

**UNICAMP**

UNIVERSIDADE ESTADUAL DE  
CAMPINAS

Instituto de Matemática, Estatística e  
Computação Científica

FELIPE SANTOS DA SILVA

**Regressão para Dados Intervalares e *Fuzzy*:  
Transversal e Longitudinal**

Campinas

2023

Felipe Santos da Silva

# **Regressão para Dados Intervalares e *Fuzzy*: Transversal e Longitudinal**

Tese apresentada ao Instituto de Matemática,  
Estatística e Computação Científica da Uni-  
versidade Estadual de Campinas como parte  
dos requisitos exigidos para a obtenção do  
título de Doutor em Matemática Aplicada.

Orientador: Laécio Carvalho de Barros

Este trabalho corresponde à versão  
final da Tese defendida pelo aluno Fe-  
lipe Santos da Silva e orientada pelo  
Prof. Dr. Laécio Carvalho de Barros.

Campinas

2023

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca do Instituto de Matemática, Estatística e Computação Científica  
Ana Regina Machado - CRB 8/5467

Si38r Silva, Felipe Santos da, 1987-  
Regressão para dados intervalares e *fuzzy* : transversal e longitudinal /  
Felipe Santos da Silva. – Campinas, SP : [s.n.], 2023.

Orientador: Laécio Carvalho de Barros.  
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de  
Matemática, Estatística e Computação Científica.

1. Números fuzzy. 2. Análise de regressão. 3. Aritmética fuzzy. 4. Análise de intervalos (Matemática). 5. Método longitudinal. I. Barros, Laécio Carvalho de, 1954-. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Informações Complementares

**Título em outro idioma:** Interval and fuzzy data regression : cross-sectional and longitudinal

**Palavras-chave em inglês:**

Fuzzy numbers

Regression analysis

Fuzzy arithmetic

Interval analysis (Mathematics)

Longitudinal method

**Área de concentração:** Matemática Aplicada

**Titulação:** Doutor em Matemática Aplicada

**Banca examinadora:**

Laécio Carvalho de Barros [Orientador]

Aurelio Ribeiro Leite de Oliveira

Caio Lucidius Naberezny Azevedo

Renata Hax Sander Reiser

Leonardo Ramos Emmendorfer

**Data de defesa:** 10-02-2023

**Programa de Pós-Graduação:** Matemática Aplicada

**Identificação e informações acadêmicas do(a) aluno(a)**

- ORCID do autor: <https://orcid.org/0009-0007-7329-9997>

- Currículo Lattes do autor: <http://lattes.cnpq.br/7157161020322155>

**Tese de Doutorado defendida em 10 de fevereiro de 2023 e aprovada  
pela banca examinadora composta pelos Profs. Drs.**

**Prof(a). Dr(a). LAÉCIO CARVALHO DE BARROS**

**Prof(a). Dr(a). AURELIO RIBEIRO LEITE DE OLIVEIRA**

**Prof(a). Dr(a). CAIO LUCIDIUS NABEREZNY AZEVEDO**

**Prof(a). Dr(a). RENATA HAX SANDER REISER**

**Prof(a). Dr(a). LEONARDO RAMOS EMMENDORFER**

A Ata da Defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós-Graduação do Instituto de Matemática, Estatística e Computação Científica.

*À minha família.*

# Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) - Código de Financiamento 001.

Agradeço ao professor Laécio pela orientação e pela sugestão do tema.

Agradeço profundamente à Olimpíada Brasileira de Matemática das Escolas Públicas, OBMEP, e também ao Programa de Iniciação Científica e Mestrado. Com a iniciação científica jr, recebida como prêmio da OBMEP, aprendi o tamanho da importância da matemática. Já o PICME me proporcionou as bolsas de fomento para o mestrado e o doutorado, com as quais pude passar o período da pós-graduação com muitas preocupações a menos.

Agradeço aos membros da banca, professores Aurelio, Caio, Renata e Leonardo, pela presença e pelas sugestões para melhorar o texto.

Agradeço aos membros suplentes da banca, professores Estevão Esmi Laureano, Luiz Alberto Díaz Rodrigues e Marcia Pasin, pela presença e pela disponibilidade.

Agradeço ao meu colega, Renan Willian Prado, pela amizade durante esses anos de doutorado.

Victor Hugo Barreiro, agradeço pela companhia, por ter me dado suporte nos momentos difíceis e aproveitado os momentos felizes ao longo do meu doutorado.

*“Mesmo com asas desajeitadas sei que consigo voar.”*  
*(Wada Kōji)*

# Resumo

Neste trabalho, apresentamos métodos de regressão para dados que possuem algum tipo de imprecisão na medida tomada. Antes de realizar a regressão, a imprecisão dos dados é modelada na forma de intervalos ou de números *fuzzy*.

Os métodos de regressão transversal seguem diretamente do método de análise sensitiva, que é definido de forma indireta através de uma definição de conjunto. Apresentamos, porém, uma expressão que usa a aritmética intervalar que fornece as previsões dadas pela regressão de forma explícita. Além disso, através da aritmética para números *fuzzy*, apresentamos métodos adequados para quando os dados possuem essa natureza.

Como método de regressão para dados longitudinais apresentamos uma adaptação do método de análise sensitiva para construir um método linear misto em dois estágios para dados imprecisos.

Apresentamos também os métodos propostos aplicados aos dados de um estudo que investiga a variação do volume corporal de um grupo de orcas-pigmeias quando se aproximaram das águas costeiras da baía de Ma'alaea, no Maui, em setembro e outubro de 2019. Um estudo sobre a poluição atmosférica na cidade de São Paulo também tem seus dados usados para construir exemplos a partir do método transversal.

**Palavras-chave:** *fuzzy*. intervalar. regressão. transversal. longitudinal.

# Abstract

In this work, we present regression methods for data with some type of inaccuracy in the measure taken. Before performing the regression, the inaccuracy of the data with the form of intervals or *fuzzy* numbers.

Cross-sectional regression methods directly follow the sensitive analysis method, which is defined indirectly through a set definition. However, we present an expression that uses interval arithmetic that provides the predictions given by the regression explicitly. Furthermore, through arithmetic for *fuzzy* numbers, we present the methods when the data have this nature.

As a regression method for longitudinal data, we show an adaptation of the sensitive analysis method to build a mixed two-stage linear method for inaccurate data.

We also present the proposed methods applied to data from a study that investigated the variation in body volume of a group of pygmy killer whales when they approached the coastal waters of Ma'alaea Bay, Maui, in September and October 2019. A study on atmospheric pollution in the city of São Paulo also has its data used to build examples from cross-sectional methods.

**Keywords:** *fuzzy*. interval. regression. cross-sectional. longitudinal.

# Sumário

	Introdução . . . . .	13
<b>I</b>	<b>REVISÃO BIBLIOGRÁFICA</b>	<b>16</b>
1	REGRESSÃO LINEAR . . . . .	17
1.1	Introdução . . . . .	17
1.2	Regressão Linear Simples . . . . .	18
1.3	Regressão Linear Múltipla . . . . .	20
1.4	Avaliação dos Modelos . . . . .	21
1.5	Conclusão . . . . .	22
2	DADOS LONGITUDINAIS . . . . .	23
2.1	Introdução . . . . .	23
2.2	Dados Longitudinais . . . . .	23
2.3	Modelo Linear Misto em Dois Estágios . . . . .	25
2.4	Exemplos . . . . .	27
2.5	Conclusão . . . . .	31
3	ARITMÉTICA INTERVALAR . . . . .	32
3.1	Introdução . . . . .	32
3.2	Operações Aritméticas entre Intervalos . . . . .	32
3.3	Ausência de Algumas Propriedades e Diferença de Hukuhara . . . . .	34
3.4	Princípio de Extensão para Intervalos . . . . .	36
3.5	Conclusão . . . . .	37
4	CONJUNTOS E NÚMEROS <i>FUZZY</i> . . . . .	38
4.1	Introdução . . . . .	38
4.2	Conjuntos <i>Fuzzy</i> . . . . .	38
4.3	Números <i>Fuzzy</i> . . . . .	39
4.4	Conclusão . . . . .	42
5	MODELANDO IMPRECISÃO ATRAVÉS DE INTERVALOS E DE NÚMEROS <i>FUZZY</i> . . . . .	43
5.1	Introdução . . . . .	43
5.2	Modelando Imprecisão . . . . .	43
5.3	Conclusão . . . . .	45

<b>6</b>	<b>MÉTODOS PARA DADOS <i>FUZZY</i></b>	<b>46</b>
6.1	Introdução	46
6.2	Métodos Possibilísticos	47
6.3	Métodos de Quadrados Mínimos	48
6.4	Regressão Quantílica e Análise Sensitiva	50
6.5	Avaliação dos Modelos	51
6.6	Exemplos	52
6.7	Conclusão	67
<b>II</b>	<b>PROPOSTA DE MODELOS PARA DADOS INTERVALARES E <i>FUZZY</i></b>	<b>68</b>
<b>7</b>	<b>PROPOSTA DE MODELOS TRANSVERSAIS PARA DADOS INTERVALARES E <i>FUZZY</i></b>	<b>69</b>
7.1	Introdução	69
7.2	Modelo Base	70
7.3	Modelo Linear Simples para Dados Intervalares e <i>Fuzzy</i>	75
7.3.1	Exemplos	77
7.4	Modelo Linear Múltiplo para Dados Intervalares e <i>Fuzzy</i>	81
7.5	Interpretando o Modelo como uma Média Ponderada	84
7.6	Conclusão	88
<b>8</b>	<b>MODELO LINEAR MISTO EM DOIS ESTÁGIOS PARA DADOS IMPRECISOS</b>	<b>90</b>
8.1	Introdução	90
8.2	Proposta de Modelo para o caso Intervalar	91
8.3	Proposta de Modelo para o caso <i>Fuzzy</i>	96
8.4	Exemplos	96
8.5	Conclusão	100
<b>III</b>	<b>APLICAÇÕES</b>	<b>103</b>
<b>9</b>	<b>VOLUME CORPORAL DE ORCAS-PIGMEIAS</b>	<b>104</b>
9.1	Introdução	104
9.2	Modelagem da Imprecisão	106
9.3	Caso Transversal	106
9.3.1	Dados na forma de números reais	107
9.3.2	Dados Imprecisos	109
9.3.2.1	Dados na Forma de Intervalos	111

9.3.2.2	Dados na Forma de Números <i>Fuzzy</i> Triangulares . . . . .	113
9.3.2.3	Dados na Forma de Números <i>Fuzzy</i> Trapezoidais . . . . .	115
9.3.2.4	Figuras e Considerações . . . . .	117
<b>9.4</b>	<b>Caso Longitudinal</b> . . . . .	<b>121</b>
9.4.1	Modelagem Clássica . . . . .	121
9.4.2	Modelagem Intervalar . . . . .	128
<b>9.5</b>	<b>Conclusão</b> . . . . .	<b>133</b>
<b>10</b>	<b>POLUIÇÃO ATMOSFÉRICA NA CIDADE DE SÃO PAULO</b> . . . . .	<b>134</b>
<b>10.1</b>	<b>Introdução</b> . . . . .	<b>134</b>
<b>10.2</b>	<b>Modelos de Regressão</b> . . . . .	<b>135</b>
<b>10.3</b>	<b>Conclusão</b> . . . . .	<b>148</b>
<b>11</b>	<b>COMENTÁRIOS FINAIS</b> . . . . .	<b>149</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>152</b>

# Introdução

Para estudar determinado fenômeno, coletamos dados referentes a ele para análise e criação de modelos. No entanto, devido a situações diversas, esses dados obtidos contém algum tipo de imprecisão ou incerteza, não fazemos distinção entre esses dois termos ao longo do texto. Uma das decisões que o pesquisador deve tomar ao iniciar o estudo sobre determinado fenômeno é qual o tratamento dará para essas imprecisões. Uma das formas é tratar essas imprecisões como parte de um erro, de uma variação, entre os dados reais e os previstos pelo modelo. Outra forma, a qual se fará presente ao longo deste texto, será o uso de conjuntos *fuzzy*, conceito introduzido por [ZADEH \(1965\)](#). Também é possível tratar essas imprecisões tratando os dados na forma de intervalos.

Para a construção de um modelo o pesquisador pode escolher usar métodos estatísticos de análise de regressão. Nesses métodos são estudados o comportamento da média dos dados juntamente da variação dos dados em torno dela. Atrelado à escolha desses métodos é necessário realizar também uma série de decisões em torno de dependência e independência entre os dados, decisões que se refletem na complexidade dos modelos ou em como se chega até eles. “Fundamentalmente, quando se procede a uma análise de dados, busca-se alguma forma de regularidade ou padrão ou, ainda, modelo, presente nas observações”([MORETTIN; BUSSAB, 2017](#)).

Outra característica dos dados que o método pode levar em consideração é quando temos mais de uma medição para cada indivíduo, gerando assim dados com medidas repetidas, os chamados dados longitudinais. Decidir tratar os dados como longitudinais traz ao modelo a possibilidade de entender melhor o comportamento isolado de um indivíduo ao longo de determinada situação, ao longo da unidade de tempo, espaço, temperatura, entre outras. Segundo [Fitzmaurice \(2011\)](#) é possível compreender as mudanças que ocorrem intra-indivíduo através de medidas repetidas dos indivíduos.

Este texto tem como objetivo trazer uma nova abordagem para problemas de regressão que trabalhem com dados que, por alguma razão, são imprecisos. Englobando tanto o caso em que todas as medições são independentes, caso transversal, quanto o caso onde os dados são longitudinais.

Na área de biomatemática usamos modelos matemáticos para explicar problemas envolvendo diversas áreas da biologia. Como exemplo motivador, usaremos o estudo apresentado por [Currie et al. \(2021\)](#). Este estudo usou vários métodos para medir o volume corporal de orcas-pigmeias (*Feresa attenuata*) ao longo de 17 dias nas águas costeiras da baía de Ma’alaea, no Maui, com o objetivo de entender o impacto da presença humana na busca por alimentos desses animais. Na Figura 2 são apresentados o volume corporal

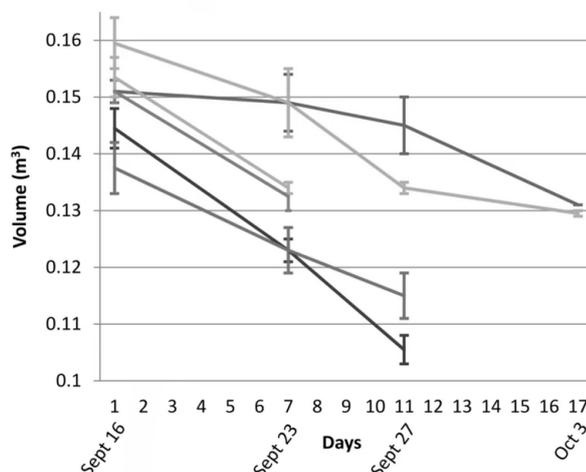
Figura 1 – Nem sempre é possível realizar as medições na natureza de forma direta.



Fonte: (Pixabay GmbH, n.d.) (adaptado)

de 6 diferentes orcas-pigmeias medidos em 4 diferentes ocasiões com diversas técnicas de medição nas águas costeiras da baía de Ma'alaea, no Maui, em setembro e outubro de 2019. Nesse exemplo temos dois indivíduos que tiveram seu volume medido nas 4 ocasiões, outros dois possuem três medidas enquanto outros dois possuem duas. Também é notório que a imprecisão das medidas, mensurada pelo desvio padrão, apresentado nas medidas varia bastante, ou seja, temos medidas bem precisas e outras com uma incerteza bem mais aparente. Essas diferenças se devem aos diferentes métodos usados para coletar essa medida. Usando imagens aéreas se tem medidas menos precisas, com uma imprecisão que pode passar de 10%. Enquanto as medidas *post-mortem*, tomadas de indivíduos encontrados mortos ao encalhar na costa, é bem mais precisa. Esse estudo aponta que esses animais

Figura 2 – Volume corporal e desvio padrão de 6 diferentes orcas-pigmeias medidos em 4 diferentes ocasiões com diversas técnicas de medição nas águas costeiras da baía de Ma'alaea, no Maui, em setembro e outubro de 2019.



Fonte: (CURRIE et al., 2021)

podem perder até 27% de seu volume corporal nesse período de 17 dias em que está sofrendo com a presença humana. Porém, podemos também criar modelos para tentar responder a este e a outros questionamentos:

- Como se comporta o volume médio da população de orcas-pigmeias ao longo desse período de tempo?
- Como se comporta o volume médio de cada orca-pigmeia ao longo desse período de tempo?
- Orcas-pigmeias com diferentes volumes corporais no início evoluem ao longo do tempo do mesmo modo? Ou o impacto é maior, ou menor dependendo do tamanho do animal ao chegar na área costeira?
- Modelar a imprecisão através de intervalos ou de números *fuzzy* oferece algum ganho para a interpretação dos dados?

## Parte I

### Revisão Bibliográfica

# 1 Regressão Linear

## 1.1 Introdução

O tema central desse trabalho é o conceito de análise de regressão. Temos como objetivo encontrar a relação entre uma variável resposta  $Y$  e certas variáveis explicativas. Podemos definir o processo de análise de regressão como (EVERITT, 2006):

Uma técnica estatística usada como base para estudar e caracterizar um sistema de interesse, formulando um modelo matemático razoável da relação entre uma variável resposta  $Y$  e um conjunto de  $q$  variáveis explicativas  $X_1, \dots, X_q$ . [...] De forma geral esses modelos podem ser considerados da forma  $y = f(\vartheta_1, \dots, \vartheta_q) + e$ , onde a função  $f$  reflete a relação verdadeira, porém, desconhecida entre  $Y$  e as variáveis explicativas. [...] O objetivo é formular uma função  $\hat{f}(\vartheta_1, \dots, \vartheta_q)$  que é uma aproximação razoável de  $f$ .

Everitt (2006) usa o termo modelo matemático, porém, pode ser mais apropriado o uso do termo modelo estatístico. Modelos na forma  $y = f(\vartheta_1, \dots, \vartheta_q) + e$  são mais corretamente chamados de modelos de regressão aditivos (BUJA; HASTIE; TIBSHIRANI, 1989). Também é possível denotar como  $Y = f(\vartheta_1, \dots, \vartheta_q) + e$ , reforçando o fato de  $Y$  ser uma variável aleatória.

Ao obter-se um modelo para análise de regressão é necessário realizar testes estatísticos para comprovar que o modelo obtido atende às hipóteses impostas sobre os dados.

Esse Capítulo trata da análise de regressão na forma de regressão linear. Usamos os conceitos de regressão linear simples e regressão linear múltipla.

Em ambos os métodos a média é apresentada através de uma função linear e é usada a hipótese de que os erros são independentes e identicamente distribuídos, ou seja, o conjunto dos erros  $e_i, i = 1, \dots, n$  tem valor esperado igual a zero e variância constante, denotamos por (FAHRMEIR et al., 2013):

$$E(e_i) = 0 \text{ e } Var(e_i) = \sigma^2, \text{ para todo } i = 1, \dots, n. \quad (1.1)$$

Para estimar os parâmetros do modelo usamos o processo de mínimos quadrados, onde o objetivo é minimizar a soma dos quadrados dos erros  $\sum_{i=1}^n e_i^2$  (FAHRMEIR et al., 2013). Mais detalhes são apresentados nas Seções 1.2 e 1.3.

Após a obter a estimativa do modelo  $\hat{f}$  não temos ainda garantia da qualidade do mesmo. Precisamos de “formas de avaliar se o modelo linear postulado é adequado ou não, dadas as suposições que fizemos sobre ele” (MORETTIN; BUSSAB, 2017). Para verificar se o modelo de fato é um bom modelo é necessário realizar testes. A Seção 1.4 apresenta o coeficiente de determinação que é a medida usada ao longo do texto para avaliar a qualidade dos modelos de regressão.

## 1.2 Regressão Linear Simples

Segundo Fahrmeir et al. (2013) a regressão linear simples busca encontrar uma função linear que modele a relação de duas variáveis aleatórias  $Y$  e  $X$ . Nesse trabalho chamamos  $Y$  de variável resposta e  $X$  de variável explicativa.

Na regressão linear simples, segundo Fahrmeir et al. (2013), dado um conjunto de observações de duas variáveis aleatórias  $Y$  e  $X$  na forma:

$$\{x_i; y_i\}, \quad i = 1, \dots, n. \quad (1.2)$$

Queremos encontrar o modelo linear:

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, \dots, n. \quad (1.3)$$

Levando em conta a hipótese dos erros serem independentes e identicamente distribuídos:

$$E(e_i) = 0 \text{ e } Var(e_i) = \sigma^2. \quad (1.4)$$

Estimados  $\beta_0$  e  $\beta_1$ , pelos valores  $\hat{\beta}_0$  e  $\hat{\beta}_1$ , respectivamente, obtemos o modelo  $f(\vartheta) = \hat{\beta}_0 + \hat{\beta}_1 \vartheta$  que estima o valor esperado de  $Y$  dado  $X = \vartheta$ . Ou seja:

$$\hat{E}(Y|X = \vartheta) = f(\vartheta) = \hat{\beta}_0 + \hat{\beta}_1 \vartheta. \quad (1.5)$$

Desta forma podemos prever  $Y$  através de  $f(\vartheta)$ . Denotamos por  $\hat{y}_i$  a estimativa do indivíduo  $i$ , ou seja,  $\hat{y}_i = \hat{f}(x_i)$ .

O modelo é construído encontrando  $\hat{\beta}_0$  e  $\hat{\beta}_1$  que minimizem a soma dos quadrados dos erros:

$$\sum_{i=0}^n e_i^2 = \sum_{i=0}^n (\hat{\beta}_0 + \hat{\beta}_1 x_i - y_i)^2. \quad (1.6)$$

Para realizar essa minimização de forma matricial, usamos a matriz de planejamento  $\mathbf{X}$  e o vetor  $\mathbf{y}$ :

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \quad (1.7)$$

e

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}. \quad (1.8)$$

Denotando

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix},$$

temos que

$$\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y} = \begin{pmatrix} \hat{\beta}_0 + \hat{\beta}_1 x_1 \\ \hat{\beta}_0 + \hat{\beta}_1 x_2 \\ \vdots \\ \hat{\beta}_0 + \hat{\beta}_1 x_n \end{pmatrix} - \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \hat{\beta}_0 + \hat{\beta}_1 x_1 - y_1 \\ \hat{\beta}_0 + \hat{\beta}_1 x_2 - y_2 \\ \vdots \\ \hat{\beta}_0 + \hat{\beta}_1 x_n - y_n \end{pmatrix}.$$

Assim podemos reescrever a Equação 1.6 como:

$$\sum_{i=0}^n e_i^2 = \sum_{i=0}^n (\hat{\beta}_0 + \hat{\beta}_1 x_i - y_i)^2 = (\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y})^T (\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y})$$

e expandir a expressão na forma:

$$\begin{aligned} (\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y})^T (\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y}) &= \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{y} - \mathbf{y}^T \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{y}^T \mathbf{y} \\ &= \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} - 2\hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{y} + \mathbf{y}^T \mathbf{y}. \end{aligned}$$

O objetivo é encontrar  $\hat{\boldsymbol{\beta}}$  que seja um ponto de mínimo de

$$s(\hat{\boldsymbol{\beta}}) = (\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y})^T (\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y}).$$

Segundo [Dineen \(2014\)](#), um ponto  $\hat{\boldsymbol{\beta}}$  será um ponto de mínimo quando o gradiente de  $s(\hat{\boldsymbol{\beta}})$ ,  $\nabla s(\hat{\boldsymbol{\beta}}) = \frac{\partial s(\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}}$  assumir o valor zero e sua matriz hessiana,  $\frac{\partial^2 s(\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}^2}$ , for uma matriz positiva definida.

Derivamos  $s(\hat{\boldsymbol{\beta}})$  em relação a  $\hat{\boldsymbol{\beta}}$  e igualamos a zero. Depois isolamos o  $\hat{\boldsymbol{\beta}}$  na equação.

$$\begin{aligned} \frac{\partial}{\partial \hat{\boldsymbol{\beta}}} \left( (\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y})^T (\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y}) \right) &= 0 \\ \frac{\partial}{\partial \hat{\boldsymbol{\beta}}} \left( \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} - 2\hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{y} + \mathbf{y}^T \mathbf{y} \right) &= 0 \\ 2\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} - 2\mathbf{X}^T \mathbf{y} + 0 &= 0 \\ 2\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} &= 2\mathbf{X}^T \mathbf{y} \\ \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} &= \mathbf{X}^T \mathbf{y} \\ \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \end{aligned}$$

Assim obtemos (FAHRMEIR et al., 2013):

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (1.9)$$

Para provar que se trata realmente de um ponto de mínimo, verificamos a matriz Hessiana:

$$\begin{aligned} \frac{\partial^2 s(\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}} \partial \hat{\boldsymbol{\beta}}} &= \\ \frac{\partial}{\partial \hat{\boldsymbol{\beta}}} \frac{\partial s(\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} &= \\ \frac{\partial}{\partial \hat{\boldsymbol{\beta}}} (2\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} - 2\mathbf{X}^T \mathbf{y}) &= 2\mathbf{X}^T \mathbf{X}, \end{aligned}$$

matriz que de fato é positiva definida. Tome um vetor  $\mathbf{w} \neq 0$ , como  $\mathbf{X}$  tem posto completo temos que  $\mathbf{X}\mathbf{w} \neq 0$  e também que:

$$\mathbf{w}^T (2\mathbf{X}^T \mathbf{X}) \mathbf{w} = 2 (\mathbf{w}^T \mathbf{X}^T \mathbf{X} \mathbf{w}) = 2 (\mathbf{X}\mathbf{w})^T (\mathbf{X}\mathbf{w}) > 0.$$

Provando assim que o  $\hat{\boldsymbol{\beta}}$  dado na Equação 1.9 é ponto de mínimo.

### 1.3 Regressão Linear Múltipla

O modelo de regressão linear múltipla é bastante similar ao modelo linear simples, a diferença mais visível é que no lugar de uma variável explicativa  $X$  teremos  $q$  variáveis explicativas. Segundo Fahrmeir et al. (2013) a regressão linear múltipla busca encontrar uma função linear que modele a relação entre uma variável resposta  $Y$  e  $q$  variáveis explicativas  $X_1, \dots, X_q$ .

A regressão linear múltipla, segundo Fahrmeir et al. (2013), dado um conjunto de observações da variável resposta  $Y$  e das variáveis explicativas  $X_1, \dots, X_q$  na forma:

$$\{x_{i,1}; \dots; x_{i,q}; y_i\}, \quad i = 1, \dots, n. \quad (1.10)$$

Queremos encontrar o modelo linear:

$$Y = \beta_0 + \beta_1 x_{i,1} + \dots + \beta_q x_{i,q} + e, \quad i = 1, \dots, n. \quad (1.11)$$

Deste modo temos:

$$y_i = \beta_0 + \beta_1 x_{i,1} + \dots + \beta_q x_{i,q} + e_i, \quad i = 1, \dots, n. \quad (1.12)$$

Levando em conta a hipótese dos erros serem independentes e identicamente distribuídos:

$$E(e_i) = 0 \text{ e } Var(e_i) = \sigma^2. \quad (1.13)$$

Estimados os  $\beta_i, i = 0, \dots, n$ , pelos valores  $\hat{\beta}_i$ , obtemos o modelo  $f(\vartheta_1, \dots, \vartheta_q) = \hat{\beta}_0 + \hat{\beta}_1\vartheta_1 + \dots + \hat{\beta}_q\vartheta_q$  que estima o valor esperado de  $Y$  dados  $X_1 = \vartheta_1, \dots, X_q = \vartheta_q$ . Ou seja:

$$\hat{E}(Y|X_1 = \vartheta_1, \dots, X_q = \vartheta_q) = f(\vartheta_1, \dots, \vartheta_q) = \hat{\beta}_0 + \hat{\beta}_1\vartheta_1 + \dots + \hat{\beta}_q\vartheta_q. \quad (1.14)$$

Desta forma podemos prever  $Y$  através de  $f(\vartheta_1, \dots, \vartheta_q)$ . Denotamos por  $\hat{y}_i$  a estimativa do indivíduo  $i$ , ou seja,  $\hat{y}_i = \hat{f}(x_{i,1}, \dots, x_{i,q})$ .

O modelo é estimado encontrando  $\beta_i, i = 1, \dots, q$ , que minimizem a soma dos quadrados dos erros:

$$\sum_{i=0}^n e_i^2 = \sum_{i=0}^n (\hat{\beta}_0 + \hat{\beta}_1 x_{i,1} + \dots + \hat{\beta}_q x_{i,q} - y_i)^2. \quad (1.15)$$

Para realizar essa minimização, usamos a matriz de planejamento  $\mathbf{X}$  e o vetor  $\mathbf{y}$ :

$$\mathbf{X} = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{1,q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \dots & x_{n,q} \end{pmatrix} \quad (1.16)$$

e

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}. \quad (1.17)$$

O objetivo é minimizar

$$\sum_{i=0}^n e_i^2 = \sum_{i=0}^n (\hat{\beta}_0 + \hat{\beta}_1 x_{i,1} + \dots + \hat{\beta}_q x_{i,q} - y_i)^2 = (\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y})^T (\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y}).$$

Do mesmo modo visto na Seção 1.2, calculamos (FAHRMEIR et al., 2013):

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_q \end{pmatrix} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (1.18)$$

## 1.4 Avaliação dos Modelos

Após se construir um modelo de regressão para um conjunto de dados  $y_1, \dots, y_n$ , com previsões  $\hat{y}_1, \dots, \hat{y}_n$ , definimos o coeficiente de determinação (MORETTIN; BUSSAB, 2017):

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (1.19)$$

onde  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ . O valor  $0 \leq R^2 \leq 1$  estima a proporção da variação de  $Y$  que é explicada pelas variáveis explicativas. Um valor de  $R^2$  mais próximo de 1 indica melhor qualidade do modelo.

Também podemos analisar se as hipóteses feitas sobre os resíduos estão sendo atendidas pelo modelo. É necessário analisar se a hipótese da média dos erros ser nula com variância constante é válida, em outras palavras, se  $E(e_i) = 0$  e  $Var(e_i) = \sigma^2$ . Segundo Dodge (2009), o gráfico quantil-quantil pode ser usado para comparar a distribuição de um conjunto de dados com uma distribuição específica. Desta forma verificamos a normalidade do conjunto formado pelos erros  $e_i = \hat{y}_i - y_i$ .

Como pode ser visto em Gutiérrez e Aguilera (2010), esse processo consiste em ordenar o conjunto dos erros em ordem crescente, obtendo o conjunto  $a_1 \leq a_2 \leq \dots \leq a_n$ . Tomamos também o conjunto com  $n$  quantis da distribuição normal com média  $\mu$  e desvio padrão  $\sigma$ ,  $b_1 \leq b_2 \leq \dots \leq b_n$ . Esses quantis são obtidos através da inversa da função de distribuição acumulada,  $F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right) dt$ , assim cada  $b_i$  é dado por  $b_i = F^{-1}\left(\frac{i - \frac{1}{2}}{n}\right)$ .

Montamos então o gráfico quantil-quantil com os pontos  $(a_i; b_i)$  juntamente da reta  $y = x$ . Se os erros  $e_i$  estão normalmente distribuídos, então os pontos  $(a_i; b_i)$  se encontram próximos da reta  $y = x$  (DODGE, 2009). Como temos por hipótese que os erros estão normalmente distribuídos com média em zero, temos que  $\mu = 0$ . Ainda precisamos de um valor para  $\sigma$ , propomos usar o desvio padrão do conjunto dos erros, dado por  $\sigma = \sqrt{\frac{\sum_{i=1}^n (e_i - M)^2}{n-1}}$ , com  $M = \frac{\sum_{i=1}^n e_i}{n}$  (DODGE, 2009).

## 1.5 Conclusão

Os modelos de regressão linear oferecem uma ferramenta simples para realizar a inferência sobre dados que pode ser resumida em construir a matriz de planejamento para o modelo, encontrar  $\hat{\beta}$  através do processo de quadrados mínimos, calcular o valor dos erros  $e_i = \hat{y}_i - y_i$  e validar a normalidade desse conjunto de erros.

Validar a normalidade dos erros é o que separa a regressão linear de técnicas de ajuste de curvas, sendo assim uma parte importante do processo.

## 2 Dados Longitudinais

### 2.1 Introdução

O capítulo 1 apresenta métodos de regressão para dados onde todas as medidas são independentes. Esses tipos de dados são chamados na literatura como dados transversais, como pode ser visto em [Davidian \(1995\)](#), [Crowder \(1990\)](#) e [Fitzmaurice \(2011\)](#). Esses mesmos autores, dentre outros, definem o que são dados longitudinais, ou também chamados de dados com medidas repetidas.

Dados longitudinais organizam os dados observados em medidas repetidas de um mesmo indivíduo, permitindo construir modelos que levam em conta que essas medidas tem essa relação.

### 2.2 Dados Longitudinais

Segundo [Everitt \(2006\)](#), dados longitudinais tratam de quando cada indivíduo representa uma variável observada em diferentes instantes de tempo<sup>1</sup>. Dessa forma, as medidas de cada indivíduo são representadas por um vetor que contém as diferentes observações deste indivíduo em diferentes pontos do tempo. Denotamos:

$$\mathbf{Y}_i = \begin{pmatrix} Y_{i,1} \\ Y_{i,2} \\ \vdots \\ Y_{i,m} \end{pmatrix}_{m_i \times 1} \quad (2.1)$$

Em que  $\mathbf{Y}_i$  é chamado de indivíduo  $i$  ou  $i$ -ésimo indivíduo e cada  $Y_{i,j}$  é a medida desse indivíduo no instante  $t_j$ . O conjunto de tempos  $t_1, \dots, t_m$  é fixo. O valor  $m_i$  é o número de linhas do vetor  $\mathbf{Y}_i$ . O indivíduo não precisa ter observações em todos os tempos  $t_1, \dots, t_m$ , alguns podem estar ausentes, isso faz o valor  $m_i$  variar para cada indivíduo.

Por exemplo, podemos ter o indivíduo  $\mathbf{Y}_3$  com as medidas tomadas nos tempos  $t_1, \dots, t_5$ :

$$\mathbf{Y}_3 = \begin{pmatrix} Y_{3,1} \\ Y_{3,2} \\ Y_{3,3} \\ Y_{3,4} \\ Y_{3,5} \end{pmatrix}, \quad (2.2)$$

<sup>1</sup> outras grandezas podem ser usadas além do tempo

onde  $m_3 = 5$ .

Agora, pode ocorrer de um indivíduo  $\mathbf{Y}_i$  não possuir medida num instante  $t_j$ , assim a medida  $Y_{i,j}$  não existe. Por exemplo, seja o indivíduo  $\mathbf{Y}_7$ , com as medidas tomadas nos tempos  $t_1, \dots, t_5$ . Caso a medida para o tempo  $t_2$  não exista, temos:

$$\mathbf{Y}_7 = \begin{pmatrix} Y_{7,1} \\ Y_{7,3} \\ Y_{7,4} \\ Y_{7,5} \end{pmatrix}, \quad (2.3)$$

onde  $m_7 = 4$ . Também podemos ter o indivíduo  $\mathbf{Y}_6$  sem as medidas em  $t_1, t_3$  e  $t_5$ . Neste caso:

$$\mathbf{Y}_6 = \begin{pmatrix} Y_{6,2} \\ Y_{6,4} \end{pmatrix}, \quad (2.4)$$

onde  $m_6 = 2$ .

Conjuntos de dados onde alguns indivíduos tem dados faltantes é chamado de um conjunto desbalanceado. Em dados de problemas reais é comum que dados estejam faltando, por isso a importância de deixar claro a estrutura da representação dos mesmos.

Dados longitudinais são caracterizados por duas medidas obtidas de um mesmo indivíduo poderem ser correlacionadas, ou seja, a correlação pode ser não nula entre as medidas  $Y_{i,j}$  e  $Y_{i,k}$  com  $j \neq k$ . Pelo fato de termos observações  $Y_{i,j}$  realizadas sobre um mesmo indivíduo, como esclarece (FITZMAURICE, 2011):

Quando temos  $m$  medidas repetidas sobre um mesmo indivíduo, não devemos assumir que essas medidas repetidas sejam independentes. Resulta disso que deve ser considerada a distribuição de probabilidade conjunta para o vetor de medidas repetidas. Observe, no entanto, que os vetores de medidas repetidas são assumidos como sendo independentes uns dos outros.

Para dados longitudinais assumimos que as variáveis aleatórias  $\mathbf{Y}_i$  são independentes e normalmente distribuídas em uma distribuição de probabilidade conjunta. Podemos assumir que essa distribuição é a distribuição normal multivariada com média  $\boldsymbol{\mu}_i \in \mathbb{R}^m$  e matriz de variância-covariância  $\boldsymbol{\Sigma}_i$  de dimensão  $m_i \times m_i$ . Denotamos por  $\mathbf{Y}_i \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ . Adotamos aqui essa abordagem, porém, existem várias outras opções para tal.

Fitzmaurice (2011) apresenta exemplos de aplicações na área da saúde, como relacionar a capacidade pulmonar ao longo do tempo de crianças com os índices de poluição das cidades, estudar o aumento do índice de gordura corporal em meninas após a primeira menstruação e em ensaios clínicos controlados aleatórios para o teste de medicamentos.

## 2.3 Modelo Linear Misto em Dois Estágios

Segundo [Fitzmaurice \(2011\)](#), o modelo linear misto em dois estágios parte da premissa de que certo subconjunto de parâmetros no modelo variam de acordo com cada indivíduo. Deste modo cada indivíduo terá o seu próprio modelo para a média, formado por dois termos, sendo um efeito fixo, que modela características pertencentes à população por inteiro, e um efeito aleatório que modela as características de cada indivíduo:

$$\mathbf{Y}_i = \underbrace{\mathbf{X}_i \boldsymbol{\beta}}_{\text{efeito fixo}} + \underbrace{\mathbf{Z}_i \mathbf{b}_i}_{\text{efeito aleatório}} + \underbrace{\mathbf{e}_i}_{\text{erro}}, \quad (2.5)$$

onde  $\mathbf{X}_i$  é a matriz de planejamento  $\mathbf{X}$  sem as linhas correspondentes aos tempos em que as observações não existem para  $\mathbf{Y}_i$ .  $\boldsymbol{\beta}$  é um vetor coluna com os coeficientes do modelo linear da média da população, esse vetor tem o número de linhas igual ao número de colunas de  $\mathbf{X}_i$ .  $\mathbf{Z}_i$  é uma outra matriz de planejamento que modela o efeito individual que o indivíduo  $\mathbf{Y}_i$  tem além da média populacional.  $\mathbf{b}_i$  é um vetor coluna com os coeficientes do modelo linear do efeito aleatório do indivíduo  $\mathbf{Y}_i$ , esse vetor tem o número de linhas igual ao número de colunas de  $\mathbf{Z}_i$  e  $\mathbf{e}_i$  é um vetor coluna em que cada elemento contém o erro entre as medidas e o valor correspondente a estas no modelo.

O efeito fixo é a média da população para cada instante de tempo, denotamos:

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}.$$

Nesse modelo, cada indivíduo  $i$  com valores observados  $\mathbf{Y}_i$  tem seu valor esperado expresso por uma função  $f_i(t)$ , com a qual é possível realizar o processo de predição para este indivíduo em específico. Por exemplo, com:

$$\mathbf{X} = \begin{pmatrix} 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_m & t_m^2 \end{pmatrix} \text{ e } \mathbf{Z} = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{pmatrix}, \quad (2.6)$$

teríamos:

$$f_i(t) = \begin{pmatrix} 1 & t & t^2 \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} 1 & t \end{pmatrix} \mathbf{b}_i = \beta_0 + \beta_1 t + \beta_2 t^2 + b_{i,0} + b_{i,1} t. \quad (2.7)$$

De forma geral, denotamos a função de inferência como:

$$f_i(t) = \mathbf{g}_x(t)^T \boldsymbol{\beta} + \mathbf{g}_z(t)^T \mathbf{b}_i. \quad (2.8)$$

Em que a  $i$ -ésima linha de  $\mathbf{X}$  é dada por  $\mathbf{g}_x(t_i)^T$  e a  $i$ -ésima linha de  $\mathbf{Z}$  é dada por  $\mathbf{g}_z(t_i)^T$ . As matrizes de planejamento  $\mathbf{X}$  e  $\mathbf{Z}$  podem possuir tantas colunas quanto o

pesquisador ache necessário para o modelo. Nesse trabalho assumimos que os elementos da primeira coluna são sempre 1 e que as colunas formam um conjunto linearmente independente.

Encontrar as funções  $f_i(t)$  servem como inferência sobre o valor esperado para o indivíduo  $i$  com as observações  $\mathbf{Y}_i$ . O valor  $f_i(t)$  dá o valor esperado para o indivíduo  $i$  no tempo  $t$ .

Apresentamos aqui um método para estimar os parâmetros desse modelo como exemplo. Outros métodos com outras suposições e outras técnicas numéricas existem na literatura.

Seja um conjunto longitudinal com  $n$  indivíduos com observações nos tempos  $t_1, \dots, t_k$ . Queremos encontrar estimativas  $\hat{\boldsymbol{\beta}}$  e  $\hat{\mathbf{b}}_1, \dots, \hat{\mathbf{b}}_n$  para o modelo:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i. \quad (2.9)$$

Este modelo pressupõe que os valores observados para os indivíduos estão identicamente distribuídos na distribuição normal:

$$\mathbf{Y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i),$$

com média  $\mathbf{X}_i\boldsymbol{\beta}$  e matriz de variância-covariância  $\boldsymbol{\Sigma}_i = \sigma^2\mathbf{I}_{m_i \times m_i} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T$ .

Os valores  $\sigma^2$  e  $\mathbf{D}$  podem ser conhecidos previamente. Nesse caso obtemos as estimativas para  $\boldsymbol{\beta}$  e  $\mathbf{b}_i$  (LAIRD; WARE, 1982):

$$\hat{\boldsymbol{\beta}} = \sum_{i=1}^n (\mathbf{X}_i^T \boldsymbol{\Sigma}_i \mathbf{X}_i)^{-1} \sum_{i=1}^n \mathbf{X}_i^T \boldsymbol{\Sigma}_i \mathbf{Y}_i \quad (2.10)$$

e

$$\hat{\mathbf{b}}_i = \mathbf{D}\mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}). \quad (2.11)$$

Caso esses valores não sejam conhecidos é possível encontrá-los pelo processo de máxima verossimilhança. Encontramos as estimativas  $\hat{\boldsymbol{\beta}}$ ,  $\hat{\sigma}^2$  e  $\hat{\mathbf{D}}$  usando a log-máxima verossimilhança (HARVILLE, 1977), (WU, 2009):

$$\arg \max_{\boldsymbol{\beta}, \sigma^2, \mathbf{D}} \left( -\frac{1}{2} \sum_{i=1}^n \left( \log(\det(2\pi\boldsymbol{\Sigma}_i)) + (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}) \right) \right), \quad (2.12)$$

com a restrição de que  $\hat{\sigma}^2 > 0$  e que  $\hat{\mathbf{D}}$  seja uma matriz positiva definida. Com esses valores os  $\hat{\mathbf{b}}_i$  podem ser encontrados pela Equação 2.11.

## 2.4 Exemplos

**Exemplo 1.** *Seja o conjunto de dados longitudinais:*

$$Y_1 = \begin{pmatrix} 10 \\ 14 \\ 19 \\ 25 \end{pmatrix}, Y_2 = \begin{pmatrix} 9 \\ 12 \\ 18 \\ 22 \end{pmatrix}, Y_3 = \begin{pmatrix} 6 \\ 8 \\ 10 \\ 12 \end{pmatrix}, Y_4 = \begin{pmatrix} 4 \\ 7 \\ 9 \\ 11 \end{pmatrix} \text{ e } Y_5 = \begin{pmatrix} 2 \\ 5 \\ 8 \\ 10 \end{pmatrix}, \quad (2.13)$$

com as medidas tomadas nos tempos  $t_1 = 1$ ,  $t_2 = 2$ ,  $t_3 = 3$  e  $t_4 = 4$ .

Como todos os indivíduos possuem todas as medidas, podemos usar os mesmos  $\mathbf{X}$  e  $\mathbf{Z}$  para todos eles, ou seja,  $\mathbf{X}_1 = \mathbf{X}_2 = \mathbf{X}_3 = \mathbf{X}_4 = \mathbf{X}$  e  $\mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{Z}_3 = \mathbf{Z}_4 = \mathbf{Z}$ .

O modelo pode tomar muitas formas distintas, dependendo de quais matrizes de planejamento  $\mathbf{X}$  e  $\mathbf{Z}$  sejam escolhidas, porém, quais são as matrizes mais apropriadas para cada caso é decisão de quem está construindo o modelo.

Para ilustrar o modelo apresentamos aqui duas situações distintas.

Na primeira situação, tomamos:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \text{ e } \mathbf{Z} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}. \quad (2.14)$$

Neste caso a matriz  $\mathbf{D}$  assumirá a forma de uma matriz com uma linha e uma coluna:

$$\mathbf{D} = \begin{pmatrix} d_{1,1} \end{pmatrix}.$$

A matriz de variância-covariância assume assim a forma:

$$\begin{aligned} \Sigma &= \sigma^2 \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} (d_{1,1}) \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} d_{1,1} + \sigma^2 & d_{1,1} & d_{1,1} & d_{1,1} \\ d_{1,1} & d_{1,1} + \sigma^2 & d_{1,1} & d_{1,1} \\ d_{1,1} & d_{1,1} & d_{1,1} + \sigma^2 & d_{1,1} \\ d_{1,1} & d_{1,1} & d_{1,1} & d_{1,1} + \sigma^2 \end{pmatrix}. \end{aligned}$$

Usando a Equação 2.12 encontramos:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} 2,80 \\ 3,30 \end{pmatrix},$$

$$\hat{\sigma}^2 = 2,67 \text{ e } \hat{D} = (17,56).$$

E com a Equação 2.11:

$$\hat{\mathbf{b}}_1 = (\hat{b}_{1,0}) = (5,73), \hat{\mathbf{b}}_2 = (\hat{b}_{2,0}) = (4,05), \hat{\mathbf{b}}_3 = (\hat{b}_{3,0}) = (-1,98),$$

$$\hat{\mathbf{b}}_4 = (\hat{b}_{4,0}) = (-3,18) \text{ e } \hat{\mathbf{b}}_5 = (\hat{b}_{5,0}) = (-4,62).$$

O gráfico referente a essa primeira escolha está presente na Figura 3(a). Note como os gráficos de  $f_i(t) = 2,80 + 3,30t + \hat{b}_{i,0}$  são paralelos. Aqui temos um modelo com intercepto 2,80 e coeficiente angular 3,30 para o valor esperado da população e cada indivíduo tem como efeito individual uma alteração desse intercepto, por exemplo, o indivíduo 1 tem um intercepto 5,73 unidades maior que o intercepto da população.

$$\begin{aligned} f_1(t) &= 2,80 + 3,30t && +5,73 \\ f_2(t) &= 2,80 + 3,30t && +4,05 \\ f_3(t) &= 2,80 + 3,30t && -1,98 \\ f_4(t) &= 2,80 + 3,30t && -3,18 \\ f_5(t) &= 2,80 + 3,30t && -4,62 \end{aligned} \tag{2.15}$$

$\underbrace{\hspace{10em}}$ 
 $\underbrace{\hspace{10em}}$

**efeito**
**efeito**

**fixo**
**aleatório**

Na segunda situação, adotamos:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \text{ e } \mathbf{Z} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix}. \tag{2.16}$$

Neste caso a matriz  $D$  assumirá a forma de uma matriz com duas linhas e duas colunas:

$$D = \begin{pmatrix} d_{1,1} & d_{1,2} \\ d_{2,1} & d_{2,2} \end{pmatrix}.$$

A matriz de variância-covariância assume assim a forma:

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} d_{1,1} & d_{1,2} \\ d_{2,1} & d_{2,2} \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{pmatrix}.$$

Usando a Equação 2.12 encontramos:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} 2,80 \\ 3,30 \end{pmatrix},$$

$$\hat{\sigma}^2 = 0,31 \text{ e } \hat{D} = \begin{pmatrix} 2,99 & 1,27 \\ 1,27 & 1,41 \end{pmatrix}.$$

E com a Equação 2.11:

$$\hat{\mathbf{b}}_1 = \begin{pmatrix} \hat{b}_{1,0} \\ \hat{b}_{1,1} \end{pmatrix} = \begin{pmatrix} 1,79 \\ 1,66 \end{pmatrix}, \quad \hat{\mathbf{b}}_2 = \begin{pmatrix} \hat{b}_{2,0} \\ \hat{b}_{2,1} \end{pmatrix} = \begin{pmatrix} 1,26 \\ 1,17 \end{pmatrix}, \quad \hat{\mathbf{b}}_3 = \begin{pmatrix} \hat{b}_{3,0} \\ \hat{b}_{3,1} \end{pmatrix} = \begin{pmatrix} 0,58 \\ -1,08 \end{pmatrix},$$

$$\hat{\mathbf{b}}_4 = \begin{pmatrix} \hat{b}_{4,0} \\ \hat{b}_{4,1} \end{pmatrix} = \begin{pmatrix} -0,89 \\ -0,96 \end{pmatrix} \text{ e } \hat{\mathbf{b}}_5 = \begin{pmatrix} \hat{b}_{5,0} \\ \hat{b}_{5,1} \end{pmatrix} = \begin{pmatrix} -2,73 \\ -0,79 \end{pmatrix}.$$

O gráfico referente a essa segunda escolha está presente na Figura 3(b). Note como os gráficos de  $f_i(t) = 2,80 + 3,30t + \hat{b}_{i,0} + \hat{b}_{i,1}t$  não são paralelos. Aqui temos um modelo com intercepto 2,80 e coeficiente angular 3,30 para o valor esperado da população e cada indivíduo tem como efeito individual uma alteração no intercepto e no coeficiente angular, por exemplo, o indivíduo 1 tem um intercepto 1,79 unidades maior que o intercepto da população com o crescimento dado pelo coeficiente angular 1,66 unidades maior que o da população.

$$\begin{aligned} f_1(t) &= 2,80 + 3,30t + 1,79 + 1,66t \\ f_2(t) &= 2,80 + 3,30t + 1,26 + 1,17t \\ f_3(t) &= 2,80 + 3,30t + 0,58 - 1,08t \\ f_4(t) &= 2,80 + 3,30t - 0,89 - 0,96t \\ f_5(t) &= \underbrace{2,80 + 3,30t}_{\text{efeito fixo}} + \underbrace{-2,73 - 0,79t}_{\text{efeito aleatório}} \end{aligned} \quad (2.17)$$

Uma dúvida que pode surgir é se todo esse processo oferece um resultado diferente do processo, que seria bem mais simples, de realizar a regressão linear por

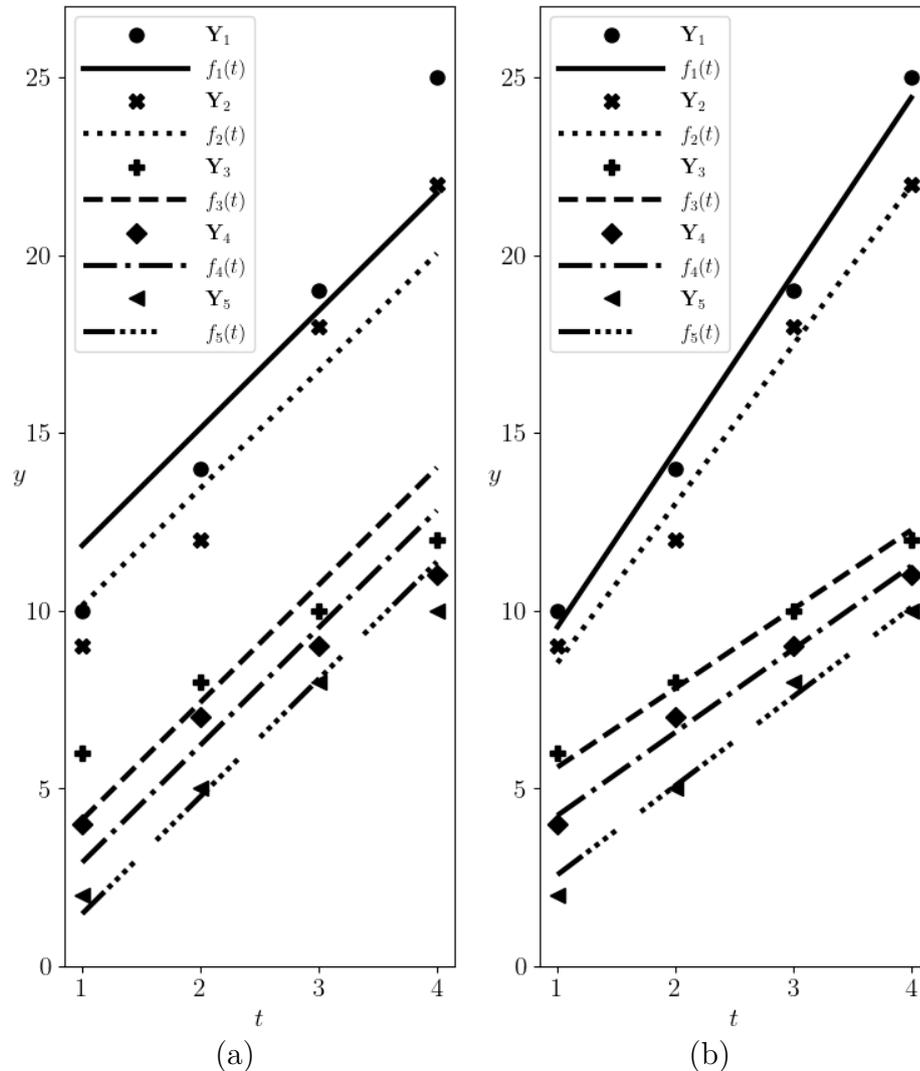
quadrados mínimos separadamente para cada indivíduo. A resposta para essa dúvida é que o resultado é sim diferente.

