X\beta, \sigma^2 I)$
what is the MLE for $Y \sim N_p(X\beta, \Sigma)$
Hint: Spectral Decomp. Thm.

Q: $\|Y - \theta\|^2$ is this a good Criterion?
$$\hat{\sigma}^2 = \frac{Y^T(I - P_X)Y}{n-p}$$ is unbiased for σ^2.
Since
$$\mathbb{E}\hat{\sigma}^2 = \frac{1}{n-p}\left[Tr(I-P_X)\sigma^2 I + \underbrace{(X\beta)^T(I-P_X)(X\beta)}_{(*)}\right]$$
$$= \sigma^2$$ $(*) = 0$ Since $(I - P_X) \perp P_X$.

But, can we do this?
$$\hat{\beta} = \underset{\theta}{\arg\min} \|Y - \theta\|_1$$

AKA Gauss-Markov Thm

$\boxed{\textbf{THM}}$ L.S.E. $\hat{\theta}$ is BLUE.

If interest is in estimating $c^T\theta$ where c is given, then:
$c^T\hat{\theta}$ is the
Best Unbiased Linear Estimator

Lec 18

Properties of LS & G-LSE

$y = X\beta + \varepsilon = \theta + \varepsilon$, $\mathbb{E}\varepsilon = 0$, $\mathrm{Cov}\,\varepsilon = I$

$\Rightarrow \hat{\beta}_{LSE} = (X^TX)^{-1}X^Ty$

Recall

$$\|y - X\hat{\beta}\|_2^2 \le \|y - X\beta\|_2^2 \quad \forall \beta \in \mathbb{R}^p$$

In terms of $\theta = X\beta \in C(X)$

$$\|y - \hat{\theta}\|_2^2 \le \|y - \theta\|_2^2$$

$$\hat{\theta} = X\hat{\beta} = X(X^TX)^{-1}X^Ty$$

$$:= P_X y$$

Note: $P_X^2 = P_X = P_X^T$.

$\mathbb{E}\hat{\beta} = \mathbb{E}(X^TX)^{-1}X^Ty = \beta$ &

$\mathbb{E}\hat{\theta} = X\mathbb{E}\hat{\beta} = X\beta = \theta$.

$S^2 = (y - \hat{y})^T(y - \hat{y})/n-p$, $\hat{y} = X\hat{\beta} = \hat{\theta}$

$$= (y^T\underbrace{(I - P_X)}_{Q_X}y)/n-p$$

$$\mathbb{E}S^2 = \frac{1}{n-p}\mathbb{E}\,y^TQ_Xy$$

$$= \frac{1}{n-p}(\theta^TQ_X\theta + \mathrm{Tr}(Q_X\sigma^2 I))$$

$$= \frac{1}{n-p}\sigma^2(n-p) = \sigma^2$$

Could also use L_1-norm

$$\sum_{i=1}^n |y_i - x_i^T\hat{\beta}| \le \sum_{i=1}^n |y_i - x_i^T\beta| \quad \forall \beta \in \mathbb{R}^p$$

$\boxed{\text{THM}}$: Gauss-Markov

Suppose $y = X\beta + \varepsilon = \theta + \varepsilon$.

If we want to estimate $c^T\theta$ for a given c, then

$$\mathrm{Var}(c^T\hat{\theta}) \le \mathrm{Var}(a^Ty)$$

where a^Ty is any linear unbiased estimate of $c^T\theta$.

Pf: a^Ty unbiased for $c^T\theta$

$\Rightarrow \mathbb{E}a^Ty = c^T\theta = a^T\theta \quad \forall \theta \in C(X)$

$\Rightarrow c^TX\beta = a^TX\beta \quad \forall \beta \in \mathbb{R}^p$

$\Rightarrow c^TX = a^TX$.

$\mathrm{Var}(c^T\hat{\theta}) - \mathrm{Var}(a^Ty)$

$= c^T\mathrm{Var}(X\hat{\beta})c - a^T(\mathrm{Cov}\,y)a$

$= \sigma^2c^TX(X^TX)^{-1}X^Tc - \sigma^2a^Ta$

$= \sigma^2a^T(P_X - I)a \le 0$

$\Rightarrow \mathrm{Var}(c^T\hat{\theta}) \le \mathrm{Var}(a^Ty)$

$\forall a$ s.t. a^Ty unbiased for $c^T\theta$.

Is $c^T\hat{\beta}$ BLUE for $c^T\beta$?

Yes, because $r(X) = p$ and this implies

$$\theta = X\beta, \quad \beta = (X^TX)^{-1}X^T\theta$$

Then pick c appropriately to show $c^T\hat{\beta}$ is BLUE for $c^T\beta$.

Dist. of $\hat{\beta}, S^2$

$$Y = X\beta + \varepsilon,\ \varepsilon \sim \mathcal{N}_n(0, \sigma^2 I)$$

$$\Rightarrow \hat{\beta} = (X^TX)^{-1}X^TY,\ Y \sim \mathcal{N}_n(X\beta, \sigma^2 I)$$

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

Thus, $(\hat{\beta} - \beta)^T \frac{X^TX}{\sigma^2} (\hat{\beta} - \beta) \sim \chi^2_p(0)$

$\Rightarrow$ confidence ellipsoid for β is

$$\left\{\beta : (\hat{\beta} - \beta)^T \frac{X^TX}{\sigma^2} (\hat{\beta} - \beta) \le \chi^2_{p,\alpha}(0)\right\}$$

provided σ^2 is known.

