PREVISÕES BAYESIANAS E MODELOS GRÁFICOS

Ralph dos Santos Silva

http://www.ralphsilva.org/previsoesbayesianasemodelosgraficos.html

Instituto de Matemática
Centro de Ciências Matemáticas e da Natureza
Universidade Federal do Rio de Janeiro

OpenBUGS e WinBUGS

- Bastante *amigável* desde o ponto de vista do usuário.
- O **OpenBUGS** está disponível no seguinte sítio: www.mrc-bsu.cam.ac.uk/software/bugs/openbugs/.
- O **WinBUGS** está disponível no seguinte sítio: www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/.
- É uma ferramenta ideal para a modelagem.
- Permite a modelagem via DAG[1], isto é, modelo gráfico direcionado.
- Permite mudar a especificação da distribuição a priori.

[1] *Directed Acyclic Graph*

Exemplo

Considere o modelo de regressão linear para $d = 2$. Utilizaremos uma representação DAG do modelo sob o enfoque bayesiano. Em seguida, aplicaremos Monte Carlo via cadeias de Markov para obter uma amostra da posteriori.

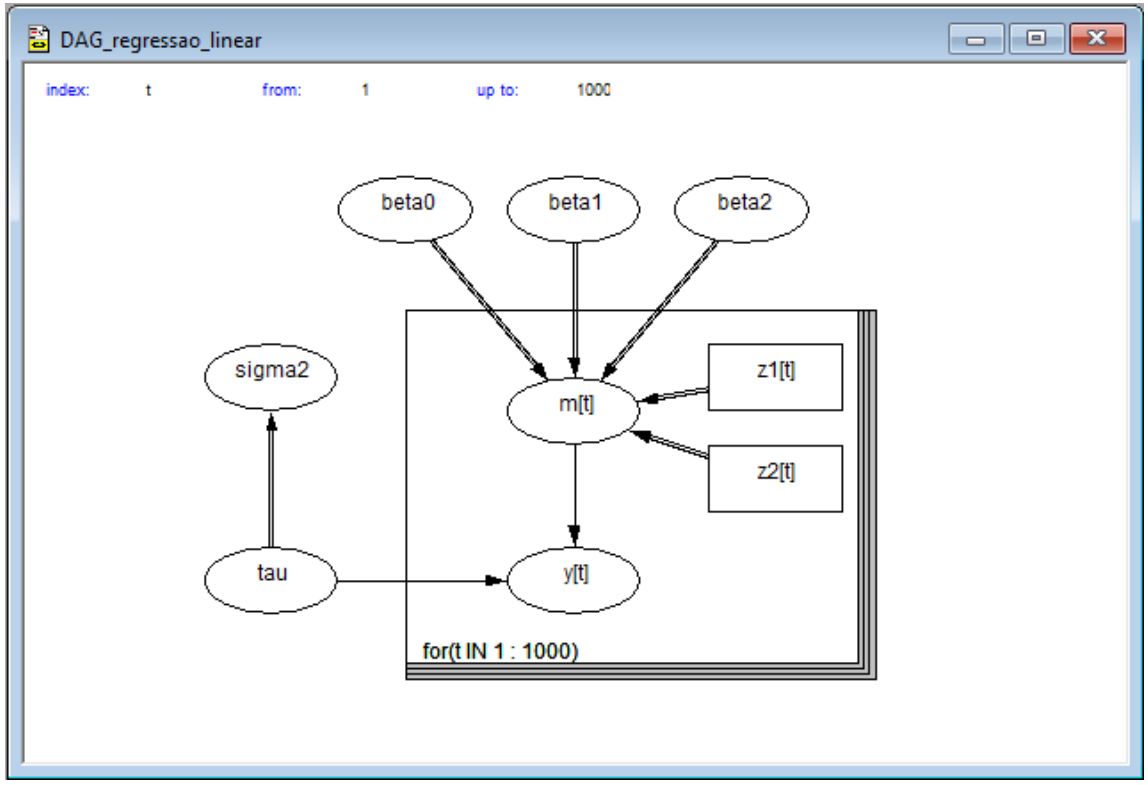

Figura 1: Representação gráfica do modelo de regressão linear em código BUGS.

