



Curve Fitting and Autoregressive Models in Admissions Data for Respiratory Diseases

Ajuste de Curvas e Modelos Autorregressivos em Dados de Internações por Doenças Respiratórias

Raphael de Oliveira Garcia¹ , Graciele Paraguaia Silveira² , Bruna Santos Silva³

Received: August 16, 2025 Received in revised form: November 20, 2025 Accepted: December 13, 2025 Available online: December 18, 2025

ABSTRACT

This study investigated hospitalization costs for respiratory diseases in the state of São Paulo between 2002 and 2025, analyzing the dynamics of these expenses within the Brazilian Unified Health System (SUS) and contextualizing the challenges of financing and managing public health in Brazil. The objective of this work was to identify historical patterns and understand expenditures by analyzing data from the DATASUS Hospital Information System. Using polynomial, logistic, and trigonometric curve fitting, as well as statistical time series models, with emphasis on the seasonal ARIMA model, seasonal trends and patterns in hospital expenditures were identified. This allowed for the capture of seasonal behaviors related to the amounts paid, specifically regarding the increase in hospitalizations for respiratory diseases during the transition between fall and winter. Finally, in the graphs associated with fluctuations around the average trend, two level transitions were observed in relation to the values paid: one associated with COVID-19 (mid-2020) and the other associated with H1N1 (in 2009). These levels correspond to the values of $K = 22.65$ million, referring to the pre-COVID logistics model, and to $K = 35$ million (logistics model that includes the COVID-19 pandemic period).

keywords hospital expenses, respiratory diseases, curve fitting, time series, ARIMA model

RESUMO

Este estudo investigou os custos de internação por doenças respiratórias no estado de São Paulo entre 2002 e 2025, analisando a dinâmica dessas despesas no âmbito do Sistema Único de Saúde (SUS) e contextualizando os desafios de financiamento e gestão da saúde pública no Brasil. O objetivo deste trabalho foi identificar padrões históricos e compreender os gastos por meio da análise de dados do Sistema de Informações Hospitalares DATASUS. Utilizando ajustes de curvas polinomiais, logísticas e trigonométricas, bem como modelos estatísticos de séries temporais, com ênfase no modelo ARIMA sazonal, foram identificadas tendências e padrões sazonais nos gastos hospitalares. Isso permitiu captar comportamentos sazonais relacionados aos valores pagos, especificamente no que diz respeito ao aumento das internações por doenças respiratórias durante a transição entre outono e inverno. Por fim, nos gráficos associados às flutuações em torno da tendência média, observaram-se duas transições de nível em relação aos valores pagos: uma associada à COVID-19 (meados de 2020) e a outra associada à H1N1 (em 2009). Esses níveis correspondem aos valores de $K = 22,65$ milhões, referentes ao modelo logístico pré-COVID, e $K = 35$ milhões (modelo logístico que inclui o período da pandemia de COVID-19).

palavras-chave gastos hospitalares, doenças respiratórias, ajuste de curvas, séries temporais, modelo ARIMA

¹Prof. Dr., Department of Actuarial Sciences, UNIFESP, Osasco, SP, Brazil. rogarcia@unifesp.br

²Department of Physics, Chemistry and Mathematics, UFSCar, Sorocaba, SP, Brazil. graciele@ufscar.br

³Graduated in Actuarial Science, UNIFESP, Osasco Campus, Osasco, SP, Brazil. bruna.santos29@unifesp.br

Introduction

Public health in Brazil faces significant challenges in ensuring universal care and the sustainability of the Unified Health System (SUS), from comprehensive support for the entire population to structural challenges that threaten its sustainability. Issues such as chronic underfunding, unequal resource distribution, growing demand due to population aging, and the impacts of external events are central to understanding the difficulties faced by the system.

Various factors can be observed, including the different types of hospitalizations due to respiratory diseases, which account for a significant portion of hospitalizations in Brazil. These pathologies, often seasonal, are aggravated by environmental factors such as pollution and climate change, and are more prevalent among vulnerable groups such as the elderly and children.

The economic impact of these diseases is considerable. Araújo et al. (2021) estimates that influenza infection, a leading cause of respiratory illness, generated total costs of R\$ 5.62 billions for the health system in 2019, taking into account both direct costs (hospitalizations and consultations) and indirect costs (loss of productivity).

Population aging also worsens the situation of respiratory diseases in Brazil. Reis et al. (2016) projects that the elderly population is more likely to be hospitalized than other age groups, due to the health characteristics of this group with a prevalence of chronic diseases. This population, being more vulnerable, demands greater attention and generates higher costs for the health system.

Another sensitive population is the newborns group, aged up to 1 year. Marques et al. (2025) estimated the direct costs and analyzed the epidemiological aspects of Hospitalizations for Primary Care-Sensitive Conditions (ICSAP) in children under one year old, in the city of São Paulo, during the period 2011-2022, taking as reference the 19 groups of diseases, with 74 diagnoses classified according to the 10th International Classification of Diseases and Related Health Problems (ICD-10).

