

Luccas Martins da Luz

**Avaliação do perfil das internações
hospitalares emergenciais em idosos com
doenças respiratórias e de possíveis fatores
associados ao desfecho**

Niterói - RJ, Brasil

11 de fevereiro de 2022

Luccas Martins da Luz

**Avaliação do perfil das internações
hospitalares emergenciais em idosos
com doenças respiratórias e de
possíveis fatores associados ao
desfecho**

Trabalho de Conclusão de Curso

Monografia apresentada para obtenção do grau de Bacharel em
Estatística pela Universidade Federal Fluminense.

Orientador(a): Prof. Dra. Núbia Karla de Oliveira Almeida

Niterói - RJ, Brasil

11 de fevereiro de 2022

Luccas Martins da Luz

**Avaliação do perfil das internações
hospitalares emergenciais em idosos com
doenças respiratórias e de possíveis fatores
associados ao desfecho**

Monografia de Projeto Final de Graduação sob o título “*Avaliação do perfil das internações hospitalares emergenciais em idosos com doenças respiratórias e de possíveis fatores associados ao desfecho*”, defendida por Luccas Martins da Luz e aprovada em 11 de fevereiro de 2022, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

Profa. Dra. Núbia Karla de Oliveira Almeida
Departamento de Estatística – UFF

Prof. Dra. Maria Cristina Bessa Moreira
Departamento de Estatística – UFF

Prof. Dr. Renan Moritz V. R. Almeida
Programa de Engenharia Biomédica - COPPE - UFRJ

Niterói, 11 de fevereiro de 2022

Ficha catalográfica automática - SDC/BIME
Gerada com informações fornecidas pelo autor

L979a Luz, Luccas Martins da
Avaliação do perfil das internações hospitalares
emergenciais em idosos com doenças respiratórias e de
possíveis fatores associados ao desfecho / Luccas Martins da
Luz ; Núbia Karla de Oliveira Almeida, orientador. Niterói,
2022.
44 f. : il.

Trabalho de Conclusão de Curso (Graduação em
Estatística)-Universidade Federal Fluminense, Instituto de
Matemática e Estatística, Niterói, 2022.

1. Regressão logística. 2. Razão de chances ajustada. 3.
Internação hospitalar emergencial. 4. Idosos. 5. Produção
intelectual. I. Oliveira Almeida, Núbia Karla de, orientador.
II. Universidade Federal Fluminense. Instituto de Matemática
e Estatística. III. Título.

CDD -

Resumo

Introdução: O cenário mundial da pandemia por Corona Vírus gerou questionamentos que o comportamento das internações hospitalares decorrentes desta doença poderia ser similar às ocorridas por doenças respiratórias. O objetivo deste trabalho é avaliar o perfil das internações emergenciais hospitalares em idosos, decorrentes de problemas respiratórios, e identificar possíveis fatores associados ao óbito. Metodologia: Modelos de regressão logística em função de características dos pacientes, das internações e do hospital foram utilizados tendo como variável resposta a variável indicadora do desfecho (alta/óbito), tomando como base registros do SIHSUS/DATASUS, do Rio de Janeiro no ano de 2019. Resultados: Alguns dos principais fatores de risco para o óbito associados ao modelo que considera os diagnósticos principais classificados em apenas 4 categorias e incorpora a informação sobre utilização da UTI foram: classe etária 80 anos ou mais (RC: 1,86), uso da UTI (RC: 4,10) e diagnóstico de insuficiência respiratória (RC: 12,26).

Palavras-chave: Regressão logística, Razão de chances ajustada, Internação hospitalar emergencial e Idosos.

Dedicatória

Dedico este Trabalho de Conclusão de Curso a minha mãe Luciana, a meu pai Altair e a meus avós José Augusto e Sidneia que sempre deram tudo por mim e agora é hora de retribuir.

Agradecimentos

Primeiramente gostaria de agradecer à minha mãe que sempre esteve ao meu lado desde o início da minha vida e foi sempre meu porto seguro em todas as etapas da minha caminhada. Agradeço também ao meu pai, que sempre esteve presente na minha vida me dando ótimos conselhos e me incentivando quando eu mais precisava. Aos meus avós que estiveram comigo quando minha mãe não pôde e hoje só sou esta pessoa e só pude estar aqui por causa deles.

Não poderia deixar de citar a minha professora, orientadora, mentora, conselheira e amiga Núbia. Ela me apoiou, me deu forças e esteve lá por mim no momento em que eu estava mais fraco e nós sabemos muito disso. Foi muito mais que uma orientadora, foi uma luz e vou ser eternamente grato a ela.

À minha namorada Isabel, que é também meu porto seguro e nos bons e maus momentos esteve sempre comigo para me auxiliar, motivar e não deixar que duvidasse da minha capacidade em nenhum momento.

Por fim, agradecer ao meu grande amigo e irmão Leonardo, que me incentivou a entrar no curso de Estatística e não estaria aqui sem o apoio e companheirismo de sempre dele. Além de meus outros grandes amigos.

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 11
1.1	Motivação	p. 11
1.2	Revisão Bibliográfica	p. 12
1.3	Objetivos	p. 13
1.4	Organização do Trabalho	p. 13
2	Materiais e Métodos	p. 14
2.1	Dados	p. 14
2.2	Metodologia	p. 16
2.2.1	Modelos Lineares Generalizados	p. 16
2.2.1.1	Introdução	p. 16
2.2.1.2	Componentes	p. 16
2.2.2	Regressão Logística	p. 17
2.2.2.1	Especificação do Modelo	p. 17
2.2.2.2	Estimação dos Parâmetros	p. 18
2.2.2.3	Qualidade do ajuste do modelo e significância das variáveis	p. 20
2.2.3	Razão de Chances	p. 22
2.2.4	Avaliação da capacidade preditiva do modelo	p. 24

3	Análise dos Resultados	p. 27
3.1	Análises Descritivas	p. 27
3.2	Regressão Logística - Ajuste e Avaliação dos Modelos	p. 31
4	Conclusão	p. 43
	Referências	p. 44

Lista de Figuras

1	Distribuição da frequência percentual do desfecho pelas características pessoais dos pacientes.	p. 29
2	Distribuição da frequência percentual do desfecho da internação pelas características do estabelecimento e de internação.	p. 31
3	Distribuição de frequência absoluta das métricas de desempenho preditivo do modelo 1 final.	p. 35
4	Distribuição de frequência absoluta das métricas de desempenho preditivo do modelo 2.	p. 37
5	Distribuição de frequência absoluta das métricas de desempenho preditivo do modelo 3.	p. 40
6	Distribuição de frequência absoluta das métricas de desempenho preditivo do modelo 4.	p. 42

Lista de Tabelas

1	Descrição das Variáveis	p. 15
2	Distribuição de frequências absoluta e percentual dos pacientes segundo suas características, de forma geral e de acordo com o desfecho ao final da internação – Rio de Janeiro, 2019	p. 28
3	Distribuição de frequências absoluta e percentual dos pacientes segundo características do esta belecimento de saúde e da internação, de forma geral e de acordo com o desfecho ao final da internação – Rio de Janeiro, 2019	p. 29
4	Resultados da regressão logística – modelo 1 – variáveis regressoras: sexo, faixa etária, município do hospital, natureza jurídica do hospital, peregrinação do paciente e diagnóstico principal	p. 33
5	Resultados da regressão logística – modelo 2 – variáveis regressoras: faixa etária, município do hospital, natureza jurídica do hospital, peregrinação do paciente, diagnóstico principal e utilização de UTI	p. 36
6	Resultados da regressão logística – modelo 3 – variáveis regressoras: sexo, faixa etária, município do hospital, natureza jurídica do hospital, peregrinação do paciente e diagnóstico principal	p. 39
7	Resultados da regressão logística – modelo 4 – variáveis regressoras: faixa etária, natureza jurídica do hospital, peregrinação do paciente, diagnóstico principal e utilização de UTI	p. 41

1 Introdução

1.1 Motivação

No ano de 2021, o mundo ainda tentava se recuperar da maior pandemia de nossa história recente, com o Brasil já ultrapassando em dezembro de 2021 as 616 mil mortes (Painel Corona Vírus Brasil, 2021), porém caminhando a passos largos para a volta à vida normal com o sucesso e a eficácia das vacinas distribuídas pelo Sistema Único de Saúde. No ano de 2020, a carência de leitos e insumos hospitalares e a alta propagação da doença foram alvo de discussões, porém há de se perceber que nenhum sistema de saúde público no mundo estaria preparado para atender a uma demanda de uma eventual pandemia. Até o SUS Brasil, que nas últimas décadas aumentou expressivamente a sua rede assistencial e capacidade de atendimentos (VIACAVA et al., 2018), não teve os recursos e material humano necessários para controlar o cenário.

A pneumonia era uma das mais graves doenças respiratórias que atingiam idosos antes da pandemia do Covid-19, podendo ser classificada como sendo do tipo viral, bacteriana ou química. A Pneumonia (OMS, 2021) é um tipo de infecção respiratória aguda que afeta o pulmão, causando dificuldade na respiração e limitando a quantidade de oxigênio inalada. De acordo com o Ministério da Saúde (MS) esta doença acomete muitos idosos, tendo sido, entre os anos 2015 e 2017, a *causa mortis* de cerca de 160 mil pessoas com idade acima de 60 anos. Além disso, 80% das pessoas neste período que foram a óbito devido à pneumonia eram idosos (Ministério da Saúde, 2019).

Em relação à Covid, em maio de 2020 cerca de 69,2% das mortes por esta enfermidade no Brasil foram referentes a idosos. No mês de janeiro de 2021 este percentual já alcançava a marca de 74,2% (Boletins Epidemiológicos, 2021), evidenciando a maior vulnerabilidade da população idosa, mas ainda sem ser atestado que os fatores associados ao óbito seriam os mesmos daqueles relativos aos casos de doenças respiratórias ocorridos antes da pandemia.

A relevância de um tema tão polêmico desperta naturalmente a curiosidade de identificar como era o perfil das internações emergenciais em idosos, acometidos de doenças respiratórias, antes da disseminação do novo Corona Vírus.

1.2 Revisão Bibliográfica

Na literatura, podemos observar diversas pesquisas relacionadas a como doenças respiratórias, principalmente a pneumonia, têm agido nos idosos. Foi estudado por Francisco, Donalisio e Lattorre (2005) o impacto da vacinação contra influenza na mortalidade por doenças respiratórias em idosos. Essa investigação abordou as técnicas de Séries Temporais e Modelos de Regressão polinomial para avaliar as tendências de taxas de mortalidade, com base em informações referentes a idosos residentes no Estado de São Paulo. Foram utilizados dados do Sistema de Informações de Mortalidade do Ministério da Saúde (DATASUS) do período de 1980 a 2000 e foi encontrada uma taxa de óbito de 5,08 óbitos por mil homens e 3,18 óbitos por mil mulheres entre os anos de 1980 e 1998 para pessoas acima de 60 anos.

