

Diego Mateus Müller

**Fatores associados ao óbito hospitalar por
COVID-19 no Rio Grande do Sul:
Uma aplicação de Regressão Logística**

Rio Grande, Rio Grande do Sul, Brasil

Janeiro, 2024

Diego Mateus Müller

**Fatores associados ao óbito hospitalar por COVID-19 no
Rio Grande do Sul:
Uma aplicação de Regressão Logística**

Trabalho de Conclusão de Curso, Matemática Aplicada Bacharelado, submetido ao Instituto de Matemática, Estatística e Física da Universidade Federal do Rio Grande.

Universidade Federal do Rio Grande - FURG
Instituto de Matemática, Estatística e Física - IMEF
Curso de Matemática Aplicada Bacharelado

Orientador: Prof^a Doutora Raquel da Fontoura Nicolette

Rio Grande, Rio Grande do Sul, Brasil
Janeiro, 2024

Diego Mateus Müller

Fatores associados ao óbito hospitalar por COVID-19 no Rio Grande do Sul Uma aplicação de Regressão Logística

Trabalho de Conclusão de Curso, Matemática Aplicada Bacharelado, submetido ao Instituto de Matemática, Estatística e Física da Universidade Federal do Rio Grande.

Trabalho Aprovado em 22 de janeiro de 2024.

Documento assinado digitalmente
 RAQUEL DA FONTOURA NICOLETTE
Data: 10/02/2025 09:43:03-0300
Verifique em <https://validar.iti.gov.br>

**Prof^a Doutora Raquel da Fontoura
Nicolette**
(Orientadora - FURG)

Documento assinado digitalmente
 ADILSON DA SILVA NUNES
Data: 29/01/2024 00:20:14-0300
Verifique em <https://validar.iti.gov.br>

Prof. Doutor Adilson da Silva Nunes
(IMEF - FURG)

Documento assinado digitalmente
 JULIANA DA SILVA RICARDO NUNES
Data: 26/01/2024 19:20:27-0300
Verifique em <https://validar.iti.gov.br>

**Prof.^a Doutora Juliana da Silva
Ricardo Nunes**
(IMEF - FURG)

Rio Grande, Rio Grande do Sul, Brasil
Janeiro, 2024

Este trabalho é dedicado aos meus pais, irmã, que não mediram esforços para que chegasse até esta etapa da minha vida. Em especial ao meu pai, que hoje a saudade só não é maior que meu amor!

Agradecimentos

Agradeço a minha família, que mesmo longe, sempre me deu segurança, me apoiou e me ajudou em todos os momentos, fazendo com que meu sonho se tornasse realidade.

A minha mãe Marcia dos Passos Müller, que mesmo querendo me manter seguro perto dela, deixou aquele menino de 17 anos ir em busca dos seus sonhos, apoiando e motivando para que eu pudesse alcançar essa conquista.

Ao meu pai João Batista Müller, que deu todo suporte necessário para que estivesse bem e conseguisse me manter seguro. Ele que lutou até o último momento, e que hoje sei que torce por mim e vai olhar sempre por nossa família.

A minha irmã Aline Graziela Müller, que foi minha inspiração durante toda minha vida, e foi quem me inspirou a seguir meus sonhos assim como ela, e de toda forma me deu o suporte necessário para chegar até esta etapa.

A minha orientadora Doutora Raquel da Fontoura Nicolette, por ter aceito me auxiliar nesta última etapa, por todo conhecimento passado, pela compreensão e atenção dedicada nesta etapa do meu processo de formação.

A todos os professores que compartilharam seus conhecimentos ao longo desta jornada de graduação.

Resumo

A regressão logística é uma das principais ferramentas de estudos nas áreas de saúde, sendo muito utilizada em estudos epidemiológicos. Desta forma, neste estudo foi utilizado o método de regressão logística para avaliar fatores associados ao óbito por Covid-19 no estado do Rio Grande do Sul (RS), Brasil.

Este método permite prever a probabilidade de ocorrência de um evento através de um modelo que relaciona uma variável dependente binária e uma ou mais variáveis independentes explicativas, com intuito de entender a influência destas variáveis explicativas na variável dependente estudada.

Foi realizado um estudo transversal com dados de 94.192 pacientes hospitalizados no RS, com a finalidade de analisar as incidências de internações hospitalares decorrentes da Síndrome Respiratória Aguda Grave (SRAG) atribuídas à COVID-19. Durante este estudo foram realizadas análises dos perfis de cada paciente hospitalizados no RS, a fim de chegar em um modelo com resultados das razões de chances destes pacientes virem a óbito.

O modelo final ajustado apresentou acurácia de 79,5%. Observou-se que os pacientes os quais tiveram internação em Unidade de Terapia Intensiva (UTI) apresentaram uma chance 18 vezes maior de risco de morte do que pacientes que não necessitaram de UTI. Entre outros resultados, observou-se ainda que a idade dos pacientes não foi relevante para o risco de morte, embora foi observado que conforme a idade aumenta, os riscos dos pacientes associados aos fatores de risco e sintomas também aumenta.

Palavras-chaves: Covid-19, Regressão Logística, SRAG.

Abstract

Logistic regression is one of the main study tools in the health areas, being widely used in epidemiological studies. Therefore, in this study we used the logistic regression method to evaluate factors associated with death from Covid-19 in state of Rio Grande do Sul (RS), Brazil.

Logistic regression is a technique that allows predicting the probability of occurrence and an event through a model that relates a binary dependent variable and/or more explanatory independent variables, with the aim of understanding the influence of these explanatory variables in the dependent variable studied.

A cross-sectional study was conducted with data from 94.192 patients hospitalized in RS, with the purpose of analyzing the incidence of hospital admissions resulting from of severe acute respiratory syndrome (SARS) attributed to COVID-19. During this study, analyzes of the profiles of each patient hospitalized in RS were conducted, to arrive at a model with results of the odds ratios of these patients coming to death.

The final combined model showed an accuracy of 79.5%. It was seen that patients' entities who were admitted to the Intensive Care Unit (ICU) presented an 18 times greater chance of death than patients who did not need ICU. Among other results, it was also seen that the age of the patients was not relevant for the risk of death, although it was seen that as age increases, the risks of patients associated with risk factors and symptoms also increases.

Key-words: Covid-19, Logistic Regression.

Lista de ilustrações

Figura 2 – Curva ROC de chances de Cura ou Óbito	29
--	----

Lista de tabelas

Tabela 1 – Idades dos pacientes	24
Tabela 2 – Escolaridade	25
Tabela 3 – Fator de Risco	25
Tabela 4 – Modelo 1 - Chances de óbito ou cura	26
Tabela 5 – Modelo 2 - Chances de óbito ou cura	27
Tabela 6 – Modelo 3 - Chances de óbito ou cura	27
Tabela 7 – Critério de Akaike - AIC	27
Tabela 8 – Tabela de Classificação	28

