

---

Universidade Federal da Paraíba  
Centro de Ciências Exatas e da Natureza  
Departamento de Estatística

Modelo de Regressão Robusto para Dados do Tipo  
Intervalo baseado no Kernel Guassiano e Seleção  
Automática do Hiper-parâmetro.

Ullysses da Nobrega Rosendo

Dezembro/2020

---

Ullysses da Nobrega Rosendo

**Modelo de Regressão Robusto para Dados do Tipo  
Intervalo baseado no Kernel Gaussiano e Seleção  
Automática do Hiper-parâmetro.**

Monografia apresentada ao Curso de Bacharelado em Estatística da Universidade Federal da Paraíba como requisito parcial para obtenção do Grau de Bacharel. Área de Concentração: Regressão.

João Pessoa  
Dezembro de 2020

**Catálogo na publicação**  
**Seção de Catalogação e Classificação**

R813m Rosendo, Ulysses da Nobrega.

Modelo de regressão robusto para dados do tipo-intervalo baseado no Kernel Gaussiano e seleção automática do hiper-parâmetro / Ulysses da Nobrega Rosendo. - João Pessoa, 2020.

42 f. : il.

Orientação: Eufrásio de Andrade Lima Neto.

TCC (Graduação/Bacharelado em Estatística) - UFPB/CCEN.

1. Regressão. 2. Funções Kernel. 3. Dado tipo-intervalo. I. Lima Neto, Eufrásio de Andrade. II. Título.

UFPB/CCEN

CDU 519.246.8(043.2)

*Dedico este trabalho a minha família e aos amigos  
que fiz durante minha trajetória acadêmica.*

## AGRADECIMENTOS

- Em primeiro lugar ao meu pai José Galberto e minha mãe Fabiana que sempre me deram um apoio incondicional em todas as minhas decisões, ao meus irmãos Guilherme e Willian pela união e parceria de sempre, ao meu tio José Damião que em muitos momentos foi como um segundo pai pra mim.
- Ao meu orientador, Prof. Dr. Eufrásio de Andrade Lima Neto, que esteve presente na minha trajetória da graduação me orientando no PIBIC, TCC I e TCC II, prestando uma espetacular orientação, depositando confiança e me incentivando sempre a ir mais longe.
- Ao Departamento de Estatística que conta com excelentes docentes.
- Aos examinadores deste Trabalho.
- Aos amigos que fiz durante o curso, em especial aos amigos Eduardo, Fernanda, Francielle, Giovani, Glezielle, Kelfanio, Kleber, Marina, Mateus, Milleny e Natan.
- Aos amigos que fiz na vida Adrielly, Kamila, Luizinho, Maria Alice, Nathalia, Paula, Rayanne e Thainah.

*Você não controla tudo o que acontece na sua vida  
mas a maneira como reage é sua responsabilidade,  
você é apenas um instante no tempo, um breve  
instante, apenas pó, veio do pó e logo voltará ao pó  
mas até lá seja, mas seja tudo o que puder ser.*

*- LetoDie (Memento Mori)*

A presença de outliers é muito comum em problemas de regressão e o uso de métodos robustos de regressão é fortemente recomendado, de tal forma que as observações discrepantes não afetem as estimativas dos parâmetros do modelo. Variáveis com valores de intervalo estão se tornando comuns em problemas de análise de dados, uma vez que esse tipo de dado representa a incerteza existente em uma medida ou a variabilidade natural presente nos dados.

Em relação à presença de outliers em conjuntos de dados com valores intervalares, diferente dos métodos já propostos na literatura, o método proposto neste trabalho engloba a estimação dos hiper-parâmetros e dos parâmetros do modelo no processo do algoritmo iterativo.

Este trabalho introduz um novo método robusto de regressão para variáveis do tipo intervalo que penaliza a presença de outliers nos pontos médios e/ou nas amplitudes das observações através do uso de funções kernel exponencial do tipo gaussiano, utilizando seleção automática do hiper-parâmetro, método denominado Modelo de Regressão Robusta com Kernel Exponencial do Tipo Gaussiano e Seleção Automática do Hiper-parâmetro para Variáveis do Tipo-intervalo (iETKRR-AS).

**Palavras-chaves:** Regressão, Funções Kernel, Dado tipo-intervalo, Estimador robusto, Hiper-parâmetro.

The presence of outliers is very common in regression problems and the use of robust regression methods is highly recommended, so that the poorly fitted observations do not affect the estimates of the model parameters. Variables with interval values are becoming common in data analysis problems, once this kind of data represents the incertity existing in a measure or the natural variability present in the data.

Regarding the presence of outliers in datasets with interval values, unlike previous proposed methods, the method proposed here includes the estimates of the hyper-parameters and the parameters of the model in the process of the iterative algorithm.

This paper introduces a new robust regression method for interval variables with penalizes the presence of outliers in the midpoints and/or in the range of the observations through the utilization of exponential-type gaussian kernel functions, using automatic selection of the hyper-parameter, witch is called Exponential-type Gaussian Kernel robust regression and Automatic Selection of the Hyper-parameters to Interval-type Variables (iETKRR-AS).

**Keywords:** Regression, Kernel functions, Interval-valued data, Robust estimator, Width hyper-parameter.

<b>Lista de Figuras</b>	<b>vii</b>
<b>Lista de Tabelas</b>	<b>viii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação . . . . .	1
1.2 Objetivos . . . . .	3
1.2.1 Gerais . . . . .	3
1.2.2 Específicos . . . . .	3
1.3 Organização do Trabalho . . . . .	4
<b>2 Referencial Teórico</b>	<b>5</b>
2.1 Variáveis com Valor do Tipo Intervalo . . . . .	5
2.2 Funções Kernel . . . . .	6
2.3 Estimadores $M$ . . . . .	8
2.4 Regressão Robusta para Dados do Tipo Intervalo . . . . .	9
2.4.1 Modelo IRR . . . . .	10
2.4.2 Modelo IQR . . . . .	11
2.4.3 Modelo SSLR . . . . .	13
2.4.4 Modelo iETKRR . . . . .	14
<b>3 Modelo iETKRR-AS</b>	<b>17</b>
3.1 Conceitos Preliminares . . . . .	17
3.2 Modelo iETKRR-AS . . . . .	18
3.3 <i>Redescending M-estimator</i> . . . . .	23
3.4 Algoritmo de Estimação dos Parâmetros do Modelo iETKRR-AS . . . . .	24

<b>4</b>	<b>Aplicações</b>	<b>26</b>
4.1	Conjunto de Dados Mushroom . . . . .	27
4.2	Conjunto de Dados Soccer . . . . .	28
4.3	Conjunto de Dados Unemployment . . . . .	29
4.4	Conjunto de Dados China Temperature . . . . .	30
4.5	Conjunto de Dados Cars . . . . .	31
4.6	Conjunto de Dados Scientific Production . . . . .	32
4.7	Resumo do desempenho do iETKRR-AS . . . . .	33
<b>5</b>	<b>Considerações Finais</b>	<b>35</b>
5.1	Conclusão . . . . .	35
	<b>Referências bibliográficas</b>	<b>37</b>
<b>A</b>	<b>Verificação Matemática</b>	<b>39</b>
A.1	Solução Equação 3.8 e por conseguinte Equação 3.9 . . . . .	39
A.2	Solução Equação 3.10 . . . . .	40
A.3	Solução Equação 3.12 e por conseguinte Equação 3.13 . . . . .	40
A.4	Prova Propriedades Redescending M-estimator . . . . .	41

## LISTA DE FIGURAS

2.1	Kernel mapeando pontos em uma dimensão superior. . . . .	7
2.2	Modelo Robusto vs Modelo Não-Robusto. . . . .	10
4.1	Conjunto de Dados Mushroom . . . . .	28
4.2	Conjunto de Dados Soccer . . . . .	29
4.3	Conjunto de Dados Unemployment . . . . .	30
4.4	Conjunto de Dados China Temperature . . . . .	31
4.5	Conjunto de Dados Cars . . . . .	32
4.6	Conjunto de Dados Scientific Production . . . . .	33

## LISTA DE TABELAS

2.1	Exemplo Conjunto de Dados do Tipo Intervalo . . . . .	5
2.2	Funções de Kernel . . . . .	7
2.3	Estimadores M . . . . .	9
4.1	Descrição dos conjuntos de dados reais com valor de intervalo . . . . .	27
4.2	Comparação entre os métodos robustos de intervalo com base na porcentagem de mudança (%) nas estimativas dos parâmetros após a eliminação dos outliers. Conjunto de dados Mushroom . . . . .	28
4.3	Comparação entre os métodos robustos de intervalo com base na porcentagem de mudança (%) nas estimativas dos parâmetros após a eliminação dos outliers. Conjunto de dados Soccer . . . . .	29
4.4	Comparação entre os métodos robustos de intervalo com base na porcentagem de mudança (%) nas estimativas dos parâmetros após a eliminação dos outliers. Conjunto de dados Unemployment . . . . .	30
4.5	Comparação entre os métodos robustos de intervalo com base na porcentagem de mudança (%) nas estimativas dos parâmetros após a eliminação dos outliers. Conjunto de dados China Temperature . . . . .	31
4.6	Comparação entre os métodos robustos de intervalo com base na porcentagem de mudança (%) nas estimativas dos parâmetros após a eliminação dos outliers. Conjunto de dados Cars . . . . .	32
4.7	Comparação entre os métodos robustos de intervalo com base na porcentagem de mudança (%) nas estimativas dos parâmetros após a eliminação dos outliers. Conjunto de dados Taobao . . . . .	33
4.8	Classificação média dos métodos nos conjuntos de dados em relação ao percentual mudança da estimação dos parâmetros após a eliminação dos outliers . . . . .	34

## 1.1 Motivação

Nos últimos anos assistimos a um crescente desenvolvimento tecnológico, com massivos conjuntos de dados de diversas áreas sendo gerados a todo momento, dados de dimensão e complexidade altas. Neste contexto dinâmico, surgem os chamados dados simbólicos. Os dados simbólicos apresentam, em sua estrutura, formas interessantes para se reduzir grandes bases de dados clássicos em novos conjuntos de dados simbólicos de tamanho reduzido, facilitando a análise de dados, e em alguns casos, sem nenhuma perda de informação. Um conjunto de dados simbólico pode conter, em uma célula de sua matriz, informações expressas por intervalos, células multi-valoradas, distribuições de frequência ou de probabilidade, etc, diferentemente de uma base de dados clássica onde cada célula assume apenas um único valor (NETO, 2008).

Em particular, em relação aos dados simbólicos do tipo intervalo, este pode representar a imprecisão e/ou incerteza existente em uma medição, mas também pode representar a variabilidade natural presente nos dados. Os dados tipo-intervalo surgem em situações como o registro de temperaturas diárias ou mensais em estações meteorológicas, preços de ações, etc. Outra fonte de dados tipo-intervalo é a agregação de grandes bases de dados em um número reduzido de grupos, cujas propriedades são descritas por variáveis de intervalo. Portanto, ferramentas para análise de dados tipo-intervalo são muito necessárias. Neste contexto, surge a necessidade de estender métodos clássicos para conjuntos de dados simbólicos. O estudo de Análise de Dados Simbólicos (Symbolic Data Analysis - SDA) é um paradigma que fornece estruturas para construir, descrever, analisar e extrair conhecimento de base de dados simbólicos. A SDA é um domínio da área de descoberta

de conhecimento e gerenciamento de dados relacionado à análise multivariada, reconhecimento de padrões e inteligência artificial. Em resumo, o objetivo da SDA é fornecer métodos adequados para gerenciar dados descritos por variáveis mais complexas (BOCK et al., 2000).

Diferentes abordagens foram introduzidas para analisar dados simbólicos. Em relação a Estatística Descritiva, Bertrand e Goupil (2000) estenderam alguns conceitos de estatística descritiva como medidas de tendência central e dispersão, Carvalho (1995) propôs histogramas para dados tipo-intervalo, Bock et al. (2000) estenderam conceitos de função de correlação e covariância para dados simbólicos. Cazes et al. (1997a) e Lauro e Palumbo (2000) introduziram métodos de análise de componentes principais considerando um conjunto de variáveis simbólicas de natureza intervalar. Ichino e Yaguchi (1994), Gowda e Diday (1991) e Gowda e Diday (1992) definiram medidas de similaridade e de dissimilaridade para mensurar distâncias entre objetos simbólicos para utilizar em métodos de análise de clusters. Bock (2002) propôs vários algoritmos de agrupamento para dados simbólicos do tipo-intervalo, Carvalho et al. (2006) propôs um algoritmo utilizando um critério de adequação baseado em distâncias adaptativas de Hausdorff. Em relação a análise fatorial simbólica, Cazes et al. (1997b) introduziu um método geométrico de classificação não-supervisionado em que as observações são descritas por intervalos. PARDOUX e Summa (2000) apresentaram uma abordagem simbólica para análise fatorial em tabelas com três entradas. Ciampi et al. (2000) generalizaram a abordagem do conceito de árvores de decisão para variáveis simbólicas. Maia, Carvalho e Ludermir (2006) são pioneiros em séries temporais para variáveis simbólicas intervalares, eles propuseram duas abordagens, a primeira considera modelos ARMA independentes para o centro e amplitude do intervalo. A segunda abordagem é considerada híbrida, pois combina um modelo ARMA e uma rede neural MLP.

Bock et al. (2000) propuseram o primeiro método de regressão para dados simbólicos levando em consideração variáveis simbólicas de natureza intervalar. Sua primeira abordagem consiste em ajustar um modelo de regressão linear aos pontos médios do intervalo, a segunda abordagem Billard e Diday (2002) consiste em estimar os limites inferior e superior dos intervalos. Neto e Carvalho (2008) propuseram o método que considera o centro e a amplitude do intervalo como uma melhoria em comparação com os métodos de regressão apresentados anteriormente.

É bem conhecido na literatura que um modelo de mínimos quadrados é sensível a outliers e que uso de técnicas não robustas normalmente leva a inferências enviesadas. Em relação aos dados simbólicos com valores de intervalo, algumas contribuições foram propostas relacionadas a modelos de regressão robustos. Domingues, Souza e Cysneiros (2010) apresentaram um modelo de regressão linear simétrico que leva em consideração

uma distribuição t de Student para os pontos médios do intervalo e uma distribuição normal para a amplitude. Fagundes, Souza e Cysneiros (2013) consideram duas regressões robustas independentes sobre os pontos médios e amplitudes dos intervalos. Os outliers são penalizados de acordo com a função de peso duplo de Tukey. Fagundes, Souza e Soares (2016) adaptaram a técnica de regressão quantílica para variáveis tipo-intervalo. Neto e Carvalho (2018) apresentaram um método de regressão robusta para variáveis tipo-intervalo baseado em funções kernel do tipo exponencial.

