

# Modelos de regressão para dados discretos (parte 4): dados binários

Prof. Caio Azevedo

## Exemplo 10: germinação de sementes de Orobanche

- Orobanche: tipo de organismo vegetal (planta).
- Os dados foram obtidos a partir de um estudo sobre germinação de duas espécies de sementes de Orobanche (*O. aegyptiaca* 75 e *O. aegyptiaca* 73), veja (Hinde e Demétrio (1998)).
- Correspondem à quantidade de sementes analisadas ( $m$ ) e germinadas ( $y$ ) e tem-se o interesse na resposta  $y/m$  (conjunto de sementes).

## Exemplo 10: germinação de sementes de Orobanche

- As sementes foram cultivadas em diluições de  $1/125$  a partir de dois tipos de extrato de raiz (pepino ou feijão), em um esquema fatorial  $2 \times 2$  com replicações.
- Objetivo: avaliar o comportamento de cada espécie, sob cada um dos tipos de extrato, em termos da capacidade de germinação.

## Exemplo 10: dados

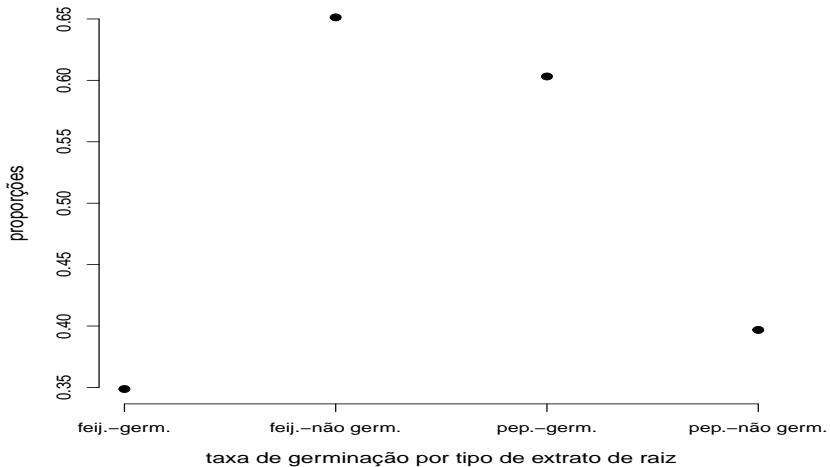
O. aegyptiaca 75				O. aegyptiaca 73			
feijão		pepino		feijão		pepino	
<i>y</i>	<i>m</i>	<i>y</i>	<i>m</i>	<i>y</i>	<i>m</i>	<i>y</i>	<i>m</i>
10	39	5	6	8	16	3	12
23	62	53	74	10	30	22	41
23	81	55	72	8	28	15	30
26	51	32	51	23	45	32	51
17	39	46	79	0	4	3	7
-	-	10	13	-	-	-	-

## Análise descritiva (considerando $y/m$ )

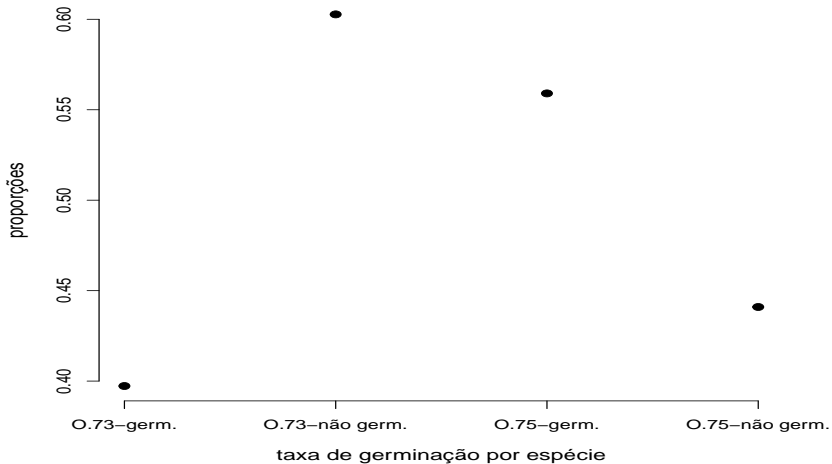
Os percentuais foram calculados dentro de cada categoria de tipo de extrato e espécie (os percentuais dentro de cada linha somam 100%).

