

UNIVERSIDADE FEDERAL DE JUIZ DE FORA

DEPARTAMENTO DE ESTATÍSTICA

CURSO DE ESTATÍSTICA

Franciele Regina Leandro

Modelo Linear Normal Assimétrico Aplicado

Juiz de Fora

2013

Franciele Regina Leandro

Modelo Linear Normal Assimétrico Aplicado

Trabalho de Conclusão de Curso apresentado ao curso de Estatística da Universidade Federal de Juiz de Fora, como requisito para obtenção do diploma de bacharel em Estatística.

Orientadora: Prof^a. Dr^a. Camila Borelli Zeller

Juiz de Fora

2013

Leandro, Franciele Regina.
Modelo Linear Normal Assimétrico Aplicado / Franciele
Regina Leandro. -- 2013.
52 p.

Orientadora: Camila Borelli Zeller
Trabalho de Conclusão de Curso (graduação) - Universidade
Federal de Juiz de Fora, Instituto de Ciências Exatas, 2013.

1. Ponto de Mudança. 2. Modelo Linear . 3. Distribuição
Normal Assimétrica. 4. Algoritmo EM. I. Zeller, Camila
Borelli, orient. II. Título.

Franciele Regina Leandro

Modelo Linear Normal Assimétrico Aplicado

Trabalho de Conclusão de Curso apresentado ao curso de Estatística da Universidade Federal de Juiz de Fora, como requisito para obtenção do diploma de bacharel em Estatística.

Aprovado em 25 de março de 2013.

BANCA EXAMINADORA

Camila Borelli Zeller

Doutora em Estatística – Universidade Estadual de Campinas

Clécio da Silva Ferreira

Doutor em Estatística – Universidade de São Paulo

Lupércio França Bessegato

Doutor em Estatística – Universidade Federal de Minas Gerais

*Dedico aos meus pais
Divino e Rute, ao meu
irmão Flávio.*

AGRADECIMENTOS

À Deus por me proteger e iluminar durante toda a minha vida.

Aos meus pais pelo amor incondicional, por estarem sempre presentes me apoiando e incentivando em todos os momentos. Ao meu irmão pelo companheirismo, dedicação e inúmeros conselhos. Ao meu primo Luís que se tornou um irmão.

À Cacau, Camila, Carol e Isabela pela amizade e por se tornarem pessoas muito especiais durante esses anos. A Raquel por compartilhar um desafio e pelos momentos de boa conversa.

Aos meus compadres Janaína e Hildeu pela grande amizade em tão pouco tempo. Aos meus queridos afilhados Álvaro, David e Sophia que apesar de tão pequenos me proporcionam muitas alegrias.

À Tamyris por ser uma grande amiga em todos os momentos bons e ruins. Aos meus amigos Ederson e Lucas por serem pessoas ao qual tenho a oportunidade de conviver.

Aos professores pelos ensinamentos durante esses anos, em especial a professora Camila por sua dedicação, paciência e ensinamentos, sendo difícil descrever a minha gratidão.

Aos meus amigos de turma que se fizeram muito importantes durante esse curso.

À todos aqueles que de alguma forma contribuíram para essa realização.

Resumo

Neste trabalho, apresentamos resultados recentes em uma área de pesquisa da estatística com uma possibilidade enorme de aplicações, que são os modelos de regressão lineares. A normalidade (simetria) dos erros aleatórios é uma suposição rotineira em modelos lineares, que pode ser não realista. Assim, relaxamos a suposição de normalidade considerando que os erros aleatórios seguem uma distribuição normal assimétrica. Esta distribuição inclui a distribuição normal como caso especial e fornece flexibilidade em capturar uma ampla variedade de comportamentos não normais, por simplesmente adicionar um parâmetro que controla o grau de assimetria. Explorando propriedades estatísticas do modelo considerado, implementamos o algoritmo EM que fornece algumas vantagens sobre a maximização direta da função log-verossimilhança.

Em um segundo estágio do trabalho, consideramos o fato de que o mesmo modelo de regressão linear pode não ser válido para todo um conjunto de dados. Isto é, o modelo pode se alterar após um ponto específico que, em geral, é desconhecido, e denominado ponto de mudança. Neste contexto, a estimação dos parâmetros do modelo será via algoritmo EM, a seleção de modelos e a detecção do ponto de mudança serão realizadas através dos critérios de informação (SIC e AIC).

Dessa forma, a principal contribuição desse trabalho está em estudar o problema de ponto de mudança associado ao modelo de regressão no contexto de assimetria, estendendo em certo sentido alguns resultados desenvolvidos por Chen (1998), Chen e Gupta (2001), Osorio e Galea (2005), Zeller et al. (2011) e Arellano et al. (2013), por exemplo.

Espera-se que este trabalho seja útil para despertar o interesse de estudantes, pesquisadores e profissionais pelo tema, que acreditamos ser de grande aplicabilidade.

Palavras-chave: Algoritmo EM, Distribuição Normal Assimétrica, Modelo Linear e Ponto de Mudança.

Abstract

At this work, we present recent results in a statistical research area with a huge possibility of applications, which are linear regression models. The normality (symmetry) of random errors is an ordinary assumption in linear models, which may be unrealistic. So, we unwind the assumption of normality considering that random errors follow a normal asymmetrical distribution. This distribution includes the normal distribution as a special case and provides flexibility in capturing a wide variety of not normal behaviors, by simply adding a parameter that controls the degree of asymmetry. Exploring statistical properties of the model considered, we implemented the algorithm EM that provides some advantages over direct maximization of the log-likelihood function.

In a second stage of this work, we considered the fact that the same linear regression model may not be valid for an entire set of data. Therefore, the model can be changed after a specific point which, in general, is unknown, and named as the changing point. In this context, the estimation of the model parameters will be via algorithm EM, the selection of models and detection of the changing point will be realized by using information criteria (SIC and AIC).

In this way, the main contribution of this work is to study the problem of the changing point associated with the regression model on the asymmetry context, extending in a way some results developed by Chen (1998), Chen and Gupta (2001), Osorio and Galea (2005), Zeller et al. (2011) and Arellano et al. (2013), for example.

It is hoped that this work will be useful to awakening the interest of students, researchers and professionals to the subject, which we believe to be of a great applicability.

Key-words: EM Algorithm, Skew-Normal Distribution, Linear Model and Change Point.

Sumário

1. Introdução	10
1.1 Motivação	10
1.2 Objetivos	11
1.3 Algoritmo EM	12
1.4 Critério de Informação	13
1.5 Descrição dos Capítulos	14
2. Distribuição Normal Assimétrica	16
2.1 Distribuição Normal Assimétrica Padrão	16
2.1.1 Propriedades	17
2.1.2 Caracterização	19
2.2 Distribuição Normal Assimétrica Localização-Escala	20
2.2.1 Caracterização	22
3. Modelo de Regressão Linear Assimétrico	23
3.1 Descrição do Modelo	23
3.2 Função Log-Verossimilhança	24
3.3 Inferência pelo Método de Máxima Verossimilhança	25
3.3.1 Representação Hierárquica	25
3.3.2 Estimação dos Parâmetros Via Algoritmo EM	26
3.4 Aplicações em Dados Reais	29
3.4.1 Aplicação 1: AIS (“Australian Institute of Sport”)	29
3.4.2 Aplicação 2: Pássaros em Extinção	33
4. Modelo de Regressão Linear Normal Assimétrico com Ponto de Mudança	37
4.1 Descrição do Modelo com Ponto de Mudança	38
4.2 Estimação dos Parâmetros Via Algoritmo EM	39
4.3 Detecção do Ponto de Mudança via Critério de Informação	41
4.4 Aplicação em Dados Simulados	42
4.4.1 Estimação por Máxima Verossimilhança	42
4.4.2 Critério de Informação	43
4.5 Aplicação em Dados Reais (Holbert’s data)	45
5. Conclusão	49
6. Perspectivas Futuras	50
7. Referências Bibliográficas	51

1. Introdução

1.1 Motivação

Modelos de regressão linear são técnicas bastante populares em pesquisa porque apresentam uma estrutura que permite aplicações em diversas áreas científicas, tais como, economia, agricultura, biologia, ciências médicas, entre outras.

Modelos de regressão linear são amplamente utilizados com o objetivo de descrever a relação entre uma variável resposta e uma ou mais variáveis explicativas. Usualmente, assume-se que o mesmo modelo de regressão linear é válido para todo um conjunto de dados, mas isso nem sempre ocorre. De fato, o modelo pode se alterar após um ponto específico que, em geral, é desconhecido, e denominado ponto de mudança.

O problema de ponto de mudança surgiu, inicialmente, no contexto de controle de qualidade e antes da introdução da hipótese de ponto de mudança associado com os modelos de regressão, pesquisadores enfrentavam dificuldades para estabelecer um modelo para alguns conjuntos de dados.

O problema de ponto de mudança tem sido um tópico de interesse permanente na literatura estatística. Em particular, na literatura, muitos autores têm estudado o problema de ponto de mudança associado com os modelos de regressão. Por exemplo, Chen (1998) e Chen e Gupta (2001), sob a suposição de normalidade, realizaram um estudo de detecção de ponto de mudança no modelo de regressão linear.

Contudo, em muitas situações, inferências sob normalidade são impróprias, por exemplo, quando os dados proveem de uma distribuição com caudas mais ou menos pesadas que a distribuição normal ou ainda assimétrica. Modelos alternativos ao modelo normal que preservam a estrutura simétrica e que permitam reduzir a influência dos “outliers” têm sido sugeridos por muitos autores. Por exemplo, Osorio e Galea (2005) realizaram um estudo de detecção de ponto de mudança no modelo de regressão linear Student-t. Essa distribuição é um membro particular de uma classe mais ampla e conhecida na literatura como distribuições elípticas; veja, por exemplo, Fang et al. (1990) para mais detalhes.

Do ponto de vista prático, muitos autores têm usado transformações de variáveis para alcançar normalidade ou pelo menos a simetria e em muitas situações, seus resultados são satisfatórios. Entretanto, Azzalini e Capitanio (1999) têm apontado alguns problemas. Por

exemplo, as variáveis transformadas são mais difíceis de serem interpretadas, especialmente quando cada variável é transformada usando uma função diferente.

Embora a classe de distribuições elípticas represente uma boa alternativa à distribuição normal, ela não é adequada em situações nas quais a distribuição das observações é assimétrica. Assim, é de interesse prático estudar distribuições que sejam menos sensíveis do que a distribuição gaussiana a certos desvios das suposições consideradas, construindo famílias paramétricas que sejam analiticamente tratáveis, que possam acomodar valores práticos de assimetria e curtose e que contenham estritamente a distribuição normal, como caso especial.

Quando a distribuição dos dados tem um comportamento assimétrico, esta pode ser modelada através de membros da classe de distribuições assimétricas, também conhecidas na literatura como distribuições skew-elípticas. Nesta classe, a distribuição normal assimétrica univariada é talvez a pioneira nesta ideia. Inicialmente, introduzida por O'hagan e Leonard (1976) como uma distribuição a priori em análise bayesiana e posteriormente, formalmente estudada por Azzalini (1985) (do ponto de vista clássico) como uma extensão natural da distribuição normal para modelar a estrutura assimétrica presente nos dados.

Estudos em modelos estatísticos sob a classe de distribuições assimétricas estão sendo aplicados com sucesso, tanto no contexto teórico quanto prático. Por exemplo, Zeller et al. (2011) realizaram um estudo de estimação e diagnóstico no modelo regressão linear sob a classe das distribuições assimétricas misturas de escala skew-normal (SMSN); veja Branco e Dey (2001) para mais detalhes sobre as distribuições assimétricas. Arellano et al. (2013) realizaram um estudo de detecção de ponto de mudança, sob o ponto de vista bayesiano, no modelo normal assimétrico localização-escala.

Nessa linha de pesquisa, propomos estudar o problema de ponto de mudança no contexto de assimetria, especificamente, no modelo de regressão linear normal assimétrico, estendendo em certo sentido alguns resultados desenvolvidos por Chen (1998), Chen e Gupta (2001), Osorio e Galea (2005), Hofrichter (2007), Zeller et al. (2011) e Arellano et al. (2013), por exemplo.

1.2 Objetivos

O objetivo geral deste trabalho é apresentar resultados recentes em uma área de pesquisa da estatística com possibilidade enorme de aplicações, que são os modelos de

regressão lineares. O enfoque do trabalho é voltado para a flexibilização do modelo linear tradicional, onde assume-se que os erros de observação tem distribuição normal. Aqui assumimos que os erros têm distribuição normal assimétrica com o intuito de acomodar desvios da suposição de normalidade como assimetria.

Podemos então relacionar os seguintes objetivos específicos:

1. A título de motivação, apresentamos alguns resultados básicos envolvendo a distribuição normal assimétrica, que é a extensão da distribuição normal incorporando assimetria.
2. Apresentar um estudo de estimação no modelo de regressão linear normal assimétrico fornecendo um suplemento necessário para a realização deste trabalho.
3. Desenvolver um estudo de detecção de ponto de mudança no modelo de regressão linear normal assimétrico, generalizando alguns resultados recentes encontrados em Osorio e Galea (2005), Zeller et al. (2011) e Arellano et al. (2013), por exemplo. É importante ressaltar que este tópico é a principal contribuição deste trabalho.

A estimação dos parâmetros será via algoritmo EM, a seleção de modelos e a detecção do ponto de mudança serão através dos critérios de informação.

