

Modelos lineares generalizados mistos: parte 2

Prof. Caio Azevedo

Exemplo 5: condição respiratória (Myers, Montgomery e Vining (2002, Seção 6.5))

- Um total de 56 pacientes foi considerado no estudo sendo que 27 receberam o tratamento com uma droga ativa enquanto que os 29 pacientes restantes receberam placebo.
- Cada paciente foi observado em quatro ocasiões (não foi informado a distância cronológica entre as visitas) em que mediu-se a condição respiratória (boa ou ruim) (0 e 1, respectivamente).
- Foram também observados o gênero e a idade (em anos) de cada paciente além da pré-existência de um nível base (sim ou não).

Cont.

- Estudo regular ou irregular: tempo entre as visitas não informado.
- Balanceado em relação a condição de avaliação e desbalanceado em relação ao grupo (29 - placebo e 27 - droga ativa).
- Completo.

Medidas resumo: por tratamento

Tratamento	Período	Média	DP	Var.	CV(%)	n
placebo	1	0.69	0.47	0.22	68.27	29
placebo	2	0,62	0,49	0,24	79,56	29
placebo	3	0,72	0,45	0,21	62,81	29
placebo	4	0,52	0,51	0,26	98,32	29
droga	1	0,81	0,40	0,16	48,58	27
droga	2	0,48	0,51	0,26	105,75	27
droga	3	0,19	0,40	0,16	213,76	27
droga	4	0,04	0,19	0,04	519,62	27

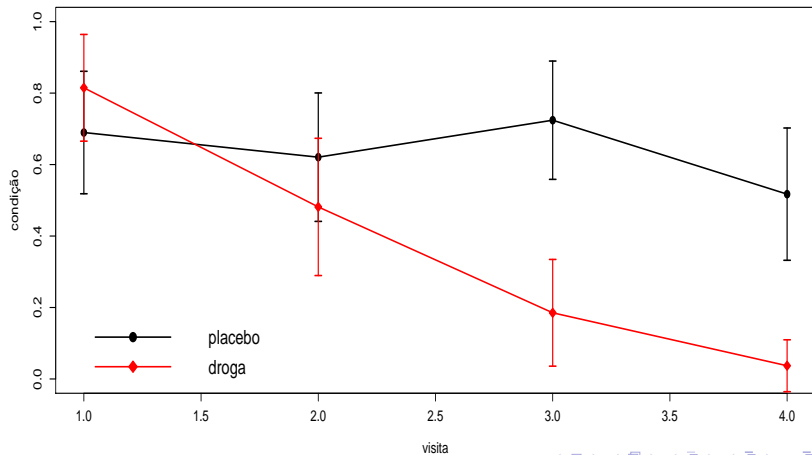
Medidas resumo: por gênero

Tratamento	Período	Média	DP	Var.	CV(%)	n
feminino	1	1,00	0,00	0,00	0,00	7
feminino	2	1,00	0,00	0,00	0,00	7
feminino	3	1,00	0,00	0,00	0,00	7
feminino	4	0,57	0,53	0,29	93,54	7
masculino	1	0,71	0,46	0,21	63,90	49
masculino	2	0,49	0,51	0,26	103,12	49
masculino	3	0,39	0,49	0,24	126,96	49
masculino	4	0,24	0,43	0,19	177,41	49

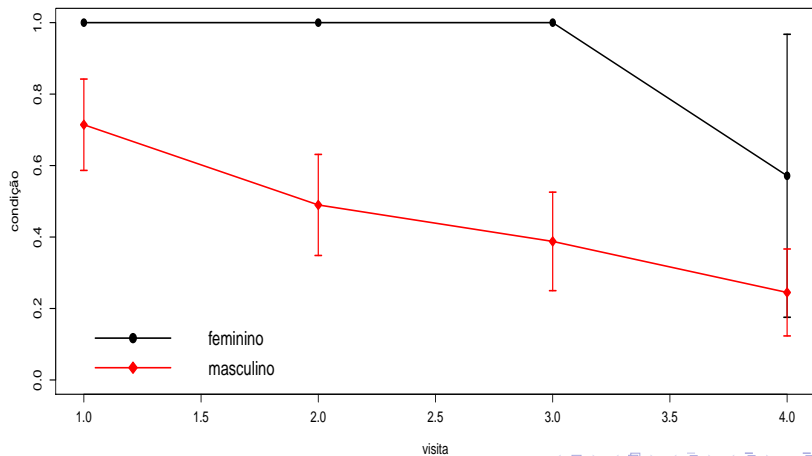
Medidas resumo: por nível base

Nível base	Período	Média	DP	Var.	CV(%)	n
ausência	1	0,71	0,49	0,24	68,31	7
ausência	2	0,71	0,49	0,24	68,31	7
ausência	3	0,71	0,49	0,24	68,31	7
ausência	4	0,43	0,53	0,29	124,72	7
presença	1	0,76	0,43	0,19	57,54	49
presença	2	0,53	0,50	0,25	95,03	49
presença	3	0,43	0,50	0,25	116,67	49
presença	4	0,27	0,45	0,20	168,13	49

