Universidade Federal de Pernambuco
Centro de Ciências Exatas e da Natureza
Programa de Pós-Graduação em Estatística

RODOLPHO JÓRDAN DOMINGOS QUINTELA

INFLUÊNCIA LOCAL EM MODELOS PARCIALMENTE LINEARES ADITIVOS GENERALIZADOS

Recife

2020

RODOLPHO JÓRDAN DOMINGOS QUINTELA

INFLUÊNCIA LOCAL EM MODELOS PARCIALMENTE LINEARES ADITIVOS GENERALIZADOS

Dissertação apresentada ao Programa de Pós-Graduação em Estatística do Centro de Ciências Exatas e da Natureza da Universidade Federal de Pernambuco, como requisito parcial à obtenção do título de mestre em Estatística.

Área de Concentração: Estatística Matemática

Orientador: Dr. Roberto Ferreira Manghi
Coorientador: Dr. Francisco José de Azevedo Cysneiros

Recife

2020

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

Q7i Quintela, Rodolpho Jórdan Domingos
Influência local em modelos parcialmente lineares aditivos generalizados / Rodolpho Jórdan Domingos Quintela. – 2020.
125 f.: il., fig., tab.

Orientador: Roberto Ferreira Manghi.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2020.
Inclui referências, apêndice e anexo.

1. Estatística matemática. 2. Diagnóstico. I. Manghi, Roberto Ferreira (orientador). II. Título.

519.5 CDD (23. ed.) UFPE- CCEN 2020 - 128

RODOLPHO JÓRDAN DOMINGOS QUINTELA

INFLUÊNCIA LOCAL EM MODELOS PARCIALMENTE LINEARES ADITIVOS GENERALIZADOS

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Aprovada em: 28 DE FEVEREIRO DE 2020

BANCA EXAMINADORA

Prof.(º) Roberto Ferreira Manghi
UFPE

Prof.(a) Audrey Helen Mariz de Aquino Cysneiros
UFPE

Prof.(º) Gilberto Alvarenga Paula
USP

À minha família, por me apoiar em todo momento e por ser um porto seguro. Mãe, seu amor e cuidado me deram forças pra continuar. Pai, és meu maior exemplo.

AGRADECIMENTOS

Acredito que a vida é fundamentada em escolhas. É claro que algumas delas não cabem a nós, ou seja, não temos controle sobre tais decisões. Notadamente, nos deparamos com questões do tipo: O que é preciso fazer para ter uma vida bem sucedida? Aristóteles dirá que é preciso fazer desabrochar as potências que são as tuas. E até chegar a este estado, inevitavelmente, nos esbarramos à fatores que não estão sob o nosso controle. Nesse momento, mediante esses choques singulares com o mundo, acabamos tomando decisões ao sabor das circunstâncias da vida.

Assim, na tentativa de encontrar a minha alegria, isto é, o desabrochar das minhas potências, entendi que não existe uma fórmula para vida. Para mim, existe uma quantidade finita de alternativas que nos são apresentadas. Obviamente optamos por aquelas que julgamos serem as melhores baseado em tudo aquilo que achamos certo. Como todo e qualquer ser humano, às vezes erramos em ter escolhido por um caminho ao invés de outro, acertamos em alguns momentos e em outras situações não resolvemos arriscar. Omitir-se é sem dúvidas a pior de todas as escolhas.

Portanto, não é fácil decidir-se sobre alguma coisa. Quase sempre estamos diante de decisões conflitantes as quais não sabemos exatamente o que fazer. Tais escolhas são complexas e ter a responsabilidade sobre elas é angustiante. Mesmo assim, diante disso tudo, temos que fazê-las. Porque também são elas que vão proporcionar instantes de vida alegres os quais não gostaríamos que acabassem tão logo, fazendo da nossa vida sempre alguma coisa digníssima de ser buscada e fantástica de ser vivida. Dessa forma, cada escolha, cada momento, cada pessoa foram importantes e, definitivamente, contribuíram para a minha formação enquanto pessoa.

Logo, em minha trajetória como aluno da pós-graduação, contei com a colaboração de inúmeras pessoas. Portanto, gostaria muito de agradecer pelo menos minimamente cada um deles como eles merecem.

- Ao professor Luis Gonzaga Pinheiro Felix. Meu primeiro amigo do mestrado. Hoje, mais que antes, tenho a plena convicção que de um jeito ou de outro, seríamos grandes amigos.
- Ranah Duarte, Larissa Lima e Adenice Ferreira, minhas conselheiras. Agradeço pelos momentos de conversas sobre relacionamentos, sobre amizade, sobre a vida. Pelas incontáveis vezes que me ajudaram nas disciplinas tornando assim, minha vida nesse mestrado menos penosa. Aprendi muito com vocês.
- Cristine Oliveira, de todos os colegas de pós, era talvez a que eu menos imaginava que nos

tornaríamos amigos. Nunca me enganei tanto e nunca fiquei tão feliz por isso. Você é uma pessoa maravilhosa.

- Calinhos, um grande amigo a quem devo muito estima.
- Fernando Luiz Maia Gomes, sempre de bem com a vida, muito obrigado pela força de sempre.
- José Jairo, o alívio cômico dessa série. Meu camarada, você é um cara que com essa alegria, bom humor e humildade é o tipo de pessoa que todo mundo quer por perto. Mas o que seria do herói sem seu fiel companheiro? João Eudes Miquéias é outro que tem alegria no coração. Continue assim que vocês vão longe.
- César Diogo foi de longe o cara mais legal que conheci na vida. Seu senso crítico e de justiça me mostrou que não é preciso ser algo que não somos só pra agradar terceiros. Meu amigo, você faz isso de uma maneira espetacularmente educada, concisa e coerente. É pra poucos!
- Nayara Luíza é a pessoa com a qual eu mais me identifiquei. Costumo dizer que ela sou eu na versão feminina, só que bem melhor.
- Anabeth Radünz e Eduardo Ensslin. Torço muito por vocês, pois os tenho como inspiração. Obrigado por tudo.
- Cris Guedes e Lucas Araújo, meus conterrâneos, obrigado por, além de tudo, trazerem um pouco da terrinha pra perto de mim. Graças a vocês, foi mais fácil lidar com a saudade de casa.
- Vinícios Scher, não tenho palavras pra descrever tanto respeito e admiração que tenho pela sua pessoa. Obrigado por todas as vezes que você quebrou meu galho. Sendo sempre muito compressível e disposto a ajudar.
- Lucas de Miranda, aqui deixo também registrado meus mais sinceros agradecimentos. Anny Kerollayne, Thalytta Evilly, Lucas David, Bruna, Tatiane Fontora, Saul De Azevêdo Souza (Shineray), Alê, Pedro Almeida, Charles, Luana. Antigos e novos amigos, serei eternamente grato à vocês. Muito Obrigado!
- Gostaria também de agradecer ao Prof. Dr. Roberto Ferreira Manghi e ao Prof. Dr. Francisco José de Azevêdo Cysneiros por me orientarem nesta dissertação.
- Um agradecimento muito especial a Profa. Dra. Audrey Helen Cysneiros. Não tenha dúvidas de que a senhora será uma ótima lembrança em minha vida. Muito obrigado por tudo.
- A todos o professores do Departamento de Estatística da UFPE e, de modo geral, a todos

os funcionários que fazem parte deste departamento. Muito Obrigado!

- Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro a este projeto.
- Não poderia deixar de agradecer a mãe do meu filho, Amélia Coelho Maciel, que me ajudou de diferentes formas para que eu pudesse terminar essa dissertação. Inclusive, nas correções, dando ideias, sugestões dentre outras coisas. A ela também devo alguns dos meus dias mais felizes, em especial, por ter me proporcionado a maior alegria de todas: ser pai. Por isso, serei eternamente grato.
- Dedico também este trabalho ao meu avô, João Oliveira (*in memorian*).
- E finalmente, gostaria de agradecer, de forma mais que especial, à minha família. À minha mãe, Maria Regina, que sempre acreditou em mim ainda que eu mesmo não acreditasse. Ao meu pai, Francisco Wilson, e minhas irmãs, Maria Jordayane e Jordânia Horrana que me deram todo apoio às minhas decisões mesmo nos momentos mais difíceis e em escolhas que possivelmente pudessem desagrada-los. Espero um dia poder retribuir tudo que vocês fizeram por mim com todo o amor que lhes é inerente.

RESUMO

Esta dissertação tem como objetivo propor resíduos e técnicas para a análise de diagnóstico nos Modelos Parcialmente Lineares Aditivos Generalizados (MPLAGs), tais como: alavancagem generalizada, análise de resíduos, dois quais propomos utilizar os resíduos de Pearson e resíduos aleatorizados, bem como medidas para análise de influência local sob os seguintes esquemas de perturbação: perturbação de de casos, perturbação na variável resposta e perturbação em uma das variáveis explicativas. Para isto, derivamos tais medidas fundamentados em uma vasta pesquisa bibliográfica e conceitual sobre tais métodos no contexto dos MPLAGs. Essas técnicas foram utilizadas em exemplos de aplicação a dados reais e os resultados foram discutidos a fim de avaliar o nosso estudo teórico. Para tanto, apresentamos as equações de estimação para os parâmetros do modelo através da função de verossimilhança penalizada, considerando como estrutura não paramétrica o uso de P-*splines*. Assim, definimos tal modelo, buscando apresentar algumas propriedades e vantagens que motivam o uso de P-*splines* no contexto de regressão não paramétrica. Por fim, o método iterativo *backfitting* (Gauss-Seidel) é utilizado para a obtenção das estimativas.

Palavras-chave: Diagnóstico. Influência Local. MPLAGs. P-splines.

ABSTRACT

This dissertation aims at the proportion of residues and techniques for diagnostic analysis in Generalized Additive Partial Linear Models (GAPLMs), such as: generalized leverage, analysis of residues, two of which we propose to use Pearson's residues and randomized residues, as well as measures for analysis of local influence under the following perturbation schemes: perturbation of cases, perturbation in response variable and pertubation in one of the variables explanatory. For this, we derive such measures based on a vast bibliographic and conceptual research on such methods in the context of GAPLMs. These techniques were used in real data application examples and the results were discussed in order to evaluate our theoretical study. For that, the estimation equations for the model parameters are provided through the penalized likelihood function, considering the use of P-*splines* as a non-parametric structure. Thus, we define such a model, searching to present some properties and advantages that motivate the use of P-splines in the context of non-parametric regression. Finally, the iterative backfitting method (Gauss-Seidel) is used to obtain the estimates.

Keywords: Diagnostics. Local Influence. GAPLMs. P-*splines*.

LISTA DE FIGURAS

Figura 1 – Gráfico de dispersão da aceleração contra o tempo. 22

Figura 2 – Contorno suave obtido com auxílio de um *splines*. 23

Figura 3 – Representação de um ajuste por meio de uma *spline* linear com nós em $\xi_1 = 15$ e $\xi_2 = 30$. 26

Figura 4 – Aproximação de curvas por meio de *splines*. 27

Figura 5 – Representação de um ajuste por meio de *spline* linear e cúbica ambas com nós em $\xi_1 = 1.5$ e $\xi_2 = 2.5$. 28

Figura 6 – Ilustração de um B-*spline* de grau 1 isolado com três nós posicionados em "x". 29

Figura 7 – Ilustração de um B-*splines* de grau 2 isolado com cinco nós posicionados em "x". 30

Figura 8 – Velocidade do impactador em função do tempo de contato entre o projétil e o alvo para uma energia de impacto constante. 54

Figura 9 – Gráficos da distribuição de frequência, frequência acumulada e boxplot, respectivamente, para a variável tempo. 54

Figura 10 – Gráficos da distribuição de frequência, frequência acumulada e boxplot, respectivamente, para a variável aceleração. 55

Figura 11 – O valor do critério VCG, curva aproximada e intervalos de confiança pontuais de 95% para a aceleração contra o tempo. 56

Figura 12 – Gráficos da distribuição de frequência, frequência acumulada e boxplot respectivos da a variável idade para o exemplo das eleições de 2006 ao Senado brasileiro. 70

Figura 13 – Gráficos com as porcentagens para cada faixa de idade e relação da idade com a quantidade de votos recebidas pelos candidatos para o exemplo das eleições de 2006 ao Senado brasileiro. 71

Figura 14 – Estes gráficos mostram a relação da idade com os gastos e diagrama de dispersão: votos segundo os gastos na campanha dos candidatos para o exemplo das eleições de 2006 ao Senado brasileiro. 71

Figura 15 – Gráficos com as porcentagens dos sexo na amostra e relação da sexo dos candidatos com as quantidades de votos por eles recebidas para o exemplo das eleições de 2006 ao Senado brasileiro. 72

Figura 16 – Gráficos com as porcentagens dos sexo na amostra e relação da sexo dos candidatos com as quantidades de votos por eles recebidas para o exemplo das eleições de 2006 ao Senado brasileiro. 73
Figura 17 – Gráfico com a relação do sexo com o estado civil dos candidatos para o exemplo das eleições de 2006 ao Senado brasileiro. 74
Figura 18 – Resultado das eleições de 2006 ao Senado brasileiro segundo o estado civil. 74
Figura 19 – Gráficos com as porcentagens dos níveis de instrução na amostra e relação das quantidades de votos por eles recebidas para o exemplo das eleições de 2006 ao Senado brasileiro 75
Figura 20 – Gráfico da curva ajustada sob distribuição normal para o exemplo das eleições de 2006 ao Senado brasileiro. 77
Figura 21 – Gráficos de resíduos referente ao modelo ajustado sob distribuição normal para o exemplo das eleições de 2006 ao Senado brasileiro. 78
Figura 22 – Gráficos de alavancagem e influência local sob os esquemas de perturbação de caso, perturbação na variável resposta e perturbação em umas das variáveis explicativas, respectivamente, referentes ao ajuste do modelo sob distribuição normal para o exemplo das eleições de 2006 ao Senado brasileiro. 80
Figura 23 – Gráficos da distribuição de frequência, frequência acumulada e boxplot respectivos da a variável preço do aluguel para o exemplo do aluguel em Munique. 83
Figura 24 – Gráficos com as proporções de cada variável categórica na amostra para o exemplo do aluguel em Munique. 84
Figura 25 – Boxplots do valor do aluguel segundo as variáveis categóricas para o exemplo do aluguel em Munique. 85
Figura 26 – Diagramas de dispersão do valor do aluguel contra as variáveis numéricas para o exemplo do aluguel em Munique. 86
Figura 27 – Gráfico das correlações entre as variáveis quantitativas para o exemplo do aluguel em Munique. 86
Figura 28 – Gráficos das curvas ajustadas sob distribuição gama para o exemplo do aluguel em Munique. 88

Figura 29 – Gráfico das curvas ajustadas para o segundo modelo sob distribuição gama para o exemplo do aluguel em Munique. 89
Figura 30 – Gráficos de resíduos referente ao modelo ajustado sob distribuição gama para o exemplo do aluguel em Munique. 90
Figura 31 – Gráficos de alavancagem e influência local sob os esquemas de perturbação de caso, perturbação na variável resposta e perturbação em umas das variáveis explicativas, respectivamente, referentes ao ajuste do modelo sob distribuição gama para o exemplo do aluguel em Munique. 92
Figura 32 – Boxplots das variáveis presentes no baco de dados considerando o exemplo do ar poluído em Chicago. 94
Figura 33 – Diagrama de dispersão do logaritmo do número de mortes contra as demais variáveis numéricas considerando o exemplo do ar poluído em Chicago. 96
Figura 34 – Gráficos das curvas ajustadas sob distribuição Poisson para o exemplo do ar poluído em Chicago. 98
Figura 35 – Gráficos de resíduos referentes ao modelo ajustado sob distribuição Poisson para o exemplo do ar poluído em Chicago. 100
Figura 36 – Gráfico de influência local sob os esquemas de perturbação de caso e perturbação na variável resposta respectivamente para o exemplo do ar poluído em Chicago. 101
Figura 37 – Gráfico que mostra como está distribuída a reprodução das aves Cotovia-de-crista em Portugal. 102
Figura 38 – Curva aproximada e intervalos de confiança pontuais de 95% referente ao modelo ajustado sob distribuição binomial para o exemplo da reprodução das aves. 104
Figura 39 – Gráfico de resíduos referente o modelo ajustado sob distribuição binomial para o exemplo da reprodução das aves. 105
Figura 40 – Gráficos de de alavancagem e influência local sob os esquemas de perturbação de caso e perturbação em uma das variáveis explicativas, respectivamente, referente ao ajuste do modelo sob distribuição binomial para o exemplo da reprodução das aves. 106

LISTA DE TABELAS

Tabela 0 – Resumo das estimativas referentes ao modelo ajustado para o exemplo do impacto do capacete. 56
Tabela 1 – Resultado das eleições de 2006 ao Senado brasileiro. 72
Tabela 2 – Resultado das eleições de 2006 ao Senado brasileiro segundo o estado civil. 73
Tabela 3 – Resultado das eleições de 2006 ao Senado brasileiro segundo o nível de instrução. 75
Tabela 4 – Resumo das estimativas referentes ao modelo ajustado sob distribuição normal para o exemplo das eleições de 2006 ao Senado brasileiro. . . . 76
Tabela 5 – Resumo das estimativas referentes ao modelo final ajustado sob distribuição normal para o exemplo das eleições de 2006 ao Senado brasileiro. . . . 77
Tabela 6 – Estimativas de máxima verossimilhança referente ao ajuste do modelo sob distribuição normal e mudança relativa em porcentagem entre parênteses para os parâmetros β_0, β_1 e β_2 após a retirada dos pontos para o exemplo das eleições de 2006 ao senado brasileiro. 81
Tabela 7 – Resumo das estimativas de máxima verosimilhança referentes ao modelo ajustado sob distribuição gama para o exemplo do aluguel em Munique. 87
Tabela 8 – Resumo das estimativas de máxima verossimilhança referentes ao segundo modelo ajustado sob distribuição gama para o exemplo do aluguel em Munique. 89
Tabela 9 – Estimativas de máxima verossimilhança referente ao ajuste do modelo sob distribuição gama e mudança relativa em porcentagem entre parênteses para os parâmetros β_0, β_1, β_2, β_3, β_4, β_5 e β_6 após a retirada dos pontos para o exemplo do aluguel em Munique. 93
Tabela 10 – Resumo das estimativas de máxima verossimilhança referentes ao modelo ajustado sob distribuição Poisson para o exemplo do ar poluído em Chicago. 97
Tabela 11 – Resumo das estimativas de máxima verossimilhança referentes ao modelo final ajustado sob distribuição Poisson para o exemplo do ar poluído em Chicago. 98

Tabela 12 – Estimativas de máxima verossimilhança referente ao ajuste do modelo sob distribuição poisson e mudança relativa em porcentagem entre parênteses para os parâmetros β_0 e β_1 após a retirada dos pontos para o exemplo do ar poluído em Chicago. 101

Tabela 13 – Resumo das estimativas de máxima verosimilhança referentes ao modelo ajustado sob distribuição binomial para o exemplo dda reprodução das aves 103

Tabela 14 – Estimativas de máxima verossimilhança e mudança relativa em porcentagem entre parênteses para os parâmetros β_0 e β_1 do modelo binomial após a retirada dos pontos para o exemplo da reprodução das aves em Portugal. 106

LISTA DE CÓDIGOS-FONTE

Código-fonte 1 – Função para ajuste de uma splines linear 111
Código-fonte 2 – Função para ajuste de uma splines cúbica 111
Código-fonte 3 – Função para ajuste do modelo 114
Código-fonte 4 – Obtenção de alpha 115
Código-fonte 5 – Plot - Pontuação VCG 117
Código-fonte 6 – Estimativas do modelo ótimo 117
Código-fonte 7 – Função para análise de influência local 118

SUMÁRIO

1 INTRODUÇÃO 18
1.1 ORGANIZAÇÃO DA DISSERTAÇÃO 20
1.2 SUPORTE COMPUTACIONAL 21
1.3 CONCEITOS PRELIMINARES 21
1.3.1 Splines 22
1.3.2 *Splines* lineares 25
1.3.3 *Splines* cúbicas 26
1.3.4 B-*splines* 29
1.3.5 P-*splines* 30
2 MODELOS PARCIALMENTE ADITIVOS GENERALIZADOS 33
2.1 INTRODUÇÃO 33
2.2 ESPECIFICAÇÃO DO MODELO 34
2.2.1 Função de verossimilhança penalizada 35
2.2.2 Restrição de identificabilidade 40
2.3 ESTIMAÇÃO DOS PARÂMETROS 46
2.4 INFERÊNCIA ESTATÍSTICA 48
2.4.1 Graus de liberdade efetivos 48
2.4.2 Seleção do parâmetro de suavização 49
2.4.3 Escolha do número de nós 50
2.4.4 Intervalos de confiança 51
2.4.5 Teste de hipóteses 52
2.4.6 Impacto do capacete da motocicleta 53
3 TÉCNICAS DE DIAGNÓSTICO NOS MPLAGS 58
3.1 ANÁLISE DE RESÍDUOS 59
3.1.1 Resíduos de Pearson 59
3.1.2 Resíduos quantílicos aleatorizados 60
3.2 MEDIDAS DE ALAVANCAGEM 60
3.3 INFLUÊNCIA LOCAL 62
3.3.1 Perturbação de casos 65
3.3.2 Perturbação da variável resposta 65
3.3.3 Perturbação em uma das variáveis explicativas 66

3.3.4 **Perturbação em vetores particionados** 67
4 **APLICAÇÕES** 69
4.1 ELEIÇÕES DE 2006 AO SENADO BRASILEIRO 69
4.1.1 **Análise exploratória dos dados** 69
4.1.2 **Ajuste do modelo** 75
4.1.3 **Diagnóstico** 78
4.2 ALUGUEL EM MUNIQUE 82
4.2.1 **Análise exploratória dos dados** 82
4.2.2 **Ajuste do modelo** 87
4.2.3 **Diagnóstico** 90
4.3 AR POLUÍDO EM CHICAGO 93
4.3.1 **Análise exploratória dos dados** 94
4.3.2 **Ajuste do modelo** 96
4.3.3 **Diagnóstico** 99
4.4 REPRODUÇÃO DE AVES EM PORTUGAL 102
4.4.1 **Ajuste do modelo** 103
4.4.2 **Diagnóstico** 104
5 **CONCLUSÕES E TRABALHOS FUTUROS** 107
REFERÊNCIAS 108
APÊNDICE A – CÓDIGOS COMPUTACIONAIS 111
ANEXO A – DERIVAÇÃO DOS ESQUEMAS DE PERTURBAÇÃO . 122

1 INTRODUÇÃO

Em geral, para descrever algum fenômeno da natureza e/ou social, considera-se algum aspecto de interesse da população em estudo como a média, variância ou algum outro parâmetro. Neste caso, busca-se alguma lei matemática capaz de modelar a influência sistemática de variáveis que se supõem ter algum efeito sobre tais parâmetros. Desta forma, assumindo como parâmetro de interesse a média das variáveis resposta, então para relacionar o efeito que determinadas variáveis explicativas tem sobre este parâmetro, pode-se considerar metodologias, tais como: regressão paramétrica, utilizada quando se sabe, mesmo que aproximadamente, algum argumento teórico que indique qual deve ser a verdadeira relação entre a média da variável resposta e as variáveis explicativas; a regressão não paramétrica, utilizada quando, por outro lado, esta relação se dá mediante a funções que se adéquem aos dados e não parâmetros; ou, como uma combinação destas metodologias, a regressão semi-paramétrica.

Neste contexto, os modelos parcialmente lineares generalizados (MPLAGs) representam uma classe de modelos de regressão semi-paramétricas que tentam estimar relações complexas entre a variável resposta e as covariáveis. A singularidade desses modelos é a sua flexibilidade, visto que não há necessidade de estabelecer antecipadamente o relacionamento entre as variáveis. Nesta classe de modelos, os dados são os que determinam a forma do relacionamento. Isso tornou os MPLAGs uma ferramenta fundamental em qualquer área de pesquisa, pois possibilitou análises estatísticas mais precisas, proporcionando vislumbrar caminhos com melhores resultados.

Dentre os principais trabalhos relacionados aos modelos aditivos generalizados, convém destacar as referências de Green e Silverman (1994). Estes autores assumem como preditor, uma soma de funções suaves de covariáveis quando a distribuição da variável resposta pertence a família exponencial. Green e Yandell (1985) apresentam o conceito de estimativas de máxima verossimilhança penalizada. Aqui, os autores também discutem critérios de seleção dos parâmetros de suavização para os modelos de regressão semi-paramétricos utilizando *splines*. Eilers e Marx (1996) estudam a estimação não paramétrica por meio de B-*splines* e P-*splines*. Além destes, Fahrmeir *et al.* (2013) e Hodges (2016) também trazem um desenvolvimento completo de modelos aditivos, principalmente no que tange a modelos mistos. Ademais, recomenda-se fortemente o livro de Wood (2017). Nele, pode-se encontrar, além de uma descrição detalhada da teoria que envolve esses modelos, uma excelente parte com aplicações no R.

Porém, conforme Cordeiro e Demétrio (2008), mesmo tendo passado por uma criteriosa escolha do modelo, pode ocorrer um resultado insatisfatório para o ajuste a um conjunto de observação. Assim sendo, uma etapa seguinte muito importante da análise de um ajuste consiste em seu diagnóstico. Segundo Paula (2013), esta etapa é responsável por verificar possíveis desvios em relação as suposições feitas para o modelo, em especial, no que diz respeito ao componente aleatório e, notadamente, para a parte sistemática, bem como a presença de observações destoantes com alguma interferência desproporcional ou inferencial nos resultados do ajuste. Em outras palavras, busca-se analisar a precisão dos resultados obtidos, tentando identificar possíveis desvios de suposições decorrentes de erros de especificação do modelo − como a escolha errônea para distribuição da variável resposta, da função de variância, função de ligação, ausência ou não do parâmetro de dispersão −, ou discrepâncias isoladas − oriundas de pontos presentes nas extremidades da amplitude de validade da variável explanatória, erros de digitação, algum fator não controlado (mas relevante) que influenciou a sua obtenção ou até mesmo multicolinearidade entre variáveis (correlação alta entre variáveis) − que podem exercer influência acentuada nas estimativas dos parâmetros do modelo (CORDEIRO; DEMÉTRIO, 2008, p.135).

Neste contexto, se faz necessário uso de técnicas para o diagnóstico do modelo, por meio da análise de resíduo, alavancagem e, sobretudo, por meio da análise de influência local, proposta por Cook (1978). Desta forma, Pulgar (2009), por exemplo, discute medidas de diagnósticos para modelos mistos aditivos semiparamétricos com erros de contornos elípticos. Emami (2016) estuda a influência local de pequenas perturbações nas estimativas obtidas pelo procedimento dos mínimos quadrados penalizados de Ridge. Manghi (2016) desenvolve técnicas de diagnóstico para os Modelos Parcialmente Lineares Aditivos Generalizados para dados correlacionados.

Dentre os trabalhos mais recentes no estudo dos MPLAGs, destaca-se o trabalho de Holanda (2018), com o título: modelos lineares parciais aditivos generalizados com suavização por meio de P-*splines*. Neste trabalho é apresentado um estudo sobre os modelos parcialmente lineares aditivos generalizados utilizando P-*splines* para descrever a relação da variável resposta com as variáveis explicativas contínuas, sendo que os P-*splines* são uma combinação entre o método de suavização B-*splines* e uma penalização de diferenças. Contudo, não foi desenvolvida a análise de diagnóstico para este modelo, bem como não há nada na literatura a respeito desse tema, sendo portanto uma lacuna a ser preenchida. Sendo assim, é oportuno desenvolver um estudo a fim de preencher tal lacuna existente. Neste caso, esta pesquisa tem a pretensão de

debruçar nas técnicas de diagnóstico por meio do estudo de resíduos, alavancagem e influência local baseada na proposta de Cook (1978), para esta classe de modelos.

Portanto, justifica-se esta pesquisa tendo em vista a necessidade de ampliar e aprofundar os estudos sobre procedimentos para o diagnóstico dos modelos parcialmente lineares aditivos generalizados. De forma mais precisa, será dedicado consideráveis esforços no que tange a análise de influência local, bem como apresentar essa técnica, mostrando seu uso em um contexto de pesquisa teórico e prático. Dessa forma, os objetivos específicos são, além do desenvolvimento teórico para análise de diagnóstico por meio da técnica de influência local, fornecer um suporte computacional no R. Além disso, buscar exemplos práticos para ilustrar a teoria desenvolvida.

1.1 ORGANIZAÇÃO DA DISSERTAÇÃO

Esta dissertação possui 4 capítulos e estão organizados da seguinte forma. No Capítulo 1 é apresentada uma introdução que visa descrever o objetivo deste trabalho. Nela, também buscamos relembrar alguns resultados importantes na literatura que utilizaram a metodologia de análise de diagnóstico por meio da técnica de influência local desenvolvida por Cook (1978) em diferentes contextos de aplicação.

No Capítulo 2 os MPLAGs são definidos e o processo de estimação dos parâmetros é apresentado. Desta forma, enfatizamos as ideias por trás da especificação desses modelos, bem como apresentamos discussões sobre processos de estimação por meio do método de máxima verossimilhança penalizada, bem como discorremos sobre o método iterativo *backfitting*.

No Capítulo 3, visando à compreensão prévia das técnicas que envolvem a análise de diagnóstico para um modelo de regressão, apresentamos e desenvolvemos uma explanação das principais estruturas para realizar a análise de diagnóstico dos MPLAGs. Para isso, sugerimos os resíduos de Pearson e os resíduos quantílicos aleatorizados com o intuito de identificar possíveis *outliers*, bem como discutimos o uso da matriz de projeção em modelos semi-paramétricos para investigar pontos de alavanca, além de desenvolver medidas de diagnóstico sob o enfoque da técnica de influência local.

Por fim, no Capítulo 4 são apresentadas algumas aplicações a banco de dados reais, seguido de comentários e conclusões do estudo.

1.2 SUPORTE COMPUTACIONAL

Destaca-se que as análises dos dados apresentados neste trabalho foram realizadas por meio do ambiente de programação R, em sua versão 3.6.1 para a plataforma **Linux**, por meio de um Ambiente de Desenvolvimento Integrado (Integrated Development Environment- IDE) chamado RStudio. A linguagem e seus pacotes podem ser obtidos no endereço **http://www.r-project.org** de forma gratuita. Para mais detalhes, consultar Ihaka e Gentleman (1996) e Team (2019). Para redigir a presente dissertação, foi utilizado o sistema tipográfico LaTeX através da IDE TeXStudio, obtida gratuitamente pelo endereço **https://www.texstudio.org**.

1.3 CONCEITOS PRELIMINARES

Em muito casos, a relação funcional entre a média da variável resposta e as variáveis explicativas se dá por meio de formas mais complexas que uma simples estrutura linear. Considere por exemplo o modelo de Michaelis-Menten para cinética enzimática. Tal modelo tem dois parâmetros e uma variável independente, relacionados por f, de tal forma que:

$$f(x, \boldsymbol{\beta}) = \frac{\beta_1 x}{\beta_2 + x}.$$

Esta função é não linear, pois não pode ser escrita como uma combinação linear dos β's. Para estimar os parâmetros dessa função poderia-se utilizar o método dos mínimos quadrados.

A abordagem acima é conhecida na literatura como regressão paramétrica não linear (ver BEALE 1960, PAYANDEH 1983 e BATES; WATTS 1988, por exemplo). A escolha dessas funções é motivada, em sua grande parte, pelo conhecimento prévio do tipo de relação entre as variáveis. Contudo, nem sempre existe esse conhecimento. Neste caso, a alternativa é deixar que a amostra, por si mesma, apresente possíveis relações entre as variáveis. A essa metodologia dá-se o nome de regressão não paramétrica.

Neste contexto, para estimar $f(\cdot)$ de forma não paramétrica, é necessário o uso de alguns métodos especiais. Dentre estes, por exemplo, estão os métodos baseados em séries ortogonais, método dos k vizinhos mais próximos (em inglês, *k-nearest neighbours, KNN*), regressão polinomial local, *splines*, *smoothing splines*, apenas para citar alguns. Estes métodos, assim como na regressão paramétrica não linear, são mais flexíveis em comparação com a regressão linear. Por captar melhor a relação funcional de $f(\cdot)$, oferecem um forte poder preditivo, bem como não há a necessidade de fazer hipóteses sobre esta relação, diferentemente da regressão paramétrica não linear. Por outro lado, são métodos cuja maior flexibilidade vem

atrelada a uma interpretabilidade mais complicada. Além disso, é necessário uma quantidade razoavelmente grande de observações a fim de obter uma aproximação para $f(\cdot)$ suficientemente precisa. Mais detalhes sobre estes procedimentos podem ser encontrados em Izbicki e Santos (2019).

Contudo, este trabalho focará em apenas um desses métodos. Para estimar a relação funcional de $f(\cdot)$, será utilizado o método de suavização por meio de P-*splines*. Este procedimento é uma combinação entre o método de B-*splines* e uma penalização de diferenças aplicadas diretamente aos coeficientes adjacentes dos B-*splines*. Assim, é necessário inicialmente apresentar uma ideia introdutória sobre *splines*, bem como os conceitos do método de B-*splines* para posteriormente definirmos os P-*splines*.

1.3.1 Splines

Considere um problema em que o interesse é investigar o comportamento de capacetes durante o impacto na cabeça em acidentes simulados. Este exemplo será retomado com mais detalhes na **Seção** 2.4.6. Por hora, é suficiente saber que neste experimento foi considerado a aceleração da cabeça em unidades de g, para determinar força sentida pelo cérebro imediatamente após o impacto. A **Figura** 1 mostra o diagrama de dispersão da aceleração contra o tempo dado em milissegundo (ms) para este experimento.

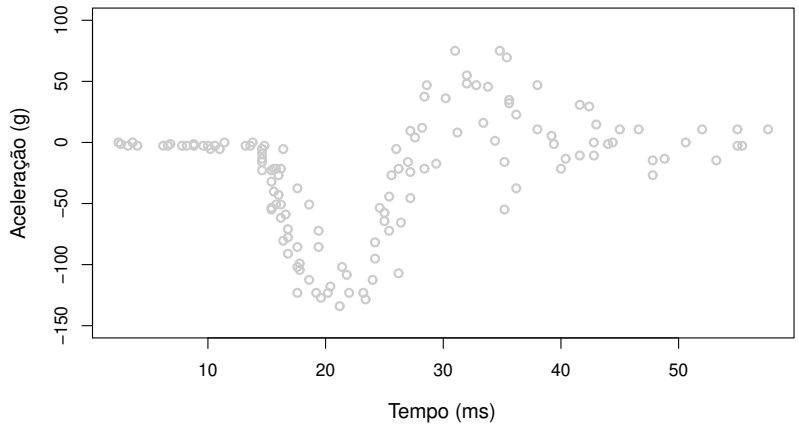


Figura 1 – Gráfico de dispersão da aceleração contra o tempo.

Como é possível observar, o comportamento da aceleração em função do tempo claramente não possui uma forma linear, tão pouco existe algum conhecimento *a priori* de como se dá essa relação. Neste caso, como brevemente exposto anteriormente, costuma-se utilizar

algum procedimento capaz de captar essa relação de modo que não seja feito qualquer suposição a seu respeito.

Contudo, talvez a forma mais usual seria tentar aproximar esse comportamento por meio de algum polinômio. Estes são bastante flexíveis e são bem simples de serem estimados. No entanto, como será visto mais adiante, a tentativa de ajustar curvas por meio de um único polinômio, principalmente os que possuem graus elevados, pode apresentar algumas dificuldades. Alternativamente pode-se usar *splines*, que são formados por pedaços de polinômios unidos em pontos chamados de nós e que são mais estáveis, sendo capazes de controlar melhor aspectos locais da relação funcional.

A palavra *spline* veio emprestada do desenho técnico e designa uma régua flexível que, antes dos computadores, era usada para desenhar curvas que passam por pontos pré-determinados com o intuito de auxiliar na etapa de delineamento de objetos, principalmente em projetos de engenharia, tais como cascos de navio, peças de avião, etc. É uma ferramenta que tenta, por meio de desenhos, comunicar visualmente como algo funciona ou é construído. Na **Figura** 2 é apresentado um contorno obtido por meio de *splines*.

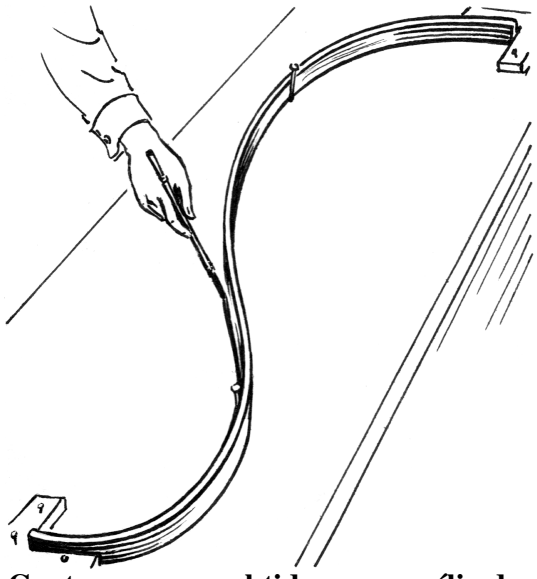


Figura 2 – Contorno suave obtido com auxílio de um *splines*.

Fonte: https://fr.wikipedia.org/wiki/Spline

Na matemática, o termo foi usado pela primeira vez em um artigo de 1946, pelo matemático romeno Isaac Jacob Schoenberg. A ideia é construir uma curva que passe por alguns pontos da amostra, chamados de nós, com o objetivo de aproximar tal curva do que se acredita ser aquela que represente adequadamente o comportamento desses dados.

De modo bastante sucinto, define-se *splines* como sedo curvas formadas por pedaços

de polinômios. Neste caso, ao invés de modelar um conjunto de observações por um único polinômio, escolhe-se pontos distintos no intervalo das observações (nós) e é definido um polinômio para cada intervalo, de forma a modelar curvas mais complexas por polinômios mais simples.

Existem algumas vantagens em se usar *splines*, dentre elas pode-se destacar a sua maior flexibilidade para o ajuste dos modelos se comparado ao modelo de regressão linear ou polinomial, a capacidade de modelar comportamentos atípicos dos dados, bem como a facilidade do ajuste quando determinados a quantidade e a posição dos nós.

Basicamente, o método consiste em substituir o vetor de entrada $\boldsymbol{x}$, por variáveis adicionais, que serão combinações de $\boldsymbol{x}$ e então utilizar a aproximação linear nesse novo espaço. Assim, seja $b_\ell(\boldsymbol{x})$ a ℓ-ésima transformação de $\boldsymbol{x}$, $\ell = 1,2,\cdots q$, então $f(\boldsymbol{x})$ assume a seguinte representação:

$$f(\boldsymbol{x}) = \sum_{\ell=1}^{q} b_\ell(\boldsymbol{x})\gamma_\ell, \tag{1.1}$$

para alguns valores dos parâmetros desconhecidos γ_ℓ. Neste caso, obtém-se uma expansão linear de bases em $\boldsymbol{x}$. Portanto, após a definição de $b_\ell(\boldsymbol{x})$, o modelo passa a ser linear nessas novas variáveis gerados pelo vetor $\boldsymbol{x}$ e consequentemente, os métodos de estimação dos parâmetros γ_ℓ podem ser aplicados nesse novo espaço. Note que, ao contrário do que se imagina, técnicas não paramétricas evolvem a estimação de muitos parâmetros. Por essa razão, Eilers e Marx (1996) preferem *"overparametric" techniques* (técnicas de muitos parâmetros), embora tais parâmetros não tenham interpretação científica.

De acordo com Wood (2017, p. 120), costuma-se escolher como representação de $b_\ell(\boldsymbol{x})$ funções capazes de aumentar a flexibilidade de $f(\boldsymbol{x})$ e, consequentemente, do modelo. Um exemplo disso são os polinômios. Os polinômios são muito úteis quando o interesse reside nas propriedades de f na vizinhança de um ponto específico. Contudo, devido a sua natureza, tendem a distorcer a realidade em regiões remotas. Considerando os interesses estatísticos, o ajuste de curvas por meio de um único polinômio geralmente acarreta em problemas de multicolinearidade, isto é, associação de alta correlação entre as variáveis explicativas. Além disso, considerando polinômios com ordens elevadas, o risco de haver *overfitting* aumenta bastante. O que queremos é uma curva que se adéque a diferentes amostras vindas de uma mesma população e que seja capaz de captar adequadamente a variabilidade dos dados. Uma ideia para lidar com essas dificuldades é ajustar polinômios de menor grau e por partes.

Assim, supondo que $\boldsymbol{x}$ seja unidimensional, então um polinômio por partes é obtido

separando o domínio de $\boldsymbol{x}$ em intervalos conectados e definir f como um polinômio por intervalo. Neste caso, pode-se definir um polinômio por partes da seguinte maneira:

$$b_\ell(\boldsymbol{x}) = h_1(\boldsymbol{x})\mathbb{I}(\boldsymbol{x} < \xi_1) + h_i(\boldsymbol{x})\mathbb{I}(\xi_i < \boldsymbol{x} < \xi_{i+1}) + h_k(\boldsymbol{x})\mathbb{I}(\xi_k < \boldsymbol{x}),\ i+1,2,\ldots,k \quad (1.2)$$

em que h_i são polinômios conhecidos, $\mathbb{I}(\cdot)$ uma função indicadora e k é a quantidade de nós. É importante destacar que não está sendo feito qualquer afirmação sobre a continuidade destas funções. Para garantir a continuidade, deve-se estabelecer restrições, de modo que $f(\xi_i^-) = f(\xi_i^+)$. Ou seja, a união dos polinômios se dá de forma suave, sem mudanças abruptas, e assim os valores da função nos nós serão os mesmo.

1.3.2 *Splines* lineares

Uma *spline* linear, também chamada de função linear segmentada, é definida por um conjunto de polinômios de grau um, unidos continuamente entre si através dos nós. Neste caso, para cada intervalo representado pelos nós, existe um segmento de reta.

Considerando $\boldsymbol{x}$ um vetor ordenado de tamanho n. Assim, definindo $n-1$ nós, $\xi_1,\xi_2,\ldots,\xi_{n-1}$, tem-se que o modelo mais simples de *spline* possui a seguinte forma:

$$f(x) = \beta_0 + \beta_1 x + \beta_2(x-\xi_1)_+ + \cdots + \beta_n(x-\xi_{n-1})_+, \quad (1.3)$$

em que

$$(x-\xi_j)_+ = \begin{cases} (x-\xi_j), & (x-\xi_j) \geq 0 \\ 0, & (x-\xi_j) < 0, \end{cases} \quad (1.4)$$

para $j = 1,2,\ldots,n-1$. Pode-se notar que, conforme o valor no eixo x aumenta, novos termos são adicionados à função. Note que, se $x < \xi_1$, a equação irá configurar esse espaço. Isto é, entre o menor valor de x e ξ_1 o valor da função será $f(x) = \beta_0 + \beta_1 x$. No entanto, quando x estiver entre os dois nós seguintes, a equação será $f(x) = \beta_0 + \beta_1 x + \beta_2(x-\xi_1)$ e assim sucessivamente. Além disso, também é possível notar que nos intervalos entre os nós a *spline* linear é contínua. Porém, sua primeira derivada é descontínua nos pontos de união representado pelos nós.

Na **Figura** 3 é apresentado um ajuste por meio de uma *spline* linear. Com base no gráfico de dispersão, foram escolhidos como ponto de quebra os nós $\xi_1 = 15$ e $\xi_2 = 30$. Neste

modelo, tem-se a seguinte formulação:

$$f(x) = \begin{cases} \hat{\beta}_0 + \hat{\beta}_1 x, & \text{se } 0 < x \leq 15 \\ \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 (x - 15), & \text{se } 15 < x \leq 30 \\ \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 (x - 15) + \beta_3 (x - 30), & \text{se } 30 < x, \end{cases} \tag{1.5}$$

em que $\hat{\beta}_0 = 2.02, \hat{\beta}_1 = 1.85, \hat{\beta}_2 = -3.81, \hat{\beta}_3 = 3.94$.

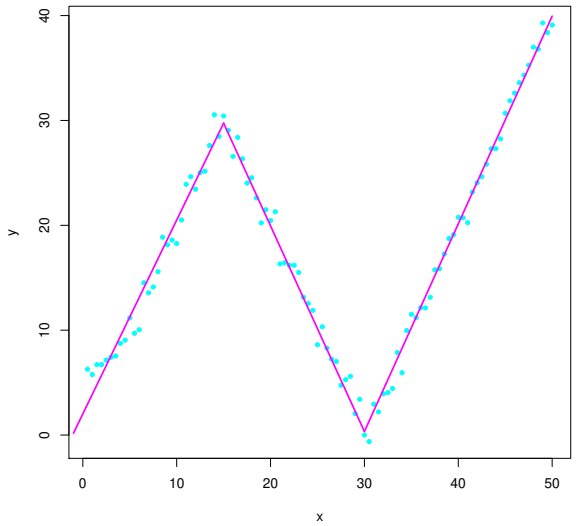


Figura 3 – Representação de um ajuste por meio de uma *spline* linear com nós em $\xi_1 = 15$ e $\xi_2 = 30$.

