

Modelos de regressão para dados discretos (parte 1): dados binários

Prof. Caio Azevedo

Exemplo 2: Estudo sobre vasoconstrição

- Dados sobre um estudo de vasoconstrição (veja [Paula, 2024](#), [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 (metodologias mais apropriadas: modelos mistos, modelos hierárquicos).

Banco de dados

Medida	Ocorrência	Volume	Razão
1	1	3,70	0,82
2	1	3,50	1,09
⋮	⋮	⋮	⋮
7	0	0,60	0,75
8	0	1,10	1,70
⋮	⋮	⋮	⋮
39	1	1,30	1,62

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

Distribuições Bernoulli e Binomial

- Bernoulli

$$f(y) = \mu^y (1 - \mu)^{(1-y)} \mathbb{1}_{\{0,1\}}(y).$$

- binomial: Seja Y^* a proporção de sucessos em m ensaios de Bernoulli independentes. Logo, temos que $mY^* \sim \text{binomial}(m, \mu)$. Nesse caso $\mathcal{E}(Y^*) = \mu, \mu \in (0, 1)$. Além disso,

$$f_{Y^*}(y^*) = \binom{n}{ny^*} \mu^{ny^*} (1 - \mu)^{n-ny^*} \mathbb{1}_{\{0,1/m,2/m,\dots,1\}}(y^*).$$

Distribuições Bernoulli e Binomial

- No Exemplo 2 temos $y_i = 1$ (Bernoulli) se ocorreu vaso constrição na i -ésima medida e 0 caso contrário.
- No Exemplo 8 temos que m_i é a quantidade de besouros expostos a i -ésima dose de CS_2 enquanto que $y_i = m_i y_i^*$ é a quantidade observada de besouros, expostos a i -ésima dose de CS_2 que morreram.

Modelo de regressão (geral) para dados binários

$$Y_i \stackrel{ind.}{\sim} \text{Bernoulli}(\mu_i)$$
$$F^{-1}(\mu_i) = \sum_{j=1}^p \beta_j x_{ji} \rightarrow \mu_i = F\left(\sum_{j=1}^p \beta_j x_{ji}\right), i = 1, 2, \dots, n$$

- Y_i : ocorrência (1) ou não (0) de algum evento.
- x_{ji} : valor da variável explicativa j associada ao indivíduo i .
- β_j : parâmetro associado ao impacto de cada covariável na probabilidade de ocorrência do supracitado evento.

Cont.

- (Cont.) Modelo com intercepto: $x_{1i} = 1, \forall i$.
- $F(\cdot)$: função de distribuição acumulada (fda) de alguma variável aleatória (contínua) com suporte em \mathcal{R} . $F^{-1}(\cdot)$ é conhecida como função de ligação.
- À rigor, $F(\cdot)$ não precisa ser uma fda, contanto que $F : \mathcal{R} \rightarrow [0, 1]$.

Modelo de regressão (geral) para dados binários (agrupados)

Análogo ao caso anterior, mas considerando que

$$Y_i = m_i Y_i^* \stackrel{ind.}{\sim} \text{binomial}(m_i, \mu_i).$$

- Y_i : quantidade de ocorrências de algum evento.
- As outras quantidades são como antes definidas.

Comentários

- Se, por exemplo, $F(\cdot)$ corresponder à fda de uma v.a. logística padrão, teremos $\ln\left(\frac{\mu_i}{1 - \mu_i}\right) = \eta_i$ (regressão logística).
- Note que em ambos os modelos podemos usar uma função de ligação que não corresponda ao inverso de uma fda.

Estimação por MV: modelo Bernoulli

- Nesse caso, $\phi = 1$.
- Lembremos que, para esse modelo, a função de ligação logito, $\ln\left(\frac{\mu_i}{1 - \mu_i}\right)$ é a função de ligação canônica.
- Ligação canônica: $\mathbf{S}(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta}, \phi)}{\partial \boldsymbol{\beta}} = \mathbf{X}'(\mathbf{y} - \boldsymbol{\mu})$, em que $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$ e $\mu_i = \frac{e^{\eta_i}}{1 + e^{\eta_i}}, i = 1, 2, \dots, n$ (veja também [aqui](#)).

Estimação por MV: modelo Bernoulli

- Ligação não-canônica:

$$\mathbf{S}(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta}, \phi)}{\partial \boldsymbol{\beta}} = \mathbf{X}' \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu}),$$

em que $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$, $\mu_i = F(\eta_i)$, $i = 1, 2, \dots, n$,

$\mathbf{V} = (V_1, \dots, V_n)$, $V_i = \mu_i(1 - \mu_i)$, $i = 1, 2, \dots, n$, $\mathbf{W} = (\omega_1, \dots, \omega_n)$ e
 $\omega_i = \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 V_i^{-1} = \frac{(f(\eta_i))^2}{\mu_i(1 - \mu_i)}$, em que $f(\eta_i) = \frac{\partial F(\eta_i)}{\partial \eta_i}$.

- Ligação canônica: $\mathbf{I}(\boldsymbol{\beta}) = \mathbf{X}' \mathbf{V} \mathbf{X}$.
- Ligação não-canônica: $\mathbf{I}(\boldsymbol{\beta}) = \mathbf{X}' \mathbf{W} \mathbf{X}$.

Estimação por MV: modelo binomial

- Lembremos que se $m_i Y_i^* \sim \text{binomial}(m_i, \mu_i)$ então $\phi_i = m_i$.
- Ligação geral: (veja [aqui](#)). Nesse caso, temos que

$$\begin{aligned} l(\beta) &= \sum_{i=1}^n \phi_i y_i^* \theta_i - \sum_{i=1}^n \phi_i b(\theta_i) + \sum_{i=1}^n c(y_i^*, \phi_i) \\ &= \sum_{i=1}^n \phi_i y_i^* \theta_i - \sum_{i=1}^n \phi_i b(\theta_i) + \text{const.} \end{aligned}$$

Estimação por MV: modelo binomial

- Além disso

$$S(\beta_j) = \sum_{i=1}^n \left\{ \sqrt{\frac{\omega_i}{V_i}} (\phi_i y_i^* - \mu_i \phi_i) X_{ji} \right\}.$$

- Assim, os resultados obtidos anteriormente podem ser usados considerando-se $y_i = \phi_i y_i^*$, $\mu_i^* = \mu_i \phi_i$ e $\phi = 1$.
- Logo, temos que sob a ligação canônica: $\mathbf{S}(\beta) = \mathbf{X}'(\mathbf{y} - \boldsymbol{\mu}^*)$ em que $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_n^*)$ e $\mu_i = F(\eta_i)$, $i = 1, 2, \dots, n$.

