

Modelos lineares generalizados hierárquicos/multiníveis

Prof. Caio Azevedo

Família exponencial bi-paramétrica

- Dizemos que uma v.a. Y (discreta ou contínua) pertence à família exponencial biparamétrica se sua fdp é dada por:

$$f(y; \theta) = \exp \{ \phi [y\theta - b(\theta)] + c(y, \phi) \} \mathbb{1}_A(y)$$

em que $\theta \in \Theta \subset \mathfrak{R}$ (espaço paramétrico de θ), $b(\theta) : \mathfrak{R} \rightarrow \mathfrak{R}$,
 $c(y, \phi) : A \times \mathfrak{R}^+ \rightarrow \mathfrak{R}$, A é um conjunto que não depende nem de θ
nem de ϕ e, por sua vez, $\phi (\phi > 0)$ é o parâmetro de precisão.

- Veja mais detalhes [aqui](#).

Família exponencial bi-paramétrica

- Vamos começar com modelos com apenas um nível.
- Consideraremos que $Y \sim FE(\theta, \phi)$ e que temos $Y_i \stackrel{ind.}{\sim} FE(\theta_i, \phi), i = 1, 2, \dots, n$, ou seja

$$f(y_i; \theta) = \exp \{ \phi [y_i \theta_i - b(\theta_i)] + c(y_i, \phi) \} \mathbb{1}_A(y_i)$$

Modelo linear generalizado

- $Y_i \sim \text{FE}(\theta_i, \phi)$, $i = 1, 2, \dots, n$, $\theta_i = h(\mu_i)$.
- $g(\mu_i) = \eta_i$, $\eta_i = \sum_{j=1}^p x_{ij}\beta_j$, $\mathcal{E}(Y_i) = \mu_i = b'(\theta_i)$, x_{ji} : covariável j associada ao indivíduo i (fixa e conhecida) e $\beta = (\beta_1, \dots, \beta_p)'$, ϕ : parâmetros desconhecidos.
- $\mathcal{V}(Y_i) = \phi^{-1}V(\mu_i)$, em que $V(\mu_i) = \frac{d\mu_i}{d\theta_i}$.
- $g(\cdot)$ é uma função de ligação (invertível e duplamente diferenciável).
Quando $\theta = g(\cdot)$ temos a função de ligação canônica.
- Veja mais informações: [aqui](#).
- Algumas distribuições seguem.

Exemplo: Binomial

Seja Y^* a proporção de sucessos em n ensaios de Bernoulli independentes. Logo, temos que $nY^* \sim \text{binomial}(n, \mu)$. Nesse caso $\mathcal{E}(Y^* = \mu), \mu \in (0, 1)$. Além disso, em termos da distribuição de Y^* (exercício), temos que $A = \{0, 1/n, 2/n, \dots, 1\}$ e

$$\begin{aligned} f_{Y^*}(y^*) &= \binom{n}{ny^*} \mu^{ny^*} (1 - \mu)^{n - ny^*} = \exp \left\{ \ln \binom{n}{ny^*} + ny^* \ln \left(\frac{\mu}{1 - \mu} \right) \right. \\ &\quad \left. + n \ln(1 - \mu) \right\} = \exp \left\{ n \left[y^* \ln \left(\frac{\mu}{1 - \mu} \right) + \ln(1 - \mu) \right] + \ln \binom{n}{ny^*} \right\} \\ &= \exp \{ \phi [y\theta - b(\theta)] + c(y, \phi) \} \end{aligned}$$

Exemplo: Poisson

$Y \sim \text{Poisson}(\lambda)$, $\lambda > 0$, então $A = \{0, 1, 2, \dots\}$ e

$$\begin{aligned}f_Y(y) &= \frac{e^{-\lambda} \lambda^y}{y!} = \exp \{y \ln(\lambda) - \lambda - \ln(y!)\} \\ &= \exp \{ \phi [y\theta - b(\theta)] + c(y, \phi) \}\end{aligned}$$

em que $\theta = \ln(\lambda)$, $b(\theta) = \exp(\theta)$, $\phi = 1$, $c(y, \phi) = -\ln(y!)$

Cont.: Binomial

em que $\theta = \ln\left(\frac{\mu}{1-\mu}\right)$, $b(\theta) = \ln(1 + e^\theta)$, $\phi = n$, $c(y^*, \phi) = \ln\binom{\phi}{y^*}$.

- Se $n = 1$, $Y^* = Y \sim \text{Bernoulli}(\mu)$.

Exemplo: Normal

Seja $Y \sim N(\mu, \sigma^2)$, temos que $A = (-\infty, \infty)$,
 $\mu \in (-\infty, \infty)$, $\sigma^2 \in (0, \infty)$ e

$$\begin{aligned}f_Y(y) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\} \\&= \exp\left\{\frac{1}{\sigma^2}\left(\mu y - \frac{\mu^2}{2}\right) - \frac{1}{2}\left[\ln(2\pi\sigma^2) + \frac{y^2}{\sigma^2}\right]\right\} \\&= \exp\{\phi[y\theta - b(\theta)] + c(y, \phi)\}\end{aligned}$$

em que $\theta = \mu$, $b(\theta) = \frac{\theta^2}{2}$, $\phi = \sigma^{-2}$, $c(y, \phi) = \frac{1}{2}\ln(\phi/2\pi) - \frac{\phi y^2}{2}$.

Exemplo: gama

Usualmente consideramos $Y \sim \text{gama}(a, b)$, $a, b > 0$, em que $\mathcal{E}(Y) = ab$, $\mathcal{V}(Y) = ab^2$ e $A = (-\infty, \infty)$. Nesse caso $f(Y) = \frac{1}{\Gamma(a)b^a} y^{a-1} e^{-y/b}$.

Contudo, consideraremos uma outra parametrização que consiste em escrever a fdp de Y em termos de $\mu = \mathcal{E}(Y)$ e do parâmetro de precisão ϕ (exercício), de modo que $CV(Y) = DP(Y)/\mathcal{E}(Y) = \phi^{-1/2}$, o que implica que $\mathcal{V}(Y) = V(\mu)/\phi$.

Exemplo: Gamma

Assim:

$$\begin{aligned} f_Y(y) &= \frac{1}{\Gamma(\phi)} \left(\frac{\phi y}{\mu}\right)^\phi \exp\left\{-\frac{\phi y}{\mu}\right\} y^{-1} = \exp\left\{\phi \left[-\frac{y}{\mu} - \ln(\mu)\right]\right. \\ &\quad \left. - \ln \Gamma(\phi) + \phi \ln(\phi y) - \ln(y)\right\} = \exp\left\{\phi \left[-\frac{y}{\mu} - \ln(\mu)\right]\right. \\ &\quad \left. + (\phi - 1) \ln(y) + \phi \ln(\phi) - \ln \Gamma(\phi)\right\} = \exp\{\phi [y\theta - b(\theta)] + c(y, \phi)\} \end{aligned}$$

em que

$$\theta = -1/\mu, \quad b(\theta) = -\ln(-\theta), \quad c(y, \phi) = (\phi - 1) \ln(y) + \phi \ln(\phi) - \ln \Gamma(\phi).$$

Se $\phi = 1$, então $Y \sim \exp(\mu)$. Se $\phi = k/2$ e $\mu = k$, então $Y \sim \chi^2_{(k)}$

Exemplo: normal inversa

Seja $Y \sim NI(\mu, \phi)$, $\mu, \phi > 0$, então $A = (0, \infty)$ e

$$f_Y(y) = \frac{\phi^{1/2}}{\sqrt{2\pi y^3}} \exp\left\{-\frac{\phi(y - \mu)^2}{2\mu^2 y}\right\}$$

Exercício: colocar na forma da família exponencial.

Principais distribuições pertencentes à FE

Distribuição	$b(\theta)$	θ	ϕ	$V(\mu)$
Normal	$\theta^2/2$	μ	σ^{-2}	1
Poisson	e^θ	$\ln \mu$	1	μ
Binomial	$\ln(1 + e^\theta)$	$\ln(\mu/(1 - \mu))$	m	$\mu(1 - \mu)$
Gama	$-\ln(-\theta)$	$-1/\mu$	$1/(CV^2)$	μ^2
N.Inversa	$-\sqrt{-2\theta}$	$-1/2\mu^2$	ϕ	μ^3

Mais detalhes sobre as parametrizações dessas distribuições, no contexto de MLG, podem ser encontradas [aqui](#), as quais são parametrizadas em termos da média (μ) e parâmetro de precisão (ϕ).

Desvio (ou função desvio)

- Defina $\mathbf{y} = (y_1, \dots, y_n)'$ e $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$.
- Sem perda de generalidade, seja $l(\boldsymbol{\mu}, \mathbf{y}) \equiv l(\boldsymbol{\beta}, \phi)$ a logverossimilhança associada ao modelo em estudo e $l(\boldsymbol{\mu}^0, \mathbf{y})$ a logverossimilhança do modelo saturado ($n=p$), ou seja, em que cada média é representada por ela mesma.
- $l(\boldsymbol{\mu}, \mathbf{y}) = \phi \left[\sum_{i=1}^n y_i \theta_i - \sum_{i=1}^n b(\theta_i) \right] + \sum_{i=1}^n c(y_i, \phi)$
- $l(\boldsymbol{\mu}^{(0)}, \mathbf{y}) = \phi \left[\sum_{i=1}^n y_i \theta_i^{(0)} - \sum_{i=1}^n b(\theta_i^{(0)}) \right] + \sum_{i=1}^n c(y_i, \phi)$
em que $\theta_i^{(0)} = h(\mu_i^{(0)})$.

Desvio (ou função desvio)

- Assim, $D^*(\mathbf{y}, \boldsymbol{\mu}) = 2(l(\boldsymbol{\mu}^{(0)}, \mathbf{y}) - l(\boldsymbol{\mu}, \mathbf{y})) = 2\phi \sum_{i=1}^n \left[y_i \left(\theta_i^{(0)} - \theta_i \right) + b(\theta_i) - b(\theta_i^{(0)}) \right] = \phi D(\mathbf{y}, \boldsymbol{\mu})$
- Sejam $\hat{\boldsymbol{\mu}}^{(0)} = \mathbf{Y}$ e $\hat{\boldsymbol{\mu}} = \mathbf{g}^{-1}(\hat{\boldsymbol{\eta}}) = \mathbf{g}^{-1}(\mathbf{X}\hat{\boldsymbol{\beta}})$ os respectivos estimadores de MV e defina $\hat{\theta}_i^{(0)} = h(\hat{\mu}_i^{(0)})$ e $\hat{\theta}_i = h(\hat{\mu}_i)$.
- Assim, o desvio não escalonado é dado por $D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \left[y_i \left(\hat{\theta}_i^{(0)} - \hat{\theta}_i \right) + b(\hat{\theta}_i) - b(\hat{\theta}_i^{(0)}) \right]$.
- O desvio escalonado é dado por $D^*(\mathbf{y}, \hat{\boldsymbol{\mu}}) = \phi D(\mathbf{y}, \hat{\boldsymbol{\mu}})$. Na prática, substituímos ϕ por algum **estimador consistente**.

Desvio: exemplos

- Bernoulli: $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = -2 \sum_{i=1}^n \{ \ln(1 - \hat{\mu}_i) I_{\{0\}}(y_i) + \ln(\hat{\mu}_i) I_{\{1\}}(y_i) \}$
- Poisson: $D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^k \{ [y_i \ln(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)] I_{\{1,2,\dots\}}(y_i) + \hat{\mu}_i \mathbb{1}_{\{0\}}(y_i) \}$
- Normal: $D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2$ e $D^*(\mathbf{y}, \hat{\boldsymbol{\mu}}) = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \hat{\mu}_i)^2$.
- Gama: $D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \{ -\ln(y_i/\hat{\mu}_i) + (y_i - \hat{\mu}_i)/\hat{\mu}_i \}$ e $D^*(\mathbf{y}, \hat{\boldsymbol{\mu}}) = \phi D(\mathbf{y}, \hat{\boldsymbol{\mu}})$.
- Mais detalhes, veja [aqui](#).

Desvio: comportamento assintótico

- Em geral $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ ou $D^*(\mathbf{y}, \hat{\boldsymbol{\mu}})$ não seguem (mesmo assintoticamente) uma distribuição $\chi^2_{(n-p)}$, sob a hipótese de que o modelo em questão é adequado. Tal convergência ($D^*(\mathbf{y}, \hat{\boldsymbol{\mu}}) \xrightarrow[D]{\phi \rightarrow \infty} \chi^2_{n-p}$) ocorre (sob a hipótese de que o modelo em questão é adequado):
 - Binomial: se $m_i \rightarrow \infty, \forall i$ e n (tamanho da amostra) é fixo. Assim, em geral, para o modelo Bernoulli ($m_i = 1, \forall i$), tal resultado não é válido.
 - Poisson: se $\mu_i \rightarrow \infty, \forall i$.
 - Nos casos em que $D^*(\mathbf{Y}, \hat{\boldsymbol{\mu}})$ depende do parâmetro de precisão, se $\phi \rightarrow \infty$.

Análise do desvio

- Podemos definir um procedimento para testar as hipóteses

$$H_0 : \beta_1 = 0 \text{ vs } H_1 : \beta_1 \neq 0.$$

- A estatística

$$Q_{AD} = \frac{(D(\mathbf{y}; \hat{\boldsymbol{\mu}}^{(0)}) - D(\mathbf{y}; \hat{\boldsymbol{\mu}})) / q}{D(\mathbf{y}; \hat{\boldsymbol{\mu}}) / (n - p)},$$

sob H_0 e para n suficientemente grande, é tal que $Q_{AD} \approx F_{(q, n-p)}$ e q é dimensão do vetor β_1

Análise do desvio

- Note que só podemos utilizar esta abordagem para modelos não saturados ($n > p$).
- Assim, rejeita-se H_0 se p -valor $\leq \alpha$, em que p -valor $\approx P(X \geq q_{AD} | H_0)$, $X \sim F_{(q, n-p)}$ e

$$q_{AD} = \frac{(D(\mathbf{y}; \tilde{\boldsymbol{\mu}}^{(0)}) - D(\mathbf{y}; \tilde{\boldsymbol{\mu}})) / q}{D(\mathbf{y}; \tilde{\boldsymbol{\mu}}) / (n - p)}.$$

Outras funções de ligação

- Seja μ a proporção de sucessos de uma binomial.
- A ligação probito é dada por

$$\Phi^{-1}(\mu) = \eta$$

ou, de modo equivalente, $\mu = \Phi(\eta)$, em que $\Phi(\cdot)$ é a fda de uma distribuição normal padrão.

