Lista 3

MI406-Regressão

Caio Gomes Alves

1 Questão 1

1.1 Pergunta

Considere o modelo definido por

$$\mathbf{Y_i} = \mu + \epsilon_\mathbf{i}, \quad i = 1, \ldots, n$$

onde $\mu \in \mathbb{R}$ é uma constante e $\epsilon_i \sim N(0, \sigma^2)$, com $\epsilon_i \perp \epsilon_j$ para todo $i \neq j$.

- **a.** Escreva o modelo descrito em forma matricial. Isto é, $\mathbf{Y} = \mathbf{X}\beta + \epsilon$, e descreva os vetores e matrizes envolvidos.
- **b.** Sabendo que o Estimador de Máxima-Verossimilhança (MV) de μ é $\bar{Y}$, mostre que o estimador de Mínimos-Quadrados $(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$ coincide com o estimador de MV.
- **c.** Descreva qual a forma da matriz de projeção $\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{\mathbf{-1}}\mathbf{X}^\top$ e interprete.

1.2 Resposta

1.2.1 a)

Temos que o modelo, em forma matricial é dado da seguinte forma:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} \qquad X = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n\times 1} \qquad \beta = \begin{bmatrix} \mu \end{bmatrix}_{1\times 1} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}_{n\times 1}$$

Aqui, a matriz Y representa o vetor de observações, onde cada observação corresponde à uma amostra da variável resposta. A matriz X indica a estrutura do modelo, que neste caso é apenas o modelo com intercepto.

A matriz β indica o vetor de parâmetros da regressão, que minimizam a soma de quadrados do modelo especificado pela matriz X (que neste caso é denotada somente por um parâmetro, μ). Por fim, a matriz ϵ indica os erros (leia-se desvios) de cada uma das observações em Y, seguindo uma distribuição Normal com média 0 e variância σ^2 (aqui, assume-se que as observações são não-correlacionadas e que essa parcela de “erro” não é explicada pelo modelo de maneira estrutural).

1.2.2 b)

Considerando a matriz X denotada anteriormente, temos que:

$$(X^\top X) = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} n \end{bmatrix}$$

$$(X^\top X)^{-1} = \begin{bmatrix} \frac{1}{n} \end{bmatrix}$$

$$(X^\top X)^{-1} X^\top = \begin{bmatrix} \frac{1}{n} \end{bmatrix} \times \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{bmatrix}$$

$$(X^\top X)^{-1} X^\top Y = \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{bmatrix} \times \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \frac{\sum_{i=1}^n y_i}{n} \end{bmatrix} = \begin{bmatrix} \bar{Y} \end{bmatrix}$$

Assim, vemos que o estimador de mínimos quadrados coincide com o estimador de máxima verossimilhança.

1.2.3 c)

Temos que:

$$X(X^\top X)^{-1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \times \begin{bmatrix} \frac{1}{n} \end{bmatrix} = \begin{bmatrix} 1/n \\ 1/n \\ \vdots \\ 1/n \end{bmatrix}$$

$$X(X^\top X)^{-1} X^\top = \begin{bmatrix} 1/n \\ 1/n \\ \vdots \\ 1/n \end{bmatrix} \times \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{bmatrix} = H$$

Assim, a matriz de projeção H é tal que estima cada um dos valores de Y_i na sua média $\bar{Y}$, ou seja, $\hat{Y}_i = \bar{Y}, \forall i \in \{1, \dots, n\}$.

2 Questão 2

2.1 Pergunta

Seja J_n uma matriz de dimensões $n \times n$ com o valor 1 em todas as entradas e H a matriz de projeção $H = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-\mathbf{1}}\mathbf{X}^\top$.

