



UNIVERSIDADE  
ESTADUAL DE LONDRINA

---

FELINTO JUNIOR DA COSTA

**GAMLSS ESPAÇO TEMPORAL PARA ENGENHARIA DE  
AVALIAÇÕES**

---

Londrina

2022

FELINTO JUNIOR DA COSTA

## **GAMLSS ESPAÇO TEMPORAL PARA ENGENHARIA DE AVALIAÇÕES**

Fascículo complementar à dissertação apresentada ao Departamento de Matemática da Universidade Estadual de Londrina, como requisito parcial para a obtenção do Título de MESTRE em Matemática Aplicada e Computacional.

Nos apêndices que se seguem são apresentados temas correlatos e desenvolvimentos complementares que, por razão de concisão, foram destacados do corpo principal dessa dissertação.

Orientador: Prof. Dr. Rodrigo Rossetto Pescim

Londrina

2022

## LISTA DE FIGURAS

B.1	Modelo de regressão linear simples com distribuição Normal para o erro $\mathbb{E}(\varepsilon_i) = 0; \text{Var}(\varepsilon_i) = \sigma^2$ . . . . .	39
B.2	Modelo linear mal ajustado processo simulado de geração da resposta por uma relação não linear $y = 0,75 + 12,50x^3 + e$ . . . . .	48
B.3	Modelo linear mal ajustado processo de geração simulada dos erros como uma função $e \sim N(0, \sigma = 3,50 + 2,50x^3)$ . . . . .	49
B.4	Relação de dependências dos elementos do algoritmo de estimação . . . . .	66
B.5	Fluxograma do algoritmo de estimação . . . . .	67
B.6	Envelope construído por simulação . . . . .	77
B.7	Regressão polinomial . . . . .	79
B.8	Diferentes regressões locais ( <i>LOWESS</i> : polinômio de grau 1) em função da amplitude da janela . . . . .	81
B.9	Diferentes regressões locais ( <i>LOESS</i> : polinômio de grau 2) em função da amplitude da janela . . . . .	82
B.10	Regressão polinomial linear por partes sem nenhuma restrição relacionada à continuidade nos nós (arbitrados em: 5,10,20,30,40) . . . . .	82
B.11	Regressão polinomial linear por partes com restrição de continuidade nos nós (arbitrados em: 5,10,20,30,40) . . . . .	84
B.12	Função suavizadora resultante da combinação linear das funções de uma base de grau 2, definidas em todo o intervalo ( $f(\cdot) = 0,15b_1 - 0,25b_2 + 0,35b_3$ ) . . .	86
B.13	Função suavizadora resultante da combinação linear das funções de base de grau 3, definidas em todo o intervalo ( $f(\cdot) = 0,15b_1 - 0,25b_2 + 0,35b_3 + 0,45b_4$ ) . .	86
B.14	Relação de recorrência de cada função da base (algoritmo de Cox-De Boor) . . . . .	88
B.15	Funções da base: relação de grau e suporte associado a cada função (expandida até $m=4$ ) . . . . .	89
B.16	Diagrama do ciclo externo do algoritmo RS de maximização do logaritmo da função de verossimilhança (penalizada) . . . . .	104
B.17	Diagrama do ciclo interno do algoritmo RS de maximização do logaritmo da função de verossimilhança (penalizada) . . . . .	108
B.18	Diagrama do algoritmo de retroajuste modificado do algoritmo RS de maximização do logaritmo da função de verossimilhança (penalizada) . . . . .	109

B.19 Interconexão (parcial) dos vários <i>loops</i> internos do algoritmo RS de maximização do logaritmo da função de verossimilhança (penalizada) . . . . .	110
B.20 Esquema do algoritmo RS de estimação dos modelos <i>GAMLSS</i> . . . . .	111
B.21 Iterações internas e externas no algoritmo CG . . . . .	112
C.1 Primeira construção em Londrina . . . . .	115

## LISTA DE TABELAS

A.1	Grau de fundamentação no caso de utilização de modelos de regressão linear . .	33
A.2	Enquadramento do laudo segundo seu grau de fundamentação no caso de utilização de modelos de regressão linear . . . . .	34
A.3	Grau de precisão nos casos de utilização de modelos de regressão linear ou do tratamento por fatores . . . . .	34
B.1	Ligações canônicas para Modelos Lineares Generalizados . . . . .	63
B.2	Distribuições contínuas disponíveis na implementação computacional dos modelos <i>GAMLSS</i> . . . . .	99
B.2	Distribuições contínuas disponíveis na implementação computacional dos modelos <i>GAMLSS</i> . . . . .	100
B.2	Distribuições contínuas disponíveis na implementação computacional dos modelos <i>GAMLSS</i> . . . . .	101
B.3	Distribuições discretas disponíveis na implementação computacional dos <i>GAMLSS</i> . . . . .	102
B.4	Distribuições mistas disponíveis na implementação computacional dos <i>GAMLSS</i>	102
B.5	Funções de ligação disponíveis na implementação computacional dos <i>GAMLSS</i>	103

## LISTA DE SIGLAS, ACRÔNIMOS E EXPRESSÕES ESTRANGEIRAS

ABNT Associação Brasileira de Normas Técnicas

AIC *Akaike Information Criteria*: Critério de Informação de Akaike (do inglês)

*Box-plot* Gráfico de caixa (do inglês)

BCT Distribuição Box&Cox “t” de probabilidade (a Distribuição Box&Cox “t” original difere apenas pela função de ligação no preditor linear do primeiro parâmetro distribucional de um *GAMLSS* na implementação computacional)

*Contour map* Mapa de Contornos ou Isolinhas (do inglês)

BIC *Bayesian Information Criteria*: Critério de Informação Bayesiano (do inglês)

CONFEA Conselho Federal de Engenharia e Agronomia

CREA Conselho Regional de Engenharia e Agronomia

CV *Cross validation*: Validação cruzada (do inglês)

Deviance Afastamento de um padrão referencial de comportamento ou estado (do inglês)

Distributional regression Regressão distribucional (do inglês)

EDF *Effective degrees of freedom*: Graus de liberdade efetivos (do inglês)

EPSG *European Petroleum Survey Group* foi uma organização científica ligada à indústria petrolífera europeia

EMV Estimativa de máxima verossimilhança

EQM Erro quadrático médio

FDP Função densidade de probabilidade

FCP Função de distribuição acumulada de probabilidade

GAM *Generalized Additive Model*: Modelo Aditivo Generalizado (do inglês)

GAMLSS *Generalized Additive Model for Location, Scale and Shape*: Modelo Aditivo Generalizado para Localização, Escala e Forma (do inglês)

GB2 Distribuição Beta Generalizada 2 (do inglês: *generalized beta 2*)

GDEV *Global deviance*: Desvio global (do inglês)

- GACI *Generalized Akaike information criteria*: critério de Akaike generalizado (do inglês)
- GLM *Generalized Linear Model*: Modelo Linear Generalizado (do inglês)
- Gray scale map* Mapa em escala de tons de cinza (do inglês)
- IBGE Instituto Brasileiro de Geografia e Estatística
- LM *Linear Model*: Modelo Linear (do inglês)
- Moving window* Janela móvel (do inglês)
- NBR Norma Brasileira
- Overfitting “Sobreajuste”, posto a palavra “sobreajuste” não constar no Vocabulário Ortográfico da Língua Portuguesa: [VOLP](#) (do inglês)
- PNAD Pesquisa Nacional por Amostragem de Domicílios
- Rought Áspero, irregular; no contexto: função muito oscilante (do inglês)
- Scatterplot* Gráfico de dispersão (do inglês)
- SI Sistema Internacional de Unidades: [INMETRO](#)
- Smooth* Suave; no contexto: função com pouca oscilação (do inglês)
- Smoother* Mais suave; no contexto: função que suaviza a variabilidade observada em um gráfico de dispersão (do inglês)
- SQR Soma dos quadrados dos resíduos
- TRVG Teste da razão de verossimilhanças generalizado
- UTM (Projeção) Universal Transversa de Mercator (do inglês: *Universal Transverse Mercator*)
- Wiggly* Aspecto que assume de uma linha com muitas curvas; no contexto: função muito oscilante (do inglês)

## LISTA DE SÍMBOLOS E NOTAÇÕES ADOTADAS

- $Y; X; Z; V$  variáveis aleatórias
- $y; x; z; v$  vetores de valores observados de variáveis aleatórias
- $x_{kj}$   $j$ -ésimo vetor de valores observados para o  $k$ -ésimo parâmetro de um *GAMLSS*
- $\varepsilon$  erro de um modelo
- $\Theta$  espaço paramétrico do parâmetro de uma distribuição
- $\theta$  parâmetro de uma distribuição ( $\theta_k$  representa o  $k$ -ésimo parâmetro de um *GAMLSS*)
- $\boldsymbol{\theta}$  vetor paramétrico de uma distribuição ( $\boldsymbol{\theta}_k$  representa o  $k$ -ésimo vetor paramétrico de um *GAMLSS*)
- $\mu$  primeiro parâmetro de uma distribuição (geralmente sua localização)
- $\sigma$  segundo parâmetro de uma distribuição (geralmente sua variabilidade)
- $\nu$  terceiro parâmetro de uma distribuição (geralmente sua forma quanto à simetria)
- $\tau$  quarto parâmetro de uma distribuição (geralmente sua forma quanto à curtose)
- $\phi$  parâmetro de dispersão de um *GLM* ou *GAM*
- $\mathbf{X}_{(n,p)}$  matriz de delineamento dos efeitos fixos de  $p$  variáveis
- $\mathbf{X}_k$  matriz de delineamento dos efeitos fixos para o  $k$ -ésimo parâmetro de um *GAMLSS*
- $S_j(x)$  polinômio cúbico para o subintervalo  $[x_j, x_{j+1}]$
- $t_i$   $i$ -ésimo nó de um *spline*
- $a(.); b(.); f(.); g(.); h(.)$  funções (suavizadoras) de uma variável
- $s_{jk}(.)$   $j$ -ésima função suavizadora para o  $k$ -ésimo parâmetro de um *GAMLSS*
- $b_j(.)$   $j$ -ésima função de uma base para uma função suavizadora



- $\mathbf{B}_{(n,j)}$  matriz de delineamento dos efeitos aleatórios de uma função suavizadora unidimensional estimada por  $j$ -funções de base
- $\mathbf{Z}_{jk}$  matriz de delineamento para o  $j$ -ésimo suavizador do  $k$ -ésimo parâmetro de um *GAMLSS*
- $B_{i,j}(\cdot)$   $i$ -ésima função (de ordem  $j$ ) *B-spline* de uma base para uma função suavizadora
- $\alpha; \beta; \delta; \epsilon; \gamma$  parâmetros de modelos (coeficientes de funções de suavização)
- $\beta_k$  vetor dos parâmetros dos efeitos fixos para o  $k$ -ésimo parâmetro de um *GAMLSS*
- $\gamma_{kj}$   $j$ -ésimo vetor de efeito aleatório para o  $k$ -ésimo parâmetro de um *GAMLSS*
- $\hat{\alpha}; \hat{\beta}; \hat{\delta}; \hat{\epsilon}; \hat{\gamma}$  estimativas dos parâmetros de modelos (coeficientes de funções de suavização)
- $\alpha; \beta; \delta; \epsilon; \gamma$  vetores paramétricos de modelos (coeficientes de funções de suavização)
- $\hat{\alpha}; \hat{\beta}; \hat{\delta}; \hat{\epsilon}; \hat{\gamma}$  estimativas de vetores paramétricos de modelos (coeficientes de funções de suavização)
- $\lambda$  parâmetro que regula a quantidade de suavização no ajuste de uma função de suavização (hiperparâmetro em um *GAMLSS*)
- $\lambda$  vetor de todos os hiperparâmetros de um *GAMLSS*
- $\mathbf{D}_{n,j}$  matriz de diferenças entre dois *B-splines* adjacentes ( $\mathbf{D}_{[d]}$  indica uma matriz de diferenças de ordem  $d$ )
- $\mathbf{G} = \mathbf{D}^T \mathbf{D}; (\mathbf{S} = \mathbf{D}^T \mathbf{D})$  matrizes de penalização
- $t_i(x_i, y_i)$   $i$ -ésima observação tomada de um ponto  $(x_i, y_i)$  no plano
- $r(\|(x_i, y_i) - (x_c, y_c)\|)$  norma entre dois pontos no plano
- $\zeta(r)$  função radial aplicada à norma “ $r$ ”
- $\mathbf{T}_{(3,n)}$  matriz dos valores avaliados das funções lineares de base  $\phi$  de um *thin-plate spline* aplicadas sobre os pontos  $t_i$
- $\mathbf{E}_{(n,n)}$  matriz dos valores avaliados da função radial  $\eta$  de um *thin-plate spline* aplicada sobre as normas  $r$
- $J_{[d]}$  funcional que quantifica a oscilabilidade de uma função suavizadora  $d$ -dimensional

$g(\mathbf{t})$  *thin plate spline* de  $\mathbf{t}$

$\mathbf{o}$  vetor de zeros

$\otimes$  operador do produto Kronecker

$\odot$  operador do produto Kronecker por linha

$D$  desvio de um modelo linear generalizado

$g_k(\cdot)$  função de ligação aplicada ao  $k$ -ésimo parâmetro de uma distribuição em um GAMLSS

$\boldsymbol{\eta}_k$   $k$ -ésimo preditor linear de um GAMLSS tal que  $\boldsymbol{\eta}_k = g_k(\boldsymbol{\theta}_k)$

$L$  função de verossimilhança

$\hat{L}_c$  função de verossimilhança para o modelo em estudo

$l$  logaritmo da função de verossimilhança

$l_p$  logaritmo da função de verossimilhança penalizada

$\kappa$  penalidade para cada grau de liberdade no modelo aplicada no cálculo do GAIC

$df$  graus de liberdade (efetivos) usados em um modelo

$Q$  Estatística Q calculado sobre os resíduos de um modelo

$Z$  Estatística Z calculado sobre os resíduos de um modelo

$\mathcal{M}$  um GAMLSS estruturado como  $\{\mathcal{D}; \mathcal{G}; \mathcal{T}; \mathcal{L}\}$

$\mathcal{D}$  a distribuição especificada para de um GAMLSS

$\mathcal{G}$  as funções de ligação para os preditores dos parâmetros distribucionais de um GAMLSS

$\mathcal{T}$  os termos explicativos de um GAMLSS

$\mathcal{L}$  a especificação dos parâmetros de suavização de um GAMLSS

$\mathbf{r} = \Phi^{-1}(\cdot)$  vetor dos resíduos quantílicos Normalizados

$\Phi^{-1}(\cdot)$  função quantílica da distribuição Normal padronizada

$F(y)$  função de distribuição acumulada de probabilidade de uma variável aleatória  $y$

$f_{Y|X}(y)$  função densidade de probabilidade condicional de uma variável aleatória  $Y$  dado  $X$

$\mathcal{E}$  família exponencial das distribuições

$\mathcal{N}(\mu, \sigma^2)$  distribuição Normal com média  $\mu$  e variância  $\sigma^2$

$\mathcal{N}(0, 1)$  distribuição Normal padronizada

$\phi(.)$  função densidade de probabilidade de uma distribuição Normal padronizada

$\Phi(.)$  função de distribuição acumulada de probabilidade de uma distribuição Normal padronizada

$\bar{\beta}^*$  valor médio das replicações *bootstrap* de uma estatística qualquer

$\delta^*$  afastamentos medidos entre uma estatística qualquer e suas replicações *bootstrap*

## SUMÁRIO

<b>Lista de Figuras</b>	<b>iii</b>
<b>Lista de Tabelas</b>	<b>v</b>
<b>Lista de Siglas, Acrônimos e Expressões Estrangeiras</b>	<b>vi</b>
<b>Lista de Símbolos e Notações Adotadas</b>	<b>viii</b>
<b>Apêndices</b>	<b>xv</b>
<b>Apêndice A Engenharia de avaliações e normatização associada</b>	<b>29</b>
A.1 Exigibilidade . . . . .	29
A.2 Definições de valor . . . . .	29
A.3 Métodos . . . . .	31
A.4 Imposições e enquadramentos normativos . . . . .	32
<b>Apêndice B Revisão da literatura</b>	<b>35</b>
B.1 Modelo Clássico de Regressão Linear ( <i>LM</i> ) . . . . .	35
B.1.1 Modelo sem distribuição especificada para o erro . . . . .	35
B.1.2 Modelo com distribuição Normal para o erro . . . . .	37
B.1.3 Estimação dos parâmetros de um Modelo Clássico de Regressão Linear	38
B.1.3.1 Método dos mínimos quadrados . . . . .	38
B.1.3.2 Método da máxima verossimilhança . . . . .	40
B.1.4 Modelo Clássico de Regressão Linear com distribuição Normal para o	
erro . . . . .	43
B.1.5 Resíduos ordinários e resíduos “estudentizados” . . . . .	43
B.1.6 Adequabilidade de um Modelo Clássico de Regressão Linear com	
distribuição Normal para o erro . . . . .	45
B.1.7 Análise da especificação do modelo . . . . .	46
B.1.8 Análise da linearidade na relação entre as covariáveis e a resposta . . .	47
B.1.9 Análise da homogeneidade da variância . . . . .	47
B.1.10 Análise quanto à independência dos resíduos . . . . .	50
B.1.11 Análise quanto à distribuição dos resíduos Normalidade . . . . .	51
B.1.12 Análise quanto à existência de observações discordantes . . . . .	54

B.2	Modelos Lineares Generalizados ( <i>GLM</i> ) . . . . .	57
B.2.1	Família exponencial de distribuições uniparamétricas . . . . .	57
B.2.2	Algumas distribuições pertencentes à família exponencial com parâmetro de dispersão . . . . .	59
B.2.3	Estrutura de um Modelo Linear Generalizado . . . . .	60
B.2.4	Especificação de um Modelo Linear Generalizado . . . . .	62
B.2.5	Estimação . . . . .	63
B.2.5.1	Elementos do algoritmo . . . . .	65
B.2.5.2	Ciclo iterativo . . . . .	66
B.2.6	Desvio ( <i>deviance</i> ) . . . . .	68
B.2.7	Comparação de modelos encaixados ( <i>nested</i> ) . . . . .	69
B.2.8	Crítério de informação de Akaike (AIC) . . . . .	71
B.2.9	Algoritmos para seleção de covariáveis . . . . .	72
B.2.10	Quantidades diagnósticas para validação de um modelo . . . . .	73
B.2.11	Validação de um Modelo Linear Generalizado . . . . .	76
B.3	Funções de suavização para métodos não paramétricos . . . . .	79
B.3.1	Regressão polinomial . . . . .	79
B.3.2	Regressão polinomial local . . . . .	80
B.3.3	Polinômios por partes . . . . .	81
B.3.4	Espaço de funções . . . . .	83
B.3.5	B-Splines . . . . .	85
B.3.6	Splines penalizados . . . . .	87
B.3.7	P-Splines: definição da penalização . . . . .	90
B.3.7.1	P-Splines: estimação do parâmetro de suavização, matrizes de penalização e suavização e graus de liberdade efetivos . . .	91
B.3.8	Modelos aditivos não paramétricos: estimação por máxima verossimilhança penalizada . . . . .	94
B.3.9	Algoritmo para estimação por mínimos quadrados ponderados sob penalização . . . . .	96
B.3.10	Funcional para penalização de <i>splines</i> de produto tensor . . . . .	96
B.4	Modelos Aditivos Generalizados para Locação, Escala e Forma . . . . .	99
B.4.1	Distribuições de probabilidade implementadas para os <i>GAMLSS</i> . . . .	99
B.4.2	Funções de ligação implementadas para os <i>GAMLSS</i> . . . . .	101
B.4.3	Algoritmo RS de estimação dos <i>GAMLSS</i> . . . . .	101
B.4.3.1	Pseudocódigo do algoritmo RS . . . . .	106
<b>Apêndice C Materiais e métodos</b>		<b>113</b>
C.1	O café, os ingleses e a cidade de Londrina . . . . .	113
C.2	Georreferenciamento . . . . .	117

C.3	Renda média dos setores censitários . . . . .	118
C.4	Termos e sua forma de introdução nos preditores dos parâmetros distribucionais	119
C.5	Funções auxiliares de controle do algoritmo de estimação . . . . .	120
C.6	Criação de um <i>grid</i> para geração de imagens . . . . .	121
<b>Referências</b>		<b>123</b>

# Apêndices





## A ENGENHARIA DE AVALIAÇÕES E NORMATIZAÇÃO ASSOCIADA

### A.1 EXIGIBILIDADE

Considerando-se a natureza dos dados experimentais obtidos e a análise à qual foram submetidos, torna-se necessário verificar o atendimento às exigências das Normas Técnicas Brasileiras NBR 14.653:1(ABNT: Associação Brasileira de Normas Técnicas, 2011) e NBR 14.653:2(ABNT: Associação Brasileira de Normas Técnicas, 2011) elaboradas pelo Comitê Brasileiro de Construção Civil da Associação Brasileira de Normas Técnicas, uma vez que o Código de Defesa do Consumidor (Lei 8078, de 11.09.1990, Regulamentada pelo Decreto 861, de 09.07.1993), na Seção IV, que trata das Práticas Abusivas, e seu Artigo 39, que no inciso VIII estabelece:

“É vedado ao fornecedor de produtos e serviços colocar, no mercado de consumo, qualquer produto ou serviço em desacordo com as normas expedidas pelos órgãos oficiais competentes ou, se normas específicas não existirem, pela Associação Brasileira de Normas Técnicas – ABNT, ou outra Entidade credenciada pelo Conselho Nacional de Metrologia, Normalização e Qualidade Industrial – CONMETRO.”.

Como não há Regulamentação Técnica específica sobre um produto ou serviço e a ABNT é a única entidade reconhecida pelo CONMETRO, as Normas ABNT passam a ser a referência para a qualidade desses itens, quando comercializados no País. Não se enquadrando esse trabalho em nenhuma das situações previstas pela Lei, as verificações aqui realizadas no âmbito da Norma são por mera liberalidade do autor.

### A.2 DEFINIÇÕES DE VALOR

As diferenças conceituais referentes a preço, valor e custo estão estabelecidas no item 0.1 da norma NBR 14653-1:2019 da ABNT (ABNT: Associação Brasileira de Normas Técnicas, 2019):

“Os conceitos de valor, preço e custo são distintos.

O valor de um bem decorre de várias características, entre as quais se destacam sua raridade e sua utilidade para satisfazer necessidades e interesses humanos e sofre influências por suas características singulares e condições de oferta e procura. Trata-se de um conceito econômico abstrato e não de um fato.

O preço é uma expressão monetária que define uma transação de um bem, de seu fruto, de um direito, ou da expectativa de sua transação. O preço é um fato concreto, relacionado às capacidades financeiras, às motivações ou aos interesses específicos do comprador ou do vendedor.

O custo tem relação com o total dos gastos diretos e indiretos necessários à produção ou aquisição de um bem, fruto ou direito.”

Em seu item 0.3 a Norma Técnica Brasileira (ABNT: Associação Brasileira de Normas Técnicas, 2019) faz algumas distinções a respeito de valor a depender do objetivo e finalidade da avaliação:

As abordagens de valor podem ser:

- abordagem pelo valor de mercado:  
É identificada a quantia mais provável pela qual um bem seria transacionado no mercado. Alguns tipos de valor podem traduzir ou se identificar com o valor de mercado:
  - o valor econômico, quando forem adotadas condições e taxas de desconto médias praticadas no mercado;
  - o valor patrimonial, quando, em situações específicas, a soma dos valores dos bens do patrimônio corresponder ao valor de mercado;
  - o valor em risco, quando o bem for segurado pelo valor de mercado (caso usual do automóvel).
- abordagem por valores específicos:  
Refere-se aos valores distintos do valor de mercado, entre eles:
  - o valor especial, que reflete características ou premissas relevantes apenas para um comprador especial, desconsideradas na identificação do valor de mercado. O valor econômico, quando forem adotadas premissas vinculadas a um comprador especial (por exemplo: taxas de desconto, isenções fiscais, projeções de receitas e despesas), adquire o caráter de um valor especial;
  - o valor patrimonial, pois a soma dos valores dos bens componentes do patrimônio usualmente não representa o seu valor de mercado;
  - o valor de liquidação forçada, que difere do valor de mercado, em razão de condições especiais destinadas a comercializar o bem em prazo inferior ao usual;
  - o valor em risco, que identifica o montante para fins de cobertura securitária do bem, de acordo com critérios estabelecidos na apólice, quando diferentes do valor de mercado;
  - o valor sinérgico, quando as sinergias estiverem disponíveis a um comprador determinado (por exemplo, caso de combinação de fluxos de caixa de empresas do mesmo segmento e união de terrenos com ganho de aproveitamento).

As definições são melhor detalhadas nos itens: 3.1.48 a 3.1.55 (ABNT: Associação Brasileira de Normas Técnicas, 2019, pp. 7–8):

3.1.47 - valor de mercado: quantia mais provável pela qual se negociaria voluntária e conscientemente um bem, em uma data de referência, dentro das condições do mercado vigente;

3.1.48 - valor econômico: valor presente da renda líquida auferível pelo empreendimento, durante sua vida econômica, a uma taxa de desconto

correspondente ao custo de oportunidade, considerados cenários previsíveis sob condições de risco;

3.1.49 - valor depreciável: diferença entre o valor do bem na condição de novo e o seu valor residual;

3.1.50 - valor de indenização: valor atribuído a danos, perdas ou prejuízos provocados, referido a uma determinada data;

3.1.51 - valor em risco: valor representativo da parcela do bem que se deseja segurar e que pode corresponder ao valor máximo segurável;

3.1.52 - valor especial: valor que considera atributos particulares de um bem ou direito, que geram interesse somente para um comprador especial ou sob as condições de uma premissa especial;

3.1.53 - valor patrimonial: valor de um bem, partes de um bem ou conjunto de bens de pessoa física ou jurídica, determinado conforme o objetivo, a finalidade e a abrangência da avaliação;

3.1.54 - valor residual: valor do bem ao final de sua vida útil ou de seu horizonte projetivo;

3.1.55 - valor sinérgico: valor resultante da interação de dois ou mais bens ou direitos, quando o valor global for maior do que a soma dos valores individuais.

### A.3 MÉTODOS

A NBR 14.653 (partes 1 e 2) estabelece vários métodos para se estimar o valor de um bem, de seus frutos e direitos:

- método comparativo de dados de mercado: ao se dispor de dados de mercado que possam compor uma amostra representativa de imóveis com características, tanto quanto possível, semelhantes ao avaliando, cujas desigualdades serão objeto de homogeneização por meio de um tratamento técnico dos atributos dos elementos comparáveis, constituintes da amostra;
- método involutivo: identifica o valor do bem, alicerçado no seu aproveitamento eficiente, baseado em modelo de estudo de viabilidade técnico-econômica, mediante hipotético empreendimento compatível com as características do bem e com as condições do mercado no qual está inserido, considerando-se cenários viáveis para execução e comercialização do produto. O método involutivo pode identificar o valor de mercado;
- método da capitalização da renda: o valor de uma propriedade é estimado a partir da montagem de um fluxo de caixa que considere as despesas de manutenção e operação e as receitas de exploração do bem, sendo o valor máximo do imóvel o valor atual do fluxo de caixa descontado sob uma taxa de atratividade mínima sob vários cenários econômicos;

- método evolutivo: o valor total do imóvel é obtida pela conjugação de métodos, a partir do valor do terreno, considerados o custo de reprodução das benfeitorias (caso a finalidade seja a identificação do valor de mercado, deve ser considerado o fator de comercialização).

#### A.4 IMPOSIÇÕES E ENQUADRAMENTOS NORMATIVOS

O item 9 da NBR 14.653-2 (ABNT: Associação Brasileira de Normas Técnicas, 2011) cita que a *especificação* de uma avaliação está “relacionada tanto com o empenho do engenheiro de avaliações, como com o mercado e as informações que possam ser dele extraídas”, e estabelece graus para a *fundamentação* e a *precisão* obtidos num trabalho avaliatório.

O *grau de fundamentação* está relacionado ao empenho profissional na etapa de pesquisa e qualificação dos elementos; o *grau de precisão* é uma decorrência exclusiva das características do mercado e da amostra.

As Tabelas 1 e 2 (ABNT: Associação Brasileira de Normas Técnicas, 2011, pp. 22–24) estabelecem os critérios de pontuação para o enquadramento de um trabalho técnico quanto ao seu *grau de fundamentação* e a Tabela 5 (ABNT: Associação Brasileira de Normas Técnicas, 2011, p. 26) estabelece os critérios de pontuação para o enquadramento de um trabalho técnico quanto ao seu *grau de precisão*.

Tabela A.1: Grau de fundamentação no caso de utilização de modelos de regressão linear

Item	Descrição	Grau		
		III	II	I
1	Caracterização do imóvel avaliando	Completa quanto a todas as variáveis analisadas	Completa quanto às variáveis utilizadas no modelo	Adoção de situação paradigma
2	Quantidade mínima de dados de mercado efetivamente utilizados	$6(k+1)$ , onde $k$ é o número de variáveis independentes	$4(k+1)$ , onde $k$ é o número de variáveis independentes	$3(k+1)$ , onde $k$ é o número de variáveis independentes
3	Identificação dos dados de mercado	Apresentação de informações relativas a todos os dados e variáveis analisados na modelagem, com foto e características observadas no local pelo autor do laudo	Apresentação de informações relativas a todos os dados e variáveis analisados na modelagem	Apresentação de informações relativas a todos os dados e variáveis efetivamente utilizados na modelagem
4	Extrapolação	Não admitida	Admitida para apenas uma variável desde que: a) as medidas das características do imóvel avaliando não sejam superiores a 100 % do limite amostral superior, nem inferiores à metade do limite amostral inferior; b) o valor estimado não ultrapasse 15 % do valor calculado no limite da fronteira amostral, para a referida variável, em módulo	Admitida desde que: a) as medidas das características do imóvel avaliando não sejam superiores a 100 % do limite amostral superior, nem inferiores à metade do limite amostral inferior; b) o valor estimado não ultrapasse 20 % do valor calculado no limite da fronteira amostral, para as referidas variáveis, <i>per si</i> e simultaneamente, e em módulo
5	Nível de significância máximo para a rejeição da hipótese nula de cada regressor (teste bicaudal)	10 %	20 %	30 %
6	Nível de significância máximo admitido para a rejeição da hipótese nula do modelo por meio do teste $F$ de <i>Snedecor</i>	1 %	2 %	5 %

Fonte: Adaptado de (ABNT: Associação Brasileira de Normas Técnicas, 2011, p. 22–23)

Tabela A.2: Enquadramento do laudo segundo seu grau de fundamentação no caso de utilização de modelos de regressão linear

	Grau		
	III	II	I
Pontos mínimos	16	10	6
Itens obrigatórios	2,4,5 e 6 no Grau III e os demais no mínimo no Grau II	2,4,5 e 6 no mínimo no Grau II e os demais no mínimo no Grau I	Todos no mínimo no Grau I

Fonte: Adaptado de (ABNT: Associação Brasileira de Normas Técnicas, 2011, p. 24)

Tabela A.3: Grau de precisão nos casos de utilização de modelos de regressão linear ou do tratamento por fatores

Descrição	Grau		
	III	II	I
Amplitude do intervalo de confiança de 80 % em torno da estimativa de tendência central	$\leq 30\%$	$\leq 40\%$	$\leq 50\%$

Fonte: Adaptado de (ABNT: Associação Brasileira de Normas Técnicas, 2011, p. 26)

## B REVISÃO DA LITERATURA

### B.1 MODELO CLÁSSICO DE REGRESSÃO LINEAR (*LM*)

#### B.1.1 Modelo sem distribuição especificada para o erro

Diferentemente de uma função matemática, um modelo estatístico busca exprimir uma relação não puramente determinística ao considerar a existência de uma aleatoriedade no processo que está a gerar a resposta. Equivale dizer que a relação funcional não é totalmente conhecida e, para exprimir todas as demais fontes de variabilidade não conhecidas, incorpora-se ao modelo um termo de erro e, ao se admitir que esse termo possa ser incorporado de modo aditivo um modelo estatístico de regressão linear assume a forma

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i \quad i = 1, 2, \dots, n, \quad (\text{B.1})$$

em que  $Y_i$  é o valor da variável resposta na  $i$ -ésima observação;  $\beta_0, \dots, \beta_k$  são os parâmetros do modelo;  $X_{ik}$  é o valor assumido pela  $k$ -ésima variável explicativa na  $i$ -ésima observação e  $\varepsilon_i$  é o erro.

A expressão “linear” não se relaciona à forma da distribuição mas sim à linearidade nos parâmetros da combinação linear estabelecida pelas covariáveis (1); ou seja, nenhum de seus parâmetros ( $\beta_0, \dots, \beta_k$ ) aparece como um expoente ou é dividido, multiplicado por algum outro parâmetro (2)

$$Y_i = \beta_0 + \beta_1 X_{i1}^2 + \beta_2 X_{i2} + \varepsilon_i \quad i = 1, 2, \dots, n; \quad (1)$$

$$Y_i = \beta_0 + e^{\beta_1} X_{i1} + \beta_2 X_{i2} + \varepsilon_i \quad i = 1, 2, \dots, n; \quad (1)$$

$$Y_i = \beta_0 + \beta_1 X_{i1}^{\beta_2} + \varepsilon_i \quad i = 1, 2, \dots, n; \quad (2)$$

$$Y_i = \beta_0 e^{\beta_1 X_{i1}} + \varepsilon_i \quad i = 1, 2, \dots, n. \quad (2)$$

O termo de erro ( $\varepsilon$ ) incorpora não apenas o efeito sobre a variável  $Y$  de todas as variáveis que (por diversas razões) foram omitidas na formulação do modelo, como:

- embasamento teórico vago, equivale dizer que a teoria existente suporta, com certeza, apenas algumas variáveis e, para essa lacuna,  $\varepsilon_i$  seria um substituto de todas as demais variáveis, excluídas do modelo;
- falta de dados disponíveis;
- princípio da parcimônia: modelo mais simples que melhor explique a relação;
- a inclusão de outras variáveis pode resultar numa forma funcional equivocada,

mas também os mais diversos tipos de erros presentes na experimentação como aqueles associados à medição, à perda de precisão instrumental (calibragem), existência de mais de um operador (equipamento).

O modelo (B.1) pode ser representado sob a forma matricial:  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$  onde

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{(n,1)} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \dots & x_{1,k} \\ 1 & x_{2,1} & x_{2,2} & \dots & x_{2,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \dots & x_{n,k} \end{bmatrix}_{(n,k+1)} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1,1)} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{(n,1)}$$

De modo geral, na  $i$ -ésima linha o vetor  $\mathbf{Y}$  apresenta a resposta dessa observação, o vetor  $\boldsymbol{\varepsilon}$  o erro a ela associado e a matriz  $\mathbf{X}$  os valores assumidos pelas  $k$ -variáveis explicativas consideradas no modelo (sua primeira coluna é inteiramente composta por “uns” posto o modelo (B.1) incorporar um intercepto:  $\beta_0$ ). O vetor  $\boldsymbol{\beta}$  apresenta os parâmetros das covariáveis.

Ao se admitir algumas premissas relativas ao termo de erro:

- $\mathbb{E}(\varepsilon_i) = 0$ ;
- $\text{Var}(\varepsilon_i) = \sigma^2$ ;
- $\varepsilon_i$  e  $\varepsilon_j$  não são correlacionados de modo a que sua covariância seja zero:  $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ ,

o modelo estatístico (B.1) apresenta as seguintes propriedades:

- decompõe a variabilidade da  $i$ -ésima observação  $Y_i$  na soma de duas partes com naturezas distintas:
  - componente sistemático ou determinístico:  $\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik}$ ;
  - termo de erro ou componente aleatório:  $\varepsilon_i$ .
- como  $\mathbb{E}(\varepsilon_i) = 0$  a resposta  $Y_i$  para a  $i$ -ésima observação associada aos valores do vetor  $[X_{i1}, X_{i2}, \dots, X_{ik}]$  vem de uma distribuição de probabilidade cujo valor esperado é  $\mathbb{E}(Y_i) = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik}$ ;
- a resposta  $Y_i$  para a  $i$ -ésima associada aos valores do vetor  $[X_{i1}, X_{i2}, \dots, X_{ik}]$  desvia-se do valor da função de regressão  $(\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik})$  por  $\varepsilon_i$ ;
- como se assume que os erros  $\varepsilon_i$  possuem variância constante  $\sigma^2$ , segue-se que as respostas  $Y_i$  tenham variância também constante e igual a  $\text{Var}(Y_i) = \sigma^2$ ;
- como se assume que os termos do erro  $\varepsilon_i$  e  $\varepsilon_j$  não são correlacionados, também não são correlacionadas as  $i$ -ésima e  $j$ -ésima respostas  $Y_i$  e  $Y_j$ .



