FDU 回归分析 Homework 01

Due: Oct. 10, 2024
姓名: 雍崔扬
学号: 21307140051

Problem 1

Suppose the experimenter postulates a model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \ (i = 1, \dots, n)$$

where β_0 is known and $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2) \ (i = 1, \dots, n)$

- (a) What is the appropriate least squares estimator of β_1? Justify your answer.
- (b) What is the variance of the estimator of β_1 in (a)?

Part (a)

What is the appropriate least squares estimator of β_1? Justify your answer.

Solution:
$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \ (i = 1, \dots, n)$ 可以写成向量形式 $y = \beta_0 1_n + \beta_1 x + \varepsilon$
残差平方和 $\mathrm{RSS}(\beta_1) := \|y - \beta_0 1_n - \beta_1 x\|_2^2$
参数 β_1 的最小均方估计量 $\hat{\beta}_1$ 是以下优化问题的最小值点:

$$\begin{array}{l}
\hat{\beta}_1 := \arg\min_{\beta_1 \in \mathbb{R}} \mathrm{RSS}(\beta_1) \\
\hline
\begin{aligned}
\nabla_{\beta_1} \mathrm{RSS}(\beta_1) &= \nabla_{\beta_1} \{\|y - \beta_0 1_n - \beta_1 x\|_2^2\} \\
&= -2x^{\mathrm{T}}(y - \beta_0 1_n - \beta_1 x)
\end{aligned} \\
\hline
\nabla_{\beta_1}^2 \mathrm{RSS}(\beta_1) = 2x^{\mathrm{T}} x \geq 0
\end{array}$$

这是一个凸优化问题，因此最小值点 $\hat{\beta}_1$ 就是驻点.
令 $\nabla_{\beta_1} \mathrm{RSS}(\beta_1) = -2x^{\mathrm{T}}(y - \beta_0 1_n - \beta_1 x) = 0$，解得:

$$\hat{\beta}_1 = \frac{x^{\mathrm{T}}(y - \beta_0 1_n)}{x^{\mathrm{T}} x} = \frac{\sum_{i=1}^n x_i (y_i - \beta_0)}{\sum_{i=1}^n x_i^2}$$

Part (b)

What is the variance of the estimator of β_1 in (a)?

Solution:
注意到 $\varepsilon \sim N(0_n, \sigma^2 I_n)$

$$\begin{array}{l}
\begin{aligned}
\hat{\beta}_1 &= \frac{x^{\mathrm{T}}(y - \beta_0 1_n)}{x^{\mathrm{T}} x} \\
&= \frac{x^{\mathrm{T}}(\beta_0 1_n + \beta_1 x + \varepsilon - \beta_0 1_n)}{x^{\mathrm{T}} x} \\
&= \beta_1 + \frac{x^{\mathrm{T}} \varepsilon}{x^{\mathrm{T}} x}
\end{aligned} \\
\hline
\begin{aligned}
(\hat{\beta}_1 - \beta_1) &= \frac{x^{\mathrm{T}} \varepsilon}{x^{\mathrm{T}} x} \\
&\sim N\left(\frac{x^{\mathrm{T}}}{x^{\mathrm{T}} x} 0_n, \frac{x^{\mathrm{T}}}{x^{\mathrm{T}} x} \sigma^2 I_n \frac{x}{x^{\mathrm{T}} x}\right) \\
&= N\left(0, \frac{\sigma^2}{x^{\mathrm{T}} x}\right)
\end{aligned}
\end{array}$$

因此 $\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma^2}{x^{\mathrm{T}} x})$，表明 $\hat{\beta}_1$ 是 β_1 的无偏估计量，且 $\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{x^{\mathrm{T}} x}$

Problem 2

Consider a model $y_i = \beta + \varepsilon_i$ where β is a constant parameter and $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)\ (i = 1, \dots, n)$
Find LSE and MLE for β

Solution:
$y_i = \beta + \varepsilon_i\ (i = 1, \dots, n)$ 可以写成紧凑形式 $y = \beta 1_n + \varepsilon$

