FDU 回归分析 期中考试 (2023 秋)

Total: 100 marks
Duration: 2 hour and 30 minutes

Problem 1

Consider the simple linear regression model $y_i = \alpha + \beta x_i + \varepsilon_i \ (i = 1, 2, \ldots, n)$ where $\{\varepsilon_i\}_{i=1}^n \overset{iid}{\sim} N(0, \sigma^2)$.

Denote $\begin{cases} \bar{y} = \frac{1}{n}\sum\limits_{i=1}^n y_i \\ s^2 = \frac{1}{n-2}\sum\limits_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} x_i)^2 \end{cases}$ where $(\hat{\alpha}, \hat{\beta})$ are LSE of parameters (α, β)

(1) Show that $\bar{y} - \mathrm{E}(\bar{y})$ is independent of $\hat{\beta} - \beta$ $[5 \text{ Marks}]$
(2) Show that $\bar{y}$ is independent of s^2 $[5 \text{ Marks}]$
(3) Find the confidence interval (with confidence level $1 - \tau$) of $\alpha + \beta x$, simultaneously for all x. $[15 \text{ Marks}]$

Preparation

我们记:

$$\begin{cases} \gamma = [\alpha, \beta]^{\mathrm{T}} \\ \varepsilon = [\varepsilon_1, \ldots, \varepsilon_n]^{\mathrm{T}} \\ x = [x_1, \ldots, x_n]^{\mathrm{T}} \\ y = [y_1, \ldots, y_n]^{\mathrm{T}} \\ X = [1_n, x] \in \mathbb{R}^{n \times 2} \end{cases}$$

则我们有:

$$y = X\gamma + \varepsilon$$

$$\begin{aligned} \hat{\gamma} &= (X^{\mathrm{T}} X)^{-1} X^{\mathrm{T}} y \\ &= \begin{bmatrix} 1_n^{\mathrm{T}} 1_n & 1_n^{\mathrm{T}} x \\ 1_n^{\mathrm{T}} x & x^{\mathrm{T}} x \end{bmatrix}^{-1} \begin{bmatrix} 1_n^{\mathrm{T}} y \\ x^{\mathrm{T}} y \end{bmatrix} \\ &= \frac{1}{n x^{\mathrm{T}} x - (1_n^{\mathrm{T}} x)^2} \begin{bmatrix} x^{\mathrm{T}} x & -1_n^{\mathrm{T}} x \\ -1_n^{\mathrm{T}} x & n \end{bmatrix} \begin{bmatrix} 1_n^{\mathrm{T}} y \\ x^{\mathrm{T}} y \end{bmatrix} \Rightarrow \begin{cases} \hat{\alpha} = \bar{y} - \bar{x}\hat{\beta} = \bar{y} - \bar{x}\frac{S_{xy}}{S_{xx}} \\ \hat{\beta} = \frac{S_{xy}}{S_{xx}} \end{cases} \\ &= \frac{1}{n(x^{\mathrm{T}} x - n\bar{x}^2)} \begin{bmatrix} x^{\mathrm{T}} x \cdot n\bar{y} - x^{\mathrm{T}} y \cdot n\bar{x} \\ n(x^{\mathrm{T}} y - n\bar{x}\bar{y}) \end{bmatrix} \\ &= \begin{bmatrix} \bar{y} - \bar{x}\frac{S_{xy}}{S_{xx}} \\ \frac{S_{xy}}{S_{xx}} \end{bmatrix} \end{aligned}$$

Part (1)

Show that $\bar{y} - \mathrm{E}(\bar{y})$ is independent of $\hat{\beta} - \beta$ $[5 \text{ Marks}]$

Solution:
我们有:

$$
\begin{aligned}
\hat{\beta} &= \frac{S_{xy}}{S_{xx}} \\
&= \frac{x^{\mathrm{T}}y - n\bar{x}\bar{y}}{S_{xx}} \\
&= \frac{(x - \bar{x}1_n)^{\mathrm{T}}y}{S_{xx}} \\
&= \frac{(x - \bar{x}1_n)^{\mathrm{T}}(\alpha 1_n + \beta x + \varepsilon)}{S_{xx}} \\
&= \frac{1}{S_{xx}}\{\alpha(x^{\mathrm{T}}1_n - \bar{x}1_n^{\mathrm{T}}1_n) + \beta(x^{\mathrm{T}}x - \bar{x}1_n^{\mathrm{T}}x) + (x - \bar{x}1_n)^{\mathrm{T}}\varepsilon\} \\
&= \frac{1}{S_{xx}}\{\alpha(n\bar{x} - n\bar{x}) + \beta(x^{\mathrm{T}}x - n\bar{x}^2) + (x - \bar{x}1_n)^{\mathrm{T}}\varepsilon\} \\
&= \frac{1}{S_{xx}}\{\alpha \cdot 0 + \beta \cdot S_{xx} + (x - \bar{x}1_n)^{\mathrm{T}}\varepsilon\} \\
&= \beta + \frac{(x - \bar{x}1_n)^{\mathrm{T}}\varepsilon}{S_{xx}}
\end{aligned}
$$

$$
\begin{aligned}
\bar{y} &= \frac{1}{n}1_n^{\mathrm{T}}y \\
&= \frac{1}{n}1_n^{\mathrm{T}}(\alpha 1_n + \beta x + \varepsilon) \\
&= \alpha + \beta\bar{x} + \frac{1}{n}1_n^{\mathrm{T}}\varepsilon \\
&= \mathrm{E}[\bar{y}] + \frac{1}{n}1_n^{\mathrm{T}}\varepsilon
\end{aligned}
$$

在正态假设下，要证明 $\bar{y} - \mathrm{E}[\bar{y}]$ 与 $\hat{\beta} - \beta$ 独立，等价于证明它们不相关:

$$
\begin{aligned}
\mathrm{Cov}(\bar{y} - \mathrm{E}[\bar{y}], \hat{\beta} - \beta) &= \mathrm{Cov}\left(\frac{1}{n}1_n^{\mathrm{T}}\varepsilon, \frac{(x - \bar{x}1_n)^{\mathrm{T}}}{S_{xx}}\varepsilon\right) \\
&= \frac{1}{n}1_n^{\mathrm{T}}\mathrm{Cov}(\varepsilon, \varepsilon)\frac{(x - \bar{x}1_n)}{S_{xx}} \\
&= \frac{1}{n}1_n^{\mathrm{T}} \cdot \sigma^2 I_n \cdot \frac{(x - \bar{x}1_n)}{S_{xx}} \\
&= \frac{\sigma^2}{S_{xx}}\left(\frac{1}{n}1_n^{\mathrm{T}}x - \frac{1}{n}1_n^{\mathrm{T}}1_n \cdot \bar{x}\right) \\
&= \frac{\sigma^2}{S_{xx}}(\bar{x} - \bar{x}) \\
&= 0
\end{aligned}
$$

因此 $\bar{y} - \mathrm{E}[\bar{y}]$ 与 $\hat{\beta} - \beta$ 相互独立.

