FDU 回归分析 期中考试 (2024 秋)

Total: 100 marks
Duration: 2 hour and 30 minutes

Problem 1

Let Y_1, Y_2, Y_3 be independent response observations satisfying:

$$Y_i = \begin{cases} \beta + \varepsilon_i & \text{if } i = 1, 2 \\ -\beta + \varepsilon_i & \text{if } i = 3 \end{cases}$$

where $\beta \in \mathbb{R}$ is an unknown parameter and $\varepsilon_1, \varepsilon_2, \varepsilon_3$ are independent $N(0, \sigma^2)$ variables for some unknown $\sigma^2 > 0$

Part (1)

The above setting decribes a linear model of the form:

$$Y \equiv \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = X\beta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$$

where $X \in \mathbb{R}^{3\times 1}$ denotes the design matrix.
Find X. $[2 \text{ Marks}]$

Solution:

$$X = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

Part (2)

Show that the least squares estimator of β is $\hat{\beta} = \frac{1}{3}(Y_1 + Y_2 - Y_3)$ $[4 \text{ Marks}]$

Solution:

$$\begin{aligned} \hat{\beta} &= (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y \\ &= \left(\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} \\ &= \frac{1}{3}(Y_1 + Y_2 - Y_3) \end{aligned}$$

Part (3)

The fitted values of Y_1, Y_2, Y_3 are given by the vector $\hat{Y} = HY$ for some matrix $H \in \mathbb{R}^{3\times 3}$.
Find H. $[2 \text{ Marks}]$

Solution:

$$\begin{aligned} H &= X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}} \\ &= \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \left(\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}^{\mathrm{T}} \\ &= \frac{1}{3} \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix} \end{aligned}$$

Part (4)

Show that the residual sum of squares has a quadratic form $Y^{\mathrm{T}}AY$ for some $A \in \mathbb{R}^{3\times 3}$.
Find A. $[4 \text{ Marks}]$

Solution:
注意到:

$$\begin{aligned}
\text{SSE} &= \|Y - \hat{Y}\|^2 \\
&= \|Y - HY\|^2 \\
&= Y^{\mathrm{T}}(I-H)^{\mathrm{T}}(I-H)Y \quad (\text{note that } \begin{cases} H^{\mathrm{T}} = H \\ H^2 = H \end{cases}) \\
&= Y^{\mathrm{T}}(I-H)Y
\end{aligned}$$

因此我们有:

$$\begin{aligned}
A &= I - H \\
&= \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix} \\
&= \frac{1}{3}\begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}
\end{aligned}$$

Part (5)

Show that the residual sum of squares and $\hat{Y}$ are independent. $[4 \text{ Marks}]$

Solution:
注意到:

$$\begin{aligned}
\text{SSE} &= \|Y - \hat{Y}\|^2 \\
&= \|Y - HY\|^2 \\
&= Y^{\mathrm{T}}(I-H)^{\mathrm{T}}(I-H)Y \\
&= Y^{\mathrm{T}}(I-H)Y \\
&= (X\beta + \varepsilon)^{\mathrm{T}}(I-H)(X\beta+\varepsilon) \quad (\text{note that } HX = X \Rightarrow (I-H)X = 0_{n\times(p+1)}) \\
&= \varepsilon^{\mathrm{T}}(I-H)\varepsilon \\
&= \|(I-H)\varepsilon\|^2
\end{aligned}$$

而 $\hat{Y} = HY = H(X\beta+\varepsilon) = X\beta + H\varepsilon$
要证明 $\text{SSE} \perp \hat{Y}$，只要证明 $(I-H)\varepsilon \perp H\varepsilon$
由于 $(I-H)\varepsilon$ 和 $H\varepsilon$ 都是 $\varepsilon \sim N(0_n, \sigma^2 I_n)$，故它们联合正态.
因此只要证明 $(I-H)\varepsilon$ 和 $H\varepsilon$ 不相关即可:

$$\begin{aligned}
\text{Cov}((I-H)\varepsilon, H\varepsilon) &= (I-H)\text{Cov}(\varepsilon)H \\
&= (I-H)\cdot\sigma^2 I_n \cdot H \\
&= \sigma^2(H - H^2) \\
&= \sigma^2(H - H) \\
&= 0_{n\times n}
\end{aligned}$$