O componente aleatório do modelo tem como objetivo manter a forma proposta para o modelo inalterada mesmo sob diferentes subconjuntos de dados coletados da população, absorvendo a aparente aleatoriedade dos valores como um resultado das diferentes fontes de variabilidade de modo parcialmente conhecido: sua média ( $\mathbb{E}(\varepsilon_i)$ ) e variância ( $\text{Var}(\varepsilon_i)$ ) são consideradas fixas; isto é apenas seus valores particulares, para uma observação qualquer é que não são conhecidos.

### B.1.2 Modelo com distribuição Normal para o erro

As suposições estatísticas anteriormente feitas são insuficientes para qualquer inferência a partir do modelo (como para a construção de intervalos de confiança para as estimativas dos parâmetros do modelo ou então das próprias estimativas realizadas com o modelo). Para tanto premissas adicionais sobre a distribuição devem ser consideradas.

A forma do modelo estatístico com erro Normal é a mesma do modelo (B.1)

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_r X_{ik} + \varepsilon_i, \quad (\text{B.2})$$

sobre o qual se assume como válidos os seguintes pressupostos (HOFFMANN, 2016, p. 43)

- há linearidade na relação entre as variáveis preditoras  $X_{ik}$  e a variável resposta  $Y_i$ : o valor esperado da variável resposta é uma combinação linear ( $\phi$ ) das variáveis preditoras observadas e dos seus parâmetros desconhecidos:  $\phi(\beta_0, \beta_1; \dots, \beta_r; x_1, \dots, x_k)$ ;
- os valores de  $X$  são fixos ou independentes do termo de erro; isto é,  $X$  não é uma variável aleatória;
- o valor médio do termo de erros é zero  $\mathbb{E}(\varepsilon_i) = 0$ ;
- a variância dos erros é constante (homocedástica) independente do valor de  $X_i$ :

$$\begin{aligned} \text{Var}(\varepsilon_i) &= \mathbb{E}(\varepsilon_i^2) = \sigma^2 \\ [\mathbb{E}[Y_i - \mathbb{E}(Y_i|X_i)]^2 &= \sigma^2; \end{aligned}$$

- não há autocorrelação entre os termos de erro: o erro de uma observações é não correlacionado ao erro de outra  $\text{Cov}(\varepsilon_i, \varepsilon_j) = \mathbb{E}(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j$ ;
- os erros têm distribuição Normal.

Os pressupostos 1, 2 e 3 permitem escrever o modelo (B.2) como  $\mathbb{E}(Y_i) = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_r X_{ik}$ . Combinado os pressupostos 3, 4 e 6 segue-se que

$$\varepsilon_i \stackrel{\text{ind}}{\sim} \mathcal{N}(0, \sigma^2),$$

o que é equivalente assumir que:

$$Y_i \stackrel{ind}{\sim} \mathcal{N}(\mathbf{X}_i\boldsymbol{\beta}, \sigma^2). \quad (\text{B.3})$$

O Modelo Clássico de Regressão Linear pressupõe, portanto três características nos dados analisados (VIEIRA, 2004, p. 19):

- a variável resposta tem distribuição Normal;
- a variância da resposta é constante;
- os efeitos das covariáveis sobre a variável resposta se combinam aditivamente.

A construção de um Modelo Hedônico de Regressão requer não apenas uma estrutura de covariáveis coerente com o que se deseja explicar e a obtenção de uma amostra compatível com a quantidade de variáveis independentes adotadas, mas também uma criteriosa análise quanto ao atendimento das premissas preliminarmente admitidas como válidas para que as inferências conduzidas com esse modelo não sejam viesadas (FIGUEIREDO FILHO et al., 2011, pp. 51–57).

As propriedades apresentadas pelos estimadores de mínimos quadrados não dependem da suposição de normalidade; todavia, se a variância não se mostrar constante, esse estimador apesar de não enviesado, não terá variância mínima e os erros-padrão dos estimadores dos coeficientes estarão superestimados, tendendo a tornar significantes as estimativas de parâmetros quando estas não o deveriam ser (VIEIRA, 2004, p. 35).

Sob as premissas anteriores, um modelo de regressão linear com distribuição Normal para o erro seria representado pela reta que passa pelo valor médio de cada nível observado das covariáveis consideradas  $X_{ik}$  e que estes possuem variâncias constantes (cf. Figura B.1).

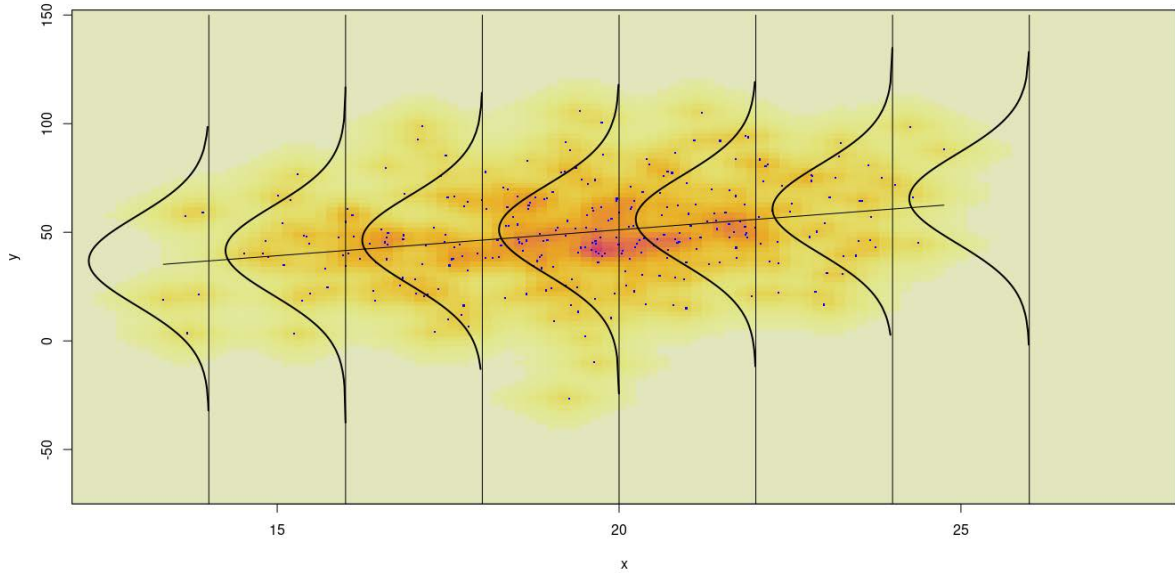
### B.1.3 Estimação dos parâmetros de um Modelo Clássico de Regressão Linear

#### B.1.3.1 Método dos mínimos quadrados

A solução do modelo (B.1) consiste na estimação do vetor  $\hat{\boldsymbol{\beta}} = \hat{\beta}_0, \dots, \hat{\beta}_k$  tal que a soma dos quadrados seja mínima. A soma dos quadrados dos erros é

$$\begin{aligned} SQR(\beta_0, \dots, \beta_k) &= \sum_{i=1}^n \varepsilon_i^2 \\ SQR(\boldsymbol{\beta}) &= \sum_{i=1}^n \left[ Y_i - \left( \beta_0 + \sum_{j=1}^k \beta_j X_{ij} \right) \right]^2 \\ &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{Y}^T \mathbf{Y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} \end{aligned} \quad (\text{B.4})$$

Figura B.1: Modelo de regressão linear simples com distribuição Normal para o erro  $\mathbb{E}(\varepsilon_i) = 0$ ;  $\text{Var}(\varepsilon_i) = \sigma^2$



Fonte: Autor

As estimadores de mínimos quadrados para  $\beta$  poderão ser obtidos derivando-se em relação a cada parâmetro e estabelecendo-se a condição:

$$\begin{aligned} \frac{\partial[SQR(\beta)]}{\partial\beta_j} \Big|_{\beta=\hat{\beta}} &= 0 \quad j = 1, \dots, p \quad \text{equivalente a:} \\ -2\mathbf{X}^T \mathbf{Y} + 2\mathbf{X}^T \mathbf{X} \hat{\beta} &= 0 \\ \mathbf{X}^T \mathbf{X} \beta &= \mathbf{X}^T \mathbf{Y}. \end{aligned}$$

Se a matriz  $\mathbf{X}^T \mathbf{X}$  for positivo definida essa equação poderá ser resolvida multiplicando-se ambos os lados por  $(\mathbf{X}^T \mathbf{X})^{-1}$ , resultando em:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (\text{B.5})$$

O Teorema de Gauss-Markov afirma que sob as suposições estatísticas admitidas sobre os dados (sua média, variância e covariância) que os estimadores de mínimos quadrados para  $\beta_0, \dots, \beta_k$  ( $\hat{\beta}_0, \dots, \hat{\beta}_k$ ) são não viesados e, entre os estimadores não viesados, são os de mínima variância (MYERS et al., 2010, p. 472).

A variância de cada estimativa e a covariância entre duas estimativas é dada pela matriz de covariância (MYERS et al., 2010, p. 15):  $\text{Cov}(\hat{\beta})$ , uma matriz simétrica e de dimensões  $(p \times p)$ , tal que:

- a variância da estimativa  $\hat{\beta}_j$  é dada pelo  $j$ -ésimo elemento de sua diagonal;
- a covariância entre duas estimativas  $\hat{\beta}_i$  e  $\hat{\beta}_j$  é dada pelo elemento na  $i$ -ésima linha e  $j$ -ésima coluna.

$$\begin{aligned}
\text{Cov}(\hat{\beta}) &= \mathbb{E}\{[\hat{\beta} - \mathbb{E}(\hat{\beta})][\hat{\beta} - \mathbb{E}(\hat{\beta})]^T\} \\
\text{Cov}(\hat{\beta}) &= \text{Var}(\hat{\beta}) \\
&= \text{Var}[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}] \\
&= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{Var}(\mathbf{Y}) (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \\
&= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\
&= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}
\end{aligned}$$

Um estimador para  $\sigma^2$  é dado por (MYERS et al., 2010, p. 18):

$$\hat{\sigma}^2 = \mathbf{Y}^T \mathbf{Y} - \hat{\beta}^T \mathbf{X}^T \mathbf{Y}$$

### B.1.3.2 Método da máxima verossimilhança

Em 17 de novembro de 1921 *Sir* Ronald Aylmer Fisher apresenta à Royal Society o primeiro de uma série de artigos que viriam a ser denominados como trabalhos seminais e tidos por muitos como as bases da Estatística moderna: *On The Mathematical Foundations of Theoretical Statistics*, com definições formais sobre consistência, eficiência, estimação, verossimilhança, suficiência (FISHER, 1922, pp. 309-310), <sup>1</sup>.

“A teoria da máxima verossimilhança é realmente muito bonita: uma abordagem conceitualmente simples para uma coleção incrivelmente ampla de problemas. Esta teoria fornece uma receita simples que pretende levar à solução ideal para todos os problemas paramétricos e mais, não apenas promete uma estimativa ótima, mas também uma avaliação geral de sua precisão. E tudo isso vem sem a necessidade de especificação de probabilidades a priori e nenhuma derivação complicada de distribuições” (STIGLER, 2007, p. 2, tradução nossa)<sup>2</sup>

O conceito primário de verossimilhança não foi originariamente concebido por Fisher. Joseph Louis Lagrange, Daniel Bernoulli, Leonard Euler, Pierre Simon Laplace e Carl Friedrich Gauss já haviam explorado o tema ainda que de modo diverso daquele que viria a sê-lo hoje (STIGLER, 2007, p. 598).

<sup>1</sup>Foi nesse artigo que Fisher utilizou o termo “parâmetro” no sentido que viria a ser doravante utilizado na Estatística moderna (STIGLER, 2005, p. 32)

<sup>2</sup>“The theory of maximum likelihood is very beautiful indeed: a conceptually simple approach to an amazingly broad collection of problems. This theory provides a simple recipe that purports to lead to the optimum solution for all parametric problems and beyond, and not only promises an optimum estimate, but also a simple all-purpose assessment of its accuracy. And all this comes with no need for the specification of a priori probabilities, and no complicated derivation of distributions”

Para estabelecer o conceito formal moderno de verossimilhança considere  $Y_1, \dots, Y_n$   $n$  variáveis aleatórias independentes com função densidade de probabilidade dadas por  $f_i(y_i; \theta), \theta \in \Theta$ . A função densidade conjunta de probabilidade de  $n$  observações independentes  $y = (y_1, \dots, y_n)$  será:

$$f(\mathbf{y}, \theta) = \prod_{i=1}^n f_i(y_i, \theta) = L(\theta|\mathbf{y}) \quad (\text{B.6})$$

A expressão (B.6), uma função do vetor paramétrico desconhecido  $\theta$  (com verdadeiro valor  $\theta_0$ ) dado o vetor de observações  $\mathbf{y}$  é denominada de função de verossimilhança e dela depreende-se que  $y = (y_1, \dots, y_n)$  serão observados com maior probabilidade sob um vetor de parâmetros  $\hat{\theta}$  em detrimento de um outro qualquer se essa função assumir um valor maior naquele primeiro; isto é, o valor para  $\theta$  deve ser tal que resulte na maior verossimilhança da função dada a amostra.

Usualmente o procedimento matemático para a maximização da função de verossimilhança é pela adoção de seu logaritmo e assim segue-se que  $\hat{\theta}$  será o estimador de máxima verossimilhança quando:

$$l(\hat{\theta}|\mathbf{y}) = \log [L(\hat{\theta}|\mathbf{y})] \geq \log [L(\theta|\mathbf{y})] \quad \text{para todo } \theta$$

$$l(\hat{\theta}, \phi|y_i) = \sum_{i=1}^n \left[ \frac{y_i \hat{\theta}_i - b(\hat{\theta}_i)}{\phi} + c(y_i, \phi) \right] \quad (\text{B.7})$$

Se (B.7) é contínua e suas duas primeiras derivadas em todo o espaço paramétrico dos valores de  $\theta$  existam, então o estimador de máxima verossimilhança poderá ser obtido pela resolução de um sistema de equações formado pelas derivadas parciais da função (B.7) em relação a cada um dos parâmetros do vetor  $\theta$  impondo-se a condição:

$$\left. \frac{\partial l(\theta|\mathbf{y})}{\partial \theta_j} \right|_{\theta=\hat{\theta}} = 0 \quad j = 1, \dots, p. \quad (\text{B.8})$$

O sistema (B.8) é composto por equações não lineares e sua resolução deve ser feita por métodos numéricos aproximativos.

O gradiente da função (B.7) usado para se maximizar a função de verossimilhança (B.6) é denominado por função Escore de Fisher ( $U(\theta)$ ) e dois importantes resultados relacionados aos dois primeiros momentos serão usados na construção da teoria dos Modelos Lineares Generalizados (CORDEIRO; DEMÉTRIO; MORAL, 2015, p. 3).

Sob condições de regularidade suficientes para algumas operações matemáticas, seguem-se alguns resultados associados à função escore de Fisher (KENDALL; STUART, 1961, p. 9)

$$\mathbb{E} \left[ \frac{\partial l(\theta|\mathbf{y})}{\partial \theta} \right] = 0. \quad (\text{B.9})$$

Assim,  $\mathbb{E}(U) = 0$ .

Pode-se também demonstrar que (WOOD, 2017, p. 407):

$$\mathcal{I} \equiv \mathbb{E} \left[ \left( \frac{\partial l(\boldsymbol{\theta}|\mathbf{y})}{\partial \theta} \right)^2 \right] = -\mathbb{E} \left[ \frac{\partial^2 l(\boldsymbol{\theta}|\mathbf{y})}{\partial \theta^2} \right], \quad (\text{B.10})$$

em que  $\mathcal{I}$  é denominada de informação sobre  $\boldsymbol{\theta}$  condida na amostra (Informação de Fisher). Como decorrência de (B.9) e (B.10) segue-se

$$\begin{aligned} \text{Var}(U) &= \mathbb{E}(U^2) - [\mathbb{E}(U)]^2 \\ &= \mathbb{E}(U^2) \\ &= -\mathbb{E} \left[ \frac{\partial^2 l(\theta)}{\partial \theta^2} \right] \\ &= -\mathbb{E}(U'). \end{aligned} \quad (\text{B.11})$$

Os estimadores de máxima verossimilhança apresentam as seguintes propriedades:

- invariância;
- suficiência;
- assintoticidade.

Seja  $\mathbf{y} = [y_1, \dots, y_n]^T$  um vetor de variáveis aleatórias com função densidade conjunta de probabilidade  $f(\mathbf{y}, \theta)$  onde  $\hat{\theta}$  é uma estimativa de máxima verossimilhança para o parâmetro  $\theta$ .

Se  $\gamma$  é um parâmetro tal que  $\gamma = g(\theta)$ , onde  $g$  é uma função qualquer inversível definida no espaço paramétrico  $\Theta$ , então a estimativa de máxima verossimilhança de  $\gamma$  será  $\hat{\gamma} = g(\hat{\theta})$ . A esta propriedade denomina-se por invariância.

Se  $T = T(y_1, \dots, y_n)$  é uma estatística suficiente<sup>3</sup> para  $\theta$  então, se houver um estimador de máxima verossimilhança de  $\theta$ , ele será uma função de  $T$ . A esta propriedade denomina-se por suficiência.

A assintoticidade dos estimadores de máxima verossimilhança sob amostras de grande tamanho e sob certas condições de regularidade atendidas (identificabilidade do modelo e existência, unicidade, consistência e normalidade assintótica dos estimadores) tem distribuição tal que

$$\sqrt{n}(\hat{\theta} - \theta) \overset{a}{\sim} \mathcal{N}\left(0, \frac{1}{\mathcal{I}(\theta)}\right),$$

e:

---

<sup>3</sup>Uma estatística  $T = T(Y)$  é dita suficiente para um parâmetro  $\theta$  quando resume toda informação sobre esse parâmetro contida na amostra. Assim, se  $T$  é suficiente para  $\theta$  então a distribuição condicional de  $Y$  dada  $T$  é independente de  $\theta$ :  $P(Y = y|T = t; \theta) = P(Y = y|T = \cdot)$ .

$$\sqrt{n} [g(\hat{\theta}) - g(\theta)] \stackrel{a}{\sim} \mathcal{N} \left( 0, \frac{[g'(\theta)]^2}{\mathcal{I}(\theta)} \right).$$

#### B.1.4 Modelo Clássico de Regressão Linear com distribuição Normal para o erro

Mantendo a premissa de que  $\varepsilon_i \stackrel{ind}{\sim} \mathcal{N}(0, 1)$  no modelo

$$y_i = f(x_i|\boldsymbol{\beta}) + \varepsilon_i,$$

a função de verossimilhança será dada por

$$L(\boldsymbol{\beta}|\sigma^2) = \prod_{i=1}^n f(x_i|\boldsymbol{\beta});$$

$$L(\boldsymbol{\beta}|\sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - f(x_i, \boldsymbol{\beta})]^2 \right\}$$

que poderá ser maximizada com a escolha de um vetor de parâmetros  $\hat{\boldsymbol{\beta}}$  tal que

$$l(\boldsymbol{\beta}|\sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - f(x_i|\boldsymbol{\beta})]^2;$$

$$\frac{1}{\sigma^2} \sum_{i=1}^n [y_i - f(x_i|\boldsymbol{\beta})] \frac{\partial f(x_i|\boldsymbol{\beta})}{\partial \beta_j} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = 0 \quad j = 1, \dots, p.$$

Seja  $\mu_i = f(x_i, \boldsymbol{\beta})$  para  $i = 1, \dots, n$  e  $D_{ij} = \partial f(x_i|\boldsymbol{\beta}) / \partial \beta_j|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}$  para  $i = 1, \dots, n$  e  $j = 1, \dots, p$ . O sistema de equações não lineares formado pode ser escrito na forma

$$\frac{1}{\sigma^2} D^T (\mathbf{y} - \hat{\boldsymbol{\mu}}) = 0, \quad (\text{B.12})$$

em que  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$ ,  $D = [D_{ij}]$  e  $\hat{\boldsymbol{\mu}}$  é o valor esperado da função  $\mu_i = f(x_i, \boldsymbol{\beta})$  onde os parâmetros  $\boldsymbol{\beta}$  foram substituídos por  $\hat{\boldsymbol{\beta}}$ . Em uma regressão linear  $\mathbf{D} = \mathbf{X}$  e  $\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}}$  e, desse modo, (B.12) fornece as mesmas estimativas que (B.5) (MYERS et al., 2010, p. 88).

#### B.1.5 Resíduos ordinários e resíduos “estudentizados”

Considerando-se que em grande parte das pesquisas é inviável construir-se um modelo estatístico sobre toda a população (chegando a ser impossível em algumas situações) o estudo é usualmente conduzido a partir de observações amostrais. Assim o modelo estatístico estimado a partir de uma amostra de tamanho  $n$  passa a ser expresso na forma

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \dots + \hat{\beta}_k X_{ik},$$

Para os valores do vetor  $[X_{i1}, X_{i2}, \dots, X_{ik}]$  associados à  $i$ -ésima observação é possível estabelecer um resíduo ordinário da estimativa do modelo (B.4) como sendo  $e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \dots - \hat{\beta}_r X_{ik}$ , que reescrito vetorialmente assume a forma

$$\begin{aligned} e &= \mathbf{Y} - \hat{\mathbf{Y}}; \\ e &= (\mathbf{I} - \mathbf{H})\mathbf{Y}, \end{aligned} \quad (\text{B.13})$$

em que  $\mathbf{Y} = (Y_1, \dots, Y_n)^T$  é o vetor resposta;  $\hat{\mathbf{Y}} = (\hat{Y}_1, \dots, \hat{Y}_n)^T$  é o vetor dos valores estimados;  $\mathbf{H}$  é a matriz de projeção e  $e = (e_1, \dots, e_n)^T$  é o vetor dos resíduos ordinários do modelo.

A relação entre  $\varepsilon$  e  $e$  se estabelece ao se substituir  $\mathbf{Y}$  por  $\mathbf{X}\beta + \varepsilon$

$$\begin{aligned} e &= (\mathbf{I} - \mathbf{H})(\mathbf{X}\beta + \varepsilon); \\ e &= (\mathbf{I} - \mathbf{H})\varepsilon, \end{aligned} \quad (\text{B.14})$$

em que  $\mathbf{X}$  é a matriz de delineamento de dimensão  $n \times p$  (com  $p = k + 1$  e tendo “uns” em sua primeira coluna se uma constante  $\beta_0$  for adicionada ao modelo e os valores das  $k$  covariáveis nas demais colunas, ou  $p = k$  caso não o seja);  $\beta = (\beta_0, \dots, \beta_k)^T$  é o vetor dos parâmetros do modelo e  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$  é o vetor dos erros do modelo.

Como a matriz de projeção em (B.13) é dada por

$$\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T,$$

em que

$$h_{ii} = \frac{1}{n} + x_i^T (\mathbf{X}^T\mathbf{X})^{-1} x_i,$$

segue-se que (B.14) possui a forma escalar

$$e_i = \varepsilon_i - \sum_{j=1}^n h_{ij} \varepsilon_j. \quad (\text{B.15})$$

A expressão (B.15) demonstra que a relação entre  $e$  e  $\varepsilon$  dependendo apenas de  $h_{ij}$  e que, assim, quando esse valor é suficientemente pequeno  $e$  pode ser tomado como uma aproximação de  $\varepsilon$  (COOK; WEISBERG, 1982, p. 11).

A matriz de projeção <sup>4</sup> é a transformação linear que projeta ortogonalmente qualquer vetor  $n$ -dimensional no espaço gerado pelas colunas da matriz de delineamento  $\mathbf{X}$  (COOK; WEISBERG, 1982, p. 11) tal que  $\hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y}$ . Como uma matriz de projeção,  $\mathbf{H}$  tem duas características (HOAGLIN; WELSCH, 1978, p. 18):

- é simétrica:  $\mathbf{H}^T = \mathbf{H}$ ;
- é idempotente:  $\mathbf{H}^2 = \mathbf{H}$ .

---

<sup>4</sup>A matriz de projeção foi alcunhada como *hat matrix* (matriz “chapéu”) por John Tukey (HOAGLIN; WELSCH, 1978, p. 17). Também é denominada por matriz de influência por seu papel na análise de resíduos.



Se  $\mathbf{H}$  é simétrica e idempotente, segue-se que  $\text{tr}(\mathbf{H}) = \text{var}(\mathbf{H}) = p$  e  $\mathbf{I} - \mathbf{H}$  também será uma matriz simétrica e idempotente.

Trazendo-se de (B.3) o pressuposto que  $\text{Var}(Y) = \sigma^2 = \sigma^2 \mathbf{I}$ , tem-se por (B.13)

$$\begin{aligned}\text{Var}(e) &= \text{Var}[(\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon}]; \\ \text{Var}(e) &= (\mathbf{I} - \mathbf{H})^2 \text{Var}(\boldsymbol{\varepsilon}); \\ \text{Var}(e) &= (\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H})\sigma^2; \\ \text{Var}(e) &= (\mathbf{I} - \mathbf{H})\sigma^2(\mathbf{I} - \mathbf{H})^T. \\ \text{Var}(e) &= \sigma^2(\mathbf{I} - \mathbf{H})\end{aligned}$$

Em forma escalar

$$\text{Var}(e_i) = \sigma^2(1 - h_{ii}).$$

A variância dos resíduos ordinários é uma função de  $\sigma^2$  e  $h_{ii}$  (B.15), assim outros tipos de resíduos independentes dessas quantidades podem ser mais úteis para o diagnóstico de um modelo de regressão.

Estudos motivados pela preocupação com observações discrepantes em uma análise de regressão linear definiram então resíduos “estudentizados” internamente do seguinte modo (COOK; WEISBERG, 1982, p. 18)

$$r_i = \frac{e_i}{\sqrt{\hat{\sigma}^2(1 - h_{ii})}}, i = 1, 2, \dots, n$$

em que  $\hat{\sigma}^2 = \sum e_i^2 / (n - p)$  é o quadrado médio dos resíduos. Se o modelo está corretamente especificado tem-se que  $r_i \stackrel{\text{ind}}{\sim} \mathcal{N}(0, 1)$ .

A preferência na utilização dos resíduos “estudentizados” - em detrimento aos resíduos ordinários - nas análises gráficas passou a ser preferida por vários autores uma vez que os padrões visuais de  $r_i$  na situação de heterocedasticidade da variância serão mais destacados do que ao se usar  $e_i$  (COOK; WEISBERG, 1982, p. 37).

#### **B.1.6 Adequabilidade de um Modelo Clássico de Regressão Linear com distribuição Normal para o erro**

A adequabilidade <sup>5</sup> de um modelo pode ser confirmada tanto por técnicas gráficas informais para exibir suas características gerais, como por testes formais para detectar desvios dos pressupostos subjacentes (procedimentos complementares) (COOK; WEISBERG, 1982, p. 10). Os desvios podem ser de natureza sistemática ou isolada.

---

<sup>5</sup>Um modelo adequado não apresenta discordâncias dos pressupostos admitidos inicialmente como válidos em sua formulação teórica

Por desvios sistemáticos entendem-se os afastamentos das premissas estabelecidas na formulação do modelo. Na etapa diagnóstica busca-se verificar se:

- o modelo de regressão não é linear;
- os resíduos não têm variância constante;
- há algumas observações não muito bem estimadas pelo modelo;
- os resíduos não são independentes;
- os resíduos não são Normalmente distribuídos; ou,
- algumas variáveis independentes importantes foram omitidas no preditor linear.

Os desvios isolados devem ser verificados para os dados amostrais a partir de seus resíduos, quando então procura-se descobrir se algumas dessas observações exerce algum tipo de interação prejudicial nas estimativas dos parâmetros do modelo ou nos valores por ele estimados. Na etapa diagnóstica buscar-se-á por:

#### **B.1.7 Análise da especificação do modelo**

Os erros de especificação relacionam-se com a estrutura de covariáveis adotada para a construção do modelo, quais serão escolhidas e como serão incorporadas ao mesmo e, assim, podem assumir uma situações abaixo (GUJARATI; PORTER, 2011, p. 469) :

- omissão de uma ou mais variáveis relevantes;
- inclusão de uma ou mais variáveis desnecessárias; ou,
- adoção da forma funcional errada: alguns termos devem ser considerados em outra forma que não a linear ( $X_1^2$ ,  $X_2^3$ , por exemplo) ou há interações que deve ser consideradas entre alguns termos ( $X_1X_2$ , por exemplo).

Tais erros acarretam em estimativas tendenciosas e inconsistente dos parâmetros do modelo e da variância do termo de erro resultando em que intervalos de confiança terão amplitude diferente da que teriam caso não houvesse tais erros e assim, os testes de hipóteses podem levar a conclusões incorretas.

Os resíduos podem ser examinados para detectar erros de especificação de modelo, como a omissão de uma variável importante ou de uma forma funcional incorreta. O padrão visual dos resíduos *versus* as variáveis não incorporadas ao modelo não pode diferir do padrão apresentado quando se plotam os resíduos contra as variáveis incorporada no modelo, não só na aleatoriedade de sua dispersão mas também nas frequências ou concentrações mostradas acima ou abaixo do eixo horizontal em zero.

### B.1.8 Análise da linearidade na relação entre as covariáveis e a resposta

A violação da linearidade é extremamente grave pois um modelo ajustado a dados não lineares leva a previsões equivocadas não somente para valores situados além das fronteiras amostrais, como se usualmente observa, mas também para valores situados dentro de seus limites.

Tal violação pode ser visualizada por meio de dois gráficos:

- resíduos *versus* as variáveis preditoras;
- resíduos *versus* valores estimados.

Os padrões visuais que se deseja emergir dos gráficos acima devem assemelhar-se respectivamente a pontos dispersos de modo constante e aproximadamente simétrico em torno do eixo horizontal em zero por toda a faixa de valores da outra variável (cf. Figura B.2).

Relações não lineares devem ser tratadas por meio da aplicação de uma transformação não linear adequada ao padrão da relação na variável resposta ou no variável preditora. Para dados estritamente positivos com uma relação não linear a transformação com a função logaritmo pode ser uma opção. Se uma a transformação com o uso da função logaritmo é aplicada apenas à variável resposta isso equivalente a assumir que ela cresce (ou decai) exponencialmente como uma função da variável preditora.

Outra possibilidade a considerar é adicionar outra variável preditora na forma de uma função não linear como, por exemplo, nos padrões de dispersão que mostrem uma curva parabólica onde pode fazer sentido regredir  $Y$  em função de  $X$  e  $X^2$ .

Não se pode descartar que a não linearidade pode também decorrer da omissão de uma ou mais variáveis independentes importantes que explicam ou corrigem o padrão não linear observado *a priori*.

### B.1.9 Análise da homogeneidade da variância

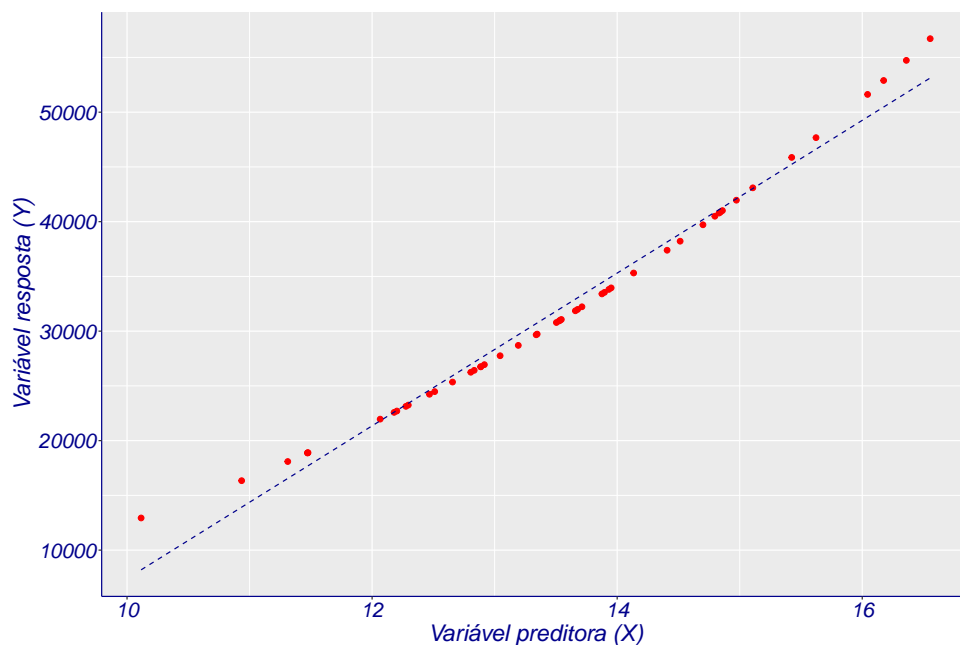
A violação do pressuposto estabelecido no modelo quanto à constância da variância dos erros (heterocedasticidade de  $\varepsilon$ ) acarreta em uma estimação imprecisa do verdadeiro desvio padrão de tal modo que os intervalos de confiança construídos para as estimativas dos parâmetros ou de valores são irreais <sup>6</sup> assim há elevada imprecisão nas inferências feitas com estatísticas baseadas na variância ( $t$ ,  $F$ ).

Sob variância constante (homocedasticidade) tem-se que  $\text{Var}(\varepsilon_i|X_i) = \sigma^2$ ; todavia o que se observa em muitas situações é que a variância está relacionada de algum modo funcional com a média ( $\sigma^2 = f(X)$ ) e, assim

---

<sup>6</sup>São mais amplos ou mais estreitos do que deveriam ser.

Figura B.2: Modelo linear mal ajustado  
 processo simulado de geração da resposta por uma relação não linear  
 $y = 0,75 + 12,50x^3 + e$



Fonte: Autor

$$\text{Var}(\varepsilon_i | X_i) = \sigma_i^2;$$

$$\mathbb{E}(\varepsilon_i^2) = \sigma_i^2.$$

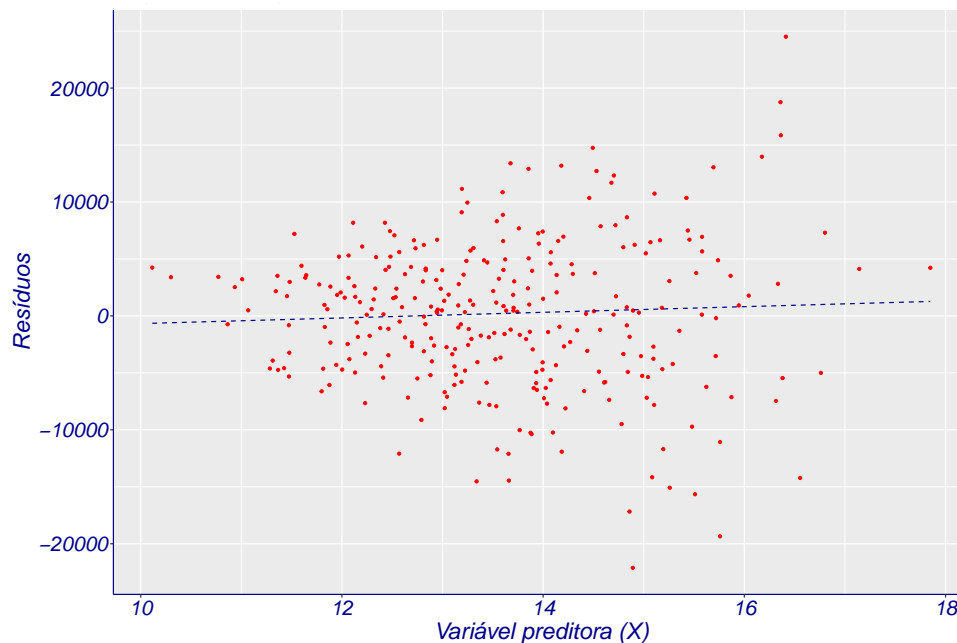
Na presença de heterocedasticidade, os estimadores de mínimos quadrados continuam sendo não viesados e consistentes, mas perdem eficiência. Equivale a dizer que haverá um outro estimador para os parâmetros do modelo, igualmente consistente e não viesado mas a menor variância

$$\text{Var}(\hat{\beta}^*) < \text{Var}(\hat{\beta}).$$

Tal violação pode ser visualizada por meio dos gráficos dos resíduos *versus* valores estimados e resíduos *versus* variáveis preditoras. Os padrões visuais desejados devem se assemelhar a pontos dispersos de modo constante a aproximadamente simétrico em torno do eixo horizontal em zero. Por constância entende-se que os resíduos não exibam, sistematicamente, nenhum padrão de crescimento ou decaimento na amplitude visual de sua dispersão (cf. Figura B.3).

A heterocedasticidade pode ser um subproduto de uma violação significativa das premissas de linearidade e/ou independência, caso em que todas essas violações podem ser conjuntamente corrigidas com a aplicação de uma transformação de potência na variável

Figura B.3: Modelo linear mal ajustado  
 processo de geração simulada dos erros como uma função  
 $e \sim N(0, \sigma = 3, 50 + 2, 50x^3)$



Fonte: Autor

dependente que terá como objetivos linearizar o ajuste tanto quanto possível; e/ou estabilizar a variância dos resíduos. Algum cuidado e discernimento é requerido pois esses dois objetivos podem conflitar entre si. Geralmente opta-se em estabilizar a variância dos resíduos primeiramente para, só então analisar linearização das relações.