Note que uma regressão linear simples não seria capaz de estimar de forma adequada a relação funcional entre as variáveis. Isto é, observando o ajuste, a regressão por partes se sai muito melhor, dado sua proximidade aos pontos observados. O banco de dados para este exemplo de aplicação considerando *splines* lineares, bem como o código para a construção do gráfico são apresentados em **Código-fonte 1**.

1.3.3 *Splines* cúbicas

As *splines* cúbicas são as de menor ordem nas quais a descontinuidade nos nós são menos evidentes. Portanto, não há muitas razões capazes de motivar o uso de *splines* com graus maiores, a não ser que por algum motivo específico seja necessário mais derivadas suavizadas.

Essas *splines* podem ser dividida em dois tipos: as *splines* cúbicas restritas e *splines* cúbicas irrestritas. No caso em que a *spline* cúbica é do tipo restrita, então suas primeiras derivadas nos nós serão zero e suas caudas, isto é, a parte do polinômio antes do primeiro nó e após o último nó, serão modeladas por meio de funções lineares. A *spline* cúbica restrita é

também chamadas de *spline* natural. Caso contrario, será uma *spline* cúbica irrestrita. Na **Figura** 4, é dado uma ilustração de uma *spline* cúbica natural.

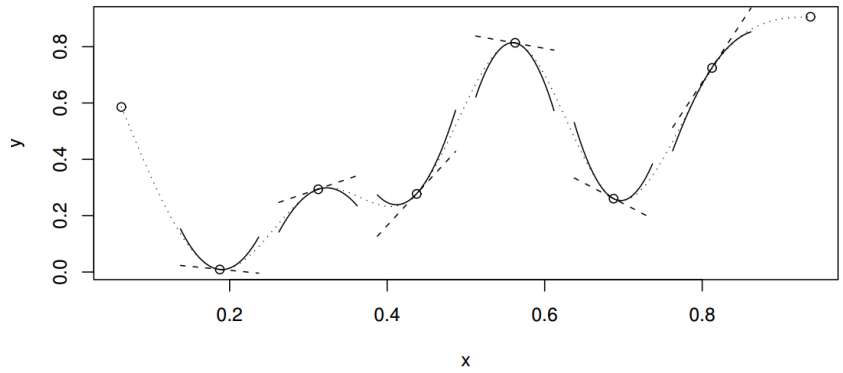


Figura 4 – Aproximação de curvas por meio de *splines*.

Fonte: Wood (2017, p. 122).

Assim, seja f uma função no intervalo $[a,b]$. Supondo que esse intervalo seja dividido em intervalos menores representado pelos nós, então define-se uma *spline* cúbica como sendo uma curva construída a partir de polinômios cúbicos, os quais são unidos de forma contínua nos nós. Ou seja, nos nós a primeira e a segunda derivadas da função serão contínuas. Além disso, cada seção do polinômio cúbico tem diferentes coeficientes, mas nos nós os valores das funções serão os mesmo.

Resumidamente, seja $[\xi_i,\xi_{i+1}]$ o intervalo delimitado pelos nós e S uma *spline* cúbica utilizada para aproximar a curva f, então S possui as seguintes propriedades:

i) para $i = 1,2\ldots,N$, seja $S(x) = s_i(x)$, onde cada s_i é um polinômio cúbico;
ii) $s_i(\xi_{i-1}) = f(\xi_{i-1})$, para $i = 1,2\ldots,N$;
iii) $s_i(\xi_i) = f(\xi_i)$, para $i = 1,2\ldots,N$;
iv) $s_i'(\xi_i) = s_{i+1}'(\xi_i)$, para $i = 1,2\ldots,N-1$;
v) $s_i''(\xi_i) = s_{i+1}''(\xi_i)$, para $i = 1,2\ldots,N$.

Dessa forma, analogamente ao que foi feito para construir os *splines* lineares, pode-se construir as *splines* cúbicas. De modo geral, a função *spline* cúbica com k nós pode ser escrita como:

$$f(x) = \beta_0 + \beta_1 x + \beta_2 (x-\xi_1)_+^3 + \cdots + \beta_{k+1}(x-\xi_k)_+^3, \tag{1.6}$$

em que

$$(x-\xi_k)_+^3 = \begin{cases} (x-\xi_k)^3, & x-\xi_k \geq 0 \\ 0, & x-\xi_k < 0. \end{cases} \tag{1.7}$$

A principal vantagem do uso de *splines* cúbicas em relação as *splines* lineares é o fato desta ser contínua até pelo menos a segunda derivada, inclusive nos nós. Isso possibilita captar pequenas ondulações no comportamento dos dados. A desvantagem é que os valores da função original não coincidem com os valores da *spline* **nas primeiras derivadas**, mesmo nos nós.

Como exemplo de aplicação, considere que os dados sejam gerados a partir da função:

$$y_i = 3 * \cos(3 + 2x_i) + 5 + \epsilon_i, \text{ para } i = 1,2,\ldots 30,$$

em que $\epsilon_i \sim \mathcal{N}(\mu = 0, \sigma^2 = 1)$ e $x_i \sim \mathcal{U}(\text{a} = 0.5, \text{b} = 3.3)$. A **Figura** 5 mostra o resultado do ajuste. Os dados gerados por meio da função anterior para este exemplo de aplicação considerando *splines* cúbica, bem como o código para a construção do gráfico é apresentado no **Código-fonte 2**.

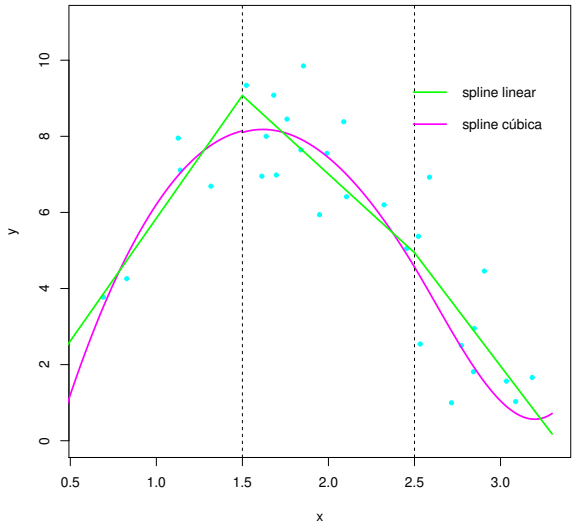


Figura 5 – Representação de um ajuste por meio de *spline* linear e cúbica ambas com nós em $\xi_1 = 1.5$ e $\xi_2 = 2.5$.

Contudo, Pulgar (2009) mostra que as *splines* cúbicas não restritas costumam apresentar comportamento inadequado nas caudas, sendo estas, muito suscetíveis a mudanças. Portanto, por convenção, as funções que vem antes e depois do primeiro nó são forçadas a terem formato linear.

1.3.4 B-*splines*

Dentre todas as funções que são contínuas em $[a,b]$ e interpola $\{x_i, y_i\}$, a *spline* cúbico é o interpolante mais suave possível através de qualquer conjunto de dados. Este resultado poder ser consultado em Wood (2017, p. 142-143). Existem várias maneiras de se representar uma *spline* cúbico. Uma delas é a representação B-*spline* proposta por De Boor (1978). Assim, neste caso, tomando $b_\ell(\cdot) = B_\ell^m(\cdot)$, define-se (1.1) como:

$$f(t_i) = \sum_{\ell=1}^{q} B_\ell^{(m)}(t_i)\gamma_\ell, \tag{1.8}$$

sendo os B-*splines*, $B_\ell^{(m)}(t_i)$, definidos como:

$$B_\ell^{(m)}(t_i) = \frac{t_i - \xi_\ell}{\xi_{\ell+m+1} - \xi_\ell} B_\ell^{(m-1)}(t_i) + \frac{\xi_{\ell+m+2} - t_i}{\xi_{\ell+m+2} - t_{\ell+1}} B_\ell^{(m-1)}(t_i) \text{ e}$$
$$B_\ell^{(0)}(t_i) = \begin{cases} 1 & \text{se } \xi_\ell \leq t_i < \xi_{\ell+1} \\ 0 & c.c. \end{cases} \tag{1.9}$$

em que t_i é uma variável que contribui de forma não linear no modelo, para $i = 1,2,\ldots,n$, $\xi_1, \xi_2, \ldots \xi_z$ são z nós equidistantes, sendo que $m = 1,2,\ldots$ é o grau dos B-*splines* e $z = q + m + 1$. De acordo com Holanda (2018), a base de De Boor é capaz de controlar aspectos locais da relação funcional através do número e da posição dos nós, bem como a quantidade de parâmetros a serem estimados.

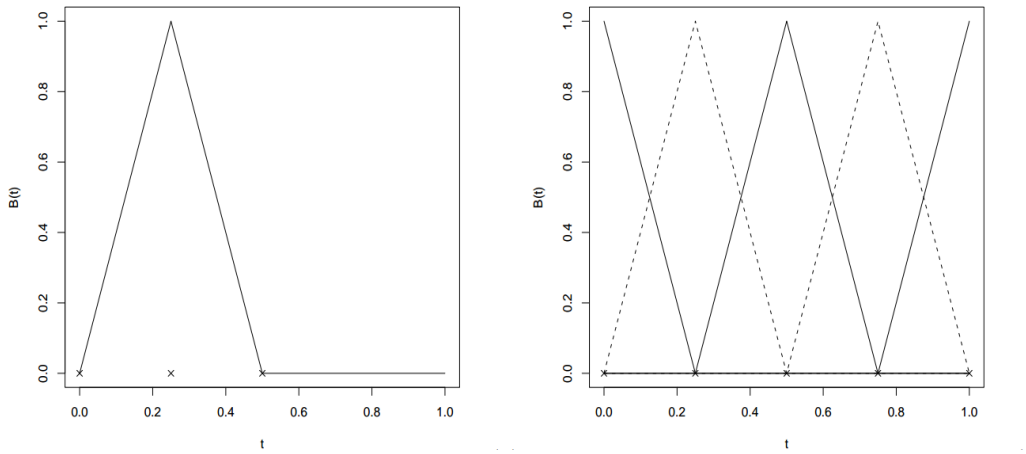


Figura 6 – Ilustração de um B-*spline* de grau 1 isolado com três nós posicionados em "x".

Fonte: Souza (2008, p. 35).

A **Figura** 6 mostra dois gráficos com B-*splines* de grau 1, ambos baseados em três nós. Na parte direita da supra citada figura, encontramos todos os B-*splines* de grau 1 no intervalo

$[0,1]$ com nós em $\{0;0.25;0.5;0.75;1\}$. Evidentemente, um número maior de B-*splines* podem ser construídos. Para isto, basta que sejam inseridos mais nós. Por sua vez, à direita da **Figura** (7) tem-se um B-*spline* de grau 2. Neste caso, a curva é formada por três pedaços de polinômios quadráticos, unidos por dois nós internos. À esquerda da **Figura** 7 o leitor pode observar todos os possíveis B-*splines* de grau 2 levando em conta os mesmo nós.

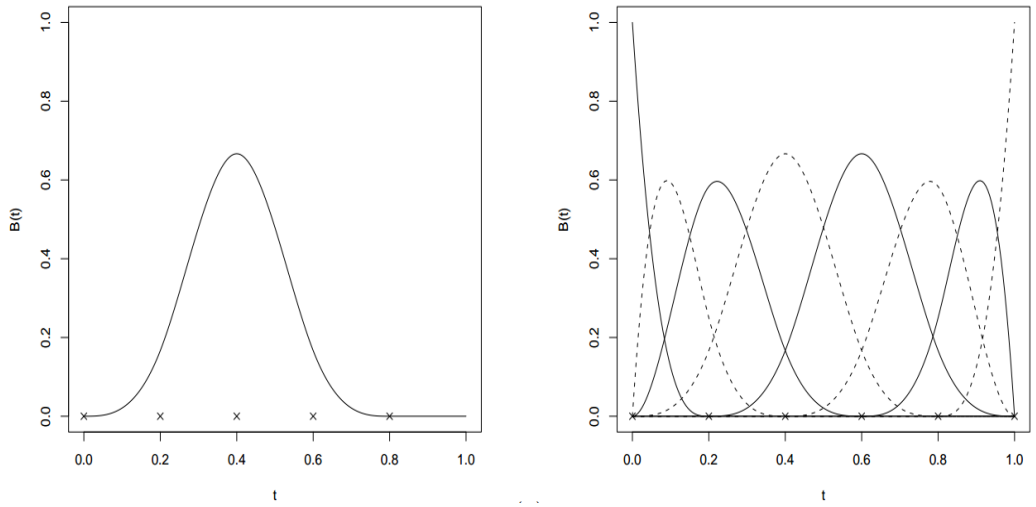


Figura 7 – Ilustração de um B-*splines* de grau 2 isolado com cinco nós posicionados em "x".

Fonte: Souza (2008, p. 35).

Novamente é interessante observar como os polinômios se encaixam, tanto na ordem como também em suas primeiras e segundas derivadas (mas não as terceires derivadas), o que evidencia o caráter contínuo, bem como o aspecto suave apresentado pela curva. Além disso, segundo Eilers e Marx (1996, p. 90), um B-*spline* de ordem m consiste em m pedaços de polinômios, cada um de grau $m-1$. Estes pedaços de polinômios se juntam em $m-1$ nós internos. Nos pontos de união, as derivadas de ordem até $m-2$ são contínuas. Além disso, o B-*spline* é positivo no domínio abrangido por $m+1$ nós, mas fora disto ele é zero.

1.3.5 P-*splines*

De acordo com Eilers e Marx (1996), os B-*splines* possuem suavização influenciada unicamente pelo número de nós que será utilizado. Muitos nós podem levar a um *overfitting*, isto é, um super ajuste dos dados. Assim, a curva ajustada se adapta muito bem a uma amostra proveniente de alguma população, mas será pouco representativa da variabilidade dos dados. Por outro lado, uma quantidade muito pequena de nós pode acarretar problemas de *underfitting*, que

ocorre quando a curva é muito suave e não se adapta bem se quer aos dados que se dispõem. Eilers e Marx (1996) afirmam que a há intensas discussões entorno da escolha dos nós (ver FRIEDMAN; SILVERMAN, 1989 e KOOPERBERG; STONE, 1992). Contudo, as pesquisas ainda não levaram a nenhum esquema que fosse considerado pelos autores como uma forma atraente.

Para contornar essa dificuldade, O'Sullivan (1986), conforme citado por Eilers e Marx (1996), propõe usar um número relativamente grande de nós. Para controlar o ajuste da curva, é utilizada uma penalidade semelhante à pioneira para *splines* de suavização proposta por Reinsch (1967). À vista disso, ao invés de definir $f(t_i) = y_i$, é melhor tratar $f(t_i)$ como n parâmetros livres da *spline* cúbica e estimá-la a fim de minimizar

$$\sum_{i=1}^{n} \{y_i - f(t_i)\}^2 + \alpha \int f''(t)^2 dt, \tag{1.10}$$

em que α é um parâmetro de suavização usado para controlar o peso relativo a ser dado aos objetivos conflitantes de combinar os dados e produzir uma função f suave. Nota-se que, portanto, a função objetivo (1.10) trata-se, na verdade, de uma penalização na função objetivo do método dos mínimos quadrados.

A ideia de utilizar o método de estimação por mínimos quadrados penalizado pela integral do quadrado da segunda derivada de f tornou-se o padrão em grande parte da literatura (ver, por exemplo, EUBANK, 1988, WAHBA, 1990 e GREEN; SILVERMAN, 1994). Contudo, Eilers e Marx (1996) sugerem usar como penalização diferenças simples aplicadas diretamente aos parâmetros adjacentes dos B-*splines*. Os autores argumentam que essa combinação, a qual chamaram de P-*splines*, apresenta propriedades interessantes, tais como: não apresentam efeitos de contorno (ou fronteira); são uma extensão direta de modelos de regressão linear (generalizados), conservam momentos (média, variância) dos dados e têm curvas polinomiais ajustadas como limites, além de serem computacionalmente mais baratos. Assim, considerando (1.10), o interesse passa a ser minimizar:

$$\sum_{i=1}^{n} \{y_i - f(t_i)\}^2 + \alpha \boldsymbol{\gamma}^\top \boldsymbol{P}^d \boldsymbol{\gamma}, \tag{1.11}$$

em que $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^\top$ são os parâmetros do B-*spline* e $\boldsymbol{P}^d$ é uma matriz obtida a partir de um operador de diferenças entre os parâmetros adjacentes em $\boldsymbol{\gamma}$. Por meio de $\boldsymbol{P}^d$ é possível controlar a dependência entre intervalos adjacentes (no domínio de t) na estimação de $f(t)$. Para

mais detalhes sobre o uso dos P-*splines*, deve-se consultar Eilers e Marx (1996) e Eilers, Marx e Maria (2015).

Nos próximos capítulos será apresentada uma classe de modelos semi-paramétricos, definidos por um preditor que é modelado tanto por parâmetros como por funções suaves, tais como (1.1). Com isso, serão discutidas algumas especificações relacionadas as condições de estimação dos parâmetros, bem como a obtenção dos graus de liberdade. Além disso, será discutido procedimentos para realização de testes de hipóteses e intervalos de confiança.

2 MODELOS PARCIALMENTE ADITIVOS GENERALIZADOS

2.1 INTRODUÇÃO

Os modelos lineares constituem a forma mais simples de se estabelecer uma relação funcional entre a variável resposta e as variáveis explicativas. Esse modelo é especialmente aplicado quando a variável resposta é contínua e apresenta distribuição próxima a normalidade. Em razão dessa simplicidade, Paula (2013) lembra que durante muitos anos essa foi a principal abordagem para descrever a maioria dos fenômenos. Porém, como exemplificado por Fahrmeir *et al.* (2013), em várias situações práticas as suposições para garantir um bom uso do modelo de regressão linear clássico não são justificadas para estabelecer uma análise adequada, como por exemplo, nos casos em que a variável resposta era discreta ou, de modo mais geral, pertencente a um sub conjunto dos números reais, ou quando é necessário considerar a heteroscedasticidade.

Assim, na tentativa de contornar essas dificuldades, muitos pesquisadores costumavam aplicar alguma transformação nos dados com o objetivo de obter a normalidade, linearidade e constância da variância. Porém, essa abordagem quase sempre apresentava algumas limitações, além de tornar difícil a interpretação dos resultados. Neste sentido, surgiram na literatura alternativas mais interessantes que visavam superar essas dificuldades específicas. A regressão logística, por exemplo, foi desenvolvida no século XIX para lidar com problemas em que a variável resposta era dicotômica. Contudo, a proposta mais inovadora é atribuída a Nelder e Wedderburn (1972). Eles mostraram que muitas técnicas estatísticas, tais como regressão logística, regressão para dados de contagem, dentre outras, que antes eram estudadas separadamente, fazem parte de uma única classe de modelos de regressão. Os autores denominaram essa classe de **Modelos Lineares Generalizados** (MLG). Esse conjunto de métodos e procedimentos possibilitou um leque bem mais diversificado para a distribuição da variável resposta y. No entanto, essa abordagem, por si só, ainda não é suficiente para tratar casos em que o efeito das covariáveis é não-linear.

Para tanto, como discutido em seções anteriores, existem algumas metodologias que podem ser levadas em consideração: a primeira seria considerar uma regressão paramétrica não linear. Para isso, seria necessário saber, pelo menos de forma aproximada, qual deve ser a verdadeira relação não linear entre a média da variável resposta e as variáveis explicativas. A segunda seria deixar que os próprios dados indiquem qual deve ser essa relação. Nessa perspectiva, considera-se o uso de funções especiais que aproximem o comportamento não

linear das variáveis sem que, para isso, seja feito qualquer suposição a respeito da verdadeira relação entre a média da variável resposta e as variáveis explicativas. Neste caso, há um maior poder preditivo do modelo. Porém, há também uma perda na interpretabilidade, bem como fazer inferência a respeito dos parâmetros torna-se impossível.

Contudo, existe um interesse na comunidade estatística em entender como a resposta é afetada pela variação dos preditores ou identificar os mais importantes na relação entre a resposta e cada um deles. Foi baseado nessa perspectiva que surgiram os Modelos Parcialmente Lineares Aditivos Generalizados (MPLAG). Esta abordagem é uma forma de encontrar um equilíbrio entre interpretabilidade e flexibilidade. Na sequência, esta classe de modelos será apresentada com mais detalhes.

2.2 ESPECIFICAÇÃO DO MODELO

Os MPLAGs fazem parte de uma família de modelos que considera, além das funções paramétricas adotadas nos modelos lineares generalizados, uma soma de funções suaves de covariáveis a fim de relaxar a suposição de linearidade. Portanto, para definir os MPLAGs, considera-se a distribuição de y_i, para $i = 1,2,\ldots,n$, dada como:

$$f(y_i;\theta_i,\phi) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i,\phi)\right\}, \tag{2.1}$$

de modo que, sob certas condições de regularidade, o valor esperado e a variância de y_i são dados, respectivamente, por $\mathbb{E}(y_i) = \mu_i$ e $\mathrm{Var}(y_i) = \phi V_i$, em que $\mu_i = \partial b(\theta_i)/\partial\theta_i$ e $V_i = \partial^2 b(\theta_i)/\partial\theta_i^2$. O parâmetro natural (ou canônico) θ_i pode ser expresso como $\theta_i = \int V_i^{-1} d\mu_i = q(\mu_i)$. Além disso, considere $g(\cdot)$ uma função de ligação monótona e duas vezes diferenciável, tal que:

$$g(\mu_i) = \eta_i = \sum_{j=1}^{p} x_{ji}\beta_j + \sum_{k=1}^{r} f_k(t_{ik}),\ i = 1,2,\ldots,n, \tag{2.2}$$

sendo $f_k(\cdot)$ uma função (suave) desconhecida da variável t_k que contribui de forma não linear no modelo. Então tem-se que um MPLAG nada mais é do que uma extensão dos MLGs, quando existe apenas a parte paramétrica no preditor.

Supõe-se que $f_k(\cdot)$ pertença a um espaço infinito-dimensional de funções duplamente contínuas e diferenciáveis e duplamente integráveis. Existem vários procedimentos para estimar esta função. Contudo, usar-se-á o método de suavização por meio de B-*splines* com função de penalização dos P-*splines*. Sendo assim, a função $f_k(t_{ik})$ é expressa da seguinte forma:

$$f_k(t_{ik}) = \sum_{\ell=1}^{q_k} B_{k\ell}^{(m_k)}(t_{ik})\gamma_{k\ell},\ i = 1,2,\ldots,n,\ \text{e}\ k = 1,2,\ldots,r, \tag{2.3}$$

no qual q_k é a dimensão da base dos k-ésimo B-*spline* (ver De BOOR, 1978), $z_k = q_k + m_k + 1$ é o número de nós, $B_{k\ell}^{(m_k)}(t_{ik})$ denota ℓ-ésimo componente da base do k-ésimo B-*spline* de ordem $m_k + 1$ no ponto t_{ik} e $\gamma_{\ell k}$ são os coeficientes a serem estimados, para $i = 1,\ldots,n$ e $k = 1,\ldots,r$.

Com relação as estimativas dos parâmetros do modelo, toma-se como principal procedimento para tal finalidade uma adaptação do *Método da Máxima Verossimilhança*. Este método consiste em, baseado nos dados da amostra, encontrar a distribuição mais plausível, dentre todas aquelas definidas pelos possíveis valores de seus parâmetros, capaz de ter gerado a amostra disponível (COLOSIMO; GIOLO, 2006, p. 85). Isto é, as estimativas dos parâmeros do modelo serão obtidos pela maximização da função de verossimilhança (ou do logaritmo da função de verossimilhança). Contudo, é importante observar que a maximização direta, sem que tenha estabelecido qualquer restrição sobre a função não paramétrica $f_k(t_{ik})$, pode gerar um superajuste aos dados. Desta forma, é necessário incorporar uma penalização no logaritmo da função de verossimilhanças para controlar o ajuste.

2.2.1 Função de verossimilhança penalizada

Conforme observado anteriormente, quando se trabalha com uma matriz de base, a princípio não é necessária uma penalização no método da máxima verossimilhança (ou mínimos quadrados). Entretanto, neste caso, a suavidade da curva estimada depende do número de nós e da dimensão da base. Para não ter que se preocupar com esta escolha, deixa-se a base e o número de nós fixados e aplica-se a penalização. Esta penalização é função tanto do parâmetro de suavização como dos nós, sendo ela responsável por evitar o super ou sob ajuste dos dados.

Na literatura, é usual a penalização por meio da integral da segunda derivada da função f. Porém, uma alternativa mais atraente é usar como critério de penalização a função de penalização dos P-*splines* proposta por Eilers e Marx (1996), definida como:

$$\boldsymbol{\gamma}^\top \boldsymbol{P}^d \boldsymbol{\gamma}, \tag{2.4}$$

em que $\boldsymbol{\gamma} = (\gamma_1,\ldots,\gamma_q)^\top$, $\boldsymbol{P}^d = (\boldsymbol{\Delta}^d)^\top(\boldsymbol{\Delta}^d)$, de modo que $\boldsymbol{\Delta}^d$ é uma matriz com ordem $(q-d)\times q$ obtida por meio do operador de diferenças Δ^d, tal que $\Delta^d\gamma_\ell = \Delta\left(\Delta^{d-1}\gamma_\ell\right)$ e $\Delta\gamma_\ell = (\gamma_\ell - \gamma_{\ell-1})$, sendo d a ordem dessas diferenças.

Portanto, seja $d = 1$, por exemplo, então tem-se que $\Delta\gamma_\ell = (\gamma_\ell - \gamma_{\ell-1})$. Assim, podemos obter um vetor coluna $\boldsymbol{\Delta\gamma}$, dado por $\boldsymbol{\Delta\gamma} = \left(\Delta\gamma_2,\ \Delta\gamma_3,\ldots,\Delta\gamma_q\right)^\top$, de onde conseguimos $\boldsymbol{\Delta\gamma} = \left(\gamma_2 - \gamma_1,\ \gamma_3 - \gamma_2,\ldots,\gamma_q - \gamma_{q-1}\right)^\top$. Observe que $\boldsymbol{\Delta\gamma}$ é um vetor de tamanho $q-1$.

Desse modo, tem-se que a função de penalização será definida por:

$$(\boldsymbol{\Delta\gamma})^\top \boldsymbol{\Delta\gamma} = \begin{bmatrix} \gamma_2-\gamma_1 & \gamma_3-\gamma_2 & \cdots & \gamma_q-\gamma_{q-1} \end{bmatrix} \begin{bmatrix} \gamma_2-\gamma_1 \\ \gamma_3-\gamma_2 \\ \vdots \\ \gamma_q-\gamma_{q-1} \end{bmatrix}$$
$$= (\gamma_2-\gamma_1)^2 + (\gamma_3-\gamma_2)^2 + \cdots + (\gamma_q-\gamma_{q-1})^2.$$

Porém, é possível notar ainda que o vetor $\boldsymbol{\Delta\gamma}$ pode ser reescrito como o produto de uma matriz por um vetor de parâmetros, de tal maneira que

$$\boldsymbol{\Delta\gamma} = \begin{bmatrix} \gamma_2-\gamma_1 \\ \gamma_3-\gamma_2 \\ \vdots \\ \gamma_q-\gamma_{q-1} \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots \\ 0 & -1 & 1 & 0 & \cdots \\ 0 & 0 & 0 & -1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_q \end{bmatrix}.$$

Desta forma, a matriz $\boldsymbol{\Delta}$, neste caso, tem dimensão $(q-1)\times q$ e será dada por:

$$\boldsymbol{\Delta} = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots \\ 0 & -1 & 1 & 0 & \cdots \\ 0 & 0 & 0 & -1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

de modo que também vale

$$\boldsymbol{\gamma}^\top \boldsymbol{\Delta}^\top \boldsymbol{\Delta\gamma} = (\gamma_2-\gamma_1)^2 + (\gamma_3-\gamma_2)^2 + \cdots + (\gamma_q-\gamma_{q-1})^2.$$

Por exemplo, tomando $q=4$, então teremos $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)^\top$ e, consequentemente $\boldsymbol{\Delta\gamma} = (\gamma_2-\gamma_1,\ \gamma_3-\gamma_2,\ \gamma_4-\gamma_3)^\top$. Ou seja:

$$\boldsymbol{\gamma}^\top \boldsymbol{\Delta}^\top \boldsymbol{\Delta\gamma} = (\gamma_2-\gamma_1)^2 + (\gamma_3-\gamma_2)^2 + (\gamma_4-\gamma_3)^2.$$

Contudo, também temos que:

$$\boldsymbol{\Delta\gamma} = \begin{bmatrix} \gamma_2-\gamma_1 \\ \gamma_3-\gamma_2 \\ \gamma_4-\gamma_3 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{bmatrix},$$

de onde obtemos

$$\boldsymbol{\Delta} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix},$$

resultando em uma matriz $\mathbf{P} = \boldsymbol{\Delta}^\top \boldsymbol{\Delta}$ dada por:

$$\mathbf{P} = \begin{bmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

$$\mathbf{P} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix},$$

o que também nos leva à função de penalização a seguir

$$\boldsymbol{\gamma}^\top \boldsymbol{\Delta}^\top \boldsymbol{\Delta} \boldsymbol{\gamma} = (\gamma_2 - \gamma_1)^2 + (\gamma_3 - \gamma_2)^2 + (\gamma_4 - \gamma_3)^2,$$

já definida anteriormente.

Adotando a ordem $d = 2$, tem-se que $\boldsymbol{\Delta}^2 \boldsymbol{\gamma} = \left(\Delta^2 \gamma_3,\ \Delta^2 \gamma_4, \ldots, \Delta^2 \gamma_q\right)^\top$. Neste caso, $\Delta^2 \gamma_\ell = \Delta(\Delta \gamma_\ell)$ em que

$$\begin{aligned} \Delta^2 \gamma_\ell &= \Delta(\gamma_\ell - \gamma_{\ell-1}) \\ &= \Delta\gamma_\ell - \Delta\gamma_{\ell-1} \\ &= (\gamma_\ell - \gamma_{\ell-1}) - (\gamma_{\ell-1} - \gamma_{\ell-2}) \\ &= \gamma_\ell - 2\gamma_{\ell-1} + \gamma_{\ell-2}, \end{aligned}$$

de onde conseguimos $\boldsymbol{\Delta}^2 \boldsymbol{\gamma} = \left(\gamma_3 - 2\gamma_2 + \gamma_1,\ \gamma_4 - 2\gamma_3 + \gamma_2, \ldots, \gamma_q - 2\gamma_{q-1} + \gamma_{q-2}\right)^\top$. Assim, $\boldsymbol{\Delta}^2 \boldsymbol{\gamma}$ é um vetor de tamanho $q - 2$. Além disso, de forma análoga ao que foi desenvolvido anteriormente para $d = 1$, é possível reescrever este vetor como

$$\boldsymbol{\Delta}^2 \boldsymbol{\gamma} = \begin{bmatrix} \gamma_3 - 2\gamma_2 + \gamma_1 \\ \gamma_4 - 2\gamma_3 + \gamma_2 \\ \vdots \\ \gamma_q - 2\gamma_{q-1} + \gamma_{q-2} \end{bmatrix} = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots \\ 0 & 1 & -2 & 1 & \cdots \\ 0 & 0 & 1 & -2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_q \end{bmatrix}.$$

Logo, a matriz $\boldsymbol{\Delta}^2$ obtida com operador de diferenças de segunda ordem assume a seguinte forma:

$$\boldsymbol{\Delta}^2 = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots \\ 0 & 1 & -2 & 1 & \cdots \\ 0 & 0 & 1 & -2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

que neste caso $\boldsymbol{\Delta}^2$ é uma matriz de dimensão $(q-2) \times q$. Consequentemente, também leva à função de penalização será dada por:

$$\boldsymbol{\gamma}^\top \left(\boldsymbol{\Delta}^2\right)^\top \left(\boldsymbol{\Delta}^2\right) \boldsymbol{\gamma} = (\gamma_1 - 2\gamma_2 + \gamma_3)^2 + (\gamma_2 - 2\gamma_3 + \gamma_4)^2 + \cdots + (\gamma_{q-2} - 2\gamma_{q-1} + \gamma_q)^2.$$

Por exemplo, tomando novamente $q = 4$, então teremos $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)^\top$ e, portanto $\boldsymbol{\Delta}^2\boldsymbol{\gamma} = (\gamma_3 - 2\gamma_2 + \gamma_1,\ \gamma_4 - 2\gamma_3 + \gamma_2)^\top$. Sendo assim, a função de penalização será dada por:

$$\boldsymbol{\gamma}^\top \boldsymbol{\Delta}^\top \boldsymbol{\Delta} \boldsymbol{\gamma} = (\gamma_2 - \gamma_1)^2 + (\gamma_3 - \gamma_2)^2 + (\gamma_4 - \gamma_3)^2.$$

Contudo, note que:

$$\boldsymbol{\Delta}^2\boldsymbol{\gamma} = \begin{bmatrix} \gamma_3 - 2\gamma_2 + \gamma_1 \\ \gamma_4 - 2\gamma_3 + \gamma_2 \end{bmatrix} = \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{bmatrix},$$

de onde obtemos

$$\boldsymbol{\Delta}^2 = \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{bmatrix},$$

resultando em uma matriz $\mathbf{P}^2 = \left(\boldsymbol{\Delta}^2\right)^\top \left(\boldsymbol{\Delta}^2\right)$ dada por:

$$\mathbf{P}^2 = \begin{bmatrix} 1 & 0 \\ -2 & 1 \\ 1 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -2 & 1 & 0 \\ -2 & 5 & -4 & 1 \\ 1 & -4 & 5 & -2 \\ 0 & 1 & -2 & 1 \end{bmatrix},$$

É claro que, como observado, pode-se calcular a função de penalização diretamente sem o intermédio de $\mathbf{P}^d$. Porém, a matriz $\mathbf{P}^d$ é necessária para o cálculo das estimativas dos parâmetros do modelo, como será desenvolvido a seguir. Além disso, os cálculos das medidas de diagnóstico que serão apresentadas no capítulo 3 também necessitam desta matriz.

Segundo Holanda (2018), a penalização P-*spline* é mais flexível, visto que independe do grau adotado para os B-*splines*. Desta maneira, é possível combinar qualquer ordem da penalidade com qualquer ordem das bases de De Boor. Além disso, ainda segundo a autora, pode-se utilizar outras penalidades a fim de obter algumas propriedades desejáveis, como, por exemplo: garantir a monotonicidade da função, para exigir uma forma convexa ou côncava ou para forçar os sinais das restrições.

Portanto, considere que $\boldsymbol{\xi}^\top = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \boldsymbol{\gamma}_2^\top, \ldots, \boldsymbol{\gamma}_r^\top) \in \Xi \subseteq \mathbb{R}^s$, em que $s = p + \sum_{k=1}^{r} q_k$. Isto posto, sob a suposição de independência, o estimador de $\boldsymbol{\xi}$ será obtido como sendo o valor que maximiza a função de log-verossimilhança penalizada dada a seguir:

$$\mathrm{L_p}(\boldsymbol{\xi}) = \mathrm{L}(\boldsymbol{\xi}) + \frac{1}{2}\sum_{k=1}^{r} \alpha_k \boldsymbol{\gamma}_k^\top \boldsymbol{P}_k^d \boldsymbol{\gamma}_k, \tag{2.5}$$

em que $\mathrm{L}(\boldsymbol{\xi}) = \sum_{i=1}^{n} \left\{\phi^{-1}[y_i\theta_i - b(\theta_i)] + c(y_i;\phi)\right\}$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_r)^\top$ é um vetor de parâmetros de suavização, de forma que $\alpha_k > 0$, $\boldsymbol{\gamma}_k$ são os coeficientes dos B-*splines* da k-ésima função não paramétrica, $\boldsymbol{P}_k^d = (\Delta_k^d)^\top(\Delta_k^d)$ e Δ_k^d é um operador de diferenças, tal que $\Delta_k^d\gamma_{k\ell} = \Delta_k\left(\Delta_k^{d-1}\gamma_{k\ell}\right)$ e $\Delta_k\gamma_{k\ell} = (\gamma_{k\ell} - \gamma_{k(\ell-1)})$, sendo d a ordem da diferença. Assim sendo, deve-se maximizar (2.5), resolvendo o seguinte sistema de equações:

$$\mathrm{U}_p(\boldsymbol{\xi}) = \frac{\partial \mathrm{L_p}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = \mathbf{0}, \tag{2.6}$$

obtendo, assim, as estimativas dos parâmetros e sendo $\mathrm{U}_p(\boldsymbol{\xi})$ o *vetor escore penalizado*.

À vista disto, tomando $\boldsymbol{X}^* = (\boldsymbol{X}^\top, \boldsymbol{B}_1^\top, \ldots, \boldsymbol{B}_r^\top)$, em que $\boldsymbol{X}$ é uma matriz $n \times p$ formada pelos elementos x_{ji}, $\boldsymbol{B}_k$ é uma matriz de dimensão $n \times q_k$ formada pelos componentes $B_k^{(m_k)}(t_{ik})$ da base de De Boor, as equações de estimação de $\boldsymbol{\xi}$ são dadas por:

$$\mathrm{U}_p(\boldsymbol{\xi}) = \boldsymbol{X}^{*\top}\boldsymbol{W}\boldsymbol{D}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) - \boldsymbol{P}(\boldsymbol{\alpha}) = \mathbf{0}, \tag{2.7}$$

sendo $\boldsymbol{W} = \boldsymbol{D}\boldsymbol{V}^{-1}\boldsymbol{D}$, com $\boldsymbol{D} = \mathrm{diag}\left\{\partial\mu_1/\partial\eta_1, \ldots, \partial\mu_n/\partial\eta_n\right\}$, $\boldsymbol{V} = \mathrm{diag}\left\{V_1, \ldots, V_n\right\}$ e $\boldsymbol{P}(\boldsymbol{\alpha}) = \mathrm{blocodiag}\left\{\mathbf{0}_{pp}, \boldsymbol{P}^d(\boldsymbol{\alpha})\right\}$, no qual $\mathbf{0}_{pp}$ é uma matriz de zeros de dimensão $p \times p$, $\boldsymbol{P}^d(\boldsymbol{\alpha}) = \boldsymbol{K}(\boldsymbol{\alpha})\boldsymbol{\gamma}$ com $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_r^\top)^\top$ e $\boldsymbol{K}(\boldsymbol{\alpha}) = \mathrm{diag}\left\{\alpha_1\boldsymbol{P}_1^d, \ldots, \alpha_r\boldsymbol{P}_r^d\right\}$.

Conforme observado por Vanegas e Paula (2016), a matriz $\boldsymbol{X}^*$ pode não ser singular. Dessa forma, Wood (2017) propõe que cada suavidade esteja sujeita a uma restrição de

centralização, aplicando uma reparametrização em termos de $q_k - 1$ novos parâmetros $\boldsymbol{\tau}_k$, de tal forma que $\boldsymbol{\gamma}_k = \boldsymbol{Z}_k \boldsymbol{\tau}_k$, em que $\boldsymbol{Z}_k$ é uma matriz $q_k \times (q_k - 1)$ obtida por meio da decomposição **QR**. A seguir, o procedimento será detalhado.

2.2.2 Restrição de identificabilidade

A incorporação de termos suaves no ajuste de um modelo aditivo generalizado geralmente acarreta em problemas de identificabilidade (mesmo para o modelo sem intercepto). De acordo com Wood (2017), isso ocorre porque os termos suaves não são ortogonais em relação a interceptação, bem como é impossível estimar um intercepto para cada suavizador. Desta forma, o autor propõe que cada função não paramétrica esteja sujeita a uma restrição de identificabilidade. Para isso, dentre as alternativas existentes, uma restrição adequada seria impor que a soma (ou a média) dos elementos de $\boldsymbol{f}_k$ seja zero, i.e., $\sum_{i=1}^{n} f_k(t_{ik}) = 0$. Deste modo, a ortogonalidade do termo suave para o intercepto estará garantida.

Com isso, note que $\boldsymbol{f}_k$ pode ser reescrito como $\boldsymbol{f}_k = \boldsymbol{B}_k \boldsymbol{\gamma}_k$. Assim sendo, pode-se redefinir $\sum_{i=1}^{n} f_k(t_{ik}) = 0$ matricialmente como:

$$\mathbf{1}^\top \boldsymbol{B}_k \boldsymbol{\gamma}_k = 0,$$

em que $\mathbf{1}^\top$ é um vetor de $1'$s. Especificamente, o autor sugere encontrar uma matriz $\boldsymbol{Z}_k$ com q_k linhas por $q_k - 1$ colunas ortogonais que satisfaz a seguinte condição:

$$\mathbf{1}^\top \boldsymbol{B}_k \boldsymbol{Z}_k = 0,$$

de modo que $\boldsymbol{\gamma}_k = \boldsymbol{Z}_k \boldsymbol{\tau}_k$. Diante disso, obtém-se uma nova matriz modelo para o k-ésimo termo, $\widetilde{\boldsymbol{B}}_k = \boldsymbol{B}_k \boldsymbol{Z}_k$, tal que $\boldsymbol{f}_k = \widetilde{\boldsymbol{B}}_k \boldsymbol{\tau}_k$ automaticamente satisfaz a condição de centralização.

A matriz $\boldsymbol{Z}_k$ não é formada explicitamente. Contudo, uma abordagem geral e elegante, porém bem simples, para obtê-la é por meio da decomposição QR (ou fatoração QR) de $\boldsymbol{C}_k^\top$, em que $\boldsymbol{C}_k = \mathbf{1}^\top \boldsymbol{B}_k$ (ver Wood (2017, p. 47), seção 1.8.1). A decomposição QR permite fatorar uma matriz $\mathbf{A}$ em um produto $\mathbf{A} = \mathbf{QR}$ de uma matriz ortogonal $\mathbf{Q}$ e uma matriz triangular superior $\mathbf{R}$. Uma matriz $\boldsymbol{Q}$ é dita ortogonal se suas colunas são vetores unitários ortogonais, isto é, se $\boldsymbol{Q}^\top \boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}^\top = \boldsymbol{I}$. De fato, encontra-se a decomposição QR de $\boldsymbol{C}_k^\top$ e as últimas $q_k - 1$ colunas da matriz ortogonal definem $\mathbf{Z}_k$.

É importante observar que toda matriz $\boldsymbol{A} \in \mathbb{R}^{m \times n} (m \geq n)$ possui uma fatoração QR. Se for de posto completo, então ela é única. Esta decomposição pode ser obtida por meio do algoritmo de ortogonalidade de Gram-Schmidt. Contudo, tal procedimento não é

recomendado, pois problemas computacionais podem resultar em uma matriz $\boldsymbol{Q}$ não ortogonal. Assim, costuma-se usar transformações adequadas de matrizes, dentre as quais encontra-se o método da triangulação de Householder (ou reflexão de Householder) e que possui a seguinte forma:

$$\boldsymbol{H} = \boldsymbol{I} - 2\frac{\boldsymbol{u}\boldsymbol{u}^\top}{\boldsymbol{u}^\top\boldsymbol{u}}, \tag{2.8}$$

em que $\boldsymbol{I}$ é uma matriz identidade de dimensão $m \times m$, $\boldsymbol{u} = \boldsymbol{x} + \text{sinal}(x_1)\alpha\boldsymbol{e}$, sendo que $\boldsymbol{x}$ é um vetor coluna de $\boldsymbol{A}$, tal que $\|\boldsymbol{x}\| = \alpha$ com $\|\cdot\|$ representando a norma euclidiana e $\boldsymbol{e} = (1,0,0,\ldots,0)^\top$.