Este trabalho apresenta um novo método de regressão robusta para variáveis com valores de intervalos, denominado Modelo de Regressão Robusta com Kernel Exponencial do tipo gaussiano e Seleção Automática do Hiper-parâmetro para Variáveis do Tipo-intervalo (iETKRR-AS). Neste método, são ajustados dois modelos de regressão baseado em funções kernel do tipo gaussiano, uma para o centro e outra para a amplitude do intervalo. Os parâmetros do modelo são estimados com uma restrição delta, onde delta é dependente de hiper-parâmetros que são estimados automaticamente no algoritmo iterativo do método.

## 1.2 Objetivos

### 1.2.1 Gerais

- Propor um novo modelo de regressão robusta baseado em kernel exponencial do tipo gaussiano e seleção automática do hiper-parâmetro para variável do tipo-intervalo.

### 1.2.2 Específicos

- Utilizar o método de otimização denominado Método dos Multiplicadores de Lagrange para encontrar as formulações para estimação de hiper-parâmetro do kernel gaussiano;
- Utilizar o denominado Truque da Distancia do Kernel para obter a função objetivo do método;
- Minimizar a função objetivo do método para obter os parâmetros do modelo;
- Mostrar matematicamente que o método proposto pertence a classe de estimadores do tipo M conhecidos na literatura como estimadores de alta robustez;
- Apresentar um algoritmo iterativo para estimação dos parâmetros do modelo;

- Aplicar o método em conjuntos de dados reais com variáveis do tipo-intervalo;
- Comparar o método proposto com outros métodos robustos para variáveis tipo-intervalo presentes na literatura.

### 1.3 Organização do Trabalho

O trabalho está organizado da seguinte forma: o Capítulo 2 revisa conceitos sobre variáveis do tipo-intervalo, funções kernel, estimadores M e apresenta alguns modelos de regressão robusta para variáveis do tipo intervalo existentes na literatura. O Capítulo 3 apresenta o novo método iETKRR-AS para variáveis do tipo intervalo, bem como o algoritmo de estimação dos parâmetros. O Capítulo 4 traz aplicações a conjuntos de dados reais do tipo-intervalo e compara o novo método iETKRR-AS com os métodos de regressão robustos apresentados no Capítulo 2. O Capítulo 5 traz as conclusões sobre o trabalho.

## 2.1 Variáveis com Valor do Tipo Intervalo

Variáveis tipo-intervalo estão se tornando comuns em problemas de análise de dados. Uma variável tipo-intervalo  $Y$  é aquela que assume valores em um intervalo; ou seja,  $Y = \mathfrak{S} = [a, b] \in \mathbb{R}$ , com  $a \leq b$  e  $a, b \in \mathbb{R}$ . O intervalo pode ser fechado ou aberto em qualquer extremidade, ou seja,  $(a, b)$ ,  $[a, b)$ ,  $(a, b]$ , ou  $[a, b]$ (DIDAY; BILLARD, 2006).

Obs	Pulso	Pressão Sistólica	Pressão Diastólica
1	[58-90]	[118-173]	[63-102]
2	[47-68]	[104-161]	[71-118]
3	[32-114]	[131-186]	[58-113]
4	[50-80]	[95-166]	[58-100]
⋮	⋮	⋮	⋮
n	[37-90]	[85-146]	[45-98]

Tabela 2.1: Exemplo Conjunto de Dados do Tipo Intervalo

O tratamento estatístico de dados tipo-intervalo é considerado no contexto da Análise de Dados Simbólicos, campo de pesquisa relacionado à análise multivariada, reconhecimento de padrões e inteligência artificial, cujo objetivo é estender os métodos estatísticos clássicos para estruturas de dados mais complexas como listas, intervalos, histogramas, etc.

Hoje em dia os sistemas de informação permitem coletar e armazenar dados de forma rápida e com baixo custo. Além disso, conjuntos de dados massivos são gerados de maneira fácil em muitas áreas como: Economia, Meteorologia, Negócios, Telecomunicações, Biologia, entre outros. Esses conjuntos de dados podem ser estruturados de forma agregada, por exemplo, por razões de confidencialidade ou porque o interesse do estudo não é a unidade individual, mas um grupo de unidades. Assim, o pesquisador não analisa um conjunto de dados clássico com valores únicos na linha real, mas um conjunto de dados complexo formado por novos tipos de dados como, por exemplo, intervalos, que oferecem informações sobre o limite inferior e superior da variável de interesse. Além disso, a agregação de dados usuais numa estrutura de intervalos pode fornecer um conjunto de dados menor e mais gerenciável.

## 2.2 Funções Kernel

Os métodos kernel mapeiam dados em espaços dimensionais superiores com o objetivo de que nesse espaço de dimensão superior os dados possam se tornar mais facilmente separados ou melhor estruturados. O mapeamento deve ser não-linear para que classes não linearmente separáveis se tornem, em um espaço de mais alta dimensão, separáveis linearmente (HAYKIN, 1994). As funções kernel devem ser contínuas, simétricas e devem ter uma matriz positiva definida.

Seja  $X = \{x_1, \dots, x_n\}$  um conjunto não vazio onde  $x_i \in \mathbb{R}^p$ . Uma função  $\mathbb{R} : X \times X \rightarrow \mathbb{R}$  é chamada de função Kernel ou Kernel Mercer se e somente se:

1.  $\mathcal{K}(x_i, x_k) = \mathcal{K}(x_k, x_i)$ .
2.  $\sum_{i=1}^n \sum_{k=1}^n c_i c_k \mathcal{K}(x_i, x_k) \geq 0 \quad \forall n \geq 2$ , onde  $c_r \in \mathbb{R}, \forall r = 1, \dots, n$ .

Seja  $\Phi : \mathbf{X} \rightarrow \mathcal{F}$  um mapeamento não-linear do espaço de entrada  $\mathbf{X}$  para um espaço de alta dimensão  $\mathcal{F}$ . Ao aplicar o mapeamento  $\Phi$ , o produto escalar  $x_i^\top x_k$  no espaço de entrada é mapeado para  $\Phi(x_i)^\top \Phi(x_k)$  no espaço de alta dimensão. Um aspecto importante é que o mapeamento não-linear  $\Phi$  não precisa ser especificado porque cada Kernel Mercer pode ser expresso como  $\mathcal{K}(x_k, x_i) = \Phi(x_i)^\top \Phi(x_k)$  (MULLER et al., 2001).

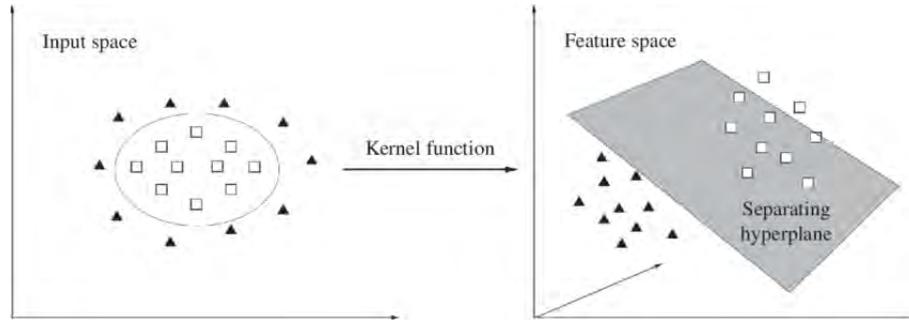


Figura 2.1: Kernel mapeando pontos em uma dimensão superior.

Um dos aspectos mais relevantes na aplicação do uso de funções kernel é que é possível calcular a distância euclidiana em  $\mathcal{F}$  sem saber a dimensão do mapeamento  $\Phi$ . Isso pode ser feito utilizando o chamado Truque da Distância do Kernel, dado por:

$$\begin{aligned}
 \|\phi(x_i) - \phi(x_k)\|^2 &= (\phi(x_i) - \phi(x_k))^\top (\phi(x_i) - \phi(x_k)) \\
 &= \Phi(x_i)^\top \Phi(x_i) - 2\Phi(x_i)^\top \Phi(x_k) + \Phi(x_k)^\top \Phi(x_k) \\
 &= \mathcal{K}(x_i, x_i) - 2\mathcal{K}(x_i, x_k) + \mathcal{K}(x_k, x_k).
 \end{aligned} \tag{2.1}$$

A Tabela 2.2 fornece algumas funções kernel que são frequentemente usadas na literatura.

Nome	Função Kernel
Kernel Linear	$\mathcal{K}(x_i, x_k) = x_i^\top x_k + c$
Kernel Polinomial	$\mathcal{K}(x_i, x_k) = (\alpha x_i^\top x_k + c)^2$
Kernel Gaussiano	$\mathcal{K}(x_i, x_k) = \exp\left(-\frac{(x_i - x_k)^2}{2\gamma^2}\right)$
Kernel Exponencial	$\mathcal{K}(x_i, x_k) = \exp\left(-\frac{(x_i - x_k)}{2\gamma^2}\right)$
Kernel Laplaciano	$\mathcal{K}(x_i, x_k) = \exp\left(-\frac{(x_i - x_k)}{\gamma}\right)$
Log-Kernel	$\mathcal{K}(x_i, x_k) = -\log(\ x_i - x_k\ ^d + 1)$
Kernel Cauchy	$\mathcal{K}(x_i, x_k) = \frac{1}{1 + \frac{\ x_i - x_k\ ^2}{\gamma^2}}$

Tabela 2.2: Funções de Kernel

## 2.3 Estimadores M

Os estimadores M podem ser vistos como uma generalização da estimação de máxima verosimilhança e foram introduzidos como estimadores robustos em problemas de regressão por Huber (1964). Assim, representam uma generalização do conhecido método dos mínimos quadrados, substituindo a função de perda quadrática por uma função simétrica  $\rho(\cdot)$ , dada por:

$$\text{Minimize}_{\beta} \sum_{i=1}^n \rho(\epsilon_i) \quad (2.2)$$

onde a função  $\rho$  fornece a contribuição de cada resíduo para a função objetivo e  $\epsilon$  o erro não-observado.

Embora não seja robusto com relação aos pontos de alavancagem, ele ainda é usado extensivamente na análise de dados para os quais se pode supor que a contaminação está principalmente na variável resposta. A função  $\rho$  é geralmente escolhida de tal forma que represente alguma ponderação do  $i$ -ésimo resíduo. Essa ponderação significa que observações periféricas têm seus pesos reduzidos e, portanto, as estimativas são menos afetadas por esse ruído. Uma ponderação de zero é equivalente à classificação como um outlier. (INSHA; QADIR; ALI, 2006)

A ponderação da observação é calculada com base na chamada função de influência, dada por:

$$\psi(x) = \frac{d\rho(x)}{dx}. \quad (2.3)$$

Agora definimos a função peso como:

$$w(x) = \frac{\psi(x)}{x}. \quad (2.4)$$

Uma função  $\rho$  para ser considerada um M-estimator deve possuir as seguintes propriedades:

1.  $\rho(0) = 0$ .
2.  $\rho(x_i) \geq 0$ .
3.  $\rho(x_i) = \rho(-x_i)$ .
4.  $\rho(x_i) \leq \rho(x_j)$  para  $|x_i| \leq |x_j|$ .
5.  $\rho$  é contínua e diferenciável.

Outra classe importante de estimadores são os Redescending M-estimators. Esta classe de estimadores possuem funções que não diminuem perto da origem, mas diminuem em direção a 0 longe da origem, possuem altos pontos de ruptura (próximo a 0.5) e são ligeiramente mais eficientes que o estimador Huber. Redescending M-estimators devem assumir todas as propriedades de um Estimador M e também satisfazer que  $\lim_{x_i \rightarrow \infty} \Psi(x_i) = 0$  onde  $\Psi = \rho'$ . A Tabela 2.3 fornece alguns estimadores M que são frequentemente utilizados na literatura.

Estimadores M	$\rho(x)$	$\psi(x)$	$w(x)$
Tukey $\begin{cases}  x  \leq c \\  x  > c \end{cases}$	$\begin{cases} \frac{c^2}{6}(1 - [1 - (x/c)^2]^3) \\ \frac{c^2}{6} \end{cases}$	$\begin{cases} x[1 - (x/c)^2]^2 \\ 0 \end{cases}$	$\begin{cases} [1 - (x/c)^2]^2 \\ 0 \end{cases}$
Huber $\begin{cases}  x  \leq k \\  x  > k \end{cases}$	$\begin{cases} \frac{x^2}{2} \\ k( x  - \frac{k}{2}) \end{cases}$	$\begin{cases} x \\ k \operatorname{sgn}(x) \end{cases}$	$\begin{cases} 1 \\ \frac{k}{ x } \end{cases}$
Cauchy	$\frac{c^2}{2} \log(1 + (x/c)^2)$	$\frac{x}{1+(x/c)^2}$	$\frac{1}{1+(x/c)^2}$

Tabela 2.3: Estimadores M

## 2.4 Regressão Robusta para Dados do Tipo Intervalo

A regressão é uma das técnicas estatísticas mais utilizadas. O método dos mínimos quadrados ordinários (OLS) é geralmente adotado devido à facilidade computacional e à ausência de pressupostos probabilísticos para obter os estimadores dos parâmetros da reta de regressão. No entanto, a estimativa OLS é afetada pela presença de outliers, não normalidade, multicolinearidade e dados ausentes. O principal objetivo da regressão robusta é fornecer estimadores resistentes à presença de outliers.

A análise de regressão robusta oferece uma alternativa ao OLS quando os pressupostos fundamentais não são cumpridos pela natureza dos dados. As propriedades de eficiência, ponto de ruptura e pontos de alta alavancagem são usados para definir o desempenho de técnicas robustas em um sentido teórico (ÖZLEM, 2011).

Um modelo de regressão robusta deve apresentar boas estimativas de parâmetros em dados com outliers ou em dados sem outliers, ou seja, os parâmetros da reta de regressão deve apresentar pequena variação quando ajustados com presença e ausência de pontos atípicos. Porém, nem sempre os outliers vão ser linearmente separáveis. Dessa necessidade que surge o uso de funções kernel e de estimadores M para ponderar as observações com o intuito de preservar as estimativas do modelo em dados contaminados com outliers.

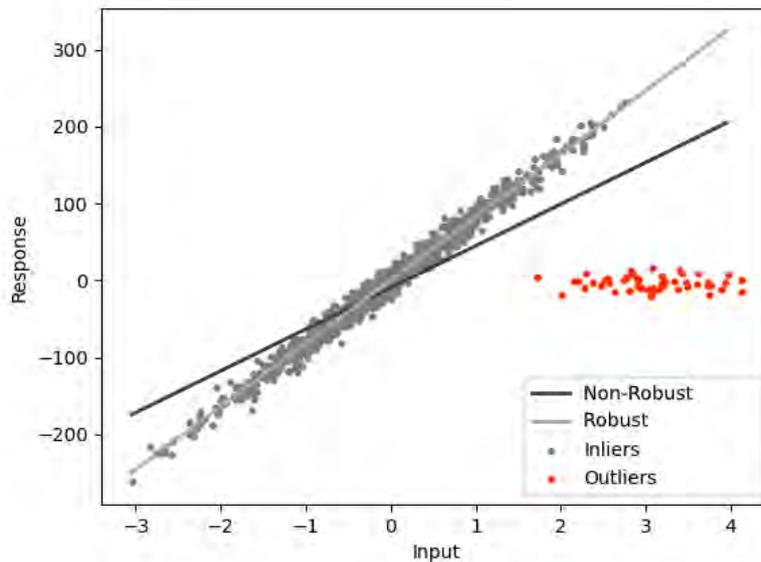


Figura 2.2: Modelo Robusto vs Modelo Não-Robusto.