- **a.** Mostre (ou justifique que) $\frac{1}{n}J_n$ é simétrica e idempotente.
- **b.** Mostre (ou justifique que)

$$\mathbf{Y}^\top \left(\mathbf{I} - \frac{\mathbf{1}}{\mathbf{n}}\mathbf{J_n}\right)\mathbf{Y} = \mathbf{Y}^\top \left(\mathbf{H} - \frac{\mathbf{1}}{\mathbf{n}}\mathbf{J_n}\right)\mathbf{Y} + \mathbf{Y}^\top(\mathbf{I} - \mathbf{H})\mathbf{Y},$$

Interprete esse resultado.

2.2 Resposta

2.2.1 a)

Sabemos que:

$$J_n = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix}_{n \times n} \Longrightarrow \frac{1}{n}J_n = \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{bmatrix}_{n \times n}$$

A simetria dessas matrizes é de fácil verificação, já que todos os elementos dela são iguais. Para ser idempotente, além de ser simétrica é necessário que as potências da matriz resultem nela mesma. Vejamos que:

$$\begin{aligned}
\left(\frac{1}{n}J_n\right) \times \left(\frac{1}{n}J_n\right) &= \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{bmatrix} \times \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{bmatrix} \\
&= \begin{bmatrix} n\left(\frac{1}{n}\right)^2 & n\left(\frac{1}{n}\right)^2 & \dots & n\left(\frac{1}{n}\right)^2 \\ n\left(\frac{1}{n}\right)^2 & n\left(\frac{1}{n}\right)^2 & \dots & n\left(\frac{1}{n}\right)^2 \\ \vdots & \vdots & \ddots & \vdots \\ n\left(\frac{1}{n}\right)^2 & n\left(\frac{1}{n}\right)^2 & \dots & n\left(\frac{1}{n}\right)^2 \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{bmatrix} \\
&= \frac{1}{n}J_n
\end{aligned}$$

Assim, a matriz $\frac{1}{n}J_n$ é idempotente.

2.2.2 b)

Temos que a decomposição da soma de quadrados total de um modelo é dada por:

$$SQT = SQReg + SQRes$$

E que, cada componente tem a seguinte forma quadrática matricial:

$$
\begin{aligned}
SQT &= \left(Y - \bar{Y}\right)^\top \left(Y - \bar{Y}\right) \\
SQRes &= \left(Y - \hat{Y}\right)^\top \left(Y - \hat{Y}\right) \\
SQReg &= \left(\hat{Y} - \bar{Y}\right)^\top \left(\hat{Y} - \bar{Y}\right)
\end{aligned}
$$

E agora, desenvolvendo a álgebra matricial necessária, chegamos em:

$$
\begin{aligned}
\left(Y - \bar{Y}\right)^\top \left(Y - \bar{Y}\right) &= \left(\hat{Y} - \bar{Y}\right)^\top \left(\hat{Y} - \bar{Y}\right) + \left(Y - \hat{Y}\right)^\top \left(Y - \hat{Y}\right) \\
\left(Y - \frac{1}{n}J_nY\right)^\top \left(Y - \frac{1}{n}J_nY\right) &= \left(HY - \frac{1}{n}J_nY\right)^\top \left(HY - \frac{1}{n}J_nY\right) + \left(Y - HY\right)^\top \left(Y - HY\right) \\
\left(\left(I - \frac{1}{n}J_n\right)Y\right)^\top \left(\left(I - \frac{1}{n}J_n\right)Y\right) &= \left(\left(H - \frac{1}{n}J_n\right)Y\right)^\top \left(\left(H - \frac{1}{n}J_n\right)Y\right) + \left((I-H)Y\right)^\top \left((I-H)Y\right) \\
Y^\top \left(I - \frac{1}{n}J_n\right)^\top \left(I - \frac{1}{n}J_n\right)Y &= Y^\top \left(H - \frac{1}{n}J_n\right)^\top \left(H - \frac{1}{n}J_n\right)Y + Y^\top \left(I-H\right)^\top \left(I-H\right)Y \qquad (1)
\end{aligned}
$$

E como vimos na questão anterior, temos que $\frac{1}{n}J_n$ é idempontente, assim como serão as matrizes $\left(I - \frac{1}{n}J_n\right)$, $\left(H - \frac{1}{n}J_n\right)$ e $(I - H)$, de modo que a expressão (1) fica:

$$
Y^\top \left(I - \frac{1}{n}J_n\right)Y = Y^\top \left(H - \frac{1}{n}J_n\right)Y + Y^\top(I-H)Y
$$

3 Questão 3

3.1 Pergunta

Seja $\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta}$.

