**Universidade Federal de Uberlândia**
**Faculdade de Matemática**

**Bacharelado em Estatística**

# AVALIAÇÃO DE FATORES DE RISCO DE MORTE POR COVID-19 EM CASOS HOSPITALIZADOS EM UBERLÂNDIA - MG

**André Luiz Rodrigues Souza**

**Uberlândia-MG**
**2021**

**André Luiz Rodrigues Souza**

# AVALIAÇÃO DE FATORES DE RISCO DE MORTE POR COVID-19 EM CASOS HOSPITALIZADOS EM UBERLÂNDIA - MG

Trabalho de conclusão de curso apresentado à Coordenação do Curso de Bacharelado em Estatística como requisito parcial para obtenção do grau de Bacharel em Estatística.

Orientador: Prof. Dr. Rogério de Melo Costa Pinto

**Uberlândia-MG**

**2021**

# Universidade Federal de Uberlândia
# Faculdade de Matemática

## Coordenação do Curso de Bacharelado em Estatística

A banca examinadora, conforme abaixo assinado, certifica a adequação deste trabalho de conclusão de curso para obtenção do grau de Bacharel em Estatística.

Uberlândia, ____ de _____ de 20____

BANCA EXAMINADORA

_____

Prof. Dr. Rogério de Melo Costa Pinto

_____

Prof. Dr. Ednaldo Carvalho Guimarães

_____

Prof. Dr. Marcelo Tavares

**Uberlândia-MG**

**2021**

# Agradecimentos

Primeiramente gostaria de agradecer a Deus por me dar sabedoria, forças e resiliência para chegar até aqui.

Aos meus pais, Joaquina Rodrigues e Otacílio da Felicidade por me darem incentivo todas as vezes que precisei ao longo da minha vida.

Ao meu orientador Prof. Dr. Rogério de melo Costa Pinto que aceitou me orientar na última etapa desse ciclo acadêmico, com sua vasta experiência acadêmica me direcionou da melhor maneira para atingir os melhores resultados nesse estudo.

A Universidade Federal de Uberlândia e todo seu corpo docente principalmente aos professores com quem cruzei nessa trajetória de graduação.

Aos colegas de curso que sempre estavam presentes para trocar informações e aprendizado ajudando para conseguir ultrapassar todos os desafios em nossa vida acadêmica.

# Resumo

Com a proliferação do vírus covid-19, diversos estudos têm mostrado que pessoas com comorbidades possuem maior risco de morte quando acometidas pela covid-19. Assim, o presente estudo foi realizado para identificar quais fatores de riscos apresentados por pacientes hospitalizados geram maior risco de morte na cidade de Uberlândia – MG. Esse trabalho consiste em uma análise de regressão logística, utilizando dados do Sistema de Saúde (Open Data SUS, 2021) para monitoramento de casos e óbitos de covid-19 hospitalizados na cidade de Uberlândia.

A variável dependente foi se houve ou não óbito e as variáveis independentes foram Idade, saturação baixa oxigênio $< 95\%$, desconforto respiratório, doença cardiovascular crônica, doença hepática crônica, asma, doença neurológica crônica, pneumopatia crônica, Imunodepressão, doença renal crônica, obesidade, internação em UTI e tomou vacina para gripe. Foi verificado que os maiores fatores de risco para óbito para covid-19 são: doença hepática crônica (OR = 7,86; p< 0,006), Imunodepressão (OR = 6,48; p<0,000), doença neurológica crônica (OR = 3,2; p<0,000), Internação em UTI (OR = 2,49; p<0,000), doença renal crônica, (OR = 2,39; p<0,000) e pneumopatia crônica (OR = 2,06; p<0,000). O fator vacina contra a gripe com (OR=0,72; p<0,005) foi o único encontrado nas variáveis significativas com características de proteção. Por meio da matriz de confusão foi constatado 75,70% de acurácia do modelo. E necessário que o sistema de saúde levante informações inferenciais da população sobre suas comorbidades para melhor se planejar para atendimento a população que tem maior risco de óbito a partir dos fatores estudados.

**Palavras-chave:** Open Data SUS, Covid-19, Regressão Logística.

# Abstract

With the proliferation of the covid-19 virus, several studies have found that people with comorbidities have a higher risk of death when affected by covid-19. Thus, the present study was carried out to identify risk factors for risk by hospitalized patients that generate a higher risk of death in the city of Uberlândia - MG. This work consists of an analysis of logistic regression, using data from the Health System (Open Data SUS, 2021) for monitoring cases and deaths of hospitalized covid-19 in the city of Uberlândia.

The dependent variable was death or not and the independent variables were Age, low oxygen saturation <95%, respiratory distress, chronic cardiovascular disease, chronic liver disease, asthma, chronic neurological disease, chronic pneumopathy, Immunodepression, chronic kidney disease, obesity, hospitalization at UTI and flu vaccine. It was found that the biggest risk factors for death for covid-19 are chronic liver disease (OR = 7.86; p <0.006), immunodepression (OR = 6.48; p <0.000), chronic neurological disease (OR = 3.2; p <0.000), ICU admission (OR = 2.49; p <0.000), chronic kidney disease, (OR = 2.39; p <0.000) and chronic lung disease (OR = 2.06; p <0.000). Through the confusion matrix, 75,70% accuracy of the model was found. It is necessary for the health system to raise inferential information from the population about their comorbidities to better plan for the care of the population that has a higher risk of death based on the factors studied.

**Keywords:** Open Data SUS, Covid-19, Logistic Regression.

# Sumário

# Lista de Figuras

# Lista de Tabelas

# 1. INTRODUÇÃO

Desde o início do surto de coronavírus (SARS-CoV-2), causador da Covid-19, houve uma grande preocupação diante de uma doença que se espalhou rapidamente em várias regiões do mundo, com diferentes impactos. De acordo com a Organização Mundial da Saúde (OMS), em 18 de março de 2020, os casos confirmados da Covid-19 já haviam ultrapassado 214 mil em todo o mundo. Não existiam planos estratégicos prontos para serem aplicados a uma pandemia de coronavírus - tudo é novo. Recomendações da OMS, do Ministério da Saúde do Brasil, do Centro de Controle e Prevenção de Doenças (CDC - Centers for Disease Control and Prevention, Estados Unidos) e outras organizações nacionais e internacionais, têm sugerido a aplicação de planos de contingência de influenza e suas ferramentas, devido às semelhanças clínicas e epidemiológicas entre esses vírus respiratórios. Esses planos de contingência preveem ações diferentes de acordo com a gravidade das pandemias (Freitas, 2020)[11]. Este cenário complexo impõe desafios adicionais à vigilância epidemiológica, às relações diplomáticas internacionais e à programação de políticas públicas, sobretudo por meio de medidas que reduzam as desigualdades de acesso aos sistemas de saúde e a condições estruturais para o autocuidado. Atentar para o comportamento desta pandemia nas distintas regiões parece ser imprescindível para a atualização das estratégias de enfrentamento desta emergência global e suas repercussões no nível local (Rafael, 2020)[20].

Com o aumento significativo de casos e consequentemente de mortes, se faz necessário entender os principais fatores que geram óbitos pelo vírus. O CDC classifica como fatores de risco de morte: doença renal crônica; doença pulmonar obstrutiva crônica (DPOC); obesidade; estado imunocomprometido (sistema imunológico enfraquecido) do transplante de órgão sólido; condições cardíacas graves, como insuficiência cardíaca, doença arterial coronariana ou cardiomiopatias; anemia falciforme e diabetes tipo 2. Outros fatores que podem aumentar o risco de uma pessoa ter doença grave e morte: asma, hipertensão e condições neurológicas (demência, AVC) (Camargo, 2020)[4].

No início da pandemia de 12 fevereiro a 28 de março de 2020 o CDC realizou um estudo nos Estados Unidos sobre a presença ou ausência de condições de saúde subjacentes e outros fatores de risco (CDC, 2020)[6]. Aproximadamente um terço desses pacientes (37,6%) apresentava pelo menos uma condição de base ou fator de risco. Diabetes (10,9%), doença pulmonar crônica (9,2%) e doença cardiovascular (9,0%) foram às condições mais frequentemente relatadas entre todos os casos.

No Brasil de acordo com Rezende (2020)[21], foi estimado que um terço (53 milhões) mais

da metade (86 milhões) dos brasileiros adultos apresentam pelo menos um fator de risco para Covid-19, o que aponta para alta prevalência de condições médicas graves em adultos mais velhos, onde a prevalência de um ou mais fatores de risco para doença grave foi de 47,3% em jovens e 75,9% em adultos mais velhos. Segundo Brasil (2021)[3], dentre os óbitos por Covid-19 na 2ª semana epidemiológica de 2021, 1.019 pessoas a faixa etária com o maior número de óbitos notificados é a de 80 a 89 anos de idade, com 252 (25,2%) óbitos.

Os estados com maior vulnerábilidade em relação a quantidade absoluta e relativa de pessoas adultas com grande risco foi São Paulo, Rio de Janeiro, Minas Gerais e Rio Grande do Sul. Existem diferenças entre as regiões Sul e Sudeste vs Norte e Nordeste que podem ser devido à estrutura etária diferente, prevalência do estado de saúde e / ou acesso a diagnóstico e cuidados médicos. Estimar a proporção da população em risco de Covid-19 grave dentro e entre países é de suma importancia para preparar o sistema de saude e previr mortes de pessoas com alto risco. Porem, até onde sabemos essas estimativas ainda estão em fase de estudo em todo o mundo. Nos EUA, estimou-se que quatro em cada dez (37,6%) adultos com mais de 18 anos podem estar em alto risco de Covid-19 grave (Rezende, 2020)[21].

Embora o Brasil tenha sido o primeiro país da América do Sul a apresentar um caso confirmado de Covid-19, este ocorreu várias semanas após a maioria dos países do hemisfério Norte (Cavalcante, 2020)[5].