Part (2)

Show that $\bar{y}$ is independent of s^2 $[5\ \mathrm{Marks}]$

Solution:

$$
\begin{aligned}
s^2 &= \frac{1}{n-2}\|y-\hat{y}\|_2^2 \\
&= \frac{1}{n-2}\|y-Hy\|_2^2 \quad (H = X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}) \\
&= \frac{1}{n-2}y^{\mathrm{T}}(I_n-H)^{\mathrm{T}}(I_n-H)y \\
&= \frac{1}{n-2}y^{\mathrm{T}}(I_n-H)y \\
&= \frac{1}{n-2}(X\gamma+\varepsilon)^{\mathrm{T}}(I_n-H)(X\gamma+\varepsilon) \\
&= \frac{1}{n-2}\varepsilon^{\mathrm{T}}(I_n-H)\varepsilon
\end{aligned}
$$

$$
\begin{aligned}
\bar{y} &= \frac{1}{n}1_n^{\mathrm{T}}y \\
&= \frac{1}{n}1_n^{\mathrm{T}}(\alpha 1_n+\beta x+\varepsilon) \\
&= \alpha+\beta\bar{x}+\frac{1}{n}1_n^{\mathrm{T}}\varepsilon \\
&= \mathrm{E}[\bar{y}]+\frac{1}{n}1_n^{\mathrm{T}}\varepsilon
\end{aligned}
$$

要证明 $\bar{y}$ 与 s^2 独立，只需证明 $\frac{1}{n}1_n^{\mathrm{T}}\varepsilon$ 与 $(I_n-H)\varepsilon$ 独立，在正态假设下，即证明它们不相关:

$$
\begin{aligned}
\mathrm{Cov}(\frac{1}{n}1_n^{\mathrm{T}}\varepsilon,(I_n-H)\varepsilon) &= \frac{1}{n}1_n^{\mathrm{T}}\mathrm{Cov}(\varepsilon,\varepsilon)(I_n-H) \\
&= \frac{1}{n}1_n^{\mathrm{T}}\cdot\sigma^2 I_n\cdot(I_n-H) \\
&= \frac{\sigma^2}{n}1_n^{\mathrm{T}}(I_n-H) \\
&= \frac{\sigma^2}{n}(1_n^{\mathrm{T}}-1_n^{\mathrm{T}}) \\
&= 0
\end{aligned}
$$

因此 $\bar{y}$ 与 s^2 独立.

Part (3)

Find the confidence interval (with confidence level $1-\tau$) of $\alpha+\beta x$, simultaneously for all x. $[15\ \text{Marks}]$

Solution:

$$
\mathrm{E}[\hat{\gamma}] = \gamma
$$

$$
\begin{aligned}
\mathrm{Cov}(\hat{\gamma}) &= \sigma^2(X^{\mathrm{T}}X)^{-1} \\
&= \sigma^2\begin{bmatrix} 1_n^{\mathrm{T}}1_n & 1_n^{\mathrm{T}}x \\ 1_n^{\mathrm{T}}x & x^{\mathrm{T}}x \end{bmatrix}^{-1} \\
&= \sigma^2\frac{1}{nx^{\mathrm{T}}x-(1_n^{\mathrm{T}}x)^2}\begin{bmatrix} x^{\mathrm{T}}x & -1_n^{\mathrm{T}}x \\ -1_n^{\mathrm{T}}x & n \end{bmatrix} \\
&= \sigma^2\frac{1}{nS_{xx}}\begin{bmatrix} x^{\mathrm{T}}x & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix} \\
&= \sigma^2\begin{bmatrix} \frac{1}{n}+\frac{\bar{x}^2}{S_{xx}} & -\frac{\bar{x}}{S_{xx}} \\ -\frac{\bar{x}}{S_{xx}} & \frac{1}{S_{xx}} \end{bmatrix}
\end{aligned}
$$

因此我们有:

$$
\mathrm{E}[\hat{\alpha}+\hat{\beta}x] = \alpha+\beta x
$$

$$
\begin{aligned}
\mathrm{Var}[\hat{\alpha}+\hat{\beta}x] &= \mathrm{Var}(\hat{\alpha})+2x\mathrm{Cov}(\hat{\alpha},\hat{\beta})+x^2\mathrm{Var}(\hat{\beta}) \\
&= \left(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right)\sigma^2+2x\cdot\left(-\frac{\bar{x}\sigma^2}{S_{xx}}\right)+x^2\frac{\sigma^2}{S_{xx}} \\
&= \sigma^2\left(\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}\right)
\end{aligned}
$$

上述问题等价于寻找一个常数 M_α 使得:

$$\mathrm{P}\left\{\frac{[(\hat{\alpha}+\hat{\beta}x)-(\alpha+\beta x)]^2}{s^2[\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}]}\le M_\tau^2 \text{ for all } x\right\}=1-\tau$$

即等价于使得:

$$\mathrm{P}\left\{\max_x\frac{[(\hat{\alpha}+\hat{\beta}x)-(\alpha+\beta x)]^2}{s^2[\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}]}\le M_\tau^2\right\}=1-\tau$$

> **(Statistical Inference 习题 11.40)**
> 若 a,b,c,d 为常数且 $c,d>0$，则 $\max_t\frac{(a+bt)^2}{c+dt^2}=\frac{a^2}{c}+\frac{b^2}{d}$
> 这个结论是下面引理的直接推论.
>
> **Lemma: (广义 Raylaigh 商)**
> 若 $b\in\mathbb{R}^n$ 为给定向量且 $A\in\mathbb{R}^{n\times n}$ 正定，则 $\max\limits_{x\ne 0_n\in\mathbb{R}^n}\frac{(b^{\mathrm{T}}x)^2}{x^{\mathrm{T}}Ax}=b^{\mathrm{T}}A^{-1}b$
>
> 证明:
>
> $$\begin{aligned}\max_{x\ne 0_n\in\mathbb{R}^n}\frac{(b^{\mathrm{T}}x)^2}{x^{\mathrm{T}}Ax}&=\max_{y\ne 0_n\in\mathbb{R}^n}\frac{(b^{\mathrm{T}}A^{-\frac{1}{2}}y)^2}{y^{\mathrm{T}}y}\quad(y:=A^{\frac{1}{2}}x)\\&=\max_{y\ne 0_n\in\mathbb{R}^n}\frac{y^{\mathrm{T}}(A^{-\frac{1}{2}}bb^{\mathrm{T}}A^{-\frac{1}{2}})y}{y^{\mathrm{T}}y}\quad(\text{Raylaigh theorem})\\&=\lambda_{\max}(A^{-\frac{1}{2}}bb^{\mathrm{T}}A^{-\frac{1}{2}})\qquad(\text{note that } A \text{ is positive definite, hence symmetric})\\&=\lambda_{\max}\{(A^{-\frac{1}{2}}b)(A^{-\frac{1}{2}}b)^{\mathrm{T}}\}\quad(\text{note that rank-one matrix } zz^{\mathrm{T}}\text{'s only non-zero eigenvalue is } \|z\|_2)\\&=\|A^{-\frac{1}{2}}b\|_2^2\\&=b^{\mathrm{T}}A^{-1}b\end{aligned}$$
>
> 其中最大值可以在 $x=\alpha A^{-1}b\ (\forall\,\alpha\ne 0\in\mathbb{R})$ 取到.

