



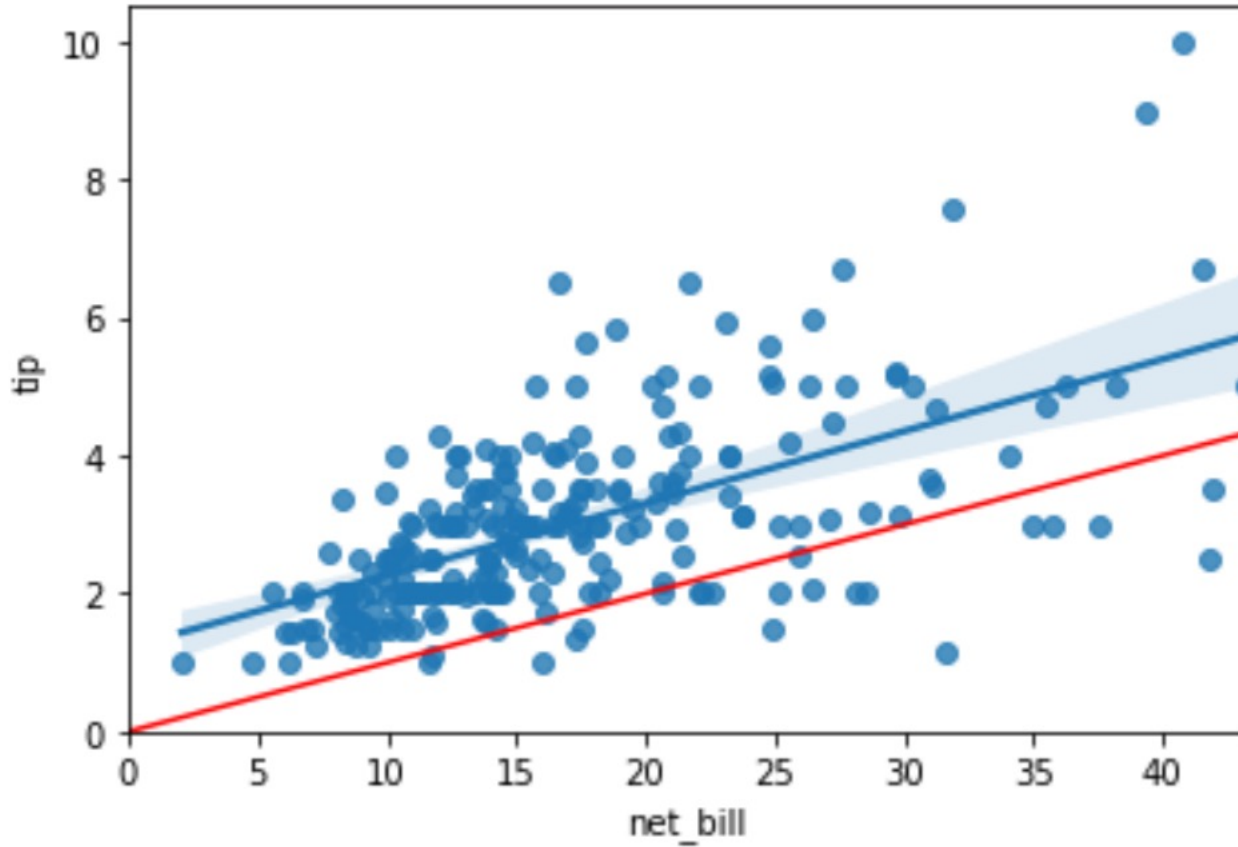
escola
britânica de
artes criativas
& tecnologia

Profissão: Cientista de Dados

Regressão Logística I

Introdução

Introdução



$Y = \text{tip_pct}$ (% de gorjeta)
 $X = \text{net_bill}$ (valor da conta)

