Econometria I
Álgebra dos Mínimos Quadrados

Victor Oliveira
Núcleo de Economia Internacional e Desenvolvimento Econômico

PPGDE – 2023

Sumário

Amostra Aleatória

- O melhor preditor de y dado $\boldsymbol{x}$ para um par de variáveis aleatórias $(y, \boldsymbol{x}) \in \mathbb{R} \times \mathbb{R}^k$ é o modelo de projeção linear.
- Agora estamos interessados em **estimar** os parâmetros desse modelo, em particular o **coeficiente de projeção**

$$\beta = \left[\mathbb{E}\left(\boldsymbol{x}\boldsymbol{x}'\right)\right]^{-1} \mathbb{E}(\boldsymbol{x}y) \tag{1}$$

- Podemos estimar β usando amostras que incluem medidas conjunta de $(y, \boldsymbol{x})$.
- Vamos diferenciar *observações (realizações)* das variáveis aleatórias.
- Variáveis aleatórias são $(y, \boldsymbol{x})$, enquanto as observações são $(y_i, \boldsymbol{x}_i)$.
- Do ponto de vista do pesquisador, os últimos são números. Do ponto de vista da teoria estatística elas são realizações de variáveis aleatórias.

- **Notação:** Para observações individuais, vamos usar um subscrito i que vai de 1 a n, portanto, a i-*ésima* observação é $(y_i, \boldsymbol{x}_i)$. O número n é o tamanho da amostra.

- **Definição:** Uma amostra aleatória é dada pelas observações

$$\{(y_1, x_1), (y_2, x_2), \ldots, (y_i, x_i), \ldots, (y_n, x_n)\} \tag{2}$$

- De forma mais simplificada temos $\{(y_i, x_i) : i = 1, ..., n\}$.
- As observações individuais podem ser extraídas de uma distribuição comum (homogênea) ou podem ser extraídas de distribuições heterogêneas. A abordagem mais simples é assumir a homogeneidade – que as observações são realizações de uma população subjacente idêntica.

Pressuposto

As variáveis $\{(\boldsymbol{y}_1, \boldsymbol{x}_1), (\boldsymbol{y}_2, \boldsymbol{x}_2), ..., (\boldsymbol{y}_i, \boldsymbol{x}_i), ..., (\boldsymbol{y}_n, \boldsymbol{x}_n)\}$ *são identicamente distribuídas e são retirados de uma distribuição comum* F.

- Vamos nos referir à distribuição comum F como sendo a **população**.
- O **modelo de projeção linear** se aplica as variáveis aleatórias $(y, \boldsymbol{x})$. É o modelo de probabilidade que descrevemos anteriormente como o melhor preditor linear.

Critério de Otimização

- O análogo empírico da esperança do erro ao quadrado é o **erro quadrático médio amostral**:

$$
\begin{aligned}
S_n(\beta) &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 \\
&= \frac{1}{n} SQE_n(\boldsymbol{\beta}) \qquad (3)
\end{aligned}
$$

em que $SQE_n(\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})$ é chamada de **função da soma dos quadrados dos erros**. Um **estimador** para β é obtido pela minimização da função $S_n(\beta)$:

$$
\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^k}{\arg\min} S_n(\boldsymbol{\beta}) \qquad (4)
$$

Estimador de MQO

- O $\widehat{\boldsymbol{\beta}}$ é conhecido como o estimador de **Mínimos Quadrados Ordinários** (MQO) de $\boldsymbol{\beta}$.
- **Distinção**: o parâmetro populacional $\boldsymbol{\beta}$ é **fixo na população**, enquanto o estimador $\widehat{\boldsymbol{\beta}}$ **varia entre as amostras**.

Caso Univariado

- Considere o caso de $k = 1$ tal que o coeficiente de $\boldsymbol{\beta}$ é um escalar.
- Assim, a SQE é dada por:

$$SQE_n(\beta) = \sum_{i=1}^{n} (y_i - x_i\beta)^2$$
$$= \left(\sum_{i=1}^{n} y_i^2\right) - 2\beta \left(\sum_{i=1}^{n} y_i x_i\right) + \beta^2 \left(\sum_{i=1}^{n} x_i^2\right) \qquad (5)$$

- O estimador de MQO minimiza esta função.

