

# Trabalho Final – Métodos Estatísticos de Previsão

Eliana, Maria Cecília, Bárbara, Sophia, Dara

2025-01-19

## DESCRIÇÃO DOS DADOS

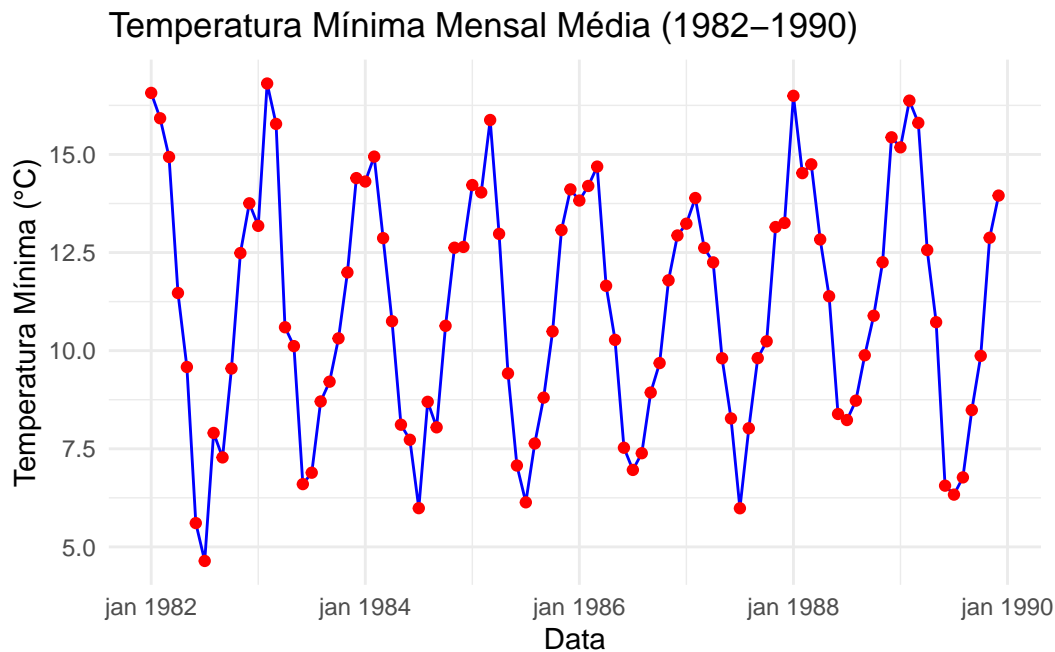
data	temperatura
jan 1982	16.567742
fev 1982	15.921429
mar 1982	14.935484
abr 1982	11.470000
mai 1982	9.583871
jun 1982	5.606667

A série temporal escolhida para esta análise é referente à temperatura mínima diária registrada na cidade de Melbourne, Austrália, entre janeiro de 1982 e dezembro de 1990. Os dados foram retirados do repositório Kaggle, especificamente do conjunto chamado Minimum Daily Temperatures in Melbourne, que pode ser acessado: <https://www.kaggle.com/datasets/ingwangdk/minimum-daily-temperatures-in-melbourne-10-years/data?select=daily-minimum-temperatures.csv>

A série original contém dados diários sobre a temperatura mínima, mas, para facilitar a análise de tendências ao longo do tempo e identificar padrões sazonais mais evidentes, decidimos transformar os dados para uma frequência mensal. Para isso, calculamos a média mensal das temperaturas mínimas, o que resultou em uma série de 108 observações mensais.

Essa transformação foi fundamental para que pudéssemos observar melhor as variações sazonais das temperaturas, como o comportamento típico das estações do ano. Com esses dados, é possível realizar análises mais profundas sobre tendências climáticas e até mesmo aplicar modelos estatísticos para prever futuras variações, utilizando técnicas como decomposição sazonal, suavização exponencial e outros métodos de previsão de séries temporais.

Foram retiradas as últimas 12 observações para o conjunto teste.



