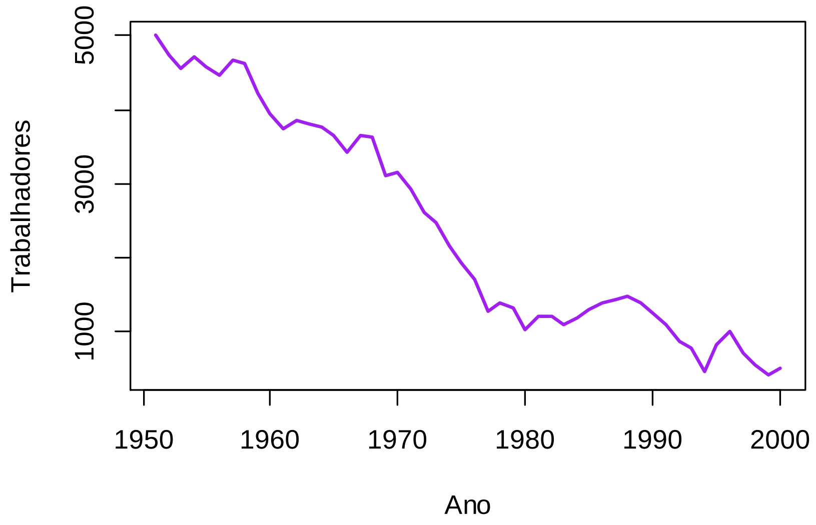
Econometria e Séries Temporais
- Aula 2 -

Prof. Mestre. Omar Barroso

Instituto Brasileiro de Educação, Pesquisa e Desenvolvimento

Lembram do último exemplo da aula anterior?

- $Employment = \alpha + \beta_1 Employment_{\boldsymbol{t-1}} + \varepsilon_t$
- Iniciamos a série com um total de 5.000 trabalhadores e simulamos a redução do emprego com um **processo autorregressivo** que apresenta um movimento descendente no longo prazo e tem erros normalmente distribuídos.
- É evidente que as observações sobre o número de empregados **não podem ser independentes neste exemplo:** o nível de emprego de hoje está **correlacionado** com o nível de emprego de amanhã. Assim, o **i.i.d.** suposição é violada (**discutiremos isso mais a diante**).



Uma série temporal não é iid!

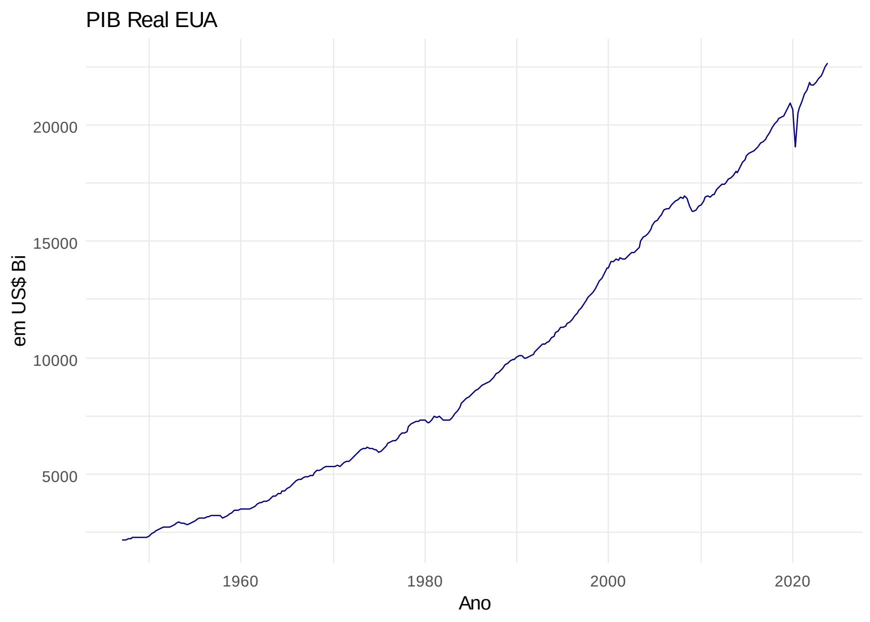
- *Razões*
- **Dependência temporal:** os pontos de dados de séries temporais são normalmente coletados em intervalos de tempo consecutivos. O valor de uma variável em um determinado momento está frequentemente relacionado aos seus valores passados. Esta dependência temporal viola a suposição de independência do i.i.d. dados.
- **Sazonalidade:** Muitas séries temporais econômicas apresentam padrões sazonais, tais como o aumento das vendas durante determinados meses ou trimestres devido a feriados ou condições meteorológicas. Estes padrões recorrentes introduzem dependência entre as observações dentro de cada ciclo sazonal. Por exemplo, os ciclos sazonais de produtos agrícolas como frutas e vegetais que crescem durante uma certa estação. Nesse mesmo contexto, os períodos de aumento de vendas como nas festas de fim de ano.
- **Tendências:** As séries cronológicas econômicas apresentam frequentemente tendências de longo prazo, tais como o aumento do PIB ao longo do tempo ou tendências inflacionárias. Essas tendências criam dependências entre as observações ao longo do tempo.