$$\begin{aligned}\max_x\frac{[(\hat{\alpha}+\hat{\beta}x)-(\alpha+\beta x)]^2}{s^2[\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}]}&=\frac{1}{s^2}\max_x\frac{[(\bar{y}-\hat{\beta}\bar{x}+\hat{\beta}x)-(\alpha+\beta\bar{x}-\beta\bar{x}+\beta x)]^2}{\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}}\\&=\frac{1}{s^2}\max_x\frac{[(\bar{y}-\alpha-\beta\bar{x})+(\hat{\beta}-\beta)(x-\bar{x})]^2}{\frac{1}{n}+\frac{1}{S_{xx}}(x-\bar{x})^2}\quad(\text{denote } t=x-\bar{x})\\&=\frac{1}{s^2}\max_t\frac{[(\bar{y}-\alpha-\beta\bar{x})+(\hat{\beta}-\beta)t]^2}{\frac{1}{n}+\frac{1}{S_{xx}}t^2}\quad(\text{use lemma})\\&=\frac{1}{s^2}\left\{\frac{(\bar{y}-\alpha-\beta\bar{x})^2}{\frac{1}{n}}+\frac{(\hat{\beta}-\beta)^2}{\frac{1}{S_{xx}}}\right\}\\&=\frac{\frac{(\bar{y}-\alpha-\beta\bar{x})^2}{\sigma^2/n}+\frac{(\hat{\beta}-\beta)^2}{\sigma^2/S_{xx}}}{s^2/\sigma^2}\\&\quad(\text{note that }\begin{cases}\bar{y}\sim N(\alpha+\beta\bar{x},\frac{\sigma^2}{n})\\\hat{\beta}\sim N(\beta,\frac{\sigma^2}{S_{xx}})\\s^2\sim\frac{1}{n-2}\sigma^2\chi^2_{(n-2)}\end{cases}\text{ and they are independent})\\&\sim\frac{\chi^2_{(2)}}{\chi^2_{(n-2)}/(n-2)}\\&=2\frac{\chi^2_{(2)}/2}{\chi^2_{(n-2)}/(n-2)}\\&=2F_{2,n-2}\end{aligned}$$

因此 $M_\tau^2=2F_{2,n-2,\tau}$，即 $M_\tau=\sqrt{2F_{2,n-2,\tau}}$
于是我们得到 $\alpha+\beta x\ (\forall\,x)$ 的 $1-\tau$ 置信区间为 $\hat{\alpha}+\hat{\beta}x\pm\sqrt{s^2[\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}]\cdot 2F_{2,n-2,\tau}}$

Problem 2

Consider the linear regression model $y = X\beta + \varepsilon$,
where $\begin{cases} \mathrm{E}(\varepsilon) = 0_n \\ \mathrm{Cov}(\varepsilon) = \sigma^2\Sigma \end{cases}$ in which Σ is a known positive definite matrix.
Find the BLUE for β and derive its variance-covariance matrix. $[10 \text{ Marks}]$

Solution:
由于 $\Sigma \in \mathbb{R}^{n\times n}$ 是已知的正定矩阵，因此平方根 $\Sigma^{\frac{1}{2}}$ 及其逆矩阵 $\Sigma^{-\frac{1}{2}}$ 都是存在的
于是我们有:

$$\begin{gathered} y = X\beta + \varepsilon \\ \Leftrightarrow \\ \Sigma^{-\frac{1}{2}} y = \Sigma^{-\frac{1}{2}} X\beta + \Sigma^{-\frac{1}{2}} \varepsilon \end{gathered}$$

定义:

$$\begin{cases} \tilde{y} = \Sigma^{-\frac{1}{2}} y \\ \tilde{X} = \Sigma^{-\frac{1}{2}} X \\ \tilde{\varepsilon} \ = \Sigma^{-\frac{1}{2}} \varepsilon \end{cases} \Rightarrow \ \tilde{y} = \tilde{X}\beta + \tilde{\varepsilon}$$

则我们有:

$$\begin{array}{c} \begin{aligned} \mathrm{E}[\tilde{\varepsilon}] &= \mathrm{E}[\Sigma^{-\frac{1}{2}}\varepsilon] \\ &= \Sigma^{-\frac{1}{2}}\mathrm{E}[\varepsilon] \\ &= \Sigma^{-\frac{1}{2}} 0_n \\ &= 0_n \end{aligned} \\ \hline \begin{aligned} \mathrm{Cov}[\tilde{\varepsilon}] &= \mathrm{Cov}[\Sigma^{-\frac{1}{2}}\varepsilon] \\ &= \Sigma^{-\frac{1}{2}}\mathrm{Cov}[\varepsilon]\Sigma^{-\frac{1}{2}} \\ &= \Sigma^{-\frac{1}{2}}\Sigma\Sigma^{-\frac{1}{2}} \\ &= I_n \end{aligned} \end{array}$$

根据 Gauss-Markov 定理可知 β 的 BLUE 即为新模型下的 LSE:

$$\begin{aligned} \hat{\beta} &:= (\tilde{X}^{\mathrm{T}}\tilde{X})^{-1}\tilde{X}^{\mathrm{T}} y \\ &= (X^{\mathrm{T}}\Sigma^{-1}X)^{-1}X^{\mathrm{T}}\Sigma^{-1}y \end{aligned}$$

其协方差矩阵为:

$$\begin{aligned} \mathrm{Cov}(\hat{\beta}) &= \mathrm{Cov}((X^{\mathrm{T}}\Sigma^{-1}X)^{-1}X^{\mathrm{T}}\Sigma^{-1}y) \\ &= (X^{\mathrm{T}}\Sigma^{-1}X)^{-1}X^{\mathrm{T}}\Sigma^{-1} \cdot \mathrm{Cov}(y) \cdot \Sigma^{-1}X(X^{\mathrm{T}}\Sigma^{-1}X)^{-1} \\ &= (X^{\mathrm{T}}\Sigma^{-1}X)^{-1}X^{\mathrm{T}}\Sigma^{-1} \cdot \mathrm{Cov}(\varepsilon) \cdot \Sigma^{-1}X(X^{\mathrm{T}}\Sigma^{-1}X)^{-1} \\ &= (X^{\mathrm{T}}\Sigma^{-1}X)^{-1}X^{\mathrm{T}}\Sigma^{-1} \cdot \sigma^2\Sigma \cdot \Sigma^{-1}X(X^{\mathrm{T}}\Sigma^{-1}X)^{-1} \\ &= \sigma^2(X^{\mathrm{T}}\Sigma^{-1}X)^{-1} \end{aligned}$$

