

Algoritmos MCMC para inferência bayesiana (Parte 2)

Prof. Caio Azevedo

Exemplo: mortalidade de besouros

- Dados relativos ao percentual de besouros mortos quando expostos à diferentes doses de disulfeto de carbono gasoso (CS_2).

Dose: $\log_{10} CS_2$	n° Besouros expostos	n° Besouros mortos
1,6907	59	6
1,7242	60	13
1,7552	62	18
1,7842	56	28
1,8113	63	52
1,8369	59	53
1,8610	62	61
1,8839	60	60

Exemplo: mortalidade de besouros

- Modelo1

$$Y_i | (\beta_0, \beta_1) \sim \text{Binomial}(m_i, p_i)$$
$$\ln \left(\frac{p_i}{1 - p_i} \right) = \beta_0 + \beta_1 x_i, i = 1, 2, \dots, 8$$

- m_i : número de besouros expostos à dose i de CS_2 .
- Y_i : número de besouros expostos à dose i de CS_2 que morreram.
- x_i : dose (log da concentração de CS_2) à que os besouros do grupo i foram expostos.

Cont. do modelo 1

- Assim, $p_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$
- β_0 é o logito $\left[\ln \left(\frac{p_i}{1-p_i} \right) \right]$ da proporção de besouros mortos submetidos à uma concentração de 1 unidade de CS_2 . Ou seja, se $x_i = \log_{10}(\text{concent}) = \log_{10}(1) = 0$ então $p_i = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$.

Cont. do modelo 1

- Sejam: $p_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$ e $p_{i+1} = \frac{e^{\beta_0 + \beta_1 (x_i + 1)}}{1 + e^{\beta_0 + \beta_1 (x_i + 1)}}$.
- Assim: $\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_i$ e $\ln\left(\frac{p_{i+1}}{1-p_{i+1}}\right) = \beta_0 + \beta_1 (x_i + 1)$.
- Logo: $\ln\left(\frac{p_{i+1}}{1-p_{i+1}}\right) - \ln\left(\frac{p_i}{1-p_i}\right) = \ln\left(\frac{p_{i+1}/(1-p_{i+1})}{p_i/(1-p_i)}\right) = \beta_1$.
- Portanto, $\frac{p_{i+1}/(1-p_{i+1})}{p_i/(1-p_i)} = e^{\beta_1}$ (conhecida como razão de chances).

Exemplo: mortalidade de besouros

■ Modelo 2

$$Y_i | (\beta_0, \beta_1) \sim \text{Binomial}(m_i, p_i)$$

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(x_i - \bar{x}), \bar{x} = \frac{1}{8} \sum_{i=1}^8 x_i, i = 1, 2, \dots, 8$$

- Neste caso, β_0 é o logito $\left[\ln\left(\frac{p_i}{1-p_i}\right)\right]$ da proporção de besouros mortos submetidos à uma concentração igual à \bar{x} unidades de CS_2 .

Ou seja, se $x_i = \log_{10}(\text{concent}) = \frac{1}{8} \sum_{i=1}^8 \log_{10}(\text{concent}_i)$ então

$$p_i = \frac{e^{\beta_0}}{1 + e^{\beta_0}}.$$

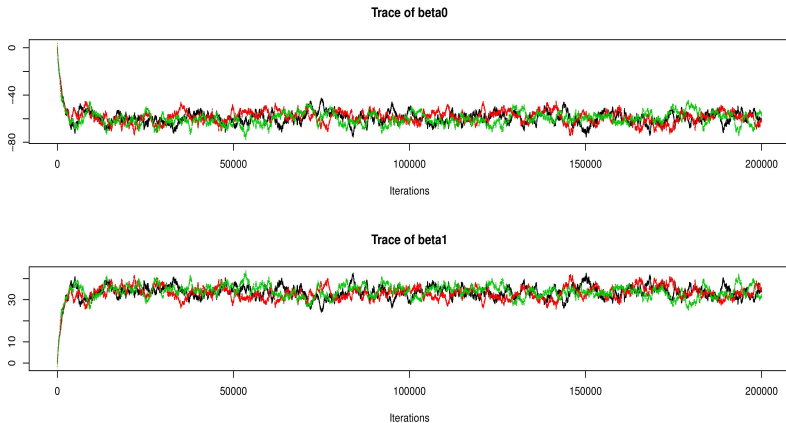
- As outras quantidades, incluindo o parâmetro β_1 , possuem as mesmas interpretações que no modelo 1, (substituindo-se x_i por $x_i - \bar{x}$).