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

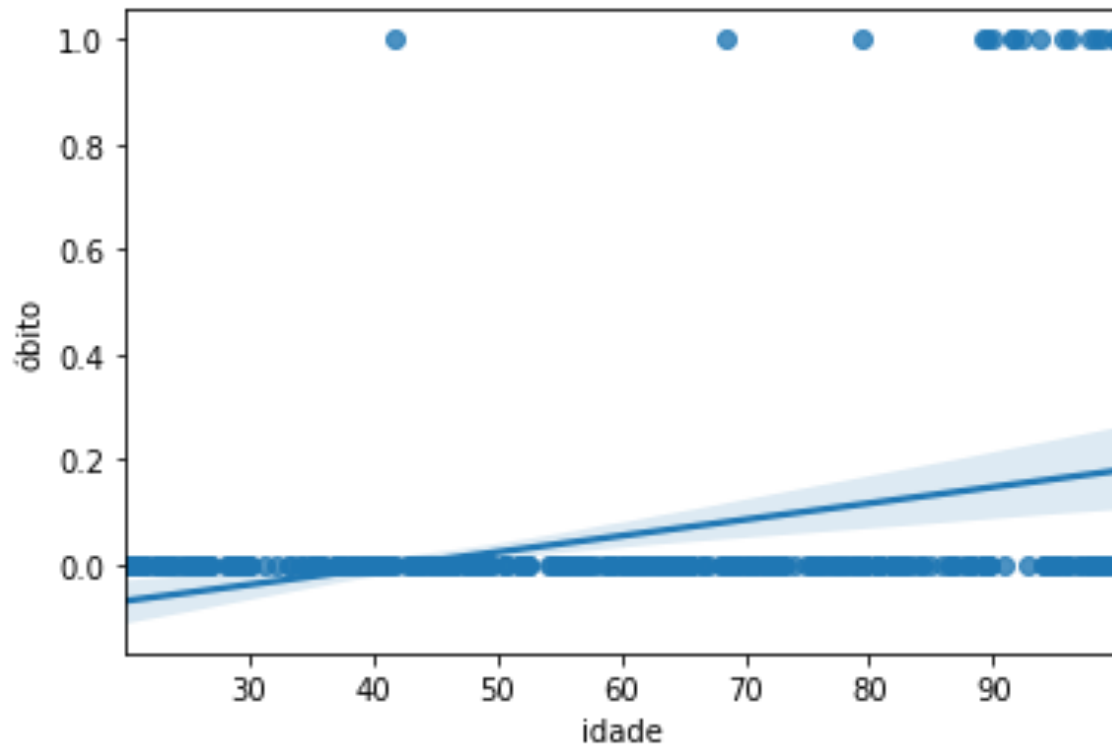
Com $i = 1, \dots, N$
 $\varepsilon_i \sim N(0, \sigma^2)$ i.i.d.

α e β são constantes desconhecidas.

Usando estimativas pontuais temos:

$$y = \hat{\alpha} + \hat{\beta}x + \hat{\varepsilon}$$

Só que a nossa resposta é binária...

