

# Exemplo Infecções de Ouvido

Gilberto A. Paula

Departamento de Estatística  
IME-USP, Brasil  
giapaula@ime.usp.br

2<sup>o</sup> Semestre 2023

- 1 Infecções de Ouvido
- 2 Análise de Dados Preliminar
- 3 Ajuste Modelo de Poisson
- 4 Ajuste Modelo com Resposta Binomial Negativa
- 5 Conclusões
- 6 Referências

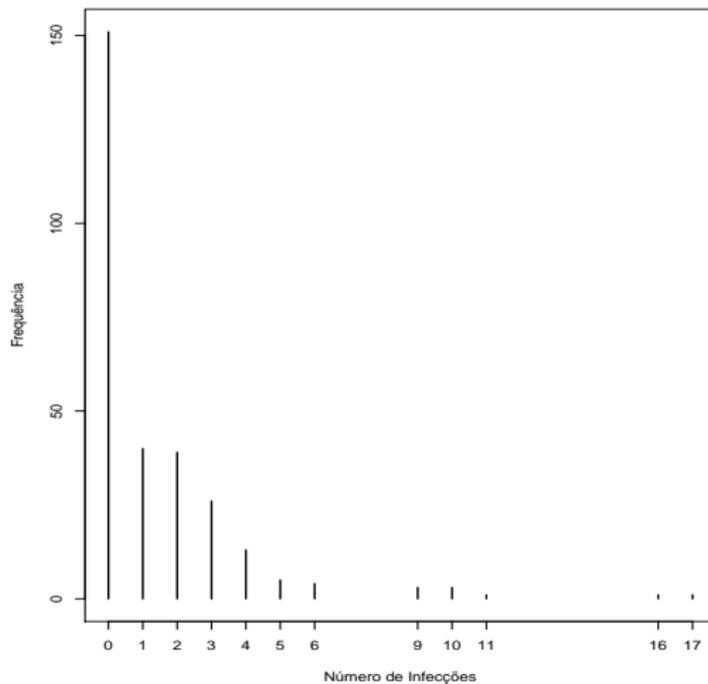
## Descrição dos Dados

Considere os dados apresentados em Hand et al. (1994) em que o número de infecções de ouvido observadas em uma amostra aleatória de  $n = 287$  recrutas norte-americanos em 1990 é relacionado com as seguintes variáveis explicativas:

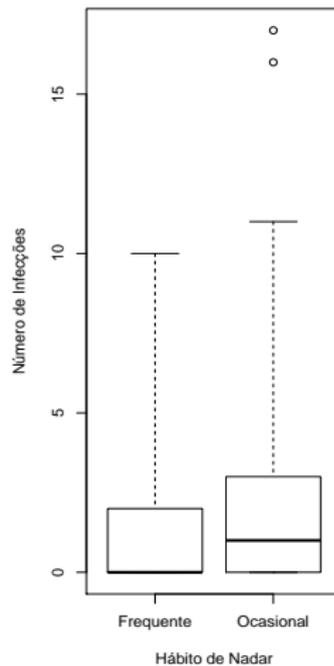
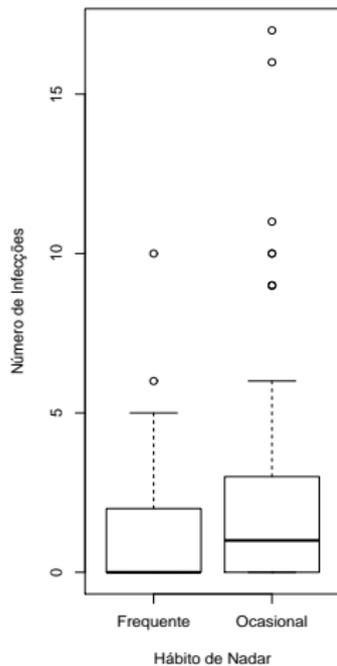
- hábito de nadar (ocasional ou frequente),
- local de nadar (praia ou piscina),
- faixa etária do recruta (15-19, 20-24, 25-29),
- gênero do recruta (feminino ou masculino).

- 1 Infecções de Ouvido
- 2 Análise de Dados Preliminar**
- 3 Ajuste Modelo de Poisson
- 4 Ajuste Modelo com Resposta Binomial Negativa
- 5 Conclusões
- 6 Referências