Comandos

- Algoritmo WinBUGS para o ajuste do modelo.

```
ComBes
probitmod<-function(){
  for( i in 1 : N ) {
    r[i] ~ dbin(p[i],n[i])
    logit(p[i]) <- beta0 + beta1 * (x[i]-1.793425)
#    logit(p[i]) <- beta0 + beta1 * (x[i])
    auxp[i] <- beta0+beta1*(x[i]-1.793425)
    auxp1[i] <-exp(auxp[i])/(1+exp(auxp[i]))
    rpredaux[i] ~ dbin(auxp1[i],n[i])
    rpred[i] <- rpredaux[i]/n[i]
  }
  beta0 ~ dnorm(0.0,0.001)
  beta1 ~ dnorm(0.0,0.001)
}
```

Traceplots para os três conjuntos de cadeias geradas: modelo 1



Autocorrelações para um dos conjuntos de cadeias: modelo

1

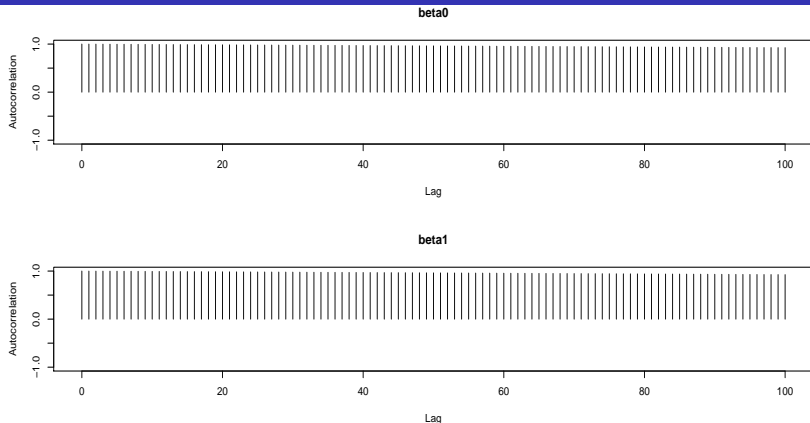


Gráfico das medianas acumuladas para um dos conjuntos de cadeias: modelo 1

