MLG para respostas positivas (assimétricas): parte 2

Prof. Caio Azevedo

Estudo de simulação de resíduos

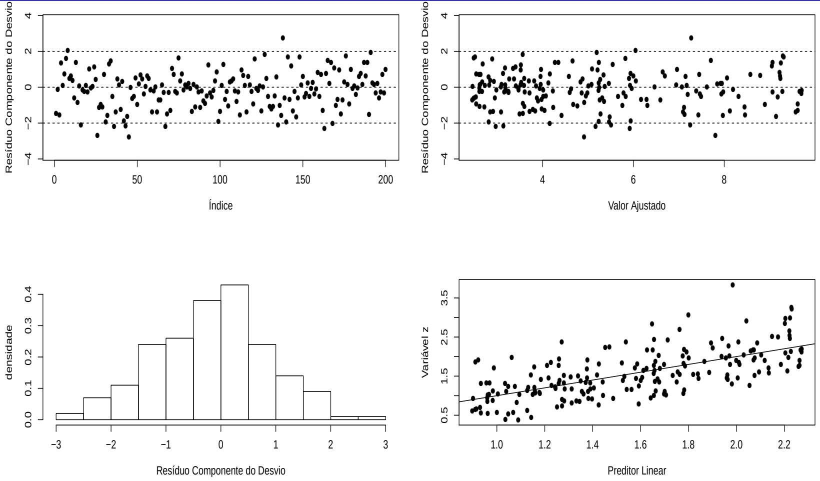
- Para os modelos de 1 a 3, $\mu_i = e^{1+1,2x_i}$, $x_i \overset{i.i.d.}{\sim} U(0,1)$, $\phi = 5$, i=1,2,....,n, n = 200.
- Para os modelos 4 e 5, $\mu_i = e^{3+1,1x_i}$ (restante igual).
- Modelo 1: $Y_i \overset{ind.}{\sim}$ gama(μ_i, ϕ) (simulado e ajustado).
- Modelo 2: $Y_i \overset{ind.}{\sim}$ gama$(\mu_i, \phi_i), \phi_i = exp(3x_i)$ (simulado); Modelo 1 (ajustado).

Estudo de simulação de resíduos

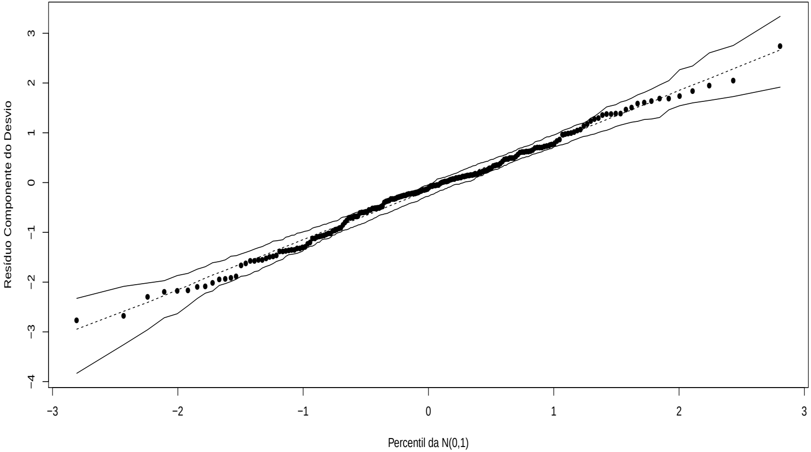
- Modelo 3: $Y_i \overset{ind.}{\sim} \text{gama}(\mu_i, \phi)$ (simulado); Modelo 1, com $\mu_i = \eta_i$ (ajustado).
- Modelo 4: $Y_i \overset{ind.}{\sim} NA(\mu_i, \phi, -0,20)$ (simulado); Modelo 1 (ajustado).
- Modelo 5: $Y_i \overset{ind.}{\sim} TA(\mu_i, \phi, -0,20, 3)$ (simulado); Modelo 1 (ajustado).