- Assim, o resultado da minimização do $SQE_n(\beta)$ é:

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2} \tag{6}$$

- O modelo com apenas o intercepto é um caso especial de $x_i = 1$. Neste caso encontramos:

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} 1} = \frac{1}{n}\sum_{i=1}^{n} y_i = \overline{y} \tag{7}$$

que é a média de y_i.

- Por esse resultado, **o estimador de MQO num modelo com apenas o intercepto é a média amostral**.
- O estimador $\widehat{\beta}$ existe apenas se o **denominador é não-zero**. Por ser uma soma dos quadrados, necessariamente é não-negativo. Portanto, $\widehat{\beta}$ existe se $\sum_{i=1}^{n} x_i^2 > 0$.

Caso Multivariado

- Considere o caso de $k > 1$ tal que o coeficiente de $\boldsymbol{\beta}$ agora é um vetor.
- Assim, a SQE é dada por:

$$SQE_n(\boldsymbol{\beta}) = \left(\sum_{i=1}^{n} y_i^2\right) - 2\boldsymbol{\beta}'\left(\sum_{i=1}^{n} y_i \boldsymbol{x}_i\right) + \boldsymbol{\beta}'\left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'\right)\boldsymbol{\beta} \quad (8)$$

- A CPO para minimização de $SQE_n(\boldsymbol{\beta})$ é:

$$\frac{\partial}{\partial \boldsymbol{\beta}} SQE_n(\widehat{\boldsymbol{\beta}}) = \mathbf{0} \quad \iff \quad = -2\sum_{i=1}^{n} \boldsymbol{x}_i y_i + 2\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \widehat{\boldsymbol{\beta}} \quad (9)$$

- Por essa expressão temos um sistema de k equações com k elementos desconhecidos $\widehat{\boldsymbol{\beta}}$.

Condições Necessárias e Suficientes

- Resolvendo esse sistema encontramos uma fórmula explícita para o estimador de MQO:

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{x}_i y_i\right) \tag{10}$$

- Este é um estimador natural do coeficiente de $\boldsymbol{\beta}$ da melhor projeção linear e também é conhecido como **estimador da projeção linear**.
- Obtendo a segunda derivada da eq. (9), temos:

$$\frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} SQE_n \boldsymbol{\beta} = 2 \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' > 0 \tag{11}$$

que é uma matriz definida positiva. Logo, $\widehat{\boldsymbol{\beta}}$ é o único que minimiza a função SQE_n.

Estimador dos Momentos

- Podemos escrever o coeficiente de projeção $\boldsymbol{\beta}$ como uma função explícita dos momentos populacionais $\boldsymbol{Q_{xy}}$ e $\boldsymbol{Q_{xx}}$. O estimador dos momentos são os momentos amostrais:

$$\widehat{\boldsymbol{Q}}_{\boldsymbol{xy}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i y_i \tag{12}$$

$$\widehat{\boldsymbol{Q}}_{\boldsymbol{xx}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \tag{13}$$

- O estimador dos momentos de $\boldsymbol{\beta}$ substitui o momento populacional pelos momentos amostrais:

$$\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{Q}}_{\boldsymbol{xx}}^{-1}\widehat{\boldsymbol{Q}}_{\boldsymbol{xy}} = \left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{x}_i y_i\right) \tag{14}$$

Definição

O estimador de mínimos quadrados $\widehat{\boldsymbol{\beta}}$ é

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^k}{\arg\min} S_n(\boldsymbol{\beta}) \tag{15}$$

em que

$$S_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right)^2 \tag{16}$$

E tem como solução

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}\left(\sum_{i=1}^{n}\boldsymbol{x}_i y_i\right) \tag{17}$$

Teorema

Se $\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' > 0$, *o estimador de Mínimos Quadrados é único e igual a*

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{x}_i y_i \right) \tag{18}$$

Resíduos

- Como um subproduto da estimação, definimos o valor ajustado

$$\widehat{y}_i = \boldsymbol{x}_i' \widehat{\boldsymbol{\beta}} \tag{19}$$

e os resíduos

$$\widehat{e}_i = y_i - \widehat{y}_i = y_i - \boldsymbol{x}_i' \widehat{\boldsymbol{\beta}} \tag{20}$$

- Note que

$$y_i = \boldsymbol{x}_i' \widehat{\boldsymbol{\beta}} + \widehat{e}_i \tag{21}$$

- **Distinção:** O termo e_i é não observável enquanto o resíduo $\widehat{e}_i$ é um subproduto da estimação.

