



**Fundação Getúlio Vargas
Escola de Matemática Aplicada**

Séries Temporais

Relatório da A1 de Séries Temporais

**Leonardo Alexandre da Silva Ferreira
Guilherme Moreira Castilho**

Rio de Janeiro
Outubro / 2025

Sumário

1	Introdução	3
2	Metodologia	3
3	Transformação de Variáveis	3
4	Decomposição STL	4
4.1	Análise dos Componentes Estruturais via Decomposição STL	4
5	Análise de Resíduos e Ajuste dos Modelos	4
6	Modelos de Regressão Linear Múltipla	5
7	Discussão	6
8	Conclusão	6

1 Introdução

O documento apresenta uma análise detalhada de séries temporais com o objetivo de modelar a variável *volume* utilizando técnicas fundamentais. A metodologia empregada englobou modelos *baseline*, transformação de variáveis, decomposição temporal e regressão linear múltipla. O propósito principal deste estudo é desenvolver um modelo preditivo robusto para a variável *volume*, levando em consideração suas características temporais, sazonalidade e tendências.

A justificativa metodológica para as escolhas feitas baseou-se na necessidade de: estabelecer **benchmarks** através de modelos *baseline* simples; **estabilizar a variância** por meio de transformações adequadas; **capturar a sazonalidade** com decomposição temporal; **desenvolver modelos preditivos** utilizando covariáveis derivadas; e **validar a adequação** dos modelos através da análise de resíduos.

2 Metodologia

Inicialmente, foram implementados quatro modelos *baseline* para estabelecer *benchmarks*: **Mean** (Média simples dos dados históricos); **Naive** (Último valor observado); **Snaive** (Último valor da mesma estação); e **Drift** (Tendência linear baseada no primeiro e último valor).

Para a avaliação dos modelos, foram utilizadas as seguintes métricas: **MAE** (Erro absoluto médio); **RMSE** (Raiz do erro quadrático médio); **MAPE** (Erro percentual absoluto médio); **MASE** (Erro absoluto escalado médio).

Implementou-se uma validação cruzada temporal com *5 folds*, respeitando a ordem cronológica dos dados para evitar vazamento de informação. As **vantagens** dessa abordagem incluem o *Realismo* (simula condições reais de previsão), a *Robustez* (evita *overfitting*) e a *Interpretabilidade* (resultados mais confiáveis).

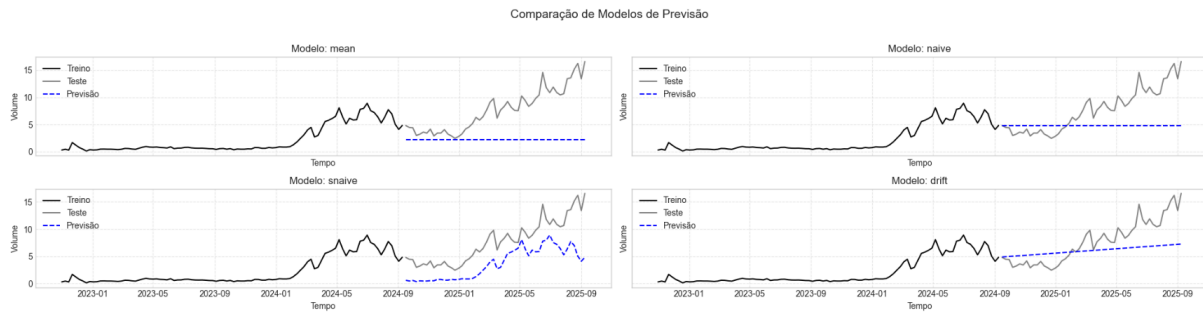
Para analisar os resíduos, utilizamos gráficos de Autocorrelação (ACF), Teste de Ljung-Box e Histograma dos Resíduos. Os dois primeiros foram utilizados para verificar a não correlação dos resíduos, enquanto o histograma permite verificar se a distribuição dos erros é aproximadamente normal e centrada em zero.