Estimação por MV: modelo binomial

- Ligação não-canônica:

$$\mathbf{S}(\boldsymbol{\beta}) = \mathbf{X}' \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu}^*),$$

em que $\mathbf{V} = (V_1, \dots, V_n)$,

$$V_i = \frac{\partial \mu_i^*}{\partial \theta_i} = m_i \frac{\partial \mu_i}{\partial \theta_i} = m_i \mu_i (1 - \mu_i), i = 1, 2, \dots, n,$$

$\mathbf{W} = (\omega_1, \dots, \omega_n)$ e

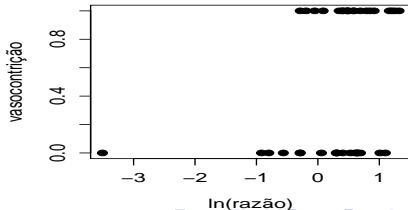
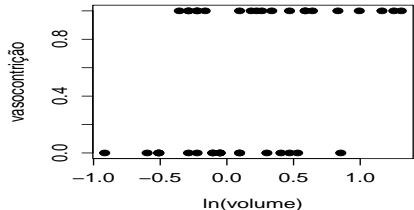
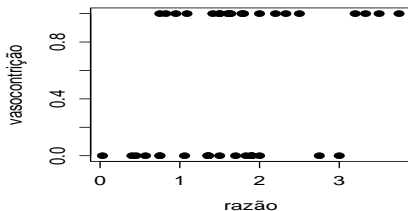
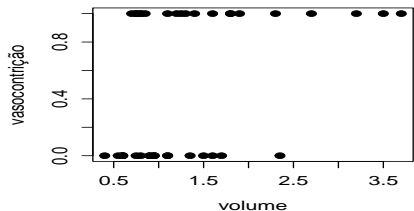
$$\omega_i = \left(\frac{\partial \mu_i^*}{\partial \eta_i} \right)^2 V_i^{-1} = \left(m_i \frac{\partial \mu_i}{\partial \eta_i} \right)^2 V_i^{-1} = m_i \frac{(f(\eta_i))^2}{\mu_i (1 - \mu_i)}.$$

- Ligação canônica: $\mathbf{I}(\boldsymbol{\beta}) = \mathbf{X}' \mathbf{V} \mathbf{X}$.
- Ligação não-canônica: $\mathbf{I}(\boldsymbol{\beta}) = \mathbf{X}' \mathbf{W} \mathbf{X}$.

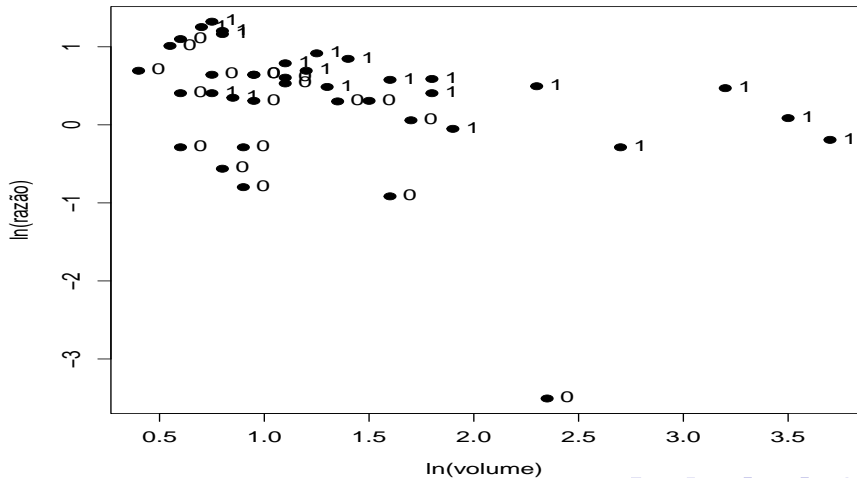
Estimação por MV

- Dessa forma, devemos utilizar o processo iterativo (algoritmo Escore de Fisher), apresentado anteriormente [aqui](#), para obtermos estimativas para β .
- As formas do desvio e do RCD para o modelo binomial/Bernoulli já foram vistas anteriormente, respectivamente: [aqui](#) e [aqui](#).

Gráficos de dispersão individuais



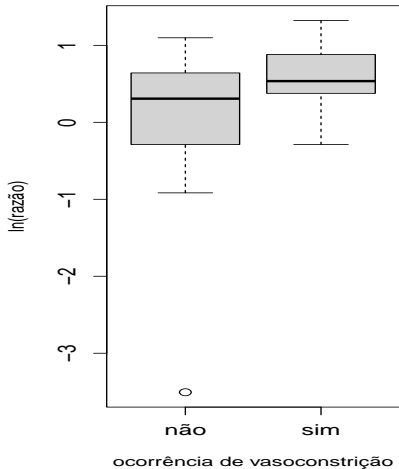
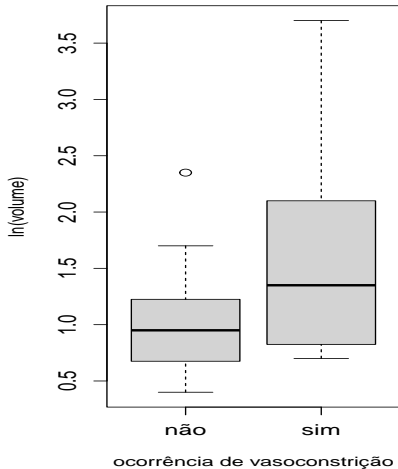
Gráficos de dispersão: $\ln(\text{razão}) \times \ln(\text{volume})$



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

Medida resumo	$\ln(\text{volume})$		$\ln(\text{razão})$	
	Resposta			
	0	1	0	1
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
ca	0,14	0,29	-2,23	-0,13
curt	2,5	1,9	8,4	2,3

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



Modelo de regressão para os dados de vasoconstricção

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

- Y_i : ocorrência (1) ou não (0) de vaso constricção.

(Cont.)

- x_{1i} : (Cont.) logaritmo natural do volume de ar inspirado da i -ésima observação.
- x_{2i} : logaritmo natural da razão de ar inspirado da i -ésima observação.
- $F(\cdot)$: corresponde à fda de uma distribuição logística padrão (portanto o nome regressão logística). Nesse caso, o $\text{logito}(\cdot)$ é a função de ligação.

Modelo de regressão para os dados de vasoconstricção

- 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**).

Modelo de regressão para os dados de vasoconstricção

- Analogamente, defina

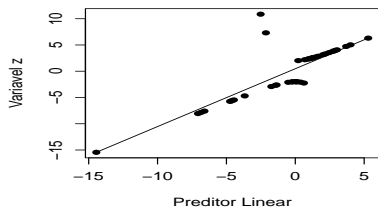
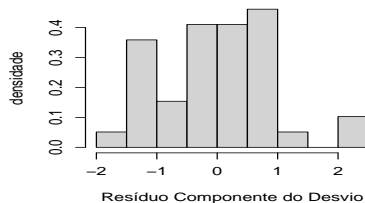
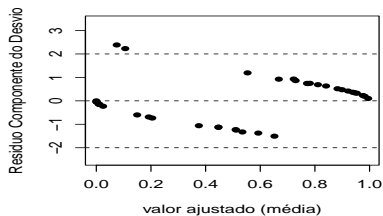
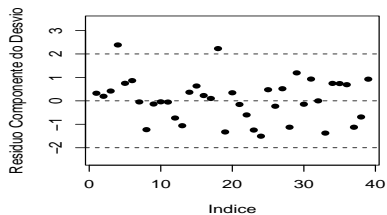
$$l_2(\mu_{i+1}) = \beta_0 + \beta_1 x_{1i} + \beta_2(x_{2i} + 1)$$

