

Modelos de regressão para dados discretos (parte 2): dados binários

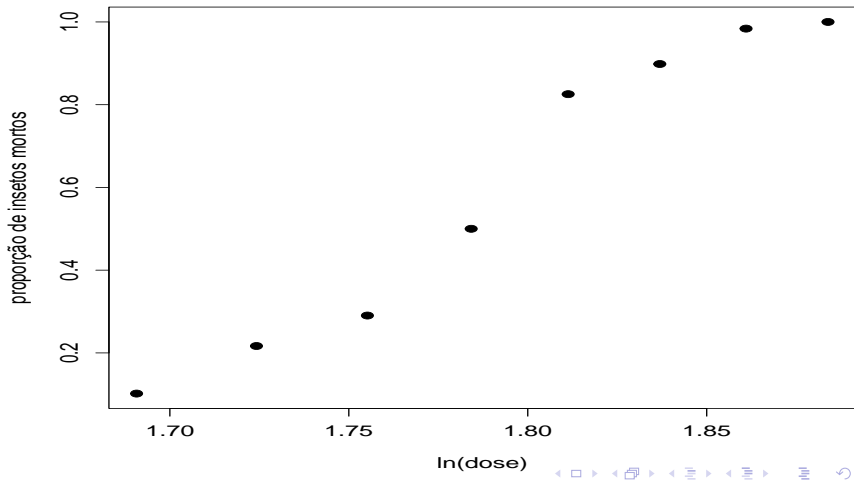
Prof. Caio Azevedo

Exemplo 8: mortalidade de besouros

- Dados relativos ao percentual de besouros mortos quando expostos à diferentes doses de disulfeto de carbono gasoso (CS_2).

Dose: $\log_{10} CS_2$	n° Besouros expostos	n° 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

Gráficos de dispersão



Modelo de regressão com ligação logito

$$Y_i \stackrel{ind.}{\sim} \text{binomial}(m_i, \mu_i)$$
$$\ln\left(\frac{\mu_i}{1 - \mu_i}\right) = \beta_0 + \beta_1(x_i - \bar{x}), \bar{x} = \frac{1}{8} \sum_{i=1}^8 x_i, i = 1, 2, \dots, 8 (n)$$

- m_i : número de besouros expostos à dose i de CS_2 .
- Y_i : número de besouros expostos à dose i de CS_2 que morreram.
- x_i : dose (log da concentração de CS_2) à que os besouros do grupo i foram expostos.

Cont.

- β_0 é o logito $\left(\ln \left(\frac{\mu_0}{1 - \mu_0} \right) \right)$ da proporção de besouros mortos submetidos à uma concentração igual à \bar{x} unidades de CS_2 . Ou seja, se $x_i = \frac{1}{8} \sum_{i=1}^8 \log_{10}(\text{concent}_i) = 1,793425$, então

$$\mu_0 = \frac{e^{\beta_0}}{1 + e^{\beta_0}}.$$

- Sejam: $\mu_i = \frac{e^{\beta_0 + \beta_1(x_i - \bar{x})}}{1 + e^{\beta_0 + \beta_1(x_i - \bar{x})}}$ e $\mu_{i+1} = \frac{e^{\beta_0 + \beta_1((x_i - \bar{x}) + 1)}}{1 + e^{\beta_0 + \beta_1((x_i - \bar{x}) + 1)}}$.

Cont.

- Assim:

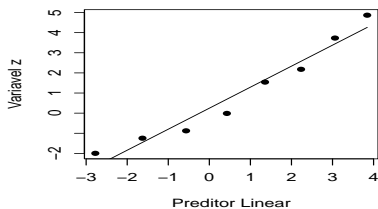
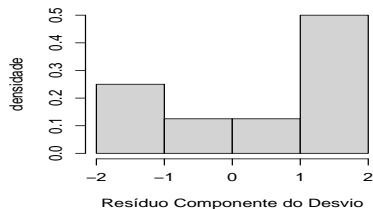
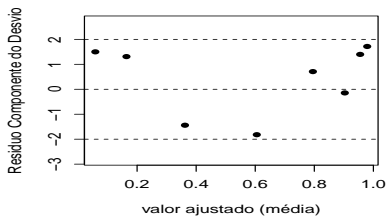
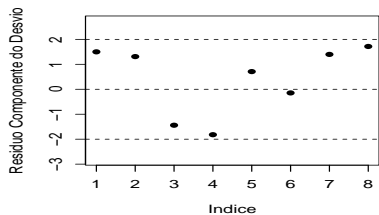
$$\ln\left(\frac{\mu_i}{1 - \mu_i}\right) = \beta_0 + \beta_1(x_i - \bar{x})$$