Na triangulação de Householder, a decomposição QR funciona encontrando-se matrizes $\boldsymbol{H}$ apropriadas de tal maneira que ao multiplicar essas matrizes pela matriz original $\boldsymbol{A}$, obtém-se uma matriz trianular superior. No entanto, $\boldsymbol{Q}$ não é obtida explicitamente, sendo necessário repetir esse precedimento na matriz $\boldsymbol{R}$ resultante um certo número de vezes até obter uma matriz triangular superior. Assim, a matriz $\boldsymbol{Q}$ será um produto escalar de cada matriz $\boldsymbol{H}$ formada sucessivamente. Isto é:

$$\boldsymbol{Q} = \boldsymbol{H}_1\boldsymbol{H}_2\cdots\boldsymbol{H}_{m-2}\boldsymbol{H}_{m-1}. \tag{2.9}$$

Seja, por exemplo, encontrar a de composição QR por meio da triangulação de Householder da matriz $\boldsymbol{A}$ dada por:

$$\boldsymbol{A} = \begin{bmatrix} 12 & -51 & 4 \\ 6 & -16 & -68 \\ -4 & 24 & -41 \end{bmatrix}. \tag{2.10}$$

Primeiramente, é necessário obter uma reflexão que transforme a primeira coluna da matriz $\boldsymbol{A}$. Neste caso, temos que $\boldsymbol{x}_1 = (12,6,-4)^\top$. Assim, $\alpha_1 = \|\boldsymbol{x}_1\| = \sqrt{12^2 + 6^2 + (-4)^2} = 14$ e $\boldsymbol{e}_1 = (1,0,0)$. Consequentemente, obtemos $\boldsymbol{u}_1 = (12,6,-4)^\top + (14,0,0)^\top = (26,6,-4)^\top$. Portanto, aplicando a equação (2.8), tem-se que:

$$\boldsymbol{H}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - 2\frac{\begin{bmatrix} 26 \\ 6 \\ -4 \end{bmatrix}[26,6,-4]}{[26,6,-4]\begin{bmatrix} 26 \\ 6 \\ -4 \end{bmatrix}}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{2}{728} \begin{bmatrix} 676 & 156 & -104 \\ 156 & 36 & -24 \\ -104 & -24 & 16 \end{bmatrix}$$
$$= \begin{bmatrix} -6/7 & -3/7 & 2/7 \\ 3/7 & 82/91 & 6/91 \\ 2/7 & 6/91 & 87/91 \end{bmatrix}.$$

Portanto, multiplicando a matriz $\boldsymbol{H}_1$ obtida anteriormente por $\boldsymbol{A}$, consegue-se o seguinte resultado:

$$\boldsymbol{H}_1\boldsymbol{A} = \begin{bmatrix} -6/7 & -3/7 & 2/7 \\ 3/7 & 82/91 & 6/91 \\ 2/7 & 6/91 & 87/91 \end{bmatrix} \begin{bmatrix} 12 & -51 & 4 \\ 6 & -16 & -68 \\ -4 & 24 & -41 \end{bmatrix}$$
$$= \begin{bmatrix} 14 & -30 & 14 \\ 0 & -451/13 & -854/13 \\ 0 & 474/13 & -553/13 \end{bmatrix}.$$

Note, porém, que a matriz obtida ainda não é triangular superior. Para isso, é necessário que o elemento $(3,2)$ da matriz $\boldsymbol{H}_1\boldsymbol{A}$ seja zero. Assim, considere $\boldsymbol{A}^{(1)}$ formada por $(1,1)$ menor da matriz $\boldsymbol{H}_1\boldsymbol{A}$, então temos que:

$$\boldsymbol{A}^{(1)} = \begin{bmatrix} -451/13 & -854/13 \\ 474/13 & -553/13 \end{bmatrix}.$$

Agora, basta aplicar o método novamente. Neste caso, temos que $\boldsymbol{x}_2 = (-451/13, 474/13)^\top$. Logo, $\alpha_2 \approx 50.4$ e $\boldsymbol{e}_2 = (1,0)$. Note que o sinal que acompanha o primeiro elemento de $\boldsymbol{x}_2$ é negativo. Consequentemente, obtemos $\boldsymbol{u}_2 = (-451/13, 474/13)^\top - (50.4, 0)^\top = (-85.0, 474/13)^\top$. Com isso,

$$\boldsymbol{H}_2^{(1)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - 2 \begin{bmatrix} 0.84 & -0.36 \\ -0.36 & 0.15 \end{bmatrix}$$
$$= \begin{bmatrix} -0.68 & 0.72 \\ 0.72 & 0.68 \end{bmatrix}$$
$$\boldsymbol{H}_2 = \begin{bmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H}_2^{(1)} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.68 & 0.72 \\ 0 & 0.72 & 0.68 \end{bmatrix}.$$

A primeira coluna $(1,0,0)^\top$ e a primeira linha $(1,0,0)$ são adicionadas para resultar em uma matriz $\boldsymbol{H}_2$ com dimensões 3×3, a fim de manter as mesmas dimensões da matriz $\boldsymbol{A}$. Então, da mesma forma que anteriormente, deve-se multiplicar $\boldsymbol{H}_2$ por $\boldsymbol{H}_1\boldsymbol{A}$, para obter uma nova matriz $\boldsymbol{R}$ que seja triangular superior. Desse modo, temos que:

$$\begin{aligned}\boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{A} &= \begin{bmatrix}1 & 0 & 0\\ 0 & -0.68 & 0.72\\ 0 & 0.72 & 0.68\end{bmatrix}\begin{bmatrix}14 & -30 & 14\\ 0 & -451/13 & -854/13\\ 0 & 474/13 & -553/13\end{bmatrix}\\ &= \begin{bmatrix}14 & -30 & 14\\ 0 & 50.33 & 14.46\\ 0 & 0 & -76.91\end{bmatrix}.\end{aligned}$$

Finalmente a matriz resultante é triangular superior. Ou seja, $\boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{A} = \boldsymbol{R}$, de modo que $\boldsymbol{Q} = \boldsymbol{H}_1\boldsymbol{H}_2$. Assim, considerando três casas decimais de aproximação, temos que:

$$\begin{aligned}\boldsymbol{Q} = \boldsymbol{H}_1\boldsymbol{H}_2 &= \begin{bmatrix}14 & -30 & 14\\ 0 & -451/13 & -854/13\\ 0 & 474/13 & -553/13\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & -0.68 & 0.72\\ 0 & 0.72 & 0.68\end{bmatrix}\\ &= \begin{bmatrix}-0.857 & 0.502 & -0.113\\ -0.428 & -0.573 & 0.698\\ 0.286 & 0.647 & 0.707\end{bmatrix}.\end{aligned}$$

Com isso, observa-se que o uso de transformações de Household é inerentemente muito simples. Além disso, é numericamente estável, sendo portanto a melhor alternativa para a obtenção da decomposição QR.

Seja agora $\boldsymbol{C} = \mathbf{1}^\top\boldsymbol{A}$. Então isso implica que $\boldsymbol{C}^\top = \boldsymbol{A}^\top\mathbf{1}$. Nosso objetivo será encontrar a decomposição QR de $\boldsymbol{C}^\top$ por meio da triangulação de Householder. Para tanto, note que:

$$\begin{aligned}\boldsymbol{C}^\top &= \begin{bmatrix}12 & 6 & -4\\ 51 & -16 & 24\\ 4 & -68 & -41\end{bmatrix}\begin{bmatrix}1\\ 1\\ 1\end{bmatrix}\\ &= \begin{bmatrix}14\\ 59\\ -105\end{bmatrix}.\end{aligned}$$

Ou seja, $\boldsymbol{C}^\top$ é um vetor coluna dado pela soma dos elementos contidos nas linas da matriz $\boldsymbol{A}$. Desse modo, $\boldsymbol{x} = (14,59,-105)^\top$. Logo, $\alpha \approx 121.252$ e $\boldsymbol{e}_1 = (1,0,0)$. Assim sendo, obtemos $\boldsymbol{u} = (14,59,-105)^\top - (121.252,0,0)^\top = (135.252,59,-105)^\top$ e, utilizando três casas decimais de aproximação, tem-se que:

$$\boldsymbol{H} = \begin{bmatrix} -0.115 & -0.487 & 0.866 \\ -0.487 & 0.788 & 0.378 \\ 0.866 & 0.378 & 0.328 \end{bmatrix}.$$

Com isso, pré multiplicando $\boldsymbol{C}^\top$ por $\boldsymbol{H}$, obtemos:

$$\boldsymbol{HC}^\top = \begin{bmatrix} -0.115 & -0.487 & 0.866 \\ -0.487 & 0.788 & 0.378 \\ 0.866 & 0.378 & 0.328 \end{bmatrix} \begin{bmatrix} 14 \\ 59 \\ -105 \end{bmatrix} = \begin{bmatrix} -121.252 \\ 0 \\ 0 \end{bmatrix},$$

finalizando o procedimento ai mesmo. Logo, $\boldsymbol{Q} = \boldsymbol{H}^\top$. Porém, note que a matriz $\boldsymbol{H}$ é simétrica. Neste caso, $\boldsymbol{H} = \boldsymbol{H}^\top$ e, portanto, $\boldsymbol{Q} = \boldsymbol{H}$.

Como discutido anteriormente, a matriz $\boldsymbol{Z}$ não é obtida explicitamente. Mas, realizando a decomposição QR de $\boldsymbol{C}^T$, obtemos $\boldsymbol{Z}$ como sendo a matriz formada pelas últimas colunas da matriz $\boldsymbol{Q}$, excluindo-se a primeira. Assim, temos que:

$$\boldsymbol{Z} = \begin{bmatrix} -0.487 & 0.866 \\ 0.788 & 0.378 \\ 0.378 & 0.328 \end{bmatrix}.$$

Desse modo, tomando o produto $\boldsymbol{CZ}$, obtém-se:

$$\boldsymbol{CZ} = \begin{bmatrix} 14 & 59 & -105 \end{bmatrix} \begin{bmatrix} -0.487 & 0.866 \\ 0.788 & 0.378 \\ 0.378 & 0.328 \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix},$$

como era de se esperar. Além disso, embora seja redundante, considere tomar a multiplicação da matriz $\boldsymbol{Z}$ pela matriz $\boldsymbol{A}$, então obtém-se o seguinte resultado:

$$\widetilde{\boldsymbol{A}} = \boldsymbol{AZ} = \begin{bmatrix} 12 & -51 & 4 \\ 6 & -16 & -68 \\ -4 & 24 & -41 \end{bmatrix} \begin{bmatrix} -0.487 & 0.866 \\ 0.788 & 0.378 \\ 0.378 & 0.328 \end{bmatrix} = \begin{bmatrix} 35.847 & 30.968 \\ -41.211 & -23.134 \\ 5.364 & -7.834 \end{bmatrix},$$

em que $\widetilde{\boldsymbol{A}}$ será a nova matriz modelo. Observe que a soma das colunas dessa matriz é zero. Consequentemente, $\widetilde{\boldsymbol{A}}$ satisfaz a condição de centralização.

Por fim, podemos definir um modelo centrado matricialmente. Assim, para cada termo de suavização, a equação (2.2) pode ser escrita como:

$$g(\mu_i) = \widetilde{\boldsymbol{X}}_i^* \tilde{\boldsymbol{\xi}}, \tag{2.11}$$

em que $\widetilde{\boldsymbol{X}}^* = \left(\boldsymbol{X} : \widetilde{\boldsymbol{B}}_1 : \cdots : \widetilde{\boldsymbol{B}}_r\right)$ e $\tilde{\boldsymbol{\xi}}^\top = \left(\boldsymbol{\beta}^\top, \boldsymbol{\tau}_1^\top, \ldots, \boldsymbol{\tau}_r^\top\right)$. Neste caso, a primeira coluna da matriz $\boldsymbol{X}$ será formada por 1's, correspondente ao intercepto do modelo. Assim, (2.11) recai em um MLG, sendo possível escrever sua função de verossimilhança. No entanto, convém lembrar que, tomando uma quantidade q_k razoável de parâmetros com o propósito de se obter uma representação adequada das funções f_k's desconhecidas, bem como as estimativas de $\tilde{\boldsymbol{\xi}}$ pela maximização da verossimilhança ordinária, provavelmente provocará um sobre ajuste, sendo necessário, portanto, a incorporação de uma penalização à função de verossimilhança.

Convenientemente, utiliza-se penalidades que assumem uma forma quadrática nos coeficientes. É fácil mostrar que, devido ao fato de $f(\cdot)$ ser linear nos parâmetros $\boldsymbol{\gamma}_k$, a penalidade baseada na integral do quadrado da segunda derivada de $f(\cdot)$ pode sempre ser escrita como uma forma quadrática em $\boldsymbol{\gamma}_k$, como pode ser visualizado a seguir

$$\int \left[f_k''(t)\right]^2 dt = \boldsymbol{\gamma}_k^\top \boldsymbol{S}_k \boldsymbol{\gamma}_k, \tag{2.12}$$

em que $\boldsymbol{S}_k$ é uma matriz de coeficientes conhecida. Porém, devido a reparametrização, essa penalidade se converte em $\boldsymbol{\tau}_k^\top \overline{\boldsymbol{S}}_k \boldsymbol{\tau}_k$, em que $\overline{\boldsymbol{S}}_k = \boldsymbol{Z}_k^\top \boldsymbol{S}_k \boldsymbol{Z}_k$. Finalmente, no caso específico dos P-*splines*, basta fazer $\boldsymbol{S}_k = \boldsymbol{P}_k^d$.

Diante do exposto, tomando a matriz $\widetilde{\boldsymbol{X}}^*$ e o vetor de parâmetros $\tilde{\boldsymbol{\xi}}$ como definidos anteriormente, tem-se que as equações de estimação à luz do modelo reparametrizado serão, portanto, dadas por:

$$\mathrm{U}_p(\tilde{\boldsymbol{\xi}}) = \widetilde{\boldsymbol{X}}^{*\top} \boldsymbol{W} \boldsymbol{D}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) - \widetilde{\boldsymbol{P}}(\boldsymbol{\alpha}) = \boldsymbol{0}, \tag{2.13}$$

sendo que $\widetilde{\boldsymbol{P}}(\boldsymbol{\alpha}) = \mathrm{blocodiag}\left\{\boldsymbol{0}_{pp}, \widetilde{\boldsymbol{P}}^d(\boldsymbol{\alpha})\right\}$, no qual $\boldsymbol{0}_{pp}$ é uma matriz de zeros de dimensão $p \times p$, $\widetilde{\boldsymbol{P}}^d(\boldsymbol{\alpha}) = \widetilde{\boldsymbol{K}}(\boldsymbol{\alpha})\boldsymbol{\tau}$ com $\boldsymbol{\tau} = (\boldsymbol{\tau}_1^\top, \ldots, \boldsymbol{\tau}_r^\top)^\top$ e $\widetilde{\boldsymbol{K}}(\boldsymbol{\alpha}) = \mathrm{diag}\left\{\alpha_1 \boldsymbol{Z}_1^\top \boldsymbol{P}_1^d \boldsymbol{Z}_1, \ldots, \alpha_r \boldsymbol{Z}_r^\top \boldsymbol{P}_r^d \boldsymbol{Z}_r\right\}$.

A título de ilustração, assumindo um modelo parcialmente aditivo generalizado com distribuição normal para a variável resposta, então segue que:

$$\boldsymbol{y} = \widetilde{\boldsymbol{X}}^* \tilde{\boldsymbol{\xi}} + \boldsymbol{\epsilon}, \tag{2.14}$$

e consequentemente, deve-se minimizar:

$$\left\| \left(\boldsymbol{y} - \widetilde{\boldsymbol{X}}^* \tilde{\boldsymbol{\xi}}\right) \right\|^2 + \tilde{\boldsymbol{\xi}}^\top \tilde{\boldsymbol{K}}^*(\boldsymbol{\alpha}) \tilde{\boldsymbol{\xi}}. \tag{2.15}$$

O que resulta precisamente na estimativa de $\tilde{\boldsymbol{\xi}}$, dada da seguinte forma:

$$\widehat{\tilde{\boldsymbol{\xi}}} = \left[\widetilde{\boldsymbol{X}}^{*\top}\widetilde{\boldsymbol{X}}^{*} + \widetilde{\boldsymbol{K}}^{*}(\boldsymbol{\alpha})\right]^{-1}\widetilde{\boldsymbol{X}}^{*\top}\boldsymbol{y}, \tag{2.16}$$

em que $\widetilde{\boldsymbol{K}}^{*}(\boldsymbol{\alpha}) = \text{blocodiag}\left\{\boldsymbol{0}_{pp}, \widetilde{\boldsymbol{K}}(\boldsymbol{\alpha})\right\}$. Este é o único caso em que é possível obter de forma fechada as estimativas de $\tilde{\boldsymbol{\xi}}$. Para outros casos é necessário um processo iterativo capaz de obtê-las numericamente, conforme será descrito a seguir.

2.3 ESTIMAÇÃO DOS PARÂMETROS

Para estimar os parâmetros dos MPLAGs, com suavização por meio de P-*splines*, é necessário encontrar os valores de $\tilde{\boldsymbol{\xi}}$ que maximizam a função de log-verossimilhança (2.5). Tendo em vista o aspecto discutido anteriormente, é necessário que o modelo seja reparametrizado a fim de que este seja identificável. A partir desse modelo, obtém-se as equações de estimação dada por (2.13). Além disso, Cordeiro e Demétrio (2008) lembram que, na grande maioria das vezes, $\mathrm{U}_p(\tilde{\boldsymbol{\xi}})$ é um sistema de equações não-lineares que não possuem uma solução analítica. Isto é, não podem ser resolvidas diretamente para $\tilde{\boldsymbol{\xi}}$. Sendo, portanto, necessário algum procedimento numérico para resolvê-las. Deste modo, Green e Silverman (1994) mostram que, para α_k e ϕ fixos, o estimador de $\tilde{\boldsymbol{\xi}}$ pode ser obtido iterativamente através do método dos mínimos quadrados reponderados penalizados da seguinte maneira:

$$\tilde{\boldsymbol{\xi}}^{(b+1)} = \left\{\widetilde{\boldsymbol{X}}^{*\top}\boldsymbol{W}^{(b)}\widetilde{\boldsymbol{X}}^{*} + \widetilde{\boldsymbol{K}}^{*}(\boldsymbol{\alpha})\right\}^{-1}\left\{\widetilde{\boldsymbol{X}}^{*}\boldsymbol{W}^{(b)}\tilde{\boldsymbol{z}}^{(b)}\right\}, \quad b = 1,2,\cdots \tag{2.17}$$

em que $\widetilde{\boldsymbol{K}}^{*}(\boldsymbol{\alpha}) = \text{blocodiag}\left\{\boldsymbol{0}_{pp}, \widetilde{\boldsymbol{K}}(\boldsymbol{\alpha})\right\}$, $\widetilde{\boldsymbol{K}}(\boldsymbol{\alpha}) = \text{diag}\left\{\alpha_1\boldsymbol{Z}_1^{\top}\boldsymbol{P}_1^{d}\boldsymbol{Z}_1, \ldots, \alpha_r\boldsymbol{Z}_r^{\top}\boldsymbol{P}_r^{d}\boldsymbol{Z}_r\right\}$ e $\boldsymbol{P}_k^{d}$ é uma matriz $q_k \times q_k$ que depende apenas das diferenças entre os coeficientes adjacentes da base de De Boor.

De acordo com Manghi (2016), o estimador de $\tilde{\boldsymbol{\xi}}$ pode ser obtido por meio do algorítimo *backfitting* (Gauss-Siedel). O processo iterativo, conforme apresentado pelo supracitado autor, é dado por:

$$\tilde{\boldsymbol{\xi}}_k^{(b+1)} = \boldsymbol{C}_k^{(b)}\boldsymbol{z}_k^{(b+1)} \quad k = 1,2,\ldots,r, \quad \text{e} \quad b = 1,2,\ldots,$$

de tal maneira que

$$\tilde{\boldsymbol{\xi}}_k^{(b+1)} = \begin{cases} \boldsymbol{\beta}^{(b+1)}, & \text{para } k = 0; \\ \widetilde{\boldsymbol{B}}_k\boldsymbol{\tau}_k^{(b+1)}, & \text{para } k = 1,2,\ldots,r; \end{cases}$$

$$\boldsymbol{C}_k^{(b)} = \begin{cases} \left\{\boldsymbol{X}^{\top}\boldsymbol{W}^{(b)}\boldsymbol{X}\right\}^{-1}\boldsymbol{X}^{\top}\boldsymbol{W}^{(b)}, & \text{para } k = 0; \\ \left\{\widetilde{\boldsymbol{B}}_k^{\top}\boldsymbol{W}^{(b)}\widetilde{\boldsymbol{B}}_k + \widetilde{\boldsymbol{K}}(\boldsymbol{\alpha})\right\}^{-1}\widetilde{\boldsymbol{B}}_k^{\top}\boldsymbol{W}^{(b)}, & \text{para } k = 1,2,\ldots,r; \end{cases}$$

$$\boldsymbol{z}_k^{(b)} = \begin{cases} \tilde{\boldsymbol{z}}^{(b)} - \sum\limits_{k=1}^{r} \widetilde{\boldsymbol{B}}_k \boldsymbol{\tau}_k^{(b)}, & \text{para} k = 0; \\ \tilde{\boldsymbol{z}}^{(b)} - \boldsymbol{X}\boldsymbol{\beta}^{(b)} - \sum\limits_{k<u}^{r} \widetilde{\boldsymbol{B}}_u \boldsymbol{\tau}_u^{(b)} - \sum\limits_{k>u}^{r} \widetilde{\boldsymbol{B}}_u \boldsymbol{\tau}_u^{(b)}, & \text{para } k,u = 1,2,\ldots,r; \end{cases}$$

em que $\tilde{\boldsymbol{z}} = \boldsymbol{D}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) + \boldsymbol{X}\boldsymbol{\beta} - \sum\limits_{k=1}^{r} \widetilde{\boldsymbol{B}}_k \boldsymbol{\tau}_k$. Neste caso, Manghi (2016) afirma que b deve ser incrementado após k ser incrementado de 0 a r. O referido autor também mostra que, para $r = 1$, o processo iterativo para obter as estimativas de $\tilde{\boldsymbol{\xi}}$ se resumem a seguinte forma:

$$\begin{aligned} \boldsymbol{\beta}^{(b+1)} &= \left\{ \boldsymbol{X}^\top \boldsymbol{W}^{(b)} \left[\boldsymbol{I} - \boldsymbol{S}^{(b)} \right] \boldsymbol{X} \right\}^{-1} \boldsymbol{X}^\top \boldsymbol{W}^{(b)} \left[\boldsymbol{I} - \boldsymbol{S}^{(b)} \right] \boldsymbol{z}^{(b)}; \\ \boldsymbol{\tau}^{(b+1)} &= \left\{ \widetilde{\boldsymbol{B}}^\top \boldsymbol{W}^{(b)} \widetilde{\boldsymbol{B}} + \widetilde{\boldsymbol{K}}(\alpha) \right\}^{-1} \boldsymbol{X}^\top \boldsymbol{W}^{(b)} \left\{ \tilde{\boldsymbol{z}}^{(b)} - \boldsymbol{X}\boldsymbol{\beta}^{(b)} \right\}, \end{aligned}$$

em que $\boldsymbol{S} = \widetilde{\boldsymbol{B}} \left\{ \widetilde{\boldsymbol{B}}^T \boldsymbol{W}^{(b)} \widetilde{\boldsymbol{B}} + \widetilde{\boldsymbol{K}}(\alpha) \right\}^{-1} \widetilde{\boldsymbol{B}} \boldsymbol{W}$ e $\boldsymbol{I}$ é a matriz identidade. Este método foi proposto por Breimane e Friedman (1985). Além deles, para mais detalhes sobre este procedimento, pode-se consultar também Buja, Hastie e Tibshirani (1989).

Por outro lado, de acordo com Wood (2017), a estimativa de $\tilde{\boldsymbol{\xi}}$ também pode ser obtida minimizando a função

$$\left\| \sqrt{\boldsymbol{W}} \left(\tilde{\boldsymbol{z}} - \widetilde{\boldsymbol{X}}^* \tilde{\boldsymbol{\xi}} \right) \right\|^2 + \tilde{\boldsymbol{\xi}}^\top \widetilde{\boldsymbol{K}}^*(\boldsymbol{\alpha}) \tilde{\boldsymbol{\xi}}, \tag{2.18}$$

cujo processo iterativo também recai em (2.17). Além disso, para computação prática, note que

$$\left\| \sqrt{\boldsymbol{W}} \left(\tilde{\boldsymbol{z}} - \widetilde{\boldsymbol{X}}^* \tilde{\boldsymbol{\xi}} \right) \right\|^2 + \tilde{\boldsymbol{\xi}}^T \widetilde{\boldsymbol{K}}^*(\boldsymbol{\alpha}) \tilde{\boldsymbol{\xi}} = \left\| \begin{bmatrix} \sqrt{\boldsymbol{W}} & 0 \\ 0 & \boldsymbol{I} \end{bmatrix} \left(\begin{bmatrix} \tilde{\boldsymbol{z}} \\ 0 \end{bmatrix} - \begin{bmatrix} \widetilde{\boldsymbol{X}}^* \\ \boldsymbol{G} \end{bmatrix} \tilde{\boldsymbol{\xi}} \right) \right\|^2, \tag{2.19}$$

onde $\boldsymbol{G}$ é uma matriz quadrada tal que $\boldsymbol{G}^\top \boldsymbol{G} = \tilde{\boldsymbol{K}}^*(\boldsymbol{\alpha})$. Esta forma é mais conveniente, tendo em vista que o autor usa a rotina da função glm do R para a obtenção das estimativas do modelo.

Ademais, para estimar o parâmetro ϕ, seja $r_i = (y_i - \mu_i)/V_i^{1/2}$, então pode-se usar a seguinte relação:

$$\text{Var}(r_i) = \mathbb{E}\left(\left[\frac{y_i - \mu_i}{V_i^{1/2}} \right]^2 \right) = \phi.$$

Dessa forma, segundo Manghi (2016), tem-se que uma estimativa de ϕ pode ser dada tomando

$$\hat{\phi} = \frac{\sum_{i=1}^{n} \hat{r}_i^2}{\text{tr}(\boldsymbol{I} - \widehat{\boldsymbol{H}^*})}, \tag{2.20}$$

em que $\text{tr}(\boldsymbol{A})$ denota o traço da matriz $\boldsymbol{A}$ e $\boldsymbol{H}^* = \widetilde{\boldsymbol{X}}^* \left\{ \widetilde{\boldsymbol{X}}^{*\top} \widetilde{\boldsymbol{W}} \widetilde{\boldsymbol{X}}^* + \widetilde{\boldsymbol{K}}^*(\boldsymbol{\alpha}) \right\}^{-1} \widetilde{\boldsymbol{X}}^{*\top} \boldsymbol{W}$.

A seguir, serão derivados os procedimentos para obtenção dos graus de liberdade e, posteriormente, procedimentos para a seleção dos parâmetros de suavização. Outrossim, será discutido a obtenção das estimativas aproximadas para os erros padrão de $\widehat{\tilde{\boldsymbol{\xi}}}$, bem como derivar medidas que evidencie as incertezas associadas às estimativas. Isto é, serão derivados os testes de hipóteses e intervalos de confiança.

2.4 INFERÊNCIA ESTATÍSTICA

2.4.1 Graus de liberdade efetivos

Em princípio, seria necessário estimar $s_v = p + \sum_{k=1}^{r} v_k$ parâmetros, em que $v_k = q_k - 1$. Este corresponderia, em tese, aos graus de liberdade referente a estimação desses parâmetros. Contudo, em razão da penalização, há uma redução do espaço paramétrico. Isso implica que os graus de liberdade também sejam reduzidos.

Para a obtenção dos graus de liberdade efetivos, Hastie e Tibshirani (1990) usam a ideia de projetores lineares ou suavizadores, obtido a partir do ajuste do MPLAG. Pra isso, observa-se que na convergência do processo iterativo $\hat{\boldsymbol{\eta}} = \widetilde{\boldsymbol{X}}^{*}\hat{\tilde{\boldsymbol{\xi}}} = \widehat{\boldsymbol{H}^{*}}\hat{\boldsymbol{z}}$, em que

$$\boldsymbol{H}^{*} = \widetilde{\boldsymbol{X}}^{*}\boldsymbol{A}^{*-1}\widetilde{\boldsymbol{X}}^{*\top}\boldsymbol{W},$$

com $\boldsymbol{A}^{*} = \widetilde{\boldsymbol{X}}^{*\top}\boldsymbol{W}\widetilde{\boldsymbol{X}}^{*} + \widetilde{\boldsymbol{K}}^{*}(\boldsymbol{\alpha})$. Com isto, Green e Silverman (1994) definem como os graus de liberdade efetivos do modelo a soma dos autovalores do suavizador. Isto é,

$$df(\boldsymbol{\alpha}) = \operatorname{tr}\left(\widehat{\boldsymbol{H}^{*}}\right).$$

Ou seja, os graus de liberdade efetivos são dados pelo traço da matriz $\widehat{\boldsymbol{H}^{*}}$.

Logo, tomando $\hat{\boldsymbol{\eta}} = \sum_{k=0}^{r}\hat{\boldsymbol{\eta}}_k$, em que $\hat{\boldsymbol{\eta}}_k = \widetilde{\boldsymbol{X}}_k\hat{\tilde{\boldsymbol{\xi}}}_k$, com $\widetilde{\boldsymbol{X}}_k = \boldsymbol{X}$, se $k = 0$ e $\widetilde{\boldsymbol{X}}_k = \widetilde{\boldsymbol{B}}_k$, para $k = 1,2,\ldots,r$. Além disso, seja $\widehat{\boldsymbol{A}_k^{*}}^{-1}$ particionada da seguinte forma $\widehat{\boldsymbol{A}_k^{*}}^{-1} = \widehat{\boldsymbol{A}_k^{*}}^{k,u}$, para $k,u = 0,1,2,\ldots,r$. Definido então $\hat{\boldsymbol{\eta}}_k = \widehat{\boldsymbol{H}_k^{*}}\hat{\boldsymbol{z}}$, Manghi (2016) mostra que pode-se decompor o suavizador $\widehat{\boldsymbol{H}_k^{*}}$ como:

$$\widehat{\boldsymbol{H}_k^{*}} = \widetilde{\boldsymbol{X}}_k\left[\sum_{u=0}^{r}\widehat{\boldsymbol{A}_k^{*}}^{k,u}\widetilde{\boldsymbol{X}}_u^{\top}\right]\widehat{\boldsymbol{W}}, \tag{2.21}$$

para $k = 0,1,2,\ldots,r$. Desse modo, para determinar os graus de liberdade efetivos associado a k-ésima componente não-paramétrica de $\hat{\boldsymbol{\eta}}$, tem-se que:

$$df_k(\alpha_k) = \operatorname{tr}\left(\widehat{\boldsymbol{H}_k^{*}}\right).$$

Observa-se que, por meio de (2.21), $\widehat{\boldsymbol{H}}_k^* = \widetilde{\boldsymbol{B}}_k \left\{ \widetilde{\boldsymbol{B}}_k^\top \widehat{\boldsymbol{W}} \widetilde{\boldsymbol{B}}_k + \widetilde{\boldsymbol{K}}(\alpha_k) \right\}^{-1} \widetilde{\boldsymbol{B}}_k^\top \widehat{\boldsymbol{W}}$, para $k = 1,2,\ldots,r$. Logo, os graus de liberdade correspondente a cada função não paramétrica serão obtidos por:

$$\begin{aligned} df_k(\alpha_k) &= \operatorname{tr}\left(\widehat{\boldsymbol{H}}_k^*\right) \\ &= \operatorname{tr}\left(\widetilde{\boldsymbol{B}}_k \left\{ \widetilde{\boldsymbol{B}}_k^\top \widehat{\boldsymbol{W}} \widetilde{\boldsymbol{B}}_k + \widetilde{\boldsymbol{K}}(\alpha_k) \right\}^{-1} \widetilde{\boldsymbol{B}}_k^\top \widehat{\boldsymbol{W}} \right). \end{aligned}$$

Em conformidade com Eilers e Marx (1996), essa quantidade ainda pode ser reescrita da seguinte maneira:

$$\begin{aligned} df_k(\alpha_k) &= \operatorname{tr}\left(\widetilde{\boldsymbol{B}}_k^\top \widehat{\boldsymbol{W}} \widetilde{\boldsymbol{B}}_k \left\{ \widetilde{\boldsymbol{B}}_k^\top \widehat{\boldsymbol{W}} \widetilde{\boldsymbol{B}}_k + \widetilde{\boldsymbol{K}}(\alpha_k) \right\}^{-1} \right) \\ &= \operatorname{tr}\left([\boldsymbol{I}_k + \alpha_k \boldsymbol{L}_k] \right) \\ &= \sum_{j=1}^{v_k} \frac{1}{1 + \alpha_k \lambda_j}, \end{aligned}$$

para $k = 1,2,\ldots r$, em que λ_j são os autovalores associados à matriz não negativa definida $\boldsymbol{L}_k = \boldsymbol{Q}_k^{-\frac{1}{2}} \boldsymbol{P}_k^d \boldsymbol{Q}_k^{-\frac{1}{2}}$ e $\boldsymbol{Q}_k^{-\frac{1}{2}}$ é uma matriz positiva definida, de forma que $\boldsymbol{Q}_k = \widetilde{\boldsymbol{B}}_k^\top \widehat{\boldsymbol{W}} \widetilde{\boldsymbol{B}}_k$. Sendo assim, tem-se uma relação inversa entre $df_k(\alpha_k)$ e α_k. Logo, sabendo um desses parâmetros, é possível obter o outro. Consequentemente, pode-se derivar um procedimento capaz de encontrar o valor ótimo para os parâmetros de suavização.

2.4.2 Seleção do parâmetro de suavização

Conforme mencionando anteriormente, o parâmetro de suavização é usado para ponderar a quantidade ideal de suavização a fim de produzir uma função f que não interpole os dados, mas que também não seja tão suave a ponto de não representá-los. Nesta perspectiva, é necessário algum procedimento capaz de encontrar o valor ótimo desse parâmetro.

Como critério para escolha do parâmetro de suavização, Wood (2017) utiliza o *método da validação cruzada generalizada*. Este consiste em usar parte dos dados para ajustar o modelo e o restante para avaliar a adequabilidade do ajuste a fim de minimizar:

$$\operatorname{VCG}(\boldsymbol{\alpha}) = \frac{(\boldsymbol{y} - \boldsymbol{\mu})^\top \boldsymbol{V}^{-1} (\boldsymbol{y} - \boldsymbol{\mu})}{\operatorname{tr}\left(\boldsymbol{I} - \widehat{\mathbf{H}}^*\right)},$$

em que $\boldsymbol{I}$ é uma matriz identidade $\sum_{k=1}^r v_k \times \sum_{k=1}^r v_k$, com $v_k = q_k - 1$.

Ademais, conforme foi visto anteriormente, existe uma relação inversa entre os graus de liberdade efetivos e o parâmetro de suavização. Além disso, sabendo um deles, pode-se obter o outro. Na prática, é mais fácil estabelecer uma faixa de valores para os graus de

liberdade. Segundo Manghi (2016), para valores pequenos de α_k, tem-se que $df_k(\alpha_k) \approx v_k$, enquanto que para altos valores de α_k, tem-se que $df_k(\alpha_k) \approx 2$. Uma vez que $df_k(\alpha_k) \in [2, v_k]$ e $\alpha_k \in [0, \infty]$, o autor sugere considerar para a seleção do vetor de parâmetros de suavização os valores $\alpha_k \in \left[\alpha_{k(min)}, \alpha_{k(max)}\right]$, tais que:

$$df_k\left(\alpha_{k(min)}\right) - df_k\left(\alpha_{k(max)}\right) = \gamma_k(v_k - 2),$$

tomando $\gamma_k \in (0,1)$. Assim, de acordo com Manghi (2016), pode-se selecionar o vetor de suavização como sendo o vetor $\boldsymbol{\alpha}_k^* \in \left[\alpha_{1(min)}, \alpha_{1(max)}\right] \times \left[\alpha_{2(min)}, \alpha_{2(max)}\right] \times \cdots \times \left[\alpha_{r(min)}, \alpha_{r(max)}\right]$ e usa-los para minimizar (2.4.2) e, assim, selecionar o vetor de parâmetros de suavização ótimo.

Além da validação cruzada generalizada, de acordo com Wood (2017), pode-se usar também medidas de informação, como o Critério de Informação de Akaike (AIC) e Critério de Informação Bayesiano (BIC). Para mais detalhes sobre esses critérios, ver Akaike (1974) e Schwarz *et al.* (1978), respectivamente.

2.4.3 Escolha do número de nós

A suavidade da curva estimada é controlada por meio de uma penalização, que é função do parâmetro de suavização α. Portanto, a escolha da posição e do número de nós não é, necessariamente, crucial. De acordo com Ruppert (2002), deve haver uma quantidade suficiente de nós para o ajuste da curva, mas qualquer incremento a partir desse número mínimo necessário de nós tem pouco efeito no ajuste. Porém, existem exemplos em que uma quantidade de nós acima do mínimo necessário aumenta moderadamente o erro quadrático médio. Assim, o autor recomenda escolher uma quantidade de nós de modo a garantir que este seja suficientemente grande para o ajuste dos dados, mas que não seja tão grande que possa resultar em um erro quadrático médio maior que o necessário.

Um dos algoritmos proposto pelo autor considera utilizar $5, 10, 20, 40, 80$ e 120 como valores de avaliação para a quantidade de nós necessária. Inicialmente, é ajustado um P-*spline* com quantidade de nós iguais a 5 e 10. Em cada caso, um α é selecionado para minimizar o critério VCG de acordo com o número de nós. Então, compara-se os valores obtidos pelo critério VCG e escolhe-se a quantidade de nós que resultou no menor valor do critério. Assim, se para a quantidade de nós igual a 10 o critério VCG for maior que 0.98 vezes o valor do critério VCG obtido com 5 nós, então conclui-se que é improvável que novos aumentos no número de nós diminuam o VCG e usa-se 5 nós ou 10 (o que tiver menor VCG). Caso contrário, calcula-se um P-*spline* com 20 nós e compara-se o valor do critério VCG obtido com o valor do critério VCG

obtido com 10 nós. Se para a quantidade de nós igual a 20 o critério VCG for maior que 0.98 vezes o valor do critério VCG obtido com 10 nós, então usa-se 10 nós ou 20 (o que tiver menor VCG). Caso contrário, calcula-se um P-*spline* com 40 nós e assim por diante.

2.4.4 Intervalos de confiança

Nas seções anteriores foram apresentados procedimentos para obter as estimativas pontuais para os parâmetros $\tilde{\boldsymbol{\xi}}$ e $\boldsymbol{\alpha}$. Entretanto, também é interessante uma evidência quantitativa a respeito da incerteza que envolve essas estimativas. Em particular, seria útil realizar testes de hipóteses a cerca dos parâmetros, bem como poder encontrar seus respectivos intervalos de confiança. Para isso, primeiramente é necessário obter os erros-padrão aproximados associados às estimativas.

Desta forma, Manghi (2016) observa que, na convergência do processo iterativo (dado $\boldsymbol{\alpha}$),

$$\widehat{\tilde{\boldsymbol{\xi}}} = \widehat{\boldsymbol{A}^*}^{-1} \widetilde{\boldsymbol{X}}^{*\top} \widehat{\boldsymbol{W}^*} \widehat{\boldsymbol{z}}. \tag{2.22}$$

Assumindo que $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ é o verdadeiro vetor de médias de $\boldsymbol{y}$ em que $\hat{\boldsymbol{\mu}} \approx \boldsymbol{\mu}_0$, tem-se que $\widehat{\tilde{\boldsymbol{\xi}}} = \boldsymbol{A}_0^{*-1} \widetilde{\boldsymbol{X}}^{*\top} \boldsymbol{W}_0 \boldsymbol{z}_0 \approx \boldsymbol{A}^{*-1} \widetilde{\boldsymbol{X}}^{*\top} \boldsymbol{W}^* \boldsymbol{z}$. Assim, sabendo que $\boldsymbol{y}$ possui matriz de variância e covariância dada por $\boldsymbol{V}^{-1}\phi$, então:

$$\begin{aligned} \operatorname{Var}\left(\widehat{\tilde{\boldsymbol{\xi}}}\right) &\approx \boldsymbol{A}^{*-1} \widetilde{\boldsymbol{X}}^{*\top} \boldsymbol{W} \boldsymbol{D}^{-1} \operatorname{Var}(\boldsymbol{y}) \boldsymbol{D}^{-1} \boldsymbol{W} \widetilde{\boldsymbol{X}}^{*} \boldsymbol{A}^{*-1} \\ &\approx \boldsymbol{A}^{*-1} \widetilde{\boldsymbol{X}}^{*\top} \boldsymbol{W} \widetilde{\boldsymbol{X}}^{*} \boldsymbol{A}^{*-1} \phi \end{aligned} \tag{2.23}$$

será a matriz de variância e covariância do estimador $\widehat{\tilde{\boldsymbol{\xi}}}$. Com efeito, a estimativa de $\operatorname{Var}\left(\widehat{\tilde{\boldsymbol{\xi}}}\right)$ será dada por:

$$\widehat{\operatorname{Var}}\left(\widehat{\tilde{\boldsymbol{\xi}}}\right) = \widehat{\boldsymbol{A}^*}^{-1} \widetilde{\boldsymbol{X}}^{*\top} \widehat{\boldsymbol{W}} \widetilde{\boldsymbol{X}}^{*} \widehat{\boldsymbol{A}^*}^{-1} \hat{\phi}. \tag{2.24}$$

Para a normalidade de $\boldsymbol{y}$ ou normalidade multivariada assintótica de $\widetilde{\boldsymbol{X}}^{*\top} \boldsymbol{W} \boldsymbol{z}$, segue-se que, aproximadamente:

$$\widehat{\tilde{\boldsymbol{\xi}}} \sim \mathrm{N}\left(\mathbb{E}\left(\widehat{\tilde{\boldsymbol{\xi}}}\right), \widehat{\operatorname{Var}}\left(\widehat{\tilde{\boldsymbol{\xi}}}\right)\right).$$

No entanto, Wood (2017) aponta para o fato de que geralmente, $\mathbb{E}\left(\widehat{\tilde{\boldsymbol{\xi}}}\right) \neq \tilde{\boldsymbol{\xi}}$ e, portanto, é inapropriado usar esse resultado para realizar testes de hipóteses e calcular os intervalos de confiança. Contudo, se $\tilde{\boldsymbol{\xi}} = 0$, então $\mathbb{E}\left(\widehat{\tilde{\boldsymbol{\xi}}}\right) = 0$ e o resultado anterior pode ser útil para testar a inclusão de termos no modelo.

Alternativamente, pode-se usar a abordagem bayesiana para estimar a incerteza. Neste caso, de acordo com Wood (2017), a matriz de variâncias e covariâncias é baseada na distribuição a posteriori dos parâmetros que, ainda segundo o autor, é dada da seguinte forma:

$$\widehat{\mathcal{V}\mathrm{ar}}\left(\widehat{\tilde{\boldsymbol{\xi}}}\right) = \widehat{\boldsymbol{A}^*}^{-1}\hat{\phi},$$

em que $\widehat{\boldsymbol{A}^*} = \widetilde{\boldsymbol{X}}^{*\top}\widehat{\boldsymbol{W}}\widetilde{\boldsymbol{X}}^* + \widetilde{\boldsymbol{K}}^*(\boldsymbol{\alpha})$, o que resulta em uma distribuição posteriori dada por:

$$\widehat{\tilde{\boldsymbol{\xi}}}|\mathbf{Y} = \mathbf{y} \sim \mathrm{N}\left(\widehat{\tilde{\boldsymbol{\xi}}}, \widehat{\mathcal{V}\mathrm{ar}}\left(\widehat{\tilde{\boldsymbol{\xi}}}\right)\right).$$

Para dados não normais a aproximação tem justificativa baseada no teorema central do limite. Logo, este resultado pode ser usado diretamente para calcular intervalos de credibilidade para os parâmetros.

Além disso, a partir dos intervalos de credibilidade para o vetor $\tilde{\boldsymbol{\xi}}$, também é possível obter os intervalos para quantidades dele derivadas, tais como seus próprios termos suaves. De fato, observe que:

$$\widehat{\mathrm{Var}}\left(\widetilde{\boldsymbol{X}}^*\widehat{\tilde{\boldsymbol{\xi}}}\right) = \widetilde{\boldsymbol{X}}^*\widehat{\boldsymbol{A}^*}^{-1}\tilde{\boldsymbol{X}}^{*T}\hat{\phi}.$$

Com isso, um intervalo de credibilidade assinótico pontual de 95% para $g(\mu_i)$ pode ser dado por: $\mathrm{IC}(g(\mu_i);0,95) = \left[\tilde{\boldsymbol{x}}_i^*\widehat{\tilde{\boldsymbol{\xi}}} \pm 2\sqrt{c_{ii}}\right]$, em que c_{ii} é o i-ésimo elemento da diagonal de $\widehat{\mathrm{Var}}\left(\widetilde{\boldsymbol{X}}^*\widehat{\tilde{\boldsymbol{\xi}}}\right)$. De igual forma, pode-se conseguir um intervalo de credibilidade para μ_i, bastando apenas aplicar a inversa da função $g(\cdot)$ de tal forma que $\mathrm{IC}(\mu_i;0,95) = \left[g^{-1}\left(\tilde{\boldsymbol{x}}_i^*\widehat{\tilde{\boldsymbol{\xi}}} \pm 2\sqrt{c_{ii}}\right)\right]$.