Ao analisar o gráfico da série temporal das temperaturas mínimas mensais entre 1982 e 1990, é possível identificar algumas características importantes:

*Sazonalidade:* A série apresenta um comportamento sazonal claro, com variações regulares a cada 12 meses. Isso é esperado devido à translação da Terra em torno do Sol, que gera mudanças de temperatura associadas às estações do ano.

*Estacionariedade:* Aparentemente as propriedades estatísticas da série (como a média e a variância) permanecem constantes ao longo do tempo. Ou seja, a série não aparenta ter tendência na parte simples nem sazonal, o que indica uma possível estacionariedade.

## AJUSTE DE MODELOS - ARIMA

Para a identificação do melhor modelo, o primeiro passo foi analisar as características da série temporal em questão. Para isso, foram utilizados os gráficos ACF (Função de Autocorrelação) e PACF (Função de Autocorrelação Parcial), que são ferramentas essenciais para verificar as dependências temporais da série.

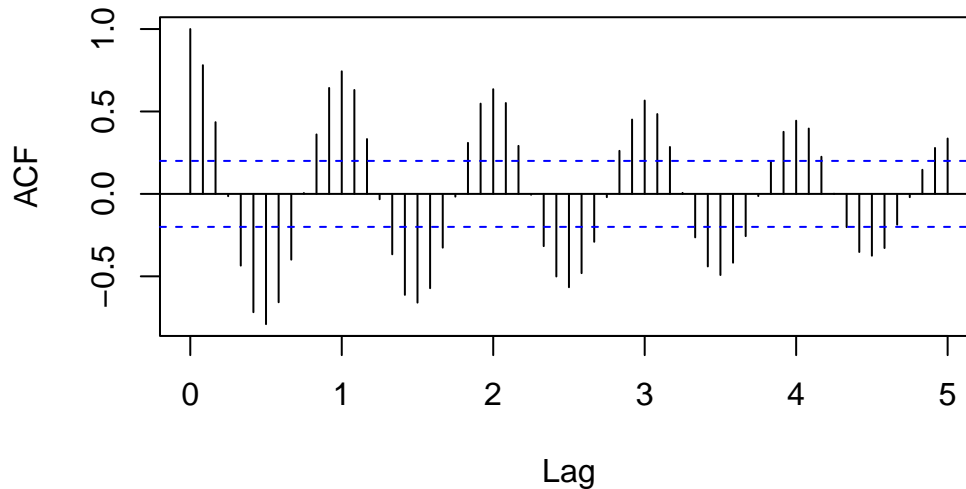
O gráfico de ACF permite observar as correlações entre os valores da série em diferentes defasagens (lags). A partir dessa análise, foi possível identificar se a série apresenta uma estrutura de dependência autoregressiva (AR) ou de média móvel (MA), além de indicar a necessidade de modelagem sazonal.

O gráfico de PACF, por sua vez, ajuda a identificar a ordem dos modelos autoregressivos, permitindo entender até que ponto os valores passados influenciam o valor presente da série. Com base nas observações dos gráficos ACF e PACF, foi possível fazer suposições iniciais sobre as ordens dos componentes AR e MA, tanto na parte sazonal quanto na parte não sazonal.