tipo de extrato	germinação	
	sim	não
feijão	0,35	0,65
pepino	0,60	0,40
espécie		
O.aegyptiaca73	0,40	0,60
Oaegyptiaca75	0,56	0,44

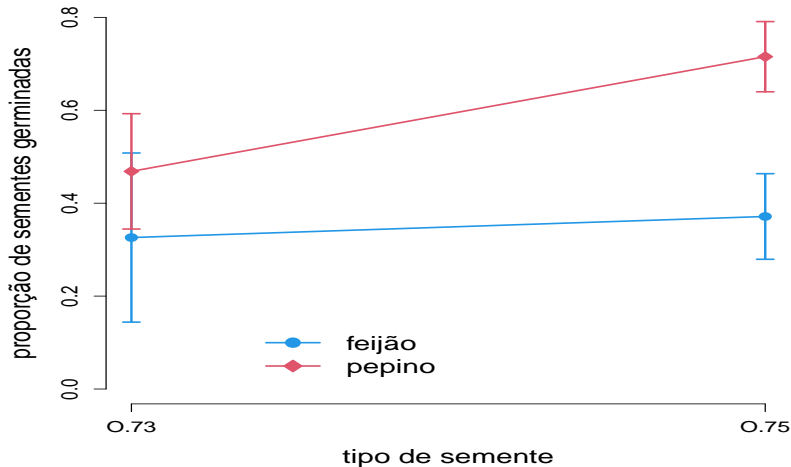
# Gráficos de proporções por tipo de extrato



# Gráficos de proporções por espécie



# Gráficos de perfis médios





# Modelos

- Estrutura geral

$$Y_{ijk} \stackrel{ind.}{\sim} \text{Binomial}(m_{ijk}, \mu_{ij})$$

$$F^{-1}(\mu_{ij}) = \alpha + \beta_i + \gamma_j + (\beta\gamma)_{ij}, i = 1, 2, j = 1, 2, k = 1, 2, \dots, n_{ij}$$

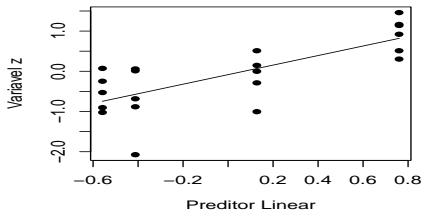
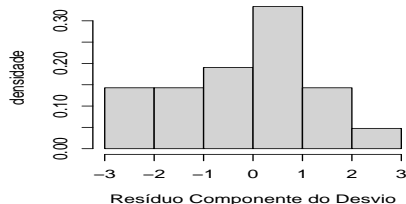
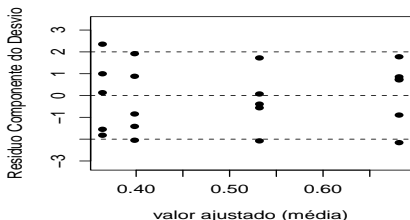
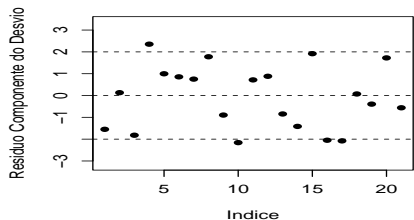
$$\beta_1 = \gamma_1 = (\beta\gamma)_{1j} = (\beta\gamma)_{i1} = 0, \forall i, j.$$

- $n_{ij}$  : número total de sementes pertencentes a  $i$ -ésima espécie (1: *O.aegyptiaca*73, 2: *O.aegyptiaca*75) e tratados com o  $j$ -ésimo tipo de extrato (1: feijão, 2: pepino),  $n_{11} = 5, n_{12} = 5, n_{21} = 5, n_{22} = 6$ .
- $Y_{ijk}$  : é o total de sementes germinadas na  $k$ -ésimo réplica, pertencentes à  $i$ -ésima espécie, e tratadas com o  $j$ -ésimo tipo de extrato vegetal.

