Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Departamento de Estatística

Amanda Romanelli

Estimação em modelos de regressão linear com ponto de mudança contínuo sob a distribuição t-Student

Juiz de Fora
2017

Amanda Romanelli

Estimação em modelos de regressão linear com ponto de mudança contínuo sob a distribuição t-Student

Monografia apresentada ao Departamento de Estatística da Universidade Federal de Juiz de Fora, na área de concentração em Estatística, como requisito parcial para obtenção do título de Bacharela em Estatística.

Orientadora: Camila Borelli Zeller

Juiz de Fora

2017

Ficha catalográfica elaborada através do programa de geração
automática da Biblioteca Universitária da UFJF,
com os dados fornecidos pelo(a) autor(a)

Romanelli, Amanda.
Estimação em modelos de regressão linear com ponto de mudança contínuo sob a distribuição t-Student / Amanda Romanelli. - 2017.
38 p. : il.

Orientadora: Camila Zeller
Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas, 2017.

1. Algoritmo EM. 2. Distribuição t-Student. 3. Modelo de Regressão Linear. 4. Observações Aberrantes. 5. Ponto de Mudança. I. Zeller, Camila, orient. II. Título.

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 linear. A normalidade dos erros aleatórios é uma suposição rotineira em modelos linear, que pode ser não realista. Assim, relaxamos a suposição de normalidade considerando que os erros aleatórios seguem uma distribuição mistura de escala normal, especificamente a distribuição t-Student. 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, denominado grau de liberdade, que controla a curtose. Além disso, 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, e a seleção de modelos será realizada através dos critérios de informação (SIC e AIC). Dessa forma, o principal objetivo deste trabalho é estudar alguns aspectos de estimação em modelos de regressão linear com ponto de mudança sob a distribuição t-Student. Finalmente, exemplos numéricos considerando dados simulados e reais são apresentados para ilustrar o modelo e os resultados inferenciais desenvolvidos. Foi utilizado o programa estatístico R. 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 t-Student. Modelo de regressão linear. Observações aberrantes. Ponto de mudança.

ABSTRACT

In this work, we present recent results in a statistical research field with a wide range of applications: linear regression models. The normality (symmetry) of random errors is an ordinary assumption in linear models, which may be unrealistic. Thus, we relax the assumption of normality and consider the case of scale mixture of normal regression,which has the normal regression as a particular case. Furthermore, this model is able to capture several non normal behaviors by applying an extra parameter, the degree of freedom, that can control the kurtosis. In addition, we consider 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 change point. In this context, the estimation of the parameters is performed via EM algorithm, and the selection of models is made by using information criteria (SIC and AIC). Finally, numerical examples considering simulated and real data are presented to illustrate the model and the inferential results developed. We believe that this is work will be useful students, researchers and practitioners to be of a great applicability.

Keywords: EM Algorithm. Student-t distribution. Linear model. Aberrant observations. Change point.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

SUMÁRIO

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 caracterizar a relação média 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 nem sempre isso é coerente, já que o modelo pode alterar seu comportamento após um ponto específico (tempo ou alguma região do domínio das variáveis preditoras, por exemplo) que, em geral, é desconhecido, e denominado ponto de mudança. O problema de ponto de mudança surgiu, inicialmente, no contexto de controle de qualidade, como demonstrado com os gráficos de Skewart (1939), 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. Dessa forma, a identificação desse ponto desempenha um importante papel. Por exemplo, em um processo de produção contínuo, é esperado que a qualidade dos produtos se mantenha estável. Entretanto, por muitas razões, o processo pode falhar na produção de produtos com a mesma qualidade. Portanto, deseja-se encontrar se há um ponto em que a partir dele a qualidade do produto começa a se deteriorar.Veja Chen & Gupta (2011) para mais detalhes.

O problema do ponto de mudança tem sido um tema de constante interesse na literatura Estatística. Principalmente porque o problema de ponto de mudança pode ser encontrado em várias áreas, tais como economia, finanças, medicina, psicologia, geologia, química, literatura, dentre outras. Do ponto de vista estatístico, o modelo pode se alterar após um ponto específico que, em geral,é desconhecido, e denominado ponto de mudança. Ou seja, no contexto de modelos de regressão linear, considerando uma sequência de observações $(x_i, Y_i), i = 1, \ldots, n$, o modelo de regressão com ponto de mudança pode ser escrito como

$$Y_i = \begin{cases} \alpha_1 + \beta_1 x_i + \varepsilon_i, \forall x_i \leq \gamma \\ \alpha_2 + \beta_2 x_i + \varepsilon_i, \forall x_i > \gamma \end{cases}, \tag{1.1}$$

onde $\alpha_1, \alpha_2, \beta_1, \beta_2$ e γ são parâmetros desconhecidos, tal que γ é o ponto de mudança, e os erros $\epsilon_i \sim N(0, \sigma^2)$. Tais modelos são divididos em dois tipos. Um tipo em que o modelo é assumido como contínuo no ponto de mudança (também conhecido como mudança gradual ou sem descontinuidade) e outro onde não é (também conhecido como mudança abrupta ou com descontinuidade). A inferência teórica é completamente diferente para

cada tipo de modelo. Estes modelos podem ser estendidos para o caso de múltiplos pontos de mudança. Muggeo (2003) adverte que o uso de muitos pontos de mudança é questionável para a maioria das aplicações práticas e que nestes casos, a modelagem de regressão não-linear poderia ser mais apropriada.

Em contraste com o modelo (1.1), uma restrição de continuidade no ponto de mudança é assumido. Assim, ambos os modelos devem prever o mesmo valor médio no ponto de mudança, o que resulta em uma transição contínua dos dois modelos. Além disso, presume-se que a variável explicativa seja classificada em ordem crescente, $x_i \leq x_{i+1}, i = 1, \ldots, n-1$. Agora, a localização do ponto de mudança não é mais restrito a um x_i observado. Em vez disso, pode ser qualquer valor dentro do intervalo $[a; b]$, onde se encontram os seguintes pontos $a = min \;\; x_i = x_{(1)}$ e $b = max \;\; x_i = x_{(n)}$. O modelo de regressão linear com um ponto de mudança contínuo pode então ser indicado como

$$Y_i = \begin{cases} \alpha_1 + \beta_1 x_i + \varepsilon_i, \forall a \leq x_i \leq \gamma \\ \alpha_2 + \beta_2 x_i + \varepsilon_i, \forall \gamma < x_i \leq b \end{cases}, \tag{1.2}$$

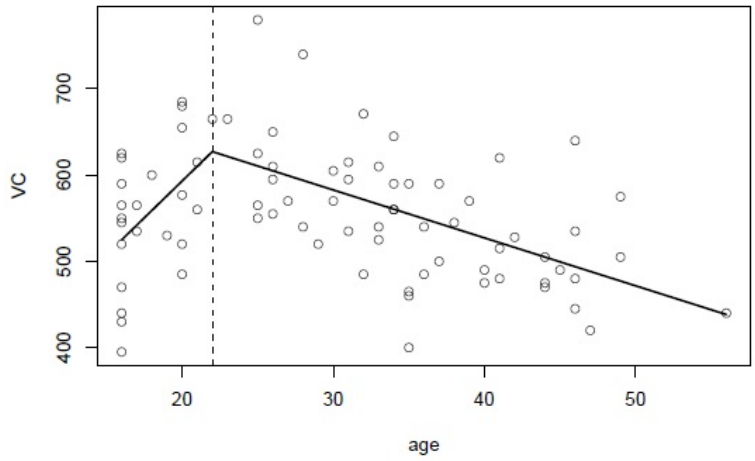
com o contraste de continuidade

$$\alpha_1 + \beta_1 \gamma = \alpha_2 + \beta_2 \gamma.$$

Hofrichter (2007) ilustra o modelo (1.2)com dados provenientes de um estudo de saúde, realizado na Estíria (Aústria). O interesse era avaliar a dependência entre a capacidade pulmonar (VC) e a idade da pessoa. Neste estudo, 79 pessoas do sexo masculino com idades entre 16 anos e 56 anos foram examinadas. Os dados mostram uma tendência clara e crescente para pessoas jovens e uma tendência decrescente para idosos (Figura 1), logo teremos duas retas distintas, uma crescente e a outra decrescente, o que indica claramente a exintência de um ponto de mudança entre estas duas retas lineares. Assim, os dois modelos devem prever o mesmo valor médio no ponto de mudança, e uma restrição de continuidade é obrigatória. O modelo ajustado com um ponto de mudança contínuo estimado na idade de $\widehat{\gamma} = 22$ é plotado na Figura 1.

A importância do ponto de mudança pode ser notada pelo grande número de artigos que vem sendo publicados em vários periódicos. Entretanto, muitos deles dissertam acerca de diferentes tipos de estruturas de ponto de mudança com descontinuidade em relação à média, à variância e à média e variância simultaneamente, de uma sequência de variáveis aleatórias independentes e normalmente distribuídas. Por exemplo, os problemas de ponto de mudança com descontinuidade na média e/ou na variância, no contexto de normalidade dos dados, já foram examinados por Chen & Gupta (1996, 1997, 1999, 2003, 2011). Portanto, os comentários acima justificam a ênfase que será dada neste trabalho ao modelo de regressão com restrição de continuidade no ponto de mudança (único).

Figura 1 – Modelo com ponto de mudança contínuo ajustado com a estimativa de ponto de mudança na idade $\widehat{\gamma} = 22$.



Agora vamos voltar nossa atenção para além dos modelos gaussianos, e estudar outros modelos importantes . A suposição de normalidade sempre foi muito atrativa para os modelos linear com respostas contínuas e, mesmo quando não era alcançada, procurava-se alguma transformação na resposta no sentido de obter pelo menos simetria. Entretanto, tal suposição pode ser irreal, especialmente quando existem evidências empíricas de que os dados seguem uma distribuição com caudas mais pesadas do que a normal. Além disso, com o passar do tempo, verificou-se que as estimativas obtidas para os coeficientes dos modelos normais mostraram-se sensíveis às observações extremas, incentivando o desenvolvimento de metodologias flexíveis. 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, Lange et al. (1989) propõem o modelo t-Student, Yamaguchi (2001) sugere usar a distribuição normal contaminada e Osorio (2006) propõe as distribuições de misturas de escala normal. Esta classe contém as distribuições normal, t-student, slash, normal contaminada, etc. e tem sido estudada com grande interesse nos últimos anos. Neste trabalho, focamos principalmente na distribuição t-Student.

