

Trabalho A1 - Series Temporais:

Resumo Executivo do Projeto

Pedro Henrique Coterli, Kauan Mariani Ferreira, Matheus Fillype Ferreira de Carvalho
Sillas Rocha da Costa, Gustavo Tironi, Livia Verly

05 de outubro de 2025

1. Contexto e Análise Exploratória

Uma **série temporal** consiste em observações $\{y_t\}$ registradas em intervalos regulares de tempo t . Neste trabalho, a série em estudo (Figura 1) representa um *volume semanal* e nosso objetivo é modelar essa variável e realizar previsões futuras. Para isso, queremos caracterizar a distribuição conjunta da sequência de variáveis aleatórias $\{Y_t\}$.

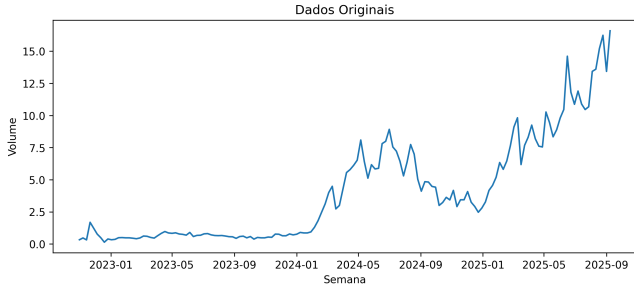


Figura 1: Gráfico da base de dados disponível

Como apenas uma realização da série é observada, a modelagem baseia-se nos **momentos de primeira e segunda ordem** — média $E[Y_t]$ e covariância $E[Y_{t+h}Y_t]$. Devido à natureza temporal dos dados, caso o modelo não capture adequadamente essa dependência, ela se manifestará nos resíduos.

Para isso, dois grupos principais de métodos foram explorados:

1. **Modelos Simples de Previsão (Baselines):** métodos simples, que servem como um "desempenho a ser batido".
2. **Modelos de Regressão Linear Múltipla:** baseados no modelo linear

$$y_t = \beta_0 + \beta_1 x_{1t} + \dots + \epsilon_t,$$

incorporando covariáveis derivadas de `week` e `volume`, como tendência, defasagens (*lags*), médias móveis e sazonalidade. A estimação dos coeficientes $\hat{\beta}$ é realizada via *Mínimos Quadrados Ordinários (OLS)*.

Iniciamos a análise investigando a presença de padrões de **tendência** e **sazonalidade** na série temporal. A **Figura 2** apresenta a decomposição do dado em suas três componentes — tendência, sazonalidade e resíduo — permitindo observar

o comportamento isolado de cada uma. Foram consideradas janelas temporais de **4 semanas** (ciclo mensal) e **52 semanas** (ciclo anual) para capturar diferentes escalas de variação.

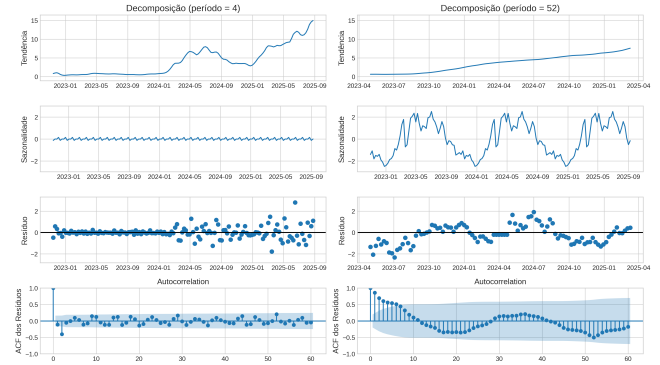


Figura 2: Tendência, sazonalidade e resíduo e autocorrelação para o ciclo mensal e anual

A análise revela um forte indício de **tendência** na série, evidenciado pela alta autocorrelação entre valores temporalmente próximos, que decai lentamente com o aumento do *lag*. Por outro lado, não há evidências relevantes de **sazonalidade**: para um período de 4 semanas (ciclo mensal), a variação sazonal é muito baixa (entre -0.1 e 0.1) e pouco significativa, além de apresentar resíduos crescentes ao longo do tempo, sugerindo heterocedasticidade. Já no período de 52 semanas (ciclo anual), embora a componente sazonal seja mais perceptível, os resíduos mantêm correlação, indicando ausência de ruído branco. Assim, em ambos os casos, a decomposição sazonal não se mostra adequada para representar a estrutura da série.