Neste caso teríamos como resultado as retas:

$$\begin{aligned}qm_1(t) &= 4,5 + 5,0t \\qm_2(t) &= 4,0 + 4,5t \\qm_3(t) &= 4,0 + 2,0t \\qm_4(t) &= 2,0 + 2,3t \\qm_5(t) &= -0,5 + 2,7t,\end{aligned}$$

que é um resultado diferente dos dois casos apresentados no exemplo. Isso se deve ao fato de que realizando o processo de quadrados mínimos separadamente não se leva em conta o efeito do comportamento da população na modelagem de cada indivíduo. Exibindo assim um pouco da importância e do ganho ao se realizar a regressão de dois estágios para o estudo de dados longitudinais.

Figura 3 – Gráficos do Exemplo 1. As pequenas figuras geométricas (  $\bullet$ ,  $\times$ ,  $+$ ,  $\blacklozenge$  e  $\blacktriangleleft$  ) apresentam os pontos referentes aos valores observados dos indivíduos em cada tempo  $t_j$ . As retas  $f_i(t)$  apresentam a reta estimada do valor esperado dos indivíduos em cada tempo  $t$ .



Fonte: O autor.

## 2.5 Conclusão

Diferente do Capítulo 1, podemos tratar os dados como sendo medidas repetidas de um mesmo indivíduo. O modelo linear misto em dois estágios, Seção 2.3, não somente leva em conta a existência de medidas repetidas como também permite a construção de um modelo para a inferência de valores para cada indivíduo.

Ainda mais, o modelo de regressão em dois estágios oferece modelos para cada indivíduo que levam em conta o comportamento da população como um todo. Isso permite realizar melhores previsões para indivíduos que tenham dados faltantes, além de várias outras vantagens apresentadas por [Fitzmaurice \(2011\)](#).

## 3 Aritmética Intervalar

### 3.1 Introdução

Nesse trabalho denotamos os intervalos fechados reais como símbolos com uma barra superior e inferior. Por exemplo, em

$$\bar{x} = [a; b], \quad (3.1)$$

$\bar{x}$  denota o intervalo real  $[a; b]$ .

Barros, Bassanezi e Lodwick (2017) exhibe o conceito de aritmética intervalar, onde temos operações de soma, subtração, multiplicação e divisão para intervalos. Esse capítulo exhibe esse conceito.

### 3.2 Operações Aritméticas entre Intervalos

Definimos o conceito de aritmética intervalar, onde temos operações de soma, subtração, multiplicação, multiplicação por escalar e divisão para intervalos.

Barros, Bassanezi e Lodwick (2017) define as operações entre intervalos reais. Sejam o número real  $\lambda$  e os intervalos  $\bar{a} = [a_1; a_2]$  e  $\bar{b} = [b_1; b_2]$ , definimos as operações:

- $\bar{a} + \bar{b} = [a_1 + b_1; a_2 + b_2]$ .
- $\bar{a} - \bar{b} = [a_1 - b_2; a_2 - b_1]$ .
- $\lambda \bar{a} = \begin{cases} [\lambda a_1; \lambda a_2], & \text{se } \lambda \geq 0 \\ [\lambda a_2; \lambda a_1], & \text{se } \lambda < 0 \end{cases}$ .
- $\bar{a}\bar{b} = [\min P; \max P]$ , onde  $P = \{a_1 b_1, a_1 b_2, a_2 b_1, a_2 b_2\}$ .
- $\bar{a}/\bar{b} = [a_1; a_2] \left[ \frac{1}{b_2}; \frac{1}{b_1} \right]$ , com  $0 \notin \bar{b}$ .

**Exemplo 2.** *Alguns exemplos de soma de intervalos:*

- $[1; 2] + [3; 4] = [(1) + (3); (2) + (4)] = [4; 6]$ .
- $[2; 5] + [-1; 1] = [(2) + (-1); (5) + (1)] = [1; 6]$ .
- $[-3; -2] + [7; 15] = [(-3) + (7); (-2) + (15)] = [4; 13]$ .
- $[-5; -3] + [-2; -1] = [(-5) + (-2); (-3) + (-1)] = [-7; -4]$ .

**Exemplo 3.** *Alguns exemplos de subtração de intervalos:*

- $[1; 2] - [3; 4] = [(1) - (4); (2) - (3)] = [-3; -1]$ .
- $[2; 5] - [-1; 1] = [(2) - (1); (5) - (-1)] = [1; 6]$ .
- $[-3; -2] - [7; 15] = [(-3) - (15); (-2) - (7)] = [-18; -9]$ .
- $[-5; -3] - [-2; -1] = [(-5) - (-1); (-3) - (-2)] = [-4; -1]$ .
- $[0; 1] - [0; 1] = [(0) - (1); (-3) - (-2)] = [-4; -1]$ .
- $[1; 2] - [1; 2] = [(1) - (2); (2) - (1)] = [-1; 1]$ .
- $[-5; 5] - [-5; 5] = [(-5) - (5); (5) - (-5)] = [-10; 10]$ .
- $[-3; -1] - [-3; -1] = [(-3) - (-1); (-1) - (-3)] = [-2; 2]$ .

**Exemplo 4.** *Alguns exemplos de multiplicação de intervalos por escalar:*

- $1[1; 2] = [1(1); 1(2)] = [1; 2]$ .
- $2[1; 2] = [2(1); 2(2)] = [2; 4]$ .
- $-3[1; 2] = [-3(2); -3(1)] = [-6; -3]$ .
- $0[1; 2] = [0(1); 0(2)] = [0; 0]$ .
- $1[-1; 1] = [1(-1); 1(1)] = [-1; 1]$ .
- $2[-1; 1] = [2(-1); 2(1)] = [-2; 2]$ .
- $-3[-1; 1] = [-3(1); -3(-1)] = [-3; 3]$ .
- $0[-1; 1] = [0(-1); 0(1)] = [0; 0]$ .
- $1[-3; -2] = [1(-3); 1(-2)] = [-3; -2]$ .
- $2[-3; -2] = [2(-3); 2(-2)] = [-6; -4]$ .
- $-3[-3; -2] = [-3(-2); -3(-3)] = [6; 9]$ .
- $0[-3; -2] = [0(-3); 0(-2)] = [0; 0]$ .

**Exemplo 5.** *Alguns exemplos de multiplicação de intervalos:*

- $[1; 2][3; 4] = [\min P, \max P] = [3; 8]$ , onde:  
 $P = \{(1) (3), (1) (4), (2) (3), (2) (4)\} = \{3, 4, 6, 8\}$ .

- $[2; 5][-1; 1] = [\min P, \max P] = [-5; 5]$ , onde:  
 $P = \{(2) (-1), (2) (1), (5) (-1), (5) (1)\} = \{-2, 2, -5, 5\}$ .
- $[-3; -2][7; 15] = [\min P, \max P] = [-45; -14]$ , onde:  
 $P = \{(-3) (7), (-3) (15), (-2) (7), (-2) (15)\} = \{-21, -45, -14, -30\}$ .
- $[-5; -3][-2; -1] = [\min P, \max P] = [3; 10]$ , onde:  
 $P = \{(-5) (-2), (-5) (-1), (-3) (-2), (-3) (-1)\} = \{10, 5, 6, 3\}$ .
- $[-1; 3][-1; 3] = [\min P, \max P] = [-3; 9]$ , onde:  
 $P = \{(-1) (-1), (-1) (3), (3) (-1), (3) (3)\} = \{1, -3, -3, 9\}$ .

**Exemplo 6.** Alguns exemplos de divisão de intervalos:

- $[1; 2]/[3; 4] = [1; 2] \left[ \frac{1}{4}; \frac{1}{3} \right] = [\min P, \max P] = \left[ \frac{1}{4}; \frac{2}{3} \right]$ , onde:  
 $P = \left\{ (1) \left( \frac{1}{4} \right), (1) \left( \frac{1}{3} \right), (2) \left( \frac{1}{4} \right), (2) \left( \frac{1}{3} \right) \right\} = \left\{ \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3} \right\}$ .
- $[2; 5]/[-1; 1]$  não é possível ser realizado, pois  $0 \in [-1; 1]$ .
- $[-3; -2]/[7; 15] = [-3; -2] \left[ \frac{1}{15}; \frac{1}{7} \right] = [\min P, \max P] = \left[ -\frac{3}{7}; -\frac{2}{15} \right]$ , onde:  
 $P = \left\{ (-3) \left( \frac{1}{15} \right), (-3) \left( \frac{1}{7} \right), (-2) \left( \frac{1}{15} \right), (-2) \left( \frac{1}{7} \right) \right\} =$   
 $= \left\{ -\frac{1}{5}, -\frac{3}{7}, -\frac{2}{15}, -\frac{2}{7} \right\}$ .
- $[-5; -3]/[-2; -1] = [-5; -3] \left[ -1; -\frac{1}{2} \right] = [\min P, \max P] = \left[ \frac{3}{2}; 5 \right]$ , onde:  
 $P = \left\{ (-5) (-1), (-5) \left( -\frac{1}{2} \right), (-3) (-1), (-3) \left( -\frac{1}{2} \right) \right\} = \left\{ 5, \frac{5}{2}, 3, \frac{3}{2} \right\}$ .

### 3.3 Ausência de Algumas Propriedades e Diferença de Hukuhara

Segundo Barros, Bassanezi e Lodwick (2017) essas operações sobre intervalos não oferecem ao espaço dos intervalos todas as propriedades de espaço vetorial. Isso se deve ao fato de não existir o elemento oposto, ou seja:

$$\bar{a} - \bar{a} \neq 0,$$

quando  $\bar{a}$  é um intervalo não degenerado. Por exemplo:

$$[1; 2] - [1; 2] = [(1) - (2); (2) - (1)] = [-1; 1] \neq 0.$$

Colocar um intervalo em evidência e uma expressão na forma  $\lambda_1 \underline{a} - \lambda_2 \underline{a}$  também pode levar a um resultado não desejado:

$$\begin{cases} 5[0; 1] - 3[0; 1] = [0; 5] - [0; 3] = [0 - 3; 5 - 0] = [-3; 5] \\ (5 - 3)[0; 1] = 2[0; 1] = [0; 2] \end{cases}.$$

Ao longo do texto os modelos sofrerão com essa rigidez nos cálculos, necessitando que nunca se realize o processo de colocar em evidência um intervalo.

Para determinadas situações pode ser mais conveniente usar outras formas de subtração. A diferença de Hukuhara, por exemplo, quando existe um intervalo  $\underline{c}$  tal que  $\underline{a} = \underline{b} + \underline{c}$  denota essa diferença como (BARROS; BASSANEZI; LODWICK, 2017):

$$\underline{a} -_H \underline{b} = \underline{c}.$$

Como a diferença de Hukuhara nem sempre existe usaremos o conceito de diferença de Hukuhara generalizada que para intervalos assume a forma (STEFANINI, 2010):

$$\underline{a} \ominus_{gH} \underline{b} = [\min\{a_1 - b_1, a_2 - b_2\}; \max\{a_1 - b_1, a_2 - b_2\}], \quad (3.2)$$

onde  $\underline{a} = [a_1; a_2]$  e onde  $\underline{b} = [b_1; b_2]$ .

**Exemplo 7.** *Alguns exemplos de diferença de Hukuhara generalizada entre intervalos:*

- $[1; 2] \ominus_{gH} [3; 4] = [\min G, \max G] = [-2; -2]$ , onde:  
 $G = \{(1) - (3), (2) - (4)\} = \{-2, -2\}$ .
- $[2; 5] \ominus_{gH} [-1; 1] = [\min G, \max G] = [3; 4]$ , onde:  
 $G = \{(2) - (-1), (5) - (1)\} = \{3, 4\}$ .
- $[-3; -2] \ominus_{gH} [7; 15] = [\min G, \max G] = [-17; -10]$ , onde:  
 $G = \{(-3) - (7), (-2) - (15)\} = \{-10, -17\}$ .
- $[-5; -3] \ominus_{gH} [-2; -1] = [\min G, \max G] = [-3; -2]$ , onde:  
 $G = \{(-5) - (-2), (-3) - (-1)\} = \{-3, -2\}$ .
- $[0; 1] \ominus_{gH} [0; 1] = [\min G, \max G] = [0; 0]$ , onde:  
 $G = \{(0) - (0), (1) - (1)\} = \{0, 0\}$ .
- $[1; 2] \ominus_{gH} [1; 2] = [\min G, \max G] = [0; 0]$ , onde:  
 $G = \{(1) - (1), (2) - (2)\} = \{0, 0\}$ .
- $[-5; 5] \ominus_{gH} [-5; 5] = [\min G, \max G] = [0; 0]$ , onde:  
 $G = \{(-5) - (-5), (5) - (5)\} = \{0, 0\}$ .

- $[-3; -1] \ominus_{gH} [-3; -1] = [\min G, \max G] = [0; 0]$ , onde:  
 $G = \{(-3) - (-3), (-1) - (-1)\} = \{0, 0\}$ .

A propriedade

$$\bar{a} \ominus_{gH} \bar{a} = [0; 0] = 0$$

faz a diferença de Hukuhara generalizada ser mais adequada para calcular o erro dos modelos com dados intervalares. Ela é usada na Seção 6.5 com esse intuito.

### 3.4 Princípio de Extensão para Intervalos

Além das operações aritméticas também é possível estender as funções reais para aplica-las em intervalos.

Seja uma função  $f : \mathcal{D} \rightarrow \mathbb{R}$ , com  $\mathcal{D} \in \mathbb{R}$ . O princípio de extensão para intervalos oferece um resultado para a função  $f$  aplicada num intervalo  $\bar{a} \in \mathcal{D}$  (BARROS; BASSANEZI; LODWICK, 2017):

$$f(\bar{a}) = \{f(a) : a \in \bar{a}\}.$$

Segundo Barros, Bassanezi e Lodwick (2017) no caso em que  $f$  é uma função contínua  $f(\bar{a})$  assume também a forma de intervalo. No caso particular de  $f$  ser contínua e crescente o princípio de extensão para intervalos assume a forma (BARROS; BASSANEZI; LODWICK, 2017):

$$f(\bar{a}) = f([a_1; a_2]) = [f(a_1); f(a_2)]. \quad (3.3)$$

**Exemplo 8.** *Alguns exemplos de princípio de extensão para intervalos:*

- $f(x) = x^2$ , função crescente e contínua na parte real positiva,  $f([2; 3]) = [4; 9]$ .
- $f(x) = \sqrt{x}$ , função crescente e contínua na parte real positiva,  $f([0; 9]) = [0; 3]$ .
- $f(x) = e^x$ , função crescente e contínua,  $f([-10; 10]) = [e^{-10}; e^{10}]$ .
- $f(x) = x^2$ , função contínua,  $f([-1; 3]) = [0; 9]$ , o resultado é um intervalo porém note que a função não é crescente na parte real negativa. Note também que  $[-1; 3] [-1; 3] = [-3; 9]$  possui um resultado diferente.

Na Seção 6.5 o princípio de extensão para intervalos é usado para calcular  $f(\bar{a})$  para  $f(x) = x^2$  e  $f(x) = \sqrt{x}$ . Explicitamente:

$$([a_1; a_2])^2 = \begin{cases} [a_1^2; a_2^2], & \text{para } a_1 \geq 0, \\ [a_2^2; a_1^2], & \text{para } a_2 \leq 0, \\ [0; \max\{a_1^2, a_2^2\}], & \text{para } a_1 < 0 \text{ e } a_2 > 0. \end{cases} \quad (3.4)$$

$$\sqrt{[a_1; a_2]} = [\sqrt{a_1}; \sqrt{a_2}]. \quad (3.5)$$

### 3.5 Conclusão

Definir operações aritméticas para intervalos oferece ferramentas para tratar dados na forma de intervalos e para a construção de modelos envolvendo esse tipo de dado.

## 4 Conjuntos e Números *Fuzzy*

### 4.1 Introdução

Conjuntos *fuzzy* são uma extensão do conceito clássico de conjunto, conceito apresentado por [ZADEH \(1965\)](#). Os chamados números *fuzzy* serão um objeto central neste trabalho, estes são um tipo particular de conjuntos *fuzzy*.

Segundo [Viertl \(2011\)](#), chamamos de *fuzzy* a todo tipo de dado que não pode ser expresso como um número preciso, tanto dados na forma de uma descrição linguística, como temperatura alta ou pressão sanguínea alta, quanto medidas de uma variável contínua que são sempre imprecisas em maior ou menor grau.

### 4.2 Conjuntos *Fuzzy*

[Barros, Bassanezi e Lodwick \(2017\)](#) explica conjuntos *fuzzy* através da função característica, começando pela função característica para conjuntos clássicos. Segundo [Barros, Bassanezi e Lodwick \(2017\)](#), dado um conjunto universo não vazio  $U$  e um subconjunto  $A$  de  $U$ , a função característica,  $\chi_A \rightarrow \{0, 1\}$ , de  $A$  é dada por:

$$\chi_A(x) = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{se } x \notin A \end{cases} \quad (4.1)$$

A função característica de um subconjunto clássico assume apenas os valores 1 ou 0, indicando a presença ou não do elemento de  $U$  no subconjunto  $A$ .

Segundo [Barros, Bassanezi e Lodwick \(2017\)](#), dado um conjunto universo não vazio  $U$ , um subconjunto  $F$  de  $U$  é caracterizado por sua função de pertinência dada por:

$$\varphi_F : U \rightarrow [0, 1], \quad (4.2)$$

onde o valor  $\varphi_F(x) \in [0, 1]$  indica o grau com que o elemento  $x$  de  $U$  pertence ao conjunto *fuzzy*  $F$ .  $\varphi_F(x) = 1$  significa que o elemento pertence certamente ao conjunto enquanto  $\varphi_F(x) = 0$  significa que o elemento não pertence ao conjunto de forma alguma. Dados dois elementos  $x$  e  $y$ , se  $\varphi_F(x) > \varphi_F(y)$ , então o elemento  $x$  é mais propenso a pertencer ao conjunto do que o elemento  $y$ .

**Exemplo 9** (Números próximos de 2). [Barros, Bassanezi e Lodwick \(2017\)](#) mostra, como exemplo, algumas formas de definir o conjunto dos números reais próximos de 2. Seja  $U = \mathbb{R}$ , o conjunto *fuzzy* dos números próximos de 2 pode ser definido pela função de

pertinência:

$$\varphi_F(x) = \begin{cases} (1 - |x - 2|) & \text{se } 1 \leq x \leq 3 \\ 0 & \text{se } x \notin [1, 3] \end{cases}, \quad (4.3)$$

$x \in \mathbb{R}$ , de onde temos que  $\varphi_F(2,001) = 0,999$ , enquanto  $\varphi_F(7) = 0$ . Dizemos então que  $x = 2,001$  é próximo a 2 com um grau de proximidade 0,999 e que  $x = 7$  não é próximo a 2.

Outra função de pertinência exibida por Barros, Bassanezi e Lodwick (2017) para o conjunto dos números próximos de 2 é:

$$v_F(x) = \exp[-(x - 2)^2], \quad (4.4)$$

onde  $v_F(2,001) = 0,99999$  e  $v_F(7) = 1,388 \times 10^{-11}$ .

Segundo Barros, Bassanezi e Lodwick (2017), dado um subconjunto *fuzzy*  $A$  de um conjunto universo  $U$ , onde  $U$  é um espaço topológico, o conjunto suporte de  $A$  é dado por:

$$\mathbf{supp} A = \{x \in U : \varphi_A(x) > 0\}, \quad (4.5)$$

enquanto para  $\alpha \in [0, 1]$ , o  $\alpha$ -nível do subconjunto  $A$  é o conjunto clássico  $[A]^\alpha$  de  $U$  definido por:

$$[A]^\alpha = \begin{cases} \{x \in U : \varphi_A(x) \geq \alpha\} & \text{para } 0 < \alpha \leq 1 \\ \overline{\mathbf{supp} A} & \text{para } \alpha = 0 \end{cases}, \quad (4.6)$$

onde  $\overline{\mathbf{supp} A}$  é o fecho de  $\mathbf{supp} A$ .

### 4.3 Números Fuzzy

Ao longo deste trabalho o conceito de número *fuzzy* tem um papel central. Números *fuzzy* são um tipo especial de conjunto *fuzzy* cujo domínio é o conjunto dos números reais. Neste trabalho denotamos os números *fuzzy* com um asterisco, notação adotada por Viertl (2011).

Segundo Barros, Bassanezi e Lodwick (2017), um número *fuzzy*  $a^*$  é um conjunto *fuzzy* onde  $U = \mathbb{R}$ , tal que:

- todos os  $\alpha$ -níveis de  $a^*$  são não vazios para  $0 \leq \alpha \leq 1$ .
- todos os  $\alpha$ -níveis de  $a^*$  são intervalos fechados de  $\mathbb{R}$ .
- $\mathbf{supp} a^* = \{x \in \mathbb{R} : \varphi_{a^*}(x) > 0\}$  é limitado.

O conjunto de todos os números *fuzzy* é denotado por  $\mathcal{F}(\mathbb{R})$ . Como notação usamos:

$$[a^*]^\alpha = [l(\alpha), r(\alpha)], \quad (4.7)$$

onde  $l : [0, 1] \rightarrow \mathbb{R}$  e  $r : [0, 1] \rightarrow \mathbb{R}$  são funções numéricas que nos dão os limites dos intervalos de cada  $\alpha$ -nível.

Tabela 1 – Tipos particulares de números *fuzzy* exibidos por Barros, Bassanezi e Lodwick (2017).

tipo	notação particular	$\alpha$ -nível	gráfico da função de pertinência
triangular	$(a; u; c)$	$[(u - a)\alpha + a; (u - c)\alpha + c]$	
trapezoidal	$(a; b; c; d)$	$[(b - a)\alpha + a; (c - d)\alpha + d]$	
<i>crisp</i>	$\hat{r}$	$[r; r]$	

Fonte: (BARROS; BASSANEZI; LODWICK, 2017). Imagens produzidas pelos autores.

Nas Tabelas 1 e 2 estão presentes alguns exemplos de tipos particulares de números *fuzzy* apresentados por Barros, Bassanezi e Lodwick (2017).

As operações aritméticas dos números reais clássicos podem se estendidas para os números *fuzzy*. Ao longo deste trabalho usamos as operações básicas de adição, subtração, multiplicação e divisão entre números *fuzzy* diversas vezes.

Barros, Bassanezi e Lodwick (2017) define as operações aritméticas de números *fuzzy* através das operações aritméticas intervalares, vistas na Seção 3.2, sobre os seus  $\alpha$ -níveis.

Sejam  $a^*$  e  $b^* \in \mathcal{F}(\mathbb{R})$  e  $\lambda \in \mathbb{R}$ , definimos as operações aritméticas entre números *fuzzy* como:

- $[a^* + b^*]^\alpha = [a^*]^\alpha + [b^*]^\alpha$ .
- $[a^* - b^*]^\alpha = [a^*]^\alpha - [b^*]^\alpha$ .
- $[a^* b^*]^\alpha = [a^*]^\alpha [b^*]^\alpha$ .

Tabela 2 – Tipos particulares de números *fuzzy* exibidos por Barros, Bassanezi e Lodwick (2017).

tipo	função de pertinência	gráfico da função de pertinência
triangular	$\varphi_{a^*}(x) = \begin{cases} \frac{x}{u-a} + \frac{a}{a-u}, & \text{se } a < x < u \\ 1, & \text{se } x = u \\ \frac{x}{u-c} + \frac{c}{c-u}, & \text{se } u < x < c \\ 0, & \text{caso contrário} \end{cases}$	
trapezoidal	$\varphi_{a^*}(x) = \begin{cases} \frac{x}{b-a} + \frac{a}{a-b}, & \text{se } a < x < b \\ 1, & \text{se } b \leq x \leq c \\ \frac{x}{c-d} + \frac{c}{c-u}, & \text{se } u < x < c \\ 0, & \text{caso contrário} \end{cases}$	
<i>crisp</i>	$\varphi_{a^*}(x) = \begin{cases} 1, & \text{se } x = r \\ 0, & \text{caso contrário} \end{cases}$	
Fixados $u$ , $a$ e $\delta$ , a função de pertinência é:		
sino	$\varphi_{a^*}(x) = \begin{cases} \exp\left(-\left(\frac{x-u}{a}\right)^2\right), & \text{se }  x-u  \leq \delta \\ 0, & \text{caso contrário} \end{cases}$	
com $\bar{\alpha} = \exp(-(\delta/a)^2)$		

Fonte: (BARROS; BASSANEZI; LODWICK, 2017). Imagens produzidas pelos autores.

- se  $0 \notin \mathbf{supp} b^*$ , então definimos  $\left[\frac{a^*}{b^*}\right]^\alpha = [a^*]^\alpha/[b^*]^\alpha$ .
- $[\lambda a^*]^\alpha = \lambda [a^*]^\alpha$ .

Os mesmos problemas apontados na Seção 3.3 para as operações com intervalos também estão presentes nas operações com números *fuzzy*. Assim também temos a ausência de elemento oposto e a falta da propriedade distributiva, deixando assim os modelos, ao longo do texto, rígidos em relação à ordem das operações.

Ao longo desse trabalho são usadas as operações de soma, subtração e multiplicação por escalar. Tomando como exemplo os números triangulares *fuzzy*  $a^* = (2; 3; 5)$  e  $b^* = (0; 1; 2)$ . Seus  $\alpha$ -níveis são dados por  $[a^*]^\alpha = [2 + \alpha; 5 - 2\alpha]$  e  $[b^*]^\alpha = [\alpha; 2 - \alpha]$ .

Para encontrar o valor de  $c^*$  tal que  $c^* = a^* + 2b^*$  tratamos dos  $\alpha$ -níveis de  $c^*$ .

$$\begin{aligned}
 c^* &= a^* + 2b^* \\
 [c^*]^\alpha &= [a^*]^\alpha + 2[b^*]^\alpha \\
 [c^*]^\alpha &= [2 + \alpha; 5 - 2\alpha] + 2[\alpha; 2 - \alpha] \\
 [c^*]^\alpha &= [2 + \alpha; 5 - 2\alpha] + [2\alpha; 4 - 2\alpha] \\
 [c^*]^\alpha &= [2 + \alpha + (2\alpha); 5 - 2\alpha + (4 - 2\alpha)] \\
 [c^*]^\alpha &= [2 + 3\alpha; 9 - 4\alpha].
 \end{aligned} \tag{4.8}$$

Assim obtemos a expressão por  $\alpha$ -níveis de  $c^* = a^* + 2b^*$ . Dado que  $[2 + 3\alpha; 9 - 4\alpha] = [(5 - 2)\alpha + 2; (5 - 9)\alpha + 9]$ , é possível observar que  $c^*$  é o número triangular *fuzzy*  $c^* = (2; 5; 9)$ .

Um conceito importante usado nesse trabalho é o processo chamado de *defuzzificação*.

Segundo Barros, Bassanezi e Lodwick (2017), o processo de defuzzificação determina um número real que melhor represente um número *fuzzy*. Dentre as diversas formas de se realizar esse processo, escolhemos arbitrariamente o método do centro de gravidade:

$$\text{defuzz}(a^*) = \frac{\int_{\mathbb{R}} u \varphi_{a^*}(u) du}{\int_{\mathbb{R}} \varphi_{a^*}(u) du}. \tag{4.9}$$

O número real  $\text{defuzz}(a^*)$  divide a função de pertinência em duas partes em que a área entre ela e o eixo dos números reais é igual. Nas aplicações o pesquisador pode escolher o método de defuzzificação mais conveniente.

## 4.4 Conclusão

Do mesmo modo que definir operações para intervalos, Capítulo 3, oferece ferramentas para a construção de modelos, também os números *fuzzy* podem ser usados com esse mesmo intuito.

## 5 Modelando Imprecisão através de Intervalos e de Números *Fuzzy*

### 5.1 Introdução

Segundo Barros, Bassanezi e Lodwick (2017), quando um dado é proveniente de uma medição ele é suscetível à imprecisão, originada do equipamento, de quem está coletando os dados ou de diversas outras fontes. Viertl (2011) reforça que toda medida tomada de uma quantidade contínua sempre está associada a algum grau de imprecisão.

Barros, Bassanezi e Lodwick (2017) propõe o uso de números *fuzzy* para modelar a imprecisão de uma medida em  $\mathbb{R}$  de um dado. Quando uma medida é aproximadamente determinado valor, usamos um número *fuzzy* como modelo.

COUSO e DUBOIS (2014) categoriza os modelos matemáticos em ônticos e epistêmicos. Um modelo ôntico é uma representação precisa da realidade mesmo que tenha algum tipo de erro. Um modelo epistêmico é uma representação que leva em conta a precisão limitada da nossa capacidade de medição.

No contexto deste trabalho usamos intervalos e números *fuzzy* como um modelo epistêmico de representar medidas que contenham imprecisão.

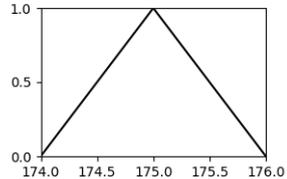
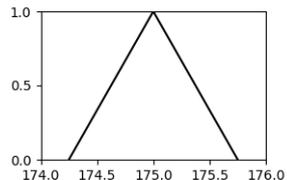
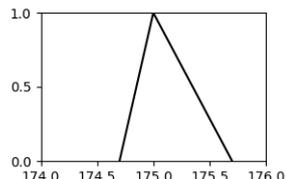
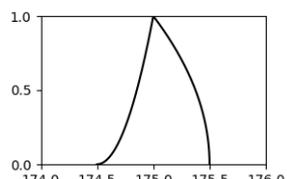
### 5.2 Modelando Imprecisão

Começando com um exemplo simples. Suponha que a altura de uma pessoa tenha sido medida com uma fita métrica com a precisão de 1 centímetro. Se a medida obtida foi de 175 centímetros podemos levar em conta que a imprecisão da fita métrica, que é de 1 centímetro, coloca um certo grau de incerteza nessa medida. A interferência de quem está medindo também pode nos dar um maior grau de imprecisão nessa medida. São 175 por que a altura estava um pouco abaixo de 175? Ou estava um pouco acima? Ou bem acima, porém, sem alcançar o ponto de 176? Todas essas características de imprecisão podem ser modeladas através de números *fuzzy*. Na Tabela 3 apresentamos diversas formas de modelar este exemplo através de números *fuzzy*.

Pinto et al. (2018) modela dados que tenham uma média  $m$  e um desvio padrão  $\sigma$  através de números *fuzzy* triangulares na forma  $(m - \sigma; m; m + \sigma)$ .

Seguindo essa mesma linha de raciocínio, podemos também criar outras formas de modelar esses dados como, por exemplo, números trapezoidais na forma  $(m - 2\sigma; m - \sigma; m + \sigma; m + 2\sigma)$ . Ou, na forma de intervalos, como  $[m - \sigma; m + \sigma]$  ou  $[m - 2\sigma; m + 2\sigma]$ .

Tabela 3 – Diferentes modos de modelar uma medida próxima de 175cm através de números *fuzzy*

descrição	número <i>fuzzy</i>	gráfico da função de pertinência
Número <i>fuzzy</i> triangular simétrico dando pertinências para os valores entre 174 e 176.	(174; 175; 176)	
Número <i>fuzzy</i> triangular simétrico dando pertinências para os valores entre 174, 25 e 175, 75.	(174, 25; 175; 175, 75)	
Número <i>fuzzy</i> triangular dando pertinências para os valores entre 174.7 e 175.7. Dando pertinências para mais valores maiores do que menores que 175.	(174, 7; 175; 175, 7)	
Número <i>fuzzy</i> genérico dando pertinências para os valores entre 174, 5 e 175, 5, porém, tendendo a dar pertinências maiores para números maiores que 175.	$[175, 5 + \frac{\sqrt{\alpha}}{2}; 176, 5 - \frac{\alpha^2}{2}]$	

Fonte: Os autores.

Segundo [Wagner e Tyler \(2011\)](#), no cuidado de cavalos no campo, diversas vezes uma balança para determinar o peso do animal não é acessível. É essencial saber a massa corporal do cavalo para a administração de medicamentos ou a prescrição de dietas ao animal e na falta de uma balança é possível estimar a mesma com o uso de uma fita métrica. A fórmula habitual para determinar a massa corporal de cavalos com uma fita métrica é ([CARROLL; HUNTINGTON, 1988](#)):

$$M = \frac{g^2 \times b}{11,880\text{cm}^3}, \quad (5.1)$$

onde  $M$  apresenta a massa em quilogramas,  $g$  o perímetro torácico em centímetros e  $b$  o comprimento do corpo do cavalo. [CARROLL e HUNTINGTON \(1988\)](#) apresenta um desvio padrão que pode variar entre 37, 2 e 42, 7 dependendo do grupo de cavalos no seu estudo.

Com essa estimativa da massa corporal temos não apenas a imprecisão da fita métrica e da medição em si, mas também o fato de ser uma estimativa coloca uma grande imprecisão nesse valor. Podemos modelar essa imprecisão também com números *fuzzy*, como, por exemplo, com o número trapezoidal  $(M - 80; M - 40; M + 40; M + 80)$ , com  $M$  dado pela equação 5.1.

Outras situações similares onde as grandezas são estimadas de forma indireta e suscetível a imprecisões podem também, a critério do pesquisador, ser modeladas na forma de números *fuzzy*. Ao longo deste trabalho vários exemplos de medidas na forma de números *fuzzy* estão presentes.

O valor do erro padrão,  $\sigma/\sqrt{n_x}$ , que é o desvio padrão da distribuição amostral, (EVERITT, 2006), também pode ser usado para modelar a imprecisão. Por exemplo, se a medida da observação foi tomada  $n_x$  vezes, com média  $m$  e desvio padrão  $\sigma$ , podemos modelar um número *fuzzy* usando o erro padrão:  $(m - \sigma/\sqrt{n_x}; m; m + \sigma/\sqrt{n_x})$ .

### 5.3 Conclusão

Os processos onde as observações são tomadas usando métodos com algum grau de imprecisão podem ser modelados usando intervalos ou números *fuzzy*. Isso permite um olhar distinto para a construção de modelos para esses conjuntos de dados.

## 6 Métodos para Dados *Fuzzy*

### 6.1 Introdução

Neste capítulo apresentamos alguns métodos presentes na literatura que estendem os conceitos de métodos lineares usuais para dados onde a variável resposta tem a forma de números *fuzzy*. Também apresenta alguns exemplos na última seção.

Para o caso simples, os métodos para dados *fuzzy* apresentam uma aproximação da relação entre a variável explicativa  $X$  e a variável resposta  $Y^*$ . Esse método estima modelos por uma função  $f^* : \mathbb{R} \rightarrow \mathcal{F}(\mathbb{R})$  que tem como argumento o número real  $\vartheta$  e como imagem um número *fuzzy*. Denotamos as observações como pares ordenados na forma  $(x_i; y_i^*)$ . Por exemplo, a primeira observação é denotada como  $(x_1; y_1^*)$ , o que significa que para  $Y^*$  foi observado um valor na forma de número *fuzzy*  $y_1^*$  com a variável  $X$  tendo o valor  $x_1$ .

Para o caso múltiplo temos  $q$  variáveis explicativas,  $X_1, \dots, X_q$  e a variável resposta  $Y^*$ . Esse método estima modelos por uma função  $f^*(\vartheta_1, \dots, \vartheta_q)$ ,  $f^* : \mathbb{R}^q \rightarrow \mathcal{F}(\mathbb{R})$ , que tem como argumento  $q$  números reais  $\vartheta_1, \dots, \vartheta_q$  e como imagem um número *fuzzy*. Denotamos as observações como tuplas ordenadas na forma  $(x_{i,1}; \dots; x_{i,q}; y_i^*)$ . Por exemplo, a primeira observação é denotada como  $(x_{1,1}; \dots; x_{1,q}; y_1^*)$ , o que significa que para  $Y^*$  foi observado um valor na forma de número *fuzzy*  $y_1^*$  com a variável  $X_1$  tendo o valor  $x_{1,1}$ , a variável  $X_2$  tendo o valor  $x_{1,2}$  e, assim por diante até  $X_q$  tendo o valor  $x_{1,q}$ .

[Diamond e Tanaka \(1998\)](#) divide os métodos para dados *fuzzy* em duas categorias: métodos possibilísticos e métodos de mínimos quadrados. Já [COUSO e DUBOIS \(2014\)](#) inclui estas e mais duas categorias: regressão quantílica e análise sensitiva. Deste modo podemos dividir os métodos de regressão *fuzzy* em quatro categorias:

- métodos possibilísticos
- métodos de quadrados mínimos
- regressão quantílica
- análise sensitiva

[COUSO e DUBOIS \(2014\)](#) usa essa classificação também para métodos onde as observações possuem a forma de intervalos.

## 6.2 Métodos Possibilísticos

Segundo COUSO e DUBOIS (2014), os métodos possibilísticos apresentam restrições na forma de inclusão de conjuntos entre os dados observados  $y_i^*$  e as previsões  $f^*(x_{i,1}, \dots, x_{i,q})$ . No caso intervalar, a restrição pode impor que o intervalo previsto  $\bar{f}(x_i)$  esteja sempre contido no intervalo  $\underline{y}_i$  observado.

COUSO e DUBOIS (2014) continua explicando alguns pontos sobre esses métodos:

- São métodos que não pressupõem uma leitura ôntica ou epistêmica dos dados. Caso os dados sejam ônticos, as previsões dadas pelo modelo se mantêm ônticas. Caso os dados sejam epistêmicos, as previsões manterão a imprecisão presente nas observações, pois o resultado da regressão é um número *fuzzy*, assim as previsões são epistêmicas com imprecisão tal qual são as observações.
- Esses métodos não estendem de forma clara os conceitos da regressão linear para dados na forma de números reais. Por exemplo, nenhuma hipótese é feita sobre os erros.
- Esses métodos apresentam modelos com um comportamento não natural ao tornar as previsões dadas por  $f^*(\vartheta_1, \dots, \vartheta_q)$  com o suporte mais largo quando o módulo dos valores atribuídos às variáveis  $\vartheta$  aumentam. Ou seja, com  $a, b \in \mathbb{R}$  se  $|a| < |b|$ , então  $\mathit{supp} f^*(b)$  é um intervalo com comprimento maior do que  $\mathit{supp} f^*(a)$ .
- Esses métodos apresentam restrição para o tipo de dado *fuzzy*. Por exemplo, alguns métodos aceitam apenas dados na forma de números *fuzzy* triangulares.

O método possibilístico *Min Problem* (TANAKA; UEJIMA; ASAI, 1982) é obtido para um conjunto de  $n$  observações  $(x_{i,1}; \dots; x_{i,q}; y_i^*)$ ,  $i = 1, \dots, n$ , onde os  $x_{i,1}, \dots, x_{i,q}$  são números reais e os  $y_i^*$  são números triangulares *fuzzy* simétricos  $y_i^* = (m_i - a_i; m_i; m_i + a_i)$ . Temos o modelo:

$$f^*(\vartheta_1, \dots, \vartheta_q) = \beta_1^* \vartheta_1 + \dots + \beta_q^* \vartheta_q, \quad (6.1)$$

onde os  $\beta_i^* = (b_i - r_i; b_i; b_i - r_i)$ , são números *fuzzy* triangulares simétricos, com  $r_i \geq 0$ , e  $\vartheta_1, \dots, \vartheta_q$  são os argumentos da função  $f^* : \mathbb{R}^q \rightarrow \mathcal{F}(\mathbb{R})$ ,  $\beta_1^*, \dots, \beta_q^*$  são os parâmetros a serem estimados para o modelo, o que é feito através de um programa de otimização linear. Fixado um  $h$  arbitrário, com  $0 \leq h < 1$ , os coeficientes  $\beta_i^*$  do modelo são encontrados com:

$$\min_{b_1, \dots, b_q, r_1, \dots, r_q} \sum_{i=1}^n (|x_{i,1}| r_1 + \dots + |x_{i,q}| r_q) \quad (6.2)$$

Sujeito a:  $r_i \geq 0$  e  $[f^*(x_{i,1}, \dots, x_{i,q})]^h \subseteq [y_i^*]^h \forall i$ .

Apesar da função objetivo  $\sum_{i=1}^n (|x_{i,1}| r_1 + \dots + |x_{i,q}| r_q)$  não apresentar os valores  $b_i$ , estes estão presentes nas restrições do problema, uma vez que  $f^*(x_{i,1}, \dots, x_{i,q}) = \beta_1^* x_{i,1} + \dots + \beta_q^* x_{i,q}$  e  $\beta_i^* = (b_i - r_i; b_i; b_i - r_i)$ .

O Exemplo 10 da Seção 6.6 trata deste método.

Fixado o valor  $h$ , a restrição diz respeito aos  $\alpha$ -níveis  $h$  dos dados observados e do valor da função do modelo no ponto correspondente, ou seja,  $[y_i^*]^h$  e  $[f^*(x_{i,1}, \dots, x_{i,q})]^h$ , respectivamente. A restrição  $[f^*(x_{i,1}, \dots, x_{i,q})]^h \subseteq [y_i^*]^h \forall i$  obriga que o  $\alpha$ -nível  $h$  da previsão  $f^*(x_{i,1}, \dots, x_{i,q})$  esteja contido na observação  $y_i^*$ .

As restrições  $[f^*(x_{i,1}, \dots, x_{i,q})]^h \subseteq [y_i^*]^h$  podem ser reescritas como desigualdades. Sejam dois números *fuzzy* triangulares simétricos,  $a^* = (m_a - r_a; m_a; m_a - r_a)$  e  $b^* = (m_b - r_b; m_b; m_b - r_b)$ , a  $[a^*]^h \subseteq [b^*]^h$  é dada pelas desigualdades:

$$\begin{cases} m_b \leq m_a + (1 - h)(r_b - r_a) \\ m_b \geq m_a - (1 - h)(r_b - r_a). \end{cases} \quad (6.3)$$

Tanaka, Hayashi e Watada (1989) apresenta também este método além de mais dois métodos possibilísticos diferentes deste.

Pinto et al. (2018) realiza um processo possibilístico para dados *fuzzy* usando definições próprias da teoria *fuzzy* para conceitos análogos à correlação e dados longitudinais. Assim como o conceito de números *fuzzy* interativos. Sendo um trabalho com o foco distinto da intenção do presente trabalho.

### 6.3 Métodos de Quadrados Mínimos

Segundo COUSO e DUBOIS (2014), os métodos de quadrados mínimos *fuzzy* são baseados em algum tipo de distância entre o modelo e os valores observados, geralmente minimizando o quadrado da distância. Explica também que este método trata os dados como entidades ônticas e que existe uma grande variedade de métodos deste tipo.

KORNER e NATHER (1998) apresenta alguns métodos de regressão para dados na forma de números *fuzzy*. Apresentamos aqui dois dos seus métodos, um que de alguma forma estende a regressão por quadrados mínimos clássica vista no Capítulo 1, já o segundo usa de fato um conceito de distância entre os números *fuzzy* no modelo.

Chamamos aqui o primeiro método apresentado de KORNER e NATHER (1998) de método clássico estendido. Neste método os valores *fuzzy* das observações são usados no lugar dos valores clássicos na fórmula de quadrados mínimos clássica, visto no Capítulo 1.

KORNER e NATHER (1998) apresenta o método clássico estendido. Seja um

conjunto de observações *fuzzy*  $(x_{1,1}, \dots, x_{1,q}; y_1^*), \dots, (x_{n,1}, \dots, x_{n,q}; y_n^*)$ , os coeficientes do modelo clássico de quadrados mínimos estendido para dados *fuzzy*,

$$f^*(x_1, \dots, x_q) = \hat{\beta}_0^* + \hat{\beta}_1^* x_1 + \dots + \hat{\beta}_q^* x_q, \quad (6.4)$$

são dados por:

$$\begin{cases} \hat{\beta}_0^* = w_{1,1} y_1^* + \dots + w_{1,n} y_n^* \\ \hat{\beta}_1^* = w_{2,1} y_1^* + \dots + w_{2,n} y_n^* \\ \vdots \\ \hat{\beta}_q^* = w_{q,1} y_1^* + \dots + w_{q,n} y_n^* \end{cases}, \quad (6.5)$$

onde

$$\mathbf{W} = \begin{pmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,n} \\ w_{2,1} & w_{2,2} & \dots & w_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{q,1} & w_{q,2} & \dots & w_{q,n} \end{pmatrix} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T, \quad (6.6)$$

sendo  $\mathbf{X}$  a matriz

$$\mathbf{X} = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{1,q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \dots & x_{n,q} \end{pmatrix}. \quad (6.7)$$

Este modelo não restringe os números *fuzzy* das observações para alguma forma específica, como, por exemplo, se limitar a apenas números *fuzzy* triangulares.

O Exemplo 11 da Seção 6.6 trata deste método.

KORNER e NATHER (1998) apresenta um método de quadrados mínimos para dados para diversas formas de números *fuzzy*. Por simplicidade apresentamos esse método para o caso particular em que os números *fuzzy* sejam triangulares simétricos.

Como caso particular do método de quadrados mínimos para números *fuzzy* apresentado por KORNER e NATHER (1998), seja um conjunto de observações *fuzzy*  $(x_{1,1}; \dots; x_{1,q}; y_1^*), \dots, (x_{n,1}; \dots; x_{n,q}; y_n^*)$ , onde os  $x_{i,1}, \dots, x_{i,q}$  são números reais e os  $y_i^*$  são números *fuzzy* triangulares simétricos  $y_i^* = (m_i - r_i; m_i; m_i + r_i)$ . O modelo

$$f^*(\vartheta_1, \dots, \vartheta_q) = \hat{\beta}_0^* + \hat{\beta}_1^* \vartheta_1 + \dots + \hat{\beta}_q^* \vartheta_q \quad (6.8)$$

tem seus coeficientes  $\hat{\beta}_i$ ,  $i = 0, 1, \dots, q$ , dados por:

$$\inf_{\hat{\beta}_0^*, \hat{\beta}_1^*, \dots, \hat{\beta}_q^*} \frac{1}{n} \sum_{i=1}^n d^2(\hat{\beta}_0^* + \hat{\beta}_1^* x_{i,1} + \dots + \hat{\beta}_q^* x_{i,q}, y_i^*), \quad (6.9)$$

onde  $d^2(a^*, b^*)$  é o quadrado da distância entre os números *fuzzy* triangulares simétricos  $a^* = (m_a - r_a; m_a; m_a + r_a)$  e  $b^* = (m_b - r_b; m_b; m_b + r_b)$  dada por:

$$d^2(a^*, b^*) = (m_a - m_b)^2 + \frac{(r_a - r_b)^2}{3}. \quad (6.10)$$

O Exemplo 12 da Seção 6.6 trata deste método.

Existe uma grande diversidade de exemplos de diferentes métodos de quadrados mínimos para números *fuzzy* na literatura (COUSO; DUBOIS, 2014). Outro exemplo mais profundo é encontrado em Sánchez (2012), que usa outros conceitos de distância e base teórica mais elaborada que garante convergência, existência e unicidade das soluções.

Neto (2012) apresenta processos por mínimos quadrados para dados intervalares. No entanto, usa as variáveis explicativas também na forma de intervalo, o que foge do nosso escopo. Caso parecido aparece em Colubi et al. (2007) que apresenta um método para variáveis explicativas *fuzzy*, porém, levando em conta os conceitos relacionados a variáveis aleatórias *fuzzy*.

## 6.4 Regressão Quantílica e Análise Sensitiva

COUSO e DUBOIS (2014) define a regressão quantílica como um método epistêmico que fornece intervalos de confiança que contém o valor  $y(x)$  com determinado grau de confiança e que pode ser modelada através de números *fuzzy* que tenham como  $\alpha$ -níveis esses intervalos. Não nos aprofundaremos nesses métodos.

Já, sobre a análise sensitiva, COUSO e DUBOIS (2014) a classifica como uma abordagem que pode ser genuinamente classificada como epistêmica e a resume como o problema que toma  $n$  observações  $(x_1; \underline{y}_1), \dots, (x_n; \underline{y}_n)$  onde  $\underline{y}_i$  são os intervalos  $\underline{y}_i = [\underline{y}_i, \bar{y}_i]$  e tem o modelo dado por:

$$\underline{f}(\vartheta) = \left\{ \begin{array}{l} \hat{a}\vartheta + \hat{b}; \\ \forall \hat{a}, \hat{b} \text{ que minimize } \sum_{i=1}^n (ax_i + b - y_i)^2, \forall y_i \in [\underline{y}_i, \bar{y}_i], i = 1, \dots, n \end{array} \right\}, \quad (6.11)$$

assim  $\underline{f}(\vartheta)$  se torna o conjunto das estimativas  $f(\vartheta)$  para todas as combinações possíveis ao se escolher um ponto do intervalo de cada observação.

$\underline{f}(\vartheta)$  terá a forma de um intervalo real. Exibindo na forma de gráfico, como vemos na Figura 6 do Exemplo 13 da Seção 6.6, se nota que os intervalos postos lado a lado formam uma região do plano onde estão os pontos encontrados nesse modelo.

Já o gráfico dessa região será a união de todas as retas na forma  $\hat{a}\vartheta + \hat{b}$   $\forall \hat{a}, \hat{b}$  que minimizem  $\sum_{i=1}^n (ax_i + b - y_i)^2, \forall y_i \in [\underline{y}_i, \bar{y}_i], i = 1, \dots, n$ . Porém, o modelo não se trata de uma família de curvas, mas sim da união de todos os pontos dessas curvas, tomando para cada  $\vartheta$  o intervalo correspondente como estimativa do modelo.

COUSO e DUBOIS (2014) expõe ainda que Loquin e Dubois (2010) usa este método para realizar o processo da geoestatística chamado de krigagem, ou krigeagem.

## 6.5 Avaliação dos Modelos

Do mesmo modo visto na Seção 1.4 é necessário ter uma forma de avaliar a qualidade de um modelo de regressão. Viertl (2011) propõe o uso do princípio da extensão de Zadeh, que estende funções reais para conjuntos *fuzzy*, para estender os testes estatísticos para modelos que tratam de dados na forma de números *fuzzy*. Desta forma o valor dado por um teste estatístico se torna um número *fuzzy* também, o qual tem o seu  $\alpha$ -nível analisado.

Usamos aqui a aritmética intervalar e o princípio da extensão para intervalos, do Capítulo 3, para estender o teste estatístico do coeficiente de determinação.

Para o coeficiente de determinação:

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

onde  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ .

No caso intervalar, com observações  $\underline{y}_1, \dots, \underline{y}_n$  e previsões  $\hat{\underline{y}}_1, \dots, \hat{\underline{y}}_n$  usamos:

$$\overline{R^2} = [1; 1] \ominus_{gH} \frac{\sum_{i=1}^n (\hat{\underline{y}}_i \ominus_{gH} \underline{y}_i)^2}{\sum_{i=1}^n (\underline{y}_i \ominus_{gH} \bar{\underline{y}})^2}, \quad (6.12)$$

onde  $\bar{\underline{y}} = \frac{1}{n} \sum_{i=1}^n \underline{y}_i$ .

No caso *fuzzy*, com observações  $y_1^*, \dots, y_n^*$  e previsões  $\hat{y}_1^*, \dots, \hat{y}_n^*$  usamos:

$$\overline{R^2} = [1; 1] \ominus_{gH} \frac{\sum_{i=1}^n ([\hat{y}_i^*]^0 \ominus_{gH} [y_i^*]^0)^2}{\sum_{i=1}^n ([y_i^*]^0 \ominus_{gH} [\bar{y}^*]^0)^2}, \quad (6.13)$$

onde  $\bar{y}^* = \frac{1}{n} \sum_{i=1}^n y_i^*$ .

Em ambos os casos  $\ominus_{gH}$  é a diferença de Hukuhara generalizada da Equação 3.2 e  $(\bar{a})^2$  é calculado usando o princípio de extensão para intervalos, Equação 3.4. Para as somas e divisões usamos as operações aritméticas entre intervalos da Seção 3.2.  $[a^*]^0$  é o  $\alpha$ -nível 0 de  $a^*$ , Equação 4.6.

Para testar a normalidade dos erros, primeiramente definimos como calcular os erros para os casos onde os dados são intervalares ou na forma de números *fuzzy*.

No caso intervalar, com observações  $\underline{y}_1, \dots, \underline{y}_n$  e previsões  $\hat{\underline{y}}_1, \dots, \hat{\underline{y}}_n$  definimos:

$$\bar{e}_i = \hat{y}_i \ominus_{gH} \bar{y}_i.$$

No caso *fuzzy*, com observações  $y_1^*, \dots, y_n^*$  e previsões  $\hat{y}_1^*, \dots, \hat{y}_n^*$  definimos:

$$\bar{e}_i = [\hat{y}_i^*]^0 \ominus_{gH} [y_i^*]^0.$$

Na Seção 1.4 definimos o gráfico quantil-quantil para o caso em que os erros são na forma de números reais. Propomos aqui uma forma de construir uma versão desse gráfico para o caso em que os erros são na forma de intervalos. Seja um conjunto com  $n$  erros na forma intervalar,  $\bar{e}_1, \dots, \bar{e}_n$ . Ordenamos esse conjunto, obtendo um conjunto  $\bar{a}_1, \dots, \bar{a}_n$  de forma que os respectivos pontos médios  $m_i$  estejam em ordem crescente. Essa ordenação não é única, pois diferentes intervalos podem ter o mesmo valor como ponto médio. Tomamos o conjunto de  $n$  quantis da distribuição normal com média zero e desvio padrão  $\sigma = \sqrt{\frac{\sum_{i=1}^n (m_i - M)^2}{n-1}}$ , com  $M = \frac{\sum_{i=1}^n m_i}{n}$ ,  $b_i = F^{-1}\left(\frac{i - \frac{1}{2}}{n}\right)$ , onde,  $F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{t^2}{2\sigma^2}\right) dt$ . Por simplicidade tomamos o desvio padrão dos pontos médios dos intervalos, outras formas de se encontrar o desvio padrão para um conjunto de intervalos não são descartadas e podem ser usadas.