e

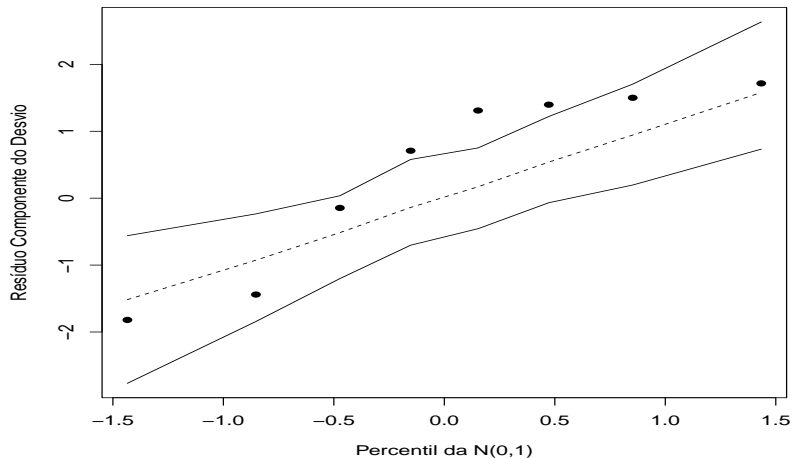
$$\ln\left(\frac{\mu_{i+1}}{1 - \mu_{i+1}}\right) = \beta_0 + \beta_1((x_i - \bar{x}) + 1).$$

- Logo: $\ln\left(\frac{\mu_{i+1}}{1 - \mu_{i+1}}\right) - \ln\left(\frac{\mu_i}{1 - \mu_i}\right) = \ln\left(\frac{\mu_{i+1}/(1 - \mu_{i+1})}{\mu_i/(1 - \mu_i)}\right) = \beta_1.$
- Portanto, $\frac{\mu_{i+1}/(1 - \mu_{i+1})}{\mu_i/(1 - \mu_i)} = e^{\beta_1}$ (razão de chances).