Na cidade de Uberlândia – MG, após o primeiro caso em 17/03/2020, foi decretado medidas de isolamento social e restrições de funcionamento de comercio, suspensão das atividades escolares em escolas, colégios, faculdades e centros universitários particulares pelo prazo de até 60 (sessenta) dias, a contar do dia 18/03/2020; excetuando-se desta recomendação as atividades relativas aos estudantes da área de saúde (Docs.Uberlândia, 2020)[8].

Com o objetivo de identificar quais os principais fatores de risco de morte em pacientes infectados pelo Covid-19 na cidade de Uberlândia - MG, foram utilizados dados do questionário (Open Data SUS, 2021)[19] aplicado a pacientes hospitalizados.

Por meio de um modelo de regressão logística pretende-se avaliar quais os principais fatores de risco de morte em pacientes infectados pelo Covid-19 e qual a chance de morte dos pacientes que possuem essas comorbidades. É de extrema importância que o sistema de saúde levante informações inferenciais da população sobre suas comorbidades, para melhor se planejar para atendimento à população que tem maior risco de óbito a partir dos fatores estudados e definir estratégias de políticas públicas no período da pandemia.

# 2. METODOLOGIA

## 2.1 CONJUNTOS DE DADOS

Os dados referentes ao estudo de fatores que oferecem maior risco de óbito quando o paciente é contaminado e internado por motivo de coronavirus, foram obtidos da ficha individual de casos de síndrome respiratória aguda grave hospitalizada com gestão do sistema de informação de vigilância epidemiológica da gripe no site (Open Data SUS, 2021)[19]. A base de dados utilizada do Open Data SUS contém informações referentes ao estado de saúde dos pacientes hospitalizados. Para o presente estudo foi segmentado as informações pelo município de Uberlândia-MG o período de coleta de informações foi referente ao período de 17/03/2020 a 05/03/2021, correspondendo a 3149 casos hospitalizados e 784 óbitos confirmados em Uberlândia-MG por Covid-19.

Foram avaliadas as seguintes variáveis:

Variável resposta:

Evolução Clínica: Óbito, recebe valor 0 e paciente curado recebe valor 1. O paciente é considerado curado quando recebe alta hospitalar.

Variáveis independentes:

Sexo: masculino recebe valor 1 e feminino recebe valor 0 conresponde.

Idade: variável contínua de 0 a 104 anos, representa quantos anos a pessoa tem na entrada da internação hospitalar.

Síndromes Gripal (SG): Caso proveniente de surto de SG que evoluiu para SRAG recebe 1 se sim e recebe 0 se não.

Nosocomial: Infecção adquirida no hospital recebe 1 se sim e recebe 0 se não.

Contato ave e suíno: Paciente trabalha ou tem contato direto com aves, suínos, ou outro animal recebe 1 se sim e recebe 0 se não.

Febre: Sinais e sintomas de febre recebem 1 se sim e recebe 0 se não.

Tosse: Sinais e sintomas de tosse recebem 1 se sim e recebe 0 se não.

Garganta: Sinais e sintomas de dor de garganta recebem 1 se sim e recebe 0 se não.

Dispneia: Sinais e sintomas de dispneia recebem 1 se sim e recebe 0 se não.

Desconforto Respiratório: Sinais e sintomas de desconforto respiratório recebem 1 se sim e recebe 0 se não.

Saturação $< 95\%$: Sinais e Sintomas de Saturação O2$< 95\%$ recebem 1 se sim e recebe 0 se não.

Diarreia: Sinais e Sintomas de diarreia recebem 1 se sim e recebe 0 se não.

Vomito: Sinais e Sintomas de vomito recebem 1 se sim e recebe 0 se não.

Doença Cardiovascular Crônica: recebem 1 se sim e recebe 0 se não.

Doença Hematológica Crônica: recebem 1 se sim e recebe 0 se não.

Síndrome de Down: recebem 1 se sim e recebe 0 se não.

Doença Hepática Crônica: recebem 1 se sim e recebe 0 se não.

Asma: recebem 1 se sim e recebem 0 se não.

Diabetes mellitus: recebem 1 se sim e recebe 0 se não.

Doença Neurológica Crônica: recebem 1 se sim e recebe 0 se não.

Doença Pneumopatia Crônica recebem 1 se sim e recebe 0 se não.

Imunodeficiência ou Imunodepressão: recebem 1 se sim e recebe 0 se não.

Doença Renal Crônica: recebem 1 se sim e recebe 0 se não. Obesidade: recebem 1 se sim e recebe 0 se não.

Vacina: recebeu vacina da gripe na última campanha recebe 1 se sim e recebe 0 se não.

Internado em UTI recebe 1 se sim e recebe 0 se não.

Tempo Clínico: tempo que o paciente ficou hospitalizado, em dias.

A partir do questionário do (Open Data SUS, 2021)[19] aplicado a pacientes hospitalizados em Uberlândia-MG, foi realizada a estatística descritiva dos dados. Foi analisada a associação entre os fatores de risco e variáveis qualitativas e quantitativas. Na sequência foi ajustado aos dados um modelo de regressão logística para avaliar quais os principais fatores de risco de morte em pacientes infectados pelo Covid-19. Se faz necessario uma breve introdução aos Modelos Lineares Generalizados antes de aprofundar em metodos de Regressão Logística.

## 2.2 Modelos Lineares Generalizados

De acordo com Demétrio (2002)[7] os modelos linearmente generalizados são uma VA(variável aleatória) Y que tem distribuição pertencente à família exponencial (1) e a sua função densidade e probabilidade (fdp) é calculada da seguinte maneira:

$$f(y/\theta,\phi) = exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\} \qquad (1)$$

Onde, em que, $\theta$ e $\phi$ são parâmetros escalares a($\phi$), b($\theta$) e c(y,$\phi$) são funções reais conhecidas.

Portanto o modelo precisa conter os três pressupostos para ser considerado MLG:

Então a estrutura do MLG é feita em três partes:

1. A componente aleatória: representada pela a variável resposta;
2. A componente sistemática: representada pela combinação linear das variáveis explicativas;
3. A função de ligação: função que linearizara a relação entre $\sum(Y)$ e $X\beta$.

No presente estudo será utilizada a variação do modelo MLG de regressão logística que tem função de ligação Logit e componente aleatória binomial que será compreendido na próxima seção.

## 2.3 Regressão Logística

A regressão logística quando de natureza binaria ou dicotômica de acordo com Figueira (2006)[10] a variável resposta possui somente uma variável dependente envolvida.

Seja um experimento aleatório Z composto por m repetições independentes de um evento dicotômico, isto é, contendo apenas dois resultados possíveis: 0 (fracasso) ou 1 (sucesso). Assim, diz-se que esse experimento é de natureza binária. Portanto as condições são as mesmas para todas as repetições, as probabilidades de cada resultado são P(1) = $\phi$ e P(0) = (1 – $\phi$) e constantes ao longo das m repetições. Chamando de Y a variável aleatória de interesse representando o número de vezes, nas m repetições, em que ocorre "sucesso", os valores que Y poderão assumir são 0, 1, 2, ..., m; a cada um desses valores está associada uma probabilidade P(Y = y) de ocorrência, sendo (y = 0, 1, 2,..., m) (Barreto, 2011)[1] . De acordo com as definições dos parâmetros $\phi$ e m para a variável aleatória Y podemos associar a função de distribuição de probabilidade como a função de probabilidade Binomial. Segue a abaixo a fórmula de cálculo de sua fdp, esperança e variância:

$$f(y,\phi) = P(Y_i = y) = \binom{m}{y}\phi^y(1-\phi)^{m-y} \qquad (2)$$

$$\sum(Y_i) = m\phi \qquad (3)$$

$$Var(Y_i) = m\phi(1-\phi) \qquad (4)$$

Nas condições de MLG, para variáveis resposta contínuas não existem delimitações com relação a resposta esperada a ser estimada pelo modelo. Contudo, no caso de variáveis respostas dicotômicas, como a VA Bernoulli Y associada ao experimento Z contando com apenas uma repetição, portanto esses valores precisão estar restritos ao intervalo [0, 1]. Em vista que em (2), para o caso Bernoulli, o valor da resposta esperada significa a probabilidade de P(Yi =1), portando existe necessidade de transformação para torna a resposta esperada a ser estimada pelo modelo em uma função não linear. A transformação mais comumente utilizada para o caso de variáveis dicotômicas é a função exponencial que, por sua vez, transforma o valor esperado da variável Y em uma função de ligação logística (Barreto, 2011)[1].

$$\sum(Y_i) = \phi = \frac{exp(ni)}{1+exp(ni)} \qquad (5)$$

Podemos verificar que o objetivo da transformação em (5) é atingido de maneira desejada

já que $n_i$ pode atingir qualquer valor real, $\phi_i$, sendo uma probabilidade e sempre continuara restrito ao intervalo de [0,1] como desejado. Dessa forma o valor estimado pelo modelo linear será $n_i$ e, para a formalização do modelo de regressão logística, ele pode ser relacionado a um modelo linear, contendo $(p-1)$ variáveis explicativas e $p$ parâmetros, resultando em uma função conhecida como função resposta logit (6):

$$n_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ ... \ + \beta_{p-1} X_{ip-1} \ ; \ i = 1, 2, ..., n \qquad (6)$$

Onde,

$n_i$: tem resposta média logit para a i-ésima observação.

$X_i$: valor da variável preditora para a i-ésima observação.

$\beta k$ (k = 0, l, ..., p - 1): coeficientes de regressão logística.