1.3 Algoritmo EM

O algoritmo EM (Dempster et. al, 1977) é um enfoque aplicado no cálculo iterativo de estimativas dos parâmetros do modelo (θ) pelo método de máxima verossimilhança. Este algoritmo é aplicado ao problema de estimação a partir de dados incompletos, aumentando o vetor de dados observados (\mathbf{y}_{obs}) com a inclusão de variáveis latentes (\mathbf{y}_{mis}), as quais não são observadas diretamente. Obtendo-se, deste modo, o vetor de dados completos $\mathbf{y}_c = (\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}})$, de tal forma que a função de log-verossimilhança é representada por $l_c(\theta|\mathbf{y}_c) = \log(f(\mathbf{y}_c|\theta))$. Cada iteração do algoritmo EM consiste em dois passos:

- **Passo E (Esperança)**

Este passo consiste em calcular a esperança da log-verossimilhança completa, denotada por $Q(\theta|\hat{\theta}^{(t)})$, condicionada no vetor de dados observados, onde $\hat{\theta}^{(t)}$ representa a

estimativa de $\boldsymbol{\theta}$ na t -ésima iteração. Ou seja, para a t -ésima iteração temos que, dado $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(t)}$,

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(t)}) = E\{l_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}_{obs}, \hat{\boldsymbol{\theta}}^{(t)}\};$$

- **Passo M (Maximização)**

Consiste em maximizar a log-verossimilhança completa em relação aos parâmetros do modelo, substituindo os dados latentes por seus valores esperados condicionais obtidos no passo E. Para a t -ésima iteração, obtemos $\hat{\boldsymbol{\theta}}^{(t+1)}$ que maximize $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^t)$, tal que

$$Q(\hat{\boldsymbol{\theta}}^{(t+1)}|\hat{\boldsymbol{\theta}}^{(t)}) > Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(t)}), \forall \boldsymbol{\theta} \in \Theta,$$

onde $\Theta \subseteq \mathbb{R}^P$ representa o espaço paramétrico.

Estes passos se alternam até convergir e geralmente são considerados como critérios de convergência $\left| \frac{\hat{\boldsymbol{\theta}}^{(t+1)} - \hat{\boldsymbol{\theta}}^{(t)}}{\hat{\boldsymbol{\theta}}^{(t)}} \right| < \epsilon$ ou $\left| \frac{l(\hat{\boldsymbol{\theta}}^{(t+1)}|\mathbf{y}_{obs}) - l(\hat{\boldsymbol{\theta}}^{(t)}|\mathbf{y}_{obs})}{l(\hat{\boldsymbol{\theta}}^{(t)}|\mathbf{y}_{obs})} \right| < \epsilon$, onde $|a|$ indica a norma do vetor a , $l(\cdot|\mathbf{y}_{obs})$ é o logaritmo da função de verossimilhança observada e $\epsilon > 0$.

Neste trabalho, consideramos

$$\left| \frac{l(\hat{\boldsymbol{\theta}}^{(t+1)}|\mathbf{y}_{obs}) - l(\hat{\boldsymbol{\theta}}^{(t)}|\mathbf{y}_{obs})}{l(\hat{\boldsymbol{\theta}}^{(t)}|\mathbf{y}_{obs})} \right| < 10^{-5} \quad (1.1)$$

como critério de parada do algoritmo EM.

1.4 Critério de Informação

Apesar de não serem testes formais, como em Zhang e Davidian (2001), verificamos a adequação dos modelos aos dados inspecionando alguns critérios de informação.

De uma forma geral, os critérios são definidos por

$$-2l(\hat{\boldsymbol{\theta}}|\mathbf{y}_{obs}) + sc_n, \quad (1.2)$$

onde $l(\cdot|\mathbf{y}_{obs})$ é o logaritmo da função de verossimilhança observada, $\hat{\boldsymbol{\theta}}$ é a estimativa de máxima verossimilhança de $\boldsymbol{\theta}$, s é o número de parâmetros do modelo e c_n é uma sequência de números positivos. Note que se $c_n = 2$, obtemos o critério de informação AIC (Akaike Information Criterion) definido por

$$AIC = -2l(\hat{\boldsymbol{\theta}}|\mathbf{y}_{obs}) + 2s, \quad (1.3)$$

Considerando $c_n = \log n$, onde n é o tamanho amostral, obtemos o critério de informação SIC (Schwarz Information Criterion) definido por

$$SIC = -2l(\hat{\boldsymbol{\theta}}|\mathbf{y}_{obs}) + \log(n) s. \quad (1.4)$$

Cada um desses critérios baseia-se numa penalização da verossimilhança na medida em que o modelo se torna mais complexo, isto é, modelos com um grande número de parâmetros. Dessa forma, o modelo que apresentar o menor valor do critério de informação será o modelo selecionado. Além disso, neste trabalho, os critérios de informação também serão utilizados na detecção do ponto de mudança associado com os modelos, conforme Chen (1998) e Osorio e Galea (2005), por exemplo. Essa linha de pesquisa será mais explorada no Capítulo 4.

1.5 Descrição dos Capítulos

O Capítulo 2 descreve a distribuição normal assimétrica introduzida por Azzalini (1985), nas formas padrão e locação-escala, destacando suas propriedades, momentos, função geradora de momentos, representação estocástica e distribuições de formas quadráticas. Neste capítulo, destacam-se os resultados referentes aos momentos dessa distribuição e também quanto à sua representação estocástica, os quais serão utilizados de forma recorrente nos demais capítulos.

O Capítulo 3 descreve o modelo de regressão linear normal assimétrico, incluindo o algoritmo EM para estimação de máxima verossimilhança dos parâmetros do modelo, conforme Ferreira (2008) e Zeller (2009), por exemplo. Por último, aplicações do modelo são realizadas considerando os conjuntos de dados reais de atletas australianos e dos pássaros em extinção, já estudados no contexto de assimetria; veja, por exemplo, Ferreira (2008) e Lachos

et al. (2013), respectivamente. Neste capítulo, destaca-se a representação hierárquica do modelo de regressão com erro assimétrico, permitindo assim a obtenção dos estimadores de máxima verossimilhança dos parâmetros do modelo via algoritmo EM.

O Capítulo 4 descreve o modelo de regressão linear normal assimétrico com ponto de mudança, incluindo o algoritmo EM para a estimação por máxima verossimilhança dos parâmetros do modelo e os critérios de informação (SIC e AIC), úteis na detecção do ponto de mudança. Exemplos numéricos considerando dados simulados e reais são apresentados para ilustrar a metodologia proposta. O quarto capítulo pode ser considerado o objetivo principal desse trabalho. Nossa principal contribuição está em estudar o problema do ponto de mudança associado ao modelo de regressão no contexto de assimetria, estendendo em certo sentido alguns resultados desenvolvidos por Chen (1998), Chen e Gupta (2001), Osorio e Galea (2005), Zeller et al. (2011) e Arellano et al. (2013), por exemplo.

Finalmente, no Capítulo 5, são apresentadas algumas considerações finais e, no Capítulo 6, as perspectivas de trabalhos futuros.

2. Distribuição Normal Assimétrica

Existe uma tendência geral na literatura estatística para encontrar métodos flexíveis que representem da forma mais verossímil possível características de fenômenos encontrados, nas mais diversas áreas da ciência. Em muitos métodos propostos se faz presente a suposição de normalidade dos dados, essa que nem sempre pode ser a mais adequada. Para modelar desvios desta suposição, muitos enfoques podem ser encontrados na literatura. Possivelmente, o método comumente adotado para alcançar a simetria é a transformação de variáveis. Embora essa estratégia possa trazer resultados razoáveis, Azzalini e Capitanio (1999) apresentam inúmeras razões para se evitar esse procedimento.

Portanto, estudar modelos paramétricos que sejam mais flexíveis que a distribuição gaussiana é de grande interesse na literatura atual, e a construção de novas famílias de distribuições que sejam capazes de incorporar assimetria, de forma analiticamente tratável, tornaram-se uma grande motivação para pesquisadores nos últimos anos. Sob essa motivação é que será proposto, neste capítulo, um estudo baseado na distribuição normal assimétrica (em inglês, “skew-normal”).

Neste capítulo, será apresentada uma revisão sobre algumas das propriedades e resultados referentes à distribuição normal assimétrica univariada. Essa distribuição será utilizada nos demais capítulos para relaxar a suposição de normalidade, geralmente, assumida nos modelos de regressão. É vasta a quantidade de trabalhos relevantes sobre a distribuição normal assimétrica na literatura, alguns deles encontram-se citados neste trabalho.

2.1 Distribuição Normal Assimétrica Padrão

Definição 2.1.1 Uma variável aleatória Y tem distribuição normal assimétrica padrão com parâmetro de forma λ ($\lambda \in \mathbb{R}$), se sua função densidade de probabilidade é dada por

$$f_Y(y|\lambda) = 2\phi(y)\Phi(\lambda y), -\infty < y < \infty, \quad (2.1)$$

onde $\phi(\cdot)$ e $\Phi(\cdot)$ representam a função densidade de probabilidade e a função distribuição acumulada da distribuição normal padrão, respectivamente. Para mais detalhes veja Azzalini (1985). Para considerações deste trabalho, denotaremos $Y \sim SN(\lambda)$ para representar uma

variável aleatória com distribuição normal assimétrica padrão. O parâmetro de forma λ caracteriza a assimetria da densidade, sendo $\lambda > 0$ assimetria positiva, $\lambda < 0$ assimetria negativa e $\lambda = 0$ simetria, coincidindo com a densidade da normal padrão. A Figura 2.1 apresenta alguns exemplos de densidades da normal assimétrica padrão para diferentes valores do parâmetro λ .

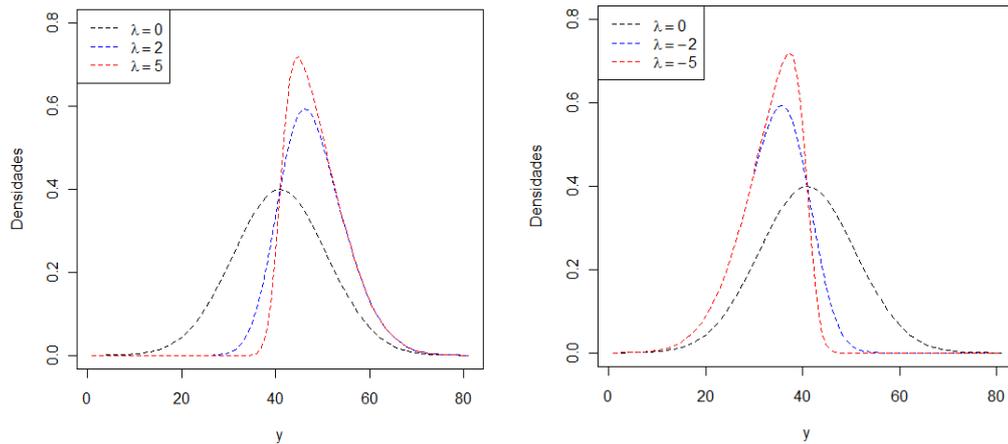


Figura 2.1: Dados simulados. Gráficos das densidades da normal assimétrica padrão para diferentes valores do parâmetro λ .

2.1.1 Propriedades

As propriedades listadas a seguir são dadas para uma variável aleatória Y com distribuição normal assimétrica padrão. As demonstrações são aqui omitidas por não serem o principal foco desse trabalho. Para uma discussão mais detalhada quanto às propriedades seguintes veja Azzalini (1985), Henze (1986) e Azzalini (2005), por exemplo.

Propriedades

Antes de apresentar uma primeira propriedade da distribuição normal assimétrica padrão, vamos definir a distribuição normal truncada (NT). Dizemos que T tem uma distribuição NT no intervalo (a, b) com parâmetros μ e σ^2 , denotada por $T \sim NT(\mu, \sigma^2, (a, b))$, quando T tiver a mesma distribuição de $X|(a < X < b)$ para alguma variável aleatória $X \sim N(\mu, \sigma^2)$. Se $a = 0$, $b = \infty$, $\mu = 0$ e $\sigma^2 = 1$, então a chamamos de half-normal padrão ($HN(0,1)$).

1. Se $Y \sim SN(\lambda)$, então $|Y| \sim HN(0,1)$ (a distribuição half-normal padrão).
2. Quando $\lambda \rightarrow \infty$, a densidade (2.1) converge para $HN(0,1)$, tal propriedade pode ser visualizada através da Figura 2.2.

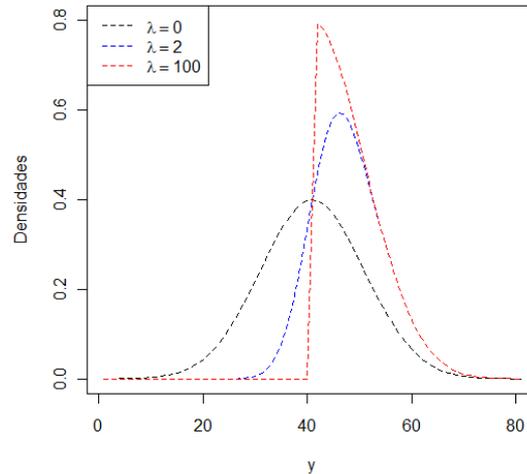


Figura 2.2: Dados Simulados. Gráfico das densidades da normal assimétrica padrão, quando

$$\lambda \rightarrow +\infty.$$

3. Se $Y \sim SN(\lambda)$ então $Y^2 \sim \chi_1^2$. A mesma propriedade é válida para o caso normal, ou seja, quando $X \sim N(0,1)$ então $X^2 \sim \chi_1^2$.

A seguir, será derivada a função geradora de momentos da distribuição normal assimétrica padrão, da qual pode-se obter medidas de interesse, tais como média e variância.

Proposição 2.1: A função geradora de momentos de $Y \sim SN(\lambda)$ é dada por

$$M_Y(t) = 2 \exp\left(\frac{t^2}{2}\right) \Phi(\delta t), t \in \mathbb{R},$$

onde $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

Demonstração: A prova desta proposição pode ser encontrada em Basso (2009), página 9.

