UNIVERSIDADE FEDERAL DO AMAZONAS
PRO REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
DEPARTAMENTO DE APOIO A PESQUISA
PROGRAMA INSTITUCIONAL DE INICIAÇÃO CIENTÍFICA

ESTIMAÇÃO BAYESIANA EM MODELOS DE REGRESSÃO ASSIMÉTRICOS

Bolsista: Vanessa Souza dos Santos, CNPq

MANAUS
2011

UNIVERSIDADE FEDERAL DO AMAZONAS
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
DEPARTAMENTO DE APOIO À PESQUISA
PROGRAMA INSTITUCIONAL DE INICIAÇÃO CIENTÍFICA

RELATÓRIO FINAL
PIB-E/0001/2010
ESTIMAÇÃO BAYESIANA EM MODELOS DE REGRESSÃO ASSIMÉTRICOS

Bolsista: Vanessa Souza dos Santos, CNPq
Orientador: Prof. Dr. Celso Rômulo Barbosa Cabral

MANAUS
2011

Resumo

Neste trabalho apresentam-se extensões do modelo de regressão linear clássico com erros de observação normais, substituindo-se a hipótese de normalidade por uma mais flexível, onde assume-se uma distribuição que estende a distribuição normal, incorporando assimetria e caudas pesadas. Para fazer inferência em torno dos parâmetros, utiliza-se a metodologia Bayesiana. Um exemplo com dados reais é apresentado, mostrando a superioridade da metodologia proposta.

Palavras chave. distribuição normal assimétrica; distribuição t de Student assimétrica; MCMC; modelos de regressão.

Conteúdo

1 Introdução

Neste trabalho apresentam-se extensões do modelo de regressão linear clássico com erros de observação normais, substituindo-se a hipótese de normalidade por uma mais flexível.

Suponhamos que para a i-ésima unidade amostral tenhamos valores fixados $x_{i1}, x_{i2}, \ldots, x_{i(p-1)}$ de $p-1$ variáveis regressoras ($p > 1$) e observamos o valor de uma variável resposta Y_i, $i = 1, \ldots, n$, $p < n$. A seguir temos a definição clássica do modelo de regressão linear.

Definição 1. *O Modelo de Regressão Linear Múltipla* é definido por observações independentes $Y_1, \ldots, Y_n$ tais que

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{(p-1)} x_{i(p-1)} + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \qquad \imath = 1, \ldots, n.$$

e $\sigma^2 > 0$.

A suposição de normalidade dos erros de observação ϵ_i pode não ser realizável na prática. Existem muitas situações onde temos claramente um comportamento de assimetria ou de observações discrepantes, que não é capturado pelo modelo normal. Isto ocorre, por exemplo, em dados da área econômica, em dados de sobrevivência e em confiabilidade de sistemas.

Apesar de vários autores terem apresentados extensões de forma implícita, o trabalho que é considerado como o marco inicial na pesquisa de distribuições assimétricas estendendo a normal é o de Azzalini (1985).

A proposta inicial deste trabalho é apresentar algumas extensões do modelo de regressão linear normal, substituindo a hipótese de normalidade pela de normalidade assimétrica. Depois, pretende-se também incorporar a possibilidade de modelar observações aberrantes, utilizando-se uma extensão assimétrica da distribuição t de Student ver, por exemplo, Branco & Dey (2001).

2 Revisão Bibliográfica

Nesta seção apresentaremos o modelo normal assimétrico e a inferência do mesmo no contexto Bayesiano.

2.1 Modelos Normal e t Assimétricos

Definição 2. *(Azzalini, 1985) Dizemos que uma variável aleatória Z tem distribuição normal assimétrica padrão com parâmtero de assimetria $\lambda \in \mathbb{R}$, denotado por $Z \sim SN(\lambda)$, se sua densidade é dada por*

$$g(z \mid \lambda) = 2\phi(z)\Phi(z)$$

em que $\phi(.)$ e $\Phi(.)$ são funções de densidade e a função de probablidade acumulada da normal simétrica padrão, respectivamente.

Definição 3. *Uma variável aleatória Y tem distribuição normal assimétrica com parâmetros de locação $\xi \in \mathbb{R}$, escala $\sigma > 0$ e assimetria $\lambda \in \mathbb{R}$, se Y é dada pela transformação*

$$Y = \xi + \sigma Z$$

onde $Z \sim SN(\lambda)$.

Dessa forma, a densidade de Y é dada por

$$f(Y \mid \xi, \sigma, \lambda) = \frac{2}{\sigma}\phi\left(\frac{y-\xi}{\sigma}\right)\Phi\left(\lambda\frac{y-\xi}{\sigma}\right). \tag{1}$$

A Figura 1 ilustra várias curvas da função densidade da distribuição normal assimétrica para os parâmetros $\xi = 50$, $\sigma = 10$, e $\lambda = \{-5, 0, 1, 3, 5\}$. Nota-se que as curvas tomam formas diferentes ao mudar o valor de λ. Para $\lambda = 0$ retorna para a curva da função densidade de uma normal simétrica.

O teorema a seguir é uma representação estocástica para uma variável aleatória com distribuição normal assimétrica, sendo muito útil para fazermos inferência e também para gerarmos amostras. A demonstração pode ser encontrada em Henze (1986).

Teorema 1. *Se T_0 e T_1 são variáveis aleatórias independentes com distribuição normal simétrica padrão e $\delta \in (-1, 1)$ e*

$$Y = \xi + \sigma\delta|T_0| + \sigma(1-\delta^2)T_1,$$

então $Y \sim SN(\xi, \sigma^2, \lambda)$, onde $\lambda = \delta/\sqrt{1-\delta^2}$.

Considerando a reparametrização

$$\omega^2 = \sigma^2(1-\delta^2), \quad \Delta = \sigma\delta, \quad \delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \quad \text{e} \quad \lambda = \frac{\Delta}{\omega},$$

temos o seguinte resultado, que é consequência imediata do anterior

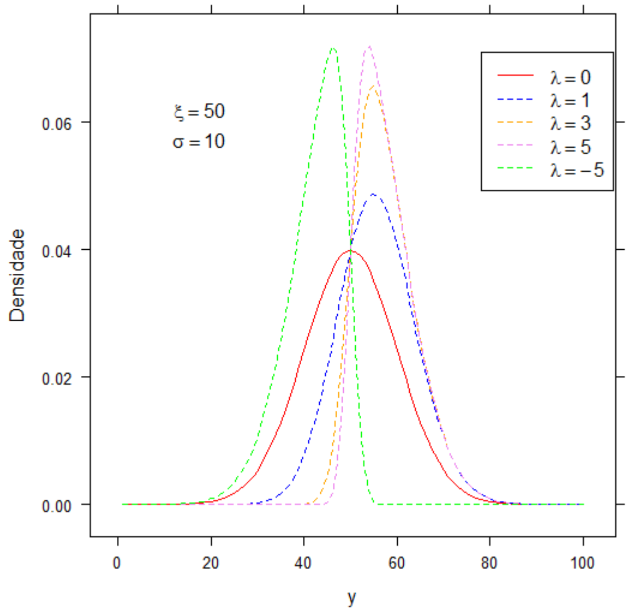

Figura 1: Densidade Normal Assimétrica

Teorema 2. *Se uma variável aleatória Y tem distribuição normal assimétrica com parâmetro de locação ξ, de assimetria Δ e de dispersão ω^2, então existe uma variável aleatória $T \sim HN(0;1)$, tal que*

$$Y \mid T = t \sim N(\xi + \Delta t, \omega^2).$$

A notação HN neste contexto significa distribuição Half Normal no intervalo $(0, \infty)$.

Agora será apresentada a versão assimétrica da distribuição t de Student que será usada neste trabalho. $\mathrm{G}(a, b)$ denota a distribuição Gamma com média a/b.

Definição 4. *Sejam $X \sim SN(0, \sigma^2, \lambda)$ e $S \sim G(\nu/2, \nu/2)$ variáveis aleatórias independentes. Seja $\xi \in \mathbb{R}$. Dizemos que a distribuição de $Y = \xi + S^{-1/2}X$ é t de Student assimétrica com parâmetro de locação ξ, parâmetro de escala σ, parâmetro de assimetria (ou de forma) λ e ν graus de liberdade.*

Denotamos por $Y \sim ST(\xi, \sigma^2, \lambda, \nu)$. A t assimétrica admite uma representação estocástica dada pelo seguinte teorema

Teorema 3. *Seja $Y \sim ST(\xi, \sigma^2, \lambda, \nu)$. Então Y admite a seguinte representação estocástica*

$$Y|T = t, S = s \sim N(\xi + \Delta t, s^{-1}\omega^2), \quad T|S = s \sim HN(0, s^{-1}), \quad S \sim G(\nu/2, \nu/2).$$

Os teoremas apresentados nesta seção serão fundamentais para obter o algoritmo de estimação dos parâmetros no modelo de regressão assimétrico.

