Econometria I

Modelo da Função de Esperança Condicional Linear

Victor Oliveira

Núcleo de Economia Internacional e Desenvolvimento Econômico

PPGDE – 2024

Sumário

Definição

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

$$\varepsilon = y - m(\boldsymbol{x}) \tag{1}$$

- Por construção, ela vem da fórmula:

$$y = m(\boldsymbol{x}) + \varepsilon \tag{2}$$

- É útil entender que o termo de erro ε é derivado da distribuição conjunta de $(y, \boldsymbol{x})$ 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 $m(\boldsymbol{x}) = \mathbb{E}[y|\boldsymbol{x}]$ e o teorema do condicionamento

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

■

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

$$\mathbb{E}[\varepsilon] = \mathbb{E}\left[\mathbb{E}\left[\varepsilon|\boldsymbol{x}\right]\right] = \mathbb{E}[0] = 0 \tag{3}$$

Teorema: Propriedades da CEF

Teorema

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

Demonstração.

Pela desigualdade de Minkowski, temos que:

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

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

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

[a] Para 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

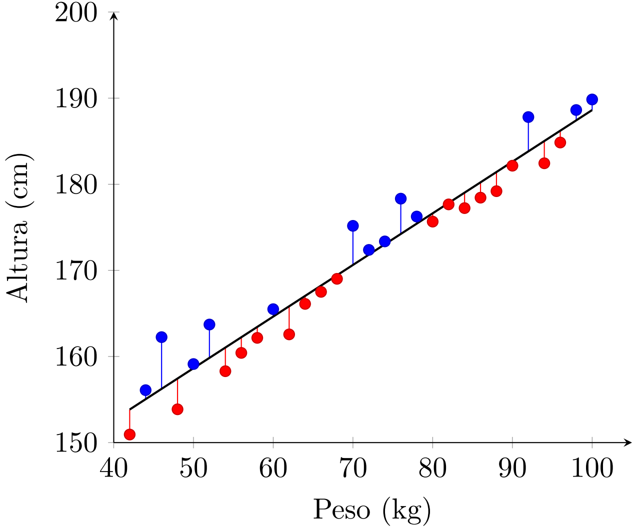
$$y = m(\boldsymbol{x}) + \varepsilon \tag{5}$$

$$\mathbb{E}[\varepsilon|\boldsymbol{x}] = 0 \tag{6}$$

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

- A condição $\mathbb{E}[\varepsilon|\boldsymbol{x}] = 0$ está implícito na definição de ε como a diferença entre y e a CEF $m(\boldsymbol{x})$.
- A equação $\mathbb{E}[\varepsilon|\boldsymbol{x}] = 0$ 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:

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

Teorema

Se $\mathbb{E}\left[y^2\right] < \infty$, *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 $\mathbb{E}[y|\boldsymbol{x}]$.

Variância Condicional

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

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

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

  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 $\mathbb{E}(y|\boldsymbol{x})$ como a "**parte explicada**" de y e $\varepsilon = y - \mathbb{E}(y|\boldsymbol{x})$ 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* $var(y) \geq var(y - \mathbb{E}(y|\boldsymbol{x}_1)) \geq var(y - \mathbb{E}(y|\boldsymbol{x}_1, \boldsymbol{x}_2))$.

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

$$\mathbb{E}\left(y - g(\boldsymbol{x})\right)^2 \tag{10}$$

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

Demonstração.

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

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

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

Teorema (Média condicional como melhor preditor)

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

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

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

- 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 $\boldsymbol{x}$ é:

$$
\begin{aligned}
\sigma^2 &= \operatorname{var}(y|\boldsymbol{x}) \\
&= \mathbb{E}\left[\left(y - \mathbb{E}(y|\boldsymbol{x})\right)^2 \middle| \boldsymbol{x}\right] \\
&= \mathbb{E}(\varepsilon^2|\boldsymbol{x}) \qquad (13)
\end{aligned}
$$

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

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

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

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.

$$u = \frac{\varepsilon}{\sigma(\boldsymbol{x})} \tag{15}$$

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

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

- Para a variância temos:

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

- 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 $\sigma^2(\boldsymbol{x})$ é uma constante e independente de $\boldsymbol{x}$.

