

**UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA**

Felipe Toledo Ferreira

Introdução à Análise de Dados Funcionais com ênfase em splines e B-splines

Juiz de Fora

2022

Felipe Toledo Ferreira

Introdução à Análise de Dados Funcionais com ênfase em splines e B-splines

Trabalho de conclusão de curso apresentado ao Departamento de Estatística da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do grau de bacharel em Estatística.

Orientador: Prof. Dr. Clécio da Silva Ferreira

Juiz de Fora

2022

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Toledo Ferreira, Felipe.

Introdução à Análise de Dados Funcionais com ênfase em splines e
B-splines / Felipe Toledo Ferreira. – 2022.

84 f. : il.

Orientador: Clécio da Silva Ferreira

Trabalho de Conclusão de Curso (graduação) – Universidade Federal de
Juiz de Fora, Instituto de Ciências Exatas. Departamento de Estatística,
2022.

1. Dados funcionais. 2. Estimacão de curvas. 3. Suavizacão penalizada.
I. Ferreira, Clécio. II. Título.

Felipe Toledo Ferreira

Introdução à Análise de Dados Funcionais com ênfase em splines e B-splines

Trabalho de conclusão de curso apresentado ao Departamento de Estatística da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do grau de bacharel em Estatística.

Aprovada em 12 de agosto de 2022

BANCA EXAMINADORA

Prof. Dr. Clécio da Silva Ferreira - Orientador
Universidade Federal de Juiz de Fora

Profa. Dra. Camila Borelli Zeller
Universidade Federal de Juiz de Fora

Prof. Dr. Tiago Maia Magalhães
Universidade Federal de Juiz de Fora

Dedico este trabalho a meus avôs e avós, José e Maria Imaculada, Alberto e Manuela.

AGRADECIMENTOS

Primeiramente, gostaria de agradecer a meus pais, Alberto e Seila, que há pouco mais de duas décadas passaram a priorizar o bem estar meu e de meu irmão acima dos próprios desejos e vontades; a eles eu devo tudo. Por extensão, agradeço também meu irmão Gabriel Toledo, quem eu amo e admiro, mas nunca irei admitir. Gostaria de agradecer a toda a minha família, como uma pequena parcela de retribuição por todo o carinho e amor com qual eles desde sempre me trataram.

Agradeço também aos meus colegas de ensino médio Antônio, Carlos Henrique, Gabriel Guilherme, Heitor, Luiz Felipe e Luciano, pelos muitos momentos em que me acompanharam, incluindo a Mostra de Profissões da Universidade Federal de Minas Gerais em 2016, quando tive meu primeiro contato com a Estatística em um ambiente acadêmico. Não fiquei muito impressionado. E agradeço também aos colegas que conheci durante a graduação, Ana Flávia, Ana Júlia, Gabriel Alves, Guilherme, Leandro, Luana, Nathan, Nicolau, Thiago Luiz, Thiago Noboa e Sabrina; sem eles, certamente a Universidade seria um lugar mais triste, e ninguém comemoraria aniversário em Julho.

Agradeço a todos os professores do Departamento de Estatística da UFJF sob quais tive a oportunidade de ser aluno: Alfredo, Ângela, Augusto, Camila, Clécio, Gustavo, Henrique, Lupércio, Marcel, Ronaldo, Tiago e Tufi. Gostaria de, em particular, destacar meu orientador, o professor Clécio, que me aceitou em sua orientação para um projeto de iniciação científica na segunda metade de 2020, sem tomar conhecimento da armadilha em que caia, e em qual continua até hoje, como meu orientador para o Trabalho de Conclusão de Curso para o Bacharelado em Estatística. Eu poderia dizer louvores sem fim quanto à paciência e a temperança com que ele soube lidar com meus muitos erros e atrasos, mas prefiro destacar toda a profunda dedicação e apreço que enxerguei e ainda enxergo nele como um estatístico e como um comunicador da ciência. Seja qual for a trajetória que eu siga de agora em diante, sei que vou me espelhar na figura deste meu caro orientador. Obrigado Clécio.

Gostaria também de agradecer aos demais professores com quais tive aulas, dos Departamentos de Matemática, Ciência da Computação, Física e Química da UFJF. Particularmente, gostaria de agradecer ao professor Sandro, do Departamento de Matemática, qual me orientou em dois projetos de iniciação científica, tão como meu Trabalho de Conclusão de Curso para o Bacharelado em Ciências Exatas. Sandro sempre mostrou-se paciente e gentil, e ativamente trabalhou para despertar em mim o interesse pela ciência, seja na Matemática ou na Estatística, e por isso eu o agradeço.

Por fim, agradeço a todos os funcionários da Universidade Federal de Juiz de Fora, em todos os setores por quais já passei; no Restaurante Universitário, nas Bibliotecas, na Central de Atendimento, nos Circulares, entre tantos outros. E agradeço a todas as

peessoas que via no meu querido percurso de casa até o ponto de ônibus, e do ponto de ônibus até a UFJF; todos os motoristas, todos os passageiros e todos os pedestres só perambulando por aí. Obrigado.

“Talvez não tenha tanto medo assim
De trovão” A. F. Elétrico (2019, (1)).

“With four parameters I can fit an elephant, and with five I can make him wiggle his trunk.”
J. V. Neumann, como citado em F. Dyson (2004, (2)).

Far better an approximate answer to the right question, which is
often vague, than the exact answer to the wrong question, which
can always be made precise. J. W. Tukey (1962, (3)).

RESUMO

O desenvolvimento de técnicas de estimação não paramétrica de curvas, qual da segunda metade do século XX à frente tem sido intimamente ligado aos avanços nas tecnologias de computação científica, tem aberto diversos caminhos a áreas previamente consideradas de alta especialização, pelas suas tangências à análise funcional e técnicas de estimação quais usualmente apresentam alto custo computacional. Nestas áreas os objetos de estudo das amostras cessam a estar limitados a quantidades finitas de valores, por assim dizer ‘discretos’, e passam a abarcar também objetos como curvas e superfícies, quais assumem valores em espaços funcionais de dimensão infinita. Neste âmbito, pode-se considerar que a Análise de Dados Funcionais (ADF) é portanto uma progressão à Análise de Dados Multivariados (ADM), em que a ADM originalmente abordava amostras compostas de vetores e outros objetos de dimensão finita. É necessário então compreender quais desenvolvimentos na estimação não paramétrica de curvas são aplicados nesta nova área de análise de dados funcionais, e como. Inicialmente, é feita uma introdução ao tópico da análise de dados funcionais, onde são apresentadas suas estatísticas descritivas, assim como a função de covariância; adiante, é feita uma breve introdução à interpolação e suavização de dados, concluindo na apresentação de splines e B-splines como funções úteis para estes propósitos, tão como a apresentação de métodos de regularização aplicados à não suavidade. Então, são introduzidos os tópicos de agrupamento funcional e modelos de regressão funcionais, quais servem de base para as aplicações finais. No percorrer do trabalho, é utilizado o pacote `fda`, desenvolvido por Ramsay *et al* (2009, (4)).

Palavras-chave: Dados funcionais. Estimação de curvas. Suavização penalizada.

ABSTRACT

The development of non parametric curve estimation, which from the second half of the 20th century onward has been closely linked to the advances in computational science, has opened many paths to fields previously deemed ‘high specialty’, given their proximity to functional analysis and usually computationally costly estimation techniques. In these areas, the elements of interest cease to be limited to finite quantities of values, ‘discrete’ so to speak, and begin to encompass objects such as curves and surfaces, all of which assume values in functional spaces of infinite dimension. Through this lens, Functional Data Analysis (FDA) can be considered the progression of Multivariate Datas Analysis (MDA), whence MDA originally encompassed samples composed of vectors and other finite dimensional objects. It is therefore necessary to comprehend which developments in non parametric curve estimation are applied within this new field of functional data analysis, and how: that is the purpose of this work. Initially, the topic of functional data analysis is introduced, covering its summary statistics, as well as the covariance function; thereafter, there is a brief introduction to data interpolation and smoothing, concluding in the presentation of splines and B-splines as useful functions for this purpose, as well as roughness penalties. Functional clustering and functional regression models are thereafter introduced, to serve as foundation to the concluding applications. The `fda` package, as developed by Ramsay *et al* (2009, (4)), is utilized throughout.

Keywords: Functional data. Curve estimation. Penalized smoothing

LISTA DE ILUSTRAÇÕES

Figura 2.1 –Amostra da absorvância de cinco diferentes carnes picadas quando submetidas a comprimentos de onda na região de 850 a 1050 nanômetros.	21
Figura 2.2 –Curva média (destacada em vermelho) da amostra completa da absorvância de 215 carnes picadas distintas submetidas a diferentes comprimentos de onda	22
Figura 2.3 –(a) Curva de variância para a amostra obtida no experimento sobre absorvâncias em carnes picadas distintas submetidas a diferentes comprimentos de onda; (b) Superfície de covariância entre diferentes comprimentos de onda para o mesmo experimento	23
Figura 3.1 –Aproximação de $f(t) = \sin(\pi t)$ via polinômios de Taylor de graus $k = 1$ até $k = 4$, com $t_0 = 0$. O quadrado colorido em vermelho destaca o ponto $(0, 0)$	33
Figura 3.2 –Spline sob uso para desenho técnico; nota-se que a spline propriamente dita trata-se apenas da tira sendo curvada.	36
Figura 3.3 –Aproximações via splines lineares e cúbicas para dados gerados para simulação da função (3.12). Os pontos coloridos em vermelho indicam os valores mensurados.	39
Figura 3.4 –(a) Funções base de B-splines de ordens 2 e 3 com nós em $\xi_1 = 0, \xi_2 = 1, \xi_3 = 2, \xi_4 = 3$ e $\xi_5 = 4$; (b) Funções base de B-splines de ordens 2 e 3 com nós em $\xi_1 = 0, \xi_2 = 1, \xi_3 = 2$ e $\xi_4 = 4$	41
Figura 3.5 –Curvas suaves geradas para aproximação da função (3.19), com distintos valores do coeficiente de regularização λ	44
Figura 4.1 –Intervalos de confiança pontuais a nível 95% para as absorvâncias mensuradas em 215 carnes picadas submetidas a diferentes comprimentos de onda, denotados pela curvas tracejadas em azul.	51
Figura 4.2 –(a) Curva estimada para $\beta(t)$ para um modelo de regressão funcional do teor de gordura de 215 carnes picadas com respeito à curva de absorvância das mesmas; (b) Valores ajustados do teor de gordura comparados aos valores reais de teor de gordura para este mesmo modelo.	60
Figura 5.1 –(a) Curvas de absorvância com respeito a intensidade do feixe estimadas para a amostra de 118 árvores do conjunto de dados Celulose; (b) <i>Boxplot</i> para os rendimentos observados para árvores do conjunto de dados de Celulose.	62
Figura 5.2 –(a) Curva $\hat{\beta}(t)$ estimada para o modelo de regressão funcional modelando o rendimento mensurado em um conjunto de árvores com respeito às curvas de absorvância por intensidade de feixe das mesmas árvores, no conjunto de dados Celulose; (b) Valores ajustados para o rendimento comparados aos valores reais de rendimento para este mesmo modelo.	64

Figura 5.3 –Curvas do número de mortes por milhão de habitantes em decorrência da COVID-19 em capitais e cidades de grande porte brasileiras, já agrupadas. Destacam-se os picos dos municípios de Belém (PA) Manaus (AM) e Juiz de Fora (MG).	67
Figura 5.4 –(a) Curvas da média por grupo do número de mortes por milhão de habitantes em decorrência da COVID-19 em capitais e cidades de grande porte brasileiras; (b) Curva da variância por grupo do número de mortes por milhão de habitantes em decorrência da COVID-19.	69
Figura 5.5 –(a) Derivadas das curvas do número de mortes por milhão de habitantes em decorrência da COVID-19 em capitais e cidades de grande porte brasileiras, já agrupadas; (b) Curvas médias por grupo da derivada das curvas de número de mortes por milhão de habitantes em decorrência da COVID-19 em capitais e cidades de grande porte brasileiras.	70
Figura 5.6 –(a) Coeficientes de silhueta médios estimados para as dissimilaridades das curvas geradas conforme (5.1), com método de ligação máximo e mínimo; (b) Realizações simuladas conforme (5.1), já agrupadas.	72
Figura 5.7 –Realizações simuladas conforme (5.2), com deslocamento de fase, e fator de deslocamento de fase $h = 1/2$, já agrupadas.. . . .	74

LISTA DE TABELAS

- Tabela 1 – (a) Tabela de idade x comparada ao total de habitantes com pelo menos x anos de idade T_x , no dia primeiro de Junho de 1900 nos Estados Unidos da América; (b) Diferenças simples de ordem de um a cinco entre T_x para cada x 34
- Tabela 2 – Comparação de medidas de qualidade de ajuste para quatro diferentes métodos de modelagem do rendimento mensurado pelas árvores com respeito às absorbâncias por intensidade de feixe distintas, para o conjunto de dados de Celulose. 63
- Tabela 3 – Atribuições por grupo do agrupamento realizado sobre as curvas de morte por milhão de habitantes em decorrência da doença COVID-19, com medida de silhueta s_i associada a cada município. 68
- Tabela 4 – Tabela de contingência comparando os grupos reais de cada realização e os grupos a quais elas foram atribuídas, via método de ligação completa. . . 73

LISTA DE ABREVIATURAS E SIGLAS

ACP	Análise de Componentes Principais
ACPF	Análise de Componentes Principais Funcionais
ADF	Análise de Dados Funcionais
ADM	Análise de Dados Multivariados
NIRS	Near-infrared spectroscopy / Espectometria no infravermelho próximo
v.a.f.	Variável aleatória funcional
SEQ	Soma de erros quadrados
SEQPEN	Soma de erros quadrados penalizada
VC	Validação cruzada
VCG	Validação cruzada generalizada
EQM	Erro Quadrático Médio

LISTA DE SÍMBOLOS

$D^m f$	m -ésima derivada de uma função univariada $f(t)$
$\mathbf{1}_{\mathcal{A}}(x)$	Função indicadora do elemento x com respeito ao conjunto \mathcal{A}
\forall	Para todo
\in	Pertence
$\sum_{i=p}^q$	Somatório de termos indexados de p até q , dados $p \leq q$
$\mathbb{E}[\cdot]$	Valor esperado de \cdot
$\text{Var}[\cdot]$	Variância de \cdot
$\text{Cov}[\cdot, \cdot]$	Covariância entre \cdot e \cdot
Δ^m	Operador de ordem m de diferenças divididas
$[a, b]$	Intervalo real aberto em a e b , dado $a \leq b$
\mathbb{I}_m	Matriz identidade de ordem m
$A \otimes B$	Produto de Kronecker entre as matrizes A e B
$\mathbf{B}_j^{(k)}$	j -ésimo B-spline de ordem k

SUMÁRIO

1	INTRODUÇÃO	16
1.1	DADOS FUNCIONAIS	16
1.2	ESTIMAÇÃO NÃO PARAMÉTRICA DE CURVAS	17
1.3	ESCOPO E OBJETIVOS	18
2	VARIÁVEIS ALEATÓRIAS FUNCIONAIS	20
2.1	DEFINIÇÃO	20
2.2	ESTATÍSTICAS DESCRITIVAS PARA VARIÁVEIS FUNCIONAIS	21
2.3	PROPRIEDADES E FUNÇÕES DE COVARIÂNCIA	22
2.4	FILTROS DE SUAVIZAÇÃO LINEAR	25
2.5	REPRESENTAÇÃO DE CURVAS VIA FUNÇÕES BASE	26
2.5.1	Estimação de curvas via mínimos quadrados	27
<i>2.5.1.1</i>	Mínimos quadrados ponderados	27
<i>2.5.1.2</i>	Propriedades de projeção ortogonal	28
<i>2.5.1.3</i>	Graus de liberdade efetivos	28
<i>2.5.1.4</i>	Comentário sobre estimação de múltiplas curvas	29
3	CLASSES DE FUNÇÃO BASE	31
3.1	FUNÇÕES POLINOMIAIS	31
3.2	INTERPOLAÇÃO E APROXIMAÇÃO	32
3.3	SPLINES	36
3.4	B-SPLINES	38
3.5	ESCOLHA DO NÚMERO DE FUNÇÕES BASE	40
3.6	REGULARIZAÇÃO APLICADA À ESTIMAÇÃO DE CURVAS	42
3.6.1	ESCOLHA DO COEFICIENTE DE REGULARIZAÇÃO	45
<i>3.6.1.1</i>	Validação Cruzada	45
<i>3.6.1.2</i>	Validação Cruzada Generalizada	46
<i>3.6.1.3</i>	Comentário sobre Validação Cruzada de múltiplas curvas	46
3.7	SUAVIZAÇÃO RESTRITA	47
3.7.1	Suavização positiva	48
4	ANÁLISE DE DADOS FUNCIONAIS	49
4.1	ESTIMAÇÃO DA FUNÇÃO DE COVARIÂNCIA	49
4.1.1	Construção do intervalo de confiança	50
4.2	AGRUPAMENTO DE DADOS FUNCIONAIS	51
4.2.1	Medidas de dissimilaridade	52
<i>4.2.1.1</i>	Distância de Hellinger	53
<i>4.2.1.2</i>	Distância de Kullback-Leibler	53
<i>4.2.1.3</i>	Diferença Quadrática Integrada (DQI)	53
4.2.2	Métodos de ligação	53

4.2.2.1	Método de ligação mínima	54
4.2.2.2	Método de ligação máxima	54
4.2.3	Seleção da quantidade de grupos	54
4.2.3.1	Método da Silhueta Média	55
4.3	MODELOS DE REGRESSÃO FUNCIONAIS	56
4.3.1	MODELO COM VARIÁVEL DE INTERESSE ESCALAR E VARIÁVEL EXPLICATIVA FUNCIONAL	56
4.3.1.1	Estimativa naïve	56
4.3.1.2	Estimativa via regularização	57
5	APLICAÇÕES	61
5.1	CONJUNTOS DE DADOS REAIS	61
5.1.1	Conjunto de dados de Celulose	61
5.1.2	Conjunto de dados de COVID-19	65
5.2	ESTUDOS DE SIMULAÇÃO	70
5.2.1	Agrupamento simulado	70
5.2.1.1	Caso sem deslocamento de fase	71
5.2.1.2	Caso com deslocamento de fase	71
6	CONCLUSÃO	75
6.1	CONSIDERAÇÕES FINAIS	75
6.2	ESTUDOS FUTUROS	75
6.2.1	Modelos de regressão funcionais com respostas funcionais	76
6.2.2	Regularização de curvas com operadores diferenciais não lineares	77
	REFERÊNCIAS	78

1 INTRODUÇÃO

1.1 DADOS FUNCIONAIS

Tradicionalmente na estatística, os eventos e fenômenos observados são mensurados e registrados como conjuntos finitos de valores, por assim dizer ‘discretos’. Existem ainda contextos experimentais onde uma amostra de eventos é medida quando submetida a condições distintas, ou em instantes distintos; este segundo caso é tradicional à área de ‘amostragem longitudinal’, visto por exemplo em Diggle (2002, (5)). É usual então a construção de modelos preditivos, que **extrapolam** os dados mensurados de forma a possibilitar que sejam feitas conclusões quanto ao comportamento de um determinado fenômeno quando submetidos a condições ‘além’ do escopo da amostra.

Tão relevante quanto a capacidade de se obter conclusões para instantes ‘além’ daqueles mensurados é a capacidade de se obter conclusões acerca o comportamento de um determinado fenômeno quando submetido às condições ‘entre’ aquelas quais ele foi mensurado. Esta área de estudo é referida como **interpolação**, cujo estudo é de alta utilidade prática pela sua capacidade de gerar funções contínuas quais podem ser aplicadas em estudos comparativos, inclusive para estimação de derivadas. A título de exemplo, em Rao (1958, (6)) o autor discute tais estudos comparativos para curvas de crescimento, estimadas a partir de uma quantia finita de medidas. Nestes contextos, para cada observação da amostra obtém-se uma curva de crescimento distinta, que assume valores em um espaço de dimensão infinita, ou espaço funcional. Cada uma destas curvas será então um dado funcional, gerado por uma variável aleatória funcional que descreve a curva de crescimento de determinada observação. Em Ferraty e Vieu (2006, (7)), os autores propõem a seguinte definição de variáveis aleatórias funcionais e dados funcionais:

Definição 1 *Uma variável aleatória \mathcal{X} é denominada ‘variável aleatória funcional’ (v.a.f.) se assume valores em um espaço de dimensão infinita, ou espaço funcional. Uma observação χ de \mathcal{X} é denominada ‘dado funcional’.*

Nota-se, porém, que o estudo de inferência sob curvas não é necessariamente novo, remetendo em grande parte a estudos anteriores nas áreas de inferência sob processos estocásticos e séries temporais, por exemplo, realizados por Grenander (1950, 1981, (8, 9)) e Parzen (1961, (10)). Porém, estes estudos em sua maioria abordavam curvas individuais, enquanto que, como visto no estudo feito por Rao, a análise de dados funcionais é aplicada sob amostras completas de dados funcionais. Para a definição de uma amostra de dados funcionais, remete-se a Ferraty e Vieu (2006, (7)).

Definição 2 *Um conjunto de dados funcionais $\chi_1, \chi_2, \dots, \chi_n$ consiste de n observações de variáveis funcionais $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_n$ identicamente distribuídas conforme \mathcal{X} .*

Em Ramsay e Dalzell (1991, (11)), artigo em qual o nome ‘Dados Funcionais’¹ foi empregado pela primeira vez, os autores sustentam que a análise de dados funcionais (ADF) trata-se de uma progressão natural à análise de dados multivariados (ADM), onde a primeira aborda observações qual assumem valores em espaços de dimensão infinita, enquanto a segunda aborda observações qual assumem valores em espaços de dimensão finita; conseqüentemente, há significativa transferibilidade entre as técnicas aplicadas nestas áreas.

Como exemplos de estudos quais aproveitaram desta transferibilidade, tem-se: Besse e Ramsay (1986, (12)), onde os autores apresentam as técnicas de análise de componentes principais (ACP) para dados funcionais, fundamentadas no contexto de processos estocásticos por Karhunen (1946, (13)) e Loève (1946, (14)), quais são denominadas técnicas de ‘análise de componentes principais funcionais’ (ACPF). Há ainda estudos na área de classificação, como Müller (2005, (15)), seja via regressão logística, como em James (2002, (16))²; via máquinas de vetores de suporte, como em Rossi e Villa (2006, (17)); ou via K vizinhos mais próximos, como em Zhang e Zhu (2022, (18)).

Ao presente trabalho são mais relevantes os desenvolvimentos na área de agrupamento de dados funcionais, onde é possível destacar os estudos de García *et al* (2015, (19)), Yamamoto e Terada (2014, (20)) e Tokushige *et al* (2007, (21)) que estendem o algoritmo de K -médias ao contexto de dados funcionais. Adicionalmente, é possível citar Ieva *et al* (2013, (22)), que utiliza do agrupamento em um estudo comparativo de curvas obtidas via eletrocardiogramas. Tokushige *et al* (2007, (21)), que estende o algoritmo de K -médias nebulosas ao contexto de dados funcionais. Finalmente, em Abraham *et al* (2003, (23)) os autores abordam o tópico de agrupamento não supervisionado utilizando B-splines como funções base às curvas estimadas, Zambom *et al* (2021, (24)) o agrupamento de superfícies, entre vários outros.

1.2 ESTIMAÇÃO NÃO PARAMÉTRICA DE CURVAS

Embora os objetos de estudos na análise de dados funcionais sejam curvas, entende-se que, na grande maioria dos experimentos, a obtenção de dados na forma de funções contínuas não é possível. Desta forma, assim como usualmente é feito na área supracitada de análise de dados longitudinais, será feita a **interpolação** dos dados mensurados em um mesmo elemento da amostra, quando submetido a condições distintas, ou instantes distintos, de forma a obter-se curvas quais descrevem a evolução de tais mensurações com respeito às variações com o tempo ou outras condições; essas, por sua vez, serão os objetos de estudo. Há uma grande variedade de métodos quais podem efetuar a interpolação, entre os mais populares, há aqueles advindos da área de estimação não paramétrica de

¹ Traduzido do inglês ‘*Functional Data*’.

² Vale notar que este artigo de James fundamenta a aplicação de modelos lineares generalizados funcionais como um todo, não apenas para regressão logística.

curvas. Ferraty e Vieu (2006, (7)) definem modelos não paramétricos, no contexto de dados funcionais, como o seguinte

Definição 3 *Dado uma variável aleatória Z qual assume valores em um espaço de dimensão infinita Ω , isto é, uma variável aleatória funcional distribuída de forma qualquer, e uma função $\phi : \Omega \rightarrow \mathbb{R}^q$ dependente na distribuição de Z , um modelo para estimação de ϕ consiste na introdução de uma restrição da forma*

$$\phi \in \mathcal{C}.$$

O modelo para estimação de ϕ é dito paramétrico se \mathcal{C} é indexado por um número finito de elementos de Ω ; caso contrário, o modelo é dito não-paramétrico.

Uma subárea fundamental à estimação não paramétrica de curvas é o estudo das classes de funções base: estas são classes de funções conhecidas, e que possuem propriedades úteis para aproximação de outras funções reais. Há uma significativa variedade de classes quais podem ser utilizadas: em Sassi (2016, (25)) e Fonseca (2021, (26)), os autores apresentam a aplicação de ondaletas sob o contexto de dados funcionais, enquanto em Saraiva (2009, (27)), o autor utiliza de B-splines em sua implementação, quais serão abordadas no presente trabalho. Em Ramsay e Silverman (2002, (28)) os autores apresentam, entre outros conceitos, um estudo de registros de estações meteorológicas, modelos via Séries de Fourier em razão da periodicidade esperada do fenômeno; estas três classes configuram as mais representadas e utilizadas na literatura de dados funcionais.

Há ainda outras formas de estimação não paramétrica de curvas, como a suavização por kernels, como abordada por Wand e Jones (1995, (29)), qual Silverman (1984, (30)) demonstrou ser equivalente à estimação via splines, sob certas condições. Enquanto entre as técnicas paramétricas para estimação de curvas, pode-se destacar por exemplo a aplicação em Konzen (2021, (31)), onde o autor contextualiza a implementação da inferência sob processos gaussianos na área de dados funcionais. Nota-se que, em Kimeldorf e Wahba (1970, (32)) e Pope (2019, (33)), os autores demonstram a equivalência entre métodos de estimação de processos gaussianos com a estimação não paramétrica de curvas via splines e ondaletas, respectivamente.

1.3 ESCOPO E OBJETIVOS

A análise de dados funcionais, assim como a Estatística e Probabilidade como um todo, é profundamente fundamentada em teorias matemáticas. Tópicos usualmente abordados nas áreas de álgebra linear e análise funcional são comumente encontrados na literatura acerca de dados funcionais e estimação não paramétrica de curvas, onde há intersecções também com áreas como aprendizado de máquina. Muito deste embasamento está contido na produção de autores como James Ramsay e Grace Wahba, e resumido em

publicações como Hsing (2015, (34)). Embora estes tópicos sejam de contínuo interesse prático a todos alunos da Estatística, optou-se por omiti-los no presente trabalho. Nota-se que, apesar de omissos no trabalho em si, a bibliografia qual o constitui apresenta uma multitude de estudos em quais estes tópicos são abordados.

O objetivo do presente trabalho é então produzir um recurso conciso e acessível tanto para o estudo da análise de dados funcionais como também para o estudo de estimação não paramétrica de curvas. É apresentada uma aplicação na área de agrupamento de dados funcionais e modelos de regressão funcionais, de forma a demonstrar a aplicabilidade dos conceitos abordados. No entremeio, a implementação de B-splines foi favorecida por estar intimamente vinculada ao histórico de desenvolvimento da área de interpolação, como descrito pelo autor em Meijering (2002, (35)), tão como pela intuitividade de suas propriedades, muitas das quais estão ligadas às propriedades de funções polinomiais.

Destacam-se entre as principais fontes para confecção do presente trabalho é possível destacar Dias (2010, (36)), Ramsay e Silverman (2005, (37)), Wang *et al* (2016, (38)) e Ferraty (2011, (39)). Em Ramsay *et al* (2009, (4)), os autores descrevem a implementação de um pacote computacional, denominado `fda`, em linguagem R, linguagem desenvolvida pelo R Core Team (2022, (40)), qual será utilizada no desenvolvimento do trabalho, tão como o supracitado pacote. Para desenvolvimento de recursos de visualização gráfica será utilizado o pacote `ggplot2`, desenvolvido por Wickham (2016, (41)).