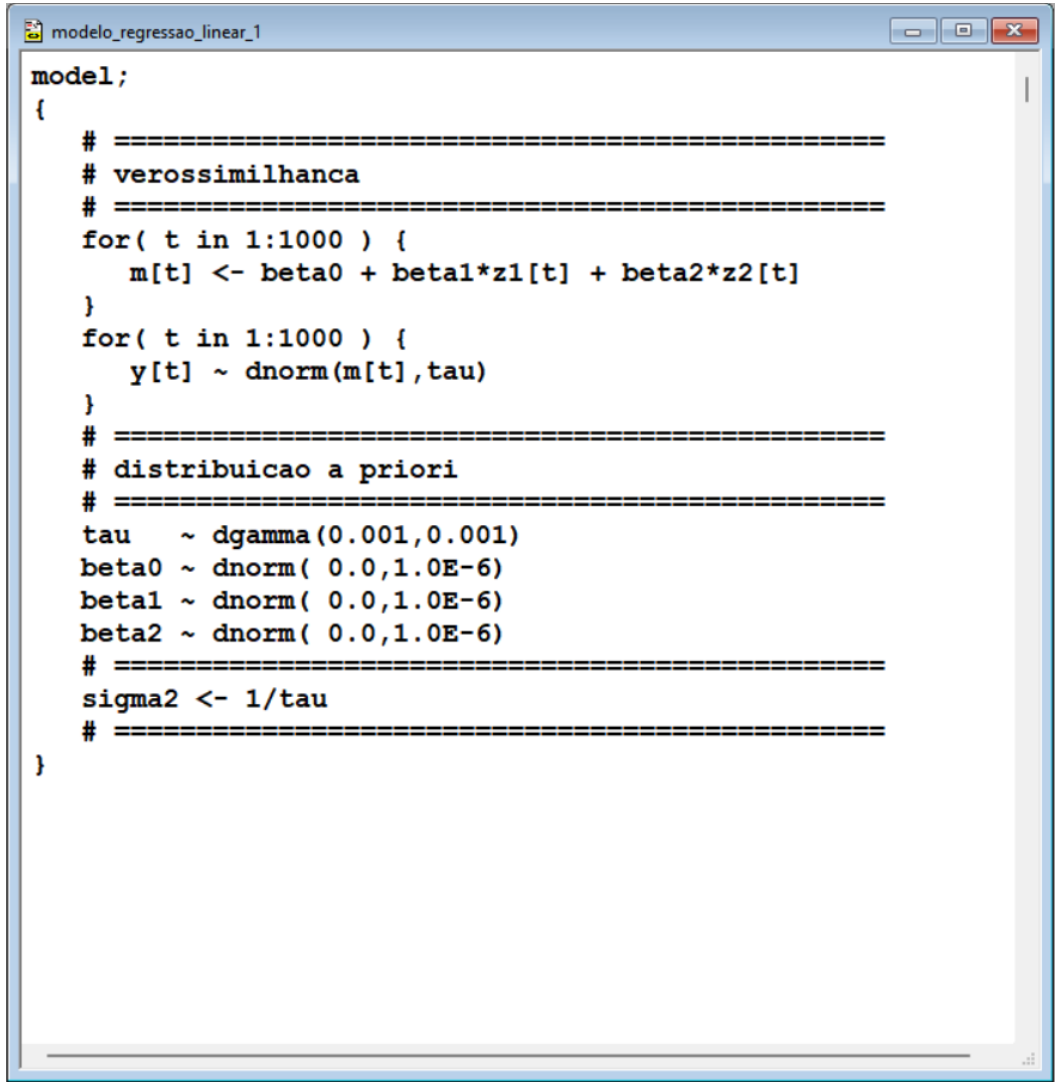

Figura 2: O modelo de regressão linear em código BUGS.