- A eq (20) implica que:

$$\sum_{i=1}^{n} \boldsymbol{x}_i \widehat{e}_i = \mathbf{0} \tag{22}$$

- Podemos demonstrar isso como:

$$\begin{aligned}
\sum_{i=1}^{n} \boldsymbol{x}_i \widehat{e}_i &= \sum_{i=1}^{n} \boldsymbol{x}_i \left(y_i - \boldsymbol{x}_i' \widehat{\boldsymbol{\beta}} \right) \\
&= \sum_{i=1}^{n} \boldsymbol{x}_i y_i - \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \widehat{\boldsymbol{\beta}} \\
&= \sum_{i=1}^{n} \boldsymbol{x}_i y_i - \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{x}_i y_i \right) \\
&= \sum_{i=1}^{n} \boldsymbol{x}_i y_i - \sum_{i=1}^{n} \boldsymbol{x}_i y_i \\
&= \mathbf{0}
\end{aligned}$$

- Quando $\boldsymbol{x}_i$ contém uma constante, uma implicação é:

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{e}_i = 0 \tag{23}$$

- Os resíduos tem média amostral zero e a correlação amostral entre os regressores e o resíduo é zero.
- Estes são resultados algébricos e se mantem verdadeiros para todas estimativas da regressão linear.

Formato Matricial

- As n equações lineares $y = \boldsymbol{x}'\boldsymbol{\beta} + e$ formam um sistema de n equações. Podemos empilhar estas n equações juntas como:

$$
\begin{aligned}
y_1 &= \boldsymbol{x}_1'\boldsymbol{\beta} + e_1 \\
y_2 &= \boldsymbol{x}_2'\boldsymbol{\beta} + e_2 \\
&\vdots \\
y_n &= \boldsymbol{x}_n'\boldsymbol{\beta} + e_n
\end{aligned}
$$

- Agora defina $\boldsymbol{y}_{n\times 1} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, $\boldsymbol{X}_{n\times k} = \begin{pmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \vdots \\ \boldsymbol{x}_n' \end{pmatrix}$, $\boldsymbol{e}_{n\times 1} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$

- O sistema de n equações pode ser compactado em uma única equação

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + e \tag{24}$$

- Somas de amostras pode ser escritas em notação matricial. Por exemplo

$$\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' = \boldsymbol{X}'\boldsymbol{X} \tag{25}$$

$$\sum_{i=1}^{n} \boldsymbol{x}_i y_i = \boldsymbol{X}'\boldsymbol{y} \tag{26}$$

- E o estimador de mínimos quadrados pode ser escrito como:

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \left(\boldsymbol{X}'\boldsymbol{y}\right) \tag{27}$$

- As versões matriciais se tornam

$$\boldsymbol{y} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \widehat{e} \tag{28}$$

ou equivalente ao vetor de resíduos

$$\widehat{e} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} \tag{29}$$

- Usando o vetor de resíduos podemos escrever

$$\boldsymbol{X}'\widehat{e} = \mathbf{0} \tag{30}$$

Propriedades

- Vamos definir a matriz:

$$\boldsymbol{P} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X} \tag{31}$$

- Observe que

$$\boldsymbol{P}\boldsymbol{X} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{X} \tag{32}$$

- Esta é uma propriedade da **matriz de projeção**. Para qualquer matriz $\boldsymbol{Z}$ que pode ser escrita como $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{\Gamma}$ para qualquer $\boldsymbol{\Gamma}$, então

$$\boldsymbol{P}\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{\Gamma} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\Gamma} = \boldsymbol{X}\boldsymbol{\Gamma} = \boldsymbol{Z}$$

- A matriz $\boldsymbol{P}$ é **simétrica** e **idempotente**. Podemos ver isso

$$\begin{aligned}
\boldsymbol{P}' &= \left(\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right)' \\
&= \left(\boldsymbol{X}'\right)'\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)\right)^{-1}\boldsymbol{X}' \\
&= \boldsymbol{X}\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)'\right)^{-1}\boldsymbol{X}' \\
&= \boldsymbol{X}\left(\left(\boldsymbol{X}\right)'\left(\boldsymbol{X}'\right)'\right)^{-1}\boldsymbol{X}' \\
&= \boldsymbol{P}
\end{aligned} \tag{33}$$