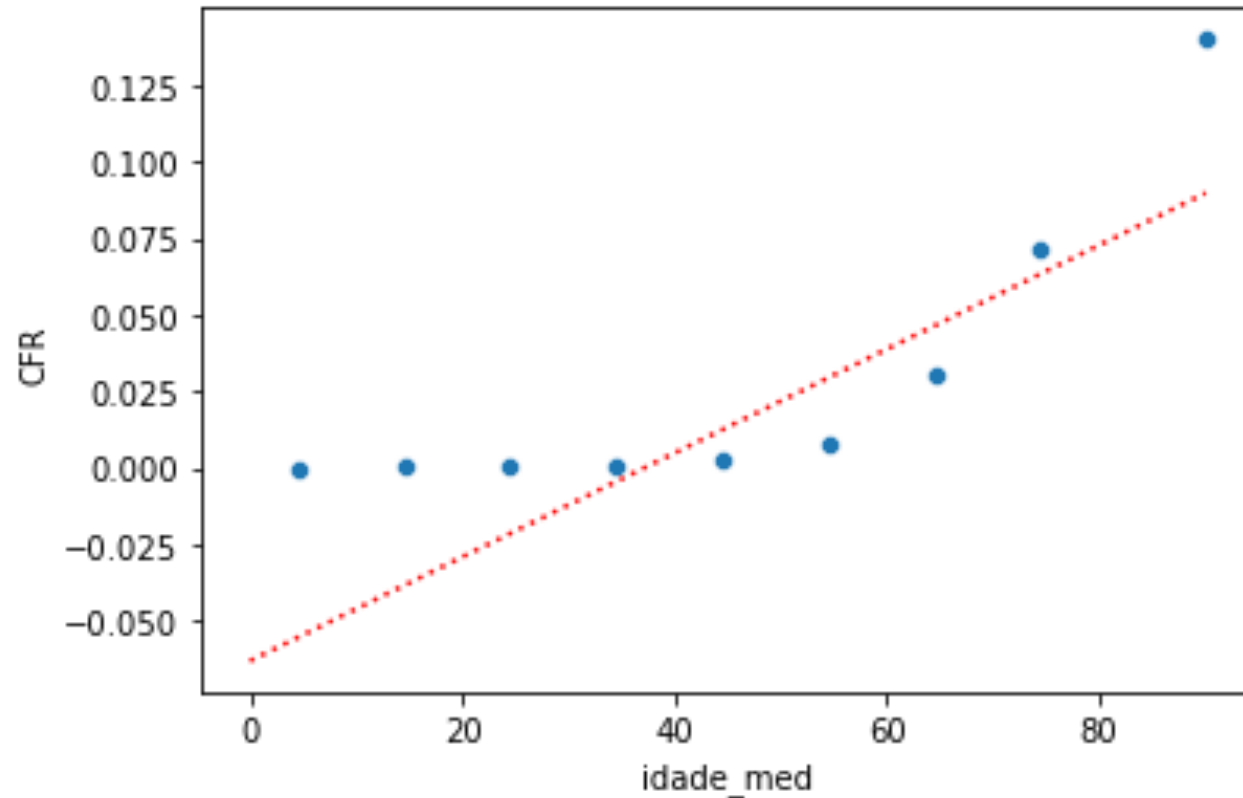


Queremos um modelo com essa cara.

Mas está sujeito a problemas.

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

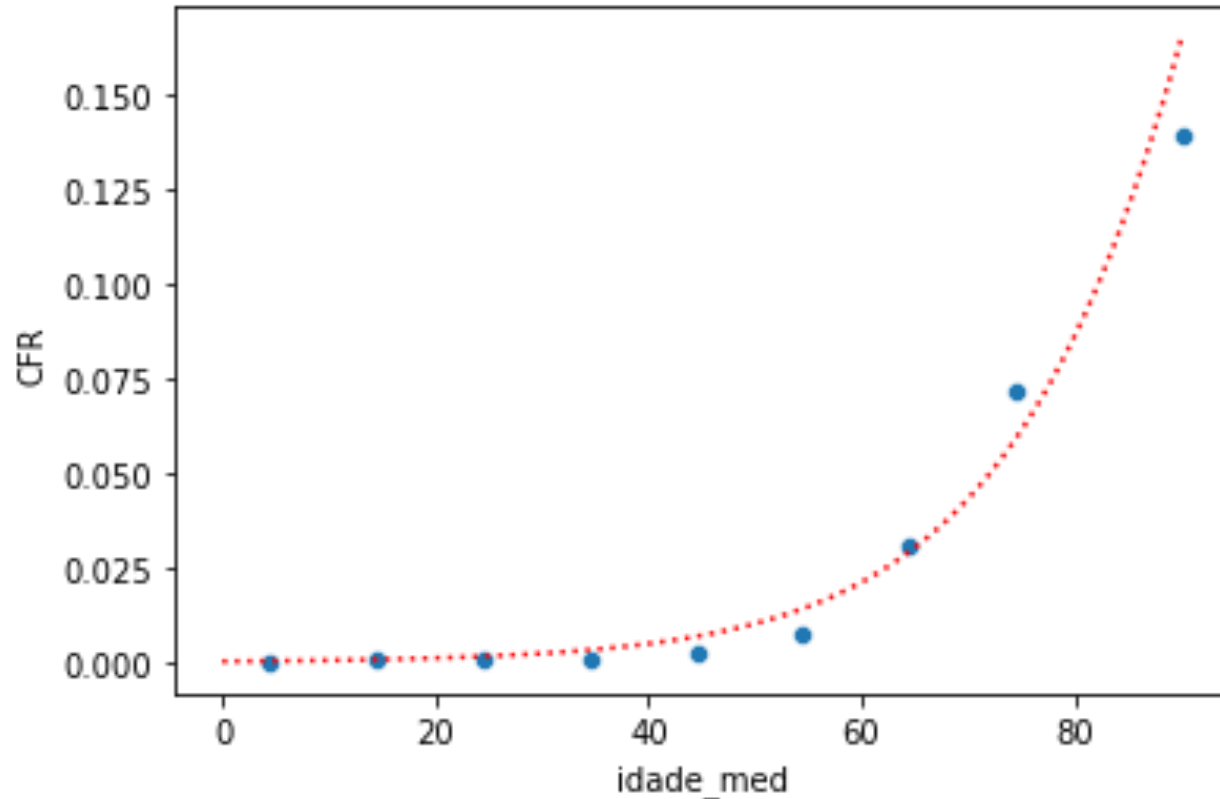
Agrupando os dados



Faixa de idade	Casos	Fatais	CFR (%)
0-9	459	0	0,00%
10-19	634	1	0,16%
20-29	3915	7	0,18%
30-39	8070	18	0,22%
40-49	9462	39	0,41%
50-59	11461	133	1,16%
60-69	10054	346	3,44%
70-79	5703	426	7,47%
80-100	2940	410	13,95%

Fonte do dado: <https://arxiv.org/abs/2005.07180>

Queremos uma preditiva mais assim



Vamos precisar de alguma alteração nesse modelo.

$$g(y_i) = \alpha + \beta x_i + \varepsilon_i$$

Fonte do dado: <https://arxiv.org/abs/2005.07180>