Problem 3

Consider a multiple linear regression model with p independent variables $y = X\beta + \varepsilon$ where $\varepsilon \sim N(0, \sigma^2 I_n)$
Denote $\begin{cases} \hat{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y \\ s^2 \ = \frac{1}{n-p-1}\|y - X\hat{\beta}\|^2 \end{cases}$
(1) Find $\mathrm{Var}(s^2)$ $[5 \text{ Marks}]$
(2) Let $A = \frac{1}{n-p-3}(I_n - XX^{\dagger})$, Find $\mathrm{E}[(y^{\mathrm{T}}Ay - \sigma^2)^2]$ $[10 \text{ Marks}]$
(3) In terms of mean square errors，which estimators of σ^2 is better, $y^{\mathrm{T}}Ay$ or s^2? $[5 \text{ Marks}]$

Part (1)

Find $\mathrm{Var}(s^2)$ $[5 \text{ Marks}]$

Solution:
Gamma 随机变量 $W \sim \mathrm{Gamma}(\alpha, \lambda)$ 的均值和方差分别为:

$$\begin{cases} \mathrm{E}[W] = \frac{\alpha}{\lambda} \\ \mathrm{Var}[W] = \frac{\alpha}{\lambda^2} \end{cases}$$

考虑 s^2 的分布:

$$
\begin{aligned}
s^2 &= \frac{1}{n-p-1}\|y-X\hat{\beta}\|_2^2 \quad (\text{recall that } \hat{y}=X\hat{\beta}=Hy \text{ where } H=X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}) \\
&= \frac{1}{n-p-1}\|y-Hy\|_2^2 \\
&= \frac{1}{n-p-1}y^{\mathrm{T}}(I_n-H)^{\mathrm{T}}(I_n-H)y \quad (\text{note that } H \text{ is symmetric and idempotent}) \\
&= \frac{1}{n-p-1}y^{\mathrm{T}}(I_n-H)y \quad (\text{note that } y=X\beta+\varepsilon) \\
&= \frac{1}{n-p-1}(X\beta+\varepsilon)^{\mathrm{T}}(I_n-H)(X\beta+\varepsilon) \quad (\text{note that } HX=X \text{ so that } (I_n-H)X=0_{n\times(p+1)}) \\
&= \frac{1}{n-p-1}\varepsilon^{\mathrm{T}}(I_n-H)\varepsilon \quad (\text{note that } \operatorname{tr}(H)=p+1 \Rightarrow \operatorname{tr}(I_n-H)=n-p-1) \\
&\sim \frac{1}{n-p-1}\sigma^2\chi^2_{n-p-1}
\end{aligned}
$$

因此其方差为:

$$
\begin{aligned}
\operatorname{Var}(s^2) &= \operatorname{Var}[\frac{1}{n-p-1}\sigma^2\chi^2_{n-p-1}] \\
&= (\frac{1}{n-p-1}\sigma^2)^2\cdot\operatorname{Var}[\chi^2_{n-p-1}] \\
&= \frac{1}{(n-p-1)^2}\sigma^4\cdot\operatorname{Var}[\operatorname{Gamma}(\frac{n-p-1}{2},\frac{1}{2})] \\
&= \frac{1}{(n-p-1)^2}\sigma^4\cdot\frac{\frac{n-p-1}{2}}{(\frac{1}{2})^2} \\
&= \frac{2\sigma^4}{n-p-1}
\end{aligned}
$$

Part (2)

Let $A=\frac{1}{n-p-3}(I_n-XX^{\dagger})$, Find $\mathrm{E}[(y^{\mathrm{T}}Ay-\sigma^2)^2]$ $[10 \text{ Marks}]$

Solution:
在 $X\in\mathbb{R}^{n\times(p+1)}$ 列满秩的假设下，Moore-Penrose 逆 $X^{\dagger}=(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$
记 $H:=XX^{\dagger}=X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$
可以证明其对称、幂等且迹 $\operatorname{tr}(H):=p+1$)，而且 $\operatorname{Range}(H)=\operatorname{Range}(XX^{\dagger})=\operatorname{Range}(X)$
则我们有:

$$
A=\frac{1}{n-p-3}(I_n-XX^{\dagger})=\frac{1}{n-p-3}(I_n-H)
$$

$$
\begin{aligned}
y^{\mathrm{T}}Ay &= y^{\mathrm{T}}\cdot\frac{1}{n-p-3}(I_n-H)\cdot y \\
&= \frac{1}{n-p-3}y^{\mathrm{T}}(I_n-H)y \\
&= \frac{1}{n-p-3}(X\beta+\varepsilon)^{\mathrm{T}}(I_n-H)(X\beta+\varepsilon) \quad (\text{note that } HX=X\Rightarrow\ (I_n-H)X=0_{n\times(p+1)}) \\
&= \frac{1}{n-p-3}\varepsilon^{\mathrm{T}}(I_n-H)\varepsilon \quad (H \text{ is symmetric and idempotent} \begin{cases}\operatorname{tr}(I_n-H)=n-p-1 \\ \operatorname{Cov}(\varepsilon)=\sigma^2 I_n\end{cases}) \\
&\sim \frac{1}{n-p-3}\sigma^2\chi^2_{(n-p-1)}
\end{aligned}
$$

因此我们有:

$$
\begin{aligned}
\mathrm{E}[y^{\mathrm{T}}Ay] &= \frac{1}{n-p-3}\sigma^2 \cdot (n-p-1) = \frac{n-p-1}{n-p-3}\sigma^2 \\
\mathrm{Var}[y^{\mathrm{T}}Ay] &= (\frac{1}{n-p-3}\sigma^2)^2 \cdot 2(n-p-1) = \frac{2(n-p-1)}{(n-p-3)^2}\sigma^4 \\
\mathrm{E}[(y^{\mathrm{T}}Ay)^2] &= \mathrm{Var}[y^{\mathrm{T}}Ay] + \{\mathrm{E}[y^{\mathrm{T}}Ay]\}^2 \\
&= \frac{2(n-p-1)}{(n-p-3)^2}\sigma^4 + (\frac{n-p-1}{n-p-3}\sigma^2)^2 \\
&= \frac{(n-p-1)(n-p+1)}{(n-p-3)^2}\sigma^4
\end{aligned}
$$

于是我们有:

$$
\begin{aligned}
\mathrm{E}[(y^{\mathrm{T}}Ay-\sigma^2)^2] &= \mathrm{E}[(y^{\mathrm{T}}Ay)^2 - 2\sigma^2 y^{\mathrm{T}}Ay + \sigma^4] \\
&= \mathrm{E}[(y^{\mathrm{T}}Ay)^2] - 2\sigma^2\mathrm{E}[y^{\mathrm{T}}Ay] + \sigma^4 \\
&= \frac{(n-p-1)(n-p+1)}{(n-p-3)^2}\sigma^4 - 2\sigma^2\frac{n-p-1}{n-p-3}\sigma^2 + \sigma^4 \\
&= \left\{\frac{(n-p-1)(n-p+1)}{(n-p-3)^2} - \frac{2(n-p-1)}{n-p-3} + 1\right\}\sigma^4 \\
&= \frac{(n-p-1)(n-p+1) - 2(n-p-1)(n-p-3) + (n-p-3)^2}{(n-p-3)^2}\sigma^4 \\
&= \frac{-(n-p-1)(n-p-7) + (n-p-3)^2}{(n-p-3)^2}\sigma^4 \\
&= \frac{2(n-p+1)}{(n-p-3)^2}\sigma^4
\end{aligned}
$$