A partir da Proposição 2.1, pode-se mostrar que

$$E(Y) = \sqrt{\frac{2}{\pi}} \delta \quad (2.2)$$

e

$$Var(Y) = 1 - \frac{2}{\pi} \delta^2. \quad (2.3)$$

Note que, quando $\lambda = 0$ e conseqüentemente $\delta = 0$, temos a média e variância da distribuição normal padrão (caso particular).

2.1.2 Caracterização

A seguir será apresentada a representação estocástica de uma variável aleatória normal assimétrica padrão. Tal representação é de suma importância no decorrer desse trabalho, visto que inúmeros resultados podem ser derivados a partir dessa caracterização.

Proposição 2.2: Seja $X_0, X_1 \stackrel{iid}{\sim} N(0,1)$, então

$$Y = \delta |X_0| + (1 - \delta^2)^{\frac{1}{2}} X_1,$$

tem distribuição $SN(\lambda)$, onde “*iid*” significa independentes e identicamente distribuídos e $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

Demonstração: A prova desta proposição pode ser encontrada em Basso (2009), página 11.

A partir da Proposição 2.2, também podemos obter a média e a variância de uma variável aleatória $Y \sim SN(\lambda)$, definidas em (2.2) e (2.3), respectivamente. Além disso, a partir dessa proposição, podemos gerar números pseudo-aleatórios de uma distribuição normal assimétrica padrão.

2.2 Distribuição Normal Assimétrica Locação-Escala

É natural estender o modelo definido em (2.1) introduzindo parâmetros de locação e escala. Neste caso, diz-se que $Y \sim SN(\mu, \sigma^2, \lambda)$.

Definição 2.2.1 Uma variável aleatória Y tem distribuição normal assimétrica com parâmetro de locação μ ($\mu \in \mathbb{R}$), parâmetro de escala σ^2 ($\sigma > 0$) e parâmetro de forma λ ($\lambda \in \mathbb{R}$), se sua função densidade de probabilidade é dada por

$$f_Y(y|\mu, \sigma^2, \lambda) = \frac{2}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right) \Phi\left(\lambda\left(\frac{y-\mu}{\sigma}\right)\right), -\infty < y < \infty, \quad (2.4)$$

onde $\phi(\cdot)$ e $\Phi(\cdot)$ representam a função densidade de probabilidade e a função distribuição acumulada da distribuição normal padrão, respectivamente. Note que, o termo $\phi(\cdot)$ depende da distância de Mahalanobis

$$d = \frac{(Y - \mu)^2}{\sigma^2}, \quad (2.5)$$

útil para a identificação de observações aberrantes e verificar o ajuste dos modelos; veja, por exemplo, Pinheiro et al. (2001), Zeller (2009) e Zeller et al. (2011).

Considerando as proposições desenvolvidas para a distribuição normal assimétrica padrão, podemos generalizá-las para a distribuição normal assimétrica locação-escala.

Proposição 2.3: Se $Z \sim SN(\lambda)$ e $Y = \mu + \sigma Z$ então $Y \sim SN(\mu, \sigma^2, \lambda)$.

Demonstração: A prova desta proposição é direta via transformação de variáveis.

A partir das Proposições 2.2 e 2.3, podemos gerar números pseudo-aleatórios de uma distribuição normal assimétrica locação-escala.

A próxima proposição nos fornece a função geradora de momentos de uma distribuição normal assimétrica locação-escala.

Proposição 2.4: A função geradora de momentos de $Y \sim SN(\mu, \sigma^2, \lambda)$ é dada por

$$M_Y(t) = 2 \exp\left(t\mu + \frac{t^2\sigma^2}{2}\right) \Phi(\delta\sigma t), \quad t \in \mathbb{R},$$

onde $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

Demonstração: A demonstração desta proposição é direta via transformação de variáveis.

Utilizando a Proposição 2.3 ou a Proposição 2.4 pode-se mostrar que a média e a variância de uma variável aleatória $Y \sim SN(\mu, \sigma^2, \lambda)$ são expressas por

$$E(Y) = \mu + \sqrt{\frac{2}{\pi}} \delta \sigma \quad (2.6)$$

e

$$Var(Y) = \sigma^2 \left(1 - \frac{2}{\pi} \delta^2\right). \quad (2.7)$$

Note que, quando $\lambda = 0$ e conseqüentemente $\delta = 0$, obtemos a função geradora de momentos de $X \sim N(\mu, \sigma^2)$, isto é, $M_X(t) = \exp\left(t\mu + \frac{t^2\sigma^2}{2}\right)$. Além disso, $E(X) = \mu$ e $Var(X) = \sigma^2$.

Uma propriedade interessante, apresentada a seguir, é que a variável aleatória normal assimétrica é invariante sob transformações lineares.

Proposição 2.5: Sejam Y e Y_2 duas variáveis aleatórias, tais que $Y \sim SN(\mu, \sigma^2, \lambda)$ e $Y_2 = a + bY$, então

$$Y_2 \sim SN(a + b\mu, b^2\sigma^2, \text{sign}(b)\lambda).$$

Demonstração: A prova desta proposição pode ser obtida em Rodríguez (2005), página 16.

O resultado descrito na Proposição 2.5 é útil na definição do modelo de regressão linear no contexto de assimetria. Considerando $\varepsilon_i \stackrel{iid}{\sim} SN(0, \sigma^2, \lambda)$ e $Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, i = 1, \dots, n$, então

$$Y_i \stackrel{ind}{\sim} SN(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2, \lambda),$$

onde “*ind*” significa independentes.

2.2.1 Caracterização

Os resultados apresentados nessa seção são de suma importância no decorrer deste trabalho e serão utilizados consistentemente nos capítulos seguintes.

Proposição 2.6: Seja $X_0, X_1 \stackrel{iid}{\sim} N(0,1)$ então

$$Y = \mu + \sigma \left[\delta |X_0| + (1 - \delta^2)^{\frac{1}{2}} X_1 \right],$$

tem distribuição normal assimétrica locação-escala, ou seja, $Y \sim SN(\mu, \sigma^2, \lambda)$, onde $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

Demonstração: A prova desta proposição é direta via os resultados das Proposições 2.2 e 2.3.

A representação estocástica dada na Proposição 2.6 além de facilitar a implementação do algoritmo EM pode ser usada também para derivar muitas propriedades da distribuição de Y , por exemplo, da forma quadrática da distância de Mahalanobis d , definida em (2.5). A partir da Proposição 2.6, a distribuição normal assimétrica locação-escala pode ser reescrita hierarquicamente como

$$Y|T = t \sim N(\mu + \delta\sigma t, \sigma^2(1 - \delta^2)) \quad (2.8)$$

$$T \sim HN(0,1), \quad (2.9)$$

onde pela notação da Proposição 2.6, temos que $T \stackrel{d}{=} |X_0|$ e $X_0 \sim N(0,1)$, sendo que “ $\stackrel{d}{=}$ ” significa equivalência em distribuição. Tal representação hierárquica será útil para realizar inferência estatística nos modelos de regressão linear, no contexto de assimetria.

3. Modelo de Regressão Linear Assimétrico

Os modelos normais de regressão linear podem ser considerados como uma das ferramentas mais populares para a análise estatística de dados. No entanto, os fenômenos sob investigação nem sempre estão de acordo com o modelo normal, seja devido a falta de assimetria ou a presença de observações atípicas. Para lidar com este problema, algumas propostas foram elaboradas na literatura substituindo a suposição de normalidade, por exemplo, por classes paramétricas mais flexíveis de distribuições.

Sob essa motivação é que, neste capítulo, será apresentado o modelo de regressão linear assimétrico, onde consideramos que a variável resposta segue uma distribuição normal assimétrica.

Assim, iniciamos este capítulo com a descrição do modelo de regressão linear assimétrico, em seguida, definimos a função de log-verossimilhança e descrevemos o algoritmo EM para a obtenção das estimativas de máxima verossimilhança dos parâmetros do modelo. Além disso, apresentamos duas aplicações considerando conjunto de dados reais, sendo uma delas a um conjunto de dados bastante conhecido na literatura, com informações provenientes de um grupo de atletas australianos.

3.1 Descrição do Modelo

O modelo de regressão linear é representado por

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, i = 1, \dots, n, \quad (3.1)$$

onde Y_i é a variável resposta para a observação i , $\mathbf{x}_i^T = (x_{1i}, \dots, x_{pi})$ representa os valores de p ($p < n$) variáveis explicativas que são assumidas fixas e conhecidas, ou seja, \mathbf{x}_i^T corresponde à i -ésima linha da matriz \mathbf{X} ($n \times p$), $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{p-1})^T$ é um vetor de parâmetros desconhecidos a serem estimados ($\boldsymbol{\beta} \in \mathbb{R}^p$). Neste trabalho, vamos substituir a usual suposição de normalidade para os erros (ε_i), pela seguinte suposição mais geral

$$\varepsilon_i \stackrel{iid}{\sim} SN(b\Delta, \sigma^2, \lambda), \quad (3.2)$$

onde $b = -\sqrt{\frac{2}{\pi}}$. Note que, a partir de (3.2), temos que o $E(\varepsilon_i) = 0$. Além disso, pela Proposição 2.5, temos que

$$Y_i \stackrel{ind}{\sim} SN(\mathbf{x}_i^T \boldsymbol{\beta} + b\Delta, \sigma^2, \lambda) \quad (3.3)$$

e assim a partir de (3.2) e (3.3), obtemos

$$E(Y_i) = \mathbf{x}_i^T \boldsymbol{\beta} \quad (3.4)$$

e

$$Var(Y_i) = Var(\varepsilon_i) = \sigma^2 - b^2 \Delta^2. \quad (3.5)$$

3.2 Função Log-Verossimilhança

A função de log-verossimilhança para $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2, \lambda)^T$, dada uma amostra Y_1, \dots, Y_n , é definida por $l(\boldsymbol{\theta}) = \sum_{i=1}^n l_i(\boldsymbol{\theta})$, em que

$$l_i(\boldsymbol{\theta}) = \log 2 - \frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 + \log k_i, \quad (3.6)$$

com $k_i = \exp\left\{-\frac{1}{2}d_i\right\} \Phi(A_i)$, onde

$$d_i = \frac{1}{\sigma^2} (Y_i - \mathbf{x}_i^T \boldsymbol{\beta} - b\Delta)^2 \quad (3.7)$$

é a distância de Mahalanobis e $A_i = (Y_i - \mathbf{x}_i^T \boldsymbol{\beta} - b\Delta) \frac{\lambda}{\sigma}$.

A função escore pode ser obtida derivando-se o logaritmo da função de verossimilhança, ou seja, a função log-verossimilhança, com respeito a cada um dos parâmetros desconhecidos. Conforme Rodríguez (2005), Ferreira (2008) e Zeller (2009), por exemplo, notamos que não existem soluções explícitas para o problema de maximização da função de log-verossimilhança dada em (3.6). Neste caso, podemos maximizar numericamente usando softwares existentes como, por exemplo, o Matlab e o R. Esses

softwares contém rotinas prontas para tratar problemas de maximização (minimização) de qualquer função.

Entretanto, similarmente ao que foi feito por Rodríguez (2005), Ferreira (2008) e Zeller (2009), por exemplo, apresentamos aqui um algoritmo EM para o cálculo dos estimadores de máxima verossimilhança nos valendo da representação estocástica dada na Proposição 2.6 e introduzindo uma reparametrização de modo a obtermos formas fechadas para o algoritmo EM no passo M.

3.3 Inferência pelo Método de Máxima Verossimilhança

Essa é umas das seções mais importantes para o desenvolvimento desse trabalho. Aqui será apresentada a representação hierárquica do modelo de regressão linear assimétrico definido em (3.1) e (3.2). Assim, a partir desta representação pode-se tratar o problema de estimação dos parâmetros via algoritmo EM. Tais resultados serão utilizados no Capítulo 4 e uma vez compreendidos nesta seção, torna-se mais fácil o entendimento do processo de estimação dos parâmetros em modelos de regressão linear assimétrico com ponto de mudança.

3.3.1 Representação Hierárquica

Utilizando os resultados da Proposição 2.6, temos que o modelo de regressão linear assimétrico, definido em (3.1) e (3.2), pode ser reescrito hierarquicamente como

$$Y_i | T_i = t_i \sim N(\mathbf{x}_i^T \boldsymbol{\beta} + \Delta t_i, \Gamma), \quad (3.8)$$

$$T_i \sim NT(b, 1, (b, +\infty)), \quad (3.9)$$

onde consideramos a seguinte reparametrização

$$\Delta = \sigma \delta \quad (3.10)$$

e

$$\Gamma = (1 - \delta^2) \sigma^2, \quad (3.11)$$

sendo que $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

3.3.2 Estimação dos Parâmetros Via Algoritmo EM

Neste processo de estimação, considere $\mathbf{y} = (y_1, \dots, y_n)^T$ o vetor de respostas observáveis para n unidades amostrais e $\mathbf{t} = (t_1, \dots, t_n)^T$. Então, sob a representação hierárquica (3.8) – (3.9), segue que a função log-verossimilhança completa associada com $\mathbf{y}_c = (\mathbf{y}^T, \mathbf{t}^T)$ é dada por

$$l_c(\boldsymbol{\theta}|\mathbf{y}_c) = c - \frac{n}{2} \log \Gamma - \frac{1}{2\Gamma} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta} - \Delta t_i)^2, \quad (3.12)$$

onde c é uma constante que não depende de $\boldsymbol{\theta}$.

