

# INLA - “Integrated Nested Laplace Approximation” para inferência bayesiana (parte 2)

Prof. Caio Azevedo

# Estudo sobre vasoconstrição

- Dados sobre um estudo de vasoconstrição (veja [Paula \(2013\)](#), [Finney \(1978\)](#) e [Pregibon \(1981\)](#)).
- Nesse estudo, foram medidos de 3 pacientes o volume e a razão de ar inspirado, como também a ocorrência ou não de vasoconstrição (contração de vasos sanguíneos) na pele dos dedos da mão. O primeiro paciente contribuiu com 9 observações, o segundo com 8 e o terceiro com 22.
- Em princípio, não seria razoável assumir independência entre as observações. Contudo, por enquanto, assumiremos independência (metodologias mais apropriadas: modelos mistos, modelos hierárquicos).

# Modelo de regressão para os dados de vasoconstricção

$$Y_i | (\beta_0, \beta_1, \beta_2) \stackrel{ind.}{\sim} \text{Bernoulli}(\mu_i)$$
$$\text{logito}(\mu_i) = \ln \left( \frac{\mu_i}{1 - \mu_i} \right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$$
$$\rightarrow \mu_i = \frac{e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}}}, i = 1, 2, \dots, n$$

- $Y_i$  : ocorrência (1) ou não (0) de vaso constricção.
- $x_{1i}$  : logaritmo natural do volume de ar inspirado da  $i$ -ésima observação;  $x_{2i}$  : logaritmo natural da razão de ar inspirado da  $i$ -ésima observação.

# Modelo de regressão para os dados de vasoconstricção

- $F(\cdot)$  : corresponde à fda de uma distribuição logística padrão (portanto o nome regressão logística). Nesse caso, o  $\text{logito}(\cdot)$  é a função de ligação.
- Interpretação dos parâmetros. Defina  $l(\mu_i) = \text{logito}(\mu_i)$ .
- Se  $x_{1j} = x_{2j} = 0$ , então  $\mu_i = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$ .
- Defina  $l_1(\mu_{i+1}) = \beta_0 + \beta_1(x_{1i} + 1) + \beta_2 x_{2i}$  e  $l_1(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$ . Então  $l_1(\mu_{i+1}) - l_1(\mu_i) = \beta_1 \rightarrow \frac{\mu_{i+1}/(1 - \mu_{i+1})}{\mu_i/(1 - \mu_i)} = e^{\beta_1}$  (**razão de chances em relação à primeira covariável**).

# Um modelo de regressão para os dados de vasoconstricção

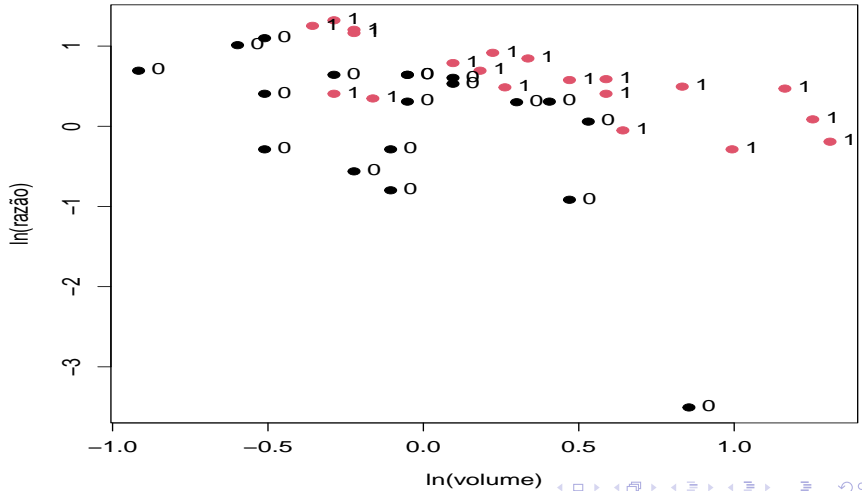
- Analogamente, defina  $l_2(\mu_{i+1}) = \beta_0 + \beta_1 x_{1i} + \beta_2(x_{2i} + 1)$  e  $l_2(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$ . Então  $l_2(\mu_{i+1}) - l_2(\mu_i) = \beta_2 \rightarrow \frac{\mu_{i+1}/(1 - \mu_{i+1})}{\mu_i/(1 - \mu_i)} = e^{\beta_2}$  (**razão de chances em relação à segunda covariável**).

