

Inferência Bayesiana para Modelos Mistos

Prof. Caio Azevedo

Concentração de bilirrubina em recém-nascidos saudáveis

- Vamos ajustar um modelo gama log-linear (com regressão segmentada).

$$Y_{ij}|b_j \stackrel{ind.}{\sim} \text{gama}(\mu_{ij}, \phi)$$

$$\begin{aligned} \log(\mu_{ij}) &= \beta_0 + \beta_1(x_{ij} - 1)I_{(i \in \{1,2\})} \\ &+ \beta_2(x_{ij} - 1)I_{(i \in \{3,4,5,6,7,8,9\})} + b_j, j = 1, 2, \dots, 89, (\text{indivíduo}) \\ &, i = 1, \dots, 9 (\text{dia (condição de avaliação)}) \end{aligned}$$

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

$$\text{em que } \mathcal{E}(Y_{ij}|b_j) = \mu_{ij}, \mathcal{V}(Y_{ij}|b_j) = \frac{\mu_{ij}^2}{\phi} \text{ e}$$

Concentração de bilirrubina em recém-nascidos saudáveis

- (densidade)

$$p(y_{ij}|\beta, \phi, b_j) = \frac{1}{\Gamma(\phi)} \left(\frac{\phi y_{ij}}{\mu_{ij}} \right)^\phi e^{(-\phi y_{ij}/\mu_{ij})} \mathbb{1}_{(0,\infty)}(y_{ij})$$

- Prioris: $\beta_i \stackrel{ind.}{\sim} N(0, 1000)$, $\phi \sim \text{gama}(0, 01; 0, 01)$,
 $\sigma^2 \sim \text{gama}(0, 01; 0, 01)$, $\mathcal{E}(\phi) = \mathcal{E}(\sigma^2) = 1$, $\mathcal{V}(\phi) = \mathcal{V}(\sigma^2) = 100$
(prioris vagas).
- Implementado no OpenBUGS via R através do pacote R2OpenBUGS.

Concentração de bilirrubina em recém-nascidos saudáveis

- Interpretação dos parâmetros. Analogamento ao modelo log-linear de Poisson misto (veja http://www.ime.unicamp.br/~cnaber/aula_MLGM_ADL_2S_2018.pdf), podemos definir médias condicionais e marginais, bem como variâncias condicionais e marginais.
- Lembremos que $x_{ij} \in \{1, 2, 3, 4, 5, 6, 8, 10, 12\}$
- $\mu_{ij}^* = e^{\beta_0 + \beta_1(x_{ij}-1)I_{(i \in \{1,2\})} + \beta_2(x_{ij}-1)I_{(i \in \{3,4,5,6,7,8,9\})}} e^{\sigma^2/2}$ (média marginal).
- $\mu_{ij} = e^{\beta_0 + \beta_1(x_{ij}-1)I_{(i \in \{1,2\})} + \beta_2(x_{ij}-1)I_{(i \in \{3,4,5,6,7,8,9\})}} e^{b_j}$.

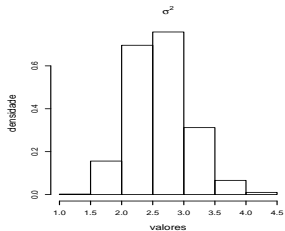
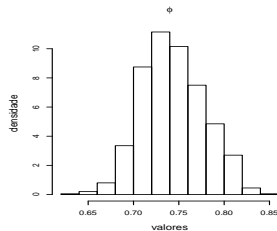
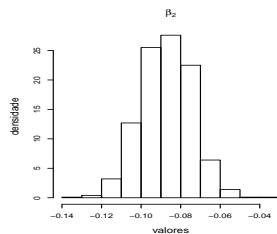
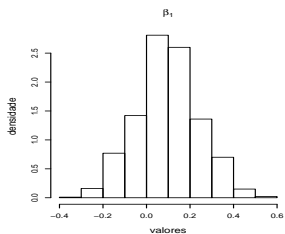
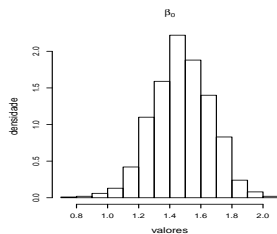
Concentração de bilirrubina em recém-nascidos saudáveis

- $\mu_{ij}|(x_{ij} = 1) = e^{\beta_0} b_j.$
- $\mu_{ij}^*|(x_{ij} = 1) = e^{\beta_0} e^{\sigma^2/2}.$
- $\mu_{ij}|(x_{ij} = 2)/\mu_{ij}|(x_{ij} = 1) = e^{\beta_1}$ (para o mesmo b_j),
- $\mu_{ij}^*|(x_{ij} = 2)/\mu_{ij}^*|(x_{ij} = 1) = e^{\beta_1}.$
- $\mathcal{V}(Y_{ij}|b_j) = \mu_{ij}^2/\phi.$
- $\mathcal{V}(Y_{ij}) = e^{2\eta_{ij} + \sigma^2} \left[\frac{1}{\phi} + e^{\sigma^2/2} (e^{\sigma^2} - 1) \right]$, em que
 $\eta_{ij} = \beta_0 + \beta_1(x_{ij} - 1)I_{(i \in \{1,2\})} + \beta_2(x_{ij} - 1)I_{(i \in \{3,4,5,6,7,8,9\})}.$