2 VARIÁVEIS ALEATÓRIAS FUNCIONAIS

2.1 DEFINIÇÃO

Para definir-se uma variável aleatória funcional, se remete à definição de processos estocásticos em espaços contínuos unidimensionais: dada uma coleção de eventos aleatórios indexados de maneira ordenada $\{X(t_j)\}_{j \in \{0,1,2,\dots\}}$, onde $a \leq t_0 < t_1 < t_2 < \dots \leq b$, conforme se acercam os instantes¹ consecutivos t_k e t_{k+1} , é possível aproximar a distribuição destes eventos por $\mathcal{X} = \{X(t); t \in \mathcal{T}\}$, em que $\mathcal{T} = [a, b] \subset \mathbb{R}$;² isto é, uma única observação, ou realização, de uma variável aleatória funcional corresponde a todos os valores atribuídos a $X(t)$ no intervalo $[a, b]$. Em conclusão, uma amostra de variáveis funcionais $\chi_1, \chi_2, \dots, \chi_n$ é composta de funções advindas de um espaço funcional Ω arbitrário. A Figura 2.1 expõe um exemplo de uma amostra de dados funcionais, obtidos em um experimento na área de Espectroscopia no Infravermelho Próximo (NIRS), disponível em <http://lib.stat.cmu.edu/datasets/tecolor>, exibindo uma amostra de cinco curvas geradas pela medição da absorvância em diferentes carnes picadas quando submetidas a diferentes comprimentos de onda, feitas via o dispositivo *Tecator Infratec Food and Feed Analyzer*. O objetivo original do trabalho era a construção de modelos de regressão qual fossem capazes de prever o teor de gordura das carnes picadas a partir destas medidas de absorvância.

Nota-se que, assim como no contexto de variáveis aleatórias tradicionais, o espaço funcional Ω é utilizado para se restringir variáveis aleatórias funcionais a certas condições. Um exemplo, visto em Ramsay e Silverman (2005, p. 111-114, (37)), é a restrição $\Omega = \{\chi \in \mathcal{H}; \chi(t) > 0, \forall t \in \mathcal{T}\}$ isto é, todas variáveis funcionais amostradas de Ω devem ser estritamente positivas; com $\mathcal{H} \supseteq \Omega$ indica um espaço funcional qualquer. Como as realizações tratam-se de funções, é possível representar esta restrição na forma da seguinte equação diferencial ordinária:

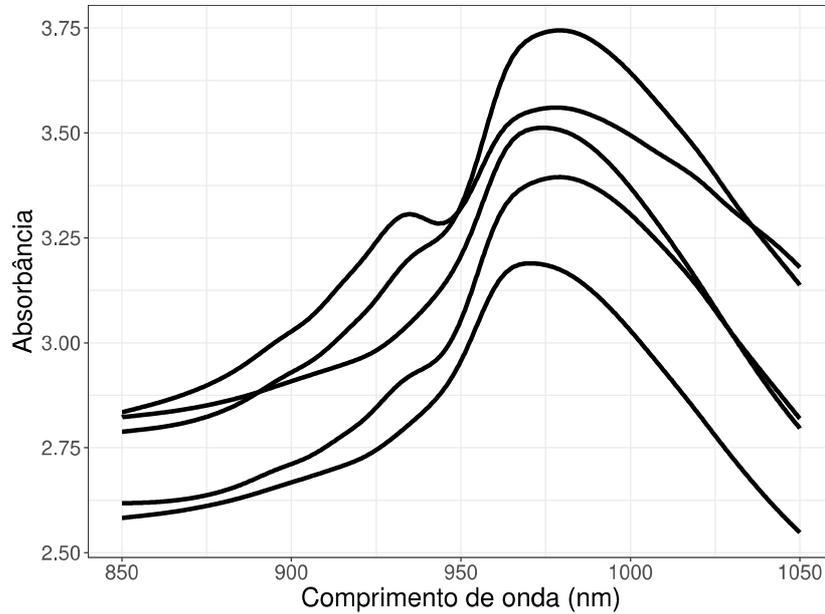
$$D\chi(t) = \omega(t)\chi(t) \quad (2.1)$$

onde $\omega(t) \in \mathbb{R}$ indica uma função qualquer definida sob o intervalo $[a, b]$. Há ainda outras restrições quais podem ser impostas às variáveis aleatórias funcionais; em Ramsay (1998, (43)), o autor propõe um método, também facilitado pela implementação de equações

¹ Uma vez que tradicionalmente processos estocásticos são tratados como eventos indexados pelo tempo, alguns se referem a pontos discretos destes como ‘instantes’. A título de exemplo, Heno (2011, (42)) apresenta uma aplicação de dados funcionais em krigagem, com dados espacialmente indexados.

² Por conveniência, aqui se adota que \mathcal{T} é um subconjunto fechado do conjunto dos reais unidimensionais (\mathbb{R}), de forma que \mathcal{X} descreve uma curva; no entanto, é possível estender a casos em quais \mathcal{X} descreve objetos de dimensões superiores, conquanto que finitas. Similarmente, é possível tratar de intervalos que apresentam descontinuidades da forma $[a, b] \cup (c, d]$, dados $a < b < c < d$.

Figura 2.1 - Amostra da absorbância de cinco diferentes carnes picadas quando submetidas a comprimentos de onda na região de 850 a 1050 nanômetros.



Fonte: Elaborado pelo autor.

diferenciais ordinárias, para manipular variáveis funcionais que são monótonas crescente ou decrescentes.

A obtenção de dados em forma explicitamente funcional, no entanto, é inviável na grande maioria dos experimentos; conseqüentemente, é necessária a utilização de técnicas de estimação não paramétrica de curvas, a qual referem-se como ‘suavização de dados funcionais’ ou também ‘regularização de dados funcionais’, para transpôr as observações de valores discretos a funções.

2.2 ESTATÍSTICAS DESCRITIVAS PARA VARIÁVEIS FUNCIONAIS

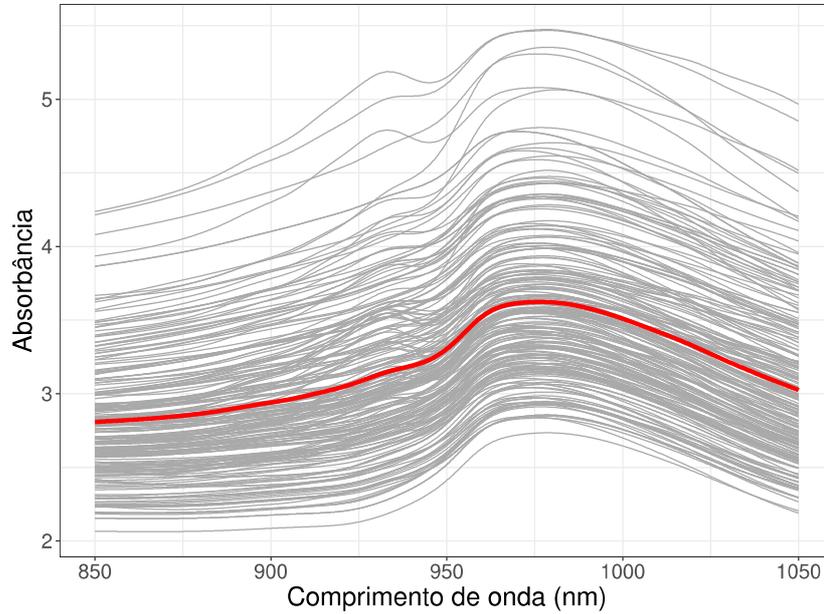
As formas de cálculo de estatísticas descritivas como a média, variância, covariância e correlação são similares às fórmulas aplicadas às variáveis aleatórias discretas. Desta forma, a partir de uma amostra de dados funcionais $\chi_1, \chi_2, \dots, \chi_n$, a função média é dada por

$$\bar{\chi}(t) = \sum_{i=1}^n \chi_i(t). \quad (2.2)$$

A Figura 2.2 exhibe a curva média de absorbância das 215 carnes picadas obtidas no experimento citado anteriormente. A função variância é dada por

$$\text{var}_{\chi}(t) = \frac{1}{n-1} \sum_{i=1}^n [\chi_i(t) - \bar{\chi}(t)]^2. \quad (2.3)$$

Figura 2.2 - Curva média (destacada em vermelho) da amostra completa da absorbância de 215 carnes picadas distintas submetidas a diferentes comprimentos de onda.



Fonte: Elaborado pelo autor.

Dada uma amostra adicional de dados funcionais $\vartheta_1, \vartheta_2, \dots, \vartheta_n$ distribuídos conforme \mathcal{V} , definida também no intervalo $\mathcal{T} = [a, b]$, e dados t_j e t_k tal que $t_j, t_k \in \mathcal{T}$, a função de covariância é dada por

$$\text{cov}_{\chi, \vartheta}(t_j, t_k) = \frac{1}{n-1} \sum_{i=1}^n [\chi_i(t_j) - \bar{\chi}(t_j)] [\vartheta_i(t_k) - \bar{\vartheta}(t_k)] \quad (2.4)$$

e por fim, a função de correlação é

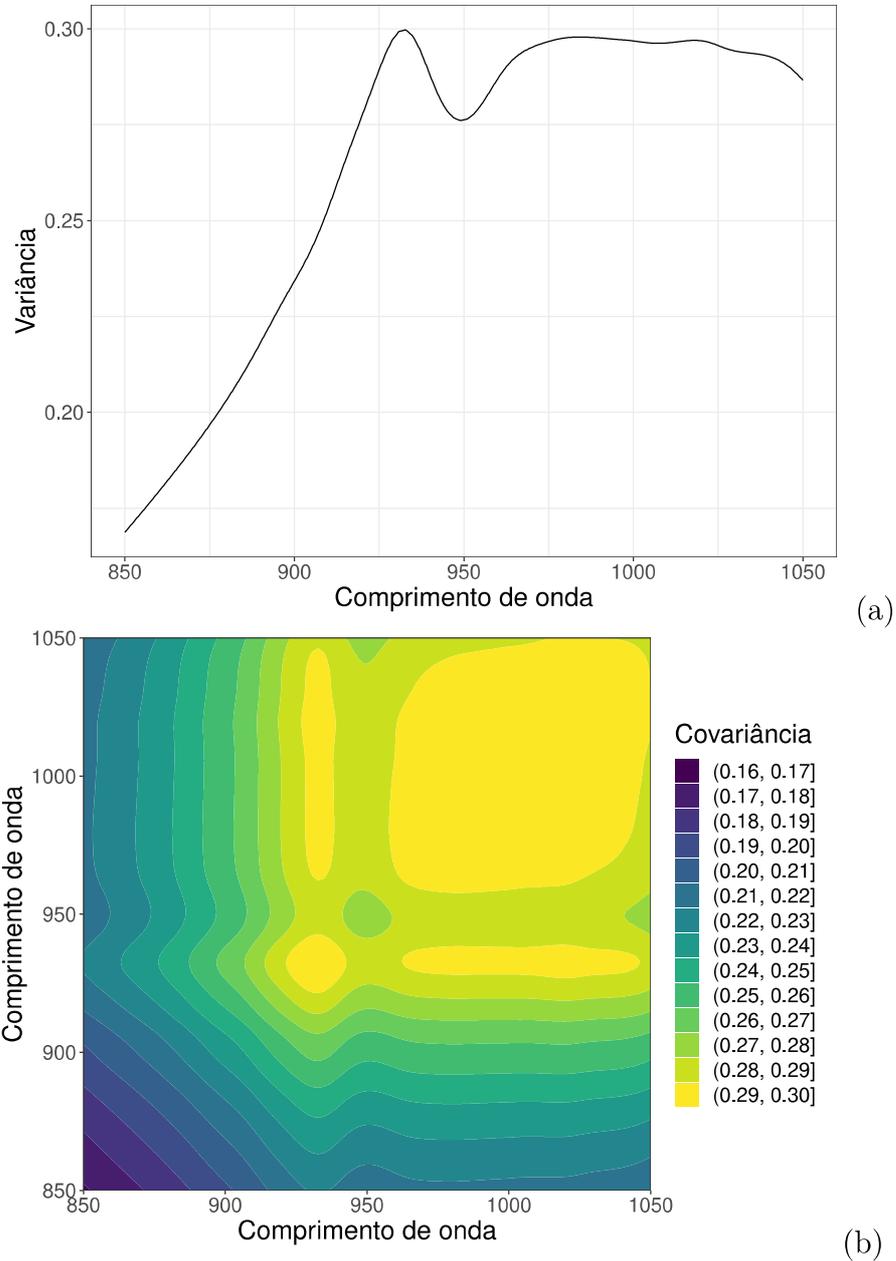
$$\text{corr}_{\chi, \vartheta}(t_j, t_k) = \frac{\text{cov}_{\chi, \vartheta}(t_j, t_k)}{\sqrt{\text{var}_{\chi}(t_j) \text{var}_{\vartheta}(t_k)}}. \quad (2.5)$$

Nota-se que as funções $\text{cov}_{\chi, \chi}(t_j, t_k)$ e $\text{corr}_{\chi, \chi}(t_j, t_k)$ podem ser utilizadas para mensuração da dependência dos valores de uma variável funcional entre instantes distintos t_j e t_k , tão como trivialmente é possível apontar que $\text{cov}_{\chi, \chi}(t, t) = \text{var}_{\chi}(t)$. A Figura 2.3 (a) exibe a curva de variância da absorbância das 215 picadas obtidas no experimento, enquanto a Figura 2.3 (b) exibe a superfície definida pela covariância da absorbância das mesmas entre instantes distintos.

2.3 PROPRIEDADES E FUNÇÕES DE COVARIÂNCIA

É possível representar todas as medidas y_j observadas de uma realização de um dado funcional como um vetor de m valores, mensurados em m instantes distintos, como

Figura 2.3 - (a) Curva de variância para a amostra obtida no experimento sobre absorbâncias em carnes picadas distintas submetidas a diferentes comprimentos de onda; (b) Superfície de covariância para o mesmo experimento.



Fonte: Elaborado pelo autor.

em

$$y_j = \underbrace{\chi(t_j)}_{(a)} + \underbrace{e_j}_{(b)}, \quad (2.6)$$

onde $j \in \{1, 2, \dots, m\}$, o termo destacado em **(a)** representa o componente sistemático, ou determinístico, da medida, enquanto o termo destacado em **(b)** representa o termo exógeno, 'ruído', ou componente probabilístico da medida. É então e_j que ocasiona a não

suavidade de nossas observações. A intenção por trás da implementação dos métodos de suavização de dados funcionais é filtrar o ruído de nossas medidas, permitindo a estimação da parte sistemática. Em forma vetorial, (2.6) reescreve-se como

$$\mathbf{y} = \chi(\mathbf{t}) + \mathbf{e},$$

onde $\mathbf{t} = [t_1, \dots, t_m]^\top$, $\mathbf{e} = [e_1, \dots, e_m]^\top$ e $\mathbf{y} = [y_1, \dots, y_m]$. Uma vez que apenas \mathbf{e} configura o componente probabilístico de \mathbf{y} , a matriz de covariâncias de \mathbf{y} corresponde à matriz de covariâncias de \mathbf{e} ; complementarmente, supõe-se que os termos exógenos são identicamente distribuídos como um ‘ruído branco’, isto é, com média zero, independentes entre si e com variância constante igual a $\sigma^2 > 0$. Então, tem-se

$$\text{Var}[\mathbf{y}] = \text{Var}[\mathbf{e}] = \Sigma_{\mathbf{e}} = \sigma^2 \mathbb{I}_m, \quad (2.7)$$

onde \mathbb{I}_m indica a matriz identidade de ordem m . Pode-se ainda substituir o modelo com ‘ruídos discretos’ dado em (2.6) por um que melhor usufrui das definições e propriedades dos processos estocásticos, ao tratar-se dos termos exógenos como variáveis funcionais $e(\cdot)$ indexadas em $[a, b]$, reformulando-o como segue

$$y_j = \chi(t_j) + e(t_j),$$

qual, reescrito no contexto vetorial, é tido

$$\mathbf{y} = \chi(\mathbf{t}) + e(\mathbf{t}).$$

Para manutenção das propriedades de ‘ruído branco’, como definidas em (2.7), a função de covariância $\Sigma(\cdot, \cdot)$ de um par de ruídos $e(t_j)$ e $e(t_k)$ é dada

$$\text{Cov}[e(t_j), e(t_k)] = \Sigma(t_j, t_k) = \sigma^2 \delta(t_j, t_k), \quad (2.8)$$

onde

$$\delta(t, s) = \begin{cases} 1, & \text{se } t = s \\ 0, & \text{se } t \neq s \end{cases}$$

Nota-se, porém, que a suposição de ‘ruído branco’ para a distribuição dos termos exógenos é vista como demasiadamente reducionista e simples, e há contextos nos quais a covariância entre instantes não pode ser ‘ignorada’. Adicionalmente, remete-se ao campo dos processos estocásticos, particularmente processos gaussianos, para refinamento da modelagem com respeito à função covariância, por exemplo, Rasmussen e Williams (2006, Cap. 4, (44)) definem uma função de covariância como uma função simétrica, tal que $\Sigma(t, s) = \Sigma(s, t)$, e positiva semi-definida; isto é, para toda função $f \in \mathcal{H}$ qualquer, tem-se que

$$\int_{\mathcal{T}} \int_{\mathcal{T}} f(t) f(s) \Sigma(t, s) dt ds \geq 0, \quad (2.9)$$

onde, no cenário em que (2.9) é estritamente superior à 0, denomina-se positiva definida. É dado também o seguinte exemplo de função de covariância

$$\text{Cov}[e(t_j), e(t_k)] = \Sigma(t_j, t_k) = \sigma^2 \exp \left\{ -\frac{\|t_j - t_k\|_2^2}{2} \right\}, \quad (2.10)$$

denominada ‘exponencial quadrática’, onde $\|\cdot\|_2$ remete à norma euclidiana. Naturalmente, neste contexto, tem-se que quão maior o intervalo entre os instantes, menor covariância há entre os ruídos, aproximando-se de zero em seus limites, embora nunca seja verdadeiramente nula. Os modelos de covariância em (2.8) e (2.10) ainda dispõem da conveniência de serem definidos por um parâmetro desconhecido σ^2 apenas. Há aplicações, como vistas por exemplo em Saraiva (2009, (27)), em quais são construídas funções de covariância assistidas por splines cúbicos, necessitando da estimação de uma quantidade maior de coeficientes.

2.4 FILTROS DE SUAUIZAÇÃO LINEAR

Operadores de suavização linear $H_j(t)$, também conhecidos como filtros de suavização linear, são utilizados como instrumentos para estimação de curvas como $\hat{\chi}(t)$ via combinações lineares, escrevendo

$$\hat{\chi}(t) = \sum_{j=1}^m H_j(t) y_j;$$

embora $H(t)$ esteja definido em todo \mathcal{T} , basta que seja avaliado em uma sequência finita de K instantes subsequentes³ $a \leq s_1 < s_2 < \dots < s_K \leq b$, tal que seja possível a construção da matriz $\mathbf{H} \in \mathbb{R}^{K \times m}$, onde $H_{k,j} = H_j(s_k)$. Dessa forma, pode-se escrever vetorialmente

$$\hat{\chi}(t) = \mathbf{H}\mathbf{y}. \quad (2.11)$$

Uma vez que a matriz \mathbf{H} trata-se de um operador linear, dado um par de constantes $d, e \in \mathbb{R}$ e um par de vetores $u, v \in \mathbb{R}^m$, é verdadeiro que

$$\mathbf{H}(d\mathbf{u} + e\mathbf{v}) = d\mathbf{H}\mathbf{u} + e\mathbf{H}\mathbf{v}.$$

Esta propriedade é particularmente útil por significativamente facilitar o cômputo numérico das estimativas $\hat{\chi}$, tão como ser fórmula de fácil manipulação em amplos contextos. Na seção seguinte é apresentada a representação de curvas via funções base, o que mais adiante será visto como forma de criação de filtros de suavização \mathbf{H} tal qual estes sejam projeções ortogonais, gozando de certas propriedades adicionais.

³ Estes instantes podem ser distintos àqueles em quais a variável funcional foi originalmente mensurada (\mathbf{t}).

2.5 REPRESENTAÇÃO DE CURVAS VIA FUNÇÕES BASE

Funções base são definidas por quaisquer classes de funções conhecidas quais desfrutam da propriedade de servirem para aproximação de outras funções através de combinações lineares das ‘bases’. Entre as classes mais utilizadas historicamente estão as classes de funções polinomiais, popularizadas mais significativamente pelo teorema originalmente proposto por Karl Weierstrass (1885, (45)) e complementado por Marshall Harvey Stone na primeira metade do século 20 (2007, (46)), qual diz

Teorema 1 (Stone-Weierstrass) *Se f é uma função real em $[a, b]$, existe uma sequência de polinômios p_k , tal que*

$$\lim_{k \rightarrow \infty} p_k(t) = f(t)$$

de maneira uniforme em $[a, b]$ ⁴ onde a sequência $\lim_{k \rightarrow \infty} p_k$ é definida como

$$1, t, t^2, \dots, t^k, \dots;$$

Entre outras sequências populares para aplicação, por exemplo, há a série de Fourier na qual, dada uma constante $\omega \in \mathbb{R}$, se escreve

$$1, \sin(\omega t), \cos(\omega t), \sin(2\omega t), \cos(2\omega t), \dots, \sin(k\omega t), \cos(k\omega t), \dots$$

Há ainda outras classes de funções, como ondaletas, exponenciais, potências e splines. Esta última servirá de foco em seções mais adiante.

Embora toda classe de funções base possua um número infinito de funções, na prática apenas as K primeiras são usadas para representação de uma curva $\chi(t)$ qualquer. Dado um vetor real de coeficientes desconhecidos $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \dots, \gamma_K]^\top$ e um vetor de funções base $\boldsymbol{\phi}(t) = [\phi_1(t), \phi_2, \dots, \phi_K(t)]^\top$, escreve-se

$$\chi(t) = \sum_{k=1}^K \gamma_k \phi_k(t),$$

ou vetorialmente

$$\chi(t) = \boldsymbol{\gamma}^\top \boldsymbol{\phi}(t) = \boldsymbol{\phi}(t)^\top \boldsymbol{\gamma}. \quad (2.12)$$

A partir deste instante, o principal foco torna-se a estimação dos parâmetros $\boldsymbol{\gamma}$ em (2.12). Na Estatística paramétrica tradicional, aplicam-se os conceitos de verossimilhança para estimação de parâmetros como $\boldsymbol{\gamma}$, enquanto na Estatística não-paramétrica utiliza-se de maneira mais geral a técnica de mínimos quadrados ordinários, vista a seguir.

⁴ Este teorema, retirado de Rudin (1976, p. 159, (47)), encontra-se aqui ligeiramente adaptado para melhor adequar-se ao contexto presente.

2.5.1 Estimação de curvas via mínimos quadrados

Como definido anteriormente em (2.6), as mensurações y_j da i -ésima realização de uma variável aleatória são constituídas de um componente sistemático e um componente probabilístico. O propósito da aplicação de filtros de suavização é a estimação do componente sistemático, e é natural que para isso busquem-se coeficientes $\boldsymbol{\gamma}$ que melhor se ajustem aos dados observados, por exemplo, tais que as distâncias $\|y_j - \boldsymbol{\gamma}^\top \boldsymbol{\phi}(t_j)\|$ sejam minimizadas para os m instantes em que foram obtidas medidas. Escreve-se a soma dos erros quadrados (SEQ) como

$$\text{SEQ}(\boldsymbol{\gamma}|\mathbf{y}) = \sum_{j=1}^m \left[y_j - \sum_{k=1}^K \gamma_k \phi_k(t_j) \right]^2$$

Se definida uma matriz $\Phi \in \mathbb{R}^{m \times K}$, onde $K \leq m$, tal que $\Phi_{j,k} = \phi_k(t_j)$ vetorialmente é possível escrever de maneira mais simples e direta

$$\text{SEQ}(\boldsymbol{\gamma}|\mathbf{y}_i) = (\mathbf{y} - \Phi\boldsymbol{\gamma})^\top (\mathbf{y} - \Phi\boldsymbol{\gamma}) \quad (2.13)$$

Disto, basta derivar (2.13) com respeito a $\boldsymbol{\gamma}$, obtendo

$$\frac{\partial \text{SEQ}(\boldsymbol{\gamma}|\mathbf{y})}{\partial \boldsymbol{\gamma}} = -2\Phi\mathbf{y} + 2\Phi^\top \Phi\boldsymbol{\gamma}. \quad (2.14)$$

Igualando a equação (2.14) a zero, trivialmente obtém-se o estimador de mínimos quadrados ordinários para $\boldsymbol{\gamma}$, qual é então

$$\hat{\boldsymbol{\gamma}}^{\text{MQ}} = (\Phi^\top \Phi)^{-1} \Phi^\top \mathbf{y}. \quad (2.15)$$

O estimador de mínimos quadrados ordinários em (2.15) é mais apropriado em contextos nos quais os erros inerentes aos dados observados obedecem às propriedades de ruído branco apresentadas anteriormente, ou seja, média zero, independência entre os termos e variância constante. Quando estas suposições aparentam ser violadas, podem ser utilizados estimadores de mínimos quadrados ponderados.

2.5.1.1 Mínimos quadrados ponderados

Para extensão da técnica de mínimos quadrados em situações nas quais algumas suposições são violadas, considere uma matriz $\mathbf{W} \in \mathbb{S}_+^m$, em que \mathbb{S}_+^m indica o conjunto de matrizes simétricas e positivas definidas de ordem m , escreve-se a soma dos erros quadrados ponderados (SEQP) como

$$\text{SEQP}(\boldsymbol{\gamma}|\mathbf{y}) = (\mathbf{y} - \Phi\boldsymbol{\gamma})^\top \mathbf{W} (\mathbf{y} - \Phi\boldsymbol{\gamma}) \quad (2.16)$$

Através do mesmo procedimento que exposto anteriormente, conclui-se que o estimador de mínimos quadrados ponderados obtido é

$$\hat{\boldsymbol{\gamma}}^{\text{MQPo}} = (\Phi^\top \mathbf{W} \Phi)^{-1} \Phi^\top \mathbf{W} \mathbf{y}. \quad (2.17)$$

Tradicionalmente, como esta técnica é implementada para remediar violações na variância dos dados, a matriz de ponderação utilizada é $\mathbf{W} = \hat{\Sigma}_{\mathbf{e}}^{-1}$, onde $\hat{\Sigma}_{\mathbf{e}}$ indica um estimador para a matriz de covariância dos dados.

2.5.1.2 Propriedades de projeção ortogonal

É possível reescrever (2.12) avaliando a função $\hat{\chi}(t)$ em cada instante \mathbf{t} em qual a curva foi mensurada como

$$\hat{\chi}(\mathbf{t}) = \Phi\hat{\boldsymbol{\gamma}}.$$

Se então os coeficientes forem substituídos pelo estimador de mínimos quadrados obtém-se

$$\hat{\chi}(\mathbf{t}) = \Phi(\Phi^{\top}\Phi)^{-1}\Phi^{\top}\mathbf{y}. \quad (2.18)$$

Define-se $\mathbf{H} = \Phi(\Phi^{\top}\Phi)^{-1}\Phi^{\top}$ onde⁵, conforme exposto em (2.11), $\hat{\mathbf{H}}$ trata-se de filtro de suavização linear. Remetendo à área adjacente de análise de regressão, a matriz \mathbf{H} é denominada matriz ‘chapéu’⁶, gozando não apenas da propriedade de linearidade como também das propriedades de uma projeção ortogonal. A saber, tem-se que

- \mathbf{H} é idempotente, isto é $\mathbf{H}^2 = \mathbf{H}$.
- \mathbf{H} é simétrica, isto é $\mathbf{H}^{\top} = \mathbf{H}$.
- O traço de \mathbf{H} é igual a K , isto é $\text{tr}(\mathbf{H}) = K$.
- O vetor de resíduos $\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{H}\mathbf{y}$ é ortogonal com respeito a $\mathbf{H}\mathbf{y}$, isto é, $\boldsymbol{\epsilon}^{\top}\mathbf{H}\mathbf{y} = 0$.⁷

A matriz chapéu e suas propriedades são consideradas no diagnóstico de medidas de influência, sendo também particularmente úteis para o cálculo dos graus de liberdade resultantes da suavização, o qual será abordado a seguir.

2.5.1.3 Graus de liberdade efetivos

Remetendo novamente à análise de regressão, entende-se que na grande maioria dos contextos experimentais, os graus de liberdade de um modelo são definidos como a diferença entre o número de dados observados e o número de parâmetros utilizados para estimação. NO contexto de estimação de curvas, subtrai-se m , o número de instantes em quais valores da curva foram mensurados, por K , o número de funções base utilizadas para aproximação, resultando em $m - K$ graus de liberdade para os resíduos.

⁵ Nota-se que no contexto de mínimos quadrados ponderados $\mathbf{H} = \Phi(\Phi^{\top}\mathbf{W}\Phi)^{-1}\Phi^{\top}\mathbf{W}$.

⁶ Traduzido do inglês ‘hat’.

⁷ Nota-se que no contexto do estimador de mínimos quadrados ponderados, esta ortogonalidade é verdadeira na forma $\boldsymbol{\epsilon}^{\top}\mathbf{W}\mathbf{H}\mathbf{y} = 0$.

Porém, conforme será visto em capítulos seguintes, a aplicação de métodos de penalização sob a estimação das curvas irá alterar o cálculo da quantidade de parâmetros do modelo, de forma que torna-se necessária uma formulação generalizada a este; utiliza-se então o seguinte

$$\text{Número de parâmetros} = \text{tr}(\mathbf{H}), \quad (2.19)$$

tal como visto anteriormente, nos contextos de mínimos quadrados ordinários e mínimos quadrados ponderados, é igual a K ; isto é, nestes âmbitos o cálculo de graus de liberdade efetivos mantém-se inalterados. Há ainda uma versão mais generalizada, como

$$\text{Número de parâmetros} = \text{tr}(\mathbf{H}\mathbf{H}^\top),$$

tal que, no contexto de mínimos quadrados ordinários e mínimos quadrados ponderados, resulta em K assim como a fórmula em (2.19), em razão das propriedades de simetria e idempotência da matriz chapéu.

