

Exemplo Infecções de Ouvido com Zeros

Gilberto A. Paula

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

2^o Semestre 2023

- 1 Infecções de Ouvido
- 2 Análise de Dados Preliminar
- 3 Modelo ZAP: Poisson Ajustado em Zero
- 4 Modelo ZANBI: Binomial Negativo Ajustado em Zero
- 5 Comparação de Modelos
- 6 Modelo Final
- 7 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 Modelo ZAP: Poisson Ajustado em Zero
- 4 Modelo ZANBI: Binomial Negativo Ajustado em Zero
- 5 Comparação de Modelos
- 6 Modelo Final
- 7 Referências

Frequência Número de Infecções

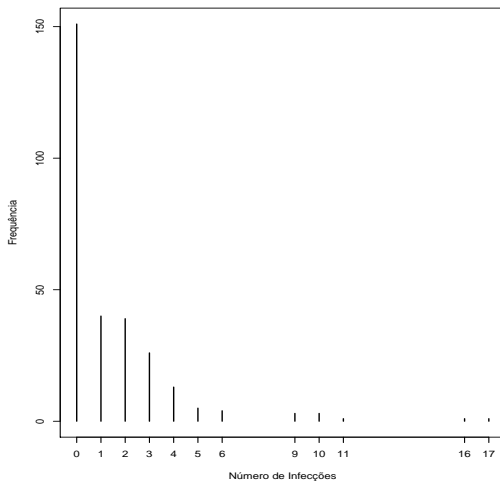


Tabela de Frequências

	Número de Infecções					Total
	0	1	2	3	>3	
Frequência	151	40	39	26	31	287

Tabela de Porcentagens

	Número de Infecções					Total
	0	1	2	3	>3	
Porcentagem	52,61	13,94	13,59	9,06	10,80	100,0

Tabela de Frequências

Hábito	Número de Infecções					Total
	0	1	2	3	>3	
Frequente	82	20	22	11	8	143
Ocasional	69	20	17	15	23	144
Total	151	40	39	26	31	287

Tabela de Porcentagens

Hábito	Número de Infecções					Total
	0	1	2	3	>3	
Frequente	57,34	13,99	15,38	7,69	5,60	100,0
Ocasional	47,92	13,89	11,80	10,42	15,97	100,0

Comentário

- há fortes indícios que recrutas com hábito frequente estejam mais propensos a não terem infecção de ouvido do que recrutas com hábito ocasional.

Tabela de Frequências

Local	Número de Infecções					Total
	0	1	2	3	>3	
Praia	90	19	19	6	13	147
Piscina	61	21	20	20	18	140
Total	151	40	39	26	31	287

Tabela de Porcentagens

Local	Número de Infecções					Total
	0	1	2	3	>3	
Praia	61,22	12,93	12,93	4,08	8,84	100,0
Piscina	43,57	15,00	14,29	14,29	12,85	100,0

Comentário

- há fortes indícios que recrutas que nadam na praia estejam mais propensos a não terem infecção de ouvido do que recrutas que nadam em piscina.

Tabela de Frequências

Faixa Etária	Número de Infecções					Total
	0	1	2	3	>3	
15-19	71	17	19	13	20	140
20-24	42	14	10	9	4	79
25-29	38	9	10	4	7	68
Total	151	40	39	26	31	287

Tabela de Porcentagens

Faixa Etária	Número de Infecções					Total
	0	1	2	3	>3	
15-19	50,71	12,14	13,57	9,29	14,29	100,0
20-24	53,17	17,72	12,66	11,39	5,06	100,0
25-29	55,88	13,24	14,71	5,88	10,29	100,0

Comentário

- à medida que aumenta a faixa etária há indícios de aumento da chance do recruta não ter infecção de ouvido.

Tabela de Frequências

Gênero	Número de Infecções					Total
	0	1	2	3	>3	
Feminino	54	12	14	9	10	99
Masculino	97	28	25	17	21	188
Total	151	40	39	26	31	287

Tabela de Porcentagens

Gênero	Número de Infecções					Total
	0	1	2	3	>3	
Feminino	54,55	12,12	14,14	9,09	10,10	100,0
Masculino	51,60	14,89	13,30	9,04	11,17	100,0

Comentário

- o grupo masculino parece menos propenso a não ter infecção de ouvido do que o grupo feminino.

