

# Caducidade em Idosos

Camille Menezes, Jeff Caponero e Michel Miler

# Sumário

Introdução . . . . .	2
Análise descritiva . . . . .	2
Modelo . . . . .	3
Resultados Inferenciais. . . . .	4
Estimativas pontuais . . . . .	4
Discussão dos resultados . . . . .	4
Função Desvio . . . . .	4
Gráficos . . . . .	5
Análise Residual. . . . .	5
Observações Atípicas . . . . .	6
Referências . . . . .	7

# Introdução

Cinquenta e quatro indivíduos considerados idosos são submetidos a um exame psiquiátrico para avaliar a ocorrência ou não de sintoma de caduquice. (Agresti, 1990, pgs. 122-123). Acredita-se que o escore obtido num exame psicológico feito previamente esteja associado com a ocorrência ou não do sintoma. Este trabalho propõe um modelo de regressão logística para estudar esta relação.

## Análise descritiva

Conduziu-se uma análise descritiva dos dados, com objetivo de entender um pouco mais as variáveis consideradas.

Tabela 1: Estatísticas-resumo para a variável score dos idosos com ou sem caducância

	Min	Q1	Median	Mean	Q3	Max	Std.Dev	CV	Skewness	Kurtosis
<b>Com Caducância</b>	4	7	8,5	8,93	11	14	3,17	0,36	0,26	-1,21
<b>Sem Caducância</b>	4	10	13,0	12,50	15	20	3,46	0,28	-0,08	-0,24

Na Tabela 1, é possível observar que a média dos scores para os idosos com caducância é menor do que para os sem caducância. A média e a mediana para ambos cenários aparentam estar bem próximas entre si, além da métrica de simetria estar próxima de zero, indicando que a distribuição dos scores é simétrica para os idosos com ou sem caducância. Para a curtose, é notável que a distribuição dos scores para os idosos com caducância é mais platicúrtica que para os idosos sem caducância.

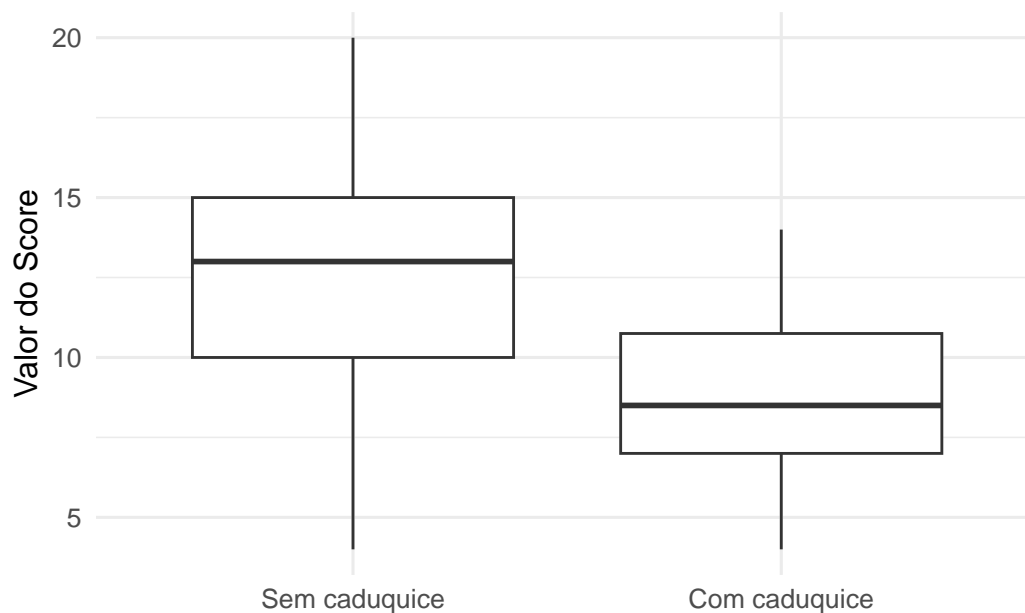


Figura 1: Boxplot para a variável score dos idosos com ou sem caducância

Através dos boxplots da Figura 1, é possível observar que para os idosos sem caducância, existe uma maior variabilidade dos scores abaixo da mediana, enquanto que para os scores dos com caducância há uma variabilidade maior entre a mediana e o terceiro quartil. Tanto pela Tabela 1 quanto pela Figura 1, já é possível notar que há uma tendência a qual menores scores estão mais associados com idosos com caducância.