- Repetir os desenvolvimentos considerando as covariáveis padronizadas, ou seja  $x_{ij}^* = \frac{x_{ji} - \bar{x}_j}{s_j}$ , em que  $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ji}$  e

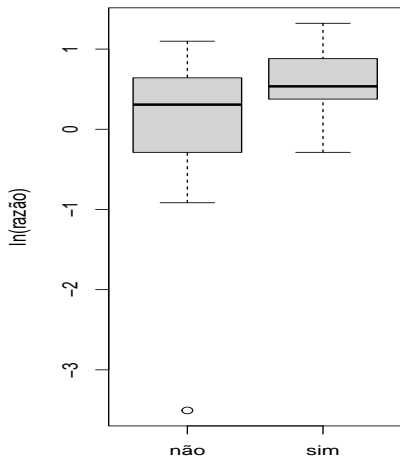
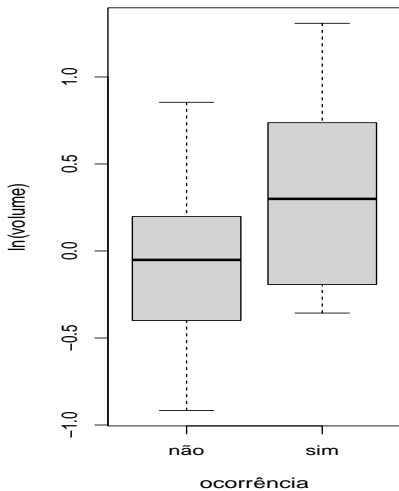
$$s_j = \sqrt{s_j^2}, \quad s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ji} - \bar{x}_j)^2.$$

- Com efeito, utilizaremos as duas covariáveis padronizadas, a fim de permitir que seus coeficientes sejam diretamente comparáveis.

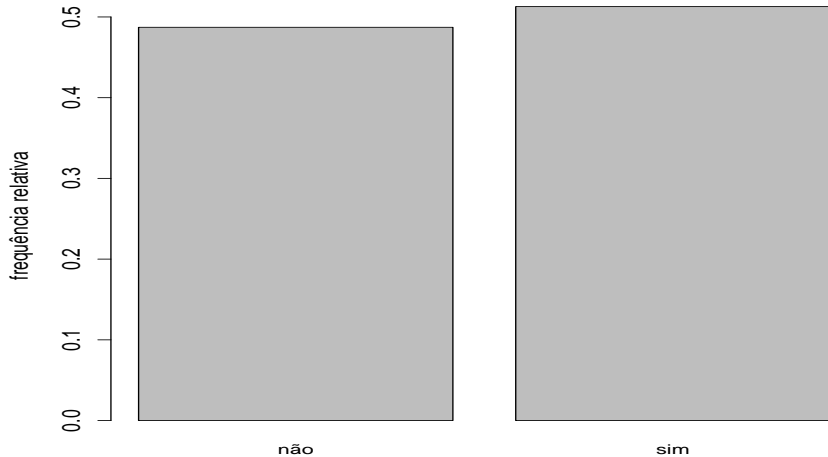
# Gráficos de dispersão: $\ln(\text{razão}) \times \ln(\text{volume})$



# Box-plots das variáveis $\ln(\text{razão}) \times \ln(\text{volume})$



# Frequências relativas de vaso-constricção





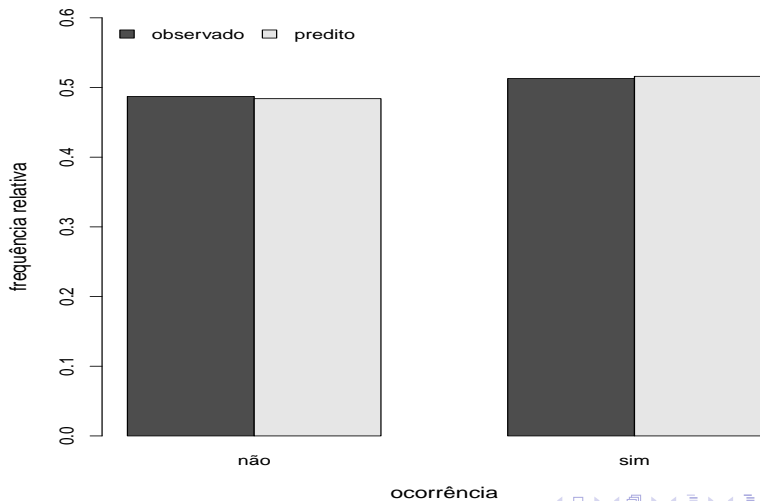
## Medidas resumo $\ln(\text{razão})$ e $\ln(\text{volume})$

Medida resumo	$\ln(\text{volume})$		$\ln(\text{razão})$	
	Resposta			
	Não	Sim	Não	Sim
Média	-0,06	0,37	0,05	0,58
Mediana	-0,05	0,30	0,31	0,54
DP	0,45	0,54	1,03	0,46
Var.	0,20	0,29	1,07	0,22
$ \text{CV}(\%) $	723,00	147,00	2223,00	81,00
Min.	-0,92	-0,36	-3,51	-0,29
Max.	0,85	1,31	1,10	1,30