- **Autocorrelação:** Autocorrelação refere-se à correlação entre uma variável e seus valores passados em diferentes intervalos de tempo. A autocorrelação é comumente observada em séries temporais econômicas, indicando que os valores atuais estão relacionados aos valores passados. Por exemplo, preços/retornos de ações em mercados de capitais e taxa de crescimento do PIB. Isso viola a suposição de independência no i.i.d. dados.
- **Rupturas Estruturais:** *Mudanças* nas políticas econômicas, avanços tecnológicos ou outras mudanças estruturais podem levar a mudanças abruptas nos dados das séries temporais. Estas quebras estruturais violam a suposição de distribuição idêntica ao longo dos períodos de tempo. Por exemplo, como a pandemia do Covid-19 afetou o crescimento econômico mundial.
- **Heterocedasticidade:** Heterocedasticidade se refere ao fenômeno onde a variabilidade de uma variável muda ao longo do tempo. As séries temporais económicas apresentam frequentemente **níveis variáveis de volatilidade**, violando a suposição de variância constante em i.i.d. dados.

- Por essas razões, os modelos econométricos para dados de séries temporais precisam levar em conta as características específicas dos dados, como **dependência temporal, sazonalidade, tendências, autocorrelação e quebras estruturais**. **Ignorar essas características pode levar a estimativas de parâmetros tendenciosas e inferências não confiáveis.** Modelos como média móvel integrada autoregressiva (**ARIMA**), heterocedasticidade condicional autoregressiva (**ARCH**) e heterocedasticidade condicional autoregressiva generalizada (**GARCH**) são comumente usados em econometria para abordar essas questões e **modelar a dinâmica** dos dados de séries temporais de forma eficaz.
- *Vamos rever esses conceitos com mais detalhes nas próximas aulas...*

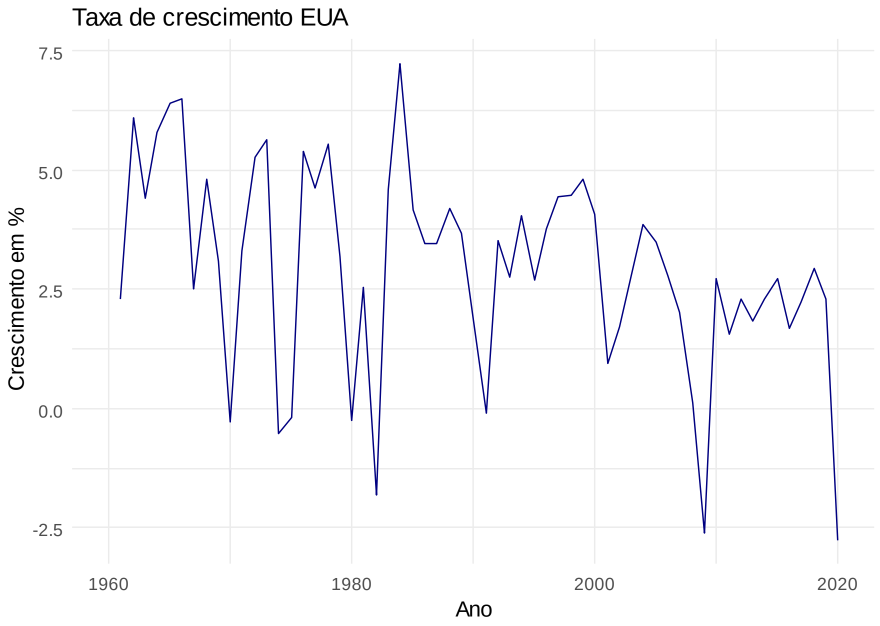
Vamos construir nossa série temporal do PIB!

- O PIB é definido como o valor total dos bens e serviços produzidos durante um determinado período de tempo.
- No exemplo ao lado estamos estimando o PIB Real estado-unidense em escala trimestral ao longo dos anos.



PIB EUA parte 2

- Agora vamos ver a taxa de crescimento do PIB ao longo dos anos.



Defasagens (lags), primeiras diferenças, logaritmos e taxas de crescimento

- Valores anteriores de uma série temporal são chamados de defasagens ou “lags”. O primeiro lag de y_t é y_{t-1} .
- Também podemos trabalhar com diferenças de série. A primeira diferença de uma série é $\Delta y_t = y_t - y_{t-1}$. Isso é equivalente a diferença entre t e t-1.
- Em alguns casos, pode ser eficiente utilizarmos o logaritmo quando trabalhando com diferenças. Nesse contexto, também podemos multiplicar o logaritmo por 100 para aproximar a mudança percentual da diferença calculada.

