

Análise exploratória e modelagem estatística de dados clínicos em pacientes com insuficiência cardíaca

Exploratory analysis and statistical modeling of clinical data in patients with heart failure

Análisis exploratorio y modelado estadístico de datos clínicos en pacientes con insuficiencia cardíaca

Recebido: 16/02/2025 | Revisado: 23/02/2025 | Aceitado: 24/02/2025 | Publicado: 27/02/2025

Dorival Mantovani

ORCID: <https://orcid.org/0009-0001-9278-0627>
Universidade de São Paulo, Brasil
E-mail: dorival.mantovani@usp.br

Laura Maria de Araújo Pereira

ORCID: <https://orcid.org/0000-0003-1897-9475>
Faculdade Maurício de Nassau, Brasil
E-mail: lauramariabiomedicina@gmail.com

Nel Cione Gomes Soares

ORCID: <https://orcid.org/0009-0005-1919-5834>
Universidade Federal de Campina Grande, Brasil
E-mail: nelcione.gomes@estudante.ufcg.edu.br

Layza Cherollin de Lima Pereira

ORCID: <https://orcid.org/0009-0004-5475-5625>
Faculdade de Ciências Médicas da Paraíba, Brasil
E-mail: layza_cherollin@hotmail.com

Marfran José Cunha Urtiga

ORCID: <https://orcid.org/0000-0003-0611-7261>
Faculdade de Ciências Médicas da Paraíba, Brasil
E-mail: marfranjose@gmail.com

Talyta Valeria Siqueira do Monte

ORCID: <https://orcid.org/0000-0002-8730-3244>
Universidade Federal de Pernambuco, Brasil
E-mail: talyta.valeri@gmail.com

José Guedes da Silva Júnior

ORCID: <https://orcid.org/0000-0003-2032-2333>
Faculdade de Ciências Médicas da Paraíba, Brasil
E-mail: zeguedescrizant@gmail.com

Resumo

A insuficiência cardíaca é uma das principais causas de mortalidade global, o que demanda estratégias para a prevenção e o tratamento. Este estudo tem como objetivo realizar uma análise exploratória de 299 pacientes com insuficiência cardíaca e investigar a influência de diferentes variáveis clínicas por meio de modelagens estatísticas. Foram aplicados modelos estatísticos de regressão linear e logística para avaliar o impacto de variáveis como idade, creatinina sérica, sódio sérico, hipertensão, anemia e tabagismo na fração de ejeção (volume sistólico) e na mortalidade dos pacientes. Os resultados indicaram que o nível de sódio sérico tem impacto significativo na fração de ejeção, enquanto idade, creatinina sérica e fração de ejeção obtiveram uma relação estatisticamente significativa com o risco de óbito. As conclusões deste estudo estão em consonância com achados prévios da literatura, destacando a importância dos biomarcadores laboratoriais na avaliação prognóstica da insuficiência cardíaca. A identificação desses fatores pode contribuir para a tomada de decisão clínica e o desenvolvimento de abordagens terapêuticas mais direcionadas, auxiliando na melhoria da sobrevida e qualidade de vida dos pacientes afetados pela doença.

Palavras-chave: Insuficiência Cardíaca; Análise de Regressão; Volume Sistólico; Mortalidade.

Abstract

Heart failure is one of the leading causes of global mortality, requiring effective strategies for prevention and treatment. This study aims to conduct an exploratory analysis of 299 patients with heart failure and investigate the influence of different clinical variables through statistical modeling. Statistical models, including linear and logistic regression, were applied to assess the impact of variables such as age, serum creatinine, serum sodium, hypertension, anemia, and smoking on ejection fraction (stroke volume) and patient mortality. The results indicated that serum sodium levels significantly influence ejection fraction, while age, serum creatinine, and ejection fraction showed a statistically

significant relationship with mortality risk. The findings of this study are consistent with previous literature, emphasizing the importance of laboratory biomarkers in the prognostic evaluation of heart failure. Identifying these factors can contribute to clinical decision-making and the development of more targeted therapeutic approaches, ultimately improving survival and quality of life for affected patients.

Keywords: Heart Failure; Regression Analysis; Stroke Volume; Mortality.

Resumen

La insuficiencia cardíaca es una de las principales causas de mortalidad a nivel mundial, lo que requiere estrategias eficaces para su prevención y tratamiento. Este estudio tiene como objetivo realizar un análisis exploratorio de 299 pacientes con insuficiencia cardíaca e investigar la influencia de diferentes variables clínicas mediante modelizaciones estadísticas. Se aplicaron modelos estadísticos de regresión lineal y logística para evaluar el impacto de variables como la edad, la creatinina sérica, el sodio sérico, la hipertensión, la anemia y el tabaquismo en la fracción de eyección (volumen sistólico) y la mortalidad de los pacientes. Los resultados indicaron que el nivel de sodio sérico tiene un impacto significativo en la fracción de eyección, mientras que la edad, la creatinina sérica y la fracción de eyección mostraron una relación estadísticamente significativa con el riesgo de mortalidad. Las conclusiones de este estudio coinciden con hallazgos previos en la literatura, destacando la importancia de los biomarcadores de laboratorio en la evaluación pronóstica de la insuficiencia cardíaca. La identificación de estos factores puede contribuir a la toma de decisiones clínicas y al desarrollo de enfoques terapéuticos más específicos, mejorando así la supervivencia y la calidad de vida de los pacientes afectados.

Palabras clave: Insuficiencia Cardíaca; Análisis de Regresión; Volumen Sistólico; Mortalidad.

1. Introdução

As doenças cardiovasculares (DCVs) são consideradas as principais causas de óbitos no mundo, estando atribuídas a elas cerca de 17 milhões de mortes todos os anos, representando assim um terço do total de óbitos, de acordo com a Organização Mundial da Saúde (World Health Organization, 2024; Tsao et al., 2023). Diante dessas complicações de cunho mais grave, foi relacionada a insuficiência cardíaca como condição mais frequente quando associada ao óbito dos pacientes (McDonagh et al., 2021; Groenewegen et al., 2020). A insuficiência cardíaca pode ser classificada de acordo com a fração de ejeção, que é um parâmetro ecocardiográfico que mede a porcentagem de sangue que é bombeada para fora do ventrículo esquerdo a cada batimento. Esta classificação é crucial, pois cada uma das formas de insuficiência cardíaca tem fatores etiológicos, tratamentos e prognósticos específicos. Nesta situação, métodos de análise avançados de dados são necessários para detectar e comprovar padrões, e assim prever conclusões clínicas pertinentes em pacientes com doenças de cunho cardiovascular (Avezum & Maia, 2012; Mebazaa et al., 2021).

Além disso, diversos fatores de risco estão intimamente ligados aos hábitos alimentares e ao estilo de vida dos pacientes, como o consumo de tabaco, dietas inadequadas e sedentarismo (Knuuti et al., 2020). Esses aspectos exercem uma influência significativa no desenvolvimento e na progressão das DCVs. Assim, algumas medidas são direcionadas a populações vulneráveis a esta condição, visando reduzir o desenvolvimento da condição. Entretanto, para indivíduos que já possuem diagnóstico de DCVs, ou estão enquadrados como “alto risco” devido à presença de hipertensão, diabetes ou dislipidemia, a identificação precoce e o manejo clínico são necessários para melhorar o prognóstico (Précoma et al., 2019).