Em suma, a maioria dos métodos que foram revisados parecem inadequados por pelo menos uma das seguintes razões: (i) o modelo é limitado a uma distribuição particular da resposta (frequentemente, gaussiana); (ii) modelo é assumido com mudança abrupta e (iii) o ponto de mudança nem sempre é tratado como "parâmetro verdadeiro" (desconhecido). Nestas linhas de pesquisa, propomos estudar o problema de ponto de mudança contínuo, na estrutura da média (isto é, nos coeficientes de regressão), em dados com distribuição t-Student. A respeito da inferência sobre modelos com ponto de mudança, devem ser considerados os seguintes aspectos: determinar a existência do ponto de mudança, localizar a posição deste ponto, estimar todos os parâmetros de

interesse do modelo e realizar análises preditivas. Se a localização do ponto de mudança é conhecida, então a estimação dos parâmetros do modelo é direta,caso contrário, um parâmetro extra (ponto de mudança, γ) deve ser estimado. O enfoque deste trabalho é voltado para a estimação dos parâmetros do modelo via algoritmo EM. O algoritmo EM tem diversas vantagens sobre a maximização direta da função de verossimihança uma vez que é facilmente implementável, numericamente estável e bastante acurado.

Na próxima seção descrevemos os objetivos e a organização do trabalho.

1.2 Objetivos

Neste trabalho, fornecemos alguns resultados adicionais para o problema de ponto de mudança sem descontinuidade, no contexto de simetria, em particular para o modelo de regressão t-Student. Inspirados pelo trabalho de Young (2004), o enfoque do trabalho é voltado para a estimação dos parâmetros do modelo via algoritmo EM. Podemos então relacionar os seguintes objetivos específicos: (i) estimar o ponto de mudança; (ii) implementar e avaliar o algoritmo EM proposto computacionalmente; e (iii) aplicar esses resultados para analisar dados reais.

1.3 Descrição dos Capítulos

O trabalho está organizado em 4 capítulos.

No Capítulo 2, apresenta-se uma revisão dos principais conceitos que serão tratados nesta monografia. Apresentamos uma breve descrição do algoritmo EM, e em seguida mostramos alguns critérios que serão utilizados para selecionar os modelos, com base na classe de misturas de escala normal. Posteriormente, apresentamos uma revisão sobre algumas das propriedades e resultados referentes à classe de distribuições de misturas de escala normal. Essa classe de distribuições será utilizada nos demais capítulos para relaxar a suposição de normalidade, geralmente, assumida nos modelos de regressão. Finalmente, descrevemos o modelo de regressão linear, onde as observações seguem distribuições de misturas de escala normal, especificamente as distribuições normal e t-Student.

No capítulo 3, estudam-se os modelos de regressão linear considerando um ponto de mudança, onde as observações seguem uma distribuição mistura de escala normal. Em particular, neste capítulo, fixamos nossa atenção à distribuição t-Student. A especificação do modelo, a estimação por máxima verossimilhança dos parâmetros, o algoritmo EM correspondente e a forma de determinação do ponto de mudança são apresentados. O terceiro capítulo pode ser considerado o objetivo principal desse trabalho. Exemplos

numéricos considerando dados simulados e reais são utilizados para ilustrar o modelo e os resultados inferenciais desenvolvidos aqui.

As conclusões finais estão no Capítulo 4.

2 PRINCIPAIS CONCEITOS

Neste capítulo apresentamos os principais conceitos que serão tratados nesta monografia. Na próxima seção, apresentamos uma breve descrição do algoritmo EM e, em seguida, mostraremos alguns critérios que serão utilizados para selecionar os modelos baseados na classe de misturas de escala normal. Posteriormente, apresentamos uma revisão sobre algumas das propriedades e resultados referentes à classe de distribuições de misturas de escala normal. Essa classe de distribuições será utilizada no próximo capítulo para relaxar a suposição de normalidade, geralmente assumida nos modelos de regressão. É vasta a quantidade de trabalhos relevantes sobre a classe de distribuições de misturas de escala normal na literatura, alguns deles são citados neste trabalho. Finalmente, descrevemos o modelo de regressão linear, onde as observações seguem distribuições de misturas de escala normal, especificamente as distribuições normal e t-Student.

2.1 Algoritmo EM

Neste trabalho, a obtenção do estimador de máxima verossimilhança (EMV) de $\boldsymbol{\theta}$ (parâmetro de interesse) será baseada no algoritmo EM. O algoritmo EM (Dempster et al., 1977) é um enfoque amplamente aplicado no cálculo iterativo de estimativas de máxima verossimilhança, sendo bastante útil para problemas com dados incompletos.

Muitos problemas em Estatística podem ser considerados utilizando uma formulação de dados aumentados permitindo assim simplificar a obtenção de estimativas de máxima verossimilhança. Os dados aumentados, também chamados dados completos, correspondem aos dados observados, em conjunto com os dados perdidos ou não observáveis. Neste contexto, as funções de verossimilhança, baseadas nos dados completos e observados, são denominadas verossimilhança de dados completos e dados incompletos, respectivamente, sendo que os dados completos são representados somente pelos dados observados. É importante salientar que a parte aumentada dos dados, referente aos dados não obeservados, não requer que eles sejam "perdidos" no sentido estrito da palavra, pois somente representam um mecanismo técnico.

Considere que $\mathbf{y}_{obs}$ denota os dados observados e $\mathbf{y}_{mis}$ denota os dados não observáveis, sendo os dados completos $\mathbf{y}_c = (\mathbf{y}_{obs}, \mathbf{y}_{mis})$. Denotamos por $f(\mathbf{y}_c|\boldsymbol{\theta})$ a função de verossimilhança dos dados completos e $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = \log(f(\mathbf{y}_c|\boldsymbol{\theta}))$ a função de log-verossimilhança dos dados completos. O algoritmo EM aborda problemas com dados incompletos indiretamente mediante a substituição da parte não observável em $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)$ por suas esperanças condicionadas a $\mathbf{y}_{obs}$, usando o ajuste atual para $\boldsymbol{\theta}$. Dessa forma, definimos $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ (função-Q) como a esperança da função de log-verossimilhança dos

dados completos, ou seja, $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = E\{\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}_{obs}, \widehat{\boldsymbol{\theta}}\}, \boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Cada iteração do algoritmo EM consiste de dois passos: esperança (passo E) e maximização (passo M). A $(t+1)$-ésima iteração do algoritmo EM é definida como

Passo E: Para $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}^{(t)}$, calcular $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(t)})$ como

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

Passo M: Obter $\widehat{\boldsymbol{\theta}}^{(t+1)}$ que maximize $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(t)})$, tal que

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

Deve-se alternar os passos E e M repetidamente até atingir a convergência. Cada iteração do algoritmo EM incrementa o logaritmo da função de verossimilhança observada $\ell(\widehat{\boldsymbol{\theta}}|\mathbf{y}_{obs})$ de modo que $\ell(\widehat{\boldsymbol{\theta}}^{(t)}|\mathbf{y}_{obs}) \leq \ell(\widehat{\boldsymbol{\theta}}^{(t+1)}|\mathbf{y}_{obs})$ e o algoritmo tipicamente converge a um máximo local ou global da função de verossimilhança. Como critério de convergência, podemos utilizar $||\widehat{\boldsymbol{\theta}}^{(t+1)} - \widehat{\boldsymbol{\theta}}^{(t)}|| < \epsilon$, onde $||\mathbf{a}||$ indica a norma do vetor $\mathbf{a}$ e $\epsilon > 0$.

Quando o passo M do algoritmo EM é complicado, este pode ser amenizado realizando o processo de maximização condicional a alguma função dos parâmetros que estão sendo estimados. Este algoritmo EM generalizado proposto por Meng & Rubin (1993) é denominado algoritmo de maximização condicional de esperança (ECM). A idéia neste caso é substituir o passo M do algoritmo EM por uma sequência de passos de maximização condicional (CM) computacionalmente mais simples, em que cada um deles maximiza a função-Q sujeita às restrições em $\boldsymbol{\theta}$. É importante notar que as propriedades de simplicidade, estabilidade e convergência monótona do algoritmo EM são compartilhadas pelo algoritmo ECM, porém com taxas de convergência mais velozes. Neste trabalho, denominamos essa variante do algoritmo EM por algoritmo tipo-EM.

2.2 Seleção de Modelos

Neste trabalho, verificamos a adequação dos modelos baseados na classe de distribuições de misturas de escala normal aos dados inspecionando dois critérios de informação: o critério de informação de Akaike (AIC) definido por $AIC = -2\ell(\widehat{\boldsymbol{\theta}}) + 2p$ e o critério de informação de Schwarz (SIC) definido por $SIC = -2\ell(\widehat{\boldsymbol{\theta}}) + p \log n$, onde $\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta} \mid \mathbf{y}_{obs})$ é a função de log-verossimilhança dos dados observados, $\widehat{\boldsymbol{\theta}}$ é a estimativa de máxima verossimilhança de $\boldsymbol{\theta}$, n é o tamanho da amostra e p é o número de parâmetros livres no modelo. 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.O modelo que apresentar o menor valor do critério de informação será o modelo selecionado.

2.3 Distribuições de Mistura de Escala Normal

O interesse pelo estudo teórico de distribuições simétricas que possuem caudas mais pesadas que a distribuição normal tem sido crescente desde a década de 70. Neste contexto, Andrews & Mallows (1974) apresentam uma classe de distribuições simétricas denominadas de Mistura de Escala Normal que agrupa distribuições simétricas, dentre elas as distribuições normal e t-Student. Nesta seção, apresentam-se as principais características desta classe de distribuições.

2.3.1 Definição e Notação

Uma variável aleatória Y tem distribuição de Mistura de Escala Normal com parâmetro de locação $\mu \in \mathbb{R}$ e parâmetro de dispersão σ^2 (positivo) se a sua função de densidade de probabilidade (f.d.p.) é dada por

$$f(y) = \int_0^\infty \phi(y; \mu, \kappa(u)\sigma^2) dH(\mu; \nu), \tag{2.1}$$

onde $\phi(.; \mu, \sigma^2)$ denota a f.d.p. de uma distribuição normal univariada com média μ e variância σ^2, $\kappa(.)$ é uma função de ponderação positiva e U é uma variável aleatória positiva com função de distribuição acumulada (f.d.a.) $H(u; \boldsymbol{\nu})$, tal que $\boldsymbol{\nu}$ é um escalar ou vetor de parâmetros da distribuição de U que controla as caudas das distribuições. Note que o termo $\phi(.; \mu, \sigma^2)$ depende da distância de Mahalanobis

$$d = \frac{(y-\mu)^2}{\sigma^2}, \tag{2.2}$$

útil na identificação de observações aberrantes; veja, por exemplo, Pinheiro et al. (2001). Esta distribuição será denotada por $SMN(\mu, \sigma^2, \boldsymbol{\nu})$.

A seguir será apresentada a representação estocástica de uma variável aleatória que tem distribuição de Mistura de Escala Normal. Tal representação é de suma importância no decorrer deste trabalho, visto que inúmeros resultados podem ser derivados a partir dessa caracterização.