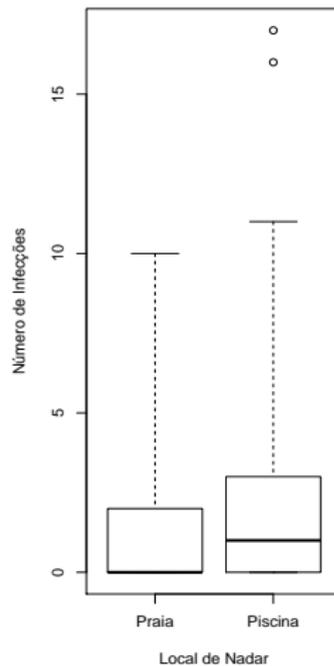
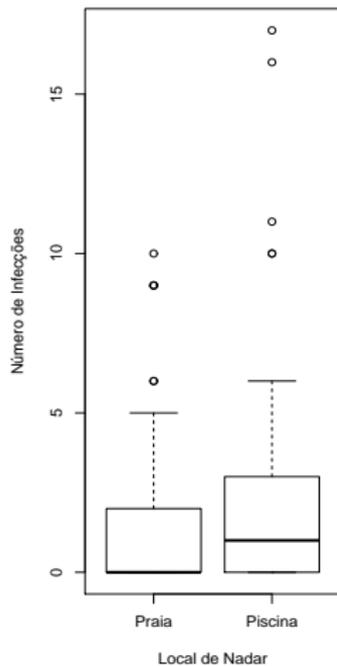
# Frequência Número de Infecções



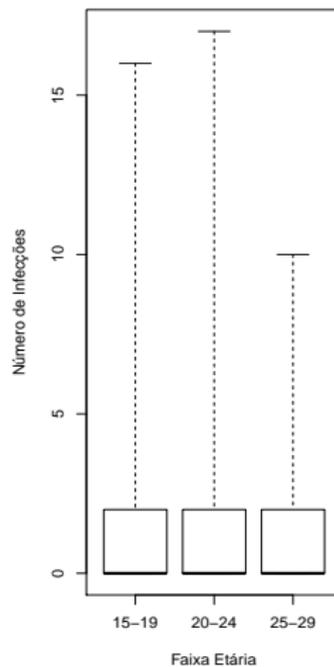
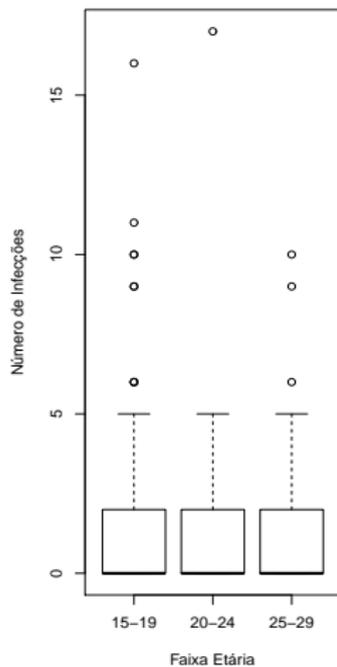
# Boxplots Número de Infecções por Hábito



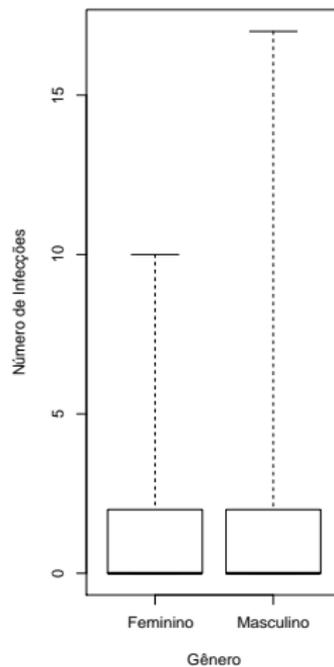
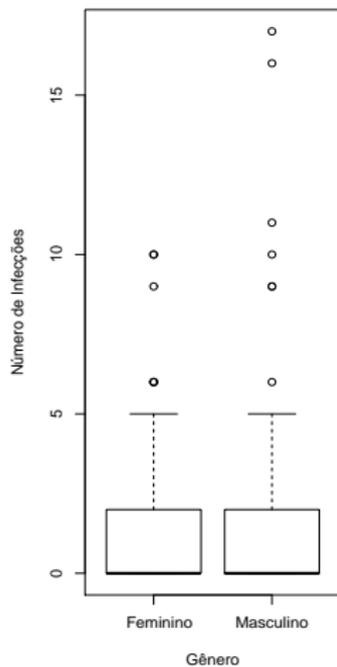
# Boxplots Número de Infecções por Local



# Boxplots Número de Infecções por Faixa Etária



# Boxplots Número de Infecções por Gênero



- 1 Infecções de Ouvido
- 2 Análise de Dados Preliminar
- 3 Ajuste Modelo de Poisson**
- 4 Ajuste Modelo com Resposta Binomial Negativa
- 5 Conclusões
- 6 Referências

## Descrição

Seja  $y_i$  o número de infecções de ouvido observadas pelo  $i$ -ésimo recruta no ano de 1990. Vamos supor inicialmente o seguinte modelo:

- $Y_i | \mathbf{x}_i \stackrel{\text{ind}}{\sim} P(\mu_i)$ ,
- $\log(\mu_i) = \alpha + \beta_1 \times \text{hábito}_i + \beta_2 \times \text{local}_i + \beta_3 \times \text{fetária}_{1i} + \beta_4 \times \text{fetária}_{2i} + \beta_5 \times \text{gênero}_i + \gamma_1 \times \text{hábito}_i * \text{local}_i + \gamma_2 \times \text{hábito}_i * \text{fetária}_{1i} + \gamma_3 \times \text{hábito}_i * \text{fetária}_{2i} + \gamma_4 \times \text{hábito}_i * \text{gênero}_i + \gamma_5 \times \text{local}_i * \text{fetária}_{1i} + \gamma_6 \times \text{local}_i * \text{fetária}_{2i} + \gamma_7 \times \text{local}_i * \text{gênero}_i + \gamma_8 \times \text{fetária}_{1i} * \text{gênero}_i + \gamma_9 \times \text{fetária}_{2i} * \text{gênero}_i$ ,

para  $i = 1, \dots, 287$ , em que **hábito**, **local**, **fetária<sub>1</sub>**, **fetária<sub>2</sub>** e **gênero** são variáveis binárias.

## Descrição

Aplicando o método de Akaike incluímos todos os efeitos principais + as interações hábito\*local, hábito\*fetária, local\*fetária e local\*gênero.

