

# Modelos lineares generalizados hierárquicos: 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 sexo 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

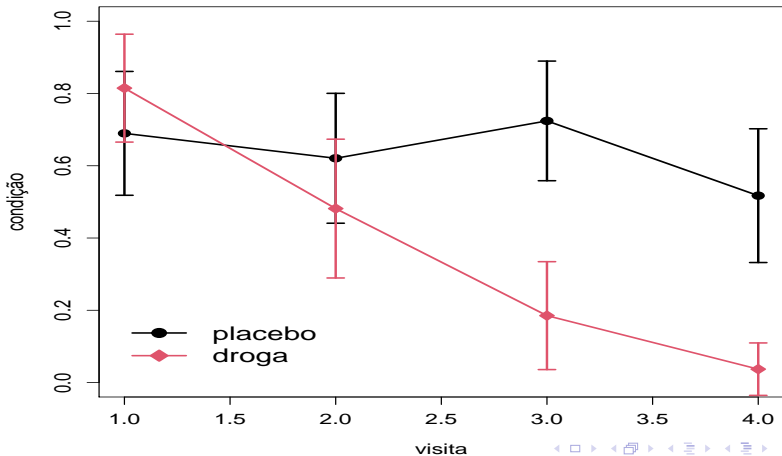
## Medidas resumo: por sexo

Sexo	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

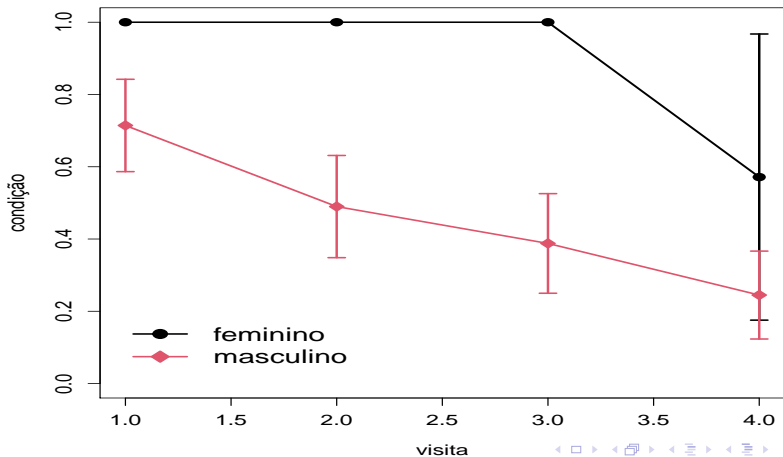
## Medidas resumo: por nível base

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

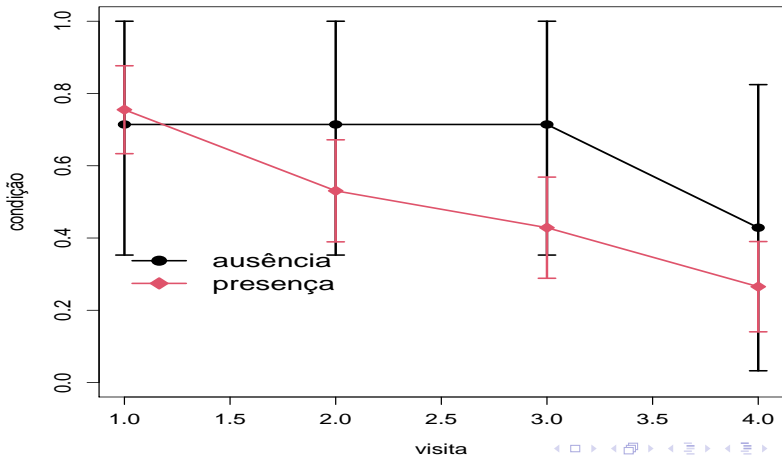


## Perfis médios: por sexo





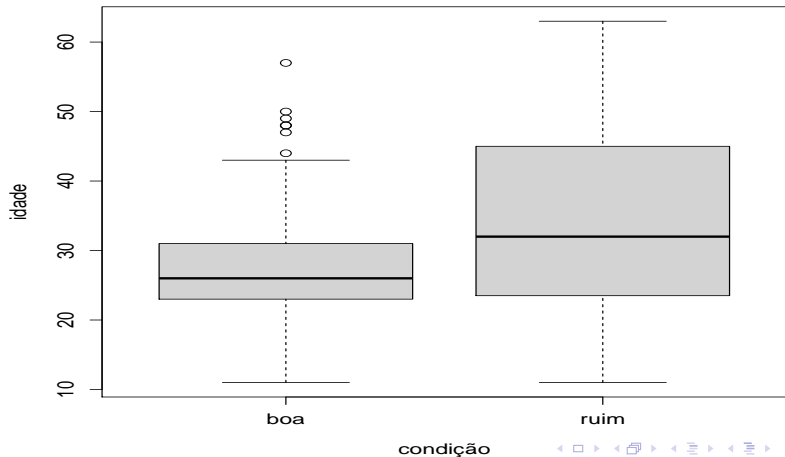
## Perfis médios: por nível base



# 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

Descrição a respeito das correlações tetracóricas: [link 1](#) e [link 2](#).

## M.R. (geral/um nível) para dados binários

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

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

- $Y_i$  : ocorrência (1) ou não (0) de algum evento.
- $x_{ji}$  : valor da variável explicativa  $j$  associada ao indivíduo  $i$ ;  $\beta_j$  : parâmetro associado ao impacto de cada covariável na probabilidade de ocorrência do supracitado evento.
- $F(\cdot)$  : (geralmente) função de distribuição acumulada de alguma variável aleatória (contínua) com suporte em  $\mathcal{R}$ .  $F^{-1}(\cdot)$  é conhecida como função de ligação. A rigor, pode ser qualquer função,  $F : \mathfrak{R} \rightarrow [0, 1]$ . 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  $l(p_{i+1}) = \beta_0 + \beta_1(x_{1i} + 1)$  e  $l(p_i) = \beta_0 + \beta_1 x_{1i}$ . 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