Definição

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

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

Definição

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

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

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 $\sigma^2(\boldsymbol{x})$ 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 $\boldsymbol{x}$, isto é, $m(\boldsymbol{x}) = \mathbb{E}(y|\boldsymbol{x})$.
- Podemos escrever a equação média como:

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

- Para facilitar a notação, podemos escrever a eq. (18) como uma função do vetor $\boldsymbol{x}$.
- Para isso, temos de aumentar o vetor $\boldsymbol{x}$ 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 $\boldsymbol{x}$ é o intercepto, então $x_k = 1$. Assim, redefinimos o vetor $\boldsymbol{x}$ como vetor $k \times 1$

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{k-1} \\ 1 \end{pmatrix} \tag{19}$$

- Com essa redefinição, a CEF passa a ser:

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

- Temos que

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} \tag{21}$$

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 $\boldsymbol{x}$.

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

$$\nabla m(\boldsymbol{x}) = \beta \tag{22}$$

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

Modelo CEF Linear

- Seja

$$y = \mathbf{x}'\beta + \varepsilon \tag{23}$$
$$\mathbb{E}(\varepsilon|\boldsymbol{x}) = 0 \tag{24}$$

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

$$y = \mathbf{x}'\beta + \varepsilon \tag{25}$$
$$\mathbb{E}(\varepsilon|\boldsymbol{x}) = 0 \tag{26}$$
$$\mathbb{E}\left(\varepsilon^2|\boldsymbol{x}\right) = \sigma^2 \tag{27}$$

- O Teorema do Melhor Preditor mostrou que a média condicional $m(\boldsymbol{x})$ 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 $\mathbb{E}\left(y^2\right) < \infty$
  CR.2 $\|\boldsymbol{x}\| < \infty$
  CR.3 $\mathbf{Q}_{\boldsymbol{xx}} = \mathbb{E}(\boldsymbol{xx}')$, em que $Q_{\boldsymbol{xx}}$ é positiva definida

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

- O melhor preditor do erro quadrático médio é:

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

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

- A partir dessas informações podemos ter a seguinte definição.

Definição

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

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

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

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

- O erro quadrático médio de predição pode ser escrito como uma função quadrática de $\boldsymbol{\beta}$:

$$S(\boldsymbol{\beta}) = \mathbb{E}y^2 - 2\boldsymbol{\beta}'\mathbb{E}(\boldsymbol{x}y) + \boldsymbol{\beta}'\mathbb{E}(\boldsymbol{x}\boldsymbol{x}')\boldsymbol{\beta} \tag{30}$$

- Derivando com relação a $\boldsymbol{\beta}$, encontramos a CPO:

$$\frac{\partial}{\partial \boldsymbol{\beta}} S(\boldsymbol{\beta}) = 0 \quad \Longleftrightarrow \quad -2\mathbb{E}(\boldsymbol{x}y) + 2\mathbb{E}(\boldsymbol{x}\boldsymbol{x}')\boldsymbol{\beta} = 0 \tag{31}$$

- Podemos reescrevendo a (31) como:

$$2\mathbb{E}(\boldsymbol{x}y) = 2\mathbb{E}(\boldsymbol{x}\boldsymbol{x}')\boldsymbol{\beta} \tag{32}$$

- Assim,

$$\boldsymbol{Q}_{\boldsymbol{xy}} = \boldsymbol{Q}_{\boldsymbol{xx}}\boldsymbol{\beta} \tag{33}$$

em que $\boldsymbol{Q}_{\boldsymbol{xy}} = \mathbb{E}(\boldsymbol{x}y)$ é um vetor $k \times 1$ e $\boldsymbol{Q}_{\boldsymbol{xx}} = \mathbb{E}(\boldsymbol{x}\boldsymbol{x}')$ é um vetor $k \times k$.