2.3.2 Representação Estocástica

Seja Y uma variável aleatória com distribuição $SMN(\mu, \sigma^2, \boldsymbol{\nu})$. Dessa forma,Y pode ser representada estocasticamente como

$$Y = \mu + \kappa^{1/2}(U)Z, \tag{2.3}$$

onde Z é uma variável aleatória normal unidimensional com média zero e variância σ^2,e U é uma variável aleatória positiva com f.d.a. $H(u; \boldsymbol{\nu})$ e f.d.p. $h(u; \boldsymbol{\nu})$, independente de Z. A representação estocástica dada em (2.3), além de facilitar a implementação do

algoritmo EM, pode ser usada também para derivar muitas propriedades da distribuição de Y, um exemplo seria a forma quadrática da distância de Mahalanobis d, definida em (2.2). Além disso, a partir de (2.3) a distribuição de mistura de escala normal pode ser reescrita hierarquicamente como apresentada na Proposição abaixo.

Proposição 2.3.1. *Sabendo que Y tem distribuição de Mistura de Escala Normal unidimensional $SMN(\mu, \sigma^2, \boldsymbol{\nu})$, então de Y condicionada a U tem distribuição normal unidimensional da forma*

$$Y \mid U = u \sim N(\mu, k(u)\sigma^2) \quad e \quad U \sim h(u; \boldsymbol{\nu}). \tag{2.4}$$

Demonstração. A prova é realizada diretamente a partir da representação estocástica dada em (2.3). □

Tal representação hierárquica será útil para obter algumas propriedades interessantes da distribução de Mistura de Escala Normal e também realizar inferência estatística nos modelos de regressão linear, no contexto de simetria. Sendo assim, algumas propriedades interessantes da distribução de Mistura de Escala Normal são:

(P1) Se a esperança existe, então $E[Y] = \mu$, se $E[\kappa^{1/2}(U)] < \infty$.

(P2) Se a variância existe, então $V[Y] = E[\kappa(U)]\sigma^2$, se $E[\kappa(U)] < \infty$.

As distribuições de Mistura de Escala Normal são constituídas por famílias paramétricas de distribuições probabilísticas que preservam a estrutura simétrica das distribuições normais. A família de distribuções normais é um elemento particular desta classe. Outras famílias conhecidas que compõem esta classe de distribuições são: t-student, Slash e Normal Contaminada. Para uma discussão mais detalhada quanto às distribuções de Mistura de Escala Normal veja, por exemplo, Andrews & Mallows (1974).

2.3.3 Algumas Distribuições Específicas

Quatro distribuições de probabilidade conhecidas, e que formam parte da classe de Mistura de Escala Normal, são detalhadas na Tabela 1, e para cada uma delas se indica as funções $\kappa(.)$ e U que as caracterizam.

Tabela 1 – Quatro distribuições de Mistura de Escala Normal univariadas.

Distribuição	Notação	$\kappa(.)$	U	
Normal	$N(\mu, \sigma^2)$	1	$Degenerada$	
t-Student	$t(\mu, \sigma^2, \nu)$	$1/u$	$Gamma(\frac{\nu}{2}, \frac{\nu}{2})$	(1)
Slash	$SL(\mu, \sigma^2, \nu)$	$1/u$	$Beta(\nu, 1)$	
Normal Contaminada	$CN(\mu, \sigma^2, \nu, \gamma)$	$1/u$	$Discreta$	(2)

(1) Considerando $Gamma(a, b)$ com média $\frac{a}{b}$.

(2) $Discreta\ com\ f.d.p.\ h(u; \boldsymbol{\nu}) = \nu \mathbb{I}_{(u=\gamma)} + (1-\nu)\mathbb{I}_{(u=1)},\ 0 \leq \nu \leq 1,\ 0 < \gamma \leq 1.$

Abaixo serão descritos os principais conceitos das duas distribuições que serão mais utilizadas neste trabalho.

2.3.3.1 Distribuição Normal

A normal é a distribuição pertencente à classe simétrica mais utilizada devido a todo o desenvolvimento teórico e aplicado estabelecido no decorrer dos anos. Se $Y \sim N(\mu, \sigma^2)$, então a f.d.p é da forma

$$f_Y(y) = \phi(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(y-\mu)^2\}, \qquad y \in \mathbb{R},$$

onde seu valor esperado e variância são, respectivamente,

$$E(Y) = \mu \qquad \text{e} \qquad Var(Y) = \sigma^2.$$

2.3.3.2 Distribuição t-Student

A variável aleatória Y tem distribuição t-Student com ν graus de liberdade, ou seja, $Y \sim t(\mu, \sigma^2, \nu)$, se sua f.d.p é dada por

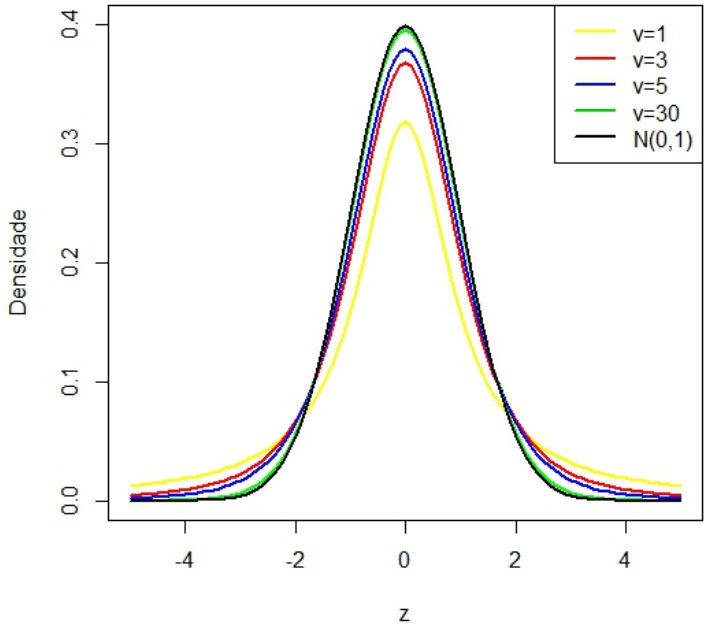
$$f_Y(y) = \frac{\Gamma(\frac{1+\nu}{2})}{\Gamma(\frac{\nu}{2})\pi^{1/2}} \nu^{-1/2} \sigma^{-1} \left(1 + \frac{d}{\nu}\right)^{-(1+\nu)/2}, \qquad y \in \mathbb{R}. \tag{2.5}$$

Dessa forma, temos que

$$E(Y) = \mu \qquad \text{e} \qquad Var(Y) = \sigma^2 \frac{\nu}{\nu - 2},\ \nu > 2.$$

Como ilustração, temos na Figura 9 as funções densidades de probabilidade das distribuições t-Student para $\nu = 1, 3, 5$ e 30, comparando com a função densidade de probabilidade da distribuição normal. Para todas as distribuições consideradas, os parâmetros de locação e escala são fixados em $\mu = 0$ e $\sigma^2 = 1$, respectivamente, apenas para ilustração. Pela Figura 9, podemos ver que a distribuição t-Student tende a uma distribuição normal quando $\nu \to \infty$. Além disso, quando $\nu = 1$, temos a distribuição Cauchy.

Figura 2 – Gráficos das distribuições t-Student com diferentes valores de ν.



2.4 Distância de Mahalanobis

Em seguida, descrevemos algumas propriedades da distância de Mahalanobis $d = \dfrac{(Y-\mu)^2}{\sigma^2}$. Mais detalhes sobre propriedades das formas quadráticas podem ser encontradas em Lange & Sinsheimer (1993), por exemplo. Dessa forma, temos que $d \sim \chi^2_1$ para o caso normal e $d \sim F(1,\nu)$ para o caso t-Student. Este resultado é interessante, pois permite avaliar os modelos estatísticos na prática. Substituindo as estimativas de máxima verossimilhança de μ e σ^2 na distância de Mahalanobis d, podemos avaliar os ajustes dos modelos através da construção de envelopes. Além disso, através de gráficos da distância de Mahalanobis e considerando como marca de referência o quantil υ da distribuição da forma quadrática d, podemos identificar "outliers". Por exemplo, para o caso normal, temos que $\upsilon = \chi^2_1(\xi)$, onde $0 < \xi < 1$.

2.5 Modelo de Regressão Linear

Nesta seção, consideramos os modelos de regressão linear, onde as observações seguem distribuições SMN, especificamente a distribuição t-Student. Em geral, o modelo de regessão linear normal é definido como

$$Y_i = \alpha + \beta x_i + \varepsilon_i, \;\; i = 1, ..., n, \tag{2.6}$$

onde os erros aleatórios são considerados independentes com distribuição $N(0, \sigma^2)$. Neste trabalho, seguindo, por exemplo, Lange et al. (1989), vamos substituir a usual suposição de normalidade para os erros, pela seguinte suposição mais flexível

$$\varepsilon_i \overset{iid}{\sim} t(0, \sigma^2, \nu),\ i = 1, ..., n. \tag{2.7}$$

Note que a partir de (2.7), temos que $E(\varepsilon_i) = 0$ e, consequentemente,

$$E(Y_i) = \alpha + \beta x_i,$$

pois

$$Y_i \overset{ind}{\sim} t(\alpha + \beta x_i, \sigma^2, \nu).$$

2.5.1 Função de Log-Verossimilhança

A função de log-verossimilhança para $\boldsymbol{\theta}$=$(\boldsymbol{\beta}\ ^\top, \sigma^2)^\top$ em que $\boldsymbol{\beta}$= $(\alpha, \beta)^\top$, onde assume-se que o parâmetro ν (associado com a variável de mistura U) é fixo, dada uma amostra $Y_1, ..., Y_n$ é definida por $\ell(\boldsymbol{\theta})$, tal que

$$\ell(\boldsymbol{\theta}) = n\ \log\ c(\nu) - \frac{n}{2} \log\ \sigma^2 - \left(\frac{\nu+1}{2}\right) \sum_{i=1}^{n} \log \left\{1 + \frac{d_i(\beta, \sigma^2)}{\nu}\right\} \tag{2.8}$$

onde

$$d_i = d_i(\boldsymbol{\beta}, \sigma^2) = \frac{\left(Y_i - \mathbf{x}_i^\top \boldsymbol{\beta}\right)^2}{\sigma^2}, i = 1, ..., n, \tag{2.9}$$

é a distância de Mahalanobis, $\mathbf{x}_i^\top$ corresponde a i-ésima linha da matriz $\mathbf{X}_{(n\times 2)}$, em que $\mathbf{x}_i^\top$= $(1, x_i)$ e $c(\nu) = \dfrac{\Gamma\left(\frac{\nu+1}{2})\right)}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})}$.

