



UNIVERSIDADE
FEDERAL DO CEARÁ



MODELAGEM E PREVISÃO DA PRODUÇÃO DE PETRÓLEO NO BRASIL

SÉRIES TEMPORAIS - 2025.1

PROF^a DR^a JENIFFER JOHANA DUARTE SANCHEZ

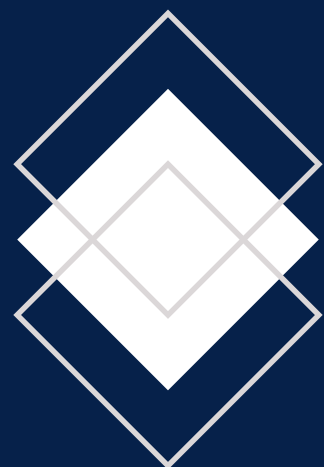
Departamento de Estatística e Matemática Aplicada - DEMA

30 de julho de 2025



Douglas Chaves Moura

Bacharelado em Estatística



Sumário:

- 03.** Introdução
- 04.** Metodologia
- 05.** Análise Exploratória
- 06.** Pré-processamento da Série
- 07.** Identificação e Estimação do Modelo
- 08.** Previsão e Avaliação do Desempenho
- 09.** Conclusão



Introdução

Introdução

ANÁLISE DA PRODUÇÃO MENSAL DE PETRÓLEO (EM M³) NO BRASIL.

◆ PERÍODO DO ESTUDO:

Janeiro de 2001 a Maio de 2025.

◆ FONTE DOS DADOS:

Agência Nacional do Petróleo (ANP) / Portal de Dados Abertos.

◆ OBJETIVO PRINCIPAL:

Identificar padrões, tendências e sazonalidades para desenvolver um modelo de previsão de valores futuros.





Metodologia

Metodologia

A ABORDAGEM

Abordagem Adotada

Metodologia **Box-Jenkins**, um método sistemático para analisar e prever séries temporais.

Análise Inicial

Estudo exploratório para compreender seus padrões fundamentais.

Modelo de Foco

Desenvolvimento de um modelo **SARIMA** para realizar as previsões.

Metodologia

FERRAMENTAS E FUNDAMENTAÇÃO

Software

- Linguagem R (Versão 4.4.2)
- RStudio (Versão 2024.12.0)

Base Teórica

- "Análise de Séries Temporais"
- Morettin e Toloí

Base Teórica

- "Time Series Analysis: Forecasting and Control"
- Box e Jenkins



Análise Exploratória

Análise Exploratória

UM RAIO-X DOS DADOS

O QUE ESTAMOS ANALISANDO?

- Volume mensal de produção de petróleo no Brasil (em m^3).
- Período: Jan/2001 a Mai/2025 (total de 293 observações).

PRINCIPAIS NÚMEROS:

- **Média Mensal:** ~10,9 milhões de m^3
- **Produção Mínima:** 5,8 milhões de m^3 (Outubro)
- **Produção Máxima:** 18,1 milhões de m^3 (Maio)
- **Alta Variabilidade:** Desvio padrão de ~3,1 milhões de m^3 indica flutuações significativas.

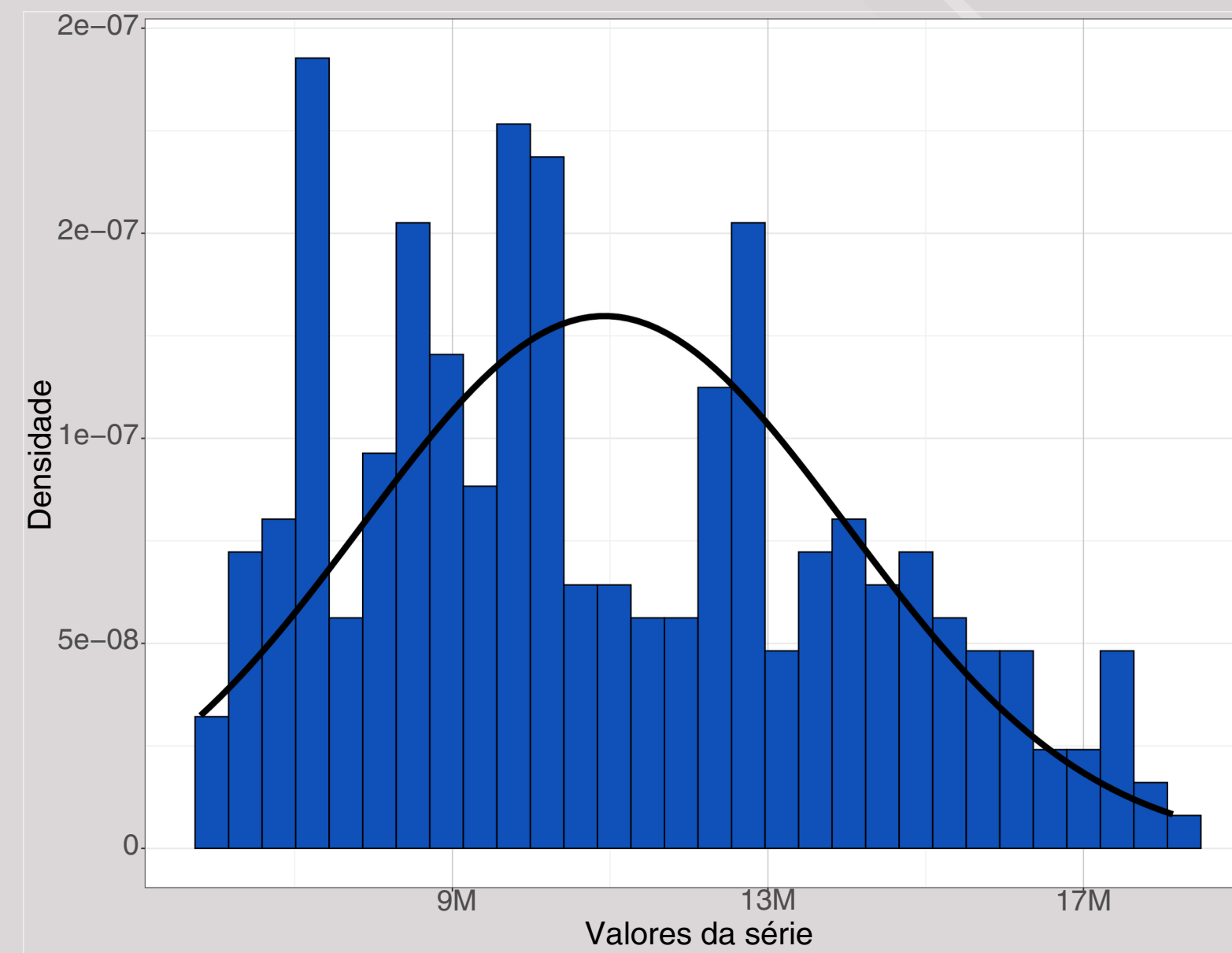


FIGURA 1: Distribuição da densidade da produção mensal de petróleo no Brasil entre jan/2001 e mai/2025.

Análise Exploratória

TRAJETÓRIA DA PRODUÇÃO DE PETRÓLEO (2001-2025)

O QUE VEMOS?

- **Clara tendência de crescimento** ao longo de mais de duas décadas.
- **Flutuações regulares** que indicam um forte componente sazonal.
- **Crescimento acentuado** entre 2021-2024, possivelmente impulsionado pela produção no pré-sal.
- A série é **visualmente não-estacionária**, o que exigirá tratamento no modelo.



Análise Exploratória

INVESTIGANDO A SAZONALIDADE

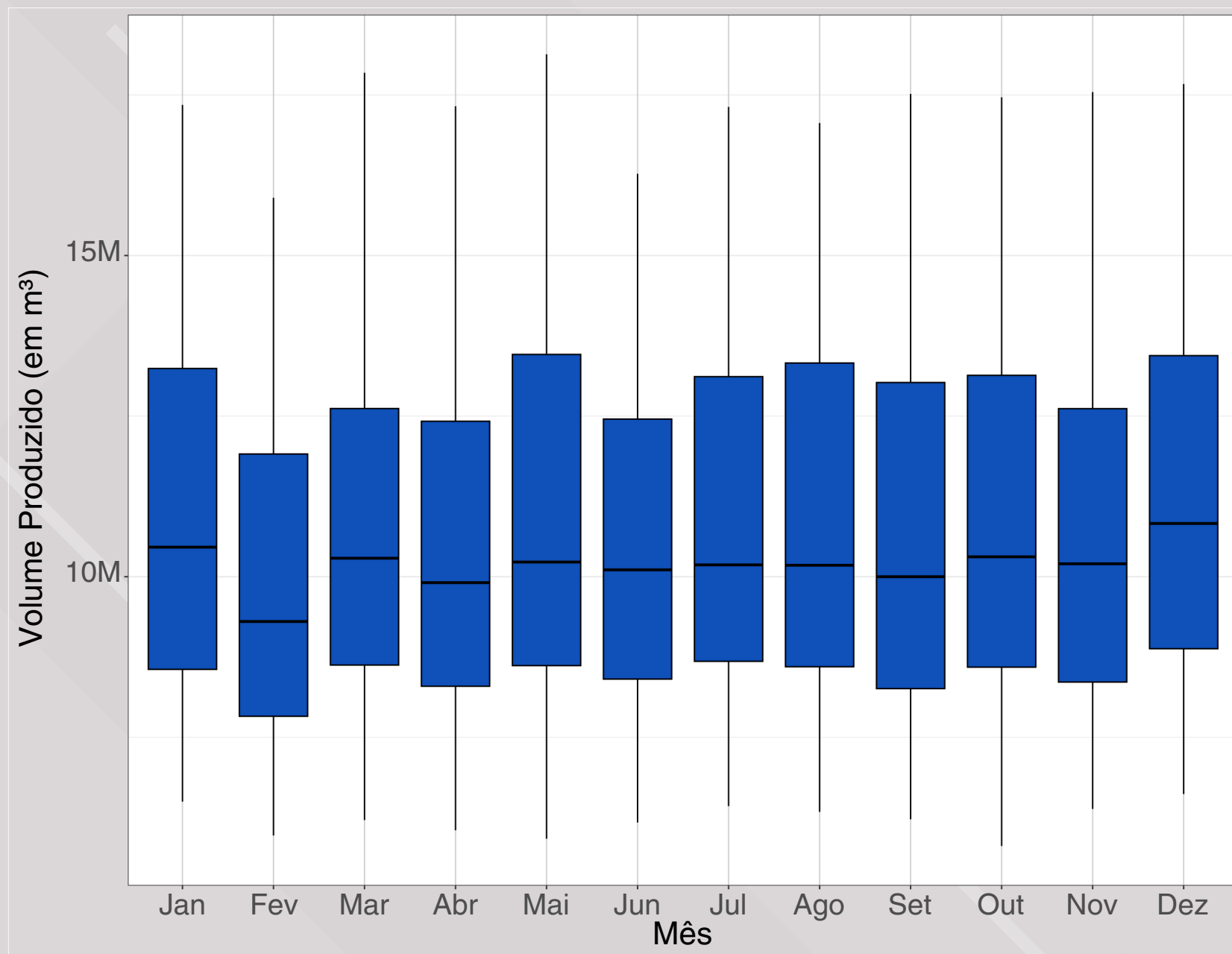


FIGURA 2: Boxplot da produção mensal de petróleo por mês.