Recall:

$$(n-p)S^2 = Y^TQ_XY,\ Y \sim \mathcal{N}(X\beta, \sigma^2 I)$$

$$\Rightarrow (n-p)S^2 \sim \chi^2_{n-p}(0).$$

Are $S^2 \perp\!\!\!\perp \hat{\beta}$?

Yes, look at $Cov(Q_XY, \hat{\beta})$

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

$$= Q_X Cov(Y, Y) X(X^TX)^{-1}$$

$$= \sigma^2 Q_X X(X^TX)^{-1} = 0.$$

$$\Rightarrow Q_XY \perp\!\!\!\perp \hat{\beta} \Rightarrow Y^TQ_XY \perp\!\!\!\perp \hat{\beta}.$$

MLEs for β & S^2

$$L(Y|\beta, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(Y-X\beta)^T(Y-X\beta)}$$

$$\ell(Y|\beta, \sigma^2) = \sum_{i=1}^{n} \left[\ln \frac{1}{\sqrt{2\pi}\sigma} - \frac{(Y-X\beta)^T(Y-X\beta)}{2\sigma^2}\right]$$

Differentiate w.r.t. β & σ^2 & Solve for them to get MLEs:

$$\hat{\beta}_{MLE} = (X^TX)^{-1}X^TY$$

$$\hat{\sigma}^2_{MLE} = \frac{\|Y - X\beta\|_2^2}{n}$$

Then $\frac{\partial^2 \ell(Y|\beta, \sigma^2)}{\partial\beta\, \partial\beta^T}$ $\frac{\partial^2 \ell(Y|\beta, \sigma^2)}{\partial\beta\, \partial\sigma^2}$

Gives us Fisher Info Matrix.

Info. matrix

$$\mathcal{I} = -\mathbb{E}\begin{pmatrix} \partial^2 & \partial^2 \\ \partial^2 & \partial^2 \end{pmatrix}_{(p+1)\times(p+1)} \ \&\ Cov\begin{pmatrix} \hat{\beta} \\ \hat{\sigma}^2 \end{pmatrix} = \mathcal{I}^{-1}$$

Weighted LS or G-LS

$Y = X\beta + \varepsilon,\ \mathbb{E}\varepsilon = 0,\ Cov\ \varepsilon = V > 0$ (known)

- $\tilde{\varepsilon} = V^{-\frac{1}{2}}\varepsilon \sim \mathcal{N}(0, I)$

$$V^{-\frac{1}{2}}Y = V^{-\frac{1}{2}}X\beta + V^{-\frac{1}{2}}\varepsilon$$

$$\tilde{Y} = \tilde{X}\beta + \tilde{\varepsilon}$$

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

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

$$\&\ Cov(\hat{\beta}_W) = (X^TV^{-1}X)^{-1}$$

Lec19 Adding Regressors

$$\mathbb{E}Y = X\beta = \underset{n\times p_1}{(X_1} \mid \underset{n\times p_2}{X_2)} \begin{pmatrix} \beta_1 \\ \hline \beta_2 \end{pmatrix}\begin{matrix} p_1 \\ p_2 \end{matrix}$$

& $\hat{\beta} = (X^TX)^{-1}X^Ty$, $\text{Cov}\,\hat{\beta} = \sigma^2(X^TX)^{-1}$

- $\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} X_1^TX_1 & X_1^TX_2 \\ X_2^TX_1 & X_2^TX_2 \end{pmatrix}^{-1} \begin{pmatrix} X_1^TY \\ X_2^TY \end{pmatrix}$

If $X_1^TX_2 = 0$, then

$$\hat{\beta}_1 = (X_1^TX_1)^{-1}X_1^Ty$$
$$\hat{\beta}_2 = (X_2^TX_2)^{-1}X_2^Ty$$

Q: How to make them orthogonal?

G model: $\mathbb{E}Y_G = \underset{n\times p}{X}\beta + \underset{n\times t}{Z}\gamma$

$= W\delta$

where $W = (X, Z)$, $\delta = \begin{pmatrix} \beta \\ \gamma \end{pmatrix}$

& $\text{rank}(Z_{n\times t}) = t$.

WLOG, assume $Z^TX = 0$

Let $\hat{\theta} = P_XY = X(X^TX)^{-1}X^Ty$

- $Q = R = I - P_X = I - P$

$\boxed{\text{Claim}}$: $\underset{t\times n}{Z^T}\underset{n\times n}{R}\underset{n\times t}{Z}$ non-singular.

Let $Z^TRZa = 0$. WTS $a = 0$

- $\|R^{\frac{1}{2}}Za\|_2^2 = 0$

$\Rightarrow R^{\frac{1}{2}}Za = 0$

$\Rightarrow RZa = 0$

$\Rightarrow (I - P_X)Za = 0$

$\Rightarrow Za = P_XZa = X(X^TX)^{-1}X^TZa$

By assumption, $C(X) \cap C(Z) = \{0\}$

$\Rightarrow Za = 0 \Rightarrow a = 0$ since $\text{rank}(Z) = 0$.