Este trabalho tem como objetivo realizar uma análise exploratória de 299 pacientes com insuficiência cardíaca e investigar a influência de diferentes variáveis clínicas por meio de modelagens estatísticas. Foi realizado uma regressão linear para avaliar se fatores como idade, creatinina sérica, sódio sérico, hipertensão, anemia e plaquetas impactam significativamente a fração de ejeção do coração. Além disso, foi empregado uma regressão logística para determinar se variáveis como idade, creatinina sérica, fração de ejeção, hipertensão, tabagismo e diabetes possuem influência significativa na mortalidade dos pacientes com insuficiência cardíaca.

2. Metodologia

Realizou-se uma pesquisa epidemiológica de natureza quantitativa (Pereira et al., 2018) e com uso de estatística

descritiva simples com histogramas, valores de média, mediana e análise estatística (Vieira, 2021; Shitsuka et al., 2014).

2.1 Dados e Variáveis

O conjunto de dados foi extraído do UCI Machine Learning Repository (Dua & Graff, 2019) e contém 13 variáveis clínicas relevantes, entre elas, a idade do paciente, presença de anemia, nível da enzima creatinina fosfoquinase no sangue (mcg/L), diagnóstico de diabetes, fração de ejeção (percentual de sangue expelido pelo coração a cada contração (%)), presença de hipertensão arterial, contagem de plaquetas no sangue (kilo-plaquetas/L), concentração de creatinina no soro sanguíneo (mg/dL), nível de sódio no soro sanguíneo (mEq/L), gênero do paciente, hábito de fumar, período de acompanhamento do paciente e desfecho clínico do paciente. Com os dados adquiridos, foi realizada uma análise exploratória (R Core Team, 2021) para entender a distribuição dos dados para cada variável e obter uma noção geral da correlação entre elas para o desenvolvimento da modelagem estatística.

2.2 Modelagem Estatística

Para a análise da insuficiência cardíaca e seus fatores de risco, foram ajustados dois modelos estatísticos: uma regressão linear para avaliar os fatores que influenciam a fração de ejeção e uma regressão logística para estimar a probabilidade de óbito dos pacientes.

2.2.1 Regressão Linear Múltipla

Visto a importância da fração de ejeção na insuficiência cardíaca, foi realizado uma regressão linear utilizando essa variável como variável resposta. O intuito é analisar estatisticamente por meio de testes de hipóteses os fatores que influenciam a fração de ejeção no coração. Para isso, as variáveis consideradas foram idade, creatinina sérica, sódio sérico, hipertensão, anemia e plaquetas.

As hipóteses são as seguintes:

Hipótese nula (H_0): Nenhuma das variáveis explicativas influencia significativamente a fração de ejeção.

Hipótese alternativa (H_A): Pelo menos uma variável explicativa influencia a fração de ejeção.

2.2.2 Regressão Logística

Para responder a seguinte pergunta: ‘Quais fatores aumentam a probabilidade de morte do paciente?’, foi ajustado um modelo de regressão logística. O modelo leva em consideração as seguintes variáveis: idade, creatinina sérica, fração de ejeção, hipertensão, tabagismo e diabetes.

As hipóteses são as seguintes:

Hipótese nula (H_0): Nenhuma das variáveis explicativas influencia a morte do paciente.

Hipótese alternativa (H_A): Pelo menos uma variável explicativa influencia a morte do paciente.

3. Resultados e Discussão

3.1 Análise Exploratória

A distribuição da idade abrange uma faixa de 40 a 95 anos, sendo a maioria dos indivíduos com idade entre 50 e 70 anos. A proporção de pacientes com anemia foi de aproximadamente 43%, o que sugere que essa condição é relevante para o estudo, uma vez que pode influenciar o prognóstico de doenças cardiovasculares. Em relação aos valores de creatina fosfoquinase, foi observado uma média de 582 mcg/L, com um desvio padrão elevado de 970 mcg/L. Notavelmente, a mediana desta variável

foi de 250 mcg/L, indicando uma distribuição não simétrica. Essa discrepância entre média e mediana sugere a presença de outliers no grupo estudado, o que eleva o valor médio.

Assim como a anemia, os casos de diabetes também apresentaram uma prevalência considerável, atingindo aproximadamente 42% dos indivíduos, com um desvio padrão de 0,49%. Outro indicador importante para a função cardíaca é a fração de ejeção, cuja média foi de 38,1%, valor inferior ao intervalo considerado normal (50–70%), de acordo com a American Heart Association (2025). A hipertensão, por sua vez, está presente em 35% dos indivíduos, configurando-se como outro fator de risco importante. A contagem de plaquetas apresentou grande variabilidade, oscilando entre 25.100 e 850.000 Kilo Plaquetas por litro, com uma média de 263.358 Kilo Plaquetas por litro e um desvio padrão de 97.804 Kilo Plaquetas por litro.

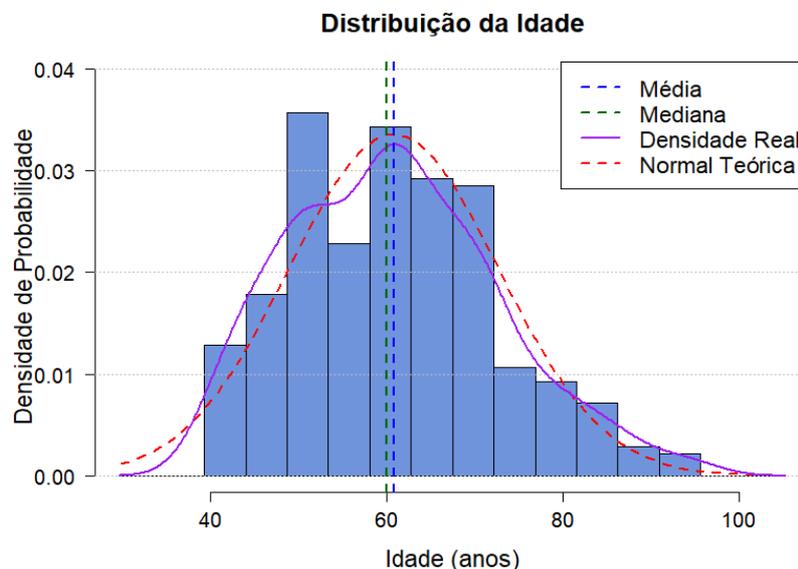
No que se refere aos parâmetros bioquímicos, a creatinina sérica apresentou uma média de 1,39 mg/dL e um desvio padrão de 1,03 mg/dL, enquanto o sódio sérico apresentou média de 137 mEq/L, com desvio padrão de 4,41 mEq/L. Quanto à distribuição de gênero, 65% dos pacientes são do sexo masculino. Além disso, aproximadamente 32% dos indivíduos são fumantes. O período médio de acompanhamento dos pacientes foi de 130 dias. Do total, cerca de 32% dos pacientes faleceram ao longo do estudo.

3.2 Distribuição por meio de Histogramas

Para avaliar a distribuição dos dados, foram construídos histogramas de algumas variáveis, representados em densidade de probabilidade. Essa abordagem permitiu comparações diretas entre variáveis normalizadas, independentemente da escala (James et al., 2021; Cleveland, 2025). Além dos histogramas, foram adicionadas a curva de densidade empírica dos dados e a curva de distribuição normal teórica, estimada com base na média e no desvio padrão da amostra.

A Figura 1 apresenta o histograma da variável idade.

Figura 1 - Histograma da distribuição da Idade.

