

ECONOMETRIA I

MODELO DA FUNÇÃO DE ESPERANÇA CONDICIONAL LINEAR

Victor Oliveira

Núcleo de Economia Internacional e Desenvolvimento Econômico

PPGDE – 2023

- 1 CEF
- 2 Variância da Regressão
- 3 Melhor Preditor
- 4 Variância Condicional
- 5 Homocedasticidade e Heterocedasticidade
- 6 Função de Esperança Condicional Linear
- 7 Modelo CEF Linear
- 8 Melhor Preditor Linear e Erro de Projeção
- 9 Propriedades do Erro de Projeção
- 10 Propriedades do Modelo de Projeção Linear

Definição

- A função esperança condicional (CEF) do erro ε é definida como a diferença entre y avaliada no vetor aleatório \mathbf{x} ,

$$\varepsilon = y - m(\mathbf{x}) \quad (1)$$

- Por construção, ela vem da fórmula:

$$y = m(\mathbf{x}) + \varepsilon \quad (2)$$

- É útil entender que o termo de erro ε é derivado da distribuição conjunta de (y, \mathbf{x}) e, assim, suas propriedades também são derivadas dessa construção.
- A principal propriedade da CEF do erro é de que ele possui média condicional igual a zero.

Propriedades

Demonstração.

Para ver isso, vamos considerar a linearidade da esperança condicional, a definição $m(\mathbf{x}) = \mathbb{E}[y|\mathbf{x}]$ e o teorema do condicionamento

$$\begin{aligned}\mathbb{E}[\varepsilon|\mathbf{x}] &= \mathbb{E}[(y - m(\mathbf{x}))|\mathbf{x}] \\ &= \mathbb{E}[y|\mathbf{x}] - \mathbb{E}[m(\mathbf{x})|\mathbf{x}] \\ &= m(\mathbf{x}) - m(\mathbf{x}) = 0\end{aligned}$$

- Este fato pode ser combinado com a lei das expectativas iteradas para mostrar que a média incondicional também é zero:

$$\mathbb{E}[\varepsilon] = \mathbb{E}[\mathbb{E}[\varepsilon|\mathbf{x}]] = \mathbb{E}[0] = 0 \quad (3)$$

Teorema: Propriedades da CEF

Teorema

- Se $\mathbb{E}|y| < \infty$ então:
 - ① $\mathbb{E}[\varepsilon|\mathbf{x}] = 0$
 - ② $\mathbb{E}[\varepsilon] = 0$
 - ③ Se $\mathbb{E}|y|^r < \infty$ para $r \geq 1$, então $\mathbb{E}|\varepsilon|^r < \infty$
 - ④ para qualquer função $h(\mathbf{x})$ tal que $\mathbb{E}|h(\mathbf{x})\varepsilon| < \infty$, então $\mathbb{E}[h(\mathbf{x})\varepsilon] = 0$
- A propriedade 4 diz que “ ε ” é não correlacionado com qualquer função dos regressores. *PROVE!*

Demonstração.

Pela desigualdade de Minkowski, temos que:

$$\begin{aligned} [\mathbb{E}|\varepsilon|^r]^{1/r} &= [\mathbb{E}|y - m(\mathbf{x})|^r]^{1/r} \\ &\leq [\mathbb{E}|y|^r]^{1/r} + [\mathbb{E}|m(\mathbf{x})|^r]^{1/r} < \infty \end{aligned} \quad (4)$$

em que as duas partes do lado direito são finitas porque $[\mathbb{E}|y|^r] < \infty$ por hipótese e $[\mathbb{E}|m(\mathbf{x})|^r] < \infty$ pela desigualdade da expectativa condicional^a.

Assim, $[\mathbb{E}|\varepsilon|^r]^{1/r} \implies [\mathbb{E}|\varepsilon|^r] < \infty$. ■

^aPara qualquer $r \geq 1$ e para quaisquer variáveis aleatórias $(Y, X) \in \mathbb{R} \times \mathbb{R}^k$ tal que $\mathbb{E}|y|^r < \infty$, então $\mathbb{E}|\mathbb{E}(y|x)|^r \leq \mathbb{E}|y|^r < \infty$.