- Para ver que a matriz é idempotente, o fato que $\boldsymbol{P}\boldsymbol{X} = \boldsymbol{X}$ implica que

$$\begin{aligned}
\boldsymbol{P}\boldsymbol{P} &= \boldsymbol{P}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' \\
&= \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' \\
&= \boldsymbol{P}
\end{aligned} \tag{34}$$

- O traço da matriz $\boldsymbol{P}$ é:

$$\begin{aligned} \operatorname{tr} \boldsymbol{P} &= \operatorname{tr}\left(\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right) \\ &= \operatorname{tr}\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\right) \\ &= \operatorname{tr}\left(\boldsymbol{I}_k\right) \\ &= k \end{aligned} \tag{35}$$

- A matriz $\boldsymbol{P}$ tem uma propriedade de gerar os valores ajustados na regressão de mínimos quadrados:

$$\boldsymbol{P}\boldsymbol{y} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{X}'\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{y}} \tag{36}$$

Teorema

A matriz de projeção $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}$ *para qualquer* $\boldsymbol{X}$ *de dimensão* $n \times k$ *com* $n \geq k$ *tem as seguintes propriedades algébricas:*

1. $\boldsymbol{P}$ *é simétrica* $(\boldsymbol{P}' = \boldsymbol{P})$.
2. $\boldsymbol{P}$ *é idempotente* $(\boldsymbol{P}\boldsymbol{P} = \boldsymbol{P})$.
3. $\operatorname{tr} \boldsymbol{P} = k$.
4. *os autovalores de* $\boldsymbol{P}$ *são 1 e zero. Existem* k *autovalores igualando a 1 e* $n - k$ *igualando a zero.*
5. $rank(\boldsymbol{P}) = k$

Propriedades

- Vamos definir

$$\begin{aligned} \boldsymbol{M} &= \boldsymbol{I}_n - \boldsymbol{P} \\ &= \boldsymbol{I}_n - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' \end{aligned} \tag{37}$$

em que $\boldsymbol{I}_n$ é uma matriz identidade $n \times n$.

- Note que

$$\begin{aligned} \boldsymbol{M}\boldsymbol{X} &= \left(\boldsymbol{I}_n - \boldsymbol{P}\right)\boldsymbol{X} \\ &= \boldsymbol{X} - \boldsymbol{P}\boldsymbol{X} = \boldsymbol{X} - \boldsymbol{X} = 0 \end{aligned} \tag{38}$$

- Então $\boldsymbol{M}$ e $\boldsymbol{X}$ são ortogonais. Chamamos $\boldsymbol{M}$ de **matriz de projeção ortogonal**.

- A matriz de projeção ortogonal $\boldsymbol{M}$ é simétrica ($\boldsymbol{M}' = \boldsymbol{M}$) e idempotente ($\boldsymbol{M}\boldsymbol{M} = \boldsymbol{M}$). Podemos calcular que

$$\operatorname{tr} \boldsymbol{M} = n - k \tag{39}$$

- Enquanto $\boldsymbol{P}$ gera os valores ajustados, $\boldsymbol{M}$ gera os resíduos:

$$\boldsymbol{M}\boldsymbol{y} = \boldsymbol{y} - \boldsymbol{P}\boldsymbol{y} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{e}} \tag{40}$$

- Podemos usar (40) para escrever uma expressão alternativa para vetor de resíduos. Substituindo $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ para dentro $\widehat{\boldsymbol{e}} = \boldsymbol{M}\boldsymbol{y}$ e usando $\boldsymbol{M}\boldsymbol{X} = 0$ encontramos

$$\widehat{\boldsymbol{e}} = \boldsymbol{M}\boldsymbol{y} = \boldsymbol{M}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}) = \boldsymbol{M}\boldsymbol{e} \tag{41}$$

Projeção e Variância

- A variância do erro $\sigma^2 = \mathbb{E}\left[e_i^2\right]$ é um momento, assim um estimador natural é o estimador de momentos. Se e_i^2 fosse observado estimaríamos σ^2 por

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \tag{42}$$

- Contudo e_i não é observado. Por isso é comum considerar uma estimação em dois passos. No primeiro passo os $\widehat{e}_i$ são calculados. No segundo passo, substituímos $\widehat{e}_i$ na eq. (42) para obter um estimador factível.

