Modelos lineares generalizados mistos

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:

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

em que $\theta \in \Theta \subset \Re$ (espaço paramétrico de θ), $b(\theta), c(y,\phi) : \Re \to \Re$, 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.

Família exponencial bi-paramétrica

- Consideraremos que $Y \sim \text{FE}(\theta, \phi)$ e que temos $Y_i \overset{ind.}{\sim} FE(\theta_i, \phi), i = 1, 2, ..., n$, ou seja

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

Modelo linear generalizado

- $Y_i \sim \text{FE}(\theta_i, \phi), i = 1, 2, ..., n$, $\theta_i = h(\mu_i)$.
- $g(\mu_i) = \eta_i$, $\eta_i = \sum_{j=1}^{p} x_{ji}\beta_j$, $\mathcal{E}(Y_i) = \mu_i = b'(\theta_i)$, x_{ji} : covariável j associada ao indivíduo i (fixa e conhecida) e $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)', \phi$: parâmetros desconhecidos.
- $\mathcal{V}(Y_i) = \phi^{-1} V(\mu_i)$, em que $V(\mu_i) = \dfrac{d\mu_i}{d\theta_i}$.
- $g(.)$ é uma função de ligação (inversível e duplamente diferenciável). Quando $\theta = g(.)$ temos a função de ligação canônica.

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

Desvio (ou função desvio)

- Sem perda de generalidade, seja $l(\boldsymbol{\mu}, \boldsymbol{y}) \equiv l(\boldsymbol{\beta}, \phi)$ a logverossimilhança associada ao modelo em estudo e $l(\boldsymbol{\mu}^0, \boldsymbol{y})$ a logverossimilhança do modelo saturado (n=p), ou seja, em que cada média é representada por ela mesma.
- $l(\boldsymbol{\mu}, \boldsymbol{y}) = \phi\left[\sum_{i=1}^{n} y_i\theta_i - \sum_{i=1}^{n} b(\theta_i)\right] + \sum_{i=1}^{n} c(y_i, \phi)$
- $l(\boldsymbol{\mu}^{(0)}, \boldsymbol{y}) = \phi\left[\sum_{i=1}^{n} y_i\theta_i^{(0)} - \sum_{i=1}^{n} b(\theta_i^{(0)})\right] + \sum_{i=1}^{n} c(y_i, \phi)$ em que $\theta_i^{(0)} = h(\mu_i^{(0)})$.
- Assim, $D^*(\boldsymbol{y}, \boldsymbol{\mu}) = 2(l(\boldsymbol{\mu}^{(0)}, \boldsymbol{y}) - l(\boldsymbol{\mu}, \boldsymbol{y})) = 2\phi\sum_{i=1}^{n}\left[y_i\left(\theta_i^{(0)} - \theta_i\right) + b(\theta_i) - b(\theta_i^{(0)})\right] = \phi D(\boldsymbol{y}, \boldsymbol{\mu})$

Desvio (ou função desvio)

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

Desvio: exemplos

- Bernoulli: $D(\boldsymbol{y};\widehat{\boldsymbol{\mu}}) = -2\sum_{i=1}^{n}\left\{\ln(1-\widehat{\mu}_i)I_{\{0\}}(y_i) + \ln(\widehat{\mu}_i)I_{\{1\}}(y_i)\right\}$
- Poisson: $D(\boldsymbol{y},\widehat{\boldsymbol{\mu}}) =$
  $2\sum_{i=1}^{k}\left\{\left[y_i\ln(y_i/\widehat{\mu}_i) - (y_i-\widehat{\mu}_i)\right]I_{\{1,2,\ldots\}}(y_i) + \widehat{\mu}_i\mathbb{1}_{\{0\}}(y_i)\right\}$
- Normal: $D(\boldsymbol{y},\widehat{\boldsymbol{\mu}}) = \sum_{i=1}^{n}(y_i-\widehat{\mu}_i)^2$ e $D^*(\boldsymbol{y},\widehat{\boldsymbol{\mu}}) = \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(y_i-\widehat{\mu}_i)^2$.
- Gama: $D(\boldsymbol{y},\widehat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{n}\left\{-\ln(y_i/\widehat{\mu}_i) + (y_i-\widehat{\mu}_i)/\widehat{\mu}_i\right\}$ e
  $D^*(\boldsymbol{y},\widehat{\boldsymbol{\mu}}) = \phi D(\boldsymbol{y},\widehat{\boldsymbol{\mu}})$.

Desvio: comportamento assintótico

- No caso do modelo de Poisson se $\mu_i \to \infty, i = 1, 2, ..., n$, sob a hipótese de que o modelo é adequado, $D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) \approx \chi^2_{(n-p)}$.
- Para os outros modelos, $D^*(\boldsymbol{y}, \widehat{\boldsymbol{\mu}})$ depende de ϕ. Então, nesses casos, utilizamos uma estimativa consistente de ϕ. Além disso, temos que

$$D^*(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) \xrightarrow[\phi \to \infty]{D} \chi^2_{n-p}$$

- Assim, podemos comparar o valor do desvio com os quantis de uma distribuição qui-quadrado, para verificar a qualidade de ajuste do modelo, quando a variabilidade dos dados for pequena.

Análise do desvio

- Podemos definir um procedimento para testar as hipóteses $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ vs $H_1 : \boldsymbol{\beta}_1 \neq \mathbf{0}$.
- A estatística $Q_{AD} = \dfrac{\left(D(\boldsymbol{y};\widehat{\boldsymbol{\mu}}^{(0)}) - D(\boldsymbol{y};\widehat{\boldsymbol{\mu}})\right)/q}{D(\boldsymbol{y};\widehat{\boldsymbol{\mu}})/(n-p)}$ sob H_0 e para n suficientemente grande, é tal que $Q_{AD} \approx F_{(q,n-p)}$ e q é dimensão do vetor $\boldsymbol{\beta}_1$.

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 $p - valor \approx P(X \geq q_{AD}|H_0)$, e $X \sim F_{(q,n-p)}$

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

Outras funções de ligação

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

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

ou, de modo equivalente, $\mu = \Phi(\eta)$, em que $\Phi(.)$ 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:

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

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

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

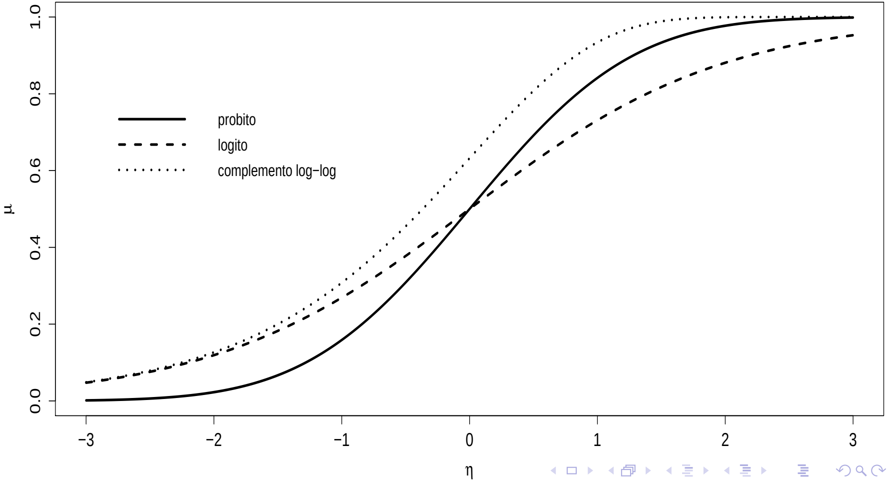
ou de modo equivalente,

$\ln(-\ln(1-\mu)) = \eta.$

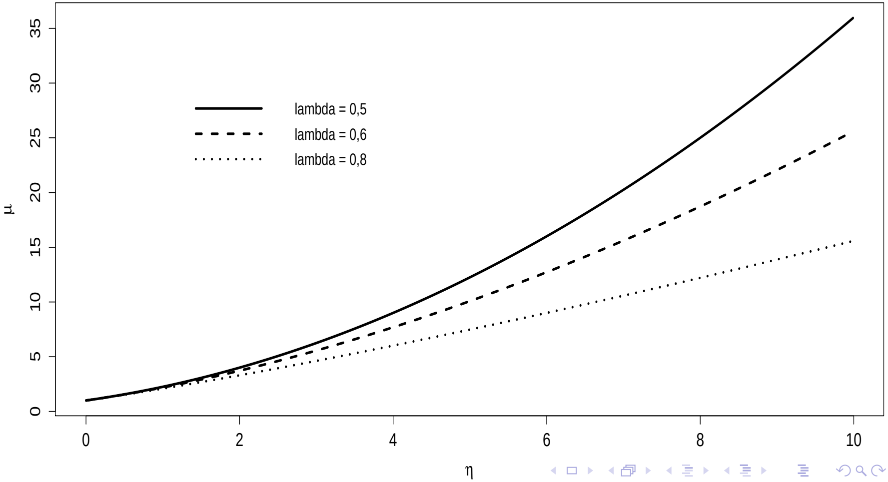
Resumo: principais funções de ligação

Distribuição	$g(\mu)$
Normal	μ,(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 $g(.)$ 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 $\boldsymbol{\beta}$ (usando a regra da cadeia)

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

em que $\boldsymbol{W} = \text{diag}(\omega_1, ..., \omega_n), \omega_i = (d\mu_i/d\eta_i)^2/V_i,\ V_i = V(\mu_i)$ e $\boldsymbol{V} = \text{diag}(V_1, ..., V_n)'$. As outras quantidades são como definidas para os modelos de regressão de Poisson e logístico.

Estimação

- Vetor Escore para ϕ

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

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

Estimação

- Informação de Fisher (usando a regra da cadeia)

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

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

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 $\boldsymbol{S}(\widehat{\boldsymbol{\beta}}) = \mathbf{0}, S(\widehat{\phi}) = 0$ 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 $I(\boldsymbol{\beta}, \phi) = -\mathcal{E}\left(\left\{\sqrt{\frac{\omega_i}{V_i}}(Y_i - \mu_i)x_{ji}\right\}\right) = \mathbf{0}$ e o produto matricial $\boldsymbol{I}^{-1}(\boldsymbol{\beta})\boldsymbol{S}(\boldsymbol{\beta})$ não depende de ϕ, a estimação é feita em duas etapas.