2.4.5 Teste de hipóteses

Os testes de hipóteses podem ser usados para a seleção de modelos, especialmente quando existem razões para favorecer modelos mais simples. No contexto dos MPLAGs, seja $\boldsymbol{\tau}_k$ o vetor que contém os coeficientes para um único termo suave, então $\mathbb{E}(\hat{\boldsymbol{\tau}}_k) \approx 0$ se $\boldsymbol{\tau}_k = 0$. Desta forma, considerando testar a hipótese $\mathrm{H}_0 : \boldsymbol{\tau}_k = \mathbf{0}$, então pode-se usar o resultado obtido em (2.4.4) para realizar esse teste. Assim, seja $\mathrm{Var}(\hat{\boldsymbol{\tau}}_k)$ denotando o bloco de $\mathrm{Var}\left(\widehat{\tilde{\boldsymbol{\xi}}}\right)$, definida em (2.23), correspondente à $\boldsymbol{\tau}_k$, tem-se que:

$$\mathrm{W}_{\boldsymbol{\tau}_k} = \hat{\boldsymbol{\tau}}_k^\top \mathrm{Var}(\hat{\boldsymbol{\tau}}_k)^{-1}\hat{\boldsymbol{\tau}}_k. \tag{2.25}$$

Que, sob H_0, segue assintoticamente uma distribuição $\chi^2_{v_k}$. Logo,

$$\hat{\boldsymbol{\tau}}_k^\top \mathrm{Var}(\hat{\boldsymbol{\tau}}_k)^{-1}\hat{\boldsymbol{\tau}}_k / v_k \sim F_{v_k, n-d}, \tag{2.26}$$

em que d e v_k são, respectivamente, as dimensões de $\tilde{\boldsymbol{\xi}}$ (desconsiderando os termos paramétricos) e $\boldsymbol{\tau}_k$.

Além disso, substituindo $(\boldsymbol{C}\hat{\boldsymbol{\tau}}_k - \boldsymbol{u})^\top \left(\boldsymbol{C}\text{Var}(\hat{\boldsymbol{\tau}}_k)\boldsymbol{C}^\top\right)^{-1}(\boldsymbol{C}\hat{\boldsymbol{\tau}}_k - \boldsymbol{u})$ no lugar de $\hat{\boldsymbol{\tau}}_k^\top \text{Var}(\hat{\boldsymbol{\tau}}_k)^{-1}\hat{\boldsymbol{\tau}}_k$, pode-se também testar hipóteses mais gerais, tais como: $\text{H}_0 : \mathbf{C}\boldsymbol{\tau}_k = \mathbf{u}$. Esse resultado é igualmente útil, quando se deseja testar casos que envolvem um único parâmetro, podendo usar equivalentemente os testes t ou $\text{N}(0,1)$ como distribuições de referência.

Porém, como penalizar a função de verossimilhança leva a uma redução do espaço paramétrico, Wood (2017) destaca que o cálculo do valor p é menos direto, sendo possível levar o teste a uma perda de poder. Portanto, o supracitado autor recomenda, entre outras propostas, utilizar $df_k(\boldsymbol{\alpha})$ no lugar de v_k.

A seguir, será apresentado um exemplo prático. O intuito será apenas destacar os pormenores no que tange os resultados obtidos até aqui. Em particular, o objetivo é descrever com certo detalhe como o processo de suavização é realizado de forma prática, com indicação dos passos a serem tomados, bem como apresentar os códigos desenvolvidos em R para a obtenção dos resultados.

2.4.6 Impacto do capacete da motocicleta

A título de exemplificação, considere novamente o exemplo exposto na **Seção** 1.3.1. Na supracitada seção, foi apresentado um experimento cujo objetivo era investigar comportamento de capacetes durante o impacto na cabeça em acidentes simulados. Para tanto, é considerada aceleração da cabeça em unidades de g, imediatamente após o impacto. A **Figura** 8 mostra o diagrama de dispersão da aceleração contra o tempo dado em milissegundo (ms) para este experimento.

Estes dados podem ser encontrados em Härdle (1990), cujo estudo é baseado em uma amostra de tamanho $n = 133$ referente a um experimento que consiste em medir a quantidade de energia absorvida por capacetes imediatamente após o impacto. Tal experimento, em geral, considera um impactador com uma determinada massa, sujeito à queda livre, ao qual é acoplada uma célula de carga que proverá pontos de dados de carga contra o tempo de contato entre o impactador e o alvo. Este experimento é descrito em detalhes por Schmidt, Mattern e Schüler (1981). Contudo, segue uma breve análise descritiva do problema.

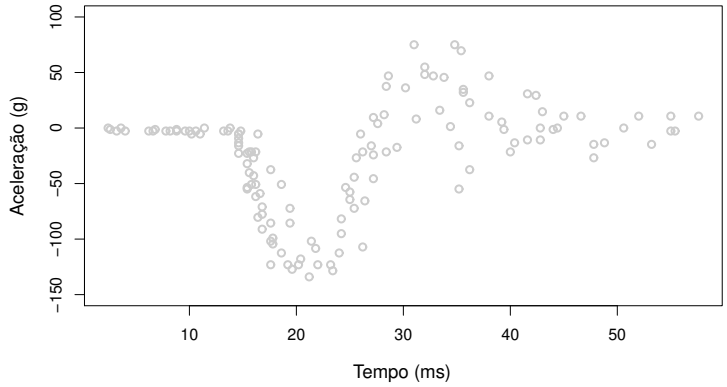
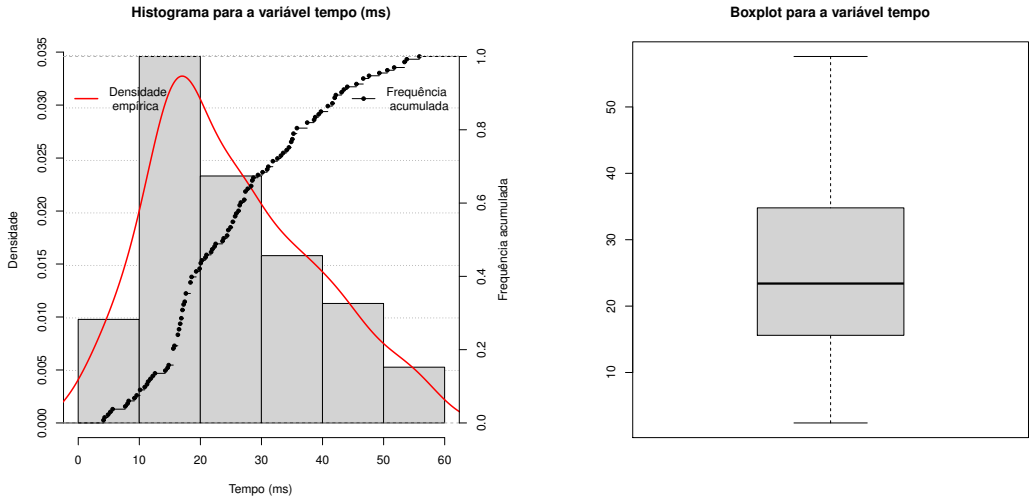


Figura 8 – Velocidade do impactador em função do tempo de contato entre o projétil e o alvo para uma energia de impacto constante.

Para este conjunto de dados, observou-se que o tempo médio foi de aproximadamente $\bar{x} = 25.18$ ms com desvio padrão de $s = 13.13$ ms, bem como coeficiente de variação dado por $cv = 52.15\%$. Além disso, o tempo mínimo observado foi de $x_{(1)} = 2.40$ ms com tempo máximo de $x_{(n)} = 57.60$ ms, com o primeiro quartil, mediana e terceiro quartil dados respectivamente por $Q_1 = 15.60$ ms, $\text{Md} = 23.40$ ms e $Q_3 = 34.80$ ms.

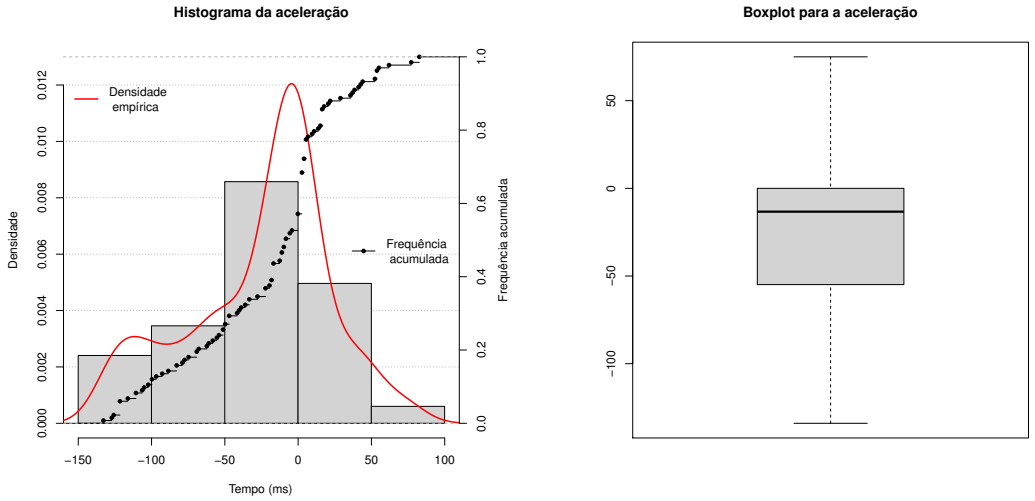


(a) **(b)**

Figura 9 – Gráficos da distribuição de frequência, frequência acumulada e boxplot, respectivamente, para a variável tempo.

De igual forma, destaca-se que a média da *aceleração* foi de aproximadamente $\bar{x} = -25.54$ g, com desvio padrão de $s = 48.32$ g, bem como a menor e a maior acelerações observadas foram de $x_{(1)} = -134$ g e $x_{(n)} = 75$ g, respectivamente. Ademais, o primeiro quartil, mediana e terceiro quartil foram dados respectivamente por $Q_1 = -54.90$ g, $\text{Md} = -13.30$ g e

$Q_3 = 0.00$ g.



(a) **(b)**

Figura 10 – Gráficos da distribuição de frequência, frequência acumulada e boxplot, respectivamente, para a variável aceleração.

Como pode-se observar, o comportamento da aceleração em função do tempo claramente dar-se de forma não linear. Com isso, assumir o modelo normal clássico, i.e., com o preditor dado de forma linear, para analisar tal comportamento é inviável. Portanto, consideraram erros aditivos e distribuição normal para ajuste do modelo cuja variável tempo é ajustada de forma não paramétrica e assumindo y_i como sendo a aceleração, o modelo adotado será dado por

$$y_i|\text{Tempo}_i \sim \text{N}(\mu_i, \sigma^2),$$
$$g(\mu_i) = \mu_i = \eta_i = f(\text{Tempo}_i),$$

em que $f(\cdot)$ é uma função suave definida como em (1.1).

Para o ajuste desse modelo, usou-se uma função em R dado pelo **Código-fonte 3**, baseado na teoria desenvolvida nas seções anteriores. Neste código, especificando os nós e α, a função psfit retorna o ajuste do modelo baseado na função glm, o valor do critério VCG, sigma e o traço da matriz H (correspondente ao grau de liberdade do modelo). Para a escolha do α ótimo pode-se usar, por exemplo, o **Código-fonte 4**. Tanto o **Código-fonte 3** quanto o **Código-fonte 4**, bem como outros códigos usados para manipulação da metodologia apresentada nas seções subsequentes, são disponibilizados no **Apêndice** A.

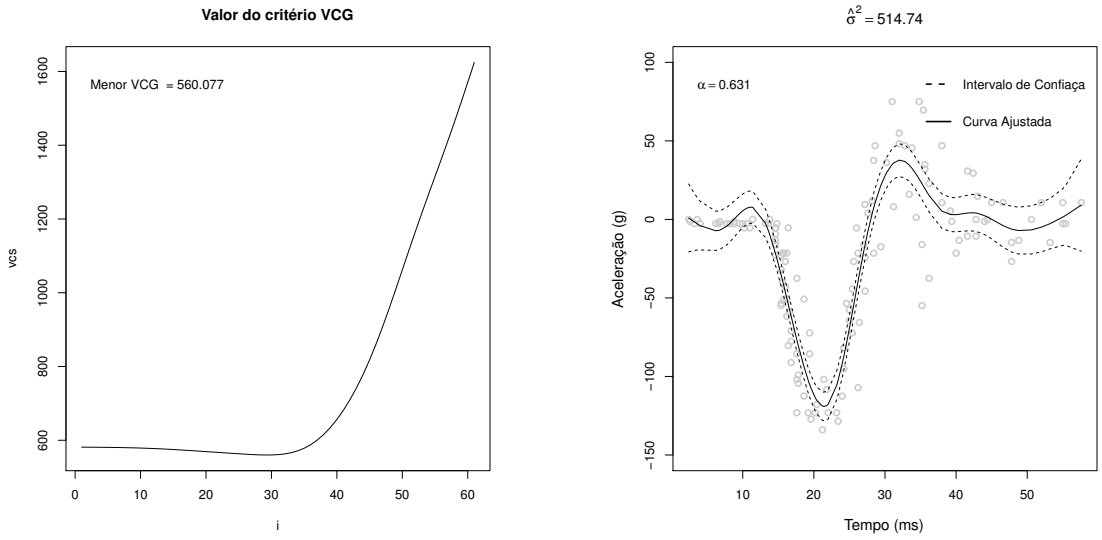
Assim, ajustando o modelo proposto, obtém-se as estimativas de máxima verossimilhança penalizadas apresentadas na **Tabela** 0. Destaca-se que para o ajuste desse modelo foi

utilizado o procedimento de P-*spline* para estimar $f(\cdot)$ por meio de um B-*spline* cúbico com 18 nós equidistante e penalização quadrática. Além disso, foi empregado o método da validação cruzada generalizada (VCG) para estimar o parâmetro de suavização α. Sobre os graus de liberdade, por definição, tem-se que $df(\alpha) = 1 + df_1(\alpha)$, em que 1 refere-se ao intercepto do modelo e $df_1(\alpha)$ corresponde ao grau de liberdade da variável suavizada.

Tabela 0 – Resumo das estimativas referentes ao modelo ajustado para o exemplo do impacto do capacete.

Efeito	**Estimativa**	**Erro-Padrão**	**Valor-p**
Intercepto	-25.546	1.967	< 0.0001
Outras medidas do ajuste			
$df(\alpha)$	10.767		
α	0.631		
σ^2	514.738		
VCG	560.077		

Portanto, por meio da **Figura** 11, verifica-se o bom ajuste da curva, bem como é possível notar uma maior variabilidade para acelerações com tempo acima de 40 e abaixo de 10 milissegundos, destacado pelo intervalo de confiança pontual de 95%, com base nos resultados apresentados na **Seção** 2.4.4.



(a) **(b)**

Figura 11 – O valor do critério VCG, curva aproximada e intervalos de confiança pontuais de 95% **para a aceleração contra o tempo.**

No entanto, para verificar a viabilidade do modelo, é necessário ainda uma análise

subsequente ao ajuste. Esta é responsável por verificar afastamento das suposições adotadas, bem como corrigi-las com objetivo de obter um modelo mais fiel à realidade do fenômeno em questão. No próximo capítulo, serão detalhados os procedimentos pelos quais se faz essa análise seguida de exemplificações.

3 TÉCNICAS DE DIAGNÓSTICO NOS MPLAGS

Modelos estatísticos são descrições aproximadas de processos bastante complexos, consequentemente podem levar a resultados imprecisos. Isto posto, surge uma importante motivação para o estudo de técnicas que avaliem essa inexatidão. Para tanto, foram desenvolvidos métodos para avaliarem a qualidade de um ajuste de regressão. Esses métodos se inserem no ramo da estatística conhecida como *análise de diagnóstico* que, de acordo com Billor e Loynes (1993), se iniciou com a análise de resíduos para detectar a presença de pontos extremos e avaliar a adequação da distribuição proposta para a variável resposta.

A ideia básica da análise de diagnóstico é verificar se, após o ajuste do modelo, o resultado obtido é satisfatório. Isso é necessário, tendo em vista que erros sistemáticos, ou até mesmo algumas observações discrepantes, podem contribuir de forma inadequada na estimação dos parâmetros do modelo. Os erros sistemáticos são identificáveis e, em geral, podem ser corrigidos. Estão associados à escolha inadequada da função de ligação ou pela escolha equivocada para a distribuição da variável resposta. Com relação a ocorrência das observações discrepantes, pode ser que, de fato, estejam erradas como resultado de uma leitura incorreta, ou porque algum fator não controlado acarretou na sua obtenção, ou porque são observações que ocorreriam com pouca frequência caso fosse realizado outro processo de amostragem para obtenção de dados provenientes de uma mesma população. Em qualquer um desses casos é necessário uma análise mais cuidadosa em torno dessas observações, que exercem um peso desproporcional na obtenção das estimativas, a fim de identificar quais foram as possíveis razões ou fatores que levaram a sua ocorrência.

Na prática, pode haver uma combinação dos diferentes tipos de erros. Esse fato, conforme observado por Cordeiro e Demétrio (2008), faz com que a verificação da adequação de um modelo para um determinado conjunto de observações seja um processo realmente difícil. Contudo, foram desenvolvidas várias técnicas com o intuito de superar essa dificuldade. A análise de resíduo, por exemplo, é uma das técnicas mais antigas e mais utilizadas. Esta técnica possibilita identificar comportamentos atípicos para o componente aleatório esperado para uma determinada especificação da distribuição da variável resposta. Uma forma bastante geral para definir resíduos pode ser vista em Cox e Snell (1968).

Para lidar com observações que influenciam nas estimativas dos parâmetros do modelo, Hoaglin e Welsch (1978) propõem estudar a matriz de projeção do modelo linear padrão $\boldsymbol{H} = \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}$. Observações influentes tendem a elevar os valores da diagonal desta

matriz. Com isso, pode-se identificar pontos que destoam da grande maioria das observações em $\boldsymbol{X}$ e causam um peso desproporcional no valor ajustado. Por essa razão, esses pontos recebem o nome de pontos de alavanca. Existem generalizações dessa medida para a classe dos MLGs.

Segundo Paula (2013), uma das técnicas mais utilizada é a deleção de pontos que visa avaliar o impacto da retirada de uma observação particular nas estimativas da regressão. Porém, ele destaca que uma das propostas mais inovadoras na área de diagnóstico é a análise de **influência local**, desenvolvida por Cook (1978). Em seus estudos, "Cook propõe avaliar a influência conjunta das observações sob pequenas mudanças (perturbações) no modelo ou nos dados, ao invés da avaliação pela retirada individual ou conjunta de pontos" (PAULA, 2013, p. 45). Para mais detalhes sobre esta técnica e as demais citadas anteriormente, sugere-se ver Paula (2013) e Cordeiro e Demétrio (2008).

3.1 ANÁLISE DE RESÍDUOS

A análise de resíduos é uma maneira eficiente de verificar a qualidade do ajuste do modelo. Nela busca-se identificar valores da variável resposta que apresentem algum comportamento atípico. Isto é, que possuem comportamento fora do padrão esperado em relação ao assumido para a variável y_i. Neste caso, os resíduos são uma grande fonte de informação, visto que por meio de seu comportamento é possível detectar a presença de pontos aberrantes, identificar a relevância de um fator omitido no modelo, assim como verificar se há indícios de afastamento sérios da distribuição assumida para o erro. Em resumo, por meio da análise de resíduos, é possível investigar possíveis falhas de suposições feitas a respeito do modelo, tais como: a escolha da função de ligação ou da distribuição da variável resposta.

3.1.1 Resíduos de Pearson

No modelo de regressão linear, os resíduos mais utilizados são os *studentizados*. Segundo Cordeiro e Demétrio (2008), estes resíduos, sob normalidade, possuem uma distribuição t de Student com $(n-p-1)$ graus de liberdade. No caso dos MLG's, os resíduos de Pearson são os mais simples e os mais comuns. Em conformidade com Green e Silverman (1994), estes resíduos também podem ser utilizados para detectar a presença de *outliers* nos MPLAG's. Com isso, sob a suposição de independência e considerando ϕ fixo, estes resíduos são dados por:

$$\hat{r}_i^P = \phi^{-1/2}\left(\frac{y_i - \hat{\mu}_i}{\hat{V}_i^{1/2}}\right),$$

com $i = 1,2,\ldots,n$. Como $\mathbb{E}(y_i) = \mu_i$ e $\text{Var}(y_i) = \phi V_i$, se $\hat{\mu}_i = \mu_i$, então $\mathbb{E}(\hat{r}_i^P) \approx 0$ e $\text{Var}(\hat{r}_i^P) \approx 1$. Portanto, de acordo com Manghi (2016), pode-se considerar como observações extremas aquelas que possuírem altos valores dos resíduos de Pearson.

3.1.2 Resíduos quantílicos aleatorizados

Alternativamente aos resíduos de Pearson, pode-se utilizar os resíduos quantílicos aleatorizados, que foram introduzidos por Dunn e Smyth (1996). Estes resíduos apresentam distribuição normal, independente da distribuição da variável resposta e de sua dispersão. Assumindo que y_i é uma variável contínua, os resíduos quantílicos aleatorizados são definidos da seguinte forma:

$$\hat{r}_i^q = \Phi^{-1}\left\{F(y_i;\hat{\mu}_i,\phi)\right\}, \text{ para } i = 1,2,\ldots,n,$$

em que Φ é a função da distribuição acumulada da normal padrão e $F(y_i;\hat{\mu}_i,\phi)$ é a função de distribuição acumulada da distribuição da variável resposta. Como se sabe, pelo Teorema da inversão, $F(y_i;\hat{\mu}_i,\phi) \sim \text{U}(0,1)$. Assim, se os parâmetros do modelo são consistentemente estimados, $\hat{r}_i^q$ convergem para uma distribuição normal.

Para o caso em que a distribuição da variável resposta é discreta, o recurso da aleatorização é aplicado de tal forma que:

$$\hat{r}_i^q = \Phi^{-1}\left\{u_i\right\},$$

em que u_i é uma variável aleatória uniformemente distribuída no intervalo $(a_i,b_i]$, com $a_i = F(y_i - 1;\hat{\mu}_i)$ e $b_i = F(y_i;\hat{\mu}_i)$, para $i = 1,2,\ldots,n$. Da mesma forma que para o caso em que a variável resposta é contínua, se os parâmetros do modelo são consistentemente estimados, então a distribuição $\hat{r}_i^q$ converge para uma normal padrão.

3.2 MEDIDAS DE ALAVANCAGEM

As medidas de alavancagem são usadas para identificar pontos extremos nos regressores. Nessas circunstâncias, de acordo com Paula (2013), essas observações são aquelas que contribuem de maneira excessiva para as estimativas dos parâmetros, associados a erros grandes (em valor absoluto) ou em razão de características dos regressores. As consequências disso são modelos que não representam adequadamente a realidade. Por isso, são pouco precisos em suas conclusões.

A principal ideia que está por trás do conceito de alavancagem consiste em conhecer a influência que cada observação exerce no próprio valor ajustado $\hat{y}_i$. Esta influência, de acordo com Wei, Hu e Fung (1998), pode ser bem representada pela derivada $\partial\hat{y}_i/\partial y_i$. Assim, no caso dos MPLAGs, tem-se que:

$$\widehat{\mathrm{GL}} = \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{y}^\top} = \left\{ \mathbf{D}_{\tilde{\boldsymbol{\xi}}} \left(-\ddot{\mathcal{F}}_{\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}} \right)^{-1} \ddot{\mathbf{L}}_{\tilde{\boldsymbol{\xi}}\boldsymbol{y}} \right\} \Bigg|_{\tilde{\boldsymbol{\xi}}=\widehat{\tilde{\boldsymbol{\xi}}}}, \tag{3.1}$$

em que $\mathbf{D}_{\tilde{\boldsymbol{\xi}}} = \partial\boldsymbol{\mu}/\partial\tilde{\boldsymbol{\xi}}^\top$, $\ddot{\mathcal{F}}_{\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}} = \partial^2 \mathrm{L}_p(\tilde{\boldsymbol{\xi}})/\partial\tilde{\boldsymbol{\xi}}^\top\partial\tilde{\boldsymbol{\xi}}^\top$ e $\ddot{\mathbf{L}}_{\tilde{\boldsymbol{\xi}}\boldsymbol{y}} = \partial^2 \mathrm{L}_p(\tilde{\boldsymbol{\xi}})/\partial\tilde{\boldsymbol{\xi}}^\top\partial\boldsymbol{y}^\top$. Observe que

$$\mathbf{D}_{\tilde{\boldsymbol{\xi}}} = \mathbf{D}\mathbf{X}^* \text{ e } \ddot{\mathbf{L}}_{\tilde{\boldsymbol{\xi}}\boldsymbol{y}} = \mathbf{X}^{*\top}\mathbf{W}\mathbf{D}^{-1}.$$

Além disso, tem-se que:

$$\ddot{\mathcal{F}}_{\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}} = \widetilde{\boldsymbol{X}}^{*\top}[\boldsymbol{M}_1 + \boldsymbol{M}_2]\widetilde{\boldsymbol{X}}^* - \widetilde{\boldsymbol{K}}^*(\boldsymbol{\alpha}),$$

que é a informação de Fisher observada desde que tenhamos $\boldsymbol{M}_1 = [\boldsymbol{D}\dot{\mathbf{V}}^{-1}\boldsymbol{D} + \dot{\mathbf{D}}\boldsymbol{V}^{-1}]\boldsymbol{C}$, $\dot{\mathbf{D}} = \mathrm{diag}\{\partial^2\mu_1/\partial\eta_1^2, \ldots, \partial^2\mu_n/\partial\eta_n^2\}$, bem como $\dot{\mathbf{V}} = \mathrm{diag}\{\partial V_1/\partial\mu_1, \ldots \partial V_n/\partial\mu_n\}$, com $\boldsymbol{C} = \mathrm{diag}\{C_1, \ldots C_n\}$, no qual $C_i = y_i - \mu_i$ e $\boldsymbol{M}_2 = -\boldsymbol{W}$.

Note que, tomando o valor esperado de $-\ddot{\mathcal{F}}_{\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}}$, isto é, o valor esperado do negativo da informação de Fisher, então temos que $-\mathbb{E}(\ddot{\mathcal{F}}_{\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}}) = \widetilde{\boldsymbol{X}}^{*\top}\boldsymbol{W}\widetilde{\boldsymbol{X}}^* + \widetilde{\boldsymbol{K}}^*(\boldsymbol{\alpha})$. Portanto, obtemos de forma aproximada

$$\widehat{\mathrm{GL}} = \widehat{\mathbf{D}}\mathbf{X}^* \left\{ \widetilde{\boldsymbol{X}}^{*\top}\widehat{\boldsymbol{W}}\widetilde{\boldsymbol{X}}^* + \widetilde{\boldsymbol{K}}^*(\boldsymbol{\alpha}) \right\}^{-1} \mathbf{X}^{*\top}\widehat{\mathbf{W}}\widehat{\mathbf{D}}^{-1}.$$

Assim, deve-se estudar os elementos da diagonal de $\widehat{\mathrm{GL}}$ a fim de identificar os pontos de alavancagem.

Contudo, a medida mais comum usada para detectar pontos extremos nos MPLAGs é baseada no estudo da diagonal de uma matriz que é análoga a matriz de projeção obtida como solução de mínimos quadrados de uma regressão normal linear ponderada. Observe que na convergência do processo iterativo 2.17 apresentado na **Seção** 2.3 referente a um MPLAG, a estimativa dos parâmetros $\tilde{\boldsymbol{\xi}}$ possui a seguinte forma:

$$\widehat{\tilde{\boldsymbol{\xi}}} = \left\{ \widetilde{\boldsymbol{X}}^{*\top}\widehat{\boldsymbol{W}}\widetilde{\boldsymbol{X}}^* + \widetilde{\boldsymbol{K}}^*(\boldsymbol{\alpha}) \right\}^{-1} \widetilde{\boldsymbol{X}}^*\widehat{\boldsymbol{W}}\widehat{\boldsymbol{z}},$$

em que $\widehat{\boldsymbol{z}} = \widehat{\boldsymbol{D}}^{-1}(\boldsymbol{y} - \widehat{\boldsymbol{\mu}}) + \widehat{\boldsymbol{\eta}}$, com $\widehat{\boldsymbol{\eta}} = \widetilde{\boldsymbol{X}}^*\widehat{\tilde{\boldsymbol{\xi}}}$. Assim, este resultado pode ser interpretado como uma regressão linear ponderada de uma variável dependente $\tilde{\boldsymbol{z}}$ sobre a matriz modelo $\widetilde{\boldsymbol{X}}^*$ usando uma matriz de pesos $\boldsymbol{W}$. Portanto, temos que:

$$\widetilde{\boldsymbol{X}}^*\widehat{\tilde{\boldsymbol{\xi}}} = \boldsymbol{X}^* \left\{ \widetilde{\boldsymbol{X}}^{*\top}\widehat{\boldsymbol{W}}\widetilde{\boldsymbol{X}}^* + \widetilde{\boldsymbol{K}}^*(\boldsymbol{\alpha}) \right\}^{-1} \widetilde{\boldsymbol{X}}^*\widehat{\boldsymbol{W}}\widehat{\boldsymbol{z}},$$

então obtemos:

$$\widehat{\boldsymbol{\eta}} = \widehat{\mathbf{H}^*}\widehat{\tilde{\mathbf{z}}},$$

chamando $\widehat{\boldsymbol{A}^*} = \left\{\widetilde{\boldsymbol{X}}^{*\top}\widehat{\boldsymbol{W}}\widetilde{\boldsymbol{X}}^* + \widetilde{\boldsymbol{K}}^*(\boldsymbol{\alpha})\right\}$ e $\widehat{\boldsymbol{H}^*} = \widetilde{\boldsymbol{X}}^*\widehat{\boldsymbol{A}^*}^{-1}\widetilde{\boldsymbol{X}}^{*\top}\widehat{\boldsymbol{W}}$ que, neste caso, representa a matriz de projeção.

Eubank (1984) discute algumas propriedades desta matriz para modelos de regressão não-paramétricos e mostra que suas propriedades são conservadas para o caso semi-paramétrico. Com isso, seja $\hat{h}^*_{ii}$ os elementos diagonais da matriz $\widehat{\boldsymbol{H}^*} = \widetilde{\boldsymbol{X}}^*\widehat{\boldsymbol{A}^*}^{-1}\widetilde{\boldsymbol{X}}^{*\top}\widehat{\boldsymbol{W}}$. Assim sendo, a distância de uma observação em relação as demais pode ser observada por $\widehat{h}^*_{ii}$. Segundo Manghi (2016), esses elementos determinam quanto o valor predito é influenciado pelo valor observado da variável resposta y_i, considerando o respectivo vetor observado das variáveis independentes.

Ainda segundo o referido autor, supondo que todos os pontos exercem a mesma influência nas estimativas dos parâmetros, espera-se que $\widehat{h}^*_{ii}$ apresente valor próximo à sua média $\overline{h}^* = \mathrm{tr}(\widehat{\boldsymbol{H}^*})/n$. Diante disso, convém examinar os pontos tais que $\widehat{h}^*_{ii} \geq 3\overline{h}^*$. Pode-se também tirar proveito do recurso gráfico, plotando os pontos $\widehat{h}^*_{ii}$ versus i, para $i = 1,2,\ldots,n$, a fim de identificar os pontos de alta alavancagem.

Sejam $\hat{w}_i$ e $\hat{v}_i$ os elementos diagonais das matrizes $\widehat{\mathbf{W}}$ e $\widehat{\mathbf{V}}$, respectivamente. Como $\hat{w}_i$ é função de $\hat{v}_i$ e que por sua vez é função de $\hat{\mu}_i$, então $\widehat{h}^*_{ii}$ também pode ser função de $\hat{\mu}_i$. Neste caso, Paula (2013) recomenda usar os gráficos de $\widehat{h}^*_{ii}$ contra o valores ajustados para detectar pontos de alavanca. Por fim, pode-se verificar que as matrizes $\widehat{\boldsymbol{H}^*}$ e $\widehat{\mathrm{GL}}$ coincidem, de modo que $\widehat{h}^*_{ii} = \widehat{\mathrm{GL}}_{ii}$, à medida que o tamanho da mostra cresce. No caso de ligação canônica essa igualdade é sempre válida para qualquer tamanho amostral.

3.3 INFLUÊNCIA LOCAL

Ajustando um modelo a um conjunto de dados, deseja-se que as estimativas obtidas não sofram mudanças substanciais quando submetidas a pequenas perturbações no modelo proposto ou nas observações. Portanto, tomando o modelo postulado como correto e comparando as estimativas obtidas através desse modelo com as estimativas decorrente de pequenas perturbações, considere o vetor de parâmetros de suavização $\boldsymbol{\alpha}$ um vetor fixo, bem como $\mathcal{F}_{\boldsymbol{\alpha}}\left(\tilde{\boldsymbol{\xi}}\right)$ uma função penalizada e pelo menos duas vezes diferenciável em $\tilde{\boldsymbol{\xi}}$ referente ao MPLAG, e, além disso, a estimativa de $\tilde{\boldsymbol{\xi}}$ seja solução da equação

$$\frac{\partial \mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}})}{\partial \tilde{\boldsymbol{\xi}}} = \Psi_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}) = \mathbf{0}. \tag{3.2}$$

Tem-se que, portanto, $\Psi_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}})$ é a função escore penalizada referente ao MPLAG. Assim, para avaliar a influência de $\boldsymbol{\omega} \in \Omega \subseteq \mathbb{R}^l$, em que $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_l)$ é um vetor de perturbação, Cook (1978) propõe a medida de influência denominada *afastamento pela verossimilhança* definida por:

$$LD(\boldsymbol{\omega}) = 2\left\{\mathcal{F}_{\boldsymbol{\alpha}}\left(\hat{\tilde{\boldsymbol{\xi}}}\right) - \mathcal{F}_{\boldsymbol{\alpha}}\left(\hat{\tilde{\boldsymbol{\xi}}}_{\boldsymbol{\omega}}\right)\right\}, \tag{3.3}$$

quando $\boldsymbol{\omega}$ varia numa vizinhança de $\boldsymbol{\omega}_0$, em que $\boldsymbol{\omega}_0$ é tal que $\mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}_{\boldsymbol{\omega}_0}) = \mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}})$. Aqui, $\hat{\tilde{\boldsymbol{\xi}}}$ e $\hat{\tilde{\boldsymbol{\xi}}}_{\boldsymbol{\omega}}$ constituem as estimativas de $\tilde{\boldsymbol{\xi}}$ sob os modelos postulado e perturbado respectivamente.

O gráfico de $LD(\boldsymbol{\omega})$ versus $\boldsymbol{\omega}$ contém informações essenciais da influência do esquema de perturbação utilizado. Esse gráfico pode ser interpretado como a superfície geométrica formado pelos valores do vetor

$$\alpha(\boldsymbol{\omega}) = \begin{pmatrix} \boldsymbol{\omega} \\ LD(\boldsymbol{\omega}) \end{pmatrix}, \tag{3.4}$$

quando $\boldsymbol{\omega}$ varia em Ω. Essa superfície é chamada de gráfico de influência. Aqui, o objetivo é avaliar como a superfície de $\alpha(\boldsymbol{\omega})$ desvia-se de seu plano tangente em torno do ponto $\boldsymbol{\omega}_0$, em que $\boldsymbol{\omega}_0$ é vetor de não perturbação.

Para isso, Cook (1978) propõe avaliar o comportamento de $LD(\boldsymbol{\omega}_0 + a\boldsymbol{d})$ versus a, com $a \in \mathbb{R}$ e $\boldsymbol{d}$ uma direção com norma unitária. Desde que $LD(\boldsymbol{\omega}_0 + a\boldsymbol{d}) = 0$ então, $LD(\boldsymbol{\omega})$ tem mínimo local em $a = 0$. Esse gráfico é chamado de linha projetada. Cada linha projetada pode ser caracterizada por uma curvatura normal $C_{\boldsymbol{d}}$ em torno de $a = 0$. Nesse caso, Cook (1978) sugere usar aquela que possui a maior curvatura.

Cook (1978) mostra que a curvatura normal correspondente à maior curvatura $C_{\boldsymbol{d}_{max}}$, é obtida considerando a direção $\boldsymbol{d}_{max}$, dada pelo maior autovetor associada ao maior autovalor da matriz definida por:

$$C_{\boldsymbol{d}} = 2\left|\boldsymbol{d}^\top \boldsymbol{F} \boldsymbol{d}\right|, \tag{3.5}$$

em que $\boldsymbol{F} = \boldsymbol{\Delta}^\top \ddot{\mathcal{F}}_{\boldsymbol{\alpha}}^{-1} \boldsymbol{\Delta}$, desde que as matrizes $\boldsymbol{\Delta}$ e $\ddot{\mathcal{F}}_{\boldsymbol{\alpha}}$ sejam definidas respectivamente como:

$$\boldsymbol{\Delta} = \frac{\partial^2 \mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial \tilde{\boldsymbol{\xi}} \partial \boldsymbol{\omega}^\top} = \frac{\partial \Psi_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\top} \tag{3.6}$$

e

$$\ddot{\mathcal{F}}_{\boldsymbol{\alpha}} = \frac{\partial^2 \mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial \tilde{\boldsymbol{\xi}} \partial \tilde{\boldsymbol{\xi}}^\top} = \frac{\partial \Psi_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial \tilde{\boldsymbol{\xi}}^\top} = \widetilde{\boldsymbol{X}}^{*\top}[\boldsymbol{M}_1 + \boldsymbol{M}_2]\widetilde{\boldsymbol{X}}^* - \widetilde{\boldsymbol{K}}^*(\boldsymbol{\alpha}), \tag{3.7}$$

sob a condição de $\boldsymbol{M}_1 = [\boldsymbol{D}\dot{\mathbf{V}}^{-1}\boldsymbol{D} + \dot{\mathbf{D}}\boldsymbol{V}^{-1}]\boldsymbol{C}$, $\dot{\mathbf{D}} = \text{diag}\left\{\partial^2\mu_1/\partial\eta_1^2,\ldots,\partial^2\mu_n/\partial\eta_n^2\right\}$, bem como $\dot{\mathbf{V}} = \text{diag}\left\{\partial V_1/\partial\mu_1,\ldots\partial V_n/\partial\mu_n\right\}$, $\boldsymbol{C} = \text{diag}\left\{C_1,\ldots C_n\right\}$, no qual $C_i = y_i - \mu_i$ e $\boldsymbol{M}_2 = -\boldsymbol{W}$ avaliados em $\tilde{\boldsymbol{\xi}} = \widehat{\tilde{\boldsymbol{\xi}}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. Note que a equação 3.7 corresponde a informação de Fisher observada para os MPLAGs. Este resultado pode ser encontrado no **Anexo** A.

Tomando f_{ii} como os elementos da diagonal principal da matriz $\boldsymbol{F}$ e, além disso, considerando a direção da i-ésima observação definida como:

$$\mathcal{C}_i = 2|f_{ii}|, \tag{3.8}$$

então pode-se utilizar o gráfico de $\mathcal{C}_i$ versus i (ordem dos dados) como técnica de influência local para avaliar a existência de observações influentes. Sugere-se considerar a i-ésima observação como influente se seu valor for maior que duas vezes a média das medidas $\mathcal{C}_i$.

Contudo, Lee, Lu e Song (2006) sugerem utilizar o desvio padrão das medidas de curvatura para definir o ponto de corte na consideração de pontos potencialmente influentes. Assim, para o ponto de corte, é sugerido utilizar $\mathcal{PC} = c_1\overline{\mathcal{C}} + c_2\text{sd}(\mathcal{C})$ em que $\overline{\mathcal{C}}$ e $\text{sd}(\mathcal{C})$ são a média e o desvio padrão das medidas $\mathcal{C}_i$, respectivamente, e c_1 e c_2 são constantes pré-fixadas. Uma possibilidade seria $c_1 = 1$ e $c_2 = 2$ ou $c_2 = 3$. Neste caso, considera-se como observações potencialmente influentes aquelas tais que $\mathcal{C}_i > \mathcal{PC}$.

Outra alternativa para definir a curvatura normal seria utilizar a negativa da informação de Fisher esperada no lugar da informação observada. Ou seja,

$$\mathcal{I}_{\boldsymbol{\alpha}} = -\mathbb{E}\left\{\ddot{\mathcal{F}}_{\boldsymbol{\alpha}}\right\} = -\mathbb{E}\left\{\frac{\partial^2\mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial\tilde{\boldsymbol{\xi}}\partial\tilde{\boldsymbol{\xi}}^\top}\right\} = \widetilde{\mathbf{X}}^{*\top}\widehat{\boldsymbol{W}}\widetilde{\mathbf{X}}^* + \widetilde{\boldsymbol{K}}^*(\boldsymbol{\alpha}). \tag{3.9}$$

Com isso, pode-se utilizar para análise de influência local baseada na curvatura normal a quantidade

$$\mathcal{C}_d\left(\tilde{\boldsymbol{\xi}}\right) = \boldsymbol{\Delta}^\top\mathcal{I}_{\boldsymbol{\alpha}}^{-1}\boldsymbol{\Delta}. \tag{3.10}$$

Adicionalmente, a fim de obter uma medida que seja invariante sob mudança de escala, Poon e Poon (1999) propõem a curvatura normal conformal definida como

$$\mathcal{B}_d\left(\tilde{\boldsymbol{\xi}}\right) = -\frac{\mathbf{d}^\top\boldsymbol{\Delta}^\top\ddot{\mathcal{F}}_{\boldsymbol{\alpha}}^{-1}\boldsymbol{\Delta}\mathbf{d}}{\sqrt{\text{tr}\left(\boldsymbol{\Delta}^\top\ddot{\mathcal{F}}_{\boldsymbol{\alpha}}^{-1}\boldsymbol{\Delta}\right)^2}}. \tag{3.11}$$

Assim, para todo direção unitária $\mathbf{d}$, $0 \leq B_d\left(\boldsymbol{\xi}\right) \leq 1$. Dessa forma, tomando a direção $\mathbf{d} = \mathbf{e}_i$, em que $\mathbf{e}_i$ é um vetor de dimensão $n \times 1$ com o valor 1 na i-ésima posição e 0 nas demais posições, pode-se investigar o gráfico de $\mathcal{B}_{\mathbf{e}_i}$ contra seus respectivos índices e considera-se como

observações potencialmente influentes aquelas tais que $\mathcal{B}_i > \mathcal{PC}$, com $\mathcal{PC} = c_1\overline{\mathcal{B}} + c_2\text{sd}(\mathcal{B})$, em que $\overline{\mathcal{B}}$ e $\text{sd}(\mathcal{B})$ são, respectivamente, a média e o desvio padrão de $\mathcal{B}_i$.

Para avaliar a qualidade do ajuste do modelo por meio da técnica de influência local, considera-se três tipos de perturbação dados pelos casos ponderados, perturbação aditiva na variável resposta e perturbação aditiva em uma das covariáveis. Billor e Loynes (1993) resumem os vários esquemas de perturbação que podem ser introduzidos através de $\boldsymbol{\omega}$, os quais podem ser divididos em dois grupos:

i) Perturbação no modelo: Neste caso, a perturbação visa modificar as suposições propostas para o modelo. Um exemplo disso, seria a violação da suposição de homoscedasticidade que, para erros normalmente distribuídos, pode ser substituída por uma suposição de heteroscedasticidade.

ii) Perturbação nos dados: Perturbar a variável resposta ou as variáveis explanatórias são exemplos de perturbação nos dados. Neste caso, as razões pelas quais se considera a perturbação nos dados residem nos possíveis erros de medida e a existência de observações aberrantes (*outliers*), em uma proporção relativamente pequena das observações, as quais podem exercer forte influência nas estimativas dos parâmetros do modelo.