Date	Value	$Value_{t-1}$	$Value_{t-2}$
1/1/2017	200	NA	NA
1/2/2017	220	200	NA
1/3/2017	215	220	200
1/4/2017	230	215	220
1/5/2017	235	230	215
1/6/2017	225	235	230
1/7/2017	220	225	235
1/8/2017	225	220	225
1/9/2017	240	225	220
1/10/2017	245	240	225

Fonte: Business Science

Autocorrelação e autocovariância

- Como em séries temporais, os valores do passado (y_{t-j}) são relacionados com o do tempo mais atual (y_t), utilizamos a autocovariância e autocorrelação de ambos para analisar padrões.
- a covariância entre Yt e o *lag* n, é chamado de a "j" auto-covariância da série y_t .
- No qual, o coeficiente de correlação "j", também pode ser chamado como **coeficiente de correlação serial (CCS)**. O CCS mede a correlação entre y_t e y_{t-j} .

$$Covariância\; n = Cov(y_t, y_{t-j})$$

$$Autocorrelação = \rho_j = \rho y_t, y_{t-j} = \frac{Cov(y_t, y_{t-j})}{\sqrt{var(y_t) var(y_{t-j})}}$$

Autocorrelação e autocovariância

- A autocovariância e autocorrelação da população podem ser estimadas como: $\widehat{Cov(y_t, y_{t-n})}$, a covariância amostral e $\widehat{\rho_n}$ como a correlação amostral. Podemos denominar da determinada maneira:
- Autocovariância de n ou “j”:

$$Cov(Y_t, Y_{t-j}),$$

Autocorrelação:

$$\rho_j = \rho_{Y_t, Y_{t-j}} = \frac{Cov(Y_t, Y_{t-j})}{\sqrt{Var(Y_t) Var(Y_{t-j})}}$$

Autocorrelação e autocovariância

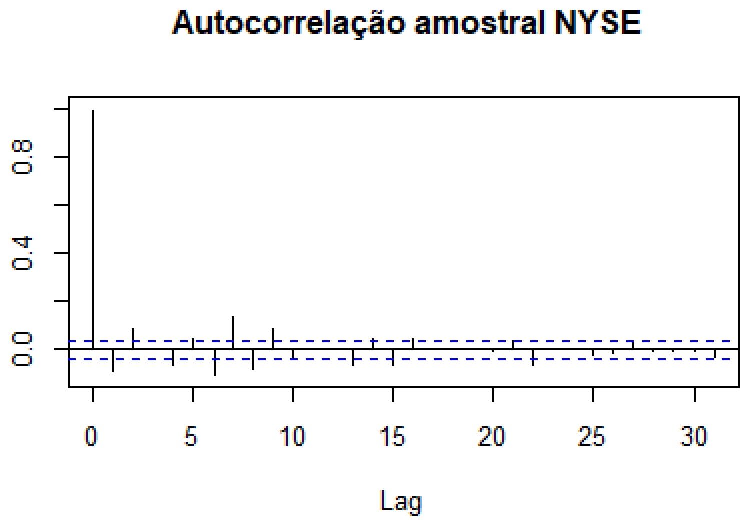
- Olhando para a autocorrelação amostral temos padrões parecidos:
- No qual, $\overline{Y_{j+1:T}}$ denota a média de $Y_{j+1} + Y_{j+2} + \cdots + Y_{j+n}$

$$Cov\widehat{(Y_t, Y_{t-j})} = \frac{1}{T}\sum_{t=j+1}^{T}(Y_t - \overline{Y}_{j+1:T})(Y_{t-j} - \overline{Y}_{1:T-j}),$$