- **a.** Mostre que
$$\hat{\mathbf{Y}} \sim N_n(X\beta, \sigma^2 H)$$
.
- **b.** Compare as distribuições de $\mathbf{Y}$ e $\hat{\mathbf{Y}}$. Interprete.
- **c.** Como podemos interpretar a entrada h_{ii} da matriz H?

3.2 Resposta

3.2.1 a)

Temos que:

$$\hat{\beta} = \left(X^\top X\right)^{-1} X^\top Y$$

$$\begin{aligned}
\mathbb{E}(\hat{\beta}) &= \mathbb{E}((X^\top X)^{-1} X^\top Y) \\
&= (X^\top X)^{-1} X^\top \mathbb{E}(Y) \\
&= (X^\top X)^{-1} X^\top X\beta \\
&= \beta
\end{aligned}$$

$$\begin{aligned}
Var(\hat{\beta}) &= Var((X^\top X)^{-1} X^\top Y) \\
&= (X^\top X)^{-1} X^\top Var(Y)((X^\top X)^{-1} X^\top)^\top \\
&= \sigma^2 I (X^\top X)^{-1} X^\top X (X^\top X)^{-1} \\
&= \sigma^2 (X^\top X)^{-1}
\end{aligned}$$

Aqui, vale notar que β é um vetor $(p+1)\times 1$ e $\sigma^2(X^\top X)^{-1}$ é uma matriz $(p+1)\times(p+1)$, de modo que a distribuição de $\hat{\beta}$ é:

$$\hat{\beta} \sim N_{p+1}\left(\beta, \sigma^2(X^\top X)^{-1}\right)$$

Disso, como $\hat{Y} = X\hat{\beta}$, temos que:

$$\begin{aligned}
\mathbb{E}(X\hat{\beta}) &= X\mathbb{E}(\hat{\beta}) \\
&= X\beta
\end{aligned}$$

$$\begin{aligned}
Var(X\hat{\beta}) &= XVar(\hat{\beta})X^\top \\
&= \sigma^2 X(X^\top X)^{-1} X^\top \\
&= \sigma^2 H
\end{aligned}$$

Neste caso, vale notar que $X\beta$ é um vetor $n\times 1$ e $\sigma^2 H$ é uma matriz $n\times n$, e assim, a distribuição de $\hat{Y} = X\hat{\beta}$ é:

$$\hat{Y} \sim N_n(X\beta, \sigma^2 H)$$

3.2.2 b)

A distribuição de Y é condicional (assim como no modelo de regressão linear simples), dada por:

$$Y|X \sim N_n(X\beta, \sigma^2 I_n)$$

Que é definida pela suposição de normalidade, independência e homocedasticidade dos ϵ_i's. Dessa forma, as médias (condicionais) são as mesmas para ambas as distribuições, mas as variâncias diferem. Enquanto que $Y|X$ possui variância constante, a variância de $\hat{Y}$ depende do valor da matriz de projeção H. Assim, a estimação de Y será mais incerta para pontos que estão distantes de $\bar{X}$, e mais precisa em pontos próximos de $\bar{X}$.

3.2.3 c)

Cada entrada h_{ii} da diagonal principal da matriz H pode ser vista como a influência que um ponto Y_i terá na sua própria estimação $\hat{Y}_i$, visto que ela é a matriz de projeção das observações Y_i no espaço gerado pelas combinações lineares das colunas de X.

