GC Revista Nacional de Gerenciamento de Cidades
National Journal of City Management

ISSN 2318-8472, v. 12, n. 86, 2024

Logistic Regression between Air Pollution and Mortality from Respiratory Diseases in Santa Gertrudes, SP (2014-2023)

Pollyane Vieira da Silva
Professor, Ph.D., UFPEL, Brazil
pollyane.silva@ufpel.edu.br

Magda Adelaide Lombardo
Professor, Ph.D., USP, Brazil
magdalombardo@yahoo.com.br

GC Revista Nacional de Gerenciamento de Cidades
National Journal of City Management

ISSN 2318-8472, v. 12, n. 86, 2024

ABSTRACT

Air pollution is a critical environmental problem that affects human health in many regions around the world. In cities with intense industrial activity, such as Santa Gertrudes, SP, the concentration of particulate matter, such as PM10, can have significant impacts on the respiratory health of the population. This study aims, using advanced statical modeling, to investigate the relationship between PM_{10} concentration and mortality from respiratory diseases. The main objective was to apply generalized linear models, specifically logistic regression, to analyze how the concentration of PM_{10} influences the mortality of patients hospitalized for respiratory diseases in the city of Santa Gertrudes during the period from 2014 to 2023. The methodology involved building and comparing several logistic models to assess the effect of PM_{10} and additional variables on mortality, such as relative humidity and wind speed. The quality of the models was assessed using the Akaike information criterion (AIC) and deviance. The results showed that $PM_{10}lag_4$ and "days of hospitalization" were the most statistically significant variables. The odds ratio associated with $PM_{10}lag_4$ indicated an increase of approximately 0.8% in the chance of death for each 1 µg/m³ increase in PM_{10} concentration. The variable "days of hospitalization" increased the chance of death by around 6.7% per additional day of hospitalization, highlighting the need for public policies aimed at reducing pollution and properly managing health conditions in the region.

KEYWORDS: Public Health. Statistical Modeling. Particulate matter.

1 INTRODUCTION

Air pollution represents one of the greatest environmental challenges on both a global and local scale. Prolonged exposure to an atmosphere contaminated by particulate matter can cause significant damage to human health, exacerbating pre-existing respiratory problems and triggering adverse conditions. This scenario is particularly worrying in heavily industrialized regions, where production activities contribute significantly to the emission of atmospheric pollutants.

The ceramics hub of Santa Gertrudes, SP, has emerged as an area of particular interest due to the concentration of long-standing industrial activities, whose emissions can directly impact local air quality and, consequently, the health of the resident population. The Santa Gertrudes ceramics hub is made up of eight cities, one of which is the city of Santa Gertrudes itself and the neighboring cities of Araras, Cordeirópolis, Ipeúna, Iracemápolis, Limeira, Piracicaba and Rio Claro.

It is important to note that this cluster plays a significant role in the national and state production of ceramic tiles. In 2022, the Brazilian industry produced 927 million square meters (m^2) of ceramic tiles, of which 736.4 million were sold on the domestic market and 130,4 million for export. This prominence on the world stage is the result of the development of the local industry's capacity, especially in the state of São Paulo. The Santa Gertrudes Pole was responsible for 70% of national production in 2020, as shown by data from the São Paulo Ceramic Tile Association (Aspacer) (ALESP, 2023).

In the floor and ceramic tile industries, the main source of air pollution is the firing oven, which produces emissions of particulate matter, nitrogen oxides, metals, chlorine and fluorides, and to a lesser extent sulfur oxide and total hydrocarbons (POLI, 2002) The size of the particles is directly associated with their potential to cause health problems, and the smaller they are, the greater the effects they cause.

Inhalable particles (PM_{10}) can be defined in a simplified way as those whose aerodynamic diameter is less than or equal to 10 µm. Depending on their size distribution in the

ISSN 2318-8472, v. 12, n. 86, 2024

0 to 10 µm range, they can be retained in the upper part of the respiratory system or penetrate more deeply, reaching the pulmonary alveoli (CETESB, 2024).