2.2 Inferência Bayesiana

Na inferência Bayesiana, os parâmetros desconhecidos são considerados variáveis aleatórias, assumindo distribuições de probabilidade fixadas a priori. No modelo normal assimétrico, pretende-se estimar no máximo três parâmeros desconhecidos, que são representados por $\theta = (\xi, \omega^2, \Delta)$. Suponha uma suposição inicial (ou crença) em relação à distribuição de θ, $\pi(\theta)$, e seja $(Y_1, Y_2, \ldots, Y_n)$ uma amostra aleatória observada de uma variável aleatória Y com densidade $f(Y|\theta)$.

O objetivo é proceder inferência Bayesiana para o vetor de parâmetros $\theta \in \Theta$, onde Θ é o espaço paramétrico, ou seja

$$\pi(\theta|Y) = \frac{L(Y|\theta)\pi(\theta)}{\pi(Y)}$$

onde $\pi(\theta|Y), L(Y|\theta)$ e $\pi(\theta)$ que estão em função de θ definem a *posteriori* de θ, a função de máxima verossimilhança e a densidade da distribuição *priori* de θ, respectivamente. A densidade $\pi(Y)$, que não depende de θ, é a densidade marginal de Y e, portanto, a *posteriori* de θ é resumida por

$$\pi(\theta|Y) \propto \pi(Y|\theta)\pi(\theta).$$

Essa relação, mais conhecida como método de Bayes, nos resultará em uma distribuição para θ dado Y. A distribuição *priori* é obtida antes da observações dos dados, por meio de invertigações passadas, ou pela propria experiência do pesquisador.

Para o modelo normal assimétrico, considera-se como *prioris* independentes ξ, ω^2 e Δ, com distribuições $N(a, b^2), N(c, d^2)$ e $Gama(e, f)$, respectivamente. Então, a densidade conjunta para θ será

$$\pi(\theta) = \pi(\xi|a, b^2)\pi(\Delta|c, d^2)\pi(\psi|e, f)$$

em que os parâmetros *a, b, c, d* e e são fixos e $\psi = 1/\omega^2$ [1].

Dessa forma, para o modelo normal assimétrico (1), obtemos

$$\pi(\theta|Y) \propto \pi(\xi|a, b^2)\pi(\Delta|c, d^2)\pi(\psi|e, f) \prod_{i=1}^{n} SN(Y_i|\xi, \omega^2, \Delta)$$

Nota-se que a solução para encontrar a *posteriori* para o modelo normal assimétrico é mais complicado, por haver métodos algébricos com manipulações complexas.

[1] ψ é a precisão

3 Métodos Utilizados

3.1 Estimação via Algoritmo de Gibbs

Para um modelo normal assimétrico, quando queremos estimar cada parâmetro isolodamente, devemos encontrar uma distribuição marginal para estes parâmetros, por exemplo, para estimar ξ, podemos encontrar a integral da forma

$$E(\xi|Y) = \int \xi f(\xi|Y) d\xi$$

Note que a solução analítica da integral acima pode ser extremamente complicada. Em geral recorremos a integração de Monte Carlo, mais especificamente, aos algoritmos do tipo MCMC (Markov Chain Monte Carlo).O algoritmo MCMC mais popular é o algoritmo de Gibbs, que gera amostras da *posteriori* a partir das distribuições condicionais completas conhecidas, que será mostrado com mais detalhes ainda na próxima seção. Para gerar amostras independentes, devemos criar uma cadeia de Marcov estacionária, isto é a probabilidade de que no estágio i assumir um valor $\theta_{(i)}$, saberemos que no estágio *i+1* a probabilidade da cadeia assumir $\theta_{(i+1)}$ no mesmo espaço paramétrico.

3.2 Os Modelos de Regressão com Erros Normais e t Assimétricos

A extensão do modelo de regressão será baseada na representação estocástica no Teorema 2.

Definição 5. *O* Modelo de Regressão Linear Normal Assimétrico *é definido por*

$$Y_i|\boldsymbol{\beta}, \omega^2, \Delta, T_i = t_i \sim N(\xi_i + \Delta t_i, \omega^2), \quad T_i \sim HN(0,1), \quad i = 1, \ldots, n,$$

onde Y_i são variáveis aleatórias que representam respostas a serem observadas em um indivíduo i, $\xi_i = \mathbf{x}_i'\boldsymbol{\beta}$, $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{i,(p-1)})'$ é um vetor com valores de $p-1$ regressores para o indivíduo i, ω^2 e Δ são parâmetros desconhecidos regulando dispersão e assimetria, respectivamente, e $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})'$ é um vetor de parâmetros desconhecidos. Além disso, $Y_1, \ldots, Y_n$ são condicionalmente independentes dados os valores dos parâmetros, variáveis latentes e regressores.

É claro que, marginalmente, temos que $Y_i \sim SN(\mathbf{x}_i'\boldsymbol{\beta}, \sigma^2, \alpha)$, $i = 1, \ldots, n$, onde $\alpha = \Delta/\omega$ e $\sigma^2 = \omega^2 + \Delta^2$. Assim, utilizando a definição 3,

$$Y_i = \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i, \quad \epsilon_i \sim SN(0, \sigma^2, \alpha), \quad i = 1, \ldots, n,$$

que é uma extensão natural do modelo de regressão linear clássico.

Para fazermos estimação dos parâmetros no modelo, precisamos fazer a especificação das distribuições a priori. Vamos assumir que

$$\boldsymbol{\beta} \sim N_p(\mathbf{a}, \mathbf{A}), \quad \Delta \sim N(b, c^2), \quad \psi = 1/\omega^2 \sim Gamma(d, e), \tag{2}$$

onde $\mathbf{a} : p \times 1$, $\mathbf{A} : p \times p$ (positiva definida), $b \in \mathbb{R}$, $c > 0$, $d > 0$ e $e > 0$ são todos fixados. Como anteriormente, assume-se que os parâmetros são a priori independentes. O procedimento álgébrico para obter um algoritmo do tipo Gibbs para gerar amostras da distribuição posterior é tão simples como antes, com pequenas diferenças no que se refere à distribuição condicional completa para $\boldsymbol{\beta}$. Vamos obtê-la então.

Primeiramente, defina $\mathbf{y} = (y_1, \ldots, y_n)'$, $\mathbf{t} = (t_1, \ldots, t_n)'$. Além disso, definamos a chamada *matriz de planejamento*, denotada por $\mathbf{X}$, em analogia ao que é feito na teoria de regressão tradicional, ou seja, a i−ésima linha de $\mathbf{X}$ é $\mathbf{x}_i'$, $i = 1, \ldots, n$.

Assim,

$$\begin{aligned}
\pi(\boldsymbol{\beta}|\mathbf{y}, \mathbf{t}, \Delta, \psi) &\propto \pi(\mathbf{y}|\boldsymbol{\beta}, \mathbf{t}, \Delta, \psi)\pi(\boldsymbol{\beta}) \\
&\propto \exp\left\{-\frac{\psi}{2}\sum_{i=1}^{n}(y_i - \mathbf{x}_i'\boldsymbol{\beta} - \Delta t_i)^2\right\}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \mathbf{a})'\mathbf{A}^{-1}(\boldsymbol{\beta} - \mathbf{a})\right\} \\
&= \exp\left\{-\frac{\psi}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \Delta\mathbf{t})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \Delta\mathbf{t})\right\}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \mathbf{a})'\mathbf{A}^{-1}(\boldsymbol{\beta} - \mathbf{a})\right\} \\
&\propto \exp\left\{-\frac{\psi}{2}[-2\boldsymbol{\beta}'\mathbf{X}'(\mathbf{y} - \Delta\mathbf{t}) + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}]\right\}\exp\left\{-\frac{1}{2}[-2\boldsymbol{\beta}'\mathbf{A}^{-1}\mathbf{a} + \boldsymbol{\beta}'\mathbf{A}^{-1}\boldsymbol{\beta}]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[-2\boldsymbol{\beta}'(\psi\mathbf{X}'(\mathbf{y} - \Delta\mathbf{t}) + \mathbf{A}^{-1}\mathbf{a}) + \boldsymbol{\beta}'(\psi\mathbf{X}'\mathbf{X} + \mathbf{A}^{-1})\boldsymbol{\beta}\right]\right\}
\end{aligned}$$