Opções básicas do OpenBUGS (WinBUGS)

1. ***Model*→ *Specification...***
2. ***check model***
   Verificar por *model is syntactically correct*
3. ***load data***
   Verificar por *data loaded*
4. ***compile***
   Verificar por *model compiled*
5. ***load inits*** ou ***gen inits***
   Verificar por *initial values generated, model initialized*
6. ***Inference*→ *Samples...***
7. ***Model*→ *Update...***

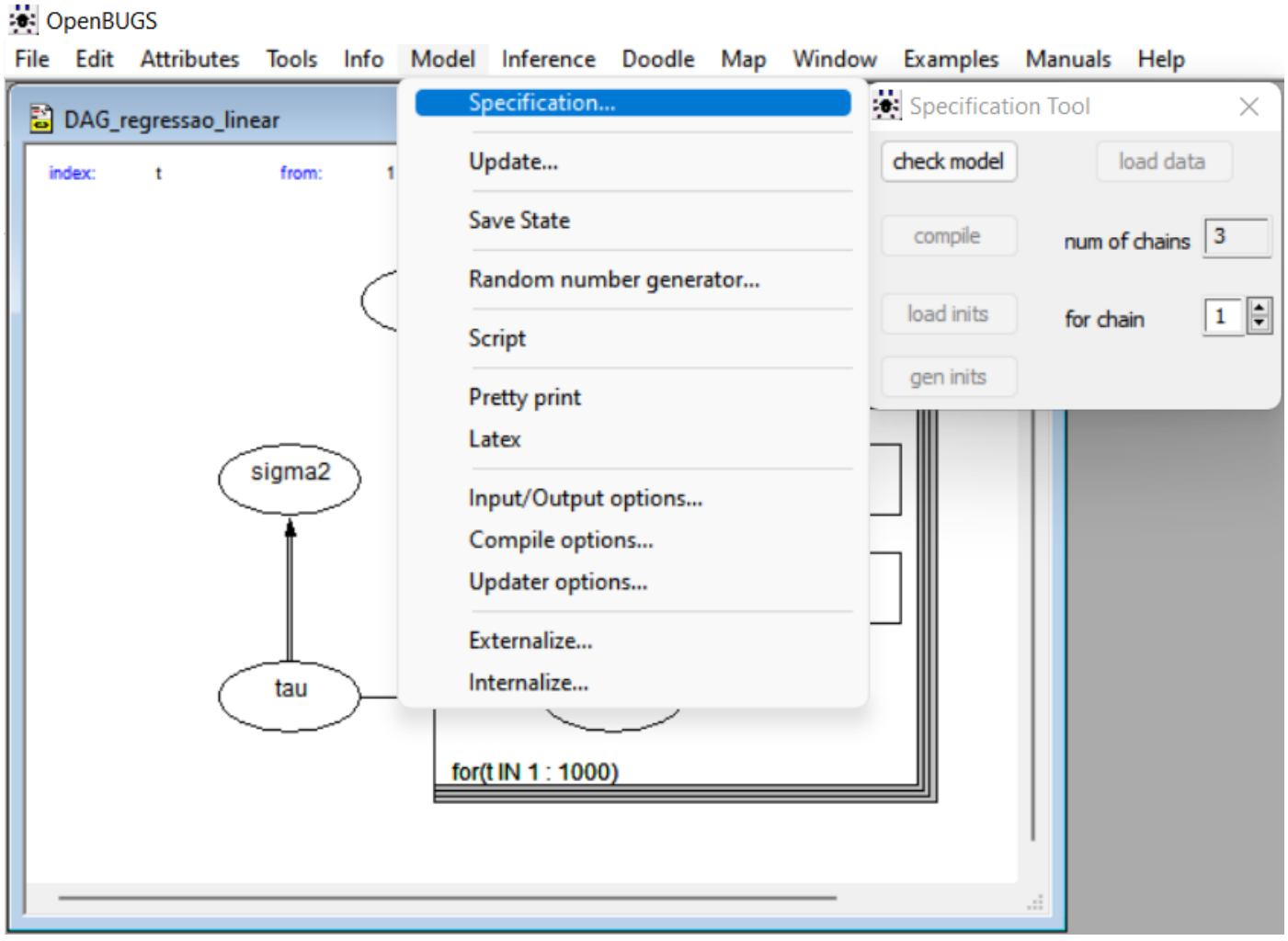

Figura 3: Executando os primeiros passos.