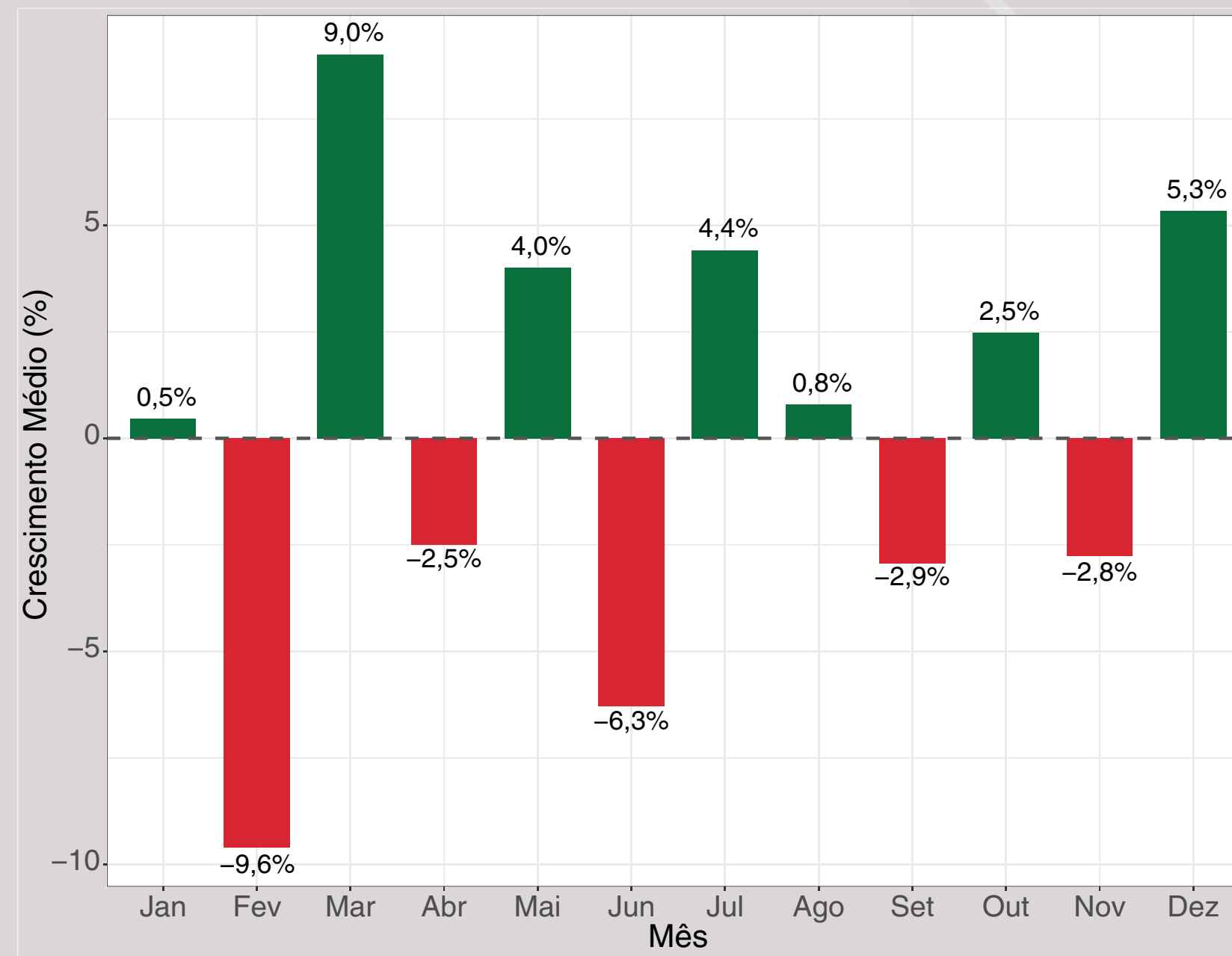


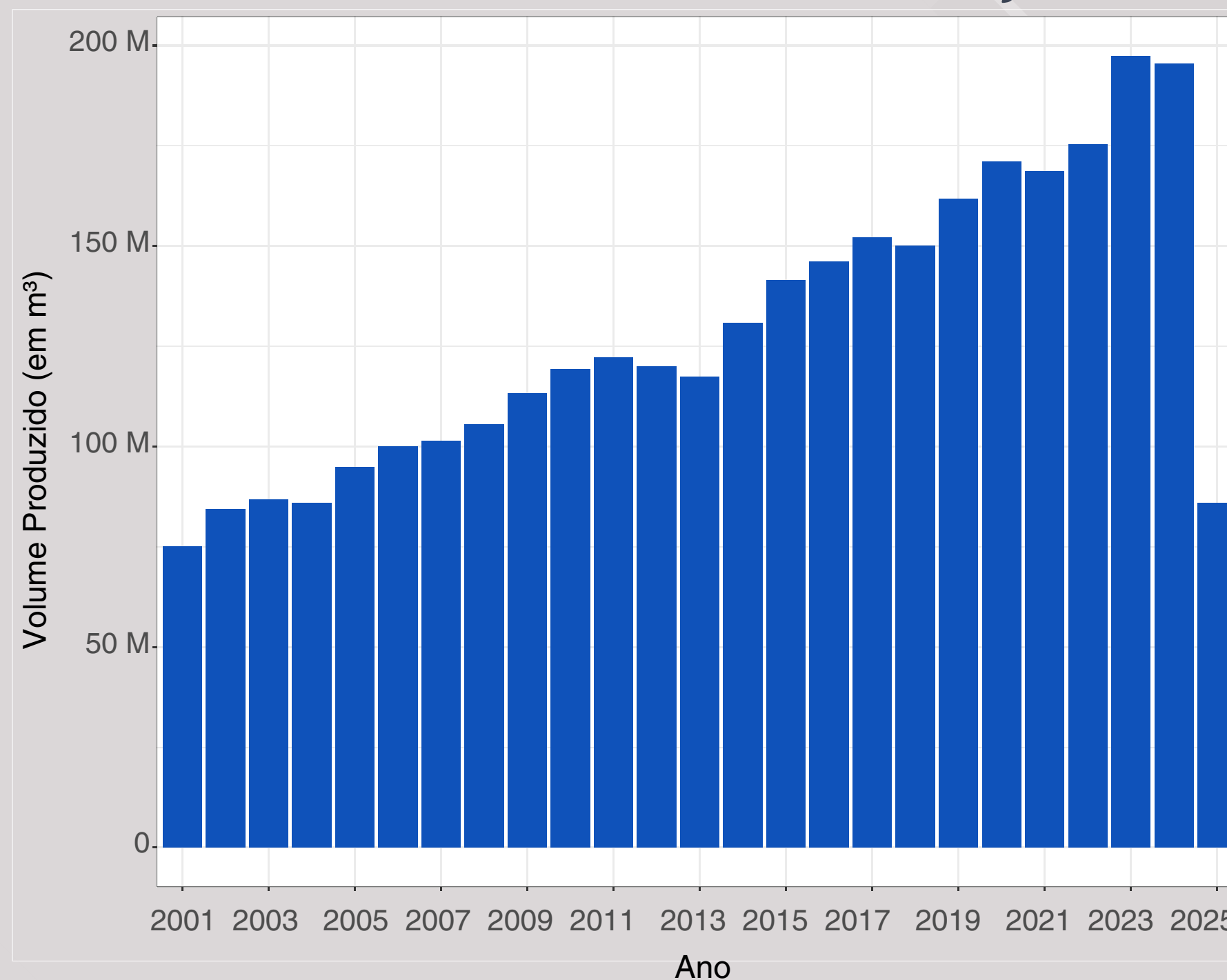
FIGURA 3: Variação percentual média da produção de petróleo em relação ao mês anterior.

Análise Exploratória

A TENDÊNCIA MACRO: CRESCIMENTO ANUAL DA PRODUÇÃO

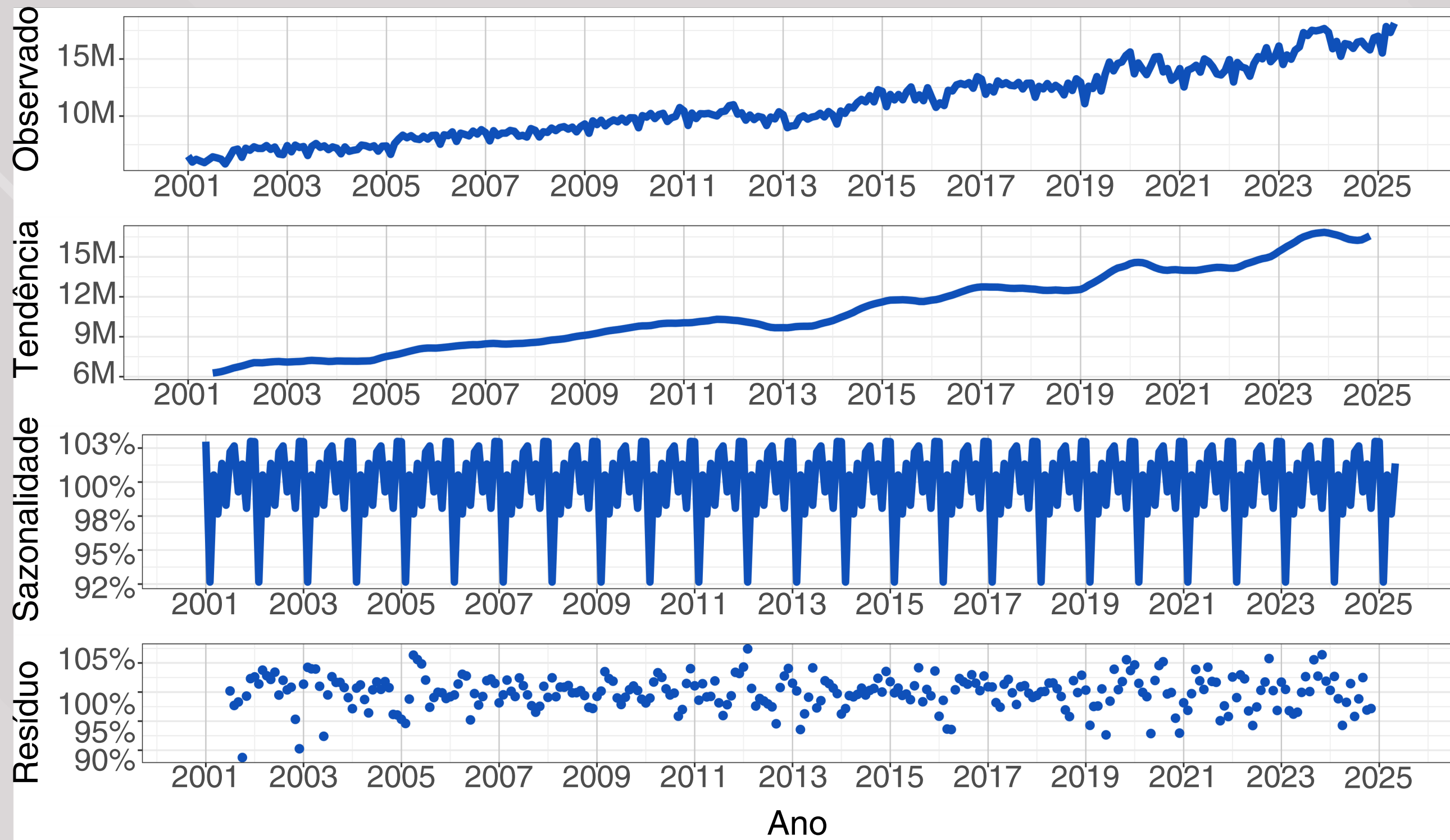
O QUE VEMOS?

- **Crescimento quase contínuo** no volume total produzido a cada ano.
- A **produção anual mais do que dobrou** desde o início da série.
- Os **picos de produção** foram atingidos nos anos mais recentes (2023 e 2024), confirmando a **forte tendência de alta**.
- **Poucos anos registraram quedas**, e todas foram de baixa magnitude.



Análise Exploratória

DECOMPOSIÇÃO DA SÉRIE TEMPORAL

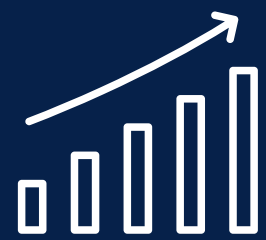




Pré-procессamento da Série

Pré-processamento

A NECESSIDADE DA ESTACIONARIEDADE



A Premissa Fundamental: A metodologia Box-Jenkins (usada no nosso modelo) exige que a série temporal seja estacionária.



O Que é Estacionariedade?

- Uma série cujas propriedades estatísticas (como média e variância) **não mudam** ao longo do tempo.
-



Nosso Desafio: A análise exploratória revelou tendência de crescimento e sazonalidade, que são claras violações da estacionariedade.

Pré-processamento

O VEREDITO DOS TESTES ESTATÍSTICOS



A SÉRIE É ESTACIONÁRIA?

Testes ADF e KPSS confirmam a presença de raiz unitária (**não estacionariedade**).



A SÉRIE POSSUI TENDÊNCIA?

Testes de Mann-Kendall e Regressão Linear apontam uma **tendência** de crescimento **positiva** e **significativa**.



A SÉRIE POSSUI SAZONALIDADE?

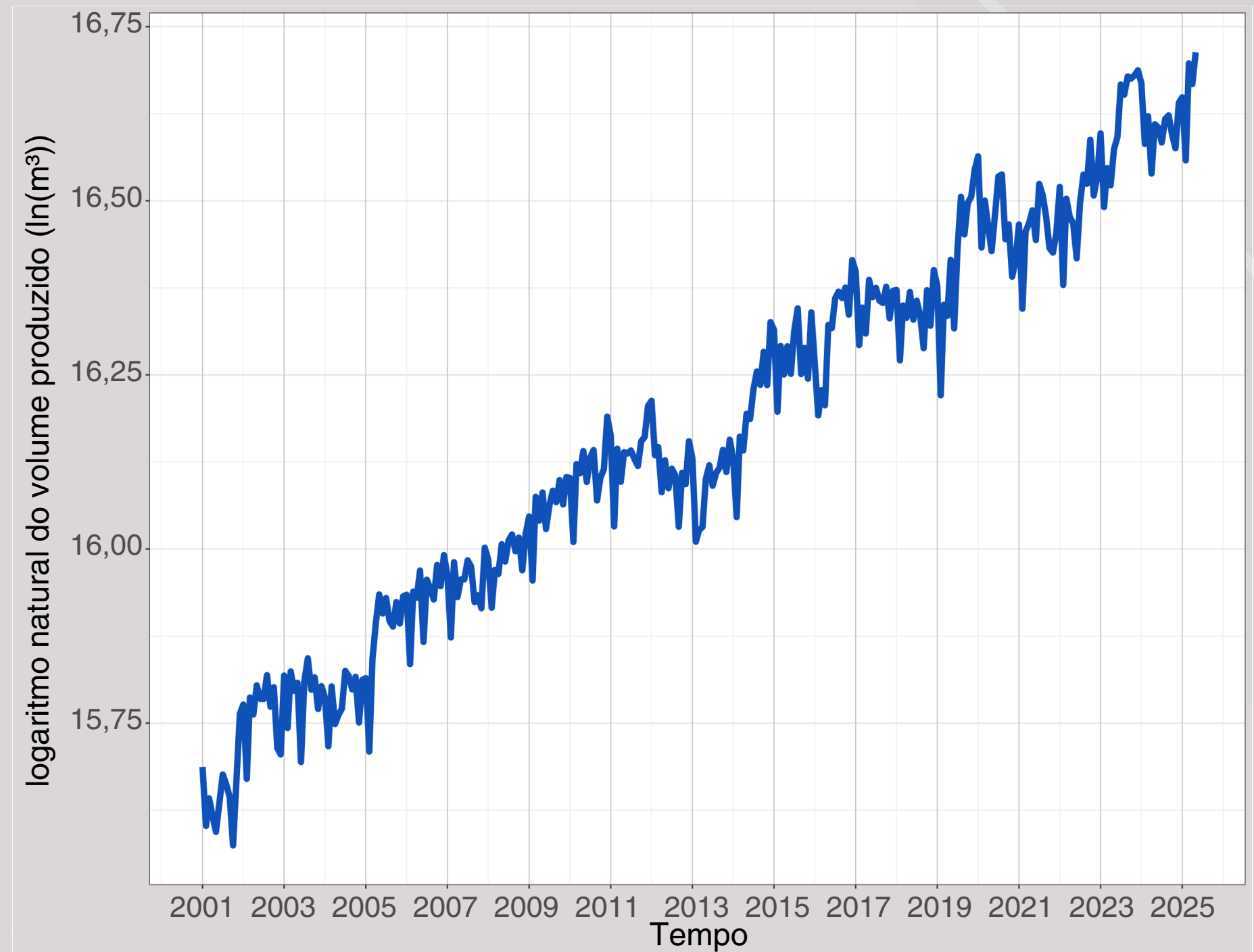
O teste QS **confirma** um forte **componente sazonal**.