Lista de abreviaturas e siglas

AIC Critério e Informação de Akaike

MS Ministério da Saúde

O.R. Odds Ratio

OMS Organização Mundial da Saúde

ROC Receiver Operating Characteristic

SARS-CoV-2 Coronavírus 2 da Síndrome Respiratória Aguda

SIVEP-Gripe Sistema de Informação de Vigilância Epidemiológica da Influenza

SRAG Síndrome Respiratória Aguda Grave

SUS Sistema Único de Saúde

Sumário

1	INTRODUÇÃO	11
1.1	Objetivos	11
1.1.1	Principal	11
1.1.2	Secundários	12
1.2	Estado da arte : Alguns estudos sobre regressão logística e a Covid	12
1.3	Estrutura da Monografia	13
2	FUNDAMENTAÇÃO MATEMÁTICA	14
2.1	Regressão Logística	14
2.2	Função Logística	14
2.3	Função de Verossimilhança	16
2.4	Estimação de parâmetros	17
2.5	Curva ROC(Receiver Operating Characteristic)	18
2.6	Critério de informação de Akaike(AIC)	18
3	METODOLOGIA	19
3.1	Tipo de pesquisa	19
3.1.1	Dados	19
3.1.2	Período de análise	19
3.1.3	População e amostra	20
3.1.4	Crterios de Inclusão	20
3.2	Análise de Dados	20
3.2.1	Filtragem	20
3.2.2	Análise Exporatória de Dados	20
3.2.3	Regressão Logística	21
3.3	Suporte Computacional	22
3.4	Ética da pesquisa	22
4	RESULTADOS E DISCUSSÃO	23
4.1	Filtragem	23
4.2	Análise Exploratória de Dados	24
4.3	Regressão Logística	25
4.3.1	Avaliação do modelo	28
5	CONSIDERAÇÕES FINAIS	30
	Referências	31

1 Introdução

Nos últimos quatro anos a palavra pandemia foi muito presente no mundo todo, devido ao novo vírus que surgiu no final do ano de 2019, SARS-CoV-2, mais conhecido como Covid-19. Porém, as epidemias são conhecidas desde o século XIV, quando houve a maior pandemia já registrada, chamada de Peste Negra(Peste Bubônica), onde estima-se que houve a morte de um terço da população da Europa [AMARAL, 2021](#) A primeira pandemia do século XXI foi a Influenza A/H1N1, a popular gripe suína, que surgiu em 2009 e foi responsável por aproximadamente 19 mil morte ao redor do mundo.

A Covid-19 se deu origem na cidade de Wuhan, na China, em dezembro de 2019. Segundo a Organização Mundial da Saúde(OMS) o nome referencial para o vírus é coronavírus 2 da síndrome respiratória aguda(SARS-CoV-2), que refere-se a um agente patogênico que causa problemas no sistema respiratório. Em 30 de janeiro de 2020, a Covid-19 foi declarada pela OMS como uma emergência de saúde pública de importância internacional, e dois meses depois foi declarada como uma pandemia.

A propagação da doença é por meio de gotículas ou contato direto com paciente infectado, e leva em torno de 5 a 6 dias para começar a sentir os primeiros sintomas, devido ao período de incubação do vírus [LIMA, 2020](#) As manifestações clínicas podem variar desde um simples resfriado a uma pneumonia grave, tal que o quadro clínico inicial é classificado como uma síndrome gripal(SG). Os principais sintomas apresentados pelos pacientes foram febre, tosse, dispneia(falta de ar), mialgia(dores musculares) e cefaléia(dores de cabeça), entre outros.

O presente estudo foi realizado com o objetivo de identificar quais os fatores de riscos apresentados pelos pacientes hospitalizados no estado do Rio Grande do Sul - RS - Brasil, levam a um maior risco de morte. Desta forma esse trabalho consiste em uma análise de regressão logística com dados provenientes do Sistema Único de Saúde - SUS .

1.1 Objetivos

1.1.1 Principal

- Analisar os fatores associados ao óbito em indivíduos internados por COVID-19 no estado do Rio Grande do Sul, Brasil.

1.1.2 Secundários

- Realizar uma análise exploratória de dados para verificar o perfil dos hospitalizados.
- Calcular a "Razão de chance" de diferentes perfis de hospitalizados.

1.2 Estado da arte : Alguns estudos sobre regressão logística e a Covid

No Brasil diversos trabalhos foram realizados fazendo uso da regressão logística na análise dos dados da COVID-19. Tais trabalhos abordam diferentes realidades dos aspectos pandêmicos.

PASTE et al. tinham como objetivo de analisar os fatores de risco associados a chance de óbito em indivíduos internados com SARS CoV 19 (COVID 19) em um hospital privado de Salvador/BA. Analisaram 400 pacientes de Fevereiro de 2020 a Janeiro de 2021. Foi realizada uma análise de regressão logística para estudar as variáveis clínicas associadas com um maior risco de morte. Um valor de $p < 0,005$ foi considerado estatisticamente significante. Concluíram que as variáveis associadas com um pior prognóstico foram a presença de Asma Brônquica e Hipertensão Arterial Sistêmica, além de lactato elevado e uso de Ventilação Mecânica Invasiva e hemodiálise.

MACIEL et al. analisaram os fatores associados ao óbito em indivíduos internados por COVID-19 em hospitais do Espírito Santo, Brasil. O estudo é do início da pandemia no estado ES até 14 de maio de 2020, 200 indivíduos receberam alta e 220 foram a óbito. Foram usados modelos de regressão logística foram empregados para estimar razões de chance (odds ratio: OR) brutas e ajustadas onde observou-se maior mortalidade em idosos, com comorbidades e usuários de hospitais públicos.

ALBUQUERQUE buscou entender como a doença pode reagir em cada pessoa de acordo com o seu sexo, idade, peso e outras características. Os resultados obtidos nessa pesquisa mostram que a equação alcançada através da regressão logística é condizente com a realidade do Brasil e com outras pesquisas publicadas, em que estas mostram as estatísticas com a COVID-19 para cura ou óbito e suas correlações com as características dos pacientes.

No Rio Grande do Sul(RS), pensando em analisar o efeito dos fatores de risco para testes positivos para SARS-CoV-2 em profissionais da saúde e não profissionais de saúde do estado, **CAMARGO** analisou dados do ESUS Notifica, contendo 5.620.062 indivíduos moradores do RS que realizaram algum tipo de teste para detectar infecção por SARS-CoV-2 entre 15 de fevereiro de 2020 e 05 de setembro de 2021. A análise utilizada foi a regressão logística com as estimativas da razão de chances e intervalos de 95% de

confiança. A conclusão do estudo é que a ocorrência de infecção pelo novo coronavírus em trabalhadores da área da saúde, no RS, mostrou-se menor do que na população em geral entre outras conclusões.

1.3 Estrutura da Monografia

Para melhor compreensão do trabalho, a monografia está dividida como segue:

Capítulo 1 - Introdução: apresenta a motivação, justificativa e objetivos deste trabalho, assim como uma leve discussão sobre a pandemia Covid-19.

Capítulo 2- Fundamentação matemática: discute os conceitos de estatística básica incluindo análise exploratória de dados, conceitos de regressão logística, função verossimilhança e estimação de parâmetros.

Capítulo 3- Metodologia: aborda o tipo de pesquisa, análise de dados, suporte computacional e a ética de pesquisa.

Capítulo 4- Resultados e discussões: neste capítulo temos os resultados das análises de dados e do método de regressão logística aplicado, bem como discussão dos resultados.

Capítulo 5- Considerações finais.

2 Fundamentação Matemática

2.1 Regressão Logística

A Regressão Logística é uma técnica estatística com objetivo de determinar a relação de várias explicativas e a probabilidade de ocorrer ou não um evento em estudo utilizada em diferentes áreas das ciências, uma delas a da saúde. Aqui dizemos que nossa variável explicativa é a nossa variável independente, ou seja, usada para prever ou explicar o resultado da variável dependente, que é o nosso evento em estudo, tal como o nosso resultado a ser previsto com base na nossa variável independente [FÁVERO; BELFIORE, 2017](#)

Desta forma, podemos criar uma variável binária que representa a ocorrência ou não do evento, que é dada por 1 , sendo sucesso, e por 0 , sendo o fracasso. Logo, pode-se dizer que a regressão logística é uma técnica que instiga a relação entre as variáveis explicativas e as variáveis categóricas binárias, destinando-se a medir a probabilidade de ocorrer um evento e identificar as características dos elementos pertencentes a cada grupo da variável categórica.