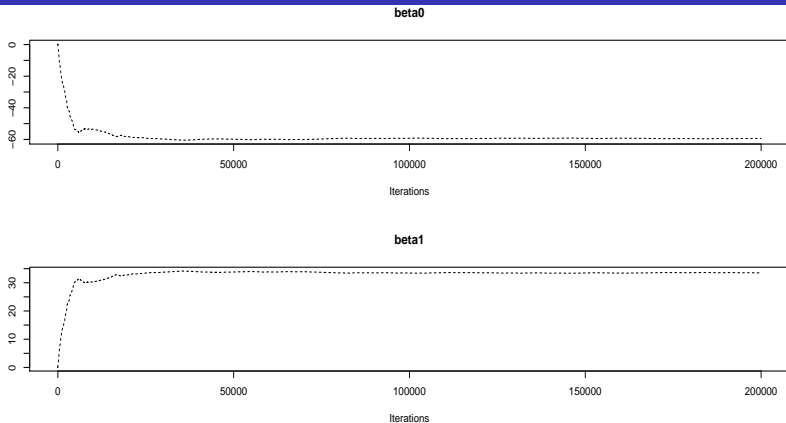


Gráfico da estatística de Geweke para um dos conjuntos de cadeias: modelo 1

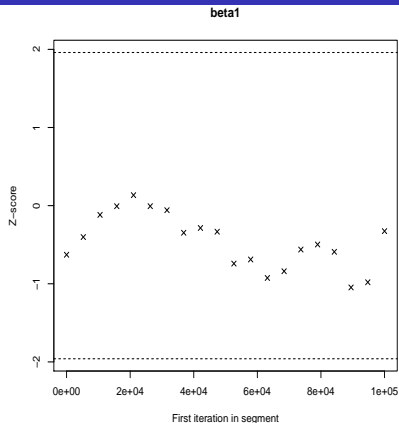
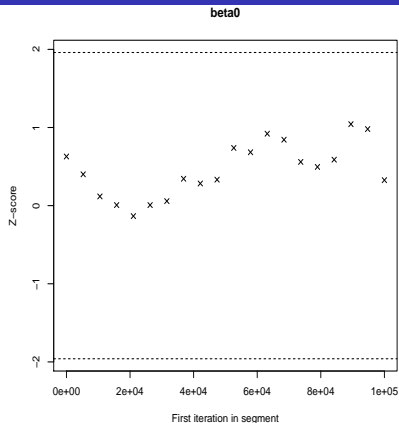
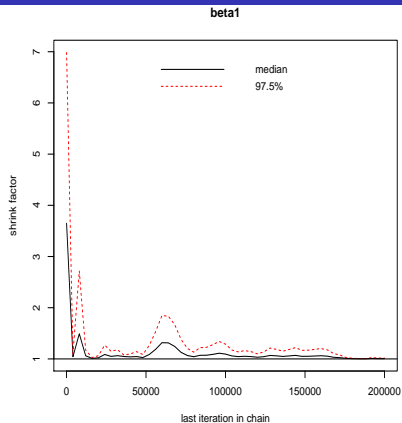
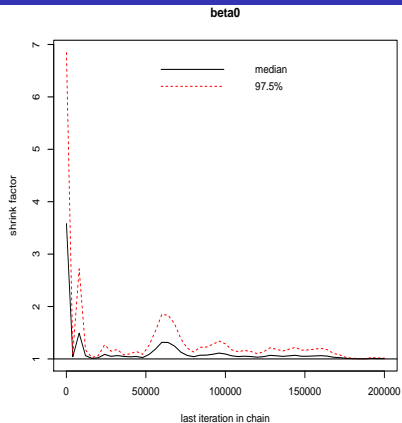
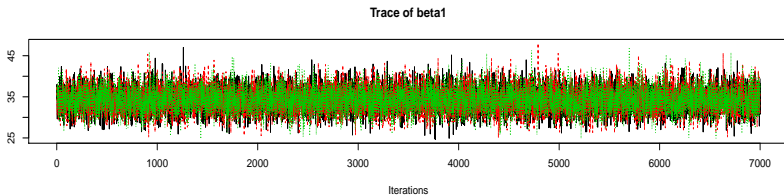
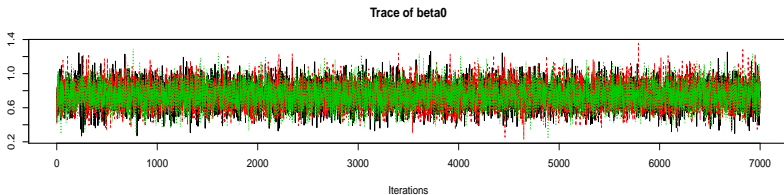


Gráfico da estatística de Gelman-Rubin utilizando os três conjuntos de cadeias: modelo 1



Traceplots para os três conjuntos de cadeias geradas: modelo 2



Autocorrelações para um dos conjuntos de cadeias: modelo

2

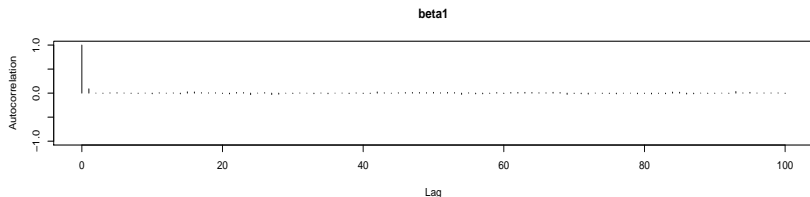
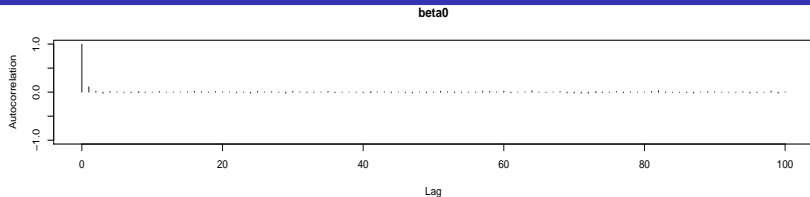


