

Notas em Econometria

Teoria e Aplicação

Gabriel Arruda

2024-08-01

Índice

Disclaimer	3
1 Introdução	4
1.1 Regressão linear	4
Estimando uma regressao no python	9
1.2 Conceitos de Convergência	11
1.3 Séries de tempo	14
1.3.1 Processo determinístico	14
1.3.2 Processo estocástico	15
1.3.3 Estacionariedade e Autocorrelação	18
1.3.4 Estacionariedade	19
1.3.5 Função de autocorrelação (FAC)	19
2 ARIMA	21
3 Modelos ARIMA	22
3.1 Modelos lineares estacionários	22
3.1.1 Modelo Autorregressivo (AR)	22
3.1.2 Modelo de Média móvel (MA)	25
3.2 Modelos lineares não-estacionários	26
4 Identificação do Modelo	28
5 Testes do Modelo	29
5.1 Testes de Raiz Unitária - Dick-Fuller	29
5.1.1 Teste de autocorrelação - Ljung-Box e Box-Pierce	30
5.1.2 Teste de heterocedasticidade - ARCH-LM	31
5.1.3 Teste de normalidade - Jarque-Bera	31
6 Modelos de volatilidade	32
6.1 Modelos lineares	32
6.1.1 ARCH	32
6.1.2 GARCH	33
Referencias	34

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 mínimo de pacotes possíveis:

```
# 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...

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

Podemos multiplicar ambos os lados por $-\frac{1}{2}$ e abrir o somatório¹.

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

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:

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

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:

¹Note 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 Equação 1.1:

$$\hat{B} = (X' X)^{-1} X' Y \quad (1.1)$$

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

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

²Vale 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 $I = (X'X)^{-1}X'X$:

$$\hat{B} = B + (X'X)^{-1}X'U \quad (1.2)$$

Observamos na equação Equação 1.2 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.

Subtraindo os dois lados da Equação 1.2 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:

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

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 igualdade³

$$\begin{aligned} &= (X'X)^{-1}X'UU'X[(X'X)^{-1}]' \\ &= (X'X)^{-1}X'UU'X(X'X)^{-1} \end{aligned}$$

³Vale lembrar que $(AB)' = B'A'$

Abrindo UU' e aplicando o operador da esperança:

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

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 homoscedasticidade⁴ e não autocorrelação, vamos ter que:

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

Então continuando, vamos ter que:

$$\begin{aligned}
 &(X'X)^{-1}X'\sigma^2X(X'X)^{-1} \\
 &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\
 &= \sigma^2(X'X)^{-1}
 \end{aligned} \tag{1.3}$$

Agora sim temos uma matriz de variância-covariância (Equação 1.3), mas percebemos que ao longo do caminho foi necessário fazer algumas hipóteses questionáveis, como a homoscedasticidade e não-autocorrelação. Outro problema dessa matriz de variância-covariância é que

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

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 . Então um estimador para a variancia dos resíduos é:

$$\hat{\sigma}^2 = \frac{SQR}{n - k} = \frac{\sum_{i=1}^n u^2}{n - k}$$

Onde:

- $n \rightarrow$ Tamanho da amostra
- $k \rightarrow$ Numero de coeficientes

Estimando uma regressao no python

Primeiro, vamos começar lendo os dados. Aqui vamos utilizar o dataset *mtcars*, juntos dos pacotes que serão usados

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm

mtcars = sm.datasets.get_rdataset('mtcars').data
```

Vamos agora criar uma função que estime os valores dos $\hat{\beta}$ e calcula o erro padrão de cada coeficiente, isso é importante pois será usado mais tarde para o calculo das estatísticas t . A Função recebe como um argumento 'data' nossa base de dados; y é a variavel dependete e x são nossas variáveis explicativas, note que da maneira que foi construida nossa função y e x deve ser strings.

```
def ols(data, y, x):
    X = data[x].values
    Y = data[y].values

    # Adiciona uma coluna de "1" para gerar o intercepto
    X = np.column_stack((np.ones(X.shape[0]), X))

    # Calcula OLS
    XtX = np.transpose(X) @ X
    XtX_inv = np.linalg.inv(XtX)
    XtY = np.transpose(X) @ Y
```

```

betas = XtX_inv @ XtY

# Calcula os resíduos
residuals = Y - X @ betas
sqr = np.sum(residuals**2) # Soma dos quadrados dos resíduos
n, k = X.shape
residual_variance = sqr / (n - k) # Variância dos resíduos

# Calcula o erro padrão dos coeficientes
standard_errors = np.sqrt(residual_variance * np.diag(XtX_inv))

return betas, standard_errors

```

tendo nossos valores em mãos, precisamos de uma função que calcule as estatísticas t

```

def t_stats(betas, standard_errors):
    # Calcula os valores-t
    t_values = betas / standard_errors
    return t_values

```

Para organizar tudo e melhorar a visibilidade dos dados, precisamos de uma função que pegue todos esses valores que crie um dataframe

```

def results_dataframe(data, y, x):
    # Calcula betas e erros padrão
    betas, standard_errors = ols(data, y, x)

    # Calcula estatísticas t
    t_values = t_stats(betas, standard_errors)

    # Cria um DataFrame para uma melhor visualização
    coef_names = ['Intercept'] + x
    results = pd.DataFrame({
        'Coeficiente': betas,
        'Erro Padrão': standard_errors,
        'Estatística t': t_values
    }, index=coef_names)

    return results

```

Temos então

```
results_dataframe(mtcars, 'mpg', ['cyl', 'disp'])
```

	Coefficiente	Erro Padrão	Estatística t
Intercept	34.660995	2.547004	13.608536
cyl	-1.587277	0.711844	-2.229809
disp	-0.020584	0.010257	-2.006696

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 (Convergência). A sequência⁵ de variáveis aleatórias, X_1, X_2, \dots , **converge em probabilidade** para uma variável aleatória X , se $\forall \varepsilon > 0$,

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

Note que se $n \rightarrow \infty \implies |X_n - X| \rightarrow 0$ 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, \dots , variáveis aleatórias iid com $\mathbb{E}[X_i] = \mu$ e $\text{Var}[X_i] = \sigma^2 < \infty$. Defina $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Então para todo $\varepsilon > 0$:*