- 1 Infecções de Ouvido
- 2 Análise de Dados Preliminar
- 3 Modelo ZAP: Poisson Ajustado em Zero**
- 4 Modelo ZANBI: Binomial Negativo Ajustado em Zero
- 5 Comparação de Modelos
- 6 Modelo Final
- 7 Referências

Descrição

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

$$f_z(z_i; \mu_i, \pi_i) = \begin{cases} \pi_i & \text{se } z_i = 0 \\ (1 - \pi_i) \frac{f_y(z_i; \mu_i)}{1 - f_y(0; \mu_i)} & \text{se } z_i = 1, 2, \dots \end{cases}$$

em que $f_y(y_i; \mu_i)$ denota a função e probabilidades uma $P(\mu_i)$ com parte sistemática dada por

- $\log(\mu_i)$ = efeitos principais + interações de 1^a ordem,
- $\log\left(\frac{\pi_i}{1 - \pi_i}\right)$ = efeitos principais + interações de 1^a ordem,

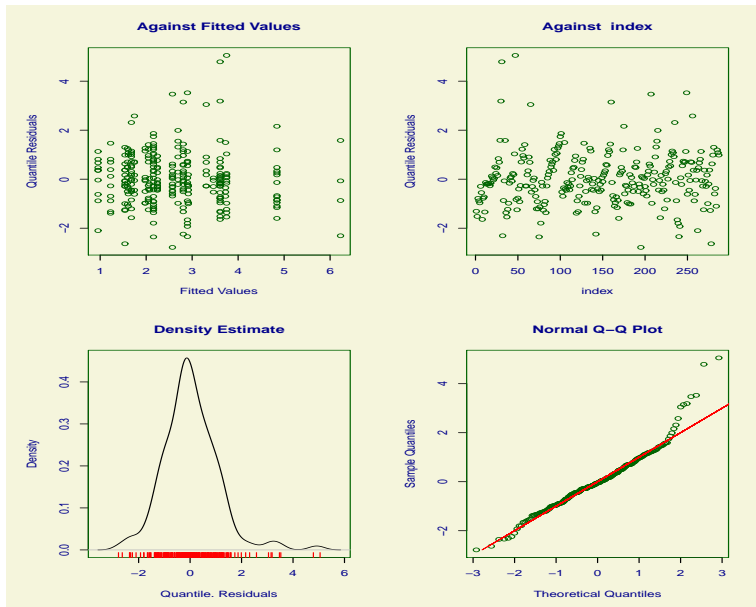
para $i = 1, \dots, 287$.

Estimativas para $\log(\mu)$

Efeito	Estimativa	E.Padrão	Valor-z
Intercepto	1,064	0,157	6,77
HábitoOcas	0,514	0,123	4,18
LocalPisc	0,251	0,129	1,96
GêneroMasc	-0,546	0,169	-3,23
Fetária20-24	-1,111	0,258	-4,32
Fetária25-29	-0,587	0,245	-2,39
F20-24*GMasc	1,150	0,308	3,73
F25-29*GMasc	0,500	0,304	1,83

Estimativas para $\log\{\pi/(1 - \pi)\}$

Efeito	Estimativa	E.Padrão	Valor-z
Intercepto	0,668	0,213	3,13
HábitoOcas	-0,405	0,242	-1,67
LocalPisc	-0,730	0,242	-3,02



- 1 Infecções de Ouvido
- 2 Análise de Dados Preliminar
- 3 Modelo ZAP: Poisson Ajustado em Zero
- 4 Modelo ZANBI: Binomial Negativo Ajustado em Zero**
- 5 Comparação de Modelos
- 6 Modelo Final
- 7 Referências

Descrição

Seja z_i o número de infecções de ouvido observadas pelo i -ésimo recruta no ano de 1990. Vamos supor agora o seguinte modelo ZANBI:

$$f_z(z_i; \mu_i, \pi_i, \nu) = \begin{cases} \pi_i & \text{se } z_i = 0 \\ (1 - \pi_i) \frac{f_y(z_i; \mu_i, \nu)}{1 - f_y(0; \mu_i, \nu)} & \text{se } z_i = 1, 2, \dots \end{cases}$$

em que $f_y(y_i; \mu_i, \nu)$ denota a função e probabilidades uma **BN**(μ_i, ν), em que $\sigma = \frac{1}{\nu}$ com parte sistemática dada por