No passo E do algoritmo EM, toma-se o valor esperado da log-verossimilhança completa dada em (3.12) condicionada no vetor de dados observados e no $\boldsymbol{\theta}$ em estado corrente. Assim, o valor esperado condicional da log-verossimilhança completa é

$$\begin{aligned} Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(t)}) &= E(l_c(\boldsymbol{\theta}, \mathbf{y}_c)|\mathbf{y}, \hat{\boldsymbol{\theta}}^{(t)}) = \\ &c - \frac{n}{2} \log \Gamma - \frac{1}{2\Gamma} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 + \frac{\Delta}{\Gamma} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta}) \hat{t}_i^{(t)} - \frac{\Delta^2}{2\Gamma} \sum_{i=1}^n \hat{t}_i^2{}^{(t)}. \end{aligned} \quad (3.13)$$

Note que as seguintes quantidades devem ser obtidas

$$\hat{t}_i^{(t)} = E(T_i|y_i, \hat{\boldsymbol{\theta}}^{(t)}) \quad (3.14)$$

e

$$\hat{t}_i^2{}^{(t)} = E(T_i^2|y_i, \hat{\boldsymbol{\theta}}^{(t)}). \quad (3.15)$$

Para o cálculo das quantidades (3.14) e (3.15), é necessário obter a distribuição condicional $T_i|Y_i = y_i$. A obtenção de tal distribuição condicional é baseada no seguinte lema.

Lema 1: Seja $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e $\mathbf{X} \sim N_q(\boldsymbol{\eta}, \boldsymbol{\Omega})$. Então,

$$\begin{aligned} \phi_p(\mathbf{y}|\boldsymbol{\mu} + \mathbf{A}\mathbf{x}, \boldsymbol{\Sigma})\phi_q(\mathbf{x}|\boldsymbol{\eta}, \boldsymbol{\Omega}) &= \phi_p(\mathbf{y}|\boldsymbol{\mu} + \mathbf{A}\boldsymbol{\eta}, \boldsymbol{\Sigma} + \mathbf{A}\boldsymbol{\Omega}\mathbf{A}^T) \\ &\times \phi_q(\mathbf{x}|\boldsymbol{\eta}, \boldsymbol{\Lambda}\mathbf{A}^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu} - \mathbf{A}\boldsymbol{\eta}), \boldsymbol{\Lambda}), \end{aligned}$$

onde $\mathbf{\Lambda} = (\mathbf{\Omega}^{-1} + \mathbf{A}^T \mathbf{\Sigma}^{-1} \mathbf{A})^{-1}$ e $\phi_k(\cdot | \boldsymbol{\mu}, \mathbf{\Sigma})$ é a função densidade de probabilidade da normal k -variada com vetor de médias $\boldsymbol{\mu}$ e matriz de covariâncias $\mathbf{\Sigma}$.

Demonstração: A prova deste lema pode ser encontrada em Lachos (2004), página 22.

Das equações (3.8) e (3.9) da representação hierárquica do modelo de regressão linear assimétrico e do Lema 1 tem-se que

$$\begin{aligned} f(y_i, t_i | \boldsymbol{\theta}) &= f(y_i | t_i, \boldsymbol{\theta}) f(t_i | \boldsymbol{\theta}) \\ &= f(y_i | \boldsymbol{\theta}) f(t_i | y_i, \boldsymbol{\theta}). \end{aligned}$$

Portanto,

$$\begin{aligned} f(y_i, t_i | \boldsymbol{\theta}) &= 2\phi(y_i | \mathbf{x}_i^T \boldsymbol{\beta} + \Delta t_i, \Gamma) \phi(t_i, b, 1) I_{(b, +\infty)}(t_i) \\ &= 2\phi(y_i | \mathbf{x}_i^T \boldsymbol{\beta} + \Delta b, \Gamma + \Delta^2) \phi(t_i | \mu_{T_i} + b, M_T^2) I_{(b, +\infty)}(t_i), \end{aligned}$$

onde

$$\mu_{T_i} = \frac{\Delta}{\Delta^2 + \Gamma} (y_i - \mathbf{x}_i^T \boldsymbol{\beta} - b\Delta)$$

e

$$M_T^2 = \frac{\Gamma}{\Delta^2 + \Gamma},$$

sendo que $I_A(\cdot)$ é a função indicadora de um conjunto A e por convecção, $\phi(\cdot | \boldsymbol{\mu}, \mathbf{\Sigma}) = \phi_1(\cdot | \boldsymbol{\mu}, \mathbf{\Sigma})$. Dessa forma, segue que

$$T_i | Y_i = y_i \sim NT \left(\mu_{T_i} + b, M_T^2, (b, +\infty) \right). \quad (3.16)$$

O próximo resultado está relacionado com as propriedades da distribuição normal truncada.

Lema 2: Seja $X \sim N(\eta, \tau^2)$. Então, para a constante real, segue que

$$E(X | X > a) = \eta + \frac{\phi_1\left(\frac{a - \eta}{\tau}\right)}{1 - \Phi_1\left(\frac{a - \eta}{\tau}\right)} \tau$$

e

$$E(X^2|X > a) = \eta^2 + \tau^2 + \frac{\phi_1\left(\frac{a-\eta}{\tau}\right)}{1 - \Phi_1\left(\frac{a-\eta}{\tau}\right)}(\eta + a)\tau.$$

Demonstração: A prova deste lema pode ser encontrada em Lachos (2004), página 44.

A partir de (3.16) e do Lema 2, temos que

$$\hat{t}_i^{(t)} = E(T_i|y_i, \hat{\boldsymbol{\theta}}^{(t)}) = \hat{\mu}_{T_i}^{(t)} + b + \hat{M}_T^{(t)} \hat{t}_{1i}^{(t)} \quad (3.17)$$

e

$$\hat{t}_i^2{}^{(t)} = E(T_i^2|y_i, \hat{\boldsymbol{\theta}}^{(t)}) = (\hat{\mu}_{T_i}^{(t)} + b)^2 + \hat{M}_T^2{}^{(t)} + \hat{M}_T^{(t)}(\hat{\mu}_{T_i}^{(t)} + 2b)\hat{t}_{1i}^{(t)}, \quad (3.18)$$

em que,

$$\hat{t}_{1i}^{(t)} = W_{\Phi}\left(\frac{\hat{\mu}_{T_i}^{(t)}}{\hat{M}_T^{(t)}}\right), \text{ sendo } W_{\Phi}(\cdot) = \frac{\phi(\cdot)}{\Phi(\cdot)}.$$

O passo M do algoritmo envolve a maximização em $\boldsymbol{\theta}$ da esperança condicional da log-verossimilhança completa dada em (3.13). Portanto, o algoritmo EM fica dado por:

Passo E: Dado $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(t)}$, obter $\hat{t}_i^{(t)}$ e $\hat{t}_i^2{}^{(t)}$, $i = 1, \dots, n$, usando (3.17) e (3.18).

Passo M: Atualizar $\hat{\boldsymbol{\theta}}^{(t)}$, maximizando $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(t)})$ sobre $\boldsymbol{\theta}$, que resulta nas seguintes expressões fechadas:

$$\hat{\boldsymbol{\beta}}^{(t+1)} = \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \sum_{i=1}^n \mathbf{x}_i (y_i - \hat{\Delta}^{(t)} \hat{t}_i^{(t)}), \quad (3.19)$$

$$\hat{\Delta}^{(t+1)} = \frac{1}{\sum_{i=1}^n \hat{t}_i^2{}^{(t)}} \left[\sum_{i=1}^n \hat{t}_i^{(t)} (y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}^{(t)}) \right], \quad (3.20)$$

$$\hat{\Gamma}^{(t+1)} = \frac{1}{n} \left[\sum_{i=1}^n (y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}^{(t)})^2 - 2\hat{\Delta}^{(t)} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}^{(t)}) \hat{t}_i^{(t)} + \hat{\Delta}^2{}^{(t)} \sum_{i=1}^n \hat{t}_i^2{}^{(t)} \right]. \quad (3.21)$$

Note que as estimativas para σ^2 e λ podem ser obtidas usando (3.20) e (3.21), e a propriedade de invariância dos estimadores de máxima verossimilhança. Dessa forma, temos que

$$\hat{\lambda}^{(t+1)} = \frac{\hat{\Delta}^{(t+1)}}{\hat{\Gamma}^{(t+1)}} \quad (3.22)$$

e

$$\hat{\sigma}^{(t+1)} = \hat{\Delta}^{2^{(t+1)}} + \hat{\Gamma}^{(t+1)}. \quad (3.23)$$

As interações são repetidas até que uma regra de convergência adequada seja satisfeita. Nesse trabalho, adotamos a regra de convergência definida em (1.1). Valores iniciais são necessários para implementar esse algoritmo. Eles são obtidos sob a suposição de normalidade e considerando como valor inicial para λ , ou seja, $\hat{\lambda}^{(0)} = \hat{\rho}$, onde $\hat{\rho}$ é o coeficiente de assimetria amostral para a variável resposta. Entretanto, com a finalidade de verificar que a estimativa de máxima verossimilhança foi encontrada, recomenda-se rodar o algoritmo EM usando uma amplitude de diferentes valores iniciais. Note que, quando $\lambda = 0$ (ou $\Delta = 0$), as equações do passo M se reduzem às equações quando assumimos a distribuição normal, então $\hat{\beta} = [X^T X]^{-1} X^T y$, onde X está definida em (3.1) e $y = (y_1, \dots, y_n)^T$.

Na próxima seção, apresentaremos aplicações em conjunto de dados reais.

3.4 Aplicações em Dados Reais

3.4.1 Aplicação 1: AIS (“Australian Institute of Sport”)

Nesta seção, o modelo de regressão linear assimétrico será aplicado a um subconjunto de variáveis do conjunto de dados do “Australian Institute of Sport” (AIS). Estes dados têm sido utilizados em ilustrações numéricas por um grande número de programas estatísticos e autores. Por exemplo, um subconjunto desses dados foi previamente analisado em Azzalini e Capitanio (2003) e Ferreira (2008), no contexto de assimetria.

Os dados estão disponíveis no endereço <http://azzalini.stat.unipd.it/index-en.html> e referem-se à 202 atletas (102 homens e 100 mulheres) dos quais foram coletadas diversas informações, tais como peso, altura, porcentagem de gordura no corpo, dentre outras. Como

em Ferreira (2008), com o intuito de ilustrar a metodologia, utilizaremos as variáveis soma de manchas de pele (ssf), variável de interesse, e porcentagem de gordura no corpo (Bfat).

Tabela 3.1: Estatísticas descritivas da variável ssf do conjunto de dados AIS.

	1° Quartil	Mediana	3° Quartil	Mínimo	Média	Máximo	Erro Padrão
ssf	43,85	58,60	90,35	28,00	69,02	200,80	32,57

Com propósitos exploratórios, a Tabela 3.1 apresenta estatísticas descritivas da variável de interesse soma de manchas na pele. Note que os resultados sugerem uma assimetria à direita (média > mediana). Sendo assim, o ajuste do modelo assimétrico para esses dados parece adequado. O modelo proposto é o seguinte

$$Y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i, i = 1, \dots, 202,$$

onde Y_i é a soma de manchas na pele para o i -ésimo atleta, x_{1i} representa a porcentagem de gordura no corpo do i -ésimo atleta e assumimos que os erros de observação seguem uma distribuição normal assimétrica. Nesta aplicação, para propósitos comparativos, iremos também considerar o modelo normal.

Ajuste do Modelo

A Tabela 3.2 apresenta as estimativas de máxima verossimilhança para os parâmetros dos dois modelos selecionados, isto é, normal e normal assimétrico, acompanhadas dos seus respectivos erros padrão obtidos via a técnica de reamostragem “Bootstrap” (100 replicações).

Tabela 3.2: Estimativas de máxima verossimilhança para os dois modelos selecionados. Os valores SE, entre parênteses, são os erros padrão estimados baseados no procedimento “Bootstrap”.

Parâmetros	Normal (SE)	Normal Assimétrica (SE)
β_0	0,5859 (1,7940)	3,0902 (1,4267)
β_1	5,0665 (0,1001)	4,8843 (0,0938)
σ^2	76,6097 (7,7251)	143,9721 (31,4690)
λ	-	1,6395 (0,6742)

Embora as estimativas do parâmetro de regressão β_1 sejam semelhantes nos dois modelos, o erro padrão do modelo normal assimétrico é menor que o obtido sob o modelo normal. Isto sugere que o modelo normal assimétrico parece produzir estimativas de máxima verossimilhança mais precisas do que o modelo normal. As estimativas para os componentes da variância (σ^2 e λ) não são comparáveis, pois estão em escalas distintas.

Em seguida, comparando os modelos normal e normal assimétrico através dos valores dos critérios de informação SIC e AIC, dados na Tabela 3.3, notamos que o modelo assimétrico apresenta um melhor ajuste que o modelo simétrico, sugerindo dessa forma, a validação da suposição usual de assimetria dos dados. Tais critérios foram calculados de acordo com as expressões definidas em (1.3) e (1.4), respectivamente.

Tabela 3.3: Alguns critérios de informação.

Modelos	$l(\hat{\theta})$	SIC	AIC
Normal	-724,8367	1457,6730	1458,8950
Normal Assimétrico	-676,8924	1363,0060	1361,7850

Análise de Diagnóstico

Com a finalidade de identificar as observações aberrantes (comumente chamadas de “outliers” ou observações extremas, anômalas ou atípicas), a distância de Mahalanobis (d_i , $i = 1, \dots, 202$) foi considerada; veja Seções 2.2 e 3.2 para mais detalhes. A Figura 3.2 ilustra tal distância para os dois modelos ajustados e as linhas de ponto de corte correspondem ao quantil v da distribuição da forma quadrática d_i . De acordo com a propriedade (3) citada na Seção 2.1.1, temos que $d_i \sim \chi_1^2$, para os modelos normal e normal assimétrico, sendo assim, o nonagésimo quinto quantil de d_i é $v = 3.841$.

Podemos observar pelos gráficos que as observações 11, 43, 51, 53, 56 e 178 são detectadas como “outliers”, para ambos os modelos, pois apresentam valores de d_i maiores ou próximos (na fronteira) da marca de referência. Além disso, Lucas (1997) ressalta que “outliers” são definidos com respeito aos modelos. Dessa forma, observações podem ser “outliers” em um modelo e serem perfeitamente regulares para outro modelo.