Pré-processamento

PARA ESTABILIZAR A VARIÂNCIA: TRANSFORMAÇÃO LOGARÍTMICA

O QUE TEMOS?

- **Problema:** A variabilidade da produção aumentava junto com o seu nível.
- **Solução:** Aplicar o logaritmo natural na série.
- **Efeito:** As flutuações se tornam mais homogêneas e a variância é drasticamente reduzida.

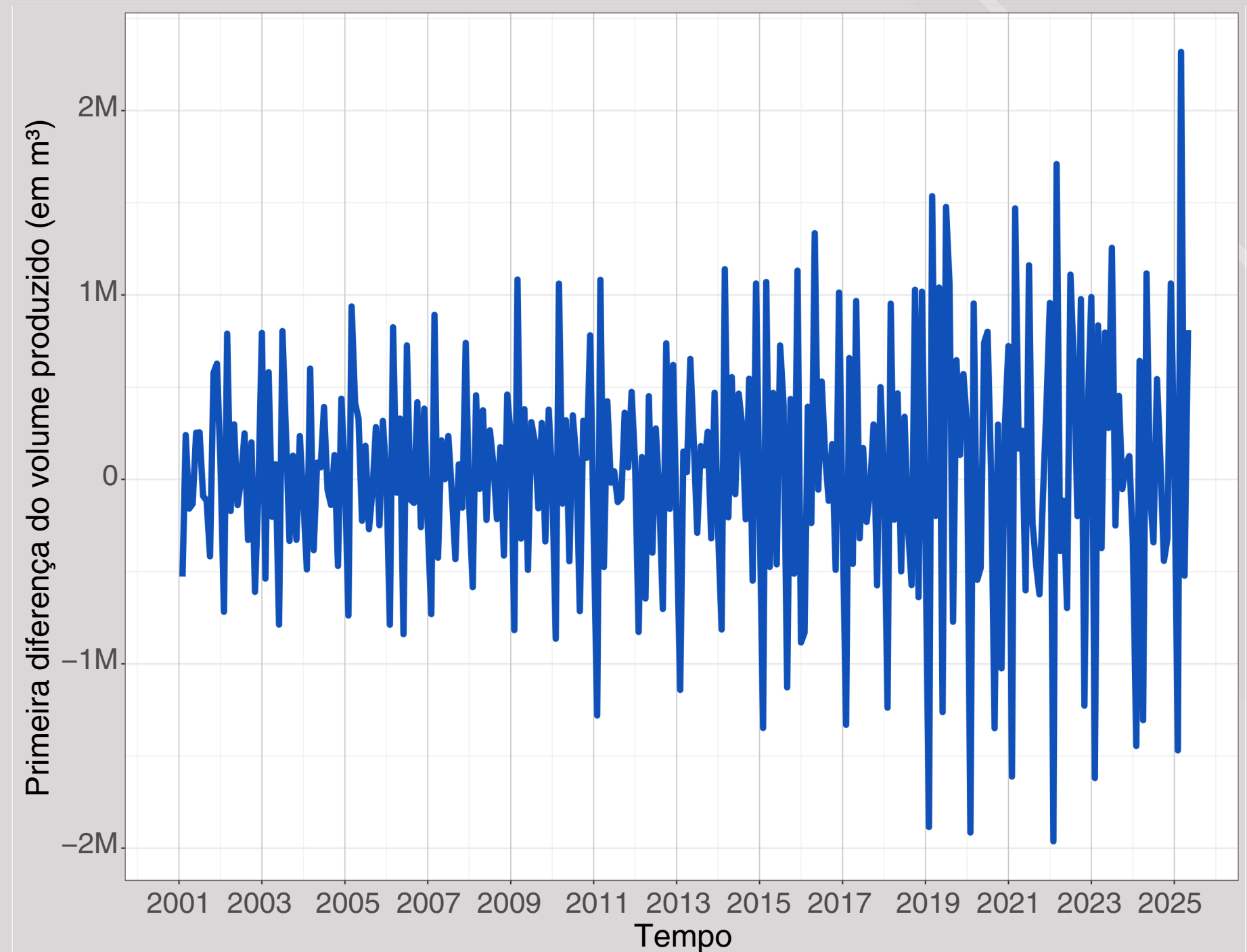


Pré-processamento

PARA REMOVER A TENDÊNCIA: DIFERENCIAÇÃO

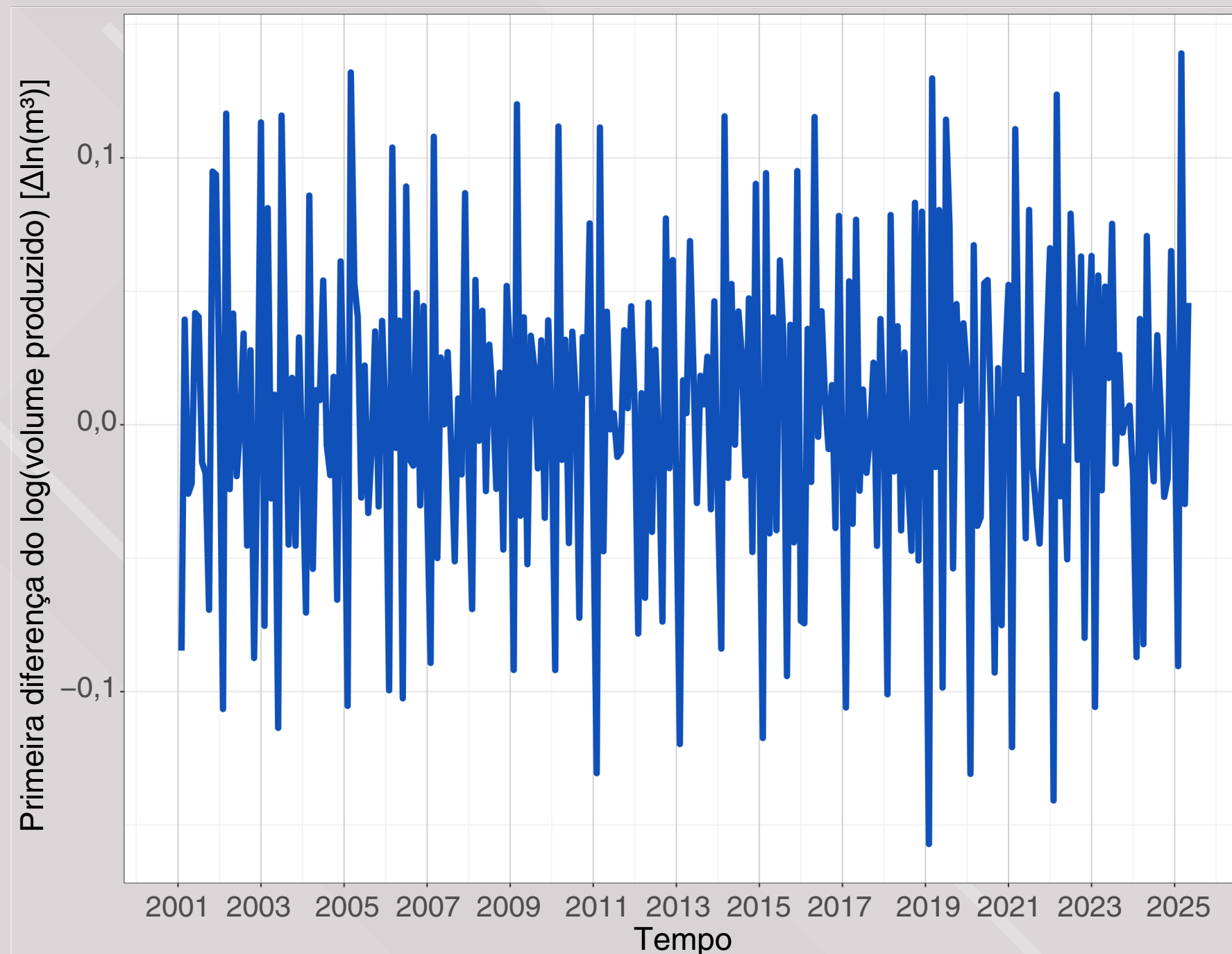
O QUE TEMOS?

- **Problema:** A média da produção cresce ao longo do tempo.
- **Solução:** Calcular a diferença entre cada mês e o mês anterior ($Y_t' = Y_t - Y_{t-1}$).
- **Efeito:** Remove a tendência de crescimento, tornando a média constante.



Pré-processamento

SÉRIE PRONTA PARA MODELAGEM



RESULTADOS:

- **Abordagem Combinada:** A solução mais robusta foi aplicar a 1ª diferença na série já transformada por logaritmo.
- **Visualmente:** O resultado é uma série estável, que flutua em torno de zero e com variabilidade homogênea.
- **Interpretação Final:** Esta nova série representa a taxa de crescimento mensal da produção de petróleo.
- **Status:** A série está estacionária na média e na variância.



Identificação e Estimação do Modelo

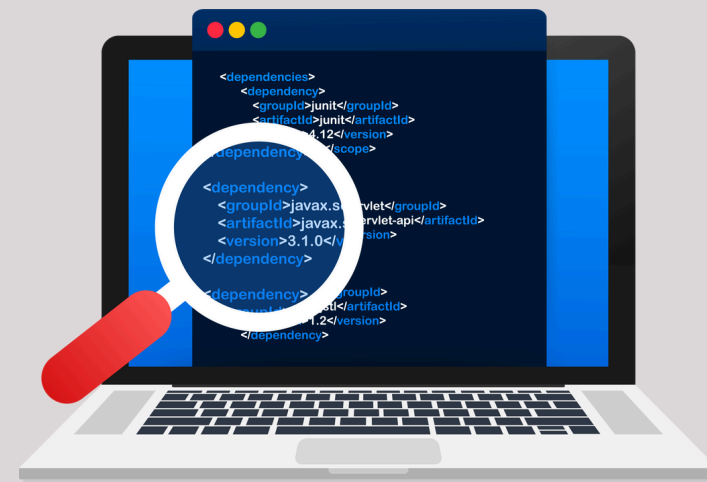
Identificação do Modelo

DEFININDO A ESTRATÉGIA

Conjunto de Treino

Janeiro de 2001 a Abril de 2021.

- Onde o modelo "aprende" os padrões históricos.