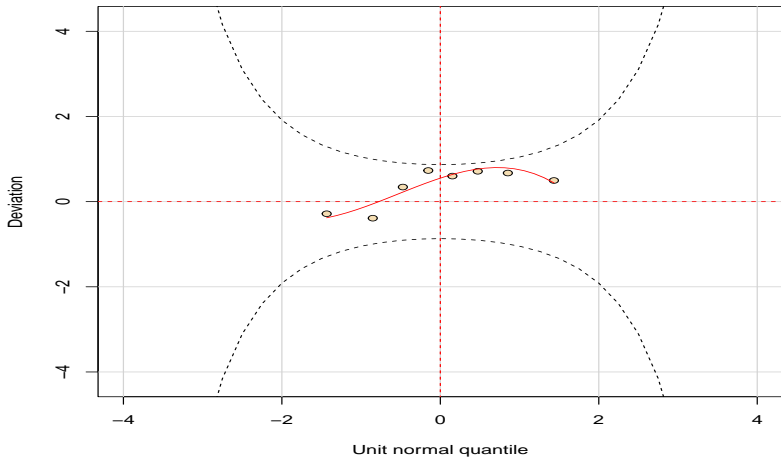
Gráficos de diagnóstico: ligação logito



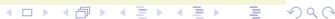
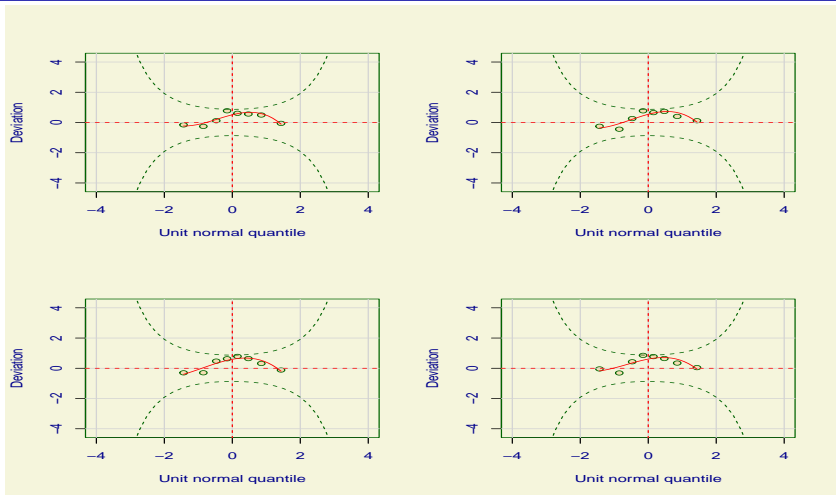
Envelope para os resíduos: ligação logito



Worm plot: ligação logito



Worm plots: ligação logito



Estimativas dos parâmetros do modelo logístico

Parâmetros	Estimativa	EP	IC(95%)	Estat. Z_t	p-valor
β_0	0,74	0,14	[0,47 ; 1,01]	5,40	< 0,0001
β_1	34,27	2,91	[28,56 ; 39,98]	11,77	0,0001

Todos os parâmetros são significativos. Além disso, $D(\mathbf{y}; \tilde{\boldsymbol{\mu}}) = 11,23$, para $k - p = 8 - 2 = 6$ graus de liberdade, o que leva à um p-valor = 0,08145, resultado que sugere um ajuste apenas razoável (considerando a aproximação pela distribuição $\chi^2_{(6)}$ razoável).

Estimativas das proporções de insetos mortos

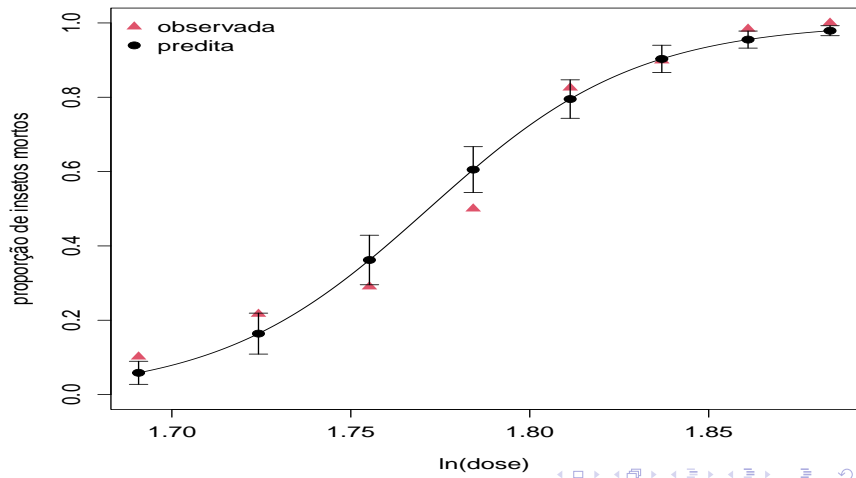
- A proporção de insetos mortos submetidos à dose x_i predita pelo modelo é dada por $\hat{\mu}_i = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1(x_i - \bar{x})}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1(x_i - \bar{x})}}$.
- Pelo método delta, para $m_i, i = 1, 2, \dots, 8$, suficientemente grandes, temos que $\hat{\mu}_i \approx N(\mu_i, \Psi_i \Sigma_{\beta} \Psi_i')$, em que

$$\Psi_i = \begin{bmatrix} \frac{\partial}{\partial \beta_0} \mu_i & \frac{\partial}{\partial \beta_1} \mu_i \end{bmatrix}$$

