

---

# Análise do Medicamento "Dobutamina" em Ecocardiografias de Estresse

---

**Marcelo Angelo**

Ciência de Dados e Inteligência Artificial  
Fundação Getulio Vargas  
Rio de Janeiro, RJ  
marangesfo@gmail.com

## Abstract

Este artigo tem por finalidade determinar se o medicamento "dobutamina" pode ser usado de forma eficaz em um teste para medir o risco de um paciente ter um evento cardíaco.

## 1 Introdução

Para pacientes mais jovens, um teste típico desse risco é chamado de "Ecocardiografia de Estresse". Este teste envolve aumentar a frequência cardíaca do paciente por meio de exercício – frequentemente fazendo o paciente correr em uma esteira – e, em seguida, realizar várias medições, como frequência cardíaca e pressão arterial, bem como medições mais complexas do coração. O problema com este teste é que muitas vezes ele não pode ser usado em pacientes mais velhos cujos corpos não suportam o estresse de exercícios intensos. No entanto, a chave para avaliar o risco é colocar o coração sob estresse antes de realizar as medições relevantes. Embora o exercício não possa ser usado para criar esse estresse em pacientes mais velhos, a dobutamina pode. Portanto, este estudo foi, em parte, uma tentativa de responder a seguinte pergunta: **O teste de ecocardiografia de estresse ainda é eficaz em prever eventos cardíacos quando o estresse no coração é produzido pela dobutamina em vez do exercício?**. Mais especificamente, este estudo procura identificar diferentes modelos considerando combinações distintas de variáveis para analisar as medições realizadas durante o teste de ecocardiografia de estresse a fim de prever se um paciente sofreria um evento cardíaco no ano seguinte. A citação completa do jornal onde os resultados do estudo foram publicados pertence à: Journal of the American College of Cardiology 33.3 (1999) 708-16. [1]

## 2 Dados

Os "eventos cardíacos" que a "Ecocardiografia de Estresse com Dobutamina" tentou prever podem ser divididos em quatro categorias:

- Infarto do miocárdio (MI)
- Revascularização por angioplastia coronária transluminal percutânea (PTCA)
- Cirurgia de revascularização do miocárdio (CABG)
- Morte cardíaca

Como o objetivo do estudo é prever qualquer evento cardíaco não vamos nos atentar às categorias acima citadas mas estas serão retomadas no fim deste artigo.

Os dados utilizados são descritos em UCLA Stress Echocardiography Data

Algumas das variáveis originais foram renomeadas e recodificadas para os conjuntos de dados. Variáveis de eventos que originalmente foram codificadas como 0=sim 1=não foram recodificadas

como 0=não 1=sim. As variáveis `equivecg` e `posecg` foram combinadas em uma nova variável chamada `ecg`. A variável original `hxofcig` tinha os valores 0 .5 1, assumidos para representar grande, moderado e nenhum. `%mphr(b)` foi renomeada para `pctMphr`.

Atributo	Descrição
<code>bhr</code>	Frequência cardíaca basal
<code>basebp</code>	Pressão arterial basal
<code>basedp</code>	Produto duplo basal (= <code>bhr</code> x <code>basebp</code> )
<code>pkhr</code>	Frequência cardíaca de pico
<code>sbp</code>	Pressão arterial sistólica
<code>dp</code>	Produto duplo (= <code>pkhr</code> x <code>sbp</code> )
<code>dose</code>	Dose de dobutamina administrada
<code>maxhr</code>	Frequência cardíaca máxima
<code>%mphr(b)</code>	% da frequência cardíaca máxima prevista atingida
<code>mbp</code>	Pressão arterial máxima
<code>dpmxdo</code>	Produto duplo na dose máxima de dobutamina
<code>dobdose</code>	Dose de dobutamina na qual ocorreu o produto duplo máximo
<code>age</code>	Idade
<code>gender</code>	Gênero (masculino = 0)
<code>baseef</code>	Fração de ejeção cardíaca basal
<code>dobef</code>	Fração de ejeção com dobutamina
<code>chestpain</code>	0 significa que experimentou dor no peito
<code>posecg</code>	Sinais de ataque cardíaco no ECG (0 = sim)
<code>equivecg</code>	ECG é inconclusivo (0 = sim)
<code>restwma</code>	Anomalia de movimento da parede no ecocardiograma (0 = sim)
<code>posse</code>	Ecocardiograma de estresse foi positivo (0 = sim)
<code>newmi</code>	Novo infarto do miocárdio (0 = sim)
<code>newptca</code>	Angioplastia recente (0 = sim)
<code>newcabg</code>	Cirurgia de bypass recente (0 = sim)
<code>death</code>	Morreu (0 = sim)
<code>hxofht</code>	Histórico de hipertensão (0 = sim)
<code>hxofdm</code>	Histórico de diabetes (0 = sim)
<code>hxofcig</code>	Histórico de tabagismo (0 = sim)
<code>hxofmi</code>	Histórico de ataque cardíaco (0 = sim)
<code>hxofptca</code>	Histórico de angioplastia (0 = sim)
<code>hxofcabg</code>	Histórico de cirurgia de bypass (0 = sim)
<code>any event</code>	Variável de resultado: "morte ou <code>newmi</code> ou <code>newptca</code> ou <code>newcabg</code> " (0 = qualquer um dos eventos)

Table 1: Explicação das Abreviações de Medições de Dados

### 3 Análise Exploratória

A análise exploratória dos dados foi realizada em notebook Python e pode ser encontrada de maneira completa no seguinte GitHub: MasFz.

