

Análise de Séries Temporais Univariadas e Multivariadas

Alunos: Iago Flávio e Hertz Rafael

Professor: Cássio da Nobrega Besarria

Universidade Federal da Paraíba
Centro de Ciências Sociais Aplicadas

Outubro de 2025



Questão 1 — Imputação de Dados Faltantes

- Seleção de uma série temporal (real ou simulada)
- Remoção proposital de períodos
- Aplicação de técnicas de imputação:
 - Interpolação Linear
 - Média Móvel
 - Outros
- Comparação entre valores reais e imputados

Exemplo dos Dados da Série

Date	Model	Estimated Deliveries	Source Type
2023-05-01	Model S	17646	Interpolated (Month)
2015-02-01	Model X	3797	Official (Quarter)
2019-01-01	Model X	8411	Interpolated (Month)
2021-02-01	Model 3	6555	Official (Quarter)
2016-12-01	Model Y	12374	Estimated (Region)

Total de observações: 2 640 linhas

Cenário 1 - Visualização da Série Estável

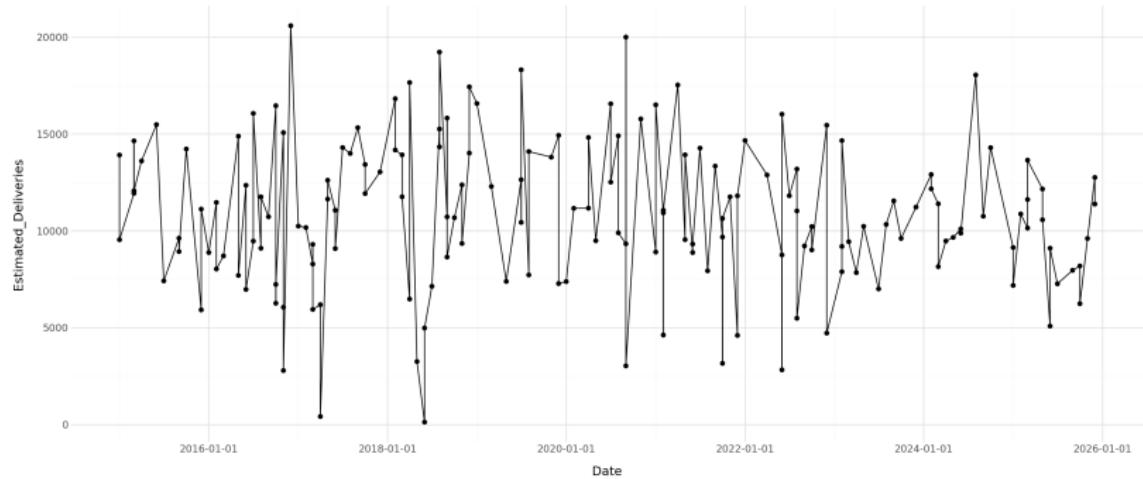


Gráfico da série temporal utilizada no processo de imputação (antes).

Cenário 1 - Visualização da Série Estável

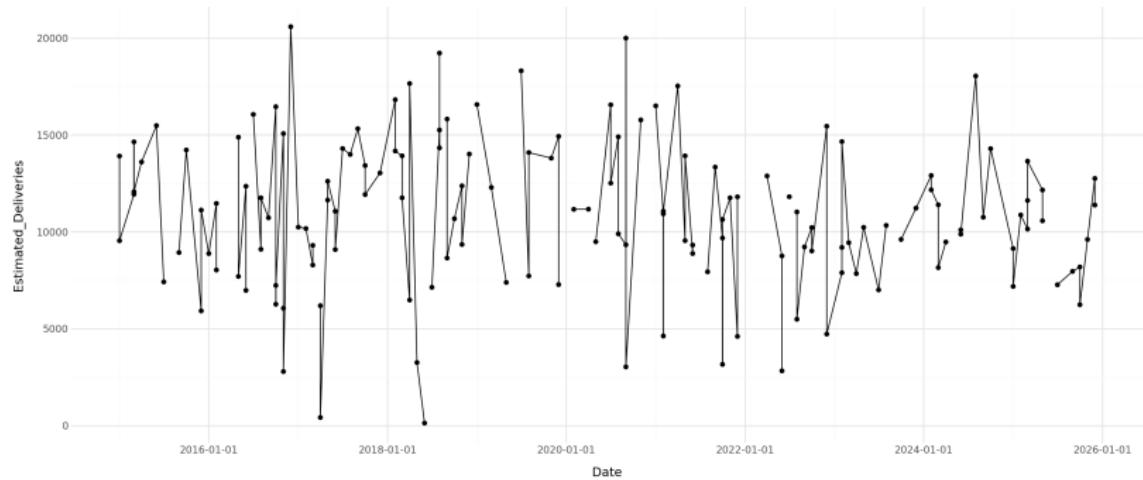


Gráfico da série temporal utilizada no processo de imputação (depois).

Métodos de Imputação Considerados

Método	Tipo de Dado	Complex.	Variab.	Comentário
Média/Mediana/Moda	Num./Cat.	Baixa	Não	Simples e rápido
Regressão	Numérico	Média	Parcial	Usa outras variáveis
KNN	Num./Cat.	Média	Sim	Mantém estrutura local
MICE	Num./Cat.	Alta	Sim	Abordagem estatística robusta
MissForest	Num./Cat.	Alta	Sim	Baseado em ML
Interpolação	Séries T.	Média	Sim	Mantém tendência temporal

Métodos utilizados no trabalho: forward fill, back fill, interpolação temporal, média.

Cenário 1 — Aplicação

Grupo	Métodos	RMSE	Melhor
Preench. simples	mean, ffill/bfill	4 944	mean (4 112)
Interp. linear	linear, time	5 372	linear (5 068)
Polin./spline	poly, quad, cubic, spline	6 542	spline (6 351)

Cenário 1 — Conclusões

- Melhor grupo: **Preenchimento simples**
- Melhor método: `mean`
- Série sem tendência → métodos simples funcionam melhor

