Notas em Econometria

Teoria e Aplicação

Gabriel Arruda

2024-07-30

Índice

Disclaimer

Este projeto teve início com base nas notas de aula do Prof. Dr. Fernando Aiube e do Prof. Dr. Francis Petterini, assim como nas notas de aula do Kotze (2019) e nos livros: Box et al. (2015), Hamilton (1994) e Enders (2014). O trabalho ainda precisa ser concluído e revisado. Vale destacar que pretendo incluir formas de aplicar os modelos em *Python*.

A idealização do projeto surgiu como uma maneira de estudo para eu aprender tanto a teoria quanto a aplicação prática de cada modelo. A implementação dos modelos será feita do zero, utilizando o mínimo de pacotes possível.

Alguns scripts estarão disponíveis dentro do texto, mas todos poderão ser acessados no meu GitHub.

Lembrando que a ideia é sempre utilizar o minimo de pacotes possiveis:

```python
# Importações globais
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

1 Introdução

1.1 Regressão linear

Regressão linear é uma ferramenta estatística usada para modelar a relação entre uma variável dependente e uma ou mais variáveis independentes, assumindo que essa relação pode ser descrita por uma linha reta. A ideia de se utilizar é uma é dado a sua simplicidade, tendo apenas um parâmetro de inclinação e um de intercepto, uma outra é que aqui se assume que as variáveis apresentam uma relação linear. A linha representa a melhor aproximação da tendência central dos dados. Aqui devemos partir de uma amostra, um par ordenado $\{x_i, y_i\}_{i=1}^{N}$, encontrar uma reta que melhor se ajusta a média dos dados, para isso, vamos partir da equação de uma reta.

$$y = \alpha + \beta x$$

Onde a ideia aqui é querer entender qual relação em que a variável x afeta a variável y, temos então que resolver dois problemas: primeiro é encontrar os parâmetros α e β que melhor se ajusta, sabendo que nem todo o y pode ser explicado pelo x, temos que adicionar uma variável à equação que consiga captar essa relação no modelo, essa variável será dada por u.

Podemos reescrever a equação acima como sendo um sistema de equações lineares

$$\begin{aligned}
y_1 &= \alpha + \beta x_1 + u_1 \\
y_2 &= \alpha + \beta x_2 + u_2 \\
y_3 &= \alpha + \beta x_3 + u_3 \\
&\vdots \\
y_n &= \alpha + \beta x_n + u_n
\end{aligned}$$

Note que esse é um sistema de n equações lineares com $n + 2$ incógnitas. E que pela regra de Cramer, sabemos que o sistema apresenta infinitas soluções. O que não nos ajuda e precisamos voltar ao problema, quais valores de α e β que melhor se ajusta? Uma maneira de se fazer isso, é minimizar a soma do erro quadrático $\left(\sum_{i=1}^{N} u_i^2\right)$ e para isso, vamos isolar o erro, elevar tudo ao quadrado e aplicar a recursividade.

$$\sum_{i=1}^{N} u_i^2 = \sum_{i=1}^{N} (y_i - \alpha - \beta x_i)^2$$

Dado isso, podemos dizer que podemos estimar valores de α e β que minimizam o erro quadrático. Seja $S(\alpha, \beta) = \sum_{i=1}^{N} u_i^2$ e sabendo que os valores dos parâmetros que zeram o gradiente $\nabla = \left(\frac{\partial S}{\partial \hat{\alpha}}, \frac{\partial S}{\partial \hat{\beta}}\right) = 0$ são os valores que minimizam o erro quadrático. Fazendo as derivadas...

$$\nabla = \begin{bmatrix} \dfrac{\partial S}{\partial \hat{\alpha}} \\ \dfrac{\partial S}{\partial \hat{\beta}} \end{bmatrix} = \begin{bmatrix} -2\sum_{i=1}^{N}(y_i - \hat{\alpha} - \hat{\beta}x_i) \\ -2\sum_{i=1}^{N}(y_i - \hat{\alpha} - \hat{\beta}x_i)(x_i) \end{bmatrix} = 0$$

Podemos multiplicar ambos os lados por $-\frac{1}{2}$ e abrir o somatório[1].

$$\begin{bmatrix} \sum_{i=1}^{N} y_i - n\hat{\alpha} - \hat{\beta}\sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} y_i x_i - \hat{\alpha}\sum_{i=1}^{N} x_i - \hat{\beta}\sum_{i=1}^{N} x_i^2 \end{bmatrix} = 0$$

Separando os termos, temos que

$$\begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} y_i x_i \end{bmatrix} = \begin{bmatrix} n\hat{\alpha} + \hat{\beta}\sum_{i=1}^{N} x_i \\ \hat{\alpha}\sum_{i=1}^{N} x_i + \hat{\beta}\sum_{i=1}^{N} x_i^2 \end{bmatrix}$$

$$\begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} y_i x_i \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 \end{bmatrix} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}$$

Podemos reorganizar da seguinte maneira:

$$\begin{bmatrix} n & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 \end{bmatrix} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} y_i x_i \end{bmatrix}$$

Pré-multiplicando ambos os lados pelo inverso da matriz que tem os valores de x:

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} y_i x_i \end{bmatrix}$$

Para termos certeza de que este é o ponto mínimo, devemos avaliar a matriz hessiana:

$$H = \begin{bmatrix} \dfrac{\partial^2 S}{\partial \alpha^2} & \dfrac{\partial^2 S}{\partial \alpha\, \partial \beta} \\ \dfrac{\partial^2 S}{\partial \alpha\, \partial \beta} & \dfrac{\partial^2 S}{\partial \beta^2} \end{bmatrix}$$

Logo:

$$H = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

Para ser um mínimo global, devemos ter que:

[1] Note que se somarmos n vezes um parâmetro é o mesmo que dizer n vezes o parâmetro, logo $\sum_{i=1}^{N} \hat{\alpha} = n\hat{\alpha}$.

- O primeiro menor principal será > 0
- O determinante do segundo menor principal será > 0

Com isso, podemos dizer que é um ponto de mínimo.

Podemos reescrever:

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

Abrindo:

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix}' \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Onde:

$$X = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix}'$$

$$\hat{B} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Então:

$$X'\hat{B} = X'Y$$