Propomos construir o gráfico quantil-quantil para o caso de erros na forma intervalar exibindo os pontos  $(a_i; b_i)$  para todo  $a_i \in \bar{a}_i$ , exibindo também em destaque os pontos  $(m_i; b_i)$  juntamente da reta  $y = x$ . Da mesma forma do caso clássico, os pontos  $(a_i; b_i)$  estarem próximos da reta  $y = x$  indica que os erros estão distribuídos normalmente.

Ao longo do texto exploramos esse tipo de gráfico em diversos exemplos. Junto ao valor do intervalo  $\bar{R}^2$  o usaremos para avaliar a qualidade dos modelos construídos.

Em adição aos gráficos quantil-quantil e ao  $R^2$  é possível definir também outras formas de avaliação dos modelos. Como o uso de bandas de confiança nos gráficos quantil-quantil. Ou o uso de outras medidas de ajuste para os modelos que o pesquisador tenha por adequado.

## 6.6 Exemplos

Esta seção tem como objetivo apresentar alguns exemplos dos quatro métodos apresentados neste capítulo. Nenhum deles terá seu significado aprofundado, este intuito estará presente na Parte de Casos de Uso, onde estão presentes casos de uso para os métodos propostos neste trabalho e comparações entre estes e os métodos da bibliografia apresentados aqui.

Usamos como exemplo os dados da Tabela 4, que é um exemplo simples e com apenas 5 observações. Deste modo somos capazes de exibir alguns detalhes nos cálculos para a obtenção dos modelos, porém, mesmo para esse caso pequeno é preferível usar ajuda computacional.

Tabela 4 – Dados usados para os Exemplos 10, 11, 12 e 14. Os valores  $y_i^*$  são números *fuzzy* triangulares simétricos.

$x_i$	$y_i^* = (m_i - a_i; m_i; m_i + a_i)$
$x_1 = 1$	$y_1^* = (6, 2; 8, 0; 9, 8)$
$x_2 = 2$	$y_2^* = (4, 2; 6, 4; 8, 6)$
$x_3 = 3$	$y_3^* = (6, 9; 9, 5; 12, 1)$
$x_4 = 4$	$y_4^* = (10, 9; 13, 5; 16, 1)$
$x_5 = 5$	$y_5^* = (10, 6; 13, 0; 15, 4)$

Fonte: Tanaka, Hayashi e Watada (1989)

**Exemplo 10** (Método Possibilístico *Min Problem*). *Sejam os dados apresentados na Tabela 4, queremos encontrar o modelo possibilístico da Equação .*

Como o objetivo é encontrar o modelo possibilístico da Equação 10:

$$f^*(\vartheta) = \hat{\beta}_0^* + \vartheta \hat{\beta}_1^*, \tag{6.14}$$

com  $\hat{\beta}_0^* = (b_0 - r_0; b_0; b_0 + r_0)$  e  $\hat{\beta}_1^* = (b_1 - r_1; b_1; b_1 + r_1)$  números *fuzzy* triangulares simétricos, adicionado o termo  $\hat{\beta}_0^*$  para o intercepto. Para encontrar os coeficientes  $\hat{\beta}_0^*$  e  $\hat{\beta}_1^*$  para  $h = 0, 5$ , resolvemos do problema de programação linear:

$$\begin{aligned} \min_{b_0, b_1, r_0, r_1} \sum_{i=1}^5 (r_0 + |x_i| r_1) \\ \text{Sujeito a: } r_i \geq 0 \text{ e } [f^*(x_i)]^h \subseteq [y_i^*]^h \forall i, \end{aligned} \tag{6.15}$$

que pode ser reescrito como

$$\begin{aligned} \min_{b_0, b_1, r_0, r_1} (r_0 + |x_1| r_1 + r_0 + |x_2| r_1 + r_0 + |x_3| r_1 + r_0 + |x_4| r_1 + r_0 + |x_5| r_1) \\ \text{Sujeito a: } r_i \geq 0 \text{ e } [f^*(x_i)]^{0,5} \subseteq [y_i^*]^{0,5} \forall i, \end{aligned} \tag{6.16}$$

$$\begin{aligned} \min_{b_0, b_1, r_0, r_1} (r_0 + 1 r_1 + r_0 + 2 r_1 + r_0 + 3 r_1 + r_0 + 4 r_1 + r_0 + 5 r_1) \\ \text{Sujeito a: } r_i \geq 0 \text{ e } [f^*(x_i)]^{0,5} \subseteq [y_i^*]^{0,5} \forall i, \end{aligned} \tag{6.17}$$

$$\begin{aligned} \min_{b_0, b_1, r_0, r_1} (5r_0 + 15 r_1) \\ \text{Sujeito a: } r_i \geq 0 \text{ e } [f^*(x_i)]^{0,5} \subseteq [y_i^*]^{0,5} \forall i. \end{aligned} \tag{6.18}$$

Usando a Equação 6.3, podemos transformar as restrições  $[f^*(x_i)]^{0,5} \subseteq [y_i^*]^{0,5}$  em desigualdades. Para  $i = 3$ , por exemplo, temos que  $[f^*(x_3)]^{0,5} \subseteq [y_3^*]^{0,5}$ . Como

$f^*(x_3) = f^*(3) = \hat{\beta}_0^* + 3\hat{\beta}_1^* = (b_0 + 3b_1 - r_0 - 3r_1; b_0 + 3b_1; b_0 + 3b_1 + r_0 + 3r_1)$  e  $y_3^* = (6, 9; 9, 5; 12, 1) = (9, 5 - 2, 6; 9, 5; 9, 5 + 2, 6)$  podemos reescrever a restrição como:

$$\begin{cases} b_0 + 3b_1 \leq 9, 5 + (1 - 0, 5)(r_0 + 3r_1 - 2, 6) \\ b_0 + 3b_1 \geq 9, 5 - (1 - 0, 5)(r_0 + 3r_1 - 2, 6). \end{cases} \quad (6.19)$$

Do mesmo modo é possível reescrever as outras restrições e assim completar o problema de programação linear.

Obtemos, com ajuda computacional, o resultado  $\hat{\beta}_0^* = (-1, 42; 4, 15; 9, 72)$  e  $\hat{\beta}_1^* = (1, 97; 1, 97; 1, 97)$ . Estimando assim o modelo:

$$f^*(\vartheta) = (-1, 42; 4, 15; 9, 72) + \vartheta (1, 97; 1, 97; 1, 97). \quad (6.20)$$

A Figura 4(a) apresenta um gráfico comparando a função  $f^*(\vartheta)$  com os dados observados na Tabela 4. A escala de cinza nos gráficos apresenta a pertinência dos números fuzzy. Cada ponto  $(\vartheta, y)$  do gráfico apresenta a pertinência do ponto  $y$  no número fuzzy  $f^*(\vartheta)$ , ou seja,  $\varphi_{f^*(\vartheta)}(y)$ . Além da escala de cinza temos também as curvas de nível para 0, 0, 5 e 1.

Os dados observados  $y_i^* = (m_i - a_i; m_i; m_i + a_i)$  estão apresentados na forma de segmentos de reta verticais com um pequeno círculo no seu centro. Essas retas verticais são formadas pelos pontos  $(x_i, y)$  tais que a pertinência de  $y$  em  $y_i^*$  seja maior que zero, ou seja,  $\varphi_{y_i^*}(y) > 0$ . O pequeno círculo apresenta o “topo” do número fuzzy  $y_i^*$ , ou seja, é o ponto  $(x_i, m_i)$ . Essas barras não seguem a escala de cinza. A Tabela 5 também compara os valores observados com as estimativas correspondentes para  $\vartheta = x_i$ . Note como o modelo superestima os suportes dos dados observados, por exemplo,  $[y_2^*]^0 = [4, 2; 8, 6]$  é um intervalo de comprimento 4, 4 enquanto a estimativa dada por  $f^*(x_2)$  tem por suporte  $f^*(x_2) = [2, 52; 13, 66]$ , um intervalo de comprimento 11, 14. O coeficiente de determinação  $\overline{R^2} = [-2, 23; 1]$  aponta para uma baixa qualidade do modelo, pois apresenta valores negativos no seu intervalo.

Podemos analisar também os erros do modelo. Usando a diferença de Hukuhara generalizada, Equação 3.2, podemos calcular esses erros do modo discutido na Seção 6.5:

$$\begin{aligned} \underline{e}_1 &= [\hat{y}_1^*]^0 \ominus_{gH} [y_1^*]^0 = \\ &= [0, 55; 11, 69] \ominus_{gH} [6, 2; 9, 8] = \\ &= [\min\{0, 55 - 6, 2; 11, 69 - 9, 8\}; \max\{0, 55 - 6, 2; 11, 69 - 9, 8\}] = \\ &= [-5, 65; 1, 89]. \end{aligned}$$

$$\underline{e}_2 = [\hat{y}_2^*]^0 \ominus_{gH} [y_2^*]^0 =$$

$$\begin{aligned}
&= [2, 52; 13, 66] \ominus_{gH} [4, 2; 8, 6] = \\
&= [\min\{2, 52 - 4, 2; 13, 66 - 8, 6\}; \max\{2, 52 - 4, 2; 13, 66 - 8, 6\}] = \\
&= [-1, 68; 5, 06].
\end{aligned}$$

$$\begin{aligned}
\bar{e}_3 &= [\hat{y}_3^*]^0 \ominus_{gH} [y_3^*]^0 = \\
&= [4, 49; 15, 63] \ominus_{gH} [6, 9; 12, 1] = \\
&= [\min\{4, 49 - 6, 9; 15, 63 - 12, 1\}; \max\{4, 49 - 6, 9; 15, 63 - 12, 1\}] = \\
&= [-2, 41; 3, 53].
\end{aligned}$$

$$\begin{aligned}
\bar{e}_4 &= [\hat{y}_4^*]^0 \ominus_{gH} [y_4^*]^0 = \\
&= [6, 46; 17, 6] \ominus_{gH} [10, 9; 16, 1] = \\
&= [\min\{6, 46 - 10, 9; 17, 6 - 16, 1\}; \max\{6, 46 - 10, 9; 17, 6 - 16, 1\}] = \\
&= [-4, 44; 1, 5].
\end{aligned}$$

$$\begin{aligned}
\bar{e}_5 &= [\hat{y}_5^*]^0 \ominus_{gH} [y_5^*]^0 = \\
&= [8, 43; 19, 57] \ominus_{gH} [10, 6; 15, 4] = \\
&= [\min\{8, 43 - 10, 6; 19, 57 - 15, 4\}; \max\{8, 43 - 10, 6; 19, 57 - 15, 4\}] = \\
&= [-2, 17; 4, 17].
\end{aligned}$$

No geral todos esses erros são intervalos com grande comprimento. Isso já indica uma baixa qualidade do modelo. Podemos investigar também se esses erros estão normalmente distribuídos, apesar do modelo possibilístico *Min Problem* não se comprometer com esse aspecto. Para tal usamos o gráfico quantil-quantil, como proposto na Seção 6.5. A Figura 5(a) apresenta esse gráfico. É possível notar que os pontos médios dos intervalos que representam os erros não estão tão distantes da reta  $y = x$ . No entanto, os extremos dos intervalos se distanciam bastante desta reta. Interpretamos assim que esses erros não estão normalmente distribuídos.

**Exemplo 11** (Método Clássico Estendido). *Sejam os dados apresentados na Tabela 4. Como exemplo queremos encontrar o modelo clássico estendido da Equação 6.4 na Seção 6.2.*

Esse modelo é expresso na forma:

$$f^*(\vartheta) = \hat{\beta}_0^* + \vartheta \hat{\beta}_1^*. \quad (6.21)$$

Tabela 5 – Comparação das estimativas dadas pelo modelo do Exemplo 10, método possibilístico *Min Problem*, com as observações correspondentes da Tabela 4.

estimativa $f^*(\vartheta)$	dados observados	distância (Equação 6.10)
$f^*(x_1) = f^*(1) = (0, 55; 6, 12; 11, 69)$	(6, 2; 8, 0; 9, 8)	2, 88
$f^*(x_2) = f^*(2) = (2, 52; 8, 09; 13, 66)$	(4, 2; 6, 4; 8, 6)	2, 57
$f^*(x_3) = f^*(3) = (4, 49; 10, 06; 15, 63)$	(6, 9; 9, 5; 12, 1)	1, 80
$f^*(x_4) = f^*(4) = (6, 46; 12, 03; 17, 60)$	(10, 9; 13, 5; 16, 1)	2, 27
$f^*(x_5) = f^*(5) = (8, 43; 14, 00; 19, 57)$	(10, 6; 13, 0; 15, 4)	2, 08
		$\overline{R^2} = [-2, 23; 1]$

Fonte: O autor.

A matriz  $\mathbf{X}$  é dada por:

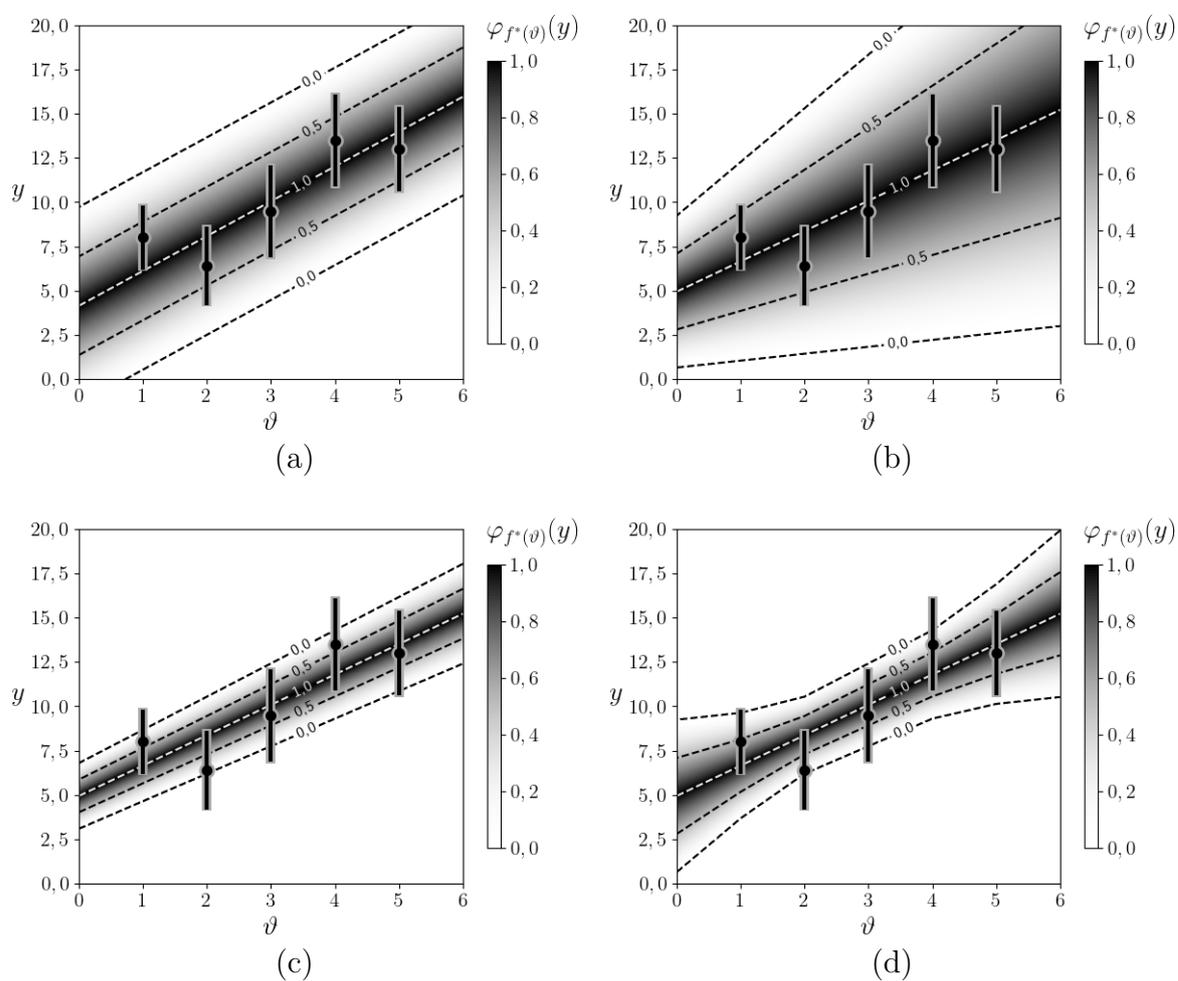
$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_5 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{pmatrix}, \quad (6.22)$$

com ela podemos encontrar a matriz  $\mathbf{W}$ :

$$\begin{aligned} \mathbf{W} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T &= \left( \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix} = \\ &= \begin{pmatrix} 5 & 15 \\ 15 & 55 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix} = \\ &= \begin{pmatrix} \frac{11}{10} & \frac{-3}{10} \\ \frac{-3}{10} & \frac{1}{10} \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix} = \\ &= \begin{pmatrix} \frac{8}{10} & \frac{5}{10} & \frac{2}{10} & \frac{-1}{10} & \frac{-4}{10} \\ \frac{-2}{10} & \frac{-1}{10} & \frac{0}{10} & \frac{1}{10} & \frac{2}{10} \end{pmatrix}. \end{aligned} \quad (6.23)$$

Figura 4 – Gráficos dos modelos  $f^*(\vartheta)$  dos Exemplos 10, 11, 12, e 14.

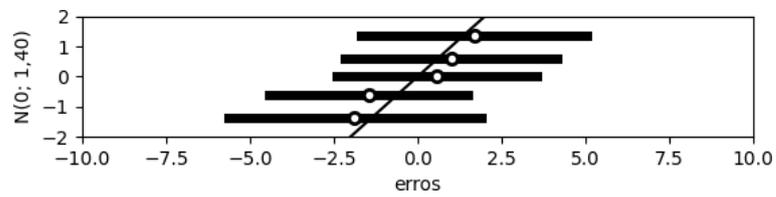
- (a) Exemplo 10.
- (b) Exemplo 11.
- (c) Exemplo 12.
- (d) Exemplo 14.



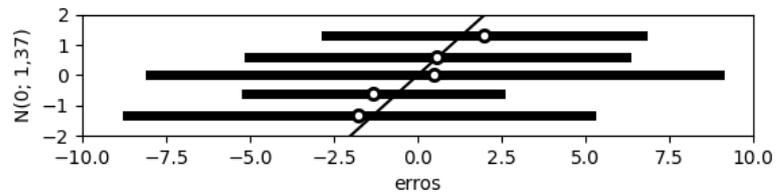
Fonte: O autor.

Figura 5 – Gráficos quantil-quantil dos erros dos modelos dos Exemplos 10, 11, 12, 13 e 14.

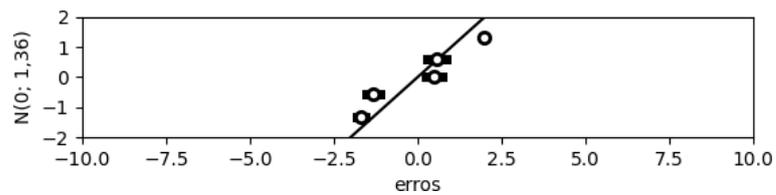
- (a) Exemplo 10 (*Min Problem*).
- (b) Exemplo 11 (clássico estendido).
- (c) Exemplo 12 (quadrados mínimos).
- (d) Exemplos 13 e 14 (análise sensitiva).



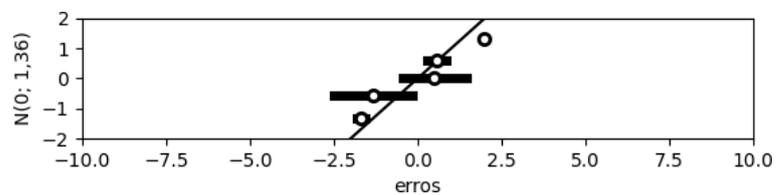
(a)



(b)



(c)



(d)

Fonte: O autor.

Assim podemos obter  $\hat{\beta}_0^*$  e  $\hat{\beta}_1^*$ :

$$\begin{aligned}
\hat{\beta}_0^* &= w_{1,1} y_1^* + w_{1,2} y_2^* + w_{1,3} y_3^* + w_{1,4} y_4^* + w_{1,5} y_5^* = \\
&= \frac{8}{10} y_1^* + \frac{5}{10} y_2^* + \frac{2}{10} y_3^* + \frac{-1}{10} y_4^* + \frac{-4}{10} y_5^* = \\
&= \frac{8}{10} (6, 2; 8, 0; 9, 8) + \frac{5}{10} (4, 2; 6, 4; 8, 6) + \frac{2}{10} (6, 9; 9, 5; 12, 1) + \\
&\quad + \frac{-1}{10} (10, 9; 13, 5; 16, 1) + \frac{-4}{10} (10, 6; 13, 0; 15, 4) = \\
&= \frac{8}{10} (6, 2; 8, 0; 9, 8) + \frac{5}{10} (4, 2; 6, 4; 8, 6) + \frac{2}{10} (6, 9; 9, 5; 12, 1) + \\
&\quad + \frac{-1}{10} (10, 9; 13, 5; 16, 1) + \frac{-4}{10} (10, 6; 13, 0; 15, 4) = \\
&= (4, 96; 6, 4; 7, 84) + (2, 10; 3, 20; 4, 3) + (1, 38; 1, 90; 2, 42) + \\
&\quad + (-1, 61; -1, 35; -1, 09) + (-6, 16; -5, 2; -4, 24) = \\
&= (0, 67; 4, 95; 9, 23) \text{ e} \\
\hat{\beta}_1^* &= w_{2,1} y_1^* + w_{2,2} y_2^* + w_{2,3} y_3^* + w_{2,4} y_4^* + w_{2,5} y_5^* = \\
&= \frac{-2}{10} y_1^* + \frac{-1}{10} y_2^* + \frac{0}{10} y_3^* + \frac{1}{10} y_4^* + \frac{2}{10} y_5^* = \\
&= (0, 39; 1, 71; 3, 03). \tag{6.24}
\end{aligned}$$

Temos assim o modelo:

$$f^*(\vartheta) = (0, 67; 4, 95; 9, 23) + \vartheta(0, 39; 1, 71; 3, 03). \tag{6.25}$$

A Figura 4(b) apresenta um gráfico comparando a função  $f^*(\vartheta)$  com os dados observados na Tabela 4. A Tabela 6 também compara os valores observados com as estimativas correspondentes para  $\vartheta = x_i$ . Note como o modelo superestima, neste caso, ainda mais que o modelo do Exemplo 10 os suportes dos dados observados, por exemplo,  $[y_5^*]^0 = [10, 6; 15, 4]$  é um intervalo de comprimento 4,8 enquanto a estimativa dada por  $f^*(x_5)$  tem por suporte  $[f^*(x_5)]^0 = [2, 62; 24, 38]$ , um intervalo de comprimento 21,76. Note que a distância entre as observações e as estimativas são maiores, como é visto na última coluna da Tabela 5 em comparação com a Tabela 6, isso é refletido na métrica do coeficiente de determinação  $\overline{R^2} = [-7, 02; 1]$  aponta para uma baixa qualidade do modelo, pois apresenta valores negativos no seu intervalo.

Para analisar os erros, calculamos do mesmo modo discutido na Seção 6.5:

$$\begin{aligned}
\overline{e}_1 &= [\hat{y}_1^*]^0 \ominus_{gH} [y_1^*]^0 = \\
&= [1, 06; 12, 26] \ominus_{gH} [6, 2; 9, 8] =
\end{aligned}$$

$$\begin{aligned}
&= [\min\{1, 06 - 6, 2; 12, 26 - 9, 8\}; \max\{1, 06 - 6, 2; 12, 26 - 9, 8\}] = \\
&= [-5, 14; 2, 46].
\end{aligned}$$

$$\begin{aligned}
\bar{e}_2 &= [\hat{y}_2^*]^0 \ominus_{gH} [y_2^*]^0 = \\
&= [1, 45; 15, 29] \ominus_{gH} [4, 2; 8, 6] = \\
&= [\min\{1, 45 - 4, 2; 15, 29 - 8, 6\}; \max\{1, 45 - 4, 2; 15, 29 - 8, 6\}] = \\
&= [-2, 75; 6, 69].
\end{aligned}$$

$$\begin{aligned}
\bar{e}_3 &= [\hat{y}_3^*]^0 \ominus_{gH} [y_3^*]^0 = \\
&= [1, 84; 18, 32] \ominus_{gH} [6, 9; 12, 1] = \\
&= [\min\{1, 84 - 6, 9; 18, 32 - 12, 1\}; \max\{1, 84 - 6, 9; 18, 32 - 12, 1\}] = \\
&= [-5, 06; 6, 22].
\end{aligned}$$

$$\begin{aligned}
\bar{e}_4 &= [\hat{y}_4^*]^0 \ominus_{gH} [y_4^*]^0 = \\
&= [2, 23; 21, 25] \ominus_{gH} [10, 9; 16, 1] = \\
&= [\min\{2, 23 - 10, 9; 21, 25 - 16, 1\}; \max\{2, 23 - 10, 9; 21, 25 - 16, 1\}] = \\
&= [-8, 67; 5, 15].
\end{aligned}$$

$$\begin{aligned}
\bar{e}_5 &= [\hat{y}_5^*]^0 \ominus_{gH} [y_5^*]^0 = \\
&= [2, 62; 24, 38] \ominus_{gH} [10, 6; 15, 4] = \\
&= [\min\{2, 62 - 10, 6; 24, 38 - 15, 4\}; \max\{2, 62 - 10, 6; 24, 38 - 15, 4\}] = \\
&= [-7, 98; 8, 98].
\end{aligned}$$

Esse modelo apresenta intervalos com comprimento ainda maiores que o exemplo anterior. Esse aspecto já é suficiente para levantar dúvidas quanto à qualidade do modelo. Na Figura 5(b) apresentamos o gráfico quantil-quantil desse conjunto de erros. Devido ao comprimento dos intervalos esse conjunto de erros possui muitos pontos longe da reta  $y = x$ , concluímos assim que esse conjunto de erros não está normalmente distribuído.

**Exemplo 12** (Quadrados Mínimos para Números Triangulares Fuzzy Simétricos). *Sejam os dados apresentados na Tabela 4. Como exemplo queremos encontrar o modelo de mínimos quadrados da Equação da Seção 6.3.*

Tabela 6 – Comparação das estimativas dadas pelo modelo do Exemplo 11, método clássico estendido, com as observações correspondentes da Tabela 4.

estimativa $f^*(\vartheta)$	dados observados	distância (Equação 6.10)
$f^*(x_1) = f^*(1) = (1, 06; 6, 66; 12, 26)$	(6, 2; 8, 0; 9, 8)	2, 57
$f^*(x_2) = f^*(2) = (1, 45; 8, 37; 15, 29)$	(4, 2; 6, 4; 8, 6)	3, 36
$f^*(x_3) = f^*(3) = (1, 84; 10, 08; 18, 32)$	(6, 9; 9, 5; 12, 1)	3, 31
$f^*(x_4) = f^*(4) = (2, 23; 11, 79; 21, 35)$	(10, 9; 13, 5; 16, 1)	4, 37
$f^*(x_5) = f^*(5) = (2, 62; 13, 50; 24, 38)$	(10, 6; 13, 0; 15, 4)	4, 92
		$\overline{R^2} = [-7, 02; 1]$

Fonte: O autor.

Esse modelo é dado por:

$$f^*(\vartheta) = \hat{\beta}_0^* + \vartheta \hat{\beta}_1^*, \quad (6.26)$$

com  $\hat{\beta}_0^* = (b_0 - r_0; b_0; b_0 + r_0)$  e  $\hat{\beta}_1^* = (b_1 - r_1; b_1; b_1 + r_1)$  números *fuzzy* triangulares simétricos.  $\hat{\beta}_0^*$  e  $\hat{\beta}_1^*$  são encontrados através da minimização:

$$\inf_{\hat{\beta}_0^*, \hat{\beta}_1^*} \frac{1}{5} \sum_1^5 d^2(\hat{\beta}_0^* + \hat{\beta}_1^* x_i, y_i^*), \quad (6.27)$$

essa minimização será realizada computacionalmente, porém, podemos exibir alguns passos para melhor entendimento. Podemos escrever:

$$\begin{aligned} & d^2(\hat{\beta}_0^* + \hat{\beta}_1^* x_i, y_i^*) = \\ & = d^2\left((b_0 - r_0; b_0; b_0 + r_0) + (b_1 - r_1; b_1; b_1 + r_1)x_i, (m_i - a_i; m_i; m_i + a_i)\right) = \\ & = d^2\left((b_0 - r_0; b_0; b_0 + r_0) + (x_i b_1 - |x_i| r_1; x_i b_1; x_i b_1 + |x_i| r_1), (m_i - a_i; m_i; m_i + a_i)\right) = \\ & = d^2\left((b_0 + x_i b_1 - (r_0 + |x_i| r_1); b_0 + x_i b_1; b_0 + x_i b_1 + (r_0 + |x_i| r_1)), (m_i - a_i; m_i; m_i + a_i)\right) = \\ & = (b_0 + x_i b_1 - m_i)^2 + \frac{(r_0 + |x_i| r_1 - a_i)^2}{3}. \end{aligned} \quad (6.28)$$

Deste modo temos:

$$\begin{aligned}
d^2(\hat{\beta}_0^* + \hat{\beta}_1^* x_1, y_1^*) &= (b_0 + 1b_1 - 8, 0)^2 + \frac{(r_0 + r_1 - 1, 8)^2}{3}, \\
d^2(\hat{\beta}_0^* + \hat{\beta}_1^* x_2, y_2^*) &= (b_0 + 2b_1 - 6, 4)^2 + \frac{(r_0 + 2r_1 - 2, 2)^2}{3}, \\
d^2(\hat{\beta}_0^* + \hat{\beta}_1^* x_3, y_3^*) &= (b_0 + 3b_1 - 9, 5)^2 + \frac{(r_0 + 3r_1 - 2, 6)^2}{3}, \\
d^2(\hat{\beta}_0^* + \hat{\beta}_1^* x_4, y_4^*) &= (b_0 + 4b_1 - 13, 5)^2 + \frac{(r_0 + 4r_1 - 2, 6)^2}{3} \text{ e} \\
d^2(\hat{\beta}_0^* + \hat{\beta}_1^* x_5, y_5^*) &= (b_0 + 5b_1 - 13, 0)^2 + \frac{(r_0 + 5r_1 - 2, 4)^2}{3}.
\end{aligned} \tag{6.29}$$

Ao se realizar a minimização computacionalmente, obtermos:

$$\hat{\beta}_0^* = (3, 11; 4, 95; 6, 79) \tag{6.30}$$

e

$$\hat{\beta}_1^* = (1, 55; 1, 71; 1, 87), \tag{6.31}$$

obtendo assim o modelo:

$$f^*(\vartheta) = (3, 11; 4, 95; 6, 79) + \vartheta (1, 55; 1, 71; 1, 87). \tag{6.32}$$

A Figura 4(c) apresenta um gráfico comparando a função  $f^*(\vartheta)$  com os dados observados na Tabela 4. A Tabela 7 também compara os valores observados com as estimativas correspondentes para  $\vartheta = x_i$ . Diferentemente dos modelos vistos nos Exemplos 10 e 11, aqui vemos um modelo que o comprimento do suporte das previsões são bem próximas ao dos dados observados. Essas características se refletem na métrica  $\overline{R^2} = [0, 67; 0, 83]$ . O coeficiente de determinação  $\overline{R^2}$  é um intervalo que contém todos os valores próximos de 1, o que indica que o modelo explica bem a variância dos dados observados.

Como definido na Seção 6.5, calculamos os erros na forma de intervalos:

$$\begin{aligned}
\overline{e}_1 &= [\hat{y}_1^*]^0 \ominus_{gH} [y_1^*]^0 = \\
&= [4, 66; 8, 66] \ominus_{gH} [6, 2; 9, 8] = \\
&= [\min\{4, 66 - 6, 2; 8, 66 - 9, 8\}; \max\{4, 66 - 6, 2; 8, 66 - 9, 8\}] = \\
&= [-1, 54; -1, 14].
\end{aligned}$$

$$\overline{e}_2 = [\hat{y}_2^*]^0 \ominus_{gH} [y_2^*]^0 =$$

$$\begin{aligned}
&= [6, 21; 10, 53] \ominus_{gH} [4, 2; 8, 6] = \\
&= [\min\{6, 21 - 4, 2; 10, 53 - 8, 6\}; \max\{6, 21 - 4, 2; 10, 53 - 8, 6\}] = \\
&= [1, 93; 2, 01].
\end{aligned}$$

$$\begin{aligned}
\bar{e}_3 &= [\hat{y}_3^*]^0 \ominus_{gH} [y_3^*]^0 = \\
&= [7, 76; 12, 4] \ominus_{gH} [6, 9; 12, 1] = \\
&= [\min\{7, 76 - 6, 9; 12, 4 - 12, 1\}; \max\{7, 76 - 6, 9; 12, 4 - 12, 1\}] = \\
&= [0, 3; 0, 86].
\end{aligned}$$

$$\begin{aligned}
\bar{e}_4 &= [\hat{y}_4^*]^0 \ominus_{gH} [y_4^*]^0 = \\
&= [9, 31; 14, 27] \ominus_{gH} [10, 9; 16, 1] = \\
&= [\min\{9, 31 - 10, 9; 14, 27 - 16, 1\}; \max\{9, 31 - 10, 9; 14, 27 - 16, 1\}] = \\
&= [-1, 83; -1, 59].
\end{aligned}$$

$$\begin{aligned}
\bar{e}_5 &= [\hat{y}_5^*]^0 \ominus_{gH} [y_5^*]^0 = \\
&= [10, 86; 16, 14] \ominus_{gH} [10, 6; 15, 4] = \\
&= [\min\{10, 86 - 10, 6; 16, 14 - 15, 4\}; \max\{10, 86 - 10, 6; 16, 14 - 15, 4\}] = \\
&= [0, 26; 0, 74].
\end{aligned}$$

Esse conjunto de erros possui comprimento bem menor que os anteriores. Na Figura 5(c) apresentamos o gráfico quantil-quantil deles. Observe que além dos intervalos serem curtos, todos os pontos estão bem próximos da reta  $y = x$ . Considerando que se trata de um conjunto com apenas 5 erros, podemos considerar satisfatória a proximidade dos erros à reta  $y = x$ , concluindo assim que a hipótese do modelo possuir erros normalmente distribuídos não pode ser descartada.

**Exemplo 13** (Análise Sensitiva). *Sejam os dados apresentados na Tabela 8. Eles são apenas uma adaptação dos dados da Tabela 4 tomando como dados os intervalos correspondentes ao  $\alpha$ -nível 0 dos dados. Como exemplo queremos encontrar o modelo por análise sensitiva da Equação 6.11.*

Ao contrário dos Exemplos 10, 11 e 12, esse modelo não é definido na forma de uma expressão. Aqui para cada  $\vartheta$  teremos um intervalo  $\bar{f}(\vartheta)$  dado por um processo infinito:

Tabela 7 – Comparação das estimativas dadas pelo modelo do Exemplo 12, quadrados mínimos para números triangulares *fuzzy* simétricos, com as observações correspondentes da Tabela 4.

estimativa $f^*(\vartheta)$	dados observados	distância (Equação 6.10)
$f^*(x_1) = f^*(1) = (4, 66; 6, 66; 8, 66)$	(6, 2; 8, 0; 9, 8)	1, 34
$f^*(x_2) = f^*(2) = (6, 21; 8, 37; 10, 53)$	(4, 2; 6, 4; 8, 6)	1, 97
$f^*(x_3) = f^*(3) = (7, 76; 10, 08; 12, 40)$	(6, 9; 9, 5; 12, 1)	0, 60
$f^*(x_4) = f^*(4) = (9, 31; 11, 79; 14, 27)$	(10, 9; 13, 5; 16, 1)	1, 71
$f^*(x_5) = f^*(5) = (10, 86; 13, 50; 16, 14)$	(10, 6; 13, 0; 15, 4)	0, 51
		$\overline{R^2} = [0, 67; 0, 83]$

Fonte: O autor.

Tabela 8 – Dados usados para o Exemplo 13. Os valores  $\underline{y}_i$  são intervalos reais.

$x_i$	$\underline{y}_i$
$x_1 = 1$	$\underline{y}_1 = [6, 2; 9, 8]$
$x_2 = 2$	$\underline{y}_2 = [4, 2; 8, 6]$
$x_3 = 3$	$\underline{y}_3 = [6, 9; 12, 1]$
$x_4 = 4$	$\underline{y}_4 = [10, 9; 16, 1]$
$x_5 = 5$	$\underline{y}_5 = [10, 6; 15, 4]$

Fonte: Tanaka, Hayashi e Watada (1989) (adaptado)

$$\underline{f}(\vartheta) = \left\{ \hat{a}\vartheta + \hat{b}; \right. \\ \left. \forall \hat{a}, \hat{b} \text{ que minimizem } \sum_{i=1}^n (ax_i + b - y_i)^2, \forall y_i \in [\underline{y}_i, \overline{y}_i], i = 1, \dots, n \right\},$$

Podemos, por exemplo, tomar os valores  $y_1 = 7 \in \underline{y}_1$ ,  $y_2 = 5 \in \underline{y}_2$ ,  $y_3 = 7 \in \underline{y}_3$ ,  $y_4 = 11 \in \underline{y}_4$  e  $y_5 = 11 \in \underline{y}_5$ , temos que a minimização de  $\sum_{i=1}^n (ax_i + b - y_i)^2$ , encontrada por mínimos quadrados, será dada por  $\hat{a} = 4$  e  $\hat{b} = \frac{7}{5}$ . Assim  $4\vartheta + \frac{7}{5} \in \underline{f}(\vartheta)$ . Porém, calcular todas as combinações de pontos é impossível. Apesar disto podemos encontrar o resultado através de problemas de programação linear.

Cada intervalo  $\underline{f}(\vartheta)$  pode ser calculado usando dois programas de otimização linear, um para cada extremo do intervalo. Como minimizar  $\sum_{i=1}^n (ax_i + b - y_i)^2$  é um problema de quadrados mínimos, podemos encontrar esse mínimo como é exibido na Seção 1.2, com a mesma matriz de planejamento apresentada no Exemplo 11. Podemos reescrever  $\underline{f}(\vartheta)$  como:

$$\begin{aligned} \underline{f}(\vartheta) = & \left\{ \hat{a}\vartheta + \hat{b}; \right. \\ & \text{com } \hat{a} = \frac{-2}{10}y_1 + \frac{-1}{10}y_2 + \frac{0}{10}y_3 + \frac{1}{10}y_4 + \frac{2}{10}y_5 \text{ e} \\ & \hat{b} = \frac{8}{10}y_1 + \frac{5}{10}y_2 + \frac{2}{10}y_3 + \frac{-1}{10}y_4 + \frac{-4}{10}y_5, \\ & \left. \forall y_i \in [\underline{y}_i, \overline{y}_i], i = 1, \dots, n \right\}, \end{aligned}$$

Assim o intervalo  $\underline{f}(\vartheta) = [f(\vartheta); \overline{f}(\vartheta)]$  onde  $f(\vartheta) = \frac{1}{10}(-2y_1 - y_2 + y_4 + 2y_5)\vartheta + \frac{1}{10}(8y_1 + 5y_2 + 2y_3 - y_4 - 4y_5)$  com os  $y_i$  dados por

$$\begin{aligned} \min_{y_1, y_2, y_3, y_4, y_5} & \frac{1}{10}(-2y_1 - y_2 + y_4 + 2y_5)\vartheta + \frac{1}{10}(8y_1 + 5y_2 + 2y_3 - y_4 - 4y_5) \\ \text{Sujeito a: } & \underline{y}_i \leq y_i \leq \overline{y}_i, i = 1, \dots, 5, \end{aligned}$$

e  $\overline{f}(\vartheta) = \frac{1}{10}(-2y_1 - y_2 + y_4 + 2y_5)\vartheta + \frac{1}{10}(8y_1 + 5y_2 + 2y_3 - y_4 - 4y_5)$  com os  $y_i$  dados por

$$\begin{aligned} \max_{y_1, y_2, y_3, y_4, y_5} & \frac{1}{10}(-2y_1 - y_2 + y_4 + 2y_5)\vartheta + \frac{1}{10}(8y_1 + 5y_2 + 2y_3 - y_4 - 4y_5) \\ \text{Sujeito a: } & \underline{y}_i \leq y_i \leq \overline{y}_i, i = 1, \dots, 5. \end{aligned}$$

Na Figura 6 apresentamos esse resultado em comparação com os intervalos observados. O coeficiente de determinação  $\overline{R}^2 = [0, 50; 0, 86]$  indica boa representação da variância dos dados.

Quanto aos erros, podemos calculá-los usando o definido na Seção 6.5:

$$\begin{aligned} \underline{e}_1 &= \underline{\hat{y}}_1 \ominus_{gH} \underline{y}_1 = \\ &= [3, 7; 9, 62] \ominus_{gH} [6, 2; 9, 8] = \\ &= [\min\{3, 7 - 6, 2; 9, 62 - 9, 8\}; \max\{3, 7 - 6, 2; 9, 62 - 9, 8\}] = \\ &= [-2, 5; -0, 18]. \end{aligned}$$

$$\begin{aligned} \underline{e}_2 &= \underline{\hat{y}}_2 \ominus_{gH} \underline{y}_2 = \\ &= [6, 21; 10, 53] \ominus_{gH} [4, 2; 8, 6] = \\ &= [\min\{6, 21 - 4, 2; 10, 53 - 8, 6\}; \max\{6, 21 - 4, 2; 10, 53 - 8, 6\}] = \\ &= [1, 93; 2, 01]. \end{aligned}$$

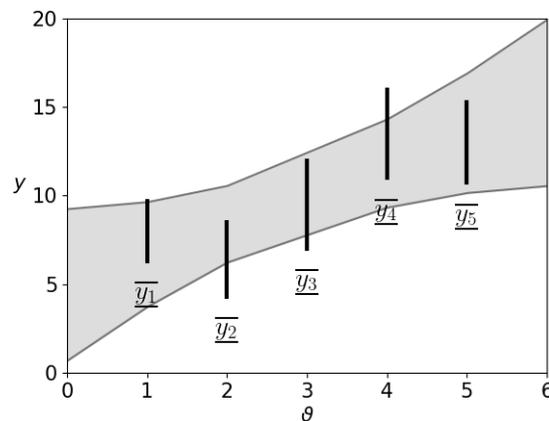
$$\begin{aligned} \bar{e}_3 &= \hat{y}_3 \ominus_{gH} \underline{y}_3 = \\ &= [7, 76; 12, 4] \ominus_{gH} [6, 9; 12, 1] = \\ &= [\min\{7, 76 - 6, 9; 12, 4 - 12, 1\}; \max\{7, 76 - 6, 9; 12, 4 - 12, 1\}] = \\ &= [0, 3; 0, 86]. \end{aligned}$$

$$\begin{aligned} \bar{e}_4 &= \hat{y}_4 \ominus_{gH} \underline{y}_4 = \\ &= [9, 31; 14, 27] \ominus_{gH} [10, 9; 16, 1] = \\ &= [\min\{9, 31 - 10, 9; 14, 27 - 16, 1\}; \max\{9, 31 - 10, 9; 14, 27 - 16, 1\}] = \\ &= [-1, 83; -1, 59]. \end{aligned}$$

$$\begin{aligned} \bar{e}_5 &= \hat{y}_5 \ominus_{gH} \underline{y}_5 = \\ &= [10, 14; 16, 86] \ominus_{gH} [10, 6; 15, 4] = \\ &= [\min\{10, 14 - 10, 6; 16, 86 - 15, 4\}; \max\{10, 14 - 10, 6; 16, 86 - 15, 4\}] = \\ &= [-0, 46; 1, 46]. \end{aligned}$$

No gráfico quantil-quantil, Figura 5(d), é possível perceber que os intervalos são próximos à reta  $y = x$ , assim a hipótese de que o conjunto de erros é normalmente distribuído não pode ser descartada.

Figura 6 – Gráfico do modelo do Exemplo 13, que modela os dados da Tabela 8 pela Equação 6.11. A faixa cinza do gráfico contém os pontos  $(\vartheta, y)$  tais que  $y \in \hat{f}(\vartheta)$ . As linhas verticais apresentam os intervalos observados  $\underline{y}_i$ , mais especificamente eles apresentam os pontos  $(x_i, y)$  tais que  $y \in \underline{y}_i$ .



Fonte: O autor.

**Exemplo 14** (Análise Sensitiva). *Sejam os dados apresentados na Tabela 4. Podemos extrapolar o método da Equação 6.11 aplicando o método para cada  $\alpha$ -nível.*

Obtemos o modelo apresentado na Figura 4(d). A Tabela 7 também compara os valores observados com as estimativas correspondentes para  $\vartheta = x_i$ .

Para esse exemplo esse modelo se mostra mais próximo dos dados observados do que os Exemplos 10 e 11, porém, não tão ajustados quanto os do Exemplo 12. O coeficiente de determinação  $\overline{R^2} = [0, 50; 0, 86]$  indica boa representação da variância dos dados. Como os erros são exatamente os mesmos do exemplo anterior, Exemplo 13, a análise dos erros é a mesma e concluímos que não podemos descartar a hipótese dos erros estarem normalmente distribuídos. A Figura 5(d) apresenta o gráfico quantil-quantil desse conjunto de erros.

Tabela 9 – Comparação das estimativas dadas pelo modelo do Exemplo 14, análise sensitiva, com as observações correspondentes da Tabela 4.

estimativa $f^*(\vartheta)$	dados observados	distância (Equação 6.10)
$f^*(x_1) = f^*(1) = (3, 70; 6, 66; 9, 62)$	(6, 2; 8, 0; 9, 8)	1, 50
$f^*(x_2) = f^*(2) = (6, 21; 8, 37; 10, 53)$	(4, 2; 6, 4; 8, 6)	1, 97
$f^*(x_3) = f^*(3) = (7, 76; 10, 08; 12, 40)$	(6, 9; 9, 5; 12, 1)	0, 60
$f^*(x_4) = f^*(4) = (9, 31; 11, 79; 14, 27)$	(10, 9; 13, 5; 16, 1)	1, 71
$f^*(x_5) = f^*(5) = (10, 14; 13, 50; 16, 86)$	(10, 6; 13, 0; 15, 4)	0, 75
		$\overline{R^2} = [0, 50; 0, 86]$

Fonte: O autor.

## 6.7 Conclusão

Na literatura encontramos uma variedade de métodos lineares que estendem os conceitos da regressão linear da estatística para os conceitos de dados intervalares ou na forma de números *fuzzy*. Alguns superestimam a imprecisão, como é o caso do método possibilístico *min problem* ou do método clássico estendido, além de apresentar o comportamento não natural das previsões longe da origem, tendo maior imprecisão, como citado por COUSO e DUBOIS (2014). Outros apresentam o problema de ambiguidade apontado por KORNER e NATHER (1998), como os problemas de quadrados mínimos que apresentam resultados distintos ao se reescalar os valores das variáveis explicativas, apesar da qualidade dos modelos se manter.

O método de análise sensitiva não possui esses problemas e oferece uma definição que facilita a extensão do modelo para casos mais complexos. É razoável, então, usá-lo como base para a construção de novos modelos. Isso não invalida a importância e a existência dos demais modelos, porém, a análise sensitiva será o foco para a construção de novos modelos ao longo deste trabalho.

## Parte II

# Proposta de Modelos para Dados Intervalares e *Fuzzy*

# 7 Proposta de Modelos Transversais para Dados Intervalares e *Fuzzy*

## 7.1 Introdução

No Capítulo 6 apresentamos diversos modelos de regressão para dados *fuzzy* ou intervalares presentes na literatura. Esses modelos são divididos em quatro categorias: modelos possibilísticos, modelos de quadrados mínimos, modelos de regressão quantílica e modelos de análise sensitiva.

Os modelos nos casos possibilístico e quadrados mínimos se diferenciam apenas na forma com que os coeficientes  $\hat{\beta}_i^*$  são encontrados e são apresentados na forma de uma função linear com coeficientes *fuzzy*

$$f^*(x_1, \dots, x_q) = \hat{\beta}_0^* + \hat{\beta}_1^* x_1 + \dots + \hat{\beta}_q^* x_q.$$

Já o modelo de análise sensitiva se apresenta na forma:

$$\underline{f}(\vartheta) = \left\{ \begin{array}{l} \hat{a}\vartheta + \hat{b}; \\ \forall \hat{a}, \hat{b} \text{ que minimize } \sum_{i=1}^n (ax_i + b - y_i)^2, \forall y_i \in [\underline{y}_i, \overline{y}_i], i = 1, \dots, n \end{array} \right\},$$

onde a função de regressão não é apresentada como uma função linear, mas sim na forma de um conjunto para cada valor da variável explicativa  $\vartheta$ . Mais explicitamente, o modelo de análise sensitiva é a união de todos os resultados de regressão linear por quadrados mínimos  $\hat{a}\vartheta + \hat{b}$  de todas as combinações de pontos tomados de cada intervalo,  $y_i \in [\underline{y}_i, \overline{y}_i]$ .

Esse modelo pode ser estendido, ou seja, a sua definição pode ser ampliada para outros modelos. Na Seção 7.4 apresentamos a extensão desse modelo para o modelo de regressão múltipla e no Capítulo 8 apresentamos a extensão para o modelo longitudinal.

Porém, partir direto da definição para calcular os conjuntos  $\underline{f}(\vartheta)$  não é eficiente, pois, mesmo usando apenas um conjunto finito de pontos de cada intervalo, o que nos levaria apenas a uma aproximação do resultado, se teria uma quantidade muito grande de operações matemáticas para se encontrar o resultado. Para o caso onde, por exemplo, o conjunto de dados fosse composto por 20 intervalos e usássemos 10 pontos de cada intervalo para realizar a aproximação precisaríamos resolver o problema de quadrados mínimos clássico  $10^{20}$  vezes. Se tratando de um valor bastante alto para uma aproximação que usa

uma quantidade de pontos bem pequena de cada intervalo. Assim resolver esse problema diretamente da definição não é eficiente de forma alguma. Uma maneira mais eficiente seria resolver, para cada  $\vartheta$ , dois problemas de programação linear, um para o máximo e outro para o mínimo, encontrando os extremos superior e inferior de cada intervalo  $\bar{f}(\vartheta)$ . Mesmo assim, ainda não temos garantia de que  $\bar{f}(\vartheta)$  seja de fato um intervalo.

O problema de encontrar os conjuntos  $\bar{f}(\vartheta)$  pode ser atacado de várias maneiras diferentes. Propomos neste capítulo uma forma de encontrar esses valores usando as operações básicas de multiplicação por escalar e soma de intervalos, chamamos essa forma de modelo base e o apresentamos na Seção 7.2. Essa abordagem não só permite encontrar  $\bar{f}(\vartheta)$  de forma mais eficiente que partindo da definição, ou até mesmo usando problemas de programação linear, como também permite usar o mesmo método para o caso em que os dados tenham a forma de números *fuzzy*.

## 7.2 Modelo Base

Esta seção apresenta um modelo base que pode ser usado para reescrever o modelo de análise sensítiva e construir modelos para o caso múltiplo e longitudinal. Inicialmente é necessário demonstrar a seguinte igualdade:

$$\sum_{i=1}^n \{a_i \bar{y}_i\} = \bigcup_{(y_1, \dots, y_n) \in \bar{y}_1 \times \dots \times \bar{y}_n} \left( \sum_{i=1}^n \{a_i y_i\} \right).$$

A seguir apresentamos alguns resultados preliminares necessários para demonstrar essa propriedade.

**Lema 1.** *Sejam o número real  $x$  e o intervalo real  $\bar{a} = [\underline{a}; \bar{a}]$ . Temos que:*

$$a \in \bar{a} \implies ax \in \bar{a}x. \quad (7.1)$$

*Demonstração.*

$$\begin{aligned} a \in \bar{a} \implies \underline{a} \leq a \leq \bar{a} \implies & \begin{cases} \underline{a}x \leq ax \leq \bar{a}x, \text{ se } x \geq 0 \\ \bar{a}x \leq ax \leq \underline{a}x, \text{ se } x < 0 \end{cases} \implies \\ \implies & \begin{cases} a \in [\underline{a}x; \bar{a}x], \text{ se } x \geq 0 \\ a \in [\bar{a}x; \underline{a}x], \text{ se } x < 0 \end{cases} \implies ax \in \bar{a}x \end{aligned}$$

□

**Lema 2.** *Sejam os intervalos reais  $\bar{a} = [\underline{a}; \bar{a}]$  e  $\bar{b} = [\underline{b}; \bar{b}]$ . Temos que:*

$$a \in \bar{a} \text{ e } b \in \bar{b} \implies a + b \in \bar{a} + \bar{b}. \quad (7.2)$$

*Demonstração.*

$$\begin{aligned} a \in \underline{a} \text{ e } b \in \underline{b} &\implies \begin{cases} \underline{a} \leq a \leq \bar{a} \\ \underline{b} \leq b \leq \bar{b} \end{cases} \implies \underline{a} + \underline{b} \leq a + b \leq \bar{a} + \bar{b} \implies \\ &\implies a + b \in [\underline{a} + \underline{b}; \bar{a} + \bar{b}] \implies a + b \in \underline{a} + \underline{b} \end{aligned}$$

□

**Lema 3.** *Sejam os números reais  $r$  e  $s$  e os intervalos reais  $\underline{a} = [\underline{a}; \bar{a}]$  e  $\underline{b} = [\underline{b}; \bar{b}]$ . Temos que:*

$$w \in \underline{a}r + \underline{b}s \implies \text{existem } a \in \underline{a} \text{ e } b \in \underline{b} \text{ tais que } w = ar + bs. \quad (7.3)$$

*Demonstração.* Tome o retângulo no plano cartesiano com o eixo das abscissas  $\vec{a}$  e o eixo das ordenadas  $\vec{b}$ . Neste plano tome a região formada pelo retângulo de vértices  $(\underline{a}, \underline{b})$ ,  $(\underline{a}, \bar{b})$ ,  $(\bar{a}, \bar{b})$  e  $(\bar{a}, \underline{b})$ .

