

**Modelos de Regressão Segmentada:
Construção, Validação e
Métodos Restritos**

Regina Shimizu Tamai Ishimoto

Dissertação apresentada ao
Instituto de Matemática e Estatística
da Universidade de São Paulo
para obtenção do grau de
Mestre em Estatística

Área de concentração: **Estatística**
Orientador: **Prof. Dr. Gilberto A. Paula**

São Paulo

2004

Modelos de Regressão Segmentada: Construção, Validação e Métodos Restritos

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Regina Shimizu Tamai Ishimoto e aprovada pela comissão julgadora.

Aprovado em: 15 de outubro de 2004.

Comissão julgadora:

- Prof. Dr. Gilberto Alvarenga Paula (Orientador)- IME/USP
- Profa. Dra. Julia Maria Pavan Soler - IME/USP
- Prof. Dr. Mário de Castro Andrade Filho - ICMC/USP

*À minha mãe Tiekô e à batiam Yukiko:
vocês partiram mas o amor permanece, é eterno.*

*Aos queridos: Letícia, Rafael Kenzo e Marcelo.
Amo vocês.*

Agradecimentos

Obrigada Santíssima Trindade. Pelas pessoas e relacionamentos, por cada descoberta, cada aprendizado, cada experiência vivida ...

A todos que diretamente ou não participaram da realização desse sonho. Em especial: Guenji Tamai, meu pai, que foi motorista, massagista ... um pouquinho de tudo. Este reencontro foi gratificante;

Minha irmã Tuca e Cristina Paulino, que acompanharam todos os bastidores;

Maria Conceição Selke, simplesmente por ser quem é;

Yoshiko Nishimura Ishimoto, minha sogra, que sempre apostou em mim;

Maria José, Sirléia, Neide, Sebastiana, Juraci que dividiram comigo as tarefas domésticas e os cuidados com meus filhos;

Mazinha e Pe. Moacir Ventura, verdadeiros mestres, que a cada dia abrem uma janela para uma vida mais significativa;

Rita Rodorigo, uma amiga e companheira de trabalho, que gentilmente acompanhou as fases finais dessa dissertação;

Alunos da Universidade Ibirapuera, que a cada ano trazem novos desafios educacionais e motivação para prosseguirmos nesta trajetória;

Dra. Cintia Pimentel Sayd, que na hora certa veio ao meu auxílio;

Dr. Minoru Koda, confiança e presença amiga em momentos difíceis;

Colegas, professores e funcionários do IME que tornaram o dia-a-dia mais gostoso;

Iracema Arashiro que por um tempo foi meu único elo com o IME. Seu incentivo, sua

determinação, as trocas de experiência foram decisivas. Uma grande pessoa;

Audrey e Francisco Cysneiros, por darem "aquela" força e pelo exemplo de dedicação, trabalho e honestidade;

Prof. Gilberto pela paciência nesses anos e por partilhar seus conhecimentos e experiências com verdadeiro desprendimento;

Rafael Kenzo, a certeza num momento incerto: "Você vai conseguir, mãe!"

Marcelo Ishimoto, que vivenciou comigo todas as fases do mestrado e, ao seu modo, colaborou de muitas formas para que eu concretizasse esse projeto.

O meu muito obrigada a todos vocês.

Resumo

O objetivo principal deste trabalho é desenvolver a construção de modelos de regressão polinomiais segmentados e apresentar alguns resultados para diagnósticos nesses modelos.

Os modelos polinomiais segmentados compõem-se de uma curva dividida em partes onde os segmentos adjacentes (“pedaços de curvas”) que a formam podem ter o mesmo grau ou não. Esses modelos são úteis quando a variável resposta não apresenta uma única tendência ao longo dos dados.

Em algumas situações, pode-se impor restrições a esses modelos para obter o grau desejado para uma determinada parte. Logo, discutimos tanto o caso irrestrito quanto o restrito desses modelos. Para avaliá-los, algumas medidas de seleção e diagnóstico são apresentadas e desenvolvidas, assim como são adaptados programas gráficos de diagnóstico e geração de envelopes em S-Plus.

Como a segmentação ocorre na parametrização da expressão $\mathbf{X}\beta$, trabalhamos também o preditor linear de um modelo linear generalizado (MLG) contemplando assim não somente os modelos normais lineares e sim, uma classe maior.

Finalmente, alguns exemplos ilustram as técnicas relatadas ao longo do texto.

Abstract

The aim of this work is to discuss some methodologies to construct segmented regression models and to present some diagnostic procedures.

Segmented models are formed by curves which are divided in adjacent segments (“pieces of the curve”) that may have (or not) the same degree. These models are useful when the response variable does not present the same tendency by along the data.

In some situations, one may impose constraints on the parameters in order to obtain an appropriate degree. Then, we discuss the unrestricted and restricted cases and to compare the fitted models under these two conditions, some diagnostic and model selection procedures as well as adapted programs in S-Plus are presented.

Since the segmentation is made in the linear predictor $\mathbf{X}\boldsymbol{\beta}$, the presented methodologies can be extended for more general models, such as generalized linear models.

Finally, in order to illustrate the methods presented in the work, some applications are given.

Índice

| | | |
|----------|---|-----------|
| 1 | Introdução | 1 |
| 2 | Modelos de Regressão Normal Polinomiais Segmentados | 5 |
| 2.1 | Definição de polinômio segmentado | 5 |
| 2.1.1 | Posição e quantidade dos pontos de mudança | 8 |
| 2.1.2 | A função “+” | 8 |
| 2.1.3 | <i>Spline</i> polinomial mista | 12 |
| 2.2 | Estimação | 19 |
| 2.2.1 | Caso 1: estimação irrestrita | 20 |
| 2.2.2 | Caso 2: estimação restrita | 23 |
| 2.3 | Seleção do modelo | 28 |
| 2.3.1 | Estatística PRESS | 28 |
| 2.4 | Técnicas de diagnóstico | 34 |
| 2.4.1 | Diagnóstico em modelos irrestritos | 34 |
| 2.4.2 | Diagnóstico em modelos restritos | 39 |
| 2.4.3 | Extensões | 43 |
| 3 | Modelos Lineares Generalizados Polinomiais Segmentados | 45 |
| 3.1 | Modelos lineares generalizados | 45 |
| 3.2 | Estimação | 48 |

| | | |
|----------|---|------------|
| 3.2.1 | Estimação irrestrita por máxima verossimilhança | 48 |
| 3.2.2 | Estimação restrita | 52 |
| 3.2.3 | Estimação restrita por máxima verossimilhança | 52 |
| 3.3 | Avaliação e seleção de modelos | 55 |
| 3.3.1 | Avaliação e seleção de MLGs irrestritos | 55 |
| 3.3.2 | Avaliação e seleção de MLGs restritos | 57 |
| 3.4 | Técnicas de diagnóstico | 58 |
| 3.4.1 | Diagnóstico em MLGs irrestritos | 58 |
| 3.4.2 | Diagnóstico em MLGs restritos | 63 |
| 3.4.3 | Envelopes | 66 |
| 3.4.4 | Resumo das medidas de diagnóstico para MLG's | 66 |
| 3.4.5 | Extensões | 67 |
| 4 | Aplicações | 68 |
| 4.1 | Crescimento de crianças | 68 |
| 4.2 | Diabetes <i>mellitus</i> | 87 |
| 5 | Considerações Finais | 98 |
| 5.1 | Conclusões | 98 |
| 5.2 | Estudos futuros | 99 |
| A | Resultados utilizados | 100 |
| A.1 | Teorema de Sherman-Morrison-Woodbury | 100 |
| A.2 | Resultado 1 | 101 |
| A.3 | Resultado 2 | 101 |
| B | Dados dos Exemplos do Capítulo 4 | 102 |
| B.1 | Crescimento de crianças | 103 |
| B.2 | Diabetes <i>mellitus</i> | 105 |

| | |
|--|-----|
| C Programa para diagnóstico em modelos normais restritos | 106 |
| D Programa para diagnóstico em modelos gama restritos | 109 |
| E Programas de Envelopes | 112 |
| Referências Bibliográficas | 117 |

Capítulo 1

Introdução

Modelos estatísticos são representações matemáticas de fenômenos reais formuladas a partir de dados e/ou informações obtidas do próprio fenômeno de interesse. Esses dados podem referir-se, por exemplo, a várias medições do fenômeno sob controle, informações clínicas de pacientes com o mesmo tipo de doença durante vários anos, estudos econômicos, etc.

Uma maneira de descrevermos a relação entre variáveis é a partir de dados envolvidos no problema, buscando a melhor modelagem para os mesmos. São vários os caminhos possíveis, várias técnicas estatísticas disponíveis e a tomada de decisão deve levar em conta a simplicidade do modelo bem como a percepção e o conhecimento do pesquisador envolvido no estudo.

Ao investigarmos a relação entre duas variáveis através de um diagrama de dispersão, não raramente surgem formas curvilíneas e alguns dos procedimentos utilizados são a transformação das variáveis a fim de se obter uma relação linear ou o ajuste de uma regressão polinomial.

Uma regressão polinomial permite-nos facilmente ajustar um modelo de grau r arbitrário, $r \geq 2$, através de técnicas de regressão linear múltipla. Porém, com a necessidade

de aumento do grau do polinômio, podem surgir problemas de mal condicionamento da matriz $\mathbf{X}^T\mathbf{X}$ devido à alta correlação entre os regressores. Além disso, podem ocorrer casos em que não obtemos uma melhora significativa no ajuste (perdem-se graus de liberdade sem uma diminuição compensadora na soma de quadrados dos resíduos), ou ainda, haver indícios como coeficientes estimados instáveis ou gráficos de resíduos com alguma estrutura sistemática, sugerindo a possibilidade da função comportar-se diferentemente em determinadas partes do campo de variação dos valores da variável X . Segundo Montgomery e Peck (1992), neste caso, a abordagem usualmente adotada é dividir a amplitude dos valores de X em segmentos e ajustar uma curva apropriada a cada um e o uso de regressão segmentada (*splines* polinomiais), seria uma alternativa interessante. Este tipo de modelagem permite-nos utilizar vários segmentos de reta (caso linear), ou várias funções de grau r (geralmente $r = 1, 2$ ou 3), uma para cada subintervalo dos dados. Uma mistura dessas funções também é possível, por exemplo, uma função linear seguida de uma quadrática ou vice-versa.

Para ilustrar, apresentaremos o estudo do crescimento de crianças do nascimento até 71,5 meses extraído de Eubank (1988). Para tanto, consideramos uma amostra de 72 crianças das mais variadas idades dentro do intervalo pré-estabelecido, em que eram avaliadas a razão entre o peso e a altura de cada uma delas. Visualizados num gráfico de dispersão (vide Figura 1.1) observamos que os dados mudam de tendência a partir de uma determinada idade sendo que até o 16^o mês, há indicações de crescimento bastante acentuado da variável resposta e então, a partir deste ponto, a resposta tenderia a ter um comportamento próximo ao linear.

Pergunta-se de que forma poderíamos ajustar um modelo para esta situação?

Modelos segmentados são opções interessantes para estes casos, em que as equações polinomiais de diferentes graus ajustam-se a esses dados seqüencialmente. Para o nosso exemplo poderíamos ter um polinômio cúbico para a primeira parte dos dados, seguido de uma reta a partir do 16^o mês. Voltaremos a esse exemplo mais adiante.

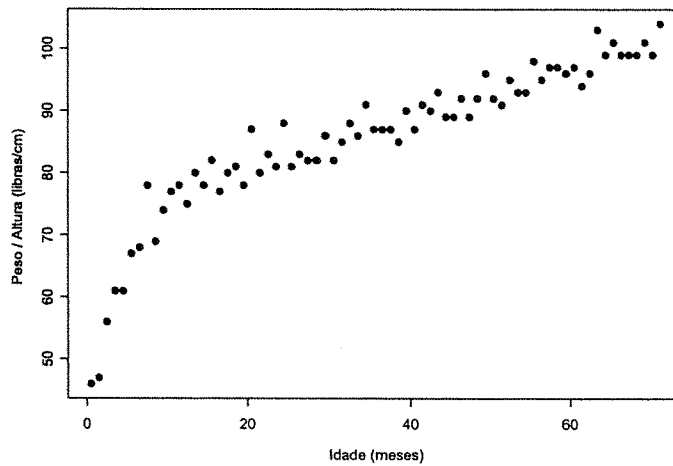


Figura 1.1: Diagrama de dispersão do crescimento de crianças (em cm) segundo a idade

Este trabalho considerará modelos segmentados, conhecidos muitas vezes como *splines*, e representam uma modelagem alternativa para situações em que temos comportamentos diferenciados em intervalos adjacentes do conjunto de dados. No capítulo 2 desenvolveremos primeiramente a construção de polinômios segmentados com o uso da função $+$. Em seguida, apresentaremos a função *spline* polinomial como um modelo linear padrão. Modelos em que há uma variação no grau das equações que compõem o polinômio segmentado serão apresentados numa outra seção, onde o uso de restrições lineares servirão como opção alternativa para obter os graus desejados que compõem a função. A estimação dos parâmetros dos modelos citados vem na seqüência. Para selecionar e avaliar o modelo e as suposições assumidas, temos nas seções 2.3 e 2.4 algumas estatísticas e técnicas de diagnóstico para modelos lineares padrão irrestrito e também para modelos lineares restritos na forma de igualdades lineares, do tipo $C\beta = d$.

No capítulo 3, abordaremos o caso em que a resposta não é mais linear nos parâmetros,

por exemplo taxas, contagens etc, mas cuja média da distribuição associada à resposta pode ser modelada e escrita como uma função linear nos parâmetros numa forma segmentada, constituindo-se um modelo linear generalizado (MLG). Serão apresentados os casos irrestrito e restrito. Métodos de seleção e técnicas de diagnóstico para esses modelos são abordados na sequência.

O 4^o capítulo compõe-se de aplicações das várias técnicas discutidas nos capítulos anteriores. Na primeira aplicação, retomamos o estudo do crescimento de crianças em que se ajustou uma *spline* cúbica-linear com distribuição normal para a variável resposta. Uma análise de diagnóstico é realizada quando então propõe-se um MLG para o ajuste de tais dados. Numa segunda aplicação temos um exemplo de modelagem quadrática-linear para um modelo gama ao propormos uma explicação para o nível de C-peptideo em crianças portadoras de diabetes melito tipo I.

Conclusões e algumas propostas de pesquisa serão apontadas no capítulo final. No apêndice apresentamos os dados e programas para diagnóstico e geração de envelopes utilizados nas aplicações.

Uma observação a ser feita é quanto ao uso do termo “spline”. Em nosso trabalho, o enfoque está em *splines* polinomiais, o que implica em utilizarmos modelos paramétricos. O termo “spline” também é utilizado em Modelos Aditivos Generalizados (Hastie e Tibishirani, 1990), indicando um possível suavizador utilizado no ajuste desses modelos. Nesse caso, o modelo não é mais paramétrico.

Capítulo 2

Modelos de Regressão Normal Polinomiais Segmentados

Neste capítulo define-se *spline* polinomial segmentada inserida no contexto de modelos lineares. Algumas técnicas de diagnóstico são também abordadas, incluindo-se os casos em que há restrições lineares nos parâmetros.