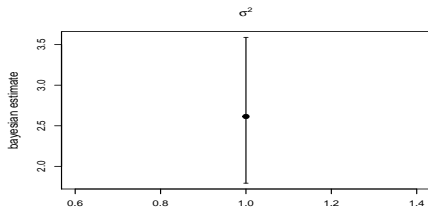
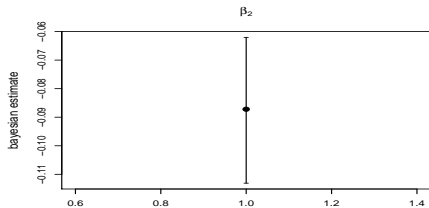
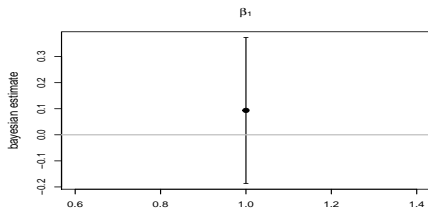
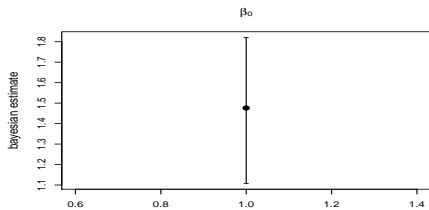
Concentração de bilirrubina em recém-nascidos saudáveis

- Análise de convergência: três conjuntos de cadeias, de tamanho 10000, foram geradas, de forma independente, a partir de pontos (diferentes) gerados de forma aleatória. Resultados <http://www.ime.unicamp.br/~cnaber/biliMGM.zip>.
- Parâmetros MCMC: burn-in = 6000, thin(lag) = 60 , número total de valor = 66000, o que gera uma amostra válida das posteriores (após burn-in, com o que devido thin) de 1000 valores.
- Sempre verificar os traceplots, gráficos de autocorrelação (ACF) e Gelman-Rubin (eventualmente também Geweke e medianas acumuladas);

Distribuições a posteriori



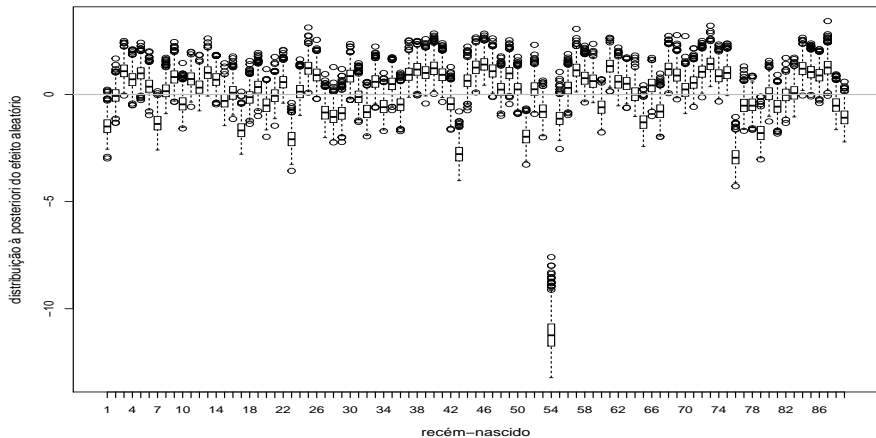
Estimativas pontuais e intervalos de credibilidade



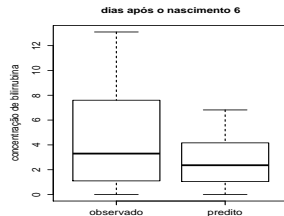
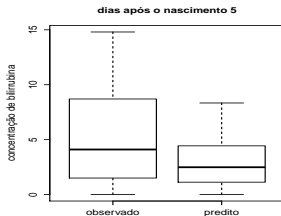
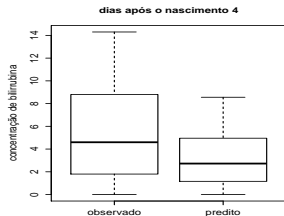
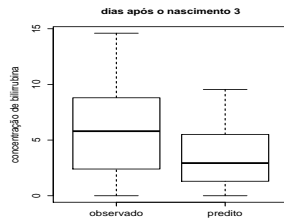
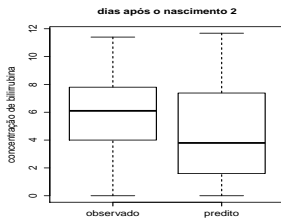
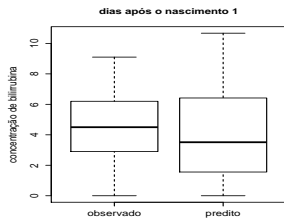
Estimativas pontuais e intervalos de credibilidade

Parâmetro	EAP	MedAP	MoAP	DPAP	IC(%95)
β_0	1,48	1,48	1,45	0,19	[1,11 ; 1,82]
β_1	0,09	0,09	0,18	0,14	[-0,19 ; 0,37]
β_2	-0,09	-0,09	-0,10	0,01	[-0,11 ; -0,06]
ϕ	0,74	0,74	0,75	0,03	[0,68 ; 0,82]
σ^2	2,61	2,59	2,45	0,47	[1,79 ; 3,59]

