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

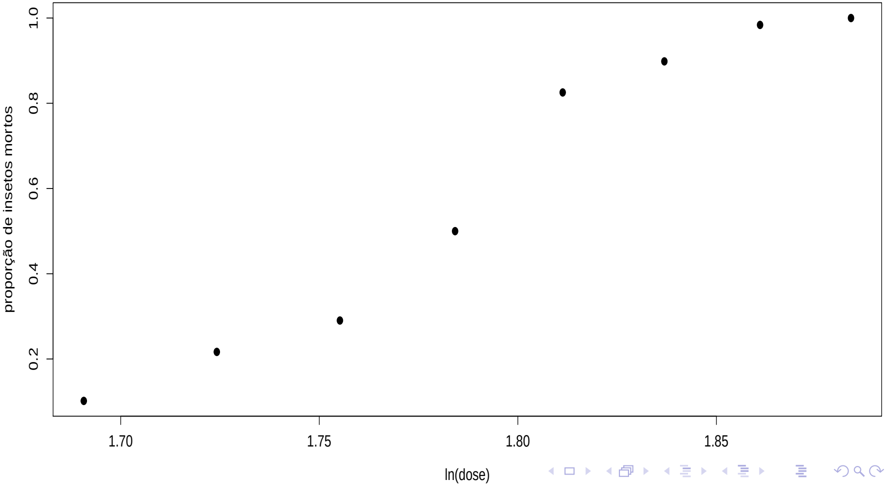
Prof. Caio Azevedo

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^o Besouros expostos	n^o 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

$$
\begin{aligned}
Y_i &\overset{ind.}{\sim} \text{binomial}(m_i, \mu_i) \\
\ln\left(\frac{\mu_i}{1-\mu_i}\right) &= \beta_0 + \beta_1(x_i - \overline{x}), \overline{x} = \frac{1}{8}\sum_{i=1}^{8} x_i, i = 1, 2, ..., 8
\end{aligned}
$$

- 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.
- 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 $\overline{x}$ unidades de CS_2. Ou seja, se $x_i = \frac{1}{8}\sum_{i=1}^{8} log_{10}(concent_i) = 1,793425$, então $\mu_0 = \dfrac{e^{\beta_0}}{1+e^{\beta_0}}$.

Cont.

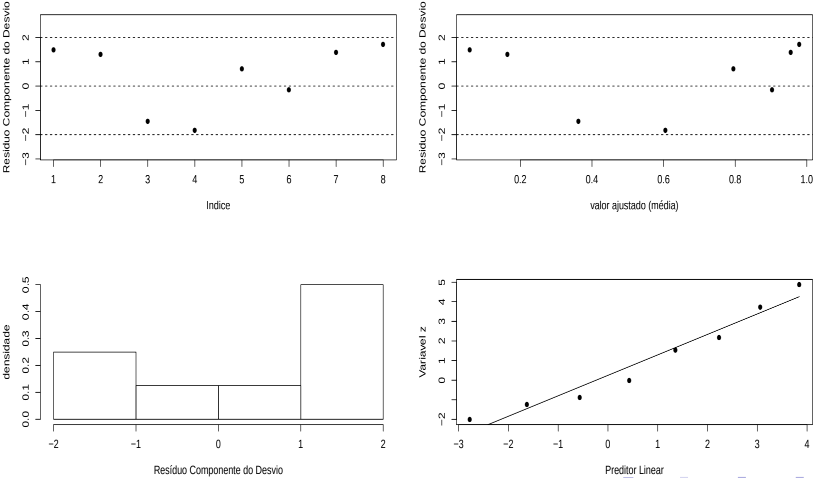
- Sejam: $\mu_i = \dfrac{e^{\beta_0+\beta_1(x_i-\overline{x})}}{1+e^{\beta_0+\beta_1(x_i-\overline{x})}}$ e $\mu_{i+1} = \dfrac{e^{\beta_0+\beta_1((x_i-\overline{x})+1)}}{1+e^{\beta_0+\beta_1((x_i-\overline{x})+1)}}$.
- Assim: $\ln\left(\dfrac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1(x_i - \overline{x})$ e
  $\ln\left(\dfrac{\mu_{i+1}}{1-\mu_{i+1}}\right) = \beta_0 + \beta_1((x_i - \overline{x}) + 1)$.
- Logo: $\ln\left(\dfrac{\mu_{i+1}}{1-\mu_{i+1}}\right) - \ln\left(\dfrac{\mu_i}{1-\mu_i}\right) = \ln\left(\dfrac{\mu_{i+1}/(1-\mu_{i+1})}{\mu_i/(1-\mu_i)}\right) = \beta_1$.
- Portanto, $\dfrac{\mu_{i+1}/(1-\mu_{i+1})}{\mu_i/(1-\mu_i)} = e^{\beta_1}$ (razão de chances).

