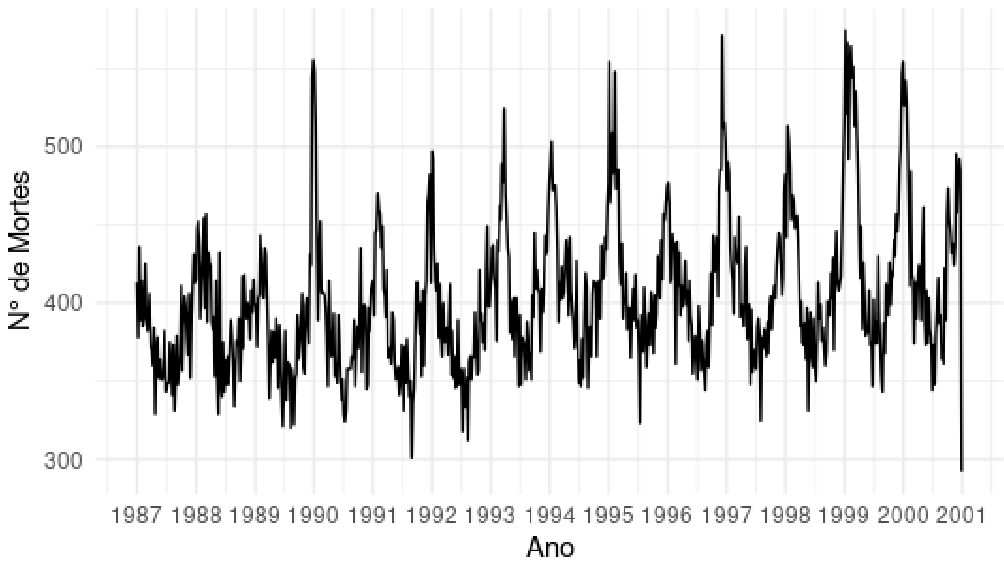
MI411 SERIES TEMPORAIS
Prova 1 / Questao 5

Adriel Wesley Nascimento Melo
RA: 258000

Campinas, 2024

O objetivo deste estudo é modelar uma série temporal de mortalidade de pessoas com mais de 75 anos na cidade de Chicago, EUA, utilizando dados semanais. A série abrange 731 observações, que representam o total de mortes em cada semana, desde 1987 até 2000. A visualização dessa série pode ser observada na Figura 1:

Figura 1: Série temporal de mortalidade semanal de pessoas com mais de 75 anos em Chicago (1987-2000)



Fonte: Autor

A Figura 1 revela uma leve tendência ao longo dos anos e uma notável sazonalidade anual. Observa-se um padrão de diminuição das mortes na metade do ano e um aumento no final e início do ano.

A Figura 2 detalha o efeito da sazonalidade na série de mortalidade, com um padrão repetitivo anualmente, incluindo alguns picos mais acentuados.

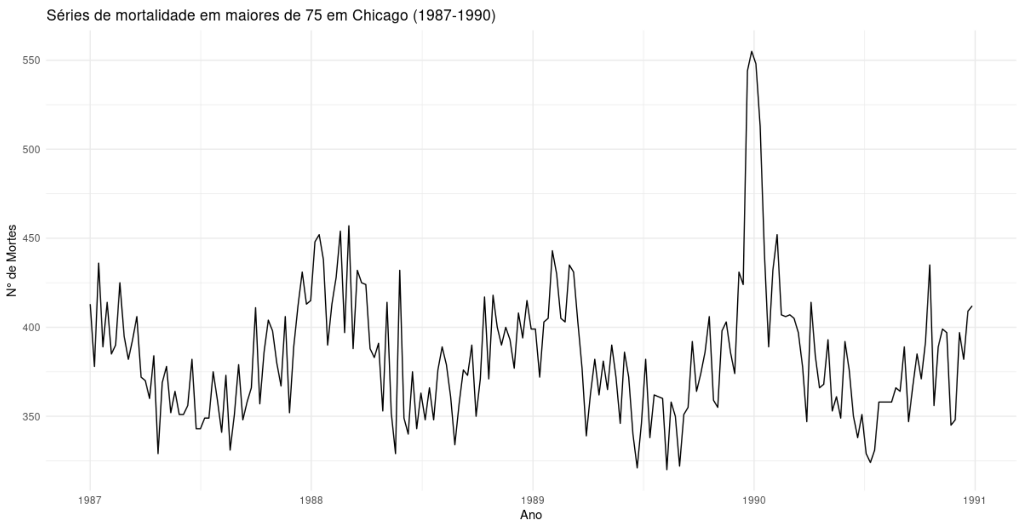
Dado que a série temporal apresenta um ciclo anual, o modelo de regressão harmônica foi escolhido. A frequência do ciclo é de 52 semanas, e a fórmula do modelo de regressão harmônica é:

$$y_t = T_t \left(\sum_{j=1}^{J} \left(a_j \sin(2\pi\nu_j t) + \beta_j \cos(2\pi\nu_j t) \right) + \epsilon_t \right),$$

onde o ciclo é $P = 52$ semanas, e para $J = \frac{52}{2} = 26$. No estudo, o modelo selecionado utilizou $J = 2$ e obteve um $R^2 = 0.58$. Os coeficientes do modelo ajustado são apresentados na tabela abaixo:

Análise dos Coeficientes:

Figura 2: Série temporal de mortalidade semanal para o período de 1987 a 1990, destacando a sazonalidade



Fonte: Autor

Tabela 1: Estimativas dos parâmetros do modelo selecionado.

	Coeficientes	Estatística t	P-valor
Intercepto	375,3	2,241	2×10^{-16}
t	0,075	0,005	2×10^{-16}
$\sin(2\pi \frac{1}{52} t)$	38,20	1,586	2×10^{-16}
$\cos(2\pi \frac{1}{52} t)$	24,20	1,579	2×10^{-16}
$\sin(2\pi \frac{2}{52} t)$	4,03	1,580	0,0109
$\cos(2\pi \frac{2}{52} t)$	7,14	1,584	$7,69 \times 10^{-6}$

Fonte: Autor

- **Intercepto**: O coeficiente é 375,3 com uma estatística t de 2,241 e um p-valor inferior a 2×10^{-16}, indicando que é estatisticamente significativo e sugere um valor base considerável para a variável dependente.

- **Variável** t: O coeficiente é 0,075 com uma estatística t de 0,005 e um p-valor inferior a 2×10^{-16}. Isso indica uma tendência linear significativa ao longo do tempo.

- **Componentes Sazonais**:

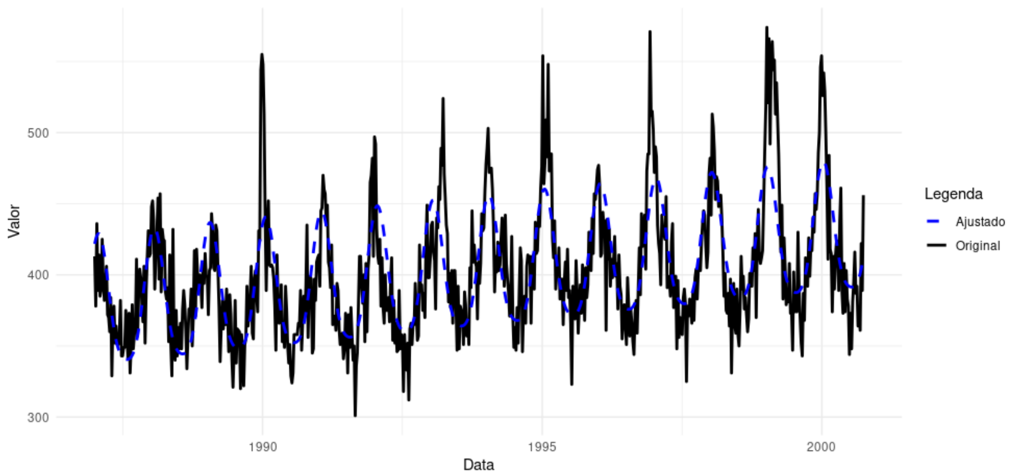
  – $\sin\left(2\pi \frac{1}{52} t\right)$: O coeficiente é 38,20 com uma estatística t de 1,586 e um p-valor muito baixo, confirmando a presença significativa desta componente sazonal.