Outras funções de ligação

- Novamente, seja μ a proporção de sucessos de uma binomial.
- A fda de uma distribuição do valor extremo padrão (ou Gumbell padrão, a qual corresponde ao logaritmo natural de uma distribuição exponencial com seu parâmetro igual a 1) é dada por:

$$F(x) = 1 - \exp\{-\exp(x)\}$$

- Assim, o modelo binomial com ligação log-log é dado por

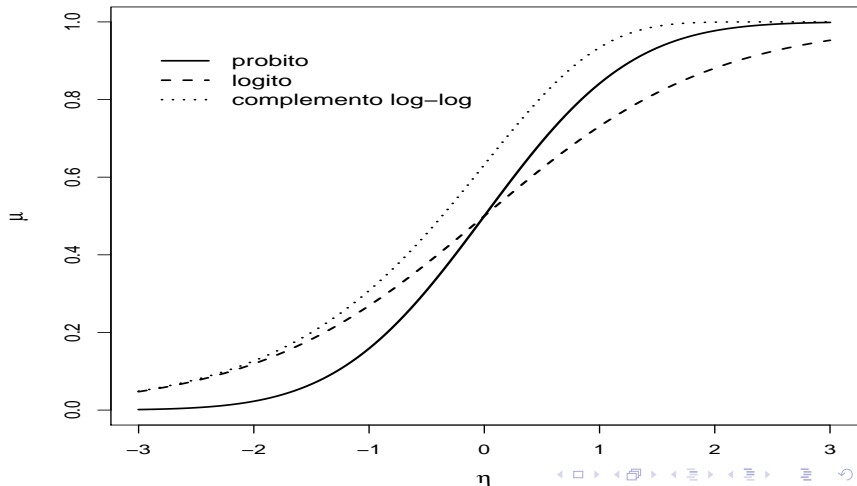
$$\mu = 1 - \exp\{-\exp(\eta)\}$$

ou de modo equivalente, $\ln(-\ln(1 - \mu)) = \eta$.

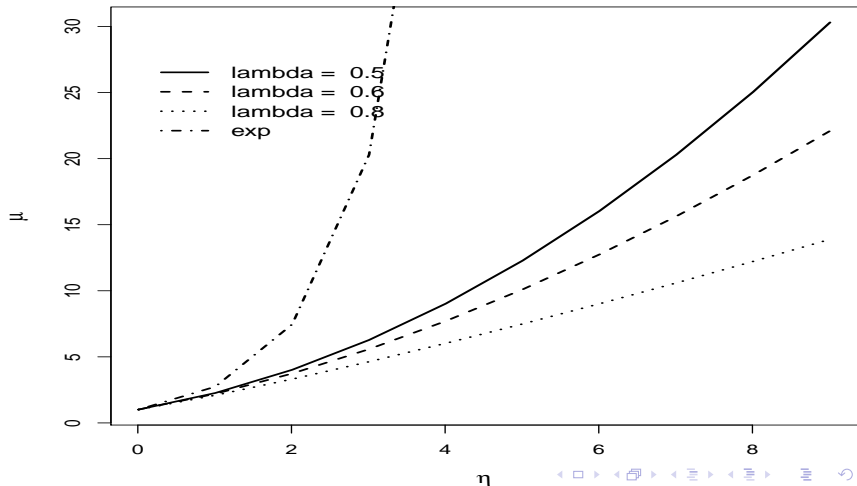
Resumo: principais funções de ligação

Distribuição	$g(\mu)$
Normal	$\mu, (\text{canônica}), 1/\mu$ e $\ln \mu$ (se $\mu > 0$)
Poisson	$\ln \mu$ (canônica), $\sqrt{\mu}$
Binomial	$\ln(\mu/(1 - \mu))$ (canônica), $\Phi^{-1}(\mu)$, $\ln(-\ln(1 - \mu))$
Gama	$1/\mu$ (canônica), $\ln \mu$
N.Inversa	$1/(2\mu^2)$ (canônica), $\ln \mu$, $1/\mu$

Funções de ligação para médias no intervalo (0,1)



Funções de ligação da família Box-Cox



Estimação

- Lembremos que $\theta_i = h(\mu_i)$, $\mu_i = g^{-1}(\eta_i)$ em que $\eta_i = \sum_{j=1}^p x_{ji}\beta_j$.
- Logo $\theta_i = h(g^{-1}(\eta_i))$. Se $g(\cdot)$ for uma função de ligação canônica, então $\theta_i = \eta_i$.

- Verossimilhança

$$L(\boldsymbol{\beta}, \phi) = \exp \left\{ \phi \left[\sum_{i=1}^n y_i \theta_i - \sum_{i=1}^n b(\theta_i) \right] + \sum_{i=1}^n c(y_i, \phi) \right\}$$

- Log-verossimilhança

$$l(\boldsymbol{\beta}, \phi) = \phi \left[\sum_{i=1}^n y_i \theta_i - \sum_{i=1}^n b(\theta_i) \right] + \sum_{i=1}^n c(y_i, \phi)$$

Estimação

- Vetor Escore para β (usando a regra da cadeia)

$$\begin{aligned} S(\beta) &= \sum_{i=1}^n \phi \left\{ y_i \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{d\eta_i}{d\beta} - \frac{db(\theta_i)}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{d\eta_i}{d\beta} \right\} \\ &= \phi \sum_{i=1}^n \left\{ \sqrt{\frac{\omega_i}{V_i}} (y_i - \mu_i) \mathbf{X}_i \right\} \\ &= \phi \mathbf{X}' \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu}) \end{aligned} \quad (1)$$

em que $\mathbf{W} = \text{diag}(\omega_1, \dots, \omega_n)$, $\omega_i = (d\mu_i/d\eta_i)^2/V_i$, $V_i = V(\mu_i)$ e $\mathbf{V} = \text{diag}(V_1, \dots, V_n)'$. As outras quantidades são como definidas anteriormente.

Estimação

- Vetor Escore para ϕ

$$S(\phi) = \sum_{i=1}^n \{y_i \theta_i - b(\theta_i)\} + \sum_{i=1}^n c'(y_i, \phi) \quad (2)$$

em que $c'(y_i, \phi) = \frac{dc(y_i, \phi)}{d\phi}$

Estimação

- Informação de Fisher (usando a regra da cadeia)

$$I(\beta, \beta) = -\mathcal{E} \left\{ \phi \left\{ \sum_{i=1}^n \frac{1}{V_i^{1/2}} (Y_i - \mu_i) \mathbf{X}_i \frac{d\omega_i^{1/2}}{d\beta'} + \omega_i^{1/2} (Y_i - \mu_i) \mathbf{X}_i \frac{dV_i^{-1/2}}{d\beta'} - \sqrt{\frac{\omega_i}{V_i}} \mathbf{X}_i \frac{d\mu_i}{d\beta'} \right\} \right\} = \phi \mathbf{X}' \mathbf{W} \mathbf{X}$$

$$I(\beta, \phi) = -\mathcal{E} \left(\left\{ \sqrt{\frac{\omega_i}{V_i}} (Y_i - \mu_i) x_{ji} \right\} \right) = 0$$

Estimação

- Informação de Fisher para ϕ

$$I(\phi, \phi) = -\mathcal{E} \left(\sum_{i=1}^n c''(Y_i, \phi) \right)$$

em que $c''(y_i, \phi) = \frac{d^2 c(y_i, \phi)}{d\phi^2}$.

Inferência para o modelo

- O sistema de equações $\mathbf{S}(\hat{\boldsymbol{\beta}}) = 0, S(\hat{\phi}) = 0$ não tem solução explícita e algum método de otimização numérica, como o algoritmo escore de Fisher, deve ser utilizado para obter-se as estimativas de MV.

- Contudo, como os parâmetros são ortogonais

$$l(\boldsymbol{\beta}, \phi) = -\mathcal{E} \left(\left\{ \sqrt{\frac{\omega_i}{V_i}} (Y_i - \mu_i) x_{ji} \right\} \right) = 0 \text{ e o produto matricial}$$

$\mathbf{I}^{-1}(\boldsymbol{\beta})\mathbf{S}(\boldsymbol{\beta})$ não depende de ϕ , a estimação é feita em duas etapas.

Estimação de β : Algoritmo escore de Fisher

- Seja $\beta^{(0)}$ uma estimativa inicial de β (chute inicial), então faça

$$\beta^{(t+1)} = \beta^{(t)} + \mathbf{I}^{-1}(\beta^{(t)})\mathbf{S}(\beta^{(t)}), t = 1, 2, \dots \quad (3)$$

até que algum critério de convergência seja satisfeito, como

$$\|\beta^{(t+1)} - \beta^{(t)}\| < \epsilon, \epsilon > 0,$$

Estimação de ϕ

- Com as estimativas de β , digamos $\tilde{\beta}$, obtidas no passo anterior, obtenha as estimativas de ϕ através de

$\hat{\phi} = \frac{n-p}{\sum_{i=1}^n \left(\frac{(Y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \right)}$ que é o estimador do método dos momentos (e consistente) de ϕ .

- O R fornece a estimativa associada ao estimador $\hat{\phi} = \frac{n-p}{D(\mathbf{y}, \hat{\boldsymbol{\mu}})}$ que não é consistente para ϕ .
- Para os modelos Poisson e Bernoulli, $\phi = 1$.
- Para mais detalhes sobre estimação em MLG veja [aqui](#) e [aqui](#).
- No R tais modelos podem ser ajustados via função [glm](#).

Análise de diagnóstico e Teste de hipótese

- O procedimento mais usual para a verificação da qualidade do MLG's baseia-se no chamado resíduo componente do desvio (RCD).
- Sob a validade das hipóteses do modelo temos que o RCD (padronizado) segue, aproximadamente, uma distribuição $N(0,1)$.
- Detalhes podem ser encontrado em Paula (2013) (link no site do curso) e [aqui \(ME720\)](#).
- Podem ser desenvolvidos testes de hipótese do tipo $C\beta = M$ ([veja aqui](#)) de modo semelhante ao que foi [visto anteriormente](#).

Exemplo 4: Ataques epiléticos (Diggle, Liang e Zeger, 1994, Seção 8.4)

- Diz respeito aos resultados de um ensaio clínico com 59 indivíduos epiléticos os quais foram aleatorizados de modo que cada um recebesse uma droga antiepilética denominada progabide ou placebo.
- Os dados de cada indivíduo consistiram de um número inicial de ataques epiléticos num período de oito semanas antes do tratamento, seguido do número de ataques em cada período de duas semanas, num total de quatro períodos, após o tratamento.

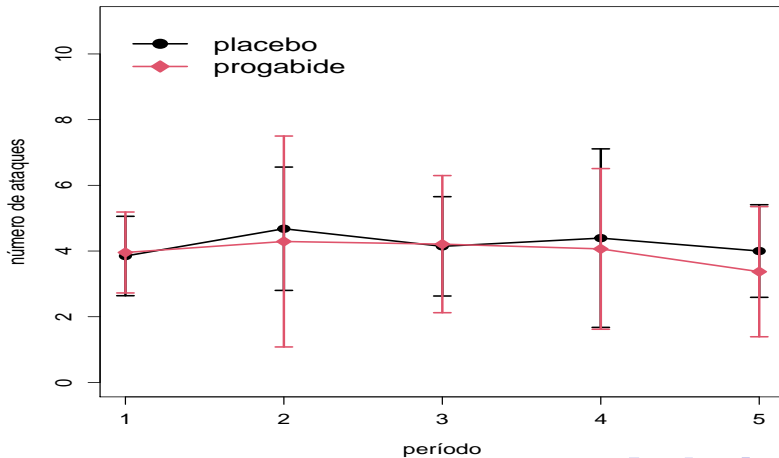
Exemplo 4: Ataques epiléticos

- O interesse da pesquisa é saber se a droga reduz a taxa de ataques epiléticos.
- Estudo irregular: número de semanas varia (8 no primeiro período e 2 nos demais).
- Balanceado em relação a condição de avaliação e desbalanceado em relação ao grupo (28 - placebo e 31 - progabide).
- Completo.

