

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

MODELOS DE REGRESSÃO LINEAR ASSIMÉTRICOS

Bolsista: Márcia Brandão de Oliveira, CNPq

MANAUS
2012

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

RELATÓRIO FINAL
PIB-E/0004/2011
MODELOS DE REGRESSÃO LINEAR ASSIMÉTRICOS

Bolsista: Márcia Brandão de Oliveira, CNPq
Orientador: Prof. Dr. Celso Rômulo Barbosa Cabral

MANAUS
2012

Resumo

Neste trabalho, será apresentado um estudo com distribuições que procuram modelar adequadamente dados que apresentam um comportamento assimétrico e de caudas mais pesadas em sua estrutura. Essas distribuições são extensões dos modelos da família de normais independentes, como o modelo Normal e t-de-Student usual, por uma classe de distribuições de Normais Assimétricas Independentes, a qual incorpora em sua estrutura um parâmetro que regula a assimetria dos dados, serão estudados os modelos Normal Assimétrico e t de Student assimetrizado, este último é bastante atrativo pois combina assimetria com caudas pesadas. Uma extensão do modelo de regressão linear clássico será realizada, substituindo a suposição de normalidade dos erros por uma mais flexível, como a de que os erros provém de uma distribuição t-de-Student assimétrica.

Palavras chave. distribuição normal assimétrica; distribuição t de Student assimétrica; algoritmo EM; modelos de regressão linear; diagnóstico de influência; exclusão de casos.

Sumário

1	Introdução	5
2	Objetivos	5
3	Revisão Bibliográfica	6
3.1	Modelo Normal Assimétrico e t de Student Assimétrico	6
4	Métodos Utilizados	8
4.1	Estimação via algoritmo EM	8
4.2	Modelos de Regressão Linear Normal e t-de-Student Assimétricos	8
4.2.1	Modelo de Regressão Linear Normal Assimétrico ou Skew-Normal	8
4.2.2	Modelo de Regressão Linear t de Student Assimétrico ou Skew-t	9
4.3	Análise de Diagnóstico do Modelo	13
4.3.1	Exclusão de Casos	13
5	Resultados e discussões	15
5.1	Ajuste do Modelo Normal	16
5.1.1	Diagnósticos para o Modelo de Regressão Normal	16
5.2	Ajuste do Modelo de Regressão t-Assimétrico	17
5.2.1	Diagnóstico para o Modelo de Regressão t de Student Assimétrico	18
6	Cronograma	20
7	Conclusão	21
	Apêndice	22
A	Códigos no R	22
	Referências	28

1 Introdução

Modelos de regressão com erros de observação normais são aplicados frequentemente para modelar dados simétricos. Entretanto, é bem sabido que vários fenômenos não suportam a suposição de normalidade, produzindo dados com uma distribuição com caudas mais pesadas. Uma boa alternativa é considerar que os erros são provenientes de uma distribuição *t* de Student. Veja Fernández & Steel (1999) para uma discussão sobre inferência neste modelo.

No entanto, existem situações onde mesmo uma distribuição com caudas mais pesadas, como a *t* de Student, não é adequada para modelar os dados. Isto pode ocorrer devido a presença de uma assimetria acentuada na distribuição dos dados, por exemplo. Desta forma, é importante a busca de modelos mais flexíveis que possam ao mesmo tempo acomodar estes afastamentos da hipótese de normalidade. Neste trabalho, sugerimos substituir a distribuição dos erros de observação por uma distribuição que estende a distribuição *t* de Student, incorporando assimetria, chamada *Distribuição t de Student Assimétrica*. Um algoritmo do tipo EM para estimação por máxima verossimilhança é sugerido.

Como o modelo normal clássico é muito sensível a observações discrepantes, a avaliação da robustez das estimativas dos parâmetros é uma questão importante. O método de exclusão de casos, que consiste em estudar o impacto das estimativas dos parâmetros após retiradas de observações individuais, é provavelmente a técnica mais empregada para detectar a influência das observações, veja Cook & Weisberg (1982). Neste trabalho estendemos a técnica usual de exclusão de casos, aplicando-a ao modelo de regressão que estamos propondo.

2 Objetivos

- Estender os modelos de regressão linear substituindo a suposição de normalidade dos erros de mensuração por uma outra distribuição da classe de distribuições assimétricas e com caudas pesadas;
- Proceder estimação de máxima verossimilhança para os modelos sugeridos utilizando algoritmo EM;
- Desenvolver técnicas de diagnóstico;
- Analisar dados simulados e reais utilizando a metodologia proposta.

3 Revisão Bibliográfica

Há diversas situações do cotidiano que apresentam dados que se afastam de uma tendência central, perturbando a simetria do modelo Normal tradicional. Entretanto, em algumas situações, dados com este tipo de comportamento são importantes. Neste caso, eles não devem ser descartados. Para manter a estrutura de modelagem utilizando normalidade, alguns analistas costumam aplicar uma transformação aos dados originais e então usar uma modelagem normal. Pode-se mostrar que a inclusão de parâmetros que regulam assimetria e modificam a distribuição tornando-a robusta a observações discrepantes é uma alternativa que se apresenta mais adequada a este tipo de situação.

3.1 Modelo Normal Assimétrico e t de Student Assimétrico

A extensão do modelo normal tradicional para uma distribuição que incorpora a assimetria ou que também apresenta caudas pesadas em sua estrutura é um problema que vem sendo amplamente discutido na literatura. Várias representações foram apresentadas e uma das mais conhecidas é a proposta por Azzalini (1985).

Definição 1 *Uma variável aleatória Y tem distribuição Skew-Normal, com parâmetros $\mu \in \mathbb{R}$, $\sigma^2 > 0$ e $\alpha \in \mathbb{R}$, se possui densidade dada por*

$$f(y|\mu, \sigma^2, \alpha) = 2\phi(y|\mu, \sigma^2)\Phi\left[\alpha\left(\frac{y-\mu}{\sigma}\right)\right], \quad (1)$$

onde $\phi(\cdot|\mu, \sigma^2)$ denota a densidade normal usual com média μ e variância σ^2 e $\Phi(\cdot)$ é a função de distribuição da normal padrão.

Adotaremos a notação $Y \sim \text{SN}(\mu, \sigma^2, \alpha)$. Esta definição da normal assimétrica é pouco intuitiva, partindo do ponto de vista de aplicações práticas é de interesse obter representações mais simples e que sejam mais aceitáveis.

Uma outra forma de definir a distribuição normal assimétrica é dada pelo seguinte teorema. Antes de apresentá-lo, vamos definir a distribuição normal truncada (NT). Dize-mos que T tem distribuição NT no intervalo (a, b) com parâmetros ξ e η^2 , denotando por $T \sim \text{NT}(\xi, \eta^2, (a, b))$, quando T tiver a mesma distribuição de $W|(a < W < b)$ para alguma variável aleatória $W \sim N(\xi, \eta^2)$. Se $a = 0$, $b = +\infty$, $\xi = 0$ e $\eta^2 = 1$, então a chamamos de half-normal ($\text{HN}(0,1)$).

Teorema 1 *Henze (1986) Sejam $T \sim \text{HN}(0,1)$ e $V \sim N(0,1)$ variáveis aleatórias independentes. Seja*

$$Y = \mu + \sigma\delta T + \sigma(1 - \delta^2)^{1/2}V.$$

Então $Y \sim \text{SN}(\mu, \sigma^2, \alpha)$, onde $\alpha = \delta/(1 - \delta^2)$ e $-1 < \delta < 1$.

Bayes & Branco (2007) propuseram uma forma mais atrativa de escrever a representação de Henze através da reparametrização:

$$\Delta = \sigma\delta, \quad \omega^2 = \sigma^2(1 - \delta^2). \quad (2)$$

Assim, podemos escrever:

$$Y = \mu + \Delta T + \omega V,$$

onde $T \sim \text{HN}(0, 1)$ e $V \sim \text{N}(0, 1)$ são independentes. Então $Y \sim \text{SN}(\mu, \omega^2, \Delta)$.

Os parâmetros originais podem ser resgatados através das fórmulas

$$\alpha = \frac{\Delta}{\omega}, \quad \sigma^2 = \omega^2 + \Delta^2.$$

Por definição $-1 < \delta < 1$, assim, $\omega^2 > 0$. Neste caso, $\Delta = 0 \Leftrightarrow \alpha = 0$. É possível assim, enunciar o seguinte teorema:

Teorema 2 *Uma variável aleatória $Y \sim \text{SN}(\mu, \omega^2, \Delta)$ admite a representação estocástica*

$$Y|T = t \sim \text{N}(\mu + \Delta t, \omega^2), \quad T \sim \text{HN}(0, 1).$$

Podemos notar que com o condicionamento da variável aleatória T obtemos uma distribuição mais simples para Y . Utilizando esta forma de representação de Y , obtém-se um modo mais fácil e simples de se gerar observações artificialmente de sua distribuição de probabilidade. Além disso, dizemos que T é uma variável latente.