Furthermore, regional inequality in access to health services also exacerbates the impact of these diseases. The World Health Organization report (World Health Organization, 2025) points out that health inequalities shorten lives by decades. Health follows a social gradient: the more disadvantaged the area in which individuals live, the lower their income, the fewer years of education, the poorer their health, and the fewer healthy years of life. These inequalities are exacerbated in populations facing discrimination and marginalization.

This report highlights that Latin America and the Caribbean continues to be the region with the highest levels of inequality in the world. To make matters worse, the COVID-19 pandemic has exacerbated the situation: in 2020, the regional economy contracted 7%, the largest decline in 120 years, leaving millions of people without income or social protection, meaning that income recovery has generally been slow (World Health Organization, 2025).

The Brazilian context is no different; the same situation recurs: regional inequality in access to health services also exacerbates the impact of these diseases. Andrade et al. (2011) point out that, while states like São Paulo concentrate the majority of highly complex services, peripheral regions suffer from a lack of adequate infrastructure. This disparity compromises equitable access to treatment and prevention of respiratory diseases. Therefore, tackling respiratory diseases in Brazil requires an integrated approach that includes investments in prevention, improvements in air quality, health promotion policies, and reducing regional inequalities. These actions are essential to mitigate the clinical and economic impacts of these conditions, promoting a more efficient and equitable healthcare system for the entire Brazilian population.

Targeting the São Paulo state, both for its high population rate and the wide hospital network available, the objective of this study was to monitor the evolution of values related to the costs of hospital admissions for respiratory diseases, between 2002 and 2025, via analysis of pre-covid data and complete data. For this purpose, curve fitting using the least squares method combining polynomial and trigonometric functions were implemented, in addition to time series statistical techniques, namely, Autoregressive Integrated Moving Average models with Seasonality (SARIMA or seasonal ARIMA).

Material and methods

The data were obtained from the SUS Hospital Information System (SIH/SUS), made available by Ministério da Saúde (2025), which gathers detailed information since 1984 on hospital admissions carried out in Brazil. Hospital Admission Authorizations (AIHs) paid by place of admission in the São Paulo state, referring to diagnoses, were considered. of ICD-10 for diseases of the respiratory system. The information analyzed refers exclusively to the hospitalization values paid by AIHs processed monthly between January 2002 and March 2025, Figure 1.

Figure 1 - Amounts spent per month, in Reais, on hospital admissions for respiratory diseases in the of São Paulo state. In blue, data from Jan/2002 to Feb/2020 and in red, data from Mar/2020 to Mar/2025. The gray rectangle represents the period of the COVID-19 pandemic.

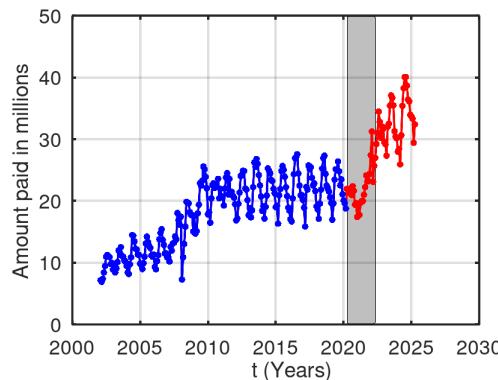


Figure 1 shows in blue the data from 2002 to February 2020. In red, we have the data from March 2000 to March 2025. The gray rectangle highlights the period of the COVID-19 pandemic in Brazil, that is, from March 2020 to April 2022, when the federal government declared the COVID-19 pandemic in Brazil and revoked the pandemic state in Brazil, respectively.

Polynomial and logistic curve fits were made to the central mean tendency, considering pre-covid data and total data. The characteristic oscillations present in the data, due to seasonal behavior, were adjusted with trigonometric sine and cosine curves. In all these listed cases, the least squares method was used, the solution of the associated linear system of which was obtained by the method GMRES - Generalized Minimum Residual Method. A polynomial basis for parabolic fitting was joined with sine and cosine functions of the type $\sin(\omega_j t)$, $\cos(\omega_j t)$, of frequencies $\omega_j = \frac{4j}{100}$, with $j = 1, \dots, 50$.

The condition number of the matrix associated with the linear system of the least squares method, with the trigonometric base, was decisive for defining the number of sine and cosine functions (Ruggiero & Lopes, 2000). With 100 functions (50 sines and 50 cosines) it is still possible to obtain a matrix with a low condition number, approximately 5, ensuring the convergence of the GMRES method and avoiding spurious results. For example, with 120 functions there are already solutions with spurious results and for example, with 180 functions, the number of conditions reaches the order of 10^6 and the GMRES method does not converge.