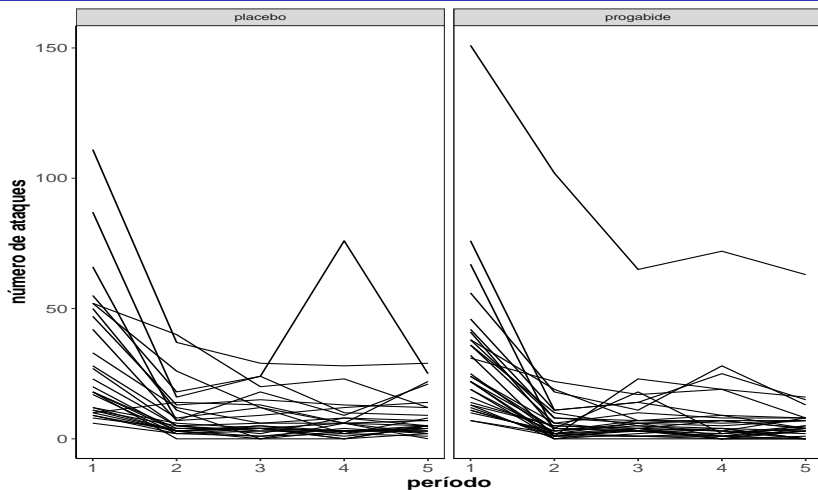
Medidas resumo

Grupo	Período	Media	DP	Var.	CV(%)	Min.	Med.	Max.	n
placebo	1	3,85	3,26	10,65	84,79	0,75	2,38	13,88	28
placebo	2	4,68	5,07	25,69	108,33	0,00	2,50	20,00	28
placebo	3	4,14	4,08	16,66	98,53	0,00	2,25	14,50	28
placebo	4	4,39	7,34	53,82	167,01	0,00	2,50	38,00	28
placebo	5	4,00	3,81	14,48	95,14	0,00	2,50	14,50	28
progabide	1	3,96	3,50	12,24	88,46	0,88	3,00	18,88	31
progabide	2	4,29	9,12	83,18	212,58	0,00	2,00	51,00	31
progabide	3	4,21	5,93	35,16	140,86	0,00	2,50	32,50	31
progabide	4	4,06	6,95	48,26	170,92	0,00	2,00	36,00	31
progabide	5	3,37	5,63	31,67	166,93	0,00	2,00	31,50	31

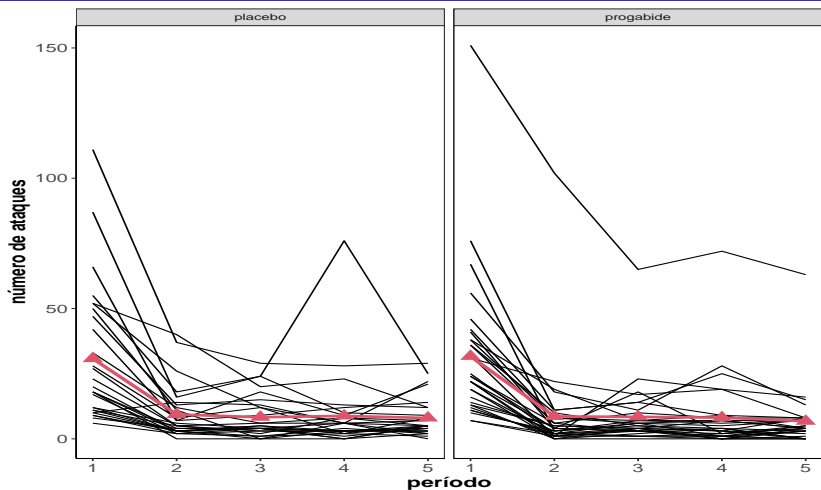
Perfis médios



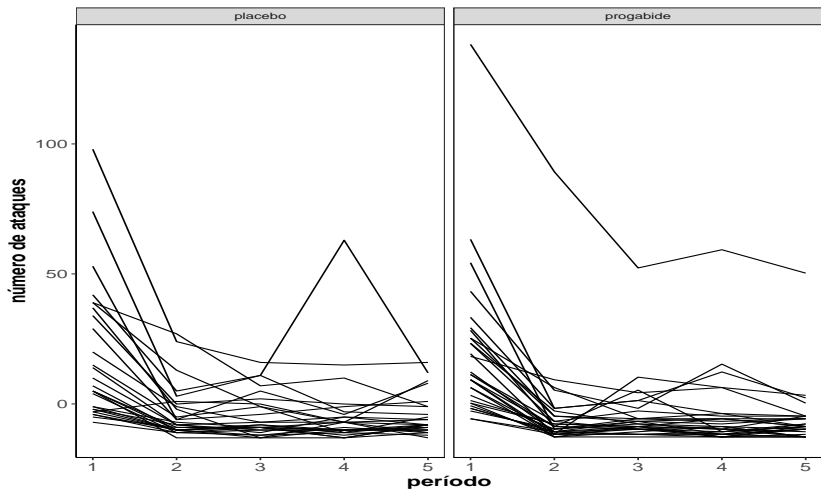
Perfis individuais



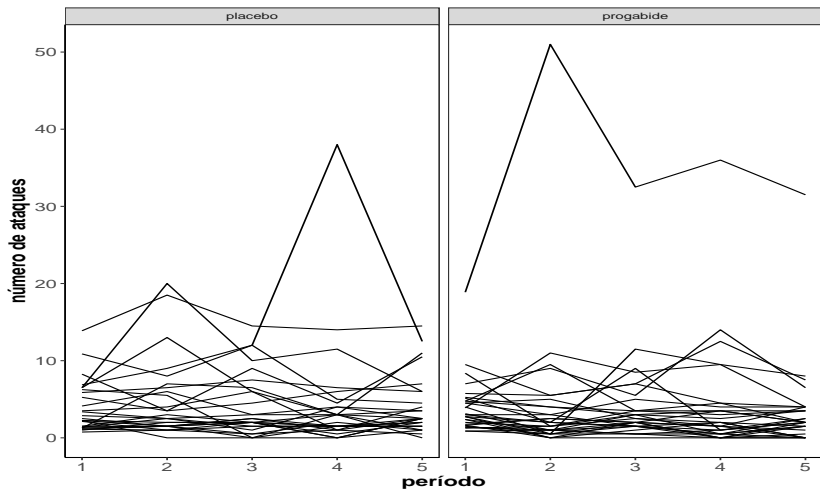
Perfis individuais e médios



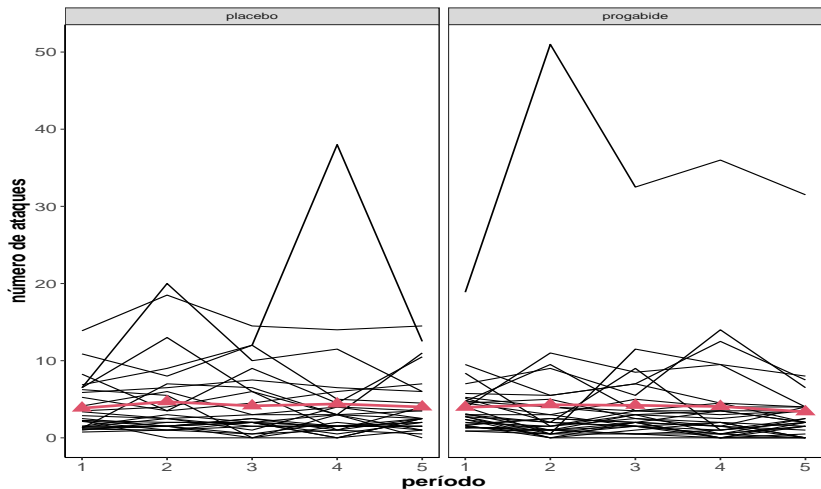
Perfis individuais centrados



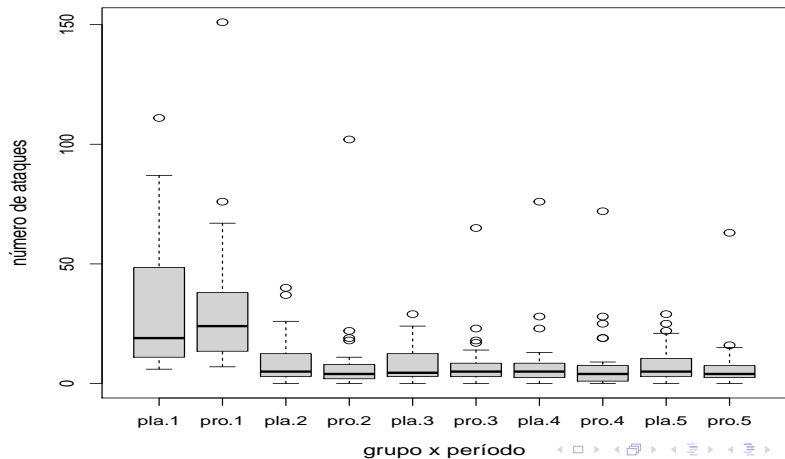
Perfis individuais relativos



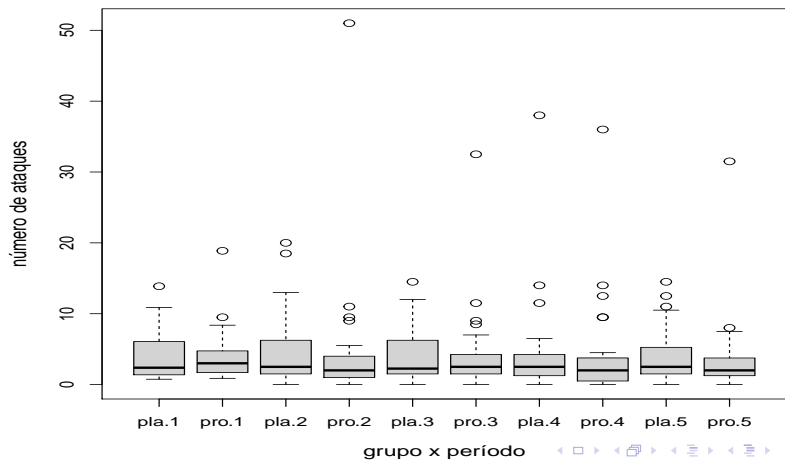
Perfis individuais e médios relativos



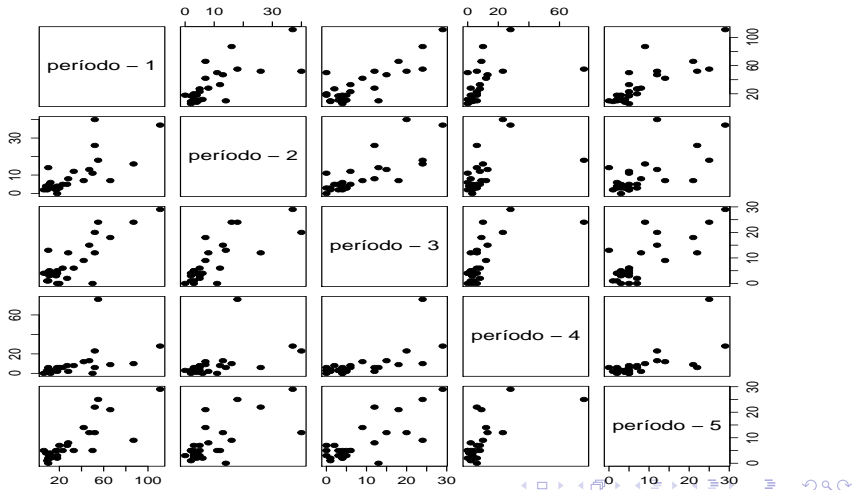
Box plot



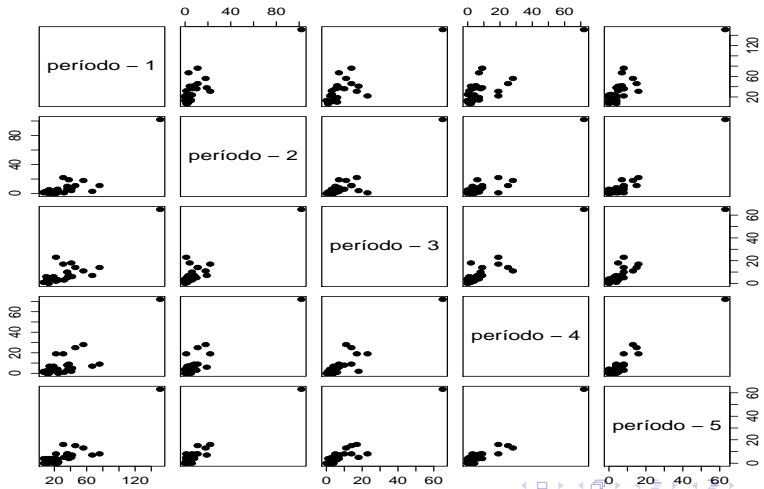
Box plot (ataques/semanas)



Matriz de diagramas de dispersão: placebo



Matriz de diagramas de dispersão: progabide



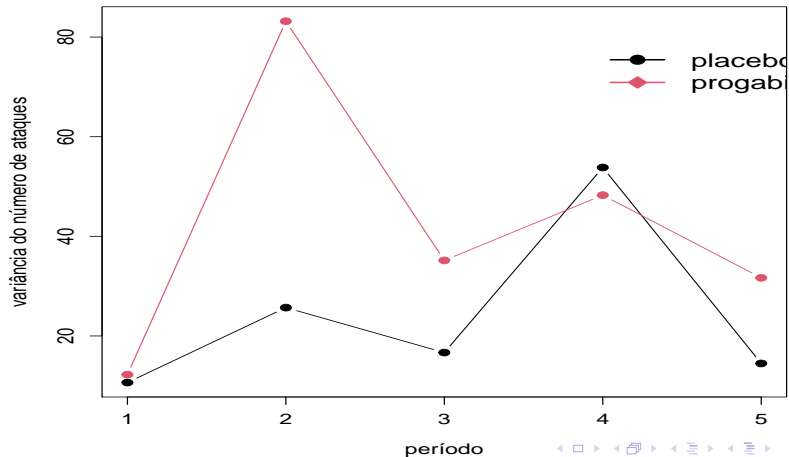
Variâncias (diagonal), correlações (acima) e covariâncias (abaixo): grupo placebo

Período				
1	2	3	4	5
681,43	0,74	0,83	0,49	0,82
196,93	102,76	0,78	0,51	0,68
177,18	64,75	66,66	0,66	0,78
188,88	75,41	79,17	215,29	0,68
162,96	52,26	48,52	75,59	57,93

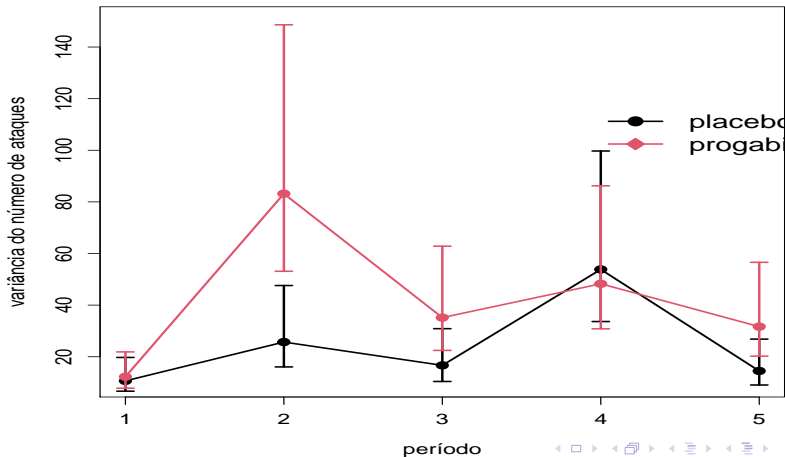
Variâncias (diagonal), correlações (acima) e covariâncias (abaixo): grupo prograbide

