Exemplo Regressão Binomial de Dose Resposta

Gilberto A. Paula

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

$2^{\underline{o}}$ Semestre 2023

Sumário

Descrição dos Dados

Como aplicação de modelos binomiais de dose resposta vamos considerar os dados descritos em Bliss (1935), em que besouros adultos são submetidos à exposição do composto

- disulfeto de carbono gasoso (CS_2)

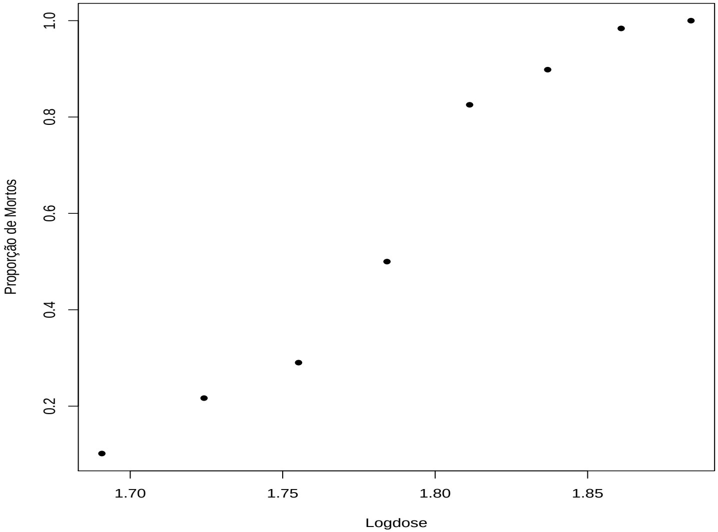
durante cinco horas.
Os resultados obtidos a partir de 481 besouros expostos segundo diferentes doses de CS_2 são apresentados a seguir. Um dos objetivos desse estudo é estimar

- DL_{100p}: dose letal que mata 100p%.

Descrição dos Dados

Dose $\log_{10} CS_2$	Besouros expostos	Besouros mortos
1,6907	59	6
1,7242	60	13
1,7552	62	18
1,7842	56	28
1,8113	63	52
1,8369	59	53
1,8610	62	61
1,8839	60	60

Proporção de Besouros Mortos versus Logdose



Sumário

Definição

Seja y_i o número de besouros mortos dentre os n_i submetidos à i-ésima logdose de CS_2, para $i = 1, \ldots, 8$. Vamos supor inicialmente o seguinte modelo:

- $Y_i | x_i \overset{\text{ind}}{\sim} \mathrm{B}\{n_i, \pi(x_i)\}$,
- $\log\left\{\frac{\pi(x_i)}{1-\pi(x_i)}\right\} = \beta_1 + \beta_2 x_i$,
- $\pi(x_i) = \frac{\exp(\beta_1+\beta_2 x_i)}{1+\exp(\beta_1+\beta_2 x_i)}$,

em que x_i denota a i-ésima logdose de CS_2.