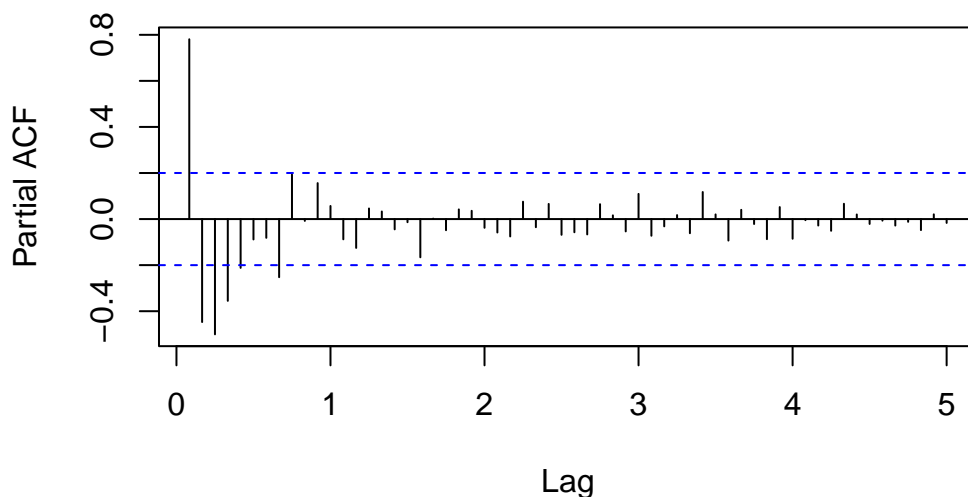
Com as características da série identificadas, seguiu-se para a construção de modelos ARIMA com diferentes combinações de ordens para os componentes AR, MA e sazonal. Para cada modelo ajustado, foi calculado o AIC (Critério de Informação de Akaike), com o objetivo de selecionar o modelo mais adequado, considerando a qualidade do ajuste e a complexidade do modelo. O modelo com o menor valor de AIC foi selecionado como o mais apropriado para a série temporal.

Em todos os testes estatísticos realizados neste estudo, foi adotado um nível de significância de 5% ( $\alpha = 0,05$ ).

### Função de Autocorrelação (ACF) da Série Temporal



## Função de Autocorrelação Parcial (PACF) da Série Temporal



Analisou-se as funções ACF e PACF para os 10 primeiros períodos, e algumas características importantes foram observadas. Tanto no gráfico da Função de Autocorrelação, quanto no gráfico da Função de Autocorrelação Parcial temos um decaimento exponencial que pode ser visualizado nos primeiros picos. Esse comportamento é característico de um modelo ARMA(1,1), por isso será o modelo inicial usado, para posteriormente sobrefixá-lo.

Analisando os picos na componente sazonal, observamos que, na ACF, ocorre um decaimento exponencial. Já na PACF, não é possível identificar um pico claro no lag 12, embora esteja próximo do limite. Com base nessa análise, o modelo mais adequado parece ser um AR(1), que oferece o melhor ajuste aos dados.

### Modelo 1

Table 2: Resultado do Teste de Coeficientes do Modelo

	Estimativa	Erro Padrão	Valor z	p-valor
ar1	0.6368315	0.1278669	4.980426	0.0000006
ma1	-0.2749314	0.1579219	-1.740932	0.0816954
sar1	0.8668429	0.0547378	15.836264	0.0000000
intercept	11.1495707	1.1649688	9.570703	0.0000000

Table 3: Valor do AIC para o Modelo mod1

Modelo	AIC
mod1	340.6309

O Modelo com nossa hipótese inicial, o parâmetro da media móvel não foi significativo com isso tiramos esse coeficiente.

### Sobrefixação da parte simples

#### Modelo 2

Table 4: Resultado do Teste de Coeficientes do Modelo

	Estimativa	Erro Padrão	Valor z	p-valor
ar1	0.4273180	0.1301849	3.282394	0.0010293
sar1	0.8403499	0.0675742	12.435954	0.0000000
intercept	11.1307511	0.9309357	11.956520	0.0000000

Table 5: Valor do AIC para o Modelo mod2

Modelo	AIC
mod2	341.5836

Assim concluirmos que o modelo 2, sem ma1 ficou melhor ajustado, uma vez que possui um bom AIC (341.5836) e todos componentes significativos.

Ao analisar  $\theta$  e  $\phi$ , comprovamos a inversibilidade e estacionariedade, uma vez que ambos são menores do que 1.