A fim de manter este artigo conciso e, apenas com os principais resultados, traremos apenas alguns trechos do notebook como argumento de autoridade para o que se afirma em seguida.

Para as variáveis numéricas, realizamos o `plot` dos histogramas e dos `boxplots` a fim de analisar distribuições dos dados e outliers. Os resultados foram os seguintes:

#### 3.1 Análise Descritiva das Variáveis Numéricas

Os histogramas e `boxplots` criados foram realizados após a higienização dos dados (remoção de dados faltantes e colunas duplicadas). Após isso podemos destacar, dentre as figuras produzidas, algumas inferências como:

# Histogramas das Variáveis Numéricas

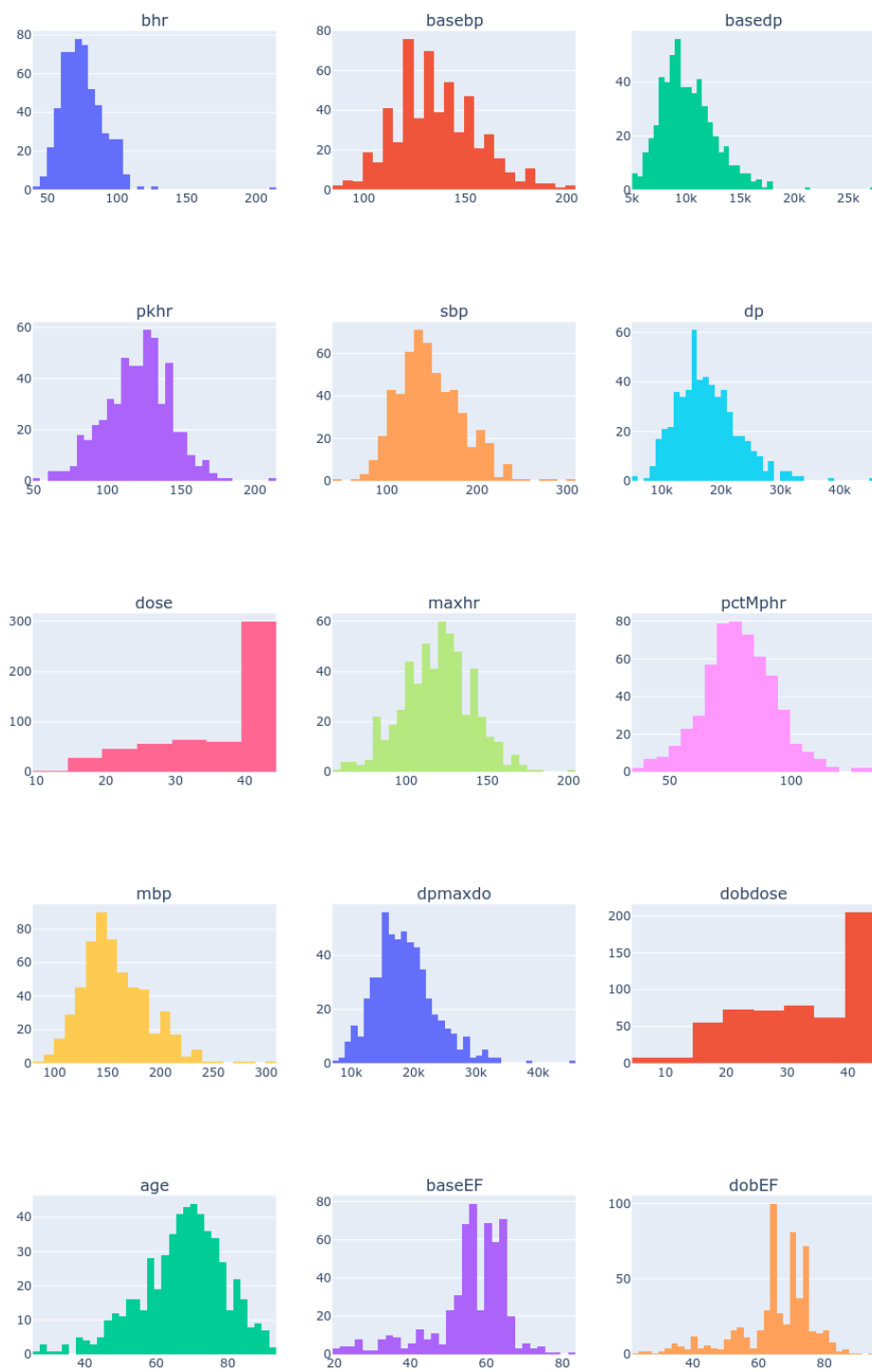


Figure 1: Histogramas

### Boxplots das Variáveis Numéricas

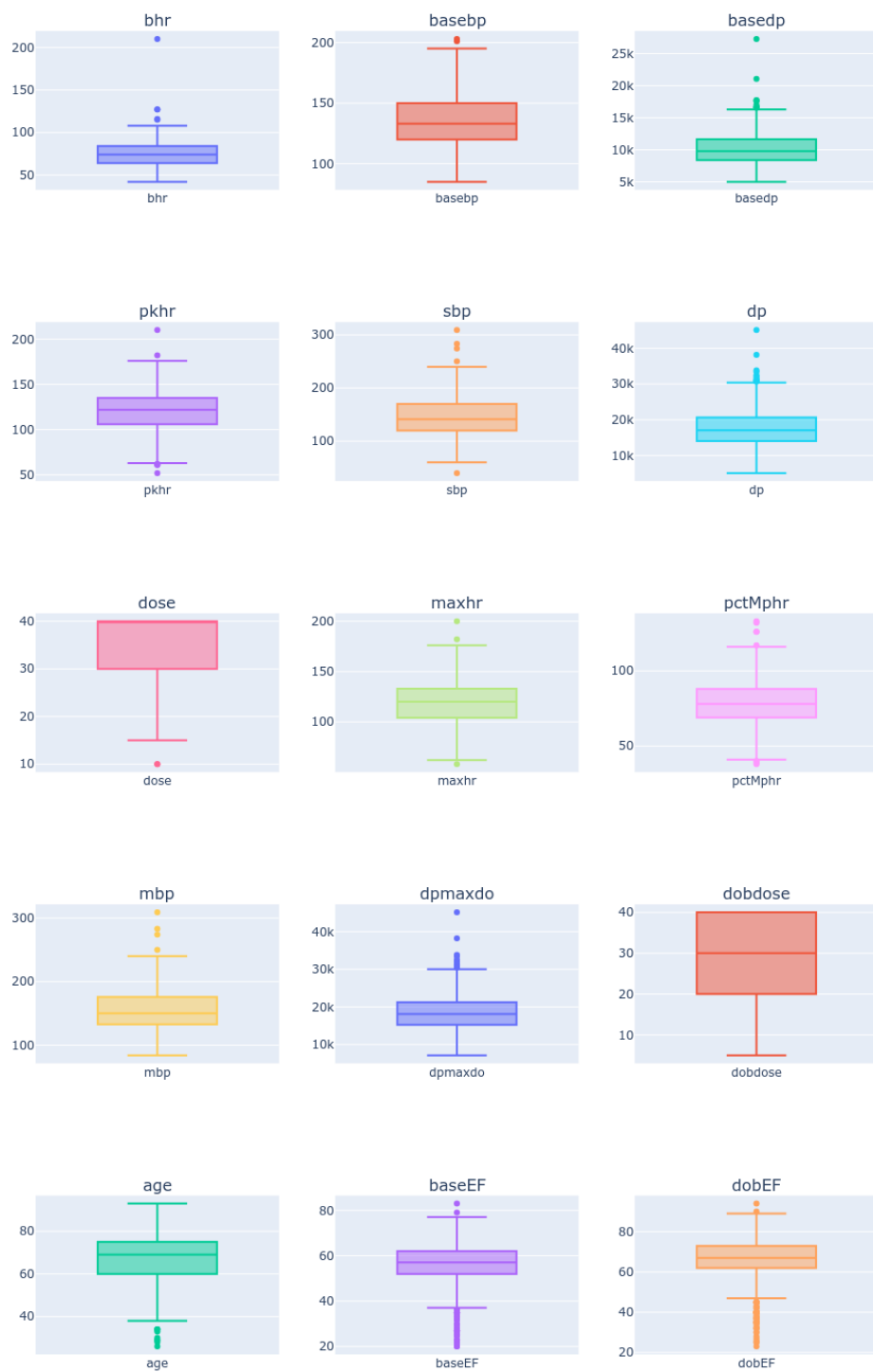


Figure 2: Boxplots

### 3.2 bhr (Frequência Cardíaca Basal)

**Histograma:** Mostra uma curva em forma de sino, indicando que a maioria dos valores está concentrada em torno de uma média, com uma diminuição simétrica nas extremidades.