A família de transformações para uma variável resposta positiva mais conhecida foi proposta por George Edward Pelham Box e David Roxbee Cox (BOX; COX, 1964, p. 214), a qual transforma um valor observado  $y$  positivo em uma nova variável  $Z$  com o propósito de produzir um comportamento aproximadamente Normal na variável e homocedasticidade nos resíduos do modelo:

$$Z = y(\lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0; \\ \log(y) & \text{se } \lambda = 0. \end{cases}$$

Os valores de  $\lambda$  que maximizam a verissimilhança perfilada resultam então no tipo de transformação a ser aplicada:

- se  $\lambda = -2 \rightarrow 1/Y^2$
- se  $\lambda = 0 \rightarrow \log(Y)$
- se  $\lambda = 2 \rightarrow Y^2$
- se  $\lambda = -1 \rightarrow 1/Y$
- se  $\lambda = 0,50 \rightarrow \sqrt{Y}$
- se  $\lambda = -0,5 \rightarrow 1/\sqrt{Y}$
- se  $\lambda = 1 \rightarrow Y$

Box e Cox (1964) propuseram também uma família de transformações apropriada para valores não positivos de  $y$

$$Z = y(\boldsymbol{\lambda}) = \begin{cases} \frac{(y+\lambda_2)^{\lambda_1}-1}{\lambda_1} & \text{se } \lambda_1 \neq 0; \\ \log(y + \lambda_2) & \text{se } \lambda_1 = 0, \end{cases}$$

em que  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^T$ .

Gráficos dos valores absolutos dos resíduos (ou do quadrado dos resíduos pois os sinais dos resíduos não são significativos para o propósito desse exame) contra a variável preditora  $X$  ou em relação aos valores ajustados também são úteis para o diagnóstico da heterocedasticidade na variância dos resíduos. Estes gráficos são recomendados quando não há muitas observações no conjunto de dados pois a plotagem dos resíduos absolutos ou seus quadrados coloca as informações sobre a alteração das suas magnitudes acima do eixo horizontal em zero o que facilita a inspeção visual de possíveis alterações de sua magnitude em relação a outra variável adotada no gráfico.

De modo complementar, vários testes estatísticos estão definidos (GUJARATI; PORTER, 2011) para a verificação da premissa da variância constante dos resíduos, dentre alguns, aqueles propostos por:

- Robert Ezra Park (PARK, 1966);
- Herbert Glejser (GLEJSER, 1969);
- Trevor Stanley Breusch e Adrian Pagan (BREUSCH; PAGAN, 1979);
- Halbert White (WHITE, 1980).

#### **B.1.10 Análise quanto à independência dos resíduos**

A dependência entre duas observações pode se manifestar sob duas formas distintas a depender da natureza do estudo espacial; ou temporal. Se a variável sob estudo é explicada ao todo ou em parte pela localização espacial das covariáveis, espera-se que respostas de unidades mais próximas terão tendência a serem mais parecidas entre si do que quando de unidades mais distantes (mantidas todas as demais variáveis constantes) (TOBLER, 2004, p. 304). Nesta situação a correlação espacial entre os resíduos poderá ser analisada usando técnicas como estudo do semivariograma experimental, pelo Diagrama de Espalhamento de Moran e seu Índice e, constatada a dependência, a localização deverá ser considerada e o modelo será um modelo de regressão espacial.

A correlação serial entre as observações é definida como a correlação entre integrantes de séries de observações ordenadas no tempo (séries temporais) quando então os resíduos de duas observações guardam correlação diferente entre si diferente de zero

$$\text{Cov}(\varepsilon_i, \varepsilon_j | x_i, x_j) \neq 0, \quad i \neq j.$$

A correlação serial pode decorrer:

- inércia: quando os efeitos na alteração da variável  $X$  demoram a se manifestar na variável  $Y$  (muito comum em dados econômicos);
- forma funcional do modelo incorreta;
- variáveis importantes foram omitidas.

Assim como na situação onde há autocorrelação espacial, há ferramentas estatísticas apropriadas para se analisar dados provenientes de séries onde a autocorrelação está presente e se trabalhar a tendência existente.

A verificação da independência dos resíduos pode ser visualmente verificada por meio de alguns gráficos entre os quais resíduos *versus* o tempo ou ordem no qual as observações foram realizadas e observações *versus* o tempo ou ordem no qual foram realizadas (um gráfico sequencial). Quando as observações da amostra são independentes o que se espera é que seus resíduos apresentem-se aleatoriamente dispersos em torno do eixo horizontal em zero quando dispostos na sequência ou local em que foram coletadas. O que se espera é que nenhuma relação funcional seja percebida.

#### **B.1.11 Análise quanto à distribuição dos resíduos Normalidade**

Sob a premissa admitida de Normalidade, os estimadores de mínimos quadrados em (B.5) são os estimadores com variância mínima dentro de toda a classe de estimadores não viesados (*BLUE: Best Linear Unbiased Estimator*). A Normalidade dos resíduos  $\hat{\varepsilon}$  pode ser verificada informalmente por meio de vários modos gráficos dentre os quais destacam-se:

- pela comparação de suas frequências às frequências esperadas de uma distribuições Normal:
  - 68 %:  $\pm 1$  desvio padrão;
  - 90 %:  $\pm 1.65$  desvio padrão;
  - 95 %:  $\pm 1.96$  desvio padrão;
- gráficos de caixas;
- histogramas;
- gráficos dos quantis teóricos da distribuição Normal padronizada contra os quantis amostrais dos resíduos (Q-Q plot);

- gráfico com envoltória simulada (envelope) dos resíduos (RIPLEY, 1977).

Se os valores de uma amostra provêm de uma distribuição Normal, então os valores das estatísticas de ordem construídas com os resíduos e os  $Z_i$  correspondentes obtidos da distribuição Normal padrão são linearmente relacionados e, assim, o gráfico dos valores deve ter o aspecto aproximado de uma reta. Todavia observam-se que alguns aspectos desse gráfico diferentes de uma reta que sugerem ausência de Normalidade têm como provável causa:

- “S”: indica distribuições com caudas muito curtas, isto é, distribuições cujos valores estão muito próximos da média;
- “S invertido”: indica distribuições com caudas muito longas e, portanto, presença de muitos valores extremos;
- “J” ou “J invertido”: indicam distribuições assimétricas, positivas e negativas, respectivamente.

A análise do modelo com respeito à Normalidade de seus resíduos é, em muitos aspectos, mais difícil do que para as outras verificações. A menos que o tamanho da amostra seja muito grande ( $n \sim 300$ ) a variação aleatória impõe sérias dificuldades para se estudar a natureza da distribuição de probabilidade da variável em estudo (NETTER; WASSERMAN; KUTNER, 1983, p. 112).

Outros tipos de desvios podem também afetar a distribuição dos resíduos como quando a função é inadequada ou quando a variância não é constante. Assim, pequenos desvios dos resíduos em relação à distribuição Normal podem ser tolerados pois não causam problemas sérios na estimação do modelo.

Para uma análise formal da Normalidade há vários testes definidos:

- $K^2$  de D’agostino (Ralph D’agostino);
- Jarque-Bera (Carlos Jarque e Anil K. Bera);
- Anderson-Darling (Theodore Wilbur Anderson e Donald Alan Darling);
- Cramer-von Mises (H. Cramer e R.E. von Mises);
- Lilliefors (Hubert W. Lilliefors);
- Shapiro-Francia (Samuel Sandford Shapiro e S. Francia);
- $X^2$  de Karl Pearson;
- Kolmogorov-Smirnov (Andrey Kolmogorov e Nikolai Smirnov);
- teste de correlação linear entre os resíduos padronizados ordenados e os quantis teóricos da distribuição Normal padronizada;



- Shapiro-Wilk (Samuel Sandford Shapiro e Martin Bradbury Wilk);

As hipóteses do teste proposto por Samuel Sandford Shapiro e Martin Bradbury Wilk sobre a procedência da distribuição de uma determinada amostra aleatória são assim estabelecidas:

$$\begin{cases} H_0 : Y_i \sim \mathcal{N}(\mu, \hat{\sigma}^2); \\ H_1 : Y_i \not\sim \mathcal{N}(\mu, \hat{\sigma}^2). \end{cases}$$

Seja  $m^T = (m_1, \dots, m_n)$  o vetor dos valores esperados de estatísticas Normais padronizadas ordenadas e seja  $V = (v_{ij})$  a correspondente matriz de covariância. Ou seja, se  $x_1 < x_2 < \dots < x_n$  representa uma amostra aleatória de tamanho  $n$  de uma variável com distribuição Normal padronizada, então

$$\begin{aligned} \mathbb{E}(x)_i &= m_i \quad i = 1, 2, \dots, n; \text{ e,} \\ \text{Cov}(x_i, x_j) &= v_{ij} \quad i, j = 1, 2, \dots, n. \end{aligned}$$

Se o vetor das observações aleatórias ordenadas  $y_{(i)} = y_{(1)}, y_{(2)}, \dots, y_{(n)}$  procede de uma distribuição tal que  $Y_i \sim \mathcal{N}(\mu, \hat{\sigma}^2)$ , então ele poderá ser expresso como

$$y_i = \mu + \sigma x_i.$$

A partir dos resultados do Teorema dos Mínimos Quadrados Generalizados Samuel Sandford Shapiro e Martin Bradbury Wilk (SHAPIRO; WILK, 1965, p. 592) deduziram os estimadores  $\hat{\mu}$  e  $\hat{\sigma}$  da regressão anterior e formalizaram a estatística  $W$  e sua função distribuição cumulativa de probabilidade como

$$W = \frac{(\sum_{i=1}^n a_i y_{(i)})^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

em que  $\bar{y}$  é a média da amostra de tamanho  $n$ ;  $a_1, \dots, a_n$  são coeficientes tais que  $a_i = m^T V^{-1} / C$  e  $C$  é a constante Normalizadora, o melhor estimador não viesado para o coeficiente angular da regressão dos valores observados ordenados  $y_i$  sobre os valores esperados  $m_i$  das estatísticas Normais padronizadas, de tal modo que os coeficientes lineares são padronizados:  $C^2 = m^T V^{-1} V^{-1} m$ .

A implementação do teste de Shapiro-Wilk sem o uso de recursos pode ser feita seguindo-se o roteiro (SHAPIRO; WILK, 1965, 602):

- ordene os valores da amostra extraída:  $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}$ ;
- calcule  $S$ , a soma dos quadrados dos desvios:  $S^2 = \sum_{i=1}^n (y_{(i)} - \bar{y})^2$ ;
- se  $n$  é par:  $b = \sum a_{(n-i-1)} (y_{(n-i+1)} - y_{(i)})$ ;

- se  $n$  é ímpar tome  $n = 2k + 1$  e:  $b = a_{(n)}(y_{(n)} - y_{(1)}) + \dots + a_{(k+2)}(y_{(k+2)} - y_{(k)})$ ;
- os coeficientes  $a_{n-i+1}$  encontram-se na Tabela 5 (SHAPIRO; WILK, 1965, p. 603);
- determine a estatística do teste:  $W = b^2/s^2$ ;
- compare o valor calculado da estatística do teste com os valores da Tabela 6 (SHAPIRO; WILK, 1965, p. 605) para vários níveis de significância da distribuição cumulativa teórica de  $W$ ;
- avaliar o nível de significância associado ao valor calculado e decidir sobre a rejeição ou não da hipótese nula.

### B.1.12 Análise quanto à existência de observações discordantes

Observações discordantes são observações extremas, muito afastadas da tendência sugerida pelas demais observações que compõem a amostra e já na análise descritiva pode-se averiguar se essa discordância decorre de variados erros associados à medição ou observação (operadores ou observadores diferentes, equipamentos descalibrados, instrumentos de medição diversos) ou até mesmo no registro das informações. Nessas situações as observações podem ser descartadas posto não se mostrarem fidedignos. Todavia na maior parte dos experimentos a identificação desse tipo de erro na etapa descritiva não é possível e, nestas situações, a análise dos resíduos gerados pelo modelo na estimação de cada observação é a principal ferramenta.

No diagnóstico do modelo quanto à presença de desvios isolados deve-se considerar três conceitos diferentes (TURKMAN; SILVA, 2000, p. 84):

- observações com grande repercussão (*leverage*): a repercussão de uma observação é a medida do efeito que ela exerce nos valores preditos pelo modelo;
- observações inconsistentes (*outliers*): uma observação consistente deve seguir a tendência sugerida pelas demais observações;
- observações com grande influência: uma observação inconsistente com grande repercussão é influente se uma ligeira alteração no seu valor (ou sua exclusão do modelo) produz alterações significativas nas estimativas dos parâmetros do modelo.

Uma técnica gráfica, para se verificar a presença desse tipo de observações, é pela análise dos resíduos *versus* valores estimados e resíduos *versus* as variáveis preditoras. Gráficos com os resíduos “estudentizados” são particularmente úteis para se distinguir as observações cujos resíduos distem muitos desvios padrão da média zero.

Os padrões visuais desejados nos gráficos acima devem assemelhar-se a pontos dispersos de modo aproximadamente simétrico em torno do eixo horizontal em zero,

que não exibam, sistematicamente, nenhum padrão de crescimento ou decaimento na amplitude visual de sua dispersão. Uma regra comum para amostras grandes ( $n > 30$ ) é considerar resíduos “estudentizados” com afastamentos em valor absoluto de quatro ou mais desvios padrão serem discordantes.

Uma questão relacionada à medida da influência que não pode abordada pelas técnicas gráficas acima é a o estudo da variação nos resultados de uma análise quando o modelo é modificado. Se uma observação for excluída, por exemplo, os resultados do novo modelo baseado do conjunto de dados reduzido podem ser bastante diferentes daqueles baseados no conjunto completo de dados e a esse estudo Cook chamou de estudo da influência (COOK; WEISBERG, 1982, p. 101).

A estatística da Distância de Cook ( $D_i$ ) mede a influência que um determinado elemento da amostra exerce nas estimativas dos parâmetros do modelo pela alteração que estes sofrem caso aquela seja suprimida da base de dados. Formalmente, a Distancia está assim definida (WOOD, 2017, p. 26)

$$D_i = \frac{\sum_{j=1}^n (\hat{Y}_j - \hat{Y}_{j(i)})^2}{p\hat{\sigma}^2} = \frac{e_i^2}{p\hat{\sigma}^2} \left[ \frac{h_{ii}}{(1 - h_{ii})^2} \right], \quad (\text{B.16})$$

em que  $\hat{Y}_j$  é a estimativa do modelo estimado com a amostra completo para  $j$ -ésima observação;  $\hat{Y}_{j(i)}$  é a estimativa do modelo estimado com a amostra reduzida (sem o  $i$ -ésimo dado) para  $j$ -ésima observação;  $\hat{\sigma}^2$  é o quadrado médio dos resíduos;  $h_{ii}$  é o  $i$ -ésimo elemento da diagonal da matriz de projeção  $\mathbf{H}$ ;  $e_i$  é o resíduo bruto associado à  $i$ -ésima observação e  $p$  é o número de parâmetros estimados no modelo ( $p = k + 1$ ).

Os limites mais usualmente empregados para os valores da estatística da Distância de Cook são

- $D_i > 1$  (COOK; WEISBERG, 1982, p. 118); alternativamente,
- $D_i > 4/n$  ou  $D_i > 4/(n-k-1)$  também são frequentemente citados.

A alavancagem é a medida da contribuição individual de uma observação nos valores estimados por um modelo. Uma observação pode ser considerada com grande capacidade de alavancar as estimativas do modelo quando esta exerce uma forte influência no seu valor ajustado.

Estando os valores dos elementos da diagonal da matriz de projeção  $\mathbf{H}$  ( $h_{ii}$ ) limitados à faixa  $(0, 1)^7$ , quanto mais próximo esta valor estiver de 1 mais a estimativa  $\hat{Y}_i$  será determinada pela  $i$ -ésima observação (BOLLEN; JACKMAN, 1985, 515).

<sup>7</sup>A faixa de valores é:  $1/n < h_{ii} < 1/c$ , onde  $n$  é o número de observações e “c” é o número de vezes que a  $i$ -ésima linha da matriz  $\mathbf{X}$  é replicada. Para modelos sem a constante  $\beta_0$ , o limite inferior deve ser substituído por zero. O máximo valor para  $h_{ii}$  será “um” quando não houver replicação de nenhum  $x_i$  (COOK; WEISBERG, 1982, p. 13). A demonstração pode ser vista em (BOLLEN; JACKMAN, 1985, p. 540)

De modo análogo à estatística da Distância de Cook, os critérios mais comuns (HOAGLIN; WELSCH, 1978, p. 18) e (BOLLEN; JACKMAN, 1985, 519) para estabelecer um valor crítico para  $h_{ii}$  são:

- $h > 2p/n$ : quando  $p > 10$  e  $n - p > 50$ ; ou, nas situações contrárias,
- $h > 3p/n$ .

## B.2 MODELOS LINEARES GENERALIZADOS (GLM)

### B.2.1 Família exponencial de distribuições uniparamétricas

A função densidade (ou cumulativa) de probabilidade característica de uma distribuição pertencente à família exponencial uniparamétrica deve ser especificada na forma

$$f(y; \theta) = h(y)e^{[\eta(\theta)t(y)-b(\theta)]}, \quad (\text{B.17})$$

em que  $\eta(\theta)$ ,  $b(\theta)$ ,  $t(y)$  e  $h(y)$  são funções específicas e não únicas com valores no subconjunto dos Reais e pode-se demonstrar pelo Teorema da Fatoração de Neyman-Fisher que  $t(y)$  é uma estatística suficiente para  $\theta$  (CORDEIRO; DEMÉTRIO; MORAL, 2015, p. 2). A família exponencial na sua forma canônica é definida admitindo-se que as funções  $\eta(\theta)$  e  $t(y)$  são iguais à função identidade. Assim, assumem a forma

$$f(y; \theta) = h(y)e^{[\theta y - b(\theta)]},$$

em que  $\theta$  é o parâmetro canônico da distribuição. Como exemplos tem-se as distribuições:

- Poisson:  $Pois(\theta)$

$$f(y; \theta) = \frac{e^{-\theta}\theta^y}{y!} = \frac{1}{y!}e^{[\log(\theta)y - \theta]},$$

em que  $e(\theta) = \log(\theta)$ ;  $b(\theta) = \theta$ ;  $t(y) = y$  e  $h(y) = 1/y!$ .

- Binomial:  $B(m, \theta)$

$$f(y; \theta) = \binom{m}{y}\theta^y(1 - \theta)^{m-y} = \binom{m}{y}e^{[\log(\frac{\theta}{1-\theta})y + m \log(1-\theta)]},$$

em que  $\eta(\theta) = \log[\theta/(1-\theta)]$ ;  $b(\theta) = -m \log(1 - \theta)$ ;  $t(y) = y$  e  $h(y) = \binom{m}{y}$ .

- Rayleigh:  $Ray(\theta)$

$$f(y; \theta) = \frac{y}{\theta^2}e^{-\frac{y^2}{2\theta^2}} = ye^{[-\frac{1}{2\theta^2}y^2 - 2 \log(\theta)]},$$

em que  $\eta(\theta) = -1/2\theta^2$ ;  $b(\theta) = 2 \log(\theta)$ ;  $t(y) = y^2$  e  $h(y) = y$ .

Admitir as funções  $\eta(\theta)$  e  $t(y)$  de (B.17) como iguais à identidade possibilita exprimir a função densidade (ou cumulativa) de probabilidade característica na sua chamada forma canônica com  $\theta$  sendo seu parâmetro canônico

$$f(y; \theta) = h(y)e^{[\theta y - b(\theta)]}. \quad (\text{B.18})$$

Para uma única observação o logaritmo da função de verossimilhança de uma variável aleatória pertencente a uma distribuição expressa sob a forma (B.18) será

$$l(\theta) = \theta y - b(\theta) + \log[h(y)].$$

A partir dos resultados associados à função Escore de Fisher em (B.9) e (B.11) segue-se que

$$\begin{aligned} \mathbb{E}(Y) &= b'(\theta); \text{ e} \\ \text{Var}(Y) &= b''(\theta). \end{aligned}$$

É possível demonstrar que a distribuição conjunta de variáveis aleatórias independentes e identicamente distribuídas  $Y_1, \dots, Y_n$  com função densidade (ou cumulativa) de probabilidade dadas na forma (B.17) também é um modelo da família exponencial

$$f(y_1, \dots, y_n; \theta) = [\prod_{i=1}^n h(y_i)] e^{[\eta(\theta) \sum_{i=1}^n t(y_i) - nb(\theta)]},$$

e que a estatística suficiente desse modelo:  $\sum_{i=1}^n T(Y_i)$  pertence à família exponencial uniparamétrica qualquer que seja  $n$  (CORDEIRO; DEMÉTRIO; MORAL, 2015, p. 4).

As distribuições de probabilidade usadas nos Modelos Lineares Generalizados possuem um parâmetro adicional de tal forma que as distribuições da família exponencial passam a ser expressas na forma

$$f(y; \theta, \phi) = e^{\left\{ \frac{[y\theta - b(\theta)]}{a(\phi)} + c(y, \phi) \right\}},$$

em que  $a(\cdot)$ ,  $b(\cdot)$  e  $c(\cdot)$  são funções conhecidas,  $\theta$  é o parâmetro canônico da locação (média) da distribuição e  $\phi > 0$  é um parâmetro extra de perturbação por vezes chamado de parâmetro de dispersão (tal como  $\sigma^2$  de uma distribuição Normal).

A derivação de expressões para os primeiro e segundo momentos da função de log-verossimilhança  $l(\theta|y)$  em termos da média e variância de  $y$  e  $a(\phi)$  parte dos resultados (B.9) e (B.11)

$$\mathbb{E} \left[ \frac{dl(\theta|y)}{d\theta} \right] = 0; \text{ e} \quad (\text{B.19})$$

$$\mathbb{E} \left[ \frac{d^2 l(\theta|y)}{d\theta^2} \right] = -\mathbb{E} \left[ \frac{dl(\theta|y)}{d\theta} \right]^2, \quad (\text{B.20})$$

em que a diferenciação sob o sinal da integração usado nas derivadas pode ser justificado pelo Teorema 9 de Lehmann (1959) (NELDER; WEDDERBURN, 1972, p. 371).

A partir de (B.19) pode-se mostrar (CORDEIRO; DEMÉTRIO; MORAL, 2015, pp. 8-9) que a média de  $y$  será dada por

$$\mu = \mathbb{E}(y) = \frac{db(\theta)}{d\theta} = b'(\theta). \quad (\text{B.21})$$

De (B.21) tem-se a função que relaciona o parâmetro canônico  $\theta$  com a média  $\mu$ , dada por  $\theta = q(\mu)$ , onde  $q(\cdot)$  é a inversa de  $b(\cdot)$ '.

De (B.20) pode-se mostrar (CORDEIRO; DEMÉTRIO; MORAL, 2015, pp. 8-9) que a variância de  $y$  é dada por :

$$\begin{aligned} \text{Var}(y) &= \frac{d^2b(\theta)}{d\theta^2} a(\phi) \\ &= a(\phi) b'' \\ &= b'' \frac{\phi}{w} \end{aligned} \quad (\text{B.22})$$

Em princípio a função  $a(\phi)$  pode assumir qualquer forma e na teoria dos modelos generalizados isso não resulta em dificuldade se  $\phi$  é conhecido. Todavia, caso  $\phi$  não o seja, torna-se mais interessante adotar  $a(\phi) = \frac{\phi}{w}$ , onde  $w$  é uma constante conhecida e isso possibilita modelos baseado na distribuição Normal com variância não constante. Para a maior parte das situações  $w = 1$ .

A variância de  $y$  pode ser decomposta em duas fontes. A que se relaciona apenas com a média e independe de  $a(\cdot)$  é dada pela função da variância com a média:

$$\begin{aligned} V(\mu) &= \frac{\text{Var}(y)}{a(\phi)} \\ &= \frac{d^2b(\theta)}{d\theta^2} \\ &= \frac{d\mu}{d\theta} \end{aligned}$$

### B.2.2 Algumas distribuições pertencentes à família exponencial com parâmetro de dispersão

- Normal:  $\mathcal{N}(\mu, \sigma^2)$

$$\begin{aligned} f(y; \mu, \sigma) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left[ \frac{-(y-\mu)^2}{2\sigma^2} \right]} \\ &= e^{\left( \frac{(y\mu - \frac{\mu^2}{2})}{\sigma^2} \right) - \frac{1}{2} \left[ \frac{y^2}{\sigma^2} + \log(2\pi\sigma^2) \right]} \end{aligned}$$

em que  $-\infty < y < \infty$ ;  $\theta = \mu$ ;  $\mu(\theta) = \theta$ ;  $a(\phi) = \phi$   $b(\theta) = \mu^2/2 = \theta^2/2$ ;  $c(y, \phi) = -(1/2)\log(2\pi\sigma^2) + y^2/2\sigma^2$ ;  $\phi = \sigma^2$  e  $V(\mu) = 1$ . A média e variância de  $y$  são dadas por

$$\mathbb{E}(Y) = \frac{db(\theta)}{d\theta} = \mu \quad \text{Var}(Y) = \frac{d^2b(\theta)}{d\theta^2} a(\phi) = \sigma^2$$

- Poisson:  $Pois(\mu)$

$$\begin{aligned} f(y; \mu) &= \frac{\mu^y e^{-\mu}}{y!} \\ &= e^{[y \log \mu - \mu - \log(y!)]} \end{aligned}$$

em que  $y = 0, 1, 2, \dots$ ;  $\theta = \log(\mu)$ ;  $\mu(\theta) = e^\theta$ ;  $a(\phi) = \phi$ ;  $b(\theta) = \mu = e^\theta$ ;  $c(y, \phi) = -\log(y!)$ ;  $\phi = 1$  e  $V(\mu) = \mu$ . A média e variância de  $y$  são dadas por:

$$\mathbb{E}(Y) = \frac{db(\theta)}{d\theta} = \mu \quad \text{Var}(Y) = \frac{d^2b(\theta)}{d\theta^2} a(\phi) = \mu$$

- Binomial:  $B(m, \pi)$

$\pi \in (0, 1)$ ;  $m = (1, 2, \dots)$ ;  $y = (1, 2, \dots, m)$  e  $\mu = m\pi$

$$\begin{aligned} f(y; \mu) &= \binom{m}{y} \mu^y (1 - \mu)^{(m-y)} \\ &= e^{\{y \log[\frac{\mu}{(1-\mu)}] + m \log(1 - \mu) + \log[\binom{m}{y}]\}} \end{aligned}$$

em que  $y = 0, 1, 2, \dots, m$ ;  $\theta = \log[\mu/(m-\mu)]$ ;  $\mu(\theta) = m[e^\theta/(1+e^\theta)]$ ;  $a(\phi) = \phi$ ;  $b(\theta) = m \log(1 - \pi) = m \log(1 + e^\theta)$ ;  $c(y, \phi) = \log[\binom{m}{y}]$ ;  $\phi = 1$  e  $V(\mu) = (\mu/m)(m - \mu)$ . A média e variância de  $y$  são dadas por:

$$\mathbb{E}(Y) = \frac{db(\theta)}{d\theta} = \mu \quad \text{Var}(Y) = \frac{d^2b(\theta)}{d\theta^2} a(\phi) = \mu(1 - \mu)$$

### B.2.3 Estrutura de um Modelo Linear Generalizado

Sob a expressão “modelo linear”, um modelo estatístico usualmente incorpora tanto os componentes sistemáticos quanto os aleatórios enquanto que sob a estrutura de um Modelo Linear Generalizado apenas o componente sistemático é considerado. Assim, a estrutura de um Modelo Linear Generalizado é formada por 3 elementos bem característicos (NELDER; WEDDERBURN, 1972, p. 372):

- o componente aleatório (variável dependente): uma variável aleatória cuja distribuição com parâmetro  $\theta$  pertence à classe da família exponencial;
- o componente sistemático: um conjunto variáveis independentes incorporadas ao modelo na forma de combinação linear (preditor linear);
- uma função de ligação adequada; isto é, uma função monótona e diferenciável, responsável pela conexão dos dois componentes anteriormente citados.



Um modelo de regressão clássico de regressão linear com distribuição Normal para o erro tem sua  $i$ -ésima observação expressa como

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_k x_{ik} + \varepsilon_i; \text{ com } \varepsilon_i \stackrel{\text{ind}}{\sim} \mathcal{N}(0, \sigma^2), i = 1, \dots, n.$$

A especificação acima é equivalente a

$$Y_i \stackrel{\text{ind}}{\sim} \mathcal{N}(\mu_i, \sigma^2); \text{ com } \mu_i = \beta_0 + \beta_1 x_{i1} + \beta_k x_{ik} i = 1, \dots, n,$$

e vetorialmente pode ser expressa como

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}; \text{ com } \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2); \\ \boldsymbol{\mu} &= \mathbf{X}\boldsymbol{\beta}; \\ \boldsymbol{\mu} &= \mathbb{E}(\mathbf{Y}), \end{aligned}$$

em que se identificam as seguinte matrizes

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{(n,1)}; \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}_{(n,1)}; \mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \dots & x_{1,k} \\ 1 & x_{2,1} & x_{2,2} & \dots & x_{2,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \dots & x_{n,k} \end{bmatrix}_{(n,k+1)}; \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1,1)}; \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{(n,1)}$$

em que  $\mathbf{Y}$  é o vetor das variáveis das respostas com dimensão  $(n \times 1)$ ;  $\boldsymbol{\mu}$  é o vetor das médias esperadas para as respostas com dimensão  $(n \times 1)$ ;  $\boldsymbol{\varepsilon}$  é o vetor dos erros admitidos como  $\sim \mathcal{N}(0, 1)$  e com dimensão  $(n \times 1)$ ;  $\boldsymbol{\beta}$  é o vetor dos parâmetros desconhecidos com dimensão  $(n \times 1)$  e  $\mathbf{X}$  é a matriz de delineamento de dimensão  $(n \times p)$  (com  $p = k + 1$  e tendo “uns” em sua primeira coluna se uma constante  $\beta_0$  for adicionada ao modelo e os valores das  $k$  covariáveis nas demais colunas, ou  $p = k$  caso não o seja).

Um Modelo Linear Generalizado tem sua  $i$ -ésima observação expressa como

$$\begin{aligned} Y_i &\stackrel{\text{ind}}{\sim} \mathcal{E}(\mu_i, \phi) \quad \text{tal que:} \\ g(\mu_i) &= \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} \quad i = 1, \dots, n, \end{aligned} \tag{B.23}$$

em que  $\mathcal{E}$  é uma distribuição pertencente à família exponencial de distribuições com parâmetro de dispersão  $\phi$ .

Em forma vetorial, o modelo  $\mathbf{Y} \stackrel{\text{ind}}{\sim} \mathcal{E}(\boldsymbol{\mu}, \phi)$  passa a ser expresso como

$$\boldsymbol{\eta} = g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta},$$

em que  $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$  é o vetor preditor linear, uma combinação linear tal que seu  $i$ -ésimo elemento é  $\eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$  e  $g(\boldsymbol{\mu})$  é o vetor da função de ligação que conecta a média  $\mu$  ao preditor linear  $\boldsymbol{\eta}$  tal que seu  $i$ -ésimo elemento é  $g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$ .

A partir de (B.23) observa-se:

- o  $i$ -ésimo elemento do vetor de médias pode ser escrito como  $\mu_i = g^{-1}(\eta_i)$ ;
- para uma variável aleatória Normal: parâmetro canônico  $\theta$  igual a  $\mu$  e variância  $\sigma^2$ ), sob uma ligação com a função identidade:  $g(\mu_i) = \eta_i$ , recai-se no Modelo Clássico de Regressão Linear com distribuição Normal para os erros;
- a função de ligação não transforma diretamente a variável ( $\mathbf{Y}$ ) mas sim seu valor esperado:  $\mathbb{E}(Y) = \mu$ .

Um Modelo Linear Generalizado vem possibilitar a análise das situações onde a estrutura aditiva do modelo clássico (B.2), a ligação pela simples função identidade, a distribuição Normal para a resposta e o pressuposto de homocedasticidade da variância não são verificados.

#### B.2.4 Especificação de um Modelo Linear Generalizado

A especificação de um Modelo Linear Generalizado envolve a escolha de uma distribuição apropriada para a variável respostas, de uma função de ligação conveniente e do modo como as variáveis explanatórias serão incorporadas ao preditor linear.

É ainda na fase de planejamento que o componente sistemático deve ter sua estrutura definida a partir do experimento a ser realizado:

- modelo de regressão (simples ou múltipla);
- análise de variância sob diversos tipos de delineamento; ou,
- análise de covariância.

A escolha da distribuição apropriada do componente aleatório deverá considerar fundamentalmente o tipo de medidas da variável resposta obtidas no experimento: contínuas: simétricas ou não em  $\mathcal{R}$  (ou ainda existindo apenas em subconjuntos), categóricas ou discretas.

As funções de ligação foram introduzidas para os Modelos Lineares Generalizados mas bem podem ser adotadas em todo tipo de regressão desde que sejam apropriadas; isto é, desde que assegurem que as estimativas dos parâmetros permaneçam no espaço paramétrico de  $\theta$ . Assim, se um parâmetro pode assumir valores apenas na faixa  $0 < \theta < \infty$ , a função logaritmo é apropriada pois  $\eta = \log(\theta)$  produz valores em  $\mathcal{R}$  mas sua inversa  $\exp(\eta)$  apenas no espaço paramétrico de  $\theta$ :  $\mathcal{R}^+$  (STASINOPOULOS et al., 2017, p. 11).

A função de ligação canônica para uma determinada distribuição é tal que  $g(\mu) = \theta = \eta$  em que  $\theta$  é o parâmetro canônico dessa distribuição. Assim, o uso da ligação canônica implica que  $\theta_i = \mathbf{X}_i\boldsymbol{\beta}$  e também que haja coincidência nos resultados da estimação dos parâmetros quer se use o Método do score de Fisher ou o Método de Newton-Raphson e em

outras propriedades teóricas interessantes por simplificar o processo de estimação por máxima verossimilhança dos parâmetros do modelo (MYERS et al., 2010, p. 208) e (WOOD, 2017, p. 111).

A tabela B.1 apresenta as ligações canônicas de algumas distribuições.

Tabela B.1: Ligações canônicas para Modelos Lineares Generalizados

Distribuição	Ligação canônica	Denominação
Normal	$\eta_i = \mu_i$	Função identidade
Binomial	$\eta_i = \log[\pi_i/(1-\pi_i)]$	Função logística
Poisson	$\eta_i = \log(\mu_i)$	Função logaritmo
Exponencial	$\eta_i = 1/\mu_i$	Função recíproca ou inversa
Gama	$\eta_i = 1/\mu_i$	Função recíproca ou inversa

Fonte: Myers et al. (2010, p. 207)

### B.2.5 Estimação

Um Modelo Linear Generalizado modela a relação de um vetor de respostas de  $n$  variáveis independentes  $Y$  tais que  $\boldsymbol{\mu} \equiv \mathbb{E}(\mathbf{Y})$ , por meio de uma função de ligação  $g(\mu_i)$  com um preditor linear  $\mathbf{X}_i\boldsymbol{\beta}$ , sendo  $Y_i \stackrel{ind}{\sim} f_{\theta_i}(y_i)$  onde  $f_{\theta_i}(y_i)$  indica uma distribuição pertencente à família exponencial com parâmetro canônico  $\theta$  determinado por  $\mu$  por meio de (B.21) e portanto, por  $\boldsymbol{\beta}$  (WOOD, 2017, pp. 105-108).

Sob independência, a função de verossimilhança para  $\boldsymbol{\beta}$  é

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n f_{\theta_i}(y_i),$$

portanto, a log-verossimilhança para  $\boldsymbol{\beta}$  será

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n \log[f_{\theta_i}(y_i)] = \sum_{i=1}^n \left\{ \frac{[y_i\theta_i - b_i(\theta_i)]}{a_i(\theta_i)} + c_i(\phi, y_i) \right\}, \quad (\text{B.24})$$

a qual estabelece a dependência de  $\boldsymbol{\beta}$  à direita da igualdade dá-se pela dependência de  $\boldsymbol{\theta}$  em  $\boldsymbol{\beta}$ . As funções  $a_i(\cdot)$ ,  $b_i(\cdot)$  e  $c_i(\cdot)$  podem variar para cada observação  $i$ , todavia  $\phi$  é admitido constante.