O Algoritmo de Gibbs para gerar a posteriori para cada parâmetro, são obtidas através dos seguintes passos:

Passo 1. Para cada $i = 1, \ldots, n$, gere t_i independentemente da distribuição $\pi(t_i|\boldsymbol{\beta}, \Delta, \psi, \mathbf{t}, \mathbf{y})$, que é igual a seguinte distribuição

$$HN\left(\frac{\Delta(y_i - \mathbf{x}_i'\boldsymbol{\beta})}{\Delta^2 + \psi^{-1}}, \frac{1}{\psi\Delta^2 + 1}\right).$$

Passo 2. Gere amostras de $\boldsymbol{\beta}$ a partir de da distribuição condicional de $\boldsymbol{\beta}|\mathbf{y}, \mathbf{t}, \Delta, \psi$, que é $N_p(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$, onde

$$\boldsymbol{\Sigma}_\beta = (\psi\mathbf{X}'\mathbf{X} + \mathbf{A}^{-1})^{-1} \ \text{ e } \ \boldsymbol{\mu}_\beta = \boldsymbol{\Sigma}_{\boldsymbol{\beta}}\left[\psi\mathbf{X}'(\mathbf{y} - \Delta\mathbf{t}) + \mathbf{A}^{-1}\mathbf{a}\right].$$

Passo 3. Obtenha uma amostra de Δ a partir da distribuição de $\pi(\Delta|\boldsymbol{\beta}, \psi, \mathbf{t}, \mathbf{y})$, que é igual a

$$N\left(\frac{c^2\psi\sum_{i=1}^{n}(y_i t_i - \mathbf{x}_i'\boldsymbol{\beta}t_i) + b}{c^2\psi\sum_{i=1}^{n}t_i^2 + 1}, \frac{c^2}{c^2\psi\sum_{i=1}^{n}t_i^2 + 1}\right);$$

Passo 4. Obtenha uma amostra de ψ a partir da distribuição de $\pi(\psi|\boldsymbol{\beta}, \Delta, \mathbf{t}, \mathbf{y})$, que é igual a

$$\text{Gamma}\left(\frac{n}{2} + d, \frac{1}{2}\sum_{i=1}^{n}(y_i - \mathbf{x}_i'\boldsymbol{\beta} - \Delta t_i)^2 + e\right);$$

A extensão do modelo de regressão para o caso em que os erros de observação seguem uma distribuição t-assimétrica é imediata, baseada na própria definição da distribuição t-assimétrica e na definição do modelo de regressão normal assimétrica que apresentamos acima. Novamente usamos a parametrização $\Delta = \sigma\delta$ e $\omega^2 = \sigma^2(1-\delta^2)$, onde $\delta = \alpha/\sqrt{1 + \alpha^2}$.

Definição 6. *O* Modelo de Regressão Linear t de Student Assimétrico *é definido por*

$$\begin{aligned} &Y_i|\boldsymbol{\beta}, \omega^2, \Delta, T_i = t_i, S_i = s_i \sim N(\xi_i + \Delta t_i, s_i^{-1}\omega^2), \;\; T_i|S_i = s_i \sim HN(0, s_i^{-1}), \\ &S \sim Gamma(\nu/2, \nu/2), \end{aligned}$$

onde Y_i*,* $i = 1, \ldots, n$*, são variáveis aleatórias que representam respostas a serem observadas em um indivíduo i,* $\xi_i = \mathbf{x}_i'\boldsymbol{\beta}$*,* $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{i,(p-1)})'$ *é um vetor com valores de* $p - 1$ *regressores para o indivíduo i,* ω^2 *e* Δ *são parâmetros desconhecidos regulando dispersão e assimetria, respectivamente,* $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})'$ *é um vetor de parâmetros desconhecidos e* ν *é conhecido. Além disso,* $Y_1, \ldots, Y_n$ *são condicionalmente independentes dados os valores dos parâmetros, variáveis latentes e regressores.*

Observe que, analogamente ao que acontece no caso normal assimétrico, temos que $Y_i \sim ST(\mathbf{x}_i'\boldsymbol{\beta}, \sigma^2, \alpha, \nu)$ onde $\alpha = \Delta/\omega$ e $\sigma^2 = \omega^2 + \Delta^2$. Assim, vem que

$$Y_i = \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i, \quad i = 1, \ldots, n,$$

onde $\epsilon_i \sim ST(0, \sigma^2, \alpha, \nu)$. Assim, esta é uma extensão natural do modelo de regressão linear incorporando erros assimétricos e com caudas pesadas.

A especificação das distribuições a priori pode seguir a mesma linha que utilizamos para o caso normal assimétrico. Para obter a distribuição condicional completa, vamos definir $\mathbf{s} = (s_1, \ldots, s_n)'$. Além disso, vamos considerar $\mathbf{M} = \text{diag}(\mathbf{s})$, a matriz diagonal onde a diaginal principal é preenchida pelo vetor $\mathbf{s}$. Assim,

$$\begin{aligned} \pi(\boldsymbol{\beta}|\mathbf{y}, \mathbf{t}, \Delta, \psi) &\propto \pi(\mathbf{y}|\boldsymbol{\beta}, \mathbf{t}, \Delta, \psi)\pi(\boldsymbol{\beta}) \\ &\propto \exp\left\{-\frac{\psi}{2}\sum_{i=1}^{n} s_i(y_i - \mathbf{x}_i'\boldsymbol{\beta} - \Delta t_i)^2\right\}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \mathbf{a})'\mathbf{A}^{-1}(\boldsymbol{\beta} - \mathbf{a})\right\} \\ &= \exp\left\{-\frac{\psi}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \Delta\mathbf{t})'\mathbf{M}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \Delta\mathbf{t})\right\}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \mathbf{a})'\mathbf{A}^{-1}(\boldsymbol{\beta} - \mathbf{a})\right\} \\ &\propto \exp\left\{-\frac{\psi}{2}[-2\boldsymbol{\beta}'\mathbf{X}'\mathbf{M}(\mathbf{y} - \Delta\mathbf{t}) + \boldsymbol{\beta}'\mathbf{X}'\mathbf{M}\mathbf{X}\boldsymbol{\beta}]\right\}\exp\left\{-\frac{1}{2}[-2\boldsymbol{\beta}'\mathbf{A}^{-1}\mathbf{a} + \boldsymbol{\beta}'\mathbf{A}^{-1}\boldsymbol{\beta}]\right\} \\ &= \exp\left\{-\frac{1}{2}\left[-2\boldsymbol{\beta}'(\psi\mathbf{X}'\mathbf{M}(\mathbf{y} - \Delta\mathbf{t}) + \mathbf{A}^{-1}\mathbf{a}) + \boldsymbol{\beta}'(\psi\mathbf{X}'\mathbf{M}\mathbf{X} + \mathbf{A}^{-1})\boldsymbol{\beta}\right]\right\} \end{aligned}$$

Para um modelo de regressão t-assimétrico, o algoritmo para gerar a posteriori é:
Passo 1. Para cada $i = 1, \ldots, n$, gere t_i independentemente da distribuição $\pi(t_i|\boldsymbol{\beta}, \Delta, \psi, \mathbf{s}, \mathbf{y})$, que é igual a seguinte distribuição

$$HN\left(\frac{(y_i - \mathbf{x}_i'\boldsymbol{\beta})\Delta\psi}{1 + \Delta^2\psi}, \frac{1}{s_i(1 + \Delta^2\psi)}\right);$$

Passo 2. Para cada $i = 1, \ldots, n$, gere s_i independentemente da distribuição $\pi(s_i|\boldsymbol{\beta}, \Delta, \psi, \mathbf{t}, \mathbf{y})$, que é igual a seguinte distribuição

$$\text{Gamma}\left(\frac{\nu + 2}{2}, \frac{1}{2}\left\{\psi[y_i - (\mathbf{x}_i'\boldsymbol{\beta} + \Delta t_i)]^2 + t_i^2 + \nu\right\}\right);$$

Passo 3. Obtenha uma amostra de $\boldsymbol{\beta}$ a partir da distribuição de $\pi(\boldsymbol{\beta}|\Delta, \psi, \mathbf{s}, \mathbf{t}, \mathbf{y})$, que é igual a

$$N_p\left(\boldsymbol{\Sigma}_\beta\left[\psi\mathbf{X}'d(\mathbf{S})(\mathbf{y}-\Delta\mathbf{t})+\mathbf{A}^{-1}\boldsymbol{a}\right],(\psi\mathbf{X}'d(\mathbf{S})\mathbf{X}+\mathbf{A}^{-1})^{-1}\right)$$