或者更简单地:

$$
\begin{aligned}
\mathrm{E}[(y^{\mathrm{T}}Ay-\sigma^2)^2] &= \mathrm{Var}[y^{\mathrm{T}}Ay-\sigma^2] + \{\mathrm{E}[y^{\mathrm{T}}Ay-\sigma^2]\}^2 \\
&= \mathrm{Var}[y^{\mathrm{T}}Ay] + \{\mathrm{E}[y^{\mathrm{T}}Ay]-\sigma^2\}^2 \\
&= \frac{2(n-p-1)}{(n-p-3)^2}\sigma^4 + \{\frac{n-p-1}{n-p-3}\sigma^2 - \sigma^2\}^2 \\
&= \frac{2(n-p-1)}{(n-p-3)^2}\sigma^4 + \{\frac{2}{n-p-3}\sigma^2\}^2 \\
&= \frac{2(n-p-1)}{(n-p-3)^2}\sigma^4 + \frac{4}{(n-p-3)^2}\sigma^4 \\
&= \frac{2(n-p+1)}{(n-p-3)^2}\sigma^4
\end{aligned}
$$

Part (3)

In terms of mean square errors, which estimators of σ^2 is better, $y^{\mathrm{T}}Ay$ or s^2? $[5 \text{ Marks}]$

Solution:
根据第 (1) 问可知 $\mathrm{Var}(s^2) = \frac{2\sigma^4}{n-p-1}$
注意到 $\mathrm{E}[s^2] = \mathrm{E}[\frac{1}{n-p-1}\sigma^2\chi^2_{n-p-1}] = \frac{1}{n-p-1}\sigma^2(n-p-1) = \sigma^2$
因此我们有:

$$
\begin{aligned}
\mathrm{MSE}(s^2) &= \mathrm{E}[(s^2-\sigma^2)^2] \quad (\text{note that } \mathrm{E}[s^2] = \sigma^2) \\
&= \mathrm{E}[(s^2-\mathrm{E}[s^2])^2] \\
&= \mathrm{Var}(s^2) \\
&= \frac{2\sigma^4}{n-p-1}
\end{aligned}
$$

根据第 (2) 问可知 $\mathrm{MSE}(y^{\mathrm{T}}Ay) = \mathrm{E}[(y^{\mathrm{T}}Ay-\sigma^2)^2] = \frac{2(n-p+1)}{(n-p-3)^2}\sigma^4$
则我们有:

$$
\begin{aligned}
\mathrm{MSE}(y^{\mathrm{T}}Ay)-\mathrm{MSE}(s^2) &= \frac{2(n-p+1)}{(n-p-3)^2}\sigma^4-\frac{2}{n-p-1}\sigma^4 \\
&= \frac{(n-p+1)(n-p-1)-(n-p-3)^2}{(n-p-3)^2(n-p-1)}\cdot(2\sigma^4) \\
&= \frac{(n-p)^2-1-(n-p)^2+6(n-p)-9}{(n-p-3)^2(n-p-1)}\cdot(2\sigma^4) \\
&= \frac{6(n-p)-10}{(n-p-3)^2(n-p-1)}\cdot(2\sigma^4) \quad (\text{note that } n>k+1) \\
&\geq \frac{6\cdot 2-10}{(n-p-3)^2(n-p-1)}\cdot(2\sigma^4) \\
&= \frac{2}{(n-p-3)^2(n-p-1)}\cdot(2\sigma^4) \\
&> 0
\end{aligned}
$$

因此从均方误差 (Mean Square Error, MSE) 的角度来说，$y^{\mathrm{T}}Ay$ 对 σ^2 的估计效果要比 s^2 差.

Problem 4

DEFINITION：
The **multiple corrlation coefficient** between a random variable Y and a random vector $(X_1,X_2,\ldots,X_k)^{\mathrm{T}}$ is the maximum correlation between Y and any linear function $\alpha_1X_1+\cdots+\alpha_kX_k$ of $(X_1,X_2,\ldots,X_k)^{\mathrm{T}}$.

Now consider a multiple linear regression model where its response variable Y and the k predictors $X_1,X_2,\ldots,X_k$ are multivariate normal. Find the multiple correlation coefficient between Y and $X_1,X_2,\ldots,X_k$. [**HINT**: use the Cauchy-Schwarz inequality for $\|\cdot\|_2$: $|u^{\mathrm{T}}v|\leq\|u\|_2\cdot\|v\|_2$] $[15\ \mathrm{Marks}]$

Solution:
记 $X=[X_1,X_2,\ldots,X_k]^{\mathrm{T}}$
根据题意可知 Y,X 服从联合多元正态分布，记为:

$$
\begin{bmatrix} Y \\ X \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_Y \\ \mu_X \end{bmatrix}, \begin{bmatrix} \sigma_{YY}^2 & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{bmatrix}\right)
$$

其中 $\Sigma_{YX}=\Sigma_{XY}^{\mathrm{T}}$ 且 Σ_{XX} 为对称阵.

记系数向量为 $\alpha=[\alpha_1,\alpha_2,\ldots,\alpha_k]^{\mathrm{T}}$
根据题干中的定义可知 Y 与 $X=[X_1,X_2,\ldots,X_k]^{\mathrm{T}}$ 之间的多重共线性系数 r 为:

$$
\begin{aligned}
r &:= \max_{\alpha\in\mathbb{R}^k}\mathrm{Corr}(Y,\alpha^{\mathrm{T}}X) \\
&= \max_{\alpha\in\mathbb{R}^k}\frac{\mathrm{Cov}(Y,\alpha^{\mathrm{T}}X)}{\sqrt{\mathrm{Var}(Y)\mathrm{Var}(\alpha^{\mathrm{T}}X)}} \\
&= \max_{\alpha\in\mathbb{R}^k}\frac{\alpha^{\mathrm{T}}\mathrm{Cov}(Y,X)}{\sqrt{\mathrm{Var}(Y)\alpha^{\mathrm{T}}\mathrm{Var}(X)\alpha}} \\
&= \max_{\alpha\in\mathbb{R}^k}\frac{\alpha^{\mathrm{T}}\Sigma_{XY}}{\sqrt{\sigma_{YY}^2\alpha^{\mathrm{T}}\Sigma_{XX}\alpha}}
\end{aligned}
$$

根据题干中的提示，我们对 $\alpha^{\mathrm{T}}\Sigma_{YX}$ 应用 Cauchy-Schwarz 不等式可得:

$$
\begin{aligned}
\alpha^{\mathrm{T}}\Sigma_{XY} &= \alpha^{\mathrm{T}}\Sigma_{XX}^{\frac{1}{2}}\Sigma_{XX}^{-\frac{1}{2}}\Sigma_{XY} \\
&= (\Sigma_{XX}^{\frac{1}{2}}\alpha)^{\mathrm{T}}(\Sigma_{XX}^{-\frac{1}{2}}\Sigma_{XY}) \quad (\text{Cauchy-Schwarz Inequality}) \\
&\leq \|\Sigma_{XX}^{\frac{1}{2}}\alpha\|_2\|\Sigma_{XX}^{-\frac{1}{2}}\Sigma_{XY}\|_2 \\
&= \sqrt{\alpha^{\mathrm{T}}\Sigma_{XX}\alpha}\cdot\sqrt{\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}} \quad (\text{note that } \Sigma_{YX}=\Sigma_{XY}^{\mathrm{T}})
\end{aligned}
$$

上述不等式当且仅当 $\Sigma_{XX}^{\frac{1}{2}}\alpha$ 和 $\Sigma_{XX}^{-\frac{1}{2}}\Sigma_{XY}$ 线性相关时取等.