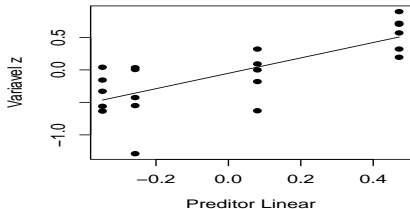
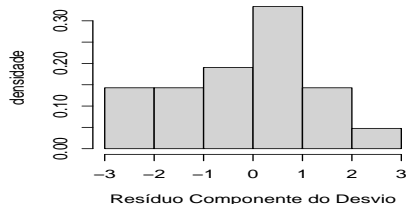
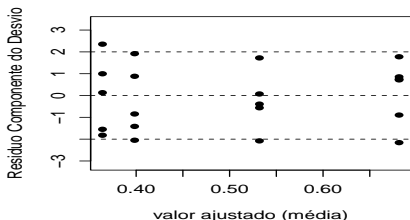
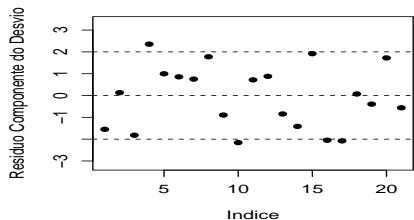
# Modelos

- Modelo 1:  $F^{-1}(\mu_{jk}) = \ln\left(\frac{\mu_{jk}}{1 - \mu_{jk}}\right)$  (logito).
- Modelo 2:  $F^{-1}(\mu_{jk}) = \Phi^{-1}(\mu_{jk})$  (probito).
- Modelo 3:  $F^{-1}(\mu_{jk}) = \tan[\pi(\mu_{jk} - 1/2)]$  (cauchito).
- Modelo 4:  $F^{-1}(\mu_{jk}) = \ln(-\ln(1 - \mu_{jk}))$  (cloglog/extremito).

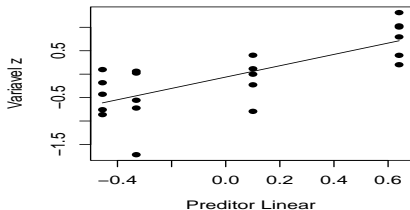
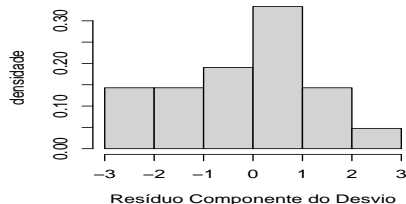
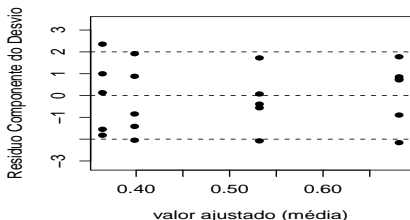
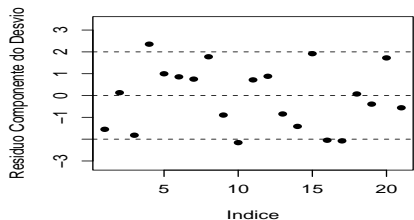
# Gráficos de diagnóstico: ligação logito