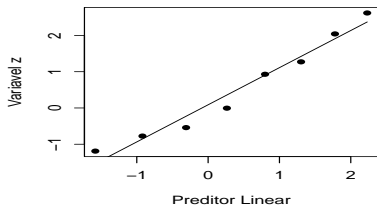
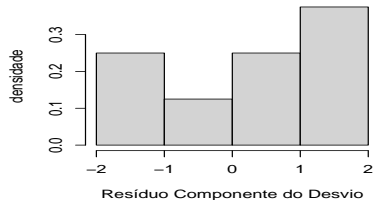
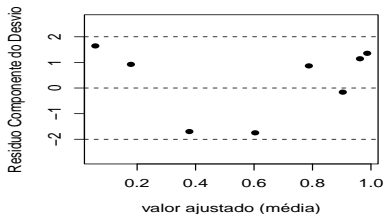
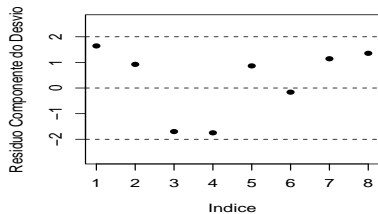
Estimativas das proporções de insetos mortos

- e $\mu_i = \frac{e^{\beta_0 + \beta_1(x_i - \bar{x})}}{1 + e^{\beta_0 + \beta_1(x_i - \bar{x})}}$. Pode-se provar que $\frac{\partial}{\partial \beta_0} \mu_i = \mu_i(1 - \mu_i)$ e $\frac{\partial}{\partial \beta_1} \mu_i = \mu_i(1 - \mu_i)x_i$ (**exercício**).
- Assim $IC(\mu_i, \gamma) = \left[\hat{\mu}_i - z_{(1+\gamma)/2} \sqrt{\hat{\psi}_i}; \hat{\mu}_i + z_{(1+\gamma)/2} \sqrt{\hat{\psi}_i} \right]$, em que $P(Z \geq z_{(1+\gamma)/2}) = \frac{1+\gamma}{2}$, $\hat{\psi}_i = \hat{\Psi}_i' \hat{\Sigma}_\beta \hat{\Psi}_i$ e $Z \sim N(0, 1)$ (lembrando que esse IC é assintótico).

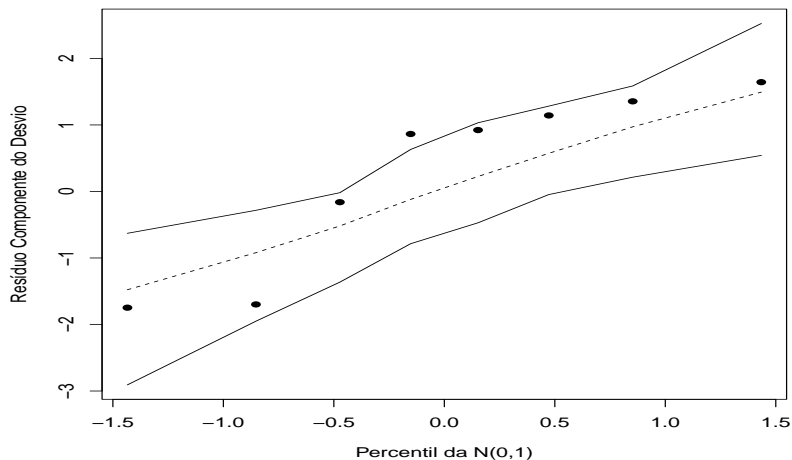
Proporções observadas \times proporções previstas pelo modelo



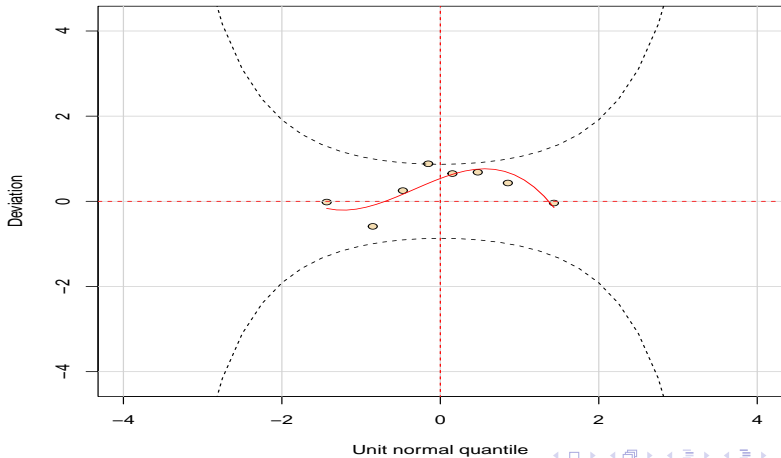
Gráficos de diagnóstico: ligação probito