- $\log(\mu_i) =$ efeitos principais + interações de 1^a ordem,
- $\log\left(\frac{\pi_i}{1 - \pi_i}\right) =$ efeitos principais + interações de 1^a ordem,

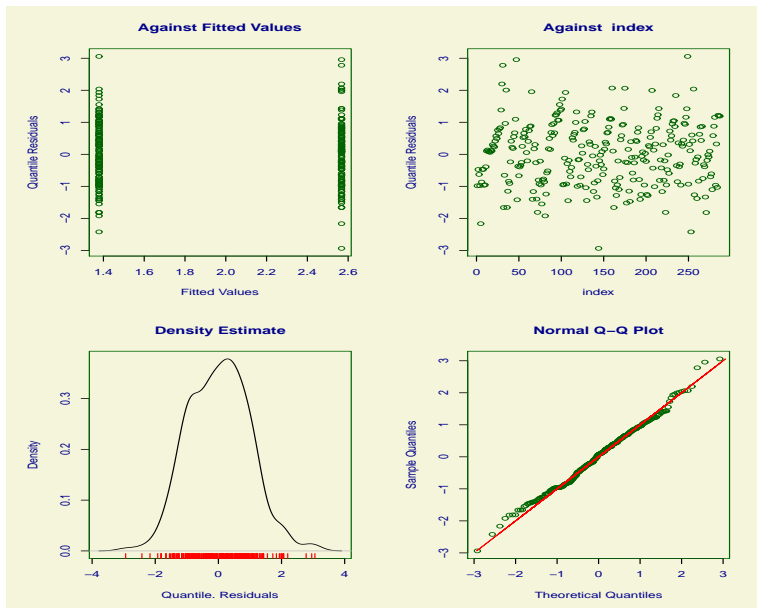
para $i = 1, \dots, 287$.

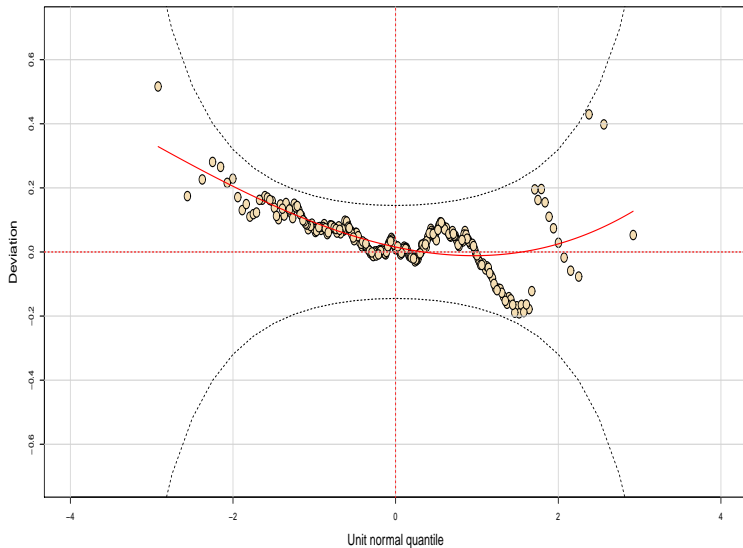
Estimativas para $\log(\mu)$

Efeito	Estimativa	E.Padrão	Valor-z
Intercepto	0,321	0,207	1,55
HábitoOcas	0,622	0,210	2,95
σ	0,831	0,328	2,53

