

# Estudo de diagnóstico de nódulos mamários



Ananda Bordignon, Brendha Lima, Giovanna Lazzarin.

# Sumário

1

Câncer de mama – Conhecendo o tema

2

Base de dados – Características dos dados

3

Análise exploratória – Características da população

4

Modelo – Ajuste de um Modelo Linear Generalizado

5

Conclusões – Interpretações

# Câncer de mama



## O que é Câncer de Mama?

Ocorre devido à perda da capacidade das células de limitar o seu próprio crescimento, multiplicando-se rapidamente e fora de controle.

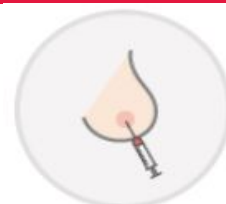
## Diagnóstico



Mamografia/  
Ultrassonografia



Ressonância nuclear  
magnética



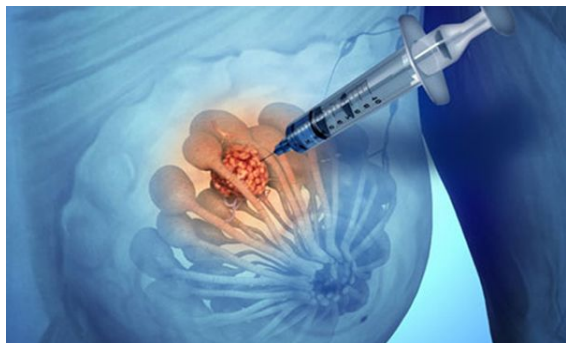
Biópsia

# Base de dados

- “Distribution of Freshwater Mussels: Coastal Rivers as Biogeographic Islands”, revista Systematic Zoology, vol. 23, #2, pp. 165–188, J.J. Sepkoski, Jr., M.A. Rex (1974).
- Dados: <http://www.stat.ufl.edu/~winner/data/mussels1.dat>
- informações: <http://www.stat.ufl.edu/~winner/data/mussels1.txt>

# Base de dados

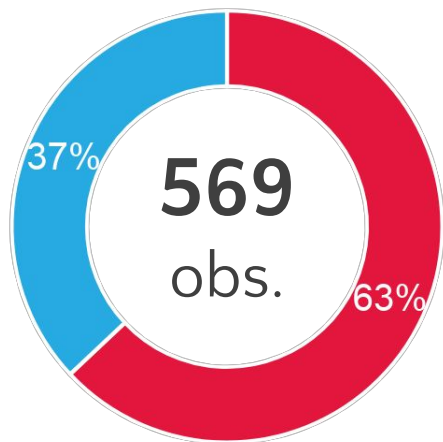
- 569 observações.
- 10 Variáveis: raio médio, textura média, perímetro médio, área médio, (regularidade) suavidade média, compacidade média, concavidade média, média de pontos côncavos , simetria média e dimensão fractal média.



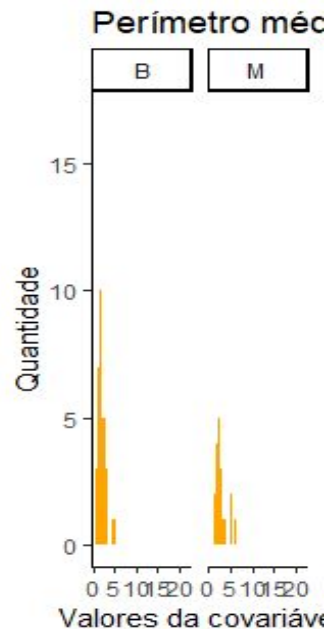
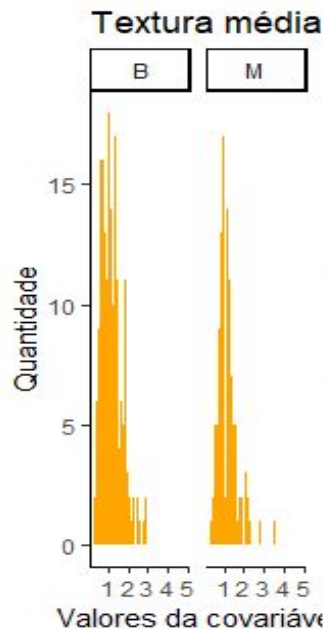
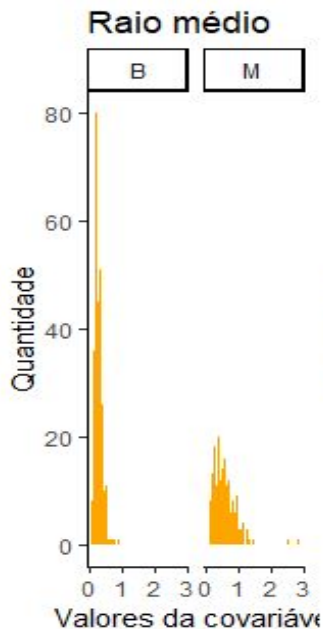
Regressão linear generalizada com família Binomial.