2.5.1.4 Comentário sobre estimação de múltiplas curvas

Ressalta-se que até este ponto discutiu-se somente a técnica de estimação de curvas para apenas uma realização de uma variável aleatória funcional. Na área de análise de dados funcionais, dada uma amostra $\chi_1, \chi_2, \dots, \chi_n$ de dados funcionais, é necessária a estimação de n de curvas simultaneamente para estudo. Adicionalmente, pode-se supôr que cada realização χ_i foi mensurada em m_i instantes distintos entre cada realização, contexto em qual a amostra é denominada ‘desbalanceada’. Com isso, são definidos os vetores de valores mensurados em cada realização como $\mathbf{y}_i = [y_{i,1}, y_{i,2}, \dots, y_{i,m_i}]^\top$, com vetor de instantes de mensuração associados $\mathbf{t}_i = [t_{i,1}, t_{i,2}, \dots, t_{i,m_i}]^\top$, vetores de erros $\mathbf{e}_i = [e_{i,1}, e_{i,2}, \dots, e_{i,m_i}]^\top$ e coeficientes associados $\boldsymbol{\gamma}_i = [\gamma_{i,1}, \gamma_{i,2}, \dots, \gamma_{i,K}]^\top$.

Uma consequência direta deste desbalanceamento é o fato das funções bases implementadas variarem de realização a realização⁸, forçando então a construção de n distintas matrizes $\Phi_i \in \mathbb{R}^{m_i \times K}$, onde $K \leq m_i \forall i$, tal que $\Phi_{i,j,k} = \phi_k(t_{i,j})$. Uma forma para estimação simultânea das múltiplas curvas seria escrever

$$\begin{array}{c} \text{(a)} \\ \left[\begin{array}{c} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_{n-1} \\ \mathbf{y}_n \end{array} \right] = \begin{array}{c} \text{(b)} \\ \left[\begin{array}{cccccc} \Phi_1 & \mathbf{0}_{m_1,K} & \mathbf{0}_{m_1,K} & \cdots & \mathbf{0}_{m_1,K} \\ \mathbf{0}_{m_2,K} & \Phi_2 & \mathbf{0}_{m_2,K} & \cdots & \mathbf{0}_{m_2,K} \\ \mathbf{0}_{m_3,K} & \mathbf{0}_{m_3,K} & \Phi_3 & \cdots & \mathbf{0}_{m_3,K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{m_n,K} & \mathbf{0}_{m_n,K} & \mathbf{0}_{m_n,K} & \cdots & \Phi_n \end{array} \right] \end{array} \begin{array}{c} \text{(c)} \\ \left[\begin{array}{c} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \\ \vdots \\ \boldsymbol{\gamma}_{K-1} \\ \boldsymbol{\gamma}_K \end{array} \right] + \begin{array}{c} \text{(d)} \\ \left[\begin{array}{c} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_{n-1} \\ \mathbf{e}_n \end{array} \right], \end{array} \quad (2.20) \end{array}$$

⁸ Nota-se também que mesmo em casos que a quantia de instantes de mensuração é igual para todas realizações, isto é, $m_i = m_j, \forall (i, j)$, a amostra pode ser dita ‘desbalanceada’ desde que os instantes de mensuração variem entre pelo menos duas realizações, isto é, $\mathbf{t}_i \neq \mathbf{t}_j$ para algum (i, j) .

em que o vetor coluna em (a) indica o vetor resultado da concatenação dos valores mensurados em cada realização, de comprimento $M = \sum_{i=1}^n m_i$, o vetor coluna em (d) indica o vetor resultado da concatenação dos erros de mensuração de cada realização, de comprimento $M = \sum_{i=1}^n m_i$, o vetor coluna em (c) indica o vetor resultado da concatenação dos coeficientes dos modelos de cada curva, de comprimento nK , a matriz em (b) indica uma matriz em blocos cujos blocos diagonais referem-se às funções base do modelo de cada curva, de dimensões $M \times nK$ e $\mathbf{0}_{p,q}$ indica uma matriz composta de zeros de dimensões $p \times q$. Trivialmente, a soma dos quadrados dos erros de (2.20) escreve-se

$$\text{SEQ}(\boldsymbol{\gamma}^* | \mathbf{y}^*) = \sum_{i=1}^n [\mathbf{y}_i - \Phi_i \boldsymbol{\gamma}_i]^\top [\mathbf{y}_i - \Phi_i \boldsymbol{\gamma}_i]$$

onde \mathbf{y}^* refere-se ao vetor em (a) na equação (2.20), $\boldsymbol{\gamma}^*$ o vetor (c) na mesma equação. Repetem-se os passos anteriores para o cálculo do estimador de mínimos quadrados, obtendo-se

$$\{\hat{\boldsymbol{\gamma}}_i = (\Phi_i^\top \Phi_i)^{-1} \Phi_i^\top \mathbf{y}_i\}_{\forall i \in \{1, 2, \dots, n\}}. \quad (2.21)$$

Em conclusão, encontrar o valor ideal de $\boldsymbol{\gamma}^*$ em (2.22) resume-se a solucionar n problemas de mínimos quadrados ordinários simultaneamente. Por conveniência e simplicidade de notação, adota-se aqui, assim como no pacote `fda` desenvolvido para linguagem R, usado extensivamente para confecção do presente trabalho, que todas amostras são ‘balanceadas’, isto é, que todas medições de todas realizações foram obtidas no mesmo instante $\mathbf{t}_i = \mathbf{t}_j \forall (i, j)$. Sob esta suposição, reescreve-se

$$\mathbf{y}^* = (\mathbb{I}_n \otimes \Phi) \boldsymbol{\gamma}^* + \mathbf{e}^* \quad (2.22)$$

onde $\cdot \otimes \cdot$ refere-se o produto de Kronecker entre \cdot e \cdot , e \mathbf{e}^* refere-se ao vetor em (d) na equação (2.20). Neste contexto, repete-se o processo via mínimos quadrados para estimação de $\boldsymbol{\gamma}^*$, com soma de erros quadrados definida por

$$\text{SEQ}(\boldsymbol{\gamma}^* | \mathbf{y}^*) = [\mathbf{y}^* - (\mathbb{I}_n \otimes \Phi) \boldsymbol{\gamma}^*]^\top [\mathbf{y}^* - (\mathbb{I}_n \otimes \Phi) \boldsymbol{\gamma}^*],$$

onde repetindo-se os passos utilizados anteriormente, concluí-se que o estimador de mínimos quadrados para múltiplas curvas obtido é

$$\hat{\boldsymbol{\gamma}}^{*:\text{MQ}} = (\mathbb{I}_n \otimes [(\Phi^\top \Phi)^{-1} \Phi^\top]) \mathbf{y}^*. \quad (2.23)$$

Nota-se então que, advindo de propriedades do produto de Kronecker, o resultado em (2.23) é equivalente ao seguinte

$$\{\hat{\boldsymbol{\gamma}}_i = (\Phi^\top \Phi)^{-1} \Phi^\top \mathbf{y}_i\}_{\forall i \in \{1, 2, \dots, n\}},$$

o resultado acima remete àquele dado em (2.21), onde observa-se que a estimação de múltiplas curvas simultaneamente, usualmente, consiste na solução de múltiplos problemas de mínimos quadrados ordinários independentes.

3 CLASSES DE FUNÇÃO BASE

3.1 FUNÇÕES POLINOMIAIS

Dado um conjunto de $k + 1$ constantes $\{\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_k\} \in \mathbb{R}^{k+1}$, escreve-se uma função p_k qualquer definida em $[a, b]$ e pertencente à classe π^k , qual diz respeito às funções polinomiais univariadas de grau no máximo k , como

$$p_k(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_k t^k$$

ou ainda, de maneira mais geral, é considerado um coeficiente adicional $t_0 \in \mathbb{R}$, tal que

$$p_k(t) = \alpha_0 + \alpha_1(t - t_0) + \alpha_2(t - t_0)^2 + \dots + \alpha_k(t - t_0)^k. \quad (3.1)$$

Utilizando da notação de somatório, pode-se escrever

$$p_k(t) = \sum_{j=0}^k \alpha_j (t - t_0)^j,$$

onde a q -ésima derivada de p_k com respeito a t é dada por

$$D^q p_k(t) = \begin{cases} \sum_{j=q}^k \frac{j!}{(j-q)!} \alpha_j (t - t_0)^{j-q}, & \text{se } q \leq k \\ 0, & \text{caso contrário} \end{cases}$$

e a integral indefinida de p_k é dada por

$$\int p_k(t) dt \propto \sum_{j=0}^k \frac{\alpha_j}{j+1} (t - t_0)^{j+1}. \quad (3.2)$$

Nota-se que as classes de funções polinomiais são definidas de tal forma que a $\pi \subset \pi^2 \subset \dots \subset \pi^{k-1} \subset \pi^k \subset \pi^{k+1} \subset \dots \subset C^\infty(\mathbb{R})$, onde $C^\infty(\mathbb{R})$ indica a classe de funções univariadas que possuem infinitas derivadas contínuas. Assim, é verdadeiro que $D^m f_k \in \pi^k$ e $\int p_k(t) dt \in \pi^{k+1}$. É observável que funções polinomiais são facilmente manipuláveis no contexto de cálculo diferencial e integral, ademais, a utilidade destas para aproximação é evidenciada desde o Teorema de Stone-Weierstrass supracitado. No entanto, observa-se que a aproximação dada no teorema é também dependente da seleção adequada de coeficientes α para o polinômio.

Dada uma função $f \in C^\infty(\mathbb{R})$ qualquer, f é dita analítica¹ em t_0 , e pertence à classe de funções analíticas univariadas $C^\omega(\mathbb{R})$ se, e somente, dado

$$\phi(t) = \sum_{j=0}^{\infty} \frac{D^j f(t_0)}{j!} (t - t_0)^j \quad (3.3)$$

¹ Pertencem à classe de funções analíticas as próprias funções polinomiais, funções trigonométricas, funções exponenciais, entre outras.

tem-se que $\phi(t)$ converge em $f(t)$ na vizinhança de t_0 . A série em (3.3) é denominada ‘Série de Taylor’², em homenagem ao matemático inglês Brook Taylor, qual a concebeu nos meados do século 18. Observa-se, portanto, que as Séries de Taylor são um caso especial de $\lim_{k \rightarrow \infty} p_k$, onde os coeficientes são definidos como $\alpha_j = \frac{D^j f(t_0)}{j!}$. É possível truncar ϕ em apenas seus k primeiros termos, em qual caso obtém-se $\phi_k(t) \in \pi^k$, qual denomina-se ‘polinômio de Taylor de grau k ’.

A Figura 3.1 exhibe os resultados da aproximação da função $f(t) = \sin(\pi t)$ via polinômios de Taylor de graus até $k = 7$, utilizando-se como centro da série $t_0 = 0$. Entende-se que as derivadas de $\sin(\pi t)$ obedecem à seguinte relação

$$D^q \sin(\pi t) = \begin{cases} -\pi^q \sin(\pi t) & \text{se } q \mid 2 \text{ e } q \nmid 4. \\ \pi^q \sin(\pi t) & \text{se } q \mid 4. \\ -\pi^q \cos(\pi t) & \text{se } (q-1) \mid 2 \text{ e } (q-1) \nmid 4. \\ \pi^q \cos(\pi t) & \text{se } (q-1) \mid 4. \end{cases}$$

onde $x \mid y$ indica que o resto da divisão de x por y é zero, isto é, x é divisível por y , enquanto $x \nmid y$ indica que x não é divisível por y . É observável na Figura 3.1 que, naturalmente, o aumento no grau do polinômio de Taylor melhora a precisão da aproximação de $f(t)$, embora conforme os valores de t se distanciam da vizinhança t_0 as divergências entre a função real e as aproximações tornem-se mais pronunciadas. Mais adiante na Seção 3.3 será apresentada uma proposta alternativa para aproximação funcional, o método de funções ‘splines’, construídas a partir de combinações lineares de segmentos de polinômios.

3.2 INTERPOLAÇÃO E APROXIMAÇÃO

A interpolação de dados trata-se de uma área de conhecimento adjacente à de aproximação funcional, que forneceu grande parte dos fundamentos para seu desenvolvimento. Um problema comum de interpolação é dado como segue: dado um conjunto de n pares ordenados $(\xi_1, y_1), (\xi_2, y_2), \dots, (\xi_m, y_m)$, onde $\xi_j < \xi_{j+1} \forall j \in \{1, 2, \dots, m\}$, busca-se uma função real $f \in C^\infty(\mathbb{R})$ tal que $f(\xi_j) = y_j \forall j \in \{1, 2, \dots, m\}$. Um método de solucionar este problema foi proposto por Isaac Newton entre o final do século 17 e os meados do século 18, via as ‘diferenças divididas’: dada uma função real qualquer f , suas diferenças divididas de primeira ordem, são definidas como

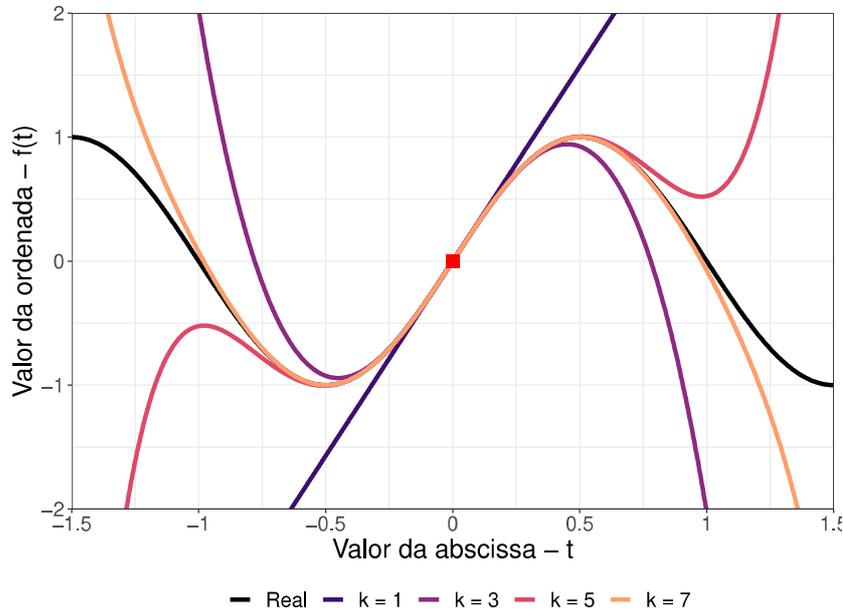
$$[\Delta f](\xi_j, \xi_{j+1}) = \frac{f(\xi_{j+1}) - f(\xi_j)}{\xi_{j+1} - \xi_j}. \quad (3.4)$$

De forma geral, as diferenças divididas de ordem $k \geq 1$, onde $k \in \mathbb{N}$, são escritas

$$[\Delta^k f](\xi_j, \xi_{j+1}, \dots, \xi_{j+k}) = \frac{[\Delta^{k-1} f](\xi_{j+1}, \dots, \xi_{j+k}) - [\Delta^{k-1} f](\xi_j, \dots, \xi_{j+k-1})}{\xi_{j+k} - \xi_j}, \quad (3.5)$$

² Similarmente, se fixado $t_0 = 0$, obtém-se uma ‘Série de Maclaurin’, também concebida no século 18, pelo matemático Colin Maclaurin.

Figura 3.1 - Aproximação de $f(t) = \sin(\pi t)$ via polinômios de Taylor de graus $k = 1$ até $k = 7$, com $t_0 = 0$. O quadrado colorido em vermelho destaca o ponto $(0, 0)$.



Fonte: Elaborado pelo autor.

onde o operador Δ^k é denominado ‘operador de ordem k de diferenças divididas’³. Observa-se que os operadores de diferenças divididas de graus superiores são construídos recursivamente a partir de operadores de diferenças divididas de graus inferiores. A Tabela 1 apresenta dados⁴ de um estudo de Glover (1910, (48)), qual buscava estimar a tabela de mortalidade nos Estados Unidos da América via ‘interpolação osculatória’.

Um resultado relevante com respeito às diferenças divididas, proposto pelo autor em Schwarz (1882, (49)) e aqui extraído de de Boor (2005, (50)), é dado como segue.

Corolário 1 (Schwarz) *Dada uma função $f \in C^k$ definida em $[a, b]$, com uma sequência $a \leq \xi_1 < \xi_2 < \dots < \xi_k < \xi_{k+1} \leq b$ de valores reais, existe $\zeta \in [\xi_1, \xi_{k+1}]$, tal que*

$$k![\Delta^k f](\xi_1, \xi_2, \dots, \xi_k, \xi_{k+1}) = D^k f(\zeta) \quad (3.6)$$

é equivalentemente verdadeiro que existe $\zeta \in [\xi_1, \xi_{k+1}]$, tal que

$$[\Delta^k f](\xi_1, \xi_2, \dots, \xi_k, \xi_{k+1}) = \frac{D^k f(\zeta)}{k!}. \quad (3.7)$$

³ O operador de ordem 0 é definido simplesmente como $[\Delta^0 f](t_j) = f(t_j)$.

⁴ Na Tabela 1–(b), embora seja utilizado o operador Δ^k , neste o autor refere-se às ‘diferenças simples’, isto é, não divididas.

Tabela 1 – (a) Tabela de idade x comparada ao total de habitantes com pelo menos x anos de idade T_x , no dia primeiro de Junho de 1900 nos Estados Unidos da América; (b) Diferenças simples de ordem de um a cinco entre T_x para cada x .

AGE.	POPULATION JUNE 1, 1900.	AGE.	POPULATION JUNE 1, 1900.
x	T_x	x	T_x
5	8,916,354	55	1,101,913
10	7,918,070	60	760,116
15	7,006,110	65	484,593
20	6,124,326	70	283,469
25	5,208,766	75	142,764
30	4,296,932	80	57,738
35	3,464,973	85	17,382
40	2,708,922	90	3,678
45	2,063,477	95	681
50	1,546,064	100	130

(a)

x	ΔT_x	$\Delta^2 T_x$	$\Delta^3 T_x$	$\Delta^4 T_x$	$\Delta^5 T_x$
5	-998,284	+ 86,324	-56,148	- 7,804	+109,258
10	-911,960	+ 30,176	-63,952	+101,454	- 62,807
15	-881,784	- 33,776	+37,502	+ 38,647	-118,763
20	-915,560	+ 3,726	+76,149	- 80,116	+118,781
25	-911,834	+ 79,875	- 3,967	+ 38,665	- 55,937
30	-831,959	+ 75,908	+34,698	- 17,272	- 54,924
35	-756,051	+110,606	+17,426	- 72,196	+156,058
40	-645,445	+128,032	-54,770	+ 83,862	-149,034
45	-517,413	+ 73,262	+29,092	- 65,172	+109,377
50	-444,151	+102,354	-36,080	+ 44,205	- 66,310
55	-341,797	+ 66,274	+ 8,125	- 22,105	+ 31,345
60	-275,523	+ 74,399	-13,980	+ 9,240	- 15,509
65	-201,124	+ 60,419	- 4,740	- 6,269	- 740
70	-140,705	+ 55,679	-11,009	- 7,009	+ 9,082
75	- 85,026	+ 44,670	-18,018	+ 2,073	+ 5,611
80	- 40,356	+ 26,652	-15,945	+ 7,684	
85	- 13,704	+ 10,707	- 8,261		
90	- 2,997	+ 2,446			
95	- 551				

(b)

Fonte: Extraído e adaptado de Glover (1910, pp. 10–11, (48)).

Em suma, a partir do Corolário de Schwarz⁵ é possível concluir-se que as diferenças divididas de ordem k de uma função f podem ser utilizadas para aproximação da m -ésima derivada da mesma função. Este resultado outrora foi particularmente útil por tratar-se de uma fórmula simples e de baixo custo computacional para cômputo aproximado de derivadas.

⁵ É possível identificar a semelhança entre este corolário proposto por Schwarz e o ‘Teorema do Valor Médio’ comumente aplicado na área da análise matemática, visto também em Rudin (1976, pp. 107–108, (47)).

A partir dos operadores de diferenças divididas é possível construir

$$\begin{aligned}
f(t) = & f(t_1) + (t - \xi_1)[\Delta f](\xi_1, \xi_2) \\
& + (t - \xi_1)(t - \xi_2)[\Delta^2 f](\xi_1, \xi_2, \xi_3) \\
& + (t - \xi_1)(t - \xi_2)(t - \xi_3)[\Delta^3 f](\xi_1, \xi_2, \xi_3, \xi_4) \quad ; \quad (3.8) \\
& + \dots \\
& + (t - \xi_1)(t - \xi_2) \dots (\xi - \xi_{m-1})[\Delta^{m-1} f](\xi_1, \xi_2, \dots, \xi_m).
\end{aligned}$$

Neste caso, têm-se que (3.8) é a solução proposta por Netwon ao problema tradicional de interpolação apresentado anteriormente. Observa-se, no entanto que esta solução apresenta uma limitação: o grau do polinômio interpolador $f(t)$ será, na grande maioria dos casos, igual a $n - 1$. Ademais, em Runge⁶ (1901, (52)), o autor aponta a existência de funções para quais o aumento no número de pares ordenados sob qual deve ser efetuada a interpolação resulta na divergência entre a função interpoladora e a função real a ser interpolada. Este fenômeno é conhecido como ‘Fenômeno de Runge’, e é particularmente estudado no contexto em que as abscissas t_i são equidistantes entre si. Percebe-se nestes contextos um padrão excessivamente oscilatório na função interpoladora, particularmente em seus limites, o que ocasiona a divergência.

Em Whittaker (1922, (53)), o autor ressalta outra consideração pertinente: o objetivo da interpolação é definir funções que se ajustam ‘perfeitamente’ aos dados mensurados, no entanto, como visto no modelo em (2.6), tradicionalmente faz-se a suposição que as medidas resultam da soma de um valor determinístico a um valor exógeno, o ruído. Dessa forma, a interpolação de todos dados não irá necessariamente refletir os valores da curva real, e ademais possivelmente irá incorrer em uma curva pouco suave. A solução dada pelo autor é denominada ‘graduação’⁷ das observações, embora contemporaneamente se denomine ‘regularização’, e resume-se a encontrar uma função f , tal que

$$\min_{f \in C^k} \left\{ \overbrace{\sum_{j=1}^m (f(t_j) - y_j)^2}^{(a)} + \lambda \overbrace{\sum_{j=1}^{m-k} \left([\Delta^k f](\xi_j, \xi_{j+1}, \dots, \xi_{j+k}) \right)^2}^{(b)} \right\}. \quad (3.9)$$

O formato em (3.9) foi definido posteriormente em Schoenberg (1964, (54)). O componente destacado em (a) define o problema de interpolação tradicional, enquanto o componente destacado em (b) caracteriza a regularização aplicada à interpolação, onde $\lambda > 0$ é denominado ‘constante de regularização’. Quando $\lambda = 0$ o problema em (3.9) se resume à interpolação. Neste mesmo trabalho, o autor veio então a demonstrar que a classe de funções splines, tópico da seguinte Seção, é uma solução generalizada para o problema dado em (3.9).

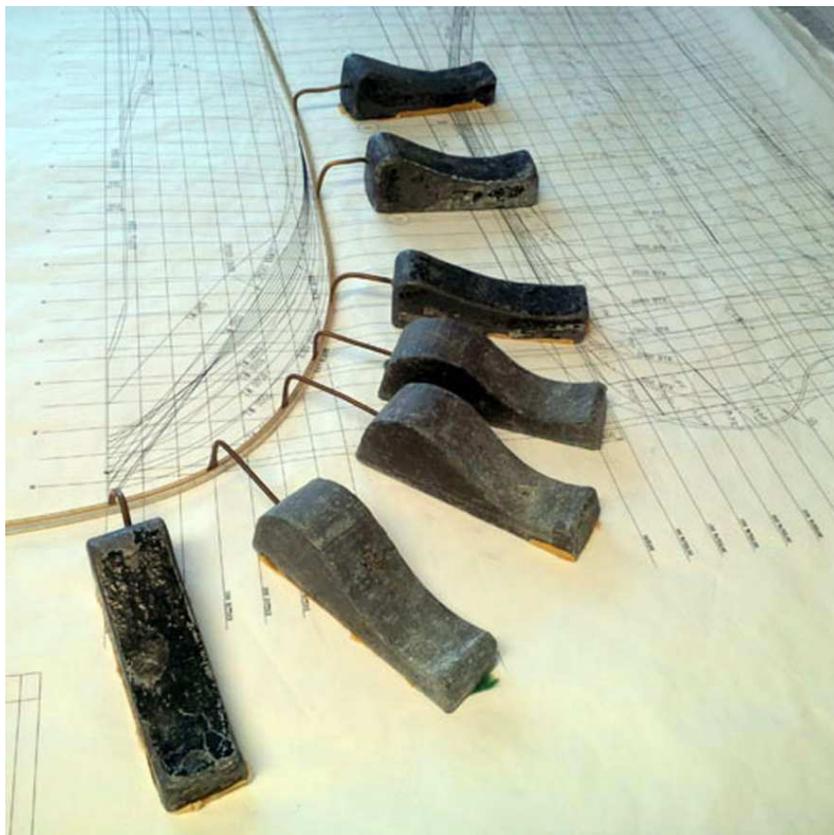
⁶ Notavelmente, de acordo com Hockey (2007, pp. 992–993, (51)), Carl Runge estudou em Berlim com Karl Weierstrass.

⁷ Traduzido do inglês ‘*graduation*’.

3.3 SPLINES

O nome ‘spline’ foi originalmente concebido em Schoenberg (1946, (55, 56)), em qual o autor notou que, sob certas condições, as funções splines se assemelham às tiras flexíveis de metal ou madeira utilizadas na confecção de curvas suaves em desenhos técnicos, denominadas ‘splines’. A Figura 3.2 exhibe o instrumento spline sendo utilizado para a confecção de uma curva qualquer.

Figura 3.2 - Spline sob uso para desenho técnico; nota-se que a spline propriamente dita trata-se apenas da tira sendo curvada.



Fonte: Extraído de Noe (2016, (57)).

Em Wahba (1990, (58)), a autora apresenta uma redefinição ao problema de graduação proposto anteriormente em (3.9), como

$$\min_{f \in C^k} \left\{ \sum_{j=1}^m (f(t_j) - y_j)^2 + \lambda \int_a^b [D^k f(t)]^2 dt \right\}. \quad (3.10)$$

Em decorrência do Corolário de Schwarz é demonstrável a equivalência aproximada entre (3.9) e (3.10). A autora então define o spline polinomial natural e univariado $S(t)$ de ordem K qual aproxima uma função $f(t)$ qualquer definida no intervalo $\mathcal{T} = [a, b]$, como uma

função real fixada através de m nós⁸ $\xi_1, \xi_2, \dots, \xi_m$, tal que $a \leq \xi_1 < \xi_2 < \dots < \xi_m \leq b$, obedientes às seguintes propriedades:

Prop. (i) $S \in \pi^{K-1}$ quando $t \in [a, \xi_1] \cup [\xi_m, b]$ ⁹;

Prop. (ii) $S \in \pi^{2K-1}$ quando $t \in \left[\bigcup_{j=1}^{m-1} [\xi_j, \xi_{j+1}] \right]$;

Prop. (iii) $S \in C^{2K-2}(\mathbb{R})$.

O termo ‘natural’¹⁰, quando utilizado para caracterizar uma função spline, refere-se à conformidade desta à condição de fronteira de Neumann, comumente abordada na área de equações diferenciais, que determina

$$D^j S(\xi_1) = D^j S(\xi_m) = 0, \quad \forall j \geq K.$$

Supondo-se que t é unidimensional, a spline natural é então definida pela soma de $m + 1$ segmentos de polinômios $s_i(t)$ delimitados pela sequência de nós $a \leq \xi_1 < \xi_2 < \dots < \xi_m \leq b$, qual escreve-se

$$S(t) = s_1(t)\mathbf{1}_{(-\infty, \xi_1]}(t) + s_2(t)\mathbf{1}_{(\xi_1, \xi_2]}(t) + \dots + s_m(t)\mathbf{1}_{(\xi_{m-1}, \xi_m]}(t) + s_{m+1}(t)\mathbf{1}_{(\xi_m, \infty)}(t),$$

onde $\mathbf{1}_{\mathcal{A}}(t)$ representa a função indicadora do elemento t com respeito ao conjunto \mathcal{A} . É necessário definir os segmentos $s_i(t)$, tal que a continuidade da spline seja garantida, isto é

$$\lim_{t \rightarrow \xi_i} S(t) = s_j(\xi_j) = s_{j+1}(\xi_j), \quad \forall j \in \{1, 2, \dots, m\}.$$

Como percebe-se nas proposições acima, a continuidade deve ser mantida até nas derivadas de ordem $2K - 2$, e conseqüentemente deve-se garantir também que

$$\lim_{t \rightarrow \xi_j} D^q S(t) = D^q s_j(\xi_j) = D^q s_{i+j}(\xi_j), \quad \forall j \in \{1, 2, \dots, m\} \text{ e } \forall q \in \{1, 2, \dots, 2K - 2\}.$$

Observa-se que, em toda a literatura acerca o tópico de splines, os mais comumente implementados são os splines lineares e os splines cúbicos¹¹, de ordens $K = 1$ e $K = 2$, respectivamente. Cada qual destes apresenta suas particularidades para utilização: por exemplo, os splines cúbicos, embora mais flexíveis e capazes de melhor refletir o comportamento de funções não lineares, ocasionalmente apresentam irregularidades antes do nó ξ_1 e após o nó ξ_m ou seja, nas ‘caudas’, quais são tradicionalmente mitigadas em splines naturais. É necessário então destacar a distinção entre splines ‘naturais’ e splines

⁸ Traduzido do inglês ‘*knots*’.