## Modelo

Queremos analisar como o valor do score obtido no exame psicológico impacta na chance de idoso apresentar caducância ou não. Desse modo, o modelo a ser definido será o modelo MLG binomial com função de ligação logito, então sendo  $Y_i$  a variável que indica se o idoso “i” apresenta caducância ou não, temos que

- $Y_i \sim \text{Binomial}(1, \mu_i)$
- $\log\left(\frac{\mu_i}{1-\mu_i}\right) = \alpha + \beta x_i$

onde

- $x_i$  é a variável score
- $\frac{\mu_i}{1-\mu_i}$  é a chance
- $\alpha$  é o efeito escalar no logaritmo da chance do idoso apresentar caducância
- $\beta$  é efeito no logaritmo da chance do idoso apresentar caducância quando uma unidade é adicionada na variável score
- $\exp(\beta)$  é efeito na razão da chance do idoso apresentar caducância quando uma unidade é adicionada na variável score
- $\alpha$  é o efeito escalar no logaritmo da chance do idoso apresentar caducância ou não.

Com base nos dados é possível avaliar um modelo de regressão logístico.

Tabela 2: Resultados para o modelo com função de ligação logito.

	Estimativa	EP	Est. z	Pr(> z )
<b>Intercepto</b>	2,4040	1,1918	2,0171	0,0437
<b>Score</b>	-0,3235	0,1140	-2,8385	0,0045

Tabela 3: Resultados para o modelo com função de ligação probito.

	Estimativa	EP	Est. z	Pr(> z )
<b>Intercepto</b>	1,3862	0,6853	2,0228	0,0431
<b>Score</b>	-0,1880	0,0630	-2,9841	0,0028

Tabela 4: Resultados para o modelo com função de ligação cauchy.

	Estimativa	EP	Est. z	Pr(> z )
<b>Intercepto</b>	3,3266	1,8072	1,8407	0,0657
<b>Score</b>	-0,4212	0,1984	-2,1234	0,0337

Os modelos apresentados nas Tabelas 2, 3 e 4 são bastante similares. Uma comparação mais acurada dos modelos pode ser verificada a partir dos valores de AIC atingidos. A Tabela 5 mostra esses valores.

Tabela 5: Valores de AIC dos modelos para cada função de ligação.

Função	AIC
<b>Logito</b>	55.0174
<b>Probit</b>	54.9836
<b>Cauchy</b>	55.1573

Desta forma, verifica-se que a diferença entre os modelos é bastante sutil e não há necessidade de se valer de uma função de ligação diferente da canônica para explicar os dados. Assim a função de ligação logito é preferível dentre as demais.

## Resultados Inferenciais.

### Estimativas pontuais

Com base no modelo escolhido é possível estimar os valores de  $\beta$  e  $\phi$  para o modelo.

Os valores estimados foram de  $\alpha = -0.2625$  e de  $\beta_1 = 0.1159$ . Com 3 iterações obteve-se a precisão de  $10^{-6}$ .

### Discussão dos resultados

Interpretação dos resultados.

### Função Desvio

Interpretação da função desvio. Compare o valor da função desvio com a estatística qui-quadrado e interprete o resultado do teste de hipótese.

## Gráficos

Apresente o gráfico da função de distribuição acumulada logística (veja Aula prática III - dados turbine). Plote o gráfico de  $\hat{z}$  versus  $\hat{\eta}$  e comente sobre as evidências de adequacidade da função de ligação.