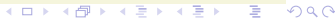
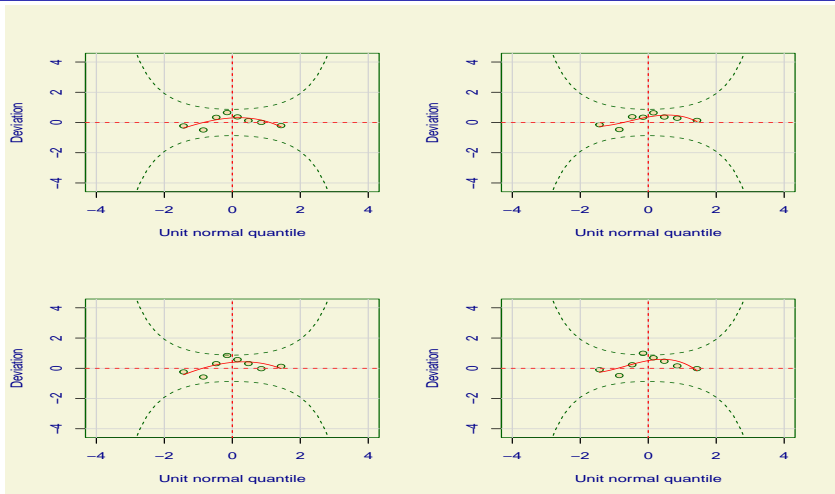
Envelope para os resíduos: ligação probito



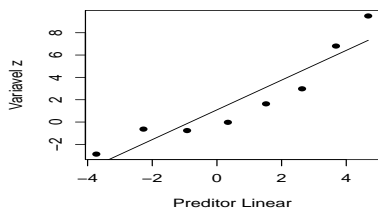
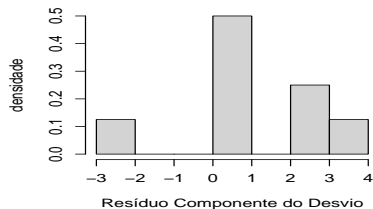
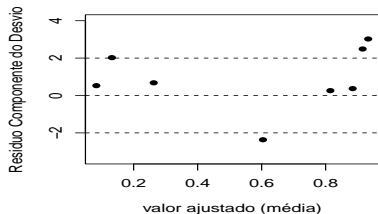
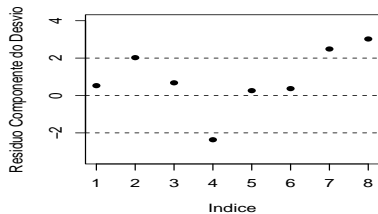
Worm plot: ligação probito