Período				
1	2	3	4	5
783,64	0,85	0,85	0,83	0,87
435,98	332,72	0,91	0,91	0,97
280,79	196,22	140,65	0,92	0,95
324,55	231,26	152,41	193,05	0,95
274,97	199,32	126,21	148,83	126,67

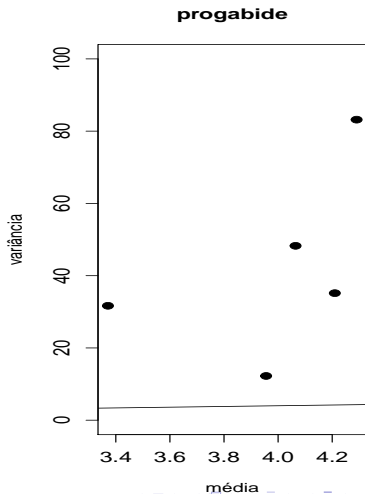
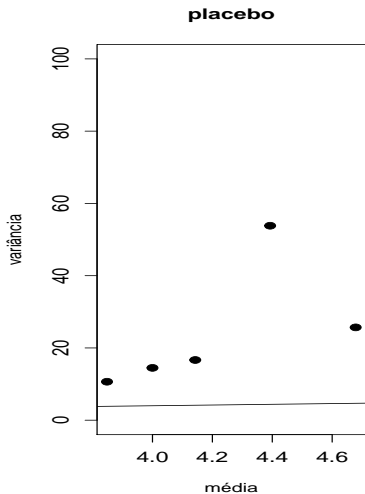
Variâncias em cada condição



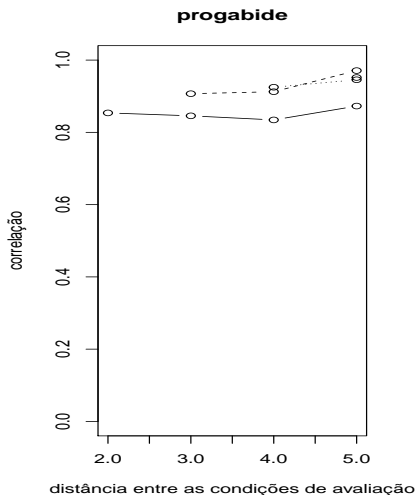
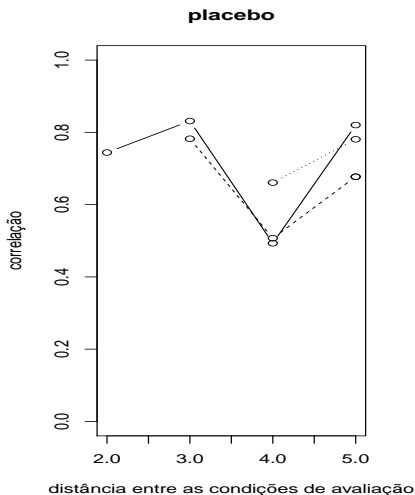
Variâncias em cada condição com intervalos de confiança



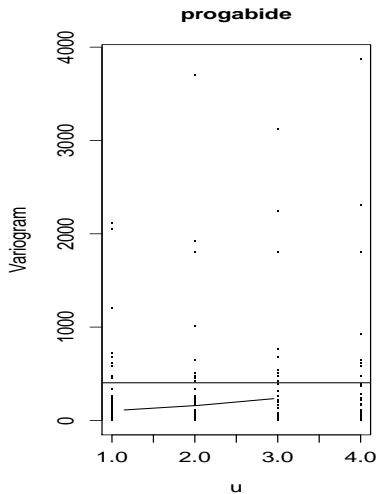
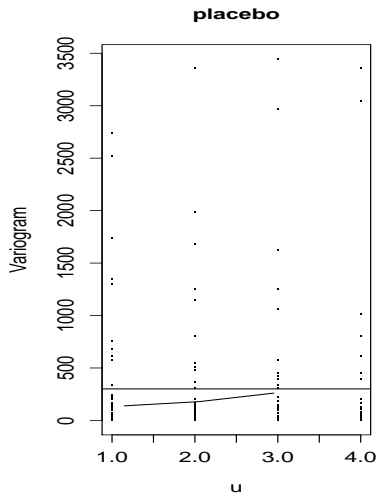
Dispersão entre as médias e as variâncias amostrais



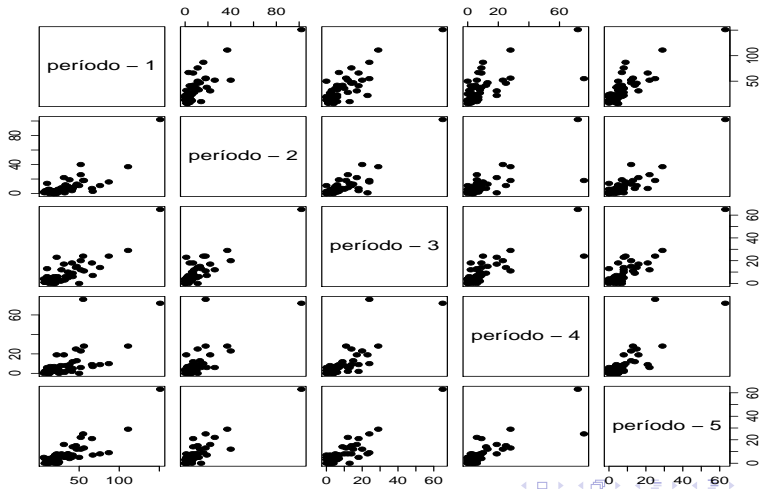
Gráficos dos perfis das linhas da matriz de correlações



Variograma



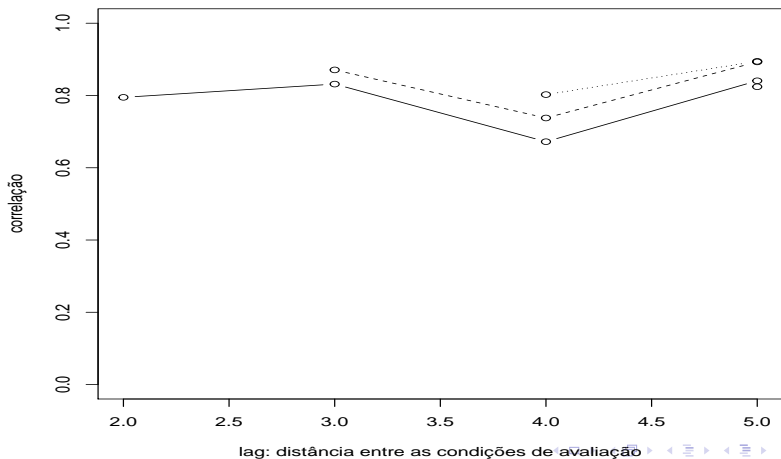
Matriz de diagramas de dispersão



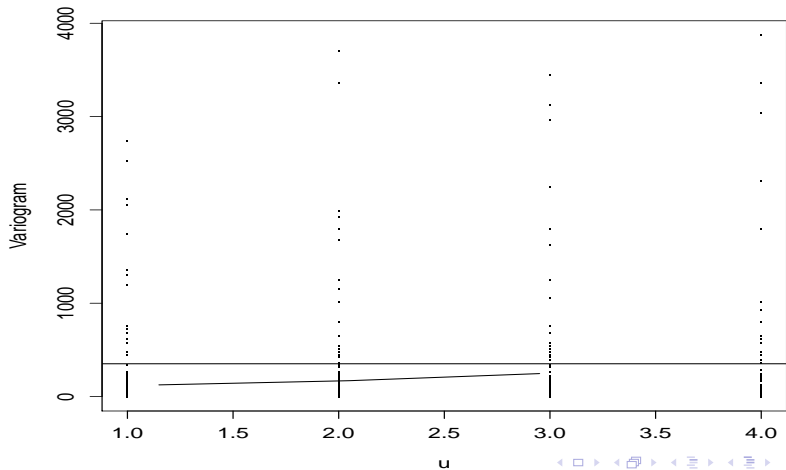
Variâncias (diagonal), correlações (acima) e covariâncias (abaixo)

Período				
1	2	3	4	5
722,74	0,80	0,83	0,67	0,84
317,01	220,08	0,87	0,74	0,89
227,74	131,60	103,78	0,80	0,90
255,65	154,85	115,67	200,18	0,82
217,81	127,67	87,83	112,38	92,88

Gráficos dos perfis das linhas da matriz de correlações



Variograma



Modelo linear generalizado

- Vimos que um MLG é dado por:

$$Y_i \stackrel{ind.}{\sim} FE(\theta_i, \phi) \quad , \quad \theta_i = h(\mu_i), i = 1, \dots, n$$
$$g(\mu_i) = \mathbf{X}_i' \boldsymbol{\beta} = \sum_{j=1}^p x_{ij} \beta_j; \mathbf{X}_i = (x_{1i}, \dots, x_{pi})'$$

Modelo linear generalizado hierárquicos (MLGH)

- O correspondente MLGH de dois níveis é dado por:

$$Y_{ji} | \beta_j \stackrel{ind.}{\sim} FE(\theta_{ji}, \phi) \quad , \quad \theta_{ji} = h(\mu_{ji}), j = 1, \dots, J, i = 1, \dots, n_j$$

$$\beta_j = \mathbf{W}_j \gamma + \mathbf{u}_j \stackrel{ind.}{\sim} N_p(\mathbf{W}_j \gamma, \Psi); \mu_{ij} = \mathcal{E}(Y_{ji} | \beta_j)$$

$$g(\mu_{ji}) = \mathbf{X}'_{ji} \beta_j = \sum_{k=1}^p x_{jik} \beta_{jk};$$

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

$$\mathbf{u}_j = (u_{1j}, \dots, u_{pj})' \stackrel{i.i.d.}{\sim} N_p(0, \Psi)$$

$$\Psi \equiv \Psi(\theta)$$

Modelo linear generalizado misto (MLGM) correspondente

- O correspondente MLGM é dado por:

$$\begin{aligned} Y_{ji} | \mathbf{u}_j &\stackrel{ind.}{\sim} FE(\theta_{ji}, \phi) \quad , \quad \theta_{ji} = h(\mu_{ji}), j = 1, \dots, J, i = 1, \dots, n_j \\ \mathbf{u}_j &\stackrel{i.i.d.}{\sim} N(0, \mathbf{\Psi}); \mu_{ji} = \mathcal{E}(Y_{ji} | \mathbf{u}_j) \\ g(\mu_{ij}) &= \mathbf{Z}'_{ji} \boldsymbol{\gamma} + \mathbf{X}'_{ji} \mathbf{u}_j \\ \mathbf{X}_{kij} &= (x_{ji1}, \dots, x_{jip})' \quad ; \quad \mathbf{Z}_{ij} = (z_{ji1}, \dots, z_{jip})' \\ \mathbf{u}_j &= (u_{1j}, \dots, u_{pj})' \\ \mathbf{\Psi} &\equiv \mathbf{\Psi}(\boldsymbol{\theta}) \end{aligned}$$

Média e estrutura de dependência

- $\mu_{ji}^* = \mathcal{E}(Y_{ji}) = \mathcal{E}(\mathcal{E}(Y_{ji}|\mathbf{u}_j)) = \mathcal{E}(\mu_{ji}) = \mathcal{E}(g^{-1}(\mathbf{Z}'_{ji}\boldsymbol{\gamma} + \mathbf{X}'_{ji}\mathbf{u}_j))$

- Variância

$$\begin{aligned}\mathcal{V}(Y_{ji}) &= \mathcal{V}(\mathcal{E}(Y_{ji}|\mathbf{u}_j)) + \mathcal{E}(\mathcal{V}(Y_{ji}|\mathbf{u}_j)) = \mathcal{V}(g^{-1}(\mathbf{Z}'_{ji}\boldsymbol{\gamma} + \mathbf{X}'_{ji}\mathbf{u}_j)) \\ &+ \mathcal{E}(\phi^{-1}V(\mu_{ji}, \mathbf{u}_j))\end{aligned}$$

em que $V(\cdot)$ é a função de variância (vista anteriormente).

- Covariância

$$\begin{aligned}\text{Cov}(Y_{ji}, Y_{j'i'}) &= \text{Cov}(\mathcal{E}(Y_{ji}|\mathbf{u}_j), \mathcal{E}(Y_{j'i'}|\mathbf{u}_j)) + \underbrace{E(\text{Cov}(Y_{ji}, Y_{j'i'}|\mathbf{u}_j))}_0 \\ &= \text{Cov}(g^{-1}((\mathbf{Z}'_{ji}\boldsymbol{\gamma} + \mathbf{X}'_{ji}\mathbf{u}_j)), g^{-1}((\mathbf{Z}'_{j'i'}\boldsymbol{\gamma} + \mathbf{X}'_{j'i'}\mathbf{u}_j)))\end{aligned}$$

Um modelo

$$Y_{jik} | u_{0j} \stackrel{ind.}{\sim} \text{Poisson}(\mu_{jik}/t_i) \text{ (nível 1)}$$

$$\ln(\mu_{jik}/t_i) = \mu_j + \alpha_k + (\beta_1 + \delta_k)(x_{jik} - x), \alpha_1 = 0, \delta_1 = 0$$

$$\ln(\mu_{jik}) = \ln(t_i) + \mu_j + \alpha_k + (\beta_1 + \delta_k)(x_{jik} - x) \text{ (nível 1),}$$

$$\mu_j = \mu + u_{0j} \text{ (nível 2)}$$

$u_{0j} \stackrel{i.i.d.}{\sim} N(0, \psi)$, $k = 1, 2$ (grupo, 1 – placebo, 2 – progabide),

$j = 1, \dots, n_i$ (indivíduo, $n_1 = 28$; $n_2 = 31$), $i = 1, 2, 3, 4, 5$ (período).

