

Exercício 5

Descrição

No arquivo **caduquice.dat** são descritos os dados de um experimento em que cinquenta e quatro indivíduos considerados idosos são submetidos a um exame psicológico para avaliar a ocorrência ou não de sintoma de caduquice. Acredita-se que o escore obtido num exame psicológico feito previamente esteja associado com a ocorrência ou não do sintoma. Os dados são apresentados na seguinte ordem:

- escore: escala no exame psicológico
- resp: ocorrência (resp = 1) ou não ocorrência (resp = 0) do sintoma.

Enunciado

- Faça inicialmente uma análise descritiva dos dados.
- Ajuste um modelo logístico para explicar a probabilidade de ocorrência do sintoma em função do escore.
- Interprete os resultados.
- Construa uma banda de confiança de 95% para a probabilidade de ocorrência do sintoma de caduquice dado o escore no exame psicológico.
- Faça uma análise de diagnóstico do modelo ajustado.

Leitura dos dados

```
caduquice <- data.frame(scan("dados/caduquice.dat",  
                             list(escore=0, resp=0)))
```

| | escore | resp |
|---|--------|------|
| 1 | 9.00 | 1.00 |
| 2 | 7.00 | 1.00 |
| 3 | 7.00 | 0.00 |
| 4 | 17.00 | 0.00 |
| 5 | 13.00 | 0.00 |

Análise descritiva

Na figura 1 apresentamos um boxplot do escore para os indivíduos que tiveram ocorrência dos sintomas e um outro para os indivíduos que não tiveram ocorrência do evento.

Neste gráfico é possível ver que os indivíduos que apresentaram os sintomas de caduquice tiveram aparentemente escores menores no teste psicológico.

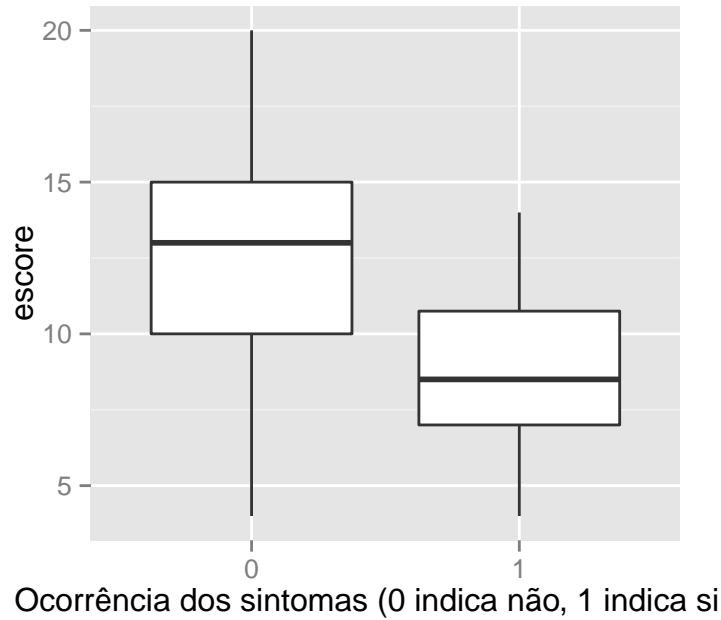


Figure 1: Boxplots do escore

Modelo

Vamos ajustar um modelo logístico que pode ser expresso da seguinte forma:

$$resp_i \sim Bernouli(p_i)$$

com:

$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta * escore_i$$

Este modelo pode ser ajustado no R usando o comando a seguir:

```
modelo <- glm(resp ~ escore, family=binomial(link = "logit"), data=caduquice)
```

As estimativas do modelo estão na tabela abaixo.

| | Estimativa | Erro Padrão | valor z | Pr(> t) |
|----------|------------|-------------|---------|----------|
| α | 2.4040 | 1.1918 | 2.02 | 0.0437 |
| β | -0.3235 | 0.1140 | -2.84 | 0.0045 |

Interpretação

Os parâmetros do modelo indicam que quanto maior o escore do indivíduo no teste psicológico, menor a chance de apresentar sintomas de caduquice. Mais precisamente, podemos dizer que com um aumento de uma unidade no escore do paciente, a chance é 72% da chance de um paciente com o escore de uma unidade menor.