e

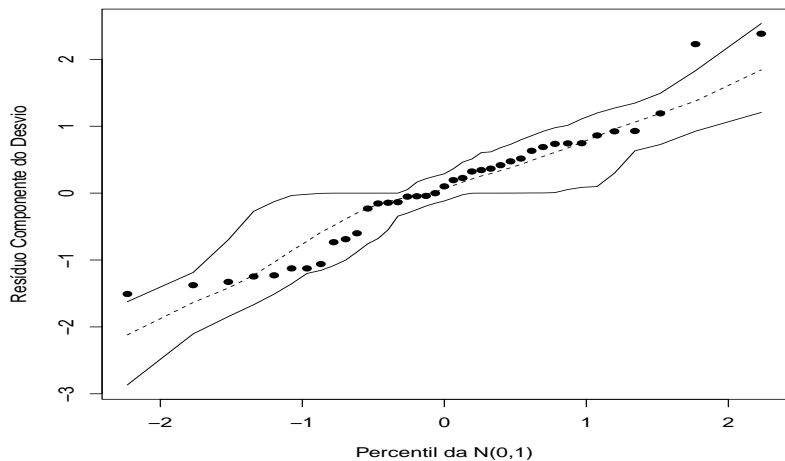
$$l_2(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}.$$

- 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**).

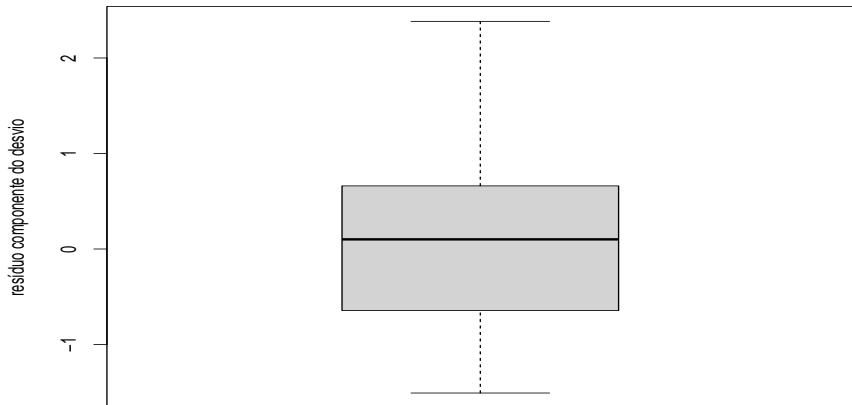
Gráficos de diagnóstico: ligação logito



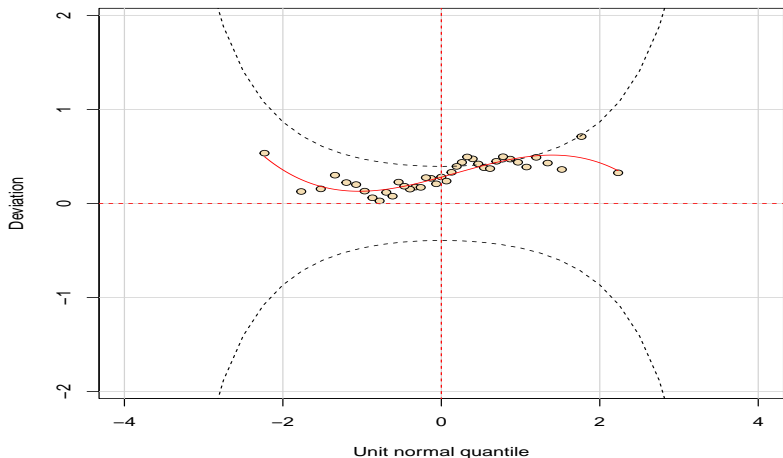
Envelope para os resíduos: ligação logito



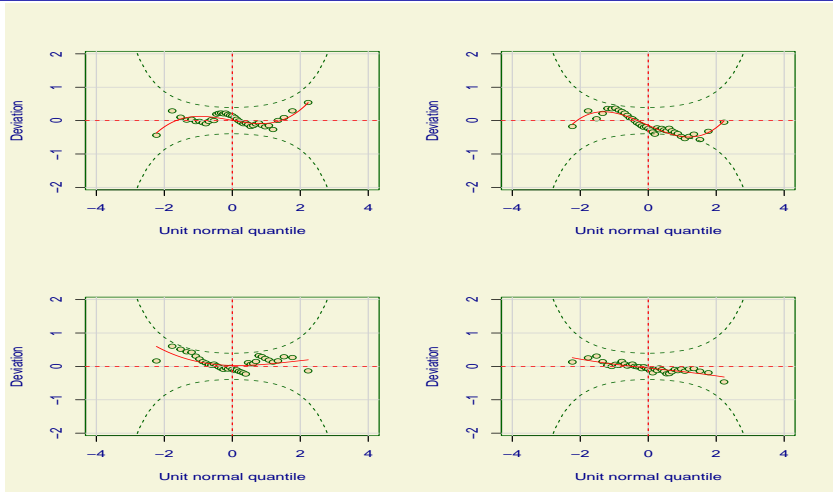
Boxplot do RCD: ligação logito



Worm plot para os resíduos: ligação logito



Worm plots para os resíduos: ligação logito



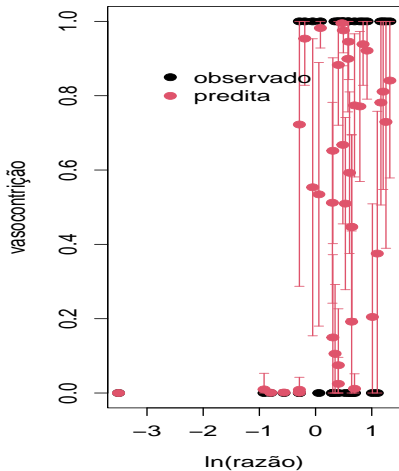
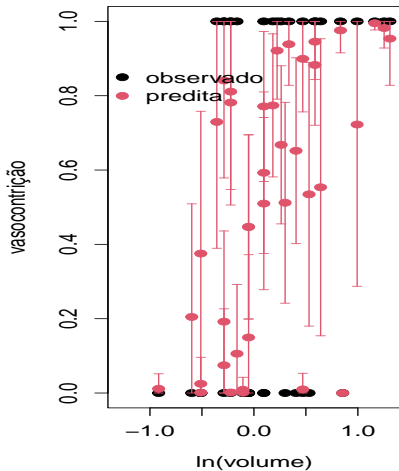
Probabilidades e valores preditos

- Probabilidades de ocorrência de vasoconstrição preditas:

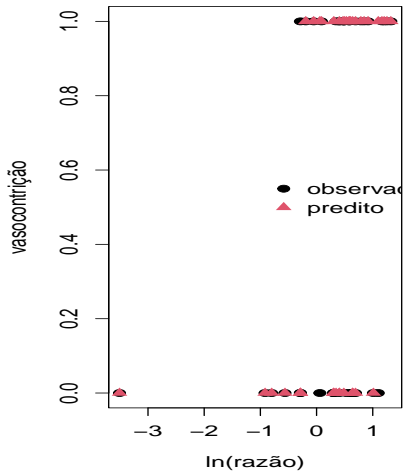
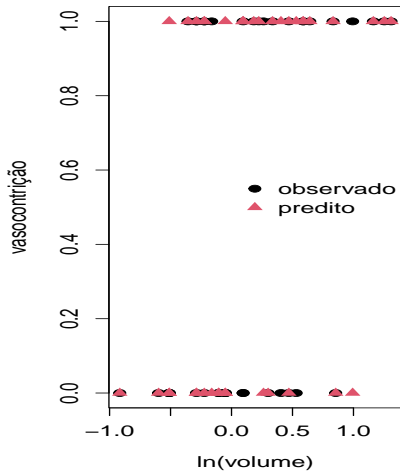
$$\tilde{\mu}_i = \frac{e^{\tilde{\beta}_0 + \tilde{\beta}_1 x_{1i} + \tilde{\beta}_2 x_{2i}}}{1 + e^{\tilde{\beta}_0 + \tilde{\beta}_1 x_{1i} + \tilde{\beta}_2 x_{2i}}}.$$

- Ocorrências de vasoconstrição preditas: simula-se u , $U \sim U(0, 1)$, se $\tilde{\mu}_i \geq u$, então $\tilde{Y}_i = 1$, caso contrário, $\tilde{Y}_i = 0$.
- Podemos comparar tanto as probabilidades quanto os valores preditos com os valores observados. Atenção: no caso dos valores preditos note que, como estamos gerando somente uma réplica, tal análise deve ser considerada com cautela.

Probabilidades previstas e valores observados: ligação logito



Valores observados e preditos: ligação logito



Perguntas

- Como gerar intervalos de confiança para $\frac{e^{\beta_0}}{1 + e^{\beta_0}}$, e^{β_1} e e^{β_2} ?
 - Método delta.
 - Fazer um IC para o parâmetro original e depois calcular o IC para a transformação.
 - Simulação/Reamostragem.

Intervalos de confiança para funções de interesse

- Sejam $g_1(\boldsymbol{\beta}) \equiv \tau_1 = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$, $g_2(\boldsymbol{\beta}) \equiv \tau_2 = e^{\beta_1}$ e $g_3(\boldsymbol{\beta}) \equiv \tau_3 = e^{\beta_2}$.
- Seja $\hat{\boldsymbol{\beta}}$ o estimador de MV de $\boldsymbol{\beta}$. Já vimos que, para n suficientemente grande, $\hat{\boldsymbol{\beta}} \approx N(\boldsymbol{\beta}, \boldsymbol{\Sigma}_\beta)$, em que $\boldsymbol{\Sigma}_\beta = \mathbf{I}^{-1}(\boldsymbol{\beta})$.
- O método delta nos diz que, para n suficientemente grande, $\hat{\tau}_i \approx N(\tau_i, \boldsymbol{\Psi}_i \boldsymbol{\Sigma}_\beta \boldsymbol{\Psi}_i')$, em que

$$\boldsymbol{\Psi}_i = \begin{bmatrix} \frac{\partial}{\partial \beta_0} g_i(\boldsymbol{\beta}) & \frac{\partial}{\partial \beta_1} g_i(\boldsymbol{\beta}) & \frac{\partial}{\partial \beta_2} g_i(\boldsymbol{\beta}) \end{bmatrix}.$$

Intervalos de confiança para funções de interesse

- Nesse caso,

$$\Psi_1 = \begin{bmatrix} \frac{e^{\beta_0}}{(1 + e^{\beta_0})^2} & 0 & 0 \end{bmatrix},$$

$$\Psi_2 = \begin{bmatrix} 0 & e^{\beta_1} & 0 \end{bmatrix},$$

$$\Psi_3 = \begin{bmatrix} 0 & 0 & e^{\beta_2} \end{bmatrix}.$$

- Assim $IC(\tau_i, \gamma) = \left[\hat{\tau}_i - z_{(1+\gamma)/2} \sqrt{\hat{\psi}_i}; \hat{\tau}_i + z_{(1+\gamma)/2} \sqrt{\hat{\psi}_i} \right]$, em que $P(Z \geq z_{(1+\gamma)/2}) = \frac{1+\gamma}{2}$ e $\hat{\psi}_i = \hat{\Psi}_i \hat{\Sigma}_\beta \hat{\Psi}_i'$, $Z \sim N(0, 1)$ (lembrando que este é um IC assintótico).

Intervalos de confiança para funções de interesse

Parâm.	Est.	IC (transf.)	IC (mét. delta)	IC (simul.)
τ_1	0.05	[< 0,01 ; 0,43]	[-0,08 ; 0,18]	[<0,01 ; 0,22]
τ_2	177,56	[4,59 ; 6862,99]	[-471,35 ; 826,48]	[19,85 ; 58917656,63]
τ_3	95,74	[2,61 ; 3511,02]	[-249,12 ; 440,61]	[12,29 ; 7914178,99]

Neste caso, os IC's obtidos através do método delta, devem ser truncados à esquerda do zero.

Estimativas relativas ao modelo de regressão logística

Parâmetro	Estimativa	EP	IC(95%)	Estat. Z_t	p-valor
β_0	-2,87	1,32	[-5,46 ; -0,29]	-2,18	0,0295
β_1	5,17	1,86	[1,52 ; 8,83]	2,78	0,0055
β_2	4,56	1,83	[0,96 ; 8,16]	2,48	0,0131

Todos os parâmetros são significativos.

Comparação com outros modelos (funções de ligação)

Modelo	AIC	BIC	AICc	SABIC	HQCIC	CAIC	DABM
Logito	35,23	40,22	35,91	30,83	37,02	43,22	0,24
Probito	35,29	40,28	35,97	30,89	37,08	43,28	0,25
Cauchito	31,08	36,07	31,76	26,68	32,87	39,07	0,16
Cloglog	32,62	37,61	33,31	28,23	34,41	40,61	0,21

$DABM = \frac{1}{n} \sum_{i=1}^n |y_i - \tilde{\mu}_i|$. Lembrando que: $\mu_i = \Phi(\eta_i)$ (probito),