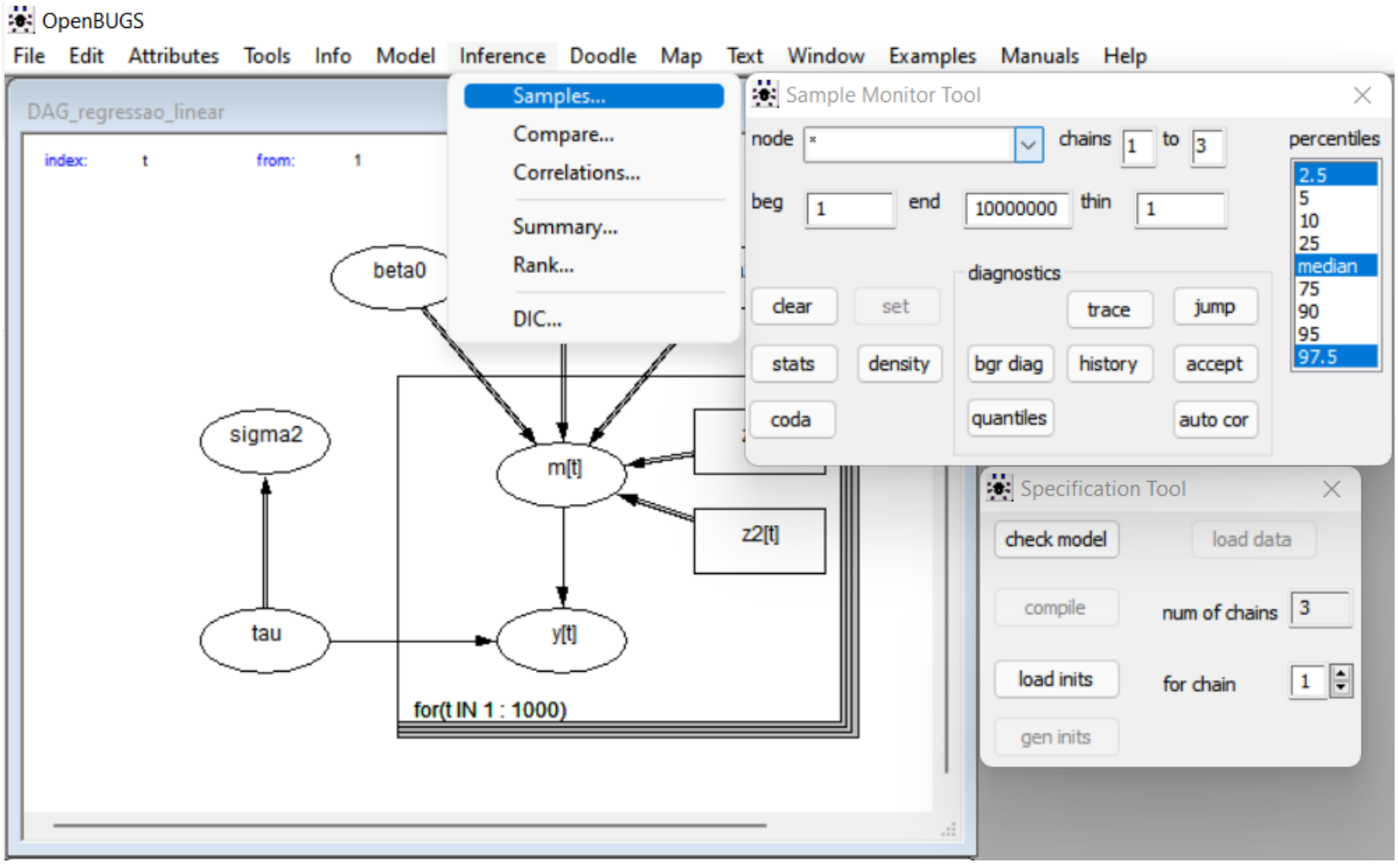

Figura 4: Opções para monitorar e analisar as quantidades de interesse.

