

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. Some nodes are highlighted with blue circles, and others with blue dots.

# Laboratório 9

A decorative network diagram in the bottom-right corner, featuring a complex web of interconnected nodes and lines. Some nodes are highlighted with blue circles, and others with blue dots.

# Universidade Federal da Bahia

## Sumário

Introdução;  
Análise Exploratória;  
Resultados;  
Considerações Finais.

## Edleide Brito

Gabriel Ferreira, Marla Lorrani e Moisés Augusto.

**Trabalho apresentado ao departamento de estatística da Universidade Federal da Bahia, como parte das avaliações da disciplina MAT229, análise de regressão, do semestre 2021.2.**

## Introdução

- Relação entre variáveis e o preço de planos de saúde dos EUA;
- Base de dados original possuía 1338 clientes; como as análises foram feitas para maiores de 20 anos, o número de observações foi reduzido para 1201;
- Criação de variável *Estado Nutricional*:

Estado Nutricional	IMC	Idade
Magreza	$\leq 18,5$	20 a 59 anos
	$< 22,0$	60 anos ou mais
Eutrofia	18,6 a 24,9	20 a 59 anos
	22,0 a 27,0	60 anos ou mais
Sobrepeso / Obesidade	$\geq 25$	20 a 59 anos
	$\geq 27,1$	60 anos ou mais

**Figura 1 – Classificação - Estado Nutricional por idade e IMC**

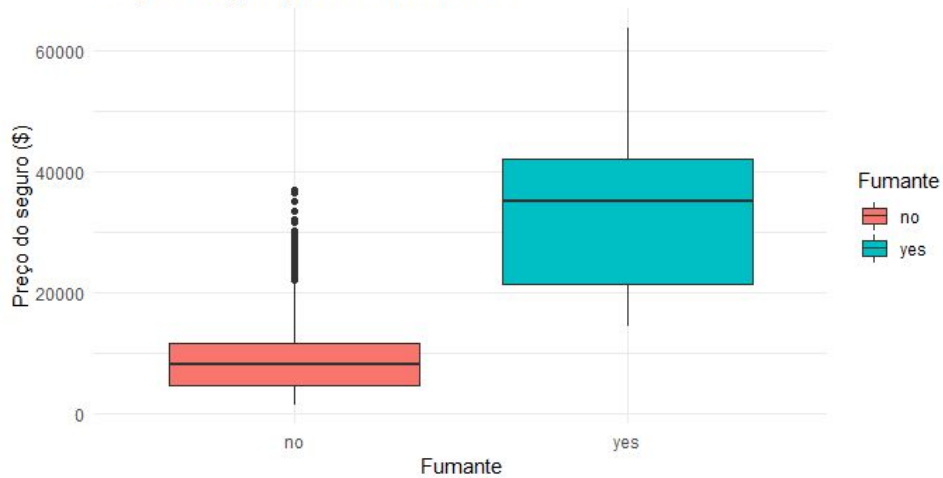
## Análise Exploratória

- 20,3% eram fumantes;
- 49,6% mulheres;
- 24,3% do nordeste, 27,0% do sudeste, 24,2% do noroeste;
- 39,4% sem filhos, 25,3% com um filho, 19,1% dois, 12,8% três e 2,0% quatro.
- 17,0% apresentaram bons índices de IMC e 1,7% estavam “magros”.