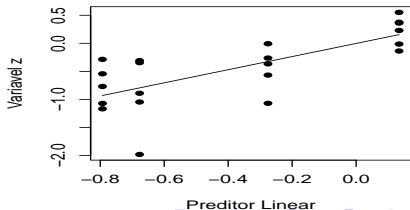
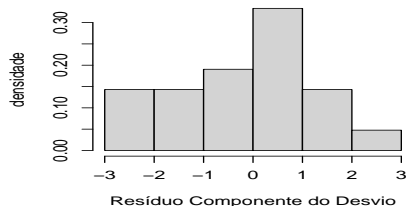
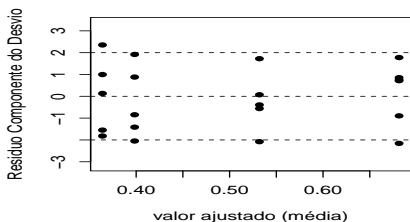
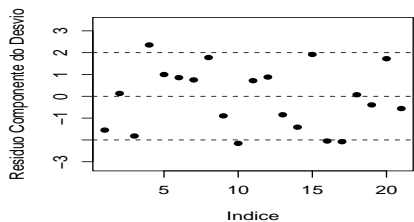
# Gráficos de diagnóstico: ligação probito



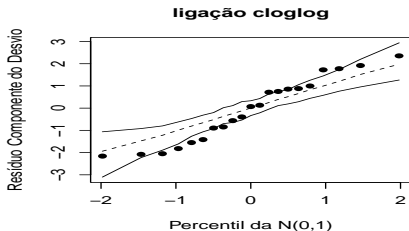
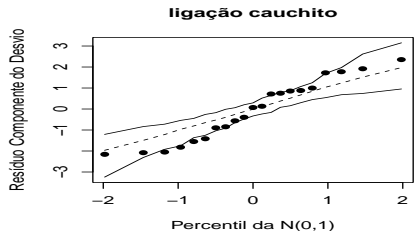
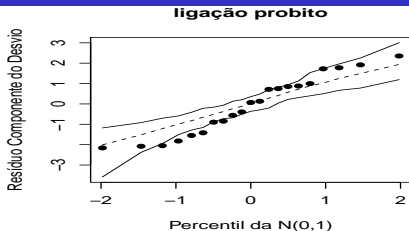
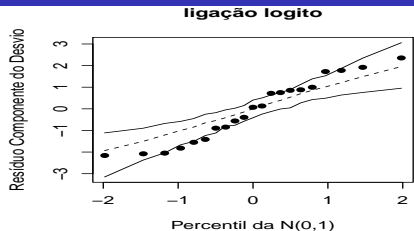
# Gráficos de diagnóstico: ligação cauchito