⁹ Vale notar que esta Propriedade não é válida no contexto de splines lineares.

¹⁰ Em Quintela (2020, (59)), entre outras referências, o autor alternativamente se refere às splines naturais como ‘splines restritas’.

¹¹ Observa-se que, para toda ordem K , os splines naturais serão polinômios de graus ímpares; de Boor (1968, (60)) apresenta resultados relevantes quanto este fato.

‘irrestritas’: além de não obrigatoriamente obedecerem à condição de fronteira de Neumann, as splines irrestritas não necessariamente condizem à Propriedade (i) citada acima.

Uma fórmula usual para funções splines cúbicos irrestritas, isto é, de ordem $m = 2$, como vista em Hastie e Tibshirani (1990, (61)) escreve-se

$$S(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 + \sum_{j=1}^m \theta_j (t - \xi_j)_+^3; \quad (3.11)$$

onde $\boldsymbol{\alpha} \in \mathbb{R}^4$ e $\boldsymbol{\theta} \in \mathbb{R}^m$ são vetores de coeficientes desconhecidos, e $(t - a)_+$ é denominada ‘função da parte positiva’¹² de $(t - a)$, tal que

$$(t - a)_+ = \begin{cases} (t - a), & \text{se } t - a \geq 0 \\ 0, & \text{se } t - a < 0 \end{cases}.$$

A Figura 3.3 exibe um exemplo da aplicação de splines para aproximação de uma função real, onde foram gerados 30 dados conforme a função

$$f(t_j) = \begin{cases} \sin(\pi t_j) + e_j, & \text{se } t \geq 0 \\ \pi t_j + e_j, & \text{se } t < 0 \end{cases}, \quad (3.12)$$

tal que $e_j \sim \mathcal{N}(0, 1/4)$ e $t_j \sim \mathcal{U}(-2, 2)$. Os nós selecionados foram $\xi_1 = -1/2$, $\xi_2 = 1/2$ e $\xi_3 = 3/2$. Nesta aplicação, já faz-se observável a divergência entre a função real e a aproximação via uma função spline cúbica irrestrita em valores quais antecedem o primeiro nó ou procedem o último nó. A aproximação via spline linear irrestrita, embora não apresente esta mesma divergência em ambas caudas, é incapaz de adequadamente refletir o comportamento do segmento trigonométrico da função.

Há ainda outras ferramentas para geração de funções splines além da fórmula em (3.12), uma das quais são as classes de funções base B-splines, abordadas por de Boor (1978, (62)), quais são tópicos da seção seguinte.

3.4 B-SPLINES

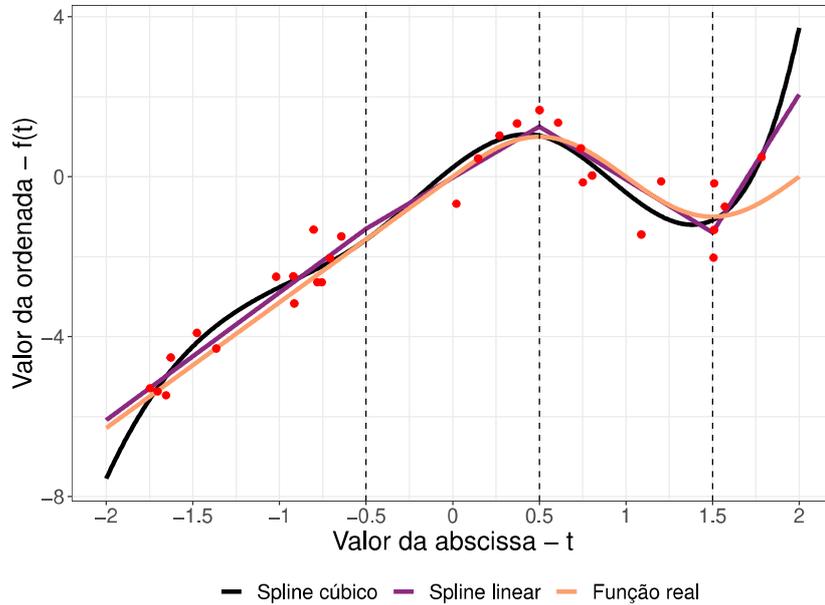
Funções B-splines foram originalmente propostas por Curry e Schoenberg (1947, (63)), onde os autores definiram, dado uma sequência de nós $a \leq \xi_1 < \xi_2 < \dots < \xi_m \leq b$, o j -ésimo B-spline de ordem k , com $k \geq 2$,¹³ como as diferenças divididas de k -ésima ordem da função da parte positiva de $(\star - t)$ elevada a $k - 1$, isto é

$$\mathbf{B}_j^{(k)}(t) = (\xi_{j+k} - \xi_j) \left[\Delta^k (\star - t)_+^{k-1} \right] (\xi_j, \xi_{j+1}, \dots, \xi_{j+k}), \quad (3.13)$$

¹² Esta função é denominada também em algumas fontes, como de Boor (1978, (62)), como ‘função de potência truncada’.

¹³ O j -ésimo B-spline de ordem $k = 1$ é definido trivialmente como $\mathbf{B}_j^{(1)}(t) = \mathbf{1}_{[\xi_j, \xi_{j+1})}(t)$.

Figura 3.3 - Aproximações via splines lineares e cúbicas para dados gerados para simulação da função (3.12). Os pontos coloridos em vermelho indicam os valores mensurados.



Fonte: Elaborado pelo autor.

onde a notação $(\star - t)_+$ foi utilizada para denotar que as diferenças divididas incidem sob o termo em \star , enquanto o t mantém-se fixo. Por exemplo, tem-se que o j -ésimo B-spline de ordem $k = 2$ escreve-se

$$\mathbf{B}_j^{(2)}(t) = \frac{(\xi_{j+2} - t)_+ - (\xi_{j+1} - t)_+}{\xi_{j+2} - \xi_{j+1}} - \frac{(\xi_{j+1} - t)_+ - (\xi_j - t)_+}{\xi_{j+1} - \xi_j}.$$

A partir da fórmula em (3.13), é possível obter-se uma fórmula recursiva para a construção de B-splines de ordem k qualquer, ainda utilizada contemporaneamente, por exemplo em Silva *et al* (2022, (64)), Galindo (2021, (65)), Wood (2006, (66)) e Eilers e Marx (1996, (67)), qual se escreve

$$\mathbf{B}_j^{(k)}(t) = \begin{cases} \omega_j^{(k)}(t)\mathbf{B}_j^{(k-1)}(t) + [1 - \omega_{j+1}^{(k)}(t)]\mathbf{B}_{j+1}^{(k-1)}(t) & \text{se } k > 1 \\ \mathbf{1}_{[\xi_j, \xi_{j+1})}(t) & \text{se } k = 1 \end{cases}, \quad (3.14)$$

onde $\omega_j^{(k)} = (t - \xi_j) / (\xi_{j+k-1} - \xi_j)$. Similarmente à forma acima, a q -ésima derivada do j -ésimo B-spline de ordem k , onde $q < k$, escreve-se recursivamente

$$D^q \mathbf{B}_j^{(k)}(t) = (k - 1) \left[\frac{D^{q-1} \mathbf{B}_j^{(k-1)}(t)}{\xi_{j+k-1} - \xi_j} - \frac{D^{q-1} \mathbf{B}_{j-1}^{(k-1)}(t)}{\xi_{j+k} - \xi_{j+1}} \right]. \quad (3.15)$$

Vale notar que, dado um número fixo de nós $\xi_1, \xi_2, \dots, \xi_m$, a quantia de funções B-splines de ordem k utilizadas para a aproximação de uma função qualquer é $m - k$. Ademais, em de Boor (1978, (62)), o autor enuncia ainda as seguintes propriedades de B-splines:

Prop. (i) $\mathbf{B}_j^{(k)}(t) = 0$ quando $t \notin (\xi_j, \xi_{j+k})$, e $\mathbf{B}_j^{(k)}(t) > 0$ caso contrário;

Prop. (ii) $\sum_{i=j+1}^{j+k-1} \mathbf{B}_j^{(k)}(\xi_i) = 1$ se todos os nós interiores ao j -ésimo B-spline são equidistantes;

Prop. (iii) $\mathbf{B}_j^{(k)} \in \pi^{k-1}$;

Prop. (iv) $\mathbf{B}_j^{(k)} \in C^{k-1}(\mathbb{R}), \forall j$.

A propriedade (i) enunciada acima é denominada ‘propriedade de positividade e suporte compacto’, e reflete o princípio qual diz que o j -ésimo B-spline de ordem k depende apenas do conjunto de $k + 1$ nós $\xi_j, \xi_{j+1}, \dots, \xi_{j+k}$. Ademais, denomina-se os nós interiores ao suporte, tal que $\xi_i \in (\xi_j, \xi_{j+k}) = \text{int}[\xi_j, \xi_{j+k}]$ de ‘nós interiores ao j -ésimo B-spline’. A propriedade (ii) é denominada ‘propriedade da partição da unidade’. Observa-se, como denotado pelos autores em Eilers e Marx (1996, (67)), que em razão das propriedades supracitadas, funções B-splines são flexíveis, e podem facilmente ser manipulados de forma a melhor localmente se adequarem aos dados.

A Figura 3.4-(a) apresenta as funções B-splines de ordens 2 e 3 com cinco nós equidistantes entre 0 e 4, enquanto a Figura 3.4-(b) apresenta as funções B-splines de ordens 2 e 3 com quatro nós não equidistantes entre 0 e 4. Observa-se que na Figura 3.4-(a) a propriedade (ii) vista acima é preservada, por exemplo, uma vez que $\mathbf{B}_1^{(3)}(\xi_2) + \mathbf{B}_1^{(3)}(\xi_2) = 1/2 + 1/2 = 1$, enquanto na Figura 3.4-(b), o B-spline $\mathbf{B}_1^{(3)}$ possui nós interiores não equidistantes, tal que $\mathbf{B}_1^{(3)}(\xi_2) + \mathbf{B}_1^{(3)}(\xi_3) = 1/2 + 2/3 \neq 1$. Adicionalmente, percebe-se que, para ambas Figuras, o B-spline $\mathbf{B}_1^{(2)}$ é igual, isto ocorre uma vez que os três primeiros nós são iguais em ambos casos, demonstrando assim o princípio que o j -ésimo B-spline depende apenas dos $k + 1$ nós definidos por $\xi_j, \xi_{j+1}, \dots, \xi_{j+k}$.

Em suma, é possível construir uma estimativa suave de uma curva $\chi(t)$ qualquer através de B-splines de ordem k definidos pelos nós $\xi_1, \xi_2, \dots, \xi_m$ ao se escrever a seguinte combinação linear

$$\hat{\chi}(t) = \sum_{j=1}^{m-k} \gamma_j \mathbf{B}_j^{(k)}(t).$$

Ao considerar-se um vetor $\mathbf{B}^{(k)} = [\mathbf{B}_1^{(k)}, \mathbf{B}_2^{(k)}, \dots, \mathbf{B}_{m-k}^{(k)}]^\top$, faz-se possível escrever na forma vetorial, equivalente àquela vista em (2.12)

$$\hat{\chi}(t) = \boldsymbol{\gamma}^\top \mathbf{B}^{(k)} = \mathbf{B}^{(k)\top} \boldsymbol{\gamma}.$$

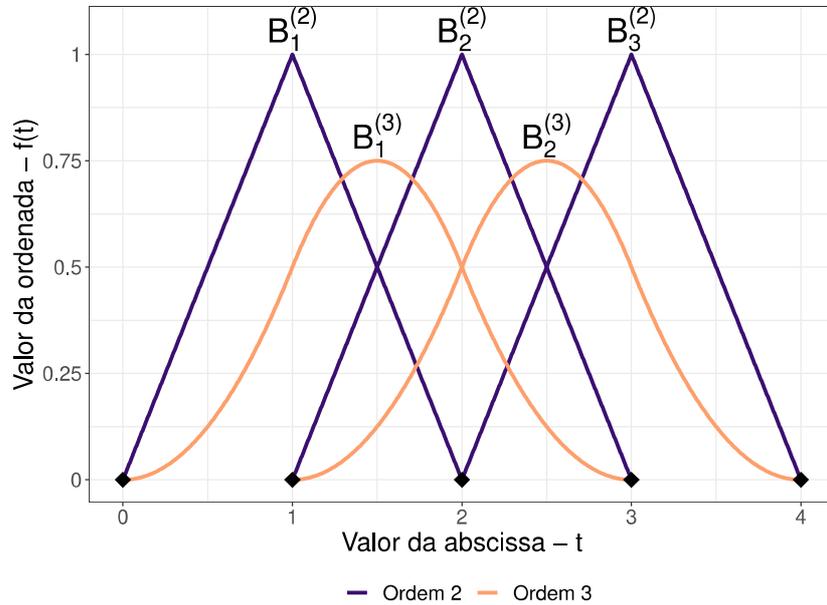
3.5 ESCOLHA DO NÚMERO DE FUNÇÕES BASE

O dilema de viés-variância¹⁴ é um problema comumente abordado na área Estatística e Probabilidade, qual trata da seguinte relação

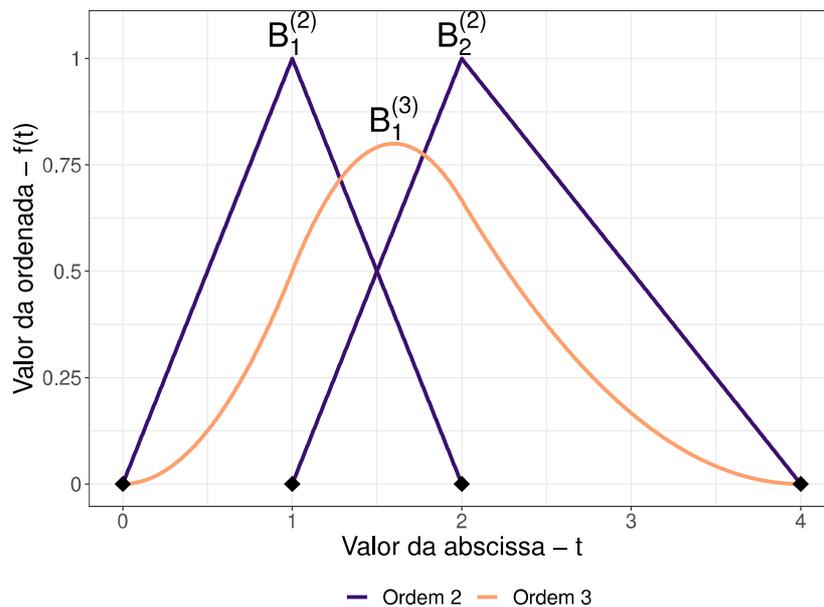
$$\text{EQM}[\hat{\chi}(t)] = \overbrace{\text{Viés}[\hat{\chi}(t)]^2}^{(a)} + \overbrace{\text{Var}[\hat{\chi}(t)]}^{(b)}, \quad (3.16)$$

¹⁴ Também conhecido em inglês como ‘*Bias-Variance Tradeoff*’.

Figura 3.4 - (a) Funções base de B-splines de ordens 2 e 3 com nós em $\xi_1 = 0$, $\xi_2 = 1$, $\xi_3 = 2$, $\xi_4 = 3$ e $\xi_5 = 4$; (b) Funções base de B-splines de ordens 2 e 3 com nós em $\xi_1 = 0$, $\xi_2 = 1$, $\xi_3 = 2$ e $\xi_4 = 4$.



(a)



(b)

Fonte: Elaborado pelo autor.

onde o termo Viés $[\hat{\chi}(t)] = \chi(t) - \mathbb{E}[\hat{\chi}(t)]$. Conforme são adicionadas mais funções base, o valor no termo (a) em (3.16) tende a diminuir, enquanto o termo (b) tende a aumentar. Em muitas situações, faz-se justificável selecionar estimativas com viés superior caso a variância das mesmas sejam significativamente inferiores, porém uma possível consequência desta seleção é que certas características locais da função aproximada podem não ser refletidas na curva estimada. Faz-se necessário então a definição de algum método para a

busca pelo valor K ideal de funções base, qual é capaz de balancear a relação em (3.16). Sousa *et al* (2022, (68)) e Anselmo *et al* (2005, (69)) propõem técnicas de penalização estocástica aplicadas aos modelos não paramétricos de estimação de curva, de forma a auxiliar tal seleção, porém não existe técnica capaz de otimamente definir o número de funções base K . Adicionalmente, a função em (3.16), denominada função de perda \mathcal{L}^2 , embora a mais popularmente aplicada, trata-se de apenas uma das diferentes funções de perda quais podem ser utilizadas para mensuração da bondade de ajuste de um estimador, onde cada qual função de perda potencialmente irá apontar um número K de funções base ideais distinto das demais.

Embora a seleção deste número não seja tópico do presente trabalho, o conceito do dilema viés-variância pode ser abordado também no contexto da seleção do coeficiente de regularização utilizado para a suavização dos dados, como visto em (3.10), qual será tópico da seção seguinte.

3.6 REGULARIZAÇÃO APLICADA À ESTIMAÇÃO DE CURVAS

Como denotado na seção anterior, o dilema viés-variância, dado por exemplo em (3.16), é ainda relevante no contexto da seleção do coeficiente de regularização λ , visto no problema de graduação proposto em (3.10). Enquanto o objetivo mantém-se em estimar a curva $\hat{\chi}(t)$ qual tenha melhor ajuste aos m dados mensurados na forma $y_j = \chi(t_j) + e_j$, como em (2.6), é necessário garantir-se também que a curva estimada mantenha-se suave, onde a ‘não suavidade’, neste contexto, corresponde à ‘oscilação’ qual pode ser ocasionalmente observada na proximidade dos pontos (t_j, y_j) interpolados pelas curvas de estimação.

Propõe-se a seguinte forma de quantificação da não suavidade

$$\text{PEN}_2(\chi) = \int_a^b [D^2 \hat{\chi}(t)]^2 dt.$$

Nota-se que funções com tendências oscilatórias tendem a possuir derivadas com altos valores absolutos e, trivialmente, a potência quadrada destas derivadas apresentará também alto valores; esta forma assemelha-se ainda ao formato da função de penalização Ridge, proposta por Hoerl e Kennard (1970, (70, 71)). De forma generalizada, define-se a quantificação da não suavidade em termos de $q \geq 2$ como

$$\text{PEN}_q(\chi) = \int_a^b [D^q \hat{\chi}(t)]^2 dt. \quad (3.17)$$

Retomando o problema de graduação visto em (3.10), tão como a SEQ vista em (2.13), é possível então escrever a seguinte soma de erros quadrados penalizada (SEQPEN)

$$\text{SEQPEN}(\gamma|\mathbf{y}, \lambda) = (\mathbf{y} - \Phi\gamma)^\top (\mathbf{y} - \Phi\gamma) + \lambda \int_a^b [D^q \hat{\chi}(t)]^2 dt. \quad (3.18)$$

A estimativa suave da curva $\chi(t)$ será então obtida através do cálculo do vetor de coeficientes γ qual minimiza (3.18). Novamente, nota-se que o valor λ é denominado ‘coeficientes de regularização’, tal que, em casos quais $\lambda = 0$, a não suavidade da curva estimada não sofre penalização, e o problema resume-se à interpolação dos dados mensurados, com solução equivalente à via mínimos quadrados vista em (2.15). Em contraste, caso $\lambda \rightarrow \infty$, a curva estimada $\hat{\chi}(t)$ tende a limitar-se à forma de um polinômio de grau $q - 1$, tal que suas derivadas superiores a q são nulas, isto é $D^q \hat{\chi}(t) = 0$.

A Figura 3.5 exhibe um exemplo da variação do coeficiente de regularização λ . Neste cenário, foram gerados 100 dados conforme a função

$$f(t_j) = \frac{\sin(2\pi t_j)}{2\pi t_j^2} + e_j, \quad (3.19)$$

onde t_j foram selecionados como 100 valores equidistantes no intervalo $[1/2, 2]$, enquanto $e_j \sim \mathcal{N}(0, 0.1^2)$. Ademais, as funções base são de ordem cinco, enquanto a penalização incide sob a terceira derivada das estimações. Utilizando-se coeficiente de regularização $\lambda = 0.0001$ obteve-se uma curva satisfatoriamente próxima à curva real geradora dos dados, enquanto que com coeficiente de regularização $\lambda = 1$, a curva gerada foi de uma parábola simples, que não reflete de forma alguma o comportamento da função aproximada. Este segundo caso é comumente denominado super-penalização, e serve de demonstração da propriedade enunciada anteriormente qual afirma que, quando a regularização é aplicada sob a q -ésima derivada, a curva resultante tende a exibir o comportamento de um polinômio do $q - 1$ -ésimo grau.

Anteriormente, foi visto que a estimativa de uma curva pode ser escrita através da seguinte combinação linear

$$\hat{\chi}(t) = \sum_{j=1}^K \gamma_j \phi_j(t) = \boldsymbol{\gamma}^\top \boldsymbol{\phi}(t) = \boldsymbol{\phi}(t)^\top \boldsymbol{\gamma}$$

e substituindo esta forma na quantificação da não suavidade generalizada em (3.17), obtém-se

$$\begin{aligned} \text{PEN}_q(\chi) &= \int_a^b [D^q \hat{\chi}(t)]^2 dt \\ &= \int_a^b [D^q \hat{\chi}(t) D^q \hat{\chi}(t)] dt \\ &= \boldsymbol{\gamma}^\top \left[\int_a^b D^q \boldsymbol{\phi}(t) D^q \boldsymbol{\phi}(t)^\top dt \right] \boldsymbol{\gamma} \\ &= \boldsymbol{\gamma}^\top \mathbf{R} \boldsymbol{\gamma}, \end{aligned} \quad (3.20)$$

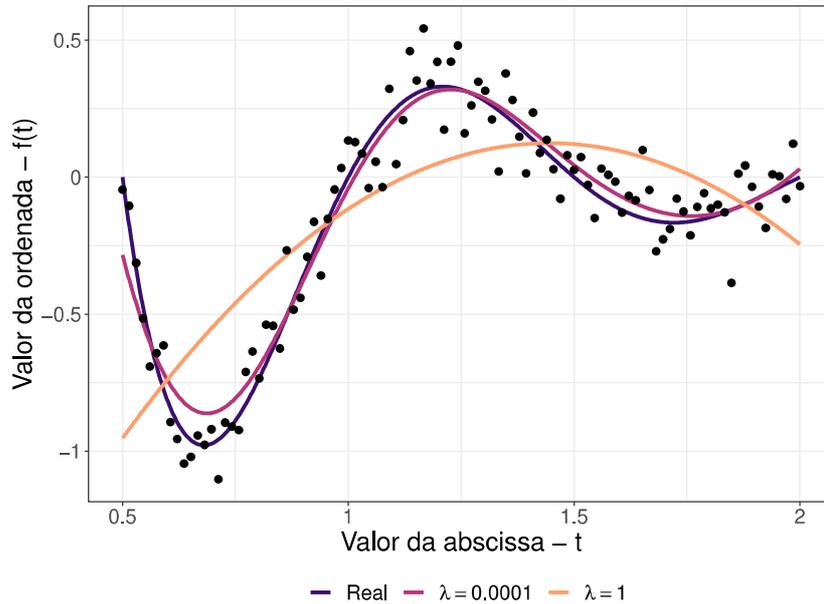
onde

$$\mathbf{R} = \int_a^b D^q \boldsymbol{\phi}(t) D^q \boldsymbol{\phi}(t)^\top dt,$$

tal que $\mathbf{R} \in \mathbb{R}^{K \times K}$ é uma matriz quadrada simétrica tal que

$$R_{i,j} = \int_a^b D^q \phi_i(t) D^q \phi_j(t) dt; \quad (3.21)$$

Figura 3.5 - Curvas suaves geradas para aproximação da função (3.19), com distintos valores do coeficiente de regularização λ .



Fonte: Elaborado pelo autor.

para algumas classes de funções base, como as B-splines, há formas analíticas de calcular os valores da matriz em (3.21). Reescreve-se a soma dos erros quadrados penalizada em (3.18) como

$$\text{SEQPEN}(\gamma|\mathbf{y}, \lambda) = (\mathbf{y} - \Phi\gamma)^\top (\mathbf{y} - \Phi\gamma) + \lambda\gamma^\top \mathbf{R}\gamma. \quad (3.22)$$

Repetindo-se o procedimento estabelecido anteriormente, deriva-se a função (3.22) com respeito a γ , qual resulta em

$$\frac{\partial \text{SEQPEN}(\gamma|\mathbf{y}, \lambda)}{\partial \gamma} = -2\Phi^\top \mathbf{y} + 2\Phi^\top \Phi^\top \gamma + 2\mathbf{R}\gamma.$$

Em conclusão, é obtido o seguinte estimador penalizado para os coeficientes da curva

$$\hat{\gamma}^{\text{Pen}} = (\Phi^\top \Phi + \lambda \mathbf{R})^{-1} \Phi^\top \mathbf{y}. \quad (3.23)$$

Neste contexto, a curva estimada é obtida como

$$\begin{aligned} \hat{\chi}(t) &= \Phi \hat{\gamma}^{\text{Pen}} \\ &= \Phi (\Phi^\top \Phi + \lambda \mathbf{R})^{-1} \Phi^\top \mathbf{y} \\ &= \mathbf{H}(\lambda) \mathbf{y}, \end{aligned} \quad (3.24)$$

onde $\mathbf{H}(\lambda) = \Phi (\Phi^\top \Phi + \lambda \mathbf{R})^{-1} \Phi^\top$ remete à matriz chapéu, ou matriz de projeção ortogonal, apresentada anteriormente. No entanto, neste contexto $\mathbf{H}(\lambda)$ não é necessariamente idempotente, como no contexto anterior, portanto é denominada matriz de sub-projeção

ortogonal ou matriz chapéu penalizada, embora ainda sirva a funcionalidade de um filtro de suavização linear. Novamente, $\mathbf{H}(\lambda)$ é utilizado no cálculo dos graus de liberdades efetivos deste modelo, qual agora são calculados como

$$\text{Número de parâmetros} = \text{tr}[\mathbf{H}(\lambda)];$$

isto é, os graus de liberdade efetivos são então dados pela diferença entre o número de instantes em quais a curva foi mensurada e o ‘número de parâmetros’ calculado acima.

Em Ramsay e Silverman (2005, (37), pp 315–316), os autores ressaltam ainda que, embora a penalização em (3.20) trate-se da mais comumente aplicada, a regularização sob splines pode ser estendida para abranger operadores diferenciais lineares de forma geral. Por exemplo, é possível utilizar do operador $L(\hat{\chi})(t) = \beta\hat{\chi} + D\hat{\chi}(t)$, onde $\beta \in \mathbb{R}$, em qual contexto a soma do quadrados dos erros penalizada seria

$$\text{SEQPEN}(\boldsymbol{\gamma}|\mathbf{y}, \lambda) = (\mathbf{y} - \Phi\boldsymbol{\gamma})^\top (\mathbf{y} - \Phi\boldsymbol{\gamma}) + \lambda \int_a^b [\beta\hat{\chi} + D\hat{\chi}(t)]^2 dt. \quad (3.25)$$

Em Wahba (1990, (58)), a autora demonstra adicionalmente que, dado uma penalização na forma de um operador diferencial $L(\hat{\chi})$ qualquer, a seguinte relação é verdadeira

$$\int_a^b \text{Viés}[\hat{\chi}(t)]^2 dt \leq \int_a^b [L(\hat{\chi})(t)]^2 dt \quad (3.26)$$

pode ser conveniente então, com finalidade de minimizar o viés, aproximar $L(\hat{\chi})(t) = 0$, o que faz-se particularmente útil quando se tem informações prévias sobre o comportamento aproximado das curvas amostradas; a título de exemplo, em casos quais os dados mensurados apresentam comportamento aproximadamente exponencial com taxa de crescimento β , a penalização em (3.25) aproxima-se de 0.

3.6.1 ESCOLHA DO COEFICIENTE DE REGULARIZAÇÃO

Para que seja feita a escolha do coeficiente de regularização λ ideal, primeiramente é necessário a definição da medida de bondade de ajuste qual será utilizada para comparações entre distintos valores para o coeficiente de regularização. Como visto anteriormente, no contexto da seleção do número de funções base K usualmente utiliza-se de funções de perda, como por exemplo \mathcal{L}^2 . Em contraste, para a seleção do coeficiente de regularização tradicionalmente são utilizados métodos de validação cruzada, como vistos em Stone (1974, (72)), quais são abordados a seguir.