$$\hat{\rho}_j = \frac{Cov\widehat{(Y_t, Y_{t-j})}}{\widehat{Var(Y_t)}}$$

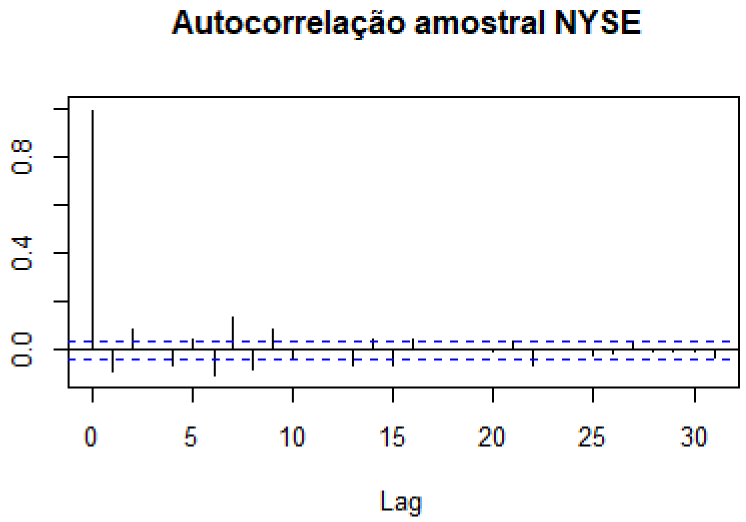
Testando a Autocorrelação da NYSE

- Um gráfico de autocorrelação é projetado para mostrar se os elementos de uma série temporal estão: positivamente correlacionados, negativamente correlacionados ou são independentes entre si.
- A função da autocorrelação (ACF *do inglês Autocorrelation Function*) no eixo vertical, varia de -1 para 1. Enquanto isso, o eixo horizontal mostra o tamanho das defasagens (lags) entre os elementos da série de tempo.
- Por exemplo, a autocorrelação com defasagem 2 é a correlação entre os elementos da série temporal e os elementos correspondentes que foram observados dois períodos anteriores.



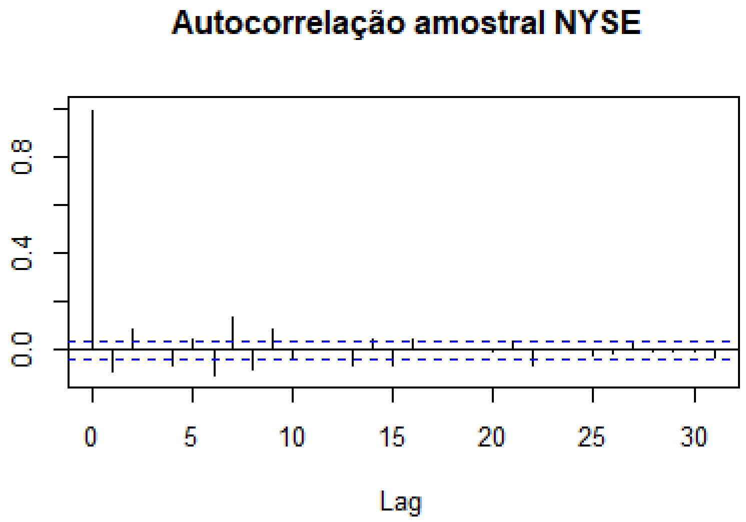
Testando a Autocorrelação da NYSE

- No gráfico, há uma linha vertical (um “pico”) correspondente a cada defasagem. A altura de cada pico mostra o valor da função de autocorrelação para cada defasagem.
- A autocorrelação com defasagem zero é sempre igual a 1, pois representa a autocorrelação entre cada termo e ele mesmo.
- Cada pico que sobe acima ou cai abaixo das linhas tracejadas (em azul) é considerado estatisticamente significativo (em um nível de 5%).
- Isso significa que o pico tem um valor significativamente diferente de zero. Se um pico for significativamente diferente de zero, isso é evidência de autocorrelação. Um pico próximo de zero é uma evidência contra a autocorrelação.