It should be noted that the concentration of pollutants is strongly related to meteorological conditions. Some of the parameters that favor high levels of pollution are a high percentage of calm, weak winds and thermal inversions at low altitude. This phenomenon is particularly common in São Paulo's winter, when the nights are cold and the temperature tends to rise rapidly during the day, causing a change in the natural cooling of the air (CETESB, 2024).

The aim of this investigation is to study the relationship between exposure to air pollutants and deaths from respiratory diseases in the municipalities of the Santa Gertrudes ceramics complex, taking into account some meteorological variables.

Therefore, it's important to use tools capable of monitoring the concentration of particulate matter (PM) in the regions under the influence of the center. Furthermore, the vast knowledge about the adverse effects of particulate matter on health makes it clear that, in addition to estimating particulate matter concentrations, the association between PM and deaths from respiratory diseases should also be analyzed.

Diseases of the respiratory system (ICD-10) were responsible for 148,773 deaths in 2020, higher than deaths from external causes (accidents and violence) and lower than diseases of the circulatory system, infectious and parasitic diseases and neoplasms, according to information from the DATASUS Mortality Information System (MOREIRA, 2023).

Among the existing respiratory diseases we have COVID-19, a disease that has caused a pandemic in recent years and has led thousands of people to die from complications of this respiratory syndrome caused by coronavirus. This research could reveal the impact of air pollution on human health and contribute to awareness-raising measures.

2 OBJECTIVES

The objective of this article was to apply statistical modeling by means of generalized linear models, using logistic regression to investigate the relationship between the concentration of particulate matter, PM_{10}, and the deaths of patients hospitalized for respiratory diseases in the city of Santa Gertrudes during the period from 2014 to 2023. In addition to assessing the potential impact on public health, this research aims to contribute to raising awareness and the adoption of mitigating measures by the ceramics industries, with a view to more sustainable environmental management and improving the quality of life in the region.

3 MATERIALS AND METHODS

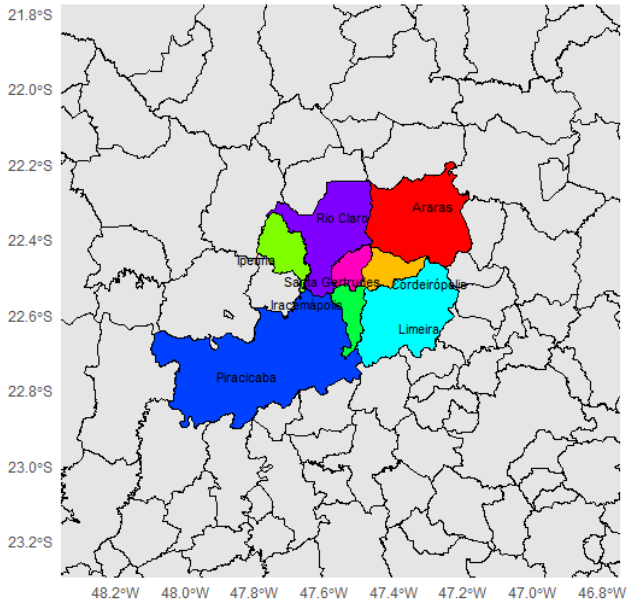
This research is classified as an ecological time-series study on the exposure of the population of Santa Gertrudes, SP, to particulate matter PM10 and the deaths resulting from respiratory diseases (with ICD codes J00 - J99 and B32.4).

Strategically located in the state of São Paulo, the Santa Gertrudes Ceramic Pole is not only a reference point in the Brazilian ceramics industry, but also a pillar of influence in the state's economic panorama. Within the ceramics cluster, the union of several cities forms a dynamic nucleus of production and innovation. Figure 1 dives into this dynamic, zooming in to

ISSN 2318-8472, v. 12, n. 86, 2024

show in detail the cities that make up this cluster.