$$(X'X)\hat{B} = X'Y$$
$$(X'X)^{-1}(X'X)\hat{B} = (X'X)^{-1}X'Y$$
$$\hat{B} = (X'X)^{-1}X'Y$$
$$\boxed{\hat{B} = (X'X)^{-1}X'Y}$$

Então, sempre que estamos falando do estimador do **MQO**, estamos nos referindo à fórmula fechada:

$$\hat{B} = (X'X)^{-1}X'Y$$

Agora, temos que pensar da seguinte maneira: dado que conseguimos construir os estimadores, como podemos criar seus intervalos de confiança? Para isso, podemos substituir Y por $XB + U$:[2]

$$\hat{B} = (X'X)^{-1}X'(XB + U)$$

[2]Vale lembrar que B sem chapéu é o melhor ajuste possível da reta, os valores que só Deus sabe.

$$= (X'X)^{-1}X'XB + (X'X)^{-1}X'U$$

Assumindo que os dados não tenham problema de multicolinearidade perfeita, a matriz $(X'X)^{-1}$ deve existir para que $I = (X'X)^{-1}X'X$:

$$\hat{B} = B + (X'X)^{-1}X'U \tag{1.1}$$

Observamos na equação Equação 1.1 que a componente do estimador influenciada pelo erro, especificamente $X'U$, ilustra uma premissa importante do modelo: $\mathbb{E}(X|U) = 0$. Isso implica que, idealmente, todas as variáveis explicativas deveriam ser exógenas, não apresentando qualquer correlação com o termo de erro. Mas, é importante reconhecer que, na prática, alcançar uma exogeneidade completa é praticamente inviável; assim, é realista esperar que qualquer modelo econômico possa manifestar algum nível, mesmo que mínimo, de endogeneidade.

Subtraind os dois lados da Equação 1.1 por $-B$ e pós-multiplicando por $(\hat{B} - B)'$:

$$(\hat{B} - B)(\hat{B} - B)' = (X'X)^{-1}X'U[(X'X)^{-1}X'U]'$$

Desenvolvendo a parte esquerda dessa igualdade, temos que:

$$\begin{bmatrix} \hat{B}_1 - B \\ \hat{B}_2 - B \end{bmatrix} \begin{bmatrix} \hat{B}_1 - B & \hat{B}_2 - B \end{bmatrix}'$$

Multiplicando e aplicando o operador da esperança:

$$\begin{bmatrix} \mathbb{E}[(\hat{B}_1 - B)^2] & \mathbb{E}[(\hat{B}_1 - B)(\hat{B}_2 - B)] \\ \mathbb{E}[(\hat{B}_1 - B)(\hat{B}_2 - B)] & \mathbb{E}[(\hat{B}_2 - B)^2] \end{bmatrix}$$

Onde a diagonal principal é a variância de $\hat{B}_1$ e o resto é a covariância, então montamos a matriz de variância-covariância:

$$\begin{bmatrix} \mathrm{Var}(\hat{B}_1) & \mathrm{Cov}(\hat{B}_1, \hat{B}_2) \\ \mathrm{Cov}(\hat{B}_1, \hat{B}_2) & \mathrm{Var}(\hat{B}_2) \end{bmatrix}$$

A partir disso, poderíamos montar um intervalo de confiança para os betas se não fosse um pequeno problema... Aqui precisamos do valor de β, e que só Deus sabe. Vamos então olhar para o lado direito da igualdade[3]

$$= (X'X)^{-1}X'UU'X[(X'X)^{-1}]'$$

$$= (X'X)^{-1}X'UU'X(X'X)^{-1}$$

[3]Vale lembrar que $(AB)' = B'A'$

Abrindo UU' e aplicando o operador da esperança:

$$UU' = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}'$$

$$= \begin{bmatrix} \mathbb{E}(u_1)^2 & \mathbb{E}(u_1, u_2) & \cdots & \mathbb{E}(u_1, u_n) \\ \mathbb{E}(u_2, u_1) & \mathbb{E}(u_2)^2 & \cdots & \mathbb{E}(u_2, u_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}(u_n, u_1) & \mathbb{E}(u_n, u_2) & \cdots & \mathbb{E}(u_n)^2 \end{bmatrix}$$

Vamos ter que na diagonal principal é a variância dos erros e $\forall\, \mathbb{E}(u_i, u_j)$ em que $i \neq j$ temos a covariância dos erros. Sob as hipóteses de homoscedasticidade[4] e não autocorrelação, vamos ter que:

$$UU' = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$= \sigma^2 I$$

Então continuando, vamos ter que:

$$(X'X)^{-1} X' \sigma^2 X (X'X)^{-1}$$

$$= \sigma^2 (X'X)^{-1} X'X (X'X)^{-1}$$

$$= \sigma^2 (X'X)^{-1}$$

Agora sim temos uma matriz de variância-covariância, mas percebemos que ao longo do caminho foi necessário fazer algumas hipóteses questionáveis, como a homoscedasticidade e

[4] Variância dos erros é constante, isto é, $\mathbb{E}(u_i^2) = \sigma^2$, $\forall i = 1, \ldots, n$

não-autocorrelação. Outro problema dessa matriz de variância-covariância é que nela precisamos da média do erro, mas só Deus sabe o erro... o máximo que podemos fazer é procurar uma estimativa para esse erro, e vamos chamá-lo de **resíduo**. Para diferenciar, o **resíduo** é a parte do modelo que não conseguimos explicar e o **erro** é tudo aquilo que afeta o Y, mas não é o X.

1.2 Conceitos de Convergência

A ideia aqui é entender o que acontece com a amostra à medida que seu tamanho vai para infinito. Embora isso seja puramente teórico, conseguimos tirar algumas ideias para o caso da amostra finita. As duas ideias principais são:

1. A **lei dos grandes números** diz que a média da amostra $X_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ **converge em probabilidade** para a expectativa $\mu = \mathbb{E}(X_i)$. Isso significa que X_n está próximo de μ com alta probabilidade.
2. O **teorema do limite central** diz que $\sqrt{n}(X_n - \mu)$ **converge em distribuição** para uma distribuição Normal. Isso significa que a média da amostra tem aproximadamente uma distribuição Normal para grandes valores de n.

Definição 1.1 (Convergencia)**.** A sequência[5] de variáveis aleatórias, $X_1, X_2, \ldots$, **converge em probabilidade** para uma variável aleatória X, se $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0 \quad \text{ou} \quad \lim_{n\to\infty} \mathbb{P}(|X_n - X| < \varepsilon) = 1$$

Note que se $n \to \infty \implies |X_n - X| \to 0$ e isso quer dizer que no limite, a sequência vai se aproximar muito da variável aleatória.

Teorema 1.1 (Teorema da Lei dos Grandes Números - Fraca)**.** *Seja $X_1, X_2, \ldots$, variáveis aleatórias iid com $\mathbb{E}[X_i] = \mu$ e $Var[X_i] = \sigma^2 < \infty$. Defina $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Então para todo $\varepsilon > 0$:*

$$\lim_{n\to\infty} \mathbb{P}(|\bar{X}_n - \mu| < \varepsilon) = 1$$

*Então, $\bar{X}_n$ **converge em probabilidade** para μ.*