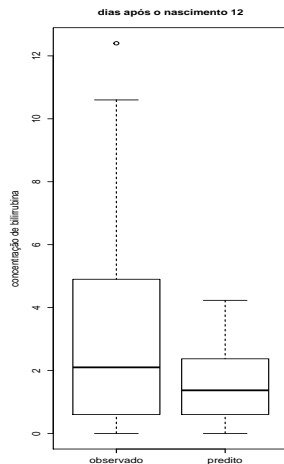
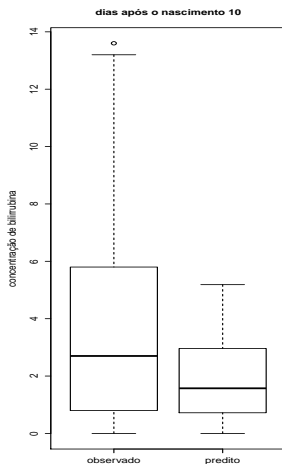
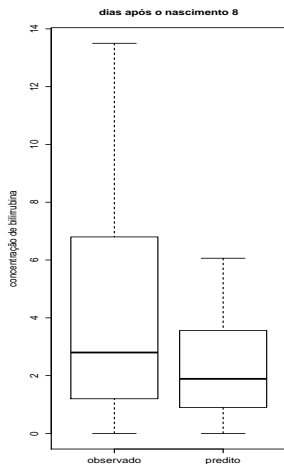
Posteriori dos efeitos aleatórios



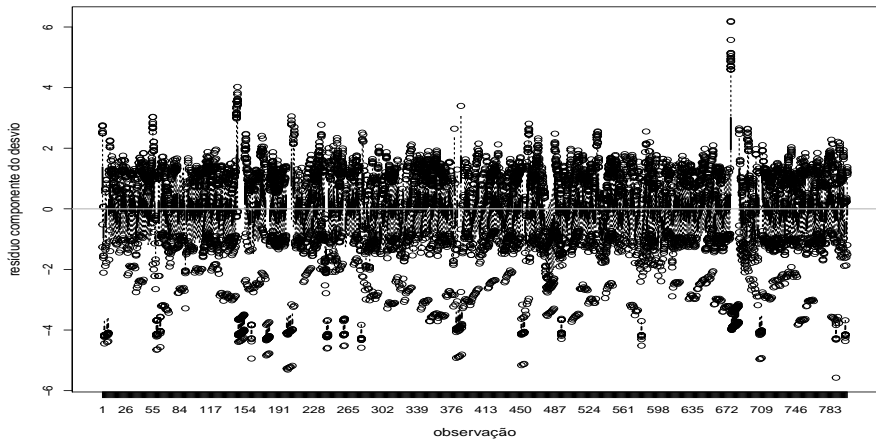
Distribuições observadas e previstas



Distribuições observadas e previstas



Medianas a posteriori e IC(95%) do RCD



Exemplo dos ataques epiléticos

- Utilizaremos o mesmo modelo considerado em http://www.ime.unicamp.br/~cnaber/aula_MLGM_ADL_2S_2018.pdf.
- Prioris: $\beta_i \stackrel{ind.}{\sim} N(0, 1000)$, $\sigma^2 \sim \text{gama}(0, 01; 0, 01)$,
 $\mathcal{E}(\phi) = \mathcal{E}(\sigma^2) = 1$, $\mathcal{V}(\sigma^2) = 100$ (prioris vagas).

Exemplo dos ataques epiléticos

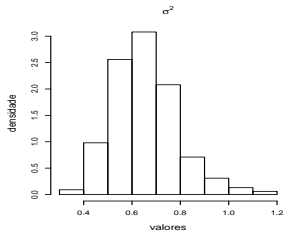
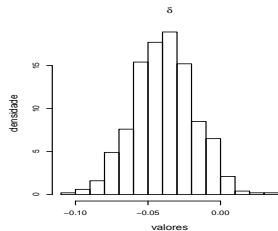
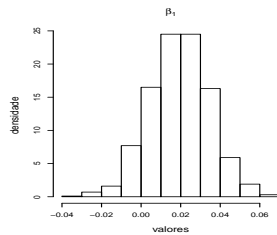
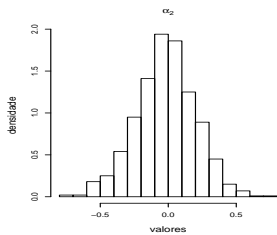
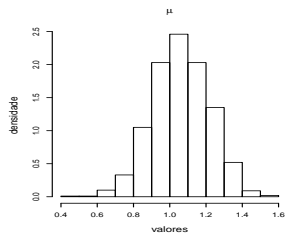
- Parâmetros MCMC: burn-in = 50000, thin(lag) = 50 , número total de iterações = 100000, o que gera uma amostra válida das posteriores (após burn-in, com o que devido thin) de 1000 valores.

Resultados `http:`

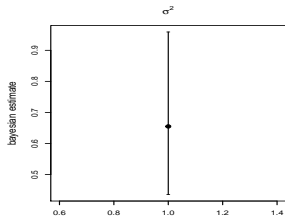
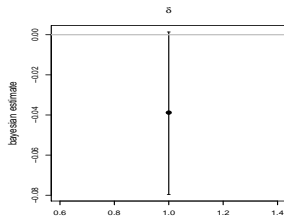
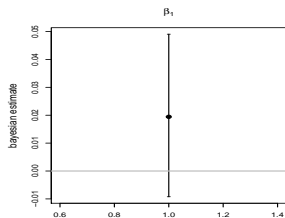
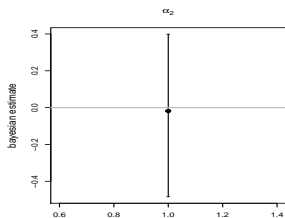
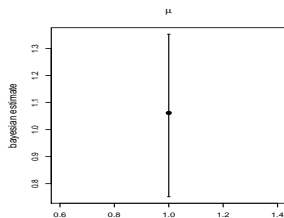
`//www.ime.unicamp.br/~cnaber/ataquesPoissonMGM.zip.`

- Problemas na convergência. Provavelmente, devido à (elevada) correlação (à posteriori) entre os interceptos fixos e os aleatórios. Possíveis soluções: utilizar algoritmos auxiliares mais apropriados, reparametrizações entre outras.