2.1 Definição de polinômio segmentado

Uma junção de vários polinômios formando uma curva (ou reta) quebrada é o que chamamos de *spline* ou polinômio segmentado. Formalmente, temos que *splines* são geralmente definidas como um conjunto de equações polinomiais de ordem r , embora uma composição de polinômios de vários graus também seja possível (ver, por exemplo, Smith, 1979). As abscissas dos pontos onde as equações se unem são denominados *knots*, os pontos de mudança, de tal forma que o polinômio segmentado tenha $k + 1$ peças polinomiais, sendo k o número de pontos de mudança. De uma maneira geral os polinômios

segmentados são expressos na forma

$$\begin{aligned}
P(x) &= \sum_{j=0}^r \beta_{0j} x^j + \sum_{j=0}^r \sum_{i=1}^k \beta_{ij} (x - t_i)_+^j \\
&= \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \dots + \beta_{0r}x^r + \\
&\quad \beta_{10}(x - t_1)_+^0 + \beta_{11}(x - t_1)_+^1 + \beta_{12}(x - t_1)_+^2 + \dots + \beta_{1r}(x - t_1)_+^r + \dots + \\
&\quad \beta_{k0}(x - t_k)_+^0 + \beta_{k1}(x - t_k)_+^1 + \beta_{k2}(x - t_k)_+^2 + \dots + \beta_{kr}(x - t_k)_+^r, \quad (2.1)
\end{aligned}$$

em que β_{ij} , $i = 1, \dots, k$ e $j = 0, \dots, r$ são os coeficientes desconhecidos a serem estimados, x é o valor da variável explicativa, t_i são os pontos de mudança e $(x - t_i)_+$ é a função “+” definida para $j = 0, 1, \dots, r$ como

$$(x - t_i)_+^j = \begin{cases} 0, & \text{se } x - t_i \leq 0 \\ (x - t_i)^j, & \text{se } x - t_i > 0. \end{cases}$$

Os polinômios conhecidos de grau r da forma $a_0 + a_1x + a_2x^2 + \dots + a_r x^r$ são considerados um caso especial de *spline* com pontos de mudança ausentes. Os casos mais extremos de polinômios segmentados incluem de um lado funções contínuas com todas as derivadas contínuas e de outro, funções com salto nos pontos de mudança. Geralmente requer-se que os valores da função e de suas $r - 1$ primeiras derivadas coincidam nos *knots* de tal modo que a *spline* seja uma função contínua com $r - 1$ derivadas contínuas. Nesse caso, deve-se impor continuidade sobre o modelo e em todas as suas derivadas de grau inferior a r . A função de grau 3 com 2 pontos de mudança dada a seguir, é um exemplo para essa situação,

$$\begin{aligned}
P(x) &= \sum_{j=0}^3 \beta_{0j} x^j + \sum_{i=1}^2 \beta_{i3} (x - t_i)_+^3 \\
&= \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \beta_{03}x^3 + \beta_{13}(x - t_1)_+^3 + \beta_{23}(x - t_2)_+^3. \quad (2.2)
\end{aligned}$$

No outro extremo temos funções descontínuas de ordem r com discontinuidades em todas as derivadas até a $(r - 1)$ -ésima ordem, caracterizando um modelo sem restrições

de continuidade. Um exemplo para esse caso de grau 3 com 2 pontos de mudança é dado por

$$\begin{aligned} P(x) &= \sum_{j=0}^3 \beta_{0j} x^j + \sum_{j=0}^3 \sum_{i=1}^2 \beta_{ij} (x - t_i)_+^j \\ &= \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \beta_{03}x^3 + \beta_{10}(x - t_1)_+^0 + \beta_{11}(x - t_1)_+^1 + \beta_{12}(x - t_1)_+^2 \\ &\quad + \beta_{13}(x - t_1)_+^3 + \beta_{20}(x - t_2)_+^0 + \beta_{21}(x - t_2)_+^1 + \beta_{22}(x - t_2)_+^2 + \beta_{23}(x - t_2)_+^3. \end{aligned}$$

Aqui, todos os termos “+” com potência inferior a 3 estão no modelo. Compare com (2.2), em que os termos com coeficientes $\beta_{10}, \beta_{11}, \beta_{12}, \beta_{20}, \beta_{21}$ e β_{22} não estão presentes.

Podemos ter casos intermediários em que os polinômios segmentados sejam submetidos a uma quantidade inferior ao número máximo de restrições de continuidade, caracterizando descontinuidade em suas derivadas até determinada ordem. Neste caso a função “+” terá um papel fundamental e na seção 2.1.2 teremos uma discussão a esse respeito. Como vimos, as equações polinomiais que compõem a *spline* podem variar em número e grau assim como o nível de suavização do ajuste. Quando falamos em imposição de restrições de continuidade, estamos na verdade querendo garantir a suavidade do modelo ao ajustá-lo aos dados. Buse e Lim (1977) discutem os testes para hipóteses sobre continuidade e presença de determinados termos na *spline*.

Na presença de um grande número de *knots*, Wold (1971, 1974) sugere o uso de *B-splines*, que é uma representação especial alternativa para *splines*, com grande eficiência computacional, mas que não garante interpretações estatísticas fáceis.

A grande flexibilidade na composição dos polinômios segmentados deve-se principalmente aos *knots*, que nos permitem obter modelagens para as mais variadas situações. Porém, abusar desta flexibilidade pode levar a um super-ajuste aos dados, acarretando outros problemas. Na verdade, a construção de um modelo segmentado requer muito de tentativa e erro em busca da melhor modelagem. É o que veremos nas seções a seguir.

2.1.1 Posição e quantidade dos pontos de mudança

Geralmente os *knots* são fixados já no início da modelagem pelo experimentador, seguindo justificativas teóricas importantes para o posicionamento dos pontos de mudança. Porém, a quantidade e a posição dos *knots* é livre em que procura-se atender ao critério da soma de quadrados residual mínima, contrabalançado com sugestões teóricas. Wold (1974) sugere que se procure colocar o menor número de pontos de mudança possível, e que cada equação polinomial contenha no mínimo, 4 ou 5 pontos dos dados. Acrescenta ainda, a importância de que cada segmento apresente no máximo um ponto extremo (máximo ou mínimo) e apenas um ponto de inflexão, sendo que os pontos extremos devem estar centrados no segmento, e os pontos de inflexão, próximos aos *knots*.

2.1.2 A função “+”

Como vimos nos polinômios apresentados, uma *spline* polinomial segmentada utiliza polinômios truncados ou funções “+” do tipo

$$u_+ = \begin{cases} 0, & \text{se } u \leq 0, \\ u, & \text{se } u > 0. \end{cases}$$

Essa função imbuete no polinômio mudanças nos coeficientes a partir de determinados pontos, indicando que uma nova equação polinomial, de mesmo grau ou não, estará combinada à equação anterior. O número de mudanças está diretamente associado ao número de *knots* e estes ao número de funções “+” .

O grau de suavização do polinômio e a existência ou não de discontinuidades na função e em suas derivadas estão relacionados e são também diretamente controlados pela função “+” nos diversos graus que compõem o polinômio. Para se ter uma idéia dos possíveis usos desta função, a Figura 2.1 mostra as variações ocorridas na curva ajustada de um modelo polinomial quando variamos o grau de suavização desejado. A idéia é construir o

modelo de forma a garantir a continuidade da função que o define ou a partir de alguma de suas derivadas.

Seja a função polinomial segmentada quadrática em x , $P_1(x)$, com apenas um ponto de mudança em t_1 ,

$$\begin{aligned} P_1(x) &= \sum_{j=0}^2 \beta_{0j} x^j + \sum_{j=0}^2 \beta_{1j} (x - t_1)_+^j \\ &= \beta_{00} + \beta_{01} x + \beta_{02} x^2 + \beta_{10} (x - t_1)_+^0 + \beta_{11} (x - t_1)_+^1 \\ &\quad + \beta_{12} (x - t_1)_+^2. \end{aligned} \tag{2.3}$$

Podemos reescrever (2.3) como,

$$P_1(x) = \begin{cases} \beta_{00} + \beta_{01} x + \beta_{02} x^2, & \text{para } x \leq t_1 \\ \beta_{00} + \beta_{01} x + \beta_{02} x^2 + \beta_{10} (x - t_1)^0 + \beta_{11} (x - t_1)^1 + \beta_{12} (x - t_1)^2, & \text{para } x > t_1, \end{cases}$$

ou equivalentemente,

$$P_1(x) = \begin{cases} \beta_{00} + \beta_{01} x + \beta_{02} x^2, & \text{para } x \leq t_1 \\ \beta_{00} + \beta_{10} - \beta_{11} t_1 + \beta_{12} t_1^2 + (\beta_{01} + \beta_{11} - 2\beta_{12} t_1) x + (\beta_{02} + \beta_{12}) x^2, & \text{para } x > t_1. \end{cases}$$

A derivada primeira dessa função com respeito a x é dada por

$$P_1^{(1)}(x) = \beta_{01} + 2\beta_{02} x + \beta_{11} (x - t_1)_+^0 + 2\beta_{12} (x - t_1)_+^1$$

ou seja,

$$P_1^{(1)}(x) = \begin{cases} \beta_{01} + 2\beta_{02} x, & \text{para } x \leq t_1 \\ \beta_{01} + 2\beta_{02} x + \beta_{11} + 2\beta_{12} (x - t_1), & \text{para } x > t_1, \end{cases}$$

em que o superescrito (1) em $P_1^{(1)}$ indica a primeira derivada de P_1 . A equação (2.3) é uma função segmentada descontínua e uma possível representação gráfica é dada pela curva 1 na Figura 2.1. Nesse caso P_1 e $P_1^{(1)}$ apresentam salto em $x = t_1$.

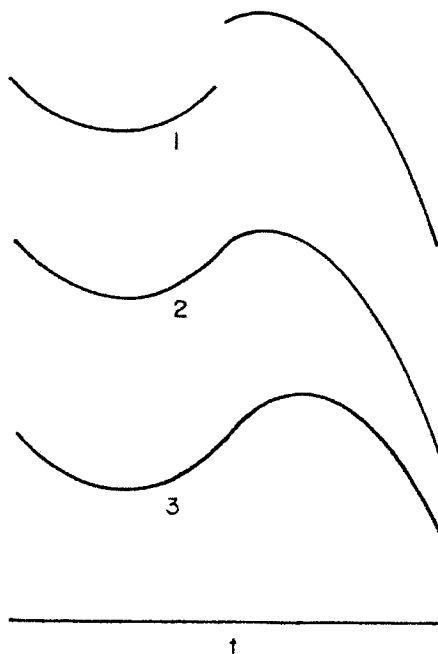


Figura 2.1: Variações na curva ajustada de um modelo polinomial.

Na medida em que necessitamos de modelos mais suaves, a imposição de restrições de continuidade é necessária. A continuidade para P_1 é obtida, eliminando-se o termo β_{10} em (2.3). Assim, a função polinomial contínua é representada por

$$P_2(x) = \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \beta_{11}(x - t_1)_+^1 + \beta_{12}(x - t_1)_+^2,$$

e separando por intervalos, temos que $P_2(x)$ é dado por

$$P_2(x) = \begin{cases} \beta_{00} + \beta_{01}x + \beta_{02}x^2, & \text{para } x \leq t_1 \\ \beta_{00} - \beta_{11}t_1 + \beta_{12}t_1^2 + (\beta_{01} - \beta_{11})x + (\beta_{02} + \beta_{12})x^2, & \text{para } x > t_1. \end{cases}$$

Já a derivada primeira de $P_2(x)$ com respeito a x apresentaria a forma,

$$P_2^{(1)}(x) = \beta_{01} + 2\beta_{02}x + \beta_{11}(x - t_1)_+^0 + 2\beta_{12}(x - t_1)_+^1.$$

Note que nesse caso a primeira derivada de P_2 é dada por uma função descontínua com um salto em $x = t_1$, dado pela parcela com coeficiente β_{11} . A representação gráfica de P_2 está na curva 2 da Figura 2.1.

Poderíamos ainda impor continuidades tanto para P_1 como para sua derivada primeira resultando na seguinte função:

$$P_3(x) = \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \beta_{12}(x - t_1)_+^2,$$

em que a eliminação dos termos com coeficientes β_{10} e β_{11} de (2.3) seriam necessários. A primeira derivada dessa função seria então dada por

$$P_3^{(1)}(x) = \beta_{01} + 2\beta_{02}x + 2\beta_{12}(x - t_1)_+^1,$$

que também é uma função contínua. Veja a curva 3 da Figura 2.1, para uma possível representação para esse polinômio.

De uma forma geral, temos que num polinômio da forma (2.1) com vários *knots*, a presença do termo genérico $\beta_{ij}(x - t_i)_+^j$ força a descontinuidade em t_i na j -ésima derivada de P , $P^{(j)}$. Isso também é válido para $j = 0$, quando $P^{(0)} = P$, em que a presença do termo $\beta_{i0}(x - t_i)_+^0$ indica que a função principal apresenta um salto no i -ésimo *knot*. Se esse termo está ausente, $P^{(j)}$ é contínua em t_i . Logo, uma função polinomial segmentada qualquer $P = P^{(0)}$ pode ser contínua em um determinado *knot* t_i por omitir de (2.1) o termo constante $\beta_{i0}(x - t_i)_+^0$, $P^{(1)}$ pode ser contínua em t_i omitindo-se o termo $\beta_{i1}(x - t_i)_+^1$, $P^{(2)}$ por omitir da função principal o termo $\beta_{i2}(x - t_i)_+^2$ e assim por diante. Desejando-se continuidade para a j -ésima derivada de P em um particular *knot* t_i , todas as derivadas de ordem inferior a j também devem ser contínuas e isto é garantido omitindo-se do polinômio (2.1) os termos $\beta_{im}(x - t_i)_+^m$, $m = 0, \dots, j$.

Também é possível termos diferentes restrições de continuidade impostas a diferentes peças polinomiais, eliminando-se os termos adequados. Quanto maior o grau de suavização desejado maior o número de termos eliminados. Conseqüentemente, um polinômio com menos parâmetros terá um pior ajuste e a curva final será mais suave, enquanto que menos

restrições de continuidade garantiria um ajuste melhor, uma vez que mais parâmetros estarão no modelo.

Cabe então, verificar quais restrições não resultam numa piora significativa do ajuste e testes podem ser realizados simultaneamente ou sequencialmente como em eliminação de variáveis através de testes de hipóteses. Se assumirmos (2.3) como a parte sistemática de um modelo, podemos realizar vários testes: $H_0 : \beta_{10} = 0$ testaria a continuidade de P_1 , $H_0 : \beta_{10} = \beta_{11} = 0$ verificaria a continuidade de P_2 e $P_2^{(1)}$ simultaneamente. Para testar a importância de todos os termos da função “+” isto é, se $\beta_{10} = \beta_{11} = \beta_{12} = 0$, devemos testar se um modelo polinomial segmentado ajusta-se significativamente melhor do que um polinômio quadrático padrão. De forma análoga, a ordem dos segmentos polinomiais também pode ser determinada através destes mesmos testes.

2.1.3 *Spline* polinomial mista

Há modelos em que as equações polinomiais são de graus variados e o modelo segmentado composto por todas as equações de mesma ordem deve ser modificado. Por exemplo, podemos querer modelar uma equação linear seguida por uma cúbica, uma cúbica seguida por uma quadrática etc. A seguir apresentaremos alguns exemplos que ilustram alguns diferentes tipos de composição com pequenas modificações, quando necessárias, que garantem os graus de cada equação da *spline*.