3.3.1 Perturbação de casos

Neste caso, o objetivo é identificar observações influentes entre todas as observações individualmente. Isto é, buscar entender a contribuição individual de cada observação dentre todas as observações no processo de estimação do modelo proposto. Neste caso, o vetor de não perturbação será $\boldsymbol{\omega}_0$, tal que $\omega_i = 1$ para todo i, $i = 1,2,\ldots,n$. Desta forma, tem-se que as equações de estimação do modelo perturbado são dadas por:

$$\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}|\boldsymbol{\omega}) = \widetilde{\boldsymbol{X}}^{*\top}\boldsymbol{W}\boldsymbol{D}^{-1}\text{diag}\{\boldsymbol{\omega}\}(\boldsymbol{y}-\boldsymbol{\mu}) - \boldsymbol{P}(\boldsymbol{\alpha}). \tag{3.12}$$

Assim, a equação (3.6) equivale a:

$$\boldsymbol{\Delta} = \widetilde{\boldsymbol{X}}^{*\top}\boldsymbol{W}\boldsymbol{D}^{-1}\boldsymbol{C}.$$

A obtenção destes resultados podem ser conferidos no **Anexo** A.

3.3.2 Perturbação da variável resposta

Seja y_i perturbado de tal maneira que $y_{\omega_i} = y_i + \omega_i V_i^{1/2}$. Aqui, o vetor de não perturbação é tal que $\omega_i = 0$ para $i = 1,2,\ldots,n$. Assim, tem-se que as equações de estimação

para este caso são dadas por:

$$\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}|\boldsymbol{\omega}) = \widetilde{\mathbf{X}}^{*\top}\boldsymbol{W}\boldsymbol{D}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_{\omega}) - \boldsymbol{P}(\boldsymbol{\alpha}), \tag{3.13}$$

em que $\boldsymbol{\mu}_{\omega} = \boldsymbol{\mu} - \boldsymbol{V}^{1/2}\boldsymbol{\omega}$. Logo,

$$\boldsymbol{\Delta} = \widetilde{\mathbf{X}}^{*\top}\boldsymbol{W}\boldsymbol{D}^{-1}\boldsymbol{V}^{1/2}.$$

A obtenção destes resultados podem ser conferidos no **Anexo** A.

3.3.3 Perturbação em uma das variáveis explicativas

Se, no entanto, o objetivo for perturbar o vetor $\boldsymbol{x}_j$ de covariáveis da matriz de planejamento $\boldsymbol{X}$, então considere a perturbação na j-ésima coluna da matriz $\boldsymbol{X}$, tal que $x_{j\omega_i} = x_{ji} + \sigma_j\omega_j$, no qual σ_j é o desvio padrão de $\boldsymbol{x}_j$, então $\mu_{\omega_i} = g^{-1}(\eta_{\omega_i})$, sendo $\eta_{\omega_i} = \beta_1 x_{1i} + \cdots + \beta_j(x_{ji} + \sigma_j\omega_i) + \sum_{k=1}^{r} f_k(t_{ik})$, para $i = 1,\ldots,n$. Assim, as equações de estimação sob o modelo perturbado serão dadas por:

$$\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}|\boldsymbol{\omega}) = \boldsymbol{X}_{\boldsymbol{\omega}}^{*\top}\boldsymbol{W}_{\boldsymbol{\omega}}\boldsymbol{D}_{\boldsymbol{\omega}}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_{\boldsymbol{\omega}}) - \boldsymbol{P}(\boldsymbol{\alpha}) = \boldsymbol{X}_{\boldsymbol{\omega}}^{*\top}\boldsymbol{D}_{\boldsymbol{\omega}}\boldsymbol{V}_{\boldsymbol{\omega}}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_{\boldsymbol{\omega}}) - \boldsymbol{P}(\boldsymbol{\alpha}), \tag{3.14}$$

em que $\boldsymbol{X}_{\boldsymbol{\omega}}^{*} = \boldsymbol{X}^{*} + \mathbf{0}_{j\boldsymbol{\omega}}$, sendo $\mathbf{0}_{j\boldsymbol{\omega}}$ uma matriz de dimensão $n \times s_v$ com $\sigma_j\boldsymbol{\omega}$ na j-ésima coluna e 0's nas demais componentes se $j = \ell$ ou $\boldsymbol{X}_{\boldsymbol{\omega}}^{*} = \boldsymbol{X}^{*}$ se $j \neq \ell$, para $j,\ell = 1,2,\ldots p$.

Portanto, temos que a matriz $\boldsymbol{\Delta}$ será dada por $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2,\ldots,\boldsymbol{\Delta}_n)$, com $\boldsymbol{\Delta}_i$ sendo obtida da seguinte forma:

$$\boldsymbol{\Delta}_i = \boldsymbol{T}_{1\omega_i} + \boldsymbol{T}_{2\omega_i},$$

em que

$$\boldsymbol{T}_{1\omega_i} = \boldsymbol{X}_{\omega}^{*\top}\boldsymbol{D}_{\boldsymbol{\omega}}\left[\frac{\partial\boldsymbol{V}_{\boldsymbol{\omega}}^{-1}}{\partial\omega_i}(\boldsymbol{y}-\boldsymbol{\mu}_{\boldsymbol{\omega}}) + \boldsymbol{V}_{\boldsymbol{\omega}}^{-1}\frac{\partial(\boldsymbol{y}-\boldsymbol{\mu}_{\boldsymbol{\omega}})}{\partial\omega_i}\right],$$

$$\boldsymbol{T}_{2\omega_i} = \left[\frac{\partial\boldsymbol{X}_{\boldsymbol{\omega}}^{*\top}}{\partial\omega_i}\boldsymbol{D}_{\boldsymbol{\omega}} + \boldsymbol{X}_{\boldsymbol{\omega}}^{*\top}\frac{\partial\boldsymbol{D}_{\boldsymbol{\omega}}}{\partial\omega_i}\right]\boldsymbol{V}_{\boldsymbol{\omega}}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_{\boldsymbol{\omega}}).$$

Neste caso, temos que:

$$\frac{\partial\boldsymbol{V}_{\boldsymbol{\omega}}^{-1}}{\partial\omega_i} = -\boldsymbol{V}_{\boldsymbol{\omega}}^{-1}\frac{\partial\boldsymbol{V}_{\boldsymbol{\omega}}}{\partial\omega_i}\boldsymbol{V}_{\boldsymbol{\omega}}^{-1},$$

$$\frac{\partial\boldsymbol{V}_{\boldsymbol{\omega}}}{\partial\omega_i} = \frac{\partial\boldsymbol{V}_{\boldsymbol{\omega}}}{\partial\mu_{\omega_i}}\boldsymbol{D}_{\boldsymbol{\omega}},$$

$$\frac{\partial\boldsymbol{V}_{\boldsymbol{\omega}}}{\partial\mu_{\omega_i}} = \beta_j\sigma_j\mathrm{diag}\{0,\ldots,\partial^2 V_{\omega_i}/\partial\mu_{\omega_i}^2,\ldots,0\}.$$

Além disso,

$$\frac{\partial(\boldsymbol{y}-\boldsymbol{\mu}_{\boldsymbol{\omega}})}{\partial\omega_i} = -\beta_j\sigma_j\left(0,\ldots,\partial\mu_{\omega_i}/\partial\eta_{\omega_i},\ldots,0\right)^\top,$$
$$\frac{\partial \boldsymbol{D}_{\boldsymbol{\omega}}^\top}{\partial\omega_i} = \beta_j\sigma_j \text{diag}\{0,\ldots,\partial^2\mu_{\omega_i}/\partial\eta_{\omega_i}^2,\ldots,0\},$$

bem como $\dfrac{\partial \boldsymbol{X}_{\boldsymbol{\omega}}^{*\top}}{\partial\omega_i} = \dfrac{\partial \boldsymbol{X}^{*\top}}{\partial\omega_i} + \dfrac{\partial \mathbf{0}_{j\boldsymbol{\omega}}^\top}{\partial\omega_i}$, resultando em

$$\frac{\partial \boldsymbol{X}_{\boldsymbol{\omega}}^{*\top}}{\partial\omega_i} = \begin{cases} \mathbf{0}, & \text{para } j \neq \ell, \\ \sigma_j \dot{\mathbf{X}}_{\omega_i}^{*\top}, & \text{para } j = \ell, \end{cases} \tag{3.15}$$

em que $\dot{\mathbf{X}}_{\omega_i}^*$ é uma matriz com dimensão $n \times s_v$ com 1 sendo o elemento (i,j) e 0 nas demais componentes.

3.3.4 Perturbação em vetores particionados

Quando há interesse em uma parte específica do conjunto de parâmetros, Cook (1978) mostra que, considerando a partição $\tilde{\boldsymbol{\xi}} = (\tilde{\boldsymbol{\xi}}_1, \tilde{\boldsymbol{\xi}}_2)$, a função de afastamento da verossimilhança será dada por:

$$LD(\boldsymbol{\omega})_s = 2\left\{\mathcal{F}_{\boldsymbol{\alpha}}\left(\widehat{\tilde{\boldsymbol{\xi}}}\right) - \mathcal{F}_{\boldsymbol{\alpha}}\left(\widehat{\tilde{\boldsymbol{\xi}}}_{1\boldsymbol{\omega}}, g\left(\widehat{\tilde{\boldsymbol{\xi}}}_{1\boldsymbol{\omega}}\right)\right)\right\},$$

sendo $\widehat{\tilde{\boldsymbol{\xi}}}_{1\boldsymbol{\omega}}$ o vetor obtido a partir de $\left(\widehat{\tilde{\boldsymbol{\xi}}}_{1\boldsymbol{\omega}}, \widehat{\tilde{\boldsymbol{\xi}}}_{2\boldsymbol{\omega}}\right)$ e $g\left(\widehat{\tilde{\boldsymbol{\xi}}}_{1\boldsymbol{\omega}}\right)$ a função que, para $\tilde{\boldsymbol{\xi}}_1$ fixado, maximiza $\mathcal{F}_{\boldsymbol{\alpha}}\left(\tilde{\boldsymbol{\xi}}_1, \tilde{\boldsymbol{\xi}}_2\right)$.

Consequentemente, a superfície será dada por:

$$\alpha_s(\boldsymbol{\omega}) = \begin{pmatrix} \boldsymbol{\omega} \\ LD(\boldsymbol{\omega})_s \end{pmatrix},$$

resultando na curvatura normal segundo uma direção unitária $\boldsymbol{d}$ dada por:

$$C\left(\boldsymbol{\xi}_1\right)_{\boldsymbol{d}} = 2\left|\boldsymbol{d}^T \boldsymbol{F}\left(\tilde{\boldsymbol{\xi}}_1\right)\boldsymbol{d}\right|,$$

para $\boldsymbol{F}\left(\tilde{\boldsymbol{\xi}}_1\right) = \Delta^T\left(\ddot{\mathcal{F}}_{\boldsymbol{\alpha}}^{-1} - \boldsymbol{B}_{22}\right)\Delta$, em que

$$\boldsymbol{B}_{22} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddot{\mathcal{F}}_{\boldsymbol{\alpha}_{22}}^{-1} \end{pmatrix},$$

com $\ddot{\mathcal{F}}_{\boldsymbol{\alpha}_{22}}^{-1}$ sendo a inversa da matriz $\ddot{\mathcal{F}}_{\boldsymbol{\alpha}_{22}}$ obtida a partir da matriz particionada

$$\ddot{\mathcal{F}}_{\boldsymbol{\alpha}} = \begin{pmatrix} \ddot{\mathcal{F}}_{\boldsymbol{\alpha}_{11}} & \ddot{\mathcal{F}}_{\boldsymbol{\alpha}_{12}} \\ \ddot{\mathcal{F}}_{\boldsymbol{\alpha}_{21}} & \ddot{\mathcal{F}}_{\boldsymbol{\alpha}_{22}} \end{pmatrix}.$$

Assim sendo, considerando a máxima curvatura $\boldsymbol{d}_{max}$ obtida pelo maior autovetor associado ao maior autovalor da matriz $\boldsymbol{F}\left(\tilde{\boldsymbol{\xi}}_1\right)$, pode-se utilizar a direção do *i*-ésimo indivíduo tal como definido em (3.8) para revelar observações que exercem forte influência nas estimativas nos valores ajustados de $\tilde{\boldsymbol{\xi}}_1$.

4 APLICAÇÕES

Neste capítulo, por intermédio de estudos a alguns bancos de dados reais, busca-se colocar em prática a teoria desenvolvida ao longo desta pesquisa. Para tanto, foi usado o pacote MGCV implementado no R por Simon Wood. Sua documentação pode ser encontrada em **https://cran.r-project.org/web/packages/mgcv/mgcv.pdf**. Para mais detalhes sobre a teoria que envolve os modelos parcialmente lineares aditivos generalizados e sua base computacional, sugere-se consultar o livro de Wood (2017).

4.1 ELEIÇÕES DE 2006 AO SENADO BRASILEIRO

Como primeira aplicação, apresentamos uma base de dados contendo informações de candidatos ao Senado do Brasil nas eleições de 2006. No referido banco de dados, encontram-se informações tais como: O Estado do candidato (UF), o partido, quantidade de votos recebido, o sexo, o resultado da eleição (eleito ou não eleito), dentre outras. No entanto, foram consideradas para análise apenas as seguintes variáveis:

1. log.votos: O logaritmo dos votos. Variável resposta;
2. log.vgastos: O logaritmo dos valores gastos por cada candidato;
3. escola: Esta variável representa o nível de escolaridade e foi recodificada de modo a reduzir o número de categorias. Sendo assim, foi criada a variável instr (*nível de instrução*), a qual encontra-se divida em duas categorias: 0 = Baixo e 1 = Alto. Na categoria "Baixo"estão desde pessoas alfabetizadas até pessoas com ensino médio completo. Na categoria "Alto"estão pessoas com ensino superior incompleto até ensino superior completo;
4. idade: Corresponde a idade do candidato;
5. est.civil: Estado civil: 0 = casado, 1 = divorciado, 2 = separado, 3 = solteiro e 4=viúvo;
6. sexo: 0 = masculino e 1 = feminino.

4.1.1 Análise exploratória dos dados

Este conjunto de dados reúne 202 observações referentes aos candidatos ao Senado brasileiro no ano de 2006. Dentre os partidos como o maior número de candidatos destaca-se o PSOL com 17, PFL com 15, PDT com 14, PSDB e PSTU ambos com 13, PCB e PMDB 12 e PT com 10. Os candidatos dessa edição possuíam uma média de idade de 52.25 anos, sendo a menor delas de 35 e a maior de 89 anos. Destaca-se também que, do total dos candidatos, 30

eram do sexo feminino e 117 do sexo masculino. A maioria casados ou, pelo menos, já foram casados, bem como a maioria possuindo no mínimo ensino superior incompleto. Este é perfil geral dos candidatos. Contudo, uma análise mais aprofundada dessas variáveis será apesentada na sequência.

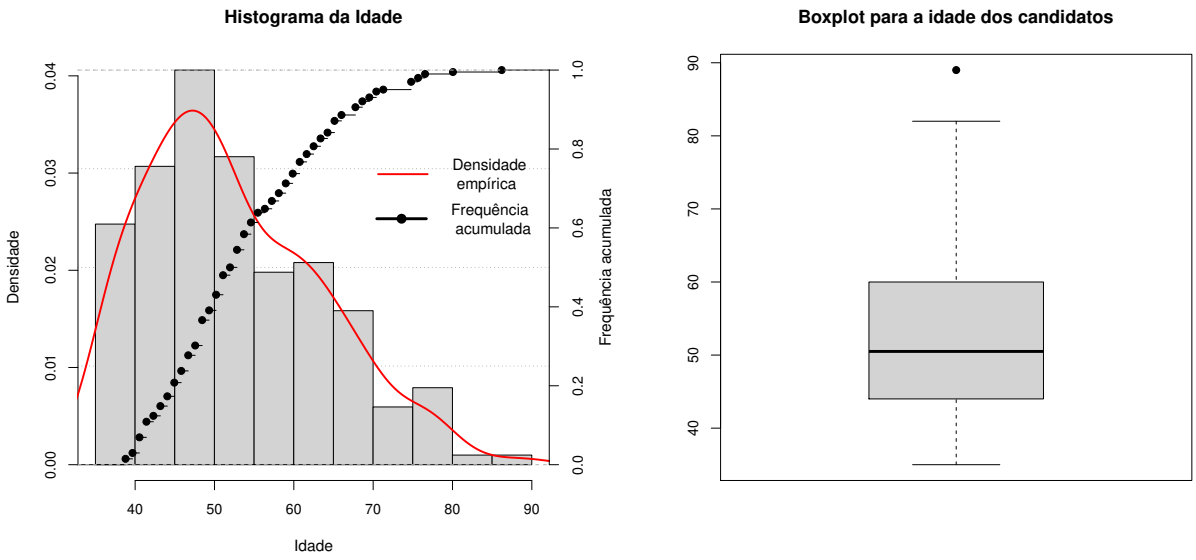


Figura 12 – Gráficos da distribuição de frequência, frequência acumulada e boxplot respectivos da a variável idade para o exemplo das eleições de 2006 ao Senado brasileiro.

Por meio do gráfico da distribuição, bem como pelo boxplot da idade dos candidatos apresentados na **Figura** 12, pode-se observar que esta variável possui comportamento assimétrico na direção das maiores idades. Além disso, o gráfico da frequência acumulada mostra outros quantis para a distribuição desta variável. Por meio dele pode-se observar que aproximadamente 93% dos candidatos possuem idades menores que 70 anos.

O gráfico das porcentagens para cada faixa de idade dos candidatos apresentado na **Figura** 13, em essência, possui o mesmo significados dos gráficos anteriores. Porém, em conjunto com o gráfico que relaciona os boxplot de cada faixa de idade com suas respectivas quantidades de votos recebidas, também apresentado na **Figura** 13, traz uma informação importante: mesmo possuindo a menor proporção, os candidatos acima de 75 anos conseguiram obter uma expressiva quantidade de votos, com a mediana maior que a dos candidatos mais jovens.

Por outro lado, como informação complementar, o gráfico da **Figura** 14a mostra que estes candidatos também foram os que mais gastaram em suas campanhas. Além disso, como os candidatos mais jovens gastaram menos em suas campanhas, há indícios de que possivelmente os

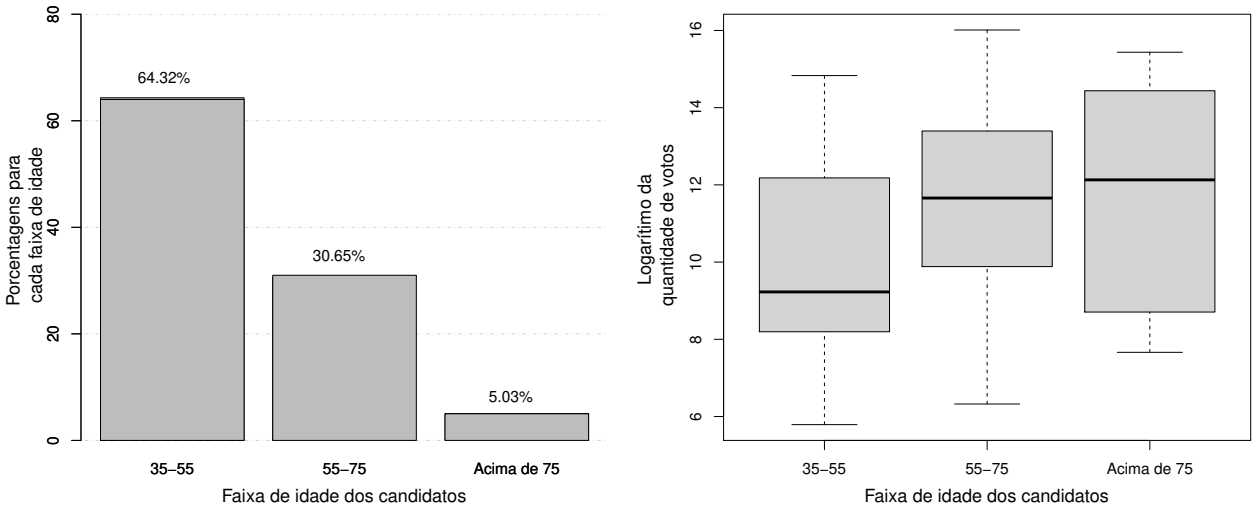
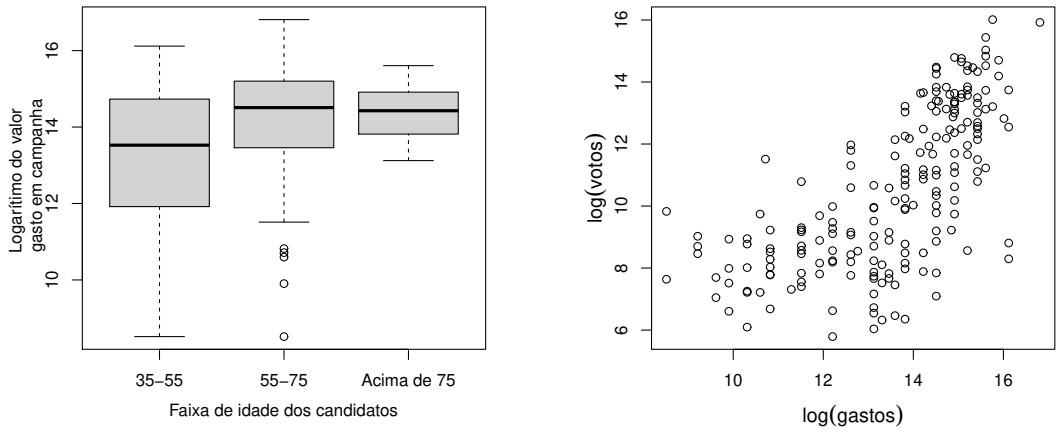


Figura 13 – Gráficos com as porcentagens para cada faixa de idade e relação da idade com a quantidade de votos recebidas pelos candidatos para o exemplo das eleições de 2006 ao Senado brasileiro.

gastos em campanha possa ter refletido na quantidade de votos recebidas. Neste caso, o gráfico de dispersão do logaritmo dos votos contra o logaritmo dos valores gastos apresentado na **Figura** 14b pode dar uma noção intuitiva dessa relação.



(a) **(b)**

Figura 14 – Estes gráficos mostram a relação da idade com os gastos e diagrama de dispersão: votos segundo os gastos na campanha dos candidatos para o exemplo das eleições de 2006 ao Senado brasileiro.

Portanto, é possível perceber uma maior concentração dos pontos nas interseções entre baixo gasto e baixo número de votos e entre alto gasto e alta votação. Ou seja, parece haver uma relação de causa e efeito entre essas variáveis, indicando que gastos maiores nas eleições

levam quantidades maiores de votos. Porém, é necessário uma análise de regressão a fim de identificar a correlação entre essas variáveis e sua significância estatística.

Tomando agora a variável sexo dos candidatos, nota-se que, embora representem apenas 14.85% da amostra, o sexo feminino consegue obter mediana de votos superior aos do sexo masculino. Porém, como evidenciado na **Tabela** 1, a maioria dos candidatos eleitos são do sexo masculino. Neste caso, é natural o questionamento sobre o fator sexo como potencialmente influente nos resultados.

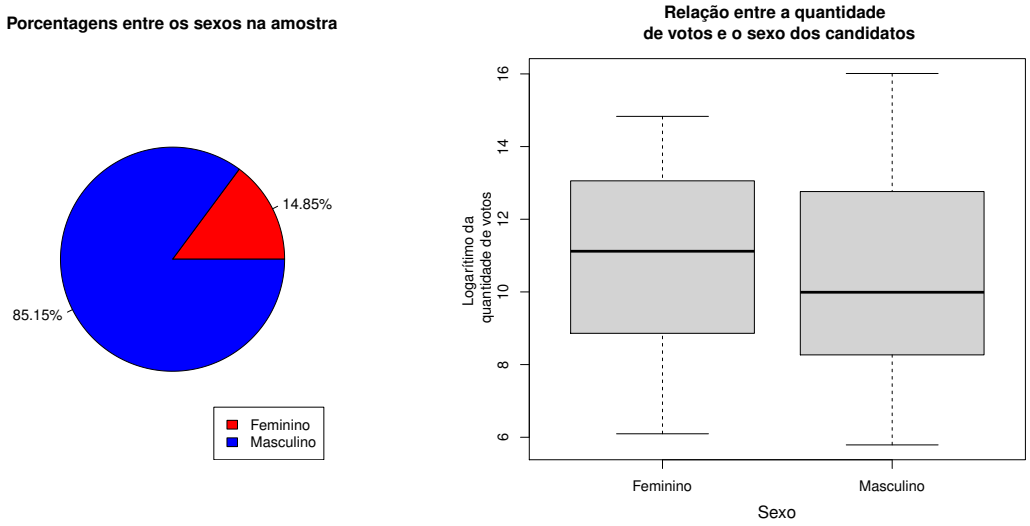


Figura 15 – Gráficos com as porcentagens dos sexo na amostra e relação da sexo dos candidatos com as quantidades de votos por eles recebidas para o exemplo das eleições de 2006 ao Senado brasileiro.

Sexo	Resultado		Total
	Eleitos	Não eleitos	
Feminino	4	26	30
Masculino	23	149	172
Total	27	175	202

Tabela 1 – Resultado das eleições de 2006 ao Senado brasileiro.

Estas constatações, no entanto, carecem de mais evidências. Considerando a normalidade dos dados, por exemplo, poderia ser aplicado um teste de hipóteses para confirmar se as médias dos votos podem ser consideradas estatisticamente diferentes em relação ao sexo ou se essa diferença é apenas resultado do processo aleatório que envolveu a obtenção dessa amostra em particular. Além disso, um teste qui-quadrado também poderia ser útil para verificar se o resultado das eleições depende do sexo do candidato.

Como o número de categorias do estado civil é relativamente grande, optou-se por reduzi-la, a fim de obter uma ideia mais clara da interação dessa variável com as demais. Neste caso, dividiu-se essa variável em duas categorias: quem era casado e que não era. Neste caso, quando a variável estado civil é relacionada com o logaritmo dos votos, percebe-se que: a proporção de casados e solteiros na amostra não difere muito, bem como os casados possuem mediana de votos superior em relação aos não casados.

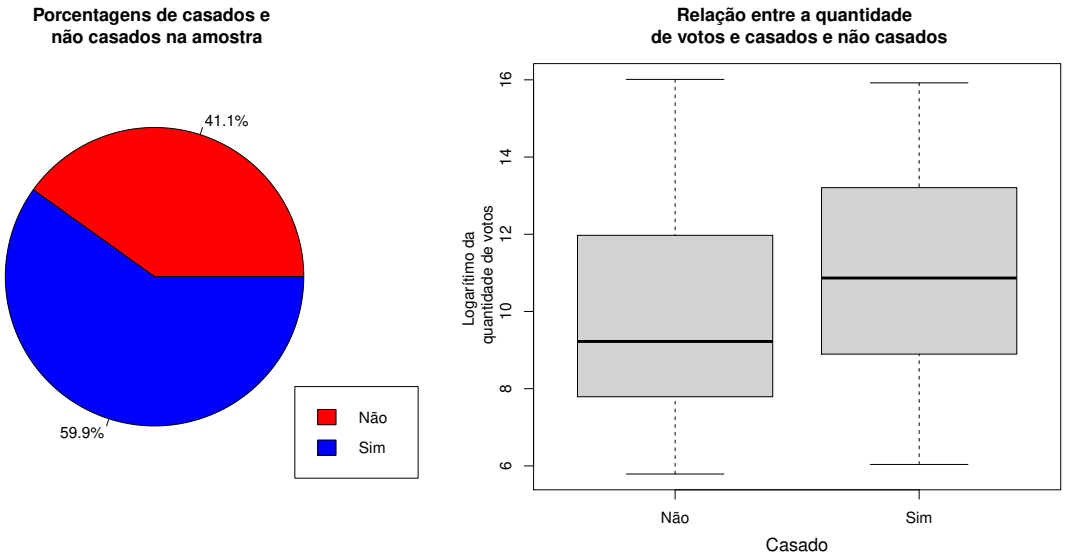


Figura 16 – Gráficos com as porcentagens dos sexo na amostra e relação da sexo dos candidatos com as quantidades de votos por eles recebidas para o exemplo das eleições de 2006 ao Senado brasileiro.

Além disso, apresentamos a **Tabela** 2 de dupla entrada com as frequências absolutas e o gráfico da **Figura** 17 que mostram quantos pontos do conjunto de dados se encaixam em cada categoria. Por meio da **Figura** 17 é possível visualizar o relacionamento entre essas variáveis graficamente. Em particular, verifica-se que a maioria dos candidatos do sexo masculino são casados e o oposto ocorre com sexo feminino.

Casado	Sexo		Total
	Feminino	Masculino	
Não	18	63	81
Sim	12	109	121
Total	30	172	202

Tabela 2 – Resultado das eleições de 2006 ao Senado brasileiro segundo o estado civil.

Assim como realizado para a variável sexo, é interessante investigar se o fato de um candidato ser casado ou não pode ter influência nos resultados. Para isso, usamos o recurso

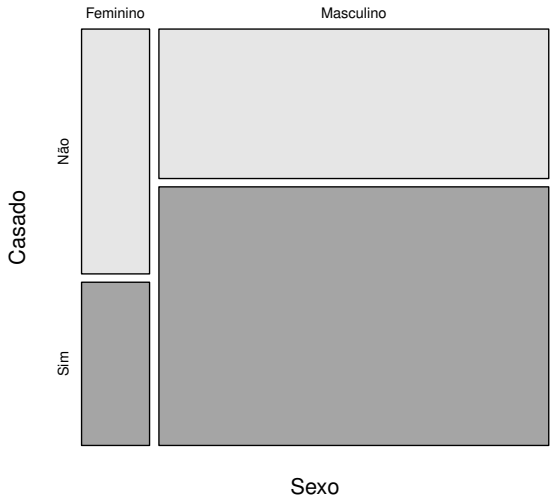


Figura 17 – Gráfico com a relação do sexo com o estado civil dos candidatos para o exemplo das eleições de 2006 ao Senado brasileiro.

gráfico apresentado na **Figura** 18. Portanto, pode-se identificar que os candidatos casados foram maioria dos eleitos. Neste caso também é interessante testes para diferenças entre médias e teste qui-quadrado a fim de obter resultados mais precisos para estas considerações.

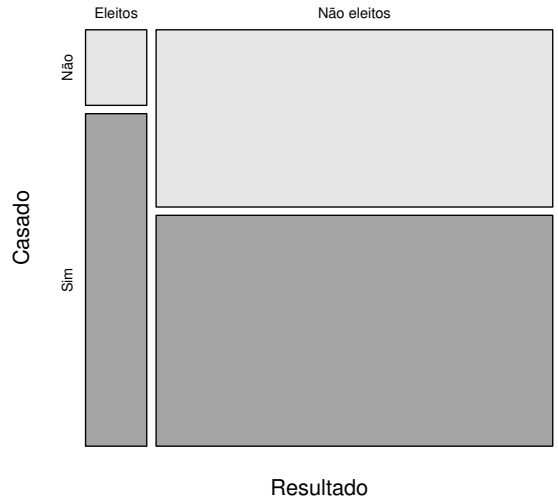


Figura 18 – Resultado das eleições de 2006 ao Senado brasileiro segundo o estado civil.

Por fim, analisando o nível de instrução e a relação com a quantidade de votos, nota-se que há uma maior concentração de votos para os candidatos que possuem ensino superior incompleto ou completo.

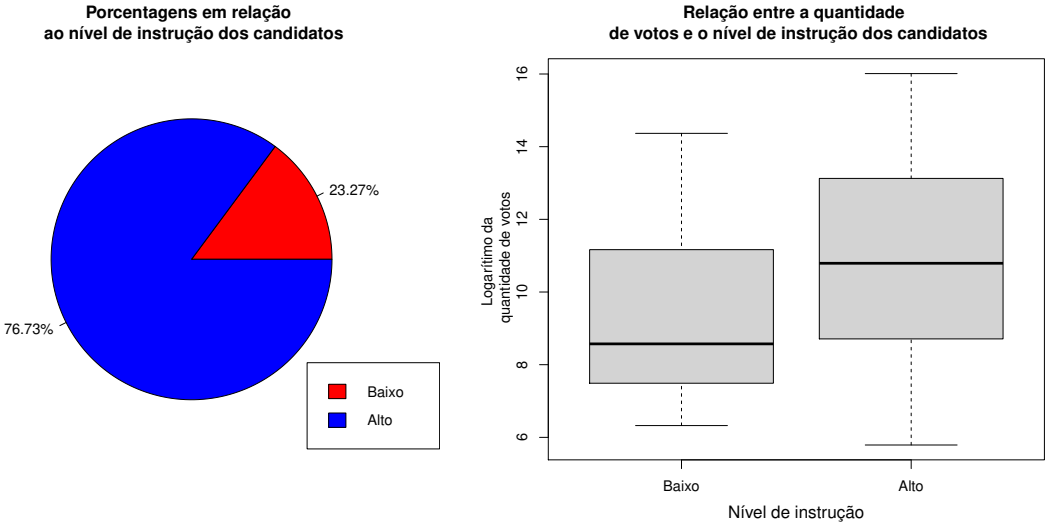


Figura 19 – Gráficos com as porcentagens dos níveis de instrução na amostra e relação das quantidades de votos por eles recebidas para o exemplo das eleições de 2006 ao Senado brasileiro

Por meio da **Tabela** 3, nota-se que, corroborando com o gráfico da **Figura** 19, há forte indício de que um nível de educação alto pode ter efeito sobre a quantidade de votos, bem como nos resultados das eleições.

Nível de Instrução	Resultado		Total
	Eleitos	Não eleitos	
Baixo	4	43	47
Alto	23	132	155
Total	27	175	202

Tabela 3 – Resultado das eleições de 2006 ao Senado brasileiro segundo o nível de instrução.

Note que as análises foram feitas de forma individual, sem que fosse considerado o efeito conjunto destas variáveis no valor médio da quantidade dos votos. O que queremos é uma forma de acomodar esta informação em um modelo capaz explicar a relação presente em cada uma das covariáveis, mostrando o impacto conjuntamente dessas variáveis na quantidade de votos. Para isso, iremos propor um modelo de regressão apropriado.

4.1.2 Ajuste do modelo

Para a análise desses dados, considerou-se um modelo de regressão com distribuição normal para a variável resposta. Sendo assim, definindo Y_i, $i = 1,2,\ldots,n$, como sendo a variável *logaritmo dos votos* para a i-ésima observação, então temos que $Y_i \sim \mathcal{N}(\mu_i, \phi = \sigma^2)$, de modo

que estabelecemos um MPLAG com preditor linear dado por:

$$\begin{aligned} g(\mu_i) = \mu_i = \eta_i & \\ &= \beta_0 + \beta_1 \text{est.civil}_1 + \beta_2 \text{est.civil}_2 + \beta_3 \text{est.civil}_3 + \beta_4 \text{est.civil}_4 \\ &+ \beta_5 \text{instr} + \beta_6 \text{sexo} + \beta_7 \text{idade} + f(t_i), \text{ para } i = 1,2,\ldots,n, \end{aligned} \tag{4.1}$$

com função de ligação *identidade*, $n = 202$, $f(\cdot)$ uma função suave para a variável log.vgastos e variável dependente dada pelo logaritmo do número de votos recebidos pelos candidatos ao Senado em 2006.

Assim, para o ajuste deste modelo, foram empregados um B-*spline* cúbico com 10 nós equidistantes e penalização quadrática. Os graus de liberdade efetivos são obtidos pela soma dos graus de liberdade referente a função não paramétrica mais oito, referindo-se ao intercepto e os β's de cada variável ajustada parametricamente. Os resultados desse ajuste estão resumidos na **Tabela** 4.

Tabela 4 – Resumo das estimativas referentes ao modelo ajustado sob distribuição normal para o exemplo das eleições de 2006 ao Senado brasileiro.

Efeito	**Estimativa**	**Erro-Padrão**	**Valor-p**
Intercepto	8.738	0.655	< 0.0001
Idade	0.032	0.011	< 0.005
instrAlto	1.139	0.271	< 0.0001
est.civil_1	-0.297	0.335	0.376
est.civil_2	-0.353	0.421	0.403
est.civil_3	-0.336	0.360	0.352
est.civil_4	-0.875	0.536	0.104
Masculino	-0.726	0.325	< 0.01
Outras medidas do ajuste			
$df(\alpha)$	12.66		
α	3.497		
ϕ	2.399		
VCG	2.571		

Nesse modelo, observa-se que todas as categorias da variável estado civil têm uma alta probabilidade de estar apenas por acaso correlacionadas com a variável dependente, pois apresentam valor $p > 0.05$. Isto é, na possibilidade de poder repetir o processo de amostragem indefinidas vezes, supondo que a variável estado civil não tenha efeito sobre o resultado das eleições, a probabilidade de se obter uma amostra que indique que há uma correlação entre essas variáveis é elevada.

Tendo em vista que em um modelo de regressão somente devem ser adicionadas variáveis cuja correlação com a variável dependente é elevada e, principalmente, teoricamente

explicável, então esta variável deve ser retirada da análise. Assim, mantendo apenas aquelas com maior contribuição. Portanto, realizou-se um ajuste sem essa variável e o resultado foi que a variável sexo também deixou de ser significativa. Com isso, foi obtido o modelo final, o qual pode ser observado na **Tabela** 5. Graficamente, o subsequente ajuste da variável suavizada pode ser visualizado pela **Figura** 20.

Tabela 5 – Resumo das estimativas referentes ao modelo final ajustado sob distribuição normal para o exemplo das eleições de 2006 ao Senado brasileiro.

Efeito	**Estimativa**	**Erro-Padrão**	**Valor-p**
Intercepto	8.148	0.574	< 0.0001
Idade	0.028	0.010	< 0.001
instrAlto	1.175	0.271	< 0.0001
Outras medidas do ajuste			
$df(\alpha)$	8.69		
α	3.082		
ϕ	2.425		
VCG	2.534		

Nota-se que a curva se adéqua muito bem a tendência dos dados. Assim, podemos identificar melhor a relação entre as variáveis quantidade de votos e valor gasto em campanha, bem como podemos observar que há uma maior variabilidade do número de votos em regiões de baixo e elevados gastos, destacado pelos intervalos de confiança pontuais.

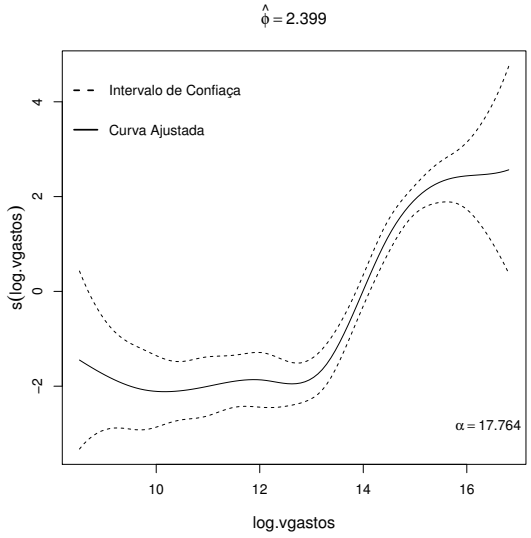


Figura 20 – Gráfico da curva ajustada sob distribuição normal para o exemplo das eleições de 2006 ao Senado brasileiro.

Na sequência, dar-se-á continuidade ao estudo desse modelo mediante o seu diagnóstico, colocando em prática a teoria desenvolvida ao longo do texto para esta finalidade, começando

pela análise de resíduos, seguida pela alavancagem e finalizando pela análise de influência local.

4.1.3 Diagnóstico

Para avaliar a adequacidade do ajuste para o modelo proposto, inicialmente deve-se observar os gráficos da **Figura** 21, em que podemos analisar os resíduos. Por meio destes gráficos, pode-se tirar conclusões a respeito do componente estocástico do modelo. Assim, qualquer comportamento atípico referente aos resíduos pode ser um indicativo de que o modelo proposto não é adequado para representar o fenômeno em estudo.

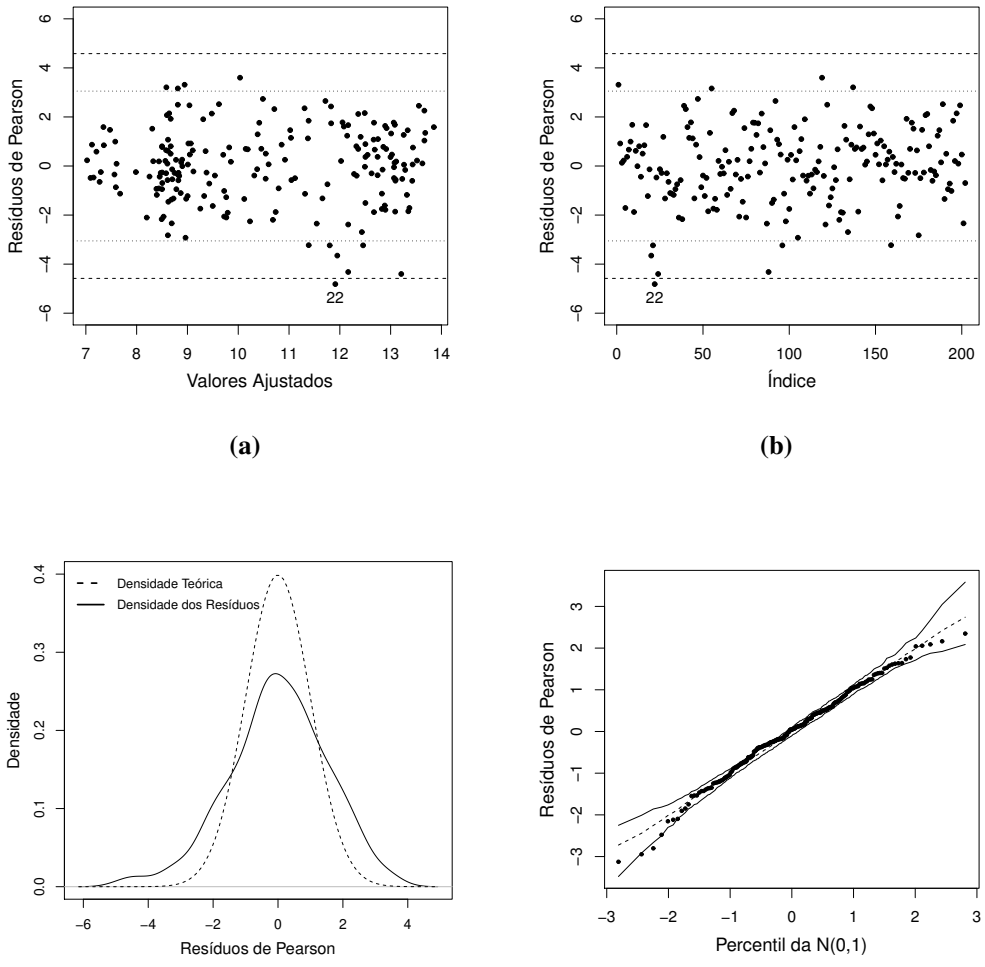


Figura 21 – Gráficos de resíduos referente ao modelo ajustado sob distribuição normal para o exemplo das eleições de 2006 ao Senado brasileiro.

Desta forma, o gráfico dos resíduos versus valores ajustados (**Figura** 21a) permite verificar a homoscedasticidade do modelo, isto é, σ^2 constante, além de possibilitar identificar

outliers em y. Assim, espera-se que este gráfico apresente os pontos distribuídos aleatoriamente em torno de zero. Se os pontos possuem alguma tendência, há indícios de que a variância dos dados não é constante, ou seja, existe indicativo de heteroscedasticidade. Já o gráfico dos resíduos versus a ordem de coleta dos dados (**Figura** 21b) avalia a hipótese de independência dos dados. Portanto, se os pontos tiverem um comportamento que se repete em determinado ponto do gráfico, temos indícios de dependência dos resíduos. Por fim, os gráficos 21c e 21d fornecem informações sobre a distribuição dos resíduos. Neste caso, espera-se um comportamento próximo a de uma distribuição normal com média zero e desvio padrão igual a um.

Portanto, é possível observar que o gráfico de resíduos quantílicos contra os valores ajustados de fato apresentam comportamento aleatório em torno de zero, indicando uma variabilidade constante dos resíduos, sendo que alguns dos pontos estão dois desvios padrão distantes da média, mas não muito distantes de três desvios padrão (linhas tracejadas), podendo representar outliers. Algumas explicações possíveis para a obtenção desses pontos, seriam erros sistemáticos, ou erro de especificação para o modelo, ou que estes pontos estão no extremo de validade para a distribuição da variável resposta. O mesmo ocorre para os resíduos versus índices, os quais não apresentam comportamento serial aparente. O único resíduo destacado foi o correspondente a observação #22 como um possível valor aberrante, porém apresenta uma leve discrepância. Além disso, os gráficos inferiores da **Figura** 21 corroboram com a hipótese de normalidade dos dados, embora a densidade dos resíduos possua uma variância um pouco maior em relação a densidade teórica esperada. Com base nessa análise de resíduos, verifica-se que as suposições para o modelo estão razoavelmente satisfeitas.