2.2 Função Logística

O modelo de função logística segue a seguinte equação:

$$f(Z) = \frac{1}{1 + e^{-Z}} \quad (2.1)$$

em que Z é definido como:

$$Z = \ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (2.2)$$

tal que p indica a probabilidade de ocorrência do evento, X indica o vetor das variáveis explicativas, α e β indicam os parâmetros do modelo e n é o tamanho da amostra. Nesta equação Z é chamada de função *Logit*, ou seja, o logaritmo da razão entre a probabilidade de ocorrência ou não do evento em estudo.

Desta forma, podemos substituir Z em $f(Z)$, assim temos:

$$f(Z) = \frac{1}{1 + e^{-(\alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}} \quad (2.3)$$

A função pode ser representado pela figura abaixo, que apresenta um comportamento da função sigmóide, que atribui um grande grau de generalidade a regressão

logística. A função apresenta valores entre 0 e 1, tal que quando a $Z \rightarrow +\infty$, então $f(Z) \rightarrow 1$ e quando $Z \rightarrow -\infty$, então $f(Z) \rightarrow 0$.

Os conceitos de chance (odds), definido na equação 2.4, e probabilidade são conceitos distintos, embora muitas vezes possam ser confundidos ou considerados sinônimos. Enquanto a probabilidade expressa a medida de quão provável é a ocorrência de um evento em relação ao total de eventos possíveis, a chance se refere à razão entre a probabilidade de um evento acontecer e a probabilidade de ele não ocorra.

$$Chance(odds) = \frac{p}{1-p} \quad (2.4)$$

Já a razão de chance, também conhecida como *odds ratio* (O.R.), é uma medida que compara as chances de um evento acontecer em dois grupos distintos. Essa medida é definida como a relação entre a probabilidade de um evento ocorrer em um grupo e a probabilidade de o mesmo evento acontecer em outro grupo. Sejam definidos dois grupos, A e B , e suas respectivas probabilidades p e q dadas pela equação 2.3, para o evento em estudo. O cálculo da razão de chance segue a equação 2.5. O conceito de chance (odds) está relacionado à determinação da probabilidade p_i de ocorrência de um dado evento, baseada nos parâmetros estimados através das variáveis explicativas sem representar a variável dependente Z .

$$O.R. = \frac{\frac{p}{1-p}}{\frac{q}{1-q}} = \frac{p(1-q)}{q(1-p)} \quad (2.5)$$

Portanto, a chance considera a relação entre a probabilidade do evento em questão e sua não ocorrência, oferecendo uma perspectiva diferente e complementar à probabilidade ao avaliar os mesmos eventos. Assim se $O.R. = 1$ indica que ambos eventos tem a mesma chance de ocorrência; $O.R. < 1$ indica que a chance de ocorrência do evento A é menor que a do evento B e se $O.R. > 1$ indica que a chance do evento A é maior que a do B .

Para exemplificar tem-se a seguinte situação fictícia: seja evento A a probabilidade de sobreviver de um paciente com covid na faixa etária de 30 a 39 anos é de 80%, assim a chance de sobreviver é de 8 para 1, enquanto o evento B é a probabilidade de cura de um paciente na faixa etária de 70+ é de 50% sendo a chance de cura de 1 para 1. Assim:

$$O.R. = \frac{0,9(0,5)}{0,5(0,1)} = 9 \quad (2.6)$$

o que significa que a probabilidade de um paciente na faixa etária de 30 a 39 anos é 9 vezes maior de não vir a óbito que um paciente na faixa etária de 70+.

Retomando as estimativas dos parâmetros, os coeficientes α e β são parâmetros desconhecidos, logo é necessário estimá-los para determinar a probabilidade de ocorrência do evento. Para encontrarmos estes valores é utilizado o método de máxima verossimilhança, tal que as variáveis explicativas permitem estabelecer a importância de cada variável para a ocorrência do evento de interesse, ou seja, calcular a probabilidade de ocorrência desse evento.

2.3 Função de Verossimilhança

Dada uma amostra $\{x, y\} = \{(x_1, y_1), \dots, (x_n, y_n)\}$, a função de verossimilhança a ser maximizada é:

$$P(\alpha, \beta | x, y) = \prod_{i=1}^n g(x_i)^{y_i} (1 - g(x_i))^{1-y_i} \quad (2.7)$$

com

$$g(x_i) = \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}} \quad (2.8)$$

Assim, aplicando o logaritmo neperiano na função, temos;

$$L(\alpha, \beta | x, y) = \sum_{i=1}^n \ln(g(x_i)^{y_i}) + \ln((1 - g(x_i))^{1-y_i}) = \sum_{i=1}^n y_i \ln(g(x_i)) + (1 - y_i) \ln(1 - g(x_i)). \quad (2.9)$$

Substituindo $g(x_i)$, temos:

$$L(\alpha, \beta | x, y) = \sum_{i=1}^n y_i \ln\left(\frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}}\right) + (1 - y_i) \ln\left(1 - \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}}\right) = \quad (2.10)$$

$$= \sum_{i=1}^n y_i \ln\left(\frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}}\right) + (1 - y_i) \ln\left(\frac{1 + e^{\alpha + \beta x_i} - e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}}\right) = \quad (2.11)$$

$$= \sum_{i=1}^n y_i \ln\left(\frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}}\right) + (1 - y_i) \ln\left(\frac{1}{1 + e^{\alpha + \beta x_i}}\right) = \quad (2.12)$$

$$= \sum_{i=1}^n y_i \ln(e^{\alpha + \beta x_i}) - y_i \ln(1 + e^{\alpha + \beta x_i}) + \ln(1) - \ln(1 + e^{\alpha + \beta x_i}) - y_i \ln(1) + y_i \ln(1 + e^{\alpha + \beta x_i}) = \quad (2.13)$$

$$= \sum_{i=1}^n y_i \ln(e^{\alpha + \beta x_i}) - \ln(1 + e^{\alpha + \beta x_i}) = \quad (2.14)$$

$$= \sum_{i=1}^n y_i (\alpha + \beta x_i) \ln(e) - \ln(1 + e^{\alpha + \beta x_i}) = \quad (2.15)$$

$$= \sum_{i=1}^n y_i (\alpha + \beta x_i) - \ln(1 + e^{\alpha + \beta x_i}) \quad (2.16)$$

Assim, temos:

$$L(\alpha, \beta | x, y) = \sum_{i=1}^n y_i (\alpha + \beta x_i) - \ln(1 + e^{\alpha + \beta x_i}) \quad (2.17)$$

2.4 Estimação de parâmetros

A modelagem da regressão logística estabelece uma relação entre as variáveis explicativas e a probabilidade de ocorrer ou não o fenômeno estudado, o que permite no nosso contexto criar uma variável ditocômica para estimar a probabilidade de um paciente sobreviver (0) ou (1) ir a óbito.

No modelo proposto a variável dependente indica a situação de sobreviver (0 ou 1), que geralmente é binária (nominal ou ordinal) e as independentes expressam fatores que acredita-se influenciarem na morbidade (comorbidades por exemplo), às quais podem ser categóricas ou contínuas, no caso das categóricas devem ser ditocomizadas.