Com relação a dados do tipo-intervalo, algumas contribuições foram propostas relacionadas a modelos robustos de regressão. Domingues, Souza e Cysneiros (2010) apresentaram um modelo de regressão linear simétrico simbólico para variáveis tipo-intervalo que leva em consideração uma distribuição t de Student para os pontos médios do intervalo e uma distribuição normal para a amplitude. Fagundes, Souza e Cysneiros (2013) considera duas regressões robustas independentes sobre os pontos médios e amplitudes dos intervalos. Os outliers são penalizados de acordo com a função de peso duplo de Tukey. Fagundes, Souza e Soares (2016) adaptou a técnica de regressão quantílica para variáveis tipo-intervalo. Neto e Carvalho (2018) apresentaram um método de regressão robusta para variáveis tipo-intervalo baseado em funções kernel do tipo exponencial que apresentou um desempenho competitivo (ou melhor) em comparação aos métodos supra-citados.

### 2.4.1 Modelo IRR

Proposto por Fagundes, Souza e Cysneiros (2013), considere  $\beta^c = (\beta_0^c, \beta_1^c, \dots, \beta_p^c)$  e  $\beta^r = (\beta_0^r, \beta_1^r, \dots, \beta_p^r)$  como sendo dois vetores com  $p + 1$  parâmetros e  $\epsilon^c = (\epsilon_1^c, \epsilon_2^c, \dots, \epsilon_n^c)$  e  $\epsilon^r = (\epsilon_1^r, \epsilon_2^r, \dots, \epsilon_n^r)$  como sendo dois vetores de erros desconhecidos para centro e a amplitude, respectivamente. Considere também duas equações de regressão linear, para o centro e a amplitude, respectivamente:

$$\begin{aligned} y_i^c &= x_i^{c\top} \beta^c + \epsilon_i^c. \\ y_i^r &= x_i^{r\top} \beta^r + \epsilon_i^r. \end{aligned} \tag{2.5}$$

Os vetores  $\beta^c = (\beta_0^c, \beta_1^c, \dots, \beta_p^c)$  e  $\beta^r = (\beta_0^r, \beta_1^r, \dots, \beta_p^r)$  são estimados minimizando uma função objetivo com base em  $\rho$  para ambos os erros  $\epsilon_i^c$  e  $\epsilon_i^r$ , dada por:

$$S = \sum_{i=1}^n \rho \left( \frac{y_i^c - x_i^{c\top} \beta^c}{s} \right) + \rho \left( \frac{y_i^r - x_i^{r\top} \beta^r}{s} \right), \quad (2.6)$$

onde  $s$  é um estimador robusto de escala e  $\rho$  é uma função particular. Uma escolha popular para  $s$  é o desvio absoluto mediano, dado por:

$$s = \frac{\text{mediana}(|\epsilon_i - \text{mediana}(\epsilon_i)|)}{0.6745}. \quad (2.7)$$

Há uma série de funções robustas populares para  $\rho$ , por exemplo, a função biweight de Tukey  $\rho(x)$  apresentada abaixo. O modelo de regressão robusta pode ser classificado pela função  $\psi(x)$  que controla o peso de cada resíduo. A função  $\psi(x)$  é a derivada primeira de  $\rho(x)$ , logo,  $\psi(x) = \rho(x)'$ .

$$\rho(x) = \begin{cases} \frac{c^2}{6} (1 - [1 - (x/c)^2]^3) & , |x| \leq c \\ \frac{c^2}{6} & , |x| > c \end{cases} \quad (2.8)$$

$$\psi(x) = \begin{cases} x[1 - (x/c)^2]^2 & , |x| \leq c \\ 0 & , |x| > c \end{cases} \quad (2.9)$$

A função (2.6) produz dois problemas de minimização independentes:

- Encontrar  $\hat{\beta}^c$ , o valor que minimiza  $\beta^c$
- Encontrar  $\hat{\beta}^r$ , o valor que minimiza  $\beta^r$

Expressões para  $\hat{\beta}^c$  e  $\hat{\beta}^r$  não podem ser obtidas de forma fechada. Portanto, alguns métodos iterativos devem ser utilizados, como Newton Raphson, BFGS (Broyden Fletcher Goldfarb Shanno), método de pontuação de Fisher, mínimos quadrados ponderados iterativo, entre outros.

## 2.4.2 Modelo IQR

A regressão quantílica ( $QR$ ) é um método de estimação motivado pelo interesse em estudar o comportamento de indivíduos "não-médios". Basicamente, a regressão quantílica estima várias retas para diferentes quantis associados. Proposto por Fagundes, Souza

e Soares (2016), a Regressão quantílica intervalar (IQR) condicionada em  $\tau$  pode ser expressa por:

$$\mathcal{Q}^\tau(Y/\mathbf{X}) = [\xi_c, \xi_r] \quad (2.10)$$

onde

$$\begin{aligned} \xi_c &= \mathcal{Q}^\tau(Y^c/\mathbf{X}^c) - \mathcal{Q}^\tau(\mathbf{Y}^r/\mathbf{X}^r) \\ \xi_r &= \mathcal{Q}^\tau(Y^c/\mathbf{X}^c) + \mathcal{Q}^\tau(\mathbf{Y}^r/\mathbf{X}^r) \end{aligned} \quad (2.11)$$

onde  $\mathcal{Q}^\tau(Y^c/\mathbf{X}^c)$  e  $\mathcal{Q}^\tau(\mathbf{Y}^r/\mathbf{X}^r)$  são regressões quantílicas para o  $\tau$ -ésimo quantil associadas ao centro e amplitude, respectivamente. Considere  $\beta^{c\tau} = (\beta_0^{c\tau}, \beta_1^{c\tau}, \dots, \beta_p^{c\tau})^\top$  e  $\beta^{r\tau} = (\beta_0^{r\tau}, \beta_1^{r\tau}, \dots, \beta_p^{r\tau})^\top$  como sendo dois vetores de  $p + 1$  parâmetros, considere  $\epsilon^{c\tau} = (\epsilon_1^{c\tau}, \epsilon_2^{c\tau}, \dots, \epsilon_n^{c\tau})^\top$  e  $\epsilon^{r\tau} = (\epsilon_1^{r\tau}, \epsilon_2^{r\tau}, \dots, \epsilon_n^{r\tau})^\top$  como sendo dois vetores de  $n$  erros desconhecidos no centro e na amplitude. Considere também duas equações de regressão  $\tau$ -ésima quantílica, para o centro e a amplitude, respectivamente:

$$\begin{aligned} y_i^c &= x_i^{c\top} \beta^{c\tau} + \epsilon_i^{c\tau} . \\ y_i^r &= x_i^{r\top} \beta^{r\tau} + \epsilon_i^{r\tau} . \end{aligned} \quad (2.12)$$

Dado que  $\beta^{c\tau}$  e  $\beta^{r\tau}$  são vetores constantes e QR é formulada de forma equivalente a uma regressão linear, QR também pressupõe que o  $\tau$ -ésimo do termo de erros seja zero. Logo, as equações do  $\tau$ -ésimo quantil para centro e amplitude podem ser rescritas como:

$$\mathcal{Q}^\tau(Y^c/\mathbf{X}^c) = \mathbf{x}_i^{c\top} \beta^{c\tau} . \quad (2.13)$$

$$\mathcal{Q}^\tau(\mathbf{Y}^r/\mathbf{X}^r) = \mathbf{x}_i^{r\top} \beta^{r\tau} . \quad (2.14)$$

Na regressão quantílica para dados clássicos, o  $\tau$ -ésimo quantil é o valor  $Q(\tau)$  que minimiza a soma das distâncias ponderadas de uma amostra de pontos, onde os pontos  $\mu_1, \dots, \mu_n$  abaixo de  $Q(\tau)$  recebem um peso de  $1 - \tau$  e pontos acima de  $Q(\tau)$  recebem um peso de  $\tau$ , ou seja, a penalização da função perda terá magnitudes diferentes para erros "para cima" e "para baixo", a depender do quantil escolhido.

Portanto, no modelo IQR, os estimadores de regressão quantílica são escolhidos para minimizar:

$$\sum_{i=1}^n d(y_i^c, Q(y_i^c/\mathbf{x}_i^c)) + \mathbf{d}(\mathbf{y}_i^r, \mathbf{Q}(\mathbf{y}_i^r/\mathbf{x}_i^r)) \quad (2.15)$$

onde:

$$d(y_i^c, Q(y_i^c/\mathbf{x}_i^c)) = \tau \sum_{\mathbf{y}_i^c \geq \mathbf{x}_i^{c\prime} \beta_c^\tau} |\mathbf{y}_i^c - \mathbf{Q}(\mathbf{y}_i^c/\mathbf{x}_i^c)| + (1 - \tau) \sum_{\mathbf{y}_i^c \leq \mathbf{x}_i^{c\prime} \beta_c^\tau} |\mathbf{y}_i^c - \mathbf{Q}(\mathbf{y}_i^c/\mathbf{x}_i^c)| \quad (2.16)$$

$$d(y_i^r, Q(y_i^r/\mathbf{x}_i^r)) = \tau \sum_{\mathbf{y}_i^r \geq \mathbf{x}_i^{r\prime} \beta_r^\tau} |\mathbf{y}_i^r - \mathbf{Q}(\mathbf{y}_i^r/\mathbf{x}_i^r)| + (1 - \tau) \sum_{\mathbf{y}_i^r \leq \mathbf{x}_i^{r\prime} \beta_r^\tau} |\mathbf{y}_i^r - \mathbf{Q}(\mathbf{y}_i^r/\mathbf{x}_i^r)| \quad (2.17)$$

Assim, temos que:

- A proporção de pontos ajustados acima da linha ajustada  $Q(y_i^c/\mathbf{x}_i^c)$  é  $\tau$  e abaixo é  $1 - \tau$ .
- A proporção de pontos ajustados acima da linha ajustada  $Q(y_i^r/\mathbf{x}_i^r)$  é  $\tau$  e abaixo é  $1 - \tau$ .

O calculo do  $\tau$ -esimo quantil que minimiza a função objetivo para estimação de  $\beta^{c^\tau}$  e  $\beta^{r^\tau}$  pode ser formulado como um problema de programação e resolvido de forma eficiente com o algoritmo simplex.

### 2.4.3 Modelo SSLR

O modelo SSLR proposto por Domingues, Souza e Cysneiros (2010), é baseado nos modelos de regressão linear simétricos. A previsão dos limites de um intervalo é realizada por meio de uma combinação de previsões do centro e amplitude dos valores do intervalo.

Suponha  $Y_1, \dots, Y_n$  como  $n$  variáveis aleatórias com suporte em  $\mathbb{R}$ . Dizemos que  $Y_i$  tem distribuição simétrica com parâmetro de locação  $\mu \in \mathbb{R}$  e de escala  $\phi > 0$  se sua função de densidade é caracterizada por:

$$f_{Y_i}(y) = \frac{1}{\sqrt{\phi}} g \left\{ \frac{(y - \mu)^2}{\phi} \right\}. \quad (2.18)$$

A função  $g : \mathbb{R} \rightarrow [0, \infty)$  tal que  $\int_0^\infty g(\mu) d\mu < \infty$  é conhecido como gerador de densidade e é denotado por  $Y_i \sim S(\mu_i, \phi, g)$ . O modelo SLR é definido como:

$$Y_i = \mu_i + \epsilon_i, \quad (2.19)$$

onde,  $\mu_i = \mathbf{x}_i^T \beta$ ,  $\beta = (\beta_0, \dots, \beta_p)^T$  é um vetor de parâmetros desconhecido,  $x_i$  é um vetor de variáveis explicativas e  $\epsilon_i \sim S(\mu_i, \phi, g)$ ,  $i = 1, \dots, n$ . Assim, temos que  $E(Y_i) = \mu$  e  $Var(Y_i) = \xi\phi$ , onde  $\xi$  é uma constante maior que zero que depende da distribuição escolhida de  $g$ . Esta classe de modelos inclui todas as distribuições contínuas simétricas, como normal, t de Student, logística, entre outras.

Para o ajuste do modelo SSLR é considerado dois modelos independentes para centro e amplitude, respectivamente. O modelo para o centro segue a distribuição t de Student para os erros e o modelo para a amplitude segue distribuição normal para os erros.

Seja  $Y^c = (y_1^c, \dots, y_n^c)$  e  $Y^r = (y_1^r, \dots, y_n^r)$  o centro e a amplitude, respectivamente, de uma variável resposta tipo-intervalo. Seja  $X_j^c = (x_{j1}^c, \dots, x_{jn}^c)$  e  $X_j^r = (x_{j1}^r, \dots, x_{jn}^r)$  ( $j = 1, \dots, p$ ) um conjunto de variáveis independentes representando centro e a amplitude dos intervalos, respectivamente.

O modelo SSLR é definido de acordo com dois modelos de regressão lineares simétricos independentes, dados por:

$$\begin{aligned} y_i^c &= x_i^{c\top} \beta^c + \epsilon_i^c \\ y_i^r &= x_i^{r\top} \beta^r + \epsilon_i^r \end{aligned} \quad (2.20)$$

onde  $\beta^c = (\beta_0^c, \beta_1^c, \dots, \beta_p^c)$  é um vetor com  $p + 1$  parâmetros desconhecidos para o centro do intervalo  $\beta^r = (\beta_0^r, \beta_1^r, \dots, \beta_p^r)$  é um vetor com  $p + 1$  parâmetros desconhecidos para a amplitude do intervalo. Os vetores de parâmetros  $\beta^c$  e  $\beta^r$  são estimados pelo método de máxima verossimilhança assumindo distribuição t de Student para os erros dos pontos médios e distribuição normal para os erros da amplitude, respectivamente.

#### 2.4.4 Modelo iETKRR

Proposto por Neto e Carvalho (2018), o modelo iETKRR pondera as observações com base na função kernel do tipo exponencial, de forma que o peso atribuído a uma observação tipo-intervalo atípica (outlier) seja o menor possível. Além disso, os parâmetros do modelo são estimados considerando um processo iterativo de modo a minimizar uma função objetivo.