3.6.1.1 Validação Cruzada

A técnica de validação cruzada é tradicionalmente utilizada no contexto da construção de modelos preditivos: dada uma amostra composta de m mensurações y_1, y_2, \dots, y_m , constrói-se um modelo preditivo para a j -ésima mensuração através das $m - 1$ demais mensurações $y_1, y_2, \dots, y_{j-1}, y_{j+1}, \dots, y_m$, qual se escreve $\hat{\chi}(t, \lambda)^{(-j)}$. Este estimador minimiza

a seguinte equação

$$\text{SEQPEN}(\boldsymbol{\gamma}|\mathbf{y}^{(j)}, \lambda) = (\mathbf{y}^{(-j)} - \Phi^{(-j)}\boldsymbol{\gamma})^\top (\mathbf{y}^{(-j)} - \Phi^{(-j)}\boldsymbol{\gamma}) + \lambda \int_a^b [D^q \hat{\chi}^{(-j)}(t, \lambda)]^2 dt.$$

Desta forma, pode-se utilizar da diferença entre o valor estimado da curva e o valor mensurado em cada instante, isto é $\hat{\chi}(t_j)^{(-j)} - y_j$, para estimar o ajuste preditivo do modelo. A validação cruzada tradicional escreve-se então como seguinte

$$\text{VC}(\lambda) = \frac{1}{m} \sum_{j=1}^m [y_j - \hat{\chi}^{(-j)}(t_j, \lambda)]^2. \quad (3.27)$$

O valor ótimo do coeficiente de regularização seria então λ qual minimiza (3.27). Observa-se que, da maneira qual está definido $\text{VC}(\lambda)$ em (3.27), é necessário a estimação de m curvas distintas para o cálculo de VC, procedimento qual pode ser computacionalmente custoso. Utiliza-se então da seguinte relação

$$y_j - \hat{\chi}^{(-j)}(t_j, \lambda) = \frac{y_j - \hat{\chi}(t_j, \lambda)}{1 - \mathbf{H}_{j,j}(\lambda)},$$

onde $\mathbf{H}_{j,j}(\lambda)$ trata-se do j -ésimo valor na diagonal da matriz de sub-projeção vista em (3.24). Desta forma, é possível substituir os resíduos em (3.27) de forma a obter-se o seguinte formato de validação cruzada

$$\text{VC}(\lambda) = \frac{1}{m} \sum_{j=1}^m \left[\frac{y_j - \hat{\chi}(t_j, \lambda)}{1 - \mathbf{H}_{j,j}(\lambda)} \right]^2, \quad (3.28)$$

para qual basta a estimação de uma curva para o cálculo da medida.

3.6.1.2 Validação Cruzada Generalizada

Outra medida de bondade de ajuste usualmente aplicada, por exemplo por Craven e Wahba (1978, (73)), tangencial à validação cruzada proposta na seção anterior trata-se da validação cruzada generalizada. Neste contexto, substituí-se os termos $1 - \hat{\mathbf{H}}_{j,j}(\lambda)$ pelos termos $1 - 1/n \cdot \text{tr}[\mathbf{H}(\lambda)]$, isto é, pela sua média, obtendo-se então o seguinte cálculo

$$\text{VCG}(\lambda) = \frac{1}{m} \sum_{j=1}^m \left[\frac{y_j - \hat{\chi}(t_j)}{1 - 1/n \cdot \text{tr}[\mathbf{H}(\lambda)]} \right]^2. \quad (3.29)$$

Trivialmente, tão como no caso da validação cruzada padrão, o objetivo é obter coeficiente de regularização λ qual minimize (3.29).

3.6.1.3 Comentário sobre Validação Cruzada de múltiplas curvas

As medidas de bondade de ajuste apresentadas nos capítulos anteriores, assim como as técnicas de regularização que as antecedem, são comumente fundamentadas e expostas no contexto da estimação de uma única curva. Porém, no contexto de dados funcionais, é

necessário levar-se em consideração a estimação de toda uma amostra $\chi_1, \chi_2, \dots, \chi_n$ de curvas. Neste contexto, é possível definir o coeficiente de regularização como um vetor $\boldsymbol{\lambda} \in \mathbb{R}^n$, onde $\lambda_i > 0 \forall i \in \{1, 2, \dots, n\}$, e definir a soma de erros quadrados penalizados como

$$\text{SEQPEN}(\boldsymbol{\gamma}^* | \mathbf{y}^*) = \sum_{i=1}^n [(\mathbf{y}_i - \Phi_i \boldsymbol{\gamma}_i)^\top (\mathbf{y}_i - \Phi_i \boldsymbol{\gamma}_i, \boldsymbol{\lambda}) + \lambda_i \boldsymbol{\gamma}_i^\top \mathbf{R}_i \boldsymbol{\gamma}_i], \quad (3.30)$$

onde aqui os termos Φ_i , \mathbf{y}_i e $\boldsymbol{\gamma}$ são definidos de forma idem à Seção 2.5.1.4, enquanto \mathbf{R}_i diz respeito à matriz construída como em (3.21) para a i -ésima curva. Derivando (3.30) com respeito $\boldsymbol{\gamma}_i \forall i \in \{1, 2, \dots, n\}$ e igualando o resultado a 0 obtém-se que os estimadores quais minimizam (3.30) são escritos como

$$\{\hat{\boldsymbol{\gamma}}_i = (\Phi_i^\top \Phi_i + \lambda_i \mathbf{R}_i)^{-1} \Phi_i^\top \mathbf{y}_i\}_{\forall i \in \{1, 2, \dots, n\}}.$$

Desta forma, a validação cruzada generalizada calcula-se como

$$\begin{aligned} \text{VCG}(\boldsymbol{\lambda}) &= \frac{1}{m} \sum_{i=1}^n \sum_{j=1}^m \left[\frac{y_{i,j} - \hat{\chi}_i(t_j)}{1 - 1/n \cdot \text{tr}[\mathbf{H}_i(\lambda_i)]} \right]^2 \\ &= \sum_{i=1}^n \text{VCG}_i(\lambda_i), \end{aligned} \quad (3.31)$$

onde $\mathbf{H}_i(\lambda_i)$ diz respeito à matriz de sub-projeção da i -ésima curva estimada. Nota-se que a minimização de (3.31) torna-se mais complexa conforme se acrescem mais curvas, uma vez que a otimização é realizada sob $\boldsymbol{\lambda} \in \mathbb{R}^n$. É comum então, por parcimônia, definir um único coeficiente de regularização $\lambda \in \mathbb{R}$ para todas curvas simultaneamente, de forma que o problema mantém-se como um problema de minimização univariado. Por exemplo, isto é realizado no pacote `fda` em `R`, qual foi utilizado na confecção do presente trabalho. Agregando a isto às suposições feitas na Seção 2.5.1.4 de balanceamento nas amostras de curvas, é possível encontrar a solução dos $\boldsymbol{\gamma}_i$'s quais minimizam a soma dos erros quadrados penalizada de todas curvas como

$$\{\hat{\boldsymbol{\gamma}}_i = (\Phi^\top \Phi + \lambda \mathbf{R})^{-1} \Phi^\top \mathbf{y}_i\}_{\forall i \in \{1, 2, \dots, n\}}.$$

3.7 SUAVIZAÇÃO RESTRITA

Como citado anteriormente, é possível ainda restringir as curvas estimadas de forma que estas obedeçam a propriedades arbitrárias quais os dados, por sua natureza, devem obedecer. Por exemplo, tem-se que os teores de um determinado minério em um conjunto espacialmente indexado, tópico comumente abordado na área de krigagem, deve ser restrito a valores positivos; complementarmente, em estudos de curvas de crescimento, tradicionalmente, espera-se que as curvas exibam comportamento de crescimento monótono. A fundamentação para estas restrições é advinda da área tangencial de equações diferenciais ordinárias, abordada por exemplo em Boyce e DiPrima (2010, (74)).

3.7.1 Suavização positiva

Dada uma função $\omega(t)$ definida em $[a, b]$, e um dado funcional $\chi(t)$ definido no mesmo intervalo, é possível explicitamente restringir $\chi(t)$ à forma de um dado funcional estritamente positivo ao escrever

$$D\chi(t) = \omega(t)\chi(t). \quad (3.32)$$

Supondo que $\omega(t)$ é uma função integrável, a solução para a equação diferencial homogênea de primeira ordem em (3.32) é dada por

$$\chi(t) = c \cdot \exp \left[\int_a^t \omega(s) ds \right],$$

onde $c \in \mathbb{R}$, $c \neq 0$ indica uma constante real tradicionalmente definida através das condições de contorno. Dessa forma, é possível então reescrever $W(t) = \int_a^t \omega(s) ds + \log c$, qual ao ser substituído na solução acima, resulta em

$$\chi(t) = \exp \{W(t)\}.$$

Uma vez que a função $W(t)$ mantém-se irrestrita, é possível estimar seu formato através de K funções base, definindo $\hat{W}(t) = \sum_{k=1}^K \gamma_k \phi_k(t)$, tal que a soma de erros quadrados penalizada será então dada por

$$\text{SEQPEN}(\boldsymbol{\gamma}|\mathbf{y}, \lambda) = (\mathbf{y} - \exp \{ \hat{W}(t) \})^\top (\mathbf{y} - \exp \{ \hat{W}(t) \}) + \lambda \int_a^b [D^q \hat{W}(t)]^2 dt.$$

Diferentemente dos métodos anteriores, este não apresenta solução analítica, e consequentemente requer a implementação de métodos computacionais para sua estimação, embora não seja por consequência intratável; para a implementação deste tipo de restrição, o pacote `fda` dispõe da função `smooth.pos()`.

4 ANÁLISE DE DADOS FUNCIONAIS

4.1 ESTIMAÇÃO DA FUNÇÃO DE COVARIÂNCIA

Retoma-se então a fórmula proposta em (2.4) de estimação da covariância em uma amostra de dados funcionais $\chi_1, \chi_2, \dots, \chi_n$, qual se escreve

$$\text{cov}_{\chi, \chi}(t_j, t_k) = \frac{1}{n-1} \sum_{i=1}^n [\chi_i(t_j) - \bar{\chi}(t_j)]^2.$$

Embora este cálculo, por si só, já apresente suficiente utilidade prática no contexto de estatística não paramétrica, ele pode também ser utilizado para a estimação dos parâmetros σ^2 vistos, por exemplo, nas funções de covariância propostas nas equações em (2.8) e (2.10). É possível definir a matriz de de covariância estimada como

$$\hat{\Sigma} = \begin{bmatrix} \text{cov}_{\chi, \chi}(t_1, t_1) & \text{cov}_{\chi, \chi}(t_1, t_2) & \dots & \text{cov}_{\chi, \chi}(t_1, t_m) \\ \text{cov}_{\chi, \chi}(t_2, t_1) & \text{cov}_{\chi, \chi}(t_2, t_2) & \dots & \text{cov}_{\chi, \chi}(t_2, t_m) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}_{\chi, \chi}(t_m, t_1) & \text{cov}_{\chi, \chi}(t_m, t_2) & \dots & \text{cov}_{\chi, \chi}(t_m, t_m) \end{bmatrix}.$$

Em que $\hat{\Sigma} \in \mathbb{S}_+^m$, isto é, pertence ao conjunto de matrizes simétricas e positivas definidas de dimensão m . Se o modelo suposto para a função de covariância é uma função $\Sigma(\cdot, \cdot)$ parametrizada por $\sigma^2 > 0$, é possível construir um estimador de σ^2 ao buscar o valor de σ^2 qual minimiza a seguinte relação

$$\|\text{vech}(\hat{\Sigma}) - \text{vech}[\Sigma(t_j, t_k)]\|_2^2 = \sum_{j=i}^m \sum_{i=1}^m [\hat{\Sigma}_{i,j} - \Sigma(t_i, t_j)]^2, \quad (4.1)$$

isto é, os mínimos quadrados ordinários, onde $\text{vech}(A)$ indica o operador de meia vetorização de uma matriz simétrica A , o qual é utilizado para criação de um vetor composto dos valores da diagonal, e superiores à diagonal, da matriz A . Por exemplo, para uma matriz simétrica A de dimensões m qualquer, tem-se

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m} \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,m} \end{bmatrix} \implies \text{vech}(A) = \begin{bmatrix} a_{1,1} \\ a_{1,2} \\ \vdots \\ a_{1,m} \\ a_{2,2} \\ a_{2,3} \\ \vdots \\ a_{m-1,m} \\ a_{m,m} \end{bmatrix}$$

Além disso, ao substituir-se a função de covariância exponencial quadrática vista em (2.10) nos mínimos quadrados ordinários em (4.1), obtêm-se a seguinte fórmula

$$\sum_{j=i}^m \sum_{i=1}^m \left[\hat{\Sigma}_{i,j} - \sigma^2 \exp \left\{ -\frac{\|t_i - t_j\|_2^2}{2} \right\} \right]^2.$$

Esta fórmula, quando diferenciada com respeito a σ^2 e igualada a 0, resulta em

$$-2 \sum_{j=i}^m \sum_{i=1}^m \exp \left\{ -\frac{\|t_i - t_j\|_2^2}{2} \right\} \left[\hat{\Sigma}_{i,j} - \sigma^2 \exp \left\{ -\frac{\|t_i - t_j\|_2^2}{2} \right\} \right] = 0.$$

Obtém-se que o estimador de mínimos quadrados para o parâmetro σ^2 da função de covariância será então

$$\hat{\sigma}^2 = \frac{\sum_{j=i}^m \sum_{i=1}^m \exp \left\{ -\frac{\|t_i - t_j\|_2^2}{2} \right\} \hat{\Sigma}_{i,j}}{\sum_{j=i}^m \sum_{i=1}^m \exp \left\{ -\|t_i - t_j\|_2^2 \right\}}.$$

Vale notar aqui que, tanto no contexto da função de covariância exponencial quadrática, quanto no contexto do ruído branco apresentado em (2.8), a variância em todo instante t é suposta como $\Sigma(t, t) = \sigma^2$, isto é, constante. Porém observa-se, por exemplo na Figura 2.3-(a), que a variância pode não ser constante para todos instantes. Uma alternativa, como vista em Saraiva (2009, (27)), é a utilização de B-splines para modelagem adicional da variância. Esta abordagem é similar àquela implementada para modelos GAMLSS¹ vistos em Rigby e Stasinopoulos (2005, (75)), capazes ainda de modelar a assimetria e curtose de um conjunto de mensurações.

4.1.1 Construção do intervalo de confiança

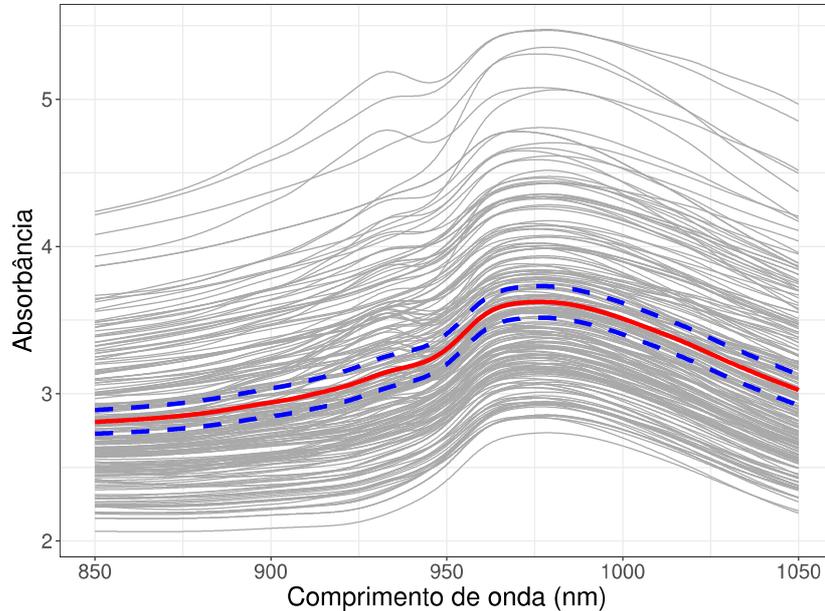
Uma vez estimada a função de covariância, a construção de intervalos de confiança para as realizações de uma determinada população faz-se simples: usualmente, é feita adicionalmente a suposição que os dados são distribuídos normalmente, de forma que é possível utilizar-se de quantis da distribuição normal para a definição dos intervalos de confiança. Para uma amostra de dados funcionais $\chi_1, \chi_2, \dots, \chi_n$, estima-se de forma pontual o intervalo de confiança a nível $\gamma\%$ para o instante t como

$$\text{IC}(\chi(t), \gamma\%) = \left[\bar{\chi}(t) + c \sqrt{\frac{\Sigma(t, t)}{n}}; \bar{\chi}(t) - c \sqrt{\frac{\Sigma(t, t)}{n}} \right], \quad (4.2)$$

onde c é o quantil γ -percentil da distribuição normal padronizada, tal que, se $Z \sim \mathcal{N}(0, 1)$, $P(-c < Z < c) = \gamma\%$. A Figura 4.1 exhibe o limite de confiança para as realizações do conjunto de dados de absorvância de carnes picadas supracitado. Embora, para este conjunto de dados, a variância não apresente comportamento constante, como visto na Figura 2.3, a variação é insignificante quanto vista na mesma escala que os valores de absorvância mensurados.

¹ Vindo do inglês ‘*Generalized Additive Modelos for Location Scale and Shape*’.

Figura 4.1 - Intervalos de confiança pontuais a nível 95% para as absorvâncias mensuradas em 215 carnes picadas submetidas a diferentes comprimentos de onda, denotados pela curvas tracejadas em azul.



Fonte: Elaborado pelo autor.

4.2 AGRUPAMENTO DE DADOS FUNCIONAIS

Também no contexto de dados funcionais, para quase todo estudo é feita a suposição que uma amostra é composta de variáveis aleatórias $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_n$ distribuídas conforme um processo qualquer \mathcal{X} apenas. Porém, há uma grande quantidade de contextos em quais uma amostra pode ser ‘contaminada’ por variáveis aleatórias funcionais distribuídas de outras formas, cada qual exibindo seu próprio padrão. Estes padrões dificilmente serão captados via métodos de visualização gráfica, de forma que o estatístico deve então recorrer a outros métodos para sua detecção.

O principal destes métodos é a ‘análise de agrupamentos’. Em Hastie *et al* (2009, (76)) os autores descrevem o agrupamento² como uma forma de aprendizado estatístico não supervisionado, uma vez que, embora apresenta variável de ‘entrada’³, quais neste contexto são os dados funcionais $\chi_1, \chi_2, \dots, \chi_n$, não apresenta variável de ‘saída’⁴ explícita sendo observada. Em Marsland (2011, (77)), o autor especifica que o propósito da análise de agrupamento é definir um conjunto finito de grupos tais que os elementos contidos nestes grupos são similares entre si, enquanto distintos dos elementos dos demais grupos; estas medidas de dissimilaridade serão definidas adiante.

² Do inglês ‘*clustering*’.

³ Do inglês ‘*input*’.

⁴ Do inglês ‘*output*’.

Entre os métodos de agrupamento, há aqueles denominados métodos por particionamento, ou não hierárquicos, como o método das K -médias, assim nomeados originalmente pelo autor em MacQueen (1967, (78)), e os métodos hierárquicos, vistos por exemplo em Ward (1963, (79)). O presente trabalho será focado na implementação dos métodos hierárquicos de agrupamento. Em particular, a abordagem ‘aglomerativa’⁵ dos métodos hierárquicos, inicialmente atribui a uma amostra de n observações, n grupos, de forma que na primeira etapa todas observações pertencem a um grupo qual contém elas mesmas, e a cada etapa seguinte são unidos os grupos dois quais apresentam menor dissimilaridade entre si. A forma em que a dissimilaridade entre grupos é calculada será vista na Seção 4.2.2.

4.2.1 Medidas de dissimilaridade

É necessário então caracterizar as medidas de dissimilaridade entre dados. Nota-se, porém, que usualmente as medidas de dissimilaridade são fundamentadas no âmbito da análise de dados multivariados, quais apresentam dimensão finita, enquanto a análise de dados funcionais trata de elementos de dimensão infinita. Faz-se necessário então a definição de medidas de dissimilaridade $d(\cdot, \cdot)$ capazes de tratar de dados funcionais. Dada uma amostra de dados funcionais $\chi_1, \chi_2, \dots, \chi_n$, estas medidas são funções $d : \Omega \times \Omega \rightarrow \mathbb{R}$ quais satisfazem as seguintes propriedades

Prop. (i) $d(\chi_i, \chi_i) = 0, \forall i \in \{1, 2, \dots, n\}$;

Prop. (ii) $d(\chi_i, \chi_j) \geq 0, \forall (i, j) \in \{1, 2, \dots, n\} \times \{1, 2, \dots, i-1, i+1, \dots, n\}$;

Prop. (iii) $d(\chi_i, \chi_j) = d(\chi_j, \chi_i), \forall (i, j) \in \{1, 2, \dots, n\} \times \{1, 2, \dots, n\}$;

Prop. (iv) $d(\chi_i, \chi_j) + d(\chi_j, \chi_k) \geq d(\chi_i, \chi_k), \forall (i, j, k) \in \{1, 2, \dots, n\} \times \{1, 2, \dots, n\} \times \{1, 2, \dots, n\}$;

onde a Propriedade (iii) acima é denominada ‘Propriedade da simetria’, enquanto a Propriedade (iv) remete à desigualdade triangular; isto é, medidas $d(\cdot, \cdot)$ mais próximas de 0 indicam mais proximidade entre duas observações. Tradicionalmente, antes de efetuar-se o agrupamento, os dados são padronizados. Em Dias e Estevam (2010, (36)), os autores definem a padronização de dados funcionais como o seguinte: dada uma função g qualquer definida no intervalo $[a, b]$, tem-se que

$$g^*(t) = \frac{g(t)^2}{\int_a^b g(t)^2 dt}, \quad (4.3)$$

tal que $g^*(t) > 0$ e $\int_a^b g^*(t) dt = 1$; equivalentemente, $g^*(t)$ é uma função de densidade de probabilidade. Também em Dias e Estevam (2010, (36)), os autores citam, entre outras, as seguintes medidas de dissimilaridade, dadas duas curvas distintas χ e ϑ definidas sob o intervalo $[a, b]$.

⁵ Também conhecida em inglês como ‘bottom-up’.

4.2.1.1 Distância de Hellinger

Proposta em Hellinger (1909, (80)) e também vista em Ferreira (2003, (81)), a Distância de Hellinger é escrita como

$$H(\chi, \vartheta) = \sqrt{\left[\int_a^b \left(\sqrt{\chi^*(t)} - \sqrt{\vartheta^*(t)} \right)^2 dt \right]}.$$

4.2.1.2 Distância de Kullback-Leibler

A distância de Kullback-Leibler, baseada nos conceitos de entropia relativa introduzidos em Kullback e Leibler (1951, (82)), é escrita como

$$\text{KL}(\chi, \vartheta) = \int_a^b [\log \chi^*(t) - \log \vartheta^*(t)] \chi^*(t) dt.$$

4.2.1.3 Diferença Quadrática Integrada (DQI)

A diferença quadrática integrada, qual é similar à norma euclidiana $\|\cdot\|_2^2$, agora estendida ao contexto de dados funcionais, é escrita como

$$\text{DQI}(\chi, \vartheta) = \int_a^b [\chi(t) - \vartheta(t)]^2.$$

Adicionalmente, em Dias e Estevam (2010, (36), pp. 114-116), os autores apresentam estudos quanto à distribuição probabilística das medidas DQI, com potenciais aplicações para testes de hipóteses.

4.2.2 Métodos de ligação

A partir das medidas de dissimilaridade de uma amostra de dados funcionais $\chi_1, \chi_2, \dots, \chi_n$, constrói-se uma matriz de dissimilaridades ou distâncias $\mathbf{d} \in \mathbb{R}^{n \times n}$, qual se escreve como

$$\mathbf{d} = \begin{bmatrix} d(\chi_1, \chi_1) & d(\chi_1, \chi_2) & \dots & d(\chi_1, \chi_n) \\ d(\chi_2, \chi_1) & d(\chi_2, \chi_2) & \dots & d(\chi_2, \chi_n) \\ \vdots & \vdots & \ddots & \vdots \\ d(\chi_n, \chi_1) & d(\chi_n, \chi_2) & \dots & d(\chi_n, \chi_n) \end{bmatrix}.$$

Nota-se que $d \in \mathbb{R}^{n \times n}$ é simétrica, como consequência direta da Propriedade (iii) supra-citada. Uma vez obtida a matriz de dissimilaridades, é necessário definir o processo de aglomeração dos grupos: com este propósito, são utilizados os denominados ‘métodos de ligação’. De maneira generalizada, procura-se estabelecer N grupos distintos para uma amostra de n realizações; define-se então uma classe \mathbf{G} de N conjuntos de índices G_j , tal que se $i \in \{1, 2, \dots, n\}$ é tal que $i \in G_j$, é dito que a i -ésima realização pertence ao j -ésimo grupo ou ‘cluster’. Vale notar que a classe \mathbf{G} configura uma partição do conjunto $1, 2, \dots, n$, e obedece às propriedades

Prop. (i) $\emptyset \notin \mathbf{G}$, isto é, todos grupos possuem pelo menos uma realização;

Prop. (ii) $\cup_{j=1}^N G_j = \{1, 2, \dots, n\}$, isto é, todas realizações pertencem a um grupo;

Prop. (iii) $G_j \cap G_k = \emptyset, \forall (j, k) \in \{1, 2, \dots, N\} \times \{1, 2, \dots, N\}$, isto é, nenhuma realização pertence a mais de um grupo.

O critério qual define a atribuição de uma realização a um determinado grupo é então definido através dos denominados ‘métodos de ligação’, quais, assim como as medidas de dissimilaridade, variam na forma de implementação e resultados, sem apresentar alternativa ‘ideal’ para implementação na grande maioria dos casos. No presente trabalho, são apresentados os métodos de ligação ‘mínima’ e ‘máxima’.

4.2.2.1 Método de ligação mínima

O método de ligação mínima, ou ‘única’, calcula a dissimilaridade ou distância entre dois grupos G_j e G_k como a menor dissimilaridade mensurada entre as realizações que pertencem a cada grupo. Destaca-se que, neste contexto, $d(G_j, G_k)$ diz respeito à dissimilaridade entre os grupos, qual no caso da ligação mínima pode ser escrita como

$$d_{\text{MIN}}(G_j, G_k) = \min_{(q,p) \in G_j \times G_k} \{d(\chi_q, \chi_p)\}.$$

Uma possível consequência da aplicação da ligação simples é a de que grandes grupos quais apresentam padrões distintos comumente serão inadvertidamente aglomerados em grupo um só; para isto ocorrer, basta que um elemento em um grupo esteja próximo de seu vizinho.

4.2.2.2 Método de ligação máxima

De maneira quase diametralmente oposta ao método de ligação mínima, o método de ligação, ou ‘completa’, máxima calcula a dissimilaridade entre dois grupos G_j e G_k como a **maior** dissimilaridade mensurada entre as realizações que pertencem a cada grupo, de forma que pode ser escrita como

$$d_{\text{MAX}}(G_j, G_k) = \max_{(q,p) \in G_j \times G_k} \{d(\chi_q, \chi_p)\}.$$

Enquanto este método alternativo não é tão propício à aglomeração de dois grupos distintos em um, ele apresenta outra limitação: grupos quais possuem realizações aberrantes, ou ‘*outliers*’, dificilmente serão aglomerados.

4.2.3 Seleção da quantidade de grupos

Assim como no contexto da regularização, para determinação do número específico de grupos quais devem ser usados no agrupamento, é necessário a definição de alguma

medida de bondade de ajuste. No contexto de agrupamento, têm-se por exemplo o índice de Dunn, proposto em Dunn (1974, (83)), e os métodos do cotovelo e da silhueta média, vistos por exemplo em Kaufman e Rousseeuw (2009, (84)). Para o presente trabalho será apresentado e implementado apenas do método da silhueta média, porém vale ressaltar que cada medida tem sua utilidade, e irá retornar resultados distintos.