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \widehat{e}_i^2 \tag{43}$$

- Considerando a notação matricial, podemos escrever eq. (42) e (43) como

$$\tilde{\sigma}^2 = n^{-1}\boldsymbol{e}'\boldsymbol{e} \tag{44}$$

e

$$\widehat{\sigma}^2 = n^{-1}\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}} \tag{45}$$

- Lembrando que a expressão $\widehat{\boldsymbol{e}} = \boldsymbol{My} = \boldsymbol{Me}$ e de (40) e (41) aplicado na eq (45) encontramos

$$\begin{aligned}
\widehat{\sigma}^2 &= n^{-1}\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}} \\
&= n^{-1}\boldsymbol{y}'\boldsymbol{MMy} \\
&= n^{-1}\boldsymbol{y}'\boldsymbol{My} \\
&= n^{-1}\boldsymbol{e}'\boldsymbol{Me} \qquad (46)
\end{aligned}$$

- Uma aplicação interessante é

$$\begin{aligned} \tilde{\sigma}^2 - \widehat{\sigma}^2 &= n^{-1}\boldsymbol{e}'\boldsymbol{e} - n^{-1}\boldsymbol{e}'\boldsymbol{M}\boldsymbol{e} \\ &= n^{-1}\boldsymbol{e}'\boldsymbol{P}\boldsymbol{e} \\ &\geq 0 \end{aligned} \tag{47}$$

- A desigualdade final mantém porque $\boldsymbol{P}$ é é positiva semi-definida e $\boldsymbol{e}'\boldsymbol{P}\boldsymbol{e}$ é uma forma quadrática.
- Este resultado mostra que o estimador de $\widehat{\sigma}^2$ é numericamente menor do que o estimador idealizado pela eq (42).

Decomposição

- Outra forma de escrever a eq (40) é

$$\boldsymbol{y} = \boldsymbol{P}\boldsymbol{y} + \boldsymbol{M}\boldsymbol{y} = \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}} \tag{48}$$

- Esta decomposição é **ortogonal**, isto é:

$$\widehat{\boldsymbol{y}}'\widehat{\boldsymbol{e}} = (\boldsymbol{P}\boldsymbol{y})'(\boldsymbol{M}\boldsymbol{y}) = \boldsymbol{y}'\boldsymbol{P}\boldsymbol{M}\boldsymbol{y} = 0 \tag{49}$$

- Segue que

$$\boldsymbol{y}'\boldsymbol{y} = \widehat{\boldsymbol{y}}'\widehat{\boldsymbol{y}} + 2\widehat{\boldsymbol{y}}'\widehat{\boldsymbol{e}} + \widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}} = \widehat{\boldsymbol{y}}'\widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}} \tag{50}$$

ou

$$\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} \widehat{y}_i^2 + \sum_{i=1}^{n} \widehat{e}_i^2 \tag{51}$$

- Vamos subtrair $\overline{y}$ de ambos os lados da eq (48) obtemos

$$\boldsymbol{y} - \mathbf{1}\overline{\boldsymbol{y}} = \widehat{\boldsymbol{y}} - \mathbf{1}\overline{\boldsymbol{y}} + \widehat{\boldsymbol{e}} \tag{52}$$

- Esta decomposição é também ortogonal quando $\boldsymbol{X}$ contém uma constante, como

$$(\widehat{\boldsymbol{y}} - \mathbf{1}\overline{\boldsymbol{y}})'\widehat{\boldsymbol{e}} = \widehat{\boldsymbol{y}}'\widehat{\boldsymbol{e}} - \overline{\boldsymbol{y}}\mathbf{1}'\widehat{\boldsymbol{e}} = 0 \tag{53}$$

sob $\frac{1}{n}\sum_{i=1}^{n}\widehat{e}_i^2 = 0$.