3 Transformação de Variáveis

A análise inicial da série temporal revelou **heterocedasticidade** (variância não constante ao longo do tempo), indicando a necessidade de transformação para estabilizar a variância. Portanto, aplicou-se a transformação Box-Cox com parâmetro $\lambda = 0.0496$.

A transformação resultou em melhorias significativas nas métricas dos modelos *baseline*. O modelo **Drift** teve uma melhoria de 86.5% (MAE Original = 2.894 para MAE Transformado = 0.391). O **Naive** melhorou em 84.0% (MAE Original = 3.763 para MAE Transformado = 0.603). O **Snaive** melhorou em 72.3% (MAE Original = 4.083 para MAE Transformado = 1.130). Por fim, o **Mean** melhorou em 67.5% (MAE Original = 5.485 para MAE Transformado = 1.784).

Comparando os resultados das previsões desses modelos com o valor real do último ano utilizando o tempo anterior como teste temos:



4 Decomposição STL

4.1 Análise dos Componentes Estruturais via Decomposição STL

A decomposição STL foi utilizada para desmembrar a série *volume* em seus componentes aditivos.

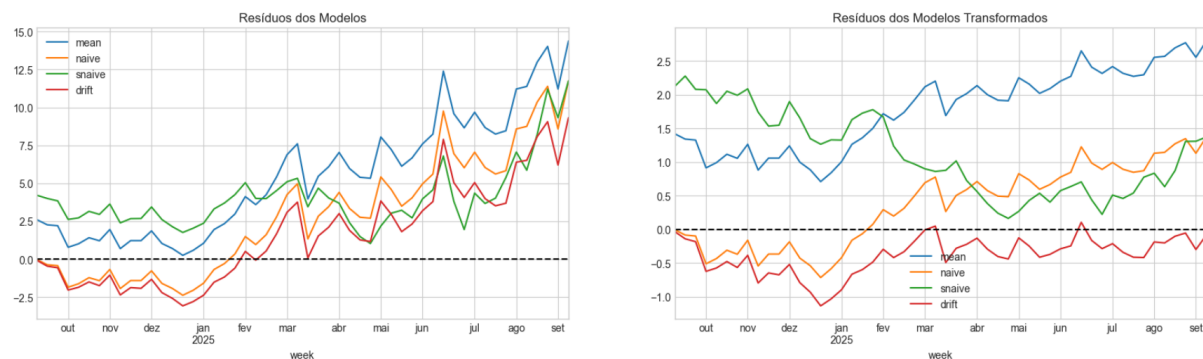
Tendência: O gráfico demonstra uma **tendência de crescimento linear e constante** ao longo de todo o período observado;

Sazonalidade: A linha do componente Sazonal não ser plana e apresentar picos e vales acentuados a cada ano **confirma a relevância da sazonalidade** na estrutura da série;

Resíduo: os resíduos apresentaram um **comportamento predominantemente aleatório**.

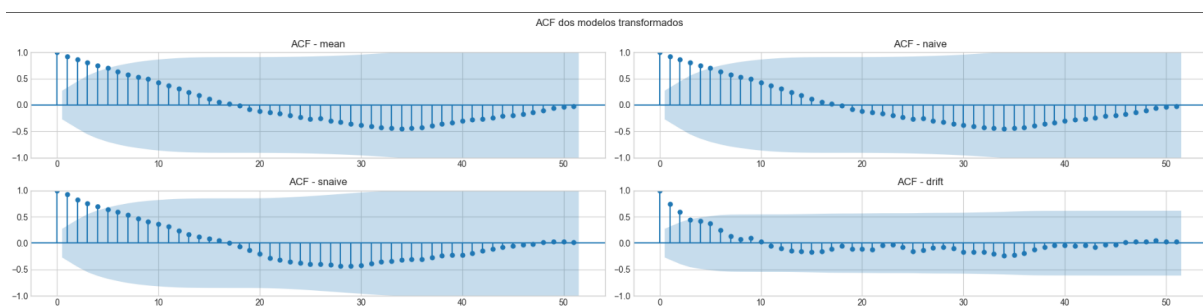
5 Análise de Resíduos e Ajuste dos Modelos