2. Ajustando os dados

A seguir, analisaremos o ajuste de modelos da forma:

$$f(\text{volume}_t) = \beta_0 + \beta_1 t + \epsilon_t$$

onde t corresponde à covariável de tendência do modelo e f a uma possível transformação.

Ao observar o comportamento da série na Figura 1, nota-se um padrão **exponencial** na variável `volume`, sugerindo a aplicação de uma **transformação logarítmica**. Portanto, ajustaremos dois modelos e avaliaremos a qualidade de seus ajustes:

1. $\text{volume}_t = \beta_0 + \beta_1 t + \epsilon_t$
2. $\log(\text{volume}_t) = \beta_0 + \beta_1 t + \epsilon_t$

A qualidade do ajuste foi avaliada utilizando o **Coefficiente de Determinação Ajustado** (R^2 ajustado), que mede a proporção da variabilidade de y_t explicada pelo modelo, penalizando o número de parâmetros. O resultado para o modelo com y linear foi 0,721 e para o modelo com y logarítmico foi 0,8142.

Com isso, o modelo com a variável `volume` transformada por logaritmo (Figura 3) apresentou um ajuste melhor e, portanto, será um dos modelos considerados para as próximas análises.

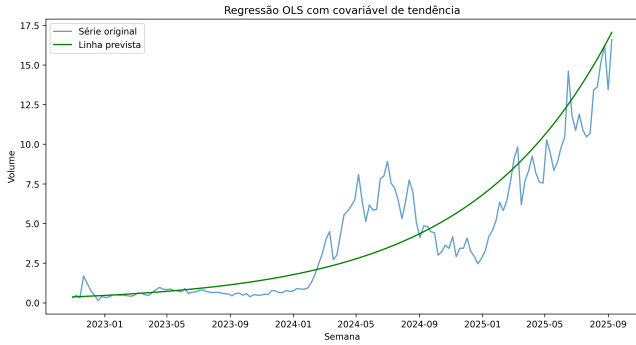


Figura 3: Ajuste com y logarítmico

2.1 Baselines

Os nossos *baselines* definem o desempenho mínimo a ser superado. Incluímos aqui Média, Random Walk sem Drift, Random Walk com Drift e Método Ingênuo Sazonal. Para obter uma abordagem mais completa, geramos uma **previsão distribucional**, que estima a distribuição de probabilidade completa dos possíveis resultados futuros (Figura 4).

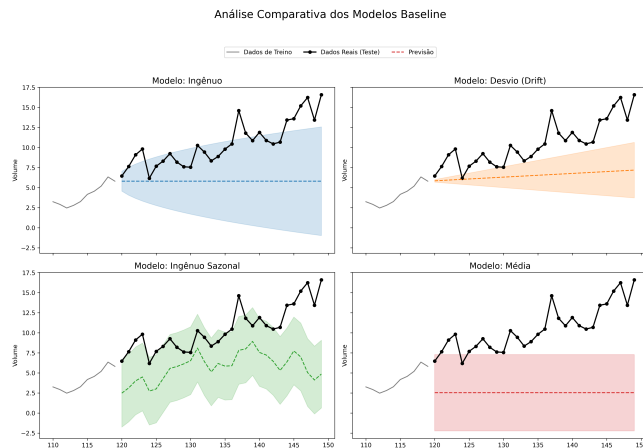


Figura 4: Previsão distribucional dos modelos Baselines

Para avaliar essa abordagem, usamos também Erro de Quantil (Quantile Score), Winkler Score e Continuous Ranked Probability Score (CRPS). Todos os *baselines* foram ajustados no conjunto de treino e avaliados no conjunto de teste. Os resultados são apresentados na tabela 1.

Tabela 1: Desempenho dos modelos *baseline*

Métrica	Média	Ingênuo	Desvio (Drift)	Ingênuo Sazonal
MAE	7.96	4.68	3.97	4.76
RMSE	8.43	5.44	4.66	5.40
MAPE (%)	74.25	40.79	34.40	44.01
MASE	2.37	1.39	1.18	1.42
Winkler Score	140.40	38.95	90.55	55.11
Avg Quantile Score	3.32	1.73	1.73	1.83
CRPS	6.61	3.48	3.45	3.67

Como fica evidente, o modelo com desvio de **Random Walk com Drift** apresentou o melhor desempenho geral entre os *baselines*. Agora, vamos tentar melhorar esses resultados.

2.2 Regressão Linear Múltipla com Covariáveis

Os modelos de **Regressão Linear Múltipla** foram desenvolvidos utilizando covariáveis baseadas nas variáveis `week` (Trend) e `volume` (Lags, Médias Móveis, Desvio Padrão Móvel). A estimação foi realizada via *Mínimos Quadrados Ordinários (OLS)*.