Definição 2 *Sejam $X \sim \text{SN}(0, \sigma^2, \alpha)$ e $S \sim \text{Gamma}(\nu/2, \nu/2)$ variáveis aleatórias independentes. Seja $\mu \in \mathbb{R}$. Dizemos que a distribuição de $Y = \mu + S^{-1/2}X$ é Skew-t ou t de Student assimétrica com parâmetro de locação μ , parâmetro de escala σ e ν graus de liberdade.*

Denotamos por $Y \sim \text{ST}(\mu, \sigma^2, \alpha, \nu)$. A representação estocástica do modelo t de Student assimétrico é dada pelo seguinte teorema:

Teorema 3 *Seja $Y \sim \text{ST}(\mu, \sigma^2, \alpha, \nu)$. Então Y admite a seguinte representação estocástica*

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

Esta representação será bastante útil, pois proporciona um modo mais simples de obtenção de um algoritmo do tipo EM para estimação dos parâmetros do modelo de regressão t de Student assimetrizado.

A Figura 1 mostra gráficos destas densidades para alguns valores especificados dos parâmetros, onde são apresentadas curvas reais das distribuições Normal, Normal Assimétrica e t de Student Assimétrica.

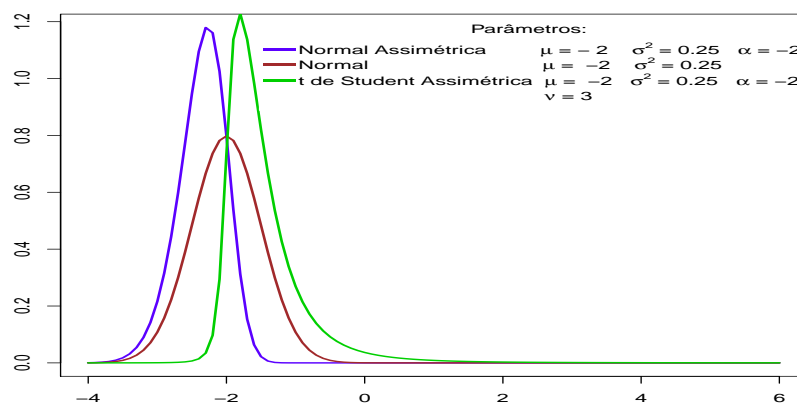


Figura 1: Curvas das densidades das distribuições Normal, Normal Assimétrica e t de Student Assimétrica.

4 Métodos Utilizados

4.1 Estimação via algoritmo EM

Para fazermos inferência, utilizaremos estimação por máxima verossimilhança via algoritmo EM (Expectation - Maximization) tanto no modelo Normal Assimétrico, quanto no t de Student Assimétrico. Este algoritmo foi proposto por Dempster *et al.* (1977), e é bastante útil para resolver problemas de estimação como este, a partir de dados incompletos. Os dados completos, também conhecidos como dados aumentados, se constituem de variáveis observadas diretamente e de variáveis não observáveis diretamente, também conhecidas como variáveis latentes, dados adicionais ou perdidos.

4.2 Modelos de Regressão Linear Normal e t-de-Student Assimétricos

Definição 3 *O modelo de regressão linear normal clássico é definido por uma variável aleatória Y_i que representa a resposta observada de determinado indivíduo i , de tal maneira que*

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad i = 1, 2, \dots, n,$$

onde, ϵ_i são erros de observação independentes e identicamente distribuídos com distribuição $N(0, \sigma^2)$, $\mathbf{x}_i^\top = (1, x_{i1}, \dots, x_{i(p-1)})$ é um vetor de $p - 1$ de variáveis chamadas regressoras associadas ao indivíduo i , $\boldsymbol{\beta}^\top = (\beta_0, \beta_1, \dots, \beta_{p-1})$ é um vetor de parâmetros desconhecidos.

O propósito do estudo com os modelos de regressão linear assimétricos e de caudas pesadas é substituir a suposição de normalidade dos erros de mensuração por uma distribuição que acomoda assimetria e caudas pesadas ao mesmo tempo. Neste caso vamos usar uma suposição de normalidade assimétrica ou t de Student assimétrica.

4.2.1 Modelo de Regressão Linear Normal Assimétrico ou Skew-Normal

A extensão da suposição de normalidade dos erros por uma distribuição normal assimétrica é dada pela representação estocástica apresentada pelo Teorema 2. Assim podemos estender tal suposição da seguinte maneira:

Definição 4 *O modelo de regressão linear normal assimétrico é definido por:*

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

onde Y_i são as variáveis respostas a serem observadas em um indivíduo i , $\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}$, $\mathbf{x}_i^\top = (1, x_{i1}, \dots, x_{i(p-1)})$ é um vetor de valores de $p - 1$ regressores para o indivíduo i , ω^2 e Δ são parâmetros desconhecidos e $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{p-1})^\top$ é um vetor de coeficientes desconhecidos da regressão. Note que Y_1, Y_2, \dots, Y_n são variáveis aleatórias condicionalmente independentes dados os valores dos parâmetros, variáveis latentes e regressores.

Marginalmente, $Y_i \sim SN(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2, \alpha)$, $i = 1, 2, \dots, n$, onde $\alpha = \Delta/\omega$ e $\sigma^2 = \omega^2 + \Delta^2$. Utilizando o fato de que transformações de localização e escala não alteram o parâmetro que regula a assimetria da distribuição, ou seja, α , tem-se:

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad \epsilon_i \sim SN(0, \sigma^2, \alpha), \quad i = 1, 2, \dots, n, \quad (3)$$

que é uma extensão do modelo apresentado na Definição 3 para um modelo de regressão linear normal.

4.2.2 Modelo de Regressão Linear t de Student Assimétrico ou Skew-t

Outra extensão do modelo de regressão linear clássico e muito mais atrativa é aquela em que se supõe que os erros provém de uma distribuição t de Student Assimétrica, neste caso, uma outra definição é apresentada.

Definição 5 *O Modelo de Regressão Linear t de Student Assimétrico é definido por*

$$Y_i | \beta, \omega^2, \Delta, T_i = t_i, S_i = s_i \sim N(\mu_i + \Delta t_i, s_i^{-1} \omega^2), T_i | S_i = s_i \sim HN(0, s_i^{-1}), S_i \sim \text{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right),$$

onde $Y_i, i = 1, \dots, n$, são variáveis aleatórias que representam as respostas a serem observadas em um indivíduo i , $\mu_i = \mathbf{x}_i^\top \beta$, $\mathbf{x}_i = (1, x_{i1}, \dots, x_{i(p-1)})^\top$ é um vetor com os valores de $(p-1)$ regressores do indivíduo i , ω^2 e Δ são parâmetros desconhecidos, $\beta = (\beta_0, \beta_1, \dots, \beta_{(p-1)})^\top$ é um vetor de parâmetros desconhecidos e ν é conhecido. Além disso Y_1, \dots, Y_n são variáveis aleatórias condicionamente independentes dados os valores dos parâmetros, variáveis latentes e regressores.

Assim, do mesmo modo como vimos no caso do modelo *Skew-Normal*, temos que $Y_i \sim \text{ST}(\mathbf{x}_i^\top \beta, \sigma^2, \alpha, \nu)$, em que os parâmetros de assimetria e escala podem ser reparametrizados por $\alpha = \Delta/\omega$ e $\sigma^2 = \omega^2 + \Delta^2$, respectivamente. Utilizando novamente o fato de que transformações de locação e escala não alteram o parâmetro que regula a assimetria, temos

$$Y_i = \mathbf{x}_i^\top \beta + \epsilon_i, \quad i = 1, 2, \dots, n,$$

onde $\epsilon_i \sim \text{ST}(0, \sigma^2, \alpha, \nu)$. Portanto, temos uma extensão do modelo de regressão linear clássico que combina erros assimétricos com caudas mais pesadas que a distribuição normal usual.