Média Condicional

- As equações

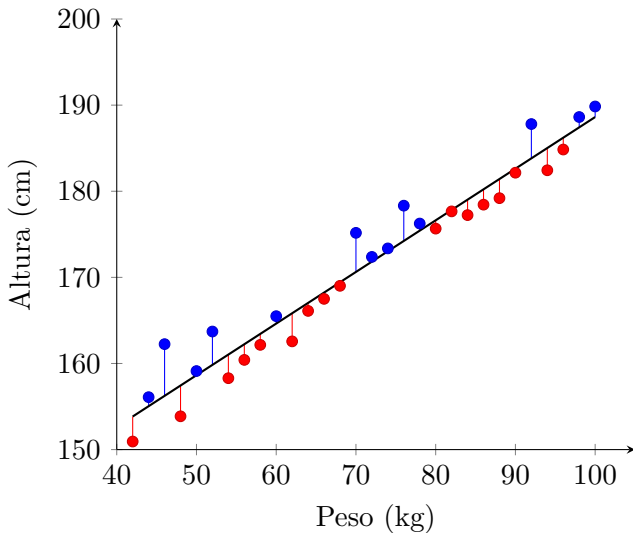
$$y = m(\mathbf{x}) + \varepsilon \quad (5)$$

$$\mathbb{E}[\varepsilon|\mathbf{x}] = 0 \quad (6)$$

juntas implicam que $m(\mathbf{x})$ é a CEF de y dado \mathbf{x} . Isso é importante para entender que isto não é uma restrição. Estas equações mantêm-se verdadeiras por definição.

- A condição $\mathbb{E}[\varepsilon|\mathbf{x}] = 0$ está implícito na definição de ε como a diferença entre y e a CEF $m(\mathbf{x})$.
- A equação $\mathbb{E}[\varepsilon|\mathbf{x}] = 0$ é as vezes chamada de restrição da média condicional, porque a média condicional do erro ε é igual a zero.

Figura 1: Simulação: Soma dos Erros é Zero



Variância da Regressão

- Uma importante medida de dispersão sobre a função CEF é a variância não condicional da CEF do erro ε . Ela pode ser escrita como:

$$\sigma^2 = \text{var}(\varepsilon) = \mathbb{E} \left[(\varepsilon - \mathbb{E}[\varepsilon])^2 \right] = \mathbb{E} \left[\varepsilon^2 \right] \quad (7)$$

Teorema

Se $\mathbb{E}[y^2] < \infty$, então $\sigma^2 < \infty$.

- Podemos chamar σ^2 a variância da regressão ou a variância do erro da regressão.
- A magnitude de σ^2 mede a quantidade da variação em y que **não** é “**explicada**” ou mensurada pela média condicional $\mathbb{E}[y|\mathbf{x}]$.

Variância Condicional

- A variância da regressão depende do \mathbf{x} . Para ver isso, considere duas regressões:

$$y = \mathbb{E}(y|\mathbf{x}_1) + \varepsilon_1 \quad (8)$$

$$y = \mathbb{E}(y|\mathbf{x}_1, \mathbf{x}_2) + \varepsilon_2 \quad (9)$$

em que ε_1 e ε_2 são diferentes. Mudar as informações do condicionante, muda a média condicional e portanto, o erro da regressão também.

- Pelas expectativas iteradas, *aumentar o conjunto de condicionante, a esperança condicional revela maior detalhes sobre a distribuição de y . Qual é a implicação para o termo de erro?*

- Há uma simples relação! Podemos pensar a média condicional $\mathbb{E}(y|\mathbf{x})$ como a “**parte explicada**” de y e $\varepsilon = y - \mathbb{E}(y|\mathbf{x})$ como a “**parte não explicada**”.
- A relação que derivamos mostra que a *variância desta parte não explicada decresce quando condicionamos em mais variáveis*. A relação é monotônica no sentido de que ao **aumentar a quantidade de informação a variância da parte não explicada irá reduzir**.

Teorema

Se $\mathbb{E}(y)^2 < \infty$, então $\text{var}(y) \geq \text{var}(y - \mathbb{E}(y|\mathbf{x}_1)) \geq \text{var}(y - \mathbb{E}(y|\mathbf{x}_1, \mathbf{x}_2))$.