Cenário 2 - Visualização da Série Acumulada

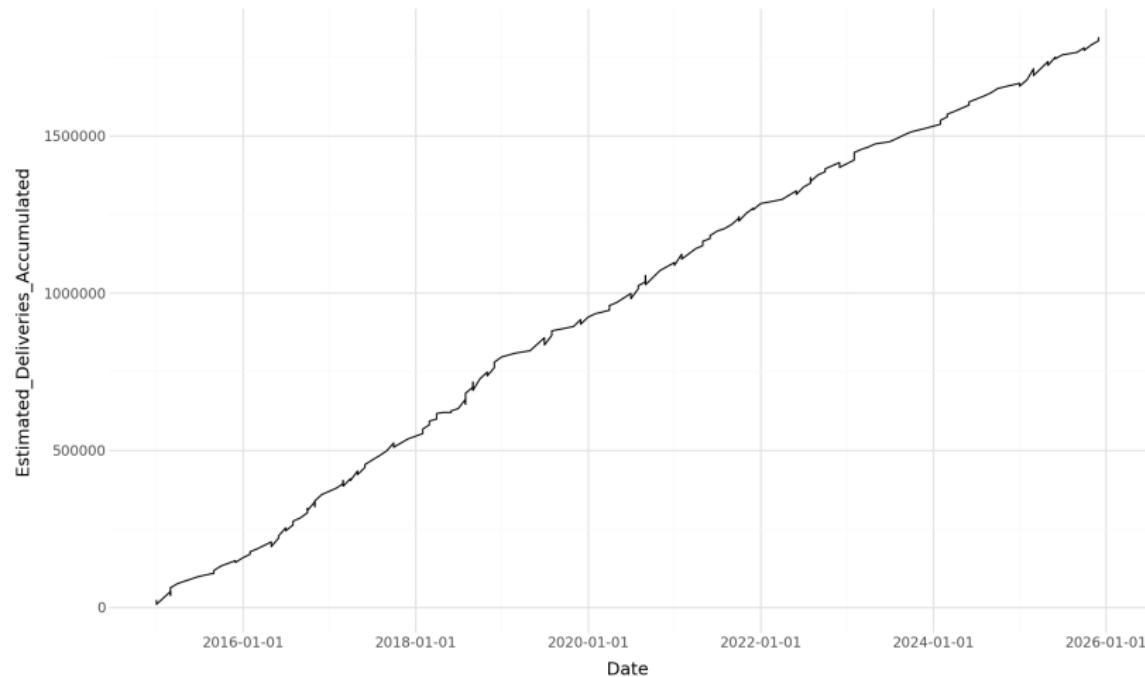


Gráfico da série temporal utilizada no processo de imputação (antes).

Cenário 2 - Visualização da Série Acumulada

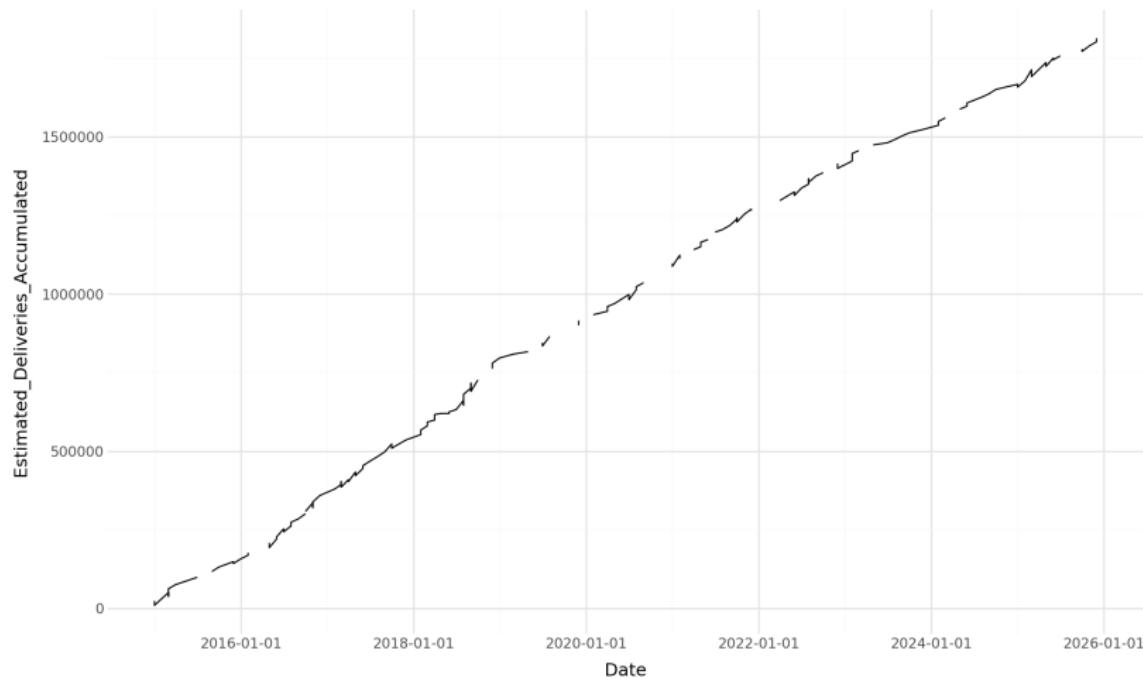


Gráfico da série temporal utilizada no processo de imputação (depois).

Cenário 2 — Série Acumulada

Grupo	Métodos	RMSE	Melhor
Preench. simples	mean, ffill/bfill	238 098	ffill/bfill (11 944)
Interp. linear	linear, time	7 122	linear (2 873)
Polin./spline	poly, quad, cubic, spline	4 811	spline (2 916)

Cenário 2 — Conclusões

- Melhor grupo: **Interpolação linear/polynomial**
- Melhor método: `interpolate_linear`
- Preserva a tendência e mantém coerência

Resumo Geral

Tipo	Grupo	Método	RMSE
Estável	Preench. simples	mean	4 100
Com tendência	Interp. linear	interpolate_linear	2 900

Questão 2 — Exemplo dos Dados Utilizados

Year	Month	Region	Model	Deliveries	Outlier Status
2023	5	Europe	Model S	17646	CO2_Saved_tons
2025	3	North America	Cybertruck	17079	CO2_Saved_tons
2020	10	Middle East	Cybertruck	18392	CO2_Saved_tons
2022	11	Middle East	Model Y	18106	Estimated_Deliveries, Production_Units
2020	2	Europe	Cybertruck	19298	CO2_Saved_tons

Coluna foco nesta etapa: Outlier_Status

Cenário 1 — Erros Claros ou Valores Impossíveis

Grupo	Métodos	Aplicação	Melhor	Comentário
Exclusão direta	drop_outliers, IQR cut	Altíssima	remover_outliers	Ideal p/ erros evidentes.
Transformação	log, sqrt	Baixa	Não recomendado	Problema é o valor, não a escala.
Imputação	median, nearest valid	Média	median	Quando não é possível excluir.

Melhor estratégia: Remover outliers impossíveis.

Cenário 2 — Assimetria Forte

Grupo	Métodos	Aplicação	Melhor	Comentário
Transformações	log, log1p, boxcox	Altíssima	log	Reduz assimetria sem remover valores.
Agrupamento	binning, categorizar	Média	faixa de valor	Mantém extremos em classes.
Exclusão	drop_outliers	Baixa	Não recomendado	Valores altos costumam ser reais.

Melhor estratégia: Transformação logarítmica.

Cenário 3 — Poucos Outliers Relevantes

Grupo	Métodos	Aplicação	Melhor	Comentário
Agrupamento	binarizar, faixa	Altíssima	binning	Mantém extremos importantes.
Transformações	log, sqrt	Média	log	Suaviza sem remover informação.
Imputação	median, nearest	Baixa	Não usar	Imputação destrói o outlier.

Melhor estratégia: Agrupar ou categorizar extremos.