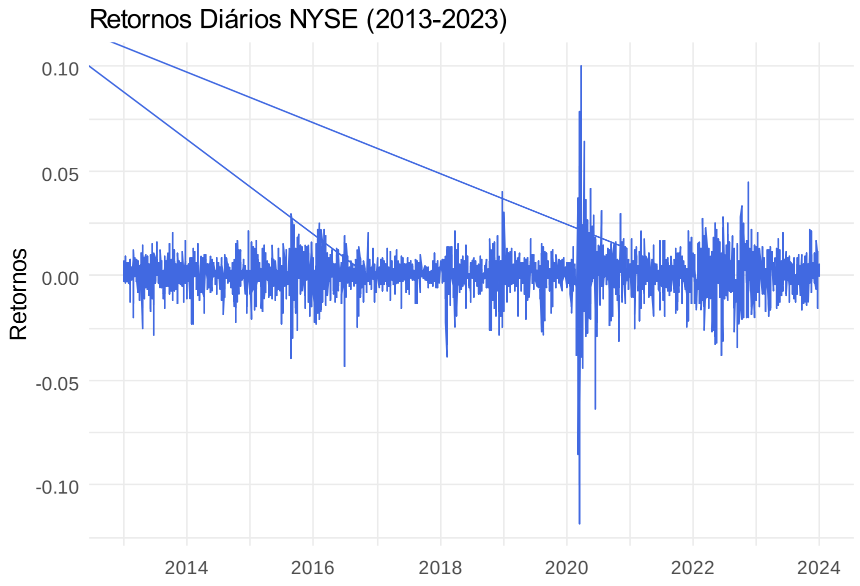
Testando a Autocorrelação da NYSE

- Podemos observar que a maioria dos picos ficam acima da linha tracejada. Ou seja, a maioria dos picos são estatisticamente significativos. Neste caso, os retornos são altamente correlacionados.
- O gráfico de autocorrelação determina se os elementos de uma série temporal são aleatórios (ou seja, se eventos aleatórios do passado afetam o momento futuro). Isto é importante porque muitos testes estatísticos envolvendo séries temporais baseiam-se nesta suposição.
- Além disso, a série NYSE, exibe o que os econometristas chamam de agrupamento de volatilidade (volatility clustering): há períodos de alta e períodos de baixa variância. Isso é comum para muitas séries temporais financeiras.



Observando os dados da bolsa de valores (NYSE)

- Usando o R podemos baixar os dados da NYSE e calcular os retornos.
- Perceba essa figura [*memorizem]*, ela segue um **padrão estacionário**.
- A série de tempo dos retornos da NYSE aparenta flutuar aleatoriamente. Todavia, esse padrão converge a zero, no qual essa é a média.
- Porém, podemos sugerir que uma série temporal segue um padrão estacionário apenas pelo padrão visual? Sim e não...



Retornos NYSE

- Dentro da função escrita no R:
- returns <- dailyReturn(NYA)
- A operação que determina o gráfico anterior [retornos], é determinada da seguinte forma:
- $Retornos\ Diários_t = \frac{P_t - P_{t-1}}{P_{t-1}}$
- $Retornos\ Diários_t$: Retornos diários no dia t.
- P_t : Preço da Ação ao final do dia t.
- P_{t-1}: $Preço\ da\ Ação\ ao\ final\ do\ dia\ t-1$

Processo Estocástico

- *Relembrem o conceito do termo de erro ε discutido na aula 1.*
- **Definição:** Um processo estocástico em um tempo discreto ou processo de série temporal é uma sequência de variáveis aleatórias't' indexadas sobre um período de tempo.

$$\{\ldots, Y_1, Y_2, \ldots, Y_t, Y_{t+1}, \ldots\} = \{Y_t\}_{t=-\infty}^{\infty}$$

- **Índice de tempo:** O índice de tempo é geralmente espaçado regularmente (por exemplo, dias, meses, anos), mas também pode ser espaçado irregularmente (por exemplo, tempos de transação intradiários).

Processo Estocástico

- **Importância da ordenação:** A ordenação imposta pelo **índice de tempo** é crucial na modelagem de dados de séries temporais para capturar relacionamentos temporais entre variáveis aleatórias.

$$\{Y_1 = y_1, Y_2 = y_2, \ldots, Y_T = y_T\} = \{y_t\}_{t=1}^{T}$$

- **Comparação com amostragem aleatória:** Na amostragem aleatória de uma população, a ordenação de variáveis aleatórias não importa porque elas são independentes.

Processo Estocástico

- **Objetivo da Modelagem de Séries Temporais:** Descrever o comportamento probabilístico do processo estocástico subjacente gerando os dados observados de forma concisa.

- **Estimativa:** Usar a amostra observada para estimar características importantes de um modelo de série temporal, como medidas de dependência temporal.

- **Suposições:** Fazer uma série de suposições sobre o comportamento conjunto das variáveis aleatórias no processo estocástico.

- **Tratamento do Processo Estocástico:** Tratar o processo estocástico de forma semelhante a como uma amostra aleatória de uma determinada população é tratada.

Processo Estocástico Estacionário

- Em muitos contextos, esperaríamos alguma dependência entre variáveis aleatórias próximas (por exemplo, X1 e X2). Ou seja, o requerimento de variáveis independente, identicamente distribuídas (iid) não precisa ser cumprido.
- Todavia, isso não quer dizer que não existe dependência entre variáveis aleatórias que se encontram distantes no tempo (por exemplo, X1 e X100). Podemos permitir esse tipo de comportamento usando os conceitos de **estacionariedade** e ergodicidade.
- Começamos com a definição de estacionariedade estrita.

Estacionariedade Estrita

- Invariância temporal: Em um processo estocástico estritamente estacionário, a distribuição conjunta de variáveis aleatórias é invariante no tempo.
- Exemplo: A distribuição conjunta de (Y1, Y5, Y7) é a mesma que a distribuição de (Y12, Y16, Y18).
- Distribuição Marginal: Todas as variáveis aleatórias individuais Yt têm a mesma distribuição Fy, implicando a mesma média, variância e Desvio Padrão (se essas quantidades existirem).