Utilizamos a covariável básica de **trend**, em conjunto com as *dummies* mensais, e, por fim, adicionamos os **lags** de 1, 4 e 12 semanas. Além disso, também incluímos as médias móveis (simples e exponenciais) de 4 e 12 semanas, e a volatilidade (desvio padrão móvel) de 4 e 12 semanas.

Visto que a série temporal não apresenta sazonalidade anual, não adicionamos variáveis muito distantes no tempo.

Portanto, as covariáveis criadas foram:

- **Trend**: Tendência linear.
- **Dummies** mensais (`mes_x`).
- **Lags**: Lag_1, Lag_4, Lag_{12} .
- **Médias Móveis** (Simples e Exponenciais): $SMA_4, SMA_{12}, EMA_4, EMA_{12}$.
- **Volatilidade** (Desvio Padrão Móvel): $StdDev_4, StdDev_{12}$.

3. Métricas e Métodos de Avaliação

As métricas de avaliação serão utilizadas para comparar o desempenho fora da amostra (*out-of-sample*) dos modelos, utilizando os erros de previsão $e_{T+h} = y_{T+h} - \hat{y}_{T+h|T}$. Consideraremos métricas de previsão **pontual**, que avaliam a precisão da estimativa média, e **distribucional**, que estima a distribuição completa dos resultados futuros.

3.1 Métricas de Previsão Pontual

- **RMSE (Raiz do Erro Quadrático Médio)**: critério de ordenação principal; penaliza mais os erros grandes e é minimizado pela média.
- **MAE (Erro Absoluto Médio)**: fácil de interpretar, pois representa o erro médio na mesma escala dos dados.

Tabela 2: Desempenho dos modelos de Regressão Linear com diferentes conjuntos e combinações de covariáveis

Modelo	MAE	RMSE	MAPE (%)	MASE	Winkler Score	Avg Quantile Score	CRPS
Médias Móveis	1.368	1.728	13.19	1.006	19.94	0.546	1.094
Lags	1.373	1.757	12.98	1.010	22.26	0.546	1.098
Desvio Padrão Móvel	3.736	4.495	33.74	2.747	78.07	1.584	3.166
Tendência	3.832	4.505	33.23	2.817	53.73	1.515	3.036
Meses	7.277	7.720	68.14	5.350	120.33	3.005	5.980
Lags + Médias Móveis	1.448	1.858	13.81	1.064	25.06	0.599	1.197
Lags + Médias Móveis + Tendência	1.444	1.877	13.68	1.062	25.72	0.596	1.192
Lags + Médias Móveis + Desvio Padrão	1.529	2.002	14.42	1.124	28.61	0.638	1.276
Tendência + Transformação log	1.802	2.075	18.93	1.324	25.89	0.798	1.756
Modelo Final	1.414	1.774	13.44	1.040	22.90	0.565	1.135

- **MASE (Erro Absoluto Médio Escalado):** essencial para comparar a performance dos modelos de Regressão Linear Múltipla com os *baselines*; é livre de escala e valores menores que 1 indicam melhor desempenho que o *baseline* ingênuo sazonal.
- **MAPE (Erro Percentual Absoluto Médio):** mostra o erro médio percentual, facilitando comparações relativas.

3.2 Métricas de Previsão Distribucional

- **Winkler Score:** avalia a qualidade de um intervalo de previsão (95% neste trabalho); penaliza a falta de calibração e favorece intervalos estreitos. Valores menores indicam intervalos mais precisos.
- **Erro de Quantil (*Quantile Score*):** mede a precisão da previsão para um quantil específico (e.g., 0.1, 0.5, 0.9); para $p = 0.5$ (mediana), reduz-se ao MAE.
- **CRPS (Continuous Ranked Probability Score):** métrica abrangente que avalia a precisão da distribuição de previsão como um todo; valores menores indicam melhor calibração e nitidez.

4. Resultados

Pelos resultados apresentados na Tabela 2, observa-se que o modelo baseado em **médias móveis** apresentou o melhor desempenho. Em seguida, o modelo com **lags** também se destacou, superando significativamente os modelos que utilizam apenas **trend** e **volatilidade**. Por outro lado, o modelo com **indicadoras mensais** apresentou o desempenho mais baixo entre todos.

Contudo, ao analisar os parâmetros dos melhores modelos apresentados na tabela de regressão (Tabela 3), observa-se um problema no modelo baseado em **médias móveis**: o parâmetro `volume_ema_12` apresenta um *p-value* elevado e seu intervalo de confiança inclui zero, indicando que não é estatisticamente significativo.