Fazendo a substituição das equações (5) e (6) formulamos o modelo de regressão logística múltipla da seguinte maneira:

$$\sum(Y_i) = \phi = \frac{exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ ... \ + \beta_{p-1} X_{ip-1})}{1 + exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ ... \ + \beta_{p-1} X_{ip-1})} \qquad (7)$$

Portanto esse modelo (7) assume que os $Y_i$ são variáveis aleatórias Bernoulli independentes com parâmetro $\mathrm{E}(Y_i) = \phi_i$.

Para um modelo de regressão logística simples seria utilizado uma única variável explicativa e dois parâmetros a serem estimados: $\beta_0$ *e* $\beta_1$.

## 2.4 Odds Ratio

Com o advento do modelo linear em (6), o valor esperado da resposta em (5) significa agora a probabilidade de que $Y_i = 1$, dados aos níveis para as preditoras em [6]; e $n_i$ se vincula a resposta logística esperada pela conhecida função de ligação logit, pois explicitando em (6) em termos de $n_i$, obtém-se:

$$n_i = ln(\frac{\phi_i}{1 - \phi_i}) \qquad (8)$$

De acordo com Barreto (2011)[1], a formulação entre parênteses em (8) é conhecido como odds de sucesso:

1. Se a probabilidade de sucesso equivaler a 0,5, o odds vale 1.
2. Se a probabilidade de sucesso for inferior a 0,5, o odds é inferior a 1.
3. Se a probabilidade de sucesso for superior a 0,5, o odds será maior do que 1.

A interpretação dos coeficientes $\beta$ em regressão logística não é tão intuitiva e nem de direta compreensão, como no caso de regressão linear, uma vez que se trata de uma função de resposta não-linear. Assim, no caso de regressão logística simples, um incremento unitário de X implicará em um efeito multiplicativo do odds estimado de sucesso, ou seja, $exp(\beta_1)$ caso de uma regressão logística múltipla, um incremento unitário em $X_1$ ocasionaria esse mesmo efeito, mantidas

constantes todas as demais variáveis do modelo $(X_2, X_3, \ldots, X_{p-1)}$. A prova das afirmações utilizadas acima pode ser verificada considerando um estado inicial para odds estimado. Seja ODDS1:

$$\widehat{odds1} = (\frac{\hat{\phi}_i}{1-\hat{\phi}_i}) \tag{9}$$

$$\widehat{odds1} = \frac{\frac{exp(\beta_0+exp\beta_{1X})}{1+exp(\beta_0+exp\beta_{1X})}}{1-\frac{exp(\beta_0+exp\beta_{1X})}{1+exp(\beta_0+exp\beta_{1X})}} \tag{10}$$

$$\widehat{odds1} = exp(\beta_0 + exp\beta_{1X}) \tag{11}$$

Agora de definimos um ODDS2, a partir de uma variação unitária em $X$:

$$\widehat{odds2} = \frac{\frac{exp(\beta_0+\beta_1(X+1))}{1+exp(\beta_0+\beta_1(X+1))}}{1-\frac{exp(\beta_0+\beta_1(X+1))}{1+exp(\beta_0+\beta_1(X+1))}} \tag{12}$$

$$\widehat{odds2} = \frac{\frac{exp(\beta_0+\beta_{1X}+\beta_1)}{1+exp(\beta_0+\beta_{1X}+\beta_1)}}{1-\frac{exp(\beta_0+\beta_{1X}+\beta_1)}{1+exp(\beta_0+\beta_{1X}+\beta_1)}} \tag{13}$$

$$\widehat{odds2} = \frac{exp(\beta_0+\beta_{1X}+\beta_1)}{1+exp(\beta_0+\beta_{1X}+\beta_1)-exp(\beta_0+\beta_{1X}+\beta_1)} \tag{14}$$

$$\widehat{odds2} = exp(\beta_0+\beta_{1X}+\beta_1) \tag{15}$$

$$\widehat{odds2} = exp(\beta_0+\beta_{1X})exp(\beta_1) \tag{16}$$

$$\widehat{odds2} = \widehat{odds1}exp(\beta_1) \tag{17}$$

Portanto temos que o odds ratio (também conhecido por razão entre os odds, ou razão chances), que mensura a taxa de variação do odds de sucesso em função da variação em X, equivale a:

$$\frac{\widehat{odds2}}{\widehat{odds1}} = exp(\beta_1) \qquad (18)$$

## 2.5 TESTE DE WALD

Para avaliar a significância dos parâmetros dos modelos é necessário utilizar o teste de Wald. De acordo com Taconeli (2015)[24], o teste de Wald baseia-se na distribuição assintótica normal dos estimadores de máxima verossimilhança dos parâmetros do modelo. Seja $\hat{\beta}_j$ o estimador de máxima verossimilhança de $\beta_j$, um particular parâmetro de um MLG. Conforme discutido anteriormente, para $n \to \infty$,

$$\hat{\beta}_j \sim Normal(\beta_j, Var(\hat{\beta}_j)) \qquad (19)$$

onde $Var(\hat{\beta}_j)$ é estimada através do correspondente termo da diagonal da matriz de covariâncias $\widehat{VAR}(\hat{\beta}) = (X'\hat{W}X)^{-1}\hat{\phi}$. É denotado por $ep(\hat{\beta}_j) = \sqrt{VAR(\hat{\beta}_j)}$ o erro padrão de $(\hat{\beta}_j)$.

Embora possam ser aplicados ao teste de hipóteses contemplando dois ou mais parâmetros, o uso mais frequente do teste de Wald contempla apenas um parâmetro por vez. Considere então o seguinte par de hipóteses:

$$H_0 : \beta_j = \beta_j^0$$
$$H_1 : \beta_j \neq \beta_j^0, \qquad (20)$$

Em que $\beta_j^0$ é algum valor postulado para $\beta_j$ (é comum tomarmos $\beta_j^0 = 0$, a fim de testarmos a nulidade de $(\beta_j)$. Então, o teste de Wald baseia-se na seguinte estatística-teste:

$$Z_t : \frac{\hat{\beta}_j - \beta_j^0}{ep(\beta_j)}, \qquad (21)$$

que, sob a hipótese nula, tem assintoticamente distribuição Normal padrão.

No software estatístico R: A estatística e o teste de Wald são apresentados no comando summary de um MLG.

## 2.6 METODOLOGIA DE PREDIÇÃO DO MODELO

Quando se ajusta os modelos de regressão logística um dos principais objetivos é verificar o quanto o modelos tem poder de discriminação e predição. Para verificar a capacidade preditiva do modelo vamos avaliar atravez da acuracia da matriz de confusão, Curva ROC, AUC, Sensibilidade e Especificidade e o Pseudo $R^2$.

## 2.7 Matriz de confusão

Atraves das probabilidades logísticas estimadas na regressão se faz necessario criar uma tabela com o resultado da classificação cruzada da variável resposta, de acordo com uma variável dicotômica em que os valores se derivam (Hosmer e Lemeshow, 2000)[14].

De acordo com Hosmer e Lemeshow (2000)[14], podemos avaliar as probabilidades que se aproximam de 1, a classificação do indivíduo pode ser estimada como $\hat{Y_(i)} = 1$, de forma inversa, se o modelo estimar probabilidades perto de 0, pode se classificar como $\hat{Y_(i)} = 0$. Para classificar a estimação utilizaremos nesse estudo o ponto de conte 0,24. Em alguns casos o corte pode ser limitado a 0,5 de acordo com (Hosmer e Lemeshow, 2000)[14].

Com a (Tabela 2.1) de matriz de confusão é determinados os avaliadores acerca da capacidade preditiva do modelo estimado:

Onde,

VP: Verdadeiro Positivo

VN: Verdadeiro Negativo

FP: Falso Positivo

FN: Falso Negativo

Tabela 2.1: Matriz de confusão para um problema com duas classes: "positivo" e "negativo".

| Previsto | Observado | |
|---|---|---|
| | Y=1 | Y=0 |
| Y=1 | Verdadeiro positivo (VP) | Falso positivo (FP) |
| Y=0 | Falso negativo (FN) | Verdadeiro negativo (VN) |

Fonte: (Silva et al., 2019)[22]

Acurácia: indica um desempenho geral do modelo. Dentre todas as classificações, quantas o modelo classificou corretamente, dado por (Silva et al., 2019)[22]:

$$\frac{VP + VN}{VP + VN + FP + FN} \qquad (22)$$

## 2.8 Pseudo $R^2$

O Pseudo $R^2$ de acordo com Field (2009)[9], é a redução proporcional no valor absoluto da medida verossimilhança-log e, portanto, é uma medida de quanto a não-aderência aumenta com o resultado da inclusão de uma variável preditora. Seu intervalor resposta é [0,1], quando valor igual a 0 indica que os preditores não agrega valor na predição da variável de saída e valor igual a 1 nosso modelo consegue mensurar de forma exata a variável de saída.

O $R^2$cs de Cox e Snell, é baseado na verossimilhança-log do modelo (VL(novo)) e a verossimilhança-log do modelo original (VL(nulo)) e o tamanho da amostra, n (Field, 2009)[9]:

$$R^2cs = 1 - e^{[-2/n(VL(novo)-VL(nulo))]}, \qquad (23)$$

Contudo, essa estatística nunca alcança o seu valor teórico máximo, 1. Portanto, Nagelkerke (1991)[18] surgiu a seguinte correção de Nagelkerke (Field, 2009)[9]:

$$R_N^2 = \frac{R^2cs}{1 - e^{[2(VL(bsico)/n)]}} \qquad (24)$$

Embora os resultados e as maneiras de calcular sejam distintas podemos considerar elas praticamente iguais pois sua interpretação pode ser comparada ao $R^2$ da regressão linear afim de verificar o quanto o modelo está aderente. Para esse estudo vamos utilizar o $R_N^2$ de Nagelkerke.