Definido Y_i a variável ditocômica que resulta em (0) o paciente pertencer a categoria sobreviver e (1) quando pertencer a categoria óbito. Seja X um vetor, no qual o primeiro elemento é constante 1 e os demais variáveis independentes do modelo $X = (1, X_1, X_2, \dots, X_n)$ e $\beta' = (\beta_0, \beta_1, \beta_2, \dots, \beta_n)$ é o vetor de parâmetros associados a cada variável. Os coeficientes $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ são estimados a partir do conjunto dados, pelo método da máxima verossimilhança

A estimação dos parâmetros pelo método de máxima verossimilhança é dada de modo que os estimadores $\hat{\alpha}$ e $\hat{\beta}$ maximizem o logaritmo da função. Assim para maximizar a função, basta derivar em relação aos parâmetros do modelo.

Desta forma, temos:

$$L(\alpha, \beta | x, y) = \sum_{i=1}^n y_i(\alpha + \beta_i x_i) - \ln(1 + e^{\alpha + \beta_i x_i}) \quad (2.18)$$

Derivando em relação aos parâmetros:

$$\begin{aligned} \frac{\partial L(\alpha, \beta)}{\partial \alpha} &= \sum_{i=1}^n y_i - \left(\frac{e^{\alpha + \beta_i x_i}}{1 + e^{\alpha + \beta_i x_i}} \right) \\ \frac{\partial L(\alpha, \beta)}{\partial \beta_1} &= \sum_{i=1}^n y_i x_{i1} - x_{i1} \left(\frac{e^{\alpha + \beta_i x_i}}{1 + e^{\alpha + \beta_i x_i}} \right) \\ &\vdots \\ \frac{\partial L(\alpha, \beta)}{\partial \beta_n} &= \sum_{i=1}^n y_i x_{in} - x_{in} \left(\frac{e^{\alpha + \beta_i x_i}}{1 + e^{\alpha + \beta_i x_i}} \right) \end{aligned} \quad (2.19)$$

Ao igualar estes termos a zero, temos:

$$\sum_{i=1}^n y_i - \frac{e^{\hat{\alpha} + \hat{\beta}_i x_i}}{1 + e^{\hat{\alpha} + \hat{\beta}_i x_i}} = 0 \quad e \quad \sum_{i=1}^n x_i \left(y_i - \frac{e^{\hat{\alpha} + \hat{\beta}_i x_i}}{1 + e^{\hat{\alpha} + \hat{\beta}_i x_i}} \right) = 0 \quad (2.20)$$

Então notamos que $\hat{\alpha}$ e $\hat{\beta}$ são os estimadores dos parâmetros de α e β .

2.5 Curva ROC(Receiver Operating Characteristic)

A análise do modelo de curva ROC é uma ferramenta muito utilizada para avaliar e especificar o desempenho com problemas relacionados a medicina [BRAGA, 2000](#). Esta análise é feita pelo método gráfico, permitindo estudar a variação de sensibilidade e especificidade, para diferentes valores. A sensibilidade avalia o percentual de resultado positivo dentre uma determinada doença ou condição avaliada, e a especificidade avalia a capacidade do mesmo teste ter resultado negativo nos indivíduos que não apresentam essa condição investigada.

Os níveis de avaliação deste resultado é dado por:

$$\begin{aligned} \text{área} < 0,7 &\rightarrow \text{baixa precisão} \\ 0,7 < \text{área} < 0,8 &\rightarrow \text{precisão aceitável} \\ 0,8 < \text{área} < 0,9 &\rightarrow \text{precisão excelente} \\ \text{área} > 0,9 &\rightarrow \text{precisão excepcional} \end{aligned}$$

2.6 Critério de informação de Akaike(AIC)

O Critério de Informações de Akaike(AIC) avalia a qualidade do ajuste dos modelos, sendo estimado pelo método de máxima verossimilhança. O AIC é definido pela função abaixo apresentada:

$$AIC = -2\ln(L(\alpha, \beta)) + 2(k) \quad (2.21)$$

onde $L(\alpha, \beta)$ é a função de máxima verossimilhança do modelo e k o número de variáveis explicativas do modelo. Assim, dado o conjunto de dados e os modelos, pode-se classificar cada um de acordo com seu AIC, tal que quanto menor o valor estimado, melhor é o ajuste do modelo [EMILIANO, 2009](#).

3 Metodologia

3.1 Tipo de pesquisa

Um estudo transversal foi conduzido com a finalidade de analisar as incidências de internações hospitalares decorrentes da Síndrome Respiratória Aguda Grave (SRAG) atribuídas à COVID-19 e documentadas no Sistema de Informação de Vigilância Epidemiológica da Influenza (SIVEP-Gripe), sob a responsabilidade do Ministério da Saúde (MS).

3.1.1 Dados

O conjunto de dados utilizados neste estudo se encontram na página oficial do DataSUS [SAÚDE, 2023] os quais contém informações sobre todos os casos notificados de SRAG(Síndrome Respiratória Aguda Grave). É possível encontrar registros desde 2009, quando ocorreu a pandemia da Influenza A(H1N1), e são atualizados todos os anos.

Os dados utilizados neste trabalho são os notificados desde o começo da pandemia do Covid-19, que se iniciou em 2020 e o qual é atualizado até o momento atual. Neste estudo serão analisados os dados dos anos de 2020 a 2023, contemplando quase quatro anos de informações.

Os dados consistem em informações sobre cada paciente notificado pela rede pública, desde informações pessoais até a evolução da SRAG em cada indivíduo registrado. Desta forma, cada paciente é uma linha do banco de dados e as colunas dizem respeito a uma ficha de registro individual, na qual são preenchidos 175 campos de dados, contendo as informações pessoais, região de notificação, sintomas, comorbidades, vacinação e a evolução de cada paciente(Anexo). A base de dados é protegida pela Lei 13.709/2018 [BRASIL, 2018] à qual passa por um tratamento de anonimização, logo são disponibilizados somente os dados que não comprometam a integridade de cada paciente.

3.1.2 Período de análise

O período de coleta de dados é de 24/3/2020 até 28/9/2023 e foram acessados dia 05/10/2023, uma quinta-feira, dado que o sistema é atualizado todas as quartas, dessa forma, mesmo os dados pretéritos podem sofrer alterações em novas atualizações do SIVEP-Gripe.

3.1.3 População e amostra

A população-alvo deste estudo foi composta por um total de 3.706.212 indivíduos hospitalizados com diagnóstico de Síndrome Respiratória Aguda Grave (SRAG) atribuída à COVID-19 no contexto nacional brasileiro, a partir da qual a amostra de análise foi selecionada. A amostra-alvo selecionada são os 198.870 casos atribuídos ao Estado do Rio Grande do Sul (RS). A amostra final, após processo de filtragem, conta com 129.037 pacientes.

3.1.4 Critérios de Inclusão

Este estudo incorporou o registro de indivíduos hospitalizados no estado do RS, de ambos os sexos, com idades superiores a 1 ano, diagnosticados tendo como classificação final Síndrome Respiratória Aguda Grave (SRAG) associada à COVID-19, provenientes de instituições hospitalares, tanto públicas quanto privadas, em território brasileiro.

3.2 Análise de Dados

A análise de dados foi composta pelas seguintes etapas:

- Filtragem;
- Análise exploratória de dados;
- Regressão logística.