Estimação de $\boldsymbol{\beta}$: Algoritmo escore de Fisher

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

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

até que algum critério de convergência seja satisfeito, como

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

Estimação de ϕ

- Com as estimativas de $\boldsymbol{\beta}$, digamos $\widetilde{\boldsymbol{\beta}}$, obtidas no passo anterior, obtenha as estimativas de ϕ através de $\widehat{\phi} = \frac{n-p}{\sum_{i=1}^{n}\left(\frac{(Y_i-\widehat{\mu}_i)^2}{V(\widehat{\mu}_i)}\right)}$ que é o estimador do método dos momentos (e consistente) de ϕ.
- O R fornece a estimativa associada ao estimador $\widehat{\phi} = \frac{n-p}{D(\boldsymbol{y},\widehat{\boldsymbol{\mu}})}$ que não é consistente para ϕ.
- Para os modelos Poisson e Bernoulli, $\phi = 1$.

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.
- Sob a validade das hipóteses do modelo temos que RCD (padronizado) segue, aproximadamente, uma distribuição N(0,1).
- Detalhes podem ser encontrado em Paula (2013) (link no site do curso) e `http://www.ime.unicamp.br/~cnaber/Material_MLG_1S_2016.htm`.
- Podem ser desenvolvidos testes de hipótese do tipo $\boldsymbol{C}\beta = \boldsymbol{M}$ de modo semelhante ao que foi visto antetiormente.

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

- Diz respeito aos resultados de um ensaio clínico com 59 indivíduos epilépticos os quais foram aleatorizados de modo que cada um recebesse uma droga antiepiléptica denominada progabide ou placebo.
- Os dados de cada indivíduo consistiram de um número inicial de ataques epilépticos 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.

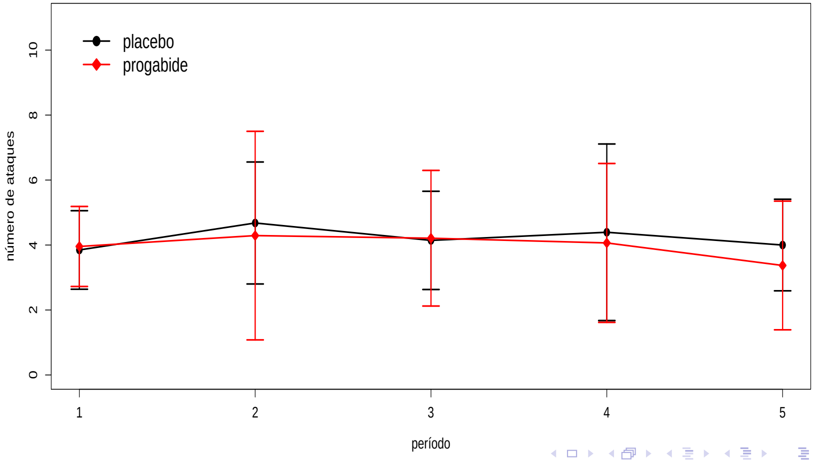
Exemplo 4: Ataques epiléticos

- O interesse da pesquisa é saber se a droga reduz a taxa de ataques epilépticos.
- 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.