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



## 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/hierárquico com dois níveis) para dados binários

$$Y_{ji} | \mathbf{u}_j \stackrel{ind.}{\sim} \text{Bernoulli}(p_{ji}) \text{ (nível 1)}$$

$$\beta_j = \mathbf{W}_j \boldsymbol{\gamma} + \mathbf{u}_j \text{ (nível 2)}$$

$$F^{-1}(p_{ji}) = \mathbf{X}'_{ji} \boldsymbol{\beta}_j = \sum_{k=1}^p \beta_{jk} x_{jik} \rightarrow p_{ji} = F \left( \sum_{k=1}^p \beta_{jk} x_{jik} \right)$$

$$\boldsymbol{\beta}_j = (\beta_{j1}, \dots, \beta_{jp})' = \mathbf{W}_j \boldsymbol{\gamma} + \mathbf{u}_j \stackrel{ind.}{\sim} N_p(\mathbf{W}_j \boldsymbol{\gamma}, \boldsymbol{\Psi})$$

$$\mathbf{u}_j = (u_{1j}, \dots, u_{pj})' \stackrel{iid}{\sim} N_p(\mathbf{z}, \boldsymbol{\Psi}), \boldsymbol{\Psi} \equiv \boldsymbol{\Psi}(\boldsymbol{\theta})$$

$$\mathbf{X}_{ji} = (x_{ji1}, \dots, x_{jip})' \quad , \quad \mathbf{W}_j = (W_{j1}, \dots, W_{jq})'$$

$$i = 1, \dots, n_j \quad ; \quad j = 1, 2, \dots, J$$

# M.R. (geral/misto, equivalente ao MLHG anterior) para dados binários

$$Y_{ji} | \mathbf{u}_j \stackrel{ind.}{\sim} \text{Bernoulli}(p_{ji})$$
$$F(p_{ji}) = \mathbf{Z}'_{ji}\boldsymbol{\gamma} + \mathbf{X}'_{ji}\mathbf{u}_j \rightarrow p_{ji} = F^{-1}(\mathbf{Z}'_{ji}\boldsymbol{\gamma} + \mathbf{X}'_{ji}\mathbf{u}_j)$$
$$\mathbf{X}_{ji} = (x_{ji1}, \dots, x_{jip})' \quad ; \quad \mathbf{Z}_{ji} = (z_{ji1}, \dots, z_{jip})'$$
$$\mathbf{u}_j = (u_{1j}, \dots, u_{pj})'$$
$$\boldsymbol{\Psi} \equiv \boldsymbol{\Psi}(\boldsymbol{\theta})$$

Todas as outras quantidades são como definidas anteriormente.

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

$$Y_{ji}|u_{0j} \stackrel{ind.}{\sim} \text{Bernoulli}(p_{ji})$$

$$\begin{aligned} \text{logito}(p_{ji}) &= \ln\left(\frac{p_{ji}}{1-p_{ji}}\right) \\ &= \alpha_j + \beta_1(\text{idade}_{ji} - a) + \beta_2 \text{trat}_j + \beta_3 \text{sexo}_j + \beta_4 \text{base}_j \\ &\quad + \beta_5(\text{periodo}_{ji} - b) \\ &\rightarrow p_{ji} = \frac{e^{\alpha_j + \beta_1(\text{idade}_{ji} - a) + \beta_2 \text{trat}_j + \beta_3 \text{sexo}_j + \beta_4 \text{base}_j + \beta_5(\text{periodo}_{ji} - b)}}{1 + e^{\alpha_j + \beta_1(\text{idade}_{ji} - a) + \beta_2 \text{trat}_j + \beta_3 \text{sexo}_j + \beta_4 \text{base}_j + \beta_5(\text{periodo}_{ji} - b)}}, \\ \alpha_j &= \alpha + u_{0j} \end{aligned}$$

$$i = 1, 2, 3, 4; j = 1, 2, \dots, 59$$

em que  $u_{0j} \stackrel{i.i.d.}{\sim} N(0, \psi)$ ,  $a = 30, 32$  (média das idades) e  $b = 1$  (período 1). (OBS: foram usadas **variáveis binárias**, para representar as covariáveis).

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

- $Y_{ji}$  : condição : 1 - ruim; 0 - boa (no instante  $i$  do paciente  $j$ ).
- $idade_{ji}$ : idade (em anos), (no instante  $i$  do paciente  $j$ ),  $trat_j$  : (=1 droga ativa, = 0 placebo);  $sexo_j$  : (=0 feminino, =1 masculino);  $base_j$  : (=0 ausência do nível base, =1 presença do nível base), do  $j$ -ésimo paciente e  $periodo_{ji}$  : instante no qual a visita foi realizada, do  $j$ -ésimo paciente no  $i$ -ésimo instante.
- Exercício: propor uma forma geral (numérica) para calcular as esperanças marginais ( $\mathcal{E}(Y_{ji})$ ).

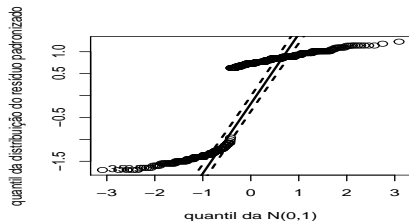
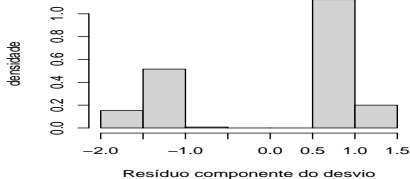
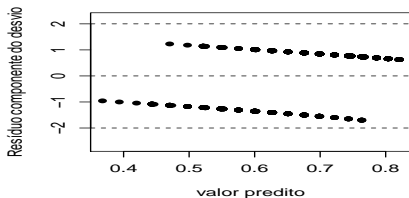
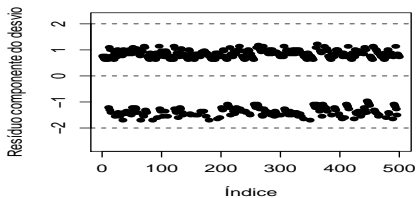
# Interpretação dos parâmetros

- 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 sexos (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:  $Y_{ji}|u_{0j} \stackrel{ind.}{\sim} \text{Bernoulli}(p_{ji}), i = 1, 2, \dots, 5; j = 1, 2, \dots, 100,$   
 $\text{logito}(p_{ji}) = 1 + 0,1x_i + u_{0j}; x_i = i; u_{0j} \stackrel{i.i.d.}{\sim} N(0; 0, 5).$
- Modelo 2:  $Y_{ji}|b_j \stackrel{ind.}{\sim} \text{Bernoulli}(p_{ji}), i = 1, 2, \dots, 5; j = 1, 2, \dots, 100,$   
 $p_{ji} = F(1 + 0,1x_i + u_{0j}, \nu = 4); x_i = i; u_{0j} \stackrel{i.i.d.}{\sim} N(0; 0, 5),$  em que  $F(\cdot, \nu = 4)$  representa a fda de uma distribuição  $t_{(4)}$ .