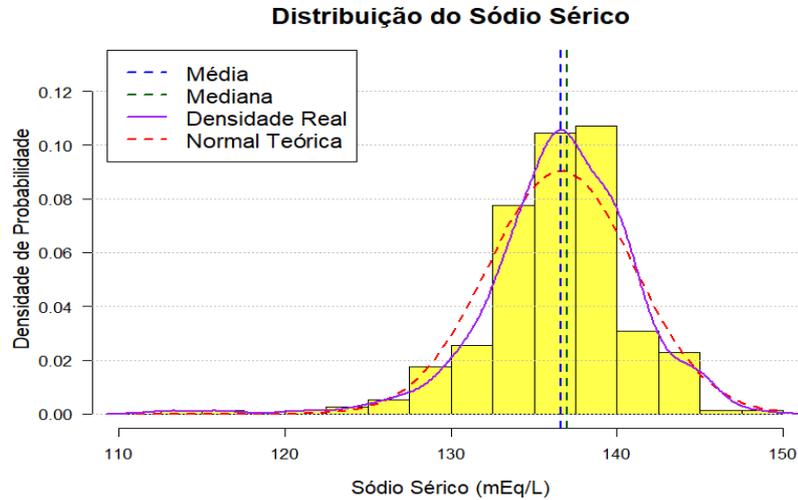


Fonte: Autoria própria.

De acordo com a Figura 1 é possível notar uma maior frequência em faixas etárias mais jovens e uma cauda estendida para idades avançadas. A curva de densidade estimada (linha roxa) ajusta-se bem aos dados, enquanto a distribuição normal teórica (linha vermelha tracejada) serve como referência para a normalidade. A proximidade da média e mediana sugere baixa

distorção na distribuição, embora a assimetria indique um leve deslocamento da média para valores superiores, devido à presença de indivíduos mais velhos.

Figura 2 - Histograma da distribuição da concentração de sódio sérico.

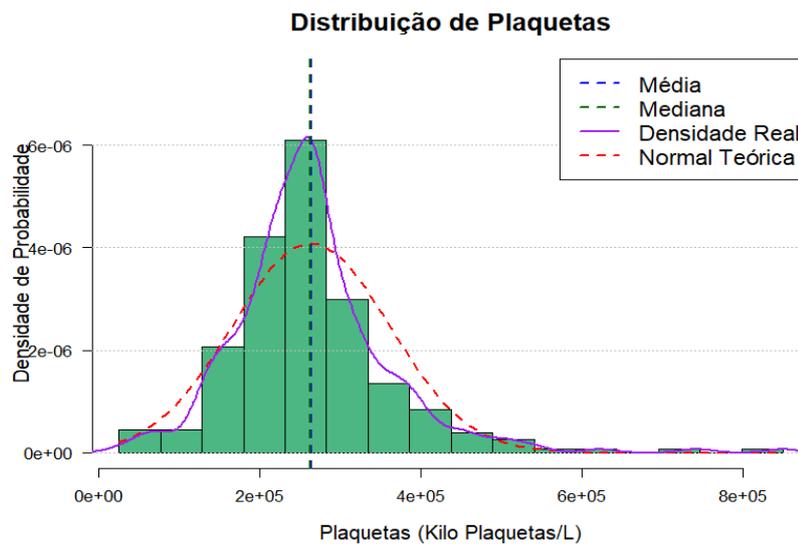


Fonte: Autoria própria.

De modo análogo ao gráfico da Figura 1, a média e a mediana no gráfico da Figura 2 sugerem uma distribuição aproximadamente simétrica. Observa-se uma maior concentração em torno de 135-140 mEq/L. É possível notar que a concentração da maioria dos indivíduos, em geral, se encontra dentro da faixa normal esperada (WHO, 2024).

O histograma da distribuição da contagem de plaquetas está representado na Figura 3.

Figura 3 - Histograma da distribuição da contagem de plaquetas.



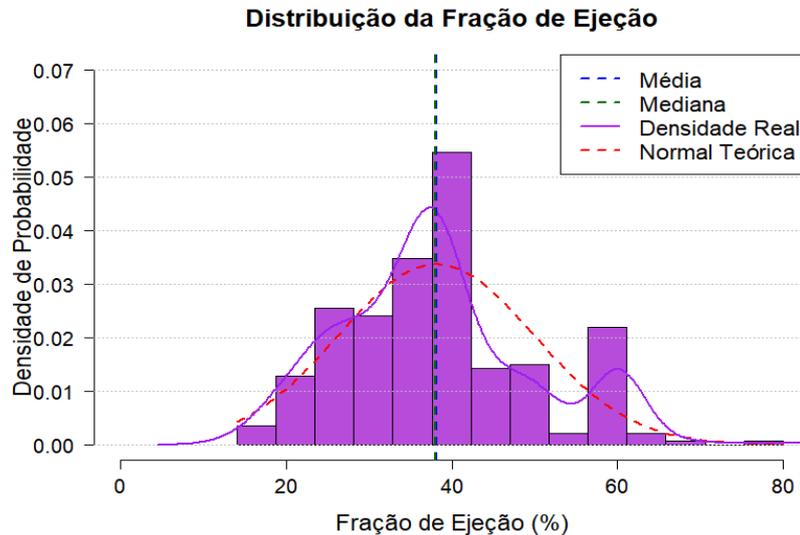
Fonte: Autoria própria.

A distribuição da contagem de plaquetas sugere uma distribuição aproximadamente simétrica. A curva da densidade real apresenta um pico mais elevado em relação à curva normal teórica. Isso sugere uma distribuição com menor dispersão, uma vez que os dados têm uma maior concentração em uma faixa estreita de valores. O intervalo da contagem de plaquetas abrange

uma ampla faixa, com maior concentração ao redor de 250.000 Kilo plaquetas/L. Esse valor está dentro do valor esperado normal de 150.000 - 450.000 Kilo plaquetas/L (WHO, 2024).

O histograma da distribuição da fração de ejeção está representado na Figura 4.

Figura 4 - Histograma da distribuição da fração de ejeção.

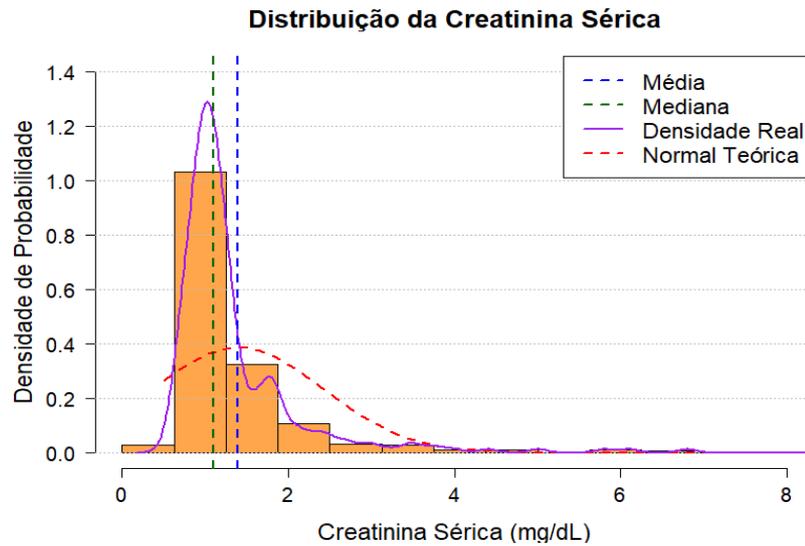


Fonte: Autoria própria.