Gráfico das medianas acumuladas para um dos conjuntos de cadeias: modelo 2

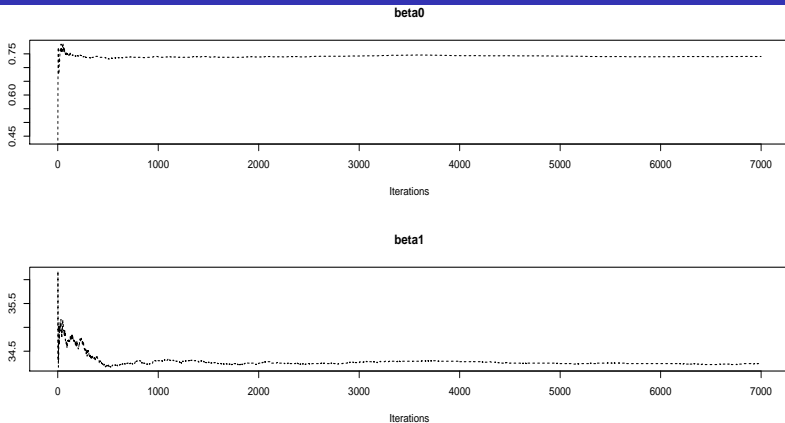


Gráfico da estatística de Geweke para um dos conjuntos de cadeias: modelo 2

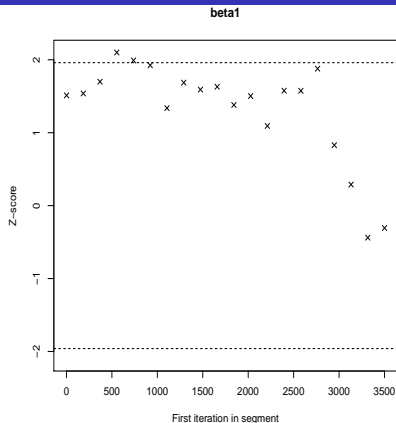
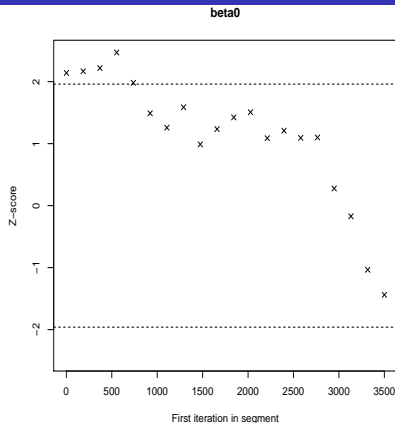
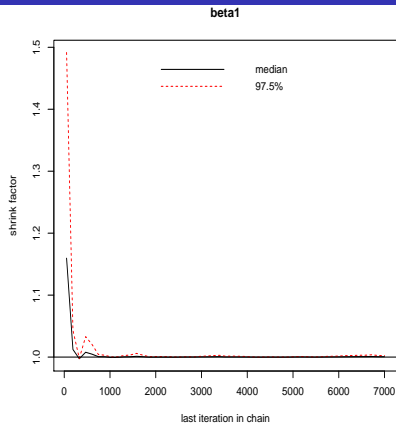
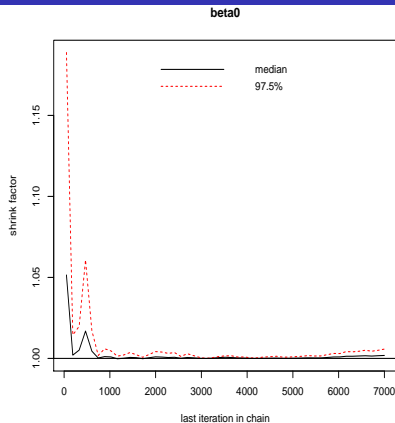
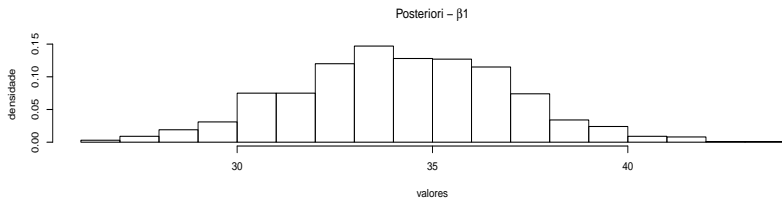
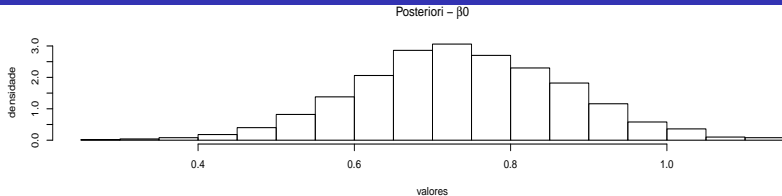


Gráfico da estatística de Gelman-Rubin utilizando os três conjuntos de cadeias: modelo 2