### Sobrefixação da sazonalidade

#### Modelo 3

Table 6: Resultado do Teste de Coeficientes do Modelo

	Estimativa	Erro Padrão	Valor z	p-valor
ar1	0.3976788	0.1123640	3.539201	0.0004013
sar1	0.4280685	0.0950530	4.503472	0.0000067
sar2	0.4975317	0.0988007	5.035710	0.0000005
intercept	11.2624592	1.0454814	10.772510	0.0000000

Table 7: Valor do AIC para o Modelo mod3

Modelo	AIC
mod3	324.077

Adicionando um novo parametro na parte sazonal referente a parte autorregressiva. Obtivemos um todos os coeficientes significativos. E um AIC menor em relação ao encontrado no modelo 2. Ao analisar  $\theta$  e  $\phi$ , comprovamos a inversibilidade e estacionariedade, uma vez que ambos são menores do que 1.

#### Modelo 4

Table 8: Resultado do Teste de Coeficientes do Modelo

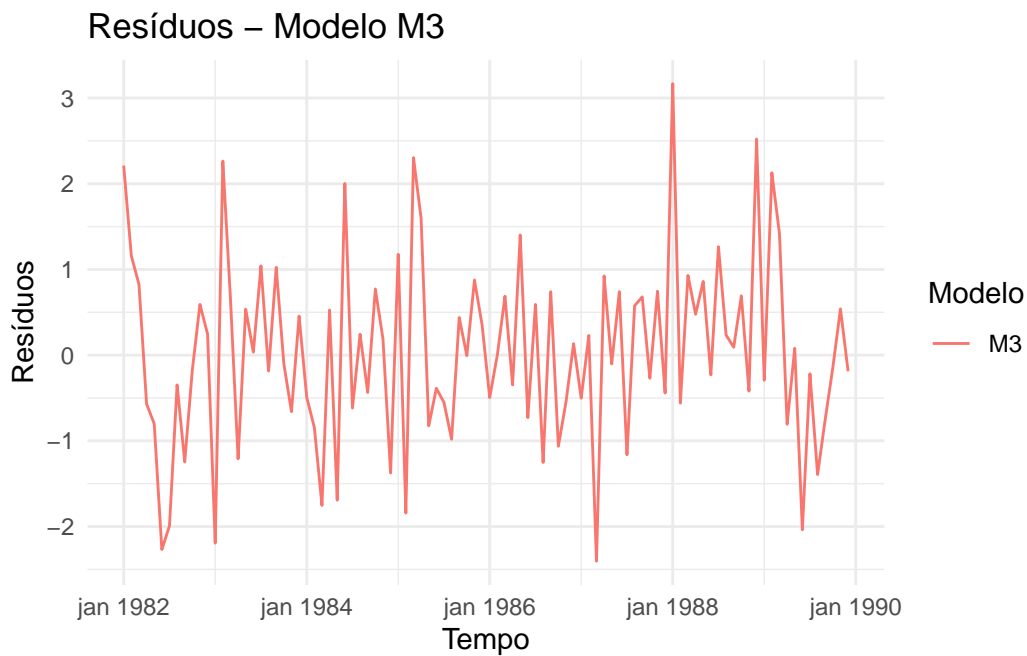
	Estimativa	Erro Padrão	Valor z	p-valor
ar1	0.4201721	0.1056930	3.975402	0.0000703
sar1	0.9990477	0.0066707	149.766514	0.0000000
sma1	-0.9084182	0.3127269	-2.904829	0.0036745
intercept	11.1144234	1.0141926	10.958888	0.0000000

Table 9: Valor do AIC para o Modelo mod4

Modelo	AIC
mod4	310.8457

Se o modelo 4 apresenta um  $\rho$  muito próximo de 1, podemos rejeitá-lo como estacionário, mesmo com um AIC menor. O modelo 3, embora tenha um AIC um pouco maior, é estacionário e mais confiável do ponto de vista teórico e estatístico.