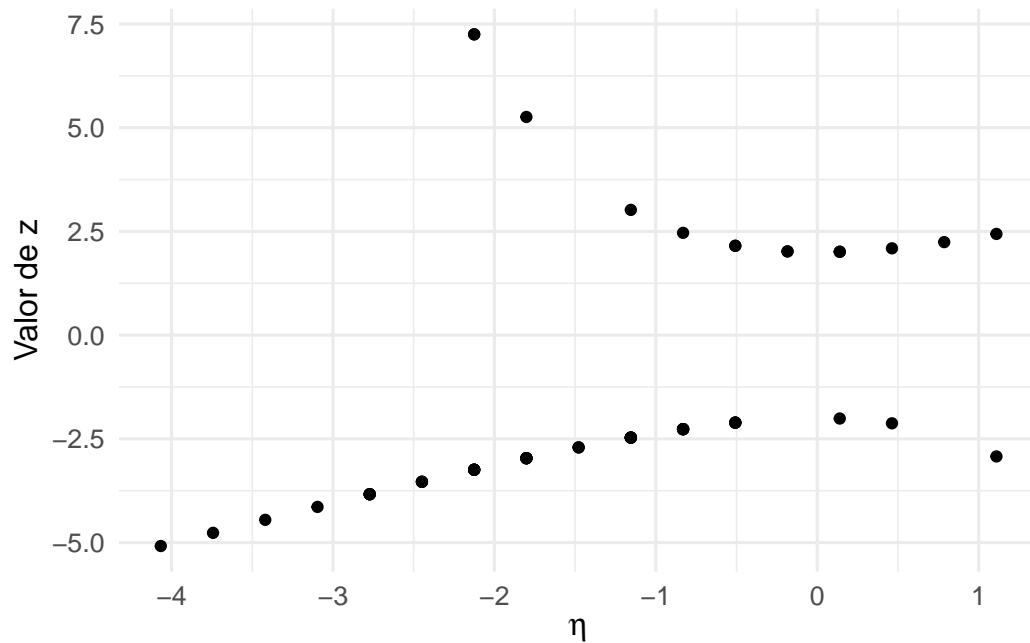


Figura 2: Função de distribuição acumulada logística

## Análise Residual.

Considerando os resíduos Studentizado (tsi) padronizado e o componente do desvio padronizado (tdi) apresente os seguintes gráficos tsi, tdi versus valores ajustados, tsi, tdi versus valores observados e os respectivos gráficos do envelope simulado.

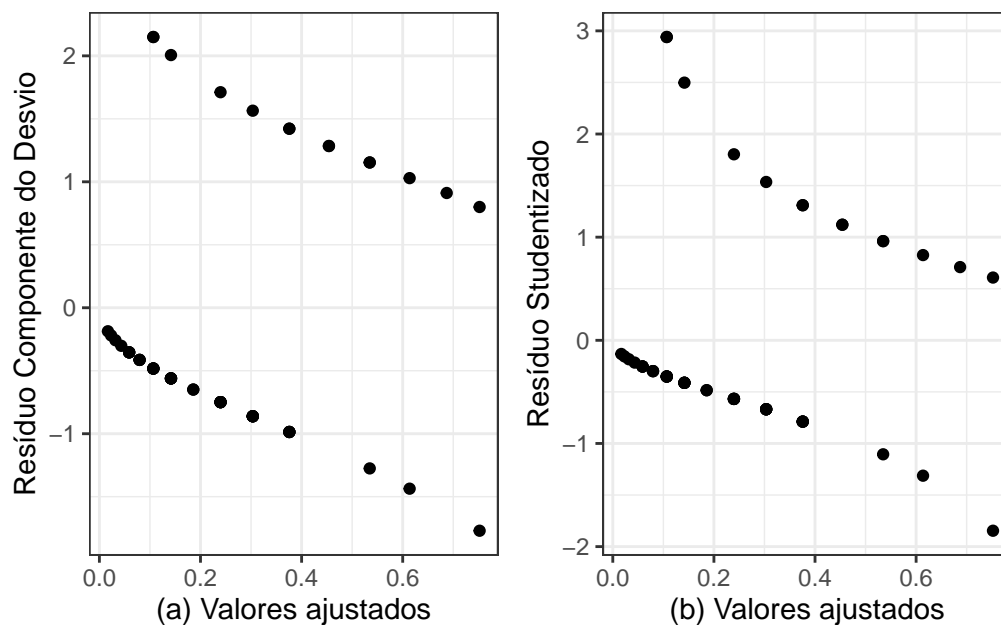


Figura 3: Análise residual. (a) Valores ajustados em função dos resíduos; (b) Valores ajustados em função dos resíduos studentizados.