Histograma da amostra válida para um dos conjuntos de cadeias: modelo 2



Estimativas bayesianas para um dos conjuntos de cadeias

■ Resumo

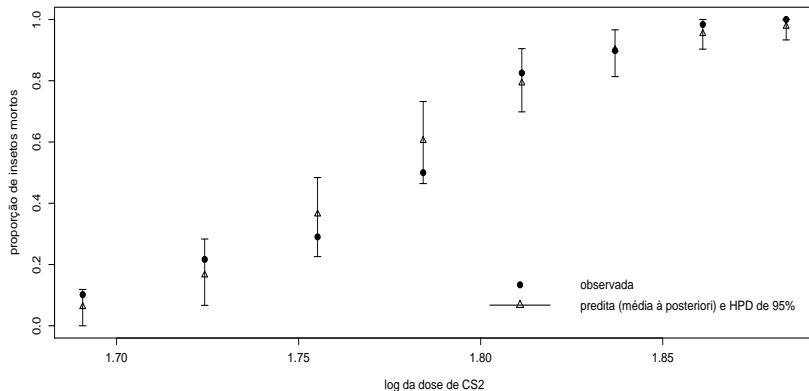
Parâmetro	EAP	DPAP	$IC_B(95\%)$	$HPD(95\%)$
β_0	0,74	0,13	[0,49 ; 1,00]	[0,49 ; 1,00]
β_1	34,19	2,78	[28,66 ; 39,68]	[29,14 ; 40,08]

- Modelo completo : Deviance = 39,4, $p_D = 2, 0$; $DIC = 41, 5$.

Modelo só com o intercepto (β_0): Deviance = 311,5,

$p_D = 1, 1$; $DIC = 312, 5$.

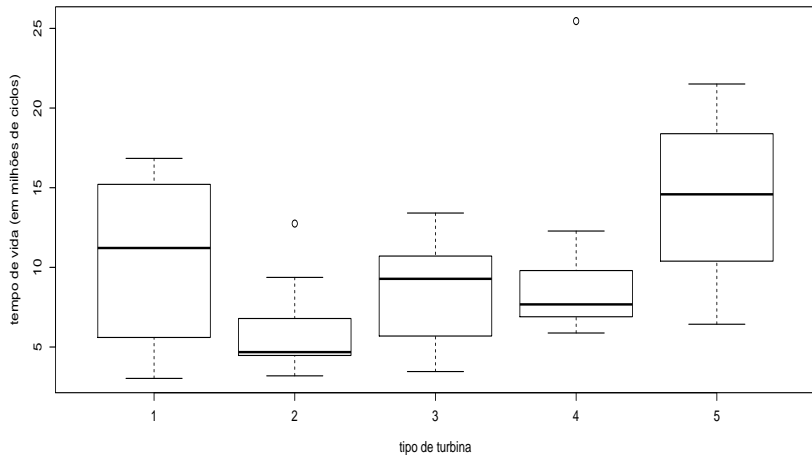
Proporções de insetos mortos observadas e previstas pelo modelo 2 para cada valor da dose



Voltando ao exemplo da potência de turbinas de aviões

- Vamos considerar os 5 tipos de turbinas analisados no experimento, doravante tipos 1, 2, 3, 4 e 5.
- $n_i = 10, \forall i$ (tamanho amostral por grupo).
- y_{ij} : tempo de vida (em milhões de ciclos) da j -ésima turbina do i -ésimo tipo.
- Suposição: $Y_{ij} | \beta, \phi \stackrel{ind.}{\sim} \text{gama}(\mu_i, \phi), i = 1, 2, 3, 4, 5, j = 1, 2, \dots, n_i$, em que $\mathcal{E}(Y_{ij} | \theta) = \mu_i$, $\mathcal{V}(Y_{ij} | \theta) = \frac{\mu_i^2}{\phi}$, e $\theta = (\beta, \phi)'$, ϕ é chamado de parâmetro de precisão.

Análise descritiva



Análise descritiva

Tipo de turbina	Média	DP	Var.	CV(%)	Mínimo	Máximo
1	10,69	4,82	23,23	45,07	3,03	16,84
2	6,05	2,92	8,50	48,18	3,19	12,75
3	8,64	3,29	10,83	38,10	3,46	13,41
4	9,80	5,81	33,71	59,26	5,88	25,46
5	14,71	4,86	23,65	33,07	6,43	21,51

Modelo 1

- Média: $\ln \mu_i = \alpha + \beta_i, \beta_1 = 0$ (casela de referência). Logo

$$\mu_i = e^{\alpha + \beta_i} \text{ e } \boldsymbol{\beta} = (\alpha, \beta_2, \beta_3, \beta_4, \beta_5)'$$

- $p(y_{ij} | \boldsymbol{\beta}, \phi) = \frac{1}{\Gamma(\phi)} \left(\frac{\phi y_{ij}}{\mu_i} \right)^\phi e^{(-\phi y_{ij} / \mu_i)} \mathbf{1}_{(0, \infty)}(y_{ij})$.