A função escore pode ser obtida derivando o logaritmo da função de verossimilhança, com respeito a cada um dos parâmetros desconhecidos. Note que não existem soluções explicitas para o problema de maximização da função de log-verossimilhança dada em (2.8). Neste caso, podemos maximizar numericamente usando, por exemplo, o Matlab e o R. Estes *softwares* contêm rotinas prontas para tratar problemas de maximização (minimização) de qualquer função.

Entretanto, neste trabalho, apresentamos o algoritmo EM para o cálculo dos estimadores de máxima verossimilhança de modo a obtermos soluções analíticas para os estimadores de $\boldsymbol{\beta}$ e σ^2.

2.5.2 Estimação dos Parâmetros via Algoritmo EM

Note que o modelo descrito em (2.6)-(2.7) pode ser descrito hierarquicamente como

$$Y_i \mid U_i = u_i \sim N\left(x_i^\top \boldsymbol{\beta}, \frac{\sigma^2}{u_i}\right) \tag{2.10}$$

$$U_i \sim Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right), i = 1, ..., n. \tag{2.11}$$

Neste processo de estimação, considere $\mathbf{y}= (y_1, ..., y_n)^\top$ o vetor de respostas observáveis para n unidades amostrais e $\mathbf{u}= (u_1, ..., u_n)^\top$. Então, sob representação hierárquica (2.10)-(2.11), segue que a função de log-verossimilhança completa associada com $\mathbf{y}_c= (\mathbf{y}^\top, \mathbf{u}^\top)$ é dada por

$$\begin{aligned} \ell_c(\boldsymbol{\theta} \mid \mathbf{y}_c) &\propto -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n} u_i(y_i - x_i^\top\boldsymbol{\beta})^2 \\ &= -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top D(\mathbf{u})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \end{aligned}$$

onde D($\mathbf{u}$)= $diag(u_1, ..., u_n)$.

Após manipulações algébricas, a esperança condicional da função de log-verossimilhança completa é dada por $Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(t)})$ e tem a forma

$$\begin{aligned} Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(t)}) &= E[\ell_c(\boldsymbol{\theta} \mid \mathbf{y}_c), \widehat{\boldsymbol{\theta}}^{(t)}] \\ &= -\frac{n}{2}\log\widehat{\sigma^2}^{(t)} - \frac{1}{2\widehat{\sigma^2}^{(t)}}\sum_{i=1}^{n}\widehat{u}_i^{(t)}(y_i - x_i^\top\widehat{\boldsymbol{\beta}}^{(t)})^2 \\ &= -\frac{n}{2}\log\widehat{\sigma^2}^{(t)} - \frac{1}{2\widehat{\sigma^2}^{(t)}}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{(t)})^\top\mathbf{D}(\widehat{\mathbf{u}}^{(t)})(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{(t)}). \end{aligned} \tag{2.12}$$

A $(t + 1)$- ésima iteração do algoritmo pode ser resumida como se segue.

PASSO E: A partir das estimativas iniciais para $\boldsymbol{\theta}$, os pesos $\widehat{u}_i^{(t)}$ são obtidos através das seguinte esperança condicional:

$$E(U_i \mid y_i, \widehat{\theta}^{(t)}) = \widehat{u}_i^{(t)} = \frac{\widehat{\nu}^{(t)} + 1}{\widehat{\nu}^{(t)} + \widehat{d}_i^{(t)}}, \tag{2.13}$$

onde $\widehat{d}_i^{(t)} = d_i(\widehat{\boldsymbol{\beta}}^{(t)}, \widehat{\sigma^2}^{(t)})$. Note que $E(U_i \mid y_i, \widehat{\theta}^{(t)})$ é inversamente proporcional à distância de Mahalanobis. Então, quanto maior o valor de $d(\widehat{\boldsymbol{\beta}}^{(t)}, \widehat{\sigma^2}^{(t)})$, temos um menor valor para $E(U_i \mid y_i, \widehat{\theta}^{(t)})$ e assim o procedimento de estimação tende a dar um menor peso para as observações atípicas no sentido da distância de Mahalanobis. Dessa forma, observamos que quando utilizamos distribuições com caudas mais pesadas que a distribuição normal, o algoritmo EM acomoda as observações atípicas atribuindo-lhes menor peso no processo de estimação.

PASSO M: Usando os pesos obtidos no passo E do algoritmo, as estimativas de máxima verossimilhança são obtidas como:

$$\widehat{\boldsymbol{\beta}}^{(t+1)} = (\mathbf{X}^\top D(\widehat{\mathbf{u}}^{(t)})\mathbf{X})^{-1}\mathbf{X}^\top D(\widehat{\mathbf{u}}^{(t)})\mathbf{y}, \tag{2.14}$$

$$\widehat{\sigma^2}^{(t+1)} = \frac{1}{n}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{(t)})^\top D(\widehat{\mathbf{u}}^{(t)})(\mathbf{y} - \mathbf{X}\widehat{\beta}^{(t)}). \tag{2.15}$$

Finalmente, de acordo com Lange et al. (1989), a função de verossimilhança perfilada é usada para determinar o valor ótimo de ν como segue: se $\boldsymbol{\theta}= (\boldsymbol{\beta}^\top, \sigma^2)^\top$ for o vetor de parâmetros de interesse e ν é o parâmetro *nuisance*, então a função de verossimilhança perfilada de $\boldsymbol{\theta}$ é $l_P(\boldsymbol{\theta}) = max_\nu \ell(\boldsymbol{\theta}, \nu)$, onde $\ell(\boldsymbol{\theta}, \nu)$ é a log-verossimilhança definida em (2.8).

Note que se $u_i = 1,\ \forall\ i = 1, ..., n$, temos que os resultados expostos acima coincidem com as estimativas de máxima verossimilhança do modelo de regressão linear sob erros normais.

2.6 *Outliers* e Robustez

Os termos *outliers* e robustez desempenham um papel importante nesta monografia. O termo *outliers* muitas vezes é usado informalmente e 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.

Um dos objetivos subjacentes às técnicas de estimação consideradas nesta monografia é o desenvolvimento de procedimentos sob a classe de distribuições SMN que sejam robustos na presença de *outliers*, ou seja, métodos estatísticos que não sejam afetados (ou menos afetados) por observações extremas, comumente chamadas de observações aberrantes (*outliers*, anômalas ou atípicas). Uma das características dessa classe rica de modelos é que as distribuições SMN podem naturalmente atribuir pesos diferentes para cada observação e consequentemente controlar a influência da observação no processo de estimação, por exemplo.

3 MODELO DE REGRESSÃO LINEAR T-STUDENT COM PONTO DE MUDANÇA

Diversos modelos estatísticos robustos considerando classes de distribuições simétricas têm sido apresentados na literatura estatística nos últimos anos.Veja por exemplo, Fernandez & Steel (2000) e Rosa et al. (2003). Estes modelos são apresentados como uma alternativa aos modelos que assumem normalidade e que se tornam restritivos em algumas situações práticas de modelagem estatística.

Neste capítulo, estudamos modelos de regressão linear robustos com erros distribuídos de acordo com uma distribuição da família de mistura de escala normal, especificamente a distribuição t-Student, e que apresentam um ponto de mudança. Assim, considera-se que os erros aleatórios apresentam uma distribuição de caudas mais pesadas que a distribuição normal. O modelo de regressão linear clássico com ponto de mudança que considera os erros normalmente distribuídos é um caso particular, pois a distribuição normal compõe a classe de distribuições considerada. Alguns trabalhos relevantes no contexto de ponto de mudança contínuo sob a suposição de normalidade são Sprent (1961), Hinkley (1969), Worsley (1983), Liu et al. (1997), Julious (2001), Muggeo (2003) e Young (2014).

Primeiramente apresenta-se de forma geral a especificação do modelo proposto com um ponto de mudança contínuo, na estrutura da média (isto é, nos coeficientes de regressão). Em seguida, é realizado um estudo de inferência com o intuito de estimar os parâmetros do modelo, incluindo o ponto de mudança. Vale ressaltar que, neste trabalho, consideramos o ponto de mudança como "parâmetro verdadeiro" (desconhecido) e este é estimado via algoritmo EM. Finalmente, exemplos numéricos considerando dados simulados e reais são utilizados para ilustrar o modelo e os resultados inferenciais desenvolvidos aqui.

3.1 Descrição do Modelo

De acordo com Muggeo (2003) e Young (2014), o modelo de regressão linear com ponto de mudança, descrito em (1.1), pode ser reescrito, no contexto em que o modelo é assumido como contínuo no ponto de mudança, isto é,

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 (x_i - \gamma)_+ + \varepsilon_i, \quad i = 1, \ldots, n, \tag{3.1}$$

onde $(x_i - \gamma)_+ = (x_i - \gamma) I_{\{x_i > \gamma\}}$, tal que $I_{\{\cdot\}}$ é a função indicadora e γ é o ponto de mudança. Note que se $x_i > \gamma$, então $(x_i - \gamma)_+ = (x_i - \gamma)$ e por sua vez, se $x_i \leq \gamma$, temos que $(x_i - \gamma)_+ = 0$. Portanto, observe que temos uma reta à esquerda do ponto de mudança dada por

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

enquanto que à direita do ponto de mudança, temos :

$$Y_i = (\beta_0 - \beta_2\gamma) + (\beta_1 + \beta_2)x_i + \varepsilon_i.$$

Ou seja, temos que β_1 é a inclinação da reta à esquerda do ponto de mudança, e $(\beta_1 + \beta_2)$ é a inclinação da reta à direita . Assim, β_2 é o parâmetro que representa a "diferença nas inclinações" dessas retas.

Neste trabalho, estendemos o modelo definido acima considerando a relação linear em (3.1) com a seguinte suposição

$$\varepsilon_i \overset{iid}{\sim} t(0, \sigma^2, \nu),\ i = 1, ..., n, \tag{3.2}$$

e consequentemente, temos que

$$Y_i \overset{ind}{\sim} t(\beta_0 + \beta_1 x_i + \beta_2(x_i - \gamma)_+, \sigma^2, \nu),\ i = 1, ..., n. \tag{3.3}$$

A seguir, os estimadores de máxima verossimilhança de $\boldsymbol{\theta} = (\gamma, \beta_0, \beta_1, \beta_2, \sigma^2)^\top$ são descritos, onde assume-se que ν é fixo. O algoritmo EM é usado para estimar os parâmetros devido a sua simplicidade para implementação.