- Y_{jik} : número de ataques do j -ésimo paciente, do k -ésimo grupo no i -ésimo período.
- t_i : número de semanas relativas ao i -ésimo período.
- $\ln(t_i)$: “offset”.

Cont.

- x_{jik} : número de semanas acumuladas ou período correspondente ao i -ésimo instante de avaliação (x é um valor de referência, podendo ser um período específico ou um número específico de semanas acumuladas).
- $\mu_{jik} = t_i e^{\mu + \alpha_k + \beta_k(x_{jik} - x)} e^{u_{0j}}$.
- $\mu_{jik}^* = t_i e^{\mu + \alpha_k + \beta_k(x_{jik} - x)} \mathcal{E}(e^{u_{0j}})$.
- $\mu_{jik}^* | (x_{jik} = x) = t_i e^{\mu + \alpha_k} \mathcal{E}(e^{u_{0j}})$.
- $\mu_{ji1}^* | (x_{ji1} = x) = t_i e^{\mu} \mathcal{E}(e^{u_{0j}})$.
- $\mu_{jik}^* | (x_{jik}^* = x_{jik} + 1) = t_{i+1} e^{\mu + \alpha_k + \beta_k(x_{jik} - x)} e^{\beta_i} \mathcal{E}(e^{u_{0j}})$.

Cont.

- $\mu_{ji1}^* = t_i e^{\mu + \beta_1(x_{ji1} - x)} \mathcal{E}(e^{u_{0j}})$ (grupo placebo).
- $\mu_{ji2}^* = t_k e^{\mu + \alpha_2 + (\beta_1 + \delta_2)(x_{ji2} - x)} \mathcal{E}(e^{u_{0j}})$ (grupo progabide).
- $\frac{\mu_{ji2}^*}{\mu_{ji1}^*} = e^{\alpha_2 + \delta_2(a - x)}$ (comparando os grupos) para $x_{1jk} = x_{ji2} = a$.
- $\frac{\mu_{ji2}^*}{\mu_{ji1}^*} = e^{\alpha_2}$ (comparando os grupos no período 1).
- $\frac{\mu_{jik}^* | (x_{jik}^* = x_{jik} + 1)}{\mu_{jik}^* | x_{jik}} = \frac{t_{i+1}}{t_i} e^{\beta_k}$ (comparando dois períodos subsequentes), em que $\beta_k = \beta_1 + \delta_k$.
- $\frac{\mu_{jik}^* | (x_{jik}^* = x_{jik} + r)}{\mu_{jik}^* | x_{jik}} = \frac{t_{i+r}}{t_i} e^{i\beta_k}$ (comparando dois períodos quaisquer).

Cont.

- Se $u_{0j} \sim N(0, \psi)$, então $e^{u_{0j}} \sim \text{log-normal}(0, \psi)$. Neste caso,

$$\begin{aligned}\mathcal{E}(Y_{jik}) &= \mathcal{E}(\mathcal{E}(Y_{jik}|u_{0j})) = t_i e^{\mu + \alpha_k + \beta_k(x_{jik} - x)} \mathcal{E}(e^{u_{0j}}) \\ &= t_i e^{\mu + \alpha_k + \beta_k(x_{jik} - x)} e^{\psi/2}\end{aligned}$$

$$\begin{aligned}\mathcal{V}(Y_{jik}) &= \mathcal{V}(\mathcal{E}(Y_{jik}|u_{0j})) + \mathcal{E}(\mathcal{V}(Y_{jik}|u_{0j})) = t_i^2 e^{2(\mu + \alpha_k + \beta_k(x_{jik} - x))} \mathcal{V}(e^{u_{0j}}) \\ &+ t_i e^{\mu + \alpha_k + \beta_k(x_{jik} - x)} \mathcal{E}(e^{u_{0j}}) \\ &= t_i^2 e^{2(\mu + \alpha_k + \beta_k(x_{jik} - x))} (e^{\psi} - 1) e^{\psi} + t_i e^{\mu + \alpha_k + \beta_k(x_{jik} - x)} e^{\psi/2} \\ &> t_i e^{\mu + \alpha_k + \beta_k(x_{jik} - x)} e^{\psi/2} = \mathcal{E}(Y_{jik})\end{aligned}$$

Estimação

- Parâmetros (efeitos aleatórios) a serem estimados: γ (efeitos fixos/parâmetros de regressão do nível 2), ϕ (parâmetro de precisão), θ (ligados a matriz de variâncias e covariâncias (Ψ) dos efeitos aleatórios) e efeitos aleatórios (\mathbf{u}).
- Verossimilhança (completa) $\mathbf{u} = (\mathbf{u}'_1, \dots, \mathbf{u}'_J)'$:

$$L(\gamma, \phi, \theta, \mathbf{u}) = \exp \left\{ \phi \left[\sum_{j=1}^J \sum_{i=1}^{n_j} y_{ji} \theta_{ij} - \sum_{j=1}^J b(\theta_{jj}) \right] + \sum_{j=1}^J \sum_{i=1}^{k_j} c(y_{ji}, \phi) \right\} \\ \times \exp \left\{ -0,5 \sum_{j=1}^J \mathbf{u}'_j \Psi^{-1} \mathbf{u}_j \right\} |\Psi|^{-n/2}$$

Verossimilhança marginal

- Verossimilhança marginal (em relação à \mathbf{u})

$$L(\boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\theta}) = \prod_{j=1}^J \int_{\mathbb{R}^p} \exp \left\{ \boldsymbol{\phi} \left[\sum_{i=1}^{n_j} y_{ji} \boldsymbol{\theta}_{ji} - b(\boldsymbol{\theta}_{ji}) \right] + \sum_{i=1}^{n_j} c(y_{ji}, \boldsymbol{\phi}) \right\} \\ \times \exp \left\{ -0,5 \mathbf{u}' \boldsymbol{\Psi}^{-1} \mathbf{u} \right\} |\boldsymbol{\Psi}|^{-1/2} d\mathbf{u}$$

- A integral acima não tem solução explícita (lembrando que $\boldsymbol{\theta}_{ij}$ é função, neste caso, de \mathbf{u}).

Resolução da integral

- Quadratura (gaussiana).
- Quadratura adaptativa.
- Aproximação de Laplace.
- Monte carlo.
- Veja mais: [aqui](#), [aqui](#) e [aqui](#).

Resolução da integral

- Quadratura (gaussiana), Quadratura adaptativa e Monte Carlo - aproximar a integral por somas de áreas de retângulos/paralelepípedos: [aqui](#) e [aqui](#).
- No R, por exemplo: [integrate](#) e [cubature](#).
- Os dois métodos acima podem ser adaptados para outras distribuições dos efeitos aleatórios, assim como os pacotes citados suportam outras distribuições.
- Veja [aqui](#) sobre o uso da quadratura adaptativa.

Verossimilhança marginal numericamente aproximada

$$L(\gamma, \phi, \theta) \approx \prod_{j=1}^J \sum_{r \in \mathcal{I}} \exp \left\{ \phi \left[\sum_{i=1}^{n_j} y_{ji} \theta_{jir} - b(\theta_{jir}) \right] + \sum_{i=1}^{n_j} c(y_{ji}, \phi) \right\} \\ \times \exp \left\{ -0,5 \mathbf{u}_r' \Psi^{-1} \mathbf{u}_r \right\} |\Psi|^{-1/2} A_r$$

em que \mathcal{I} é o conjunto de todas as k-uplas, A_r (essencialmente é o volume do paralelepípedo) são os chamados pesos de quadratura e \mathbf{u}_r o vetor de pontos de quadraturas.

Verossimilhança marginal numericamente aproximada: um único efeito aleatório

$$L(\gamma, \phi, \theta) \approx \prod_{j=1}^J \sum_{r=1}^m \exp \left\{ \phi \left[\sum_{i=1}^{n_j} y_{ji} \theta_{jir} - b(\theta_{jir}) \right] + \sum_{i=1}^{n_j} c(y_{ji}, \phi) \right\} \\ \times \exp \left\{ \frac{-u_r^2}{2\psi} \right\} \psi^{-1/2} A_r$$

Equações de estimação

- Dado que aproximamos a integral de interesse (por algum dos métodos anteriores), temos que derivar

$$l(\boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\theta}) = \sum_{j=1}^J \ln \int_{\mathbb{R}^p} \exp \left\{ \boldsymbol{\phi} \left[\sum_{i=1}^{n_j} y_{ji} \boldsymbol{\theta}_{ji} - b(\boldsymbol{\theta}_{ji}) \right] + \sum_{i=1}^{n_j} c(y_{ji}, \boldsymbol{\phi}) \right\} \\ \times \exp \left\{ -0,5 \mathbf{u}' \boldsymbol{\Psi}^{-1} \mathbf{u} \right\} |\boldsymbol{\Psi}|^{-1/2} d\mathbf{u}$$

em relação à $\boldsymbol{\gamma}$, $\boldsymbol{\phi}$ e $\boldsymbol{\theta}$.

Equações de estimação

- As derivadas com relação à γ e ϕ assemelham-se à (1) e (2), respectivamente.
- Com relação à θ **depende da estrutura adotada** (o que também depende do modelo).
- Mesmo a equação de estimação com relação à θ não tem, em geral, solução explícita.

Equações de estimação

- O sistema de equações gerado por

$$\begin{cases} \mathbf{S}(\tilde{\gamma}) = 0 \\ \mathbf{S}(\tilde{\phi}) = 0 \\ \mathbf{S}(\tilde{\theta}) = 0 \end{cases}$$

não tem solução explícita e algum [algoritmo de otimização](#) tem de ser usado (Newton-Raphson, Score de Fisher, FBGS, Nelder-Mead etc) ([link](#)).

Equações de estimação

- Preditores para os efeitos aleatórios podem ser considerados usando-se a metodologia vista para os **modelos lineares mistos**.
- Nesse caso, a distribuição “à posteriori” dos efeitos aleatórios não é obtível analiticamente e **métodos numéricos apropriados** têm de ser utilizados.

Aproximação de Laplace

- Trabalha-se com

$$\begin{aligned}L(\gamma, \phi, \theta) &= \prod_{j=1}^J \int_{\mathbb{R}^p} \exp \left\{ \phi \left[\sum_{i=1}^{n_j} y_{ji} \theta_{ji} - b(\theta_{ji}) \right] + \sum_{i=1}^{n_j} c(y_{ji}, \phi) \right\} \\ &\times \exp \left\{ -0,5 \mathbf{u}' \Psi^{-1} \mathbf{u} \right\} |\Psi|^{-1/2} d\mathbf{u} \\ &= |\Psi|^{-J/2} \prod_{j=1}^n \int_{\mathbb{R}^p} \exp(Q_j(\mathbf{u})) d\mathbf{u} = I\end{aligned}\quad (4)$$

- Em que:

$$\begin{aligned}Q_j(\mathbf{u}) &= \phi \left[\sum_{i=1}^{n_j} y_{ji} \theta_{ji} - b(\theta_{ji}) \right] + \sum_{i=1}^{n_j} c(y_{ji}, \phi) \\ &\times -0,5 \mathbf{u}' \Psi^{-1} \mathbf{u}\end{aligned}$$

Aproximação de Laplace

- Seja $\hat{\mathbf{u}}_j$ o valor que maximiza $Q_j(\mathbf{u})$ e considere a aproximação de $Q_j(\mathbf{u})$ por uma expansão de segunda ordem em séries de Taylor, ou seja:

$$Q_j(\mathbf{u}) \equiv Q(\mathbf{u}_j) = Q(\hat{\mathbf{u}}_j) + \frac{1}{2}(\mathbf{u}_j - \hat{\mathbf{u}}_j)' Q''(\hat{\mathbf{u}}_j)(\mathbf{u}_j - \hat{\mathbf{u}}_j)$$

- Assim, temos a seguinte aproximação (de (4)):

$$\begin{aligned} I &\approx |\Psi|^{-J/2} \prod_{j=1}^J \exp(Q(\hat{\mathbf{u}}_j)) \int_{\mathbb{R}^p} \exp \left\{ \frac{1}{2}(\mathbf{u} - \hat{\mathbf{u}}_j)' Q''(\hat{\mathbf{u}}_j)(\mathbf{u} - \hat{\mathbf{u}}_j) \right\} d\mathbf{u} \\ &= |\Psi|^{-J/2} \prod_{j=1}^J ((2\pi)^{-p/2} | - Q''(\hat{\mathbf{u}}_j) |^{-1/2} \exp(Q(\hat{\mathbf{u}}_j))) \end{aligned} \quad (5)$$

Aproximação de Laplace

- Em que

$$-Q''(\mathbf{u}_j) = \Psi^{-1} + \phi^{-1} \sum_{i=1}^{k_j} \mathbf{x}_{ji} b''(\mu_{ji}, \mathbf{u}_{ji}) \mathbf{x}'_{ji}, b(\mu_{ji}, \mathbf{u}_{ji}) \equiv b(\theta_{ji})$$

- A maximização numérica da logverossimilhança é feita de modo iterativo alternando-se as atualizações de \mathbf{u}_j (maximizando-se $Q(\mathbf{u}_j)$) e $(\gamma, \phi, \boldsymbol{\theta})$ (maximizando-se (5)).