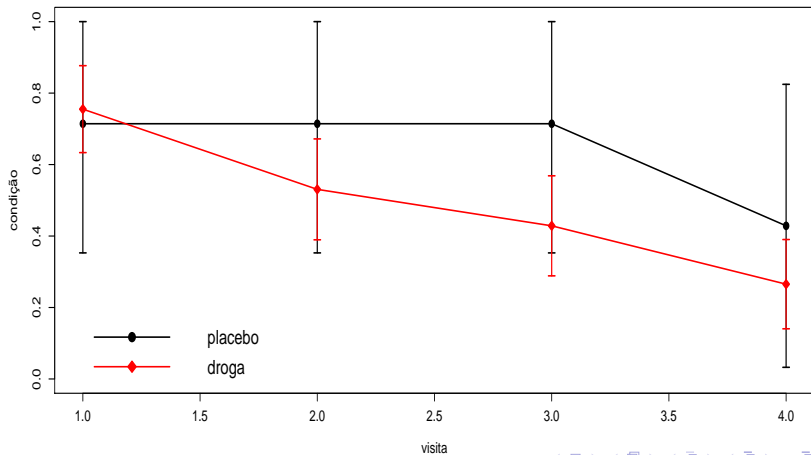
Perfis médios: por tratamento



Perfis médios: por gênero



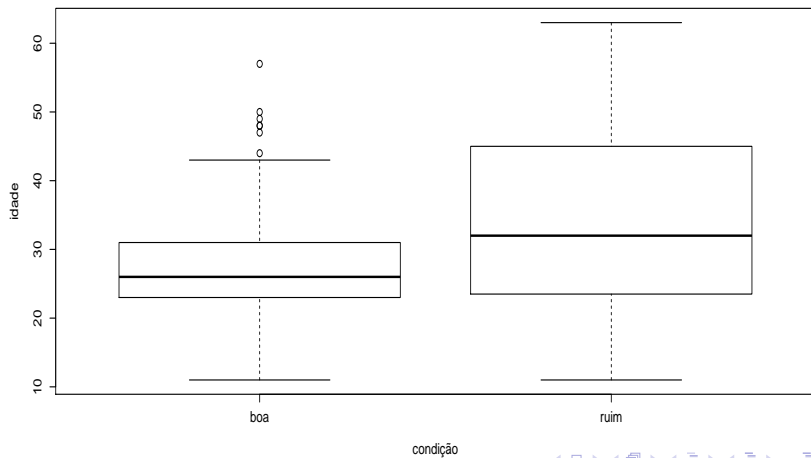
Perfis médios: por nível base



Medidas resumo idade por condição

	condição	
Estatística	boa	ruim
media	27,17	33,31
dp	8,44	13,17
vari	71,19	173,30
cv	31,06	39,52
minimo	11,00	11,00
mediana	26,00	32,00
maximo	57,00	63,00
n	109	115

Box plot: idade em função da condição



Correlações tetracóricas

	1	2	3	4
1	1,000	0,740	0,354	0,556
2	0,740	1,000	0,685	0,417
3	0,354	0,685	1,000	0,778
4	0,556	0,417	0,778	1,000

M.R. (geral/efeitos fixos) para dados binários

$$Y_i \stackrel{ind.}{\sim} \text{Bernoulli}(p_i)$$

$$F^{-1}(p_i) = \sum_{j=1}^p \beta_j x_{ji} \rightarrow p_i = F\left(\sum_{j=1}^p \beta_j x_{ji}\right), i = 1, 2, \dots, n$$

- 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.
- $F(\cdot)$: função de distribuição acumulada de alguma variável aleatória (contínua) com suporte em \mathcal{R} . $F^{-1}(\cdot)$ é conhecida como função de ligação.
- Modelo com intercepto: $x_{1i} = 1, \forall i$.

M.R. logística para dados binários com uma única covariável

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

- Y_i : ocorrência (1) ou não (0) de algum evento.
- $F(\cdot)$: corresponde à fda de uma distribuição logística padrão (portanto o nome regressão logística). Nesse caso, o $\text{logito}(\cdot)$ é a função de ligação.

Cont.

- Interpretação dos parâmetros. Defina $l(p_i) = \text{logito}(p_i)$.
- Se $x_{1j} = 0$, então $p_i = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$.
- Defina $l(p_{i+1}) = \beta_0 + \beta_1(x_{1i} + 1) + \beta_2 x_{2i}$ e
 $l(p_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$. Então
 $l(p_{i+1}) - l(p_i) = \beta_1 \rightarrow \frac{p_{i+1}/(1 - p_{i+1})}{p_i/(1 - p_i)} = e^{\beta_1}$ (**razão de chances**).

M.R. para dados binários com um único fator

$$\begin{aligned} Y_{ij} &\stackrel{ind.}{\sim} \text{Bernoulli}(p_{ij}) \\ \text{logito}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) &= \mu + \alpha_i, \alpha_1 = 0 \\ \rightarrow p_i &= \frac{e^{\mu+\alpha_i}}{1+e^{\mu+\alpha_i}}, i = 1, 2, \dots, k; j = 1, 2, \dots, n_j \end{aligned}$$

Cont.

- Interpretação dos parâmetros. Defina $l(p_i) = \text{logito}(p_i)$.

- $p_1 = \frac{e^\mu}{1 + e^\mu}$.

- Defina $l(p_i) = \mu + \alpha_i$. Então

$$l(p_i) - l(p_{i'}) = \alpha_i - \alpha_{i'} \rightarrow \frac{p_i/(1 - p_i)}{p_{i'}/(1 - p_{i'})} = e^{\alpha_i - \alpha_{i'}} \text{ (razão de chances)}.$$

- Especificamente, $\frac{p_i/(1 - p_i)}{p_1/(1 - p_1)} = e^{\alpha_i}$ (comparando o grupo de referência com os demais).

M.R. (geral/efeitos aleatórios) para dados binários

$$Y_{ij} \stackrel{ind.}{\sim} \text{Bernoulli}(p_{ij})$$
$$F^{-1}(p_{ij}) = \sum_{k=1}^p \beta_k x_{kij} + \sum_{r=1}^q z_{rij} b_{rj} \rightarrow p_{ij} = F \left(\sum_{k=1}^p \beta_k x_{kij} + \sum_{r=1}^q z_{rij} b_{rj} \right)$$

, $i = 1, \dots, k_j; j = 1, 2, \dots, n$

- Y_{ij} : ocorrência (1) ou não (0) de algum evento.
- x_{kij} : valor da variável explicativa k associada ao indivíduo j no instante i ;
- β_k : parâmetro associado ao impacto de cada covariável na probabilidade de ocorrência do supracitado evento (mantendo-se a parte aleatória do preditor fixa).

M.R. (geral/efeitos aleatórios) para dados binários

- $F(\cdot)$: função de distribuição acumulada de alguma variável aleatória (contínua) com suporte em \mathcal{R} . $F^{-1}(\cdot)$ é conhecida como função de ligação.
- Modelo com intercepto: $x_{1i} = 1, \forall i$.
- As suposições para \mathbf{b}_j são as mesmas feitas anteriormente (para o MLGM geral).