3.2 Estimação dos Parâmetros via Algoritmo EM

Note que o modelo descrito em (3.1)-(3.2) pode ser descrito hierarquicamente como

$$Y_i \mid U_i = u_i \sim N\left(\beta_0 + \beta_1 x_i + \beta_2(x_i - \gamma)_+, \frac{\sigma^2}{u_i}\right), \tag{3.4}$$

$$U_i \sim Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right), i = 1, ..., n. \tag{3.5}$$

Neste processo de estimação, considere $\mathbf{y} = (y_1, ..., y_n)^\top$ o vetor de respostas observáveis para n unidades amostrais e $\mathbf{u} = (u_1, ..., u_n)^\top$. Então, sob a representação hierárquica (3.4)-(3.5), segue que a função de log-verossimilhança completa associada com $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{u}^\top)$ é dada por

$$\ell_c(\boldsymbol{\theta} \mid \mathbf{y}_c) \ \propto\ -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n} u_i(y_i - \beta_0 + \beta_1 x_i + \beta_2(x_i - \gamma)_+)^2.$$

Após manipulações algébricas, a esperança condicional da função de log-verossimilhança completa, $Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(t)})$, tem a mesma expressão matricial definida em (2.12), dada por

$$\begin{aligned} Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(t)}) &= E[\ell_c(\boldsymbol{\theta} \mid \mathbf{y}_c), \widehat{\boldsymbol{\theta}}^{(t)}] \\ &= -\frac{n}{2}\log\widehat{\sigma^2}^{(t)} - \frac{1}{2\widehat{\sigma^2}^{(t)}}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{(t)})^\top \mathbf{D}(\hat{\mathbf{u}}^{(t)})(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{(t)}. \end{aligned} \tag{3.6}$$

Porém, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^\top$ e $X = \begin{bmatrix} 1 & x_1 & (x_1-\gamma)_+ \\ 1 & x_2 & (x_2-\gamma)_+ \\ \vdots & \vdots & \vdots \\ 1 & x_n & (x_n-\gamma)_+ \end{bmatrix}$, com $D(\mathbf{u}) = diag(u_1, ..., u_n)$.
Para $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^\top)^\top$, onde $\theta_1 = \gamma$ e $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}^\top, \sigma^2)^\top$, a $(t+1)$- ésima iteração do algoritmo pode ser resumida como se segue.

PASSO E: A partir das estimativas iniciais para $\boldsymbol{\theta}$, os pesos $\widehat{u}_i^{(t)}$ são calculados como em (2.13).

Usando os pesos obtidos no **PASSO E**, as estimativas de máxima verossimilhança são obtidas como:

PASSO1-CM: Neste passo, ocorre a estimação do ponto de mudança, onde assumimos que o ponto de mudança deve ocorrer dentro do domínio de x, e calculamos

$$\widehat{\theta}_1^{(t+1)} = \underset{\theta_1}{\operatorname{argmax}}\, Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(t)})$$

com θ_2 fixado como $\widehat{\boldsymbol{\theta}}_2^{(t)}$.

PASSO2-CM: Neste passo, calculamos

$$\widehat{\theta}_2^{(t+1)} = \underset{\theta_2}{\operatorname{argmax}}\, Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(t)})$$

com θ_1 fixado como $\widehat{\boldsymbol{\theta}}_1^{(t)}$. Dessa forma, atualizar $\widehat{\theta}_2^{(t+1)}$ consiste em obter

$$\widehat{\boldsymbol{\beta}}^{(t+1)} = (\mathbf{X}^\top D(\widehat{\mathbf{u}}^{(t)})\mathbf{X})^{-1}\mathbf{X}^\top D(\widehat{\mathbf{u}}^{(t)})\mathbf{y}, \tag{3.7}$$

$$\widehat{\sigma^2}^{(t+1)} = \frac{1}{n}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{(t)})^\top D(\widehat{\mathbf{u}}^{(t)})(\mathbf{y} - \mathbf{X}\widehat{\beta}^{(t)}). \tag{3.8}$$

Note que se $u_i = 1, i = 1, ..., n$, temos que os resultados expostos acima coincidem com as estimativas de máxima verossimilhança do modelo de regressão linear sob erros normais.

Finalmente, de acordo com Lange et al. (1989), a função de verossimilhança perfilada é usada para determinar o valor ótimo de ν, como descrito na Seção 2.5.2, onde ressaltamos que $\boldsymbol{\theta} = (\gamma, \boldsymbol{\beta}^\top, \sigma^2)^\top$, com $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^\top$. Deve-se alternar os passos E e M repetidamente até atingir a convergência. Neste capítulo, usamos o critério de convergência, definido na Seção 2.1, com $\epsilon = 10^{-6}$. Além disso, valores iniciais são necessários para implementar este algoritmo. Como "chute inicial"para o ponto de mudança γ, utilizamos a mediana dos 100 valores gerados aleatoriamente de uma uniforme no intervalo [a,b], domínio correspondente de x. Fixado este ponto de mudança, ajustamos o modelo de regressão linear via mínimos quadrados e obtivemos os "chutes"iniciais para os demais parâmetros. Entretanto, para garantir que as estimativas de máxima verossimilhanças foram encontradas, recomendamos a execução do algoritmo EM usando uma variedades de diferentes valores iniciais.

3.3 Exemplos Numéricos

Nesta seção, aplicações a dados reais e estudos de simulação são apresentados a fim de ilustrar o modelo e os resultados inferenciais desenvolvidos.

3.3.1 Estudos de Simulação

A qualidade do método de estimação dos parâmetros do modelo e o desempenho desses estimadores será avaliado usando dados simulaçoes de Monte Carlo.

3.3.1.1 Experimento 1: Estimação via Algoritmo EM

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 t-Student com ponto de mudança, sem descontinuidade, aos dados que foram gerados artificialmente.

Neste caso, geramos 500 amostras de tamanho $n = 25, 50, 100$ e 200 provenientes do modelo de regressão linear t-Student com ponto de mudança contínuo, considerando $\gamma = 4, 6$ e 8.A variável explicativa utilizada foi $x_i \sim U(1, 10)$ e os erros $\epsilon_i \sim t(0, 1, 3)$, com os seguintes valores para os parâmetros: $\beta_0 = 2, \beta_1 = -1$ e $\beta_2 = 3$. Os valores médios, medianos e os correspondentes desvios padrão das estimativas obtidas via algoritmo EM em todas as amostras foram calculados e os resultados estão apresentados na Tabela 2.

Note que o algoritmo proposto é eficiente para a estimação dos parâmetros do modelo de regressão linear t-Student, no contexto de ponto de mudança, considerando diferentes posições de ponto de mudança e tamanhos amostrais.

Tabela 2 – Média (Md), Mediana (Med) e os desvios padrãoes (SD) das estimativas obtidas via algoritmo EM baseadas em 500 amostras provenientes do modelo de regressão linear t-Student com ponto de mudança contínuo. Os valores verdadeiros dos parâmetros estão entre parênteses.

n=25		$\gamma = 4$			$\gamma = 6$			$\gamma = 8$	
Parâmetros	Md	SD	Med	Md	SD	Med	Md	SD	Med
$\beta_0(2)$	1.808	2.706	1.758	2.235	1.104	2.175	2.351	1.004	2.270
$\beta_1(-1)$	-0.917	1.161	-0.854	-1.072	0.303	-1.047	-1.120	0.275	-1.085
$\beta_2(3)$	2.986	1.156	2.928	3.122	0.550	3.089	2.571	1.430	2.402
$\sigma^2(1)$	0.772	0.455	0.716	0.698	0.399	0.653	0.761	0.442	0.666
ν	2.858	1.113	3.000	2.796	1.139	2.875	2.858	1.100	3.000
γ	4.248	0.601	4.171	5.955	0.427	5.965	7.160	1.162	7.523

n=50		$\gamma = 4$			$\gamma = 6$			$\gamma = 8$	
Parâmetros	Md	SD	Med	Md	SD	Med	Md	SD	Med
$\beta_0(2)$	1.807	1.089	1.831	2.132	0.724	2.107	2.187	0.556	2.181
$\beta_1(-1)$	-0.914	0.421	-0.920	-1.038	0.198	-1.024	-1.057	0.144	-1.048
$\beta_2(3)$	2.949	0.424	2.960	3.039	0.312	3.042	2.838	1.222	2.776
$\sigma^2(1)$	0.884	0.321	0.857	0.883	0.322	0.845	0.913	0.350	0.872
ν (3)	3.011	0.899	3.000	2.975	0.909	3.000	2.986	0.895	3.000
γ	4.119	0.363	4.074	5.966	0.292	5.959	7.622	0.754	7.833

n=100		$\gamma = 4$			$\gamma = 6$			$\gamma = 8$	
Parâmetros	Md	SD	Med	Md	SD	Med	Md	SD	Med
$\beta_0(2)$	1.890	0.766	1.898	2.009	0.480	2.019	2.077	0.350	2.087
$\beta_1(-1)$	-0.952	0.292	-0.953	-1.017	0.127	-1.005	-1.024	0.078	-1.021
$\beta_2(3)$	2.972	0.298	2.971	2.995	0.224	2.981	2.873	0.620	2.912
$\sigma^2(1)$	0.953	0.239	0.937	0.945	0.239	0.919	0.957	0.236	0.939
ν (3)	3.075	0.755	3.000	3.077	0.771	3.000	3.094	0.758	3.000
γ	4.069	0.225	4.038	5.975	0.186	5.986	7.823	0.440	7.922

n=200		$\gamma = 4$			$\gamma = 6$			$\gamma = 8$	
Parâmetros	Md	SD	Med	Md	SD	Med	Md	SD	Med
$\beta_0(2)$	1.950	0.463	1.956	2.044	0.320	2.050	2.031	0.241	2.030
$\beta_1(-1)$	-0.975	0.184	-0.975	-1.015	0.086	-1.017	-1.007	0.048	-1.006
$\beta_2(3)$	2.983	0.189	2.982	3.000	0.144	3.003	2.991	0.379	2.999
$\sigma^2(1)$	0.965	0.172	0.955	0.984	0.181	0.973	0.970	0.173	0.957
ν (3)	3.041	0.602	3.000	3.124	0.635	3.000	3.090	0.629	3.000
γ	4.031	0.142	4.029	5.971	0.134	5.977	7.966	0.189	7.986

3.3.1.2 Experimento 2: Desempenho dos Estimadores de Máxima Verossimilhança

O segundo objetivo é avaliar o desempenho dos estimadores de máxima verossimilhança dos parâmetros do modelo de regressão linear t-Student com ponto de mudança contínuo.

Consideramos os mesmos cenários simulados no experimento 1 (Seção 3.3.1.1),

variando os tamanhos amostrais e as posições do ponto de mudança.