Resumo — Tratamento de Outliers

Situação	Grupo	Método	Descrição
Erro claro / impossível	Remoção direta	drop_outliers	Limpa sem distorcer.
Muito alto, porém real	Transformação	log	Mantém e reduz impacto.
Extremo relevante	Agrupamento	binning	Preserva informação.

AR(1) com $\phi_1 = -2$



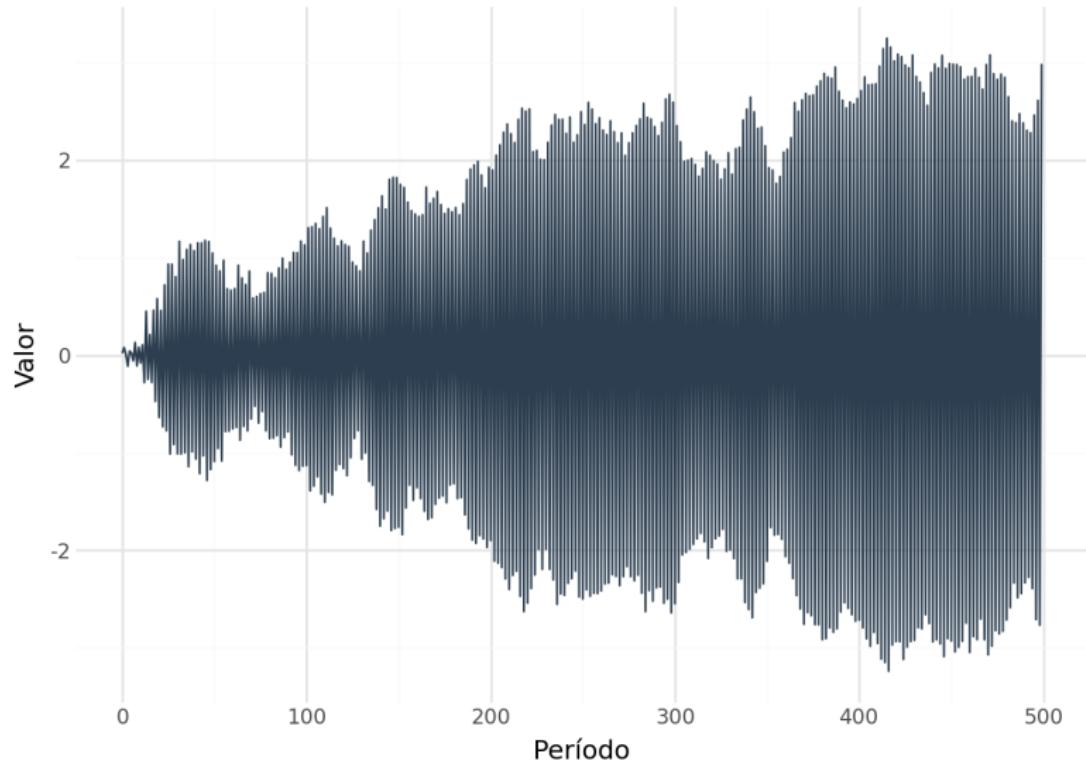
AR(1) com $\phi_1 = -2$ — Explicação

Processo explodindo e oscilando:

- Não-estacionário: $|\phi_1| > 1$.
- Oscila mudando de sinal a cada período.
- Amplitude cresce exponencialmente.
- Pequenos choques se ampliam rapidamente.

AR(1) com $\phi_1 = -1$

AR(1) com $\phi_1 = -1, \sigma = 0.1$



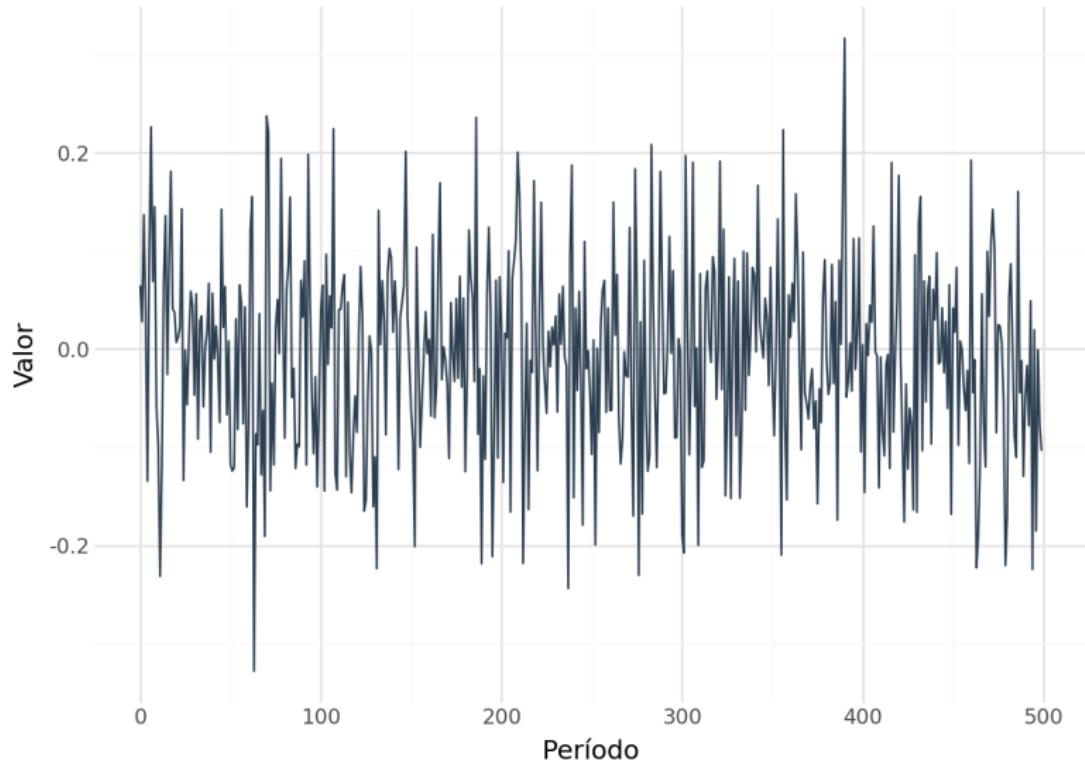
AR(1) com $\phi_1 = -1$ — Explicação

Raiz unitária oscilatória:

- Não-estacionário: choques têm efeito permanente.
- Alterna sinais a cada período.
- Oscilações aumentam gradualmente.
- Sem força de reversão à média.