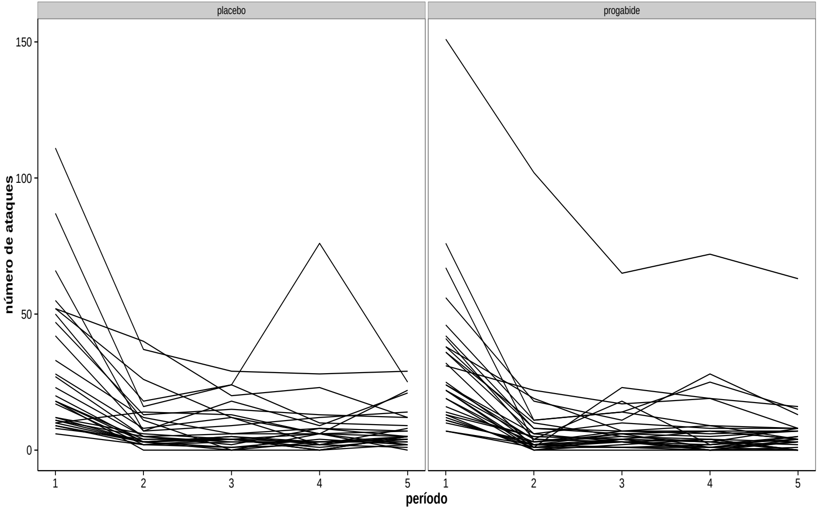
Medidas resumo

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

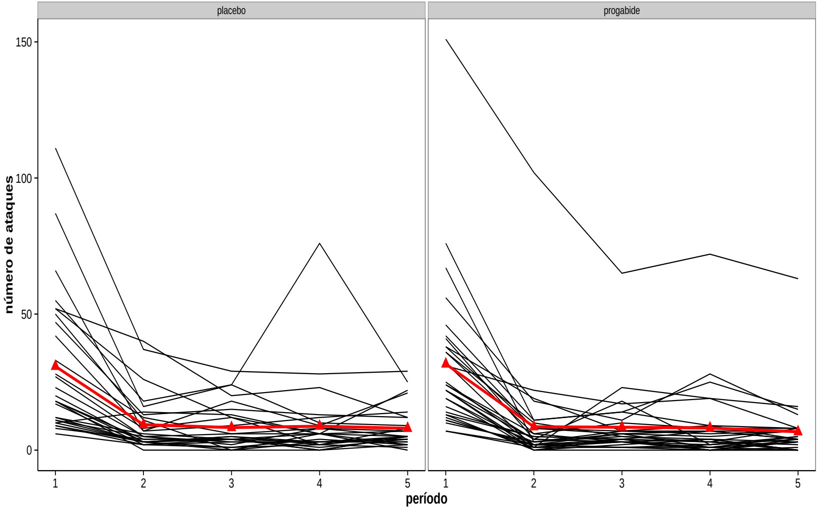
Perfis médios



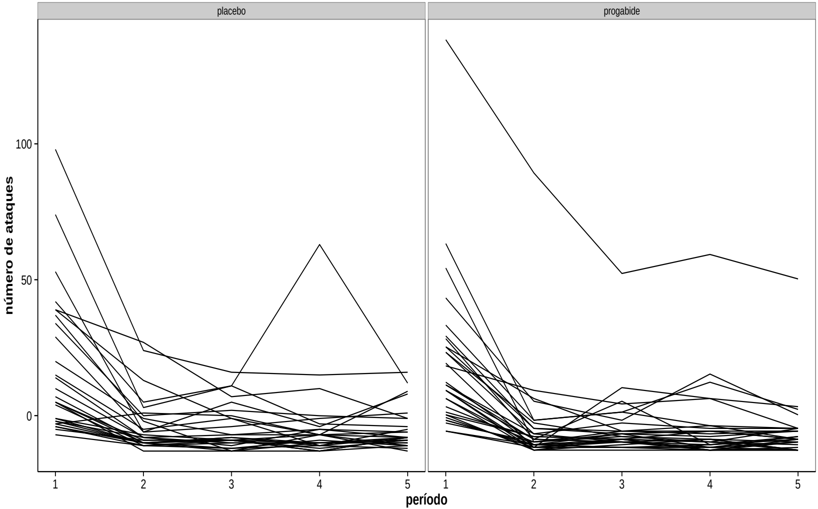
Perfis individuais



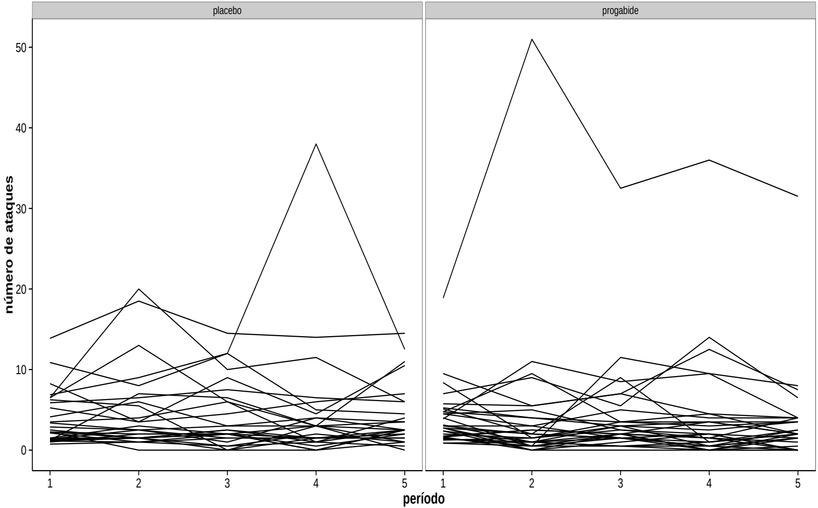
Perfis individuais e médios



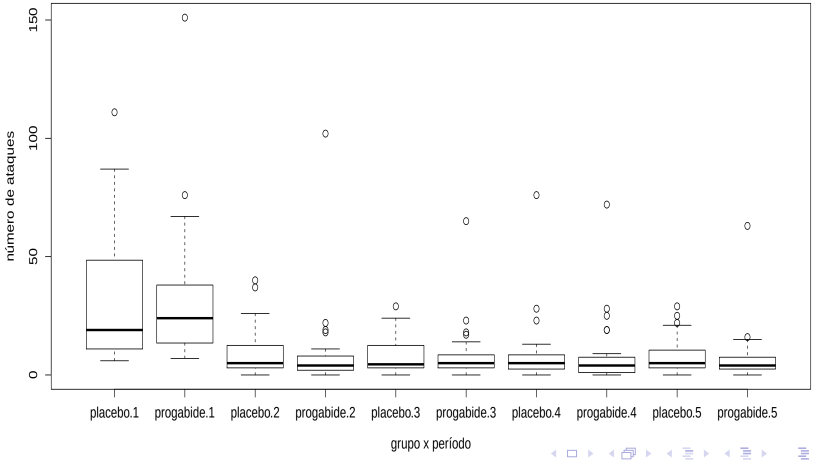
Perfis individuais centrados



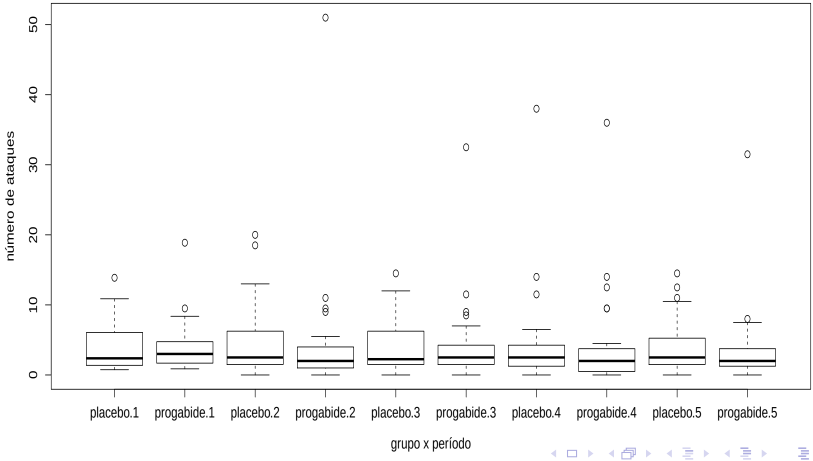
Perfis individuais relativos



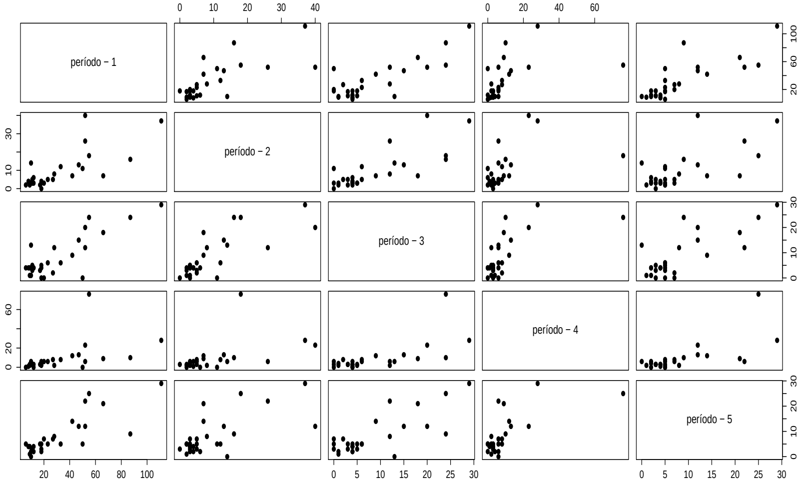
Box plot


número de ataques
0
50
100
150
placebo.1
progabide.1
placebo.2
progabide.2
placebo.3
progabide.3
placebo.4
progabide.4
placebo.5
progabide.5
grupo x período

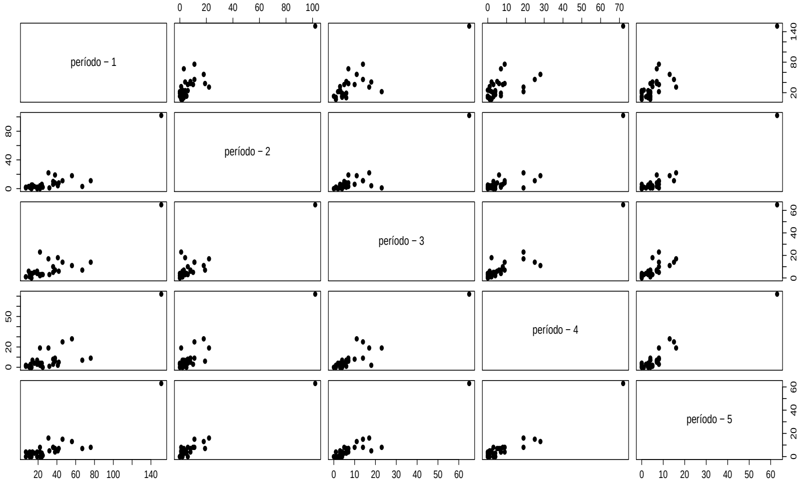
Box plot (ataques/semanas)



Matriz de diagramas de dispersão: placebo



Matriz de diagramas de dispersão: progabide



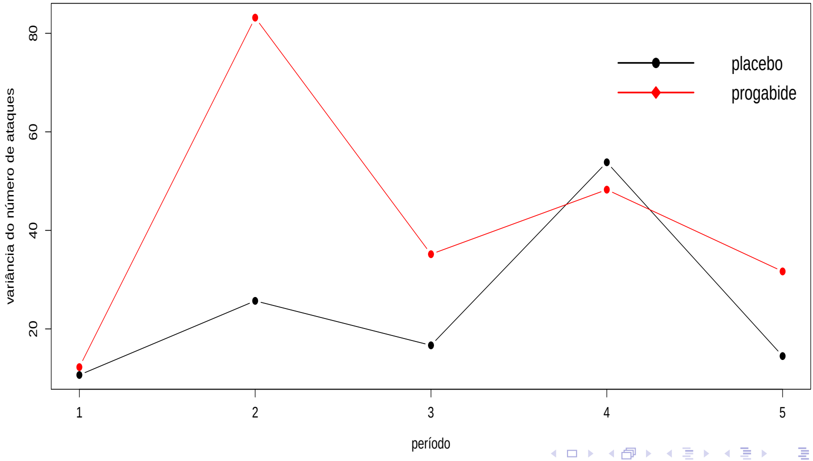
Variâncias (diagonal), correlações (acima) e covariâncias (abaixo): grupo placebo

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

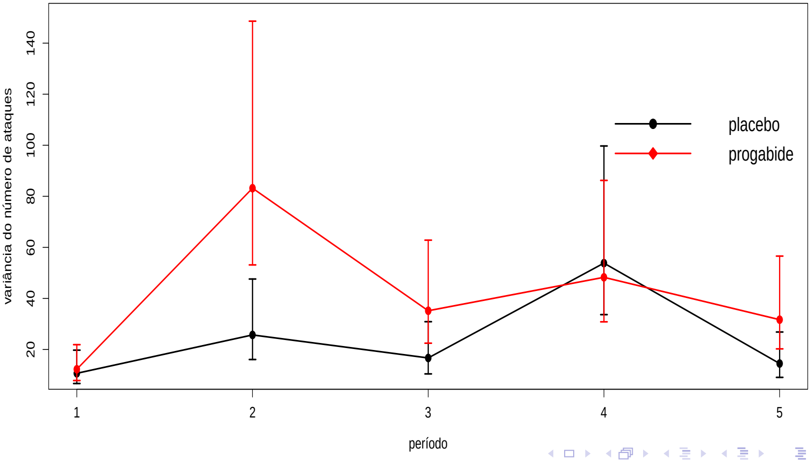
Variâncias (diagonal), correlações (acima) e covariâncias (abaixo): grupo prograbide

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

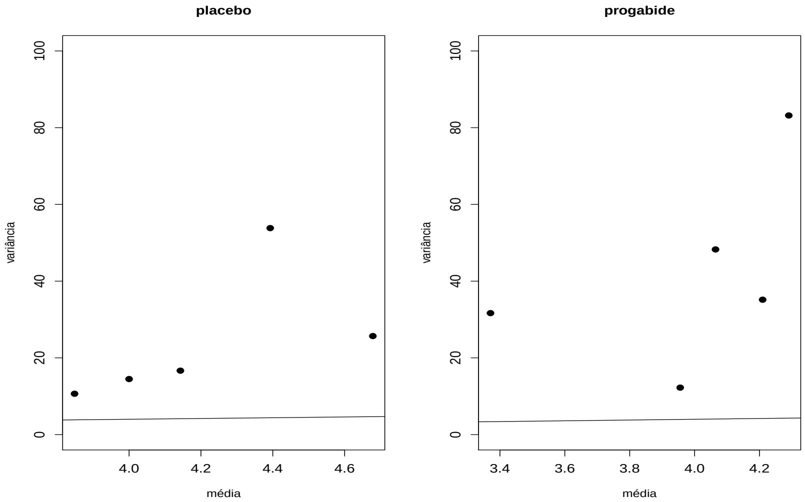
Variâncias em cada condição



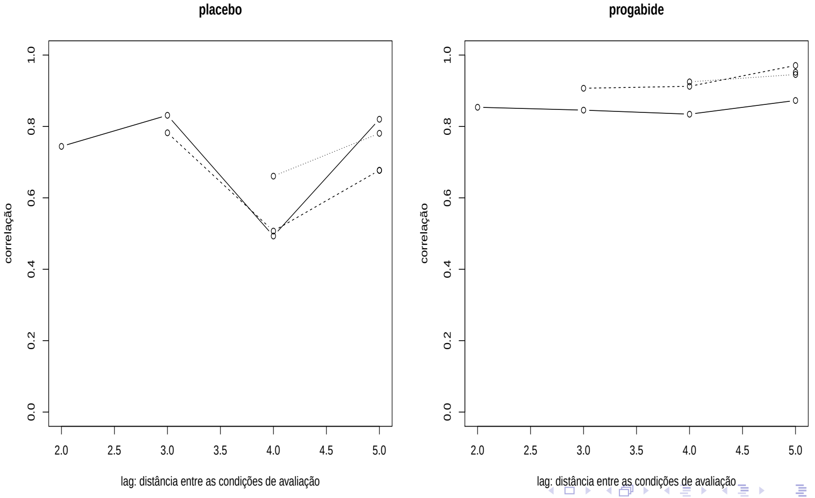
Variâncias em cada condição com intervalos de confiança