首先推导 β 的 LSE:

$$\hat{\beta}_{\mathrm{LSE}} := \arg\min_{\beta \in \mathbb{R}} \|y - \beta 1_n\|_2^2$$

令 $\nabla_\beta \{\|y - \beta 1_n\|_2^2\} = -1_n^{\mathrm{T}} \cdot 2(y - \beta 1_n) = 0$，解得 $\hat{\beta}_{\mathrm{LSE}} = \frac{1_n^{\mathrm{T}} y}{1_n^{\mathrm{T}} 1_n} = \frac{1}{n} 1_n^{\mathrm{T}} y = \bar{y}$

其次推导 β 的 MLE:
注意到 $y = \beta 1_n + \varepsilon \sim N(\beta 1_n, \sigma^2 I_n)$，故对数似然函数为:

$$\begin{aligned} L(\beta, \sigma^2) &:= \log\{\mathrm{P}\{N(\beta 1_n, \sigma^2 I_n) = y\}\} \\ &= \log\left\{\frac{1}{\sqrt{(2\pi)^n \det(\sigma^2 I_n)}} \exp\{-\frac{1}{2}(y - \beta 1_n)^{\mathrm{T}} (\sigma^2 I_n)^{-1} (y - \beta 1_n)\}\right\} \\ &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\|y - \beta 1_n\|_2^2 \end{aligned}$$

对于任意固定的 $\sigma^2 > 0$，则我们有:

$$\begin{aligned} \hat{\beta}_{\mathrm{MLE}} &:= \arg\max_{\beta \in \mathbb{R}} L(\beta, \sigma^2) \\ &= \arg\max_{\beta \in \mathbb{R}} \left\{-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\|y - \beta 1_n\|_2^2\right\} \\ &= \arg\min_{\beta \in \mathbb{R}} \|y - \beta 1_n\|_2^2 \\ &= \hat{\beta}_{\mathrm{LSE}} \\ &= \bar{y} \end{aligned}$$

固定 $\beta = \hat{\beta}_{\mathrm{MLE}} = \bar{y}$，则我们有:

$$\begin{aligned} \hat{\sigma}^2_{\mathrm{MLE}} &:= \arg\max_{\sigma^2 > 0} L(\bar{y}, \sigma^2) \\ &= \arg\max_{\sigma^2 > 0} \left\{-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\|y - \bar{y} 1_n\|_2^2\right\} \\ &= \text{solution of } \left\{-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\|y - \bar{y} 1_n\|_2^2 = 0\right\} \\ &= \frac{1}{n}\|y - \bar{y} 1_n\|_2^2 \end{aligned}$$

Problem 3

A sample of n boys and n girls is taken from a school and their heights are measured.
Let $y_1, \dots, y_n$ denote the heights of the n girls, and $y_{n+1}, \dots, y_{2n}$ denote those of the n boys.
It is believed that $y_i = \alpha + \beta x_i + \varepsilon_i$ where $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)\ (i = 1, \dots, 2n)$ and $x_i = \begin{cases} -1 & i = 1, \dots, n \\ +1 & i = n+1, \dots, 2n \end{cases}$
Find the least squares estimators of β

- Note: x is a dummy variable used to represent gender.

Solution:
我们记

$$x=\begin{bmatrix}x_1\\ \vdots\\ x_n\\ x_{n+1}\\ \vdots\\ x_{2n}\end{bmatrix}=\begin{bmatrix}-1\\ \vdots\\ -1\\ +1\\ \vdots\\ +1\end{bmatrix}\quad X=[1_{2n};x]\quad \gamma=\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\quad \varepsilon=\begin{bmatrix}\varepsilon_1\\ \vdots\\ \varepsilon_n\\ \varepsilon_{n+1}\\ \vdots\\ \varepsilon_{2n}\end{bmatrix}$$

$$y^{(1)}=\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix}\quad y^{(2)}=\begin{bmatrix}y_{n+1}\\ \vdots\\ y_{2n}\end{bmatrix}\quad y=\begin{bmatrix}y^{(1)}\\ y^{(2)}\end{bmatrix}$$