The ARIMA (AutoRegressive Integrated Moving Average) model is widely used for time series modeling due to its ability to capture temporal dependence patterns and smooth seasonal variations and trends. This model combines three main components: (i) autoregressive (AR), which represents the linear relationship between past observations of the series; (ii) integration (I), which transforms a non-stationary series into a stationary one through differentiation; and (iii) moving average (MA), which incorporates past errors to improve the prediction of future values Morettin and Toloi (2018). A description of the ARIMA model applied to time series associated with biological problems can be found in Silva et al. (2021).

The AR(p) (p-order Autoregressive) processes are characterized by a gradually decaying FAC, which can exhibit exponential or oscillatory behavior, depending on the model coefficients. This type of decay is infinite in length, reflecting the prolonged dependence between the pasts of the series values and its future states.

MA(q) processes (moving averages of order q) have a distinct structure, in which the FAC is finite and presents a cutoff after q lags. This occurs because the values of the series are influenced only by past residuals, and not by the values of the dependent variable itself, which limits the persistence of autocorrelation.

ARMA(p,q) models combine autoregressive and moving average components, resulting in a FAC that exhibits a decay pattern that can be complex, often composed of a combination of exponential and damped sinusoidal terms. This makes them more flexible in capturing different time-dependence structures. Details about the ARMA model can be found in Wheeler and Ionides (2024).

Identifying the appropriate model is based on the analysis of the FAC and FACP. The FAC is particularly useful for recognizing patterns characteristic of MA models, given that its structure is finite and allows for the direct determination of the q-order. However, for AR and ARMA models, the FAC can present behavior that is more difficult to interpret, making FACP analysis an essential tool. The FACP helps identify the p-order of autoregressive models, as it exhibits a cutoff after p-lags in AR processes, while in MA processes, the FACP behaves more diffusely.

When the series presents recurring periodic variations, it is necessary to extend this model to include seasonality, resulting in the seasonal ARIMA (SARIMA) model. The model SARIMA (Morettin & Toloi, 2018) is defined by notation ARIMA (p, d, q)(P, D, Q)[s], where the parameters p, d, q, correspond to the orders of the autoregressive component (AR), the differentiation (I) and the moving average (MA) of the non-seasonal part, respectively and explained previously. P, D, Q represent the orders of these same components applied to the seasonal part, and [s] indicates the seasonality period.

The general equation of this model, using a seasonal period of 12, can be expressed as:

$$\phi(B)\Phi(B^{12})\Delta^d\Delta_{12}^D Z_{12} = \theta(B)\Theta(B^{12})a_t, \quad (1)$$

in which B is the lag operator; $\phi(B)$ and $\theta(B)$ are polynomials that represent the autoregressive and non-seasonal moving average parts; $\Phi(B^{12})$ and $\Theta(B^{12})$ are polynomials that represent the seasonal autoregressive and moving average parts; Δ^d and Δ_{12}^D are differentiations applied to remove trends and seasonal patterns (Morettin & Toloi, 2018).

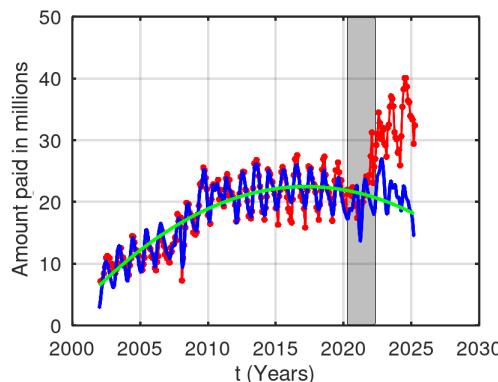
In summary, analysis of the autocorrelation (FAC) and partial autocorrelation (PFAC) functions provides relevant subsidies for modeling time series. Recognizing the characteristics of FAC in different models facilitates the selection of the most appropriate approach, allowing the capture of dynamic patterns in the data and improving forecast accuracy. Therefore, correct model specification is a crucial step in time series analysis, ensuring that statistical assumptions are met and that the results obtained are reliable. This work used the *arima* package from the *RStudio* software, whose documentation can be found at Wheeler et al. (2025).

Results and discussion

Adjustments 1 - pre-covid-19 pandemic

Considering the adjustments with the polynomial central tendency, for the pre-covid data from January 2002 to February 2020, Figure 2 was obtained.

Figure 2 - Curve Fitting via the Least Squares Method. In red we have the total data, in green the polynomial fit for pre-covid data, and in blue the polynomial fit with sines and cosines for pre-covid data.