Em outro estudo (BONFADA et al., 2017) foi adotada a técnica de análise de sobrevivência a dados oriundos de 457 prontuários de idosos internados em Unidades de Terapia Intensiva na cidade de Natal, Rio Grande do Norte. Foi observado (utilizando o estimador de Kaplan-Meier para calcular as funções de sobrevivência e o modelo múltiplo de riscos proporcionais de Cox para a identificação do efeito independente dos preditores de sobrevivência) que a sobrevivência dos idosos internados em UTI de Natal é afetada por fatores prognósticos de origem demográfica, clínica, tipo de internação e rede de serviços de saúde.

No que se remete ao tema abordado neste texto, foi evidenciado que as doenças respiratórias são responsáveis por boa parte das internações no grupo de pessoas com mais de 60 anos (TEIXEIRA; BASTOS; SOUZA, 2017). Foi identificado, com base numa amostra de 186.202 internações de idosos ocorridas de 2008 a 2015, que em pessoas com mais de 80 anos, o grupo de doenças do aparelho respiratório é a segunda maior causa de internações, entre 70 e 79 é a 3^a e de 60 a 69 é a 5^a. Logo, pode ser observado que é um problema que atinge grande parte dos idosos.

Por último vale mencionar o estudo de Cacciatore et al. (2017) que identificou que nas hospitalizações em idosos a pneumonia é a 4^a maior causa de morte nesse grupo de idade e a principal causa de morte oriunda de doenças infecciosas. A taxa de mortalidade para pacientes que foram hospitalizados tendo como causa a pneumonia é preocupante, pois,

segundo o artigo, 20% dos idosos que passam por essa situação acabam morrendo e para pessoas desse grupo de idade que desenvolveram a doença enquanto estavam hospitalizados, essa taxa sobe para 50-70%, sendo maior nas mulheres. A revisão aqui apresentada indica que, apesar das doenças respiratórias serem causas frequentes de internação e morte em idosos, não foram relatados resultados descrevendo o comportamento do desfecho de internação emergencial de idosos acometidos de doenças respiratórias.

1.3 Objetivos

Este trabalho tem como objetivo principal investigar o perfil das internações emergenciais por doenças respiratórias em idosos residentes no estado do Rio de Janeiro no ano de 2019 e identificar possíveis fatores associados ao desfecho óbito, sendo os objetivos secundários:

- identificar o perfil das internações em relação às características pessoais dos idosos e das hospitalizações;
- avaliar a relação do desfecho (óbito/alta) com características dos idosos e das internações;
- investigar um modelo de regressão logística que descreva o comportamento do desfecho das internações em função das características pessoais dos idosos e das hospitalizações;
- averiguar a existência de fatores de risco que aumentem a chance do desfecho óbito.

1.4 Organização do Trabalho

Este Trabalho de Conclusão de Curso apresenta no Capítulo 1 a motivação do estudo, a revisão bibliográfica de artigos relacionados ao tema de interesse e os objetivos geral e específicos estabelecidos na investigação. No Capítulo 2 são apresentadas informações sobre os dados a serem considerados e a metodologia estatística que será adotada. O Capítulo 3 apresenta as análises dos resultados e no quarto Capítulo estão registradas as principais conclusões obtidas.

2 Materiais e Métodos

2.1 Dados

Para a realização deste estudo, foram considerados dados do Sistema de Informação Hospitalar do DATASUS (DATASUS/SIHSUS) no ano de 2019 no estado do Rio de Janeiro. Os dados são restritos a internações emergenciais de pessoas idosas cuja causa da internação possa ser classificada de acordo com o código internacional de doenças (CID-10) em uma das doenças a seguir, as quais pertencem ao conjunto de doenças do aparelho respiratório (J00-J99).

- I) J00-J06 - Infecções agudas das vias aéreas superiores
- II) J09-J11 - Influenza (gripe)
- III) J12-J18 - Pneumonia
- IV) J20-J22 - Outras infecções agudas das vias aéreas inferiores
- V) J30-J39 - Outras doenças das vias aéreas superiores
- VI) J40-J47 - Doenças crônicas das vias aéreas inferiores
- VII) J60-J70 - Doenças pulmonares devidas a agentes externos
- VIII) J80-J84 - Outras doenças respiratórias que afetam principalmente o interstício
- IX) J85-J86 - Afecções necróticas e supurativas das vias aéreas inferiores
- X) J90-J94 - Outras doenças da pleura
- XI) J95-J99 - Outras doenças do aparelho respiratório

Neste Trabalho de Conclusão de Curso, foram consideradas 15 variáveis, apresentadas no Tabela 1 com destaque para a variável resposta "MORTE", que descreve o desfecho da internação.

Tabela 1: Descrição das Variáveis

Variável	Descrição	Categorias/Valores Originais	Categorias/Valores do Trabalho
MUNIC.RES	Município de Residência	Município de Residência	Município do Rio de Janeiro Outro município
SEXO	Sexo do paciente	Masculino / Feminino	Masculino / Feminino
UTLMES.TO	Quantidade de dias de UTI no mês	0,1,2,...	0 / Maior que 0
DIAG.PRINC	Diagnóstico Principal	Código CID-10	Pneumonia Outra doença respiratória
MORTE	Desfecho da internação	Morte Não Morte (Transferência / Alta)	Morte / Alta
CAR.INT	Caráter da Internação	Caráter da Internação	Internações Emergenciais
RACACOR	Raça/Cor	Branca / Preta / Parda / Amarela / Indígena	Branca / Não Branca
NATUREZA	Tipo de vínculo da unidade hospitalar com o SUS	Administração Pública Entidades empresariais Entidades sem fins lucrativos	Administração Pública Privada
MUNIC.MOV	Município do Estabelecimento	Município do Estabelecimento	Município do Rio de Janeiro Outro município
COD.IDADE	Unidade de medida da idade	Horas / Dias / Meses / Anos completos / Outra	Anos completos
IDADE	Idade	Idade	60 a 69 anos 70 a 79 anos Acima de 80 anos

Deve ser esclarecido que originalmente a variável COD_IDADE tem 5 categorias, referentes à unidade de medida da idade, que poderia ser horas, dias, meses, anos completos e outra, sendo esta última aplicável apenas para pacientes acima de 100 anos. A existência dessas 5 categorias acontece porque havia no banco de dados original pacientes bebês, que poderiam ter nascido há poucas horas, dias ou meses, bem como indivíduos com pelo menos 1 ano completo de vida. Foi considerada apenas a categoria anos completos, uma vez que o estudo avalia somente idosos (60 anos completos ou mais). A posterior categorização indicada no Tabela 1 é usual na divulgação de estatísticas sobre hospitalizações divulgadas no DATASUS.

A partir das variáveis originais MUNIC_RES e MUNIC_MOV foi possível criar a variável categórica chamada PEREGRINAÇÃO, que indica se a pessoa foi atendida num hospital no seu município de residência ou não.

2.2 Metodologia

2.2.1 Modelos Lineares Generalizados

2.2.1.1 Introdução

O modelo de regressão linear tem como principal objetivo identificar relações entre a variável resposta e as variáveis explicativas. Entretanto, um dos pressupostos para a utilização desta técnica envolve a distribuição Normal. Os modelos lineares generalizados entram com uma maneira de expandir as possibilidades de distribuições envolvidas nos pressupostos, pois admitem outras distribuições além da Normal.

Esses modelos (NELDER; WEDDERBURN, 1972) possuem uma variável aleatória Y de interesse chamada resposta ou dependente, pressuposta como tendo uma distribuição pertencente à família exponencial, além de k variáveis explicativas, denotadas por x_1, \dots, x_k , que também são chamadas de covariáveis ou variáveis independentes que tentam explicar a variabilidade de Y .

2.2.1.2 Componentes

Os modelos lineares generalizados, uma extensão do modelo linear clássico, admite que a variável resposta Y tenha distribuição de probabilidades qualquer pertencente a família exponencial e além disso, apesar de manter a linearidade, a função que relaciona o valor esperado e o vetor de covariáveis pode ser qualquer função diferenciável (TURKMAN; SILVA, 2000).

Os modelos lineares generalizados possuem as seguintes componentes:

i) Componente aleatória

A componente aleatória de um modelo linear generalizado é composta por um conjunto de variáveis respostas independentes e identicamente distribuídas Y_1, \dots, Y_n , sendo esta distribuição pertencente à família exponencial e que possuem $E(Y_i | \mathbf{x}_i) = \mu_i = b'(\theta_i)$.

ii) Componente sistemática

A componente sistemática está relacionada com o preditor linear $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$, no qual $\mathbf{x}_i^T = (1, x_{1i}, \dots, x_{ki})$ é o vetor que contém as k variáveis explicativas (ou covariáveis) presentes no modelo referente a i -ésima observação, enquanto $\boldsymbol{\beta}$, denominado vetor de

parâmetros do modelo, pode ser expresso por:

$$\boldsymbol{\beta}^T = (\beta_0, \beta_1, \beta_2, \dots, \beta_k) \quad (2.1)$$

As variáveis explicativas associadas ao modelo podem ser divididas em variáveis aleatórias quantitativas e categóricas. As variáveis categóricas que possuem c categorias são separadas em uma categoria de referência para o modelo e $c - 1$ variáveis binárias chamadas de *dummies*. Elas são uma forma de representação das variáveis qualitativas presentes nas análises as quais representam a presença ou ausência de determinada característica.

iii) Função de ligação

A função de ligação de um modelo linear generalizado é uma função monótona e diferenciável que associa a média da componente aleatória (μ_i) à componente sistemática (preditor linear do modelo) pela seguinte expressão:

$$g(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$$

2.2.2 Regressão Logística

2.2.2.1 Especificação do Modelo

A regressão logística é um dos principais modelos utilizados quando o problema consiste em averiguar como um conjunto de variáveis independentes podem explicar o comportamento de uma variável resposta binária (DOBSON; BARNETT, 2018). O modelo de regressão logística tem como suas principais vantagens ser fácil de usar e muito flexível do ponto de vista matemático, além de apresentar interpretações simples para os parâmetros. Existe no caso então uma variável Y que segue uma distribuição de Bernoulli (π) e haverá uma função de ligação que relacione a média π com as variáveis explicativas.

Em relação à notação, é utilizado $\pi(\mathbf{x}) = E(Y/\mathbf{x})$ para indicar a esperança condicional de Y dado \mathbf{x} quando trabalhamos com a regressão logística. O modelo de regressão logística é dado da seguinte forma:

$$\pi(\mathbf{x}) = \frac{\exp(\mathbf{x}^T \boldsymbol{\beta})}{1 + \exp(\mathbf{x}^T \boldsymbol{\beta})}$$

em que, conforme mencionado anteriormente, \mathbf{x} é o vetor que contém as covariáveis e $\boldsymbol{\beta}$

o vetor de parâmetros do modelo.