Passo 4. Obtenha uma amostra de Δ a partir da distribuição de $\pi(\Delta|\boldsymbol{\beta}, \psi, \mathbf{s}, \mathbf{t}, \mathbf{y})$, que é igual a

$$N\left(\frac{\psi c^2\sum_{i=1}^n s_i(y_i t_i-\mathbf{x}_i'\boldsymbol{\beta}t_i)+b}{\psi c^2\sum_{i=1}^n s_i t_i^2+1},\frac{c^2}{\psi c^2\sum_{i=1}^n s_i t_i^2+1}\right)$$

Passo 5. Obtenha uma amostra de ψ a partir da distribuição de $\pi(\psi|\boldsymbol{\beta}, \Delta, \mathbf{s}, \mathbf{t}, \mathbf{y})$, que é igual a

$$\text{Gamma}\left(\frac{n+2d}{2},\frac{1}{2}\sum_{i=1}^n s_i(y_i-(\mathbf{x}_i'\boldsymbol{\beta}+\Delta t_i))^2+e\right).$$

4 Resultados e Discussões

4.1 Apresentação dos Dados

Como aplicação da metodologia proposta aqui, utilizaremos um conjunto de dados citado no texto de Albert (2009). Todos os procedimentos computacionais que aparecem a seguir foram realizados utilizando o software R (R Development Core Team, 2011).

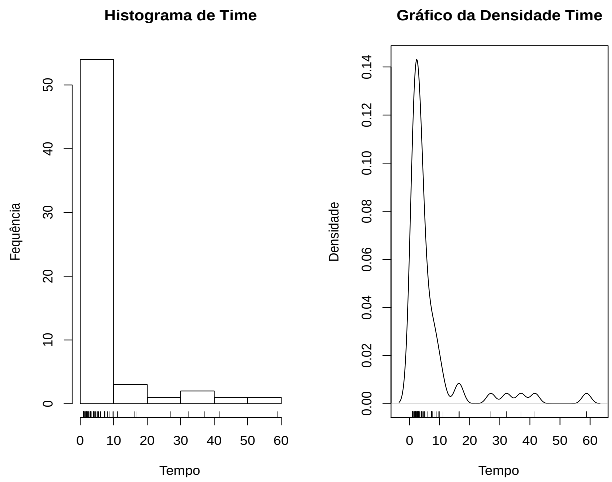

Figura 2: Histograma e densidade de Time.

Este conjunto de dados consiste em medidas feitas em pares de espécies coletadas em 16 ilhas próximas à Grã-Bretanha ao longo de várias décadas. Para cada espécie, o conjunto de dados contém as variáveis:

1. *Time* - tempo médio de extinção na ilha onde a espécie apareceu;

2. *Nesting* - número médio de pares aninhados;

3. *Size* - tamanho da espécie (grande ou pequeno);

4. *Status* - status migratório da espécie (migrantes ou residentes).

O objetivo é estudar o efeito das covariáveis *Nesting*, *Size* e *Status* na variação no tempo de extinção. Espera-se que as espécies com grande número de pares aninhados tenderão a permanecer mais tempo antes de se tornarem extintas.

4.2 Ajuste para o Modelo Normal

O modelo a ser ajustado é o seguinte:

$$time_i = \beta_0 + \beta_1 nesting_i + \beta_2 size_i + \beta_3 status_i + \epsilon_i,$$

onde
β_0 é o intercepto;
β_1 é o efeito do número de pares aninhados no período de tempo antes da extinção da espécie;
β_2 é o efeito do tamanho da espécie no período de tempo antes da extinção da espécie;
β_3 é o efeito do status migratótrio da espécie no período de tempo antes da extinção da espécie;
$\epsilon_i \sim N(0, \sigma^2)$ é o erro de observação.

Tabela 1: Estimativas para o modelo normal

Coeficientes	Estimativas	Erro	Estatística t	p-valor
β_0	-0,6276	2,9115	-0,216	0,83008
β_1	2,1337	0,5173	4,125	0,00012 *
β_2	-3,9275	2,3435	-1,676	0,09914
β_3	3,5298	2,568	1,375	0,17456

A Tabela 1 mostra as estimativas de máxima verossimilhança para os parâmetros de regressão linear normal. O ajuste é feito utilizando a função `lm` do R. Nota-se que o efeito do número médio de pares aninhados (*nesting*) é claramente significativo, com p-valor muito pequeno. As outras covariáveis *size* e *status* não apresentaram significância para o modelo.

A relação entre *time* e *nesting* está apresentada na Figura 3, onde não aparenta haver falta de linearidade. A Figura 2 mostra com mais clareza um comportamento assimétrico para a variável resposta *Time*. Além disso, especificamente na curva da densidade, temos caudas pesadas, o que se leva a sugerir um ajuste do modelo t assimétrico para esse dados. As Figuras 4 e 5 também confirmam que há evidências de não normalidade para os erros.

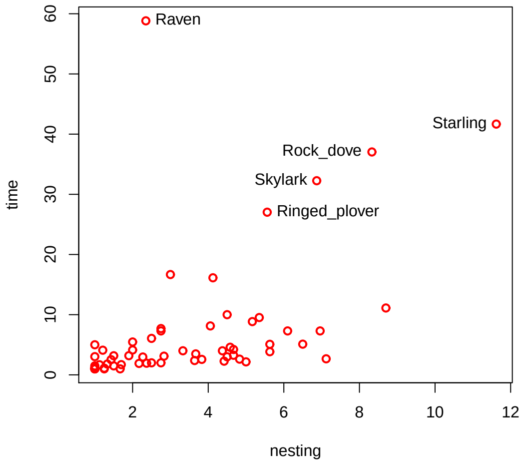

Figura 3: Gráfico de dispersão *Nesting* × *Time*.

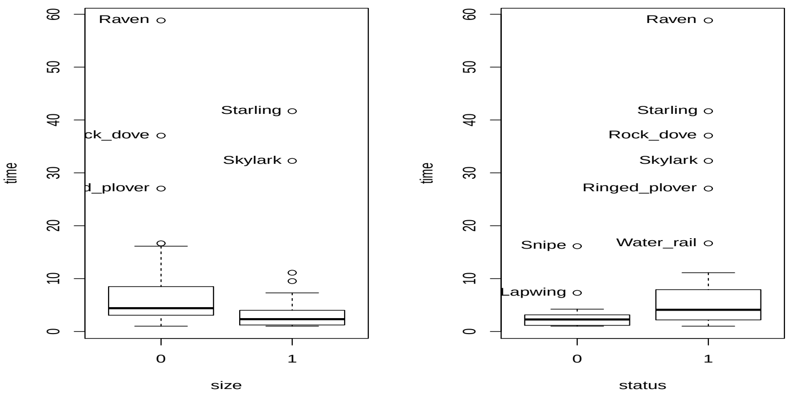

Figura 4: Box-plot size e status.

4.3 Ajuste para os Modelos Normal Assimétrico e t-Assimétrico

O Método de estimação por máxima verossimilhança para o modelo normal assimétrico é ultilizado para obter as estimativas apresentadas na Tabela 2. Foi utilizada a função `msn.fit`, que faz parte do pacote `sn` do R (Azzalini, 2010). O pacote fornece opções bem conhecidas de algoritmos para maximização da verossimilhança, como *BFGS*, $Nelder-Mead$ e *Nlminb*. No entanto, só há convergência no primeiro caso. Porém, analisando a suposição de que os erros seguem distribuição normal assimétrica, as Figuras 6 e 8 mostram que o modelo não está adequado.

Agora, usando o algoritmo de *BFGS*, a Figura 9 mostra que o modelo t-assimétrico está mais adequado. As estimativas dos parâmetros são mostradas na Tabela 3.