Para  $r > 0$  e  $s > 0$ :

Se  $w \in \underline{a}r + \underline{b}s$ ,  $r > 0$  e  $s > 0$ , temos que:

$$\underline{a}r + \underline{b}s \leq w \leq \bar{a}r + \bar{b}s. \quad (7.4)$$

Tome as retas dadas pelas equações:

$$ar + bs = \underline{a}r + \underline{b}s, \quad (7.5)$$

$$ar + bs = w \quad (7.6)$$

e

$$ar + bs = \bar{a}r + \bar{b}s, \quad (7.7)$$

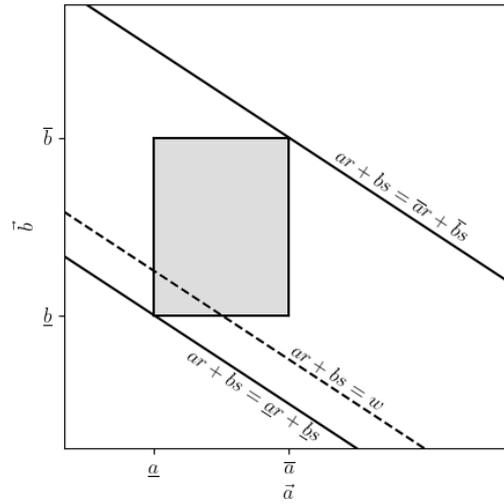
como  $\underline{a}r + \underline{b}s \leq w \leq \bar{a}r + \bar{b}s$ , a reta  $ar + bs = w$  está na região delimitada pelas outras duas retas. E do fato de  $r$  e  $s$  serem os coeficientes das três retas segue que elas são paralelas.

Note também que a reta  $ar + bs = \underline{a}r + \underline{b}s$  passa pelo ponto  $(\underline{a}, \underline{b})$  e a reta  $ar + bs = \bar{a}r + \bar{b}s$  passa pelo ponto  $(\bar{a}, \bar{b})$ . Como esses pontos são vértices opostos do retângulo formado pelos vértices  $(\underline{a}, \underline{b})$ ,  $(\underline{a}, \bar{b})$ ,  $(\bar{a}, \bar{b})$  e  $(\bar{a}, \underline{b})$  temos que esse está contido na região entre as retas  $ar + bs = \underline{a}r + \underline{b}s$  e  $ar + bs = \bar{a}r + \bar{b}s$  e como o retângulo é uma região contínua segue que a interseção entre ele e a reta  $ar + bs = w$  é não vazia. Provando assim o resultado para  $r > 0$  e  $s > 0$ . A Figura 7 apresenta um exemplo dessa situação.

Para outras combinações dos sinais de  $r$  e  $s$  a demonstração do resultado é análogo. Por exemplo, para  $r > 0$  e  $s < 0$  teremos  $\underline{a}r + \bar{b}s \leq w \leq \bar{a}r + \underline{b}s$  e os vértices tocados pelas retas serão os  $(\underline{a}, \bar{b})$  e  $(\bar{a}, \underline{b})$ .

□

Figura 7 – Imagem ilustrando a demonstração do Lema 3.



Fonte: O autor.

**Lema 4.** *Sejam os números reais  $r_i, i = 1, \dots, n$  e os intervalos reais  $\bar{a}_i = [\underline{a}_i; \bar{a}_i]$ . Temos que:*

$$\begin{aligned}
 & w \in \sum_{i=1}^n \bar{a}_i r_i, i = 1, \dots, n \implies \\
 \implies & \text{ existem } a_i \in \bar{a}_i, i = 1, \dots, n \text{ tais que } w = \sum_{i=1}^n a_i r_i.
 \end{aligned}
 \tag{7.8}$$

*Demonstração.* De forma análoga ao Lema 3, tome o espaço  $\mathbb{R}^n$  com coordenadas  $\bar{a}_i$ . Tome o hiperretângulo que contem os pontos  $(c_1, \dots, c_n)$  tais que  $\underline{a}_i \leq c_i \leq \bar{a}_i, i = 1, \dots, n$ .

Para  $r_i > 0, i = 1, \dots, n$ :

Tome os três hiperplanos:

$$\sum_1^n a_i r_i = \sum_1^n \underline{a}_i r_i, \tag{7.9}$$

$$\sum_1^n a_i r_i = w \tag{7.10}$$

e

$$\sum_1^n a_i r_i = \sum_1^n \bar{a}_i r_i. \tag{7.11}$$

Note que esses três hiperplanos são paralelos, com o hiperplano  $\sum_1^n a_i r_i = w$  estando no espaço delimitado pelos outros dois. O hiperplano  $\sum_1^n a_i r_i = \sum_1^n \underline{a}_i r_i$  passa pelo ponto

$(\underline{a}_1, \dots, \underline{a}_n)$  enquanto o hiperplano  $\sum_1^n a_i r_i = \sum_1^n \bar{a}_i r_i$  passa pelo ponto  $(\bar{a}_1, \dots, \bar{a}_n)$ . Esses pontos são dois vértices opostos do hiperretângulo, assim esse fica contido entre esses dois hiperplanos, tocando ambos e como o hiperretângulo é uma figura contínua ele terá interseção não vazia com o hiperplano  $\sum_1^n a_i r_i = w$  que também está entre os dois. Assim provamos o resultado para  $r_i > 0, i = 1, \dots, n$ .

Para outras combinações dos sinais de  $r_i$  a demonstração é análoga.

□

Com esses resultados é possível demonstrar a propriedade desejada.

**Proposição 1.** *Seja um conjunto de  $n$  intervalos  $\underline{y}_i$  e  $n$  números reais  $a_i$ , temos que:*

$$\sum_{i=1}^n \{a_i \underline{y}_i\} = \bigcup_{(y_1, \dots, y_n) \in \underline{y}_1 \times \dots \times \underline{y}_n} \left( \sum_{i=1}^n \{a_i y_i\} \right).$$

*Demonstração.* Para demonstrar esse resultado dividimos a proposição em duas inclusões de conjuntos.

$$\text{Primeiro demonstramos que } \sum_{i=1}^n \{a_i \underline{y}_i\} \subseteq \bigcup_{(y_1, \dots, y_n) \in \underline{y}_1 \times \dots \times \underline{y}_n} \left( \sum_{i=1}^n \{a_i y_i\} \right).$$

Seja  $w \in \sum_{i=1}^n \{a_i \underline{y}_i\}$  segue do Lema 4 que existem  $r_i \in \underline{y}_i, i = 1, \dots, n$  tais que  $w = \sum_{i=1}^n \{a_i r_i\}$ . Porém, como  $(r_1, \dots, r_n) \in \underline{y}_1 \times \dots \times \underline{y}_n$ , temos também que

$$\sum_{i=1}^n \{a_i r_i\} \subseteq \bigcup_{(y_1, \dots, y_n) \in \underline{y}_1 \times \dots \times \underline{y}_n} \left( \sum_{i=1}^n \{a_i y_i\} \right)$$

Segue assim que

$$w \in \sum_{i=1}^n \{a_i \underline{y}_i\} \implies w \in \bigcup_{(y_1, \dots, y_n) \in \underline{y}_1 \times \dots \times \underline{y}_n} \left( \sum_{i=1}^n \{a_i y_i\} \right),$$

o que demonstra essa primeira inclusão.

$$\text{A segunda inclusão é } \sum_{i=1}^n \{a_i \underline{y}_i\} \supseteq \bigcup_{(y_1, \dots, y_n) \in \underline{y}_1 \times \dots \times \underline{y}_n} \left( \sum_{i=1}^n \{a_i y_i\} \right).$$

Seja  $w \in \bigcup_{(y_1, \dots, y_n) \in \underline{y}_1 \times \dots \times \underline{y}_n} \left( \sum_{i=1}^n \{a_i y_i\} \right)$ , existem assim  $y_1 \in \underline{y}_1, \dots, y_n \in \underline{y}_n$  tais que  $w = \sum_{i=1}^n \{a_i y_i\}$ .

Segue do Lema 1 que

$$y_i \in \underline{y}_i \implies a_i y_i \in a_i \underline{y}_i,$$

e do Lema 2 que

$$\sum_{i=1}^n \{a_i y_i\} \in \sum_{i=1}^n \{a_i \underline{y}_i\}.$$

Segue assim que

$$w \in \bigcup_{(y_1, \dots, y_n) \in \underline{y}_1 \times \dots \times \underline{y}_n} \left( \sum_{i=1}^n \{a_i y_i\} \right) \implies w \in \sum_{i=1}^n \{a_i \underline{y}_i\},$$

o que demonstra essa segunda inclusão e completa a demonstração.  $\square$

Dados  $n$  intervalos  $\underline{y}_i$  e  $n$  funções reais  $f_i : \mathbb{R}^m \rightarrow \mathbb{R}$  nas  $m$  variáveis  $\vartheta_i$ . Definimos como modelo base:

$$\underline{f}(\vartheta_1, \dots, \vartheta_m) = \sum_{i=1}^n \{f_i(\vartheta_1, \dots, \vartheta_m) \underline{y}_i\}. \quad (7.12)$$

Lembrando que não existe a propriedade associativa para intervalos, logo esse somatório deve ser calculado sempre sem o uso dessa propriedade.

Segue da Proposição 1 que:

$$\sum_{i=1}^n \{f_i(\vartheta_1, \dots, \vartheta_m) \underline{y}_i\} = \bigcup_{(y_1, \dots, y_n) \in \underline{y}_1 \times \dots \times \underline{y}_n} \left( \sum_{i=1}^n \{f_i(\vartheta_1, \dots, \vartheta_m) y_i\} \right).$$

Ou seja, o modelo base é equivalente à união de todas as possibilidades de soma  $\sum_{i=1}^n \{f_i(\vartheta_1, \dots, \vartheta_m) y_i\}$  onde  $y_i \in \underline{y}_i$ . Desta forma qualquer modelo que tenha como definição a união dos resultados de todas as possibilidades de escolha de pontos  $y_i$  de intervalos  $\underline{y}_i$  e que possa ser reescrito na forma  $\sum_{i=1}^n \{f_i(\vartheta_1, \dots, \vartheta_m) \underline{y}_i\}$  permite usar essa equação para calcular o modelo.

Como as operações aritméticas dos números *fuzzy* vem diretamente das operações aritméticas de intervalos, como visto na Seção 4.3, o modelo base pode ser reescrito para o caso em que são dados  $n$  números *fuzzy* e  $n$  funções reais  $f_i : \mathbb{R}^m \rightarrow \mathbb{R}$  nas  $m$  variáveis  $\vartheta_i$ :

$$f^*(\vartheta_1, \dots, \vartheta_m) = \sum_{i=1}^n \{f_i(\vartheta_1, \dots, \vartheta_m) y_i^*\}. \quad (7.13)$$

Em termos dos  $\alpha$ -níveis, segue da Proposição 1:

$$\begin{aligned} & [f^*(\vartheta_1, \dots, \vartheta_m)]^\alpha = \\ & \sum_{i=1}^n \{f_i(\vartheta_1, \dots, \vartheta_m) [y_i^*]^\alpha\} = \\ & \bigcup_{(y_1, \dots, y_n) \in [y_1^*]^\alpha \times \dots \times [y_n^*]^\alpha} \left( \sum_{i=1}^n \{f_i(\vartheta_1, \dots, \vartheta_m) [y_i^*]^\alpha\} \right). \end{aligned}$$

Nas próximas seções o modelo base, tanto na forma intervalar como na forma de números *fuzzy*, é usada para definir, para os casos intervalar e *fuzzy*, o modelo linear simples e múltiplo assim como o modelo linear em dois estágios para dados longitudinais.

### 7.3 Modelo Linear Simples para Dados Intervalares e Fuzzy

Na Seção 1.2 é apresentado o modelo linear simples, onde um conjunto de dados na forma  $\{x_i; y_i\}, i = 1, \dots, n$  é modelado através de uma função  $f(\vartheta) = \hat{\beta}_0 + \hat{\beta}_1\vartheta$ . Para encontrar os termos  $\hat{\beta}_0$  e  $\hat{\beta}_1$  usamos a matriz de planejamento  $\mathbf{X}$  e o vetor  $\mathbf{y}$ :

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

e

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Denotando

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix},$$

obtemos:

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}.$$

Suponha agora que os dados observados tem a forma de intervalos, ou seja:

$$\{x_i; \overline{y_i}\}, i = 1, \dots, n.$$

O modelo de análise sensitiva, Equação 6.11, se apresenta na forma:

$$\underline{f}(\vartheta) = \left\{ \hat{a}\vartheta + \hat{b}; \right. \\ \left. \forall \hat{a}, \hat{b} \text{ que minimize } \sum_{i=1}^n (ax_i + b - y_i)^2, \forall y_i \in [\underline{y}_i, \overline{y}_i], i = 1, \dots, n \right\},$$

onde  $\underline{f}(\vartheta)$  é a união de todos os resultados da regressão linear simples para todas as possibilidades de escolha de  $y_i \in \underline{y}_i$ .

Esse modelo pode ser reescrito de forma a atender ao modelo base  $\sum_{i=1}^n \{f_i(\vartheta) \underline{y}_i\}$ . Desta forma a união de todas as possibilidades se resume a calcular o intervalo resultante do modelo base.

Para reescrever esse modelo, inicialmente reescrevemos o modelo clássico  $f(\vartheta) = \hat{\beta}_0 + \hat{\beta}_1\vartheta$ .

$$f(\vartheta) = \hat{\beta}_0 + \hat{\beta}_1\vartheta = \hat{\beta}^T \begin{pmatrix} 1 \\ \vartheta \end{pmatrix} = \left( (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \right)^T \begin{pmatrix} 1 \\ \vartheta \end{pmatrix}.$$

Seja  $\mathbf{W} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ , temos:

$$\left( (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \right)^T \begin{pmatrix} 1 \\ \vartheta \end{pmatrix} = (\mathbf{W} \mathbf{y})^T \begin{pmatrix} 1 \\ \vartheta \end{pmatrix} = \mathbf{y}^T \mathbf{W}^T \begin{pmatrix} 1 \\ \vartheta \end{pmatrix} = \\ \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}^T \begin{pmatrix} w_{1,1} + \vartheta w_{2,1} \\ w_{1,2} + \vartheta w_{2,2} \\ w_{1,3} + \vartheta w_{2,3} \\ \vdots \\ w_{1,n} + \vartheta w_{2,n} \end{pmatrix} = \sum_{i=1}^n \{(w_{1,i} + \vartheta w_{2,i}) y_i\}.$$

Definindo  $f_i(\vartheta) = w_{1,i} + \vartheta w_{2,i}$  obtemos:

$$f(\vartheta) = \sum_{i=1}^n \{(w_{1,i} + \vartheta w_{2,i}) y_i\} = \sum_{i=1}^n \{f_i(\vartheta) y_i\}.$$

Com essa expressão podemos reescrever o modelo de análise sensitiva como:

$$\underline{f}(\vartheta) = \bigcup_{(y_1, \dots, y_n) \in \underline{y}_1 \times \dots \times \underline{y}_n} \left( \sum_{i=1}^n \{f_i(\vartheta) y_i\} \right).$$

Segue então da Proposição 1 que:

$$\bar{f}(\vartheta) = \sum_{i=1}^n \{f_i(\vartheta) \underline{y}_i\}.$$

Com isso podemos definir o modelo para o caso intervalar. Note que este é o modelo da Equação 6.11 reescrito de outra forma. Seja um conjunto de dados na forma  $\{x_i; \underline{y}_i\}$ ,  $i = 1, \dots, n$ . O modelo linear simples para dados intervalares é dado por:

$$\bar{S}(\vartheta) = \sum_{i=1}^n \{(w_{1,i} + \vartheta w_{2,i}) \underline{y}_i\}, \quad (7.14)$$

onde

$$\mathbf{W} = \begin{pmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,n} \\ w_{2,1} & w_{2,2} & \dots & w_{2,n} \end{pmatrix} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T,$$

com

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}.$$

Do mesmo modo podemos definir o modelo com os dados na forma de números *fuzzy*. Seja um conjunto de dados na forma  $\{x_i; y_i^*\}$ ,  $i = 1, \dots, n$ . O modelo linear simples para dados *fuzzy* é dado por:

$$S^*(\vartheta) = \sum_{i=1}^n \{(w_{1,i} + \vartheta w_{2,i}) y_i^*\}, \quad (7.15)$$

onde  $\mathbf{W} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ , com

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}.$$

### 7.3.1 Exemplos

Tendo definidos o modelo linear simples para dados intervalares e para dados *fuzzy*, equações 7.14 e 7.15, apresentamos dois exemplos com a intenção de apresentar somente como o modelo funciona para fins didáticos. Logo não há intenção de interpretar os modelos aqui. Interpretações e reflexões estão presentes sobre os modelos apresentados nas aplicações da Parte III.

**Exemplo 15** (modelo linear simples para dados intervalares). *Sejam os dados apresentados na Tabela 10.*

*Como exemplo queremos encontrar o modelo linear simples para dados intervalares da Equação 7.14.*

Tabela 10 – Dados usados para o Exemplo 15. Os valores  $\underline{y}_i$  são intervalos reais.

$x_i$	$\underline{y}_i$
$x_1 = 1$	$\underline{y}_1 = [6, 2; 9, 8]$
$x_2 = 2$	$\underline{y}_2 = [4, 2; 8, 6]$
$x_3 = 3$	$\underline{y}_3 = [6, 9; 12, 1]$
$x_4 = 4$	$\underline{y}_4 = [10, 9; 16, 1]$
$x_5 = 5$	$\underline{y}_5 = [10, 6; 15, 4]$

Fonte: Tanaka, Hayashi e Watada (1989) (adaptado)

A matriz  $\mathbf{X}$  é dada por:

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ 1 & x_5 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{pmatrix}, \quad (7.16)$$

com ela podemos encontrar a matriz  $\mathbf{W}$ :

$$\mathbf{W} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T = \begin{pmatrix} \frac{8}{10} & \frac{5}{10} & \frac{2}{10} & \frac{-1}{10} & \frac{-4}{10} \\ -2 & -1 & 0 & 1 & 2 \\ \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} \end{pmatrix}. \quad (7.17)$$

Assim temos o modelo:

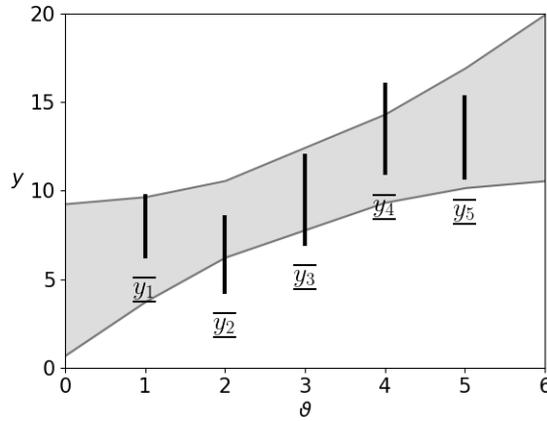
$$\begin{aligned} \underline{S}(\vartheta) &= (w_{1,1} + \vartheta w_{2,1}) \underline{y}_1 + \dots + (w_{1,5} + \vartheta w_{2,5}) \underline{y}_n = \\ &= \left( \frac{8}{10} + \vartheta \times \frac{-2}{10} \right) \underline{y}_1 + \left( \frac{5}{10} + \vartheta \times \frac{-1}{10} \right) \underline{y}_2 + \left( \frac{2}{10} + \vartheta \times \frac{0}{10} \right) \underline{y}_3 + \\ &+ \left( \frac{-1}{10} + \vartheta \times \frac{1}{10} \right) \underline{y}_4 + \left( \frac{-4}{10} + \vartheta \times \frac{2}{10} \right) \underline{y}_5. \end{aligned} \quad (7.18)$$

Podemos, por exemplo, calcular  $\underline{S}(2)$ :

$$\begin{aligned} \underline{S}(2) &= \left( \frac{8}{10} + 2 \times \frac{-2}{10} \right) \underline{y}_1 + \left( \frac{5}{10} + 2 \times \frac{-1}{10} \right) \underline{y}_2 + \left( \frac{2}{10} + 2 \times \frac{0}{10} \right) \underline{y}_3 + \\ &+ \left( \frac{-1}{10} + 2 \times \frac{1}{10} \right) \underline{y}_4 + \left( \frac{-4}{10} + 2 \times \frac{2}{10} \right) \underline{y}_5 = \\ &= \left( \frac{4}{10} \right) \underline{y}_1 + \left( \frac{3}{10} \right) \underline{y}_2 + \left( \frac{2}{10} \right) \underline{y}_3 + \left( \frac{1}{10} \right) \underline{y}_4 + \left( \frac{0}{10} \right) \underline{y}_5 = \\ &= [6, 21; 10, 53] \end{aligned} \quad (7.19)$$

A Figura 8 mostra o gráfico desse exemplo.

Figura 8 – Gráfico do modelo do Exemplo 15. A faixa cinza do gráfico contém os pontos  $(\vartheta, y)$  tais que  $y \in \overline{S}(\vartheta)$ . As linhas verticais apresentam os intervalos observados  $\overline{y}_i$ , mais especificamente eles apresentam os pontos  $(x_i, y)$  tais que  $y \in \overline{y}_i$ . Esse gráfico coincide com o gráfico da Figura 6.



Fonte: O autor.

Tabela 11 – Dados usados para o Exemplo 16.

$x_i$	$y_i^* = (m_i - a_i; m_i; m_i + a_i)$
$x_1 = 1$	$y_1^* = (6, 2; 8, 0; 9, 8)$
$x_2 = 2$	$y_2^* = (4, 2; 6, 4; 8, 6)$
$x_3 = 3$	$y_3^* = (6, 9; 9, 5; 12, 1)$
$x_4 = 4$	$y_4^* = (10, 9; 13, 5; 16, 1)$
$x_5 = 5$	$y_5^* = (10, 6; 13, 0; 15, 4)$

Fonte: Tanaka, Hayashi e Watada (1989)

**Exemplo 16** (modelo linear simples para dados fuzzy). *Sejam os dados apresentados na Tabela 11.*

*Como exemplo queremos encontrar o modelo linear simples para dados fuzzy da Equação 7.15.*

*Como no Exemplo 15, obtemos a matriz W:*

$$W = \begin{pmatrix} \frac{8}{10} & \frac{5}{10} & \frac{2}{10} & \frac{-1}{10} & \frac{-4}{10} \\ -2 & -1 & 0 & 1 & 2 \\ \frac{-1}{10} & \frac{1}{10} & \frac{-4}{10} & \frac{2}{10} & \frac{2}{10} \end{pmatrix}. \tag{7.20}$$

*Assim o modelo toma a forma:*

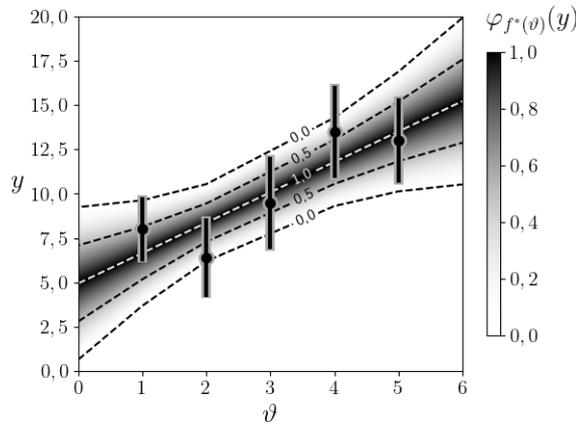
$$\begin{aligned} S^*(\vartheta) &= (w_{1,1} + \vartheta w_{2,1}) y_1^* + \dots + (w_{1,5} + \vartheta w_{2,5}) y_n^* = \\ &= \left( \frac{8}{10} + \vartheta \times \frac{-2}{10} \right) y_1^* + \left( \frac{5}{10} + \vartheta \times \frac{-1}{10} \right) y_2^* + \left( \frac{2}{10} + \vartheta \times \frac{0}{10} \right) y_3^* + \\ &+ \left( \frac{-1}{10} + \vartheta \times \frac{1}{10} \right) y_4^* + \left( \frac{-4}{10} + \vartheta \times \frac{2}{10} \right) y_5^*. \end{aligned} \tag{7.21}$$

Podemos, por exemplo, calcular:

$$\begin{aligned}
 S^*(1) &= \left(\frac{8}{10} + 1 \times \frac{-2}{10}\right) y_1^* + \left(\frac{5}{10} + 1 \times \frac{-1}{10}\right) y_2^* + \left(\frac{2}{10} + 1 \times \frac{0}{10}\right) y_3^* + \\
 &+ \left(\frac{-1}{10} + 1 \times \frac{1}{10}\right) y_4^* + \left(\frac{-4}{10} + 1 \times \frac{2}{10}\right) y_5^* = \\
 &= \left(\frac{6}{10}\right) y_1^* + \left(\frac{4}{10}\right) y_2^* + \left(\frac{2}{10}\right) y_3^* + \left(\frac{0}{10}\right) y_4^* + \left(\frac{-2}{10}\right) y_5^* = \\
 &= (3, 70; 6, 66; 9, 62),
 \end{aligned} \tag{7.22}$$

valor próximo ao valor de  $y_1^* = (6, 2; 8, 0; 9, 8)$ . A Figura 4 mostra o gráfico desse exemplo.

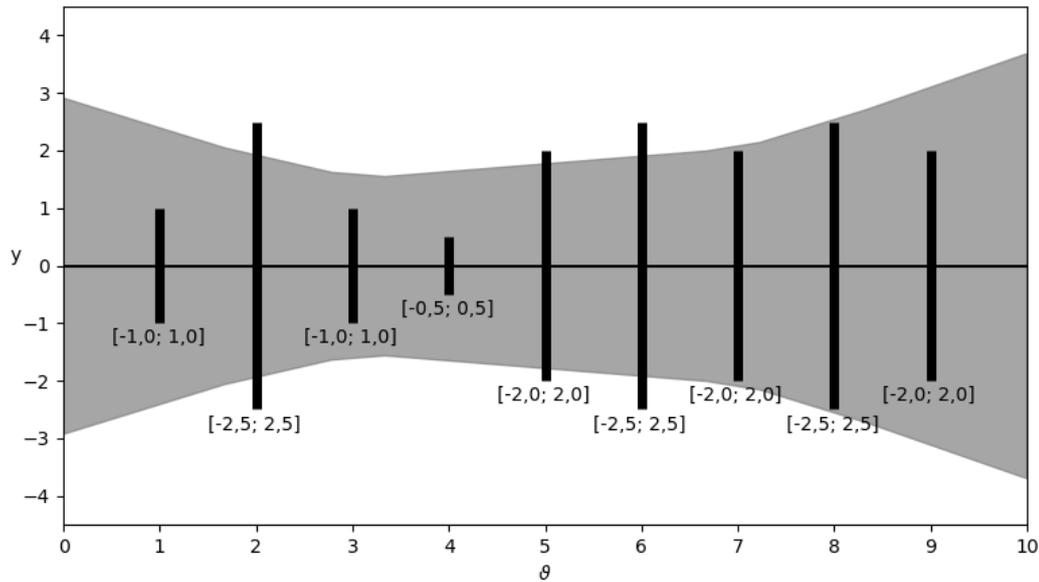
Figura 9 – Gráfico do modelo  $S^*(\vartheta)$  do Exemplo 16. A escala de cinza nos gráficos apresenta a pertinência dos números fuzzy. Cada ponto  $(\vartheta, y)$  do gráfico apresenta a pertinência do ponto  $y$  no número fuzzy  $S^*(\vartheta)$ , ou seja,  $\varphi_{S^*(\vartheta)}(y)$ . Além da escala de cinza temos também as curvas de nível para 0, 0,5 e 1. Os dados observados  $y_i^* = (m_i - a_i; m_i; m_i + a_i)$  estão apresentados na forma de segmentos de reta verticais com um pequeno círculo no seu centro. Essas retas verticais são formadas pelos pontos  $(x_i, y)$  tais que a pertinência de  $y$  em  $y_i^*$  seja maior que zero, ou seja,  $\varphi_{y_i^*}(y) > 0$ . O pequeno círculo apresenta o “topo” do número fuzzy  $y_i^*$ , ou seja, é o ponto  $(x_i, m_i)$ . Essas barras não seguem a escala de cinza. Esse gráfico coincide com o da Figura 4(d) da Seção 6.6.



Fonte: O autor.

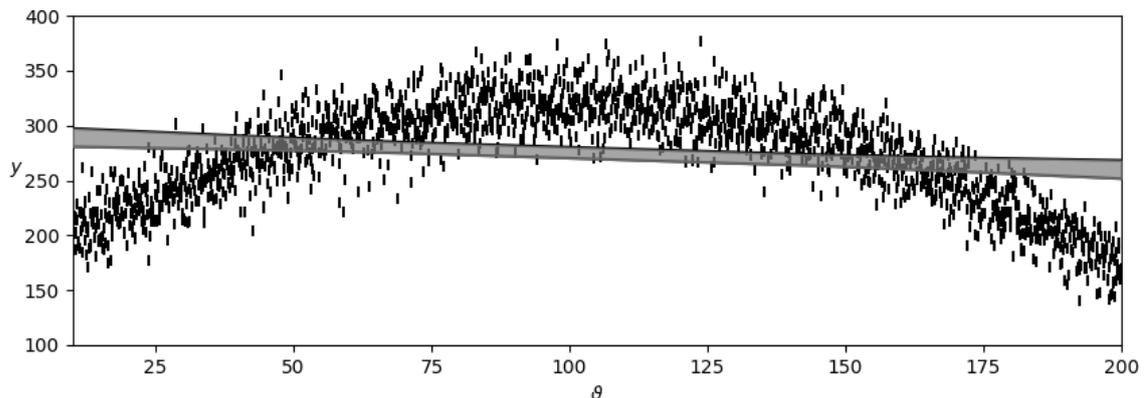
A Figura 10 apresenta o gráfico para um modelo linear com 9 dados intervalares tais que o centro de cada um dos intervalos fica na origem, exemplificando que esse caso não é exceção. Já a Figura 11 apresenta o gráfico para um conjunto de dados com 1900 intervalos, note que claramente o modelo linear simples não é o mais adequado para esse conjunto de dados e que o conjunto de intervalos não forma uma sequência monótona. Com esses dois exemplos reforçamos o fato do método estar bem definido para qualquer conjuntos de dados intervalares.

Figura 10 – Gráfico para um modelo linear com 9 dados intervalares centrados na origem.



Fonte: O autor.

Figura 11 – Gráfico para um modelo com 1900 dados intervalares.



Fonte: O autor.

## 7.4 Modelo Linear Múltiplo para Dados Intervalares e Fuzzy

Na Seção 1.3 é apresentado o método linear múltiplo, onde um conjunto de dados na forma  $\{x_{i,1}; \dots; x_{i,q}; y_i\}$ ,  $i = 1, \dots, n$  é modelado através da função  $f(\vartheta_1, \dots, \vartheta_q) = \hat{\beta}_0 + \hat{\beta}_1\vartheta_1 + \dots + \hat{\beta}_q\vartheta_q$ . Para encontrar os termos  $\hat{\beta}_0, \dots, \hat{\beta}_q$  usamos a matriz de planejamento  $\mathbf{X}$  e o vetor  $\mathbf{y}$ :

$$\mathbf{X} = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{1,q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \dots & x_{n,q} \end{pmatrix}$$

e

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Denotando

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_q \end{pmatrix},$$

obtemos:

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_q \end{pmatrix} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}.$$

Usando

$$\mathbf{W} = \begin{pmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,n} \\ w_{2,1} & w_{2,2} & \dots & w_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{q,1} & w_{q,2} & \dots & w_{q,n} \\ w_{q+1,1} & w_{q+1,2} & \dots & w_{q+1,n} \end{pmatrix} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T,$$

reescrevemos o modelo:

$$\begin{aligned} f(\vartheta_1, \dots, \vartheta_q) &= \hat{\beta}_0 + \hat{\beta}_1 \vartheta_1 + \dots + \hat{\beta}_q \vartheta_q = \\ &= \hat{\boldsymbol{\beta}}^T \begin{pmatrix} 1 \\ \vartheta_1 \\ \vdots \\ \vartheta_q \end{pmatrix} = \left( (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \right)^T \begin{pmatrix} 1 \\ \vartheta_1 \\ \vdots \\ \vartheta_q \end{pmatrix} = (\mathbf{W} \mathbf{y})^T \begin{pmatrix} 1 \\ \vartheta_1 \\ \vdots \\ \vartheta_q \end{pmatrix} = \mathbf{y}^T \mathbf{W}^T \begin{pmatrix} 1 \\ \vartheta_1 \\ \vdots \\ \vartheta_q \end{pmatrix} = \\ &= \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}^T \begin{pmatrix} w_{1,1} + \vartheta_1 w_{2,1} + \dots + \vartheta_q w_{q+1,1} \\ w_{1,2} + \vartheta_1 w_{2,2} + \dots + \vartheta_q w_{q+1,2} \\ w_{1,3} + \vartheta_1 w_{2,3} + \dots + \vartheta_q w_{q+1,3} \\ \vdots \\ w_{1,n} + \vartheta_1 w_{2,n} + \dots + \vartheta_q w_{q+1,n} \end{pmatrix} = \sum_{i=1}^n \{ (w_{1,i} + \vartheta_1 w_{2,i} + \dots + \vartheta_q w_{q+1,i}) y_i \}. \end{aligned}$$

Assim temos a regressão linear múltipla reescrita na forma:

$$f(\vartheta_1, \dots, \vartheta_q) = \sum_{i=1}^n \{ (w_{1,i} + \vartheta_1 w_{2,i} + \dots + \vartheta_q w_{q+1,i}) y_i \} \quad (7.23)$$

Tendo reescrito dessa forma o modelo clássico de regressão linear múltipla podemos definir o modelo para o caso em que os dados observados tem a forma de

intervalos. Dado um conjunto de  $n$  dados  $\{x_{i,1}; \dots; x_{i,q}; \underline{y}_i\}$ ,  $i = 1, \dots, n$ , tomamos como modelo a união de todas as possibilidades de regressão linear múltipla ao se tomar um elemento  $y_i$  pertencente a cada intervalo  $\underline{y}_i$ , ou seja,

$$\underline{S}(\vartheta_1, \dots, \vartheta_q) = \left\{ \begin{array}{l} \hat{a}_0 + \hat{a}_1\vartheta_1 + \dots + \hat{a}_q\vartheta_q; \\ \forall \hat{a}_0, \dots, \hat{a}_q \text{ que minimize } \sum_{i=1}^n (\hat{a}_0 + \hat{a}_1\vartheta_1 + \dots + \hat{a}_q\vartheta_q - y_i)^2, \\ \forall y_i \in \underline{y}_i, i = 1, \dots, n \end{array} \right\},$$

onde  $\underline{S}(\vartheta_1, \dots, \vartheta_q)$  é a união de todos os resultados da regressão linear múltipla para todas as possibilidades de escolha de  $y_i \in \underline{y}_i$ .

Usando a Equação 7.23 podemos reescrever como:

$$\underline{S}(\vartheta_1, \dots, \vartheta_q) = \bigcup_{(y_1, \dots, y_n) \in \underline{y}_1 \times \dots \times \underline{y}_n} \left( \sum_{i=1}^n \{(w_{1,i} + \vartheta_1 w_{2,i} + \dots + \vartheta_q w_{q+1,i}) y_i\} \right).$$

Porém, pela Proposição 1, esse modelo é equivalente a

$$\underline{S}(\vartheta_1, \dots, \vartheta_q) = \sum_{i=1}^n \{(w_{1,i} + \vartheta_1 w_{2,i} + \dots + \vartheta_q w_{q+1,i}) \underline{y}_i\}.$$

Desta forma definimos o modelo linear múltiplo para dados intervalares. Seja um conjunto de dados na forma  $\{x_{i,1}; \dots; x_{i,q}; \underline{y}_i\}$ ,  $i = 1, \dots, n$ , o modelo linear múltiplo para dados intervalares é dado por:

$$\underline{S}(\vartheta_1, \dots, \vartheta_q) = \sum_{i=1}^n \{(w_{1,i} + \vartheta_1 w_{2,i} + \dots + \vartheta_q w_{q+1,i}) \underline{y}_i\}, \quad (7.24)$$

onde

$$\mathbf{W} = \begin{pmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,n} \\ w_{2,1} & w_{2,2} & \dots & w_{2,n} \\ \vdots & & & \\ w_{q+1,1} & w_{q+1,2} & \dots & w_{q+1,n} \end{pmatrix} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T,$$

com

$$\mathbf{X} = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{1,q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \dots & x_{n,q} \end{pmatrix}.$$

Do mesmo modo podemos definir o modelo com os dados na forma de números *fuzzy*. Seja um conjunto de dados na forma  $\{x_{i,1}; \dots; x_{i,q}; y_i^*\}$ ,  $i = 1, \dots, n$ , o modelo linear

simples para dados *fuzzy* é dado por:

$$S^*(\vartheta_1, \dots, \vartheta_q) = \sum_{i=1}^n \{(w_{1,i} + \vartheta_1 w_{2,i} + \dots + \vartheta_q w_{q+1,i}) y_i^*\}, \quad (7.25)$$

onde  $\mathbf{W} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ , com

$$\mathbf{X} = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{1,q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \dots & x_{n,q} \end{pmatrix}.$$

## 7.5 Interpretando o Modelo como uma Média Ponderada

É importante interpretar o significado dos modelos. O modelo linear simples para dados intervalares,  $\underline{S}(\vartheta) = \sum_{i=1}^n \{(w_{1,i} + \vartheta w_{2,i}) \underline{y}_i\}$ , e o modelo linear múltiplo para dados intervalares,  $\underline{S}(\vartheta_1, \dots, \vartheta_q) = \sum_{i=1}^n \{(w_{1,i} + \vartheta_1 w_{2,i} + \dots + \vartheta_q w_{q+1,i}) \underline{y}_i\}$  podem ser interpretados como uma média aritmética ponderada. Segundo [Everitt \(2006\)](#) uma média aritmética ponderada é:

Uma média de valores aos quais uma série de pesos foram atribuídos para permitir uma consideração apropriada da importância relativa desses valores. Por exemplo, uma média aritmética ponderada de um conjunto de observações,  $a_1; a_2; \dots; a_n$ , com pesos  $p_1; p_2; \dots; p_n$ , é dada por  $\frac{\sum_{i=1}^n p_i a_i}{\sum_{i=1}^n p_i}$ .

Podemos reescrever  $\frac{\sum_{i=1}^n p_i a_i}{\sum_{i=1}^n p_i}$  adotando  $u_i = \frac{p_i}{\sum_{j=1}^n p_j}$ . Deste modo a média aritmética ponderada é dada por  $\sum_{i=1}^n u_i a_i$  e obtemos a propriedade de que  $\sum_{i=1}^n u_i = 1$ :

$$\sum_{i=1}^n u_i = \sum_{i=1}^n \left( \frac{p_i}{\sum_{j=1}^n p_j} \right) = \frac{\sum_{i=1}^n p_i}{\sum_{j=1}^n p_j} = 1.$$

Assim, se provarmos que  $\sum_{i=1}^n (w_{1,i} + \vartheta w_{2,i}) = 1$ , para o caso linear simples e que  $\sum_{i=1}^n (w_{1,i} + \vartheta_1 w_{2,i} + \dots + \vartheta_q w_{q+1,i}) = 1$  para o caso linear múltiplo segue que os modelos lineares simples e múltiplo para dados intervalares podem ser interpretados como uma média ponderada.

Retomando o modelo linear múltiplo para dados intervalares, seja um conjunto de dados na forma  $\{x_{i,1}; \dots; x_{i,q}; \underline{y}_i\}$ ,  $i = 1, \dots, n$ , o modelo linear múltiplo para dados

intervalares é dado por:

$$\bar{S}(\vartheta_1, \dots, \vartheta_q) = \sum_{i=1}^n \{(w_{1,i} + \vartheta_1 w_{2,i} + \dots + \vartheta_q w_{q+1,i}) \bar{y}_i\},$$

onde

$$\mathbf{W} = \begin{pmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,n} \\ w_{2,1} & w_{2,2} & \dots & w_{2,n} \\ \vdots & & & \\ w_{q+1,1} & w_{q+1,2} & \dots & w_{q+1,n} \end{pmatrix} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T,$$

com

$$\mathbf{X} = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{1,q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \dots & x_{n,q} \end{pmatrix}.$$

Queremos mostrar que  $\sum_{i=1}^n (w_{1,i} + \vartheta_1 w_{2,i} + \dots + \vartheta_q w_{q+1,i}) = 1$ . Para isso, seja a matriz  $\mathbf{E}$  é um vetor coluna com todas as  $n$  entradas iguais a 1, temos que:

$$\mathbf{W}\mathbf{E} = \begin{pmatrix} \sum_{i=1}^n w_{1,i} \\ \sum_{i=1}^n w_{2,i} \\ \vdots \\ \sum_{i=1}^n w_{q,i} \\ \sum_{i=1}^n w_{q+1,i} \end{pmatrix}. \quad (7.26)$$

Por outro lado podemos demonstrar que:

$$\begin{aligned} \mathbf{E} = \mathbf{X} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} &\implies \mathbf{X}^T \mathbf{E} = \mathbf{X}^T \mathbf{X} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \implies \\ (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{E} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \implies (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{E} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \implies \\ &\implies \mathbf{W}\mathbf{E} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \end{aligned} \quad (7.27)$$

Das equações 7.26 e 7.27 segue que:

$$\begin{pmatrix} \sum_{i=1}^n w_{1,i} \\ \sum_{i=1}^n w_{2,i} \\ \vdots \\ \sum_{i=1}^n w_{q,i} \\ \sum_{i=1}^n w_{q+1,i} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}.$$

Ou seja, temos:

$$\begin{aligned} \sum_{i=1}^n w_{1,i} &= 1, \\ \sum_{i=1}^n w_{2,i} &= 0, \\ &\vdots \\ \sum_{i=1}^n w_{q,i} &= 0 \text{ e} \\ \sum_{i=1}^n w_{q+1,i} &= 0. \end{aligned} \tag{7.28}$$

Desta forma finalmente provamos:

$$\begin{aligned} &\sum_{i=1}^n (w_{1,i} + \vartheta_1 w_{2,i} + \cdots + \vartheta_q w_{q+1,i}) = \\ &= \sum_{i=1}^n w_{1,i} + \vartheta_1 \sum_{i=1}^n w_{2,i} \cdots + \vartheta_q \sum_{i=1}^n w_{q+1,i} = 1 + \vartheta_1 0 \cdots + \vartheta_q 0 = 1. \end{aligned} \tag{7.29}$$

Desta forma concluímos que o modelo linear múltiplo para dados intervalares,  $\overline{\mathcal{S}}(\vartheta_1, \dots, \vartheta_q) = \sum_{i=1}^n \{(w_{1,i} + \vartheta_1 w_{2,i} + \cdots + \vartheta_q w_{q+1,i}) \underline{y}_i\}$ , é uma média ponderada dos dados  $\underline{y}_i$ . Tanto o caso simples quanto o caso *fuzzy*, e a té mesmo o caso real, são demonstrados de forma análoga.

Sabendo que os modelos se tratam de médias ponderadas dos dados observados podemos tirar algumas conclusões sobre os modelos. Primeiro o fato que fixados os valores dos  $\vartheta_i$ , temos como resultado de  $\overline{\mathcal{S}}(\vartheta_1, \dots, \vartheta_q)$  a soma ponderada dos dados observados onde o indivíduo  $i$  tem peso  $w_{1,i} + \vartheta_1 w_{2,i} + \cdots + \vartheta_q w_{q+1,i}$ . Quanto maior o peso mais o dado observado tem influência no resultado final naquele ponto.

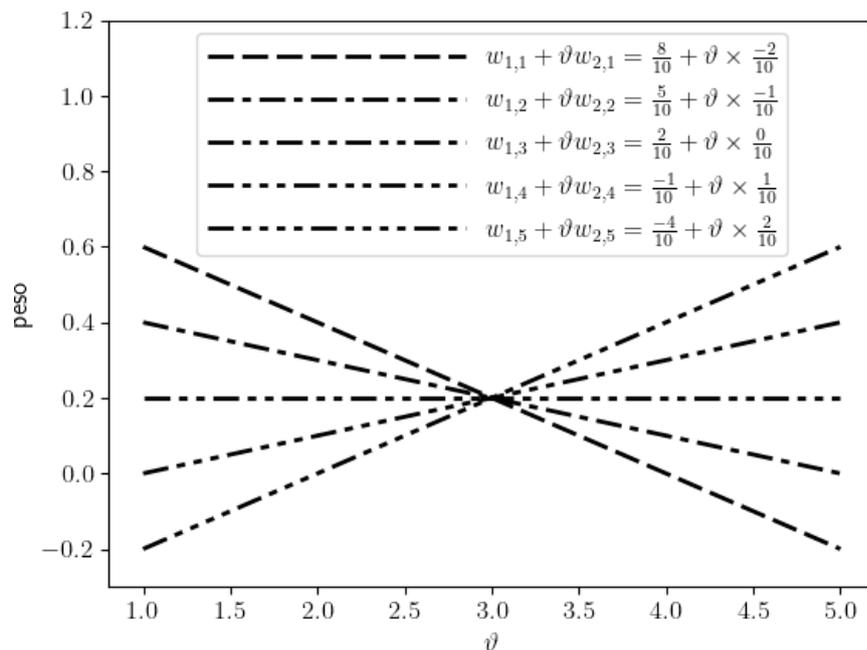
Pesos negativos são presentes nos modelos e o significado destes precisa ser investigado. Uma possibilidade é que eles incorporam uma maior imprecisão ao modelo nas regiões onde estão presentes. Retomando o Exemplo 15 onde o modelo tem a forma:

$$\begin{aligned} \bar{S}(\vartheta) &= (w_{1,1} + \vartheta w_{2,1}) \underline{y}_1 + \dots + (w_{1,5} + \vartheta w_{2,5}) \underline{y}_n = \\ &= \left( \frac{8}{10} + \vartheta \times \frac{-2}{10} \right) \underline{y}_1 + \left( \frac{5}{10} + \vartheta \times \frac{-1}{10} \right) \underline{y}_2 + \left( \frac{2}{10} + \vartheta \times \frac{0}{10} \right) \underline{y}_3 + \\ &\quad + \left( \frac{-1}{10} + \vartheta \times \frac{1}{10} \right) \underline{y}_4 + \left( \frac{-4}{10} + \vartheta \times \frac{2}{10} \right) \underline{y}_5. \end{aligned}$$

Fixado  $\vartheta$ , temos que  $\frac{8}{10} + \vartheta \times \frac{-2}{10}$  é o peso de  $\underline{y}_1$ ,  $\frac{5}{10} + \vartheta \times \frac{-1}{10}$  é o peso de  $\underline{y}_2$ ,  $\frac{2}{10} + \vartheta \times \frac{0}{10}$  é o peso de  $\underline{y}_3$ ,  $\frac{-1}{10} + \vartheta \times \frac{1}{10}$  é o peso de  $\underline{y}_4$  e  $\frac{-4}{10} + \vartheta \times \frac{2}{10}$  é o peso de  $\underline{y}_5$ .

Note que para valores de  $\vartheta$  menores do que 2 o peso de  $\underline{y}_5$  é menor do que zero, assim como que para valores maiores que 4 o peso de  $\underline{y}_1$  se torna menor que zero. Como os dados observados estão no intervalo  $\vartheta \in [1; 5]$  o fato de ter pesos negativos neste intervalo se destaca. Na Figura 8 podemos notar o aumento do comprimento dos intervalos  $\bar{S}(\vartheta)$  a medida que  $\vartheta$  diminui para valores menores que 2 ou aumenta para valores maiores que 4. Tendo esse fato em vista podemos supor que os pesos negativos agregam no modelo regiões onde a imprecisão da previsão  $\bar{S}(\vartheta)$  aumenta.

Figura 12 – Gráfico apresentando os valores dos pesos dados a cada observação quando interpretamos o modelo  $\bar{S}(\vartheta) = \left( \frac{8}{10} + \vartheta \times \frac{-2}{10} \right) \underline{y}_1 + \left( \frac{5}{10} + \vartheta \times \frac{-1}{10} \right) \underline{y}_2 + \left( \frac{2}{10} + \vartheta \times \frac{0}{10} \right) \underline{y}_3 + \left( \frac{-1}{10} + \vartheta \times \frac{1}{10} \right) \underline{y}_4 + \left( \frac{-4}{10} + \vartheta \times \frac{2}{10} \right) \underline{y}_5$  como uma média ponderada.



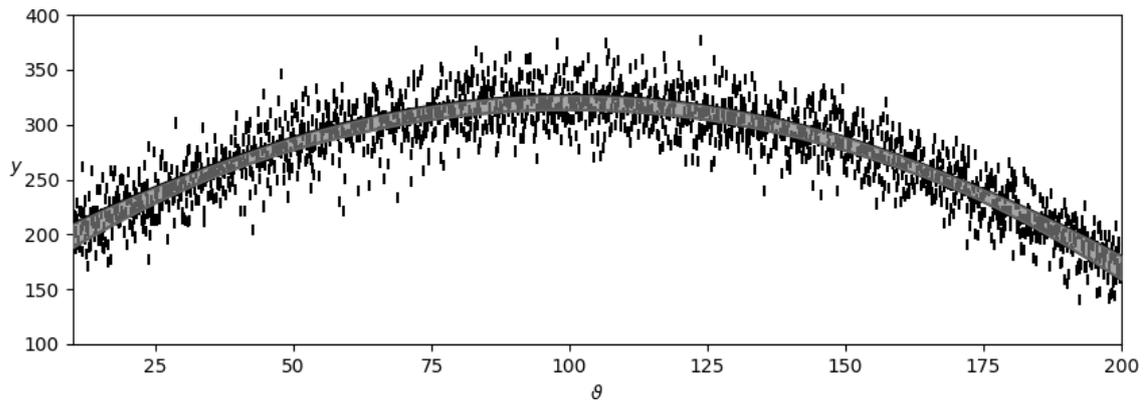
Fonte: O autor.

Na Figura 13 temos o mesmo exemplo da Figura 11, com dados na forma  $(\underline{y}_i; x_i)$ , porém, usando  $\vartheta_1 = \vartheta$  e  $\vartheta_2 = \vartheta^2$  para realizar a regressão na forma polinomial,

usamos a matriz de planejamento:

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_{1900} & x_{1900}^2 \end{pmatrix}.$$

Figura 13 – Exemplo polinomial com 1900 dados. Note o melhor ajuste em comparação com a Figura 11.



Fonte: O autor.

## 7.6 Conclusão

Esse capítulo apresentou dois modelos lineares para dados intervalares e dois para dados na forma de números *fuzzy*. Apesar do modelo linear simples,  $\bar{\mathcal{S}}(\vartheta) = \sum_{i=1}^n \{(w_{1,i} + \vartheta w_{2,i}) \bar{y}_i\}$ , Equação 7.14, possuir o mesmo resultado do modelo de análise sensitiva, Equação 6.11, oferece, no entanto, uma fórmula fechada para encontrar os intervalos  $\bar{\mathcal{S}}(\vartheta)$ .

Esses modelos vem da união de todas as possibilidades de casos clássicos tomando valores reais pertencentes aos intervalos, ou aos  $\alpha$ -níveis dos números *fuzzy*, o que oferece a esses modelos certa robustez herdada dos modelos clássicos.

Podemos realizar uma pequena análise da quantidade de operações que são necessárias para calcular esse modelo, tendo assim ideia de quanto trabalho terá um computador para calculá-lo.

Primeiro é necessário, uma única vez, calcular a matriz  $\mathbf{W} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ . Calcular  $\mathbf{X}^T \mathbf{X}$  tem um número de operações proporcional a  $n$ , pois são realizadas 4 vezes o produto escalar entre dois vetores de tamanho  $n$ . A matriz resultante é uma matriz  $2 \times 2$ , que tem um número de operações fixo para realizar a sua inversão e o produto desta inversa por  $\mathbf{X}^T$  também é proporcional ao número de colunas de  $\mathbf{X}^T$ , que é  $n$ . Logo a quantidade de operações para calcular a matriz  $\mathbf{W}$  é proporcional ao número  $n$  de observações.