O modelo iETKRR considera a variável resposta tipo intervalo  $\mathbf{Y} = [Y_1, Y_2]$  e a estimação dos parâmetros é guiada pela minimização de uma função objetiva levando em conta o truque da distancia de kernel (MULLER et al., 2001). Desta forma, a soma dos erros

quadrados é definida não no espaço original, mas em um espaço de maior dimensão  $\mathcal{F}$  através de um mapeamento não-linear  $\Phi$  aplicado, respectivamente, na variável resposta observada  $\mathbf{y}_i = (y_{1i}, y_{2i})$  e a sua média correspondente  $\boldsymbol{\mu}_i = (\mu_{1i}, \mu_{2i})$ .

Com base no truque da distância do kernel, a soma dos erros quadrados definidos no espaço original é computada em um espaço alta dimensão como segue:

$$\begin{aligned} S &= \sum_{i=1}^n (\|\Phi_1(y_{1i}) - \Phi_1(\mu_{1i})\|^2 + \|\Phi_2(y_{2i}) - \Phi_2(\mu_{2i})\|^2) = \\ &= \sum_{i=1}^n [\mathcal{K}_1(y_{1i}, y_{1i}) - 2\mathcal{K}_1(y_{1i}, \mu_{1i}) + \mathcal{K}_1(\mu_{1i}, \mu_{1i}) + \mathcal{K}_2(y_{2i}, y_{2i}) - 2\mathcal{K}_2(y_{2i}, \mu_{2i}) + \mathcal{K}_2(\mu_{2i}, \mu_{2i})]. \end{aligned} \quad (2.21)$$

Seja  $\mathcal{K}_G$  a função do Kernel Gaussiano, as seguintes propriedades são verificadas:  $\mathcal{K}_G(\mathbf{y}_i, \mathbf{y}_i) = \mathcal{K}_G(\boldsymbol{\mu}_i, \boldsymbol{\mu}_i) = 1$  ( $1 \leq i \leq n$ ) e a função objetivo (2.21) pode ser reescrita como:

$$\begin{aligned} S &= \sum_{i=1}^n 2[2 - (\mathcal{K}_{G_1}(y_{1i}, \mu_{1i}) + \mathcal{K}_{G_2}(y_{2i}, \mu_{2i}))] \\ S &= \sum_{i=1}^n 2 \left[ 2 - \left[ \exp\left(-\frac{1}{2} \left(\frac{y_1 - \mu_1}{\gamma_1}\right)^2\right) + \exp\left(-\frac{1}{2} \left(\frac{y_2 - \mu_2}{\gamma_2}\right)^2\right) \right] \right] \end{aligned} \quad (2.22)$$

Assim, quanto mais próximo os termos  $y_{1i}$  e  $\mu_{1i}$  ( $y_{2i}$  e  $\mu_{2i}$ ),  $\forall i = 1, \dots, n$ , tão perto do mínimo se torna a função objetiva para o modelo iETKRR. Diferenciando a função (2.22) em relação a cada elemento dos parâmetros  $\beta_1$  e  $\beta_2$ , obtemos, respectivamente, o conjunto de equações normais para o modelo, dado por:

$$\begin{aligned} \frac{\partial S}{\partial \beta_{10}} &= -2 \sum_{i=1}^n \mathcal{K}_{G_1}(y_{1i}, \hat{\mu}_{1i}) \frac{(y_{1i} - \hat{\mu}_{1i})}{\gamma_1} = 0, \\ \frac{\partial S}{\partial \beta_{1j}} &= -2 \sum_{i=1}^n \mathcal{K}_{G_1}(y_{1i}, \hat{\mu}_{1i}) \frac{x_{1ij}(y_{1i} - \hat{\mu}_{1i})}{\gamma_1} = 0, \\ \frac{\partial S}{\partial \beta_{20}} &= -2 \sum_{i=1}^n \mathcal{K}_{G_2}(y_{2i}, \hat{\mu}_{2i}) \frac{(y_{2i} - \hat{\mu}_{2i})}{\gamma_2} = 0, \\ \frac{\partial S}{\partial \beta_{2j}} &= -2 \sum_{i=1}^n \mathcal{K}_{G_2}(y_{2i}, \hat{\mu}_{2i}) \frac{x_{2ij}(y_{2i} - \hat{\mu}_{2i})}{\gamma_2} = 0 \end{aligned} \quad (2.23)$$

A solução do sistema de equações (3.16) é obtido através de mínimos quadrados com

reponderação iterativa, com base no modelo de regressão:

$$\mathbf{y}^* = \mathbf{X}^* \boldsymbol{\beta} + \boldsymbol{\epsilon}^*,$$

onde

$$\mathbf{y}^* = \begin{bmatrix} \mathbf{y}_1^* \\ \mathbf{y}_2^* \end{bmatrix}, \quad \mathbf{X}^* = \begin{bmatrix} \mathbf{X}_1^* & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2^* \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}, \quad \text{e } \boldsymbol{\epsilon}^* = \begin{bmatrix} \boldsymbol{\epsilon}_1^* \\ \boldsymbol{\epsilon}_2^* \end{bmatrix},$$

onde

- $\mathbf{y}_1^* = \mathbf{K}^{1/2} \mathbf{y}_1$  e  $\mathbf{y}_2^* = \mathbf{K}^{1/2} \mathbf{y}_2$  sendo vetores ponderados  $n \times 1$  das variáveis resposta  $Y_1$  e  $Y_2$ , respectivamente;
- $\mathbf{X}_1^* = \mathbf{K}^{1/2} \mathbf{X}_1$  e  $\mathbf{X}_2^* = \mathbf{K}^{1/2} \mathbf{X}_2$  sendo matrizes ponderadas  $n \times p$  das variáveis explicativas  $X_1$  e  $X_2$ , respectivamente;
- $\boldsymbol{\beta}_1$  e  $\boldsymbol{\beta}_2$  sendo os respectivos vetores de parâmetros da regressão;
- $\boldsymbol{\epsilon}_1^* = \mathbf{K}^{1/2} \boldsymbol{\epsilon}_1$  e  $\boldsymbol{\epsilon}_2^* = \mathbf{K}^{1/2} \boldsymbol{\epsilon}_2$  sendo vetores de erros independentes ponderados  $n \times 1$ ;
- $\mathbf{K}_1 = \text{diag}(k_{11}, k_{12}, \dots, k_{1n})$  e  $\mathbf{K}_2 = \text{diag}(k_{21}, k_{22}, \dots, k_{2n})$  sendo matrizes diagonal  $n \times n$  de pesos do kernel  $k_{1i}$  definido como  $k_{1i} = \mathcal{K}(y_{1i}, \mu_{1i})$  e  $k_{2i} = \mathcal{K}(y_{2i}, \mu_{2i})$

### 3.1 Conceitos Preliminares

Seja  $E = \{e_1, \dots, e_n\}$  um conjunto de exemplos que são descritos por  $p + 1$  variáveis do tipo intervalo  $Z, W_1, \dots, W_p$ . A variável com valor de intervalo  $Z$  é uma variável dependente e está relacionada a um conjunto de variáveis com valor do tipo intervalo  $W_j$  ( $j = 1, 2, \dots, p$ ), conhecidas como variáveis independentes. Cada exemplo  $e_i \in E$  ( $1 \leq i \leq n$ ) é representado por um vetor com valor de intervalo  $(\mathbf{w}_i, z_i)$ , com  $\mathbf{w}_i = (w_{i1}, \dots, w_{ip})$ , onde  $w_{ij} = [a_{ij}, b_{ij}] \in \mathfrak{S} = \{[a, b] : a, b \in \mathfrak{R}, a \leq b\}$  ( $1 \leq j \leq p$ ) e  $z_i = [z_{Li}, z_{Ui}] \in \mathfrak{S}$  são os valores observados de  $W_j$  e  $Z$ , respectivamente.

Agora, seja  $Y_1, X_{1j}$  e  $Y_2, X_{2j}$  ( $j = 1, 2, \dots, p$ ) variáveis quantitativas que representam os limites inferior e superior ou os pontos médios e a amplitude dos intervalos definidos pelas variáveis com valor de intervalo  $Z$  e  $W_j$  respectivamente. No caso em que as variáveis quantitativas,  $Y_1, X_{1j}$  e  $Y_2, X_{2j}$  ( $j = 1, 2, \dots, p$ ) representam, respectivamente, os limites inferior e superior das variáveis de intervalo  $Z$  e  $W_j$ , nós denotamos  $Y_1 = Y^L, X_{1j} = X_j^L, Y_2 = Y^U, X_{2j} = X_j^U$ , e cada exemplo  $e_i \in E$  ( $i = 1, \dots, n$ ) será representado por dois vetores:  $(\mathbf{x}_i^L, y_i^L)$  e  $(\mathbf{x}_i^U, y_i^U)$ , com  $\mathbf{x}_i^L = (x_{i1}^L, \dots, x_{ip}^L)$  e  $\mathbf{x}_i^U = (x_{i1}^U, \dots, x_{ip}^U)$ , onde  $x_{ij}^L = a_{ij}, x_{ij}^U = b_{ij}, y_i^L = z_{Li}$  e  $y_i^U = z_{Ui}$  são os valores observados das variáveis quantitativas  $X_j^L, X_j^U, Y^L$  e  $Y^U$ , respectivamente.

Da mesma forma, para o caso em que as variáveis quantitativas,  $Y_1, X_{1j}$  e  $Y_2, X_{2j}$  ( $1 \leq j \leq p$ ) representam, respectivamente, o ponto médio e amplitude do intervalo das variáveis de intervalo  $Z$  e  $W_j$  nós denotamos  $Y_1 = Y^c, X_{1j} = X_j^c, Y_2 = Y^r, X_{2j} = X_j^r$ ,

e cada exemplo  $e_i \in E$  ( $1 \leq i \leq n$ ) será representado pelos vetores  $(\mathbf{x}_i^c, y_i^c)$  e  $(\mathbf{x}_i^r, y_i^r)$ , com  $\mathbf{x}_i^c = (x_{i1}^c, \dots, x_{ip}^c)$  e  $\mathbf{x}_i^r = (x_{i1}^r, \dots, x_{ip}^r)$ , onde  $x_{ij}^c = (a_{ij} + b_{ij})/2$ ,  $x_{ij}^r = (b_{ij} - a_{ij})/2$ ,  $y_i^c = (z_{Li} + z_{Ui})/2$  e  $y_i^r = (z_{Ui} - z_{Li})/2$  são os valores observados das variáveis  $X_j^c$ ,  $X_j^r$ ,  $Y^c$  e  $Y^r$ , respectivamente.

## 3.2 Modelo iETKRR-AS

Seja o vetor da variável resposta  $\mathbf{y}_i = (y_{1i}, y_{2i})$  e seu correspondente vetor médio  $\boldsymbol{\mu}_i = (\mu_{1i}, \mu_{2i})$ , onde os subscritos 1 e 2 representam o limite inferior (ou o ponto médio) e o limite superior (ou a amplitude) dos intervalos, respectivamente, com  $\mu_{1i} = \mathbf{x}_{1i}^\top \boldsymbol{\beta}_1 = \beta_{10} + \sum_{j=1}^p \beta_{1j} x_{1ij}$  e  $\mu_{2i} = \mathbf{x}_{2i}^\top \boldsymbol{\beta}_2 = \beta_{20} + \sum_{j=1}^p \beta_{2j} x_{2ij}$  representando a média das variáveis de resposta  $Y_1$  e  $Y_2$ , respectivamente, com  $\mathbf{x}_{1i} = (1, x_{1i1}, \dots, x_{1ip})$  e  $\mathbf{x}_{2i} = (1, x_{2i1}, \dots, x_{2ip})$  relacionados aos valores do conjunto das variáveis explicativas  $X_{1j}$  e  $X_{2j}$  ( $1 \leq j \leq p$ ), respectivamente, e como  $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11}, \dots, \beta_{1p})$  e  $\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21}, \dots, \beta_{2p})$  representando o respectivo vetor de parâmetros a serem estimados.

Modelo de Regressão Robusta com Kernel Exponencial do tipo Gaussiano e Seleção Automática do Hiper-parâmetro para Variáveis do Tipo-intervalo (iETKRR-AS) considera as variáveis  $Y_1$  e  $Y_2$  e utiliza o chamado Truque da Distancia do Kernel (Eq. 2.1). A estimação dos parâmetros do modelo será feita pelo método de otimização denominado Método dos Multiplicadores de Lagrange, considerando a função  $L_1$  e guiada pela otimização da função  $S_1$  sujeita a uma restrição  $\delta$ ,  $\delta > 0$ . Assim,

$$L_1 = S_1 - \theta \left[ \frac{1}{\gamma_1^2} \frac{1}{\gamma_2^2} - \delta \right], \text{ onde} \quad (3.1)$$

$$S_1 = \sum_{i=1}^n 2[1 - \mathcal{K}(y_{1i}, \mu_{1i}, \gamma_1^2)] + \sum_{i=1}^n 2[1 - \mathcal{K}(y_{2i}, \mu_{2i}, \gamma_2^2)] \quad (3.2)$$

Sujeito a:

$$\frac{1}{\gamma_1^2} \frac{1}{\gamma_2^2} = \delta \text{ constante, onde } \frac{1}{\gamma_1^2} > 0, \frac{1}{\gamma_2^2} > 0 \quad (3.3)$$

Temos que  $\mathcal{K}$  é um Kernel Guassiano, dado por:

$$\mathcal{K}(y_{1i}, \mu_{1i}, \gamma_1^2) = \text{Exp} \left\{ -\frac{1}{2} \frac{(y_{1i} - \mu_{1i})^2}{\gamma_1^2} \right\} \quad , \quad \mu_{1i} = x_{1i}^T \beta_1 = \beta_{10} + \sum_{j=1}^p \beta_{1j} x_{1ij}. \quad (3.4)$$

$$\mathcal{K}(y_{2i}, \mu_{2i}, \gamma_2^2) = \text{Exp} \left\{ -\frac{1}{2} \frac{(y_{2i} - \mu_{2i})^2}{\gamma_2^2} \right\} \quad , \quad \mu_{2i} = x_{2i}^T \beta_2 = \beta_{20} + \sum_{j=1}^p \beta_{2j} x_{2ij}. \quad (3.5)$$

Note que pela restrição  $\delta$  proposta no modelo, temos:

$$\frac{1}{\gamma_1^2} \frac{1}{\gamma_2^2} = \delta \Rightarrow \frac{1}{\gamma_1^2} = \frac{\delta}{(1/\gamma_2^2)} \quad \text{ou} \quad \frac{1}{\gamma_2^2} = \frac{\delta}{(1/\gamma_1^2)}. \quad (3.6)$$

Diferenciando a equação 3.1 com relação a cada elemento, temos:

$$\frac{\partial L_1}{\partial (1/\gamma_1^2)} = 0 \quad , \quad \frac{\partial L_1}{\partial (1/\gamma_2^2)} = 0 \quad , \quad \frac{\partial L_1}{\partial \theta} = 0. \quad (3.7)$$