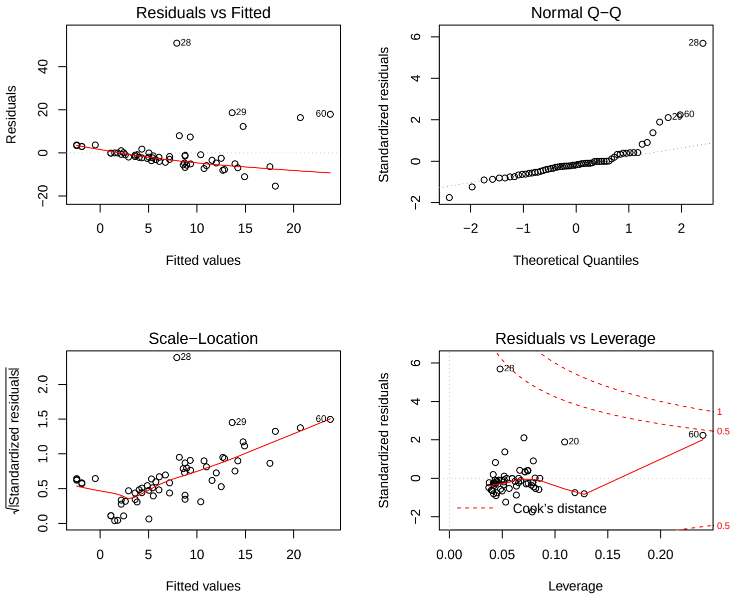

Figura 5: Análise de Resíduos.

Tabela 2: Estimativas para o Modelo Normal Assimétrico

Parâmetros	Nelder-Mead	BFGS	nlminb
β_0	0,0739	0,3283	0,3870
β_1	0,2940	0,3105	0,3093
β_2	-0,3712	0,2172	-0,2262
β_3	0,3514	0,0828	0,0774
σ^2	138,5660	135,0096	134,4638
λ	38,0969	574,2210	592686,3
Convergência	Não	Sim	Não

Tabela 3: Estimativas para o Modelo t-Assímetrico

Parâmetros	Nelder-Mead	BFGS	nlminb
β_0	1,8733	0,0989	0,6955
β_1	0,4971	0,5109	0,4818
β_2	-1,4358	-0,2020	-0,5918
β_3	0,4253	0,5008	0,34718
σ^2	0,2965	1,1297	0,8118
λ	-0,0366	6,6235	3,4756
ν	0,6598	0,7565	0,7241
Convergência	Não	Sim	Sim

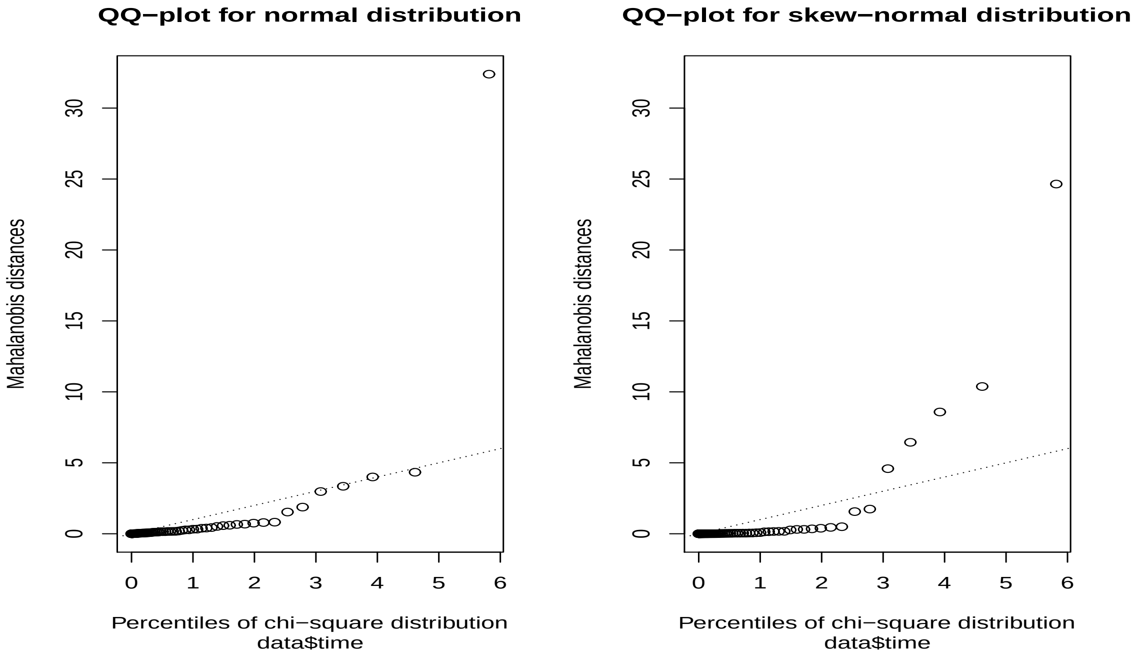

Figura 6: Gráfico Comparativo: QQ-Plot modelos Normal e Normal Assimétrico .

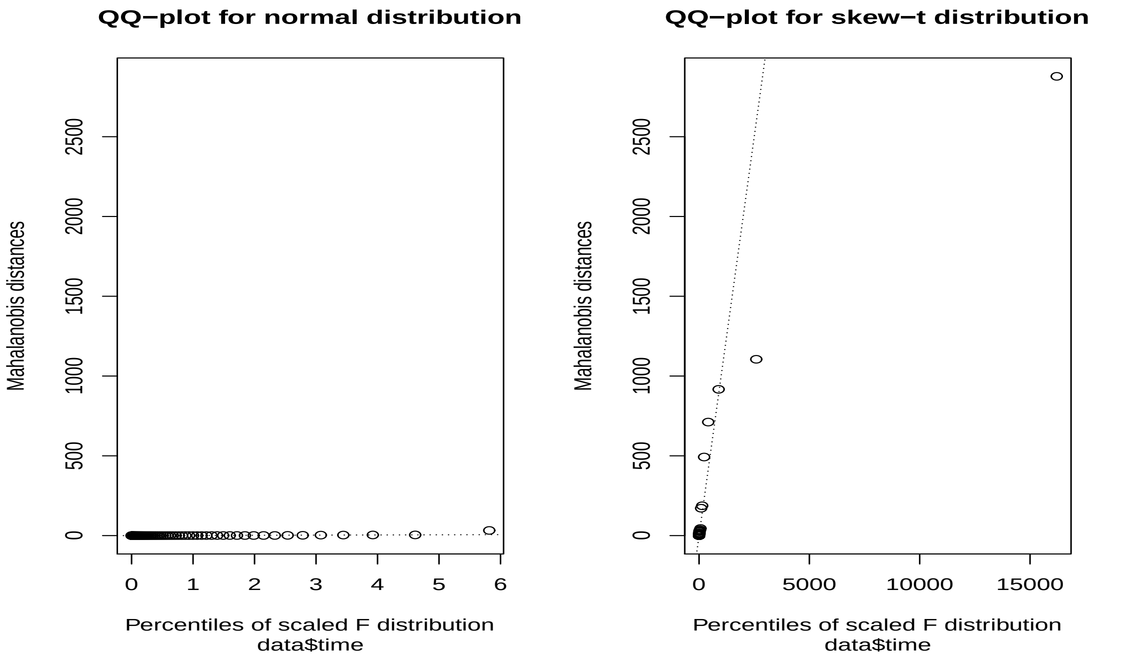

Figura 7: Gráfico Comparativo: QQ-Plot modelos Normal e t-Assimétrico .

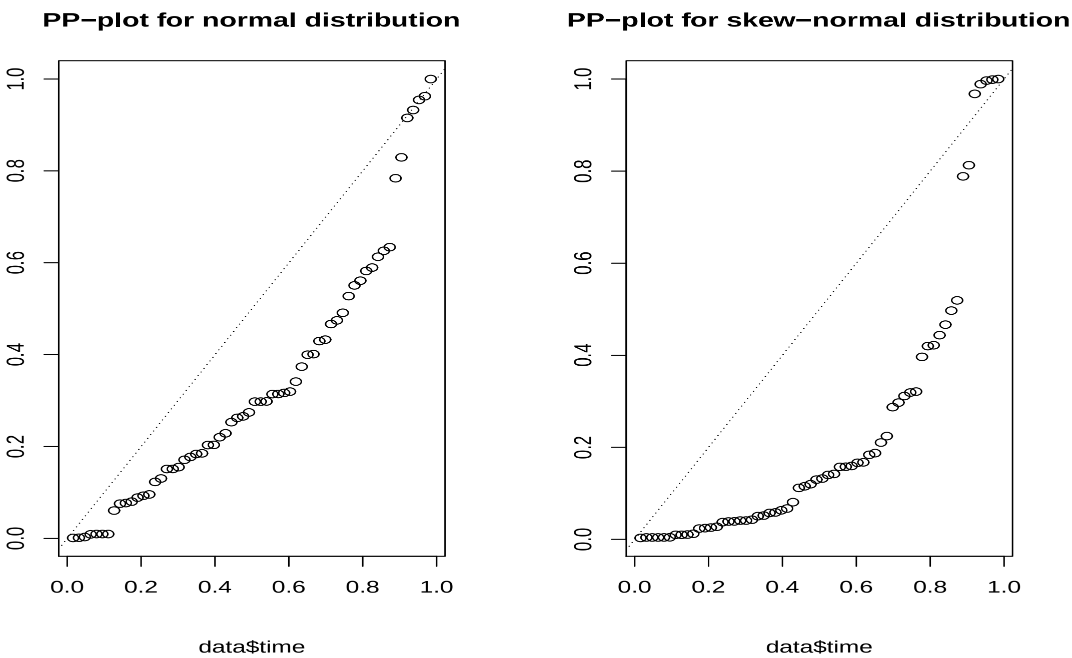

Figura 8: Gráfico Comparativo: PP-Plot modelos Normal e Normal Assimétrico .