**Boxplot:** Provavelmente tem uma mediana central e quartis bem definidos, com poucos outliers.

### 3.3 basedp (Produto Duplo Basal)

**Histograma:** Curva em forma de sino, com a maioria dos valores centralizados em torno da média.

**Boxplot:** Mediana central, com distribuição simétrica dos quartis e poucos outliers.

### 3.4 pkhr (Frequência Cardíaca de Pico)

**Histograma:** Distribuição normal com um pico central.

**Boxplot:** Simetria em torno da mediana e poucos outliers.

### 3.5 sbp (Pressão Arterial Sistólica)

**Histograma:** Forma de sino, indicando valores concentrados em torno de uma média.

**Boxplot:** Mediana central, com quartis bem distribuídos e poucos outliers.

### 3.6 dp (Produto Duplo)

**Histograma:** Distribuição normal, com valores centrados.

**Boxplot:** Quartis simétricos e poucos outliers.

### 3.7 maxhr (Frequência Cardíaca Máxima)

**Histograma:** Forma de sino.

**Boxplot:** Mediana central e distribuição simétrica dos quartis.

### 3.8 pctMphr (% da Frequência Cardíaca Máxima Prevista Atingida)

**Histograma:** Curva simétrica.

**Boxplot:** Quartis bem definidos e simétricos.

### 3.9 mbp (Pressão Arterial Máxima)

**Histograma:** Distribuição centrada em torno de uma média.

**Boxplot:** Mediana central e poucos outliers.

### 3.10 dpmaxdo (Produto Duplo na Dose Máxima de Dobutamina)

**Histograma:** Forma de sino.

**Boxplot:** Quartis simétricos e poucos outliers.

### 3.11 age (Idade)

**Histograma:** Distribuição normal centrada.

**Boxplot:** Mediana central com quartis bem distribuídos e poucos outliers.

### 3.12 Análise Descritiva das Variáveis Categóricas

Para as variáveis categóricas foi realizada a frequência de ocorrência de cada variável:

### 3.13 Matriz de Significância

Após a descrição de nossas variáveis, partimos para análise de significância entre as variáveis:

Variável	Categoria	Contagem
gender	female	338
	male	220
hxofCig	non-smoker	298
	moderate	138
	heavy	122
ecg	normal	311
	equivocal	176
	MI	71

Table 2: Distribuição das variáveis categóricas



Figure 3: Trecho da Matriz de Significância

A Matriz completa de significância está no notebook Python.

Ainda foram realizados o Teste t nas variáveis numéricas e o Teste Qui-Quadrado para as variáveis categóricas, todas em relação à variável 'any.event'.

#### Ranking de Significância das Variáveis com relação a any.event

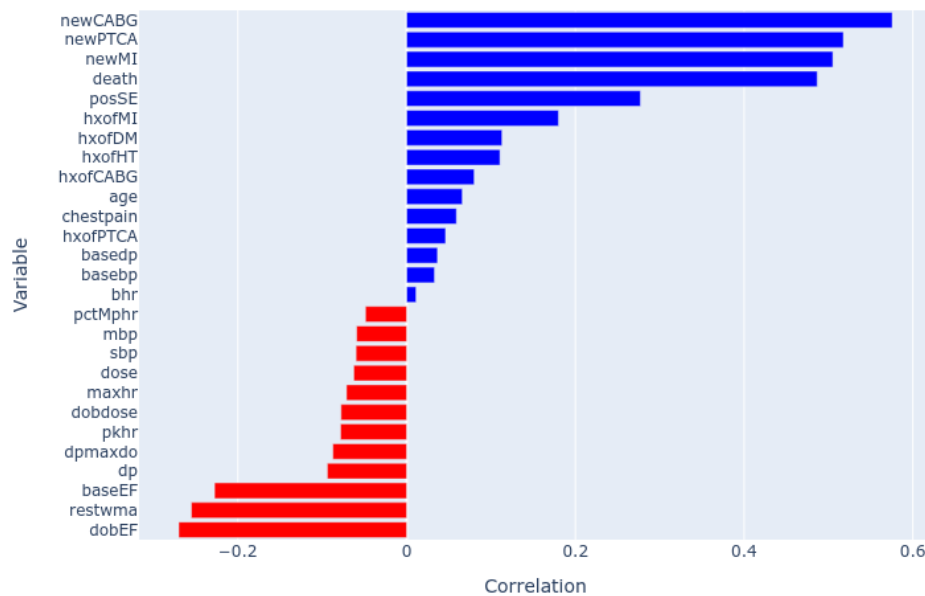


Figure 4: Ranking de Sinificância de cada cariável com algum evento cardíaco

### 3.14 Teste t

O teste t compara as médias das variáveis numéricas entre dois grupos (com e sem eventos cardíacos). Um valor p pequeno (geralmente  $< 0.05$ ) indica que há uma diferença estatisticamente significativa entre as médias dos dois grupos.

t-statistic: Indica a magnitude da diferença entre as médias. Valores maiores (em valor absoluto) indicam maior diferença. p-value: Indica a probabilidade de que a diferença observada tenha ocorrido por acaso. Valores menores que 0.05 indicam significância estatística.