Já o histograma da Figura 4 mostra que a distribuição da fração de ejeção não segue perfeitamente uma normal, evidenciado pela discrepância entre a curva de densidade real e a normal teórica. A densidade real apresenta assimetria e múltiplos picos, indicando uma heterogeneidade nos dados. A média e a mediana estão próximas, sugerindo uma distribuição relativamente simétrica, mas a normal teórica subestima a cauda esquerda e superestima a centralização dos valores. Isso sugere que os dados contém subgrupos distintos e que a distribuição possui caudas mais pesadas do que uma normal. Um desses subgrupos pode ser evidenciado próximo a 60% de fração de ejeção, o que corresponde a um grupo significativo com um valor elevado. A maior frequência de valores se dá ao redor de 40%.

O histograma da distribuição da creatinina sérica está representado na Figura 5.

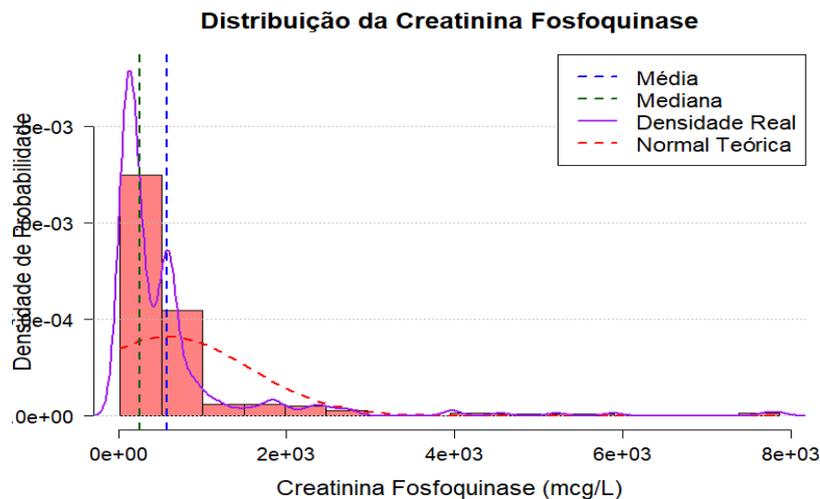
Figura 5 - Histograma da distribuição da creatinina sérica.



De acordo com o gráfico, a distribuição da creatinina sérica apresenta forte assimetria à direita, evidenciada pela diferença entre a densidade real e a normal teórica. Outro ponto que reforça a assimetria é o fato da mediana estar à esquerda da média, sugerindo que alguns valores elevados de creatinina aumentam a média de forma desproporcional. Isso indica a presença de um subgrupo com níveis elevados de creatinina. A densidade real mostra um pico bem definido abaixo de 1 mg/dL, indicando que a maioria dos valores se concentra nessa faixa, enquanto a cauda longa sugere uma menor, mas significativa, ocorrência de níveis elevados.

O histograma da distribuição da creatinina fosfoquinase está representado na Figura 6.

Figura 6 - Histograma da distribuição da creatinina fosfoquinase.



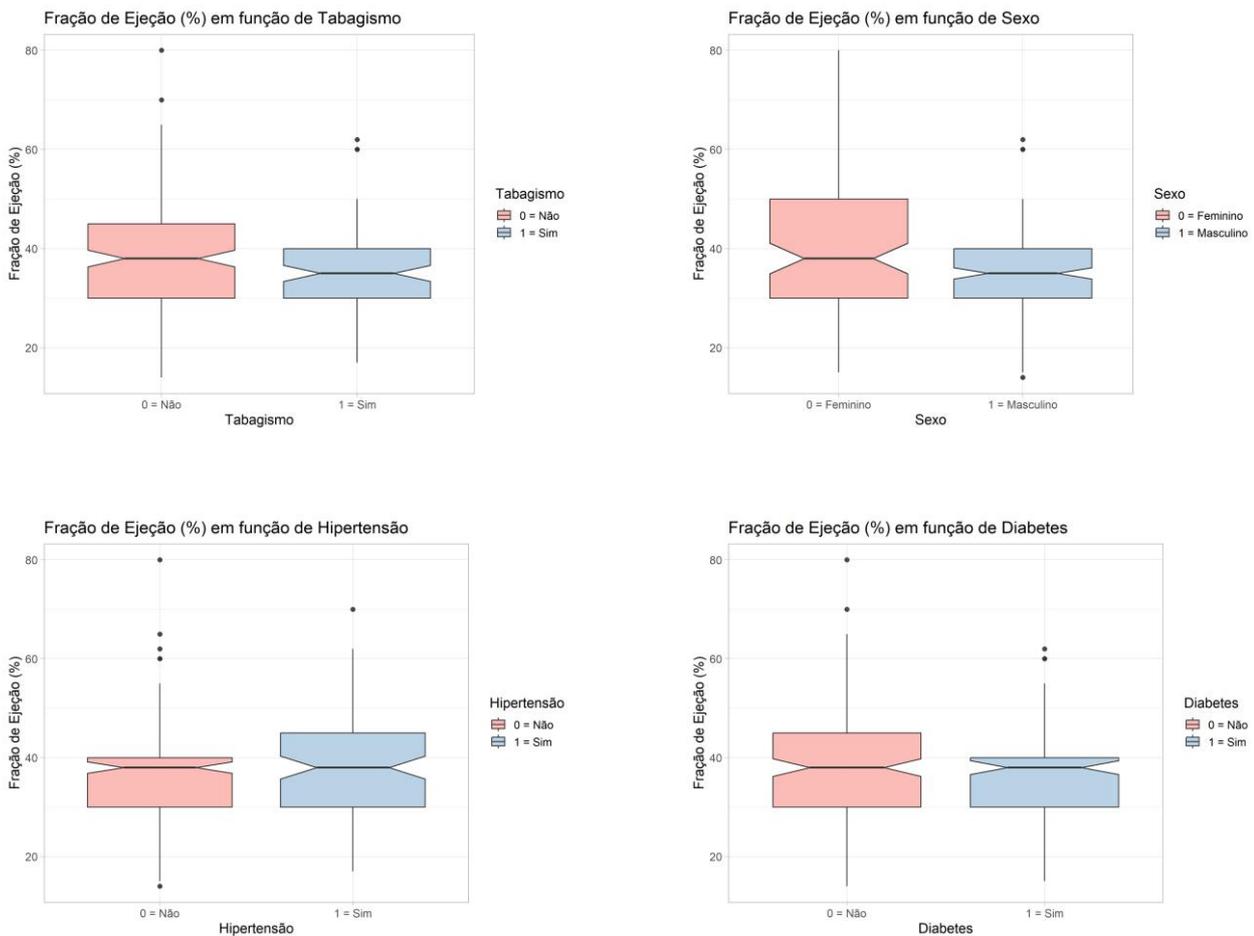
Por fim, a distribuição da creatinina fosfoquinase mostra uma distribuição assimétrica à direita, com a maior concentração de valores em torno de 250 mcg/L. A média está à frente da mediana, indicando que a distribuição não é simétrica e é afetada por valores elevados à direita, que puxam a média para cima. A curva de densidade apresenta um pico pronunciado