则我们可将 $y_i=\alpha+\beta x_i+\varepsilon_i\ (i=1,\dots,2n)$ 记为 $y=\alpha 1_{2n}+\beta x+\varepsilon=X\gamma+\varepsilon$

$$\begin{aligned}\hat{\gamma}&:=\arg\min_{\gamma\in\mathbb{R}^2}\mathrm{RSS}(\gamma)\\ &=\arg\min_{\gamma\in\mathbb{R}^2}\|y-X\gamma\|_2^2\\ &=\text{solution of }\nabla_\gamma\{\|y-X\gamma\|_2^2\}=-2X^{\mathrm{T}}(y-X\gamma)=0_2\\ &=(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y\\ &=\begin{bmatrix}1_{2n}^{\mathrm{T}}1_{2n} & 1_{2n}^{\mathrm{T}}x\\ 1_{2n}^{\mathrm{T}}x & x^{\mathrm{T}}x\end{bmatrix}^{-1}\begin{bmatrix}1_{2n}^{\mathrm{T}}y\\ x^{\mathrm{T}}y\end{bmatrix}\\ &=\begin{bmatrix}2n & 0\\ 0 & 2n\end{bmatrix}^{-1}\begin{bmatrix}\bar{y}\\ x^{\mathrm{T}}y\end{bmatrix}\\ &=\frac{1}{2n}\begin{bmatrix}\bar{y}\\ -1_n^{\mathrm{T}}y^{(1)}+1_n^{\mathrm{T}}y^{(2)}\end{bmatrix}\end{aligned}$$

因此 $\hat{\beta}$ 作为 $\hat{\gamma}$ 的第 2 个分量，它为:

$$\hat{\beta}=\frac{1}{2n}(-1_n^{\mathrm{T}}y^{(1)}+1_n^{\mathrm{T}}y^{(2)})=\frac{1}{2n}\sum_{i=1}^n(y_{n+i}-y_i)$$

Problem 4

Consider a simple linear regression model:

$$y_i=\alpha+\beta x_i+\varepsilon_i\ (i=1,\dots,n)$$

where $x_1,\dots,x_n$ are constants, α,β are parameters, $\varepsilon_i\overset{iid}{\sim}N(0,\sigma^2)\ (i=1,\dots,n)$
A weighted least square estimators for α,β are obtained by minimizing the Risidual Sum of Squares:

$$\mathrm{RSS}(\alpha,\beta):=\sum_{i=1}^n w_i(y_i-\alpha-\beta x_i)^2$$

where $w_1,\dots,w_n$ are some predefined known **positive** constant values and $\sum_{i=1}^n w_i=1$
Find the weighted least square estimators for α,β and their covariance matrix.

Solution:
$y_i=\alpha+\beta x_i+\varepsilon_i\ (i=1,\dots,n)$ 可以写成紧凑形式 $y=\alpha 1_n+\beta x+\varepsilon$，其中 $\varepsilon\sim N(0_n,\sigma^2 I_n)$

记 $\begin{cases}W=\mathrm{diag}\{w_1,\dots,w_n\}\\ w=[w_1,\dots,w_n]^{\mathrm{T}}\\ \gamma=[\alpha,\beta]^{\mathrm{T}}\\ X=[1_n;x]\end{cases}$ 则我们有:

$$\begin{array}{c}y=\alpha 1_n+\beta x+\varepsilon=X\gamma+\varepsilon\\ \hline \mathrm{RSS}(\alpha,\beta):=\sum_{i=1}^n w_i(y_i-\alpha-\beta x_i)^2=(y-X\gamma)^{\mathrm{T}}W(y-X\gamma)\text{ where }W\text{ satisfies }\begin{cases}W\succ 0\\ \mathrm{tr}\,(W)=1\end{cases}\end{array}$$

于是 γ 的最小二乘估计量为如下优化问题的全局最小点:

