

Análise da incidência de câncer por gênero em Belo Horizonte (2021-2025) usando ARIMA

Analysis of cancer incidence by gender in Belo Horizonte (2021-2025) using ARIMA model

Análisis de la incidencia de cáncer por género en Belo Horizonte (2021-2025) utilizando ARIMA

 Celso Correia de Souza¹

 João Bosco Arbués Carneiro
Júnior João²

 Raul Asseff Castelão¹

 Victor Rocha Pires de
Oliveira³

 Gisela Maria Jorgino Crespo¹

 Juliana Maria Correia de
Souza Vieira¹

¹Universidade Anhanguera Uniderp
Campo Grande, MS, Brasil

²Universidade Federal de
Rondonópolis
Rondonópolis, MT, Brasil

³Associação Beneficente Casa Rosa
Campo Grande, MS, Brasil

Autor correspondente:

João Bosco Arbués Carneiro Júnior
João
jbacj@hotmail.com

Submissão: 30 jul 2025

Aceite: 24 fev 2026

RESUMO. Objetivo: este estudo objetivou analisar a incidência de câncer por gênero em Belo Horizonte entre 2021 e 2025, utilizando o modelo ARIMA para prever a evolução dos casos. **Métodos:** a análise foi baseada em dados históricos de 2000 a 2020 fornecidos pelo Instituto Nacional de Câncer (INCA). **Resultados:** as correlações entre o aumento da população e os casos de câncer foram fortes: 0,93 para os homens e 0,95 para as mulheres. A correlação entre as incidências de câncer entre os gêneros foi de 0,98, indicando uma relação quase linear. Para aplicar o modelo ARIMA, foram calculados os parâmetros p, d e q, resultando em bons ajustes. O R² foi de 0,98 para a população, e de 0,86 e 0,82 para os gêneros masculino e feminino. **Conclusão:** as previsões indicam que, até 2025, a incidência de câncer continuará a crescer, evidenciando a eficácia do modelo e a necessidade de políticas públicas para prevenção e controle.

Descritores: Análise Temporal; Modelagem ARIMA; Câncer por Gênero; Saúde Pública; Valor Previsto de Incidência.

ABSTRACT. Objective: this study analyzed cancer incidence by gender in Belo Horizonte between 2021 and 2025, using the ARIMA model to forecast the evolution of cases. **Methods:** the analysis was based on historical data from 2000 to 2020 provided by the National Cancer Institute (INCA). **Results:** correlations between population growth and cancer cases were strong: 0.93 for men and 0.95 for women. The correlation between cancer incidences across genders was 0.98, indicating an almost linear relationship. To apply the ARIMA model, the parameters p, d, and q were calculated, resulting in good model fits. The R² value was 0.98 for the population, and 0.86 and 0.82 for the male and female groups, respectively. **Conclusion:** the forecasts indicate that, by 2025, cancer incidence will continue to rise, demonstrating both the effectiveness of the model and the need for public policies aimed at prevention and control.

Descriptors: Time Series Analysis; ARIMA Modeling; Gender-specific Cancer; Public Health; Predicted Value.

RESUMEN. Objetivo: este estudio analizó la incidencia de cáncer por género en Belo Horizonte entre 2021 y 2025, utilizando el modelo ARIMA para prever la evolución de los casos. **Métodos:** el análisis se basó en datos históricos de 2000 a 2020 proporcionados por el Instituto Nacional del Cáncer (INCA). **Resultados:** las correlaciones entre el aumento de la población y los casos de cáncer fueron fuertes: 0,93 para los hombres y 0,95 para las mujeres. La correlación entre las incidencias de cáncer entre los géneros fue de 0,98, lo que indica una relación casi lineal. Para aplicar el modelo ARIMA, se calcularon los parámetros p, d y q, obteniéndose buenos ajustes. El R² fue de 0,98 para la población total, y de 0,86 y 0,82 para los géneros masculino y femenino, respectivamente. **Conclusión:** las previsiones indican que, hasta 2025, la incidencia de cáncer continuará creciendo, lo que evidencia la eficacia del modelo y la necesidad de políticas públicas para la prevención y el control.

Descriptores: Análisis temporal; Modelización ARIMA; Cáncer por género; Salud pública; Valor previsto de incidencia.

INTRODUÇÃO

A incidência de câncer é uma preocupação crescente no cenário da saúde pública no Brasil, especialmente em grandes centros urbanos como Belo Horizonte. Com o aumento das taxas de prevalência dessa doença, torna-se essencial o desenvolvimento de estratégias eficazes para seu monitoramento e prevenção¹. O câncer é uma das principais causas de morte no país, afetando diferentes faixas etárias, gênero, raça, etc².

No contexto brasileiro, a literatura científica sobre a previsão da incidência de câncer tem avançado, com várias publicações científicas em veículos científicos nacionais ou internacionais, mas existem limitações de pesquisas que utilizam modelos estatísticos robustos para prever as tendências e previsões dessa doença³.

Este artigo se propõe a utilizar o modelo ARIMA (AutoRegressive Integrated Moving Average), uma ferramenta amplamente utilizada para análise de séries temporais, como método para prever a incidência de câncer em Belo Horizonte, com base em dados históricos da neoplasia no município, no período de 2000 a 2020, segmentada por gênero, para o planejamento de políticas públicas de saúde. A previsão da incidência de câncer pode fornecer informações essenciais para a alocação de recursos, o planejamento de campanhas de conscientização e a melhoria dos serviços de saúde⁴.

No entanto, um grande desafio é a obtenção de dados organizados e completos sobre essa neoplasia no Brasil. A falta de registros consistentes, a subnotificação em algumas regiões e a desigualdade no acesso aos serviços de saúde dificultam uma análise mais precisa da distribuição e da incidência do câncer, comprometendo o planejamento de políticas públicas eficazes e a alocação adequada de recursos para o combate à doença².