因此 $\text{SSE} \perp \hat{Y}$

Part (6)

Based on the result in part (5),
test whether or not $\beta = 0$ at the 5% level of significance given that $\begin{cases} Y_1 = 8 \\ Y_2 = 12 \\ Y_3 = -13 \end{cases} \quad [4 \text{ Marks}]$

Solution:
注意到:

$$
\begin{aligned}
\hat{\beta} &= (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y \\
&= (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}(X\beta+\varepsilon) \\
&= \beta+(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\varepsilon \\
&= \beta+N((X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}0_n,(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\cdot\sigma^2I_n\cdot X(X^{\mathrm{T}}X)^{-1}) \\
&= \beta+N(0_n,\sigma^2(X^{\mathrm{T}}X)^{-1}) \\
&= N(\beta,\sigma^2(X^{\mathrm{T}}X)^{-1})
\end{aligned}
\Rightarrow \frac{\hat{\beta}-\beta}{\sigma\sqrt{(X^{\mathrm{T}}X)^{-1}}}\sim N(0,1)
$$

代入 $X=\begin{bmatrix}1\\1\\-1\end{bmatrix}$ 和 $Y=\begin{bmatrix}Y_1\\Y_2\\Y_3\end{bmatrix}=\begin{bmatrix}8\\12\\-13\end{bmatrix}$ 可知:

$$
\hat{\beta}=(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y=\frac{1}{3}(Y_1+Y_2-Y_3)=\frac{1}{3}(8+12-(-13))=11
$$

$$
s^2=\frac{1}{2}\mathrm{SSE}=\frac{1}{2}Y^{\mathrm{T}}AY=\frac{1}{2}\begin{bmatrix}8\\12\\-13\end{bmatrix}^{\mathrm{T}}\left(\frac{1}{3}\begin{bmatrix}2&-1&1\\-1&2&1\\1&1&2\end{bmatrix}\right)\begin{bmatrix}8\\12\\-13\end{bmatrix}=7
$$

注意到 $\operatorname{tr}(A)=\operatorname{tr}\left(\frac{1}{3}\begin{bmatrix}2&-1&1\\-1&2&1\\1&1&2\end{bmatrix}\right)=2$，因此我们有 $s^2\sim\sigma^2\chi^2_{(2)}$

因此我们有:

$$
\frac{\hat{\beta}-\beta}{s\sqrt{(X^{\mathrm{T}}X)^{-1}}}=\frac{\frac{\hat{\beta}-\beta}{\sigma\sqrt{(X^{\mathrm{T}}X)^{-1}}}}{\frac{s}{\sigma}}\sim\frac{N(0,1)}{\sqrt{\chi^2_{(2)}/2}}=t_{(2)}
$$

在零假设 $H_0:\beta=0$ 下我们有:

$$
\left|\frac{\hat{\beta}-\beta}{s\sqrt{(X^{\mathrm{T}}X)^{-1}}}\right|\overset{H_0}{=}\left|\frac{11-0}{\sqrt{7}\cdot\sqrt{\frac{1}{3}}}\right|=\left|11\sqrt{\frac{3}{7}}\right|\approx 7.20
$$

注意到 $F_{1,2}(0.05)\approx 18.51$，因此 $t_2(0.025)=\sqrt{F_{1,2}(0.05)}=\sqrt{18.51}\approx 4.30$
于是我们有:

$$
\left|\frac{\hat{\beta}-\beta}{s\sqrt{(X^{\mathrm{T}}X)^{-1}}}\right|\overset{H_0}{=}\left|\frac{11-0}{\sqrt{7}\cdot\sqrt{\frac{1}{3}}}\right|=\left|11\sqrt{\frac{3}{7}}\right|\approx 7.20>4.30=t_2(0.025)
$$