Figure 1: Cities that make up the Santa Gertrudes ceramics cluster, SP



Source: Map created with R software using the ggplot2 package.

Data from Aspacer estimates that in 2006 the state's production of ceramic tiles amounted to 380 million square meters, of which 323 million square meters were produced in the Santa Gertrudes hub.

With a population of 23,611 people and a demographic density of approximately 240.22 inhabitants per square kilometer in a territorial area of 98.291km², Santa Gertrudes was, according to IBGE (2022), chosen as the main study region given the significant concentration of ceramic industries in this municipality in relation to the others in the hub.

For the scenario defined in the study, public data was selected from the government research bodies Companhia Ambiental do Estado de São Paulo (CETESB) and Departamento de Informática do Sistema Único de Saúde (DATASUS) for a period of 10 years, from 2014 to 2023.

Data on PM_{10} concentration and meteorological variables was extracted from the CETESB website, and were selected the study period, station (municipality) and parameters of interest. The system limits you to choosing three parameters at a time, and in this study we chose: PM_{10}, Relative Humidity (RH) and Wind Speed (WS).

The DATASUS system was searched for information regarding deaths and hospitalizations from respiratory diseases during the period in question. The historical table of hospitalizations and deaths was compiled exclusively with data from Santa Gertrudes. The first selection criterion used was the municipality of residence code (MUNIC RES), since our focus is on people who live in Santa Gertrudes, regardless of whether they were hospitalized in neighboring cities.

In order to optimize the DATASUS table for this study, only the relevant attributes were selected, with filtering on the main diagnosis using ICD 10 codes from J00 to J99 (diseases of the respiratory system) and B32.4 (coronavirus infection of unspecified location). Generalized linear

ISSN 2318-8472, v. 12, n. 86, 2024

models (GLMs) via logistic regression were used to work with the daily data on deaths from respiratory diseases. The best models were selected using the Akaike (AIC) and Deviance values. To check the quality of the fit, a simulated envelope graph was generated using the hnp package (MORAL; HINDE; DEMÉTRIO, 2017) in R (R CORE TEAM, 2024).

4 RESULTS AND DISCUSSION

Firstly, a logistic model (M_1) was fitted with the outcome variable "death" as the dependent variable and the variable "PM_{10}" together with a 7-day lag of PM_{10}, the lags, as the predictor variables. The estimated coefficients, standard errors and respective p-values for M_1 are shown in Table 1.

Table 1 - Estimated coefficients, standard errors and respective p-values for M_1

Variable	Estimate	Standard error	p-value
Intercept	-3.439	0.658	**$1.73e^{-7}$**
PM10	0.004	0.013	0.743
$PM10lag_1$	-0.001	0.016	0.975
$PM10lag_2$	-0.022	0.017	0.185
$PM10lag_3$	0.002	0.015	0.913
$PM10lag_4$	0.037	0.014	**0.008**
$PM10lag_5$	-0.006	0.015	0.719
$PM10lag_6$	-0.006	0.017	0.716
$PM10lag_7$	-0.003	0.013	0.841

Source: The authors (2024).

As shown in Table 1, only the variable "$PM_{10}lag_4$" was statistically significant in the first model. Therefore, in the second model (M_2), the variable "$PM_{10}lag_4$" was kept, and the meteorological variables "relative humidity (RH)" and "wind speed (WS)" were added to investigate the impact of weather conditions on deaths.

The estimated coefficients, standard errors and p-values of the M_2 model are detailed in Table 2.

Table 2 - Estimated coefficients, standard errors and respective p-values for M_2

Variable	Estimate	Standard error	p-value
Intercept	-6.569	1.714	**$1.27e^{-4}$**
$PM10lag_4$	0.025	0.007	**$2.26e^{-4}$**
RH	0.036	0.018	**0.047**
WS	-0.140	0.499	0.779

Source: The authors (2024).