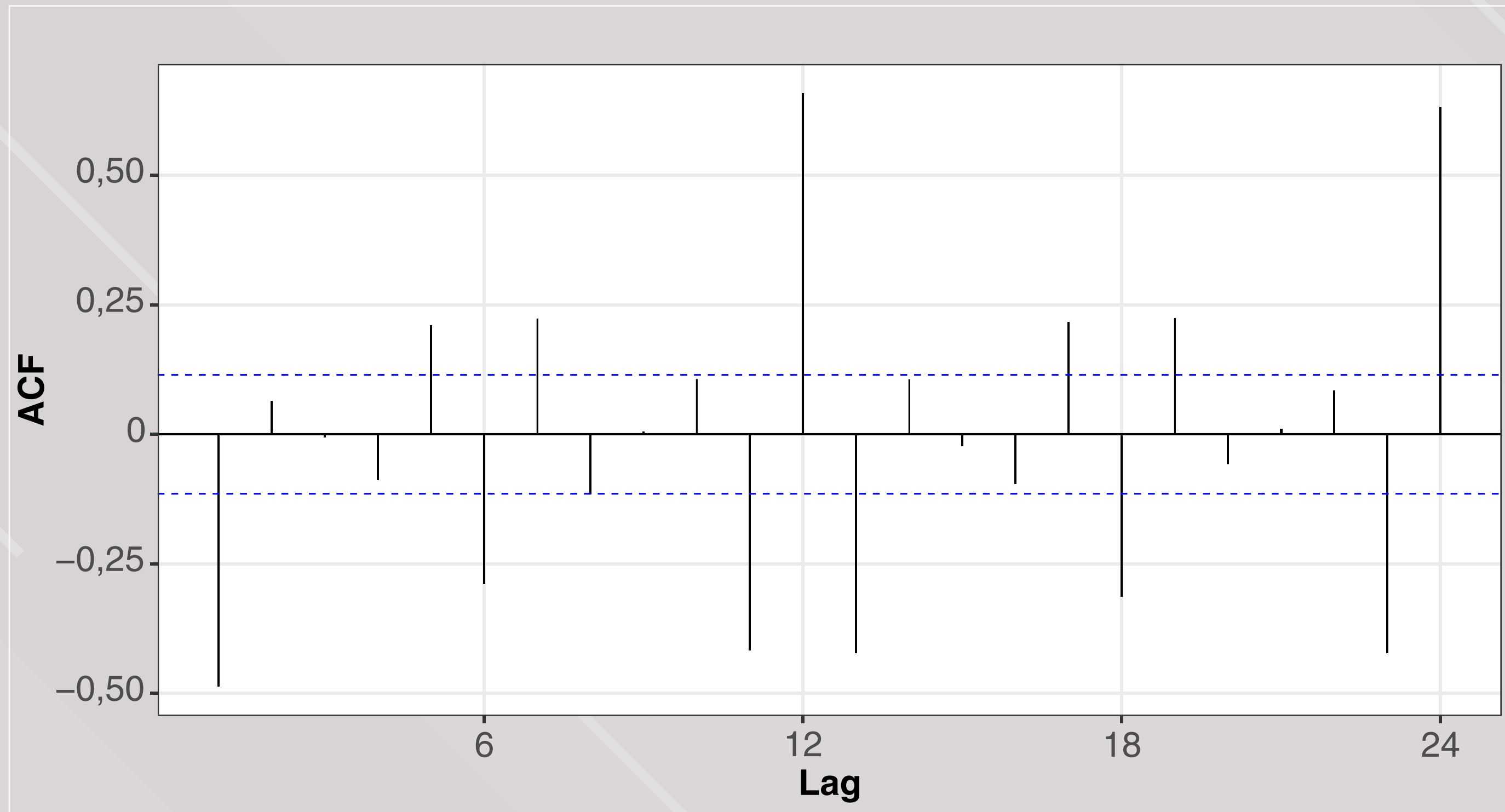
Conjunto de Teste

Maio de 2021 a Maio de 2025.

- Dados "inéditos" para validar a capacidade de previsão do modelo em um período mais recente.

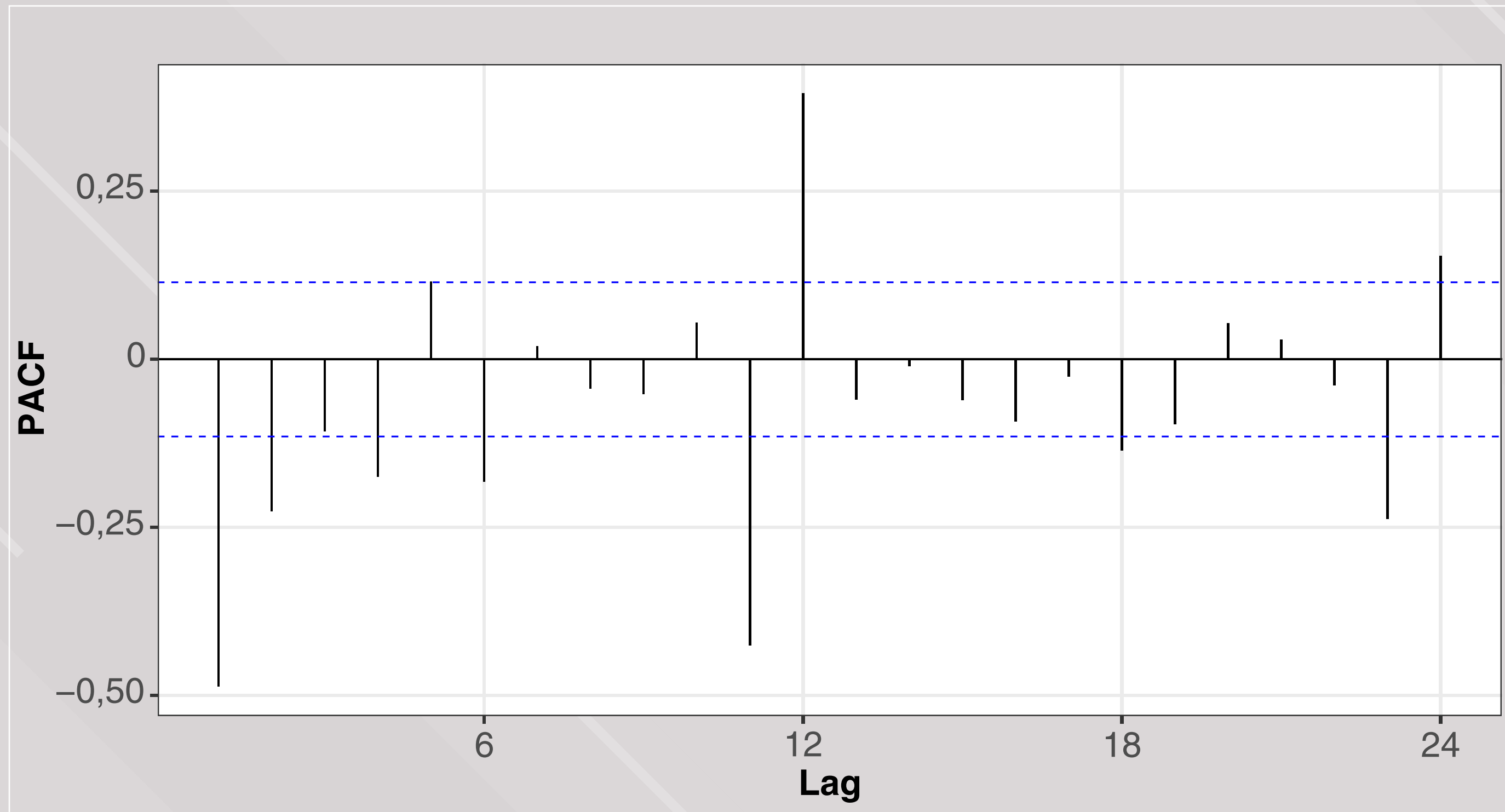
Identificação do Modelo

AUTOCORRELAÇÃO DA SÉRIE DE TREINO TRANSFORMADA



Identificação do Modelo

AUTOCORRELAÇÃO PARCIAL DA SÉRIE DE TREINO TRANSFORMADA



Identificação do Modelo

PRIMEIRO CANDIDATO: MODELO SARIMA(0,1,1)(0,1,1)₁₂

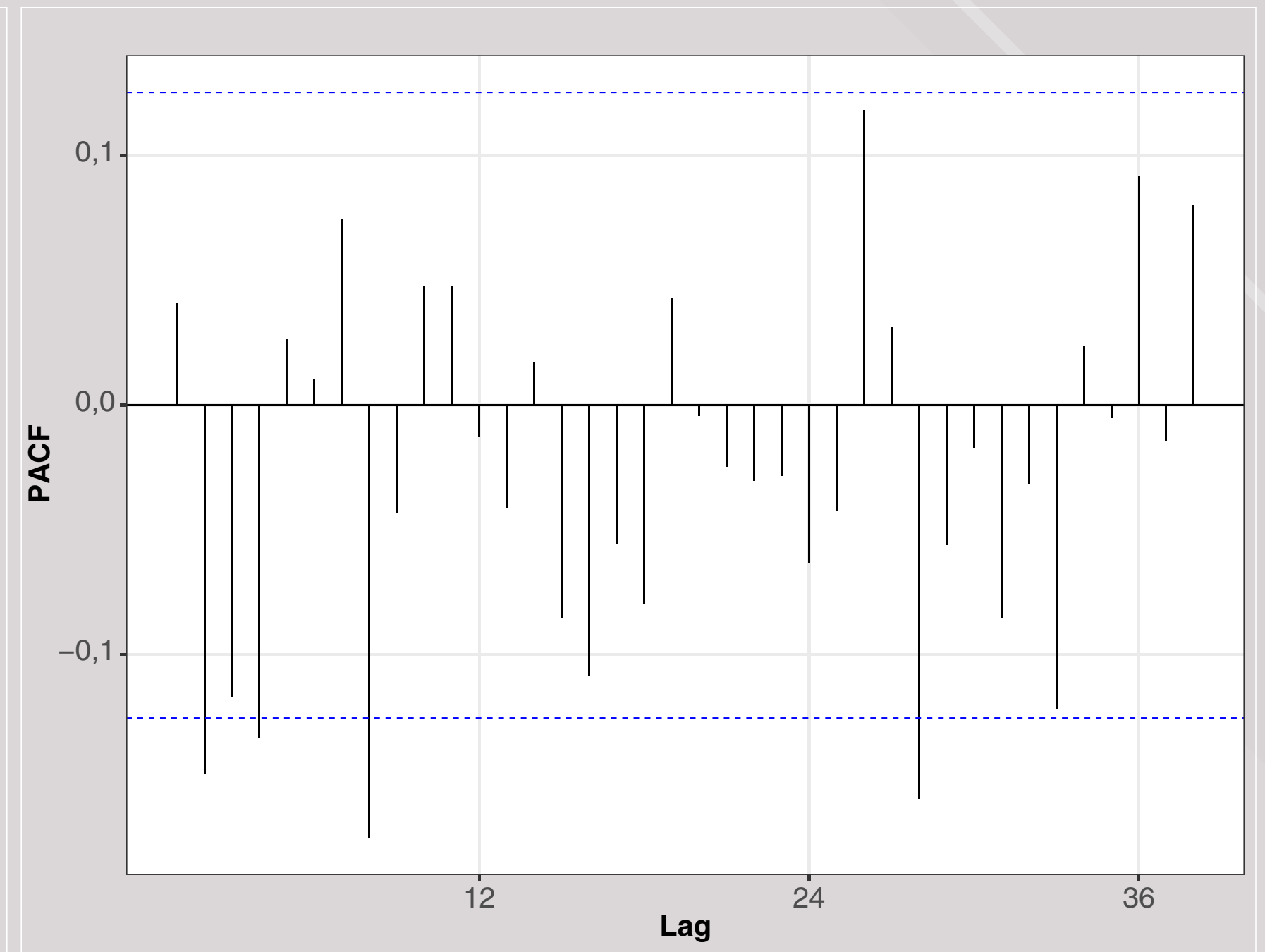
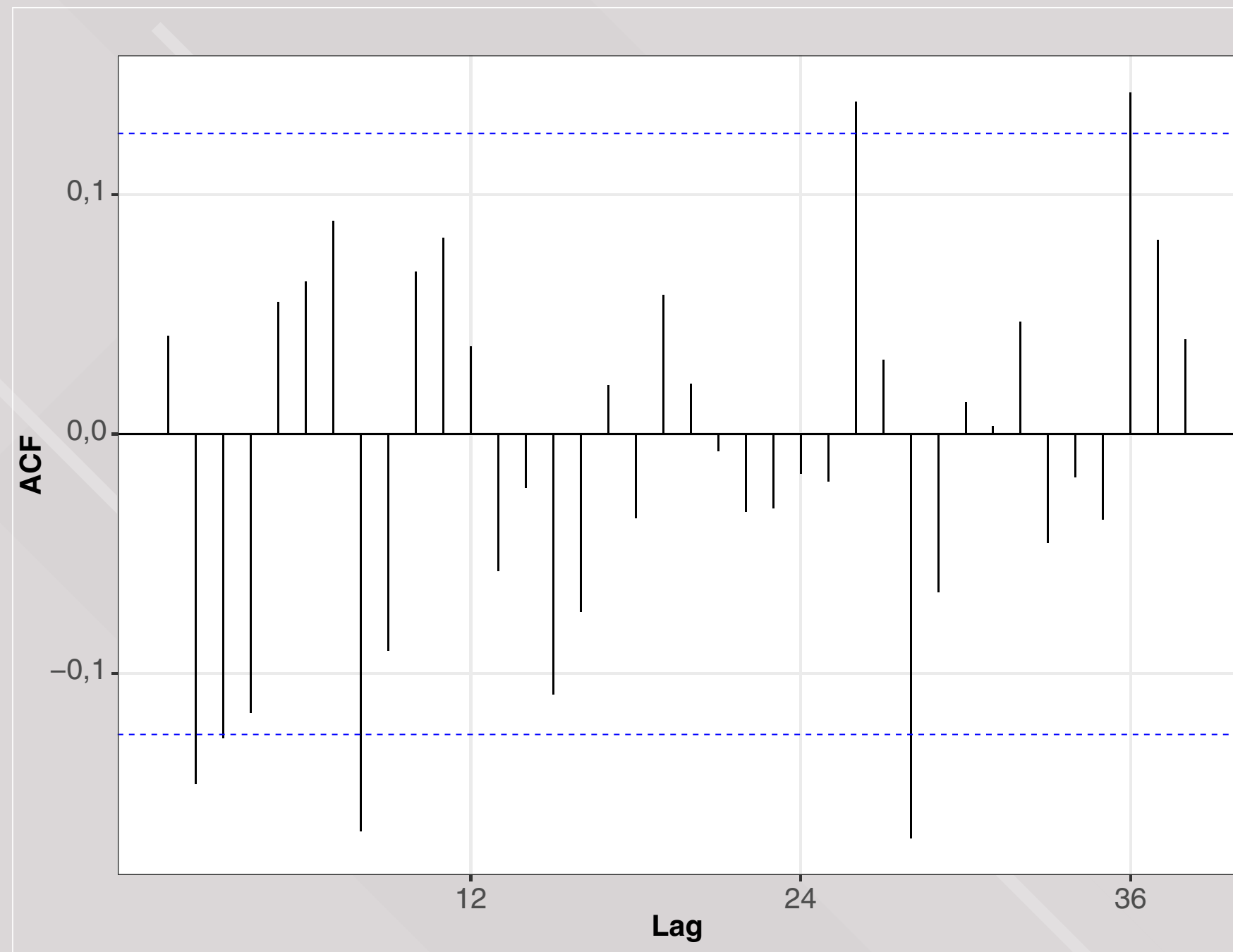
A REPRESENTAÇÃO DO MODELO SARIMA(0,1,1)(0,1,1)₁₂ PARA A SÉRIE $Z_t = \text{LOG}(\text{PRODUÇÃO}_t)$ É:

$$(1 - B)(1 - B^{12})Z_t = (1 - \theta_1 B)(1 - \Theta_1 B^{12})a_t$$

$$Z_t = Z_{t-1} + Z_{t-12} - Z_{t-13} + a_t - \theta_1 a_{t-1} - \Theta_1 a_{t-12} + \theta_1 \Theta_1 a_{t-13}$$