$NA(\mu, \phi, \lambda)$ e $TA(\mu, \phi, \lambda, \nu)$ representam, respectivamente, a distribuição normal assimétrica e t de Student assimétrica com parâmetro de localização μ, de dispersão ψ, de assimetria λ e graus de liberdade ν.

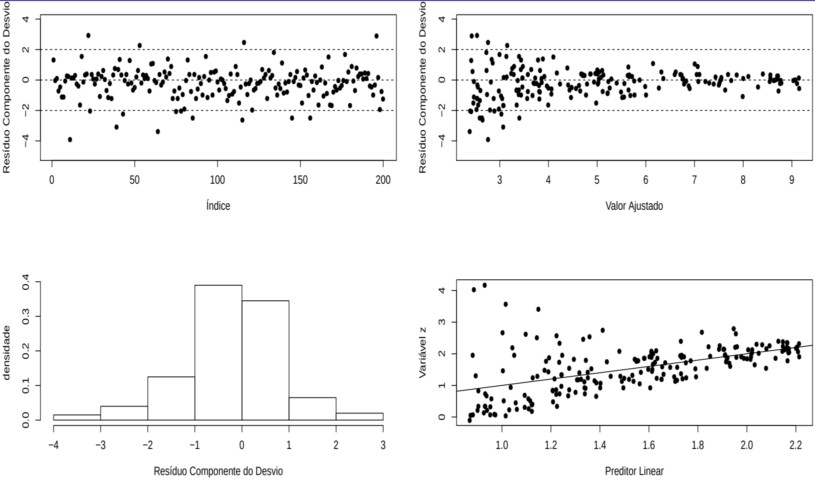
Modelo 1: gráficos de diagnóstico



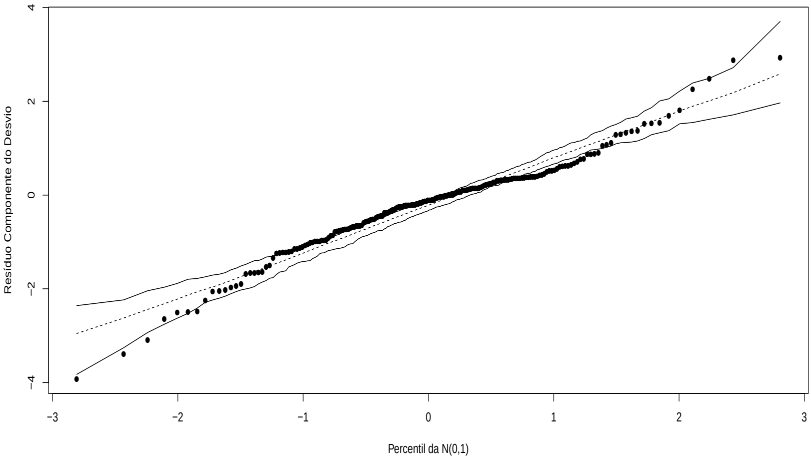
Modelo 1: gráficos de envelope



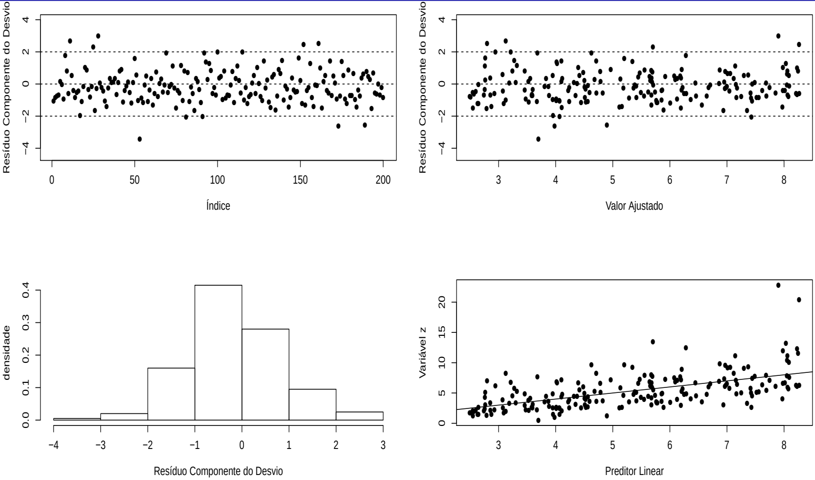
Modelo 2: gráficos de diagnóstico