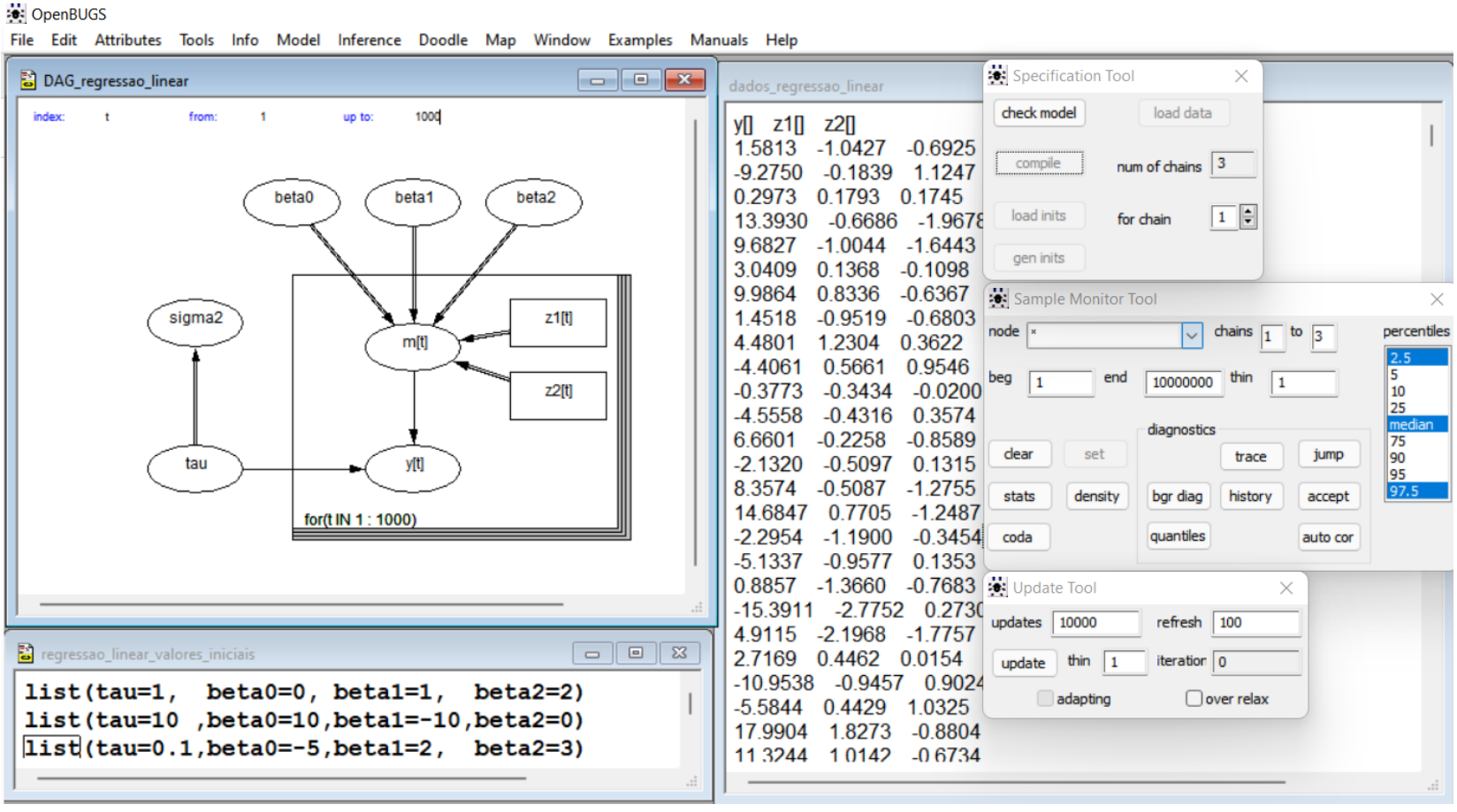

Figura 5: Visão geral das diversas janelas e funcionalidades básicas.