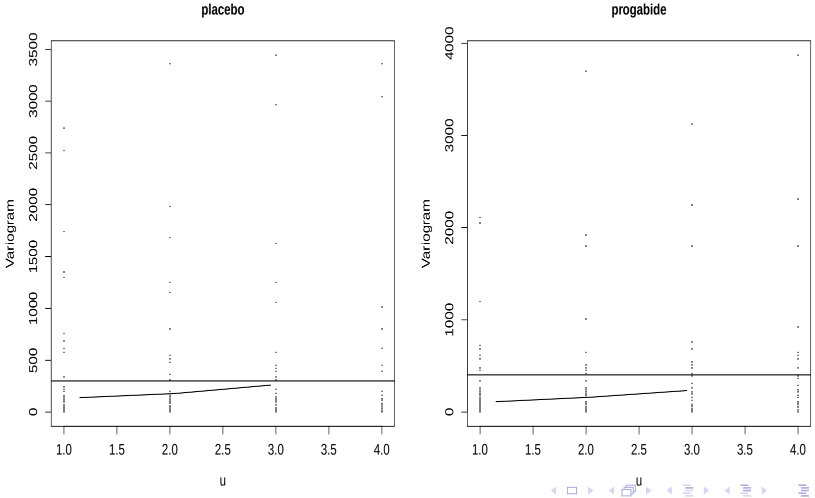
Dispersão entre as médias e as variâncias amostrais



Gráficos dos perfis das linhas da matriz de correlações

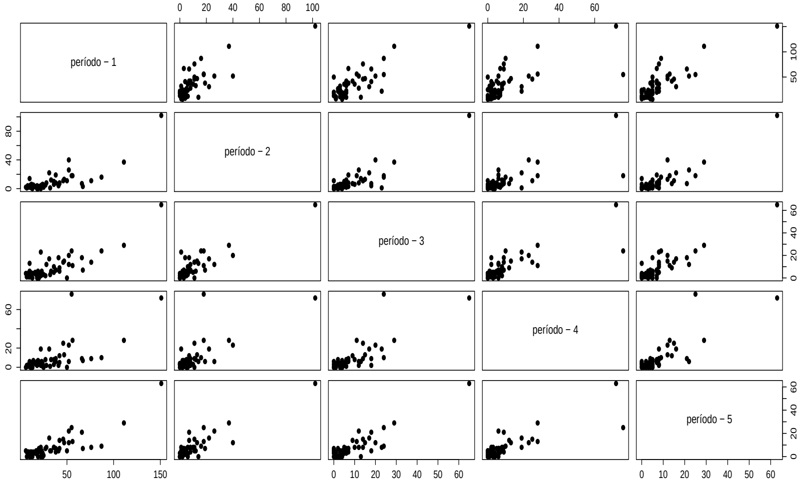


Variograma



placebo
progabide
Variogram
Variogram
0
500
1000
1500
2000
2500
3000
3500
0
1000
2000
3000
4000
1.0
1.5
2.0
2.5
3.0
3.5
4.0
u
u

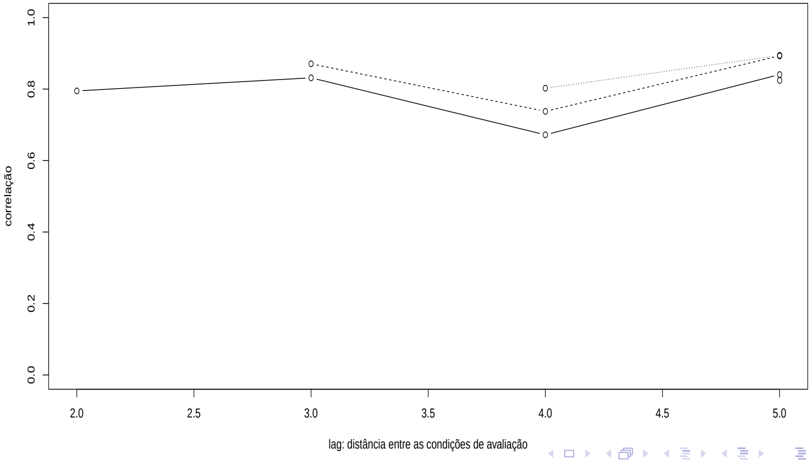
Matriz de diagramas de dispersão



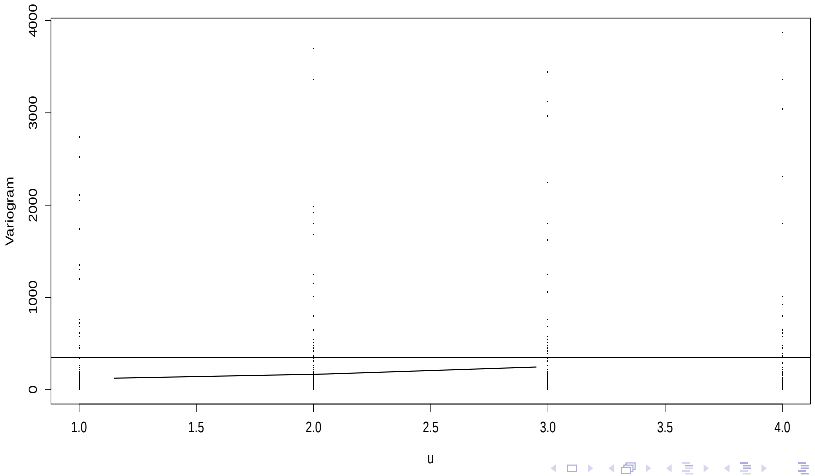
Variâncias (diagonal), correlações (acima) e covariâncias (abaixo)

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:

$$
\begin{aligned}
Y_i \overset{ind.}{\sim} FE(\theta_i, \phi) \quad &, \quad \theta_i = h(\mu_i), i = 1, ..., n \\
g(\mu_i) \quad &= \quad \boldsymbol{X}_i'\boldsymbol{\beta} = \sum_{j=1}^{p} x_{ji}\beta_j ; \boldsymbol{X}_i = (x_{1i}, ..., x_{pi})'
\end{aligned}
$$

Modelo linear generalizado misto (MLGM)

- O correspondente MLGM é dado por:

$$
\begin{aligned}
Y_{ij}|\boldsymbol{b}_j \overset{ind.}{\sim} FE(\theta_{ij},\phi) \quad &, \quad \theta_{ij} = h(\mu_{ij}), j = 1,...,n, i = 1,...,k_j \\
\boldsymbol{b}_j \quad &\overset{i.i.d.}{\sim} \quad N(\boldsymbol{0},\boldsymbol{\Psi}); \mu_{ij} = \mathcal{E}(Y_{ij}|\boldsymbol{b}_j) \\
g(\mu_{ij}) \quad &= \quad \boldsymbol{X}'_{ij}\beta + \boldsymbol{Z}'_{ij}\boldsymbol{b}_j = \sum_{k=1}^{p} x_{kij}\beta_k + \sum_{r=1}^{q} z_{rij}b_{rj};, \\
\boldsymbol{X}_{kij} = (x_{1ij},...,x_{pij})' \quad &; \quad \boldsymbol{Z}_{ij} = (z_{1ij},...,z_{qij})' \\
\boldsymbol{b}_j \quad &= \quad (b_{1j},...,b_{qj})' \\
\boldsymbol{\Psi} \quad &\equiv \quad \boldsymbol{\Psi}(\boldsymbol{\theta})
\end{aligned}
$$

Média e estrutura de dependência

- $\mu_{ij}^{*} = \mathcal{E}(Y_{ij}) = \mathcal{E}(\mathcal{E}(Y_{ij}|\boldsymbol{b}_j)) = \mathcal{E}(\mu_{ij}) = \mathcal{E}(g^{-1}(\boldsymbol{X}_{ij}'\beta + \boldsymbol{Z}_{ij}\boldsymbol{b}_j))$
- Variância

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

em que $V(.)$ é a função de variância (veja slide 4).

- Covariância

$$\begin{aligned} Cov(Y_{ij}, Y_{i'j}) &= Cov(\mathcal{E}(Y_{ij}|\boldsymbol{b}_j), \mathcal{E}(Y_{i'j}|\boldsymbol{b}_j)) + E(\underbrace{Cov(Y_{ij}, Y_{i'j}|\boldsymbol{b}_j)}_{0}) \\ &= Cov(g^{-1}(\boldsymbol{X}_{ij}'\beta + \boldsymbol{Z}_{ij}\boldsymbol{b}_j), g^{-1}(\boldsymbol{X}_{i'j}'\beta + \boldsymbol{Z}_{i'j}\boldsymbol{b}_j)) \end{aligned}$$

Um possível modelo

$$Y_{ijk}|b_j \overset{ind.}{\sim} \text{Poisson}(\mu_{ijk}/t_k), i = 1, 2(\text{grupo}, 1 - \textit{placebo}, 2 - \textit{progabide}),$$

$$j = 1, ..., n_i(\text{indivíduo}, n_1 = 28; n_2 = 31), k = 1, 2, 3, 4, 5(\text{período})$$

$$\ln(\mu_{ijk}/t_k) = \mu + \alpha_i + (\beta_1 + \delta_i)(x_{ijk} - x) + b_j, \alpha_1 = 0, \delta_1 = 0$$

$$\ln(\mu_{ijk}) = \ln(t_k) + \mu + \alpha_i + (\beta_1 + \delta_i)(x_{ijk} - x) + b_j,$$

$$b_j \overset{i.i.d.}{\sim} N(0, \sigma^2)$$

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

Cont.

- x_{ijk} : número de semanas acumuladas ou período correspondente ao k-ésimo instante de avaliação (x é um valor de referência, podendo ser um período específico ou um número específico de semanas acumuladas).
- $\mu_{ijk} = t_k e^{\mu+\alpha_i+\beta_i(x_{ijk}-x)} e^{b_j}$.
- $\mu_{ijk}^* = t_k e^{\mu+\alpha_i+\beta_i(x_{ijk}-x)} \mathcal{E}(e^{b_j})$.
- $\mu_{ijk}^* | (x_{ijk} = x) = t_k e^{\mu+\alpha_i} \mathcal{E}(e^{b_j})$.
- $\mu_{1jk}^* | (x_{1jk} = x) = t_k e^{\mu} \mathcal{E}(e^{b_j})$.
- $\mu_{ijk}^* | (x_{ijk}^* = x_{ijk} + 1) = t_{k+1} e^{\mu+\alpha_i+\beta_i(x_{ijk}-x)} e^{\beta_i} \mathcal{E}(e^{b_j})$.