Quase-verossimilhança penalizada

- Trabalha-se com uma aproximação das observações (Y_{ji}) usando a média $\mathcal{E}(Y_{ji}|\mathbf{u}_j)$ e erros (normalmente distribuídos) com média 0 e variância $\mathcal{V}(Y_{ji}|\mathbf{u}_j)$ em torno de estimativas provisórias de (γ, \mathbf{u}_j) digamos $\tilde{\gamma}$ e $\tilde{\mathbf{u}}_j$ e considerando $h(\cdot) = g^{-1}(\cdot)$, ou seja:

$$\begin{aligned} Y_{ji} &\approx h(\mu_{ji}, \mathbf{u}_j) + \xi_{ji} \approx h(\tilde{\mu}_{ji}, \tilde{\mathbf{u}}_j) + h'(\tilde{\mu}_{ji}, \tilde{\mathbf{u}}_j) \mathbf{z}'_{ji} (\gamma - \tilde{\gamma}) \\ &+ h'(\tilde{\mu}_{ji}, \tilde{\mathbf{u}}_j) \mathbf{x}'_{ji} (\mathbf{u}_j - \tilde{\mathbf{u}}_j) + \xi_{ji} \\ &= \tilde{\mu}_{ji} + V(\tilde{\mu}_{ji}) \mathbf{z}'_{ji} (\gamma - \tilde{\gamma}) + V(\tilde{\mu}_{ji}) \mathbf{z}'_{ji} (\mathbf{u}_j - \tilde{\mathbf{u}}_j) + \xi_{ji} \end{aligned}$$

Quase-verossimilhança penalizada

- Matricialmente

$$\mathbf{Y}_j = \tilde{\boldsymbol{\mu}}_j + \tilde{\mathbf{V}}_j \mathbf{Z}_j (\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}) + \tilde{\mathbf{V}}_j \mathbf{X}_j (\mathbf{u}_j - \tilde{\mathbf{u}}_j) + \boldsymbol{\xi}_j$$

em que $\tilde{\boldsymbol{\mu}}_j = (\tilde{\mu}_{j1}, \dots, \tilde{\mu}_{jn_j})'$, $\tilde{\mathbf{V}}_j$ é uma matriz diagonal com elementos $V(\tilde{\mu}_{ji})$ e \mathbf{X}_j e \mathbf{Z}_j contêm os vetores \mathbf{x}_{ji} e \mathbf{z}_{ji} .

- Rearranjando os termos da equação acima e pré-multiplicando por $\tilde{\mathbf{V}}_j^{-1}$ temos que:

$$\mathbf{Y}_j^* := \tilde{\mathbf{V}}_j^{-1} (\mathbf{Y}_j - \tilde{\boldsymbol{\mu}}_j) + \mathbf{Z}_j \tilde{\boldsymbol{\gamma}} + \mathbf{X}_j \tilde{\mathbf{u}}_j \quad (6)$$

$$\approx \mathbf{Z}_j \boldsymbol{\gamma} + \mathbf{X}_j \mathbf{u}_j + \boldsymbol{\xi}_j^* \quad (7)$$

Quase-verossimilhança penalizada

- Em que $\xi_j^* = \tilde{\mathbf{V}}_j^{-1} \xi_j$.
- O modelo (7) pode ser visto como um modelo linear misto para os pseudo-dados \mathbf{Y}_j^* .
- Algoritmo
 - Passo A: Para valores atualizados de γ, ϕ e θ calculam-se valores preditos para \mathbf{u}_j e os pseudo-dados \mathbf{Y}_j^* (Equação (6)).
 - Passo B: Para os pseudo-dados \mathbf{Y}_j^* ajusta-se o modelo (7) e as estimativas para γ, ϕ e θ são atualizadas.

Comentários

- Para qualquer um dos métodos anteriores de estimação, em geral, a convergência em distribuição para normal multivariada dos estimadores de (γ, ϕ, θ) , sob certas condições de regularidade, são observadas.
- Em geral, a informação de Fisher associada (observada ou esperada) fornece boas aproximações para a matriz de covariâncias assintóticas dos respectivos estimadores.
- Intervalos de confiança e testes de hipótese podem ser construídos de modelo semelhante ao que foi feito para os [modelos lineares mistos](#) e [modelos lineares generalizados](#).

Laplace

- A aproximação de Laplace tende a ser mais rápida do que as abordagens de quadratura adaptativa e QVP (quase-verossimilhança penalizada).
- A aproximação tende a se melhor para valores elevados do número de unidades no nível 1 (n_j) e para respostas menos “discretizadas”, por exemplo, tende a ser melhor para modelos de regressão de Poisson do que para dados binários.
- A aproximação pode ser melhorada considerando-se mais termos na expansão em séries de Taylor.

QVP

- Tal abordagem tem esse nome por se basear numa quase-verossimilhança envolvendo apenas o primeiro e o segundo momentos condicionais mais um termo de penalização devido aos efeitos aleatórios.
- O método de QVP usa uma aproximação da logverossimilhança e é exato apenas para MLM. Funciona melhor à medida que a distribuição da variável resposta se aproxima da normal, por exemplo, distribuição de Poisson com médias elevadas ou binomial com valores elevados para o número de repetições.
- As estimativas são inconsistentes. Existem algumas correções para o viés ([Breslow & Lin, Biometrika, 1995](#), [Lin & Breslow, JASA, 1996](#)).

Comentários e outros métodos de estimação

- Mais detalhes sobre os processos de inferência apresentados podem ser encontrados [aqui](#), [aqui](#), [aqui](#) e [aqui](#).
- Outros métodos de estimação
 - Algoritmo EM.
 - Algoritmo MCEM (Monte Carlo EM).
 - Algoritmo SEM (Stochastic EM).
 - Algoritmo SAEM (Stochastic approximation EM).
 - Algoritmo CADEM.

Resultados

Laplace

Parâmetro	Estimativa	Erro-padrão	IC(95%)	Estat. z	p-valor
μ	1,07	0,15	[0,77 ; 1,36]	7,03	< 0,0001
α_2	-0,03	0,21	[-0,44 ; 0,38]	-0,13	0,8955
β_1	0,02	0,02	[-0,01 ; 0,05]	1,23	0,2202
δ_2	-0,04	0,02	[-0,08 ; 0,00]	-1,75	0,0815

Os resultados indicam que os pacientes dos dois grupos são equivalentes no primeiro período ($\alpha_2 = 0$) e que a droga tem um efeito (marginal) decrescente no número de ataques ($\delta_2 = -0,04$) e que não há um “efeito placebo” ($\beta_1 = 0$). Além disso, os efeitos aleatórios parecem ser significativos ($\tilde{\psi} = 0,61$).

Resultados

Quadratura adaptativa

Parâmetro	Estimativa	Erro-padrão	IC(95%)	Estat. z	p-valor
μ	1,07	0,15	[0,77 ; 1,36]	7,03	< 0,0001
α_2	-0,03	0,21	[-0,44 ; 0,38]	-0,13	0,8995
β_1	0,02	0,02	[-0,01 ; 0,05]	1,23	0,2185
δ_2	-0,04	0,02	[-0,08 ; 0,00]	-1,75	0,0804

Os resultados indicam que os pacientes dos dois grupos são equivalentes no primeiro período ($\alpha_2 = 0$) e que a drogram tem um efeito (marginal) decrescente no número de ataques ($\delta_2 = -0,04$) e que não há um “efeito placebo” ($\beta_1 = 0$). Além disso, os efeitos aleatórios parecem ser significativos ($\tilde{\psi} = 0,61$).

Resultados

QVP

Parâmetro	Estimativa	Erro-padrão	IC(95%)	Estat. z	p-valor
μ	1,11	0,15	[0,81 ; 1.41]	7,20	<0,0001
α_2	-0,02	0,21	[-0,44 ; 0.39]	-0,11	0,9121
β_1	0,02	0,03	[-0,03 ; 0.07]	0,72	0,4715
δ_2	-0,04	0,04	[-0,11 ; 0.03]	-1,02	0,3065

Os resultados indicam que os pacientes dos dois grupos são equivalentes ao longo de todos os períodos ($\alpha_2 = \beta_1 = \delta_2 = 0$). Além disso, os efeitos aleatórios parecem ser significativos ($\tilde{\psi} = 0,57$).

Tipos de resíduos em modelos lineares generalizados (mistos)

- Resíduo ordinário (“response residual”)

$$R_{ji}^R = Y_{ji} - \hat{\mu}_{ji}$$

μ_{ji} : média condicional.

- Resíduo de Pearson (“Pearson residual”)

$$R_{ji}^P = \frac{Y_{ji} - \hat{\mu}_{ji}}{\sqrt{\mathcal{V}(\hat{\mu}_{ji})}}$$

$V(\mu_{ji})$: função de variância.

Tipos de resíduos em modelos lineares generalizados (mistos)

- Resíduo componente do desvio (RCD) (“deviance residual”)

$$R_{ji}^D = \left(\text{sign}(Y_{ji} - \hat{\mu}_{ji}) \sqrt{\hat{d}_{ji}} \right)$$

em que d_{ji} é a a função desvio (para os MLGM o desvio é calculado, em geral, usando-se a log-verossimilhança marginal).

Comentários: para os MLG

- Em geral, não se espera que os resíduos ordinário e de Pearson tenham distribuição $N(0,1)$ (nem mesmo aproximadamente), salvo em alguns poucos casos.
- Por exemplo, se $Y \sim \text{Poisson}(\lambda)$, então $Z = \frac{Y - \lambda}{\sqrt{\lambda}} \xrightarrow[\lambda \rightarrow \infty]{D} N(0, 1)$.
- Espera-se, sob o bom ajuste do modelo, que o RCD (padronizado, veja [Paula \(2013\)](#)), tenha aproximadamente distribuição $N(0,1)$, sob certas condições (em geral tamanho de amostra suficientemente grande, $J \rightarrow \infty$, $n_j \rightarrow \infty, \forall j$ e/ou $\phi \rightarrow \infty$).

Comentários: para os MLG

- Resíduos ordinários: detecção de outliers (má predição do valor observado).
- Resíduos de Pearson: homocedasticidade.
- Resíduos componente do desvio: detecção de outliers, homocedasticidade e adequação do modelo (distribuição da resposta, função de ligação, preditor linear).

Comentários: para os MLGM

- Problema: o preditor linear é aleatório.
- Em relação aos resíduos ordinário e de Pearson espera-se a mesma habilidade vista para os MLG, desde que devidamente adaptados.
- Contudo, não se espera normalidade do RCD mesmo sob o bom ajuste do modelo, no caso dos MLGM.

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 simular 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 (de referência) 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 $(\tilde{\gamma}', \tilde{\phi}')'$ e os valores preditos (\tilde{u}_j) e calcule o RCD para cada observação, $(t_{D_{ji}}), j = 1, 2, \dots, J, i = 1, \dots, n_j$.

Cont.

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

Cont.

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

$$\mathbf{T}_1 = \begin{bmatrix} t_{D_{111}}^* & t_{D_{112}}^* & \cdots & t_{D_{11m}}^* \\ t_{D_{121}}^* & t_{D_{122}}^* & \cdots & t_{D_{12m}}^* \\ \vdots & \vdots & \ddots & \vdots \\ t_{D_{Jn_j1}}^* & t_{D_{Jn_j2}}^* & \cdots & t_{D_{Jn_jm}}^* \end{bmatrix}$$

Cont.

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

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

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

Cont.

5) Na prática considera-se $t_{(ji)l}^* = \frac{t_{D_{(ji)(2)}}^* + t_{D_{(ji)(3)}}^*}{2}$ e

$t_{(ji)s}^* = \frac{t_{D_{(ji)(m-2)}}^* + t_{D_{(ji)}}^*}{2}$ (refinamento das estimativas dos limites do envelope), em que $t_{D_{(ji)(r)}}^*$ é a r -ésima estatística de ordem dentro de cada linha, $j = 1, 2, \dots, J, i = 1, 2, \dots, n_j$.

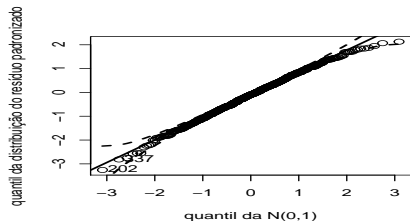
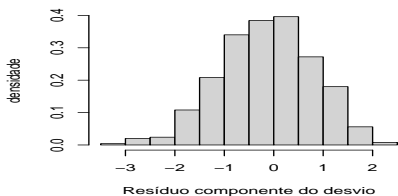
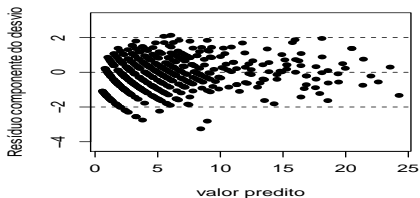
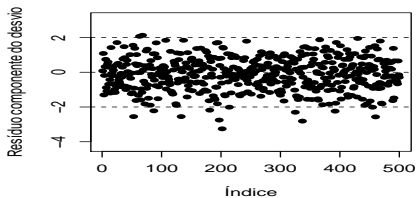
- Além disso, consideramos como a linha de referência

$$t_{(ji)}^* = \frac{1}{m} \sum_{r=1}^m t_{D_{(ji)r}}^*, j = 1, 2, \dots, J; i = 1, \dots, n_j.$$

Estudo de simulação

- Simulou-se dois modelos.
- Modelo 1: $Y_{ji}|u_j \stackrel{ind.}{\sim} \text{Poisson}(\mu_{ji}), i = 1, 2, \dots, 5; j = 1, 2, \dots, 100,$
 $\ln \mu_{ji} = 1 + 0,1x_i + u_j; x_i = i; u_j \stackrel{i.i.d.}{\sim} N(0; 0,5).$
- Modelo 2:
 $Y_{ji}|u_j \stackrel{ind.}{\sim} \text{Binomial-negativa}(\mu_{ji}, \phi), i = 1, 2, \dots, 5; j = 1, 2, \dots, 100,$
 $\ln \mu_{ji} = 1 + 0,1x_i + b_j; x_i = i; b_j \stackrel{i.i.d.}{\sim} N(0; 0,5)$ (veja [Paula \(2013\)](#)).
- Nos gráficos de envelope sem título, a simulação dos envelopes via distribuição normal padrão.
- Nos gráficos de envelope com título, a simulação dos envelopes via modelo.