O modelo também pode ser apresentado através da função de ligação *logit* abaixo, em que o logaritmo da fração entre os termos fornece um modelo linear:

$$\ln \left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right] = \mathbf{x}^T \boldsymbol{\beta} \quad (2.2)$$

2.2.2.2 Estimação dos Parâmetros

Para estimar os parâmetros do modelo em que a variável resposta é dicotômica é utilizado o método de máxima verossimilhança. Nesse método as estimativas dos parâmetros são obtidas de forma a maximizar a função de verossimilhança. A função de verossimilhança, no caso de uma variável aleatória discreta Y com função de probabilidade $p(y|\theta)$, baseada em uma amostra de n observações independentes e identicamente distribuídas é definida por:

$$l(\theta) = \prod_{i=1}^n p(y_i/\theta) \quad (2.3)$$

Como no caso da regressão logística Y é uma variável resposta binária com 0 ou 1 como valores possíveis, ou seja, segue uma distribuição de Bernoulli com probabilidade de sucesso $\pi(\mathbf{x})$, então a função de verossimilhança, considerando n observações independentes e identicamente distribuídas, será dada por:

$$l(\boldsymbol{\beta}) = \prod_{i=1}^n \pi(\mathbf{x}_i)^{y_i} [1 - \pi(\mathbf{x}_i)]^{1-y_i} \quad (2.4)$$

$$l(\boldsymbol{\beta}) = \prod_{i=1}^n \left(\frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta})} \right)^{y_i} \left[1 - \left(\frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta})} \right) \right]^{1-y_i}$$

Para facilitar a obtenção dos valores de β_j , com j variando de 0, 1, ..., k que maximizam a função em 2.4, usualmente é aplicada função logaritmo (\ln). Assim, o log da função de verossimilhança é dado por:

$$L(\boldsymbol{\beta}) = \ln [l(\boldsymbol{\beta})] = \sum_{i=1}^n \{y_i \ln[\pi(\mathbf{x}_i)] + (1 - y_i) \ln[1 - \pi(\mathbf{x}_i)]\} \quad (2.5)$$

$$L(\boldsymbol{\beta}) = \sum_{i=1}^n \left\{ y_i \ln \left[\frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta})} \right] + (1 - y_i) \ln \left[1 - \frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta})} \right] \right\}$$

A obtenção de valores de máximo dos parâmetros do modelo recai em um sistema de $k + 1$ equações onde, em cada uma delas, será calculada a derivada parcial do logaritmo da função de verossimilhança (2.5) em relação ao um particular β_j e igualada a zero, ou seja,

$$\begin{cases} \frac{\delta L(\boldsymbol{\beta})}{\delta \beta_0} = 0 \\ \frac{\delta L(\boldsymbol{\beta})}{\delta \beta_1} = 0 \\ \frac{\delta L(\boldsymbol{\beta})}{\delta \beta_2} = 0 \\ \dots \\ \frac{\delta L(\boldsymbol{\beta})}{\delta \beta_k} = 0 \end{cases} \quad (2.6)$$

Além disso, para que sejam obtidos os valores que maximizam a função de verossimilhança, deve-se descobrir as expressões das derivadas parciais de segunda ordem do logaritmo da função de verossimilhança. As expressões decorrentes desse procedimento não são lineares e a solução deve ser obtida por meio de métodos iterativos, como o de Newton-Rapson já implementados em diversos *softwares*. Uma vez obtidas as estimativas dos parâmetros do modelo, a estimativa pontual da probabilidade da resposta de interesse ($Y = 1$), consideradas as variáveis explicativas, é dada por:

$$\hat{\pi}(\mathbf{x}) = \frac{\exp(\mathbf{x}^T \hat{\boldsymbol{\beta}})}{1 + \exp(\mathbf{x}^T \hat{\boldsymbol{\beta}})} \quad (2.7)$$

2.2.2.3 Qualidade do ajuste do modelo e significância das variáveis

Para avaliar a qualidade do ajuste do modelo, basicamente existem três etapas a serem consideradas. Em um primeiro momento deve-se avaliar se o modelo de forma global é satisfatório ou se existe alguma indicação da existência de pelo menos uma variável não relevante. Para tanto será considerada a estatística *deviance* (D) (JR; LEMESHOW; STURDIVANT, 2013). Esta estatística compara as verossimilhanças dos modelos ajustado e saturado (número de parâmetros igual ao número de observações, n) e é dada pela seguinte expressão:

$$D = -2\ln\left(\frac{\text{verossimilhança do modelo ajustado}}{\text{verossimilhança do modelo saturado}}\right) \quad (2.8)$$

com D , tendo assintoticamente uma distribuição Qui-Quadrado com $n - (p + 1)$ graus de liberdade. O *deviance* tem papel parecido com o coeficiente de determinação (R^2) presente na regressão linear, mas para sua interpretação, diferentemente do R^2 , quanto menor o valor de D melhor a qualidade do ajuste do modelo. Para tirar conclusões sobre a qualidade do ajuste, é essencial a utilização do operador $-2\ln$ na equação 2.8 para que se possa trabalhar com uma distribuição de probabilidades conhecida no teste de hipóteses sobre a significância conjunta das variáveis consideradas no modelo. O teste, no caso da regressão logística e através de 2.3 e 2.7, se baseará na estatística D , expressa por:

$$D = -2 \sum_{i=1}^n \left[y_i \ln\left(\frac{\hat{\pi}_i}{y_i}\right) + (1 - y_i) \ln\left(\frac{1 - \hat{\pi}_i}{1 - y_i}\right) \right] \quad (2.9)$$

sendo $\hat{\pi}_i = \hat{\pi}(x_i)$

Como os resultados possíveis para a variável resposta são 0 ou 1, a verossimilhança do modelo saturado é 1, dado que $\hat{\pi}_i = y_i$. Logo, a *deviance* pode ser escrita como:

$$D = -2\ln(\text{verossimilhança do modelo ajustado}) \quad (2.10)$$

Em um segundo passo para a verificação da qualidade do modelo, a significância de cada uma das variáveis deverá ser avaliada, através do teste de Wald. Ele testa se cada parâmetro do modelo pode ser considerado nulo e a estatística de teste compara a estimativa de máxima verossimilhança $\hat{\beta}_j$ (que para grandes amostras, segue aproximadamente

uma distribuição Normal com média β_j e variância $Var(\hat{\beta}_j)$ com o erro padrão desta estimativa. Com isso, as hipóteses do teste são definidas da seguinte forma:

$$H_0 : \beta_j = 0 \quad (2.11)$$

$$H_1 : \beta_j \neq 0 \quad (2.12)$$

E a estatística de teste, W , dada por:

$$W = \frac{\hat{\beta}_j}{EP(\hat{\beta}_j)} \quad (2.13)$$

Para a regra de decisão do teste, será dito que a variável associada ao parâmetro em questão é relevante no modelo se o *p-valor* for inferior ao nível de significância α onde:

$$p - valor = P(|W| > W_{obs} | H_0 \text{ verdadeiro}) \quad (2.14)$$

Finalmente a última etapa de qualidade de ajuste do modelo poderá ser considerada, caso haja necessidade de comparar modelos distintos, aqui denotados por \mathbf{M}_1 e \mathbf{M}_2 , onde o segundo tem mais variáveis que o primeiro e além disso todas as variáveis de \mathbf{M}_1 estejam presentes também em \mathbf{M}_2 . Nesta etapa, o teste denominado razão de verossimilhança é utilizado através da estatística G , definida por:

$$G = -2\ln \left(\frac{\text{verossimilhança sem as variáveis}}{\text{verossimilhança com as variáveis}} \right) \quad (2.15)$$

$$G = -2\ln \left(\frac{\text{verossimilhança de } \mathbf{M}_1}{\text{verossimilhança de } \mathbf{M}_2} \right)$$

A estatística G segue distribuição Qui-Quadrado com g graus de liberdade, que representa a diferença no número de variáveis presentes em \mathbf{M}_2 e \mathbf{M}_1 . Assim, esta estatística avalia a significância conjunta das variáveis que fazem parte de \mathbf{M}_2 mas que não estão em \mathbf{M}_1 . Portanto G será a estatística de teste para avaliar as hipóteses:

$$H_0 : \beta_i = \beta_j = \dots = 0, \text{ sendo } i, j, \dots \text{ as variáveis testadas} \quad (2.16)$$

$$H_1 : \text{pelo menos um dos coeficientes é não nulo} \quad (2.17)$$

A regra de decisão envolverá novamente o *p-valor*, que é expresso por:

$$p - \text{valor} = P(|G| > G_{obs} | H_0 \text{ verdadeiro}) \quad (2.18)$$

Com isso, sendo o *p-valor* menor que o nível de significância estabelecido, há evidências que pelo menos uma das variáveis testadas é significativa no modelo. Por outro lado, caso o *p-valor* seja maior que α , a hipótese nula não poderá ser rejeitada, indicando que as variáveis explicativas em questão não são relevantes para o modelo. Por tudo que foi mencionado, não é difícil perceber que o teste de razão de verossimilhança baseado na estatística G pode ser usado para avaliar, de forma alternativa ao teste de Wald, a significância de cada uma das variáveis do modelo. Neste caso, a diferença no número de variáveis de M_1 e M_2 seria 1.

2.2.3 Razão de Chances

A razão de chances é uma medida associativa entre uma condição e o desfecho de interesse ($Y = 1$) que representa a chance de que o desfecho de interesse ocorra dada uma condição particular, comparado com a chance deste resultado ocorrer dada a ausência dessa condição particular.

Dessa forma, no contexto de regressão logística, onde a probabilidade de ocorrer o evento de interesse é função das variáveis explicativas do modelo, a chance é expressa por:

$$\text{chance} = \frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \quad (2.19)$$

A razão de chances (RC) entre dois grupos (1 e 2) é definida por:

$$RC = \frac{\text{chance}_1}{\text{chance}_2} \quad (2.20)$$

onde os grupos 1 e 2 diferem apenas quanto à presença e ausência de uma determinada característica.

Para o modelo logístico linear simples (MENARD, 2002), por exemplo, temos que a probabilidade de sucesso tem relação com a variável explicativa x da seguinte forma:

$$\ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x \quad (2.21)$$

sendo β_0 e β_1 parâmetros desconhecidos.

Considerando que x é uma variável categórica, vamos atribuir o valor $x = 1$ aos indivíduos com a presença de determinado fator e $x = 0$, àqueles com ausência do fator. Dessa maneira, a chance de ocorrer sucesso na variável resposta para um indivíduo com a presença do fator é:

$$\frac{\pi(1)}{1 - \pi(1)} = e^{\beta_0 + \beta_1 x} \quad (2.22)$$

e por outro lado, a chance de ocorrer sucesso na variável resposta para um indivíduo com ausência do fator é:

$$\frac{\pi(0)}{1 - \pi(0)} = e^{\beta_0} \quad (2.23)$$

Portanto, a razão de chances é expressa da seguinte maneira:

$$RC = \frac{\frac{\pi(1)}{1 - \pi(1)}}{\frac{\pi(0)}{1 - \pi(0)}} = e^{\beta_1} \quad (2.24)$$

isto é, dependendo apenas do parâmetro β_1 .

O resultado em 2.24 é válido quando há várias variáveis explicativas no modelo, sendo elas categóricas ou não. Dessa forma, identifica-se que, quando ajustamos o modelo de regressão logística, $exp(\beta_j)$ para $j = 1, \dots, k$ representam as razões de chance associadas a cada uma das k variáveis do modelo, cujas estimativas pontuais são $exp(\hat{\beta}_j)$.