Para tanto, este estudo se baseará em dados do Instituto Nacional de Câncer (INCA), que oferece informações sobre a prevalência de câncer por gênero em diferentes regiões do Brasil, particularmente, em Belo Horizonte. A escolha do ARIMA justifica-se pela sua capacidade de lidar com séries temporais e prever tendências futuras a partir de dados históricos, permitindo uma análise detalhada da evolução da incidência de câncer nos anos anteriores².

A questão de pesquisa deste trabalho é: "Quais são as tendências de incidência de câncer por gênero em Belo Horizonte nos próximos anos, e como essas previsões podem auxiliar no planejamento de saúde pública?". Desse modo, o objetivo principal deste estudo FOI O DE analisar a incidência de câncer por gênero em Belo Horizonte entre 2021 e 2025, utilizando o modelo ARIMA para prever a evolução dos casos, com o intuito de fornecer subsídios para a formulação de políticas públicas mais eficazes na prevenção e no tratamento da doença. A pesquisa visa também identificar

as principais tendências de aumento ou diminuição dos casos de câncer, observando as diferenças entre os gêneros masculino e feminino.

Os resultados esperados incluem a previsão da incidência de câncer para o período de 2021 a 2025, com base nos dados históricos de 2000 a 2020. Espera-se que os resultados ofereçam uma visão clara das tendências da doença, permitindo que gestores de saúde pública adotem medidas mais proativas no combate ao câncer, ajustando suas ações de prevenção, diagnóstico precoce e tratamento de acordo com as projeções feitas pelo modelo. Além disso, a pesquisa contribuirá para o campo da saúde pública no Brasil, ao demonstrar como ferramentas quantitativas podem ser aplicadas na gestão de doenças crônicas, como o câncer.

MÉTODOS

Este estudo adota uma abordagem aplicada e descritiva, com o objetivo de utilizar o modelo ARIMA (AutoRegressive Integrated Moving Average) para prever a incidência de câncer por gênero em Belo Horizonte, contribuindo para o planejamento de políticas públicas de prevenção e controle da doença. A pesquisa foi desenvolvida com base em uma análise de séries temporais, com o intuito de projetar a evolução da incidência de câncer entre a população masculina e feminina. Segundo⁵, a modelagem de séries temporais é uma ferramenta poderosa para prever tendências e analisar dados de longo prazo, sendo particularmente útil em áreas como a saúde pública.

A pesquisa é quantitativa e utiliza dados secundários fornecidos pelo Instituto Nacional de Câncer (INCA), abrangendo o período de 2000 a 2020, não necessitando de parecer do Comitê de Ética. As variáveis analisadas foram a população de Belo Horizonte, a incidência de câncer, discriminada por gênero, e o tempo, representado pelos anos de 2000 a 2020. A variável dependente é a quantidade de casos de câncer diagnosticados, enquanto o tempo é a variável independente, usada para modelar e prever as tendências da incidência da doença ao longo do período de estudo. Essa abordagem de séries temporais permite captar as flutuações sazonais e tendências de longo prazo, que são cruciais para políticas de saúde pública⁶.

A incidência de câncer é um dos maiores desafios para a saúde pública mundial, sendo responsável por aproximadamente 10 milhões de mortes anuais, o que representa quase 1 em cada 6 óbitos no mundo. Estima-se que a incidência da doença continue a crescer, especialmente em países em desenvolvimento, devido ao envelhecimento da população e à adoção de estilos de vida que favorecem fatores de risco como o tabagismo, a obesidade e o sedentarismo. As projeções indicam que, até 2030, o número de novos casos de câncer e de mortes pela doença continuará a aumentar, o que torna imperativo o investimento em estratégias de prevenção e controle eficazes⁷.

No Brasil, o câncer também representa um grave problema de saúde pública. Dados do Instituto Nacional de Câncer (INCA) mostram que a doença é a segunda principal causa de morte no país, atrás apenas das doenças do aparelho circulatório.² Entre os tipos mais comuns de câncer no Brasil estão o câncer de mama, pulmão, cólon e reto, e próstata, com diferenças notáveis na prevalência entre homens e mulheres. A cada ano, o país registra centenas de milhares de novos casos e mortes, o que evidencia a necessidade de intensificar as políticas de prevenção, diagnóstico precoce e tratamento⁸.

Em Belo Horizonte, a situação não é diferente, pois, a evolução da doença na capital do estado de Minas Gerais, reflete as tendências observadas a nível nacional, com um crescente número de casos diagnosticados, que impacta tanto os sistemas de saúde pública quanto o planejamento das políticas de saúde. De acordo com dados do INCA, o câncer de mama é o tipo mais frequente entre as mulheres em Belo Horizonte, enquanto o câncer de próstata predomina entre os homens². Além disso, o aumento da incidência de cânceres relacionados ao estilo de vida, como os de pulmão e cólon, também tem gerado preocupação².

O Brasil tem avançado em estratégias de prevenção, com campanhas educativas voltadas para a conscientização sobre fatores de risco e a importância do diagnóstico precoce. No entanto, os desafios ainda são grandes. A cobertura de exames de rastreamento, como a mamografia e o exame de Papanicolau, ainda é desigual, com desigualdades regionais e dificuldades no acesso aos serviços de saúde, principalmente em áreas mais periféricas⁹. Além disso, é necessário fortalecer as políticas públicas de promoção de hábitos saudáveis, como a prática regular de atividades físicas, a alimentação balanceada e o combate ao tabagismo¹⁰.

Em Belo Horizonte, a preocupação com a prevenção e o controle do câncer tem levado a iniciativas locais de conscientização, com a promoção de campanhas educativas e a ampliação da oferta de serviços de saúde. A cidade tem investido na melhoria do acesso aos serviços de diagnóstico e tratamento, mas a demanda crescente por esses serviços ainda representa um desafio considerável para o sistema público de saúde¹.