$\Rightarrow Z^TRZ$ non-singular.

Write $Y = X\beta + Z\gamma$

as $X\beta + P_XZ\gamma + (I - P_X)Z\gamma$

$= X\alpha + RZ\gamma$

where

$\alpha = \beta + (X^TX)^{-1}X^TZ\gamma$

$R = I - P_X$

& notice

$X^T(RZ) = 0$.

- $\hat{\alpha} = (X^TX)^{-1}X^Ty$

$= \hat{\beta}_G + (X^TX)^{-1}X^TZ\hat{\gamma}_G$

& $\hat{\beta}_G = (X^TX)^{-1}X^T(y - Z\hat{\gamma}_G)$

& $\hat{\gamma}_G = (Z^TRZ)^{-1}Z^TRy$

Residuals from the G-model

$$\hat{Y}_G = X\hat{\beta}_G + Z\hat{\gamma}_G$$

The residual vector is

$$e_G = Y - \hat{Y}_G = Y - X(\hat{\beta} - (X^TX)^{-1}X^TZ\hat{\gamma}_G) - Z\hat{\gamma}_G$$
$$= Y - X\hat{\beta} + (P_X - I)Z\hat{\gamma}_G$$
$$= R(Y - Z\hat{\gamma}_G)$$

$$SSE_G = e_G^Te_G = (Y - Z\hat{\gamma}_G)^TR(Y - Z\hat{\gamma}_G)$$
$$= Y^TRY - 2\hat{\gamma}_G^TZ^TRY + \hat{\gamma}_G^TZ^TRZ\hat{\gamma}_G$$
$$= Y^TRY - \hat{\gamma}^TZ^TRY - \hat{\gamma}_G^TZ^T(RY - RZ\hat{\gamma}_G)$$
$$= Y^TRY - \hat{\gamma}^TZ^TRY + 0$$

$\hat{\gamma}_G = ((RZ)^TRZ)^{-1}Z^TR^TY$

i.e. $Z^TRZ\hat{\gamma}_G = Z^TRY$

$$\hat{\gamma}^TZ^TRY = ((Z^TRZ)^{-1}(Z^TRY))^TZ^TRY$$
$$= Y^TRZ(Z^TRZ)^{-1}Z^TRY$$
$$\geq 0 \quad \text{Since } r(Z) = t$$

$\therefore$ SSReg. cannot be smaller when we add regressors. However,

$$Cov(\hat{\beta}_G) \geq Cov(\hat{\beta})$$

- $Cov(\hat{\beta}_G) = Cov(\hat{\beta} - (X^TX)^{-1}X^TZ\hat{\gamma}_G)$

$$= \sigma^2(X^TX)^{-1} + \sigma^2(X^TX)^{-1}X^TZ(Z^TRZ)^{-1}Z^TX(X^TX)^{-1}$$
$$\geq \sigma^2(X^TX)^{-1}$$

Since $Cov(\hat{\beta}, (X^TX)^{-1}X^TZ\hat{\gamma}_G)$

$$= Cov((X^TX)^{-1}X^TY, (X^TX)^{-1}X^TZ\hat{\gamma}_G)$$

$\hat{\gamma}_G = (Z^TRZ)^{-1}Z^TRY$

$$= 0 \quad \text{Since } X^TR = X^T(I - P_X) = 0$$

Lec 20 Lagrangian Multiplier & Orthogonal Projection

Recall least squares est. for β in $EY = X\beta$.

$$\hat{\beta} = \arg\min_{\beta} \|Y - X\beta\|_2^2$$

$$\hat{\theta} = X\hat{\beta} = X(X^TX)^{-1}X^Ty = P_X y$$

Constrained least-squares:

$$\min \|Y - X\beta\|_2^2$$
$$\text{s.t.} \quad \underset{q\times p}{A}\,\underset{p\times 1}{\beta} = \underset{q\times 1}{c}$$

where $r(A) = q$ & A, c are given.

<u>Motivation:</u> Suppose a Quadratic model is of interest.

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = Y = \begin{pmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix}\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

- $Y = X\beta + \varepsilon$

It may be known that

$$\begin{pmatrix} y_0 \\ \tilde{y}_0 \end{pmatrix} = \begin{pmatrix} 1 & x_0 & x_0^2 \\ 1 & \tilde{x}_0 & \tilde{x}_0^2 \end{pmatrix}\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

So $A = \begin{pmatrix} 1 & x_0 & x_0^2 \\ 1 & \tilde{x}_0 & \tilde{x}_0^2 \end{pmatrix}$, $c = \begin{pmatrix} y_0 \\ \tilde{y}_0 \end{pmatrix}$ & $A\beta = c$.

- $\min \|Y - X\beta\|_2^2$
  s.t. $A\beta = c$

$$\mathcal{U} = \|Y - X\beta\|_2^2 + \lambda(A\beta - c)$$

$$\left.\begin{aligned} \frac{d\mathcal{U}}{d\beta} &= 0 \\ \frac{d\mathcal{U}}{d\lambda} &= 0 \end{aligned}\right\} \Rightarrow \text{If } (\hat{\beta}_c, \hat{\lambda}_c) \text{ are solutions, then verify max or min.}$$

$\boxed{\text{Lagrangian Multiplier Method}}$

Orthogonal Projections

Suppose V is a r-dim sbsp. of Ω which is n-dim. i.e. $V \subseteq \Omega$.