Estimativas dos parâmetros do modelo logístico

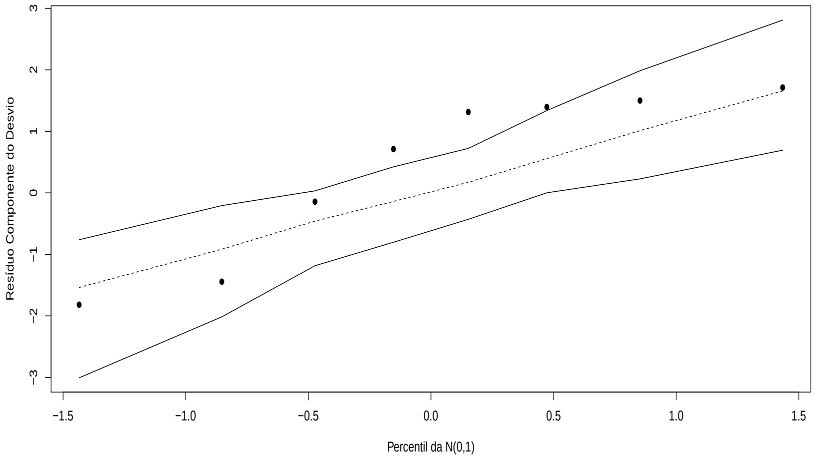
Parâmetros	Estimativa	EP	IC(95%)	Estat. Z_t	p-valor
β_0	0,74	0,14	[0,47 ; 1,01]	5,40	$<$ 0,0001
β_1	34,27	2,91	[28,56 ; 39,98]	11,77	0,0001

Todos os parâmetros são significativos. Além disso, $D(\boldsymbol{y};\widetilde{\boldsymbol{\mu}}) = 11,23$, 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).

Gráficos de diagnóstico: ligação logito



Envelope para os resíduos: ligação logito



Estimativas das proporções de insetos mortos

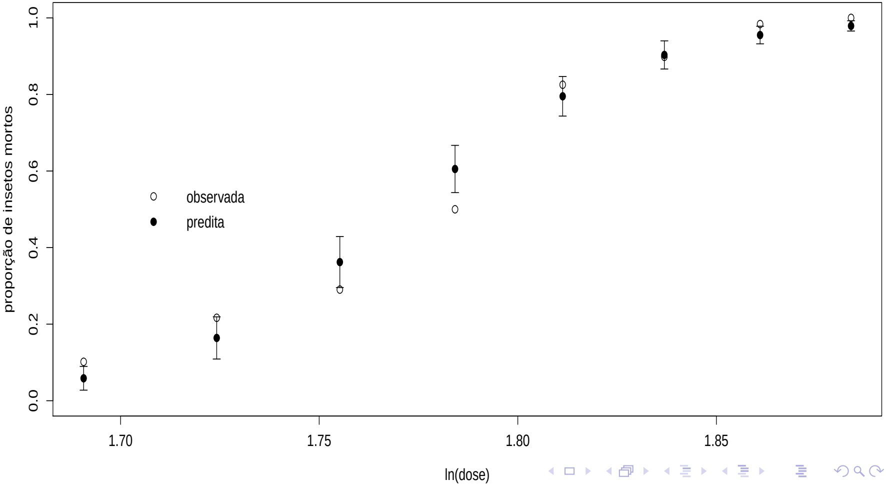
- A proporção de insetos mortos submetidos à dose x_i predita pelo modelo é dada por $\widehat{\mu}_i = \dfrac{e^{\widehat{\beta}_0+\widehat{\beta}_1(x_i-\overline{x})}}{1+e^{\widehat{\beta}_0+\widehat{\beta}_1(x_i-\overline{x})}}$.
- Pelo método delta, para $m_i, i = 1, 2, ..., 8$, suficientemente grandes,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