Resultados: modelo 1 - RCD



Resultados: modelo 1 - RCD

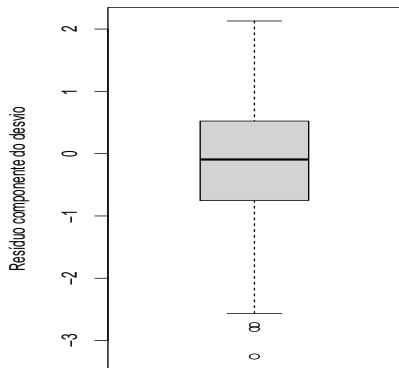
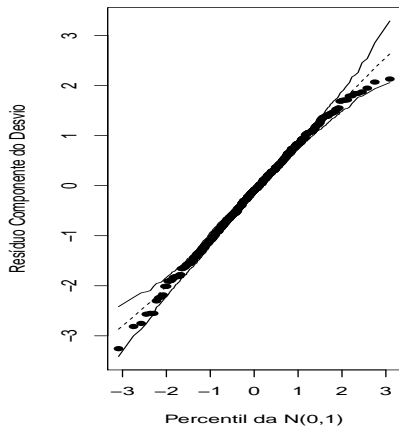
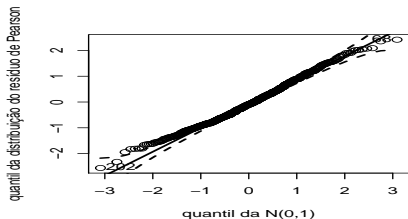
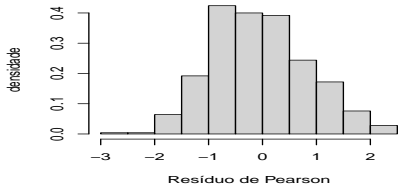
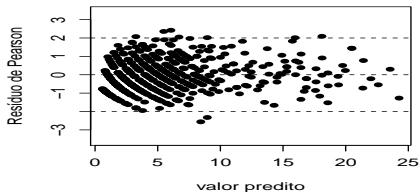
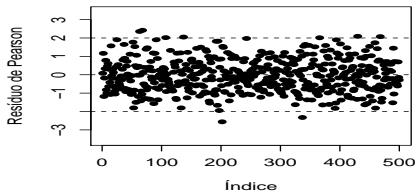


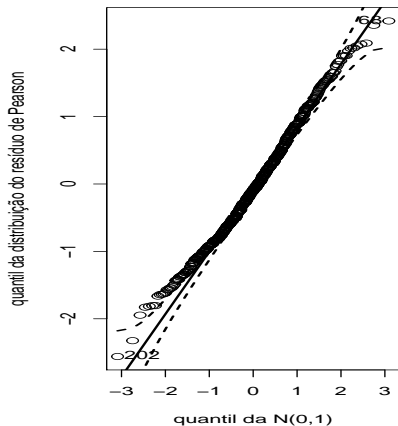
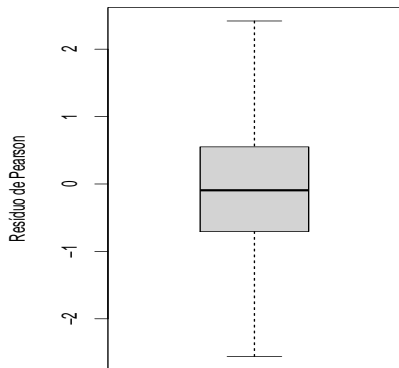
Gráfico de quantil-quantil normal



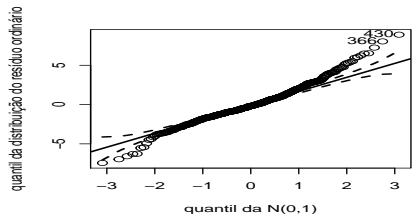
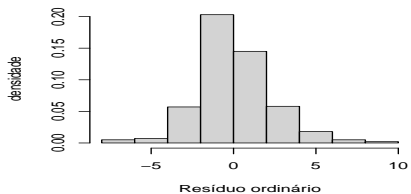
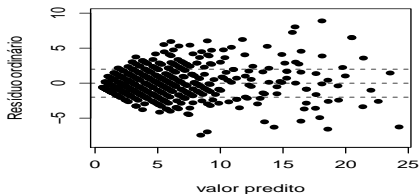
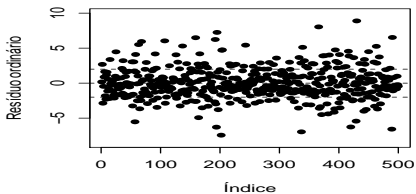
Resultados: modelo 1 - Pearson



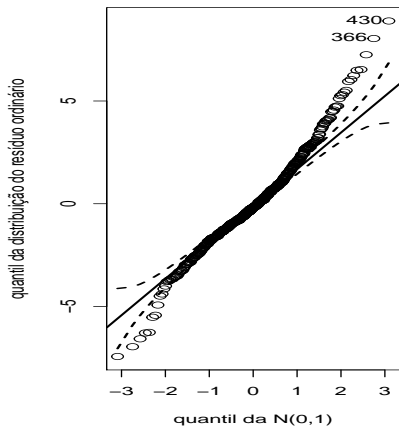
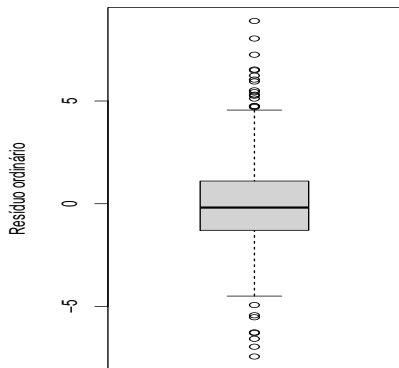
Resultados: modelo 1 - Pearson



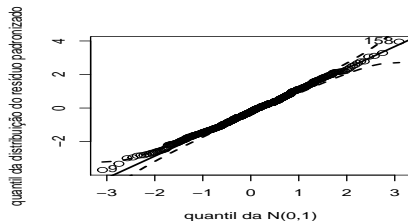
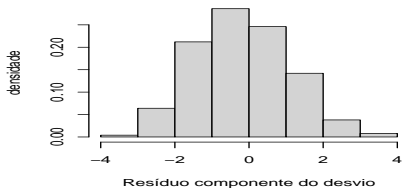
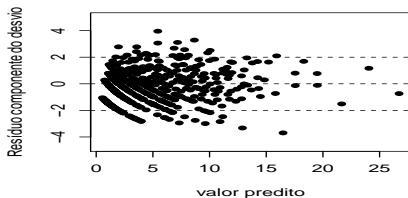
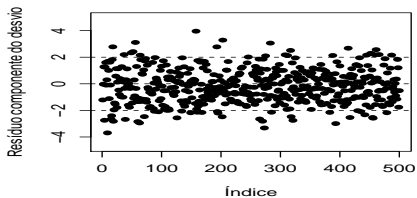
Resultados: modelo 1 - Ordinário



Resultados: modelo 1 - Ordinário



Resultados: modelo 2 - RCD



Resultados: modelo 2 - RCD

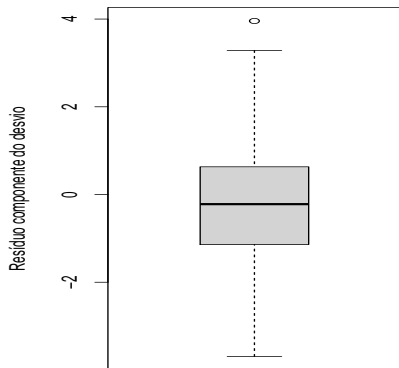
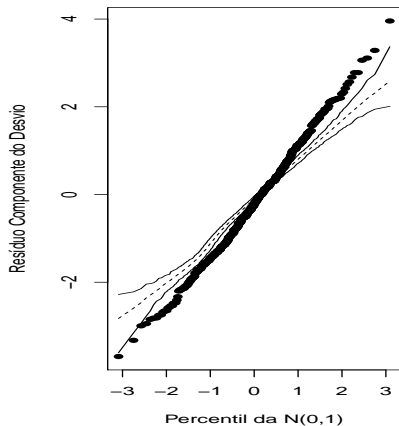
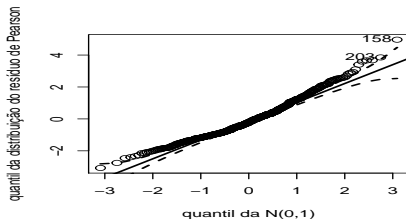
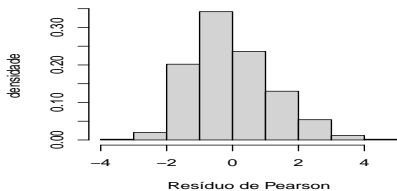
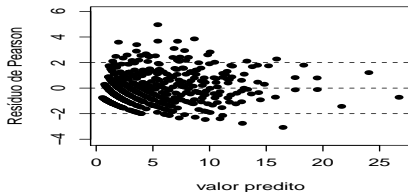
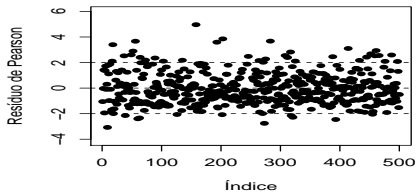


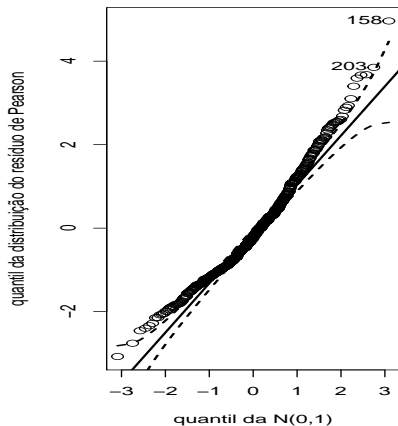
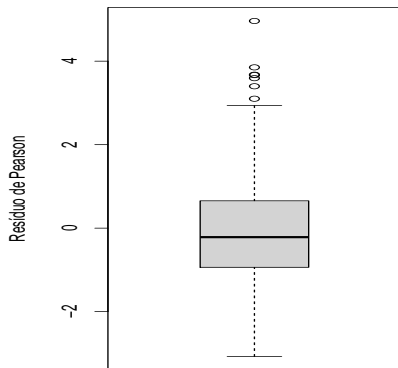
Gráfico de quantil-quantil normal



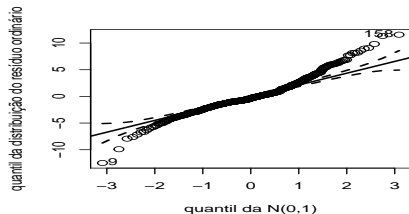
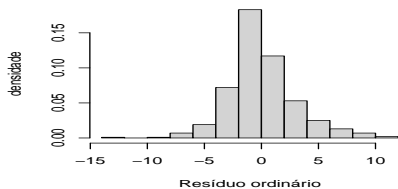
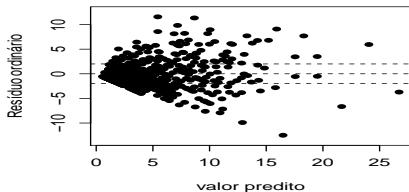
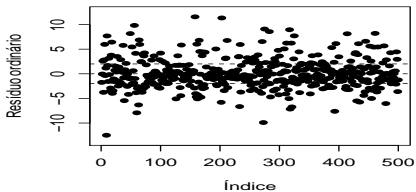
Resultados: modelo 2 - Pearson



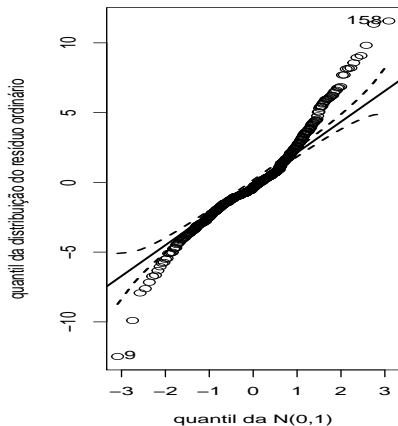
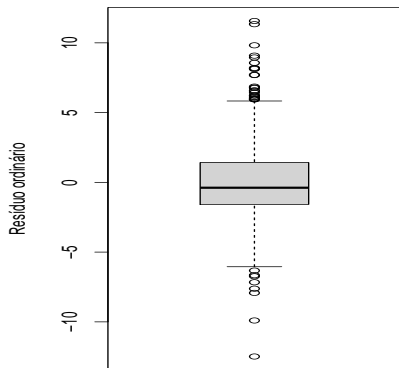
Resultados: modelo 2 - Pearson



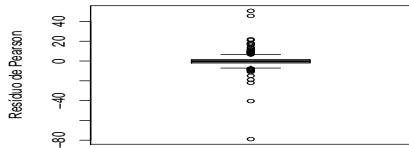
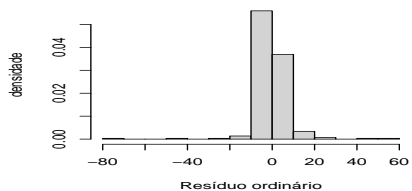
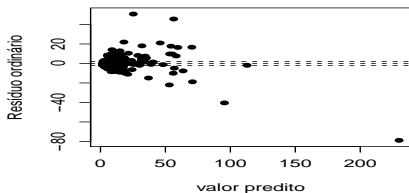
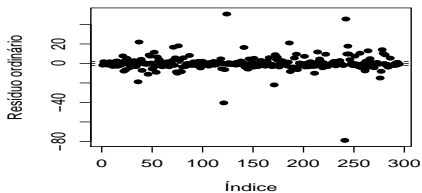
Resultados: modelo 2 - Ordinário