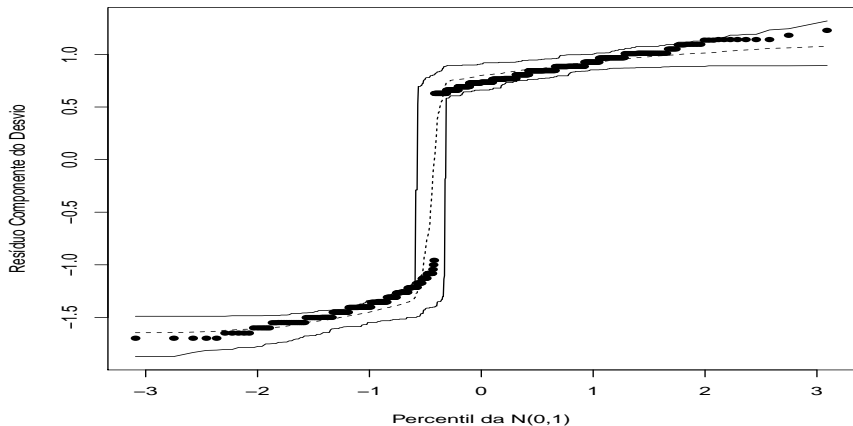
# Resultados: modelo 1 - RCD



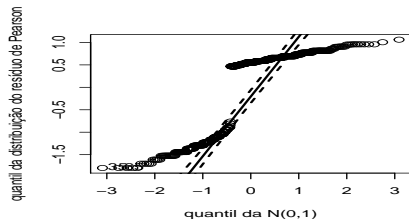
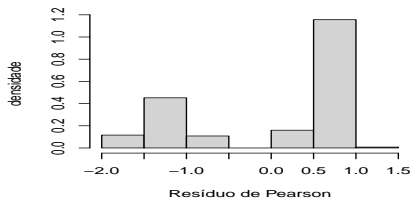
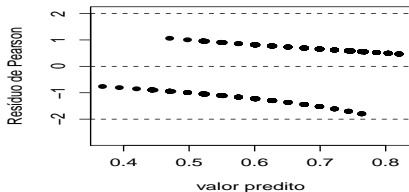
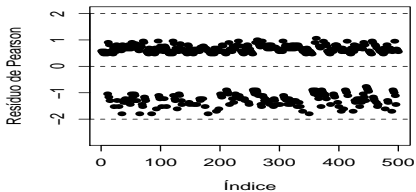


# Resultados: modelo 1 - RCD

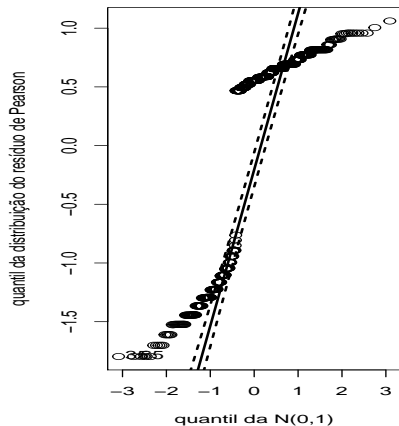
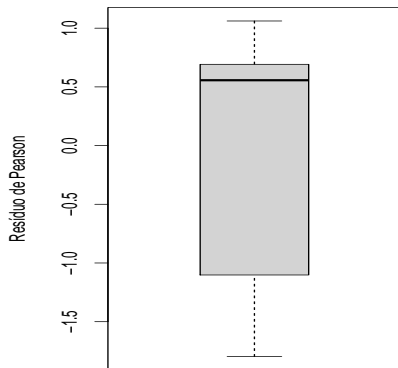
Gráfico de quantil-quantil normal



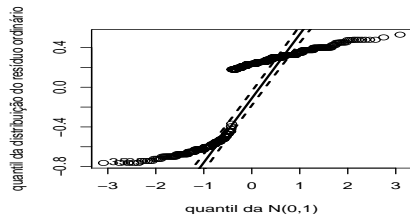
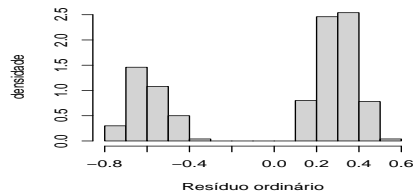
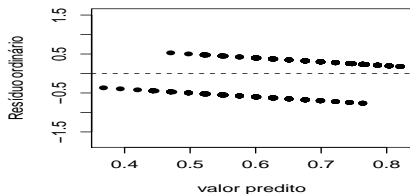
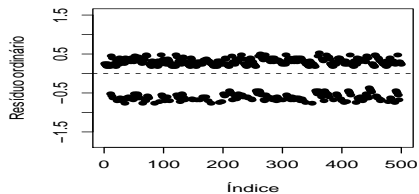
# Resultados: modelo 1 - Pearson



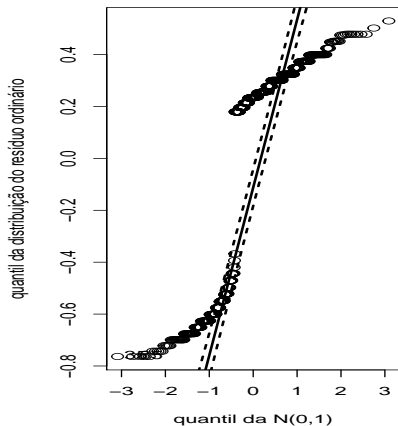
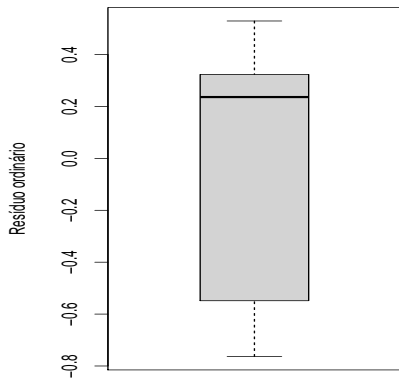
# Resultados: modelo 1 - Pearson