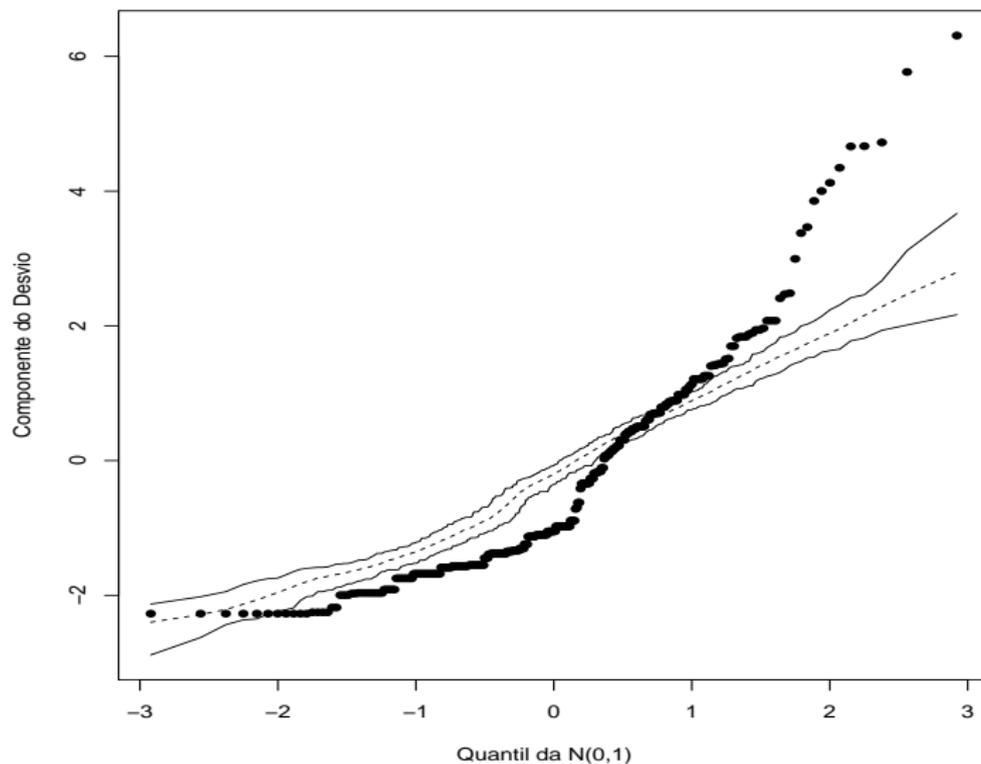
## Desvio

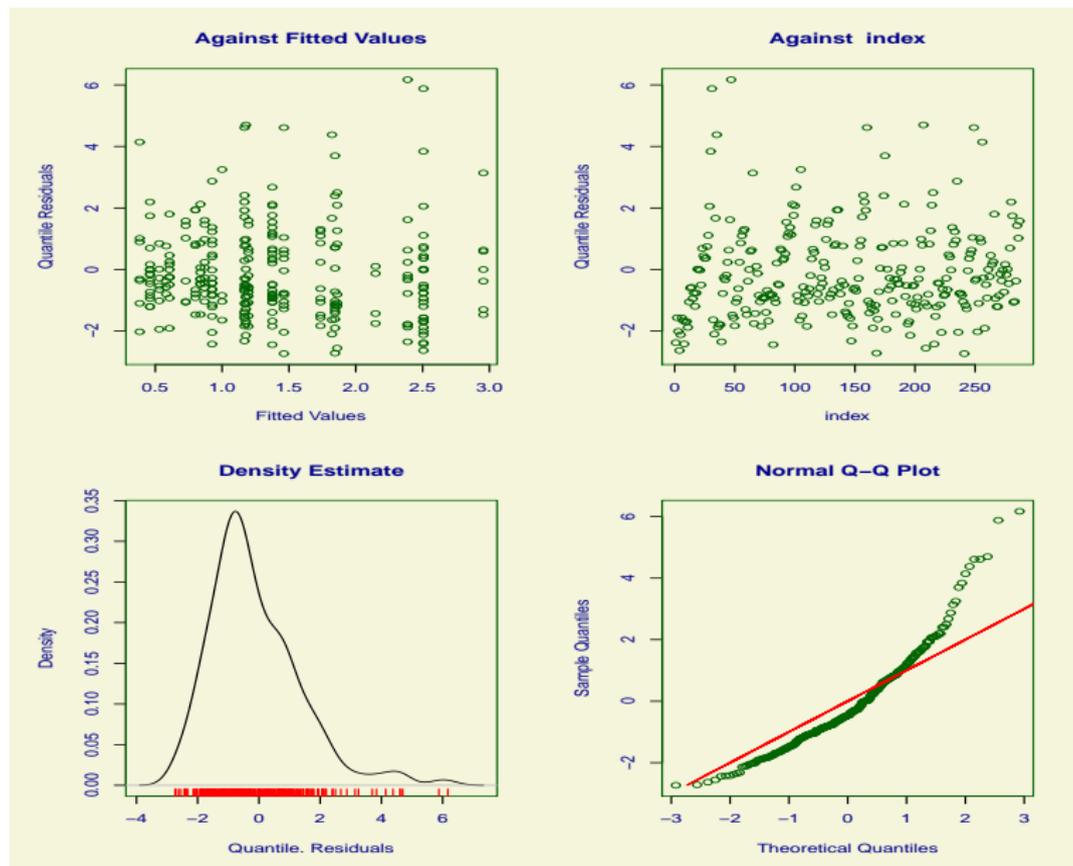
O desvio do modelo é dado por  $D(\mathbf{y}; \hat{\mu}) = 732,16$  (275 g.l.) com valor-P dado por  $P=0,00$  (rejeitamos o modelo).

## Descrição

Efeito	Estimativa	E.Padrão	Valor-z
Intercepto	0,380	0,169	2,25
HábitoOcas	0,232	0,189	1,23
LocalPisc	-0,380	0,258	-1,47
Fetária20-24	-0,880	0,289	-3,04
Fetária25-29	-0,698	0,277	-2,52
GêneroMasc	-0,457	0,163	-2,81
HOcas*LPisc	0,367	0,222	1,65
HOcas*F20-24	0,091	0,261	0,35
HOcas*F25-30	0,706	0,293	2,41
LPisc*F20-24	0,741	0,294	2,52
LPisc*F25-29	0,156	0,265	0,59
LPisc*GMasc	0,776	0,234	3,32

# Resíduos Modelo Ajustado Poisson





- 1 Infecções de Ouvido
- 2 Análise de Dados Preliminar
- 3 Ajuste Modelo de Poisson
- 4 Ajuste Modelo com Resposta Binomial Negativa**
- 5 Conclusões
- 6 Referências

## Descrição

A sobredispersão detectada sob o modelo de Poisson sugere o seguinte modelo com resposta binomial negativa:

- $Y_i | \mathbf{x}_i \stackrel{\text{ind}}{\sim} \text{BN}(\mu_i, \nu)$ ,
- $\log(\mu_i) = \alpha + \beta_1 \times \text{hábito}_i + \beta_2 \times \text{local}_i + \beta_3 \times \text{fetária}_{1i} + \beta_4 \times \text{fetária}_{2i} + \beta_5 \times \text{gênero}_i + \gamma_1 \times \text{hábito}_i * \text{local}_i + \gamma_2 \times \text{hábito}_i * \text{fetária}_{1i} + \gamma_3 \times \text{hábito}_i * \text{fetária}_{2i} + \gamma_4 \times \text{hábito}_i * \text{gênero}_i + \gamma_5 \times \text{local}_i * \text{fetária}_{1i} + \gamma_6 \times \text{local}_i * \text{fetária}_{2i} + \gamma_7 \times \text{local}_i * \text{gênero}_i + \gamma_8 \times \text{fetária}_{1i} * \text{gênero}_i + \gamma_9 \times \text{fetária}_{2i} * \text{gênero}_i$ ,