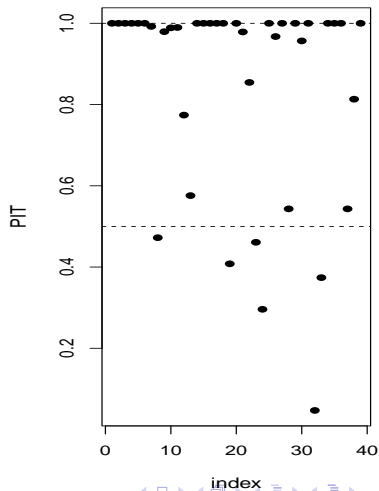
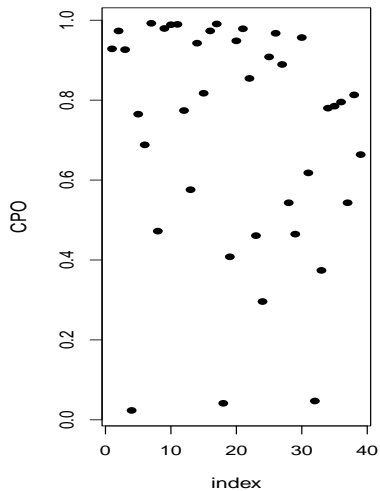
# Comentários

- Pelos resultados anteriores vemos que individual e conjuntamente o  $\log(\text{volume})$  e o  $\log(\text{razão})$  afetam, positivamente, a ocorrência de vaso constrição.
- As frequências relativas observadas (sem considerar as covariáveis) de ocorrência e não ocorrência foram próximas.

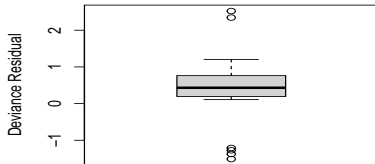
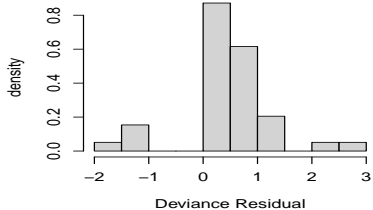
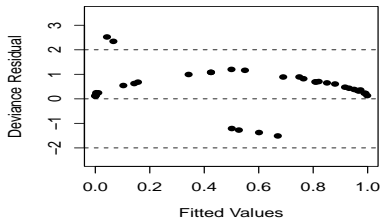
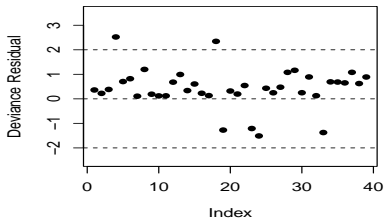
# Dados observados e distribuição preditiva (logito)



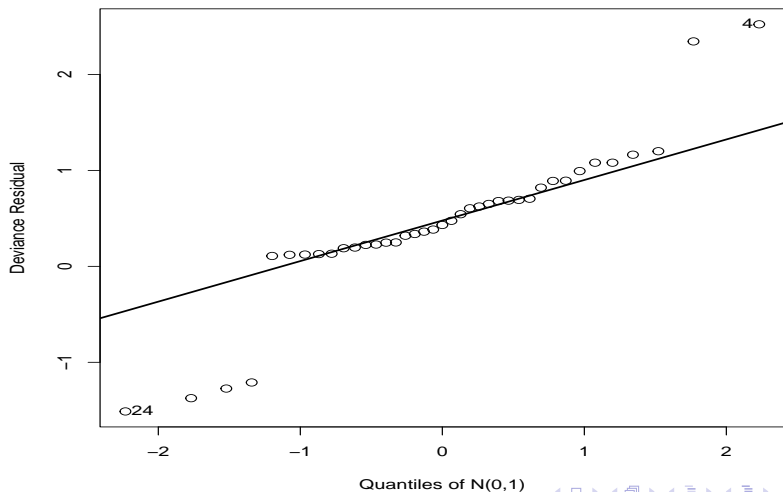
# CPO e PIT (logito)



# Diagnóstico dos resíduos (logito)



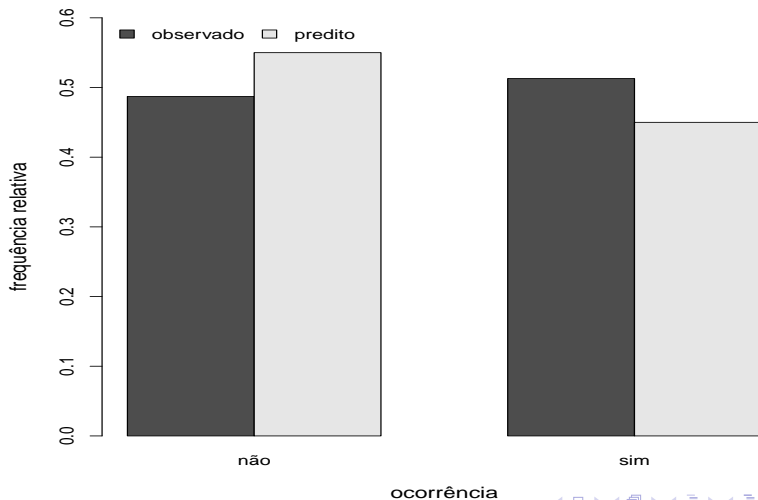
# QQplot dos resíduos (logito)