Nas situações onde  $a_i(\phi)$  possa ser expressa como  $a_i(\phi) = \phi/w_i$  onde  $w_i$  é uma constante conhecida (usualmente um), segue-se de (B.24)

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n w_i \left\{ \frac{[y_i\theta_i - b_i(\theta_i)]}{\phi} + c_i(\phi, y_i) \right\}. \quad (\text{B.25})$$

A maximização da função (B.25) pelo Método de Newton (WOOD, 2017, p. 76) requer o vetor gradiente e a matriz Hessiana de  $l$

$$\frac{\partial l}{\partial \beta_j} = \frac{1}{\phi} \sum_{i=1}^n w_i \left[ y_i \frac{\partial \theta_i}{\partial \beta_j} - b'_i(\theta_i) \frac{\partial \theta_i}{\partial \beta_j} \right],$$

e pela regra da cadeia

$$\begin{aligned} \frac{\partial \theta_i}{\partial \beta_j} &= \frac{d\theta_i}{d\mu_i} \frac{\partial \mu_i}{\partial \beta_j} \\ &= \frac{d\theta_i}{d\mu_i} \frac{\partial \eta_i}{\partial \beta_j} \frac{d\mu_i}{d\eta_i} \\ &= \frac{X_{ij}}{g'(\mu_i) b''(\theta_i)}. \end{aligned}$$

A última igualdade final decorre do fato de que  $d\eta_i/d\mu_i = g'(\mu_i)$  (de modo que  $d\mu_i/d\eta_i = 1/g'(\mu_i)$ ),  $\partial \eta_i/\partial \beta_j = X_{ij}$  e diferenciando-se (B.21),  $d\theta_i/d\mu_i = 1/b''(\theta_i)$ .

Após alguma manipulação e trazendo-se (B.21) e (B.22) segue-se

$$\begin{aligned} \frac{\partial l}{\partial \beta_j} &= \frac{1}{\phi} \sum_{i=1}^n \frac{y_i - b'_i(\theta_i)}{g'(\mu_i) b''(\theta_i)/w_i} X_{ij} \\ &= \frac{1}{\phi} \sum_{i=1}^n \frac{y_i - \mu_i}{g'(\mu_i) V(\mu_i)} X_{ij}. \end{aligned}$$

Diferenciando-se novamente segue-se

$$\begin{aligned} \frac{\partial^2 l}{\partial \beta_j \partial \beta_k} &= -\frac{1}{\phi} \sum_{i=1}^n \left\{ \frac{X_{ik} X_{ij}}{g'(\mu_i)^2 V(\mu_i)} + \frac{(y_i - \mu_i) V'(\mu_i) X_{ik} X_{ij}}{g'(\mu_i)^2 V(\mu_i)^2} + \frac{(y_i - \mu_i) X_{ij} g''(\mu_i) X_{ik}}{g'(\mu_i)^3 V(\mu_i)} \right\} \\ &= -\frac{1}{\phi} \sum_{i=1}^n \frac{X_{ik} X_{ij} \alpha(\mu_i)}{g'(\mu_i)^2 V(\mu_i)}, \end{aligned}$$

em que  $\alpha(\mu_i) = 1 + (y_i - \mu_i) \{V'(\mu_i)/V(\mu_i) + g''(\mu_i)/g'(\mu_i)\}$ . Observa-se que a expressão para  $E(\partial^2 l/\partial \beta_j \partial \beta_k)$  é a mesma, apenas com  $\alpha(\mu_i) = 1$ .

Assim, definindo-se a matriz  $\mathbf{W} = \text{diag}(w_i)$  com  $w_i = \alpha(\mu_i)/[g'(\mu_i)^2 V(\mu_i)]$ , a Hessiana da log-verossimilhança é  $(1/\phi)(-\mathbf{X}^T \mathbf{W} \mathbf{X})$ , enquanto que a Hessiana desejada é obtida fazendo-se  $\alpha(\mu_i) = 1$ . Os pesos computados com  $\alpha(\mu_i) = 1$  serão denominados como pesos de Fisher.

Definindo-se  $\mathbf{G} = \text{diag}[g'(\mu_i)/\alpha(\mu_i)]$  o gradiente do vetor de log-verossimilhança pode ser expresso como  $(1/\phi)[\mathbf{X}^T \mathbf{W} \mathbf{G}(\mathbf{y} - \boldsymbol{\mu})]$ .

Assim, uma nova iteração no Método de Newton será expressa como

$$\begin{aligned} \boldsymbol{\beta}^{[m+1]} &= \boldsymbol{\beta}^{[m]} + (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{G}(\mathbf{y} - \boldsymbol{\mu}) \\ &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \{ \mathbf{G}(\mathbf{y} - \boldsymbol{\mu}) + \mathbf{X} \boldsymbol{\beta}^{[m]} \} \\ &= (\mathbf{X}^T \mathbf{W}^{[m]} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{[m]} \mathbf{z}^{[m]} \end{aligned}$$

em que  $z_i = g'(\mu_i)(y_i - \mu_i)/\alpha(\mu_i) + \eta_i$  (lembrando que  $\boldsymbol{\eta} \equiv \mathbf{X} \boldsymbol{\beta}$ ).

Essa nova iteração possui a mesma forma quer se use o método de Newton ou o escore de Fisher, no qual a esperança da Hessiana (a matriz de informação de Fisher) substitui a Hessiana e por essa razão fez-se  $\alpha(\mu_i) = 1$ .

Em qualquer uma das situações deve-se notar que as equações das iterações são estimativas de mínimos quadrados de  $\beta$  decorrentes da minimização de uma soma quadrados ponderados

$$\sum_{i=1}^n w_i (z_i - \mathbf{X}_i \beta)^2.$$

Assim, os Modelos Lineares Generalizados podem ser estimados iterativamente por um algoritmo que minimize a soma de quadrados cujos pesos serão atualizados a cada ciclo.

### B.2.5.1 Elementos do algoritmo

Para a implementação do algoritmo de estimação parâmetros (mínimos quadrados ponderados iterativo) considere os elementos a seguir onde  $[m]$  é a  $m$ -ésima iteração e  $\hat{\beta}^{[m]}$  o vetor das estimativas dos parâmetros nessa iteração particular (cf. Figura B.4)

$$\hat{\beta}^{[m]} = \begin{bmatrix} \hat{\beta}_0^{[m]} \\ \hat{\beta}_1^{[m]} \\ \vdots \\ \hat{\beta}_k^{[m]} \end{bmatrix}_{(k+1,1)}.$$

A matriz  $\mathbf{X}$  é a matriz de delineamento

$$\mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \dots & x_{1,k} \\ 1 & x_{2,1} & x_{2,2} & \dots & x_{2,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \dots & x_{n,k} \end{bmatrix}_{(n,k+1)}.$$

A matriz  $\mathbf{W}$  é a matriz de pesos

$$\mathbf{W}^{[m]} = \begin{bmatrix} w_{11}^{[m]} & 0 & 0 & \dots & 0 \\ 0 & w_{22}^{[m]} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & w_{nn}^{[m]} \end{bmatrix}_{(n,n)},$$

cujo  $i$ -ésimo elemento da diagonal principal  $w_{ii}^{[m]}$  na  $m$ -ésima iteração é dado por

$$w_{ii}^{[m]} = \frac{1}{\text{Var}(y_i)} \left[ \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^{[m]} \right]^2 \quad i = 1, \dots, n.$$

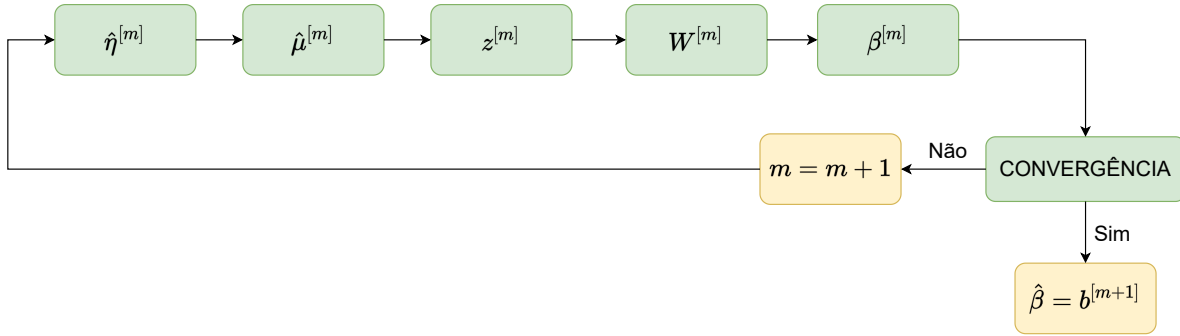
O vetor de ajuste  $z$  para a  $m$ -ésima iteração

$$z^{[m]} = \begin{bmatrix} z_1^{[m]} \\ z_2^{[m]} \\ \vdots \\ z_n^{[m]} \end{bmatrix}_{(n,1)},$$

cujo  $i$ -ésimo elemento  $z_i^{[m]}$  na  $m$ -ésima iteração é dado por

$$z_i^{[m]} = \hat{\eta}_i^{[m]} + (y_i - \hat{\mu}_i^{[m]}) \left( \frac{\partial \eta_i}{\partial \mu_i} \right)^{[m]} \quad i = 1, \dots, n.$$

Figura B.4: Relação de dependências dos elementos do algoritmo de estimação



Fonte: Adaptado de Vieira (2004, p. 75)

### B.2.5.2 Ciclo iterativo

A estimação de dos parâmetros de um Modelo Linear Generalizado segue assim uma sequência de passos para a minimização da soma de quadrados ponderados por pesos calculados de modo iterativo de tal modo que a diferença entre duas estimativas sucessivas  $\hat{\beta}^{[m+1]}$  e  $\hat{\beta}^{[m]}$  atenda a alguma medida estabelecida como tolerância para a convergência, como a *deviance* ou a norma dos vetores das estimativas (cf. Figura B.5).

- uma estimativa inicial para  $\hat{\beta}$  é calculada tomando-se como uma estimativa inicial para o vetor de médias as próprias observações (NELDER; WEDDERBURN, 1972, p. 374):  $\hat{\mu}_i^{[0]} = y_i + \delta_i$  e onde  $\delta_i$ , comumente é fixado em zero, pode ser arbitrado como uma pequena constante de modo a assegurar que  $\hat{\eta}_i$  seja finito; com esse resultado pode-se então calcular uma estimativa inicial para  $\hat{\eta}_i^{[0]}$ :

$$\hat{\eta}_i^{[0]} = g(\hat{\mu}_i^{[0]})$$

- a seguir calculam-se os elementos do vetor de ajuste neste primeiro ciclo:

$$z_i^{[0]} = \hat{\eta}_i^{[0]} + g'(\hat{\mu}_i^{[0]})(y_i - \hat{\mu}_i^{[0]})$$

e os elementos da matriz pesos para essa etapa:

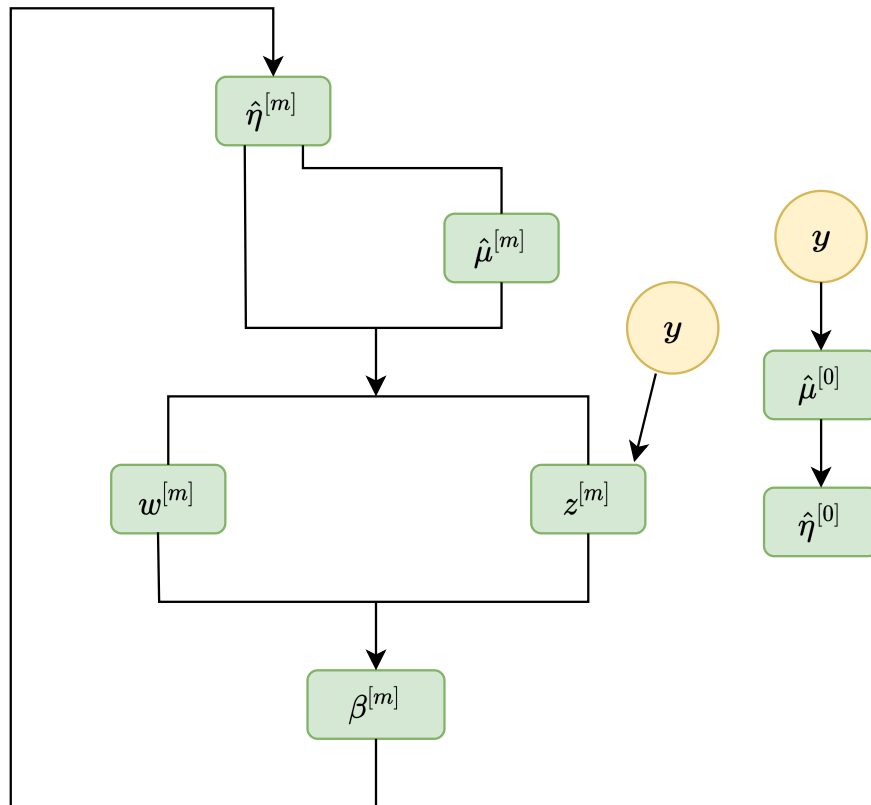
$$w_i^{[0]} = \frac{\alpha(\hat{\mu}_i^{[0]})}{V(\hat{\mu}_i^{[0]})g'(\hat{\mu}_i^{[0]})^2}$$

- estima-se a primeira aproximação para  $\hat{\beta}$  a qual minimiza a soma dos quadrados ponderados pelo vetor  $z$  e pela matriz de pesos  $\mathbf{W}$ :

$$\hat{\beta}^{[1]} = (\mathbf{X}^T \mathbf{W}^{[0]} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{[0]} \mathbf{z}^{[0]}$$

- atualizar  $\hat{\eta}^{[0]}$  para  $\hat{\eta}^{[1]} = \mathbf{X}\hat{\beta}^{[1]}$  e  $\hat{\mu}_i^{[1]} = g^{-1}(\hat{\eta}_i^{[1]})$ .

Figura B.5: Fluxograma do algoritmo de estimação



Fonte: Adaptado de Vieira (2004, p. 75)

### B.2.6 Desvio (*deviance*)

Para qualquer conjunto de dados pode-se estabelecer dois modelos extremos. O primeiro deles é usualmente denominado por modelo nulo (minimal), um modelo que ajusta apenas a grande média e assim, efetivamente, irá consignar toda a variabilidade dos dados no componente aleatório ( $\hat{\mu} = g^{-1}(\beta_0)$ ). O modelo completo (também chamado saturado) ajusta toda a variabilidade de modo exato incorporando um parâmetro a cada observação ( $\hat{\mu} = \mathbf{y}$ ) e concentrando-a no componente sistemático. Ambos os modelos são de pouca aplicabilidade prática, sendo considerados como referenciais nos processos de escolha de modelos.

Um modelo linear é dito ordenado quando a estimação de seus parâmetros  $\beta$  é promovida na mesma sequência na qual eles são declarados no modelo e assim, o processo de modelagem deve consistir em se propor um modelo convenientemente situado entre o modelo nulo e o modelo completo pois, se por um lado para cada termo incorporado ao preditor linear desse modelo há um ganho na qualidade do ajuste aos dados, por outro lado há um incremento na complexidade do modelo e, por conseguinte, em sua interpretação (*tradeoff*). Assim, torna-se útil a determinação de uma quantidade associada a um Modelo Linear Generalizado que possa ser interpretada de modo similar à soma dos quadrados dos resíduos de um Modelo Clássico de Regressão Linear.

Com a estimação dos parâmetros do modelo proposto promovida pela maximização do logaritmo da função de verossimilhança pode-se, a cada estágio (termo incorporado) definir-se essa quantidade como uma medida da discrepância desse modelo em relação aos dados determinada a partir do modelo completo. A esta quantidade denominou-se por desvio em relação ao modelo saturado (*deviance*) (NELDER; WEDDERBURN, 1972, p. 375).

Considere  $\hat{l}_n$  e  $\hat{l}_p$  como os máximos dos logaritmos das funções de verossimilhança e  $\tilde{\theta}_i = q(y_i)$  e  $\hat{\theta}_i = q(\hat{\mu}_i)$  as estimativas de máxima verossimilhança dos parâmetros canônicos de dois modelos: o modelo completo ( $l_n$ ) e o modelo proposto ( $l_p$ ), respectivamente. Assim, o desvio escalonado para o modelo proposto:  $D_p^*$  (*scaled deviance*) é dado por  $D_p^* = 2(\hat{l}_n - \hat{l}_p)$  onde

$$\hat{l}_n = \frac{1}{\phi} \sum_{i=1}^n \left[ y_i \tilde{\theta}_i - b(\tilde{\theta}_i) \right] + \sum_{i=1}^n c(y_i, \phi); \text{ e } \hat{l}_p = \frac{1}{\phi} \sum_{i=1}^n \left[ y_i \hat{\theta}_i - b(\hat{\theta}_i) \right] + \sum_{i=1}^n c(y_i, \phi).$$

Assim

$$D_p^* = \frac{1}{\phi} \left\{ 2 \sum_{i=1}^n \left[ y_i (\tilde{\theta}_i - \hat{\theta}_i) + b(\hat{\theta}_i) - b(\tilde{\theta}_i) \right] \right\},$$

e o desvio do modelo proposto:  $D_p$  (*deviance*) será

$$D_p = \frac{D_p^*}{\phi} = 2 \sum_{i=1}^n \left[ y_i (\tilde{\theta}_i - \hat{\theta}_i) + b(\hat{\theta}_i) - b(\tilde{\theta}_i) \right], \quad (\text{B.26})$$



o qual assume, por exemplo, as seguintes expressões para as distribuições:

$$\begin{aligned} \text{Normal} &: \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \hat{\mu}_i)^2; \\ \text{Poisson} &: 2 \sum_{i=1}^n [y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) + (\hat{\mu}_i - m_i)]; \\ \text{Binomial} &: 2 \sum_{i=1}^n [y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) + (m_i - y_i) \log\left(\frac{m_i - y_i}{m_i - \hat{\mu}_i}\right)]. \end{aligned}$$

Associado a cada modelo há uma quantidade  $p$  definida como graus de liberdade dada pelo posto da matriz de delineamento  $\mathbf{X}$ , equivalente ao número de parâmetros linearmente independentes a serem estimados. Para uma amostra de  $n$  observações independentes, o desvio para o modelo proposto possui  $(n - p)$  graus de liberdade para os resíduos.

### B.2.7 Comparação de modelos encaixados (*nested*)

A inferência sobre Modelos Lineares Generalizados é baseada em certas propriedades dos estimadores de máxima verossimilhança e em testes de razão de verossimilhança cujos resultados se observam de modo assintótico em grandes amostras (MYERS et al., 2010, p. 132).

Seja  $M_{p_1}, \dots, M_{p_r}$  uma sequência de Modelos Lineares Generalizados com matrizes de delineamento  $X_{p_1} \subset \dots \subset X_{p_r}$  com número de parâmetros  $p_1 < \dots < p_r$  e desvios  $D_{p_1} < \dots < D_{p_r}$ . Sob esta definição estes modelos são ditos encaixados ou aninhados (*nested*) pela imposição de sucessivas restrições no modelo de maior ordem.

No contexto de Modelos Lineares Generalizados a hipótese nula é usualmente formulada considerando-se que o modelo correto é uma versão simplificada do modelo enunciado na hipótese alternativa e assim, o teste de hipóteses é proposto de modo a dar aos fatos uma chance de refutar a hipótese nula (FISHER, 1935, p. 19, tradução nossa)<sup>8</sup> assumindo a forma:

$$\begin{cases} H_0 : g(\mu) = \mathbf{X}_0 \boldsymbol{\beta}_0 \\ H_1 : g(\mu) = \mathbf{X}_1 \boldsymbol{\beta}_1, \end{cases}$$

em que  $\mu$  é o valor esperado do vetor de respostas  $\mathbf{Y}$ , cujos elementos são variáveis aleatórias independentes de uma mesma distribuição pertencente à família exponencial,  $\boldsymbol{\beta}_0$  e  $\boldsymbol{\beta}_1$  são os parâmetros desses modelos e  $X_0 \subset X_1$  suas matrizes de delineamento, estando assim o modelo  $M_0$  encaixado no modelo  $M_1$ .

<sup>8</sup>“Em relação a qualquer experimento nós devemos falar dessa hipótese como a hipótese nula e deve-se salientar que a hipótese nula nunca é provada ou estabelecida; mas é, possivelmente, refutada, no curso da experimentação. Todo experimento, deve ser afirmado, existe apenas para dar aos fatos uma chance de refutar a hipótese nula”

Pode-se definir uma estatística baseada na razão das verossimilhanças desses modelos:  $L(\beta_0)$  e  $L(\beta_1)$

$$\begin{aligned} W &= 2 \log \left( \frac{L(\beta_1)}{L(\beta_0)} \right) \\ &= 2 \left\{ l(\hat{\beta}_1) - l(\hat{\beta}_0) \right\} \end{aligned} \quad (\text{B.27})$$

Para grandes amostras, se  $H_0$  é verdadeira tem-se então que (WOOD, 2017, pp. 411-414)

$$W \sim \chi^2_{(p_1 - p_0)}, \quad (\text{B.28})$$

em que  $p_1$  é o número de parâmetros identificáveis  $\beta_i$  do modelo  $i$ . Se a hipótese nula é falsa, então o modelo  $M_1$  tenderá a ter uma verossimilhança substancialmente superior tal que o dobro da diferença para com o modelo  $M_0$  não será consistente com a distribuição  $\chi^2$  com  $(p_1 - p_0)$  graus de liberdade.

A estatística (B.27) é diretamente útil apenas se o logaritmo das funções de verossimilhança desses modelos puder ser calculado o que, no caso de Modelos Lineares Generalizados com parâmetros estimados pelo método dos mínimos quadrados ponderados iterativamente, verifica-se apenas quando o parâmetro de dispersão  $\phi$  é conhecido.

Pode-se redefinir a distribuição da estatística (B.28) em termos equivalentes à diferença dos desvios dos modelos  $M_0$  e  $M_1$  (ainda mantendo-se sua aplicabilidade apenas em inferências onde o parâmetro de dispersão seja conhecido) (WOOD, 2017, pp. 108-109)

$$D_0 - D_1 \sim \chi^2_{(p_1 - p_0)}, \quad (\text{B.29})$$

em que  $D_i$  é o desvio do modelo  $i$  com  $p_i$  parâmetros identificáveis.

Sob certas condições o resultado (B.29) possibilita a construção de um teste preliminar de um Modelo Linear Generalizado pela comparação de seu desvio com o valor crítico da distribuição  $\chi^2$  com  $(n - p)$  graus de liberdade sob um nível de significância arbitrado  $\alpha$  (JØRGENSEN, 1987, p. 135).

Assim, se

$$D_i \leq \chi^2_{(n - p_i; 1 - \alpha)}, \quad (\text{B.30})$$

haverá evidências de que o modelo proposto seja adequado todavia, recomendando-se ainda de modo suplementar, uma análise de resíduos.

Para modelos nos quais não se tenha conhecimento do parâmetro de dispersão  $\phi$ , tratar  $(D_0 - D_1)$  e  $D_i$  nas aproximações (B.29) e (B.30) como assintoticamente independentes implica em

$$F = \frac{(D_0 - D_1) / (p_1 - p_0)}{\hat{\phi}} \sim F_{(p_1 - p_0, n - p_1)}, \quad (\text{B.31})$$

onde  $\hat{\phi}$  é uma estimativa consistente para  $\phi$  que pode ser obtida pelo Método do desvio, Método de Pearson ou ainda pelo Método da Máxima Verossimilhança (WOOD, 2017, p. 110)

$$\begin{aligned} \text{Método do desvio: derivado de (B.26):} \quad \hat{\phi}_i &= \frac{\hat{D}_i}{(n - p_i)}; \\ \text{Método de Pearson: derivado da estatística de Pearson:} \quad \hat{\phi}_i &= \frac{\hat{X}_i^2}{(n - p_i)}, \end{aligned}$$

em que  $X_i^2$  é a estatística generalizada de Pearson cuja expressão é:

$$X_p^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \quad (\text{B.32})$$

Assim, se  $H_0$  é verdadeira a estatística (B.31) tem valor inferior ao valor crítico da distribuição F de Fisher-Snedecor com  $(p_1 - p_0; n - p_1)$  graus de liberdade sob um nível de significância arbitrado  $\alpha$ .

### B.2.8 Critério de informação de Akaike (AIC)

O critério para seleção de Modelos Lineares Generalizados encaixados baseado na verossimilhança geralmente não é satisfatório uma vez que um modelo com um maior número de parâmetros terá sempre uma maior verossimilhança posto que cada parâmetro adicional possibilita ao modelo uma melhor aproximação aos dados especificamente usados na modelagem.

O Critério de Informação de Akaike (AKAIKE, 1977) <sup>9</sup> é uma métrica alternativa para a decisão entre modelos (sejam eles encaixados ou não) no qual o “melhor modelo” se apresenta como aquele tendo o menor valor para a divergência de Kullback-Leibler entre os modelos proposto e “verdadeiro”. Sua fundamentação pode ser dividida em duas partes (WOOD, 2017, p. 414):

- a performance de um modelo estimado por máxima verossimilhança em prever novas observações pode ser aferida pelo valor esperado da divergência de Kullback-Leibler <sup>10</sup> entre esse modelo e o “modelo verdadeiro”. Isto pode ser mostrado pelo

<sup>9</sup>AIC Akaike Information Criteria)

<sup>10</sup>Há situações em que se necessita comparar duas distribuições de probabilidade. Mais especificamente, pode-se ter uma única variável aleatória e duas distribuições de probabilidade diferentes para ela: como uma distribuição verdadeira e uma aproximação. Em situações como essa, pode ser útil quantificar a diferença entre as distribuições. Geralmente, isso é referido como o problema de calcular a distância estatística entre dois objetos estatísticos, por exemplo, distribuições de probabilidade. Uma abordagem é calcular uma medida de distância entre as duas distribuições. Isso pode ser desafiador, pois pode ser difícil interpretar a medida. Alternativamente, é mais comum calcular uma divergência entre duas distribuições de probabilidade. Uma divergência é como uma medida, mas não é simétrica. Isso significa que uma divergência é uma pontuação de como uma distribuição difere de outra, em que o cálculo da divergência para as distribuições P e Q daria uma pontuação diferente de Q e P. As pontuações de divergência são uma base importante para muitos cálculos diferentes na teoria da informação e, mais geralmente, no aprendizado de máquina. Por exemplo, eles fornecem atalhos para calcular pontuações, como informações mútuas (ganho de informações) e entropia cruzada usada como uma função de perda para modelos de classificação. Dois

menor valor da divergência que o modelo proposto pode obter para qualquer valor dos parâmetros mais uma constante estimável;

- o valor esperado do negativo do logaritmo da função de verossimilhança do modelo proposto pode ser expresso como função do menor valor que se poderia obter da divergência de Kullback-Leibler a menos da mesma constante estimável.

Formalmente, o AIC é definido como

$$\text{AIC} = -2\hat{l} + 2p,$$

em que  $\hat{l}$  é o máximo do logaritmo da função de verossimilhança do modelo proposto e  $p$  é o número de parâmetros do modelo que devem ser estimados, uma quantidade que atua como uma penalização do valor do AIC pela complexidade de um modelo com mais variáveis. Por esse critério, o modelo como menor valor de AIC é escolhido.

Alternativamente, o Critério de Informação de Schwarz (BIC) (SCHWARZ, 1978) impõe uma penalização mais aguda a modelos mais complexos para amostras de maiores dimensões

$$\text{SBC} = -2\hat{l} + p \log(n),$$

em que  $n$  é o tamanho da amostra.

### B.2.9 Algoritmos para seleção de covariáveis

Para um conjunto de  $p$  variáveis há, potencialmente,  $2^p$  possíveis escolhas para serem incorporadas ao preditor linear e assim, algoritmos iterativos para seleção de variáveis baseados em alguma métrica (usualmente o Critério de informação de Akaike: AIC) para comparação entre os diversos modelos foram desenvolvidos e, basicamente, consistem em três variantes (WEISBERG, 2014, p. 239):

- método de seleção direta (*forward selection*): considera como subconjuntos todos os modelos candidatos que consistem em um termo adicional além do modelo inicial (usualmente o modelo nulo) e calcula a AIC para cada um desses modelos candidatos. Se o AIC para todos os modelos candidatos exceder o AIC para o modelo atual o processo é interrompido e o modelo atual é selecionado. Por outro lado, seleciona-se o modelo que minimiza o AIC como novo modelo, repetindo-se esta etapa para cada variável. Se o total de variáveis além do intercepto é  $p$ , esse algoritmo poderá considerar no máximo:  $p + (p - 1) + \dots + 1 = p + (p+1)/2$  do total  $2^p$  de possíveis subconjuntos. Assim, para  $p = 10$ , esse algoritmo analisa apenas 45 dos 1024 possíveis subconjuntos de variáveis.

---

escores de divergência comumente usados pela teoria da informação são os Critérios de Divergência de Kullback-Leibler e Jensen-Shannon. Uma divergência de Kullback-Leibler igual a zero indica que as distribuições P e Q são muito parecidas (ou até mesmo iguais), enquanto que uma divergência de um sugere que são distribuições distintas.

- método de eliminação reversa (*backward selection*): considere os modelos candidatos que diferem do modelo inicial pela exclusão de um termo. Se o AIC para todos os modelos candidatos excede a AIC do modelo inicial mantém-se o modelo inicial; caso contrário, define-se como modelo atual o modelo candidato que apresentar o menor AIC. O processo é interrompido caso nenhuma variável permanecer no modelo atual; caso contrário, repete-se esta etapa. Tal como acontece com o método de seleção direta, apenas  $p(p+1)/2$  do total de subconjuntos são considerados.
- método de seleção por partes (*stepwise selection*): os algoritmos anteriores podem ser combinados em um método por partes (*stepwise selection*) onde a cada etapa o subconjunto considerado é verificado pela adição ou supressão de uma variável. Os modelos candidatos consistem em todos os subconjuntos obtidos a partir do subconjunto inicial adicionando ou excluindo uma variável. Aceita-se como o novo modelo candidato o subconjunto com o menor AIC. Se o novo candidato foi o mesmo da última etapa o processo é interrompido; caso contrário, repete-se a etapa anterior.

#### B.2.10 Quantidades diagnósticas para validação de um modelo

Assim como para os modelos clássicos de regressão linear, a análise de resíduos segue como uma ferramenta fundamental na verificação da adequabilidade também para um Modelo Linear Generalizado, quando então poderão ser validadas:

- adequação da distribuição adotada para a resposta;
- adequação da parte sistemática;
- adequação da função de ligação;
- verificação de observações inconsistentes, influentes e com capacidade de alavancagem;

A principal razão pela qual não se recomenda a adoção dos resíduos brutos ( $\hat{\varepsilon}_i = y_i - \hat{\mu}_i$ ) para a verificação da adequabilidade de um Modelo Linear Generalizado reside na dificuldade para a validação da relação média e variância estabelecida pela distribuição adotada com esse tipo de resíduos. No modelo Poisson, à guisa de exemplo, a variância dos resíduos deve crescer de modo diretamente proporcional à magnitude dos valores ajustados ( $V(\hat{\mu}) = \hat{\mu}$ ) e necessita-se uma notável habilidade para se depreender essa relação em um gráfico dos resíduos brutos pelos valores estimados (WOOD, 2017, p. 112). A padronização dos resíduos de um Modelo Linear Generalizado facilita a verificação da adequabilidade do modelo ao se mostrarem com uma variância relativamente constante.

Denominam-se os componentes da estatística generalizada de Pearson (B.32) como resíduos de Pearson e assim são definidos

$$\hat{\varepsilon}_i^p = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}.$$

A denominação adotada decorre do fato de que para uma Distribuição Poisson a soma de seus quadrados resulta na estatística de Pearson (NELDER; WEDDERBURN, 1989, p. 37)

$$\sum_{i=1}^n (\hat{\varepsilon}_i^p)^2 = X^2.$$

Os resíduos de Pearson podem ser padronizados (“estudentizados”) pela simples divisão desses, por uma quantidade proporcional ao desvio padrão do modelo ajustado

$$\hat{\varepsilon}_i^{p'} = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\phi}V(\hat{\mu}_i)(1 - \hat{h}_{ii})}},$$

em que  $\hat{h}_{ii}$  é o  $i$ -ésimo elemento da diagonal da matriz de projeção  $\mathbf{H}$   $\mathbf{H} = \sqrt{\mathbf{W}}\mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\sqrt{\mathbf{W}}$ .

Outro tipo de resíduo com propriedades mais apropriadas à análise de um Modelo Linear Generalizado é o resíduo do desvio (*deviance residual*) obtidos a partir da raiz quadrada dos componentes do desvio

$$D_p = \sum_{i=1}^n d_i; \text{ com } \hat{\varepsilon}_i^d = \text{sign} [y_i - \hat{\mu}_i] \sqrt{d_i}.$$

Os resíduos componentes do desvio podem ser padronizados (estudentizados) pela expressão

$$\hat{\varepsilon}_i^{d'} = \frac{\hat{\varepsilon}_i^d}{\sqrt{\hat{\phi}(1 - \hat{h}_{ii})}}.$$

Infelizmente os resíduos componentes do desvio não são, em geral, bem definidos (por exemplo, ao modelar múltiplos parâmetros para a distribuição de resposta), enquanto os resíduos de Pearson podem estar longe de serem Normalmente distribuídos para variáveis resposta com distribuições altamente assimétricas ou com elevada curtose (STASINOPOULOS et al., 2017, p. 418).

Resíduos quantílicos, tal como propostos, apresentam uma distribuição exatamente Normal e são obtidos invertendo a distribuição ajustada para cada valor de resposta e determinando o quantil Normal padrão equivalente. Para dados discretos, a classe de resíduos proposta inclui alguma randomização (DUNN; SMITH, 1996, p. 236).

Seja  $F(y; \mu; \phi)$  a função distribuição cumulativa de probabilidade de  $\mathcal{P}(\mu; \phi)$ . Se  $F$  é contínua, então  $F(y_i; \mu_i; \phi)$  é uniformemente distribuída em  $[0,1]$ . Neste caso, os resíduos quantílicos são definidos como

$$\varepsilon_i^q = \Phi^{-1} \left\{ F(y_i; \hat{\mu}_i; \hat{\phi}) \right\} \sim \mathcal{N}(0, 1),$$

em que  $\Phi$  é a função distribuição cumulativa de probabilidade Normal padronizada. Independentemente da variabilidade amostral em  $\hat{\mu}_i$  e  $\hat{\phi}$ ,  $\varepsilon_i^q$  seguem exatamente uma distribuição

Normal padronizada. Isso implica que  $\hat{\varepsilon}_i^q$  converge para uma distribuição Normal padronizada se  $\beta$  e  $\phi$  foram estimados de modo consistente (DUNN; SMITH, 1996, p. 239).

A modificação proposta por Atkinson (1981, p. 15) na distância de Cook (B.16) tem como vantagens sua padronização para diferentes configurações da matriz de projeção  $\mathbf{X}$  e destaca pontos extremos com mais intensidade

$$T_i = \left\{ \left[ \frac{(n-p)}{p} \right] \left[ \frac{h_i}{(1-h_i)} \right] \right\}^{1/2} |r_i^*| r_i^*,$$

em que

$$r_i^* = \frac{r_i'}{\left\{ \left[ \frac{(n-p)}{(n-p-1)} - \frac{r_i'}{(n-p-1)} \right]^{1/2} \right\}}$$

e  $r_i'$  são os resíduos padronizados.

A adaptação dessa medida para seu uso em Modelos Lineares Generalizados fica expressa como (NELDER; WEDDERBURN, 1989, p. 407)

$$D_i' = \frac{(\beta_{(i)} - \hat{\beta})(\mathbf{X}^T \mathbf{W} \mathbf{X})(\beta_{(i)} - \hat{\beta})}{p \phi},$$

em que  $\beta_{(i)}$  é o vetor dos parâmetros estimados com a deleção da  $i$ -ésima observação.

Uma abordagem geral para a influência foi proposta por Cook e Weisberg (1982, p. 182), onde a influência da  $i$ -ésima observação poderia ser medida pela alteração que sua remoção implica nas estimativas de máxima verossimilhança dos parâmetros.

Seja  $L(\hat{\theta})$  a verossimilhança do modelo baseado nos dados completos. Define-se como distância da verossimilhança  $L(D_i)$  (*likelihood distance*) à quantidade

$$LD_i = 2 \left[ L(\hat{\theta}) - L(\hat{\theta}_{(i)}) \right],$$

em que  $L(\hat{\theta}_{(i)})$  é a verossimilhança do modelo com a exclusão da  $i$ -ésima observação. O afastamento da verossimilhança pode ser expresso em termos dos resíduos padronizados de Pearson ( $\hat{\varepsilon}_i^{p'}$ ): e dos elementos da diagonal matriz de projeção ( $\hat{h}_{ii}$ ), resultando na expressão:

$$LD_i = \frac{\hat{h}_{ii}}{(1 - \hat{h}_{ii})} (\hat{\varepsilon}_i^{p'})^2.$$

Valores de afastamento muito destoantes dos demais despertam atenção e requerem análise individual.

A avaliação da adequação da função de ligação pode ser realizada pela análise da variação no desvio entre o modelo proposto e um modelo alternativo ajustado com a inclusão, como variável explicativa, de  $\hat{\eta}^2$ , onde  $\hat{\eta}$  é o preditor linear do modelo proposto. Se a redução observada for expressiva há indícios de inadequação da função de ligação adotada no modelo proposto (CORDEIRO; DEMÉTRIO; MORAL, 2015, p. 164).

Alternativamente, pelo gráfico dos elementos do vetor de ajuste ( $\hat{z}_{ii}$ ) *versus* os elementos do preditor linear ( $\hat{z}_i$ ), no qual aguarda-se uma relação linear como indício confirmatório da adequação da função de ligação do modelo proposto.

