

Aqui você coloca o título

Aqui você coloca o autor(s)

## Resumo

Aqui você escreve o resumo

## Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Modelagem</b>	<b>1</b>
2.1	Dados de teste e treino . . . . .	1
2.2	Reajuste aos dados completos . . . . .	2
2.3	Equações dos modelos . . . . .	3

## 1 Introdução

## 2 Modelagem

### 2.1 Dados de teste e treino

Realizamos a divisão da base de dados completa em outras duas: Dados de treino e Dados de teste, para assim pode verificar se os modelos propostos são bons para previsão fora da amostra, sem problemas de **overfitting** e **underfitting**. Assim obtivemos as seguintes medidas para os dados de teste utilizando os modelos ajustados com os dados de treino:

Tabela 1: Estatísticas do Modelo 1 ajustado com os dados de treino

term	estimate	std.error	statistic	p.value
(Intercept)	-1.392	0.289	-4.81	0
Age	-0.037	0.009	-4.15	0
Pclass_X1	2.610	0.328	7.95	0
Pclass_X2	1.207	0.279	4.33	0
Sex_female	2.571	0.239	10.74	0

Percebemos que esse modelo é um modelo de regressão logística comum, em que conseguimos obter erro padrão, estatística e p-valores, podendo fazer uma robusta análise inferencial.

Tabela 2: Estatísticas do Modelo 2 ajustado com os dados de treino

term	estimate	penalty
(Intercept)	-1.244	0
Age	-0.025	0
Pclass_X1	1.879	0
Pclass_X2	0.817	0
Sex_female	2.124	0

Ao contrário do modelo de cima, esse é um modelo focado para previsão dos dados, envolvendo mais técnicas de aprendizado de máquina, assim não podemos realizar a análise inferencial.

Tabela 3: Métricas de Avaliação do Modelo 1 nos Dados de Treino

.metric	.estimate
accuracy	0.777
kap	0.525
precision	0.786
sensitivity	0.860
specificity	0.653

Tabela 4: Métricas de Avaliação do Modelo 2 nos Dados de Treino

.metric	.estimate
accuracy	0.771
kap	0.507
precision	0.770
sensitivity	0.879
specificity	0.611

Percebemos que as métricas nos dois Modelos são muito próximas, porém o Modelo 1 que utiliza a engina glm padrão, Precisão e Especificidade e o Kappa de Cohen. O Modelo 2 tem Sensibilidade superior, mas quanto a Acurácia os dois modelos empatam. No entanto, são boas métricas para ambos os modelos assim seguiremos com eles, porém agora unindo os dados de treino e de teste.

## 2.2 Reajuste aos dados completos

Tabela 5: Estatísticas do Modelo 1

term	estimate	std.error	statistic	p.value
(Intercept)	-1.326	0.248	-5.35	0
Age	-0.037	0.008	-4.83	0
Pclass_X1	2.581	0.281	9.17	0
Pclass_X2	1.271	0.244	5.21	0
Sex_female	2.523	0.207	12.16	0

Ajustando o Modelo 1, com todos os dados ainda obtemos significância em todas as variáveis.

Tabela 6: Métricas de Avaliação do Modelo 1

.metric	.estimate
accuracy	0.789
kap	0.558
precision	0.811
sensitivity	0.840
specificity	0.714

Em relação as métricas, acontece algo curioso em relação ao modelo ajustado com os dados de teste, em que apenas a Sensibilidade acaba aumentando, mas ainda são bons valores, a Acurácia é de 78,9%.

Tabela 7: Estatísticas do Modelo 2

term	estimate	penalty
(Intercept)	-1.210	0
Age	-0.024	0
Pclass_X1	1.858	0
Pclass_X2	0.870	0
Sex_female	2.097	0

Ajustando o Modelo 2, possuímos valores extremamente semelhantes para as estimativas dos  $\beta'$ s, assim as previsões devem permanecer parecidas.

Tabela 8: Métricas de Avaliação do Modelo 2

.metric	.estimate
accuracy	0.789
kap	0.558
precision	0.811
sensitivity	0.840
specificity	0.714

Para o modelo 2, aconteceu uma conjuntura semelhante ao modelo 1, em que as métricas de avaliação acabam diminuindo do Modelo com os Dados de Treino para o Modelo com todos os dados, porém nesse caso nem a Sensibilidade aumentou. No entanto, analisando os valores percebemos que eles são semelhantes aos valores do Modelo 1.

## 2.3 Equações dos modelos

*Modelo 1:*

$$\log \left[ \frac{P(\widehat{..y = 1})}{1 - P(\widehat{..y = 1})} \right] = -1.39 - 0.04(\text{Age}) + 2.61(\text{Pclass\_X1}) + 1.21(\text{Pclass\_X2}) + 2.57(\text{Sex\_female}) \quad (1)$$