# Comentários

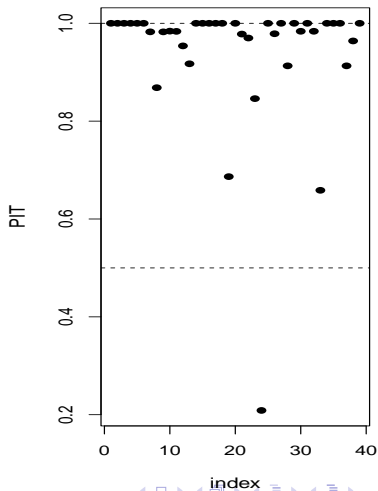
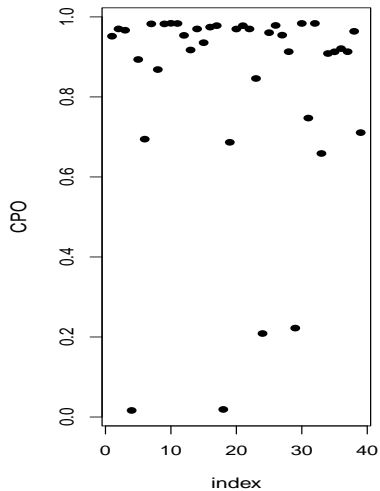
- O modelo logito não se adequou bem aos dados, com exceção da predição da distribuição daqueles.
- Devido ao comportamento dos resíduos - existência de vários outliers e uma assimetria (talvez provocada pela presença de outliers), devemos tentar outras funções de ligação: com caudas pesadas ou caudas pesadas e assimétrica (veja [aqui](#)).
- Devido à [uma análise anterior](#) e as opções de escolha de [função de ligação do INLA](#) (`names(inla.models())$link`), utilizarmos como alternativa o **modelo cauchito**: utilizar a função de distribuição acumulada da Cauchy(0,1) (padrão), ao invés da fda da distribuição logística padrão ([aqui](#)).

# Dados observados e distribuição preditiva (cauchito)

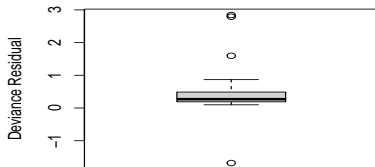
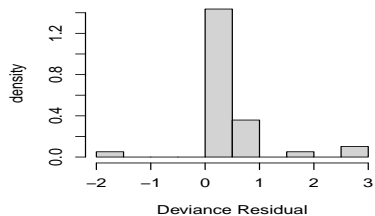
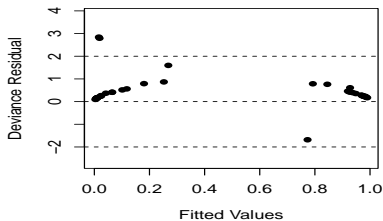
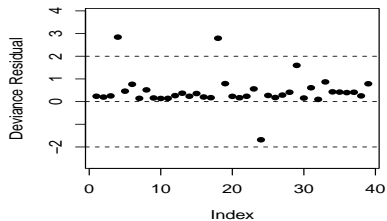




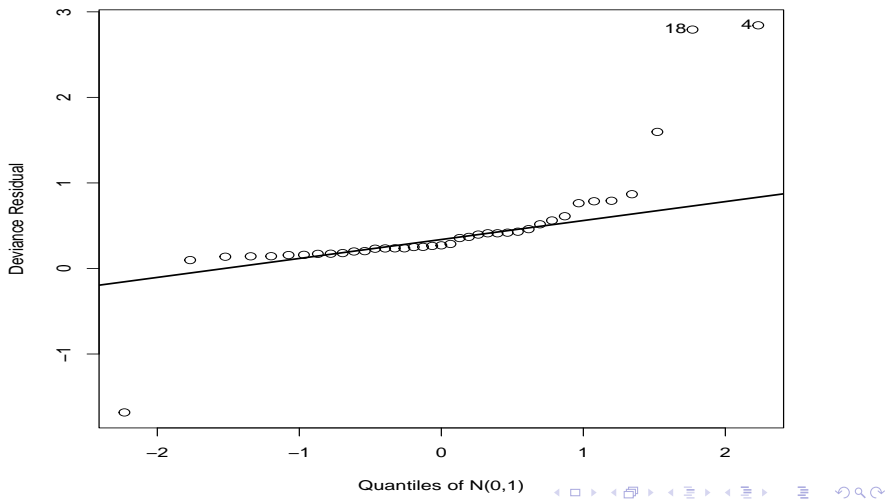
# CPO e PIT (cauchito)