### B.2.11 Validação de um Modelo Linear Generalizado

As técnicas para validação de um Modelo Linear Generalizado podem ser também, como para um Modelo Clássico de Regressão Linear, formais ou informais e, tal qual para aquela classe de modelos, as técnicas informais recaem na habilidade de se perceber padrões nos resíduos: se há algum padrão visível em gráficos de resíduos então há um outro modelo que será mais adequado (NELDER; WEDDERBURN, 1989, p. 392). Não obstante a flagrante subjetividade e dificuldade para conjunto de dados de pequeno tamanho, verificações informais ainda são importantes.

Para ser possível interpretar quão extrema alguma quantidade medida é, seja ela algum tipo de resíduo (*i.e.*, medida de consistência), influência (distância de Cook) ou alavancagem (*leverage*, gráficos dessas quantidades contra os valores esperados das estatísticas de ordem de uma amostra vinda de uma variável com distribuição Normal são usualmente empregados sob dois tipos: gráfico Seminormal (*half-Normal plot*) ou Normal (*full-Normal plot*). O primeiro tipo apropriado para quantidades não negativas como a medida de alavancagem ( $h_{ii}$ ) ou de influência (distância de Cook) e o último para quantidades como a consistência (resíduos).

Para cada gráfico a medida ordenada é justaposta aos valores esperados da estatística de ordem de uma amostra vinda de uma variável com distribuição Normal, que podem ser obtidos de modo aproximado por <sup>11</sup> (NELDER; WEDDERBURN, 1989, p. 407)

$$\begin{aligned} \text{Normal (full-Normal plot): } \quad \mathbb{E}(z_i) &= \Phi^{-1} \left( \frac{i - 3/8}{n + 1/4} \right) \quad i = 1, \dots, n; \\ \text{Seminormal (half-Normal plot): } \quad \mathbb{E}(z_i) &= \Phi^{-1} \left( \frac{n + i + 1/2}{2n + 9/8} \right) \quad i = 1, \dots, n. \end{aligned} \quad (\text{B.33})$$

À exceção dos resíduos quantílicos, os demais tipos de resíduos (*i.e.*, medidas de consistência) e quantidades relativas à influência ( $D_i$ : distância de Cook) e alavancagem (*leverage*:  $h_{ii}$ ) não seguem uma distribuição Normal de tal modo que o uso de envelopes simulados (NELDER; WEDDERBURN, 1989, p. 408) torna mais confiável a avaliação visual do padrão assumido em gráficos onde se aguarda um padrão Normal de distribuição.

A primeira abordagem para a exploração das flutuações amostrais usando simulações para criar regiões críticas para análises foi proposta por Ripley (1977, p. 181) e, pouco mais tarde, Atkinson (1981) apresentou uma técnica para se detectar observações

<sup>11</sup> A expressão acima para uma aproximação do valor esperado da estatística de ordem de uma amostra vinda de uma distribuição Normal completa foi proposta originalmente na tese de Blom (1958, p. 145).

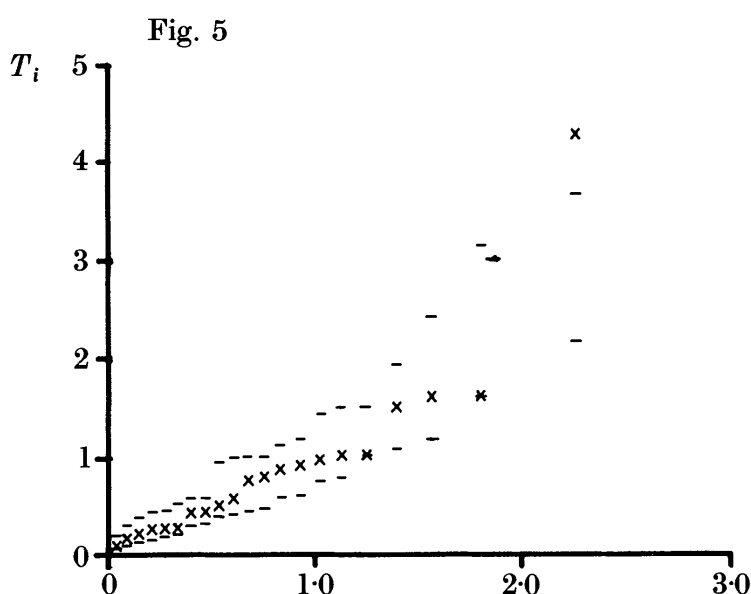


extremas que extrapolassem os limites de bandas de confiança (envelope) construídas por meio de simulações.

A sequência a seguir ilustra a construção de envelopes (cf. Figura B.6) para a distribuição Seminormal (*half-Normal*) (ATKINSON, 1981, p. 15) :

- ordenar os valores absolutos das quantidades diagnósticas calculadas (v.g: resíduos:  $\hat{\varepsilon}_i$ , distância de Cook:  $\hat{D}_i$ , leverage:  $\hat{h}_{ii}$ );
- com a matriz de variáveis explicativas  $\mathbf{X}$  dos dados e o modelo sob análise simular 19 amostras de valores aleatórios ( $y_i$ );
- para cada amostra estimar um modelo, obter e ordenar os valores absolutos das quantidades diagnósticas associadas a cada amostra;
- para cada elemento  $i$  de cada simulação calcular os percentis associados 0,05, 0,50 e 0,95;
- justapor em um gráfico os valores esperados para a estatística de ordem de uma variável com distribuição Normal (B.33), os valores absolutos ordenados das quantidades diagnósticas e seus percentis.

Figura B.6: Envelope construído por simulação



Fonte: Adaptação Atkinson (1981, p. 17)

Algumas análises informais baseadas em técnicas gráficas, nas quais espera-se que nenhum padrão visual sugestivo de tendência possa ser identificado:

- resíduos contra valores estimados;
- resíduos contra variáveis explanatória consideradas no modelo;
- resíduos contra variáveis explanatórias não incluídas no modelo;
- resíduos contra os índices  $i$  (ordem na qual foram obtidas as amostras).

Adicionalmente um gráfico Seminormal (*half-Normal plot*) dos resíduos quantílicos sob envelopes simulados e das quantidades diagnósticas:

- $\hat{h}_{ii}$ ;
- $\hat{D}'_i$ ;

### B.3 FUNÇÕES DE SUAVIZAÇÃO PARA MÉTODOS NÃO PARAMÉTRICOS

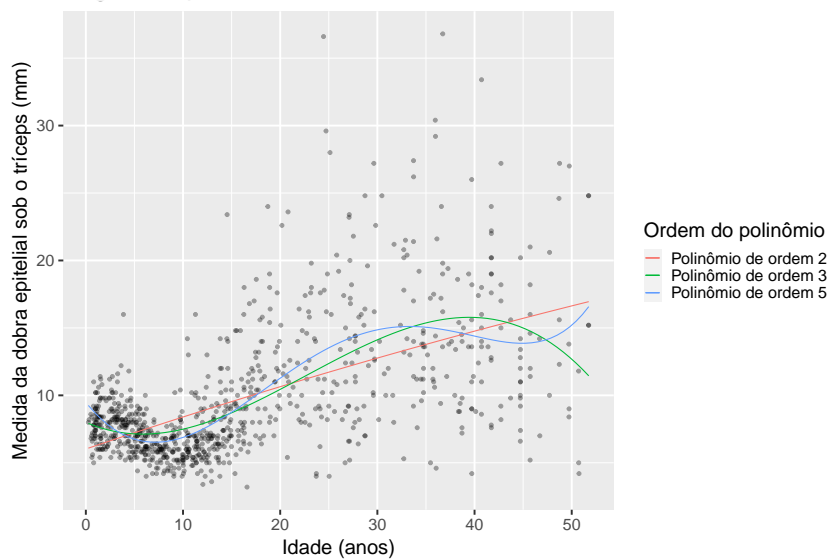
#### B.3.1 Regressão polinomial

Polinômios formam uma das classes mais úteis e conhecidas de funções de  $\mathcal{R}$  em  $\mathcal{R}$  posto possibilitarem aproximar, de modo uniforme, funções contínuas (BURDEN; FAIRES; BURDEN, 2017, p. 116).

A regressão polinomial ilustrada na Figura B.7 consiste no ajuste matemático de uma única função a um intervalo definido e fechado pela incorporação de transformações na variável preditora  $\mathbf{X}$  na forma de polinômio de grau  $p$  como variáveis auxiliares, tal como

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p.$$

Figura B.7: Regressão polinomial



Fonte: Autor

Os Polinômios de Taylor aproximam tão bem quanto possível uma função concentrando sua precisão em um ponto específico na sua vizinhança (Teorema de Taylor). Todavia, para conjuntos de dados com súbitas alterações na relação da resposta com a variável explicativa, observa-se que na interpolação polinomial a função passa a ter um comportamento bastante errático decorrente do atendimento de sua continuidade e de suas derivadas em todos os pontos e tal comportamento pode dificultar a compreensão da real tendência existente (WOOD, 2017, p. 163).

### B.3.2 Regressão polinomial local

Como alternativa a se impor uma única estrutura para se obter uma estimativa  $\hat{y}$ , o conceito de uma regressão local desenvolveu-se a partir do trabalho de Watson (1964) e se baseia em segmentar, de modo sucessivo, o intervalo de valores no processo do ajuste polinomial, estimando-se subestruturas baseadas apenas na fração dos dados situados imediatamente adjacentes a um ponto focal  $x_0$  e usadas localmente na estimação de  $\hat{y}$  neste ponto focal.

As subestruturas são determinadas por uma janela móvel de observação (*moving window*) a qual, excetuando-se junto às fronteiras amostrais da variável explanatória, possui um igual número de observações  $x_i$  à esquerda e à direita do ponto focal (simetria).

Em uma regressão polinomial local a amplitude da janela (*span*)  $span = (2k+1)/n$  influi diretamente no perfil final da suavização da tendência obtido: janelas mais estreitas ( $span \sim 0$ ) resultam em uma suavização da tendência da variável resposta com baixo viés todavia, com muitas flutuações (*wiggly*) e elevada variabilidade; de modo contrário, janelas muito amplas ( $span \sim 2$ ) apresentam um ajuste suave (*smooth*) com pouca oscilação e reduzida variabilidade mas com um elevado viés.

A amplitude da janela está diretamente associada ao conceito de penalização do efeito de suavização da função por algum parâmetro a ser estimado ou arbitrado de modo a se obter um equilíbrio entre a variabilidade e o enviesamento da função de suavização da regressão, como mostrado nas Figuras B.9 e B.8. O equilíbrio entre tendenciosidade e variância (*bias-variance trade-off*) é empregada por (HASTIE; TIBSHIRANI, 1986, p. 299) para sugerir que estas duas características da função suavizadora não podem ser otimizadas simultaneamente. Usualmente denotado por  $\lambda$  sua estimação é um dos mais importantes aspectos relacionados às técnicas de suavização (STASINOPOULOS et al., 2017, p. 262).

Em uma regressão polinomial local três aspectos precisam ser antecipadamente definidos: a escolha da amplitude da janela (parâmetro de suavização); o grau do polinômio e se as observações serão ponderadas ou não na regressão polinomial. A sequência abaixo exemplifica as etapas básicas de um processo de regressão local:

- ordenar os pares de valores  $(y_i, x_i)$  para  $i = 1, \dots, n$  com relação a  $x_i$ ;
- especificar a amplitude da janela de observação;
- focar em uma única observação  $x_0$ , ajustar a função polinomial (com ou sem pesos) para as observações dentro da janela e estimar  $\hat{y}_0$ ;
- repetir para as demais observações.

Para o ajuste de uma regressão polinomial com pesos locais (LOWESS: *Robust Locally Weighted Regression and Smoothing Scatterplots*) (CLEVELAND, 1979)

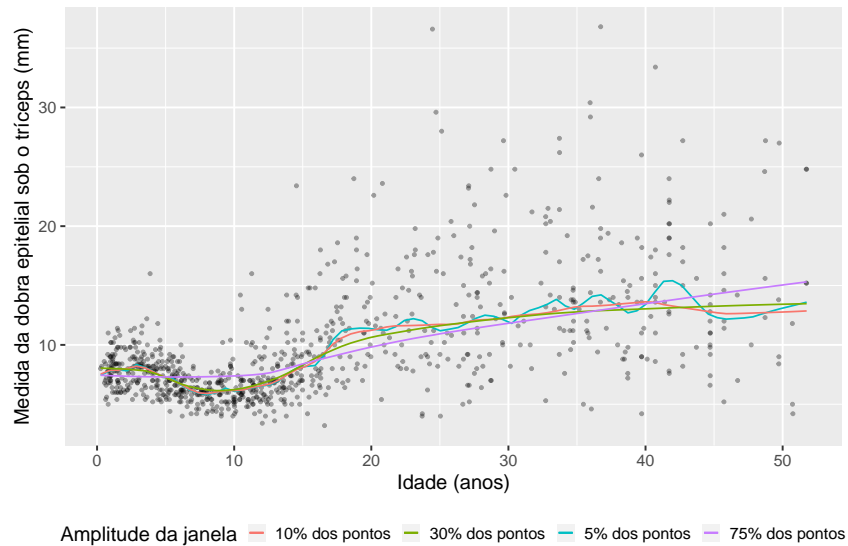
usualmente utiliza-se como função de pesos uma função positiva e simétrica tal como (REGUERA, 2021, p. 15):

$$W(z) = \begin{cases} (1 - |z|^3)^3 & \text{se } |z| < 1 \\ 0 & \text{se } |z| \geq 1, \end{cases}$$

em que  $z_i = \frac{(x_i - x_0)}{h}$  e  $h$  é a metade da amplitude da janela. O ajuste do polinômio aos dados internos à janela é promovido usando-se o método dos mínimos quadrados ponderados e algumas rotinas computacionais usam polinômios de graus diferentes para as regressões: grau 2 sob a denominação *LOESS* e grau 1 sob a denominação *LOWESS*.

A amplitude da janela pode ser considerada constante ou variável em todo o intervalo de valores (REGUERA, 2021, p. 17): quando a abertura é constante apenas as observações dentro dessa faixa de valores é usada. Esse é um método apropriado se a distribuição de “x” for aproximadamente uniforme ou o tamanho da amostra for grande (nas situações contrárias poderão ocorrer casos em que se tenham poucos ou nenhum vizinho ao ponto focal) ou escolher os vizinhos mais próximos do ponto focal por uma certa proporção do total dos valores de “x”.

Figura B.8: Diferentes regressões locais (*LOWESS*: polinômio de grau 1) em função da amplitude da janela

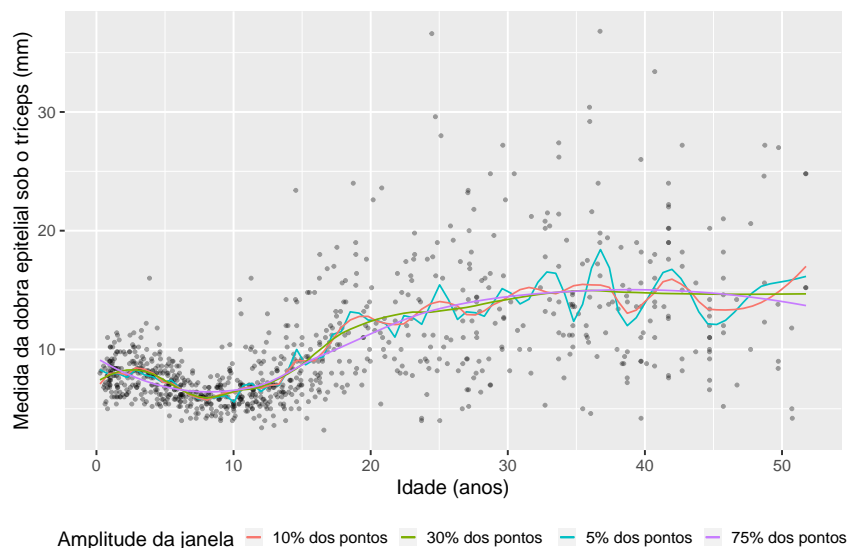


Fonte: Autor

### B.3.3 Polinômios por partes

Polinômios por partes são estimados usando apenas a informação da variável contida em intervalos delimitados por pontos arbitrariamente escolhidos (nós) e, sem a

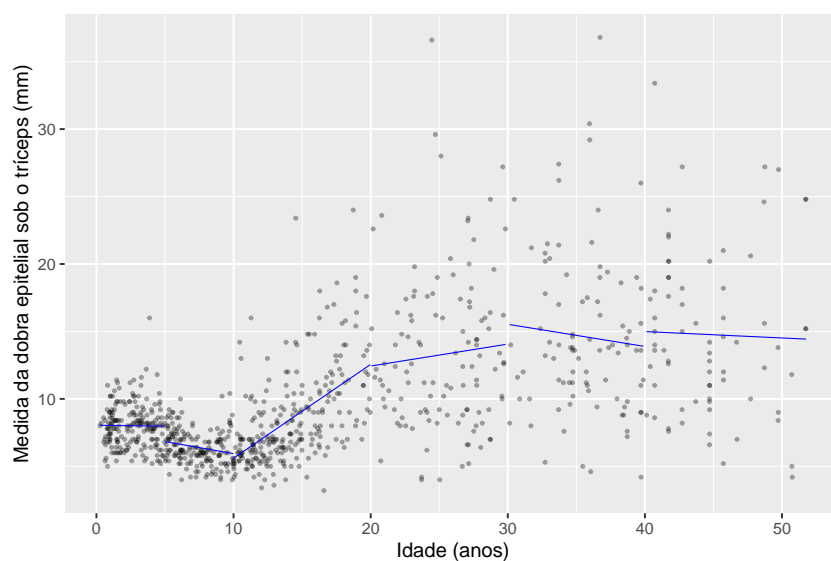
Figura B.9: Diferentes regressões locais (*LOESS*: polinômio de grau 2) em função da amplitude da janela



Fonte: Autor

imposição de qualquer tipo de restrição relacionada à continuidade nos nós, como mostrado na Figura B.10.

Figura B.10: Regressão polinomial linear por partes sem nenhuma restrição relacionada à continuidade nos nós (arbitrados em: 5,10,20,30,40)



Fonte: Autor

Uma maneira de se contornar esse efeito é pelo uso de polinômios definidos em partes (*piecewise polynomials*); isto é, funções com suporte limitado a intervalos delimitados pela localização dos pontos de descontinuidade denominados nós, assumindo valor zero no restante do domínio.

Considere os nós assim localizados  $\{x_j^* : 1, \dots, k\}$  e admita que  $x_j^* > x_{j-1}^*$ . Assim, para  $j = 1, \dots, (k-1)$  as seguintes funções lineares permitem expressar de modo isolado pelos nós a relação (WOOD, 2017, p. 164):

$$b_j(x) = \begin{cases} (x-x_{(j-1)}^*)/(x_j^*-x_{(j-1)}^*) & x_{(j-1)}^* < x \leq x_j^* \\ (x_{(j+1)}^*-x)/(x_{(j+1)}^*-x_j^*) & x_j^* < x \leq x_{(j+1)}^* \\ 0 & \text{caso contrário,} \end{cases}$$

com

$$b_1(x) = \begin{cases} (x_2^*-x)/(x_2^*-x_1^*) & x < x_2^* \\ 0 & \text{caso contrário,} \end{cases}$$

e

$$b_k(x) = \begin{cases} (x-x_{(k-1)}^*)/(x_k^*-x_{(k-1)}^*) & x > x_{(k+1)}^* \\ 0 & \text{caso contrário,} \end{cases}$$

Assim,  $b_j$  é zero em qualquer valor que não esteja no intervalo imediatamente anterior e posterior a  $x^*$ . Nesse intervalo,  $b_j$  cresce linearmente começando em zero no nó  $x_{(k-1)}^*$  até o valor de 1 no nó  $x_k^*$ , após o qual ela decresce até 0 no nó  $x_{(k+1)}^*$ .

Com a imposição de uma restrição relacionada à continuidade nos nós, as funções polinomiais por partes são habitualmente denominadas de *splines* e o aspecto final assumido pode ser observado na Figura B.11.

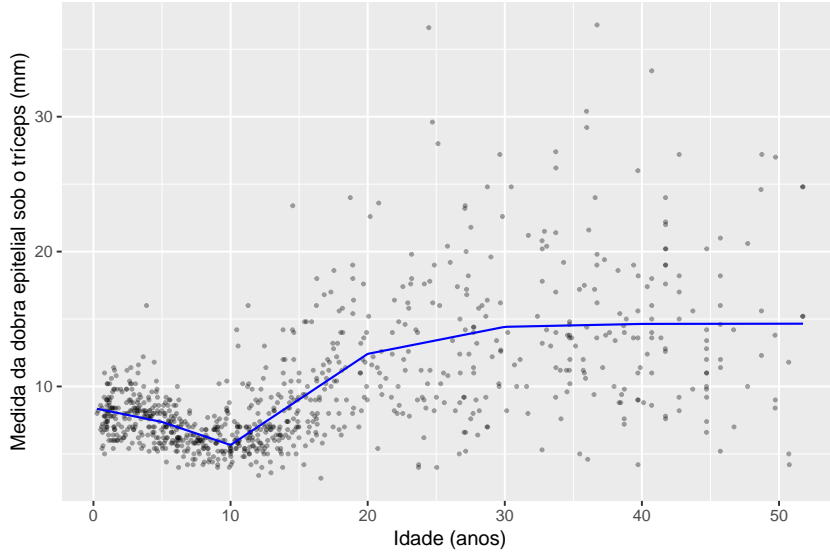
### B.3.4 Espaço de funções

Considere  $n$  observações de uma variável resposta  $\mathbf{Y} = (y_1, \dots, y_n)^T$  e uma única variável explanatória  $\mathbf{X} = (x_1, \dots, x_n)^T$ . Um suavizador de dispersão univariado pode ser expresso formalmente como um modelo estatístico (WOOD, 2017):

$$f(x) = \mathbb{E}(Y|x) \quad \text{ou, de modo equivalente como} \\ y_i = f(x_i) + \varepsilon_i \quad i = 1, \dots, n. \quad (\text{B.34})$$

De acordo com o Teorema de Weierstrass há sempre um polinômio arbitrariamente próximo a qualquer função contínua, todavia o grau desse polinômio pode ser muito grande. Uma estratégia para a estimação é usar vários polinômios de menor ordem cada um, estimando apenas um pequeno segmento da função a ser aproximada (PRICE, 2018, p. 9).

Figura B.11: Regressão polinomial linear por partes com restrição de continuidade nos nós (arbitrados em: 5,10,20,30,40)



Fonte: Autor

Admitindo-se que a função de suavização a ser estimada pertença a uma coleção infinita de funções, sua escolha é equivalente à escolha de uma base de funções que apresentem propriedades úteis e gerem um espaço vetorial de funções. Assim, expressar (B.34) na forma de uma combinação linear torna possível tanto sua estimação, quanto a realização de inferências pelos métodos tradicionais abordados anteriormente. A definição desse espaço vetorial é de particular interesse.

Diz-se pertencer ao *Espaço de Banach* uma função numérica  $f$  definida em um subconjunto aberto  $\Omega \in \mathcal{R}^n$  representado por  $L^p(\Omega)$ , onde  $1 \leq p < \infty$  com norma integrável em  $\Omega$  tal que

$$\|f\|_{L^p(\Omega)} = \left( \int_{\Omega} |f(x)|^p dx \right)^{1/p}.$$

Na situação particular em que  $\Omega = [a, b]$  e  $p = 2$  tem-se que

$$\|f\|_{L^2([a,b])} = \left( \int_a^b |f(x)|^2 dx \right)^{1/2},$$

é denominado como *Espaço de Hilbert*.

Para que uma função de suavização  $f(\cdot) \in L^2([a, b])$  seja suave, torna-se necessário que essa possua  $m$  derivadas contínuas. As funções com essa característica pertencem ao espaço denominado por *Espaço de Sobolev*, isto é

$$W_2^m([a, b]) = \{f : f^m \text{ é absolutamente contínua, e } f^m \in L^2([a, b])\}$$



Um modelo linear não paramétrico de ordem  $k$  de uma função não conhecida  $f(\cdot)$  consiste em um espaço linear  $k$ -dimensional  $\mathcal{F}$  que tem por funções de base  $B_j$ ,  $j = 1, \dots, k$ , em que cada elemento  $f \in \mathcal{F}$  pode ser escrito como uma combinação linear única (WOOD, 2017, p. 162), isto é

$$f(x) = \beta_1 B_1(x) + \dots + \beta_k B_k(x) = \sum_{j=1}^k \beta_j B_j(x). \quad (\text{B.35})$$

A escolha das funções da base deve ser coerente com o que se sabe sobre a natureza dos dados e, neste sentido, pode-se propor dentre outras, funções radiais; funções unitárias retificadas (RELU: *rectified linear units*); funções tangentes hiperbólicas; soma de séries de senos e cossenos de Fourier.

Admitindo-se que a função suavizadora  $f(\cdot)$  seja um polinômio de grau 2, uma base para o espaço vetorial dos polinômios de grau  $\leq 2$  pode ser dada pelas seguintes funções  $\hat{\beta}_1(x) = 1$ ;  $\hat{\beta}_2(x) = x$  e  $\hat{\beta}_3(x) = x^2$  e a equação (B.35) pode ser expressa como

$$\hat{f}(x) = \hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 x^2$$

e assim o modelo (B.34) passa a ser denotado como

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + \hat{\beta}_3 x_i^2 + \hat{\varepsilon}_i.$$

As Figuras B.12 e B.13 mostram o suavizador procurado a partir da determinação de  $f(\cdot)$  proveniente da estimação de  $\beta$ .

Observa-se que as funções de base ilustradas nas Figuras B.12 e B.13 estão definidas sobre todo o domínio da variável estudada. Uma maneira de se obter uma base de funções para o espaço vetorial (ao qual acredita-se que a função de suavização pertença) definidas apenas em partes do domínio da variável é pelo uso de *B-splines*.

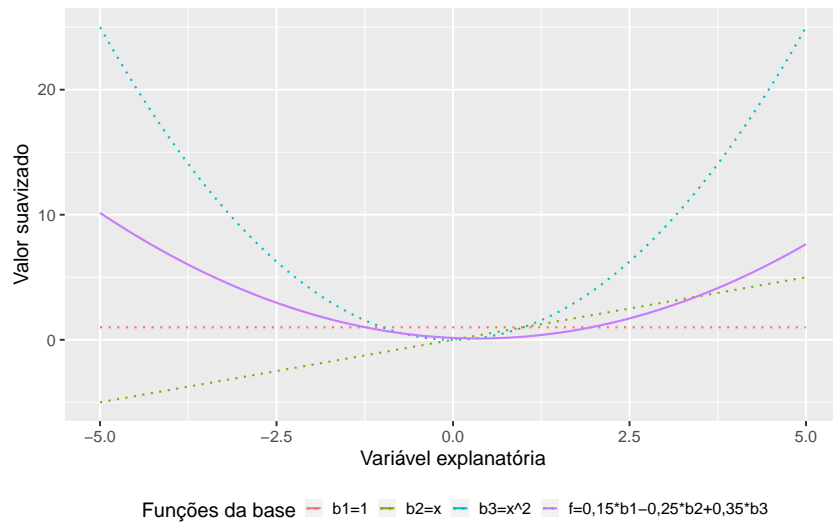
### B.3.5 B-Splines

Para a definição de uma base *B-spline* de  $k$  parâmetros define-se inicialmente o vetor com  $(k + m + 2)$  nós  $x_1 < x_2 < \dots < x_{k+m+2}$  sobre o intervalo onde o *spline* será avaliado, isto é,  $[x_{m+2}, x_{k+1}]$  (a localização do primeiro nó e do nó  $(m + 1)$  são essencialmente arbitrárias).

Assim, um *spline* de ordem  $(m + 1)$  pode ser representado na forma de uma combinação linear de bases *B-splines*  $B_i^m$  com  $\beta_i$  parâmetros

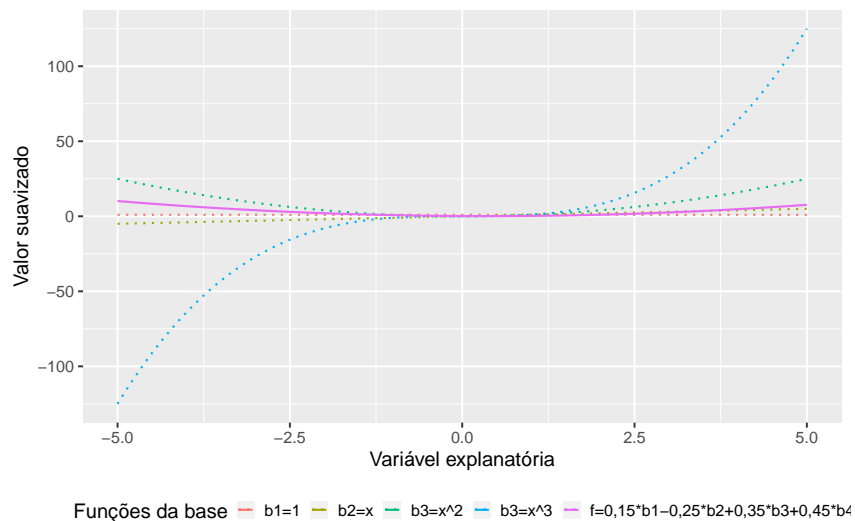
$$f(x) = \sum_{i=1}^k B_i^m(x) \beta_i,$$

Figura B.12: Função suavizadora resultante da combinação linear das funções de uma base de grau 2, definidas em todo o intervalo ( $f(\cdot) = 0,15b_1 - 0,25b_2 + 0,35b_3$ )



Fonte: Autor

Figura B.13: Função suavizadora resultante da combinação linear das funções de base de grau 3, definidas em todo o intervalo ( $f(\cdot) = 0,15b_1 - 0,25b_2 + 0,35b_3 + 0,45b_4$ )



Fonte: Autor

em que  $(m + 1)$  é a ordem dos polinômios da base de B-splines;  $k$  é o número de parâmetros a serem estimados (número de funções base) e  $B_i^m$  é a  $i$ -ésima função B-spline da base definida de modo recursivo na forma (WOOD, 2017)

$$B_i^{(m)}(x) = \frac{x - x_i}{x_{(i+m+1)} - x_i} B_i^{(m-1)}(x) + \frac{(i+m+2) - x}{x_{(i+m+2)} - x_{(i+1)}} B_{(i+1)}^{(m-1)}(x), \quad i = 1, 2, \dots, k \quad (\text{B.36})$$

e

$$B_i^{(-1)}(x) = \begin{cases} 1 & \text{se } x_i \leq x \leq x_{i+1}, \\ 0 & \text{caso contrário.} \end{cases} \quad (\text{B.37})$$

As equações B.36 e B.37 apresentam as seguintes propriedades:

- positividade:  $B_i^m(x) > 0$ ;
- suporte local:  $B_i^m(x) = 0$  fora do intervalo limitado a  $(m + 3)$  nós adjacentes;
- partição da unidade:  $\sum_{i=1}^k B_i^m(x) = 1$ ;
- continuidade:  $C^{(m+1)-2}$  em cada nó;
- recursividade: as expressões (B.36) e (B.37) são conhecidas como fórmulas recursivas de Cox-De Boor.

A relação recursiva das funções da base em termos de seu grau e suporte podem expressas na forma mostrada na Figura B.14 e seu cálculo implementado pelo algoritmo de Cox-De Boor (considerado uma generalização do algoritmo de Casteljau)<sup>12</sup>.

À guisa de exemplo, considere o vetor de nós  $x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4$  os intervalos delimitados pelos nós serão  $[1, 2), [2, 3), [3, 4)$ . Desse modo, as funções da base com grau zero serão:  $B_1^0(x) = 1$  para  $x \in [1, 2)$  (zero no restante),  $B_2^0(x) = 1$  para  $x \in [2, 3)$  (zero no restante) e  $B_3^0(x) = 1$  para  $x \in [3, 4)$  (zero no restante). As funções de uma base com grau superior terão como suporte o intervalo definido pelos intervalos das funções de ordem imediatamente inferior e assim sucessivamente. (cf. Figura B.15).