O projeto Associação Beneficente Casa Rosa está sendo desenvolvido em Campo Grande com o objetivo de promover a prevenção e o tratamento do câncer em mulheres. Sob a coordenação de um médico cirurgião oncológico, a iniciativa busca oferecer suporte integral às mulheres diagnosticadas com câncer, desde o diagnóstico até o tratamento e acompanhamento. A Casa Rosa conta com a colaboração da Santa Casa de Misericórdia de Campo Grande, que fornece suporte logístico e assistencial. A associação oferece exames de rastreamento, orientação, apoio psicológico e terapias complementares para melhorar a qualidade de vida das pacientes^{11, 12, 13}.

Agora, a próxima etapa do projeto é a criação da Casa Azul, uma extensão da Casa Rosa, mas com foco no atendimento ao público masculino. Assim como a Casa Rosa, a Casa Azul terá como objetivo a prevenção e o tratamento do câncer em homens, principalmente o câncer da próstata, atendendo às suas necessidades específicas, com suporte médico, psicológico e assistencial, buscando reduzir a mortalidade masculina por esse tipo de câncer¹⁴.

O uso de modelos de previsão, como o ARIMA, pode ser uma ferramenta importante para antecipar tendências e ajudar na alocação eficiente de recursos para o controle e prevenção do câncer, contribuindo assim para a redução de sua carga sobre a população e o sistema de saúde¹⁵.

O modelo ARIMA foi escolhido devido à sua capacidade de modelar séries temporais com tendências e de prever valores futuros com base em dados históricos. O ARIMA é um modelo caracterizado por três parâmetros: p (ordem autoregressiva), d (ordem de diferenciação) e q (ordem da média móvel). A primeira etapa para aplicar o ARIMA consiste em garantir que a série temporal seja estacionária. Uma série é estacionária quando suas propriedades estatísticas, como média e variância, não variam ao longo do tempo. Caso a série não seja estacionária, um processo de diferenciação é realizado para torná-la estacionária. A diferenciação é feita subtraindo-se o valor anterior do valor atual da série, repetindo o processo até que a série atenda à condição de estacionariedade.

Após garantir a estacionariedade da série, é necessário determinar os parâmetros p, d e q. O parâmetro d é determinado pelo número de diferenciações necessárias para tornar a série estacionária. Os parâmetros p e q são definidos com base na análise das funções de autocorrelação (FAC) e autocorrelação parcial (FACP). A função de autocorrelação é usada para identificar a ordem da média móvel (q), enquanto a função de autocorrelação parcial ajuda na escolha da ordem autoregressiva (p). A escolha ideal desses parâmetros pode ser feita por meio da análise visual dos gráficos da FAC e FACP ou utilizando critérios de seleção de modelos, como o AIC (Akaike Information Criterion) e o BIC (Bayesian Information Criterion)¹⁶.

De acordo com⁵, um indicador comum de erro de previsão é a diferença entre o valor real (Z_t) e o valor previsto (\hat{Z}_t) no tempo correspondente ($\theta_t = Z_t - \hat{Z}_t$). Existem várias ferramentas para avaliar a qualidade do ajuste e o desempenho de um modelo de previsão, incluindo o coeficiente de ajustamento R^2 , a raiz quadrada do erro médio quadrático (RMSE - Root Mean Squared Error), o erro médio absoluto percentual (MAPE), o erro médio absoluto (MAE) e o BIC normalizado.

As equações (1) e (2) representam matematicamente os erros de ajustamentos do modelo.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(y_{i,obs} - y_{i,pred})^2}{N}} \quad (1)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_{i,obs} - y_{i,pred}|}{y_i} \quad (2)$$

Onde n é o número de períodos da série temporal; $|*|$ é o valor absoluto da variável $*$; $y_{i,obs}$: valor observado i e $y_{i,pred}$: valor predito i .

O que é considerado um "RMSE baixo" ou "alto" depende da escala dos dados e do contexto. Um RMSE é baixo quando o valor está em uma faixa pequena em relação à magnitude dos dados, e alto quando as previsões estão distantes dos valores reais. Por exemplo, em séries com valores na ordem de milhões, um RMSE de 10.000 pode ser aceitável, mas em séries de valores menores, esse erro seria significativo. É importante comparar o RMSE com outras métricas, como o MAPE, e com o RMSE de outros modelos para avaliar o desempenho. Valores de RMSE entre 5% a 10% do valor médio da variável geralmente são considerados razoáveis.

Já, segundo¹⁷, menores valores de MAPE indicam melhor desempenho do modelo de previsão, que pode ser apresentado em níveis de aceitação do erro: quando o MAPE é inferior a 5%, a precisão é considerada **excelente**, indicando alta confiabilidade nas previsões. Um MAPE entre 5% e 10% é classificado como **bom**, o que significa precisão moderada. Quando o MAPE está entre 10% e 20%, a precisão é considerada **regular**, refletindo uma confiabilidade razoável. Por fim, se o MAPE for superior a 20%, a precisão do modelo é considerada **ruim**, indicando baixa confiabilidade nas previsões.

O BIC normalizado, também conhecido como Critério de Schwarz, é um critério de informação bayesiano que considera a densidade a posteriori dos parâmetros do modelo ajustado. O melhor modelo é aquele que apresenta o menor valor do BIC¹⁸.

A equação (3) representa matematicamente o grau de ajustamento do modelo pelo BIC normalizado.

$$BIC = T \ln(\sigma_{\epsilon,ML}^2) + n \ln(T) \quad (3)$$

Onde " n " é o número de parâmetros estimados, " T " é o número de observações da série temporal, $\ln(.)$ representa o logaritmo natural de $(.)$, e $\sigma_{\epsilon,ML}^2$ é a estimativa de máxima verossimilhança de ϵ . Quanto menores essas medidas, melhor será o desempenho do modelo de previsão, indicando um menor erro nas previsões.