Com o intuito de avaliar as propriedades de consistência e eficiência assintótica dos estimadores de máxima verossimilhança, calculamos o viés médio (Vies Md), viés mediano (Vies Med), erro quadrático médio (EQM) e o desvio absoluto mediano (DAM), definidos por:

$$\begin{aligned} Vies\ \ Md_i &= \frac{1}{m}\sum_{j=1}^{m}(\widehat{\theta}_{ij}-\theta_i), \\ Vies\ \ Med_i &= \underset{1\leq j\leq m}{\text{med}}\,(\widehat{\theta}_{ij}-\theta_i), \\ EQM_i &= \sqrt{\frac{1}{m}\sum_{j=1}^{m}\left(\widehat{\theta}_{ij}-\theta_i\right)^2} \ \text{ e} \\ DAM_i &= \underset{1\leq j\leq m}{\text{med}}\left|\widehat{\theta}_{ij}-\underset{1\leq l\leq m}{\text{med}}\,\theta_{il}\right|, \end{aligned}$$

$i = 1, \ldots, k$, onde m é o número de simulações realizadas e k é a dimensão do vetor de parâmetros $\boldsymbol{\theta}$.

As Tabelas 3 e 4 apresentam os valores do viés médio (Vies Md), viés mediano (Vies Med), erro quadrático médio (EQM) e o desvio absoluto mediano (DAM) das estimativas obtidas via algoritmo EM baseadas em 500 amostras provenientes do modelo de regressão linear t-Student com ponto de mudança contínuo. Observa-se que, em geral, o víes e a variablidade das estimativas dos parâmetros decrescem quando o tamanho da amostra aumenta, como esperado. Ou seja, isto concorda essencialmente com as propriedades assintóticas dos estimadores de máxima verossimilhança.

Estes resultados também podem ser visualizados pela Figura 3 que mostra os boxplots das estimativas dos parâmetros do modelo de regressão linear t-Student com ponto de mudança simulado na posição $\gamma = 8$. Para os outros modelos, os resultados são similares, e então não serão mostrados aqui para salvar espaço.

Tabela 3 – Viés médio (Vies Md), viés mediano (Vies Med), erro quadrático médio (EQM) e o desvio absoluto mediano (DAM) das estimativas obtidas via algoritmo EM baseadas em 500 amostras provenientes do modelo de regressão linear t-Student com ponto de mudança contínuo, considerando $n = 25$ e $n = 50$. Os valores verdadeiros dos parâmetros estão entre parênteses.

		n=25				**n=25**		
		$\gamma = 4$				$\gamma = 6$		
Parâmetros	Vies Md	Vies Med	EQM	DAM	Vies Md	Vies Med	EQM	DAM
$\beta_0(2)$	0.191	0.241	2.73	1.756	0.235	0.175	1.200	1.569
β_1(-1)	0.082	0.145	1.164	3.027	0.072	0.047	0.312	3.542
$\beta_2(3)$	0.013	0.071	1.157	1.257	0.122	0.089	0.661	1.710
$\sigma^2(1)$	0.227	0.284	0.508	1.804	0.301	0.346	0.500	1.948
ν	0.141	0.00	1.191	1.178	0.203	0.125	1.292	1.824
γ	0.248	0.171	1.161	2.027	0.045	0.034	0.505	3.516
		n=25				**n=50**		
		$\gamma = 8$				$\gamma = 4$		
Parâmetros	Vies Md	Vies Med	EQM	DAM	Vies Md	Vies Med	EQM	DAM
$\beta_0(2)$	0.351	0.270	1.226	1.472	0.192	0.168	1.155	1.405
β_1(-1)	0.120	0.085	0.300	3.453	0.085	0.079	0.430	3.215
$\beta_2(3)$	0.428	0.597	1.923	1.730	0.050	0.039	0.450	1.051
$\sigma^2(1)$	0.238	0.333	0.502	1.751	0.115	0.142	0.341	1.877
ν	0.141	0.000	1.179	1.729	0.011	0.000	0.900	1.287
γ	0.839	0.476	6.818	5.105	0.119	0.074	0.600	1.823
		n=50				**n=50**		
		$\gamma = 6$				$\gamma = 8$		
Parâmetros	Vies Md	Vies Med	EQM	DAM	Vies Md	Vies Med	EQM	DAM
$\beta_0(2)$	0.132	0.107	0.771	1.310	0.187	0.181	0.670	1.219
β_1(-1)	0.038	0.024	0.201	3.530	0.057	0.048	1.315	3.604
$\beta_2(3)$	0.039	0.042	0.335	1.534	0.161	0.223	0.661	1.536
$\sigma^2(1)$	0.116	0.154	0.342	1.865	0.086	0.127	0.361	1.839
ν	0.033	0.000	0.355	1.654	0.013	0.000	0.896	1.513
γ	0.033	0.040	0.355	3.461	0.377	0.166	3.112	5.256

3.4 Dados Reais

Vamos considerar um conjuntos de dados reais que já foi analisado no contexto de ponto de mudança sob o modelo normal. Veja Julious (2001) e Chen et al. (2011) para mais detalhes. Nesta seção, expandimos a análise desses dados reais, no contexto de ponto de mudança, sob o modelo t-Student. Realizamos estimação por máxima verossimilhança dos parâmetros do modelo, via algoritmo EM, e por fim, comparamos os ajustes dos modelos normal e t-Student com ponto de mudança via SIC e AIC.

3.4.1 Exemplo Prático

Quando as pessoas se exercitam, precisam produzir energia e existem diferentes caminhos metabólicos através dos quais essa energia é obtida (aeróbica e anaeróbica). Para um determinado indivíduo, é importante saber se há alteração no caminho metabólico

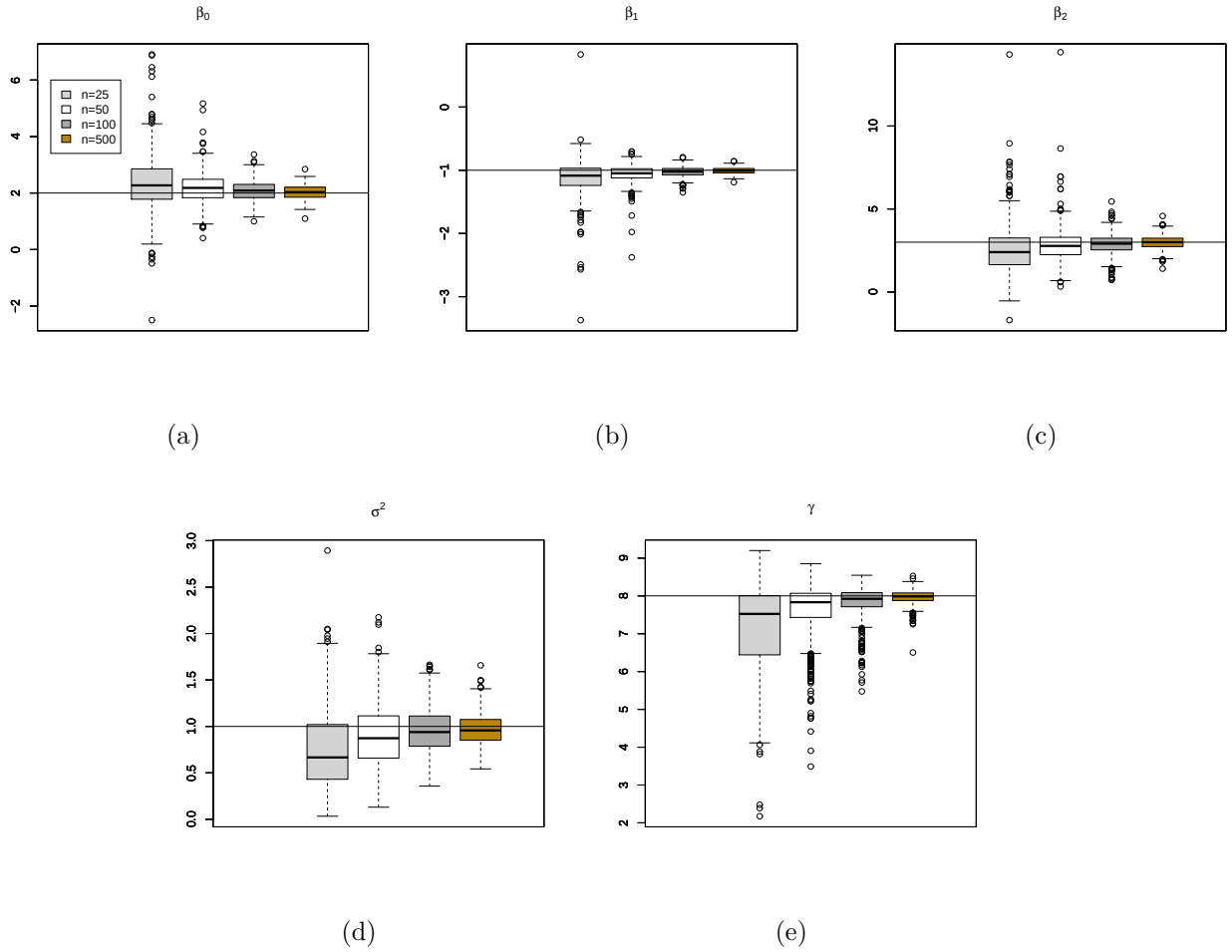
Tabela 4 – Viés médio (Vies Md), viés mediano (Vies Med), erro quadrático médio (EQM) e o desvio absoluto mediano (DAM) das estimativas obtidas via algoritmo EM baseadas em 500 amostras provenientes do modelo de regressão linear t-Student com ponto de mudança contínuo, considerando $n = 100$ e $n = 200$. Os valores verdadeiros dos parâmetros estão entre parênteses.

		n=100				**n=100**		
		$\gamma = 4$				$\gamma = 6$		
Parâmetros	Vies Md	Vies Med	EQM	DAM	Vies Md	Vies Med	EQM	DAM
$\beta_0(2)$	0.109	0.101	0.797	1.342	0.009	0.019	0.480	1.261
β_1(-1)	0.047	0.046	0.296	3.337	0.007	0.005	0.128	3.568
$\beta_2(3)$	0.027	0.028	0.309	1.068	0.004	0.018	0.224	1.526
$\sigma^2(1)$	0.046	0.062	0.244	1.876	0.054	0.080	0.245	1.885
ν	0.075	0.00	0.788	1.062	0.077	0.000	0.805	1.655
γ	0.069	0.038	0.356	1.748	0.024	0.013	0.237	3.406
		n=100				**n=200**		
		$\gamma = 8$				$\gamma = 4$		
Parâmetros	Vies Md	Vies Med	EQM	DAM	Vies Md	Vies Med	EQM	DAM
$\beta_0(2)$	0.077	0.087	0.382	1.080	0.049	0.043	0.473	1.140
β_1(-1)	0.024	0.021	0.082	3.538	0.024	0.025	0.186	3.361
$\beta_2(3)$	0.126	0.087	0.727	1.489	0.016	0.017	0.196	1.052
$\sigma^2(1)$	0.042	0.060	0.240	1.880	0.034	0.044	0.176	1.916
ν	0.094	0.00	0.809	1.662	0.041	0.000	0.615	1.029
γ	0.176	0.077	1.478	5.344	0.031	0.029	0.188	1.626
		n=200				**n=200**		
		$\gamma = 6$				$\gamma = 8$		
Parâmetros	Vies Md	Vies Med	EQM	DAM	Vies Md	Vies Med	EQM	DAM
$\beta_0(2)$	0.044	0.050	0.332	1.156	0.031	0.030	0.249	1.116
β_1(-1)	0.015	0.017	0.087	3.543	0.007	0.006	0.049	3.558
$\beta_2(3)$	0.000	0.003	0.144	1.491	0.008	0.000	0.380	1.515
$\sigma^2(1)$	0.015	0.026	0.181	1.933	0.029	0.042	0.176	1.945
ν	0.124	0.00	0.737	1.526	0.090	0.000	0.685	1.719
γ	0.028	0.022	0.217	3.375	0.033	0.013	0.328	5.330

durante o exercício e, em caso afirmativo, quando. Uma maneira de detectar isso é através da análise da relação entre duas variáveis metabólicas ao longo do tempo, enquanto a pessoa está se exercitando. Neste exemplo específico, um remador foi conectado a um equipamento de medição que lê certas respostas físicas ao longo do tempo, e sabe-se que a carga de trabalho (resistência da máquina de remo) foi aumentada ao longo do tempo.