Assim temos que:

$$\frac{\partial L_1}{\partial (1/\gamma_1^2)} = 0 \Rightarrow \frac{1}{\gamma_1^2} = \frac{\theta \delta}{(\partial S_1 / \partial (1/\gamma_1^2))}. \quad (3.8)$$

Da mesma forma, temos que:

$$\frac{\partial L_1}{\partial (1/\gamma_2^2)} = 0 \Rightarrow \frac{1}{\gamma_2^2} = \frac{\theta \delta}{(\partial S_1 / \partial (1/\gamma_2^2))}. \quad (3.9)$$

Temos também que:

$$\frac{\partial L_1}{\partial \theta} = 0 \Rightarrow \theta \delta = \delta^{1/2} \left( \frac{\partial S_1}{\partial (1/\gamma_1^2)} \frac{\partial S_1}{\partial (1/\gamma_2^2)} \right)^{1/2}. \quad (3.10)$$

Desse modo, substituindo a Equação 3.10 na Equação 3.8, temos:

$$\frac{1}{\gamma_1^2} = \frac{\theta\delta}{(\partial S_1/\partial(1/\gamma_1^2))} = \frac{\delta^{1/2} \left( \frac{\partial S_1}{\partial(1/\gamma_1^2)} \frac{\partial S_1}{\partial(1/\gamma_2^2)} \right)^{1/2}}{(\partial S_1/\partial(1/\gamma_1^2))} = \frac{\delta^{1/2} \left( \frac{\partial S_1}{\partial(1/\gamma_2^2)} \right)^{1/2}}{(\partial S_1/\partial(1/\gamma_1^2))^{1/2}} \quad (3.11)$$

onde:

$$\frac{\partial S_1}{\partial(1/\gamma_1^2)} = \sum_{i=1}^n (y_{1i} - \mu_{1i})^2 \mathcal{K}(y_{1i}, \mu_{1i}, \gamma_1^2) \quad (3.12)$$

Da mesma forma, temos que:

$$\frac{\partial S_1}{\partial(1/\gamma_2^2)} = \sum_{i=1}^n (y_{2i} - \mu_{2i})^2 \mathcal{K}(y_{2i}, \mu_{2i}, \gamma_2^2) \quad (3.13)$$

Logo

$$\frac{1}{\gamma_1^2} = \frac{\delta^{1/2} [\sum_{i=1}^n (y_{2i} - \mu_{2i})^2 \mathcal{K}(y_{2i}, \mu_{2i}, \gamma_2^2)]^{1/2}}{[\sum_{i=1}^n (y_{1i} - \mu_{1i})^2 \mathcal{K}(y_{1i}, \mu_{1i}, \gamma_1^2)]^{1/2}}. \quad (3.14)$$

$$\frac{1}{\gamma_2^2} = \frac{\delta^{1/2} [\sum_{i=1}^n (y_{1i} - \mu_{1i})^2 \mathcal{K}(y_{1i}, \mu_{1i}, \gamma_1^2)]^{1/2}}{[\sum_{i=1}^n (y_{2i} - \mu_{2i})^2 \mathcal{K}(y_{2i}, \mu_{2i}, \gamma_2^2)]^{1/2}}. \quad (3.15)$$

Diferenciando a equação 3.2 com relação a cada elemento dos vetores dos parâmetros  $\beta_1$  e  $\beta_2$  obtemos, respectivamente, o conjunto de equações normais para o modelo iETKRR-AS, dado por:

$$\begin{aligned}
\frac{\partial S_1}{\partial \beta_{10}} &= -2 \sum_{i=1}^n \mathcal{K}(y_{1i}, \mu_{1i}, \gamma_1^2) \frac{(y_{1i} - \mu_{1i})}{\gamma_1^2} = 0, \\
\frac{\partial S_1}{\partial \beta_{1j}} &= -2 \sum_{i=1}^n \mathcal{K}(y_{1i}, \mu_{1i}, \gamma_1^2) \frac{x_{1ij}(y_{1i} - \mu_{1i})}{\gamma_1^2} = 0, (1 \leq j \leq p) \\
\frac{\partial S_1}{\partial \beta_{20}} &= -2 \sum_{i=1}^n \mathcal{K}(y_{2i}, \mu_{2i}, \gamma_2^2) \frac{(y_{2i} - \mu_{2i})}{\gamma_2^2} = 0, \\
\frac{\partial S_1}{\partial \beta_{2j}} &= -2 \sum_{i=1}^n \mathcal{K}(y_{2i}, \mu_{2i}, \gamma_2^2) \frac{x_{2ij}(y_{2i} - \mu_{2i})}{\gamma_2^2} = 0, (1 \leq j \leq p)
\end{aligned} \tag{3.16}$$

A solução para o sistema de equações normais 3.16 pode ser obtida pelo seguinte modelo de regressão:

$$\mathbf{y}^* = \mathbf{X}^* \boldsymbol{\beta} + \boldsymbol{\epsilon}^*$$

com

$$\mathbf{y}^* = \begin{bmatrix} \mathbf{y}_1^* \\ \mathbf{y}_2^* \end{bmatrix}, \quad \mathbf{X}^* = \begin{bmatrix} \mathbf{X}_1^* & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2^* \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}, \quad \text{e } \boldsymbol{\epsilon}^* = \begin{bmatrix} \boldsymbol{\epsilon}_1^* \\ \boldsymbol{\epsilon}_2^* \end{bmatrix},$$

onde:

- $\mathbf{y}_1^* = \mathbf{K}_1^{1/2} \mathbf{y}_1$  e  $\mathbf{y}_2^* = \mathbf{K}_2^{1/2} \mathbf{y}_2$  sendo vetores ponderados  $n \times 1$  das variáveis resposta  $Y_1$  e  $Y_2$ , respectivamente;
- $\mathbf{X}_1^* = \mathbf{K}_1^{1/2} \mathbf{X}_1$  e  $\mathbf{X}_2^* = \mathbf{K}_2^{1/2} \mathbf{X}_2$  sendo matrizes ponderadas  $n \times p$  das variáveis explicativas  $X_1$  e  $X_2$ , respectivamente;
- $\boldsymbol{\beta}_1$  e  $\boldsymbol{\beta}_2$  sendo os respectivos vetores de parâmetros da regressão. A estimativa para os vetores de parâmetros da regressão  $\boldsymbol{\beta}_1$  e  $\boldsymbol{\beta}_2$  é feita utilizando mínimos quadrados ponderados, dado por:

$$\beta_1^{(m)} = (X_1^T K_1^{(m-1)} X)^{-1} X^T K_1^{(m-1)} y_1 \tag{3.17}$$

$$\beta_2^{(m)} = (X_2^T K_2^{(m-1)} X)^{-1} X^T K_2^{(m-1)} y_2 \tag{3.18}$$

- $\boldsymbol{\epsilon}_1^* = \mathbf{K}_1^{1/2} \boldsymbol{\epsilon}_1$  e  $\boldsymbol{\epsilon}_2^* = \mathbf{K}_2^{1/2} \boldsymbol{\epsilon}_2$  sendo vetores de erros independentes ponderados  $n \times 1$ ;
- $\mathbf{K}_1 = \text{diag}(k_{11}, k_{12}, \dots, k_{1n})$  sendo uma matriz diagonal  $n \times n$  de pesos do kernel  $k_{1i}$

definido como:

$$k_{1i} = \mathcal{K}(y_{1i}, \mu_{1i}, \gamma_1^2) = \text{Exp} \left\{ -\frac{1}{2} \frac{(y_{1i} - \mu_{1i})^2}{\gamma_1^2} \right\} \quad (3.19)$$

para  $1 \leq i \leq n$ , onde  $0 \leq k_{1i} \leq 1$ ;

- $\mathbf{K}_2 = \text{diag}(k_{21}, k_{22}, \dots, k_{2n})$  sendo uma matriz diagonal  $n \times n$  de pesos do kernel  $k_{2i}$  definido como:

$$k_{2i} = \mathcal{K}(y_{2i}, \mu_{2i}, \gamma_2^2) = \text{Exp} \left\{ -\frac{1}{2} \frac{(y_{2i} - \mu_{2i})^2}{\gamma_2^2} \right\} \quad (3.20)$$

para  $1 \leq i \leq n$ , onde  $0 \leq k_{2i} \leq 1$ ;

Finalmente, dada uma nova observação  $e$  descrita por um vetor de intervalos  $(\mathbf{w}, z)$ , O valor  $z = [z_L, z_U]$  de  $Z$  que será previsto, considerando a regra proposta por Xu (2010), dada por:

- Se  $e$  for descrito em termos dos limites de intervalo inferior e superior, então  $\hat{z}_L = \min\{\mathbf{x}_1^\top \hat{\boldsymbol{\beta}}_1, \mathbf{x}_2^\top \hat{\boldsymbol{\beta}}_2\}$  e  $\hat{z}_U = \max\{\mathbf{x}_1^\top \hat{\boldsymbol{\beta}}_1, \mathbf{x}_2^\top \hat{\boldsymbol{\beta}}_2\}$ , onde  $\mathbf{x}_1^\top = (1, x_{11}, \dots, x_{1p})$  e  $\mathbf{x}_2^\top = (1, x_{21}, \dots, x_{2p})$ ,  $x_{1j} = a_j$  e  $x_{2j} = b_j$  ( $1 \leq j \leq p$ );
- Se  $e$  é descrito em termos do ponto médio e meio intervalo do intervalo, então  $\hat{z}_L = \min\{\hat{y}^c - \hat{y}^r, \hat{y}^c + \hat{y}^r\}$  e  $\hat{z}_U = \max\{\hat{y}^c - \hat{y}^r, \hat{y}^c + \hat{y}^r\}$ , onde:  $\hat{y}^c = \mathbf{x}_1^\top \hat{\boldsymbol{\beta}}_1$  e  $\hat{y}^r = \mathbf{x}_2^\top \hat{\boldsymbol{\beta}}_2$ , com  $\mathbf{x}_1^\top = (1, x_{11}, \dots, x_{1p})$  e  $\mathbf{x}_2^\top = (1, x_{21}, \dots, x_{2p})$ ,  $x_{1j} = \frac{a_j + b_j}{2}$  e  $x_{2j} = \frac{b_j - a_j}{2}$  ( $1 \leq j \leq p$ ).

### 3.3 Redescending M-estimator

Nessa seção é demonstrado que os estimadores do modelo iETKRR-AS se enquadram dentro da classe de estimadores M. Considere a função objetivo  $S_1$  definida na seção 3.2. Seja  $\rho(\epsilon_{qi}) = 1 - \exp(-\frac{\epsilon_{qi}^2}{2\gamma_q^2})$ , com  $\epsilon_{qi} = y_{qi} - \mu_{qi}$  com  $i = (1, \dots, n)$  e  $q = (1, 2)$ , onde os subscritos  $q$  representam o limite inferior (ou o ponto médio) e o limite superior (ou a amplitude) dos intervalos. Dessa forma temos que a função objetivo do método iETKRR-AS é dada por:

$$S_1 = \sum_{i=1}^n 2\rho(\epsilon_{1i}) + \sum_{i=1}^n 2\rho(\epsilon_{2i}) \quad (3.21)$$

Uma função  $\rho$  considerada um *Redescending M-estimator* são estimadores M que são capazes de rejeitar outliers extremos completamente. Esta classe de estimadores possuem funções que não diminuem perto da origem, mas diminuem em direção a 0 longe da origem, possuem altos pontos de ruptura (próximo a 0.5). A função  $\rho$  acima satisfaz as 6 propriedades para ser considerada um Redescending M-estimator. (Prova das propriedades encontra-se no Apêndice A.4)

1.  $\rho(0) = 0$ .
2.  $\rho(\epsilon_{qi}) \geq 0$ .
3.  $\rho(\epsilon_{qi}) = \rho(-\epsilon_{qi})$ .
4.  $\rho(\epsilon_{qi}) \leq \rho(\epsilon_{qj})$ . para  $|\epsilon_{qi}| \leq |\epsilon_{qj}|$ .
5.  $\rho$  é contínua e diferenciável.
6.  $\lim_{\epsilon_i \rightarrow \infty} \Psi(\epsilon_{qi}) = 0$  onde  $\Psi = \rho'$ .

### 3.4 Algoritmo de Estimação dos Parâmetros do Modelo iETKRR-AS

A estimação dos vetores de parâmetros  $\beta_1$  e  $\beta_2$  do modelo iETKRR-AS ocorre com base no algoritmo de mínimos quadrados ponderados iterativo apresentado abaixo

#### Algoritmo 1: Estimação dos Parâmetros do Modelo iETKRR-AS

**Início** :  $\mathbf{X}_1, \mathbf{X}_2, \mathbf{y}_1, \mathbf{y}_2, \epsilon$  (Tolerância),  $T$  (Iterações)

**Saída** :  $\hat{\beta}_1, \hat{\beta}_2, \hat{\mu}_1 = (\mu_{11}, \dots, \mu_{1n})^\top, \hat{\mu}_2 = (\mu_{21}, \dots, \mu_{2n})^\top$

Início:

Estabeleça  $t = 0$

Calcule  $\hat{\beta}_1^{(OLS)} = (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{y}_1$  e  $\hat{\beta}_2^{(OLS)} = (\mathbf{X}_2^\top \mathbf{X}_2)^{-1} \mathbf{X}_2^\top \mathbf{y}_2$ ;

Calcule  $\hat{\mu}_1^{(OLS)} = \mathbf{X}_1 \hat{\beta}_1^{(OLS)}$  e  $\hat{\mu}_2^{(OLS)} = \mathbf{X}_2 \hat{\beta}_2^{(OLS)}$ ;

Calcule  $\gamma_1^{2(0)}$  e  $\gamma_2^{2(0)}$ ;

Calcule  $\delta$ ;

$\mathbf{K}_1^{(0)} = \text{diag}\{k_{11}^{(0)}, \dots, k_{1n}^{(0)}\}$ , onde  $k_{1i}^{(0)} = \mathcal{K}(\mathbf{y}_{1i}, \boldsymbol{\mu}_{1i}^{(OLS)}, \gamma_1^{2(0)})$  e

$\mathbf{K}_2^{(0)} = \text{diag}\{k_{21}^{(0)}, \dots, k_{2n}^{(0)}\}$ , onde  $k_{2i}^{(0)} = \mathcal{K}(\mathbf{y}_{2i}, \boldsymbol{\mu}_{2i}^{(OLS)}, \gamma_2^{2(0)})$  onde  $\mathcal{K}$  é o Kernel Gaussiano;

Calcule  $\hat{\beta}_1^{(0)} = (\mathbf{X}_1^\top \mathbf{K}_1^{(0)} \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{K}_1^{(0)} \mathbf{y}_1$  e

$\hat{\beta}_2^{(0)} = (\mathbf{X}_2^\top \mathbf{K}_2^{(0)} \mathbf{X}_2)^{-1} \mathbf{X}_2^\top \mathbf{K}_2^{(0)} \mathbf{y}_2$ ;