Exemplo 1

Seja uma função segmentada constituída por 2 equações, sendo a primeira linear e a segunda quadrática, escrita como

$$P(x) = \beta_{00} + \beta_{01}x + \beta_{11}(x - t_1)_+^1 + \beta_{12}(x - t_1)_+^2.$$

Ao separar por intervalos, vemos mais facilmente os graus das equações,

$$P(x) = \begin{cases} \beta_{00} + \beta_{01}x, & \text{para } x \leq t_1, \\ \beta_{00} - \beta_{11}t_1 + \beta_{12}t_1^2 + (\beta_{01} + \beta_{11} - 2\beta_{12}t_1)x + \beta_{12}x^2, & \text{para } x > t_1. \end{cases}$$

Nesse caso a construção foi direta e nenhuma intervenção na função inicial foi utilizada.

Exemplo 2

Considere agora o caso em que desejamos um polinômio segmentado formado pela primeira equação quadrática e a segunda linear, o contrário do exemplo anterior. Uma primeira tentativa na construção dessa função é dada por

$$P(x) = \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \beta_{11}(x - t_1)_+^1.$$

Tal função necessita de modificações, uma vez que separando por intervalos temos a contribuição do termo quadrático para valores superiores a t_1 , como observamos a seguir:

$$P(x) = \begin{cases} \beta_{00} + \beta_{01}x + \beta_{02}x^2, & \text{para } x \leq t_1, \\ (\beta_{00} - \beta_{11}t_1) + (\beta_{01} + \beta_{11})x + \beta_{02}x^2, & \text{para } x > t_1. \end{cases}$$

Com o intuito de linearizar a equação para valores $x > t_1$, nossa primeira idéia seria tomar $\beta_{02} = 0$. Porém, β_{02} não pode ser nulo por formar uma equação de grau 2 para $x \leq t_1$. Uma *spline* alternativa seria incorporar à função anterior um termo quadrático através da função “+”, ou seja,

$$P(x) = \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \beta_{11}(x - t_1)_+^1 + \beta_{12}(x - t_1)_+^2. \quad (2.4)$$

Assim, separando por intervalos, temos as seguintes equações:

$$P(x) = \begin{cases} \beta_{00} + \beta_{01}x + \beta_{02}x^2, & \text{para } x \leq t_1, \\ (\beta_{00} - \beta_{11}t_1 + \beta_{12}t_1^2) + (\beta_{01} + \beta_{11} - 2\beta_{12}t_1)x + (\beta_{02} + \beta_{12})x^2, & \text{para } x > t_1. \end{cases}$$

Observe que nesta função, para valores de $x > t_1$ há uma contribuição quadrática dada por $(\beta_{02} + \beta_{12})x^2$. Para linearizá-la basta impor a restrição

$$\beta_{02} + \beta_{12} = 0.$$

A imposição da restrição acima pode ser realizada de duas formas diferentes. A primeira, partindo da própria restrição e isolando o parâmetro β_{12} para então substituí-lo na função principal pelo valor $-\beta_{02}$. Logo, a equação (2.4) seria reescrita como

$$P(x) = \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \beta_{11}(x - t_1)_+^1 - \beta_{02}(x - t_1)_+^2.$$

Desta forma, a função principal separada por intervalos ficaria dada por

$$P(x) = \begin{cases} \beta_{00} + \beta_{01}x + \beta_{02}x^2, & \text{para } x \leq t_1, \\ \beta_{00} - \beta_{11}t_1 - \beta_{02}t_1^2 + (\beta_{01} + \beta_{11} + 2\beta_{02}t_1)x, & \text{para } x > t_1, \end{cases}$$

garantindo as formas quadrática e linear para as equações.

Uma outra maneira é tomar a equação (2.4) sujeita à restrição $\beta_{02} + \beta_{12} = 0$ que pode ser escrita na forma matricial $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$ com

$$\mathbf{C} = [0, 0, 1, 0, 1],$$

$$\boldsymbol{\beta}^T = [\beta_{00}, \beta_{01}, \beta_{02}, \beta_{11}, \beta_{12}]$$

e

$$\mathbf{d} = 0,$$

e utilizar métodos que considerem tal restrição, como será visto na seção 2.2.2.

Portanto, quando uma equação de grau r é seguida por outra de grau inferior, para que tenhamos coeficientes nulos em determinada ordem de X , duas alternativas são possíveis:

F1) - modificar a função passo-a-passo, reformulando os parâmetros até conseguir com que todas as equações da função segmentada apresentem o grau de interesse (Smith, 1979). Desta forma um novo modelo linear, com as restrições embutidas será formado;

F2) - partir do modelo completo com todas as equações de mesma ordem e impor outros tipos de restrições, agora sobre os parâmetros $\beta_1, \beta_2, \dots, \beta_p$ (Tarpey, 2000). A idéia aqui é garantir que certas potências de x não estejam presentes em determinados intervalos de dados da *spline* polinomial, e isto é obtido impondo-se que determinados coeficientes, que são formados geralmente por somas algébricas dos parâmetros β_j , sejam nulos. As restrições que trabalharemos serão lineares da forma $\mathbf{C}_j^T \boldsymbol{\beta} = d_j$, $j = 1, \dots, q$ em que \mathbf{C}_j é um vetor $p \times 1$ de coeficientes, $\boldsymbol{\beta}$ é o vetor $p \times 1$ de parâmetros desconhecidos, d_j um número real qualquer, e que simultaneamente estarão na forma $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$. Para ajustar o modelo, utiliza-se algum método de estimação restrita.

Exemplo 3

Considere agora uma *spline* contínua com 2 pontos de mudança em t_1 e t_2 dada por

$$S(x) = \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \beta_{03}x^3 + \beta_{11}(x - t_1)_+^1 + \beta_{12}(x - t_1)_+^2 + \beta_{13}(x - t_1)_+^3 + \beta_{21}(x - t_2)_+^1 + \beta_{22}(x - t_2)_+^2 + \beta_{23}(x - t_2)_+^3. \quad (2.5)$$

Separando por intervalos temos que $S(x)$ é expressa por

$$S(x) = \begin{cases} \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \beta_{03}x^3, & \text{para } x \leq t_1, \\ \beta_{00} - \beta_{11}t_1 + \beta_{12}t_1^2 - \beta_{13}t_1^3 \\ + (\beta_{01} + \beta_{11} - 2\beta_{12}t_1 + 3\beta_{13}t_1^2)x \\ + (\beta_{02} - 3\beta_{13}t_1 + \beta_{12})x^2 \\ + (\beta_{03} + \beta_{13})x^3, & \text{para } t_1 < x \leq t_2, (2.6) \\ \beta_{00} - \beta_{11}t_1 + \beta_{12}t_1^2 - \beta_{13}t_1^3 - \beta_{21}t_2 + \beta_{22}t_2^2 \\ + (\beta_{01} + \beta_{11} - 2\beta_{12}t_1 + 3\beta_{13}t_1^2 + \beta_{21} - 2\beta_{22}t_2 + 3\beta_{23}t_2^2)x \\ + (\beta_{02} - 3\beta_{13}t_1 + \beta_{12} + \beta_{22} - 3\beta_{23}t_2)x^2 \\ + (\beta_{03} + \beta_{13} + \beta_{23})x^3, & \text{para } x > t_2. \end{cases}$$

Podemos partir de (2.5) e modificar esta função de modo a obter uma *spline* polinomial formada não apenas por 3 equações cúbicas como acima e sim, por outra formada pelas seguintes equações: a primeira cúbica, a segunda quadrática e a terceira linear. Utilizaremos as 2 formas apresentadas em F1) e F2) para obter esta função.

F1) modificando a função passo-a-passo para embutir as restrições. Ao eliminarmos os termos com β_{13} e β_{12} não removemos a contribuição cúbica nas 2 últimas equações da *spline* porque ainda existe o termo $\beta_{03}x^3$. Observando os termos quadráticos que envolvem a função “+” nos intervalos de $x > t_1$ em (2.6), vemos que ao eliminarmos o termo com coeficiente β_{23} e impormos a condição de que $\beta_{03} + \beta_{13} = 0$, ou seja, $\beta_{13} = -\beta_{03}$, as últimas

equações não serão mais cúbicas. Assim, teríamos

$$S(x) = \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \beta_{03}[x^3 - (x - t_1)_+^3] + \beta_{11}(x - t_1)_+ + \beta_{12}(x - t_1)_+^2 + \beta_{21}(x - t_2)_+ + \beta_{22}(x - t_2)_+^2, \quad (2.7)$$

que é equivalente a

$$S(x) = \begin{cases} \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \beta_{03}x^3, & \text{para } x \leq t_1, \\ \beta_{00} - \beta_{11}t_1 + \beta_{12}t_1^2 + \beta_{03}t_1^3 \\ + (\beta_{01} + \beta_{11} - 2\beta_{12}t_1 - 3\beta_{03}t_1^2)x \\ + (\beta_{02} + 3\beta_{03}t_1 + \beta_{12})x^2, & \text{para } t_1 < x \leq t_2, \\ \beta_{00} - \beta_{11}t_1 + \beta_{12}t_1^2 + \beta_{03}t_1^3 - \beta_{21}t_2 + \beta_{22}t_2^2 \\ + (\beta_{01} + \beta_{11} - 2\beta_{12}t_1 - 3\beta_{03}t_1^2 + \beta_{21} - 2\beta_{22}t_2)x \\ + (\beta_{02} + 3\beta_{03}t_1 + \beta_{12} + \beta_{22})x^2, & \text{para } x > t_2. \end{cases}$$

Finalmente, para obter a terceira equação linear impomos a restrição

$$\beta_{02} + 3\beta_{03}t_1 + \beta_{12} + \beta_{22} = 0,$$

e, isolando o parâmetro β_{22} ,

$$\beta_{22} = -\beta_{02} - 3\beta_{03}t_1 - \beta_{12},$$

podemos substituir essa expressão em (2.7) e assim, obter a equação final

$$S(x) = \beta_{00} + \beta_{01}x + \beta_{02}[x^2 - (x - t_2)_+^2] + \beta_{03}[x^3 - (x - t_1)_+^3 - 3t_1(x - t_2)_+^2] + \beta_{11}(x - t_1)_+ + \beta_{12}[(x - t_1)_+^2 - (x - t_2)_+^2] + \beta_{21}(x - t_2)_+.$$

Por intervalos teríamos

$$S(x) = \begin{cases} \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \beta_{03}x^3, & \text{para } x \leq t_1, \\ \beta_{00} - \beta_{11}t_1 + \beta_{12}t_1^2 + \beta_{03}t_1^3 \\ + (\beta_{01} + \beta_{11} - 2\beta_{12}t_1 - 3\beta_{03}t_1^2)x \\ + (\beta_{02} + 3\beta_{03}t_1 + \beta_{12})x^2, & \text{para } t_1 < x \leq t_2, \\ \beta_{00} - \beta_{11}t_1 + \beta_{12}t_1^2 + \beta_{03}t_1^3 - \beta_{21}t_2 - \beta_{02}t_2^2 - 3\beta_{03}t_1t_2^2 - \beta_{12}t_2^2 \\ + (\beta_{01} + \beta_{11} - 2\beta_{12}t_1 - 3\beta_{03}t_1^2 \\ + \beta_{21} + 2\beta_{02}t_2 + 6\beta_{03}t_1t_2 + 2t_2\beta_{12})x, & \text{para } x > t_2. \end{cases}$$

F2) A segunda forma, considera a função (2.5) sujeito às restrições lineares que constituem a matriz \mathbf{C} de forma a obter os graus desejados para as equações 2 e 3 da *spline* polinomial. Neste caso, analisando a função por intervalos (2.6), para que a segunda equação seja de grau 2, devemos ter

$$\beta_{03} + \beta_{13} = 0,$$

e para valores $x > t_2$, em que desejamos uma equação linear, tomamos de (2.6)

$$\beta_{02} - 3\beta_{13}t_1 + \beta_{12} + \beta_{22} - 3\beta_{23}t_2 = 0$$

e

$$\beta_{03} + \beta_{13} + \beta_{23} = 0.$$

Podemos então, escrever as 3 restrições acima na forma matricial $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$, na qual

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & -3t_1 & 0 & 1 & -3t_2 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix},$$

$$\boldsymbol{\beta}^T = [\beta_{00}, \beta_{01}, \beta_{02}, \beta_{03}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{21}, \beta_{22}, \beta_{23}]$$

e

$$\mathbf{d} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Quanto mais complexas forem as modificações a serem feitas para garantir os graus das equações, há vantagens em proceder pela forma F2, bastando utilizar os métodos de estimação restrita. Um ponto a se lembrar é que a matriz \mathbf{C} deve ser construída de modo a ter posto completo.

2.2 Estimação

A estimação dos parâmetros de um modelo polinomial segmentado envolvendo a função “+” com pontos de mudança fixos e conhecidos pode ser realizada através do método de mínimos quadrados ordinários. Caso os pontos de mudança sejam considerados parâmetros a serem estimados, o problema torna-se não-linear. Gallant e Fuller(1973) trazem uma discussão a respeito. Este trabalho considerará apenas modelos com esses pontos fixos.