Para estimação do vetor de parâmetros $\theta = (\beta^\top, \omega^2, \Delta)^\top$ do modelo de regressão skew-t será utilizada a abordagem de dados incompletos, em que se aumenta o vetor de dados observados $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$ com a inclusão das variáveis latentes $\mathbf{t} = (t_1, t_2, \dots, t_n)^\top$ e $\mathbf{s} = (s_1, s_2, \dots, s_n)^\top$, obtendo assim dados aumentados ou completos. A função de log-verossimilhança completa é dada por

$$\begin{aligned} l_c(\theta | \mathbf{y}, \mathbf{t}, \mathbf{s}) &= c + \frac{1}{2} \sum_{i=1}^n \log s_i - \frac{n}{2} \log \omega^2 - \frac{1}{2\omega^2} \sum_{i=1}^n s_i (y_i - \mathbf{x}_i^\top \beta - \Delta t_i)^2 + \sum_{i=1}^n \log s_i \\ &\quad - \frac{1}{2} \sum_{i=1}^n s_i t_i^2 + \left(\frac{\nu-2}{2}\right) \sum_{i=1}^n \log s_i - \left(\frac{\nu}{2}\right) \sum_{i=1}^n s_i, \end{aligned} \quad (4)$$

onde c é uma constante que não depende do vetor de parâmetros θ .

No passo E do algoritmo EM é tomada a esperança condicional da log-verossimilhança dado o vetor de dados observados, $Q(\theta | \theta') = E_{\theta'} \{l_c(\theta | \mathbf{y}, \mathbf{t}, \mathbf{s}) | \mathbf{y}\}$, onde se substitui as variáveis

latentes pelas esperanças condicionadas a seguir:

$$\begin{aligned} S_{1i} &= E_{\theta'}[S_i|Y_i = y_i]; \\ S_{2i} &= E_{\theta'}[S_i T_i|Y_i = y_i]; \\ S_{3i} &= E_{\theta'}[S_i T_i^2|Y_i = y_i]. \end{aligned}$$

onde $E_{\theta'}$ é a esperança calculada supondo que θ' é o verdadeiro valor do parâmetro. A partir deste ponto, vamos retirar o subscrito θ' da esperança a fim de facilitar a notação.

Para obter a primeira esperança condicional temos que encontrar

$$E[S_i|Y_i = y_i] = \int_0^\infty s_i \pi(s_i|Y_i = y_i) ds_i$$

Usando o fato de que $Y_i|S_i = s_i \sim \text{SN}(\mathbf{x}_i^\top \boldsymbol{\beta}, s_i^{-1} \sigma^2, \alpha)$ e o Teorema de Bayes temos que

$$\begin{aligned} E[S_i|Y_i = y_i] &= \int_0^\infty s_i \frac{\pi(y_i|s_i)\pi(s_i)}{\pi(y_i)} ds_i \\ &= \frac{1}{\pi(y_i)} \int_0^\infty \frac{2}{\sigma^2} s_i \phi \left[\frac{s_i(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})}{\sigma^2} \right] \Phi \left[\alpha \frac{s_i^{1/2}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})}{\sigma} \right] \pi(s_i) ds_i \end{aligned}$$

Note que $s_i \phi \left[\frac{s_i(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})}{\sigma^2} \right] \pi(s_i) ds_i$ é a densidade conjunta de (Z_i, S_i) , onde $Z_i|S_i = s_i \sim \text{N}(\mathbf{x}_i^\top \boldsymbol{\beta}, s_i^{-1} \sigma^2)$, $S_i \sim \text{Gamma}(\nu/2, \nu/2)$ e marginalmente $Z_i \sim \text{t}(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2, \nu)$. Assim,

$$\begin{aligned} E[S_i|Y_i = y_i] &= \frac{2}{\pi(y_i)} \int_0^\infty s_i \pi_{Z_i, S_i}(y_i, s_i) \Phi \left[\alpha \frac{s_i^{1/2}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})}{\sigma} \right] ds_i \\ &= \frac{2}{\pi(y_i)} \int_0^\infty s_i \pi(s_i|Z_i = z_i) \pi_{Z_i}(y_i) \Phi \left[\alpha \frac{s_i^{1/2}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})}{\sigma} \right] ds_i \\ &= \frac{2\pi_{Z_i}(y_i)}{\pi(y_i)} \int_0^\infty s_i \Phi \left[\alpha \frac{s_i^{1/2}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})}{\sigma} \right] \pi(s_i|Z_i = z_i) ds_i. \end{aligned}$$

Observe que $\pi_{Z_i}(y_i)$ é a densidade $\text{t}(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2, \nu)$ e $\pi(y_i)$ é a densidade $\text{ST}(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2, \nu, \alpha)$. Definamos $b(y_i) = \frac{\pi_{Z_i}(y_i)}{\pi(y_i)}$, $a(y_i) = \frac{\alpha(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})}{\sigma}$, $\alpha = (1 + \nu)/2$ e $\beta(y_i) = \frac{1}{2} \left[\nu + \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\sigma^2} \right]$. Segue que

$$E[S_i|Y_i = y_i] = 2b(y_i) \int_0^\infty s_i \Phi \left[\alpha \frac{s_i^{1/2}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})}{\sigma} \right] \pi(s_i|Z_i = z_i) ds_i.$$

Usando a transformação $U_i = 2\beta(y_i)s_i$, temos que

$$S_{1i} = E[S_i|Y_i = y_i] = 2 \frac{b(y_i)}{\beta(y_i)} \frac{\Gamma(\kappa/2)}{\Gamma(\kappa/2 - 1)} T(x_i|\kappa), \quad (5)$$

onde $\kappa = 2(\alpha + 1)$ e $x_i = a(y_i) \sqrt{\frac{\kappa}{2\beta(y_i)}}$.

Fazendo uma analogia ao que fora realizado no cálculo de S_{1i} para encontrar S_{2i} , temos

$$\begin{aligned} E[T_i S_i | Y_i = y_i] &= \int_0^\infty \int_0^\infty s_i t_i \pi(s_i, t_i | y_i) dt_i ds_i \\ &= \int_0^\infty s_i \left[\int_0^\infty t_i \pi(s_i, t_i | y_i) dt_i \right] ds_i \\ &= \int_0^\infty s_i \left[\int_0^\infty t_i \pi(t_i | s_i, y_i) \pi(s_i | y_i) dt_i \right] ds_i \\ &= \int_0^\infty s_i \pi(s_i | y_i) \left[\int_0^\infty t_i \pi(t_i | s_i, y_i) dt_i \right] ds_i. \end{aligned}$$

Para dar continuidade a demonstração, precisamos saber qual é a distribuição de $T|S = s, Y = y$. O resultado apresentado no teorema a seguir é de fácil demonstração.

Teorema 4 Utilizando a Definição 5 e definindo

$$M^2 = \frac{\omega^2}{\omega^2 + \Delta^2} \quad e \quad m(y) = \frac{M^2 \Delta (y - \mathbf{x}^\top \boldsymbol{\beta})}{\omega^2},$$

temos que

$$T|S = s, Y = y \sim NT(m(y), s^{-1}M^2, (0, \infty)).$$

Neste caso, dizemos que a distribuição de $T|S = s, Y = y$ é Normal Truncada no intervalo $(0, \infty)$ com parâmetros $m(y)$ e $s^{-1}M^2$.

Para o próximo resultado vamos definir

$$\mathcal{R}(\cdot) = \frac{\phi(\cdot|0, 1)}{\Phi(\cdot)}$$

onde $\phi(\cdot|0, 1)$ denota a densidade da normal padrão e $\Phi(\cdot)$ é a função de distribuição da normal padrão.

Corolário 1 A média e a variância condicional da distribuição de $T|S = s, Y = y$, são dadas respectivamente por

$$\begin{aligned} E[T|S = s, Y = y] &= m(y) + s^{-1/2} M \mathcal{R}(s^{1/2} a(y)), \\ Var[T|S = s, Y = y] &= s^{-1/2} M^2 \{1 - \mathcal{R}(s^{1/2} a(y))\} [\mathcal{R}(s^{1/2} a(y)) + s^{1/2} a(y)]. \end{aligned}$$

Retornando ao cálculo de S_{2i} , temos

$$\begin{aligned} S_{2i} = E[T_i S_i | Y_i = y_i] &= \int_0^\infty s_i \pi(s_i | y_i) m(y_i) ds_i + M \int_0^\infty s_i^{1/2} \mathcal{R}(s_i^{1/2} a(y_i)) \pi(s_i | y_i) ds_i \\ &= m(y_i) E[S_i | Y_i = y_i] + M E[s_i^{1/2} \mathcal{R}(s_i^{1/2} a(y_i)) | y_i] \\ &= m(y_i) S_{1i} + M \mathcal{F}(y_i), \end{aligned} \tag{6}$$

$$\text{onde } \mathcal{F}(y_i) = \frac{(2/\pi)^{-1/2} b(y_i) \Gamma(\alpha + \frac{1}{2}) [\beta(y_i)]^\alpha}{\Gamma(\alpha) [\beta(y_i) + \frac{1}{2} a(y_i)^2]^{\alpha + \frac{1}{2}}}.$$