Identificação do Modelo

ACF E PACF DOS RESÍDUOS DO MODELO SARIMA(0,1,1)(0,1,1)₁₂

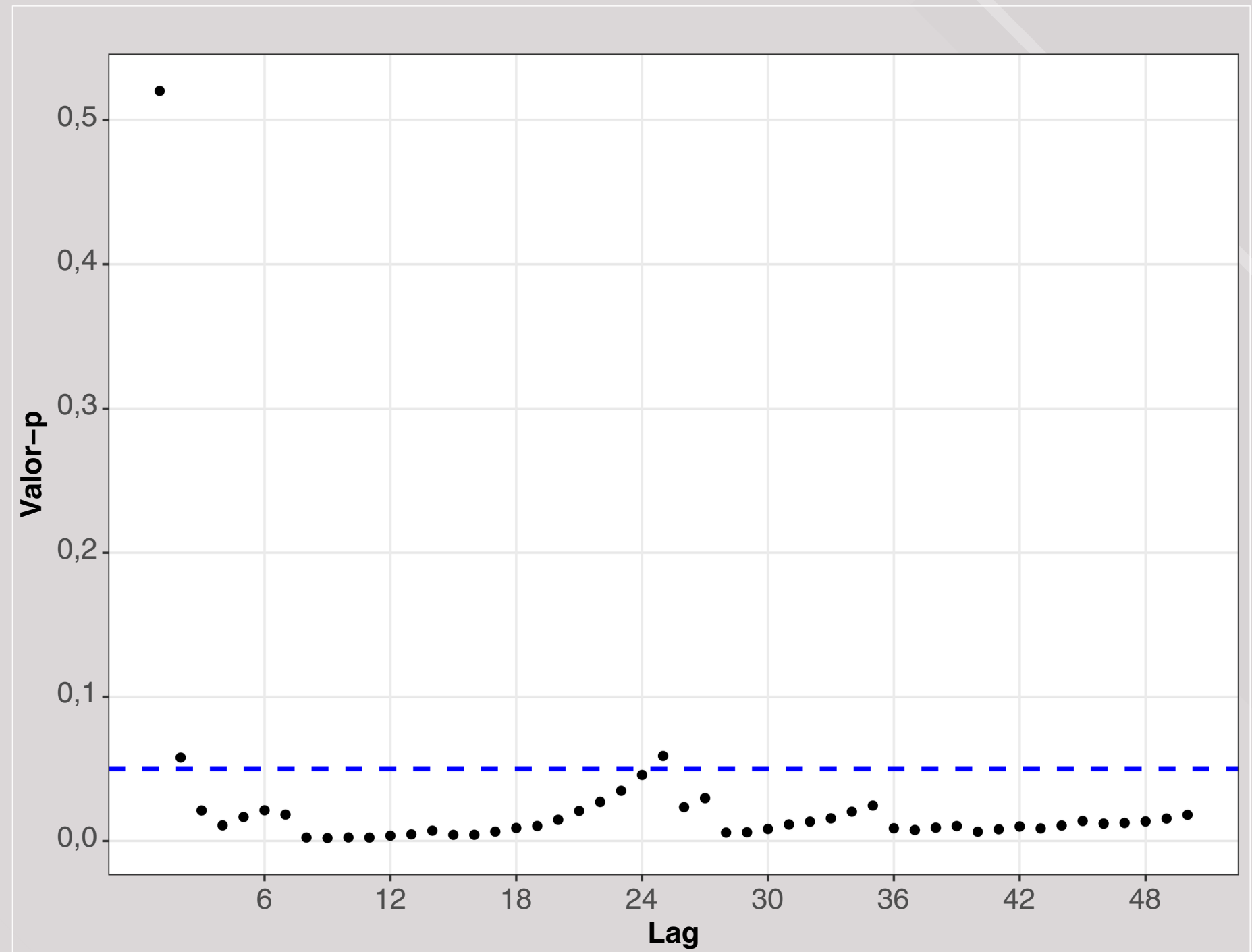


Identificação do Modelo

TESTE DE LJUNG-BOX DO MODELO SARIMA(0,1,1)(0,1,1)₁₂

RESULTADOS

- A análise da autocorrelação do modelo mostrou que ainda **havia correlação não capturada**.
- O teste de Ljung-Box confirmou que os **resíduos não são independentes**.



Identificação do Modelo

SEGUNDO CANDIDATO: MODELO SARIMA(1,1,1)(0,1,1)₁₂

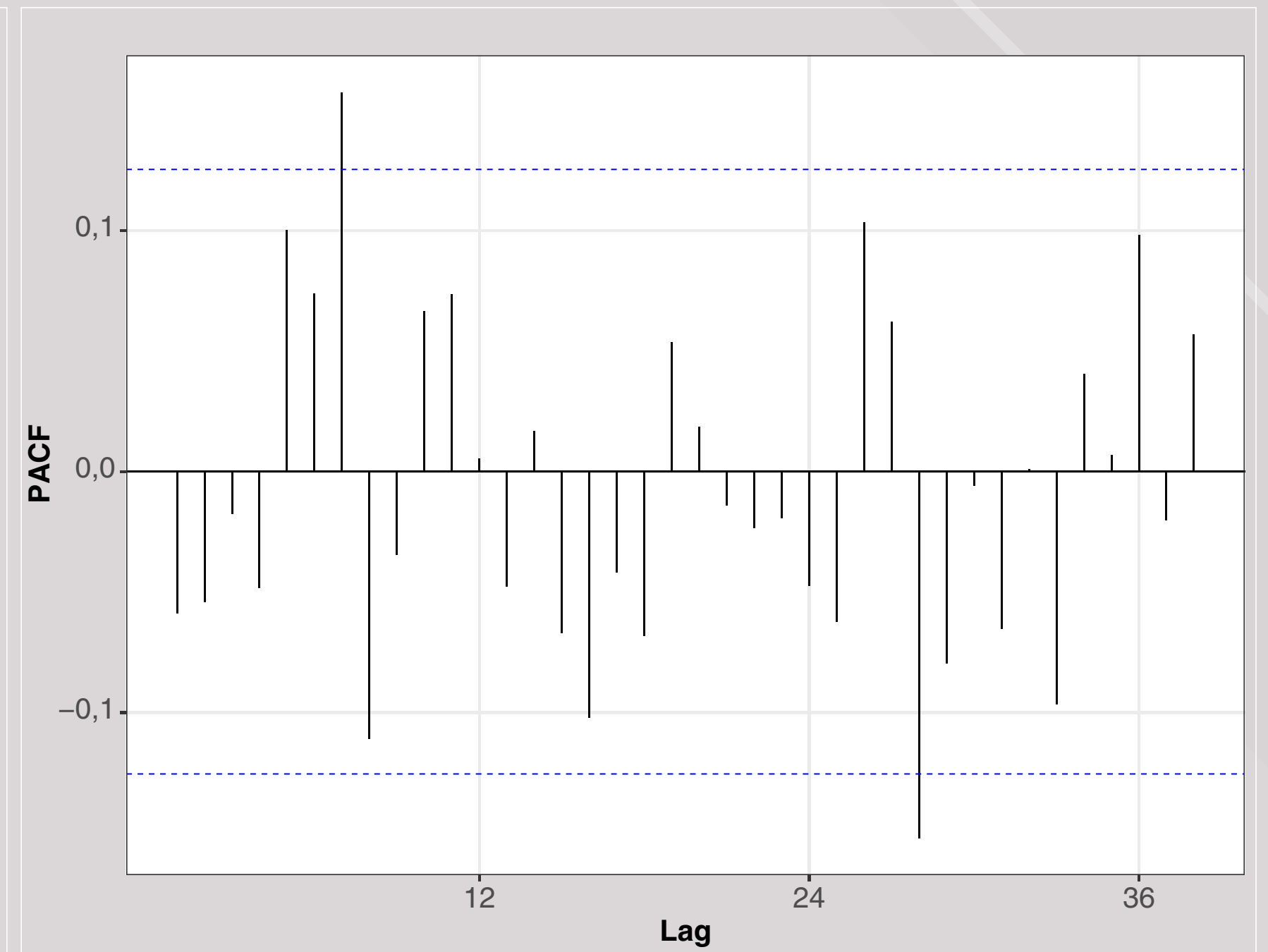
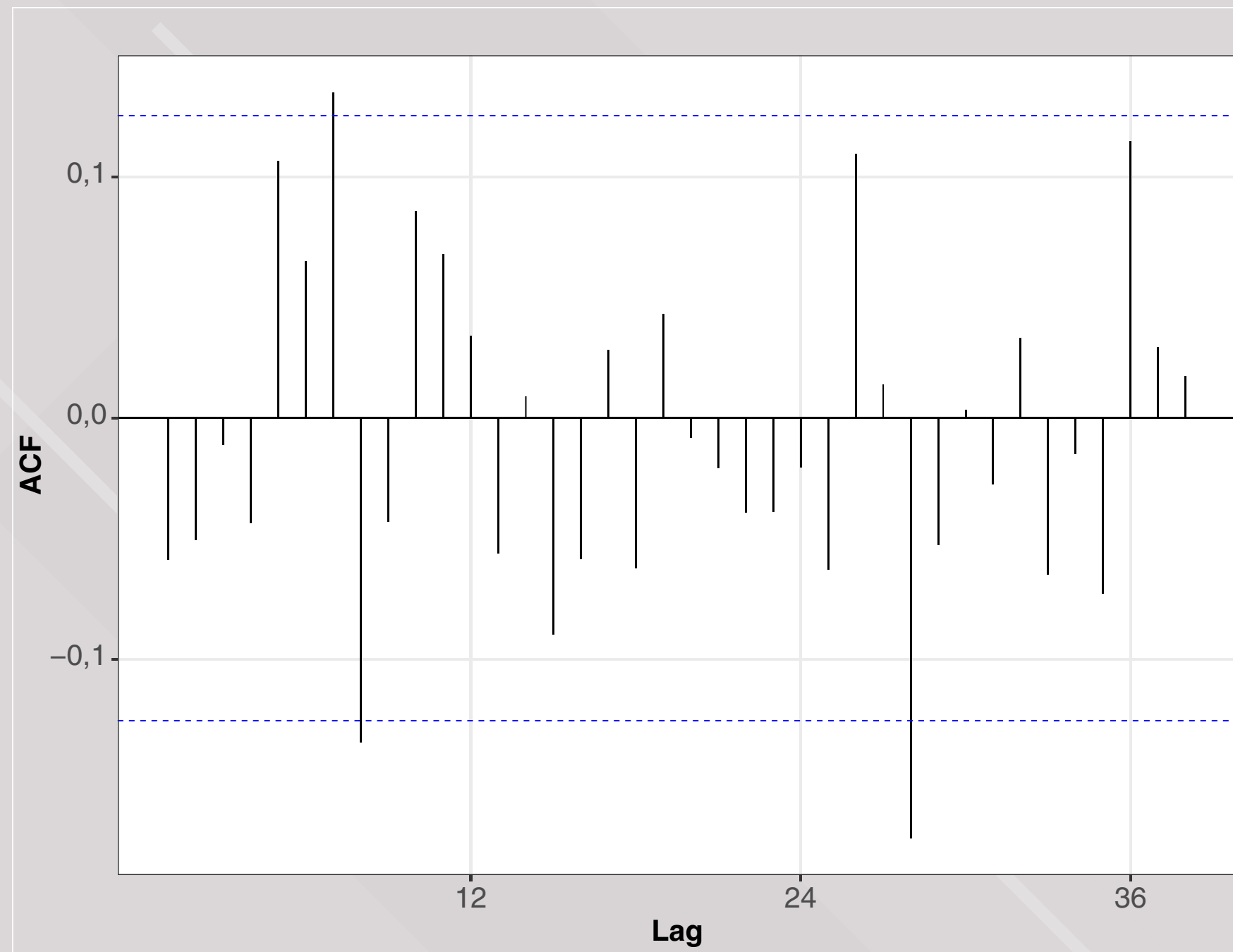
A REPRESENTAÇÃO DO MODELO SARIMA(1,1,1)(0,1,1)₁₂ PARA A SÉRIE $Z_t = \text{LOG}(\text{PRODUÇÃO}_t)$ É:

$$(1 - \phi_1 B)(1 - B)(1 - B^{12})Z_t = (1 - \theta_1 B)(1 - \Theta_1 B^{12})a_t$$

$$Z_t = \phi_1 Z_{t-1} + Z_{t-12} - Z_{t-13} + a_t - \theta_1 a_{t-1} - \Theta_1 a_{t-12} + \theta_1 \Theta_1 a_{t-13}$$