### Análise resíduos:

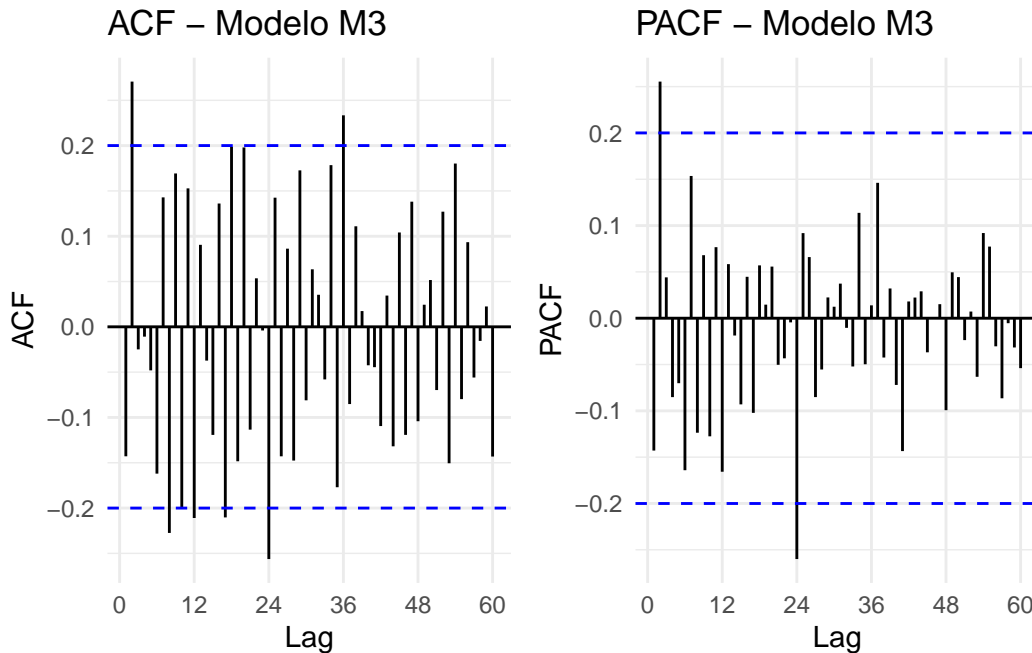


O pressuposto da variância constante ao longo do tempo foi atendido, garantindo que os resíduos do modelo não apresentam padrões sistemáticos ou mudanças na dispersão.

Table 10: Resultado do Teste de Normalidade de Shapiro-Wilk para os Resíduos do Modelo

	Estatística de Teste	Valor p
W	0.9878979	0.5307165

No teste de Shapiro-Wilk, nossa hipótese nula é a de os dados seguem a distribuição normal. O p-valor do teste foi igual a 0.5307165. Utilizando o nível de significância de 5%, não rejeitaremos a hipótese nula. Logo, o pressuposto de normalidade também foi atendido.



Em relação aos gráficos dos resíduos, a função de autocorrelação (ACF) apresenta cinco picos significativos, indicando possíveis autocorrelações nos primeiros lags. Por outro lado, a função de autocorrelação parcial (PACF) revelou um pico significativo, indicando que o valor atual está fortemente relacionado ao valor imediatamente anterior.

A análise da autocorrelação dos resíduos é essencial para verificar se o modelo ajustado ARIMA capturou adequadamente as dependências temporais da série. Se houver autocorrelação significativa nos resíduos, isso indica que o modelo não está ajustado corretamente. No teste de Box-Ljung, a hipótese nula é de que não existe autocorrelação significativa nos resíduos, ou seja, os resíduos são independentes. O p-valor obtido no teste para o primeiro lag foi de 0.1551, o que indica que, para este lag específico, não podemos rejeitar a hipótese nula, sugerindo ausência de autocorrelação significativa. Isso indica que o modelo ARIMA ajustado (mod3) não conseguiu explicar completamente as dependências temporais presentes na série.