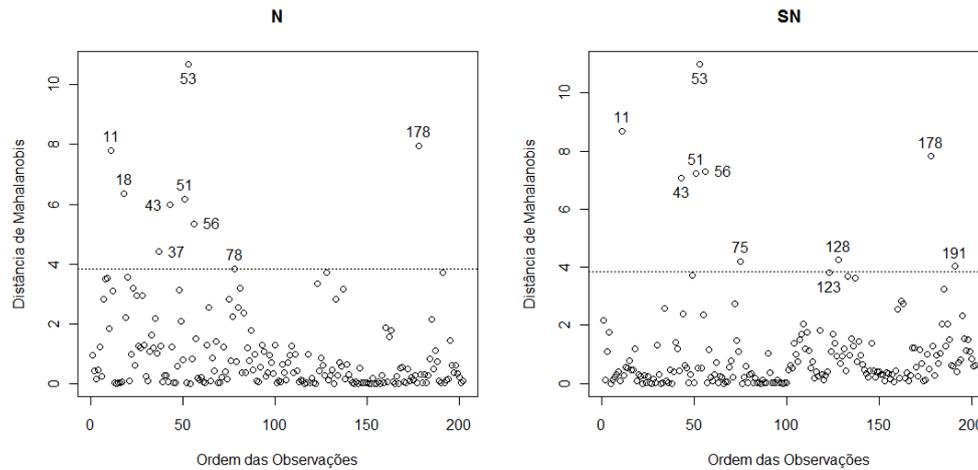


Figura 3.2: Gráfico da Distância de Mahalanobis versus Ordem das Observações para os modelos normal (N) e normal assimétrico (SN).

Como sugerido por Lee et al.(2006), usamos as quantidades TRC e MRC para revelar o impacto das observações atípicas detectadas na estimativa de máxima verossimilhança de θ . Elas são definidas por

$$TRC = \sum_{j=1}^{n_p} \left| \frac{\hat{\theta}_j - \hat{\theta}_{[i]j}}{\hat{\theta}_j} \right| \text{ e } MRC = \max_{j=1, \dots, n_p} \left| \frac{\hat{\theta}_j - \hat{\theta}_{[i]j}}{\hat{\theta}_j} \right|, \quad (3.24)$$

onde n_p é a dimensão de θ e o subscrito $[i]$ indica que a estimativa de verossimilhança de θ foi obtida excluindo i -ésima observação, y_i . Na Tabela 3.4, as comparações dessas medidas são dadas, considerando os modelos selecionados, quando excluimos as observações 11, 43, 51, 53, 56 e 178. Note que a maior mudança ocorreu sob a distribuição normal. Assim, sob o modelo normal assimétrico, as observações atípicas parecem ter menos impacto no processo de estimação comparado com o modelo normal.

Tabela 3.4: Comparação das mudanças relativas nas estimativas de máxima verossimilhança em termos de quantidades TRC e MRC.

Modelos	TRC	MRC
Normal	6,1955	5,8894
Normal Assimétrico	1,6273	0,7196

3.4.2 Aplicação 2: Pássaros em Extinção

Nesta seção, vamos analisar o seguinte conjunto de dados citado no texto de Albert (2007, Ex. 9.2.6). Trata-se de um estudo sobre pássaros em extinção, originalmente publicado em Pimm et al. (1988), realizado ao longo de várias décadas, observando pares de espécies de pássaros em 16 ilhas próximas à Grã-Betanha. Para cada espécie as seguintes variáveis foram observadas:

1. Tempo (variável de interesse): tempo médio, em anos, de extinção na ilha onde a espécie apareceu;
2. Aninhamento: número médio a pares aninhados;
3. Tamanho: tamanho da espécie (0 = pequeno ou 1 = grande);
4. Status: status migratório da espécie (0 = migrantes ou 1 = residente).

Os dados referem-se à 62 observações, no qual a variável de interesse $Y = \text{tempo}$ está relacionada com outras três variáveis, denominadas $x_1 = \text{aninhamento}$, $x_2 = \text{tamanho}$ e $x_3 = \text{status}$. Os dados estão disponíveis no pacote “LearnBayes” do software R, com o nome “birdextinct”. Albert (2007) modela este conjunto de dados fazendo uma transformação logarítmica na variável resposta, com o objetivo de atingir a normalidade. Aqui, nós vamos evitar este tipo de procedimento e fazer um ajuste baseado no modelo de regressão linear assimétrico proposto anteriormente. O objetivo é mostrar que o modelo assimétrico fornece um ajuste mais adequado que um modelo menos flexível, aqui utilizamos o normal como parâmetro de comparação.

Tabela 3.5: Estatísticas descritivas da variável tempo do conjunto de dados dos Pássaros em Extinção.

	1° Quartil	Mediana	3° Quartil	Mínimo	Média	Máximo	Erro Padrão
Tempo	1,91	3,18	6,99	1,00	6,96	58,82	10,70

Com propósitos exploratórios, a Tabela 3.5 apresenta as estatísticas descritivas da variável de interesse tempo. Note que os resultados sugerem um padrão assimétrico (média > mediana). Sendo assim, o modelo proposto é o seguinte:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i, i = 1, \dots, 62,$$

onde assumiremos que os erros de observação seguem distribuição normal assimétrica.

Ajuste do Modelo

A Tabela 3.6 contém as estimativas de máxima verossimilhança para os parâmetros dos dois modelos selecionados, isto é, normal e normal assimétrico, acompanhados dos seus respectivos erros padrão obtidos via a técnica de reamostragem “Bootstrap” (100 replicações).

Pela Tabela 3.6, notamos que há evidências de assimetria. Além disso, Lachos et al. (2013) analisaram o mesmo conjunto de dados considerando que os erros seguem uma distribuição na classe das distribuições assimétricas misturas de escala skew-normal, no contexto bayesiano. Novamente, observe que os erros padrão sob o modelo normal assimétrico são menores que os obtidos sob o modelo normal. De um modo geral, estes resultados favorecem a evidência de que o modelo assimétrico é mais apropriado para o conjunto de dados Pássaros em Extinção.

Tabela 3.6: Estimativas de máxima verossimilhança para os dois modelos selecionados. Os valores SE, entre parênteses, são os erros padrão baseados no procedimento “Bootstrap”.

Parâmetros	Normal (SE)	Normal Assimétrica (SE)
β_0	-0,6276 (2,5794)	4,9238 (1,3607)
β_1	2,1337 (0,5200)	1,0041 (0,2675)
β_2	-3,9275 (2,2436)	-1,1728 (1,1410)
β_3	3,5298 (2,1334)	1,5061 (1,2554)
σ^2	78,7233 (16,1414)	127,5961 (18,1298)
λ	-	6,5820 (6,4080)

Em seguida, comparamos os modelos normal e normal assimétrico, inspecionando os critérios de informação descritos na Seção 1.4. Comparando os modelos simétrico e assimétrico através dos valores dos critérios de informação dadas na Tabela 3.7, notamos que o modelo assimétrico apresenta um ajuste melhor do que o modelo simétrico.

Tabela 3.7: Alguns critérios de informação.

Modelos	$l(\hat{\theta})$	SIC	AIC
Normal	-223,3183	457,3910	458,6370
Normal Assimétrico	-190,6478	393,2956	392,0500

Análise de Diagnóstico

Identificamos as observações atípicas considerando a distância de Mahalanobis d_i definida em (2.5). A Figura 3.3 mostra tal distância para os dois modelos ajustados e as linhas de ponto de corte correspondem ao 95° quantil, $v = 3.841$, da forma quadrática d_i . Dessa forma, as observações que apresentam valores de d_i maiores ou próximos (na fronteira) da marca de referência serão identificados como “outliers”. Note que as observações 20, 28, 29 e 60 aparecem como “outliers” para os modelos normal e normal assimétrico.

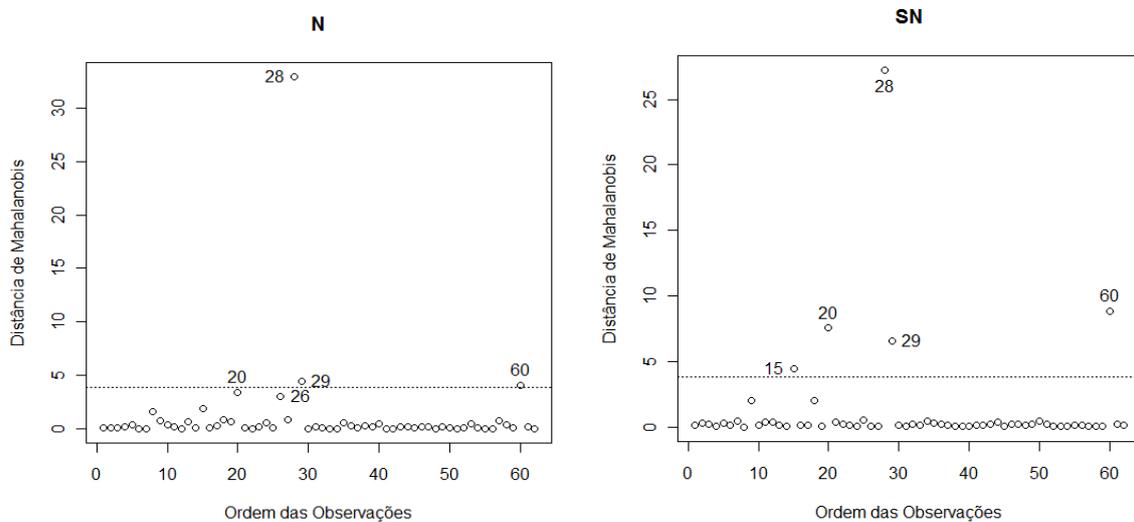


Figura 3.3: Gráfico da Distância de Mahalanobis versus Ordem das Observações para os modelos normal (N) e normal assimétrico (SN).

Tabela 3.8: Comparação das mudanças relativas nas estimativas de máxima verossimilhança em termos de quantidades TRC e MRC.

Modelos	TRC	MRC
Normal	7,4094	5,1687
Normal Assimétrico	2,5004	0,8004

Com o intuito de avaliar o impacto das observações atípicas detectadas na estimativa de máxima verossimilhança de θ , as quantidades TRC e MRC, definidas em (3.24), foram calculadas para os dois modelos selecionados. Na Tabela 3.8, as comparações dessas medidas são dadas quando excluímos as observações 20, 28, 29 e 60. Observe que a maior mudança

ocorreu sob a distribuição normal. Como esperado, os resultados indicam que as estimativas de máxima verossimilhança sob o modelo assimétrico são menos sensíveis à presença de dados atípicos.

4. Modelo de Regressão Linear Normal Assimétrico com Ponto de Mudança

O problema de ponto de mudança tem sido um tópico de interesse permanente na literatura estatística. Muitos autores têm estudado o problema de ponto de mudança associado com os modelos de regressão, sob o contexto de assimetria dos dados. Por exemplo, Chen (1998) propõe o uso do SIC para detectar o ponto de mudança no modelo de regressão linear sob a suposição de normalidade. Veja também Chen e Gupta (2001) para mais detalhes. Por sua vez o uso da distribuição Student-t, como alternativa à distribuição normal, tem frequentemente sido sugerido na literatura. No contexto de ponto de mudança, Osorio e Galea (2005) propõe o uso do SIC para detectar o ponto de mudança no modelo de regressão linear Student-t.

Quando a distribuição dos dados tem um comportamento assimétrico, esta pode ser modelada através de membros da classe de distribuições assimétricas. Nesta classe, a distribuição normal assimétrica é talvez a pioneira nessa ideia. Recentemente, Arellano et al. (2013) sob o ponto de vista bayesiano, realizaram um estudo de detecção de ponto de mudança no modelo normal assimétrico localização-escala.

Na literatura estatística, existem poucos trabalhos aplicados sobre ponto de mudança no contexto de assimetria, onde se baseia o nosso objetivo.

Considere a sequência de observações $(x_1^T, Y_1), \dots, (x_n^T, Y_n)$. O objetivo deste trabalho é testar as hipóteses da forma

$$H_0: Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, i = 1, \dots, n, \quad (4.1)$$

isto é, os coeficientes de regressão não se alteram, versus

$$H_1: Y_i = \mathbf{x}_i^T \boldsymbol{\beta}_1 + \varepsilon_i, i = 1, \dots, k, \quad (4.2)$$

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta}_2 + \varepsilon_i, i = k + 1, \dots, n,$$

onde $\boldsymbol{\beta}_1 = (\beta_0, \beta_1, \dots, \beta_{p-1})^T$ e $\boldsymbol{\beta}_2 = (\beta_0^*, \beta_1^*, \dots, \beta_{p-1}^*)^T$, ou seja, existe mudança (nos coeficientes de regressão) em uma posição desconhecida k , denominada ponto de mudança.

Nessa linha de pesquisa, desenvolvemos alguns resultados adicionais para o problema de ponto de mudança no contexto de assimetria. Em particular, propomos estudar a detecção de ponto de mudança no modelo de regressão linear normal assimétrico.

Assim, pretendemos contribuir positivamente para o desenvolvimento da área dos modelos lineares, estendendo alguns resultados encontrados na literatura, por exemplo, Chen (1998), Chen e Gupta (2001), Osorio e Galea (2005) e Arellano et al. (2013).

4.1 Descrição do Modelo com Ponto de Mudança

Considere o modelo de regressão, descrito na Seção 3.1,

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, i = 1, \dots, n,$$

onde Y_i é a variável resposta para a observação i , $\mathbf{x}_i^T = (x_{1i}, \dots, x_{pi})$ representa os valores de p ($p < n$) variáveis explicativas que são assumidas fixas e conhecidas, ou seja, \mathbf{x}_i^T corresponde à i -ésima linha da matriz \mathbf{X} ($n \times p$), $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{p-1})^T$ é um vetor de parâmetros desconhecidos a serem estimados ($\boldsymbol{\beta} \in \mathbb{R}^p$). Neste trabalho, assume-se que os erros aleatórios $\varepsilon_i \stackrel{iid}{\sim} SN(b\Delta, \sigma^2, \lambda)$, e consequentemente $Y_i \stackrel{ind}{\sim} SN(\mathbf{x}_i^T \boldsymbol{\beta} + b\Delta, \sigma^2, \lambda)$, $i = 1, \dots, n$, tal que $b = -\sqrt{\frac{2}{\pi}}$.