Identificação do Modelo

ACF E PACF DOS RESÍDUOS DO MODELO SARIMA(1,1,1)(0,1,1)₁₂

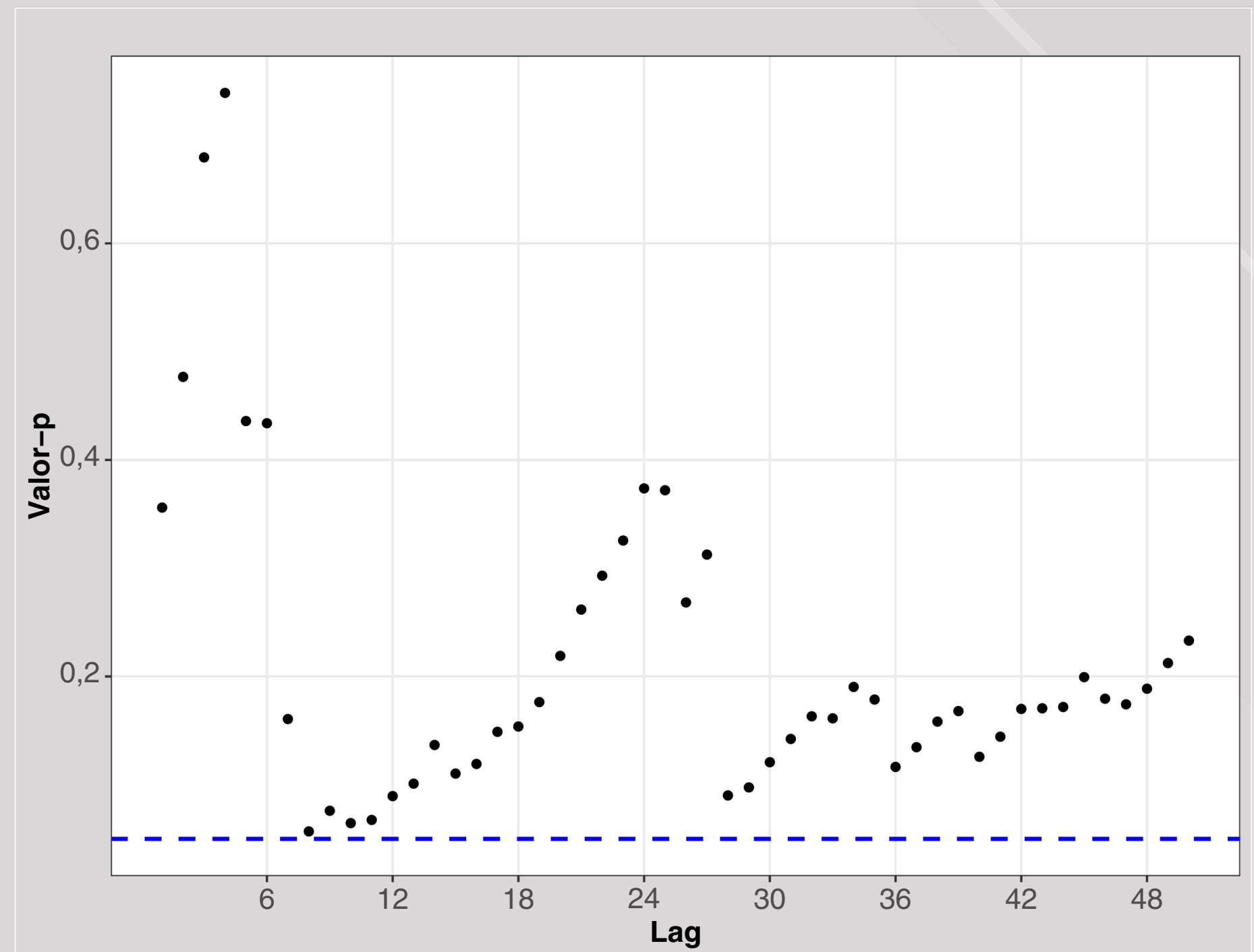


Identificação do Modelo

TESTE DE LJUNG-BOX DO MODELO SARIMA(1,1,1)(0,1,1)₁₂

RESULTADOS

- Houve uma **melhora** considerável na **captura de correlação** dos lags do modelo nos gráficos de autocorrelação.
- O teste de Ljung-Box **confirmou a independência dos resíduos**.



Identificação do Modelo

SELEÇÃO DO MODELO FINAL



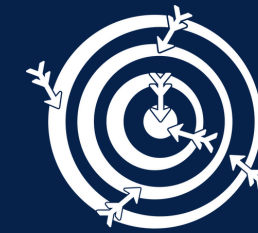
A Busca pelo Melhor Modelo:

Mesmo com um modelo adequado, testamos **outras 5 configurações** para buscar um ajuste ainda mais preciso.



Critérios de Informação (AIC, AICc e BIC):

Buscam o melhor equilíbrio entre ajuste e complexidade.



Métricas de Erro (ME, RMSE, MAE, MPE, MAPE, MASE, ACFI):

Medem o quão perto o modelo se ajusta aos dados.

Identificação do Modelo

CRITÉRIOS DE INFORMAÇÃO

	AIC	AICc	BIC
SARIMA(0,1,1)(0,1,1) ₁₂	-890,19	-890,09	-879,87
SARIMA(1,1,1)(0,1,1) ₁₂	-898,42	-898,25	-884,65
SARIMA(2,1,1)(0,1,1) ₁₂	-897,55	-897,29	-880,34
SARIMA(1,1,2)(0,1,1) ₁₂	-897,82	-897,55	-880,60
SARIMA(3,1,2)(0,1,1) ₁₂	-895,12	-894,62	-871,02
SARIMA(2,1,3)(0,1,1) ₁₂	-895,51	-895,01	-871,41
SARIMA(2,1,3)(0,1,1) ₁₂ com $\varphi_1 = 0$	-897,43	-897,06	-876,78

Identificação do Modelo

CRITÉRIOS DE INFORMAÇÃO

	AIC	AICc	BIC
SARIMA(0,1,1)(0,1,1) ₁₂	-890,19	-890,09	-879,87
SARIMA(1,1,1)(0,1,1) ₁₂	-898,42	-898,25	-884,65
SARIMA(2,1,1)(0,1,1) ₁₂	-897,55	-897,29	-880,34
SARIMA(1,1,2)(0,1,1) ₁₂	-897,82	-897,55	-880,60
SARIMA(3,1,2)(0,1,1) ₁₂	-895,12	-894,62	-871,02
SARIMA(2,1,3)(0,1,1) ₁₂	-895,51	-895,01	-871,41
SARIMA(2,1,3)(0,1,1)₁₂ com $\varphi_1 = 0$	-897,43	-897,06	-876,78



Identificação do Modelo

MÉTRICAS DE ERRO

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
SARIMA(0,1,1)(0,1,1)₁₂	-0,0008	0,0324	0,0241	-0,0054	0,1501	0,3886	0,0409
SARIMA(1,1,1)(0,1,1)₁₂	-0,0029	0,0312	0,0229	-0,0184	0,1424	0,3688	-0,0587
SARIMA(2,1,1)(0,1,1)₁₂	-0,0027	0,0311	0,0229	-0,0172	0,1425	0,3691	-0,0089
SARIMA(1,1,2)(0,1,1)₁₂	-0,0027	0,0312	0,0229	-0,0169	0,1426	0,3694	0,0030
SARIMA(3,1,2)(0,1,1)₁₂	-0,0027	0,0312	0,0228	-0,0169	0,1419	0,3674	-0,0074
SARIMA(2,1,3)(0,1,1)₁₂	-0,0026	0,0312	0,0229	-0,0165	0,1423	0,3684	0,0028
SARIMA(2,1,3)(0,1,1)₁₂ com $\varphi_1 = 0$	-0,0026	0,0312	0,0229	-0,0168	0,1421	0,3680	0,0022

Identificação do Modelo

Modelo Final: SARIMA(2,1,3)(0,1,1)₁₂ com $\varphi_1 = 0$

REPRESENTAÇÃO DO MODELO SARIMA(2,1,3)(0,1,1)₁₂ COM $\Phi_1 = 0$:

$$(1 - \phi_2 B^2)(1 - B)(1 - B^{12})Z_t = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3)(1 - \Theta_1 B^{12})a_t$$

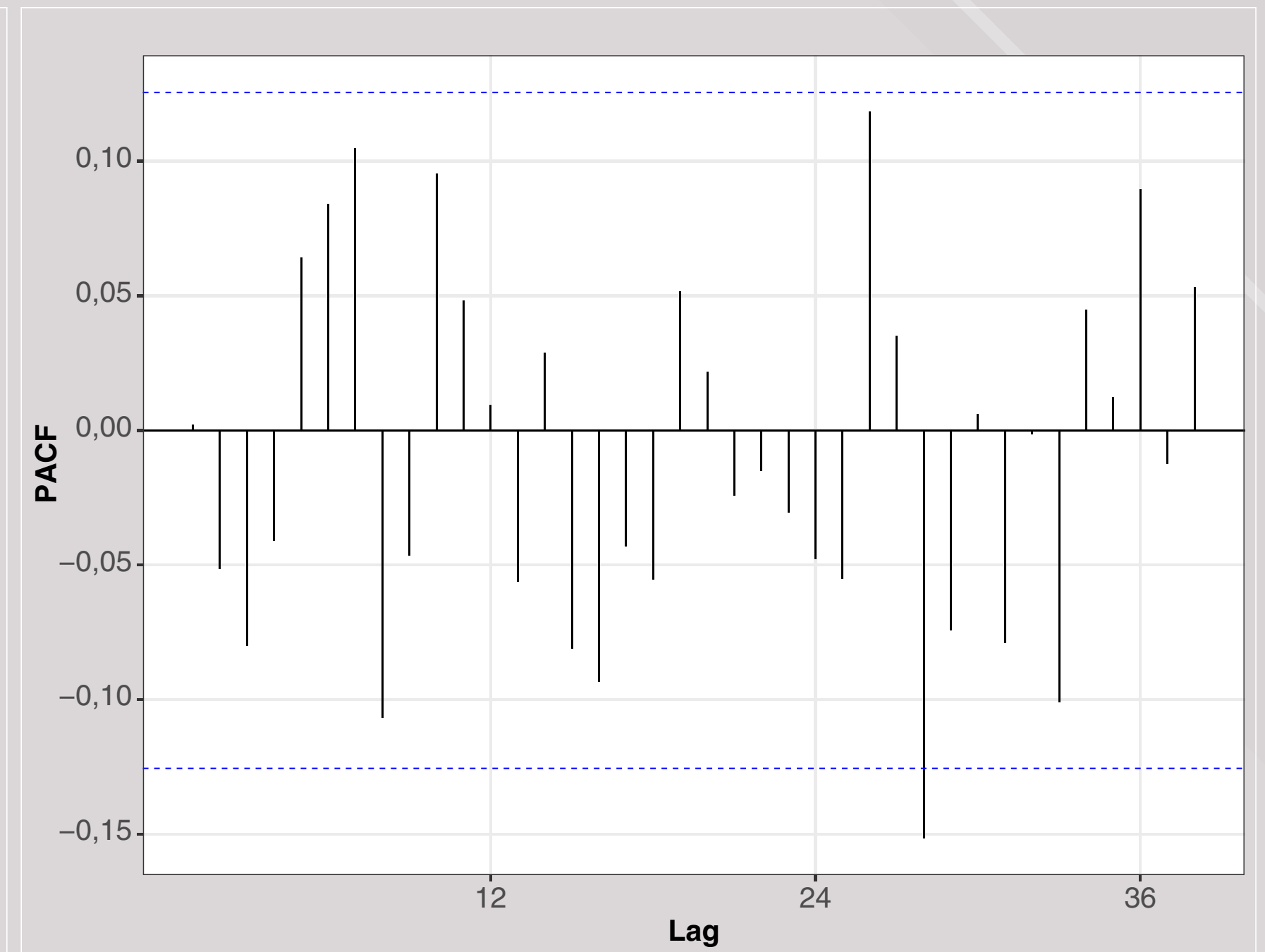
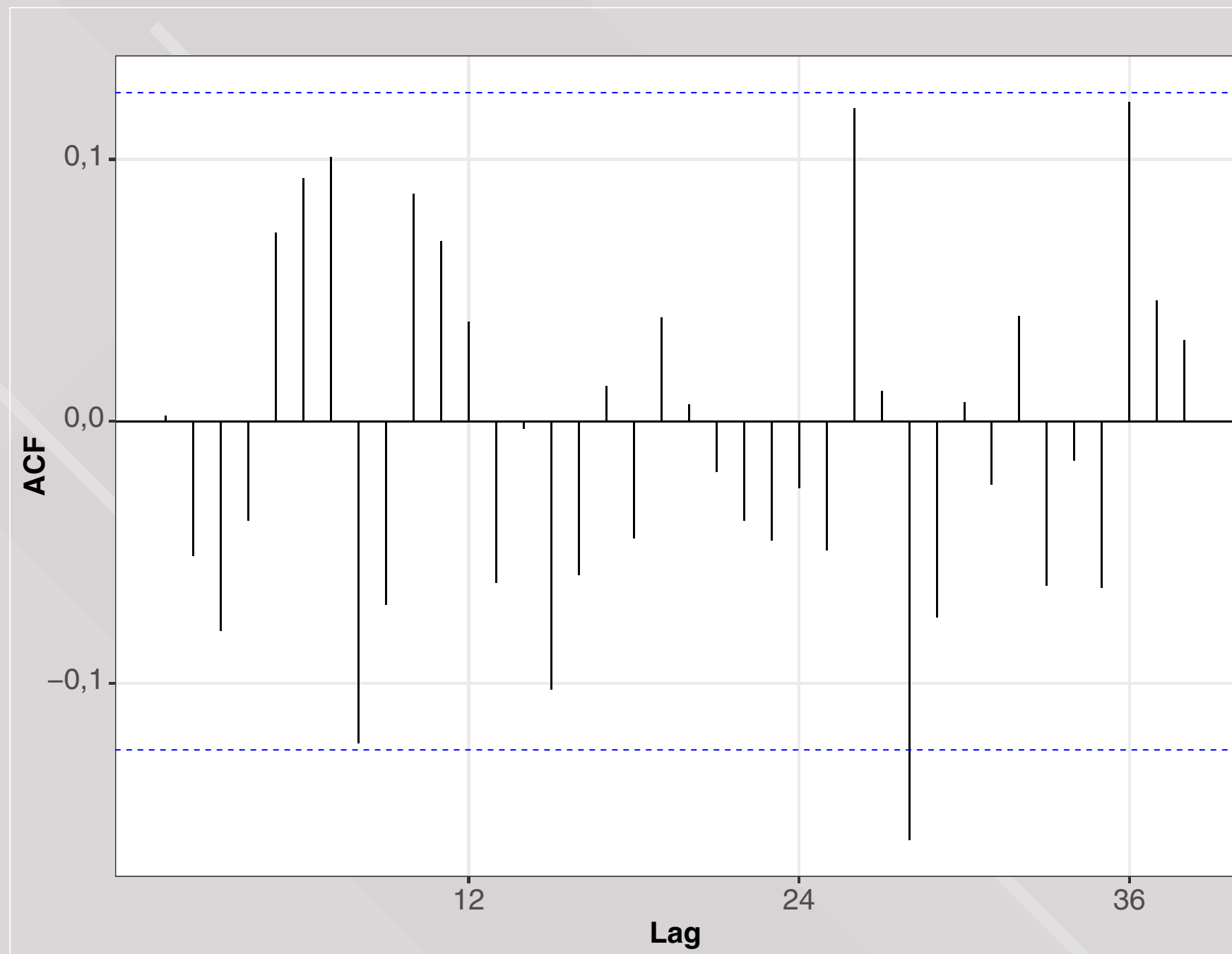
SUA EQUAÇÃO, COM OS COEFICIENTES ESTIMADOS, É:

