

# **Análise da reprovação em Apc (Algoritmo de Programação de Computadores) pelos alunos do curso de Ciência da Computação**

Análise prévia

JOÃO VITOR RODRIGUES

LETÍCIA FONSECA

MAXSUELL NINA

# INTRODUÇÃO

## Tópicos

- Análise descritiva
- Regressão logística
- Matriz de confusão
- Resultados

# Variáveis do Banco de dados

## Quantitativa Discreta

- Qual sua idade?

## Qualitativa Ordinal

- Como você se descreveria em termos de programação?
- Fez o ensino médio em escola pública ou particular?
- Quanto tempo demora para chegar na Universidade?

## Qualitativa Nominal

- Por quais motivos você decidiu entrar no curso?
- Já fez uma Graduação ou Curso Técnico na área de TI?
- É a sua primeira vez cursando Algoritmos e Programação de Computadores (APC)?
- Já sabia programar antes de entrar na UnB?
- Já está trabalhando ou estagiando na área de TI?
- Você viu alguma coisa de programação na escola?
- Você veio de outra cidade para estudar na UnB?
- Consegue estudar em casa com tranquilidade?
- Consegue estudar na faculdade com tranquilidade?
- Possui computador/notebook próprio?
- Utiliza alguma instalação da UnB com computador?
- Possui internet de qualidade no lugar onde mora?

# O que é uma regressão logística?

A regressão logística é um método estatístico utilizado para modelar a relação entre uma variável dependente categórica (ou binária) e uma ou mais variáveis independentes, que podem ser categóricas ou contínuas. Ela é comumente utilizada para prever a probabilidade de ocorrência de um evento binário com base em um conjunto de variáveis independentes.

Ao contrário da regressão linear, que é usada para prever valores contínuos, a regressão logística é adequada para problemas em que a variável dependente é binária, ou seja, possui apenas duas categorias, como "sim" ou "não", "verdadeiro" ou "falso", "1" ou "0".

O modelo de regressão logística utiliza a função logística (também conhecida como função sigmóide) para transformar a soma ponderada das variáveis independentes em uma probabilidade entre 0 e 1. Essa probabilidade representa a chance de a variável dependente pertencer a uma das categorias.

## Equação geral do modelo

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}}$$

# Regressão

**Os dados fornecidos parecem ser resultados de uma regressão linear múltipla que tenta prever se os alunos estão cursando Algoritmos e Programação de Computadores (APC) pela primeira vez com base em várias variáveis independentes. Vamos analisar os resultados:**

# Resultados:

OLS Regression Results			
Dep. Variable:	É a sua primeira vez cursando Algoritmos e Programação de Computadores (APC)?	R-squared:	0.315
Model:	OLS	Adj. R-squared:	-0.059
Method:	Least Squares	F-statistic:	0.8418
Date:	Wed, 07 Feb 2024	Prob (F-statistic):	0.669
Time:	23:57:07	Log-Likelihood:	-34.323
No. Observations:	89	AIC:	118.6
Df Residuals:	44	BIC:	174.5
Df Model:	24		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Qual sua idade?	-0.0046	0.009	-0.518	0.607	-0.023	0.013
Já fez uma Graduação ou Curso Técnico na área de TI?	0.2744	0.193	1.422	0.162	-0.115	0.663
Já sabia programar antes de entrar na UnB?	0.0886	0.207	0.429	0.670	-0.328	0.505
Já está trabalhando ou estagiando na área de TI?	-0.1459	0.223	-0.653	0.517	-0.596	0.304
Você viu alguma coisa de programação na escola?	0.0592	0.216	0.274	0.786	-0.377	0.495
Você veio de outra cidade para estudar na UnB?	-0.0062	0.217	-0.028	0.978	-0.444	0.432
Consegue estudar em casa com tranquilidade?	0.2676	0.166	1.613	0.114	-0.067	0.602
Consegue estudar na faculdade com tranquilidade?	-0.0450	0.167	-0.270	0.789	-0.382	0.292
Possui computador/notebook próprio?	-0.1085	0.225	-0.482	0.632	-0.562	0.345
Utiliza alguma instalação da UnB com computador?	-0.1884	0.261	-0.723	0.474	-0.714	0.337
Possui internet de qualidade no lugar onde mora?	0.4141	0.293	1.412	0.165	-0.177	1.005
Por quais motivos você decidiu entrar no curso? _Conveniência de horário	-0.0314	0.227	-0.138	0.891	-0.489	0.426
Por quais motivos você decidiu entrar no curso? _Desenvolvimento pessoal/profissional	0.0358	0.174	0.206	0.838	-0.315	0.387
Por quais motivos você decidiu entrar no curso? _Dinheiro	0.1107	0.217	0.510	0.613	-0.327	0.548
Por quais motivos você decidiu entrar no curso? _Dinheiro	0.2553	0.480	0.531	0.598	-0.713	1.224
Por quais motivos você decidiu entrar no curso? _Outros	-0.4552	0.274	-1.663	0.103	-1.007	0.097
Por quais motivos você decidiu entrar no curso? _Por gostar do curso/Afinidade	0.0616	0.149	0.414	0.681	-0.238	0.361
Por quais motivos você decidiu entrar no curso? _novos conhecimentos	0.1971	0.560	0.352	0.726	-0.931	1.325
Como você se descreveria em termos de programação? _Não tem experiência em programação	0.2295	0.183	1.251	0.218	-0.140	0.599
Como você se descreveria em termos de programação? _Não tem muita experiência em programação	0.0997	0.152	0.657	0.514	-0.206	0.405
Como você se descreveria em termos de programação? _Tem muita experiência em programação	-0.1198	0.271	-0.442	0.661	-0.666	0.427
Como você se descreveria em termos de programação? _Tem um pouco de experiência em programação	-0.0356	0.171	-0.208	0.836	-0.380	0.309
Fez o ensino médio em escola pública ou particular? _Particular	0.0302	0.236	0.128	0.899	-0.445	0.505
Fez o ensino médio em escola pública ou particular? _Público	0.1436	0.211	0.682	0.499	-0.281	0.568
Quanto tempo demora para chegar na Universidade? _Entre 1 hora e 2 horas	-0.0022	0.174	-0.013	0.990	-0.352	0.348
Quanto tempo demora para chegar na Universidade? _Entre 30 e 1 hora	0.0706	0.205	0.345	0.732	-0.342	0.483
Quanto tempo demora para chegar na Universidade? _Mais de 2 horas	-0.0394	0.348	-0.113	0.910	-0.740	0.661
Quanto tempo demora para chegar na Universidade? _Menos de 30 minutos	0.1449	0.225	0.644	0.523	-0.309	0.599

## Conclusões:

**Em resumo, os resultados sugerem que o modelo não é muito bom em prever se os alunos estão cursando APC pela primeira vez com base nas variáveis independentes fornecidas. Além disso, a alta probabilidade de multicolinearidade e a possível violação das suposições de normalidade dos resíduos podem afetar a confiabilidade dos resultados.**