Para a implementação do modelo ARIMA e a realização das previsões, foi utilizado o software SPSS (Statistical Package for the Social Sciences). O SPSS oferece uma interface amigável para a realização de análises estatísticas e a implementação de modelos de séries temporais, incluindo o ARIMA, facilitando a análise e a interpretação dos resultados. O SPSS também permite a realização de testes de estacionariedade e a visualização das funções de autocorrelação, proporcionando um ambiente completo para a construção e avaliação do modelo¹⁹.

O modelo ARIMA foi escolhido devido à sua capacidade de modelar séries temporais e fazer previsões de valores futuros. O processo de aplicação do ARIMA consistiu nas seguintes etapas: A primeira etapa foi garantir que a série temporal fosse estacionária, ou seja, suas propriedades estatísticas não deveriam variar ao longo do tempo. Caso a série não fosse estacionária, seria realizado um processo de diferenciação para torná-la estacionária. Essa técnica é fundamental, pois modelos de séries temporais estacionárias têm desempenho superior em termos de previsibilidade²⁰.

Após a obtenção da estacionariedade da série, os parâmetros do modelo ARIMA (p, d, q) foram identificados. O parâmetro d foi determinado com base no número de diferenciações necessárias para garantir a estacionariedade. Os parâmetros p (ordem autoregressiva) e q (ordem da média móvel) foram definidos com base nas funções de autocorrelação (FAC) e autocorrelação parcial (FACP), além de critérios de seleção de modelos, como o BIC normalizado (Bayesian Information Criterion). Segundo²¹, a análise de autocorrelação é uma ferramenta chave na identificação de padrões dentro de séries temporais.

A qualidade do ajuste do modelo foi avaliada utilizando diversos critérios, como o coeficiente de ajuste R^2 , o erro médio quadrático (RMSE) e o erro médio absoluto percentual (MAPE). O BIC normalizado, equação (3), também foi utilizado para selecionar o melhor modelo, com base na comparação entre modelos ajustados. De acordo com¹⁷, o AIC (Akaike Information Criterion) e o BIC são amplamente utilizados para a seleção de modelos, especialmente quando se deseja comparar a complexidade de diferentes ajustes de modelos.

A implementação do modelo ARIMA e a realização das previsões foram realizadas no software SPSS (Statistical Package for the Social Sciences). O SPSS foi escolhido devido à sua capacidade de realizar análises estatísticas avançadas e à sua interface amigável, que facilita tanto a construção do modelo quanto a avaliação dos resultados. O SPSS também foi utilizado para realizar testes de estacionariedade e para a visualização das funções de autocorrelação, oferecendo um ambiente completo para a modelagem e interpretação dos dados. A escolha do SPSS é válida devido à sua robustez e ampla utilização em estudos de saúde pública, especialmente para análise de séries temporais²².

RESULTADOS E DISCUSSÃO

Nesta seção apresentaremos a análise da incidência de câncer em Belo Horizonte, com base nos dados de 2000 a 2020, discriminados por gênero. O quadro 1 mostra, segundo os^{2,23}, a evolução dos casos registrados de câncer nas populações masculina e feminina durante este período. A partir desses dados, será possível identificar tendências e padrões, além de possíveis implicações para implementações de políticas de saúde pública e estratégias de prevenção.

Quadro 1. População de Belo Horizonte e incidência de câncer, por gênero (masculina e feminina) na cidade, no período de 2000 a 2020.

Ano	População	Masculina	Feminina
2000	2238526	2798	3169
2001	2246819	3263	3850
2002	2255292	3334	3721
2003	2264021	3473	3663
2004	2273011	3582	3604
2005	2282348	3415	3799
2006	2291969	3497	3934
2007	2301885	3914	4452
2008	2312029	3726	4468
2009	2322458	4508	4956
2010	2375151	4553	5591
2011	2384246	4352	5278
2012	2393918	4982	5587
2013	2403754	4305	5435
2014	2413813	5023	5535
2015	2424051	5163	5776
2016	2434472	4871	5744
2017	2445059	6070	7093
2018	2455879	6086	7078
2019	2466901	6536	7609
2020	2478184	5364	6239

Fonte: Elaborado pelos Autores com dados de².

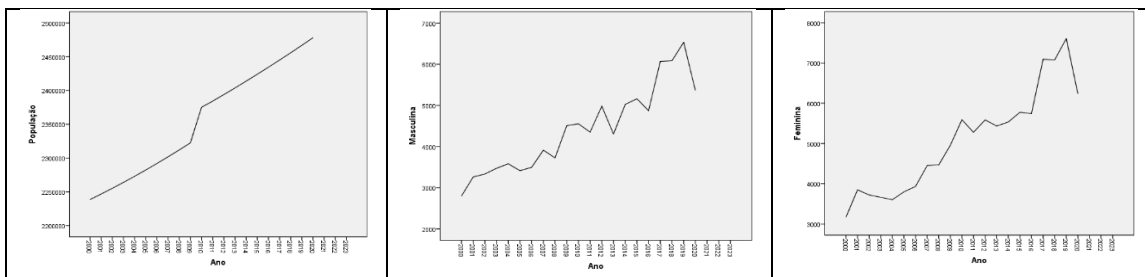
Inicialmente, foi calculada a matriz de correlação entre as variáveis “população”, “masculina” e “feminina”, tabela 1, que trata das incidências de câncer por gênero em Belo Horizonte, para verificar a associação entre elas, com base nos dados apresentados no quadro 2. A análise de correlação tem como objetivo identificar a intensidade e a direção das relações entre o número de casos de câncer nas populações masculina e feminina e o tamanho da população total ao longo dos anos. A partir dos resultados dessa análise, será possível entender se o aumento da população ao longo do tempo tem alguma relação direta com o crescimento das taxas de incidência de câncer em ambos os gêneros, fornecendo informações para a interpretação dos padrões observados.