Calcule  $\hat{\mu}_1^{(0)} = \mathbf{X}_1 \hat{\beta}_1^{(0)}$  e  $\hat{\mu}_2^{(0)} = \mathbf{X}_2 \hat{\beta}_2^{(0)}$ ;

Calcule  $S_1^{(0)}$  de acordo com a Eq. (3.2);

Estimação dos Parâmetros:

**Repita** :  $|S_1^{(t)} - S_1^{(t-1)}| \leq \epsilon$  ou  $t \geq T$

Estabeleça  $t = t + 1$ ;

Calcule  $\gamma_1^{2(t)}$  e  $\gamma_2^{2(t)}$  de acordo com a Eq. (3.14) e Eq. (3.15), respectivamente;

$\mathbf{K}_1^{(t)} = \text{diag}\{k_{11}^{(t)}, \dots, k_{1n}^{(t)}\}$ , onde  $k_{1i}^{(t)} = \mathcal{K}(\mathbf{y}_{1i}, \boldsymbol{\mu}_{1i}^{(t-1)}, \gamma_1^{2(t)})$  e

$\mathbf{K}_2^{(t)} = \text{diag}\{k_{21}^{(t)}, \dots, k_{2n}^{(t)}\}$ , onde  $k_{2i}^{(t)} = \mathcal{K}(\mathbf{y}_{2i}, \boldsymbol{\mu}_{2i}^{(t-1)}, \gamma_2^{2(t)})$  onde  $\hat{\mu}_{1i}^{(t-1)} = \mathbf{x}_{1i}^\top \hat{\beta}_1^{(t-1)}$  e  $\hat{\mu}_{2i}^{(t-1)} = \mathbf{x}_{2i}^\top \hat{\beta}_2^{(t-1)}$ ;

Calcule  $\hat{\beta}_1^{(t)} = (\mathbf{X}_1^\top \mathbf{K}_1^{(t)} \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{K}_1^{(t)} \mathbf{y}_1$  e  $\hat{\beta}_2^{(t)} = (\mathbf{X}_2^\top \mathbf{K}_2^{(t)} \mathbf{X}_2)^{-1} \mathbf{X}_2^\top \mathbf{K}_2^{(t)} \mathbf{y}_2$ ;

Calcule  $\hat{\mu}_1^{(t)} = \mathbf{X}_1 \hat{\beta}_1^{(t)}$  e  $\hat{\mu}_2^{(t)} = \mathbf{X}_2 \hat{\beta}_2^{(t)}$ ;

Calcule  $S_1^{(t)}$  de acordo com a Eq. (3.2);

O algoritmo parte da estimação dos parâmetros  $\beta_1$  e  $\beta_2$  utilizando o método dos mínimos quadrados ordinários para obter as estimativas de  $\mu_1$  e  $\mu_2$ ,  $\gamma_1^2$  e  $\gamma_2^2$  são computados na iteração inicial no algoritmo e é definida uma restrição  $\delta$  com base nos valores de  $\gamma_1^2$  e  $\gamma_2^2$ . Então com os valores iniciais de  $\mu_1$ ,  $\mu_2$ ,  $\gamma_1^2$  e  $\gamma_2^2$  estimados, é calculado o peso para cada observação utilizando o Kernel Gaussiano. Inicia-se a estimação dos parâmetros  $\beta_1$  e  $\beta_2$  usando mínimos quadrados ponderados iterativo, os parâmetros  $\gamma_1^2$  e  $\gamma_2^2$  são atualizados com base nas Equações 3.14 e 3.15, respectivamente. O peso de cada observação é atualizado e o processo alterna entre essas duas fases principais até a convergência da função objetivo.

Este Capítulo considera a aplicação do método proposto, iETKRR-AS, em conjuntos de dados reais com o objetivo de destacar sua utilidade em comparação com os métodos robustos IRR, IQR, SSLR e iETKRR. Em relação ao método SSLR, para o centro é considerado a distribuição t de Student e para a amplitude é considerada a distribuição Normal, a seleção do grau da distribuição t de Student usa o AIC (Akaike Information Criterion) em uma grade de valores de 2 a 10. O método IQR considera a mediana ( $\tau = 0,5$ ) para o centro e para a amplitude.

Para avaliar a robustez de cada método é calculado a porcentagem de mudança das estimativas dos parâmetros do modelo quando os outliers são suprimidos. Este critério é útil para identificar as abordagens menos sensíveis à contaminação dos dados por outliers. É utilizado o critério proposto por Fagundes, Souza e Cysneiros (2013) para identificar outlier. De acordo com este critério, uma observação com seu resíduo estudentizado maior que 2,0 é considerada um outlier. Uma observação pode ser considerada um outlier em relação ao centro, amplitude ou ambos.

Em relação ao estimadores para o hiper-parâmetro do kernel gaussiano, este Capítulo apresenta apenas os estimadores que apresentaram o melhor desempenho. Foram considerados o estimador S1 proposto por Caputo et al. (2002). O estimador S2, que representa a mediana dos valores de  $r_{ji} = (y_i - \mu_j^{OLS}), \forall r_{ji} \neq 0, i = 1, \dots, n$ . O estimador S3 é definido como:

$$\frac{1}{n-p-1} \sum_{i=1}^n (y_i - \mu_i^{OLS})^2 \quad (4.1)$$

onde  $p$  é o número de variáveis explicativas. Finalmente, o estimador S4 é baseado no critério de informação de Akaike melhorado ( $AIC_c$ ) proposto por Hurvich, Simonoff e Tsai (2002).

A restrição  $\delta$  do modelo iETKRR-AS é calculada com base no hiper-parâmetro, como mostrado na seção 3.2. A estimação da restrição  $\delta$  pode ser o mínimo entre os hiper-parâmetros, o máximo, a média aritmética ou a média geométrica. O rótulo iETKRR-AS $_{(H,R)}$ , significa que o melhor modelo foi obtido com os hiper-parâmetros estimados por H, com H podendo ser S1, S2, S3 e S4 e restrição R, com R podendo ser MIN(mínimo), MAX(máximo), MEAN(média aritmética), GEOM(média geométrica).

A tabela abaixo apresenta uma visão geral sobre os 6 conjuntos de dados considerados para avaliar os métodos.

Conjunto de Dados	Tamanho (n)	Variáveis Explicativas (p)	Outliers
Mushroom	23	2	1 (4%)
Soccer	20	2	1 (5%)
Unemployment	58	1	4 (7%)
China Temperature	225	3	21 (9%)
Cars	27	3	5 (18%)
Scientific Production	5630	2	153 (3%)

Tabela 4.1: Descrição dos conjuntos de dados reais com valor de intervalo

## 4.1 Conjunto de Dados Mushroom

Conjunto de dados de 23 espécies de cogumelos da família Amanita descritos por três variáveis de valor de intervalo: comprimento do estipe (W1), espessura do estipe (W2) como variáveis preditoras e a variável resposta comprimento do píleo (Z). Este conjunto de dados tem um único outlier (observação 12) no centro e na amplitude.

A Tabela 4.2 fornece a porcentagem de mudança (%) nas estimativas dos parâmetros após a supressão da observação do outlier para cada método. O método iETKRR-AS $_{(S2,MIN)}$  com estimador de hiper-parametros S2 para  $\gamma_1^2$  e  $\gamma_2^2$  e a restrição o mínimo dos parâmetros apresentou menor percentual de variação, ou seja, melhor rendimento. Seguido do iETKRR $_{(S4,S2)}$  com S4 e S2 como estimadores dos hiper-parametros  $\gamma_1$  e  $\gamma_2$ , respectivamente. Dentre os métodos, o método IQR teve o pior desempenho.

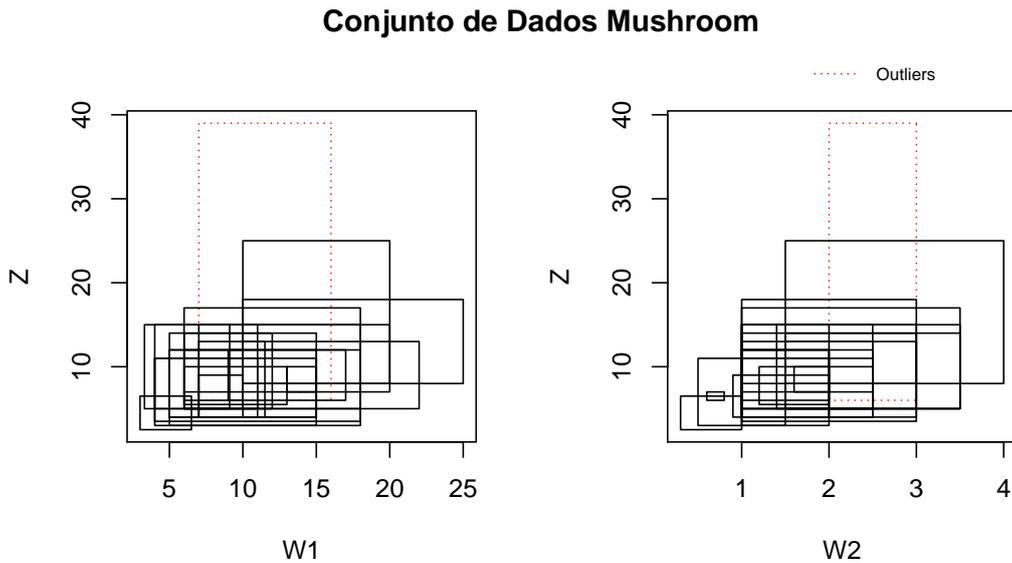


Figura 4.1: Conjunto de Dados Mushroom

Modelo	$\beta_0^c$	$\beta_1^c$	$\beta_2^c$	$\beta_0^r$	$\beta_1^r$	$\beta_2^r$	Total
iETKRR-AS <sub>(S2,MIN)</sub>	0.43	0.16	0.00	1.14	0.25	0.12	2.10
iETKRR <sub>(S4,S2)</sub>	8.61	2.96	0.00	1.66	0.35	0.16	13.75
IRR	7.14	5.10	0.16	2.56	2.86	2.53	20.34
IQR	1362.01	9.52	21.86	53.85	17.95	2.56	1467.76
SSLR (3,3)	10.91	0.49	3.11	20.27	4.68	2.04	41.50

Tabela 4.2: Comparação entre os métodos robustos de intervalo com base na porcentagem de mudança (%) nas estimativas dos parâmetros após a eliminação dos outliers. Conjunto de dados Mushroom

## 4.2 Conjunto de Dados Soccer

Este conjunto de dados é de cerca de 531 jogadores de futebol do Campeonato French Football Professional (NETO; CORDEIRO; CARVALHO, 2011). Os jogadores são agrupados em 20 times e são descritos por três variáveis de valor de intervalo: peso como a variável resposta ( $Z$ ), altura ( $W1$ ) e idade ( $W2$ ) são variáveis explicativas. Foi identificado um único outlier de centro (observação 14).

O método iETKRR<sub>(S4,S3)</sub> com estimadores de hiper-parâmetro  $S4$  e  $S3$  apresentou menor variação com estimação dos parâmetros da regressão após a supressão da observação do outlier, seguido pelo iETKRR-AS<sub>(S1,MIN)</sub>. Observe que o método iETKRR-AS<sub>(S1,MIN)</sub> apresenta variação inferior nos parâmetros de centro do modelo. Entre os métodos, IRR teve o pior desempenho.

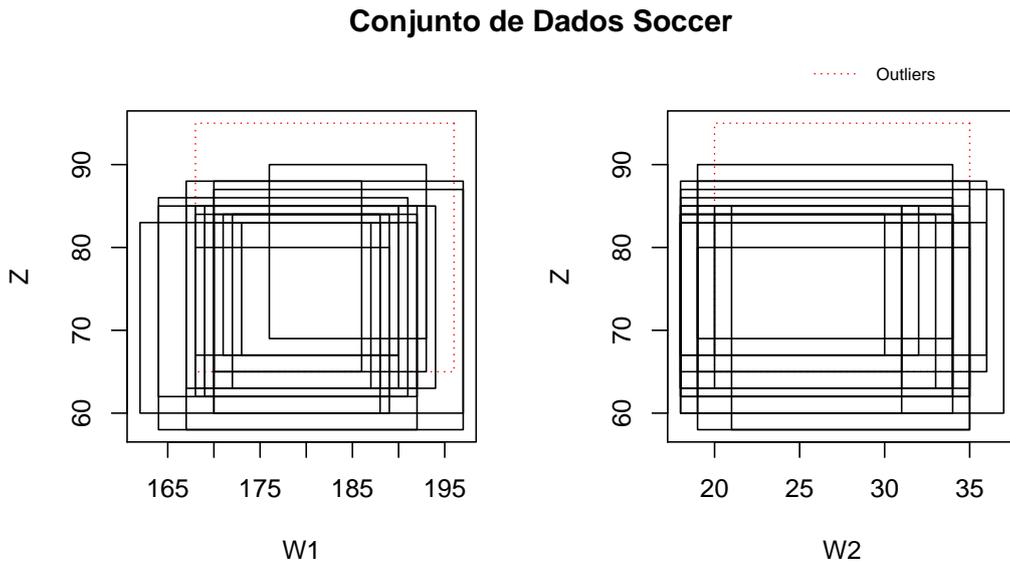


Figura 4.2: Conjunto de Dados Soccer

Modelo	$\beta_0^c$	$\beta_1^c$	$\beta_2^c$	$\beta_0^r$	$\beta_1^r$	$\beta_2^r$	Total
iETKRR-AS <sub>(S1,MIN)</sub>	0.61	0.48	0.65	142.48	13.29	4.22	161.72
iETKRR <sub>(S4,S3)</sub>	1.45	0.84	0.64	20.91	6.54	1.49	31.86
IRR	674.11	7.20	16.32	110.00	10.32	1.61	819.55
IQR	308.75	16.93	2.02	95.56	10.90	24.32	458.48
SSLR <sub>(2,2)</sub>	36.96	7.40	5.29	188.97	10.94	0.72	250.27

Tabela 4.3: Comparação entre os métodos robustos de intervalo com base na porcentagem de mudança (%) nas estimativas dos parâmetros após a eliminação dos outliers. Conjunto de dados Soccer

### 4.3 Conjunto de Dados Unemployment

Este conjunto de dados é referente a desempregados de Portugal (DIAS; BRITO, 2017). Possui 58 classes de indivíduos agrupados de acordo com gênero, região, idade e educação. Cada classe é descrita por duas variáveis de intervalo: a variável resposta log-tempo de desemprego ( $Z$ ) e a variável explicativa tempo que as pessoas trabalharam anteriormente ( $W$ ). Este conjunto de dados possui 4 outliers, observações 12 e 17 foram considerados outliers no centro e na amplitude dos intervalos, enquanto as observações 20 e 57 foram considerados outliers apenas no centro e apenas na amplitude, respectivamente

O método iETKRR<sub>(S1,S4)</sub> com estimadores de hiper-parâmetro S1 e S4 apresentou menor variação com estimação do parâmetros da regressão após a supressão da observação do outlier, seguido pelo iETKRR-AS<sub>(S4,MIN)</sub>. Entre os métodos, IQR teve o pior desempenho.