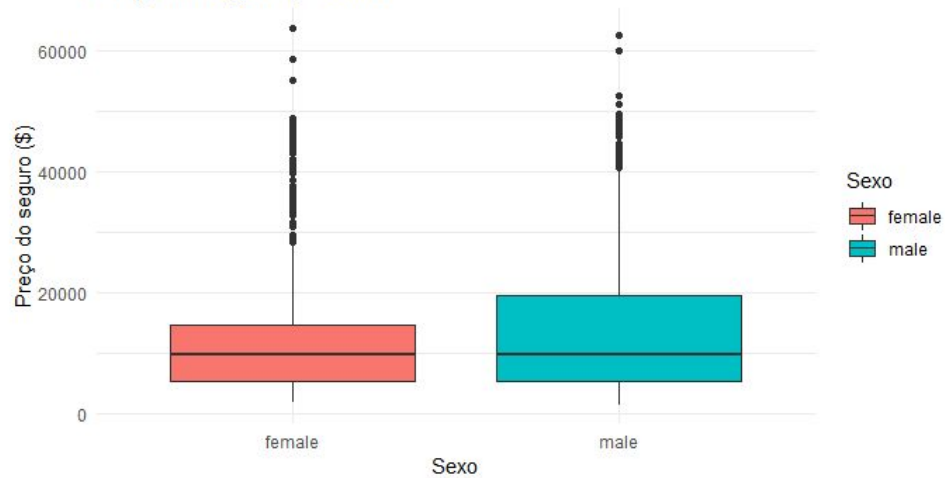
Tabela 1 – Estatísticas descritivas dos dados

Descritiva	Idade	IMC	Preço do Seguro (\$)
Mínimo	20,00	16,82	1392,00
1º Quartil	30,00	26,40	5439,00
Mediana	42,00	30,50	9801,00
Média	41,57	30,74	13825,00
3º Quartil	52,00	34,77	17180,00
Máximo	64,00	52,58	63770,00