## 2.9 MULTICOLINEARIDADE

A multicolinearidade pode ser detectada de várias maneiras. Duas medidas mais comumente utilizadas são o valor de tolerância ou seu inverso, chamada de fatores de inflação da variância (VIF) que vamos utilizar nesse estudo definido pela equação (Miloca et al., 2008)[17]:

$$VIF_i = \frac{1}{1 - R_i^2}, \qquad (25)$$

A (VIF) é uma medida do grau em que cada variável independente é explicada pelas demais variáveis independentes. Quanto maior for o fator de inflação da variância, mais severa será multicolinearidade. Quando o valor de inflação da variância exceder 10, então a multicolinearidade causar a efeitos nos coeficientes de regressão. O problema de multicolinearidade torna a estimativa dos parâmetros imprecisa, por conta de um alto valor do erro padrão, o que não é conveniente estatisticamente (Kutner et al., 2005)[15] e (Tamhane et al., 2000)[25].

Algumas formas de resolver o problema de multicolinearidade é:

1 - Excluir uma ou mais variáveis independentes altamente correlacionadas e identificar outras variáveis independentes para ajudar na previsão. Esse procedimento deve ser feito com cautela pois, neste caso, há o descarte de informações, contida nas variáveis removidas;

2 - Usar o modelo com variáveis independentes altamente correlacionadas apenas para ´ previsão, ou seja, não interpretar os coeficientes de regressão;

3 - Usar as correlações simples entre cada variável independente e a dependente para ´ compreender a relação entre variáveis independentes e dependente;

4 - Usar um método mais sofisticado de análise como a regressão Bayesiana ou a regressão sobre componentes principais para obter um modelo que reflita mais claramente os efeitos simples das variáveis independentes.

Para a realização das análises foi utilizado o software estatístico R. O nível de significância considerado foi 0,05.

## 2.10 CURVA ROC

A curva ROC (Receiver Operating Characteristic) é uma abordagem gráfica no qual os seus eixos são representados pela fração de verdadeiros positivos de um classificador (sensibilidade), no eixo das ordenadas, e pela fração de falsos positivos (1-especificidade), no eixo das abcissas, em que cada ponto é gerado por um valor limite diferente, ou seja, ponto de corte (Goksuluk et al., 2016)[13]. Entende-se por sensibilidade como sendo a proporção de positivos que foram identificados corretamente e por especificidade como sendo a proporção de negativos que foram identificados corretamente. Uma das principais tarefas passa por determinar o valor ideal do ponto de corte que corresponda aos valores razoáveis de FVP (fração de verdadeiros positivos) e FFP (fração de falsos positivos). Assim, recorre-se à curva ROC de modo a encontrar o ponto de corte ideal localizado na curva, que é o ponto mais próximo do canto superior esquerdo. Contudo, esse ponto de corte ideal nem sempre corresponde àquele em que se maximiza a sensibilidade e a especificidade (Goksuluk et al., 2016)[13].

Na curva ROC o critério de avaliação por Favero (2009)[12] obedece às seguintes faixas de avaliação para AUC que mede toda a área bidimensional abaixo de toda a curva ROC:

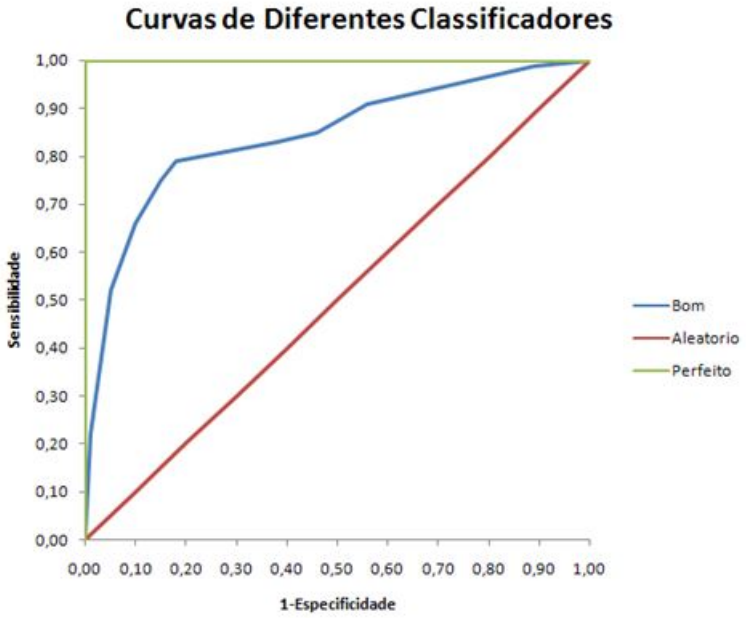
Valores menores ou igual a 0,5 não existe nenhuma discriminação;

Valores entre 0,7 e 0,8 tem discriminação aceitável.

Valores acima de 0,8 tem boa discriminação.

Vê-se na Figura 2.1 a ilustração por Souza (2009)[23] da Curva ROC.

Figura 2.1: Modelo de curva ROC com suas classificações.




Fonte:(Souza, 2009)[23]

Como comentado anteriormente as definições de Sensibilidade e Especificidade segue a forma de cálculo abaixo: Fórmula de cálculo de Sensibilidade:

$$Sensibilidade = \frac{PV}{PV + FN} \qquad (26)$$

Fórmula de cálculo de Especificidade:

$$Especificidade = \frac{VN}{VN + FP} \qquad (27)$$

# 3. Resultados e Discussões

De acordo com os dados coletados no período estudado, 3.149 pessoas foram hospitalizadas e dessas, 784 foram a óbito, resultando em uma prevalência de 24,89% em Uberlândia-MG.

Tem-se observado na literatura que a idade é um importante fator da causa de morte de pacientes acometidos pela Covid-19. No presente trabalho foi possível observar que a média de idade das pessoas que vieram à óbito é maior que 70 anos (Tabela 3.1).

Tabela 3.1: Estatísticas descritivas da variável idade.

| **Parâmetro** | **Idade Hospitalizada** | **Idade Óbitos** |
|---|---|---|
| Valor Máximo | 104 | 101 |
| Valor Mínimo | 0 | 16 |
| Mediana | 60 | 72 |
| Média | 59,35 | 70,84 |
| Coeficiente de Variação | 30,37 | 20,46 |
| Variância | 325,08 | 210,25 |
| Desvio Padrão | 18,03 | 14,5 |

Adultos mais velhos apresentam mais fatores de riscos esse resultado reflete o que é encontrado na literatura por Rezende (2020)[21], cuja porcentagem de mortalidade é maior em pacientes idosos, onde no Brasil na 2ª semana epidemiológica de 2021, a faixa etária com mais óbitos foram de 80 a 89 anos (Brasil, 2021)[3].

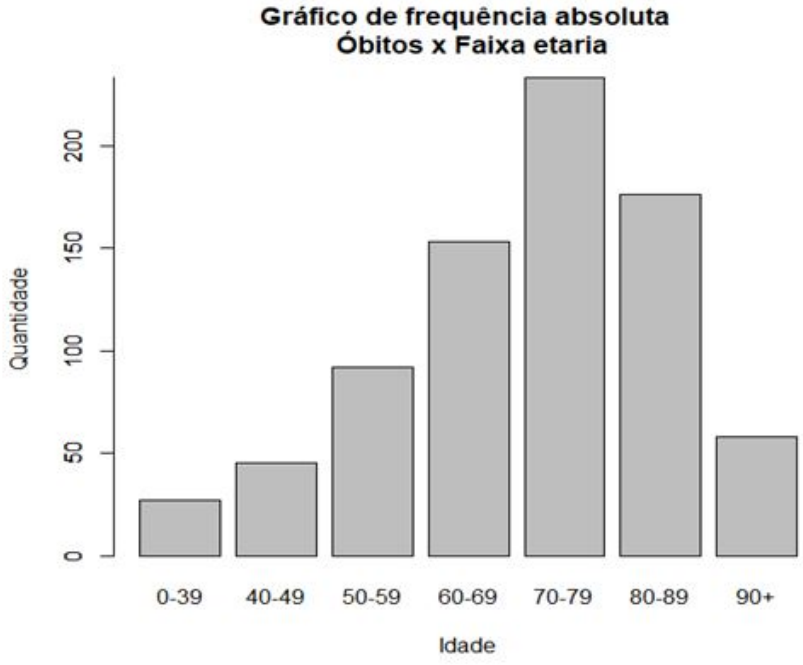
Analisando a idade dos pacientes hospitalizados em Uberlândia (Tabela 3.1) pode-se observar que a idade tem o mínimo de 0 ano até o máximo 104 anos com média de 59,35 anos. Verifica-se que os pacientes que vieram a óbito, cuja idade máxima foi de 101 anos e a mínima de 16, apresentaram média de idade (70,84 anos). Na (Tabela 3.2) será verificado a frequência de casos hospitalizados e casos hospitalizados que vieram a óbito.