- Resumindo:

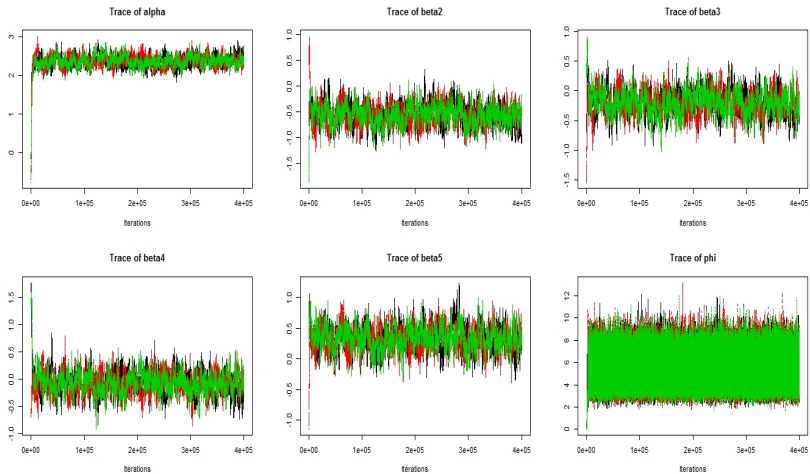
$$Y_{ij} | \boldsymbol{\beta}, \phi \stackrel{ind.}{\sim} \text{gama}(\mu_i, \phi)$$

$$\ln \mu_i = \alpha + \beta_i, \beta_1 = 0$$

$$\mathcal{E}(Y_{ij} | \boldsymbol{\theta}) = \mu_i \quad ; \quad \mathcal{V}(Y_{ij} | \boldsymbol{\theta}) = \frac{\mu_i^2}{\phi}$$

- Prioris: $\boldsymbol{\beta}: \beta_i \sim N(0, 100), \phi \sim \text{gama}(0, 01, 0, 01)$ (vaga)
(parametrização do curso), mutuamente independentes entre si.

