

Caducidade em Idosos

Camille Menezes, Jeff Caponero e Michel Miler

Sumário

Introdução	2
Análise descritiva	2
Modelo de Regressão Logístico	4
Definição	4
Resultados Inferenciais	4
Estimativas pontuais	5
Discussão dos resultados	5
Função Desvio	6
Gráficos	6
Análise Residual	6
Observações Atípicas	8
Referências	8

Introdução

É importante a compreensão das mudanças psicológicas associadas ao envelhecimento. Por isso, este trabalho tem como objetivo analisar um grupo de cinquenta e quatro indivíduos considerados idosos, submetidos a um exame psiquiátrico para avaliar a presença ou ausência de sintomas de caduquice. Acredita-se que os escores obtidos em exames psicológicos prévios podem estar relacionados com a ocorrência desses sintomas. Assim, a presente pesquisa propõe a utilização de um modelo de regressão logística para investigar essa relação.

Os dados são provenientes de (Agresti, 1990, pgs. 122-123). Ao longo deste trabalho, será realizada uma análise descritiva com objetivo de entender um pouco mais as variáveis consideradas, a construção do modelo de regressão com diferentes funções de ligação, a apresentação de resultados inferenciais e estimativas pontuais. Além disso, também será apresentado a análise de resíduos, com foco na identificação de observações atípicas.

Análise descritiva

Na amostra de idosos incluídos no estudo, 74% não demonstraram sinais de caduquice. A Tabela 1 apresenta um resumo dos escores registrados por esses idosos.

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
Com Caducância	4	7	8,5	8,93	11	14	3,17	0,36	0,26	-1,21
Sem Caducância	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.

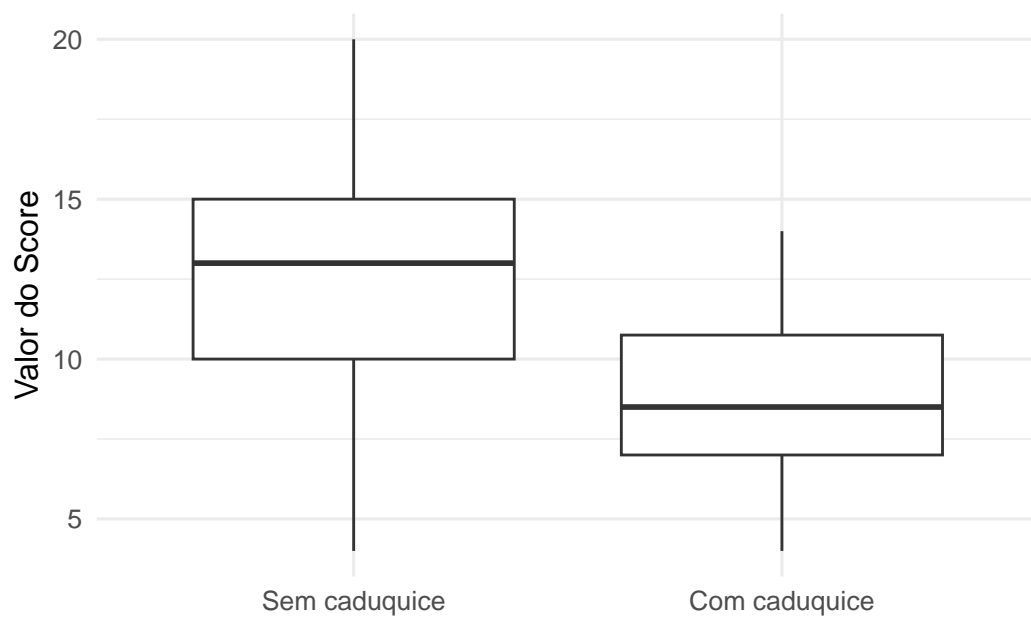


Figura 1: Boxplot para a variável score dos idosos com ou sem caducância

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 de Regressão Logístico

Definição

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
- α é o efeito escalar no logaritmo da chance do idoso apresentar caducância
- β 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
- α é o efeito escalar no logaritmo da chance do idoso apresentar caducância ou não.

Resultados Inferenciais

Na Tabela 2 abaixo, podemos notar as estimativas do intercepto (α) e do coeficiente da variável score (β) foram significativas ao nível de 5%. O valor do coeficiente do score foi igual a -0,3235, ou seja, adicionar uma unidade na variável score é responsável por uma diminuição de 0,3235 no logaritmo da chance do idoso apresentar caducância. Outra interpretação é que adicionar uma unidade na variável score é responsável por uma diminuição de $\exp(1)$

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

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

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

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

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

	Estimativa	EP	Est. z	$\Pr(> z)$
Intercepto	3,3266	1,8072	1,8407	0,0657
Score	-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
Logito	55.0174
Probit	54.9836
Cauchy	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.