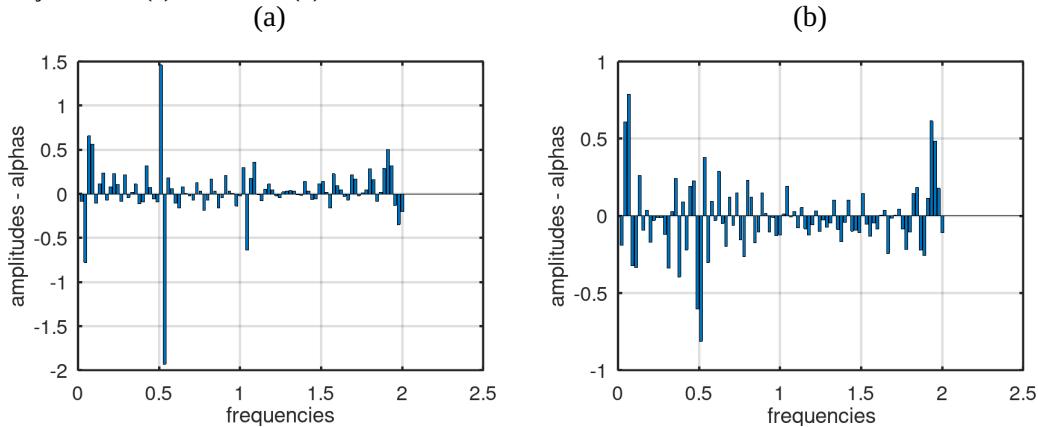


In Figure 2, the total data is in red and the pandemic period is in gray. The polynomial fit of degree 2 (green color), whose parabolic function obtained was

$$p_2(t) = (6.492774624692798 \times 10^6) + (1.751793352771674 \times 10^5)t - (4.794424843972461 \times 10^2)t^2. \quad (2)$$

Adding the adjustment of the oscillations (blue color), the amplitudes obtained from the sine and cosine functions with the respective frequencies are represented in Figure 3.

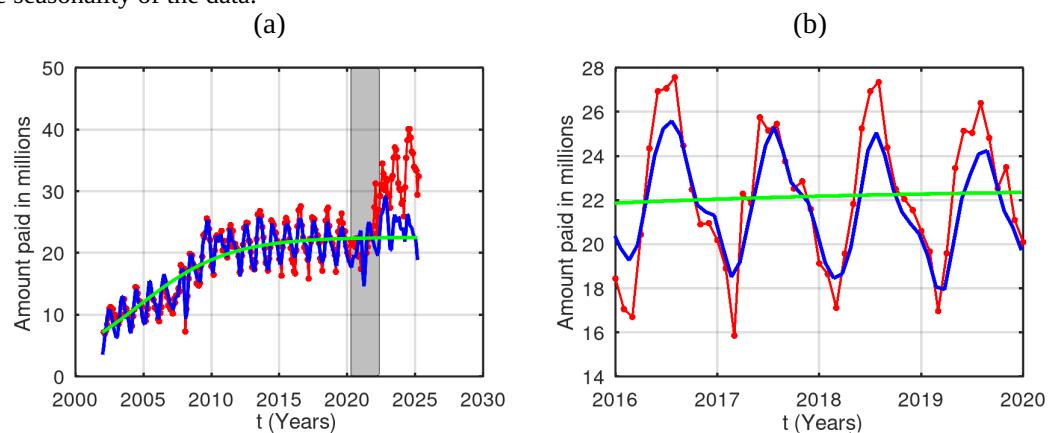
Figure 3 - Amplitudes obtained in relation to the frequencies of the trigonometric functions considered in the adjustment: (a) sines, and (b) cosines.



Furthermore, Figure 2 shows the projection of values up to March 2025. Note that the adjustment of the fluctuations followed the average trend (polynomial adjustment). The projection did not capture a stagnation trend in the data around 22 million. This stagnation is from pre-COVID data, as when observing the data after the start of the pandemic, the values change levels (see values in red).

Given this stagnation trend in values, a curve capable of capturing this trend is the logistic curve. Thus, considering a logistic fit to the mean trend of pre-COVID data, Figure 4 is obtained.

Figure 4 - Curve Fitting via the Least Squares Method: (a) In red is shown the total data, in green the logistic fit, and in blue the logistic fit with sines and cosines, and (b) a zoom from 2016 to 2020 where we see seasonality of the data.



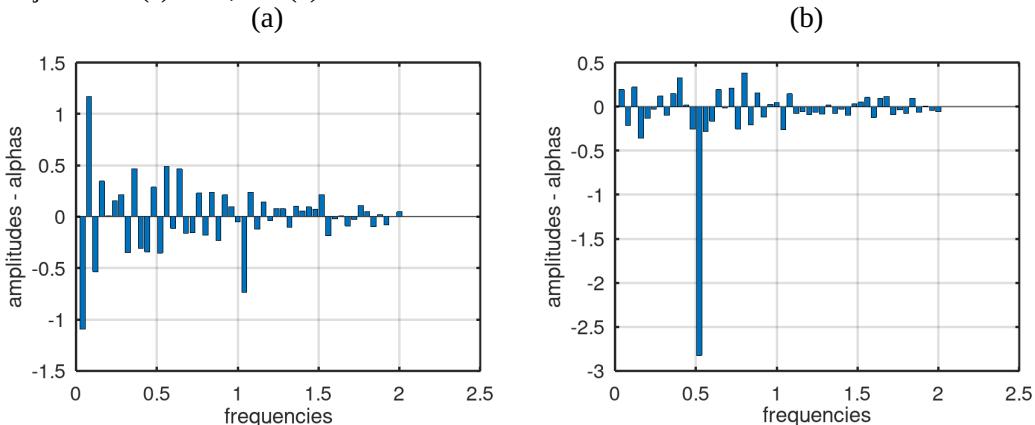
In Figure 4(a) the total data are in red and the pandemic period is in gray, and a zoom is shown for highlight the seasonality of the data (Figure 4(b)). The logistic fit (green color), whose function was obtained, was