- Esse teorema diz que a variância da diferença entre y e sua média condicional (fracamente) decresce quando uma variável adicional é incluída na informação condicionante.

Predição

- Suponha que dado um valor realizado de \mathbf{x} , queremos gerar uma previsão de y .
- Podemos escrever qualquer preditor como uma função $g(\mathbf{x})$ de \mathbf{x} . O erro de previsão é a diferença realizada $y - g(\mathbf{x})$.
- Uma medida não estocástica da magnitude do erro de previsão é a esperança do seu quadrado:

$$\mathbb{E}(y - g(\mathbf{x}))^2 \quad (10)$$

- Podemos definir o melhor preditor como a função $g(\mathbf{x})$ que minimiza $\mathbb{E}(y - g(\mathbf{x}))^2$.
- Qual função é o melhor preditor? Acontece que a resposta é CEF $m(\mathbf{x})$. Isso é válido independentemente da distribuição conjunta de (y, \mathbf{x}) .

Demonstração.

Para ver isso, note que o erro quadrado médio de um preditor $g(\mathbf{x})$ é dado por:

$$\begin{aligned}\mathbb{E}[y - g(\mathbf{x})]^2 &= \mathbb{E}[\varepsilon + m(\mathbf{x}) - g(\mathbf{x})]^2 \\ &= \mathbb{E}[\varepsilon]^2 + 2\mathbb{E}[\varepsilon(m(\mathbf{x}) - g(\mathbf{x}))] + \mathbb{E}[m(\mathbf{x}) - g(\mathbf{x})]^2 \\ &= \mathbb{E}[\varepsilon]^2 + \mathbb{E}[m(\mathbf{x}) - g(\mathbf{x})]^2 \\ &\geq \mathbb{E}[\varepsilon]^2 \\ &= \mathbb{E}[y - m(\mathbf{x})]^2\end{aligned}\tag{11}$$

em que $y = m(\mathbf{x}) + \varepsilon$. ■

Teorema (Média condicional como melhor preditor)

Se $\mathbb{E}(y)^2 < \infty$, então para qualquer preditor $g(\mathbf{x})$,

$$\mathbb{E}[y - g(\mathbf{x})]^2 \geq \mathbb{E}[y - m(\mathbf{x})]^2 \quad (12)$$

quando $m(\mathbf{x}) = \mathbb{E}[y|\mathbf{x}]$.

- Embora a média condicional seja uma boa medida da localização de uma distribuição condicional, ela não fornece informações sobre a extensão da distribuição.
- Uma medida comum da dispersão é a **variância condicional**.

Definição

Se $\mathbb{E}(y)^2 < \infty$, a **variância condicional** de y dada por \mathbf{x} é:

$$\begin{aligned}\sigma^2 &= \text{var}(y|\mathbf{x}) \\ &= \mathbb{E} \left[(y - \mathbb{E}(y|\mathbf{x}))^2 \mid \mathbf{x} \right] \\ &= \mathbb{E}(\varepsilon^2|\mathbf{x})\end{aligned}\tag{13}$$

- O desvio-padrão condicional é $\sigma(\mathbf{x}) = \sqrt{\sigma^2(\mathbf{x})}$.

- A variância não condicional e a variância condicional são relacionadas pela lei das expectativas iteradas.

$$\begin{aligned}\sigma^2 &= \mathbb{E}(\varepsilon^2) = \mathbb{E}(\mathbb{E}(\varepsilon^2|\mathbf{x})) \\ &= \mathbb{E}(\sigma^2(\mathbf{x}))\end{aligned}\tag{14}$$

isto é, a **variância não condicional** é a **média da variância condicional**.

