① Examples of Linear Models

② Linear Algebra : - Vector space

- Projection

\+ LSE

③ Estimability , Imposing Conditions

④ Gauss Markov -

Linear Algebra : Spectral decomposition of square, symmetric positive definite matrix

$Cov(e) = \sigma^2 V$

- HW # 1 & #2 w/ solutions
- HW#3
- Exam 1 : 04/26 (Tue)

$$-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)$$

♣ Def: A *quadratic form* in the n variables is $\mathbf{x}^T\mathbf{A}\mathbf{x}$ where n−dim vector $\mathbf{x}$ and $\mathbf{A}$ is a $n \times n$ symmetric matrix.

Observe

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \sum_{i,j} a_{ij}x_ix_j,$$

that is, it has only squared terms x_i^2 and product terms, x_ix_j.

† Th: A symmetric matrix $\mathbf{A}$ is *positive* (*nonnegative*) *definite* if, for nay nonzero vector $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{v}^T\mathbf{A}\mathbf{v}$ is positive (nonnegative).

$$0 \le \mathbf{x}^T\mathbf{A}\mathbf{x} \qquad \text{for all } \mathbf{x},$$
$$0 < \mathbf{x}^T\mathbf{A}\mathbf{x} \qquad \text{for all } \mathbf{x}(\neq \mathbf{0}).$$

① Let $A = I$

$x^T I x = x_1^2 + \cdots + x_n^2 > 0$ for all x $(x \neq 0)$

I: p.d

② Let $A = J_{n\times n} = \underset{n\times 1}{1_n}\,\underset{1\times n}{1_n^T}$ $\quad 2^T 2 = 2^2$

$x^T J x = x^T 1_n 1_n^T x = (\underbrace{1_n^T x}_{(x_1+\cdots+x_n)})^T (1_n^T x) = (x_1+\cdots+x_n)^2 \geq 0$

for any x,

$x = [1-n, 1, \cdots, 1]^T \Rightarrow x^T J x = 0$

$\Rightarrow J$: nonnegative definite

- Let's connect the spectral decomposition of a symmetric matrix **A** into $\mathbf{A}^{-}$.

$\Lambda = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \quad \Lambda^{-1} = \begin{bmatrix} 1/2 & 0 \\ 0 & 0 \end{bmatrix}$

From the previous slide,

$$\mathbf{A}^{-1} = \mathbf{Q}\Lambda^{-1}\mathbf{Q}^T$$

where Λ is the diagonal matrix with λ_j.

† Th: Let **A** be a $n \times n$ symmetric matrix of rank r ($r \leq n$). Let Λ_r be the diagonal matrix containing its nonzero eigenvalues (in decreasing order of magnitude), and let $\mathbf{Q}_r$ be the $n \times r$ matrix whose columns are the eigenvectors corresponding to the nonzero λ_j of **A**. Then,

$$\mathbf{A}^{-} = \boxed{\mathbf{Q}_r}\,\overbrace{\Lambda_r^{-1}}^{r\times r}\,\overbrace{\mathbf{Q}_r^T}^{r\times n}.$$

$n \times r$

Or this is the same as letting Λ^{-1} with $1/\lambda_j$ for $\lambda_j \neq 0$ and 0 for $\lambda_j = 0$.

$$= \underbrace{Q_r \Lambda_r^{-1/2} Q_r^T}_{(A^{1/2})^{-}}\, \underbrace{Q_r}_{I_r} \underbrace{\Lambda_r^{-1/2} Q_r^T}_{(A^{1/2})^{-}}$$

† The form of linear models is

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e},$$

- $\mathbf{y}$: $n \times 1$ vector of observations (random)
- $\mathbf{X}$: $n \times p$ matrix of known constants (*design* matrix) with $r(\mathbf{X}) = r$
- β: $p \times 1$ vector of unobservable parameters
- $\mathbf{e}$: $n \times 1$ vector of unobservable random errors

★ Assumptions (Gauss-Markov model assumption):

- $\mathrm{E}(\mathbf{e}) = \mathbf{0}$ ($\Leftrightarrow \mathrm{E}(\mathbf{y}) = \mathbf{X}\beta$)
- $\mathrm{Cov}(\mathbf{e}) = \sigma^2 I$ where σ^2 is some unknown parameter ($\sigma^2 > 0$)

$\mathbf{y}$: $n \times 1$ random vector

$$E(\mathbf{y}) = \begin{bmatrix} E(y_1) \\ \vdots \\ E(y_n) \end{bmatrix} = \mu$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$\underbrace{Cov(\mathbf{y}, \mathbf{y})}_{n \times n} = Var(\mathbf{y}) = E(\underbrace{(\mathbf{y} - \mu)}_{n \times 1}(\mathbf{y} - \mu)^T)$$

$$Cov(\underset{n \times 1}{y}, \underset{s \times 1}{W}) = E(\underbrace{(y - \mu_y)}_{n \times 1}(W - \mu_W)^T)$$

$n \times s$

♠ Remind!