- Segue que

$$(\boldsymbol{y} - \mathbf{1}\overline{\boldsymbol{y}})'(\boldsymbol{y} - \mathbf{1}\overline{\boldsymbol{y}}) = (\widehat{\boldsymbol{y}} - \mathbf{1}\overline{\boldsymbol{y}})'(\widehat{\boldsymbol{y}} - \mathbf{1}\overline{\boldsymbol{y}}) + \widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}} \tag{54}$$

$$\sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2 + \sum_{i=1}^{n}\widehat{e}_i^2 \tag{55}$$

Variância e Ajuste do Modelo

- As eq. (54) e (55) são chamadas de fórmula da análise da variância para regressão de mínimos quadrados.
- Uma medida de ajuste da regressão é o coeficiente de determinação ou $\boldsymbol{R}^2$:

$$R^2 = \frac{\sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} = 1 - \frac{\sum_{i=1}^{n}\widehat{e}_i^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{56}$$

- O $\boldsymbol{R}^2$ é frequentemente descrito como a fração da variância amostral de y_i que é explicada pelo ajuste dos mínimos quadrados.
- **Atenção**: Uma restrição/crítica ao $\boldsymbol{R}^2$ é que ele aumenta quando incluímos mais regressores na regressão.

- Considere a partição $\boldsymbol{X} = [\boldsymbol{X}_1 \boldsymbol{X}_2]$ e $\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}$.
- Assim, o modelo de regressão pode ser reescrito como

$$\boldsymbol{y} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{e}} \tag{57}$$

- O estimador de MQO de $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ é obtido pela regressão de $\boldsymbol{y}$ em $\boldsymbol{X} = [\boldsymbol{X}_1 \boldsymbol{X}_2]$ e pode ser escrito como

$$\boldsymbol{y} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{e}} = \boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 + \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 + \widehat{\boldsymbol{e}} \tag{58}$$

- Estamos interessados na expressão algébrica para $\widehat{\boldsymbol{\beta}}_1$ e $\widehat{\boldsymbol{\beta}}_2$.

- A álgebra aqui é idêntica a que fizemos anteriormente para os coeficientes da população.
- Particionar $\widehat{\boldsymbol{Q}}_{\boldsymbol{xx}}$ e $\widehat{\boldsymbol{Q}}_{\boldsymbol{xy}}$ como

$$\widehat{\boldsymbol{Q}}_{\boldsymbol{xx}} = \begin{bmatrix} \widehat{\boldsymbol{Q}}_{11} & \widehat{\boldsymbol{Q}}_{12} \\ \widehat{\boldsymbol{Q}}_{21} & \widehat{\boldsymbol{Q}}_{22} \end{bmatrix} = \begin{bmatrix} \frac{1}{n}(\boldsymbol{X}_1'\boldsymbol{X}_1) & \frac{1}{n}(\boldsymbol{X}_1'\boldsymbol{X}_2) \\ \\ \frac{1}{n}(\boldsymbol{X}_2'\boldsymbol{X}_1) & \frac{1}{n}(\boldsymbol{X}_2'\boldsymbol{X}_2) \end{bmatrix} \tag{59}$$

e

$$\widehat{\boldsymbol{Q}}_{\boldsymbol{xy}} = \begin{bmatrix} \widehat{\boldsymbol{Q}}_{1y} \\ \widehat{\boldsymbol{Q}}_{2y} \end{bmatrix} = \begin{bmatrix} \frac{1}{n}(\boldsymbol{X}_1'\boldsymbol{y}) \\ \\ \frac{1}{n}(\boldsymbol{X}_2'\boldsymbol{y}) \end{bmatrix} \tag{60}$$

- Calculando a matriz inversa para $\widehat{\boldsymbol{Q}}_{\boldsymbol{xx}}$

$$\widehat{\boldsymbol{Q}}_{\boldsymbol{xx}}^{-1} = \begin{bmatrix} \widehat{\boldsymbol{Q}}_{11} & \widehat{\boldsymbol{Q}}_{12} \\ \widehat{\boldsymbol{Q}}_{21} & \widehat{\boldsymbol{Q}}_{22} \end{bmatrix}^{-1} \stackrel{def}{=} \begin{bmatrix} \widehat{\boldsymbol{Q}}^{11} & \widehat{\boldsymbol{Q}}^{12} \\ \widehat{\boldsymbol{Q}}^{21} & \widehat{\boldsymbol{Q}}^{22} \end{bmatrix}$$
$$= \begin{bmatrix} \widehat{\boldsymbol{Q}}_{11.2}^{-1} & -\widehat{\boldsymbol{Q}}_{11.2}^{-1}\widehat{\boldsymbol{Q}}_{12}\widehat{\boldsymbol{Q}}_{22}^{-1} \\ -\widehat{\boldsymbol{Q}}_{22.1}^{-1}\widehat{\boldsymbol{Q}}_{21}\widehat{\boldsymbol{Q}}_{11}^{-1} & \widehat{\boldsymbol{Q}}_{22.1}^{-1} \end{bmatrix} \tag{61}$$