$$(1 - 0,7088B^2)(1 - B)(1 - B^{12})Z_t = (1 + 0,2711B + 0,8748B^2 - 0,1460B^3)(1 + 0,9117B^{12})a_t$$

$$Z_t : \log(\text{Producao}_t); \quad a_t \sim \text{RB}(0, 0, 00105).$$

Identificação do Modelo

ACF e PACF dos resíduos do modelo SARIMA(2,1,3)(0,1,1)₁₂ com $\varphi_1 = 0$

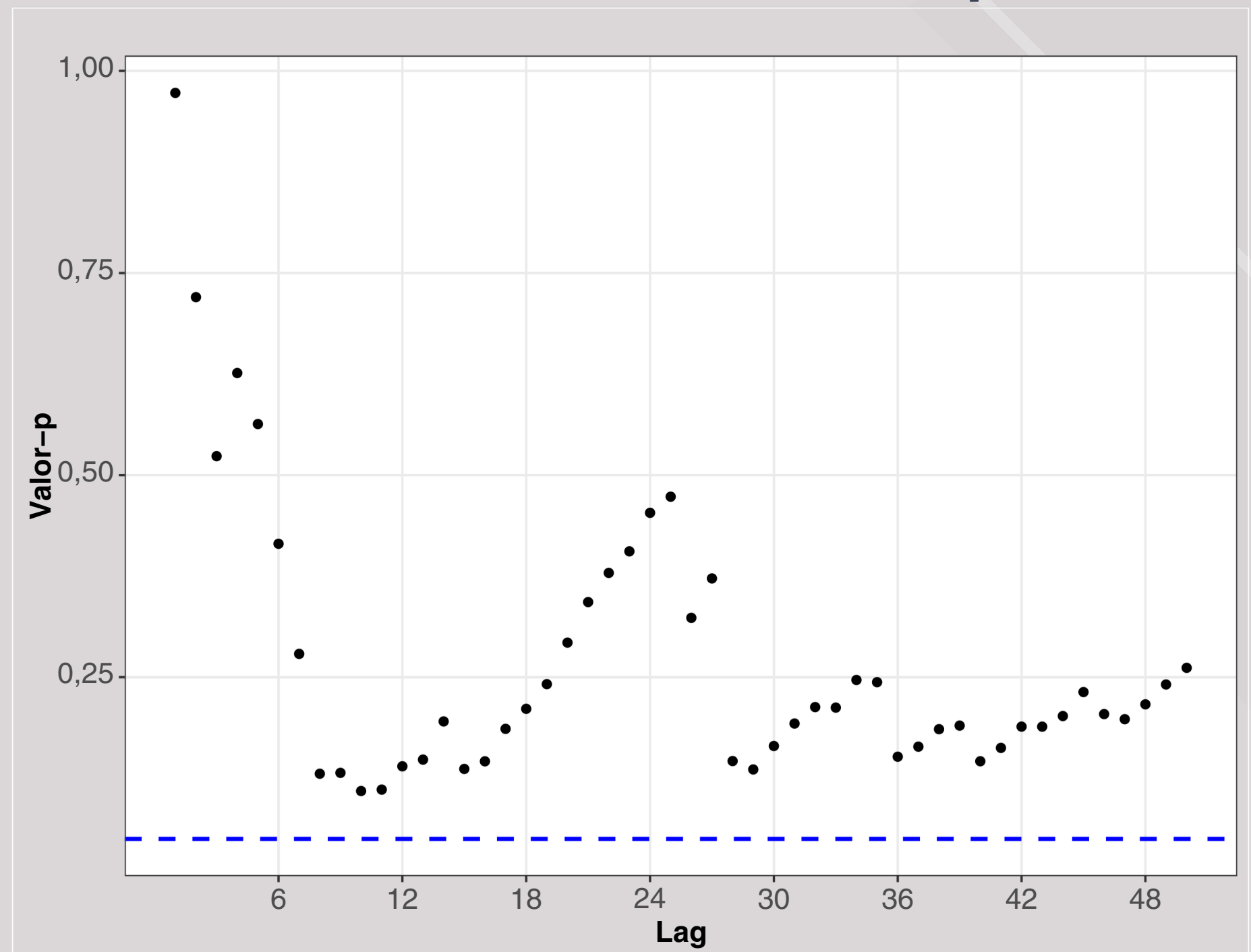


Identificação do Modelo

Teste de Ljung-Box do Modelo SARIMA(2,1,3)(0,1,1)₁₂ com $\varphi_1 = 0$

RESULTADOS

- O modelo **capta com sucesso** a estrutura de dependência temporal da série.
- O teste de Ljung-Box **confirmou a independência dos resíduos**.



Identificação do Modelo

NORMALIDADE DOS RESÍDUOS

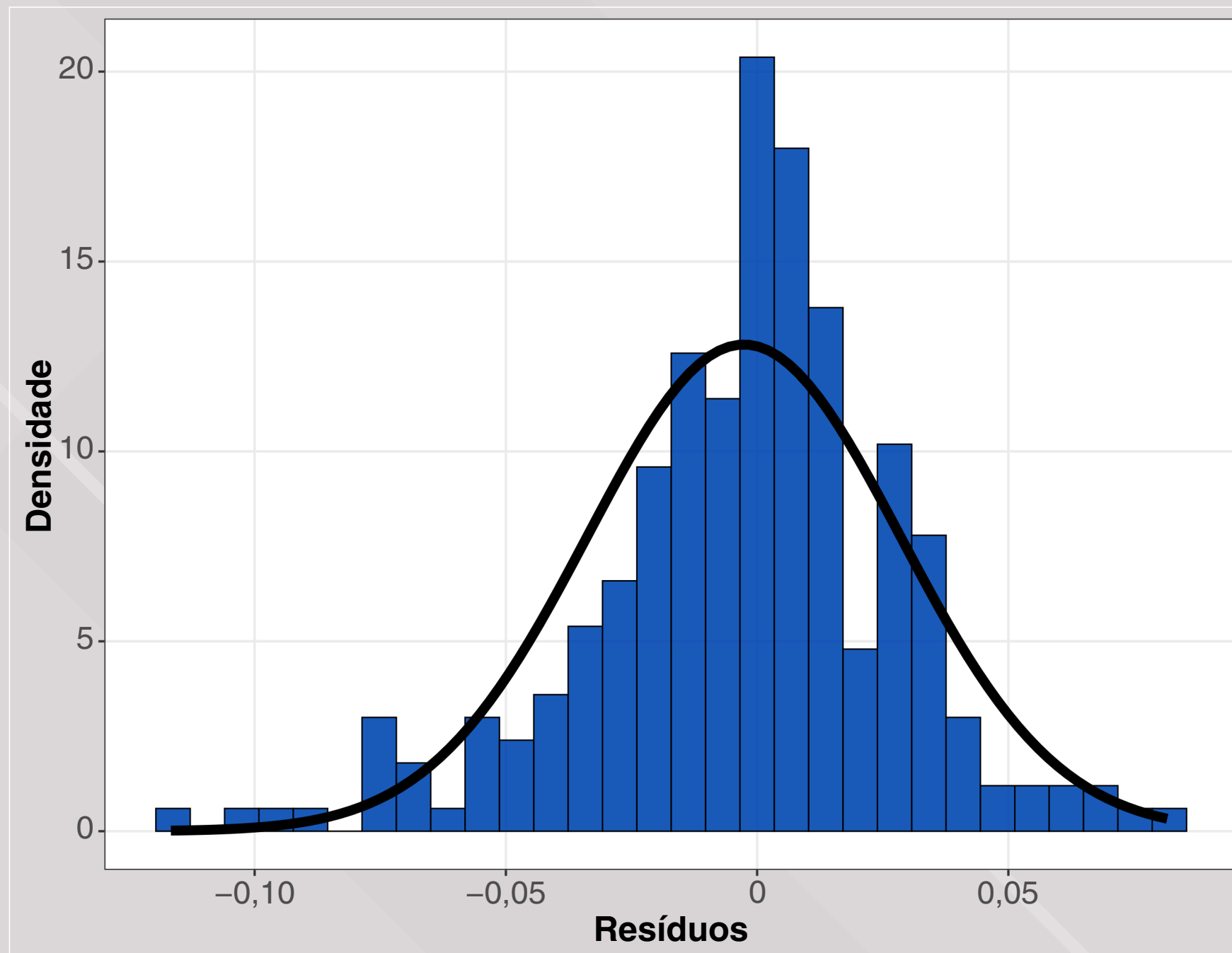


FIGURA 4: Histograma com curva de densidade.