Segundo o gráfico de alavancagem, destacado pela **Figura** 22a, não há pontos considerados com alta alavancagem. Ainda tomando o modelo definido anteriormente, considerando a análise de influência local sob os esquemas de perturbação de casos, perturbação na variável resposta e perturbação em uma das variáveis explicativas, tem-se que para os pontos destacados pelas **Figura** 22b e 22c são considerado pontos potencialmente influentes, que podem estar exercendo um peso desproporcional nas estimativas do modelo.

Desse modo, é necessário avaliar como as estimativas dos parâmetros do modelo estão sendo afetadas por essas observações, tentando verificar se existe interferências substanciais, bem como mudanças inferenciais. Neste caso, deve-se ter uma atenção especial sobre estas observações, investigando quais são suas características e quais as possíveis respostas para entender o porquê de estarem influenciando de forma tão crítica nas estimativas. Assim, deve-se retirar individualmente cada uma dessas observações, realizar um novo ajuste e saber qual o

impacto da retirada dessa observação potencialmente influente nas novas estimativas.

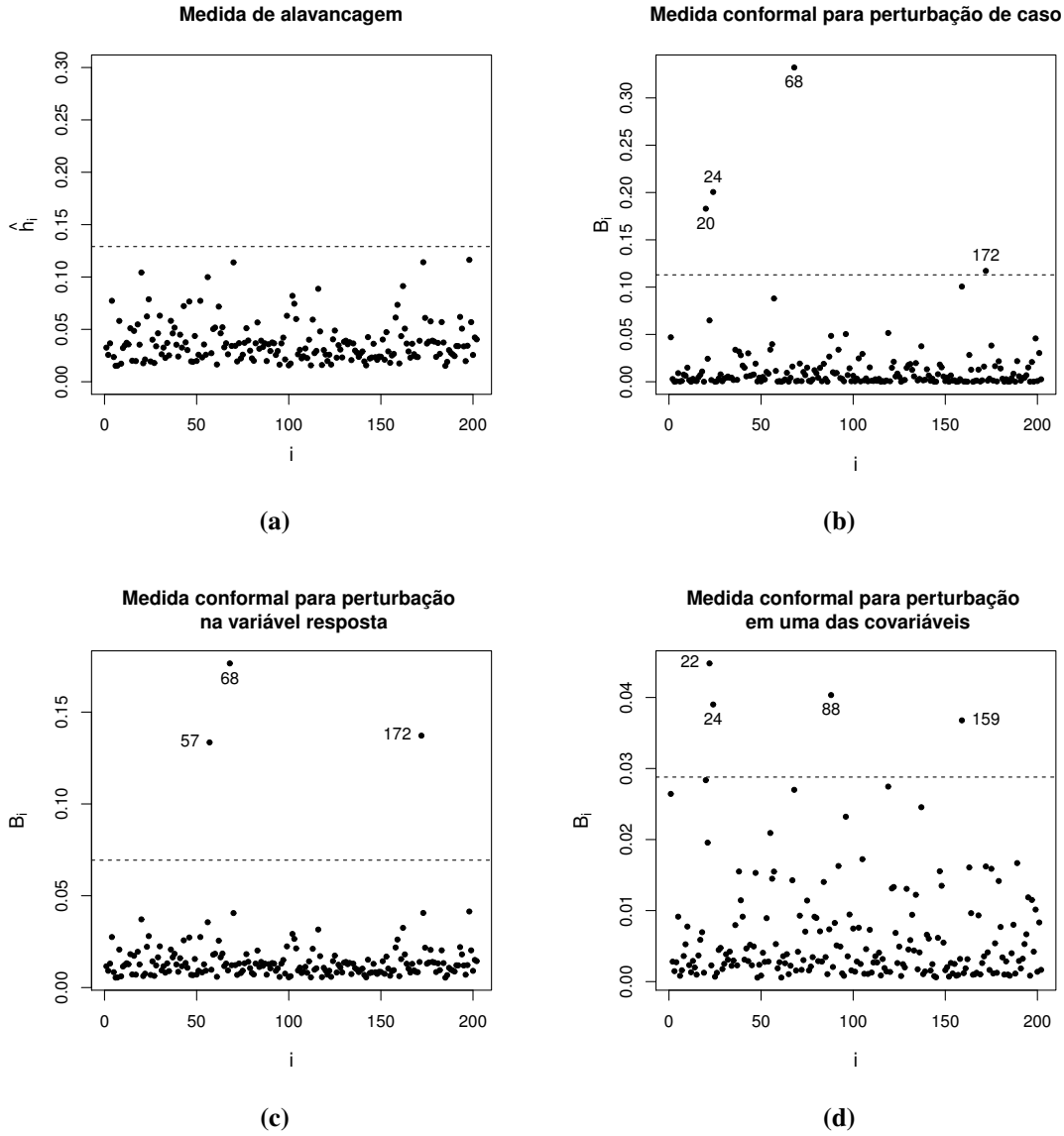


Figura 22 – Gráficos de alavancagem e influência local sob os esquemas de perturbação de caso, perturbação na variável resposta e perturbação em umas das variáveis explicativas, respectivamente, referentes ao ajuste do modelo sob distribuição normal para o exemplo das eleições de 2006 ao Senado brasileiro.

Nota-se, de acordo com os gráficos, que as observações potencialmente influentes foram: #20, #22, #24, #57,#68, #88, #159 e #172. A fim de analisar o impacto das observações destacadas sob as estimativas dos parâmetros, realizamos uma análise de sensibilidade destas observações. Como já destacado, esta análise confirmatória consiste em reajustar o modelo, eliminando individualmente cada observação considerada possivelmente influente e, em seguida,

calcular as taxas de variação das estimativas dos parâmetros segundo a expressão:

$$\text{Taxa}(\theta) = \left| \frac{\tilde{\theta}^{(i)} - \hat{\theta}}{\hat{\theta}} \right| \times 100, \tag{4.2}$$

em que $\hat{\theta}$ e $\tilde{\theta}^{(i)}$ são as EMV no modelo completo e no modelo sem a i-ésima observação respectivamente. Com isso, obteve-se a tabela a **Tabela** 6, em que temos as observações retiradas seguidas dos valores das estimativas e as respectivas mudanças relativas em porcentagem.

Tabela 6 – Estimativas de máxima verossimilhança referente ao ajuste do modelo sob distribuição normal e mudança relativa em porcentagem entre parênteses para os parâmetros β_0, β_1 e β_2 após a retirada dos pontos para o exemplo das eleições de 2006 ao senado brasileiro.

Observação retirada	Estimativas dos parâmetros		
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
#20	8.30 (1.92%)	0.03 (4.55%)	1.07 (9.20%)
#22	8.19 (0.56%)	0.03 (4.16%)	1.22 (3.70%)
#24	8.23 (1.00%)	0.03 (5.74%)	1.19 (1.40%)
#57	8.19 (0.48%)	0.03 (3.03%)	1.20 (2.19%)
#68	8.18 (0.38%)	0.03 (4.01%)	1.18 (0.46%)
#88	8.10 (0.62%)	0.03 (2.83%)	1.21 (2.75%)
#159	7.87 (3.41%)	0.03 (19.48%)	1.19 (1.17%)
#172	8.18 (0.40%)	0.03 (4.92%)	1.23 (4.32%)
#68 e #172	8.21 (0.80%)	0.03 (9.04%)	1.23 (4.74%)

O valor de referência para comparar a mudança relativa das estimativas após a retirada de um ponto potencialmente influente é $(p/n) \times 100$, em que p é a quantidade de parâmetros e n o tamanho da amostra. Sendo assim, espera-se que na retirada de uma observação a mudança relativa seja de aproximadamente 0.01%. Então, baseado na **Tabela** 6, temos que as observações causam impactos desproporcionais principalmente nas estimativas de β_1 e β_2. Porém, não houve mudança inferencial com a retirada destas observações.

Por fim, deve-se investigar os motivos pelos quais as observações apresentadas na **Tabela** 6 se destacam das demais. Assim, por exemplo, para a observação #20, tem-se que é um candidato do sexo masculino, possui 48 anos e ensino médio completo (nível de instrução

baixa). Além disso, teve o quarto maior gasto na campanha, porém tendo obtido apenas 4013 (ou $\log(4013) = 8.3$) votos, ficando entre os 50 candidatos menos votados. Neste caso, há um contraste em relação a tendência apresentada no modelo.

Já para a observação #22, também acontece algo semelhante, pois tem-se que é um candidato do sexo masculino, com 48 anos e nível superior completo (nível de instrução alto). Além disso, teve gastos elevado na campanha (logaritmo igual 14.50), porém ficando também entre os 15 candidatos com menores votos recebidos.

Tomando agora a observação #159, tem-se que é um candidato do sexo feminino, possui 89 anos (a maior do banco de dados), com nível superior completo (nível de instrução alto). Além disso, teve gastos elevado na campanha (logaritmo igual 13.82), porém ficando também entre os 50 candidatos com menores votos recebidos.

4.2 ALUGUEL EM MUNIQUE

Neste segundo exemplo será utilizado o banco de dados disponível no R denominado *rent*. Tal banco de dados é sobre uma pesquisa realizada em abril de 1993, com o objetivo de estudar o valor do aluguel em Munique. Para este exemplo, serão consideradas as seguintes variáveis:

1. R: Valor mensal do Aluguel (em euros). Variável resposta;
2. Fl: Tamanho do apartamento (em metros quadrados);
3. A: Ano de construção (período);
4. loc: Qualidade da Localização (1: Média, 2: Boa e 3: Excelente);
5. B: Qualidade do banheiro (0: Padrão e 1: Premium) e
6. L: Qualidade da Cozinha (0: Padrão e 1: Premium).

Antes de propor um modelo para explicar a relação das variáveis explicativas com a média do aluguel mensal em Munique, será apresentada uma análise descritivas dos dados. O objetivo é verificar e resumir as principais características dos dados para obter um entendimento básico sobre as relações existentes entre as variáveis analisadas.

4.2.1 Análise exploratória dos dados

Inicialmente, será descrito o comportamento de cada variável segundo medidas resumo, tais como: média, desvio padrão, mediana, etc., afim de obter uma ideia básica sobre o comportamento dos dados. Além disso, para melhor entendimento, serão apresentados métodos

visuais, tais como gráficos, bem como comentários sobre aspectos particulares das análises.

Para este conjunto de dados, a variável *valor mensal do aluguel* possui as seguintes características: o valor médio do aluguel mensal foi de $\bar{x} = 804.4\,€$, com desvio padrão igual a $s = 364.708\,€$ e, consequentemente, coeficiente de variação dado por $cv = 45.34\%$. Além disso, os valores do menor e maior aluguel foram de $x_{(1)} = 167.2\,€$ e $x_{(n)} = 2869.9\,€$, respectivamente, bem como os primeiro e terceiros quartis dados por $Q_1 = 536.5\,€$ e $Q_3 = 1015.8\,€$ com mediana igual a $\text{Md} = 716.5\,€$. A **Figura** 23 apresentam um resumo gráfico dessas características.

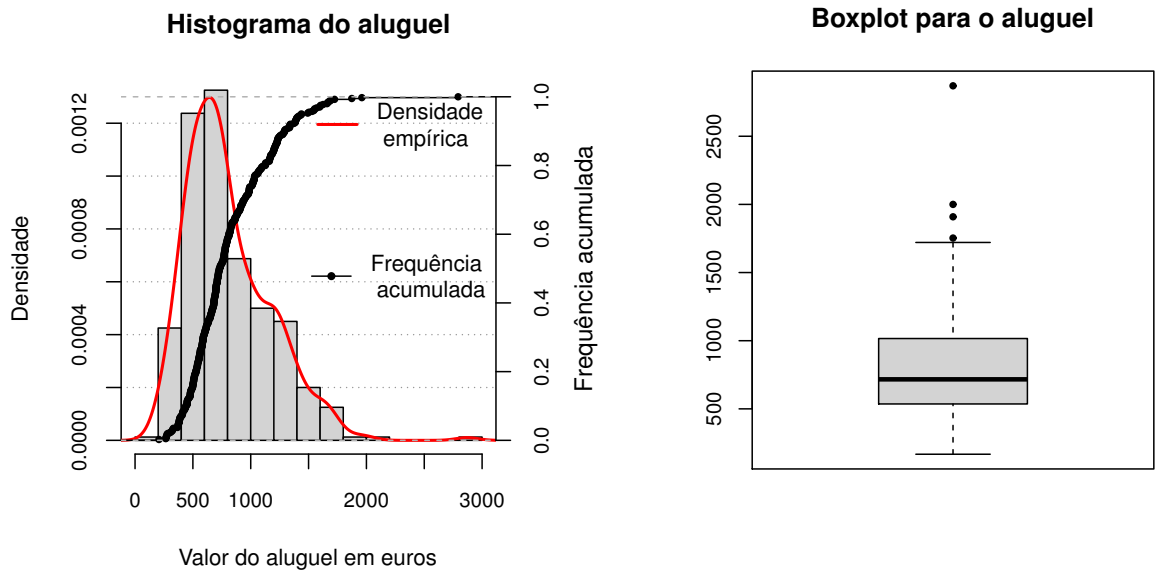


Figura 23 – Gráficos da distribuição de frequência, frequência acumulada e boxplot respectivos da a variável preço do aluguel para o exemplo do aluguel em Munique.

Pode-se observar que a distribuição do valor mensal do aluguel apresenta uma assimetria à direita, bem como quatro pontos que se destacam das demais observações. Os pontos destacados são: #299, #236, #312, e #26, que correspondem respectivamente aos seguintes valores dos aluguéis: 1753.2; 1909.0; 2000.0 e 2869.9. A esses aluguéis estão associadas as seguintes variáveis: para a observação #299, correspondente ao aluguel 1753.2 – apartamento com 83 metros quadrados, para o ano de 1957, com excelente localização, além de banheiro padrão e cozinha premium; para a observação #236, correspondente ao aluguel 1909.2 – apartamento com 104 metros quadrados, para o ano de 1981, com boa localização, além de banheiro padrão e cozinha premium; para a observação #312, correspondente ao aluguel 2000.0 – apartamento com 114 metros quadrados, para o ano de 1985, com excelente localização,

além de banheiro e cozinha padrão; para a observação #26, correspondente ao aluguel 2869.9 – apartamento com 90 metros quadrados, para o ano de 1985, com excelente localização, além de banheiro e cozinha padrão.

Também foi observado que as variáveis categóricas possuem suas proporções na amostra como apresentadas nos gráficos da **Figura** 24. Desta forma, pode-se concluir que na amostra os tipos mais comuns de localidade, banheiro e cozinha são, respetivamente, boa e padrão. Além disso, segundo os boxplots apresentados na **Figura** 25, nota-se que os preços dos alugueis aumentam quando a qualidade da localidade e da cozinha melhoram.

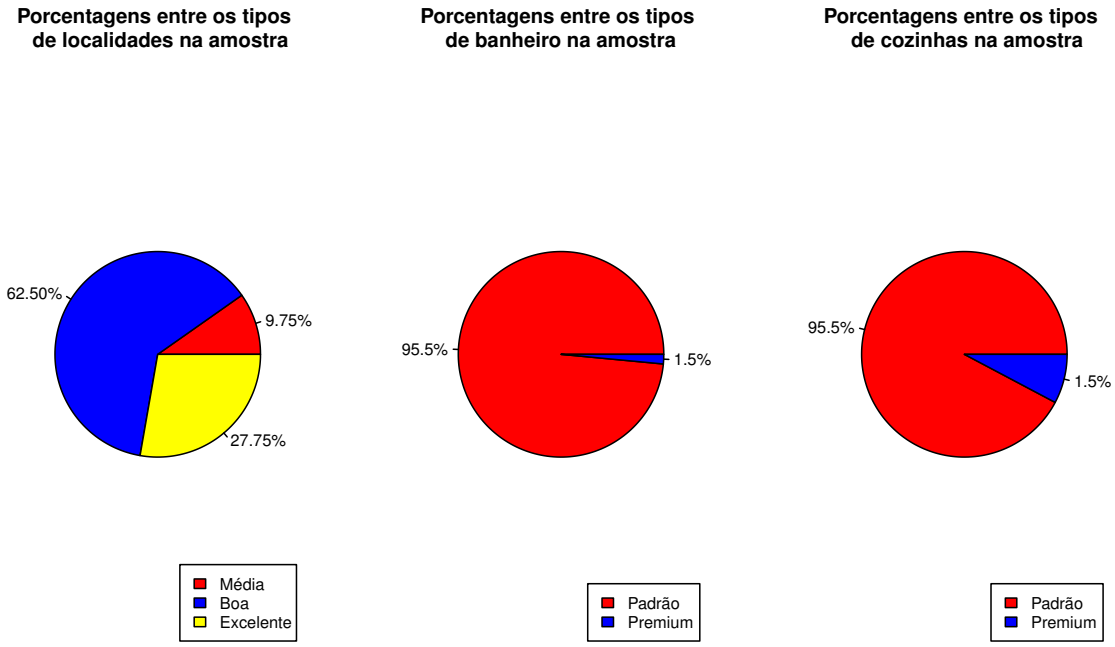


Figura 24 – Gráficos com as proporções de cada variável categórica na amostra para o exemplo do aluguel em Munique.

Os alugueis médios, segundo os tipos de localidade, foram de 595.40 € com desvio padrão igual a 169.97 € para o tipo de localidade média; 792.23 € com desvio padrão igual a 348.00 € para o tipo de localidade boa; 884.32 € com desvio padrão igual a 411.25 € para o tipo de localidade excelente. Consequentemente, os coeficientes de variação são: 28.54%, 43.92% e 46.5% respectivamente. Conclui-se, portanto, que a localidade com denominação média corresponde a categoria mais homogênea em relação as demais. De igual forma, temos que o aluguel médio segundo o tipo de banheiro foram de 811.13 € com desvio padrão igual a

363.25 € para o tipo padrão; 362.76 € com desvio padrão igual a 73.40 € para o tipo premium. Consequentemente, os coeficientes de variação são: 44.78% e 20.23%, respectivamente. Finalmente, temos que o aluguel médio segundo o tipo de cozinha foram de 778.29 € com desvio padrão igual a 349.65 € para o tipo padrão; 1085.47 € com desvio padrão igual a 408.66 € para o tipo premium. Consequentemente, os coeficientes de variação são: 44.92% e 50.80%, respectivamente.

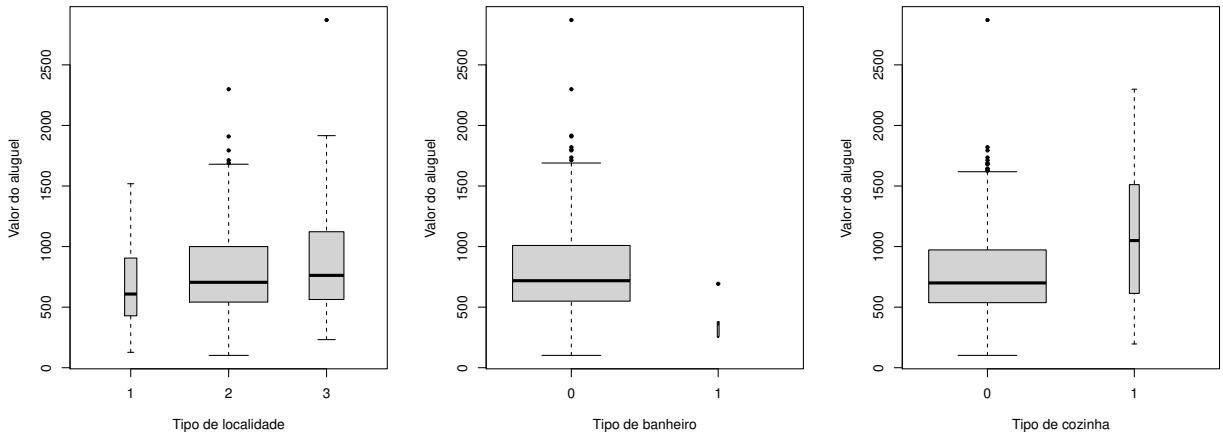


Figura 25 – Boxplots do valor do aluguel segundo as variáveis categóricas para o exemplo do aluguel em Munique.

Além disso, a variável *tamanho do apartamento* possui as seguintes características: tamanho médio de $\bar{x} = 65.98\text{m}^2$, com desvio padrão igual a $s = 20.50\text{m}^2$ e, consequentemente, coeficiente de variação dado por $cv = 31.08\%$. Os valores da menor e maior área foram de $x_{(1)} = 30.00\text{m}^2$ e $x_{(n)} = 120.00\text{m}^2$, respectivamente, bem como os primeiro e terceiros quartis dados por $Q_1 = 50.0\text{m}^2$ e $Q_3 = 80.0\text{m}^2$ com mediana igual a $\text{Md} = 65.50\text{m}^2$. Portanto, os tamanhos dos apartamentos são aproximadamente homogêneos. Por fim, apresentamos os gráficos de dispersão, **Figura** 26 e medidas da correlação entre as variáveis numéricas, **Figura** 27. Esses gráficos buscam dar uma ideia visual da relação presente entre as variáveis.

Assim, nota-se que, por meio do gráfico de dispersão, parece que o valor do aluguel aumenta com o passar dos anos e com o tamanho do apartamento. Além disso, de acordo com o gráfico da **Figura** 27, há uma correlação moderada entre o valor do aluguel e as variáveis ano e área do apartamento. Logo, há indícios de possível relação de causalidade entre essas variáveis. Contudo, apenas análises estatísticas mais avançadas, tais como testes de hipóteses ou modelos de regressão, poderá dizer a natureza aproximada dessa relação.

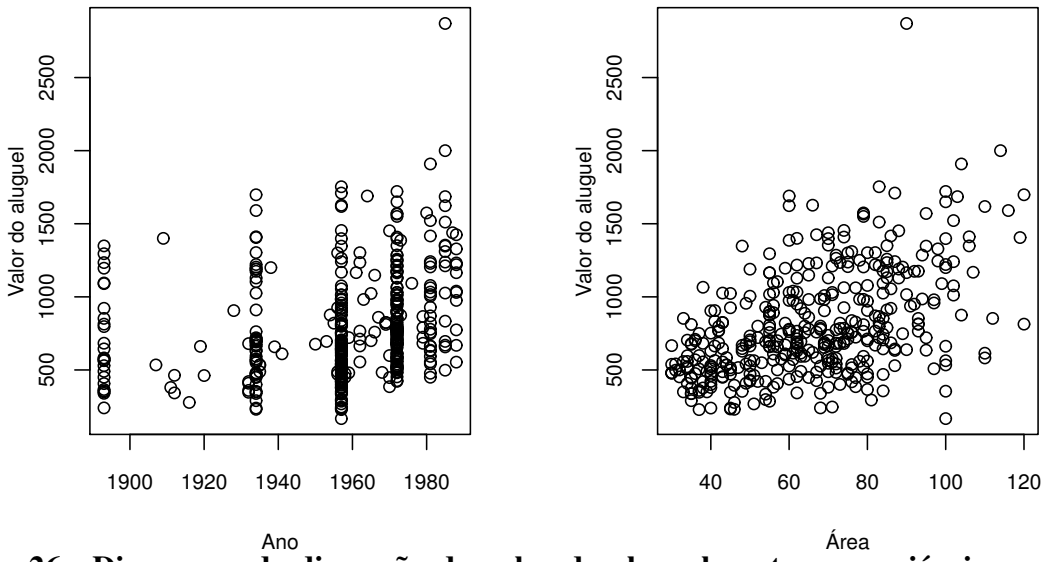


Figura 26 – Diagramas de dispersão do valor do aluguel contra as variáveis numéricas para o exemplo do aluguel em Munique.

Note que as análises foram feitas de forma individual. Isto é, observa-se cada variável de forma separada, desconsiderando o efeito que a presença de outras variáveis desempenham no valor médio do aluguel. Em outaras palavras, essas análises não levam em conta a relação da média do aluguel segundo a presença conjunta dessas variáveis. Isso só será possível por meio de um modelo de regressão apropriado, que mostre o impacto conjuntamente dessas variáveis no preço do aluguel em Munique.

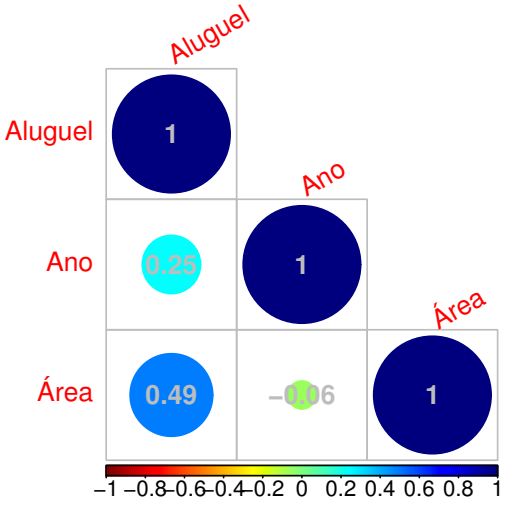


Figura 27 – Gráfico das correlações entre as variáveis quantitativas para o exemplo do aluguel em Munique.

Diante de todo o exposto, o objetivo daqui em diante passa a ser encontrar um modelo de regressão apropriado de modo a acomodar o efeito simultâneo dessas variáveis. Por fim, realizar a análise de diagnóstico, a fim de obter indícios de afastamento das suposições para o modelo, bem como fazer inferências a respeito dos parâmetros sob o modelo postulado.

4.2.2 Ajuste do modelo

Visto que a variável resposta, *valor do aluguel* (R), é contínua e assume valores nos reais positivos, então faz sentido assumir uma distribuição gama como parte estocástica do modelo. Sendo assim, definindo Y_i, $i = 1,2,\ldots,n$ como sendo o *valor do aluguel* para a i-ésima observação, então temos que $Y_i \sim \ G(\mu_i,\phi)$, de modo que a parte sistemática será dada por:

$$\begin{aligned} g(\mu_i) = \mu_i^{-1} &= \eta_i \\ &= \beta_1 + \beta_2 loc_{2i} + \beta_3 loc_{3i} + \beta_4 L_i + \beta_5 B_i + f_1(A) + f_2(Fl), \end{aligned}$$

em que $\mu_i = \mathbb{E}(\mathrm{R_i})$, f_k, com $k = 1,2$ são funções suaves das covariáveis *ano de construção* (A) e *tamanho do apartamento* (Fl), além de uma amostra de tamnho $n = 400$. Tem-se também que a função de ligação adotada será a *inversa* e os resultados do ajuste podem ser conferidos na **Tabela** 7 com seus respectivos erros padrão e valor p. Assim, destaca-se que as variáveis

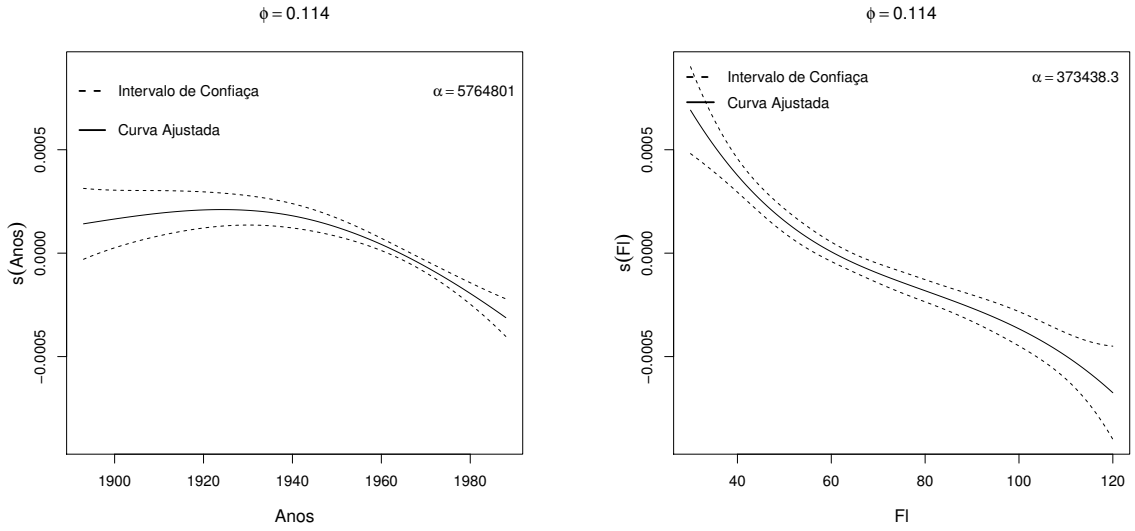
Tabela 7 – Resumo das estimativas de máxima verosimilhança referentes ao modelo ajustado sob distribuição gama para o exemplo do aluguel em Munique.

Efeito	**Estimativa**	**Erro-Padrão**	**Valor-p**
Intercepto	0.0017	0.00010	< 0.0001
loc2	-0.00034	0.00011	< 0.005
loc3	-0.00048	0.00011	< 0.001
B1	0.0011	0.00038	< 0.001
L1	-0.00013	0.00005	< 0.05
Outras medidas do ajuste			
$df_1(\boldsymbol{\alpha})$	2.10		
$df_2(\boldsymbol{\alpha})$	3.08		
$\boldsymbol{\alpha}$	$(5764801.0, 373438.3)$		
ϕ	0.114		
VCG	0.120		

"A"e "Fl"foram ambas ajustadas segundo um P-*spline* considerando um B-*spline* cúbico com 5 nós equidistantes para cada suavizador e diferenças de segunda ordem. Neste caso, observou-se que uma quantidade de nós acima de 5 aumentava o valor do critério VCG. Além disso, os

respectivos parâmetros de suavização foram obtidos por meio do método VCG. Observa-se que todas as variáveis ajustadas parametricamente foram significativas para o modelo.

Com relação aos graus de liberdade efetivos, nota-se que $df(\boldsymbol{\alpha}) = 5 + df_1(\boldsymbol{\alpha}) + df_2(\boldsymbol{\alpha})$, em que 5, por definição, corresponde aos graus de liberdade referente a parte paramétrica e $df_k(\boldsymbol{\alpha})$ correspondendo a parte não paramétrica do modelo. Neste caso, tem-se que $df_1(\boldsymbol{\alpha}) =$ 2.10 para a variável "A"e $df_2(\boldsymbol{\alpha}) = 3.08$ para a variável "Fl", em que, respectivamente, $\alpha_1 = 5{,}764{,}801$ e $\alpha_2 = 373{,}438.3$. Além disso, foi obtida uma pontuação VCG de 0.1203. As estimativas estão resumidas na **Tabela** 7. Na sequência, os gráficos do ajuste para as variáveis suavizadas são apresentados na **Figura** 28. Como a função de ligação utilizada foi a inversa,



(a) **(b)**

Figura 28 – Gráficos das curvas ajustadas sob distribuição gama para o exemplo do aluguel em Munique.

então significa que a relação entre as variáveis se dá de maneira inversa. Desta forma, como era de se esperar, a tendência é que os preços médios dos alugueis aumentem de acordo com o tamanho do imóvel, o que é evidenciado pelo gráfico da **Figura** 28b. Note que, no entanto, esta variável se relaciona com o valor médio do aluguel de maneira aproximadamente linear, sendo necessário reavaliar a possibilidade de realizar um novo ajuste, porém considerando esta variável de forma paramétrica. Sendo assim, foi realizado outro ajuste, agora seguindo esta mudança. Isto é, considerar que a variável Fl seja ajustada de forma paramétrica. Os resultados podem ser consultado na **Tabela** 8.

Tabela 8 – Resumo das estimativas de máxima verossimilhança referentes ao segundo modelo ajustado sob distribuição gama para o exemplo do aluguel em Munique.

Efeito	**Estimativa**	**Erro-Padrão**	**Valor-p**
Intercepto	0.0024	0.000130	< 0.0001
Fl	-0.00001	0.000001	< 0.0001
loc2	-0.00033	0.000108	< 0.001
loc3	-0.00046	0.000100	< 0.001
B1	0.00116	0.000400	< 0.01
L1	-0.00011	0.000059	< 0.05
	Outras medidas do ajuste		
$df(\alpha)$	2.12		
α	5439268		
ϕ	0.117		
VCG	0.122		

Desta forma, ao observar o gráfico da **Figura** 29, pode-se notar que, de 1900 à 1940 os preços dos alugueis apresentaram queda. Esse período foi marcado por crises, recessões e guerras na Europa, prováveis fatores que levaram a esta redução dos alugueis em Munique. Após as destruições provocadas neste conturbado período, foi implantado pelos Estados Unidos o plano de recuperação da Europa, também conhecido como Plano Marshall. Este plano consistia em uma política de ajuda econômica para a reconstrução dos países aliados, tais como a Alemanha Ocidental. Com isto, a economia de Munique pôde se recuperar, retomando, dentre outros aspectos da economia, o aquecimento do mercado imobiliário.

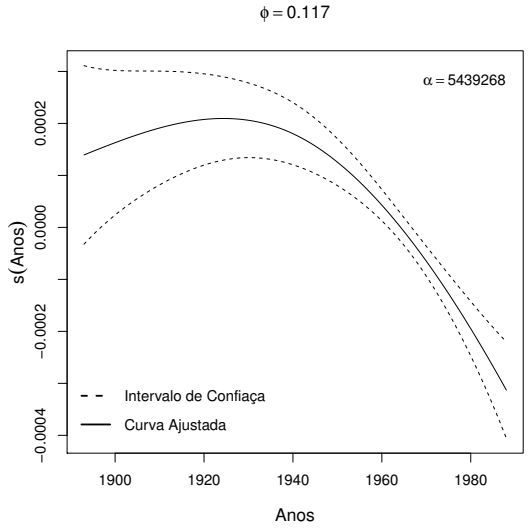


Figura 29 – Gráfico das curvas ajustadas para o segundo modelo sob distribuição gama para o exemplo do aluguel em Munique.

Com base no exposto, por meio da análise gráfica, nota-se que o modelo é um forte candidato a uma boa representação para estudarmos a relação do aluguel médio em função das variáveis explicativas. A seguir, será dado o seu diagnóstico a fim de confirmar a viabilidade do

modelo proposto.

4.2.3 Diagnóstico

A análise de resíduos para este modelo encontra-se na **Figura** 30. Verifica-se que o gráfico dos resíduos contra os valores ajustados apresentam comportamento aleatório e nenhum indício de tendência. De modo semelhante, o gráfico dos resíduos contra os índices também não apresenta comportamento atípico ao esperado. Além disso, apenas alguns poucos pontos encontram-se distantes dois desvios padrão da média, mas os resíduos associados as observações #67 e #390 ultrapassam 3 desvios padrão.

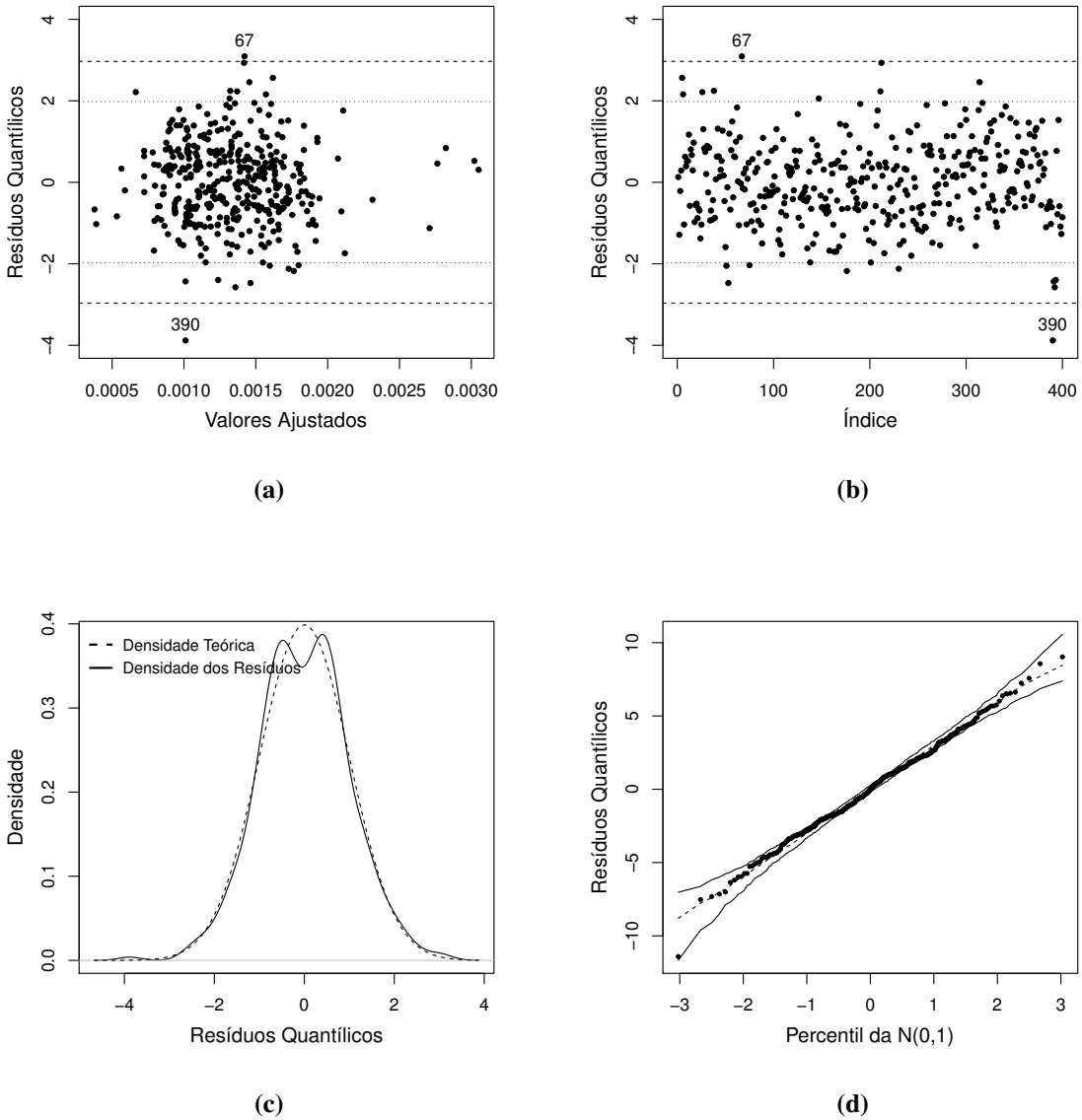


Figura 30 – Gráficos de resíduos referente ao modelo ajustado sob distribuição gama para o exemplo do aluguel em Munique.

A observação #67 corresponde ao aluguel é de 1689.7 (um dos alugueis mais caros presentes no banco de dados), do ano de 1964 cuja área é de 60 metros quadrados. Dos que possuem esse tamanho, essa é a observação com o aluguel mais caro. Além disso, possui boa localidade, banheiro e cozinha padrão. Por outro lado, a observação #390 corresponde ao aluguel de 167.2 (o menor da mostra), com área de 100 metros quadrados, também uma das maiores áreas presentes na amostra. Além disso, possui boa localidade, cozinha e banheiro padrão, para o ano de 1957.

Como visto anteriormente, se o modelo for adequando, os resíduos quantílicos aleatorizados possuem uma distribuição próxima a de uma normal com média zero e desvio padrão igual a um. Logo, os gráficos inferiores da **Figura** 30 dizem respeito a essa característica. Portanto, espera-se que, para um ajuste adequado, os resíduos observados contra os teóricos devem ficar concentrados dentro do envelope simulado e a densidade dos resíduos quantílicos devem se aproximar de uma normal padrão a medida que o tamanho da amostra cresce.

Dessa maneira, pode-se observar que a densidade dos resíduos apresentada está bem próxima a densidade teórica esperada, além de nenhum dos pontos estarem fora do envelope. Ou seja, não há indicações de que a função de ligação é inadequada. Portanto, segundo a análise de resíduo, o modelo proposto não apresenta desvios graves de suposição, sendo apto ao uso. Porém, seguiremos a análise por meio das medidas de alavancagem e influência local para finalizar o diagnóstico.

Segundo o gráfico de alavancagem, destacam-se na **Figura** (31a) as observações que são candidatos a pontos de alavancagem, ou aberrantes. Portanto, é necessário uma análise mais cuidadosa destas observações. Considerando a análise de influência sob os esquemas de perturbação de caso (**Figura** 31b), perturbação na variável resposta (**Figura** 31c) e perturbação em uma das variáveis explicativas (**Figura** 31d), considera-se como observações potencialmente influentes aquelas distantes da linha tracejada, que corresponde a média mais três vezes o desvio padrão de B_i. Desta forma, destacam-se no gráfico os pontos com suposta influência nas estimativas dos parâmetros, sendo necessário realizar uma análise mais cuidadosa dessas observações.

Portanto, assim como desenvolvido no exemplo anterior, devemos remover essas observações e realizar novo ajuste. Em seguida, calcular a taxa de variação das estimativas deste ajuste em relação ao ajuste com todas as observações, segundo a expressão 4.2 e observar se as estimativas mudaram de maneira muito drástica, bem como se houve mudança inferencial. Com isso, obteve-se a **Tabela** 9, com as estimativas do modelo sem as observações destacadas e

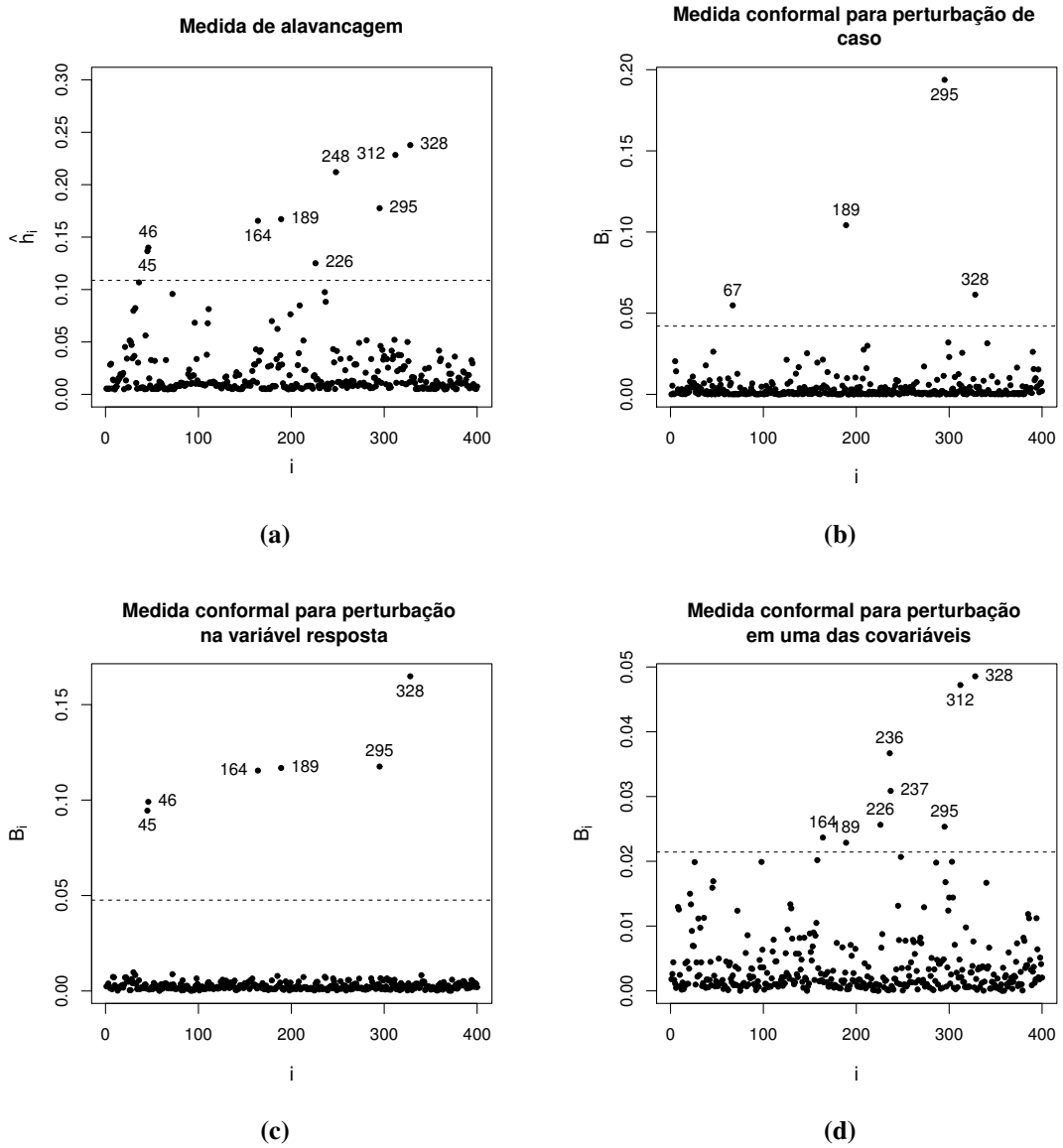


Figura 31 – Gráficos de alavancagem e influência local sob os esquemas de perturbação de caso, perturbação na variável resposta e perturbação em umas das variáveis explicativas, respectivamente, referentes ao ajuste do modelo sob distribuição gama para o exemplo do aluguel em Munique.

mudança relativa das estimativas.