Distribuições a posteriori



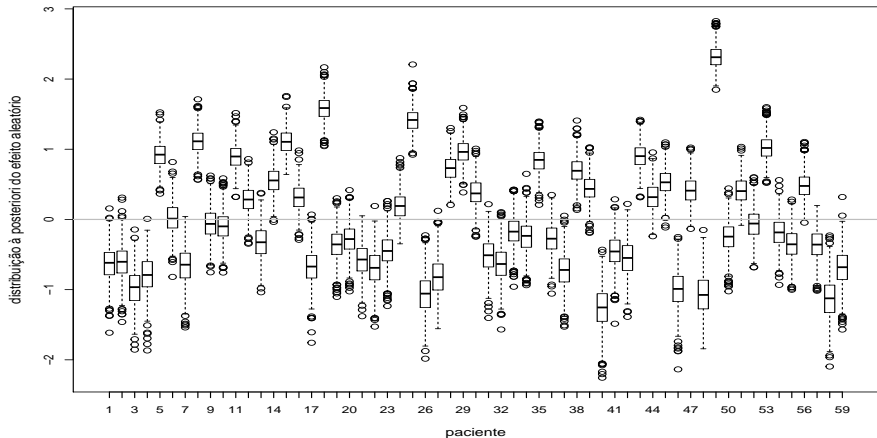
Estimativas pontuais e intervalos de credibilidade



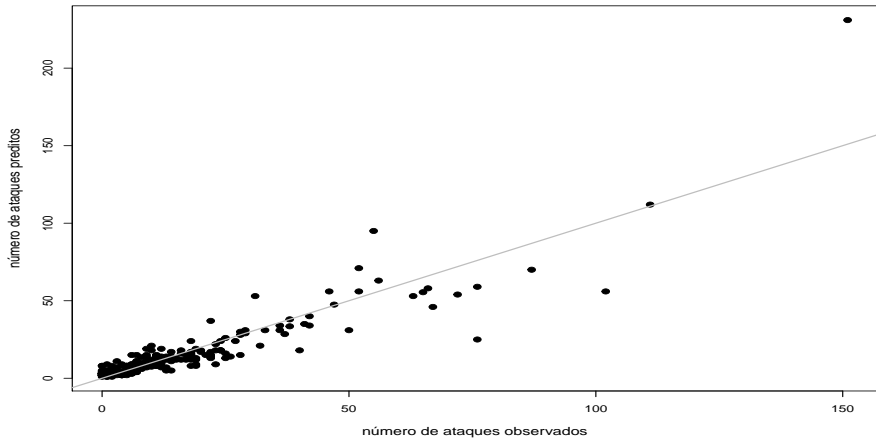
Estimativas pontuais e intervalos de credibilidade

Parâmetro	EAP	MedAP	MoAP	DPAP	IC(%95)
μ	1,06	1,06	1,01	0,16	[0,75 ; 1,35]
α_2	-0,02	-0,01	-0,03	0,22	[-0,48 ; 0,40]
β_1	0,02	0,02	0,02	0,02	[-0,01 ; 0,05]
δ	-0,04	-0,04	-0,05	0,02	[-0,08 ; 0,00]
σ^2	0,65	0,64	0,74	0,13	[0,44 ; 0,96]