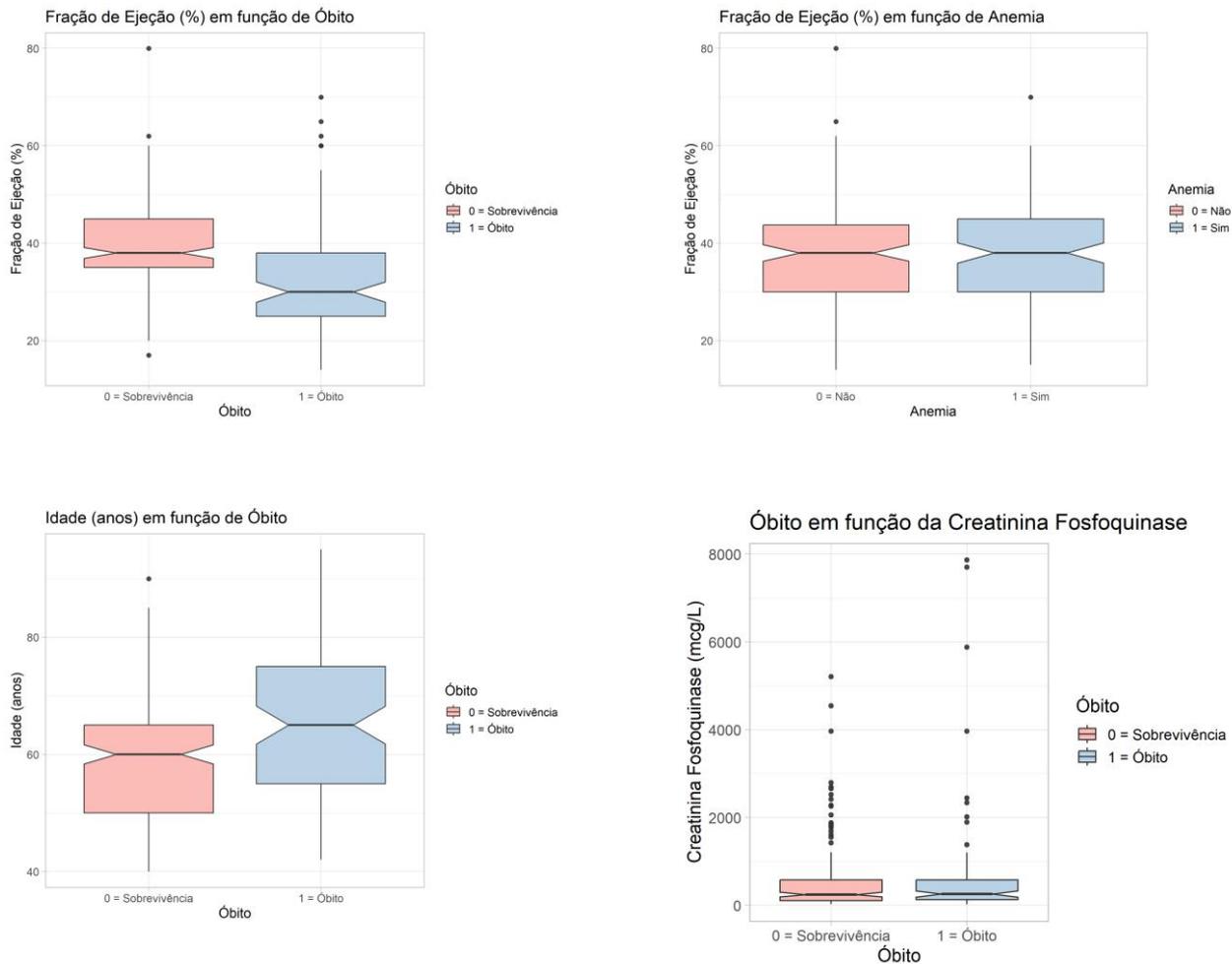
próximo à mediana e outro menor na média, refletindo a concentração central dos dados. Além disso, a cauda direita do histograma se estende até quase 8000 mcg/L, sugerindo a presença de outliers com valores elevados.

3.3 Análise por Meio de Boxplots

A fim de compreender a distribuição dos dados em relação a variável de interesse: fração de ejeção e óbitos, foram construídos boxplots. Os gráficos estão representados na Figura 7.

Figura 7 - Gráfico dos boxplots das variáveis de interesse.





Fonte: Autoria própria.

No primeiro gráfico, foi observado a relação entre a fração de ejeção e o tabagismo. O grupo de não fumantes apresenta uma mediana ligeiramente superior à dos fumantes, indicando que, em média, não fumantes tendem a ter uma melhor fração de ejeção. Além disso, a dispersão dos valores é maior entre os não fumantes, com a presença de alguns outliers acima do intervalo interquartil.

Já no gráfico onde a fração de ejeção é analisada em relação ao sexo, as mulheres apresentam uma mediana superior em comparação aos homens, sugerindo que, em média, a função cardíaca é melhor no grupo feminino. Assim como no primeiro gráfico, há maior variabilidade nos dados femininos, com alguns outliers acima do intervalo interquartil.

Para a Fração de ejeção e a hipertensão, observa-se que indivíduos hipertensos apresentam uma mediana inferior em comparação aos não hipertensos, demonstrando que a hipertensão está associada a uma pior função cardíaca (Gallo & Savoia, 2024; Ziaean & Fonarow, 2016). Além disso, os não hipertensos apresentam uma maior dispersão dos dados, com alguns valores atípicos acima do intervalo interquartil.

Quanto à fração de ejeção em relação ao diabetes, verifica-se uma tendência semelhante à observada no primeiro gráfico. Indivíduos com diabetes apresentam uma mediana inferior à dos não diabéticos, indicando que a presença dessa condição pode estar associada a uma pior fração de ejeção. A dispersão dos valores é relativamente semelhante entre os dois grupos, mas os não diabéticos apresentam alguns outliers acima do intervalo interquartil.

Além disso, observa-se que a fração de ejeção dos indivíduos que faleceram tende a ser menor do que a dos sobreviventes. A mediana da fração de ejeção no grupo de óbito está mais baixa, e a dispersão dos dados também sugere maior variabilidade nesse grupo, com presença de valores atípicos. Isso indica que uma menor fração de ejeção pode estar associada a um maior risco de óbito, como esperado pela literatura.

Já no gráfico que analisa a fração de ejeção em função da presença ou ausência de anemia, observa-se que os dois grupos possuem distribuições semelhantes, sem diferenças visíveis significativas na mediana ou dispersão. Mostrando assim que a presença de anemia pode não influenciar diretamente a fração de ejeção, pelo menos dentro da amostra analisada.

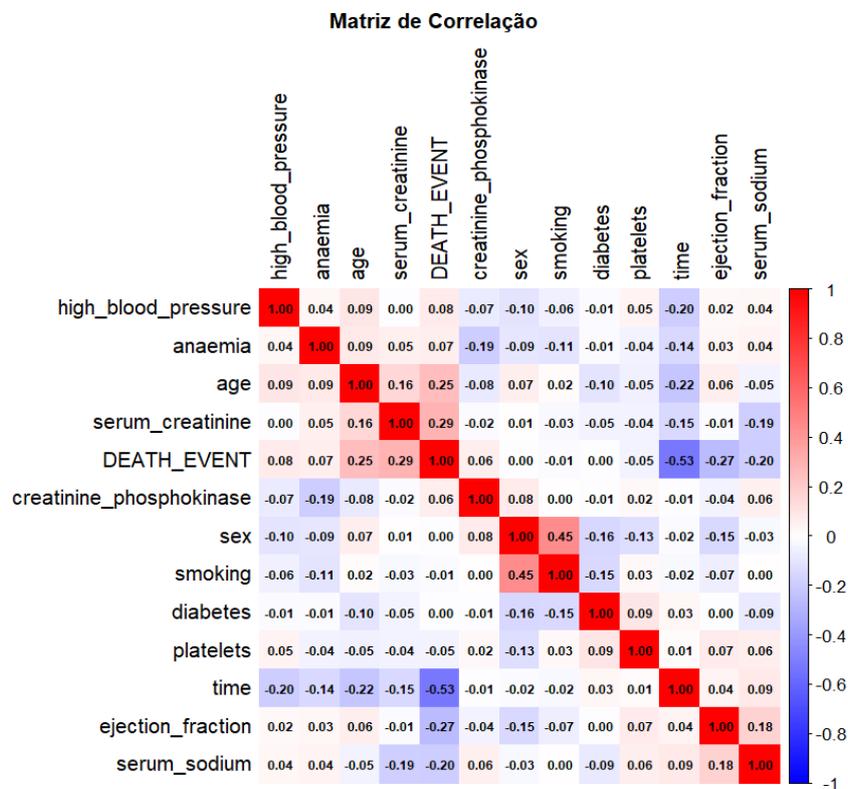
No gráfico que representa a idade em função do óbito, observa-se que os indivíduos que faleceram são mais velhos. Além disso, a distribuição dos valores para o grupo de óbito parece mais ampla, demonstrando que a idade avançada pode estar associada a um maior risco de mortalidade.