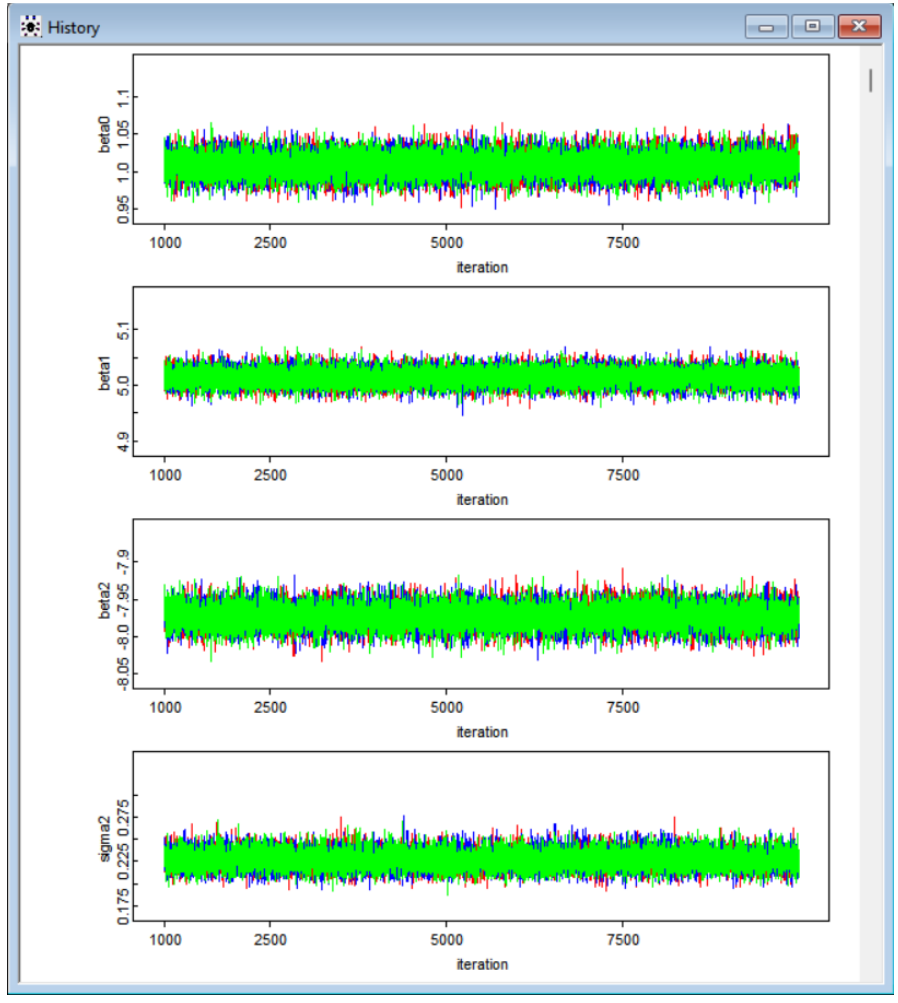

Figura 6: Cadeias geradas exibindo indícios de convergência - cada um dos parâmetros, $\{\beta_0, \beta_1, \beta_2, \sigma^2\}$, está sendo gerado em torno de um valor comum nas três cadeias (azul, verde e vermelha) e com variabilidade constante.