```python
import numpy as np
import matplotlib.pyplot as plt
#plt.rcParams['figure.figsize'] = (1, 1)
```

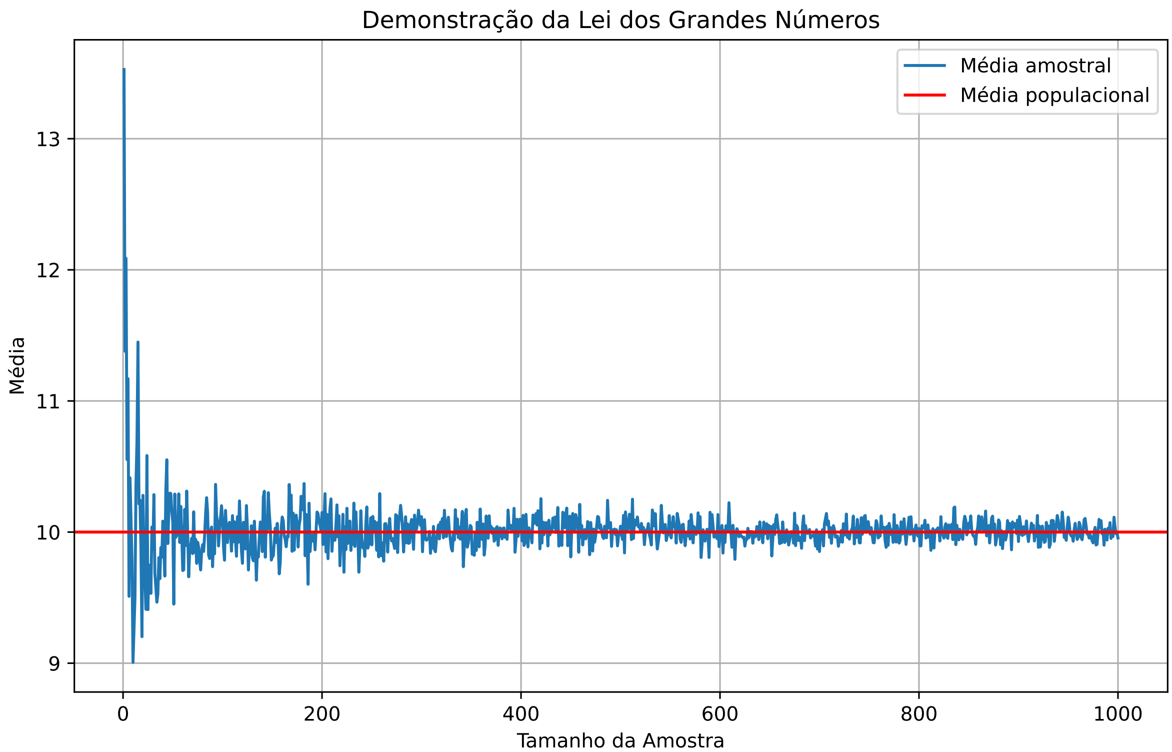
[5] Lembre-se da ideia de convergência de uma sequência. Dado um $\varepsilon > 0$, dizemos que $x_k \to x$ se existir um k_0, em que $\forall k \geq k_0 \implies |x_k - x| < \varepsilon$