Given $y \in \Omega$, $\hat{y}$ is an orthogonal proj. of y onto V if

(i) $\hat{y} \in V$

(ii) $y - \hat{y} \in V^\perp$ i.e.

$$\langle y, v\rangle = \langle \hat{y}, v\rangle \quad \forall v \in V$$

Q: How to find $\hat{y}$?

Suppose V has $\{x_1, \dots, x_r\}$ as an orthonormal basis. Extend it to n-dim (Ω)

$$\{x_1, \dots, x_r, x_{r+1}, \dots, x_n\}$$

an orthonormal basis for Ω. Then

$$y \in \Omega: \quad y = \sum_{i=1}^n \alpha_i x_i$$

$$= \underbrace{\sum_{i=1}^r \alpha_i x_i}_{\in V} + \underbrace{\sum_{j=r+1}^n \alpha_j x_j}_{\in V^\perp}$$

$$= \hat{y} + \sum_{j=r+1}^n \alpha_j y_j$$

$$y - \hat{y} = \sum_{j=r+1}^n \alpha_j v_j \in V^\perp.$$

$\boxed{\text{Claim}}$ $\hat{y}$ is unique.

Pf: Let $\hat{y}_1, \hat{y}_2$ be 2 ortho. proj. of y onto V.

$$y = \hat{y}_1 + \hat{y}_1^\perp = \hat{y}_2 + \hat{y}_2^\perp$$

$$\Rightarrow \hat{y}_2 - \hat{y}_1 = \hat{y}_1^\perp - \hat{y}_2^\perp$$

Note LHS $\in V$ & RHS $\in V^\perp$

But $V \cap V^\perp = \{0\} \Rightarrow$

$$\hat{y}_1 = \hat{y}_2 \qquad \square.$$

Can we write $\hat{y}$ more explicitly?

$$y = \sum_{i=1}^r \alpha_i x_i$$

If $1 \le j \le r$, $\langle y, x_j\rangle = \sum_{i=1}^r \alpha_i \langle x_i, x_j\rangle = \alpha_j$

$$\Rightarrow \quad y = \sum_{i=1}^r \langle x_j, y\rangle x_i$$

$$= (x_1 \cdots x_r)\begin{pmatrix} x_1^T y \\ \vdots \\ x_r^T y\end{pmatrix}$$

$$= TT^T y$$

$$= Py$$

where $T = (x_1 \cdots x_r)$

- $\boxed{P = TT^T}$ is an orthogonal projection matrix.

NOTE: This O.P is unique (Given the sbsp. of interest). If P_1 & P_2 are O.P. matrices onto V, then $\forall y \in \Omega$:

$$P_2 y = \hat{y} = P_1 y \quad \forall y$$

$$\text{or} \quad (P_1 - P_2)y = 0 \quad \forall y$$

- This $P = TT^T$ is unique.

Another Ex concerns $(X^T X)^-$ which is not unique. Remarkably,

$$X(X^T X)^- X^T$$

is UNIQUE.

$\boxed{\text{Lemma:}}$

If P is the O.P. onto V, then

$$C(P_V) = V$$

If $w \in C(P_V)$, $w = (P_V z \text{ for some } z) \in V$ $\Rightarrow C(P_V) \subseteq V$. Conversely, if $w \in V$, then $P_V w = w \in C(P_V)$. $\Rightarrow C(P_V) \supseteq V$.

Some charaterization of O.P.

$$\hat{Y} = TT^T y \quad \forall y$$

& TT^T is the O.P. onto $C(TT^T)$

① TT^T is idempotent & symmetric iff TT^T is the O.P.

Pf: $T\underbrace{T^T T}_{I} T^T = TT^T$ $(\Leftarrow)$

$(\Rightarrow)$ $y = Py + (I-P)y$

$= Y_1 + Y_2$ &

$Y_1 \perp Y_2$ Since $y^T P^T (I-P) y = 0$

So P is an O.P. iff $P^2 = P = P^T$.

$\boxed{\text{Facts about O.P. matrix}}$

① If $W \subseteq \Omega$, $P_W P_\Omega = P_\Omega P_W = P_W$

Because $\forall Y$

$P_\Omega(P_W Y) = P_W Y$

② $P_\Omega - P_W$? Since $\Omega = W + W^\perp \cap \Omega$ for $W \subseteq \Omega$.

$\Rightarrow P_\Omega = P_W + P_{W^\perp \cap \Omega}$

If $W \subseteq \Omega$ we have

$$P_\Omega - P_W = P_{W^\perp \cap \Omega}$$

Let $W = N(A_1) \cap \Omega$

$\boxed{\text{Claim:}}$ $W^\perp \cap \Omega = C(P_\Omega A_1^T)$

Thus, $P_{W^\perp \cap \Omega} = (P_\Omega A_1^T)(A_1 P_\Omega A_1^T)^{-1} A_1 P_\Omega$

Since $P_X = X(X^TX)^{-1}X$. & $P_X \cdot P_X = P_X$.