- Um resultado interessante que normalmente usamos em análise de regressão: $\mathbb{E}(u) = 0$ e $\sigma^2 = 1$.
- Dado a variância condicional podemos definir o erro da seguinte forma.

$$u = \frac{\varepsilon}{\sigma(\mathbf{x})} \quad (15)$$

- Podemos fazer essa conta, uma vez que $\sigma(\mathbf{x})$ é uma função de \mathbf{x} :

$$\begin{aligned} \mathbb{E}(u|\mathbf{x}) &= \mathbb{E}\left[\frac{\varepsilon}{\sigma(\mathbf{x})} \middle| \mathbf{x}\right] \\ &= \frac{1}{\sigma(\mathbf{x})} \mathbb{E}(\varepsilon|\mathbf{x}) \\ &= 0 \end{aligned} \quad (16)$$

- Para a variância temos:

$$\begin{aligned}\text{var}(u|\mathbf{x}) &= \mathbb{E}\left(u^2|\mathbf{x}\right) \\ &= \mathbb{E}\left[\frac{\varepsilon^2}{\sigma^2(\mathbf{x})}\middle|\mathbf{x}\right] \\ &= \frac{1}{\sigma^2(\mathbf{x})}\mathbb{E}\left(\varepsilon^2|\mathbf{x}\right) \\ &= \frac{\sigma^2(\mathbf{x})}{\sigma^2(\mathbf{x})} \\ &= 1\end{aligned}\tag{17}$$

- Então u possui uma média condicional zero e uma variância condicional igual a 1.

Homocedasticidade e Heterocedasticidade

- Um caso especial importante que obtemos é quando a variância condicional $\sigma^2(\mathbf{x})$ é uma constante e independente de \mathbf{x} .

Definição

O erro é **homocedástico** se $\mathbb{E}(\varepsilon^2|\mathbf{x}) = \sigma^2$. Isto é, não depender de \mathbf{x} .

- No caso geral onde $\sigma^2(\mathbf{x})$ depende de \mathbf{x} , dizemos que o erro é **heterocedástico**.

Definição

O erro é **heterocedástico** se $\mathbb{E}(\varepsilon^2|\mathbf{x}) = \sigma^2(\mathbf{x})$. Isto é, depende de \mathbf{x} .

- Por definição, a variância não condicional σ^2 é uma constante e independente dos regressores \mathbf{x} . Assim, quando pensamos na variância como função dos regressores, estamos falando na variância condicional $\sigma^2(\mathbf{x})$.

Observação

Definir heterocedasticidade como o caso onde “a variância de e varia entre as observações” é uma forma pobre e confusa de definição.

- É mais elegante/construtivo entender que heterocedasticidade significa que a variância condicional $\sigma^2(\mathbf{x})$ depende das observáveis.

Observação

Visão de que homocedasticidade é o componente e uma especificação correta da regressão e descrever heterocedasticidade como uma exceção ou desvio.

- A visão correta é que heterocedasticidade é genérico e “padrão”, enquanto homocedasticidade é incomum e excepcional. O padrão no trabalho empírico deve ser assumir que os erros são heteroscedásticos, e não o contrário.

Função de Esperança Condicional Linear

- Um caso especial é quando a CEF é linear em \mathbf{x} , isto é, $m(\mathbf{x}) = \mathbb{E}(y|\mathbf{x})$.
- Podemos escrever a equação média como:

$$m(\mathbf{x}) = x_1\beta_1 + x_2\beta_2 + \cdots + x_k\beta_k + \beta_{k+1} \quad (18)$$

- Para facilitar a notação, podemos escrever a eq. (18) como uma função do vetor \mathbf{x} .
- Para isso, temos de aumentar o vetor \mathbf{x} listando o número **1** como um elemento. Vamos chamar de constante e o coeficiente associado será chamado de intercepto.

- Vamos assumir que o último elemento do vetor \mathbf{x} é o intercepto, então $x_k = 1$. Assim, redefinimos o vetor \mathbf{x} como vetor $k \times 1$

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{k-1} \\ 1 \end{pmatrix} \quad (19)$$

- Com essa redefinição, a CEF passa a ser:

$$m(\mathbf{x}) = x_1\beta_1 + x_2\beta_2 + \cdots + x_k\beta_k = \mathbf{x}'\boldsymbol{\beta} \quad (20)$$

- Temos que

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} \quad (21)$$

é um vetor de coeficiente $k \times 1$.

- Este é o **modelo da função de esperança condicional linear**.
- É também chamado de **modelo de regressão linear** ou regressão de y e \mathbf{x} .