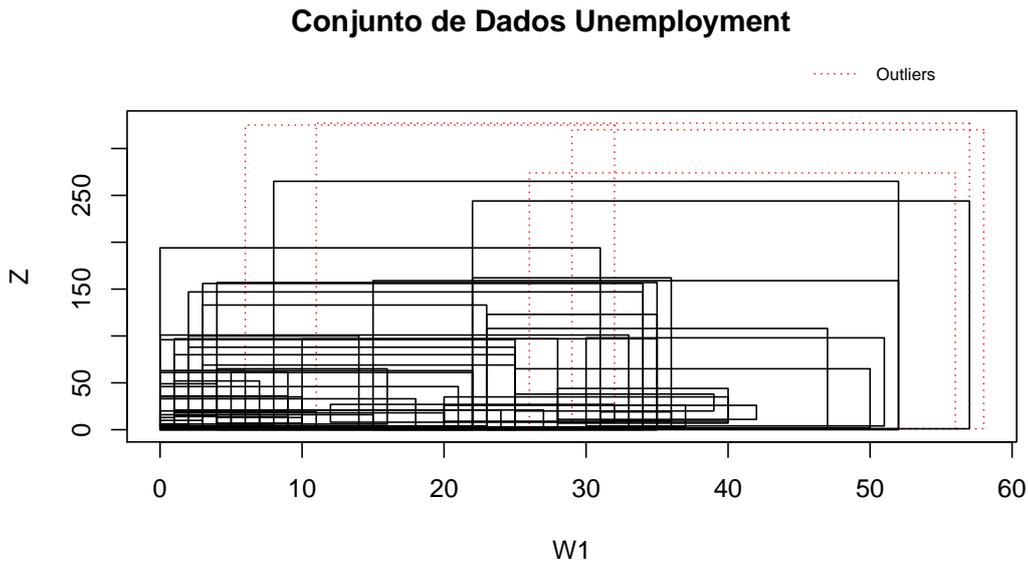


Figura 4.3: Conjunto de Dados Unemployment

Modelo	$\beta_0^c$	$\beta_1^c$	$\beta_0^r$	$\beta_1^r$	Total
iETKRR-AS <sub>(S4,MIN)</sub>	4.60	9.55	50.78	20.84	85.77
iETKRR <sub>(S1,S4)</sub>	2.89	13.84	12.29	0.55	29.57
IRR	18.04	24.86	34.64	10.09	87.63
IQR	204.75	46.08	47.48	15.44	313.75
SSLR <sub>(2,2)</sub>	27.59	34.24	35.22	11.62	108.66

Tabela 4.4: Comparação entre os métodos robustos de intervalo com base na porcentagem de mudança (%) nas estimativas dos parâmetros após a eliminação dos outliers. Conjunto de dados Unemployment

## 4.4 Conjunto de Dados China Temperature

Este conjunto de dados, da biblioteca R MAINT.DATA, é sobre os intervalos de temperatura (escala Celsius) no leste da China nos quatro trimestres, Q1 a Q4, de 1974 a 1988. Ele registra 225 observações com valores de intervalo descritas por quatro variáveis de valor de intervalo: a variável de resposta Z intervalo de temperatura no trimestre Q4 e três variáveis explicativas, em relação aos trimestres anteriores Q1 (W1), Q2 (W2) e Q3 (W3). Foram identificados 21 outliers, entre eles 10 somente no centro, 9 somente na amplitude e 2 no centro e na amplitude.

A Tabela 4.5 mostra o desempenho dos métodos. O método iETKRR<sub>(S4,S2)</sub>, seguido do iETKRR-AS<sub>(S2,MAX)</sub> apresentaram melhor performance, menor variação dos parâmetros após a supressão das observações outliers. Os métodos IRR, QIR e SSLR<sub>(10,10)</sub> não tiveram um bom desempenho, principalmente nos parâmetros da amplitude.

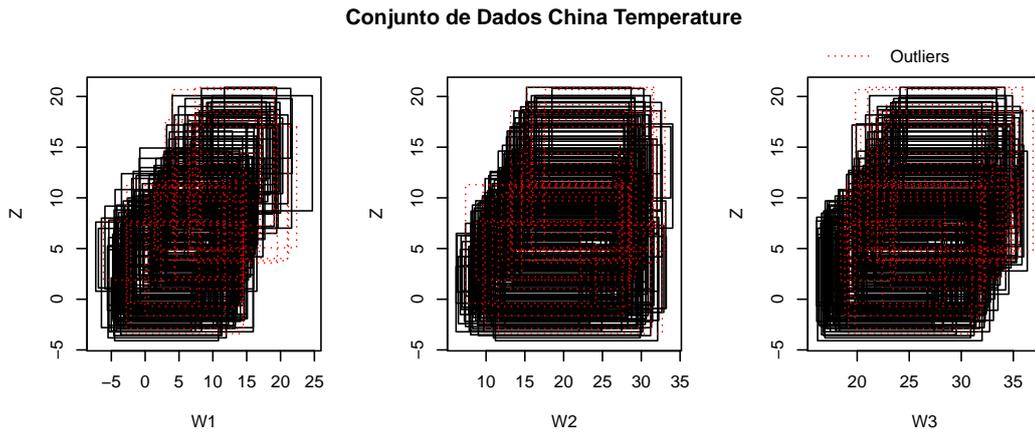


Figura 4.4: Conjunto de Dados China Temperature

Modelo	$\beta_0^c$	$\beta_1^c$	$\beta_2^c$	$\beta_3^c$	$\beta_0^r$	$\beta_1^r$	$\beta_2^r$	$\beta_3^r$	Total
iETKRR-AS <sub>(S2,MAX)</sub>	1.33	0.47	1.17	1.19	3.18	16.03	40.09	3.09	66.55
iETKRR <sub>(S4,S2)</sub>	13.24	0.53	8.34	5.05	1.92	1.54	4.81	1.03	36.46
IRR	20.42	0.82	12.50	7.14	7.37	44.72	77.31	13.97	184.24
IQR	4.63	1.69	0.36	2.65	4.97	108.41	98.86	3.63	225.21
SSLR <sub>(10,10)</sub>	17.09	0.24	12.09	4.97	8.64	47.70	68.89	12.34	171.95

Tabela 4.5: Comparação entre os métodos robustos de intervalo com base na porcentagem de mudança (%) nas estimativas dos parâmetros após a eliminação dos outliers. Conjunto de dados China Temperature

## 4.5 Conjunto de Dados Cars

Este conjunto de dados, disponível na biblioteca R MAINT.DATA, tem 27 modelos de carros descrito por quatro variáveis de valor de intervalo: a variável de resposta Preço (Z) e três variáveis explicativas, capacidade do motor (W1), velocidade máxima (W2) e aceleração (W3). Foram identificados 5 outliers, entre eles 2 somente no centro e 3 somente na amplitude.

A Tabela 4.6 mostra o desempenho dos métodos. O método iETKRR<sub>(S4,S1)</sub>, seguido do iETKRR-AS<sub>(S1,MIN)</sub> apresentaram melhor performance em relação à porcentagem de mudança nas estimativas dos parâmetros após a supressão das observações outliers. O método IRR apresentou pior desempenho.

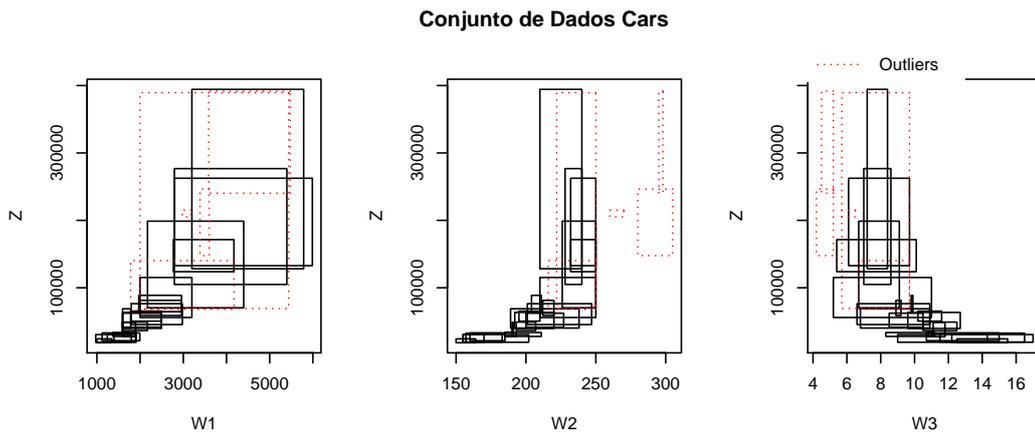


Figura 4.5: Conjunto de Dados Cars

Modelo	$\beta_0^c$	$\beta_1^c$	$\beta_2^c$	$\beta_3^c$	$\beta_0^r$	$\beta_1^r$	$\beta_2^r$	$\beta_3^r$	Total
iETKRR-AS <sub>(S1,MIN)</sub>	28.97	60.19	61.97	12.48	45.18	6.65	19.62	2.30	237.36
iETKRR <sub>(S4,S1)</sub>	0.00	0.00	0.00	0.00	3.97	23.15	67.85	70.26	165.23
IRR	144.43	1.91	31.64	64.87	44.90	5.82	382.93	55.38	731.88
IQR	196.45	5.62	205.91	30.49	9.62	23.96	118.66	90.66	681.38
SSLR <sub>(2,2)</sub>	100.74	1.80	73.98	79.18	92.54	1.16	45.89	110.75	506.03

Tabela 4.6: Comparação entre os métodos robustos de intervalo com base na porcentagem de mudança (%) nas estimativas dos parâmetros após a eliminação dos outliers. Conjunto de dados Cars

## 4.6 Conjunto de Dados Scientific Production

Este conjunto de dados trata-se da produção científica de instituições de pesquisa e universidades brasileiras. A base de dados original possui 141.260 pesquisadores brasileiros descrito por variáveis de valor real representando médias de valores de produção durante três anos (2006 a 2008) para cada pesquisador. Desse banco de dados, Pimentel e Souza (2014) construíram um conjunto de dados com valor de intervalo que agrega as pesquisas por instituição e sub-área do conhecimento, resultando em 5.630 itens relacionados à produção científica das instituições de acordo com a área disciplinar do conhecimento.

Consideramos esses 5.630 itens descritos por três variáveis de valor de intervalo: uma variável de resposta  $Z$  número de publicações de artigos e duas variáveis explicativas,  $W1$  número de alunos de doutorados orientados,  $W2$  número de alunos de mestrado orientados. Foram identificados 153 outliers, entre eles 9 outliers somente no centro, 12 outliers somente na amplitude e 132 outliers em ambos.

A Tabela 4.7 fornece a porcentagem de mudança (%) nas estimativas dos parâmetros após a supressão da observação do outlier para cada método. O método iETKRR-

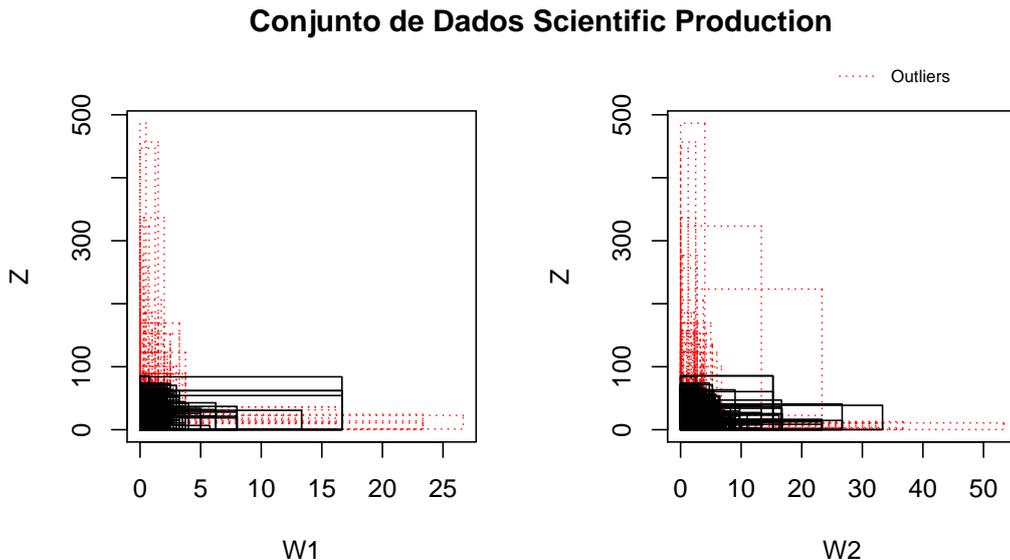


Figura 4.6: Conjunto de Dados Scientific Production

$AS_{(S1,MIN)}$  com estimador de hiper-parametros S1 e a restrição o mínimos dos parâmetros apresentou menor percentual de variação, ou seja, melhor rendimento. Seguindo do  $iETKRR_{(S1,S1)}$  com S1 como estimadores para ambos hiper-parametros  $\gamma_1$  e  $\gamma_2$ . Dentre os métodos, o método IQR teve o pior desempenho.

Modelo	$\beta_0^c$	$\beta_1^c$	$\beta_2^c$	$\beta_0^r$	$\beta_1^r$	$\beta_2^r$	Total
$iETKRR-AS_{(S1,MIN)}$	0.65	0.01	0.27	0.10	0.34	0.03	1.39
$iETKRR_{(S1,S1)}$	0.67	0.40	0.72	0.81	0.44	0.21	3.25
IRR	1.30	1.43	3.33	1.08	0.82	0.63	8.59
IQR	7.69	9.52	12.49	6.67	8.89	4.99	50.26
$SSLR_{(2,2)}$	3.13	2.38	9.09	3.03	1.41	4.81	23.86

Tabela 4.7: Comparação entre os métodos robustos de intervalo com base na porcentagem de mudança (%) nas estimativas dos parâmetros após a eliminação dos outliers. Conjunto de dados Taobao

## 4.7 Resumo do desempenho do $iETKRR-AS$

O método proposto  $iETKRR-AS$  foi avaliado em comparação com métodos de regressão robustos em sete conjuntos de dados com valores de intervalo. A Tabela 4.8 fornece a classificação dos métodos em cada conjunto de dados com base na porcentagem de mudança das estimativas de parâmetro, classificação 1 para a menor porcentagem, 2 para segunda menor porcentagem e assim por diante.