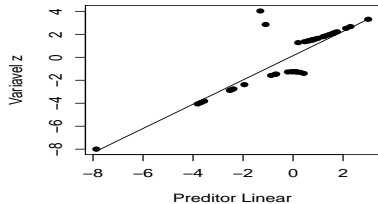
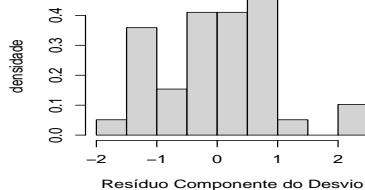
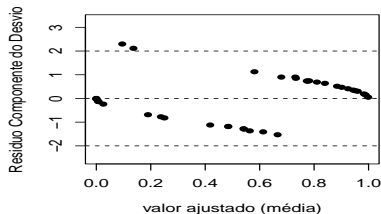
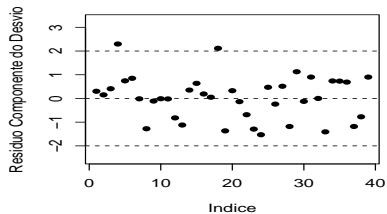
$\mu_i = \frac{1}{\pi} \arctan(\eta_i) + \frac{1}{2}$ (cauchito) e $\mu_i = 1 - e^{-e^{\eta_i}}$ (cloglog) e $\Phi(\cdot)$ é a

fda da normal padrão. Exercício, escrever os demais modelos e

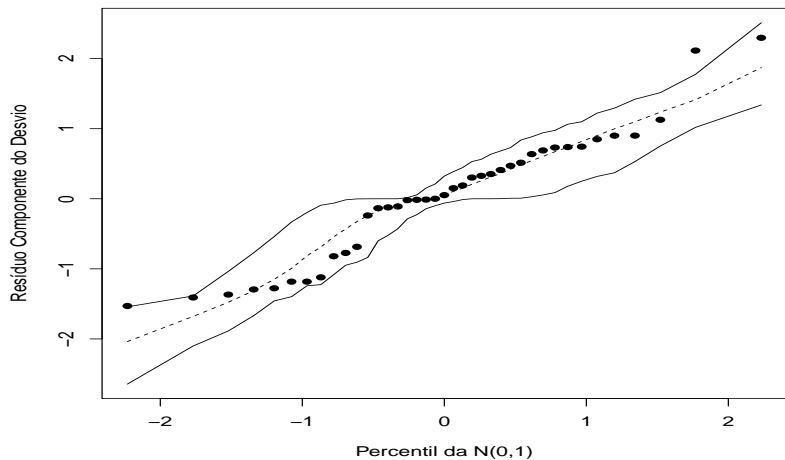
interpretações de funções de seus parâmetros e a forma das razões de

chances.

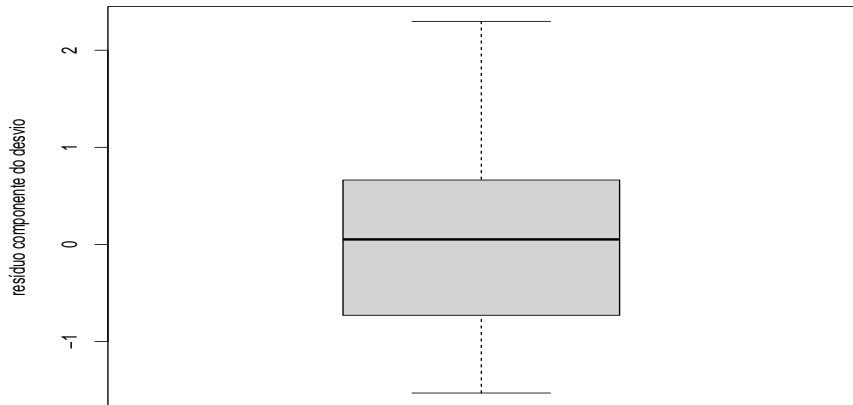
Gráficos de diagnóstico: ligação probito



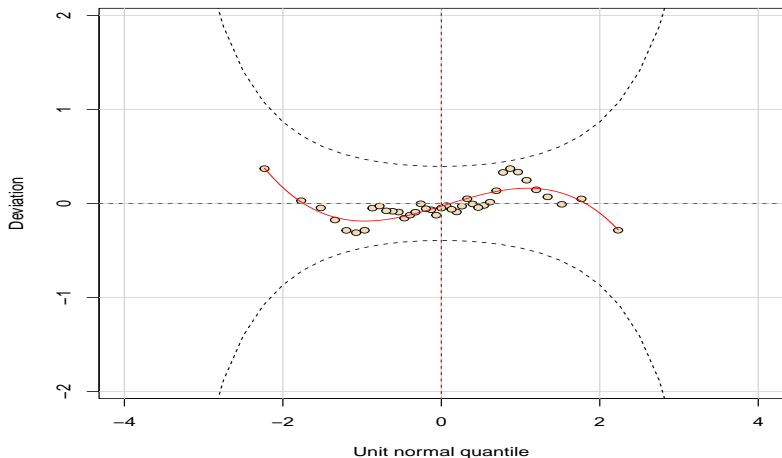
Envelope para os resíduos: ligação probito



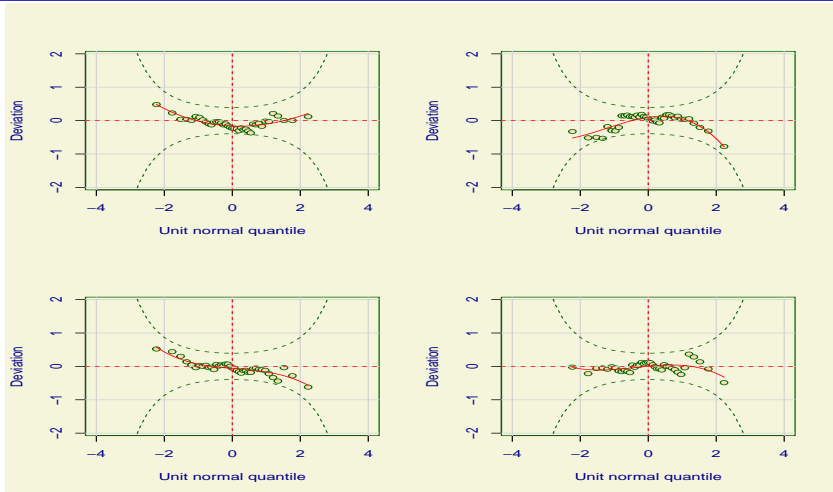
Boxplot do RCD: ligação probito



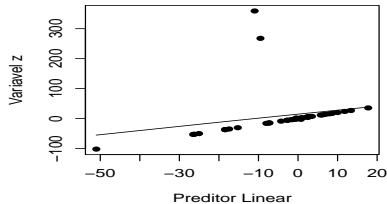
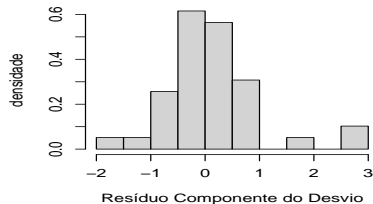
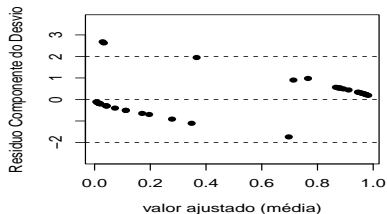
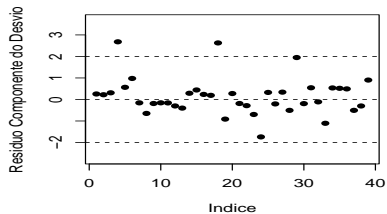
Worm plot para os resíduos: ligação probito