知乎@Snivellus Snape

$$
\hat{\gamma} := \arg\min_{\gamma \in \mathbb{R}^2} (y - X\gamma)^{\mathrm{T}} W (y - X\gamma)
$$

$$
\begin{aligned}
\nabla_\gamma \{(y - X\gamma)^{\mathrm{T}} W (y - X\gamma)\} &= \nabla_\gamma \{y^{\mathrm{T}} W y - 2y^{\mathrm{T}} W X\gamma + \gamma X^{\mathrm{T}} W X \gamma\} \\
&= -2X^{\mathrm{T}} W y + 2X^{\mathrm{T}} W X \gamma \\
\nabla^2_\gamma \{(y - X\gamma)^{\mathrm{T}} W (y - X\gamma)\} &= 2X^{\mathrm{T}} W X \succ 0
\end{aligned}
$$

因此这是一个无约束凸优化问题，其全局最小点即为驻点.
令 $\nabla_\gamma \{(y - X\gamma)^{\mathrm{T}} W (y - X\gamma)\} = -2X^{\mathrm{T}} W y + 2X^{\mathrm{T}} W X \gamma = 0_2$ 可得:

$$
\begin{aligned}
\hat{\gamma} &= (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W y \\
&= \begin{bmatrix} 1_n^{\mathrm{T}} W 1_n & 1_n^{\mathrm{T}} W x \\ x^{\mathrm{T}} W 1_n & x^{\mathrm{T}} W x \end{bmatrix}^{-1} \begin{bmatrix} 1_n^{\mathrm{T}} W y \\ x^{\mathrm{T}} W y \end{bmatrix} \quad (\text{note that } \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \text{ if } ad - bc \neq 0) \\
&= \frac{1}{(1_n^{\mathrm{T}} W 1_n)(x^{\mathrm{T}} W x) - (1_n^{\mathrm{T}} W x)(x^{\mathrm{T}} W 1_n)} \begin{bmatrix} x^{\mathrm{T}} W x & -1_n^{\mathrm{T}} W x \\ -x^{\mathrm{T}} W 1_n & 1_n^{\mathrm{T}} W 1_n \end{bmatrix} \begin{bmatrix} 1_n^{\mathrm{T}} W y \\ x^{\mathrm{T}} W y \end{bmatrix} \quad (\text{note that } \begin{cases} 1_n^{\mathrm{T}} W 1_n = \sum_{i=1}^n w_i = 1 \\ W 1_n = w \end{cases}) \\
&= \frac{1}{x^{\mathrm{T}} W x - (w^{\mathrm{T}} x)^2} \begin{bmatrix} (x^{\mathrm{T}} W x)(w^{\mathrm{T}} y) - (w^{\mathrm{T}} x)(x^{\mathrm{T}} W y) \\ -(w^{\mathrm{T}} x)(w^{\mathrm{T}} y) + x^{\mathrm{T}} W y \end{bmatrix}
\end{aligned}
$$

(根据上述讨论我们也知道 $(X^{\mathrm{T}} W X)^{-1} = \frac{1}{x^{\mathrm{T}} W x - (w^{\mathrm{T}} x)^2} \begin{bmatrix} x^{\mathrm{T}} W x & -w^{\mathrm{T}} x \\ -w^{\mathrm{T}} x & 1 \end{bmatrix}$)
因此 α, β 的加权最小二乘估计量 $\hat{\alpha}, \hat{\beta}$ 分别为 $\hat{\gamma}$ 的两个分量:

$$
\begin{aligned}
\hat{\alpha} &= \frac{(x^{\mathrm{T}} W x)(w^{\mathrm{T}} y) - (w^{\mathrm{T}} x)(x^{\mathrm{T}} W y)}{x^{\mathrm{T}} W x - (w^{\mathrm{T}} x)^2} \\
&= \frac{(\sum_{i=1}^n w_i x_i^2)(\sum_{i=1}^n w_i y_i) - (\sum_{i=1}^n w_i x_i)(\sum_{i=1}^n w_i x_i y_i)}{\sum_{i=1}^n w_i x_i^2 - (\sum_{i=1}^n w_i x)^2} \\
\hat{\beta} &= \frac{x^{\mathrm{T}} W y - (w^{\mathrm{T}} x)(w^{\mathrm{T}} y)}{x^{\mathrm{T}} W x - (w^{\mathrm{T}} x)^2} \\
&= \frac{\sum_{i=1}^n w_i x_i y_i - (\sum_{i=1}^n w_i x_i)(\sum_{i=1}^n w_i y_i)}{\sum_{i=1}^n w_i x_i^2 - (\sum_{i=1}^n w_i x)^2}
\end{aligned}
$$