- $\mathrm{E}(\mathbf{a}^T\mathbf{y}) = \mathbf{a}^T\mathrm{E}(\mathbf{y})$
- $\mathrm{Var}(\mathbf{a}^T\mathbf{y}) = \mathbf{a}^T\mathrm{Cov}(\mathbf{y})\mathbf{a}$ for fixed $\mathbf{a}$
- $\mathrm{Cov}(\mathbf{a}^T\mathbf{y}, \mathbf{c}^T\mathbf{y}) = \mathbf{a}^T\mathrm{Cov}(\mathbf{y})\mathbf{c}$ for fixed $\mathbf{a}$ and $\mathbf{c}$
- $\mathrm{Cov}(\mathbf{A}^T\mathbf{y}) = \mathbf{A}^T\mathrm{Cov}(\mathbf{y})\mathbf{A}$ for fixed $\mathbf{A}$

$Cov(A^T y + b) = A^T Cov(y) A$

- $Cov(y)$ is nonnegative definite for any random ~~variable~~ vector y

Claim: for arbitrary vector u, $u^T Cov(y) u \geq 0$

$$u^T Cov(y) u = u^T E((y - \mu)(y - \mu)^T) u$$

$$= E(\underbrace{(\underbrace{(y-\mu)^T}_{\substack{n \times 1 \\ 1 \times n}} \underset{n \times 1}{u})^T}_{} \underbrace{(y-\mu)^T u}_{= s})$$

$$= E(s^2) \geq 0$$

- y is nonsingular if $Cov(y)$ is nonsingular
  y is singular if $cov(y)$ is singular

◇ Cor For a symmetric matrix $\mathbf{A}$, there exists $\mathbf{A}^-$ such that $\mathbf{A}^-\mathbf{A}\mathbf{A}^- = \mathbf{A}^-$ and $(\mathbf{A}^-)^T = \mathbf{A}^-$.

$A^- = Q_r \Lambda_r^{-1} Q_r^T$

In layman's terms, there exists a reflexive ($\mathbf{A}^-\mathbf{A}\mathbf{A}^- = \mathbf{A}^-$), symmetric generalized inverse for any symmetric matrix $\mathbf{A}$.

$E(\lambda^T\hat{\beta}) = \lambda^T\beta$

$= E(\lambda^T (X^TX)^- X^Ty) = E(a^TX(X^TX)^- X^T y)$

$= a^T \underbrace{X(X^TX)^- X^TX}_{=X}\beta = a^TX\beta$

$= \underline{\underline{\lambda^T\beta}}$

- $XGX^TX = XHX^TX = X$

(G & H are a g-inverse of (X^TX))

♠ Consider a linear model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e},$$

$\lambda \in \mathcal{C}(X^T)$

$\exists\, a$ s.t. $\lambda = X^Ta$

$(\Leftrightarrow \lambda^T = a^TX)$

where $\mathrm{E}(\mathbf{e}) = \mathbf{0}$ and $\mathrm{Var}(\mathbf{e}) = \sigma^2 I$. Suppose $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is estimable. Find the variance of $\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}$.