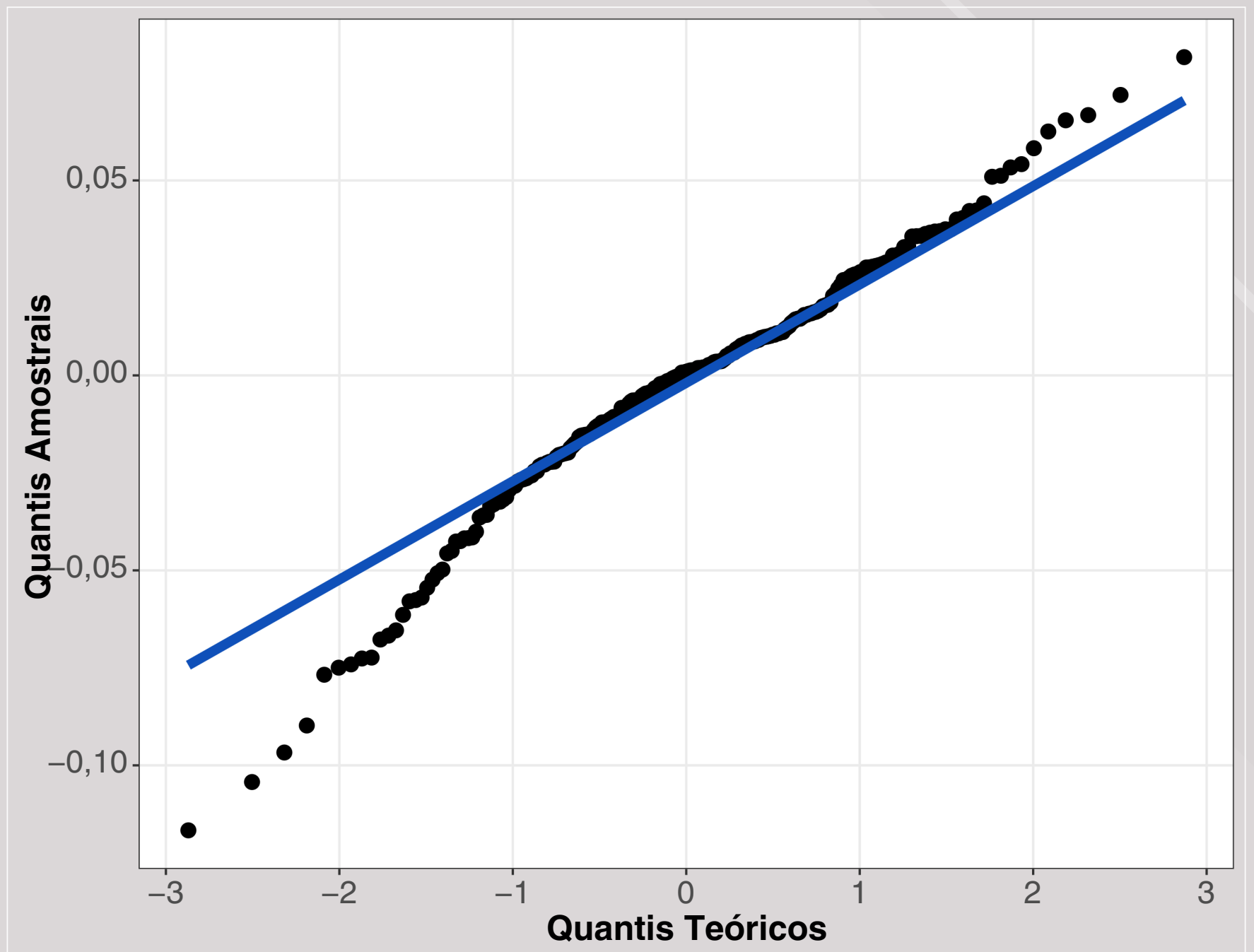


FIGURA 5: Gráfico Q-Q de normalidade.

Identificação do Modelo

TESTES DE NORMALIDADE

Normalidade dos Resíduos?

- Os resíduos **não** seguem uma distribuição perfeitamente normal (caudas mais "pesadas").

Teste de Shapiro-Wilk


- H_0 : Os resíduos seguem uma distribuição normal.
- **Valor-p < 1%.**

Teste de Jarque-Bera

- H_0 : Os resíduos seguem uma distribuição normal.
- **Valor-p < 1%.**



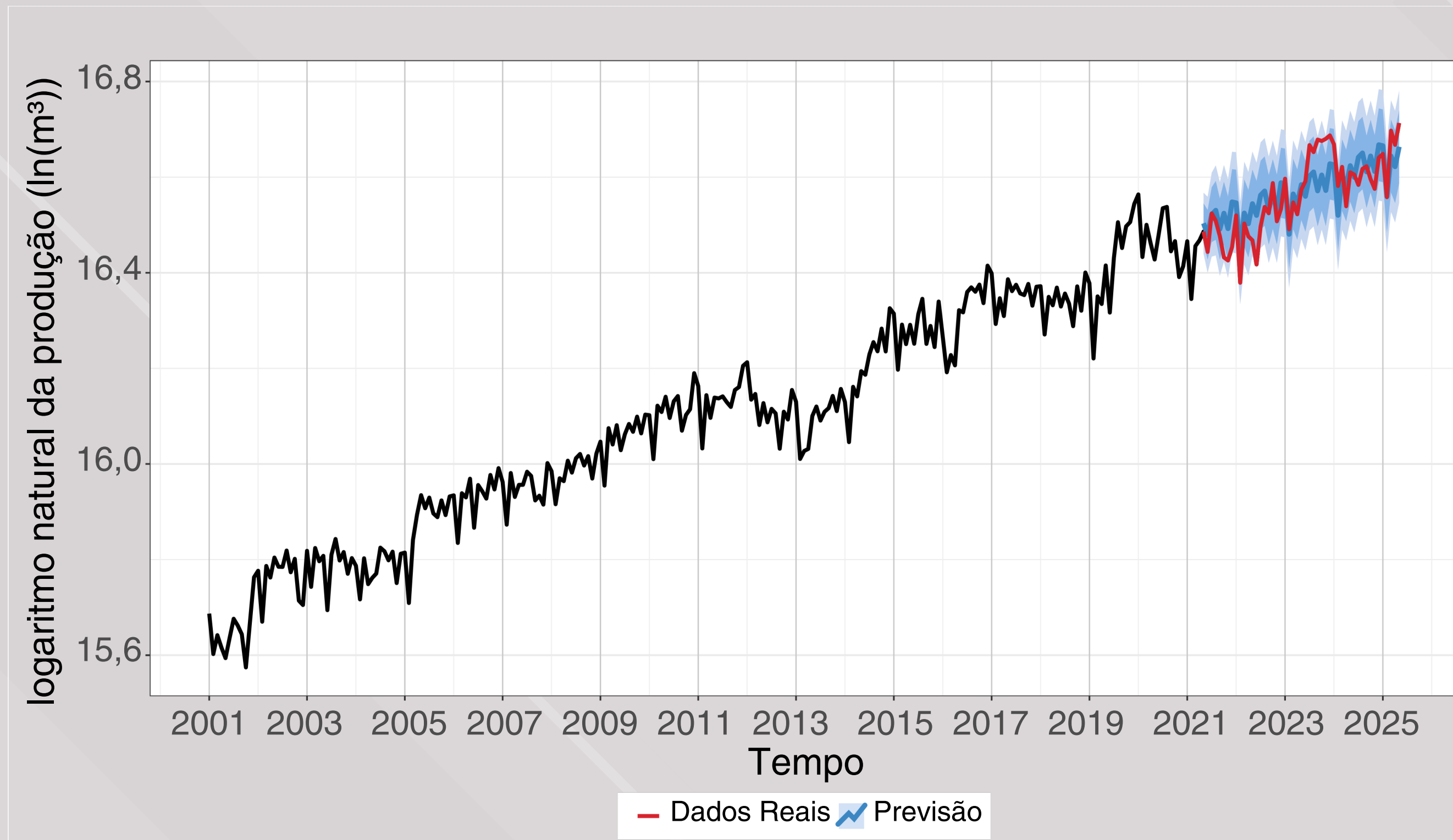
A falta de normalidade não invalida o modelo para previsões pontuais.



Previsão e Avaliação do Desempenho

Previsão e Avaliação

COLOCANDO O MODELO À PROVA (2021-2025)



Previsão e Avaliação

COMENTÁRIOS



Adesão Geral: As previsões do modelo (SARIMA(2,1,3)(0,1,1)₁₂) acompanham bem a tendência e a sazonalidade dos dados reais.



Pontos de Atenção:

- O modelo tende a subestimar os picos de produção no período de alta de 2023-2024.
 - Apesar disso, os valores reais permanecem dentro dos intervalos de confiança.
-



Conclusão Visual: O modelo demonstra uma performance preditiva acurada e consistente.

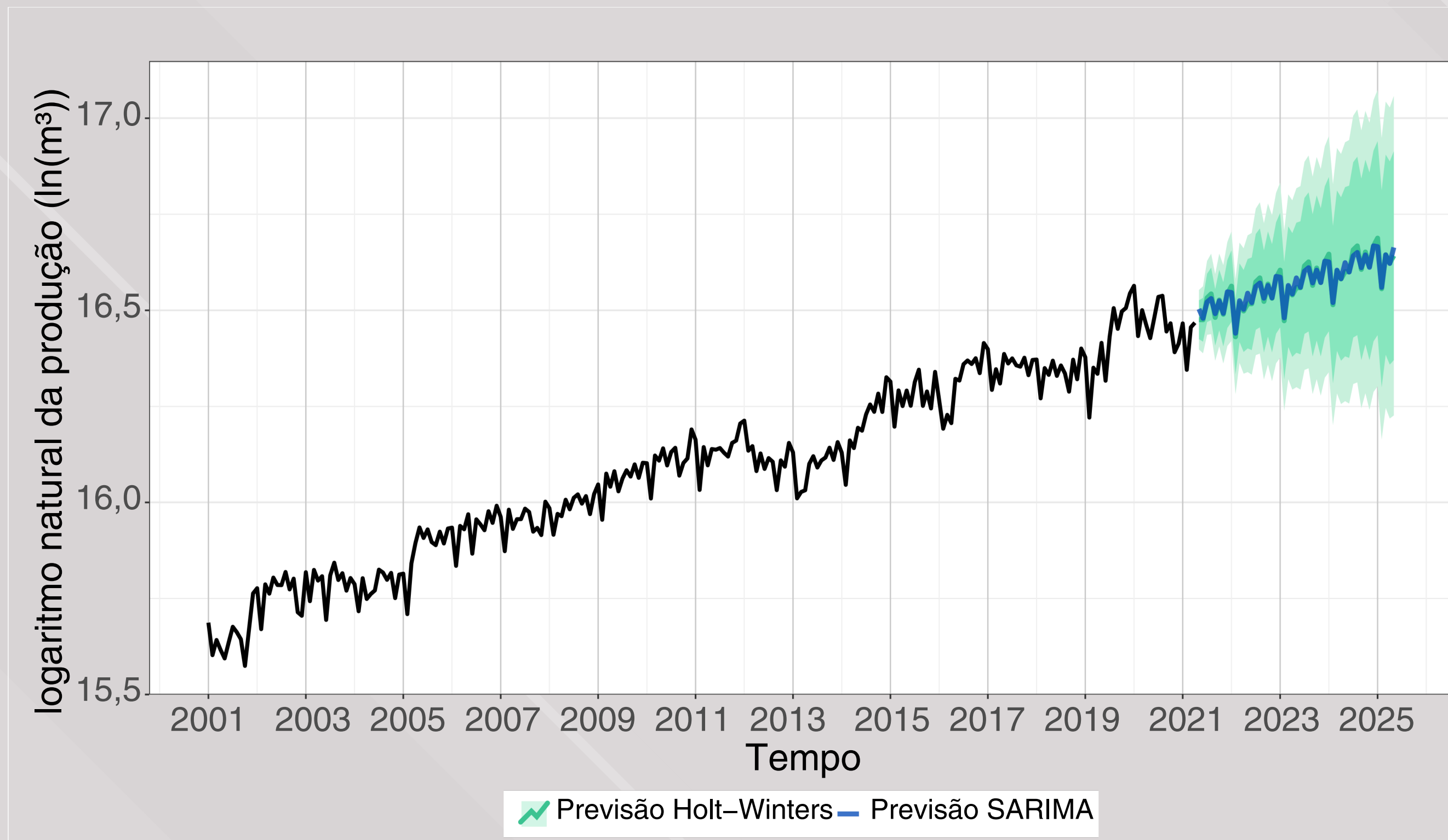
Previsão e Avaliação

ANÁLISE DE ACURÁCIA: JUSTIFICANDO A ESCOLHA FINAL

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
SARIMA(0,1,1)(0,1,1)₁₂	-0,0013	0,0489	0,0393	-0,0091	0,2371	0,7030	0,8487
SARIMA(1,1,1)(0,1,1)₁₂	-0,0077	0,0498	0,0409	-0,0477	0,2470	0,7070	0,8637
SARIMA(2,1,1)(0,1,1)₁₂	-0,0075	0,0496	0,0408	-0,0464	0,2461	0,7056	0,8618
SARIMA(1,1,2)(0,1,1)₁₂	-0,0074	0,0496	0,0407	-0,0456	0,2455	0,7046	0,8604
SARIMA(3,1,2)(0,1,1)₁₂	-0,0074	0,0496	0,0407	-0,0460	0,2457	0,7033	0,8615
SARIMA(2,1,3)(0,1,1)₁₂	-0,0071	0,0494	0,0405	-0,0442	0,2445	0,7013	0,8585
SARIMA(2,1,3)(0,1,1)₁₂ com $\varphi_1 = 0$	-0,0072	0,0495	0,0406	-0,0448	0,2450	0,7019	0,8598

Previsão e Avaliação

SARIMA VS. HOLT-WINTERS



Previsão e Avaliação

COMENTÁRIOS



Comparação com um Modelo Alternativo:

- Comparamos o modelo SARIMA com o Holt-Winters, uma técnica mais simples de suavização exponencial.
-



Previsões Pontuais: Os dois modelos geram previsões notavelmente próximas.

Incerteza: O modelo SARIMA apresenta intervalos de confiança muito mais estreitos, sugerindo maior precisão.



Conclusão

Conclusão

CONSIDERAÇÕES FINAIS

Modelo Final

O SARIMA(2,1,3)(0,1,1)₁₂ com $\varphi_1=0$ provou ser um modelo eficaz, preciso e estatisticamente válido para prever a produção de petróleo no Brasil.

Desempenho

O modelo foi capaz de capturar a complexa estrutura de tendência e sazonalidade da série.

Recomendação Final

A escolha entre os métodos depende do objetivo: a agilidade da suavização exponencial versus a precisão e o rigor da metodologia Box-Jenkins.

“Essencialmente, todos os modelos estão errados, mas alguns são úteis.”

George E. P. Box



**OBRIGADO
PELA
ATENÇÃO!**