Para as hipóteses de ponto de mudança (equações (4.1) e (4.2)) considere a seguinte notação, onde $k = p, \dots, n - p$,

$$Y_1 = (Y_1, Y_2, \dots, Y_n)^T, \quad Y_2 = (Y_{k+1}, Y_2, \dots, Y_n)^T,$$

$$X_1 = \begin{pmatrix} x_1^T \\ \vdots \\ x_k^T \end{pmatrix} \quad \text{e} \quad X_2 = \begin{pmatrix} x_{k+1}^T \\ \vdots \\ x_n^T \end{pmatrix}.$$

Sob H_0 , há um modelo que não apresenta qualquer alteração nos coeficientes de regressão. Por outro lado, sob H_1 há uma relação de modelos com pontos de mudanças nas posições p ou $p + 1$ ou $n - p$. O objetivo é, portanto, selecionar um modelo proveniente dos grupos anteriores.

A seguir, os estimadores de máxima verossimilhança de $\boldsymbol{\theta}$ sob ambas as hipóteses serão descritos. O algoritmo EM é usado para estimar os parâmetros devido a sua simplicidade para implementação.

4.2 Estimação dos Parâmetros Via Algoritmo EM

Em um primeiro estágio do trabalho, os estimadores de máxima verossimilhança de θ sob H_0 estão apresentados na Seção 3.3.2 (equações (3.19) - (3.23)). Observe que sob H_0 temos $\theta = (\beta^T, \sigma^2, \lambda)^T$.

Agora, considere o modelo sob a hipótese alternativa, isto é, com ponto de mudança na posição k ($k = p, \dots, n - p$). Neste caso, para $\theta = (\beta_1^T, \beta_2^T, \sigma^2, \lambda)^T$ e sob a representação hierárquica (3.8) – (3.9), onde

$$\mu_i = \begin{cases} \mathbf{x}_i^T \beta_1, & i = 1, \dots, k, \\ \mathbf{x}_i^T \beta_2, & i = k + 1, \dots, n, \end{cases}$$

segue que a função log-verossimilhança completa associada com $\mathbf{y}_c = (\mathbf{y}^T, \mathbf{t}^T)$, tal que $\mathbf{y} = (y_1, \dots, y_n)^T$ e $\mathbf{t} = (t_1, \dots, t_n)^T$, é dada por

$$l_c(\theta | \mathbf{y}_c) = c - \frac{n}{2} \log \Gamma - \frac{1}{2\Gamma} \left[\sum_{i=1}^k (y_i - \mathbf{x}_i^T \beta_1 - \Delta t_i)^2 + \sum_{i=k+1}^n (y_i - \mathbf{x}_i^T \beta_2 - \Delta t_i)^2 \right], \quad (4.3)$$

onde c é uma constante que não depende de θ . Assim, a função Q é dada por

$$\begin{aligned} Q(\theta | \hat{\theta}^{(t)}) &= c - \frac{n}{2} \log \Gamma - \frac{1}{2\Gamma} \sum_{i=1}^k (y_i - \mathbf{x}_i^T \beta_1)^2 + \frac{\Delta}{2\Gamma} \sum_{i=1}^k (y_i - \mathbf{x}_i^T \beta_1) \hat{t}_i^{(t)} - \frac{\Delta^2}{2\Gamma} \sum_{i=1}^k \hat{t}_i^{2(t)} \\ &\quad - \frac{1}{2\Gamma} \sum_{i=k+1}^n (y_i - \mathbf{x}_i^T \beta_2)^2 + \frac{\Delta}{2\Gamma} \sum_{i=k+1}^n (y_i - \mathbf{x}_i^T \beta_2) \hat{t}_i^{(t)} - \frac{\Delta^2}{2\Gamma} \sum_{i=k+1}^n \hat{t}_i^{2(t)}. \end{aligned} \quad (4.4)$$

O algoritmo EM consiste em dois passos, descritos a seguir:

Passo E: Dado $\theta = \hat{\theta}^{(t)}$, obter $\hat{t}_i^{(t)}$ e $\hat{t}_i^{2(t)}$, sendo que, conforme as equações (3.17) e (3.18), temos

$$\hat{t}_i^{(t)} = \hat{\mu}_{T_i}^{(t)} + b + \widehat{M}_T^{(t)} \hat{t}_{1i}^{(t)} \quad (4.5)$$

e

$$\hat{t}_i^2(t) = (\hat{\mu}_{T_i}(t) + b)^2 + \widehat{M}_T^2(t) + \widehat{M}_T(t)(\hat{\mu}_{T_i}(t) + 2b)\hat{t}_{1i}(t), \quad (4.6)$$

onde

$$\hat{\mu}_{T_i}(t) = \begin{cases} \frac{\widehat{\Delta}(t)}{\widehat{\Delta}^2(t) + \widehat{\Gamma}(t)}(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_1 - b\Delta), & i = 1, \dots, k, \\ \frac{\widehat{\Delta}(t)}{\widehat{\Delta}^2(t) + \widehat{\Gamma}(t)}(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_2 - b\Delta), & i = k + 1, \dots, n. \end{cases}$$

Passo M: Usando os estimadores obtidos no passo E, os estimadores de máxima verossimilhança de $\boldsymbol{\theta}$ são calculados por:

$$\widehat{\boldsymbol{\beta}}_1^{(t+1)} = \left(\sum_{i=1}^k \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \sum_{i=1}^k \mathbf{x}_i (y_i - \widehat{\Delta}(t) \hat{t}_i^{(t)}), \quad (4.7)$$

$$\widehat{\boldsymbol{\beta}}_2^{(t+1)} = \left(\sum_{i=k+1}^n \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \sum_{i=k+1}^n \mathbf{x}_i (y_i - \widehat{\Delta}(t) \hat{t}_i^{(t)}), \quad (4.8)$$

$$\widehat{\Delta}^{(t+1)} = \frac{1}{\sum_{i=1}^n \hat{t}_i^2(t)} \left[\sum_{i=1}^k (y_i - \mathbf{x}_i^T \widehat{\boldsymbol{\beta}}_1^{(t)}) \hat{t}_i^{(t)} + \sum_{i=k+1}^n (y_i - \mathbf{x}_i^T \widehat{\boldsymbol{\beta}}_2^{(t)}) \hat{t}_i^{(t)} \right], \quad (4.9)$$

$$\begin{aligned} \widehat{\Gamma}^{(t+1)} = & \frac{1}{n} \left[\sum_{i=1}^k (y_i - \mathbf{x}_i^T \widehat{\boldsymbol{\beta}}_1^{(t)})^2 - 2\widehat{\Delta}(t) \sum_{i=1}^k (y_i - \mathbf{x}_i^T \widehat{\boldsymbol{\beta}}_1^{(t)}) \hat{t}_i^{(t)} \right. \\ & + \sum_{i=k+1}^n (y_i - \mathbf{x}_i^T \widehat{\boldsymbol{\beta}}_2^{(t)})^2 - 2\widehat{\Delta}(t) \sum_{i=k+1}^n (y_i - \mathbf{x}_i^T \widehat{\boldsymbol{\beta}}_2^{(t)}) \hat{t}_i^{(t)} \\ & \left. + \widehat{\Delta}^2(t) \sum_{i=1}^n \hat{t}_i^2(t) \right], \quad (4.10) \end{aligned}$$

$$\widehat{\lambda}^{(t+1)} = \frac{\widehat{\Delta}^{(t+1)}}{\widehat{\Gamma}^{(t+1)}} \quad (4.11)$$

e

$$\widehat{\sigma}^{(t+1)} = \widehat{\Delta}^{2(t+1)} + \widehat{\Gamma}^{(t+1)}. \quad (4.12)$$

Os estimadores dados acima podem ser escritos na forma matricial e seguem como:

$$\widehat{\boldsymbol{\beta}}_1^{(t)} = (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{Y}_1 - \widehat{\Delta}^{(t)} (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \widehat{\mathbf{t}}_1^{(t)}, \quad (4.13)$$

$$\widehat{\boldsymbol{\beta}}_2^{(t)} = (\mathbf{X}_2^T \mathbf{X}_2)^{-1} \mathbf{X}_2^T \mathbf{Y}_2 - \widehat{\Delta}^{(t)} (\mathbf{X}_2^T \mathbf{X}_2)^{-1} \mathbf{X}_2^T \widehat{\mathbf{t}}_2^{(t)}, \quad (4.14)$$

$$\widehat{\Delta}^{(t)} = \left(\mathbf{1}_n^T \widehat{\mathbf{t}}_2^{(t)} \right)^{-1} \left[\left(\mathbf{Y}_1 - \mathbf{X}_1 \widehat{\boldsymbol{\beta}}_1^{(t)} \right)^T \widehat{\mathbf{t}}_1^{(t)} + \left(\mathbf{Y}_2 - \mathbf{X}_2 \widehat{\boldsymbol{\beta}}_2^{(t)} \right)^T \widehat{\mathbf{t}}_2^{(t)} \right], \quad (4.15)$$

$$\begin{aligned} \widehat{\Gamma}^{(t)} = & \frac{1}{n} \left[\left(\mathbf{Y}_1 - \mathbf{X}_1 \widehat{\boldsymbol{\beta}}_1^{(t)} \right)^T \left(\mathbf{Y}_1 - \mathbf{X}_1 \widehat{\boldsymbol{\beta}}_1^{(t)} \right) - 2 \widehat{\Delta}^{(t)} \left(\mathbf{Y}_1 - \mathbf{X}_1 \widehat{\boldsymbol{\beta}}_1^{(t)} \right)^T \widehat{\mathbf{t}}_1^{(t)} \right. \\ & + \left. \left(\mathbf{Y}_2 - \mathbf{X}_2 \widehat{\boldsymbol{\beta}}_2^{(t)} \right)^T \left(\mathbf{Y}_2 - \mathbf{X}_2 \widehat{\boldsymbol{\beta}}_2^{(t)} \right) - 2 \widehat{\Delta}^{(t)} \left(\mathbf{Y}_2 - \mathbf{X}_2 \widehat{\boldsymbol{\beta}}_2^{(t)} \right)^T \widehat{\mathbf{t}}_2^{(t)} \right. \\ & \left. + \widehat{\Delta}^{2(t)} \mathbf{1}_n^T \widehat{\mathbf{t}}_2^{(t)} \right], \quad (4.16) \end{aligned}$$

onde $\widehat{\mathbf{t}}_1 = (\widehat{t}_1, \dots, \widehat{t}_n)^T$, $\widehat{\mathbf{t}}_2 = (\widehat{t}_{k+1}, \dots, \widehat{t}_n)^T$, $\widehat{\mathbf{t}}_2 = (\widehat{t}_1^2, \dots, \widehat{t}_n^2)^T$ e $\mathbf{1}_n^T$ é um vetor de uns de dimensão $1 \times n$. O algoritmo alterna entre os passos E e M até que a sequência $\widehat{\boldsymbol{\theta}}^{(t)}$ convirja. Observe que quando $\Delta = 0$ (ou $\lambda = 0$), temos que as equações do passo M se reduzem as equações obtidas por Chen (1998) e Osorio e Galea (2005, Seção 2), sob o modelo normal com ponto de mudança.

4.3 Detecção do Ponto de Mudança via Critério de Informação

O trabalho de Chen (1998) propõe transformar o processo de testar hipóteses em um procedimento de seleção de modelos usando o critério de informação SIC, definido em (1.4), dado por $SIC = -2l(\widehat{\boldsymbol{\theta}}|\mathbf{y}_{obs}) + \log(n) s$. Neste trabalho, nós propomos também utilizar o critério de informação AIC, definido em (1.3), dado por $AIC = -2l(\widehat{\boldsymbol{\theta}}|\mathbf{y}_{obs}) + 2s$ para a detecção do ponto de mudança associado com os modelos. Note que maximizar a função log-verossimilhança é equivalente a minimizar os critérios de informação SIC e AIC.

Assim, o critério de informação irá escolher o modelo com ponto de mudança na posição k , se para algum k

$$SIC(n) > SIC(k), \quad (4.17)$$

onde $SIC(n)$ e $SIC(k)$ denotam os critérios sob H_0 e sob H_1 , respectivamente, para $k = p, \dots, n - p$.

Quando a hipótese nula é rejeitada, o estimador de máxima verossimilhança do ponto de mudança nos coeficiente de regressão, denotado por \hat{k} , deve satisfazer

$$SIC(\hat{k}) = \min\{SIC(k): p \leq k \leq n - p\}. \quad (4.18)$$

De forma análoga, o procedimento proposto por Chen (1998), descrito em (4.17) e (4.18), pode ser reescrito considerando o critério AIC.

Na próxima seção, com o intuito de exemplificar a teoria apresentada neste capítulo, serão realizadas aplicações em dados simulados e reais.

4.4 Aplicação em Dados Simulados

4.4.1 Estimação por Máxima Verossimilhança

O primeiro objetivo é verificar se podemos recuperar os valores dos parâmetros reais quando usamos o algoritmo EM proposto, ajustando o modelo de regressão linear normal assimétrico com ponto de mudança aos dados que foram gerados artificialmente. Neste caso, sem perda de generalidade, geramos 100 amostras de tamanho $n = 30, 50, 100, 200, 300$ e 500 provenientes do modelo de regressão linear normal assimétrico com ponto de mudança, considerando k (posição conhecida) na metade da quantidade de n (Tabela 4.1) e na outra situação considerando k em uma posição arbitrária de n (Tabela 4.2), a variável explicativa utilizada foi $x_i \sim U(0,1)$ e os erros $\varepsilon_i \sim SN(0.76, 1, -3)$, da seguinte forma:

$$y_i = \begin{cases} 4 + 7x_i + \varepsilon_i, & i = 1, \dots, k, \\ -1 + 3x_i + \varepsilon_i, & i = k + 1, \dots, n. \end{cases}$$

Tabela 4.1: Média das estimativas dos parâmetros baseada nas 100 amostras provenientes do modelo normal assimétrico com ponto de mudança na metade de n .