em que $\widehat{\boldsymbol{Q}}_{11.2} = \widehat{\boldsymbol{Q}}_{11} - \widehat{\boldsymbol{Q}}_{12}\widehat{\boldsymbol{Q}}_{22}^{-1}\widehat{\boldsymbol{Q}}_{21}$ e $\widehat{\boldsymbol{Q}}_{22.1} = \widehat{\boldsymbol{Q}}_{22} - \widehat{\boldsymbol{Q}}_{21}\widehat{\boldsymbol{Q}}_{11}^{-1}\widehat{\boldsymbol{Q}}_{12}$.

- Assim,

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \widehat{\boldsymbol{\beta}}_1 \\ \widehat{\boldsymbol{\beta}}_2 \end{pmatrix}$$

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\boldsymbol{Q}}_{11.2}^{-1} & -\widehat{\boldsymbol{Q}}_{11.2}^{-1}\widehat{\boldsymbol{Q}}_{12}\widehat{\boldsymbol{Q}}_{22}^{-1} \\ -\widehat{\boldsymbol{Q}}_{22.1}^{-1}\widehat{\boldsymbol{Q}}_{21}\widehat{\boldsymbol{Q}}_{11}^{-1} & \widehat{\boldsymbol{Q}}_{22.1}^{-1} \end{bmatrix} \times \begin{bmatrix} \widehat{\boldsymbol{Q}}_{1y} \\ \widehat{\boldsymbol{Q}}_{2y} \end{bmatrix}$$

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \widehat{\boldsymbol{Q}}_{11.2}^{-1}\widehat{\boldsymbol{Q}}_{1y.2} \\ \widehat{\boldsymbol{Q}}_{22.1}^{-1}\widehat{\boldsymbol{Q}}_{2y.1} \end{pmatrix} \tag{62}$$

- Agora

$$\begin{aligned} \widehat{\boldsymbol{Q}}_{11.2} &= \widehat{\boldsymbol{Q}}_{11} - \widehat{\boldsymbol{Q}}_{12}\widehat{\boldsymbol{Q}}_{22}^{-1}\widehat{\boldsymbol{Q}}_{21} \\ &= \frac{1}{n}\boldsymbol{X}_1'\boldsymbol{X}_1 - \frac{1}{n}\boldsymbol{X}_1'\boldsymbol{X}_2\left(\frac{1}{n}\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\frac{1}{n}\left(\boldsymbol{X}_2'\boldsymbol{X}_1\right) \\ &= \frac{1}{n}\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1 \end{aligned} \tag{63}$$

- Em que

$$\boldsymbol{M}_2 = \boldsymbol{I}_n - \left(\boldsymbol{X}_2\boldsymbol{X}_2'\boldsymbol{X}\right)^{-1}\boldsymbol{X}_2 \tag{64}$$

é a matriz de projeção ortogonal para $\boldsymbol{X}_2$.

- De forma similar $\widehat{\boldsymbol{Q}}_{22.1} = \frac{1}{n}\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2$, em que

$$\boldsymbol{M}_1 = \boldsymbol{I}_n - \left(\boldsymbol{X}_1\boldsymbol{X}_1'\boldsymbol{X}\right)^{-1}\boldsymbol{X}_1 \tag{65}$$

é a matriz de projeção ortogonal para $\boldsymbol{X}_1$.