于是我们有:

$$
\begin{aligned}
r &= \max_{\alpha\in\mathbb{R}^k} \frac{\alpha^{\mathrm{T}}\Sigma_{XY}}{\sqrt{\sigma_{YY}^2\alpha^{\mathrm{T}}\Sigma_{XX}\alpha}} \\
&= \frac{\sqrt{\alpha^{\mathrm{T}}\Sigma_{XX}\alpha}\cdot\sqrt{\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}}}{\sqrt{\sigma_{YY}^2\alpha^{\mathrm{T}}\Sigma_{XX}\alpha}} \\
&= \frac{\sqrt{\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}}}{\sigma_{YY}}
\end{aligned}
$$

根据多元线性回归模型的假设，我们有 $Y=\beta_0+\beta^{\mathrm{T}}X+\varepsilon$
其中 $\beta\in\mathbb{R}^k$ 且 $\varepsilon\sim N(0,\sigma^2)$ 与 X 独立.
因此 $\begin{bmatrix} Y \\ X \end{bmatrix} = \begin{bmatrix} \beta_0+\beta^{\mathrm{T}}X+\varepsilon \\ X \end{bmatrix} = \begin{bmatrix} \beta_0 \\ 0_k \end{bmatrix} + \begin{bmatrix} 1 & \beta^{\mathrm{T}} \\ 0_k & I_k \end{bmatrix}\begin{bmatrix} \varepsilon \\ X \end{bmatrix}$

其协方差矩阵为:

$$
\begin{aligned}
\mathrm{Cov}\left(\begin{bmatrix} Y \\ X \end{bmatrix}\right) &= \mathrm{Cov}\left(\begin{bmatrix} \beta_0 \\ 0_k \end{bmatrix} + \begin{bmatrix} 1 & \beta^{\mathrm{T}} \\ 0_k & I_k \end{bmatrix}\begin{bmatrix} \varepsilon \\ X \end{bmatrix}\right) \\
&= \begin{bmatrix} 1 & \beta^{\mathrm{T}} \\ 0_k & I_k \end{bmatrix}\mathrm{Cov}\left(\begin{bmatrix} \varepsilon \\ X \end{bmatrix}\right)\begin{bmatrix} 1 & \beta^{\mathrm{T}} \\ 0_k & I_k \end{bmatrix}^{\mathrm{T}} \\
&= \begin{bmatrix} 1 & \beta^{\mathrm{T}} \\ 0_k & I_k \end{bmatrix}\begin{bmatrix} \sigma^2 & \\ & \Sigma_{XX} \end{bmatrix}\begin{bmatrix} 1 & \beta^{\mathrm{T}} \\ 0_k & I_k \end{bmatrix}^{\mathrm{T}} \\
&= \begin{bmatrix} \sigma^2 & \beta^{\mathrm{T}}\Sigma_{XX} \\ 0_k & \Sigma_{XX} \end{bmatrix}\begin{bmatrix} 1 & 0_k^{\mathrm{T}} \\ \beta & I_k \end{bmatrix} \\
&= \begin{bmatrix} \sigma^2+\beta^{\mathrm{T}}\Sigma_{XX}\beta & \beta^{\mathrm{T}}\Sigma_{XX} \\ \Sigma_{XX}\beta & \Sigma_{XX} \end{bmatrix}
\end{aligned}
$$

因此我们有:

$$
\begin{aligned}
\sigma_{YY}^2 &= \sigma^2+\beta^{\mathrm{T}}\Sigma_{XX}\beta \\
\Sigma_{XY} &= \Sigma_{XX}\beta \\
\Sigma_{YX} &= \beta^{\mathrm{T}}\Sigma_{XX}
\end{aligned}
$$

$$
r = \frac{\sqrt{\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}}}{\sigma_{YY}} = \frac{\sqrt{\beta^{\mathrm{T}}\Sigma_{XX}\Sigma_{XX}^{-1}\Sigma_{XX}\beta}}{\sqrt{\sigma^2+\beta^{\mathrm{T}}\Sigma_{XX}\beta}} = \sqrt{\frac{\beta^{\mathrm{T}}\Sigma_{XX}\beta}{\sigma^2+\beta^{\mathrm{T}}\Sigma_{XX}\beta}}
$$

Problem 5

Given the multiple linear regression model with k independent variables:
$y=X\beta+\varepsilon$ where $\varepsilon\sim N(0,\sigma^2 I_n)$
Consider the following hypothesis testing:

$$
H_0: C\beta=h \quad \Leftrightarrow \quad H_1: C\beta\neq h
$$

where $C\in\mathbb{R}^{m\times(k+1)}$ is a $\text{rank-}m$ constant matrix and $h\in\mathbb{R}^m$ is a constant vector.

(1) Find the LSE (least squares estimator) of β under H_0 $[10\text{ marks}]$
(2) Find the $\mathrm{F}\text{-test}$ for testing H_0
(Write the test statistic and prove its null distribution) $[10\text{ marks}]$

Part (1)

Find the LSE (least squares estimator) of β under H_0 $[10\text{ marks}]$

Solution:
考虑求解线性约束最小二乘问题 $\min\limits_{C\beta=h}\|y-X\beta\|_2^2$

注意到目标函数 $f(\beta)=\|y-X\beta\|_2^2$ 是关于 β 的凸函数，而问题只有线性等式约束 $C\beta=h$
因此这是一个标准形式的凸优化问题，其最优解即为 KKT 点.