  – $\cos\left(2\pi \frac{1}{52} t\right)$: O coeficiente é 24,20 com uma estatística t de 1,579 e um p-valor também muito baixo, indicando que esta componente cossenoidal semanal é significativa.

– $\sin\left(2\pi\frac{2}{52}t\right)$: O coeficiente é 4,03 com uma estatística t de 1,580 e um p-valor de 0,0109, sugerindo significância, mas com menor intensidade comparada às primeiras ordens.
– $\cos\left(2\pi\frac{2}{52}t\right)$: O coeficiente é 7,14 com uma estatística t de 1,584 e um p-valor muito baixo, confirmando a importância desta componente cossenoidal de segunda ordem.

Na Figura 3 pode-se observar os valores reais e os ajustados para série de mortalidade.

Figura 3: Valores ajustados



Fonte: Autor

A análise dos resíduos do modelo, mostrada na Figura 4, indica não haver padrão nos resíduos, sugerindo serem independentes.

O teste de Ljung-Box para a independência dos resíduos resultou em um p-valor de 0,9658, levando a não rejeição da hipótese nula de que os resíduos são independentes.

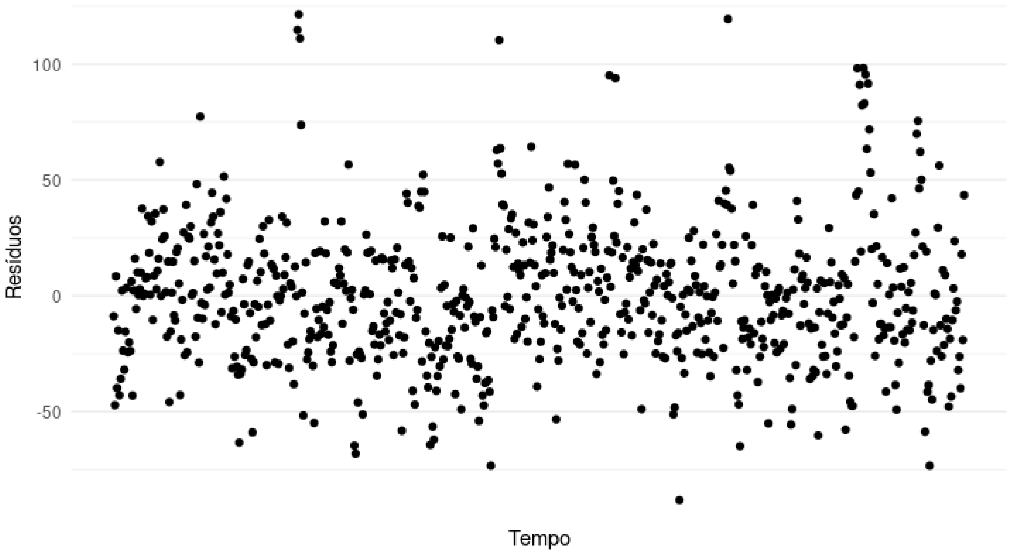
O modelo de regressão revela que a variável t e as componentes sazonais (seno e cosseno) são significativas para explicar a variável dependente. As variáveis sazonais têm coeficientes significativos, indicando que a sazonalidade semanal e suas interações de segunda ordem são importantes para capturar a variação na série temporal analisada.

Na Figura 5 pode-se observar os valores observados contra preditos, alguns valores ficaram bem próximos dos valores reais, porém alguns ficaram longes distantes, como na ultima observação.