### 3.15 Teste Qui-Quadrado

O teste qui-quadrado verifica a independência entre variáveis categóricas e eventos cardíacos. Um valor p pequeno (geralmente  $< 0.05$ ) indica que há uma associação estatisticamente significativa entre a variável categórica e os eventos cardíacos.

chi2: Indica a magnitude da associação entre as variáveis. Valores maiores indicam maior associação. p-value: Indica a probabilidade de que a associação observada tenha ocorrido por acaso. Valores menores que 0.05 indicam significância estatística.

### 3.16 Conclusão

Com base nos resultados obtidos para as variáveis podemos selecionar as variáveis de fato significativas.

Variáveis Numéricas Significativas: Variáveis como dp, dpmxdo, baseEF, dobEF, restwma, posSE, newMI, newPTCA, newCABG, death, hxofHT, hxofDM e hxofMI são significativamente diferentes entre os grupos com e sem eventos cardíacos. Isso sugere que essas variáveis podem ser bons preditores de eventos cardíacos.

Variável	t-statistic	p-value
bhr	-0.2635	0.7923
basebp	-0.7798	0.4358
basedp	-0.8633	0.3884
pkhr	1.8534	0.0644
sbp	1.4177	0.1568
dp	2.2278	0.0263
dose	1.4819	0.1389
maxhr	1.6856	0.0924
pctMphr	1.1528	0.2495
mbp	1.4015	0.1616
dpmxdo	2.0707	0.0389
dobdose	1.8414	0.0661
age	-1.5601	0.1193
baseEF	5.5142	5.37e-08
dobEF	6.6151	8.74e-11
chestpain	-1.3937	0.1640
restwma	6.2243	9.53e-10
posSE	-6.8010	2.69e-11
newMI	-13.8040	1.75e-37
newPTCA	-14.2657	1.43e-39
newCABG	-16.5947	1.56e-50
death	-13.1358	1.61e-34
hxofHT	-2.6254	0.0089
hxofDM	-2.6822	0.0075
hxofMI	-4.3144	1.89e-05
hxofPTCA	-1.0896	0.2764
hxofCABG	-1.8947	0.0586

Table 3: Resultados do Teste t para Variáveis Numéricas

Variável	chi2	p-value
gender	3.0739	0.0796
hxofCig	2.0794	0.3536
ecg	19.8010	5.01e-05

Table 4: Resultados do Teste Qui-Quadrado para Variáveis Categóricas

Variáveis Categóricas Significativas: A variável ecg mostra uma associação significativa com eventos cardíacos, indicando que o resultado do eletrocardiograma pode ser um preditor importante.

## 4 Metodologia

Em um primeiro momento utilizaremos todas as colunas significantes em nosso modelo. Como estamos lidando com 14 colunas e um pouco mais de 500 registros esperamos encontrar um overfitting dos dados por conta da complexidade do modelo.

### 4.0.1 Escolha dos Métodos Estatísticos

Optamos por utilizar a **Regressão Logística** para prever a probabilidade de um evento binário (evento cardíaco) com base em variáveis independentes. As razões para essa escolha incluem:

- Adequação para o tipo de dados (binário).
- Permite interpretar os coeficientes como razões de chances.
- É um método robusto e amplamente utilizado em estudos clínicos.

Para lidar com a **Inovação e Complexidade**, adicionamos técnicas de regularização (Lasso, Ridge) para lidar com a multicolinearidade.



## 4.1 Implementação dos Métodos Numéricos

### 4.1.1 Escolha do Método Numérico para Regressão Logística

Implementamos os seguintes métodos numéricos para a regressão logística:

- **Método de Newton-Raphson:** Um método iterativo para encontrar zeros de uma função.
- **Iterative Weighted Least Squares (IWLS):** Usado em regressão logística para encontrar os coeficientes do modelo.
- **Fisher Scoring:** Uma variante do método de Newton-Raphson, geralmente mais estável.

### 4.1.2 Validação do Modelo

Para validar nosso modelo, utilizamos:

- **Curva ROC e AUC:** Avaliamos a capacidade discriminativa do modelo.

Com base nesta análise chegamos em:

Classe	Precisão (Precision)	Recall	F1-Score	Suporte (Support)
0.0	1.00	1.00	1.00	469
1.0	0.99	1.00	0.99	89
Acurácia (Accuracy)	1.00			
Média Macro (Macro Avg)	0.99	1.00	1.00	558
Média Ponderada (Weighted Avg)	1.00	1.00	1.00	558