这表明在因子水平 $\alpha=0.05$ 下我们可以拒绝零假设 $H_0:\beta=0$

Problem 2

Consider the linear regression model:

$$
Y\equiv\begin{bmatrix}Y_1\\\vdots\\Y_n\end{bmatrix}\sim N(X\beta,\sigma^2I_n)
$$

$$
Y_\star\equiv\begin{bmatrix}Y_1^\star\\\vdots\\Y_m^\star\end{bmatrix}\sim N(X_\star\beta,\sigma^2I_m)
$$

where $X\in\mathbb{R}^{n\times d}$, $X_\star\in\mathbb{R}^{m\times d}$ are given design matrices, and observation $Y\in\mathbb{R}^n$ is given.
Let $a=[a_1,\ldots,a_m]^{\mathrm{T}}$ be a vector of m known constants.
Suppose we are interested in predicting $l=a^{\mathrm{T}}Y_\star=\sum_{j=1}^m a_jY_j^\star$
Give the $1-\alpha$ prediction interval for l $[10\text{ Marks}]$

Solution:
模型的最小二乘估计量为 $\hat{\beta}:=(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y\sim N(\beta,\sigma^2(X^{\mathrm{T}}X)^{-1})$
定义 $\hat{l}:=a^{\mathrm{T}}\hat{Y}_\star=a^{\mathrm{T}}X_\star\hat{\beta}=a^{\mathrm{T}}X_\star(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y$
则我们有:

$$
\begin{aligned}
\hat{l}-l &= a^{\mathrm{T}}X_\star\hat{\beta}-a^{\mathrm{T}}Y_\star \\
&= a^{\mathrm{T}}X_\star\hat{\beta}-a^{\mathrm{T}}(X_\star\beta+\varepsilon_\star) \\
&= a^{\mathrm{T}}X_\star(\hat{\beta}-\beta)-a^{\mathrm{T}}\varepsilon_\star \quad (\text{note that } \begin{cases}\hat{\beta}-\beta\sim N(0_d,\sigma^2(X^{\mathrm{T}}X)^{-1}) \\ \varepsilon_\star\sim N(0_m,\sigma^2 I_m) \\ \hat{\beta}\perp\varepsilon_\star\end{cases}) \\
&\sim N(a^{\mathrm{T}}X_\star 0_d-a^{\mathrm{T}}0_m, a^{\mathrm{T}}X_\star\cdot\sigma^2(X^{\mathrm{T}}X)^{-1}\cdot X_\star^{\mathrm{T}}a+a^{\mathrm{T}}\cdot\sigma^2 I_m\cdot a) \\
&= N(0,\sigma^2 a^{\mathrm{T}}[X_\star(X^{\mathrm{T}}X)^{-1}X_\star^{\mathrm{T}}+I_m]a)
\end{aligned}
$$

其中 $\hat{\beta}\perp\varepsilon_\star$ 是因为 $\hat{\beta}$ 依赖于前 n 个样本，与新的 m 个样本的随机噪音 $\varepsilon_\star$ 相互独立.

因此我们有:

$$
\frac{l-\hat{l}}{\sigma\sqrt{a^{\mathrm{T}}[X_\star(X^{\mathrm{T}}X)^{-1}X_\star^{\mathrm{T}}+I_m]a}}\sim N(0,1)
$$

现考虑 $s^2=\frac{1}{n-d}\|Y-\hat{Y}\|^2=\frac{1}{n-d}\|Y-HY\|^2=\frac{1}{n-d}Y^{\mathrm{T}}(I_n-H)Y$ (其中 $H=X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$)

我们有 $\begin{cases}s^2\sim\frac{1}{n-d}\sigma^2\chi^2_{(n-d)} \\ s^2\perp\hat{\beta} \\ s^2\perp\varepsilon_\star\end{cases}$ (其中 $s^2\perp\varepsilon_\star$ 是因为 s^2 依赖于前 n 个样本，与新的 m 个样本的随机噪音 $\varepsilon_\star$ 相互独立)

于是我们有:

$$
\frac{l-\hat{l}}{s\sqrt{a^{\mathrm{T}}[X_\star(X^{\mathrm{T}}X)^{-1}X_\star^{\mathrm{T}}+I_m]a}}=\frac{\frac{l-\hat{l}}{\sigma\sqrt{a^{\mathrm{T}}[X_\star(X^{\mathrm{T}}X)^{-1}X_\star^{\mathrm{T}}+I_m]a}}}{\frac{s}{\sigma}}\sim\frac{N(0,1)}{\sqrt{\chi^2_{(n-d)}/(n-d)}}=t_{(n-d)}
$$