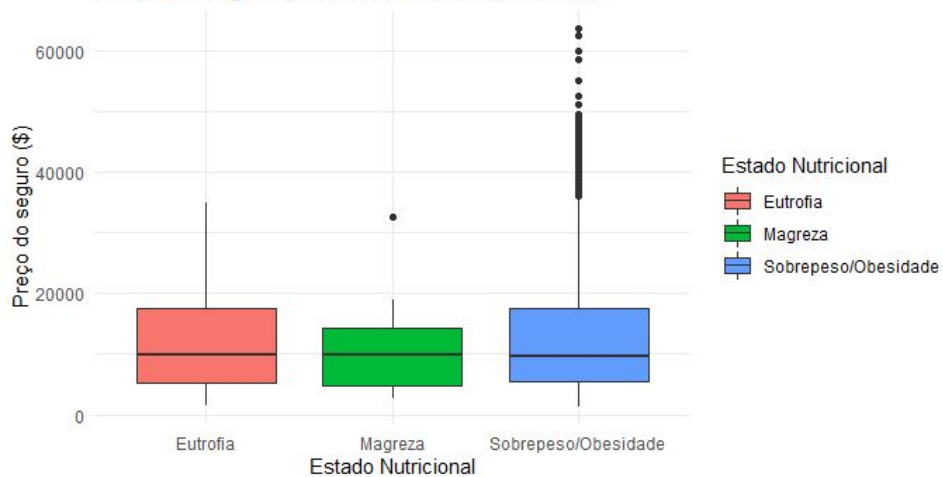
Preço do seguro por hábito de fumo



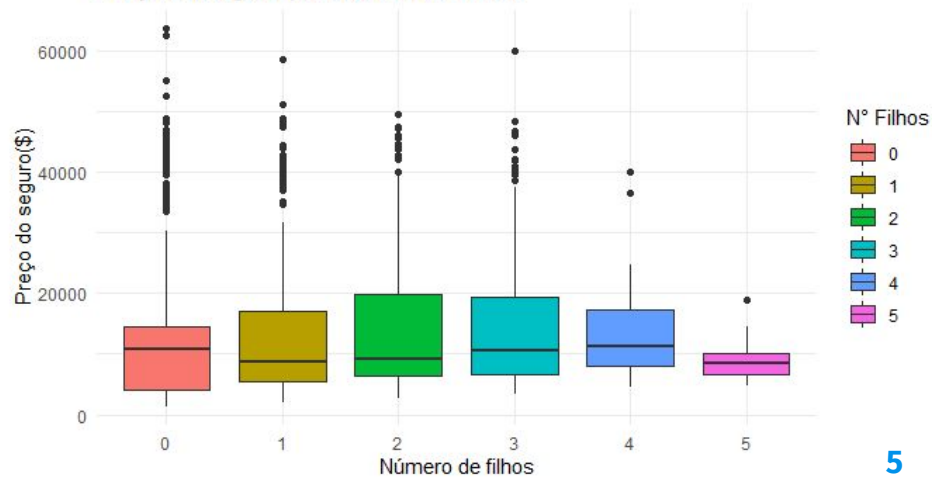
Preço do seguro por sexo



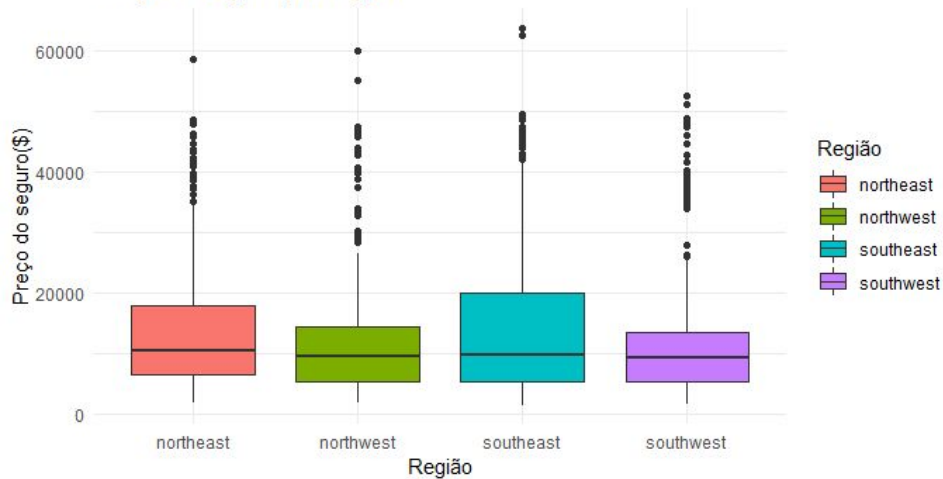
Preço do seguro por estado nutricional (IMC)