A razão de chances é frequentemente usada para identificar quando uma condição particular é um fator de risco para um certo desfecho de interesse e também para comparar a magnitude de diversos fatores de risco para aquele evento. Sendo assim, se:

$RC = 1$ a condição particular não afeta a chance de um desfecho

$RC > 1$ a condição particular associada é denominada de fator de risco para o desfecho

$RC < 1$ a condição particular associada é denominada de fator de proteção para o desfecho

Percebemos que é mais fácil olhar para a razão de chances, no que diz respeito à interpretação da influência que as variáveis explicativas têm em relação à variável resposta, do que observar apenas as estimativas pontuais dos parâmetros β_j . Portanto, nada mais razoável e adequado que se obter também as estimativas das razões de chances por intervalos de confiança (IC), cujos limites são dados por:

$$\exp \left(\hat{\beta}_j \pm z_{1-\alpha/2} \sqrt{\widehat{Var}(\hat{\beta}_j)} \right) \quad (2.25)$$

considerando-se que, para n grande, $\hat{\beta}_j$ tem distribuição amostral aproximadamente Normal com média β_j e variância dada por $\widehat{Var}(\hat{\beta}_j)$, sendo esta estimada por métodos iterativos. Além disso, $P(Z \geq z_\alpha = \alpha)$ onde Z tem distribuição normal padronizada

Intervalos de confiança de RC que incluam o valor 1 indicam que a variável em questão não representa nem um fator de risco ou de proteção para o fenômeno investigado, ou seja não afeta o desfecho. Dessa forma, a interpretação do IC de RC é uma forma alternativa de avaliar a relevância de uma variável no modelo.

2.2.4 Avaliação da capacidade preditiva do modelo

Obtida a expressão do modelo que melhor relaciona a variável dependente Y com as variáveis explicativas X 's é possível obter estimativas de Y para novos registros não observados previamente. Se a previsão se aproximar com o valor verdadeiro observado, dizemos que o modelo previu bem.

Na regressão logística, como mencionado anteriormente, trabalha-se com uma variável resposta binária, tendo como possíveis valores para esta variável 1 e 0, referentes à presença do desfecho de interesse ou ausência, respectivamente. Denota-se aqui as estimativas de Y como sendo positivo (+), se a previsão indicar a ocorrência do desfecho de interesse e negativo (-), caso a previsão não indique tal ocorrência.

A matriz de confusão é composta pelo número de acertos e o número de erros do modelo, tendo os acertos na diagonal principal e os erros na secundária. Há duas possibilidades de acertos do modelo, sendo eles os verdadeiros positivos e os verdadeiros negativos. O primeiro trata-se de quando $Y = 1$ e o valor predito é (+), já o segundo é o cenário em que $Y = 0$ e o valor predito é (-). Para os erros preditivos do modelo, tem-se os falsos positivos e falsos negativos. Os falsos negativos são encontrados quando $Y = 1$ e o valor predito é (-) e os falsos positivos quando o resultado da previsão é (+) e $Y = 0$. Com base nos acertos preditivos do modelo, podem ser definidas as seguintes medidas que caracterizam a qualidade preditiva de um modelo:

1) Acurácia (Taxa global de acerto de classificação): verifica a proporção de acertos globais do modelo e é expressa por:

$$\begin{aligned} \text{Acurácia} &= P(\text{verdadeiro positivo}) + P(\text{verdadeiro negativo}) \\ &= P(Y = 1 \cap +) + P(Y = 0 \cap -) \end{aligned} \quad (2.26)$$

2) Sensibilidade: verifica a proporção de acertos no grupo de indivíduos onde o desfecho de interesse foi observado, sendo expressa por:

$$\text{Sensibilidade} = P(+ | Y = 1) = \frac{P(\text{verdadeiro positivo})}{P(Y = 1)} \quad (2.27)$$

3) Especificidade: avalia a proporção de acertos no grupo de indivíduos onde o desfecho de interesse não foi observado, sendo sua expressão dada por:

$$\text{Especificidade} = P(- | Y = 0) = \frac{P(\text{verdadeiro negativo})}{P(Y = 0)} \quad (2.28)$$

Um modelo com boa capacidade preditiva deveria ter então acurácia, sensibilidade e especificidade de valores elevados. No entanto, é usual em pesquisas da área de saúde priorizar a sensibilidade ou a especificidade do modelo, dependendo das possíveis consequências dos erros preditivos.

Como forma de melhorar o desempenho do modelo, foi utilizada a curva ROC. A curva ROC é uma curva formada plotando-se no eixo y a sensibilidade e no eixo x 1-especificidade do modelo, considerando-se diversos valores de probabilidade a partir da qual um indivíduo será classificado como tendo $Y=1$. Tais valores são denominados pontos de corte. O ponto de corte de uma curva ROC geralmente é o ponto que maximiza a soma da especificidade e sensibilidade, porém foi utilizado neste trabalho o ponto de corte que promovesse valores de sensibilidade e especificidade elevados (no mínimo de 50%), visando um aumento na capacidade preditiva. Além disso, a área sob a curva ROC ajuda a comparar a capacidade preditiva de modelos distintos, uma vez que em geral admite-se que quanto maior é esta área, maior a capacidade preditiva alcançada.

Para viabilizar a aplicação da estatística deste estudo, não foi usada a totalidade dos dados para evitar problemas de *overfitting*, que se trata de quando um modelo se adequa perfeitamente aos dados observados, o que pode fazer com que, para novos dados a serem obtidos posteriormente, este modelo não tenha um bom desempenho preditivo. Portanto, os registros de internações foram separados em 2 conjuntos: treino e teste. O conjunto de treino engloba as informações referentes a 90% dos registros de internações, obtido de forma aleatória em cada um dos grupos da variável desfecho de interesse (óbito/alta).

Esse conjunto será utilizado na etapa de ajustes e seleção de modelos. Já o conjunto de teste, formado com os demais 10% dos registros, será utilizado para avaliar a capacidade preditiva do modelo selecionado como mais adequado para explicar o fenômeno de interesse.

3 Análise dos Resultados

Este trabalho de conclusão de curso visa identificar o perfil das internações emergenciais por doenças respiratórias em idosos no estado do Rio de Janeiro e possíveis fatores associados ao desfecho óbito, utilizando dados do Sistema de Informação Hospitalar do DATASUS no ano de 2019 no estado do Rio de Janeiro. O banco de dados inicial possuía 754.725 registros, mas após algumas restrições e filtros serem implementados como limitação a doenças respiratórias, somente internações emergências, faixas etárias para idades acima de 60 anos e exclusão de dados faltantes, a base final ficou estabelecida com 13.847 registros e 9 variáveis separadas em 2 grupos: características pessoais do paciente e características do estabelecimento de saúde e da internação. A variável "RACACOR" que a priori seria utilizada, foi excluída das análises pois possui quase 1/4 de seus valores sendo faltantes e, sendo assim. Logo não é interessante utilizá-la porque reduziria muito o número de informações da base de dados caso houvesse a exclusão desses registros, podendo causar a descaracterização de possíveis relações entre as variáveis.

3.1 Análises Descritivas

A análise descritiva dos dados fornecerá um panorama inicial do comportamento das variáveis consideradas neste estudo, sugerindo a identificação de possíveis associações com o óbito do paciente.

A Tabela 2 apresenta as distribuições de frequências das características dos pacientes, relacionadas com o desfecho da internação (óbito/alta). Como se pode observar, as mulheres foram maioria para os dois grupos de desfecho, porém não há uma dissimilaridade grande na quantidade de pessoas do sexo feminino e masculino. Em relação à faixa etária, percebe-se que pessoas com 80 ou mais anos foram maioria. Além disso, dentre os pacientes que foram a óbito, 44% referiam-se a idosos com 80 anos ou mais, enquanto para pacientes que receberam alta, este grupo representa um pouco menos de 34%, uma diferença considerável. Para o município de residência, pessoas que residiam

fora do Rio de Janeiro foram grande maioria considerando todas as categorias do desfecho da internação.

Tabela 2: Distribuição de frequências absoluta e percentual dos pacientes segundo suas características, de forma geral e de acordo com o desfecho ao final da internação – Rio de Janeiro, 2019

Característica	Geral (n = 13.847) n(%)	Óbito (n = 4.950) n(%)	Alta (n = 8.897) n(%)
Sexo			
Feminino	7.128 (51,5)	2.581 (52,1)	4.547 (51,1)
Masculino	6.719 (48,5)	2.369 (47,9)	4.350 (48,9)
Faixa Etária			
60 a 69	4.291 (31,0)	1.283 (25,9)	3.008 (33,8)
70 a 79	4.396 (31,7)	1.514 (30,6)	2.882 (32,4)
80 ou mais	5.160 (37,3)	2.153 (43,5)	3.007 (33,8)
Município de Residência			
Rio de Janeiro	2.834 (20,5)	1.182 (23,9)	1.652 (18,6)
Outro Município	11.013 (79,5)	3.768 (76,1)	7.245 (81,4)

No intuito de avaliar que fatores possam contribuir para o óbito do paciente, a Figura 1 ilustra a distribuição de frequência do desfecho da internação, segundo características dos idosos. Como se pode identificar, o comportamento do desfecho em homens e mulheres foi muito semelhante, o que pode sugerir que a variável sexo não seja relevante para explicar a morte do paciente. Já para a faixa etária, observa-se que quanto mais elevada a classe etária, maiores foram os percentuais de óbitos, chegando a quase 41% na faixa de 80 anos ou mais. Com relação ao município de residência, pode-se visualizar que pacientes que residiam fora do Rio de Janeiro tiveram mais o desfecho alta hospitalar que moradores da cidade do Rio.

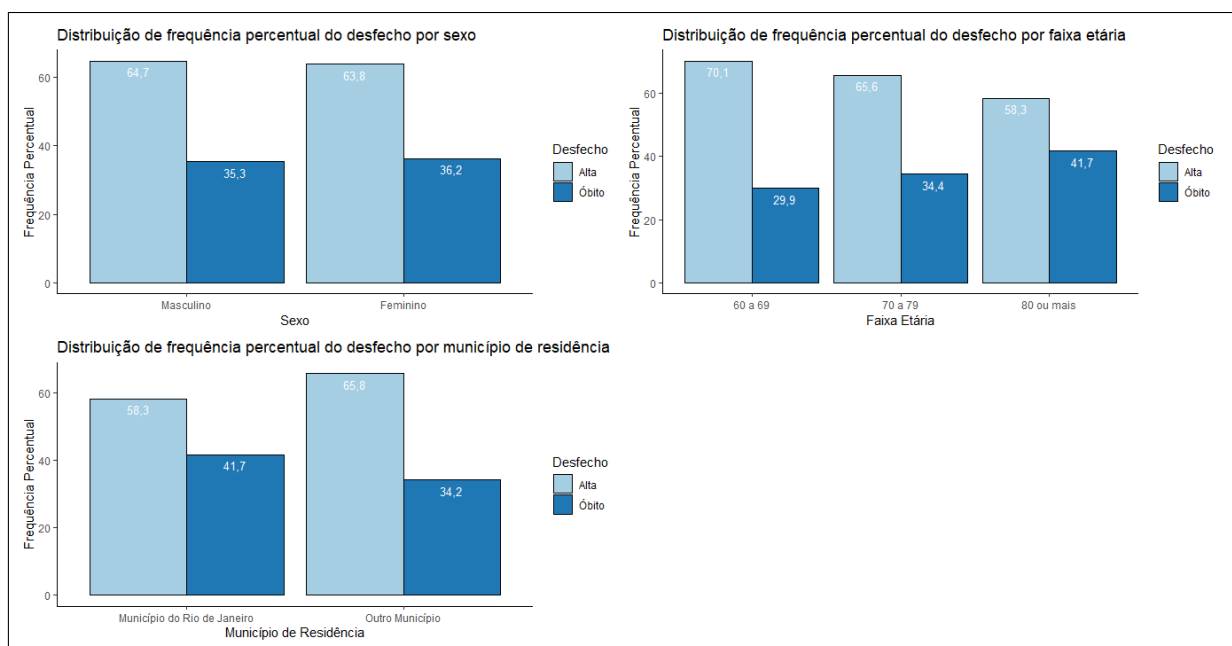


Figura 1: Distribuição da frequência percentual do desfecho pelas características pessoais dos pacientes.