Observou-se que a correlação entre a população e a incidência de câncer na população masculina é de 0,93, e entre a população e a incidência de câncer na população feminina, de 0,95,

indicando que as variações de uma acompanham as variações das outras. A correlação entre incidências de câncer nas populações masculina e feminina é de 0,98, sugerindo que as variações em ambos os grupos de gênero são quase perfeitamente sincronizadas. Portanto, a estrutura populacional da cidade manteve um equilíbrio de incidências de câncer entre os gêneros durante esse período.

Para aplicar o método ARIMA (p, d, q) em previsões, considera-se, inicialmente, $p = d = q = 0$, sendo essencial verificar a estacionariedade das variáveis. Nesse sentido, optou-se por verificar a estacionariedade graficamente. Ao construir os gráficos das variáveis população e incidências de câncer nas populações masculina e feminina, no período de 2000 a 2020, obteve-se a figura 1.

Figura 1. Gráficos das variáveis “População” e incidências de câncer nas populações “Masculina” e Feminina”, no período de 2000 a 2020.

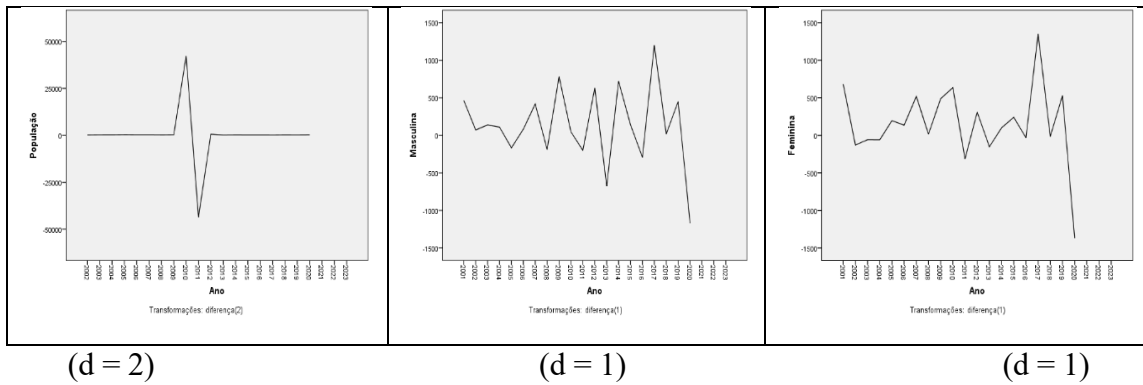


Fonte: Autores (2025).

Como apresentadas na figura 1, todas as três variáveis apresentam gráficos com **tendências de crescimentos** ao longo do período analisado, de 2000 a 2020, mas sem uma periodicidade clara, indicando que **não são estacionárias**. As oscilações de curto prazo nas populações masculina e feminina não alteram o padrão geral de aumento, evidenciando que o número de casos de câncer nas duas populações tem seguido um crescimento sustentado. A diferenciação (d) das séries seria necessária para tratar essas não estacionariedade das séries e possibilitar a aplicação do modelo ARIMA para previsões em cada uma delas.

Para tornar as três séries estacionárias foram realizadas as seguintes diferenciações: a variável população foi submetida a 2 diferenciações ($d = 2$), enquanto as populações masculina e feminina passaram por 1 diferenciação cada ($d = 1$, tornando-se estacionárias, conforme ilustrado nos gráficos da figura 2.

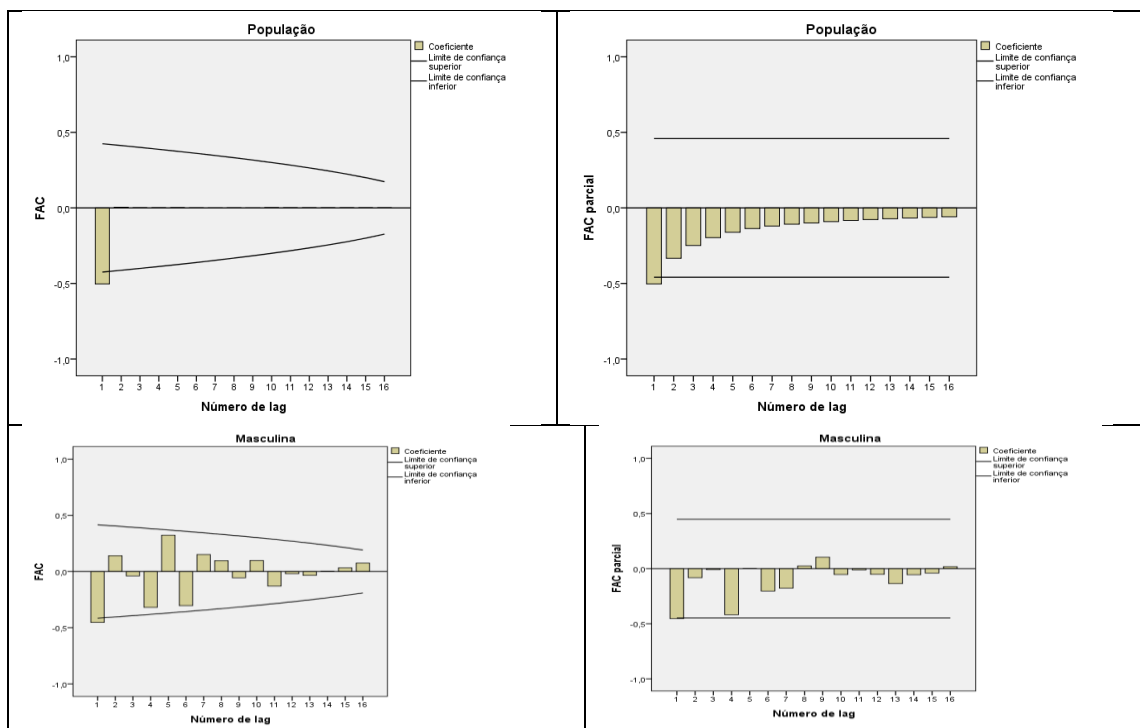
Figura 2. Gráficos das variáveis população de BH ($d = 2$) e masculina e feminina ($d = 1$).