M.R. para os dados do exemplo (regressão logística)

$$\begin{aligned} Y_{ij} &\stackrel{ind.}{\sim} \text{Bernoulli}(p_{ij}) \\ \text{logito}(p_{ij}) &= \ln\left(\frac{p_{ij}}{1-p_{ij}}\right) \\ &= \alpha + \beta_1 idade_j + \beta_2 trat_j + \beta_3 gen_j + \beta_4 base_j + \beta_5(\text{periodo}_{ij} - a) \\ &\quad + b_j \\ \rightarrow p_{ij} &= \frac{e^{\alpha + \beta_1(idade_j - a) + \beta_2 trat_j + \beta_3 gen_j + \beta_4 base_j + \beta_5(\text{periodo}_{ij} - b) + b_j}}{1 + e^{\alpha + \beta_1(idade_j - a) + \beta_2 trat_j + \beta_3 gen_j + \beta_4 base_j + \beta_5(\text{periodo}_{ij} - b) + b_j}}, \\ &\quad i = 1, 2, 3, 4; j = 1, 2, \dots, 59 \end{aligned}$$

em que $b_j \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$, $a = 30,32143$ (média das idades) e $b = 1$ (período 1).

M.R. para os dados do exemplo (regressão logística)

- Y_{ij} : condição : 1 - ruim; 0 - boa.
- $idade_j$: idade (em anos), $trat_j$: (=1 droga ativa, = 0 placebo); gen_j : (=0 feminino, =1 masculino); $base_j$: (=0 ausência do nível base, =1 presença do nível base), do i -ésimo paciente e $periodo_{ij}$: instante no qual a visita foi realizada, do j -ésimo paciente no i -ésimo instante.

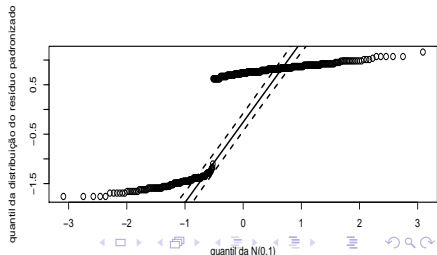
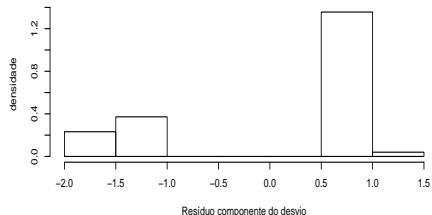
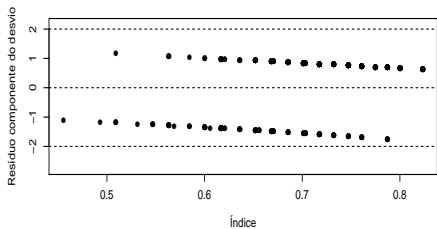
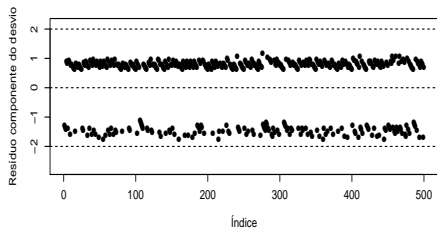
Interpretação dos parâmetros

- Para cada covariável, se considerarmos o mesmo indivíduo, os logitos funcionam de forma parecida com o caso anterior, mantendo-se cada uma das outras covariáveis fixas.
- Razão de chances entre tratamentos (placebo/droga): $\psi_t = e^{\beta_2}$.
- Razão de chances entre gêneros (masculino/feminino): $\psi_g = e^{\beta_3}$.
- Razão de chances entre bases (presença/ausência): $\psi_b = e^{\beta_4}$.
- Razão de chances entre para o aumento em uma ano na idade:
 $\psi_i = e^{\beta_1}$.
- Razão de chances entre um determinado período e o subsequente:
 $\psi_p = e^{\beta_5}$.

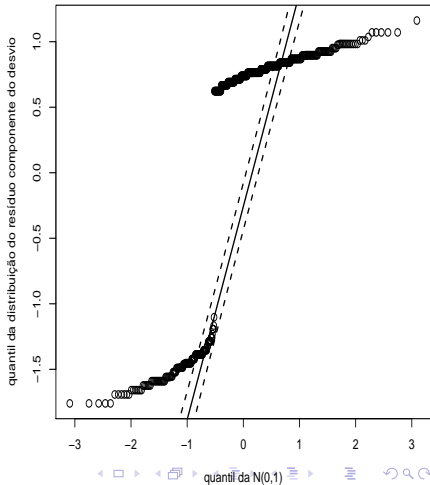
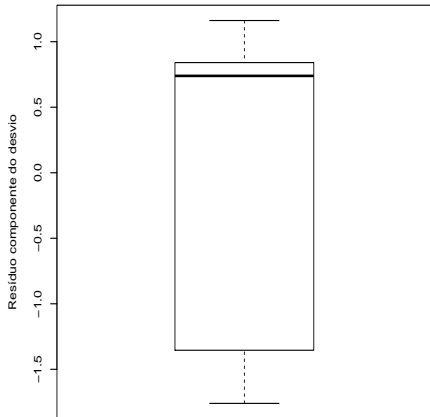
Estudo de simulação

- Simulou-se dois modelos.
- Modelo 1: $Y_{ij}|b_j \stackrel{ind.}{\sim} \text{Bernoulli}(p_i)$, $i = 1, 2, \dots, 5; j = 1, 2, \dots, 100$,
 $\text{logito}(p_i) = 1 + 0,1x_i + b_j$; $x_i = i$; $b_j \stackrel{i.i.d.}{\sim} N(0; 0, 5)$.
- Modelo 2: $Y_{ij}|b_j \stackrel{ind.}{\sim} \text{Bernoulli}(p_i)$, $i = 1, 2, \dots, 5; j = 1, 2, \dots, 100$,
 $\mu_i = F(1 + 0,1x_i + b_j, \nu = 4)$; $x_i = i$; $b_j \stackrel{i.i.d.}{\sim} N(0; 0, 5)$, em que
 $F(., \nu = 4)$ representa a fda de uma distribuição $t_{(4)}$.