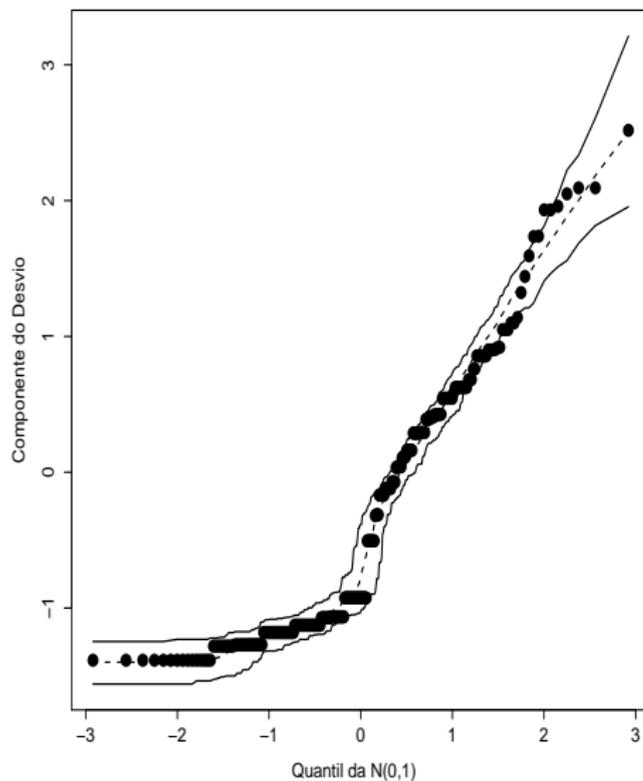
para  $i = 1, \dots, 287$ , em que **hábito**, **local**, **fetária<sub>1</sub>**, **fetária<sub>2</sub>** e **gênero** são variáveis binárias.

## Descrição

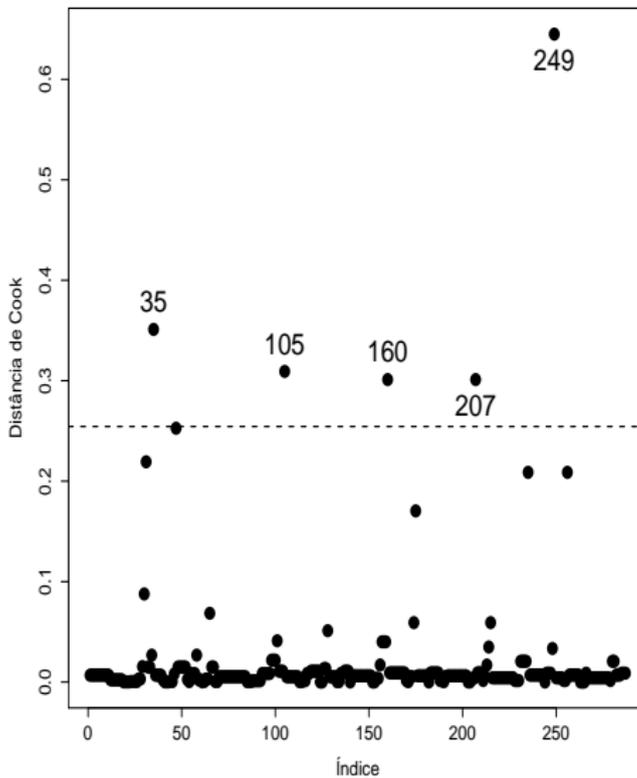
Aplicando o método de Akaike incluímos apenas a interação entre local e gênero, além dos efeitos principais hábito, local e gênero.

Efeito	Estimativa	E.Padrão	Valor-z
Intercepto	-0,064	0,229	-0,28
HábitoOcas	0,593	0,190	3,12
LocalPisc	0,007	0,330	0,02
GêneroMasc	-0,407	0,275	-1,48
LPisc*GMasc	0,745	0,408	1,83
$\nu$	0,576	0,091	6,33

O desvio do modelo é dado por  $D(\mathbf{y}; \hat{\mu}) = 269,54$  (282 g.l.) com valor-P dado por  $P=0,69$  (não rejeitamos o modelo).



# Distância de Cook Modelo Ajustado BN



## Descrição

Estimativas eliminando a observação #249.

Efeito	Estimativa	E.Padrão	Valor-z
Intercepto	-0,263	0,233	-1,13
HábitoOcas	0,669	0,188	3,56
LocalPisc	0,163	0,329	0,50
GêneroMasc	-0,250	0,275	-0,91
LPisc*GMasc	0,591	0,405	<b>1,46</b>
$\nu$	0,606	0,098	6,18