Comparando os resíduos dos 4 modelos baseline com e sem a transformação:

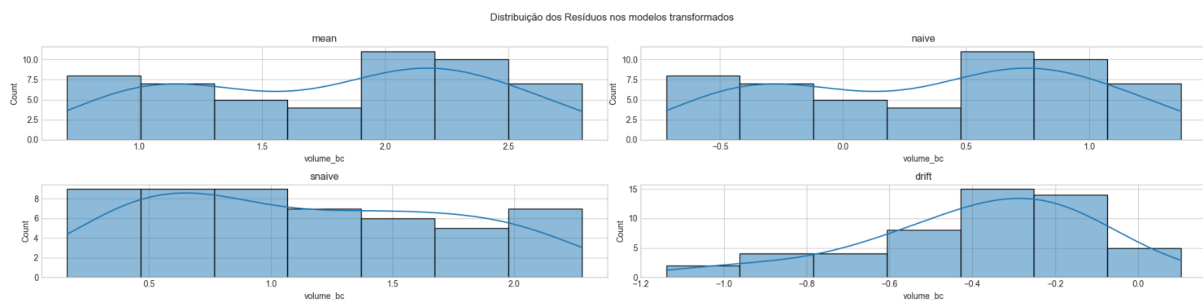


Perceba que os resíduos na transformação se aproximam muito mais de 0, e variam bem menos que sem a transformação - isso vale para todos os modelos.

Comparando agora o ACF para os modelos com transformação, vemos que o modelo com Drift é o que melhor performa aqui:



Além disso, seus resíduos são os que mais se aproximam de uma distribuição normal centrada em 0.



6 Modelos de Regressão Linear Múltipla

A regressão por Mínimos Quadrados Ordinários (Ordinary Least Squares – OLS) foi utilizada como abordagem principal para modelagem da série. OLS é um método clássico e amplamente consolidado, cujo objetivo é estimar os coeficientes de forma a minimizar a soma dos quadrados dos resíduos.

Para enriquecer o modelo e capturar diferentes aspectos da dinâmica da série, foram incluídas variáveis derivadas, sendo elas: variáveis de calendário (ano, mês, semana do ano, dia da semana) que permitem calcular efeitos sazonais e padrões recorrentes associados ao calendário; defasagem (lags de 1, 2 e 52 períodos) que incorporam dependência temporal de curto e longo prazo; Médias móveis (rolling mean de 4 e 12 períodos) que suavizam flutuações de curto prazo e ajudam a representar tendências locais.

Para avaliar a robustez do modelo e selecionar o melhor conjunto de variáveis, foi aplicada a lógica de validação cruzada leave-one-out adaptada ao contexto temporal. Diferente da LOOCV tradicional (que remove uma observação aleatória), aqui cada ponto futuro foi previsto a partir de todo o histórico anterior, respeitando a ordem cronológica. Essa estratégia garante que não haja vazamento de informação e simula de forma realista o processo de previsão em tempo real. A vantagem é obter uma medida mais confiável do erro fora da amostra, reduzindo o risco de overfitting e permitindo comparar diferentes especificações de forma justa.

Os modelos levaram em conta um set de previsão/teste de um ano e todo o tempo que antecede esse período fori para o set de treino.

Seguindo a ideia de Leave-One-Out, foram testadas todas as combinações de features das derivações feitas. O ranqueamento foi feito pela média e desvio padrão dos erros de MAE e RMSE em cada set da Cross-Validation. Com isso, as features selecionadas foram "year", "lag1" e "lag2".