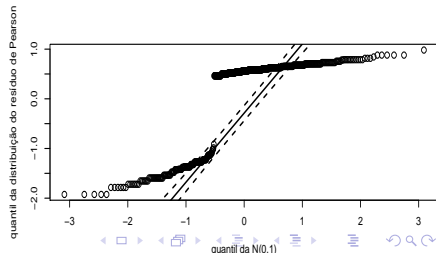
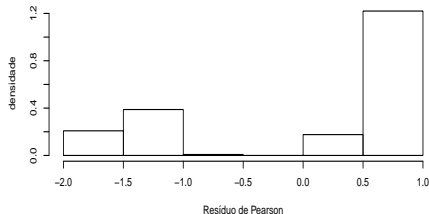
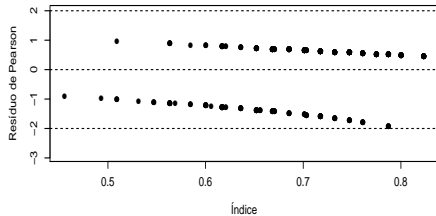
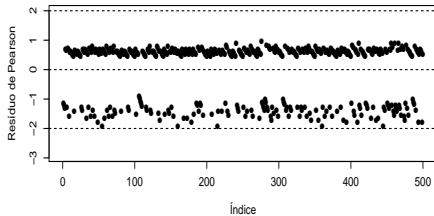
Resultados: modelo 1 - RCD



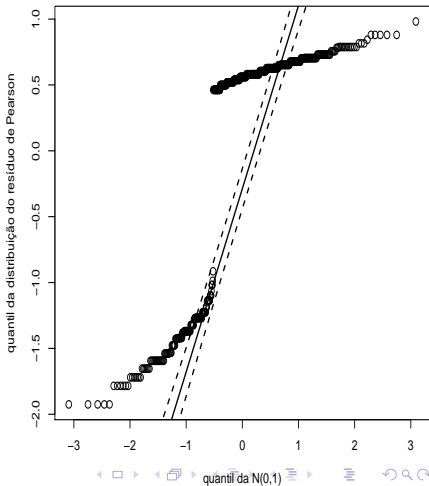
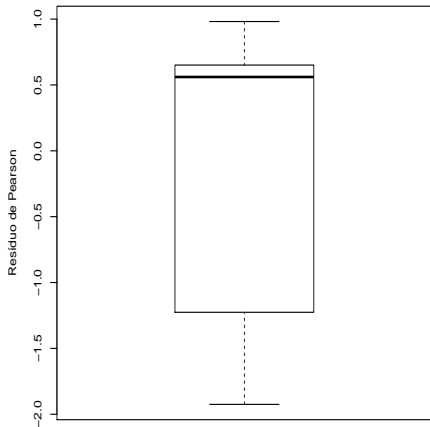
Resultados: modelo 1 - RCD



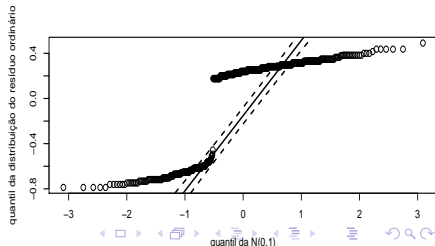
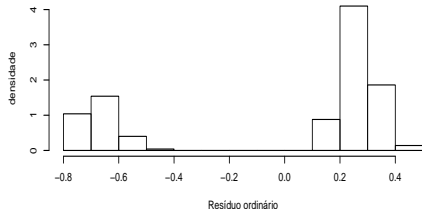
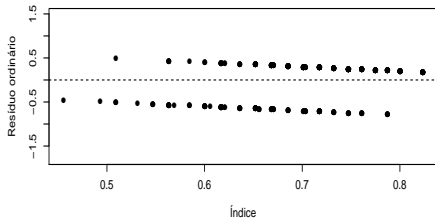
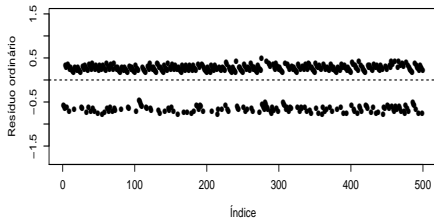
Resultados: modelo 1 - Pearson



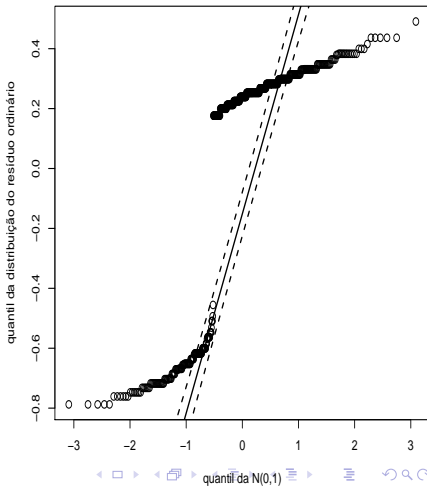
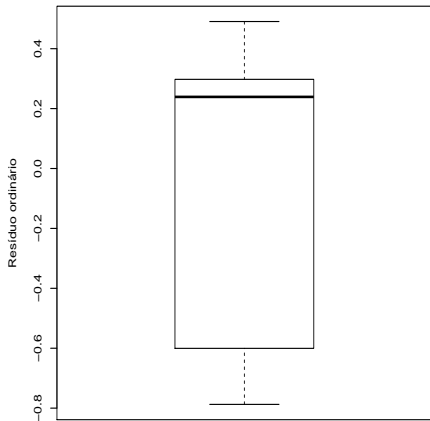
Resultados: modelo 1 - Pearson



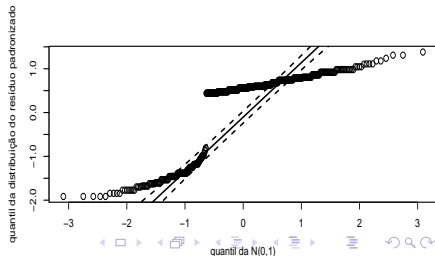
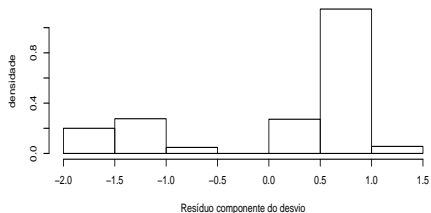
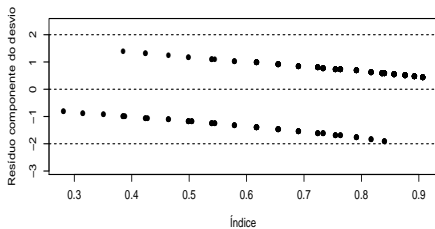
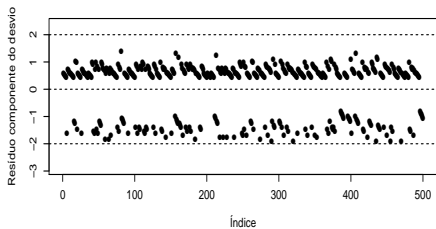
Resultados: modelo 1 - Ordinário