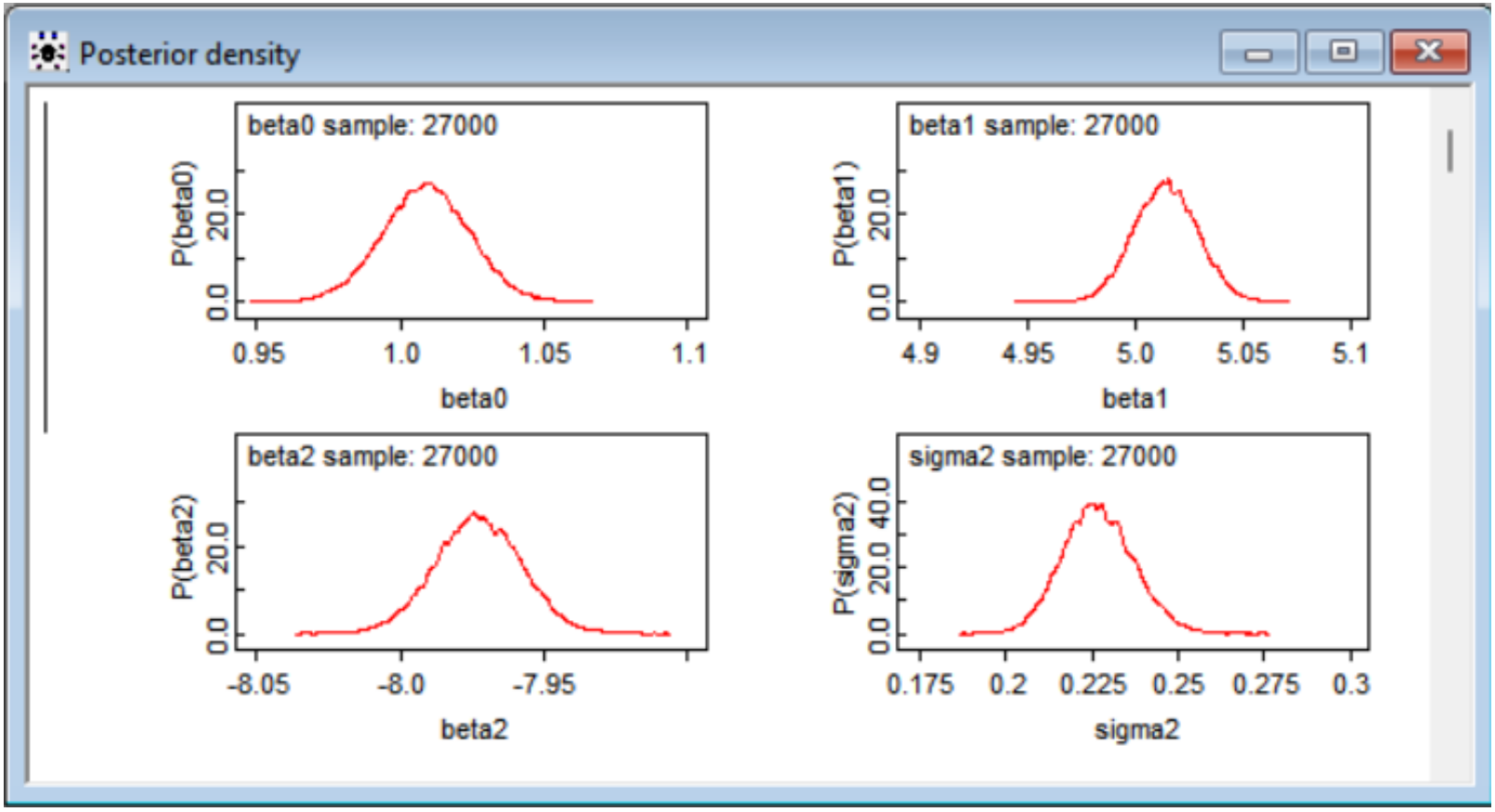

Figura 7: Distribuições marginais a posteriori estimadas.

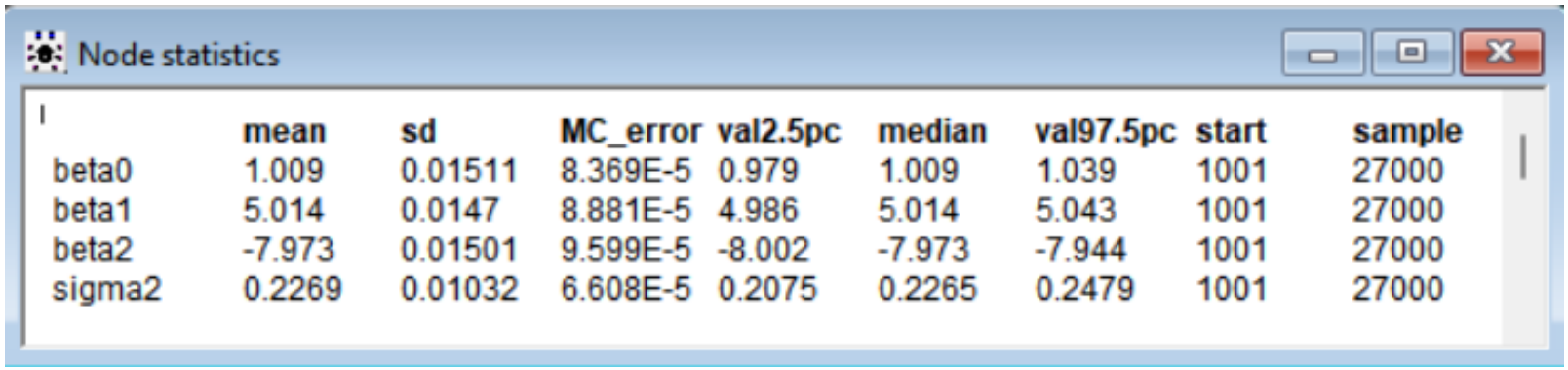

Figura 8: Resumo da distribuição a posteriori.