O valor de referência para comparar a mudança relativa das estimativas após a retirada de um ponto potencialmente influente é de 1.5%. Sendo assim, espera-se que na retirada de uma observação a mudança relativa seja de aproximadamente essa porcentagem, se esta não causa influência nas estimativas. Então, na **Tabela** 9 temos que as observações causam impactos desproporcionais, principalmente nas estimativas de β_4, correspondente ao tipo de banheiro, especialmente quando foram retiradas as observações #189, #295 e #328, bem como nas estimativas de β_5, correspondente a qualidade da cozinha. Em particular, quando é retirada a observação #328, a estimativa desse parâmetro aumentar 10.6%. Além disso, quando as

observações #236 e #312 foram retiradas do ajuste, houve mudança inferencial ao nível de 5% para este mesmo parâmetro.

Tabela 9 – Estimativas de máxima verossimilhança referente ao ajuste do modelo sob distribuição gama e mudança relativa em porcentagem entre parênteses para os parâmetros β_0, β_1, β_2, β_3, β_4, β_5 e β_6 após a retirada dos pontos para o exemplo do aluguel em Munique.

Observação retirada	Estimativas dos parâmetros					
	$\hat{\beta_0}$	$\hat{\beta_1}$	$\hat{\beta_2}$	$\hat{\beta_3}$	$\hat{\beta_4}$	$\hat{\beta_5}$
#164	2.46×10^{-3} (0.04%)	1.19×10^{-5} (0.02%)	3.33×10^{-4} (0.28%)	4.62×10^{-4} (0.13%)	1.2×10^{-3} (5.87%)	1.19×10^{-4} (0.13%)
#189	2.46×10^{-3} (0.3%)	1.19×10^{-5} (0.15%)	3.43×10^{-4} (2.84%)	4.72×10^{-4} (2.05%)	1.32×10^{-3} (12.91%)	1.19×10^{-4} (0.49%)
#236	2.46×10^{-3} (0.07%)	1.19×10^{-5} (0.18%)	3.32×10^{-4} (0.37%)	4.62×10^{-4} (0.05%)	1.17×10^{-3} (0.12%)	1.15×10^{-4} (2.71%)
#295	2.47×10^{-3} (0.09%)	1.19×10^{-5} (0.03%)	3.36×10^{-4} (0.75%)	4.65×10^{-4} (0.68%)	9.69×10^{-4} (17.07%)	1.19×10^{-4} (0.01%)
#312	2.48×10^{-3} (0.56%)	1.12×10^{-5} (2.07%)	3.27×10^{-4} (1.94%)	4.66×10^{-4} (0.82%)	1.17×10^{-3} (0.17%)	1.06×10^{-4} (10.60%)
#328	2.47×10^{-3} (0.09%)	1.19×10^{-5} (0.14%)	3.34×10^{-4} (0.32%)	4.64×10^{-4} (0.43%)	1.05×10^{-3} (10.16%)	1.18×10^{-4} (0.31%)

Sendo assim, foi destacado que a observação #189 corresponde ao aluguel de 450.4 € com cozinha padrão e banheiro premium, dentre os quais o aluguel é o segundo maior. Por outro lado, a observação #236 corresponde a um apartamento com banheiro padrão e cozinha premium cujo aluguel é de 1909 €, sendo este o maior valor dentre os apartamentos com cozinha premium. Para a observação #295, foi constatado que corresponde a um apartamento com banheiro e cozinha padrão, porém em uma ótima localidade, bem como um aluguel de 233.0 €. O quarto aluguel mais baixo da amostra. Com relação a observação #312, temos que corresponde a um apartamento com banheiro e cozinhas padrão, porém em uma ótima localidade cujo aluguel, avaliado no ano de 1985, foi de 2000.0 €. O segundo maior valor da amostra. Por fim, tomando a observação #328, constatou-se que se trata de um apartamento com cozinha padrão, ótima localidade e banheiro premiu, dentre os quais possui o maior aluguel.

4.3 AR POLUÍDO EM CHICAGO

Nesta aplicação, considera-se um conjunto de dados apresentado em Wood (2017, p. 243) disponível no pacote *gamair* do R e que relaciona a taxa de mortes diárias, ao longo de vários anos, com os níveis de ozônio, níveis de dióxido de enxofre, temperatura média diária e níveis de material particulado[1]. Além dessas variáveis, a taxa de mortalidade tende a variar

[1] São partículas muito finas de sólidos ou líquidos suspensos no ar, como as que são liberados pelo escapamento de motores a diesel, por exemplo.

conforme o passar do tempo, em especial ao longo do ano, por razões que têm pouco ou nada a ver com a qualidade do ar conforme observa Wood (2017, p. 243). Assim, o objetivo desta aplicação é apresentar uma análise estatística, por meio de um modelo parcialmente linear generalizado para dados de contagem, bem como apresentar as medidas de diagnósticos a fim de verificar a qualidade do ajuste, pondo em prática as técnicas desenvolvidas ao longo das seções anteriores.

4.3.1 Análise exploratória dos dados

Este conjunto de dados reúne uma amostra de 300 observações. Observamos alguns boxplots que facilitam a visualização dos dados. Nota-se que algumas variáveis têm um ou mais pontos discrepantes, sendo necessário investigar o porquê do comportamento atípico.

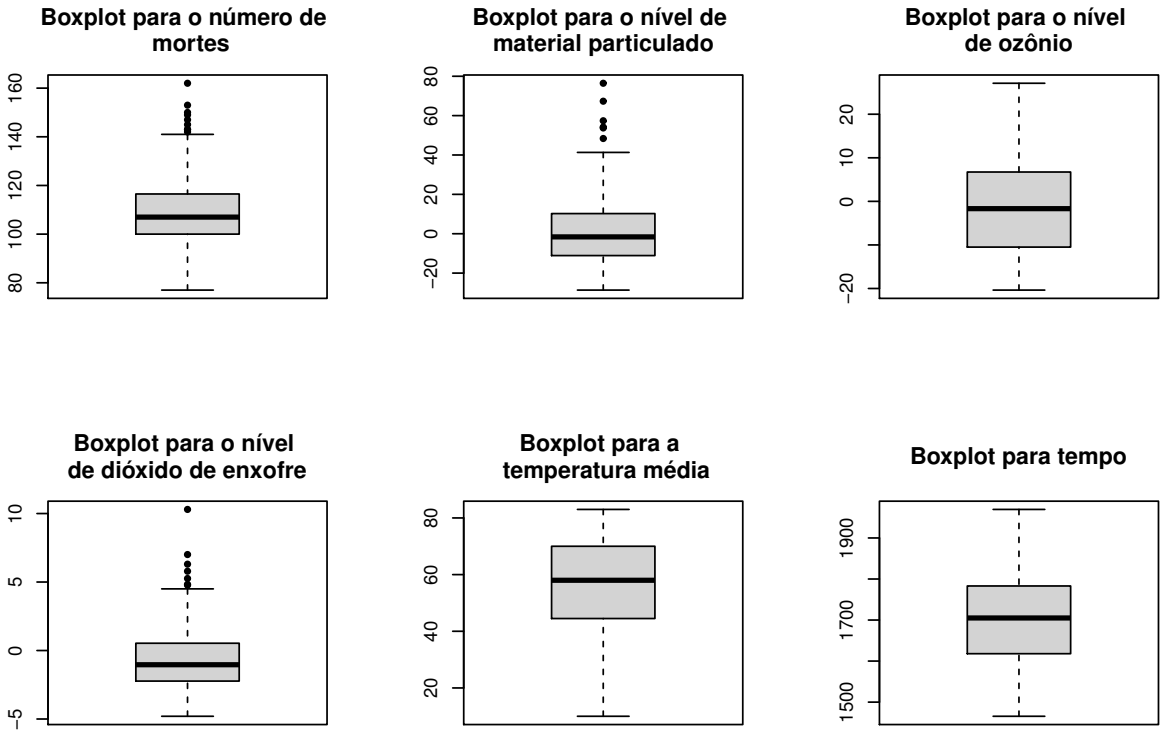


Figura 32 – Boxplots das variáveis presentes no baco de dados considerando o exemplo do ar poluído em Chicago.

Cada uma das variáveis possuem as seguintes características: O *número de mortes diária* possui valor médio dado por $\bar{x} = 109.1$ (por dia) , com desvio padrão igual a $s = 14.17$ (por dia). Além disso, os valores correspondentes ao menor e maior valor para esta variável foram de $x_{(1)} = 77$ (por dia) e $x_{(n)} = 162$ (por dia), respectivamente, bem como os primeiro e terceiros quartis dados por $Q_1 = 100$ (por dia) e $Q_3 = 116.2$ (por dia) com mediana igual a $\text{Md} = 107.0$ (por dia); O *material particulado* possui valor médio dado por $\bar{x} = 1.03\ \text{mg/m}^3$, com desvio padrão igual a $s = 17.17\ \text{mg/m}^3$. Além disso, os valores correspondentes ao menor e maior valor para esta variável foram de $x_{(1)} = -28.63\ \text{mg/m}^3$ e $x_{(n)} = 76.43\ \text{mg/m}^3$, respectivamente, bem como os primeiro e terceiros quartis dados por $Q_1 = -11.082\ \text{mg/m}^3$ e $Q_3 = 10.179\ \text{mg/m}^3$ com mediana igual a $\text{Md} = -1.60\ \text{mg/m}^3$; o *nível de ozônio* possui valor médio dado por $\bar{x} = -1.41$ ppb, com desvio padrão igual a $s = 10.402$ ppb. Além disso, os valores correspondentes ao menor e maior valor para esta variável foram de $x_{(1)} = -20.36$ ppb e $x_{(n)} = 27.11$ ppb, respectivamente, bem como os primeiro e terceiros quartis dados por $Q_1 = -10.49$ ppb e $Q_3 = 6.690$ ppb com mediana igual a $\text{Md} = -1.68$ ppb; a *temperatura média diária* possui média $\bar{x} = 56.18\,^{\circ}\text{F}$, com desvio padrão igual a $s = 16.28\,^{\circ}\text{F}$. Além disso, os valores correspondentes ao menor e maior valor para esta variável foram de $x_{(1)} = 10.00\,^{\circ}\text{F}$ e $x_{(n)} = 83.00\,^{\circ}\text{F}$, respectivamente, bem como os primeiro e terceiros quartis dados por $Q_1 = 44.75\,^{\circ}\text{F}$ e $Q_3 = 70.00\,^{\circ}\text{F}$ com mediana igual a $\text{Md} = 58.00\,^{\circ}\text{F}$; o nível de *dióxido de enxofre* possui valor médio dado por $\bar{x} = -0.6634$ ppb, com desvio padrão igual a $s = 2.35$ ppb. Além disso, os valores correspondentes ao menor e maior valor para esta variável foram de $x_{(1)} = -4.80$ ppb e $x_{(n)} = 10.29$ ppb, respectivamente, bem como os primeiro e terceiros quartis dados por $Q_1 = -2.22$ ppb e $Q_3 = 0.51$ ppb com mediana igual a $\text{Md} = -1.04$ ppb e, finalmente, o tempo em dias possui o valor médio dado por $\bar{x} = 1708$ (por dia), com desvio padrão igual a $s = 112.39$ (por dia). Além disso, os valores correspondentes ao menor e maior valor para esta variável foram de $x_{(1)} = 1466$ (por dia) e $x_{(n)} = 1970$ (por dia), respectivamente, bem como os primeiro e terceiros quartis dados por $Q_1 = 1618$ (por dia) e $Q_3 = 1783$ (por dia) com mediana igual a $\text{Md} = 1705$ (por dia).

Em alguns casos pode ser conveniente trabalhar com o log da variável para obter uma maior simetria. Assim, plotamos os gráficos de dispersão para cada variável segundo o logaritmo da variável número de mortes e o resultado pode ser conferido na **Figura** 33. Assim, podemos notar que parece existir uma relação entre essas variáveis. Nota-se também que a relação entre a variável número de mortes e temperatura, bem como número de mortes e tempo possuem uma relação aparentemente não linear.

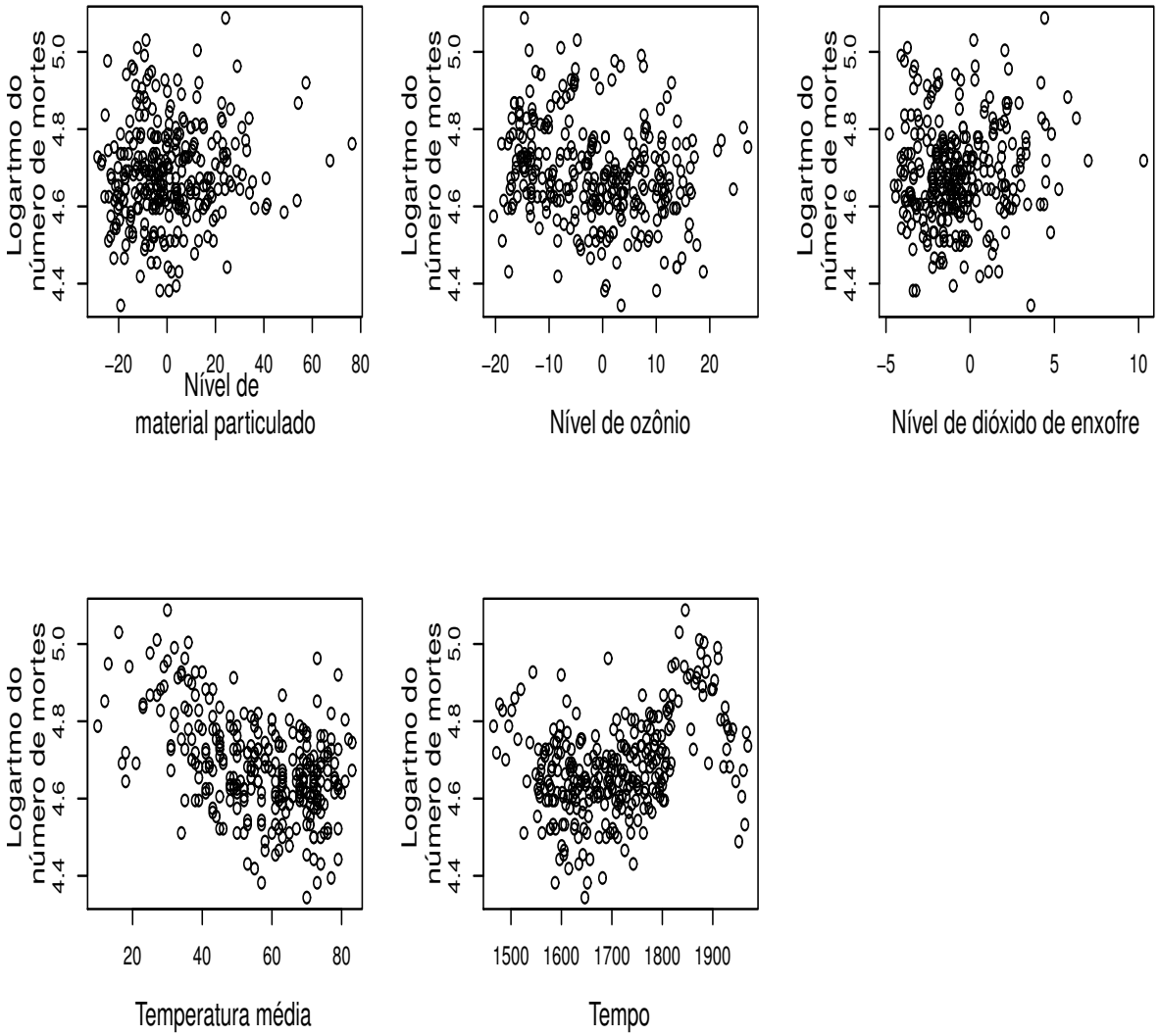


Figura 33 – Diagrama de dispersão do logaritmo do número de mortes contra as demais variáveis numéricas considerando o exemplo do ar poluído em Chicago.

Assim como realizado em exemplos anteriores, neste também será proposto um modelo de regressão. O objetivo é o de acomodar conjuntamente a informação contida nas variáveis explicativas e como elas impactam a mádia da variável resposta. Posteriormente a escolha do modelo de regressão adequando, será realizado uma análise de diagnóstico, visando identificar possíveis afastamentos de suposição, bem como prováveis erros sistemáticos.

4.3.2 Ajuste do modelo

Neste exemplo, desejamos modelar uma variável de contagem, ou seja, uma variável discreta com suporte no conjunto dos inteiros não negativos. Para problemas deste tipo,

comumente a primeira alternativa de modelagem via modelo parcialmente lineares aditivos generalizados faz uso da distribuição Poisson com função de ligação logarítmica. Portanto, para ajuste desse modelo, considera-se que o número de mortes segue uma distribuição de Poisson com valor médio da taxa de mortalidade produto dessas variáveis que medem a qualidade do ar. Dessa forma, o modelo será dado por:

$$\begin{aligned} g(\mu_i) &= \log(\mu_i) = \eta_i \\ &= \beta_0 + \beta_1 MP_i + \beta_2 \mathrm{SO}_{2i} + f_1(\mathrm{O}_{3i}) + f_2(\mathrm{tmpd}_i) + f_3(\mathrm{Tempo}_i) \end{aligned}$$

em que $\mu_i = \mathbb{E}(\mathrm{Morte}_i)$, para $i = 1,2,3,\ldots,300$, onde $f_k(\cdot)$, com $k = 1,2$, são funções suaves das covariáveis ozônio, temperatura média diária e tempo. As quais, acredita-se que possuem relação não linear e, portanto, devem ser estimadas de forma não paramétrica. Além disso, foi utilizada a função de ligação logarítmica.

Assim, para o ajuste deste modelo, destaca-se que foram empregados o método de P-*spline* para o ajuste da curva $f_k(\cdot)$, baseado em um B-*spline* cúbico com 10 nós equidistantes em todas as curvas e penalização quadrática. Os graus de liberdade efetivos são obtidos pela soma dos graus de liberdade referente a função não paramétrica mais os graus de liberdade referente a parte ão paramétrica do modelo. Além disso, foi utilizado o método VCG para estimar o vetor de parâmetros de suavização $\boldsymbol{\alpha}$. A **Tabela** 10 apresenta as estimativas de máxima verossimilhança penalizada para ajuste do modelo.

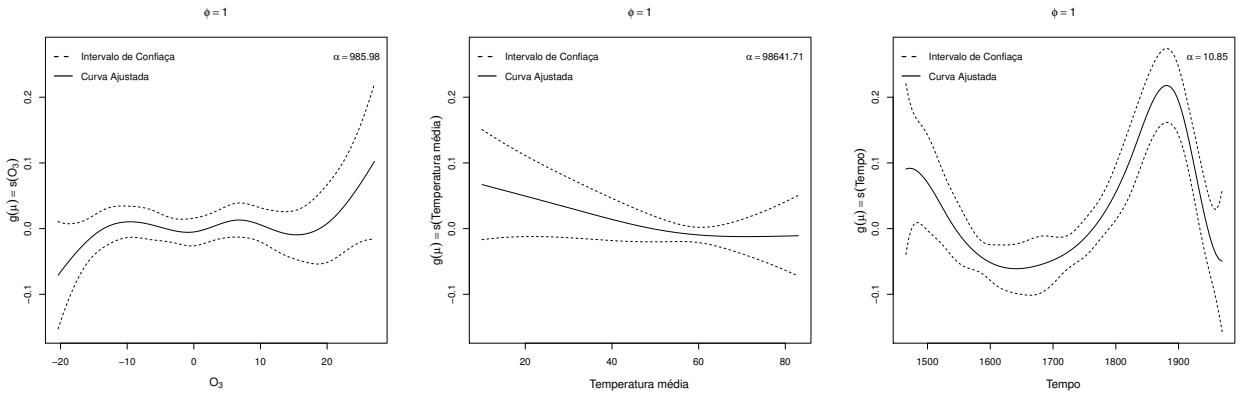
Tabela 10 – Resumo das estimativas de máxima verossimilhança referentes ao modelo ajustado sob distribuição Poisson para o exemplo do ar poluído em Chicago.

Efeito	**Estimativa**	**Erro-Padrão**	**Valor-p**
Intercepto	4.68	0.006	< 0.05
MP	0.001	0.0004	< 0.05
SO_2	-0.0007	0.003	0.8
Outras medidas do ajuste			
$df_1(\boldsymbol{\alpha})$	5.03		
$df_2(\boldsymbol{\alpha})$	1.73		
$df_3\boldsymbol{\alpha})$	8.10		
$\boldsymbol{\alpha}$	(985.98, 98641.71, 10.85)		
ϕ	1		
VCG	0.017		

Note que a variável SO_2 não foi significativa para o modelo. Além disso, o custo total para a estimação dos parâmetros dos componentes paramétricos e não paramétricos foi

$df(\boldsymbol{\alpha}) = 3 + df_1(\boldsymbol{\alpha}) + df_2(\boldsymbol{\alpha}) + df_3(\boldsymbol{\alpha}) = 17.85$, bem como critério VCG igual a 0.0222. Por meio da **Figura** 34, observamos as estimativas das curvas ajustada de forma não paramétrica.

Por fim, realizou-se mais um ajuste. No entanto, considerando apenas as variáveis material particulado, níveis de ozônio e tempo. Assim, a **Tabela** final 11 resultante apresenta um resumo das estimativas e outras quantidades do referido ajuste.



(a) **(b)** **(c)**

Figura 34 – Gráficos das curvas ajustadas sob distribuição Poisson para o exemplo do ar poluído em Chicago.

Como a variável temperatura média possui grau de liberdade pequeno, bem como a curva ajustada é aproximadamente linear, então resolveu-se fazer um novo ajuste retirando a variável SO_2, que não foi significativa, assim como tomou-se a variável temperatura média sem suavizá-la. Entretanto, a avariável temperatura média também não foi significativa ao nível de significância de 5% para este segundo modelo.

Tabela 11 – Resumo das estimativas de máxima verossimilhança referentes ao modelo final ajustado sob distribuição Poisson para o exemplo do ar poluído em Chicago.

Efeito	**Estimativa**	**Erro-Padrão**	**Valor-p**
Intercepto	4.68	0.005	< 0.0001
MP	0.001	0.0003	< 0.0001
Outras medidas do ajuste			
$df_1(\boldsymbol{\alpha})$	5.05		
$df_2(\boldsymbol{\alpha})$	8.28		
$\boldsymbol{\alpha}$	(965.7,7.85)		
ϕ	1		
VCG	0.018		

Desta forma, o modelo final será explicado de forma paramérica pela variável material particulado, sendo possível realizar inferências sobre a relação ou saber qual o efeito da mudança de valor dessa variável em relação ao número de mortos. Por outro lado, também será explicado pelas componentes não paramétricas, perdendo um pouco do aspecto inferencial, mas ganhando muito no que tange a predição de novas observações.

4.3.3 Diagnóstico

Pela análise de resíduo tem-se que, se o modelo for adequando, os resíduos quantílicos aleatorizados possuem uma distribuição próxima a de uma normal com média zero e desvio padrão igual a um. Dessa forma, espera-se que os resíduos quantílicos contra os valores ajustados demonstrados no gráfico apresentem comportamento aleatório em tono de zero e que os resíduos versus índices demonstrados no outro gráfico não apresentem comportamento serial aparente. Além disso, espera-se também que tanto a densidade quanto o envelope simulado dos resíduos apresentem um comportamento característico para essa distribuição. Assim, é possível observar que, por meio da **Figura** (35), os gráficos dos resíduos contra os valores ajustados e o dos resíduos contra os índices, assim como os gráficos da densidade e envelope simulado apresentam comportamentos adequados.

Ademais, considerando pontos extremos aqueles que apresentam resíduos com 3 desvios padrão da média, então estes gráficos não apresentam pontos extremos. No entanto, apresentam alguns pontos de alarmes, que são merecedores de análise mais cuidadosa. Ou seja, pontos que estão acima de dois desvios padrão da média, porém abaixo de três desvios padrão. Portanto, pode-se tomar o modelo, tal como foi proposto, como um forte candidato para representar adequadamente uma aproximação da verdadeira natureza geradora dos dados.

Seguindo a análise de diagnóstico, considerando as medidas de alavancagem para o modelo proposto, observa-se que os pontos em destaque são os que possuem as alavancagens mais altas ilustradas na **Figura** (36a). Portanto, sugere-se uma análise mais cautelosa destes pontos. Considerando a análise de influência sob os esquemas de perturbação de caso e em uma das variáveis explicativas, foram obtidos, respectivamente para cada tipo de observação, os gráficos da curvatura conformal contra a ordem das observações apresentado nas **Figuras**. 36b e 36b.

Neste caso, deve-se retirar esses pontos, realizar um novo ajuste e comparar as estimativas do novo modelo com as do modelo anterior. Assim, pela retirada desses pontos e

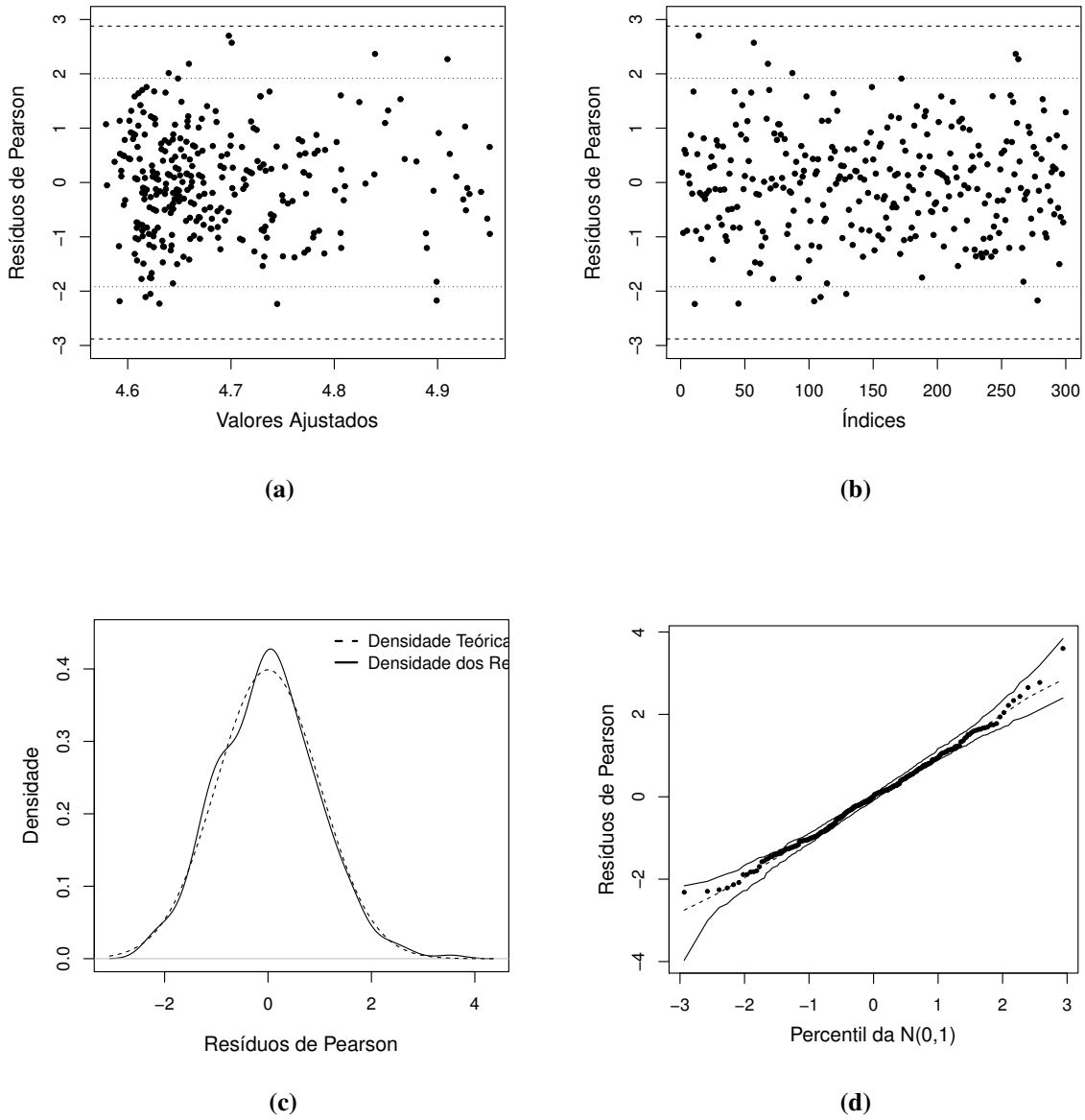
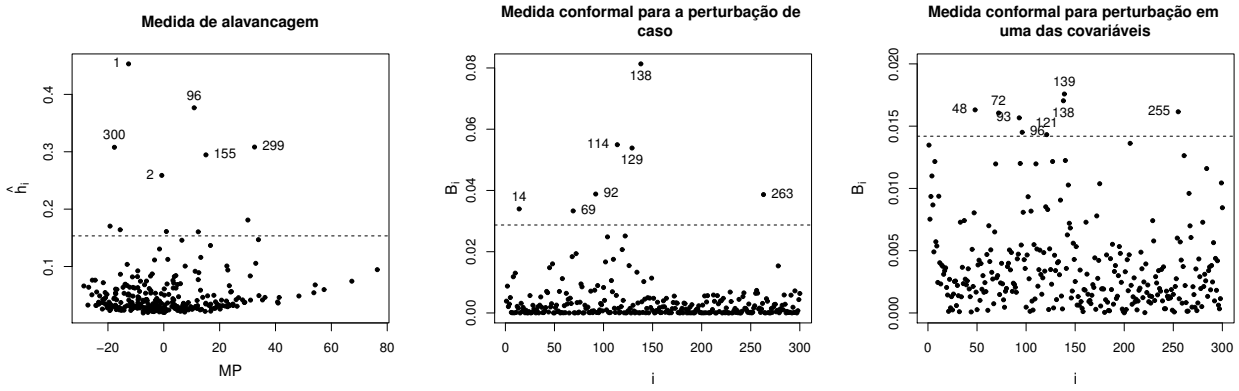


Figura 35 – Gráficos de resíduos referentes ao modelo ajustado sob distribuição Poisson para o exemplo do ar poluído em Chicago.

novo ajuste, obteve-se os resultados presentes na **Tabela** 12, onde podemos observar a mudança relativa entre o modelo completo e o modelo sem as observações potencialmente influentes. Após a confirmação de que as observações destacadas são influentes, devemos saber o porquê destas impactarem de forma desproporcional nas estimativas dos parâmetros.

O valor de referência para comparar as mudanças relativas das estimativas para este exemplo é de 0.66%. Assim, espera-se que na retirada de uma observação, as mudanças relativas sejam próximas a essa magnitude se a observação não for influente. Desta forma, nota-se que as estimativas do intercepto pouco mudam com a retirada individual destas observações, bem



(a) **(b)** **(c)**

Figura 36 – Gráfico de influência local sob os esquemas de perturbação de caso e perturbação na variável resposta respectivamente para o exemplo do ar poluído em Chicago.

como não mudou a inferência. Isto é, os parâmetros permaneceram significativos. Porém, as estimativas do parâmetro associado variável material particulado, β_1, aumenta quase 10% com a retirada da observação #138. Ao se investigar esta observação, nota-se que corresponde a maior quantidade de material particulado associado a um total de 143 mortes, um dos sete dias que mais houveram mortes.

Tabela 12 – Estimativas de máxima verossimilhança referente ao ajuste do modelo sob distribuição poisson e mudança relativa em porcentagem entre parênteses para os parâmetros β_0 e β_1 após a retirada dos pontos para o exemplo do ar poluído em Chicago.

Observação retirada	Estimativas dos parâmetros	
	$\hat{\beta_0}$	$\hat{\beta_1}$
#1	4.68 (0.01%)	0.0013 (0.15%)
#96	4.69 (0.0%)	0.001 (0.06%)
#114	4.69 (0.02%)	0.001 (1.54%)
#129	4.69 (0.02%)	0.001 (0.81%)
#138	4.69 (0.02%)	0.001 (10.37%)
#139	4.69 (0.01%)	0.001 (0.20%)
#255	4.69 (0.01%)	0.001 (0.00%)

4.4 REPRODUÇÃO DE AVES EM PORTUGAL

Nesta aplicação, considera-se parte de um conjunto de dados apresentado em Wood (2017). Este conjunto de dados descreve como é distribuída a reprodução de aves da espécie Cotovia-de-poupa em Portugal. O objetivo será, neste exemplo, estudar como a probabilidade de reprodução dessa espécie de ave varia no país, considerando, para isso, a localidade: "x"(em km considerando a direção leste de uma origem) e "y"(em km considerando a direção norte de uma origem). O objetivo é identificar onde as aves Cotovia-de-crista costumam se reproduzir em Portugal, considerando subdivisões do país em quadras de 2km por 2km. Para este exemplo, foi considerada uma mostra de 250 localidades. O resultado pode ser observado no gráfico da **Figura** 37.

Neste gráfico, as coordenadas x e y são usadas para indicar a presença ou ausência de aves da espécies Cotovia-de-crista se reproduzindo em Portugal. Portanto, os pontos destacados em preto indicam a presença de aves se reproduzindo naquela região (ou para aquelas coordenadas). Nota-se que para o lado leste e em regiões próximas a fronteira, bem como, principalmente, na região sul do país há uma maior concentração dessas aves se reproduzindo.

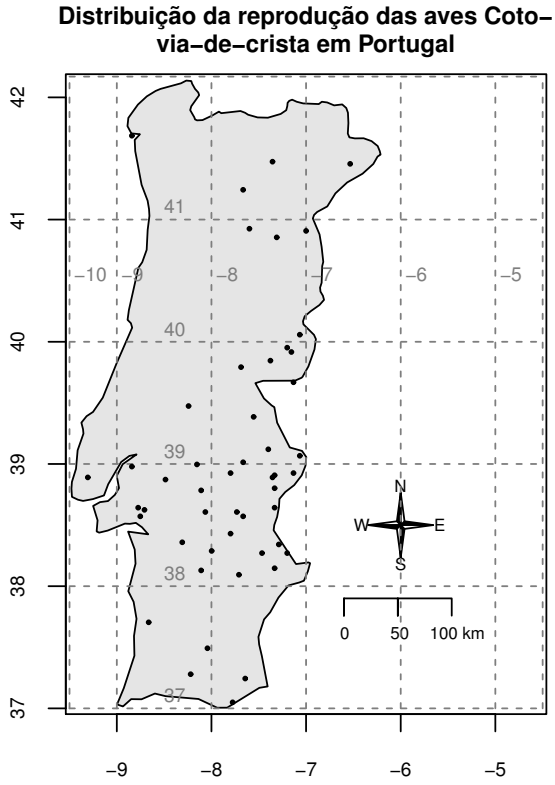


Figura 37 – Gráfico que mostra como está distribuída a reprodução das aves Cotovia-de-crista em Portugal.

A seguir, será proposto um modelo de regressão para tentar estimar a relação da reprodução dessas aves segundo a região definida no mapa. Em seguida, serão apresentadas as análises de diagnóstico para este modelo por meio da análise de resíduos, alavancagem e análise de influência local sob os esquemas de pertubação definidos e exemplificados em seções anteriores.

4.4.1 Ajuste do modelo

Para o ajuste desse modelo, considera-se que a presença de aves nessas subdivisões segue uma distribuição binomial com média subjacente produto dessas variáveis. Ou seja, considera-se que

$$g(\mu_i) = \eta_i \tag{4.3}$$

$$= \beta_0 + \beta_1 y_i + f(x_i),\ i = 1,2...n. \tag{4.4}$$

Destaca-se que foram empregados o método de P-*spline* para o ajuste da curva $f(\cdot)$ baseado em um B-*spline* cúbico com 10 nós equidistantes e penalização quadrática. Os graus de liberdade efetivos são obtidos pela soma dos graus de liberdade referente a função não paramétrica mais os graus de liberdade referente a parte paramétrica do modelo. Além disso, foi utilizado o método VCG para estimar o vetor de parâmetros de suavização α, bem como a função de ligação utilizada logit. A **Tabela** (13) apresenta as estimativas de máxima verossimilhanças penalizada para ajuste do modelo.

Tabela 13 – Resumo das estimativas de máxima verosimilhança referentes ao modelo ajustado sob distribuição binomial para o exemplo dda reprodução das aves

Efeito	**Estimativa**	**Erro-Padrão**	**Valor-p**
Intercepto	22.69	5.954	< 0.0001
MP	-5.546×10^{-6}	1.38×10^{-6}	< 0.0001
Outras medidas do ajuste			
$df(\alpha)$	5.01		
α	17.764		
ϕ	1		
VCG	-0.094		

Note que tanto o intercepto quanto a variável y foram significativas para o modelo. Além disso, os graus de liberdade efetivos foram estimados em $df(\alpha) = 2 + df_1(\alpha) = 5.01$ e

que o grau de liberdade referente a variável suavizada x foi estimado em 3.01. Por fim, destaca-se que o critério VCG obtido foi de -0.094.

Por meio da **Figura** 38, podemos observar a representação gráfica do supracitado ajuste. Neste gráfico, nota-se que, de fato, a relação entre as variáveis se dá de forma mais complexa que uma simples estrutura linear. Assim, o modelo será bastante preciso em análises de predição, embora perca um pouco da interpretabilidade.

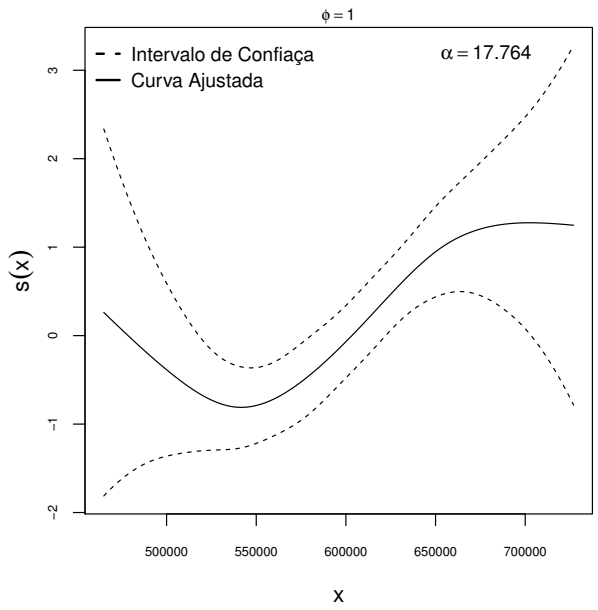


Figura 38 – Curva aproximada e intervalos de confiança pontuais de 95% **referente ao modelo ajustado sob distribuição binomial para o exemplo da reprodução das aves.**

Nosso estudo agora se volta para verificar a validade das suposições assumidas para o modelo. Neste caso, será empregada técnicas de diagnósticos, por meio das quais será possível identificar desvios de suposições. Para tanto, faremos uso da análise de resíduos, alavancagem e análise de influência local. Superada essa fase, nosso modelo estará apto para uso.

4.4.2 Diagnóstico

Para análise de resíduo, considera-se a **Figura** 39 apresentada a seguir. Portanto, observa-se que, como visto anteriormente, o modelo será adequado se os resíduos quantílicos aleatorizados possuírem uma distribuição próxima a de uma normal com média zero e desvio padrão igual a um. Dessa maneira, pode-se observar que a densidade dos resíduos apresentada está bem próxima a densidade teórica esperada, além de nenhum dos pontos estarem fora do envelope.

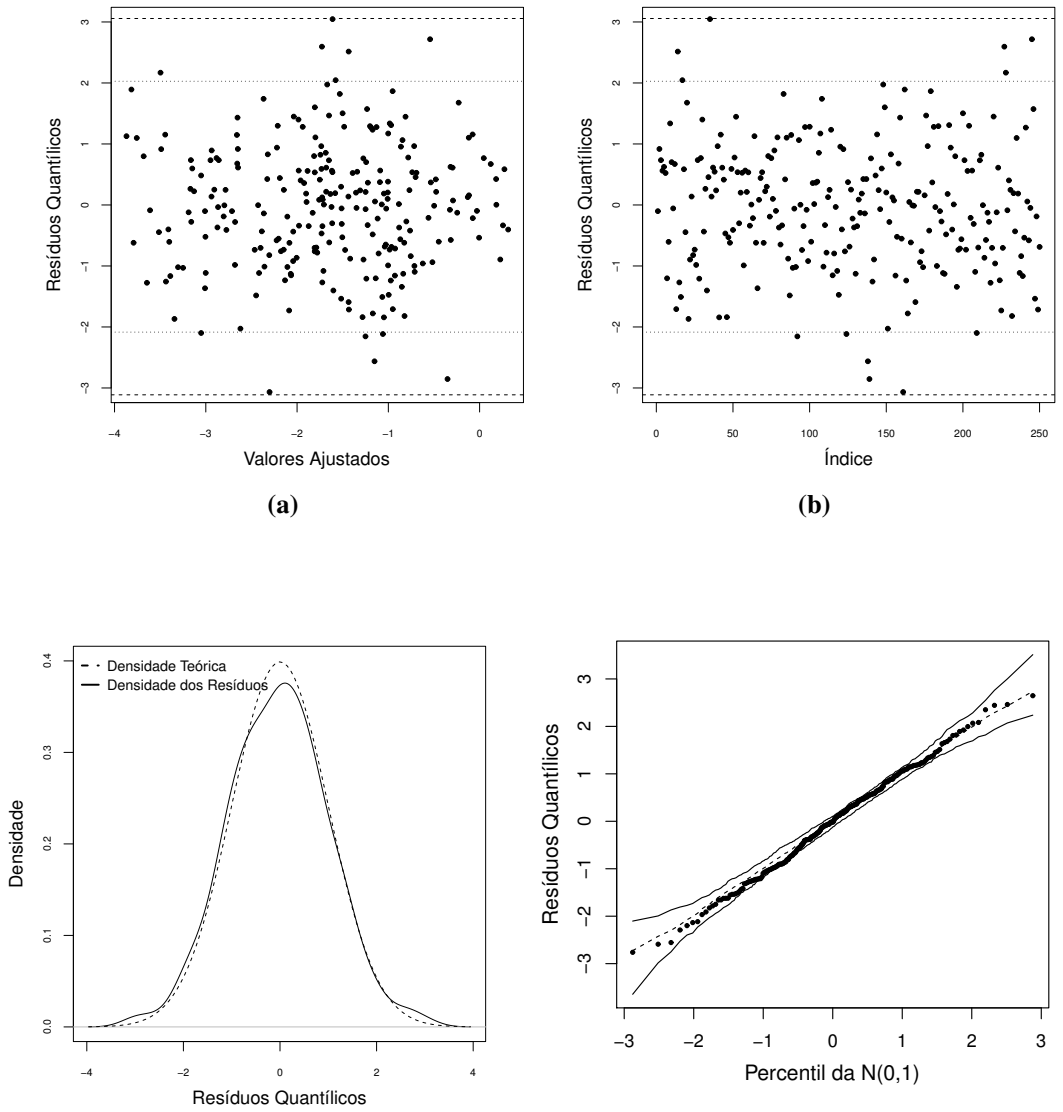


Figura 39 – Gráfico de resíduos referente o modelo ajustado sob distribuição binomial para o exemplo da reprodução das aves.

Portanto, segundo a análise de resíduo, o modelo proposto está apto ao uso. Porém, seguiremos a análise por meio das medidas de alavancagem e influência local para finalizar o diagnóstico. Dessa forma, pode-se tomar o modelo tal qual foi proposto como um forte candidato para representar adequadamente uma aproximação da verdadeira natureza geradora dos dados.

Seguindo a análise de diagnóstico, considerando as medidas de alavancagem para o modelo proposto, observa-se que os pontos em destaque são os que possuem as alavancagens mais altas ilustrada na **Figura** 40. Portanto, sugere-se uma análise mais cautelosa destes. Considerando a análise de influência sob os esquemas de perturbação de caso e na variável explicativa, observa-se que existem observações como possíveis observações influentes para

o gráfico correspondente a perturbação de caso. Por outro lado, não houveram observações influentes paro o caso em que perturbamos a variável explicativa.