$\mathrm{Var}(\lambda^T\hat{\beta}) = \mathrm{Var}(\lambda^T(X^TX)^- X^Ty) = (\lambda^T(X^TX)^- X^T)\,\overbrace{\mathrm{Var}(y)}^{\sigma^2 I}\,(\lambda^T(X^TX)^- X^T)^T = \sigma^2\lambda^T \underbrace{(X^TX)^- X^TX\{(X^TX)^-\}^T}_{(X^TX)^-}\lambda$

$= \underline{\sigma^2\lambda^T(X^TX)^-\lambda}$

$\star$ Note: From Exercise 4.2, $\mathrm{Var}(\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}})$ does not depend on the choice of g-inverse $(\mathbf{X}^T\mathbf{X})^-$.

$$\mathrm{Var}(d^T y) \ge \mathrm{Var}(d^T y - a^T P y) + \mathrm{Var}(a^T P y)$$
$$= \underbrace{\mathrm{Var}(d^T y - \lambda^T \hat\beta)}_{\ge 0} + \mathrm{Var}(\lambda^T \hat\beta)$$
$$\ge \mathrm{Var}(\lambda^T \hat\beta)$$

$$E(d^T y) = \lambda^T \beta$$
$$\Leftrightarrow \quad d^T X \beta = \lambda^T \beta$$
$$\Leftrightarrow \quad d^T X = \lambda^T = a^T X$$

The Gauss-Markov Theorem

♣ Th 4.1: Under the assumptions of the Gauss-Markov model,

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e}, \quad \text{where} \quad E(\mathbf{e}) = \mathbf{0}, \ \ \mathrm{Var}(\mathbf{e}) = \sigma^2 I,$$

if $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is estimable, then $\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}$ is the best (*minimum variance*) linear unbiased estimator (BLUE) of $\boldsymbol{\lambda}^T\boldsymbol{\beta}$, where $\hat{\boldsymbol{\beta}}$ solves the NEs $\mathbf{X}^T\mathbf{X}\beta = \mathbf{X}^T\mathbf{y}$.

Suppose $d^T y$ is another unbiased estimator of $\lambda^T \beta$

Goal: want to show $\mathrm{Var}(d^T y) \ge \mathrm{Var}(\lambda^T \hat\beta)$.

since $\lambda^T\beta$ is estimable, $\lambda \in \mathcal{C}(X^T) \Leftrightarrow \exists\, a$ s.t. $\lambda = X^T a$ $(\Leftrightarrow \lambda^T = a^T X)$

$$\lambda^T\hat\beta = \lambda^T (X^T X)^- X^T y = a^T \underbrace{X (X^T X)^- X^T}_{=P} y = a^T P y$$

$$\mathrm{Var}(d^T y) = \mathrm{Var}(d^T y - \overbrace{a^T P y}^{=\lambda^T\hat\beta} + a^T P y) \qquad (*) = 0$$

$$= \underbrace{\mathrm{Var}(d^T y - a^T P y)}_{\ge 0} + \underbrace{\mathrm{Var}(a^T P y)}_{= \mathrm{Var}(\lambda^T\hat\beta)} + \overbrace{2\,\mathrm{Cov}(d^T y - a^T P y,\ a^T P y)}$$

$= (d^T - a^T P) y$ $\qquad = \sigma^2 I$

$$(*) = \mathrm{Cov}(d^T y - a^T P y,\ a^T P y) = (d^T - a^T P)\,\mathrm{Var}(y)\,(a^T P)^T$$
$$= \sigma^2 (d^T - a^T P) P a = \sigma^2 \underbrace{(d^T P - a^T P)}_{=0} a = 0$$

will show

$$d^T P - a^T P = \underbrace{d^T X}_{= a^T X} (X^T X)^- X^T - a^T X (X^T X)^- X^T = 0$$

$r(X)=r$

♠ Let's consider multiple $\boldsymbol{\lambda}_j$, $j = 1, \ldots, m$ together. Let the columns $\boldsymbol{\lambda}_j$ of Λ ($p \times m$) where $\boldsymbol{\lambda}_j$ are linearly independent; then

- The LSE of $\Lambda^T\beta$ is $\Lambda^T\hat{\beta}$ (BLUE) $\Lambda^T = \begin{bmatrix} \lambda_1^T \\ \vdots \\ \lambda_m^T \end{bmatrix}$
- Its variance is

$$\begin{aligned}
\mathrm{Cov}(\Lambda^T\hat{\beta}) &= \mathrm{Cov}(\Lambda^T(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y}) \\
&= \sigma^2\Lambda^T(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\Lambda \\
&= \sigma^2\Lambda^T(\mathbf{X}^T\mathbf{X})^-\Lambda.
\end{aligned}$$

For any unbiased estimator, $\mathbf{C}^T\mathbf{y}$,

$$\mathrm{Cov}(\mathbf{C}^T\mathbf{y}) - \mathrm{Cov}(\Lambda^T\hat{\beta}) \geq 0.$$

$\Rightarrow \Lambda^T\hat{\beta}$ is the BLUE.