- A solução é obtida invertendo a matriz $\boldsymbol{Q}_{\boldsymbol{xx}}$ e fazendo:

$$\begin{aligned} \boldsymbol{\beta} &= \boldsymbol{Q}_{\boldsymbol{xx}}^{-1}\boldsymbol{Q}_{\boldsymbol{xy}} \\ \boldsymbol{\beta} &= [\mathbb{E}(\boldsymbol{x}\boldsymbol{x}')]^{-1}\mathbb{E}(\boldsymbol{x}y) \end{aligned} \tag{34}$$

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:

$$\mathcal{P}(y|\boldsymbol{x}) = \boldsymbol{x}' \left[\mathbb{E}(\boldsymbol{x}\boldsymbol{x}')\right]^{-1} \mathbb{E}(\boldsymbol{x}y) \tag{35}$$

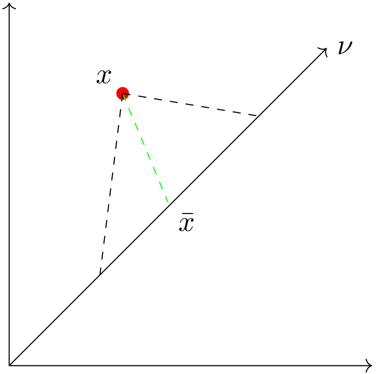
- Essa expressão podemos chamar de **projeção linear** de y em $\boldsymbol{x}$.
- O **erro de projeção** é dado por

$$\varepsilon = y - \boldsymbol{x}'\boldsymbol{\beta} \tag{36}$$

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:

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

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

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

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

- Logo,

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

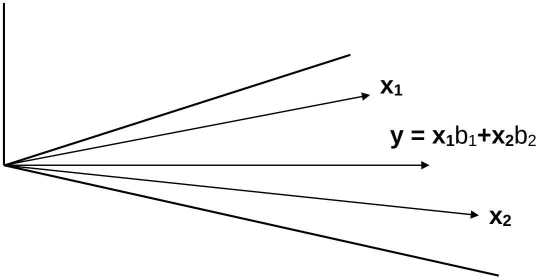
em que $\bar{x} = \nu c = \nu\left(\nu'\nu\right)^{-1}\nu' x = \mathcal{P}x$ e $\mathcal{P}_\nu = \nu\left(\nu'\nu\right)^{-1}\nu'$.

Intuição Gráfica para a Projeção Linear

- **Caso 1**: y está no espaço coluna de $\boldsymbol{x}$.
- Isso significa que y pode ser expresso exatamente como uma combinação linear das colunas de $\boldsymbol{x}$.
- Exemplo: considere que $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ 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 $\boldsymbol{x}$



40/50

Intuição Gráfica para a Projeção Linear

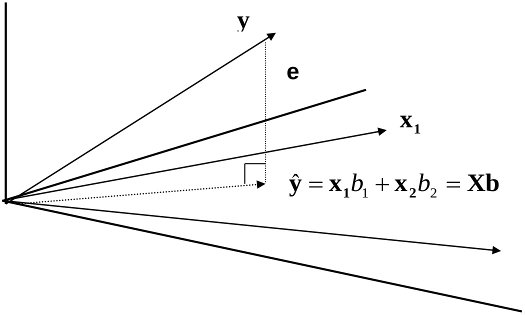
- **Caso 2**: y não está no espaço coluna de $\boldsymbol{x}$.
- Isso significa que y **não pode ser expresso como uma combinação linear das colunas de $\boldsymbol{x}$**. Ou seja, não há um vetor $\boldsymbol{b}$.
- Mas podemos escrever:

$$y = \boldsymbol{x}'\boldsymbol{\beta} + \varepsilon \tag{39}$$

  em que "ε" é dado por $\varepsilon = y - \boldsymbol{x}'\boldsymbol{\beta}$.
- Exemplo: considere que $\boldsymbol{x}_1$ e $\boldsymbol{x}_2$ 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 $\boldsymbol{x}$



Propriedades do Erro de Projeção

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

$$y = \boldsymbol{x}'\beta + \varepsilon \tag{40}$$

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

Observação

Propriedade 1

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

Demonstração.