A Tabela 3 exibe as distribuições de frequência das características do estabelecimento de saúde e da internação de modo geral e em relação a cada grupo de desfecho. Percebe-se que a grande maioria dos pacientes foi atendida fora da cidade do Rio de Janeiro. Além disso, a maior parte deles se consultaram em hospitais públicos, não precisaram se deslocar de município para serem atendidos, tiveram a pneumonia como diagnóstico principal da sua doença, não utilizaram a UTI e receberam alta como sendo o desfecho de sua internação.

Tabela 3: Distribuição de frequências absoluta e percentual dos pacientes segundo características do estabelecimento de saúde e da internação, de forma geral e de acordo com o desfecho ao final da internação – Rio de Janeiro, 2019

Característica	Geral ($n = 13.847$) $n(\%)$	Óbito ($n = 4.950$) $n(\%)$	Alta ($n = 8.897$) $n(\%)$
Município do Hospital			
Rio de Janeiro	3.108 (22,4)	1.263 (25,5)	1.845 (20,7)
Outro Município	10.739 (77,6)	3.687 (74,5)	7.052 (79,3)
Natureza jurídica do hospital			
Pública	10.592 (76,5)	4.142 (83,7)	6.450 (72,5)
Não Pública	3.255 (23,5)	808 (16,3)	2.447 (27,5)
Peregrinação do Paciente			
Sim	2.203 (15,9)	884 (17,9)	1.319 (14,8)
Não	11.644 (84,1)	4.066 (82,1)	7.578 (85,2)

Tabela 3: (Continuação)			
Diagnóstico Principal			
Grupo 1			
Pneumonia	9.111 (65,8)	3.049 (61,6)	6.062 (68,1)
Insuficiência Respiratória	1.028 (7,4)	857 (17,3)	171 (1,9)
Outra doença respiratória	3.708 (26,8)	1.044 (21,1)	2.664 (29,9)
Grupo 2			
III	9.111 (65,8)	3.049 (61,6)	6.062 (68,1)
I	81 (0,6)	16 (0,3)	65 (0,7)
II	137 (1,0)	38 (0,8)	99 (1,1)
IV	50 (0,4)	10 (0,2)	40 (0,4)
V	111 (0,8)	2 (0)	109 (1,2)
VI	1.384 (10,0)	317 (6,4)	1.067 (12,0)
VII	112 (0,8)	37 (0,7)	75 (0,8)
VIII	1.157 (8,4)	424 (8,6)	733 (8,2)
IX	45 (0,3)	12 (0,2)	33 (0,4)
X	522 (3,8)	166 (3,4)	356 (4,0)
XI	1.137 (8,2)	879 (17,8)	258 (2,9)
Utilização de UTI			
Sim	1.328 (9,6)	823 (16,6)	505 (5,7)
Não	12.519 (90,4)	4.127 (83,4)	8.392 (94,3)
Desfecho			
Alta	8.897 (64,3)	0 (00,0)	8.897 (100,0)
Óbito	4.950 (35,7)	4.950 (100,0)	0 (00,0)

Para identificar o perfil do desfecho em relação às características do estabelecimento e internação, a Figura 2 ilustra as distribuições de frequências de morte/alta para estas características. Como se pode observar, pacientes que foram atendidos em hospitais fora do Rio de Janeiro tiveram um percentual de óbito menor do que aqueles que foram atendidos na capital do estado. Percebe-se que há uma grande disparidade na porcentagem de óbitos se compararmos hospitais públicos e privados, já que 39,1% das pessoas que foram internadas em hospitais públicos acabaram morrendo, enquanto para hospitais particulares, esse número cai para 24,8%. Percebe-se para a peregrinação uma diferença considerável na porcentagem de óbitos entre quem foi internado na sua cidade de residência e quem não foi, o que sugere que o deslocamento do paciente até o hospital pode influenciar nesse desfecho. Em relação ao diagnóstico principal, para o grupo 1, insuficiência respiratória aguda possui uma porcentagem de óbitos muito maior que as outras duas categorias, enquanto que para o grupo 2, outras doenças respiratórias têm a maior porcentagem de mortes. Para o uso de UTI, pode-se observar que o óbito foi bem mais frequente para pacientes que utilizaram a mesma em relação aos que não usaram.

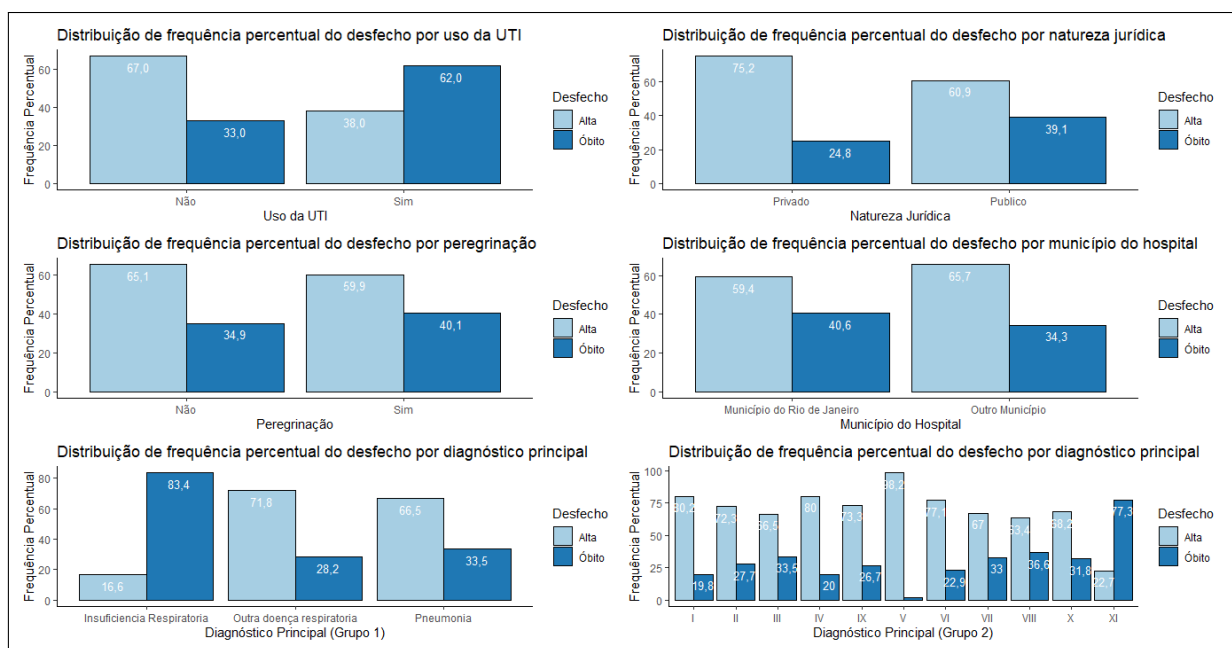


Figura 2: Distribuição da frequência percentual do desfecho da internação pelas características do estabelecimento e de internação.

3.2 Regressão Logística - Ajuste e Avaliação dos Modelos

Nesta etapa da análise será verificado como as variáveis referentes a características do paciente, do hospital e da internação podem conjuntamente influenciar a chance do óbito do idoso e quais são os principais fatores risco associados a esse desfecho. Para tanto são considerados 4 modelos principais, que diferem entre si em relação às variáveis regressoras que serão contempladas. Para definição das categorias de referência dos modelos, foi utilizada a de menor porcentagem de óbito dentre todas as categorias da mesma variável com exceção da variável diagnóstico principal nos modelos que incluem o grupo 2 de doenças, em que a categoria III (pneumonia) foi utilizada como referência para destacar sua relação com as demais enfermidades. Portanto as categorias de referência nos modelos são sexo masculino, classe etária de 60 a 69 anos, hospitais privados, localização do hospital fora do Rio, não haver peregrinação e diagnóstico principal de pneumonia para os modelos do grupo 2 e de outra doença respiratória para os do grupo 1. É oportuno lembrar para facilitar a interpretação dos resultados que as categorias de diagnósticos principais do grupo 1 são dadas por:

- I) Infecções agudas das vias aéreas superiores
- II) Influenza (gripe)
- III) Pneumonia
- IV) Outras infecções agudas das vias aéreas inferiores
- V) Outras doenças das vias aéreas superiores
- VI) Doenças crônicas das vias aéreas inferiores
- VII) Doenças pulmonares devidas a agentes externos
- VIII) Outras doenças respiratórias que afetam principalmente o interstício
- IX) Afecções necróticas e supurativas das vias aéreas inferiores
- X) Outras doenças da pleura
- XI) Outras doenças do aparelho respiratório

A Tabela 4 exibe as variáveis regressoras, estimativas dos coeficientes, resultado do teste de Wald e estimativas de razão de chances ajustadas (com intervalo de 95% de confiança) para o modelo inicialmente ajustado (com todas as variáveis regressoras) e para o modelo final (com somente variáveis relevantes). Estatísticas de desempenho do modelo final também são apresentadas. Não é difícil identificar no modelo final que a chance de óbito aumenta conforme o idoso pertence a uma classe etária mais avançada, sendo 84% maior para aqueles com 80 anos ou mais, comparativamente aos pacientes de 60 a 69 anos. Também pode ser identificado que a chance de óbito é 81% maior para idosos que foram atendidos em hospitais públicos, 20% mais elevada para aqueles que tiveram que peregrinar para serem internados e 12% maior para as internações ocorridas fora do Rio, comparativamente às respectivas categorias de referência. Em relação ao diagnóstico principal, pacientes com outras doenças das vias aéreas superiores e doenças crônicas das vias aéreas inferiores têm 96% e 36% de chance menor de morte que aqueles com pneumonia, respectivamente. Já pacientes com outras doenças respiratórias que afetam principalmente o interstício (VIII) e com outras doenças do aparelho respiratório (XI), têm 26% e 168% de chances de virem a óbito maior que aqueles com pneumonia.