Por ser uma matriz de projeção, os elementos da diagnoal principal de H estarão entre 0 e 1, e quanto mais próximo de 0 um elemento h_{ii} estiver, menos influente é a observação Y_i na estimação de $\hat{Y}$, e quanto mais próximo de 1, mais influente ele será, e será um ponto que "domina" a estimação de $\hat{Y}$. Isso leva a uma outra definição da matriz H, que também é conhecida como a matriz de alavancagem do modelo.

Para além disso, cada elemento h_{ii} também mostra o quão influente um ponto será na variância da estimação naquele ponto.

4 Questão 4

4.1 Pergunta

Sabendo que a existência dos estimadores de Mínimos Quadrados e da Matriz de Projeção dependem da inversa $(\mathbf{X}^\top\mathbf{X})^{-\mathbf{1}}$.

- **a.** Apresente a forma geral da matriz $\mathbf{X}^\top\mathbf{X}$ no contexto de regressão linear simples e descreva as condições necessárias para a existência de sua inversa.
- **b.** Dê um exemplo de valores das variáveis $x_1, x_2, \ldots, x_n$ para os quais a inversa não existe.
- **c.** Do ponto de vista da interpretação dos parâmetros, explique o motivo pelo qual o cenário do exemplo do item b não nos permite ter estimadores únicos para β_0 e β_1.

4.2 Resposta

4.2.1 a)

Sendo o modelo $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, ou de maneira matricial $Y = X\beta + \epsilon$, temos que:

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}_{n\times 2} \qquad X^\top = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_n \end{bmatrix}_{2\times n}$$

$$(X^\top X) = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} = \begin{bmatrix} n & n\bar{x} \\ \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{bmatrix}$$

Neste caso, para que $(X^\top X)$ seja inversível, é necessário que o determinante dessa matriz não seja igual a 0, ou seja:

$$n\sum_{i=1}^{n} x_i^2 - n^2\bar{x}^2 \neq 0$$
$$n\left(\sum_{i=1}^{n} x_i^2 - \bar{x}^2\right) \neq 0$$
$$\sum_{i=1}^{n}(x_i - \bar{x})^2 \neq 0$$

Assim, para o caso da regressão linear simples, não pode ocorrer da soma de quadrados total do modelo ser 0, e isso só irá ocorrer se todos os valores dos x_i's forem iguais.

4.2.2 b)

Para o caso de regressão linear simples, a matriz $X^\top X$ não terá inversa quando todos os valores de x forem iguais. Ou seja, considere por exemplo que $x_i = k,\ \forall i \in \{1, \ldots, n\} \Rightarrow \bar{x} = k \Rightarrow \sum_{i=1}^{n}(x_i - \bar{x})^2 = 0$. Nesse caso, não haverá inversa para a matriz $X^\top X$, de modo que não será possível ajustar um modelo de regressão linear simples por meio do método de mínimos quadrados.

Caso considerem-se modelos com mais de uma covariável (digamos que X seja uma matriz com p covariáveis), a matriz $X^\top X$ não terá inversa em casos onde o seu determinante seja igual a zero, como por exemplo:

1. Caso X tenha mais covariáveis do que observações ($n < p$);
2. Caso X tenha colunas que sejam combinações lineares de outras colunas (i.e. colunas linearmente dependentes);
3. Colunas ou linhas em que todas as entradas sejam iguais a zero;
4. Dentre outras.

4.2.3 c)

Para o caso de regressão linear simples, caso todos os valores de x sejam iguais (diga $x_i = k\ \forall i$), então temos que o modelo pode ser escrito como $Y_i = \beta_0 + \beta_1 k + \epsilon_i$. Como o ponto $(\bar{x}, \bar{Y})$ pertence à reta de regressão, existirão infinitas retas que passam por esse ponto (já que $x_i = \bar{x},\ \forall i$ e $\bar{Y}$ não depende de x) e nenhuma delas pode ser dita como a que minimiza a soma de quadrados.

Dessa forma, se todos os valores de um parâmetro forem iguais não há como explicar como o aumento/diminuição dele afeta a resposta, visto que não há variação no valor dos parâmetros.