

Explorando a relação entre dores no peito e doenças cardíacas - uma análise de Regressão Logística

Nicole dos Santos de Souza¹

¹Escola de Matemática Aplicada - FGV EMap, Rio de Janeiro, RJ

Junho de 2023

RESUMO

Este estudo investiga a relação entre a presença de dores no peito e doenças cardíacas, usando um modelo de regressão logística por meio de uma abordagem frequentista. Através da análise de 9 atributos, incluindo idade, sexo, angina induzida por exercício e outros, buscamos determinar se as dores no peito são um indicativo relevante para o diagnóstico dessas condições. Os dados utilizados são provenientes de um conjunto disponível no repositório da UC Irvine Machine Learning. O objetivo é melhorar os métodos de diagnóstico e tratamento das doenças cardíacas, visando resultados clínicos mais efetivos para os pacientes.

PALAVRAS-CHAVE: Logística; Dores no peito; Doenças cardíacas; Análise; Diagnóstico.

1 Introdução

As doenças cardiovasculares são a principal causa de morte não só no Brasil, mas em todo o mundo¹, sendo de extrema importância compreender os fatores associados a essas condições para um diagnóstico precoce e um tratamento efetivo. Dentre os sintomas frequentemente relatados pelos pacientes, as dores no peito têm despertado particular interesse devido à sua potencial relação com doenças cardíacas. Esse fato pode ser melhor entendido no artigo [1] que explora a importância das dores no peito como um sintoma relevante na avaliação e manejo de pacientes com suspeita de síndrome coronariana aguda.

Será melhor explorado aqui a angina: geralmente descrita como uma sensação de aperto, pressão, queimação ou desconforto no peito. A angina ocorre quando há uma redução temporária no suprimento de sangue rico em oxigênio para o músculo cardíaco devido à obstrução parcial das artérias coronárias. Geralmente, esse tipo de sintoma é, por senso comum, um alarmante para o estado da saúde do coração - mas será que isso se justifica estatisticamente?

O presente estudo busca explorar a relação entre alguns tipos de angina e a presença de doenças cardíacas, utilizando uma abordagem frequentista com um modelo de regressão logística. Através dessa análise, pretendemos investigar se a presença de dores no peito pode ser considerada um atributo relevante para o diagnóstico dessas condições, bem como avaliar o impacto dessa variável na acurácia do modelo preditivo.

Ao compreender melhor essa relação, poderemos aprimorar os métodos de diagnóstico e tratamento, proporcionando melhores resultados clínicos para os pacientes com doenças cardíacas.

¹<https://www.gov.br/pt-br/noticias/saude-e-vigilancia-sanitaria/2022/09/doencas-cardiovasculares-principal-caoa-de-morte-no-mundo\protect\@normalcr\relax-pode-ser-prevenida>

1.1 Dados

Inicialmente, planejava-se extrair os dados de um conjunto disponível no Kaggle². No entanto, após realizar algumas análises, suspeitou-se que os dados pudessem conter erros. As suposições iniciais, como a hipótese "dor no peito induzida por exercício físico está mais presente em pacientes diagnosticados com doença cardíaca", revelaram-se inconsistentes, e os resultados obtidos para cada variável analisada eram contrários às expectativas intuitivas.

Diante disso, foi realizado um levantamento de informações na própria plataforma do Kaggle, por meio de discussões relacionadas ao conjunto de dados em questão, e foi encontrado o artigo intitulado "Investigating Heart Disease Datasets and Building Predictive Models" [2]. Nele, há uma análise detalhada dos mesmos dados provenientes de fontes diferentes: o conjunto de dados inicialmente considerado problemático e também um conjunto de dados proveniente do repositório da UC Irvine Machine Learning (UCI)³. A conclusão obtida nesse estudo é que a classificação binária dos dados do conjunto do Kaggle estava realmente invertida, confirmando a suspeita anterior.

Portanto, para o desenvolvimento do modelo, os dados utilizados serão os provenientes do repositório da UCI. Essa decisão baseia-se na validação e confirmação obtidas por meio da análise comparativa dos dados entre as duas fontes mencionadas.

O dataset em questão contém 4 bases provenientes de diferentes países: Cleveland, Hungary, Switzerland, and the VA Long Beach. Posteriormente será justificada a escolha de apenas 2 dessas bases para a construção do modelo.

A base contém 76 atributos, mas analisaremos apenas 14 deles, que dispõem abaixo, na tabela:

Atributo	Descrição	Tipo de Variável	Domínio
age	Idade da pessoa	Numérica	em anos
sex	Sexo da pessoa	Catégorica	0 = mulher 1 = homem
cp	Tipo de dor no peito	Catégorica	1 = angina típica 2 = angina atípica 3 = dor não anginosa 4 = assintomático
trestbps	Pressão arterial em repouso	Numérica	mm Hg
chol	Colesterol medido via sensor BMI	Numérica	mg/dl
fbs	Açúcar no sangue em jejum	Catégorica	1 = verdadeiro 0 = falso
restecg	Resultados do eletrocardiograma em repouso	Catégorica	0 = normal 1 = anormalidade na onda ST-T 2 = provável ou definitiva hipertrofia ventricular esquerda
thalach	Frequência cardíaca máxima atingida	Numérica	-
exang	Angina induzida pelo exercício	Catégorica	1 = sim, 0 = não
oldpeak	Depressão de ST induzida pelo exercício em relação ao repouso	Numérica	-
slope	Inclinação do segmento ST no pico do exercício	Numérica	-
ca	Número de vasos principais coloridos por fluoroscopia	Numérica	-
thal	Resultado da cintilografia com Talio	Catégorica	3 = normal 6 = defeito fixo 7 = defeito reversível
num	Diagnóstico de doença cardíaca	Catégorica	0: estreitamento < 50% 1, 2, 3, 4: estreitamento > 50% (em qualquer vaso principal)

Tabela 1: Informações do conjunto de dados

²<https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset?protect\@normalc\relaxdatasetId=216167&searchQuery=sex>

³<https://archive.ics.uci.edu/dataset/45/heart+disease>

2 Metodologia

Nesta seção, descreveremos a metodologia utilizada para explorar a relação entre dores no peito e doenças cardíacas por meio de uma análise de regressão logística utilizando a biblioteca StatsModels em Python.

2.1 Análise Estatística

Para explorar a relação entre dores no peito e doenças cardíacas, optamos por realizar uma regressão logística. A regressão logística é uma técnica estatística adequada para modelar variáveis dependentes binárias, como é o caso da presença ou ausência de doença cardíaca.

Os modelos serão ajustados seguindo a abordagem frequentista, através da aplicação do método de máxima verossimilhança. A escolha do enfoque frequentista é amplamente adotada na análise de regressão logística e oferece resultados inferenciais sólidos, permitindo a interpretação dos coeficientes estimados com base na probabilidade associada aos eventos em estudo.

Na análise de regressão logística, utilizamos a distribuição binomial para modelar a variável dependente, que possui duas categorias discretas. A distribuição binomial é apropriada para modelar dados binários, pois descreve a probabilidade de sucesso (presença de doença cardíaca) em um determinado número de ensaios (indivíduos da amostra).

2.2 Atributos escolhidos

Abaixo constam todas as variáveis que foram úteis para a análise:

Variável Dependente

Presença de doença cardíaca (binária: 0 para ausência e 1 para presença).

Variáveis Independentes

As variáveis com maior interpretabilidade no modelo proposto (as mais relevantes) são:

Tipo de dor no peito (categórica: 1 para angina típica, 2 para angina atípica, 3 para dor não-anginosa e 4 para assintomático).

Angina induzida por exercício (binária : 0 para ausência e 1 para presença)

Entretanto, outras variáveis clínicas foram incluídas como preditores adicionais, pois se mostraram relevantes para a previsão:

Depressão de ST induzida pelo exercício em relação ao repouso (numérica)

Sexo (binária: 0 para mulher e 1 para homem)

Idade (numérica)

Frequência cardíaca máxima atingida (numérica)

Pressão arterial em repouso (numérica)

Colesterol medido via sensor BMI (numérica)

Açúcar no sangue em jejum (numérica)

2.2.1 Tratando variável categórica

Nota-se que um dos atributos principais, correspondete ao tipo de dor no peito do indivíduo, possui 4 classes. Por isso, foi utilizado a codificação *dummy*, que é uma técnica comum para lidar com variáveis categóricas em regressão logística. Ela envolve a criação de variáveis dummy (binárias) para cada categoria da variável categórica. Cada variável dummy representa uma categoria específica, com um valor de 0 ou 1, indicando a ausência ou presença da categoria em um determinado caso.

Além disso, a saída (num) foi transformada, onde os níveis de doença cardíaca 0,1,2,3 e 4 foram convertidos em uma variável binária: ausência ou presença.

2.3 Implementação em Python

A análise de regressão logística foi implementada em Python utilizando a biblioteca StatsModels. Essa biblioteca oferece uma variedade de funções e modelos estatísticos, assim como métodos avaliativos. Abaixo consta um trecho exemplo de um dos modelos implementados:

```
import statsmodels.api as sm
import statsmodels.formula.api as smf

# criando um modelo de regressão logística
model1 = smf.glm(formula='y ~ x1 + x2 + x4 + x5 + x6 + x7 + x8 +
x9 + x3_1 + x3_2 + x3_3 + x3_4', data=df_train, family=sm.families.Binomial()).fit()

# usa o modelo para fazer previsões
y_proba = model1.predict(df_test)

# transforma as previsões em 0 ou 1
y_pred = y_proba.apply(lambda x: 1 if x > 0.5 else 0)
```

A variável x_3 corresponde ao tipo de dor no peito (CP - Chest Pain) e foi codificada para ser implementada no modelo, por isso está dividida em 4 subgrupos.

2.4 Avaliação

Para avaliar o desempenho dos modelos propostos e compará-los, utilizou-se das seguintes métricas:

2.4.1 Acurácia (Accuracy)

A acurácia é uma métrica simples e intuitiva que mede a taxa de acertos do modelo em relação ao número total de exemplos.