Traceplots para os três conjuntos de cadeias geradas



Autocorrelações para um dos conjuntos de cadeias

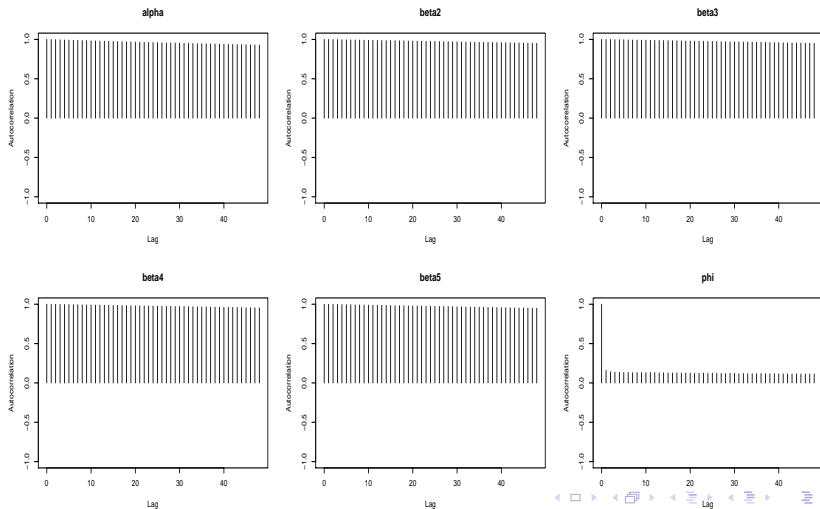


Gráfico da estatística de Geweke para um dos conjuntos de cadeias

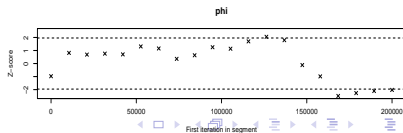
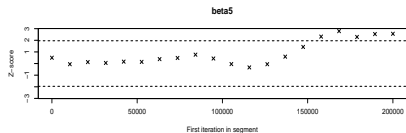
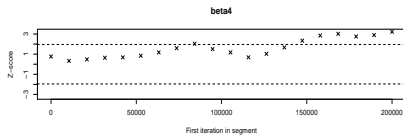
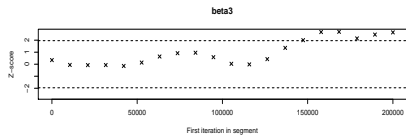
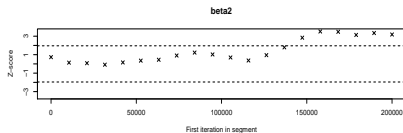
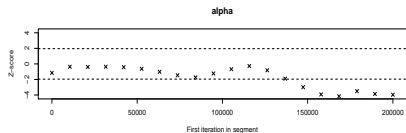
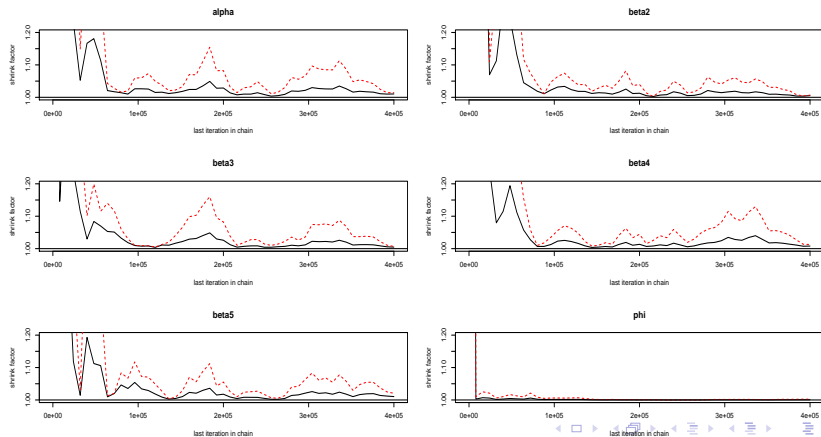
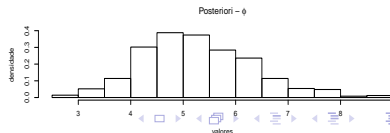
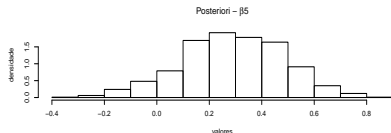
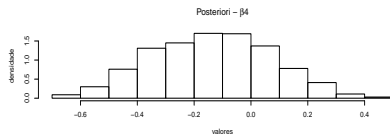
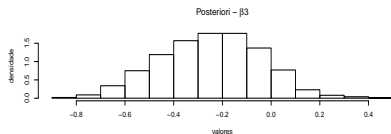
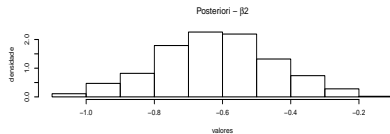
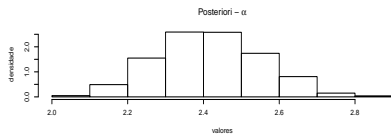


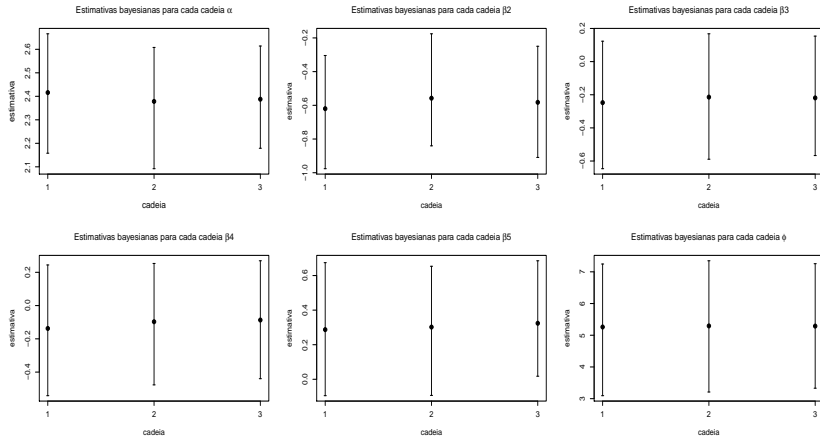
Gráfico da estatística de Gelman-Rubin utilizando os três conjuntos de cadeias