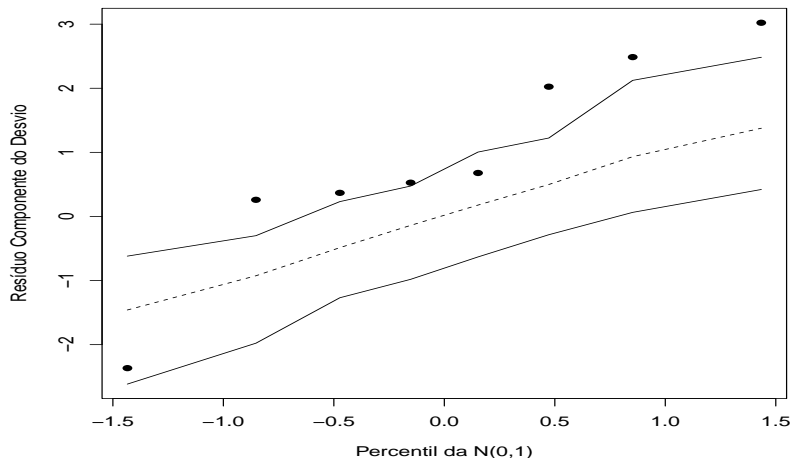
Worm plots: ligação probito



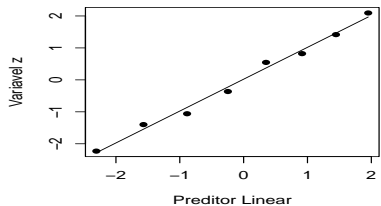
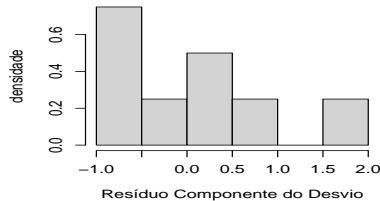
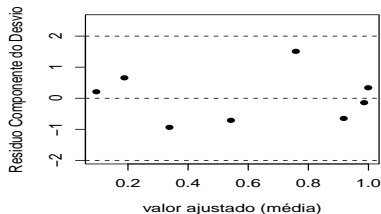
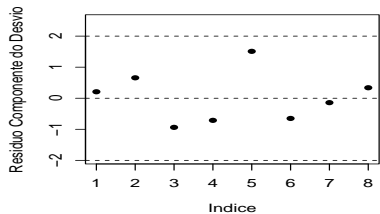
Gráficos de diagnóstico: ligação cauchito



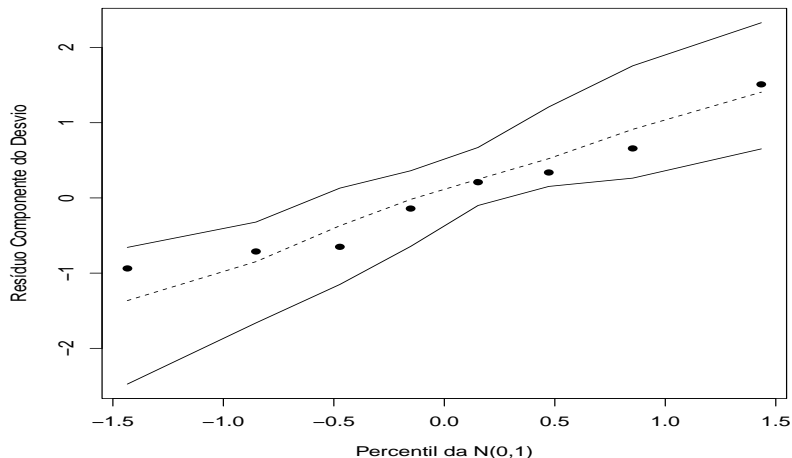
Envelope para os resíduos: ligação cauchito



Gráficos de diagnóstico: ligação complemento log-log



Envelope para os resíduos: ligação complemento log-log

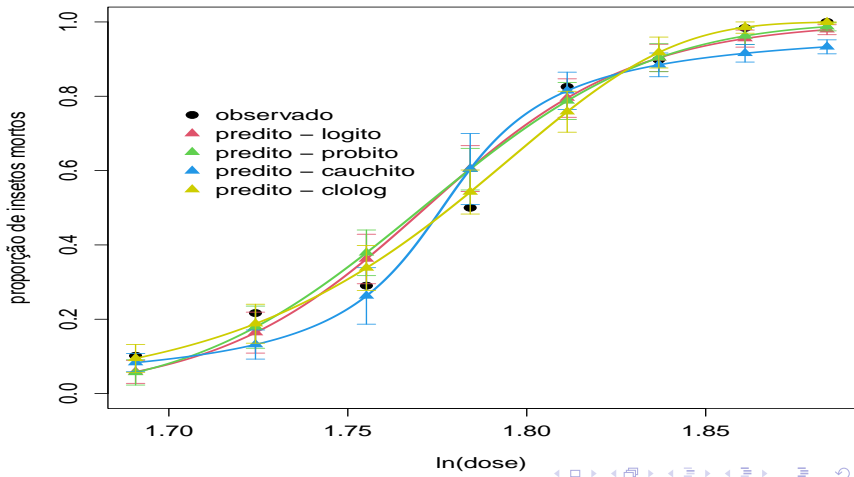


Comparação com outros modelos (funções de ligação)

Modelo	AIC	BIC	AICc	SABIC	HQCIC	CAIC	DABM	p-valor
logito	41,43	41,59	43,83	35,68	40,36	43,59	0,045	0,0815
probito	40,32	40,48	42,72	34,57	39,25	42,48	0,044	0,1197
cauchito	50,36	50,52	52,76	44,61	49,28	52,52	0,049	0,0026
cloglog	33,64	33,80	36,04	27,89	32,57	35,80	0,027	0,7511