Para encontrarmos a esperança condicionada S_{3i} , precisamos calcular

$$\begin{aligned}
E[T_i^2 S_i | Y_i = y_i] &= \int_0^\infty \int_0^\infty t_i^2 s_i \pi(t_i, s_i | y_i) dt_i ds_i \\
&= \int_0^\infty \int_0^\infty t_i^2 s_i \pi(t_i | s_i, y_i) \pi(s_i, y_i) dt_i ds_i \\
&= \int_0^\infty s_i \left[\int_0^\infty t_i^2 \pi(t_i | s_i, y_i) dt_i \right] \pi(s_i, y_i) ds_i \\
&= \int_0^\infty s_i \left[s_i^{-1} M^2 - s^{-1/2} \mathcal{R}(s^{1/2} a(y_i)) (a(y_i) M^2 + m(y_i) M) + m(y_i)^2 \right] \pi(s_i | y_i) ds_i \\
&= M^2 + m(y_i)^2 E[S_i | Y_i = y_i] + M m(y_i) E[s_i^{1/2} \mathcal{R}(s_i^{1/2} a(y_i)) | y_i] \\
&= M^2 + m(y_i)^2 S_{1i} + M m(y_i) \mathcal{F}(y_i)
\end{aligned} \tag{7}$$

O Passo M consiste na maximização da função $Q(\theta | \theta^{(k)})$ em relação ao vetor de parâmetros θ . A partir deste ponto, uma quantidade indexada pelo superescrito (k) , indica o valor da estimativa desta quantidade obtida no passo (k) do algoritmo EM.

$$Q(\theta | \theta^{(k)}) = c - \frac{n}{2} \log \omega^2 - \frac{1}{2\omega^2} \sum_{i=1}^n \left[(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2 S_{1i} - 2(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) \Delta S_{2i} + \Delta^2 S_{3i} \right]. \tag{8}$$

Para facilitar o cálculo do estimador do vetor de coeficientes da regressão vamos definir $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$ um vetor $(n \times 1)$ de variáveis respostas do modelo de regressão, o vetor com os valores das esperanças $\mathbf{S}_i = (S_{i1}, \dots, S_{in})$, $i = 1, 2, 3$, $\mathbf{A} = \text{diag}(\mathbf{S}_1)$ é uma matriz diagonal com o vetor \mathbf{S}_1 na diagonal principal, $\mathbf{1} = (1, \dots, 1)^\top$ é um vetor $n \times 1$ e a matriz

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

é a matriz de planejamento com todos os regressores do modelo. Sabendo que a seguinte igualdade é verdadeira:

$$\sum_{i=1}^n S_{1i} (Y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2 = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{A} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}),$$

então podemos reescrever a função em (8), a menos de constantes não importantes, da seguinte maneira:

$$Q(\theta | \theta^{(k)}) = -\frac{n}{2} \log \omega^2 - \frac{1}{2\omega^2} \left[(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{A} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right] + \frac{1}{2\omega^2} \left[2\Delta \mathbf{S}_2^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) - \Delta^2 \mathbf{1}^\top \mathbf{S}_3 \right].$$

Deste modo é possível encontrar expressões de forma fechada para atualização dos parâmetros envolvidos no modelo de regressão. Portanto podemos definir o seguinte algoritmo EM que segue

Passo-E: Dado $\theta = \theta^{(k)}$, calcula-se $\mathbf{S}_1^{(k)}$, $\mathbf{S}_2^{(k)}$, $\mathbf{S}_3^{(k)}$ e consequentemente $\mathbf{A}^{(k)}$.

Passo-M: Atualiza-se $\theta^{(k)} = (\boldsymbol{\beta}^{(k)}, \Delta^{(k)}, \omega^{2(k)})^\top$ do modelo de regressão linear skew-t, usando as expressões de forma fechada que seguem

$$\begin{aligned}\boldsymbol{\beta}^{(k+1)} &= (\mathbf{X}^\top \mathbf{A}^{(k)} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{A}^{(k)} \mathbf{Y} - \Delta^{(k)} (\mathbf{X}^\top \mathbf{A}^{(k)} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{S}_2^{(k)} \\ \Delta^{(k+1)} &= (\mathbf{1}^\top \mathbf{S}_3^{(k)})^{-1} \mathbf{S}_2^{\top(k)} (\mathbf{Y} - \mathbf{X} \boldsymbol{\beta}^{(k+1)}) \\ \omega^{2(k+1)} &= \frac{1}{n} \left[(\mathbf{Y} - \mathbf{X} \boldsymbol{\beta}^{(k+1)})^\top \mathbf{A}^{(k)} (\mathbf{Y} - \mathbf{X} \boldsymbol{\beta}^{(k+1)}) - 2 \Delta^{(k+1)} \mathbf{S}_2^{\top(k)} (\mathbf{Y} - \mathbf{X} \boldsymbol{\beta}^{(k+1)}) \right] \\ &\quad + \frac{1}{n} \left[\Delta^{2(k+1)} \mathbf{1}^\top \mathbf{S}_3^{(k)} \right]\end{aligned}$$

Utilizando as expressões acima é possível resgatar os parâmetros de escala e assimetria do modelo,

$$\begin{aligned}\sigma^{2(k+1)} &= \omega^{2(k+1)} + (\Delta^{(k+1)})^2; \\ \alpha^{(k+1)} &= \frac{\Delta^{(k+1)}}{\sqrt{\omega^{2(k+1)}}}.\end{aligned}$$

É sabido que o algoritmo EM é um método iterativo de estimação, o mesmo repete os passos E e M até que um critério de convergência seja atingido, neste caso, quando a diferença entre as estimativas das log-verossimilhanças $|l(\theta^{(k+1)}) - l(\theta^{(k)})|$ seja bastante pequena. Para mais detalhes sobre estimação via algoritmo EM em modelos assimétricos ver Zeller (2009) e Cabral *et al.* (2012).

4.3 Análise de Diagnóstico do Modelo

A técnica de análise de diagnóstico consiste na avaliação das estimativas dos parâmetros de um modelo de regressão quando certas perturbações ocorrem tanto no conjunto de dados, quanto nas suposições do modelo. Existem dois meios de detecção da influência das observações, utilizando a técnica de exclusão de casos Cook (1977) ou a de influência local Cook (1986). Neste trabalho, estudaremos apenas a primeira.

4.3.1 Exclusão de Casos

A técnica de exclusão de casos é uma abordagem comum quando se deseja estudar o efeito da retirada da i -ésima observação do conjunto de dados nas estimativas dos parâmetros do modelo. Neste caso, definimos uma quantidade com a subscrição “[i]”, que denota a quantidade original com o i -ésimo caso excluído. Seja $\widehat{\boldsymbol{\theta}}_{[i]} = (\widehat{\boldsymbol{\beta}}_{[i]}^\top, \widehat{\Delta}_{[i]}, \widehat{\omega}_{[i]}^2)^\top$ o vetor de estimativas dos parâmetros que maximizam a função $Q_{[i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = E_{\widehat{\boldsymbol{\theta}}} \{l_c(\boldsymbol{\theta}|\mathbf{y}_{[i]}, \mathbf{t}_{[i]}, \mathbf{s}_{[i]}|\mathbf{y})\}$, onde $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\Delta}, \widehat{\omega}^2)^\top$ é o estimador EM de $\boldsymbol{\theta}$. Para avaliar a influência da i -ésima observação na estimativa de $\boldsymbol{\theta}$, vamos verificar a diferença entre $\widehat{\boldsymbol{\theta}}_{[i]}$ e $\widehat{\boldsymbol{\theta}}$. Se a exclusão causar muita influência nas estimativas, maior atenção deve ser dada a este caso. Assim, se as estimativas de $\widehat{\boldsymbol{\theta}}_{[i]}$ e $\widehat{\boldsymbol{\theta}}$ forem muito diferentes, neste caso, essa observação é dita ser uma observação influente. Um esforço computacional muito grande é necessário para encontrar as estimativas $\widehat{\boldsymbol{\theta}}_{[i]}$ para todo i . Portanto, uma pseudo-aproximação $\widehat{\boldsymbol{\theta}}_{[i]}^1$ foi proposta para reduzir este esforço (Cook, 1986),

$$\widehat{\boldsymbol{\theta}}_{[i]}^1 = \widehat{\boldsymbol{\theta}} + \{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1} \dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \quad (9)$$

onde $\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ é a matriz hessiana e $\dot{Q}_{[i]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = \frac{\partial Q_{[i]}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ é o vetor escore, que têm seus elementos dados por