De acordo com Julious (2001) e Chen et al. (2011), as variáveis consideradas foram os volumes de oxigênio inalado (x) e de dióxido de carbono exalado (y) em 1 minuto. As medidas foram realizadas a cada 30 segundos até um máximo de 17.5 minutos. Neste contexto, o intuito é saber se há uma relação linear entre as duas variáveis e se há uma mudança na inclinação quando um nível crítico de inalação de oxigênio é alcançado. Caso sim, o ponto de mudança representaria o ponto em que o indivíduo altera as vias metabólicas, desde aeróbica até anaeróbica.

Figura 3 – (a)-(e) Boxplots das estimativas dos parâmetros (linha indica os valores verdadeiros dos parâmetros) do modelo de regressão linear t-Student com ponto de mudança simulado em $\gamma = 8$. Legenda está no painel (a).



A Figura 4 apresenta o gráfico de dispersão desses dados, e pode-se notar uma relação linear entre os volumes de oxigênio inalado (x) e de dióxido de carbono exalado (y) em 1 minuto. Julious (2001) propõe inicialmente ajustar aos dados um modelo de regressão linear clássico (suposição de normalidade dos dados). Com propósitos exploratórios, construímos a Figura 5 que mostra o gráfico Q-Q juntamente com os envelopes simulados para este modelo clássico ajustado. Observe que o modelo gaussiano parece não fornecer um bom ajuste aos dados. Além disso, a Figura 6 apresenta o gráfico da distância de Mahalanobis (d_i) para o modelo gaussiano ajustado, e a linha de ponto de corte corresponde ao nonagésimo quinto percentil, $\upsilon = 3.841$, da forma quadrática d_i. Podemos observar pela Figura 6 que as osbervações 1 ($x_1 = 12.5$ e $y_1 = 0.75$) e 23 ($x_{23} = 48.4$ e $y_{23} = 2.96$) são detectadas como *outliers* para o modelo gaussiano ajustado, pois apresentam valores de d_i maiores que a marca de referência.

Figura 4 – Volume de dióxido de carbono exalado (litros por minuto) vs. volume de oxigênio inalado (litros por minuto).

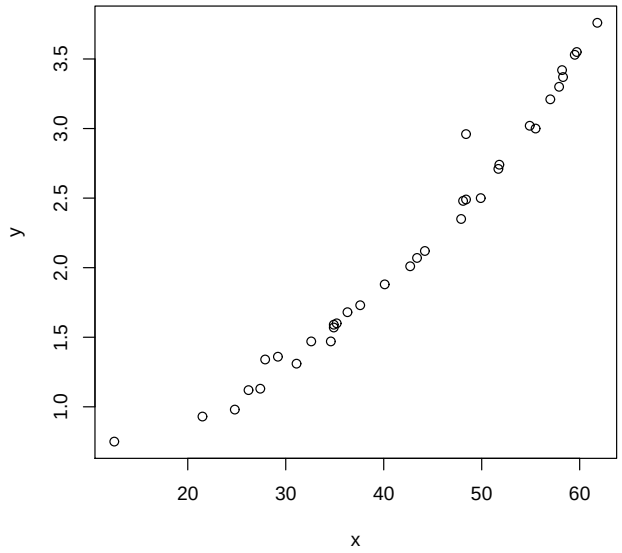
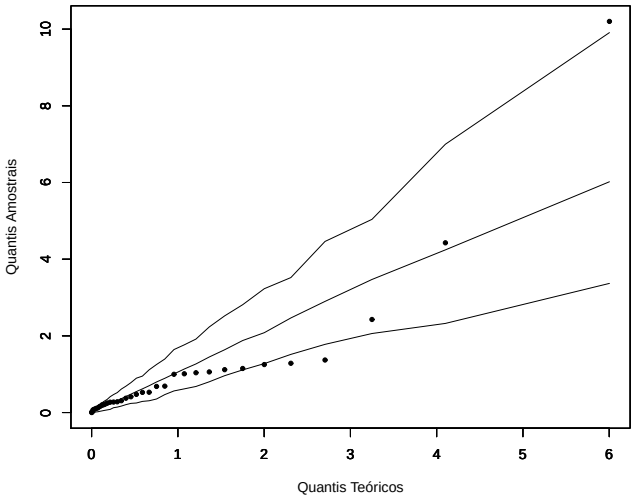
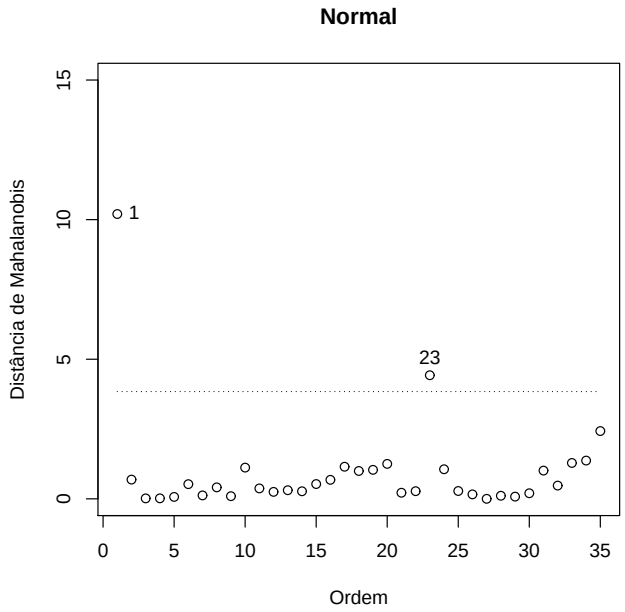


Figura 5 – Gráfico Q-Q juntamente com os envelopes simulados para modelo de regressão linear clássico ajustado.



Um dos objetivos subjacentes às técnicas de estimação consideradas nesta monografia é o desenvolvimento de procedimentos sob a classe de distribuições SMN que sejam robustos na presença de *outliers*, ou seja, métodos estatísticos que não sejam afetados (ou menos afetados) por observações extremas, comumente chamadas de observações aberrantes (*outliers*, anômalas ou atípicas). Dessa forma, nesta aplicação, revisitamos este conjunto de dados e ajustamos o modelo de regressão linear sob a

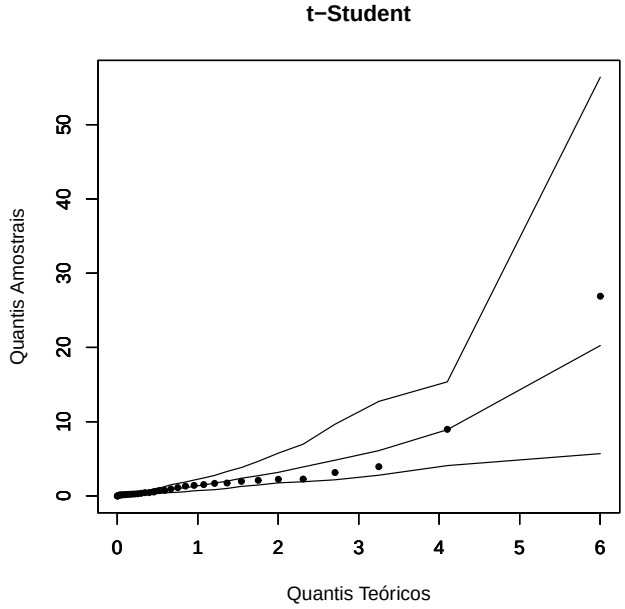
Figura 6 – Gráfico da distância de Mahalanobis vs. ordem das observações.



distribuição t-Student. Em seguida, comparamos os modelos gaussiano e t-Student inspecionando os valores do AIC. Obtivemos os seguintes valores do AIC para os modelos normal e t-Student, -16.660 e -17.919, respectivamente. Note que, pelo AIC, temos que o modelo t-Student se ajusta melhor aos dados. Adicionalmente, construímos a Figura 7 que mostra o gráfico Q-Q juntamente com os envelopes simulados para o modelo de regressão t-Student ajustado. Pela Figura 7 há evidências de que a distribuição t-Student (distribuição com caudas pesadas) é mais apropriada para este conjunto de dados do que a distribuição normal. Com a finalidade de identificar as observações aberrantes, a distância de Mahalanobis foi considerada. A Figura 8 (a) ilustra tal distância para o modelo t-Student ajustado e a linha de ponto de corte corresponde ao nonagésimo quinto percentil, $\upsilon = 7.361$, da forma quadrática d_i. Podemos observar que as observações 1 e 23 também são detectadas como *outliers* para o modelo t-Student ajustado. Entretanto, pela Figura 8 (b), note que os pesos dessas observações atípicas ($\widehat{u}_1 = 0.168$ e $\widehat{u}_{23} = 0.397$) são menores para o modelo ajustado com caudas pesadas, confirmando a robustez das estimativas de máxima verossimilhança do modelo t-Student contra observações atípicas. Para o modelo normal, temos que $u_i = 1, 2, ..., 35$.

Segundo Julious (2001) e Chen et al. (2011), em particular, é importante saber se, e quando, uma pessoa alterna as vias metabólicas de aeróbica para anaeróbica. Neste contexto, ajustamos, então, a este conjunto de dados, o modelo de regressão t-Student com ponto de mudança contínuo. Salientamos novamente que o ponto de mudança é, portanto, a mudança nas vias metabólicas. Nesta aplicação, para propósitos comparativos,

Figura 7 – Gráfico Q-Q juntamente com os envelopes simulados para modelo de regressão linear t-Student ajustado.



iremos também considerar o modelo de regressão normal com ponto de mudança contínuo. A Tabela 5 contém as estimativas de máxima verossimilhança para os parâmetros dos modelos normal e t-Student com ponto de mudança contínuo. Como sugerido por Lange et al. (1989), a função log-verossimilhança perfilada foi usada para a escolha de ν. Encontramos $\nu = 1.25$ para o modelo t-Student.