Cont.

- $\mu^*_{1jk} = t_k e^{\mu + \beta_1 (x_{1jk} - x)} \mathcal{E}(e^{b_j})$ (grupo placebo).
- $\mu^*_{2jk} = t_k e^{\mu + \alpha_2 + (\beta_1 + \delta_2)(x_{2jk} - x)} \mathcal{E}(e^{b_j})$ (grupo progabide).
- $\dfrac{\mu^*_{2jk}}{\mu^*_{1jk}} = e^{\alpha_2 + \delta_2 (a - x)}$ (comparando os grupos) para $x_{1jk} = x_{2jk} = a$.
- $\dfrac{\mu^*_{2jk}}{\mu^*_{1jk}} = e^{\alpha_2}$ (comparando os grupos no período 1).
- $\dfrac{\mu^*_{ijk} | (x^*_{ijk} = x_{ijk} + 1)}{\mu^*_{ijk} | x_{ijk}} = \dfrac{t_{k+1}}{t_k} e^{\beta_i}$ (comparando dois períodos subsequentes), em que $\beta_i = \beta_1 + \delta_i$.
- $\dfrac{\mu^*_{ijk} | (x^*_{ijk} = x_{ijk} + r)}{\mu^*_{ijk} | x_{ijk}} = \dfrac{t_{k+r}}{t_k} e^{r\beta_i}$ (comparando dois períodos quaisquer).

Cont.

- Se $b_j \sim N(0,\sigma^2)$, então $e^{b_j} \sim$ log-normal$(0,\sigma^2)$. Neste caso,

$$
\begin{aligned}
\mathcal{E}(Y_{ijk}) &= \mathcal{E}(\mathcal{E}(Y_{ijk}|b_j)) = t_k e^{\mu+\alpha_i+\beta_i(x_{ijk}-x)}\mathcal{E}(e^{b_j}) \\
&= t_k e^{\mu+\alpha_i+\beta_i(x_{ijk}-x)} e^{\sigma^2/2}
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{V}(Y_{ijk}) &= \mathcal{V}(\mathcal{E}(Y_{ijk}|b_j)) + \mathcal{E}(\mathcal{V}(Y_{ijk}|b_j)) = t_k^2 e^{2(\mu+\alpha_i+\beta_i(x_{ijk}-x))}\mathcal{V}(e^{b_j}) \\
&+ t_k e^{\mu+\alpha_i+\beta_i(x_{ijk}-x)}\mathcal{E}(e^{b_j}) \\
&= t_k^2 e^{2(\mu+\alpha_i+\beta_i(x_{ijk}-x))}(e^{\sigma^2}-1)e^{\sigma^2} + t_k e^{\mu+\alpha_i+\beta_i(x_{ijk}-x)} e^{\sigma^2/2} \\
&> t_k e^{\mu+\alpha_i+\beta_i(x_{ijk}-x)} e^{\sigma^2/2} = \mathcal{E}(Y_{ijk})
\end{aligned}
$$

Estimação

- Verossimilhança (completa) $\boldsymbol{b} = (\boldsymbol{b}_1, ..., \boldsymbol{b}_n)'$.

$$
\begin{aligned}
L(\boldsymbol{\beta}, \phi, \boldsymbol{b}, \boldsymbol{\theta}) &= \exp\left\{\phi\left[\sum_{j=1}^{n}\sum_{i=1}^{k_j} y_{ij}\theta_{ij} - \sum_{j=1}^{n} b(\theta_{ij})\right] + \sum_{j=1}^{n}\sum_{i=1}^{k_j} c(y_{ij}, \phi)\right\} \\
&\times \quad exp\left\{-0,5\sum_{j=1}^{n} \boldsymbol{b}_j'\boldsymbol{\Psi}^{-1}\boldsymbol{b}_j\right\} |\boldsymbol{\Psi}|^{-n/2}
\end{aligned}
$$

Verossimilhança marginal

$$
\begin{aligned}
L(\boldsymbol{\beta}, \phi, \boldsymbol{\theta}) &= \prod_{j=1}^{n} \int_{\Re^q} \exp\left\{ \phi \left[\sum_{i=1}^{k_j} y_{ij}\theta_{ij} - b(\theta_{ij}) \right] + \sum_{i=1}^{k_j} c(y_{ij}, \phi) \right\} \\
&\times \quad exp\left\{ -0,5 \boldsymbol{b}_j' \boldsymbol{\Psi}^{-1} \boldsymbol{b}_j \right\} |\boldsymbol{\Psi}|^{-1/2} d\boldsymbol{b}
\end{aligned}
$$

A integral acima não tem solução explícita (lembrando que θ_{ij} é função de $\boldsymbol{b}_j$).

Resolução da integral

- Quadratura (gaussiana).
- Quadratura adaptativa.
- Aproximação de Laplace.
- Monte carlo.

Resolução da integral

- Quadratura (gaussiana), Quadratura adaptativa e Monte Carlo - aproximar a integral por somas de áreas de retângulos:

  `http://www.ime.unicamp.br/~cnaber/aula_IN.pdf`

  e

  `http://www.ime.unicamp.br/~cnaber/aula_IN_2.pdf`

Verossimilhança marginal numericamente aproximada

$$
\begin{aligned}
L(\boldsymbol{\beta},\phi,\boldsymbol{\theta}) &\approx \prod_{j=1}^{n}\sum_{r\in\mathcal{I}} \exp\left\{\phi\left[\sum_{i=1}^{k_j} y_{ij}\theta_{ijr} - b(\theta_{ijr})\right] + \sum_{i=1}^{k_j} c(y_{ij},\phi)\right\} \\
&\times \quad exp\left\{-0,5\boldsymbol{b}_r'\boldsymbol{\Psi}^{-1}\boldsymbol{b}_r\right\}|\boldsymbol{\Psi}|^{-1/2}A_r
\end{aligned}
$$

em que $\mathcal{I}$ é o conjunto de todas as k-uplas, A_r são os chamados pesos de quadratura e $\boldsymbol{b}_r$ o vetor de pesos de quadraturas.

Verossimilhança marginal numericamente aproximada: um único efeito aleatório

$$
\begin{aligned}
L(\boldsymbol{\beta}, \phi, \boldsymbol{\theta}) &\approx \prod_{j=1}^{n} \sum_{r=1}^{m} \exp\left\{ \phi \left[\sum_{i=1}^{k_j} y_{ij}\theta_{ijr} - b(\theta_{ijr}) \right] + \sum_{i=1}^{k_j} c(y_{ij}, \phi) \right\} \\
&\times \quad exp\left\{ \frac{-b_r^2}{2\psi} \right\} \psi^{-1/2} A_r
\end{aligned}
$$

Equações de estimações

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

$$
\begin{aligned}
l(\boldsymbol{\beta}, \phi, \boldsymbol{\theta}) &= \sum_{j=1}^{n} \ln \int_{\Re^q} \exp\left\{ \phi \left[\sum_{i=1}^{k_j} y_{ij}\theta_{ij} - b(\theta_{ij}) \right] + \sum_{i=1}^{k_j} c(y_{ij}, \phi) \right\} \\
&\times \quad exp\left\{-0,5\boldsymbol{b}_j'\boldsymbol{\Psi}^{-1}\boldsymbol{b}_j\right\} |\boldsymbol{\Psi}|^{-1/2} d\boldsymbol{b}_j
\end{aligned}
$$

em relação à $\boldsymbol{\beta}$, ϕ e $\boldsymbol{\theta}$.

- As derivadas com relação à $\boldsymbol{\beta}$ e ϕ assemelham-se à (1) e (2), respectivamente. Com relação à $\boldsymbol{\theta}$ depende da estrutura adotada (o que também depende do modelo). Mesmo a equação de estimação com relação à $\boldsymbol{\theta}$ não tem, em geral, solução explícita.

Equações de estimações

- O sistema de equações gerado por

$$\begin{cases} \boldsymbol{S}(\widetilde{\boldsymbol{\beta}}) = \mathbf{0} \\ \boldsymbol{S}(\widetilde{\phi}) = 0 \\ \boldsymbol{S}(\widetilde{\boldsymbol{\theta}}) = \mathbf{0} \end{cases}$$

não tem solução explícita e algum algoritmo de otimização tem de ser usado (Newton-Raphson, Score de Fisher, FBGS, Nelder-Mead etc). Preditores para os efeitos aleatórios podem considerados usando-se a metodologia vista para os modelos lineares mistos. Nesse caso, a distribuição "à posteriori" dos efeitos aleatórios não é obtenível analitacamente e métodos numéricos apropriados têm de ser utilizados.