### B.3.6 Splines penalizados

Admita-se o ajuste de  $n$  pontos  $(x_j, y_j)$  por uma função suavizadora que tem  $k$  B-splines como base, expressa na forma

$$f(x) = \sum_{i=1}^k B_i^m(x_j) \beta_i, \quad x_j \in [x_{m+2}, x_{k+1}],$$

ou na forma vetorial

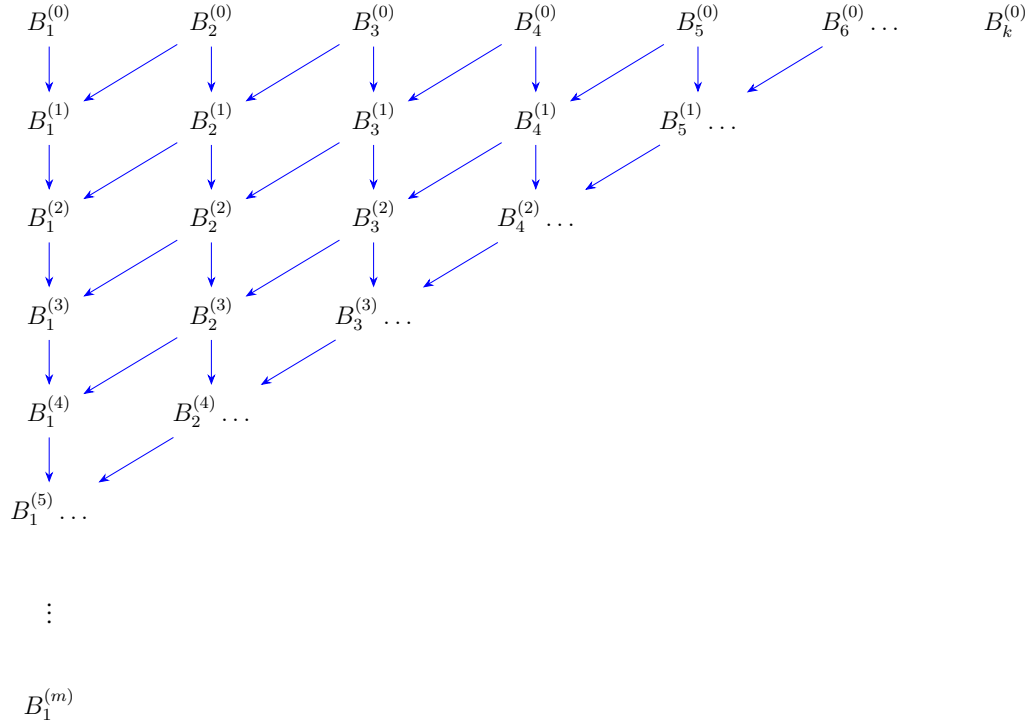
$$\mathbf{y} = \mathbf{B} \boldsymbol{\beta},$$

---

<sup>12</sup>Desenvolvido por Maurice G. Cox e Carl Willhem Reinhold De Boor

Figura B.14: Relação de recorrência de cada função da base

(algoritmo de Cox-De Boor)



Fonte: Autor

em que os elementos de  $\mathbf{B}$  são  $B_i^m(x_j)$ . A estimação do vetor  $\beta$  passa pela minimização da função:

$$S = \sum_{j=1}^n \left[ y_j - \sum_{i=1}^k \beta_i B_i^m(x_j) \right]^2$$

$$S = \|\mathbf{y} - \beta \mathbf{B}\|^2, \quad (\text{B.38})$$

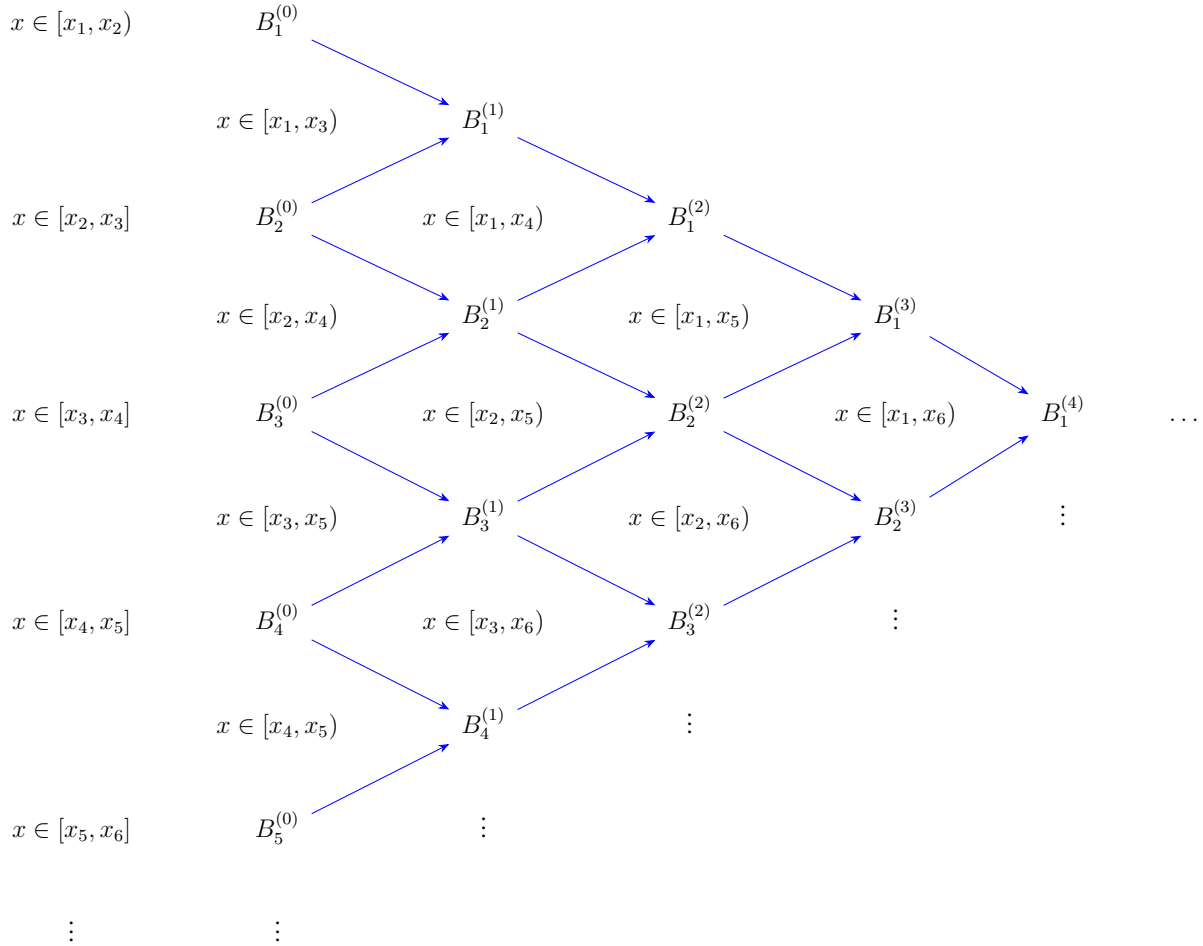
em que  $\mathbf{B} = [b_{ij}]$  é denominada como matriz das bases e a Equação (B.38) reside essencialmente na solução de um problema de regressão linear e a estimativa de mínimos quadrados para  $\beta$  é dada por

$$\hat{\beta} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{y}.$$

O'Sullivan (1986, p. 504) verificou que se  $f(\cdot)$  é estimada como uma combinação linear de B-splines a medida usual de aspereza, o quadrado integrado da segunda derivada poderia ser expresso como uma função quadrática dos coeficientes da base e o critério a ser minimizado pode ser rescrito a partir de (B.38) derivando-se uma matriz de banda  $\mathbf{P}$  (matriz de penalização) tal que  $\mathbf{R} = \beta^T \mathbf{P} \beta$  (EILERS; MARX; REGUERA, 2015, p. 153)

Figura B.15: Funções da base: relação de grau e suporte associado a cada função

(expandida até  $m=4$ )



Fonte: Autor

$$\begin{aligned}
 Q &= S + \lambda R \\
 &= S + \lambda \beta^T P \beta \\
 &= ||\mathbf{y} - \beta||^2 + \lambda \beta^T P \beta,
 \end{aligned} \tag{B.39}$$

em que  $\lambda$  é o parâmetro de suavização;  $\mathbf{y}$  é o vetor das respostas  $y_i$  e  $\beta$  é o vetor dos parâmetros a serem estimados para que minimize  $Q$ . Assim a solução de (B.39) será

$$\hat{\beta} = (B^T B + \lambda P)^{-1} B^T \mathbf{y},$$

e o suavizador procurado será dado por

$$\hat{f}(x) = \sum_{j=1}^n B_j(x) \hat{\beta}_j.$$

O cálculo de  $\mathbf{P}$  não é trivial e um algoritmo é apresentado por Wood (2016, p. 986). Assim considere exprimir a expansão das funções de base na forma que se segue e que  $m_1 = 3$  denote um *spline* cúbico

$$\mathbf{P}_{ij} = \int_a^b B_{m_1,i}^{[m_2]}(x) B_{m_1,j}^{[m_2]}(x) dx.$$

Todavia, por definição,  $B_{m_1,i}^{[m_2]}$  é construído por segmentos de polinômios de ordem  $p = m_1 - m_2$  entre nós  $x_l$  adjacentes do *spline*. Segue-se então que:

$$\begin{aligned} \mathbf{P}_{ijl} &= \int_{x_l}^{x_{l+1}} B_{m_1,i}^{[m_2]}(x) B_{m_1,j}^{[m_2]}(x) dx \\ &= \frac{h_l}{2} \int_{-1}^1 \sum_{i=0}^t a_i x^i \sum_{j=0}^t d_j x^j dx, \end{aligned}$$

para alguns coeficientes  $a_i$  e  $d_j$  e onde  $h_l = x_{l+1} - x_l$ . Estes coeficientes são a solução obtida ao se avaliar  $B_{m_1,i}^{[m_2]}(x)$  em  $t + 1$  pontos igualmente espaçados de  $x_l$  a  $x_{l+1}$  para se obter o vetor das derivadas avaliadas  $\mathbf{g}_a$  e assim resolver  $\mathbf{T}\mathbf{a} = \mathbf{g}_a$  onde  $T_{ij} = \{[-1+2(i-1)]/p\}^2$  (e de modo semelhante para  $\mathbf{T}\mathbf{d} = \mathbf{g}_d$ ). Então  $P_{ij} = \sum_l S_{ijl}$ .

Dado que

$$\int_{-1}^1 x^q dx = \{[1+(-1)^q]/(q+1)\}^q,$$

segue-se que

$$P_{ijl} = h_l \mathbf{g}_a^T \mathbf{T}^{-T} \mathbf{H} \mathbf{T}^{-1} \left( \frac{\mathbf{g}_d}{2} \right)$$

Denote por  $\mathbf{G}$  a matriz que mapeia  $\beta$  aos gradientes dos vetores para todos os intervalos e  $\mathbf{W}$  uma matriz diagonal de blocos dados por  $(h_l \mathbf{T}^{-T} \mathbf{H} \mathbf{T}^{-1})/2$  ter-se-á que

$$P_{ij} = \mathbf{G}_i^T \mathbf{W} \mathbf{G}_j,$$

em que  $\mathbf{G}_i$  é a  $i$ -ésima linha de  $\mathbf{G}$ . Seguem-se em (WOOD, 2016, p. 986) os passos das iterações do algoritmo com os elementos anteriormente definidos.

### B.3.7 P-Splines: definição da penalização

Considere um *spline* estimado por  $n$  B-splines aplicado como função suavizadora a  $m$  pontos. Com a penalização proposta a função a ser minimizada passa a ser expressa na forma (EILERS; MARX, 1996, p. 91):

$$SQR(f, \lambda) = \sum_{i=1}^m \left[ y_i - \sum_{j=1}^n \beta_j B_j(x_i) \right]^2 + \lambda \sum_{j=d+1}^n (\Delta^d \beta_j)^2,$$

em que as diferenças de ordem  $d$  entre os coeficientes de dois B-splines adjacentes,  $\Delta_{\beta_j}^d$  são dadas por

$$\begin{aligned}\Delta_{\beta_j} &= \beta_j - \beta_{j-1}; \\ \Delta_{\beta_j}^2 &= \beta_j - 2\beta_{j-1} + \beta_{j-2}; \\ &\vdots \\ \Delta_{\beta_j}^d &= \Delta \left( \Delta_{\beta_j}^{d-1} \right).\end{aligned}\tag{B.40}$$

Para  $d = 0$  recai-se em uma penalização de cumeieira e (B.39) passa a  $\mathbf{Q} = \|\mathbf{y} - \boldsymbol{\beta}\|^2 + \lambda \boldsymbol{\beta}^T \boldsymbol{\beta}$ .

Em geral, utiliza-se uma penalização de ordem  $d=2$  e (B.40) passa a (REGUERA, 2021, p. 29)

$$(\beta_1 - 2\beta_2 + \beta_3)^2 + \cdots + (\beta_{k-2} - 2\beta_{k-1} + \beta_k)^2 = \boldsymbol{\beta}^T \mathbf{D}^{2T} \mathbf{D}^2 \boldsymbol{\beta}$$

em que

$$\mathbf{D}^2 = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \dots \\ 0 & 1 & -2 & 1 & 0 & \dots \\ 0 & 0 & 1 & -2 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}_{(k-2,k)}.$$

Substituindo-se a penalização por  $\lambda \|\mathbf{D}^d \boldsymbol{\beta}\|^2 = \lambda \boldsymbol{\beta}^T \mathbf{D}^{dT} \mathbf{D}^d \boldsymbol{\beta} = \lambda \boldsymbol{\beta}^T \mathbf{P} \boldsymbol{\beta}$  obtém uma construção semelhante a (B.39).

### B.3.7.1 P-Splines: estimação do parâmetro de suavização, matrizes de penalização e suavização e graus de liberdade efetivos

A validação cruzada tem um alto custo computacional. Todavia, valendo-se da relação  $\hat{\mathbf{y}} = \mathbf{H} \mathbf{y}$  pode-se obter  $y_i - \hat{y}_{i-1} = (y_i - \hat{y}_i)/(1 - h_{ii})$  (EILERS; MARX; REGUERA, 2015, p. 158) com  $\mathbf{H} = \mathbf{B}(\mathbf{B}^T \mathbf{B} + \lambda \mathbf{P})^{-1} \mathbf{B}^T$ .

Um *spline* suavizador com parâmetro de suavização  $\lambda$  pré-escolhido é um exemplo de suavizador linear considerando que os parâmetros estimados em (B.39) são uma combinação linear de  $y_i$ .

Considere  $\hat{\mathbf{f}}$  o vetor dos valores estimados  $\hat{f}(x_i)$

$$\begin{aligned}\hat{\mathbf{f}} &= \mathbf{B}(\mathbf{B}^T \mathbf{B} + \lambda \mathbf{P})^{-1} \mathbf{B}^T \mathbf{y} \\ &= \mathbf{S}_\lambda \mathbf{y},\end{aligned}$$

o *spline* suavizador ainda segue linear em  $\mathbf{y}$  e  $\mathbf{S}_\lambda$  é conhecido como matriz de projeção<sup>13</sup> do modelo. Observa-se que a estimação de  $\hat{\mathbf{f}}$  não depende de  $\mathbf{y}$ ;  $\mathbf{S}_\lambda$  depende apenas de  $x_i$  e  $\lambda$

<sup>13</sup>Matriz de suavização ou de influência.

e assim, portanto, ele pode ser incorporado à estrutura de um Modelo Linear Generalizado compondo uma nova classe de modelos denominada Modelos Generalizados Aditivos <sup>14</sup> que herda muitos dos princípios daquela.

Para um *spline* suavizador a validação cruzada por ser expressa como

$$CV_\lambda = \frac{1}{n} \sum_{i=1}^n \left[ \frac{(y_i - \hat{f}_\lambda(x_i))}{1 - S_{ii}(\lambda)} \right]^2.$$

Na prática é costume substituir  $S_{ii}$  por seu valor médio  $[tr(\mathbf{S})/n]$ , resultando no processo denominado de validação cruzada generalizada (WOOD, 2017, p. 171)

$$CV_\lambda = \frac{n \sum_{i=1}^n [y_i - \hat{f}_\lambda(x_i)]^2}{[n - tr(\mathbf{S})]^2}$$

Considere  $\mathbf{B}_\xi$  como uma matriz de dimensões (nxm) ( $m \ll n$ ) de uma base composta por  $m$  *splines* cúbicos avaliados em  $n$  pontos  $x_j$  com uma sequência de nós  $\xi$ . Assim, o vetor dos valores estimados será dado por (HASTIE; TIBSHIRANI; FRIEDMAN, 2009, p. 153)

$$\begin{aligned} \hat{\mathbf{f}}(x) &= \mathbf{B}_\xi (\mathbf{B}_\xi^T \mathbf{B}_\xi)^{-1} \mathbf{B}_\xi^T \mathbf{y}; \\ &= \mathbf{H}_\xi \mathbf{y}, \end{aligned}$$

em que  $\mathbf{H}_\xi$  é o operador linear conhecido como matriz de projeção (*hat matrix*) dos sistemas de regressão linear.

Hastie, Tibshirani e Friedman (2009) assinalam as semelhanças e diferenças entre  $\mathbf{H}_\xi$  e  $\mathbf{S}_\lambda$ , dentre as quais:

- ambas são matrizes simétricas e positivo semidefinidas;
- $\mathbf{H}_\xi$  é idempotente ( $\mathbf{H}_\xi \mathbf{H}_\xi = \mathbf{H}_\xi$ ); todavia,  $\mathbf{S}_\lambda \mathbf{S}_\lambda \preceq \mathbf{S}_\lambda$ , significando que o lado direito excede o lado esquerda por uma matriz definida semipositiva;
- $\mathbf{H}_\xi$  tem posto  $m$ ;  $\mathbf{S}_\lambda$  tem posto  $n$ ;
- $tr(\mathbf{H}_\xi) = m$ , onde  $m$  é a dimensão do espaço de projeção: o número de funções da base e assim o número de parâmetros envolvidos na estimação de  $f(\cdot)$ .

Define-se <sup>15</sup> como graus de liberdade efetivos de um *spline* suavizador como a soma dos autovalores de  $\mathbf{S}_\lambda$

<sup>14</sup>GAM: *Generalised Additive Models*

<sup>15</sup>Esta definição não é a única. Hastie e Tibshirani (1990, p. 53) mencionam ainda  $df_\lambda = n - tr(2\mathbf{S}_\lambda - \mathbf{S}_\lambda \mathbf{S}_\lambda^T)$  e  $tr(\mathbf{S}_\lambda \mathbf{S}_\lambda^T)$ .



$$df_\lambda = \text{tr}(\mathbf{S}_\lambda)$$

e esta é uma das quantidades utilizadas para calibrar a quantidade de suavização realizada pela função suavizadora. A matriz  $\mathbf{S}_\lambda$  pode ser expressa na forma

$$\mathbf{S}_\lambda = (\mathbf{I} + \lambda \mathbf{K})^{-1}$$

em que  $\mathbf{K}$  não depende de  $\lambda$ .

A solução para  $\hat{\mathbf{f}} = \mathbf{S}_\lambda \mathbf{y}$  será a minimização de uma soma de quadrados penalizados

$$\min_f (\mathbf{y} - \mathbf{f})^T (\mathbf{y} - \mathbf{f}) + \lambda \mathbf{f}^T \mathbf{K} \mathbf{f},$$

em que  $\mathbf{K}$  é denominada de matriz de penalização.

Se a matriz de projeção  $\mathbf{S}_\lambda$  é simétrica e positivo definida, então a solução para  $\hat{\mathbf{f}} = \mathbf{S}_\lambda \mathbf{y}$  será uma minimização de uma soma de quadrados penalizados na forma:

$$\min_f (\mathbf{y} - \mathbf{f})^T (\mathbf{y} - \mathbf{f}) + \lambda (\mathbf{S}_\lambda^- - \mathbf{I}) \mathbf{f},$$

em que  $\mathbf{S}_\lambda^-$  é a inversa generalizada e  $\mathbf{S}_\lambda$  (HASTIE; TIBSHIRANI, 1990, p. 72).

Os graus de liberdade efetivos fornecem uma ótima maneira de se quantificar a complexidade de um *spline* suavizador. Ele sintetiza as influências combinadas do número de funções B-splines da base, a intensidade da penalização e o valor do parâmetro de suavização (EILERS; MARX; REGUERA, 2015, p. 160).

A partir dos graus de liberdade efetivos do *spline* suavizador uma estimativa da variância  $\sigma^2$  dos erros considerados no modelo não paramétrico (B.34) pode ser obtida por

$$\hat{\sigma}^2 = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)}{(m - df_\lambda)}.$$

Há situações onde a variância não possa ser considerada como constante como o foi em (B.34)

$$y_i = f(x_i) + \varepsilon_i,$$

em que  $f(\cdot)$  é o componente sistemático, uma função arbitrária que se assume existir e que possua, em algumas situações, propriedades conhecidas e  $\varepsilon$  é o componente não sistemático tal que  $\varepsilon_i \stackrel{\text{ind}}{\sim} \mathcal{N}(0, \sigma_i^2)$ .

Nestas situações, admitindo-se  $\sigma_i^2$  conhecido, pode-se atribuir um peso  $w_i = 1/\sigma_i^2$  à  $i$ -ésima observação no processo de estimação de um suavizador ponderado. Assim, considere a matriz  $\mathbf{W} = \text{diag}(\mathbf{w})$ , uma matriz diagonal cujos elementos são os pesos e seja a matriz  $\mathbf{D} = \text{diag}(\mathbf{S}\mathbf{w})$ .

O suavizador ponderado será dado por (HASTIE; TIBSHIRANI, 1990, p. 73)

$$D^{-1}SW$$

e a solução para  $\hat{f} = S_\lambda y$  será uma minimização de uma soma de quadrados ponderados penalizados na forma

$$\min_f (y - f)^T (W)(y - f) + \lambda(f^T K f)$$

com:

$$S_\lambda = (W + \lambda K)^{-1}W.$$

### B.3.8 Modelos aditivos não paramétricos: estimação por máxima verossimilhança penalizada

Um modelo aditivo não paramétrico para o vetor de observações  $Y$  pode ser proposto a partir da estrutura de um Modelo Linear Generalizado

$$\begin{aligned}\mathbb{E}(Y_i) &= \mu_i = b'(\theta_i); \\ h(\mu_i) &= \eta_i = t_i^T \beta; \\ h(b'(\theta_i)) &= \eta_i = t_i^T \beta,\end{aligned}$$

com o relaxamento da relação entre a função de ligação e o preditor na forma (GREEN; SILVERMAN, 1994, p. 98)

$$\theta_i = g(t_i),$$

em que  $\mathbb{E}(Y_i) \sim \mathcal{E}(\mu_i, \phi)$  é uma distribuição pertencente à família exponencial de distribuições com parâmetro de dispersão  $\phi$  e  $t_i$  são os valores observados de uma variável unidimensional, distintos e ordenados dentro do intervalo  $[a, b] : a < t_1 < \dots < t_n < b$  e  $g(\cdot)$  é uma função suavizadora desconhecida.

Essa flexibilidade e conveniência introduzem dois novos problemas teóricos: como representar e como determinar quão suaves serão as funções (STASINOPOULOS et al., 2017, p. 15).

De modo semelhante, quando se impõe uma penalização no processo de estimação da função suavizadora, na maximização do logaritmo da função de verossimilhança para a estimação da função  $g(\cdot)$  será procedido de igual modo, com a imposição de uma penalização, pois sempre será possível existir uma função  $g(\cdot)$  extremamente oscilante que interpole os dados <sup>16</sup> (GREEN; SILVERMAN, 1994, pp. 99–100). Portanto, no lugar de se maximizar o logaritmo da função de verossimilhança expressa na forma:

<sup>16</sup>No sentido de que seus valores estimados sejam iguais aos observados  $Y_i = b'\{g(t_i)\}$

$$l(g, \phi | y_i) = \sum_{i=1}^n \left\{ \frac{y_i g(t_i) - b[g(t_i)]}{\phi} + c(y_i, \phi) \right\} \quad (\text{B.41})$$

a função estimada  $\hat{g} \in [a, b]$  será aquela maximize o logaritmo da função de verossimilhança penalizada e expressa na forma

$$l(g, \phi) - \frac{1}{2} \lambda \int g''(t)^2 dt \quad (\text{B.42})$$

em que  $\lambda$  é o parâmetro de suavização e a função  $\hat{g}$  é a estimativa de máxima verossimilhança penalizada.

Tomando-se  $\alpha = \phi \lambda$  e multiplicando-se (B.42) por  $\phi$  faz a maximização do logaritmo da função de verossimilhança equivaler a maximizar:

$$\sum_{i=1}^n \left\{ y_i g(t_i) - b[g(t_i)] - \frac{1}{2} \alpha \int g''(t)^2 dt \right\}.$$

Definindo-se

$$l_1(g) = \sum_{i=1}^n \{ y_i g(t_i) - b[g(t_i)] \} \quad (\text{B.43})$$

de modo que (B.43) e (B.42) diferem apenas por um termo que não depende de  $g$ .

O fator  $1/2$  em (B.41) decorre que no modelo linear Normal:

$$l(g, \phi) = -\frac{1}{2\sigma^2} \sum_{i=1}^n [y_i g(t_i)].$$

Desde que  $\phi = \sigma^2$ , segue-se que  $\phi l(g, \phi) = -\frac{1}{2} \sum_{i=1}^n [y_i - g(t_i)]^2$  e a maximização de (B.49) será equivalente a se minimizar

$$\sum_{i=1}^n [y_i - g(t_i)]^2 + \alpha \int g''(t)^2 dt \quad (\text{B.44})$$

Para a maximização do logaritmo da função de verossimilhança penalizada (B.44) Green e Silverman (1994, p. 100) demonstram por teorema o método do escore de Fisher como solução

$$g_{(i+1)} = (W + \alpha K)^{-1} W z$$

com

$$z_{(i)} = g_{(i)} + \frac{y_{(i)} - b'(g_{(i)})}{b''(g_{(i)})}$$

e  $W$  é uma matriz diagonal

$$W_{(ii)} = b''(g_{(i)}).$$

### B.3.9 Algoritmo para estimação por mínimos quadrados ponderados sob penalização

Considere o logaritmo da função de verossimilhança penalizada, isto é

$$l_p(\beta) = l(\beta) - \frac{1}{2\phi} \sum_j \lambda_j \beta^T S_{[j]} \beta. \quad (\text{B.45})$$

Dado o vetor de parâmetros de suavização arbitrados  $\lambda$ , a solução de (B.45) pode ser obtida pelo método dos mínimos quadrados penalizados iterativamente ponderados (*penalized iteratively re-weighted least squares*) (WOOD, 2017, p. 251):

- tome  $\hat{\eta} = g(y) + \delta$  (usualmente  $\delta = 0$ , todavia, para assegurar que  $\hat{\eta}$  seja finito, pode-se arbitrar um pequeno valor constante;
- iterativamente prossiga até obter convergência:
  - calcule os falsos dados:  $z_i = \frac{g'(\hat{\mu}_i)(y_i - \hat{\mu}_i)}{\alpha(\mu_i) + \hat{\eta}_i}$  e os pesos iterativos:  $w_i = \frac{\alpha(\mu_i)}{\{V(\hat{\mu}_i)g'(\hat{\mu}_i)^2\}}$ ;
  - calcule uma estimativa revisada de  $\beta$  pela minimização dos quadrados mínimos ponderados

$$\hat{\beta} = \arg. \min_{\beta} \sum_i w_i (z_i - X_i \beta)^2 + \sum_j \lambda_j \beta^T S_{[j]} \beta;$$

- atualize  $\hat{\eta} = X\hat{\beta}$  e  $\hat{\mu}_i = g^{-1}(\hat{\eta}_i)$ ;
- $V(\mu)$  é a função variância determinada pela distribuição exponencial adotada;
- para o método de Newton<sup>17</sup>:  $\alpha(\mu_i) = 1 + (y_i - \hat{\mu}_i) \left( \frac{V'_i}{V_i} + \frac{g''_i}{g_i} \right)$  e para o método do escore de Fisher:  $\alpha(\mu_i) = 1$ .

### B.3.10 Funcional para penalização de *splines* de produto tensor

Considere que cada função suavizadora marginal tenha o funcional usado para quantificar a variabilidade da função que estimam, expresso na forma

$$\begin{aligned} J_x(f_x) &= \alpha^T S_x \alpha; \\ J_z(f_z) &= \delta^T S_z \delta; \text{ e} \\ J_v(f_v) &= \beta^T S_v \beta, \end{aligned}$$

em que  $J(f) = \int f''(\cdot)^2 d\cdot$  é o funcional de penalização dos *splines* cúbicos,  $S$  as matrizes de penalização e  $\alpha$ ,  $\delta$  e  $\beta$  os vetores dos coeficientes estimados dos suavizadores marginais.

<sup>17</sup>No método do escore de Fisher a Hessiana do logaritmo da função de verossimilhança é substituída por seu valor esperado.

Para a construção do funcional de penalização de  $f_{xzv}$  considere  $f_{x|zv}(x)$  e  $f_{v|xv}(v)$  como funções suavizadoras apenas de  $x$  e  $v$ , quando se mantêm constantes  $z$  e  $v$ ,  $x$  e  $v$ , respectivamente. Assim, o funcional desejado pode ser expresso na forma

$$J(f_{xzv}) = \lambda_x \int_{z,v} J_x(f_{x|zv}) dz dv + \lambda_z \int_{x,v} J_z(f_{z|xv}) dx dv + \lambda_v \int_{x,z} J_v(f_{v|xz}) dx dz, \quad (\text{B.46})$$

em que  $\lambda$  são os parâmetros de suavização que mediam o equilíbrio entre interpolação e suavização nas diferentes direções, permitindo que a penalização seja invariante com relação à escala das covariáveis. Com o funcional de penalização dos *splines* cúbicos para as funções marginais segue-se (WOOD, 2017, p. 229)

$$J(f) = \int_{x,z,v} \lambda_x \left( \frac{\partial^2 f}{\partial x^2} \right)^2 + \lambda_z \left( \frac{\partial^2 f}{\partial z^2} \right)^2 + \lambda_v \left( \frac{\partial^2 f}{\partial v^2} \right)^2 dx dz dv.$$

Para a avaliação das integrais em (B.46) considere expressar as funções marginais  $f_{x|zv}(x)$ ,  $f_{z|xv}(z)$  e  $f_{v|xz}(v)$  na forma (WOOD, 2017, pp. 230-231)

$$\begin{aligned} f_{x|zv}(x) &= \sum_{i=1}^I \alpha_i(z, v) a_i(x); \\ f_{z|xv}(z) &= \sum_{l=1}^L \delta_l(x, v) d_l(z); \text{ e} \\ f_{v|xz}(v) &= \sum_{k=1}^K \beta_k(x, z) b_k(v). \end{aligned}$$

Com o reordenamento realizado no vetor  $\beta^T = (\beta_{111}, \beta_{112}, \dots, \beta_{11K}, \beta_{121}, \beta_{122}, \dots, \beta_{1LK}, \beta_{ILK})$  determine as matrizes  $M$  tais que

$$\begin{aligned} \alpha(z, v) &= M_{zv} \beta; \\ \delta(x, v) &= M_{xv} \beta; \text{ e} \\ \beta(x, z) &= M_{xz} \beta, \end{aligned}$$

e assim segue-se que

$$\begin{aligned} J_x &= \alpha(z, v)^T S_x \alpha(z, v) = \beta^T M_{zv}^T S_x M_{zv} \beta; \\ J_z &= \delta(x, v)^T S_z \delta(x, v) = \beta^T M_{xv}^T S_z M_{xv} \beta; \text{ e} \\ J_v &= \beta(x, z)^T S_v \beta(x, z) = \beta^T M_{xz}^T S_v M_{xz} \beta, \end{aligned} \quad (\text{B.47})$$

e portanto

$$\begin{aligned}\int_{z,v} J_x(f_{x|zv}) dz dv &= \boldsymbol{\beta}^T \int_{z,v} \mathbf{M}_{zv}^T \mathbf{S}_x \mathbf{M}_{zv} dz dv \boldsymbol{\beta}; \\ \int_{x,v} J_z(f_{z|xv}) dx dv &= \boldsymbol{\beta}^T \int_{x,v} \mathbf{M}_{xv}^T \mathbf{S}_z \mathbf{M}_{xv} dx dv \boldsymbol{\beta}; \text{ e} \\ \int_{x,z} J_v(f_{v|xz}) dx dz &= \boldsymbol{\beta}^T \int_{x,v} \mathbf{M}_{xz}^T \mathbf{S}_v \mathbf{M}_{xz} dx dz \boldsymbol{\beta}.\end{aligned}$$

As integrais à direita em (B.47) podem ser resolvidas numericamente.

Em Wood (2017, p. 230) é apresentado um modo alternativo mediante a reparametrização de  $f_x$ ,  $f_z$  e  $f_v$ , obtendo-se uma solução aproximada

$$\begin{aligned}\alpha'_i &= f_x(x_i^*) \quad i = 1, \dots, I; \\ \delta'_i &= f_z(z_i^*) \quad i = 1, \dots, L; \text{ e} \\ \beta'_i &= f_v(v_i^*) \quad i = 1, \dots, K.\end{aligned}$$

Sob esta reparametrização observa-se que

$$\begin{aligned}\boldsymbol{\alpha}' &= \boldsymbol{\Gamma} \boldsymbol{\alpha}; \\ \boldsymbol{\delta}' &= \boldsymbol{\Gamma} \boldsymbol{\delta}; \text{ e} \\ \boldsymbol{\beta}' &= \boldsymbol{\Gamma} \boldsymbol{\beta},\end{aligned}$$

em que  $\Gamma_{ij} = a_i(x_j^*)$ ;  $\Gamma_{ij} = d_i(z_j^*)$  e  $\Gamma_{ij} = b_i(v_j^*)$ . As matrizes de delineamento marginais passaram a ser expressas na forma

$$\begin{aligned}\mathbf{B}'_x &= \mathbf{B}_x \boldsymbol{\Gamma}^{-1}; \\ \mathbf{B}'_z &= \mathbf{B}_z \boldsymbol{\Gamma}^{-1}; \text{ e} \\ \mathbf{B}'_v &= \mathbf{B}_v \boldsymbol{\Gamma}^{-1},\end{aligned}$$

e as matrizes dos coeficientes de penalização como:

$$\begin{aligned}\mathbf{S}'_x &= \boldsymbol{\Gamma}^{-T} \mathbf{S}_x \boldsymbol{\Gamma}^{-1}; \\ \mathbf{S}'_z &= \boldsymbol{\Gamma}^{-T} \mathbf{S}_z \boldsymbol{\Gamma}^{-1}; \text{ e} \\ \mathbf{S}'_v &= \boldsymbol{\Gamma}^{-T} \mathbf{S}_v \boldsymbol{\Gamma}^{-1}.\end{aligned}$$

Seguem-se então as aproximações

$$\begin{aligned}
\int_{z,v} J_x(f_{x|zv}) dz dv &\approx h \sum_{l,k} J_x f(f_{x|z_l^* v_k^*}); \\
\int_{x,v} J_z(f_{z|xv}) dx dv &\approx h \sum_{i,k} J_z f(f_{z|x_i^* v_k^*}); \text{ e} \\
\int_{x,z} J_v(f_{v|xz}) dx dz &\approx h \sum_{i,l} J_v f(f_{v|x_i^* z_l^*}),
\end{aligned} \tag{B.48}$$

em que  $h$  é alguma constante de proporcionalidade relacionada ao espaçamento entre de  $x_i^*$ ,  $z_l^*$  e  $v_k^*$ . Assim, as Equações B.48 são representadas por

$$\begin{aligned}
J_x^*(f_{xzv}) &= \beta^T \tilde{\mathbf{S}}_x \beta; \\
J_z^*(f_{xzv}) &= \beta^T \tilde{\mathbf{S}}_z \beta; \text{ e} \\
J_v^*(f_{xzv}) &= \beta^T \tilde{\mathbf{S}}_v \beta,
\end{aligned}$$

em que

$$\begin{aligned}
\tilde{\mathbf{S}}_x &= \mathbf{S}'_x \otimes \mathbf{I}_L \otimes \mathbf{I}_K; \\
\tilde{\mathbf{S}}_z &= \mathbf{I}_I \otimes \mathbf{S}'_z \otimes \mathbf{I}_K; \text{ e} \\
\tilde{\mathbf{S}}_v &= \mathbf{I}_L \otimes \mathbf{I}_I \otimes \mathbf{S}'_v,
\end{aligned} \tag{B.49}$$

resultando na aproximação

$$J(f_{xzv}) \approx J^*(f_{xzv}) = \lambda_x J_x^*(f_{xzv}) + \lambda_z J_z^*(f_{xzv}) + \lambda_v J_v^*(f_{xzv}).$$

## B.4 MODELOS ADITIVOS GENERALIZADOS PARA LOCAÇÃO, ESCALA E FORMA

### B.4.1 Distribuições de probabilidade implementadas para os *GAMLSS*

Tabela B.2: Distribuições contínuas disponíveis na implementação computacional dos modelos *GAMLSS*

Distribuição	Denominação	Faixa de valores	Funções de ligação			
			$\mu$	$\sigma$	$\nu$	$\tau$
beta	BE	(0,1)	<i>logit</i>	<i>logit</i>	-	-
Box&Cox Cole&Green	BCCG	$(0, \infty)$	ident.	log	ident.	-
Box&Cox Cole&Green or.	BCCGo	$(0, \infty)$	log	log	ident.	-
Box&Cox pot. exp.	BCPE	$(0, \infty)$	ident.	log	ident.	log
Box&Cox pot. exp. original	BCPEo	$(0, \infty)$	log	log	ident.	log
Box&Cox “t”	BCT	$(0, \infty)$	ident.	log	ident.	log
Box&Cox “t” original	BCTO	$(0, \infty)$	log	log	ident.	log
exponencial	EXP	$(0, \infty)$	log	-	-	-

Tabela B.2: Distribuições contínuas disponíveis na implementação computacional dos modelos *GAMLSS*

(continuação)

Distribuição	Denominação	Faixa de valores	Funções de ligação			
			$\mu$	$\sigma$	$\nu$	$\tau$
Gaussiana exponencial	exGAUS	$(-\infty, \infty)$	ident.	log	log	-
exponencial beta gen.	EGB2()	$(-\infty, \infty)$	ident.	log	log	log
gamma	GA	$(0, \infty)$	log	log	-	-
beta gen. tipo I	GB1	$(0, 1)$	<i>logit</i>	<i>logit</i>	log	log
beta gen. tipo II	GB2	$(0, \infty)$	log	log	log	log
gama gen.	GG	$(0, \infty)$	log	log	ident.	-
Gaussiana inv. gen.	GIG	$(0, \infty)$	log	log	ident.	-
“t”generalizada	GT	$(-\infty, \infty)$	ident.	log	log	log
Gumbel	GU	$(-\infty, \infty)$	ident.	log	-	-
gama inversa	IGAMMA	$(0, \infty)$	log	log	-	-
inversa Gaussiana	IG	$(0, \infty)$	log	log	-	-
Johnson SU rep.	JSU	$(-\infty, \infty)$	ident.	log	ident.	log
Johnson SU or..	JSUo	$(-\infty, \infty)$	ident.	log	ident.	log
logística	LO	$(-\infty, \infty)$	ident.	log	-	-
logística Normal	LOGITNO	$(0,1)$	ident.	log	-	-
log-Normal	LOGNO	$(0, \infty)$	ident.	log	-	-
log Normal 2	LOGNO2	$(0, \infty)$	log	log	-	-
log Normal (Box&Cox)	LNO	$(0;1)$	ident.	log	fixo	-
NET	NET	$(-\infty, \infty)$	ident.	log	fixo	-
Normal	NO,NO2	$(-\infty, \infty)$	ident.	log	-	-
família Normal	NOF	$(-\infty, \infty)$	ident.	log	-	-
Pareto 2	PARETO2	$(0, \infty)$	log	log	-	-
Pareto 2 original	PARETO2o	$(0, \infty)$	log	log	-	-
Pareto 2 repar	GP	$(0, \infty)$	log	log	-	-
power exponential	PE	$(-\infty, \infty)$	ident.	log	log	-
reverse gen. extrema	RGE	$y > \mu - (\sigma/\nu)$	ident.	log	log	-
reverse Gumbel	RG	$(-\infty, \infty)$	ident.	log	-	-
sinh-arcsinh	SHASH	$(-\infty, \infty)$	ident.	log	log	log
sinh-arcsinh original	SHASHo	$(-\infty, \infty)$	ident.	log	ident.	log
sinh-arcsinh original 2	SHASHo2	$(-\infty, \infty)$	ident.	log	ident.	log
skew Normal tipo 1	SN1	$(-\infty, \infty)$	ident.	log	ident.	-
skew Normal tipo 2	SN2	$(-\infty, \infty)$	ident.	log	log	-
skew power exp. tipo 1	SEP1	$(-\infty, \infty)$	ident.	log	ident.	log



Tabela B.2: Distribuições contínuas disponíveis na implementação computacional dos modelos *GAMLSS*

(continuação)

Distribuição	Denominação	Faixa de valores	Funções de ligação			
			$\mu$	$\sigma$	$\nu$	$\tau$
skew power exp. tipo 2	SEP2	$(-\infty, \infty)$	ident.	log	ident.	log
skew power exp. tipo 3	SEP3	$(-\infty, \infty)$	ident.	log	log	log
skew power exp. tipo 4	SEP4	$(-\infty, \infty)$	ident.	log	log	log
skew t tipo 1	ST1	$(-\infty, \infty)$	ident.	log ident.	log	
skew t tipo 2	ST2	$(-\infty, \infty)$	ident.	log	ident.	log
skew t tipo 3	ST3	$(-\infty, \infty)$	ident.	log	log	log
skew t tipo 3 repar	SST	$(-\infty, \infty)$	ident.	log	log	log-2
skew t tipo 4	ST4	$(-\infty, \infty)$	ident.	log	log	log
skew t tipo 5	ST5	$(-\infty, \infty)$	ident.	log	ident.	log
família t	TF	$(-\infty, \infty)$	ident.	log	log	-
família t repar	TF2	$(-\infty, \infty)$	ident.	log	log-2	-
Weibull	WEI	$(0, \infty)$	log	log	-	-
Weibull (PH)	WEI2	$(0, \infty)$	log	log	-	-
Weibull (com média $\mu$ )	WEI3	$(0, \infty)$	log	log	-	-