定义其 Lagrange 函数 $L(\beta,\lambda)$ 为:

$$
\begin{aligned}
L(\beta,\lambda) &= f(\beta) - \lambda^{\mathrm{T}}(C\beta - h) \\
&= \|y - X\beta\|_2^2 - \lambda^{\mathrm{T}}(C\beta - h) \\
\hline
\mathrm{dom}\{L\} &= \mathbb{R}^{p+1} \times \mathbb{R}^m
\end{aligned}
$$

Lagrange 函数 $L(\beta,\lambda)$ 关于 β 的梯度为:

$$
\begin{aligned}
\nabla_\beta L(\beta,\lambda) &= \nabla_\beta\{\|y - X\beta\|_2^2 - \lambda^{\mathrm{T}}(C\beta - h)\} \\
&= -X^{\mathrm{T}} \cdot 2(y - X\beta) - (\lambda^{\mathrm{T}}C)^{\mathrm{T}} \\
&= -2X^{\mathrm{T}}y + 2X^{\mathrm{T}}X\beta - C^{\mathrm{T}}\lambda
\end{aligned}
$$

KKT 条件为:

$$
\begin{cases}
\nabla_\beta L(\beta,\lambda) = -2X^{\mathrm{T}}y + 2X^{\mathrm{T}}X\beta - C^{\mathrm{T}}\lambda = 0_{p+1} & ① \\
C\beta = h & ②
\end{cases}
$$

① 式左乘 $(X^{\mathrm{T}}X)^{-1}$ 可得 $-2(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y + 2\beta - (X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}\lambda = 0_{p+1}$
于是有 $\beta = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y + \frac{1}{2}(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}\lambda$
代入 ② 式即得 $C\beta = C(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y + \frac{1}{2}C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}\lambda = h$
解得 $\lambda_{\mathrm{KKT}} = 2[C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-1}[h - C(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y]$
因此我们有:

$$
\begin{aligned}
\hat{\beta}_{\text{reduced}} &= \beta_{\mathrm{KKT}} \\
&= (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y + \frac{1}{2}(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}\lambda_{\mathrm{KKT}} \\
&= (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y + \frac{1}{2}(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}} \cdot 2[C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-1}[h - C(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y] \\
&= \hat{\beta}_{\text{full}} - (X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}[C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-1}(C\hat{\beta}_{\text{full}} - h)
\end{aligned}
$$

其中 $\hat{\beta}_{\text{full}} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y$

Part (2)

Find the F-test for testing H_0
(Write the test statistic and prove its null distribution) $[10 \text{ marks}]$

Solution:
下面我们计算简约模型误差平方和 $\mathrm{SSE}_{\text{reduced}} = \|y - X\hat{\beta}_{\text{reduced}}\|_2^2$

$$
\begin{aligned}
\mathrm{SSE}_{\text{reduced}} &= \|y - X\hat{\beta}_{\text{reduced}}\|_2^2 \\
&= \|y - X\{\hat{\beta}_{\text{full}} - (X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}[C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-1}(C\hat{\beta}_{\text{full}} - h)\}\|_2^2 \\
&= \|y - X\hat{\beta}_{\text{full}} - X(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}[C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-1}(C\hat{\beta}_{\text{full}} - h)\|_2^2 \\
&= \|y - X\hat{\beta}_{\text{full}} - A(C\hat{\beta}_{\text{full}} - h)\|_2^2 \quad (\text{denote } A := X(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}[C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-1}) \\
&= \|y - X\hat{\beta}_{\text{full}}\|_2^2 - 2(y - X\hat{\beta}_{\text{full}})^{\mathrm{T}}A(C\hat{\beta}_{\text{full}} - h) + \|A(C\hat{\beta}_{\text{full}} - h)\|_2^2
\end{aligned}
$$

其中 $A := X(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}[C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-1} \in \mathbb{R}^{n\times m}$
考虑交叉项 $(y - X\hat{\beta}_{\text{full}})^{\mathrm{T}}A(C\hat{\beta}_{\text{full}} - h)$:

$$
\begin{aligned}
&(y - X\hat{\beta}_{\text{full}})^{\mathrm{T}}A(C\hat{\beta}_{\text{full}} - h) \\
&= (y - Hy)^{\mathrm{T}}A(C\hat{\beta}_{\text{full}} - h) \quad (\text{recall that } X\hat{\beta}_{\text{full}} = Hy \text{ where } H = X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y) \\
&= y^{\mathrm{T}}(I_n - H)X(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}[C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-1}(C\hat{\beta}_{\text{full}} - h) \quad (\text{note that } HX = X \text{ so that } (I_n - H)X = 0_{n\times(p+1)}) \\
&= y^{\mathrm{T}}0_{n\times(p+1)}(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}[C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-1}(C\hat{\beta}_{\text{full}} - h) \\
&= 0
\end{aligned}
$$

因此我们有:

$$
\begin{aligned}
\mathrm{SSE}_{\text{reduced}} &= \|y - X\hat{\beta}_{\text{full}}\|_2^2 - 2(y - X\hat{\beta}_{\text{full}})^{\mathrm{T}}A(C\hat{\beta}_{\text{full}} - h) + \|A(C\hat{\beta}_{\text{full}} - h)\|_2^2 \\
&= \mathrm{SSE}_{\text{full}} - 2\cdot 0 + \|A(C\hat{\beta}_{\text{full}} - h)\|_2^2 \\
&= \mathrm{SSE}_{\text{full}} + \|A(C\hat{\beta}_{\text{full}} - h)\|_2^2
\end{aligned}
$$

我们定义**额外误差平方和** (extra sum of squares, ESS) 为从全模型到简约模型增加的误差平方和:

$$\mathrm{ESS} = \mathrm{SSE}_{\text{reduced}} - \mathrm{SSE}_{\text{full}} = \|A(C\hat{\beta}_{\text{full}} - h)\|_2^2$$

其中 $\begin{cases} A = X(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}[C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-1} \\ \hat{\beta}_{\text{full}} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y \end{cases}$

考虑第一类型错误概率界限为 α 的检验问题 $H_0 : C\beta = h \ \leftrightarrow \ H_1 : C\beta \neq h$
下面我们研究**额外误差平方和** $\mathrm{ESS} = \mathrm{SSE}_{\text{reduced}} - \mathrm{SSE}_{\text{full}} = \|A(C\hat{\beta}_{\text{full}} - h)\|_2^2$ 在零假设 $H_0 : C\beta = h$ 下的分布.

$$\begin{aligned}
\mathrm{ESS} &= \mathrm{SSE}_{\text{reduced}} - \mathrm{SSE}_{\text{full}} \\
&= \|A(C\hat{\beta}_{\text{full}} - h)\|_2^2 \\
&= (C\hat{\beta}_{\text{full}} - h)^{\mathrm{T}}A^{\mathrm{T}}A(C\hat{\beta}_{\text{full}} - h) \quad (\text{recall that } A = X(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}[C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-1}) \\
&= (C\hat{\beta}_{\text{full}} - h)^{\mathrm{T}}\{X(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}[C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-1}\}^{\mathrm{T}}\{X(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}[C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-1}\}(C\hat{\beta}_{\text{full}} - h) \\
&= (C\hat{\beta}_{\text{full}} - h)^{\mathrm{T}}\{[C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-1}C(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\} \cdot \{X(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}[C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-1}\}(C\hat{\beta}_{\text{full}} - h) \\
&= (C\hat{\beta}_{\text{full}} - h)^{\mathrm{T}}[C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-1}(C\hat{\beta}_{\text{full}} - h) \\
&= \eta^{\mathrm{T}}\eta \quad (\text{denote } \eta = [C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-\frac{1}{2}}(C\hat{\beta}_{\text{full}} - h))
\end{aligned}$$

注意到 $\hat{\beta}_{\text{full}} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y \sim N(\beta, \sigma^2(X^{\mathrm{T}}X)^{-1})$
于是我们有:

$$\begin{aligned}
C\hat{\beta}_{\text{full}} - h &\sim N(C\beta - h, \sigma^2 C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}) \\
&\overset{H_0}{=} N(0_m, \sigma^2 C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}) \quad (\text{where } H_0 : C\beta = h)
\end{aligned}$$