2.2.1 Caso 1: estimação irrestrita

Seja o modelo de regressão linear padrão dado por

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2.8)$$

em que \mathbf{y} é um vetor $n \times 1$ com valores observados da variável resposta, \mathbf{X} é a matriz $n \times p$ de posto completo com os valores de n casos de $p - 1$ variáveis explicativas do modelo, $\boldsymbol{\beta}$ é o vetor $p \times 1$ de coeficientes desconhecidos e $\boldsymbol{\varepsilon}$ é o vetor $n \times 1$ de erros. Supondo $\boldsymbol{\varepsilon} \sim \mathbf{N}_n(0, \sigma^2 \mathbf{I}_n)$, com σ^2 desconhecido, a estimação de $\boldsymbol{\beta}$ por mínimos quadrados ordinários é dada pela minimização da soma de quadrados dos erros. Seja o vetor $n \times 1$ com erros dado por

$$\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta},$$

o estimador de $\boldsymbol{\beta}$ é encontrado minimizando-se a função

$$\begin{aligned} S(\boldsymbol{\beta}) &= \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} \\ &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{y}^T \mathbf{y} - 2\mathbf{y}^T \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta}. \end{aligned} \quad (2.9)$$

Derivando-se (2.9) em relação a $\boldsymbol{\beta}$ e igualando-se o resultado a zero obtemos $-2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X}\boldsymbol{\beta} = 0$, e assumindo que \mathbf{X} tenha posto completo extraímos o estimador irrestrito de mínimos quadrados ordinários de $\boldsymbol{\beta}$,

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

A matriz de derivadas segundas $\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = 2(\mathbf{X}^T \mathbf{X})$ é definida positiva, o que garante que $\hat{\boldsymbol{\beta}}$ minimiza $S(\boldsymbol{\beta})$. O valor esperado e a matriz de variâncias-covariâncias desse estimador são dados, respectivamente, por

$$\mathbf{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

e

$$\text{Var}(\hat{\beta}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}. \quad (2.10)$$

Tomando-se $\mathbf{Q} = (\mathbf{X}^T\mathbf{X})^{-1}$, temos que,

$$\text{Var}(\hat{\beta}_j) = \sigma^2 Q_{jj},$$

e

$$\text{Cov}(\hat{\beta}_i, \hat{\beta}_j) = \sigma^2 Q_{ij}, \quad i \neq j,$$

em que Q_{ij} é o elemento (i, j) da matriz \mathbf{Q} . O estimador irrestrito de σ^2 é obtido da soma de quadrados residual dada por

$$\begin{aligned} \text{SQR} &= (\mathbf{y} - \hat{\mathbf{y}})^T(\mathbf{y} - \hat{\mathbf{y}}) \\ &= \mathbf{e}^T\mathbf{e} \\ &= \mathbf{y}^T\mathbf{y} - 2\hat{\beta}^T\mathbf{X}^T\mathbf{y} + \hat{\beta}^T\mathbf{X}^T\mathbf{X}\hat{\beta}, \end{aligned}$$

uma vez que $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\beta}$. Como $\mathbf{X}^T\mathbf{X}\hat{\beta} = \mathbf{X}^T\mathbf{y}$, temos que a expressão acima pode ser reescrita como

$$\text{SQR} = \mathbf{y}^T\mathbf{y} - \hat{\beta}^T\mathbf{X}^T\mathbf{y}.$$

A soma de quadrados residual (SQR) apresenta $n - p$ graus de liberdade, assim, podemos definir o quadrado residual médio por

$$\text{QMR} = \frac{\text{SQR}}{n - p},$$

em que n é o número de observações e p o número de parâmetros do modelo. Temos portanto que um estimador não viesado de σ^2 dado por $s^2 = \text{QMR}$, ou seja,

$$\begin{aligned} s^2 &= \frac{\text{SQR}}{n - p} \\ &= \frac{\mathbf{y}^T\mathbf{y} - \hat{\beta}^T\mathbf{X}^T\mathbf{y}}{n - p}. \end{aligned} \quad (2.11)$$

Obtido esse valor, podemos obter a variância estimada dos estimadores de β , dada por

$$\widehat{\text{Var}}(\hat{\beta}_j) = s^2 Q_{jj}.$$

O modelo de regressão linear dado por (2.8) inclui a classe de modelos de regressão polinomial segmentado em uma única variável X com *knots* fixos. Genericamente, para um modelo completo sem restrição temos

$$\mathbf{y} = \sum_{j=0}^r \beta_{0j} \mathbf{x}^j + \sum_{j=0}^r \sum_{i=1}^k \beta_{ij} (\mathbf{x} - \mathbf{t}_i)_+^j + \varepsilon, \quad (2.12)$$

em que os vetores $\mathbf{x} = (x_1, \dots, x_n)^T$ e $\mathbf{t}_i = (t_i, \dots, t_i)^T$ são de dimensão $n \times 1$, com $x_1 \leq x_2 \leq \dots \leq x_n$. Reescrevendo-se $\mathbf{x}^j = (x_1^j, \dots, x_n^j)^T$ como \mathbf{x}_{0j} e $(\mathbf{x} - \mathbf{t}_i)^j = ((x_1 - t_i)^j, \dots, (x_n - t_i)^j)^T$ como \mathbf{x}_{ij} , para $j = 0, \dots, r$ e $i = 1, \dots, k$, obtemos os regressores de uma regressão linear múltipla em que o vetor de parâmetros de dimensão $(r+1)(k+1) \times 1$ é dado por $\beta = (\beta_{00}, \dots, \beta_{0r}, \beta_{10}, \dots, \beta_{1r}, \beta_{k0}, \dots, \beta_{kr})^T$. A equação (2.12) transforma-se num problema de regressão linear dado por

$$\begin{aligned} \mathbf{y} &= \beta_{00} \mathbf{1} + \beta_{01} \mathbf{x}_{01} + \beta_{02} \mathbf{x}_{02} + \dots + \beta_{0r} \mathbf{x}_{0r} + \beta_{10} \mathbf{x}_{10} + \beta_{11} \mathbf{x}_{11} + \dots + \beta_{1r} \mathbf{x}_{1r} \\ &\quad + \dots + \beta_{k0} \mathbf{x}_{k0} + \beta_{k1} \mathbf{x}_{k1} + \dots + \beta_{kr} \mathbf{x}_{kr} + \varepsilon \\ &= \mathbf{X}\beta + \varepsilon, \end{aligned}$$

sendo $\mathbf{1} = (1, \dots, 1)^T$ o vetor de 1's com dimensão n e $\mathbf{X} = \begin{bmatrix} \mathbf{1} & \mathbf{x}_{01} & \mathbf{x}_{02} & \dots & \mathbf{x}_{kr} \end{bmatrix}$.

Como exemplo, considere o modelo segmentado cúbico contínuo com um ponto de mudança em t_1 , expresso por

$$\mathbf{y} = \beta_{00} \mathbf{1} + \beta_{01} \mathbf{x} + \beta_{02} \mathbf{x}^2 + \beta_{03} \mathbf{x}^3 + \beta_{11} (\mathbf{x} - \mathbf{t})_+^1 + \beta_{12} (\mathbf{x} - \mathbf{t})_+^2 + \beta_{13} (\mathbf{x} - \mathbf{t})_+^3 + \varepsilon,$$

sendo $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ e $\mathbf{t} = (t_1, \dots, t_1)^T$. Tal modelo pode ser escrito na forma (2.8) sendo a matriz modelo \mathbf{X} e o vetor de parâmetros β dados, respectivamente, por

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & 0 & 0 & 0 \\ 1 & x_2 & x_2^2 & x_2^3 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{t_1} & x_{t_1}^2 & x_{t_1}^3 & 0 & 0 & 0 \\ 1 & x_{t_1+1} & x_{t_1+1}^2 & x_{t_1+1}^3 & (x_{t_1+1} - t_1) & (x_{t_1+1} - t_1)^2 & (x_{t_1+1} - t_1)^3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & (x_n - t_1) & (x_n - t_1)^2 & (x_n - t_1)^3 \end{bmatrix}$$

ou seja,

$$\mathbf{X} = \begin{bmatrix} \mathbf{1} & \mathbf{x}_{01} & \mathbf{x}_{02} & \mathbf{x}_{03} & \mathbf{x}_{11} & \mathbf{x}_{12} & \mathbf{x}_{13} \end{bmatrix}$$

e

$$\boldsymbol{\beta} = [\beta_{00}, \beta_{01}, \beta_{02}, \beta_{03}, \beta_{11}, \beta_{12}, \beta_{13}]^T.$$

Definindo-se as matrizes acima, o estimador $\hat{\boldsymbol{\beta}}$ é obtido facilmente.

2.2.2 Caso 2: estimação restrita

Para os casos de modelos polinomiais segmentados mistos em que a ordem das equações polinomiais variam, vimos que foi necessário impor restrições do tipo $\mathbf{C}_j^T \boldsymbol{\beta} = d_j$, $j = 1, \dots, q$. Considerando-se necessárias q restrições simultâneas, obtemos a forma matricial

$$\mathbf{C}\boldsymbol{\beta} = \mathbf{d}, \quad (2.13)$$

em que \mathbf{C} é uma matriz $q \times p$ ($q < p$) de posto completo, $\boldsymbol{\beta}$ o vetor $p \times 1$ de parâmetros a serem estimados e $\mathbf{d} = (d_1, \dots, d_q)^T$ um vetor $q \times 1$ de constantes. Para esse caso, um dos possíveis métodos de estimação dos parâmetros é o método de Lagrange aplicado a mínimos quadrados ordinários, que consiste em minimizar a soma de quadrados dos erros

sujeita a (2.13). Para derivarmos o estimador restrito de β por esse método, considere o vetor $q \times 1$ de multiplicadores de Lagrange, λ . A função a ser minimizada é dada por

$$\begin{aligned} S_c(\beta) &= (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta) - 2\lambda^T(\mathbf{C}\beta - \mathbf{d}) \\ &= \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\beta - \beta^T\mathbf{X}^T\mathbf{y} + \beta^T\mathbf{X}^T\mathbf{X}\beta - 2\lambda^T\mathbf{C}\beta + 2\lambda^T\mathbf{d} \\ &= \mathbf{y}^T\mathbf{y} - 2\beta^T\mathbf{X}^T\mathbf{y} + \beta^T\mathbf{X}^T\mathbf{X}\beta - 2\beta^T\mathbf{C}^T\lambda + 2\mathbf{d}^T\lambda. \end{aligned}$$

Diferenciando-se parcialmente $S_c(\beta)$ com relação a β , igualando-se o resultado a zero obtemos

$$\begin{aligned} -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\tilde{\beta} - 2\mathbf{C}^T\tilde{\lambda} &= \mathbf{0} \\ \tilde{\beta} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T\tilde{\lambda} \\ \tilde{\beta} &= \hat{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T\tilde{\lambda}. \end{aligned} \quad (2.14)$$

Fazendo o mesmo com relação a λ , obtemos

$$\begin{aligned} -2\mathbf{C}\tilde{\beta} + 2\mathbf{d} &= \mathbf{0} \\ \mathbf{C}\tilde{\beta} &= \mathbf{d}. \end{aligned} \quad (2.15)$$

Portanto (2.14) e (2.15) formam um sistema de equações de onde se obtém

$$\tilde{\lambda} = -\{\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T\}^{-1}(\mathbf{C}\hat{\beta} - \mathbf{d}). \quad (2.16)$$

Finalmente, a expressão acima pode ser substituída em (2.14) de modo a obter-se o estimador de mínimos quadrados restrito de β , dado por

$$\tilde{\beta} = \hat{\beta} - (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T\{\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T\}^{-1}(\mathbf{C}\hat{\beta} - \mathbf{d}), \quad (2.17)$$

sendo $\hat{\beta}$ o estimador irrestrito de mínimos quadrados de β . Definindo-se as matrizes

$$\mathbf{Z} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T \quad (2.18)$$

e

$$\mathbf{M} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T (\mathbf{Z}^T \mathbf{Z})^{-1}, \quad (2.19)$$

podemos reescrever (2.16) e (2.17) em função da matriz \mathbf{Z} , obtendo-se

$$\begin{aligned} \tilde{\boldsymbol{\lambda}} &= (\mathbf{Z}^T \mathbf{Z})^{-1} \{\mathbf{d} - \mathbf{C}\hat{\boldsymbol{\beta}}\} \\ &= (\mathbf{Z}^T \mathbf{Z})^{-1} (\mathbf{Z}^T \mathbf{y} - \mathbf{d}) \end{aligned} \quad (2.20)$$

e

$$\begin{aligned} \tilde{\boldsymbol{\beta}} &= \hat{\boldsymbol{\beta}} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T (\mathbf{Z}^T \mathbf{Z})^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d}). \\ &= \hat{\boldsymbol{\beta}} - \mathbf{M}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d}). \end{aligned} \quad (2.21)$$

O vetor com os valores esperados é dado por

$$\begin{aligned} E(\tilde{\boldsymbol{\beta}}) &= E(\hat{\boldsymbol{\beta}}) - \mathbf{M}(E(\mathbf{C}\hat{\boldsymbol{\beta}}) - \mathbf{d}) \\ &= \boldsymbol{\beta} - \mathbf{M}(\mathbf{C}\boldsymbol{\beta} - \mathbf{d}) \\ &= \boldsymbol{\beta}, \end{aligned} \quad (2.22)$$

uma vez que $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$. A matriz de variâncias-covariâncias de $\tilde{\boldsymbol{\beta}}$ é dada por

$$\begin{aligned} \text{Var}(\tilde{\boldsymbol{\beta}}) &= \text{Var}\{\hat{\boldsymbol{\beta}} - \mathbf{M}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})\} \\ &= \text{Var}\{\hat{\boldsymbol{\beta}}\} + \text{Var}\{\mathbf{M}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})\} - 2\text{Cov}\{\hat{\boldsymbol{\beta}}, \mathbf{M}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})\}. \end{aligned} \quad (2.23)$$

De (2.10) sabemos que $\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$. Agora,

$$\begin{aligned} \text{Var}\{\mathbf{M}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})\} &= \text{Var}(\mathbf{M}\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{M}\mathbf{d}) \\ &= \text{Var}(\mathbf{M}\mathbf{C}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{M}\mathbf{C}\text{Var}(\hat{\boldsymbol{\beta}})\mathbf{C}^T\mathbf{M}^T \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{C} \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{C} (\mathbf{X}^T \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{C} (\mathbf{X}^T \mathbf{X})^{-1}. \end{aligned}$$

Também temos que

$$\begin{aligned}
\text{Cov}\{\hat{\beta}, M(C\hat{\beta} - d)\} &= E\left\{[\hat{\beta} - E(\hat{\beta})][M(C\hat{\beta} - d) - E\{M(C\hat{\beta} - d)\}]^T\right\} \\
&= E\left\{[\hat{\beta} - \beta][M(C\hat{\beta} - d) - 0]^T\right\} \\
&= E\left\{[(X^T X)^{-1}X^T y - \beta][M\{C(X^T X)^{-1}X^T y - d\}]^T\right\} \\
&= E\left\{[(X^T X)^{-1}X^T(X\beta + \varepsilon) - \beta][M\{C(X^T X)^{-1}X^T y - d\}]^T\right\} \\
&= E\left\{[\beta + (X^T X)^{-1}X^T \varepsilon - \beta][M\{C\beta + C(X^T X)^{-1}X^T \varepsilon - d\}]^T\right\} \\
&= E\left\{[(X^T X)^{-1}X^T \varepsilon][M\{C\beta + C(X^T X)^{-1}X^T \varepsilon - d\}]^T\right\} \\
&= E\left\{(X^T X)^{-1}X^T \varepsilon \beta^T C^T M^T - (X^T X)^{-1}X^T \varepsilon d^T M^T\right. \\
&\quad \left.+ (X^T X)^{-1}X^T \varepsilon \varepsilon^T X (X^T X)^{-1} C^T M^T\right\} \\
&= (X^T X)^{-1}X^T E(\varepsilon) \beta^T C^T M^T - (X^T X)^{-1}X^T E(\varepsilon) d^T M^T \\
&\quad + (X^T X)^{-1}X^T E(\varepsilon \varepsilon^T) X (X^T X)^{-1} C^T M^T \\
&= \sigma^2 (X^T X)^{-1} C^T M^T \\
&= \sigma^2 (X^T X)^{-1} C^T (Z^T Z)^{-1} C (X^T X)^{-1},
\end{aligned}$$

uma vez que $E(\varepsilon) = 0$ e $\text{Var}(\varepsilon) = E(\varepsilon \varepsilon^T) = \sigma^2 I_n$.