Por fim, foram analisados os níveis de creatinina fosfoquinase em função do óbito. Nota-se que a mediana dos valores é semelhante entre os grupos, mas o grupo de óbito apresenta uma dispersão muito maior, com diversos valores atípicos em níveis elevados. Informando que a maioria dos indivíduos possui níveis relativamente baixos de CPK, alguns casos com valores extremamente altos estão mais frequentemente associados ao óbito, possivelmente indicando maior gravidade da condição clínica.

3.4 Matriz de Correlação

Em seguida, foi construído uma matriz de correlação representado na Figura 8.

Figura 8 - Matriz de correlação das variáveis.



Fonte: Autoria própria.

A partir da matriz de correlação, pode se notar que a idade está relacionada positivamente com a creatinina sérica e o óbito dos indivíduos. Além disso, a fração de ejeção mostrou correlação negativa com o evento de morte, indicando que

indivíduos com menor fração de ejeção tiveram maior risco de óbito. Também foi observado que a creatinina sérica apresentou correlação negativa com o sódio sérico e a fração de ejeção, sugerindo que níveis elevados de creatinina sérica estavam associados a menor fração de ejeção e a concentrações mais baixas de sódio sérico.

Assim, por meio dos padrões observados na análise exploratória, foram desenvolvidos os modelos para confirmar se as variáveis estão estatisticamente relacionadas.

3.5 Modelagem Estatística

3.5.1 Regressão Linear Múltipla

A seguir, o Quadro 1 apresenta os valores dos resíduos:

Quadro 1 - Valores dos resíduos.

Min	1Q	Median	3Q	Max
-23.299	-8.092	-0.941	6.154	43.102

Fonte: Autoria própria.

Os valores dos resíduos indicam uma dispersão significativa dos dados em torno da reta de regressão, com alguns resíduos relativamente grandes, tanto negativos quanto positivos, devido à presença de outliers.

Em relação aos coeficientes das variáveis: o valor do intercepto foi de -33.14, o que representa a estimativa da fração de ejeção quando todas as variáveis independentes são zero. No entanto, este valor possui um p-valor de 0.132, o que não é significativo (acima de 0.05), indicando que o intercepto não é estatisticamente diferente de zero. O coeficiente de idade foi de 0.0667, indicando que, para cada aumento de um ano na idade, a fração de ejeção aumenta em média 0.0667 unidades. Contudo, o p-valor de 0.25319 sugere que a idade não tem um impacto estatisticamente significativo sobre a fração de ejeção. Clinicamente, o que se espera é que com o aumento da idade, a fração de ejeção diminua, devido ao enfraquecimento do músculo cardíaco no bombeamento de sangue (Dunlay et al., 2017).

O coeficiente da creatinina sérica, por sua vez, foi igual a 0.1492, mas o p-valor de 0.82571 indica que a creatinina sérica não é uma variável explicativa significativa para a fração de ejeção. O coeficiente do sódio sérico foi de 0.4727, indicando que, à medida que o sódio sérico aumenta, a fração de ejeção também aumenta. O p-valor de 0.00286 é altamente significativo (menor que 0.05), o que sugere que o sódio sérico tem uma relação estatisticamente significativa com a fração de ejeção. O coeficiente da pressão alta foi de 0.1895, mas o p-valor de 0.89462 indica que a pressão alta não tem um efeito significativo na fração de ejeção. O coeficiente da anemia foi de 0.4814, mas o p-valor de 0.72703 indica também que a anemia não tem um impacto significativo na fração de ejeção. Por fim, o coeficiente das plaquetas foi de $7,96,10^{-6}$, mas o p-valor de 0.25483 sugere que as plaquetas também não influenciam significativamente a fração de ejeção.

Quanto às estatísticas globais do modelo, o R-quadrado múltiplo foi igual a 0.04046, o que significa que aproximadamente 4.05% da variação na fração de ejeção pode ser explicada pelas variáveis independentes do modelo. Este valor é muito baixo, indicando que o modelo não explica bem os dados. Já o R-quadrado ajustado (0.02074), sugere que o modelo ainda não é eficaz na explicação da variação da fração de ejeção.

Quanto a estatística F (2.052) com p-valor de 0.0589, está muito perto de 0.05, indicando uma significância marginal. Isso pode indicar que, em um nível de significância de 0.05, o modelo global não é significativamente diferente de zero, ou seja, as variáveis explicativas no modelo não têm um efeito significativo.

Assim sendo, para avaliar as hipóteses, foi examinado os p-valores das variáveis explicativas:

A variável sódio sérico apresenta um p-valor significativo (0.00286), o que indica que ela tem um efeito significativo sobre a fração de ejeção. Enquanto as outras variáveis não apresentaram p-valores significativos. Dessa forma, o p-valor global do modelo (0.0589) é marginalmente significativo, o que sugere que, com base neste modelo, pode haver alguma influência coletiva das variáveis independentes sobre a fração de ejeção, mas a evidência não é forte o suficiente para rejeitar a hipótese nula com um nível de significância de 0.05. Assim, com base nos resultados, podemos não rejeitar a hipótese nula de que nenhuma das variáveis explicativas têm um impacto significativo sobre a fração de ejeção no modelo. A exceção é o sódio sérico, que tem um efeito significativo. No entanto, o baixo valor de R-quadrado sugere que o modelo geral não é capaz de explicar bem a variação na fração de ejeção.

Foi realizada também a análise de variância (ANOVA).

Quadro 2 - Valores da soma dos quadrados (*Sum Sq*), os quadrados médios (*Mean Sq*), o valor de F (*F value*) e o p-valor (*Pr(>F)*) para cada variável.

Variáveis	Sum Sq	Mean Sq	F value	Pr (>F)
Idade	151	150,75	1,0991	0,2953
Creatinina Sérica	19	18,65	0,1360	0,7126
Sódio Sérico	1323	1322,57	9,6426	0,0021
Hipertensão	6	5,66	0,0412	0,8392
Anemia	12	12,41	0,0905	0,7638
Plaquetas	179	178,54	1,3017	0,2548
Resíduos	40050	137,16	-	-

Fonte: Autoria própria.

Os valores resumem a contribuição de cada variável explicativa na variação da fração de ejeção. Vamos analisar as variáveis significativas e não significativas e os resíduos.