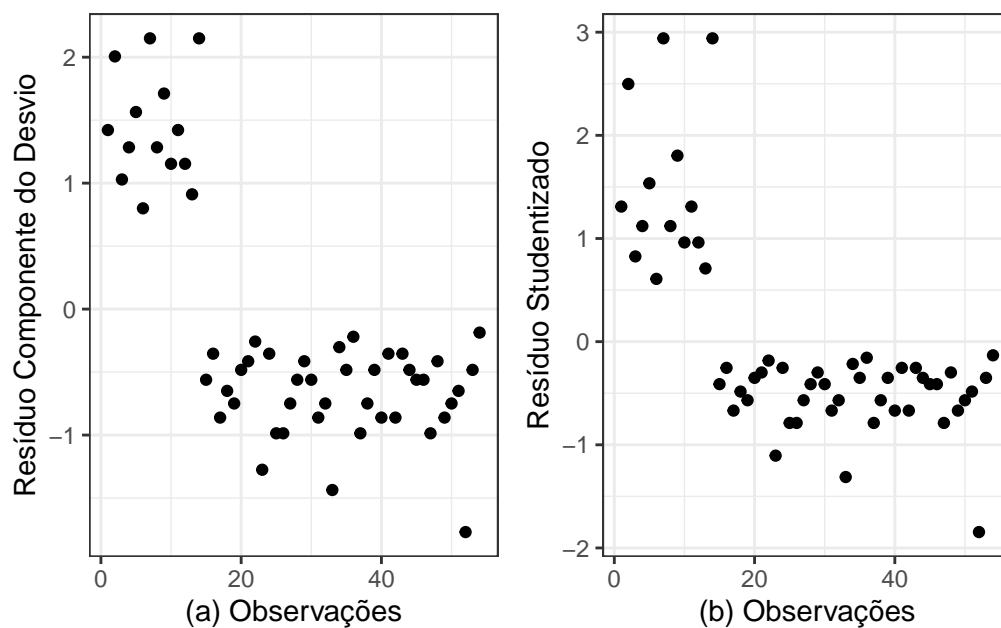


Figura 4: Análise residual. (a) Observações em função dos resíduos; (b) Observações em função dos resíduos studentizados.

## Observações Atípicas

Identifique as observações atípicas. Comente cada gráfico.

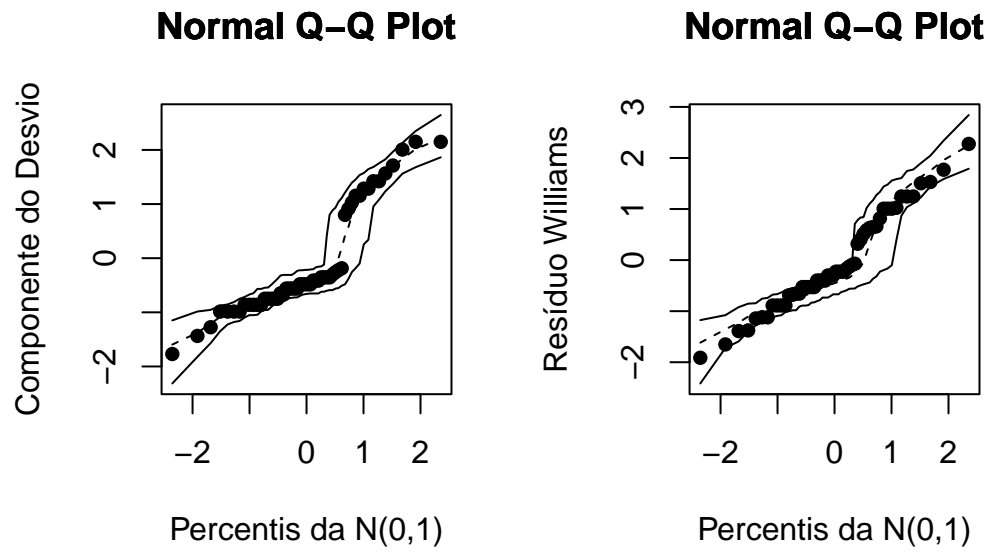


Figura 5: Análise de ajuste do modelo. (a) Componentes do desvio; (b) Resíduo de Williams.

## Referências

Agresti A. (1990). Categorical Data Analysis. John Wiley, New York.