Já, para cada intervalo  $\underline{S}(\vartheta)$  a expressão  $(w_{1,i} + \vartheta w_{2,i})\bar{y}_i$  é calculada  $n$  vezes, assim no total temos  $n$  multiplicações por escalar,  $n$  operações de soma e  $n$  operações de multiplicação de um intervalo por um número real. Após essas operações é necessário realizar a soma dos  $n$  intervalos resultantes, o que significa realizar  $n - 1$  somas de intervalos. Deste modo a quantidade de operações ao se realizar essa soma é proporcional ao número  $n$  de observações.

Para o caso múltiplo a matriz  $\mathbf{X}^T \mathbf{X}$  será de dimensão  $(q + 1) \times (q + 1)$ , o que dificulta um pouco o cálculo de sua inversa, porém, é possível evitar esse cálculo ao se realizar o cálculo através da solução do sistema linear  $(\mathbf{X}^T \mathbf{X})\mathbf{M} = \mathbf{X}^T$ , onde a matriz  $\mathbf{M}$  é a incógnita. Porém, continuamos com um número de operações proporcionais a  $n$ . Apesar do caso múltiplo ter mais operações de soma e multiplicação por escalar para cada expressão  $(w_{1,i} + \vartheta_1 w_{2,i} + \dots + \vartheta_q w_{q+1,i})$ , essa expressão se repete  $n$  vezes, o restante do cálculo é idêntico ao realizado no caso simples, sendo assim também proporcional a  $n$ .

Outra vantagem que os modelos apresentados nesse capítulo tem é que os modelos definidos para dados na forma de intervalos servem também para o caso em que os dados tem a forma de números *fuzzy*, pelo fato das operações aritméticas destes derivarem das operações daqueles.

Encerramos assim não só a apresentação dos modelos lineares para dados transversais que tenham os dados na forma de intervalos ou números *fuzzy* mas também a discussão sobre a utilidade dos mesmos. Na Parte III são exibidos alguns exemplos de aplicações dos modelos tratados nesse capítulo.

## 8 Modelo Linear Misto em Dois Estágios para Dados Imprecisos

### 8.1 Introdução

Na Seção 2.3 é apresentado o modelo linear misto em dois estágios para dados longitudinais clássico, isto é, os indivíduos tem suas medidas tomadas de forma repetida em diferentes instantes de tempo,  $t_1, \dots, t_m$ , porém, não necessariamente em todos os instantes  $t_j$ . Representa-se os dados de um indivíduo como:

$$\mathbf{Y}_i = \begin{pmatrix} Y_{i,1} \\ Y_{i,2} \\ \vdots \\ Y_{i,m} \end{pmatrix},$$

onde  $Y_{i,j}$  é a medida do indivíduo  $i$  tomada no tempo  $t_j$ .

O modelo de dois estágios assume a forma:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i,$$

com as suposições de que  $\mathbf{Y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i)$ , em que  $\boldsymbol{\Sigma}_i = \sigma^2\mathbf{I}_{m_i \times m_i} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T$ ,  $\mathbf{e}_i \sim N(0, \sigma^2\mathbf{I}_{m_i \times m_i})$  e  $\mathbf{b}_i \sim N(0, \mathbf{D})$ .

Este capítulo tem como objetivo propor um modelo onde as medidas  $Y_{i,j}$  são intervalos ou números *fuzzy*, que representam medidas imprecisas, como discutido no Capítulo 5.

Para o caso intervalar os dados tem a forma:

$$\overline{\mathbf{Y}}_i = \begin{pmatrix} \overline{Y_{i,1}} \\ \overline{Y_{i,2}} \\ \vdots \\ \overline{Y_{i,m}} \end{pmatrix},$$

onde  $\overline{Y_{i,j}}$  é a medida intervalar do indivíduo  $i$  tomada no tempo  $t_j$ .

Para o caso de dados na forma de números *fuzzy* representamos aqui esses dados na forma:

$$\mathbf{Y}_i^* = \begin{pmatrix} Y_{i,1}^* \\ Y_{i,2}^* \\ \vdots \\ Y_{i,m}^* \end{pmatrix},$$

onde  $Y_{i,j}^*$  é a medida do indivíduo  $i$  tomada no tempo  $t_j$ .

Para construir esses modelos escolhemos adaptar o processo de análise sensitiva da Equação 6.11 da Seção 6.4 e utilizar a Proposição 1 da Seção 7.2 para escrever o modelo como soma de intervalos ou como soma de números *fuzzy*.

## 8.2 Proposta de Modelo para o caso Intervalar

Na Equação 6.11 da Seção 6.4 é apresentado o processo de análise sensitiva, que trata do caso transversal simples. Nesse caso, tomando  $n$  observações  $(x_1; \underline{y}_1), \dots, (x_n; \underline{y}_n)$  onde  $\underline{y}_i$  são os intervalos  $\underline{y}_i = [\underline{y}_i, \overline{y}_i]$ , esse modelo é dado por:

$$\underline{f}(\vartheta) = \left\{ \begin{array}{l} \hat{a}\vartheta + \hat{b}; \\ \forall \hat{a}, \hat{b} \in \mathbb{R} \text{ que minimize } \sum_{i=1}^n (ax_i + b - y_i)^2, \forall y_i \in [\underline{y}_i, \overline{y}_i], i = 1, \dots, n \end{array} \right\}.$$

Assim, o intervalo  $\underline{f}(\vartheta)$  é o conjunto das estimativas  $f(\vartheta)$  para todas as combinações possíveis ao se escolher um ponto do intervalo de cada observação.

Na Seção 7.4 esse processo foi adaptado para ser usado no caso transversal múltiplo. Onde dado um conjunto de  $n$  dados  $\{x_{i,1}; \dots; x_{i,q}; \underline{y}_i\}$ ,  $i = 1, \dots, n$ , tomamos como modelo a união de todas as possibilidades de regressão linear múltipla ao se tomar um elemento  $y_i$  pertencente a cada intervalo  $\underline{y}_i$ , ou seja,

$$\underline{S}(\vartheta_1, \dots, \vartheta_q) = \left\{ \begin{array}{l} \hat{a}_0 + \hat{a}_1\vartheta_1 + \dots + \hat{a}_q\vartheta_q; \\ \forall \hat{a}_0, \dots, \hat{a}_q \text{ que minimize } \sum_{i=1}^n (\hat{a}_0 + \hat{a}_1\vartheta_1 + \dots + \hat{a}_q\vartheta_q - y_i)^2, \\ \forall y_i \in \underline{y}_i, i = 1, \dots, n \end{array} \right\},$$

onde  $\underline{S}(\vartheta_1, \dots, \vartheta_q)$  é a união de todos os resultados da regressão linear múltipla para todas as possibilidades de escolha de  $y_i \in \underline{y}_i$ .

Em ambos os casos as equações do caso clássico reescritas foram usadas para apresentar os modelos na forma do modelo base da Equação 7.13 através da Proposição 1. O caso simples assume a forma  $\underline{f}(\vartheta) = \sum_{i=1}^n \{f_i(\vartheta) \underline{y}_i\}$  e o caso múltiplo a forma  $\underline{f}(\vartheta_1, \dots, \vartheta_m) = \sum_{i=1}^n \{f_i(\vartheta_1, \dots, \vartheta_m) \underline{y}_i\}$ .

Do mesmo modo feito para o caso transversal múltiplo, é possível adaptar o modelo de análise sensitiva para o caso longitudinal e reescrever o modelo na forma do modelo base.

Começando pelo modelo clássico, o caso em que as observações são números reais. Na Seção 2.3 é apresentado o modelo linear misto em dois estágios para dados longitudinais. Supondo conhecidos ou estimados de alguma forma prévia, os valores de  $\hat{\sigma}^2$  e  $\hat{\mathbf{D}}, \hat{\Sigma}_i = \sigma^2 \mathbf{I}_{m_i \times m_i} + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T$ , e com as matrizes de planejamento

$$\mathbf{X} = \begin{pmatrix} 1 & g_{x,1}(t_1) & \dots & g_{x,k_x}(t_1) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & g_{x,1}(t_m) & \dots & g_{x,k_x}(t_m) \end{pmatrix} = \begin{pmatrix} \mathbf{g}_x(t_1)^T \\ \vdots \\ \mathbf{g}_x(t_m)^T \end{pmatrix}$$

e

$$\mathbf{Z} = \begin{pmatrix} 1 & g_{z,1}(t_1) & \dots & g_{z,k_z}(t_1) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & g_{z,1}(t_m) & \dots & g_{z,k_z}(t_m) \end{pmatrix} = \begin{pmatrix} \mathbf{g}_z(t_1) \\ \vdots \\ \mathbf{g}_z(t_m) \end{pmatrix},$$

tais que as suas linhas formem um conjunto linearmente independente. Assim as estimativas dos coeficientes  $\beta$  e  $\mathbf{b}_i$ ,  $i = 1, \dots, n$ , do modelo  $\mathbf{Y}_i = \mathbf{X}_i \beta + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i$  são dadas por

$$\hat{\beta} = \left( \sum_{i=1}^n \mathbf{X}_i^T \hat{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^n \mathbf{X}_i^T \hat{\Sigma}_i^{-1} \mathbf{Y}_i$$

e

$$\hat{\mathbf{b}}_i = \hat{\mathbf{D}} \mathbf{Z}_i^T \hat{\Sigma}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}).$$

Para o caso em que  $n$  indivíduos tem observações em forma de intervalos nos tempos  $t_1, \dots, t_m$ :

$$\underline{\mathbf{Y}}_i = \begin{pmatrix} \overline{Y_{i,1}} \\ \overline{Y_{i,2}} \\ \vdots \\ \overline{Y_{i,m}} \end{pmatrix}$$

propomos realizar um processo semelhante ao realizado no modelo de análise sensitiva, Equação 6.11, avaliando todas as possibilidades onde os dados reais estão contidos nos

intervalos:

$$\overline{S}_i(t) = \left\{ \begin{array}{l} \mathbf{g}_x(t)^T \hat{\boldsymbol{\beta}} + \mathbf{g}_z(t)^T \hat{\mathbf{b}}_i, \\ \text{onde:} \\ \hat{\boldsymbol{\beta}} = \left( \sum_{l=1}^n \mathbf{X}_l^T \hat{\boldsymbol{\Sigma}}_l^{-1} \mathbf{X}_l \right)^{-1} \sum_{p=1}^n \mathbf{X}_p^T \hat{\boldsymbol{\Sigma}}_p^{-1} \mathbf{Y}_p \\ \text{e} \\ \hat{\mathbf{b}}_i = \hat{\mathbf{D}} \mathbf{Z}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \left( \mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}} \right); \\ \forall \mathbf{Y}_k \text{ onde } Y_{k,j} \in \overline{Y}_{k,j}, \\ k = 1, \dots, n \text{ e } j = 1, \dots, m, \text{ tais que } \exists \overline{Y}_{k,j}. \end{array} \right\}. \quad (8.1)$$

Podemos reescrever a expressão  $\mathbf{g}_x(t)^T \hat{\boldsymbol{\beta}} + \mathbf{g}_z(t)^T \hat{\mathbf{b}}_i$  em termos dos  $Y_{k,j}$ .

$\mathbf{g}_x(t)^T \hat{\boldsymbol{\beta}}$  pode ser reescrito como:

$$\begin{aligned} \mathbf{g}_x(t)^T \hat{\boldsymbol{\beta}} &= \mathbf{g}_x(t)^T \left( \left( \sum_{l=1}^n \mathbf{X}_l^T \hat{\boldsymbol{\Sigma}}_l^{-1} \mathbf{X}_l \right)^{-1} \sum_{k=1}^n \mathbf{X}_k^T \hat{\boldsymbol{\Sigma}}_k^{-1} \mathbf{Y}_k \right) = \\ &= \left( \mathbf{g}_x(t)^T \left( \sum_{l=1}^n \mathbf{X}_l^T \hat{\boldsymbol{\Sigma}}_l^{-1} \mathbf{X}_l \right)^{-1} \sum_{k=1}^n \mathbf{X}_k^T \hat{\boldsymbol{\Sigma}}_k^{-1} \mathbf{Y}_k \right) = \\ &= \sum_{k=1}^n \left( \mathbf{g}_x(t)^T \left( \sum_{l=1}^n \mathbf{X}_l^T \hat{\boldsymbol{\Sigma}}_l^{-1} \mathbf{X}_l \right)^{-1} \mathbf{X}_k^T \hat{\boldsymbol{\Sigma}}_k^{-1} \mathbf{Y}_k \right) = \\ &= \sum_{k=1}^n \left( \mathbf{g}_x(t)^T \underbrace{\left( \mathbf{W}_1^{(k)} \quad \mathbf{W}_2^{(k)} \quad \dots \quad \mathbf{W}_m^{(k)} \right)}_{\text{apenas } \mathbf{W}_j^{(k)} \text{ tais que } \exists \overline{Y}_{k,j}} \mathbf{Y}_k \right) = \\ &= \sum_{k=1}^n \left( \mathbf{g}_x(t)^T \left( \mathbf{W}_1^{(k)} \quad \mathbf{W}_2^{(k)} \quad \dots \quad \mathbf{W}_m^{(k)} \right) \mathbf{Y}_k \right) = \\ &= \sum_{k=1}^n \left( \mathbf{g}_x(t)^T \left( \mathbf{W}_1^{(k)} \quad \mathbf{W}_2^{(k)} \quad \dots \quad \mathbf{W}_m^{(k)} \right) \begin{pmatrix} Y_{k,1} \\ \vdots \\ Y_{k,m} \end{pmatrix} \right) = \\ &= \sum_{k=1}^n \left( \sum_{j \in \overline{Y}_{k,j}} \left\{ \mathbf{g}_x(t)^T \mathbf{W}_j^{(k)} Y_{k,j} \right\} \right). \end{aligned}$$

Já  $\mathbf{g}_z(t)^T \hat{\mathbf{b}}_i$  pode ser dividido em duas parcelas:

$$\begin{aligned} \mathbf{g}_z(t)^T \hat{\mathbf{b}}_i &= \mathbf{g}_z(t)^T \left( \hat{\mathbf{D}} \mathbf{Z}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \left( \mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}} \right) \right) = \\ &= \mathbf{g}_z(t)^T \left( \hat{\mathbf{D}} \mathbf{Z}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Y}_i \right) - \mathbf{g}_z(t)^T \left( \hat{\mathbf{D}} \mathbf{Z}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{X}_i \hat{\boldsymbol{\beta}} \right) \end{aligned}$$

$$\begin{aligned}
 \mathbf{g}_z(t)^T \left( \hat{\mathbf{D}} \mathbf{Z}_i^T \hat{\Sigma}_i^{-1} \mathbf{Y}_i \right) &= \mathbf{g}_z(t)^T \underbrace{\left( \mathbf{V}_1^{(i)} \quad \mathbf{V}_2^{(i)} \quad \dots \quad \mathbf{V}_m^{(i)} \right)}_{\text{apenas } \mathbf{V}_j^{(i)} \text{ tais que } \exists \overline{Y_{k,j}}.} \mathbf{Y}_i = \\
 &= \mathbf{g}_z(t)^T \left( \mathbf{V}_1^{(i)} \quad \mathbf{V}_2^{(i)} \quad \dots \quad \mathbf{V}_m^{(i)} \right) \begin{pmatrix} Y_{i,1} \\ \vdots \\ Y_{i,m} \end{pmatrix} = \\
 &= \sum_{j | \exists \overline{Y_{i,j}}} \left\{ \mathbf{g}_z(t)^T \mathbf{V}_j^{(i)} Y_{i,j} \right\}.
 \end{aligned}$$

$$\begin{aligned}
 \mathbf{g}_z(t)^T \left( \hat{\mathbf{D}} \mathbf{Z}_i^T \hat{\Sigma}_i^{-1} \mathbf{X}_i \hat{\beta} \right) &= \\
 &= \mathbf{g}_z(t)^T \left( \hat{\mathbf{D}} \mathbf{Z}_i^T \hat{\Sigma}_i^{-1} \mathbf{X}_i \left( \sum_{l=1}^n \mathbf{X}_l^T \hat{\Sigma}_l^{-1} \mathbf{X}_l \right)^{-1} \sum_{k=1}^n \mathbf{X}_k^T \hat{\Sigma}_k^{-1} \mathbf{Y}_k \right) = \\
 &= \sum_{k=1}^n \left( \mathbf{g}_z(t)^T \mathbf{V}^{(i)} \mathbf{X}_i \left( \sum_{l=1}^n \mathbf{X}_l^T \hat{\Sigma}_l^{-1} \mathbf{X}_l \right)^{-1} \mathbf{X}_k^T \hat{\Sigma}_k^{-1} \mathbf{Y}_k \right) = \\
 &= \sum_{k=1}^n \left( \mathbf{g}_z(t)^T \mathbf{V}^{(i)} \mathbf{X}_i \underbrace{\left( \mathbf{W}_1^{(k)} \quad \mathbf{W}_2^{(k)} \quad \dots \quad \mathbf{W}_m^{(k)} \right)}_{\text{apenas } \mathbf{W}_j^{(k)} \text{ tais que } \exists \overline{Y_{k,j}}.} \mathbf{Y}_k \right) = \\
 &= \sum_{k=1}^n \left( \mathbf{g}_z(t)^T \mathbf{V}^{(i)} \mathbf{X}_i \left( \mathbf{W}_1^{(k)} \quad \mathbf{W}_2^{(k)} \quad \dots \quad \mathbf{W}_m^{(k)} \right) \begin{pmatrix} Y_{k,1} \\ \vdots \\ Y_{k,m} \end{pmatrix} \right) = \\
 &= \sum_{k=1}^n \left( \sum_{j | \exists \overline{Y_{k,j}}} \left\{ \mathbf{g}_z(t)^T \mathbf{V}^{(i)} \mathbf{X}_i \mathbf{W}_j^{(k)} Y_{k,j} \right\} \right)
 \end{aligned}$$

$$\begin{aligned}
 \mathbf{g}_x(t)^T \hat{\beta} + \mathbf{g}_z(t)^T \hat{\mathbf{b}}_i &= \\
 &= \sum_{k=1}^n \sum_{j | \exists \overline{Y_{k,j}}} \left( \mathbf{g}_x(t)^T \mathbf{W}_j^{(k)} + \delta_{i,k} \mathbf{g}_z(t)^T \mathbf{V}_j^{(i)} - \mathbf{g}_z(t)^T \mathbf{V}^{(i)} \mathbf{X}_i \mathbf{W}_j^{(k)} \right) Y_{k,j} = \\
 &= \sum_{k=1}^n \sum_{j | \exists \overline{Y_{k,j}}} \left( \mathbf{g}_x(t)^T \mathbf{W}_j^{(k)} + \mathbf{g}_z(t)^T \left( \delta_{i,k} \mathbf{V}_j^{(i)} - \mathbf{V}^{(i)} \mathbf{X}_i \mathbf{W}_j^{(k)} \right) \right) Y_{k,j},
 \end{aligned}$$

$$\text{onde } \delta_{i,k} = \begin{cases} 1 & \text{se } i = k \\ 0 & \text{se } i \neq k \end{cases}.$$

Assim é possível reescrever  $\overline{S}_i(t)$  como:

$$\overline{S}_i(t) = \left\{ \begin{array}{l} \sum_{k=1}^n \sum_{j|\exists \overline{Y}_{k,j}} \left( \mathbf{g}_x(t)^T \mathbf{W}_j^{(k)} + \mathbf{g}_z(t)^T \left( \delta_{i,k} \mathbf{V}_j^{(i)} - \mathbf{V}^{(i)} \mathbf{X}_i \mathbf{W}_j^{(k)} \right) \right) \overline{Y}_{k,j}, \\ \text{com} \\ \mathbf{W}^{(k)} = \left( \sum_{l=1}^n \mathbf{X}_l^T \hat{\Sigma}_l^{-1} \mathbf{X}_l \right)^{-1} \mathbf{X}_k^T \hat{\Sigma}_k^{-1} \text{ e} \\ \mathbf{V}^{(i)} = \hat{\mathbf{D}} \mathbf{Z}_i^T \hat{\Sigma}_i^{-1}; \\ \forall Y_{k,j} \in \overline{Y}_{k,j}, \quad k = 1, \dots, n \text{ e } j = 1, \dots, m, \text{ tais que } \exists \overline{Y}_{k,j}. \end{array} \right. \quad (8.2)$$

Pela Proposição 1 da Subseção 7.2 temos que

$$\sum_{i=1}^n \{a_i \underline{y}_i\} = \bigcup_{(y_1, \dots, y_n) \in \underline{y}_1 \times \dots \times \underline{y}_n} \left( \sum_{i=1}^n \{a_i y_i\} \right),$$

logo podemos reescrever como:

$$\overline{S}_i(t) = \sum_{k=1}^n \sum_{j|\exists \overline{Y}_{k,j}} \left( \mathbf{g}_x(t)^T \mathbf{W}_j^{(k)} + \mathbf{g}_z(t)^T \left( \delta_{i,k} \mathbf{V}_j^{(i)} - \mathbf{V}^{(i)} \mathbf{X}_i \mathbf{W}_j^{(k)} \right) \right) \overline{Y}_{k,j}, \quad (8.3)$$

com

$$\mathbf{W}^{(k)} = \left( \sum_{l=1}^n \mathbf{X}_l^T \hat{\Sigma}_l^{-1} \mathbf{X}_l \right)^{-1} \mathbf{X}_k^T \hat{\Sigma}_k^{-1}$$

e

$$\mathbf{V}^{(i)} = \hat{\mathbf{D}} \mathbf{Z}_i^T \hat{\Sigma}_i^{-1}.$$

Porém, é necessário levantar a questão sobre quais  $\hat{\sigma}^2$  e  $\hat{\mathbf{D}}$  usar nesse modelo. Uma primeira opção é usar valores previamente calculados ou conhecidos. Porém, nem sempre esses valores estarão disponíveis. Neste caso optamos por usar os mesmos valores obtidos pelo modelo linear misto em dois estágios visto na Seção 2.3 para o conjunto de dados:

$$\mathring{\mathbf{Y}}_i = \begin{pmatrix} \mathring{Y}_{i,1} \\ \vdots \\ \mathring{Y}_{i,m} \end{pmatrix}, \quad i = 1, \dots, n, \quad (8.4)$$

onde  $\mathring{Y}_{i,j}$  é o ponto médio do intervalo  $\overline{Y}_{i,j}$ . Desta forma adotamos o valor de um caso clássico como aproximação para o caso intervalar. Porém, o método apresentado aqui independe da forma como  $\hat{\sigma}^2$  e  $\hat{\mathbf{D}}$  foram obtidos.

### 8.3 Proposta de Modelo para o caso *Fuzzy*

Definimos o modelo para o caso em que  $n$  indivíduos tem observações em forma de números *fuzzy* nos tempos  $t_1, \dots, t_m$ :

$$\mathbf{Y}_i^* = \begin{pmatrix} Y_{i,1}^* \\ Y_{i,2}^* \\ \vdots \\ Y_{i,m}^* \end{pmatrix},$$

como:

$$S_i^*(t) = \sum_{k=1}^n \sum_{j|\exists Y_{k,j}^*} \left( \mathbf{g}_x(t)^T \mathbf{W}_j^{(k)} + \mathbf{g}_z(t)^T \left( \delta_{i,k} \mathbf{V}_j^{(i)} - \mathbf{V}^{(i)} \mathbf{X}_i \mathbf{W}_j^{(k)} \right) \right) Y_{k,j}^*, \quad (8.5)$$

com

$$\mathbf{W}^{(k)} = \left( \sum_{l=1}^n \mathbf{X}_l^T \hat{\Sigma}_l^{-1} \mathbf{X}_l \right)^{-1} \mathbf{X}_k^T \hat{\Sigma}_k^{-1}$$

e

$$\mathbf{V}^{(i)} = \hat{\mathbf{D}} \mathbf{Z}_i^T \hat{\Sigma}_i^{-1}.$$

Escolhidas as funções  $\mathbf{g}_x(t)$  e  $\mathbf{g}_z(t)$  podemos montar as matrizes de planejamento  $\mathbf{X}$  e  $\mathbf{Z}$ :

$$\mathbf{X} = \begin{pmatrix} \mathbf{g}_x^T(t_1) \\ \vdots \\ \mathbf{g}_x^T(t_m) \end{pmatrix} \text{ e } \mathbf{Z} = \begin{pmatrix} \mathbf{g}_z^T(t_1) \\ \vdots \\ \mathbf{g}_z^T(t_m) \end{pmatrix}.$$

É preciso também encontrar  $\hat{\sigma}^2$  e  $\hat{\mathbf{D}}$ , caso não sejam conhecidas previamente. Como no caso intervalar propomos usar os valores encontrados no modelo clássico com

$$\hat{\mathbf{Y}}_i = \begin{pmatrix} \hat{Y}_{i,1} \\ \vdots \\ \hat{Y}_{i,m} \end{pmatrix}, \quad i = 1, \dots, n,$$

onde  $\hat{Y}_{i,j} = \text{defuzz}(Y_{i,j}^*)$ , pelo processo de *defuzzificação* visto na Equação 4.9 da Seção 4.3.

### 8.4 Exemplos

Tendo definidos o modelo linear misto em dois estágios para dados intervalares e também para dados na forma de números *fuzzy*, Equações 8.3 e 8.5 apresentamos um exemplo com a intenção de apresentar somente como o modelo funciona para fins didáticos.

**Exemplo 17.** *Seja um conjunto de dados longitudinais com as observações na forma de intervalos:*

$$\underline{\mathbf{Y}}_1 = \begin{pmatrix} [12, 0 & 8, 0] \\ [15, 5 & 12, 5] \\ [20, 0 & 18, 0] \\ [26, 5 & 23, 5] \end{pmatrix}, \underline{\mathbf{Y}}_2 = \begin{pmatrix} [10, 0 & 8, 0] \\ [14, 0 & 10, 0] \\ [19, 0 & 17, 0] \\ [24, 5 & 19, 5] \end{pmatrix}, \underline{\mathbf{Y}}_3 = \begin{pmatrix} [8, 0 & 4, 0] \\ [10, 0 & 6, 0] \\ [11, 5 & 8, 5] \\ [13, 0 & 11, 0] \end{pmatrix}, \quad (8.6)$$

$$\underline{\mathbf{Y}}_4 = \begin{pmatrix} [5, 5 & 2, 5] \\ [8, 5 & 5, 5] \\ [11, 0 & 7, 0] \\ [11, 5 & 10, 5] \end{pmatrix} \text{ e } \underline{\mathbf{Y}}_5 = \begin{pmatrix} [2, 5 & 1, 5] \\ [6, 5 & 3, 5] \\ [10, 0 & 6, 0] \\ [11, 0 & 9, 0] \end{pmatrix}. \quad (8.7)$$

O conjunto de dados longitudinais com os pontos médios é dado por:

$$\dot{\mathbf{Y}}_1 = \begin{pmatrix} 10 \\ 14 \\ 19 \\ 25 \end{pmatrix}, \dot{\mathbf{Y}}_2 = \begin{pmatrix} 9 \\ 12 \\ 18 \\ 22 \end{pmatrix}, \dot{\mathbf{Y}}_3 = \begin{pmatrix} 6 \\ 8 \\ 10 \\ 12 \end{pmatrix}, \dot{\mathbf{Y}}_4 = \begin{pmatrix} 4 \\ 7 \\ 9 \\ 11 \end{pmatrix} \text{ e } \dot{\mathbf{Y}}_5 = \begin{pmatrix} 2 \\ 5 \\ 8 \\ 10 \end{pmatrix}. \quad (8.8)$$

Adotando

$$\mathbf{g}_x(t) = \begin{pmatrix} 1 \\ t \end{pmatrix} \text{ e } \mathbf{g}_z(t) = \begin{pmatrix} 1 \\ t \end{pmatrix}$$

obtemos as matrizes de planejamento

$$\mathbf{X} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \text{ e } \mathbf{Z} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix}, \quad (8.9)$$

podemos tomar os  $\hat{\sigma}^2$  e  $\hat{\mathbf{D}}$  obtidos no modelo linear misto de dois estágios, como no Exemplo 1 da Subseção 2.4:

$$\hat{\sigma}^2 = 0,31 \text{ e } \hat{\mathbf{D}} = \begin{pmatrix} 2,99 & 1,27 \\ 1,27 & 1,41 \end{pmatrix}. \quad (8.10)$$

Usando estas  $\hat{\sigma}^2$  e  $\hat{\mathbf{D}}$  definimos o modelo para cada indivíduo na forma:

$$\overline{S}_i(t) = \sum_{k=1}^n \sum_{j|\exists \overline{Y}_{k,j}} \left( \mathbf{g}_x(t)^T \mathbf{W}_j^{(k)} + \mathbf{g}_z(t)^T \left( \delta_{i,k} \mathbf{V}_j^{(i)} - \mathbf{V}^{(i)} \mathbf{X}_i \mathbf{W}_j^{(k)} \right) \right) \overline{Y}_{k,j},$$

com

$$\mathbf{W}^{(k)} = \left( \sum_{l=1}^n \mathbf{X}_l^T \hat{\Sigma}_l^{-1} \mathbf{X}_l \right)^{-1} \mathbf{X}_k^T \hat{\Sigma}_k^{-1}$$

e

$$\mathbf{V}^{(i)} = \hat{\mathbf{D}} \mathbf{Z}_i^T \hat{\Sigma}_i^{-1}.$$

Como  $\mathbf{X}_1 = \dots = \mathbf{X}_5 = \mathbf{X}$ ,  $\mathbf{Z}_1 = \dots = \mathbf{Z}_5 = \mathbf{Z}$  e  $\hat{\Sigma}_1 = \dots = \hat{\Sigma}_5 = \hat{\Sigma}$ , temos de forma mais simples:

$$\overline{S}_i(t) = \sum_{k=1}^n \sum_{j|\exists \overline{Y}_{k,j}} (\mathbf{g}_x(t)^T \mathbf{W}_j + \mathbf{g}_z(t)^T (\delta_{i,k} \mathbf{V}_j - \mathbf{V} \mathbf{X} \mathbf{W}_j)) \overline{Y}_{k,j},$$

com

$$\mathbf{W} = \left( \sum_{l=1}^n \mathbf{X}^T \hat{\Sigma}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \hat{\Sigma}^{-1}$$

e

$$\mathbf{V} = \hat{\mathbf{D}} \mathbf{Z}^T \hat{\Sigma}^{-1}.$$

Obtemos as matrizes:

$$\mathbf{W} = \left( \mathbf{W}_1 \quad \mathbf{W}_2 \quad \mathbf{W}_3 \quad \mathbf{W}_4 \right) = \begin{pmatrix} 0,2 & 0,1 & 0 & -0,1 \\ -0,06 & -0,02 & 0,02 & 0,06 \end{pmatrix}$$

e

$$\mathbf{V} = \left( \mathbf{V}_1 \quad \mathbf{V}_2 \quad \mathbf{V}_3 \quad \mathbf{V}_4 \right) = \begin{pmatrix} 0,695 & 0,361 & 0,027 & -0,306 \\ -0,192 & -0,051 & 0,09 & 0,231 \end{pmatrix}.$$

A Figura 14 apresenta os gráficos de  $\overline{S}_i(t)$ .

$$\begin{aligned} \overline{S}_1(t) = & (0.756 - 0.214t) \overline{Y}_{1,1} + (0.389 - 0.061t) \overline{Y}_{1,2} + \\ & +(0.022 + 0.092t) \overline{Y}_{1,3} + (-0.345 + 0.245t) \overline{Y}_{1,4} + \\ & +(0.061 - 0.022t) \overline{Y}_{2,1} + (0.028 - 0.010t) \overline{Y}_{2,2} + \\ & +(-0.005 + 0.002t) \overline{Y}_{2,3} + (-0.039 + 0.014t) \overline{Y}_{2,4} + \\ & +(0.061 - 0.022t) \overline{Y}_{3,1} + (0.028 - 0.010t) \overline{Y}_{3,2} + \\ & +(-0.005 + 0.002t) \overline{Y}_{3,3} + (-0.039 + 0.014t) \overline{Y}_{3,4} + \\ & +(0.061 - 0.022t) \overline{Y}_{4,1} + (0.028 - 0.010t) \overline{Y}_{4,2} + \\ & +(-0.005 + 0.002t) \overline{Y}_{4,3} + (-0.039 + 0.014t) \overline{Y}_{4,4} + \\ & +(0.061 - 0.022t) \overline{Y}_{5,1} + (0.028 - 0.010t) \overline{Y}_{5,2} + \\ & +(-0.005 + 0.002t) \overline{Y}_{5,3} + (-0.039 + 0.014t) \overline{Y}_{5,4}. \end{aligned}$$

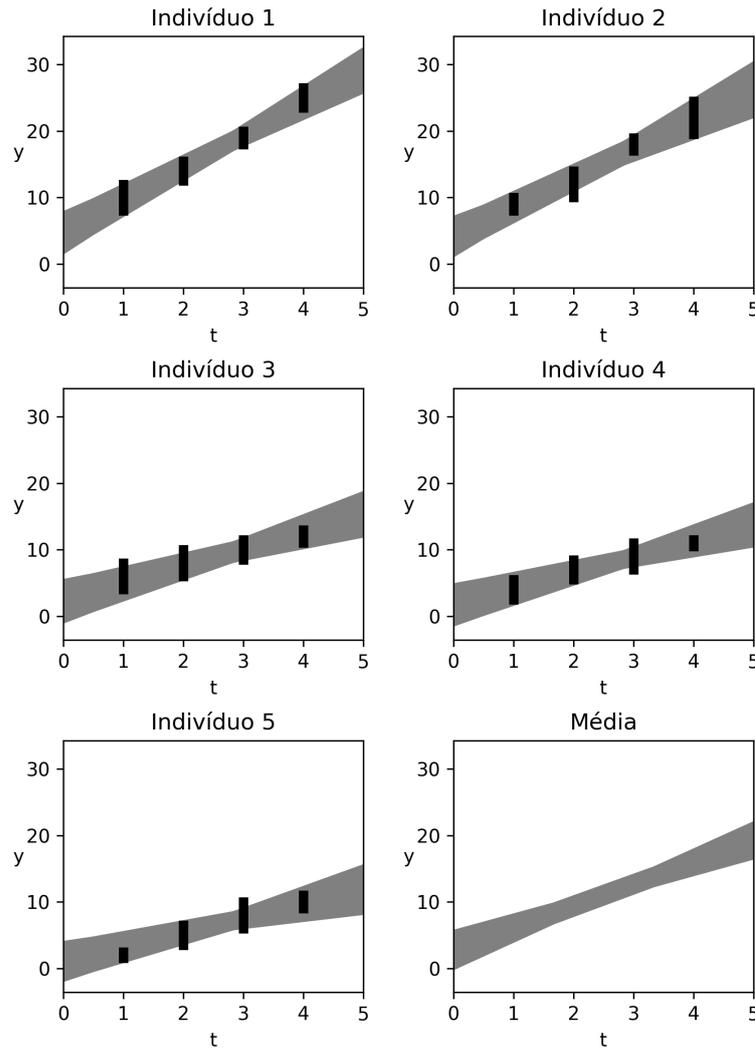
Para exemplificar o cálculo dos termos do modelo, tome o termo  $(0.756 - 0.214t) \overline{Y_{1,1}}$ , que é dado por  $(\mathbf{g}_x(t)^T \mathbf{W}_1 + \mathbf{g}_z(t)^T (\delta_{1,1} \mathbf{V}_1 - \mathbf{V} \mathbf{X} \mathbf{W}_1)) \overline{Y_{1,1}}$ .

$$\begin{aligned}
& (\mathbf{g}_x(t)^T \mathbf{W}_1 + \mathbf{g}_z(t)^T (\delta_{1,1} \mathbf{V}_1 - \mathbf{V} \mathbf{X} \mathbf{W}_1)) \overline{Y_{1,1}} = \\
& = \left( (1 \ t) \begin{pmatrix} 0,2 \\ -0,06 \end{pmatrix} + (1 \ t) \left( 1 \begin{pmatrix} 0,695 \\ -0,192 \end{pmatrix} - \mathbf{V} \mathbf{X} \begin{pmatrix} 0,2 \\ -0,06 \end{pmatrix} \right) \right) \overline{Y_{1,1}} = \\
& = \left( (1 \ t) \begin{pmatrix} 0,2 \\ -0,06 \end{pmatrix} + (1 \ t) \left( \begin{pmatrix} 0,695 \\ -0,192 \end{pmatrix} - \begin{pmatrix} 0,139 \\ -0,038 \end{pmatrix} \right) \right) \overline{Y_{1,1}} = \\
& = \left( (1 \ t) \begin{pmatrix} 0,2 \\ -0,06 \end{pmatrix} + (1 \ t) \begin{pmatrix} 0,556 \\ -0,154 \end{pmatrix} \right) \overline{Y_{1,1}} = \\
& = ((0,2 - 0,06t) + (0,556 - 0,154t)) \overline{Y_{1,1}} = \\
& = (0,756 - 0,214t) \overline{Y_{1,1}}.
\end{aligned}$$

Na Figura 14 apresentamos o gráfico  $\overline{S}_i(t)$  para cada indivíduo  $i$ , assim como o gráfico da média  $\sum_{k=1}^n \sum_{j \in \overline{Y_{k,j}}} (\mathbf{g}_x(t)^T \mathbf{W}_j) \overline{Y_{k,j}}$ . Da mesma forma como discutido na Seção 2.3 essa média diz respeito apenas ao efeito fixo presente em toda a população. Note que cada indivíduo possui o seu próprio modelo  $\overline{S}_i(t)$ , porém, o modelo trata de toda a população para criá-lo.

Figura 14 – Gráficos  $\overline{S}_i(t)$  para os indivíduos 1, 2, 3, 4 e 5 do Exemplo 17. A área cinza apresenta os pontos  $(t, y)$  tais que  $y \in \overline{S}_i(t)$ . As linhas verticais apresentam os intervalos observados  $\overline{Y}_{i,j}$ , ou seja, os dados, mais especificamente eles apresentam os pontos  $(t_j; \overline{y}_{i,j}), i = 1, \dots, 5$  e  $j = 1, \dots, 4$ , tais que  $y_{i,j} \in \overline{Y}_{i,j}$ . O último gráfico apresenta o valor médio da população, ou seja,

$$\sum_{k=1}^n \sum_{j|\exists \overline{Y}_{k,j}} (\mathbf{g}_x(t)^T \mathbf{W}_j) \overline{Y}_{k,j}.$$



Fonte: O autor.

## 8.5 Conclusão

Esse capítulo apresentou um modelo linear misto em dois estágios para dados intervalares e um para dados na forma de números *fuzzy*. Fixadas as matrizes  $\hat{\Sigma}$  e  $\hat{D}$ , esses modelos vem da união de todas as possibilidades de casos clássicos tomando valores reais pertencentes aos intervalos, ou aos  $\alpha$ -níveis dos números *fuzzy*, o que oferece a esses modelos certa robustez herdada dos modelos clássicos.

Para a escolha das matrizes  $\hat{\Sigma}$  e  $\hat{D}$  é proposto tomar as estimativas do modelo clássico tomando os pontos médios dos intervalos ou usando os valores defuzzificados. Para o propósito de apresentar o modelo linear misto em dois estágios deste capítulo esse método é suficiente. Porém, essa não é a única forma possível. Por exemplo, [Denoeux \(2011\)](#) propõe um método de máxima verossimilhança para dados *fuzzy*. Esse método talvez possa ser adaptado para realizar a máxima verossimilhança com a restrição de que os  $\hat{\Sigma}_i$  tenham a forma  $\hat{\Sigma}_i = \sigma^2 \mathbf{I}_{m_i \times m_i} + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T$ .

Estimar os  $\hat{\Sigma}_i$  e o  $\hat{D}$  precisa ser realizado apenas uma vez, porém, dependendo da quantidade de dados e da forma escolhida para realizar esses cálculos podemos ter maior ou menor esforço computacional nesse passo inicial.

Ainda antes de calcular  $\overline{S}_i(t)$  ainda é necessário calcular as matrizes  $\mathbf{W}^{(k)}$  e  $\mathbf{V}^{(i)}$ .

$$\mathbf{W}^{(k)} = \left( \sum_{l=1}^n \mathbf{X}_l^T \hat{\Sigma}_l^{-1} \mathbf{X}_l \right)^{-1} \mathbf{X}_k^T \hat{\Sigma}_k^{-1}$$

e

$$\mathbf{V}^{(i)} = \hat{D} \mathbf{Z}_i^T \hat{\Sigma}_i^{-1}.$$

Para calcular as inversas de cada  $\hat{\Sigma}_i$  temos a quantidade inerente de operações para calcular a inversa de uma matriz  $m_i \times m_i$ . No entanto, a estrutura  $\hat{\Sigma}_i = \sigma^2 \mathbf{I}_{m_i \times m_i} + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T$  pode facilitar e reduzir o número de operações necessárias.

Os cálculos de todas as matrizes  $\mathbf{W}^{(k)}$  possuem em comum a parte

$$\left( \sum_{l=1}^n \mathbf{X}_l^T \hat{\Sigma}_l^{-1} \mathbf{X}_l \right)^{-1}$$

assim essa expressão precisa ser calculada apenas uma vez. O cálculo de  $\mathbf{X}_l^T \hat{\Sigma}_l^{-1} \mathbf{X}_l$  independe de  $n$ , pois é uma multiplicação de matrizes de dimensão  $(k_x + 1) \times m_l$ ,  $m_l \times m_l$  e  $m_l \times (k_x + 1)$ . O número de operações da soma dessas  $n$  matrizes por sua vez é proporcional a  $n$ . O cálculo dessa inversa pode ser substituído pela resolução do sistema linear  $\left( \sum_{l=1}^n \mathbf{X}_l^T \hat{\Sigma}_l^{-1} \mathbf{X}_l \right) \mathbf{W}^{(k)} = \mathbf{X}_k^T \hat{\Sigma}_k^{-1}$ , onde a matriz  $\mathbf{W}^{(k)}$  é a incógnita da equação. É necessário calcular  $n$ , ou menos, matrizes  $\mathbf{W}^{(k)}$ , tornando o cálculo de todas elas proporcional a  $n$ .

Calcular as matrizes  $\mathbf{V}^{(i)} = \hat{D} \mathbf{Z}_i^T \hat{\Sigma}_i^{-1}$  se trata de uma multiplicação de matrizes de dimensão  $(k_z + 1) \times (k_z + 1)$ ,  $(k_z + 1) \times m_i$  e  $m_i \times m_i$ . Essa multiplicação tem um número de operações independente de  $n$ . Porém, é necessário calcular  $n$ , ou menos, matrizes  $\mathbf{V}^{(i)}$ , tornando o cálculo de todas elas proporcional a  $n$ .

Dessa forma, para calcular as matrizes necessárias para calcular os  $\overline{S}_i(t)$  tem um número de operações proporcional a  $n$ , ao número de indivíduos.

Enfim, fixado  $t$ , para calcular

$$\overline{S}_i(t) = \sum_{k=1}^n \sum_{j|\exists \overline{Y}_{k,j}} \left( \mathbf{g}_x(t)^T \mathbf{W}_j^{(k)} + \mathbf{g}_z(t)^T \left( \delta_{i,k} \mathbf{V}_j^{(i)} - \mathbf{V}^{(i)} \mathbf{X}_i \mathbf{W}_j^{(k)} \right) \right) \overline{Y}_{k,j},$$

é necessário fazer  $nm$ , ou menos, vezes a operação

$$\mathbf{g}_x(t)^T \mathbf{W}_j^{(k)} + \mathbf{g}_z(t)^T \left( \delta_{i,k} \mathbf{V}_j^{(i)} - \mathbf{V}^{(i)} \mathbf{X}_i \mathbf{W}_j^{(k)} \right),$$

onde  $m$  é o máximo dos valores  $m_i$ .  $\mathbf{g}_x(t)^T \mathbf{W}_j^{(k)}$  e  $\mathbf{g}_z(t)^T \left( \delta_{i,k} \mathbf{V}_j^{(i)} - \mathbf{V}^{(i)} \mathbf{X}_i \mathbf{W}_j^{(k)} \right)$  são produtos internos entre vetores do  $\mathbb{R}_i^m$ . Desta forma a soma final terá uma quantidade de operações proporcionais a  $n$ .

Para o caso em que as observações tem a forma de números *fuzzy* também temos um número de operações proporcionais a  $n$ . Desta forma os modelos possuem uma complexidade proporcional ao número de indivíduos, o que indica um baixo esforço computacional.

Encerramos assim a apresentação do modelo linear misto em dois estágios para os casos em que os dados tenham a forma de intervalos ou na forma de números *fuzzy*. Na Parte III são exibidos alguns exemplos de aplicações dos modelos tratados nesse capítulo.

# Parte III

## Aplicações

## 9 Volume Corporal de Orcas-Pigmeias

### 9.1 Introdução

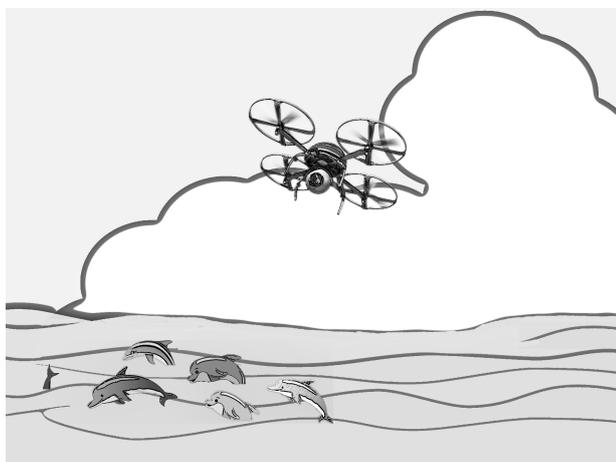
O estudo apresentado por [Currie et al. \(2021\)](#), usado como exemplo motivador no capítulo de Introdução, página 13, trata de um ótimo exemplo de estudo longitudinal com imprecisão nos valores das medidas. Em quatro ocasiões diferentes, dias 1, 7, 11 e 17 o volume corporal de orcas-pigmeias de seis diferentes indivíduos foi medido. Em cada ocasião as medições algum grau de imprecisão, dependendo do método de medição utilizado juntamente das condições do momento em que a medida foi tomada.

Um dos métodos usados por [Currie et al. \(2021\)](#) para as medições foi a fotogrametria usando um veículo aéreo não tripulado. Para determinar o volume corporal do indivíduo  $i$  no dia  $j$  foram tomadas fotografias ao longo do dia. A partir de cada fotografia o volume corporal do indivíduo naquele dia foi calculado, gerando assim medidas repetidas no mesmo dia. Foram calculados então a média e o desvio padrão destas medidas. Deste modo os valores apresentados tomam a forma de uma média acompanhada de um desvio padrão:

$$\text{volume}(m^3) \pm SD.$$

Como discutido no Capítulo 5 podemos modelar essa imprecisão na forma de intervalos ou em forma de números *fuzzy*, como é realizado por [Pinto et al. \(2018\)](#).

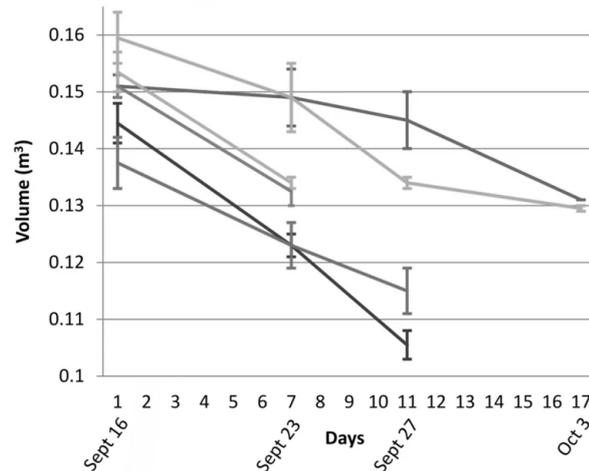
Figura 15 – [Currie et al. \(2021\)](#), através de imagens tomadas por um veículo aéreo não tripulado, estimou o volume corporal de um grupo de orcas-pigmeias. Para cada foto o volume corporal é estimado e ao final são usados o valor médio e o desvio-padrão destas estimativas.



Fonte: ([Pixabay GmbH, n.d.](#))(adaptado)

A Figura 16 apresenta o gráfico original de Currie et al. (2021). A Tabela 12 apresenta uma aproximação dos valores apresentados no estudo, em decímetros cúbicos. Os valores apresentados não são os valores exatos do trabalho original, tendo como objetivo apresentar o problema com valores mais didáticos.

Figura 16 – Volume corporal e desvio padrão de 6 diferentes orcas-pigmeias medidos em 4 diferentes ocasiões com diversas técnicas de medição nas águas costeiras da baía de Ma'alaea, no Maui, em Setembro e Outubro de 2019.



Fonte: (CURRIE et al., 2021)

Tabela 12 – Valores aproximados dos apresentados em Currie et al. (2021).

Medição	Indivíduo	Dia	Volume ( $dm^3$ )	Desvio Padrão
1	1	1	159,4	4,5
2	2	1	151,1	1,9
3	3	1	153,3	3,7
4	4	1	151,1	1,9
5	5	1	144,3	3,7
6	6	1	137,4	4,6
7	1	7	149,1	6,0
8	2	7	149,1	4,9
9	3	7	134,1	0,9
10	4	7	132,5	2,5
11	5	7	123,0	1,9
12	6	7	123,0	3,9
13	1	11	134,1	0,9
14	2	11	144,9	5,1
15	5	11	105,6	2,5
16	6	11	115,1	3,9
17	1	17	129,5	0,5
18	2	17	131,2	0,0

Fonte: Currie et al. (2021) (adaptado)

## 9.2 Modelagem da Imprecisão

A Tabela 12 apresenta os dados do problema junto ao desvio padrão que representa a imprecisão da medida obtida.  $v$  é o valor do volume de uma medida e  $d$  o desvio padrão da mesma, uma proposta de como incluir essa imprecisão ao modelo é na forma de intervalo  $[v - d; v + d]$ . Usando números *fuzzy* é possível modelar essa imprecisão também e de diversas formas. Como primeira forma optamos por usar números *fuzzy* triangulares na forma  $(v - d; v; v + d)$ , que tem centro e suporte coincidentes ao da forma intervalar usada. Na forma *fuzzy* trapezoidal optamos por  $\left(v - \frac{3}{2}d; v - \frac{1}{2}d; v + \frac{1}{2}d; v + \frac{3}{2}d\right)$ , que tem um suporte um pouco maior que o triangular  $(v - d; v; v + d)$ , sendo assim um modelo mais permissivo que admite valores fora do intervalo  $[v - d; v + d]$ . A Tabela 13 apresenta os dados modelados nessas três formas.

Tabela 13 – Dados do problema modelados na forma de intervalos, números *fuzzy* triangulares e números *fuzzy* trapezoidais.

medição	indivíduo	dia	intervalo	<i>fuzzy</i> triangular	<i>fuzzy</i> trapezoidal
1	1	1	[154, 9; 163, 9]	(154, 9; 159, 4; 163, 9)	(152, 6; 157, 1; 166, 1; 161, 6)
2	1	2	[149, 2; 153, 0]	(149, 2; 151, 1; 153, 0)	(148, 3; 150, 2; 154, 0; 152, 0)
3	1	3	[149, 6; 157, 0]	(149, 6; 153, 3; 157, 0)	(147, 7; 151, 4; 158, 9; 155, 1)
4	1	4	[149, 2; 153, 0]	(149, 2; 151, 1; 153, 0)	(148, 3; 150, 2; 154, 0; 152, 0)
5	1	5	[140, 6; 148, 0]	(140, 6; 144, 3; 148, 0)	(138, 8; 142, 5; 149, 8; 146, 2)
6	1	6	[132, 8; 142, 0]	(132, 8; 137, 4; 142, 0)	(130, 5; 135, 1; 144, 3; 139, 7)
7	7	1	[143, 1; 155, 1]	(143, 1; 149, 1; 155, 1)	(140, 1; 146, 1; 158, 1; 152, 1)
8	7	2	[144, 2; 154, 0]	(144, 2; 149, 1; 154, 0)	(141, 8; 146, 7; 156, 4; 151, 6)
9	7	3	[133, 2; 135, 0]	(133, 2; 134, 1; 135, 0)	(132, 8; 133, 6; 135, 4; 134, 6)
10	7	4	[130, 0; 135, 0]	(130, 0; 132, 5; 135, 0)	(128, 8; 131, 2; 136, 2; 133, 8)
11	7	5	[121, 1; 124, 9]	(121, 1; 123, 0; 124, 9)	(120, 2; 122, 0; 125, 8; 124, 0)
12	7	6	[119, 1; 126, 9]	(119, 1; 123, 0; 126, 9)	(117, 2; 121, 0; 128, 9; 125, 0)
13	11	1	[133, 2; 135, 0]	(133, 2; 134, 1; 135, 0)	(132, 8; 133, 6; 135, 4; 134, 6)
14	11	2	[139, 8; 150, 0]	(139, 8; 144, 9; 150, 0)	(137, 2; 142, 4; 152, 6; 147, 4)
15	11	5	[103, 1; 108, 1]	(103, 1; 105, 6; 108, 1)	(101, 8; 104, 4; 109, 4; 106, 8)
16	11	6	[111, 2; 119, 0]	(111, 2; 115, 1; 119, 0)	(109, 2; 113, 2; 121, 0; 117, 0)
17	17	1	[129, 0; 130, 0]	(129, 0; 129, 5; 130, 0)	(128, 8; 129, 2; 130, 2; 129, 8)
18	17	2	[131, 2; 131, 2]	(131, 2; 131, 2; 131, 2)	(131, 2; 131, 2; 131, 2; 131, 2)

Fonte: Currie et al. (2021) (adaptado)

## 9.3 Caso Transversal

Chamamos de transversal o caso onde os dados são tratados como todos independentes. Usando os dados na forma de números reais, isto é, o caso clássico, podemos realizar o processo de regressão linear simples, como visto na Seção 1.2. Para os casos intervalar e *fuzzy* usamos os modelos apresentados na Seção 7.3.