$$p_{logist}(t) = \frac{Kp_0}{p_0 + (K - p_0)e^{-rt}}, \quad (3)$$

in which $p_0 = 7.17061881$, $K = 22.56392655$ and $r = 0.025008259126605942$.

Adding the adjustment of the oscillations (blue color), the amplitudes obtained from the sine and cosine functions with the respective frequencies are represented in Figure 5.

Figure 5 - Amplitudes obtained in relation to the frequencies of the trigonometric functions considered in the adjustment: (a) sines, and (b) cosines.



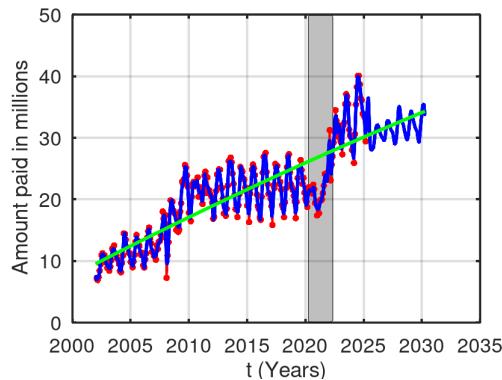
The logistic model, equation (3), contains the parameter $K = 22.56392655$ which means that the amounts paid in millions tend to stabilize asymptotically in K (green color - Figure 4) and the oscillations follow this value (blue color - Figure 4).

It can be seen that the peaks and troughs of the oscillations coincide with the peaks of hospitalizations due to *ICD-10* that usually occur in the transition between fall and winter. The lowest hospitalizations occur during the transition between spring and summer. Thus, the spacing between crests is approximately one year (Figure 4(b)).

Adjustments 2 - complete data

Considering the adjustments with the polynomial central tendency, for the data from January 2002 to March 2025, that is, including the COVID-19 pandemic period, the Figure 6 was obtained.

Figure 6 - Curve Fitting via the Least Squares Method. In red we have the total data, in green the polynomial fit for total data, and in blue the polynomial fit with sines and cosines for total data.



In Figure 6 the total data are in red and the pandemic period is in gray. The degree 2 polynomial fit is in green, whose parabolic function was

$$p_2(t) = (9.557698770580137 \times 10^6) + (8.093186200341591 \times 10^4)t - (2.347834234476136 \times 10^1)t^2. \quad (4)$$

Adding the adjustment of the oscillations (blue color), the amplitudes achieved by the sine and cosine functions, with their respective frequencies, are represented in Figure 7.

Furthermore, Figure 6 shows the projection of values up to March 2030. Note that the fluctuations adjusted in line with the average trend (polynomial adjustment). In this case, there is a slight stabilization trend, but it is not yet clear. However, it was evident that after the pandemic, the values exceeded 30 million.

A graph of interest in this case is one that contains only the oscillations, that is, the data subtracted from the average trend values. Thus, we have Figure 8.

Figure 7 - Amplitudes obtained in relation to the frequencies of the trigonometric functions considered in the adjustment: (a) sines, and (b) cosines.

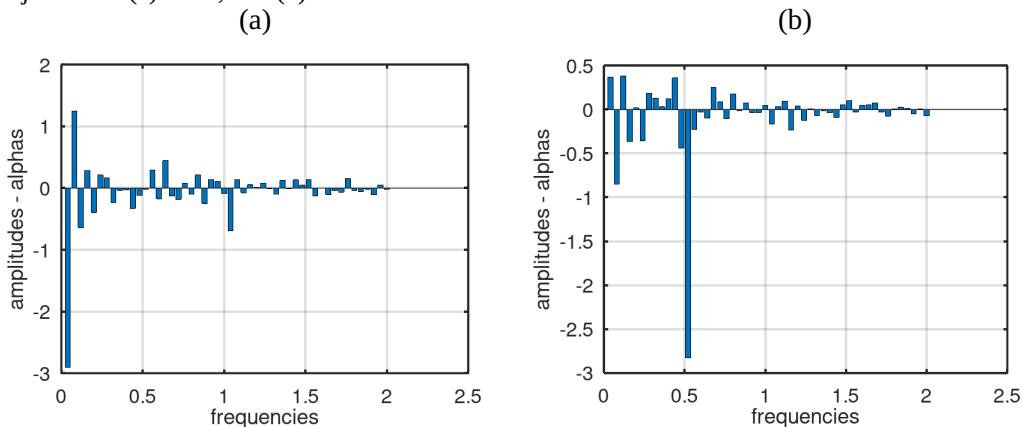


Figure 8 - Comparison between oscillations (black color) and adjustments (blue color). In red we have the total data, in blue the adjustment and the projection. The gray rectangle covers the period from mid-2009 to mid-2020.