因此 l 的 $100(1-\alpha)\%$ 预测区间为 $[\hat{l}\pm t_{n-d,\frac{\alpha}{2}}\cdot s\sqrt{a^{\mathrm{T}}[X_\star(X^{\mathrm{T}}X)^{-1}X_\star^{\mathrm{T}}+I_m]a}]$
其中 $\hat{l}:=a^{\mathrm{T}}\hat{Y}_\star=a^{\mathrm{T}}X_\star\hat{\beta}=a^{\mathrm{T}}X_\star(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y$
而 $t_{n-d,\frac{\alpha}{2}}$ 是 $t_{(n-d)}$ 分布的 $1-\frac{\alpha}{2}$ 分位数.

Problem 3

Suppose that $\mu\equiv\mathrm{E}[Y]=X\beta$ and $\mathrm{Cov}(Y)=\sigma^2 I_n$ in the Linear regression model, where $X\in\mathbb{R}^{n\times(p+1)}$ and $\beta\in\mathbb{R}^{p+1}$
Let $\tilde{\phi}=c^{\mathrm{T}}Y$ be any unbiased linear estimator of $\phi=t^{\mathrm{T}}\beta$, where $t\in\mathbb{R}^{p+1}$ is an arbitrary vector.
Prove that $\mathrm{Var}(\tilde{\phi})\geq\mathrm{Var}(\hat{\phi})$
where $\hat{\phi}=t^{\mathrm{T}}\hat{\beta}$ and $\hat{\beta}=(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y$ is the least squares estimator of β. $[10\text{ Marks}]$

Solution:
根据 $\tilde{\phi}=c^{\mathrm{T}}Y$ 的无偏性可知:

$$
\mathrm{E}[\tilde{\phi}]=\mathrm{E}[c^{\mathrm{T}}Y]=c^{\mathrm{T}}\mathrm{E}[Y]=c^{\mathrm{T}}(X\beta)=t^{\mathrm{T}}\beta
$$

因此 $c\in\mathbb{R}^n$ 一定满足 $X^{\mathrm{T}}c=t$

目标函数 $\mathrm{Var}(\tilde{\phi})=\mathrm{Var}(c^{\mathrm{T}}Y)=c^{\mathrm{T}}\mathrm{Var}(Y)c=c^{\mathrm{T}}\cdot\sigma^2 I_n\cdot c=\sigma^2 c^{\mathrm{T}}c$
考虑求解最优化问题:

$$
\min_{X^{\mathrm{T}}c=t} c^{\mathrm{T}}c
$$

定义 Lagrange 函数为 $L(c,\lambda)=c^{\mathrm{T}}c-\lambda^{\mathrm{T}}(X^{\mathrm{T}}c-t)$ (其中 Lagrange 乘子 $\lambda\in\mathbb{R}^{p+1}$)
注意到上述问题是凸优化问题，故其最优解即 KKT 点.
其 KKT 条件为:

$$
\begin{cases}\nabla_c L(c,\lambda)=2c-X\lambda=0_n \\ X^{\mathrm{T}}c=t\end{cases}
$$

将 $c=\frac{1}{2}X\lambda$ 代入 $X^{\mathrm{T}}c=t$ 即得 $\lambda_\star=2(X^{\mathrm{T}}X)^{-1}t$
因此 $c_\star=\frac{1}{2}X\lambda_\star=X(X^{\mathrm{T}}X)^{-1}t$
这意味着使得 $\mathrm{Var}(\tilde{\phi})=\sigma^2 c^{\mathrm{T}}c$ 达到最小值的 $\tilde{\phi}_\star=c_\star^{\mathrm{T}}Y=t^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y=t^{\mathrm{T}}\hat{\beta}=\hat{\phi}$
所以我们有 $\mathrm{Var}(\tilde{\phi})\geq\mathrm{Var}(\hat{\phi})$ 恒成立.