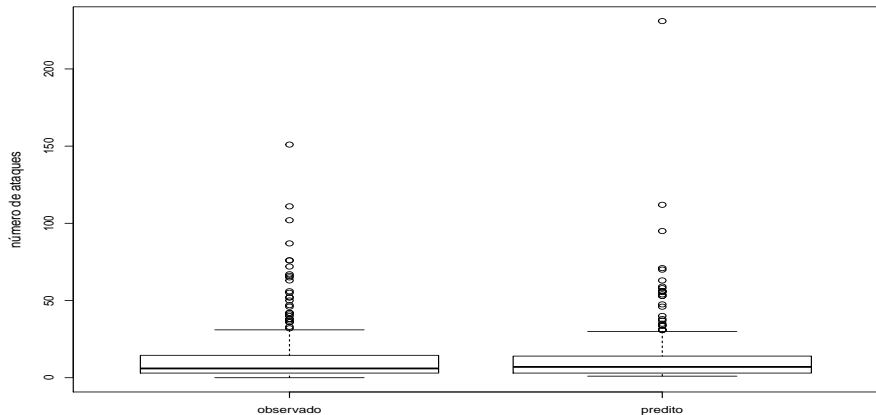
Posteriori dos efeitos aleatórios



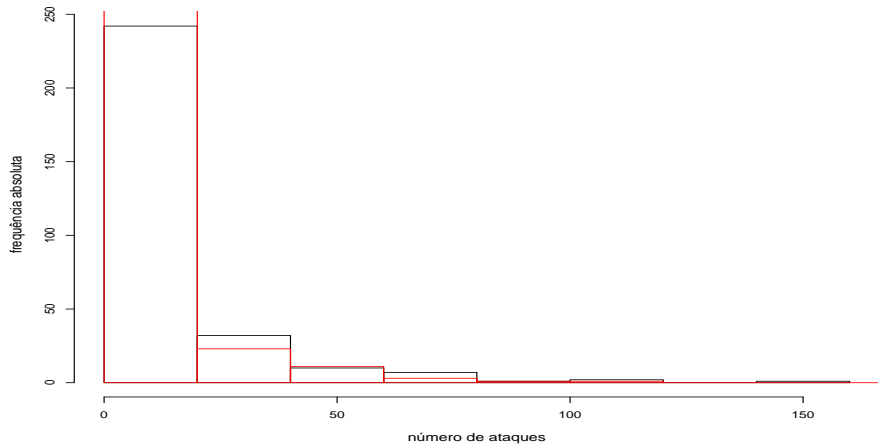
Valores observados e preditos



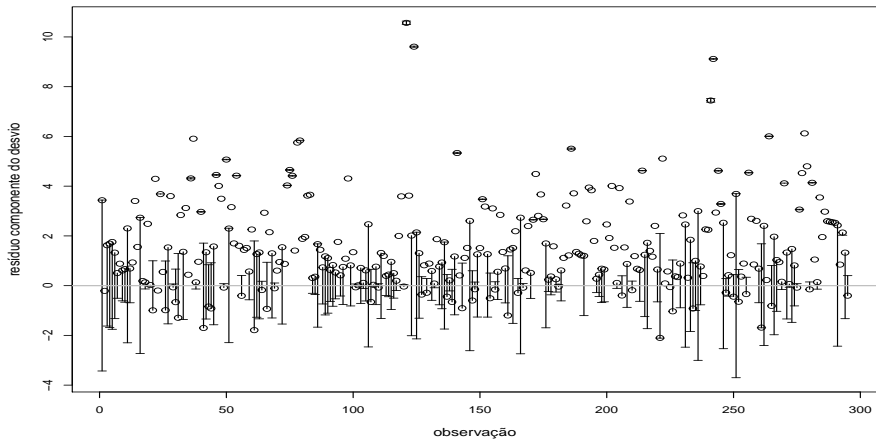
Distribuições observadas e previstas



Distribuições observadas (preto) e previstas (vermelho)



Medianas a posteriori e IC(95%) do RCD



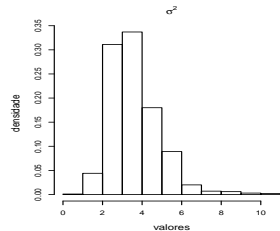
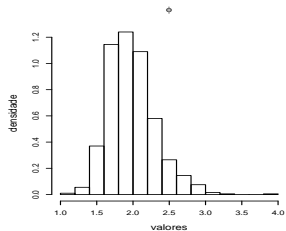
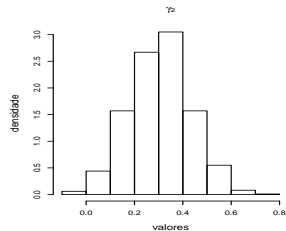
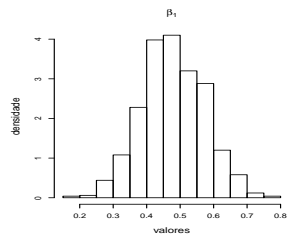
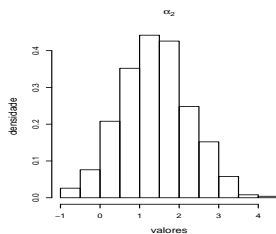
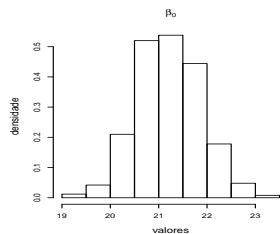
Exemplo dos dados de Potthof and Roy

- $Y_{ijk} = \mu_{ijk} + \xi_{ijk}$, $j = 1, 2, \dots, n_{ik}$, (indivíduo), $i = 1, 2, 3, 4$ (ano (condição de avaliação)),
 $k = 1, 2$ (gênero - 1: feminino, 2: masculino), $n_{i1} = 11$; $n_{i2} = 16$, $\forall i$
- $\mu_{ijk} = \beta_0 + \alpha_k + (\beta_1 + \gamma_k)(x_{ijk} - 8) + b_{jk}$.
- $\xi_{ijk} \stackrel{i.i.d.}{\sim} N(0, \psi)$, $b_{jk} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$.
- Outros modelos foram ajustados (gama e log normal), porém apresentaram problemas (numéricos).

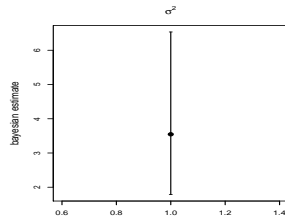
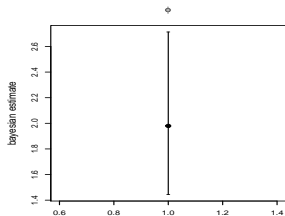
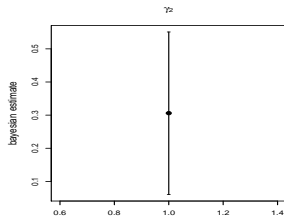
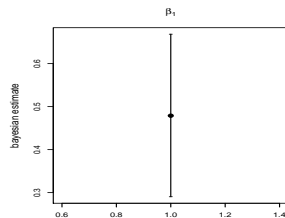
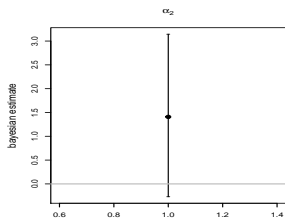
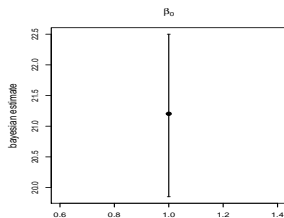
Exemplo dos dados de Potthof and Roy

- Parâmetros MCMC: burn-in = 5000, thin(lag) = 50, número total de iterações = 55000, o que gera uma amostra válida das posteriores (após burn-in, com o que devido thin) de 1000 valores. Resultados <http://www.ime.unicamp.br/~cnaber/PotRoyMGMG.zip>.
- Problemas na convergência. Provavelmente, devido à (elevada) correlação (à posteriori) entre os interceptos fixos e os aleatórios. Possíveis soluções: utilizar algoritmos auxiliares mais apropriados, reparametrizações entre outras.