The statistical significance of the "RH" variable in the M_2 model corroborates other existing studies and is in line with some CETESB recommendations. In winter, days with low air humidity and a high concentration of pollutants are common, which aggravates respiratory problems. In these conditions, dry mucous membranes can lead to nosebleeds, dry skin and eye irritation, favoring the onset of respiratory complications.

In the third model (M_3), the statistically significant variables from the previous models were maintained and a health-related variable, "days spent in hospital", was added. The

ISSN 2318-8472, v. 12, n. 86, 2024

estimated coefficients, standard errors and p-values for M_3 are shown in Table 3, providing a broader view of the impact of these variables on mortality from respiratory diseases.

Table 3 - Estimated coefficients, standard errors and respective p-values for M_3

Variable	Estimate	Standard error	p-value
Intercept	-6.247	1.265	**$7.85e^{-7}$**
$PM_{10}lag_4$	0.021	0.006	**$2.35e^{-4}$**
RH	0.021	0.015	0.151
Days of hospitalization	0.168	0.028	**$2.15e^{-9}$**

Source: The authors (2024).

With the M_3 configuration, the significant variables were $PM_{10}lag_4$ and Days of stay. Thus, the final model (FM) was defined as the one that includes only these significant variables. The expression of the final model is given by:

$$death = \beta_0 + \beta_1 \,\mathrm{PM10lag4} + \beta_2 \mathrm{Days\ of\ hospitalization} + \epsilon$$

This model suggests that an increase in the concentration of PM_{10}, with a lag of 4 days, and the number of days of hospitalization are directly associated with an increase in the probability of death from respiratory diseases.

The final model was tested using three different link functions: probit, logit and log-log complement. With all these link functions, the variables "$PM_{10}lag_4$" and "days of hospitalization" remained statistically significant. The quality of the fit was assessed using the Akaike information criterion (AIC) and the deviance, as shown in Table 4.

Table 4 - AIC and Deviance values of the Final Model (FM)

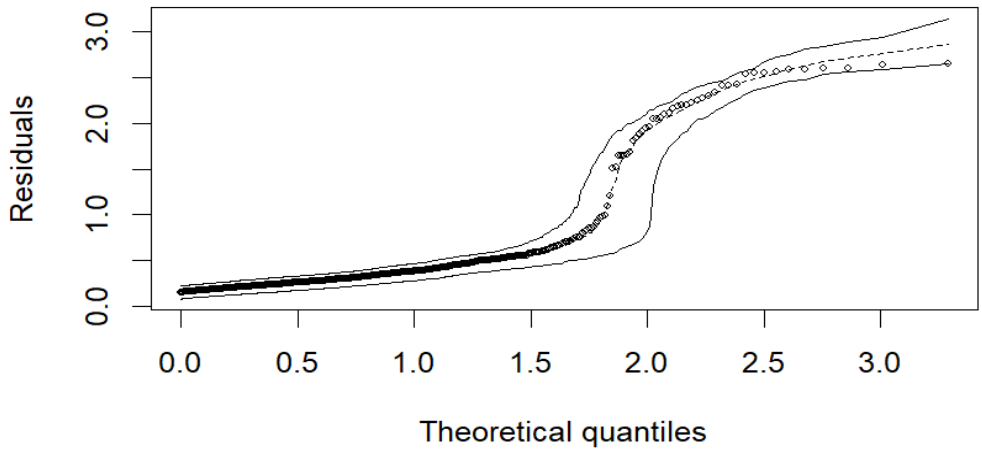
Link function	AIC	Deviance
Probit	**275,20**	**269,20**
Logit	276,32	270,32
Complemento log-log	276,75	270,75

Source: The authors (2024).

The model with the probit link function showed lower AIC and deviance compared to the other models, suggesting a better fit to the data. In addition, the analysis of residuals with the Half-Normal Plot (hnp) (Figure 2) reinforced the model's suitability, indicating a good distribution of residuals.