Problem 4

Consider a multiple linear regression model:

$$Y=\beta_0 1_n+X\beta+\varepsilon \text{ where } \begin{cases}\mathrm{E}[\varepsilon]=0_n\\ \mathrm{Cov}[\varepsilon]=\sigma^2 I_n\end{cases}$$

where $Y\in\mathbb{R}^n$ is the observation vector and $X\in\mathbb{R}^{n\times k}$ design matrix with full rank.
Find the F-statistics for the following two hypothesis testing problems:

- ① $H_0:\beta_1=\cdots=\beta_k=c$ where $c\in\mathbb{R}$ is some given constant. $[10\text{ Marks}]$
- ② $H_0:\beta_1=\cdots=\beta_k$ $[10\text{ Marks}]$

Part (1)

$H_0:\beta_1=\cdots=\beta_k=c$ where $c\in\mathbb{R}$ is some given constant. $[10\text{ Marks}]$

Solution:
零假设 $H_0:\beta_1=\cdots=\beta_k=c$ 可以等价写为:

$$[0_k,I_k]\begin{bmatrix}\beta_0\\ \beta\end{bmatrix}=c1_k$$

表明等式约束的个数为 k

在 H_0 成立条件下的简约模型为 $Y=\beta_0 1_n+X\cdot c1_k+\varepsilon$
我们记 $\tilde{Y}:=Y-X\cdot c1_k$ 即得 $\tilde{Y}=\beta_0 1_n+\varepsilon$
因此关于简约模型我们有:

$$\hat{\beta}_0:=(1_n^{\mathrm{T}}1_n)^{-1}1_n^{\mathrm{T}}\tilde{Y}=\frac{1}{n}1_n^{\mathrm{T}}\tilde{Y} \text{ where } \tilde{Y}=Y-X\cdot c1_k$$
$$\mathrm{SSE}_{\text{reduced}}:=\|\tilde{Y}-\hat{\beta}_0 1_n\|^2$$

而关于全模型我们有:

$$\tilde{X}:=[1_n,X]\in\mathbb{R}^{n\times(k+1)}$$
$$\hat{\beta}_{\text{full}}:=(\tilde{X}^{\mathrm{T}}\tilde{X})^{-1}\tilde{X}^{\mathrm{T}}Y$$
$$\mathrm{SSE}_{\text{full}}:=\|Y-\tilde{X}\hat{\beta}_{\text{full}}\|^2$$

额外平方和 $\mathrm{ESS}:=\mathrm{SSE}_{\text{reduced}}-\mathrm{SSE}_{\text{full}}\sim\sigma^2\chi^2_{(k)}$ (考虑到等式约束的个数为 k 所以其自由度为 k)
因此我们可以定义 F-统计量:

$$F:=\frac{\mathrm{ESS}/k}{\mathrm{SSE}_{\text{full}}/(n-k-1)}\sim F_{k,n-k-1}$$

检验法为:
若 $F>F_{k,n-k-1}(\alpha)$，就拒绝零假设 $H_0:\beta_1=\cdots=\beta_k=c$
其中 $F_{k,n-k-1}(\alpha)$ 为 $F_{k,n-k-1}$ 分布的 $1-\alpha$ 分位数.

Part (2)

$H_0:\beta_1=\cdots=\beta_k$ $[10\text{ Marks}]$

Solution:
零假设 $H_0:\beta_1=\cdots=\beta_k$ 可以等价写为:

$$C\begin{bmatrix}\beta_0\\ \beta\end{bmatrix}=\begin{bmatrix}0&1&-1&&&\\0&&1&-1&&\\ \vdots&&&\ddots&\ddots&\\0&&&&1&-1\end{bmatrix}\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \vdots\\ \beta_k\end{bmatrix}=\begin{bmatrix}0\\0\\ \vdots\\0\end{bmatrix}$$

其中系数矩阵 $C\in R^{(k-1)\times(k+1)}$，表明等式约束的个数为 $k-1$

在 H_0 成立条件下的简约模型为 $Y=\beta_0 1_n + X1_k\beta_1+\varepsilon$
我们记 $X_1 := [1_n, X1_k]\in\mathbb{R}^{n\times 2}$ 即得 $Y=X_1\begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix}+\varepsilon$
因此关于简约模型我们有:

$$\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{bmatrix} := (X_1^{\mathrm{T}}X_1)^{-1}X_1^{\mathrm{T}}Y \text{ where } X_1 := [1_n, X1_k]$$

$$\mathrm{SSE}_{\text{reduced}} := \left\|Y - X_1\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{bmatrix}\right\|^2$$