3.2.1 Filtragem

O processo de filtragem baseia-se em escolher o subconjunto de dados (amostra) que se irá trabalhar, mantendo o conjunto de dados original (população). Com a seleção desse subconjunto segue-se a análise de dados. Neste processo podem ser feitas exclusões de informações que não são pertinentes no contexto analisado.

Essa filtragem precisa seguir uma sequência lógica dos dados que complementam cada subconjunto escolhido, para obtermos um resultado coerente diante de cada análise que será efetuada.

3.2.2 Análise Exploratória de Dados

A análise exploratória de dados consiste em organizar e interpretar os dados coletados antes de aplicar quaisquer método estatístico, tendo pleno entendimento das variáveis trabalhadas e as relações que possuem entre si.

As variáveis são divididas em dois grupos, sendo eles:

- Variáveis quantitativas: são as variáveis que podem ser contadas ou mensuradas, que apresentam dois subgrupos:
 - Contínuas: que apresentam valores em escala contínua, podendo assumir infinitos valores. Como por exemplo temperatura corporal.
 - Discretas: que apresentam valores finitos ou infinitos contáveis, sendo eles valores inteiros. Como por exemplo número de filhos e número de casos de gripe.
- Variáveis qualitativas: são as variáveis categóricas, que não possuem valores quantitativos, representando uma classificação dos dados, que apresentam também dois subgrupos:
 - Nominais: são as variáveis que não apresentam ordenação, apenas para identificar a categoria. Como por exemplo sexo e endereço.
 - Ordinais: são as variáveis que apresentam ordenação das categorias. Como por exemplo faixa etária e escolaridade.

As variáveis ainda podem ser classificadas, de acordo com o interesse de cada estudo, como dependentes ou independentes. As variáveis independentes são ajustadas em conformidade com as variáveis dependentes selecionadas, exercendo influência sobre os resultados obtidos por meio das variáveis dependentes. Assim, através de cada metodologia analítica empregada na análise dos dados, é possível examinar esse nível de influência entre as variáveis, permitindo a obtenção de resultados estatisticamente significativos em cada estudo realizado.

3.2.3 Regressão Logística

Regressão logística é a análise de dados entre dois ou mais fatores, distintos em uma variável binária dependente e uma ou mais variáveis independentes, na qual cada variável independente escolhida influencia nos resultados da variável dependente analisada. Segundo [FÁVERO; BELFIORE, 2017](#) Regressão Logística é um método de modelagem destinado a elucidar a relação entre uma variável dependente binária e variáveis independentes, as quais podem ser contínuas ou categóricas. Desta forma, essa abordagem mostra a regressão logística como uma forma de regressão múltipla que se propõe a investigar o impacto das variáveis às quais indivíduos ou sujeitos estão expostos na probabilidade de ocorrência de um evento específico de interesse. Essa análise nos resulta em uma tabela que apresenta como cada variável dependente tem resultado significativo em relação a cada variável independente, ou seja, qual a significância daquela variável para que o evento ocorra.

3.3 Suporte Computacional

Para executar as filtrações e análises descritivas dos dados apresentadas neste trabalho foi utilizado o *Software R 4.11*, o qual é uma ferramenta amplamente utilizada para análises e manipulação de dados. O R é um software de análise de dados de código aberto que teve sua origem em 1996, fruto do trabalho dos professores de estatística Ross Ihaka e Robert Gentleman, da Universidade de Auckland. Eles conceberam uma nova linguagem computacional, assemelhada à linguagem S criada por John Chambers.

O *Software R* oferece uma vasta gama de técnicas estatísticas e gráficas, abrangendo desde funções básicas matemáticas e estatística até modelagem linear e não linear, bem como testes estatísticos tradicionais, análise de séries temporais, classificação, agrupamento e outras ferramentas diversas. Possui uma comunidade bastante ativa, a qual contribui de forma significativa com novos pacotes e funções. A grande capacidade de extensão do R possibilita aos usuários a criação de pacotes para aplicações ou campos de estudo específicos.

A regressão logística foi efetuada no software *Jamovi*. O Jamovi é um software de código aberto para análise de dados e testes estatísticos, derivado do projeto JASP, baseado na linguagem computacional do R, com uma interface gráfica, sendo útil para explorar dados, bem como efetuar análises descritivas e inferenciais. Desenvolvido por Jonathon Love, Damian Dropmann e Ravi Selker em 2017, o software oferece uma interface gráfica para a linguagem R, sendo aplicável em pesquisas estatísticas diversas.

3.4 Ética da pesquisa

O trabalho aqui realizado, por utilizar dados secundários de domínio público e não envolver diretamente seres humanos, não necessitou ser submetido à avaliação do Comitê de Ética em Pesquisa.

4 Resultados e Discussão

A análise realizada contempla dados dos quatro anos divulgados pela página oficial do DataSUS, sendo os anos de 2020 à 2023, no mês de agosto, somando 175 colunas de variáveis e 3.706.212 casos de pacientes hospitalizados notificados. Todo procedimento foi realizado no software R.

4.1 Filtragem

O primeiro passo foi a leitura dos 4 conjuntos de dados, onde observou-se que os anos de 2020 e 2021 não possuíam o mesmo número de variáveis que os outros anos. Em tais anos faltavam informações sobre as vacinas específicas do Covid-19. Essas colunas faltantes foram adicionados a estes anos com entradas NA, ou seja, um valor que apesar de existir não está disponível, para que no momento que fosse unificar as planilhas não houvesse erro. Totalizando 3.706.212 registros de pacientes em todo o Brasil.

Com os conjuntos de dados ajustados foram selecionados os dados com base em uma coluna variável específica SG_UF_NOT (Unidades Federativas onde se localiza a Unidade que realizou a notificação). Neste trabalho foi utilizado os conjuntos notificados no Rio Grande do Sul/RS iniciando o banco de dados com 198.820 registros.

A partir disso selecionou-se as 67 variáveis analisadas nesse estudo, sendo elas variáveis de região, sintomas, comorbidades, vacinas e evolução dos pacientes. Com estes conjuntos ajustados pode-se compilar os quatros anos em uma única planilha, somando-se 198.870 pacientes hospitalizados registrados em unidades de saúde do estado do RS.

Ajustou-se a variável, SEM_NOT(Semana Epidemiológica do preenchimento da ficha de notificação), pois estava numerada de 1 a 52 em cada um dos anos de 2020 à 2023. Após este ajuste, somam 197 semanas desde o começo dos registros.

A variável NU_IDADE_N(Idade dos pacientes), também serviu de base para as filtragens, pois as idades foram computadas com números inteiros, sendo calculadas automaticamente pelo sistema, considerando o intervalo entre a data de nascimento e a data dos primeiros sintomas, e há o tipo de idade, sendo em dias, se a diferença do registro ser entre 0 e 30 dias, em meses, se a diferença de registro for entre 1 e 11 meses e em anos, se a diferença for maior ou igual a 12 meses. Com essas especificações foi utilizado somente idades registradas a partir de 1 ano, assim resultaram 198.535 pacientes.

Com isto, foi realizada a filtragem dos dados mais relevantes para a pesquisa segundo a variável EVOLUCAO(Evolução do caso), a qual apresenta 4 categorias:cura, óbito, óbito por outras causas e ignorado. Os dados relevantes para a pesquisa são os

de cura e óbito, assim foi feita outra filtragem a partir desta variável, o qual resultou no conjunto de 186.378 pacientes hospitalizados registrados.