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

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

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

```
def lei_grandes_numeros(pop_mean, pop_std, sample_sizes):
    np.random.seed(0) # Para reprodutibilidade

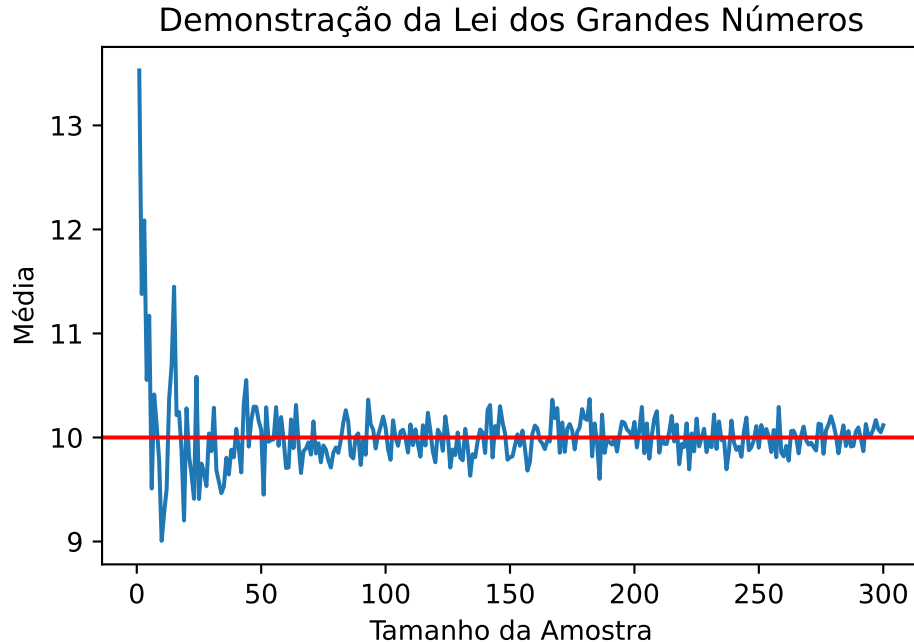
    sample_means = [np.random.normal(pop_mean, pop_std, size).mean() for size in sample_sizes]

    return np.array(sample_means)

pop_mean = 10
pop_std = 2
n_simulations = 300
sample_sizes = range(1, n_simulations + 1)

# Gerando os dados
sample_means = lei_grandes_numeros(pop_mean, pop_std, sample_sizes)

# Visualização dos resultados
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.show()
```



Teorema 1.2 (Teorema do Limite Central). *Sejam X_1, \dots, 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,*

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

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

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

```
def tcl(N, pop):
    size_sample = 100
    sample_mean = [
        np.random.choice(pop, size=size_sample, replace=True).mean() for _ in range(N)
    ]
    return np.array(sample_mean)

def plot_tcl(N_values, pop):

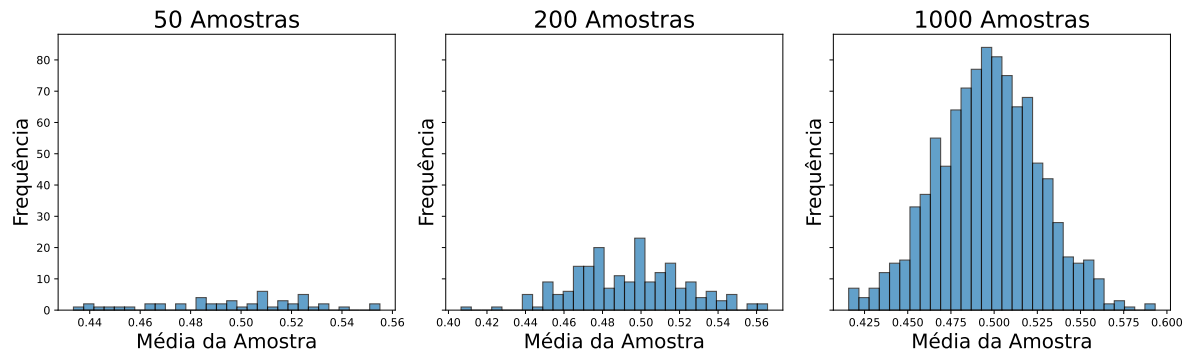
    fig, axs = plt.subplots(1, len(N_values), figsize=(15, 5), sharey=True)

    for i, N in enumerate(N_values):
        sample_means = tcl(N, pop)
        axs[i].hist(sample_means, bins=30, edgecolor='k', alpha=0.7)
        axs[i].set_title(f'{N} Amostras', fontsize=21)
        axs[i].set_xlabel('Média da Amostra', fontsize=18)
        axs[i].set_ylabel('Frequência', fontsize=18)

    plt.tight_layout(rect=[0, 0.03, 1, 0.95])
    plt.show()

np.random.seed(0)
# Gerar a população
pop = np.random.uniform(size=1000)

# Plota um histograma com diferentes números de amostras
plot_tcl([50, 200, 1000], pop)
```



1.3 Séries de tempo

Uma série de tempo é caracterizada por uma sequência de observações tomada sequencialmente no tempo. Podemos chamar essa sequência de observações $(y_{\{1\}}, y_{\{2\}}, y_{\{3\}}, \dots)$ como sendo variáveis aleatórias, onde $y = f(t) + \varepsilon_t$, em outras palavras, cada observação é uma função do tempo mais um fator aleatório (ruído branco). Como esse y depende do tempo, vamos ter um valor diferente de y para cada valor de $t = 0, 1, 2, \dots, T$; $y^* = f(t^*) + \varepsilon_{t^*}$. Então y_1 é o valor de y no período 1, y_2 é o valor de y no período 2 e assim vai. Uma série de tempo pode ser dividida entre um processo determinístico ou estocástico.