ISSN 2318-8472, v. 12, n. 86, 2024

Figure 2 - Final Model hnp graph with probit link function



Source: The authors (2024).

The final estimated model was:

$$death = -2.387 + 0.008\,\mathrm{PM10lag4} + 0.067\,\text{Days of hospitalization}$$

The estimated coefficients for the final model indicated that the odds ratio associated with the variable "$PM_{10}lag_4$" was 1.008 for each one-unit increase in PM_{10}. This means that for each increase of 1 µg/m³ in the concentration of PM_{10}, the chance of death increases by approximately 0.8%, keeping the other variables constant.

The variable "days of hospitalization" had an estimated coefficient of 0.067, indicating that for each additional day of hospitalization, the chance of death increases by around 6.7% according to the odds ratio of 1.067. The variable "days of hospitalization" had an estimated coefficient of 0.067, indicating that for each additional day of hospitalization, the chance of death increases by around 6.7%, according to the odds ratio of 1.067.

5 CONCLUSION

This study used statistical modeling through logistic regression to investigate the association between air pollution, meteorological conditions and mortality from respiratory diseases in the city of Santa Gertrudes, SP, from 2014 to 2023. The results indicated that the concentration of PM_{10} with a lag of 4 days ($PM_{10}lag_4$) was an important predictor of death, suggesting that for each increase of 1 µg/m³ in the concentration of PM10, the chance of death increased by approximately 0.8%.

The addition of meteorological variables such as relative humidity (RH) and wind speed (WS), in the second model (M_2), corroborated the literature, showing that adverse weather conditions, such as low humidity during the winter, can aggravate respiratory problems, which is in line with CETESB recommendations.

In the final model (FM), the variables "$PM_{10}lag_4$" and "days spent in hospital" remained significant. The odds ratio analysis revealed that for each additional day of hospitalization, the chance of death increased by 6.7%, reinforcing the relevance of prolonged hospital care for

ISSN 2318-8472, v. 12, n. 86, 2024

patients with respiratory diseases.

The comparison between different link functions (probit, logit and log-log complement) showed that all the variables remained significant, with the Akaike criterion (AIC) and deviance indicating that the model with the probit link function showed the best fit to the data.

This study highlights the significant influence of air pollution, especially PM_{10}, and meteorological and clinical factors on mortality from respiratory diseases. These findings reinforce the need for public policies that control air quality, particularly during critical periods of low humidity, and for more effective interventions to reduce the length of hospitalization in order to minimize the risks of fatal complications.

6 ACKNOWLEDGEMENTS

The authors would like to thank the University of São Paulo (USP) and the Department of Forestry Sciences for the support and infrastructure provided during the Post-Doctoral Program, which resulted in the collection and analysis of the data presented in this study.

7 BIBLIOGRAPHIC REFERENCE

ALESP. **Com 92% da produção estadual, Santa Gertrudes se destaca como "Capital da Cerâmica de Revestimento"**. São Paulo: ALESP, 2023. Disponível em: https://www.al.sp.gov.br/noticia/?id=456564. Acesso em: 16 set. 2024.

ASPACER. Associação Paulista das Cerâmicas de Revestimento. 2024. Disponível em: https://www.aspacer.com.br/. Acesso em: 16 set. 2024.

CETESB. Companhia Ambiental do Estado de São Paulo. 2024. Disponível em: https://cetesb.sp.gov.br/. Acesso em: 12 jul. 2023.

CETESB. Companhia Ambiental do Estado de São Paulo. Índices de Qualidade do Ar. 2024. Disponível em: https://cetesb.sp.gov.br/ar/poluentes/#:~:text=O%20material%20particulado%20pode%20tamb%C3%A9m,de%20rea%C3%A7%C3%B5es%20qu%C3%ADmicas%20no%20ar. Acesso em: 16 set. 2024.

DATASUS. Departamento de Informação e Informática do Sistema Único de Saúde. 2015. Disponível em: http://www.datasus.gov.br. Acesso em: 12 jul. 2023.