### 9.3.1 Dados na forma de números reais

Para o caso clássico podemos realizar uma regressão linear simples, usando a variável  $t$  como o tempo medido em dias:  $t_1 = 1$ ,  $t_2 = 7$ ,  $t_3 = 11$  e  $t_4 = 17$ . Para encontrar um modelo na forma  $f(t) = \beta_0 + \beta_1 t$ , usamos a matriz de planejamento:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & 7 \\ \vdots & \vdots \\ 1 & 11 \\ \vdots & \vdots \\ 1 & 17 \\ \vdots & \vdots \end{pmatrix} \begin{array}{l} \left. \vphantom{\begin{matrix} 1 \\ \vdots \end{matrix}} \right\} \text{seis linhas} \\ \left. \vphantom{\begin{matrix} 1 \\ \vdots \end{matrix}} \right\} \text{seis linhas} \\ \left. \vphantom{\begin{matrix} 1 \\ \vdots \end{matrix}} \right\} \text{quatro linhas} \\ \left. \vphantom{\begin{matrix} 1 \\ \vdots \end{matrix}} \right\} \text{duas linhas} \end{array}.$$

Obtemos então a matriz

$$\mathbf{W} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T = (\mathbf{W}_1 \quad \mathbf{W}_2 \quad \dots \quad \mathbf{W}_{18}).$$

Realizando as operações básicas de matrizes na expressão  $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$  obtemos a matriz  $\mathbf{W}$ . Essa matriz possui 18 colunas, denotadas  $\mathbf{W}_1, \dots, \mathbf{W}_{18}$ , muitas delas possuem valores iguais, apresentamos aqui os valores dessas colunas:

$$\begin{aligned} \mathbf{W}_1 = \dots = \mathbf{W}_6 &= \begin{pmatrix} 0,14306 \\ -0,01250 \end{pmatrix}, \\ \mathbf{W}_7 = \dots = \mathbf{W}_{12} &= \begin{pmatrix} 0,05556 \\ 0 \end{pmatrix}, \\ \mathbf{W}_{13} = \dots = \mathbf{W}_{16} &= \begin{pmatrix} -0,00278 \\ 0,00833 \end{pmatrix} \text{ e} \\ \mathbf{W}_{17} = \mathbf{W}_{18} &= \begin{pmatrix} -0,09028 \\ 0,02083 \end{pmatrix}. \end{aligned} \tag{9.1}$$

Usando os dados do problema na forma de um vetor coluna  $\mathbf{y}$ :

$$\mathbf{y} = \begin{pmatrix} 159,4 \\ 151,1 \\ 153,3 \\ 151,1 \\ 144,3 \\ 137,4 \\ 149,1 \\ 149,1 \\ 134,1 \\ 132,5 \\ 123,0 \\ 123,0 \\ 134,1 \\ 144,9 \\ 105,6 \\ 115,1 \\ 129,5 \\ 131,2 \end{pmatrix},$$

podemos usar a Equação 1.9,  $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ , para encontrar a estimativa  $\hat{\beta}$ :

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = \mathbf{W} \mathbf{y} = \begin{pmatrix} 148,3846 \\ -1,612 \end{pmatrix},$$

obtendo assim o modelo:

$$\hat{f}(t) = 148,3846 - 1,612t. \quad (9.2)$$

O coeficiente angular do modelo  $-1,612$  informa que com o passar do tempo o volume corporal dos indivíduos diminui. A Figura 18(a) apresenta o gráfico desse modelo. A Tabela 14 apresenta os valores dos dados observados,  $y_i$ , das estimativas,  $\hat{y}_i$  e os erros,  $e_i = \hat{y}_i - y_i$ . A Figura 19(a) apresenta esses erros de forma gráfica. Já a Figura 20(a) apresenta o gráfico quantil-quantil para esse conjunto de erros, onde a proximidade dos pontos à reta  $y = x$  é satisfatória.

Note que o caso clássico trata os dados como números reais, portanto sem uma modelagem da imprecisão. Dessa forma temos o valor esperado, a média, ou seja, o valor  $f(t)$ , apresentada pelo modelo da mesma forma como um número real. É necessário levar em conta que esse modelo foi feito de forma transversal, ou seja, medidas que pertencem a um mesmo indivíduo estão sendo tratadas como independentes. Esse mesmo problema se faz presente nos modelos onde a imprecisão dos dados é modelada.

Tabela 14 – Dados, estimativas e erros dados pelo modelo da Equação 9.2,  $\hat{f}(t) = 148,3846 - 1,612t$ .

medição	dado	estimativa	erro
1	$y_1 = 159,4$	$\hat{y}_1 = \hat{f}(1) = 146,77$	$e_1 = -12,63$
2	$y_2 = 151,1$	$\hat{y}_2 = \hat{f}(1) = 146,77$	$e_2 = -4,33$
3	$y_3 = 153,3$	$\hat{y}_3 = \hat{f}(1) = 146,77$	$e_3 = -6,53$
4	$y_4 = 151,1$	$\hat{y}_4 = \hat{f}(1) = 146,77$	$e_4 = -4,33$
5	$y_5 = 144,3$	$\hat{y}_5 = \hat{f}(1) = 146,77$	$e_5 = 2,47$
6	$y_6 = 137,4$	$\hat{y}_6 = \hat{f}(1) = 146,77$	$e_6 = 9,37$
7	$y_7 = 149,1$	$\hat{y}_7 = \hat{f}(7) = 137,10$	$e_7 = -12,00$
8	$y_8 = 149,1$	$\hat{y}_8 = \hat{f}(7) = 137,10$	$e_8 = -12,00$
9	$y_9 = 134,1$	$\hat{y}_9 = \hat{f}(7) = 137,10$	$e_9 = 3,00$
10	$y_{10} = 132,5$	$\hat{y}_{10} = \hat{f}(7) = 137,10$	$e_{10} = 4,60$
11	$y_{11} = 123,0$	$\hat{y}_{11} = \hat{f}(7) = 137,10$	$e_{11} = 14,10$
12	$y_{12} = 123,0$	$\hat{y}_{12} = \hat{f}(7) = 137,10$	$e_{12} = 14,10$
13	$y_{13} = 134,1$	$\hat{y}_{13} = \hat{f}(11) = 130,65$	$e_{13} = -3,45$
14	$y_{14} = 144,9$	$\hat{y}_{14} = \hat{f}(11) = 130,65$	$e_{14} = -14,25$
15	$y_{15} = 105,6$	$\hat{y}_{15} = \hat{f}(11) = 130,65$	$e_{15} = 25,05$
16	$y_{16} = 115,1$	$\hat{y}_{16} = \hat{f}(11) = 130,65$	$e_{16} = 15,55$
17	$y_{17} = 129,5$	$\hat{y}_{17} = \hat{f}(17) = 120,98$	$e_{17} = -8,52$
18	$y_{18} = 131,2$	$\hat{y}_{18} = \hat{f}(17) = 120,98$	$e_{18} = -10,22$

Fonte: O autor.

### 9.3.2 Dados Imprecisos

Para os três casos em que a imprecisão está modelada como intervalo, número *fuzzy* triangular ou trapezoidal temos casos muito semelhantes. Nos três casos é usada a matriz  $\mathbf{W}$  da Equação 9.1 para obter o modelo da Equação 7.14 ou o modelo da Equação 7.15. A diferença entre os três casos é a forma como os dados estão modelados.

Nos três casos cada dado  $i$  tem peso  $(W_{1,i} + t W_{2,i})$  no instante  $t$ :

$$\begin{aligned}
 W_{1,1} + t W_{2,1} &= 0,14306 - 0,01250t, \\
 &\vdots \\
 W_{1,6} + t W_{2,6} &= 0,14306 - 0,01250t, \\
 W_{1,7} + t W_{2,7} &= 0,05556, \\
 &\vdots \\
 W_{1,12} + t W_{2,12} &= 0,05556, \\
 W_{1,13} + t W_{2,13} &= -0,00278 + 0,00833t, \\
 &\vdots \\
 W_{1,16} + t W_{2,16} &= -0,00278 + 0,00833t, \\
 W_{1,17} + t W_{2,17} &= -0,09028 + 0,02083t \text{ e} \\
 W_{1,18} + t W_{2,18} &= -0,09028 + 0,02083t.
 \end{aligned} \tag{9.3}$$

O gráfico apresentando esses pesos está presente na Figura 17.

O peso dado às medidas 1 a 6 no instante  $t$  é dado por  $0,14306 - 0,01250t$ . Assim observamos que a influência que estes indivíduos tem no resultado final diminui ao se aumentar o valor de  $t$ . Para  $t = 1$  o peso é de  $0,13056$ , esse peso decresce linearmente assumindo valor zero em  $t = 11,4448$  e assume em  $t = 17$  o valor de  $-0,069440$ . Interpretamos assim que estes dados possuem maior influência no resultado final quando  $t$  é pequeno. Notamos que para  $t > 11,4448$  o peso dado a essas observações é negativo, nessa região esses dados possuem então o papel de aumentar a imprecisão do modelo, tornando os intervalos, ou conjunto suporte nos casos *fuzzy*, com comprimento maior.

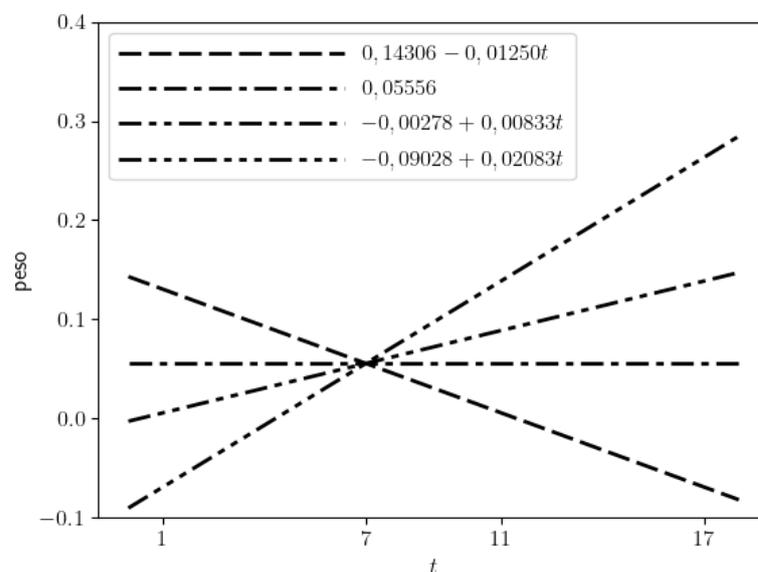
O peso dado às medidas 7 a 12 é constante igual a  $0,05556$ . Desta forma para qualquer valor de  $t$  esses indivíduos possuem sempre a mesma importância naquele ponto para o modelo.

As medidas 13 a 16 possuem o peso  $W_{2,13} = -0,00278 + 0,00833t$ . Aqui temos valores negativos para os pesos para valores de  $t$  menores que  $0,3337$ , agregando maior imprecisão aos valores dados pelo modelo nessa região. Quando  $t$  aumenta o peso dado a esses indivíduos também aumenta, assim a sua influência no modelo também aumenta.

As medidas 17 e 18 possuem pesos negativos para  $t < 4,334$ . Assim essas medidas contribuem para o aumento da imprecisão para essa região. Seu peso também aumenta a medida que  $t$  aumenta.

Essa análise dos termos  $W_{1,i} + t W_{2,i}$  é válida para os três modelos apresentados nessa seção.

Figura 17 – Gráfico dos pesos  $W_{1,i} + t W_{2,i}$  da Equação 9.3.



Fonte: O autor.

### 9.3.2.1 Dados na Forma de Intervalos

No caso onde as observações estão modeladas como intervalos, usamos os dados na forma  $[v - d; v + d]$ , onde  $v$  é o valor do volume de uma medida e  $d$  o desvio padrão da mesma. Temos assim o conjunto de dados:

$$\begin{aligned}
\overline{Y}_1 &= [154, 9; 163, 9], \\
\overline{Y}_2 &= [149, 2; 153, 0], \\
\overline{Y}_3 &= [149, 6; 157, 0], \\
\overline{Y}_4 &= [149, 2; 153, 0], \\
\overline{Y}_5 &= [140, 6; 148, 0], \\
\overline{Y}_6 &= [132, 8; 142, 0], \\
\overline{Y}_7 &= [143, 1; 155, 1], \\
\overline{Y}_8 &= [144, 2; 154, 0], \\
\overline{Y}_9 &= [133, 2; 135, 0], \\
\overline{Y}_{10} &= [130, 0; 135, 0], \\
\overline{Y}_{11} &= [121, 1; 124, 9], \\
\overline{Y}_{12} &= [119, 1; 126, 9], \\
\overline{Y}_{13} &= [133, 2; 135, 0], \\
\overline{Y}_{14} &= [139, 8; 150, 0], \\
\overline{Y}_{15} &= [103, 1; 108, 1], \\
\overline{Y}_{16} &= [111, 2; 119, 0], \\
\overline{Y}_{17} &= [129, 0; 130, 0] \text{ e} \\
\overline{Y}_{18} &= [131, 2; 131, 2].
\end{aligned}$$

Com a matriz  $\mathbf{W}$  da Equação 9.1, segue diretamente da fórmula dada pela Equação 7.14 o modelo:

$$\overline{S(t)} = \sum_{i=1}^{18} (W_{1,i} + t W_{2,i}) \overline{Y}_i. \quad (9.4)$$

A Figura 18(b) exhibe o gráfico obtido nesse caso enquanto a Figura 19(b) mostra de forma gráfica o valor dos erros  $\overline{e}_i = \widehat{Y}_i \ominus_{gH} \overline{Y}_i$ , conforme o discutido na Seção 6.5. Por exemplo, o dado  $\overline{Y}_1 = [154, 9; 163, 9]$  se refere a uma medida tomada no tempo  $t = 1$ .

Assim a previsão dada pelo modelo é  $\widehat{Y}_1 = \overline{S(1)} = [142, 902; 150, 643]$ . O erro nesse caso é dado por:

$$\begin{aligned} \overline{e}_1 &= \widehat{Y}_1 \ominus_{gH} \overline{Y}_1 = \\ &= [\min\{142, 902 - 154, 9; 150, 643 - 163, 9\}; \max\{142, 902 - 154, 9; 150, 643 - 163, 9\}] = \\ &= [-13, 257; -11, 998]. \end{aligned}$$

A Figura 20(b) apresenta o gráfico quantil-quantil para esse conjunto de erros, onde a proximidade dos pontos à reta  $y = x$  é satisfatória. A Tabela 15 apresenta os valores observados na forma de intervalos, os valores estimados no instante correspondente e o valor dos erros. Note que neste caso os dados, as estimativas e os erros aparecem na forma de intervalos. Também é notável a queda no volume corporal dos golfinhos, passando de  $\overline{S(1)} = [142, 902; 150, 643]$  no primeiro dia para  $\overline{S(17)} = [116, 599; 125, 360]$  no décimo sétimo dia. No primeiro dia o modelo indica que o valor esperado do volume corporal dos golfinhos está no intervalo  $[142, 902; 150, 643]$ , assim sendo podemos dizer que o valor esperado é maior que  $142, 902dm^3$ . Já no dia 17 o valor esperado está no intervalo  $[116, 599; 125, 360]$ , o que indica que o valor esperado é menor que  $125, 360$ . O fato de no primeiro dia termos os valores maiores que  $142, 902$  e no dia 17 valores menores que  $125, 360$  evidenciando que os valores diminuíram nesse período.

É importante apontar que o modelo na forma  $\sum_{i=1}^{18} (W_{1,i} + t W_{2,i}) \overline{Y}_i$  não possui nenhum parâmetro que indique imediatamente uma tendência de queda ou de subida dos valores. Todos os parâmetros indicam apenas como se comporta o peso dado a cada dado no modelo final.

Tabela 15 – Dados, estimativas e erros dados pelo modelo da Equação 9.4,  $\overline{S(t)} =$

$$\sum_{i=1}^{18} (W_{1,i} + t W_{2,i}) \overline{Y}_i.$$

medição	dado	estimativa	erro ( $\overline{e}_i = \hat{Y}_i \ominus_{gH} \overline{Y}_i$ )
1	$\overline{Y}_1 = [154, 9; 163, 9]$	$\hat{Y}_1 = \overline{S(1)} = [142, 902; 150, 643]$	$\overline{e}_1 = [-13, 257; -11, 998]$
2	$\overline{Y}_2 = [149, 2; 153, 0]$	$\hat{Y}_2 = \overline{S(1)} = [142, 902; 150, 643]$	$\overline{e}_2 = [-6, 298; -2, 357]$
3	$\overline{Y}_3 = [149, 6; 157, 0]$	$\hat{Y}_3 = \overline{S(1)} = [142, 902; 150, 643]$	$\overline{e}_3 = [-6, 698; -6, 357]$
4	$\overline{Y}_4 = [149, 2; 153, 0]$	$\hat{Y}_4 = \overline{S(1)} = [142, 902; 150, 643]$	$\overline{e}_4 = [-6, 298; -2, 357]$
5	$\overline{Y}_5 = [140, 6; 148, 0]$	$\hat{Y}_5 = \overline{S(1)} = [142, 902; 150, 643]$	$\overline{e}_5 = [2, 302; 2, 643]$
6	$\overline{Y}_6 = [132, 8; 142, 0]$	$\hat{Y}_6 = \overline{S(1)} = [142, 902; 150, 643]$	$\overline{e}_6 = [8, 643; 10, 102]$
7	$\overline{Y}_7 = [143, 1; 155, 1]$	$\hat{Y}_7 = \overline{S(7)} = [134, 139; 140, 061]$	$\overline{e}_7 = [-15, 039; -8, 961]$
8	$\overline{Y}_8 = [144, 2; 154, 0]$	$\hat{Y}_8 = \overline{S(7)} = [134, 139; 140, 061]$	$\overline{e}_8 = [-13, 939; -10, 061]$
9	$\overline{Y}_9 = [133, 2; 135, 0]$	$\hat{Y}_9 = \overline{S(7)} = [134, 139; 140, 061]$	$\overline{e}_9 = [0, 939; 5, 061]$
10	$\overline{Y}_{10} = [130, 0; 135, 0]$	$\hat{Y}_{10} = \overline{S(7)} = [134, 139; 140, 061]$	$\overline{e}_{10} = [4, 139; 5, 061]$
11	$\overline{Y}_{11} = [121, 1; 124, 9]$	$\hat{Y}_{11} = \overline{S(7)} = [134, 139; 140, 061]$	$\overline{e}_{11} = [13, 039; 15, 161]$
12	$\overline{Y}_{12} = [119, 1; 126, 9]$	$\hat{Y}_{12} = \overline{S(7)} = [134, 139; 140, 061]$	$\overline{e}_{12} = [13, 161; 15, 039]$
13	$\overline{Y}_{13} = [133, 2; 135, 0]$	$\hat{Y}_{13} = \overline{S(11)} = [128, 251; 133, 053]$	$\overline{e}_{13} = [-4, 949; -1, 947]$
14	$\overline{Y}_{14} = [139, 8; 150, 0]$	$\hat{Y}_{14} = \overline{S(11)} = [128, 251; 133, 053]$	$\overline{e}_{14} = [-16, 947; -11, 549]$
15	$\overline{Y}_{15} = [103, 1; 108, 1]$	$\hat{Y}_{15} = \overline{S(11)} = [128, 251; 133, 053]$	$\overline{e}_{15} = [24, 953; 25, 151]$
16	$\overline{Y}_{16} = [111, 2; 119, 0]$	$\hat{Y}_{16} = \overline{S(11)} = [128, 251; 133, 053]$	$\overline{e}_{16} = [14, 053; 17, 051]$
17	$\overline{Y}_{17} = [129, 0; 130, 0]$	$\hat{Y}_{17} = \overline{S(17)} = [116, 599; 125, 360]$	$\overline{e}_{17} = [-12, 401; -4, 640]$
18	$\overline{Y}_{18} = [131, 2; 131, 2]$	$\hat{Y}_{18} = \overline{S(17)} = [116, 599; 125, 360]$	$\overline{e}_{18} = [-14, 601; -5, 840]$

Fonte: O autor.

### 9.3.2.2 Dados na Forma de Números *Fuzzy* Triangulares

Com os dados modelados na forma de números *fuzzy* triangulares:

$$Y_1^* = (154, 9; 159, 4; 163, 9),$$

$$Y_2^* = (149, 2; 151, 1; 153, 0),$$

$$Y_3^* = (149, 6; 153, 3; 157, 0),$$

$$Y_4^* = (149, 2; 151, 1; 153, 0),$$

$$Y_5^* = (140, 6; 144, 3; 148, 0),$$

$$Y_6^* = (132, 8; 137, 4; 142, 0),$$

$$Y_7^* = (143, 1; 149, 1; 155, 1),$$

$$Y_8^* = (144, 2; 149, 1; 154, 0),$$

$$Y_9^* = (133, 2; 134, 1; 135, 0),$$

$$\begin{aligned}
Y_{10}^* &= (130, 0; 132, 5; 135, 0), \\
Y_{11}^* &= (121, 1; 123, 0; 124, 9), \\
Y_{12}^* &= (119, 1; 123, 0; 126, 9), \\
Y_{13}^* &= (133, 2; 134, 1; 135, 0), \\
Y_{14}^* &= (139, 8; 144, 9; 150, 0), \\
Y_{15}^* &= (103, 1; 105, 6; 108, 1), \\
Y_{16}^* &= (111, 2; 115, 1; 119, 0), \\
Y_{17}^* &= (129, 0; 129, 5; 130, 0) \text{ e} \\
Y_{18}^* &= (131, 2; 131, 2; 131, 2).
\end{aligned}$$

Usando a matriz  $\mathbf{W}$  da Equação 9.1, segue diretamente da fórmula da Equação 7.15 o modelo:

$$f^*(t) = \sum_{i=1}^{18} (W_{1,i} + t W_{2,i}) Y_i^*. \quad (9.5)$$

A Figura 18(c) exibe o gráfico obtido nesse caso e Figura 19(c) exibe de forma gráfica o valor dos erros  $\bar{e}_i = [\hat{Y}_i^*]^0 \ominus_{gH} [Y_i^*]^0$ . Já a Figura 20(b) apresenta o gráfico quantil-quantil para esse conjunto de erros, onde a proximidade dos pontos à reta  $y = x$  é satisfatória. A Tabela 16 apresenta os valores observado modelados na forma de intervalos, os valores estimados no instante correspondente e o valor dos erros. Note que os erros coincidem com os do modelo intervalar, isso se deve ao fato de que os números *fuzzy* triangulares tem, nesse caso, o  $\alpha$ -nível zero idêntico aos intervalos do modelo anterior. A queda do volume corporal dos golfinhos também é visto nos valores dados pelo modelo pode ser observado do mesmo modo que foi feito no caso intervalar. Reforçamos que nenhum dos parâmetros do modelo  $\sum_{i=1}^{18} (W_{1,i} + t W_{2,i}) Y_i^*$  tem como objetivo apresentar a tendência de queda ou de subida dos dados ao longo do tempo, mas sim cada função linear  $W_{1,i} + t W_{2,i}$  indica o peso dado a cada dado para a média ponderada para cada valor de  $t$ .

Tabela 16 – Dados, estimativas e erros dados pelo modelo da Equação 9.5,

$$f^*(t) = \sum_{i=1}^{18} (W_{1,i} + t W_{2,i}) Y_i^*.$$

i	dado ( $Y_i^*$ )	estimativa ( $\hat{Y}_i^*$ )	erro ( $\bar{e}_i = [\hat{Y}_i^*]^0 \ominus_{gH} [Y_i^*]^0$ )
1	(154, 9; 159, 4; 163, 9)	(142, 902; 146, 773; 150, 643)	[-13, 257; -11, 998]
2	(149, 2; 151, 1; 153, 0)	(142, 902; 146, 773; 150, 643)	[-6, 298; -2, 357]
3	(149, 6; 153, 3; 157, 0)	(142, 902; 146, 773; 150, 643)	[-6, 698; -6, 357]
4	(149, 2; 151, 1; 153, 0)	(142, 902; 146, 773; 150, 643)	[-6, 298; -2, 357]
5	(140, 6; 144, 3; 148, 0)	(142, 902; 146, 773; 150, 643)	[2, 302; 2, 643]
6	(132, 8; 137, 4; 142, 0)	(142, 902; 146, 773; 150, 643)	[8, 643; 10, 102]
7	(143, 1; 149, 1; 155, 1)	(134, 139; 137, 100; 140, 061)	[-15, 039; -8, 961]
8	(144, 2; 149, 1; 154, 0)	(134, 139; 137, 100; 140, 061)	[-13, 939; -10, 061]
9	(133, 2; 134, 1; 135, 0)	(134, 139; 137, 100; 140, 061)	[0, 939; 5, 061]
10	(130, 0; 132, 5; 135, 0)	(134, 139; 137, 100; 140, 061)	[4, 139; 5, 061]
11	(121, 1; 123, 0; 124, 9)	(134, 139; 137, 100; 140, 061)	[13, 039; 15, 161]
12	(119, 1; 123, 0; 126, 9)	(134, 139; 137, 100; 140, 061)	[13, 161; 15, 039]
13	(133, 2; 134, 1; 135, 0)	(128, 251; 130, 652; 133, 053)	[-4, 949; -1, 947]
14	(139, 8; 144, 9; 150, 0)	(128, 251; 130, 652; 133, 053)	[-16, 947; -11, 549]
15	(103, 1; 105, 6; 108, 1)	(128, 251; 130, 652; 133, 053)	[24, 953; 25, 151]
16	(111, 2; 115, 1; 119, 0)	(128, 251; 130, 652; 133, 053)	[14, 053; 17, 051]
17	(129, 0; 129, 5; 130, 0)	(116, 599; 120, 979; 125, 360)	[-12, 401; -4, 640]
18	(131, 2; 131, 2; 131, 2)	(116, 599; 120, 979; 125, 360)	[-14, 601; -5, 840]

Fonte: O autor.

### 9.3.2.3 Dados na Forma de Números *Fuzzy* Trapezoidais

Outra forma de modelar a imprecisão desse caso é usando números *fuzzy* trapezoidais na forma  $\left(v - \frac{3}{2}d; v - \frac{1}{2}d; v + \frac{1}{2}d; v + \frac{3}{2}d\right)$ . Diferente dos casos intervalar e triangular usamos nesse modelo um  $\alpha$ -nível zero um pouco mais largo, representando uma possível maior imprecisão nos dados. Usando  $Q$  no lugar de  $Y$  apenas para diferenciar do modelo com números *fuzzy* triangulares da subseção anterior, apresentamos esses dados na forma:

$$Q_1^* = (152, 6; 157, 1; 161, 6; 166, 1),$$

$$Q_2^* = (148, 3; 150, 2; 152, 0; 154, 0),$$

$$Q_3^* = (147, 7; 151, 4; 155, 1; 158, 9),$$

$$Q_4^* = (148, 3; 150, 2; 152, 0; 154, 0),$$

$$Q_5^* = (138, 8; 142, 5; 146, 2; 149, 8),$$

$$Q_6^* = (130, 5; 135, 1; 139, 7; 144, 3),$$

$$Q_7^* = (140, 1; 146, 1; 152, 1; 158, 1),$$

$$Q_8^* = (141, 8; 146, 7; 151, 6; 156, 4),$$

$$\begin{aligned}
Q_9^* &= (132, 8; 133, 6; 134, 6; 135, 4), \\
Q_{10}^* &= (128, 8; 131, 2; 133, 8; 136, 2), \\
Q_{11}^* &= (120, 2; 122, 0; 124, 0; 125, 8), \\
Q_{12}^* &= (117, 2; 121, 0; 125, 0; 128, 9), \\
Q_{13}^* &= (132, 8; 133, 6; 134, 6; 135, 4), \\
Q_{14}^* &= (137, 2; 142, 4; 147, 4; 152, 6), \\
Q_{15}^* &= (101, 8; 104, 4; 106, 8; 109, 4), \\
Q_{16}^* &= (109, 2; 113, 2; 117, 0; 121, 0), \\
Q_{17}^* &= (128, 8; 129, 2; 129, 8; 130, 2) \text{ e} \\
Q_{18}^* &= (131, 2; 131, 2; 131, 2; 131, 2).
\end{aligned}$$

Novamente com a matriz  $\mathbf{W}$  da Equação 9.1, segue diretamente da fórmula da Equação 7.15 o modelo:

$$q^*(t) = \sum_{i=1}^{18} (W_{1,i} + t W_{2,i}) Q_i^*. \quad (9.6)$$

A Figura 18(d) exibe o gráfico obtido nesse caso. A Figura 19(d) exibe de forma gráfica o valor dos erros  $\bar{e}_i = [\hat{Q}_i^*]^0 \ominus_{gH} [Q_i^*]^0$ . Já a Figura 20(c) apresenta o gráfico quantil-quantil para esse conjunto de erros, onde a proximidade dos pontos à reta  $y = x$  é satisfatória. A Tabela 17 apresenta os valores observado modelados na forma de intervalos, os valores estimados no instante correspondente e o valor dos erros. Na A Figura 18(d) é visível que a faixa de valores do modelo se torna mais larga, sendo assim um modelo com mais possibilidades para o peso real dos indivíduos em cada instante. Apesar dessas diferenças esse modelo também apresenta uma tendência de queda no volume corporal dos indivíduos com o passar do tempo. Para  $t = 1$  temos  $q^*(1) = (140, 97; 144, 84; 148, 71; 152, 58)$ , situação em que apenas valores maiores que 140, 97 tem pertinência maior que zero enquanto para  $t = 17$  temos  $q^*(17) = (114, 41; 118, 79; 123, 17; 127, 55)$ , onde apenas valores menores que 127, 55 possuem pertinência maior que zero no número *fuzzy*. Deste modo é evidente que o valor previsto para  $t = 1$  é maior que o valor previsto para  $t = 17$ .

Tabela 17 – Dados, estimativas e erros dados pelo modelo da Equação 9.6,  $q^*(t) = \sum_{i=1}^{18} (W_{1,i} + t W_{2,i}) Q_i^*$ .

i	dado ( $Q_i^*$ )	estimativa ( $\hat{Q}_i^*$ )	erro ( $\bar{e}_i = [\hat{Q}_i^*]^0 \ominus_{gH} [Q_i^*]^0$ )
1	(152, 65; 157, 15; 161, 65; 166, 15)	(140, 97; 144, 84; 148, 71; 152, 58)	[-13, 57; -11, 68]
2	(148, 25; 150, 15; 152, 05; 153, 95)	(140, 97; 144, 84; 148, 71; 152, 58)	[-7, 28; -1, 37]
3	(147, 75; 151, 45; 155, 15; 158, 85)	(140, 97; 144, 84; 148, 71; 152, 58)	[-6, 78; -6, 27]
4	(148, 25; 150, 15; 152, 05; 153, 95)	(140, 97; 144, 84; 148, 71; 152, 58)	[-7, 28; -1, 37]
5	(138, 75; 142, 45; 146, 15; 149, 85)	(140, 97; 144, 84; 148, 71; 152, 58)	[2, 22; 2, 73]
6	(130, 50; 135, 10; 139, 70; 144, 30)	(140, 97; 144, 84; 148, 71; 152, 58)	[8, 28; 10, 47]
7	(140, 10; 146, 10; 152, 10; 158, 10)	(132, 66; 135, 62; 138, 58; 141, 54)	[-16, 56; -7, 44]
8	(141, 75; 146, 65; 151, 55; 156, 45)	(132, 66; 135, 62; 138, 58; 141, 54)	[-14, 91; -9, 09]
9	(132, 75; 133, 65; 134, 55; 135, 45)	(132, 66; 135, 62; 138, 58; 141, 54)	[-0, 09; 6, 09]
10	(128, 75; 131, 25; 133, 75; 136, 25)	(132, 66; 135, 62; 138, 58; 141, 54)	[3, 91; 5, 29]
11	(120, 15; 122, 05; 123, 95; 125, 85)	(132, 66; 135, 62; 138, 58; 141, 54)	[12, 51; 15, 69]
12	(117, 15; 121, 05; 124, 95; 128, 85)	(132, 66; 135, 62; 138, 58; 141, 54)	[12, 69; 15, 51]
13	(132, 75; 133, 65; 134, 55; 135, 45)	(127, 05; 129, 45; 131, 85; 134, 25)	[-5, 70; -1, 20]
14	(137, 25; 142, 35; 147, 45; 152, 55)	(127, 05; 129, 45; 131, 85; 134, 25)	[-18, 30; -10, 20]
15	(101, 85; 104, 35; 106, 85; 109, 35)	(127, 05; 129, 45; 131, 85; 134, 25)	[24, 90; 25, 20]
16	(109, 25; 113, 15; 117, 05; 120, 95)	(127, 05; 129, 45; 131, 85; 134, 25)	[13, 30; 17, 80]
17	(128, 75; 129, 25; 129, 75; 130, 25)	(114, 41; 118, 79; 123, 17; 127, 55)	[-14, 34; -2, 70]
18	(131, 20; 131, 20; 131, 20; 131, 20)	(114, 41; 118, 79; 123, 17; 127, 55)	[-16, 79; -3, 65]

Fonte: O autor.

### 9.3.2.4 Figuras e Considerações

Apresentamos quatro modelos nessa seção. O modelo clássico com os dados na forma de números reais,  $\hat{f}(t) = 148,3846 - 1,612t$ . O modelo intervalar,  $\overline{S}(t) = \sum_{i=1}^{18} (W_{1,i} + t W_{2,i}) \overline{Y}_i$ . O modelo com números *fuzzy* triangulares,  $f^*(t) = \sum_{i=1}^{18} (W_{1,i} + t W_{2,i}) Y_i^*$ .

E o modelo com números *fuzzy* trapezoidais,  $q^*(t) = \sum_{i=1}^{18} (W_{1,i} + t W_{2,i}) Q_i^*$ . Tanto o modelo intervalar quanto os dois modelos *fuzzy* diferem apenas na forma em que os dados são apresentados, os pesos  $W_{1,i} + t W_{2,i}$  são os mesmos. Deste modo não é visível nenhuma forma de ganho em usar os dados de forma intervalar ou *fuzzy*, porém, a forma como a imprecisão se faz presente no modelo agrega informação ao caso real. Obtemos tanto a informação da imprecisão em cada ponto quanto a informação do peso que cada dado tem em cada instante  $t$  no resultado final.

Todos os modelos apresentados nessa seção tratam os dados observados como independentes, porém, é possível usar o conhecimento de que certas medidas pertencem a um mesmo indivíduo para realizar outro tipo de modelo. A próxima seção apresenta modelos que tratam esses dados de forma longitudinal, o que oferece uma visão diferente do problema.

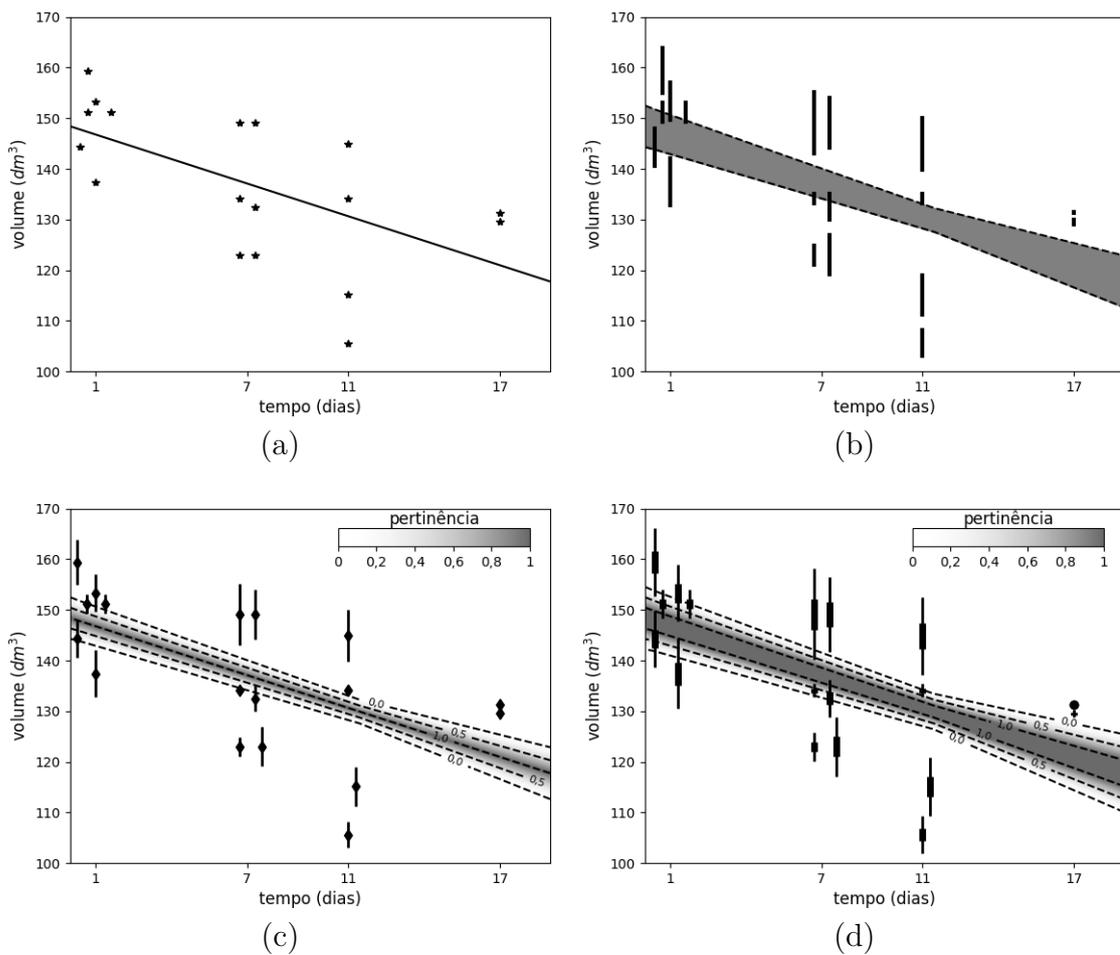
Figura 18 – Gráficos para as regressões em diferentes formas de modelagem dos dados.

(a) Dados reais.

(b) Dados modelados como intervalos reais. Linhas verticais representam os intervalos, a faixa cinza apresenta o intervalo do modelo de regressão para cada ponto  $t$  no tempo.

(c) Dados modelados como números *fuzzy* triangulares. As linhas verticais apresentam o  $\alpha$ -nível zero dos números *fuzzy*, o ponto no centro dos mesmos apresenta o seu  $\alpha$ -nível 1. A curva de nível apresenta a pertinência dos números *fuzzy* do modelo de regressão em cada ponto  $(t, y)$ .

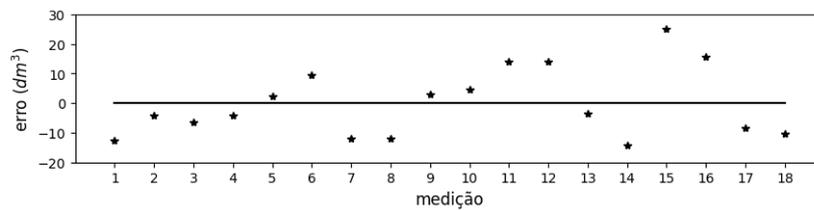
(d) Dados modelados como números *fuzzy* trapezoidais. As linhas verticais mais finas apresentam o  $\alpha$ -nível zero dos números *fuzzy*, as linhas mais grossas no centro dos mesmos apresenta o seu  $\alpha$ -nível 1. A curva de nível apresenta a pertinência dos números *fuzzy* do modelo de regressão em cada ponto  $(t, y)$ . Os dados foram tomados nos tempos 1, 7, 11 e 17, nos gráficos eles apresentam alguma variação nesses valores para melhor visualização.



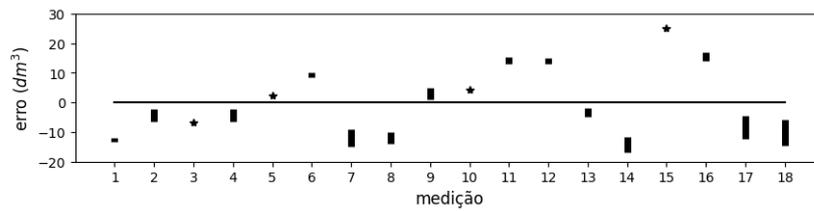
Fonte: O autor.

Figura 19 – Gráficos exibindo os erros para as regressões em diferentes formas de modelagem dos dados.

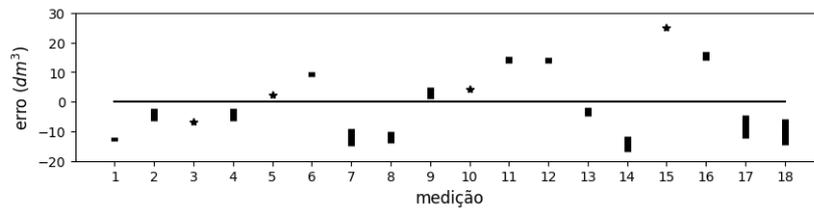
- (a) Dados reais.
- (b) Dados modelados como intervalos reais.
- (c) Dados modelados como números *fuzzy* triangulares.
- (d) Dados modelados como números *fuzzy* trapezoidais.



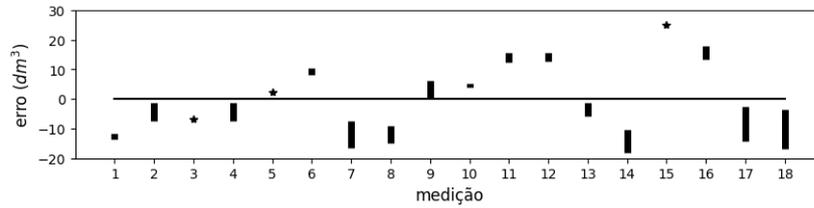
(a)



(b)



(c)



(d)

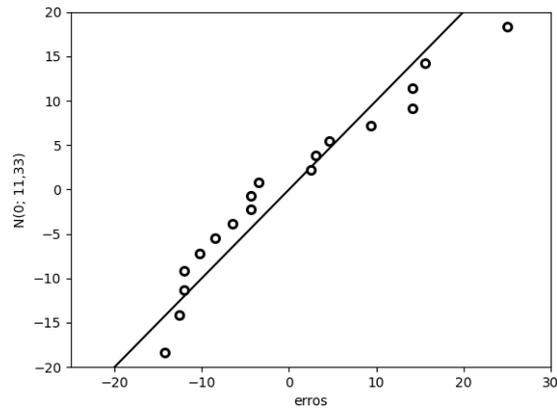
Fonte: O autor.

Figura 20 – Gráficos quantil-quantil para as regressões em diferentes formas de modelagem dos dados.

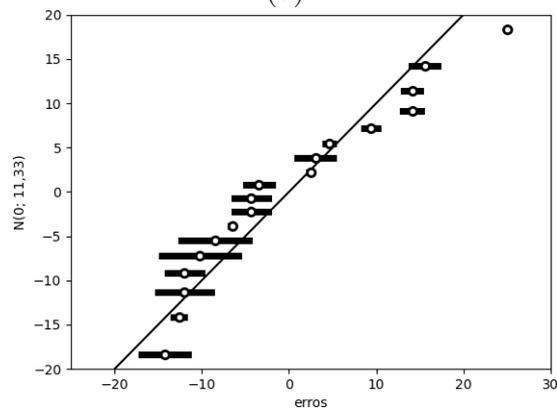
(a) Dados reais.

(b) Dados modelados como intervalos reais e dados modelados como números *fuzzy* triangulares.

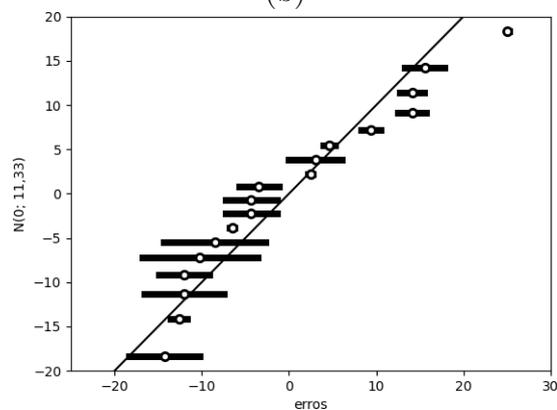
(c) Dados modelados como números *fuzzy* trapezoidais.



(a)



(b)



(c)

Fonte: O autor.

## 9.4 Caso Longitudinal

Outra forma de tratar o problema do volume corporal das orcas-pigmeias é levando em conta que medidas de um mesmo indivíduo não sejam independentes, considerando que existe certa correlação entre as mesmas. Os dados tratados dessa forma são denominados longitudinais, como discutido no Capítulo 2, para o caso clássico, e no Capítulo 8, com os casos intervalar e *fuzzy*.

### 9.4.1 Modelagem Clássica

Para o caso longitudinal clássico denotamos os dados como:

$$\mathbf{Y}_1 = \begin{pmatrix} Y_{1,1} \\ Y_{1,2} \\ Y_{1,3} \\ Y_{1,4} \end{pmatrix} = \begin{pmatrix} 159, 4 \\ 149, 1 \\ 134, 1 \\ 129, 5 \end{pmatrix}, \quad \mathbf{Y}_2 = \begin{pmatrix} Y_{2,1} \\ Y_{2,2} \\ Y_{2,3} \\ Y_{2,4} \end{pmatrix} = \begin{pmatrix} 151, 1 \\ 149, 1 \\ 144, 9 \\ 131, 2 \end{pmatrix}, \quad \mathbf{Y}_3 = \begin{pmatrix} Y_{3,1} \\ Y_{3,2} \end{pmatrix} = \begin{pmatrix} 153, 3 \\ 134, 1 \end{pmatrix},$$

$$\mathbf{Y}_4 = \begin{pmatrix} Y_{4,1} \\ Y_{4,2} \end{pmatrix} = \begin{pmatrix} 151, 1 \\ 132, 5 \end{pmatrix}, \quad \mathbf{Y}_5 = \begin{pmatrix} Y_{5,1} \\ Y_{5,2} \\ Y_{5,3} \end{pmatrix} = \begin{pmatrix} 144, 3 \\ 123, 0 \\ 105, 6 \end{pmatrix} \quad \text{e} \quad \mathbf{Y}_6 = \begin{pmatrix} Y_{6,1} \\ Y_{6,2} \\ Y_{6,3} \end{pmatrix} = \begin{pmatrix} 137, 4 \\ 123, 0 \\ 115, 1 \end{pmatrix},$$

onde  $Y_{i,j}$  é a medida tomada do indivíduo  $i$  no tempo  $t_j$ . Nesse problema temos  $t_1 = 1$ ,  $t_2 = 7$ ,  $t_3 = 11$  e  $t_4 = 17$ .

Como discutido na Seção 2.3, usamos o modelo na forma:

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i.$$

Levando em conta que os vetores  $\mathbf{Y}_i$  são identicamente distribuídos com distribuições normais  $\mathbf{Y}_i \sim N(\mathbf{X}_i \boldsymbol{\beta}, \boldsymbol{\Sigma}_i)$ , com média  $\mathbf{X}_i \boldsymbol{\beta}$  e matriz de variância-covariância  $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}_{m_i \times m_i} + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T$ . Como os  $\mathbf{Y}_i$  são independentes encontramos estimativas para os valores  $\boldsymbol{\beta}$ ,  $\mathbf{b}_i$ ,  $\sigma^2$  e  $\mathbf{D}$  através de máxima verossimilhança, como discutido na Seção 2.3.

Escolhemos arbitrariamente as funções  $\mathbf{g}_x(t)$  e  $\mathbf{g}_z(t)$  para assim obter as matrizes de planejamento  $\mathbf{X}_i$  e  $\mathbf{Z}_i$ . Escolhemos assim as funções:

$$\mathbf{g}_x(t) = \begin{pmatrix} 1 \\ t \end{pmatrix} \quad \text{e} \quad \mathbf{g}_z(t) = \begin{pmatrix} 1 \end{pmatrix}.$$

Desta forma  $\mathbf{g}_x(t)$  indica que foi escolhido para a média geral do modelo uma função linear na forma  $\beta_0 + \beta_1 t$  e  $\mathbf{g}_z(t)$  que o efeito individual é dado por apenas  $b_i$ .

Com  $\mathbf{g}_x(t)$  e  $\mathbf{g}_z(t)$  obtemos as matrizes de planejamento:

$$\mathbf{X}_1 = \mathbf{X}_2 = \begin{pmatrix} \mathbf{g}_x(t_1) \\ \mathbf{g}_x(t_2) \\ \mathbf{g}_x(t_3) \\ \mathbf{g}_x(t_4) \end{pmatrix} = \begin{pmatrix} \mathbf{g}_x(1) \\ \mathbf{g}_x(7) \\ \mathbf{g}_x(11) \\ \mathbf{g}_x(17) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 7 \\ 1 & 11 \\ 1 & 17 \end{pmatrix},$$

$$\mathbf{Z}_1 = \mathbf{Z}_2 = \begin{pmatrix} \mathbf{g}_z(t_1) \\ \mathbf{g}_z(t_2) \\ \mathbf{g}_z(t_3) \\ \mathbf{g}_z(t_4) \end{pmatrix} = \begin{pmatrix} \mathbf{g}_z(1) \\ \mathbf{g}_z(7) \\ \mathbf{g}_z(11) \\ \mathbf{g}_z(17) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix},$$

$$\mathbf{X}_3 = \mathbf{X}_4 = \begin{pmatrix} \mathbf{g}_x(t_1) \\ \mathbf{g}_x(t_2) \end{pmatrix} = \begin{pmatrix} \mathbf{g}_x(1) \\ \mathbf{g}_x(7) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 7 \end{pmatrix},$$

$$\mathbf{Z}_3 = \mathbf{Z}_4 = \begin{pmatrix} \mathbf{g}_z(t_1) \\ \mathbf{g}_z(t_2) \end{pmatrix} = \begin{pmatrix} \mathbf{g}_z(1) \\ \mathbf{g}_z(7) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

$$\mathbf{X}_5 = \mathbf{X}_6 = \begin{pmatrix} \mathbf{g}_x(t_1) \\ \mathbf{g}_x(t_2) \\ \mathbf{g}_x(t_3) \end{pmatrix} = \begin{pmatrix} \mathbf{g}_x(1) \\ \mathbf{g}_x(7) \\ \mathbf{g}_x(11) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 7 \\ 1 & 11 \end{pmatrix}$$

e

$$\mathbf{Z}_5 = \mathbf{Z}_6 = \begin{pmatrix} \mathbf{g}_z(t_1) \\ \mathbf{g}_z(t_2) \\ \mathbf{g}_z(t_3) \end{pmatrix} = \begin{pmatrix} \mathbf{g}_z(1) \\ \mathbf{g}_z(7) \\ \mathbf{g}_z(11) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Realizando assim a máxima verossimilhança, Equação 2.12 da Seção 2.3:

$$\arg \max_{\beta, \sigma^2, \mathbf{D}} \left( -\frac{1}{2} \sum_{i=1}^6 \left( \log(\det(2\pi \Sigma_i)) + (\mathbf{Y}_i - \mathbf{X}_i \beta)^T \Sigma_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \beta) \right) \right),$$

onde  $\Sigma_i = \sigma^2 \mathbf{I}_{m_i \times m_i} + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T$  é uma matriz positiva definida. Obtemos então as estimativas:

$$\hat{\beta} = \begin{pmatrix} 150, 33 \\ -2, 05 \end{pmatrix},$$

$$\hat{\sigma}^2 = 30,76$$

e

$$\hat{\mathbf{D}} = (91, 90).$$

Desta forma, usando  $\Sigma_i = \sigma^2 \mathbf{I}_{m_i \times m_i} + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T$ , temos as matrizes de variância-covariância para cada indivíduo:

$$\hat{\Sigma}_1 = \hat{\Sigma}_2 = \begin{pmatrix} 122,66 & 91,9 & 91,9 & 91,9 \\ 91,9 & 122,66 & 91,9 & 91,9 \\ 91,9 & 91,9 & 122,66 & 91,9 \\ 91,9 & 91,9 & 91,9 & 122,66 \end{pmatrix},$$

$$\hat{\Sigma}_3 = \hat{\Sigma}_4 = \begin{pmatrix} 122,66 & 91,9 \\ 91,9 & 122,66 \end{pmatrix}$$

e

$$\hat{\Sigma}_5 = \hat{\Sigma}_6 = \begin{pmatrix} 122,66 & 91,9 & 91,9 \\ 91,9 & 122,66 & 91,9 \\ 91,9 & 91,9 & 122,66 \end{pmatrix}.$$

Todas as variâncias tem o valor de 122,6 e as covariâncias um valor de 91,9. O que nos fornece uma correlação de  $\frac{91,9}{\sqrt{122,6}\sqrt{122,6}} = 0,75$  entre medidas de tempos  $t_i$  distintos.