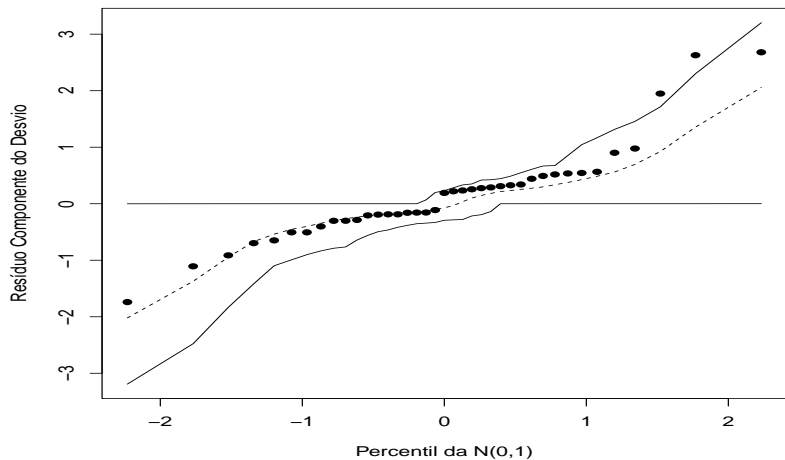
Worm plots para os resíduos: ligação probito



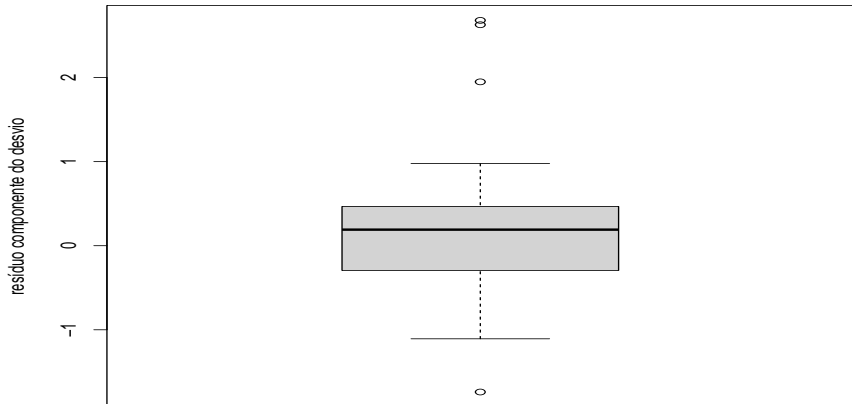
Gráficos de diagnóstico: ligação cauchito



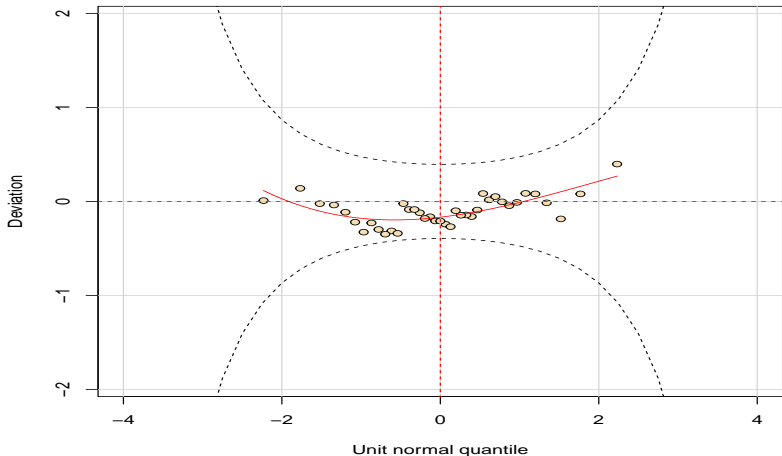
Envelope para os resíduos: ligação cauchito



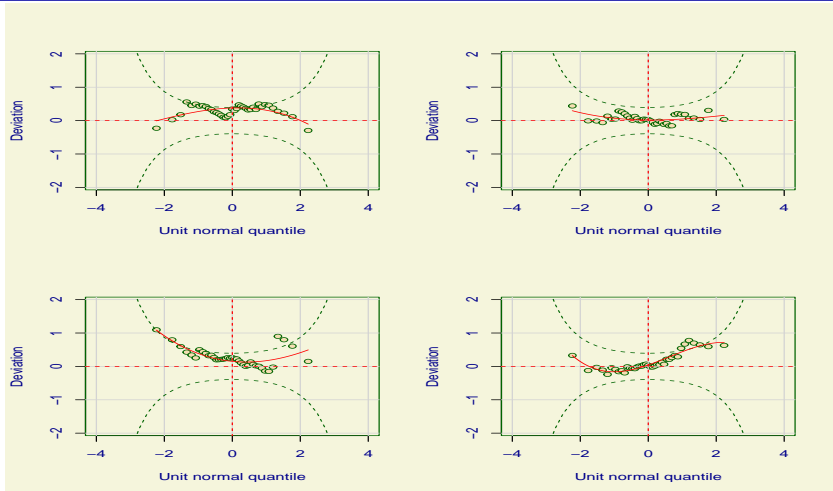
Boxplot do RCD: ligação cauchito



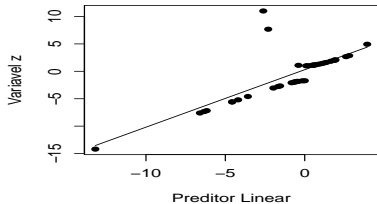
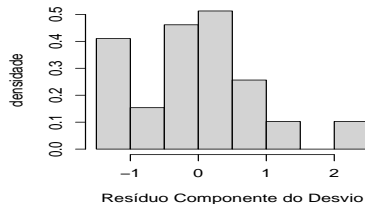
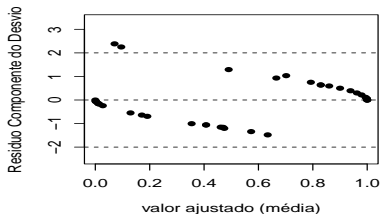
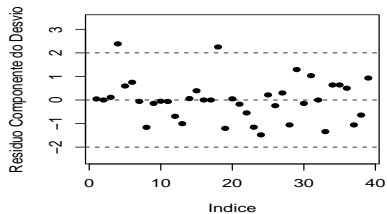
Worm plot para os resíduos: ligação cauchito



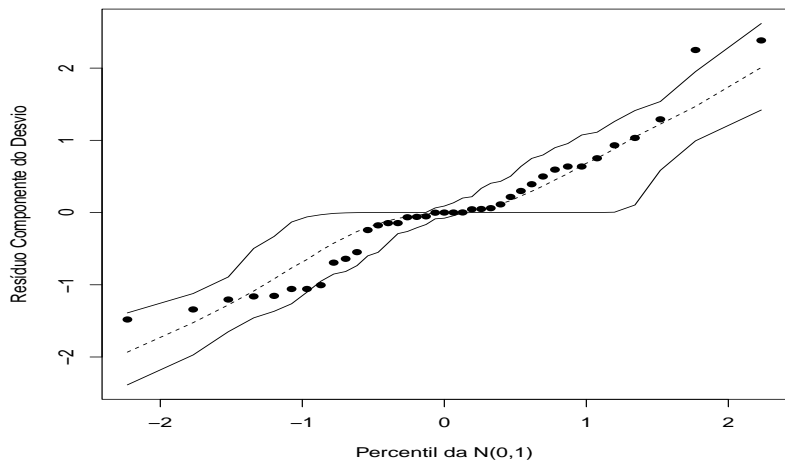
Worm plots para os resíduos: ligação cauchito