Logo, substituindo as expressões acima em (2.23), obtemos

$$\begin{aligned}
\text{Var}(\tilde{\beta}) &= \sigma^2 (X^T X)^{-1} + \sigma^2 (X^T X)^{-1} C^T (Z^T Z)^{-1} C (X^T X)^{-1} \\
&\quad - 2\sigma^2 (X^T X)^{-1} C^T (Z^T Z)^{-1} C (X^T X)^{-1} \\
&= \sigma^2 (X^T X)^{-1} - \sigma^2 (X^T X)^{-1} C^T (Z^T Z)^{-1} C (X^T X)^{-1} \\
&= \sigma^2 (X^T X)^{-1} [I_p - C^T (Z^T Z)^{-1} C (X^T X)^{-1}]. \tag{2.24}
\end{aligned}$$

A variância estimada de $\tilde{\beta}$ é obtida substituindo-se σ^2 por um estimador consistente. Para tanto, vamos obter a expressão para a soma de quadrados de resíduos de um modelo

linear restrito. Utilizando-se (2.19) e (2.21), podemos escrever SQR_c como

$$\begin{aligned}
SQR_c &= (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})^T(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) \\
&= \{\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}\mathbf{M}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})\}^T\{\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}\mathbf{M}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})\} \\
&= \{\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})\}^T\{\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})\} \\
&= \{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T + (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})^T\mathbf{X}^T\}\{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + \mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})\} \\
&= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T\mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) + \\
&\quad + (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})^T\mathbf{X}^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})^T\mathbf{X}^T\mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}). \tag{2.25}
\end{aligned}$$

Na expressão acima reconhecemos a soma de quadrados irrestrita dos resíduos dada por $SQR = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$, e que $\mathbf{X}^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = 0 = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T\mathbf{X}$. Também podemos retomar a expressão (2.17), ligeiramente modificada e assim escrever (2.25) como

$$\begin{aligned}
SQR_c &= SQR + (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})^T\mathbf{X}^T\mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) \\
&= SQR + [\mathbf{M}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})]^T\mathbf{X}^T\mathbf{X}[\mathbf{M}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})] \\
&= SQR + (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})^T\mathbf{M}^T\mathbf{X}^T\mathbf{X}\mathbf{M}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d}) \\
&= SQR + (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})^T\{\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T\}^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d}) \\
&= SQR + (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})^T(\mathbf{Z}^T\mathbf{Z})^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d}).
\end{aligned}$$

Uma vez que $E(SQR_c) = (n - p + q)\sigma^2$, temos que o estimador restrito não-viesado da variância do erro é dado por

$$s_c^2 = \frac{SQR_c}{(n - p + q)}. \tag{2.26}$$

Dessa forma, a variância estimada de $\tilde{\boldsymbol{\beta}}$ pode ser obtida através de

$$\hat{\text{Var}}(\tilde{\boldsymbol{\beta}}) = s_c^2(\mathbf{X}^T\mathbf{X})^{-1}[\mathbf{I}_p - \mathbf{C}^T(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}].$$

2.3 Seleção do modelo

A seleção de um modelo de regressão pode ser efetuada por vários critérios: R^2 , R^2 ajustado, métodos *stepwise*, *forward* e *backward*, método de Akaike etc (ver, por exemplo, Paula, 2004). Quando o interesse principal concentra-se em realizar previsões, a estatística PRESS (soma de quadrados de resíduos preditos) pode ser uma boa opção. Inclusive esta estatística é útil para comparar *splines* irrestritas com aquelas em que há restrições nos parâmetros.

2.3.1 Estatística PRESS

A estatística PRESS foi proposta por Allen (1974) e baseia-se na técnica *jackknife* em que se determinam os coeficientes de um modelo de regressão excluindo-se o i -ésimo caso $(x_{i1}, \dots, x_{ip}, y_i)$ da amostra, e então, a partir deste modelo ajustado, prevê-se a resposta $\hat{y}_{(i)}$ no ponto (x_{i1}, \dots, x_{ip}) .

A diferença entre a i -ésima resposta observada e a i -ésima resposta prevista é definida como resíduo PRESS, ou seja, $e_{(i)} = y_i - \hat{y}_{(i)}$. Repete-se o ajuste acima para cada amostra possível, ou seja, excluindo-se um caso por vez. Então, tem-se que a estatística PRESS é definida por,

$$\text{PRESS} = \sum_{i=1}^n e_{(i)}^2.$$

Porém, pode-se determinar os resíduos $e_{(i)}$ sem realizar os n ajustes. Para tanto, considere a expressão para o vetor de parâmetros ajustados sem o i -ésimo caso (Cook e Weisberg, 1982)

$$\hat{\beta}_{(i)} = \hat{\beta} - \frac{e_i(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i}{1 - h_{ii}}, \quad (2.27)$$

em que $e_i = y_i - \hat{y}_i = y_i - \mathbf{x}_i^T \hat{\beta}$ é o resíduo ordinário associado ao i -ésimo caso, \mathbf{x}_i^T é a i -ésima linha da matriz \mathbf{X} e h_{ii} é o i -ésimo elemento da diagonal principal da matriz de projeção de vetores do \mathcal{R}^n sobre o subespaço gerado pelas colunas da matriz \mathbf{X} , $\mathbf{P}_\mathbf{X} =$

$\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \mathbf{H}$, dado por

$$h_{ii} = \mathbf{x}_i^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i, \quad (2.28)$$

em que $\frac{1}{n} \leq h_{ii} \leq \frac{1}{c}$ sendo c o número de linhas de \mathbf{X} idênticas a \mathbf{x}_i^T (Cook e Weisberg, 1982). Portanto, o resíduo PRESS pode ser escrito como

$$\begin{aligned} e_{(i)} &= y_i - \hat{y}_{(i)} \\ &= y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{(i)} \\ &= y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \frac{e_i}{1 - h_{ii}} \\ &= e_i + h_{ii} \frac{e_i}{1 - h_{ii}} \\ &= \frac{e_i(1 - h_{ii}) + e_i h_{ii}}{1 - h_{ii}} \\ &= \frac{e_i}{1 - h_{ii}}, \end{aligned}$$

de onde se obtém

$$e_{(i)} = \frac{y_i - \hat{y}_i}{1 - h_{ii}}, \quad i = 1, \dots, n. \quad (2.29)$$

Podemos determinar a estatística PRESS para qualquer subconjunto de variáveis e assim selecionar o modelo que apresenta o menor valor para a estatística. Também é possível comparar modelos irrestritos com modelos restritos, utilizando-se uma versão dessa estatística para modelos com restrições. Tarpey (2000) apresenta uma forma similar para calcular os resíduos PRESS da mesma forma que (2.29) para o caso de modelos da forma $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, em que $\boldsymbol{\beta}$ está sujeito a restrições do tipo $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$. Sua metodologia consiste em equiparar 2 modelos: um restrito na forma $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$, com um irrestrito que possa embutir as restrições na própria matriz modelo \mathbf{X} , obtendo uma matriz \mathbf{X} modificada. Assim, obtém-se resultados de interesse para casos restritos através da teoria já desenvolvida para modelos irrestritos. Neste trabalho, adotaremos outra estratégia. Vamos deduzir a expressão para o estimador do vetor de parâmetros restritos sem a i -ésima observação, $\tilde{\boldsymbol{\beta}}_{(i)}$ e obter o resíduo PRESS restrito $\tilde{e}_{(i)} = y_i - \tilde{y}_{(i)}$, sendo $\tilde{y}_{(i)} = \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}_{(i)}$, o i -ésimo valor ajustado de um modelo restrito no qual a i -ésima observação foi excluída.

Estatística PRESS em modelos restritos, PRESSc

Em modelos restritos o vetor de valores ajustados é dado por $\tilde{y} = X\tilde{\beta}$, sendo $\tilde{\beta}$ dado por (2.17). Considerando-se as matrizes X e C de posto completo, podemos substituir a expressão (2.21) e, utilizando também as expressões para Z e M definidas em (2.18) e (2.19), reescrevemos o vetor com os valores ajustados da seguinte forma:

$$\begin{aligned}
 \tilde{y} &= X[\hat{\beta} - MC\hat{\beta} + Md] \\
 &= X\hat{\beta} - XMC\hat{\beta} + XMd \\
 &= X(X^T X)^{-1}X^T y - Z(Z^T Z)^{-1}Z^T y + Z(Z^T Z)^{-1}d \\
 &= P_X y - P_Z y + Z(Z^T Z)^{-1}d,
 \end{aligned} \tag{2.30}$$

em que $P_X = X(X^T X)^{-1}X^T$ e $P_Z = Z(Z^T Z)^{-1}Z^T$. Nota-se que P_X e P_Z são matrizes de projeção de vetores do \mathbb{R}^n nos subespaços gerados pelas colunas de X e Z , respectivamente. Assumindo que as matrizes X e C tenham posto completo, isto é, $\text{posto}(X) = p$ e $\text{posto}(C) = q$, temos que $P_X = H$ e nomeando $P_Z = G$, podemos escrever (2.30) como

$$\tilde{y} = (H - G)y + Z(Z^T Z)^{-1}d.$$

Para o caso em que $d = 0$, isto é, $C\beta = 0$ a expressão acima resume-se a

$$\tilde{y} = (H - G)y. \tag{2.31}$$

Portanto, a matriz $H - G$ é a responsável por fornecer o valor ajustado do vetor y em modelos restritos quando $d = 0$.

Para obter uma forma direta para os resíduos PRESS em modelos restritos, partimos da estimativa do vetor de parâmetros sem o i -ésimo caso deduzida abaixo.

Excluindo-se o i -ésimo caso da expressão para $\tilde{\beta}$ dada em (2.14) podemos escrever

$$\tilde{\beta}_{(i)} = \hat{\beta}_{(i)} + (X_{(i)}^T X_{(i)})^{-1} C^T \tilde{\lambda}_{(i)},$$

em que a forma para $\hat{\beta}_{(i)}$ está em (2.27), e as demais matrizes são dadas por

$$(\mathbf{X}_{(i)}^T \mathbf{X}_{(i)})^{-1} = (\mathbf{X}^T \mathbf{X})^{-1} + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1}}{1 - h_{ii}} \quad (\text{Cook e Weisberg, 1982})$$

e

$$\begin{aligned} \tilde{\lambda}_{(i)} &= (\mathbf{Z}_{(i)}^T \mathbf{Z}_{(i)})^{-1} (\mathbf{d} - \mathbf{Z}_{(i)}^T \mathbf{y}_{(i)}) \\ &= \tilde{\lambda} + \frac{(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{z}_i v_i}{1 - h_{ii} + g_{ii}}, \end{aligned} \quad (\text{Paula, 1993})$$

sendo $\tilde{\lambda}$ o vetor de multiplicadores de Lagrange estimados dado por (2.20) e $v_i = y_i - \bar{y}_i$.

Assim, segue que

$$\begin{aligned} \tilde{\beta}_{(i)} &= \hat{\beta} - \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i}{1 - h_{ii}} + \left[(\mathbf{X}^T \mathbf{X})^{-1} + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1}}{1 - h_{ii}} \right] \mathbf{C}^T \\ &\quad \times \left[\tilde{\lambda} + \frac{(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{z}_i v_i}{1 - h_{ii} + g_{ii}} \right] \\ &= \hat{\beta} - \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i}{1 - h_{ii}} + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T \tilde{\lambda} + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{z}_i v_i}{1 - h_{ii} + g_{ii}} \\ &\quad + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T \tilde{\lambda}}{1 - h_{ii}} \\ &\quad + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{z}_i v_i}{(1 - h_{ii})(1 - h_{ii} + g_{ii})}. \end{aligned} \quad (2.32)$$

Como $\mathbf{Z} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T$ e $\mathbf{G} = \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T$ temos que a i -ésima linha da matriz \mathbf{Z} e o i -ésimo elemento da diagonal principal de \mathbf{G} são dados, respectivamente, por

$$\mathbf{z}_i^T = \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T$$

e

$$g_{ii} = \mathbf{z}_i^T (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{z}_i. \quad (2.33)$$

Assim, podemos reescrever (2.32) como

$$\begin{aligned}
\tilde{\beta}_{(i)} &= \hat{\beta} + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T \tilde{\lambda} - \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i}{1 - h_{ii}} + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{z}_i^T \tilde{\lambda}}{1 - h_{ii}} \\
&\quad + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{z}_i v_i}{1 - h_{ii} + g_{ii}} + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i v_i g_{ii}}{(1 - h_{ii})(1 - h_{ii} + g_{ii})} \\
&= \tilde{\beta} - \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i}{1 - h_{ii}} + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{z}_i^T \tilde{\lambda}}{1 - h_{ii}} \\
&\quad + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{z}_i v_i}{1 - h_{ii} + g_{ii}} + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i v_i g_{ii}}{(1 - h_{ii})(1 - h_{ii} + g_{ii})} \tag{2.34}
\end{aligned}$$

uma vez que $\tilde{\beta} = \hat{\beta} + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T \tilde{\lambda}$. Também dessa expressão obtemos, ao pré-multiplicar pela matriz \mathbf{X}

$$\begin{aligned}
\mathbf{X} \tilde{\beta} &= \mathbf{X} \hat{\beta} + \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T \tilde{\lambda}, \\
\mathbf{X} \tilde{\beta} - \mathbf{X} \hat{\beta} &= \mathbf{Z} \tilde{\lambda}, \\
(\mathbf{y} - \mathbf{X} \hat{\beta}) - (\mathbf{y} - \mathbf{X} \tilde{\beta}) &= \mathbf{Z} \tilde{\lambda}, \\
\mathbf{e} - \mathbf{v} &= \mathbf{Z} \tilde{\lambda},
\end{aligned}$$

que resulta em,

$$e_i - v_i = \mathbf{z}_i^T \tilde{\lambda}, \quad i = 1, \dots, n.$$

Substituindo a expressão acima em (2.34) obtemos

$$\begin{aligned}
\tilde{\beta}_{(i)} &= \tilde{\beta} - \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i}{1 - h_{ii}} + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i (e_i - v_i)}{1 - h_{ii}} + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{z}_i v_i}{1 - h_{ii} + g_{ii}} \\
&\quad + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i v_i g_{ii}}{(1 - h_{ii})(1 - h_{ii} + g_{ii})} \\
&= \tilde{\beta} + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{z}_i v_i}{1 - h_{ii} + g_{ii}} - \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i v_i}{1 - h_{ii}} + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i v_i g_{ii}}{(1 - h_{ii})(1 - h_{ii} + g_{ii})} \\
&= \tilde{\beta} - \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i v_i}{1 - h_{ii} + g_{ii}} + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{z}_i v_i}{1 - h_{ii} + g_{ii}}.
\end{aligned}$$

Definindo-se $\mathbf{a}_i = \mathbf{C}^T(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{z}_i$, temos que a estimativa do vetor de parâmetros sem a i -ésima observação é dada, num modelo linear restrito, por

$$\tilde{\boldsymbol{\beta}}_{(i)} = \tilde{\boldsymbol{\beta}} - \frac{v_i}{1 - h_{ii} + g_{ii}}(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{x}_i - \mathbf{a}_i). \quad (2.35)$$

Assim, temos que o resíduo PRESS para modelos restritos é dado por

$$\begin{aligned} \tilde{e}_{(i)} &= y_i - \tilde{y}_{(i)} \\ &= y_i - \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}_{(i)} \\ &= y_i - \mathbf{x}_i^T \left\{ \tilde{\boldsymbol{\beta}} - \frac{v_i}{1 - h_{ii} + g_{ii}}(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{x}_i - \mathbf{a}_i) \right\} \\ &= y_i - \mathbf{x}_i^T \tilde{\boldsymbol{\beta}} + \mathbf{x}_i^T (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{x}_i \frac{v_i}{1 - h_{ii} + g_{ii}} - \mathbf{x}_i^T (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{a}_i \frac{v_i}{1 - h_{ii} + g_{ii}}. \end{aligned}$$

Uma vez que $\mathbf{x}_i^T (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{a}_i = \mathbf{z}_i^T (\mathbf{Z}^T\mathbf{Z})^{-1} \mathbf{z}_i = g_{ii}$, podemos reescrever a expressão acima como

$$\begin{aligned} \tilde{e}_{(i)} &= v_i + \frac{h_{ii}v_i}{1 - h_{ii} + g_{ii}} - \frac{g_{ii}v_i}{1 - h_{ii} + g_{ii}} \\ &= \frac{v_i}{1 - h_{ii} + g_{ii}}, \end{aligned}$$

em que $v_i = y_i - \tilde{y}_i$ é o i -ésimo resíduo restrito. Finalmente, temos que a estatística PRESS para modelos restritos pode ser obtida diretamente por

$$\text{PRESS}_c = \sum_{i=1}^n \tilde{e}_{(i)}^2 = \sum_{i=1}^n \left(\frac{y_i - \tilde{y}_i}{1 - h_{ii} + g_{ii}} \right)^2. \quad (2.36)$$

em que h_{ii} e g_{ii} foram dados, respectivamente, em (2.28) e (2.33). Podemos então utilizar facilmente a estatística PRESS para comparar modelos polinomiais segmentados restritos e irrestritos.