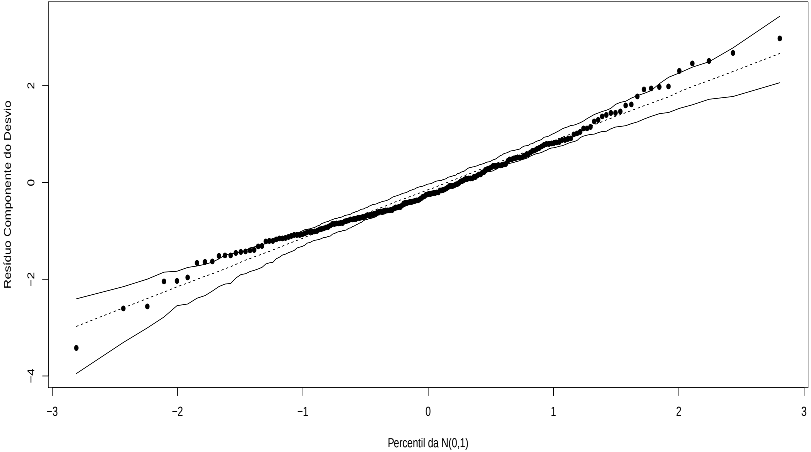
Modelo 2: gráficos de envelope



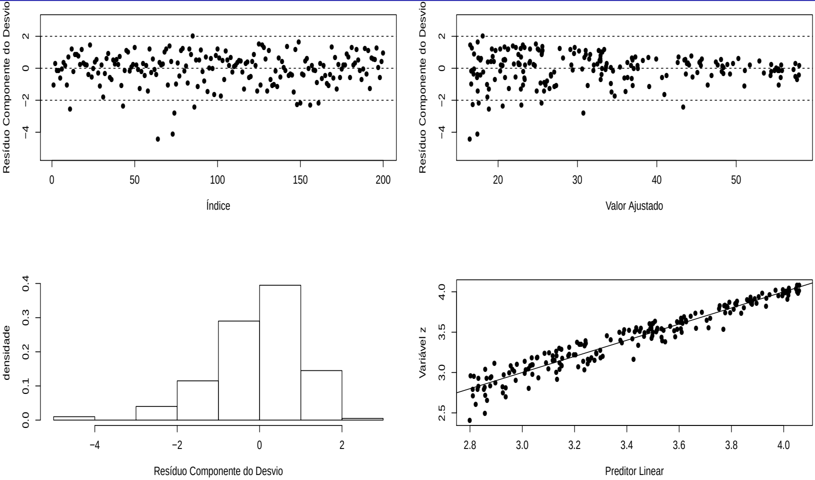
Modelo 3: gráficos de diagnóstico



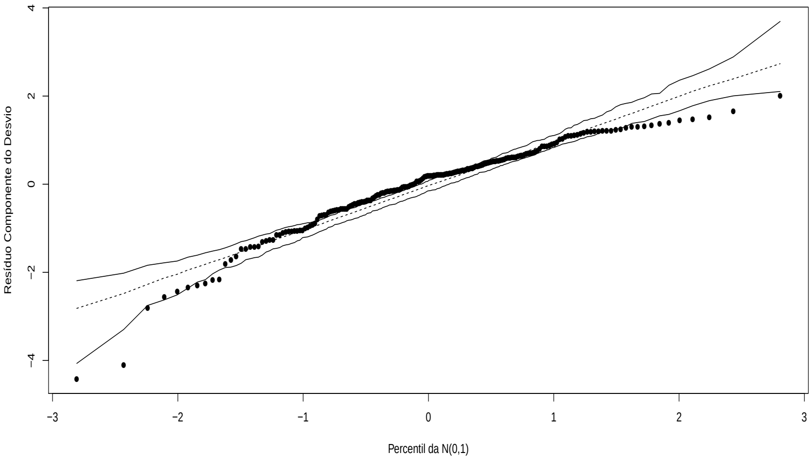
Modelo 3: gráficos de envelope



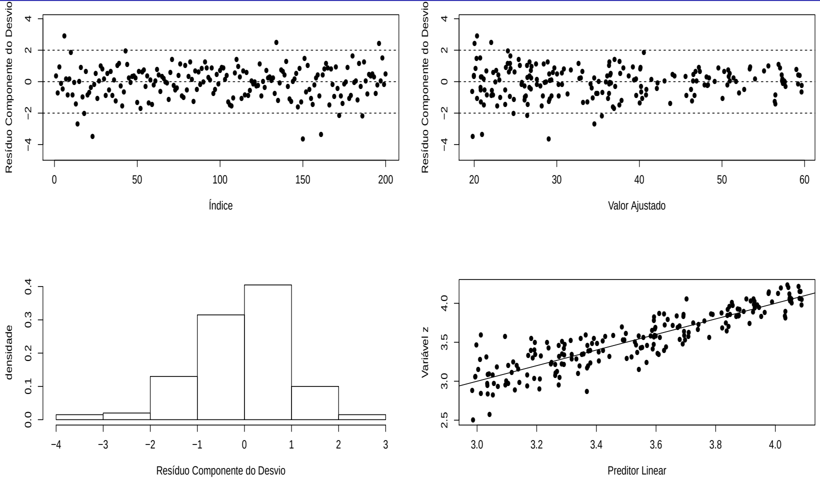
Modelo 4: gráficos de diagnóstico



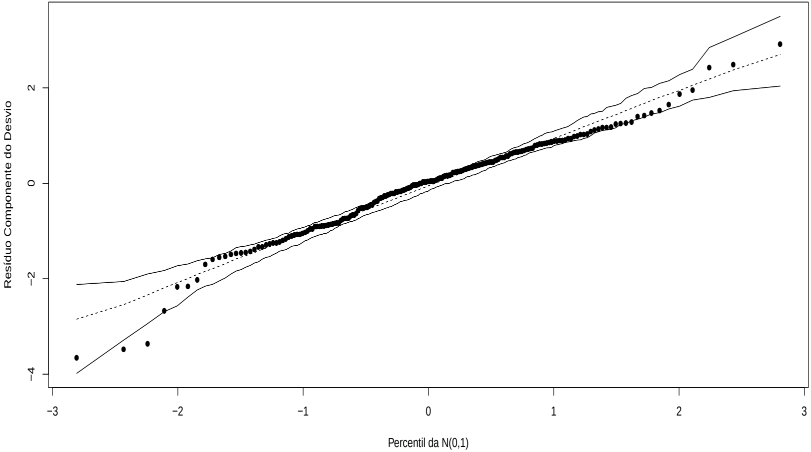
Modelo 4: gráficos de envelope



Modelo 5: gráficos de diagnóstico



Modelo 5: gráficos de envelope



Exemplo 6: pagamentos de seguros na Austrália

- Vamos considerar parte dos dados descritos em de Jong e Heller (2008, pgs. 14-15) referentes aos valores pagos de seguros individuais (em dólares australianos) por danos com acidentes pessoais no período de julho de 1989 a junho de 1999 (veja também Paula 2013).
- As análises serão restritas ao período de janeiro de 1998 a junho de 1999, correspondendo a um total de 769 seguros pagos.
- Variável reposta: valor pago ao segurado - (vpago).