而关于全模型我们有:

$$\begin{aligned}\tilde{X} &:= [1_n, X]\in\mathbb{R}^{n\times(k+1)}\\ \hat{\beta}_{\text{full}} &:= (\tilde{X}^{\mathrm{T}}\tilde{X})^{-1}\tilde{X}^{\mathrm{T}}Y\\ \mathrm{SSE}_{\text{full}} &:= \|Y-\tilde{X}\hat{\beta}_{\text{full}}\|^2\end{aligned}$$

额外平方和 $\mathrm{ESS} := \mathrm{SSE}_{\text{reduced}} - \mathrm{SSE}_{\text{full}} \sim \sigma^2\chi^2_{(k-1)}$ (考虑到等式约束的个数为 $k-1$ 所以其自由度为 $k-1$)
(实际上根据 Problem 5 的结论我们可以写出确切的表达式: $\mathrm{ESS} = (C\hat{\beta}_{\text{full}})^{\mathrm{T}}[C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-1}(C\hat{\beta}_{\text{full}})$)
因此我们可以定义 F-统计量:

$$F := \frac{\mathrm{ESS}/(k-1)}{\mathrm{SSE}_{\text{full}}/(n-k-1)} \sim F_{k-1,n-k-1}$$

检验法为:
若 $F > F_{k-1,n-k-1}(\alpha)$，就拒绝零假设 $H_0:\beta_1=\cdots=\beta_k$
其中 $F_{k-1,n-k-1}(\alpha)$ 为 $F_{k-1,n-k-1}$ 分布的 $1-\alpha$ 分位数.

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 F-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 the signal-plus-noise model such that

$$
Y = \mu + \varepsilon \text{ where } \begin{cases} \mu \in \mathbb{R}^n \\ \varepsilon \sim N(0_n, \sigma^2 I_n) \end{cases}
$$

(1) Find the BLUE of μ and its MSE $[10 \text{ Marks}]$
(2) Consider another shrinkage estimator $\hat{\mu} := [\hat{\mu}_1, \ldots, \hat{\mu}_n]^{\mathrm{T}}$ such that $\hat{\mu}_i = Y_i(1 - \frac{a}{\|Y\|^2})$ $(i = 1, \ldots, n)$
Find $a \in \mathbb{R}$ such that the MSE of $\hat{\mu}$ is smaller than the MSE of the BLUE $[10 \text{ Marks}]$
Hint: use Stein's lemma: If $X \sim N(\mu, \sigma^2)$ then $\mathrm{E}[(X-\mu)h(X)] = \sigma^2\mathrm{E}[h'(X)]$ with $h(X)$ continuous.

Part (1)

Find the BLUE of μ and its MSE $[10 \text{ Marks}]$

Solution:
考虑 μ 的线性估计量 $\hat{\mu} = C^{\mathrm{T}}Y$
要满足无偏性 $\mathrm{E}[\hat{\mu}] = \mathrm{E}[C^{\mathrm{T}}Y] = C^{\mathrm{T}}\mathrm{E}[Y] = C^{\mathrm{T}}\mu = \mu \ (\forall \ \mu \in \mathbb{R}^n)$
矩阵 $C \in \mathbb{R}^{n\times n}$ 必须满足 $C^{\mathrm{T}}\mu = \mu \ (\forall \ \mu \in \mathbb{R}^n)$
这意味着 C 只能是单位阵 I_n
因此 $\hat{\mu}_\star = Y$ 作为 μ 唯一的线性无偏估计量，一定是最佳线性无偏估计量 (BLUE)