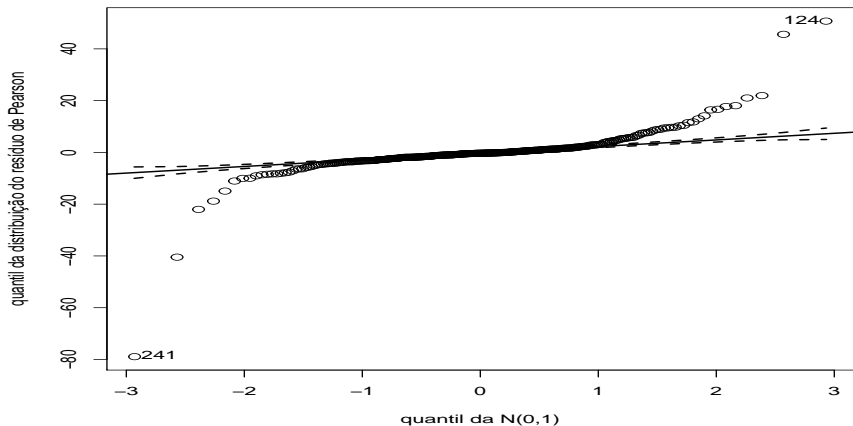
Resultados: modelo 2 - Ordinário



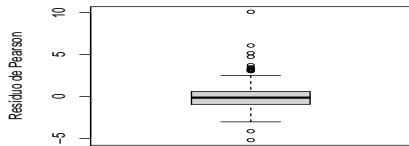
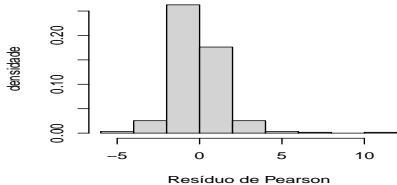
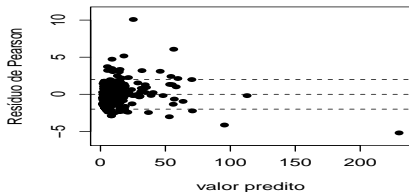
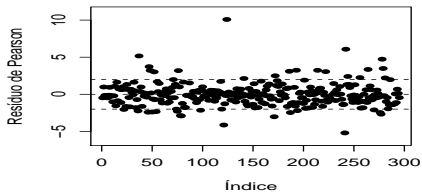
Dados reais: modelo ajustado com QA - Ordinário



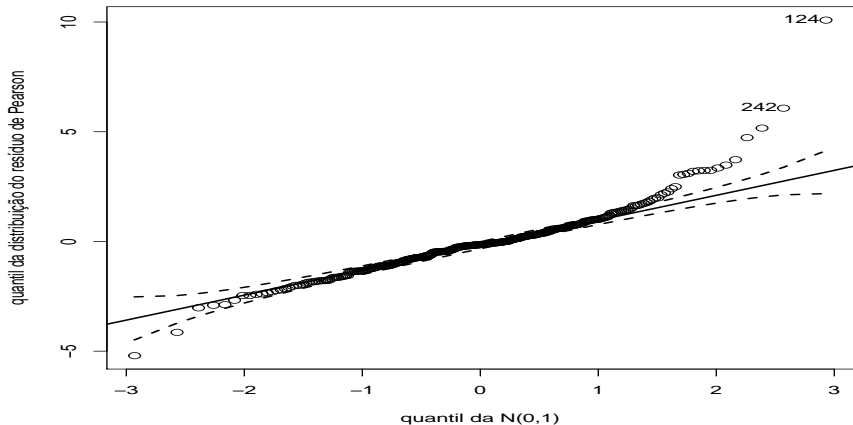
Dados reais: modelo ajustado com QA - Ordinário



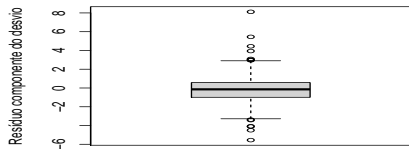
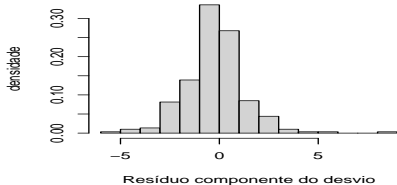
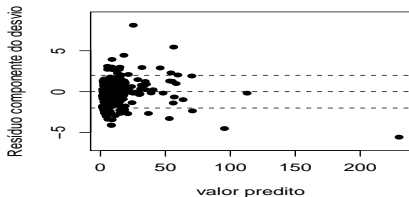
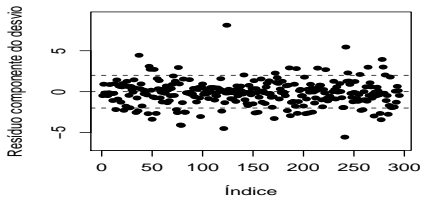
Dados reais: modelo ajustado com QA - Pearson



Dados reais: modelo ajustado com QA - Pearson

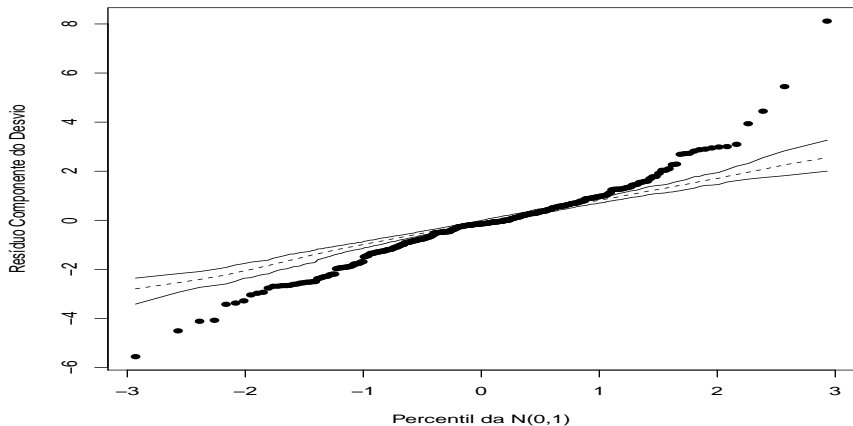


Dados reais: modelo ajustado com QA - RCD

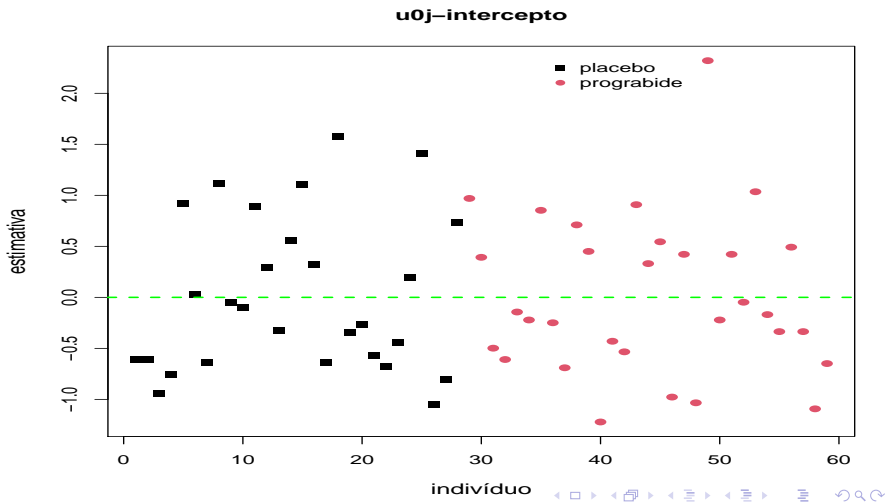


Dados reais: modelo ajustado com QA - RCD

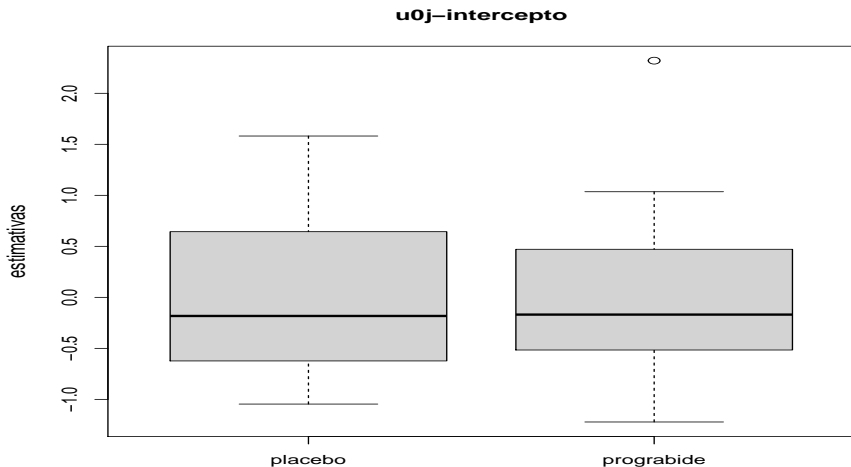
Gráfico de quantil-quantil normal



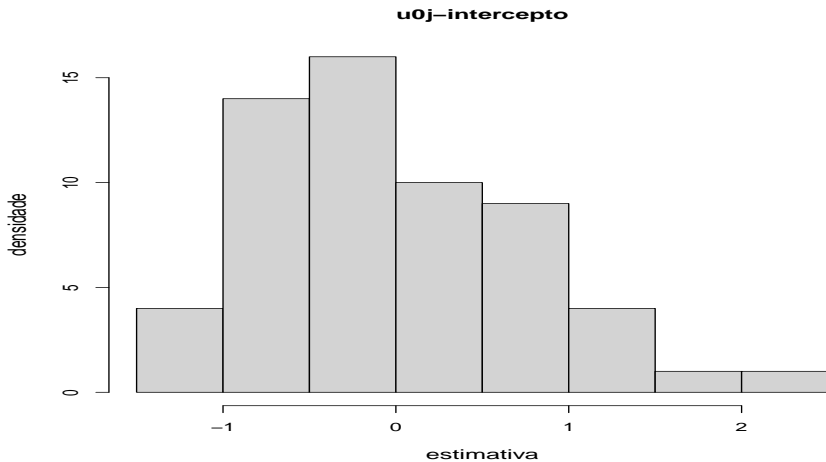
Dados reais: efeitos aleatórios por grupo



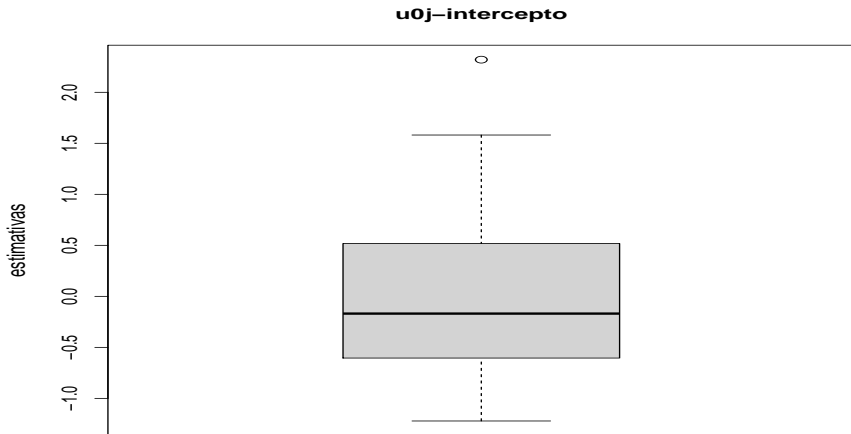
Dados reais: box plot dos efeitos aleatórios por grupo



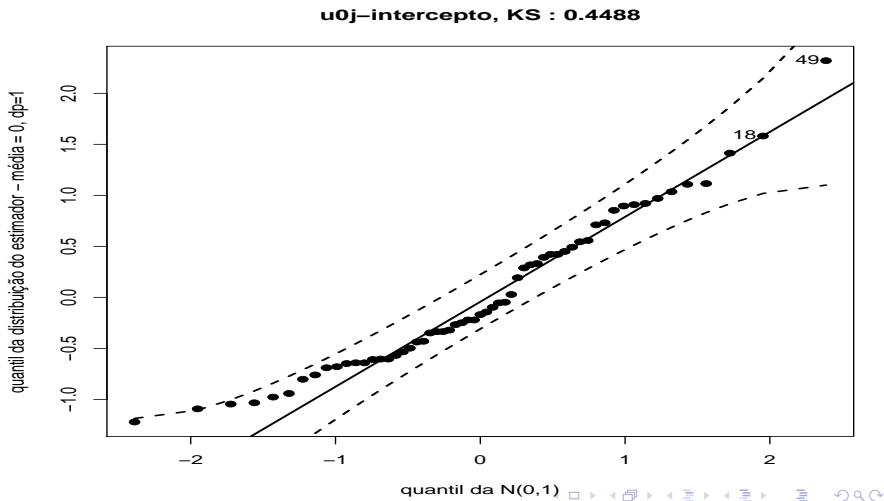
Dados reais: histograma aleatórios por grupo



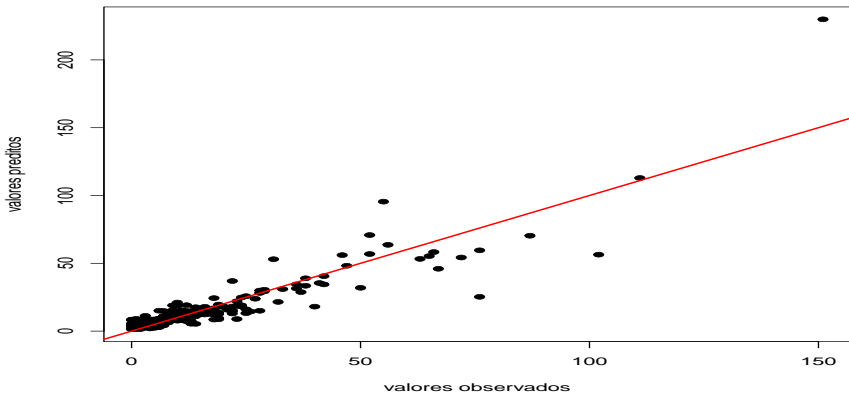
Dados reais: boxplot dos efeitos aleatórios



Dados reais: QQplot com envelopes dos efeitos aleatórios



Dados reais: valores preditos e observados



Comentários

- O modelo log-linear de Poisson não se ajustou bem aos dados.
Possíveis motivos: excesso de zeros e/ou superdispersão.
- Aparentemente, os grupos são equivalente no período 1 e tendem a se diferenciar ao longo do tempo, com um número médio de ataques decrescente para o progabide, e constante, para o placebo.

Comentários

- Os efeitos aleatórios são significativos (diversos valores preditos são diferentes de zero), com aparente fuga da normalidade. Há uma aparente e módica diferença entre os efeitos aleatórios dos dois grupos.
- Alternativas: modelos inflacionados de zeros e/ou binomial negativo (antes de modificar a distribuição dos efeitos aleatórios).