Tabela 3.2: Frequências de hospitalizados e hospitalizados que vieram a óbito.

| Idade | Hospitalizados | Vieram a óbito |
|---|---|---|
| 0-39 | 481 (15,2%) | 27 (3,4%) |
| 40-49 | 433 (13,7%) | 45 (5,7%) |
| 50-59 | 621 (19,7%) | 92 (11,7%) |
| 60-69 | 636 (20,2%) | 153 (19,5%) |
| 70-79 | 538 (17,1%) | 233 (29,7%) |
| 80-89 | 348 (11,0%) | 176 (22,4%) |
| 90+ | 92 (2,9%) | 58 (7,4%) |
| Total | 3149 | 784 |

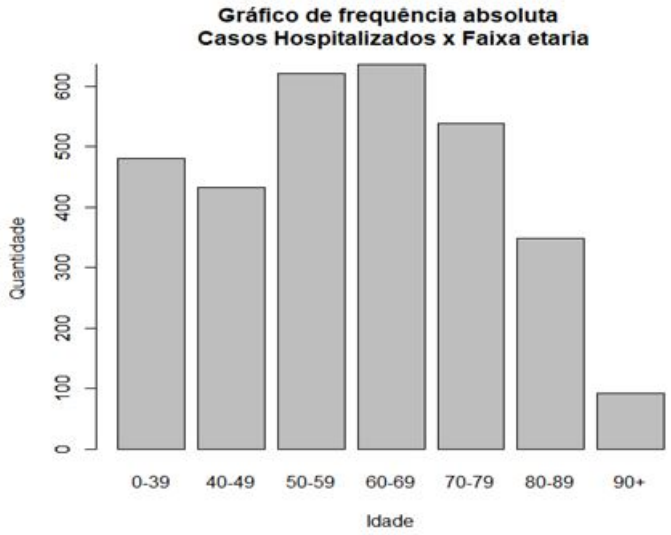
De acordo com a variável idade por faixa etária (Tabela 3.2), observa-se que as faixas etárias mais atingidas por internações foram pessoas na faixa de 60 a 69 anos com 20,2% e logo em seguida da faixa de 50 a 59 anos com 19,7%. Porém, quando se olham as faixas de idade que mais pessoas foram a óbito, é a faixa de 70 a 79 anos de idade com 29,7% dos óbitos. Para melhor visualização, na Figura 3.1 observa-se o número de internações em frequência absoluta por faixa de idade comparado a Figura 3.2 mortes por coronavirus hospitalizados.

Figura 3.1: Histograma de frequência por faixa de idade por pessoas hospitalizadas que vieram a óbito




Fonte: Autoria própria 2021

Figura 3.2: Histograma de frequência por faixa de idade por pessoas hospitalizadas.




Fonte: Autoria própria 2021

Pode-se observar claramente na Figura 3.2 que o número de obitos para as faixas de 0 a 69 anos não mantem a mesma distribuição que os casos hospitalizados da Figura 3.1, assim, podemos inferir que a morte para pessoas com idade até 69 anos é menor que pessoas acima de 69 anos.

Para realizar a comparação de proporção de faixas etárias dentro e entre os grupos, foi utilizado o teste de comparação múltipla de proporções (BIASE FERREIRA, 2009)[2].

Para a interpretação da (Tabela 3.3) temos uma matriz cruzada cujo na diagonal principal vê-se as proporções de óbitos por faixa etária, acima da diagonal a estimativa da estatística e abaixo da diagonal principal os p-valores de comparações estatísticas entre os grupos de idade.

Tabela 3.3: Matriz de proporção de mortes por Covid-19 dentro de cada faixa etária e suas diferenças estatísticas.

| | **0-39** | **40-49** | **50-59** | **60-69** | **70-79** | **80-89** | **90+** |
|---|---|---|---|---|---|---|---|
| **0-39** | 5,61% | 2,78 | 12,27 | 49,82 | 192,99 | 218,30 | 136,22 |
| **40-49** | 0,835 | 10,39% | 2,67 | 25,72 | 139,01 | 166,60 | 112,49 |
| **50-59** | 0,056 | 0,849 | 14,81% | 14,35 | 125,17 | 152,52 | 99,68 |
| **60-69** | 0,000 | 0,000 | 0,026 | 24,06% | 57,77 | 84,59 | 65,34 |
| **70-79** | 0,000 | 0,000 | 0,000 | 0,000 | 43,31% | 5,97 | 16,36 |
| **80-89** | 0,000 | 0,000 | 0,000 | 0,000 | 0,427 | 50,57% | 6,05 |
| **90+** | 0,000 | 0,000 | 0,000 | 0,000 | 0,012 | 0,418 | 63,04% |

Nível de significância estatística a 5%.

Avaliando a proporção de infectados por óbito dentro de cada faixa etária (Tabela 3.3), observa-se que o percentual de morte cresce consideravelmente com o aumento da faixa etária das pessoas hospitalizadas. Verifica-se que pessoas de 0 a 39 anos tem (5,61%) de probabilidade de vir a óbito, já a faixa etária mais crítica, são os idosos a partir de 90 anos com probabilidade de óbito de (63,04%) dos casos de entrada em alguma unidade de saúde para internação. Realizado o teste de comparações de proporções múltiplas afim de confirmar na (Tabela 3.4) se existe diferença estatística entre as faixas etárias constatamos que nos grupos dentre as faixas etárias de 0-59 anos não houve diferença estatística com alfa a 5% de significância estatística, o grupo de 60-69 anos foi diferente dos demais grupos, o grupo de 70-79 foi igual ao grupo de 80-90 porém a faixa etária de 90 anos ou mais ficou igual estatisticamente de 80-89, segue abaixo uma tabela de agrupamento de grupos para melhor visualização.

Tabela 3.4: Agrupamento por faixas etárias de idade.

| | | | | |
|---|---|---|---|---|
| **0-39** | A | | | |
| **40-49** | A | | | |
| **50-59** | A | | | |
| **60-69** | | B | | |
| **70-79** | | | C | |
| **80-89** | | | C | D |
| **90+** | | | | D |

Observação: Faixas etárias com a mesma letra são estatisticamente iguais.

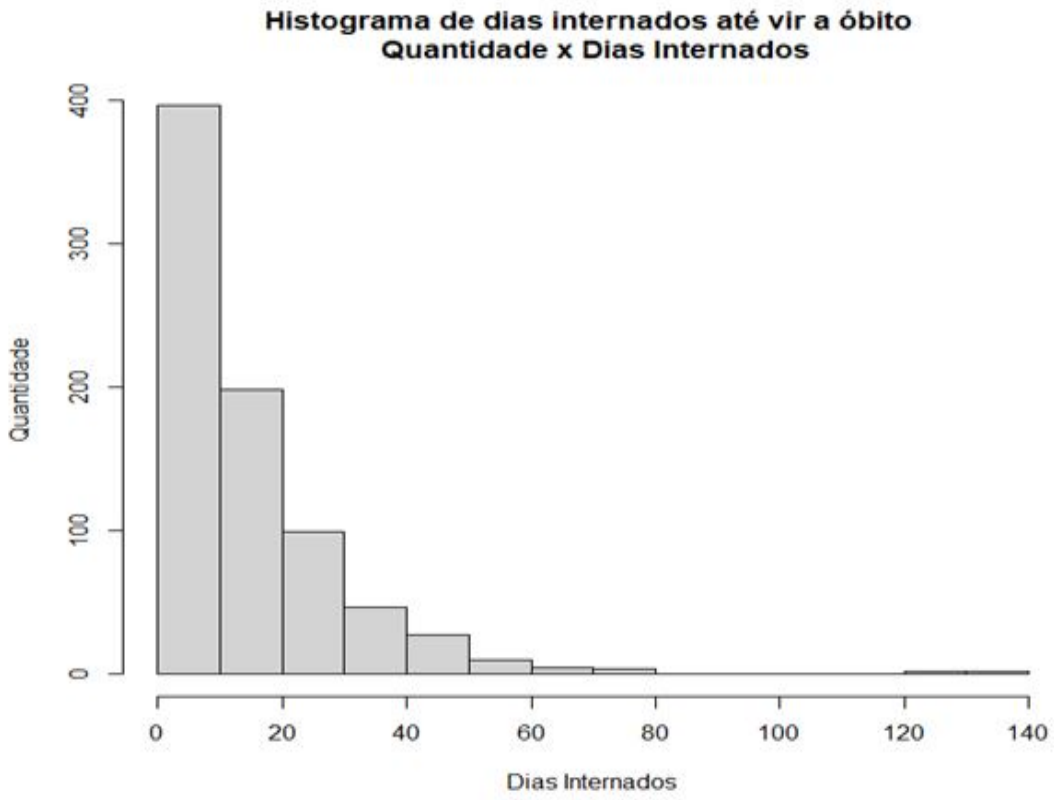
Analisando as estatísticas descritivas de internação de tempo em dias dos pacientes que vieram a óbito em Uberlândia (Tabela 3.5), pode-se observar que a média de tempo de internações até o paciente vir a óbito é 14,42 dias e a mediana igual a 10, o tempo máximo observado foi de 140 dias internado e o mínimo de 0 dias, com o paciente dando entrada no hospital e indo a óbito no mesmo dia. Quando comparamos com os números da China em março de 2020 início da pandemia, o tempo médio de letalidade do vírus é de 17,8 dias (Verity et al., 2020)[26]. Podemos confirmar que o tempo de letalidade do vírus em Uberlândia é menor que na China em seu início de pandemia.

Tabela 3.5: Estatísticas descritivas de tempo em dias de internação até óbito.

| **Parâmetro** | **Tempo até óbito** |
|---|---|
| Valor Máximo | 140 |
| Valor Mínimo | 0 |
| Mediana | 10 |
| Média | 14,42 |
| Variância | 212,28 |
| Desvio Padrão | 14,57 |

Para melhor visualização, na Figura 3.3 é ilustrado o histograma de frequência de tempo em dias de internação até o óbito do paciente. Nota-se que 50% dos pacientes internados ficaram até no 10 dias internados até vir a óbito, acima de 80 dias apenas 2 pessoas ficaram internadas até o óbito.

Figura 3.3: Histograma de frequência de tempo em dias de internação até o óbito do paciente.




Fonte: Autoria própria

Podemos constatar na (Tabela 3.6) abaixo que os fatores de risco que tem maior frequência nos casos hospitalizados foram dispneia (70,66%), tosse (68,43%), saturação $< 95\%$ (67,70%), desconforto respiratório (63%), febre (57.48%), doenças cardíacas (40,74%) e diabetes (24,33%). Comparando esses achados com o estudo do (CDC, 2020)[6] nos Estados Unidos, os principais fatores de risco que aparecem nos infectados aparecem também comumente em Uberlândia.