Diante disso, propomos a criação de novos modelos, removendo esse parâmetro e combinando diferentes grupos de va-

Tabela 3: Parâmetros do modelo de Médias Móveis

Coef.	coef	std err	t	P> t	[0.025	0.975]
const	0.134	0.102	1.316	0.191	-0.068	0.336
volume_sma_4	-1.331	0.364	-3.660	0.000	-2.053	-0.610
volume_sma_12	-0.752	0.430	-1.747	0.084	-1.605	0.102
volume_ema_4	2.638	0.386	6.838	0.000	1.873	3.403
volume_ema_12	0.420	0.540	0.777	0.439	-0.651	1.491

riáveis, a fim de avaliar o impacto no desempenho. Os resultados foram também colocados na Tabela 2.

A partir disso, ao analisar os diferentes grupos de modelos, observa-se que nenhum deles apresenta melhora significativa se comparados aos modelos de **lags** e **médias móveis** básicos. Além disso, a adição de variáveis de **tendência** e **volatilidade** não trouxe melhorias significativas, indicando que sua inclusão não é necessária para aprimorar o modelo.

Novamente, vamos analisar os parâmetros do modelo de Lags apresentados na tabela de regressão (Tabela 4).

Tabela 4: Parâmetros do modelo de Lags

Coef.	coef	std err	t	P> t	[0.025	0.975]
const	0.157	0.091	1.717	0.089	-0.024	0.338
lag_1	0.898	0.057	15.719	0.000	0.785	1.011
lag_4	0.166	0.072	2.321	0.022	0.024	0.308
lag_12	-0.109	0.044	-2.505	0.014	-0.195	-0.023

Observa-se que seus parâmetros são todos estatisticamente significativos, mas possuem p-valores maiores que o modelo de médias móveis simplificado.

A Figura 5 apresenta uma representação visual de todos os modelos ajustados, incluindo suas previsões no conjunto de teste.

4.1 Análise dos resíduos e ajuste dos modelos Baseline

Avaliando os modelos baseline, observamos os seguintes pontos:

Métricas de erro:

O Modelo Média apresentou desempenho claramente infe-

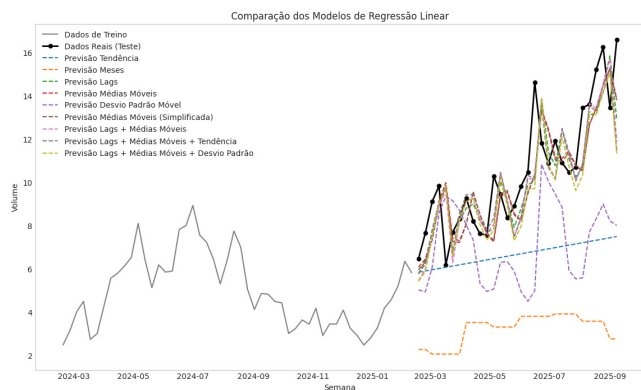


Figura 5: Modelos de Regressão Linear no conjunto de teste

rior, com MAE de 7.96, RMSE elevado e MAPE de 75%, indicando erros substanciais. Os demais modelos (Ingênuo, Desvio/Drift e Ingênuo Sazonal) apresentam métricas similares, com MAE próximo a 4 e MAPE em torno de 35%, demonstrando ajuste melhor, porém ainda limitado.

Média dos resíduos:

A média dos resíduos varia de 3.97 (Drift) a 7.96 (Média), representando entre 37% e 76% da média dos dados (10.49). Idealmente, a média dos resíduos deveria estar próxima de zero, indicando que nenhum modelo baseline capturou adequadamente a série.

Distribuição dos resíduos:

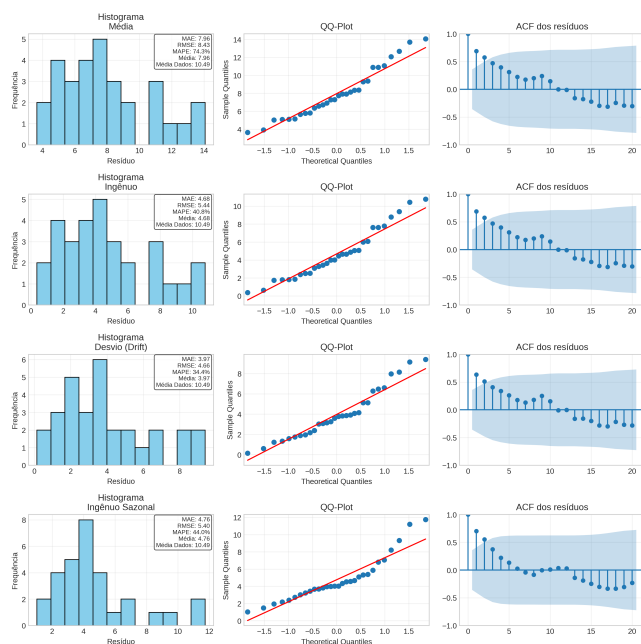


Figura 6: Distribuição dos resíduos dos modelos baseline.