Aproximação de Laplace

- Trabalha-se com

$$
\begin{aligned}
L(\boldsymbol{\beta}, \phi, \boldsymbol{\theta}) &= \prod_{j=1}^{n} \int_{\Re^q} \exp\left\{ \phi \left[\sum_{i=1}^{k_j} y_{ij}\theta_{ij} - b(\theta_{ij}) \right] + \sum_{i=1}^{k_j} c(y_{ij}, \phi) \right\} \\
&\times \; exp\left\{ -0,5 \boldsymbol{b}_j' \boldsymbol{\Psi}^{-1} \boldsymbol{b}_j \right\} |\boldsymbol{\Psi}|^{-1/2} d\boldsymbol{b}_j \\
&= |\boldsymbol{\Psi}|^{-n/2} \prod_{j=1}^{n} \int_{\Re^q} \exp(Q(\boldsymbol{b}_j)) d\boldsymbol{b}_j = I
\end{aligned}
$$

Aproximação de Laplace

- Em que:

$$
\begin{aligned}
Q(\boldsymbol{b}_j) &= \phi\left[\sum_{i=1}^{k_j} y_{ij}\theta_{ij} - b(\theta_{ij})\right] + \sum_{i=1}^{k_j} c(y_{ij},\phi) \\
&\times -0,5\boldsymbol{b}_j'\boldsymbol{\Psi}^{-1}\boldsymbol{b}_j
\end{aligned}
$$

Aproximação de Laplace

- Seja $\widehat{\boldsymbol{b}}_j$ o valor que maximiza $Q(\boldsymbol{b}_j)$ e considere a aproximação de $Q(\boldsymbol{b}_j)$ por uma expansão de segunda ordem em séries de Taylor:

$$Q(\boldsymbol{b}_j) = Q(\widehat{\boldsymbol{b}}_j) + \frac{1}{2}(\boldsymbol{b}_j - \widehat{\boldsymbol{b}}_j)' Q^{''}(\widehat{\boldsymbol{b}}_j)(\boldsymbol{b}_j - \widehat{\boldsymbol{b}}_j)$$

- Assim, temos a seguinte aproximação

$$\begin{aligned} I &\approx |\boldsymbol{\Psi}|^{-n/2} \prod_{j=1}^{n} exp(Q(\widehat{\boldsymbol{b}}_j)) \int_{\Re^q} exp\left\{\frac{1}{2}(\boldsymbol{b}_j - \widehat{\boldsymbol{b}}_j)' Q^{''}(\widehat{\boldsymbol{b}}_j)(\boldsymbol{b}_j - \widehat{\boldsymbol{b}}_j)\right\} d\boldsymbol{b}_j \\ &= |\boldsymbol{\Psi}|^{-n/2} \prod_{j=1}^{n} ((2\pi)^{q/2} |-Q''(\widehat{\boldsymbol{b}}_j)|^{-1/2} exp(Q(\widehat{\boldsymbol{b}}_j))) \end{aligned} \quad (4)$$

Aproximação de Laplace

- Em que

$$-Q''(\boldsymbol{b}_j) = \boldsymbol{\Psi}^{-1} + \phi^{-1}\sum_{i=1}^{k_j} \boldsymbol{z}_{ij} b''(\mu_{ij}, \boldsymbol{b}_{ij})\boldsymbol{z}'_{ij}, b(\mu_{ij}, \boldsymbol{b}_{ij}) \equiv b(\theta_{ij})$$

- A maximização numérica da logverossimilhança é feita de modo iterativo alternando-se as atualizações de $\boldsymbol{b}_j$ (maximizando-se $Q(\boldsymbol{b}_j)$) e $(\boldsymbol{\beta}, \boldsymbol{\theta})$ (maximizando-se (4)).

Quase-verossimilhança penalizada

- Trabalha-se com uma aproximação das observações (Y_{ij}) usando a média $\mathcal{E}(Y_{ij}|\boldsymbol{b}_j)$ e erros (normalmente distribuídos) com média 0 e variância $\mathcal{V}(Y_{ij}|\boldsymbol{b}_j)$ em torno de estimativas provisórias de $(\boldsymbol{\beta}, \boldsymbol{b}_j)$ digamos $\widetilde{\boldsymbol{\beta}}$ e $\widetilde{\boldsymbol{b}}_j$ e considerando $h(.) = g^{-1}(.)$, ou seja:

$$
\begin{aligned}
Y_{ij} &\approx h(\mu_{ij}, \boldsymbol{b}_j) + \xi_{ij} \approx h(\widetilde{\mu}_{ij}, \widetilde{\boldsymbol{b}}_j) + h'(\widetilde{\mu}_{ij}, \widetilde{\boldsymbol{b}}_j)\boldsymbol{x}'_{ij}(\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}) \\
&+ h'(\widetilde{\mu}_{ij}, \widetilde{\boldsymbol{b}}_j)\boldsymbol{z}'_{ij}(\boldsymbol{b}_j - \widetilde{\boldsymbol{b}}_j) + \xi_{ij} \\
&= \widetilde{\mu}_{ij} + V(\widetilde{\mu}_{ij})\boldsymbol{x}'_{ij}\left(\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}\right) + V(\widetilde{\mu}_{ij})\boldsymbol{z}'_{ij}\left(\boldsymbol{b}_j - \widetilde{\boldsymbol{b}}_j\right) + \xi_{ij}
\end{aligned}
$$

Quase-verossimilhança penalizada

- Matricialmente

$$\boldsymbol{Y}_j = \widetilde{\boldsymbol{\mu}}_j + \widetilde{\boldsymbol{V}}_j \boldsymbol{X}_j \left(\beta - \widetilde{\beta}\right) + \widetilde{\boldsymbol{V}}_j \boldsymbol{Z}_j \left(\boldsymbol{b}_j - \widetilde{\boldsymbol{b}}_j\right) + \boldsymbol{\xi}_j$$

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

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

$$\begin{aligned} \boldsymbol{Y}_j^* &:= \widetilde{\boldsymbol{V}}_j^{-1}(\boldsymbol{Y}_j - \widetilde{\boldsymbol{\mu}}_j) + \boldsymbol{X}_j\widetilde{\beta} + \boldsymbol{Z}_j\widetilde{\boldsymbol{b}}_j \quad (5) \\ &\approx \boldsymbol{X}_j\beta + \boldsymbol{Z}_j\boldsymbol{b}_j + \xi_j^* \quad (6) \end{aligned}$$

Quase-verossimilhança penalizada

- Em que $\boldsymbol{\xi}_j^* = \widetilde{\boldsymbol{V}}_j^{-1}\boldsymbol{\xi}_j$.
- O modelo (6) pode ser visto como um modelo linear misto para os pseudo-dados $\boldsymbol{Y}_j^*$.
- Algoritmo
  - Passo A: Para valores atuais de $\boldsymbol{\beta}$ e $\boldsymbol{\theta}$ calculam-se valores preditos para $\boldsymbol{b}_j$ e os pseudo-dados $\boldsymbol{Y}_j^*$ (Equação (5)).
  - Passo B: Para os pseudo-dados $\boldsymbol{Y}_j^*$ ajusta-se o modelo (6) e as estimativas para $\boldsymbol{\beta}$ e $\boldsymbol{\theta}$ são atualizadas.

Comentários

- Para qualquer um dos métodos anteriores de estimação, em geral, a convergência em distribuição para normal multivariada dos estimadores de $(\boldsymbol{\beta}, \phi, \boldsymbol{\theta})$, 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.

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 medidas repetidas (k_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).

Outros métodos de estimação

- Algoritmo EM.
- Algoritmo MCEM (Monte Carlo EM).
- Algoritmo SAEM (Stochastic approximation EM).
- Algoritmo SEM (Stochastic EM).

Resultados

Laplace

Parâmetro	Estimativa	Erro-padrão	Estat. z	p-valor
μ	1,07	0,15	7,03	$<$ 0,0001
α_2	-0,03	0,21	-0,13	0,8995
β_1	0,02	0,02	1,23	0,2185
δ	-0,04	0,02	-1,75	0,0804

$\sigma^2 = 0,61$

Resultados

Quadratura adaptativa

Parâmetro	Estimativa	Erro-padrão	Estat. z	p-valor
μ	1,07	0,15	7,03	$<$ 0,0001
α_2	-0,03	0,21	-0,13	0,8995
β_1	0,02	0,02	1,23	0,2185
δ	-0,04	0,02	-1,75	0,0804

$\sigma^2 = 0,61$

Resultados

QVP

Parâmetro	Estimativa	Erro-padrão	Estat. z	p-valor
μ	1,11	0,15	7,20	0,0000
α_2	-0,02	0,21	-0,11	0,9121
β_1	0,02	0,03	0,72	0,4715
δ	-0,04	0,04	-1,02	0,3065

$\sigma^2 = 0,57$

Tipos de resíduos em modelos lineares generelizados (mistos)

- Resíduo ordinário ("response residual")

$$R_{ij}^{R} = Y_{ij} - \widehat{\mu}_{ij}$$

μ_{ij} : média condicional.

- Resíduo de Pearson ("Pearson residual")

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

$V(\mu_{ij})$: função de variância.

Tipos de resíduos em modelos lineares generelizados (mistos)

- Resíduo componente do desvio (RCD) ("deviance residual")

$$R_{ij}^{D} = \left(\text{sinal}(Y_{ij} - \widehat{\mu}_{ij})\sqrt{\widehat{d}_{ij}}\right)$$

em que d_{ij} 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 $Y \sim Poisson(\lambda)$, então $Z = \frac{Y-\lambda}{\sqrt{\lambda}} \xrightarrow[\lambda \to \infty]{D} N(0,1)$.
- 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 e/ou $\phi \to \infty$).

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 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 ($\widetilde{\beta}$) e os valores preditos ($\widetilde{\boldsymbol{b}}_j$) e calcule o RCD para cada observação, $(t_{D_{ij}}), j = 1, 2, ..., n, i = 1, ..., k_j$.

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. $FE(\widetilde{\theta}_{ij}, \widetilde{\phi})$, com $\widetilde{\theta}_{ij} = h(g^{-1}(\boldsymbol{X}'_{ij}\widetilde{\boldsymbol{\beta}} + \boldsymbol{Z}_{ij}\widetilde{\boldsymbol{b}}_j))$.
   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,...,n,, i=1,..,k_j, (amostra), r=1,...,m (réplica).

$$\boldsymbol{T}_1 = \begin{bmatrix} t^*_{D_{111}} & t^*_{D_{112}} & \ldots & t^*_{D_{11m}} \\ t^*_{D_{121}} & t^*_{D_{212}} & \ldots & t^*_{D_{21m}} \\ \vdots & \vdots & \ddots & \vdots \\ t^*_{D_{k_n n1}} & t^*_{D_{k_n n2}} & \ldots & t^*_{D_{k_n nm}} \end{bmatrix}$$

Cont.

4) Dentro de cada amostra, ordena-se, de modo crescente, os RCD's, obtendo-se $t^*_{D_{(ij)r}}$ (estatísticas de ordem):

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

5) Obtem-se os limites $t^*_{(ij)I} = \min\limits_{1 \leq r \leq m} t^*_{D_{(ij)r}}$ e $t^*_{(ij)S} = \max\limits_{1 \leq r \leq m} t^*_{D_{(ij)r}}$, $r = 1, 2, ..., m$.