1.3.1 Processo determinístico

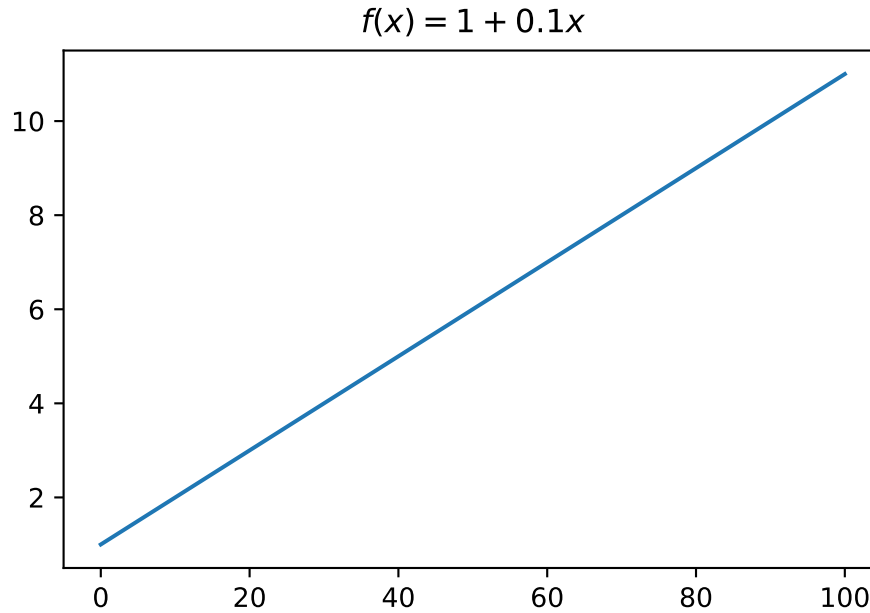
É um processo que não depende de um termo aleatório e sempre irá ter o mesmo resultado dado um valor inicial. O processo $T_t = 1 + 0,1t$ é um processo determinístico, pois não depende de nenhum fator aleatório, estando sempre acompanhado de uma constante (1) e um termo de tendência determinística ($0,1t$).

Podemos observar essa função melhor, gerando o seguinte gráfico no **Python**.

```
def trend(f):
    x = np.linspace(0, 100, 1000)
    y = f(x)

    plt.clf()
    plt.plot(x, y)
    plt.title(r'$f(x) = 1 + 0.1x$')
    plt.show()

# f(x) = 1 + 0.1*x
trend(lambda x: 1 + 0.1 * x)
```



1.3.2 Processo estocástico

O que diferencia um processo estocástico do determinístico é o fator aleatório, por mais que esse fator seja aleatório ele apresenta uma função de distribuição de probabilidade, por mais que se saiba qual é o valor inicial de uma série, não dá para saber com exatidão o próximo valor, mas podemos atribuir probabilidades a diferentes valores.

1.3.2.1 Ruído Branco

Um ruído branco é uma sequência serial de variáveis aleatórias não correlacionadas com média zero e variância finita e constante, seria aquele erro (u) da primeira parte do curso. Assumimos duas condições importantes para o ruído branco: ele deve ser independente um do outro e deve apresentar uma distribuição normal⁶. Sendo escrito da seguinte forma:

$$\varepsilon_t \sim i.i.d.\mathcal{N}(0, \sigma^2)$$

$$\varepsilon_t \sim RB(0, \sigma^2)$$

O ruído branco assume 3 propriedades:

- $\mathbb{E}[\varepsilon_t] = \mathbb{E}[\varepsilon_t \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \dots] = 0$
- $\mathbb{E}[\varepsilon_t \varepsilon_{t-j}] = \text{cov}[\varepsilon_t \varepsilon_{t-j}] = 0$

⁶Chamada também de *Distribuição normal gaussiana*.

- $\text{var}[\varepsilon_t] = \text{cov}[\varepsilon_t \varepsilon_t] = \sigma_{\varepsilon_t}^2$

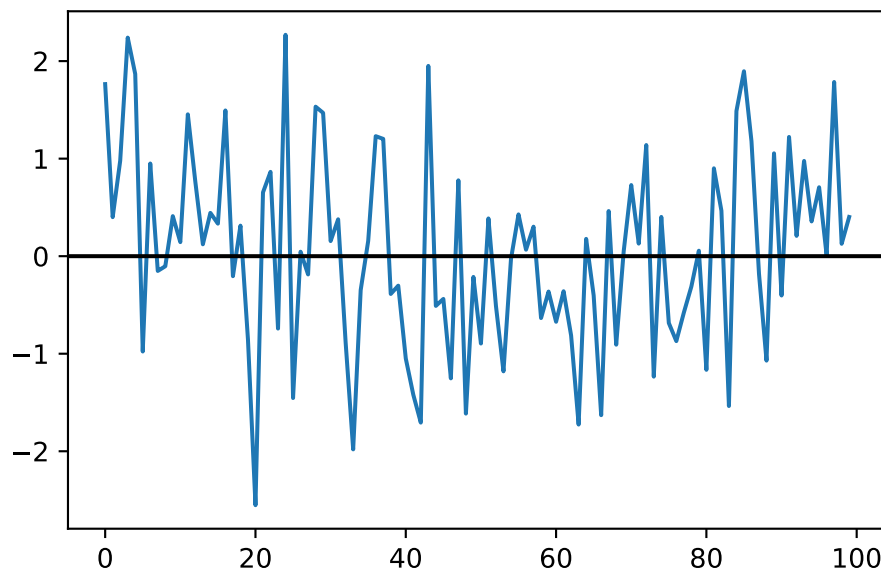
As duas primeiras propriedades dizem respeito à impossibilidade de preditividade e à ausência de autocorrelação. A terceira diz respeito à homocedasticidade, a variância ser constante.

Para visualizar isso, criei uma função em **Python** que gera um conjunto de dados de ruído branco usando a função `np.random.normal()` que cria um conjunto aleatório de dados de uma distribuição normal, onde n é o tamanho da amostra desejado.

```
def white_noise(n):
    np.random.seed(0)
    white_noise = np.random.normal(0, 1, n)

    plt.clf()
    plt.plot(white_noise)
    plt.axhline(0, color = 'black')
    plt.show()

white_noise(100)
```



1.3.2.2 Passeio aleatório

Esse tempo passeio aleatório se diz pelo fato de que o valor de uma variável em um determinado período é igual ao seu valor no período passado mais um fator aleatório determinado por ε_t , sendo descrito por:

$$y_t = y_{t-1} + \varepsilon_t$$

Vamos supor um valor inicial para y_t como sendo y_1 , então:

$$y_1 = y_0 + \varepsilon_1$$

$$y_2 = y_1 + \varepsilon_2$$

$$y_2 = y_0 + \varepsilon_1 + \varepsilon_2$$

Resolvendo isso recursivamente, temos:

$$y_t = y_0 + \sum_{i=1}^t \varepsilon_i$$

Se $y_0 \sim 0$, podemos então dizer que o passeio aleatório é uma soma acumulada de ruídos brancos:

$$y_t = \sum_{i=1}^t \varepsilon_i$$

Pode acontecer o caso do passeio aleatório ter uma constante γ adicionada:

$$y_1 = \gamma + y_0 + \varepsilon_1$$

Logo, quando $y_0 \sim 0$, vamos ter que:

$$y_t = \gamma t + \sum_{i=1}^t \varepsilon_i$$

```
def random_walk(trend, n):
    np.random.seed(0)

    yt = np.zeros(n)
    epsilon = np.random.normal(0,1,n)
    for t in np.arange(1,n):
        yt[t] = t * trend + np.sum(epsilon[:t+1])

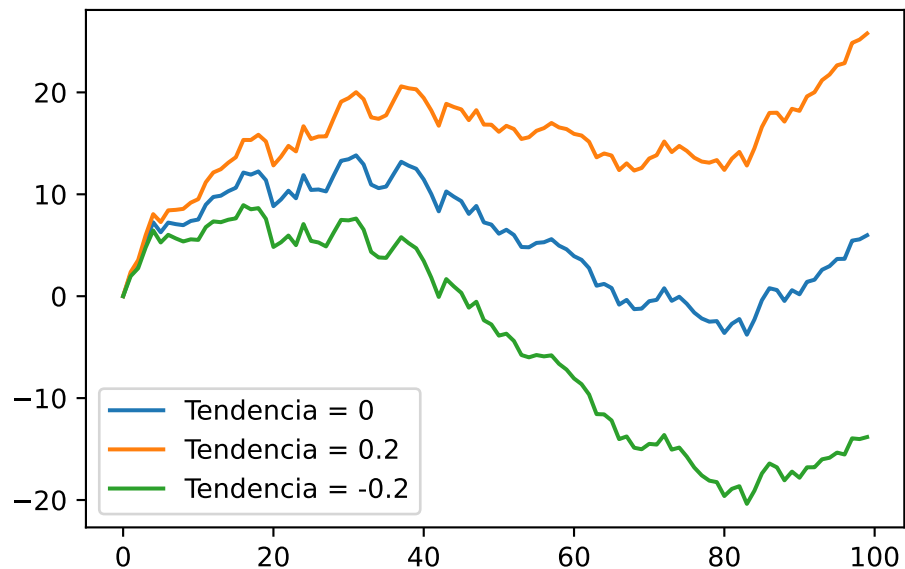
    return yt

def plot_rw(trend_list, t):
```

```
plt.clf()
for trend in trend_list:
    rw = random_walk(trend, t)
    plt.plot(rw, label=f'Tendencia = {trend}')

plt.legend()
plt.show()

trend_list = [0, 0.2, -0.2]
plot_rw(trend_list, t=100)
```



1.3.3 Estacionariedade e Autocorrelação

Antes de entrar no assunto de autocorrelação, devemos ter em mente algumas propriedades referentes ao primeiro momento (esperança) e segundo momento (variância) de um passeio aleatório. O primeiro momento de um passeio aleatório pode ser calculado pela sua média:

$$\mathbb{E}[y_t] = \mathbb{E}[y_0 + \sum_{i=1}^t \varepsilon_i] = \mathbb{E}[y_0] + \mathbb{E}[\sum_{i=1}^t \varepsilon_i] = \mathbb{E}[y_0] + 0 = y_0 = \mu$$

Nota-se que a esperança (média) de y_t não depende de t , sendo uma constante e é por isso que estou chamando de μ .

O segundo momento é a variância:

$$\begin{aligned}
\text{Var}[y_t] &= \mathbb{E}[y_t^2] = \mathbb{E}[(y_t - \mathbb{E}[y_t])^2] \\
&= \mathbb{E}[y_t^2] - \mathbb{E}[y_0]^2 \\
&= \mathbb{E}[(y_0 + \sum \varepsilon_i)^2] \\
&= \mathbb{E}[y_0^2 + 2y_0 \sum \varepsilon_i + \sum \varepsilon_i^2] - \mu^2 \\
&= \mathbb{E}[y_0^2] + 2y_0 \sum \mathbb{E}[\varepsilon_i] + \sum \mathbb{E}[\varepsilon_i^2] - \mu^2 \\
&= \mu^2 + 0 + \sum \sigma^2 - \mu^2 = t\sigma^2
\end{aligned}$$

Podemos observar que a variância de um processo aleatório não é constante, pois vai depender de t .

Outra parada importante é definir a covariância para k lags:

$$\text{Cov}[y_t, y_{t-k}] = \mathbb{E}\{(y_t - \mathbb{E}[y_t])(y_{t-k} - \mathbb{E}[y_{t-k}])\} = \mathbb{E}[(y_t - \mu)(y_{t-k} - \mu)]$$

1.3.4 Estacionariedade

A hipótese da estacionariedade é um caso particular do **processo estocástico**, no qual se assume que o processo está em um estado de equilíbrio. O caso dos processos estritamente estacionários (estacionariedade forte) é quando **todas** as suas propriedades (momentos) não são afetadas pelo tempo. Como essas condições são muito restritas, na grande maioria das vezes nos referimos a esse processo estocástico como **fracamente estacionário**, ou **covariância-estacionário**, ou **estacionário de segunda ordem**. Isso quer dizer que apenas a média e a variância devem ser constantes e que a covariância deve depender apenas do número de lags (k):

$$\begin{aligned}
\mathbb{E}[y_t] &= \mu \\
\text{Var}[y_t] &= \mathbb{E}[(y_t - \mu)^2] = \sigma^2 \\
\text{Cov}[y_t, y_{t+k}] &= \mathbb{E}[(y_t - \mu)(y_{t+k} - \mu)] = \gamma_k
\end{aligned}$$

1.3.5 Função de autocorrelação (FAC)

Ao assumir a hipótese da estacionariedade, vamos ter que a função conjunta de probabilidade de y_{t_1} e y_{t_2} será a mesma para todo o tempo t_1 e t_2 , que será constante. Isso implica que a **covariância** entre y_t e y_{t+k} será separada apenas por k intervalos de tempo (lag). Assim, a *autocovariância* ao lag k é definida por:

$$\gamma_k = \text{Cov}[y_t, y_{t+k}] = \mathbb{E}[(y_t - \mu)(y_{t-k} - \mu)]$$

Da mesma forma, teremos que a **autocorrelação** é dada por:

$$\rho_k = \frac{\mathbb{E}[(y_t - \mu)(y_{t-k} - \mu)]}{\sqrt{\mathbb{E}[(y_t - \mu)^2]\mathbb{E}[(y_{t-k} - \mu)^2]}} = \frac{\mathbb{E}[(y_t - \mu)(y_{t-k} - \mu)]}{\sigma^2}$$

Como em um processo estacionário a variância é constante, vamos ter que $\gamma_0 = \sigma^2$. Podemos então dizer que a autocorrelação no lag k é:

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

E isso implica que $\rho_0 = 1$ se $k = 0$.