Estacionariedade Estrita

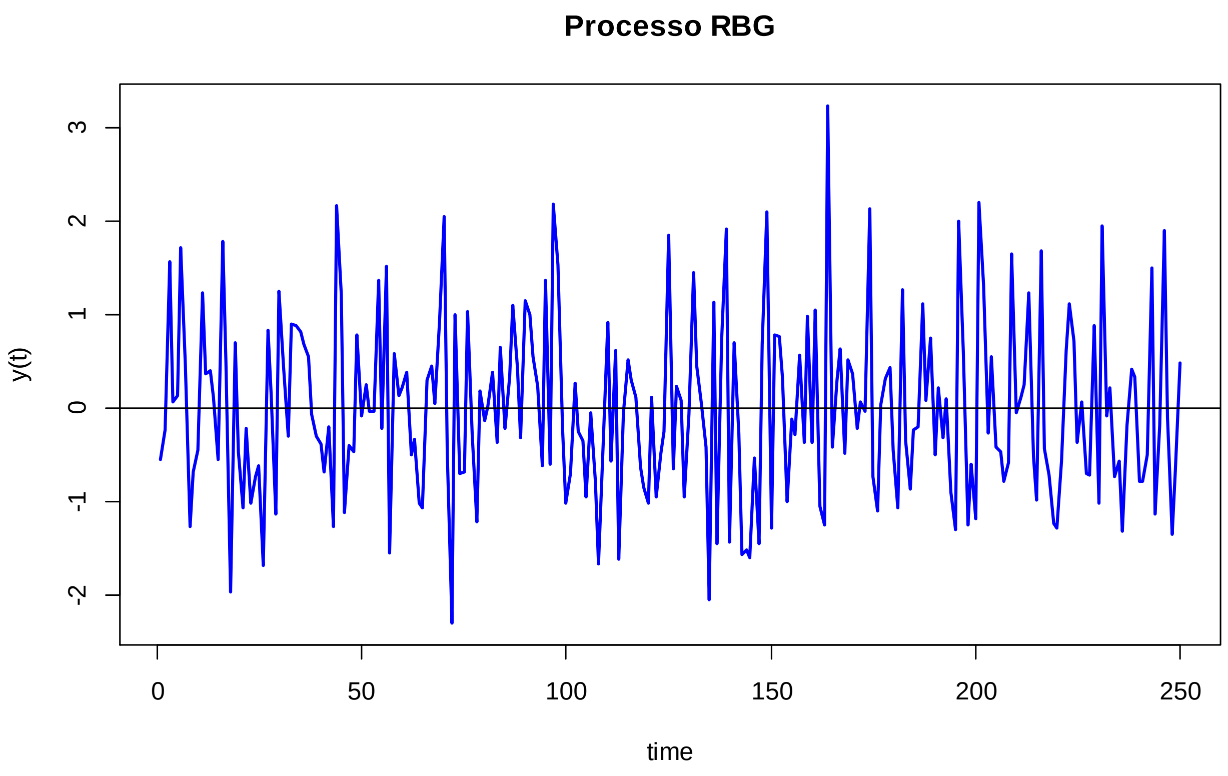
- Correlação: Um processo estocástico estritamente estacionário não assume correlações específicas entre variáveis, exceto que a correlação depende na diferença temporal (t-tr), e não especificamente entre tempo t e tr.
- Dependência Temporal: A estacionariedade estrita permite dependência temporal geral entre as variáveis aleatórias.
- Preservação da Transformação: A estacionariedade estrita é preservada sob transformações gerais.

Proposições

- Dado que {Yt} é estritamente estacionário (ES) e g(.) é uma função dos elementos em {Yt}, assim {g(Yt)} também é estritamente estacionário.
- *Exemplo*
- Se {Yt} é ES, então $\{y_t^2\}$ e $\{y_t, y_{t-1}\}$ também são ES.

Ruído Branco Gaussiano (RBG)

- Dado que $y_t \sim N(0, \sigma^2)$ asiim {Yt} é um RBG denominado como $y_t \sim RBG(0, \sigma^2)$.
- Assim,
- $E[y_t] = 0$ é independente de t;
- $Var(y_t) = \sigma^2$é independente de t;
- $cov\left(y_t, y_{t-j}\right) = 0 \ (no\ qual,\ \ j > 0) e\ indpendente\ de\ t \ \forall\ j$
- Tal que, {Yt} satisfaz as propriedades de um processo estacionário de covariância. A característica definidora de um processo RBG é a ausência de qualquer padrão previsível ao longo do tempo nos valores realizados do processo.