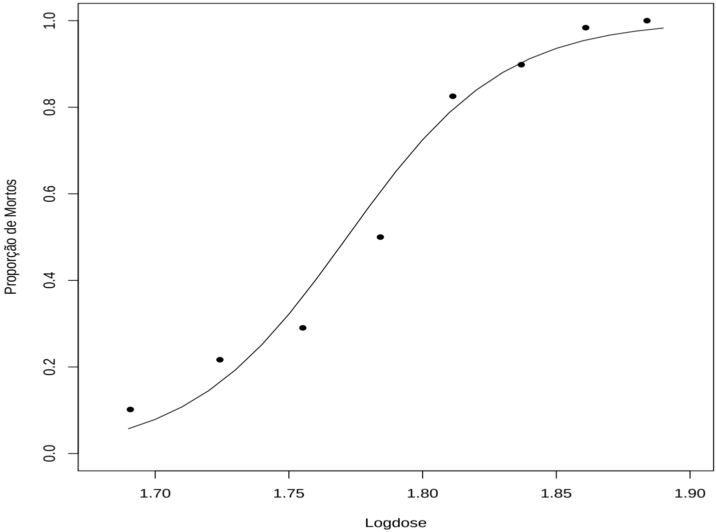
Sumário

Estimativas

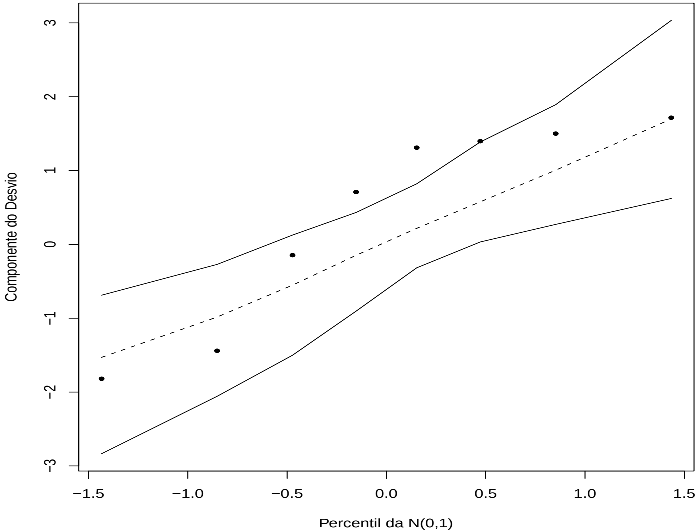
Efeito	Estimativa	E.Padrão	Valor-z	Valor-P
Constante	-60,72	5,18	-11,72	0,00
Logdose	34,27	2,91	11,78	0,00

O desvio do modelo é dado por $D(\mathbf{y}; \hat{\mu}) = 11,23$ (6 g.l.), com nível descritivo $P = 0,081$. Embora os efeitos sejam altamente significativos, parece que o modelo não está bem ajustado. Contudo, desde que $D(\mathbf{y}; \bar{\mathbf{y}}) = 282,20$, tem-se que $R^2 = 1 - \frac{11,23}{282,20} = 0,96$.