A matriz Hessiana

$$\begin{aligned} \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= -\frac{1}{\omega^2} \mathbf{X}^\top \mathbf{A} \mathbf{X}; \\ \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \Delta \partial \Delta} &= -\frac{1}{\omega^2} \mathbf{1}^\top \mathbf{S}_3; \\ \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \omega^2 \partial \omega^2} &= \frac{n}{2\omega^4} - \frac{1}{\omega^6} \left[(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{A} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) - 2\Delta \mathbf{S}_2^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \Delta^2 \mathbf{1}^\top \mathbf{S}_3 \right] \\ \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \Delta} &= -\frac{1}{\omega^2} \mathbf{X}^\top \mathbf{S}_2; \\ \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \omega^2} &= -\frac{1}{\omega^4} \left[\mathbf{X}^\top \mathbf{A} \mathbf{X} \boldsymbol{\beta} - \mathbf{X}^\top \mathbf{A} \mathbf{Y} + \Delta \mathbf{X}^\top \mathbf{S}_2 \right]; \\ \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \Delta \partial \omega^2} &= -\frac{1}{\omega^4} \left[\mathbf{S}_2^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) - \Delta \mathbf{1}^\top \mathbf{S}_3 \right]; \end{aligned}$$

O Vetor Escore

$$\begin{aligned} \frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta}} &= \frac{1}{\omega^2} \left[\mathbf{X}_{[i]}^\top \mathbf{A}_{[i]} \mathbf{X}_{[i]} \boldsymbol{\beta} - \mathbf{X}_{[i]}^\top \mathbf{A}_{[i]} \mathbf{Y}_{[i]} + \Delta \mathbf{X}_{[i]}^\top \mathbf{S}_{2[i]} \right] \\ \frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \Delta} &= \frac{1}{\omega^2} \left[(\mathbf{Y}_{[i]} - \mathbf{X}_{[i]} \boldsymbol{\beta})^\top \mathbf{S}_{2[i]} - \Delta \mathbf{1}_{[i]}^\top \mathbf{S}_{3[i]} \right] \\ \frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \omega^2} &= -\frac{1}{\omega^2} \left\{ \frac{(n-1)}{2} - \frac{1}{2\omega^2} \left[(\mathbf{Y}_{[i]} - \mathbf{X}_{[i]} \boldsymbol{\beta})^\top \mathbf{A}_{[i]} (\mathbf{Y}_{[i]} - \mathbf{X}_{[i]} \boldsymbol{\beta}) \right] \right\} \\ &\quad + \frac{1}{2\omega^4} \left[\Delta^2 \mathbf{1}_{[i]}^\top \mathbf{S}_{3[i]} - 2\Delta \mathbf{S}_{2[i]}^\top (\mathbf{Y}_{[i]} - \mathbf{X}_{[i]} \boldsymbol{\beta}) \right] \end{aligned}$$

Baseado na métrica que mensura a distância entre $\hat{\boldsymbol{\theta}}_{[i]}^1$ e $\hat{\boldsymbol{\theta}}$, proposta por Zhu & Lee (2001), baseada no algoritmo EM, vamos considerar a *distância de Cook generalizada* que segue:

$$GD_i = (\hat{\boldsymbol{\theta}}_{[i]}^1 - \hat{\boldsymbol{\theta}})^\top \ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})^{-1} (\hat{\boldsymbol{\theta}}_{[i]}^1 - \hat{\boldsymbol{\theta}}), \quad i = 1, 2, \dots, n. \quad (10)$$

Substituindo (9) em (10), obtemos a seguinte expressão:

$$GD_i = \dot{Q}_{[i]}^\top \{ -\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) \}^{-1} \dot{Q}_{[i]}, \quad i = 1, 2, \dots, n. \quad (11)$$

Para identificar as observações influentes vamos construir um gráfico com os valores das *distâncias de Cook generalizadas* e traçar nele um ponto de corte (*benchmark*, em inglês), $BM = 2(p+1)/n$, ver Barros *et al.* (2010). Considera-se ponto influente aquele em que a *distância de Cook generalizada* for maior que o *benchmark* BM, indicando que esta i -ésima observação causa impactos nas estimativas de máxima verossimilhança do vetor de parâmetros do modelo.

5 Resultados e discussões

Para emprego e análise da metodologia proposta, utilizou-se um conjunto de dados citado no texto de Albert(2000), o qual propõe um estudo com pares de espécies de pássaros coletados em 16 ilhas próximas à Grã-Bretanha ao longo de várias décadas. Para cada espécie são associadas as seguintes variáveis:

- *Time* – Tempo médio de extinção na ilha onde a espécie apareceu;
- *Nesting* – Número médio de pares aninhados;
- *Size* – Tamanho da espécie (grande ou pequeno);
- *Status* – Status migratório da espécie (migrantes ou residentes).

O objetivo do estudo é verificar se o tempo médio de extinção na ilha, *Time*, sofre variação por efeito das covariáveis *Nesting*, *Size* e *Status*.

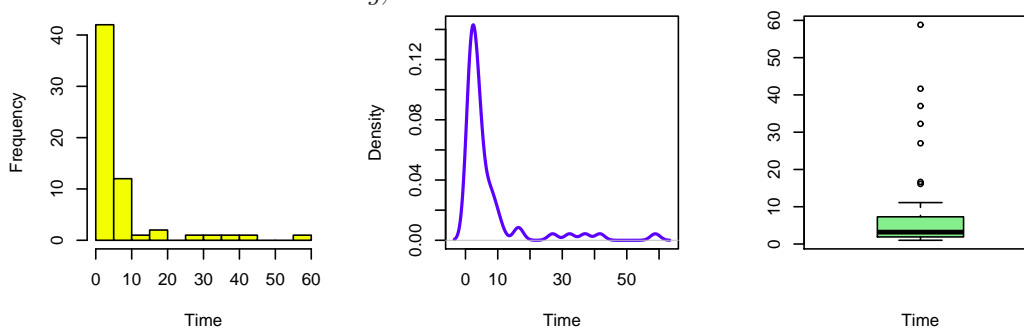


Figura 2: *Histograma, estimativa da densidade e boxplot da variável resposta Time.*

Como pode ser observado na Figura 2, a variável *Time* apresenta comportamento bastante assimétrico, com a presença de muitas observações discrepantes, sugerindo um modelo que combina assimetria e acomode em suas caudas mais pesadas tais observações.

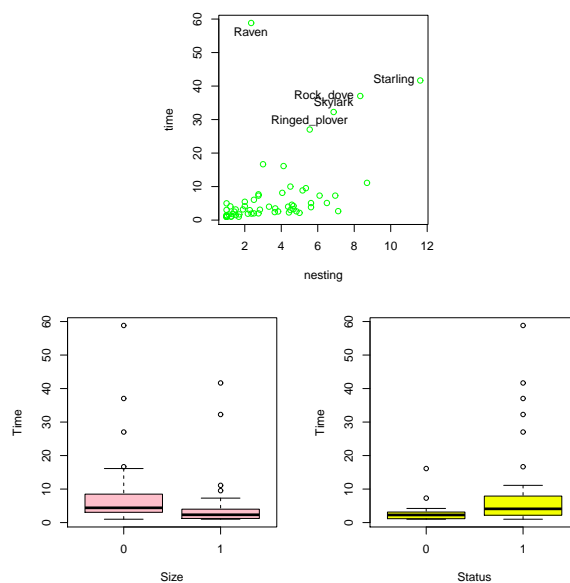


Figura 3: *Gráficos de dispersão dos regressores X variável resposta Time.*

Observando os gráficos apresentados na Figura 3, vemos que com a variável Nesting, há um comportamento quase que linear com a variável Time, já os boxplots inerentes a variável Time estratificada pelos níveis das variáveis Size e Status sugerem um comportamento de não normalidade, além de vários *outliers*.

A proposta do estudo é modelar o tempo médio de extinção dos pássaros na ilha como função das demais variáveis apresentadas no estudo, assim o seguinte modelo estatístico é proposto:

$$\text{Time}_i = \beta_0 + \beta_1 \text{Nesting}_i + \beta_2 \text{Size}_i + \beta_3 \text{Status}_i + \epsilon_i, \quad i = 1, 2, \dots, 62. \quad (12)$$

onde, β_i , $i = 0, 1, 2, 3$, são os coeficientes da regressão a serem estimados e ϵ_i são erros de observação, que serão modelados primeiramente pela distribuição Normal, $\epsilon_i \sim N(0, \sigma^2)$ e posteriormente t-assimétrica, $\epsilon_i \sim \text{ST}(0, \sigma^2, \alpha, \nu)$.