Exemplo 6: pagamentos de seguros na Austrália

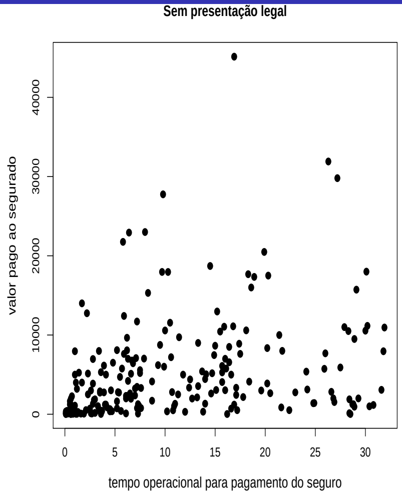
- Variáveis explicativas:
  - *legrep*, representação legal - (0: não, 1: sim) e *optime* - tempo operacional para pagamento do seguro.
  - Essa última variável assume valores no intervalo (0, 100) e por exemplo um valor 23 significa que 23% dos seguros foram pagos antes do seguro em análise.
  - Como estamos considerando apenas parte dos dados (referentes aos últimos 18 meses), os valores de optime variam de 0,1 a 31,9.
- Veja o arquivo insurance.prn.

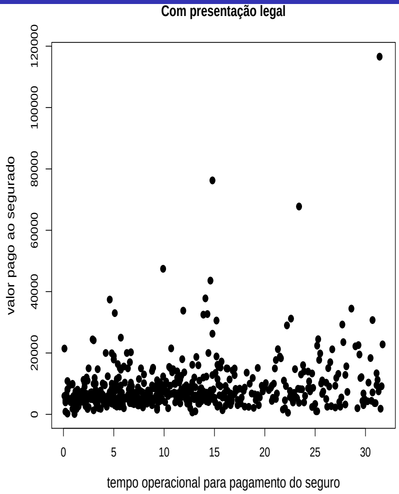
Banco de dados

Observação	vpago	optime	legrep
1	119,60	0,10	0
2	290,00	0,10	0
3	30,00	0,10	0
4	40,00	0,10	0
5	21450,00	0,10	1
⋮	⋮	⋮	⋮
769	10947,00	31,90	0

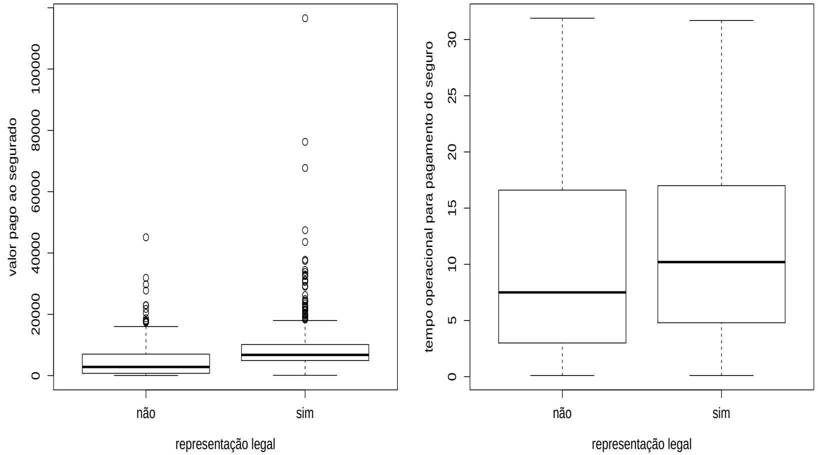
Vamos considerar que cada observação corresponde à uma ocorrência (observações independentes).

Gráficos de dispersão





Boxplots


valor pago ao segurado
0
20000
40000
60000
80000
100000
não
sim
representação legal
tempo operacional para pagamento do seguro
0
5
10
15
20
25
30
não
sim
representação legal

Medidas resumo

	Variável			
	vpago		optime	
Medida-resumo	nao	sim	nao	sim
média	5069,27	8996,41	10,74	11,86
DP	6338,01	8790,71	9,18	8,53
Var.	40170367,59	77276620,67	84,26	72,69
CV(%)	125,02	97,71	85,45	71,90
CA	2,51	5,58	0,73	0,63
Min.	30,0	109,00	0,10	0,10
Max.	45132,03	116586,72	31,90	31,70