- No modelo CEF linear, a derivada da regressão é simplesmente o vetor de coeficientes, isto é:

$$\nabla m(\mathbf{x}) = \beta \quad (22)$$

- Uma das vantagens desse modelo é poder interpretar o coeficiente como efeito marginal de uma variável, mantendo as demais constantes.

Modelo CEF Linear

- Seja

$$y = \mathbf{x}'\beta + \varepsilon \quad (23)$$

$$\mathbb{E}(\varepsilon|\mathbf{x}) = 0 \quad (24)$$

- Se o erro for homocedástico, temos um modelo CEF homocedástico linear, isto é,

$$y = \mathbf{x}'\beta + \varepsilon \quad (25)$$

$$\mathbb{E}(\varepsilon|\mathbf{x}) = 0 \quad (26)$$

$$\mathbb{E}(\varepsilon^2|\mathbf{x}) = \sigma^2 \quad (27)$$

- O Teorema do Melhor Preditor mostrou que a média condicional $m(\mathbf{x})$ é o melhor preditor no sentido de que ele possui o menor erro ao quadrado entre todos os preditores.
- Podemos definir uma aproximação para a CEF pela função linear com o menor erro quadrado entre todos preditores lineares.
- Para fazermos essa derivação vamos precisar das condições de regularidade:

CR.1 $\mathbb{E}(y^2) < \infty$

CR.2 $\|\mathbf{x}\| < \infty$

CR.3 $\mathbf{Q}_{xx} = \mathbb{E}(\mathbf{x}\mathbf{x}')$, em que \mathbf{Q}_{xx} é positiva definida

- Na condição de regularidade 2, usamos a notação $\|\mathbf{x}\|$ para denotar o comprimento euclidiano do vetor \mathbf{x} .
- As condições de regularidade 1 e 2 implicam que as variáveis y e \mathbf{x} tem média, variância e covariância finitas.
- A condição de regularidade 3 equivale impor que as colunas da $\mathbf{Q}_{xx} = \mathbb{E}(\mathbf{x}\mathbf{x}')$ são linearmente independentes. Isto é, a matriz \mathbf{Q}_{xx} tem inversa.
- Um preditor linear para y é uma função da forma $\mathbf{x}'\boldsymbol{\beta}$ para algum $\boldsymbol{\beta} \in \mathbb{R}^k$

- O melhor preditor do erro quadrático médio é:

$$S(\beta) = \mathbb{E}[y - \mathbf{x}'\beta]^2 \quad (28)$$

- O melhor preditor linear de y dado \mathbf{x} , escrito como $\mathcal{P}(y|\mathbf{x})$, é obtido selecionando o vetor β que minimiza $S(\beta)$.

- A partir dessas informações podemos ter a seguinte definição.

Definição

O melhor preditor linear de y dado \mathbf{x} é

$$\mathcal{P}(y|\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta} \quad (29)$$

em que $\boldsymbol{\beta}$ minimiza o erro quadrático médio predito.

- O minimizador $\boldsymbol{\beta} = \arg \min_{\mathbf{b} \in \mathbb{R}^k} S(\mathbf{b})$ é chamado de **coeficiente de projeção linear**.

- O erro quadrático médio de predição pode ser escrito como uma função quadrática de β :

$$S(\beta) = \mathbb{E}y^2 - 2\beta'\mathbb{E}(xy) + \beta'\mathbb{E}(xx')\beta \quad (30)$$

- Derivando com relação a β , encontramos a CPO:

$$\frac{\partial}{\partial \beta} S(\beta) = 0 \quad \Longleftrightarrow \quad -2\mathbb{E}(xy) + 2\mathbb{E}(xx')\beta = 0 \quad (31)$$

- Podemos reescrevendo a (31) como:

$$2\mathbb{E}(xy) = 2\mathbb{E}(xx')\beta \quad (32)$$

- Assim,

$$Q_{xy} = Q_{xx}\beta \quad (33)$$

em que $Q_{xy} = \mathbb{E}(xy)$ é um vetor $k \times 1$ e $Q_{xx} = \mathbb{E}(xx')$ é um vetor $k \times k$.