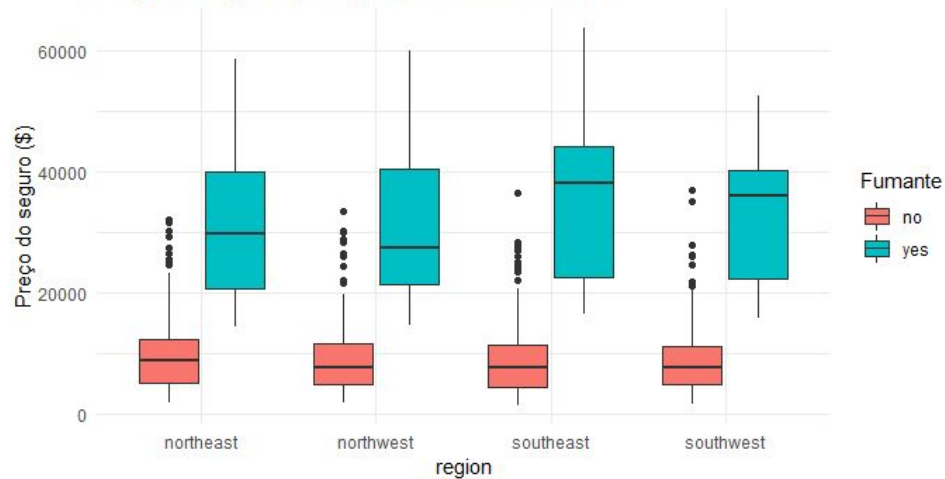
Preço do seguro por número de filhos



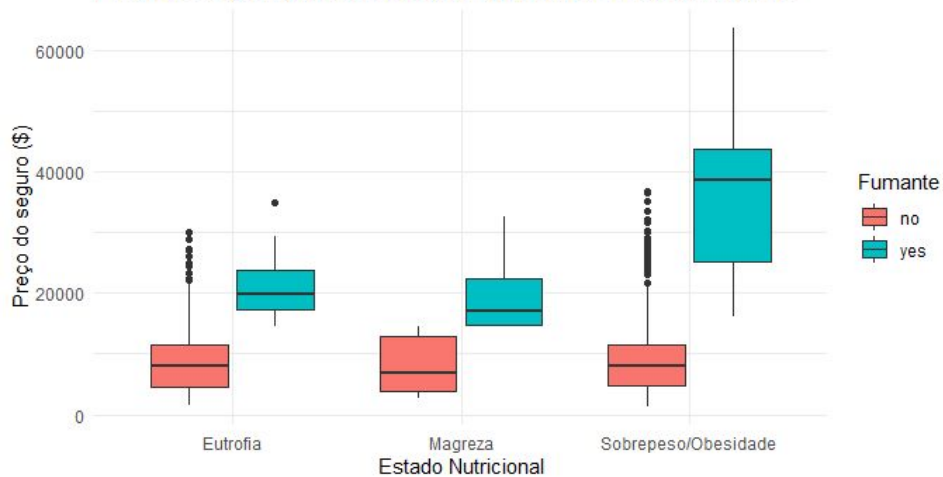
Preço do seguro por região



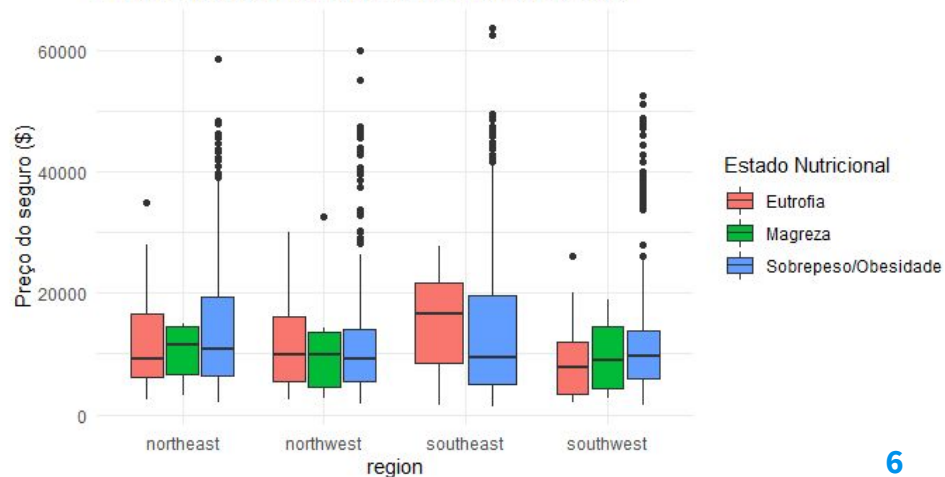
Preço do seguro por região e hábito de fumo



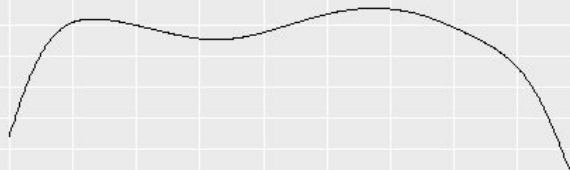
Preço do seguro por estado nutricional (IMC) e hábito de fumo