AR(1) com $\phi_1 = 0$

AR(1) com $\phi_1 = 0, \sigma = 0.1$



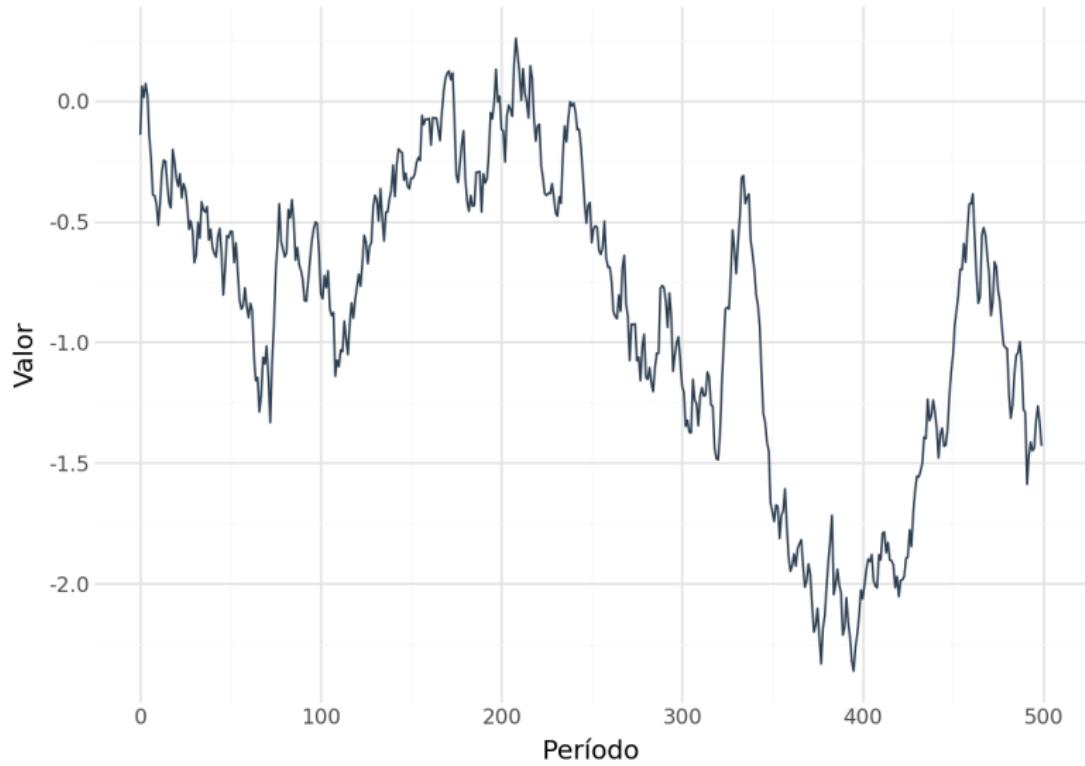
AR(1) com $\phi_1 = 0$ — Explicação

Ruído branco:

- Série totalmente aleatória: $Y_t = \varepsilon_t$.
- Sem memória do valor anterior.
- Estacionária (média e variância constantes).
- Sem tendência ou padrão visível.

AR(1) com $\phi_1 = 1$

AR(1) com $\phi_1 = 1, \sigma = 0.1$



AR(1) com $\phi_1 = 1$ — Explicação

Passeio aleatório (Random Walk):

- Não-estacionário: variância cresce no tempo.
- Acumula choques, vagando sem reversão à média.
- Choques têm efeito permanente.
- Série apresenta longas "tendências" falsas.

AR(1) com $\phi_1 = 2$

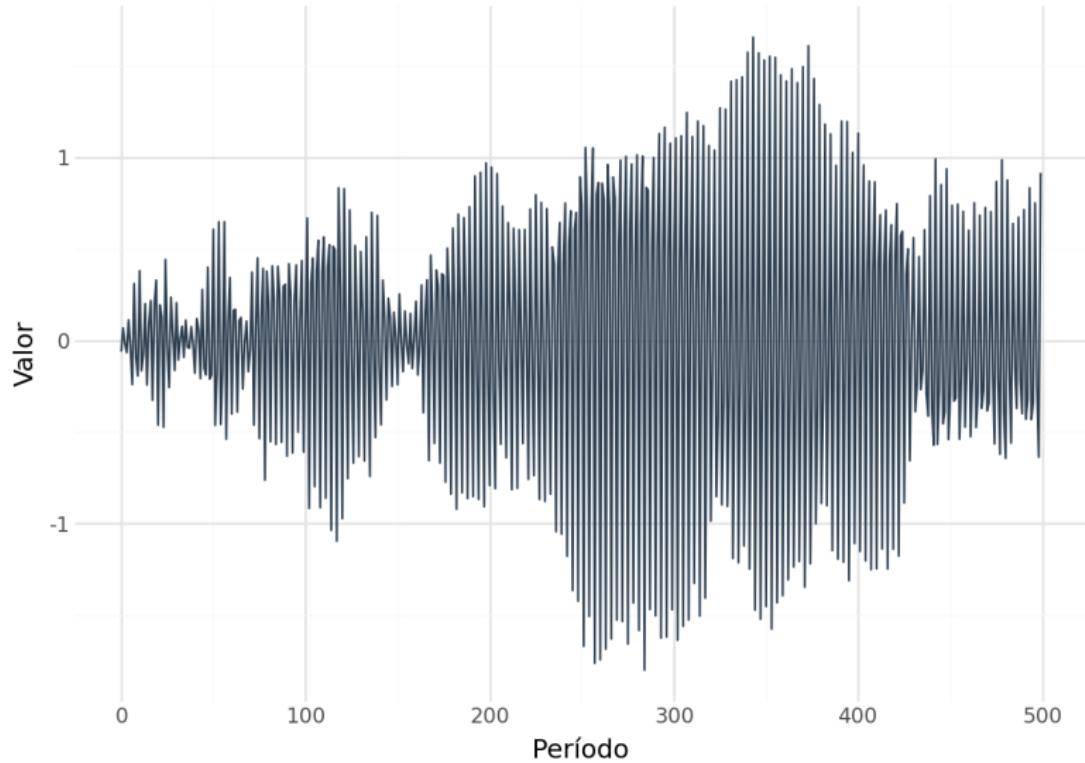


Processo explosivo positivo:

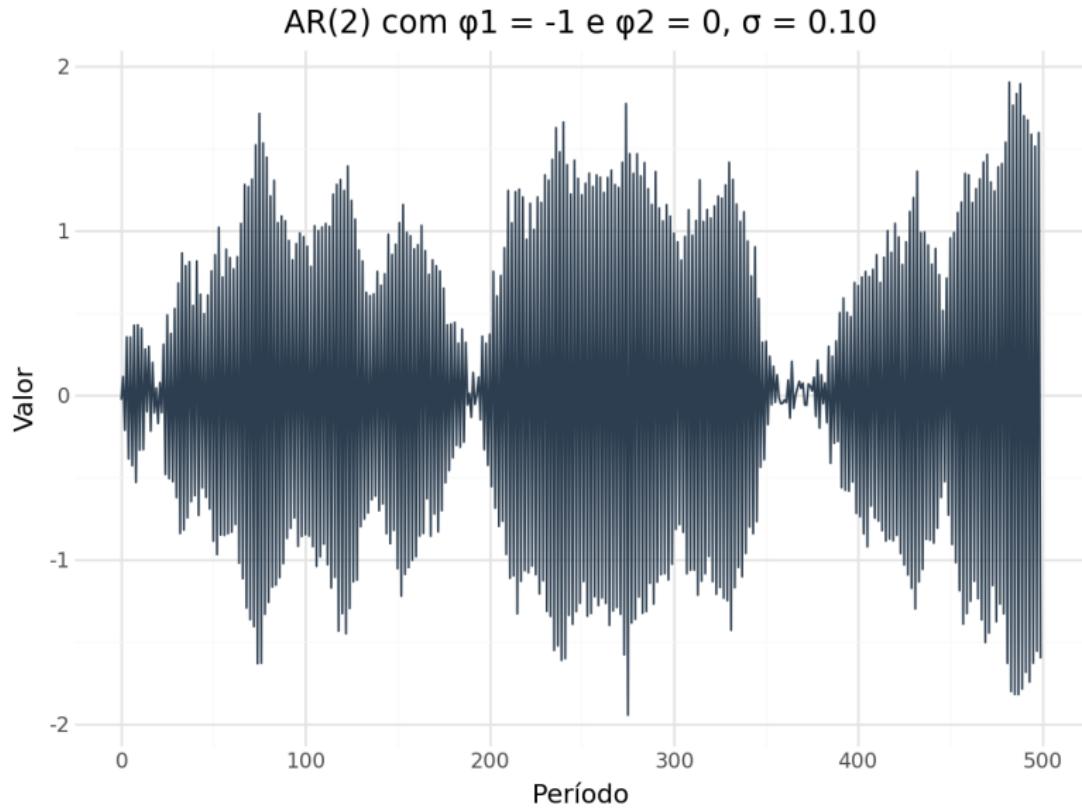
- Não-estacionário: $|\phi_1| > 1$.
- Diverge rapidamente para $\pm\infty$.
- Pequenos ruídos tornam-se grandes em poucos períodos.
- A série explode verticalmente após algum tempo.