Tarpey (2000) chega ao resultado (2.36) através de outro método o qual consiste em obter medidas de interesse num modelo restrito por intermédio de um modelo equivalente irrestrito.

2.4 Técnicas de diagnóstico

Selecionado um modelo, uma etapa posterior na análise de regressão é realizar uma análise de diagnóstico, basicamente gráfica, para verificar se existem afastamentos sérios das posições assumidas para o modelo, como também, destacar do rol de dados, aquelas observações que são consideradas atípicas principalmente por influenciarem desproporcionalmente os resultados do ajuste. Preocupar-nos-emos em destacar os seguintes pontos atípicos: a) **aberrantes**, que possuem um alto resíduo, podendo alterar principalmente a estimativa do intercepto; b) **de alavanca**, que estão mais afastados dos demais, ou seja, estão localizados numa região remota no subespaço formado pelas colunas da matriz X , podendo influenciar ou não, as estimativas dos coeficientes; c) **influentes**, que influenciam desproporcionalmente as estimativas dos coeficientes, alterando a inclinação da reta ajustada. Outra questão na análise de diagnóstico é avaliar a adequabilidade da distribuição proposta para a variável resposta (ou para o erro). Para esse caso Atkinson (1981,1985) sugere a construção de um tipo de banda de confiança denominada envelope, que é informativa quanto à existência de pontos discrepantes, sobre a falta de homogeneidade das variâncias, adequabilidade do erro e da relação funcional entre a média e as variáveis explicativas. Assim como a estatística PRESS, a análise de diagnóstico poderá ser útil na escolha entre um modelo segmentado restrito e irrestrito.

2.4.1 Diagnóstico em modelos irrestritos

As principais técnicas para diagnóstico em modelos de regressão polinomial segmentados quando não há restrições adicionais sobre os parâmetros coincidem com aquelas utilizadas em modelos normais lineares padrão (vide, por exemplo, Belsley et al 1990). A seguir, apresentaremos algumas dessas técnicas.

Resíduos

Resíduos são utilizados em vários procedimentos estatísticos, sendo que um deles é quantificar a discordância entre os dados observados e ajustados pelo modelo, trazendo informação a respeito das suposições assumidas pelo modelo.

Então, tem-se que o resíduo para a i -ésima observação pode ser definido como uma função do valor observado, y_i , e do valor ajustado, \hat{y}_i . No caso de modelos segmentados irrestritos, o vetor de valores ajustados é dado por $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{H}\mathbf{y}$, tal que o vetor de resíduos ordinários do modelo irrestrito seja expresso na forma

$$\begin{aligned}\mathbf{e} &= \mathbf{y} - \hat{\mathbf{y}} \\ &= (\mathbf{I}_n - \mathbf{H})\mathbf{y}.\end{aligned}$$

Assumindo que $\mathbf{y} \sim \mathbf{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$, temos que

$$\begin{aligned}E(\mathbf{e}) &= (\mathbf{I}_n - \mathbf{H})E(\mathbf{y}) \\ &= (\mathbf{I}_n - \mathbf{H})\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{X}\boldsymbol{\beta} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{0}\end{aligned}$$

e

$$\begin{aligned}\text{Var}(\mathbf{e}) &= (\mathbf{I}_n - \mathbf{H})\text{Var}(\mathbf{y})(\mathbf{I}_n - \mathbf{H})^T \\ &= \sigma^2(\mathbf{I}_n - \mathbf{H})(\mathbf{I}_n - \mathbf{H})^T \\ &= \sigma^2(\mathbf{I}_n - \mathbf{H}),\end{aligned}$$

uma vez que a matriz $\mathbf{I}_n - \mathbf{H}$ é simétrica e idempotente. Portanto, temos que o i -ésimo resíduo e_i tem distribuição normal de média zero, com variâncias e covariâncias dadas por

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

e

$$\text{Cov}(e_i, e_j) = -\sigma^2 h_{ij}, i \neq j.$$

Uma vez que as variâncias dos e_i 's são dependentes de escala, isto é, dependem de σ^2 e de h_{ii} , podendo variar para cada i , $i = 1, \dots, n$, seria interessante obter uma versão padronizada dos mesmos. Cook e Weisberg (1982), seguindo as propostas de Margolin(1977), definem o processo de studentização como sendo a divisão de uma estatística dependente de escala, que no nosso caso é o resíduo ordinário, pela estimativa do desvio padrão, tal que a razão resulte numa distribuição que seja livre de parâmetro de ruído (σ^2). David (1981) vai além, distinguindo studentização interna, em que a estatística e a estimativa de seu desvio padrão são originários dos mesmos dados e portanto dependentes, da studentização externa, em que eles são independentes.

Dividindo-se e_i pela estimativa do respectivo desvio padrão, $dp(e_i) = \sqrt{s^2(1 - h_{ii})}$, em que s^2 é o quadrado médio residual definido em (2.11), obtém-se o resíduo studentizado internamente,

$$t_i = \frac{e_i}{s(1 - h_{ii})^{1/2}}, \quad i = 1, \dots, n. \quad (2.37)$$

Em Cook e Weisberg (1982) prova-se que $t_i^2/(n - p)$ tem uma distribuição beta com parâmetros $\frac{1}{2}$ e $(n - p - 1)/2$, sendo que nesse caso e_i e s são dependentes. Para obtermos um estimador de σ^2 que seja independente de e_i , no caso de assumirmos normalidade para os erros, podemos seguir a sugestão de Elleberg(1973) e encontrar uma expressão em forma quadrática em e que seja independente de $e_{(i)}$, obtendo conseqüentemente, e_i e $s_{(i)}^2$ independentes. Tal expressão é dada por

$$\text{SQR}_{(i)} = \mathbf{e}^T \mathbf{e} - e_i^2 (1 - h_{ii})^{-1},$$

sendo que $\text{SQR}_{(i)}$ representa a soma de quadrados dos resíduos sem o i -ésimo caso, com $\text{SQR}_{(i)} \sim \sigma^2 \chi_{(n-p-1)}^2$. Uma vez que $s^2 = \mathbf{e}^T \mathbf{e} / (n - p)$ e $s_{(i)}^2 = \text{SQR}_{(i)} / (n - p - 1)$, temos

que a expressão acima pode ser reescrita como

$$(n - p - 1)s_{(i)}^2 = (n - p)s^2 - e_i^2/(1 - h_{ii}). \quad (2.38)$$

Agora, dividindo-se ambos os lados da equação (2.38) por σ^2 e reorganizando a equação acima obtemos,

$$\frac{(n - p)s^2}{\sigma^2} = \frac{(n - p - 1)s_{(i)}^2}{\sigma^2} + \frac{e_i^2}{\sigma^2(1 - h_{ii})},$$

com $(n - p - 1)s_{(i)}^2/\sigma^2 \sim \chi_{(n-p-1)}^2$, $e_i^2(1 - h_{ii})^{-1}/\sigma^2 \sim \chi_{(1)}^2$ e $(n - p)s^2/\sigma^2 \sim \chi_{(n-p)}^2$, o que também garante a independência entre $s_{(i)}^2$ e e_i (Teorema de Fisher-Cochran).

Portanto, um resíduo studentizado externamente é obtido substituindo-se s por $s_{(i)}$ em (2.37). Desse modo,

$$t_{S_i} = \frac{e_i}{s_{(i)}(1 - h_{ii})^{1/2}},$$

que segue uma distribuição t de Student com $n - p - 1$ graus de liberdade. Para obter-se $s_{(i)}^2$ basta isolá-lo em (2.38) até se chegar à seguinte expressão:

$$s_{(i)}^2 = s^2 \left(\frac{n - p - t_i^2}{n - p - 1} \right). \quad (2.39)$$

Podemos também substituir (2.39) em (??), e assim encontrar a relação entre t_{S_i} e t_i expressa por

$$t_{S_i} = t_i \left(\frac{n - p - 1}{n - p - t_i^2} \right).$$

Os resíduos studentizados externamente t_{S_i} , poderão ser utilizados para:

- a) destacar pontos aberrantes, fazendo um gráfico de t_{S_i} contra a ordem das observações;
- b) para detectar heteroscedasticidade, através do gráfico de t_{S_i} contra os valores ajustados, \hat{y}_i ;
- c) falta de algum termo extra, fazendo um gráfico de t_{S_i} contra \hat{y}_i ou contra as covariáveis que estão ou não foram incluídas no modelo;

d) afastamento da normalidade, através da construção do gráfico normal de probabilidades com bandas de confiança simuladas, denominado envelope (Atkinson, 1985). Os envelopes também são úteis para detectar pontos aberrantes, verificar se há constância nas variâncias das observações, como também, avaliar a qualidade do ajuste do modelo.

Alavancagem

Dos elementos da diagonal principal da matriz \mathbf{H} , obtém-se um valor importante para a construção de técnicas de diagnóstico. Lembrando que o i -ésimo valor ajustado fica dado por

$$\hat{y}_i = h_{ii}y_i + \sum_{j \neq i} h_{ji}y_j,$$

nota-se que para valores altos de h_{ii} predomina na expressão acima a influência de y_i sobre o correspondente valor ajustado. Dessa forma, utiliza-se h_{ii} como medida da influência da i -ésima observação sobre o próprio valor ajustado. Caso todos os pontos da amostra de dados exerçam a mesma influência sobre os valores ajustados, espera-se que todos os h_{ii} estejam próximos de $\text{tr} \frac{\mathbf{H}}{n} = \frac{p}{n}$. Então, os pontos que estejam muito distantes desse valor, devem ser cuidadosamente examinados, uma vez que estão localizados em regiões remotas no subespaço gerado pelas colunas da matriz \mathbf{X} . Sugere-se utilizar o valor $\frac{2p}{n}$ num gráfico de t_{S_i} contra a ordem das observações, como limite inferior para se considerar pontos de alavanca.

Distância de Cook

A distância de Cook (Cook, 1977) é uma medida de influência que mede quanto uma perturbação $\delta = (\delta_1, \dots, \delta_n)^T$ afasta $\hat{\beta}_\delta$ de $\hat{\beta}$, segundo a métrica $\mathbf{X}^T \mathbf{X}$, sendo definida por

$$D_\delta = \frac{(\hat{\beta} - \hat{\beta}_\delta)^T (\mathbf{X}^T \mathbf{X}) (\hat{\beta} - \hat{\beta}_\delta)}{ps^2}, \quad (2.40)$$

em que $\hat{\beta}_\delta$ é a estimativa de máxima verossimilhança supondo que o logaritmo da função de verossimilhança para β seja expresso na forma

$$L_\delta(\beta, \mathbf{y}) = \sum_{j=1}^n \delta_j l(\beta; y_j),$$

com δ_j sendo uma perturbação tal que $0 \leq \delta_j \leq 1$. Em particular, quando o i -ésimo ponto é excluído, (2.40) fica dado por,

$$\begin{aligned} D_{(i)} &= \frac{(\hat{\beta} - \hat{\beta}_{(i)})^T (\mathbf{X}^T \mathbf{X}) (\hat{\beta} - \hat{\beta}_{(i)})}{ps^2} \\ &= \frac{\{(\mathbf{X}\hat{\beta} - \mathbf{X}\hat{\beta}_{(i)})\}^T \mathbf{X}(\hat{\beta} - \hat{\beta}_{(i)})}{ps^2} \\ &= \left(\frac{e_i}{1 - h_{ii}}\right) \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \left(\frac{e_i}{1 - h_{ii}}\right) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \frac{1}{ps^2} \\ &= \left(\frac{e_i}{1 - h_{ii}}\right)^2 \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \frac{1}{ps^2} \\ &= \frac{t_i^2}{p} \frac{h_{ii}}{(1 - h_{ii})}. \end{aligned}$$

que é obtido utilizando-se uma pequena modificação da expressão para $\hat{\beta}_{(i)}$ dada em (2.27). Portanto, observações com h_{ii} alto e/ou t_i alto devem em geral levar a valores altos para $D_{(i)}$, indicando uma influência maior dessas observações na estimativa dos coeficientes em comparação às demais. Um gráfico de $D_{(i)}$ contra a ordem das observações será útil para comparar esses valores. Outras medidas de influência tais como DFFITS_{*i*} têm sido sugeridas, contudo não refletem da mesma forma a influência das observações nos coeficientes do modelo como $D_{(i)}$ (vide discussão, por exemplo, em Paula, 2004).

2.4.2 Diagnóstico em modelos restritos

Os métodos utilizados em modelos de regressão irrestrita nem sempre podem ser estendidos diretamente para modelos restritos, a menos que as estimativas restritas e irrestritas

coincidam. Apresentamos a seguir algumas técnicas de diagnóstico para modelos normais lineares com restrições na forma $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$, em particular para $\mathbf{d} = \mathbf{0}$. Muitos dos resultados são adaptações do trabalho de Paula (1993) onde são apresentadas algumas técnicas de diagnóstico para modelos normais lineares com restrições na forma $\mathbf{C}\boldsymbol{\beta} \geq \mathbf{d}$.

Resíduos

Do modelo (2.8) sujeito a $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ segue que o vetor de valores ajustados é dado por $\tilde{\mathbf{y}} = \mathbf{X}\tilde{\boldsymbol{\beta}}$, e utilizando-se (2.31), temos que o vetor de resíduos ordinários para o modelo normal linear restrito é dado por

$$\begin{aligned}\mathbf{v} &= \mathbf{y} - \tilde{\mathbf{y}} \\ &= \mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}} \\ &= \mathbf{y} - (\mathbf{H} - \mathbf{G})\mathbf{y} \\ &= (\mathbf{I} - \mathbf{H} + \mathbf{G})\mathbf{y}.\end{aligned}$$

De maneira similar à seção anterior, vamos apresentar uma versão studentizada para o vetor de resíduos \mathbf{v} . Uma vez que $\tilde{\boldsymbol{\beta}}$ tem distribuição normal p -variada com média e variância dadas por (2.22) e (2.24), segue que o resíduo \mathbf{v} tem distribuição normal com média e variância dadas por

$$\begin{aligned}\mathbf{E}(\mathbf{v}) &= \mathbf{E}(\mathbf{y}) - \mathbf{X}\mathbf{E}(\tilde{\boldsymbol{\beta}}) \\ &= \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{0}\end{aligned}$$

e

$$\begin{aligned}\text{Var}(\mathbf{v}) &= (\mathbf{I} - \mathbf{H} + \mathbf{G})\text{Var}(\mathbf{y})(\mathbf{I} - \mathbf{H} + \mathbf{G})^T \\ &= \sigma^2(\mathbf{I} - \mathbf{H} + \mathbf{G})(\mathbf{I} - \mathbf{H} + \mathbf{G})^T \\ &= \sigma^2\{\mathbf{I} - \mathbf{H} + \mathbf{G} - \mathbf{H} + \mathbf{H}\mathbf{H} - \mathbf{H}\mathbf{G} + \mathbf{G} - \mathbf{G}\mathbf{H} + \mathbf{G}\mathbf{G}\},\end{aligned}$$

e como $\mathbf{HG} = \mathbf{GH} = \mathbf{G}$ e as matrizes \mathbf{H} e \mathbf{G} são idempotentes, temos que

$$\text{Var}(\mathbf{v}) = \sigma^2(\mathbf{I} - \mathbf{H} + \mathbf{G}).$$

uma vez que a matriz $(\mathbf{I} - \mathbf{H} + \mathbf{G})$ é simétrica e idempotente, assim como $\mathbf{I} - \mathbf{H}$.