2.4.2 Precisão (Precision)

A precisão é a proporção de exemplos positivos classificados corretamente em relação ao total de exemplos classificados como positivos.

2.4.3 Recall (Recall)

O recall é a proporção de exemplos positivos classificados corretamente em relação ao total de exemplos que realmente são positivos. No contexto do diagnóstico de doenças, o recall é geralmente considerado mais importante do que a precisão. Isso ocorre porque, nesse cenário, é fundamental identificar corretamente todos os casos positivos (pacientes doentes) para evitar falsos negativos, ou seja, garantir que o mínimo possível de pacientes com doenças seja classificado erroneamente como saudáveis. É preferível ter alguns falsos positivos (que classifica como erro do tipo 1) a ter falsos negativos (que classifica como erro do tipo 2) nesse contexto, pois o tratamento adequado e o cuidado precoce são cruciais para a saúde dos pacientes.

2.4.4 AUC (Area Under the Curve)

A AUC (Área Sob a Curva) é uma métrica amplamente utilizada na análise de curvas ROC (Receiver Operating Characteristic) e é uma medida da qualidade do modelo em separar exemplos positivos de negativos em diferentes pontos de corte. A curva ROC é uma representação gráfica da taxa de verdadeiros positivos (recall) em relação à taxa de falsos positivos (especificidade) para diferentes valores de threshold de classificação. Uma AUC-ROC mais próxima de 1 indica que o modelo tem uma boa capacidade de separar as classes, enquanto

um valor próximo de 0.5 indica que o modelo tem um desempenho semelhante ao de um classificador aleatório.

Quando se trata do diagnóstico de doenças, a curva ROC e a AUC são especialmente relevantes. Uma curva ROC bem ajustada e uma AUC alta indicam que o modelo tem uma capacidade robusta de distinguir entre indivíduos doentes e saudáveis, minimizando tanto os falsos positivos quanto os falsos negativos. No entanto, no cenário atual em que os custos associados aos falsos negativos são muito altos, é desejável ter um alto recall para minimizar a ocorrência de falsos negativos. Nesse caso, a curva ROC pode ser usada para selecionar o ponto de corte que otimize o trade-off entre recall e taxa de falsos positivos. Uma AUC-ROC mais próxima de 1 indica que o modelo tem uma boa capacidade de separar as classes, enquanto um valor próximo de 0.5 indica que o modelo tem um desempenho semelhante ao de um classificador aleatório.

2.4.5 AIC (Akaike Information Criterion)

O AIC é uma métrica que avalia a qualidade do modelo em relação à sua complexidade. Quanto menor o valor do AIC, melhor o modelo está ajustado aos dados. Essa métrica é útil quando se está interessado em selecionar o modelo mais parcimonioso, que fornece um bom ajuste sem incluir variáveis desnecessárias.

2.5 Pipeline

O procedimento de construção do modelo consistiu nos seguintes passos:

a) Pré-processamento dos dados: Isso incluiu a limpeza e estruturação do conjunto de dados, verificação de valores ausentes, codificação adequada das variáveis categóricas e o particionamento dos dados em conjunto de treinamento e teste.

b) Análise Exploratória dos Dados (AED): Com objetivo de compreender melhor as características da base e identificar possíveis padrões ou relações entre as variáveis, foi utilizado ferramentas de visualização, como gráficos e histogramas, para fazer algumas escolhas. Essa etapa permitiu formular hipóteses e fazer escolhas sobre quais atributos incluir no modelo de regressão logística.

c) Ajuste do modelo de regressão logística: Utilizamos a função apropriada da biblioteca StatsModels para ajustar o modelo de regressão logística à nossa amostra de dados. Isso envolveu a definição da variável dependente e das variáveis independentes, bem como a escolha da distribuição binomial.

d) Avaliação do modelo: Avaliamos a qualidade do modelo ajustado por meio de medidas estatísticas, como acurácia, precisão e recall, além da área sob a curva ROC e a métrica AIC. Essas medidas nos permitem avaliar a adequação do modelo aos dados e a importância das variáveis independentes na explicação da variável dependente.

3 Resultados

3.1 Higienização dos Dados

O dataset escolhido possui dados provenientes de 4 fontes e estavam separados em arquivos distintos, por isso, o primeiro passo foi juntar toda a informação em um único Dataframe que pudesse ser trabalhado. Foi preciso realizar uma conversão nos tipos dos valores, além de ajustar as colunas corretamente pois cada arquivo tinha um separador diferente. Com relação aos dados ausentes, foi feita uma análise mais detalhada pois eram numerosos. A conclusão foi que as fontes *switzerland* e *long-beach-va* haviam dados fundamentais faltando e, por isso, foram retiradas do Dataframe principal.

O passo seguinte foi verificar se os dados de cada fonte (no caso, sobraram *cleveland* e *hungarian*) tinham uma distribuição semelhante, para que a análise pudesse ser feita incluindo todos eles e não fazendo a distinção. Após observar o gráfico das distribuições para cada uma das variáveis, algumas foram excluídas por não terem distribuições muito semelhantes para as duas fontes julgando que isso pudesse comprometer o modelo.