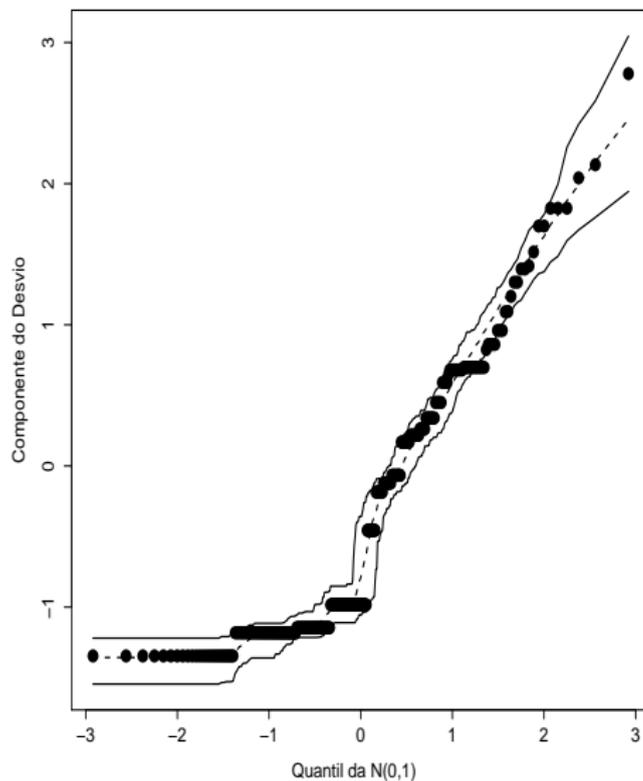
Portanto, **há mudança inferencial**, essa observação está induzindo a interação entre local e gênero. A observação #249 tem um perfil atípico, refere-se a uma recruta que nada com frequência na praia e registrou 10 infecções no ano de 1990. A retirada das demais observações não muda a inferência.

## Descrição

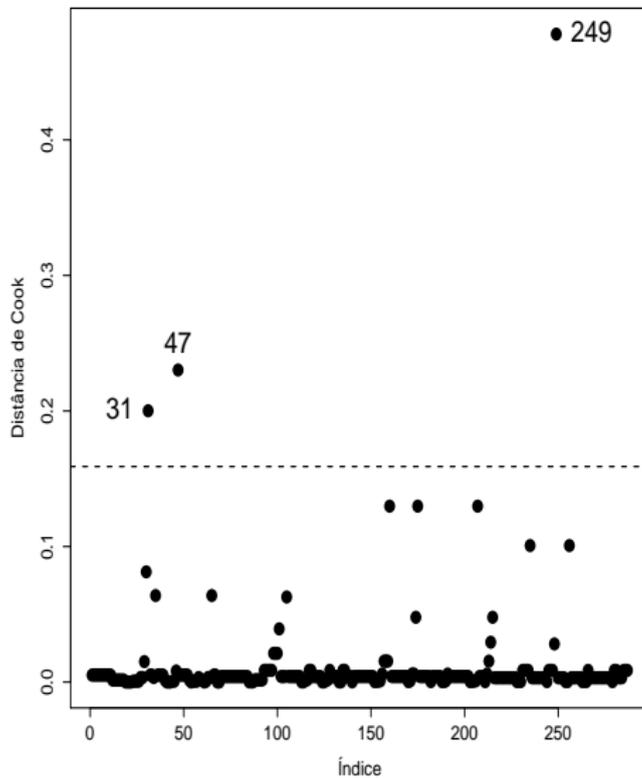
O modelo final tem apenas os efeitos principais de hábito e local. As estimativas são apresentadas abaixo.

Efeito	Estimativa	E.Padrão	Valor-z
Intercepto	-0,278	0,174	-1,59
HábitoOcas	0,590	0,191	3,10
LocalPisc	0,481	0,190	2,53
$\nu$	0,560	0,087	6,43

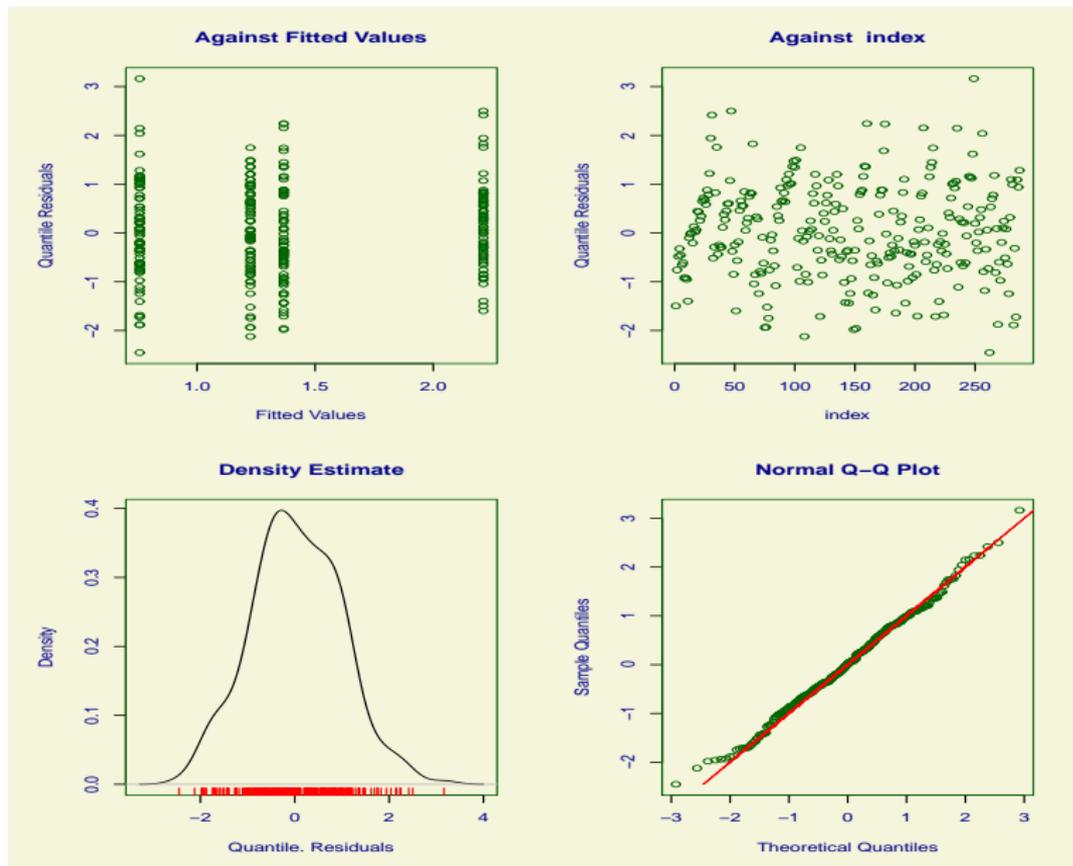
O desvio do modelo é dado por  $D(\mathbf{y}; \hat{\mu}) = 269,10$  (284 g.l.) com valor-P dado por  $P=0,73$  (não rejeitamos o modelo).