Pf: WTS $W^\perp \cap \Omega = C(P_\Omega A_1^T)$ when $W = N(A_1) \cap \Omega$.

$(\supseteq)$: Let $x \in C(P_\Omega A_1^T)$, thus

$$x = P_\Omega A_1^T z \in \Omega \text{ for some } z$$

Next if $u \in W \subseteq \Omega$

$u^T x = u^T P_\Omega A_1^T z$

$= P_\Omega (A_1 u)^T z = 0$

Since $u \in N(A_1) \cap \Omega$. $\Rightarrow x \in \Omega \cap W^\perp$

$\therefore C(P_\Omega A^\perp) \subseteq W^\perp \cap \Omega$.

$(\subseteq)$: $x \in W^\perp \cap \Omega$

$= \{N(A_1) \cap \Omega\}^\perp \cap \Omega$

$= \{N(A_1)^\perp + \Omega^\perp\} \cap \Omega$

$= \{C(A_1^T) + \Omega^\perp\} \cap \Omega$

$x = P_\Omega \{A_1^T z + (I - P_\Omega) v\}$ for some vectors z & v.

$\Rightarrow x = P_\Omega A_1^T z \in C(P_\Omega A_1^T)$ □.

Apply these results to obtain constrained L.S.E. for β in the following problem:

$$\min \|Y - X\beta\|_2^2 \quad \text{s.t. } A\beta = c$$

Let $A\beta_0 = c$, $\mathbb{E}Y = X\beta$, $\mathbb{E}\tilde{Y} = \mathbb{E}(Y - X\beta_0)$

$= X(\beta - \beta_0) = X\gamma$

Further, $\in C(X) = \Omega$

$A(X^TX)^{-1}X^TX\gamma = A\beta - A\beta_0$

$\beta, \beta_0 \in \Omega \cap N(A_1)$ where $A_1 = A(X^TX)^{-1}X^T$

$\therefore$ By the principle of L.S.
the sought L.S.e. of β is

$$\hat{\theta}_{H} = X\hat{\beta}_{H} = P_{\omega}\tilde{Y}$$

$$= (P_{\Omega} - P_{\omega^{\perp}\cap\Omega})\tilde{Y}$$

$$= P_{\Omega}\tilde{Y} - P_{C(P_{\Omega}A_1)}\tilde{Y}$$

obtain $\hat{\beta}_{H}$ as a fct. of

$\hat{\beta}$, A, C, & X.

Lecture 21 Constrained Least Squares.

- $\mathbb{E}Y = X\beta = \theta \in C(X) \triangleq \Omega$

constraints are

$$A\beta = C \; ; \; q = r(\underset{q\times p}{A})$$

Let β_0 be a particular sol. s.t. $A\beta_0 = C$.

- Reparametrize:

$$\mathbb{E}(Y - X\beta_0) = X\beta - X\beta_0 = X\gamma, \; \gamma = \beta - \beta_0$$

$$\mathbb{E}\tilde{Y} = \underbrace{X\gamma}_{\theta} \in C(X) = \Omega \text{ where } \tilde{Y} = Y - X\beta_0.$$

Then $\underline{A(X^TX)^{-1}X^T}\underbrace{\theta}_{X\gamma} = A(X^TX)^{-1}X^TX\gamma = 0.$

$\Rightarrow \boxed{\theta \in \mathcal{N}(A(X^TX)^{-1}X^T\theta) \cap \Omega}$

$\therefore \hat{\theta} = P_W\tilde{Y}$ where $W = \mathcal{N}(A(X^TX)^{-1}X^T\theta) \cap \Omega$.

$= (P_\Omega - P_{\Omega \cap W^\perp})\tilde{Y}$, Call $A_1 := A(X^TX)^{-1}X^T$

$= P_{C(X)}\tilde{y} - P_{C(P_\Omega A_1^T)}\tilde{y}$

$= X(X^TX)^{-1}X^T\tilde{y} - P_\Omega A_1^T(A_1 P_\Omega A_1^T)^{-1}A_1 P_\Omega \tilde{y}$

NOTE that $\hat{\theta} = X[\hat{\beta}_H - \beta_0]$, $X\tilde{\beta} = X(X^TX)^{-1}X^Ty$. thus,

$$\Rightarrow X\hat{\beta}_H = X\tilde{\beta} - X(X^TX)^{-1}A^T(A(X^TX)^{-1}A^T)^{-1}(A\tilde{\beta} - c).$$

Since $r(\underset{n\times p}{X}) = p$, premultiply by $(X^TX)^{-1}X^T$:

- $\boxed{\hat{\beta}_H = \tilde{\beta} - (X^TX)^{-1}A^T(A(X^TX)^{-1}A^T)^{-1}(A\tilde{\beta} - c)}$.

Q: why $A(X^TX)^{-1}A^T$ invertible?