因此我们有:

$$\begin{aligned}
\eta &= [C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-\frac{1}{2}}(C\hat{\beta}_{\text{full}} - h) \\
&\overset{H_0}{\sim} N([C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-\frac{1}{2}} \cdot 0_m, [C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-\frac{1}{2}}\sigma^2 C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}\{[C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-\frac{1}{2}}\}^{\mathrm{T}}) \quad (\text{where } H_0 : C\beta = h) \\
&= N(0_m, \sigma^2 I_m)
\end{aligned}$$

于是我们有 $\mathrm{ESS} = \eta^{\mathrm{T}}\eta \overset{H_0}{\sim} \sigma^2\chi^2_{(m)}$

现在我们可以构造线性约束检验问题 $H_0 : C\beta = h \ \leftrightarrow \ H_1 : C\beta \neq h$ 的检验统计量了:

$$\begin{aligned}
F &:= \frac{(\mathrm{SSE}_{\text{reduced}} - \mathrm{SSE}_{\text{full}})/m}{\mathrm{SSE}_{\text{full}}/n-p-1} \\
&= \frac{\mathrm{ESS}/m}{\mathrm{SSE}_{\text{full}}/n-p-1} \\
&\quad (\text{note that } \begin{cases} \mathrm{ESS} \overset{H_0}{\sim} \sigma^2\chi^2_{(m)} \\ \mathrm{SSE}_{\text{full}} \sim \sigma^2\chi^2_{(n-p-1)} \end{cases} \text{where } H_0 : C\beta = h) \\
&\overset{H_0}{\sim} \frac{\sigma^2\chi^2_{(m)}/m}{\sigma^2\chi^2_{(n-p-1)}/(n-p-1)} \\
&= F_{m,n-p-1}
\end{aligned}$$

其中分子 $\frac{1}{m}\mathrm{ESS} = \frac{1}{m}\|A(C\hat{\beta}_{\text{full}} - h)\|_2^2$ 与分母 $s^2_{\text{full}} = \frac{1}{n-p-1}\|y - X\hat{\beta}_{\text{full}}\|_2^2$ 的独立性由 $\hat{\beta}_{\text{full}} \perp s^2_{\text{full}}$ 保证.
(其中 s^2_{full} 记为全模型中根据 σ^2 的极大似然估计量构造出的无偏估计量)

我们记 $F_{m,n-p-1,\alpha}$ 为 $F_{m,n-p-1}$ 分布的 $1-\alpha$ 分位数.
则线性约束 $C\beta = h$ 的显著性检验的 F-检验法为:

- (F-检验法)
  若 $F = \frac{\mathrm{ESS}/m}{\mathrm{SSE}_{\text{full}}/n-p-1} = \frac{\|A(C\hat{\beta}_{\text{full}}-h)\|_2^2/m}{\|y-X\hat{\beta}_{\text{full}}\|_2^2/(n-p-1)} > F_{m,n-p-1,\alpha}$
  则我们拒绝零假设 $H_0 : C\beta = h$，即我们认为线性先验关系 $C\beta = h$ 不成立.
  其中 $\begin{cases} A = X(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}[C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-1} \\ \hat{\beta}_{\text{full}} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y \end{cases}$

Problem 6

Consider a linear regression model with Normal errors and take σ as known.
Show that the model with the largeset AIC statistic is the model with the lowest C_p statistic. $[10 \text{ marks}]$

Solution:
给定某个 p 元解释变量子集，设其设计矩阵为 $X_p \in \mathbb{R}^{n\times(p+1)}$
我们记其回归参数向量 β_p 的估计量 $\hat{\beta}_p = (X_p^{\mathrm{T}}X_p)^{-1}X_p y$ (这里的下标只代表模型的解释变量个数，不代表分量)
误差平方和 $\mathrm{SSE}_p = \|y-\hat{y}_p\|_2^2 = \|y - X_p\hat{\beta}_p\|_2^2$ (自由度为 $n-p-1$)

在正态假设下，我们有 $y \sim N(X_p\beta_p, \sigma^2 I_n)$ (其中 σ^2 反映的是随机噪音的强度，与模型无关)
其概率密度函数为:

$$\begin{aligned} f(y) &= \frac{1}{(\sqrt{2\pi})^n|\sigma^2 I_n|^{\frac{1}{2}}}\exp\{-\frac{1}{2}(y-X_p\beta_p)^{\mathrm{T}}(\sigma^2 I_n)^{-1}(y-X_p\beta_p)\} \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^n}\exp\{-\frac{1}{2\sigma^2}\|y-X_p\beta_p\|_2^2\} \end{aligned}$$

似然函数 $L(\beta_p,\sigma^2|X_p,y)$ 和对数似然函数 $\log L(\beta_p,\sigma^2|X_p,y)$ 为:

$$\frac{L(\beta_p,\sigma^2|X_p,y) = f(y) = \frac{1}{(\sqrt{2\pi}\sigma)^n}\exp\{-\frac{1}{2\sigma^2}\|y-X_p\beta_p\|_2^2\}}{\log L(\beta_p,\sigma^2|X_p,y) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\|y-X_p\beta_p\|_2^2}$$

根据多元线性回归的结论可知，最大化对数似然函数 $\log L(\beta_p,\sigma^2|X_p,y)$ 得到的似然解为:
(假设 σ^2 已知)

$$\begin{aligned} \max_{\beta,\sigma^2}\log L(\beta_p,\sigma^2|X_p,y) &= \log L(\hat{\beta}_p,\sigma^2|X_p,y) \quad (\text{where } \hat{\beta}_p = (X_p^{\mathrm{T}}X_p)^{-1}X_p^{\mathrm{T}}y) \\ &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\|y-X_p\hat{\beta}_p\|_2^2 \\ &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\mathrm{SSE}_p \end{aligned}$$

我们希望最大化:
(王勤文老师采用的定义，其中模型复杂度由解释变量个数来表征，而最大对数似然函数舍去了与 p 无关的项)

$$\begin{aligned} \mathrm{AIC}_p &:= \text{maximum\_log-likelihood} - \text{model\_complexity} \\ &= -\frac{1}{2\sigma^2}\mathrm{SSE}_p - p \end{aligned}$$

在所有回归子集中 Akaike 信息量 AIC_p 最大者对应的回归模型就是最优模型.

在 σ^2 已知的条件下，我们可以构造 C_p 统计量:

$$C_p = \frac{\mathrm{SSE}_p}{\sigma^2} - n + 2(p+1)$$

在所有回归子集中，我们尽量选择 Mallows 指数 $C_p \approx p+1$ 的子集.
若有多个子集满足 $C_p \approx p+1$ (它们对应的 p 可能是不同的)，则我们尽量选取 C_p 值较小的子集.
简单起见，我们认为在所有回归子集中 C_p 统计量最小者对应的回归模型就是最优模型.
因此最大化 AIC_p 和最小化 C_p 所选择的模型是一致的.

The End