Preço do seguro por região e estado nutricional

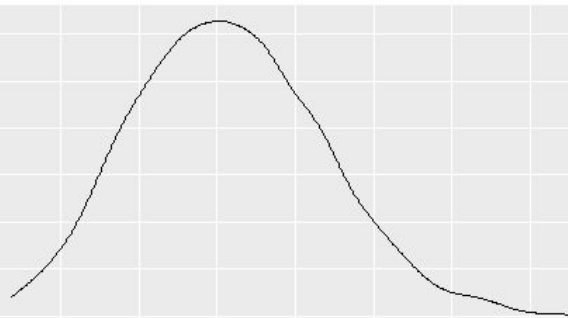


age



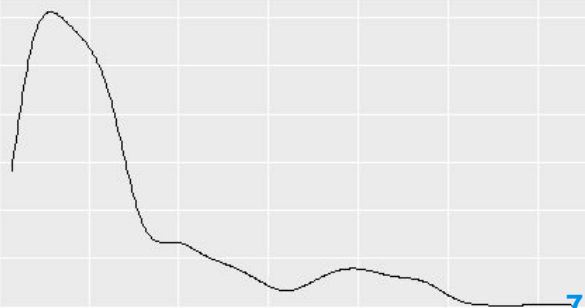
bmi

Corr:  
0.112\*\*\*



charges

Corr:  
0.282\*\*\*



age

bmi

charges

## Resultados - Modelo 1

$$\begin{aligned}\hat{Y} = & -5632,19 + 24040,59 \text{fumante} + 281,59 \text{idade} - 63,55 \text{masculino} + 445,98 \text{filhos1} \\ & + 1794,55 \text{filhos2} + 1081,17 \text{filhos3} + 2992,84 \text{filhos4} + 1353,29 \text{filhos5} \\ & - 669,45 \text{noroeste} - 273,68 \text{sudeste} - 885,09 \text{sudoeste} \\ & - 152,16 \text{magreza} + 3300,87 \text{sobrepeso}\end{aligned}\quad (1)$$

### Interpretação

Não fumar, não ter filhos, ser mulher, ser do Nordeste e eutrófico foram utilizados como referência.

### Ao nível de 95%

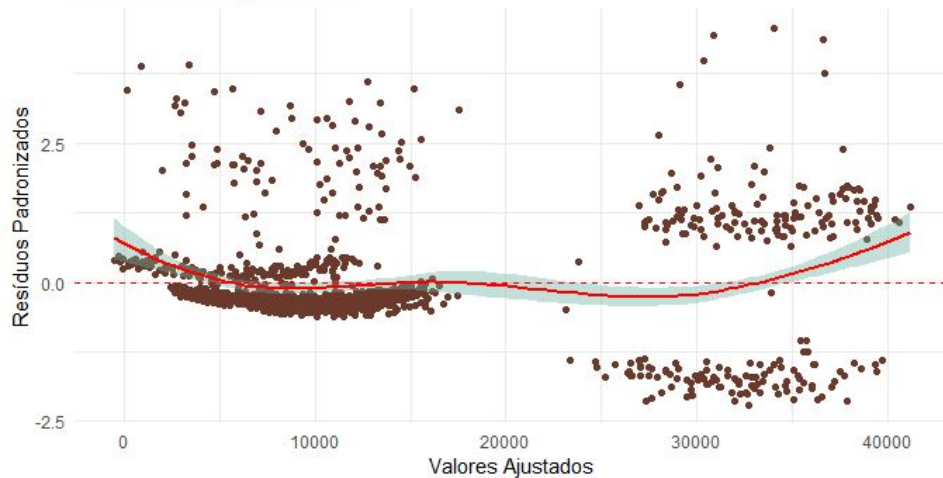
Ser fumante, ter dois ou quatro filhos, ou possuir sobrepeso/obesidade, foram considerados significantes para o modelo.

### Pressupostos

Resíduos não normais; Goldfeld-Quandt apontando para homocedasticidade e Breush-Pagan para heterocedasticidade; resíduos independentes.