5.1 Ajuste do Modelo Normal

Para ajustar um modelo normal, algumas suposições inerentes a este modelo devem ser consideradas, como a de que os erros aleatórios são provenientes de uma distribuição normal com média 0 e variância σ^2 constante e que não há multicolineariedade entre as colunas da matriz de regressores. Desta forma, utilizando a função `lm` no pacote `stats` do software R^1 , obtém-se as seguintes estimativas:

Tabela 1 *Resultados referentes ao ajuste do modelo normal.*

Coeficientes	Estimativa	Erro Padrão
Intercepto	-0.6276	2.9115
Nesting	2.1337	0.5173
Size(Grande)	-3.9275	2.3435
Status(Residente)	3.5298	2.5680
AIC	456.63	–
BIC	467.27	–

Na Tabela 1 temos as estimativas de máxima verossimilhança dos coeficientes do ajuste do modelo normal.

5.1.1 Diagnósticos para o Modelo de Regressão Normal

Para verificar se o modelo normal ajusta-se adequadamente aos dados, foi realizada uma análise nos resíduos deste modelo, os resultados obtidos podem ser melhor observados a seguir.

¹Sobre o software R , ver R Development Core Team (2012)

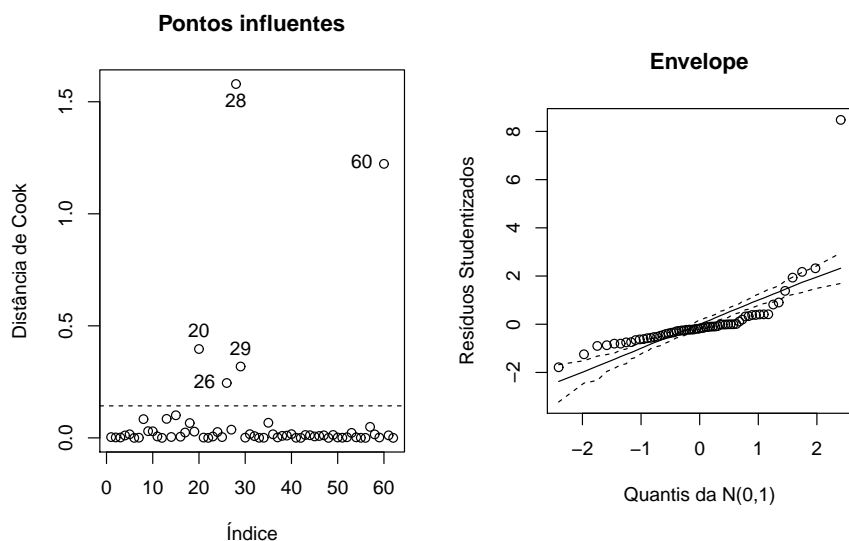


Figura 4: Gráficos da análise dos resíduos do modelo de regressão normal.

O diagnóstico do modelo normal apresentado na Figura 4, sugere vários pontos influentes no modelo. Dentre os influentes, foram identificadas as observações 20, 26, 28, 29 e 60, as quais se referem respectivamente às espécies *Rock-dove*, *Jackdaw*, *Raven*, *Skylark* e *Starling*. O envelope normal construído, mostra que o modelo normal não é adequado e sugere um modelo com distribuição com caudas mais pesadas.

5.2 Ajuste do Modelo de Regressão t-Assimétrico

Como o ajuste do modelo de regressão normal apresentou vários problemas na presença de observações discrepantes, nesta seção será realizada a estimação de máxima verossimilhança do modelo de regressão apresentado em (12). Neste caso $\epsilon_i \sim ST(0, \sigma^2, \alpha, \nu)$, ou seja, parâmetro de locação 0, escala σ^2 , assimetria α e ν graus de liberdade.

Fixou-se os graus de liberdade em $\nu = 3$, pois Lucas(1997) observa que somente no pressuposto de graus de liberdade fixos, as estimativas dos parâmetros se comportam de forma robusta contra observações extremas.

Tabela 2 Resultados referentes ao ajuste do modelo skew-t via algoritmo EM.

Coefficientes	Estimativa	Erro Padrão
Intercepto	-0.0146	0.4893
Nesting	0.4895	0.1428
Size	-0.4423	0.3739
Status	0.6179	0.3609
α	8.4851	4.5654
AIC	424.73	—
BIC	437.50	—

Na Tabela 2 acima temos as estimativas de máxima verossimilhança obtidas por intermédio do algoritmo EM, apresentado na seção (4.2.2), construído a fim de estimar os parâmetros do modelo de regressão proposto. Observando as estimativas dos desvios, vemos que as mesmas são muito menores que as obtidas no modelo normal. Avaliando os critérios de seleção de modelos, obtidos nos dois ajustes, vemos que tanto o AIC quanto o BIC obtidos no modelo t-assimétrico são menores, evidenciando que este modelo melhor ajusta os dados.

Tabela 4 *Intervalos de Confiança dos coeficientes do modelo t-assimétrico.*

Coeficientes	Lim. Inferior	Lim. Superior
β_0	-0.1363695	0.1072339
β_1	0.4539939	0.5250687
β_2	-0.5353715	-0.3492369
β_3	0.5281448	0.7077972

Através da Tabela 4 é possível observar os limites de variação construídos ao nível de 95% de confiança para os coeficientes do modelo de regressão t-assimétrico. É possível perceber que todos os regressores do modelo são marginalmente significativos, ou seja, as variáveis Nesting, Size e Status são variáveis que podem explicar o tempo médio de extinção de espécies de pássaros nas ilhas da Grã-Bretanha.

5.2.1 Diagnóstico para o Modelo de Regressão t de Student Assimétrico

Para realizar o diagnóstico no modelo proposto para o estudo de extinção de pássaros, foram calculadas *distancias de Cook generalizadas*, a fim de identificar possíveis pontos influentes que possam estar interferindo na estimação dos parâmetros e também mostrar a superioridade da metodologia proposta no uso do modelo *skew-t*.

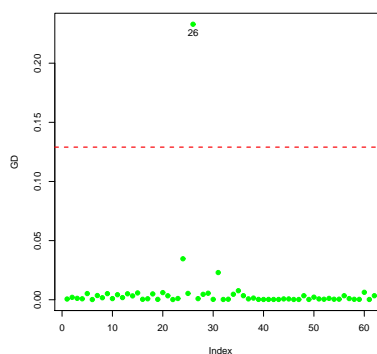


Figura 5: *Distancias de cook generalizadas - modelo t de Student Assimétrico.*

Através da Figura 5 é possível observar as *distâncias de Cook* calculadas para o modelo proposto. A linha tracejada em vermelho indica o ponto de corte, $BM = 0.1290$, mostrando apenas a observação 26 referente a espécie *Jackdaw* como um possível ponto influente, evidenciando a superioridade do modelo proposto, pois no ajuste do modelo normal, várias observações influentes foram evidenciadas, enquanto que no modelo t de Student Assimétrico, que possui caudas mais pesadas que a distribuição normal, foi apontado somente um ponto influente no ajuste.

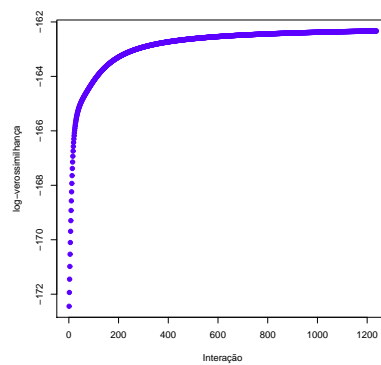


Figura 6: *Gráfico do comportamento da função de log-verossimilhança.*

A Figura 6 ilustra o comportamento da função de log-verossimilhança durante o processo de estimação via algoritmo EM. É possível observar que a cada interação a função assume um valor maior que o calculado na interação anterior. Quando o critério de parada apresentado na seção (4.2.2) é atingido, ou seja, a diferença entre a log-verossimilhança atualizada e a anterior é na ordem de 10^{-6} , o algoritmo se estabiliza, de modo que a diferença entre a verossimilhança atual e a anterior é muito pequena, indicando assim que o algoritmo convergiu, conforme abordado na literatura.

6 Cronograma

Nesta seção apresentamos o cronograma de atividades propostas e que foram ou não desenvolvidas no decorrer do projeto.