Modelo

$$
\begin{aligned}
Y_{ij} &\overset{ind.}{\sim} \text{gama}(\mu_{ij}, \phi), \\
i &= 1, 2\,(\text{legpre}, 1 - \text{não}, 2 - \text{sim});\,; j = 1, 2, ..., n_i \\
& \quad (\text{observação}, n_1 = 227, n_2 = 542) \\
\ln(\mu_{ij}) &= \alpha + \alpha_i + (\beta + \beta_i)(x_{ij} - 0,1) + (\gamma + \gamma_i)(x_{ij} - 0,1)^2, \\
\alpha_1 &= \beta_1 = \gamma_1 = 0 \\
\mathcal{E}(Y_{ij}) = \mu_i \quad &; \quad \mathcal{V}(Y_{ij}) = \frac{\mu_i^2}{\phi}
\end{aligned}
$$

Y_{ij}: valor do seguro pago na j-ésima ocorrência sob a i-ésima representação legal. x_{ij}: tempo operacional para pagamento do seguro na j-ésims ocorrência sob a i-ésima representação legal.

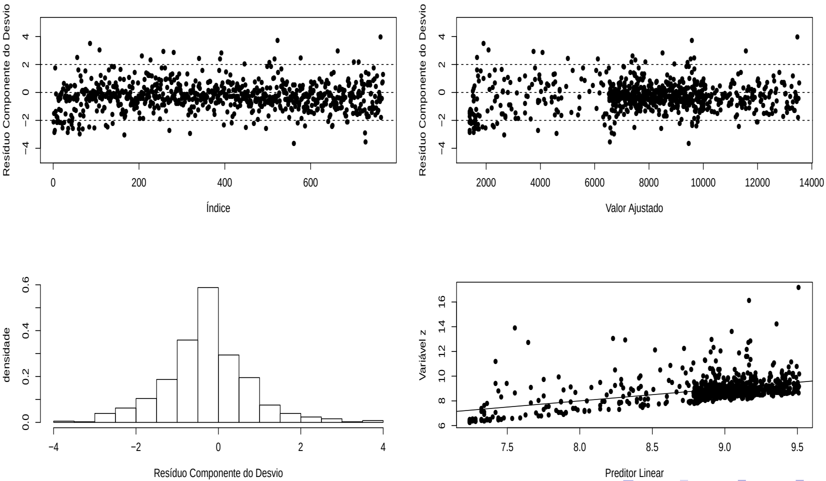
Modelo

- e^{α} : valor médio do seguro pago sem representação legal com o tempo operacional igual à 0,1.
- $e^{\alpha+\alpha_2}$: valor médio do seguro pago com representação legal com o tempo operacional igual à 0,1.
- $-\dfrac{\beta}{2\gamma}+0,1$: valor do tempo operacional para o qual o log do valor do esperado (consequentemente o próprio valor esperado) esperado do seguro pago é máximo/mínimo, sem representação legal.