3.5.2 Variáveis Significativas

Sódio Sérico: F value = 9.6426 e Pr(>F) = 0.002088, o que indica que a variável é estatisticamente significativa na explicação da fração de ejeção, com um p-valor inferior a 0,05. Isso demonstra que, para este modelo, a concentração de sódio no sangue tem um impacto relevante na fração de ejeção do coração.

3.5.3 Variáveis Não Significativas

A seguir está representado o quadro dos valores do p-valor das variáveis.

Quadro 3 - Valores do p-valor das variáveis.

Variável	Idade	Creatinina sérica	Hipertensão	Anemia	Plaquetas
p-valor	0.295327	0.712575	0.839236	0.763760	0.254833

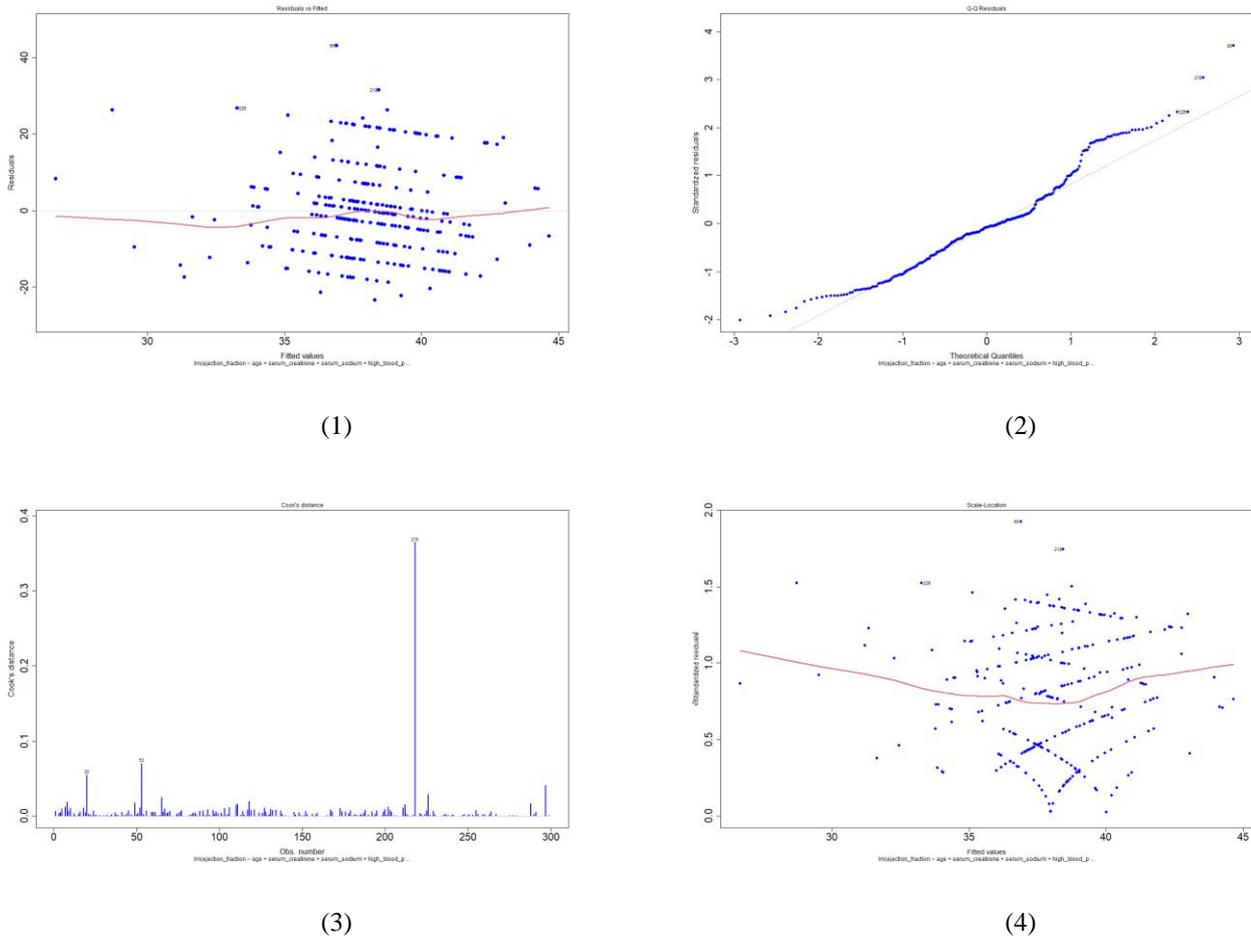
Fonte: Autoria própria.

Todas essas variáveis possuem um p-valor acima de 0,05 indicando que não são estatisticamente significativas na explicação da fração de ejeção.

3.5.4 Resíduos

As linhas Residuais mostram que a soma dos quadrados dos resíduos foi de 40050, revelando que a variabilidade não é explicada pelas variáveis independentes no modelo. O quadrado médio dos resíduos (137.16) indica que a variação média não é explicada pelo modelo. Os gráficos de diagnóstico do modelo estão representados na Figura 9.

Figura 9 - Gráficos de diagnóstico do modelo linear.



Fonte: Autoria própria.

O gráfico de Resíduos vs Ajustados (1) revelou que os resíduos não estão mostrando uma tendência óbvia de aumento ou diminuição da variância conforme os valores ajustados mudam. Isso indica que a variância dos resíduos parece ser constante em todos os níveis de valores ajustados. Outra questão é a quantidade de resíduos mais distantes, indicando um grande número de outliers. O gráfico da Normal Q-Q (2) sugere que a distribuição dos resíduos possui maiores valores extremos do que uma distribuição normal, isto é, há uma maior frequência de valores residuais muito baixos do que seria esperado em uma distribuição normal. Além disso, a distribuição tem uma menor ocorrência de valores muito altos, sugerindo que a distribuição dos resíduos pode ser mais concentrada em torno de zero do que a normal. Esses desvios podem ser devido a quantidade de outliers, o que afetou a qualidade do modelo. Já no gráfico da distância de Cook (3) foi possível notar três pontos pronunciados que possivelmente são outliers que estão distorcendo o resultado dos modelos. Por fim, no gráfico de leverage (4), a dispersão dos

dados mostrou que a maioria das observações do conjunto de dados não estão exercendo uma grande influência sobre o modelo de regressão

Assim, a partir dos resultados do modelo linear, foi concluído que, em um nível de significância de 5%, apenas o sódio sérico é uma variável significativa para a fração de ejeção do coração. As outras variáveis não têm impacto significativo.

A Hipótese nula (H_0), que afirmava que nenhuma das variáveis explicativas teria um impacto significativo sobre a fração de ejeção, pode ser rejeitada, uma vez que o sódio sérico mostrou um efeito significativo. Isso sugere que pelo menos uma variável explicativa tem influência na fração de ejeção, como previsto pela hipótese alternativa (H_A).