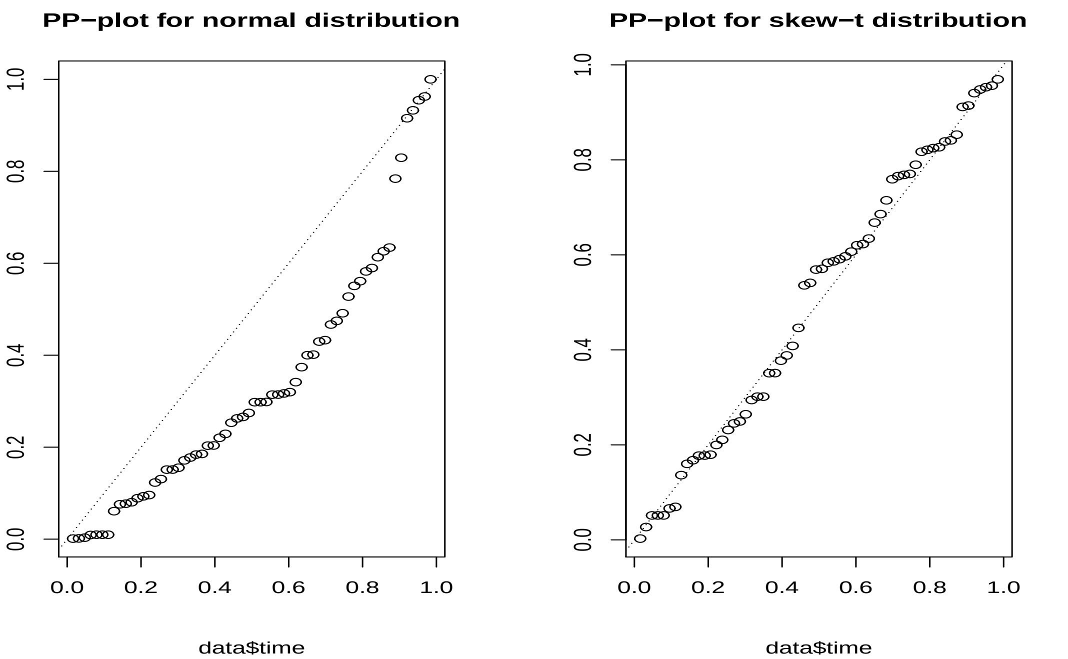

Figura 9: Gráfico Comparativo: PP-Plot modelos Normal e t-Assimétrico .

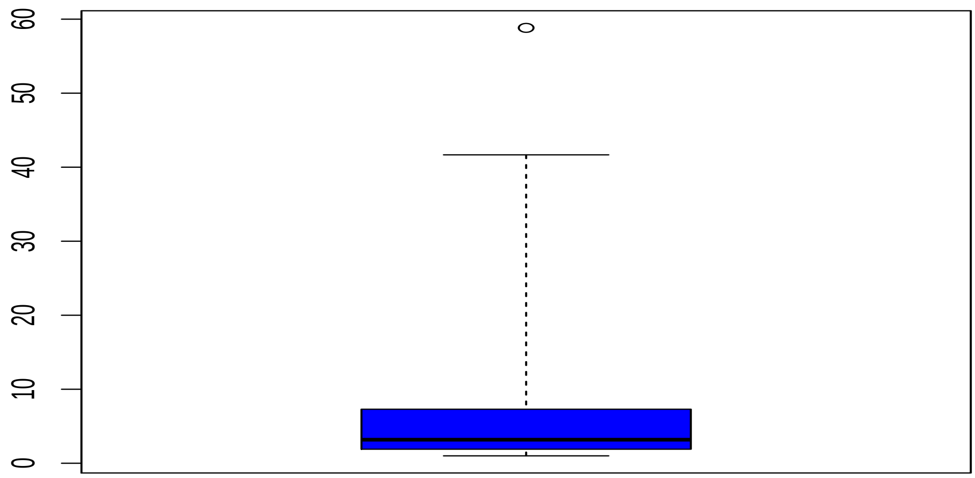

Figura 10: Box Plot - Modelo Normal Assimétrico .

A Figura 10 ilustra um gráfico mais adequado para o conjunto de dados, ao contrário da Figura 4, onde o gráfico do box plot foi construído por métodos convecionais para um modelo de regressão normal, apresentando vários pontos fora dos limites.

4.4 Ajuste para os Modelos Bayesianos Normal, Normal Assimétricos e t-Assimétricos

Como vimos no exemplo acima, a modelagem por máxima verossimilhança pode apresentar muitos problemas como, por exemplo, a existência de múltiplas modas e singularidades. Além disso, pode ser muito complicado obtermos o erro padrão para as estimativas dos parâmetros, já que estes são calculados em função da matriz de informação de Fisher, o que pode exigir amostras muito grandes. Neste sentido, o método Bayesiano mostra-se uma alternativa bastante acaeitável.

Para procedermos a análise do mesmo conjunto de dados utilizando metodologia Bayesiana, foi utilizado o software WinBugs, com interface no software R, para gerar 31000 amostras de Gibbs da densidade a posteriori conjunta, como encontrados na seção 3.2, e destas foram descartadas as 1000 primeiras amostras *("burn-in samples")* com a finalidade de eliminar o efeito dos valores iniciais usados no algoritmo de simulação. Além disso, foram consideradas as iterações $5^a, 10^a, 15^a, \ldots,$ resultando uma amostra final de 6000 amostras para cada parâmetro. Foram usados como valores iniciais para parâmetro, valores próximos às estimativas dos parâmetros encontrados na Tabela 3, de acordo com o algoritmo *BFGS*, sendo β_1 = 0,5, β_2 = -0,2, β_3 = 0,5, $\sigma^2 = 1$ e $\lambda = 4$. O número de graus de liberdade foi fixado em $\nu = 3$.

Tabela 4: Estimativas para o modelos Normal e Normal, t assimétrico

Parâmetros	Normal Assimétrico	t-Assimétrico	Normal
β_1	0,29037	0,3954	2,07072
β_2	-0,8095	-0,3759	-3,9366
β_3	0,1343	0,5123	3,0435
λ	71,9917	17,8047	-
σ^2	142,4572	11,6722	82,89588
DIC	403,6805	334,9306	454,5675
PD	2,7394	2,1596	3,8248

Para comparar os três modelos ajustados, foram obtidos os respectivos valores do Deviance Information Criterion (DIC), ver Celeux *et al.* (2006). O modelo t-Assimétrico obteve um menor DIC comparado aos outros modelos, isto é, significa que este modelo proporciona um melhor ajuste aos dados, conforme mostrado na Tabela 4. A Figura 11 mostra traceplots para as amostras MCMC geradas para cada parâmetro do modelo t-assimétrico, para as 6000 interações.

Podemos observar que todos os intervalos de credibilidade para os β's na Tabela 5 contem o zero. Portanto, nenhuma das covariáveis em consideração pode ser tomada como influente para explicar a variabilidade do tempo de extinção. Além disso, vemos que há uma forte componente assimétrica nestes dados, em função do intervalo observado para λ.