# Diagnóstico dos resíduos (cauchito)



# QQplot dos resíduos (cauchito)



# Cr terios de informa o e predi es individuais

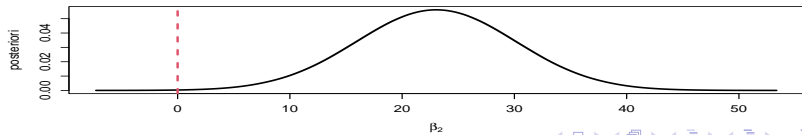
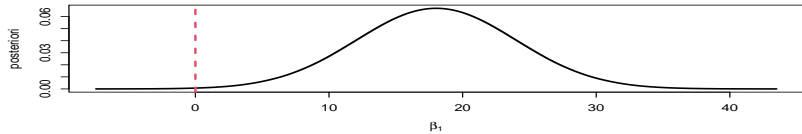
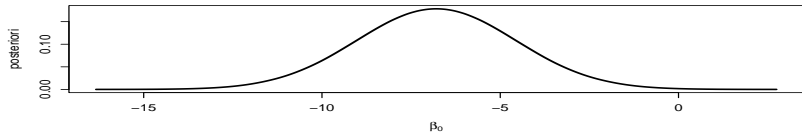
Modelo	DIC	WAIC	CPO	MLIK
logito	34,77	35,76	20,92	-21,50
cauchito	28,26	28,78	14,53	-16,75

	modelo (predito)			
	logito		cauchito	
	N�o	Sim	N�o	Sim
N�o	0,79	0,21	0,95	0,05
Sim	0,10	0,90	0,15	0,85

# Comentários

- O modelo cauchito se adequou de modo razoável (e melhor do que o modelo logito) aos dados, embora a predição da distribuição não tenha sido tão boa quanto à do modelo logito.
- No entanto, a predição das observações individuais foi, em geral, melhor.
- Sugestão: utilizar um link correspondente à fda de uma distribuição [Skew t](#) (devido ao fato do ajuste do modelo ainda não estar satisfatório, com a presença de uma leve assimetria positiva e caudas pesadas).

# Resultados inferencias (cauchito)



# Resultados inferencias

Mod.	Parâm.	EAP	MeAP	DPAP	IC(95%)	HPD(95%)
logito	$\beta_0$	-0,78	-0,78	0,61	[-1,99 ; 0,42]	[-1,99 ; 0,42]
	$\beta_1$	3,38	3,38	1,00	[1,43 ; 5,34]	[1,43 ; 5,34]
	$\beta_2$	4,61	4,61	1,52	[1,64 ; 7,58]	[1,64 ; 7,58]
cauchito	$\beta_0$	-6,79	-6,80	2,24	[-11,18 ; -2,42]	[-11,18 ; -2,42]
	$\beta_1$	18,03	18,01	5,97	[6,32 ; 29,71]	[6,32 ; 29,71]
	$\beta_2$	23,04	23,02	7,10	[9,10 ; 36,94]	[9,10 ; 36,94]

# Resultados inferencias

- Os resultados inferenciais são substancialmente diferentes entre os modelos logito e cauchito.
- Vemos que o impacto do  $\log(\text{razão})$  é maior (aproximadamente, 1,27 vezes) do que o impacto do  $\log(\text{volume})$ , sendo ambos positivos, na ocorrência de vasoconstricção.
- Exercício: Pesquisar sobre a razão de chances no modelo cauchito.



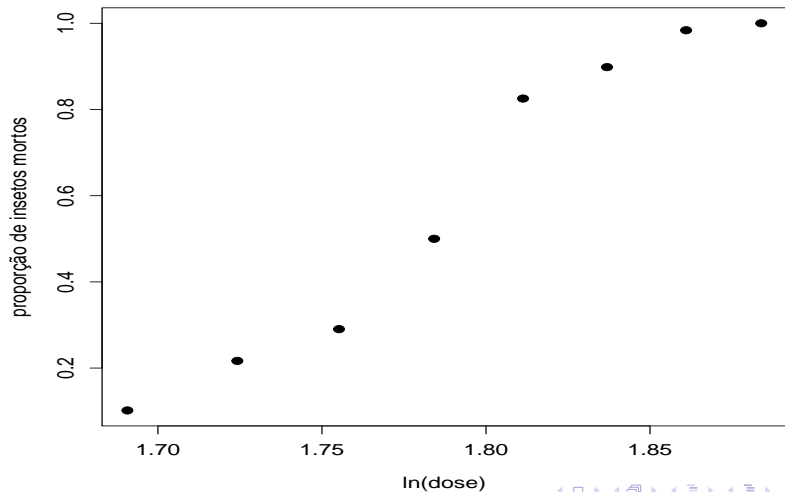
# Estudo sobre mortalidade de besouros

- Dados relativos ao percentual de besouros mortos quando expostos à diferentes doses de disulfeto de carbono gasoso ( $CS_2$ ).
- Objetivo: modelar a probabilidade morte dos insetos pela exposição ao referido gás, em função da respectiva dose.