Valores Verdadeiros	(4;7)	(-1;3)	1	-3
Estimativas	$\hat{\beta}_1 = (\hat{\beta}_0, \hat{\beta}_1)$	$\hat{\beta}_2 = (\hat{\beta}_0^*, \hat{\beta}_1^*)$	$\hat{\sigma}^2$	$\hat{\lambda}$
$n=30$	(3,9835; 7,0027)	(-1,0443; 3,0682)	0,6790	-2,1990
$n=50$	(4,0089; 6,9994)	(-1,0379; 3,0014)	0,8494	-3,1648
$n=100$	(3,9879; 7,0035)	(-0,9851; 2,9793)	1,0017	-3,7430
$n=200$	(4,0057; 7,0046)	(-0,9872; 2,9891)	0,9629	-3,2079
$n=300$	(4,0107; 6,9980)	(-0,9923; 2,9760)	0,9849	-3,2222
$n=500$	(4,0016; 7,0010)	(0,9925; 2,9850)	0,9903	-3,0720

Tabela 4.2: Média das estimativas dos parâmetros baseada nas 100 amostras provenientes do modelo normal assimétrico com ponto de mudança k em diferentes posições de n .

Valores Verdadeiros	(4;7)	(-1;3)	1	-3
Estimativas	$\hat{\beta}_1 = (\hat{\beta}_0, \hat{\beta}_1)$	$\hat{\beta}_2 = (\hat{\beta}_0^*, \hat{\beta}_1^*)$	$\hat{\sigma}^2$	$\hat{\lambda}$
$n=30$ ($k=10$)	(3,9816; 7,0298)	(-1,0286; 2,9986)	0,7824	-2,8348
$n=50$ ($k=15$)	(3,9598; 7,0262)	(-1,0230; 3,0153)	0,8925	-3,4105
$n=100$ ($k=25$)	(3,9898; 6,9727)	(-0,9995; 3,0149)	0,9620	-3,4829
$n=200$ ($k=115$)	(3,9930; 7,0173)	(-0,9823; 2,9627)	1,0011	-3,4208
$n=300$ ($k=100$)	(4,0186; 6,9770)	(-0,9921; 2,9888)	0,9924	-3,2128
$n=500$ ($k=165$)	(3,9991; 6,9884)	(-0,9978; 2,9946)	1,0070	-3,1324

Note que o algoritmo proposto é eficiente para a estimação dos parâmetros do modelo de regressão linear normal assimétrico, no contexto de ponto de mudança com a posição k na metade da quantidade de n e para k com diferentes posições em n , independente do tamanho da amostra.

4.4.2 Critério de Informação

Usualmente, o ponto de mudança é desconhecido. Portanto, o objetivo deste estudo de simulação é avaliar os critérios de informação SIC e AIC como ferramentas para detectar o ponto de mudança, no contexto de assimetria. Neste caso, sem perda de generalidade, utilizamos amostras geradas na seção anterior de tamanho 30 ($k = 10$ e 15) e 300 ($k = 100$ e 150), e avaliamos se os critérios SIC e AIC são ferramentas eficientes na detecção de tais pontos de mudança.

Um resumo dos resultados se encontra nas Figuras 4.1 e 4.2. Note que é possível detectar pontos de mudança nas posições 10 e 15 para $n = 30$ e, 100 e 150 para $n = 300$, considerando os critérios SIC e AIC. Como esperado, os critérios de informação SIC e AIC detectaram o mesmo ponto de mudança, uma vez que para um mesmo modelo ajustado, neste caso, o modelo normal assimétrico, a comparação entre critérios se reduz à comparação das funções log-verossimilhança, pois a constante de penalização de ambos os critérios depende do mesmo número de parâmetros.

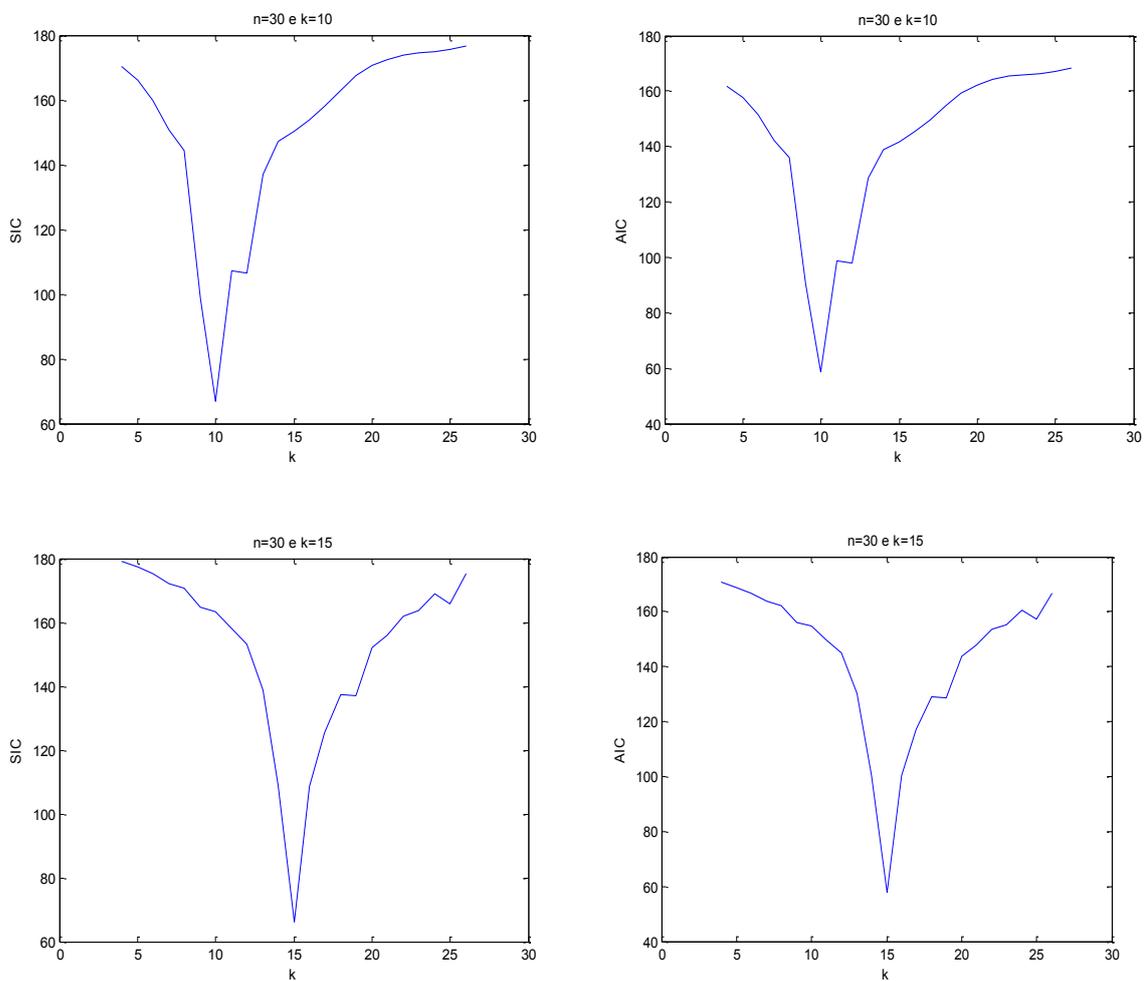


Figura 4.1: Gráficos de ponto de mudança para $n = 30$ com $k = 10$ e 15 para os critérios SIC e AIC.

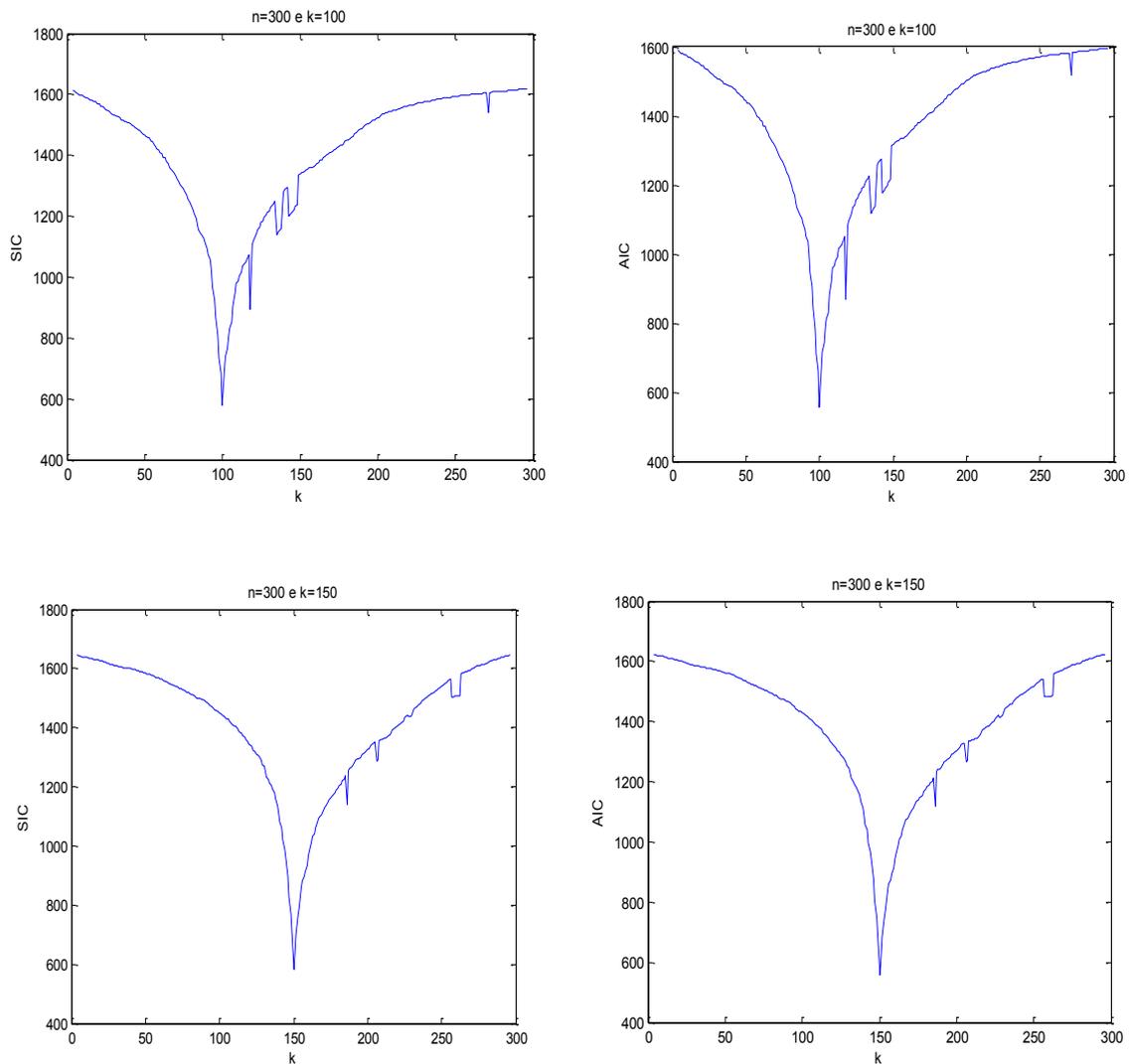


Figura 4.2: Gráficos de ponto de mudança para $n = 300$ com $k = 100$ e 150 para os critérios SIC e AIC.

Dessa forma, os estudos de simulação confirmam a necessidade do desenvolvimento de procedimentos para realizar inferência no contexto do modelo de regressão com ponto de mudança e estrutura assimétrica dos dados.

4.5 Aplicação em Dados Reais (Holbert's data)

Nesta seção, será utilizado o conjunto de dados analisados por Holbert (1982). Sob o ponto de vista bayesiano, Holbert (1982) usa o modelo de regressão linear normal com ponto de mudança nos coeficientes. Posteriormente, Chen (1998) analisa estes dados usando o

critério SIC para detectar mudanças na média para o modelo de regressão linear normal. Ainda no contexto de simetria dos dados, Osorio e Galea (2005) utilizam o critério SIC para a detecção do modelo de regressão linear Student-t. Neste trabalho, os dados Holbert serão utilizados para ilustrar os critérios SIC e AIC como procedimentos para detecção de ponto de mudança no modelo de regressão linear normal assimétrico.

Os dados foram coletados no período de janeiro de 1967 e novembro de 1969, onde foram consideradas as variáveis volume de venda mensal em milhões de dólares pela Bolsa de Valores de Boston (BSE) e o volume de venda mensal em milhões de dólares das Bolsas de Valores Americanas com a de New York (NYAMSE).

O modelo de regressão linear simples, sob H_0 , é definido por

$$Y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i, i = 1, \dots, 35,$$

onde Y_i é o BSE para o i -ésimo mês, x_{1i} representa NYAMSE do i -ésimo mês e assumimos que os erros de observação seguem as distribuições normal e normal assimétrica. Utilizaremos as duas distribuições com a finalidade de comparações. A Figura 4.3 apresenta o gráfico de dispersão desses dados. Observe que apenas um modelo de regressão linear não parece ajustar adequadamente os dados, pois nota-se que a massa de dados se divide em dois grupos, sendo o possível ponto de mudança na observação 23. Portanto, os critérios SIC e AIC serão utilizados como ferramentas para avaliarmos a existência do ponto de mudança na posição 23, informação obtida através da análise exploratória dos dados.