Cont.

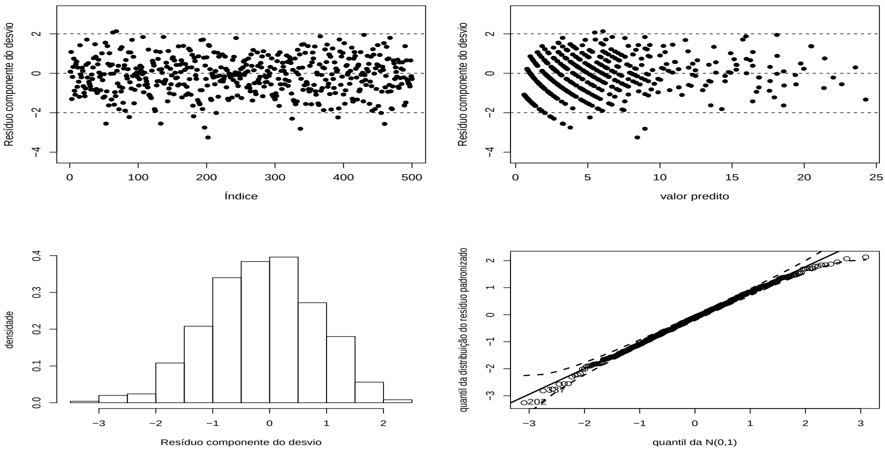
5) Na prática considera-se $t^*_{(ij)I} = \dfrac{t^*_{D_{(ij)(2)}} + t^*_{D_{(ij)(3)}}}{2}$ e $t^*_{(ij)S} = \dfrac{t^*_{D_{(ij)(m-2)}} + t^*_{D_{ij(m-1)}}}{2}$ (refinamento das estimativas dos limites do envelope), em que $t^*_{D_{(ij)(r)}}$ é a r-ésima estatística de ordem dentro de cada linha, $j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

- Além disso, consideramos como a linha de referência $t^*_{(ij)} = \frac{1}{m}\sum_{r=1}^{m} t^*_{D_{(ij)_r}}, j = 1, 2, ..., n; i = 1, ..., k_j$.

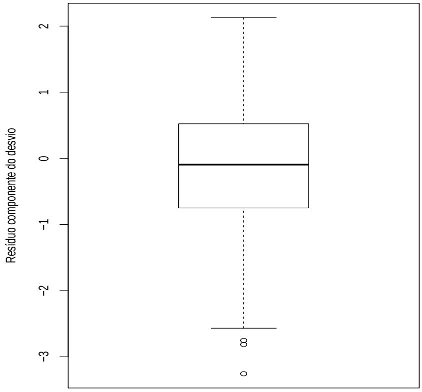
Estudo de simulação

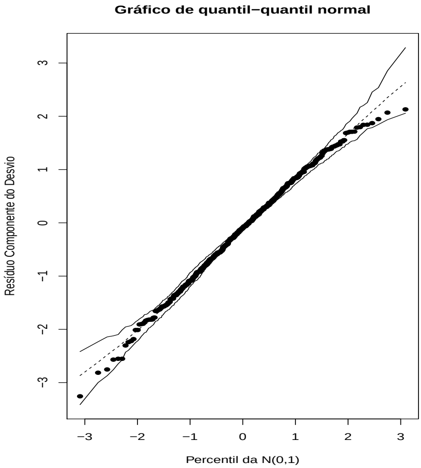
- Simulou-se dois modelos.
- Modelo 1: $Y_{ij}|b_j \overset{ind.}{\sim} Poisson(\mu_i), i = 1, 2, .., 5; j = 1, 2, ..., 100,$
  $\ln \mu_i = 1 + 0,1x_i + b_j; x_i = i;\ b_j \overset{i.i.d.}{\sim} N(0; 0,5).$
- Modelo 2:
  $Y_{ij}|b_j \overset{ind.}{\sim}$ Binomial-negativa$(\mu_i, \phi), i = 1, 2, ..,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

Resultados: modelo 1 - RCD

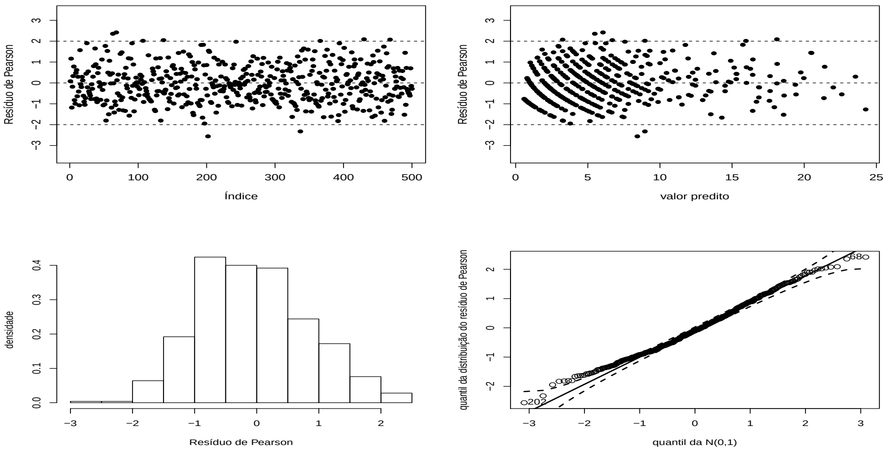


Resultados: modelo 1 - RCD

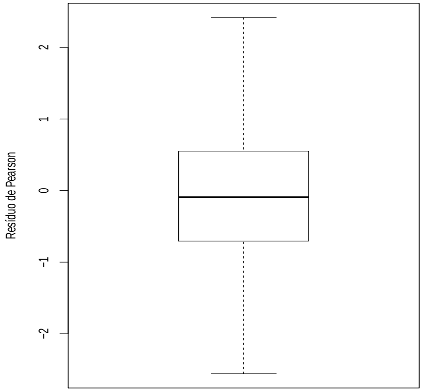


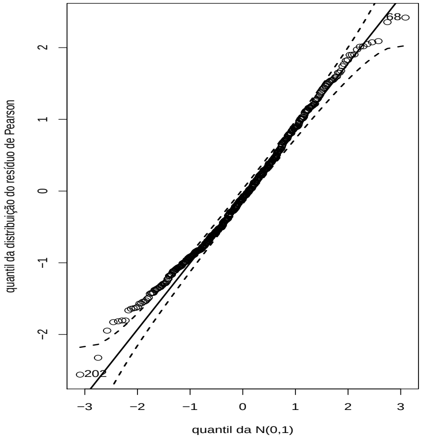


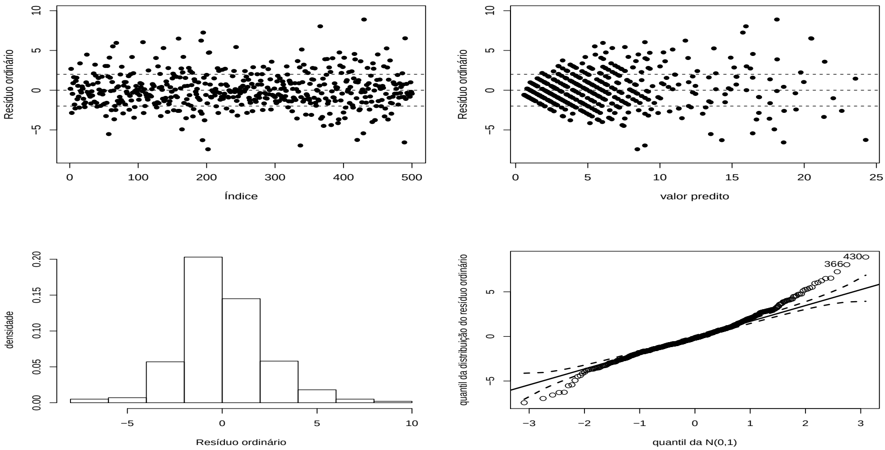
Resultados: modelo 1 - Pearson



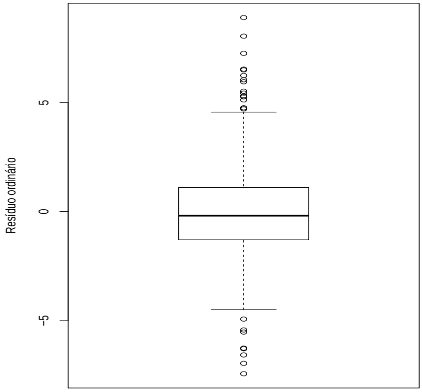
Resultados: modelo 1 - Pearson

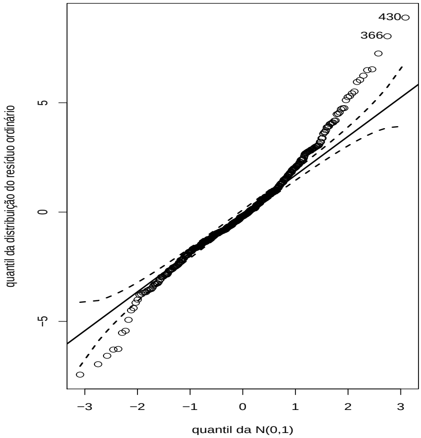




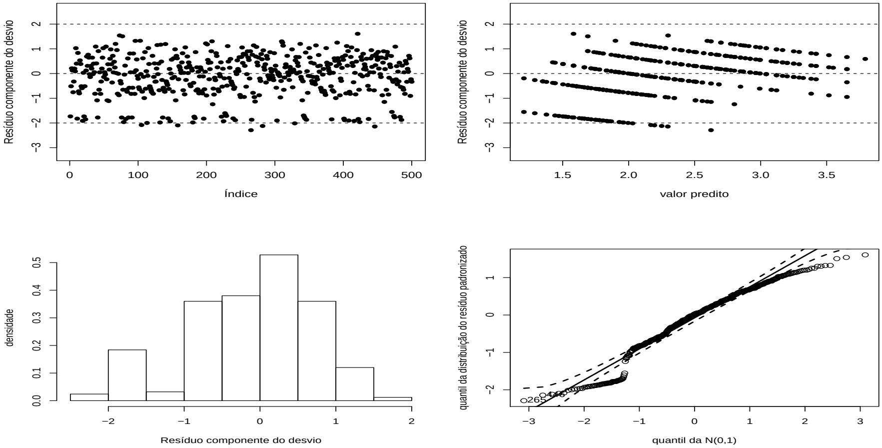

Resíduo ordinário
Índice
Resíduo ordinário
valor predito
densidade
Resíduo ordinário
quantil da distribuição do resíduo ordinário
quantil da N(0,1)
430
366