Usando as matrizes  $\hat{\Sigma}_i$  podemos obter os  $\hat{\mathbf{b}}_i = \hat{\mathbf{D}} \mathbf{Z}_i^T \hat{\Sigma}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})$ :

$$\hat{\mathbf{b}}_1 = (10, 31), \quad \hat{\mathbf{b}}_2 = (11, 28), \quad \hat{\mathbf{b}}_3 = (1, 35),$$

$$\hat{\mathbf{b}}_4 = (-0, 27), \quad \hat{\mathbf{b}}_5 = (-11, 72) \quad \text{e} \quad \hat{\mathbf{b}}_6 = (-10, 94).$$

Essa escolha nos fornece um modelo na forma de uma função afim para a média global, o valor esperado para um indivíduo desse conjunto de dados para cada instante  $t$ . Essa média para os valores  $t_i$  é dada por:

$$\hat{\boldsymbol{\mu}} = \mathbf{X} \hat{\boldsymbol{\beta}} = \begin{pmatrix} 1 & 1 \\ 1 & 7 \\ 1 & 11 \\ 1 & 17 \end{pmatrix} \begin{pmatrix} 150,33 \\ -2,05 \end{pmatrix} = \begin{pmatrix} 148,28 \\ 135,98 \\ 127,78 \\ 115,48 \end{pmatrix},$$

ou seja, a média para o instante  $t_1 = 1$  é de 148,28, para o instante  $t_2 = 7$  é de 135,98, para o instante  $t_3 = 11$  é de 127,78 e para  $t_4 = 17$  é de 115,48. Para outros instantes de tempo podemos usar a função  $\mu(t) = \mathbf{g}_x(t)^T \hat{\boldsymbol{\beta}} = 150,33 - 2,05t$ . O gráfico dessa função está presente na Figura 27 junto das medidas dos indivíduos.

Obtemos também as estimativas  $\hat{\mathbf{Y}}_i = \mathbf{X}_i \hat{\boldsymbol{\beta}} + \mathbf{Z}_i \hat{\mathbf{b}}_i$  para cada um dos indivíduos:

$$\hat{\mathbf{Y}}_1 = \begin{pmatrix} \hat{Y}_{1,1} \\ \hat{Y}_{1,2} \\ \hat{Y}_{1,3} \\ \hat{Y}_{1,4} \end{pmatrix} = \begin{pmatrix} 158, 59 \\ 146, 29 \\ 138, 09 \\ 125, 79 \end{pmatrix}, \quad \hat{\mathbf{Y}}_2 = \begin{pmatrix} \hat{Y}_{2,1} \\ \hat{Y}_{2,2} \\ \hat{Y}_{2,3} \\ \hat{Y}_{2,4} \end{pmatrix} = \begin{pmatrix} 159, 56 \\ 147, 26 \\ 139, 06 \\ 126, 76 \end{pmatrix}, \quad \hat{\mathbf{Y}}_3 = \begin{pmatrix} \hat{Y}_{3,1} \\ \hat{Y}_{3,2} \end{pmatrix} = \begin{pmatrix} 149, 63 \\ 137, 33 \end{pmatrix},$$

$$\hat{\mathbf{Y}}_4 = \begin{pmatrix} \hat{Y}_{4,1} \\ \hat{Y}_{4,2} \end{pmatrix} = \begin{pmatrix} 148, 01 \\ 135, 71 \end{pmatrix}, \quad \hat{\mathbf{Y}}_5 = \begin{pmatrix} \hat{Y}_{5,1} \\ \hat{Y}_{5,2} \\ \hat{Y}_{5,3} \end{pmatrix} = \begin{pmatrix} 136, 56 \\ 124, 26 \\ 116, 06 \end{pmatrix} \text{ e } \hat{\mathbf{Y}}_6 = \begin{pmatrix} \hat{Y}_{6,1} \\ \hat{Y}_{6,2} \\ \hat{Y}_{6,3} \end{pmatrix} = \begin{pmatrix} 137, 34 \\ 125, 04 \\ 116, 84 \end{pmatrix}.$$

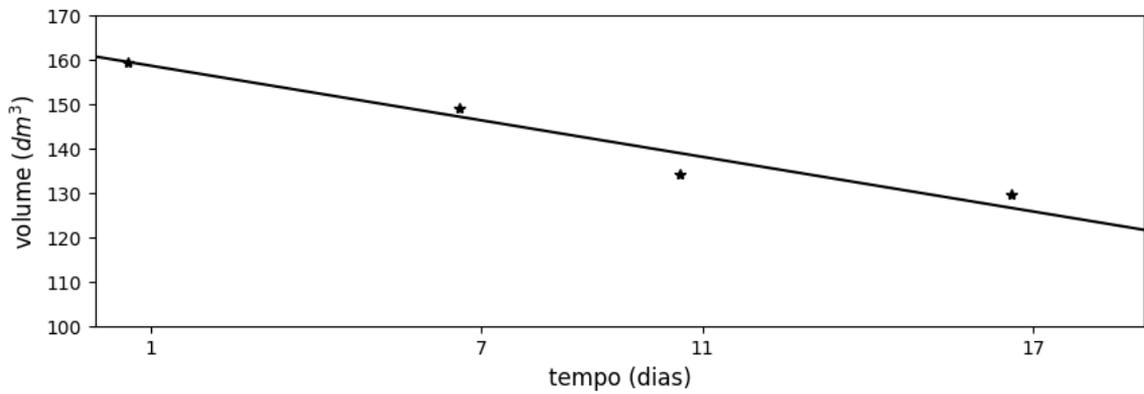
Ou na forma de função  $f_i(t) = \mathbf{g}_x(t)^T \hat{\boldsymbol{\beta}} + \mathbf{g}_z(t)^T \hat{\mathbf{b}}_i$ :  $f_1(t) = \mathbf{g}_x(t)^T \hat{\boldsymbol{\beta}} + \mathbf{g}_z(t)^T \hat{\mathbf{b}}_1$ ,  $f_2(t) = \mathbf{g}_x(t)^T \hat{\boldsymbol{\beta}} + \mathbf{g}_z(t)^T \hat{\mathbf{b}}_2$ ,  $f_3(t) = \mathbf{g}_x(t)^T \hat{\boldsymbol{\beta}} + \mathbf{g}_z(t)^T \hat{\mathbf{b}}_3$ ,  $f_4(t) = \mathbf{g}_x(t)^T \hat{\boldsymbol{\beta}} + \mathbf{g}_z(t)^T \hat{\mathbf{b}}_4$ ,  $f_5(t) = \mathbf{g}_x(t)^T \hat{\boldsymbol{\beta}} + \mathbf{g}_z(t)^T \hat{\mathbf{b}}_5$  e  $f_6(t) = \mathbf{g}_x(t)^T \hat{\boldsymbol{\beta}} + \mathbf{g}_z(t)^T \hat{\mathbf{b}}_6$ . Obtemos:

$$\begin{aligned} f_1(t) &= 150,33 - 2,05t + 10,31 = 160,64 - 2,05t, \\ f_2(t) &= 150,33 - 2,05t + 11,28 = 161,61 - 2,05t, \\ f_3(t) &= 150,33 - 2,05t + 1,35 = 151,68 - 2,05t, \\ f_4(t) &= 150,33 - 2,05t - 0,27 = 150,06 - 2,05t, \\ f_5(t) &= 150,33 - 2,05t - 11,72 = 138,61 - 2,05t \quad \text{e} \\ f_6(t) &= 150,33 - 2,05t - 10,94 = 139,39 - 2,05t. \end{aligned}$$

As Figuras 21 até 26 apresentam os gráficos dessas funções juntamente das medidas dos indivíduos. Note que nesse modelo todas as funções possuem o mesmo coeficiente angular,  $-2,05$ , porém, cada indivíduo possui o seu próprio valor para o intercepto. Esse modelo prevê que cada indivíduo chega ao litoral com o seu próprio volume corporal, porém, todos diminuem com a mesma taxa de  $-2,05$  ao dia.

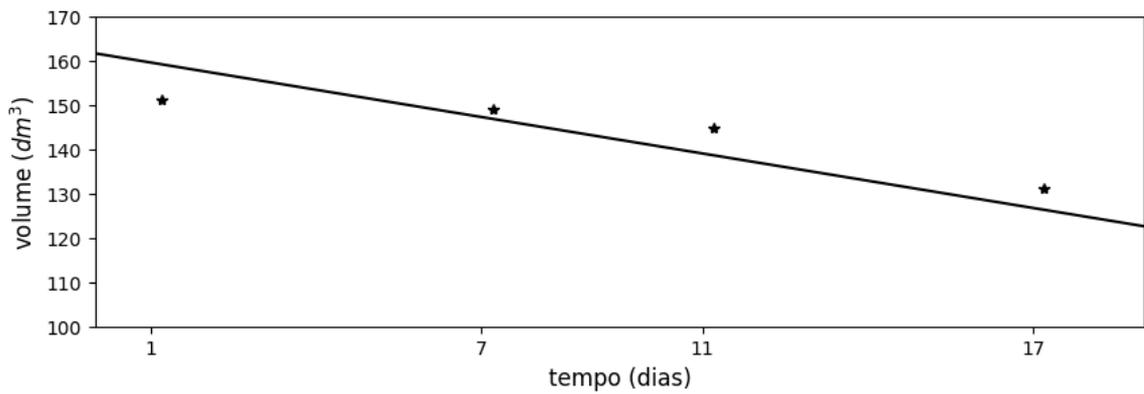
Diferente de como tratamos os dados na forma transversal, note que com a metodologia longitudinal obtemos uma função de regressão para cada indivíduo, assim as previsões, sendo individuais, respeitam as particularidades de cada um. E também mesmo a média global,  $\mu(t) = 150,33 - 2,05t$ , atende melhor às características de cada indivíduo pelo fato de levar em consideração a correlação presente entre medidas de um mesmo indivíduo.

Figura 21 – Gráfico de  $f_1(t) = 160,64 - 2,05t$ , linha cheia. Os \* apresentam as medidas do indivíduo 1.



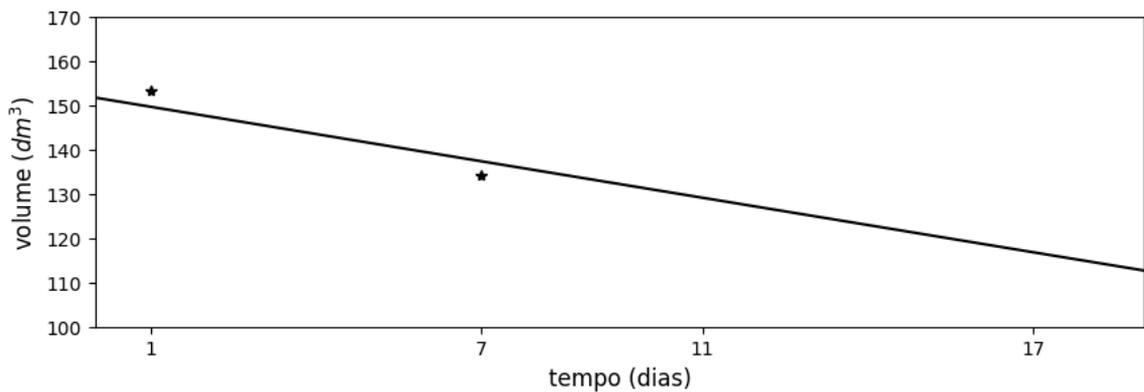
Fonte: O autor.

Figura 22 – Gráfico de  $f_2(t) = 161,61 - 2,05t$ , linha cheia. Os \* apresentam as medidas do indivíduo 2.



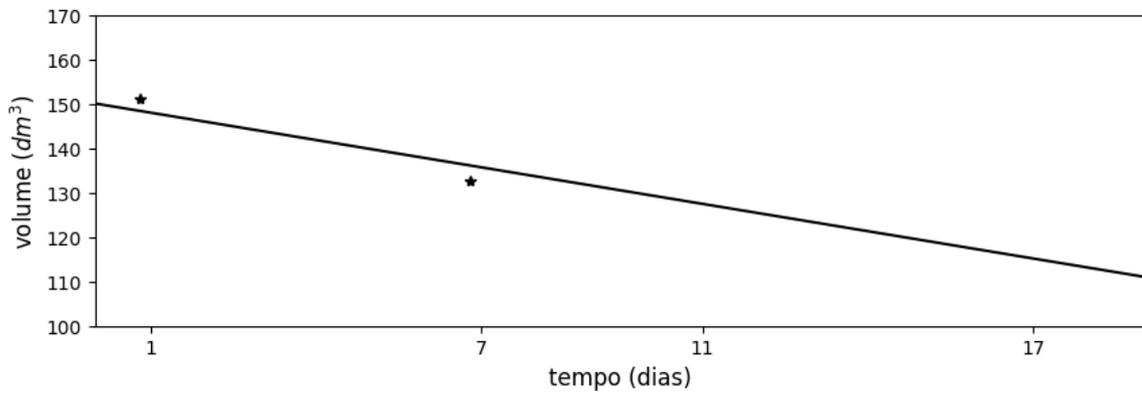
Fonte: O autor.

Figura 23 – Gráfico de  $f_3(t) = 151,68 - 2,05t$ , linha cheia. Os \* apresentam as medidas do indivíduo 3.



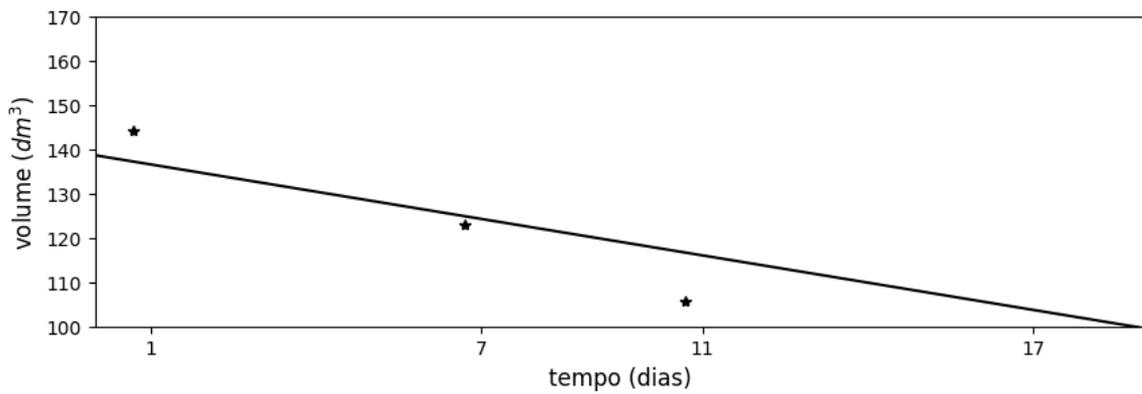
Fonte: O autor.

Figura 24 – Gráfico de  $f_4(t) = 150,06 - 2,05t$ , linha cheia. Os \* apresentam as medidas do indivíduo 4.



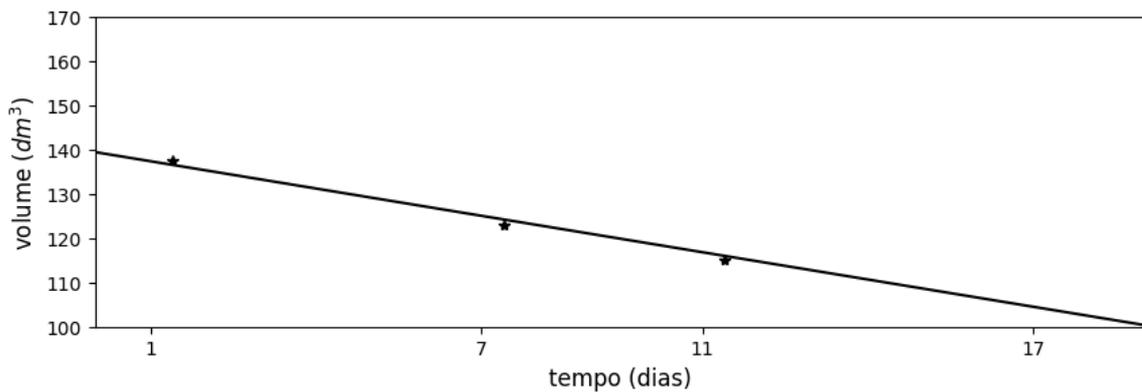
Fonte: O autor.

Figura 25 – Gráfico de  $f_5(t) = 138,61 - 2,05t$ , linha cheia. Os \* apresentam as medidas do indivíduo 5.



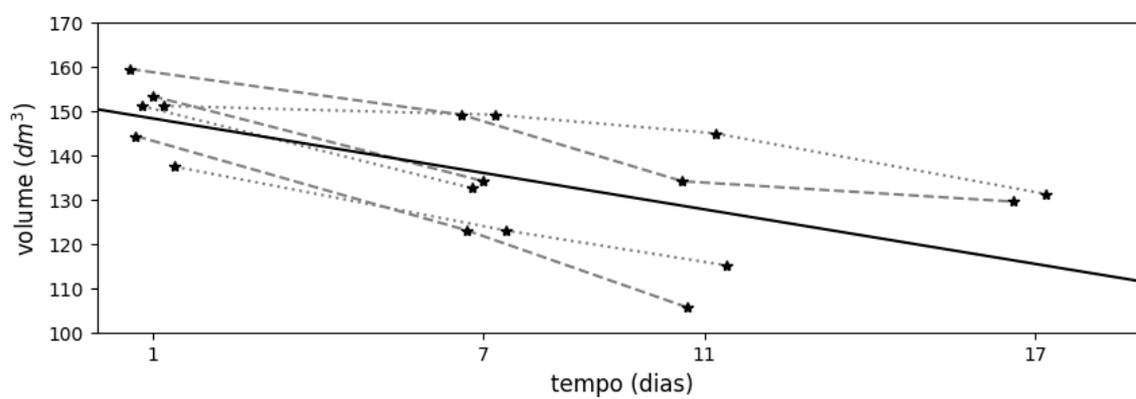
Fonte: O autor.

Figura 26 – Gráfico de  $f_6(t) = 139,39 - 2,05t$ , linha cheia. Os \* apresentam as medidas do indivíduo 6.



Fonte: O autor.

Figura 27 – Gráfico de  $\mu(t) = 150,33 - 2,05t$ , linha cheia. Os \* apresentam as medidas dos indivíduos, os quais estão ligados por linhas tracejadas às medidas que pertencem ao mesmo indivíduo. .



Fonte: O autor.

### 9.4.2 Modelagem Intervalar

Como apresentado no Capítulo 8 também podemos tratar esses dados como intervalos e com a metodologia longitudinal. Modelando as observações na forma de intervalos  $[v - d; v + d]$ , onde  $v$  é o valor do volume de uma medida e  $d$  o desvio padrão da mesma, obtemos as medidas na forma  $\overline{Y_{i,j}}$ . Organizando as medidas dos indivíduos  $\overline{\mathbf{Y}_1}$  obtemos os dados longitudinais intervalares:

$$\begin{aligned} \overline{\mathbf{Y}_1} &= \begin{pmatrix} \overline{Y_{1,1}} \\ \overline{Y_{1,2}} \\ \overline{Y_{1,3}} \\ \overline{Y_{1,4}} \end{pmatrix} = \begin{pmatrix} [154, 9; 163, 9] \\ [143, 1; 155, 1] \\ [133, 2; 135, 0] \\ [129, 0; 130, 0] \end{pmatrix}, \\ \overline{\mathbf{Y}_2} &= \begin{pmatrix} \overline{Y_{2,1}} \\ \overline{Y_{2,2}} \\ \overline{Y_{2,3}} \\ \overline{Y_{2,4}} \end{pmatrix} = \begin{pmatrix} [149, 2; 153, 0] \\ [144, 2; 154, 0] \\ [139, 8; 150, 0] \\ [131, 2; 131, 2] \end{pmatrix}, \\ \overline{\mathbf{Y}_3} &= \begin{pmatrix} \overline{Y_{3,1}} \\ \overline{Y_{3,2}} \end{pmatrix} = \begin{pmatrix} [149, 6; 157, 0] \\ [133, 2; 135, 0] \end{pmatrix} \\ \overline{\mathbf{Y}_4} &= \begin{pmatrix} \overline{Y_{4,1}} \\ \overline{Y_{4,2}} \end{pmatrix} = \begin{pmatrix} [149, 2; 153, 0] \\ [130, 0; 135, 0] \end{pmatrix}, \\ \overline{\mathbf{Y}_5} &= \begin{pmatrix} \overline{Y_{5,1}} \\ \overline{Y_{5,2}} \\ \overline{Y_{5,3}} \end{pmatrix} = \begin{pmatrix} [140, 6; 148, 0] \\ [121, 1; 124, 9] \\ [103, 1; 108, 1] \end{pmatrix} \text{ e} \\ \overline{\mathbf{Y}_6} &= \begin{pmatrix} \overline{Y_{6,1}} \\ \overline{Y_{6,2}} \\ \overline{Y_{6,3}} \end{pmatrix} = \begin{pmatrix} [132, 8; 142, 0] \\ [119, 1; 126, 9] \\ [111, 2; 119, 0] \end{pmatrix}. \end{aligned}$$

Escolhemos arbitrariamente, como na subseção anterior, as funções  $\mathbf{g}_x(t)$  e  $\mathbf{g}_z(t)$ :

$$\mathbf{g}_x(t) = \begin{pmatrix} 1 \\ t \end{pmatrix} \text{ e } \mathbf{g}_z(t) = \begin{pmatrix} 1 \end{pmatrix}.$$

Obtemos assim os mesmos  $\mathbf{X}_i$  e  $\mathbf{Z}_i$ . E usando os valores de máxima verossimilhança  $\hat{\sigma}^2 = 30,76$  e  $\hat{\mathbf{D}} = (91, 90)$  obtemos os mesmos  $\hat{\Sigma}_i$ .

Usamos então a expressão da Equação 8.3 da Seção 8.2:

$$\overline{S_i}(t) = \sum_{k=1}^n \sum_{j|\exists \overline{Y_{k,j}}} \left( \mathbf{g}_x(t)^T \mathbf{W}_j^{(k)} + \mathbf{g}_z(t)^T \left( \delta_{i,k} \mathbf{V}_j^{(i)} - \mathbf{V}^{(i)} \mathbf{X}_i \mathbf{W}_j^{(k)} \right) \right) \overline{Y_{k,j}},$$

com

$$\mathbf{W}^{(k)} = \left( \sum_{l=1}^n \mathbf{X}_l^T \hat{\Sigma}_l^{-1} \mathbf{X}_l \right)^{-1} \mathbf{X}_k^T \hat{\Sigma}_k^{-1}$$

e

$$\mathbf{V}^{(i)} = \hat{\mathbf{D}} \mathbf{Z}_i^T \hat{\Sigma}_i^{-1}.$$

Calculando os  $\mathbf{W}^{(k)}$ , obtemos:

$$\mathbf{W}^{(1)} = \mathbf{W}^{(2)} = \begin{pmatrix} 0,16 & 0,07 & 0,01 & -0,08 \\ -0,02 & 0 & 0,01 & 0,02 \end{pmatrix},$$

$$\mathbf{W}^{(3)} = \mathbf{W}^{(4)} = \begin{pmatrix} 0,13 & 0,04 \\ -0,01 & 0,01 \end{pmatrix}$$

e

$$\mathbf{W}^{(5)} = \mathbf{W}^{(6)} = \begin{pmatrix} 0,14 & 0,05 & -0,02 \\ -0,01 & 0 & 0,01 \end{pmatrix}.$$

Assim como os  $\mathbf{V}^{(i)}$ :

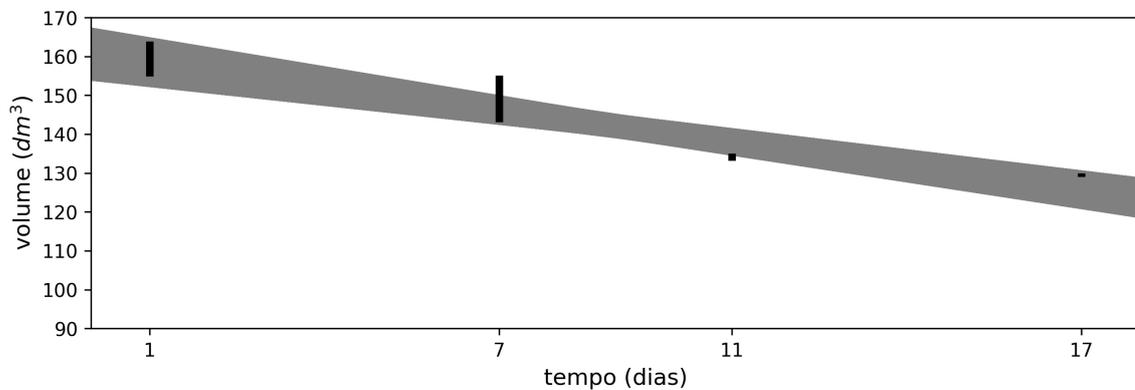
$$\mathbf{V}^{(1)} = \mathbf{V}^{(2)} = \begin{pmatrix} 0,23 & 0,23 & 0,23 & 0,23 \end{pmatrix},$$

$$\mathbf{V}^{(3)} = \mathbf{V}^{(4)} = \begin{pmatrix} 0,43 & 0,43 \end{pmatrix}$$

e

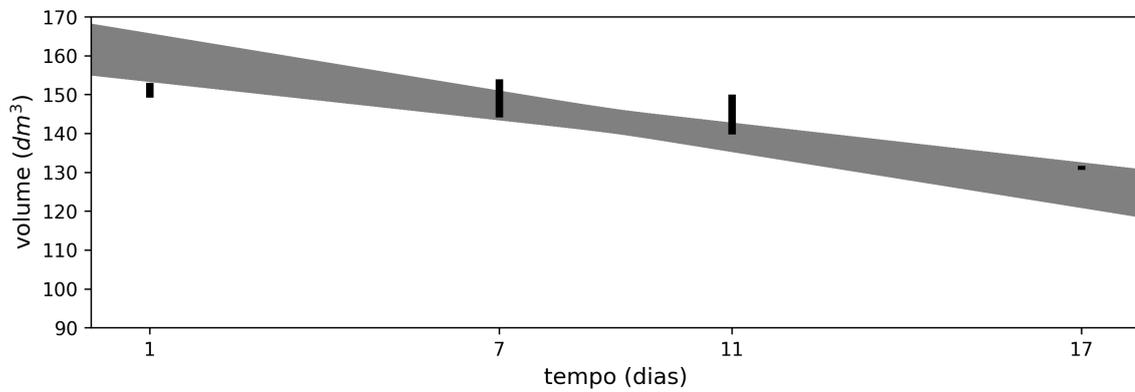
$$\mathbf{V}^{(5)} = \mathbf{V}^{(6)} = \begin{pmatrix} 0,3 & 0,3 & 0,3 \end{pmatrix}.$$

Figura 28 – Gráfico de  $\overline{S}_1(t)$ , área cinza. As linhas verticais em preto apresentam as medidas do indivíduo 1.



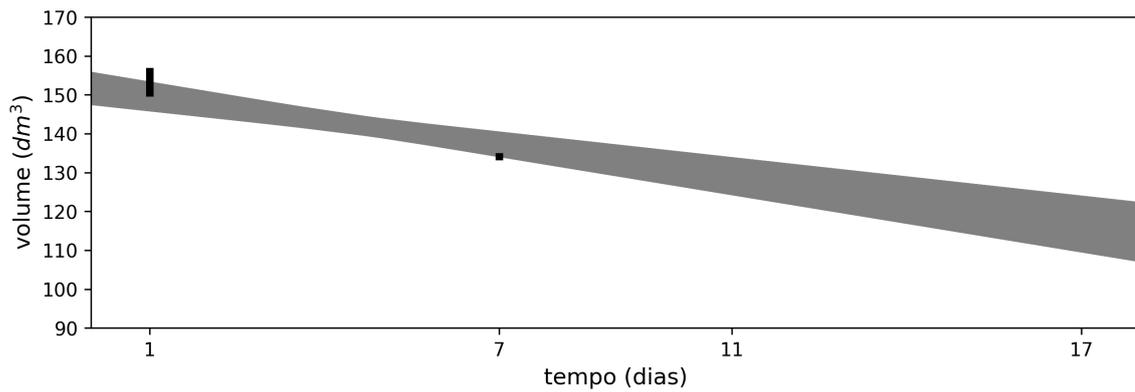
Fonte: O autor.

Figura 29 – Gráfico de  $\overline{S_2}(t)$ , área cinza. As linhas verticais em preto apresentam as medidas do indivíduo 2.



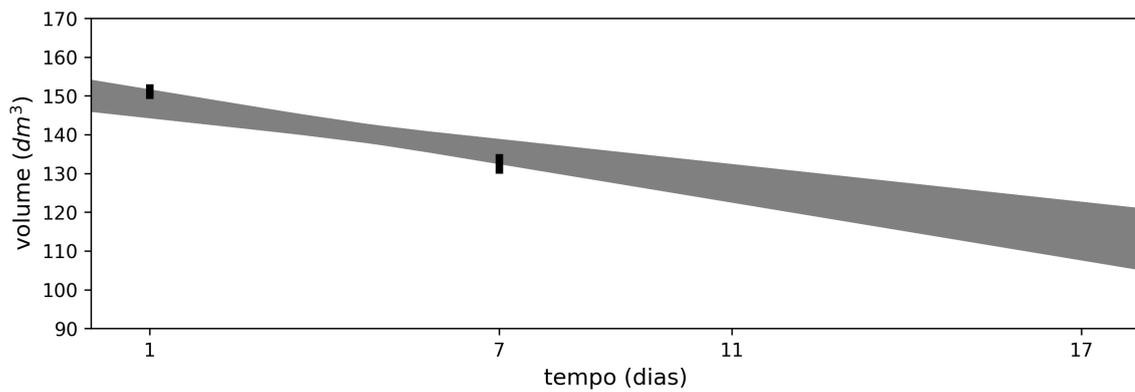
Fonte: O autor.

Figura 30 – Gráfico de  $\overline{S_3}(t)$ , área cinza. As linhas verticais em preto apresentam as medidas do indivíduo 3.



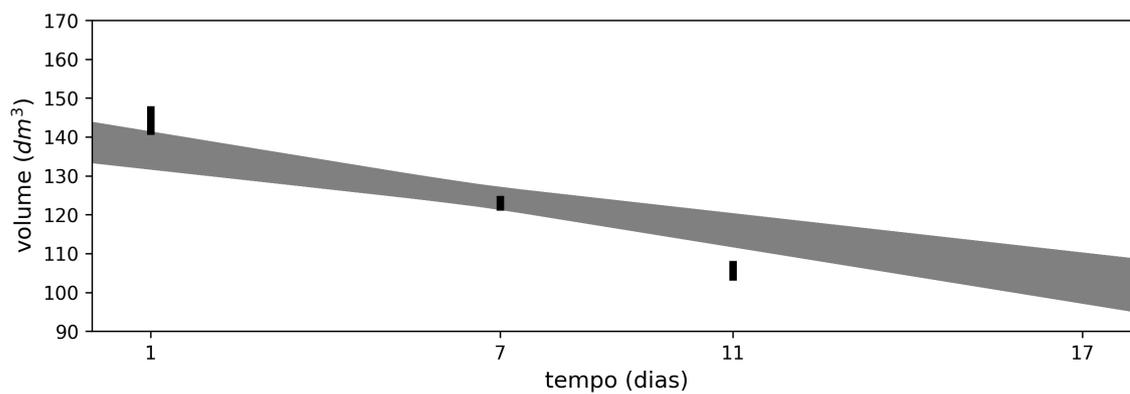
Fonte: O autor.

Figura 31 – Gráfico de  $\overline{S_4}(t)$ , área cinza. As linhas verticais em preto apresentam as medidas do indivíduo 4.



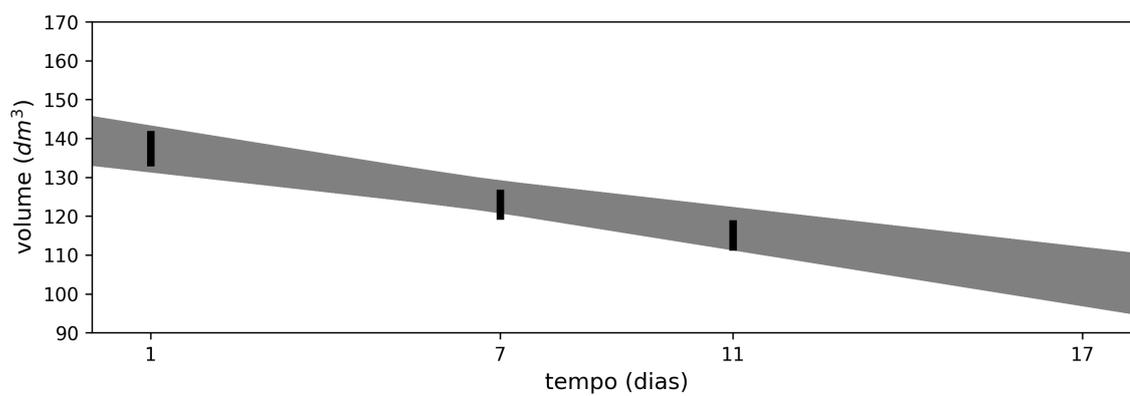
Fonte: O autor.

Figura 32 – Gráfico de  $\overline{S}_5(t)$ , área cinza. As linhas verticais em preto apresentam as medidas do indivíduo 5.



Fonte: O autor.

Figura 33 – Gráfico de  $\overline{S}_6(t)$ , área cinza. As linhas verticais em preto apresentam as medidas do indivíduo 6.



Fonte: O autor.

A Tabela 18 apresenta os valores das observações comparadas com os valores das estimativas, apresentando o erro. Por exemplo, podemos calcular o erro  $\overline{e_{1,1}}$ :

$$\begin{aligned}\overline{e_{1,1}} &= \widehat{\overline{Y_{1,1}}} \ominus_{gH} \overline{Y_{1,1}} = \\ &= [152, 416; 164, 753] \ominus_{gH} [154, 9; 163, 9] = \\ &= [\min\{152, 416 - 154, 9; 164, 753 - 163, 9\}; \max\{152, 416 - 154, 9; 164, 753 - 163, 9\}] = \\ &= [-2, 484; 0, 853].\end{aligned}$$

Uma análise interessante é a estimativa para os indivíduos que não tiveram as medidas tomadas no tempo  $t_4 = 17$ .  $\widehat{\overline{Y_{3,4}}} = [109, 705; 123, 871]$ ,  $\widehat{\overline{Y_{4,4}}} = [107, 837; 122, 483]$ ,  $\widehat{\overline{Y_{5,4}}} = [97, 374; 110, 051]$  e  $\widehat{\overline{Y_{6,4}}} = [97, 1; 111, 884]$ . Note que todas essas estimativas são ainda mais pessimistas que a apresentada na subseção anterior, Equação 9.4,  $\overline{S(17)} = [116, 599; 125, 360]$ . Por pessimistas nos referimos ao fato do volume corporal das orcas-pigmeias previsto estar em um intervalo com valores ainda menores.

Tabela 18 – Tabela comparando as observações  $\overline{Y_{i,j}}$  com as estimativas  $\widehat{Y_{i,j}}$ .

observação	estimativa ( $\widehat{Y_{i,j}} = \overline{S_i}(t_j)$ )	erro ( $\overline{e_{i,j}} = \widehat{Y_{i,j}} \ominus_{gH} \overline{Y_{i,j}}$ )
$\overline{Y_{1,1}} = [154, 9; 163, 9]$	$\widehat{Y_{1,1}} = [152, 416; 164, 753]$	$\overline{e_{1,1}} = [-2, 484; 0, 853]$
$\overline{Y_{1,2}} = [143, 1; 155, 1]$	$\widehat{Y_{1,2}} = [143, 669; 150, 805]$	$\overline{e_{1,2}} = [-4, 295; 0, 569]$
$\overline{Y_{1,3}} = [133, 2; 135, 0]$	$\widehat{Y_{1,3}} = [124, 437; 133, 772]$	$\overline{e_{1,3}} = [-8, 763; -1, 228]$
$\overline{Y_{1,4}} = [129, 0; 130, 0]$	$\widehat{Y_{1,4}} = [107, 837; 122, 483]$	$\overline{e_{1,4}} = [-21, 163; -7, 517]$
$\overline{Y_{2,1}} = [149, 2; 153, 0]$	$\widehat{Y_{2,1}} = [152, 416; 164, 753]$	$\overline{e_{2,1}} = [3, 216; 11, 753]$
$\overline{Y_{2,2}} = [144, 2; 154, 0]$	$\widehat{Y_{2,2}} = [143, 669; 150, 805]$	$\overline{e_{2,2}} = [-3, 195; -0, 531]$
$\overline{Y_{2,3}} = [139, 8; 150, 0]$	$\widehat{Y_{2,3}} = [124, 437; 133, 772]$	$\overline{e_{2,3}} = [-16, 228; -15, 363]$
$\overline{Y_{2,4}} = [131, 2; 131, 2]$	$\widehat{Y_{2,4}} = [107, 837; 122, 483]$	$\overline{e_{2,4}} = [-23, 363; -8, 717]$
$\overline{Y_{3,1}} = [149, 6; 157, 0]$	$\widehat{Y_{3,1}} = [152, 416; 164, 753]$	$\overline{e_{3,1}} = [2, 816; 7, 753]$
$\overline{Y_{3,2}} = [133, 2; 135, 0]$	$\widehat{Y_{3,2}} = [143, 669; 150, 805]$	$\overline{e_{3,2}} = [10, 469; 15, 805]$
não existe	$\widehat{Y_{3,3}} = [124, 437; 133, 772]$	não existe
não existe	$\widehat{Y_{3,4}} = [109, 705; 123, 871]$	não existe
$\overline{Y_{4,1}} = [149, 2; 153, 0]$	$\widehat{Y_{4,1}} = [152, 416; 164, 753]$	$\overline{e_{4,1}} = [3, 216; 11, 753]$
$\overline{Y_{4,2}} = [130, 0; 135, 0]$	$\widehat{Y_{4,2}} = [143, 669; 150, 805]$	$\overline{e_{4,2}} = [13, 669; 15, 805]$
não existe	$\widehat{Y_{4,3}} = [122, 743; 132, 21]$	não existe
não existe	$\widehat{Y_{4,4}} = [107, 837; 122, 483]$	não existe
$\overline{Y_{5,1}} = [140, 6; 148, 0]$	$\widehat{Y_{5,1}} = [152, 416; 164, 753]$	$\overline{e_{5,1}} = [11, 816; 16, 753]$
$\overline{Y_{5,2}} = [121, 1; 124, 9]$	$\widehat{Y_{5,2}} = [143, 669; 150, 805]$	$\overline{e_{5,2}} = [22, 569; 25, 905]$
$\overline{Y_{5,3}} = [103, 1; 108, 1]$	$\widehat{Y_{5,3}} = [124, 437; 133, 772]$	$\overline{e_{5,3}} = [21, 337; 25, 672]$
não existe	$\widehat{Y_{5,4}} = [97, 374; 110, 051]$	não existe
$\overline{Y_{6,1}} = [132, 8; 142, 0]$	$\widehat{Y_{6,1}} = [152, 416; 164, 753]$	$\overline{e_{6,1}} = [19, 616; 22, 753]$
$\overline{Y_{6,2}} = [119, 1; 126, 9]$	$\widehat{Y_{6,2}} = [143, 669; 150, 805]$	$\overline{e_{6,2}} = [23, 905; 24, 569]$
$\overline{Y_{6,3}} = [111, 2; 119, 0]$	$\widehat{Y_{6,3}} = [124, 437; 133, 772]$	$\overline{e_{6,3}} = [13, 237; 14, 772]$
não existe	$\widehat{Y_{6,4}} = [97, 1; 111, 884]$	não existe

Fonte: O autor.

## 9.5 Conclusão

Quanto à situação das orcas-pigmeias a conclusão é a mesma apresentada por Currie et al. (2021) de que os animais perderam volume corporal ao longo dos dias em que foram observadas. Pela Equação 9.2,  $\hat{f}(t) = 148,3846 - 1,612t$ , do método de regressão linear da Subseção 9.3.1.

Os modelos para dados intervalares ou *fuzzy* estimados nos dois casos, transversal e longitudinal, apresentam comportamento similar aos modelos para números reais. Consequência do fato dos modelos imprecisos apresentados serem uma extensão dos modelos clássicos. Ou seja, caso se use dados na forma de números reais para os métodos intervalares ou *fuzzy* propostos neste trabalho se obterá o mesmo resultado dos métodos clássicos.

# 10 Poluição Atmosférica na Cidade de São Paulo

## 10.1 Introdução

Conceição et al. (2001) apresenta dados sobre a poluição atmosférica na cidade de São Paulo entre os anos de 1994 e 1997. O trabalho, obviamente, não apresenta um valor para a poluição propriamente dita, mas sim a concentração de determinados poluentes no ar, sendo eles o ozônio ( $O_3$ ), as partículas sólidas no ar menores que  $10\mu m$  ( $PM_{10}$ ), o monóxido de carbono ( $CO$ ) e o dióxido de enxofre ( $SO_2$ ).

Os dados sobre as concentrações de ozônio ( $O_3$ ) são usados por Pinto et al. (2018) que modela os valores na forma de números *fuzzy* triangulares simétricos. Usando a média  $m$  e o desvio padrão  $\sigma$  anual dos dados apresentados constrói os números *fuzzy* triangulares na forma  $(m - \sigma; m; m + \sigma)$ . Cabe ressaltar que nesse caso não temos exatamente uma medida imprecisa mas sim uma modelagem da média anual na forma de número *fuzzy*.

No Capítulo 6 discutimos a classificação dos modelos para dados *fuzzy* em quatro categorias: modelos possibilísticos, modelos de quadrados mínimos, regressão quantílica e análise sensitiva. O modelo apresentado por Pinto et al. (2018) se encaixa na categoria de modelo possibilístico. Usando conceitos próprios apenas da teoria *fuzzy* para a construção do mesmo. Desta forma usamos aqui apenas a forma como o trabalho modelou os dados na forma de números *fuzzy*, deixando de lado a função  $f^* : \mathbb{R} \rightarrow \mathcal{F}(\mathbb{R})$  que é apresentada.

É possível modelar todos os valores apresentados por Conceição et al. (2001) do mesmo modo que Pinto et al. (2018) fez para os dados do ozônio. A Tabela 19 apresenta a modelagem desses dados na forma de números *fuzzy* triangulares.

O objetivo desse capítulo é explorar melhor o modelo linear simples para dados *fuzzy*,  $S^*(\vartheta) = \sum_{i=1}^n \{(w_{1,i} + \vartheta w_{2,i}) y_i^*\}$ , visto na Seção 7.3. Assim como usar o modelo linear múltiplo para dados *fuzzy*,  $S^*(\vartheta_1, \dots, \vartheta_q) = \sum_{i=1}^n \{(w_{1,i} + \vartheta_1 w_{2,i} + \dots + \vartheta_q w_{q+1,i}) y_i^*\}$ , da Seção 7.4 para realizar regressão polinomial nesses dados.

Esses dados não se encaixam na definição de dados longitudinais então não é possível construir modelos linear mistos para estes como os apresentados no Capítulo 8.

Tabela 19 – Dados de poluição na cidade de São Paulo modelados como números *fuzzy* triangulares.

	1994 ( $x_1 = 1$ )	1995 ( $x_2 = 2$ )	1996 ( $x_3 = 3$ )	1997 ( $x_4 = 4$ )
$O_3$	$o_1^* =$ (17, 6; 57; 96, 4)	$o_2^* =$ (25, 3; 60, 7; 96, 1)	$o_3^* =$ (34, 8; 76, 3; 117, 8)	$o_4^* =$ (29, 5; 63; 96, 5)
$PM_{10}$	$p_1^* =$ (30, 6; 67, 0; 103, 4)	$p_2^* =$ (43, 0; 73, 8; 104, 6)	$p_3^* =$ (33, 7; 63, 8; 93, 9)	$p_4^* =$ (35, 1; 60, 3; 85, 5)
$CO$	$c_1^* =$ (2, 7; 5, 1; 7, 5)	$c_2^* =$ (2, 7; 5, 1; 7, 5)	$c_3^* =$ (1, 9; 3, 9; 5, 9)	$c_4^* =$ (2, 1; 3, 7; 5, 3)
$SO_2$	$s_1^* =$ (7, 5; 14, 9; 22, 3)	$s_2^* =$ (14, 7; 27, 5; 40, 3)	$s_3^* =$ (11, 8; 23; 34, 2)	$s_4^* =$ (9, 3; 19, 6; 29, 9)

Fonte: (PINTO et al., 2018) e Conceição et al. (2001) (adaptado).

## 10.2 Modelos de Regressão

Para cada um dos conjuntos de dados da Tabela 19 apresentaremos um modelo simples e dois modelos polinomiais usando o modelo de regressão múltiplo.

Os dados foram observados em quatro diferentes anos. Adotamos  $x_1 = 1$ ,  $x_2 = 2$ ,  $x_3 = 3$ ,  $x_4 = 4$  para representar os anos 1994, 1995, 1996 e 1997, respectivamente. Dessa forma para o modelo de regressão *fuzzy* simples utilizamos a matriz de planejamento:

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix},$$

com a qual construímos a matriz:

$$\mathbf{W} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T = \begin{pmatrix} 1 & 0,5 & 0 & -0,5 \\ -0,3 & -0,1 & 0,1 & 0,3 \end{pmatrix}.$$

Usando essa matriz  $\mathbf{W}$  podemos construir modelos para cada um dos conjuntos de dados usando a Equação 7.15  $S^*(\vartheta) = (w_{1,1} + \vartheta w_{2,1}) y_1^* + \dots + (w_{1,n} + \vartheta w_{2,n}) y_n^*$ .

Para os dados  $o_1^*$ ,  $o_2^*$ ,  $o_3^*$  e  $o_4^*$  referentes ao Ozônio na Tabela 19, obtemos o modelo:

$$\mathcal{O}_1^*(x) = (1 - 0,3x)o_1^* + (0,5 - 0,1x)o_2^* + (0 + 0,1x)o_3^* + (-0,5 + 0,3x)o_4^*.$$

Do mesmo modo obtemos modelos para os demais poluentes.

Partículas  $PM_{10}$ :

$$\mathcal{P}_1^*(x) = (1 - 0,3x)p_1^* + (0,5 - 0,1x)p_2^* + (0 + 0,1x)p_3^* + (-0,5 + 0,3x)p_4^*.$$

Monóxido de carbono ( $CO$ ):

$$\mathcal{C}_1^*(x) = (1 - 0,3x)c_1^* + (0,5 - 0,1x)c_2^* + (0 + 0,1x)c_3^* + (-0,5 + 0,3x)c_4^*.$$

Dióxido de enxofre ( $SO_2$ ):

$$\mathcal{S}_1^*(x) = (1 - 0,3x)s_1^* + (0,5 - 0,1x)s_2^* + (0 + 0,1x)s_3^* + (-0,5 + 0,3x)s_4^*.$$

O gráfico de  $\mathcal{O}_1^*(x)$  é apresentado na Figura 37(a). O gráfico de  $\mathcal{P}_1^*(x)$  é apresentado na Figura 38(a). O gráfico de  $\mathcal{C}_1^*(x)$  é apresentado na Figura 39(a). O gráfico de  $\mathcal{S}_1^*(x)$  é apresentado na Figura 40(a).

A Figura 34 apresenta os gráficos com os pesos desses modelos. No que se refere aos pesos dados às observações os quatro modelos são idênticos, desta forma podemos avaliar apenas  $\mathcal{O}_1^*(x)$ . O peso dado à  $o^*1$  é dado pela reta  $1 - 0,3x$ , uma reta decrescente. Assim este peso decresce ao longo de  $x$ , por exemplo, ao se calcular  $\mathcal{O}_1^*(1)$  o peso dado a  $o^*1$  é 0,7, para  $\mathcal{O}_1^*(2)$  o peso é 0,4, para  $\mathcal{O}_1^*(3)$  é 0,1 e para  $\mathcal{O}_1^*(4)$  é  $-0,2$ . Do mesmo modo o peso dado para  $o_2^*$  também é uma reta decrescente,  $0,5 - 0,1x$ , enquanto os pesos de  $o_3^*$  e  $o_4^*$  é dado por retas crescentes,  $0 + 0,1x$  e  $-0,5 + 0,3x$ , respectivamente.

Deste modo os dados tendem a possuir maior influência para  $x$  próximo do valor de  $x$  no qual ele foi tomado. Deste modo  $o_1^*$  possui maior influência em  $x = 1$  e  $o_4^*$  possui maior influência em  $x = 4$ . O mesmo não acontece com  $o_2^*$ , em relação a  $x = 2$  e  $o_3^*$ , para  $x = 3$ , com o mesmo impacto devido à limitação que uma linha reta possui de não poder ser maior em  $x = 2$  do que em  $x = 1$  e  $x = 3$  ao mesmo tempo.

Tabela 20 – Pesos do modelo  $\mathcal{O}_1^*(x)$  para alguns valores de  $x$ .

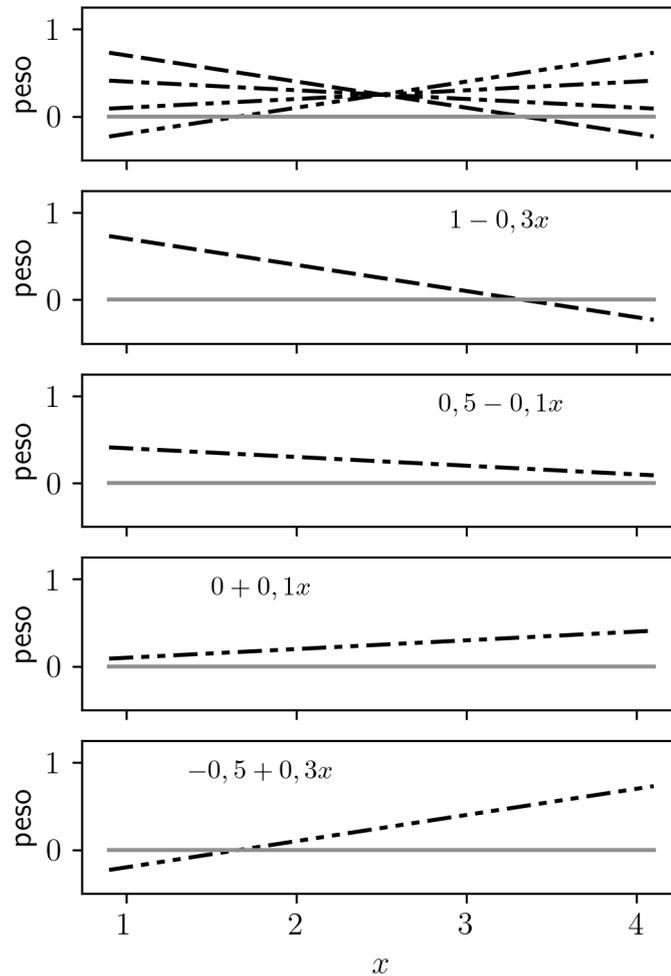
	$x = 1$	$x = 2$	$x = 3$	$x = 4$
$1 - 0,3x$	0,7	0,4	0,1	$-0,2$
$0,5 - 0,1x$	0,4	0,3	0,2	0,1
$0 + 0,1x$	0,1	0,2	0,3	0,4
$-0,5 + 0,3x$	$-0,2$	0,1	0,4	0,7

Fonte: O autor.

Usando a regressão múltipla para dados *fuzzy* da Equação 7.25 da Seção 7.4

$$f^*(\vartheta_1, \dots, \vartheta_q) = (w_{1,1} + \vartheta_1 w_{2,1} + \dots + \vartheta_q w_{q+1,1}) y_1^* + \dots + (w_{1,n} + \vartheta_1 w_{2,n} + \dots + \vartheta_q w_{q+1,n}) y_n^*.$$

Figura 34 – Pesos dos modelos  $\mathcal{O}_1^*(x)$ ,  $\mathcal{P}_1^*(x)$ ,  $\mathcal{C}_1^*(x)$  e  $\mathcal{S}_1^*(x)$ . O primeiro gráfico apresenta os pesos  $1 - 0,3x$ ,  $0,5 - 0,1x$ ,  $0 + 0,1x$  e  $-0,5 + 0,3x$  num mesmo gráfico e os demais apresentam cada uma dessas funções em separado.



Fonte: O autor.

usando  $\vartheta_1 = x$  e  $\vartheta_2 = x^2$  montamos a matriz de planejamento

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ 1 & x_4 & x_4^2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{pmatrix},$$

com a qual construímos a matriz:

$$\mathbf{W} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T = \begin{pmatrix} 2,25 & -0,75 & -1,25 & 0,75 \\ -1,55 & 1,15 & 1,35 & -0,95 \\ 0,25 & -0,25 & -0,25 & 0,25 \end{pmatrix}.$$

Para os dados da Tabela 19, obtemos os modelos.

Ozônio  $O_3$ :

$$\mathcal{O}_2^*(x) = (2, 25 - 1, 55x + 0, 25x^2)o_1^* + (-0, 75 + 1, 15x - 0, 25x^2)o_2^* + (-1, 25 + 1, 35x - 0, 25x^2)o_3^* + (0, 75 - 0, 95x + 0, 25x^2)o_4^*.$$

Partículas  $PM_{10}$ :

$$\mathcal{P}_2^*(x) = (2, 25 - 1, 55x + 0, 25x^2)p_1^* + (-0, 75 + 1, 15x - 0, 25x^2)p_2^* + (-1, 25 + 1, 35x - 0, 25x^2)p_3^* + (0, 75 - 0, 95x + 0, 25x^2)p_4^*.$$

Monóxido de carbono ( $CO$ ):

$$\mathcal{C}_2^*(x) = (2, 25 - 1, 55x + 0, 25x^2)c_1^* + (-0, 75 + 1, 15x - 0, 25x^2)c_2^* + (-1, 25 + 1, 35x - 0, 25x^2)c_3^* + (0, 75 - 0, 95x + 0, 25x^2)c_4^*.$$

Dióxido de enxofre ( $SO_2$ ):

$$\mathcal{S}_2^*(x) = (2, 25 - 1, 55x + 0, 25x^2)s_1^* + (-0, 75 + 1, 15x - 0, 25x^2)s_2^* + (-1, 25 + 1, 35x - 0, 25x^2)s_3^* + (0, 75 - 0, 95x + 0, 25x^2)s_4^*.$$

O gráfico de  $\mathcal{O}_2^*(x)$  é apresentado na Figura 37(b). O gráfico de  $\mathcal{P}_2^*(x)$  é apresentado na Figura 38(b). O gráfico de  $\mathcal{C}_2^*(x)$  é apresentado na Figura 39(b). O gráfico de  $\mathcal{S}_2^*(x)$  é apresentado na Figura 40(b).