Tabela 5: Intervalo de Credibilidade (95%)

Parâmetros	Inferior	Superior
β_1	-0,5364	0,6807
β_2	-0,5945	0,6324
β_3	-0,5716	0,5421
σ^2	4,4612	20,3230
λ	11,0728	22,9451

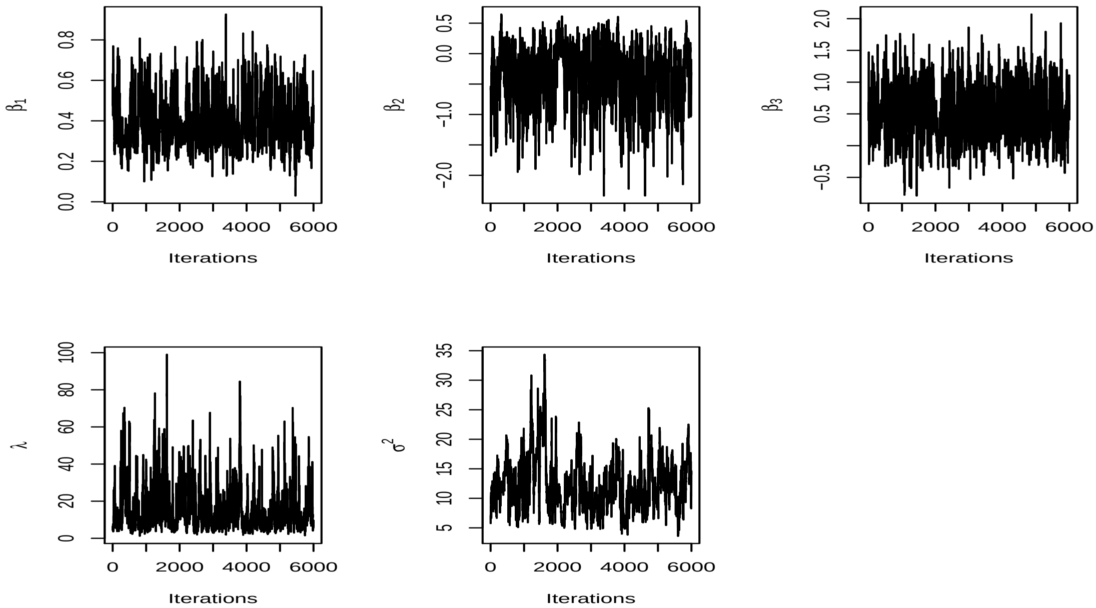

Figura 11: traceplot via Algoritmo do tipo Gibbs para modelo t-assimétrico .

5 Conclusões

Este relatório teve por objetivo propor um modelo de regressão, onde os erros seguem distribuições assimétricas, fazendo comparações entre os modelos de regressão normal, normal assimétrico e t assimétrico sob o ponto de vista Bayesiano e Frequentista, de Azzalini (1985).

O esforço principal desse trabalho foi dado no processo de estimação Bayesiana dos parâmetros para os modelos normal, normal assimétrico e t assimétrico via algorítmo de Gibbs, através de condicionais completas fechadas. A preocupação que se teve foi em propor um algorítmo simples de ser implementado em qualquer ambiente de programação e bons aspectos computacionais, podendo ser implementada sem maiores dificuldades com o software R. Portanto, através de toda a metodologia proposta nesse relatório, o modelo mais adequado para o conjunto de dados sobre a extinção de espécies de passáros, foi o modelo de regressão t-assimétrico.

6 Apêndice

Comandos para os Modelos Frequentistas

```
### DADOS ###
data<-read.table("C:\\Users\\VANESSA\\Documents\\PROJETO\\RELATORIO_PIBIC\\birdextinct.r" ,h=T)
############ MODELO FREQUENTISTA ############################################################
size<-factor(data$size)
status<-factor(data$status)
#---dispersão
plot(data$nesting,data$time, col="red", lwd=2,lty=2)
identify(data$nesting,data$time, label=data$species,n=5)

#--- ajuste para o modelo de regressão normal
data.mod<-lm(data$time ~ data$nesting+ data$size+ data$status, data= data,x=TRUE, y= TRUE)
summary(data.mod)

par(mfrow=c(1,2))
hist(time, ylab="Fequência", main="Histograma de Time", xlab="Tempo")
rug(time)
plot(density(time), ylab="Densidade", xlab="Tempo", main="Gráfico da Densidade Time")
rug(time)

#--box plot
par(mfrow=c(1,2))
plot(data$time~data$size, xlab="size", ylab="time")
identify(data$time~data$size, label=data$species)
plot(data$time~data$status, xlab="status", ylab="time")
identify(data$time~data$status, label=data$species)

par(mfrow=c(2,2))
plot(data.mod)

## MODELO FREQUENTISTA ST ##
library(sn)
library(msm)
mod1=mst.fit(X = cbind(rep(1,62),data$nesting, data$size, data$status), y = data$time,
                                              plot.it = T,algorithm = c("Nelder-Mead"))
mod2=mst.fit(X = cbind(rep(1,62),data$nesting, data$size, data$status), y = data$time,
                                                     plot.it = T,algorithm = c("BFGS"))
mod3=mst.fit(X = cbind(rep(1,62),data$nesting, data$size, data$status), y = data$time,
                                                                        plot.it = F)
## MODELO FREQUENTISTA ST ##
modn1=msn.fit(X = cbind(rep(1,62),data$nesting, data$size, data$status), y = data$time,
                                                  plot.it = T,algorithm = c("Nelder-Mead"))
modn2=msn.fit(X = cbind(rep(1,62),data$nesting, data$size, data$status), y = data$time,
                                                     plot.it = T,algorithm = c("BFGS"))
modn3=msn.fit(X = cbind(rep(1,62),data$nesting, data$size, data$status), y = data$time,
                                                                         plot.it = T)
```

```
}
\normalsize{
\subsection*{Algoritmo para as Condicionais Completas - Modelo Normal Assimétrico}
}
\footnotesize{
\begin{verbatim}

lmSN=function(m,Y.vec,X,beta.ini,delta.ini,psi.ini,a,A,c,d,e,f,cut,lag){

  # m é Nº de iterações do algoritmo de gibbs
  # Y.vec é o vetor aleatório  t-assimétrico
  # X é a matriz de planejamento do modelo

  # beta.ini é o vetor de valores iniciais para Beta
  # delta.ini é o valor inicial do parâmetro delta
  # psi.ini é o valor inicial do parâmetro psi
  # cut é a quantidade de observações descartadas no algoritmo de gibbs, de 1 à cut
  # lag informa que serão guardadas apenas amostras geradas de lag em lag interações

  #(a, A) são hiperparâmetros da priori de Beta, A é uma matriz pxp e a um vetor px1
  #(c, d) são hiperparâmetros da priori de delta
  #(e, f) são hiperparâmetros da priori de psi

  if(cut<m && lag<m){
  if(length(X[1,])==length(A[1,]) && length(X[1,])==length(a)
   && length(X[1,])==length(beta.ini) ){

  library(msm)
  library(mnormt)
  n=length(Y.vec)

  #Valores Iniciais
  ti            = delta.vec = psi.vec = vector()
  beta.mat      = matrix(NA,length(X[1,]),m+1)
  beta.mat[,1]  = beta.ini          # Beta
  delta.vec     = delta.ini         # Delta
  psi.vec       = psi.ini           # psi

  for(i in 1:m){
    # Passo 1 - Gerar ti i.i.d para i=1,...,n
      mu.vec = X %*% beta.mat[,i]
      mu.t   = ((Y.vec - mu.vec)*delta.vec[i] ) / ( delta.vec[i]^2 + 1/psi.vec[i] )
      sig.t  = 1 / ( psi.vec[i] * (delta.vec[i]^2 + 1/psi.vec[i]) )
      ti     = rtnorm(n,mu.t,sqrt(sig.t),lower=0,upper=Inf)

    # Passo 2 - Gerar Beta
      sig.beta        = solve(psi.vec[i]*t(X)%*%X + solve(A))
      mu.beta         = sig.beta%*%(psi.vec[i] * (t(X)%*%(Y.vec - delta.vec[i]*ti))
        + solve(A)%*%a)
```

```
      beta.mat[,i+1]   = rmnorm(1,mu.beta,sig.beta)

    # Passo 3 - Gerar Delta
      mu.vec         = X %*% beta.mat[,i+1]
      mu.delta       = ( sum((Y.vec - mu.vec) * ti) + (((1/psi.vec[i]) * c)/(d^2)))/
          ( sum(ti^2) + ((1/psi.vec[i])/(d^2)) )
      sig.delta      = 1 / (psi.vec[i] * (sum(ti^2) + ((1/psi.vec[i])/(d^2))))
      delta.vec[i+1] = rnorm(1,mu.delta,sqrt(sig.delta))

    # Passo 4 - Gerar Psi
      mu.vec         = X %*% beta.mat[,i+1]
      v1.psi       = (e + n/2)
      v2.psi       = (f + 1/2 * sum((Y.vec - (mu.vec + delta.vec[i+1] * ti))^2))
      psi.vec[i+1] = rgamma(1,shape=v1.psi,rate=v2.psi)
  }

     l=list(BETA=t(beta.mat[,-(1:(cut+1))][,seq(1,m-cut,lag)]),
            delta=delta.vec[-(1:(cut+1))][seq(1,m-cut,lag)],
            psi=psi.vec[-(1:(cut+1))][seq(1,m-cut,lag)])
      return(l)

 }else{print("ERRO...Hiperparâmetros ou beta inicial com dimensões erradas")}
 }else{print("ERRO!, cut ou lag são maiores que a quantidade de iterações")}
GS=lmSN(31000,Y.vec,X,beta.ini=c(0.5,-0.2,0.5),delta.ini=1,psi.ini=4,a=c(0,0,0),
     A=matrix(c(100,0,0,0,100,0,0,0,100),3,3),c=0,d=100,e=0.01,f=0.01,cut=1000,lag=5)

#Médias ---------------------------------------------------------------------
beta.hat   = colMeans(GS$BETA)
print(beta.hat)
delta.hat  = mean(GS$delta)
print(delta.hat)
psi.hat    = mean(GS$psi)
print(psi.hat)

omega2 = 1/psi.hat
sigma2 = omega2 + delta.hat^2
print(sigma2)
lambda = delta.hat/sqrt(omega2)
print(lambda)

OMEGA2 = 1/GS$psi
SIGMA2 =OMEGA2+ (GS$delta)^2
LAMBDA = GS$delta/sqrt(OMEGA2)

par(mfrow=c(2,3))

traceplot(mcmc(GS$BETA[,1]),ylab=expression(beta[1]))
traceplot(mcmc(GS$BETA[,2]),ylab=expression(beta[2]))
traceplot(mcmc(GS$BETA[,3]),ylab=expression(beta[3]))
```

```
traceplot(mcmc(LAMBDA),ylab=expression(lambda))
traceplot(mcmc(SIGMA2),ylab=expression(sigma^2))
}
```