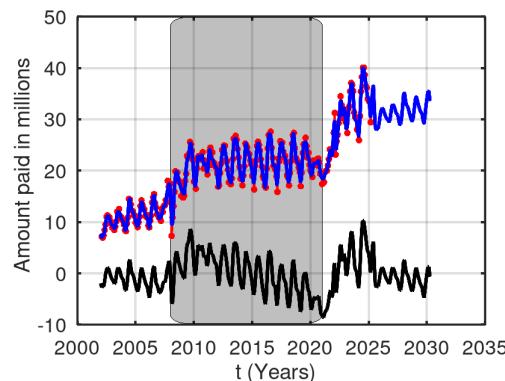
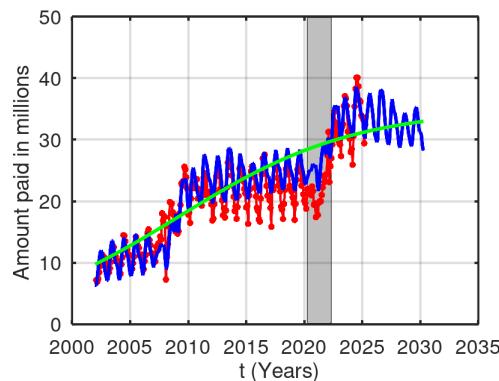


Figure 8 shows the oscillations in relation to the average trend in black. This graph highlights a period of two sharp declines in payment amounts, followed by two sharp increases, shifting the level of payment amounts. The gray rectangle covers this period, which began in mid-2009 and ended in mid-2020.

When performing a logistic adjustment for the mean trend on the complete data, the result is what appears in Figure 9.

Figure 9 - Curve Fitting via the Least Squares Method. Curve Fitting via the Least Squares Method. In red we have the total data, in green the logistic fit for total data, and in blue the logistic fit with sines and cosines for total data.



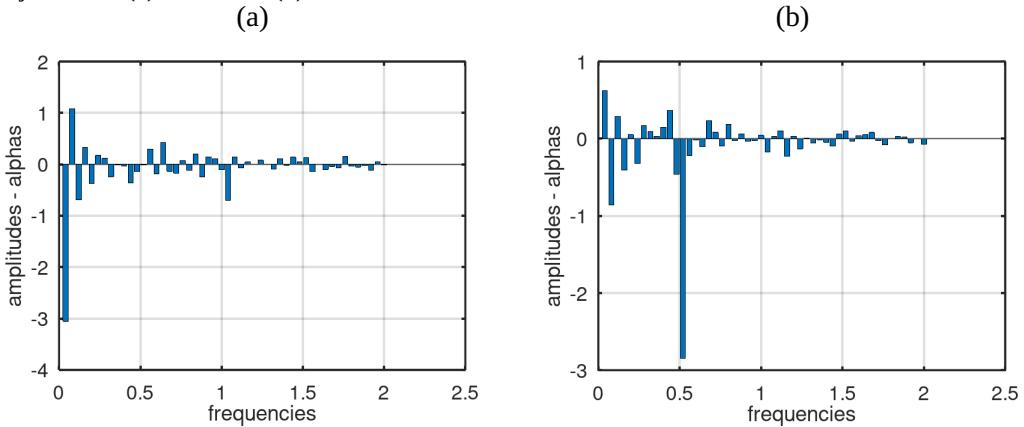
In Figure 9 the total data are in red and the pandemic period is in gray. The logistic adjustment is shown in green and its respective function was

$$p_{\text{logist}}(t) = \frac{Kp_0}{p_0 + (K - p_0)e^{-rt}}, \quad (5)$$

where $p_0 = 9.762687549611824$, $K = 35.01329728652672$ and $r = 1.099967023262771e - 02$. Note that the logistic model suggests a stagnation of the amounts paid around 35 million.

Including the adjustment of the oscillations (blue color), the amplitudes of the sine and cosine functions, with their respective frequencies, are represented in Figure 10.

Figure 10 - Amplitudes obtained in relation to the frequencies of the trigonometric functions considered in the adjustment: (a) sines, and (b) cosines.



The logistic model, equation (5), contains the parameter $K = 35.01329728652672$ which means that the amounts paid in millions tend to stabilize asymptotically at K (green color - Figure 9) and the oscillations follow this value (blue color - Figure 9).