# Mortalidade de besouros

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

# Proporção de besouros mortos em função da dose



## Modelo de regressão com ligação logito

$$Y_i | (\beta_0, \beta_1) \stackrel{ind.}{\sim} \text{binomial}(m_i, \mu_i)$$

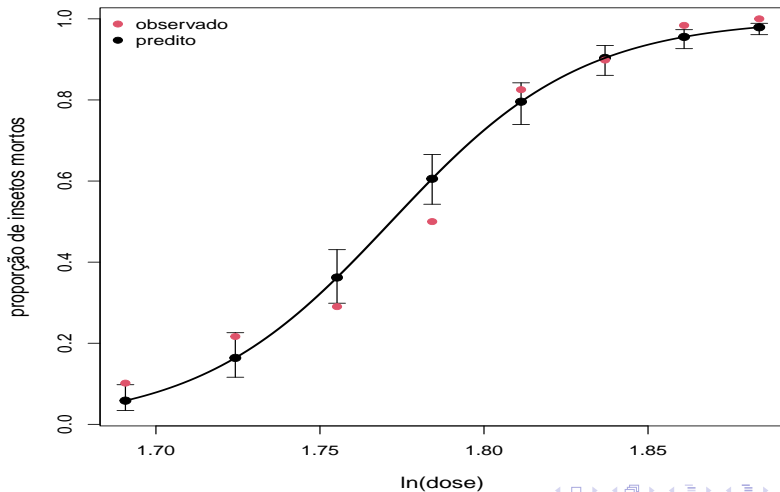
$$\ln\left(\frac{\mu_i}{1 - \mu_i}\right) = \beta_0 + \beta_1(x_i - \bar{x}), \bar{x} = \frac{1}{8} \sum_{i=1}^8 x_i, i = 1, 2, \dots, 8$$

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

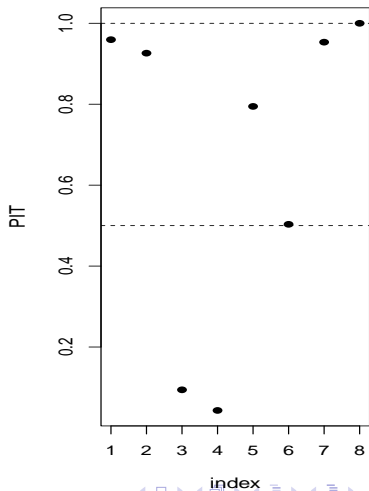
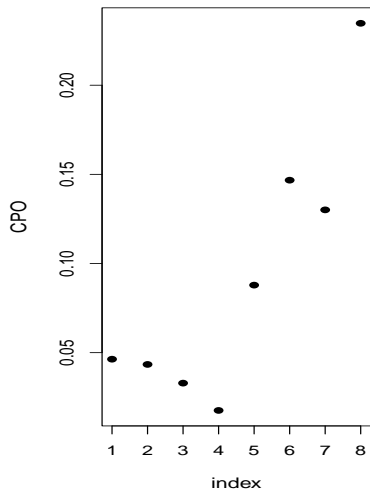
## Cont.

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

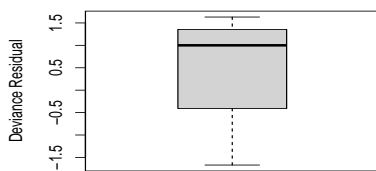
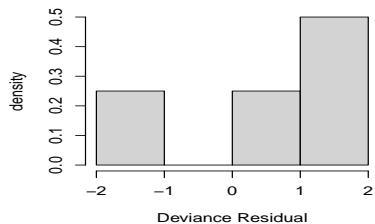
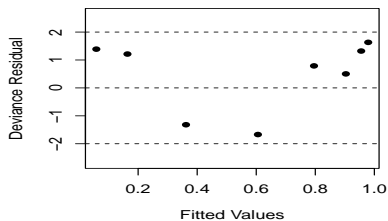
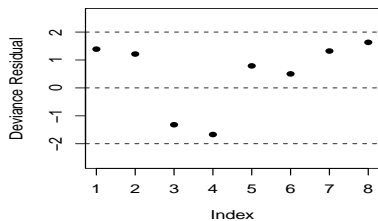
# Dados observados e distribuição preditiva (logito)