- A solução é obtida invertendo a matriz Q_{xx} e fazendo:

$$\begin{aligned} \beta &= Q_{xx}^{-1} Q_{xy} \\ \beta &= [\mathbb{E}(xx')]^{-1} \mathbb{E}(xy) \end{aligned} \quad (34)$$

Observação

A condição de regularidade 3 é fundamental para que eq. (34) possa existir!

Melhor Preditor Linear e Erro de Projeção

- Temos uma solução explícita para o melhor preditor linear:

$$\mathcal{P}(y|\mathbf{x}) = \mathbf{x}' [\mathbb{E}(\mathbf{x}\mathbf{x}')]^{-1} \mathbb{E}(\mathbf{x}y) \quad (35)$$

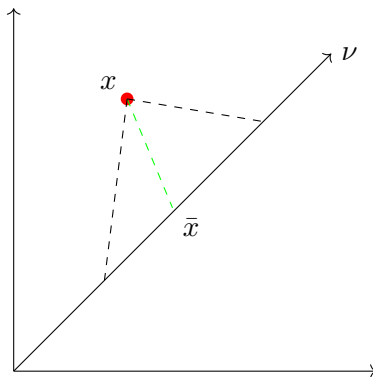
- Essa expressão podemos chamar de **projeção linear** de y em \mathbf{x} .
- O **erro de projeção** é dado por

$$\varepsilon = y - \mathbf{x}'\beta \quad (36)$$

Projeção

- Formalmente, uma projeção \mathcal{P} é uma função linear em um espaço vetorial, de modo que, quando aplicada a si mesma, obtém-se o mesmo resultado, ou seja, $\mathcal{P}^2 = \mathcal{P}$.
- Considere um vetor ν em duas dimensões. ν é uma linha reta finita apontando em uma determinada direção.
- Suponha que haja algum ponto x que não está nessa linha reta, mas no mesmo espaço bidimensional.
- A projeção de x , isto é, $\mathcal{P}x$, é uma função que retorna o ponto “mais próximo” de x ao longo da linha vetorial ν .
- Esse ponto será denotado por \bar{x} .
- Na maioria dos contextos, o mais próximo refere-se à distância euclidiana, ou seja, $\sqrt{\sum_i (x_i - \bar{x}_i)^2}$.

Figura 2: Projeção Ortogonal



- Resumindo, a projeção é uma forma de simplificar algum espaço n -dimensional comprimindo informações em um (hiper)plano.
- Isso é útil especialmente em ambientes de ciências sociais onde a complexidade dos fenômenos que estudamos significa que a previsão exata é impossível.
- Em vez disso, muitas vezes queremos construir modelos que comprimam dados e variáveis em explicações mais simples e parcimoniosas.
- A projeção é o método estatístico para conseguir isso – ela pega todo o espaço e o simplifica em relação a um certo número de dimensões.
- Embora a Figura acima seja (razoavelmente) intuitiva, vale a pena explicar a matemática por trás da projeção, até porque ajuda a demonstrar a conexão entre projeção linear e regressão linear.

- Para começar, podemos pegar algum ponto no espaço n -dimensional, x , e a reta vetorial ν ao longo da qual queremos projetar x . O objetivo é o seguinte:

$$\begin{aligned}\arg \min_c (\|\bar{x} - x\|) &= \arg \min_c \sqrt{\sum_i (\bar{x}_i - x_i)^2} \text{ [pela norma } L^2\text{]} \\ &= \arg \min_c \sum_i (\bar{x}_i - x_i)^2 \text{ } [\sqrt{\cdot} \text{ é monotônica}] \\ &= \arg \min_c \sum_i (c\nu_i - x_i)^2 \text{ } [\bar{x} = c\nu]\end{aligned}$$

- Esta forma é quadrática em c , então atinge mínimo quando:

$$\begin{aligned}
 \frac{d}{dc} \sum_i (c\nu_i - x_i)^2 &= \sum_i 2\nu_i(c\nu_i - x_i) \\
 &= 2 \left(\sum_i c\nu_i^2 - \sum_i \nu_i x_i \right) \\
 &= 2(c\nu'\nu - \nu'x)
 \end{aligned} \tag{37}$$

no mínimo, $\frac{d}{dc} \sum 2(c\nu'\nu - \nu'x) = 0$.