Banda de confiança da probabilidade

A banda de confiança da probabilidade de apresentar os sintomas dado o escore no teste psicológico está apresentada na figura 2.

```
ggplot(caduquice, aes(y = resp, x = escore)) +  
  stat_smooth(method = "glm", family = binomial(link = logit))
```

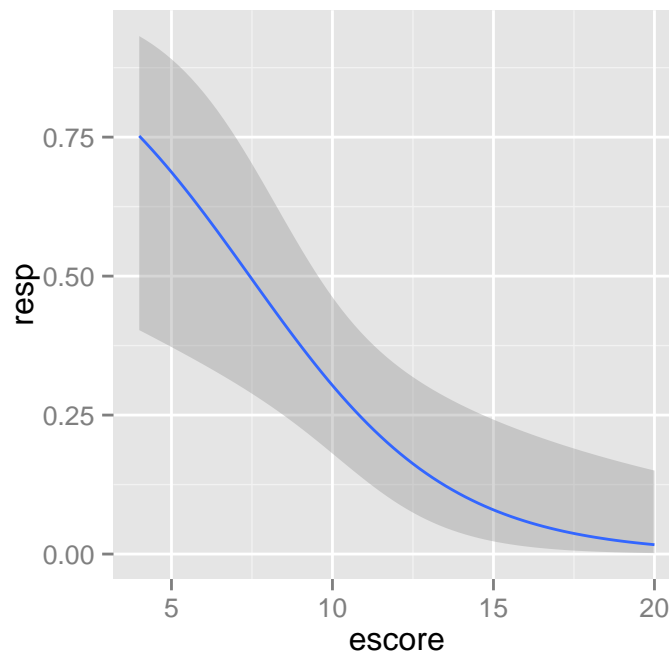


Figure 2: Banda de 95% de confiança da probabilidade de apresentar sintomas

Análise de diagnóstico

Pelo gráfico da figura 3 vemos que a suposição do modelo de que os dados seguem uma distribuição de Bernouli é razoável e pode ser aceita.

Pelos demais gráficos (figura 4) vemos que o modelo parece estar bem ajustado:

- os resíduos não apresentam nenhum padrão, indicando que não existe correlação entre as observações.
- não existe nenhuma observação com altíssima influência o que poderia causar instabilidade nos parâmetros.
- os resíduos tem valores entre -2 e 2, o que é considerado aceitável.

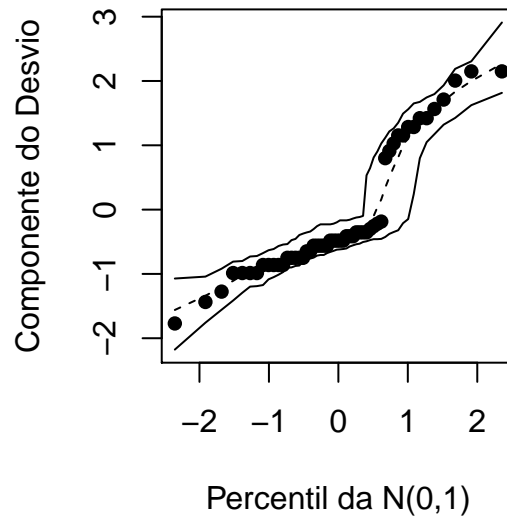


Figure 3: Gráfico Quantil-Quantil

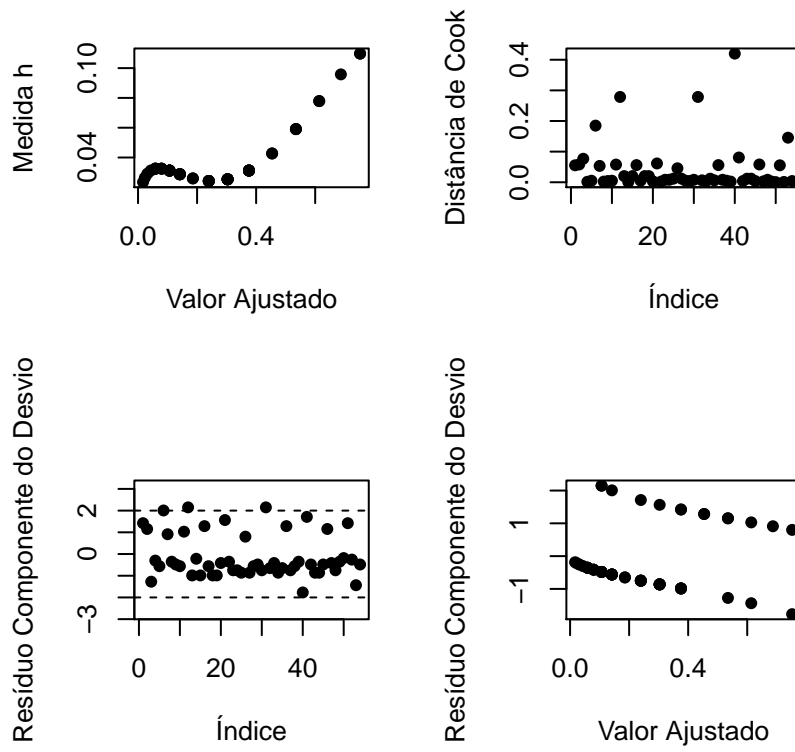


Figure 4: Gráficos de Diagnóstico