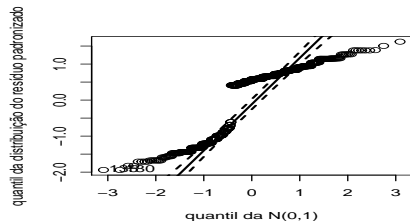
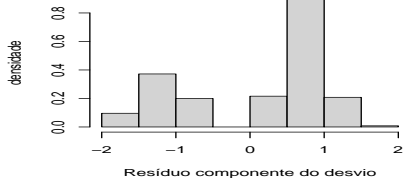
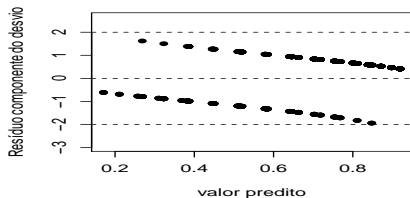
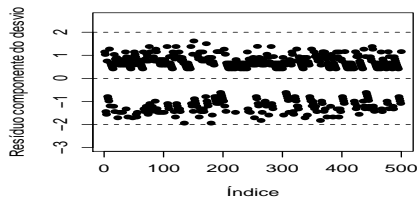
# Resultados: modelo 1 - Ordinário



## Resultados: modelo 1 - Ordinário

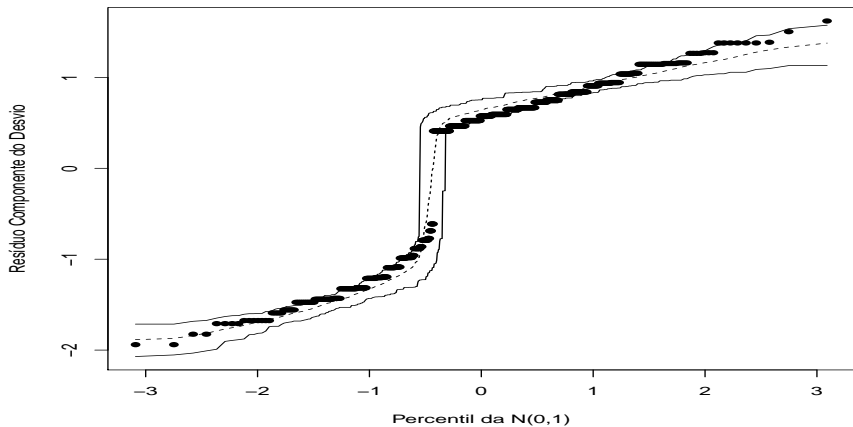


# Resultados: modelo 2 - RCD

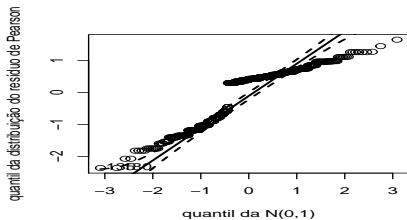
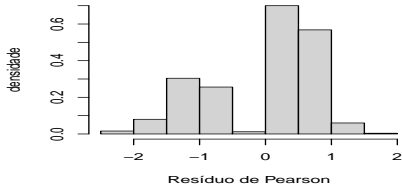
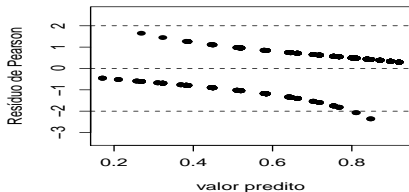
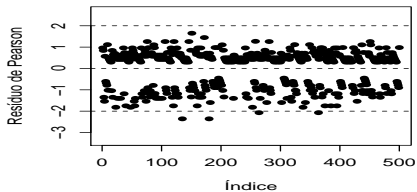


# Resultados: modelo 2 - RCD

Gráfico de quantil-quantil normal

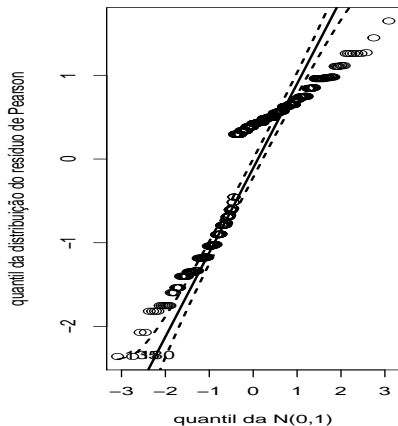
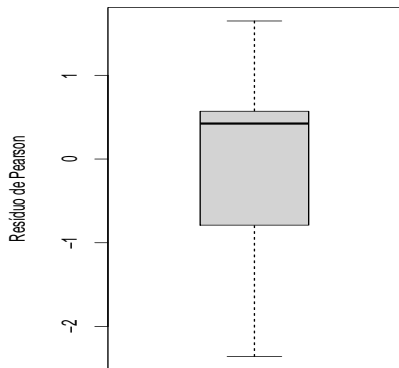


# Resultados: modelo 2 - Pearson

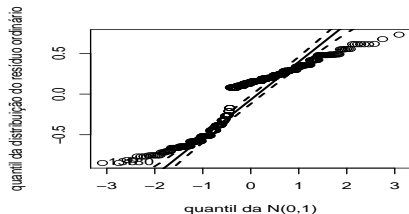
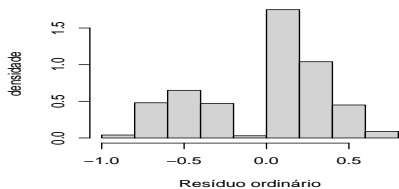
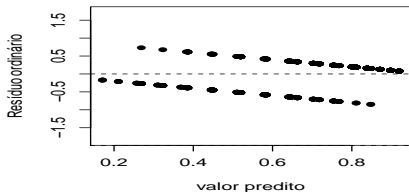
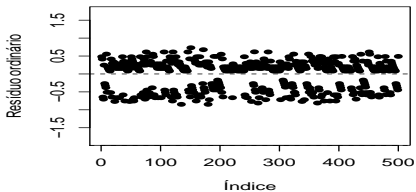




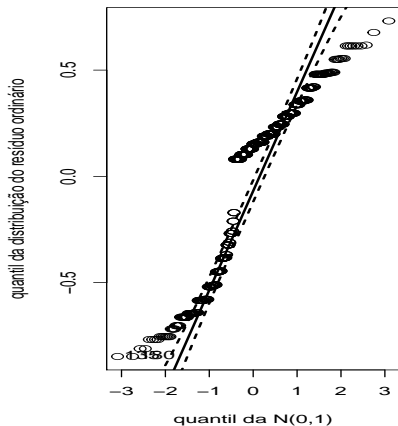
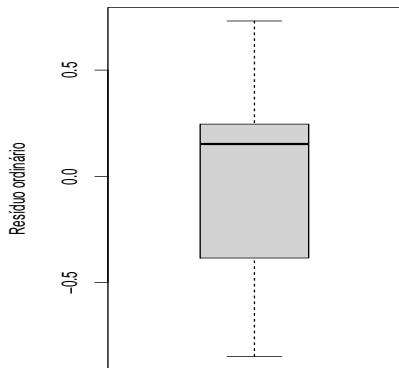
## Resultados: modelo 2 - Pearson