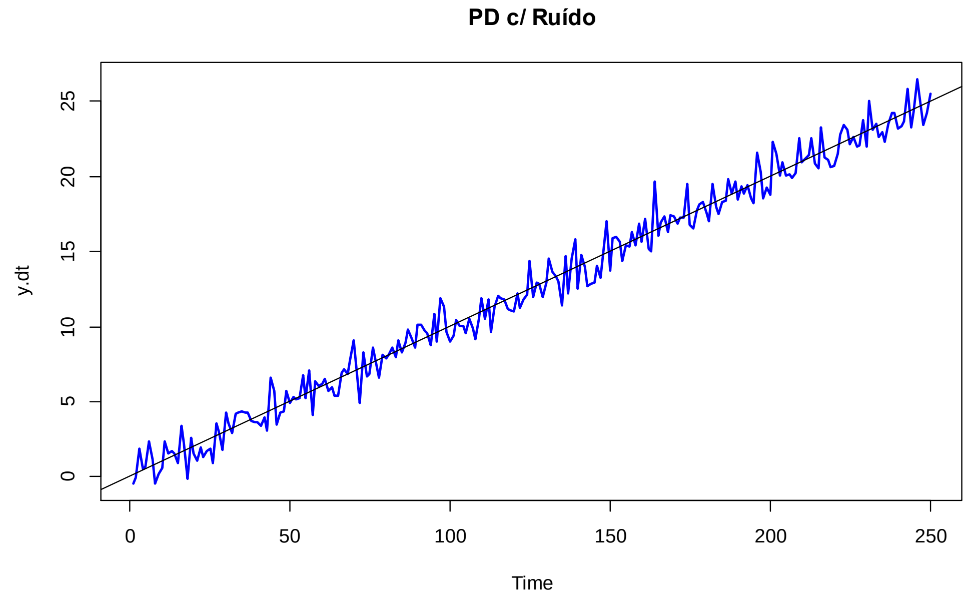

Processo RBG
y(t)
3
2
1
0
-1
-2
0
50
100
150
200
250
time

Não-Estacionariedade (NE)

- O conceito de estacionariedade é violado quando uma série exibe uma tendência ou quebras de padrões. Isso resulta em complicações na análise econométrica que dependerá que alguma análise "não-estacionária".
- Em um processo estocástico estacionário de covariância, assume-se que as médias, variâncias e auto-covariâncias são independentes do tempo. Em um processo não estacionário, uma ou mais dessas suposições não são verdadeiras.
- Em muitos casos, o pesquisador deve transformar uma séries "não-estacionária" em estacionária.

Padrão Determinístico (PD)

- Suponha que {Yt} é decorrente de um PD, desta maneira:
- $y_t = \beta_0 + \beta_{1t} + \varepsilon_t, tal\ que, \varepsilon_t \sim RBG(0, \sigma^2); t = 0,1,2, \ldots$
- Assim, {Yt} é NE, dado que as variáveis dependem de t:$E[y_t] = \beta_0 + \beta_{1t}$



Passeio Aleatório (PA)

- Técnica interessante para associar esse fenômeno: No inglês, os pesquisadores apelidaram este padrão como '*Drunken Walk*' ou a caminhada do bêbado. Dado que quando uma pessoa está muito embriagada, ela costuma a andar aleatoriamente . Todavia, ela chega a algum destino (independente de qual seja).

Fonte: Hergé

Passeio Aleatório

- Podemos modelar uma série temporal y_t [projeção]que demonstra um padrão estocástico como um PA.
- $y_t = y_{t-1} + u_t$
- u_t são erros iid com $E(u_t|y_{t-1}, y_{t-2}, \dots) = 0$
- Assim,
- $E(y_t|y_{t-1}, y_{t-2}, \dots) = E(y_{t-1}|y_{t-1}, y_{t-2}, \dots) + E(u_t|y_{t-1}, y_{t-2}, \dots) = y_{t-1}$
- Ou seja, a projeção de hoje y_t depende dos dados de ontem y_{t-1} e o caminho para chegar a projeção depende de u_t. Em outras palavras, para projetar o amanhã, dependemos de ontem, mas chegaremos lá com um caminho ‘desconhecido’.