# Análise exploratória



■ Benigno ■ Maligno



# Modelo

Função de ligação

Regressão Logística

Métodos

Ajuste	AIC	Verossim.
logito	287,92	-132,96
probito	286,74	-132,37
cloglog	295,09	-136,54
cauchy	300,06	-139,03

$$\ln \left( \frac{\pi_i}{1 - \pi_i} \right) = \beta_0 + \beta_1 x_i.$$

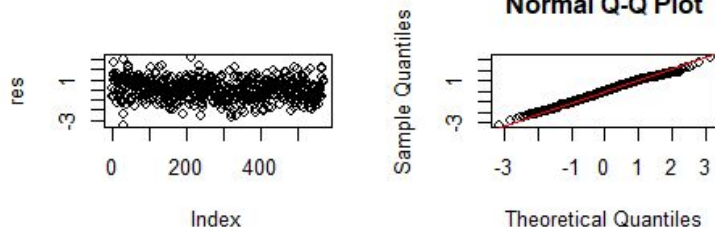
- Forward
- Backward
- Stepwise

# Modelo

## Ajuste 1

$$\ln\left(\frac{\pi_i}{1 - \pi_i}\right) = -3,545 +$$

0,4039 . Área média +  
- 28,73 . Raio médio +  
74,4 . Compacidade Média +  
-356 . Dimensão fractal



## Ponto de alavancagem

Observação 291 é ponto influente.

	Model 1	Model 2
## (Intercept)	-3.545	-4.058
## SE	0.476	0.523
##		
## mean_area	0.4039	0.4227
## SE	0.0501	0.0521
##		
## mean_radius	-28.73	-29.65
## SE	4.63	4.72
##		
## mean_compactness	74.4	56.5
## SE	20.4	20.5
##		
## mean_fractal_dimension	-356	-142
## SE	147	152



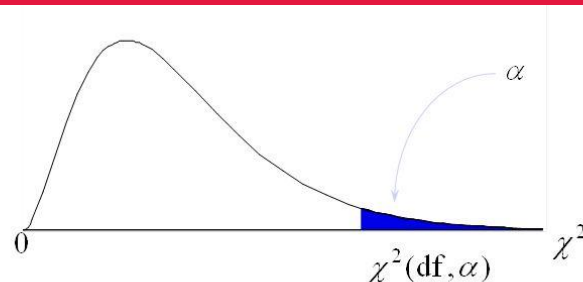
# Modelo

## Ajuste 2

$$\ln\left(\frac{\pi_i}{1 - \pi_i}\right) = -4,136 +$$

0,4340 . Área média +  
- 30,661 . Raio médio +  
30,807 . Compacidade Média

## Qualidade de ajuste



Valores de Qui-Quadrado com 95% e 565 df

- Tabelado: 619,31
- Obtidos: 261,1

**Não rejeita** a hipótese do modelo estar bem ajustado

# Conclusões

- Variáveis Área Média e Compacidade Média têm relação positiva com o aumento da probabilidade de ter câncer de mama.
- Variável Raio Médio possui relação negativa com o aumento de probabilidade.

	Coefficientes	Valores
Intercepto	-4,136	
Área	0,434	15
Raio	-30,661	0,11
Compacidade	39,807	0,02
-0,20257		
Probabilidade estimada		45%

	Coefficientes	Valores
Intercepto	-4,136	
Área	0,434	17
Raio	-30,661	0,11
Compacidade	39,807	0,02
0,66543		
Probabilidade estimada		66%

Ao fixarmos as variáveis e **aumentarmos em 2 unidades** a área média, o impacto na probabilidade estimada de maus vai de **45%** para **66%**.