# Resultados: modelo 2 - Ordinário



## Resultados: modelo 2 - Ordinário



# Resultados do ajuste do modelo

## Laplace

Parâmetro	Estimativa	EP	IC(95%)	Estat. z	p-valor
$\beta_0$	4,56	1,13	[2,34 ; 6,78 ]	4,03	0,0001
$\beta_2$ (trat)	-1,62	0,55	[-2,69 ; -0,5]	-2,97	0,0030
$\beta_3$ (sexo)	-3,09	1,05	[1,03 ; -2,94]	-2,94	0,0033
$\beta_4$ (base)	0,79	0,88	[-0,94 ; 2,52]	0,90	0,3683
$\beta_1$ (idade)	0,07	0,02	[0,02 ; 0,11 ]	2,92	0,0034
$\beta_5$ (período)	-1,05	0,21	[-1,46 ; -0,65]	-5,08	< 0,0001

$$\tilde{\psi} = 1,538$$

# Resultados do ajuste do modelo

QA

Parâmetro	Estimativa	EP	IC(95%)	Estat. z	p-valor
$\beta_0$	4,61	1,16	[2.33 ; 6.90]	3,96	0,0001
$\beta_2$ (trat)	-1,65	0,57	[-2.76 ; -0.54]	-2,91	0,0037
$\beta_3$ (sexo)	-3,13	1,09	[-5.26 ; -1.00 ]	-2,88	0,0040
$\beta_4$ (base)	0,81	0,92	[-0.99 ; 2.62]	0,89	0,3757
$\beta_1$ (idade)	0,07	0,02	[0.02 ; 0.12]	2,85	0,0043
$\beta_5$ (período)	-1,07	0,21	[-1.48 ; -0.66]	-5,10	< 0,0001

$$\tilde{\psi} = 1,777$$

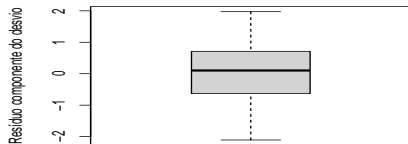
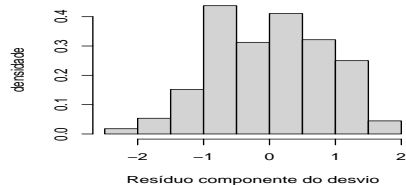
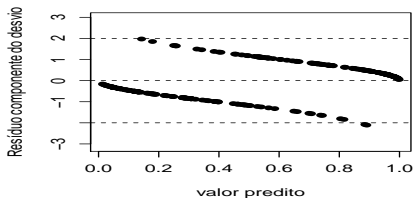
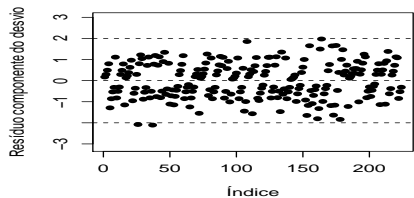
# Resultados do ajuste do modelo

QVP

Parâmetro	Estimativa	EP	IC(95%)	Estat. z	p-valor
$\beta_0$	4,29	0,99	[2,36 ; 6,23]	4,35	< 0,0001
$\beta_2$ (trat)	-1,55	0,52	[-2,56 ; -0,54]	-3,00	0,0041
$\beta_3$ (sexo)	-2,86	0,96	[-4,74 ; -0,97]	-2,97	0,0045
$\beta_4$ (base)	0,78	0,87	[-0,93 ; 2,48]	0,89	0,3771
$\beta_1$ (idade)	0,06	0,02	[0,02 ; 0,11 ]	2,86	0,0061
$\beta_5$ (período)	-1,04	0,15	[-1,33 ; -0,74]	-6,98	< 0,0001

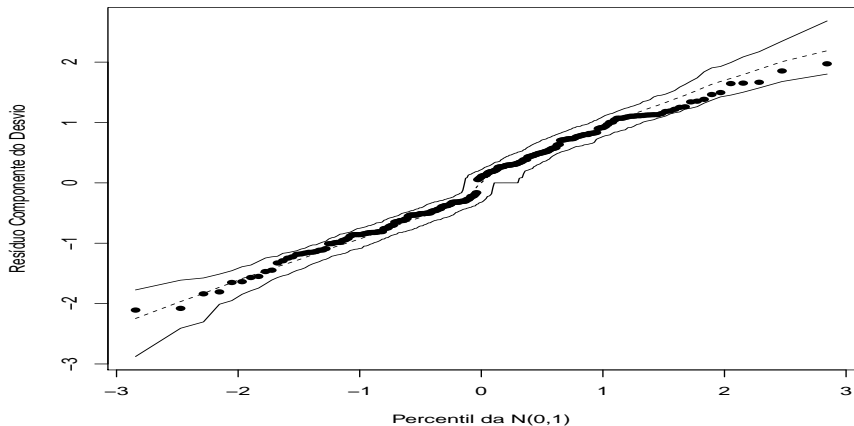
$$\tilde{\psi} = 2,059$$

# Resultados: modelo ajustado com QA - RCD



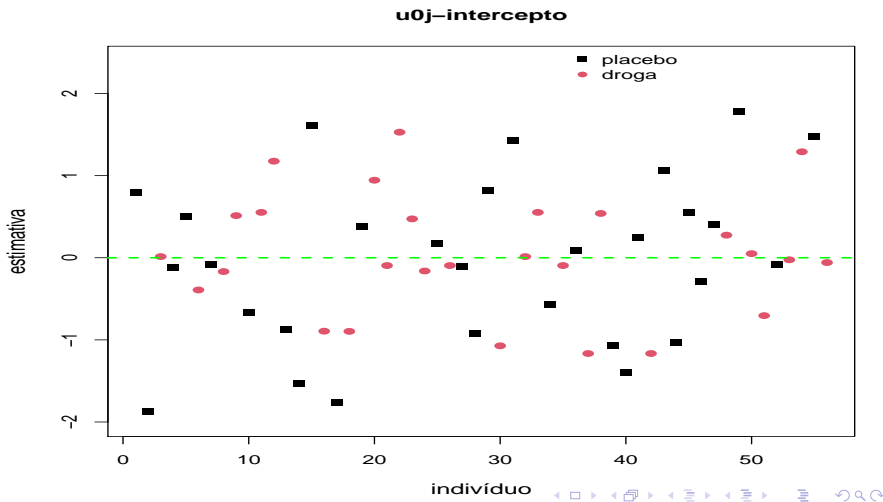
# Resultados: modelo ajustado com QA - RCD