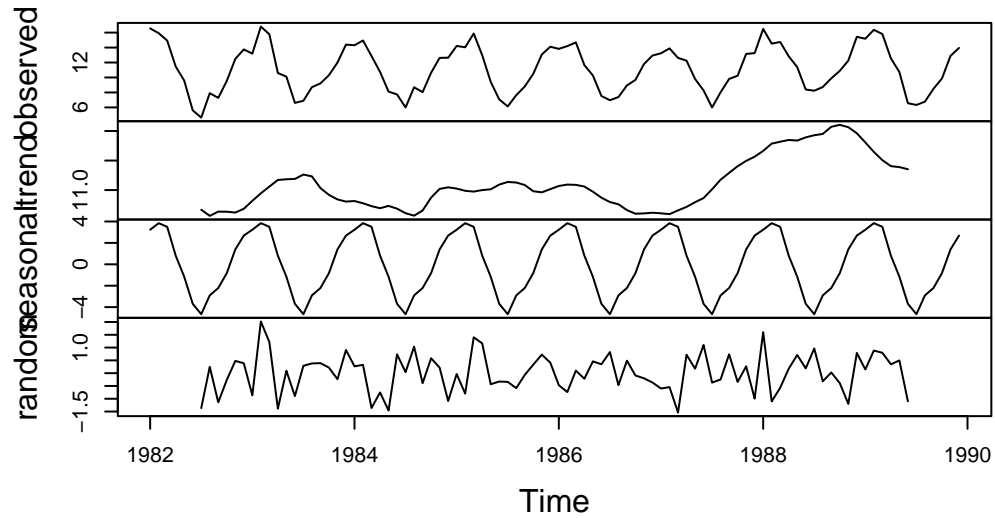
No entanto, para os lags 2, 11, 12 e 24, os p-valores foram menores que 0,05 (0.009315, 0.001525, 0.0004527 e 4.154e-06, respectivamente), indicando que nesses lags há evidências de autocorrelação significativa nos resíduos, e a hipótese nula de independência deve ser rejeitada.

Portanto, os resíduos apresentam correlação significativa em determinados lags, sugerindo que o modelo ajustado pode não ter capturado completamente todas as dependências temporais nos dados. Nesse caso, outro modelo, que não faz parte do grupo dos ARIMAs estudados na disciplina, poderia ser ajustado aos dados. Não tendo as ferramentas para tal, consideraremos o modelo 7 aquele com melhor ajuste entre os analisados.”

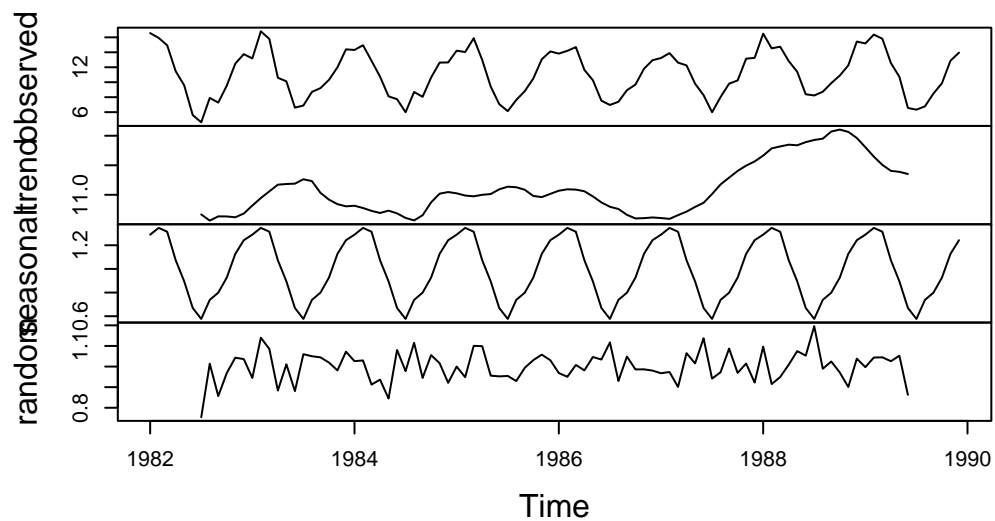


## AJUSTE DE MODELOS - Alisamento Exponencial

### Decomposition of additive time series



### Decomposition of multiplicative time series



Os plots acima decompõe a série temporal em três componentes principais: tendência, sazonalidade e resíduo:

*Tendência:* A série parece mostrar um aumento nas temperaturas mínimas a partir de 1987, sugerindo uma tendência de aquecimento nas últimas décadas do período analisado. No entanto, é importante realizar testes estatísticos.