Os métodos IRR e SSLR exibiram um desempenho bastante semelhante. Além disso, iETKRR-AS apresentou o segundo melhor desempenho. O método iETKRR em geral foi o melhor nesses conjuntos de dados reais. Os resultados também destacaram a utilidade de diferentes estimadores para o hiper-parâmetro de largura a ser utilizado, como resultado de diferentes tipos de outliers que estão presentes nos conjuntos de dados reais.

Modelo	Mushroom	Soccer	Unempl.	Temp.	Cars	Scientif.	Média
iETKRR-AS	1	2	2	2	2	1	1.67
iETKRR	2	1	1	1	1	2	1.33
IRR	3	5	3	4	5	3	3.83
IQR	5	4	5	5	4	5	4.67
SSLR	4	3	4	3	3	4	3.50

Tabela 4.8: Classificação média dos métodos nos conjuntos de dados em relação ao percentual mudança da estimação dos parâmetros após a eliminação dos outliers

## 5.1 Conclusão

Um novo método de regressão linear robusto para variáveis com valor do tipo intervalo foi proposto neste trabalho, denominado Modelo de Regressão Robusta com Kernel Exponencial do tipo Gaussiano e Seleção Automática do Hiper-parâmetro para Variáveis do Tipo-intervalo (iETKRR-AS). Neste método, o modelo é baseado em funções kernel exponencial do tipo Gaussiano, os parâmetros do modelo e os hiper-parâmetros do kernel são estimados na iteração do método, a partir dessa estimacão é definida uma restrição  $\delta$  com base nos hiper-parâmetros, sendo os outliers penalizados utilizando a função kernel gaussiana, considerada um Redescending M-estimator. Tal classe de estimadores possuem propriedades interessantes, garantindo uma alta robustez do método.

O algoritmo de estimacão de parâmetros proposto é baseado em um processo de mínimos quadrados iterativos ponderados, sendo necessários valores iniciais para o vetor de parâmetros do ponto médio e da amplitude dos intervalos, bem como para o estimador do hiper-parâmetros do kernel gaussiano.

Um estudo comparativo entre o método iETKRR-AS e outras abordagens de regressão robustas para variáveis de valor de intervalo em 6 conjuntos de dados reais demonstrou a utilidade do método proposto. Para avaliar a robustez, foi calculado a porcentagem de mudançã das estimativas dos parâmetros do modelo quando os outliers são suprimidos. O método proposto mostrou um desempenho satisfatório em comparacão com os métodos de regressão robustos iETKRR, IRR, QIR e SSLR.

Em relação a trabalhos futuros, apesar do bom desempenho em bases de dados reais, é preciso ser feito um estudo controlado onde o modelo proposto, iETKRR-AS, será avaliado com conjuntos de dados sintéticos em termos de viés e erro quadrático médio (EQM) das estimativas de parâmetros levando em consideração outliers no espaço X, outliers no espaço Y, pontos de alavancagem, tamanhos de amostra diferentes e porcentagem de outliers na amostra.

## REFERÊNCIAS BIBLIOGRÁFICAS

- BERTRAND, P.; GOUPIL, F. Descriptive statistics for symbolic data. 2000.
- BILLARD, L.; DIDAY, E. Symbolic regression analysis. 2002.
- BOCK et al. Analysis of symbolic data: Exploratory methods for extracting statistical information from complex data. 2000.
- BOCK, H. Clustering algorithms and kohonen maps for symbolic data. 2002.
- CAPUTO, B. et al. Appearance-based object recognition using svms: Which kernel should i use. 2002.
- CARVALHO, F. D. Histograms in symbolic data analysis. 1995.
- CARVALHO, F. de A.T. de et al. Adaptive hausdorff distances and dynamic clustering of symbolic data. 2006.
- CAZES, P. et al. Extension de l'analyse en composantes principales des donnes de type intervalle. 1997.
- CAZES, P. et al. Extensions de l'analyse en composantes principales a des donnés de type intervalle. 1997.
- CIAMPI, A. et al. Growing a tree classifier with imprecise data. 2000.
- DIAS, S.; BRITO, P. Off the beaten track: A new linear model for interval data. 2017.
- DIDAY, E.; BILLARD, L. Symbolic data analysis: Conceptual statistics and data mining. 2006.
- DOMINGUES, M. A.; SOUZA, R. M. de; CYSNEIROS, F. J. A. A robust method for linear regression of symbolic interval data. 2010.
- FAGUNDES, R.; SOUZA, R.; SOARES, Y. Quantile regression of interval-valued data. 2016.

- FAGUNDES, R. A.; SOUZA, R. M. de; CYSNEIROS, F. J. A. Robust regression with application to symbolic interval data. 2013.
- Gowda, K. C.; Diday, E. Symbolic clustering using a new dissimilarity measure. 1991.
- Gowda, K. C.; Diday, E. Symbolic clustering using a new similarity measure. 1992.
- HAYKIN, S. S. Neural networks: A comprehensive foundation. 1994.
- HURVICH, C.; SIMONOFF, J.; TSAI, C.-L. Smoothing parameter selection in nonparametric regression using an improved akaike information criterion. 2002.
- ICHINO, M.; YAGUCHI, H. Generalized minkowski metrics for mixed feature type data analysis. 1994.
- INSHA, U.; QADIR, M.; ALI, A. Insha's redescending m-estimator for robust regression: A comparative study. 2006.
- LAURO, C.; PALUMBO, F. Principal component analysis of interval data: a symbolic data analysis approach1. 2000.
- MAIA, A. L. S.; CARVALHO, F. de A.T. de; LUDERMIR, T. B. A hybrid model for symbolic interval time series forecasting. 2006.
- MULLER, K. . R. et al. An introduction to kernel-based learning algorithms. 2001.
- NETO, E. de A. L. Modelos de regressão para dados simbólicos de natureza intervalar. 2008.
- NETO, E. de A. L.; CARVALHO, F. de A.T. de. An exponential-type kernel robust regression model for interval-valued variables. 2018.
- NETO, E. L.; CARVALHO, F. Centre and range method for fitting a linear regression model to symbolic interval data. 2008.
- NETO, E. L.; CORDEIRO, G.; CARVALHO, F. Bivariate symbolic regression models for interval-valued variables. 2011.
- PARDOUX, C.; SUMMA, M. G. Symbolic approaches for three-way data. 2000.
- PIMENTEL, B. A.; SOUZA, R. M. de. A weighted multivariate fuzzy c-means method in interval-valued scientific production data. 2014.
- XU, W. Symbolic data analysis: interval-valued data regression. 2010.
- ÖZLEM, G. A. Comparison of robust regression methods in linear regression. 2011.

**A.1 Solução Equação 3.8 e por conseguinte Equação 3.9**

$$\begin{aligned}
 \frac{\partial L_1}{\partial(1/\gamma_1^2)} &= 0 & (A.1) \\
 \frac{\partial S_1}{\partial(1/\gamma_1^2)} - \theta \frac{1}{\gamma_2^2} &= 0 \\
 \frac{\partial S_1}{\partial(1/\gamma_1^2)} - \theta \frac{\delta}{(1/\gamma_1^2)} &= 0 \\
 \frac{\partial S_1}{\partial(1/\gamma_1^2)} - \frac{\theta\delta}{(1/\gamma_1^2)} &= 0 \\
 \frac{\partial S_1}{\partial(1/\gamma_1^2)} &= \frac{\theta\delta}{(1/\gamma_1^2)} \\
 \frac{1}{\gamma_1^2} &= \frac{\theta\delta}{(\partial S_1/\partial(1/\gamma_1^2))}
 \end{aligned}$$

## A.2 Solução Equação 3.10

$$\begin{aligned}
 \frac{\partial L_1}{\partial \theta} = 0 &\Rightarrow - \left[ \frac{1}{\gamma_1^2} \frac{1}{\gamma_2^2} - \delta \right] = 0 & (A.2) \\
 &\frac{1}{\gamma_1^2} \frac{1}{\gamma_2^2} = \delta \\
 \frac{\theta \delta}{(\partial S_1 / \partial (1/\gamma_1^2))} &= \frac{\theta \delta}{(\partial S_1 / \partial (1/\gamma_2^2))} = \delta \\
 (\theta \delta)^2 &= \delta \left( \frac{\partial S_1}{\partial (1/\gamma_1^2)} \frac{\partial S_1}{\partial (1/\gamma_2^2)} \right) \\
 \theta \delta &= \delta^{1/2} \left( \frac{\partial S_1}{\partial (1/\gamma_1^2)} \frac{\partial S_1}{\partial (1/\gamma_2^2)} \right)^{1/2}
 \end{aligned}$$

## A.3 Solução Equação 3.12 e por conseguinte Equação 3.13

$$\begin{aligned}
 \frac{\partial S_1}{\partial (1/\gamma_1^2)} &= \left( \sum_{i=1}^n 2[1 - \mathcal{K}(y_{1i}, \mu_{1i}, \gamma_1^2)] + \sum_{i=1}^n 2[1 - \mathcal{K}(y_{2i}, \mu_{2i}, \gamma_2^2)] \right)' & (A.3) \\
 &= \left( \sum_{i=1}^n 2[1 - \mathcal{K}(y_{1i}, \mu_{1i}, \gamma_1^2)] \right)' + \left( \sum_{i=1}^n 2[1 - \mathcal{K}(y_{2i}, \mu_{2i}, \gamma_2^2)] \right)' \\
 &= \left( \sum_{i=1}^n 2[1 - \mathcal{K}(y_{1i}, \mu_{1i}, \gamma_1^2)] \right)' \\
 &= \left( -2 \sum_{i=1}^n \mathcal{K}(y_{1i}, \mu_{1i}, \gamma_1^2) \right)' \\
 &= \left( -2 \sum_{i=1}^n \text{Exp} \left\{ -\frac{1}{2} \frac{(y_{1i} - \mu_{1i})^2}{\gamma_1^2} \right\} \right)' \\
 &= -2 \sum_{i=1}^n \text{Exp} \left\{ -\frac{1}{2} \frac{(y_{1i} - \mu_{1i})^2}{\gamma_1^2} \right\} \left( -\frac{(y_{1i} - \mu_{1i})^2}{2} \right) \\
 &= \sum_{i=1}^n (y_{1i} - \mu_{1i})^2 \text{Exp} \left\{ -\frac{1}{2} \frac{(y_{1i} - \mu_{1i})^2}{\gamma_1^2} \right\} \\
 &= \sum_{i=1}^n (y_{1i} - \mu_{1i})^2 \mathcal{K}(y_{1i}, \mu_{1i}, \gamma_1^2)
 \end{aligned}$$

## A.4 Prova Propriedades Redescending M-estimator

- **Prova 1:**

$$\rho(0) = 1 - \exp\left\{\frac{-|0|^2}{2\gamma^2}\right\} = 1 - \exp\{-0\} = 1 - 1 = 0$$

- **Prova 2:**

$$\text{Temos que para } \epsilon = \begin{cases} 0, & \exp\left\{\frac{-|\epsilon_i|^2}{2\gamma^2}\right\} = 1 \\ \neq 0, & \exp\left\{-\frac{|\epsilon_i|^2}{2\gamma^2}\right\} \in (0, 1) \end{cases}, \text{ portanto } \rho(\epsilon) = \begin{cases} 0, & \epsilon = 0 \\ \geq 0, & \epsilon \neq 0 \end{cases}$$

- **Prova 3:**

$$\rho(-\epsilon) = 1 - \exp\left\{-\frac{|-\epsilon|^2}{2\gamma^2}\right\} = 1 - \exp\left\{-\frac{|\epsilon|^2}{2\gamma^2}\right\} = \rho(\epsilon)$$

- **Prova 4:**

Considere  $f(\epsilon) = \exp\left\{-\frac{|\epsilon|^2}{2\gamma^2}\right\}$ , um valor  $\epsilon$  grande fará  $f(\epsilon) \rightarrow 0$  rapidamente, logo, se temos  $0 < \epsilon_1 \leq \epsilon_2$  então  $f(\epsilon_1) \geq f(\epsilon_2)$ , portanto,

$$\rho(r_1) = 1 - f(r_1) \leq \rho(r_2) = 1 - f(r_2) \Rightarrow \rho(r_1) \leq \rho(r_2)$$

- **Prova 5**

Por definição, uma função é contínua e diferencial, se  $\lim_{n \rightarrow c} \rho(\epsilon) = \rho(c)$  e  $\lim_{h \rightarrow 0} \frac{f(a+h) - f(h)}{h}$  existir, temos então

$$\lim_{n \rightarrow c} \rho(\epsilon) = 1 - \exp\left\{-\frac{|c|^2}{2\gamma^2}\right\} = \rho(c)$$

alem disso,

$$\begin{aligned} \rho'(\epsilon) &= \lim_{h \rightarrow 0} \frac{\rho(\epsilon+h) - \rho(\epsilon)}{h} = \lim_{h \rightarrow 0} \frac{1 - \exp\left\{-\frac{|\epsilon+h|^2}{2\gamma^2}\right\} - \left(1 - \exp\left\{-\frac{|\epsilon|^2}{2\gamma^2}\right\}\right)}{h} \\ &= \lim_{h \rightarrow 0} \frac{1 - \exp\left\{-\frac{|\epsilon+h|^2}{2\gamma^2}\right\} - 1 + \exp\left\{-\frac{|\epsilon|^2}{2\gamma^2}\right\}}{h} \\ &= \lim_{h \rightarrow 0} \frac{-\exp\left\{-\frac{|\epsilon+h|^2}{2\gamma^2}\right\} \exp\left\{-\frac{|\epsilon|^2}{2\gamma^2}\right\}}{h} \\ &= \lim_{h \rightarrow 0} \frac{\exp\left\{-\frac{|\epsilon|^2}{2\gamma^2} / -\frac{|\epsilon+h|^2}{2\gamma^2}\right\}}{h} \end{aligned}$$

$$= \lim_{h \rightarrow 0} \frac{\exp\left\{\frac{|\epsilon|^2}{|\epsilon+h|^2}\right\}}{h}$$

Devido a indeterminação  $\frac{\exp\{1\}}{0}$ , por L'Hospital, temos  $\frac{f(x)'}{g(x)'}$  onde  $f(\epsilon) = \exp\left\{\left(\frac{\epsilon}{\epsilon+h}\right)^2\right\}$  e  $g(x) = h$ , Portando

$$\rho'(\epsilon) = \lim_{h \rightarrow 0} \frac{\rho(\epsilon + h) - \rho(\epsilon)}{h} = \lim_{h \rightarrow 0} \frac{-2\epsilon^2 \exp\left\{\left(\frac{\epsilon}{\epsilon+h}\right)^2\right\}}{(\epsilon + h)^3} = \frac{-2\exp\{1\}}{\epsilon}$$

• **Prova 6**

Temos  $\Psi(\epsilon_i) = \rho'(\epsilon_i) = \frac{-2\exp\{1\}}{\epsilon_i}$ , logo  $\lim_{\epsilon_i \rightarrow \infty} \Psi(\epsilon_i) = 0$