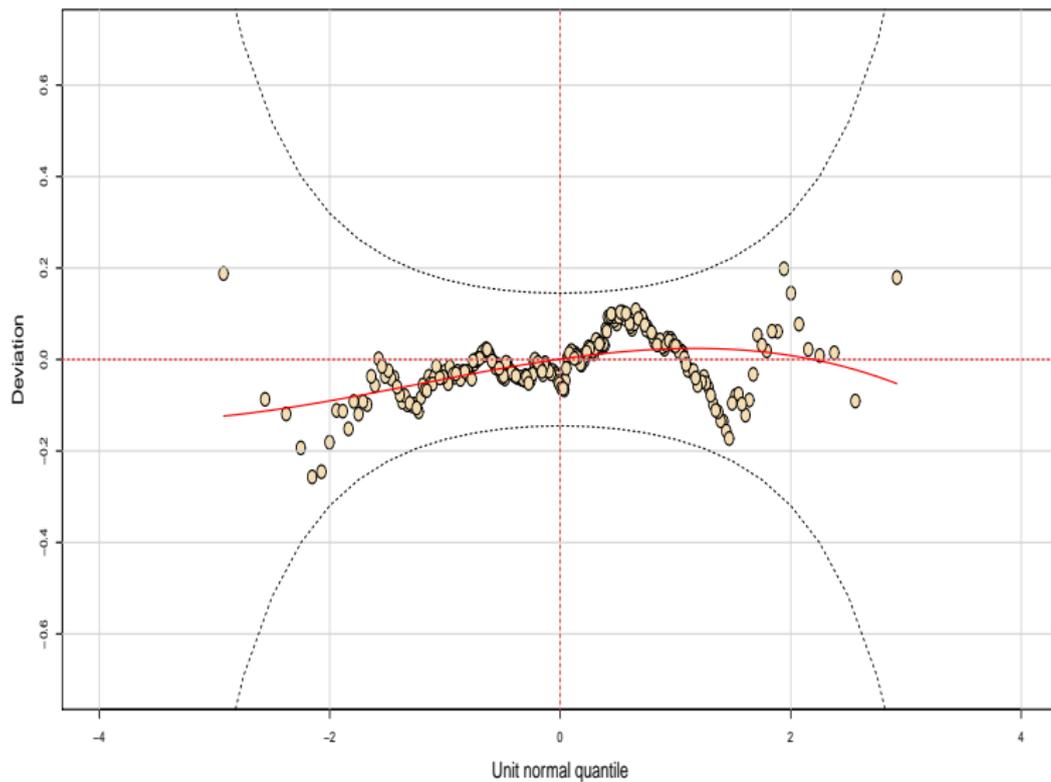


# Distância de Cook Modelo Final BN

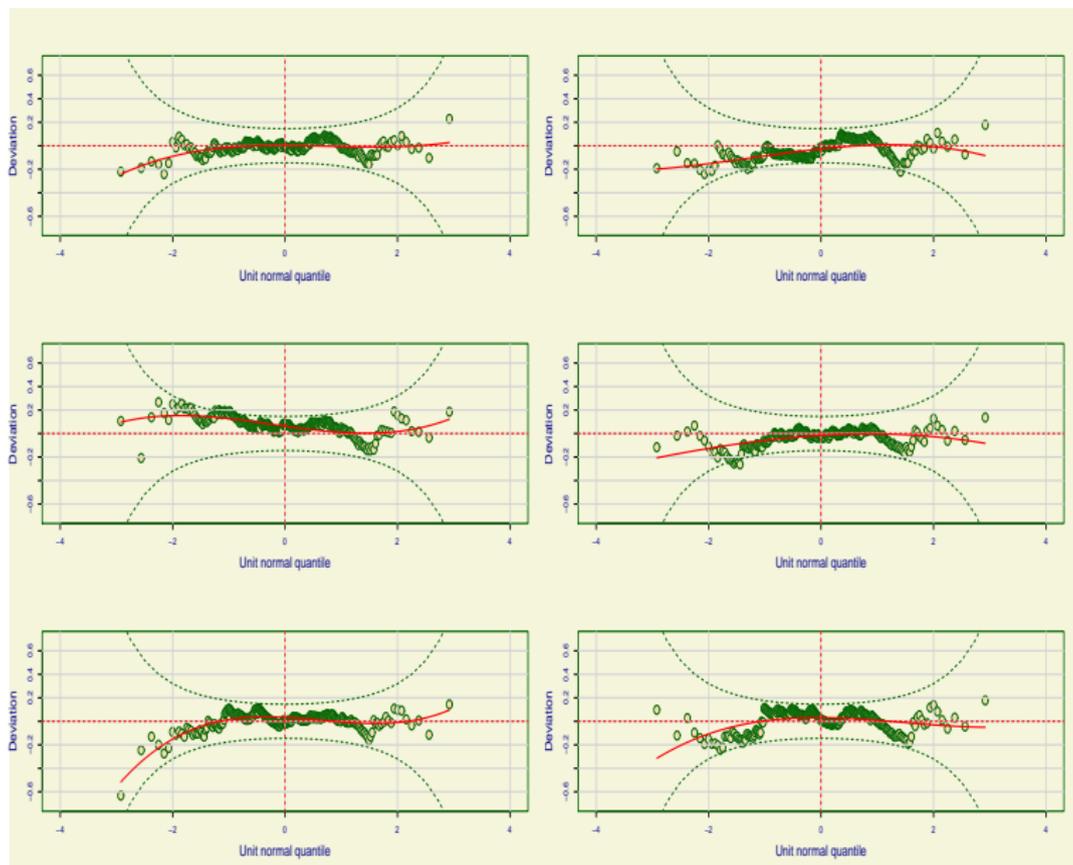


# Resíduos GAMLSS, em que $\sigma = \frac{1}{\nu}$





# Worm Plot GAMLSS



## Descrição

Estimativas eliminando a observação #249.

Efeito	Estimativa	E.Padrão	Valor-z
Intercepto	-0,403	0,176	-2,29
HábitoOcas	0,675	0,189	3,56
LocalPisc	0,561	0,188	2,98
$\nu$	0,594	0,095	6,25

O desvio do modelo é dado por  $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 268,81$  (284 g.l.) com valor-P dado por  $P=0,72$  (não rejeitamos o modelo). Nota-se aumento das estimativas e significância dos efeitos de hábito ocasional e local piscina. O intercepto fica significativo. A retirada das demais observações não muda a inferência.

### Interpretações Modelo Final

- Recrutas com hábito de nadar ocasional têm aproximadamente **80%** ( $\exp(0,59) = 1,80$ ) mais infecções de ouvido do que recrutas com hábito de nadar frequente.
- Recrutas que nadam em piscina têm aproximadamente **62%** ( $\exp(0,481) = 1,62$ ) mais infecções de ouvido do que recrutas que nadam na praia.
- Recrutas com hábito de nadar ocasional em piscina têm aproximadamente **192%** ( $\exp(0,590 + 0,481) = 2,92$ ) mais infecções de ouvido do que recrutas com hábito de nadar frequente na praia.