- Recall $\underset{q\times p}{A}$ has rank q.

$$\left.\begin{aligned}
&z^T A(X^TX)^{-1}A^T z = 0 \text{ iff} \\
&y^T(X^TX)^{-1}y = 0,\ y = A^T z \text{ iff} \\
&y = 0 \text{ or } A^T z = 0 \text{ iff} \\
&z = 0 \text{ since } r(A) = q.
\end{aligned}\right\} \begin{aligned}
&\Rightarrow A(X^TX)^{-1}A^T \text{ positive definite} \\
&\Rightarrow \lambda(A(X^TX)^{-1}A^T) > 0.
\end{aligned}$$

- $\underset{n\times p}{A}\,\underset{p\times q}{B}$, $r(A)=p$. $r(B)=q$. $r(AB)=?$

Lec 24 Linear Models when X has less than full Col. rank

$$\underset{n\times1}{Y} = \underset{n\times p}{X}\underset{p\times1}{\beta} + e, \quad r(X) = r < p$$

$\mathbb{E}e = 0$ & $Cov(e) = \sigma^2 I_n$.

- Recall $X^TX\beta = X^TY$ (normal equations)

Always has a sol. Since $r(X^TX \mid X^TY) = r(X^TX)$.

Sol. is $\hat{\beta} = (X^TX)^- X^TY$.

Let $\theta = \mathbb{E}Y = X\beta$, then $\hat{\theta} = X\hat{\beta} = X(X^TX)^- X^TY$

- $\hat{\theta}$ is the O.P. of Y on $\Omega = C(X)$

Further, $X(X^TX)^- X^T$ is unique.

Q: Can I find an unbiased $\underset{\text{(Linear)}}{}$ estimator of $\underset{p\times1}{\beta}$. i.e. is there a matrix C s.t. $\mathbb{E}CY = \beta$?

A: not possible.

$\boxed{LHS = \mathbb{E}CY = CX\beta = \beta \ \forall \beta \Rightarrow \underset{p\times n}{C}\underset{n\times p}{X} = \underset{p\times p}{I}, \text{ But } rank(CX) \le r < p}$

NOTE: If a^TY is unbiased for $c^T\beta$, then:

$$a^TX\beta = \mathbb{E}a^TY = c^T\beta \quad \forall \beta.$$

$$\Rightarrow X^Ta = c. \text{ or } c \in N(X)^{\perp}$$

$$\text{or } c \in C(X^T)$$

$\boxed{\text{DEF}}$ $c^T\beta$ is estimable if $\exists\, a$ s.t. $\mathbb{E}\, a^TY = c^T\beta$.

$\boxed{\text{THM}}$ $c^T\beta$ is estimable iff $c^T = c^T(X^TX)^- X^TX$.

$(\Leftarrow)$: $\mathbb{E}c^T(X^TX)^- X^TY = c^T(X^TX)^- X^TX\beta \underset{\text{ass.}}{=} c^T\beta$, Let $a^T = c^T(X^TX)^- X^T$, then

$\mathbb{E}a^TY = c^T\beta$.

($\Rightarrow$): If $c^T\beta$ is estimable, then $c = X^T\gamma$ for some γ,

$$c^T(X^TX)^-X^TX = \gamma^TX(X^TX)^-X^TX = \gamma^TP_\Omega X = \gamma^TX = c^T.$$

Ex $Y_{ij} = \alpha + \tau_i + \varepsilon_{ij}, \quad j=1,2,\cdots,n_i;\ i=1,2,\cdots,K.$

$$\underset{Kn_K\times 1}{\mathbb{E}Y} = X\beta = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & \vdots & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix}\alpha\\ \tau_1 \\ \vdots \\ \tau_K\end{bmatrix}, \quad \underbrace{\alpha+\tau_1+\tau_2}_{\text{NOT Interested}} = c^T\beta = [1\ 1\ 1\ 0]\begin{pmatrix}\alpha\\ \tau_1\\ \vdots \\ \tau_K\end{pmatrix}$$

- $c^T = c^T(X^TX)^-X^TX$

Find all c^T's s.t. the above eq. holds.

If $\underset{n\times n}{A} = A^T$ & A has rank r.

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22}\end{pmatrix}, \quad A^- = \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & 0\end{pmatrix}$$

$\underbrace{r(A_{11}) = r}$

Answer: If $c^T = (c_1, \cdots, c_p)$. $c^T = (\sum_{i=2}^{K} c_i, c_2, \cdots, c_p)$.

G inverses: $AA^-A = A$. A^- is a g-inv if eq. holds

verify (1) $r(A) = r(AA^-) = r(A^-A)$

(2) $r(A) \le r(A^-)$

(3) $r(A) = r(AA^+) = r(AA^+)$.

If A has full ~~row~~ col. rank, $A^+ = (A^TA)^{-1}A$.

If A has full row rank, $A^+ = A^T(AA^T)^{-1}$.

Q: If $A = A^T$, $(A^-)^T = A^-$? $AA^-A = A \Rightarrow$

$(AA^-A)^T = A^T(A^-)^TA^T = A(A^-)^TA = A = A^T$.

$\Rightarrow (A^T)^- = (A^-)^T$, but $(A^T)^- = A^-$. □.

- $(AB)^- = B^-A^-$? If so, $(AB)(AB)^-(AB) = (AB)$, $ABB^-A^-AB = AB$

If $\begin{cases} B \text{ has full } \text{row}~~\text{col}~~ \text{rank}, \\ A \text{ column} \end{cases}$ $B^- = B^T(BB^T)^{-1}$, $A^- = (A^TA)^{-1}A^T$, then $ABB^T(BB^T)^{-1}(A^TA)^{-1}A^TAB = AB$.