Tabela 1: Resumo da distribuição a posteriori.

Parâmetro	Média	D.P	2,5%	Mediana	97,5%
β_0	1,009	0,0151	0,979	1,009	1,039
β_1	5,014	0,0147	4,986	5,014	5,043
β_2	-7,973	0,0150	-8,002	-7,973	-7,944
σ^2	0,227	0,0102	0,208	0,227	0,248

Exemplo

Considere o modelo de **regressão logística** (útil em classificação) dado por

$$\begin{aligned} f(y_i|\boldsymbol{\beta}) &= \pi_i^{y_i}(1-\pi_i)^{1-y_i}, \quad y_i \in \{0,1\} \\ \text{logit}(\pi_i) &= \ln\left(\frac{\pi_i}{1-\pi_i}\right) \\ &= \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \cdots + \beta_d z_{id}, \quad i = 1, 2, \ldots, n. \end{aligned}$$

Para completar o modelo bayesiano, tomamos

$$\boldsymbol{\beta} \sim \mathcal{N}_d(\mathbf{0};\, 10^6 \boldsymbol{I}) \iff \beta_j \sim \mathcal{N}(0;\, 10^6), \text{ para } j = 1, 2, \ldots, d.$$

Note que poderíamos reescrever o modelo como

$$\begin{aligned} f(y_i = 1|\boldsymbol{\beta}) &= \frac{\exp\left(\beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \cdots + \beta_d z_{id}\right)}{1 + \exp\left(\beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \cdots + \beta_d z_{id}\right)} \\ f(y_i = 0|\boldsymbol{\beta}) &= \frac{1}{1 + \exp\left(\beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \cdots + \beta_d z_{id}\right)} \end{aligned}$$

para $i = 1, 2, \ldots, n$.

Modelo gráfico de regressão logística e previsão

No modelo gráfico abaixo, os vértices Y_i em vermelho representam os dados observados, $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^\top$ com $y_i \in \{0, 1\}$, para $i = 1, 2, \ldots, n$.

As análises são condicionais ao conhecimento de $\boldsymbol{z}_i = (1, z_{i1}, z_{i2}, \ldots, z_{id})^\top$, par $i = 1, 2, \ldots, n$. Estas quantidades são tratadas como constantes na representação gráfica. O mesmo vale para $\boldsymbol{z}_k^\star$, $k = 1, 2, \ldots, m$, utilizadas na previsão.

Os vértices $Y_i^\star$ representam dados não observados que desejamos prever ou classificar (condicional a $\boldsymbol{z}_k^\star$), para $k = 1, 2, \ldots, m$.

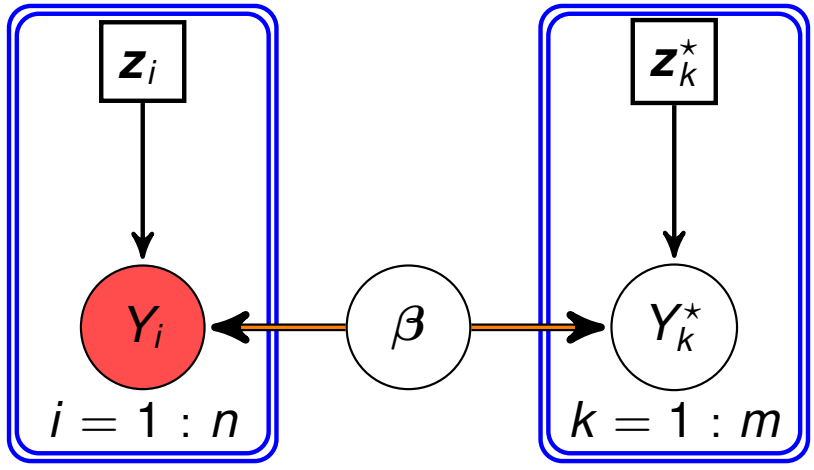

Figura 9: Modelo gráfico de regressão logística sob o enfoque bayesiano.