*Sazonalidade:* A sazonalidade é claramente visível, com picos e vales que se repetem a cada 12 meses. Isso reflete as variações típicas de temperatura associadas às estações do ano, com temperaturas mais altas no verão e mais baixas no inverno.

*Resíduo:* O resíduo apresenta um padrão aleatório, com valores positivos e negativos distribuídos de forma equilibrada. Isso sugere que o modelo de decomposição conseguiu capturar a maior parte da variabilidade da série, deixando apenas flutuações aleatórias, ou seja, o modelo foi bem-sucedido em explicar os principais componentes da série.

Para determinar qual modelo de decomposição (aditivo ou multiplicativo) é mais adequado, é necessário realizar alguns testes e análises adicionais:

Table 11: Parâmetros e Métricas do Modelo Aditivo de Holt-Winters

Parâmetro	Valor
Alpha	0.2278362
Beta	0.0169791
Gamma	0.5369961
RMSE	0.9957728
SSE	95.1900868

Table 12: Parâmetros e Métricas do Modelo Multiplicativo de Holt-Winters

Parâmetro	Valor
Alpha	0.2870191
Beta	0.0074449
Gamma	0.6553478
RMSE	1.0740517
SSE	110.7443662

- RMSE (Root Mean Squared Error): Mede a precisão do modelo. Quanto menor, melhor o modelo ajusta-se aos dados.
- SSE (Soma dos Erros ao Quadrado): Mede a soma dos quadrados dos resíduos (erros), sendo uma indicação do quão bem o modelo está ajustado.

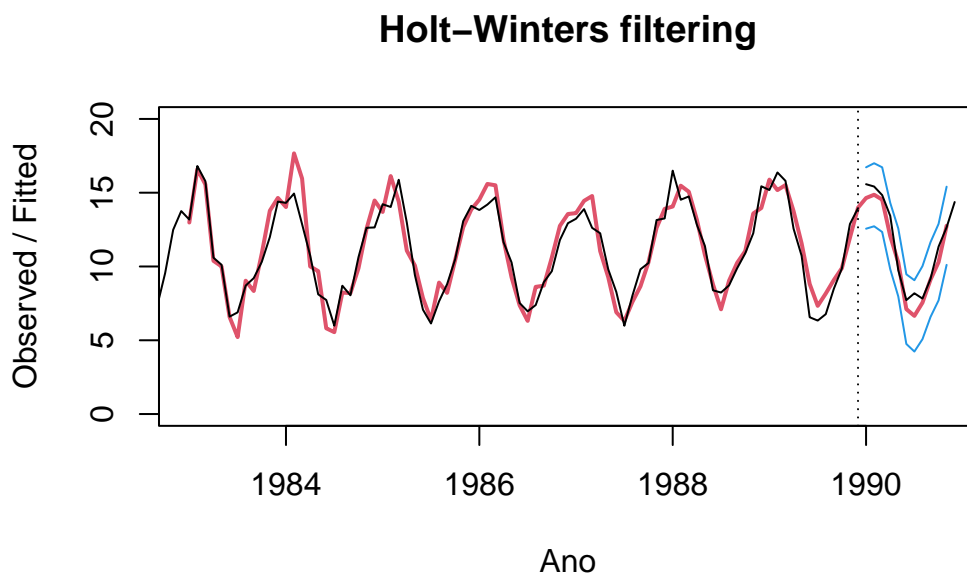
O modelo aditivo apresenta um alpha de 0.2278, o que indica uma suavização moderada da série temporal, ou seja, o modelo atribui um peso razoável aos valores mais recentes sem exagerar. O beta de 0.0169 sugere uma tendência de crescimento ou decrescimento muito suave, enquanto o gamma de 0.5370 aponta para uma sazonalidade com um impacto considerável na modelagem.