$\beta_0 = E(y)$ at $X=0$

$y_i = \beta_0 + \beta_1 x_i + e_i$

$\Leftrightarrow \quad y_i = \beta_0 + \beta_1(x_i - \bar{x} + \bar{x}) + e_i$

$= \underbrace{\beta_0 + \bar{x}\beta_1} + \beta_1(x_i - \bar{x}) + e_i$

$= \gamma_0 + \gamma_1 x_i^* + e_i$

$\gamma_0 = E(y)$ at $x^* = 0 \Leftrightarrow x_i = \bar{x}$

height vs weight

$x \quad y$

♠ Ex 1 Consider a simple linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + e_i,$$

where $E(\mathbf{e}) = \mathbf{0}$ and $Var(\mathbf{e}) = \sigma^2 I$. Find the variance of $\hat{\boldsymbol{\beta}}$. $= \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$

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

$= I$

$\hat{\beta} = \Lambda^T \hat{\beta} = \hat{\beta}$

$Var(\hat{\beta}) = \sigma^2 \Lambda^T (X^T X)^- \Lambda$

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

$= \sigma^2 \left(\begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \right)^{-1}$

$= \sigma^2 \begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix}^{-1}$

$= \dfrac{\sigma^2}{n\Sigma x_i^2 - (\Sigma x_i)^2} \begin{bmatrix} \Sigma x_i^2 & -\Sigma x_i \\ -\Sigma x_i & n \end{bmatrix}$

$= \dfrac{\sigma^2}{n \underbrace{\Sigma (x_i - \bar{x})^2}_{= S_{xx}}} \begin{bmatrix} \Sigma x_i^2 & -\Sigma x_i \\ -\Sigma x_i & n \end{bmatrix}$

$Var(\hat{\gamma})$

- Now let's estimate σ^2. $\hat{e} = (I-P)y$

Recall $SSE = ||\hat{\mathbf{e}}||^2 = \hat{\mathbf{e}}^T\hat{\mathbf{e}} = \mathbf{y}^T(I-\mathbf{P})\mathbf{y}$.

★ Result 4.2: Consider the Gauss-Markov model. An unbiased estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{SSE}{n-r} = \frac{\mathbf{y}^T(I-\mathbf{P})\mathbf{y}}{n-r},$$

where $r = rank(\mathbf{X})$.

$E(\hat{\sigma}^2) = \sigma^2$ $\qquad E(z^T A z) = \mu^T A \mu + tr(A\Sigma)$

$\Leftrightarrow E(SSE) = (n-r)\sigma^2$

$\Rightarrow E(SSE) = E(\underbrace{y}_{=z}^T \overbrace{(I-P)}^{=A} y)$

$= (X\beta)^T(I-P)(X\beta) + tr((I-P)\cdot\sigma^2 I)$

$(X\beta \in C(X))$, $C(X)^\perp$ $n-r$, r

$= 0 + \sigma^2 \underbrace{tr(I-P)}_{=(n-r)} = \sigma^2(n-r)$

$(n-r)$ 1's and r 0's

$A = Q\Lambda Q^T$ by spectral decomposition

$A^2 = A \Leftrightarrow Q\Lambda Q^T = Q\Lambda \overbrace{Q^T Q}^{=I} \Lambda Q^T = Q\Lambda^2 Q^T$

$\lambda_i = \lambda_i^2 \Leftrightarrow \lambda_i = 0, 1$

$\#$ of nonzero $\lambda_i = r(A)$

Lemma 4.1 Let z be a vector random variable with $\mathrm{E}(z) = \mu$ and $\mathrm{Cov}(z) = \Sigma$. Then $\mathrm{E}(z^T \mathbf{A} z) = \mu^T \mathbf{A} \mu + tr(\mathbf{A}\Sigma)$.

- The Aitken Model and Generalized Least Squares (GLS)

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e}, \quad \text{where} \quad \mathrm{E}(\mathbf{e}) = \mathbf{0}, \;\; \mathrm{Var}(\mathbf{e}) = \sigma^2\mathbf{V}.$$

Here we assume $\mathbf{V}$ is a known positive definite matrix.

♣ What is coming next?