AR(2): $\phi_1 = -1$, $\phi_2 = -1$

AR(2) com $\phi_1 = -1$ e $\phi_2 = -1$, $\sigma = 0.10$



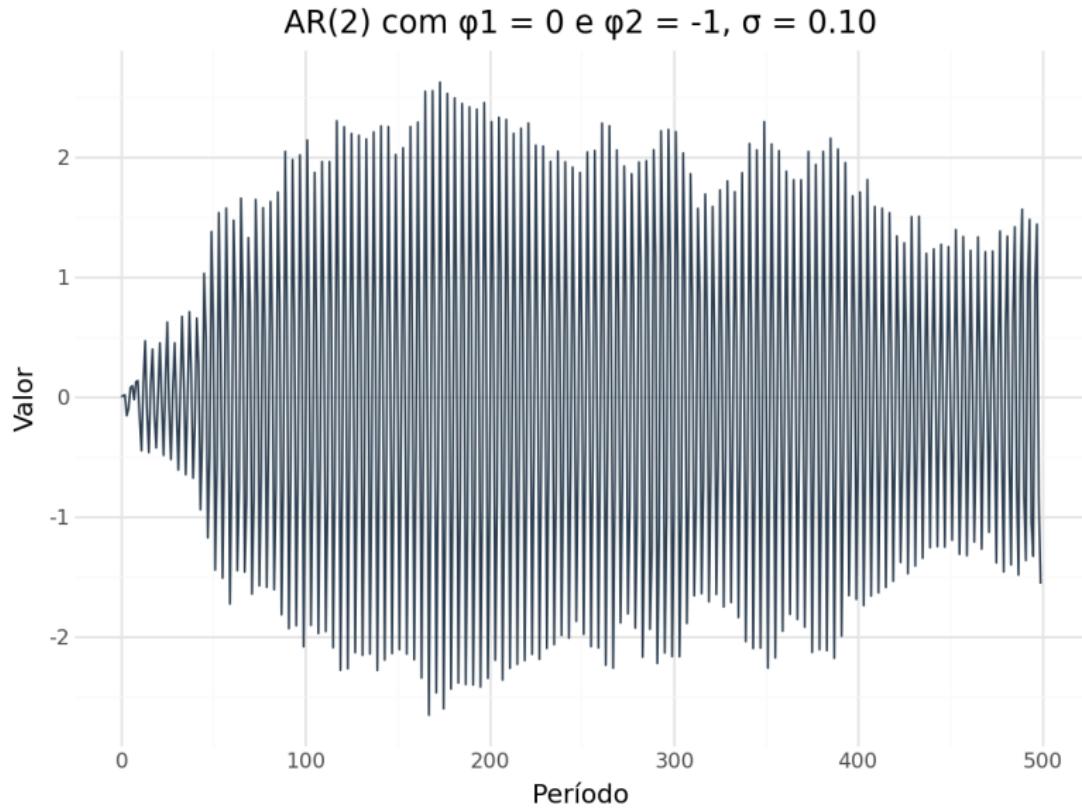
AR(2): $\phi_1 = -1$, $\phi_2 = 0$



AR(2): $\phi_1 = -1$, $\phi_2 = 1$

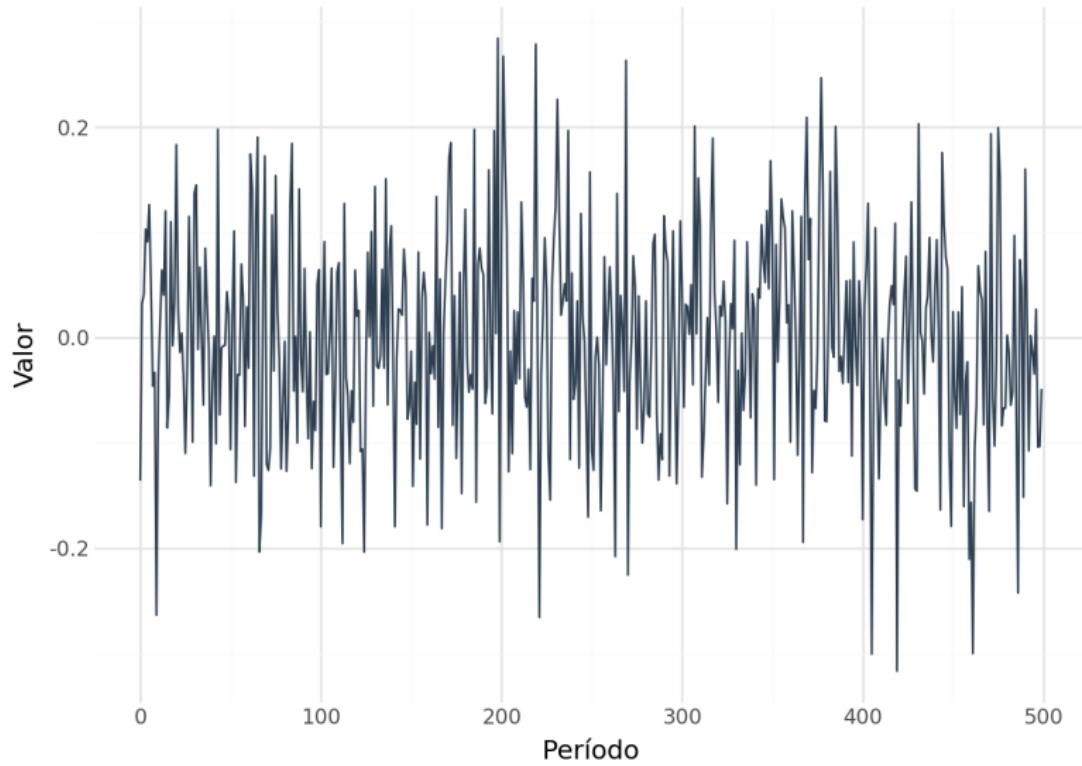


AR(2): $\phi_1 = 0$, $\phi_2 = -1$



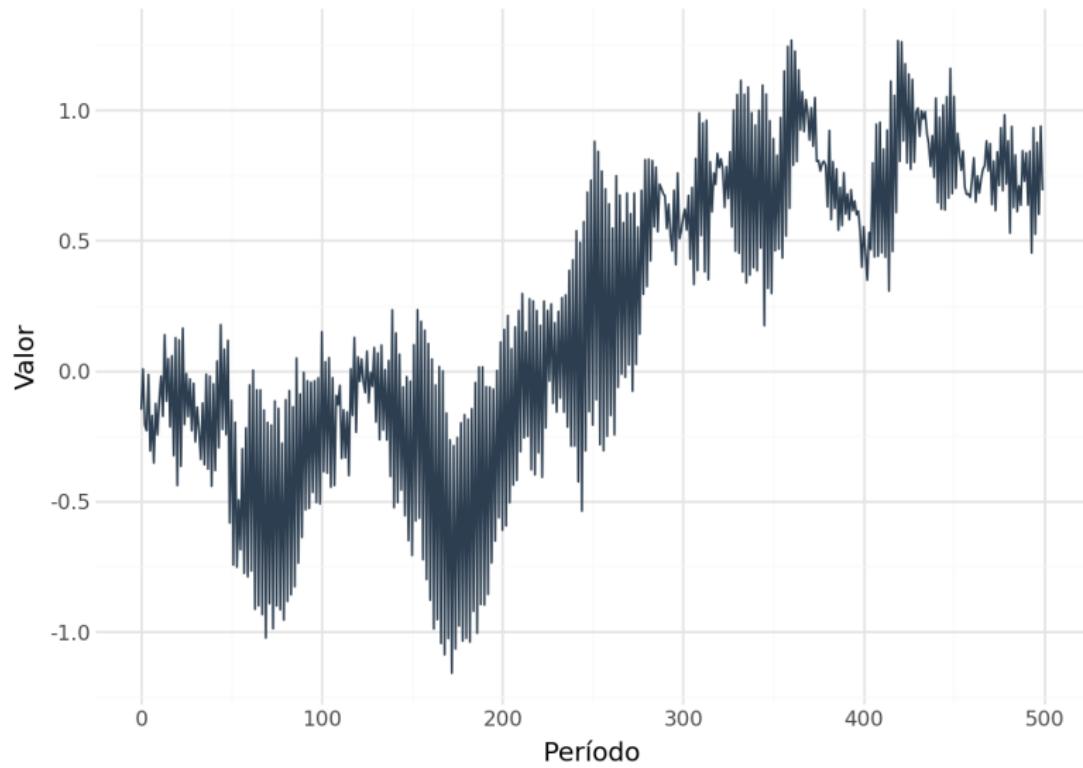
AR(2): $\phi_1 = 0$, $\phi_2 = 0$

AR(2) com $\varphi_1 = 0$ e $\varphi_2 = 0$, $\sigma = 0.10$



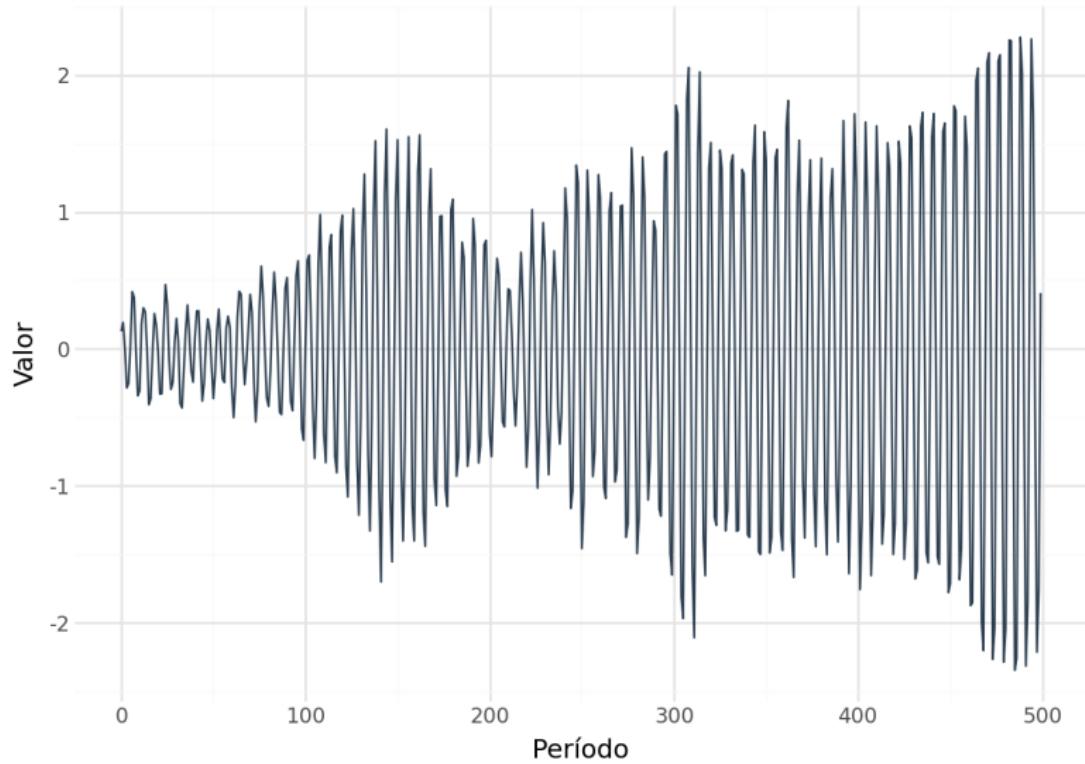
AR(2): $\phi_1 = 0$, $\phi_2 = 1$

AR(2) com $\varphi_1 = 0$ e $\varphi_2 = 1$, $\sigma = 0.10$



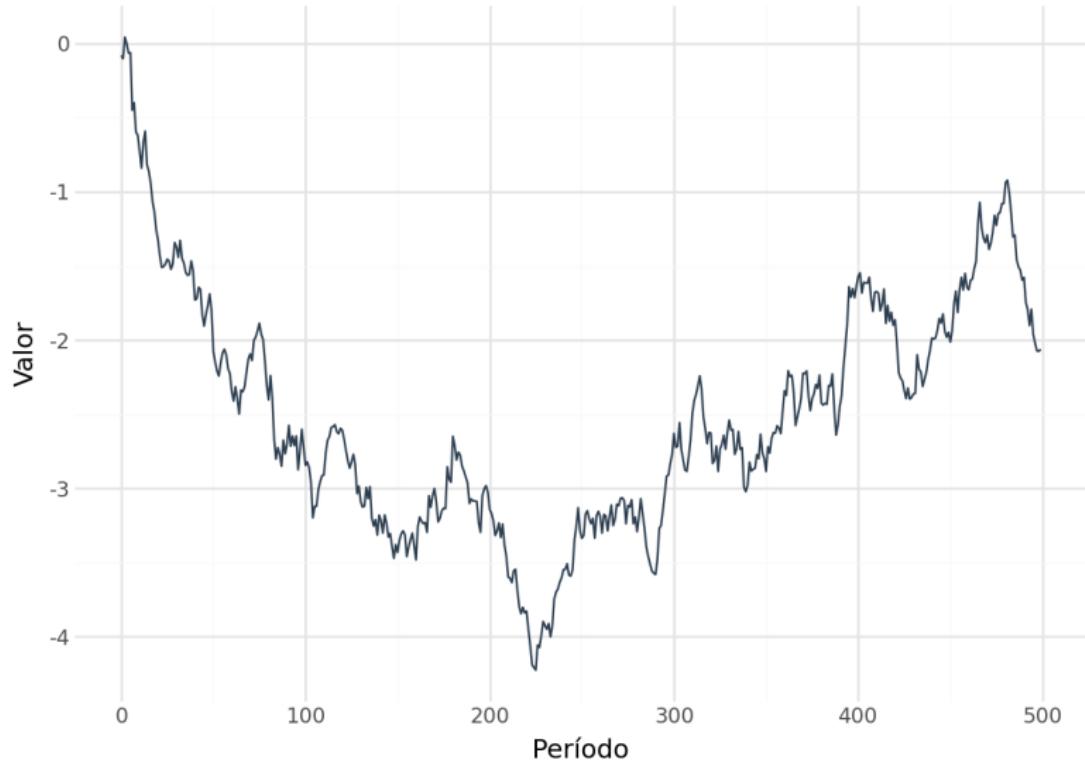
AR(2): $\phi_1 = 1$, $\phi_2 = -1$

AR(2) com $\phi_1 = 1$ e $\phi_2 = -1$, $\sigma = 0.10$



AR(2): $\phi_1 = 1$, $\phi_2 = 0$

AR(2) com $\varphi_1 = 1$ e $\varphi_2 = 0$, $\sigma = 0.10$



AR(2): $\phi_1 = 1$, $\phi_2 = 1$



Resumo do AR(2) — Padrões Gerais

Modelo AR(2):

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

Comportamentos observados nos gráficos:

- **Oscilações cíclicas:** aparecem quando $\phi_2 < 0$.
- **Persistência elevada:** quando ϕ_1 está próximo de 1.
- **Ruído quase branco:** quando $\phi_1 = \phi_2 = 0$.
- **Oscilação amortecida:** se $\phi_2 > 0$ e $|\phi_1| < 1$.

Resumo do AR(2) — Estabilidade e Explosão

Condição-chave: Estacionariedade

$$1 - \phi_1 z - \phi_2 z^2 = 0$$