Gráfico de quantil-quantil normal

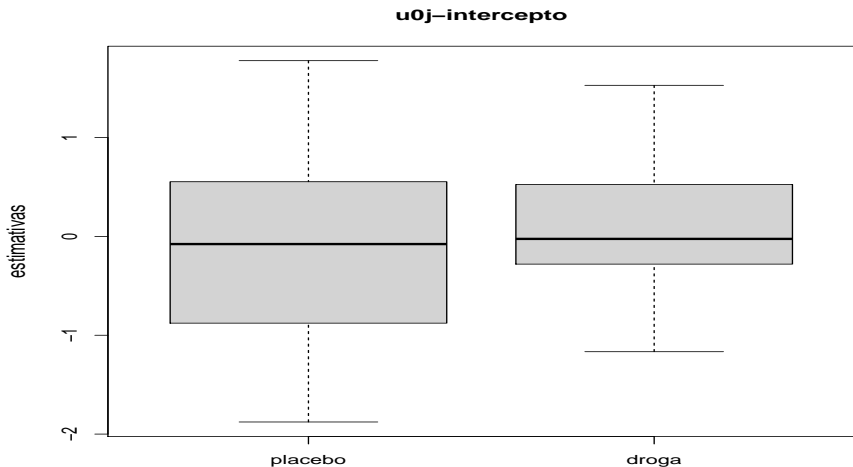




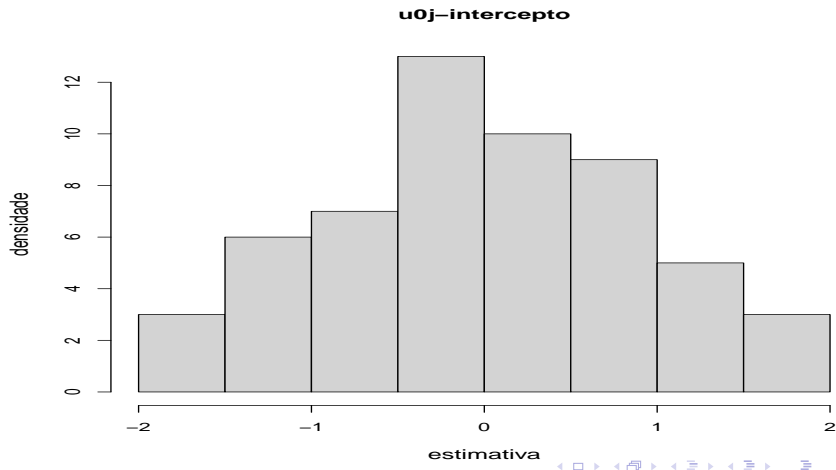
# Dados reais: efeitos aleatórios por grupo



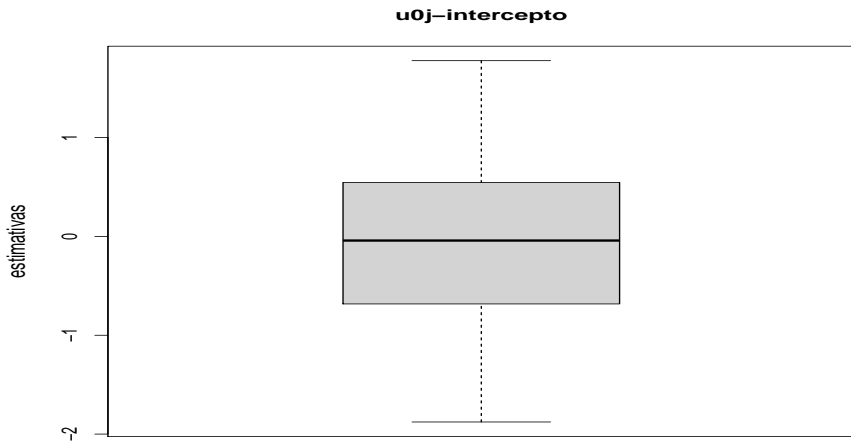
# Dados reais: box plot dos efeitos aleatórios por grupo



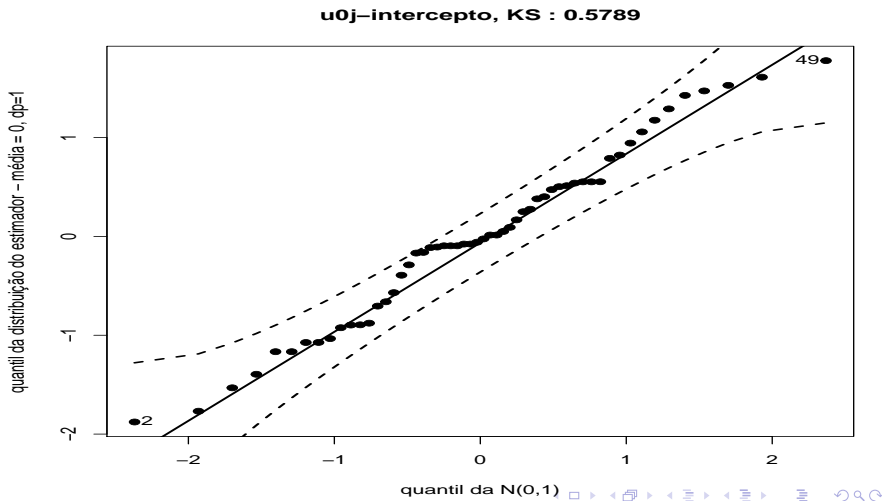
# Dados reais: histograma dos efeitos aleatórios



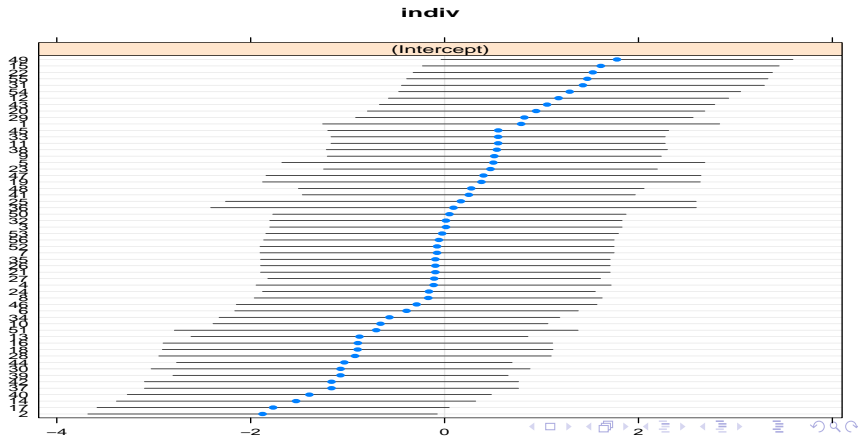
# Dados reais: boxplot dos efeitos aleatórios