O desempenho apresentado pelo modelo resultante sobre a transformação box-cox da variável "volume" foi:

MAE: 0.15441879382273171
RMSE: 0.19422264333295294

OLS Regression Results

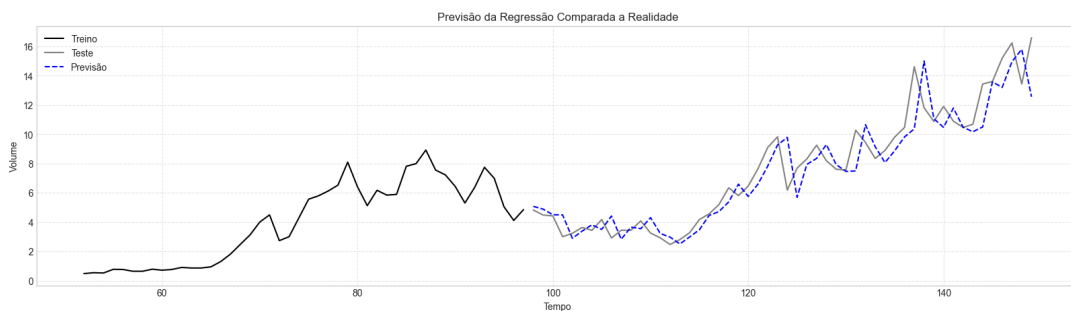
```
=====
Dep. Variable:          volume      R-squared (uncentered):      0.983
Model:                  OLS         Adj. R-squared (uncentered):    0.982
Method:                 Least Squares   F-statistic:                822.8
Date:                  dom, 05 out 2025   Prob (F-statistic):         5.55e-38
Time:                  23:02:03         Log-Likelihood:             10.267
No. Observations:      46            AIC:                        -14.53
Df Residuals:          43            BIC:                        -9.049
Df Model:              3
Covariance Type:       nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
year	4.386e-05	2.07e-05	2.117	0.040	2.08e-06	8.56e-05
lag1	1.1403	0.149	7.638	0.000	0.839	1.441
lag2	-0.1853	0.146	-1.272	0.210	-0.479	0.108

```
=====
Omnibus:                1.895      Durbin-Watson:              1.898
Prob(Omnibus):          0.388      Jarque-Bera (JB):           1.553
Skew:                   -0.447      Prob(JB):                   0.460
Kurtosis:               2.902      Cond. No.                   1.42e+04
=====
```

Veja que o modelo apresenta baixos RMSE e MAE, além de altos valores de R^2 .

E a partir do plot de comparação entre o valor real e o valor predito vemos o bom comportamento do modelo.



7 Discussão

Os resultados demonstram que: 1. A Transformação Box-Cox foi fundamental para estabilizar a variância e melhorar a performance dos modelos. 2. O Modelo Drift foi o melhor baseline, capturando a tendência de crescimento. 3. A Regressão Linear Múltipla superou significativamente os baselines. 4. A Variável lag1 foi a mais importante, indicando forte autocorrelação de primeira ordem.

8 Conclusão

Este estudo demonstrou a eficácia de uma abordagem metodológica sistemática para análise de séries temporais, resultando em um modelo preditivo robusto. A **Transformação de Variáveis (Box-Cox com $\lambda = 0.0496$)** foi crucial para estabilizar a variância, resultando em uma melhoria de até 86.5% no MAE dos modelos *baseline* e adequação aos pressupostos estatísticos. A **Decomposição Temporal (STL)** revelou um padrão sazonal e uma tendência de crescimento, facilitando a interpretação. Entre

os Modelos de Referência, o **Drift** apresentou a melhor performance ($MAE = 0.391$), estabelecendo um *benchmark* adequado. O modelo de **Regressão Linear Múltipla** final obteve $R^2 = 0.998$, com a variável lag1 sendo altamente significativa ($p < 0.001$) e os resíduos mostrando adequação estatística.