# Gráficos de diagnóstico: ligação cloglog

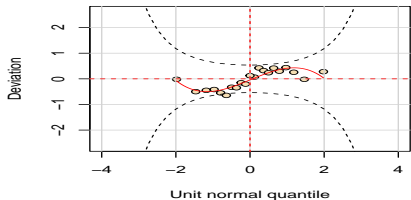


# Gráficos de envelopes para os quatro modelos

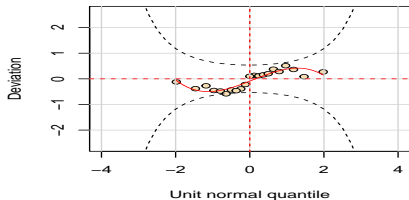


# Wormplots para os quatros modelos

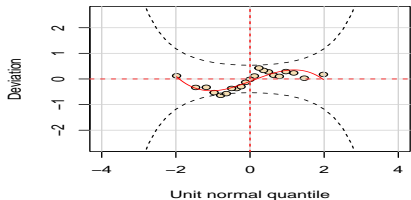
**ligação logito**



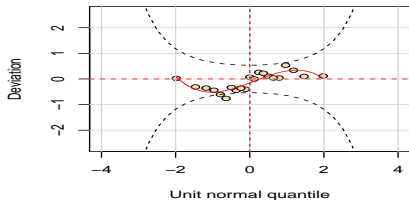
**ligação probito**



**ligação cauchito**

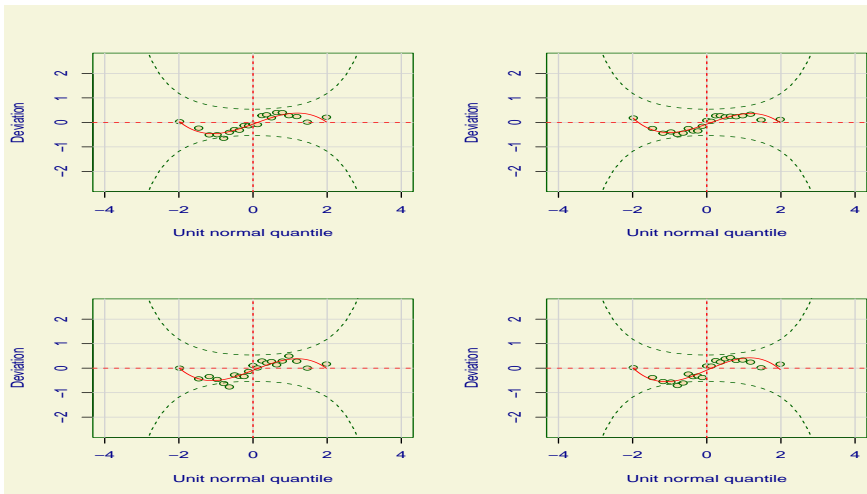


**ligação cloglog**

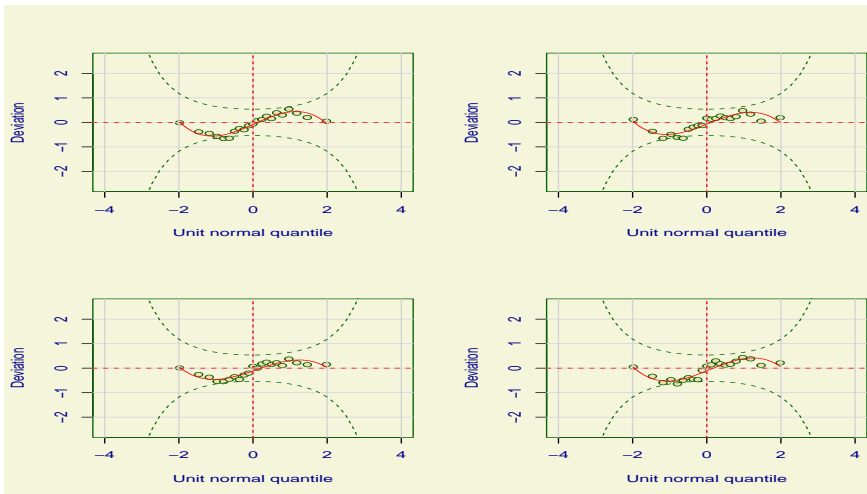




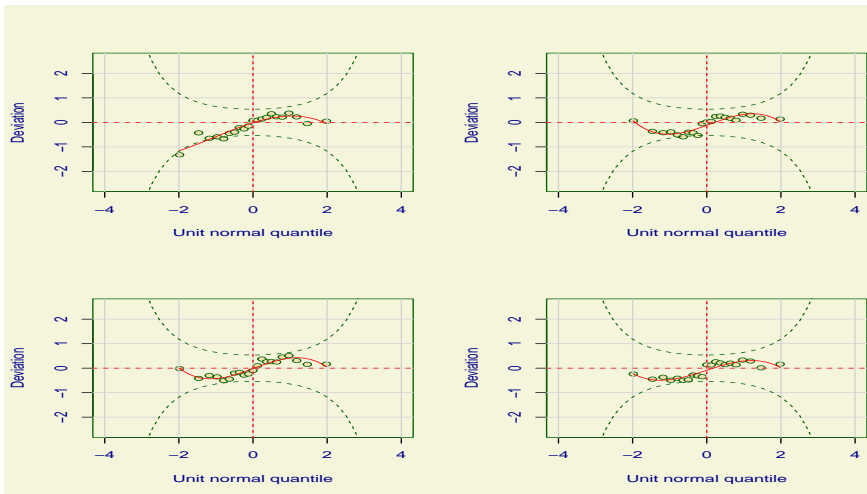
# Wormplots para o modelo logito



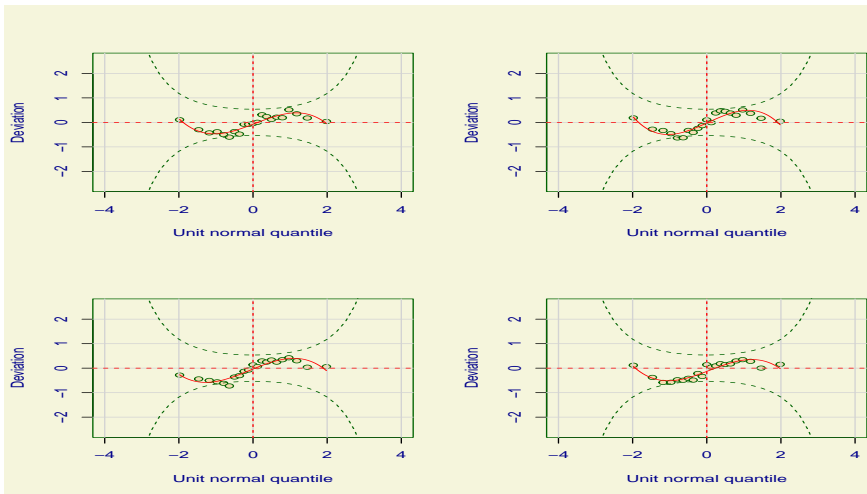
# Wormplots para o modelo probito



# Wormplots para o modelo cauchito



# Wormplots para o modelo cloglog

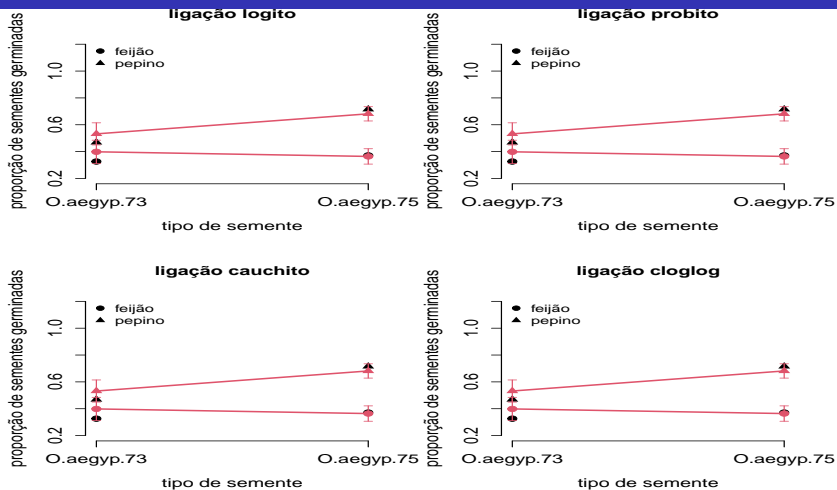


## Comparação entre os modelos (funções de ligação)

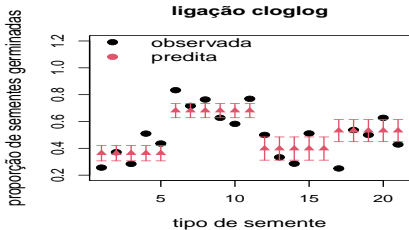
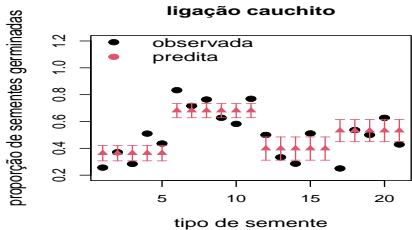
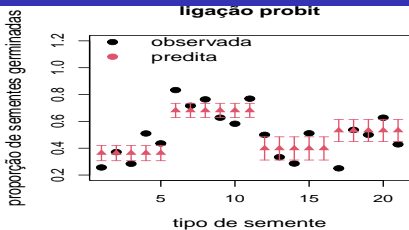
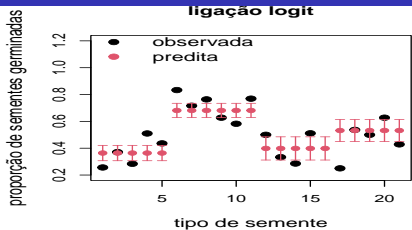