Abaixo, na esquerda, está disposto o gráfico das distribuições antes da seleção das variáveis e, na direita, depois.

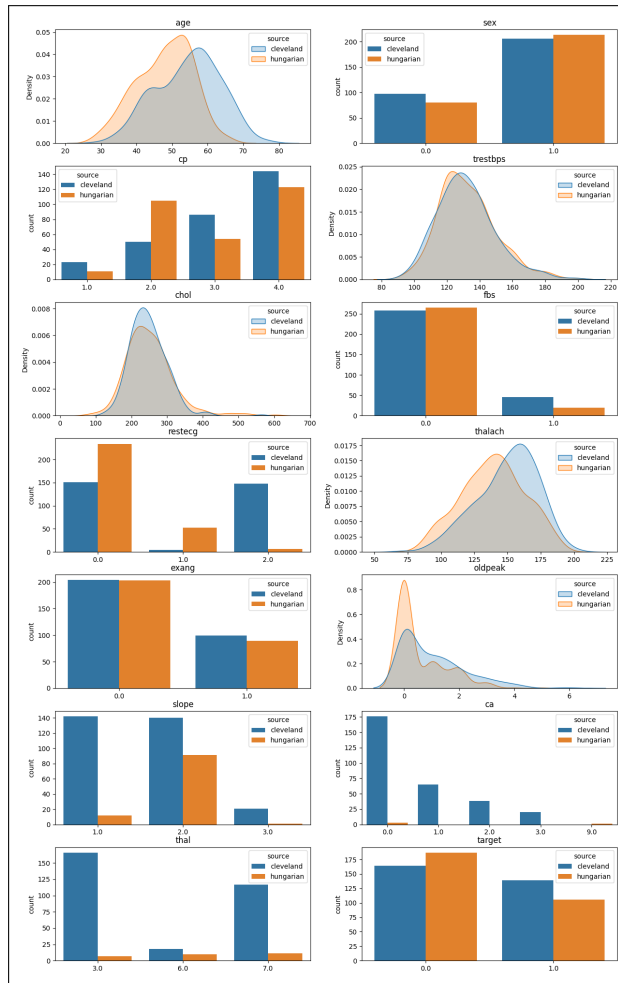


Figura 1: Distribuição dos dados para todas as variáveis

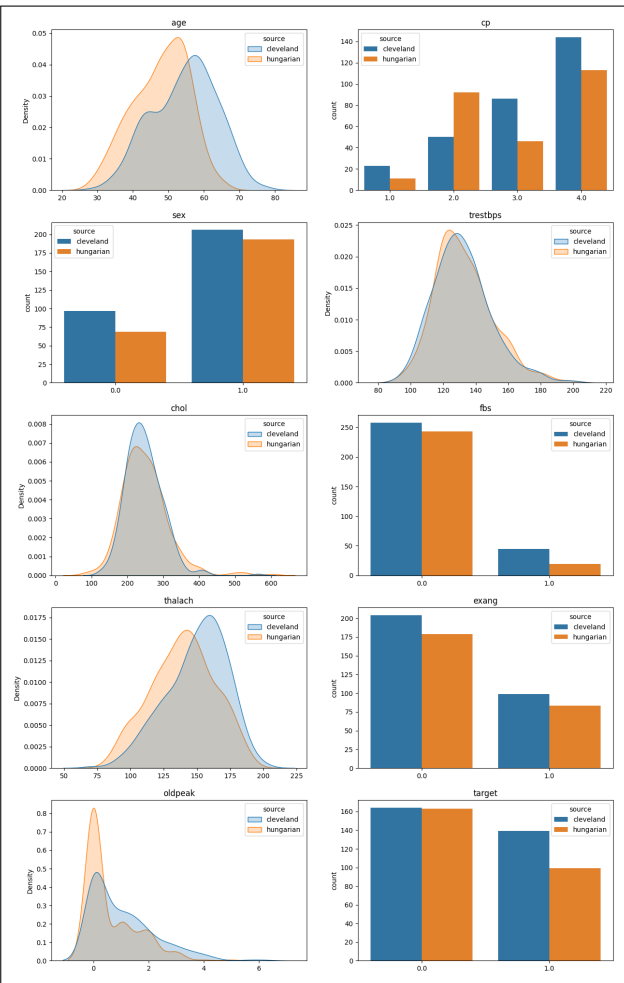


Figura 2: Distribuição dos dados para as variáveis selecionadas

Assim, foi obtido um Dataframe limpo e coerente para realizar as análises.