Histograma da amostra válida para um dos conjuntos de cadeias



EAP e HPD(95%) para todas as cadeias

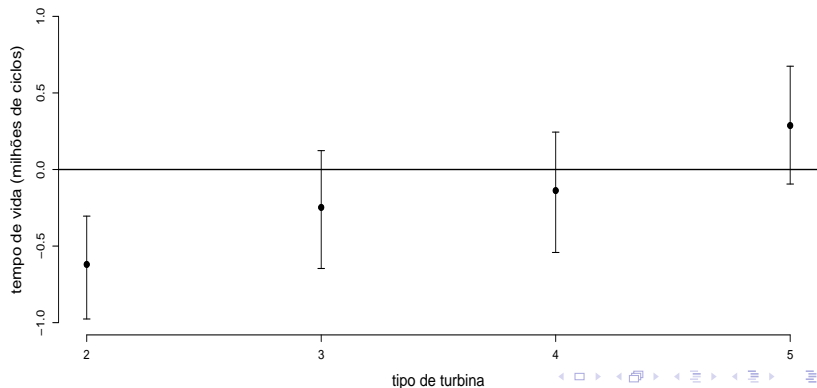


Estimativas bayesianas para um dos conjuntos de cadeias

■ Resumo

Parâmetro	EAP	DPAP	$IC_B(95\%)$	$HPD(95\%)$
α	2,42	0,14	[2,17 ; 2,68]	[2,16 ; 2,67]
β_2	-0,62	0,17	[-0,96 ; -0,28]	[-0,98 ; -0,30]
β_3	-0,25	0,21	[-0,64 ; 0,14]	[-0,65 ; 0,12]
β_4	-0,14	0,21	[-0,53 ; 0,27]	[-0,54 ; 0,24]
β_5	0,29	0,20	[-0,12 ; 0,65]	[-0,09 ; 0,67]
ϕ	5,26	1,05	[3,38 ; 7,64]	[3,09 ; 7,25]

EAP e HPD(95%) para um dos conjuntos de cadeias para os parâmetros $(\beta_2, \beta_2, \beta_4, \beta_5)$



Comparação de modelos

- Vamos comparar os seguintes modelos:
 - Modelo 1: modelo completo.
 - Modelo 2: modelo 1 com $\beta_4 = 0$.
 - Modelo 3: modelo 1 com $\beta_4 = \beta_3 = 0$.
 - Modelo 4: modelo 1 com $\beta_5 = \beta_4 = \beta_3 = 0$.

Comparação de modelos

■ Resultados

	Modelo			
Estatística	1	2	3	4
Deviance	280,1	279,4	279,1	285,5
p_D	5,9	5,1	3,8	3,0
DIC	286,1	284,5	282,9	288,6

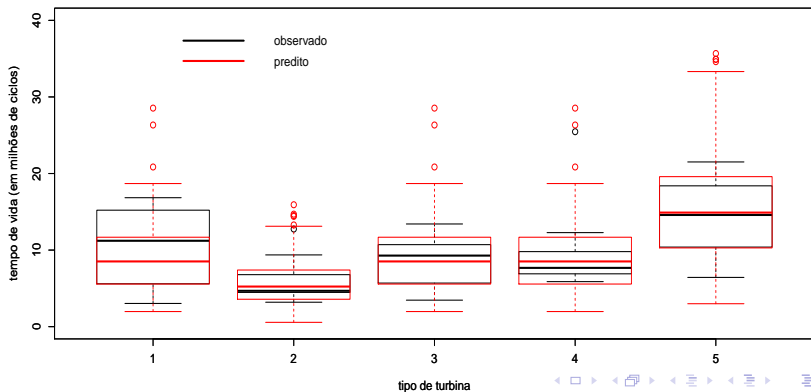
- O modelo 3 ($\beta_4 = \beta_3 = 0$) foi selecionado .

Estimativas Bayesianas para um dos conjuntos de cadeias: modelo 3

■ Resumo

Parâmetro	EAP	DPAP	$IC_B(95\%)$	$HPD(95\%)$
α	2,27	0,08	[2,10 ; 2,43]	[2,12 ; 2,44]
β_2	-0,46	0,17	[-0,77 ; -0,10]	[-0,78 ; -0,13]
β_5	0,43	0,17	[0,14 ; 0,77]	[0,13 ;0,75]
ϕ	5,28	1,07	[3,44 ; 7,84]	[3,35 ; 7,36]
μ_{134}	9,74	0,82	[8,19 ; 11,40]	[8,31 ; 11,45]
μ_2	6,16	0,91	[4,72 ; 8,12]	[4,66 ; 7,92]
μ_5	15,14	2,27	[11,40 ; 19,99]	[11,18 ; 19,60]

Boxplots dos tempos de vidas observados e preditos pelo modelo 3



Tempos médios de vida ajustados pelo modelo 3