Tabela 3.6: Fatores com maiores frequências nos casos hospitalizados.

| **Fatores** | **Qtd. Fatores** | **%. Com fatores** |
|---|---|---|
| Dispneia | 2225 | 70,66% |
| Tosse | 2155 | 68,43% |
| Saturação <95% | 2132 | 67,70% |
| Desconforto Respiratório | 1984 | 63,00% |
| Febre | 1810 | 57,48% |
| Cardiopatia | 1283 | 40,74% |
| Diabetes | 766 | 24,33% |
| Vacina | 681 | 21,63% |
| Internado em UTI | 590 | 18,74% |
| Diarreia | 547 | 17,37% |
| Garganta | 506 | 16,07% |
| Vômito | 387 | 12,29% |
| Obesidade | 333 | 10,57% |
| Doença Neurológica | 189 | 6,00% |
| Pneumopatia | 174 | 5,53% |
| Síndrome Gripal | 171 | 5,43% |
| Doença Renal | 160 | 5,08% |
| Imunodepressão | 117 | 3,72% |
| Asma | 100 | 3,18% |
| Hematologia | 26 | 0,83% |
| Doença Hepática | 14 | 0,44% |
| Nosocomial | 11 | 0,35% |
| Contato Ave e Suíno | 10 | 0,32% |
| Síndrome de Down | 6 | 0,19% |

Base: 3149 pessoas

Vê-se na (Tabela 3.7) que o percentual de pessoas que obtinha o fator de risco específico em relação à quantidade de pessoas que foram hospitalizadas pelo mesmo critério em Uberlândia, em nosso estudo constatamos que quando se tem doença hepática crônica, o risco de morte em relação aos que também tem é de (78,57%), e foi o maior entre os fatores, seguidos de nosocomial (72,73%), Imunodepressão (65.81%) e doença neurológica (61,90%). Comparando com a 2ª semana epidemiológica de 2021, entre os 1.019 óbitos de SRAG por covid-19 notificados, 630 (61,8%) apresentavam pelo menos uma comorbidade. Cardiopatia e diabetes foram às condições mais frequentes, sendo que a maior parte destes indivíduos que evoluiu a óbito e apresentava alguma comorbidade possuía 60 anos ou mais de idade (Brasil, 2021)[3].

Tabela 3.7: Razão por fatores de casos hospitalizados versos óbitos.

| Fatores | Qtd. Óbitos | %. Óbitos |
|---|---|---|
| Dispneia | 610 | 27,42% |
| Tosse | 528 | 24,50% |
| Saturação <95% | 643 | 30,16% |
| Desconforto Respiratório | 573 | 28,88% |
| Febre | 427 | 23,59% |
| Cardiopatia | 457 | 35,62% |
| Diabetes | 250 | 32,64% |
| Vacina | 167 | 24,52% |
| Internado em UTI | 263 | 44,58% |
| Diarreia | 135 | 24,68% |
| Garganta | 95 | 18,77% |
| Vômito | 83 | 21,45% |
| Obesidade | 109 | 32,73% |
| Doença Neurológica | 117 | 61,90% |
| Pneumopatia | 97 | 55,75% |
| Síndrome Gripal | 57 | 33,33% |
| Doença Renal | 89 | 55,63% |
| Imunodepressão | 77 | 65,81% |
| Asma | 29 | 29,00% |
| Hematologia | 13 | 50,00% |
| Doença Hepática | 11 | 78,57% |
| Nosocomial | 8 | 72,73% |
| Contato Ave e Suíno | 4 | 40,00% |
| Síndrome de Down | 2 | 33,33% |

Base: 784 pessoas

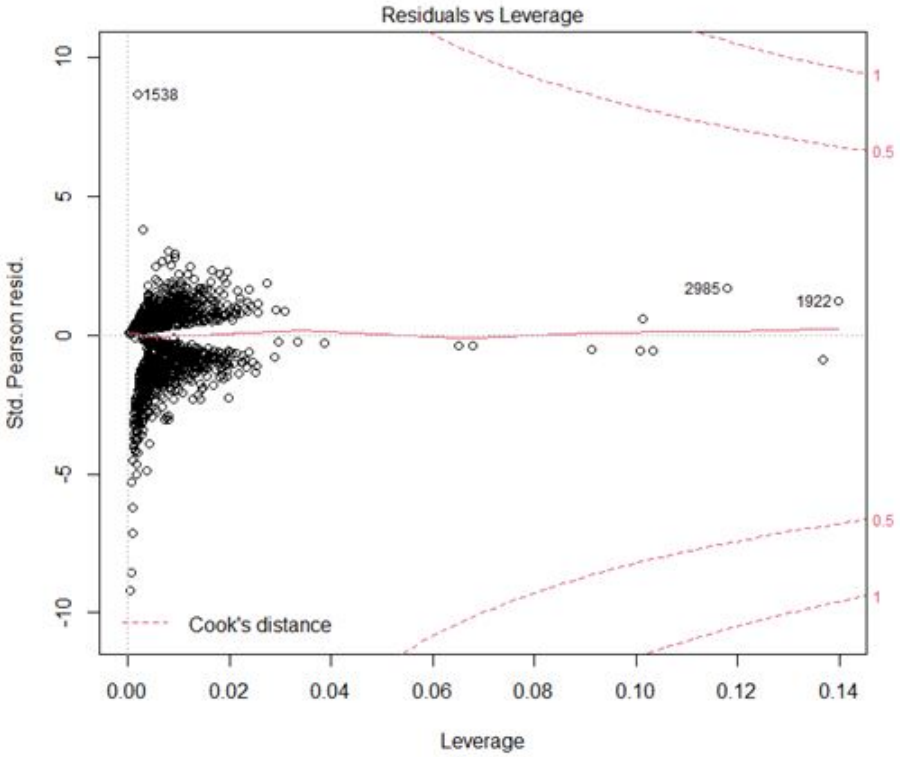
Para a realização da regressão logística visando identificar quais variáveis são fatores de risco de morte para pacientes acometidos pela Covid-19, foram analisadas as pressuposições do modelo para a realização das análises.

No primeiro momento foi verificado se o modelo possui outliers ou pontos de alavancagem analisando o gráfico dos resíduos padronizados (Figura 3.4), verificamos que não existem pontos de alavancagem e nem outliers. Portanto não existem valores ultrapassando os limites de Cooks distance, e dessa forma, seguiu-se com o modelo estudado.

Na avaliação da pressuposição de multicolinearidade entre as variáveis não dicotômicas (Tabela 3.8), foi utilizado o critério de análise de VIF (fator de inflação da Variância).

Foi constatado apartir os dados da (Tabela 3.8), que todos os valores de VIF estão abaixo de 5, indicando que não existe multicolinearidade entre as variáveis. Quando os valores acima de 10, indica que tem alta correlação e existe multicoliearidade entre as variáveis.

Figura 3.4: Verificação de pontos de alavancagem dos resíduos padronizados com Cooks distance.




Fonte: Autoria própria

Tabela 3.8: Valores de VIF para as variáveis do modelo final.

| **Fator Influência** | **VIF** |
|---|---|
| Sexo | 1,053 |
| Saturação <95% | 1,180 |
| Desconforto Respiratório | 1,181 |
| Idade | 1,188 |
| Cardiopatia | 1,105 |
| Doença Hepática | 1,005 |
| Asma | 1,035 |
| Doença Neurológica | 1,037 |
| Pneumopatia | 1,017 |
| Imunodepressão | 1,023 |
| Doença Renal | 1,011 |
| Obesidade | 1,059 |
| Vacina | 1,024 |
| Internado em UTI | 1,018 |

Após verificadas as pressuposições do modelo, foi realizada a regressão logística para avaliar quais variáveis estatísticas foram significativas e com isso, definir o modelo final com as variáveis significativas (Tabela 3.9).

Tabela 3.9: Estatística representando o ajuste de modelo de regressão logística com as variáveis de fatores significativas de modelo.

| **Parâmetros** | **Estimativa** | **Erro Padrão** | **Wald** | **Valor p** | **Pseudo $R^2$** |
|---|---|---|---|---|---|
| (Intercept) | -6,150 | 0,278 | -22,099 | 0,000 | 0,364 |
| Sexo | 0,527 | 0,127 | 4,160 | 0,000 | |
| Idade | 0,053 | 0,004 | 14,824 | 0,000 | |
| Desc. Respiratório | 0,329 | 0,114 | 2,883 | 0.003 | |
| Saturação <95% | 0,585 | 0,102 | 5,743 | 0,000 | |
| Cardiopatia | 0,412 | 0,102 | 4,022 | 0,000 | |
| Doença Hepática | 2,063 | 0,752 | 2,742 | 0.006 | |
| Asma | 0,746 | 0,273 | 2,732 | 0.006 | |
| Doença Neurológica | 1,166 | 0,180 | 6,468 | 0,000 | |
| Pneumopatia | 0,725 | 0,184 | 3,946 | 0,000 | |
| Imunodepressão | 1,869 | 0,242 | 7,721 | 0,000 | |
| Doença Renal | 0,871 | 0,194 | 4,479 | 0,000 | |
| Obesidade | 0,583 | 0,151 | 3,847 | 0.000 | |
| Vacina | -0,326 | 0,118 | -2,760 | 0.005 | |
| Internado em UTI | 0,915 | 0,114 | 8,060 | 0,000 | |

O Pseudo $R^2$ de Nagelkerke foi usado para verificar a qualidade do modelo (Tabela 3.9). Para um bom ajuste do modelo os valores devem estar entre 0,2 e 0,4 (Louviere et al., 2000)[16]. No modelo considerado somente com as variáveis que foram significativas, o Pseudo $R^2$ foi de 0,364, indicando boa qualidade de ajuste.

O modelo final foi composto pelas vaiáveis sexo, pacientes com desconforto respiratório, saturação $< 95\%$, idade, doença cardiovascular crônica, doença hepática crônica, asma, Doença neurológica, doença pneumopatia crônica, imunodepressão, doença renal crônica, obesidade, vacina para e internados em UTI.