# CPO e PIT (logito)

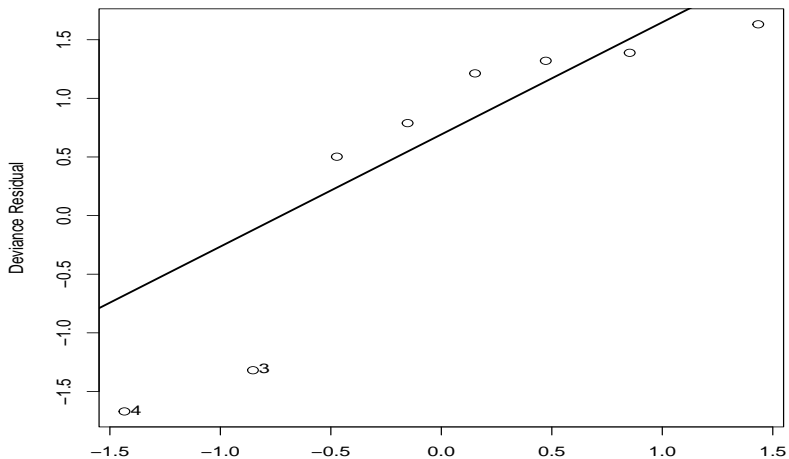


# Diagnóstico dos resíduos (logito)





## QQplot dos resíduos (logito)

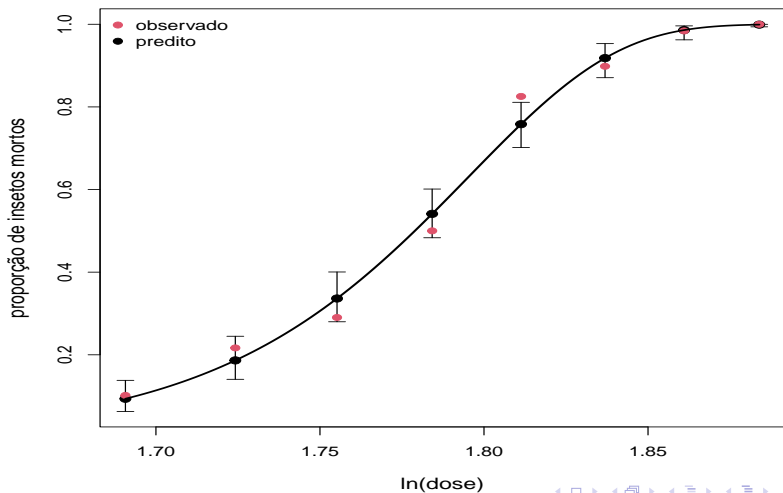


Quantiles of  $N(0,1)$

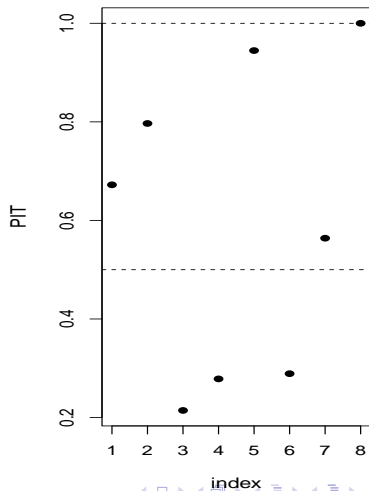
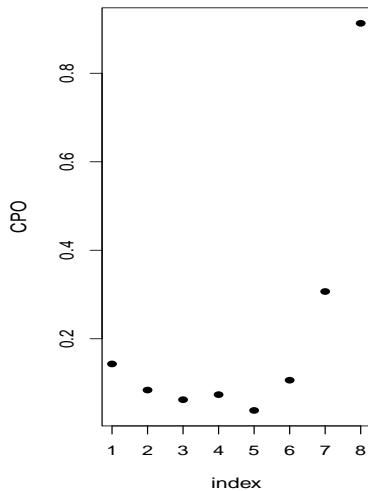
# Comentários

- Como temos poucas observação ( $n=8$ ), as análises dos gráficos de resíduos requerem cuidado adicional,
- Percebemos que, em geral, o modelo não prevê de forma adequada as proporções observadas.
- Além disso, os resíduos apresentam assimetria com a presença de outliers.
- Sugestão: ajustar um modelo com função de ligação complemento log-log (cloglog, que é assimétrica), que corresponde a utilizar a fda de uma distribuição de valor extremo.

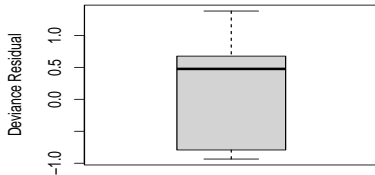
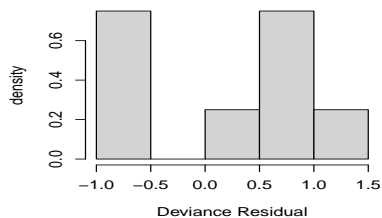
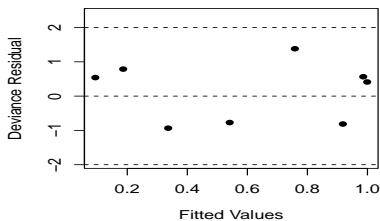
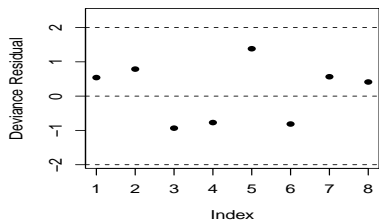
# Dados observados e distribuição preditiva (cloglog)