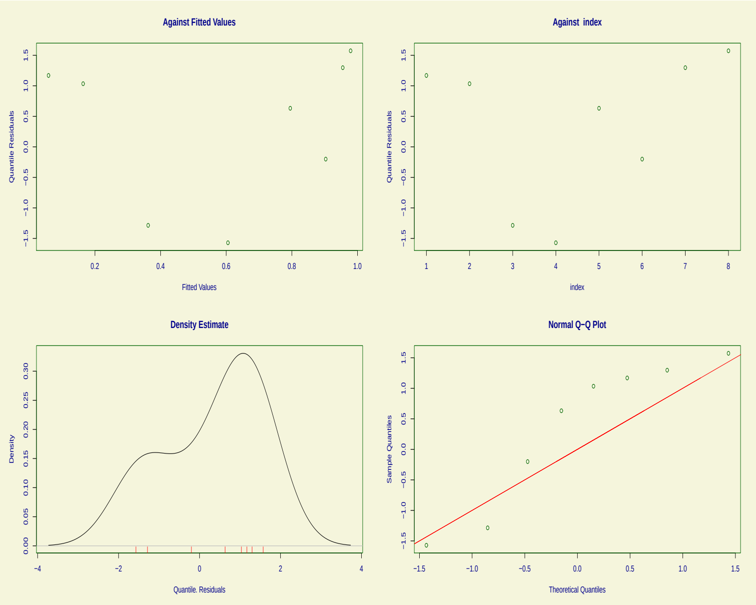
Curva Logística Ajustada



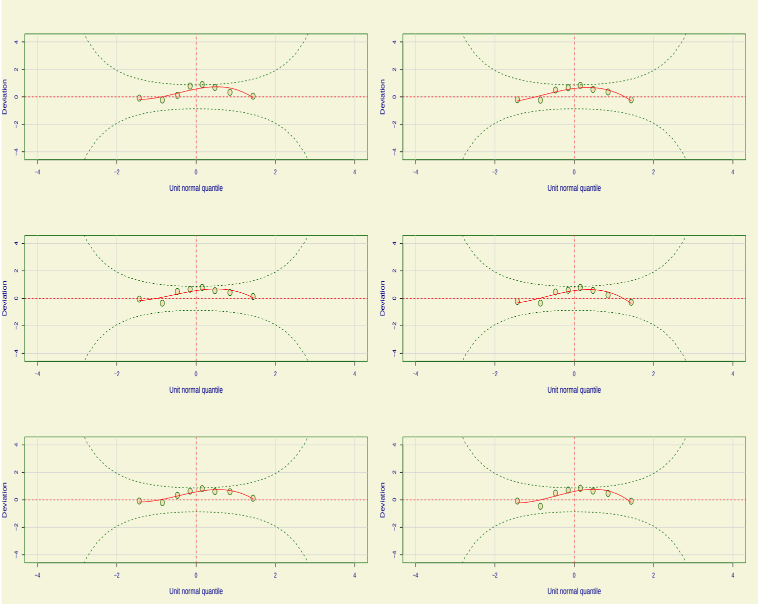
Resíduos Modelo Logístico



Resíduos GAMLSS



Worm Plot GAMLSS



Sumário

Definição

Seja y_i o números de besouros mortos dentre os n_i submetidos à i-ésima logdose de CS_2, para $i = 1, \ldots, 8$.Vamos supor agora o seguinte modelo:

- $Y_i|x_i \stackrel{\text{ind}}{\sim} \text{B}\{n_i, \pi(x_i)\}$,
- $\log[-\log\{1 - \pi(x_i)\}] = \beta_1 + \beta_2 x_i$,
- $\pi(x_i) = 1 - \exp\{-\exp(\beta_1 + \beta_2 x_i)\}$,

em que x_i denota a i-ésima logdose de CS_2.

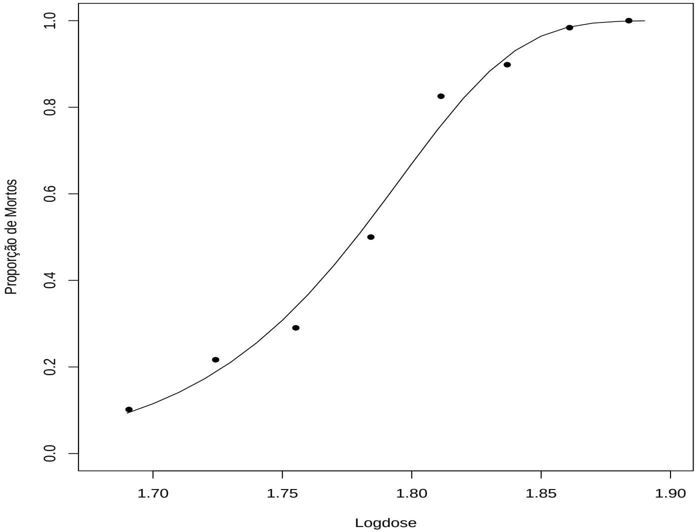
Sumário

Estimativas

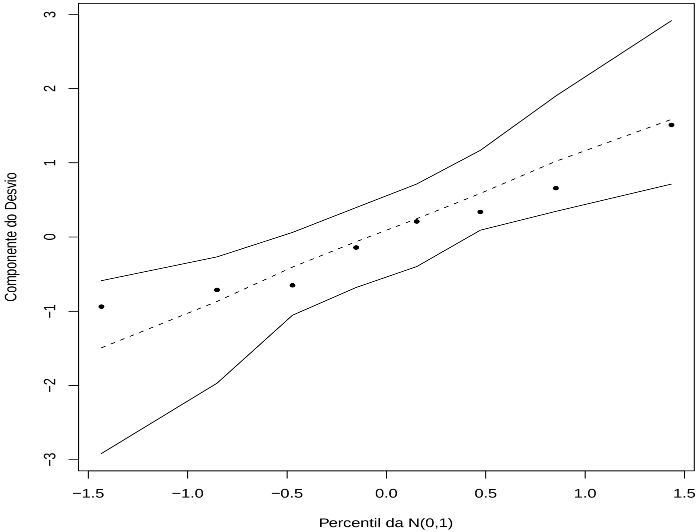
Efeito	Estimativa	E.Padrão	Valor-z	Valor-P
Constante	-39,57	3,24	-12,21	0,00
Logdose	22,04	1,80	12,25	0,00

O desvio do modelo é dado por $D(\mathbf{y}; \hat{\mu}) = 3,45$ (6 g.l.), com nível descritivo $P = 0,75$. Temos agora indícios de que o modelo está bem ajustado. Desde que $D(\mathbf{y}; \bar{\mathbf{y}}) = 282,20$, tem-se agora $R^2 = 1 - \frac{3,45}{282,20} = 0,99$ (excelente ajuste).