Tabela 5 – Estimativas de máxima verossimilhança para os dois modelos ajustados.

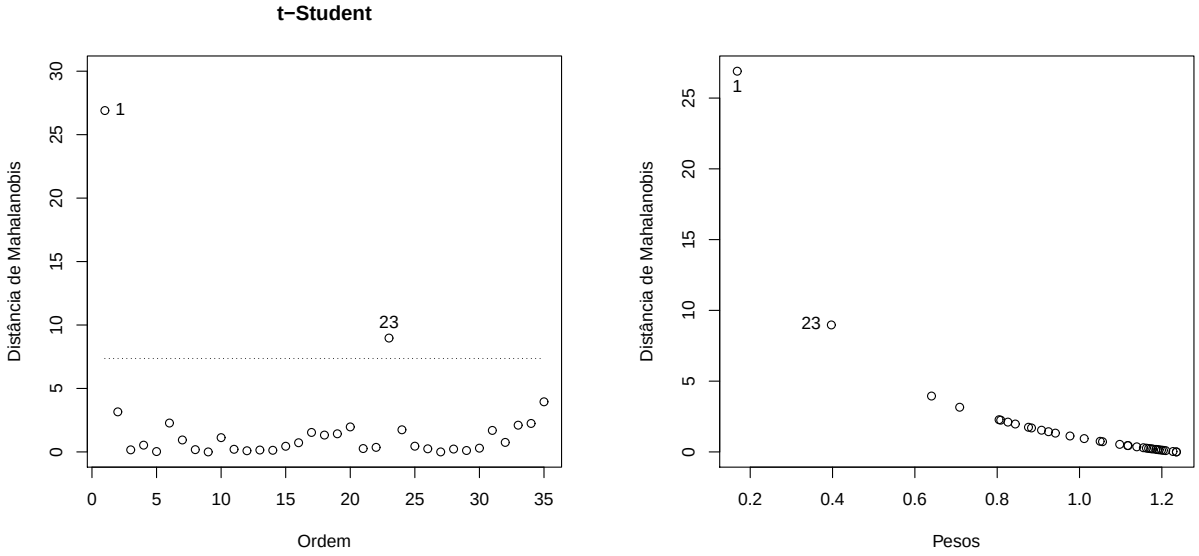
Parâmetros	β_0	β_1	β_2	σ^2	γ
Modelo Normal	0.076	0.042	0.044	0.011	39.463
Modelo t-Student	-0.351	0.055	0.046	0.001	47.368

Em seguida, comparando os modelos normal e t-Student com ponto de mudança contínuo através dos valores de SIC e AIC, exibidos na Tabela 6, notamos que o modelo t-Student apresenta um melhor ajuste que o modelo normal.

Tabela 6 – Alguns critérios de informação.

Modelos	$\ell(\widehat{\boldsymbol{\theta}})$	AIC	SIC	$\widehat{\gamma}$
Normal	29.058	-48.115	-40.339	39.46
t-Student	37.913	-63.826	-54.493	47.37

Figura 8 – (a) Gráfico da distância de Mahalanobis vs. ordem das observações. (b) Pesos vs. distância de Mahalanobis para o modelo t-Student ajustado.



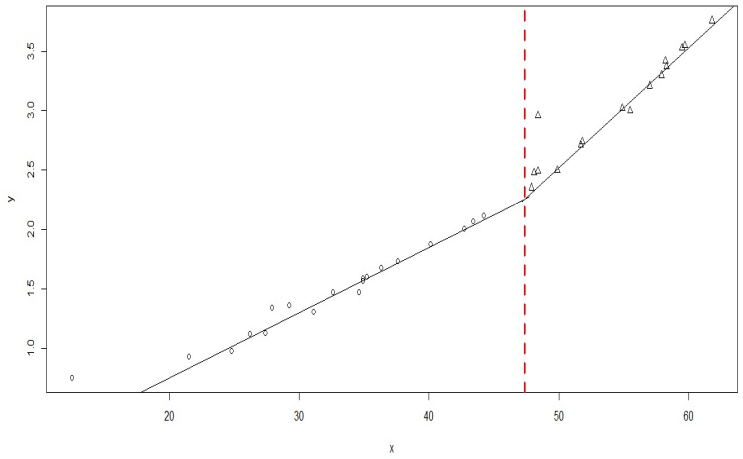
(a) (b)

Dessa forma, a Figura 9 mostra o gráfico de dispersão dos dados com o seguinte modelo t-Student ajustado, considerando ponto de mudança contínuo,

$$\widehat{y}_i = \begin{cases} -0.351 + 0.055x_i, & 12.5 \leq x_i \leq 47.37 \\ -2.530 + 0.101x_i, & 47.37 \leq x_i \leq 61.8. \end{cases} \tag{3.9}$$

Há, portanto, fortes evidências que sugerem que a relação linear entre as quantidades de dióxido de carbono exalada e a de oxigênio inalada muda uma vez que a quantidade de oxigênio excede cerca de 47.37 (litros por minuto). Segundo Julious (2001), isso pode ser devido ao fato de que, no início do exercício, durante a produção aeróbia de energia, o oxigênio é usado, mas à medida que o exercício se torna mais difícil, o requisito de energia do remador excede a quantidade que pode ser produzida através da via aeróbica sozinha. Neste ponto, o remador começa a usar a produção de energia anaeróbica e isso causa a mudança súbita na relação linear entre os volumes de dióxido de carbono e oxigênio.

Figura 9 – Gráfico de dispersão dos dados com o modelo t-Student ajustado com ponto de mudança contínuo em $\widehat{\gamma} = 47.37$.



4 CONCLUSÃO

Neste trabalho discutimos modelos de regressão linear com ponto de mudança, onde as observações seguem uma distribuição mistura de escala normal. Em particular, fixamos nossa atenção à distribuição t-Student. Os modelos baseados nesta classe de distribuições apresentam alternativas robustas do que modelos desenvolvidos sob normalidade. Para a estimação via máxima verossimilhança, o algoritmo EM foi proposto, explorando as propriedades estatísticas discutidas para a classe de distribuições mistura de escala normal. Adicionalmente, mostramos que os critérios de informação usuais, tais como o AIC e SIC, podem ser usados para detectar afastamentos da normalidade.

Os resultados obtidos foram aplicados em conjunto de dados reais e/ou simulados. Foi utilizado o programa estatístico R para a programação do procedimento de estimação 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 aos dados, tais que as estimativas de máxima verossimilhança dos parâmetros de interesse são obtidas via algoritmo EM. 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.

REFERÊNCIAS

[1] Andrews, D. F. & Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society*, B, 36, 99-102.

[2] Chen, J. & Gupta, A. K. (1996). Detecting changes of mean in multidimensional normal sequences with application to literature and geology. *Computational Statistics*, 11, 211-221.

[3] Chen, J. & Gupta, A. K. (1997). Testing and locating variance change points with application to stock prices. *Journal of the American Statistical Association*, 92, 739-747.

[4] Chen, J. & Gupta, A. K. (1999). Change point analysis of a gaussian model. *Statistical Papers*, 40, 323-333.

[5] Chen, J. & Gupta, A. K. (2001). On change point detection and estimation. *Communications in Statistics - Simulation and Computation*, 30(3), 665–697.

[6] Chen, J. & Gupta, A. K. (2003). Information-theoretic approach for detecting change in the parameters of a normal model. *Mathematical Methods of Statistics*, 12, 116-130.

[7] Chen, J. & Gupta, A. K. (2011). *P*arametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance. Springer Science & Business Media, New York.

[8] Chen, C. W. S, Chan, J. S. K., Gerlach R. & Hsieh, W. Y. L. (2011). A comparison of estimators for regression models with change points. *Statistics and Computing*, 21 (3), 395-414.

[9] Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Jornal of the Royal Statistical Society*, B, 39, 1-22.

[10] Fernandez, C. & Steel, M. (2000). Bayesian regression analysis whith Scale Mixture of Normals. *Econometric Theory*, 16, 80-101.

[11] Friedl, H. (1998). *C*omputer statistics. Lecture Notes, Graz University of Technology.

[12] Hinkley, D. V. (1969). Inference about the intersection in two-phase regression. *Biometrika*, 56(3), 495–504.

[13] Hofrichter, J. (2007). *C*hange Point Detection in Generalized Linear Models. Dissertation zur erlangung des akademischen grades doktor der technischen wissenschaften, Graz University of Technology.

[14] Julious, S. A. (2001). Inference and estimation in a changepoint regression problem. *Journal of the Royal Statistical Society*, D, 50(1), 51–61.

[15] Lange, K . L., Little, R. & Taylor, J. (1989). Robust Statistical modeling using t distribuition. *Journal of the American Statistical Association*, 84, 881-896.

[16] Lange, K. L. & Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics*, 2, 175-198.

[17] Liu, S., Wu, S. & Zidek, J. V. (1997). On segmented multivariate regression. *Statistica Sinica*, 7(2), 497–525.

[18] Lucas, A. (1997). Robustness of the Student t based M estimator. *Communications in Statistics: Theory and Methods*, 26, 1165-1182.

[19] Meng, X. L. & Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, 80, 267-278.

[20] Muggeo, V. M. R. (2003). Estimating regression models with unknown breakpoints. *Statistcs in Medicine*, 22(19), 3055-3071.

[21] Pinheiro, J. C., Liu, C. H. & 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.

[22] Rosa, G., Padovani, C. & Gianola, D. (2003). Robust Linear Mixed Models whith Normal/Independent Distribution and Bayesian MCMC Implementation. *Biometrical Journal*, 45, 573-590.

[23] Shewhart, W. A. (1939). Statistical Method from the Viewpoint of Quality Control. Dover, Washington.

[24] Sprent, P. (1961). Some hypotheses concerning two phase regression lines. *Biometrics*, 17(4), 634–645.

[25] Osorio, F. (2006). *Diagnóstico de Influência em Modelos Elípticos com Efeitos Mistos*. Tese de doutorado, Departamento de Estatística, IME-USP.

[26] Worsley, K. J. (1983). Testing for a two-phase multiple regression. *Technometrics*, 25(1), 35–42.

[27] Yamaguchi, K. (2001). Analysis of repeated measurements with multivariate t or contaminated multivariate normal errors. *Bulletin of the Computational Statistics of Japan*, 3, 1-18.

[28] Young, D. S. (2014). Mixtures of regressions with changepoints. *Statistics and Computing*, 24, 265-281.