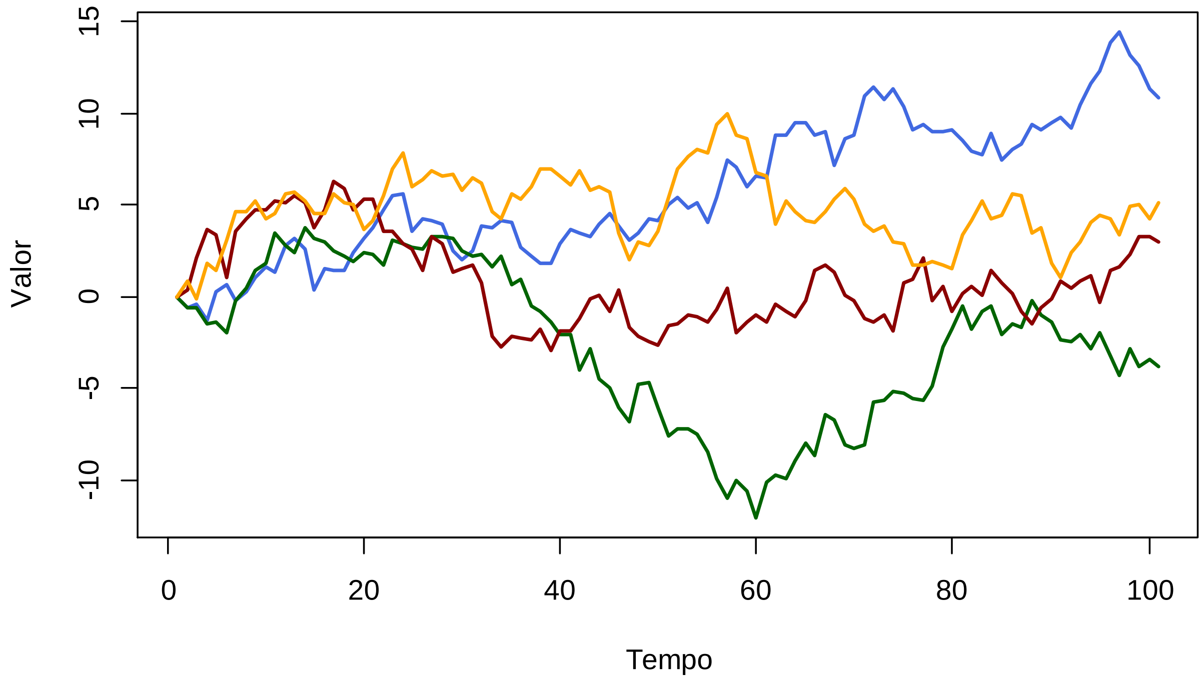
Passeio Aleatório

- Exemplo, uma série Y_0começando em ponto 0 com PA.
- Podemos visualizer a variância demonstrando um PA na seguinte operação.

$$\begin{aligned} Y_0 &= 0 \\ Y_1 &= 0 + u_1 \\ Y_2 &= 0 + u_1 + u_2 \\ &\vdots \\ Y_t &= \sum_{i=1}^{t} u_i. \end{aligned}$$

$$\begin{aligned} Var(Y_t) &= Var(u_1 + u_2 + \cdots + u_t) \\ &= t\sigma_u^2. \end{aligned}$$

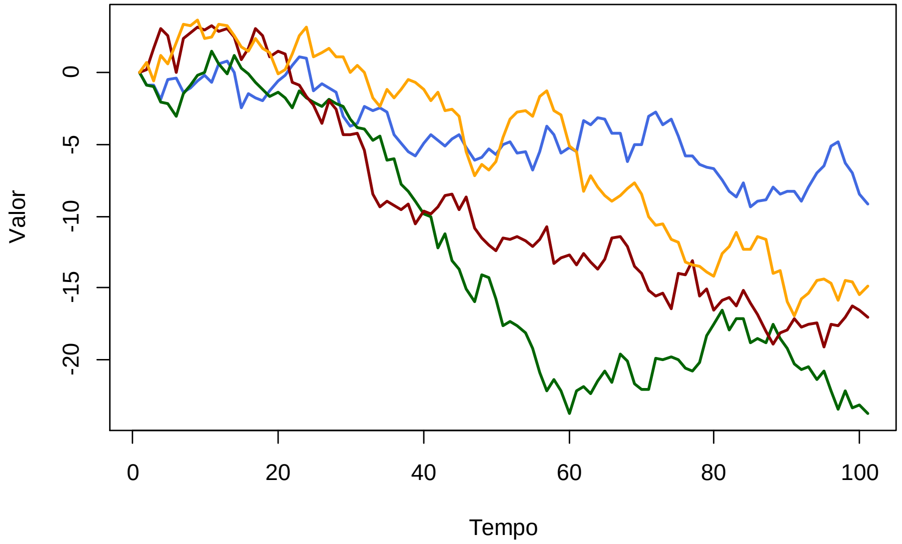
Passeio Aleatório (visualização)



Passeio Aleatório (c/ leve flutuação constante)

- Nesse caso, adicionamos um termo constante β_0 para verificar algum tipo de padrão no PA.
- Se β_0 é positivo a série tem um padrão positivo que gradualmente declina ao longo do tempo.

$$Y_t = \beta_0 + Y_{t-1} + u_t$$



Questões para refletir (não precisa entregar)

- Qual é o papel da autocorrelação e como isso difere uma série temporal de uma série que segue observações iid (independentes identicamente distribuídas)?
- O que caracteriza um processo estocástico estacionário e um processo de não-estacionariedade?
- O que é um Ruído Branco Gaussiano?
- O que é um passeio aleatório em uma série temporal?

Referências

- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74, 427–431[1]

Obrigado!