```python
# Configurações da simulação
np.random.seed(0)  # Para reprodutibilidade
pop_mean = 10  # Média populacional
pop_std = 2  # Desvio padrão populacional
n_simulations = 1000  # Número de amostras/simulações
sample_sizes = np.arange(1, n_simulations + 1)  # Tamanhos das amostras

# Simulação da Lei dos Grandes Números
sample_means = []
for size in sample_sizes:
    sample = np.random.normal(pop_mean, pop_std, size)
    sample_mean = np.mean(sample)
    sample_means.append(sample_mean)

# Conversão para array do numpy
sample_means = np.array(sample_means)

# Visualização dos resultados
plt.figure(figsize=(10, 6))
plt.plot(sample_sizes, sample_means, label='Média amostral')
plt.axhline(y=pop_mean, color='r', linestyle='-', label='Média populacional')
plt.xlabel('Tamanho da Amostra')
plt.ylabel('Média')
plt.title('Demonstração da Lei dos Grandes Números')
plt.legend()
plt.grid(True)
plt.show()
```



Teorema 1.2 (Teorema do Limite Central)**.** *Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas com média μ e variância σ^2. Seja $X_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Então,*

$$Z_n = \frac{X_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\sqrt{n}(X_n - \mu)}{\sigma} \xrightarrow[n\to\infty]{} Z$$

onde Z tem uma distribuição normal padrão. Em outras palavras,

$$\lim_{n\to\infty} \mathbb{P}(Z_n \leq z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx.$$

Referencias

Box, George EP, Gwilym M Jenkins, Gregory C Reinsel, e Greta M Ljung. 2015. *Time series analysis: forecasting and control*. John Wiley & Sons.

Enders, W. 2014. *Applied Econometric Times Series*. Wiley Series em Probability e Statistics. Wiley. https://books.google.com.br/books?id=lmr9oQEACAAJ.

Hamilton, James Douglas. 1994. *Time series analysis*. Princeton university press.

Kotze, Kevin. 2019. «Time Series Analusis». 2019. https://www.economodel.com/time-series-analysis-2019.