Estimativas pontuais

Com base no modelo escolhido é possível estimar os valores de β e ϕ 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

Todos os coeficientes foram significativos ao nível de 5%. Como a variável score não assume valor zero, o coeficiente do intercepto não apresenta interpretabilidade. A variável score apresenta um coeficiente de -0.3235. Isso indica que, a chance do idoso apresentar caduquice com o aumento de uma unidade na variável score é $\exp(-0.3235) = 0.7236$.

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 adequidade da função de ligação.

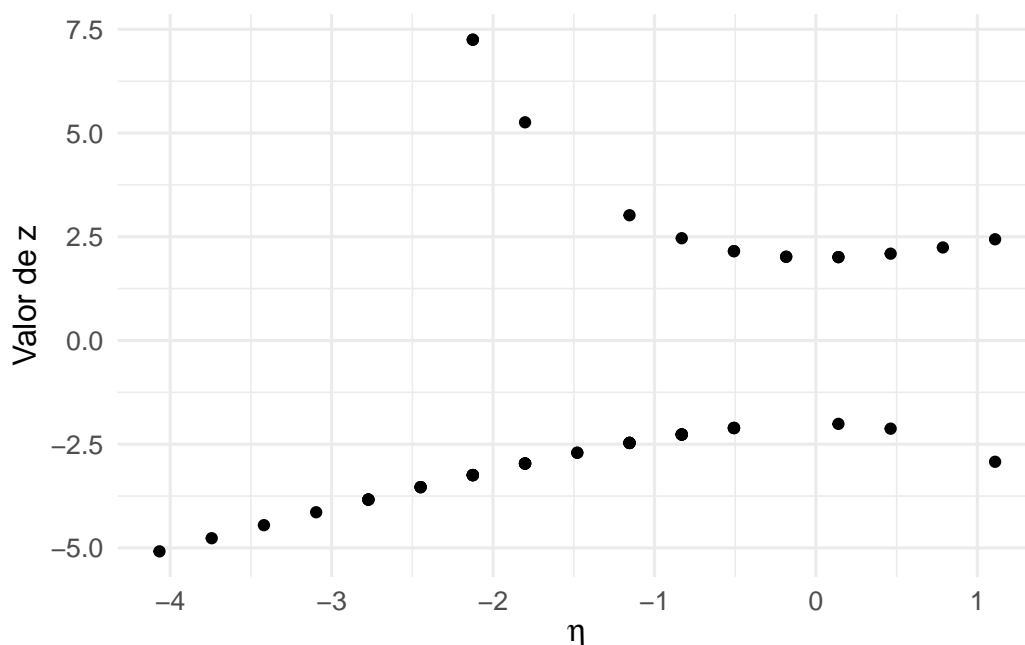


Figura 2: Função de distribuição acumulada logística

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.

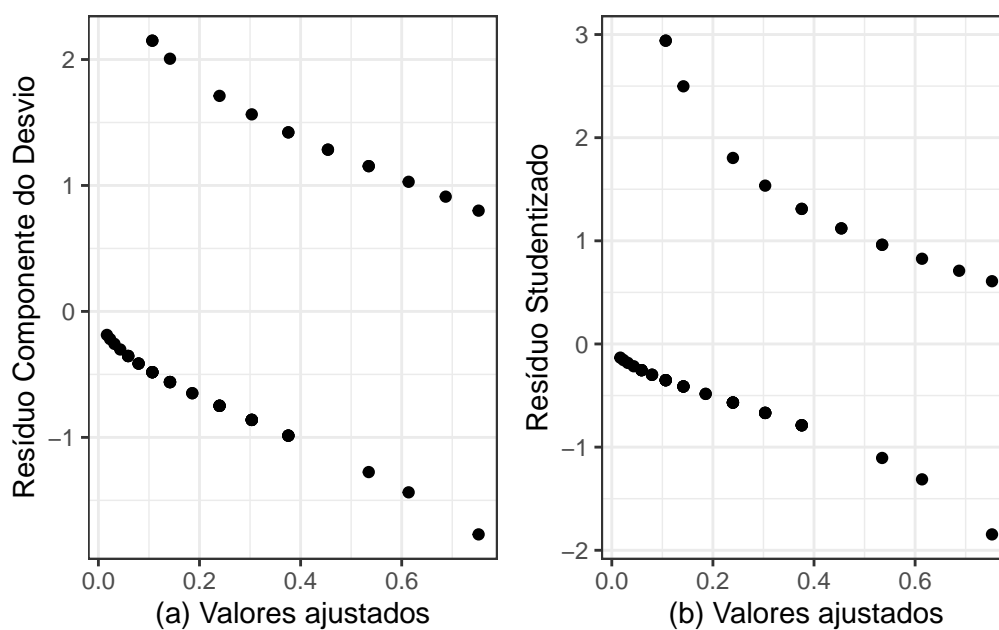


Figura 3: Análise residual. (a) Valores ajustados em função dos resíduos; (b) Valores ajustados em função dos resíduos studentizados.