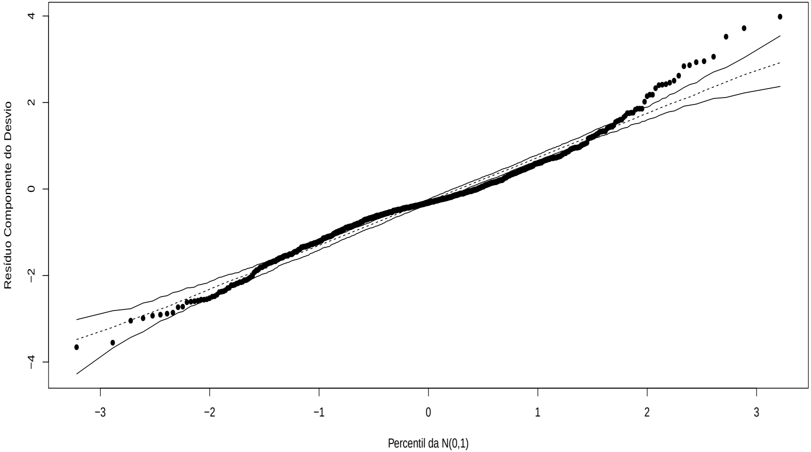
Modelo

- $-\dfrac{\beta + \beta_2}{2(\gamma + \gamma_2)} + 0,1$: valor do tempo operacional para o qual o log do valor esperado (consequentemente o próprio valor esperado) esperado do seguro pago é máximo/mínimo, com representação legal.
- Exercício: suponha que $\gamma = \gamma_2 = 0$. Inteprete e^{β} e $e^{\beta+\beta_2}$.
- $D(\boldsymbol{y}; \widetilde{\boldsymbol{\mu}}) = 856,25 (p = 0,0104)$ (considerando-se a aproximação pela $\chi^2_{(763)}$ adequada), o que indica que o modelo não se ajustou de modo satisfatório aos dados.
- Neste caso, a estimativa de ϕ é pequena e a aproximação pela χ^2 pode não ser razoável.

Gráficos de diagnóstico



Envelope para os resíduos



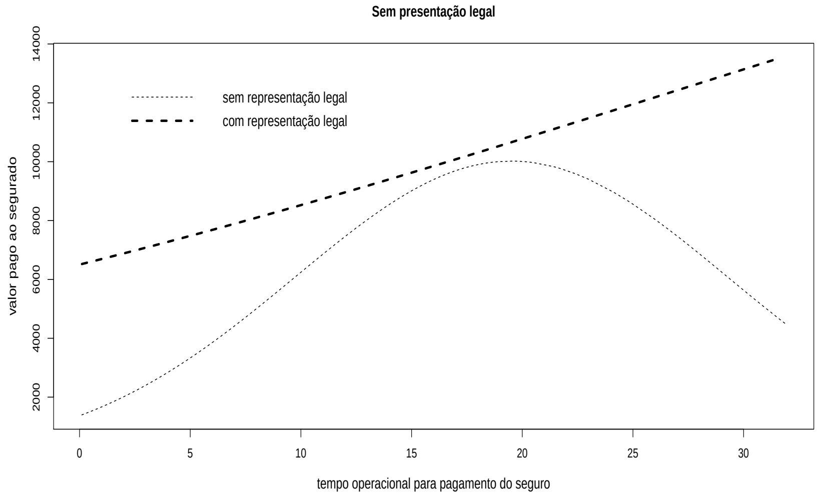
Estimativas dos parâmetros

Parâmetro	Estimativa	EP	IC(95%)	Estat. Z	p-valor
α	7,241	0,112	[7,021 ; 7,462]	64,440	<0,0001
β	0,203	0,021	[0,162 ; 0,244]	9,679	<0,0001
γ	-0,005	0,001	[-0,007 ; -0,004]	-7,313	<0,0001
α_2	1,542	0,144	[1,260 ; 1,823]	10,726	<0,0001
β_2	-0,174	0,026	[-0,225 ; -0,124]	-6,732	<0,0001
γ_2	0,005	0,001	[0,003 ; 0,007]	5,774	<0,0001
ϕ	1,407	0,065	[1,280 ; 1,535]	-	-