Tabela 4: Resultados da regressão logística – modelo 1 – variáveis regressoras: sexo, faixa etária, município do hospital, natureza jurídica do hospital, peregrinação do paciente e diagnóstico principal

Variáveis do modelo	Estimativa dos coeficientes	p-valor do teste Wald	RC (IC95%)
-Modelo inicial: com todas as variáveis regressoras-			
Intercepto	-1,52291	< 0,001	0,22 (0,19 ; 0,25)
Sexo Masculino(ref) Feminino	-0,05719	0,15137	0,94 (0,87 ; 1,02)
Faixa etária 60 a 69 (ref) 70 a 79 80 ou mais	0,25361 0,58868	< 0,001 < 0,001	1,29 (1,16 ; 1,42) 1,80 (1,63 ; 1,99)
Município do Hospital Outro Município(ref) Rio de Janeiro	0,09875	0,04363	1,10 (1,01 ; 1,21)
Natureza jurídica do hospital Privado(ref) Pública	0,60308	< 0,001	1,83 (1,65 ; 2,02)
Peregrinação do Paciente Não(ref) Sim	0,22569	< 0,001	1,25 (1,13 ; 1,39)
Diagnóstico Principal III(ref) I II IV V VI VII VIII IX X XI	-0,54515 -0,27664 -0,60084 -3,19159 -0,44614 0,03261 0,22873 -0,32653 0,05300 1,98356	0,07166 0,16691 0,10996 < 0,001 < 0,001 0,87680 0,00106 0,37855 0,61001 < 0,001	0,58 (0,31 ; 1,02) 0,76 (0,51 ; 1,11) 0,55 (0,25 ; 1,10) 0,04 (0,01 ; 0,13) 0,64 (0,55 ; 0,74) 1,03 (0,68 ; 1,55) 1,26 (1,10 ; 1,44) 0,72 (0,33 ; 1,44) 1,05 (0,86 ; 1,29) 7,27 (6,22 ; 8,52)
-Características gerais: Deviance = 14976,62 (p-valor \cong 1)-			
-Modelo final: sem as variáveis regressoras não significativas-			
Intercepto	-1,56215	< 0,001	0,21 (0,19 ; 0,23)
Faixa etária 60 a 69 (ref) 70 a 79 80 ou mais	0,26723 0,60990	< 0,001 < 0,001	1,31 (1,18 ; 1,44) 1,84 (1,67 ; 2,02)
Município do Hospital Outro Município(ref) Rio de Janeiro	0,11508	0,015747	1,12 (1,02 ; 1,23)
Natureza jurídica do hospital Privado(ref) Pública	0,60157	< 0,001	1,82 (1,65 ; 2,02)
Peregrinação do Paciente Não(ref) Sim	0,18039	0,000589	1,20 (1,08 ; 1,33)

Tabela 4: (Continuação)			
Diagnóstico Principal			
III(ref)			
V	-3,17103	< 0,001	0,04 (0,01 ; 0,13)
VI	-0,44145	< 0,001	0,64 (0,56 ; 0,74)
VIII	0,23078	< 0,001	1,26 (1,10 ; 1,44)
XI	0,98734	< 0,001	2,68 (2,42 ; 2,98)

—Características gerais: *Deviance* = 15.409,63 (*p*-valor $\cong 1$) / *G*=433,01 (*p*-valor < 0,001)/ Ponto de corte = 0,355
 Área sob a curva ROC = 0,644

A estatística *deviance* do modelo final indica uma qualidade de ajuste relevante (*p*-valor > 0,05), permitindo então que o modelo possa ser utilizado para realizar previsões de novas admissões de internação. A curva ROC indicou o ponto de corte para a classificação de uma caso de óbito como sendo a probabilidade de 0,355. Para avaliar então a capacidade preditiva do modelo foi utilizado o conjunto de teste. A partir deste conjunto foram selecionadas aleatoriamente 100 amostras de tamanho 50, mantendo as proporções originais do desfecho da internação. A Figura 3 exibe as distribuições de frequências da acurácia, especificidade e sensibilidade. Como pode ser observado, a média de todas as métricas (linha pontilhada) é acima de 60% e os comportamentos da especificidade e da sensibilidade parecem simétricos. Os primeiros quartis das sensibilidades, especificidades e acurácias são respectivamente, 0,5625, 0,6111 e 0,6000, indicando que apenas 25% das amostras tiveram sensibilidade, especificidade e acurácia inferior a 56,25%, 61,11% e 60,00%. Além disso, foi verificado que das 100 amostras avaliadas 89% delas obtiveram simultaneamente sensibilidade e especificidade acima de 50%, mas apenas 43% alcançaram ambas as estatísticas acima de 60%.

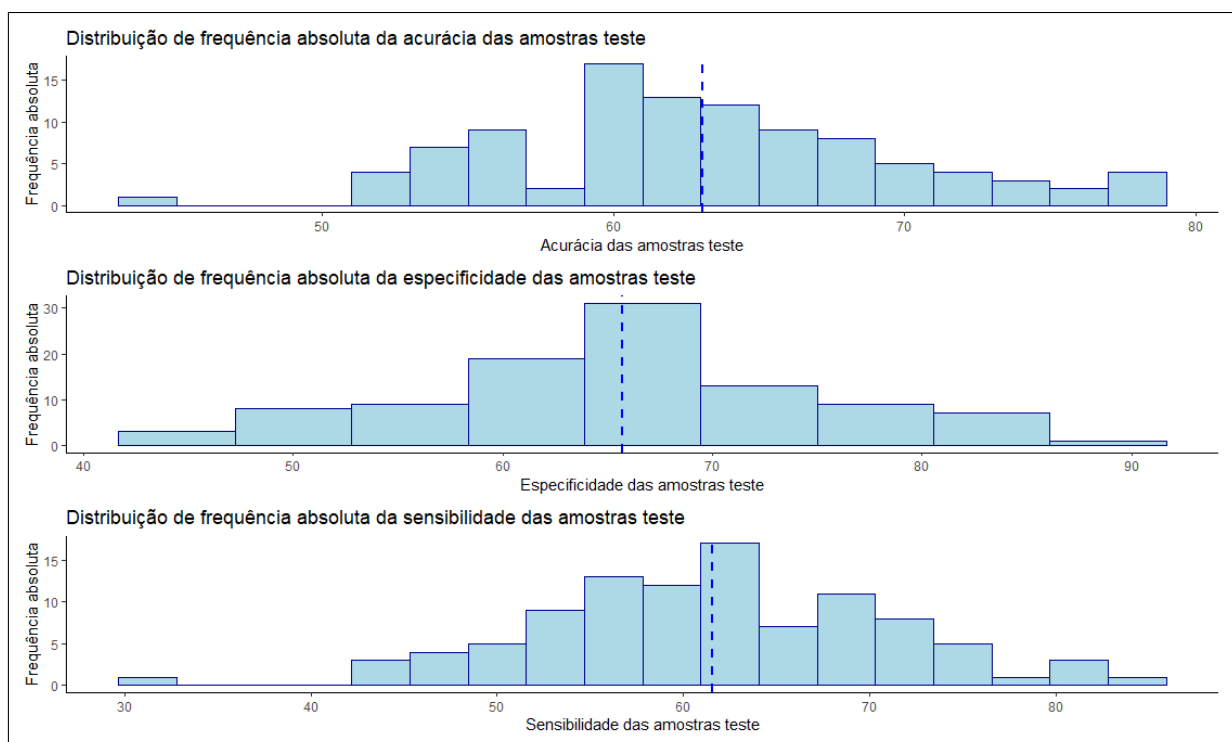


Figura 3: Distribuição de frequência absoluta das métricas de desempenho preditivo do modelo 1 final.

A utilização da UTI é uma variável que nem sempre é possível ser previamente conhecida na admissão hospitalar, porém caso esta informação seja obtida em algum momento da internação e for de interesse gerar um novo modelo para avaliar a chance de óbito de um paciente, essa informação pode tornar-se relevante. Com isso, foi adicionada essa variável ao modelo 1 final (Tabela 4) e gerado um modelo 2 com a não utilização da UTI como categoria de referência, cujos resultados são exibidos na Tabela 5. Analisando os valores de RC's ajustados, identifica-se que há uma tendência ao aumento da chance do óbito quanto mais velho o paciente, visto que para aqueles de 80 ou mais, a chance de óbito aumenta 93% em relação aos de 60 a 69 anos. Além disso, idosos que foram hospitalizados no município do Rio de Janeiro têm 10% maior chance de óbito e pacientes que foram atendidos em hospitais públicos têm 122%. Os idosos que tiveram que sair do seu município para serem atendidos tiveram 19% mais chances de óbitos que os que não tiveram esse tipo de peregrinação. Em relação ao diagnóstico principal, as chances de morte foram bem similares ao modelo final da Tabela 4. Para a nova variável utilização da UTI, podemos observar que foi muito significativa através do teste de Wald e os pacientes que utilizaram a unidade de terapia intensiva tiveram 287% de chance a mais de morrer que os que não utilizaram, o que mostra que a pessoas ter utilizado UTI influi consideravelmente no desfecho da internação. Importante destacar que a redução

no valor da estatística *deviance* com a inclusão da variável utilização da UTI indica que o modelo se tornou mais adequado para explicar o comportamento do óbito. Importante destacar que a redução no valor da estatística *deviance* e o teste de razão de verossimilhança ($G=418,77 - p - valor < 0,001$) indicam que o modelo com a inclusão da variável utilização da UTI se tornou mais adequado para explicar o comportamento do óbito.

Tabela 5: Resultados da regressão logística – modelo 2 – variáveis regressoras: faixa etária, município do hospital, natureza jurídica do hospital, peregrinação do paciente, diagnóstico principal e utilização de UTI

Variáveis do modelo	Estimativa dos coeficientes	p-valor do teste Wald	RC (IC95%)
Intercepto	-1,85856	< 0,001	0,16 (0,14 ; 0,18)
Faixa etária			
60 a 69 (ref)			
70 a 79	0,28221	< 0,001	1,33 (1,20 ; 1,47)
80 ou mais	0,66027	< 0,001	1,93 (1,76 ; 2,13)
Município do Hospital			
Outro Município(ref)			
Rio de Janeiro	0,09814	0,04228	1,10 (1,00 ; 1,21)
Natureza jurídica do hospital			
Privado(ref)			
Pública	0,79951	< 0,001	2,22 (2,00 ; 2,47)
Peregrinação do Paciente			
Não(ref)			
Sim	0,17521	0,00101	1,19 (1,07 ; 1,32)
Diagnóstico Principal			
III(ref)			
V	-3,07517	< 0,001	0,05 (0,01 ; 0,15)
VI	-0,43763	< 0,001	0,65 (0,56 ; 0,74)
VIII	0,14118	0,04804	1,15 (1,00 ; 1,32)
XI	0,91439	< 0,001	2,49 (2,24 ; 2,78)
Utilização de UTI			
Não(ref)			
Sim	1,35300	< 0,001	3,87 (3,39 ; 4,42)

Características gerais: *Deviance* = 14.990,86 (p -valor $\cong 1$) / $G=418,77$ (p -valor < 0,001) / Ponto de corte = 0,345

Área sob a curva ROC = 0,682

Para ilustrar a capacidade preditiva do modelo final exibido na Tabela 5, a Figura 4 apresenta a distribuição absoluta das métricas de desempenho do mesmo. Como se pode observar, a sensibilidade e especificidade apresentam comportamento simétrico, a média da acurácia é superior a 65%, enquanto a média da especificidade é 68,7% e da sensibilidade é 64,2% para as 100 amostras utilizadas como teste. Além disso, valores dos quartis 1 indicam que somente 25% das amostras têm acurácia abaixo de 62,00%, sensibilidade abaixo de 59,38% e especificidade abaixo de 61,11%. Também foi verificado que das 100