- Logo,

$$\begin{aligned}
 c\nu'\nu &= \nu'x \\
 c &= (\nu'\nu)^{-1} \nu'x
 \end{aligned} \tag{38}$$

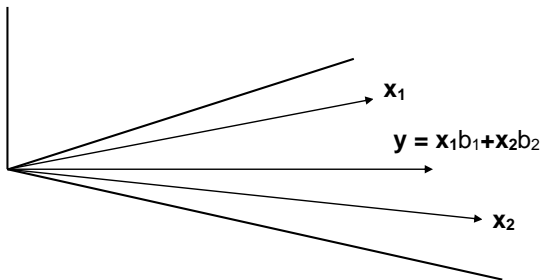
em que $\bar{x} = \nu c = \nu (\nu'\nu)^{-1} \nu'x = \mathcal{P}x$ e $\mathcal{P}_\nu = \nu (\nu'\nu)^{-1} \nu'$.

Intuição Gráfica para a Projeção Linear

- **Caso 1:** y está no espaço coluna de x .
- Isso significa que y pode ser expresso exatamente como uma combinação linear das colunas de x .
- Exemplo: considere que x_1 , x_2 e y são vetores com a terceira coordenada igual (isto é, estão no mesmo hiperplano).

- Graficamente, temos:

Figura 3: Caso 1, y está no espaço coluna de x



Intuição Gráfica para a Projeção Linear

- **Caso 2:** y não está no espaço coluna de x .
- Isso significa que y **não pode ser expresso como uma combinação linear das colunas de x** . Ou seja, não há um vetor b .
- Mas podemos escrever:

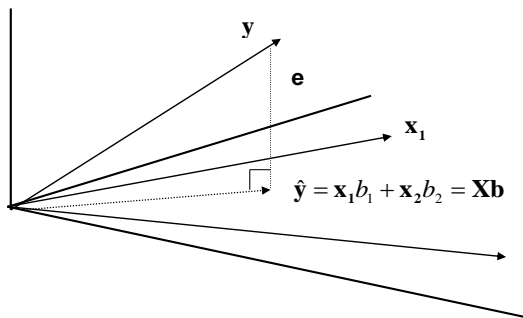
$$y = x'\beta + \varepsilon \quad (39)$$

em que “ ε ” é dado por $\varepsilon = y - x'\beta$.

- Exemplo: considere que x_1 e x_2 são vetores com a terceira coordenada igual a zero. Considere ainda que y é um vetor com a terceira coordenada não nula.

- Graficamente, temos:

Figura 4: Caso 2, y não está no espaço coluna de x



Propriedades do Erro de Projeção

- Considere a decomposição de y entre preditor linear e o erro:

$$y = \mathbf{x}'\beta + \varepsilon \quad (40)$$

- Da eq. (40) podemos chamar $\mathbf{x}'\beta$ de melhor preditor linear de y em \mathbf{x} ou projeção linear de y em \mathbf{x} . Ela frequentemente também é chamada de **regressão** de y em \mathbf{x} .

Observação

Propriedade 1

$$\mathbb{E}(\mathbf{x}\varepsilon) = \mathbf{0} \quad (41)$$

Demonstração.

Veja que

$$\begin{aligned}\mathbb{E}(\mathbf{x}\varepsilon) &= \mathbb{E}[\mathbf{x}(y - \mathbf{x}'\beta)] \\ &= \mathbb{E}(\mathbf{x}y) - \mathbb{E}(\mathbf{x}\mathbf{x}')\mathbb{E}[\mathbf{x}\mathbf{x}']^{-1}\mathbb{E}(\mathbf{x}y) = \mathbf{0}\end{aligned}\quad (42)$$

A eq. (42) é um conjunto de k equações, uma para cada regressor:

$$\mathbb{E}(\mathbf{x}_j\varepsilon) = 0 \quad (43)$$

para $j = 1, \dots, k$. O regressor \mathbf{x} contém uma constante, isto é $x_k = 1$. Neste caso, para $j = k$ é o mesmo que

$$\mathbb{E}(\varepsilon) = 0 \quad (44)$$

Assim, o erro de projeção tem uma média zero quando o regressor contém uma constante.