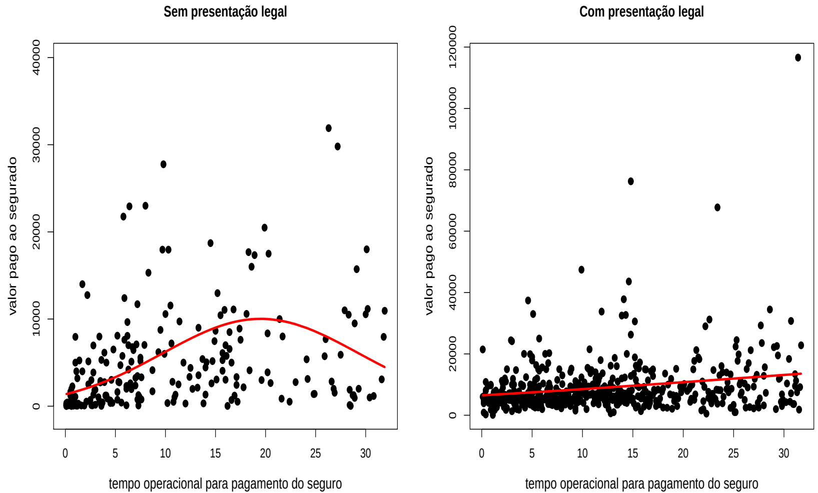
Estimativas dos parâmetros

- O modelo gama é preferível ao modelo exponencial.
- Todos os parâmetros de regressão significativos.
- Curvas: $\ln(\widetilde{\mu}_{1j}) = 7,241 + 0,203(x_{1j} - 0,1) - 0,005(x_{1j} - 0,1)^2$ (sem representação legal) e $\ln(\widetilde{\mu}_{2j}) = 8,783 + 0,029(x_{2j} - 0,1) - 0,00018(x_{2j} - 0,1)^2$ (com representação legal),
- A curva do primeiro grupo (sem representação legal) apresenta um ponto de máximo, enquanto que a do segundo é, essencialmente, uma reta.

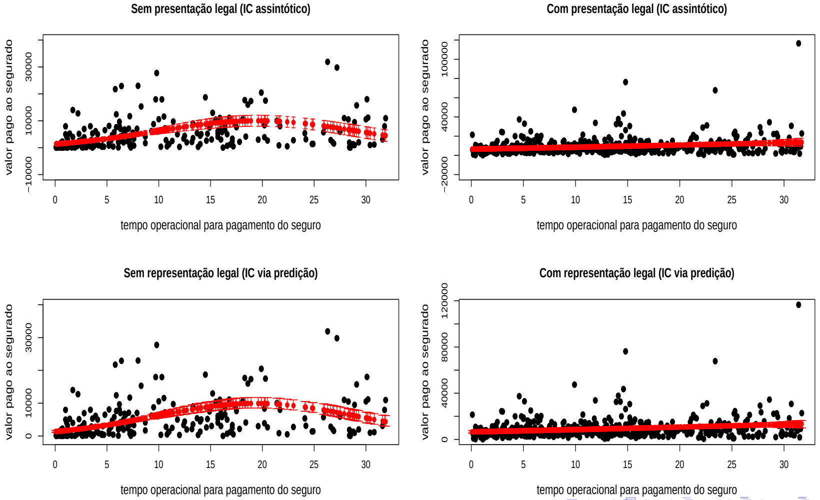
Médias preditas pelo modelo



Médias preditas pelo modelo



Médias preditas pelo modelo com intervalos de confiança



Conclusões

- O modelo não se ajustou bem aos dados. Alternativas: modelo normal assimétrico, modelo t de Student assimétrico, modelo Birnbaum-Saunders baseado na normal assimétrica ou na t de Student assimétrica (com parte do preditor linear correspondendo à uma estrutura não paramétrica, i.e., splines, ondaletas etc).

Conclusões

- Contudo, supondo que o modelo tivesse se ajustado adequadamente, poderíamos concluir que:
  - Para o grupo com representação é esperado que quanto maior o tempo operacional para pagamento do seguro, maior será o valor pago, enquanto que para o grupo sem representação há um valor do tempo operacional para o qual o valor é máximo, ou seja, 19,51 [17,67;21,35].
  - Exercício: utilizar o método Delta para encontrar o intervalo de confiança para $-\dfrac{\boldsymbol{\beta}}{2\gamma} + 0,1$.

Conclusões

- O valor máximo do seguro pago é da ordem de: 10.021,04 [8.326,24 ; 11.805,85].
- Exercício: utilizar o método Delta para encontrar o intervalo de confiança para o valor máximo do seguro pago.