$$\boldsymbol{\Psi}_i = \left[\begin{array}{cc} \dfrac{\partial}{\partial \beta_0}\mu_i & \dfrac{\partial}{\partial \beta_1}\mu_i \end{array} \right]$$

Estimativas das proporções de insetos mortos

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

Proporções observadas × proporções preditas pelo modelo

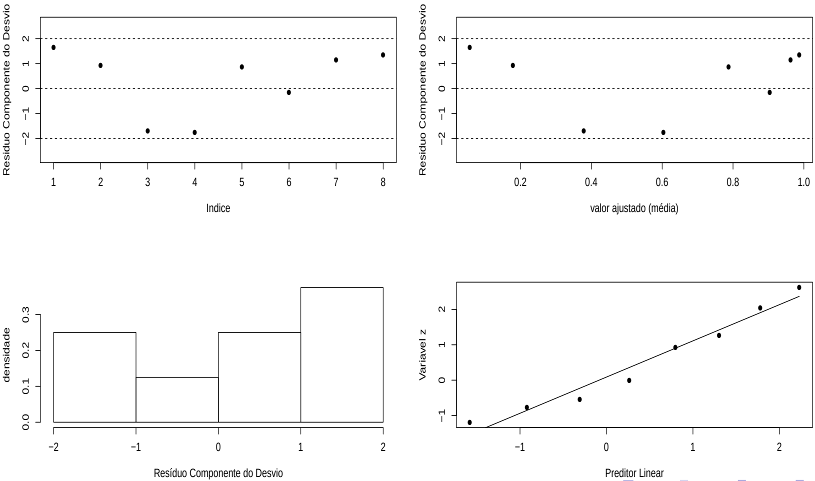


Comparação com outros modelos (funções de ligação)

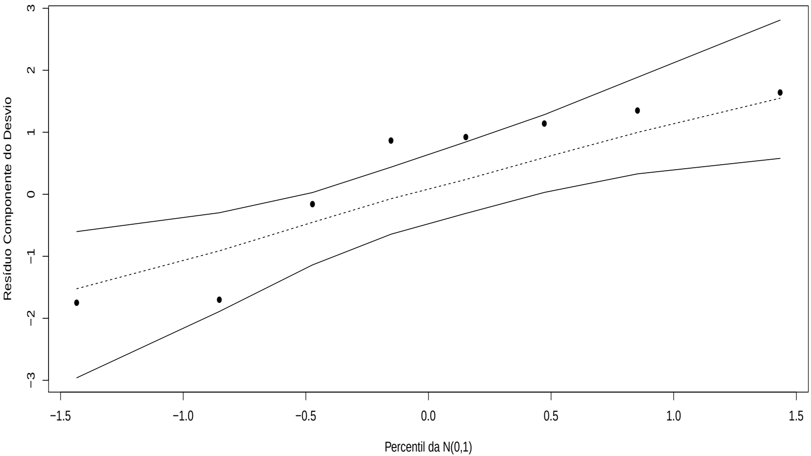
Função de ligação	AIC	BIC	DABM	Desvio	p-valor (desvio)
logito	41,43	41,59	0,04	11,23	0,0815
probito	40,32	40,48	0,04	10,12	0,1197
cauchito	50,36	50,52	0,05	20,16	0,0026
cloglog	33,64	33,80	0,03	3,45	0,7511