The peaks and troughs of the oscillations coincide with the peaks of hospitalizations due to *ICD-10*, which occur during the transition between fall and winter. The lowest hospitalizations occur during the transition between spring and summer. Thus, the spacing between peaks encompasses an interval of approximately one year. This fact was observed in all fits with oscillations, as the amplitude of the cosine functions stood out more in all cases and is associated with the frequency 0.5.

Figure 11 - Comparison between oscillations (black color) and adjustments (blue color). In red we have the total data, in blue the adjustment and the projection. The gray rectangle covers the period from mid-2009 to mid-2020.

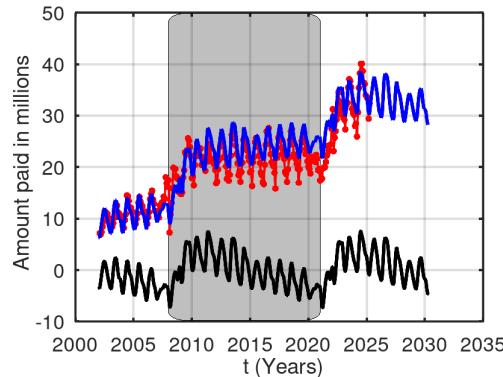


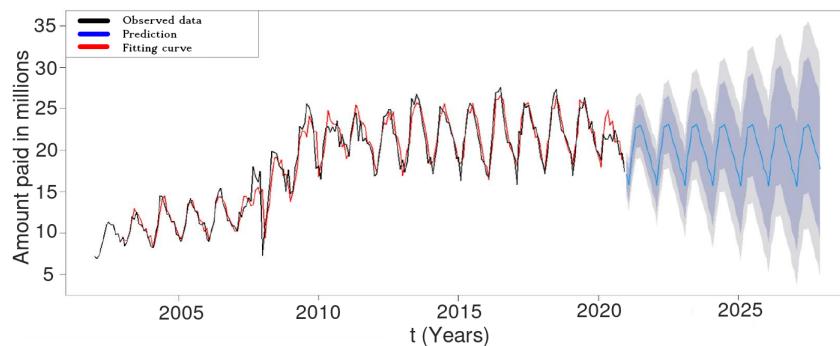
Figure 11 shows the fluctuations relative to the average trend in black. This graph further highlights the period highlighted in Figure 9, with the highs even more intense near the edges of the rectangle. The gray rectangle shows this period, which began in mid-2009 and ended in mid-2020.

When researching epidemics in Brazil, in 2009 there was an H1N1 pandemic that hit the country (CEE - Fiocruz, 2021) Fiocruz (2021). It should be noted that both the H1N1 and COVID-19 pandemics caused ICD-10-related payments to change. Furthermore, after pandemics, payments tend to fluctuate due to the seasonal characteristics of respiratory diseases, with an average tendency to stabilize. The two projections until 2030, Figures 9 and 11 (blue color) show such a tendency of stagnation in the amounts spent.

Adjustments 3 - ARIMA

In R language, the ARIMA package was used, with values $ARIMA(p, d, q)(P, D, Q)$ [12], which present a structure composed of non-seasonal and seasonal components. The non-seasonal components include an autoregressive term ($p=1$), which considers the dependence between the current values and the immediately preceding periods; a simple differentiation ($d=1$), used to transform the series into stationary by removing the long-term trend; and two moving average terms ($q=1$), which take into account the effects of residual errors from two previous periods. The seasonal component of the model incorporates a seasonal autoregressive term ($P=1$) to capture annual dependencies; a seasonal differencing term ($D=1$); and a seasonal moving average term ($Q=1$) to model fluctuations in seasonal errors. With this configuration, the model was able to capture both the overall trends of the series and recurring seasonal fluctuations, improving the data range for forecasts through 2025. See Figure 12.

Figure 12 - Forecasting via seasonal ARIMA model.



The metrics associated with the ARIMA model are found in Table 1 and were extracted from the ARIMA package. The results found show that the model performed satisfactorily, although there is still room for refinements and improvements. Interpreting the mean error (ME) of -0.0965, it may suggest that the model does not have a significant bias, that is, it has not systematically increased or decreased values. This indicates that the forecast tends to balance around the actual values over time. However, the root mean square error (RMSE) of 1.3272 indicates that there are changes in the residuals that may affect the model's accuracy.

Table 1 - Model application metrics ARIMA(1,1,2)(1,1,1)[12].

Metric	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Value	-0.0965	1.3272	0.9307	-1.0353	5.3442	0.5782	-0.021

The confidence intervals, Figure 12, represented by the shaded areas, gradually widen as the forecasts progress, indicating greater uncertainty in the results as we move away from the historical period. However, a key point worth highlighting is the difference between the 2020 forecast and the actual values. It is interesting to observe this specific period.

Regarding the future forecasts for the 80 months following the end of 2020, illustrated in Figure 12, there is a persistence of the observed seasonal behavior, with regular and consistent cycles over time. Note that this regularity in the forecasts is quite remarkable with respect to the forecast of average behavior (blue line).