# Dados reais: QQplot com envelopes dos efeitos aleatórios



# Dados reais: valores preditos e intervalos de confiança dos efeitos aleatórios



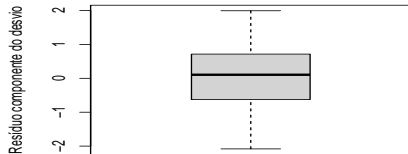
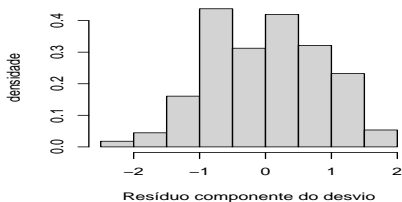
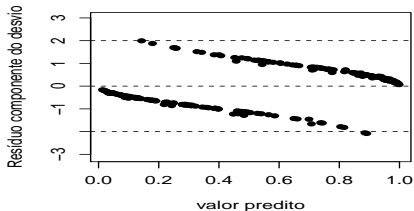
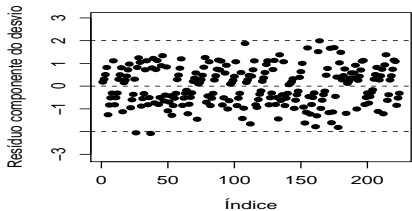
# Resultados do ajuste do modelo reduzido - MR (sem nível base)

QA

Parâmetro	Estimativa	EP	IC(95%)	Estat. z	p-valor
$\beta_0$	4,97	1,12	[2,78 ; 7,17]	4,44	< 0,0001
$\beta_2$ (trat)	-1,54	0,55	[-2,61 ; -0,46]	-2,79	0,0052
$\beta_3$ (sexo)	-2,79	0,99	[-4,73 ; -0,85]	-2,82	0,0047
$\beta_1$ (idade)	0,07	0,02	[0,02 ; 0,12]	2,85	0,0044
$\beta_5$ (período)	-1,07	0,21	[-1,48 ; -0,66]	-5,10	< 0,0001

$$\tilde{\psi} = 1,813$$

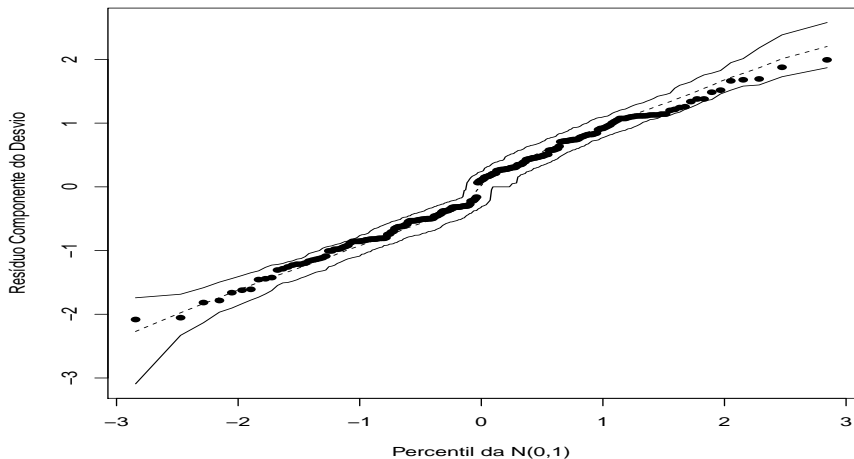
# MR ajustado com QA - RCD (diagnóstico)