$$DABM = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widetilde{\mu}_i|$$

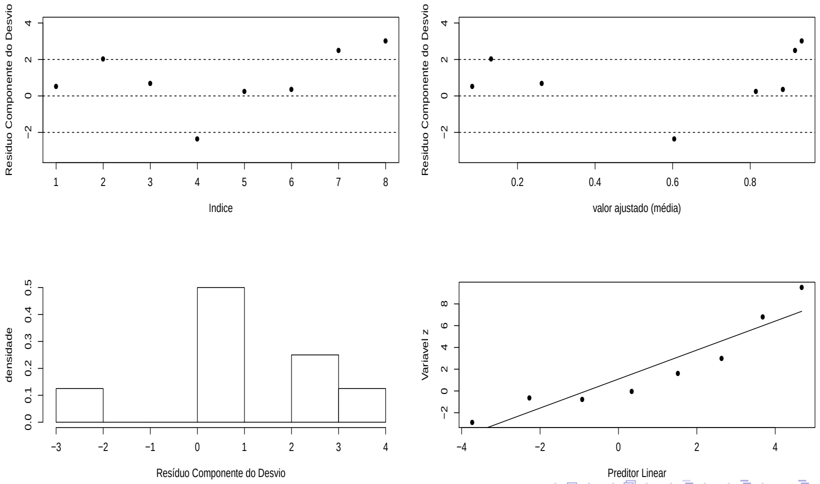
Gráficos de diagnóstico: ligação probito



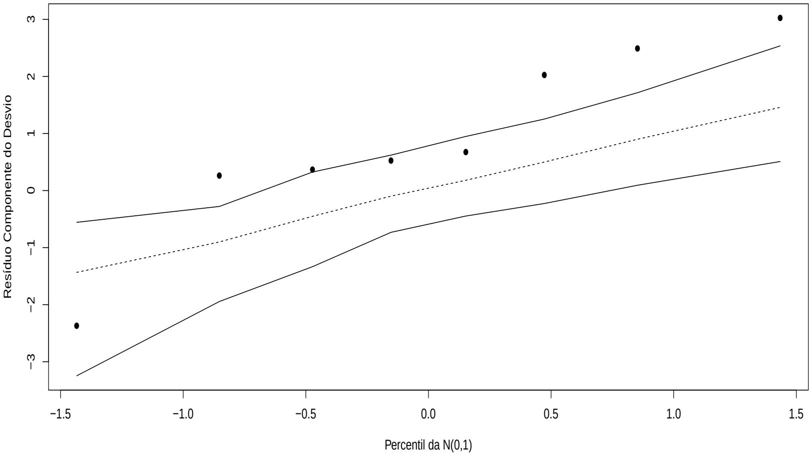
Envelope para os resíduos: ligação probito



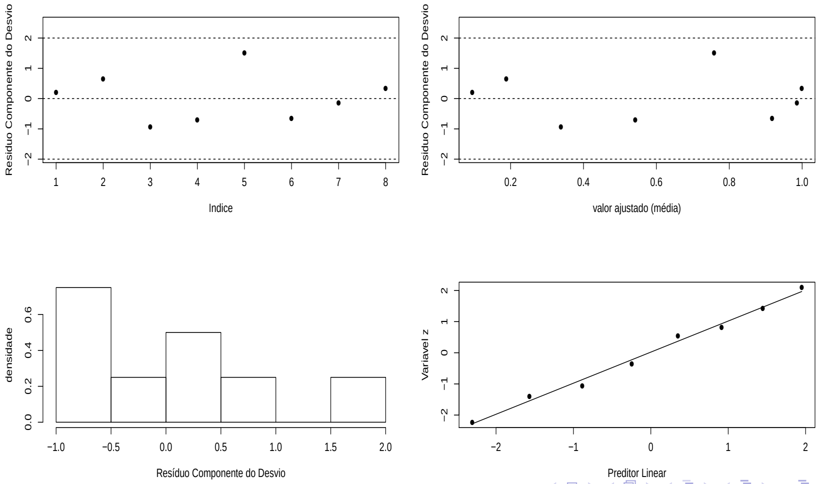
Gráficos de diagnóstico: ligação cauchito