Table 5: Relatório de Classificação

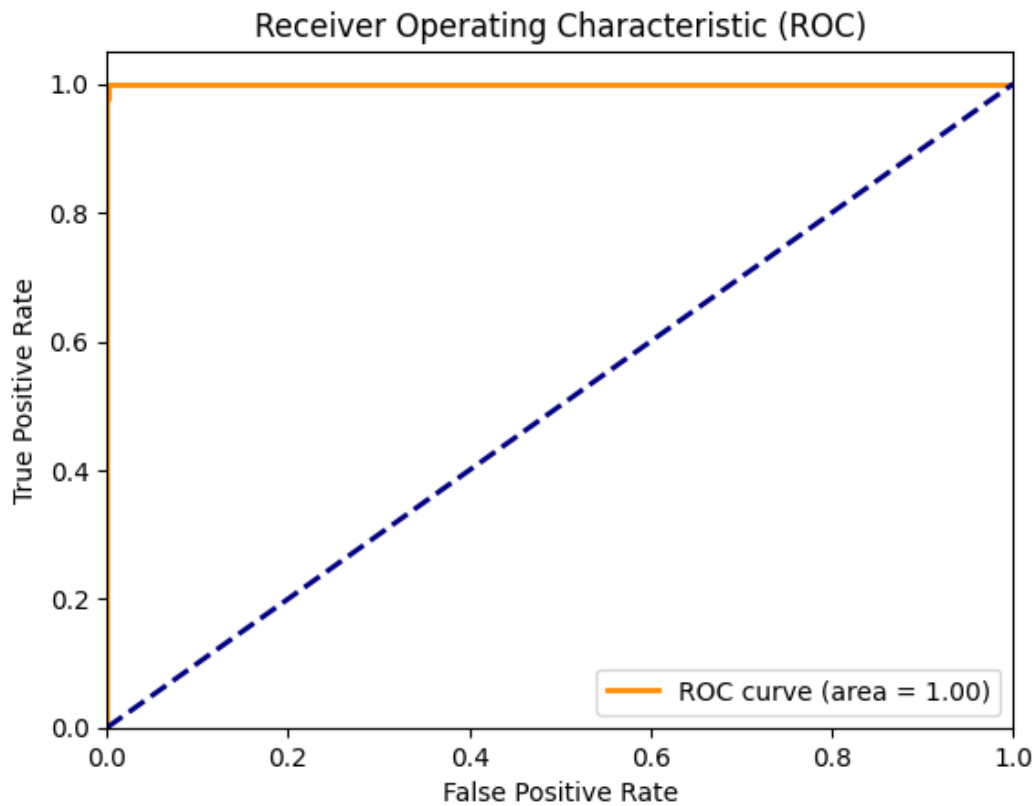


Figure 5: ROC - Primeiro Modelo

Como esperado, encontramos um modelo muito complexo para a quantidade de nossos dados. Em nosso notebook, tentamos reduzir a complexidade de nosso modelo mas sem sucesso, ainda foi cogitado utilizar técnicas de Machine Learning como Random Forest mas optou-se por utilizar algumas dessas técnicas na modelagem subsequente. Dessa maneira optou-se por reduzir o número de variáveis em 9 modelos distintos e computar a validação de cada um desses modelos.

Os resultados obtidos foram:

Modelo	Acurácia	Precisão	Recall	AIC
Model 1	0.928571	1.000000	0.555556	317.295250
Model 2	0.922619	1.000000	0.518519	324.116276
Model 3	0.916667	1.000000	0.481481	319.345090
Model 4	0.886905	0.900000	0.333333	310.130421
Model 5	0.886905	0.900000	0.333333	308.838248
Model 6	0.880952	0.888889	0.296296	312.369648
Model 7	0.839286	0.000000	0.000000	329.864108
Model 8	0.839286	0.000000	0.000000	323.732169
Model 9	0.880952	1.000000	0.259259	313.486258