Fonte: Adaptado de Stasinopoulos et al. (2017)

#### B.4.2 Funções de ligação implementadas para os *GAMLSS*

#### B.4.3 Algoritmo RS de estimação dos *GAMLSS*

Sob a premissa de que as observações são independentes, um *GAMLSS* estritamente paramétrico é estimado pela maximização do logaritmo de sua função de verossimilhança, e para um modelo não paramétrico, pela maximização do logaritmo de sua função de verossimilhança penalizada, respectivamente expressas nas formas (STASINOPOULOS et al., 2017)

$$l = \sum_{i=1}^n \log f(y_i | \mu_i, \sigma_i, \nu_i, \tau_i),$$

e:

$$l_p = l - \frac{1}{2} \sum_{k=1}^4 \sum_{j=1}^{J_k} \gamma_{jk}^T \mathbf{G}_{jk}(\lambda_{jk}) \gamma_{jk}. \quad (\text{B.50})$$

Tabela B.3: Distribuições discretas disponíveis na implementação computacional dos *GAMLSS*

Distribuição	Denominação	Faixa de valores	Funções de ligação		
			$\mu$	$\sigma$	$\nu$
beta binomial	BB	$\{0, 1, \dots, n\}$	<i>logit</i>	log	-
binomial	BI	$\{0, 1, \dots, n\}$	<i>logit</i>	-	-
Delaporte	DEL	$\{0, 1, 2, \dots\}$	log	log	<i>logit</i>
geometric	GEOM	$\{0, 1, 2, \dots\}$	log	-	-
logaritmica	LG	$\{1, 2, 3, \dots\}$	<i>logit</i>	-	-
binomial negativa tipo I	NBI	$\{0, 1, 2, \dots\}$	log	log	-
binomial negativa tipo II	NBII	$\{0, 1, 2, \dots\}$	log	log	-
Poisson	PO	$\{0, 1, 2, \dots\}$	log	-	-
Poisson inversa Gaussiana	PIG	$\{0, 1, 2, \dots\}$	log	log	-
Sichel	SI	$\{0, 1, 2, \dots\}$	log	log	ident.
Sichel (com média $\mu$ )	SICHEL	$\{0, 1, 2, \dots\}$	log	log	ident.
Waring	WARING	$\{0, 1, 2, \dots\}$	log	log	-
Yule (com média $\mu$ )	YULE	$\{0, 1, 2, \dots\}$	log	-	-
beta binomial zero alterada	ZABB	$\{0, 1, 2, \dots, n\}$	<i>logit</i>	log	<i>logit</i>
binomial zero alterada	ZABI	$\{0, 1, 2, \dots, n\}$	<i>logit</i>	<i>logit</i>	-
logaritmica zero alterada	ZALG	$\{0, 1, 2, \dots\}$	<i>logit</i>	<i>logit</i>	-
binomial negativa zero alterada	ZANBI	$\{0, 1, 2, \dots\}$	log	log	<i>logit</i>
Poisson zero alterada	ZAP	$\{0, 1, 2, \dots, n\}$	<i>logit</i>	<i>logit</i>	-
beta binomial inflada de zeros	ZIBB	$\{0, 1, 2, \dots\}$	<i>logit</i>	log	<i>logit</i>
binomial inflada de zeros	ZIBI	$\{0, 1, \dots, n\}$	<i>logit</i>	<i>logit</i>	-
binomial negativa inflada de zeros	ZINBI	$\{0, 1, 2, \dots\}$	log	log	<i>logit</i>
Poisson inflada de zeros	ZIP	$\{0, 1, 2, \dots\}$	log	<i>logit</i>	-
Poisson inflada de zeros (com média $\mu$ )	ZIP2	$\{0, 1, 2, \dots\}$	log	<i>logit</i>	-
Poisson inversa Gaussiana inflada de zeros	ZIPIG	$\{0, 1, 2, \dots\}$	log	log	<i>logit</i>

Fonte: Adaptado de Stasinopoulos et al. (2017)

Tabela B.4: Distribuições mistas disponíveis na implementação computacional dos *GAMLSS*

Distribuição	Denominação	Faixa de valores	Funções de ligação			
			$\mu$	$\sigma$	$\nu$	$\tau$
beta inflada (em 0)	BE0I	$[0, 1)$	<i>logit</i>	log	<i>logit</i>	-
beta inflada (em 0)	BEINF0	$[0, 1)$	<i>logit</i>	<i>logit</i>	log	-
beta inflada (em 1)	BEZI	$(0, 1]$	<i>logit</i>	log	<i>logit</i>	-
beta inflada (em 1)	BEINF1	$(0, 1]$	<i>logit</i>	<i>logit</i>	log	-
beta inflada (em 0 e 1)	BEINF	$[0, 1]$	<i>logit</i>	<i>logit</i>	log	log
Gaussiana ajustada de zero	ZAGA	$[0, \infty]$	log	log	<i>logit</i>	-
Inversa Gaussiana ajustada de zero	ZAIG	$[0, \infty]$	log	log	<i>logit</i>	-

Fonte: Adaptado de Stasinopoulos et al. (2017, p. 160)

Rigby e Stasinopoulos (2005) demonstraram que a maximização de (B.50) conduz à matriz de suavização (reduzida)  $\mathbf{S}_{jk}$  aplicada aos resíduos parciais  $\epsilon_{jk}$  para se obter a atualização da estimativa do preditor aditivo  $\mathbf{Z}_{jk}\gamma_{jk}$  dentro do algoritmo de retroajuste

$$\mathbf{S}_{jk} = \mathbf{Z}_{jk}(\mathbf{Z}_{jk}^T \mathbf{W}_{kk} \mathbf{Z}_{jk} + \mathbf{G}_{jk})^{-1} \mathbf{Z}_{jk}^T \mathbf{W}_{kk}, \quad (\text{B.51})$$

Tabela B.5: Funções de ligação disponíveis na implementação computacional dos *GAMLSS*

Faixa de valores	Funções de ligação	Fórmula ( $g(\theta)$ )
$-\infty \leq \theta \leq \infty$	identidade	$\theta$
$0 \leq \theta \leq \infty$	logaritmo	$\log(\theta)$
	raiz quadrada	$\sqrt{\theta}$
	inversa	$1/\theta$
	quadrática inversa	$1/\theta^2$
	quadrática	$\theta^2$
$0 \leq \theta \leq 1$	<i>logit</i>	$\log(\theta/(1-\theta))$
	<i>probit</i>	$\Phi^{-1}(\theta)$
	<i>cauchit</i>	$\tan(\pi(\theta - 0,05))$
	<i>cloglog</i>	$\log(-\log(1 - \theta))$
$1 \leq \theta \leq \infty$	<i>logshiftto1</i>	$\log(\theta - 1)$
$2 \leq \theta \leq \infty$	<i>logshiftto2</i>	$\log(\theta - 2)$
$0,00001 \leq \theta \leq \infty$	<i>logshiftto0</i>	$\log(\theta - 0,00001)$

Fonte: Adaptado de Stasinopoulos et al. (2017)

para  $j = 1, \dots, J_k$  e  $k = 1, \dots, p$ , em que  $\mathbf{W}_{kk}$  é a matriz diagonal de pesos. Diferentes formas de  $\mathbf{Z}_{jk}$  e  $\mathbf{G}_{jk}$  resultam em diferentes termos no preditor  $\boldsymbol{\eta}_k$  para  $k = 1, \dots, p$ . Os hiperparâmetros  $\boldsymbol{\lambda}$  podem ser estimados ou arbitrados (fixos).

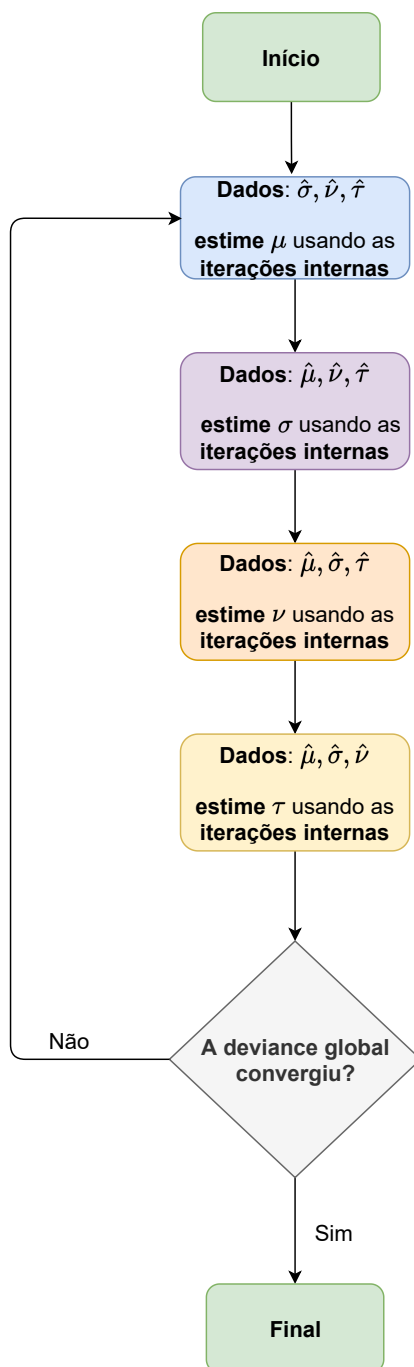
A estimação pela maximização (a *posteriori*) do logaritmo da função de verossimilhança penalizada para  $\boldsymbol{\beta}$  e  $\boldsymbol{\gamma}$ , dado  $\boldsymbol{\lambda}$ , pode se dar por meio de dois algoritmos (RIGBY; STASINOPOULOS, 2005, p. 538) CG: Coole Green (COLE; GREEN, 1992) e RS: Rigby e Stasinopoulos. Nos modelos estritamente paramétricos, os algoritmos retornam estimativas de máxima verossimilhança (EMV) para  $\boldsymbol{\beta}$  ao passo que nos modelos não paramétricos as estimativas máximas *a posteriori* obtidas para  $\boldsymbol{\beta}$  e  $\boldsymbol{\gamma}$  são para valores fixos de  $\boldsymbol{\lambda}$ . Nessa última situação, a probabilidade *a posteriori* é dada por

O algoritmo RS é composto por três *loops* aninhados a saber:

- o *loop* das iterações externas, detalhado na Figura B.16;
- o *loop* das iterações internas ou escore local, detalhado na Figura B.17 (GLIM: *Generalized Linear Interactive Modeling*, desenvolvido pelo grupo de trabalho sobre computação estatística da *Royal Statistical Society* presidido inicialmente por John Nelder e lançada em 1993 por Brian Francis; Mick Green; Clive Payne);
- o *loop* de retroajuste modificado, detalhado na Figura B.18.

O ciclo externo do algoritmo RS maximiza o logaritmo da função de verossimilhança (penalizada) com respeito aos vetores  $\boldsymbol{\beta}_k$  e  $\boldsymbol{\gamma}_{jk}$  para  $j = 1, \dots, J_k$  de modo sucessivo para cada parâmetro da distribuição  $\theta_k$  com  $k = 1, \dots, p$ . A cada cálculo do algoritmo todos os novos valores (atualizados) são usados.

Figura B.16: Diagrama do ciclo externo do algoritmo RS de maximização do logaritmo da função de verossimilhança (penalizada)



Fonte: Adaptado de Stasinopoulos et al. (2017)

Considere a seguinte notação:  $\theta_1 = \mu$ ,  $\theta_2 = \sigma$ ,  $\theta_3 = \nu$  e  $\theta_4 = \tau$ . O ciclo externo do algoritmo RS promove a estimação sucessiva de  $\hat{\mu}$ ,  $\hat{\sigma}$ ,  $\hat{\nu}$  e  $\hat{\tau}$  pela chamada do algoritmo do ciclo interno para o ajuste específico do preditor de cada um dos parâmetros da distribuição da resposta (cf. Figura B.19), um algoritmo baseado na implementação

computacional *GLIM*. O algoritmo do ciclo interno encontra-se baseado na ideia de obter convergência do desvio global pelo processo conhecido como Método Iterativo dos Mínimos Quadrados Ponderados.

Admita  $z_k$  como sendo as variáveis dependentes ajustadas e  $u_k$  como as funções escore para a estimação do  $k$ -ésimo parâmetro  $\theta_k$  da distribuição da resposta (STASINOPOULOS; RIGBY, 2007)

$$z_k = \eta_k + W_{kk}^{-1} u_k = \eta_k + w_k^{-1} \circ u_k,$$

em que  $z_k = z_{k1}, \dots, z_{kn}^T$ ,  $\eta_k = \eta_{k1}, \dots, \eta_{kn}^T$ ,  $w_k = w_{k1}, \dots, w_{kn}^T$  e  $u_k = u_{k1}, \dots, u_{kn}$  são vetores de comprimento  $n$ , para  $k = 1, \dots, p$ ;  $W_{kk}$  é uma matriz diagonal cujos elementos são os pesos iterativos  $w_k$ ;  $\eta_k = g_k(\theta_k)$  é o vetor preditor do  $k$ -ésimo parâmetro  $\theta_k$  para  $k = 1, 2, 3, 4$  correspondentes a  $\mu, \sigma, \nu$  e  $\tau$ ; ( $\circ$ ) define o operador do produto *Hadamard* (também denominado de produto elemento a elemento ou produto *Schur*) é um operador que toma duas matrizes  $A_{(n,m)}$  e  $B_{(n,m)}$  e retorna uma terceira  $C_{(n,m)}$ , onde cada um de seus elementos  $c_{ij}$  corresponde ao produto dos elementos  $a_{ij}$  e  $b_{ij}$  das matrizes originais. Recebem estes nomes em homenagem aos matemáticos Jacques Hadamard e Issai Schur respectivamente) elemento a elemento de  $w_k^{-1}$  por  $u_k$  como  $w_k^{-1} \circ u_k = (w_{k1}^{-1} u_{k1}, w_{k2}^{-1} u_{k2}, \dots, w_{kn}^{-1} u_{kn})^T$  e  $u_k$  é o vetor da  $k$ -ésima função escore (a primeira derivada do logaritmo da função de verossimilhança com relação ao preditor  $\eta_k$ ) dado por

$$u_k = \frac{\partial l}{\partial \eta_k} = \left( \frac{\partial l}{\partial \theta_k} \right) \circ \left( \frac{d\theta_k}{d\eta_k} \right),$$

tais que

$$\begin{aligned} \frac{\partial l}{\partial \eta_k} &= \left( \frac{\partial l_1}{\partial \eta_{k1}}, \dots, \frac{\partial l_n}{\partial \eta_{kn}} \right)^T; \\ \frac{\partial l}{\partial \theta_k} &= \left( \frac{\partial l_1}{\partial \theta_{k1}}, \dots, \frac{\partial l_n}{\partial \theta_{kn}} \right)^T; \text{ e} \\ \frac{d\theta_k}{d\eta_k} &= \left( \frac{d\theta_{k1}}{d\eta_{k1}}, \dots, \frac{d\theta_{kn}}{d\eta_{kn}} \right)^T, \end{aligned}$$

para  $k = 1, \dots, p$ .

Os pesos iterativos  $w_k$  para  $k = 1, \dots, p$  são assim definidos

$$w_k = -f_k \circ \left( \frac{d\theta_k}{d\eta_k} \right) \circ \left( \frac{d\theta_k}{d\eta_k} \right),$$

em que  $f_k$  pode ser definida de três modos diferentes

$$\begin{aligned} \mathbf{f}_k &= \mathbb{E} \left( \frac{\partial^2 l}{\partial \boldsymbol{\theta}_k^2} \right) \quad (\text{se a esperança existe}); \\ \mathbf{f}_k &= \frac{\partial^2 l}{\partial \boldsymbol{\theta}_k^2}; \\ \mathbf{f}_k &= - \left( \frac{\partial l}{\partial \boldsymbol{\theta}_k} \right) \circ \left( \frac{\partial l}{\partial \boldsymbol{\theta}_k} \right), \end{aligned}$$

tal que

$$\frac{\partial^2 l}{\partial \boldsymbol{\theta}_k^2} = \left( \frac{\partial^2 l_1}{\partial \theta_{k1}^2}, \dots, \frac{\partial^2 l_n}{\partial \theta_{kn}^2} \right)^T,$$

para  $k = 1, \dots, p$ .

Depreende-se assim que as funções escore  $\mathbf{u}_k$  podem assumir três formas distintas a depender:

- da informação esperada, resultando em um algoritmo pelo método do escore de Fisher;
- da informação observada, resultando em um algoritmo pelo método de Newton-Raphson; ou,
- do produto das funções escore, resultando em um algoritmo pelo método *quasi*-Newton-Raphson,

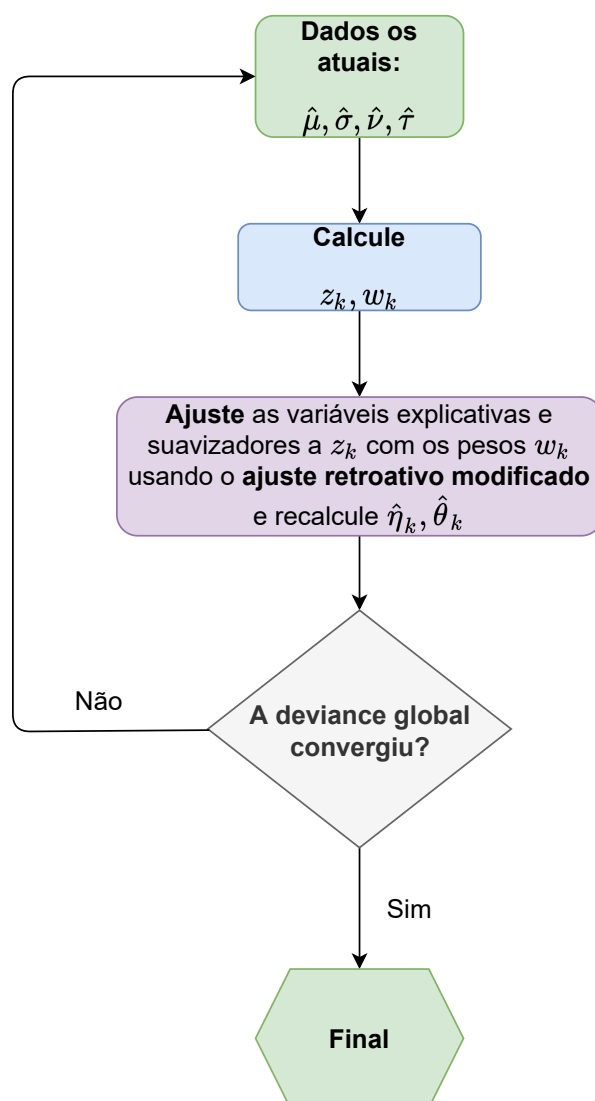
#### B.4.3.1 Pseudocódigo do algoritmo RS

- Passo 01: Inicie os valores ajustados:  $\boldsymbol{\theta}_k^{[1,1]}$  e os efeitos aleatórios:  $\boldsymbol{\gamma}_{jk}$ , para  $j = 1, \dots, J_k$  e  $k = 1, \dots, p$ . Avalie os preditores iniciais  $\boldsymbol{\eta}_k^{[1,1]} = \mathbf{g}_k(\boldsymbol{\theta}_k^{[1,1]})$  para  $k = 1, \dots, p$ .
- Passo 02: Inicie o ciclo externo ( $r = 1, 2, \dots$ ) até obter convergência. Para  $k = 1, \dots, p$ :
  - Inicie o ciclo interno ( $i = 1, 2, \dots$ ) até obter convergência:
    - \* avalie os atuais  $\mathbf{u}_k^{[r,i]}$ ,  $\mathbf{W}_{kk}^{[r,i]}$  e  $\mathbf{z}_k^{[r,i]}$ ;
    - \* inicie o ciclo de retroajuste ( $m = 1, \dots$ ) até obter convergência;
    - \* regrida os atuais resíduos parciais  $\epsilon_{0k}^{[r,i,m]} = \mathbf{z}_k^{[r,i]} - \sum_{j=1}^{J_k} \mathbf{Z}_{jk} \boldsymbol{\gamma}_{jk}^{[r,i,m]}$  contra a matriz de delineamento  $\mathbf{X}_k$  usando os pesos iterativos  $\mathbf{W}_{kk}^{[r,i]}$  para obter as estimativas atualizadas dos parâmetros  $\boldsymbol{\beta}_k^{[r,i,(m+1)]}$ ;
    - \* para  $j = 1, \dots, J_k$  suavize os resíduos parciais  $\epsilon_{jk}^{[r,i,m]} = \mathbf{z}_k^{[r,i]} - \mathbf{X}_k \boldsymbol{\beta}_k^{[r,i,(m+1)]} - \sum_{t=1, t \neq j}^{J_k} \mathbf{Z}_{tk} \boldsymbol{\gamma}_{tk}^{[r,i,c]}$  usando a matriz de suavização  $\mathbf{S}_{jk}$  dada por (B.51) para obter a atualização (e atual) termo do preditor aditivo  $\mathbf{Z}_{jk} \boldsymbol{\gamma}_k^{[r,i,(m+1)]}$ ;
    - \* na convergência de  $\boldsymbol{\beta}_k^{[r,i,.]}$  e  $\mathbf{Z}_{jk} \boldsymbol{\gamma}_{jk}^{[r,i,.]}$  fim do ciclo de retroajuste e faça  $\boldsymbol{\beta}_k^{[r,(i+1)]} = \boldsymbol{\beta}_k^{[r,.]}$  e  $\boldsymbol{\gamma}_{jk}^{[r,(i+1)]} = \boldsymbol{\gamma}_{jk}^{[r,i,.]}$  para  $j = 1, \dots, J_k$ ; caso contrário, atualize  $m$  e prossiga no ciclo de retroajuste;

- \* calcule e atualize  $\boldsymbol{\eta}_k^{[r,(i+1)]}$  e  $\boldsymbol{\theta}_k^{[r,(i+1)]}$ ;
- Fim do ciclo interno na convergência de  $\boldsymbol{\beta}_k^{r,\cdot}$  e dos termos aditivos  $\mathbf{Z}_{jk}\boldsymbol{\gamma}_{jk}^{[r,\cdot]}$ , faça  $\boldsymbol{\beta}_k^{[(r+1),1]} = \boldsymbol{\beta}_k^{[r,\cdot]}$ ,  $\boldsymbol{\gamma}_{jk}^{[(r+1),1]} = \boldsymbol{\gamma}_{jk}^{[r,\cdot]}$ , para  $j = 1, \dots, J_k$ ,  $\boldsymbol{\eta}_k^{[(r+1),1]} = \boldsymbol{\eta}_k^{[r,\cdot]}$  e  $\boldsymbol{\theta}_k^{[(r+1),1]} = \boldsymbol{\theta}_k^{[r,\cdot]}$ ; caso contrário atualize  $i$  e prossiga no ciclo interno;
- Passo 03: Atualize o valor de  $k$ ;
- Passo 04: Fim do ciclo externo se a mudança no logaritmo da função de verossimilhança (penalizada) foi suficientemente reduzida; caso contrário, atualize  $r$  e prossiga no ciclo externo.

O algoritmo RS difere do CG porque a matriz diagonal de pesos  $\mathbf{W}_{kk}$  é avaliada (atualizada) durante (*dentro*) a estimação de cada parâmetro  $\boldsymbol{\theta}_k$  no qual todas as matrizes  $\mathbf{W}_{ks}$  para  $k = 1, \dots, p$  e  $s = 1, \dots, p$  são avaliadas *depois* de se estimar todos os parâmetros  $\boldsymbol{\theta}_k$  para  $k = 1, \dots, p$  (RIGBY; STASINOPOULOS, 2005).

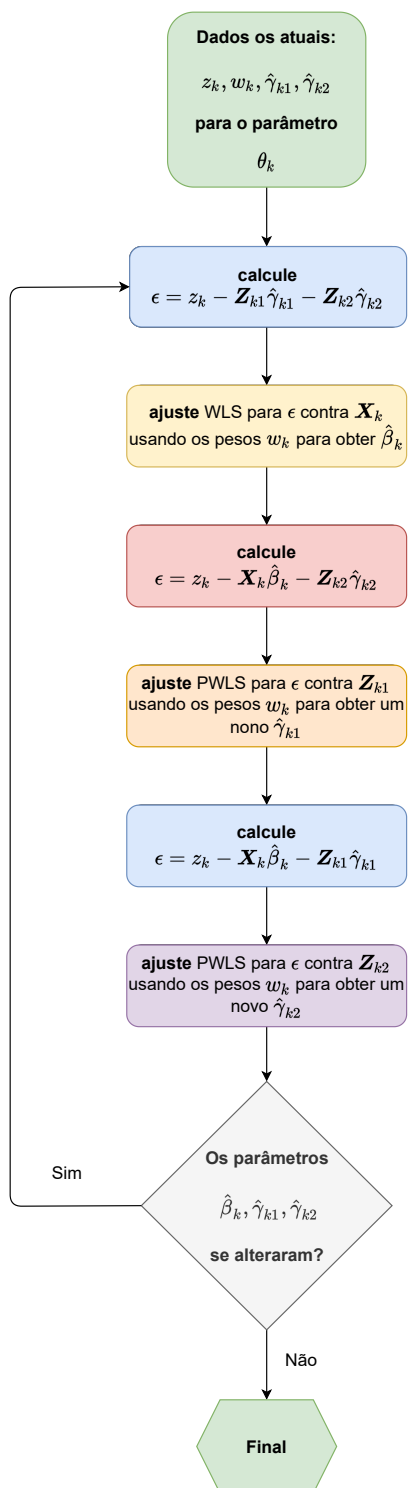
Figura B.17: Diagrama do ciclo interno do algoritmo RS de maximização do logaritmo da função de verossimilhança (penalizada)



Fonte: Adaptado de Stasinopoulos et al. (2017)

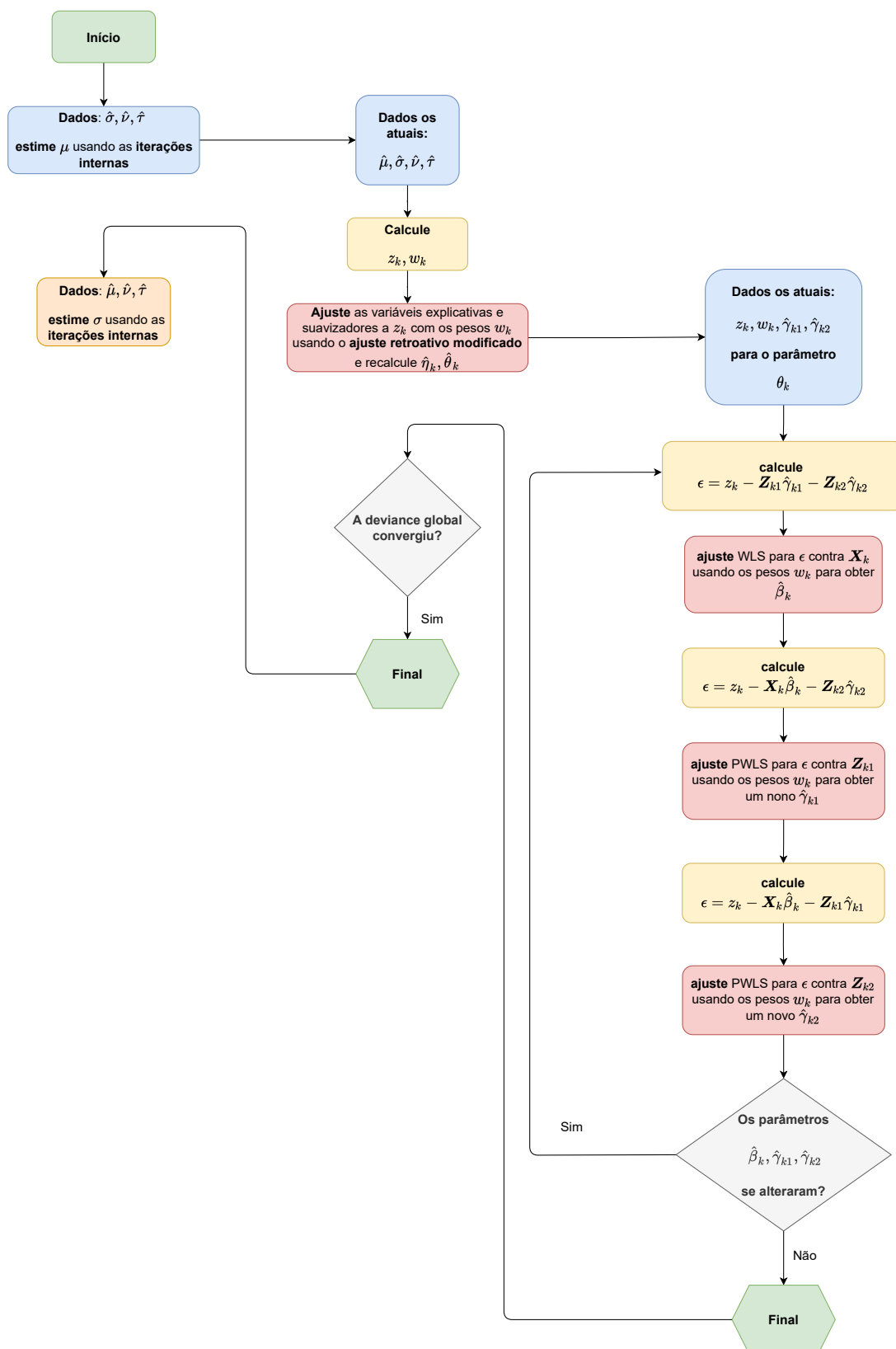


Figura B.18: Diagrama do algoritmo de retroajuste modificado do algoritmo RS de maximização do logaritmo da função de verossimilhança (penalizada)



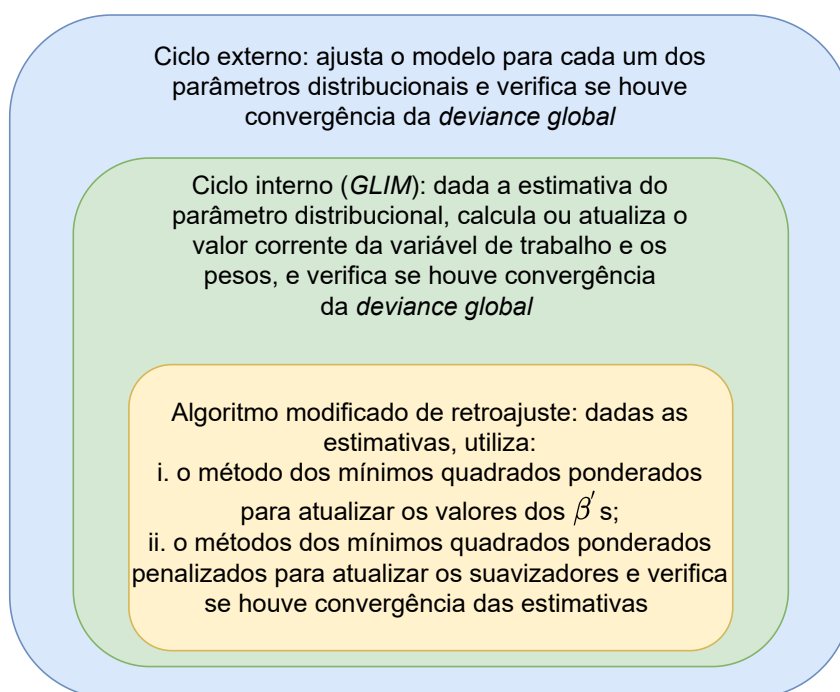
Fonte: Adaptado de Stasinopoulos et al. (2017)

Figura B.19: Interconexão (parcial) dos vários *loops* internos do algoritmo RS de maximização do logaritmo da função de verossimilhança (penalizada)



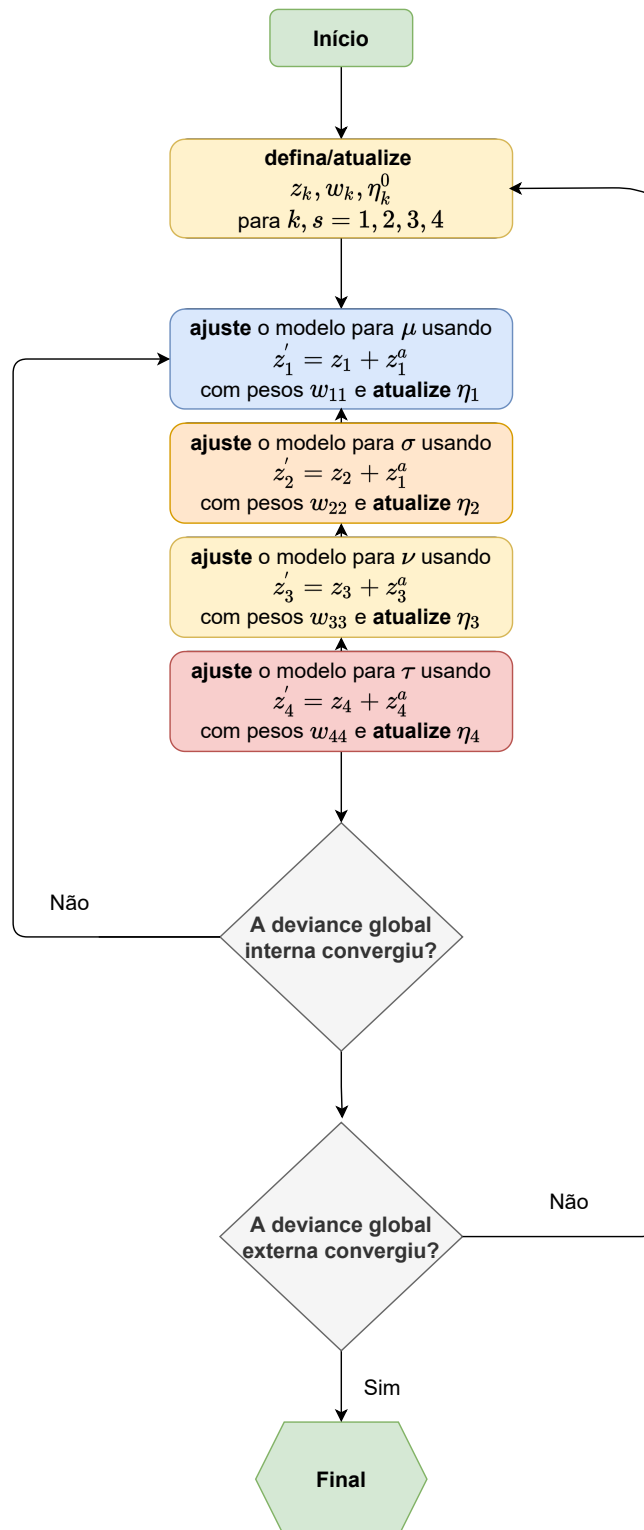
Fonte: Adaptado de Stasinopoulos et al. (2017)

Figura B.20: Esquema do algoritmo RS de estimação dos modelos *GAMLSS*



Fonte: Adaptado de Thomas (2017)

Figura B.21: Iterações internas e externas no algoritmo CG



Fonte: Adaptado de Stasinopoulos et al. (2017)

## C MATERIAIS E MÉTODOS

### C.1 O CAFÉ, OS INGLESES E A CIDADE DE LONDRINA

Planta originária das terras quentes do nordeste da África, a primeira menção ao café consta em manuscritos do Iêmen (c. 575) para onde a planta foi levada de Kafa, no interior da Etiópia.

Da península arábica onde as primeiras plantações comerciais se desenvolveram, o hábito que aí floresceu expandiu-se ao Egito, Turquia, Itália, Holanda, França, Áustria e Inglaterra pelo porto de Moka no Iêmen, onde os frutos, sem o pergaminho (para impedir sua germinação e manter o monopólio árabe), eram embarcados.

A bebida só chegou à Europa em 1615, mais especificamente em Veneza, e de lá espalhou-se pelo continente onde, como oportuna alternativa ao chá, pois era uma planta sob monopólio Inglês. Pela sua rápida aceitação social, logo despertou o interesse dos holandeses e franceses pelo seu cultivo que obtiveram as primeiras mudas em 1616 e 1644, respectivamente.

As primeiras experiências no cultivo se deram na ilha de Java (1658) pela Holanda e na ilha de Bourbon (1715) pela França. Na América do Sul, o cultivo do café foi introduzido nas colônias francesas (Guiana) e holandesas (Suriname, 1718) (MARTINS, 2008, pp. 17–28).