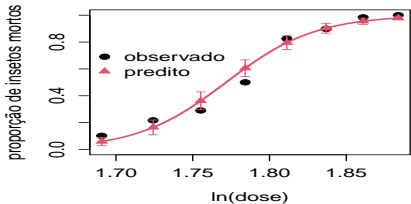
$$DABM = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i}{m_i} - \tilde{\mu}_i \right|.$$

Probab. preditas e observadas pelos modelos ajustados

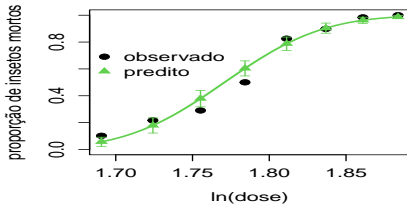


Probab. preditas e observadas pelos modelos ajustados

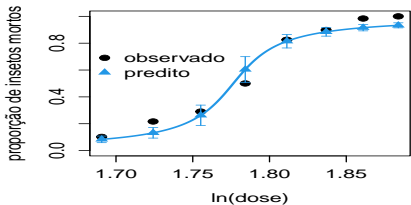
modelo logito



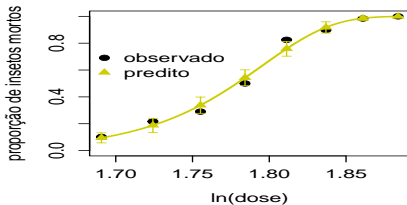
modelo probito



modelo cauchito



modelo cloglog



Estimativas dos parâmetros dos modelos

Modelo	Parâm.	Est.	EP	IC(95%)	Estat. Z_t	p-valor
logito	β_0	0,74	0,14	[0,47 ; 1,01]	5,40	<0,0001
	β_1	34,27	2,91	[28,56 ; 39,98]	11,77	<0,0001
probito	β_0	0,45	0,08	[0,30 ; 0,60]	5,81	<0,0001
	β_1	19,73	1,49	[16,81; 22,64]	13,27	<0,0001
cauchito	β_0	0,74	0,19	[0,37 ; 1,12]	3,87	0,0001
	β_1	43,53	6,38	[31,02 ; 56,03]	6,82	<0,0001
cloglog	β_0	-0,04	0,08	[-0,20 ; 0,11]	-0,54	0,5914
	β_1	22,04	1,80	[18,51 ; 25,57]	12,25	<0,0001

Estimativas da dose letal

- Sabemos que a probabilidade (proporção) estimada de insetos mortos para uma certa dose (x) é dada por $\tilde{\mu} = F(\tilde{\beta}_0 + \tilde{\beta}_1 x)$.
- A notação usual para uma dose letal de 100p% é DL_{100p} . Dessa forma a proporção de sucessos para essa dose é dada por $\tilde{p} = F(\tilde{\beta}_0 + \tilde{\beta}_1 DL_{100p})$.
- Portanto o estimador para DL_{100p} é dado por
$$\widehat{DL}_{100p} = \frac{1}{\widehat{\beta}_1} \left[F^{-1}(p) - \widehat{\beta}_0 \right].$$

Estimativas da dose letal

- Para cada um dos quatro modelos, temos:

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

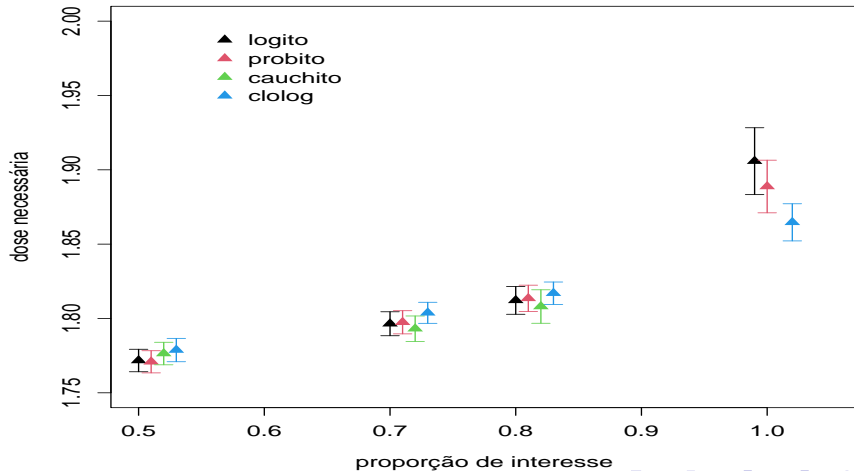
- $\widehat{DL}_{100p} = \frac{1}{\widehat{\beta}_2} \left[\Phi^{-1}(p) - \widehat{\beta}_1 \right].$