Algoritmo para as Condicionais Completas - Modelo t-assimétrico

```
lmST=function(m,Y.vec,X,v,beta.ini,delta.ini,psi.ini,a,A,b,c2,d,e,cut,lag){

  # m é Nº de iterações do algoritmo de gibbs
  # Y.vec é o vetor aleatório t assimétrico
  # X é a matriz de planejamento do modelo

  #(a, A) são hiperparâmetros da priori de Beta, A é uma matriz pxp e a um vetor px1
  #(c, d) são hiperparâmetros da priori de delta
  #(e, f) são hiperparâmetros da priori de psi

  # beta.ini é o vetor de valores iniciais para Beta
  # delta.ini é o valor inicial do parâmetro delta
  # psi.ini é o valor inicial do parâmetro psi
  # cut é a quantidade de observações descartadas no algoritmo de gibbs, de 1:cut
  # lag informa que serão guardadas apenas amostras geradas de lag em lag interções

  if(cut<m && lag<m){
  if(length(X[1,])==length(A[1,]) && length(X[1,])==length(a)
     && length(X[1,])==length(beta.ini) ){

  library(msm)
  library(mnormt)
  n=length(Y.vec)

 #Valores Iniciais ---------------------------------------------------------------
  ti            = delta.vec = psi.vec = v2.psi = vector()
  beta.mat      = matrix(,length(beta.r),m+1)
  beta.mat[,1]  = beta.r           # Beta
  delta.vec     = delta.r          # Delta
  psi.vec       = psi.r            # psi
  s.vec = rgamma(n,v/2,v/2)        # s_i

  for(i in 1:m){
   # Passo 1 - Gerar ti i.i.d para i=1,...,n
    mu.vec = X %*% beta.mat[,i]
    mu.t   = ((Y.vec - mu.vec)*delta.vec[i]*psi.vec[i] ) /
             ( psi.vec[i]*delta.vec[i]^2 + 1)
    sig.t  = 1 / ( s.vec * (psi.vec[i]*delta.vec[i]^2 + 1 ))
    ti     = rtnorm(n,mu.t,sqrt(sig.t),lower=0,upper=Inf)

   # Passo 2 - Gerar Beta
    sig.beta         = solve(psi.vec[i]*t(X)%*%diag(s.vec)%*%X + solve(A))
    mu.beta          = sig.beta%*%(psi.vec[i]*t(X)%*% diag(s.vec) %*%
```

```
                       (Y.vec - delta.vec[i]*ti) + solve(A)%*%a)
    beta.mat[,i+1]   = rmnorm(1,mu.beta,sig.beta)

   # Passo 3 - Gerar delta
    mu.vec         =  X %*% beta.mat[,i+1]
    mu.delta       =  (psi.vec[i]*c2*(sum(ti*s.vec*(Y.vec-mu.vec))+b))/
                      (psi.vec[i]*c2*sum(s.vec*(ti^2))+1)
    sig.delta      =  c2/(psi.vec[i]*c2*sum(s.vec*(ti^2))+1)
    delta.vec[i+1] =  rnorm(1,mu.delta,sqrt(sig.delta))

   # Passo 4 - Gerar si i.i.d para i=1,...,n
    v1.s    =   (2+v)/2
    v2.s    =   0.5*(psi.vec[i]*((Y.vec - (mu.vec + delta.vec[i]*ti))^2)+ ti^2+ v)
    s.vec   =   rgamma(n,shape=v1.s,rate=v2.s)

   # Passo 5 - Gerar psi
    v1.psi       =  (n+2*d)/2
    v2.psi       =   0.5*sum(s.vec*((Y.vec -(mu.vec+delta.vec[i]*ti))^2))+e
    psi.vec[i+1] =   rgamma(1,shape=v1.psi,rate=v2.psi)

 }
  l=list(BETA=t(beta.mat[,-(1:(cut+1))][,seq(1,m-cut,lag)]),
         delta=delta.vec[-(1:(cut+1))][seq(1,m-cut,lag)],
         psi=psi.vec[-(1:(cut+1))][seq(1,m-cut,lag)])
  return(l)

 }else{print("ERRO...Hiperparâmetros ou beta inicial com dimensões erradas")}
 }else{print("ERRO!, cut ou lag são maiores que a quantidade de iterações")}

GS=lmST(31000,Y.vec,X,v=3,beta.ini=c(0.5,-0.2,0.5),delta.ini=1,psi.ini=4,a=c(0,0,0),
         A=matrix(c(100,0,0,0,100,0,0,0,100),3,3),b=0,c2=100,d=0.01,e=0.01,cut=1000,lag=5)

#Médias -------------------------------------------------------------------------
beta.hat   = colMeans(GS$BETA)
print(beta.hat)
delta.hat  = mean(GS$delta)
print(delta.hat)
psi.hat    = mean(GS$psi)
print(psi.hat)

omega2 = 1/psi.hat
sigma2 = omega2 + delta.hat^2
print(sigma2)
lambda = delta.hat/sqrt(omega2)
print(lambda)

OMEGA2 = 1/GS$psi
SIGMA2 =OMEGA2+ (GS$delta)^2
LAMBDA = GS$delta/sqrt(OMEGA2)
```

```
par(mfrow=c(2,3))

traceplot(mcmc(GS$BETA[,1]),ylab=expression(beta[1]))
traceplot(mcmc(GS$BETA[,2]),ylab=expression(beta[2]))
traceplot(mcmc(GS$BETA[,3]),ylab=expression(beta[3]))
traceplot(mcmc(LAMBDA),ylab=expression(lambda))
traceplot(mcmc(SIGMA2),ylab=expression(sigma^2))

#################
#REPARAMETRIZAÇÃO

omegar = 1/GS$psi
sigma2r = omega2 + (GS$delta)^2
lambdar = GS$delta/sqrt(omega2)

# Transformando a amostra em amostra MCMC
mcSN=mcmc(data.frame(GS$BETA[1,], GS$BETA[2,], GS$BETA[3,], sigma2r,lambdar))

#Obtendo os intervalos de credibilidade
HPDinterval(mcSN)
 }}
```

Referências

Albert, J. (2009). *Bayesian Computation with R*. Springer, second edition.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171–178.

Azzalini, A. (2010). *R package `sn`: The skew-normal and skew-t distributions (version 0.4-16)*. Università di Padova, Italia.

Branco, M. D. & Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, **79**, 99–113.

Celeux, G., Forbesy, F., Robert, C. P. & Titterington, D. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, **79**, 651–674.

Henze, N. (1986). A probabilistic representation of the skew-normal distribution. *Scandinavian Journal of Statistics*, **13**, 271–275.

R Development Core Team (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.