No Brasil o café foi introduzido pelo comerciante português Francisco de Mello Palheta em seu retorno ao país em 1727 de uma expedição à Guiana Francesa sob o pretexto oficial de mediar uma disputa de fronteira entre a França e a Holanda.

A cultura iniciada no final do séc. XVIII em São Luiz do Maranhão chegou nas primeiras décadas do século seguinte até o Vale do Rio Paraíba em São Paulo, mas ainda longe de ser um cultivo com a relevância econômica que se veria nas décadas seguintes, pelas dificuldades que cultivos e produtos mais tradicionais (algodão, arroz, açúcar, fumo, couro) experimentariam com o surgimento de novos mercados produtores mais atraentes e a extinção do tráfico de escravos.

De um cultivo secundário, cuja produção tinha por objetivo apenas a satisfação de um segmento mais abastado da população desejosa em imitar hábitos europeus, o café passou a responder em 1830 por 18 % do total das exportações brasileiras e já em 1850 ascendia ao posto de produto com maior participação na exportações (40 %). Em 1900 o Brasil respondia por 80 % da produção mundial.

Historicamente denominada como *Norte do Paraná*, esta extensão territorial com área aproximada de 100 mil quilômetros quadrados é a soma dos vales férteis formados pelos afluentes situados na margem esquerda dos rios Paraná e Paranapanema, no arco por eles formado entre as cidades de Cambará e Guaíra, com solos de terra roxa iguais aos do Planalto de

Piratininga, no interior paulista, que perfizeram uma feliz união com a cultura cafeeira, meados do século XIX.

A necessidade de escoamento da produção fez surgir na província de São Paulo uma demanda crescente por aço para a construção de ferrovias e pontes para a criação de uma malha de serviços de transportes que despertou na Inglaterra o interesse no fornecimento de materiais e capital, surgindo em 1886 a São Paulo *Railway* ligando Jundiaí ao porto de Santos e que que, em poucos anos, tornou-se a empresa ferroviária mais lucrativa da América Latina (SANTOS, 2013, p. 19).

A disponibilidade de capitais por parte dos produtores de café e o clima de internacional de confiança dos investidores nos investimentos realizados facilitaram a expansão da malha e com ela o desenvolvimento sócio-econômico dessa região do país. Os lucros obtidos pela alta produtividade das plantações na terra roxa eram criteriosamente reinvestidos não só em novas plantações (“a marcha para o Oeste”) mas também em novas ferrovias e indústrias.

Em 1908 a ferrovia chega à cidade de Ourinhos onde o Major Antônio Barbos Ferraz adquire uma extensa gleba junto à divisa com o estado do Paraná onde passa a se dedicar ao cultivo de um milhão de pés de café, após vender sua próspera propriedade nas imediações de Ribeirão Preto.

Junto a outros produtores fundam a Companhia Ferroviária São Paulo-Paraná com o propósito de estender um ramal ferroviária de Ourinhos a Cambará, empreitada que demandaria aporte de capital inglês.

Em 1924 desembarca no Brasil uma missão econômica chefiada por Edwin Montagu (*Lord Montagu*), convidada pelo presidente Arthur Bernardes para equacionar a dívida nacional com aquele país. Dentre seus membros estava Simon Joseph Fraser (*Lord Lovat*) que veio com a missão de aplicar capitais da *Sudan Plantations* com o propósito exclusivo de obter algodão e suprir a crescente demanda da indústria têxtil inglesa (SANTOS, 2013, p. 39).

Pode-se entender que o nascimento de Londrina decorreu da recusa do Major Barbosa Ferraz à oferta de Simon Fraser (*Lord Lovat*) para adquirir suas terras feita num jantar na sede da fazenda Água do Bugre em janeiro de 1924, nos arredores de Cambará, do qual participaram também o prefeito de Cambará, Willie Davids e seu engenheiro Gastão de Mesquita Filho.

Para ele e tantos outros que o seguiram na abertura de novas frentes agrícolas, a terra representava tudo quanto era de mais precioso, “[...] é parte do corpo e da alma de quem a possui e a usa para semear e colher [...]” (SANTOS, 2013, p. 46).

Ainda perplexo com a recusa, foi Gastão de Mesquita que desperta a atenção de Fraser para as férteis terras existentes no Paraná, vendidas a preço muito baixo pelo governo estadual e que deveriam se valorizar grandemente com a conclusão de extensão do ramal ferroviário entre Ourinhos e Cambará e, numa etapa seguinte, seu prolongamento além dos rios Cinzas, Laranjinhas, Tibagi e Ivaí.

Após reunir-se em Londres com o gerente da *Sudan Plantations*, Arthur Hugh Miller Thomas, que lhe comunica sobre a decisão da empresa em investir no Brasil com a fundação da *Brazil Plantations Syndicate Ltd.* e, em 1925, sua subsidiária brasileira, a Companhia de Terras Norte do Paraná.

Gradualmente os interesses primários da *Brazil Plantations* até então limitados ao plantio de algodão foram se expandido para o cultivo de café e, por fim, na colonização pela venda de terras com a fundação da *Parana Plantations Ltd.* em Londres.

Figura C.1: Primeira construção em Londrina



Primeira construção em Londrina, em 1929: hotel da CTNP

Fonte: Santos (2013)

O modelo de colonização adotado pela Companhia estipulava a subdivisão da área em pequenos lotes, de área variável, demarcados de tal maneira a incluírem todos eles uma parte de baixada (córrego para suprir água) e uma de espigão (estrada para prover o acesso). Geralmente na baixada o colono construía sua casa, plantava um pomar e uma horta, cercava um pasto e, com isso, garantia sua subsistência. Nos trechos mais elevados e menos propensos a serem atingidos por geadas, estendia-se o cafezal, sua fonte de renda (SANTOS, 2013, p. 104).

O registro da primeira expedição dos funcionários da Companhia de Terras Norte do Paraná rumo à região onde pouco tempo depois seria fundada Londrina, no depoimento do George Craig Smith:

“Partimos na madrugada do dia 20 de agosto de 1929, e à tardinha do mesmo dia chegamos à cidade maleitosa de Jataí, onde dormimos no racho de palmitos construídos por Ian Fraser, escocês, funcionário da Companhia Maxwell. Por ser época de seca, fizemos uma viagem relativamente boa e rápida. Mais tarde,

em épocas de chuvas, a mesma viagem chegou a levar até três dias, visto que a estrada de Cambará até Jataí era das mais primitivas e havia inúmeros atoleiros de barro mal-cheiroso, nos quais os caminhões e outros veículos ficavam atolados até os estribos. Somente à custa de esforços sobre-humanos e com grande sacrifício pessoal conseguíamos vencer todos os obstáculos. A famosa Serra Morena era o pavor dos motoristas, pois o caminho estreito que ladeava a montanha era cheio de pedras soltas e buracos. Se os veículos não subissem de uma só arrancada corriam o perigo de rolar abismo abaixo, como de fato aconteceu com alguns caminhões. Em Jataí tratamos logo de comprar uma tropa de burros de carga e montaria para prosseguir viagem até as terras da Companhia, situadas 22 quilômetros além da margem esquerda do rio Tibagi. Com dificuldade e pagando um preço excessivo conseguimos comprar os burros necessários. Contratamos também os serviços de um índio para nos servir de guia, e gostávamos de vê-lo falar mansamente com os animais, chamando cada um pelo nome. E eles misteriosamente obedeciam. Bem cedo, no dia 21 de agosto de 1929, tratamos de atravessar o majestoso rio Tibagi. Como não havia ponte nem balsa, todos os animais atravessaram a nado, um por um. Enquanto um de nós ia remando numa canoa feita de tronco de árvore, outro segurava o burro pelo cabresto e guiava-o até a outra margem. Com várias travessias perigosas foram transpostos os animais, os mantimentos e todo o pessoal para a margem esquerda do Tibagi, de onde iniciamos a dura caminhada até o local denominado Patrimônio Três Bocas, onde se iniciavam as terras da Companhia de Terras Norte do Paraná. Foi uma viagem feita com grande sacrifício, em marcha lenta, caminhando por um picadão escuro, barrento e cheio de tocos e buracos. Depois de muitas peripécias – tais como a fuga dos burros assustados, que derrubavam as cargas e procuravam ganhar as picadas laterais que se encontravam pelo caminho – chegamos à tarde na divisa das terras da Companhia, onde o engenheiro Alexandre Razgulaeff, orgulhosamente, fincou o primeiro marco de madeira e disse: chegamos. Aqui começam as terras da Companhia de Terras Norte do Paraná. Descarregamos os burros no meio daquela floresta densa e assustadora e amarramos os animais para que não fugissem. Sem perda de tempo o Sr. Alberto Loureiro, homem dinâmico, ordenou que seus camaradas, com foices e machados, fizessem uma pequena clareira e construíssem os dois primeiros ranchos. Isso foi na tarde do dia 21 de agosto de 1929.”(SANTOS, 2013, pp. 57–61).

Concebida originalmente pelos colonizadores para ser uma cidade com uma população projetada de 30.000 habitantes Londrina alcançou esse valor poucos anos depois de sua emancipação do 1934 (dados disponíveis em: [Prefeitura do Município de Londrina](#)).

Ao longo das décadas que se seguiram à sua fundação em 1929, a cidade experimentou diversos ciclos de expansão urbana em regiões e épocas distintas (JANUZZI, 2005) ressaltando a natureza da variação do valor da terra como não apenas espacial mas igualmente temporal; em outras palavras, o espaço urbano experimenta taxas de valorização mais ou menos intensas de modo distinto, tanto sob o aspecto da localização, quanto do tempo a que se refere.

As transformações urbanas ocorridas e que serão apresentadas a seguir foram compiladas do portal mantido pela Universidade Estadual de Londrina ([Atlas Ambiental da Cidade de Londrina](#)).



### Principais períodos:

- 1929-1945: implantação do primeiro plano de organização urbana e a importância da exploração cafeeira no desenvolvimento da cidade;
- 1946-1970: adensamento populacional e início do processo de verticalização no centro da cidade;
- 1971-1997: a expansão territorial com ênfase na região norte com a construção de conjuntos habitacionais;
- 1998-2017: a expansão da região sul-sudoeste.

A elevada produtividade da cafeicultura regional atraiu um considerável contingente populacional em direção à cidade entre 1950 e 1960, resultando numa explosão demográfica. O elevado afluxo populacional que se dirigia à cidade e o seu rápido crescimento, contribuíram para uma impressionante expansão da cidade.

Em julho de 1975 ocorre o fenômeno climático extremo que passou a ser conhecido como “geada negra”, marcando o fim do ciclo do café na história da cidade e causando um intenso êxodo rural em toda a região.

Entre 1978 e 1980 ocorre a implantação de quase 9.000 mil unidades residenciais, grande parte na região norte da cidade onde os primeiros conjuntos habitacionais foram construídos em 1979

Na década de 1980 ocorreram importantes transformações relacionadas ao uso do solo na área urbana de Londrina, especialmente na região central, dando início ao processo de verticalização com as construções de edifícios residenciais no eixo central da cidade.

Após 2000 a expansão vertical que havia se estendido do centro para os bairros periféricos atinge a região sul onde, com a legislação estabelecida no anos de 1990 em relação ao fechamento de áreas residenciais, tendência predominante foi a construção de condomínios horizontais fechados de alto padrão, notadamente com a valorização introduzida na região sul-sudoeste pela construção de um grande *Shopping Center*

## C.2 GEORREFERENCIAMENTO

Uma vantagem da adoção do sistema UTM para o georreferenciamento dos dados espaciais é a possibilidade de perfazer operações matemáticas com suas coordenadas de modo mais simples. Em contrapartida, uma de suas desvantagens é requerer transformações apropriadas nas situações onde os dados se distribuam por mais de um fuso.

O Grupo Europeu de Pesquisa de Pesquisas Petrolíferas (*EPSG: European Petroleum Survey Group*, 1986 - 2005) foi uma organização científica ligada à indústria petrolífera europeia composta por especialistas que trabalhavam em geodésia aplicada, levantamento topográfico e cartografia, relacionados com a exploração de petróleo.

A codificação de sistemas de georeferenciamento “EPSG” é um registro público de *datums* geodésicos, sistemas de referência espacial, elipsóides terrestres, transformações de coordenadas e unidades de medida relacionadas. Cada registro recebe um código “EPSG” entre 1024-32767, junto com uma representação de texto conhecida e legível por máquina padrão.

Neste trabalho, os dados encontram-se referenciados no sistema de referência geodésico regional para a América do Sul (SAD69: *South American Datum*, 1969) o *datum* geodésico regional para a América do Sul oficial no Brasil até fevereiro de 2005, quando o Sistema de Referência Geocêntrico para as Américas (SIRGAS2000) foi oficializado como novo referencial geodésico para o Sistema Geodésico Brasileiro pela publicação da resolução 01/2005 do Fundação Instituto de Geografia e Estatística (IBGE).

Os códigos “EPSG” relativos a esses dois *datums* são:

- EPSG 4674: SIRGAS2000 (Sistema de Referência Geocêntrico para as Américas);
- EPSG 29192: SAD69 (*South American Datum*).

### C.3 RENDA MÉDIA DOS SETORES CENSITÁRIOS

As planilhas com os dados agregados de setores referentes ao censo de 2000 para o estado do Paraná estão disponíveis no portal [IBGE: dados dos setores e documentação sobre os dados \(censo 2000\)](#); e as referentes aos dados do censo de 2010 no portal [IBGE: dados dos setores \(censo 2010\)](#) e [IBGE: documentação sobre os dados \(censo 2010\)](#)

O arquivo compactado “Agregado\_de\_setores\_2000\_PR.zip” contém o documento: “Censo Demográfico 2000: Agregado por Setores Censitários dos Resultados do Universo 2ª edição” que em sua página 28 define a variável “Var02: Total do rendimento nominal mensal das pessoas responsáveis por domicílios particulares permanentes” da planilha “Basico\_PR.xls” e, na página seguinte, define a variável “Var003: Domicílios particulares permanentes” na planilha “Domicilio\_PR.xls”. A renda média domiciliar é o resultado da divisão de “Var02” por “Var003”.

O arquivo compactado “Documentacao\_Agregado\_dos\_Setores\_2010416.zip” contém o documento “Base de informações do Censo Demográfico 2010: resultados do Universo por setor censitário” que em sua página 45 define a variável “V001: Domicílios particulares permanentes ou pessoas responsáveis por domicílios particulares permanentes” da planilha “Basico\_PR.xls” e na página 153 define a variável “V001: Total do rendimento nominal mensal dos domicílios particulares” da planilha “DomicilioRenda\_PR.xls”. A renda média domiciliar é o resultado da divisão de “V001” da planilha “DomicilioRenda\_PR.xls” por “V001” da planilha “Basico\_PR.xls”.

O código de identificação de um setor censitário (menor unidade geográfica de um censo) é formado por 15 dígitos compondo a sequência: “UF+MMMMM+DD+SD+SSSS” onde os códigos identificam:

- UF: unidade da federação;
- MMMMM: o município do estado;
- DD: o distrito do município;
- SD: o subdistrito do distrito;
- SSSS: a identificação do setor censitário.

O arquivo digital com a malha dos setores censitários referentes ao senso realizado em 2010 está disponível no portal mantido pelo [IBGE: setores censitários de Londrina \(senso 2000\)](#); e os referentes ao senso de 2010 no portal [IBGE: setores censitários do Paraná \(senso 2010\)](#).

Os arquivos digitais das malhas dos setores censitários incorporam em treze camadas, informações cadastrais e geométricas (espaciais, os polígonos delimitadores dos setores):

- |               |               |               |
|---------------|---------------|---------------|
| • ID;         | • CD_GEOCODS; | • NM_MUNICIP; |
| • CD_GEOCODI; | • NM_SUBDIST; |               |
| • TIPO;       | • CD_GEOCODD; | • NM_MICRO;   |
| • CD_GEOCODB; | • NM_DISTRIT; |               |
| • NM_BAIRRO;  | • CD_GEOCODM; | • NM_MESO.    |

#### C.4 TERMOS E SUA FORMA DE INTRODUÇÃO NOS PREDITORES DOS PARÂMETROS DISTRIBUCIONAIS

Assim, as estimativas dos parâmetros distribucionais serão dadas por seus preditores que assumem as seguintes estruturas gerais aditivas em um modelo maximal:

$$\begin{aligned} \log(\hat{\mu}) = & \beta_{10} + s_{11}(\text{AT}) + s_{12}(\text{DATA}) + s_{13}(\text{UTM\_X}, \text{UTM\_Y}) + s_{14}(\text{UTM\_X}, \text{UTM\_Y}, \text{DATA}) \\ & \beta_{11}(\text{se RELEVO=ACLIVE}) + \beta_{12}(\text{se RELEVO=DECLIVE}) + \\ & \beta_{13}(\text{se NATUREZA=TRANSAÇÃO}) + \\ & \beta_{14}(\text{se PAVIMENTACAO=ASFÁLTICA}) + \\ & \beta_{15}(\text{se IMPLANTACAO=CONDOMÍNIO}) \end{aligned}$$

$$\begin{aligned} \log(\hat{\sigma}) = & \beta_{20} + s_{21}(\text{AT}) + s_{22}(\text{DATA}) + s_{23}(\text{UTM\_X}, \text{UTM\_Y}) + s_{24}(\text{UTM\_X}, \text{UTM\_Y}, \text{DATA}) \\ & \beta_{21}(\text{se RELEVO=ACLIVE}) + \beta_{22}(\text{se RELEVO=DECLIVE}) + \\ & \beta_{23}(\text{se NATUREZA=TRANSAÇÃO}) + \\ & \beta_{24}(\text{se PAVIMENTACAO=ASFÁLTICA}) + \\ & \beta_{25}(\text{se IMPLANTACAO=CONDOMÍNIO}) \end{aligned}$$

$$\begin{aligned}\hat{\nu} = & \beta_{30} + s_{31}(\text{AT}) + s_{32}(\text{DATA}) + s_{33}(\text{UTM\_X}, \text{UTM\_Y}) + s_{34}(\text{UTM\_X}, \text{UTM\_Y}, \text{DATA}) \\ & \beta_{31}(\text{se RELEVO=ACLIVE}) + \beta_{32}(\text{se RELEVO=DECLIVE}) + \\ & \beta_{33}(\text{se NATUREZA=TRANSAÇÃO}) + \\ & \beta_{34}(\text{se PAVIMENTACAO=ASFÁLTICA}) + \\ & \beta_{35}(\text{se IMPLANTACAO=CONDOMÍNIO})\end{aligned}$$

$$\begin{aligned}\log(\hat{\tau}) = & \beta_{40} + s_{41}(\text{AT}) + s_{42}(\text{DATA}) + s_{43}(\text{UTM\_X}, \text{UTM\_Y}) + s_{44}(\text{UTM\_X}, \text{UTM\_Y}, \text{DATA}) \\ & \beta_{41}(\text{se RELEVO=ACLIVE}) + \beta_{42}(\text{se RELEVO=DECLIVE}) + \\ & \beta_{43}(\text{se NATUREZA=TRANSAÇÃO}) + \\ & \beta_{44}(\text{se PAVIMENTACAO=ASFÁLTICA}) + \\ & \beta_{45}(\text{se IMPLANTACAO=CONDOMÍNIO})\end{aligned}$$

em que as funções suavizadoras podem ser codificadas como :

- $s_{i1}(\text{AT})$  e (o efeito principal de)  $s_{i2}(\text{DATA})$ :
  - $\text{gamlss}(\dots + \text{ga}(\sim \text{ti}([\text{AT}/\text{DATA}], \text{bs} = \text{"tp"}, \text{np} = \text{FALSE}, \dots), \dots))$ , com igual resultado que:
 
$$\text{gamlss}(\dots + \text{ga}(\sim s(\text{AT}/\text{DATA}], \text{bs} = \text{"tp"}, \dots), \dots);$$
  - $\text{gamlss}(\dots + \text{ga}(\sim \text{ti}(\text{AT}/\text{DATA}], \text{bs} = \text{"cr"}, \dots), \dots))$ , com igual resultado que:
 
$$\text{gamlss}(\dots + \text{ga}(\sim s(\text{AT}/\text{DATA}], \text{bs} = \text{"cr"}, \dots), \dots);$$
  - $\text{gamlss}(\dots + \text{pb}([\text{AT}/\text{DATA}], \dots))$ ;
  - $\text{gamlss}(\dots + \text{cs}([\text{AT}/\text{DATA}], \dots))$ ; ou,
  - $\text{gamlss}(\dots + \text{bs}([\text{AT}/\text{DATA}], \dots))$ .
- $s_{i3}(\text{UTM\_X}, \text{UTM\_Y})$ :
 
$$\text{gamlss}(\dots + \text{ga}(\sim \text{ti}(\text{UTM\_X}, \text{UTM\_Y}), \text{bs} = \text{"tp"}, \dots), \dots);$$
- $s_{i4}(\text{UTM\_X}, \text{UTM\_Y}, \text{DATA})$ :
 
$$\text{gamlss}(\dots + \text{ga}(\sim \text{ti}(\text{UTM\_X}, \text{UTM\_Y}, \text{DATA}), \text{bs} = c(\text{"tp"}, \text{"cr"}), d = c(2, 1), \dots), \dots).$$

## C.5 FUNÇÕES AUXILIARES DE CONTROLE DO ALGORITMO DE ESTIMAÇÃO

Para os vários modelos analisados, houve a necessidade de se alterar alguns dos parâmetros internos de duas funções de controle auxiliar do algoritmo de estimação chamadas internamente pela função *gamlss()*:

- *gamlss.control()*: controla o ciclo externo;
- *glim.control()*: controla o ciclo interno,

detalhadas respectivamente em [gamlss.control\(\)](#) e [glim.control\(\)](#). Em tempo, considerando o custo computacional de se estimar o modelo proposto repetidas vezes, as funções auxiliares de controle do algoritmo de estimação (*glim.control* e *gamlss.control*) foram deixadas em suas configurações *default* nas simulações para a estimação de intervalos de confiança *bootstrap* e na validação cruzada *leave one out* para análise das alterações da medida do AIC. Por esta razão, os valores nelas estabelecidos para os critérios de convergência (0,001) nem sempre se verificaram ao cabo do número de iterações estabelecidas por *default* (20 para o ciclo externo, 50 para o ciclo interno e 30 para o retroajuste) e assim as estimativas dos parâmetros são menos precisas.

## C.6 CRIAÇÃO DE UM *grid* PARA GERAÇÃO DE IMAGENS

Pode-se melhor entender tal procedimento como uma sucessiva estimação de um imóvel que possui as características descritas na situação paradigma e na data arbitrada, em cada localização definida no *grid* que, no estudo foi definido com um espaçamento entre pontos de 50m em cada direção cartesiana.

A criação de um *grid* é um procedimento bastante comum na área de geoestatística quando da geração de superfícies de valor no qual um denso conjunto de pontos a serem usados na estimação de vários locais é gerado compondo uma forma inicialmente retangular tendo como limites a área do estudo e posteriormente tal forma é recortada para que apenas aqueles situados dentro dos limites geográficos da área permaneçam ao final).

Os arquivos digitais com informações espaciais relacionados à cidade de Londrina encontram-se disponibilizados no portal SIGLON Sistema de Informação Geográfica, mantido pela Prefeitura e a leitura do arquivo em formato *shp* dá-se ela função *readOGR()* disponível na biblioteca [rgdal](#):

- [SIGLON: perímetro urbano de Londrina](#), contendo informações relacionadas aos bairros do município(polígonos delimitadores, denominação, áreas e região a que pertencem:
- [SIGLON: bairros de Londrina](#), contendo a poligonal e a área.

O *shapefile* do perímetro urbano de Londrina possui internamente as coordenadas limítrofes (*bounding box*) que podem ser usadas para a delimitação do *grid* a ser criado pela função *expand.grid()* da biblioteca [base](#) e projetado para o mesmo sistema de referência (*CRS Coordinate Reference System*) do *shapefile* do perímetro urbano com a função *proj4string()* da biblioteca [sp](#).

Na sequência, para que a superfície a ser estimada apresente a mesma conformação espacial que aquela delimitada pelo perímetro urbano municipal, projeta-se ou sobrepoem-se a poligonal que delimita o perímetro urbano (informada no *shapefile*) ao *grid* de pontos inicialmente gerado. Tal procedimento pode ser realizado por várias maneiras a depender das classes a que pertencem os objetos e da informação que se deseja extrair.

Para a situação em estudo os objetos pertencem às classes *SpatialPointsDataFrame* e *SpatialPolygonsDataFrame*, respectivamente e um procedimento simples de *overlay* como o descrito em (PEBESMA, 2021, p. 3), que retorna um objeto de mesma classe que a original cujos pontos localizam-se exclusivamente dentro dos limites do perímetro urbano municipal, cujas coordenadas serão utilizadas na geração dos múltiplos locais, cada um replicando a situação paradigma antes estabelecida, a serem estimados pelo modelo proposto.

## REFERÊNCIAS

- ABNT: Associação Brasileira de Normas Técnicas. *NBR 14.653-2:2011*: Avaliação de bens parte 2: imóveis urbanos. Rio de Janeiro, RJ, 2011. 54 p.
- ABNT: Associação Brasileira de Normas Técnicas. *NBR 14.653-1:2019*: Avaliação de bens parte 1: procedimentos gerais. Rio de Janeiro, RJ, 2019. 19 p.
- AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v. 19, n. 6, p. 716–723, 1977. Disponível em: [<recurso>](#). Acesso em: abr. 2021.
- ATKINSON, A. B. Two graphical displays for outlying and influential observations in regression. *Biometrika*, v. 38, p. 13–20, 1981. Disponível em: [<recurso>](#). Acesso em: mai. 2021.
- BLOM, G. *Statistical Estimates and Transformed Beta-Variables*. New York, USA: John Wiley & Sons, 1958. Disponível em: [<recurso>](#). Acesso em: mai. 2021.
- BOLLEN, K. A.; JACKMAN, R. W. Regression diagnostics: an exploratory treatment of outliers and influential cases. *Sociological Methods & Research*, v. 13, n. 4, p. 510–542, 1985. Disponível em: [<recurso>](#). Acesso em: abr. 2021.
- BOX, G. E. P.; COX, D. R. An analysis of transformations. *Journal of the Royal Statistical Society*, v. 26, n. 2, p. 211–252, 1964. Disponível em: [<recurso>](#). Acesso em: abr. 2021.
- BREUSCH, T. S.; PAGAN, A. A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, v. 47, n. 5, p. 1287–1294, 1979. Disponível em: [<recurso>](#). Acesso em: abr. 2021.
- BURDEN, R. L.; FAIRES, D. J.; BURDEN, A. M. *Análise Numérica*. 3. ed. São Paulo, SP: Cengage, 2017. 879 p.
- CLEVELAND, W. S. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, v. 74, p. 829–836, 1979. Disponível em: [<recurso>](#). Acesso em: mai. 2021.
- COLE, T. J.; GREEN, P. J. Smoothing reference centile curves: The ims method and penalized likelihood. *Statistics in Medicine*, v. 11, p. 1305–1319, 1992. Disponível em: [<recurso>](#). Acesso em: jun. 2021.
- COOK, R. D.; WEISBERG, S. *Residuals and Influence in Regression*. London, EN: Chapman and Hall, 1982. 240 p. Disponível em: [<recurso>](#). Acesso em: abr. 2021.
- CORDEIRO, G. M.; DEMÉTRIO, C. G. B.; MORAL, R. de A. *Modelos Lineares Generalizados e Extensões*. Piracicaba, SP: Escola Superior de Agricultura Luiz de Queiroz, 2015.
- DUNN, P. K.; SMITH, G. K. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, v. 5, p. 236–244, 1996. Disponível em: [<recurso>](#). Acesso em: abr. 2021.

EILERS, P. H. C.; MARX, B. D. Flexible smoothing with b-splines and penalties. *Statistical Science*, v. 11, n. 2, p. 89–121, 1996. Disponível em: <recurso>. Acesso em: mai. 2021.

EILERS, P. H. C.; MARX, B. D.; REGUERA, M. L. D. Twenty years of p-splines. *SORT-Statistics and Operations Research Transactions*, Catalunha, SP, v. 39, n. 2, p. 149–186, 2015. Disponível em: <recurso>. Acesso em: jun. 2021.

FIGUEIREDO FILHO, D. et al. O que fazer e o que não fazer com a regressão: pressupostos e aplicações do modelo linear de mínimos quadrados ordinários (mqo). *Revista Política Hoje*, v. 20, n. 1, p. 44–99, 2011. Disponível em: <recurso>. Acesso em: jun. 2014.

FISHER, R. A. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical*, v. 222, p. 594–604, 1922. Disponível em: <recurso>. Acesso em: maio 2021.

FISHER, R. A. *The Design of Experiments*. London, UK: Oliver and Boyd, 1935. Disponível em: <recurso>. Acesso em: abr. 2021.

GLEJSER, H. A new test for heteroskedasticity. *Journal of the American Statistical Society*, v. 64, n. 325, p. 316–323, 1969. Disponível em: <recurso>. Acesso em: abr. 2021.

GREEN, P. J.; SILVERMAN, B. W. *Nonparametric Regression and Generalized Linear Models a Roughness Penalty Approach*. Boca Raton, FL: Chapman & Hall/CRC Texts in Statistical Science, 1994. 194 p.

GUJARATI, D. N.; PORTER, D. C. *Econometria Básica*. São Paulo, SP: AMGH Editora, 2011. Disponível em: <recurso>. Acesso em: abr. 2021.

HASTIE, T. J.; TIBSHIRANI, R. J. Generalized additive models. *Statistical Science*, v. 1, n. 3, p. 297–318, 1986. Disponível em: <recurso> Acesso em: abr. 2021.

HASTIE, T. J.; TIBSHIRANI, R. J. *Generalized additive models*. London, EN: Chapman & Hall, 1990.

HASTIE, T. J.; TIBSHIRANI, R. J.; FRIEDMAN, J. *The Elements of Statistical Learning: Data mining, inference and prediction*. 2nd. ed. New York, NY: Springer, 2009. Disponível em: <recurso>. Acesso em: abr. 2021.

HOAGLIN, D. C.; WELSCH, R. E. The hat matrix in regression and anova. *The American Statistician*, v. 32, n. 1, p. 17–22, 1978. Disponível em: <recurso>. Acesso em: abr. 2021.

HOFFMANN, R. *Análise de Regressão: Uma introdução à econometria*. São Paulo, SP: Universidade de São Paulo, 2016. Disponível em: <recurso>. Acesso em: abr. 2021.

JANUZZI, D. de C. R. O desenvolvimento de londrina e as transformações nos espaços públicos da região central. *Semina: Ciências Humanas e Sociais*, v. 26, p. 87–94, 2005. Disponível em: <recurso>. Acesso em: mai. 2021.

JØRGENSEN, B. Exponential dispersion models. *Journal of the Royal Statistics Society: Series B (Methodological)*, v. 49, p. 127–145, 1987. Disponível em: <recurso>. Acesso em: abr. 2021.

KENDALL, M. G.; STUART, A. *The Advanced Theory of Statistics*. 3rd. ed. London , EN: Butler & Tanner, 1961. v. 2. 684 p. Disponível em: <recurso>. Acesso em: abr. 2021.



- MARTINS, A. L. *História do café*. São Paulo, SP: Editora Contexto, 2008. 316 p.
- MYERS, R. H. et al. *Generalized Linear Models with Applications in Engineering and the Sciences*. 2nd. ed. Hoboken, NJ: Wiley, 2010. 521 p. Disponível em: <recurso>. Acesso em: abr. 2021.
- NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society*, v. 135, n. 3, p. 370–384, 1972. Disponível em: <recurso>. Acesso em: abr. 2021.
- NELDER, J. A.; WEDDERBURN, R. W. M. *Generalized Linear Models*. 2nd. ed. Boca Raton, FL: Chapman & Hall/CRC Texts in Statistical Science, 1989. 532 p.
- NETTER, J.; WASSERMAN, W.; KUTNER, M. H. *Applied Linear Statistical Models*. 1<sup>st</sup>. ed. Homewood, Il: Richard D. Irwin, 1983. Disponível em: <recurso>. Acesso em: abr. 2021.
- O’SULLIVAN, F. A statistical perspective on ill-posed inversed problems. *Statistical Science*, v. 1, p. 502–527, 1986. Disponível em: <recurso>. Acesso em: mai. 2021.
- PARK, R. E. Estimation with heterocedastic error terms. *Econometrica*, v. 34, n. 4, p. 888, 1966. Disponível em: <recurso>. Acesso em: abr. 2021.
- PEBESMA, E. *Map overlay and spatial aggregation in sp*. [S.l.], 2021. Disponível em: <recurso>. Acesso em: set. 2021.
- PRICE, M. J. *Penalized b-splines and their application with an in depth look at the bivariate tensor product penalized b-spline*. Tese (Doctor of Philosophy) — Iowa State University, Ames, IA, 2018. Disponível em: <recurso>. Acesso em: abr. 2021.
- REGUERA, M. L. D. *Modelos Aditivos Generalizados con Psplines*. Madrid, SP: Universidade Carlos III, 2021. Disponível em: <recurso>. Acesso em: mai. 2021.
- RIGBY, R. A.; STASINOPOULOS, D. M. Generalized additive models for location, scale and shape. *Journal of the Royal Statistics Society*, v. 54, n. 3, p. 507–554, 2005. Disponível em: <recurso>. Acesso em: jun. 2021.
- RIPLEY, B. D. Modelling spatial patterns. *Journal of the Royal Statistical Society*, v. 39, n. 2, p. 172–212, 1977. Disponível em: <recurso>. Acesso em: abr. 2021.
- SANTOS, R. R. dos (org.). *Colonização e Desenvolvimento do Norte do Paraná: Depoimentos sobre a maior obra do gênero realizada por uma empresa privada (24 de setembro de 1975)*. 3. ed. São Paulo, SP: Editora Melhoramentos, 2013. 256 p. Disponível em: <recurso>. Acesso em: abr. 2021.
- SCHWARZ, G. Estimating the dimension of a model. *The Annals of Statistics*, v. 6, p. 461–464, 1978. Disponível em: <recurso>. Acesso em: abr. 2021.
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, v. 52, n. 3/4, p. 591–611, 1965. Disponível em: <recurso>. Acesso em: abr. 2021.
- STASINOPOULOS, D. M.; RIGBY, R. A. Generalized additive models for locations scale and shape (gamlss) in r. *Journal of Statistics Software*, v. 23, n. 7, p. 46, 2007. Disponível em: <recurso>. Acesso em: jun. 2021.

STASINOPOULOS, D. M. et al. *Flexible Regression and Smoothing Using GAMLSS in R*. Boca Raton, FL: Chapman & Hall/CRC Texts in Statistical Science, 2017. 549 p.

STIGLER, S. M. Fisher in 1921. *Statistical Science*, v. 20, n. 1, p. 32–49, 2005. Disponível em: [<recurso>](#). Acesso em: maio 2021.

STIGLER, S. M. The epic story of maximum likelihood. *Statistical Science*, v. 22, n. 4, p. 598–620, 2007. Disponível em: [<recurso>](#). Acesso em: maio 2021.

THOMAS, G. *GAMLSS with applications to zero inflated and hierarchical data*. Dissertação (Mestrado em Estatística e Experimentação Agronômica) — Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, SP, 2017. Disponível em: [<recurso>](#). Acesso em: jul. 2021.

TOBLER, W. R. On the first law of geography: A reply. *Annals of the Association of American Geographers*, v. 94, n. 2, p. 304–310, 2004. Disponível em: [<recurso>](#). Acesso em: abr. 2021.

TURKMAN, M. A. da C. A.; SILVA, G. L. Modelos lineares generalizados - da teoria à prática. In: *VIII Congresso Anual da Sociedade Portuguesa de Estatística*. Peniche, PT: SPE - Sociedade Portuguesa de Estatística, 2000. Disponível em: [<recurso>](#). Acesso em: abr. 2021.

VIEIRA, A. F. de C. *Análise da Média e Dispersão em Experimentos Fatoriais não Replicados para Otimização de Processos Industriais*. Tese (Doutorado em Engenharia de Produção) — Pontifícia Universidade Católica, Rio de Janeiro, RJ, 2004. Disponível em: [<recurso>](#). Acesso em: abr. 2021.

WATSON, G. S. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics*, v. 26, n. 4, p. 359–372, 1964. Disponível em: [<recurso>](#). Acesso em: jun. 2021.

WEISBERG, S. *Applied Linear Regression*. Minneapolis, MN: John Wiley & Sons, 2014. Disponível em: [<recurso>](#). Acesso em: mai. 2021.

WHITE, H. A heteroskedasticity consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, v. 48, n. 4, p. 817–838, 1980. Disponível em: [<recurso>](#). Acesso em: abr. 2021.

WOOD, S. N. P-splines with derivative based penalties and tensor product smoothing of unevenly distributed data. *Statistics and Computing*, v. 27, p. 985–989, 2016. Disponível em: [<recurso>](#). Acesso em: abr. 2021.

WOOD, S. N. *Generalized Additive Models: an introduction with R*. Second edition. Boca Raton, FL: Chapman & Hall/CRC Texts in Statistical Science, 2017. 476 p.