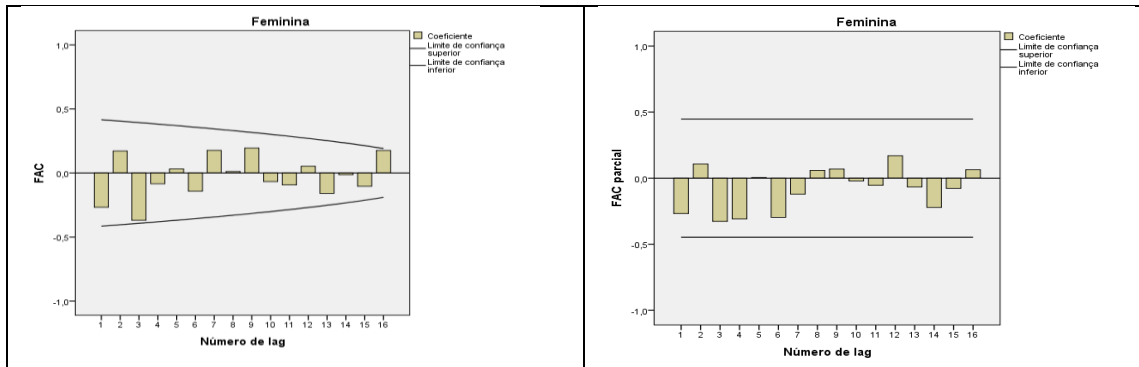


Fonte: Autores (2025).

Na figura 2, observa-se que os três gráficos das variáveis analisadas apresentam estacionaridade, com oscilações em torno de zero. Com isso, é possível prosseguir com o processo de previsão utilizando as variáveis diferenciadas, considerando os respectivos valores de d . O próximo passo é determinar os valores de p e q para cada variável, a fim de obter os modelos ARIMA correspondentes. A escolha dos valores de p e q foi baseada na análise gráfica das funções de autocorrelação (FAC) e das funções de autocorrelação parcial (FACP), observando os lags que extrapolam os intervalos de confiança. A figura 3 apresenta os gráficos utilizados para a determinação dos valores de p e q das variáveis.

Figura 3. Gráficos das funções de autocorrelações (FAC) e de autocorrelações parciais (FACP) das variáveis “população”, incidências de câncer nas populações “masculina” e “feminina”.





Fonte: Autores (2025).

Da observação dos gráficos da figura 3, p e q são determinados observando os lags que ultrapassam os intervalos de confiança. Assim, com relação à variável “população”, tem-se os modelos: ARIMA (1, 2, 1) para a variável “população”; ARIMA (0, 1, 1) para a variável masculina e; ARIMA (0, 1, 0) para a variável feminina. Construídos os modelos, a próxima fase é determinar as previsões para cada variável com o uso do SPSS.

Realizou-se a previsão de casos futuros, de 2021 a 2025, com os modelos ARIMA citados acima, para cada série de dados, de 2000 a 2020. Como ferramenta estatística utilizou-se o software Statistical Package for the Social Sciences (SPSS) para realizar essa previsão.²⁰ Os resultados estatísticos do ajuste do modelo, bem como o teste de Ljung-Box, podem ser encontrados na tabela 1.

Tabela 1. Estatísticas do ajuste do modelo da variável ETB, com preditores, estatísticas do modelo, teste de Ljung-Box e número de valores discrepantes.

Variáveis	Modelo ARIMA	Estatísticas de ajuste do modelo		Ljung-Box Q(18)			Erros	
		R ²	BIC normalizado	Estatísticas	gl	p	MAPE (%)	RMSE
População	(1, 2, 1)	0,98	9,25	1,159	16	1,000	0,17	11.077,76
Masculina	(0, 1, 1)	0,86	12,44	14,102	17	0,660	6,48	401,63
Feminina	(0, 1, 0)	0,82	12,87	27,630	18	0,068	6,86	537,11

Fonte: Autores (2025).

A tabela 1 apresenta as estatísticas do ajuste do modelo ARIMA para as variáveis população, masculina e feminina. O modelo ARIMA (1, 2, 1) da população, com um R² de 0,98, indica alta precisão e bom ajuste. Para a população masculina, o modelo (0, 1, 1) apresentou um R² de 0,86, e para a feminina, com o modelo (0, 1, 0), o R² foi de 0,82, sugerindo uma menor precisão para essas

variáveis. O BIC normalizado para a população foi 9,25, refletindo um bom ajuste, enquanto para as variáveis masculina e feminina, o BIC foi mais alto, com 12,44 e 12,87, respectivamente, indicando uma adequação um pouco inferior. O teste de Ljung-Box mostrou que, para a população, o valor de $p = 1,000$ indica ausência de autocorrelação nos resíduos, enquanto para as séries masculina e feminina, os valores de p foram 0,660 e 0,068, sugerindo possível autocorrelação residual, o que pode indicar a necessidade de ajustes no modelo.