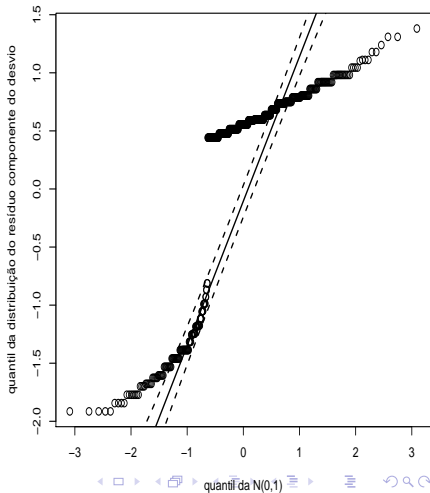
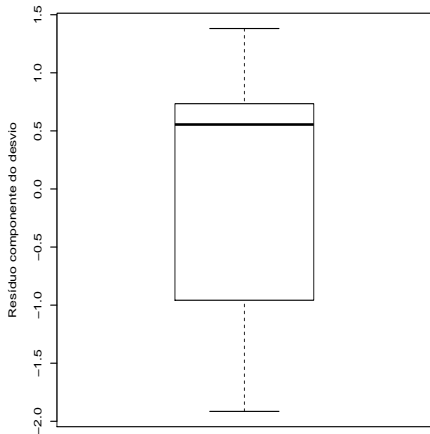
Resultados: modelo 1 - Ordinário



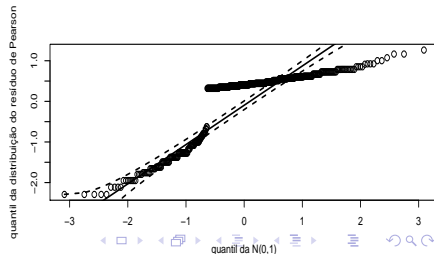
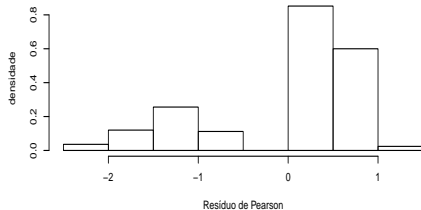
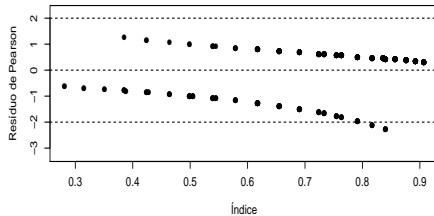
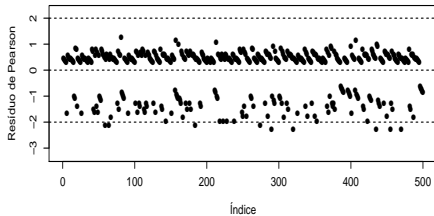
Resultados: modelo 2 - RCD



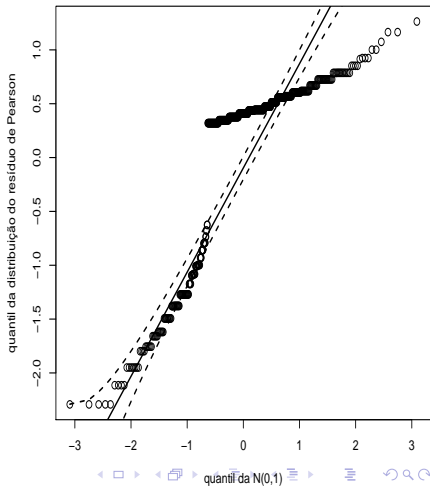
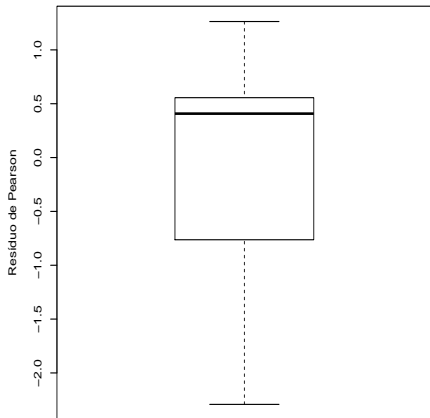
Resultados: modelo 2 - RCD



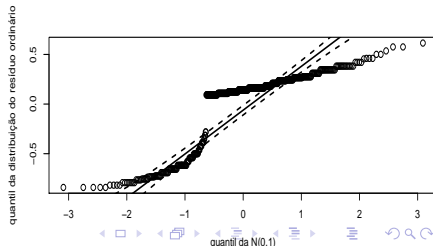
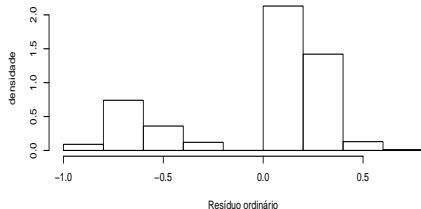
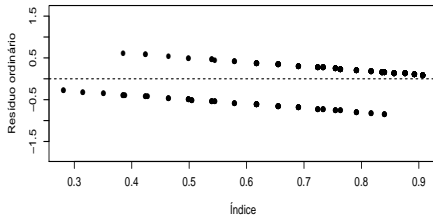
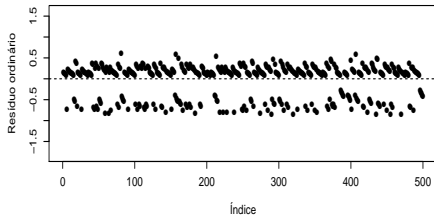
Resultados: modelo 2 - Pearson



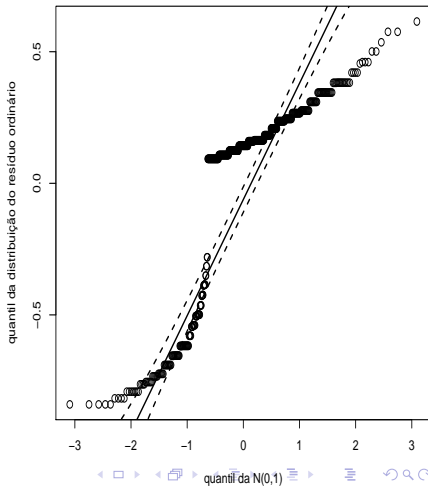
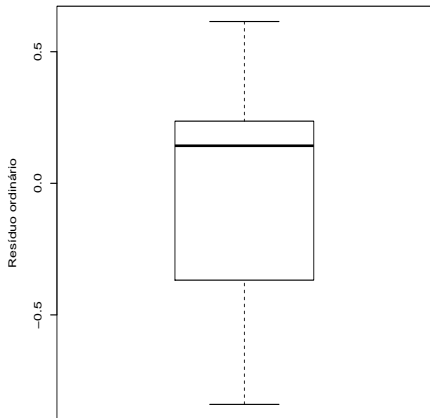
Resultados: modelo 2 - Pearson