Distribuições a posteriori



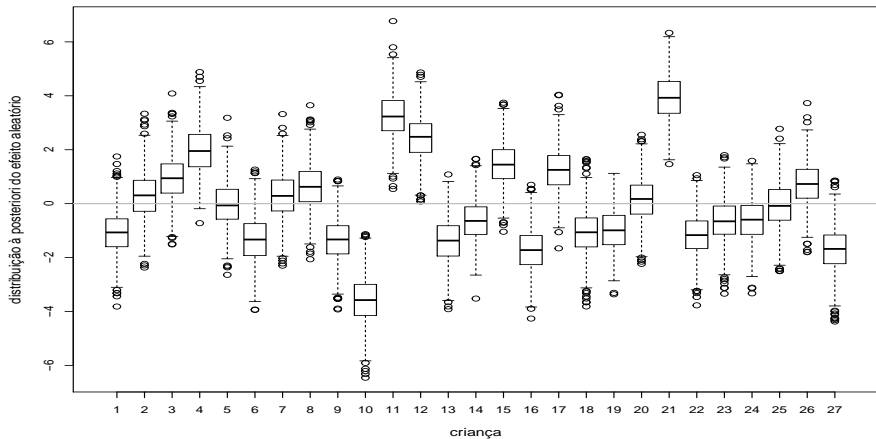
Estimativas pontuais e intervalos de credibilidade



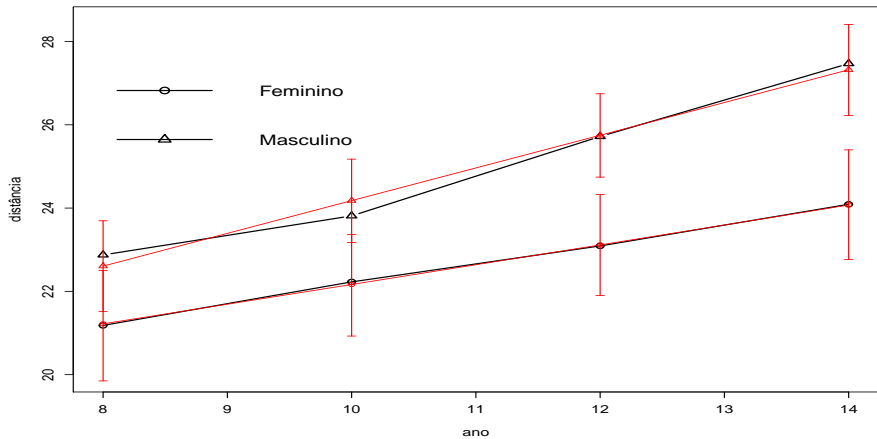
Estimativas pontuais e intervalos de credibilidade

Parâmetro	EAP	MedAP	MoAP	DPAP	IC(%95)
β_0	21,20	21,22	21,31	0,67	[19,85 ; 22,50]
α_2	1,41	1,39	1,33	0,87	[-0,27 ; 3,14]
β_1	0,48	0,48	0,50	0,10	[0,29 ; 0,67]
γ_2	0,31	0,31	0,29	0,12	[0,06 ; 0,55]
ϕ	1,98	1,95	1,85	0,32	[1,44 ; 2,71]
σ^2	3,55	3,33	3,24	1,24	[1,79 ; 6,53]

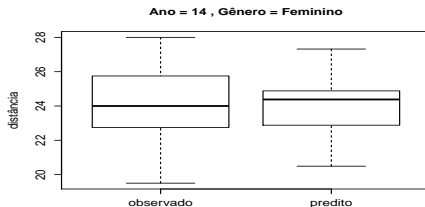
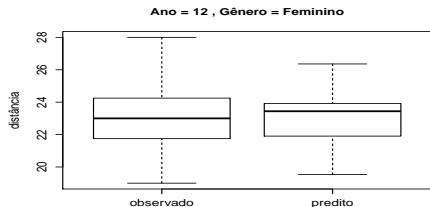
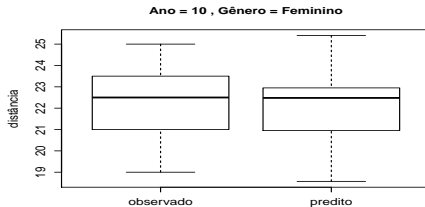
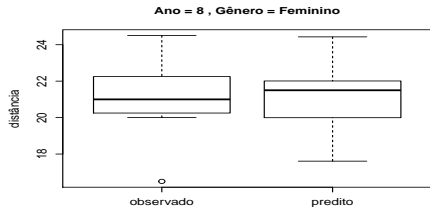
Posteriori dos efeitos aleatórios