Resíduos Vs. Ajustados



Escala - Locação

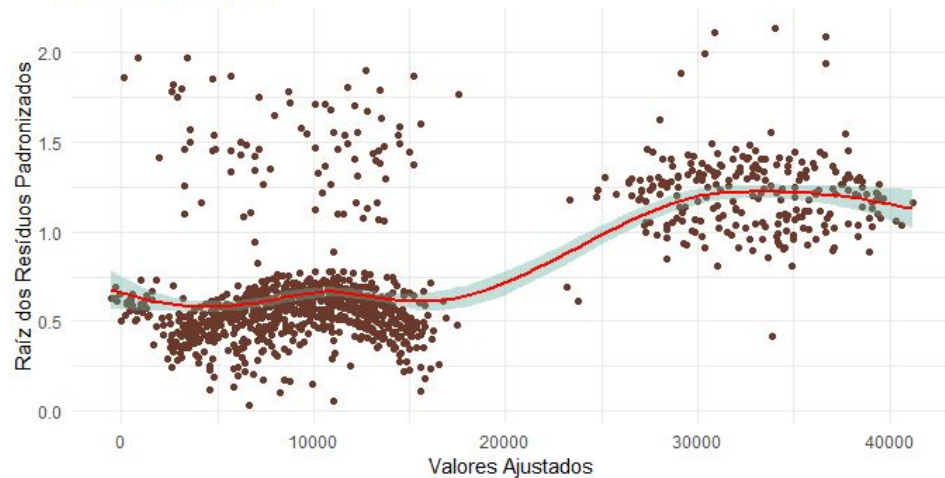
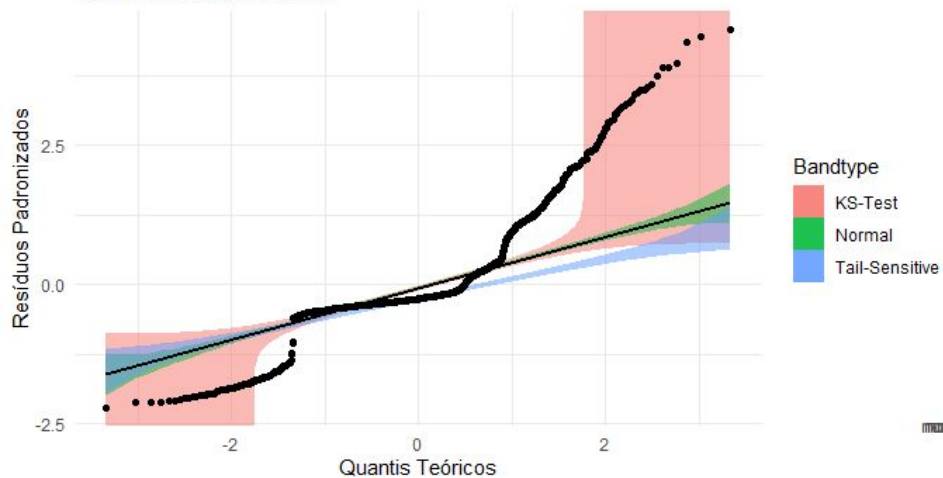
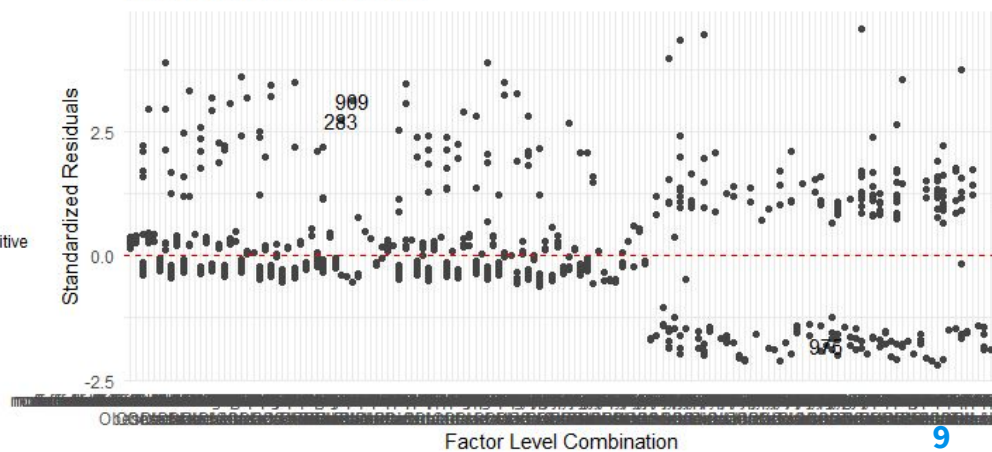