Tabela 5: *Cronograma de atividades PIBIC - 2011/ 2012*

it	Descrição	JUL	AGO	SET	OUT	NOV	DEZ
1	Leitura de textos básicos em distribuições normais e t de Student assimétricas	R	R	R	R		
2	Formulação de modelos lineares com erros normais assimétricos				R	R	
3	Desenvolvimento e implementação de algoritmo EM para estimação no modelo Skew-Normal					R	R
		JAN	FEV	MAR	ABR	MAI	JUN
4	Elaboração de relatório parcial	R					
5	Formulação de modelos lineares com erros t de Student assimétricos		R	R			
6	Desenvolvimento e implementação de algoritmo EM para estimação no modelo Skew-t				R	R	
7	Desenvolvimento de técnicas de diagnóstico em um modelo de regressão Skew-t						R
8	Elaboração de relatório Final						R

Legenda:

R: Realizado

NR: Não Realizado

7 Conclusão

O uso de distribuições assimétricas como a *Normal Assimétrica* e *t de Student Assimétrica* são alternativas bastante atrativas, pois como visto, há situações no cotidiano em que se deseja estudar um conjunto de observações para tentar descrever fenômenos da natureza, como o estudo de extinção de pássaros nas ilhas da Grã-Bretanha, e o uso dos modelos de regressão linear têm sido bastante utilizado nesta tarefa. Entretanto, os modelos de regressão linear clássicos baseados na distribuição normal usual, apresentam falhas na presença de observações com aspecto assimétrico e/ou com presença de observações discrepantes. Neste contexto, os modelos assimétricos são empregados, pois utilizam a distribuição normal como um caso particular, são modelos mais robustos, como o *t de Student Assimétrico*, que combina assimetria com caudas pesadas. Estes modelos são atrativos pelo fato de possibilitar ao pesquisador a realização de inferências mais robustas a respeito do comportamento de seus dados, não sendo necessário realizar transformações nos mesmos.

O exemplo utilizado para aplicação da metodologia proposta neste estudo, ilustra com bastante eficácia a superioridade de um modelo mais robusto como o modelo de regressão *t de Student* assimétrico, pois estima com mais qualidade os parâmetros do modelo, que por combinar assimetria com caudas pesadas, não sofre influência de observações discrepantes.

A metodologia utilizada para detecção de observações influentes no modelo de regressão t de Student Assimétrico, apresentada na subseção (4.3.1), trata-se de uma técnica inédita, tendo em vista que não foi encontrado nenhum estudo publicado, referente a identificação de observações influentes no modelo de regressão t de Student Assimétrico, utilizando a técnica da exclusão de observações. Este vem a ser o grande ganho científico com a realização deste projeto.

A Códigos no R

Nesta seção apresentamos os códigos criados no software estatístico R para simulação de dados e análise durante o estudo.

```

#-----
# Ajuste modelo normal
#-----

library(LearnBayes)

data(birdextinct)
dados <- (birdextinct)
attach(dados)
names(dados) # "species" "time" "nesting" "size" "status"

# Análise descritiva
par(mfrow=c(1,3),pty='s')
hist(time,breaks=10,main="",col="yellow",xlab="Time")
plot(density(time),main="",col="blue",xlab="Stack-Loss",lwd=2)
boxplot(time,col="lightgreen",xlab="Time")

# observando linearidade dos regressores com var. resposta
layout(matrix(c(1,1,2,3), 2, 2, byrow=TRUE),respect=TRUE)
par(mar=c(5.1, 12.1, 1.4, 10.1), pty='s')
plot(nesting,time,col="green")
identify(dados$nesting,dados$time, label=dados$species,n=5)
par(mar=c(5.1, 4.1, 1.4, 2.1))
boxplot(time ~ size, col="pink",main="",xlab="Size",ylab="Time")
# identify(dados$time~dados$size, label=dados$species)
boxplot(time ~ status,col="yellow",xlab="Size",ylab="Time")
# identify(dados$time~dados$status, label=dados$species)

#ajuste mod. normal
ajuste1<-lm( time ~ nesting + factor(size) + factor(status))
summary(ajuste1)

#dados
n <- length(species)
y.teste<-dados$time
X <- cbind(dados$nesting,dados$size,dados$status)

## Matriz modelo e chapéu
X <- model.matrix(ajuste1) ## matriz modelo
Xr <- as.matrix(X[,-1]) ## matriz modelo só com regressores
X <- Xr
h <- hat(Xr) ## diagonal de matriz H = X%*%solve(t(X)%*%X)%*%t(X)

## Resíduos
e <- residuals(ajuste1, type="pearson") ## resíduos ordinários (de Pearson)

n <- nrow(X) ## número de observações
p <- ncol(X) ## número de parâmetros de regressão + intercepto
k <- ncol(Xr) ## número de regressores
S2 <- sum(e^2)/(n-p) ## estimativa da variância dos erros
phi <- 1/S2 ## parâmetro de dispersão
rp <- e*sqrt(phi) ## resíduos padronizados

ts <- rp/sqrt(1-h) ## resíduos studentizados

```

```
#####
## Verificando Violação de Normalidade dos Erros ##
#####
qqnorm(rp, main="", xlab="Quantis Teóricos", ylab="Quantis Amostrais")
qqline(rp, col = 2)

jarque.bera.test(rp)      ## teste de Jarque-Bera

#####
## Análise de Diagnóstico ##
#####

limiar.h <- 3*p/n          ## para pontos de alavanca

ma.ts <- max(ts)
mi.ts <- min(ts)

Cd <- (rp^2)*h/(1-h)      ## Distância de Cook
limiar.Cd <- 8/(n-(2*p))

#####
# Envelope #
#####

envelope.normal<-function(form=form,k=k,alfa=alfa){
  alfa1<-ceiling(k*alfa)
  alfa2<-ceiling(k*(1-alfa))
  glm1<-lm(formula=form)
  X<-model.matrix(glm1)
  n<-nrow(X)
  p<-ncol(X)
  H<-X%*%solve(t(X)%*%X)%*%t(X)
  h<-diag(H)
  lmi<-lm.influence(glm1)
  si<-lmi$sigma
  rp<-residuals(glm1)
  ts<-rp/(si*(1-h)^0.5)
  ident<-diag(n)
  epsilon<-matrix(0,n,k)
  e<-matrix(0,n,k)
  e1<-numeric(n)
  e2<-numeric(n)
  for(i in 1:k){
    epsilon[,i]<-rnorm(n,0,1)
    e[,i]<-(ident-H)%*%epsilon[,i]
    u<-diag(ident-H)
    e[,i]<-e[,i]/(u^0.5)
    e[,i]<-sort(e[,i])
  }
  for(i in 1:n){
    eo<-sort(e[,i])
    e1[i]<-eo[alfa1]
    e2[i]<-eo[alfa2]
  }
  xb<-apply(e,1,mean)
  faixa<-range(ts,e1,e2)
  par(pty="s")
  qqnorm(e1,axes=F,xlab="",ylab="",type="l",ylim=faixa,lty=2, main="")
  par(new=T)
  qqnorm(e2,axes=F,xlab="",ylab="",type="l",ylim=faixa,lty=2, main="")
  par(new=T)
  qqnorm(xb,axes=F,xlab="",ylab="",type="l",ylim=faixa,lty=1, main="")
}
```

```

par(new=T)
qqnorm(ts,xlab="Quantis da N(0,1)",ylab="Resíduos Studentizados", ylim=faixa, main="Envelope")
}

#####
# Gráficos de Diagnóstico #
#####

par(mfrow = c(2,2),pty="s")
plot(h, xlab="Índice", ylab="h", main="Pontos de Alavanca",ylim=c(0,1))
abline(limiar.h,0,lty=2)
identify(h)
plot(ts,xlab="Índice", ylab="Resíduos Studentizados", main="Pontos aberrantes", ylim=c(mi.ts-1,ma.ts+1))
abline(2,0,lty=2)
abline(-2,0,lty=2)
identify(ts)
plot(Cd, xlab="Índice", ylab="Distância de Cook", main="Pontos influentes")
abline(limiar.Cd,0,lty=2)
identify(Cd)
form <- (time ~ nesting + factor(size) + factor(status))
envelope.normal(form,k=100,alfa=0.05)

#-----
## Ajuste modelo Skew-t
#-----
library(LearnBayes)
data(birdextinct)
dados <- (birdextinct)

attach(dados)
names(dados)
[1] "species" "time"      "nesting" "size"      "status"

n <- length(time)
y.teste <- vector()
y.teste<-dados$time
X <- cbind(nesting,size,status)

# valores iniciais dos Parâmetros do modelo
beta <- c(0.2, 0.1, 0.1)
nu <- 3
sigma2 <- 1
alpha <- 1
tol <- 0.000001

## Densidade da ST com locação escala
# Parametrização original
dt.ls <- function(y,mu, sigma2,shape,nu){
d <- (y - mu)/sqrt(sigma2)
dens <- 2*dt(d, df = nu)*pt(sqrt((1+nu)/(d^2+nu))*d*shape,1+nu)/sqrt(sigma2)
return(dens)
}

# Algoritmo EM para estimar os parâmetros no modelo de regressão ST
est.regST <- function(y,X,beta,sigma2,alpha,nu,tol){
sigma2 <- c(sigma2)
alpha <- c(alpha)
nu <- c(nu)
beta <- matrix(beta)
lk <- vector()
lkante <- 1
n <- length(y)