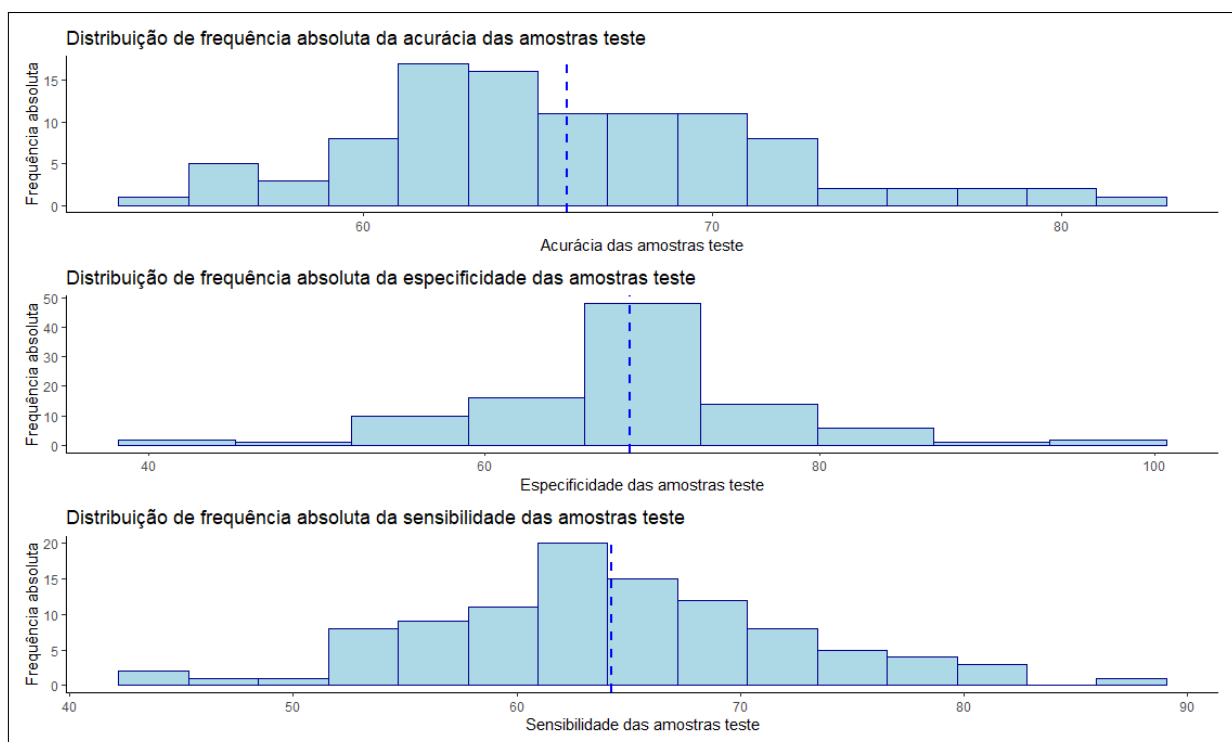


Figura 4: Distribuição de frequência absoluta das métricas de desempenho preditivo do modelo 2.

amostras avaliadas 95% obtiveram simultaneamente sensibilidade e especificidade acima de 50% e em 57% delas alcançaram ambas as estatísticas acima de 60%.

Devido à alta porcentagem de óbito na categoria de outras doenças respiratórias, foi averiguado e descoberto que insuficiência respiratória aguda representava boa parte dessa porcentagem, como pode ser observado na Figura 2. Com isso, essa doença foi separada do grupo de outras doenças respiratórias e um novo modelo foi considerado com apenas 3 categorias (Pneumonia, insuficiência respiratória aguda e outra doença respiratória) para verificar se comparando com outras doenças respiratórias sem a insuficiência respiratória, a pneumonia teria uma chance de óbito maior. Os resultados do modelo com essa nova classificação de diagnóstico do paciente são exibidos na Tabela 6. Como se pode observar no modelo 3 final, quanto maior a faixa etária, maior a chance de morte assim como nos modelos anteriores, com 80 ou mais tendo 76% de chance maior de óbito que a classe etária de 60 a 69 anos. Pacientes atendidos em hospitais públicos tiveram 83% de chances mais elevadas de morrerem e os que precisaram se deslocar de seu município para serem internados aumentaram sua chance de óbito em 23%, comparativamente às suas respectivas categorias de referência. Em relação ao diagnóstico principal, identifica-se que os pacientes com pneumonia tem 17% mais chances de morrerem que os de outra doença respiratória. Como já se suspeitava, idosos que sofreram insuficiência respiratória aguda

tiveram uma chance de morte absurdamente maior que por qualquer outra doença respiratória. Também deve ser observado que o teste de razão de verossimilhança ($G=3,57$) obteve $p - valor=0,17$, indicando que o modelo final 3 (somente com variáveis significativas) se ajusta melhor aos dados que o modelo inicial 3. Outro resultado interessante é o da redução no valor da estatística *deviance*, comparativamente aos do modelo final 1 e modelo 2, indicando que o modelo 3 tem uma melhor qualidade de ajuste. O modelo pode ser descrito pela seguinte expressão:

$$\begin{aligned} \ln \left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right] = & - 1,67730 + 0,24232 * \textit{faixa etária}_{70 \text{ a } 79} & (3.1) \\ & + 0,56691 * \textit{faixa etária}_{80 \text{ ou mais}} \\ & + 0,60634 * \textit{natureza jurídica}_{Pública} \\ & + 0,20612 * \textit{peregrinação}_{Sim} \\ & + 0,16107 * \textit{diagnóstico principal}_{Pneumonia} \\ & + 2,51418 * \textit{diagnóstico principal}_{Insuficiência Respiratória} \end{aligned}$$

Tabela 6: Resultados da regressão logística – modelo 3 – variáveis regressoras: sexo, faixa etária, município do hospital, natureza jurídica do hospital, peregrinação do paciente e diagnóstico principal

Variáveis do modelo	Estimativa dos coeficientes	p-valor do teste Wald	RC (IC95%)
Modelo inicial: com todas as variáveis regressoras			
Intercepto	-1,66275	< 0,001	0,19 (0,17 ; 0,22)
Sexo Masculino(ref) Feminino	-0,03986	0,31758	0,96 (0,89 ; 1,04)
Faixa etária 60 a 69 (ref) 70 a 79 80 ou mais	0,24552 0,57597	< 0,001 < 0,001	1,28 (1,16 ; 1,41) 1,78 (1,61 ; 1,96)
Município do Hospital Outro Município(ref) Rio de Janeiro	0,07928	0,10263	1,08 (0,98 ; 1,19)
Natureza jurídica do hospital Privado(ref) Pública	0,58365	< 0,001	1,79 (1,62 ; 1,98)
Peregrinação do Paciente Não(ref) Sim	0,21190	< 0,001	1,24 (1,11 ; 1,37)
Diagnóstico Principal Outra doença respiratória(ref) Pneumonia Insuficiência respiratória	0,16146 2,51056	< 0,001 < 0,001	1,17 (1,07 ; 1,29) 12,31 (10,19 ; 14,95)
Características gerais: Deviance = 14,952,94 (p -valor \cong 1)			
- Modelo final: sem as variáveis regressoras não significativas			
Intercepto	-1,67730	< 0,001	0,19 (0,16 ; 0,21)
Faixa etária 60 a 69 (ref) 70 a 79 80 ou mais	0,24232 0,56691	< 0,001 < 0,001	1,27 (1,15 ; 1,41) 1,76 (1,60 ; 1,94)
Natureza jurídica do hospital Privado(ref) Pública	0,60634	< 0,001	1,83 (1,66 ; 2,02)
Peregrinação do Paciente Não(ref) Sim	0,20612	< 0,001	1,23 (1,11 ; 1,36)
Diagnóstico Principal Outra doença respiratória(ref) Pneumonia Insuficiência respiratória	0,16107 2,51418	< 0,001 < 0,001	1,17 (1,07 ; 1,29) 12,36 (10,23 ; 15,01)
Características gerais: Deviance = 14,956,5 (p -valor \cong 1) / $G=3,57$ (p -valor=0,17) / Ponto de corte = 0,335 Área sob a curva ROC = 0,657			

A Figura 5 exibe a distribuição de frequência absoluta das métricas da capacidade preditiva do modelo final 3 e assim como nos modelos anteriores, a média de todas as métricas é maior que 60%. Novamente pelos valores dos quartis 1, identifica-se que apenas

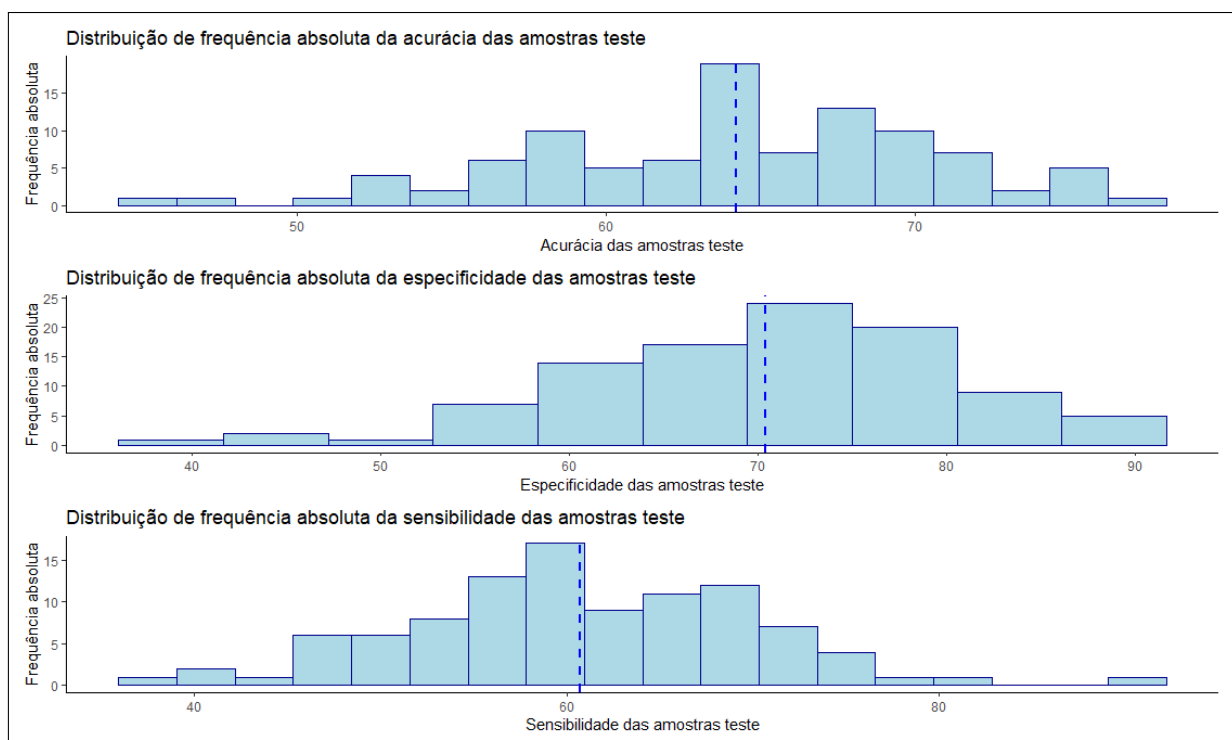


Figura 5: Distribuição de frequência absoluta das métricas de desempenho preditivo do modelo 3.