下面我们考虑 $\hat{\gamma} = [\hat{\alpha}, \hat{\beta}]^{\mathrm{T}}$ 的分布:

$$
\begin{aligned}
\hat{\gamma} &= (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W y \\
&= (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W (X\gamma + \varepsilon) \\
&= \gamma + (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W \varepsilon \\
\hat{\gamma} - \gamma &= (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W \varepsilon \\
&\sim N((X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W \cdot 0_n, (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W \cdot \sigma^2 I_n \cdot [(X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W]^{\mathrm{T}}) \\
&= N(0_n, \sigma^2 \cdot (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W^2 X (X^{\mathrm{T}} W X)^{-1})
\end{aligned}
$$

因此 $\hat{\alpha}, \hat{\beta}$ 的协方差矩阵为:

$$
\begin{aligned}
\mathrm{Cov}(\hat{\gamma}) &= \sigma^2 \cdot (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W^2 X (X^{\mathrm{T}} W X)^{-1} \quad (\text{note that } (X^{\mathrm{T}} W X)^{-1} = \frac{1}{x^{\mathrm{T}} W x - (w^{\mathrm{T}} x)^2} \begin{bmatrix} x^{\mathrm{T}} W x & -w^{\mathrm{T}} x \\ -w^{\mathrm{T}} x & 1 \end{bmatrix}) \\
&= \frac{\sigma^2}{[x^{\mathrm{T}} W x - (w^{\mathrm{T}} x)^2]^2} \begin{bmatrix} x^{\mathrm{T}} W x & -w^{\mathrm{T}} x \\ -w^{\mathrm{T}} x & 1 \end{bmatrix} \begin{bmatrix} 1_n^{\mathrm{T}} W^2 1_n & 1_n^{\mathrm{T}} W^2 x \\ x^{\mathrm{T}} W^2 1_n & x^{\mathrm{T}} W^2 x \end{bmatrix} \begin{bmatrix} x^{\mathrm{T}} W x & -w^{\mathrm{T}} x \\ -w^{\mathrm{T}} x & 1 \end{bmatrix}
\end{aligned}
$$

其中 $\begin{cases} W = \mathrm{diag}\{w_1, \ldots, w_n\} \\ w = [w_1, \ldots, w_n]^{\mathrm{T}} \end{cases}$ (不想继续化简了，这没有什么意义)

Problem 5

Consider a simple linear regression model:

$$
y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \ (i = 1, \ldots, n)
$$

where $x_1, \ldots, x_n$ are constants, α, β are parameters, $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)\ (i = 1, \ldots, n)$
Suppose $n = 3$ and three data points: $(-1, y_1), (0, y_2), (1, y_3)$

- (a) Find the design matrix X
- (b) Express the least squares estimators of β_0, β_1 in terms of y_1, y_2, y_3

- (c) Find out the expression of the regression sum of squares in terms of y_1, y_2, y_3
- (d) Suppose $y_3 = y_1 + 2$ and the total sum of squares $\mathrm{SST} := S_{yy} = \sum_{i=1}^{3}(y_i - \bar{y})^2 = 2.5$
  What is the value of the coefficient of determination R^2?
  Does the regression line fit the three data points well?
- (e) Under the condition of (d), find the 95% confidence intervals for β_0, β_1 respectively.

Part (a)

Find the design matrix X

Solution:

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$$

于是 $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \ (i = 1, 2, 3)$ 可以写成紧凑形式 $y = X\beta + \varepsilon$ (其中 $\beta = [\beta_0, \beta_1]^{\mathrm{T}}$)

Part (b)

Express the least squares estimators of β_0, β_1 in terms of y_1, y_2, y_3

Solution:

$$\begin{aligned} \hat{\beta} &:= \arg\min_{\beta \in \mathbb{R}^2} \|y - X\beta\|_2^2 \\ &= \text{solution of } \{-2X^{\mathrm{T}}(y - X\beta) = 0_2\} \\ &= (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y \\ &= \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{3}(y_1 + y_2 + y_3) \\ \frac{1}{2}(y_3 - y_1) \end{bmatrix} \end{aligned}$$