3.6 Regressão Logística

Por meio de seis variáveis preditoras: idade, creatinina sérica, fração de ejeção, hipertensão, tabagismo e diabetes, o modelo de regressão logística foi ajustado para prever o óbito dos indivíduos. Dessa forma, a análise identificou três preditores estatisticamente significativos. A variável idade ($p < 0,001$) mostrou uma associação positiva, com cada aumento de um ano associado a um aumento de 0,0526 nos log-odds do evento de óbito, indicando que indivíduos mais velhos têm maior risco de morte. A creatinina sérica ($p < 0,001$) também demonstrou uma forte associação positiva, com cada aumento unitário relacionado a um aumento de 0,6876 nos log-odds do evento de óbito, sugerindo que níveis elevados de creatinina sérica são um fator de risco significativo para mortalidade. Por outro lado, a fração de ejeção ($p < 0,001$) exibiu uma associação negativa, com cada aumento unitário associado a uma redução de 0,0713 nos log-odds do evento de óbito, destacando que uma melhor função cardíaca é protetora contra a morte (Redfield, 2016). Os demais preditores hipertensão ($p = 0,145$), tabagismo ($p = 0,814$) e diabetes ($p = 0,493$) não foram estatisticamente significativos, indicando que não há evidências estatisticamente significativas de sua associação com o evento de óbito neste modelo. O ajuste do modelo foi adequado, conforme evidenciado pela redução significativa da deviance de 375,35 (nulo) para 302,51 (residual) e pelo Critério de Informação de Akaike (AIC) de 316,51.

Assim, o modelo destaca a idade, creatinina sérica e fração de ejeção como preditores clinicamente relevantes para mortalidade. Para complementar foi realizado um teste de razão de verossimilhança.

O teste de razão de verossimilhança foi conduzido para comparar o modelo de regressão logística ajustado (modelo 1) com um modelo nulo (modelo 2) contendo apenas o intercepto. O modelo 1 inclui os seis preditores já citados, enquanto o modelo 2 é um modelo reduzido sem preditores. Os resultados do teste mostram uma diferença significativa entre os dois modelos, com uma estatística qui-quadrado de 72,843 e um p-valor de $1,066e-13$ ($p < 0,001$). Isso indica que a inclusão dos preditores no modelo 1 melhora significativamente o ajuste do modelo em comparação com o modelo nulo. O alto valor do qui-quadrado e o p-valor baixo fornecem evidências de que pelo menos um dos preditores tem uma relação estatisticamente significativa com o evento de óbito.

Assim, a hipótese nula (H_0) de que nenhuma das variáveis explicativas influencia o desfecho de morte é rejeitada e foi concluído que pelo menos uma das variáveis preditoras tem uma relação estatisticamente significativa com o evento de morte, como a idade, os níveis de creatinina sérica e a fração de ejeção.

4. Conclusão

Dessa forma, por meio da modelagem estatística foi possível concluir que o sódio sérico é uma variável estatisticamente significativa para a fração de ejeção do coração com um p-valor de 0.002. Isso vai ao encontro com a literatura, uma vez que níveis alterados de sódio sérico são observados em pacientes com insuficiência cardíaca, pois reflete a ativação neuro-hormonal e retenção hídrica, apesar de outros fatores subjacentes influenciarem esse quadro (McMurray et al., 2022). Além disso, foi constatado por meio de uma regressão logística que a mortalidade está relacionada estatisticamente a idade, creatinina sérica e fração de ejeção.

Referências

- American Heart Association. (2025). Cardiovascular health and lifestyle changes. American Heart Association. <https://www.heart.org/>
- Avezum, Á., Maia, L. N., & Nakazone, M. (2012). Cenário das doenças cardiovasculares no mundo moderno. In *Manual de cardiologia* (pp. 1–5).
- Cleveland, W. S. (2025). *The elements of graphing data*. Hobart Press.
- Dua, D., & Graff, C. (2019). UCI Machine Learning Repository. University of California, Irvine, School of Information and Computer Sciences. <https://archive.ics.uci.edu/ml>
- Dunlay, S. M., Roger, V. L., & Redfield, M. M. (2017). Epidemiology of heart failure with preserved ejection fraction. *Nature Reviews Cardiology*, 14(10), 591-602. <https://doi.org/10.1038/nrcardio.2017.65>
- Gallo, G., & Savoia, C. (2024). Hypertension and heart failure: From pathophysiology to treatment. *International Journal of Molecular Sciences*, 25(12), 6661. <https://doi.org/10.3390/ijms25126661>
- Groenewegen, A., Rutten, F. H., Mosterd, A., & Hoes, A. W. (2020). Epidemiology of heart failure. *European Journal of Heart Failure*, 22(8), 1342-1356. <https://doi.org/10.1002/ejhf.1858>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: with applications in R*. Springer. <https://doi.org/10.1007/978-1-0716-1418-1>
- Knuuti, J., Wijns, W., Saraste, A., Capodanno, D., Barbato, E., Funck-Brentano, C., ... & Bax, J. J. (2020). 2020 ESC guidelines for the diagnosis and management of chronic coronary syndromes. *European Heart Journal*, 41(3), 407-477. <https://doi.org/10.1093/eurheartj/ehz425>
- McDonagh, T. A., Metra, M., Adamo, M., Gardner, R. S., Baumbach, A., Böhm, M., ... & Seferović, P. M. (2021). 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. *European Heart Journal*, 42(36), 3599-3726. <https://doi.org/10.1093/eurheartj/ehab368>
- McMurray, J. J. V., Solomon, S. D., Inzucchi, S. E., Køber, L., Kosiborod, M. N., Martinez, F. A., ... & De Boer, R. A. (2022). Dapagliflozin in patients with heart failure and reduced ejection fraction. *New England Journal of Medicine*, 381(21), 1995-2008. <https://doi.org/10.1056/NEJMoa1911303>
- Mebazaa, A., Yilmaz, M. B., Levy, P., Ponikowski, P., Peacock, W. F., Laribi, S., ... & Harjola, V. P. (2021). Recommendations on pre-hospital & early hospital management of acute heart failure. *European Journal of Heart Failure*, 23(7), 1005-1022. <https://doi.org/10.1002/ejhf.2238>
- Pereira A. S. et al. (2018). *Metodologia da pesquisa científica*. [free e-book]. Santa Maria/RS. Ed. UAB/NTE/UFSM.
- Précoma DB, Oliveira GMM, Simão AF, Dutra OP, Coelho OR, Izar MCO, et al. Updated cardiovascular prevention guideline of the Brazilian Society of Cardiology - 2019. *Arq Bras Cardiol* 2019; 113(4): 787-891. <https://doi.org/10.5935/abc.20190204>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Redfield, M. M. (2016). Heart failure with preserved ejection fraction. *The New England Journal of Medicine*, 375(19), 1868–1877. <https://doi.org/10.1056/NEJMcp1511175>
- Shitsuka, R. et al. (2014). *Matemática fundamental para tecnologia*. 2ed. Editora Erica.
- Tsao, C. W., Aday, A. W., Almarazooq, Z. I., Anderson, C. A. M., Arora, P., Avery, C. L., Baker-Smith, C. M., Beaton, A. Z., Boehme, A. K., Buxton, A. E., Commodore-Mensah, Y., Elkind, M. S. V., Evenson, K. R., Eze-Nliam, C., Fugar, S., Generoso, G., Heard, D. G., Hiremath, S., Ho, J. E., ... Martin, S. S. (2023). Heart disease and stroke statistics—2023 update: A report from the American Heart Association. *Circulation*, 147(8). <https://doi.org/10.1161/CIR.0000000000001123>
- Vieira, S. (2021). *Introdução à bioestatística*. Ed. GEN/Guanabara Koogan.
- World Health Organization. (2024, Jan 19). Cardiovascular diseases. <https://www.who.int/health-topics/cardiovascular-diseases>
- Ziaeian, B., & Fonarow, G. (2016). Epidemiology and aetiology of heart failure. *Nature Reviews Cardiology*, 13(6), 368–378. <https://doi.org/10.1038/nrcardio.2016.25>