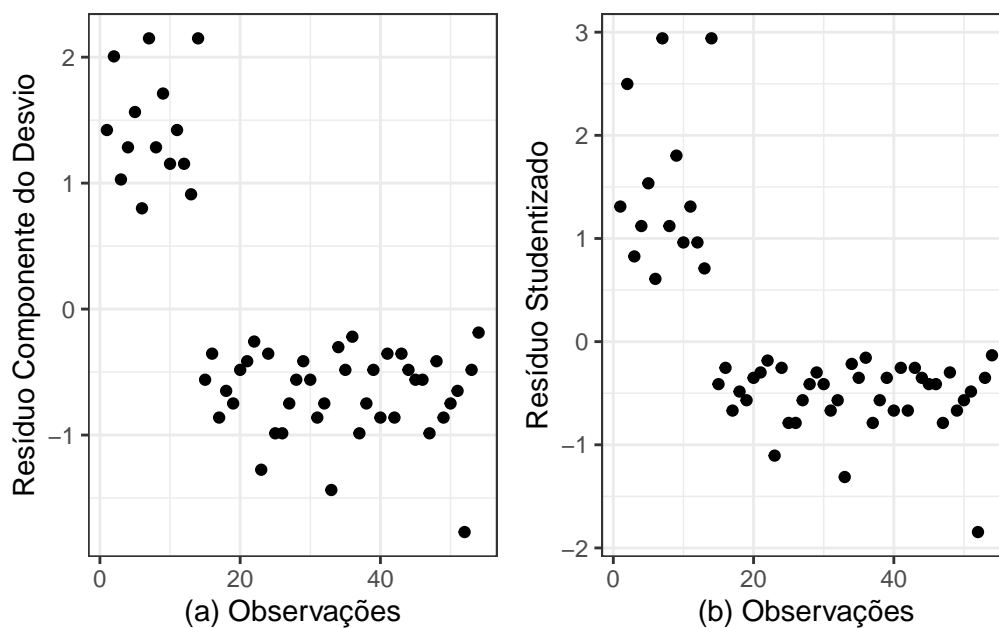


Figura 4: Análise residual. (a) Observações em função dos resíduos; (b) Observações em função dos resíduos studentizados.

Observações Atípicas

Identifique as observações atípicas. Comente cada gráfico.

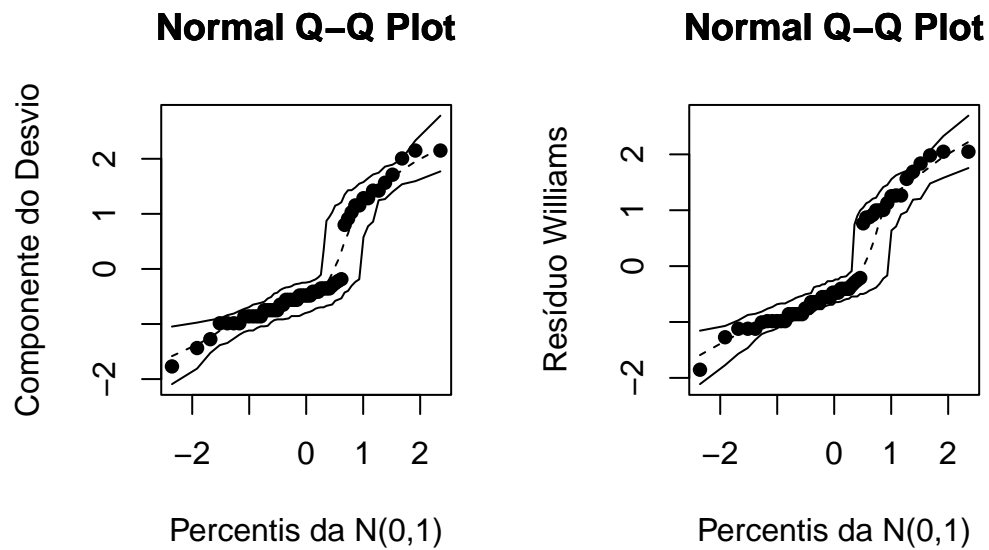


Figura 5: Análise de ajuste do modelo. (a) Componentes do desvio; (b) Resíduo de Williams.

Referências

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