Modelo	AIC	BIC	AICc	SABIC	HQCIC	CAIC	p-valor	DABM
logito	117,87	122,05	120,37	109,70	118,78	126,05	0,11	0,0104
probito	117,87	122,05	120,37	109,70	118,78	126,05	0,11	0,0104
cauchito	117,87	122,05	120,37	109,70	118,78	126,05	0,11	0,0104
cloglog	117,87	122,05	120,37	109,70	118,78	126,05	0,11	0,0104

$$\text{DABM} = \frac{1}{n} \sum_{j=1}^2 \sum_{k=1}^2 \sum_{i=1}^{n_{jk}} \left| \frac{y_{ijk}}{m_{ijk}} - \tilde{\mu}_{ijk} \right|.$$

# Médias das proporções observadas e previstas



# Proporções individuais observadas e previstas



## Estimativas dos parâmetros: modelos 1 e 2

Modelo	Parâm.	Est.	EP	IC(95%)	Estat. $Z_t$	p-valor
logito	$\alpha$	-0,41	0,18	[-0,77 ; -0,05]	-2,24	0,0252
	$\beta_2$	-0,15	0,22	[-0,58 ; 0,29]	-0,65	0,5132
	$\gamma_2$	0,54	0,25	[ 0,05 ; 1,03]	2,16	0,0306
	$(\beta\gamma)_{22}$	0,78	0,31	[0,18 ; 1,38]	2,54	0,0111
probito	$\alpha$	-0,26	0,11	[-0,48 ; -0,03]	-2,25	0,0243
	$\beta_2$	-0,09	0,14	[-0,36 ; 0,18]	-0,65	0,5137
	$\gamma_2$	0,34	0,16	[ 0,03 ; 0,64]	2,17	0,0301