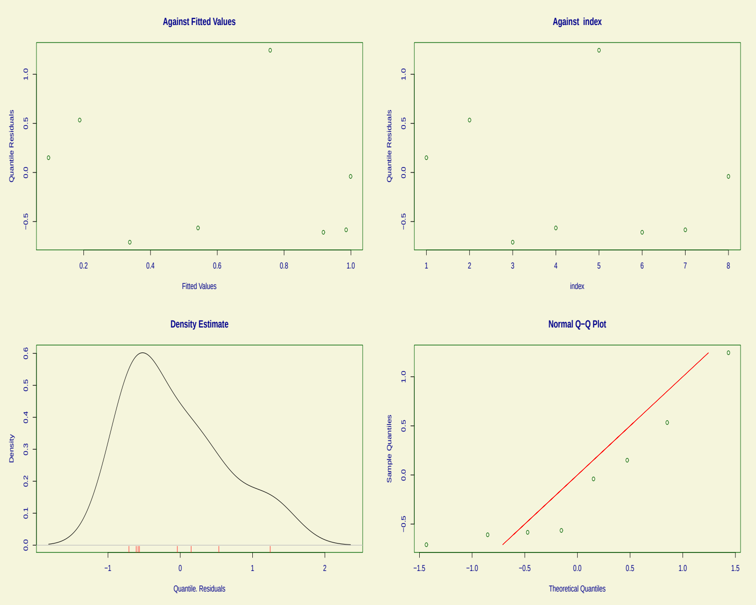
Curva Cloglog Ajustada



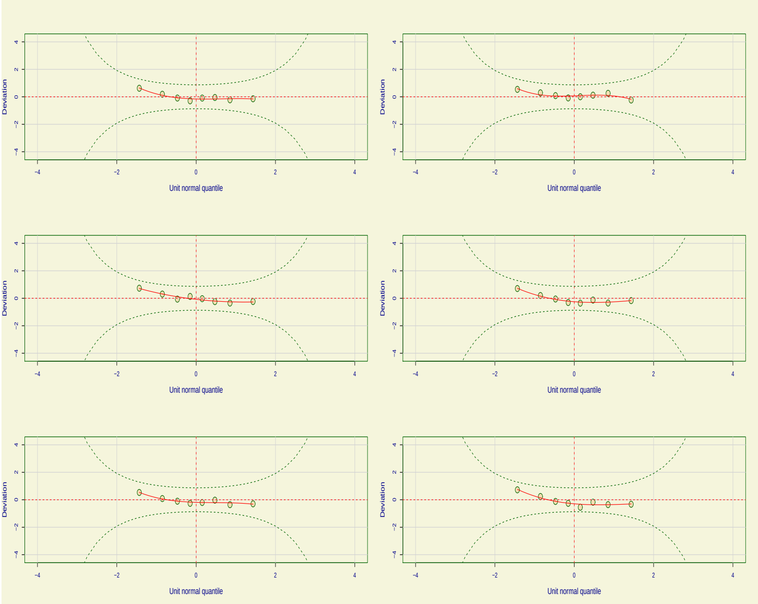
Resíduos Modelo Complementar Log-Log



Resíduos GAMLSS



Worm Plot GAMLSS



Sumário

Estimativa Pontual

A dose letal DL_{100p} num modelo binomial de dose resposta é obtida da expressão $g(p) = \beta_1 + \beta_2 DL_{100p}$. Logo, tem-se que

$$DL_{100p} = \frac{g(p) - \beta_1}{\beta_2},$$

em que $g(p)$ denota a função de ligação, que por exemplo pode ser logito, probito ou complementar log-log.