LEC 26: More on $r(X)<p$
LEC 28: OP & Hypothesis Testing
LEC 29: HT: Ex

$\mathbb{E}Y = X\beta$, $r(X)=p$,

$H_0: \underset{q\times p}{A}\,\underset{p\times 1}{\beta} = c$,

$$\hat\beta_H = \underset{A\beta=c}{\arg\min} \|Y-X\beta\|_2^2$$

$$\hat\beta = \underset{\theta\in C(X)}{\arg\min} \|Y-\underbrace{X\beta}_{\theta}\|_2^2$$

Express $\hat\beta_H$ in terms of $\hat\beta$.

$$\|Y-X\hat\beta_H\|_2^2 = \|\underbrace{Y-X\hat\beta}_{\in\Omega^\perp} + \underbrace{X\hat\beta - X\hat\beta_H}_{\in\Omega}\|_2^2$$

$$\|X\hat\beta - X\hat\beta_H\|_2^2 = SSE_0 - SSE$$

$$= (A\hat\beta - A\beta)^T \frac{A(X^TX)^{-1}A^T}{\sigma^2}(A\hat\beta-A\beta) \quad (\Delta)$$

$$A\hat\beta - c \sim \mathcal{N}(A\beta-c,\ \sigma^2 A(X^TX)^{-1}A^T)$$

$$\Rightarrow (\Delta) \sim \chi^2_q(\delta^2)$$

- $$\frac{(SSE_0-SSE)/q}{SSE/(n-q)} \overset{H_0}{\sim} F_{q,n-p}(0)$$

$$\begin{cases} Y = \hat\alpha_1 + \hat\beta_1 X \\ Y = \hat\alpha_2 + \hat\beta_2 X. \end{cases}$$

Problem 1: $\beta_1 = \beta_2$?

- $Y_{ki} = \alpha_k + \beta_k X_{ki} + \varepsilon_{ki}$, $i=1,\dots,n_k$, $k=1,\dots,K$

$$\begin{bmatrix} Y_{11}\\ \vdots \\ Y_{1n_1} \\ \vdots \\ Y_{k1}\\ \vdots\\ Y_{kn_k}\end{bmatrix} = Y = \begin{bmatrix} 1 & 0 & X_{11} & 0 \\ \vdots & & \vdots & \\ 1 & & X_{1n_1} & \\ & 1 & \cdots & X_{k1}\\ & \vdots & & \vdots \\ & 1 & & X_{kn_k}\end{bmatrix}\begin{bmatrix}\alpha_1\\ \vdots\\ \alpha_k\\ \beta_1\\ \vdots \\ \beta_k\end{bmatrix}$$

$$\underset{(k-1)\times 2k}{A}\,\underset{2k\times 1}{\beta} = \left[\begin{array}{c|cccc} 0 & 1 & & & -1 \\ & & 1 & & -1 \\ & & & \ddots & \vdots \\ & & & 1 & -1 \end{array}\right]\begin{bmatrix}\alpha_1\\ \vdots\\ \alpha_k\\ \beta_1\\ \vdots \\ \beta_k\end{bmatrix}$$

$r(A) = k-1$, $p = 2k$, $n = N$

$SSE_0 - SSE$, $(A\beta=0)$ $\quad F = \dfrac{(SSE_0-SSE)/(k-1)}{SSE/(N-2k)}$

$$SSE_0 = \sum_i\sum_j (Y_{ij} - \alpha_i - \beta X_{ij})^2$$

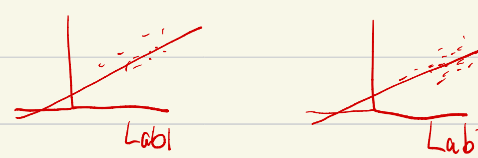
$$SSE = \sum_i\sum_j (Y_{ij} - \alpha_i - \beta_i X_{ij})^2$$

Problem 2: Test Reg lines:

$$0 = A\beta = \left(\begin{array}{c|c} 0 & \begin{matrix} 1 & \cdots & -1 \\ & \ddots & \vdots \\ & 1 & -1 \end{matrix} \\ \hline \begin{matrix} 1 & \cdots & -1 \\ & \ddots & \vdots \\ & 1 & -1 \end{matrix} & 0 \end{array}\right)\begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_k\\ \beta_1\\ \vdots \\ \beta_k\end{pmatrix}$$

$2(k-1)\times 2k$

Problem 3:



$\mathbb{E}Y_i = \alpha_k + \beta_k x_i$

$H_0: \alpha_i + \beta_i \phi$ = constant for $i=1,2,\dots,k$.

Testing hypothesis when $r(X)=r<p$.

Test $A\beta = c$. (H_0). $(\mathbb{E}Y = X\beta = \theta \in \Omega = C(X))$

Under H_0: $\theta \in \omega \subseteq \Omega$. & ω is a $(r-q)$-dim sbsp.

$$SSE = Y^T(I-P_\Omega)Y$$

$$SSE_0 = Y^T(I-P_\omega)Y$$

$$SSE_0 - SSE = Y^T(P_\Omega - P_\omega)Y$$

$\therefore\ \varepsilon^T(P_\Omega - P_\omega)\varepsilon \overset{H_0}{\sim} \chi^2_q(0)$

$\varepsilon^T(I - P_\Omega)\varepsilon \overset{H_0}{\sim} \chi^2_{n-r}$

Under H_0: $\dfrac{(SSE_0 - SSE)/q}{SSE/(n-r)} \sim F_{q,n-r}(0)$

If $W = N(A(X^TX)^{-1}X^T) \cap \Omega$, when $r(X) = p$.