因此我们有:

$$\begin{cases} \hat{\beta}_0 = \frac{1}{3}(y_1 + y_2 + y_3) = \bar{y} \\ \hat{\beta}_1 = \frac{1}{2}(y_3 - y_1) \end{cases}$$

Part (c)

Find out the expression of the regression sum of squares in terms of y_1, y_2, y_3

Solution:

知乎@Snivellus Snape

$$
\begin{aligned}
\mathrm{SSR} &= \|\hat{y}-\bar{y}1_3\|_2^2 \\
&= \|X\hat{\beta}-\frac{1}{3}1_3^{\mathrm{T}}y1_3\| \quad (\text{note that } \hat{\beta}=(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y) \\
&= \|[X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}-\frac{1}{3}1_31_3^{\mathrm{T}}]y\|_2^2 \\
&= y^{\mathrm{T}}[X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}-\frac{1}{3}1_31_3^{\mathrm{T}}]y \\
&= y^{\mathrm{T}}\left(\begin{bmatrix}1 & -1\\ 1 & 0\\ 1 & 1\end{bmatrix}\begin{bmatrix}3 & 0\\ 0 & 2\end{bmatrix}^{-1}\begin{bmatrix}1 & -1\\ 1 & 0\\ 1 & 1\end{bmatrix}^{\mathrm{T}}-\frac{1}{3}\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix}\right)y \\
&= y^{\mathrm{T}}\left(\begin{bmatrix}1 & -1\\ 1 & 0\\ 1 & 1\end{bmatrix}\begin{bmatrix}\frac{1}{3} & \\ & \frac{1}{2}\end{bmatrix}\begin{bmatrix}1 & 1 & 1\\ -1 & 0 & 1\end{bmatrix}-\frac{1}{3}\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix}\right)y \\
&= y^{\mathrm{T}}\left(\begin{bmatrix}\frac{5}{6} & \frac{1}{3} & -\frac{1}{6}\\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3}\\ -\frac{1}{6} & \frac{1}{3} & \frac{5}{6}\end{bmatrix}-\frac{1}{3}\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix}\right)y \\
&= y^{\mathrm{T}}\begin{bmatrix}\frac{1}{2} & 0 & -\frac{1}{2}\\ 0 & 0 & 0\\ -\frac{1}{2} & 0 & \frac{1}{2}\end{bmatrix}y \\
&= \frac{1}{2}(y_1-y_3)^2
\end{aligned}
$$

另法:

$$
\begin{aligned}
\mathrm{SSR} &= \|\hat{y}-\bar{y}1_n\|_2^2 \\
&= \|\hat{\beta}_01_n+\hat{\beta}_1x-\bar{y}1_n\|_2^2 \\
&= \|(\bar{y}-\hat{\beta}_1\bar{x})1_n+\hat{\beta}_1x-\bar{y}1_n\|_2^2 \\
&= \|\hat{\beta}_1(x-\bar{x}1_n)\|_2^2 \\
&= \hat{\beta}_1^2S_{xx} \\
&= \left[\frac{1}{2}(y_3-y_1)\right]^2\cdot 2 \\
&= \frac{1}{2}(y_1-y_3)^2
\end{aligned}
$$

Part (d)

Suppose $y_3=y_1+2$ and the total sum of squares $\mathrm{SST}:=S_{yy}=\sum_{i=1}^3(y_i-\bar{y})^2=2.5$
What is the value of the coefficient of determination R^2?
Does the regression line fit the three data points well?

Solution:
由于 $y_3=y_1+2$，故 $\mathrm{SSR}=\frac{1}{2}(y_1-y_3)^2=\frac{1}{2}\cdot 2^2=2$
因此 $R^2=\frac{\mathrm{SSR}}{\mathrm{SST}}=\frac{2}{2.5}=0.8$
这个数值很接近于 1，故模型较好地拟合了 3 个数据点.

Part (e)

Under the condition of (d), find the 95% confidence intervals for β_0,β_1 respectively.