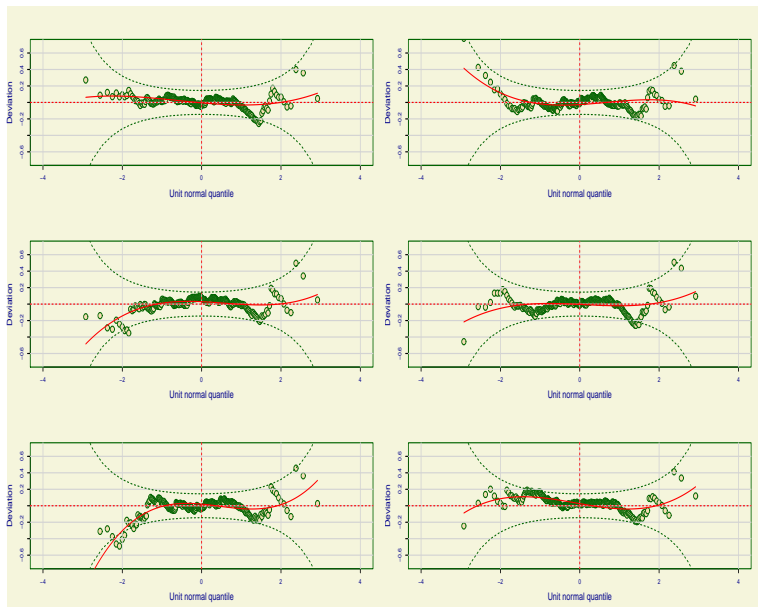
Estimativas para $\log\{\pi/(1 - \pi)\}$

Efeito	Estimativa	E.Padrão	Valor-z
Intercepto	0,668	0,213	3,13
HábitoOcas	-0,405	0,242	-1,67
LocalPisc	-0,730	0,242	-3,02





Worm Plot GAMLSS



Estimativas para $\log(\mu)$

Efeito	Estimativa	E.Padrão	Valor-z
Intercepto	0,260	0,202	1,29
HábitoOcas	0,716	0,208	3,44
σ	0,727	0,284	2,56

Estimativas para $\log\{\pi/(1 - \pi)\}$

Efeito	Estimativa	E.Padrão	Valor-z
Intercepto	0,699	0,215	3,24
HábitoOcas	-0,426	0,243	-1,75
LocalPisc	-0,750	0,243	-3,09

- 1 Infecções de Ouvido
- 2 Análise de Dados Preliminar
- 3 Modelo ZAP: Poisson Ajustado em Zero
- 4 Modelo ZANBI: Binomial Negativo Ajustado em Zero
- 5 Comparação de Modelos**
- 6 Modelo Final
- 7 Referências

Comparação de Modelos

- Modelo de Poisson ajustado em zero: **GAIC=949,9532**
- Modelo binomial negativo ajustado em zero: **GAIC=899,8217**

Comparando os gráficos normais de probabilidade com os resíduos quantílicos nota-se um melhor ajuste com o **modelo binomial negativo ajustado em zero**.

- 1 Infecções de Ouvido
- 2 Análise de Dados Preliminar
- 3 Modelo ZAP: Poisson Ajustado em Zero
- 4 Modelo ZANBI: Binomial Negativo Ajustado em Zero
- 5 Comparação de Modelos
- 6 Modelo Final**
- 7 Referências

Interpretações

- Recrutas que nadam ocasionalmente têm em média mais infecções de ouvido do que recrutas que nadam com frequência.
- A chance de um recruta que nada com frequência, não ter infecção de ouvido, é $\exp(0,405)=1,50$ a chance de um recruta que nada ocasionalmente.
- A chance de um recruta que nada na praia, não ter infecção de ouvido, é $\exp(0,730)=2,08$ a chance de um recruta que nada em piscina.

Conclusões Finais

- Este é um exemplo com sobredispersão e excesso de zeros em que o modelo binomial negativo ajustado em zero apresenta um ajuste melhor com relação ao modelo de Poisson ajustado em zero.
- As interpretações entre o modelo binomial negativo e o modelo binomial negativo ajustado em zero são similares, contudo o último modelo permite interpretar as razões de chances de não infecção de ouvido entre os fatores selecionados.
- A observação discrepante #249 que aparecia como influente no modelo binomial negativo não é mais influente no modelo binomial negativo ajustado em zero.

- 1 Infecções de Ouvido
- 2 Análise de Dados Preliminar
- 3 Modelo ZAP: Poisson Ajustado em Zero
- 4 Modelo ZANBI: Binomial Negativo Ajustado em Zero
- 5 Comparação de Modelos
- 6 Modelo Final
- 7 Referências**

Referências

- 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.
- Zeileis, A., Kleiber, C. e Jackman, M. (2008). Regression models for count data in R. *Journal of Statistical Software* 27(8),1-25.