Tabela 1: Comparação entre Valores Reais e Previstos com Erro

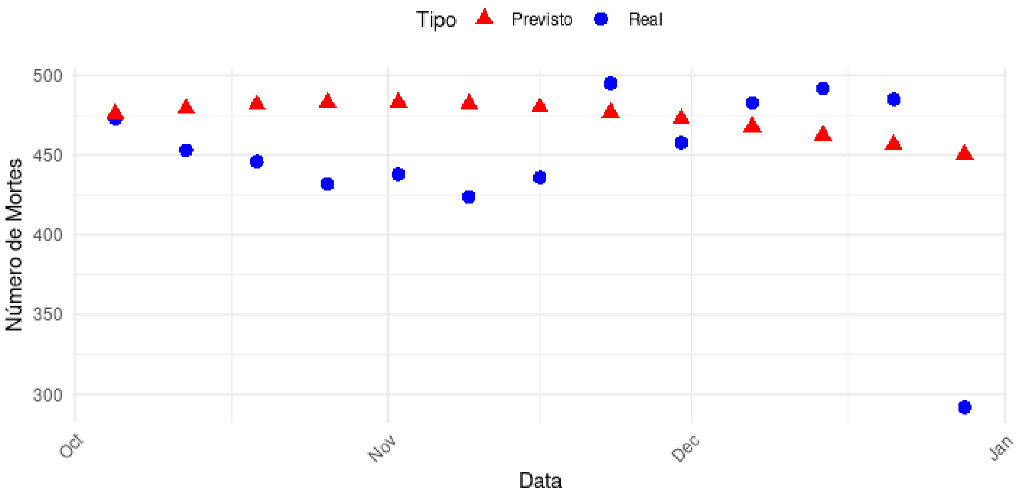
A Tabela 2 mostra a comparação entre os valores reais e os valores previstos pelo

Figura 4: Resíduos do Modelo de Regressão Harmônica



Fonte: Autor

Figura 5: Comparação entre Valores Reais e Previstos



Fonte: Autor

modelo, além dos erros correspondentes. Observa-se que os erros variam consideravelmente ao longo das semanas, com alguns erros positivos e outros negativos.

- Erros Significativos: Notam-se erros particularmente altos em algumas datas, como em 28/12/2000, onde o erro é de -158.51. Isso sugere que o modelo teve uma previsão significativamente imprecisa nessa data. A diferença entre os valores reais e previstos é considerável, o que pode indicar a necessidade de revisão do modelo ou da abordagem para melhorar a precisão das previsões.

- Erros Menores: Em outras semanas, como 05/10/2000, o erro é relativamente pequeno, -2.90, indicando uma previsão mais próxima do valor real. Isso mostra que o modelo consegue fornecer previsões razoavelmente precisas em alguns períodos.

Tabela 2: Comparação entre Valores Reais e Previstos com Erro

Data	**Real**	**Previsto**	**Erro**
05/10/2000	473	475.90	-2.90
12/10/2000	453	479.28	-26.28
19/10/2000	446	481.64	-35.64
26/10/2000	432	482.90	-50.90
02/11/2000	438	483.03	-45.03
09/11/2000	424	482.02	-58.02
16/11/2000	436	479.93	-43.93
23/11/2000	495	476.82	18.18
30/11/2000	458	472.80	-14.80
07/12/2000	483	468.00	15.00
14/12/2000	492	462.57	29.43
21/12/2000	485	456.68	28.32
28/12/2000	292	450.51	-158.51

Fonte: Autor

Tabela 2: Métricas de Erro
A Tabela 3 resume as principais métricas de erro para a avaliação do modelo:

- Erro Absoluto Médio (MAE): O MAE é 35,31, o que indica que, em média, o modelo está errando por 35,31 unidades em relação aos valores reais. Embora o MAE forneça uma visão geral da precisão do modelo, não captura a magnitude dos erros individuais.

- Erro Quadrático Médio (RMSE): O RMSE é 43.67, o que considera o quadrado dos erros. O RMSE é sensível a erros grandes e, portanto, pode refletir a presença de previsões muito imprecisas, como aquelas observadas em algumas datas específicas.

- Erro Percentual Absoluto Médio (MAPE): O MAPE é 7.82%, o que fornece uma medida da precisão do modelo em termos percentuais. Um MAPE abaixo de 10% geralmente é considerado bom, então, com um MAPE de 7.82%, o modelo mostra um desempenho razoável na previsão percentual dos valores.

Tabela 3: Métricas de Erro

Métrica	**Valor**
Erro Absoluto Médio (MAE)	35.31
Erro Quadrático Médio (RMSE)	43.67
Erro Percentual Absoluto Médio (MAPE)	7.82%

Fonte: Autor

Os resultados mostram que o modelo de regressão harmônica fornece previsões que são geralmente razoáveis, mas há espaço para melhoria. Erros significativos em algumas datas sugerem que o modelo pode não capturar todos os padrões da série temporal. As métricas de erro fornecem uma visão quantitativa do desempenho do modelo e ajudam a identificar áreas onde ajustes ou modelos alternativos podem ser necessários para melhorar a precisão das previsões.