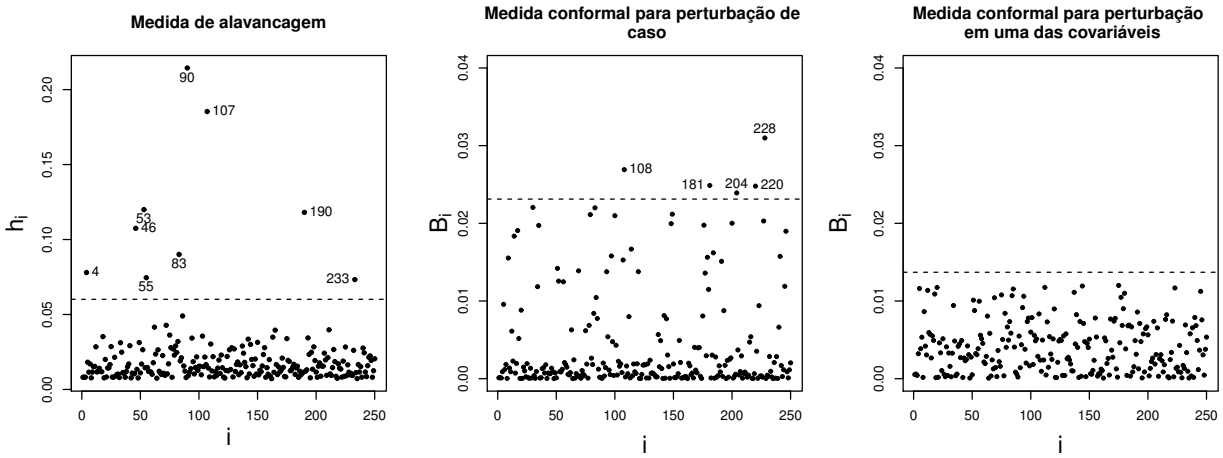


Figura 40 – Gráficos de de alavancagem e influência local sob os esquemas de perturbação de caso e perturbação em uma das variáveis explicativas, respectivamente, referente ao ajuste do modelo sob distribuição binomial para o exemplo da reprodução das aves.

Assim, novamente, analisamos o impacto das observações destacadas na **Figura** 39 sob as estimativas dos parâmetros, por meio de uma análise de sensibilidade. Desta forma, consideramos a **Tabela** 14 em que temos as observações retiradas seguidas dos valores das estimativas e as respectivas mudanças relativas em porcentagem para a análise de influência local. Para este exemplo, o valor de referência para comparar a mudança relativa das estimativas é de 0.8%. Portanto, mudanças relativas distantes desse valor confirma que a observação destacada é, de fato, influente nas estimativas dos parâmetros do modelo. Sendo assim, temos que as observações destacadas causam mudanças relativas substanciais nas estimativas dos parâmetros. Porém, não houveram mudanças inferenciais.

Tabela 14 – Estimativas de máxima verossimilhança e mudança relativa em porcentagem entre parênteses para os parâmetros β_0 e β_1 do modelo binomial após a retirada dos pontos para o exemplo da reprodução das aves em Portugal.

Estimativas dos parâmetros	Observação retirada				
	#108	#181	#204	#220	#228
$\hat{\beta}_0$	24.667 (8.705%)	24.270 (6.957%)	23.439 (3.294%)	24.228 (6.770%)	25.527 (12.495%)
$\hat{\beta}_1$	-60.09×10^{-7} (8.339%)	-59.16×10^{-7} (6.663%)	-57.27×10^{-7} (3.257%)	-59.07×10^{-7} (6.499%)	-62.10×10^{-7} (11.965%)

5 CONCLUSÕES E TRABALHOS FUTUROS

Na presente dissertação, discutimos o desenvolvimento da teoria que envolve os modelos parcialmente lineares aditivos generalizados com suavização por meio de P-*slines*. Neste contexto, desenvolvemos medidas para análise de diagnóstico por meio das principais estruturas conhecidas na literatura para este fim. Assim, sugerimos os resíduos de Pearson e os resíduos quantílicos aleatorizados com o intuito de identificar possíveis *outliers*, bem como discutimos o uso da matriz de projeção em modelos semi-paramétricos para investigar pontos de alavanca. Além disso, derivamos as expressões para realizarmos a análise de influência local, nossa principal contribuição teórica, baseando-se na proposta de Cook. Com isso, destaca-se que foram elaborados três esquemas de perturbação, a saber: perturbação no modelo (caso ponderado), perturbação na variável resposta e perturbação em uma das covariáveis, com o objetivo de estudar as estimativas do modelo sob tais esquemas de perturbação. Neste sentido, derivamos os resultados teóricos, além de elaborarmos uma implementação por intermédio do *software* R a fim de viabilizar o uso da técnica. Por fim, apresentamos quatro exemplos para que fosse possível vislumbrar o uso do método em diferentes distribuições.

Por meio dos exemplos apresentados, foi possível pôr em prática as técnicas desenvolvidas neste trabalho, bem como observar o desempenho destas medidas. Assim sendo, vimos que as técnicas de diagnóstico são extremamente úteis para ajudar a entender melhor as suposições inerentes ao modelo e são imprescindíveis para quaisquer análises estatísticas mais avançadas. Desta forma, as medidas desenvolvidas nesta dissertação são contribuições muito importantes para a comunidade estatística.

Em termos computacionais, a implementação dessas medidas a princípio parecem complicadas. Porém, o *software* R oferece suporte computacional bem sofisticado e de compreensão razoável, o que possibilitou desenvolver uma função para realizar os disgnósticos, baseada nas saídas do pacote **mgcv**. Estas funções e outros códigos usados nesta dissertação encontrasse no apêndice A, que podem motivar a criação de um pacote. Por fim, seria interessante investigar o desempenho dessas técnicas em diferentes cenários, o que só seria possível por meio de um estudo de simulação.

REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. In: **Selected Papers of Hirotugu Akaike**. [S.l.]: Springer, 1974. p. 215–222.

BATES, D. M.; WATTS, D. G. Nonlinear regression: iterative estimation and linear approximations. **Nonlinear Regression Analysis and Its Applications, John Wiley & Sons, Inc., Hoboken, NJ, USA, doi**, v. 10, n. 9780470316757, p. 9393–9441, 1988.

BEALE, E. Confidence regions in non-linear estimation. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 22, n. 1, p. 41–76, 1960.

BILLOR, N.; LOYNES, R. Local influence: a new approach. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 22, n. 6, p. 1595–1611, 1993.

BOOR, C. D. A practical guide to splines. **Springer**, Berlin, 1978.

BREIMANE, L.; FRIEDMAN, J. H. Estimating optimal transformations for multiple regression and correlation. **Journal of the American Statistical Association**, v. 80, p. 580–298, 1985.

BUJA, A.; HASTIE, T.; TIBSHIRANI, R. Linear smoother and additive models. **The Annals of Statistics**, v. 17, p. 453–555, 1989.

COLOSIMO, A. E.; GIOLO, S. R. **Análise de Sobrevivênvia Aplicada**. [S.l.]: Associação Brasileira de Estatística, 2006. v. 1.

COOK, R. D. Assessment of local influence. **Journal of the Royal Statistical Society - Series B (Methodological)**, 1978.

CORDEIRO, G. M.; DEMÉTRIO, C. G. B. **Modelos Lineares Generalizados e Extensões**. Recife - PE: Departamento de Estatística e Informática - UFRPE, 2008.

COX, D. R.; SNELL, E. J. A general definition of residuals (with discussion). **Journal of the Royal Statistical Society B**, v. 30, p. 248–275, 1968.

DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. **Journal of Computational and Graphical Statistics**, v. 5, p. 236–244, 1996.

EILERS, P. H. C.; MARX, B. D. Flexible smoothing with b-splines and penalties. **Statistical Science**, 1996.

EILERS, P. H. C.; MARX, B. D.; MARIA, D. Twenty years of p-splines. Statistics and Operations Research Transactions, v. 39(2), p. 149–186, 2015.

EMAMI, H. Local influence in ridge semiparametric models. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 86, n. 17, p. 3357–3370, 2016.

EUBANK, R. L. The hat matrix for smoothing splines. **Statistics and Probability Letters**, v. 2, p. 9–14, 1984.

EUBANK, R. L. **Spline Smoothing and Nonparametric Regression**. [S.l.]: Marcel Dekker, 1988.

FAHRMEIR, L.; KNEIB, T.; LANG, S.; MARX, B. **Regression: Models, Methods and Applications**. [S.l.]: Springer, 2013.

FRIEDMAN, J.; SILVERMAN, B. W. Flexible parsimonious smoothing and additive modeling (with discussion). **Technometrics**, v. 31, p. 2–39, 1989.

GREEN, P. J.; SILVERMAN, B. W. Nonparametric regression and generalized linear models. Chapman and Hall, 1994.

GREEN, P. J.; YANDELL, B. S. Semi-parametric generalized linear models. **Lecture Notes in Statistics**, v. 32, p. 44–55, 1985.

HÄRDLE, W. Applied nonparametric regression. Cambridge Univ. Press, 1990.

HASTIE, T.; TIBSHIRANI, R. Generalized additive models. **Chapman and Hall**, 1990.

HOAGLIN, D. C.; WELSCH, R. E. The hat matrix in regression and anova. **The American Statistician**, v. 32, p. 17–22, 1978.

HODGES, J. S. **Richly parameterized linear models: additive, time series, and spatial models using random effects**. [S.l.]: Chapman and Hall/CRC, 2016.

HOLANDA, A. A. Modelos lineares parciais aditivos generalizados com suavização por meio de p-splines. Depertamento de Estatística - USP, São Paulo - SP, Maio 2018.

IHAKA, R.; GENTLEMAN, R. R: A language for data analysis and graphics. **Journal of Computational and Graphical Statistics**, v. 5, p. 299–314, 1996.

IZBICKI, R.; SANTOS, T. M. **Machine Learning sob a ótica estatística: Uma abordagempreditivista para a Estatística com Exemplos em R**. [S.l.]: Depertamento de Estatística - UFSCAR, 2019.

KOOPERBERG, C.; STONE, C. J. Logspline density estimation for censored data. **J. Comput. Graph. Statist**, v. 1, p. 301–328, 1992.

LEE, S.-Y.; LU, B.; SONG, X.-Y. Assessing local influence for nonlinear structural equation models with ignorable missing data. **Computational statistics & data analysis**, Elsevier, v. 50, n. 5, p. 1356–1377, 2006.

MANGHI, R. F. Técnicas de diagnóstico em modelos parcialmente lineares aditivos generalizados para dados correlacionados. Universidade Federal de Pernambuco - Departamento de Estatística, 2016.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. **Journal of the Royal Statistical Society A**, v. 135, p. 370–384, 1972.

O'SULLIVAN, F. A statistical perspective on ill-posed inverse problems (with discussion). **Statist. Sci.**, v. 1, p. 505–527, 1986.

PAULA, G. A. **Modelos de Regressão com apoio computacional**. Universidade de São Paulo: Instituto de Matemática e Estatística, 2013. 2 p.

PAYANDEH, B. Some applications of nonlinear regression models in forestry research. **The Forestry Chronicle**, NRC Research Press Ottawa, Canada, v. 59, n. 5, p. 244–248, 1983.

PULGAR, G. M. I. **Modelos mistos aditivos semiparamétricos de contornos elípticos**. Tese (Doutorado) — Universidade de São Paulo, 2009.

REINSCH, C. Smoothing by spline functions. **Numer. Math**, v. 10, p. 177–183, 1967.

RUPPERT, D. Selecting the number of knots for penalized splines. **Journal of computational and graphical statistics**, Taylor & Francis, v. 11, n. 4, p. 735–757, 2002.

SCHMIDT, G.; MATTERN, R.; SCHÜLER, F. Biomechanical investigation to determine physical and traumatological differentiation criteria for the maximum load capacity of head and vertebral column with and without protective helmet under effects of impact. **Final report phase III, Project**, v. 65, 1981.

SCHWARZ, G. *et al.* Estimating the dimension of a model. **The annals of statistics**, Institute of Mathematical Statistics, v. 6, n. 2, p. 461–464, 1978.

SOUZA, C. P. E. Testes de hipôteses para dados funcionais baseados em distâncias: um estudo usando splines. 2008.

TEAM, R. D. C. **An Introduction to R**. [s.n.], 2019. Disponível em: <https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>.

VANEGAS, L. H.; PAULA, G. A. An extension of log-symmetric regression models: R codes and application. **Journal of Statistical Simulation and Computation**, v. 86, p. 1709–1735, 2016.

WAHBA, G. Spline models for observational data. SIAM, 1990.

WEI, B.-C.; HU, Y.-Q.; FUNG, W.-K. Generalized leverage and its applications. **Scandinavian Journal of statistics**, Wiley Online Library, v. 25, n. 1, p. 25–37, 1998.

WOOD, S. N. **Generalized Additive Models: an Introduction with R**. [S.l.]: CRC Press, 2017. v. 2.

APÊNDICE A – CÓDIGOS COMPUTACIONAIS

Neste apêndice, são apresentados alguns códigos computacionais que foram implementados, utilizando a sintaxe do R, para obtenção dos resultados expostos nessa dissertação.

CÓDIGO EM R PARA CONSTRUIR UMA SPLINE LINEAR E CÚBICA

```
1
2   pkg <- c("segmented",
3   "latticeExtra",
4   "RSADBE")
5   sapply(pkg, require,
6   character.only = TRUE)
7
8
9   data("PW_Illus")
10  data <- PW_Illus
11  names(data) <- c("x", "y")
12  data$x2 <- with(data,
13  ifelse(x <= 15, 0, x - 15))
14  data$x3 <- with(data,
15  ifelse(x <= 30, 0, x - 30))
16  m0 <- lm(y ~ x+x2+x3, data)
17  coef.m0 <- coefficients(m0)
18
19  plot(data$x,data$y, pch = 20, col = "cyan", ylab = "y", xlab = "x")
20  curve(coef.m0[1]+{coef.m0[2]*x}, add=T,
21  col = 6, lwd = 2, from = -1, to = 15)
22
23  curve(coef.m0[1]+{coef.m0[2]*x+coef.m0[3]*(x-15)},
24  col = 6, lwd = 2, add = T, from = 15, to = 30)
25  curve(coef.m0[1]+{coef.m0[2]*x+coef.m0[3]*(x-15)+coef.m0[4]*(x-30)},
26  col = 6, lwd = 2, add = T, from = 30)
```

Código-fonte 1 – Função para ajuste de uma splines linear

```
1
```

```
set.seed(19)
x = sort(round(runif(30, 0.5,3.3),3))
y = 3*cos(3+2*x) + 5  + rnorm(30, 0, 1)


data <- data.frame(y,x)
data$x2 <- with(data,
ifelse(x <= 1.5, 0, x - 1.5))

data$x3 <- with(data,
ifelse(x <= 2.5, 0, x - 2.5))

m0 <- lm(y ~ x+I(x**2)+I(x**3)+I(x2**3)+I(x3**3),
data)
coef.m0 <- coefficients(m0)

plot(x,y, pch = 20, col = "cyan", ylim = c(0,11), xlim = c(0.6,3.3))
curve(coef.m0[1]+coef.m0[2]*x+coef.m0[3]*x**2+{coef.m0[4]*x**3},
col = 6,
lwd = 2, add = T, from = 0, to = 1.5)

curve(coef.m0[1]+coef.m0[2]*x+coef.m0[3]*x**2+{coef.m0[4]*x**3+coef.m0
    [5]*(x-2)**3},
col = 6,
lwd = 2, add = T, from = 1.5, to = 2.5)

curve(coef.m0[1]+coef.m0[2]*x+coef.m0[3]*x**2+{coef.m0[4]*x**3+
coef.m0[5]*(x-2)**3+coef.m0[6]*(x-2.5)**3},
col = 6,
lwd = 2, add = T, from = 2.5)
abline(v=1.5, lty = 2)
abline(v=2.5, lty = 2)

m1 <- lm(y ~ x+x2+x3,
data)
coef.m1 <- coefficients(m1)

curve(coef.m1[1]+{coef.m1[2]*x}, add=T,
col = "green",
```

```
40   lwd = 2, from = 0, to = 1.5)
41
42   curve(coef.m1[1]+{coef.m1[2]*x+coef.m1[3]*(x-1.5)},
43   col = "green",
44   lwd = 2, add = T, from = 1.5, to = 2.5)
45   curve(coef.m1[1]+{coef.m1[2]*x+coef.m1[3]*(x-1.5)+coef.m1[4]*(x-2.5)},
46   col = "green",
47   lwd = 2, add = T, from = 2.5)
48
49   legend(2.4,10,legend = c("spline linear","spline cubica"),
50   bty = "n",lwd = 2, lty=c(1,1), col =  c("green",6))
```

Código-fonte 2 – Função para ajuste de uma splines cúbica

FUNÇÃO PARA AJUSTE DE UMA FUNÇÃO NÃO PARAMÉTRICA

Exemplo 1: Impacto do capacete da motocicleta

```
################## Função para ajuste do modelo ############################
#                                                                          #
# ------ Função para estimar as quantidades de um modelo adtivo ---------#
#       generalizado para uma única função não paramétrica                 #
#                                                                          #
############################################################################
# ========================================================================= #
#                                                                          #
#                          DETALHES DA FUNÇÃO                              #
#                                                                          #
#   A função recebe 4 objetos de entrada:                                  #
#                                                                          #
#   1 - A variável explicativa                                             #
```

```
#   2 - A variável resposta                                              #
#   3 - Um vetor contendo os nós                                         #
#   4 - Um parâmetro desuavização                                        #
#                                                                        #
#     ------- (NESSA ORDEM) --------                                     #
#   Obs.: Note que esta função é facilmente estendida para ajustes       #
#         considerando mais de uma função não paramétrica                #
#                                                                        #
# ====================================================================== #
```

```r
library(MASS)
library(mosaic)
data(mcycle)
attach(mcycle)

# Funcao para o ajuste do modelo
psfit <- function(x,y,pord=2,k,alpha=1){
X <- splineDesign(knots = k, x) # Matriz de base B-spline
nb <- ncol(X)
P <- diff(diag(nb),differences=pord)
MatC <- rep(1, nrow(X)) %*% X
qrc <- qr(t(MatC))
Z <- qr.Q(qrc,complete=TRUE)[,(nrow(MatC)+1):ncol(MatC)]
B <- cbind(1,X %*% Z) # Nova matriz modelo

# Matriz de penalidades
Matsqrt <- function(S)
{
d <- eigen(S,symmetric=TRUE)
rS <- d$vectors %*% abs(diag(d$values))^0.5 %*% t(d$vectors)
return(rS)
}

S <- t(Z) %*% t(P) %*% P %*% Z
rS <- Matsqrt(S) * sqrt(alpha/16)
O <- matrix(0,1,1)
rS <- bdiag(O,rS)
# Ajuste
```

```
  X1 <- as.matrix(rbind(B,rS))
  k <- ncol(B)
  n <- nrow(B)
  y1 <- y;y1[(n+1):(n+k)]<-0
  f <- glm(y1~X1-1,family = gaussian(link = "identity"))
  h <- hat(f$qr)[1:n]
  mu <- fitted.values(f)[1:n]
  # Validacao Cruzada e dispersao

  trH <- sum(h)
  rss <- sum((y-mu)^2/f$family$variance(mu)[1:n])
  gcv <- n*rss/(n-trH)^2
  sigma <- sqrt(rss / (n - trH))

  output <- list(gcv=gcv,sigma=sigma,f=f,trH=trH, S=S, B=B)
  return(output)
  }
  knots <- c(-12.739855,-7.711636,-2.683418,2.344800,7.373018,
             12.401236,17.429455,22.457673,27.485891,32.514109,
               37.542327,42.570545,47.598764,52.626982,57.655200,
             62.683418,67.711636,72.739855)
  fit <- psfit(times,accel,k=knots, alpha = 1)
```

Código-fonte 3 – Função para ajuste do modelo

```
############################## Obtendo Alpha ###################################
#                                                                              #
#                                                                              #
# ------ Esta parte do código é utilizada para encontrar o ---------------#
#            valor do parâmetro de suavização ótimo                            #
#                                                                              #
################################################################################
```

```
lla <- seq(-3, 3, by = 0.10)
```

```
cvs <- 0 * lla
alpha<-1
for (k in 1:length(lla)) {
  alpha = 10^lla[k]
  pn =  psfit(times,accel,k=knots,alpha = alpha)
  cvs[k] = pn$gcv
}
# alpha escolhido por cv generalizado
alpha.cv <- 10^(lla[which.min(cvs)])
```

Código-fonte 4 – Obtenção de alpha

```
#Menor VCG
cvs[which.min(cvs)]
plot(1:length(lla),cvs,type="l",main="Pontuacao VCG",
                                xlab="i", ylab = "vcs")
legend("topleft",legend = "Menor VCG  = 560.077",bty = "n")
```

Código-fonte 5 – Plot - Pontuação VCG

```
################## Estimativas do modelo ótimo ##########################
#                                                                       #
#      Esta parte do código é usada para obtenção do ajuste baseado no  #
#      valor ótimo de alpha, bem como obter outras medidas importantes,  #
#      tais como: Matriz de variância e covariância, intervalos de con-  #
#      fiança, dentre outras.                                           #
#                                                                       #
#########################################################################
```

```
# Ajuste baseado no alpha otimo
fit.cv <- psfit(times,accel,k=knots,lambda=lam.cv)
# Graus de liberdade efetivos
df<-fit.cv$trH
# medidas para obter os erros padrao e intervalos de confianca
n<-length(accel)
mu <- fitted.values(fit.cv$f)[1:n]
eta <- fit.cv$f$family$linkinv(mu)[1:n]
D <- diag(fit.cv$f[]$family$mu.eta(eta)[1:n],n)
O <- matrix(0,1,1)
S<-fit.cv$S
K <- bdiag(lam.cv*S);K<-bdiag(O,K)
W<-diag(fit.cv$f$weights[1:n],n)
A<-solve(t(fit.cv$B) %*% W %*% fit.cv$B + K)*fit.cv$sigma^2
Variace<-fit.cv$B %*% A %*% t(fit.cv$B)
pred.int <- fit.cv$B %*% coef(fit.cv$f)
upr <- pred.int + (2 * sqrt(diag(Variace)))
lwr <- pred.int - (2 * sqrt(diag(Variace)))
```

```
plot(times,accel,lwd=2,lty = 1,
col = "gray80", ann = F, ylim = c(-150,100))
lines(fitted(fit.cv$f)[order(times)]~times[order(times)], col= 1, lwd=1)
lines(upr~times, lty = 2, col = 1)
lines(lwr~times, lty = 2, col = 1)
mtext(side = 1,text = "Tempo (ms)",line=3,cex = 1.2)
mtext(side = 2,text = "Aceleracao (g)",line=3,cex = 1.2)
fit.cv$sigma^2
title(expression(hat(sigma)^2==514.74))
cor = c(1,1)
legend("topright",legend = c("Intervalo de Confiaca","Curva Ajustada"),
                              bty = "n",lwd = 2, lty=c(2,1), col =  cor)
legend("topleft",legend = expression(alpha==0.631),bty = "n")

erros_padrao<-sqrt(diag(A))

t_value<-coef(fit.cv$f)[1]/erros_padrao[1]
```

Código-fonte 6 – Estimativas do modelo ótimo

```
library(ISLR)
library(splines)
library(mgcv)
library (splines)
library(mosaicCalc)
library(mosaic)
library(MASS)
library(matlib)


gaminfluence.diag <- function(fit){
y <- fit$y
w <- fit$weights
mu <- fitted.values(fit)
eta <- fit$family$linkfun(mu)
lambda <- fit$sp
family <- fit$family
#Matriz modelo
```

```
19 X<-model.matrix(fit)
20 # quantidade de termos suavizados
21 d <- length(fit$sp)
22 # quantidades de parametros na parte parametrica do modelo
23 p <-length(summary(fit)$p.coeff)
24 n<-length(y)
25 #derivada da funcao de variacia
26
27 deriv.varience <- D(family$variance(mu)~mu)
28
29 # Derivada de g(eta)
30 # detalhes - mu.eta e a derivada da inversa da funcao de ligacao:
31 #               mu = ginv(eta) e mu.eta = d(ginv(eta))/d(eta) = d(mu)/d(
       eta)
32
33 deriv.mu.eta <- D(family$mu.eta(eta)~eta)
34
35 S<-D<-list()
36 for (i in 1:d){
37 #--------------- Matriz S para a penalizacao de cada termo suavisado
       ------------------------
38 D[i]<-fit$smooth[[1]]$S
39 S[[i]] <- lambda[i]*D[[i]]
40 }
41
42
43 #----------------------------  Matriz O de zeros
       ----------------------------------------
44 O <- matrix(0,p,p)
45
46
47 #------------------------------- Matriz K de penalizacao
       ---------------------------------
48
49 K <- bdiag(S);K<-bdiag(O,K)
50
51
52 #-------------------------------- Matriz L
       ------------------------------------------------
```

```
53 D <- diag(family$mu.eta(eta),n)
54 W <- diag(w,n)
55 V <- diag(family$variance(mu),n)
56 dV <- diag(deriv.varience(mu),n)
57 dD <- diag(deriv.mu.eta(eta),n)
58 C <- diag(y - mu,n)
59 M1 <- (D %*% ginv(dV) %*% D + dD %*% ginv(V)) %*% C
60 M2 <- -W
61 L <- t(X) %*% (M1+M2) %*% X - K;L<- as.matrix(L)
62 # ------------------------------ Pertubarcao de caso
       -------------------------------------
63 Deltac <- t(X) %*% W %*% ginv(D) %*% C
64 Fc<- -t(Deltac) %*% ginv(L) %*% Deltac # Matriz F
65 Cic<-2*abs(diag(Fc))
66 deFc<-eigen(Fc)
67 Cdmaxc<-deFc$val[1]  #maior autovalor
68 dmaxc<-deFc$vec[,1] #maior autovetor associado
69 dmaxc<-dmaxc/sqrt(Cdmaxc)
70 dmaxc<-abs(dmaxc)
71 #curvatura normal conformal
72 Bc<- abs(diag(Fc)/sqrt(sum(deFc$values^2)))
73
74
75 #-------------------- Perturbacao na variavel Resposta
       -------------------------------------
76 Deltar <- t(X) %*% W %*% ginv(D) %*% sqrt(V)
77 Fr<- -t(Deltar) %*% ginv(L) %*% Deltar # Matriz F
78 Cir<-2*abs(diag(Fr))
79 deFr<-eigen(Fr)
80 Cdmaxr<-deFr$val[1]  #maior autovalor
81 dmaxr<-deFr$ver[,1] #maior autovetor associado
82 dmaxr<-dmaxr/sqrt(Cdmaxr)
83 dmaxr<-abs(dmaxr)
84 #curvatura normal conformal
85 Br<- abs(diag(Fr)/sqrt(sum(deFr$values^2)))
86
87 list(Fc=Fc,Cic = Cic, Cdmaxc = Cdmaxc,dmaxc=dmaxc,Bc=Bc,Fr=Fr,Cir = Cir,
       Cdmaxr = Cdmaxr,dmaxr=dmaxr, Br=Br)
88 }
```

Código-fonte 7 – Função para análise de influência local

ANEXO A – DERIVAÇÃO DOS ESQUEMAS DE PERTURBAÇÃO

Os métodos de diagnósticos por meio da análise de influência local consistem nos seguintes esquemas de perturbação: casos ponderados, perturbação na variável resposta e perturbação de uma covariável. Portanto, considere que $\tilde{\boldsymbol{\xi}} = (\boldsymbol{\beta}, \boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \ldots, \boldsymbol{\tau}_r)^\top \in \Xi \subseteq \mathbb{R}^s_v$, em que $s_v = p + \sum_{k=1}^r v_k$ e $v_k - q_k - 1$. Isto posto, sob a suposição de independência, a função de log-verossimilhança penalizada será dada por:

$$\mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}) = \mathrm{L}(\tilde{\boldsymbol{\xi}}) + \frac{1}{2}\sum_{k=1}^{r} \alpha_k \boldsymbol{\tau}_k^\top \overline{\boldsymbol{S}}_k \boldsymbol{\tau}_k, \tag{A.1}$$

em que $\mathrm{L}(\tilde{\boldsymbol{\xi}}) = \sum_{i=1}^{n} \left\{\phi^{-1}[y_i\theta_i - b(\theta_i)] + c(y_i;\phi)\right\}$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_r)^\top$ é um vetor de parâmetros de suavização, de forma que $\alpha_k > 0$, $\boldsymbol{\tau}_k$ são os coeficientes dos B-*splines* da k-ésima função não paramétrica centralizada, $\overline{\boldsymbol{S}}_k = \boldsymbol{Z}_k^\top \boldsymbol{P}_k^d \boldsymbol{Z}_k$, $\boldsymbol{P}_k^d = (\Delta_k^d)^\top (\Delta_k^d)$ e Δ_k^d é um operador de diferenças, tal que $\Delta_k^d \gamma_{kl} = \Delta_k \left(\Delta_k^{d-1}\gamma_{kl}\right)$ e $\Delta_k \gamma_{kl} = (\gamma_{kl} - \gamma_{k(l-1)})$, sendo d a ordem da diferença.

Para fazer uso da teoria de influência local, devemos calcular a matriz de informação observada, $\ddot{\mathcal{F}}_{\boldsymbol{\alpha}}$. Para tanto, considerando $\boldsymbol{\alpha}$ e ϕ fixados, tem-se que $\ddot{\mathcal{F}}_{\boldsymbol{\alpha}}$ é definida como:

$$\ddot{\mathcal{F}}_{\boldsymbol{\alpha}} = \frac{\partial^2 \mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega}_0)}{\partial\tilde{\boldsymbol{\xi}}\partial\tilde{\boldsymbol{\xi}}^\top} = \begin{pmatrix} \ddot{\mathcal{F}}_{\boldsymbol{\alpha}_{\boldsymbol{\beta}\boldsymbol{\beta}}} & \ddot{\mathcal{F}}_{\boldsymbol{\alpha}_{\boldsymbol{\beta}\boldsymbol{\tau}_k}} \\ \ddot{\mathcal{F}}_{\boldsymbol{\alpha}_{\boldsymbol{\tau}_k\boldsymbol{\beta}}} & \ddot{\mathcal{F}}_{\boldsymbol{\alpha}_{\boldsymbol{\tau}_k\boldsymbol{\tau}_k}} \end{pmatrix}, \tag{A.2}$$

em que

$$\begin{aligned}
\ddot{\mathcal{F}}_{\boldsymbol{\alpha}_{\boldsymbol{\beta}\boldsymbol{\beta}}} = \frac{\partial^2 \mathcal{F}_{\boldsymbol{\alpha}}(\boldsymbol{\xi}|\boldsymbol{\omega}_0)}{\partial\beta_j\partial\beta_l} &= \phi^{-1}\sum_{i=1}^{n}(y_i-\mu_i)\frac{d^2\theta_i}{d\mu_i^2}\left(\frac{d\mu_i}{d\eta_i}\right)^2 x_{ij}x_{il} \\
&\quad + \phi^{-1}\sum_{i=1}^{n}(y_i-\mu_i)\frac{d\theta_i}{d\mu_i}\frac{d^2\mu_i}{d\eta_i^2}x_{ij}x_{il} \\
&\quad - \phi^{-1}\sum_{i=1}^{n}\frac{d\theta_i}{d\mu_i}\left(\frac{d\mu_i}{d\eta_i}\right)^2 x_{ij}x_{il}, \\
\ddot{\mathcal{F}}_{\boldsymbol{\alpha}_{\boldsymbol{\tau}_k\boldsymbol{\tau}_k}} = \frac{\partial^2 \mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega}_0)}{\partial\tau_{jk}\partial\tau_{kl}} &= \phi^{-1}\sum_{i=1}^{n}(y_i-\mu_i)\frac{d^2\theta_i}{d\mu_i^2}\left(\frac{d\mu_i}{d\eta_i}\right)^2 \tilde{b}_{ijk}\tilde{b}_{ilk} \\
&\quad + \phi^{-1}\sum_{i=1}^{n}(y_i-\mu_i)\frac{d\theta_i}{d\mu_i}\frac{d^2\mu_i}{d\eta_i^2}\tilde{b}_{ijk}\tilde{b}_{ilk} \\
&\quad - \phi^{-1}\sum_{i=1}^{n}\left\{\frac{d\theta_i}{d\mu_i}\left(\frac{d\mu_i}{d\eta_i}\right)^2 \tilde{b}_{ijk}\tilde{b}_{ilk} - \alpha_k\bar{s}_{ijk}\right\}, \quad \text{para} \quad j,l = 1,2,...,v_k. \\
\ddot{\mathcal{F}}_{\boldsymbol{\alpha}_{\boldsymbol{\beta}\boldsymbol{\tau}_k}} = \frac{\partial^2 \mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega}_0)}{\partial\beta_j\partial\tau_{kl}} &= -\phi^{-1}\sum_{i=1}^{n}(y_i-\mu_i)\frac{d^2\theta_i}{d\mu_i^2}\left(\frac{d\mu_i}{d\eta_i}\right)^2 x_{ij}\tilde{b}_{ilk} \\
&\quad - \phi^{-1}\sum_{i=1}^{n}(y_i-\mu_i)\frac{d\theta_i}{d\mu_i}\frac{d^2\mu_i}{d\eta_i^2}x_{ij}\tilde{b}_{ilk}
\end{aligned}$$

$$+\phi^{-1}\sum_{i=1}^{n}\frac{d\theta_i}{d\mu_i}\left(\frac{d\mu_i}{d\eta_i}\right)^2 x_{ij}\tilde{b}_{ilk},$$

$\ddot{\mathcal{F}}_{\boldsymbol{\alpha}\boldsymbol{\beta}\boldsymbol{\tau}_k} = \ddot{\mathcal{F}}^{\top}_{\boldsymbol{\alpha}\boldsymbol{\tau}_k\boldsymbol{\beta}}$ no qual $\mu_i = g^{-1}(\eta_i)$, $\eta_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \tilde{\boldsymbol{b}}_{1i}^{\top}\boldsymbol{\tau}_1 + \tilde{\boldsymbol{b}}_{2i}^{\top}\boldsymbol{\tau}_2 + \cdots, \tilde{\boldsymbol{b}}_{ri}^{\top}\boldsymbol{\tau}_r$. Escrevendo em foma de matriz, temos:

$$\ddot{\mathcal{F}}_{\boldsymbol{\alpha}} = \frac{\partial\Psi_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial\tilde{\boldsymbol{\xi}}^{\top}} = \widetilde{\boldsymbol{X}}^{*\top}[\boldsymbol{M}_1 + \boldsymbol{M}_2]\widetilde{\boldsymbol{X}}^{*} - \widetilde{\boldsymbol{K}}^{*}(\boldsymbol{\alpha}), \tag{A.3}$$

sob a condição de $\boldsymbol{M}_1 = [\boldsymbol{D}\dot{\mathbf{V}}^{-1}\boldsymbol{D} + \dot{\mathbf{D}}\boldsymbol{V}^{-1}]\boldsymbol{C}$, $\dot{\mathbf{D}} = \text{diag}\left\{\partial^2\mu_1/\partial\eta_1^2, \ldots, \partial^2\mu_n/\partial\eta_n^2\right\}$, bem como $\dot{\mathbf{V}} = \text{diag}\left\{\partial V_1/\partial\mu_1, \ldots \partial V_n/\partial\mu_n\right\}$, $\boldsymbol{C} = \text{diag}\left\{C_1, \ldots C_n\right\}$, no qual $C_i = y_i - \mu_i$, $\boldsymbol{M}_2 = -\boldsymbol{W}$ e $\widetilde{\boldsymbol{K}}^{*}(\boldsymbol{\alpha}) = \text{blocodiag}\left\{\boldsymbol{O}_{pp}, \overline{\boldsymbol{S}}(\boldsymbol{\alpha})\right\}$ avaliados em $\tilde{\boldsymbol{\xi}} = \widehat{\tilde{\boldsymbol{\xi}}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$.

CASO PONDERADO

Para avaliar a influência das perturbações de casos, o logaritmo da função de verossimilhança perturbada é definido por:

$$\mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}) = \text{L}(\tilde{\boldsymbol{\xi}}) + \frac{1}{2}\sum_{k=1}^{r}\alpha_k\boldsymbol{\tau}_k^{\top}\overline{\boldsymbol{S}}_k\boldsymbol{\tau}_k, \tag{A.4}$$

Contudo, perturbamos $\text{L}(\tilde{\boldsymbol{\xi}})$ de tal forma que $\text{L}(\tilde{\boldsymbol{\xi}}) = \sum_{i=1}^{n}\omega_i\left\{\phi^{-1}[y_i\theta_i - b(\theta_i)] + c(y_i;\phi)\right\}$. Aqui, objetivo é identificar pontos influentes entre todas as observações. Neste caso, o vetor correspondente à não perturbação é o vetor $\boldsymbol{\omega}_0 = (1,1,\ldots,1)^T$ n-dimensional. Note que $\boldsymbol{\Delta} = \left(\boldsymbol{\Delta}_{\boldsymbol{\beta}}^{\top}, \boldsymbol{\Delta}_{\boldsymbol{\tau}_1}^{\top}, \boldsymbol{\Delta}_{\boldsymbol{\tau}_2}^{\top}, \cdots, \boldsymbol{\Delta}_{\boldsymbol{\tau}_k}^{\top}\right)^{\top}$ é uma matriz $(p+s_v)\times n$, avaliada em $\tilde{\boldsymbol{\xi}} = \widehat{\tilde{\boldsymbol{\xi}}}$ e em $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, onde:

$$\begin{aligned}
\frac{\partial\mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial\beta_j} &= \phi^{-1}\sum_{i=1}^{n}\omega_i\left\{y_i\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{\partial\eta_i}{\partial\beta_j} - \frac{db(\theta_i)}{d\theta_i}\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{\partial\eta_i}{\partial\beta_j}\right\} \\
&= \phi^{-1}\sum_{i=1}^{n}\omega_i\left\{\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}(y_i - \mu_i)x_{ij}\right\} \quad \text{e, portanto,} \\
\frac{\partial^2\mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial\beta_j\partial\omega_i} &= \frac{\partial}{\partial\omega_i}\left[\phi^{-1}\sum_{i=1}^{n}\omega_i\left\{\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}(y_i - \mu_i)x_{ij}\right\}\right] \\
&= \phi^{-1}\sum_{i=1}^{n}\left\{\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}(y_i - \mu_i)x_{ij}\right\}.
\end{aligned}$$

Da mesma forma, tem-se que:

$$\begin{aligned}
\frac{\partial\mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial\tau_{kj}} &= \phi^{-1}\sum_{i=1}^{n}\left\{y_i\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{\partial\eta_i}{\partial\gamma_{kj}} - \frac{db(\theta_i)}{d\theta_i}\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{\partial\eta_i}{\partial\gamma_{kj}}\right\} + \frac{1}{2}\sum_{k=1}^{r}\alpha_k\overline{\boldsymbol{S}}_k\boldsymbol{\tau}_k \\
&= \phi^{-1}\sum_{i=1}^{n}\omega_i\left\{\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}(y_i - \mu_i)\tilde{b}_{kij}\right\} + \frac{1}{2}\sum_{k=1}^{r}\alpha_k\overline{\boldsymbol{S}}_k\boldsymbol{\tau}_k
\end{aligned}$$

e, portanto,

$$\frac{\partial^2 \mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial \tau_{kj} \partial \omega_i} = \frac{\partial}{\partial \omega_i}\left[\phi^{-1}\sum_{i=1}^{n}\omega_i\left\{\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}(y_i-\mu_i)\tilde{b}_{kij}\right\} + \frac{1}{2}\sum_{k=1}^{r}\alpha_k\overline{\boldsymbol{S}}_k\boldsymbol{\tau}_k\right]$$
$$= \phi^{-1}\sum_{i=1}^{n}\left\{\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}(y_i-\mu_i)\tilde{b}_{kij}\right\}.$$

Colocando na forma matricial, obtemos

$$\boldsymbol{\Delta}_{\boldsymbol{\beta}} = \frac{\partial^2 \mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top}\phi^{-1}\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{D}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}), \tag{A.5}$$

$$\boldsymbol{\Delta}_{\boldsymbol{\tau}_k} = \frac{\partial^2 \mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial \boldsymbol{\tau}_k \partial \boldsymbol{\omega}^\top} = \phi^{-1}\widetilde{\boldsymbol{B}}_k^\top\boldsymbol{W}\boldsymbol{D}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}), \tag{A.6}$$

em que $\widetilde{\boldsymbol{B}}_k$ é uma matriz de dimensão $(n \times v_k)$ formada pelos componentes $B_k^{(m_k)}(t_{ik})$ da base de De Boor (1978) após ser centralizada, $i = 1,2,\ldots,n$ e $j = 1,2,\ldots,v_k$.

ESQUEMA DE PERTURBAÇÃO NA VARIÁVEL RESPOSTA

Para este caso, considera-se que a variável y_i é perturbada de tal maneira que $y_{\omega_i} = y_i + V_i^{1/2}$ em que o vetor de não perturbação $\boldsymbol{\omega}_0$ é tal que $\omega_i = 0\ \forall i\ i = 1,2,\ldots,n$. Assim como no caso anterior, $\boldsymbol{\Delta} = \left(\boldsymbol{\Delta}_{\boldsymbol{\beta}}^\top, \boldsymbol{\Delta}_{\boldsymbol{\tau}_1}^\top, \boldsymbol{\Delta}_{\boldsymbol{\tau}_2}^\top, \cdots, \boldsymbol{\Delta}_{\boldsymbol{\tau}_k}^\top\right)^\top$, porém

$$\frac{\partial \mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial \beta_j} = \phi^{-1}\sum_{i=1}^{n}\left\{y_{\omega_i}\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{\partial \eta_i}{\partial \beta_j} - \frac{db(\theta_i)}{d\theta_i}\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{\partial \eta_i}{\partial \beta_j}\right\}$$
$$= \phi^{-1}\sum_{i=1}^{n}\left\{\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}(y_{\omega_i}-\mu_i)x_{ij}\right\} \quad \text{e, portanto,}$$
$$\frac{\partial^2 \mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial \beta_j \partial \omega_i} = \frac{\partial}{\partial \omega_i}\left[\phi^{-1}\sum_{i=1}^{n}\left\{\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}(y_{\omega_i}-\mu_i)x_{ij}\right\}\right]$$
$$= \phi^{-1}\sum_{i=1}^{n}\left\{\frac{\partial}{\partial \omega_i}\left[\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\right](y_{\omega_i}-\mu_i)x_{ij}\right.$$
$$\left. + \frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{\partial}{\partial \omega_i}\left[(y_{\omega_i}-\mu_i)x_{ij}\right]\right\}$$
$$= \phi^{-1}\sum_{i=1}^{n}\left\{\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}V_i^{1/2}x_{ij}\right\}.$$

Da mesma forma, tem-se que:

$$\frac{\partial \mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial \tau_{kj}} = \phi^{-1}\sum_{i=1}^{n}\left\{y_{\omega_i}\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{\partial \eta_i}{\partial \tau_{kj}} - \frac{db(\theta_i)}{d\theta_i}\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{\partial \eta_i}{\partial \tau_{kj}}\right\} + \frac{1}{2}\sum_{k=1}^{r}\alpha_k\overline{\boldsymbol{S}}_k\boldsymbol{\tau}_k$$
$$= \phi^{-1}\sum_{i=1}^{n}\left\{\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}(y_{\omega_i}-\mu_i)\tilde{b}_{kij}\right\} + \frac{1}{2}\sum_{k=1}^{r}\alpha_k\overline{\boldsymbol{S}}_k\boldsymbol{\tau}_k$$

e, portanto,

$$\frac{\partial^2 \mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial \tau_{kj} \partial \omega_i} = \frac{\partial}{\partial \omega_i}\left[\phi^{-1}\sum_{i=1}^{n}\left\{\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}(y_{\omega_i}-\mu_i)\tilde{b}_{kij}\right\} + \frac{1}{2}\sum_{k=1}^{r}\alpha_k \overline{\boldsymbol{S}}_k \boldsymbol{\tau}_k\right]$$
$$= \phi^{-1}\sum_{i=1}^{n}\left\{\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}V_i^{1/2}\tilde{b}_{kij}\right\}.$$

Considerando que $\tilde{\boldsymbol{\xi}} = \widehat{\tilde{\boldsymbol{\xi}}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, temos que, em forma matricial

$$\boldsymbol{\Delta}_{\boldsymbol{\beta}} = \frac{\partial^2 \mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial \beta \partial \omega_i} = \phi^{-1}\boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{D}^{-1}\boldsymbol{V}^{1/2} \tag{A.7}$$

$$\boldsymbol{\Delta}_{\boldsymbol{\tau}_k} = \frac{\partial^2 \mathcal{F}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial \tau_{kj} \partial \omega_i} = \phi^{-1}\widetilde{\boldsymbol{B}}_k^\top \boldsymbol{W}\boldsymbol{D}^{-1}\boldsymbol{V}^{1/2}. \tag{A.8}$$