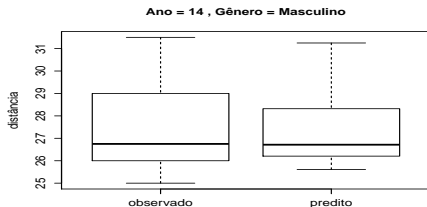
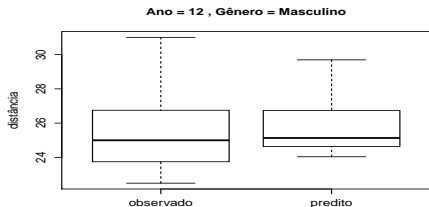
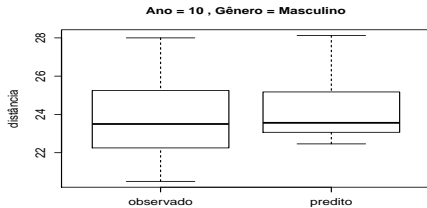
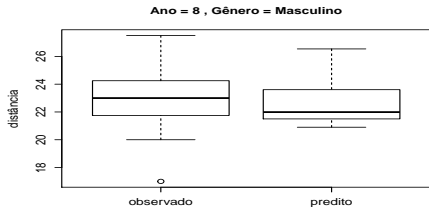
Valores observados (preto) e preditos (vermelho)



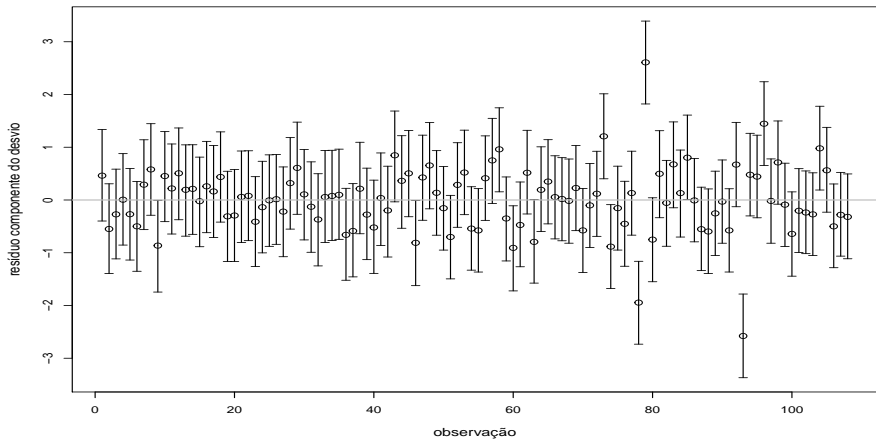
Valores observados e preditos



Valores observados e preditos



Medianas a posteriori e IC(95%) do RCD



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

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

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

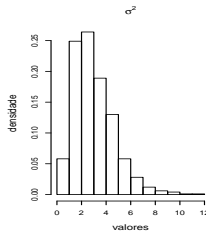
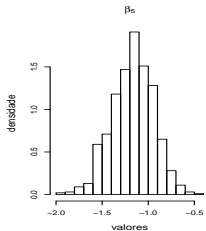
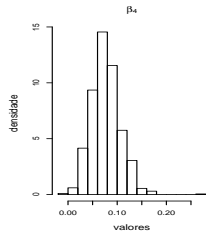
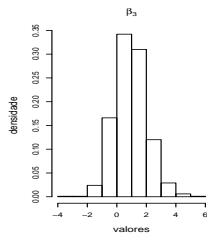
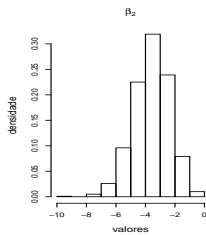
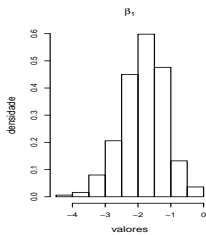
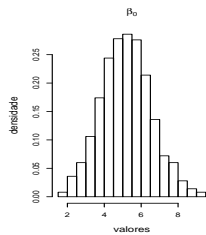
M.R. para os dados do exemplo (regressão logística)

- Y_{ij} : condição : 1 - ruim; 0 - boa.
- $idade_j$: idade (em anos), $trat_j$: (=1 droga ativa, = 0 placebo); gen_j : (=0 feminino, =1 masculino); $base_j$: (=0 ausência do nível base, =1 presença do nível base), do i -ésimo paciente e $periodo_{ij}$: instante no qual a visita foi realizada, do j -ésimo paciente no i -ésimo instante.
- Exercício: propor uma forma geral (numérica) para calcular as esperanças marginais ($\mathcal{E}(Y_{ij})$).
- Para mais detalhes, veja http://www.ime.unicamp.br/~cnaber/aula_MLGM_ADL_P2_2S_2018.pdf.

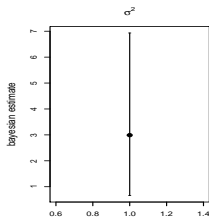
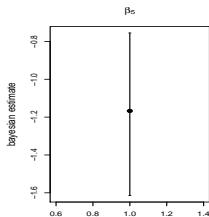
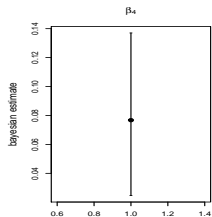
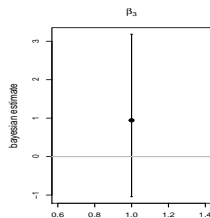
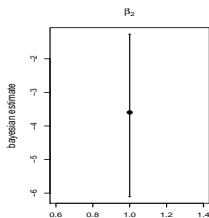
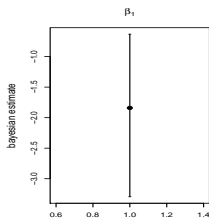
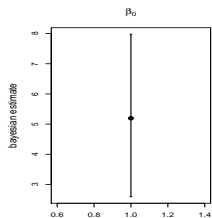
Exemplo dos dados da condicação respiratória

- Parâmetros MCMC: burn-in = 5000, thin(lag) = 50, número total de iterações = 55000, o que gera uma amostra válida das posteriores (após burn-in, com o que devido thin) de 1000 valores. Resultados <http://www.ime.unicamp.br/~cnaber/respiratorioMGM.zip>.
- Problemas na convergência. Provavelmente, devido à (elevada) correlação (à posteriori) entre os interceptos fixos e os aleatórios. Possíves soluções: utilizar algoritmos auxiliares mais apropriados, reparametrizações entre outras.