Por fim incorporou-se a variável CLASS_FIN(Classificação final do caso), que nos mostra o diagnóstico final do paciente, sendo dividido em cinco categorias: SRAG por influenza; SRAG por outro vírus respiratório; SRAG por outro agente etiológico(especificando qual agente); SRAG não especificado; SRAG por Covid-19. No projeto foi utilizado a categoria de covid-19. Esta variável só foi implementada no SIVEP-GRIPE em janeiro de 2021, assim os dados efetivamente analisados são de 03 de janeiro de 2021 até 28 de setembro de 2023 finalizando em 94.192 pacientes.

4.2 Análise Exploratória de Dados

A população de pacientes para o estudo contou com 43.495 do sexo feminino (46,18%) e 50.697 do sexo masculino(53,82%), com idades entre 1 e 112 anos, como apresentados na tabela 1:

Tabela 1 – Idades dos pacientes

	Sexo	Idade
Média	F	60,3
	M	58,1
Mediana	F	62
	M	59
Moda	F	59
	M	57

Fonte: Elaborado pelo autor

Analisando a raça desta população observamos que 85% dos pacientes são brancos, somando 80.092, 3,6% são pretos, somando 3.349 pacientes, 0,3% amarelos, somando 239 pacientes, 3,8% pardos, somando 3.613 pacientes e 0,1% indígenas, somando 134 pacientes, com 7,2% dos pacientes com informação de raça não declarada, somando 6.765 pacientes.

Quanto a escolaridade podemos observar na tabela 2 que a maioria dos pacientes do sexo feminino tinha apenas o fundamental 1 (1° a 5° série) completo, sendo um total de 5.599 mulheres, e do sexo masculino apenas o fundamental 2 (6° a 9° série) completo, sendo um total de 6.167 homens, e apenas 4.392 pacientes tinham superior completo, dentre homens e mulheres.

Onde *Não se aplica* é preenchido para pacientes com idade menor que 7 anos, e *Ignorado* são pacientes que não preencheram este campo. Os níveis de Fundamental 1, Fundamental 2 e Médio são considerados a partir da última série ou ano concluído.

Tabela 2 – Escolaridade

Sexo	Escolaridade	Contagem	% Total
F	Analfabeto	891	1,6%
	Fundamental 1	5.599	10,3%
	Fundamental 2	4.800	8,8%
	Médio	4.748	8,7%
	Superior	1.855	3,4%
	Não se aplica	191	0,4%
	Ignorado	6.840	12,6%
M	Analfabeto	817	1,5%
	Fundamental 1	5.992	11%
	Fundamental 2	5.483	10,1%
	Médio	6.167	11,4%
	Superior	2.537	4,7%
	Não se aplica	218	0,4%
	Ignorado	8.186	15,1%

Fonte: Elaborado pelo autor

Analisando os dados do grupo de fator de risco podemos notar que 66.063 pacientes pertenciam a este grupo, somando 70,1% da população total. Destes 34,4% são do sexo feminino e 35,7% do sexo masculino, sendo 32.412 mulheres e 33.651 homens respectivamente. Na tabela 3 podemos observar as idades deste grupo:

Tabela 3 – Fator de Risco

	Fator de Risco	Idade
Média	Sim	63,1
	Não	49,7
Mediana	Sim	65
	Não	49
Moda	Sim	70
	Não	56

Fonte: Elaborado pelo autor

Dentro do grupo de Fator de Risco podemos observar que dos pertencentes a este grupo 41,1% foram curados da Covid-19 e 29% vieram a óbito, e os que não estavam incluso no grupo 25,3% foram curados e 4,5% foram a óbito.

4.3 Regressão Logística

Para a estimação do modelo de Regressão Logística foram feitas análises consideradas importantes até chegar o modelo final utilizado na pesquisa. Essas análises foram construídas com base nas informações dos grupos demográficos, sintomas, comorbidades, vacinação e evolução de cada paciente.

Primeiramente foram construídas regressões em cada grupo separadamente, e dentre eles verificando quais possuíam valores mais significativos para uma análise de regressão final. Com isso foi analisado a curva ROC em cada uma das regressões, observando quais tinham maior acurácia para um modelo final.

Dentre esses modelos realizados, foram selecionados os que obtinham maior significância dentro de cada um dos grupos, os quais poderiam nos fornecer um modelo apropriado para a regressão final. Todos eles foram estimados com base na variável independente EVOLUCAO, para obter os resultados significativos com cada variável independente dos grupos.

Num primeiro momento chegamos a um modelo com variáveis dos diversificados grupos iniciais, contendo 6 variáveis explicativas, sendo elas: VACINA_COV (paciente que recebeu a vacina Covid-19), NU_IDADE_N (idade paciente), DISPNEIA, DESC_RESP (pacientes que tiverem desconforto respiratório), UTI (Unidade de Terapia Intensiva) e SATURACAO (paciente que apresentou saturação $O_2 < 95\%$), como mostrado abaixo:

Tabela 4 – Modelo 1 - Chances de óbito ou cura

Preditor	Coefficiente Estimado	Erro-padrão	Z	p-valor	OR
Intercepto	-6,0484	0,10220	-59,18	<0,001	0,00236
VACINA_COV	-0,3451	0,04012	-8,60	<0,001	0,70813
NU_IDADE_N	0,0571	0,00127	44,88	<0,001	1,05877
DISPNEIA	0,3818	0,05827	6,55	<0,001	1,46485
DESC_RESP	0,1895	0,04621	4,10	<0,001	1,20868
UTI	2,3687	0,03740	63,34	<0,001	10,68367
SATURACAO	0,4706	0,05350	8,80	<0,001	1,60103

Fonte: Elaborado pelo autor

Observando os resultados apresetados, podemos avaliar o impacto de cada uma das variáveis de acordo com seus coeficientes, de maneira que os valores positivos representam pacientes com maior probabilidade de virem a óbito. Da mesma maneira avaliamos a acurácia do modelo, que neste primeiro momento representa 79,2% de acerto na classificação geral, com especificidade de 89,8% e sensibilidade de 52,9%.

Com isso, avaliamos outro modelo com interações entre duas variáveis explicativas, sendo a variável UTI com a variável DISPNEIA, nos resultando no Modelo 1 apresentado na tabela 5.

Com o Modelo 2, observamos que a acurácia permaneceu a mesma, sendo de 79,2%, porém com especificidade de 89,7% e sensibilidade de 53,2%, nos apresentando um AIC menor que o primeiro, o que implica em um resultado melhor que o Modelo 1.

A partir disso chegou-se a um modelo final, o Modelo 3, com um conjunto de dados com 24.028 objetos, contendo 6 variáveis explicativas e sendo 2 iterações: variável UTI com a variável DISPNEIA e a variável UTI com a variável VACINA_COV.

Tabela 5 – Modelo 2 - Chances de óbito ou cura

Preditor	Coefficiente Estimado	Erro-padrão	Z	p-valor	OR
Intercepto	-6,1267	0,10690	-57,31	<0,001	0,00218
VACINA_COV	-0,3446	0,04012	8,59	<0,001	0,70848
NU_IDADE_N	0,0571	0,00127	44,86	<0,001	1,05875
DISPNEIA	0,4812	0,06974	6,90	<0,001	1,61809
DESC_RESP	0,888	0,04629	4,08	<0,001	1,20783
UTI	2,6570	0,11249	23,62	<0,001	14,25382
SATURACAO	0,4654	0,05368	8,67	<0,001	1,59259
UTI*DISPNEIA	-0,3200	0,11752	-2,72	0,006	0,72612

Fonte: Elaborado pelo autor

Os resultados do Modelo 3 podem ser observados na tabela 6.