A **Função de autocorrelação** é quando se plota um gráfico ρ_k contra o lag k .

2 ARIMA

3 Modelos ARIMA

Antes de chegar no modelo ARIMA, vou passar pelo AR, dps o MA.

3.1 Modelos lineares estacionários

3.1.1 Modelo Autorregressivo (AR)

Um processo é dito autoregressivo quando ele carrega consigo msm uma certa persistência, quer dizer que o passado influencia o presente, quando apenas o período anterior influencia o atual vamos ter um AR(1)

$$y_t = \phi y_{t-1} + \varepsilon_t$$

onde ϕ é um coeficiente tbm chamado de **peso**, seria o peso da influencia de um período no outro. Um caso especial do processo *autoregressivo* é o AR(p)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

onde todos os pesos ϕ são diferentes de 0.

Podemos reescrever esse AR(p) usando o **operador lag**

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) y_t = \varepsilon_t$$

ou então

$$\phi(L) y_t = \varepsilon_t$$

onde $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$.

O processo *autoregressivo* deve respeitar algumas condições para ser estacionário. Seja o processo AR(1) para exemplificar

$$(1 - \phi L) y_t = \varepsilon_t$$

e para que esse processo seja estacionário, devemos ter que todas as raízes de $|\phi(L)| > 1$ e isso implica que o parâmetro ϕ do processo AR(1) deve ser $|\phi| < 1$ para ser estacionário. Como a raiz de $1 - \phi L = 0$ é $L = \phi^{-1}$, isso quer dizer que a raiz **está fora do círculo unitário**, já que $|\phi| < 1$.

O mesmo raciocínio pode ser levado para o processo autorregressivo de ordens superiores, seja o caso do AR(p)

$$\phi(L)y_t = (1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)y_t = \varepsilon_t$$

onde a condição de estacionaridade é que **TODAS AS RAÍZES** de $\phi(L) = 0$ devem ser menores que 1 em módulo, estar **fora do círculo unitário**. Existem alguns testes para ver se uma série é estacionária ou não, como o do Dickey-Fuller e KPSS, que serão abordados mais à frente.

Um processo autorregressivo é dito **inversível** quando podemos reescrever o **AR(1)** em forma de **MA(∞)**, isso é usado em alguns casos como o das funções de impulso e resposta que são usadas quando queremos avaliar o quanto um choque em uma variável afeta a outra. Seja então o processo AR(1)

$$y_t = \phi y_{t-1} + \varepsilon_t$$

escrevendo na notação do operador lag

$$(1 - \phi L)y_t = \varepsilon_t$$

$$y_t = \frac{\varepsilon_t}{(1 - \phi L)}$$

podemos ver que essa expressão tem a forma de um somatório de PG infinita e a única condição para que isso seja verdade é a de que $\phi \leq 1$ para que ela seja convergente. Podemos então dizer que no caso da **série ser estacionária podemos dizer que ela é inversível**, sendo ε_t o primeiro elemento e ϕL a razão. Então é a mesma coisa que dizer que isso é

$$y_t = \varepsilon_t + \phi L \varepsilon_t + \phi^2 L^2 \varepsilon_t + \dots$$

ou ainda

$$y_t = \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \dots$$

$$y_t = \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}$$

Autocorrelação: De acordo com Box et al. (2015, 58), temos que o caso de um AR(1) satisfaz a seguinte condição

$$\rho_k = \phi_1 \rho_{k-1}$$

então para $k = 1$

$$\rho_1 = \phi_1$$

e resolvendo recursivamente

$$\rho_2 = \phi_1 \rho_1 = \phi_1^2$$

$$\rho_3 = \phi_1 \rho_2 = \phi_1 \phi_1^2 = \phi_1^3$$

$$\therefore \rho_k = \phi_1^k$$

Da mesma maneira, podemos ver que a autocorrelação de um **AR(2)** satisfaz a seguinte condição

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}$$

então para $k = 1$

$$\rho_1 = \phi_1 + \phi_2 \rho_{-1}$$

como a correlação é simétrica, podemos trocar $\rho_{-1} \rightarrow \rho_1$

$$\rho_1 = \phi_1 + \phi_2 \rho_1$$

$$\rho_1 = \frac{\phi_1}{1 - \phi_2}$$

e resolvendo recursivamente

$$\rho_2 = \phi_1 \rho_1 + \phi_2 \rho_0$$

$$\rho_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2$$

$$\rho_3 = \phi_1 \rho_2 + \phi_2 \rho_1$$

$$\rho_3 = \phi_1 \left[\frac{\phi_1^2}{1 - \phi_2} + \phi_2 \right] + \frac{\phi_1 \phi_2}{1 - \phi_2}$$

\vdots

*** terminar essa parte mais tarde****

Variância: Uma maneira fácil de se calcular a variância de um processo AR(1) é usando o operador da variância **Var[.]**:

$$\text{Var}[y_t] = \text{Var}[\phi y_{t-1}] + \text{Var}[\varepsilon_t]$$

$$\text{Var}[y_t] = \phi^2 \text{Var}[y_t] + \text{Var}[\varepsilon_t]$$

$$\text{Var}[y_t] = \phi^2 \text{Var}[y_t] + \sigma_\varepsilon^2$$

$$(1 - \phi^2) \text{Var}[y_t] = \sigma_\varepsilon^2$$

$$\sigma_y^2 = \frac{\sigma_\varepsilon^2}{(1 - \phi^2)}$$

se vc observar o valor de $2 + 2$ a variância é negativa, oq seria um absurdo; se $2 + 2$ a variância vai para infinito e é por isso que quando o processo é não estacionário ele apresenta variância infinita.

EXEMPLO: Seja a modelo $y_t = -0,3y_{t-1} + 0,6y_{t-2} + \varepsilon_t$, verifique se é estacionário.