Distribuições a posteriori



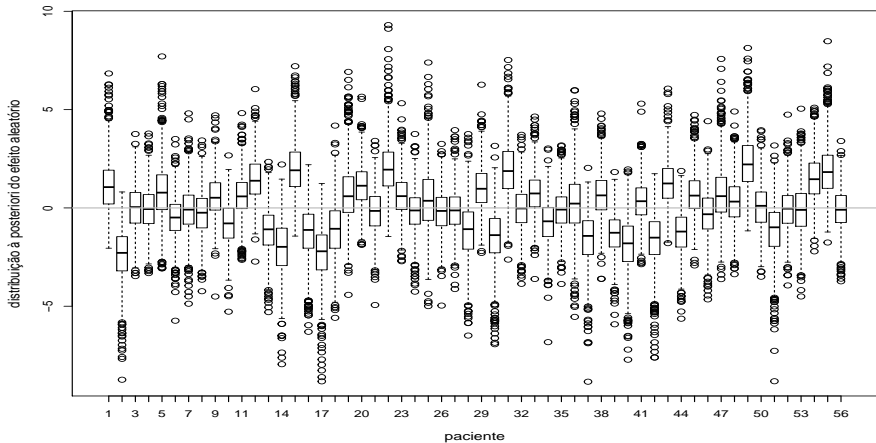
Estimativas pontuais e intervalos de credibilidade



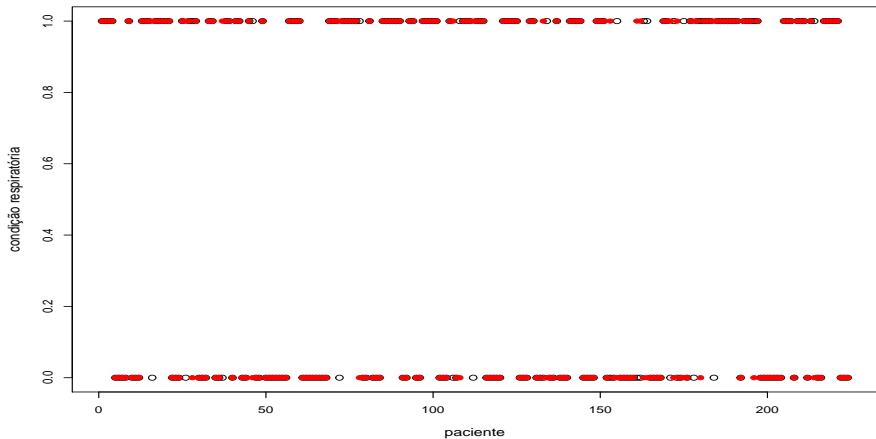
Estimativas pontuais e intervalos de credibilidade

Parâmetro	EAP	MedAP	MoAP	DPAP	IC(%95)
β_0	5,19	5,18	5,47	1,36	[2,59 ; 7,98]
β_1	-1,84	-1,80	-1,59	0,67	[-3,30 ; -0,63]
β_2	-3,60	-3,53	-3,63	1,25	[-6,11 ; -1,27]
β_3	0,94	0,91	1,33	1,08	[-1,04 ; 3,18]
β_4	0,08	0,08	0,11	0,03	[0,02 ; 0,14]
β_5	-1,17	-1,15	-1,18	0,23	[-1,61 ; -0,76]
σ^2	2,99	2,64	1,88	1,63	[0,66 ; 6,93]

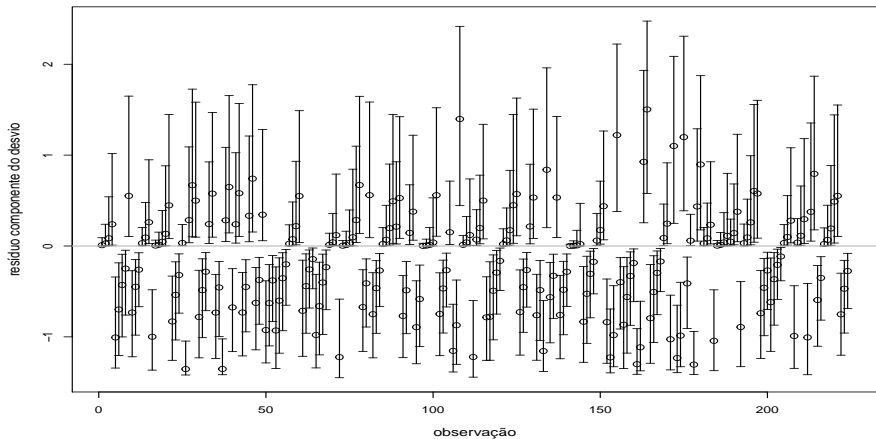
Posteriori dos efeitos aleatórios



Valores observados (preto) e preditos (vermelho)



Medianas a posteriori e IC(95%) do RCD



Exercício

- Analise o conjunto de dados constante na Prova I (disponível no R, no pacote “lme4”, sob o nome “sleepstudy”) utilizando algum modelo apropriado, através da metodologia bayesiana. Realize todas as análises de diagnóstico apropriadas.