DATASUS. Departamento de Informação e Informática do Sistema Único de Saúde. Morbidade Hospitalar do SUS. Disponível em: https://datasus.saude.gov.br/acesso-a-informacao/morbidade-hospitalar-do-sus-sih-sus/. Acesso em: 19 out. 2023.

MORAL, R. A.; HINDE, J.; DEMÉTRIO, C. G. B. Half-normal plots and overdis- persed models in R: the hnp package. **Journal of Statistical Software**, California, v. 81, n. 10, p. 1-23, 2017.

MOREIRA, M. **Doenças respiratórias são a quarta causa de morte evitável no país.** Disponível em: https://medicinasa.com.br/doencas-respiratorias-morte/ . Acesso em: 16 set. 2024.

POLI, M. C. **Emissões de flúor nas indústrias de pisos cerâmicos dos municípios de Cordeirópolis e Santa Gertrudes situação e recomendações.** Monografia de Especialização. UNICAMP. 2002. Disponível em: https://cetesb.sp.gov.br/escolasuperior/wp-content/uploads/sites/30/2016/06/Maria_Cristina_Poli.pdf. Acesso em: 16 set. 2024.

R CORE TEAM. **R: a language and environment for statistical computing.** Vienna: R Foundation for Statistical Computing, 2024. Disponível em: https://www.R-project.org/. Acesso em: 16 set. 2024.

GC Revista Nacional de Gerenciamento de Cidades
National Journal of City Management

ISSN 2318-8472, v. 12, n. 86, 2024

DECLARAÇÕES

Contribuição de Cada Autor
Concepção e Design do Estudo: Pollyane Vieira da Silva, sob a supervisão de Magda Adelaide Lombardo.
Curadoria de Dados: Pollyane Vieira da Silva organizou e verificou os dados, garantindo sua qualidade.
Análise Formal: Pollyane Vieira da Silva foi responsável por realizar as análises estatísticas aplicadas.
Aquisição de Financiamento: Não aplicável/não houve aquisição de financiamento externo para este estudo (ou incluir o responsável, se houver).
Investigação: Pollyane Vieira da Silva conduziu a investigação e coleta de dados necessários.
Metodologia: Pollyane Vieira da Silva desenvolveu e ajustou as metodologias utilizadas no estudo.
Redação - Rascunho Inicial: Pollyane Vieira da Silva redigiu a primeira versão do manuscrito.
Redação - Revisão Crítica: Magda Adelaide Lombardo revisou o texto, contribuindo para a clareza e coerência do conteúdo.
Revisão e Edição Final: Magda Adelaide Lombardo realizou a revisão e edição final do manuscrito, garantindo a conformidade com as normas da revista.
Supervisão: Magda Adelaide Lombardo supervisionou a pesquisa, coordenando o trabalho e assegurando a qualidade geral do estudo.

Declaração de Conflitos de Interesse

Nós, Pollyane Vieira da Silva e Magda Adelaide Lombardo**, declaramos que o manuscrito intitulado "Regressão Logística entre Poluição do Ar e Mortalidade por Doenças Respiratórias em Santa Gertrudes, SP (2014-2023)"**

1. **Vínculos Financeiros**: Não possui vínculos financeiros que possam influenciar os resultados ou interpretação do trabalho. (Detalhe aqui, se aplicável: “Nenhuma instituição ou entidade financiadora esteve envolvida no desenvolvimento deste estudo”).
2. **Relações Profissionais**: Não possui relações profissionais que possam impactar na análise, interpretação ou apresentação dos resultados. (Detalhe aqui, se aplicável: "Nenhuma relação profissional relevante ao conteúdo deste manuscrito foi estabelecida”).
3. **Conflitos Pessoais**: Não possui conflitos de interesse pessoais relacionados ao conteúdo do manuscrito. (Detalhe aqui, se aplicável: “Nenhum conflito pessoal relacionado ao conteúdo foi identificado”).