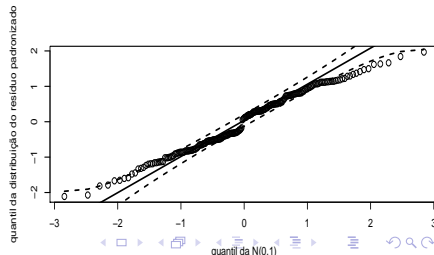
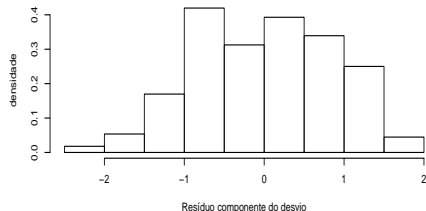
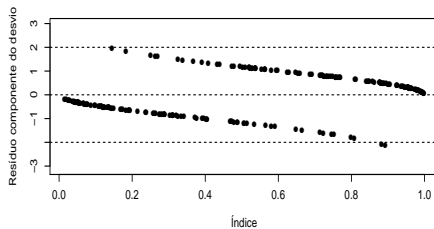
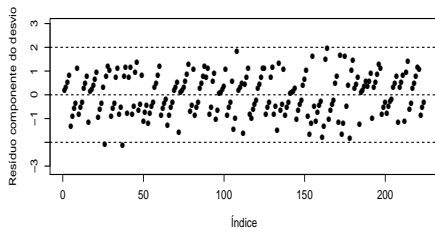
Resultados: modelo 2 - Ordinário



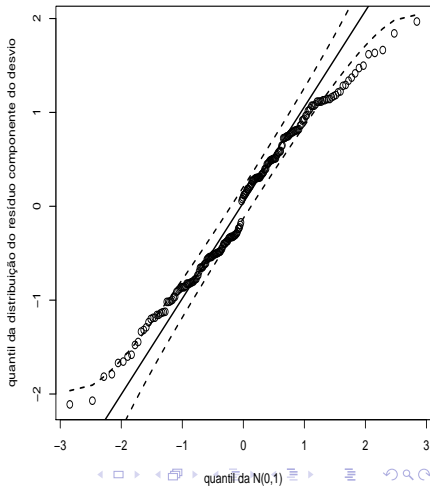
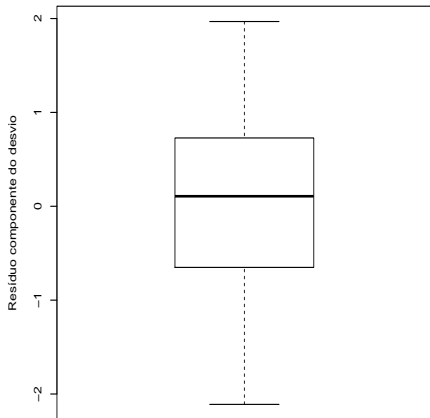
Resultados: modelo 2 - Ordinário



Resultados: modelo ajustado com QA - RCD



Resultados: modelo ajustado com QA - RCD



Alternativa para a construção do gráfico de envelopes

- Em geral, quando temos modelos de regressão, a forma mais apropriada para se construir os envelopes é simular do próprio modelo ao invés de similar da distribuição esperada para os resíduos sob a validade das hipóteses do modelo.
- Tal abordagem é ainda mais útil quando não estamos certos à respeito da distribuição dos resíduos (mesmo sob as validades da hipótese do modelo).

Procedimento para se gerar o gráfico de envelopes com o RCD

- 1) Ajuste o modelo de regressão (estima-se os parâmetros do modelo) obtendo-se as estimativas de MV ($\tilde{\beta}$) e os valores preditos ($\tilde{\mathbf{b}}_j$) e calcule o RCD para cada observação, $(t_{D_{ij}}), j = 1, 2, \dots, n, i = 1, \dots, k_j$.

Cont.

- 2) De posse das estimativas de MV e dos valores preditos, repita os passos (a) e (b) m vezes.
- a) Simule n vetores aleatórios ind. $FE(\tilde{\theta}_{ij}, \tilde{\phi})$, com
$$\tilde{\theta}_{ij} = h(g^{-1}(\mathbf{X}'_{ij}\tilde{\beta} + \mathbf{Z}_{ij}\tilde{\mathbf{b}}_j)).$$
 - b) Ajuste o modelo de regressão considerando os vetores simulados no item a) e obtenha o RCD para cada observação (i,j) em cada réplica (r).

Cont.

- 3) Ao final teremos uma matriz com os RCD's, ou seja $t_{D_{ijr}}^*$, $j=1,\dots,n,$,
 $i=1,\dots,k_j$, (amostra), $r=1,\dots,m$ (réplica).

$$\mathbf{T}_1 = \begin{bmatrix} t_{D_{111}}^* & t_{D_{112}}^* & \cdots & t_{D_{11m}}^* \\ t_{D_{121}}^* & t_{D_{212}}^* & \cdots & t_{D_{21m}}^* \\ \vdots & \vdots & \ddots & \vdots \\ t_{D_{k_n n 1}}^* & t_{D_{k_n n 2}}^* & \cdots & t_{D_{k_n n m}}^* \end{bmatrix}$$

Cont.

- 4) Dentro de cada amostra, ordena-se, de modo crescente, os RCD's, obtendo-se $t_{D_{(ij)r}}^*$ (estatísticas de ordem):

$$\mathbf{T}_2 = \begin{bmatrix} t_{D_{(11)1}}^* & t_{D_{(11)2}}^* & \cdots & t_{D_{(11)m}}^* \\ t_{D_{(21)1}}^* & t_{D_{(21)2}}^* & \cdots & t_{D_{(21)m}}^* \\ \vdots & \vdots & \ddots & \vdots \\ t_{D_{(k_n n)1}}^* & t_{D_{(k_n n)2}}^* & \cdots & t_{D_{(k_n n)m}}^* \end{bmatrix}$$

- 5) Obtem-se os limites $t_{(ij)I}^* = \min_{1 \leq r \leq m} t_{D_{(ij)r}}^*$ e $t_{(ij)S}^* = \max_{1 \leq r \leq m} t_{D_{(ij)r}}^*$,
 $r = 1, 2, \dots, m$.