- **Histograma:** A distribuição dos resíduos não segue a normal, apresentando assimetria e concentração à esquerda.

- **QQ-Plot:** Pontos fora da diagonal, especialmente nas extremidades, confirmam caudas mais pesadas que o esperado.

Autocorrelação (ACF): Todos os modelos apresentam autocorrelação positiva nos primeiros lags, seguida de queda gradual. Isso indica dependência temporal não capturada pelos

modelos, evidenciando que os baselines estão subajustados.

Conclusão: Os modelos baseline servem como uma referência mínima, mas não são capazes de capturar a estrutura temporal da série.

4.2 Análise dos resíduos e ajuste dos modelos de regressão

Métricas de erro:

Entre os modelos avaliados, os de melhor desempenho são aqueles que incluem **lags** e **médias móveis**, com MAE entre 1.37 e 1.45, RMSE entre 1.73 e 1.86 e MAPE próximo de 13–14%. O **Modelo Médias Móveis (Simplificado)**, que exclui parâmetros não significativos, manteve desempenho consistente (MAE = 1.3870, RMSE = 1.7325, MAPE = 13.41%). Os modelos **Meses**, **Tendência** e **Desvio Padrão Móvel** se destacam negativamente, com erros significativamente maiores, sendo incapazes de capturar a correlação temporal.

Média dos resíduos:

O **Modelo Final** apresenta média de resíduos próxima de zero (0.50), enquanto os modelos fracos (Meses, Tendência e Desvio Padrão) possuem médias mais altas, confirmando ajuste ruim e captura insuficiente da estrutura da série.

Distribuição dos resíduos:

- **Histograma:** Apenas Tendência e Meses apresentam distorção clara da normalidade.

- **QQ-Plot:** Pontos fora da diagonal nas extremidades, reforçando caudas mais pesadas em alguns modelos.

Autocorrelação (ACF): Modelos que combinam Lags e Médias Móveis capturam adequadamente a dependência temporal, enquanto os modelos fracos apresentam autocorrelação residual positiva, indicando subajuste.

Por fim, apresentamos a distribuição dos resíduos do **Modelo Final** na Figura 7. As distribuições dos demais modelos podem ser consultadas no material completo disponível no GitHub.

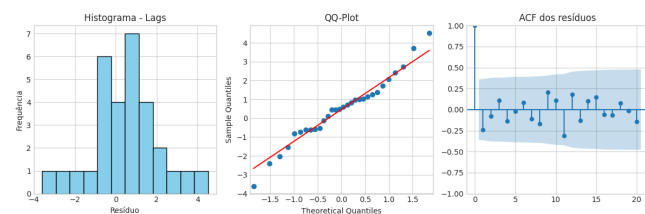


Figura 7: Distribuição dos resíduos do Modelo Final de Regressão Linear no conjunto de teste.

Conclusão: O **Modelo Final**, baseado em Lags e Médias Móveis, consegue capturar a estrutura temporal da série, apresenta resíduos equilibrados e erros reduzidos, sendo a melhor escolha para previsões futuras.

5. Conclusão

Os modelos de baseline efetivamente serviram apenas como referência, apresentando desempenho insatisfatório. Entre os modelos de regressão múltipla, os modelos **Meses**, **Tendência** e **Desvio Padrão Móvel** mostraram desempenho fraco em todas as métricas e, portanto, não devem ser utilizados, especialmente considerando o bom desempenho de outros modelos.

O modelo **Lags + Médias Móveis + Desvio Padrão** apresentou média de resíduos maior que os demais e desempenho ligeiramente inferior em outras métricas, não sendo a escolha ideal.

Os quatro últimos modelos apresentaram desempenho bastante semelhante. Dentre eles, o modelo **Lags** se destaca, com excelente MAPE e média dos resíduos, e MAE e RMSE muito próximos do melhor. Portanto, o modelo **Lags** é considerado o de melhor ajuste para os dados analisados.