3.2 Análise Exploratória dos Dados - AED

Inicialmente, foi investigada a correlação entre as variáveis (entre elas e com a target). Para isso, alguns gráficos foram feitos, como o *pairplot* (para analisar como cada par de variáveis se relaciona) e um *heatmap* com as correlações, que ficou de melhor visualização:

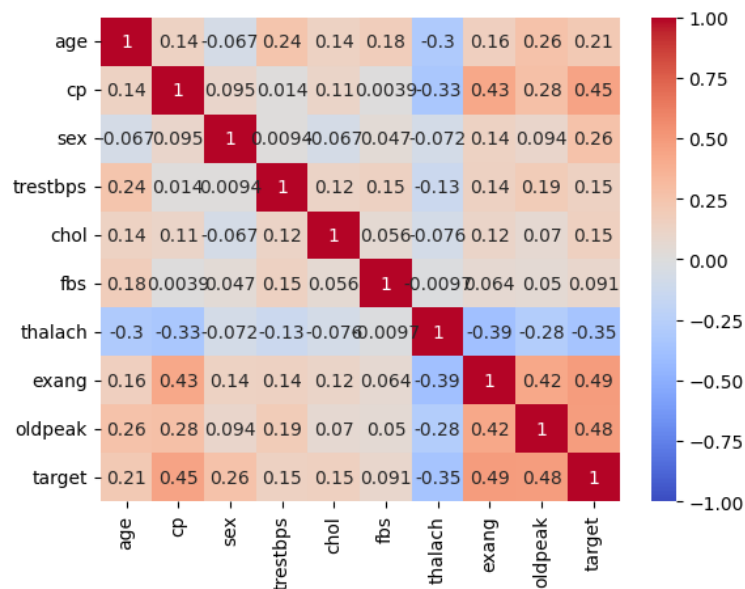


Figura 3: Correlação entre as variáveis

Foi feito também um código que calcula as correlações da variável alvo com todas as outras variáveis do DataFrame, seleciona as maiores correlações em termos de valores absolutos e gera um gráfico de barras para visualizá-las. Isso pode ajudar a identificar quais variáveis possuem uma correlação mais forte com a variável alvo.

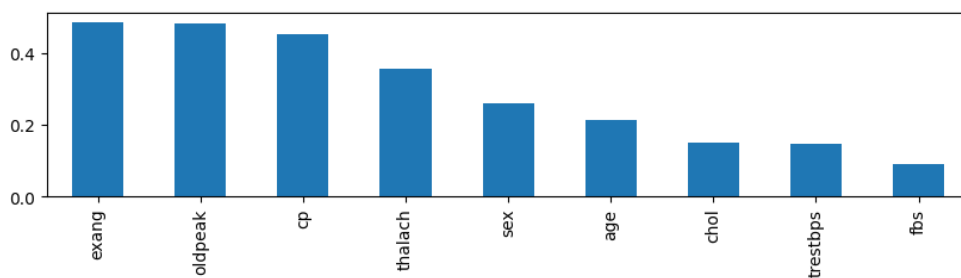


Figura 4: Ordem de relevância entre os atributos

Assim, as variáveis mais relevantes foram exploradas, principalmente as que respondem as nossas perguntas iniciais: a variável 'cp' com os tipos de dores no peito e a variável 'exang' que indica se o indivíduo tem ou não angina induzida por exercício. Faz sentido pensar que essas variáveis possuem correlação, e a figura 3 confirma essa hipótese. Para melhor visualização dessa relação, então, foi feito o seguinte gráfico, onde a escala de cores vai de 0 (verde escuro) e 1 (vermelho escuro) indicando a probabilidade da presença da doença cardíaca:

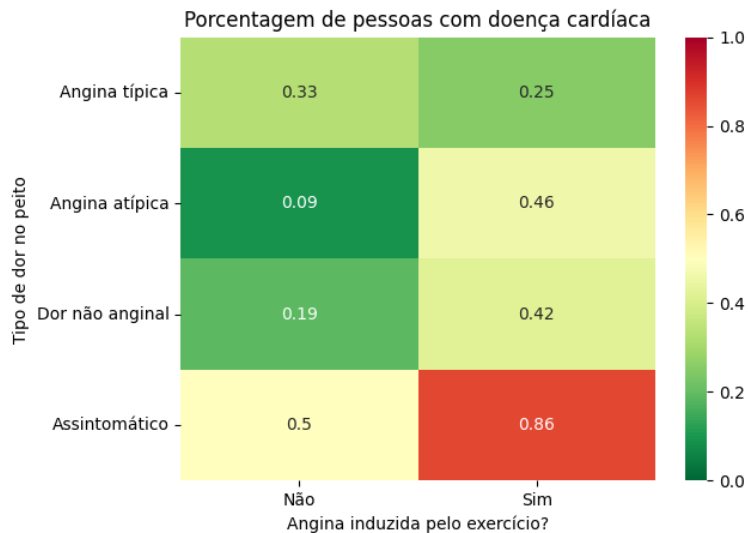


Figura 5: Relação cp x exang

Com isso, nota-se um comportamento não intuitivo: as pessoas que possuem angina atípica parecem estar menos propensas a ter doença cardíaca. Uma pessoa assintomática, por outro lado (que não costuma sentir dores no peito em repouso) que sente angina induzida por exercício físico, parece estar mais propensa. Esse é um comportamento curioso, que será melhor trabalhado posteriormente. Cabe também verificar a relação de alguns atributos especificamente com a saída.