```



```

criterio <- 1
count <- 0
y <- matrix(y)
mu <- X%*%beta
#-----
#Iniciando as iterações do EM
#-----
delta <- alpha/sqrt(1+alpha^2)
Delta <- sqrt(sigma2)*delta
omega2 <- sigma2*(1-delta^2)
while(criterio > tol){
  count <- count + 1
  print(count)

#-----
# Passo E
#-----
# Calculando E[S_i|y_i]
by <- ((1/sqrt(sigma2)) * dt((y-mu)/sqrt(sigma2),nu))/ dt.ls(y,mu, sigma2,alpha,nu)
betay <- 0.5*(((y-mu)^2)/sigma2) + nu)
ay <- alpha*(y-mu)/sqrt(sigma2)
kappa <- nu+3
xx <- ay*sqrt(kappa/(2*betay))
S1 <- 2* (by/betay) * (gamma(kappa/2)/gamma(kappa/2-1)) * pt(xx,kappa)
#-----
# Calculando E[S_i,T_i|y_i]
M2 <- 1/(1+alpha^2)
my <- (alpha/(sqrt(sigma2*(1+alpha^2))))*(y-mu)
Fy <- (sqrt(2/pi)*by*gamma((nu+2)/2)*betay^((nu+1)/2))/(gamma((nu+1)/2)*(betay+0.5*ay^2)^((nu+2)/2))
S2 <- my*S1 + sqrt(M2)*Fy
#-----
# Calculando E[S_i,T_i^2|y_i]
S3 <- M2 + (my^2)*S1 + sqrt(M2)*my*Fy
#-----
# PASSO M
#-----
A <- diag(as.vector(S1))
baba <- solve(crossprod(X,A%*%X))
beta <- baba%*%crossprod(X,A)%*%y - Delta * baba%*%crossprod(X,as.vector(S2))
mu <- X%*%beta
Delta <- crossprod(y-mu,S2)/sum(S3)
omega2 <- (mahalanobis(t(y),t(mu),A,TRUE)-2*Delta*crossprod(y-mu,S2)+Delta^2*sum(S3))/n
sigma2 <- omega2 + Delta^2
alpha <- Delta/sqrt(omega2)

sigma2 <- c(sigma2)
alpha <- c(alpha)
Delta <- c(Delta)
omega2 <- c(omega2)
#=====
# Critérios
for (i in 1:n){
  lk[i] <- log(dt.ls(y[i],mu[i],sigma2,alpha,nu))
}
lks <- sum(lk)
criterio <- abs( lks/ lkante-1)
lkante <- lks
#print(lks)
}
return(list(beta=beta,Delta=Delta,omega2=omega2,sigma2=sigma2,alpha=alpha,S2=S2,S3=S3,A=A,mu=mu,lks=lks))
}

```

```

est.teste <- est.regST(y.teste,X,beta,sigma2,alpha,nu,tol)

#-----
# Programa para calcular AIC e BIC no modelo de regressão skew-t
#-----
criteria.ST <- function(y,X,beta,sigma2,alpha,nu){
  sigma2 <- c(sigma2)
  alpha <- c(alpha)
  nu <- c(nu)
  beta <- matrix(beta)
  lk <- vector()
  n <- length(y)
  y <- matrix(y)
  mu <- X%*%beta
  for (i in 1:n){
    lk[i] <- log(dt.ls(y[i],mu[i],sigma2,alpha,nu))
  }
  lks <- sum(lk)
  AIC <- -2*lks+2*(nrow(beta)+2)
  BIC <- -2*lks+log(n)*(nrow(beta)+2)
  return(list(AIC=AIC,BIC=BIC))
}

criteria.ST <- criteria.ST(y.teste,X,beta,sigma2,alpha,nu)

#####
## Diagnóstico de influencia - Deleção de casos ##
#####

n<-length(y.teste)
p<-length(beta)
S2 = est.teste$S2
S3 = est.teste$S3
A = est.teste$A
beta = est.teste$beta
omega2 = est.teste$omega2
Delta = est.teste$Delta
sigma2 <- est.teste$sigma2
alpha <- est.teste$alpha
mu = est.teste$mu
um <-rep(1,n)
y <- y.teste
theta.est <- c(beta,Delta,omega2)

# matriz derivadas segundas
q11 <- -(1/omega2)*crossprod(X,A%*%X)
q12 <- -(1/omega2)*crossprod(X,S2)
#q13<- -(1/omega2^2)*(crossprod(X,A%*%X)%*%beta - crossprod(X,A%*%y) + Delta*crossprod(X,S2))
q13 <- matrix(0,p,1)
q22 <- -(1/omega2)*crossprod(um,S3)
#q23<- -(1/omega2^2)*(crossprod(S2,(y-mu)) - Delta*crossprod(um,S3))
q23 <- 0
q33 <- -(1/omega2^2)*(-(n/2) + (1/omega2)*(mahalanobis(t(y),t(mu),A,TRUE)-2*Delta*crossprod(y-mu,S2)+
Delta^2*crossprod(um,S3)))
#-----
Qdp <- rbind(cbind(q11,matrix(c(q12,q13),p,2)),cbind(rbind(t(q12),t(q13)),matrix(c(q22,q23,q23,q33),2,2)))
Qup <- list()
#-----

#derivadas primeiras
theta1 <- matrix(,n,p+2)

```

```

GD1 <- vector()
for(j in 1:n){
  print(j)
  X. <- X[-j,]
  y. <- y[-j]
  A. <- A[-j,-j]
  S2. <- S2[-j]
  S3. <- S3[-j]
  mu. <- X.%*%beta
  um. <- um[-j]

  k1 <- (1/omega2)*(crossprod(X.,A.%*%X.)%*%beta - crossprod(X.,A.%*%y.) +Delta*crossprod(X.,S2.))
  k2 <- (1/omega2)*(crossprod((y.-mu.),S2.) -Delta*crossprod(um.,S3.) )
  k3 <- -(n-1)/(2*omega2) + (1/(2*omega2^2))*(mahalanobis(t(y.),t(mu.),A.,TRUE)-
    2*Delta*crossprod(y.-mu.,S2.)+Delta^2*crossprod(um.,S3.))

  Qup[[j]] <- matrix(c(k1,k2,k3),p+2,1)

  #Aproximação=====
  theta1[j,] <- theta.est + solve(-Qdp)%*%Qup[[j]]
  #=====

  GD1[j] <- crossprod(Qup[[j]],solve(-Qdp)%*%Qup[[j]])
}

#####
# GD - Generalized cook Distance #
#####
GD = vector()

for(i in 1:n){
  GD[i] <- t(theta1[i,] - theta.est)%*%(-Qdp)%*%(theta1[i,] - theta.est)
}

##### GRÁFICOS #####
plot(GD,pch=19,col="green")
criterio <- 2*(p+1)/n
abline(criterio,0,lwd="2",col="red",lty=2)
identify(GD)
#-----

```

Referências

- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171–178.
- Barros, M., Galea, M. & Leiva, V. (2010). Influence diagnostics in the tobit censored response model. *Statistics Methods & Applications*, **19**, 379–397.
- Bayes, C. L. & Branco, M. D. (2007). Bayesian inference for the skewness parameter of the scalar skew-normal distribution. *Brazilian Journal of Probability and Statistics*, **21**, 141–163.
- Cabral, C. R. B., Lachos, V. H. & Prates, M. O. (2012). Multivariate mixture modeling using skew-normal independent distributions. *Computational Statistics and Data Analysis*, **56**, 126–142.
- Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, pages 15–18.
- Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society, Series B*, **48**, 133–169.
- Cook, R. D. & Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman & Hall/CRC, Boca Raton, FL.
- Dempster, A., Laird, N. & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1–38.
- Fernández, C. & Steel, M. J. F. (1999). Multivariate Student-t regression models: pitfalls and inference. *Biometrika*, **86**, 153.
- Henze, N. (1986). A probabilistic representation of the skew-normal distribution. *Scandinavian Journal of Statistics*, **13**, 271–275.
- R Development Core Team (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Zeller, C. B. (2009). *Distribuições misturas de scala skew-normal: Estimação e diagnóstico em modelos lineares*. Ph.D. thesis, Universidade Estadual de Campinas.
- Zhu, H. & Lee, S. (2001). Local influence for incomplete-data models. *Journal of the Royal Statistical Society, Series B*, **63**, 111–126.