◇ Let's find an estimator (generalized least square estimator) for an estimable function, $\boldsymbol{\lambda}^T\beta$, $\boldsymbol{\lambda}^T\hat{\beta}_{GLS}$.

◇ Then we will show it is the BLUE for $\boldsymbol{\lambda}^T\beta$.

- Recall the Aitken model,

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e}, \quad \text{where} \quad \mathrm{E}(\mathbf{e}) = \mathbf{0}, \;\; \mathrm{Var}(\mathbf{e}) = \sigma^2\mathbf{V}. \tag{2}$$

Here we assume $\mathbf{V}$ is a known positive definite matrix. We can write $\mathbf{V}^{-1} = \mathbf{R}\mathbf{R}^T$ for some nonsingular matrix $\mathbf{R}$.

$V = R^{-1}(R^T)^{-1}$

In particular, consider $\mathbf{V} = \mathbf{Q}\Lambda\mathbf{Q}^T$ and let $\mathbf{R} = \mathbf{Q}\Lambda^{-1/2}\mathbf{Q}^T$ (square root matrix, symmetric & nonsingular).

$y = X\beta + e, \quad Cov(e) = \sigma^2 V$

- Now consider an equivalent model,

$$\underbrace{\mathbf{Ry}}_{y^*} = \underbrace{\mathbf{RX}}_{X^*}\beta + \underbrace{\mathbf{Re}}_{e^*}. \tag{3}$$

$\hat{y}^*$ becomes the orthogonal projection onto $\mathcal{C}(X^*) = \mathcal{C}(RX)$

Check!

$\surd$ E(**Re**) $= R\,E(e) = 0$

$\surd$ Var(**Re**) $= R\,Var(e)\underbrace{R^T}_{=R} = \sigma^2 R \underline{V} R$

$\underbrace{V^{1/2}}_{=R^{-1}} V^{1/2}$

$= \sigma^2 R\,(\cancel{V^{-1/2}})(\cancel{V^{-1/2}})\,R$

$= \sigma^2 R\,(R^{-1})\,R^{-1} R = \sigma^2 I$

$\Rightarrow$ Gauss-Markov model!!.

- Under transformed model, LSE minimizes

$$||\overbrace{\mathbf{Ry}}^{y^*} - \overbrace{\mathbf{RX}}^{X^*}\beta||^2 = (\mathbf{Ry} - \mathbf{RX}\beta)^T(\mathbf{Ry} - \mathbf{RX}\beta).$$

$$= (R(y - X\beta))^T R(y - X\beta)$$

$$= (y - X\beta)^T R^T R (y - X\beta)$$

$$= (y - X\beta)^T V^{-1} (y - X\beta) \leftarrow$$

$$\underline{X^T X \beta = X^T y}$$

$\Rightarrow$ the NEs becomes

$$(X^*)^T X^* \beta = (X^*)^T y^*$$

$$\Leftrightarrow \quad (RX)^T RX\beta = (RX)^T Ry$$

$$X^T R^T R X \beta = X^T R^T R y$$

$$X^T V^{-1} X \beta = X^T V^{-1} y \leftarrow$$

- How $\mathbf{V}$ works?

Minimizes

$$(\mathbf{y}-\mathbf{X}\beta)^T\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\beta).$$

In words, instead of minimizing the squared distance between $\mathbf{y}$ and $\mathbf{X}\beta$, we minimize a generalized distance determined by $\mathbf{V}^{-1}$.

★: Ex Consider a simple linear regression and $\mathbf{V}$ diagonal matrix. Observe

$$(\mathbf{y}-\mathbf{X}\beta)^T\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\beta) = \sum_i \frac{1}{v_{ii}}(y_i-\beta_0-\beta_1 x_i)^2$$

(Known as weighted least squares!)

$Var(e_i) = \sigma^2 V_{ii}$

$Cov(e_i, e_{i'}) = 0$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

$$\begin{bmatrix} y_1-\beta_0-\beta_1x_1 \\ \vdots \\ y_n-\beta_0-\beta_1x_n \end{bmatrix}^T \begin{bmatrix} \frac{1}{V_{11}} & & & 0 \\ & \frac{1}{V_{22}} & & \\ & & \ddots & \\ 0 & & & \frac{1}{V_{nn}} \end{bmatrix} \begin{bmatrix} y_1-\beta_0-\beta_1x_1 \\ \vdots \\ y_n-\beta_0-\beta_1x_n \end{bmatrix}$$