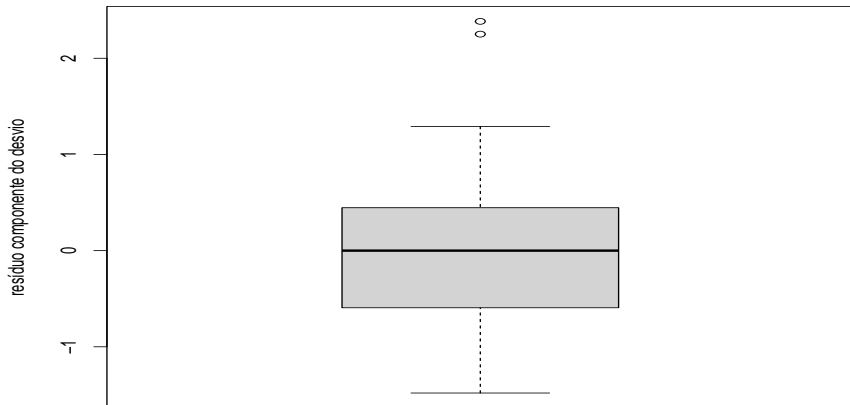
Gráficos de diagnóstico: ligação complemento log-log



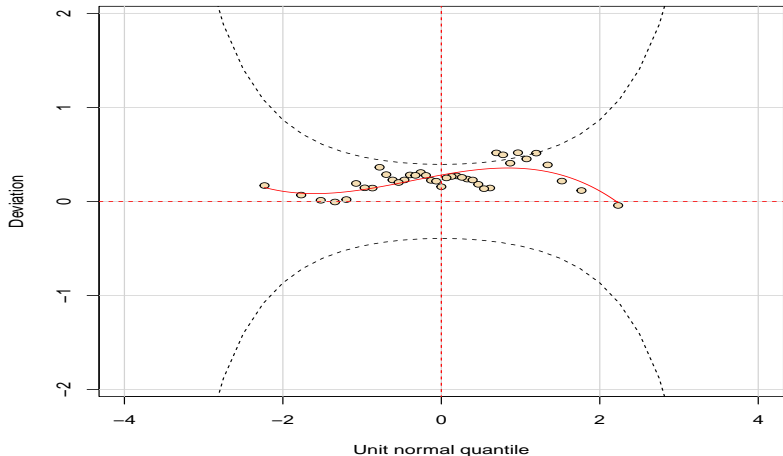
Envelope para os resíduos: ligação complemento log-log



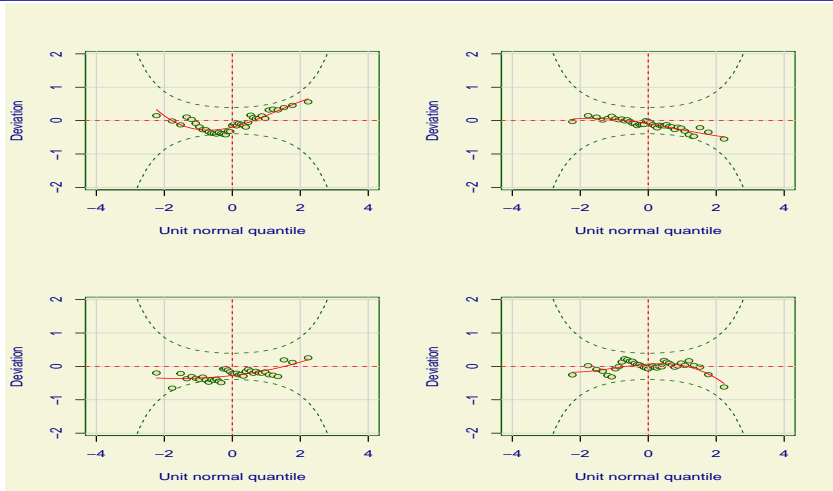
Boxplot do RCD: ligação cloglog



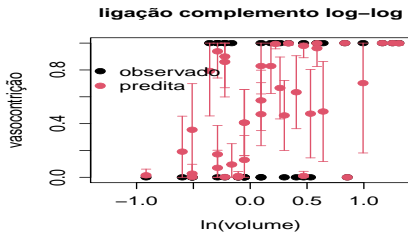
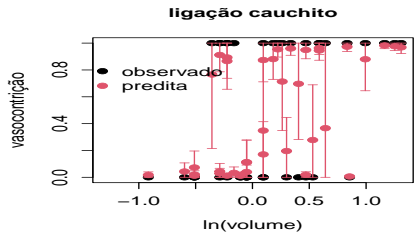
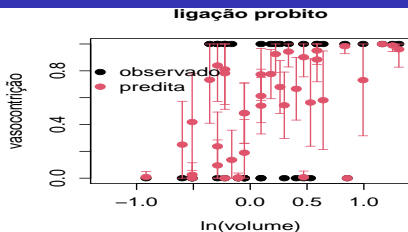
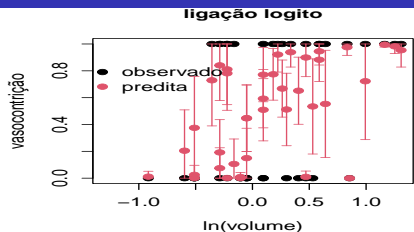
Worm plot para os resíduos: ligação cloglog



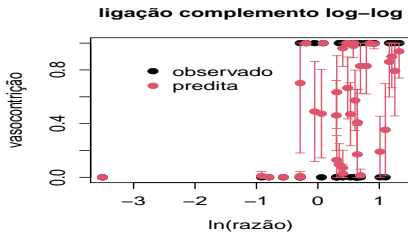
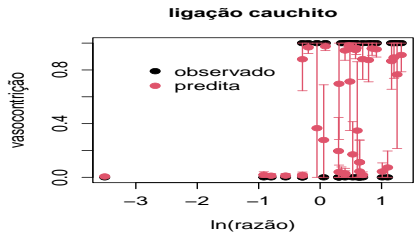
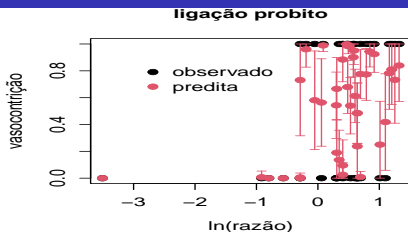
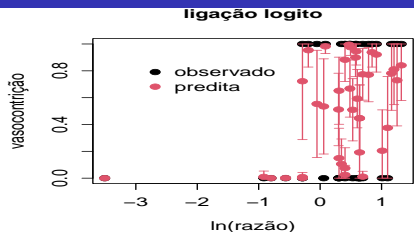
Worm plots para os resíduos: ligação cloglog