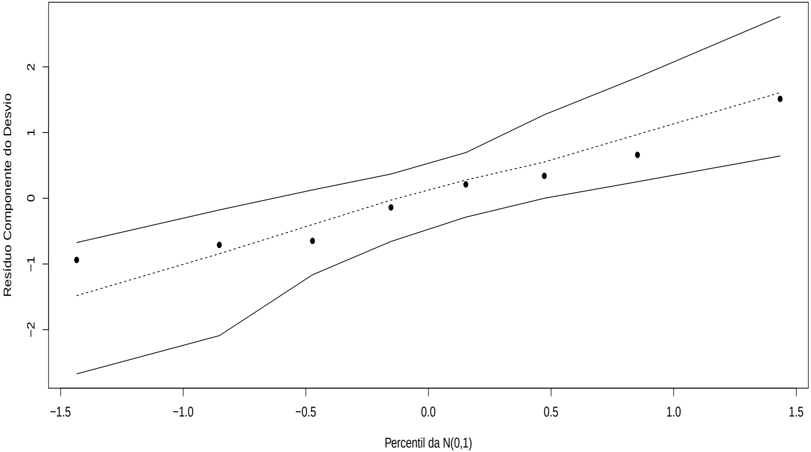
Envelope para os resíduos: ligação cauchito



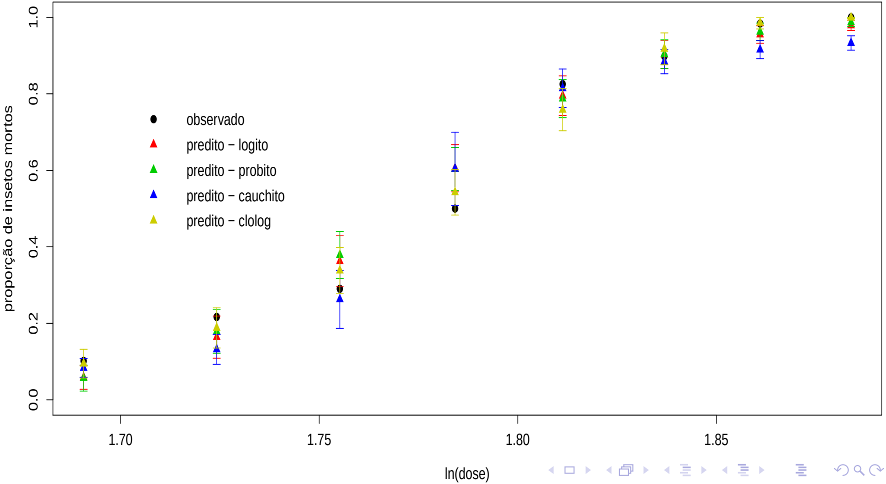
Gráficos de diagnóstico: ligação complemento log-log