Todas essas variáveis foram significativas pelo teste de Wald (p-valor<0,05). As variáveis analisadas que não foram significativas foram retiradas do modelo, foi inserido na (Tabela 3.10) as codificações de $x_n$ juntamente com os nomes das variáveis para visualizar a equação do modelo final (28).

Tabela 3.10: Razão de chances dos principais fatores de óbitos em casos hospitalizados.

| | **Odds Ratio** | **IC 95%** | |
|---|---|---|---|
| (Intercept) | - | - | - |
| Sexo (x1) | 1,795 | 1,470 | 2,192 |
| Idade (x2) | 1,055 | 1,047 | 1,062 |
| Desc. Respiratório (x10) | 1,389 | 1,111 | 1,737 |
| Saturação <95% (x11) | 1,693 | 1,321 | 2,170 |
| Cardiopatia (x14) | 1,509 | 1,235 | 1,845 |
| Doença Hepática (x17) | 7,866 | 1,801 | 34,362 |
| Asma (x18) | 2,108 | 1,235 | 3,601 |
| Doença Neurológica (x20) | 3,209 | 2,254 | 4,568 |
| Pneumopatia (x21) | 2,065 | 1,440 | 2,960 |
| Imunodepressão (x22) | 6,482 | 4,033 | 10,417 |
| Doença Renal (x23) | 2,390 | 1,632 | 3,498 |
| Obesidade (x24) | 1,791 | 1,331 | 2,410 |
| Vacina (x25) | 0,722 | 0,573 | 0,910 |
| Internado em UTI (x26) | 2,498 | 1,999 | 3,121 |

Com 3.149 pessoas hospitalizadas e dessas 784 foram a óbito com prevalência de 24,89% dos totais dos casos em Uberlândia-MG. Os seguintes fatores de riscos foram evidenciados na (Tabela 3.10): Na variável Sexo homens tem maior risco de morte que mulheres com (OR = 1,79); Saturação abaixo de 95% (OR = 1,69); Desconforto ao respirar (OR = 1,38); Idade tem 5,2% mais risco de morte quanto maior sua idade (OR = 1,05); Doença Cardiovascular (OR = 1,5); Doença Hepática teve o maior fator de risco para morte (OR = 7,86); Asma (OR = 2,10); Doença Neurológica (OR = 3,20); Pneumopatia crônica (OR = 2,06); Pessoa com Imunidade baixa e Imunodepressão tem o segundo maior fator de risco para óbito com (OR = 6,48); Doença Renal (OR = 2,39); Obesidade (OR = 1,79); Pessoas que tomaram a Vacina da gripe tem um risco de 27,8% a menos de óbito com (OR = 0,72) sendo assim entre todas os fatores significativos um fator de proteção; Pessoas que foram levadas a UTI têm 2,5 vezes mais chances de óbito que pessoas que não foram a UTI. Após mapear os fatores de risco temos o modelo logístico múltiplo:

$$ln\left[\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right] = -6,15 + 0,52x1 + 0,05x2 + 0,32x10 + 0,58x11 + 0,41x14 - 2,06x17 +$$

$$+ 0{,}75\text{x}18 + 1{,}17\text{x}20 + 0{,}73\text{x}21 + 1{,}87\text{x}22 + 0{,}87\text{x}23 + 0{,}58\text{x}24 - 0{,}33\text{x}25 +$$
$$+ 0{,}92\text{x}26. \qquad (28)$$

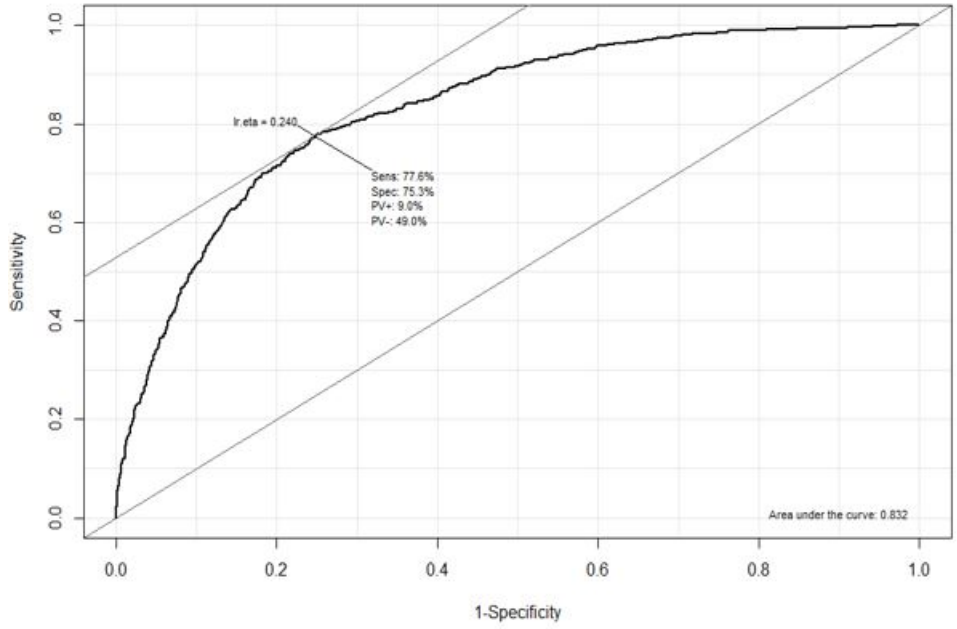
Após a constatar o grau de aderência do modelo verificou-se na matriz de confusão a acurácia do modelo:

Tabela 3.11: Matriz de Confusão

| **Valor Estimado** | **Valor Observado** | | **Total** |
|---|---|---|---|
| | Óbito | Cura | |
| Óbito | 609 | 590 | 1199 |
| Cura | 175 | 1775 | 1950 |
| **Total** | 784 | 2365 | 3149 |

A acurácia do modelo no presente estudo foi de 75,70%, sendo assim considerado um bom poder de previsão. Através do package Epi em software R na função ROC, obteve-se o Gráfico da curva ROC com os valores de AUC de 83,2%, garantindo uma boa discriminação para o modelo, com sensibilidade de (77,6%) e especificidade de (75,3%) para garantir o melhor ajuste, e o cutoff do modelo foi ajustado em 0,24, gerando assim o mais assertivo ponto de corte para o modelo estudado.

Figura 3.5: Curva ROC com valores de Sensibilidade, Especificidade e AUC.




Fonte: Autoria própria

Com o avanço da pandemia atingindo pessoas de todas as idades e com diversos fatores naturais e de saúde, o presente estudo pode servir de base para políticas de saúde públicas para preparar o sistema de atendimento hospitalar, mapeando e entendo quais são o volume de pessoas que apresentam fatores de riscos apresentados, como doenças hepáticas, imunodepressão, neurológica e renal que foram os de maiores chances de óbitos apresentados, assim fazendo um acompanhamento frequente sobre sua condição de saúde e realizando conscientização através de canais de comunicação direta sobre os cuidados de saúde para evitar o contato com o coronavirus e priorizando leitos para pessoas com esse histórico de enfermidade.

# 4. CONCLUSÕES

Esse estudo mostrou que a regressão logística pode ser utilizada para classificar os fatores de riscos associados ao óbito pelo coronavirus em casos hospitalizados na cidade de Uberlândia-MG, e por meio da matriz de confusão foi constatado 75,70% de acurácia do modelo. Foi verificado que os maiores fatores de risco para óbito para covid-19 são: doença hepática crônica (OR=7,86; p< 0,006), Imunodepressão (OR=6,48; p<0,000), doença neurológica crônica (OR=3,2; p<0,000), Internação em UTI (OR=2,49; p<0,000), doença renal crônica, (OR=2,39; p<0,000) e pneumopatia crônica (OR=2,06; p<0,000). O único fator de proteção encontrado nas variáveis significativas foi Vacina contra a gripe com (OR=0,72; p<0,005).

# Referências Bibliográficas

[1] Barreto, A. S.: *Modelos de Regressão: Teoria e Aplicação com o Programa Estatístico*. Edição do Autor, 1ª ed., 2011.

[2] Biase, N. G. e Ferreira, D. F.: *Comparações múltiplas e testes simultâneos para parâmetros binomiais de K populações independentes*. Revista Brasileira de Biometria, 27(3):301–323, 2009.

[3] Brasil, M. S.: *Boletim epidemiológico nº 46. Secretaria da Vigilância em Saúde. Ministério da Saúde.*, 2021. https://www.gov.br/saude/pt-br/media/pdf/2021/janeiro/22/boletim_epidemiologico_covid_46-final.pdf, acessado em 20/03/2021.

[4] Camargo, E. B.: *Nota rápida de evidência: observações sobre condições de risco para o agravamento ou morte por COVID 19*. 2020. https://www.arca.fiocruz.br/handle/icict/42575.

[5] Cavalcante, J. A.: *COVID-19 no Brasil: evolução da epidemia até a semana epidemiológica 20 de 2020.* Epidemiol Serv Saude, p. 5, 2020. http://dx.doi.org/10.5123/s1679-49742020000400010.

[6] CDC: *Preliminary Estimates of the Prevalence of Selected Underlying Health Conditions Among Patients with Coronavirus Disease 2019 United States, February 12March 28, 2020, Centers for Disease Control and Prevention*, 2020. http://dx.doi.org/10.15585/mmwr.mm6913e2, acessado em 17/03/2021.

[7] Demétrio, C. G. B.: *Modelos Lineares Generalizados em Experimentação Agronômica.* 2002. http://www.lce.esalq.usp.br/clarice/Apostila.pdf.

[8] Docs.Uberlândia: *Comunicado COVID-19, Uberlandia.mg.gov*, 2020. http://docs.uberlandia.mg.gov.br/wpcontent/uploads/2020/03/Comunicado-Comitê-COVID-19-17.03.2020.pdf, acessado em 08/10/2020.