Probabilidades previstas e valores observados



Probabilidades previstas e valores observados



Predição (total por linha)

■ Ligação logito:

	Predito	
Obs.	0	1
0	68,42	31,58
1	20,00	80,00

■ Ligação probito:

	Predito	
Obs.	0	1
0	89,47	10,53
1	15,00	85,00

Predição (total por linha)

- Ligação cauchito:

	Predito	
Obs.	0	1
0	94,74	5,26
1	20,00	80,00

- Ligação cloglog:

	Predito	
Obs.	0	1
0	73,68	26,32
1	15,00	85,00

Predição (número total de observações)

- Percentual de classificação correta (zeros e uns)

logito	probito	cauchito	cloglog
74,36	87,18	87,18	79,49

Estimativas dos parâmetros dos modelos

Modelo	Parâm.	Est.	EP	IC(95%)	Estat. Z_t	p-valor
logito	β_0	-2,88	1,32	[-5,46 ; -0,29]	-2,18	0,0295
	β_1	5,18	1,86	[1,52 ; 8,83]	2,78	0,0055
	β_2	4,56	1,84	[0,96 ; 8,16]	2,48	0,0131
probito	β_0	-1,50	0,68	[-2,85 ; -0,16]	-2,20	0,0279
	β_1	2,86	0,94	[1,02 ; 4,70]	3,05	0,0023
	β_2	2,51	0,95	[0,65 ; 4,38]	2,64	0,0083
cauchito	β_0	-11,89	7,79	[-27,16 ; 3,39]	-1,53	0,1272
	β_1	19,09	12,87	[-6,14 ; 44,31]	1,48	0,1380
	β_2	15,80	10,12	[-4,03 ; 35,64]	1,56	0,1183
c. log-log	β_0	-2,97	1,09	[-5,12 ; -0,83]	-2,72	0,0065
	β_1	4,34	1,56	[1,28 ; 7,39]	2,78	0,0054
	β_2	3,97	1,38	[1,26 ; 6,68]	2,87	0,0041

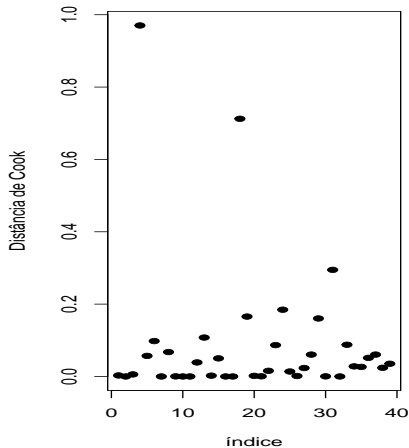
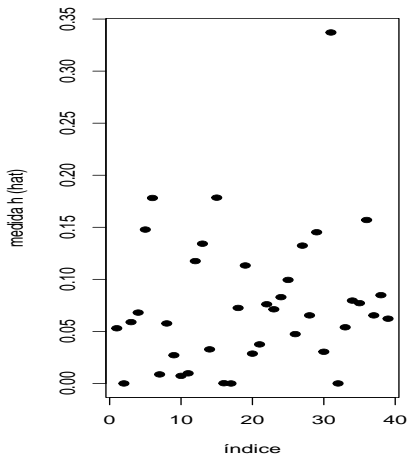
Conclusões

- O modelo melhor ajustado é o que considera a função de ligação cauchito. Entretanto, muito provavelmente devido ao fato de que este modelo impõe probabilidades maiores para valores mais extremos, os erros-padrão associados às estimativas indicam uma não significância dos parâmetros, adicionalmente ao tamanho reduzido da amostra.
- Além disso, os ajustes foram muito similares.
- No geral, nenhum ajuste foi satisfatório.

Conclusões

- Em princípio podemos conduzir as inferências com base no modelo logito. Se optarmos por utilizar o modelo cauchito, as quantidades de interesse (incluindo as razões de chances) devem ser calculadas usando a fda da distribuição Cauchy.
- Fórmula geral para a razão de chances: $\frac{F(\eta_{i+1})/(1 - F(\eta_{i+1}))}{F(\eta_i)/(1 - F(\eta_i))}$.
- Outras possibilidades: utilizar funções de ligação baseadas nas distribuições normal assimétrica e t de Student assimétrica.

Análise de influência



Análise de Sensibilidade

Parâmetros β_0

Obs.	Est.	EP	IC(95%)	Estat. Z_t	p-valor
todas	-2,88	1,32	[-5,46;-0,29]	-2,18	0,0295
-#4	-5,21	2,21	[-9,53;-0,88]	-2,36	0,0184
-#18	-4,76	2,03	[-8,74;-0,78]	-2,34	0,0191
-#31	-3,04	1,36	[-5,70;-0,38]	-2,24	0,0248

Análise de Sensibilidade

Parâmetros β_1

Obs.	Est.	EP	IC(95%)	Estat. Z_t	p-valor
todas	5,18	1,86	[1,52;8,83]	2,78	0,0055
-#4	8,47	3,31	[1,98;14,96]	2,56	0,0106
-#18	7,67	2,98	[1,84;13,50]	2,58	0,0100
-#31	4,97	1,82	[1,39;8,54]	2,72	0,0065

Análise de Sensibilidade

Parâmetros β_2

Obs.	Est.	EP	IC(95%)	Estat. Z_t	p-valor
todas	4,56	1,84	[0,96;8,16]	2,48	0,0131
-#4	7,45	2,98	[1,61;13,30]	2,50	0,0124
-#18	6,88	2,75	[1,49;12,27]	2,50	0,0123
-#31	4,76	1,88	[1,07;8,46]	2,53	0,0115

Análise de Sensibilidade

- Apesar de não terem mudado de forma significativa, a significância dos parâmetros, cada observação, ao ser retirada individualmente, mudou (substancialmente, em termos percentuais), as estimativas pontuai, erros-padrão e IC's.
- Perfis:

Observação	Resposta	Log(vol)	Log(razão)
4	1	-0,29	0,41
18	1	-0,16	0,35
31	1	0,99	-0,29

Análise de Sensibilidade

■ Medidas resumo

Medida	Log(vol)	Log(razão)
média	0,16	0,32
min.	-0,92	-3,51
1Q	-0,22	0,07
mediana	0,10	0,49
3Q	0,50	0,69
Máx.	1,31	1,32

Análise de Sensibilidade

- A obs. 4 tem $\text{Log}(\text{vol})$ abaixo da média e do 1Q e $\text{Log}(\text{razão})$ abaixo da mediana.
- A obs. 18 tem $\text{Log}(\text{vol})$ abaixo da mediana e próxima do 1Q e $\text{Log}(\text{razão})$ abaixo da mediana.
- A obs. 31 tem $\text{Log}(\text{razão})$ abaixo da média e do 1Q e $\text{Log}(\text{volume})$ acima do 3Q.
- Deve-se investigar melhor essas observações.
- Alternativa em termos de modelagem: utilizar funções de ligação com caudas pesadas (e com assimetria).