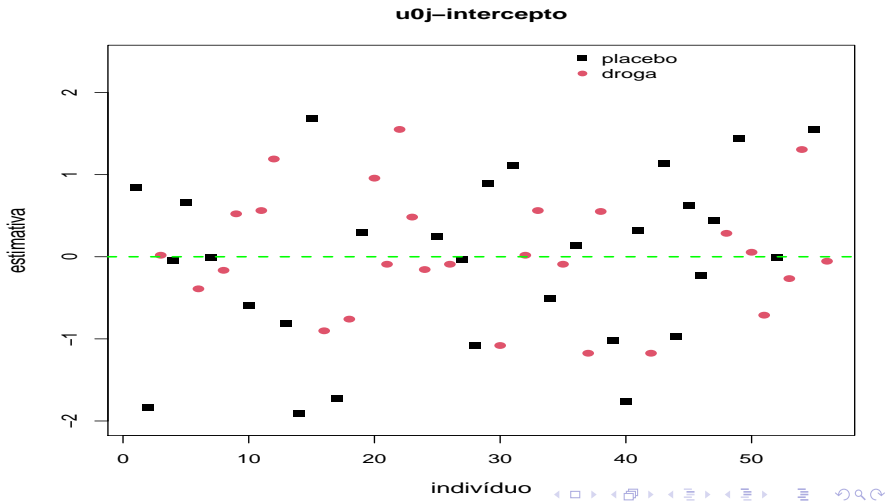


# MR ajustado com QA - RCD (env. sim. do modelo)

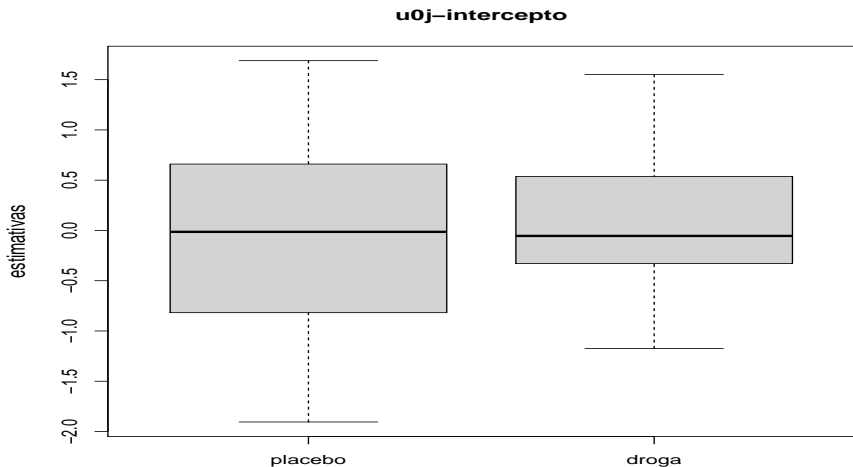
Gráfico de quantil-quantil normal



# Dados reais: efeitos aleatórios por grupo (MR)



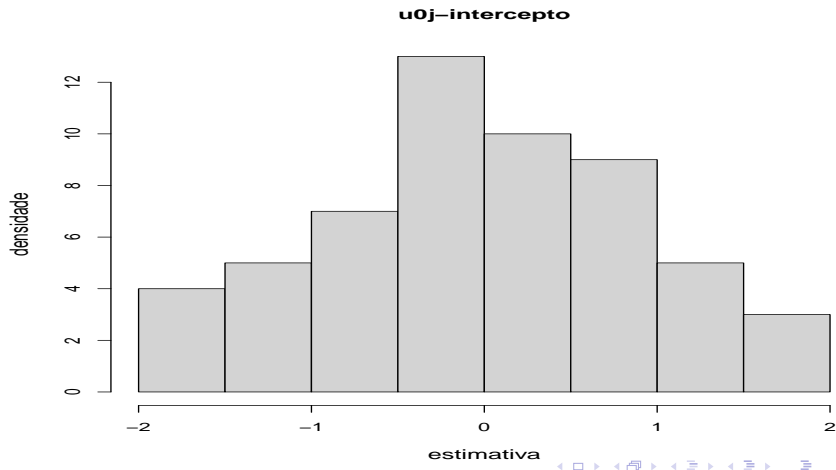
# Dados reais: boxplot dos efeitos aleatórios por grupo (MR)



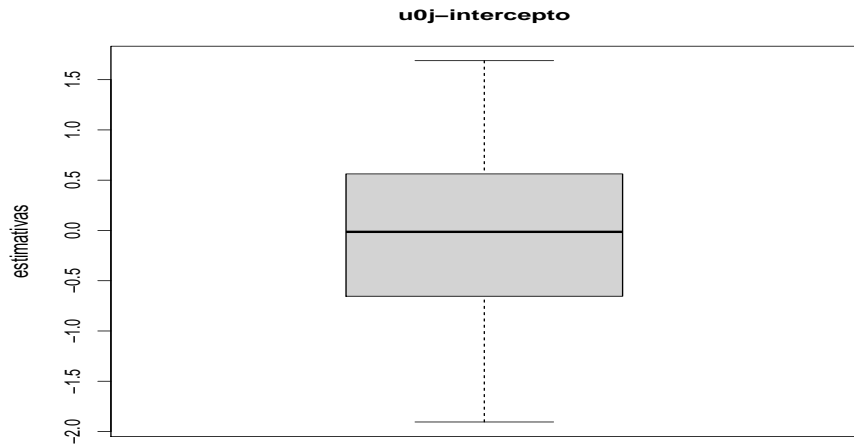
grupo



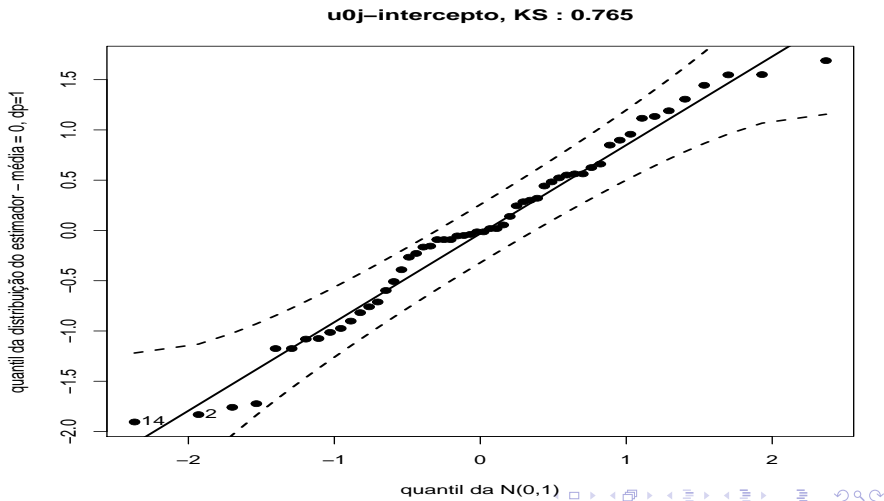
# Dados reais: histograma aleatórios por grupo (MR)



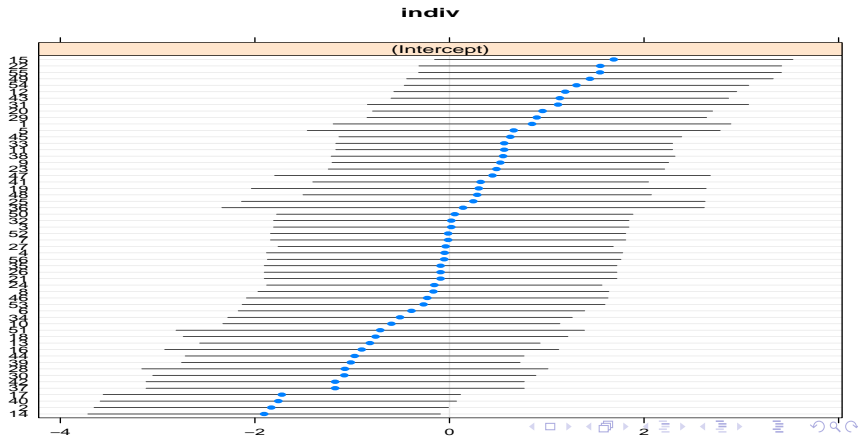
# Dados reais: boxplot dos efeitos aleatórios (MR)



# Dados reais: QQplot com env. dos efeitos aleatórios (MR)



# Dados reais: valores preditos e intervalos de confiança dos efeitos aleatórios (MR)



# Comentários

- Todos os fatores, com exceção do nível base são significativos.
- Os modelos apresentaram, me geral, um bom ajuste.
- O grupo que usou a droga apresentou melhora (significativa) na condição respiratória.
- Homens e/ou jovens apresentam, em geral, melhor condição respiratória.
- A suposição de normalidade parece ser razoável para modelar os efeitos aleatórios, os quais parecem ser significativos.