Podemos reescrever esse modelo usando o operador lag

$$(1 + 0,3L - 0,6L^2)y_t = \varepsilon_t$$

temos então aqui nesse exemplo que $\phi(L) = 1 + 0,3L - 0,6L^2$ e como a condição para que o processo seja estacionário as raízes de $\phi(L) = 0$ tem que ser maior que 1 em módulo. Vamos ter que

$$\begin{aligned} L &= \frac{-0,3 \pm \sqrt{0,3^2 - 4(-0,6)}}{2(-0,6)} \\ &= \frac{-0,3 \pm 1,578}{-1,2} \Rightarrow L = -1,065 \text{ e } 1,565 \end{aligned}$$

sendo todas as raízes maiores que um em modulo, então o processo é estacionário.

3.1.2 Modelo de Média móvel (MA)

O modelo de média móvel é uma extensão do processo de ruído branco, onde os parâmetros de peso são representados pelo símbolo θ . O modelo é escrito da seguinte forma:

$$y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}$$

O motivo pelo qual se usa esse tipo de modelo é que, em economia, estamos diante de diversos choques ao longo do tempo, seja por uma greve, desastres naturais, decisões do governo, etc. Esses choques normalmente não terão apenas um efeito imediato na série, muitas vezes apresentando um efeito contemporâneo, demorando a ser dissipado e tendo efeito ao longo do tempo em vários períodos consecutivos.

Seja o modelo de MA(q) com $\mu = 0$ com esperança

$$\begin{aligned} \mathbb{E}[y_t] &= \mathbb{E}[\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}] \\ &= 0 \end{aligned}$$

e variância

$$\begin{aligned} \text{Var}[y_t] &= \text{Var}[\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}] \\ &= \text{Var}[\varepsilon_t] + \text{Var}[\theta_1\varepsilon_{t-1}] + \text{Var}[\theta_2\varepsilon_{t-2}] + \cdots + \text{Var}[\theta_q\varepsilon_{t-q}] \\ &= \sigma_\varepsilon^2 + \theta_1^2\sigma_\varepsilon^2 + \theta_2^2\sigma_\varepsilon^2 + \cdots + \theta_q^2\sigma_\varepsilon^2 \\ &= (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)\sigma_\varepsilon^2 \end{aligned}$$

nota-se então que a variância de um modelo MA(2) é igual a $\text{Var}[y_t] = (1 + \theta_1^2 + \theta_2^2)\sigma_\varepsilon^2$.

Invertibilidade: As condições de invertibilidade do modelo são iguais às condições para que um modelo AR(p) seja estacionário, ou seja, $|\theta| < 1$, ou que as raízes de $\theta(L) = 0$ sejam maiores que 1 em módulo.

Função de autocorrelação: Vamos ter a seguinte fórmula para o caso geral de um modelo MA(q),

$$\rho_k = \begin{cases} \frac{-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2} & k = 1, 2, \dots, q \\ 0 & k > q \end{cases}$$

onde temos que $\gamma_0 = (1 + \theta_1^2 + \dots + \theta_q^2)\sigma_\varepsilon^2$ e que $\gamma_k = (-\theta_k + \theta_1\theta_{k+1} + \dots + \theta_{q-k}\theta_q)\sigma_\varepsilon^2$.

Uma coisa a se notar nessa equação é que, quando tivermos que o número de lags da função de autocorrelação for maior que a ordem do modelo, a FAC será igual a zero. Seja um modelo MA(2), então $\rho_3, \dots, \rho_k = 0$.

Essa equação pode ser aplicada para modelos MA(1) também:

$$\rho_k = \begin{cases} \frac{-\theta_1}{1 + \theta_1^2} & k = 1 \\ 0 & k > 1 \end{cases}$$

De forma análoga, para um MA(2) vamos ter:

$$\rho_1 = \frac{-\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}$$

$$\rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}$$

3.2 Modelos lineares não-estacionários

A maioria das séries temporais apresenta não estacionariedade, o que significa que em um modelo ARMA, se as raízes de $\phi(L) = 0$ estiverem fora do círculo unitário, a série é dita estacionária, mas quando estão fora do círculo unitário, ela terá um comportamento explosivo, sendo não estacionária. Um modelo ARMA sem sazonalidade pode ser escrito da seguinte forma:

$$\phi(L)y_t = \theta(L)\varepsilon_t$$

Se $\phi(L)$ for um operador autorregressivo não estacionário, tal que d raízes de $\phi(L) = 0$ são unitárias e o resto está fora do círculo unitário, esse modelo pode ser escrito como:

$$\phi(L)(1 - L)^d y_t = \theta(L)\varepsilon_t$$

onde $\phi(L)$ se torna um operador autorregressivo estacionário. Nesse caso, foi feita a separação de todas as raízes que estão fora do círculo unitário das raízes que estão dentro do círculo unitário. Sabendo que a diferenciação torna as raízes estacionárias, podemos usar o operador $\nabla^d = (1 - L)^d$ para $d \geq 1$; aqui ∇ é chamado de operador de diferenciação.

$$\phi(L)\nabla^d y_t = \theta(L)\varepsilon_t$$

Esse processo será conhecido como integração, o I do ARIMA(p, d, q), onde a ordem de integração desse processo é dada pelo valor de d . E se chamarmos $w_t = \nabla^d y_t$, o processo passa a ser escrito da seguinte forma:

$$\phi(L)w_t = \theta(L)\varepsilon_t$$

Abaixo podemos ver alguns exemplos de processo ARIMA:

1. O modelo ARIMA(0,1,1):

$$\begin{aligned}\nabla y_t &= \varepsilon_t - \theta_1 \varepsilon_{t-1} \\ &= (1 - \theta_1 L) \varepsilon_t\end{aligned}$$

onde $p = 0$, $d = 1$, $q = 1$, $\phi(L) = 1$, $\theta(L) = (1 - \theta_1 L)$.

2. O modelo ARIMA(1,1,1):

$$\begin{aligned}\nabla y_t - \phi_1 \nabla y_{t-1} &= \varepsilon_t - \theta_1 \varepsilon_{t-1} \\ (1 - \phi_1 L) \nabla y_t &= (1 - \theta_1 L) \varepsilon_t\end{aligned}$$

onde $p = 1$, $d = 1$, $q = 1$, $\phi(L) = (1 - \phi_1 L)$, $\theta(L) = (1 - \theta_1 L)$.

EXAMPLE: Seja a modelo $y_t = 0,1y_{t-1} + 0,9y_{t-2} + \varepsilon_t$, verifique se é estacionário.