4.2.3.1 Método da Silhueta Média

Para o método da silhueta média, obtém-se, a partir das medidas de dissimilaridade calculadas, valores quais buscam refletir a adequação de cada realização com o grupo a qual foi atribuída. Primeiramente, para uma realização qualquer $i \in G_p$, define-se

$$a_i = \frac{1}{\text{card}(G_p) - 1} \sum_{\substack{j \in G_p \\ j \neq i}} d(\chi_i, \chi_j),$$

onde o operador $\text{card}(\cdot)$ diz respeito à cardinalidade de um conjunto qualquer. Nota-se que a_i calcula então a dissimilaridade média da i -ésima realização com as demais realizações no mesmo grupo a qual esta realização foi atribuída. A seguir, ainda com $i \in G_p$, define-se

$$b_i = \min_{\substack{q \in \{1, 2, \dots, N\} \\ q \neq p}} \left\{ \frac{1}{\text{card}(G_q)} \sum_{j \in G_q} d(\chi_i, \chi_j) \right\}.$$

Observa-se que b_i calcula a menor dissimilaridade média qual a i -ésima realização apresenta com um grupo a qual não pertence. Desta forma, o coeficiente de silhueta da i -ésima realização é dado por s_i , qual se escreve

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}}. \quad (4.4)$$

Observa-se que, casos em quais $a_i < b_i$ indicam que a i -ésima realização se ‘ajusta’ melhor ao grupo qual foi atribuída que a qualquer outro grupo, enquanto $b_i > a_i$ indica que existe pelo menos um grupo ao qual a i -ésima realização é mais ‘adequada’ que aquele a qual ela foi atribuída. É possível ainda reescrever (4.4) como

$$s_i = \begin{cases} b_i/a_i - 1 & \text{se } a_i > b_i \\ 0 & \text{se } a_i = b_i \\ 1 - a_i/b_i & \text{se } a_i < b_i \end{cases}.$$

Trivialmente, tem-se que $-1 \leq s_i \leq 1$; quão mais próximos os valores de s_i estão de 1, melhor é a adequação da i -ésima realização ao grupo a qual esta foi atribuída. É possível então escrever o coeficiente de silhueta médio⁶ para o agrupamento de n realizações qualquer como

$$\bar{s} = \frac{1}{n} \sum_{i=1}^n s_i. \quad (4.5)$$

Trivialmente, busca-se então definir o número de grupos N tal que \bar{s} é maximizado.

⁶ Traduzido do inglês ‘Mean Silhouette Coefficient’.

4.3 MODELOS DE REGRESSÃO FUNCIONAIS

Na Estatística e Probabilidade, considera-se de interesse não apenas estudar o comportamento dos dados e variáveis em si, como também vir a entender como estes podem interagir entre si, e potencialmente ser utilizados para explicar a aleatoriedade de outros fenômenos. Com este propósito, são tradicionalmente construídos modelos, quais assumem diversas formas: modelos de regressão linear, modelos aditivos, modelos lineares generalizados, entre demasiados outros. Em Müller e Yao (2008, (85)) e James (2002, (16)) os autores estendem estas aplicações ao contexto de dados funcionais, fundamentando os denominados ‘modelos de regressão funcionais’.

Um modelo de regressão funcional pode ser classificado em uma de três situações: apenas sua variável de interesse é um dado funcional; pelo menos uma de suas variáveis explicativas é um dado funcional, porém não a variável de interesse; tanto sua variável de interesse quanto pelo menos uma de suas variáveis explicativas são dados funcionais. O presente trabalho optará por apresentar o segundo caso, com duas aplicações distintas.

4.3.1 MODELO COM VARIÁVEL DE INTERESSE ESCALAR E VARIÁVEL EXPLICATIVA FUNCIONAL

Dado uma matriz de planejamento $X \in \mathbb{R}^{n \times k}$, tal que $X = [x_1, x_2, \dots, x_n]^\top$, onde $x_i \in \mathbb{R}^k$, e vetor de respostas $z \in \mathbb{R}^n$, têm-se que o modelo de regressão linear usual, desconsiderando o intercepto, com vetor de parâmetros desconhecidos $\beta \in \mathbb{R}^k$, escreve-se

$$z_i = \overbrace{\sum_{j=1}^k \beta_j x_{i,j}}^{(a)} + e_j, \quad \forall i \in \{1, 2, \dots, n\}, \quad (4.6)$$

onde $e \in \mathbb{R}^n$ refere-se ao termo de erro, qual usualmente define-se como distribuído com média zero, variância constante e independente entre si. No contexto de dados funcionais, é possível redefinir o vetor de valores desconhecidos como uma curva de valores desconhecidos, particularmente, é possível reescrever o modelo (4.6) acima ao substituir o termo em (a) pela integral de Riemman-Stieltjes, como segue

$$z_i = \int_a^b \beta(t) \chi_i(t) + e_j, \quad \forall i \in \{1, 2, \dots, n\}. \quad (4.7)$$

Torna-se então necessário definir algum método para estimar a função desconhecida $\beta(t)$; entre outros, Ramsay (2009, (4)) define alguns que são enunciados nas seções seguintes.

4.3.1.1 Estimativa naïve

A estimativa naïve usualmente proposta trata-se simplesmente da discretização dos dados funcionais em k valores, de forma que o modelo de regressão linear funcional em

(4.7) retorna ao formato do modelo de regressão linear usual em (4.6). Define-se $k < \infty$, tal que se permita escrever

$$z_i = \sum_{j=1}^k \beta_j \chi_i(t_j) + e_j, \quad \forall i \in \{1, 2, \dots, n\}. \quad (4.8)$$

Naturalmente, este modelo apresenta diversas limitações, por exemplo, a estimação via mínimos quadrados é inviabilizada quanto $k > n$, caso em qual o problema é dito possuir ‘posto incompleto’. Como alternativa, é possível aplicar sob este modelo outras formas de estimativa do vetor $\beta \in \mathbb{R}^k$, seja através de métodos de regularização como a regressão LASSO⁷, proposta em Santosa e Symes (1986, (86)) e assim nomeada em Tibshirani(1996, (87)) ou regressão Ridge, citada anteriormente, como também via métodos de redução de dimensionalidade, como a regressão via componentes principais ou mínimos quadrados parciais. É também possível escolher k tal que $k \leq n$, embora ao assim fazer os valores estimados para o vetor β possivelmente serão incapazes de refletir com acurácia o comportamento da função $\beta(t)$ do modelo.

A estrutura em (4.8) permite ainda notar o seguinte problema: quando $k = n$, é possível estimar um vetor $\hat{\beta}$ tal que $z_i - \sum_{j=1}^k \hat{\beta}_j \chi_i(t_j) = 0 \forall i \in \{1, 2, \dots, n\}$, de forma que, em suma, os dados são interpolados pelo modelo resultante. Adicionalmente, no caso supracitado de um problema de posto incompleto, isto é, $k > n$, há um número infinito de vetores $\hat{\beta}$ satisfazendo esta mesma relação. Esta afirmação segue verdadeira quando β assume valores em um espaço de dimensão infinita, isto é, quanto $k \rightarrow \infty$, contexto equivalente à integral de Riemman-Stieltjes.

4.3.1.2 Estimativa via regularização

Como supracitado, ao tratar-se de $\beta(t)$ como uma função, o modelo fica sujeito a sobreajuste⁸, uma vez que são utilizadas variáveis de dimensão infinita para modelagem de um número finito de observações. Remete-se então, novamente, aos métodos de representação de curvas via funções de base, tal que a função $\beta(t)$ é reescrita a combinação linear de K_β funções base. Dado um vetor de funções base conhecidas $\psi(t) = [\psi_1(t), \psi_2(t), \dots, \psi_{K_\beta}(t)]^\top$ e um vetor de coeficientes desconhecidos $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_{K_\beta}]^\top$, escreve-se

$$\hat{\beta}(t) = \sum_{j=1}^{K_\beta} \alpha_j \psi_j(t) = \alpha^\top \psi(t) = \psi(t)^\top \alpha. \quad (4.9)$$

⁷ Do inglês ‘*Least Absolute Shrinkage Selection Operator*’.

⁸ Traduzido do inglês ‘*Overfitting*’.

Desta forma, é simples então definir os valores ótimos de $\boldsymbol{\alpha}$ como aqueles quais minimizam a seguinte soma de erros quadrados penalizada

$$\begin{aligned} \text{SEQPEN}(\boldsymbol{\alpha}|\mathbf{z}, \lambda) &= \sum_{i=1}^n \left(z_i - \int_a^b \hat{\beta}(t) \chi_i(t) dt \right)^2 + \rho \int_a^b [L\hat{\beta}(t)]^2 dt \\ &= \sum_{i=1}^n \left(z_i - \boldsymbol{\alpha}^\top \int_a^b \boldsymbol{\psi}(t) \chi_i(t) dt \right)^2 + \rho \int_a^b [L\boldsymbol{\alpha}^\top \boldsymbol{\psi}(t)]^2 dt, \end{aligned}$$

onde L indica um operador diferencial qualquer aplicado sob a função estimada $\hat{\beta}$, enquanto ρ indica o coeficiente de regularização neste contexto. Ao utilizar-se de fundamentos similares àqueles apresentados na Seção 3.6, é possível definir o seguinte

$$\Theta_i = \int_a^b \boldsymbol{\psi}(t) \chi_i(t) dt \quad (4.10)$$

tão como

$$\mathbf{R} = \int_a^b L\boldsymbol{\psi}(t) L\boldsymbol{\psi}(t)^\top dt. \quad (4.11)$$

Ao diferenciar a soma de erros quadrados penalizada acima com respeito a $\boldsymbol{\alpha}$ e igualar o resultado a 0, obtém-se o estimador mínimos quadrados penalizados como

$$\hat{\boldsymbol{\alpha}}^{\text{Pen}} = (\boldsymbol{\Psi}^\top \boldsymbol{\Psi} + \rho \mathbf{R} \boldsymbol{\alpha})^{-1} \boldsymbol{\Psi}^\top \mathbf{z}, \quad (4.12)$$

onde $\boldsymbol{\Psi} \in \mathbb{R}^{n \times K_\beta}$ é uma matriz tal que $\boldsymbol{\Psi} = [\Theta_1, \Theta_2, \dots, \Theta_n]^\top$. Assim como no contexto de regularização de dados funcionais, é necessário a implementação de alguma medida de bondade de ajuste, como a validação cruzada, para a definição do coeficiente de regularização ideal ρ . Remetendo à forma de validação cruzada vista anteriormente em (3.27), define-se a validação cruzada do modelo funcional como

$$\text{VC}(\lambda) = \frac{1}{n} \sum_{i=1}^n \left[z_i - \int_a^b \hat{\beta}^{(-i)}(t, \rho) \chi_i(t) dt \right]^2,$$

onde $\hat{\beta}^{(-i)}(t, \rho)$ indica a estimativa da curva $\beta(t)$ obtida com coeficiente de regularização ρ numa amostra em qual foi excluída a i -ésima realização.

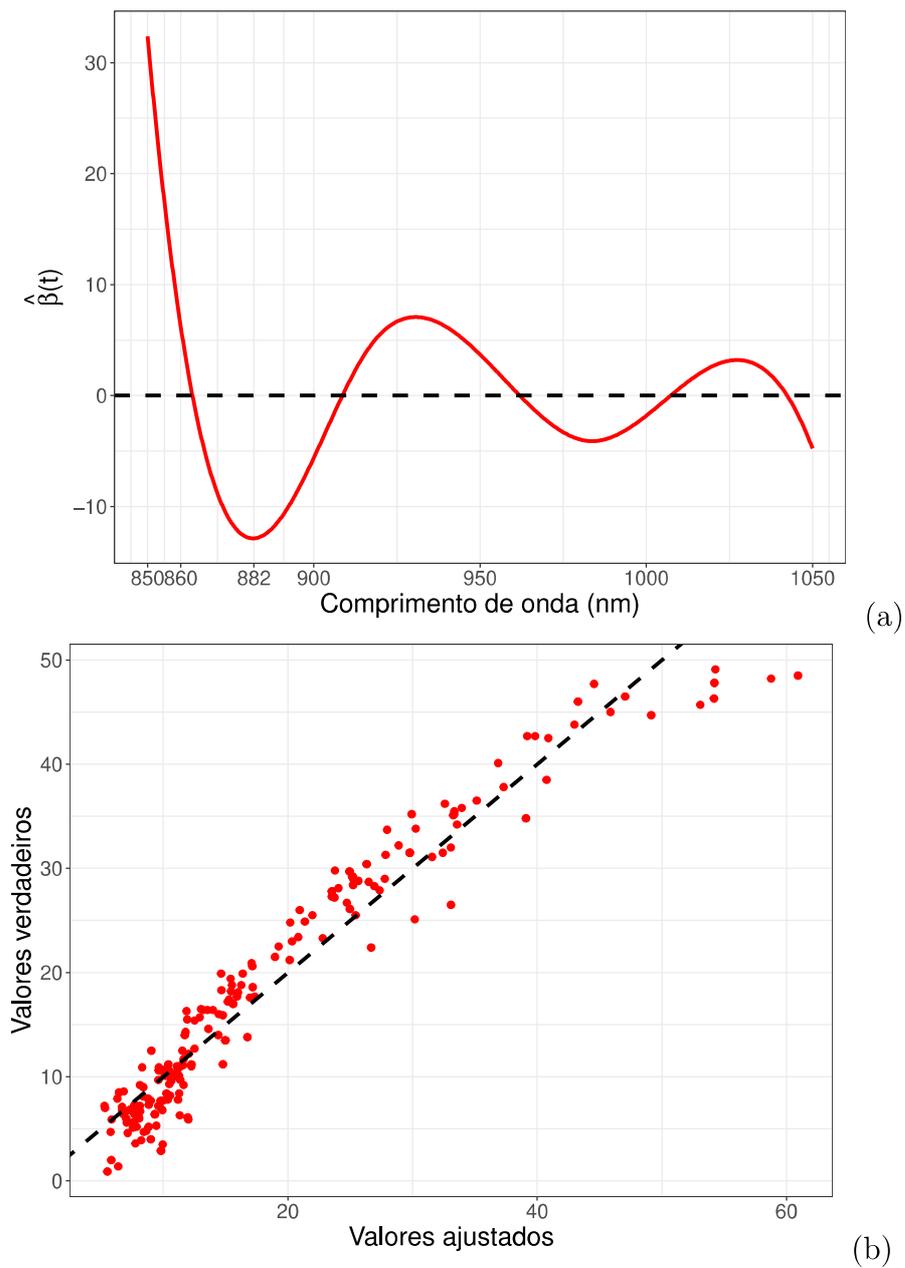
Ilustrativamente, o conjunto de dados supracitado da *Tecator Infratec Food and Feed Analyzer*, qual é composto de 215 curvas quais descrevem a absorvância de um conjunto de carnes picadas quanto submetidas a diferentes comprimentos de onda por um dispositivo analisador de rações. Adicionalmente a estas curvas, para cada carne picada era mensurado seu teor de gordura, em porcentagem. O objetivo do estudo foi então criar um modelo qual fosse capaz de prever o teor de gordura das carnes picadas a partir de suas curvas de absorvância, o que pode ser efetuado através do ponto de vista de um modelo de regressão funcional com variável resposta escalar. As Figuras 4.2-(a) e 4.2-(b) indicam, respectivamente, a curva $\hat{\beta}(t)$ ajustada para este problema, e um gráfico comparativo entre os valores ajustados via o modelo e os valores reais dos teores de gordura para cada carne picada.

Observa-se que valores altos de absorvância entre 850 nanômetros até aproximadamente 860 nanômetros são os mais positivamente correlacionados ao teor de gordura, decaindo até aproximadamente 882 nanômetros, em qual vizinhança a absorvância é mais negativamente correlacionada ao teor de gordura; daí em diante, os valores de $\hat{\beta}(t)$ tendem a oscilar em torno de zero, indicando correlação menos significativa. Neste contexto, seria útil a aplicação de testes de hipóteses para averiguar tal significância, porém estes fogem ao escopo do presente trabalho. É possível também estimar, por exemplo, o coeficiente de determinação de um modelo de regressão funcional como

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

onde $\hat{z}_i = \int_a^b \hat{\beta}(t) \chi_i(t) dt$; observa-se que o coeficiente obtido para o modelo de regressão funcional *Tecator* é aproximadamente de $R^2 = 0,937$.

Figura 4.2 - (a) Curva estimada para $\beta(t)$ para um modelo de regressão funcional do teor de gordura de 215 carnes picadas com respeito à curva de absorvância das mesmas; (b) Valores ajustados do teor de gordura comparados aos valores reais de teor de gordura para este mesmo modelo.



Fonte: Elaborado pelo autor.

5 APLICAÇÕES

Enquanto os Capítulos anteriores foram dedicados à fundamentação de um conjunto de estudos feitos na área de análise de dados funcionais, o presente Capítulo irá focar na implementação destas mesmas metodologias tanto em conjuntos de dados reais quanto em estudos de simulações confeccionados para o mesmo. Particularmente, a Seção 5.1.1 é voltada à aplicação de modelos de regressão funcionais sob um conjuntos de dados real. As Seções 5.1.2 e 5.2.1 são dedicadas ao tópico de agrupamento de dados funcionais.

5.1 CONJUNTOS DE DADOS REAIS

5.1.1 Conjunto de dados de Celulose

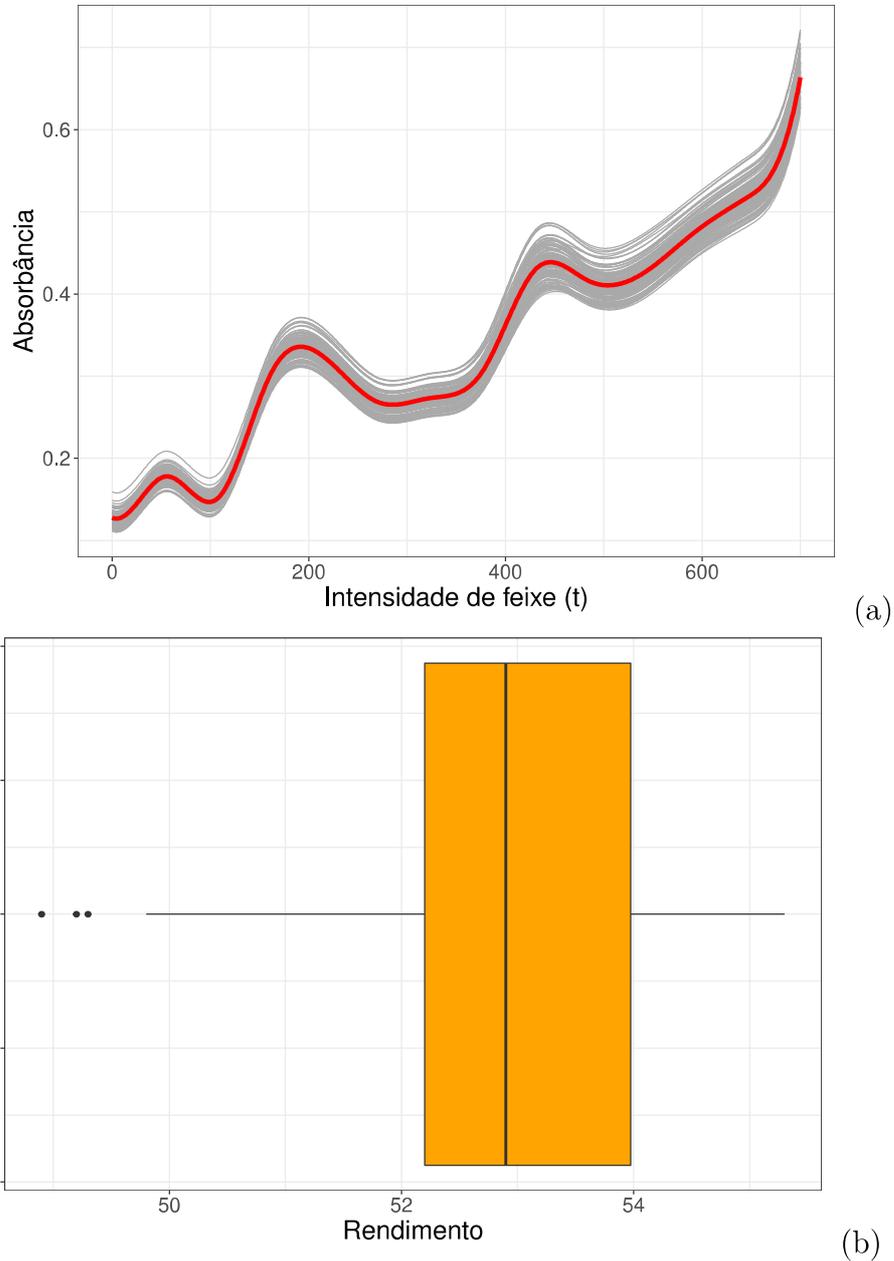
Primeiramente será implementada a técnica de modelos de regressão funcionais sob um base de dados de uma empresa produtora de celulose. Este estudo será feito de forma comparativa àquele feito em Penna (2021, (88)), onde a autora utiliza das técnicas de regularização LASSO e Ridge, tão como redução de dimensionalidade via mínimos quadrados parciais, para modelagem.

O conjunto de dados consiste de 118 árvores, cada qual teve seu rendimento mensurado (em unidade não especificada), e foi submetida a uma forma de tratamento espectroscópico no infravermelho próximo (NIR), qual mediu a absorbância das árvores quando submetidas a feixes de luz de 700 intensidades distintas. Ao utilizar-se das 700 distintas intensidades como variáveis explicativas distintas para modelar-se o rendimento, no contexto de regressão linear multivariada, o problema é dito ser ‘mal-condicionado’, de forma que não é possível construir estimativa via mínimos quadrados tradicionais, apenas através de métodos de regularização ou redução de dimensionalidade.

Sob a estrutura da análise de dados funcionais, as 700 mensurações de absorbância por árvore podem ser utilizadas para a estimação de curvas de absorbância para cada árvore, e assim para a construção de modelos de regressão funcionais onde a variável de interesse é o rendimento de cada árvore. A Figura 5.1-(a) expõe as curvas de absorbância ajustadas para a amostra de 118 árvores, assim como a curva média, enquanto a Figura 5.1-(b) expõe um *boxplot* do rendimento mensurado para as mesmas 118 árvores. As estimativas das curvas foram construídas a partir de 16 funções B-spline, de ordem 4, configuração selecionada de forma empírica, por melhor ajustar-se á regressão funcional, que não reflete perfeitamente o comportamento das curvas aproximadas. Nota-se que a distribuição do rendimento é altamente concentrada entre os valores de 52 e 54, com apenas três valores *outliers* abaixo de 50.

Tratando das observações do rendimento como um vetor de interesse $\mathbf{z} \in \mathbb{R}^{118}$, e as curvas de absorbância vistas na Figura 5.1-(a) como uma amostra de dados funcionais

Figura 5.1 - (a) Curvas de absorvância com respeito a intensidade do feixe estimadas para a amostra de 118 árvores do conjunto de dados Celulose; (b) *Boxplot* para os rendimentos observados para árvores do conjunto de dados de Celulose.



Fonte: Elaborado pelo autor.

$\chi_1, \chi_2, \dots, \chi_{118}$, constrói-se um modelo de regressão funcional como em (4.7), de forma a estimar-se a função de regressão $\beta(t)$. A Figura 5.2-(a) exibe a curva estimada de regressão $\hat{\beta}$, enquanto a Figura 5.2-(b) exibe um gráfico comparativo entre os valores ajustados de rendimento via o modelo de regressão funcional (4.7) e os valores de rendimento verdadeiros. A curva $\hat{\beta}(t)$ estimada exibe, assim como no modelo para o conjunto de

dados da absorvância de carnes picadas supracitado, comportamento oscilatório, embora na Figura 5.2-(a) observa-se que apenas após a intensidade de feixe de 400 a oscilação se aproxima de 0, a ponto de possivelmente indicar não significatividade destes na modelagem do rendimento mensurado das árvores. Novamente, a intensidade de feixe mínima indica maior correlação positiva com rendimento, seguida de uma alta correlação negativa na proximidade da intensidade de feixe 20, e então novamente uma alta correlação positiva próxima da intensidade de feixe 70. Aproximadamente na intensidade de feixe 320 observa-se a maior correlação negativa com o rendimento.

É possível então comparar os resultados aqui encontrados àqueles encontrados em Penna (2021, (88)), ao construir-se uma tabela qual expõe o coeficiente de determinação e erro quadrático médio dos resultados obtidos via estimação de modelos de regressão linear via regularização Ridge e LASSO, e redução de dimensionalidade, como mínimos quadrados parciais (PLS)¹. Neste contexto, o erro quadrático médio é calculado como

$$\text{EQM}(\hat{\mathbf{z}}) = \sum_{i=1}^n (z_i - \hat{z}_i)^2 = (\mathbf{z} - \hat{\mathbf{z}})^\top (\mathbf{z} - \hat{\mathbf{z}}).$$

A Tabela 2 apresenta uma comparação dos métodos LASSO, Ridge, PLS e de regressão funcional (denotado como MRF) quanto ao coeficiente de determinação e erro quadrático médio das estimativas, onde os dois primeiros foram estimados via funções do pacote `glmnet` e `pls` em R, desenvolvidos em Friedman *et al* (2010, 2011, (89, 90)) e Mevik e Wehrens (2007, (91)), respectivamente. Destaca-se que os resultados são similares àqueles vistos em Penna (2021, (88)), uma vez que a regressão via regularização Ridge apresentou pior ajuste, seguida da regularização LASSO, enquanto a redução de dimensionalidade via PLS apresentou melhor ajuste. No entanto, nota-se que o modelo de regressão funcional apresentou ajuste quase tão bom quanto o via mínimos quadrados parciais, indicando seu alto potencial para aplicação em conjuntos de dados de alta dimensão com medidas altamente correlacionadas, como é o caso da área da espectroscopia.

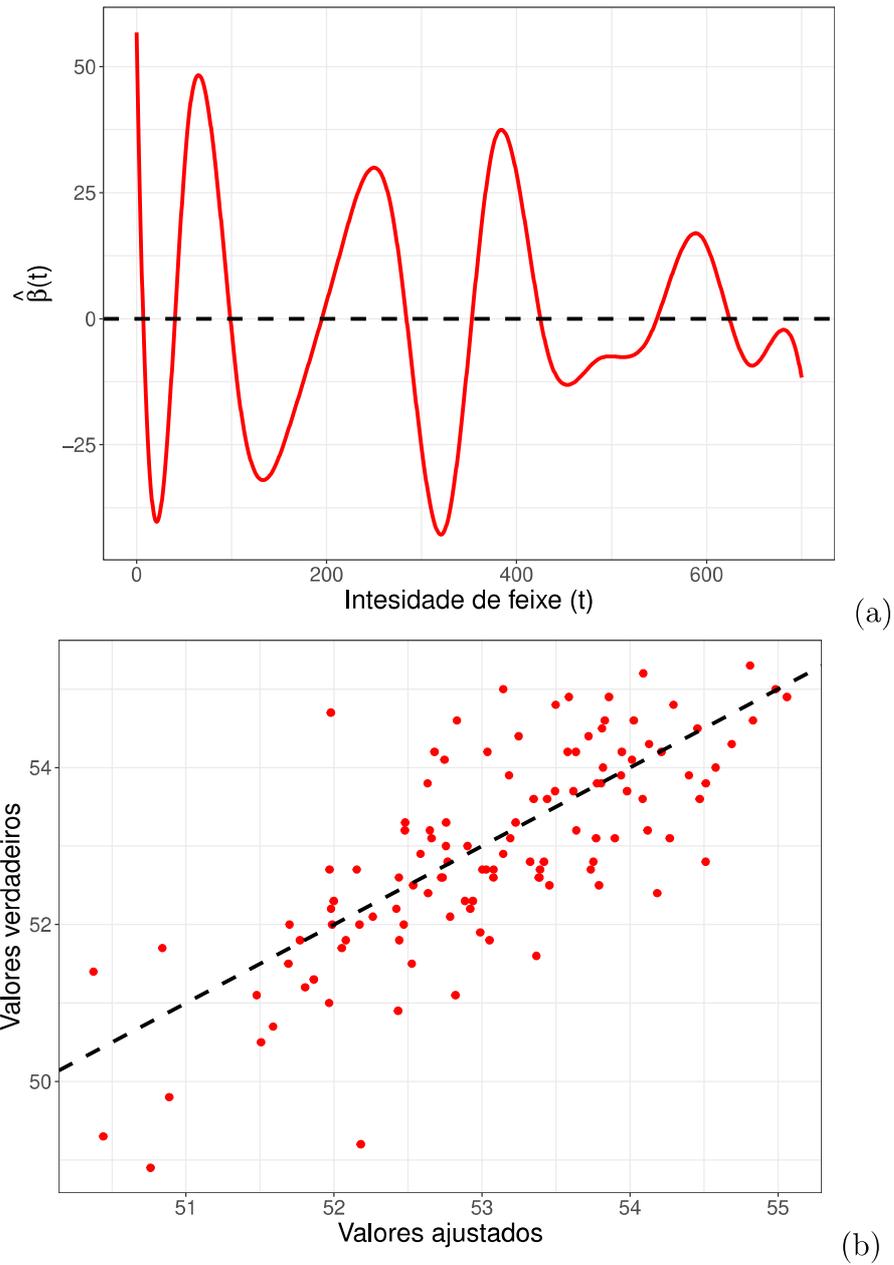
Tabela 2 – Comparação de medidas de qualidade de ajuste para quatro diferentes métodos de modelagem do rendimento mensurado pelas árvores com respeito às absorvâncias por intensidade de feixe distintas, para o conjunto de dados de Celulose.

	LASSO	Ridge	PLS	MRF
Coefficiente de determinação	0,391	0,0391	0,566	0,555
Erro quadrático médio	122,72	193,767	87,449	89,811

Fonte: Elaborado pelo autor.

¹ Do inglês ‘*Partial Least Squares*’.

Figura 5.2 - (a) Curva $\hat{\beta}(t)$ estimada para o modelo de regressão funcional modelando o rendimento mensurado em um conjunto de árvores com respeito às curvas de absorvância por intensidade de feixe das mesmas árvores, no conjunto de dados Celulose; (b) Valores ajustados para o rendimento comparados aos valores reais de rendimento para este mesmo modelo.