25% dos pacientes têm sensibilidade abaixo de 56,25%, especificidade abaixo de 65,28% e acurácia inferior a 59,50%. Também foi verificado que nas 100 amostras de teste, 87% delas obtiveram simultaneamente sensibilidade e especificidade acima de 50%, mas apenas 41% alcançaram ambas as estatísticas acima de 60%.

Assim como no modelo 1, foi acrescentada a variável utilização de UTI ao modelo 3 e ajustado o modelo 4, cujos resultados são apresentados na Tabela 7. As interpretações são bem similares ao modelo 3 em relação à identificação dos fatores de risco, havendo em geral elevação nos valores das RC's. Já em relação à nova variável, percebe-se que quem precisou ir para a UTI teve 310% de chance de morte maior que os que não chegaram a ir. Vale destacar que o teste de razão de verossimilhança ($G=433,95$) obteve $p - valor < 0,001$, indicando que o modelo final 3 (sem a variável UTI) não se ajusta melhor aos dados do que o modelo 4, que pode ser escrito por:

$$\ln \left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right] = - 2,03736 + 0,25950 * faixa\ etária_{70\ a\ 79} \quad (3.2)$$

$$+ 0,61929 * faixa\ etária_{80\ ou\ mais}$$

$$+ 0,79991 * natureza\ jurídica_{Pública}$$

$$+ 0,20129 * peregrinação_{Sim}$$

$$+ 0,21785 * diagnóstico\ principal_{Pneumonia}$$

$$+ 2,50674 * diagnóstico\ principal_{Insuficiência\ Respiratória}$$

$$+ 1,41160 * utilização\ da\ UTI_{Sim}$$

Tabela 7: Resultados da regressão logística – modelo 4 – variáveis regressoras: faixa etária, natureza jurídica do hospital, peregrinação do paciente, diagnóstico principal e utilização de UTI

Variáveis do modelo	Estimativa dos coeficientes	p-valor do teste Wald	RC (IC95%)
Intercepto	-2,03736	< 0,001	0,13 (0,11 ; 0,14)
Faixa etária			
60 a 69 (ref)			
70 a 79	0,25950	< 0,001	1,30 (1,17 ; 1,44)
80 ou mais	0,61929	< 0,001	1,86 (1,68 ; 2,05)
Natureza jurídica do hospital			
Privado(ref)			
Pública	0,79991	< 0,001	2,22 (2,01 ; 2,47)
Peregrinação do Paciente			
Não(ref)			
Sim	0,20129	< 0,001	1,22 (1,10 ; 1,36)
Diagnóstico Principal			
Outra doença respiratória(ref)			
Pneumonia	0,21785	< 0,001	1,24 (1,13 ; 1,36)
Insuficiência respiratória	2,50674	< 0,001	12,26 (10,13 ; 14,94)
Utilização de UTI			
Não(ref)			
Sim	1,41160	< 0,001	4,10 (3,59 ; 4,70)

Características gerais: *Deviance* = 14.522,55 (*p-valor* \cong 1) / *G*=433,95 (*p-valor* < 0,001) / Ponto de corte = 0,334
Área sob a curva ROC = 0,696

Para mostrar a distribuição de frequência absoluta das métricas de desempenho do modelo 4, foram construídos histogramas, que estão exibidos na Figura 6. Esse modelo foi o que teve o melhor desempenho, com a média de todas as métricas de avaliação sendo maiores que 65%. Além disso, com relação ao primeiro quartil, para a acurácia, sensibilidade e especificidade somente 25% dos idosos tiveram essas métricas menores que

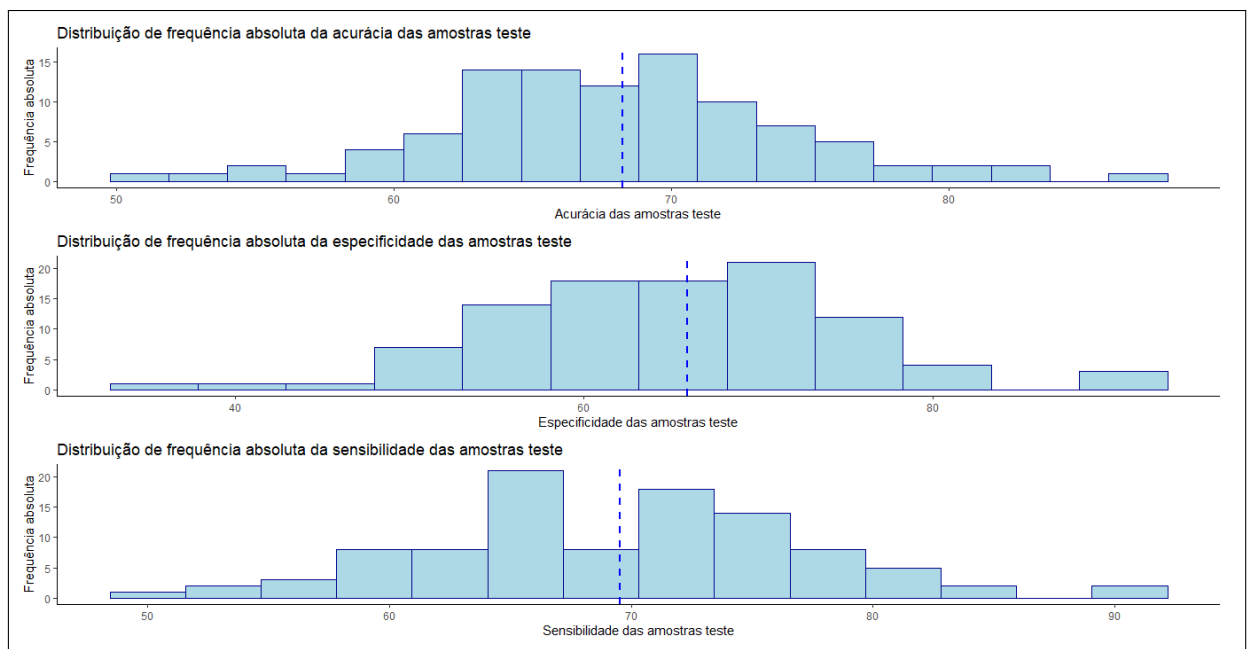


Figura 6: Distribuição de frequência absoluta das métricas de desempenho preditivo do modelo 4.

64,00%, 65,62% e 61,11%. Por último foi identificado que das 100 amostras de testes, 97% delas obtiveram simultaneamente sensibilidade e especificidade acima de 50% e em 64% delas ambas as estatísticas tiveram valores acima de 60%.

Os resultados dos modelos indicaram que a informação sobre a utilização da UTI é muito importante, mas para o momento da admissão hospitalar (onde pode não se ter conhecimento sobre a utilização da UTI), o modelo final 3 é o que melhor descreve o comportamento do desfecho da internação.

4 Conclusão

Nesta seção são destacadas as principais conclusões obtidas neste Trabalho de Conclusão de Curso, cujo objetivo principal era investigar o perfil das internações emergenciais por doenças respiratórias em idosos residentes no estado do Rio de Janeiro e identificar possíveis fatores associados ao desfecho óbito. Verificou-se que de todos os modelos aqui apresentados, os modelos finais 3 e 4 obtiveram os menores valores da estatística *deviance*, indicando serem eles os que melhor explicam o comportamento dos desfechos das internações. Pôde ser observado no modelo 4, que ter mais de 80 anos, ser atendido em hospitais públicos, ter que se deslocar de seu município para ser internado, ter como diagnóstico principal a pneumonia ou principalmente insuficiência respiratória aguda e utilizar a UTI durante a internação são fatores de risco para o óbito do paciente. Deve-se ressaltar, no entanto, que a utilização da UTI não deve ser interpretada por si só como um fator que aumente o risco de óbito. Na verdade retrata o encaminhamento de um paciente com maior gravidade de saúde para uma unidade de assistência intensiva. Consequentemente é razoável esperar que para estes casos o risco de óbito seja maior.

Uma das limitações desse estudo foi a ausência de informações clínicas do paciente que poderiam auxiliar na obtenção de um modelo com desempenho e capacidade preditiva melhores.

Entende-se que esta pesquisa contribui para uma investigação que viabilize a comparação dos comportamentos de internações emergenciais por doenças respiratórias, síndromes agudas graves e por covid em idosos.

Referências

- Boletins Epidemiológicos. [S.l.], 2021. Acessado em: 03 mai. 2021. Disponível em: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/boletins-epidemiologicos/covid-19>.
- BONFADA, D. et al. Análise de sobrevida de idosos internados em unidade de terapia intensiva. *Revista Brasileira de Geriatria e Gerontologia*, SciELO Brasil, v. 20, n. 2, p. 197–205, 2017.
- CACCIATORE, F. et al. Pneumonia and hospitalizations in the elderly. *Geriatric Care*, v. 3, n. 1, 2017.
- DATASUS/SIHSUS. *Sistema de Informação Hospitalar do Departamento de Informática do Sistema Único de Saúde*. 2019. Acessado em: 03 mai. 2021. Disponível em: <https://datasus.saude.gov.br/transferencia-de-arquivos/>.
- DOBSON, A. J.; BARNETT, A. G. An introduction to generalized linear models capa. CRC Press, 2018.
- FRANCISCO, P. M. S. B.; DONALISIO, M. R. de C.; LATTORRE, M. do Rosário Dias de O. Impacto da vacinação contra influenza na mortalidade por doenças respiratórias em idosos. *Revista de Saúde Pública*, SciELO Brasil, v. 39, n. 1, p. 75–81, 2005.
- JR, D. W. H.; LEMESHOW, S.; STURDIVANT, R. X. Applied logistic regression. John Wiley Sons, v. 398, 2013.
- MENARD, S. Applied logistic regression analysis. Sage, v. 106, 2002.
- Ministério da Saúde. [S.l.], 2019. Acessado em: 03 mai. 2021. Disponível em: <https://covid.saude.gov.br/>.
- NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, Wiley Online Library, v. 135, n. 3, p. 370–384, 1972.
- OMS. *Pneumonia*. 2021. Acessado em: 14 fev. 2022. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/pneumonia>.
- Painel Corona Vírus Brasil. [S.l.], 2021. Acessado em: 22 dez. 2021. Disponível em: <https://covid.saude.gov.br/>.
- TEIXEIRA, J. J. M.; BASTOS, M. G. C. F. C.; SOUZA, A. C. L. de. Perfil de internação de idosos. *Revista da Sociedade Brasileira de Clínica Médica*, v. 15, n. 1, p. 15–20, 2017.

TURKMAN, M. A. A.; SILVA, G. L. Modelos lineares generalizados: da teoria á prática. *VIII Congresso Anual da Sociedade Portuguesa de Estatística, Lisboa, 2000.*

VIACAVA, F. et al. Sus: oferta, acesso e utilização de serviços de saúde nos últimos 30 anos. *Ciênc. saúde colet.*, 2018.