Estimativa Pontual

Em particular, a estimativa pontual para DL_{100p} sob o modelo binomial de dose resposta com ligação complementar loglog fica dada por

$$\widehat{DL}_{100p} = \frac{1}{\hat{\beta}_2}\left[\log\{-\log(1-p)\} - \hat{\beta}_1\right].$$

Variância Assintótica

A variância aproximada de $\widehat{\text{DL}}_{100p}$ fica dada por

$$\text{Var}_A[\widehat{\text{DL}}_{100p}] = \mathbf{D}(\boldsymbol{\beta})^\top(\mathbf{X}^\top\mathbf{W}\mathbf{X})^{-1}\mathbf{D}(\boldsymbol{\beta}),$$

em que **X** é a matriz modelo, **W** é uma matriz diagonal de pesos dados por $\omega_i = n_i\pi_i^{-1}(1-\pi_i)\{\log(1-\pi_i)\}^2$, para $i = 1, \ldots, k$, com

$$\mathbf{D}(\boldsymbol{\beta})^\top = (-\beta_2^{-1}, \beta_2^{-2}[\beta_1 - \log\{-\log(1-p)\}]).$$

Estimativa Pontual

$$
\begin{aligned}
\widehat{DL}_{50} &= \frac{1}{\hat{\beta}_2}\left[\log\{-\log(1-0,5)\} - \hat{\beta}_1\right] \\
&= \frac{1}{22,04}(-0,3665 + 39,57) \\
&= 1,779.
\end{aligned}
$$

Estimativa Intervalar

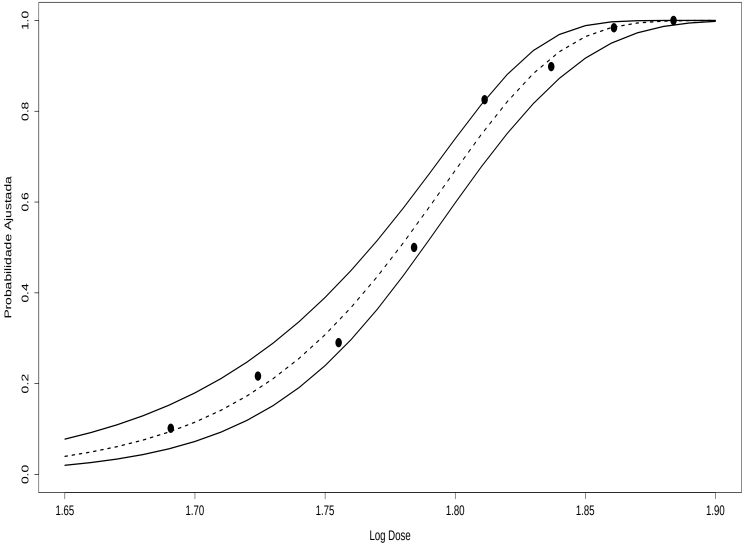
A variância aproximada de $\widehat{DL}_{50}$ fica dada por

$$\begin{aligned}\widehat{\text{Var}}(\widehat{DL}_{50}) &= (-0,0454, -0,0807)^T(\mathbf{X}^\top\widehat{\mathbf{W}}\mathbf{X})^{-1}\begin{pmatrix}-0,0454\\-0,0807\end{pmatrix}\\ &= 0,00001606.\end{aligned}$$

Logo, a estimativa intervalar de 95% para a dose letal fica dada por

$$\begin{aligned}&[1,779 \pm 1,96\sqrt{0,00001606}]\\ = &\ [1,771; 1,787].\end{aligned}$$

Banda de Confiança de 95%



Sumário

Considerações Finais

Neste exemplo, em que ajustamos a probabilidade de morte de besouros expostos a CS_2, nota-se que

- o modelo binomial logístico não se ajusta bem aos dados.
- O modelo binomial complementar log-log tem um ajuste muito superior.
- Essa superioridade do modelo complementar log-log deve-se ao fato do comportamento da proporção de mortes aumentar mais rapidamente após DL_{50}.
- Modelos binomiais de dose resposta com parâmetros na ligação (Aranda-Ordaz, 1981; Stukel, 1988) têm sido utilizados para ajustar esses dados.

Sumário

Referências

- Aranda-Ordaz, F.J.(1981). On two families of transformations to additivity for binary response data. *Biometrika* **68**, 357-364.
- Bliss, C.I.(1935). The calculation of the dosage-mortality curve. *Annals of Applied Biology* **22**, 134-167.
- Piegorsch, W. W. e Casella, G. (1988). Confidence bands for logistic regression with restricted predictor variables. *Biometrics* **44**, 739-750.
- Stukel, T.A.(1988). Generalized logistic models. *Journal of the American Statistical Association* **83**, 426-431.