[9] Field, A.: *Descobrindo a estatística usando SPSS.* Artmed, 2ª ed., 2009.

[10] Figueira, C. A.: *Modelos de regressão logística.* Mestrado em Matemática, 2006.

[11] Freitas, A. R. R.: *Análise da gravidade da pandemia de Covid-19.* Epidemiol Serv Saude, p. 1, 2020. https://doi.org/10.5123/S1679-49742020000200008.

[12] Fávero, L. P. L., Belfiore, P. P., Silva, F. L. da e Chan, B. L.: *Análise de dados: modelagem multivariada para tomada de decisões*. Campos, 1ª ed., 2009.

[13] Goksuluk, D., Korkmaz, S., Zararsiz, G. e Karaagaoglu, A. E.: *easyROC: An Interactive Web-tool for ROC Curve Analysis Using R Language Environment*. The R Journal, 8(2):213–230, 2016. https://doi.org/10.32614/RJ-2016-042.

[14] Hosmer, D. W. e Lemeshow, S.: *Applied Logistic Regression*. John Wiley, 2ª ed., 2000.

[15] Kutner, M. H. e al. et: *M Applied linear models*. McGraw-Hill Irwin, 5ª ed., 2005.

[16] Louviere, J. J., Hensher, D. A. e Swait, J. D.: *Stated choice methods: Analysis and Applications*. Cambridge University Press, 1ª ed., 2000.

[17] Miloca, S. A. e Conejo, P. D.: *Multicolinearidade em Modelos de Regressao*. Universidade Estadual do Oeste do Parana, 2008. https://www.ime.usp.br/~yambar/MI404-Metodos%20Estatisticos/Aula%208-9%20Regress%E3o%20mult%20dim/inete%20adicional%20-%20multicolinearidade%20em%20modelos%20de%20regressao.pdf.

[18] Nagelkerke, N.: *A note on a general definition of the coefficient of determination*. Biometrika, 78(3):691–692, 1991. https://www.jstor.org/stable/2337038.

[19] Open.Data.Sus: *Ficha de registro individual de casos de sindrome respiratória aguda grave hospitalizados*, 2020. https://opendatasus.saude.gov.br/dataset/bd-srag-2020, acessado em 29/07/2020.

[20] Rafael, R. M. R.: *Epidemiologia, políticas públicas e Covid-19: o que esperar no Brasil?* Resvista Enfermagem Uerj, p. 1, 2020. https://doi.org/10.12957/reuerj.2020.49570.

[21] Rezende, L. F. M.: *Adults at high-risk of severe coronavirus disease-2019 (Covid-19) in Brazil*. Revista Saúde Publica, p. 4, 2020. https://doi.org/10.11606/s1518-8787.2020054002596.

[22] Silva, L. C. C., Silva, A. W. S., Filho, A. N. e Filho, A. O. B.: *Estudo Comparativo de Métodos de Aprendizagem de Máquina Aplicados em Sistemas de Detecção de Intrusão*. Escola regional de computação ceara, p. 5, 2019. https://sol.sbc.org.br/index.php/ercemapi/article/view/8855/8756.

[23] Souza, C.: *Análise de Poder Discriminativo Através de Curvas ROC*, 2009. http://crsouza.com/2009/07/13/analise-de-poder-discriminativo-atraves-de-curvas-roc/, acessado em 13/04/2021.

[24] Taconeli, C.: *Testes de hipóteses e intervalo de confiança em modelos lineares generalizados*. Edição do Autor, 2015. https://docs.ufpr.br/~taconeli/CE225/Aula9.pdf.

[25] Tamhane, A. C. e Dunlop, D. D.: *Statistics and Data Analysis: from elementary to intermediate*. Pearson, 1ª ed., 2000.

[26] Verity, R. et al.: *Estimates of the severity of coronavirus disease 2019: a model-based analysis*, 2020. https://doi.org/10.1016/S1473-3099(20)30243-7, acessado em 05/04/2021.

# A. APÊNDICE

Regressão Logística

Passo 1: Carregamento dos pacotes que serão utilizados Caso não instalados utilizar

if(!require(pacman)) install.packages("pacman")

library(pacman)

pacman::p$_l oad(dplyr, psych, car, MASS, DescTools, QuantPsyc, ggplot2)$

install.packages("pROC")

install.packages("ISwR")

install.packages("caret")

install.packages("generics")

install.packages("sampling")

install.packages("descr")

install.packages("ROCR")

install.packages("Epi")

install.packages("reshape2")

library(reshape2)

library(Epi)

library(ROCR)

library(descr)

library(rpart)

library(survival)

library(sampling)

library(ISwR)

library(dplyr)

library(readxl)

library(pROC)

library(caret)

library(generics)

Passo 2: Carregar o banco de dados

Selecionar o diretório de trabalho (working directory)

Diretorio: Session > Set Working Directory > Choose Directory

dados<-read.table("CORO.txt",sep="",header=T,dec=",")

Vizualizar dados:

View(dados)

Visualização de um resumo dos dados

glimpse(dados)

Passo 3: Análise das frequências das categorias da EVOLUÇÃO = ÓBITO

table(dados$EVOLUCAO)

Gráfico de óbitos por faixa etária

barplot(table(dados$CLA$_I DADE[dados\$EVOLUCAO == "1"]))$title("Frequnciasconjuntasabsolut$
"$Idade", ylab = "Quantidade")axis(2, at = seq(0, 500, by = 50))$

Passo 4: Checagem das categorias de referência nossa referência é igual óbito Y=1

levels(dados$EVOLUCAO)

Passo 5: Checagem dos pressupostos

1. Verificar se variável dependente é dicotômica (categorias mutuamente exclusivas)

2. Verificar independência das observações e consistencia dos dados (sem medidas repetidas)

Construção do modelo:

mod <- glm(EVOLUCAO IDADE + CS$_S EXO + SURTO_S G + NOSOCOMIAL + AVE_S UINO + FEBRE + TOSSE + GARGANTA + DISPNEIA + DESC_R ESP + SATURACAO + DIARREIA + VOMITO + CARDIOPATI + HEMATOLOGI + SIND_D OWN + HEPATICA + ASMA + DIABETES + NEUROLOGIC + PNEUMOPATI + IMUNODEPRE + RENAL + OBESIDADE + VACINA + UTI, family = binomial(link =' logit'), data = dados)

3. Ausência de outliers/ pontos de alavancagem

plot(mod, which = 5)

summary(stdres(mod))

4. Ausência de multicolinearidade

vif(mod)

Se VIF > 10 temos Multicolinearidade

5. Relação linear entre cada VI contínua e o logito da Variavel dependente

Interação entre a VI contínua e o seu log não significativa (Box-Tidwell)

intlog <- dados$IDADE * log(dados$IDADE)

dados$intlog <- intlog

modint <- glm(EVOLUCAO IDADE + CS$_S EXO + SURTO_S G + NOSOCOMIAL + AVE_S UINO + FEBRE + TOSSE + GARGANTA + DISPNEIA + DESC_R ESP + SATURACAO + DIARREIA + VOMITO + CARDIOPATI + HEMATOLOGI + SIND_D OWN + HEPATICA + ASMA + DIABETES + NEUROLOGIC + PNEUMOPATI + IMUNODEPRE + RENAL + OBESIDADE + VACINA + UTI + intlog, family = binomial(link =' logit'), data = dados)

summary(modint)

Cálculo do logito

logito <- mod$linear.predictors

dados$logito <- logito

Verificar gráficamente a relação linear

dev.off()

ggplot(dados, aes(logito, IDADE)) + geom$_p oint(size = 0.5, alpha = 0.5) + geom$_s mooth(method = "loess") + theme$_c lassic()$

Passo 6: Análise do modelo

Overall effects

Anova(mod, type = 'II', test = "Wald")

Efeitos específicos

summary(mod)

Obtenção de Odds Ratio (razões de chance) com IC 95% (usando log-likelihood)

OR1=exp(mod$coefficients);OR1

ICOR1=exp(cbind(OR = coef(mod), confint(mod)));ICOR1

round((cbind(OR1, ICOR1)),3)

Passo 7: Criação e análise do modelo final com as variaveis significativas

mod2 <- glm(EVOLUCAO SATURACAO + CS$_S EXO + DESC_R ESP + IDADE + CARDIOPATI + HEPATICA + ASMA + NEUROLOGIC + PNEUMOPATI + IMUNODEPRE+RENAL+OBESIDADE+VACINA+UTI, family = binomial(link =' logit'), data = dados)

Overall effects

Anova(mod2, type="II", test="Wald")

Efeitos específicos

summary(mod2)

Obtenção de Odds Ratio (razões de chance) com IC 95% (usando log-likelihood)

OR2=exp(mod2$coefficients);OR2

ORSP2=exp(cbind(OR = coef(mod2), confint.default(mod2)))

round((cbind(OR2, ORSP2)),3)

Passo 8: Avaliação da qualidade do ajuste de modelos

8.1 Pseudo R-quadrado de Nagelkerke

PseudoR2(mod2, which = "Nagelkerke")

Matriz de confusão com Acuráciae cutoff iguak a 0.24

ClassLog(mod2, dados$EVOLUCAO, 0.24)

Gerando a Curva ROC com os valores de sensibilidade e especificidade e AUC.

ROC(form=dados$EVOLUCAO dados$SATURACAO + dados$CS$_S EXO+dados$DESC_R ESP+ dados$IDADE+dados$CARDIOPATI+dados$HEPATICA+dados$ASMA+dados$NEUROLOGIC dados$PNEUMOPATI+dados$IMUNODEPRE+dados$RENAL+dados$OBESIDADE+ dados$VACINA + dados$UTI, , MI = "FALSE", plot = "ROC")