- Ainda

$$\begin{aligned}\widehat{\boldsymbol{Q}}_{1y.2} &= \widehat{\boldsymbol{Q}}_{1y} - \widehat{\boldsymbol{Q}}_{12}\widehat{\boldsymbol{Q}}_{22}^{-1}\widehat{\boldsymbol{Q}}_{2y} \\ &= \frac{1}{n}\boldsymbol{X}_1'\boldsymbol{y} - \frac{1}{n}\boldsymbol{X}_1'\boldsymbol{X}_2\left(\frac{1}{n}\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\frac{1}{n}\boldsymbol{X}_2'\boldsymbol{y} \\ &= \frac{1}{n}\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{y} \end{aligned} \tag{66}$$

- E $\widehat{\boldsymbol{Q}}_{2y.1} = \frac{1}{n}\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{y}$.
- Portanto,

$$\widehat{\boldsymbol{\beta}}_1 = \left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)^{-1}\left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{y}\right) \tag{67}$$

e

$$\widehat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{y}) \tag{68}$$

- Estes são as expressões algébricas para as estimativas da equação particionada.

- Como mostrou Frisch e Waugh (1933), as expressões dadas pelas eqs. (67) e (68) podem ser utilizadas para mostrar que o estimadores de Mínimos Quadrados $\widehat{\boldsymbol{\beta}}_1$ e $\widehat{\boldsymbol{\beta}}_2$ podem ser obtidos por um procedimento de regressão em dois passos.
- Considere a eq. (68). Como $\boldsymbol{M}_1$ é idempotente, $\boldsymbol{M}_1 = \boldsymbol{M}_1\boldsymbol{M}_1$ e assim

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_2 &= \left(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2\right)^{-1}\left(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{y}\right)\\ &= \left(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{M}_1\boldsymbol{X}_2\right)^{-1}\left(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{M}_1\boldsymbol{y}\right)\\ &= \left(\tilde{\boldsymbol{X}}_2'\tilde{\boldsymbol{X}}_2\right)^{-1}\left(\tilde{\boldsymbol{X}}_2'\tilde{\boldsymbol{e}}_1\right)\end{aligned} \tag{69}$$

em que $\tilde{\boldsymbol{X}}_2 = \boldsymbol{M}_1\boldsymbol{X}_2$ e $\tilde{\boldsymbol{e}}_1 = \boldsymbol{M}_1\boldsymbol{y}$.

- Vamos começar pelo configuração mais simples: modelo somente com intercepto.

$$y_i = \mu + e_i \tag{70}$$
$$\mathbb{E}[e] = 0 \tag{71}$$

- Esse modelo é equivalente a um modelo de regressão com $k = 1$ e $x_i = 1$. No modelo somente com intercepto, $\mu = \overline{y}$ é a média de y_i.
- O estimador de mínimos quadrados $\widehat{\mu} = \overline{y}$ iguala a média amostral.
- Vamos calcular a esperança e a variância do estimador $\overline{y}$. Como a a média amostral é uma função das observações, sua esperança é simples de calcular:

$$\mathbb{E}[\overline{y}] = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) \tag{72}$$

- Assim,

$$\mathbb{E}[\overline{y}] = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\overline{y}_i] = \mu \tag{73}$$

- Temos que o valor esperado do estimador de mínimos quadrados (média amostral) iguala o coeficiente de projeção (média populacional).
- Um estimador com a propriedade de que sua esperança é igual ao parâmetro que ele está estimando é conhecido como **estimador não viesado**.

Definição

Um estimador $\widehat{\theta}$ para θ é **não viesado** se $\mathbb{E}[\widehat{\theta}] = \theta$.

Variância amostral

- Vamos calcular a variância do estimador $\overline{y}$ considerando a definição acima. Usando $y_i = \mu + e_i$, temos:

$$\overline{y} - \mu = \frac{1}{n}\sum_{i=1}^{n} e_i \tag{74}$$

- Assim,

$$\begin{aligned} \text{var}[\overline{y}] &= \mathbb{E}\left(\overline{y} - \mu\right)^2 \\ &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} e_i\right)\left(\frac{1}{n}\sum_{j=1}^{n} e_j\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}[e_i e_j] \end{aligned} \tag{75}$$

- Logo,

$$\begin{aligned} \mathrm{var}[\overline{y}] &= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 \\ &= \frac{1}{n}\sigma^2 \end{aligned} \tag{76}$$

- **Nota:** $\mathbb{E}[e_i e_j] = \sigma^2$ para $i = j$, $\mathbb{E}[e_i e_j] = 0$ para $i \neq j$ em virtude da independência.

Econometria I
Álgebra dos Mínimos Quadrados

Victor Oliveira
Núcleo de Economia Internacional e Desenvolvimento Econômico

PPGDE – 2023