- $\widehat{DL}_{100p} = \frac{1}{\widehat{\beta}_2} \left[\tan \left(\pi \left(p - \frac{1}{2} \right) \right) - \widehat{\beta}_1 \right].$

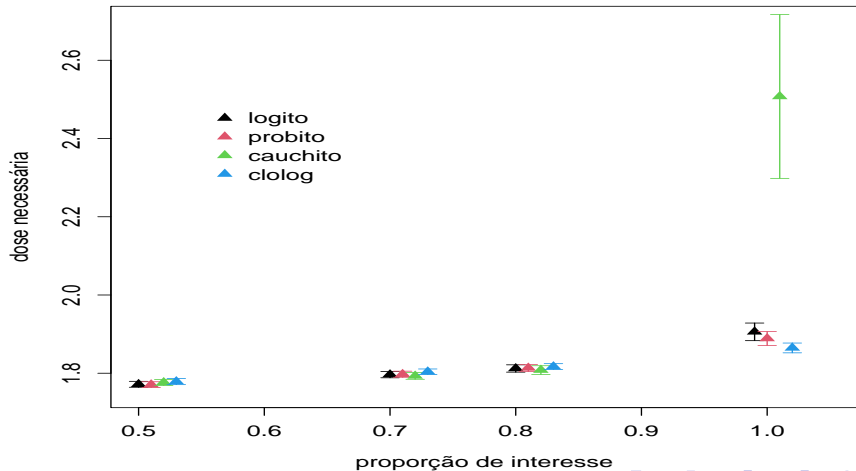
- $\widehat{DL}_{100p} = \frac{1}{\widehat{\beta}_2} \left[\ln(-\ln(1-p)) - \widehat{\beta}_1 \right].$

- Exercício: obter a distribuição assintótica do emv da dose letal para cada modelo.

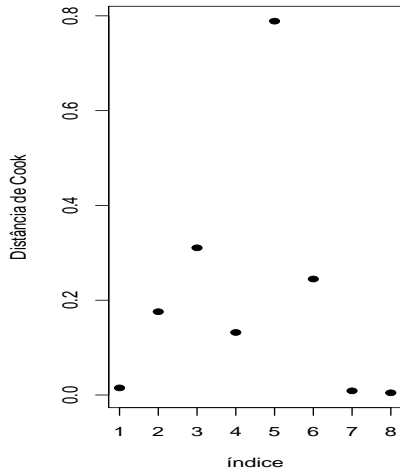
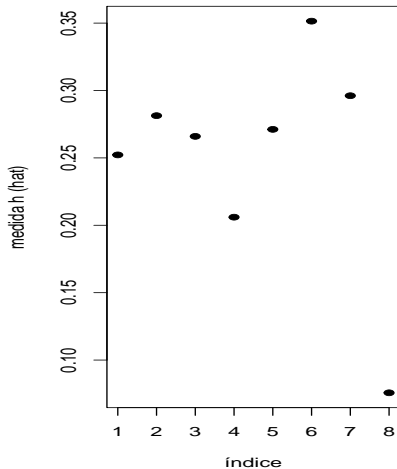
Estimativas da dose letal



Estimativas da dose letal



Análise de Influência



Comentários

- O modelo cloglog apresentou o melhor ajuste sob todos os critérios: resíduos, análise preditiva e critérios de informação, como esperado.
- As estimativas do impacto da dose de CS_2 diferem entre os modelos, bem como as estimativas das doses letais, para uma mesma proporção.
- O modelo cloglog apresentou um ajuste bastante razoável.
- Não há necessidade de se avaliar a sensibilidade à retirada de nenhum ponto.