A Figura 35 apresenta os gráficos com os pesos desses modelos. No que se refere aos pesos dados às observações os quatro modelos são idênticos, desta forma podemos avaliar apenas  $\mathcal{O}_2^*(x)$ . Os pesos nesse modelo estão na forma de parábolas.

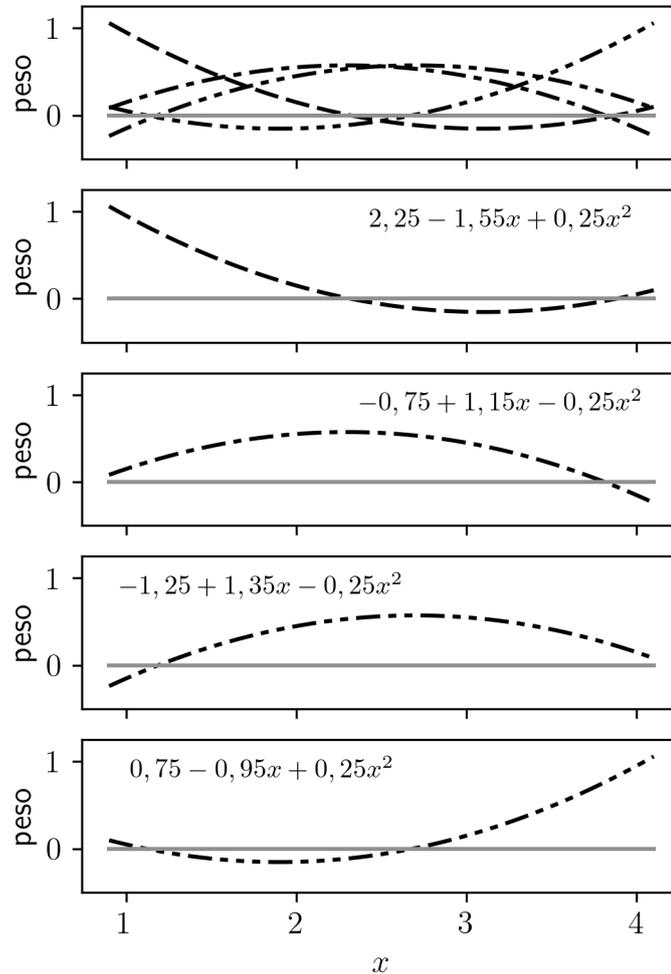
O peso dado à  $o_1^*$  é dado pela parábola  $2, 25 - 1, 55x + 0, 25x^2$ . Esse dado foi tomado para  $x = 1$ , onde para  $\mathcal{O}_2^*(1)$  o peso de  $o_1^*$  é de 0,95. Note, na Figura 35 e na Tabela 21, que ao menos na região  $1 \leq x \leq 4$  o peso de  $o_i^*$  é maior para valores de  $x$  próximos de  $x_i$ , porém, o peso dos demais dados continua tendo influência no valor de  $\mathcal{O}_2^*(x_i)$ .

Assim o modelo  $\mathcal{O}_2^*(x)$  aparenta representar melhor os dados observados do que o modelo  $\mathcal{O}_1^*(x)$ . Nas Figuras 41(a) e 41(b) estão presentes os gráficos quantil-quantil de  $\mathcal{O}_1^*(x)$  e  $\mathcal{O}_2^*(x)$ , respectivamente. A maior proximidade dos intervalos à reta  $y = x$  no gráfico quantil-quantil de  $\mathcal{O}_2^*(x)$  em relação ao de  $\mathcal{O}_1^*(x)$  também indica melhor qualidade do modelo dado por  $\mathcal{O}_2^*(x)$ .

Apresentamos também uma terceira forma usando  $\vartheta_1 = x$ ,  $\vartheta_2 = x^2$  e  $\vartheta_3 = x^3$  montamos a matriz de planejamento

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ 1 & x_3 & x_3^2 & x_3^3 \\ 1 & x_4 & x_4^2 & x_4^3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ 1 & 4 & 16 & 64 \end{pmatrix},$$

Figura 35 – Pesos dos modelos  $\mathcal{O}_2^*(x)$ ,  $\mathcal{P}_2^*(x)$ ,  $\mathcal{C}_2^*(x)$  e  $\mathcal{S}_2^*(x)$ . O primeiro gráfico apresenta os pesos  $2, 25 - 1, 55x + 0, 25x^2$ ,  $-0, 75 + 1, 15x - 0, 25x^2$ ,  $-1, 25 + 1, 35x - 0, 25x^2$  e  $0, 75 - 0, 95x + 0, 25x^2$  num mesmo gráfico e os demais apresentam cada uma dessas funções em separado.



Fonte: O autor.

com a qual construímos a matriz:

$$\mathbf{W} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T = \begin{pmatrix} 4 & -6 & 4 & -1 \\ -\frac{13}{3} & \frac{19}{2} & -7 & \frac{11}{6} \\ \frac{3}{2} & -4 & \frac{7}{2} & -1 \\ -\frac{1}{6} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{6} \end{pmatrix}.$$

Para os dados da Tabela 19, obtemos os modelos.

Tabela 21 – Pesos do modelo  $\mathcal{O}_2^*(x)$  para alguns valores de  $x$ .

	$x = 1$	$x = 2$	$x = 3$	$x = 4$
$2,25 - 1,55x + 0,25x^2$	0,95	0,15	-0,15	0,05
$-0,75 + 1,15x - 0,25x^2$	0,15	0,55	0,45	-0,15
$-1,25 + 1,35x - 0,25x^2$	-0,15	0,45	0,55	0,15
$0,75 - 0,95x + 0,25x^2$	0,05	-0,15	0,15	0,95

Fonte: O autor.

Ozônio  $O_3$ :

$$\begin{aligned} \mathcal{O}_3^*(x) = & \left(4 - \frac{13}{3}x + \frac{3}{2}x^2 - \frac{1}{6}x^3\right) o_1^* + \left(-6 + \frac{19}{2}x - 4x^2 + \frac{1}{2}x^3\right) o_2^* + \\ & + \left(4 - 7x + \frac{7}{2}x^2 - \frac{1}{2}x^3\right) o_3^* + \left(-1 + \frac{11}{6}x - x^2 + \frac{1}{6}x^3\right) o_4^*. \end{aligned}$$

Partículas  $PM_{10}$ :

$$\begin{aligned} \mathcal{P}_3^*(x) = & \left(4 - \frac{13}{3}x + \frac{3}{2}x^2 - \frac{1}{6}x^3\right) p_1^* + \left(-6 + \frac{19}{2}x - 4x^2 + \frac{1}{2}x^3\right) p_2^* + \\ & + \left(4 - 7x + \frac{7}{2}x^2 - \frac{1}{2}x^3\right) p_3^* + \left(-1 + \frac{11}{6}x - x^2 + \frac{1}{6}x^3\right) p_4^*. \end{aligned}$$

Monóxido de carbono ( $CO$ ):

$$\begin{aligned} \mathcal{C}_3^*(x) = & \left(4 - \frac{13}{3}x + \frac{3}{2}x^2 - \frac{1}{6}x^3\right) c_1^* + \left(-6 + \frac{19}{2}x - 4x^2 + \frac{1}{2}x^3\right) c_2^* + \\ & + \left(4 - 7x + \frac{7}{2}x^2 - \frac{1}{2}x^3\right) c_3^* + \left(-1 + \frac{11}{6}x - x^2 + \frac{1}{6}x^3\right) c_4^*. \end{aligned}$$

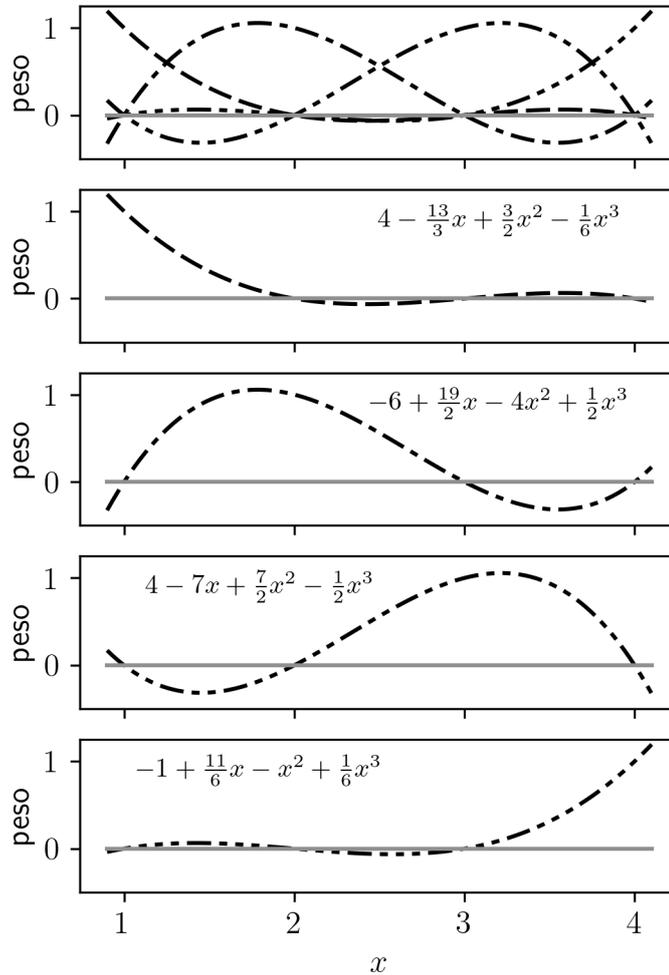
Dióxido de enxofre ( $SO_2$ ):

$$\begin{aligned} \mathcal{S}_3^*(x) = & \left(4 - \frac{13}{3}x + \frac{3}{2}x^2 - \frac{1}{6}x^3\right) s_1^* + \left(-6 + \frac{19}{2}x - 4x^2 + \frac{1}{2}x^3\right) s_2^* + \\ & + \left(4 - 7x + \frac{7}{2}x^2 - \frac{1}{2}x^3\right) s_3^* + \left(-1 + \frac{11}{6}x - x^2 + \frac{1}{6}x^3\right) s_4^*. \end{aligned}$$

O gráfico de  $\mathcal{O}_3^*(x)$  é apresentado na Figura 37(c). O gráfico de  $\mathcal{P}_3^*(x)$  é apresentado na Figura 38(c). O gráfico de  $\mathcal{C}_3^*(x)$  é apresentado na Figura 39(c). O gráfico de  $\mathcal{S}_3^*(x)$  é apresentado na Figura 40(c).

A Tabela 23 apresenta os valores de  $\overline{R^2}$  segundo o definido na Subseção 6.5. Os modelos  $\mathcal{O}_3^*(x)$ ,  $\mathcal{P}_3^*(x)$ ,  $\mathcal{C}_3^*(x)$  e  $\mathcal{S}_3^*(x)$  apresentam  $\overline{R^2} = 1$ , pois eles tem estimativas

Figura 36 – Pesos dos modelos  $\mathcal{O}_3^*(x)$ ,  $\mathcal{P}_3^*(x)$ ,  $\mathcal{C}_3^*(x)$  e  $\mathcal{S}_3^*(x)$ . O primeiro gráfico apresenta os pesos  $4 - \frac{13}{3}x + \frac{3}{2}x^2 - \frac{1}{6}x^3$ ,  $-6 + \frac{19}{2}x - 4x^2 + \frac{1}{2}x^3$ ,  $4 - 7x + \frac{7}{2}x^2 - \frac{1}{2}x^3$  e  $-1 + \frac{11}{6}x - x^2 + \frac{1}{6}x^3$  num mesmo gráfico e os demais apresentam cada uma dessas funções em separado.



Fonte: O autor.

exatamente iguais às observações, porém, longe de ser um sinal de ser um bom modelo aqui se trata de um caso de modelo sobre-ajustado, que é um modelo com mais parâmetros do que deveria (VIERTL, 2011). Isso é evidenciado na Tabela 22 onde vemos que para cada  $x_i$  o peso dado para  $o_i^*$  é exatamente 1 enquanto para os demais dados o peso dado é zero.

A Figura 41 apresenta os gráficos quantil-quantil dos erros associados aos modelos  $\mathcal{O}_1^*(x)$ ,  $\mathcal{O}_2^*(x)$ ,  $\mathcal{P}_1^*(x)$ ,  $\mathcal{P}_2^*(x)$ ,  $\mathcal{C}_1^*(x)$ ,  $\mathcal{C}_2^*(x)$ ,  $\mathcal{S}_1^*(x)$  e  $\mathcal{S}_2^*(x)$ . Comparando os gráficos quantil-quantil de  $\mathcal{O}_1^*(x)$  e  $\mathcal{O}_2^*(x)$ , Figura 41 (a) e (b), é notável que o segundo possui os pontos centrais dos intervalos mais próximos da reta  $y = x$ . O comprimento dos intervalos que representam os erros do segundo são mais uniformes, mais parecidos uns com os outros, podemos supor que essa característica é indício de melhor qualidade do modelo,

Tabela 22 – Pesos do modelo  $\mathcal{O}_3^*(x)$  para alguns valores de  $x$ .

	$x = 1$	$x = 2$	$x = 3$	$x = 4$
$4 - \frac{13}{3}x + \frac{3}{2}x^2 - \frac{1}{6}x^3$	1	0	0	0
$-6 + \frac{19}{2}x - 4x^2 + \frac{1}{2}x^3$	0	1	0	0
$4 - 7x + \frac{7}{2}x^2 - \frac{1}{2}x^3$	0	0	1	0
$-1 + \frac{11}{6}x - x^2 + \frac{1}{6}x^3$	0	0	0	1

Fonte: O autor.

porém, não temos base teórica para tal afirmação. Comparado os gráficos quantil-quantil de  $\mathcal{P}_1^*(x)$  e  $\mathcal{P}_2^*(x)$ , Figura 41 (c) e (d), de  $\mathcal{C}_1^*(x)$  e  $\mathcal{C}_2^*(x)$ , Figura 41 (e) e (f), e de  $\mathcal{S}_1^*(x)$  e  $\mathcal{S}_2^*(x)$ , Figura 41 (e) e (f), chegamos às mesmas conclusões. Concluimos que os erros dos modelos  $\mathcal{O}_1^*(x)$ ,  $\mathcal{P}_1^*(x)$ ,  $\mathcal{C}_1^*(x)$  e  $\mathcal{S}_1^*(x)$  aparentam seguir melhor uma distribuição normal com média em zero do que os modelos  $\mathcal{O}_2^*(x)$ ,  $\mathcal{P}_2^*(x)$ ,  $\mathcal{C}_2^*(x)$  e  $\mathcal{S}_2^*(x)$ .

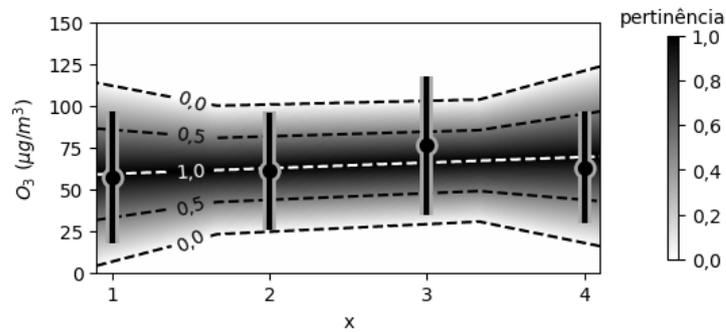
Para os modelos  $\mathcal{O}_3^*(x)$ ,  $\mathcal{P}_3^*(x)$ ,  $\mathcal{C}_3^*(x)$  e  $\mathcal{S}_3^*(x)$  tem todos os erros iguais a zero, devido ao problema de sobre-ajuste. Para esses modelos não é interessante realizar o testes sobre a normalidade dos erros, pois é imediato que não se trata de um conjunto normalmente distribuído.

Tabela 23 – Avaliação dos modelos segundo os métodos da Subseção 6.5. O coeficiente de determinação  $\overline{R}^2$ , Equação 6.13, indica a proporção da variação dos dados que é explicada pelo modelo. Valores mais próximos de 1 indicam melhor qualidade dos modelos.

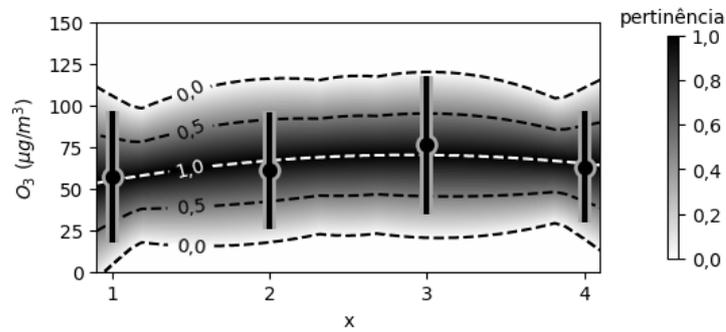
modelo	$\overline{R}^2$
$\mathcal{O}_1^*(x)$	$[-10, 35; 0, 92]$
$\mathcal{O}_2^*(x)$	$[-9, 23; 1, 0]$
$\mathcal{O}_3^*(x)$	$[1, 0; 1, 0]$
$\mathcal{P}_1^*(x)$	$[-7, 54; 0, 91]$
$\mathcal{P}_2^*(x)$	$[-8, 17; 1, 0]$
$\mathcal{P}_3^*(x)$	$[1, 0; 1, 0]$
$\mathcal{C}_1^*(x)$	$[-3, 8; 0, 97]$
$\mathcal{C}_2^*(x)$	$[-4, 85; 1, 0]$
$\mathcal{C}_3^*(x)$	$[1, 0; 1, 0]$
$\mathcal{S}_1^*(x)$	$[-8, 05; 0, 9]$
$\mathcal{S}_2^*(x)$	$[-2, 4; 1, 0]$
$\mathcal{S}_3^*(x)$	$[1, 0; 1, 0]$

Fonte: O autor.

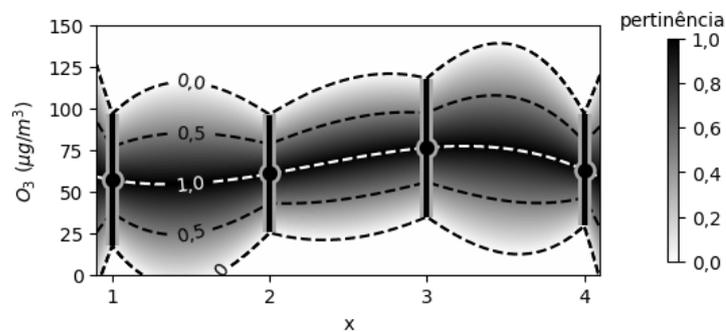
Figura 37 – Gráficos dos modelos  $\mathcal{O}_1^*(x)$ , (a),  $\mathcal{O}_2^*(x)$ , (b), e  $\mathcal{O}_3^*(x)$ , (c), na forma de escala de cinza e curvas de nível. Os dados observados  $o_i^*$  estão apresentados na forma de segmentos de reta verticais com um pequeno círculo no seu centro. Essas retas verticais são formadas pelos pontos  $(x_i, y)$  tais que a pertinência de  $y$  em  $o_i^*$  seja maior que zero, ou seja,  $\varphi_{o_i^*}(y) > 0$ . O pequeno círculo apresenta o “topo” do número fuzzy  $o_i^*$ .



(a)



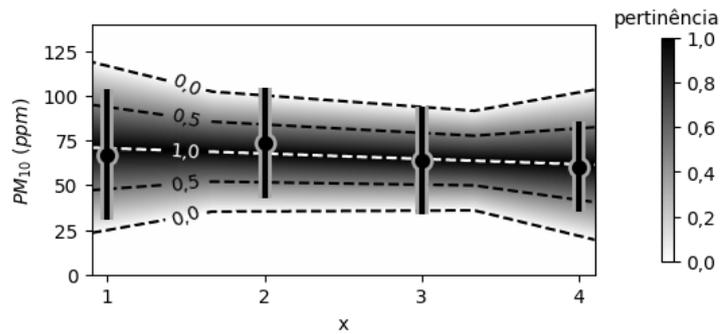
(b)



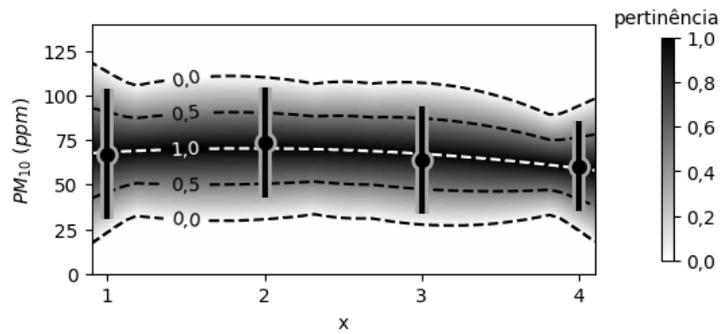
(c)

Fonte: O autor.

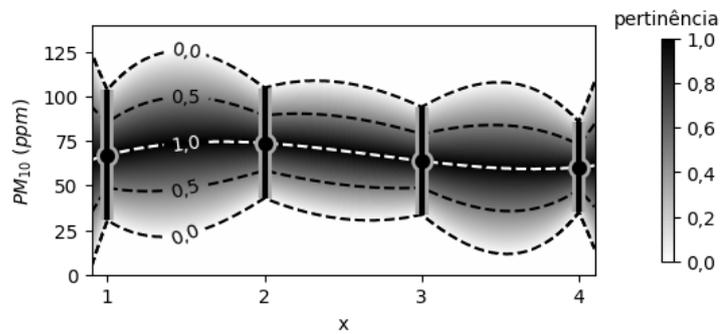
Figura 38 – Gráficos dos modelos  $\mathcal{P}_1^*(x)$ , (a),  $\mathcal{P}_2^*(x)$ , (b), e  $\mathcal{P}_3^*(x)$ , (c), na forma de escala de cinza e curvas de nível. Os dados observados  $p_i^*$  estão apresentados na forma de segmentos de reta verticais com um pequeno círculo no seu centro. Essas retas verticais são formadas pelos pontos  $(x_i, y)$  tais que a pertinência de  $y$  em  $p_i^*$  seja maior que zero, ou seja,  $\varphi_{p_i^*}(y) > 0$ . O pequeno círculo apresenta o “topo” do número fuzzy  $p_i^*$ .



(a)



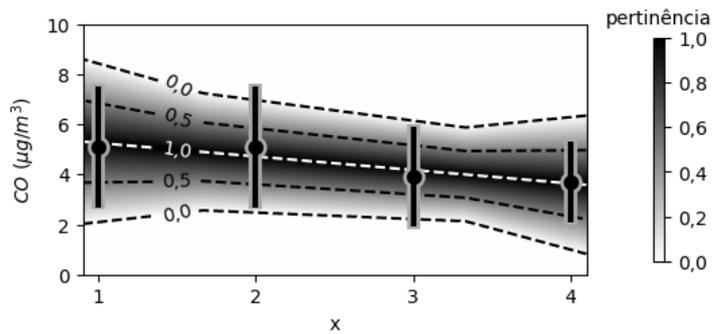
(b)



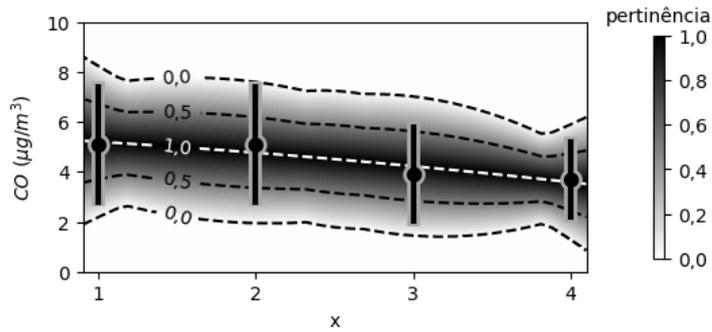
(c)

Fonte: O autor.

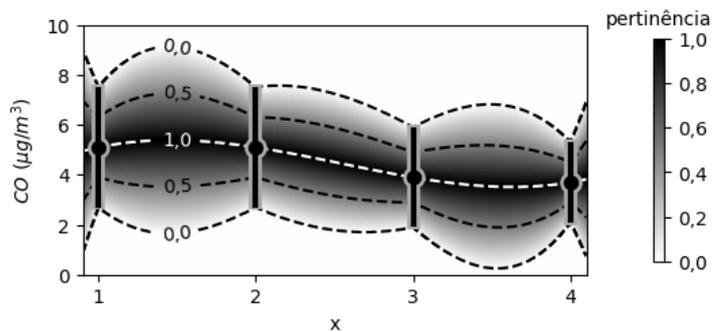
Figura 39 – Gráficos dos modelos  $\mathcal{C}_1^*(x)$ , (a),  $\mathcal{C}_2^*(x)$ , (b), e  $\mathcal{C}_3^*(x)$ , (c), na forma de escala de cinza e curvas de nível. Os dados observados  $c_i^*$  estão apresentados na forma de segmentos de reta verticais com um pequeno círculo no seu centro. Essas retas verticais são formadas pelos pontos  $(x_i, y)$  tais que a pertinência de  $y$  em  $c_i^*$  seja maior que zero, ou seja,  $\varphi_{c_i^*}(y) > 0$ . O pequeno círculo apresenta o “topo” do número fuzzy  $c_i^*$ .



(a)



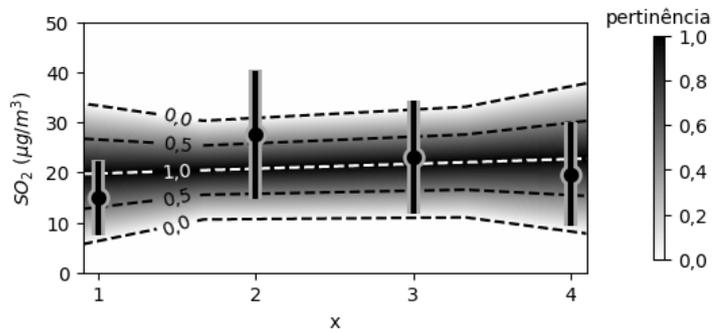
(b)



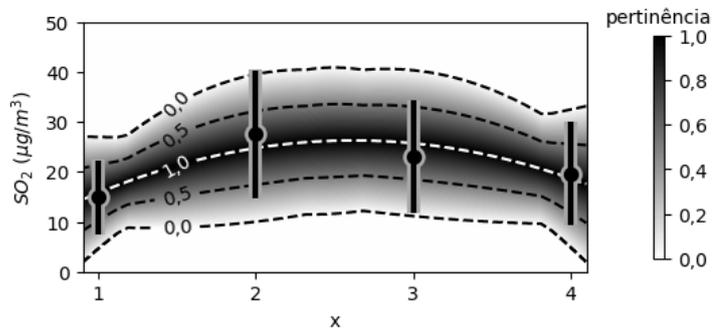
(c)

Fonte: O autor.

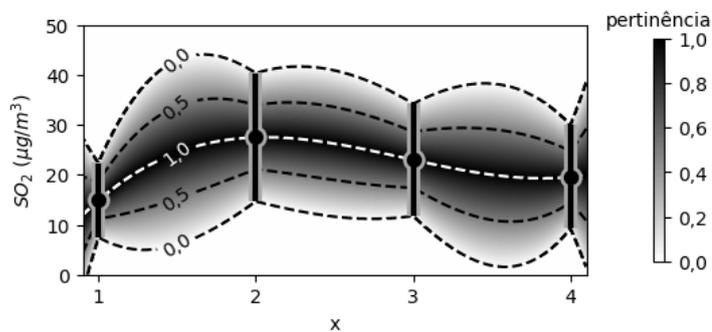
Figura 40 – Gráficos dos modelos  $\mathcal{S}_1^*(x)$ , (a),  $\mathcal{S}_2^*(x)$ , (b), e  $\mathcal{S}_3^*(x)$ , (c), na forma de escala de cinza e curvas de nível. Os dados observados  $s_i^*$  estão apresentados na forma de segmentos de reta verticais com um pequeno círculo no seu centro. Essas retas verticais são formadas pelos pontos  $(x_i, y)$  tais que a pertinência de  $y$  em  $s_i^*$  seja maior que zero, ou seja,  $\varphi_{s_i^*}(y) > 0$ . O pequeno círculo apresenta o “topo” do número fuzzy  $s_i^*$ .



(a)



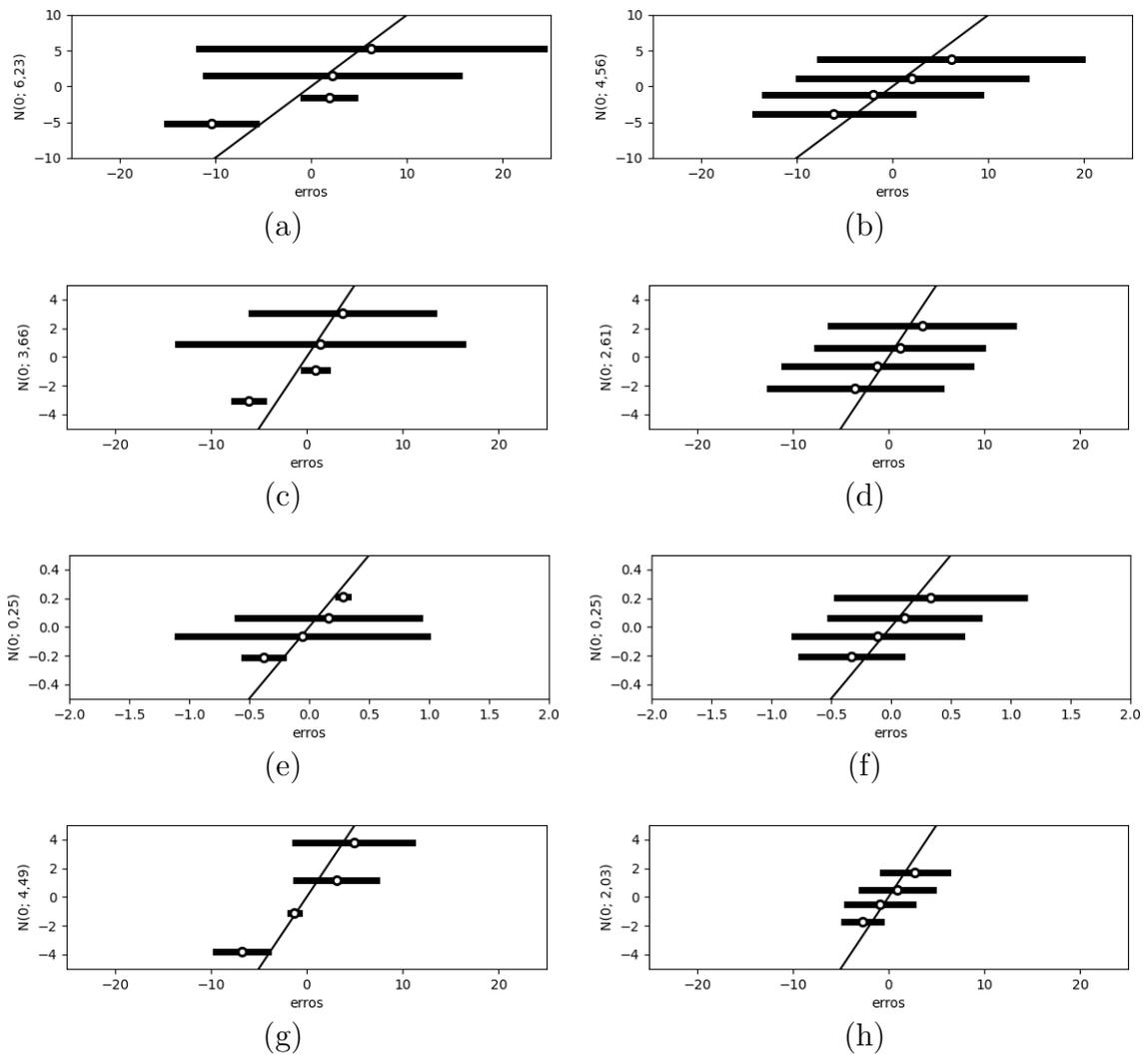
(b)



(c)

Fonte: O autor.

Figura 41 – Gráficos quantil-quantil dos erros associados aos modelos  $\mathcal{O}_1^*(x)$ , (a),  $\mathcal{O}_2^*(x)$ , (b),  $\mathcal{P}_1^*(x)$ , (c),  $\mathcal{P}_2^*(x)$ , (d),  $\mathcal{C}_1^*(x)$ , (e),  $\mathcal{C}_2^*(x)$ , (f),  $\mathcal{S}_1^*(x)$ , (g), e  $\mathcal{S}_2^*(x)$ , (h). Conforme definido na Subseção 6.5.



Fonte: O autor.

## 10.3 Conclusão

O presente capítulo investigou mais profundamente os modelos propostos no Capítulo 7 assim como o uso do coeficiente de determinação e de gráficos quantil-quantil adaptados para o caso de erros na forma de intervalos, discutidos na Subseção 6.5.

De um modo geral os gráficos quantil-quantil apontam que os modelos  $\mathcal{O}_2^*(x)$ ,  $\mathcal{P}_2^*(x)$ ,  $\mathcal{C}_2^*(x)$  e  $\mathcal{S}_2^*(x)$  representam melhor os dados do que  $\mathcal{O}_1^*(x)$ ,  $\mathcal{P}_1^*(x)$ ,  $\mathcal{C}_1^*(x)$  e  $\mathcal{S}_1^*(x)$ , respectivamente. Enquanto os valores obtidos como coeficiente de determinação não são tão distintos, assim nesse quesito todos esses modelos tem avaliação semelhante. O sobre-ajuste dos modelos  $\mathcal{O}_3^*(x)$ ,  $\mathcal{P}_3^*(x)$ ,  $\mathcal{C}_3^*(x)$  e  $\mathcal{S}_3^*(x)$ , evidenciado pelo fato de que  $\overline{R^2} = [1, 0; 1, 0]$ , leva ao descarte desses modelos.

Concluimos que dentre os modelos construídos aqui os  $\mathcal{O}_2^*(x)$ ,  $\mathcal{P}_2^*(x)$ ,  $\mathcal{C}_2^*(x)$  e  $\mathcal{S}_2^*(x)$  tem qualidade superior. Observando o peso dado a cada observação nesses exemplos podemos chegar à conclusão de que bons modelos possuem pesos maiores próximos do  $x$  onde o dado foi observado, porém, ainda possui certa influência dos dados tomados em outros valores de  $x$ .

Como os números *fuzzy* dos dados observado foram obtidos, de certa forma, da média anual dos valores de cada poluente não cabe uma interpretação destes modelos como um problema real.

# 11 Comentários Finais

Reservamos este capítulo para os comentários finais, dado que cada capítulo ao longo do texto possui suas conclusões específicas na seção “Conclusão”. Apresentamos aqui uma breve linha do tempo dos trabalhos relacionados a este trabalho.

Nos séculos 19 e 20 a pesquisa na área de regressão teve boa parte do seu desenvolvimento. Partindo do ajuste de curvas, evoluindo para as questões estatísticas de regressão e para a análise de dados longitudinais.

O método de quadrados mínimos foi proposto por Adrien-Marie Legendre em 1805. Neste ponto o processo era completamente algébrico, sem relação com estatística. Esse ajuste de curvas foi justificado estatisticamente mais tarde por Carl Friedrich Gauss, Pierre-Simon Laplace, Augustin-Louis Cauchy, Thorvald Nicolai Thiele, entre outros, ([FAREBROTHER, 2001](#)).

Segundo [Stanton \(2001\)](#), em 1875, Sir Francis Galton relacionou a massa de sementes de ervilha de cheiro de duas gerações diferentes. Dividiu as sementes “mães” em 7 grupos, cada grupo com sementes de massa semelhante entre si. Percebeu assim que a relação entre a mediana da massa das sementes “filhas” e a massa das suas sementes “mães” era bem explicada por uma função linear. Mais tarde, em 1896, Karl Pearson tratou de forma mais rigorosa os conceitos de regressão e correlação.

Alguns dos primeiros métodos propostos para analisar dados longitudinais foram propostos por Ronald Aylmer Fisher no início do século 20, ([FITZMAURICE, 2011](#)).

A partir de 1965 a teoria *fuzzy* tomou forma e começou a ser aplicada nas mais diversas áreas. Na década de 1980 os primeiros modelos de regressão possibilística, que podemos ver como ajustes de curva usando a teoria *fuzzy*, surgiram. Ao final dos anos 1990 a teoria se apresenta mais firme, com a presença dos modelos de quadrados mínimos e com o surgimento de ideias mais diversas para o tema.

A teoria de conjuntos *fuzzy* foi introduzida por Lotfali Askar-Zadeh em 1965, ([ZADEH, 1965](#)). Um conjunto *fuzzy* permite que a relação de um elemento pertencer ou não a um conjunto deixe de ser binária e seja expressa como um número real do intervalo  $[0; 1]$ , chamado de pertinência. Mais tarde, em 1975, com a finalidade de usar esses conjuntos para expressar variáveis linguísticas, introduziu o conceito de números *fuzzy*, ([ZADEH, 1975](#)). Deste modo é possível expressar termos como muito jovem, jovem, não tão jovem, não tão velho, velho, muito velho, entre outros, através de números *fuzzy*.

Em 1982 Hideo Tanaka apresentou uma proposta de regressão linear onde as variáveis resposta tem a forma de números *fuzzy* triangulares simétricos, ([TANAKA;](#)

UEJIMA; ASAI, 1982). O modelo é apresentado como uma função linear *fuzzy*, ou seja,  $y^*(x_1, \dots, x_n) = a_1^*x_1 + \dots + a_n^*x_n$ , onde os parâmetros  $a_i^*$  são números *fuzzy* triangulares simétricos obtidos através de um problema de programação linear que otimiza o modelo usando restrições que se referem à teoria *fuzzy*. No ano de 1989 propôs mais dois modelos, (TANAKA; HAYASHI; WATADA, 1989), os quais usam restrições diferentes do anterior. Esse tipo de modelo é batizado como regressão linear possibilística nesse segundo trabalho.

Em 1998 Ralf Körner e Wolfgang Näther apresentam um estudo sobre regressão linear para dados *fuzzy* discutindo o modelo clássico estendido assim como modelos de quadrados mínimos, (KORNER; NATHER, 1998). Os modelos apresentados são também na forma de função linear *fuzzy* e são explorados os problemas de o conjunto de todos os números *fuzzy*,  $\mathcal{F}(\mathbb{R})$ , não ser um espaço linear.

Também em 1998, Phil Diamond, junto a Hideo Tanaka, compilam os modelos de regressão para dados *fuzzy* dividindo nas categorias de modelos possibilísticos e de quadrados mínimos, (DIAMOND; TANAKA, 1998).

Inés Couso Blanco e Luciano Sánchez Ramos, ainda em 1998, propõem um modelo classificado como regressão quantílica, (RAMOS; BLANCO, 1998). Mais tarde, em 2014, com Didier Dubois, classifica os modelos nas quatro categorias: quadrados mínimos, possibilísticos, quantílicos e por análise sensitiva, (COUSO; DUBOIS, 2014).

O presente trabalho propõe métodos para estimar modelos na forma  $f^* : \mathbb{R}^q \rightarrow \mathcal{F}(\mathbb{R})$  para realizar regressão para dados onde as observações tem a forma de números *fuzzy* e as variáveis explicativas são números reais. No Capítulo 7 são apresentados métodos para modelos simples e múltiplo para dados transversais. A ideia dos métodos em si não é inédita, ambos provém diretamente dos métodos classificados como análise sensitiva. No entanto, apresentamos os modelos na forma de expressões fechadas, para o caso intervalar e *fuzzy*, e adicionamos a interpretação dos modelos como médias ponderadas das observações. Já no Capítulo 8 apresentamos um método para estimar o modelo linear misto em dois estágios para dados longitudinais intervalares e na forma de números *fuzzy*. Um ponto forte dos métodos apresentados é que os modelos herdam as características dos modelos clássicos, por conta de serem a união de todas as possibilidades em se tomar valores reais de cada observação. O que também implica que para o caso de todas as observações serem valores reais clássicos o modelo coincide com os vigentes usados na estatística.

Outra contribuição aparece na Seção 6.5 com a proposta de uma forma de gráfico quantil-quantil para testar a normalidade de um conjunto de erros na forma de intervalos. Temos assim uma ferramenta visual para avaliar a qualidade dos modelos de regressão transversais com dados intervalares ou na forma de números *fuzzy*. Apontamos a clara necessidade de melhorar os testes estatísticos para esse caso.

Como possíveis trabalhos futuros podemos citar a adequação de diversos testes

estatísticos para os modelos construídos com os métodos propostos ou a produção de métodos semelhantes para outras formas de dados, como intersecção de números *fuzzy* ou para números *fuzzy* tipo-2.

O problema do volume corporal das orcas-pigmeias apresentado no Capítulo 9 possui a característica da medição dos dados ter um caráter impreciso. Essa imprecisão vem do objetivo de não causar problemas ambientais, como causar incômodo, ferimentos ou até mesmo a morte dos indivíduos estudados. O volume corporal das orcas-pigmeias sendo estimado por imagens aéreas, permite que os dados sejam coletados sem estressar, ferir ou levar o animal a óbito. Árvores podem ter a sua massa estimada sem a necessidade de removê-las do chão. Os modelos apresentados possuem a capacidade de incorporar essa imprecisão aos dados observados, desta forma vários problemas na área de biomatemática podem se beneficiar deles. Outras áreas em que os dados tenham natureza imprecisa também podem se beneficiar dos modelos.

Outra possível aplicação é que através do modelo base, apresentado na Seção 7.2, outros métodos estatísticos podem ser adaptados para dados intervalares ou na forma de números *fuzzy*. Problemas de outras áreas talvez também possam ser tratados através do modelo base como, por exemplo, problemas da área de equações diferenciais ordinárias. Problemas esses que tem também diversas aplicações em biomatemática.

## Referências

- BARROS, L. C. de; BASSANEZI, R. C.; LODWICK, W. A. *A First Course in Fuzzy Logic, Fuzzy Dynamical Systems, and Biomathematics*. Springer Berlin Heidelberg, 2017. Disponível em: <<https://doi.org/10.1007/978-3-662-53324-6>>. Citado 10 vezes nas páginas 32, 34, 35, 36, 38, 39, 40, 41, 42 e 43.
- BUJA, A.; HASTIE, T.; TIBSHIRANI, R. Linear smoothers and additive models. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 17, n. 2, p. 453–510, 1989. ISSN 00905364. Disponível em: <<http://www.jstor.org/stable/2241560>>. Citado na página 17.
- CARROLL, C. L.; HUNTINGTON, P. J. Body condition scoring and weight estimation of horses. Wiley, v. 20, n. 1, p. 41–45, jan. 1988. Disponível em: <<https://doi.org/10.1111/j.2042-3306.1988.tb01451.x>>. Citado na página 44.
- COLUBI, A.; CORRAL, N.; GONZALEZ-RODRIGUEZ, G.; MONTENEGRO, M. A determination coefficient for fuzzy random variables in a fuzzy arithmetic-based linear model. In: *2007 IEEE International Fuzzy Systems Conference*. [S.l.: s.n.], 2007. p. 1–4. Citado na página 50.
- CONCEIÇÃO, G. M.; MIRAGLIA, S. G.; KISHI, H. S.; SALDIVA, P. H.; SINGER, J. M. Air pollution and child mortality: a time-series study in são paulo, brazil. *Environ. Health Perspect.*, Environmental Health Perspectives, v. 109 Suppl 3, n. suppl 3, p. 347–350, jun. 2001. Citado 2 vezes nas páginas 134 e 135.
- COUSO, I.; DUBOIS, D. Statistical reasoning with set-valued information: Ontic vs. epistemic views. *International Journal of Approximate Reasoning*, v. 55, n. 7, p. 1502–1518, 2014. ISSN 0888-613X. Special issue: Harnessing the information contained in low-quality data sources. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0888613X13001618>>. Citado 8 vezes nas páginas 43, 46, 47, 48, 50, 51, 67 e 150.
- CROWDER, M. J. *Analysis of repeated measures*. London New York: Chapman and Hall, 1990. ISBN 0-412-31830-X. Citado na página 23.
- CURRIE, J. J.; ASWEGEN, M. van; STACK, S. H.; WEST, K. L.; VIVIER, F.; BEJDER, L. Rapid weight loss in free ranging pygmy killer whales (*feresa attenuata*) and the implications for anthropogenic disturbance of odontocetes. *Sci. Rep.*, v. 11, n. 1, p. 8181, 2021. Citado 6 vezes nas páginas 13, 14, 104, 105, 106 e 133.
- DAVIDIAN, M. *Nonlinear models for repeated measurement data*. London New York: Chapman & Hall, 1995. ISBN 0412983419. Citado na página 23.
- DENOEU, T. Maximum likelihood estimation from fuzzy data using the EM algorithm. *Fuzzy Sets and Systems*, Elsevier, v. 183, n. 1, p. 72–91, 2011. Disponível em: <<https://hal.archives-ouvertes.fr/hal-00654118>>. Citado na página 101.
- DIAMOND, P.; TANAKA, H. Fuzzy regression analysis. In: SŁOWIŃSKI, R. (Ed.). *Fuzzy Sets in Decision Analysis, Operations Research and Statistics*. Boston,

- MA: Springer US, 1998. p. 349–387. ISBN 978-1-4615-5645-9. Disponível em: <[https://doi.org/10.1007/978-1-4615-5645-9\\_11](https://doi.org/10.1007/978-1-4615-5645-9_11)>. Citado 2 vezes nas páginas 46 e 150.
- DINEEN, S. *Multivariate Calculus and Geometry*. Springer London, 2014. Disponível em: <<https://doi.org/10.1007/978-1-4471-6419-7>>. Citado na página 19.
- DODGE, Y. *The Concise Encyclopedia of Statistics*. New York, NY: Springer New York, 2009. ISBN 9781441913906. Citado na página 22.
- EVERITT, B. *The Cambridge dictionary of statistics*. Cambridge, UK New York: Cambridge University Press, 2006. ISBN 9780511243233. Citado 4 vezes nas páginas 17, 23, 45 e 84.
- FAHRMEIR, L.; KNEIB, T.; LANG, S.; MARX, B. *Regression*. Springer Berlin Heidelberg, 2013. Disponível em: <<https://doi.org/10.1007/978-3-642-34333-9>>. Citado 4 vezes nas páginas 17, 18, 20 e 21.
- FAREBROTHER, R. W. Adrien-marie legendre. In: \_\_\_\_\_. *Statisticians of the Centuries*. New York, NY: Springer New York, 2001. p. 101–104. ISBN 978-1-4613-0179-0. Disponível em: <[https://doi.org/10.1007/978-1-4613-0179-0\\_20](https://doi.org/10.1007/978-1-4613-0179-0_20)>. Citado na página 149.
- FITZMAURICE, G. *Applied longitudinal analysis*. Hoboken, N.J: Wiley, 2011. ISBN 978-0-470-38027-7. Citado 6 vezes nas páginas 13, 23, 24, 25, 31 e 149.
- GUTIÉRREZ, S. C.; AGUILERA, E. D. L. Q-q plot normal. los puntos de posicion gráfica. *Iniciación a la Investigación*, n. 2, mar. 2010. Disponível em: <<https://revistaselectronicas.ujaen.es/index.php/ininv/article/view/259>>. Citado na página 22.
- HARVILLE, D. A. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, [American Statistical Association, Taylor & Francis, Ltd.], v. 72, n. 358, p. 320–338, 1977. ISSN 01621459. Disponível em: <<http://www.jstor.org/stable/2286796>>. Citado na página 26.
- KORNER, R.; NATHER, W. Linear regression with random fuzzy variables: extended classical estimates, best linear estimates, least squares estimates. *Information Sciences*, Elsevier BV, v. 109, n. 1-4, p. 95–118, ago. 1998. Disponível em: <[https://doi.org/10.1016/s0020-0255\(98\)00010-3](https://doi.org/10.1016/s0020-0255(98)00010-3)>. Citado 4 vezes nas páginas 48, 49, 67 e 150.
- LAIRD, N. M.; WARE, J. H. Random-effects models for longitudinal data. *Biometrics*, [Wiley, International Biometric Society], v. 38, n. 4, p. 963–974, 1982. ISSN 0006341X, 15410420. Disponível em: <<http://www.jstor.org/stable/2529876>>. Citado na página 26.
- LOQUIN, K.; DUBOIS, D. Kriging and epistemic uncertainty: A critical discussion. In: JEANSOULIN, R.; PAPINI, O.; PRADE, H.; SCHOCKAERT, S. (Ed.). *Methods for Handling Imperfect Spatial Information*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. p. 269–305. ISBN 978-3-642-14755-5. Disponível em: <[https://doi.org/10.1007/978-3-642-14755-5\\_11](https://doi.org/10.1007/978-3-642-14755-5_11)>. Citado na página 51.
- MORETTIN, P.; BUSSAB, W. *ESTATÍSTICA BÁSICA*. Saraiva Educação S.A., 2017. ISBN 9788502207172. Disponível em: <<https://books.google.com.br/books?id=vDhnDwAAQBAJ>>. Citado 3 vezes nas páginas 13, 18 e 21.

- NETO, E. de A. L. *MODELOS DE REGRESSÃO PARA DADOS SIMBÓLICOS DE NATUREZA INTERVALAR*. Tese (Tese de doutorado) — Universidade Federal de Pernambuco, 2012. Citado na página 50.
- PINTO, N. J. B.; WASQUES, V. F.; ESMI, E.; BARROS, L. C. Least squares method with interactive fuzzy coefficient: Application on longitudinal data. In: BARRETO, G. A.; COELHO, R. (Ed.). *Fuzzy Information Processing*. Cham: Springer International Publishing, 2018. p. 132–143. ISBN 978-3-319-95312-0. Citado 5 vezes nas páginas 43, 48, 104, 134 e 135.
- Pixabay GmbH. *Pixabay*. n.d. <<https://pixabay.com>>. Acessado em Maio de 2022. Citado 2 vezes nas páginas 14 e 104.
- RAMOS, L. S.; BLANCO, I. C. Learning from imprecise examples with ga-p algorithms. Universitat Politècnica de Catalunya. Secció de Matemàtiques i Informàtica, 1998. Disponível em: <<http://hdl.handle.net/2099/3533>>. Citado na página 150.
- STANTON, J. M. Galton, pearson, and the peas: A brief history of linear regression for statistics instructors. *Journal of Statistics Education*, Taylor & Francis, v. 9, n. 3, p. null, 2001. Citado na página 149.
- STEFANINI, L. A generalization of hukuhara difference and division for interval and fuzzy arithmetic. *Fuzzy Sets and Systems*, v. 161, n. 11, p. 1564–1584, 2010. ISSN 0165-0114. Theme: Decision Systems. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0165011409002644>>. Citado na página 35.
- SÁNCHEZ, A. D. B. *ALGUMAS CONTRIBUIÇÕES NA TEORIA DE APROXIMAÇÃO E OTIMIZAÇÃO COM VARIÁVEIS FUZZY*. Tese (Tese de doutorado) — Universidade Estadual de Campinas, 2012. Citado na página 50.
- TANAKA, H.; HAYASHI, I.; WATADA, J. Possibilistic linear regression analysis for fuzzy data. Elsevier BV, v. 40, n. 3, p. 389–396, jun. 1989. Disponível em: <[https://doi.org/10.1016/0377-2217\(89\)90431-1](https://doi.org/10.1016/0377-2217(89)90431-1)>. Citado 6 vezes nas páginas 48, 53, 64, 78, 79 e 150.
- TANAKA, H.; UEJIMA, S.; ASAI, K. Linear regression analysis with fuzzy model. Institute of Electrical and Electronics Engineers (IEEE), v. 12, n. 6, p. 903–907, 1982. Disponível em: <<https://doi.org/10.1109/tsmc.1982.4308925>>. Citado 2 vezes nas páginas 47 e 150.
- VIERTL, R. *Statistical methods for fuzzy data*. Chichester, West Sussex: Wiley, 2011. ISBN 978-0-470-69945-4. Citado 5 vezes nas páginas 38, 39, 43, 51 e 141.
- WAGNER, E. L.; TYLER, P. J. A comparison of weight estimation methods in adult horses. Elsevier BV, v. 31, n. 12, p. 706–710, dez. 2011. Disponível em: <<https://doi.org/10.1016/j.jjevs.2011.05.002>>. Citado na página 44.
- WU, L. *Mixed Effects Models for Complex Data*. [S.l.: s.n.], 2009. ISBN 9781420074024. Citado na página 26.
- ZADEH, L. Fuzzy sets. *Information and Control*, v. 8, n. 3, p. 338–353, 1965. ISSN 0019-9958. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S001999586590241X>>. Citado 3 vezes nas páginas 13, 38 e 149.

---

\_\_\_\_\_. The concept of a linguistic variable and its application to approximate reasoning—i. *Information Sciences*, v. 8, n. 3, p. 199–249, 1975. ISSN 0020-0255. Disponível em: <<https://www.sciencedirect.com/science/article/pii/0020025575900365>>. Citado na página 149.