Cont.

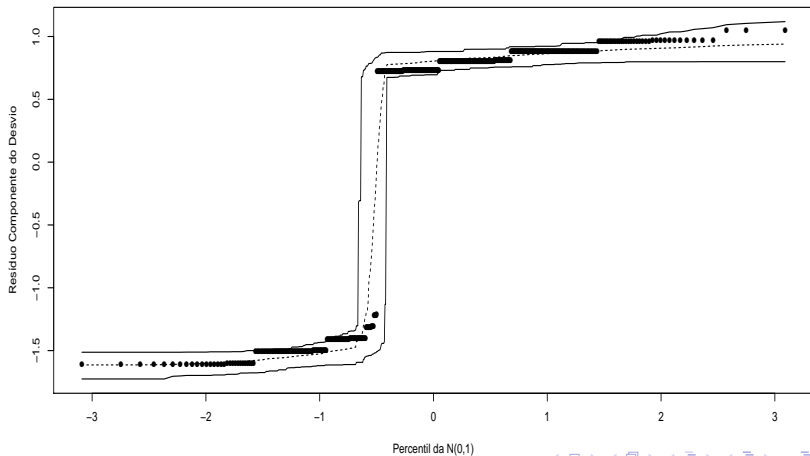
- 5) Na prática considera-se $t_{(ij)l}^* = \frac{t_{D_{(ij)(2)}}^* + t_{D_{(ij)(3)}}^*}{2}$ e
- $t_{(ij)s}^* = \frac{t_{D_{(ij)(m-2)}}^* + t_{D_{(ij)(m-1)}}^*}{2}$ (refinamento das estimativas dos limites do envelope), em que $t_{D_{(ij)(r)}}^*$ é a r -ésima estatística de ordem dentro de cada linha, $j = 1, 2, \dots, n$, $i = 1, 2, \dots, k_j$.

- Além disso, consideramos como a linha de referência

$$t_{(ij)}^* = \frac{1}{m} \sum_{r=1}^m t_{D_{(ij)r}}^*, j = 1, 2, \dots, n; i = 1, \dots, k_j.$$

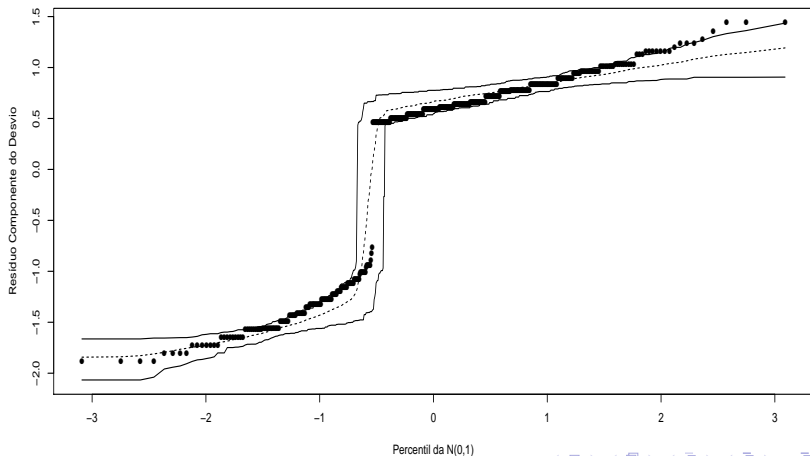
Modelo 1 (envelope simulado do modelo)

Gráfico de quantil-quantil normal



Modelo 2 (envelope simulado do modelo)

Gráfico de quantil-quantil normal



Resultados do ajuste do modelo

Laplace

Parâmetro	Estimativa	EP	Estat. z	p-valor
β_0	4,56	1,13	4,03	0,0001
β_2 (trat)	-1,62	0,55	-2,97	0,0030
β_3 (gen)	-3,09	1,05	-2,94	0,0033
β_4 (base)	0,79	0,88	0,90	0,3683
β_1 (idade)	0,07	0,02	2,92	0,0034
β_5 (período)	-1,05	0,21	-5,08	< 0,0001

$$\sigma^2 = 1,538$$

Resultados do ajuste do modelo

QA

Parâmetro	Estimativa	EP	Estat. z	p-valor
β_0	4,61	1,16	3,96	0,0001
β_2 (trat)	-1,65	0,57	-2,91	0,0037
β_3 (gen)	-3,13	1,09	-2,88	0,0040
β_4 (base)	0,81	0,92	0,89	0,3757
β_1 (idade)	0,07	0,02	2,85	0,0043
β_5 (período)	-1,07	0,21	-5,10	< 0,0001

$$\sigma^2 = 1,777$$

Resultados do ajuste do modelo

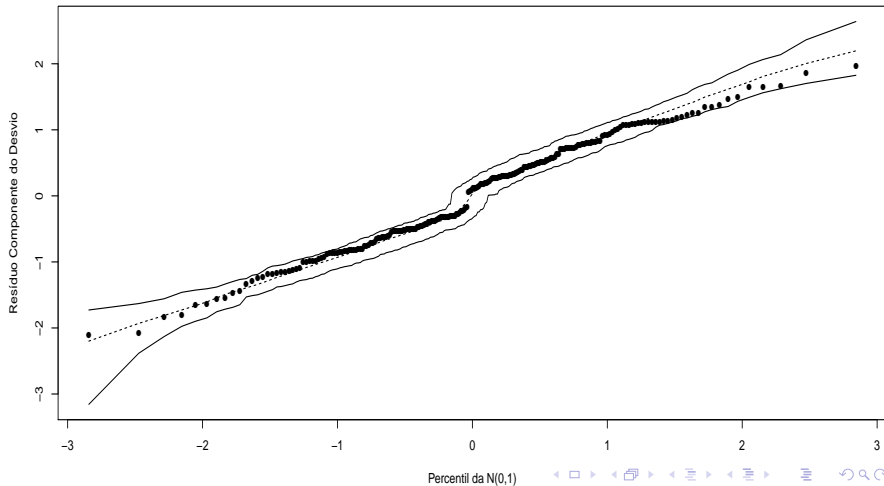
QVP

Parâmetro	Estimativa	EP	Estat. z	p-valor
β_0	4,29	0,99	4,35	< 0,0000
β_2 (trat)	-1,55	0,52	-3,00	0,0041
β_3 (gen)	-2,86	0,96	-2,97	0,0045
β_4 (base)	0,78	0,87	0,89	0,3771
β_1 (idade)	0,06	0,02	2,86	0,0061
β_5 (período)	-1,04	0,15	-6,98	< 0,0001

$$\sigma^2 = 2,059$$

Modelo ajustado com QA - RCD (env. sim. do modelo)