	$(\beta\gamma)_{22}$	0,48	0,19	[0,11 ; 0,85]	2,54	0,0112



## Estimativas dos parâmetros: modelos 3 e 4

Modelo	Parâm.	Est.	EP	IC(95%)	Estat. $Z_t$	p-valor
cauchito	$\alpha$	-0,33	0,15	[ -0,63 ; -0,03 ]	-2,15	0,0316
	$\beta_2$	-0,12	0,19	[ -0,50 ; 0,25 ]	-0,66	0,5100
	$\gamma_2$	0,43	0,20	[ 0,03 ; 0,83 ]	2,12	0,0342
	$(\beta\gamma)_{22}$	0,66	0,26	[ 0,15 ; 1,18 ]	2,55	0,0109
cloglog	$\alpha_2$	-0,68	0,14	[ -0,96 ; -0,39 ]	-4,69	< 0,0001
	$\beta_2$	-0,12	0,18	[ -0,46 ; 0,23 ]	-0,66	0,5112
	$\gamma_2$	0,40	0,19	[ 0,04 ; 0,77 ]	2,15	0,0315
	$(\beta\gamma)_{22}$	0,53	0,23	[ 0,08 ; 0,97 ]	2,34	0,0195

# Comentários

- A qualidade do ajuste dos modelos foram igualmente ruins.
- Uma vez que eles diferem apenas em termos das funções de ligação, o que estaria causando o mal ajuste.
  - Pontos aberrantes (provavelmente não).
  - Ausência de covariáveis (talvez) ou de termos extras (provavelmente não) no preditor linear.
  - Super/subdispersão (muito provavelmente).