Por exemplo, essa é a relação do tipo de dor no peito com a presença de doença cardíaca:

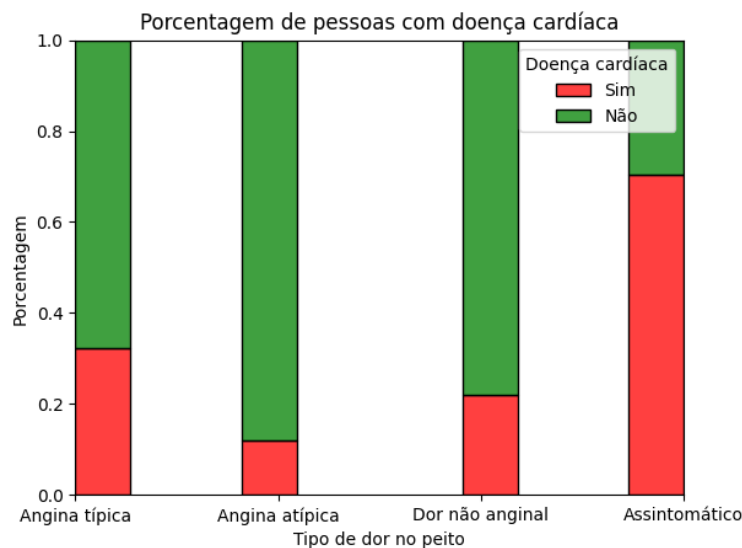


Figura 6: Relação cp x target

Nota-se que a Angina típica parece ter uma relação contrária com a doença cardíaca, o que não é tão natural de pensar. É interessante estudarmos os motivos desse fato estar ocorrendo.

Já a angina induzida por exercício físico está relacionada com a variável alvo como se é esperado: uma frequência maior de indivíduos que possuem a doença cardíaca sentem dores no peito após fazerem atividades físicas.

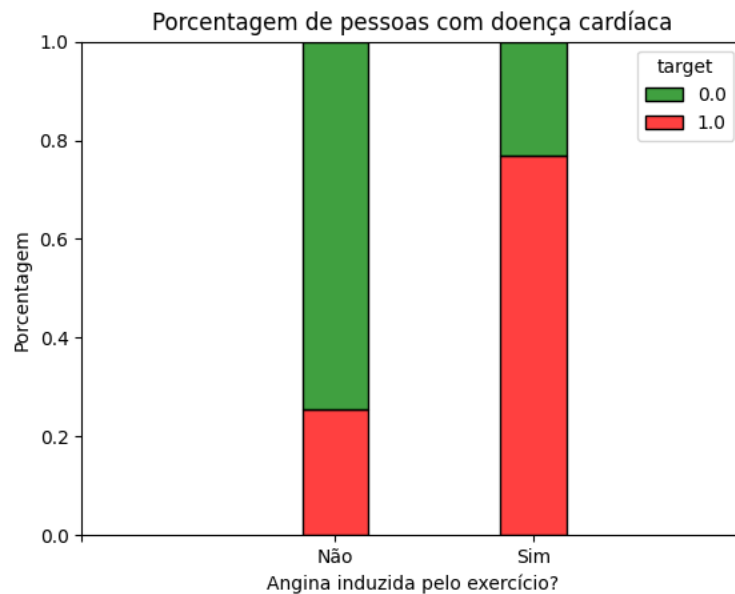


Figura 7: Relação exang x target

As outras variáveis também foram exploradas visualmente, e podem ser visualizadas abaixo:

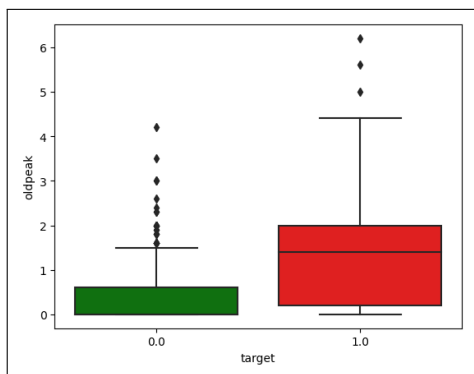


Figura 8: Relação oldpeak x target

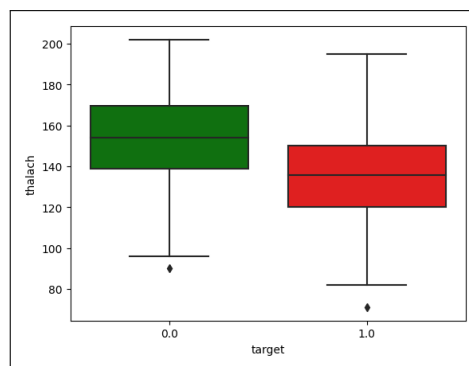


Figura 9: Relação thalach x target

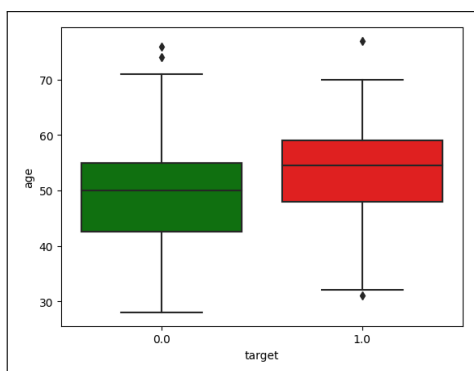


Figura 10: Relação age x target

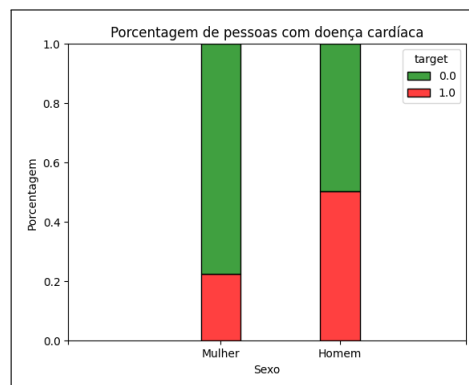


Figura 11: Relação sex x target

3.3 Modelos

Feito isso, alguns modelos de regressão logística foram construídos, cada um com alguma pequena alteração para avaliar o impacto no resultado final.

3.3.1 Modelo 1

Nesse primeiro modelo, todas as variáveis foram consideradas, e as correlações entre elas não foram corrigidas, ou seja, foi feito de forma linear somando todas. A variável 'cp' que possuía 4 categorias foi transformada em 4 sub-variáveis binárias.

Assim, obtemos essas métricas:

Acurácia: 0.7705882352941177
Recall: 0.7746478873239436
Precision: 0.7051282051282052

3.3.2 Modelo 2

Neste foi feita uma pequena alteração: foi incluída no modelo a correlação entre as variáveis 'exang' e 'cp' que se mostrou bastante alta na análise anterior. Entretanto, não notou-se mudanças nas métricas:

Acurácia: 0.7705882352941177
Recall: 0.7746478873239436
Precision: 0.7051282051282052

3.3.3 Modelo 3

Neste modelo a correlação entre 'oldprek' e 'exang' que foi incluída. Notamos um pequeno aumento na precisão, mas o *recall* foi diminuído, e como foi fundamentado antes, para o nosso modelo, o recall é mais interessante.

Acurácia: 0.7764705882352941
Recall: 0.7605633802816901
Precision: 0.72

3.3.4 Modelo 4

Dessa vez a variável 'cp' foi excluída do modelo, e as outras adicionadas de maneira linear sem representar nenhuma correlação. Notamos um curioso aumento significativo na acurácia e na precisão, mas o recall, nosso maior interesse, diminuiu:

Acurácia: 0.8058823529411765
Recall: 0.7323943661971831
Precision: 0.7878787878787878

3.3.5 Modelo 5

A diferença entre este modelo e o anterior é que a correlação entre 'exang' e 'oldprek' (que é possível verificar na figura 4 que é relativamente alta) foi considerada. Assim, aumentamos um pouquinho as métricas, mas a recall continua menor do que considerando a variável de tipos de dores no peito.

Acurácia: 0.8235294117647058
Recall: 0.7464788732394366
Precision: 0.8153846153846154

A partir das métricas que observamos, os modelos que consideram a variável de dores no peito apresentam uma acurácia e precisão melhor, mas o recall, que é nossa principal métrica, já que estamos tratando de diagnóstico de uma doença, é maior quando consideramos este atributo. Vamos aprofundar nossa avaliação introduzindo mais métricas para avaliar os modelos. Entretanto, utilizaremos apenas os mais corretos para cada caso: o modelo 3 que considera as correlações mais fortes mas ainda possui a variável 'cp', e o modelo 5 que não possui a variável mas considera as correlações.

3.3.6 Área sob a curva ROC

É possível visualizar abaixo, no mesmo gráfico, as curvas correspondentes a cada modelo juntamente com suas respectivas áreas:

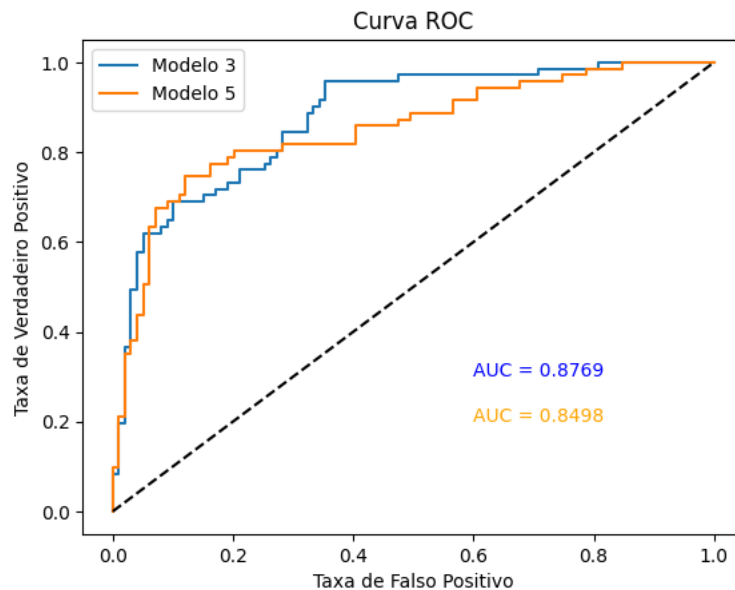


Figura 12: Curva ROC para cada modelo

Notamos que o modelo 3 (que considera os tipos de dor no peito na regressão) resulta em uma curva com maior área. A AUC é uma excelente métrica para comparar modelos, então desconfia-se de que os tipos de dores no peito realmente sejam relevantes para a predição.