O processo é estacionário quando ambas as raízes têm módulo maior que 1.

Padrões observados:

- **Explosão:** ocorre quando as raízes têm módulo < 1 .
- **Oscilação explosiva:** $\phi_2 < 0$ e $|\phi_1|$ grande.
- **Oscilação suave e persistente:** $\phi_2 > 0$.

Resumo visual:

- $\phi_2 < 0 \rightarrow$ séries alternam sinais e criam ciclos.
- $\phi_2 = 0 \rightarrow$ comportamento semelhante ao AR(1).
- $\phi_2 > 0 \rightarrow$ oscilações mais suaves e estáveis.

Questão 4 — Introdução aos Dados e Modelos

Nesta questão aplicamos três técnicas clássicas de **suavização exponencial**:

- **SES** – Suavização Exponencial Simples (nível)
- **Holt** – Dupla Suavização (nível + tendência)
- **Holt-Winters** – Tripla Suavização (nível + tendência + sazonalidade)

Cada método foi aplicado a uma série temporal apropriada para sua estrutura.

O objetivo é comparar o desempenho dos modelos e interpretar as previsões obtidas.

Questão 4 — Séries Utilizadas em Cada Método

1) Suavização Exponencial Simples (SES)

- Utiliza a série da **Tesla**, já utilizada nas questões anteriores.
- Séries estacionárias ou com pouca estrutura → ideal para SES.

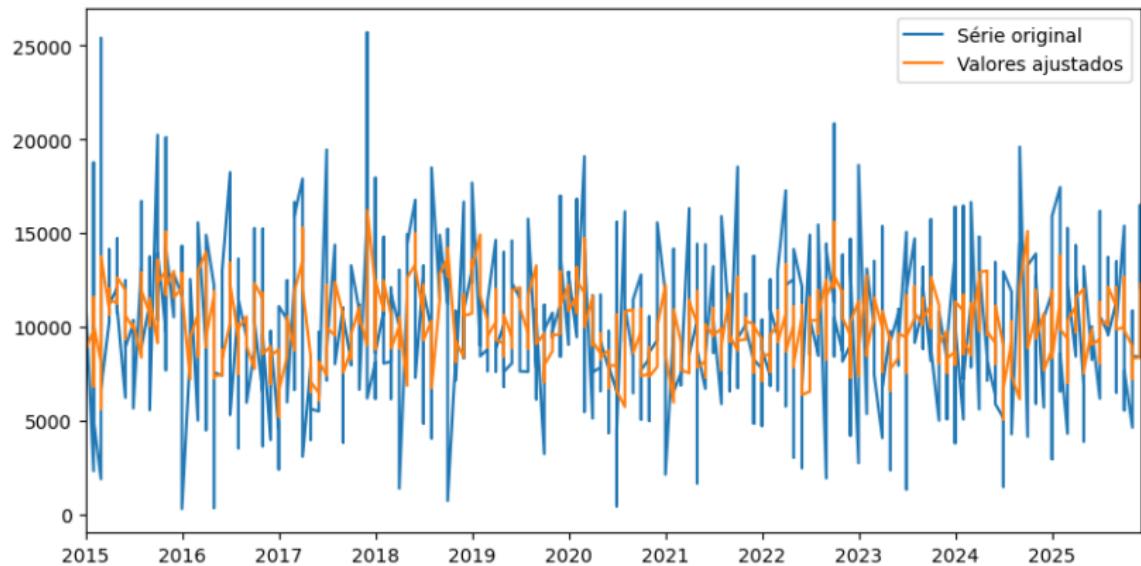
2) Holt (Nível + Tendência)

- Utiliza dados do índice **S&P 500** (ticker **GSPC**).
- Série com forte tendência crescente.