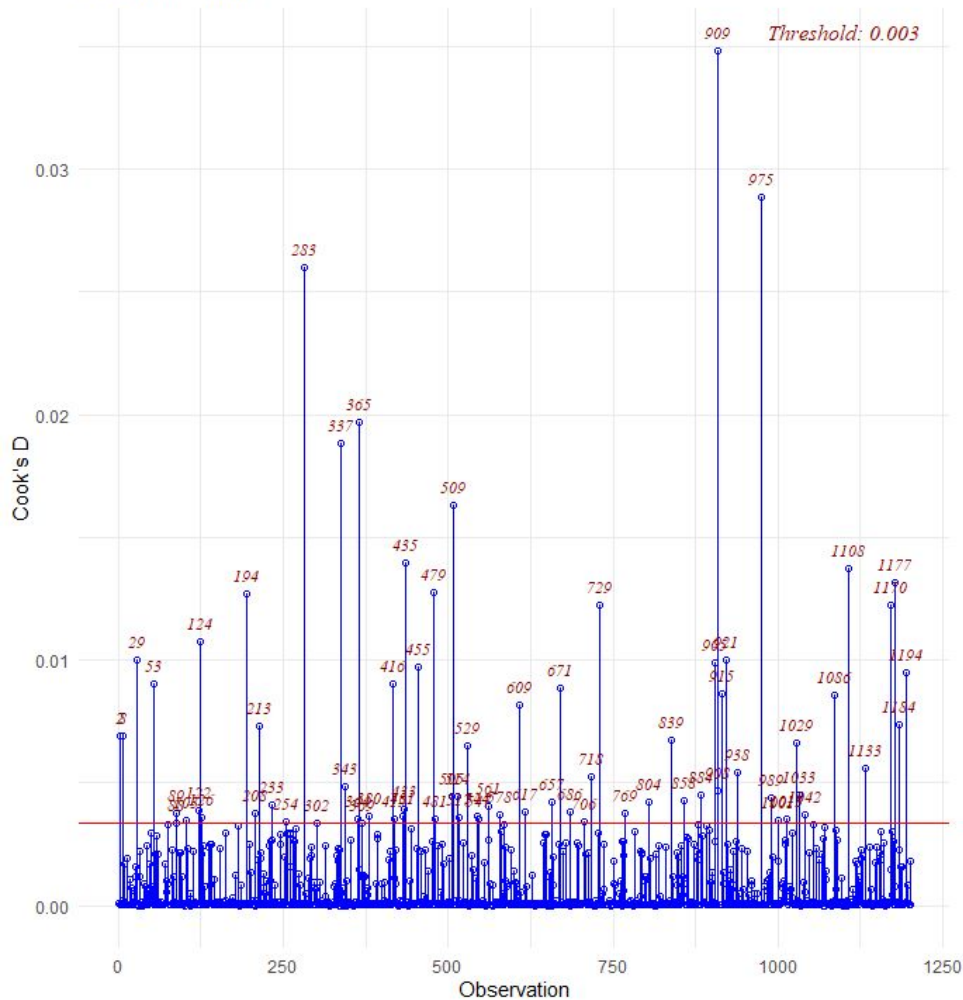
Gráfico Quantil Quantil



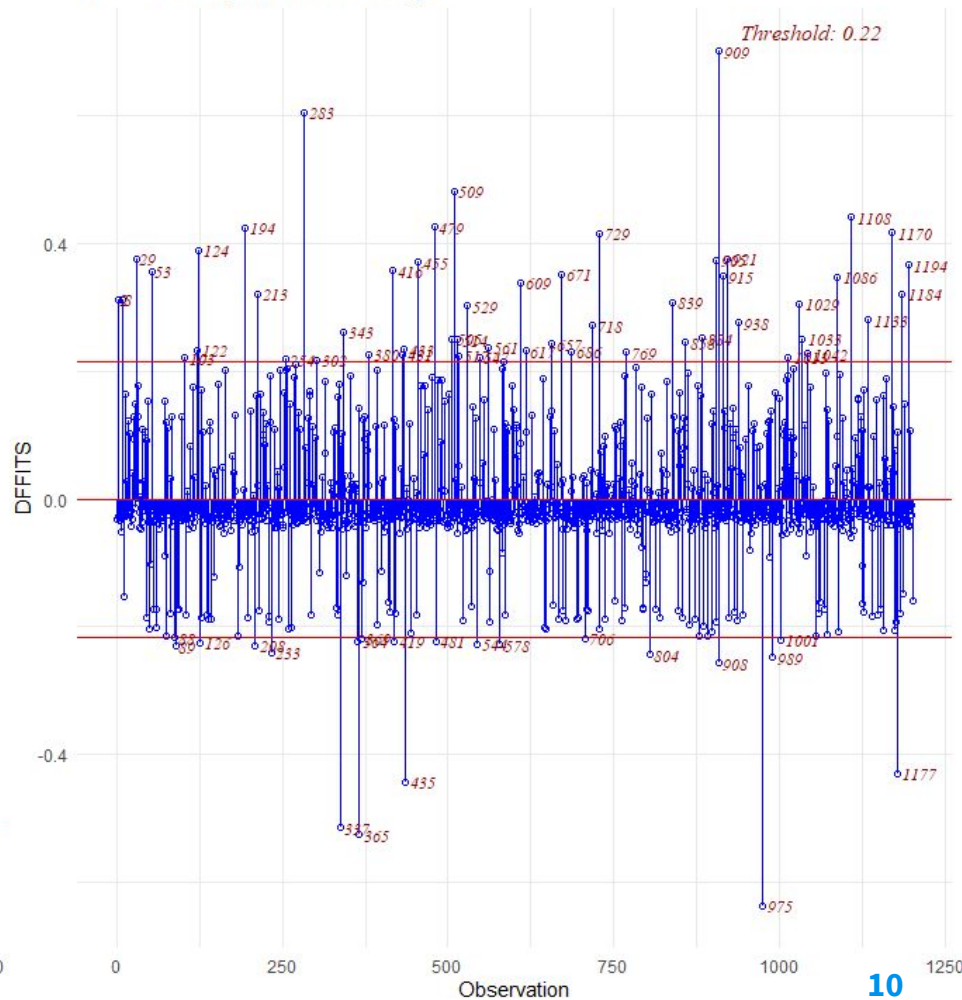
Constant Leverage:  
Residuals vs Factor Levels



Cook's D Chart



Influence Diagnostics for charges



## Resultados - Modelo 2

$$\begin{aligned}\hat{Y} = & -5536,33 + 24040,81 \textit{fumante} + 279,22 \textit{idade} - 258,50 \textit{masculino} + 446,03 \textit{filhos1} \\ & + 1797,58 \textit{filhos2} + 1083,44 \textit{filhos3} + 3002,47 \textit{filhos4} + 1362,47 \textit{filhos5} \\ & - 669,32 \textit{norooeste} - 274,48 \textit{sudeste} - 885,81 \textit{sudoeste} \\ & - 155,88 \textit{magreza} + 3303,52 \textit{sobrepeso} + 4,68 \textit{interação_idade_sexo}\end{aligned}\quad (2)$$

### Interpretação

Nem sexo nem interação entre idade e sexo foram significantes;

### Ao nível de 95%

A significância das demais variáveis se manteve como no Modelo 1;

### Pressupostos

Os resultados dos pressupostos não foram melhores que no Modelo 1.

A decorative graphic at the top of the slide featuring a network of interconnected nodes and lines, with a central blue double quote icon inside a dashed circle.

“

*Considerando o mencionado, um novo ajuste desconsiderando algumas das co-variáveis consideradas não significativas poderia melhorar algumas das análises dos pressupostos dos modelos;*

Obrigado!

Perguntas?