3.3.7 AIC

Cabe calcular também a AIC (Akaike Information Criterion)⁴ que é uma métrica boa para indicar a qualidade do modelo em relação à sua complexidade

AIC do modelo 3: 335.72213496067036

AIC do modelo 5: 368.7767662468726

Comparativamente é possível perceber que o modelo 3 também se destaca com relação à importância das variáveis independentes na explicação da variável dependente.

⁴https://en.wikipedia.org/wiki/Akaike_information_criterion

4 Conclusão

O objetivo inicial de estudar a relação das dores no peito com o diagnóstico de doença cardíaca foi feito através de modelos de regressão logística com pequenas alterações entre um e outro, conforme os estudos provenientes da análise exploratória.

A conclusão foi que, de fato, as dores no peito são um atributo relevante para a presença de doença cardíaca, e essa relação foi estudada através da comparação entre dois modelos com boa acurácia: um considerando essa variável e o outro a omitindo. Foi visto que, apesar da acurácia no modelo que desconsidera esse atributo ser maior, outras métricas estatísticas mais relevantes para o contexto desse estudo se mostraram melhores no modelo que o contém.

Em contrapartida, voltando à análise exploratória, notou-se que os tipos de dores no peito pareciam se comportar de maneira curiosa e não-intuitiva quando relacionados com a presença de doença cardíaca. Isso pode ser explicado por diferentes motivos, alheios ao conjunto de dados e aos modelos. Por exemplo, no artigo[3] se discute as limitações da classificação tradicional de dor no peito em angina típica, angina atípica e dor torácica atípica. Os autores argumentam que essa taxonomia tem grandes limitações e pode levar a decisões clínicas equivocadas - isso parece estar bem relacionado ao contexto que estamos discutindo.

Se tratando de limitações do modelo, um ponto importante é a alta correlação entre as variáveis 'cp' e 'exang', que poderiam ter sido melhor exploradas para aprimorar a predição. Em [4] é feita uma investigação da relação entre sintomas de dor no peito e eventos de doença cardíaca isquêmica em homens, onde eles categorizaram a dor no peito em "angina definitiva" (dor no peito que preenchia todos os critérios da OMS) e "angina possível" (dor no peito durante o esforço sem todos os critérios da OMS). Os resultados mostraram que tanto a angina definitiva quanto a angina possível estavam associadas a resultados semelhantes de doença cardíaca isquêmica. Portanto, os pesquisadores combinaram essas duas categorias em uma única categoria de angina. Diante disso, no nosso modelo, a necessidade de tratar melhor a correlação entre essas duas variáveis se faz presente. Esse mesmo artigo também é um alerta para a possibilidade de estudar os modelos para cada sexo separadamente, que poderia ser um fator contribuinte e que não foi considerado dessa forma (o sexo estava presente apenas como uma variável binária entre as variáveis independentes).

Outra limitação clara são os dados utilizados. Eles necessitaram de uma higienização bastante rigorosa e muitos dados tiveram que ser desconsiderados. No final, o dataframe trabalhado tinha apenas 565 linhas, e teve que ser dividido em treino e teste. Dessa forma, seria interessante fazer o mesmo estudo para um conjunto com maior qualidade e quantidade.

Em conclusão, esse estudo demonstrou que a presença de dor no peito é um atributo relevante para o diagnóstico de doença cardíaca. Apesar de algumas limitações no modelo utilizado, a inclusão da variável de dor no peito mostrou-se estatisticamente significativa. Portanto, a resposta para o questionamento: dor no peito é um alarmante para a doença cardíaca? - pode ser melhor desenvolvida em estudos futuros para confirmar a hipótese aqui feita que é: sim.

5 Referências

- [1] Jaffe AS, Goodship TH, Hall RJ. Chest Pain in Patients with Non-ST Elevation Acute Coronary Syndrome—Prognosis and Management. *Circulation*. 2011; 124(23): 2910-2917. doi: 10.1161/CIRCULATIONAHA.110.000861.
- [2] Simmons II, B. (2021). Investigating Heart Disease Datasets and Building Predictive Models. (Tese de mestrado, Elizabeth City State University).
- [3] Sagristà-Sauleda, J., Barrabés, J. A. (ano). Typical angina, atypical angina, and atypical chest pain: is it time to change this terminology?
- [4] Lampe, F. C., Whincup, P. H., Wannamethee, S. G., Ebrahim, S., Walker, M., Shaper, A. G. (1998). Chest pain on questionnaire and prediction of major ischaemic heart disease events in men. *European Heart Journal*, 19(1), 63-73.