3) Holt-Winters (Nível + Tendência + Sazonalidade)

- Utiliza o clássico dataset de **passageiros de avião**, disponível no GitHub.
- Série com tendência e comportamento sazonal marcado.

Questão 4 — Série Estacionária



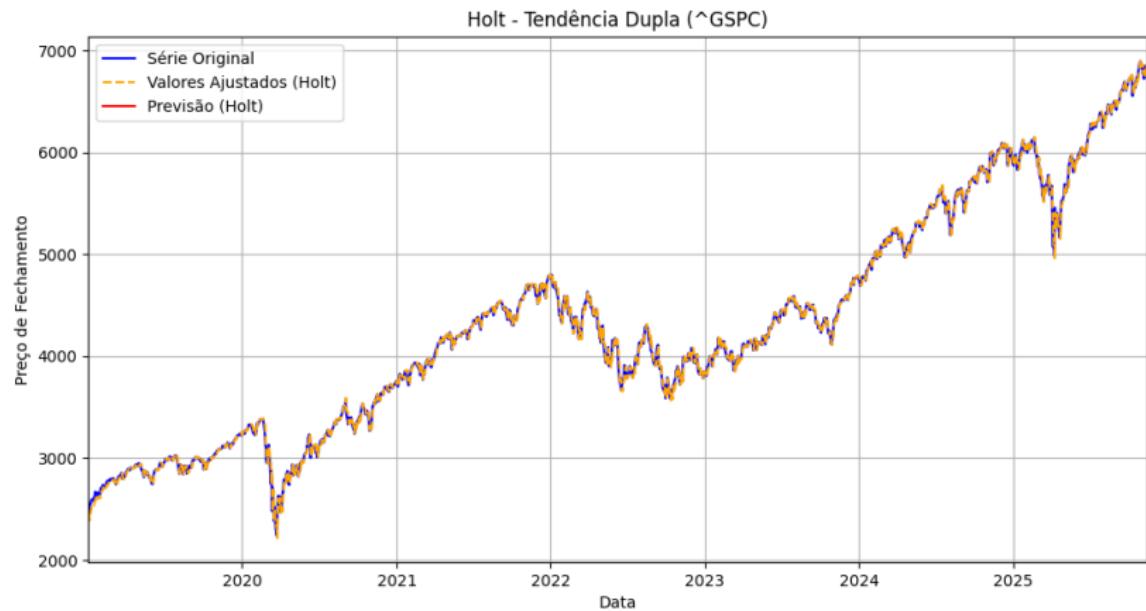
Questão 4 — Interpretação (SES)

Modelo aplicado: Suavização Exponencial Simples (SES)

- Funciona melhor quando a série só tem **nível**, sem tendência e sem sazonalidade.
- Modelo mais parcimonioso: captura apenas a média móvel exponencial.
- Apresentou o **menor RMSE** no cenário estacionário.

Conclusão: O SES é o melhor para séries estacionárias ou de ruído branco, onde modelos mais complexos só adicionariam erro.

Questão 4 — Série com Tendência



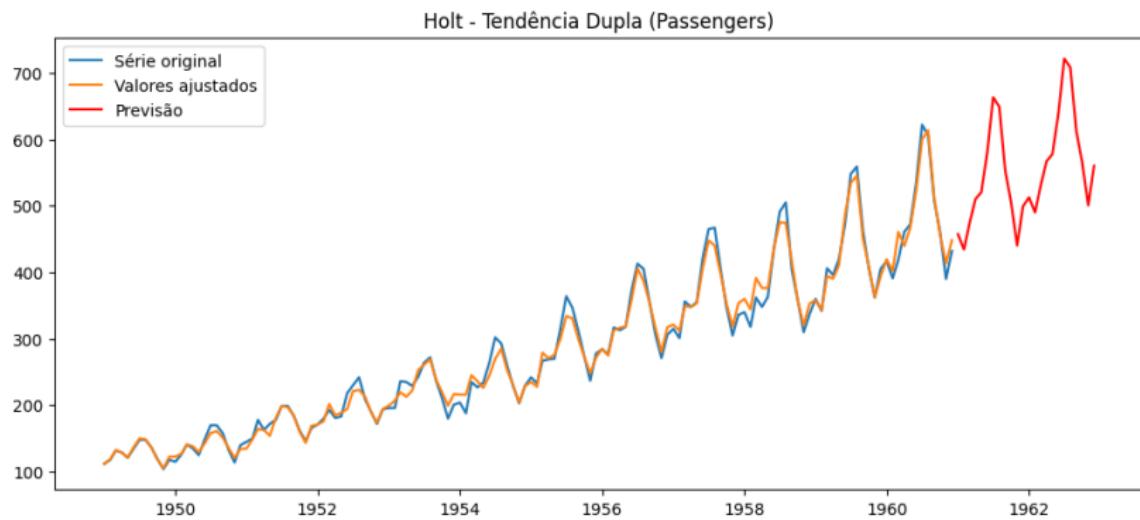
Questão 4 — Interpretação (Holt)

Modelo aplicado: Holt (nível + tendência)

- Projetado para séries com **tendência linear**.
- O SES falha totalmente aqui, pois prevê apenas o nível.
- Holt capturou bem a tendência, mesmo com ruído elevado.
- RMSE médio, mas **melhor desempenho entre os métodos**.

Conclusão: O modelo de Holt é ideal quando existe tendência clara, mesmo com forte volatilidade.

Questão 4 — Série com Tendência e Sazonalidade



Questão 4 — Interpretação (Holt-Winters)

Modelo aplicado: Holt-Winters Multiplicativo

- Indicado quando há **nível + tendência + sazonalidade**.
- O SES falha totalmente e o Holt não captura padrões sazonais.
- O Holt-Winters Multiplicativo teve o **melhor RMSE**.
- Modelo consegue acompanhar crescimento sazonal crescente.

Conclusão: Para séries com sazonalidade clara, o Holt-Winters é o modelo natural e mais eficiente.

Resumo Global das Técnicas — Parte 1

Comparação geral dos modelos de suavização exponencial:

1) Séries com apenas Nível (Estacionárias / Ruído)

- **Modelo ideal: SES**
- **RMSE típico: Baixo**
- **Por quê?** Holt e Holt-Winters introduzem complexidade desnecessária.

2) Séries com Nível + Tendência

- **Modelo ideal: Holt**
- **RMSE típico: Médio**
- **Limitação dos outros:**
 - SES ignora tendência.
 - Holt-Winters é mais complexo sem necessidade.

Resumo Global das Técnicas — Parte 2

3) Séries com Nível + Tendência + Sazonalidade

- **Modelo ideal:** Holt-Winters
- **RMSE típico:** Baixo
- **Razão:** Holt não modela sazonalidade, gerando erro estrutural.

Regra geral de aplicação:

- Se a série tem **apenas nível** → use **SES**.
- Se tem **tendência** → use **Holt**.
- Se tem **sazonalidade** → use **Holt-Winters**.

Conclusão final: Cada modelo deve respeitar a estrutura da série temporal. Modelos mais complexos só funcionam bem quando a série realmente exige essa complexidade.