Tabela 6 – Modelo 3 - Chances de óbito ou cura

Preditor	Coefficiente Estimado	Erro-padrão	Z	p-valor	OR
Intercepto	-6,2643	0,10910	-57,42	<0,001	0,00190
VACINA_COV	-0,0852	0,05192	-1,64	0,101	0,91832
NU_IDADE_N	0,0566	0,00128	44,31	<0,001	1,05822
DISPNEIA	0,5098	0,07010	7,27	<0,001	1,66490
DESC_RESP	0,1860	0,04628	4,02	<0,001	1,20440
UTI	2,9773	0,11952	24,91	<0,001	19,63526
SATURACAO	0,4687	0,05368	8,73	<0,001	1,59798
UTI*DISPNEIA	-0,3785	0,11739	-3,22	0,001	0,68488
UTI*VACINA_COV	-0,5620	0,07049	-7,97	<0,001	0,57004

Fonte: Elaborado pelo autor

Com este modelo, podemos observar que a acurácia foi maior, sendo de 79,5%, e com especificidade de 89,7% e sensibilidade de 53,2%, sendo um modelo melhor ajustado.

Com isso, observando a tabela 7, temos que o menor AIC é o do Modelo 3, sendo o que melhor ajusta o nosso modelo.

Tabela 7 – Critério de Akaike - AIC

Modelo	Desviância	AIC
Modelo 1	20.881	20.895
Modelo 2	20.874	20.890
Modelo 3	20.810	20.828

Fonte: Elaborado pelo autor

A tabela 6 mostra as variáveis selecionadas e as estatísticas gerada pelo modelo logístico, com nível de significância de 5%. Sendo dada pela função matemática abaixo:

$$f(Z) = \frac{e^Z}{1 + e^Z} \quad (4.1)$$

$$\begin{aligned}
Z = & - 6,2643 - 0,0852 \times \text{VACINA_COV} + 0,0566 \times \text{NU_IDADE_N} \\
& + 0,5098 \times \text{DISPNEIA} + 0,1860 \times \text{DESC_RESP} + 2,9773 \times \text{UTI} \\
& + 0,4687 \times \text{SATURACAO} - 0,3785 \times (\text{UTI} \star \text{DISPNEIA}) \\
& - 0,5620 \times (\text{UTI} \star \text{VACINA_COV})
\end{aligned}$$

onde

Observa-se que a variável mais significativa para o caso de óbitos do modelo é o da UTI, na qual há 18 vezes(1.800%) mais chances do paciente vir a óbito, seguidas das variáveis DISPNEIA e SATURACAO, com 66% e 59% mais chances respectivamente.

Por outro lado, temos as variáveis com coeficientes negativos, que nos dizem que há uma redução na chances de óbitos destes pacientes, sendo elas UTI★DISPNEIA e UTI★VACINA_COV, com 32% e 43% menos chances respectivamente. Ou seja, mesmo com as chances elevadas de óbito de pacientes que necessitam da UTI, aquelas que fizeram a Vacina contra a Covid-19 reduzem suas chances de virem a óbito. O mesmo acontece com quem possuía dispnéia, as chances de virem a óbito quem teve o sintoma e foi para a UTI é reduzida.

Observa-se também que há uma redução no risco de óbitos quando o paciente havia feito a vacina para o Covid-19, sendo 9,9% menor o risco. Assim como quem teve desconforto respiratório tem maior chances de vir a óbito, sendo 20% mais elevadas.

4.3.1 Avaliação do modelo

Com nosso modelo finalizado e ajustado, podemos analisar a acurácia para conseguirmos avaliar a eficiência dos resultados encontrados.

O modelo encontrado pelo método de Regressão Logística apresentou um percentual de acerto de classificação geral de 79,5%, nos dizendo que há um bom resultado de classificação.

Tabela 8 – Tabela de Classificação

Observado	1	0	% Correta
1	15.338	1.792	89,5%
0	3.130	3.768	54,6%

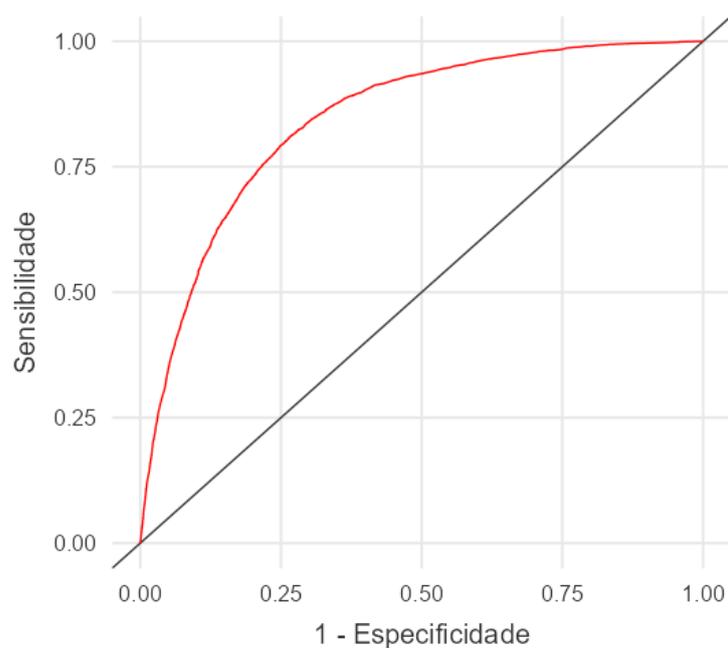
Fonte: Elaborado pelo autor

A sensibilidade do Modelo 3, de acordo com a tabela 8, que é quando o modelo classifica como óbito os pacientes que realmente vieram a óbito, nos mostra que o modelo acertou 54,6% corretamente, bem como a especificidade, que é a capacidade de classificar como cura os que realmente foram curados foi de 89,5% corretamente.

Este modelo foi considerado aceitável, o seja, adequado e ajustado segundo a curva ROC apresentada por ele, sendo uma acurácia de 0,795, com nível de significância

de 5%. Segundo BRAGA, 2000 o modelo possui precisão aceitável quando a área da curva estiver entre 0,7 e 0,8, como nos apresenta a figura 2 a seguir. Sendo assim, como os ajustes do modelo são aceitáveis, a acurácia do modelo também é aceitável.

Figura 2 – Curva ROC de chances de Cura ou Óbito



Fonte: Elaborado pelo autor.

5 Considerações finais

Com este estudo, o objetivo foi por meio de Regressão Logística encontrar fatores associados ao óbito de pacientes internados por Covid-19 no estado do Rio Grande do Sul. Para isso, foi apresentado conceitos e definições estatísticas, como função logística, função de verossimilhança, estimação de parâmetros, curva ROC e critério de Akaike. Para a construção do modelo de Regressão Logística que se utilizou no decorrer do estudo foram feitas regressões múltiplas, afim de encontrar o modelo que melhor se ajustava de acordo com o Critério de Akaike.

Para estimar o modelo que mais se ajustava foram feitas análises iniciais de acordo com grupos específicos de dados. A partir disso, foram selecionados dados relevantes em cada grupo específico, e analisando os valores estimados e a razão de chances, a fim de chegar no modelo ideal. Desta forma, encontramos um modelo com precisão aceitável, de 79,5% de acurácia, nos dando um bom resultado de classificação para o modelo.