- É útil observar que:

$$\text{cov}(x_j, \varepsilon) = \mathbb{E}(x_j \varepsilon) - \mathbb{E}(x_j) \mathbb{E}(\varepsilon) \quad (45)$$

- Considerando o resultado da eq. (43) e (44) temos que x_j e ε são **não correlacionadas**.

Propriedades do Modelo de Projeção Linear

- ❶ Os momentos $\mathbb{E}(\mathbf{x}\mathbf{x}')$ e $\mathbb{E}(\mathbf{x}y)$ existe com elementos finitos.
- ❷ O coeficiente de projeção linear, β , existe, é único, e igual

$$\beta = [\mathbb{E}(\mathbf{x}\mathbf{x}')]^{-1}\mathbb{E}(\mathbf{x}y) \quad (46)$$

- ❸ O melhor preditor linear de y dado \mathbf{x} é

$$\mathcal{P}(y|\mathbf{x}) = \mathbf{x}'(\mathbb{E}(\mathbf{x}\mathbf{x}'))^{-1}\mathbb{E}(\mathbf{x}y) \quad (47)$$

- ❶ O erro de projeção $\varepsilon = y - \mathbf{x}'\beta$ existe e satisfaz

$$\mathbb{E}(\varepsilon^2) < \infty \quad (48)$$

e

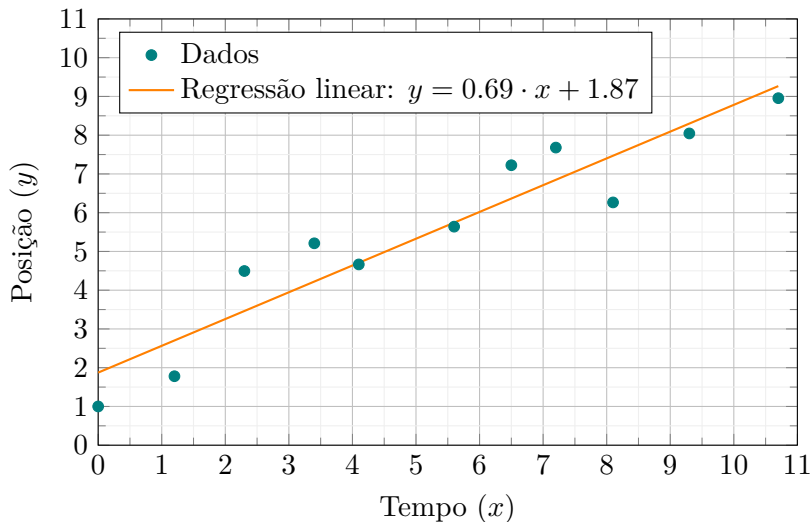
$$\mathbb{E}(\mathbf{x}\varepsilon) = \mathbf{0} \quad (49)$$

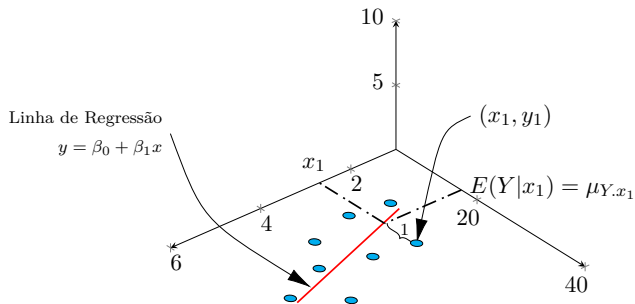
- ❷ Se \mathbf{x} contém uma constante, então

$$\mathbb{E}(\varepsilon) = 0 \quad (50)$$

- ❸ Se $\mathbb{E}|y|^r < \infty$ e $\mathbb{E}\|\mathbf{x}\|^r < \infty$ para $r \geq 1$, então $\mathbb{E}|\varepsilon|^r < \infty$.

Simulação





ECONOMETRIA I

MODELO DA FUNÇÃO DE ESPERANÇA CONDICIONAL LINEAR

Victor Oliveira

Núcleo de Economia Internacional e Desenvolvimento Econômico

PPGDE – 2023