A RMSE de 0.9957 e o SSE de 95.1901 são relativamente baixos, sugerindo que o modelo aditivo consegue capturar bem a variabilidade dos dados e ajustar-se de forma eficaz à série.

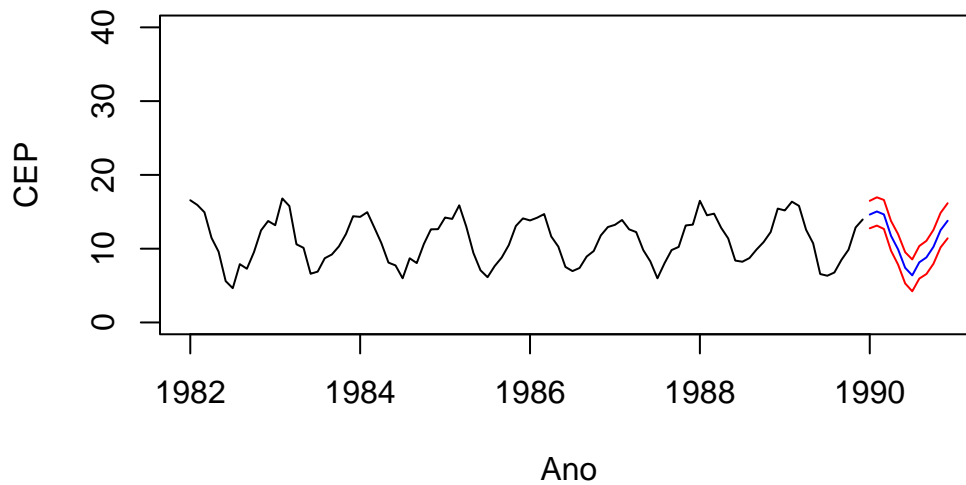
Por outro lado, o modelo multiplicativo apresenta um  $\alpha$  um pouco maior (0.2870), o que indica que ele atribui um peso ligeiramente maior aos dados mais recentes, ajustando-se mais rapidamente a mudanças nas observações. O  $\beta$  de 0.0074 é bem baixo, o que significa que a tendência do modelo é muito suave, com uma variação muito pequena. O  $\gamma$  de 0.6553, por sua vez, sugere uma sazonalidade mais forte, ou seja, o modelo multiplicativo ajusta-se a variações sazonais mais acentuadas. No entanto, a RMSE de 1.0741 e o SSE de 110.7444 são maiores, indicando que o modelo multiplicativo tem um desempenho inferior, com maior erro de ajuste e menor precisão.

Dado que o modelo aditivo apresenta valores de RMSE e SSE menores, indicando um ajuste mais preciso e menos erro, ele será o modelo escolhido para esta série temporal.

### Previsões usando AEHW aditivo



## Previsão do modelo ARIMA



## COMPARAÇÃO DOS MODELOS

Modelo	MAE	RMSE	MAPE
ARIMA	0.6662216	0.8713011	6.088299
Alisamento Exponencial	0.6322774	0.7858409	5.827719

Com base nos resultados das métricas de erro, o modelo de Alisamento Exponencial apresentou um desempenho superior ao ARIMA. Ele obteve valores menores de MAE (0.6323 contra 0.6662), RMSE (0.7858 contra 0.8713) e MAPE (5.8277 contra 6.0883). Essas diferenças indicam que o Alisamento Exponencial foi mais eficaz em reduzir os erros absolutos e percentuais, além de lidar melhor com variações maiores nos dados.

Portanto, para o conjunto de dados testado, o Alisamento Exponencial se mostra uma escolha mais adequada, oferecendo previsões mais precisas. Ele é particularmente vantajoso para cenários em que erros percentuais e grandes desvios devem ser minimizados, garantindo maior confiabilidade nas estimativas realizadas.