- 1 Infecções de Ouvido
- 2 Análise de Dados Preliminar
- 3 Ajuste Modelo de Poisson
- 4 Ajuste Modelo com Resposta Binomial Negativa
- 5 Conclusões**
- 6 Referências

### Conclusões Finais

- Este é um exemplo com sobredispersão em que o modelo com resposta binomial negativa ajusta melhor os dados do que o modelo de Poisson.
- A observação discrepante #249 induz a interação entre hábito e gênero no modelo binomial negativo. Essa observação aumenta a significância dos efeitos do modelo final bem como o valor das estimativas dos efeitos **hábito ocasional** e **local piscina**. Esse resultado é esperado levando-se em conta o perfil da observação.

- 1 Infecções de Ouvido
- 2 Análise de Dados Preliminar
- 3 Ajuste Modelo de Poisson
- 4 Ajuste Modelo com Resposta Binomial Negativa
- 5 Conclusões
- 6 Referências**

## Referência

- Hand, D. J., Daly, F., Lunn, A. D., McConway, K. J. e Ostrowski, E. (1994). *A Handbook of Small Data Sets*. Chapman and Hall, London.
- Stasinopoulos, M. D., Rigby, R. A., Gillian, Z. A., Voudouris, V. e de Bastiani, F. (2017). *Flexible Regression and Smoothing Using GAMLSS in R*. Chapman and Hall/CRC.