Resultados: modelo 1 - Ordinário

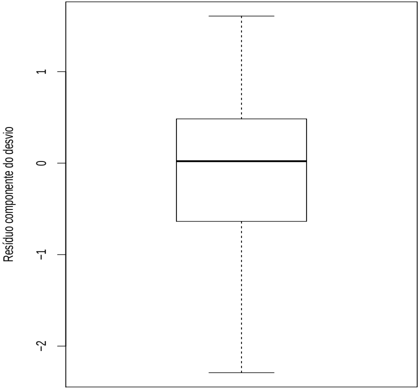


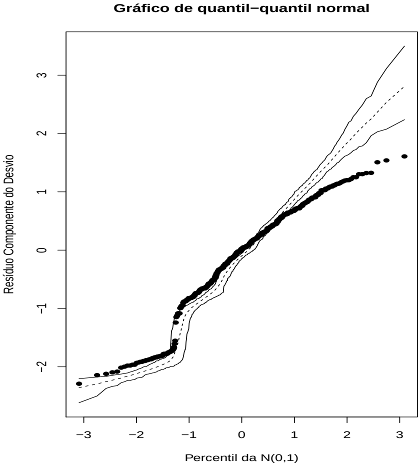


Resultados: modelo 2 - RCD

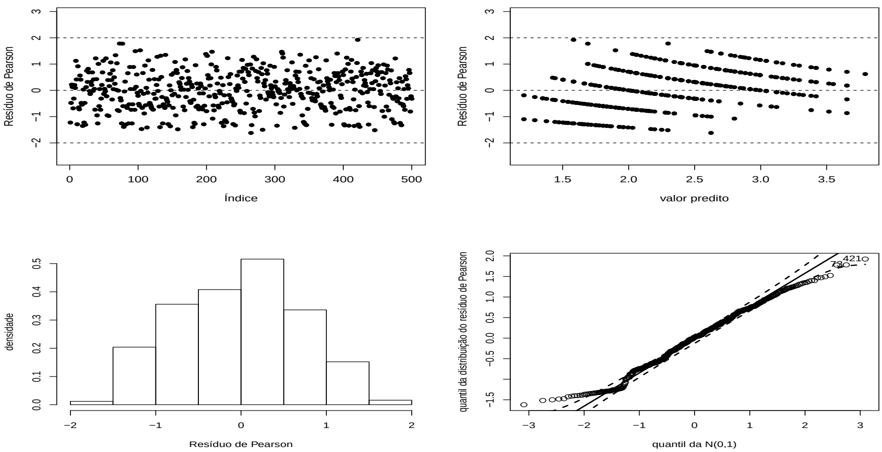


Resultados: modelo 2 - RCD

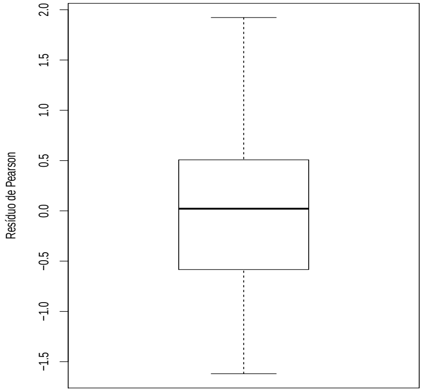


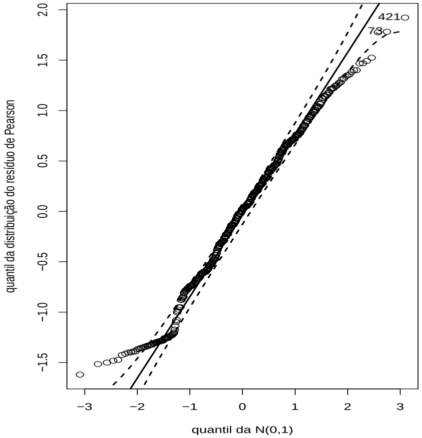


Resultados: modelo 2 - Pearson

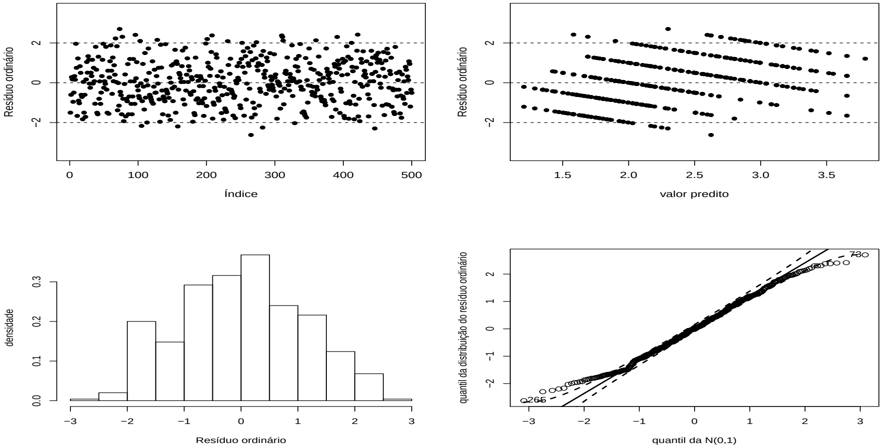


Resultados: modelo 2 - Pearson

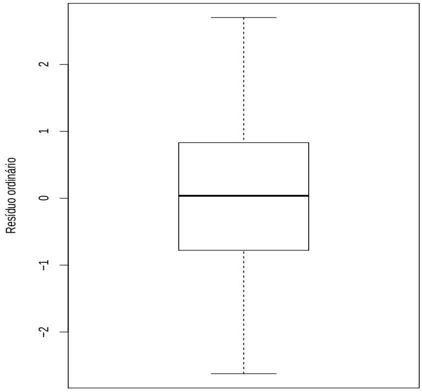


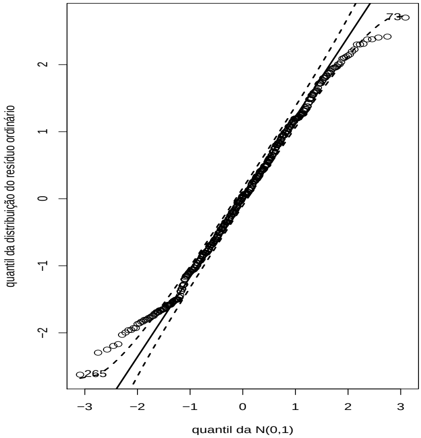


Resultados: modelo 2 - Ordinário

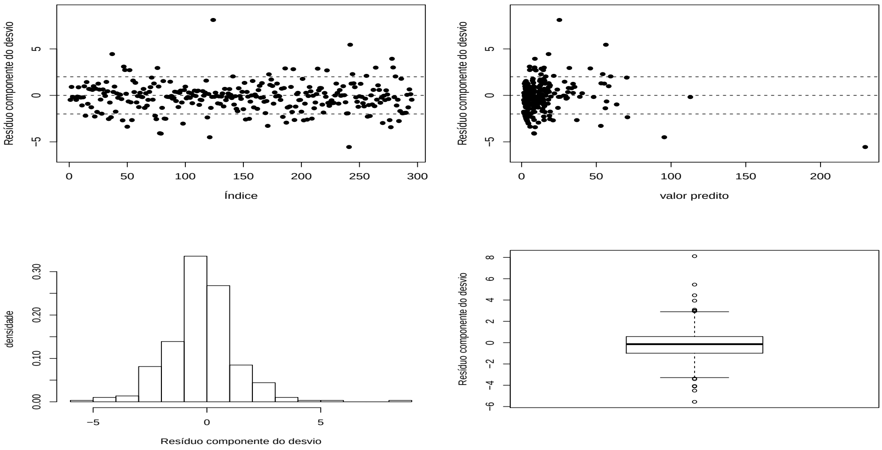


Resultados: modelo 2 - Ordinário

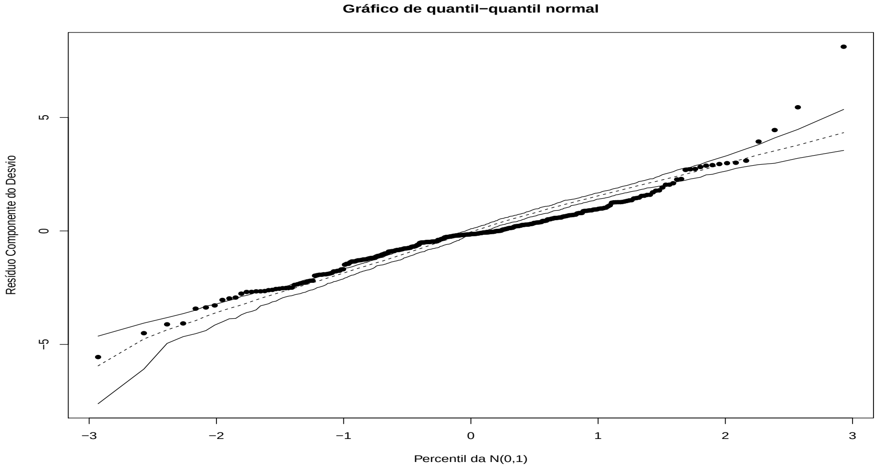




Dados reais: modelo ajustado com QA - RCD



Dados reais: modelo ajustado com QA - RCD



Comentários

- O modelo log-linear de Poisson não se ajustou bem aos dados.
- Possíveis motivos: excesso de zeros e/ou superdispersão.
- Alternativas: modelos inflacionados de zeros e/ou binomial negativo.