Reescrevendo usando o operador lag:

$$\begin{aligned}(1 - 0,1L - 0,9L^2)y_t &= \varepsilon_t \\ (0,9)(1,1111 - 0,1111L - L^2)y_t &= \varepsilon_t\end{aligned}$$

Usando a fórmula quadrática, vamos encontrar que as raízes de $\phi(L)$ são -1,1111 e 1; como uma das raízes é igual a 1, esse processo é não estacionário. Reescrevendo, temos:

$$\begin{aligned}(0,9)(L + 1,1111)(L - 1)y_t &= \varepsilon_t \\ (1 + 0,9L)(L - 1)y_t &= \varepsilon_t \\ (1 + 0,9L)(1 - L)y_t &= \varepsilon_t \\ (1 + 0,9L)\nabla y_t &= \varepsilon_t\end{aligned}$$

Assim sendo um modelo ARIMA(1,1,1) pois temos ∇y_t que corresponde à integração de ordem 1 (I(1)). Também pode ser reescrito da seguinte maneira:

$$(1 + 0,9L)w_t = \varepsilon_t$$

Mas note, que como aqui não temos o operador de diferenciação, este modelo é um ARMA(1,1).

4 Identificação do Modelo

Uma das formas de se identificar qual o melhor modelo usar é por meio dos critérios de informação de Akaike (AIC) ou o critério de informação Schwartz (*Bayesian Information Criterion* - BIC). A ideia por trás do critério de informação é tentar mensurar a qualidade de um modelo de acordo com a informação que ele carrega e também eles prezam pela simplicidade, quer dizer que quanto mais parâmetros e mais lags for adicionado, mais penalizado será o modelo de acordo com o critério de informação, prezando sempre por um modelo mais parcimonioso.

O ideal é que o AIC e BIC sejam o menor possível (vale notar que pode ser negativo). Quanto melhor for o ajuste do modelo, mais AIC e BIC irão se aproximar de $-\infty$. O critério de informação pode ser usado para comparar dois modelos; Um modelo A pode ser dito melhor que o B, se o critério de informação (AIC ou BIC) do A for menor que o do B, lembrando que para essa comparação, o ideal é que os modelos tenham o mesmo tamanho de amostra. São calculados da seguinte maneira:

$$\begin{aligned} \text{AIC} &= \frac{-2 \ln(L)}{T} + \frac{2n}{T} \\ \text{BIC} &= \frac{-2 \ln(L)}{T} + \frac{n \ln(T)}{T} \end{aligned}$$

Onde $n = p + q + 1$ que é a soma dos parâmetros do modelo; T é o tamanho da amostra; L é o valor máximo da função de máxima verossimilhança.

Note que a diferença entre eles é que no BIC vamos ter $\ln(T)$ ao invés do 2 multiplicando o n e como $\ln(T)$ sempre será maior que 2, quer dizer que o BIC é mais parcimonioso que o AIC; o custo marginal de se adicionar regressores é maior no BIC do que no AIC.

5 Testes do Modelo

5.1 Testes de Raiz Unitária - Dick-Fuller

O teste de Dick-Fuller é o teste mais usado para detectar a presença de raiz unitária, na qual a hipótese nula é se a série é um passeio aleatório, contra a hipótese alternativa se a série é estacionária. Para mostrar o teste, vamos usar um AR(1),

$$y_t = \phi y_{t-1} + \varepsilon_t, \quad \text{onde } \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2)$$

onde já sabemos que se $\phi = 1$ o processo é um passeio aleatório e se $\phi < 1$ ele é um processo estacionário. Aqui vamos ter que esse ϕ é um parâmetro estimado por MQO, por isso precisamos aplicar um teste a esse parâmetro para saber o quão confiável é esse coeficiente que foi estimado, para isso, Dick-Fuller aplicaram um teste de regressão que é derivado desse modelo AR(1), subtraindo y_{t-1} de ambos os lados:

$$\begin{aligned} y_t - y_{t-1} &= \phi y_{t-1} - y_{t-1} + \varepsilon_t \\ \nabla y_t &= (\phi - 1)y_{t-1} + \varepsilon_t \end{aligned}$$

Substituindo $\delta = \phi - 1$, temos:

$$\nabla y_t = \delta y_{t-1} + \varepsilon_t$$

Usando essa equação, pode-se investigar que se $\phi = 1 \Rightarrow \delta = 0$ e isso implica que $\nabla y_t = \varepsilon_t$ sendo um passeio aleatório; caso contrário, se $\phi < 1 \Rightarrow \delta \neq 0$ sendo um processo estacionário. Então a ideia desse teste de DF é olhar para o δ , onde:

$$H_0 : \delta = 0$$

Se a hipótese nula não for rejeitada (satisfeita ou aceita), vamos ter que y_t é integrado de ordem 1, tal que $y_t \sim I(1)$, pois será estacionário após uma defasagem. A hipótese alternativa tem a seguinte forma:

$$H_a : \delta \neq 0$$

Então se a hipótese nula for rejeitada, implica que $y_t \sim I(1)$, sendo a série estacionária. O interessante aqui é que com essa regressão de teste, conseguimos aplicar uma estatística de teste (τ) em δ , podendo assim comparar essa estatística do modelo com valores já calculados por Dick-Fuller, chamados de τ_c (tau calculado), assim: - Se $\tau < \tau_c \Rightarrow$ rejeita H_0 (não estacionário) - Se $\tau > \tau_c \Rightarrow$ **NÃO** rejeita H_0 (estacionário)

Esse valor de teste é um simples teste t , onde:

$$\hat{t}_{DF} = \frac{\hat{\delta}}{SE(\hat{\delta})} = \frac{\phi - 1}{SE(\phi)}$$

SE corresponde ao *standard error* (erro padrão).

Esse teste de DF pode ser modificado para caso a série tenha um drift, adicionando uma constante e um termo de tendência, assumindo a seguinte forma

$$\nabla y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \varepsilon_t$$

a hipótese nula permanece a mesma. Para saber se essa constante e o termo de tendência são significantes, temos que olhar a estatística de teste t de cada beta, ou olhar para seus *p-valores*.