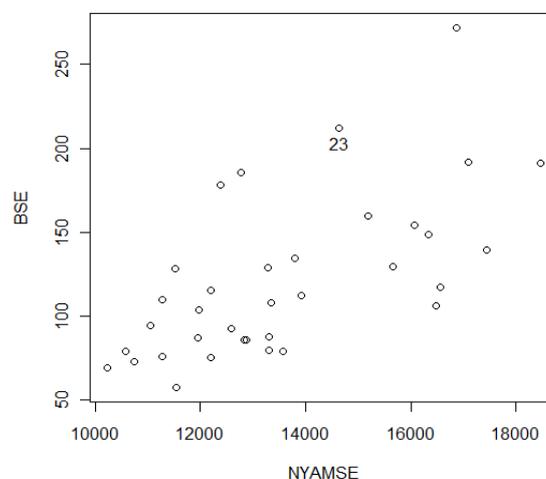


Figura 4.3: Gráficos de dispersão Holbert's data.

Os critérios SIC e AIC foram utilizados para a detecção do ponto de mudança nos modelos normal e normal assimétrico. Um sumário dos resultados se encontra nas Figuras 4.4 e 4.5.

Sob o modelo normal assimétrico, os gráficos da Figura 4.4 evidenciam a suspeita de existir um ponto de mudança no conjunto de dados, indicando a posição 23. Já sob o modelo normal, os gráficos da Figura 4.5, também indicam que existe um ponto de mudança na posição 23. Tais resultados estão em concordância com a análise realizada por Osorio e Galea (2005) que também identificou a posição 23 sob o modelo de regressão linear Student-t.

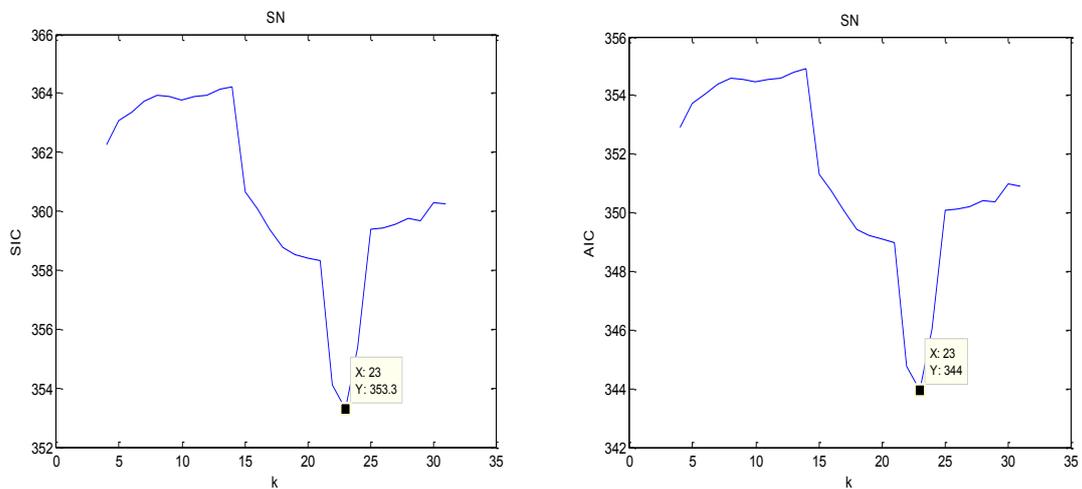


Figura 4.4: Gráficos de critérios de informação SIC e AIC para normal assimétrica.

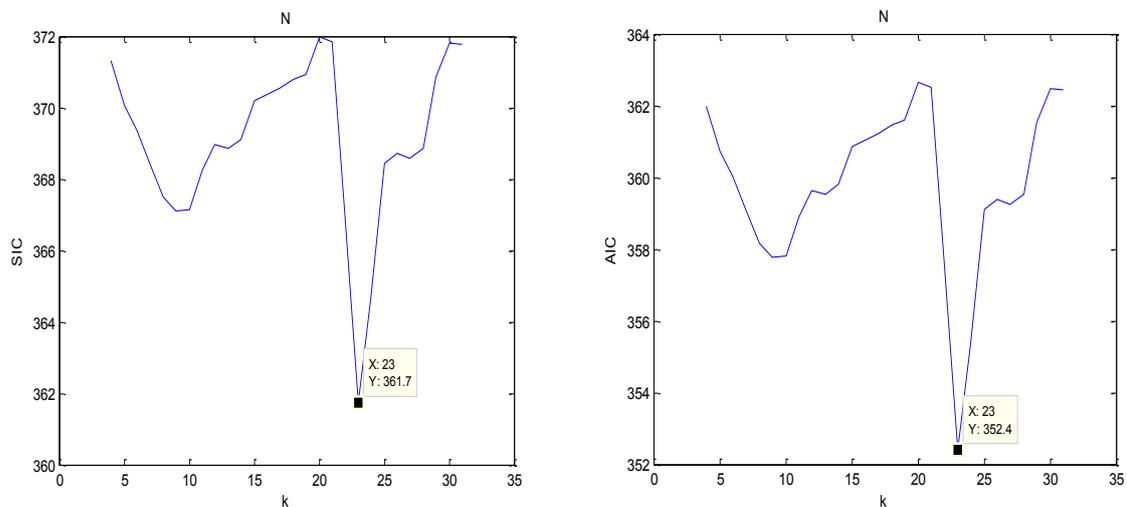


Figura 4.5: Gráficos de critérios de informação SIC e AIC para normal.

Considerando o ponto de mudança na posição 23, as estimativas dos parâmetros dos modelos normal e normal assimétrico estão apresentadas na Tabela 4.3. Note que as estimativas dos coeficientes de regressão dos modelos normal e normal assimétrico são bastante distintas, a partir do ponto de mudança, sendo que as estimativas dos interceptos alternam o sinal (caso normal: 11,0747 e caso normal assimétrico: -12,6296). Através da Tabela 4.4, podemos inspecionar os critérios de informação SIC e AIC. Note que o modelo de regressão normal assimétrico se ajusta melhor ao conjunto de dados.

Tabela 4.3: Estimativas dos parâmetros dos modelos ajustados com ponto de mudança na posição 23.

	Normal	Normal Assimétrico
β_1 (1 a 23)	(-110,5475; 0,0178)	(-116,1910; 0,0180)
β_2 (24 a 35)	(11,0747; 0,0067)	(-12,6296; 0,0910)

Tabela 4.4: Alguns critérios de informação.

Modelos	$l(\hat{\theta})$	SIC	AIC
Normal	-171,2020	351,6684	354,4040
Normal Assimétrico	-165,9820	341,2284	343,9640

Além disso, usamos a estatística da Razão de Verossimilhanças (RV) para testar a hipótese nula $H_0: \lambda = 0$, isto é, para verificar se o modelo simétrico (modelo normal) é suficiente. Obtivemos um P-Valor = 0,0012, portanto há evidências suficientes nos dados para rejeitar a hipótese nula e assim, a distribuição normal assimétrica é mais apropriada para este conjunto de dados.

5. Conclusão

Em resumo, neste trabalho discutimos vários aspectos envolvendo modelos de regressão lineares sob a distribuição normal assimétrica. Um dos aspectos abordados foi o problema de ponto de mudança associado com os modelos de regressão. Para a estimação via máxima verossimilhança, a densidade marginal das quantidades envolvidas foi obtida de modo que os estimadores podem ser usados diretamente através de alguma técnica de maximização, usando programas estatísticos, tais como R e o Matlab, por exemplo. Alternativamente, o algoritmo EM foi proposto, explorando as propriedades estatísticas discutidas para a distribuição normal assimétrica. Adicionalmente, mostramos que os critérios de informação usuais, tais como SIC e AIC, podem ser usadas para detectar afastamentos da normalidade e detectar também ponto de mudança associado com os modelos de regressão lineares.

Outro aspecto abordado foi o desenvolvimento de métodos de diagnósticos para os modelos de regressão lineares assimétricos. A distância de Mahalanobis foi considerada para identificar possíveis observações atípicas que podem prejudicar o ajuste dos modelos bem como as inferências estatísticas. Temos notado que, para algumas aplicações, os modelos lineares assimétricos tendem a acomodar melhor as observações aberrantes e/ou influentes.

Os resultados obtidos foram aplicados em conjunto de dados reais e/ou simulados. Foram utilizados os programas estatísticos R e Matlab para as programações dos procedimentos de estimação e diagnóstico dos modelos ajustados.

Para o leitor interessado em aplicar a metodologia desenvolvida nesta monografia com o objetivo de analisar um conjunto de dados, recomendamos como passo inicial uma análise exploratória dos dados. De acordo com as observações acima, ajustamos um modelo para o conjunto de dados, tais que as estimativas de máxima verossimilhança dos parâmetros de interesse são obtidas via algoritmo EM e seus respectivos erros padrão são calculados via a técnica de reamostragem “Bootstrap”. Para avaliar o ajuste do modelo proposto, sugerimos a inspeção de alguns critérios de informação, tais como SIC e AIC, por exemplo.

Concluindo, esta monografia é um esforço inicial para estudar alguns tópicos nesta área de pesquisa e divulgar a utilidade da mesma.

6. Perspectivas Futuras

Vários trabalhos de pesquisa podem ser derivados e/ou direcionados a partir dos resultados desta monografia, entre eles podemos mencionar os seguintes:

- Estender o modelo de regressão linear com ponto de mudança para a classe de distribuições assimétricas mistura de escala skew-normal.
- Considerar outros critérios para detecção de ponto de mudança no modelo de regressão linear, no contexto de assimetria.
- Estudar o modelo de regressão linear com múltiplos pontos de mudança, no contexto de assimetria.

7. Referências Bibliográficas

- [1] Albert, J. (2007). Bayesian Computation with R. Springer.
- [2] Arellano, R. B. V., Castro, L. M. e Loschi, R. H. (2013). Change Point in the Skew-Normal Model Parameters. *Communications in Statistics - Theory & Methods*, **42**, 603-618.
- [3] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171-178.
- [4] Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Foundation of the Scandinavian Journal of Statistics*, Vol 32: 159–188.
- [5] Azzalini, A. e Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society*, **61**, 579-602.
- [6] Azzalini, A. e Capitanio, A. (2003). Distributions generated and perturbation of symmetry with emphasis on the multivariate skew-t distribution. *Journal of the Royal Statistical Society*, **61**, 367-389.
- [7] Basso, R. M. Misturas finitas de misturas de escala skew-normal. Dissertação (Mestrado em Estatística) – Universidade Estadual de Campinas, Campinas, 2009.
- [8] Branco, M. D. e Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, **79**, 99-113.
- [9] Chen, J. (1998). Testing for a change point in linear regression models. *Communications in Statistics - Theory & Methods*, **27**, 2481-2493.
- [10] Chen, J. e Gupta, A. K. (2001). On change point detection and estimation. *Communications in Statistics – Simulation and Computation*, **30**, 665-607.
- [11] Dempster, A. P., Laird, N. M. e Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1-22.
- [12] Fang, K. T., Kotz, S. e Ng. K. W. (1990). *Simetric multivariate and relates distributions*. Champman & Hall, London.
- [13] Ferreira, C. S. Inferência e diagnóstico em modelos assimétricos. Tese (Doutorado em Estatística) – Universidade de São Paulo, São Paulo, 2008.
- [14] Henze, N. (1986). A probabilistic representation of the “Skew-normal” distribution. *Scand J Statist* **13**, 271-275.

- [15] Hofrichter. J. (2007). Change Point Detection in Generalized Linear Model. Dissertação – Graz University of Technology.
- [16] Holbert. D. (1982). A Bayesian Analysis of a Switching Linear Model. *Journal of Econometrics*, **19**, 77-87.
- [17] Lachos, V. H. Modelos lineares mistos assimétricos. Tese (Doutorado em Estatística) – Universidade de São Paulo, São Paulo, 2004.
- [18] Lachos, V. H. D., Cabral C. R. D. e Garay A. W. M. (2013). Modelos Não Lineares Assimétricos. Escola de Modelos de Regressão, 13º, Maresias, São Paulo.
- [19] Lee, S.Y., Lu, B. e Song, X. Y. (2006). Assessing local influence for nonlinear structural equation models with ignorable missing data. *Computational Statistics and Data Analysis*, **50**, 1356-1377.
- [20] Lucas, A. (1997) Robustness of the student t based m-estimator. *Communications in Statistics - Theory & Methods*, **26**, 1165-1182.
- [21] O'hagan, A. e Leonard, T. (1976). Bayes estimation subject to uncertainty about constraints. *Biometrika*, **63**, 201-203.
- [22] Osorio, F. e Galea, M. (2005). Detection of a change-point in student-t linear regression models. *Statistical Papers*, **45**, 31-48.
- [23] Pimm, S., Jones, H. e Diamond, J. (1988). On the risk of extinction. *American Naturalist*, **132**, 757-785.
- [24] Pinheiro, J. C., Liu, C. H. e Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using a multivariate t-distribution. *Journal of Computational and Graphical Statistics*. **10**, 249-276.
- [25] Rodríguez, C. L. B. Inferência Bayesiana no Modelo Normal Assimétrico. Dissertação (Mestrado em Estatística) - Universidade de São Paulo, São Paulo, 2005.
- [26] Zeller, C. B. Distribuições Misturas de Escala Skew-Normal: Estimção e Diagnóstico em Modelos Lineares. Tese (Doutorado em Estatística) - Universidade Estadual de Campinas, Campinas, 2009.
- [27] Zeller, C. B., Lachos, V. H. e Vilca, L. F. (2011). Local influence analysis for regression models with scale mixtures of skew-normal distributions. *Journal Applied Statistics*, **38** (2), 348-363.
- [28] Zhang, D. e Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*. **57**, 795-802.