Solution:
根据 $\begin{cases}y_3=y_1+2\\ S_{yy}=\sum_{i=1}^3(y_i-\bar{y})^2=y_1^2+y_2^2+y_3^2-3\bar{y}^2=2.5\\ \bar{y}=\frac{1}{3}(y_1+y_2+y_3)\end{cases}$ 我们可以得到:

知乎 @Snivellus Snape

$$
\begin{aligned}
2.5 &= y_1^2 + y_2^2 + (y_1+2)^2 - 3 \cdot \frac{1}{9}(2y_1 + y_2 + 2)^2 \\
&= 2y_1^2 + y_2^2 + 4y_1 + 4 - \frac{1}{3}(4y_1^2 + y_2^2 + 4 + 4y_1y_2 + 8y_1 + 4y_2) \\
&= \frac{2}{3}y_1^2 + \frac{2}{3}y_2^2 + \frac{4}{3}y_1 - \frac{4}{3}y_2 - \frac{4}{3}y_1y_2 + \frac{8}{3} \\
&= \frac{2}{3}(y_1^2 + y_2^2 + 1 + 2y_1 - 2y_2 - 2y_1y_2) + 2 \\
&= \frac{2}{3}(y_1 - y_2 + 1)^2 + 2
\end{aligned}
$$

$$
(y_1 - y_2 + 1)^2 = \frac{3}{4} \ \Rightarrow \ (y_1 - y_2) = -1 \pm \frac{\sqrt{3}}{2}
$$

但遗憾的是，根据已有数据我们仍然无法计算出 $\hat{\beta}_0 = \bar{y}$，我们就当 $\bar{y}$ 是已知的就好了.
根据 **Part (b)** 我们有:

$$
\begin{aligned}
\hat{\beta} &:= \arg\min_{\beta \in \mathbb{R}^2} \|y - X\beta\|_2^2 \\
&= \text{solution of } \{-2X^{\mathrm{T}}(y - X\beta) = 0_2\} \\
&= (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y \\
&= \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{3}(y_1 + y_2 + y_3) \\ \frac{1}{2}(y_3 - y_1) \end{bmatrix} \quad (\text{note that } y_3 = y_1 + 2) \\
&= \begin{bmatrix} \bar{y} \\ 1 \end{bmatrix}
\end{aligned}
$$

因此我们有:

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

$$
\begin{aligned}
\hat{\beta} - \beta &= (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\varepsilon \\
&\sim N((X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}0_n, (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}} \cdot \sigma^2 I_n \cdot [(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}]^{\mathrm{T}}) \\
&= N(0_2, \sigma^2(X^{\mathrm{T}}X)^{-1}) \\
&= N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}^{-1}\right) \\
&= N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{3}\sigma^2 & 0 \\ 0 & \frac{1}{2}\sigma^2 \end{bmatrix}\right)
\end{aligned}
$$

而 σ^2 的无偏估计量 $s^2 = \frac{1}{n-2}\mathrm{SSE} = \frac{1}{n-2}(\mathrm{SST} - \mathrm{SSR}) = \frac{1}{3-2}(2.5 - 2) = 0.5 \sim \sigma^2\chi^2_{(1)}$
因此用于对 β_0, β_1 进行区间估计的枢轴量为:

$$
\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\frac{1}{3}s^2}} = \sqrt{6}(\bar{y} - \beta_0) \sim t_1
$$

$$
\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{1}{2}s^2}} = 2(1 - \beta_1) \sim t_1
$$

因此 β_0 的 95% 置信区间为 $[\bar{y} - \frac{\sqrt{6}}{6}t_{1,0.975}, \bar{y} + \frac{\sqrt{6}}{6}t_{1,0.975}]$ (其中我们假设 $\bar{y}$ 已知)
而 β_1 的 95% 置信区间为 $[1 - \frac{1}{2}t_{1,0.975}, 1 + \frac{1}{2}t_{1,0.975}]$
其中 $t_{1,0.975} \approx 12.71$ 为自由度为 1 的 t 分布的 $1 - \frac{1}{2}(1 - 0.95) = 0.975$ 分位数.

The End

知乎@Shivellus Snape

知乎@Snivellus Snape