Gráfico de quantil-quantil normal



Resultados do ajuste do modelo reduzido (sem nível base)

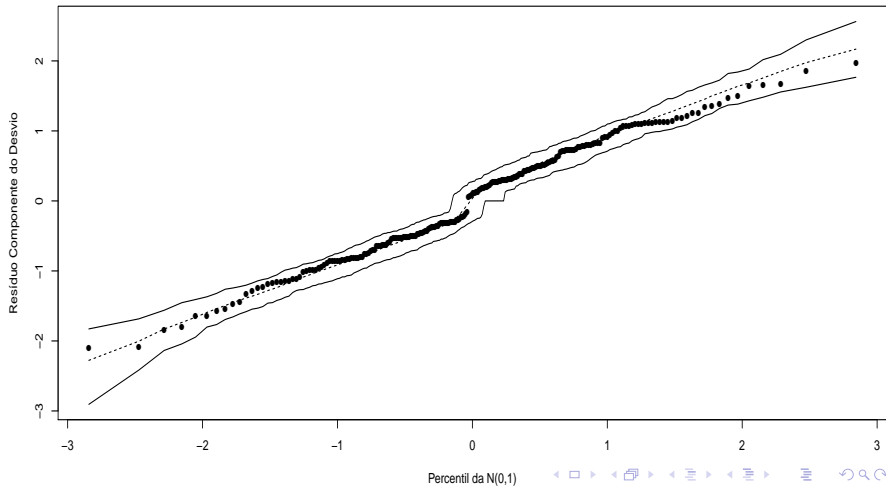
QA

Parâmetro	Estimativa	EP	Estat. z	p-valor
β_0	4.97	1.12	4.44	0.00
β_2 (trat)	-1.54	0.55	-2.79	0.01
β_3 (gen)	-2.79	0.99	-2.82	0.00
β_1 (idade)	0.07	0.02	2.85	0.00
β_5 (período)	-1.07	0.21	-5.10	0.00

$$\sigma^2 = 1,813$$

Mod. red. ajustado com QA - RCD (env. sim. do modelo)

Gráfico de quantil-quantil normal



Estudo de simulação anterior (modelo de Poisson)

- Simulou-se dois modelos.

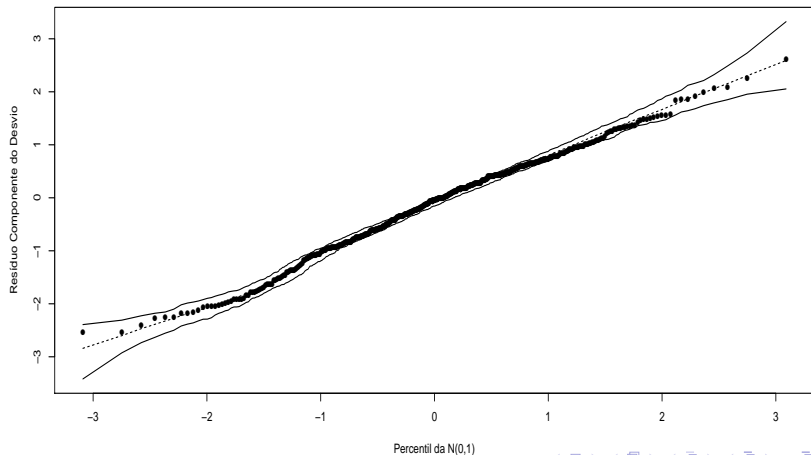
- Modelo 1: $Y_{ij}|b_j \stackrel{ind.}{\sim} \text{Poisson}(\mu_i), i = 1, 2, \dots, 5; j = 1, 2, \dots, 100,$
 $\ln \mu_i = 1 + 0,1x_i + b_j; x_i = i; b_j \stackrel{i.i.d.}{\sim} N(0; 0, 5).$

- Modelo 2:

$Y_{ij}|b_j \stackrel{ind.}{\sim} \text{Binomial-negativa}(\mu_i, \phi), i = 1, 2, \dots, 5; j = 1, 2, \dots, 100,$
 $\ln \mu_i = 1 + 0,1x_i + b_j; x_i = i; b_j \stackrel{i.i.d.}{\sim} N(0; 0, 5)$ (veja Paula (2013)).

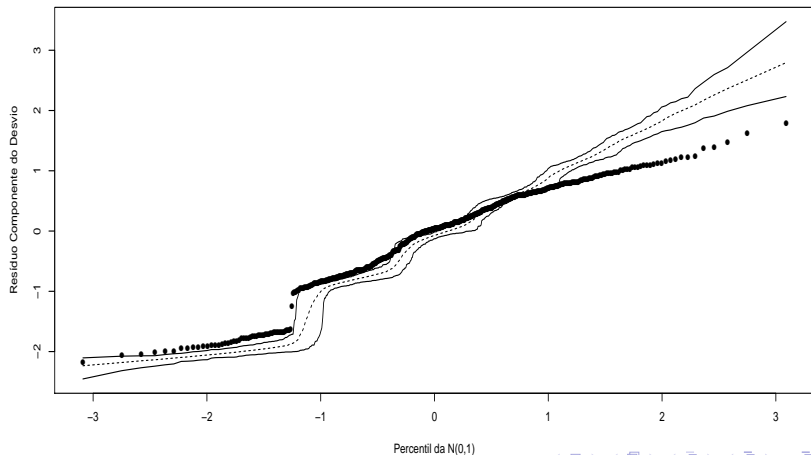
Modelo 1 (envelope simulado do modelo)

Gráfico de quantil-quantil normal



Modelo 2 (envelope simulado do modelo)

Gráfico de quantil-quantil normal



Dados reais (envelope simulado do modelo)

Gráfico de quantil-quantil normal