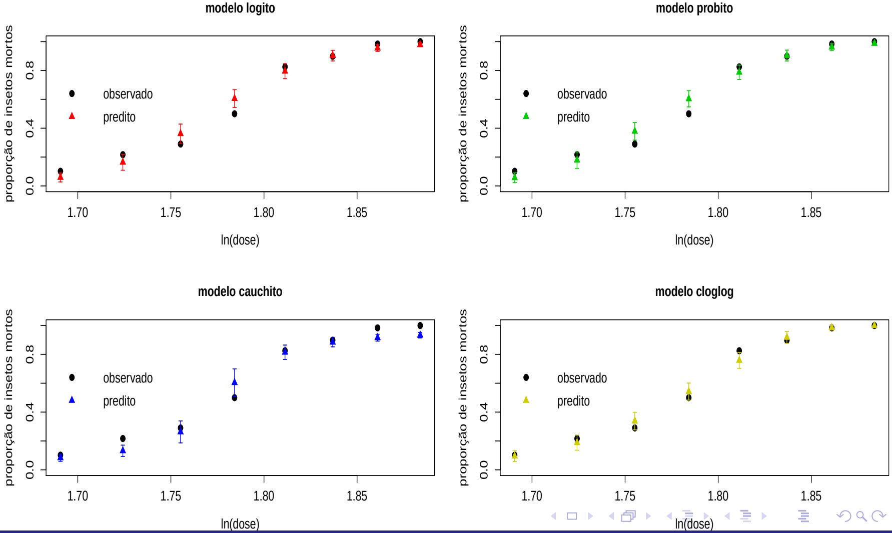
Envelope para os resíduos: ligação complemento log-log



Probab. preditas e observadas pelos modelos ajustados



Probab. preditas e observadas pelos modelos ajustados



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
c, log-log	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

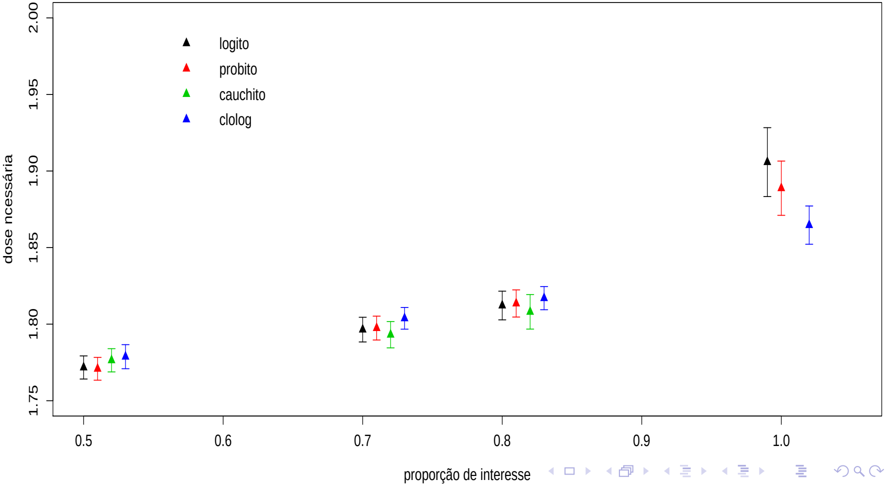
Estimativas da dose letal

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

Estimativas da dose letal

- Para cada um dos quatro modelos, temos:
  - $\widehat{DL}_{100p} = \frac{1}{\widehat{\beta}_2}\left[\ln\left(\frac{p}{1-p}\right) - \widehat{\beta}_1\right].$
  - $\widehat{DL}_{100p} = \frac{1}{\widehat{\beta}_2}\left[\Phi^{-1}(p) - \widehat{\beta}_1\right].$
  - $\widehat{DL}_{100p} = \frac{1}{\widehat{\beta}_2}\left[\tan\left(\pi\left(p - \frac{1}{2}\right)\right) - \widehat{\beta}_1\right].$
  - $\widehat{DL}_{100p} = \frac{1}{\widehat{\beta}_2}\left[\ln\left(-\ln\left(1-p\right)\right) - \widehat{\beta}_1\right].$
- Exercício: obter a distribuição assintótica do emv da dose letal.

Estimativas da dose letal



Estimativas da dose letal