O erro MAPE de ajuste foi de 0,17% para a população, classificando-se como excelente, enquanto para as variáveis masculina e feminina, os valores de MAPE foram 6,48% e 6,86%, respectivamente, o que é considerado bom. O RMSE para a população foi de 11.077,76, um valor alto, refletindo a grande magnitude dos dados, enquanto para a população masculina e feminina, o RMSE foi significativamente mais baixo, 401,63 e 537,11, respectivamente. No geral, o modelo ARIMA teve um bom desempenho, mas ajustes podem ser necessários, especialmente para as variáveis masculina e feminina, para melhorar a precisão das previsões.

Na tabela 2 estão relacionados os valores observados e previstos da população de Belo Horizonte para os anos (2021-2025), bem como o erro MAPE determinado entre valores observados e previstos. Observa-se que os valores previstos foram obtidos até 2025, portanto para os anos 2024 e 2025 não houve comparação pela ausência de dados observados para 2024 e 2025.

Tabela 2. Valores Observados da População de Belo Horizonte (2021-2023) e dados Previstos (2021-2025), com Cálculo do Erro MAPE até 2023.

Ano	Valores observados da população de Belo Horizonte*	Valores previstos da população de Belo Horizonte**
2021	2 530 701	2.490.167
2022	2 305 560	2.502.150
2023	2 416 339	2.514.133
2024	-	2.526.116
2025	-	2.538.098
MAPE (2021-2023)		4,77%

*Fonte: IBGE²⁴. **Fonte: Autores (2025).

A tabela 2 apresenta os valores da população de Belo Horizonte: observados (21-23) e previstos (21-25). Calculou-se o erro MAPE (2021-2023), que foi de 4,77% indicando que as previsões produzidas pelo modelo são excelentes, segundo o quadro 1, com um erro relativamente baixo, sugerindo boa qualidade no ajuste do modelo ARIMA para esses dados.

A tabela 3 apresenta a previsão da evolução, feita com o modelo ARIMA, da incidência de câncer por gênero em Belo Horizonte entre 2021 e 2025.

Tabela 3. Evolução prevista pelo modelo ARIMA da Incidência de Câncer por Gênero em Belo Horizonte (2021-2025).

Ano	Masculino	Feminino
2021	6375	6269
2022	6591	6288
2023	6812	6295
2024	7038	6290
2025	7270	6274

Fonte: Autores (2025).

Os dados da tabela 3 mostram as previsões do número de casos de câncer diagnosticados para as populações masculina e feminina de Belo Horizonte ao longo dos anos de 2021 a 2025, evidenciando o crescimento contínuo da incidência para ambos os gêneros durante o período projetado. No entanto, não foi possível encontrar na bibliografia dados observados de câncer por gênero em Belo Horizonte a partir de 2021, o que, infelizmente, não permitiu o cálculo do erro MAPE, como foi feito na tabela 2.

CONCLUSÃO

Este estudo teve como objetivo utilizar o modelo ARIMA para prever a incidência de câncer por gênero em Belo Horizonte entre 2021 e 2025, com base em dados de 2000 a 2020 fornecidos pelo INCA (2023). A análise revelou uma correlação significativa entre a população e a incidência de câncer nos gêneros masculino e feminino, com valores de 0,93 e 0,95, respectivamente, o que indica que o aumento da população é acompanhada pelo crescimento das taxas de câncer. A correlação entre as incidências de câncer nos dois gêneros foi ainda mais alta, atingindo 0,98, sugerindo uma relação quase perfeita entre os dois grupos.

A metodologia ARIMA foi aplicada após a diferenciação das séries temporais, já que as variáveis apresentavam tendência de crescimento e não eram estacionárias. O modelo ARIMA (1, 2, 1) para a população, (0, 1, 1) para o gênero masculino e (0, 1, 0) para o feminino foram ajustados com base nas funções de autocorrelação. A partir desses modelos, foram feitas as previsões para os anos de 2021 a 2025, com destaque para o bom desempenho da previsão da RNA_MLP para a

população, com um MAPE de ajuste do modelo de 0,17%, classificando-se como excelente, e valores de RMSE mais elevados, como esperado devido à magnitude dos dados populacionais.

Contudo, para os subgrupos masculino e feminino, os modelos apresentaram um desempenho ligeiramente inferior, com MAPEs de ajustes dos modelos de 6,48% e 6,86%, respectivamente, ainda dentro de uma faixa considerada boa, conforme a classificação de¹³. No entanto, as previsões geradas para a incidência de câncer por gênero mostraram um crescimento contínuo da doença entre 2021 e 2025, mas, infelizmente, a falta de dados observacionais após 2021 impossibilitou a validação do modelo para esses anos, o que impediu a avaliação do erro MAPE entre valores observados e previstos para esse período.

Dessa forma, o objetivo deste estudo foi parcialmente alcançado, já que as previsões geradas foram adequadas para os ajustes dos modelos, mas a ausência de dados observacionais atualizados impediu a análise comparativa do erro entre valores observados e previstos para a incidência de câncer por gênero, o que limita a avaliação total da acuracidade do modelo. Este estudo, entretanto, proporciona uma base importante para o planejamento de políticas públicas de saúde, destacando a necessidade de disponibilização de dados mais atualizados para um acompanhamento mais preciso da evolução do câncer em Belo Horizonte.