Fonte: Elaborado pelo autor.

5.1.2 Conjunto de dados de COVID-19

Para esta aplicação, é abordado um conjunto de dados qual diz respeito ao avanço do Coronavírus por todo o Brasil, disponível em <https://brasil.io/dataset/covid19/>. Este conjunto de dados é composto pelos boletins diários de novos casos confirmados do vírus, tão como mortes em decorrência do mesmo, na grande maioria dos municípios brasileiros. Uma vez que estes dados são repetidamente mensurados por cidade diariamente, eles são apropriados para a aplicação de séries temporais e, por extensão, dados funcionais, como para o presente contexto.

Primeiramente, é necessário manipular e tratar os dados de forma a favorecer a aplicação dos métodos de dados funcionais, tão como de forma a facilitar a visualização gráfica dos resultados. Por parcimônia, foram removidos todos municípios quais contêm menos de meio milhão de habitantes, com base na população estimada de 2019, ou não são capitais estaduais. Sob os dados restantes foi efetuada a limpeza de dados de forma usual, filtrando valores irrealistas, como por exemplo, registros negativos de novos casos confirmados ou mortes. Ademais, para reduzir-se o número de registros por cidade, os dados de boletins diários foram somados por ‘semana epidemiológica’, onde anualmente há cinquenta e duas semanas epidemiológicas.

Feito isso, com intuito de padronizar o ‘eixo temporal’ dos conjuntos de dados, tanto a curva de novos casos confirmados, quanto a curva de mortes em decorrência da doença, foram alteradas de forma que os dados só passam a ser contados na semana em que é contabilizado o primeiro caso em cada qual município, isto é, daqui por diante, ao destacar-se os registros no instante t da i -ésima cidade, refere-se ao dados contabilizados na t -ésima semana após o primeiro caso da doença registrado na i -ésima cidade, incluindo a primeira semana. Para todas as cidades foram estimadas curvas apenas para as 100 primeiras semanas da pandemia. Por fim, para padronização dos próprios valores registrados, o número de casos novos e mortes em decorrência da COVID-19 foram transformados ao formato de casos novos e mortes por milhão de habitantes, onde a estimativa da população em cada município é do ano de 2019.

Optou-se então, antes de construir-se os gráficos das realizações, efetuar um agrupamento das curvas de mortes em decorrência da COVID-19 por milhão de habitantes. Para este modelo, foram selecionados dez grupos para o agrupamento, utilizando como medida de dissimilaridade a diferença quadrática integrada e do método de ligação máxima. As curvas foram estimadas via suavização positiva, uma vez que os registros tem de ser positivos, com 104 funções base B-splines da sexta ordem penalizadas sob a quarta derivada, e com coeficiente de regularização $\lambda = 0,05$; estas configurações foram selecionadas de forma empírica.

A Figura 5.3 expõe as curvas de mortes por milhão de habitantes em decorrência da COVID-19. Observa-se que, embora dez grupos tenham sido selecionados para o

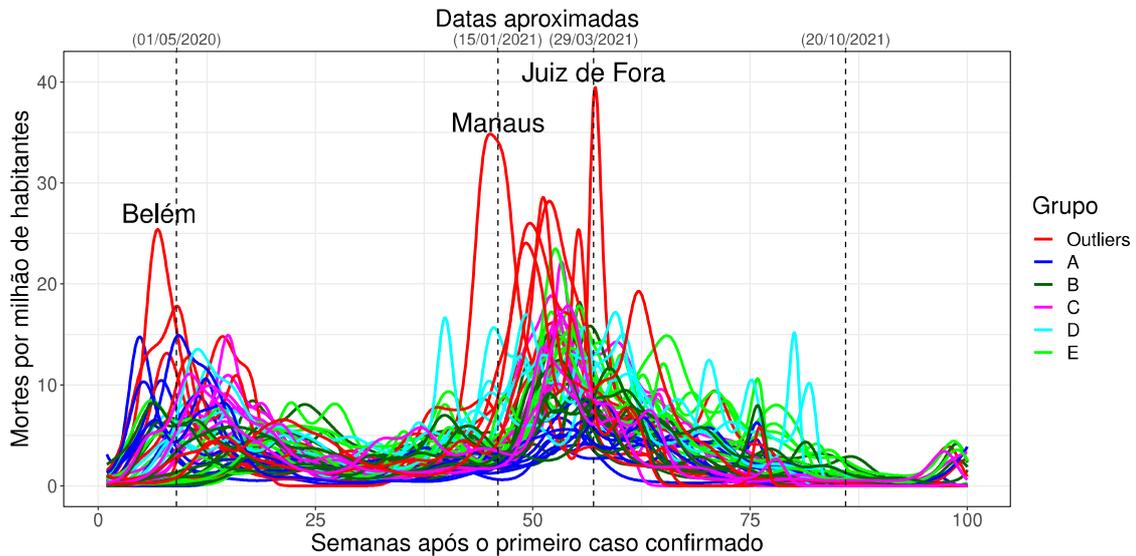
agrupamento, destes dez grupos, cinco eram compostos de apenas uma ou duas realizações. Isto aponta então que as dissimilaridades entre estes grupos e os demais são significativas, de forma que as realizações pertencentes a estes grupos ‘pequenos’ são consideradas *outliers*. Nota-se então, a partir dos dados, que as realizações pertencentes ao grupo de *outliers* apresentam picos em pontos no número de mortes por milhão de habitantes em instantes distintos das demais, possivelmente razão por qual são consideradas realizações ‘aberrantes’.

Têm-se a curva mortes de Belém, no estado do Pará, qual apresentou seu primeiro pico de mortes logo após a confirmação do primeiro caso de COVID-19 na cidade. De fato, em Limão e Carneiro (2020, (92)), uma notícia publicada no portal de notícias G1 no dia primeiro de maio de 2020, os autores ressaltam que, de uma semana à seguinte, o número de mortes em decorrência da COVID-19 no Pará havia triplicado, com metade destas mortes atribuídas à capital, aproximadamente coincidindo com o pico observado. O pico observado posteriormente para o município de Manaus, no estado do Amazonas, por sua vez, pode ser explicado por coincidir com a crise do oxigênio por qual o município passou. De fato, uma notícia publicada também no portal de notícias G1 (2021, (93)) datada ao dia 15 de janeiro de 2021, descreve a tragédia, qual também coincide aproximadamente com o pico observado. Por fim, o pico observado posteriormente em Juiz de Fora, no estado de Minas Gerais, embora não necessariamente diretamente explicado por tal, coincide com um marco de altas em morte no Brasil em decorrência da COVID-19. Como visto em uma retrospectiva em Madeiro (2022, (94)), notícia publicada no portal de notícias Uol, no dia 29 de março de 2021 foram registradas 3541 mortes em todo o Brasil, o máximo até e desde então, também coincidindo aproximadamente com o pico de mortes no município de Juiz de Fora.

Observam-se também marcos positivos, tais como visto em Tavares (2021, (95)), uma notícia publicada no portal de notícias Correio Braziliense, no dia 10 de outubro de 2021, o Brasil chegou à atingir 50% da população já vacinada com pelo menos duas doses contra o vírus COVID-19. Nota-se que, após esta semana, não apenas o número de mortes por milhão de habitantes decresce, como também sua variância parece decrescer. Embora não seja possível extrair conclusões diretamente desta aparente relação, ela representa potenciais hipóteses quais podem ser postuladas e testadas via a análise de dados funcionais. Ademais, uma notícia publicada no portal de notícias CNN (2021, (96)), qual remete a uma pesquisa sobre as cidades mais vulneráveis e quais melhor combateram as mortes pela COVID-19, aponta entre as cinco capitais mais eficazes os municípios de Florianópolis, São Paulo, Palmas, Belo Horizonte e Curitiba; na Tabela 3 observa-se o resultado dos agrupamentos por cidade, onde nota-se que destas, apenas Curitiba não foi atribuída ao grupo B.

Já nas Figuras 5.4-(a) e 5.4-(b), estão dispostas, respectivamente, a curva de média e de variância por grupo das mortes por milhão de habitantes por COVID em cada cidade.

Figura 5.3 - Curvas do número de mortes por milhão de habitantes em decorrência da COVID-19 em capitais e cidades de grande porte brasileiras, já agrupadas. Destacam-se os picos dos municípios de Belém (PA) Manaus (AM) e Juiz de Fora (MG).



Fonte: Elaborado pelo autor.

Destas, é possível aferir certas conclusões quanto aos padrões observados em cada grupo: as realizações pertencentes ao grupo A e B apresentam curva média inferior às demais, potencialmente indicando que cidades atribuídas a estes grupos foram eficazes na prevenção e tratamento da COVID-19. Remete-se à notícia que indicava quatro integrantes do grupo B como entre as capitais que melhor combateram o vírus. Ademais, observa-se que o grupo E destaca-se não apenas por apresentar o maior pico com respeito à média de mortes por milhão de habitantes, como também pela sua alta variância quando comparado aos demais grupos, como visto na Figura 5.4-(b). Por sua vez, observa-se que no grupo D, as únicas cidades quais apresentam coeficiente silhueta acima de 0, são Rio de Janeiro e Niterói, cidades vizinhas no estado do Rio de Janeiro, indicando que a proximidade espacial destes municípios potencialmente ocasionou curvas de mortes por milhões de habitantes similares.

Nota-se, porém, que a proximidade espacial não apresenta demasiada influência nos resultados do agrupamento: cidades da região metropolitana do Rio de Janeiro, como Belford Roxo, Duque de Caxias e Nova Iguaçu, embora todas atribuídas ao grupo A, não foram agrupadas com a própria cidade do Rio de Janeiro. Similarmente Contagem, pertencente à região metropolitana de Belo Horizonte, não foi agrupada com Belo Horizonte. Adicionalmente, para a região metropolitana de São Paulo, composta, entre outros, por Guarulhos, Osasco, Santo André e São Bernardo do Campo, nota-se que apenas Osasco e São Bernardo do campo foram atribuídos ao mesmo grupo, no caso o grupo E, embora

ambos apresentem coeficiente de silhueta negativo, e nenhuma destas cidades foi agrupada com a própria cidade de São Paulo.

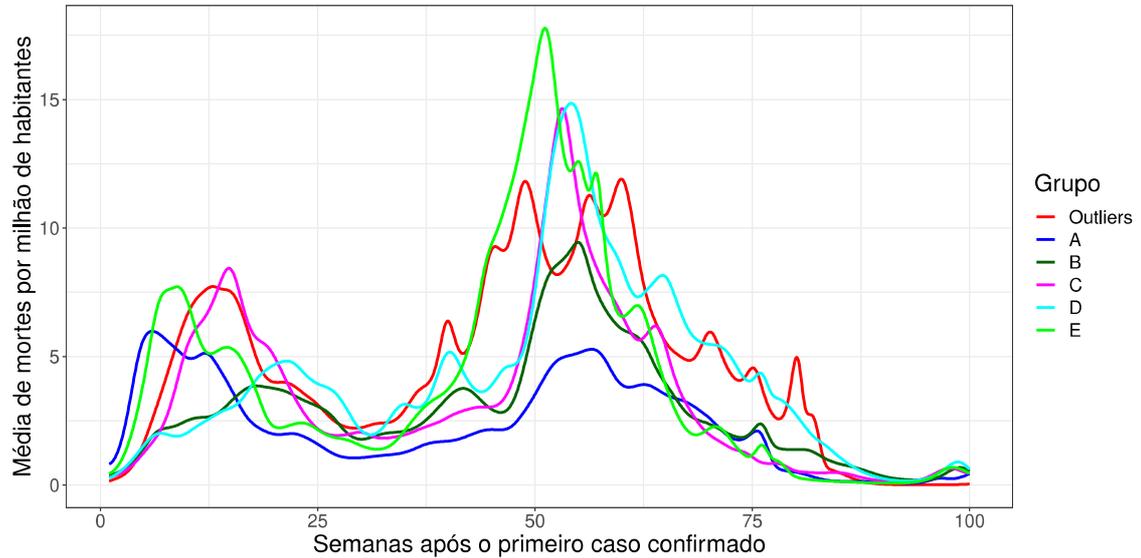
Tabela 3 – Atribuições por grupo do agrupamento realizado sobre as curvas de morte por milhão de habitantes em decorrência da doença COVID-19, com medida de silhueta s_i associada a cada município.

A	s_i	B	s_i
Ananindeua	0,337	Aparecida de Goiânia	0,0171
Belford Roxo	0,421	Belo Horizonte	0,265
Duque de Caxias	0,524	Brasília	0,0755
Feira de Santana	-0,00142	Caxias do Sul	0,0641
Jaboatão dos Guararapes	0,522	Florianópolis	0,15
Maceió	0,108	Macapá	-0,113
Nova Iguaçu	0,277	Palmas	0,271
Recife	0,272	São Gonçalo	-0,0798
São Luís	0,494	São José dos Campos	0,276
-		São Paulo	-0,0419
-		Serra	-0,188
C	s_i	D	s_i
Aracaju	0,237	Boa Vista	-0,0585
Campinas	0,186	Niterói	0,128
Contagem	-0,00135	Ribeirão Preto	-0,0376
Guarulhos	0,389	Rio de Janeiro	0,242
João Pessoa	0,414	Santo André	-0,0073
Natal	0,375	-	
Salvador	0,0423	-	
Teresina	0,408	-	
Vitória	0,313	-	
E	s_i	<i>Outliers</i>	
Campo Grande	0,398	Belém	
Campos dos Goytacazes	0,115	Cuiabá	
Curitiba	0,353	Fortaleza	
Goiânia	0,413	Juiz de Fora	
Joinville	0,0508	Manaus	
Londrina	0,313	Porto Velho	
Osasco	-0,157	Sorocaba	
Porto Alegre	0,144	Uberlândia	
São Bernardo do Campo	-0,269	-	

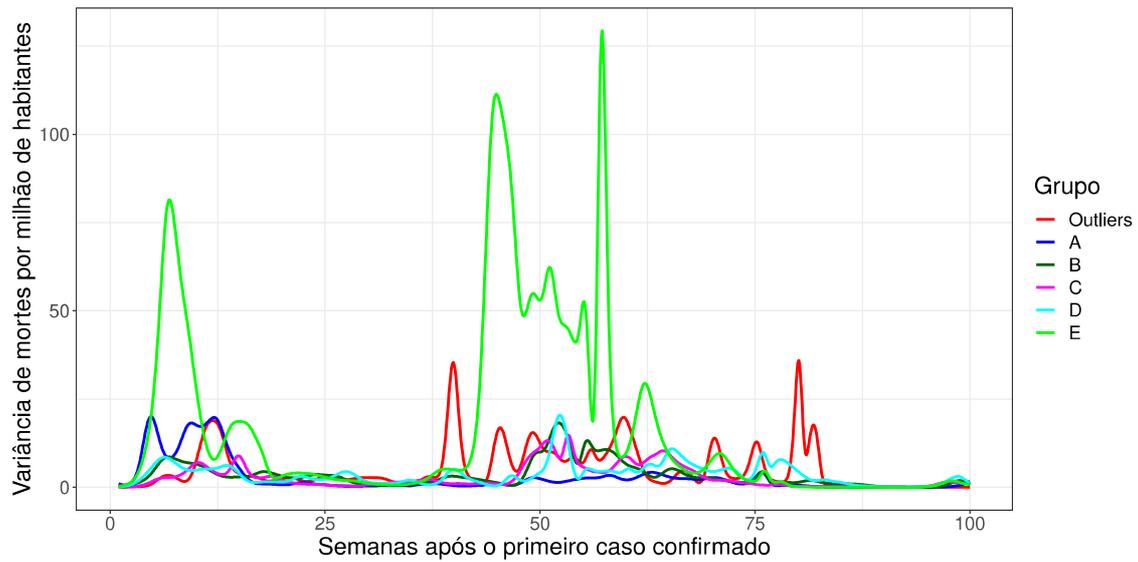
Fonte: Elaborado pelo autor.

Por fim, observa-se na Figura 5.5-(a) as derivadas com respeito às semanas das curvas do número de mortes por milhão de habitantes em decorrência da COVID-19. Embora para esta aplicação tenha sido utilizado o agrupamento sobre as curvas originais, é possível, e em certos contextos é até mais útil, efetuar o agrupamento sobre a derivada destas curvas, uma vez que estas podem apontar a ‘velocidade de crescimento’, por exemplo, do número de mortes por milhão de habitantes. Desta forma, o agrupamento das curvas derivadas é ideal para, por exemplo, a previsão de novas ondas da doença.

Figura 5.4 - (a) Curvas da média por grupo do número de mortes por milhão de habitantes em decorrência da COVID-19 em capitais e cidades de grande porte brasileiras; (b) Curva da variância por grupo do número de mortes por milhão de habitantes em decorrência da COVID-19.



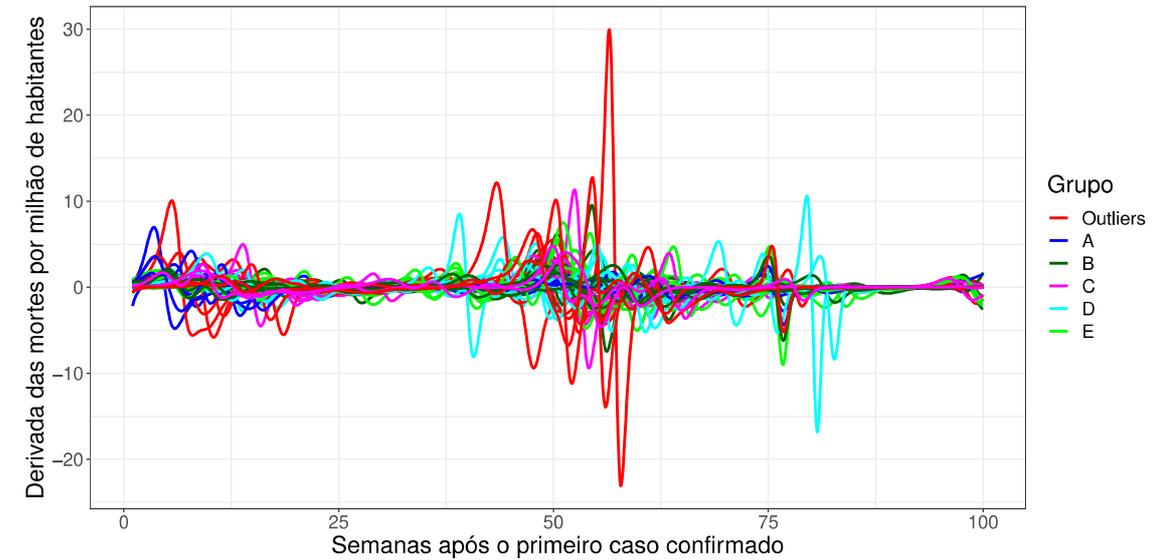
(a)



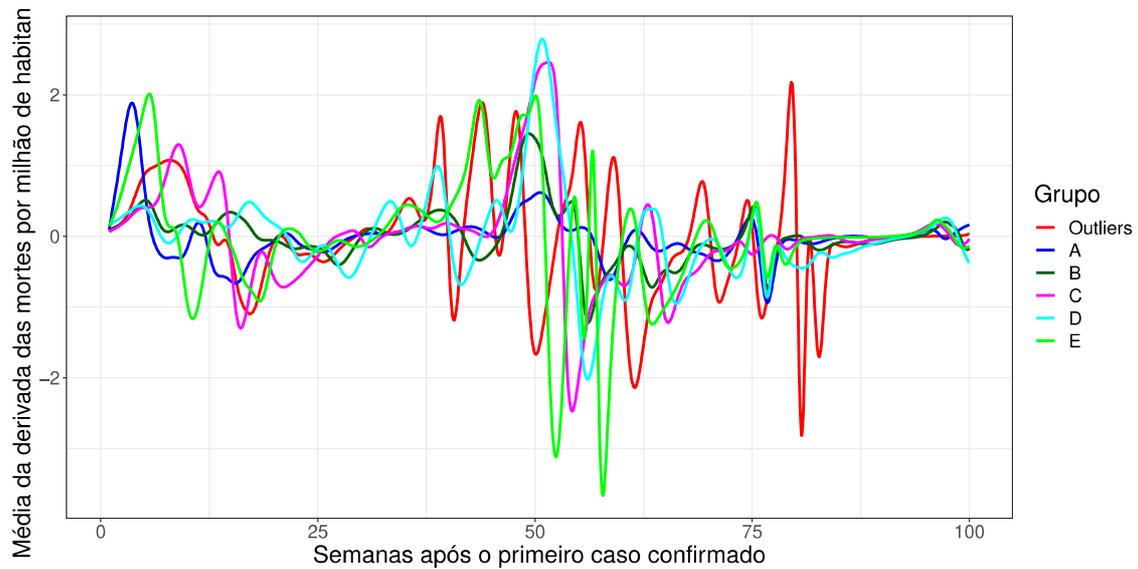
(b)

Fonte: Elaborado pelo autor.

Figura 5.5 - (a) Derivadas das curvas do número de mortes por milhão de habitantes em decorrência da COVID-19 em capitais e cidades de grande porte brasileiras, já agrupadas; (b) Curvas médias por grupo da derivada das curvas de número de mortes por milhão de habitantes em decorrência da COVID-19.



(a)



(b)

Fonte: Elaborado pelo autor.

5.2 ESTUDOS DE SIMULAÇÃO

5.2.1 Agrupamento simulado

Nesta seção, apresentam-se dois estudos de cenários simulados para os métodos de agrupamento, procurando demonstrar a importância de um conceito relevante na área

de dados funcionais, denominado ‘deslocamento de fase’². Como destacado em Anselmo (2005, (69)), grande parte dos métodos de análise de dados funcionais dependem da pressuposição que não há deslocamento de fase entre todas as realizações de uma amostra de dados funcionais, onde pode-se dizer que um conjunto de dados aleatórios funcionais $\chi_1, \chi_2, \dots, \chi_n$ definidos no intervalo $[a, b]$ não apresenta deslocamento de fase se o estado de todas estas realizações no instante a é equivalente. Nota-se que em Ramsay e Li (1998, (97)) os autores apresentam um método para padronização de casos com deslocamento de fase, denominado ‘*registration*’.

5.2.1.1 Caso sem deslocamento de fase

Inicialmente, são geradas 100 curvas, cada qual com 100 mensurações distribuídas de forma equidistante entre os instantes 0 e 3, e pertencente a um de cinco grupos, onde as funções geradas são da forma:

$$f_i(t_j) = (i - t_j)\mathbf{1}_{[i-1, i)}(t_j) + \mathbf{1}_{[i-2, i-1)}(t_j) + (t_j - i + 3)\mathbf{1}_{[i-3, i-2)}(t_j) + e_{k,j}$$

$$= e_{k,j} + \begin{cases} (i - t_j), & \text{se } i - 1 \leq t_j < i \\ 1, & \text{se } i - 2 \leq t_j < i - 1 \\ (t_j - i + 3), & \text{se } i - 3 \leq t_j < i - 2 \\ 0, & \text{caso contrário,} \end{cases} \quad (5.1)$$

Em que $i \in \{1, 2, 3, 4, 5\}$ indica o índice da função gerada para cada um dos cinco grupos, enquanto $e_{k,j}$ indica o erro da j -ésima mensuração para a k -ésima curva simulada, onde $e_{k,j} \sim \mathcal{N}(0, 0.05^2)$. Uma vez geradas as 100 realizações, elas foram suavizadas através de 102 funções base B-splines de ordem quatro, com regularização incidindo sob a segunda derivada e coeficiente de regularização $\lambda = 0,05$.

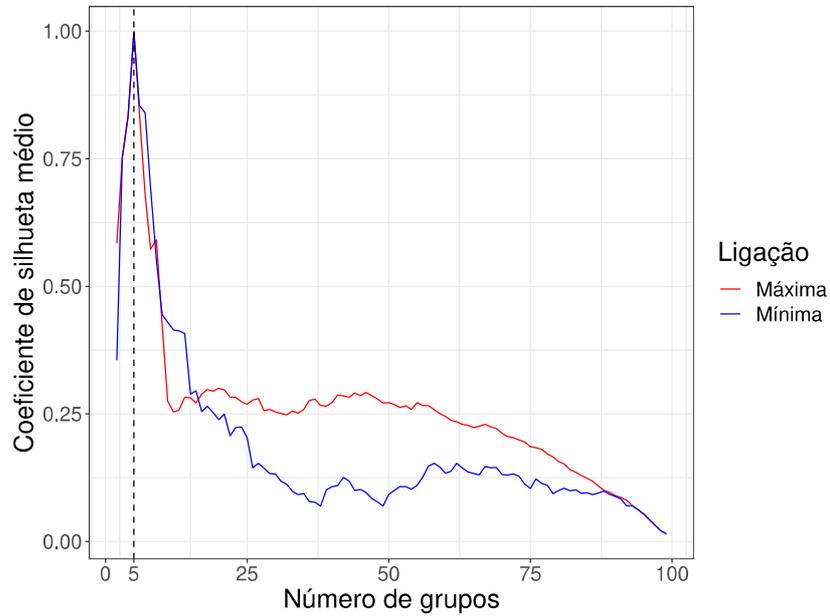
Em seguida, foi efetuado o agrupamento dos dados, utilizando como medida de dissimilaridade a diferença quadrática integrada. A Figura 5.6-(a) expõe os coeficientes de silhueta médios estimados para cada quantia de grupos. Nota-se que o número de grupos qual incorreu máximo para este foram cinco grupos, para qual o coeficiente de silhueta médio foi $\bar{s} = 0,999$. Observa-se que, neste contexto, todas as realizações foram ‘corretamente’ agrupadas, isto é, os cinco grupos gerados via agrupamento captaram perfeitamente os cinco distintos padrões de curvas gerados. A Figura 5.6-(b) expõe então o resultado do agrupamento, novamente salientando o bom funcionamento do algoritmo quando não há deslocamento de fase nos dados.

5.2.1.2 Caso com deslocamento de fase

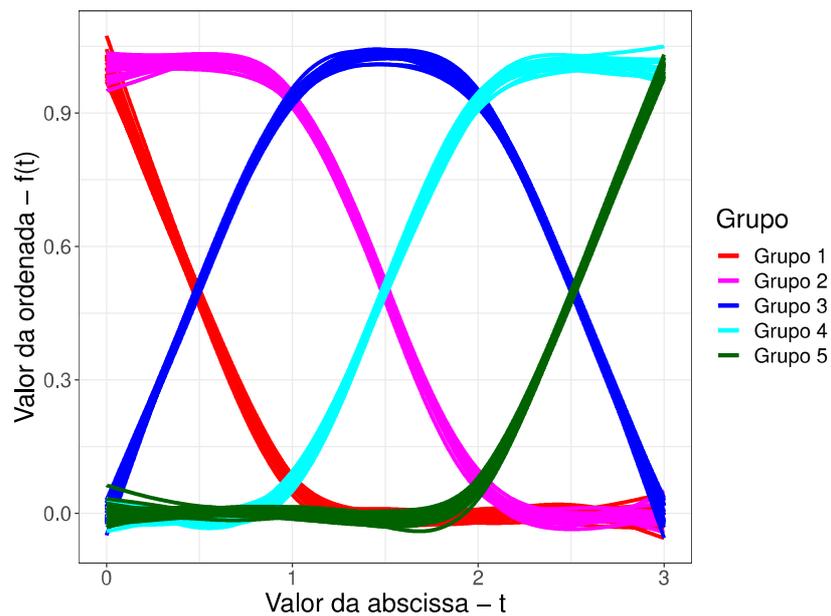
Enquanto o exemplo anterior demonstra o bom funcionamento do agrupamento quando a pressuposição em respeito à ausência de deslocamento de fase é respeitada, este

² Do inglês ‘phase shift’.

Figura 5.6 - (a) Coeficientes de silhueta médios estimados para as dissimilaridades das curvas geradas conforme (5.1), com método de ligação máxima e mínima; (b) Realizações simuladas conforme (5.1), já agrupadas.