Veja que

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

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

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

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

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

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

0

- É útil observar que:

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

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

Propriedades do Modelo de Projeção Linear

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

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

3. O melhor preditor linear de y dado $\boldsymbol{x}$ é

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

4. O erro de projeção $\varepsilon = y - \boldsymbol{x}'\beta$ existe e satisfaz

$$\mathbb{E}\left(\varepsilon^2\right) < \infty \tag{48}$$

e

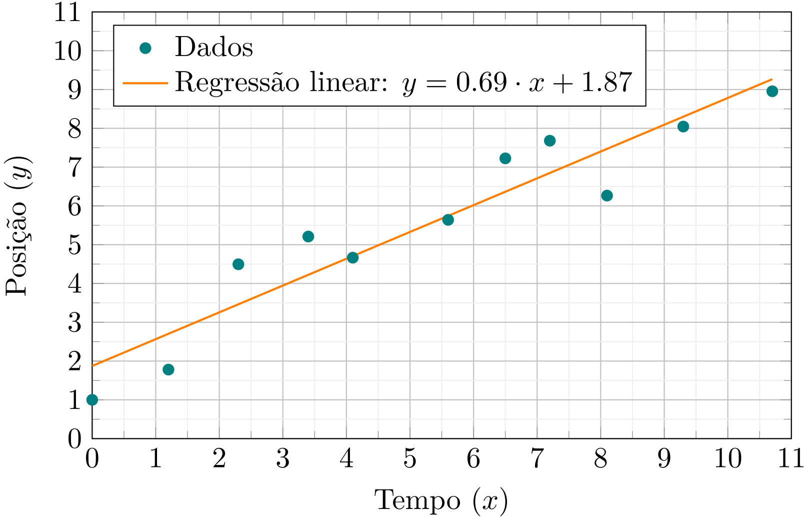
$$\mathbb{E}(\boldsymbol{x}\varepsilon) = \mathbf{0} \tag{49}$$

5. Se $\boldsymbol{x}$ contém uma constante, então

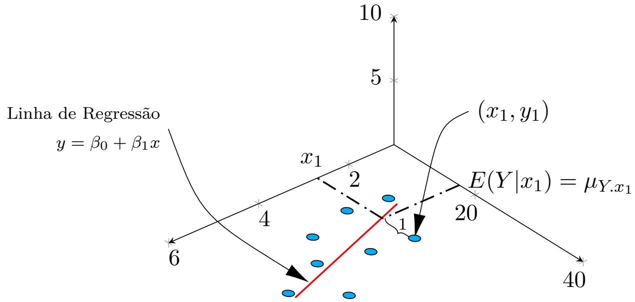
$$\mathbb{E}(\varepsilon) = 0 \tag{50}$$

6. Se $\mathbb{E}|y|^r < \infty$ e $\mathbb{E}\|\boldsymbol{x}\|^r < \infty$ para $r \geqslant 1$, então $\mathbb{E}|\varepsilon|^r < \infty$.

Simulação



Dados
Regressão linear: $y = 0.69 \cdot x + 1.87$
Posição (y)
Tempo (x)
0 1 2 3 4 5 6 7 8 9 10 11
0 1 2 3 4 5 6 7 8 9 10 11


Linha de Regressão
$y = \beta_0 + \beta_1 x$
10
5
(x_1, y_1)
x_1
2
4
6
1
20
40
$E(Y|x_1) = \mu_{Y.x_1}$

Econometria I

Modelo da Função de Esperança Condicional Linear

Victor Oliveira

Núcleo de Economia Internacional e Desenvolvimento Econômico

PPGDE – 2024