Com o modelo acurado, pode-se observar riscos elevados para pacientes que tiveram que fazer uso da UTI, sendo de 18 vezes mais elevado o risco de morte para estes pacientes. Assim como resultados significativos encontrados para pacientes com sintomas de Dispneia, na qual tinham 66% mais chances de óbito que outros pacientes avaliados, e também para pacientes com níveis de saturação abaixo do considerado normal, de $O_2 < 95\%$, com 59% a mais de chances de virem a óbito. A idade destes pacientes se mostrou como menos relevante para o risco de mortalidade, mas nos mostra que conforme a idade aumenta, os riscos dos pacientes associados aos fatores de risco e sintomas também aumentam.

Diante disso, também observou-se que pacientes que se vacinaram contra a Covid-19 reduziam o risco de mortalidade, tal que quem havia se vacinado reduzia o seu risco em 9,9%. Já quem havia feito a vacina e precisou fazer uso da UTI, tinham chances maiores de cura durante o tratamento, sendo o risco de mortalidade reduzido em 43%.

Após o estudo aqui realizado pode-se concluir que as ferramentas matemáticas e estatísticas são ferramentas essenciais para avaliar risco e fatores associados a morbidades na área da saúde.

Referências

- ALBUQUERQUE, H. A. d. *A covid-19 e os fatores que influenciam a probabilidade de mortalidade: uma análise de regressão logística com dados do Sistema de Único de Saúde do Brasil*. 2022. Disponível em: <https://repositorio.ufersa.edu.br/items/b7ff6d97-2f2a-437c-8144-633b85696ba2/full>. Acesso em: 20 Nov 2023. Citado na página [12](#).
- AMARAL, G. V. d. S. A. C. *MANIFESTAÇÕES NEUROCOMPORTAMENTAIS DA COVID-19: uma revisão bibliográfica*. 2021. Disponível em: <https://www.unifan.edu.br/unifan/aparecida/wp-content/uploads/sites/2/2023/03/MANIFESTACOES-NEUROCOMPORTAMENTAIS-DA-COVID-19-uma-revisao-bibliografica.pdf>. Acesso em: 20 Nov 2023. Citado na página [11](#).
- BRAGA, A. C. da S. *Curvas ROC: aspectos funcionais e aplicações*. Tese (Doutorado) — Universidade do Minho, 2000. Citado 2 vezes nas páginas [18](#) e [29](#).
- BRASIL. Lei nº 12.651/2012. dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural. *Diário Oficial da República Federativa do Brasil*, Brasília, DF, 2018. ISSN 1677-7042. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Citado na página [19](#).
- CAMARGO, L. *Comparação entre a ocorrência da infecção por SARS-CoV-2 em profissionais da saúde e não profissionais da saúde no Rio Grande do Sul*. 2021. Disponível em: <https://lume.ufrgs.br/handle/10183/253678>. Acesso em: 20 Nov 2023. Citado na página [12](#).
- EMILIANO, P. C. *Fundamentos e Aplicações dos Critérios de Informação: Akaike e Bayesiano*. Tese (Doutorado) — Universidade Federal de Lavras, 2009. Citado na página [18](#).
- FÁVERO, L.; BELFIORE, P. *Manual de Análise de Dados: Estatística e Modelagem Multivariada com Excel®, SPSS® e Stata®*. Elsevier Editora Ltda., 2017. ISBN 9788535285055. Disponível em: <https://books.google.com.br/books?id=SmlaDwAAQBAJ>. Citado 2 vezes nas páginas [14](#) e [21](#).
- GUIMARÃES, A. C. N. I. A. *RECONHECIMENTO DE PADRÕES: METODOLOGIAS ESTATÍSTICAS EM CRÉDITO AO CONSUMIDOR*. 2002. Disponível em: https://www.researchgate.net/figure/Figura-3-Grafico-da-Funcao-Sigmoide-y-1_fig1_228600220. Acesso em: 12 Set 2023. Nenhuma citação no texto.
- LIMA, C. M. A. de O. *Informações sobre o novo coronavírus (COVID-19)*. 2020. Disponível em: <https://doi.org/10.1590/0100-3984.2020.53.2e1>. Acesso em: 20 Nov 2023. Citado na página [11](#).

MACIEL, E. L. et al. Fatores associados ao óbito hospitalar por covid-19 no espírito santo, 2020. *Epidemiologia e Serviços de Saúde*, Secretaria de Vigilância em Saúde e Ambiente - Ministério da Saúde do Brasil, v. 29, n. 4, p. e2020413, 2020. ISSN 2237-9622. Disponível em: <https://doi.org/10.1590/S1679-49742020000400022>. Citado na página 12.

PASTE, A. et al. Análise de fatores associados a mortalidade em pacientes internados com a covid 19 em um hospital privado em salvador-bahia. *The Brazilian journal of infectious diseases*, Elsevier España, S.L.U, v. 26, p. 102005, 2022. ISSN 1413-8670. Citado na página 12.

SAÚDE, M. da. *SRAG*. 2023. Disponível em: <https://opendatasus.saude.gov.br/dataset?tags=SRAG>. Acesso em: 04 Set 2023. Citado na página 19.



Universidade Federal do Rio Grande – FURG

Instituto de Matemática, Estatística e Física

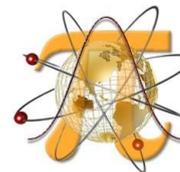
Curso de Bacharelado em Matemática Aplicada

Av. Itália km 8 Bairro Carreiros

Rio Grande-RS CEP: 96.203-900 Fone (53)3293.5411

e-mail: imef@furg.br

Sítio: www.imef.furg.br



Ata de Defesa de Monografia

No dia 22 do mês de janeiro de 2024, às 14h, foi realizada a apresentação pública da defesa do Trabalho de Conclusão de Curso do acadêmico Diego **Mateus Müller**, sob orientação da Prof^ª. Doutora Raquel da Fontoura Nicolette, deste instituto, e intitulada **Fatores associados ao óbito hospitalar por COVID-19 no Rio Grande do Sul - Uma aplicação de Regressão Logística**. Para participar da banca avaliadora junto a orientadora foram convidados o Prof. Doutor Adilson da Silva Nunes IMEF/FURG, e a Prof^ª. Doutora Juliana da Silva Ricardo Nunes - IMEF - FURG IMEF/FURG. Concluídos os trabalhos de apresentação e arguição, o candidato foi: (X) aprovado por unanimidade; () aprovado somente após satisfazer as exigências que constam na folha de modificações, no prazo fixado pela banca; () reprovada. Na forma regulamentar, foi lavrada a presente ata, que é abaixo assinada pelos membros da banca, na ordem acima relacionada.

Documento assinado digitalmente

gov.br

RAQUEL DA FONTOURA NICOLETTE

Data: 10/02/2025 09:43:03-0300

Verifique em <https://validar.iti.gov.br>

Prof^ª. Doutora Raquel da Fontoura Nicolette
(Orientadora-FURG)

Documento assinado digitalmente

gov.br

Documento assinado digitalmente

JULIANA DA SILVA RICARDO NUNES

Data: 26/01/2024 19:19:24-0300

Verifique em <https://validar.iti.gov.br>

Prof^ª. Doutora Juliana da Silva Ricardo
Nunes
(Avaliadora – IMEF-FURG)

Documento assinado digitalmente

gov.br

ADILSON DA SILVA NUNES

Data: 29/01/2024 00:20:14-0300

Verifique em <https://validar.iti.gov.br>

Prof. Doutor Adilson da Silva Nunes
(Avaliador - IMEF - FURG)