REFERÊNCIAS

1. Santos MO, Lima FCS, Martins LFL, Oliveira JFP, Almeida LM, Cancela MC. Estimativa de incidência de câncer no Brasil, 2023-2025. *Rev Bras Cancerol.* 2023;69(1):e-213700. doi:10.32635/2176-9745.RBC.2023v69n1.3700.
2. Instituto Nacional de Câncer (INCA). Câncer no Brasil: dados e informações. 2023. Disponível em: <https://www.gov.br/inca/pt-br>. Acesso em: 25 jan. 2025.
3. Oncoguia. Estudos apontam tendência de aumento de casos de câncer em adultos até 2030. 2023. Disponível em: <https://www.oncoguia.org.br/conteudo/estudos-apontam-tendencia-de-aumento-de-casos-de-cancer-em-adultos-ate-2030/16658/7/>. Acesso em: 27 jan. 2025.
4. Morais BL, Meira GM, Xavier BB, Lima AP, Afonso MPD. Avaliação do rastreamento para câncer de colo do útero em Belo Horizonte de 2018 a 2023: uma análise de séries temporais. *Contrib Ciênc Soc.* 2024;17(8):1-22. doi:10.55905/revconv.17n.8-085.
5. Montgomery DC, Jennings CL, Kulahci M. *Introduction to time series analysis and forecasting.* Hoboken (NJ): John Wiley & Sons; 2008.
6. Box GEP, Jenkins GM, Reinsel GC. *Time series analysis: forecasting and control.* 5th ed. Hoboken (NJ): Wiley; 2015.
7. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. *Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in*

- 185 countries. CA Cancer J Clin. 2022;72(3):209-49. doi:10.3322/caac.21834. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/38572751/>. Acesso em: 15 jan. 2025.
8. Brasil. Ministério da Saúde. Instituto Nacional de Câncer (INCA). Estimativa 2020: incidência de câncer no Brasil. 2020. Disponível em: <https://www.gov.br/>. Acesso em: 20 jan. 2025.
 9. Pizão P. Câncer: desigualdade de acesso, diagnóstico precoce e tratamento. 2024. Disponível em: <https://medicinas.com.br/combate-desigualdade/>. Acesso em: 25 jan. 2025.
 10. Horbus E, Costa P. Políticas públicas de saúde para o tratamento do câncer no Brasil e a concentração regional das unidades de tratamento. Rev Bras Geogr Econ. 2023. Disponível em: <https://journals.openedition.org/espacoeconomia/23273>. Acesso em: 25 jan. 2025.
 11. Câmara Municipal de Campo Grande. Casa Rosa: oferece apoio às mulheres na luta contra o câncer de mama. 2024. Disponível em: <https://camara.ms.gov.br/vereador-dr-victor-rocha/casa-rosa-oferece-apoio-as-mulheres-na-luta-contr-o-cancer-de-mama/>. Acesso em: 15 jan. 2025.
 12. Câmara Municipal de Campo Grande. Casa Rosa: há um ano na luta contra o câncer de mama, fortalecendo o SUS em Campo Grande. 2024. Disponível em: <https://camara.ms.gov.br/vereador-dr-victor-rocha/casa-rosa-ha-um-ano-na-luta-contr-o-cancer-de-mama-fortalecendo-o-sus-em-campo-grande/>. Acesso em: 15 jan. 2025.
 13. O Estado. Quer saber mais sobre a Casa Rosa? 2024. Disponível em: <https://www.facebook.com/drvictorrochavr/videos/quer-saber-mais-sobre-a-casa-rosaesse-projeto-tem-por-objetivo-reduzir-a-mortali/475962337010015/>. Acesso em: 15 jan. 2025.
 14. Folha MS. Dr. Victor Rocha apresenta projeto Casa Azul na Câmara Municipal de Campo Grande. 2025. Disponível em: <https://folhams.com.br/2025/01/10/dr-victor-rocha-apresenta-projeto-casa-azul-na-camara-municipal-de-campo-grande/>. Acesso em: 15 jan. 2025.
 15. Anffe RCM, Espinosa MM, Souza PCF, Galvão ND. Modelos para previsão das taxas de incidência e mortalidade do câncer do colo do útero. J Health NPEPS. 2022;7(2):e10446. doi:10.30681/2526101010446.
 16. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, editors. Proceedings of the 2nd International Symposium on Information Theory; 1973; Budapest. Budapest: Akademiai Kiado; 1973. p. 267-281. doi:10.1007/978-1-4612-1694-0_15.
 17. Carneiro Júnior JBA, Souza CC. Aplicação de redes neurais artificiais na previsão do produto interno bruto do Mato Grosso do Sul em função da produção de cana-de-açúcar, açúcar e etanol. Rev Ibero-Am Cienc Ambient. 2019;10(5):1-10. doi:10.6008/CBPC2179-6858.2019.005.0019.
 18. Souza CC, Carneiro Júnior JBA, Cristaldo MF, Castelão RA, Frainer DM, Viganó HHG. Previsão de produção de cana-de-açúcar e seus derivados através de ARIMA no período de 2020 a 2025 na macrorregião Centro-Sul do Brasil. Res Soc Dev. 2022;11(4):e27048. doi:10.33448/rsd-v11i427048.
 19. IBM Corp. IBM Forecasting 28: previsão. 2024. Disponível em: https://www.ibm.com/docs/en/SSLVMB_28.0.0/pdf/IBM_SPSS_Forecasting.pdf. Acesso em: 10 dez. 2024.

20. Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice. 2nd ed. Melbourne (AU): OTexts; 2018. Disponível em: <https://otexts.com/fpp2/>. Acesso em: 25 jan. 2025.
21. Chatfield C. The analysis of time series: an introduction. 6th ed. Boca Raton (FL): Chapman & Hall/CRC; 2003.
22. Field A. Discovering statistics using SPSS. 5th ed. Thousand Oaks (CA): SAGE Publications; 2018.
23. Brasil. Ministério da Saúde. Instituto Nacional de Câncer (INCA). Estimativa 2023: incidência de câncer no Brasil. 2023. Disponível em: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files/media/document/estimativa-2023.pdf>. Acesso em: 10 jan. 2025.
24. Instituto Brasileiro de Geografia e Estatística (IBGE). Projeções da população. 2025. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9109-projecao-da-populacao.html?edicao=41053>. Acesso em: 11 jan. 2025.