Then we obtain earlier results.

LEC 30. Coefficient of Determination

REVIEW & Questions:

$$\mathbb{E}Y = X\beta,\quad r(\underset{n\times p}{X}) = r < p$$

$$\text{s.t. } A\beta = c$$

Find Constrained LSE for β. ($\hat\beta_H$)

We showed

$$\underset{P_\omega Y}{\underset{\|}{X\hat\beta_H}} = \underset{P_\Omega Y}{\underset{\|}{X\hat\beta}} - X(X^TX)^- A^T(A(X^TX)^-A^T)^{-1}(A\hat\beta - c)$$

Recall if $r(X) = p$.

$$\hat\beta_H = \hat\beta - (X^TX)^{-1}A^T(A(X^TX)^{-1}A^T)^{-1}(A\hat\beta - c)$$

where $\hat\beta = (X^TX)^{-1}X^Ty$

$$X\hat\beta_H = X\hat\beta - X(X^TX)^{-1}d,\quad d = A^T(A(X^TX)^-A^T)^{-1}(A\hat\beta - c)$$

$$\&\quad \hat\beta_H = (X^TX)^-(X^TX\hat\beta - X^TX(X^TX)^- d)$$

$$(Ax = b \Rightarrow x = A^-b,\ \hat x = A^-b + (I - A^-A)z)$$

$$\Rightarrow \hat\beta_H = (X^TX)^- X^Ty - X^TXd$$

- X^TX Sym $\Rightarrow \exists T$ s.t. $T^TX^TXT = \begin{pmatrix} D_r & 0 \\ 0 & 0 \end{pmatrix}$

$$\Rightarrow (X^TX)^- = \begin{pmatrix} D_r^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

⑤ $Y = \theta + \varepsilon,\ \varepsilon \sim \mathcal{N}_4(0, \sigma^2 I_4)$

$1^T\theta = 0,\ \theta = (\theta_1\ \theta_2\ \theta_3\ \theta_4)$

$H_0: \theta_1 = \theta_3.$

$X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{pmatrix},\quad Y = X\begin{pmatrix}\theta_1\\ \theta_2 \\ \theta_3\end{pmatrix} + \varepsilon$

- $0 = A\beta = (1\ \ 0\ \ -1)\begin{pmatrix}\theta_1\\ \theta_2 \\ \theta_3\end{pmatrix}$

Coefficient of Determinant R^2

$X_i^T\hat\beta = \hat Y_i,\ Y_i,\ i = 1, 2, \cdots, n$

$Corr(Y, \hat Y) = \dfrac{\sum_i (Y_i - \bar Y)(\hat Y_i - \bar{\hat Y})}{\sqrt{\sum_i (Y_i - \bar Y)^2}\sqrt{\sum_i (\hat Y_i - \bar{\hat Y})^2}}$

Define $R^2 = r^2$

$\hat Y = \begin{pmatrix}\hat Y_1 \\ \vdots \\ \hat Y_n\end{pmatrix}$ is model-based Since

$\hat Y_i = X_i^T\hat\beta = (\text{———})\hat\beta$

$\bar Y = \sum_i Y_i / n$ is not model-based.

① $\hat Y = P_\Omega Y,\ e = Y - \hat Y = (I - P_\Omega)Y = QY$

$\sum_i e_i = 1^T e = 1^T Q Y = 1^T(I - P_\Omega)Y$

$= (1^T - (P_\Omega 1)^T)Y = 0$ if $1 \in \Omega$

② $e^T(PY) = (QY)^T PY = Y^T \underline{Q^T P} Y = 0$

③ $1^T(PY)$ = sum of Predicted value

$= (P1)^T Y = 1^T y$ if $1 \in \Omega.$

Recall: $SSE = TSS - SSReg.$

Verify: $R^2 = \dfrac{\sum(\hat Y_i - \bar Y)^2}{\sum(Y_i - \bar Y)^2} = 1 - \dfrac{SSE}{\sum(Y_i - \bar Y)^2}$

(Ex) $\mathbb{E}Y = X\beta = \underset{n\times p}{X}\begin{pmatrix}\beta_0\\ \vdots \\ \beta_{p-1}\end{pmatrix}$

Overall Test $\quad \underset{(p-1)\times p}{A}\beta = 0,\ A = \begin{pmatrix} 0 & \\ \vdots & I_{p-1} \\ 0 & \end{pmatrix}$

Omibus

$F = \dfrac{(SSE_0 - SSE)/q}{SSE/(n-p)},\quad q = p-1.$

$SSE_0 = \sum_i (Y_i - \hat Y_i)^2$

$= \sum_i (Y_i - \hat\beta_0)^2 = \sum_i (Y_i - \bar Y)^2$

$\Rightarrow F = \dfrac{R^2}{1-R^2}\ \dfrac{n-p}{p-1}$

Rej. H if $F > F_{p-1, n-p, \alpha}$