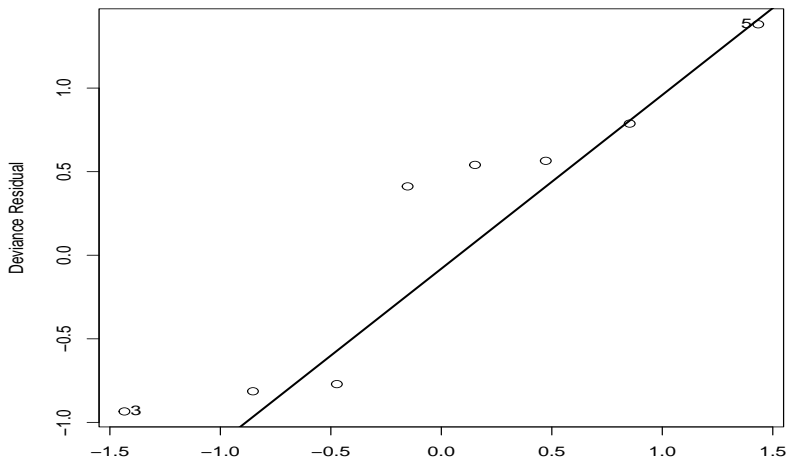
# CPO e PIT (cloglog)



# Diagnóstico dos resíduos (cloglog)



# QQplot dos resíduos (cloglog)

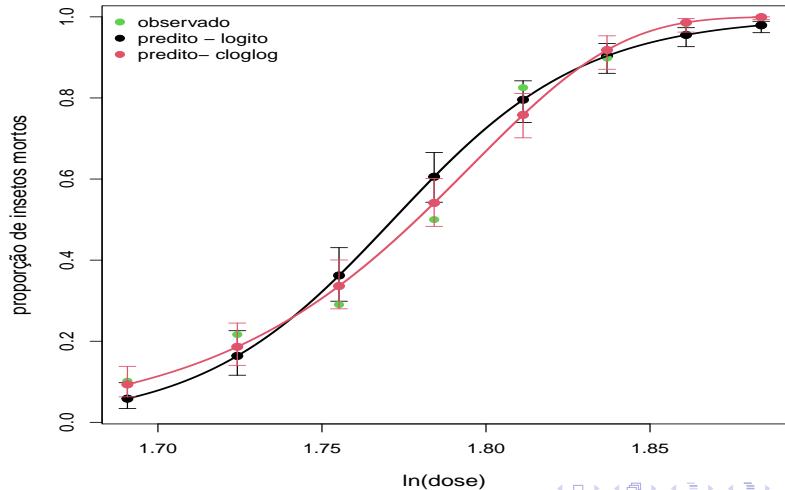


Quantiles of N(0,1)

# Critérios de informação

Modelo	DIC	WAIC	CPO	MLIK
logito	41,38	42,64	21,50	-22,81
cloglog	33,63	33,01	16,59	-19,55

# Comparação entre os modelos





# Resultados inferencias

Mod.	Parâm.	EAP	MeAP	DPAP	IC(95%)	HPD(95%)
logito	$\beta_0$	0,75	0,75	0,14	[0,48 ; 1,01]	[0,48 ; 1,01]
	$\beta_1$	34,32	34,31	2,88	[28,67 ; 39,95]	[28,67 ; 39,95]
cloglog	$\beta_0$	-0,05	-0,05	0,08	[-0,20 ; 0,11]	[-0,20 ; 0,11]
	$\beta_1$	22,14	22,14	1,78	[18,64 ; 25,64]	[18,64 ; 25,64]

# Comparação entre os modelos

$\log_{10} CS_2$	PropObs	logito		cloglog	
		EAP	HPD(95%)	EAP	HPD(95%)
1,69	0,10	0,06	[0,03 ; 0,10]	0,09	[0,06 ; 0,14]
1,72	0,22	0,16	[0,12 ; 0,23]	0,19	[0,14 ; 0,24]
1,76	0,29	0,36	[0,30 ; 0,43]	0,34	[0,28 ; 0,40]
1,78	0,50	0,61	[0,54 ; 0,67]	0,54	[0,48 ; 0,60]
1,81	0,83	0,80	[0,74 ; 0,84]	0,76	[0,70 ; 0,81]
1,84	0,90	0,90	[0,86 ; 0,93]	0,92	[0,87 ; 0,95]
1,86	0,98	0,96	[0,93 ; 0,97]	0,99	[0,96 ; 1,00]
1,88	1,00	0,98	[0,96 ; 0,99]	1,00	[0,99 ; 1,00]

# Comentários

- Todos os critérios de informação indicaram o modelo cloglog.
- O ajuste do modelo cloglog foi razoável e melhor do que aquele relativo ao modelo logito, tanto do ponto de vista residual quanto do ponto de vista preditivo.
- Sugestão: utilizar um link correspondente à fda de uma distribuição [família de escala normal assimétrica](#) (para tentar melhorar o ajuste do modelo).
- Exercício: verificar como fica o cálculo da razão de chances sob o modelo cloglog.