Table 6: Resultados dos Modelos

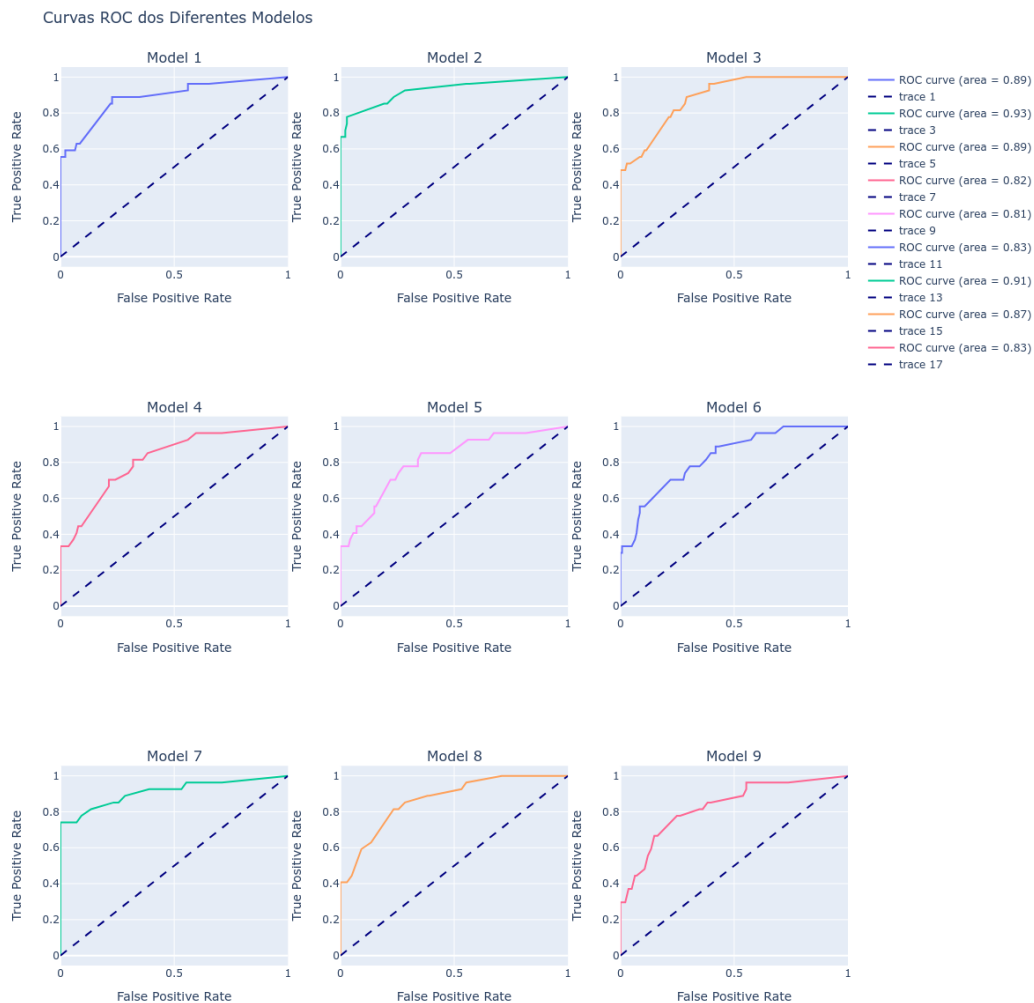


Figure 6: ROC - Diferentes Modelos

## 4.2 Interpretação dos Resultados

### 4.3 Modelos 1 a 3

- **Acurácia:** Alta, variando de 0.9167 a 0.9286.
- **Precisão:** Perfeita (1.0000), indicando que todos os casos previstos como positivos são verdadeiramente positivos.
- **Recall:** Moderado, variando de 0.4815 a 0.5556, indicando que o modelo está identificando entre 48.15% a 55.56% dos verdadeiros positivos.
- **AIC:** Menor nos Modelos 1 e 3, sugerindo que esses modelos são melhores ajustados em termos de complexidade e ajuste aos dados.

### 4.4 Modelos 4 a 6

- **Acurácia:** Moderada, em torno de 0.8869 a 0.8809.
- **Precisão:** Alta (0.9000 a 0.8889), mas não perfeita.
- **Recall:** Baixo, variando de 0.2963 a 0.3333, indicando que o modelo está identificando uma menor proporção dos verdadeiros positivos.
- **AIC:** Menor nos Modelos 4 e 5, sugerindo que esses modelos são melhores ajustados em termos de complexidade e ajuste aos dados do que o Modelo 6.

### 4.5 Modelos 7 e 8

- **Acurácia:** Mais baixa (0.8393).
- **Precisão e Recall:** Zero, indicando que o modelo não identificou corretamente nenhum dos casos positivos.
- **AIC:** Alto, indicando um ajuste ruim do modelo aos dados.

### 4.6 Modelo 9

- **Acurácia:** Moderada (0.880952).
- **Precisão:** Perfeita (1.0000), indicando que todos os casos previstos como positivos são verdadeiramente positivos.
- **Recall:** Muito baixo (0.259259), indicando que o modelo está identificando apenas 25.93% dos verdadeiros positivos.
- **AIC:** Relativamente alto, indicando um ajuste ruim do modelo aos dados comparado aos Modelos 1 a 5.

## 4.7 Conclusão

- **Melhores Modelos:** Os Modelos 1, 2 e 3 têm a melhor combinação de alta acurácia e precisão, juntamente com recall moderado e AIC relativamente baixo, tornando-os os mais equilibrados.
- **Modelos com Baixo Desempenho:** Os Modelos 7 e 8 apresentaram o pior desempenho com precisão e recall zero.
- **Modelos Intermediários:** Os Modelos 4, 5 e 6 têm boa precisão, mas o recall mais baixo sugere que estão perdendo muitos casos positivos.

Para os Modelos acima a seguinte metodologia foi seguida:

## 4.8 Modelos Usados

Nos modelos anteriores, utilizamos o algoritmo de Gradient Boosting Classifier com validação cruzada e ajuste de hiperparâmetros através de GridSearchCV. Este método utiliza algoritmos de otimização como métodos de gradiente, que são implementados internamente pelo scikit-learn.

## 4.9 Discussão sobre Métodos Numéricos

**Gradiente:** Este método é adequado para problemas de classificação binária e é especialmente útil quando se lida com um número razoável de variáveis e dados. Ele fornece uma convergência rápida e é amplamente utilizado em práticas de machine learning.