Esse teste de Dick-Fuller é simples e não permite testar a persistência do processo, pois os resíduos podem ser autocorrelacionados. Assim, foi desenvolvido o teste de **Dick-Fuller Aumentado** (ADF) para controlar essa autocorrelação nos resíduos, possibilitando testar valores de lag maiores (AR(.) de ordem superior). Vamos considerar a regressão de teste para um AR(1)

$$\nabla y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \varepsilon_t$$

se acharmos que ∇y_t segue um AR(2), a regressão de teste será

$$\nabla y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \gamma \nabla y_{t-1} + \varepsilon_t$$

generalizando isso para um processo AR(p), temos

$$\nabla y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \sum_{i=0}^m \gamma_i \nabla y_{t-i} + \varepsilon_t$$

Aqui, a hipótese é a mesma, sendo $H_0 : \delta = 0$ não estacionário e $H_a : \delta < 0$ estacionário. A significância dos γ pode ser testada através de um teste t .

5.1.1 Teste de autocorrelação - Ljung-Box e Box-Pierce

O teste de Ljung-Box pode ser definido como

- H_0 : os erros são independentemente distribuídos
- H_A : os erros **NÃO** são independentemente distribuídos; existe correlação serial

Não há muito mais o que comentar nesse teste, apenas mostrar que a estatística de teste do Ljung-Box é

$$LB(k) = N(N+2) \sum_{j=1}^k (N-j)^{-1} \hat{\rho}_j^2(\hat{\varepsilon})$$

e que o teste Q (Box-Pierce) é

$$Q(k) = N \sum_{j=1}^k \hat{\rho}_j^2(\hat{\varepsilon})$$

onde ρ_j é a correlação dos resíduos ao lag j e N é o tamanho da amostra. Como aqui as estatísticas de teste se distribuem como uma qui-quadrado com $k - p - q$. Rejeita-se H_0 se

$$Q, LB > \tau$$

sendo $\tau = \chi_{\alpha}^2(k - p - q)$, em que α é o nível de significância.

5.1.2 Teste de heterocedasticidade - ARCH-LM

O teste **ARCH lagrange multiplier** ou **ARCH-LM** é usado para detectar heterocedasticidade e é bem simples. Ele testa se os erros da regressão são variantes no tempo sem de fato estimar um parâmetro *ARCH*. A estimação desse teste pode seguir a seguinte metodologia em dois passos:

1. Estimar via MQO a equação da regressão mais adequada ou usar um modelo ARMA e assumir que $\hat{\varepsilon}_t^2$ seja o erro quadrático ajustado.
2. Regressar os erros quadráticos contra uma constante e com q lags desses valores $\hat{\varepsilon}_{t-1}^2, \hat{\varepsilon}_{t-2}^2, \dots, \hat{\varepsilon}_{t-q}^2$, estimando a seguinte regressão

$$\hat{\varepsilon}_t^2 = \alpha_0 + \alpha_1 \hat{\varepsilon}_{t-1}^2 + \alpha_2 \hat{\varepsilon}_{t-2}^2 + \dots + \alpha_q \hat{\varepsilon}_{t-q}^2$$

Se não houver efeito ARCH ou GARCH, os valores estimados de α_1 até α_q têm que ser zero. Essa regressão apresentará pouco poder explicatório, o que significa que o R^2 será pequeno. Usando um tamanho de amostra T , a estatística de teste TR^2 (vale ressaltar que isso é um produto) converge para uma distribuição χ^2 com q graus de liberdade. Por isso, esse tipo de teste entra na categoria de testes assintóticos, que dependem do tamanho da amostra. Se a amostra for muito pequena, o valor do teste TR^2 será superestimado, mas para amostras grandes, ele funciona bem. A hipótese nula aqui é $H_0 : \varepsilon_t$ sem efeito ARCH (**sem heterocedasticidade**); $H_a : \varepsilon_t$ tem efeito ARCH (**tem heterocedasticidade**). A hipótese nula é rejeitada se $TR^2 > \tau$.

5.1.3 Teste de normalidade - Jarque-Bera

Difícilmente dados empíricos passam no teste de normalidade dos dados. O mais famoso teste para testar se os dados estão normalmente distribuídos é o teste de **Jarque-Bera**.
CONTINUAR

6 Modelos de volatilidade

6.1 Modelos lineares

6.1.1 ARCH

Seja o modelo

$$y_t = \mathbb{E}[y_t | I_{t-1}] + \nu_t$$

onde $\nu_t \sim i.i.d(0, h_t)$ representa o erro do modelo, com média zero e uma variância h_t que é condicional ao conjunto de informações passadas $I_{t-1} = \{y_1, x_1, \dots, y_{t-1}, x_{t-1}\}$ e ν_t pode ser separado em duas partes, uma é a parte estocástica $\varepsilon_t \sim i.i.d(0, 1)$ e a outra é o desvio padrão condicional que depende do tempo $h_t^{\frac{1}{2}}$. Logo, a variância de ν_t é

$$\mathbb{E}[\nu_t^2 | I_{t-1}] = h_t$$

Assumindo que a média é zero, podemos escrever o modelo ARCH básico da seguinte forma

$$y_t = h_t^{\frac{1}{2}} \varepsilon_t$$

como h é a variância condicional ao conjunto de informação, ele é definido como

$$h_t = w + \alpha y_{t-1}^2$$

onde ele depende de uma constante e do quadrado da informação passada.

Note que, como está ao quadrado a informação passada, o modelo ARCH, que é da classe dos lineares, **atribui peso igual para volatilidade tanto negativa quanto positiva**. A ideia central desse tipo de modelo é conseguir separar a volatilidade condicional do ruído branco e modelar essa volatilidade condicional h_t .

Modelo AR-ARCH: Como aqui nos modelos de volatilidade nós estamos considerando o y_t como uma série de retornos, pode acontecer de que os retornos sejam correlacionados, ainda que fracamente, sendo então adicionado um AR(1) ao modelo

$$\begin{aligned} y_t &= \phi y_{t-1} + h_t^{\frac{1}{2}} \varepsilon_t \\ h_t &= w + \alpha y_{t-1}^2 \end{aligned}$$

em que ϕ mede a correlação entre os retornos no tempo t e $t - 1$.

6.1.2 GARCH

Esse modelo é uma extensão do modelo ARCH. Aqui, ele inclui dentro da variância sua própria defasagem. Podemos escrever o modelo GARCH(1,1) da seguinte maneira

$$\begin{aligned}y_t &= \phi y_{t-1} + h_t^{\frac{1}{2}} \varepsilon_t \\h_t &= w + \alpha y_{t-1}^2 + \beta h_{t-1}\end{aligned}$$

e um modelo GARCH(p,q) é dado por

$$h_t = w + \sum_{i=1}^q \alpha_i y_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i}$$

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 Analisis». 2019. <https://www.economodel.com/time-series-analysis-2019>.