(a)



(a)

Fonte: Elaborado pelo autor.

segundo exemplo, exemplo busca demonstrar as consequências quando o deslocamento de fase não é respeitado. Similarmente ao anterior, são geradas 100 curvas, cada qual com 100 mensurações distribuídas de forma equidistante entre os instantes 0 e 3, e pertencente

a um de cinco grupos, onde as funções geradas são da forma:

$$\begin{aligned}
 f_i(t_j) &= (U_{i,k} - t_j)\mathbf{1}_{[U_{i,k}-1, U_{i,k}]}(t_j) + \mathbf{1}_{[U_{i,k}-2, U_{i,k}-1]}(t_j) + (t_j - U_{i,k} + 3)\mathbf{1}_{[U_{i,k}-3, U_{i,k}-2]}(t_j) + e_{k,j} \\
 &= e_{k,j} + \begin{cases} (U_{i,k} - t_j), & \text{se } U_{i,k} - 1 \leq t_j < U_{i,k} \\ 1, & \text{se } U_{i,k} - 2 \leq t_j < U_{i,k} - 1 \\ (t_j - U_{i,k} + 3), & \text{se } U_{i,k} - 3 \leq t_j < U_{i,k} - 2 \\ 0, & \text{caso contrário,} \end{cases}
 \end{aligned} \tag{5.2}$$

onde $i \in \{1, 2, 3, 4, 5\}$ indica o índice da função gerada para cada um dos cinco grupos, enquanto $e_{k,j}$ indica o erro da j -ésima mensuração para a k -ésima curva simulada, onde $e_{k,j} \sim \mathcal{N}(0, 0.05^2)$. Adicionalmente, para causar o deslocamento de fase, os termos que em (5.1) foram escritos como i , nestes experimentos serão gerados através de variáveis aleatórias uniforme $U_{i,k} \sim \mathcal{U}(i - h, i + h)$, onde h será denominado ‘fator de deslocamento de curva’, tal que $U_{i,k}$ denotará o deslocamento da k -ésima realização, pertencente ao i -ésimo grupo, de forma a horizontalmente deslocar as realizações de forma aleatória. Trivialmente, quando $h = 0$, tem-se o caso sem deslocamento de curva, enquanto que, conforme o h aumentar, maior será deslocamento de curva. Utilizou-se primeiramente $h = 1/2$. Neste contexto, via o método de ligação mínima foi 25, com coeficiente de silhueta médio $\bar{s} = 0,693$, enquanto via o método de ligação máxima foi 19, com coeficiente de silhueta médio $\bar{s} = 0,756$. Observa-se que ambos números superestimam significativamente o número real de grupos quais geraram os dados. Se de toda forma forem utilizados cinco grupos para o agrupamento, a Tabela 4 expõe um cruzamento entre os grupos reais de cada realização e os grupos a quais estas foram atribuídas. Embora nenhum grupo tenha

Tabela 4 – Tabela de contingência comparando os grupos reais de cada realização e os grupos a quais elas foram atribuídas, via método de ligação completa.

		Grupo real				
		1	2	3	4	5
Grupo atribuído	A	18	0	0	0	0
	B	2	20	8	0	0
	C	0	0	8	0	0
	D	0	0	4	6	0
	E	0	0	0	14	20

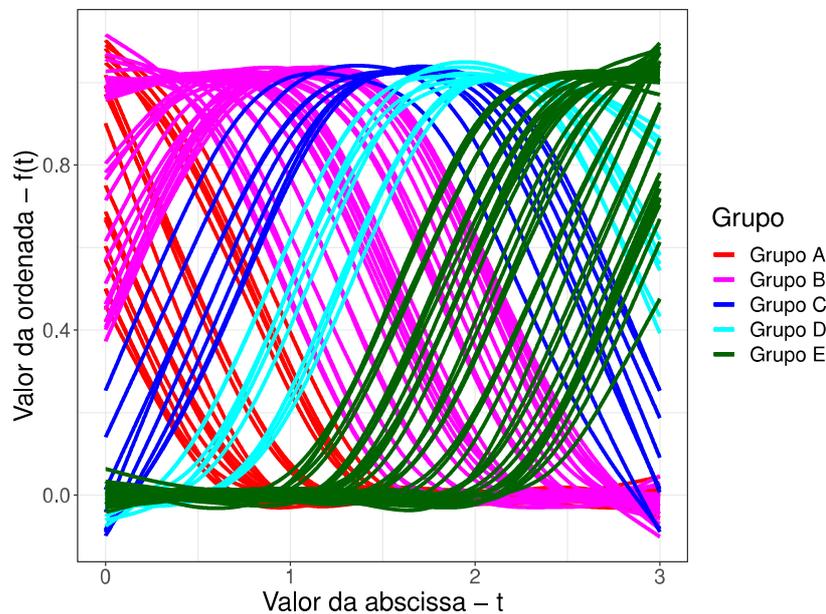
Fonte: Elaborado pelo autor.

captado todo e apenas todo o padrão de algum dos grupos, os grupos A, B e E mostram forte correlação com os grupos reais 1, 2 e 5, respectivamente. Nota-se, além disto, que, para o modelo gerado em (5.2), grupos reais vizinhos (por exemplo 1 e 2, 2 e 3, etc.) podem ser aproximados por consequência do deslocamento de fase, o que justifica o fato do grupo B agrupar realizações originalmente dos grupos 1, 2, 3, assim como o grupo D

agrupar realizações originalmente dos grupos 3 e 4, e grupo E realizações originalmente dos grupos 4 e 5. De toda forma, observa-se que a introdução do deslocamento de fase prejudica o agrupamento.

A Figura 5.7 expõe os efeitos do deslocamento de fase, demonstrando então como este fenômeno pode negativamente impactar não apenas a aplicação de métodos de análise de dados funcionais como o agrupamento, como também a própria visualização gráfica dos dados, dificultando a capacidade de visualmente observar-se um padrão.

Figura 5.7 - Realizações simuladas conforme (5.2), com deslocamento de fase, e fator de deslocamento de fase $h = 1/2$, já agrupadas.



Fonte: Elaborado pelo autor.

Como destacado anteriormente, em Ramsay e Li (1998, (97)), os autores propõem técnicas computacionais para tratamento do deslocamento de fase, denominada *'registering'*, porém nem sempre esta será necessária. Por exemplo, na aplicação na Seção 5.1.2, para tratar-se do possível deslocamento de fase, todas as curvas foram transformadas de forma que o instante t refere-se à t -ésima semana após o primeiro caso confirmado de COVID-19 em cada cidade. Porém, padronizações desta forma dependem do contexto e características dos dados em questão, e devem ser implementadas com cuidado.

6 CONCLUSÃO

6.1 CONSIDERAÇÕES FINAIS

O intuito do presente trabalho foi então formular uma introdução a um destes objetos: o dado funcional. A Análise de Dados Funcionais, embora fundamentada em tópicos abordados até mesmo na primeira metade do século XX, só teve seu nome estabelecido em data tão recente quanto 1991. Seu desenvolvimento é apresentado e contextualizado à área mais generalizada de estimação não paramétrica de curvas, com qual um aluno de graduação em Estatística não é usualmente familiarizado. Em conclusão, espera-se que este trabalho fomente entre alunos da graduação interesse nesta áreas com quais tem tão pouco contato.

Além de estudos com dados espacialmente ou temporalmente indexados, aplicações utilizando modelos de regressão funcionais, como a vista na Seção 5.1.1, demonstram a versatilidade da ADF em grandes bases de dados com variáveis altamente correlacionadas, possivelmente como alternativa aos métodos de redução de dimensionalidade ou regularização comumente utilizados. Nota-se porém, que até conjuntos de dados como o conjunto de dados da COVID-19, visto na Seção 5.1.2, podem valer da transferibilidade supracitada entre a Análise de Dados Funcionais e a Análise de Dados Multivariados, onde estudos em análise de agrupamentos ou análise de componentes principais podem ser feitos. Pela pluralidade de referências disponíveis, é possível afirmar que há ainda diversas outras formas de aplicações na área de ADF, quais de toda forma houveram de ser omissos no presente trabalho.

Uma das grandes dificuldades no desenvolvimento deste trabalho foi a baixa praticidade computacional do mesmo, uma vez que o armazenamento e manipulação de dados funcionais equivalem a estes mesmos procedimentos realizados sob grandes quantias de dados discretos. No contexto de dados funcionais, o cálculo das matrizes de dissimilaridades para o agrupamento de n realizações, como visto nas Seções 5.1.2 e 5.2.1, requer o cálculo de $\frac{n(n+1)}{2}$ integrais, quais normalmente são estimadas computacionalmente. Adicionalmente, embora o pacote `fda`, disponível em linguagem R, tenha certamente facilitado a confecção do trabalho, este pacote é notadamente ultrapassado em algumas de suas funcionalidades, e não é de simples uso.

6.2 ESTUDOS FUTUROS

Nesta seção, denotam-se aplicações futuras na área de análise de dados funcionais, quais podem conferir continuidade aos conceitos apresentados no presente trabalho.

6.2.1 Modelos de regressão funcionais com respostas funcionais

Enquanto o modelo de regressão funcional apresentado na Seção 4.3.1 é limitado no sentido de considerar apenas variáveis resposta escalares, em Ramsay (2005, (37), pp. 220–221), o autor apresenta também modelos de regressão funcionais onde as variáveis resposta são também funcionais. Diferentemente do caso onde a variável resposta é escalar, estes apresentam diversas formas de aplicação.

Dado duas amostras de dados funcionais $\chi_1, \chi_2, \dots, \chi_n$, distribuída conforme \mathcal{X} , e $\vartheta_1, \vartheta_2, \dots, \vartheta_n$, distribuída conforme \mathcal{V} , onde tanto \mathcal{X} quanto \mathcal{V} estão definidos no intervalo $[a, b]$, tem-se que o modelo ‘concomitante’¹ é escrito como

$$\vartheta_i(t) = \beta(t)\chi_i(t) + e_i(t), \quad \forall i \in \{1, 2, \dots, n\}, \quad (6.1)$$

onde este modelo parte do pressuposto que o dado funcional ϑ_i , no instante t , pode ser modelado apenas pelo valores do dado funcional χ_i no mesmo instante; assim como no contexto de 4.7, $\beta(t)$ é uma função definida em $[a, b]$.

Há ainda o modelo de regressão funcional ‘total’, qual escreve-se

$$\vartheta_i(t) = \int_a^b \beta(s, t)\chi_i(s)ds + e_i(t), \quad \forall i \in \{1, 2, \dots, n\}, \quad (6.2)$$

onde este modelo busca estender a influência do ponto t do dado funcional χ_i sob todo o intervalo do dado funcional a ser modelado ϑ_i . Observa-se neste contexto que a função $\beta(s, t)$ a ser estimada será então uma superfície definida em $[a, b] \times [a, b]$, e conseqüentemente devem ser abordadas novas técnicas para sua estimação.

Por fim, de forma geral, é possível definir o modelo de regressão funcional denominado ‘modelo de influência local’, qual escreve-se

$$\vartheta_i(t) = \int_{\Theta_t} \beta(s, t)\chi_i(s)ds + e_i(t), \quad \forall i \in \{1, 2, \dots, n\}, \quad (6.3)$$

onde Θ_t denota um subconjunto de $[a, b]$ dependente de t . Este formato pode ser generalizado ao modelo de regressão funcional ‘total’ visto em (6.2), basta definir $\forall t \in [a, b]$, $\Theta_t = [a, b]$. Naturalmente, neste contexto, a facilidade de estimação da superfície $\beta(s, t)$ é dependente do formato imposto pelo modelo a Θ_t .

REssalta-se que, a partir deste modelo, é possível estender a aplicação vista na Seção 5.1.2, com respeito às curvas de mortes por milhão de habitantes em decorrência da COVID-19, ao modelar a curva de mortes via as curvas de casos novos. Uma vez que há estudos que apontam um período de uma a duas semanas entre a detecção da doença e morte, utilizando os mesmos dados da Seção 5.1.2, é possível definir $\Theta_t = [t - 2, t]$, e então implementar um modelo como em (6.3). É esperado que assim, os valores $\beta(s, t)$ apontem intensidade com que casos novos confirmados impactam o número de mortos nas duas semanas seguintes aos seus registros.

¹ Traduzido do inglês ‘concurrent’.

6.2.2 Regularização de curvas com operadores diferenciais não lineares

Resgata-se então um resultado visto em Wahba (1990, (58)) e supracitado em (3.26), qual diz

$$\int_a^b \text{Viés}[\hat{\chi}(t)]^2 dt \leq \int_a^b [L(\hat{\chi})(t)]^2 dt$$

onde $L(\hat{\chi})$ indica um operador diferencial linear utilizado para regularização da estimação da curva χ . Desta forma, caso exista conhecimento prévio da curva χ qual indica que esta é, aproximadamente, solução de uma equação diferencial linear homogênea do tipo

$$a_0\chi + a_1D\chi(t) + a_2D^2\chi(t) + \dots + a_kD^k\chi(t) = 0, \quad (6.4)$$

onde $[a_0, a_1, a_2, \dots, a_k] \in \mathbb{R}^{k+1}$ são coeficientes fixos, é possível selecionar operador diferencial linear $L(\chi)$ igual à parte esquerda de (6.4), tal que a relação do viés integrado vista acima aproxima-se de zero.

Embora este resultado apresente suficiente flexibilidade, é possível ainda estendê-lo ao considerar operadores diferenciais não lineares. Por exemplo, tem-se a curva de Gompertz, assim nomeada em Gompertz (1815, (98)), comumente utilizada para modelagem de funções de crescimento, como em Winsor (1932, (99)). Esta pode ser obtida como solução da seguinte equação diferencial

$$D\chi(t) - \beta \log\left(\frac{\kappa}{\chi(t)}\right)\chi(t) = 0, \quad (6.5)$$

onde $\beta \in \mathbb{R}$ e $\kappa > 0$. Supondo que a relação em (3.26) mantém-se válida para operadores diferenciais não lineares, utilizando a parte esquerda da equação em (6.5) como regularização à estimação de uma curva χ , quando presume-se que esta é aproximadamente Gompertz com parâmetros β e $\kappa > 0$, é possível então minimizar o viés da estimativa. A mesma lógica pode ser aplicada à função logística generalizada obtida via a equação diferencial de Richards, vista em Richards (1959, (100)).

No todo, uma boa sequência ao estudo do presente trabalho seria fundamentar métodos que busquem selecionar o operador diferencial ideal para regularização, que seria capaz de minimizar o viés.

REFERÊNCIAS

- 1 A. F. Elétrico, “Saudade,” 2019. Álbum: Little Electric Chicken Heart.
- 2 F. Dyson, “A meeting with enrico fermi.,” *Nature*, vol. 427, p. 297, Jan 2004.
- 3 J. W. Tukey, “The future of data analysis,” *The Annals of Mathematical Statistics*, pp. 1–67, 1962.
- 4 J. O. Ramsay, G. Hooker, and S. Graves, *Functional data analysis with R and MATLAB*. Springer, 2009.
- 5 P. Diggle, P. Heagerty, K.-Y. Liang, and S. Zeger, *Analysis of longitudinal data*. No. 25, Oxford University Press, 2002.
- 6 C. R. Rao, “Some statistical methods for comparison of growth curves,” *Biometrics*, vol. 14, no. 1, pp. 1–17, 1958.
- 7 F. Ferraty and P. Vieu, *Nonparametric functional data analysis: theory and practice*. Springer, 2006.
- 8 U. Grenander, “Stochastic processes and statistical inference,” *Arkiv för Matematik*, vol. 1, no. 3, pp. 195 – 277, 1950.
- 9 U. Grenander, *Abstract Inference*. John Wiley & Sons, 1981.
- 10 E. Parzen, “An Approach to Time Series Analysis,” *The Annals of Mathematical Statistics*, vol. 32, no. 4, pp. 951 – 989, 1961.
- 11 J. O. Ramsay and C. J. Dalzell, “Some tools for functional data analysis,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 53, no. 3, pp. 539–572, 1991.
- 12 P. Besse and J. O. Ramsay, “Principal components analysis of sampled functions,” *Psychometrika*, vol. 51, pp. 285—311, 1986.
- 13 K. Karhunen, “Zur spektraltheorie stochastischer prozesse,” *Ann. Acad. Sci. Fennicae, Ser. A*, vol. 34, pp. 1–7, 1946.
- 14 M. Loève, “Fonctions aléatoires à décomposition orthogonale exponentielle,” *La Revue Scientifique*, vol. 84, pp. 159–162, 1946.
- 15 H.-G. Müller, “Functional modelling and classification of longitudinal data,” *Scandinavian Journal of Statistics*, vol. 32, no. 2, pp. 223–240, 2005.
- 16 G. M. James, “Generalized linear models with functional predictors,” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 64, no. 3, pp. 411–432, 2002.
- 17 F. Rossi and N. Villa, “Support vector machine for functional data classification,” *Neurocomputing*, vol. 69, no. 7, pp. 730–742, 2006. New Issues in Neurocomputing: 13th European Symposium on Artificial Neural Networks.

- 18 J.-T. Zhang and T. Zhu, “A new k-nearest neighbors classifier for functional data,” *Statistics and Its Interface*, vol. 15, no. 2, pp. 247–260, 2022.
- 19 M. Luz López García, R. García-Ródenas, and A. González Gómez, “K-means algorithms for functional data,” *Neurocomputing*, vol. 151, pp. 231–245, 2015.
- 20 M. Yamamoto and Y. Terada, “Functional factorial k-means analysis,” *Computational Statistics & Data Analysis*, vol. 79, pp. 133–148, 2014.
- 21 S. Tokushige, H. Yadohisa, and K. Inada, “Crisp and fuzzy k-means clustering algorithms for multivariate functional data,” *Computational Statistics*, vol. 22, no. 1, pp. 1–16, 2007.
- 22 F. Ieva, A. M. Paganoni, D. Pigoli, and V. Vitelli, “Multivariate functional clustering for the morphological analysis of electrocardiograph curves,” *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 62, no. 3, pp. 401–418, 2013.
- 23 C. Abraham, P. A. Cornillon, E. Matzner-Løber, and N. Molinari, “Unsupervised curve clustering using b-splines,” *Scandinavian Journal of Statistics*, vol. 30, no. 3, pp. 581–595, 2003.
- 24 A. Z. Zambom, Q. Wang, and R. Dias, “A basis approach to surface clustering,” *Statistics, Optimization & Information Computing*, vol. 10, pp. 339–351, Feb. 2022.
- 25 G. P. Sassi, *Estimação de modelos geoestatísticos com dados funcionais usando ondaletas*. PhD thesis, Universidade de São Paulo, 2016.
- 26 R. V. Fonseca, *Modelos de Análise de Dados Funcionais por Ondaletas: Fundamentos e Aplicações*. PhD thesis, Universidade de Campinas, 2021.
- 27 M. A. Saraiva, “Análise não-paramétrica de dados funcionais: Uma aplicação à quimiometria,” Master’s thesis, Universidade de Campinas, 2009.
- 28 J. O. Ramsay and B. W. Silverman, *Applied functional data analysis: methods and case studies*, vol. 77. Springer, 2002.
- 29 M. P. Wand and M. C. Jones, “Kernel smoothing,” 1995.
- 30 B. W. Silverman, “Spline smoothing: The equivalent variable kernel method,” *The Annals of Statistics*, vol. 12, no. 3, pp. 898–916, 1984.
- 31 E. Konzen, Y. Cheng, and J. Q. Shi, “Gaussian process for functional data analysis: The gpfd package for r,” 2021.
- 32 G. Kimeldorf and G. Wahba, “A correspondence between bayesian estimation on stochastic processes and smoothing by splines,” *The Annals of Mathematical Statistics*, vol. 41, no. 2, pp. 495–502, 1970.
- 33 C. A. Pope, *The Confluence of Gaussian Process Emulation and Wavelets*. PhD thesis, University of Leeds, 2019.
- 34 T. Hsing and R. Eubank, “Theoretical foundations of functional data analysis, with an introduction to linear operators,” 2015.

- 35 E. Meijering, “A chronology of interpolation: from ancient astronomy to modern signal and image processing,” *Proceedings of the IEEE*, vol. 90, no. 3, pp. 319–342, 2002.
- 36 R. Dias and C. P. Estevam, “Introdução à análise de dados funcionais,” vol. 19 of *SINAPE*, 2010.
- 37 J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*. Springer-Verlag, 2 ed., 2005.
- 38 J.-L. Wang, J.-M. Chiou, and H.-G. Müller, “Functional data analysis,” *Annual Review of Statistics and Its Application*, vol. 3, no. 1, pp. 257–295, 2016.
- 39 F. Ferraty, ed., *Recent Advances in Functional Data Analysis and Related Topics*, Contributions to Statistics, Physica Heidelberg, 2011.
- 40 R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022.
- 41 H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- 42 R. Giraldo, P. Delicado, and J. Mateu, “Ordinary kriging for function-valued spatial data,” *Environmental and Ecological Statistics*, vol. 18, pp. 411–426, 2011.
- 43 J. O. Ramsay, “Estimating smooth monotone functions,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 60, no. 2, pp. 365–375, 1998.
- 44 C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Mit Pr, 2006.
- 45 K. Weierstrass, *Über die analytische Darstellbarkeit sogenannter willkürlicher Functionen reeller Argumente*, vol. 3 of *Cambridge Library Collection - Mathematics*, p. 1–38. Cambridge University Press, 2013.
- 46 M. H. Stone and P. Ross, *The Generalized Weierstrass Approximation Theorem*, p. 35–44. Mathematical Association of America, 2007.
- 47 W. Rudin, *Principles of Mathematical Analysis*, vol. 3. McGraw-Hill New York, 1976.
- 48 J. W. Glover, “Derivation of the united states mortality table by osculatory interpolation,” *Publications of the American Statistical Association*, vol. 12, no. 90, pp. 85–109, 1910.
- 49 H. A. Schwarz, “Démonstration élémentaire d’une propriété fondamentale des fonctions interpolaires,” *Torino Atti*, vol. 17, pp. 740–743, 1882.
- 50 C. de Boor, “Divided differences,” *Surveys in Approximation Theory*, pp. 46–69, 2005.
- 51 T. Hockey, ed., *Biographical Encyclopedia of Astronomers*. Springer New York, 1 ed., 2007.

- 52 C. D. T. Runge, “Über empirische funktionen und die interpolation zwischen äquidistanten ordinaten,” *Zeitschrift für Mathematik und Physik*, vol. 46, pp. 224–243, 1901.
- 53 E. T. Whittaker, “On a new method of graduation,” *Proceedings of the Edinburgh Mathematical Society*, vol. 41, p. 63–75, 1922.
- 54 I. J. Schoenberg, “Spline functions and the problem of graduation,” *Proceedings of the National Academy of Sciences*, vol. 52, no. 4, pp. 947–950, 1964.
- 55 I. J. Schoenberg, “Contributions to the problem of approximation of equidistant data by analytic functions: Part a.—on the problem of smoothing or graduation. a first class of analytic approximation formulae,” *Quarterly of Applied Mathematics*, vol. 4, no. 1, pp. 45–99, 1946.
- 56 I. J. Schoenberg, “Contributions to the problem of approximation of equidistant data by analytic functions: Part b—on the problem of osculatory interpolation. a second class of analytic approximation formulae,” *Quarterly of Applied Mathematics*, vol. 4, no. 2, pp. 112–141, 1946.
- 57 Noe, R., “When splines were physical objects.”
<https://www.core77.com/posts/55368/When-Splines-Were-Physical-Objects>, 2016. Acessado em: 2022-07-13.
- 58 G. Wahba, *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics, 1990.
- 59 R. J. D. Quintela, “Influência local em modelos parcialmente lineares aditivos generalizados,” Master’s thesis, Universidade Federal de Pernambuco, 2020.
- 60 C. de Boor, “On the convergence of odd-degree spline interpolation,” *Journal of Approximation Theory*, vol. 1, no. 4, pp. 452–463, 1968.
- 61 T. Hastie and R. Tibshirani, *Generalized additive models*. Chapman & Hall/CRC, 1990.
- 62 C. D. Boor, *A Practical Guide to Splines*. New York, Springer-Verlag, 1978.
- 63 H. B. Curry and I. J. Schoenberg, *On Pólya Frequency Functions IV: The Fundamental Spline Functions and their Limits*, pp. 347–383. Boston, MA: Birkhäuser Boston, 1988.
- 64 C. d. Silva Ferreira, G. A. Paula, and G. C. Lana, “Estimation and diagnostic for partially linear models with first-order autoregressive skew-normal errors,” *Computational Statistics*, vol. 37, no. 1, pp. 445–468, 2022.
- 65 A. F. L. Galindo, “Modelos semi-paramétricos para dados independentes e longitudinais com resposta limitada,” Master’s thesis, Universidade de Campinas, 2021.
- 66 S. Wood, *Generalized additive models: an introduction with R*. CRC Press, 2006.
- 67 P. H. C. Eilers and B. D. Marx, “Flexible smoothing with B-splines and penalties,” *Statistical Science*, vol. 11, no. 2, pp. 89 – 121, 1996.

- 68 P. H. Sousa, C. P. Estevam, and R. Dias, “Adaptive bayesian selection of basis functions for functional data via stochastic penalization.” 2022.
- 69 R. Anselmo, Cezar A.F. and Dias and N. L. Garcia, “Adaptive basis selection for functional data analysis via stochastic penalization,” *Computational & Applied Mathematics*, vol. 24, no. 2, pp. 209 – 229, 2005.
- 70 A. E. Hoerl and R. W. Kennard, “Ridge regression: Applications to nonorthogonal problems,” *Technometrics*, vol. 12, no. 1, pp. 69–82, 1970.
- 71 A. E. Hoerl and R. W. Kennard, “Ridge regression: Biased estimation for nonorthogonal problems,” *Technometrics*, vol. 42, no. 1, pp. 80–86, 2000.
- 72 M. Stone, “Cross-validatory choice and assessment of statistical predictions,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 36, no. 2, pp. 111–133, 1974.
- 73 P. Craven and G. Wahba, “Smoothing noisy data with spline functions,” *Numerische Mathematik*, vol. 31, no. 4, pp. 377–403, 1978.
- 74 W. E. Boyce and R. C. DiPrima, *Equações Diferenciais Elementares e Problemas de Valores de Contorno*. LTC, 9 ed., 2010.
- 75 R. A. Rigby and D. M. Stasinopoulos, “Generalized additive models for location, scale and shape,(with discussion),” *Applied Statistics*, vol. 54, pp. 507–554, 2005.
- 76 T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer, feb 2009.
- 77 S. Marsland, *Machine learning: an algorithmic perspective*. CRC Press, 2011.
- 78 J. MacQueen, “Some methods for classification and analysis of multivariate observations,” 1967.
- 79 J. H. W. Jr., “Hierarchical grouping to optimize an objective function,” *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244, 1963.
- 80 E. Hellinger, “Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen,” *Journal für die reine und angewandte Mathematik*, vol. 1909, no. 136, pp. 210–271, 1909.
- 81 C. da Silva Ferreira, “Estimação de máxima verossimilhança penalizada para funções de regressão com erros perpendiculares,” Master’s thesis, Universidade de Campinas, 2003.
- 82 S. Kullback and R. A. Leibler, “On Information and Sufficiency,” *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79 – 86, 1951.
- 83 J. C. Dunn, “Well-separated clusters and optimal fuzzy partitions,” 1974.
- 84 L. Kaufman and P. J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*, vol. 344. John Wiley & Sons, 2009.
- 85 H.-G. Müller and F. Yao, “Functional additive models,” *Journal of the American Statistical Association*, vol. 103, no. 484, pp. 1534–1544, 2008.

- 86 F. Santosa and W. W. Symes, “Linear inversion of band-limited reflection seismograms,” *SIAM Journal on Scientific and Statistical Computing*, vol. 7, no. 4, pp. 1307–1330, 1986.
- 87 R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- 88 Isabela Penna, “Seleção de modelos de regressão linear em bases de alta dimensão,” 2021.
- 89 J. Friedman, T. Hastie, and R. Tibshirani, “Regularization paths for generalized linear models via coordinate descent,” *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.
- 90 N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, “Regularization paths for cox’s proportional hazards model via coordinate descent,” *Journal of Statistical Software*, vol. 39, no. 5, pp. 1–13, 2011.
- 91 B.-H. Mevik and R. Wehrens, “The pls package: Principal component and partial least squares regression in r,” *Journal of Statistical Software*, vol. 18, no. 2, p. 1–23, 2007.
- 92 Ana Carolina Limão and Taymã Carneiro, “Pará vê número de mortes por coronavírus quase triplicar em uma semana.” <https://g1.globo.com/pa/para/noticia/2020/05/01/para-ve-numero-de-mortes-por-coronavirus-subir-em-uma-semana.ghtml>, 2020.
- 93 G1, “Crise do oxigênio no amazonas: entenda quanto falta e as ações para repor o insumo.” <https://g1.globo.com/am/amazonas/noticia/2021/01/15/crise-do-oxigenio-no-amazonas-entenda-o-quanto-falta-e-as-acoes-para-repor-o-insumo.ghtml>, 2021.
- 94 Carlos Madeiro, “Dia mais letal da pandemia no país em 2021 teve mais mortes que dezembro.” <https://noticias.uol.com.br/saude/ultimas-noticias/redacao/2022/01/17/dia-mais-letal-da-pandemia-no-pais-teve-mais-mortes-que-dezembro-de-2021.html>, 2022.
- 95 João Vitor Tavares, “Brasil atinge marca de 50% da população totalmente vacinada contra covid.” <https://www.correiobraziliense.com.br/brasil/2021/10/4956713-brasil-atinge-marca-de-50-da-populacao-totalmente-vacinada-contra-covid.html>, 2021.
- 96 Douglas Porto and Victória Cocolo and Manuela Niclewicz, “Índice aponta as melhores capitais no combate à covid-19; veja ranking.” <https://www.cnnbrasil.com.br/saude/indice-aponta-as-melhores-capitais-no-combate-a-covid-19-veja-ranking/>, 2021.
- 97 J. O. Ramsay and X. Li, “Curve registration,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 60, no. 2, pp. 351–363, 1998.
- 98 B. Gompertz, “On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies.,” Jan. 1815.

- 99 C. P. Winsor, "The gompertz curve as a growth curve," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 18, no. 1, pp. 1–8, 1932.
- 100 F. J. RICHARDS, "A Flexible Growth Function for Empirical Use," *Journal of Experimental Botany*, vol. 10, pp. 290–301, 06 1959.