Dada a complexidade e o tamanho dos dados, a escolha do Gradient Boosting com métodos de otimização internos, como os métodos de gradiente, faz-se de grande importância. Ele equilibra a necessidade de precisão e eficiência computacional, garantindo um bom desempenho na modelagem de dados complexos.

## 5 Tratamento da Incerteza

### 5.1 Estimativas e Predições

**Incerteza nas Estimativas:** Nos modelos anteriores, a incerteza foi abordada através da validação cruzada e da avaliação das métricas de desempenho (acurácia, precisão, recall e AIC). Estas métricas fornecem uma visão sobre a variabilidade e a confiança nas predições.

Para uma abordagem ainda mais robusta, quantificar as várias fontes de incerteza (como variância dos parâmetros, intervalos de confiança, e incerteza nos dados de entrada) e discutir como elas impactam as predições seria ideal.

## 6 Resumo da Modelagem

**Métodos Numéricos:** A escolha do Gradient Boosting parece apropriada, fornecendo um equilíbrio entre precisão e eficiência computacional.

**Tratamento da Incerteza:** A incerteza foi tratada através da validação cruzada e das métricas de desempenho. Para uma abordagem mais robusta, a quantificação explícita da incerteza através de métodos Bayesianos poderia ser considerada no futuro.

## 7 Conclusão

Em nossa análise concluímos que de fato a utilização do medicamento dobutamina é possível e traz uma boa previsão com relação à eventos cardíacos.

Esta análise consistiu de 2 maneiras diferentes sendo ambas descritas em detalhes na seção de metodologia. A primeira considera todas as variáveis significantes enquanto a segunda seleciona de maneira arbitrária 8 variáveis significantes e constrói o modelo em cima dessas variáveis.

Com relação à segunda abordagem podemos destacar ainda que a bondade do ajuste dos modelos foi avaliada utilizando várias métricas de desempenho:

Os modelos foram avaliados com base na acurácia, precisão e recall. Modelos com alta acurácia, como os Modelos 1 a 3, mostram um bom ajuste geral. A precisão perfeita (1.0000) nesses modelos indica que todos os casos previstos como positivos são verdadeiramente positivos, enquanto o recall moderado sugere que o modelo está identificando entre 48.15% a 55.56% dos verdadeiros positivos.

A métrica AUC-ROC foi utilizada para medir a capacidade discriminativa dos modelos. Modelos com AUC-ROC alta indicam uma boa capacidade de discriminação entre classes, com o Modelo 1 apresentando uma área sob a curva de 0.9286.

O Akaike Information Criterion (AIC) foi calculado para avaliar a qualidade do ajuste em termos de complexidade do modelo. Modelos com AIC menor, como os Modelos 1 e 3, sugerem um ajuste melhor em termos de complexidade e ajuste aos dados.

Dessa maneira acredita-se ter alcançado todos os objetivos da atividade. Uma vez que foram feitas diferentes modelagens para um mesmo problema sendo a primeira puramente estatística e as demais com recursos de Machine Learning a fim de aprimorar o que foi feito na primeira etapa.

## 7.1 Limitações

Primeiramente, a sensibilidade (recall) é moderada a baixa em alguns modelos, indicando que nem todos os eventos positivos são corretamente identificados, o que pode ser crítico em contextos de saúde. Além disso, a complexidade dos modelos e o uso intensivo de parâmetros podem levar a overfitting.

A falta de técnicas adicionais para tratar a incerteza, como métodos bayesianos, também pode limitar a robustez das previsões. Por fim, a dependência de um único tipo de modelo sem comparações extensivas com outras abordagens pode restringir a generalização dos resultados.

## 7.2 Direções Futuras

Além de implementações futuras a fim de reduzir limitações levantadas no tópico anterior e nos demais tópicos ao longo do texto, existem outros estudos que podem ser realizados com base no Ecocardiografia de Estresse.

Por exemplo, o nosso estudo se baseia se o paciente teve algum evento cardíaco num intervalo de 1 ano após os exames, estudos mais recentes iniciaram as análises considerando um tempo maior após os exames.[2]

## Agradecimentos

Gostaria de agradecer ao professor Luiz Max Fagundes de Carvalho pelas sugestões da melhor forma de conduzir este trabalho, bem como as aulas de Modelagem Estatística ao longo do período, as quais foram cruciais para adquirir o conhecimento necessário para a realização deste estudo.

## Referências

- [1] Janine Krivokapich et al. “Prognostic value of dobutamine stress echocardiography in predicting cardiac events in patients with known or suspected coronary artery disease”. In: *Journal of the American College of Cardiology* 33.3 (1999), pp. 708–716.
- [2] Nobutaka Noto et al. “Long-term prognostic impact of dobutamine stress echocardiography in patients with Kawasaki disease and coronary artery lesions: a 15-year follow-up study”. In: *Journal of the American College of Cardiology* 63.4 (2014), pp. 337–344.