下面计算 $\hat{\mu}_\star = Y$ 的 MSE:

$$
\begin{aligned}
\mathrm{MSE}(\hat{\mu}_\star) &= \mathrm{E}[\|\hat{\mu}_\star - \mu\|^2] \\
&= \mathrm{E}[\|Y - \mu\|_2^2] \\
&= \mathrm{E}[\|\varepsilon\|^2] \\
&= \mathrm{E}[\mathrm{tr}\left(\varepsilon^{\mathrm{T}}\varepsilon\right)] \\
&= \mathrm{E}[\mathrm{tr}\left(\varepsilon\varepsilon^{\mathrm{T}}\right)] \\
&= \mathrm{tr}\left(\mathrm{E}[\varepsilon\varepsilon^{\mathrm{T}}]\right) \\
&= \mathrm{tr}\left(\mathrm{Cov}(\varepsilon)\right) \\
&= \mathrm{tr}\left(\sigma^2 I_n\right) \\
&= n\sigma^2
\end{aligned}
$$

Part (2)

Consider another shrinkage estimator $\hat{\mu} := [\hat{\mu}_1, \ldots, \hat{\mu}_n]^{\mathrm{T}}$ such that $\hat{\mu}_i = Y_i(1 - \frac{a}{\|Y\|^2}) \ (i = 1, \ldots, n)$
Find $a \in \mathbb{R}$ such that the MSE of $\hat{\mu}$ is smaller than the MSE of the BLUE $[10 \text{ Marks}]$
Hint: use Stein's lemma: If $X \sim N(\mu, \sigma^2)$ then $\mathrm{E}[(X-\mu)h(X)] = \sigma^2\mathrm{E}[h'(X)]$ with $h(X)$ continuous.

Solution:
考虑 $\hat{\mu} = (1 - \frac{a}{\|Y\|^2})Y$，我们有:

$$
\begin{aligned}
\mathrm{MSE}(\hat{\mu}) - \mathrm{MSE}(\hat{\mu}_\star) &= \mathrm{E}[\|\hat{\mu} - \mu\|^2] - \mathrm{E}[\|\hat{\mu}_\star - \mu\|^2] \\
&= \mathrm{E}\left[\left\|(1 - \frac{a}{\|Y\|^2})Y - \mu\right\|^2\right] - \mathrm{E}[\|Y - \mu\|^2] \\
&= \mathrm{E}\left[\|Y - \mu\|^2 + a^2\frac{\|Y\|^2}{\|Y\|^4} - 2a(Y-\mu)^{\mathrm{T}}\frac{Y}{\|Y\|^2}\right] - \mathrm{E}[\|Y - \mu\|^2] \\
&= a^2\mathrm{E}\left[\frac{1}{\|Y\|^2}\right] - 2a\mathrm{E}\left[(Y-\mu)^{\mathrm{T}}\frac{Y}{\|Y\|^2}\right] \\
&= a^2\mathrm{E}\left[\frac{1}{\|Y\|^2}\right] - 2a\sum_{i=1}^n \mathrm{E}\left[(Y_i - \mu_i)\frac{Y_i}{\|Y\|^2}\right] \quad (\text{use Stein's lemma and note that } Y_i \sim N(\mu_i, \sigma^2)) \\
&= a^2\mathrm{E}\left[\frac{1}{\|Y\|^2}\right] - 2a\sum_{i=1}^n \sigma^2\mathrm{E}\left[\frac{\|Y\|^2 - 2Y_i^2}{\|Y\|^4}\right] \\
&= a^2\mathrm{E}\left[\frac{1}{\|Y\|^2}\right] - 2a\sigma^2\mathrm{E}\left[\frac{n\|Y\|^2 - 2\sum_{i=1}^n Y_i^2}{\|Y\|^4}\right] \\
&= a^2\mathrm{E}\left[\frac{1}{\|Y\|^2}\right] - 2a\sigma^2\mathrm{E}\left[\frac{n\|Y\|^2 - 2\|Y\|^2}{\|Y\|^4}\right] \\
&= a^2\mathrm{E}\left[\frac{1}{\|Y\|^2}\right] - 2a(n-2)\sigma^2\mathrm{E}\left[\frac{1}{\|Y\|^2}\right] \\
&= a(a - 2(n-2)\sigma^2)\mathrm{E}\left[\frac{1}{\|Y\|^2}\right]
\end{aligned}
$$

要令 $\mathrm{MSE}(\hat{\mu}) - \mathrm{MSE}(\hat{\mu}_\star) < 0$，等价于令 $a(a - 2(n-2)\sigma^2) < 0$
解得 $0 < a < 2(n-2)\sigma^2$

The End