Temos então que a variância do i -ésimo elemento de \mathbf{v} é dada por

$$\text{Var}(v_i) = \sigma^2\{(1 - h_{ii}) + g_{ii}\},$$

e ainda

$$\text{Cov}(v_i, v_j) = \sigma^2(g_{ij} - h_{ij}), \text{ para } i \neq j,$$

em que h_{ij} e g_{ij} são os elementos (i, j) das matrizes \mathbf{H} e \mathbf{G} , respectivamente.

Paula (1993) apresenta uma versão studentizada externamente para v_i que é obtida dividindo-se v_i pela estimativa do erro padrão residual que seja independente de v_i . Temos então o resíduo studentizado externamente dado por

$$t_{v_i} = \frac{v_i}{s_{(i)}\{(1 - h_{ii}) + g_{ii}\}^{1/2}},$$

em que $s_{(i)}^2$ é o quadrado médio residual do modelo irrestrito calculado sem a i -ésima observação, dado em (2.39). É possível mostrar que t_{v_i} segue uma distribuição t de Student com $(n - p + q)$ graus de liberdade. Esses resíduos poderão ser utilizados para destacar pontos aberrantes, detectar variância não constante e afastamentos da normalidade através da construção de envelopes, da mesma maneira que foi proposto para modelos irrestritos, apenas substituindo t_{S_i} por t_{v_i} .

Pontos de alavanca

Como vimos anteriormente o vetor de valores ajustados do modelo restrito é dado por $\hat{\mathbf{y}} = (\mathbf{H} - \mathbf{G})\mathbf{y}$. Assim, podemos considerar os elementos da diagonal principal da matriz $\mathbf{H} - \mathbf{G}$, dados por $h_{ii} - g_{ii}$, como sendo medidas para avaliar a influência de cada observação

no próprio valor ajustado de um modelo restrito de forma análoga ao caso irrestrito. Desde que $\text{traço}(\mathbf{H} - \mathbf{G}) = p - q$, então deve-se olhar com mais atenção aqueles pontos tais que $(h_{ii} - g_{ii}) \geq \frac{2(p-q)}{n}$.

Distância de Cook

Para obtermos a distância de Cook para modelos lineares restritos utilizamos a expressão para $\tilde{\boldsymbol{\beta}}_{(i)}$ dado em (2.35). Assim, quando o i -ésimo ponto é excluído, obtemos

$$\begin{aligned}
 D_{(i)}^c &= \frac{(\tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}_{(i)})^T (\mathbf{X}^T \mathbf{X}) (\tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}_{(i)})}{(p-q)s_c^2} \\
 &= \frac{[\mathbf{X}(\tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}_{(i)})]^T \mathbf{X}(\tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}_{(i)})}{(p-q)s_c^2} \\
 &= \frac{v_i}{1 - h_{ii} + g_{ii}} (\mathbf{x}_i^T - \mathbf{a}_i^T) (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \frac{v_i}{1 - h_{ii} + g_{ii}} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{x}_i - \mathbf{a}_i) \frac{1}{(p-q)s_c^2} \\
 &= \left(\frac{v_i}{1 - h_{ii} + g_{ii}} \right)^2 \frac{(\mathbf{x}_i^T - \mathbf{a}_i^T) (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{x}_i - \mathbf{a}_i)}{(p-q)s_c^2} \\
 &= \left(\frac{v_i}{1 - h_{ii} + g_{ii}} \right)^2 \frac{1}{(p-q)s_c^2} \left\{ \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i - \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{a}_i \right\} \\
 &= \frac{v_i^2}{(p-q)s_c^2} \frac{(h_{ii} - g_{ii})}{(1 - h_{ii} + g_{ii})^2}, \tag{2.41}
 \end{aligned}$$

uma vez que o i -ésimo elemento da diagonal principal da matriz \mathbf{G} pode também ser expresso por $g_{ii} = \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{a}_i = \mathbf{a}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i = \mathbf{a}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{a}_i$. O valor de s_c^2 poderá ser substituído por s^2 , uma vez que ambas são estimativas consistentes de σ^2 . Chamando

$$t_{c_i} = \frac{v_i}{s_c \sqrt{1 - h_{ii} + g_{ii}}},$$

podemos reexpressar (2.41) na forma

$$D_{(i)}^c = \frac{t_{c_i}^2}{(p-q)} \frac{(h_{ii} - g_{ii})}{(1 - h_{ii} + g_{ii})}.$$

Similarmente ao caso irrestrito, $D_{(i)}^c$ será grande se t_{c_i} e/ou $h_{ii} - g_{ii}$ forem grandes. Recomenda-se os gráficos de $h_{ii} - g_{ii}$, $D_{(i)}^c$ e t_{c_i} contra a ordem das observações.

Na Tabela 2.1 apresentamos um resumo com algumas medidas de diagnóstico encontradas para modelos normais irrestritos e restritos. Nota-se uma semelhança na forma das expressões, diferenciado apenas pelo número de graus de liberdade (há um ganho de q graus de liberdade) e pela presença da matriz \mathbf{G} e dos elementos de sua diagonal principal g_{ii} , no caso restrito.

Tabela 2.1 - Medidas de diagnóstico para modelos lineares irrestritos e restritos na forma $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$

| Medida | caso irrestrito | caso restrito |
|---------------------------|---|---|
| valor ajustado | $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$ | $\tilde{\mathbf{y}} = (\mathbf{H} - \mathbf{G})\mathbf{y}$ |
| resíduo ordinário | $\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{y}$ | $\mathbf{v} = (\mathbf{I} - \mathbf{H} + \mathbf{G})\mathbf{y}$. |
| resíduo studentizado int. | $t_i = \frac{e_i}{s(1 - h_{ii})^{1/2}}$ | $t_{c_i} = \frac{v_i}{s_c(1 - h_{ii} + g_{ii})^{1/2}}$, |
| resíduo studentizado ext. | $t_{S_i} = \frac{e_i}{s_{(i)}(1 - h_{ii})^{1/2}}$ | $t_{v_i} = \frac{v_i}{s_{(i)}(1 - h_{ii} + g_{ii})^{1/2}}$, |
| alavanca, (corte) | $h_{ii}, \left(\frac{2p}{n}\right)$ | $h_{ii} - g_{ii}, \left(\frac{2(p - q)}{n}\right)$ |
| distância Cook | $D_{(i)} = \frac{t_i^2}{p} \frac{h_{ii}}{(1 - h_{ii})}$ | $D_{(i)}^c = \frac{t_{c_i}^2}{(p - q)} \frac{(h_{ii} - g_{ii})}{(1 - h_{ii} + g_{ii})}$ |
| resíduo PRESS | $e_{(i)} = \frac{e_i}{1 - h_{ii}}$ | $\tilde{e}_{(i)} = \frac{v_i}{1 - h_{ii} + g_{ii}}$ |

2.4.3 Extensões

Sem perda de generalidade podemos estender os resultados desse capítulo para modelos de regressão linear em que a estrutura polinomial aparece na forma de covariáveis. Ou seja, podemos considerar o modelo

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon},$$

em que \mathbf{X}_1 é uma matriz $n \times p_1$ com valores de variáveis explicativas, por exemplo, a representação de grupos e \mathbf{X}_2 é uma matriz $n \times (p - p_1)$ com a representação polinomial. Assim, chamando $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ e $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$, todos os resultados desenvolvidos nas seções anteriores são estendidos para esse caso supondo ou não restrições do tipo $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$. Uma aplicação seria a comparação de f grupos controlando-se os efeitos através de alguma covariável, por exemplo, o tempo. O modelo poderia ser expresso na forma

$$y_{ij} = \alpha_i + P(x_i, \boldsymbol{\gamma}) + \epsilon_{ij},$$

em que $i = 1, \dots, f$ e $j = 1, \dots, n_i$, α_i denota o efeito do i -ésimo grupo, $P(x_i, \boldsymbol{\gamma})$ o polinômio correspondente e $\boldsymbol{\gamma} = (\gamma_0, \dots, \gamma_s)$ denota os coeficientes polinomiais do vetor $x_i = (1, x_{i1}, x_{i2}^2, \dots, x_{is}^s)^T$.

Capítulo 3

Modelos Lineares Generalizados Polinomiais Segmentados

Neste capítulo estendemos os resultados obtidos no capítulo 2 para modelos lineares generalizados (MLG's). Nesse caso, o preditor linear que compõe a parte sistemática de um MLG poderá ter a forma de uma *spline* polinomial segmentada, e estar sujeito à imposição de restrições sobre os parâmetros, sendo assim, uma opção de modelagem. Algumas técnicas de diagnóstico são também sugeridas.

3.1 Modelos lineares generalizados

Sejam y_1, \dots, y_n observações realizadas de um conjunto de variáveis aleatórias independentes, cada qual tendo função de probabilidade (caso discreto) ou função densidade de probabilidade (caso contínuo) pertencente à família exponencial, cuja forma é dada por

$$f_{Y_i}(y_i; \theta_i) = \exp[\phi\{y_i\theta_i - b(\theta_i)\} + c(y_i, \phi)], \quad (3.1)$$

em que $b(\cdot)$ e $c(\cdot)$ são funções especificadas, θ_i é o parâmetro desconhecido que pode variar para cada distribuição e $\phi^{-1} > 0$ é o parâmetro de dispersão geralmente conhecido e considerado constante em todas as distribuições.

Simultaneamente, temos que as variáveis Y_1, \dots, Y_n têm distribuição conjunta dada por

$$f_{Y_1, \dots, Y_n}(y_1, \dots, y_n; \theta_1, \dots, \theta_n) = \prod_{i=1}^n f_{Y_i}(y_i; \theta_i) = f(\mathbf{y}; \boldsymbol{\theta}), \quad (3.2)$$

em que $\mathbf{y} = (y_1, \dots, y_n)^T$ e $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)^T$. A função (3.2) pode ser vista como função dos parâmetros $\theta_1, \dots, \theta_n$ com $\mathbf{y} = (y_1, \dots, y_n)^T$ considerado dado, e será denotada por $L(\boldsymbol{\theta}; \mathbf{y})$. Definida dessa forma, $L(\boldsymbol{\theta}; \mathbf{y})$ é chamada de função de verossimilhança, sendo o logaritmo dessa função dado por

$$\log L(\boldsymbol{\theta}; \mathbf{y}) = l(\boldsymbol{\theta}; \mathbf{y}). \quad (3.3)$$

A derivada primeira de (3.3) em relação a θ_i , $U_i = dl/d\theta_i$, $i = 1, \dots, n$, é denominada função score, sendo que seus primeiros momentos apresentam as seguintes propriedades:

$$E(U_i) = 0$$

e

$$\text{Var}(U_i) = E(U_i^2) = E(-U_i'),$$

em que $U_i' = dU_i/d\theta_i$, sendo que $\text{Var}(U_i)$ é chamada de informação de Fisher de θ_i . Dos momentos acima obtemos

$$E(Y_i) = \mu_i = b'(\theta_i) = db(\theta_i)/d\theta_i$$

e

$$\text{Var}(Y_i) = b''(\theta_i) = d^2b(\theta_i)/d\theta_i^2 = \phi^{-1}V_i,$$

em que $V_i = d\mu_i/d\theta_i$ é denominada função de variância e caracteriza a classe de distribuições da qual $f_{Y_i}(\cdot)$ pertence.

Como os parâmetros θ_i 's podem ser diferentes para cada observação, não há um interesse direto sobre os mesmos na especificação do modelo. Para esse fim, utilizamos um conjunto menor de parâmetros β_1, \dots, β_p , $p < n$, associado a variáveis explicativas, tal que uma combinação linear desses elementos seja equivalente a uma função do valor esperado de Y_i , ou seja,

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} = \eta_i, \quad (3.4)$$

em que $g(\cdot)$ é uma função monótona e diferenciável chamada função de ligação, $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$ denota um vetor $p \times 1$ de valores de variáveis explicativas associadas a cada valor observado y_i , $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ é um vetor $p \times 1$ de parâmetros desconhecidos a serem estimados e $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ é o preditor linear. Portanto, temos que os modelos lineares generalizados são definidos pelos seguintes componentes (ver, por exemplo, Cordeiro, 1986; McCullagh e Nelder, 1989):

- 1) variáveis respostas Y_1, \dots, Y_n independentes, cada uma com função de probabilidade ou densidade dada por (3.1);
- 2) uma função de ligação na forma (3.4), que relaciona a média da variável resposta aos parâmetros β_1, \dots, β_p através do preditor linear.

Matricialmente temos que os MLGs são definidos por

$$\begin{aligned} i) \quad & E(\mathbf{Y}) = \boldsymbol{\mu} \quad e \\ ii) \quad & \mathbf{g}(\boldsymbol{\mu}) = \mathbf{X}^T \boldsymbol{\beta} = \boldsymbol{\eta}, \end{aligned}$$

em que $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ é o vetor $n \times 1$ composto por variáveis aleatórias independentes cuja função de distribuição de cada variável é dada por (3.1), $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ é o vetor

$n \times 1$ de médias, $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^T$ e \mathbf{X} é a matriz modelo $n \times p$ definida por,

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix},$$

sendo que cada linha \mathbf{x}_i^T é como definida acima e cada coluna representa os valores de uma variável explicativa para os n casos.

Para o nosso interesse o preditor linear η poderá ter a seguinte forma:

$$\begin{aligned} \eta = & \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \dots + \beta_{0r}x^r + \\ & + \beta_{10}(x - t_1)_+^0 + \beta_{11}(x - t_1)_+^1 + \beta_{12}(x - t_1)_+^2 + \dots + \beta_{1r}(x - t_1)_+^r + \dots + \\ & + \beta_{k0}(x - t_k)_+^0 + \beta_{k1}(x - t_k)_+^1 + \beta_{k2}(x - t_k)_+^2 + \dots + \beta_{kr}(x - t_k)_+^r, \end{aligned}$$

ou seja, η poderá representar uma *spline* polinomial segmentada e estar sujeito à imposição de restrições lineares. Discutiremos este tópico na seção 3.2.2.