The gray band represents a 95% confidence interval; that is, when performing infinite future stochastic trajectories, 95% of the trajectories will be entirely within the gray band. As uncertainty increases as we increase the final forecast time, the gray band becomes thicker to maintain the same 95% confidence interval.

Conclusions

This research enabled an analysis of data associated with hospitalization costs for respiratory diseases in the state of São Paulo over time. Statistical modeling allowed us to identify significant patterns, such as increased payments made in the early years of the historical series and the presence of seasonality and variations over time, which highlight the complexity involved in these costs.

Polynomial and logistic regressions helped identify the average growth trend in hospital costs, along with stabilization trends. The addition of sinusoidal and cosinoidal adjustments helped refine the model and capture seasonal fluctuations. These adjustments were made using pre-COVID data and complete data from 2002 to 2025.

Seasonal behaviors regarding payment amounts, associated with seasonal characteristics, were captured. In other words, there is an increase in hospitalizations for respiratory diseases during the transition between fall and winter.

Furthermore, in the graphs that include fluctuations around the average trend, two level transitions stand out in relation to the values paid: one linked to COVID-19 (mid-2020) and the other associated with H1N1 (in 2009). These transitions occur at values of $K = 22.65$ million, referring to the pre-COVID logistic model, and $K = 35$ million (logistic model that added the period of the COVID-19 pandemic).

Finally, studies that analyze cost data behavior can assist future research in resource allocation and strategic planning in public management. Furthermore, they can help in understanding the behavior of these costs and can support the formulation of public policies aimed at the sustainability of the system and the efficient management of health resources.

Author Contributions

R. O. Garcia and **G. P. Silveira** contributed to supervision, validation, formal analysis, visualization, and manuscript review and editing. **B. S. Silva** was responsible for conceptualization, data curation, investigation, methodology, and preparation of the original draft.

Conflicts of Interest

The authors declare no conflict of interest.

References

Andrade, C. L. L. d., Victora, C. G., Mendonça, M. H. P. d., & Giovanella, L. (2011). Financiamento, gasto e oferta de serviços de saúde em grandes centros urbanos do estado de São Paulo (Brasil). *Ciência & Saúde Coletiva*, 16, 1875–1885. <https://doi.org/10.1590/S1413-81232011000300022>

Araújo, R., Watanabe, S., Boiron, L., Pereira, A. C., & Asano, E. (2021). Impacto econômico da infecção por influenza no Brasil: Uma análise sob a perspectiva dos sistemas de saúde e da sociedade em 2019. *Jornal Brasileiro de Economia da Saúde*, 13(3), 300–309. <https://doi.org/10.21115/JBES.v13.n3.p300-9>

Fiocruz. (2021). *Combate à epidemia de H1N1: um histórico de sucesso*. <https://cee.fiocruz.br/?q=node/1314>

Marques, L. J. P., Pereira, A. C., & Raimundo, A. C. S. (2025). Custos e características das internações por condições sensíveis à atenção primária em menores de um ano em São Paulo, Brasil. *Ciência & Saúde Coletiva*, 30(1), 1–14. <https://doi.org/10.1590/1413-81232025301.15512023>

Ministério da Saúde. (2025). *DataSUS*. <https://datasus.saude.gov.br/>

Morettin, P. A., & Toloi, C. M. C. (2018). *Análise de séries temporais: modelos lineares univariados* (3a ed., pp. 591–612). Blucher.

Reis, C. S. d., Noronha, K., & Wajnman, S. (2016). Envelhecimento populacional e gastos com internação do SUS: Uma análise realizada para o Brasil entre 2000 e 2010. *Revista Brasileira de Estudos de População*, 33(3), 591–612. <https://doi.org/10.20947/S0102-30982016c0007>

Ruggiero, M. A. G., & Lopes, V. L. d. R. (2000). *Cálculo numérico: Aspectos teóricos e computacionais*. Pearson.

Silva, L. M. d., Alvarez, G. B., Christo, E. d. S., Pelén Sierra, G. A., & Garcia, V. d. S. (2021). Time series forecasting using ARIMA for modeling of glioma growth in response to radiotherapy. *Semina: Ciências Exatas e Tecnológicas*, 42(1), 3–12. <https://doi.org/10.5433/1679-0375.2021v42n1p3>

Wheeler, J., & Ionides, E. L. (2024). Likelihood based inference for ARMA models. *ArXiv*, 5, 1-25. <https://arxiv.org/pdf/2310.01198v5.pdf>

Wheeler, J., McAllister, N., Sylvertooth, D., Ionides, E., & Ripley, B. (2025). Arima2: Likelihood based inference for ARIMA modeling. <https://doi.org/10.32614/CRAN.package.arima2>

World Health Organization. (2025). *World report on social determinants of health equity*. <https://www.who.int/publications/i/item/9789240107588>