3.2 Estimação

3.2.1 Estimação irrestrita por máxima verossimilhança

A estimação dos parâmetros do modelo pelo método de máxima verossimilhança considera a função de verossimilhança para respostas independentes dada por

$$L(\theta_1, \dots, \theta_n; y_1, \dots, y_n) = \exp \left[\phi \left\{ \sum y_i \theta_i - \sum b(\theta_i) \right\} + \sum c(y_i, \phi) \right],$$

que é algebricamente igual à função de distribuição conjunta de Y_1, \dots, Y_n , dada por (3.2).

A estimativa de máxima verossimilhança de $\boldsymbol{\theta}$ é o valor que maximiza a função $l(\boldsymbol{\theta})$ dada por (3.3), ou a função $L(\cdot)$ acima, tendo a condição de que a segunda derivada de $l(\boldsymbol{\theta})$ seja uma matriz positiva definida.

Como as distribuições da família exponencial têm a propriedade de satisfazerem as condições de regularidade, o máximo global da função $l(\boldsymbol{\theta})$ é único e dado pelas soluções das equações

$$\frac{\partial l}{\partial \theta_i} = 0, \quad i = 1, \dots, n.$$

No entanto, como o interesse principal está em estimarmos os parâmetros β_1, \dots, β_p , podemos, equivalentemente, resolver as p equações

$$\frac{\partial l}{\partial \beta_j} = 0, \quad j = 1, \dots, p.$$

Em geral, podemos escrever

$$\begin{aligned} \frac{\partial l}{\partial \beta_j} &= U_{\beta_j} \\ &= \sum_{i=1}^n \frac{(y_i - \mu_i)}{\text{Var}(Y_i)} x_{ij} \frac{\partial \mu_i}{\partial \eta_i}, \quad j = 1, \dots, p, \end{aligned}$$

em que x_{ij} é o j -ésimo elemento de \mathbf{x}_i^T . Quando a ligação é canônica, isto é, quando $\theta_i = \eta_i$, a solução dessas equações é única.

Geralmente as equações $U_{\beta_j} = 0$, são não-lineares em $\boldsymbol{\beta}$ e técnicas iterativas são necessárias para obter-se uma estimativa para $\boldsymbol{\beta}$. O método de Newton-Raphson é uma opção a ser utilizada, em que a estimativa irrestrita $\hat{\boldsymbol{\beta}}$, é obtida através de um processo iterativo dado por

$$\begin{aligned} \boldsymbol{\beta}^{(m+1)} &= \boldsymbol{\beta}^{(m)} - \left[\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right]_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(m)}}^{-1} \mathbf{U}^{(m)} \\ &= \boldsymbol{\beta}^{(m)} - \left[\mathbf{U}'(\boldsymbol{\beta}^{(m)}) \right]^{-1} \mathbf{U}^{(m)}, \end{aligned}$$

sendo que $\mathbf{U}^{(m)}$ é a função score $\mathbf{U} = (U_{\beta_1}, \dots, U_{\beta_p})^T$ calculada no m -ésimo passo da estimativa de $\boldsymbol{\beta}^{(m)}$ e \mathbf{U}' é a matriz de segundas derivadas de $l(\boldsymbol{\theta})$. Como a matriz de segundas derivadas pode não ser positiva definida, podemos substituí-la pela matriz de

valores esperados $E\left(\frac{-\partial^2 l}{\partial \beta \partial \beta^T}\right)$ e assim obter as estimativas de β pelo método de *scoring* de Fisher, dado por

$$\beta^{(m+1)} = \beta^{(m)} + (\mathbf{K}^{(m)})^{-1} \mathbf{U}^{(m)}, \quad (3.5)$$

em que $\mathbf{K}^{(m)}$ é a matriz de informação de Fisher calculada em $\beta^{(m)}$.

Matricialmente (ver, por exemplo, Paula, 2004), a função escore e a matriz de informação de Fisher para β são dadas, respectivamente, por

$$\mathbf{U}(\beta) = \phi \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})$$

e

$$\mathbf{K} = \phi \mathbf{X}^T \mathbf{W} \mathbf{X},$$

em que $\mathbf{V} = \text{diag}\{V_1, \dots, V_n\}$ e $\mathbf{W} = \text{diag}\{w_1, \dots, w_n\}$ é a matriz com pesos $w_i = (d\mu_i/d\eta_i)^2/V_i$. Assim, substituindo-se as expressões acima em (3.5) e trabalhando-se o lado direito da mesma expressão, temos que a $(m+1)$ -ésima aproximação da estimativa irrestrita de β , é dada por

$$\beta^{(m+1)} = (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{z}^{(m)},$$

em que $\mathbf{z}^{(m)} = \mathbf{X} \beta^{(m)} + \mathbf{W}^{(m)-1/2} \mathbf{V}^{(m)-1/2} (\mathbf{y} - \boldsymbol{\mu}^{(m)})$. Essa expressão pode ser vista como uma representação de equações normais de um modelo linear obtido por mínimos quadrados ponderados a serem obtidos iterativamente, uma vez que \mathbf{z} e \mathbf{W} dependem do valor de β que muda a cada passo do processo iterativo. Nesse caso, teríamos que \mathbf{z} é a variável dependente modificada e \mathbf{X} é a matriz modelo, ponderados pela matriz \mathbf{W} . Esse procedimento é repetido até a convergência, ou seja, até que $\beta^{(m+1)}$ esteja muito próximo de $\beta^{(m)}$ o que implica que as estimativas de máxima verossimilhança de um MLG são obtidas por mínimos quadrados ponderados. A convergência do processo iterativo acima ocorre em um número finito de passos, independente dos valores utilizados inicialmente. Sob condições de regularidade suaves, Fahrmeir e Kaufmann(1985) mostram que o estimador

irrestrito $\hat{\beta}$ é consistente e assintoticamente normal com matriz de variâncias-covariâncias dada por

$$\text{Var}(\hat{\beta}) = \phi^{-1}(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}.$$

Sendo a regressão normal um caso particular de um MLG, em que $\mathbf{W} = \mathbf{I}_n$, temos que $\text{Var}(\hat{\beta}) = \phi^{-1}(\mathbf{X}^T \mathbf{X})^{-1}$, que coincide com (2.10).

Quanto à estimação do parâmetro de dispersão (ϕ^{-1}), vale lembrar que β e ϕ são ortogonais, o que leva à independência assintótica entre seus estimadores (ver, por exemplo, Paula, 2004).

Para as distribuições normal e normal inversa, a estimativa de máxima verossimilhança para ϕ é dada por

$$\hat{\phi} = \frac{n}{D(\mathbf{y}; \hat{\boldsymbol{\mu}})},$$

em que $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ é o desvio do modelo sob investigação definido por

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \{y_i(\hat{\theta}_i^0 - \hat{\theta}_i) + b(\hat{\theta}_i) - b(\hat{\theta}_i^0)\} \quad (3.6)$$

sendo que $\hat{\theta}_i^0 = \theta(y_i)$ e $\hat{\theta}_i = \theta(\hat{\mu}_i)$ indicam, respectivamente, as estimativas de máxima verossimilhança de θ_i sob o modelo saturado (n parâmetros) e com p parâmetros ($p < n$).

Para a distribuição gama, obtém-se a estimativa de máxima verossimilhança através da solução da equação

$$2n\{\log \hat{\phi} - \psi(\hat{\phi})\} = D(\mathbf{y}; \hat{\boldsymbol{\mu}})$$

em que $\psi(\phi) = \Gamma'(\phi)/\Gamma(\phi)$ é a função digama. Essa estimativa e o respectivo desvio padrão podem ser obtidos diretamente do S-Plus através da "biblioteca MASS", especificamente através do comando

```
gamma.shape(ajuste)
```

em que "ajuste" representa os resultados de uma regressão realizada através da função `glm()`. A estimativa obtida diretamente do ajuste `glm()` é dado por

$$\hat{\phi}^{-1} = D(\mathbf{y}; \hat{\boldsymbol{\mu}})/(n - p)$$

e não é consistente para ϕ . Outras estimativas podem ser encontradas em Cordeiro e McCullagh(1991) e Paula (2004).

3.2.2 Estimação restrita

Há situações em que existe interesse em se impor restrições lineares sobre os parâmetros por questões de modelagem ou mesmo por interesses inferenciais em se testar hipóteses. Neste trabalho em especial, as restrições lineares serão utilizadas para controlar os graus das equações polinomiais formadoras da *spline* segmentada que constituirão o próprio preditor linear η . O uso dessas *splines* para o preditor linear é indicado se, ao realizarmos o gráfico da função de ligação na variável resposta contra a covariável, $g(y) \times x$, obtivermos um padrão que sugere segmentação como na Figura 1.1. Ou seja, ao identificarmos essa situação, poderemos impor restrições lineares sobre os parâmetros β_j 's, de forma a garantir o grau desejado das equações que o compõem, semelhante ao que foi realizado para o caso linear. Assim, as q restrições lineares da forma $C_j^T \beta = d_j$ compõem uma equação matricial da forma

$$C\beta = \mathbf{d},$$

sendo que C é uma matriz $q \times p$ composta pelos coeficientes das restrições, β é o vetor $p \times 1$ de parâmetros desconhecidos, e \mathbf{d} é o vetor solução de dimensão $q \times 1$.

3.2.3 Estimação restrita por máxima verossimilhança

O processo de estimação de β em que a maximização do logaritmo da função de verossimilhança está sujeito a $C\beta = \mathbf{d}$, consiste em um problema de maximização restrita. Nyquist(1991) propõe um processo iterativo para obter essas estimativas restritas de β para MLG's utilizando o método da função penalizada (vide, por exemplo, Ryan, 1974). Primeiramente, obtém-se uma estimativa irrestrita $\hat{\beta}(\lambda)$ através da função penalizada quadrática seguida por um processo iterativo. Então, a estimativa restrita $\tilde{\beta}$ é obtida por

intermédio de uma relação existente entre essa estimativa e a estimativa irrestrita $\hat{\beta}(\lambda)$. A seguir, apresentaremos alguns dos resultados principais desse método de estimação, sendo que os detalhes encontram-se em Cysneiros (1997).

Considere a função penalizada quadrática

$$P(\beta, \lambda) = \sum_{i=1}^n \phi \{y_i \theta_i - b(\theta_i)\} + \sum_{i=1}^n c(y_i, \phi) - \frac{1}{2} \sum_{j=1}^k \lambda_j (d_j - C_j^T \beta)^2.$$

Para cada λ_j fixo e positivo, $j = 1, \dots, q$, obtemos a solução $\hat{\beta}(\lambda)$, maximizando-se a função $P(\beta, \lambda)$ em relação a β , em que

$$\frac{\partial P(\beta, \lambda)}{\partial \beta} = Q(\beta, \lambda) = \phi X^T W^{1/2} V^{-1/2} (y - \mu) + C^T \Lambda (d - C\beta), \quad (3.7)$$

sendo Λ a matriz diagonal com elementos λ_j , $j = 1, \dots, q$. Portanto, o cálculo de $\hat{\beta}(\lambda)$ se resume a um simples problema de estimação irrestrita, sendo necessário o uso de métodos iterativos para obtê-la, o que nos permite seguir o mesmo raciocínio desenvolvido anteriormente.

Aplicando-se o método *scoring* de Fisher chega-se a

$$\beta(\lambda)^{(m+1)} = \beta(\lambda)^{(m)} + S(\beta^{(m)}, \lambda)^{-1} Q(\beta^{(m)}, \lambda), \quad (3.8)$$

sendo $Q(\beta^{(m)}, \lambda)$ o vetor $p \times 1$ com os valores da função score para β dado por (3.7) e $S(\beta^{(m)}, \lambda)$ a matriz de informação de Fisher de dimensão $p \times p$ dada por

$$S(\beta^{(m)}, \lambda) = X^T W^{(m)} X + C^T \Lambda C.$$

Após trabalharmos a expressão (3.8) e substituirmos as formas matriciais, chegamos ao processo iterativo

$$\begin{aligned} \beta(\lambda)^{(m+1)} &= (X^T W^{(m)} X)^{-1} X^T W^{(m)} z^{(m)} + (X^T W^{(m)} X)^{-1} C^T \\ &\quad \times \{ \Lambda^{-1} + C (X^T W^{(m)} X)^{-1} C^T \}^{-1} \{ d - C (X^T W^{(m)} X)^{-1} X^T W^{(m)} z^{(m)} \}. \end{aligned} \quad (3.9)$$

A convergência do processo iterativo (3.9) leva à solução $\hat{\beta}(\lambda)$. A estimativa restrita por igualdades lineares é então dada por

$$\tilde{\beta} = \lim_{\lambda_1, \dots, \lambda_q \rightarrow \infty} \hat{\beta}(\lambda).$$

Assim, utilizando-se o resultado obtido em (3.9) temos que a $(m+1)$ -ésima aproximação da estimativa de máxima verossimilhança restrita de β , é dada por

$$\begin{aligned} \tilde{\beta}^{(m+1)} &= \lim_{\lambda_1, \dots, \lambda_q \rightarrow \infty} \hat{\beta}(\lambda)^{(m+1)} \\ &= (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{z}^{(m)} + (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{C}^T \{ \mathbf{C} (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{C}^T \}^{-1} \times \\ &\quad \times \{ \mathbf{d} - \mathbf{C} (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{z}^{(m)} \} \\ &= \tilde{\beta}^{*(m+1)} + (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{C}^T \{ \mathbf{C} (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{C}^T \}^{-1} (\mathbf{d} - \mathbf{C} \tilde{\beta}^{*(m+1)}) \\ &= \tilde{\beta}^{*(m+1)} - (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{C}^T \{ \mathbf{C} (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{C}^T \}^{-1} (\mathbf{C} \tilde{\beta}^{*(m+1)} - \mathbf{d}) \end{aligned}$$

em que $\tilde{\beta}^{*(m+1)}$ assume a mesma forma da estimativa irrestrita, embora seja avaliada na estimativa restrita de m -ésimo passo. Portanto, $\tilde{\beta}^{(m+1)}$ pode ser considerada como uma estimativa de mínimos quadrados reponderados dada por $\tilde{\beta}^{*(m+1)}$ mais um termo de correção.

Supondo satisfeitas as condições de regularidade sob as restrições $\mathbf{C}\beta = \mathbf{d}$, temos por Fahrmeir e Kaufmann(1994) que $\tilde{\beta}$ é consistente e assintoticamente normal com matriz de variâncias-covariâncias dada por

$$\begin{aligned} \text{Var}(\tilde{\beta}) &= \lim_{\lambda_1, \dots, \lambda_q \rightarrow \infty} (\mathbf{X}^T \mathbf{W} \mathbf{X} + \mathbf{C}^T \lambda \mathbf{C})^{-1} \\ &= \phi^{-1} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} [\mathbf{I}_p - \mathbf{C}^T \{ \mathbf{C} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{C}^T \}^{-1} \mathbf{C} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}]. \end{aligned}$$

Para o caso normal em que $\mathbf{W} = \mathbf{I}$, obtemos

$$\text{Var}(\tilde{\beta}) = \phi^{-1} (\mathbf{X}^T \mathbf{X})^{-1} [\mathbf{I}_p - \mathbf{C}^T \{ \mathbf{C} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T \}^{-1} \mathbf{C} (\mathbf{X}^T \mathbf{X})^{-1}],$$

que coincide com (2.23), uma vez que $\phi^{-1} = \sigma^2$. Para obter a variância estimada, basta substituir ϕ por um estimador consistente.

Em Cysneiros (1997) há o desenvolvimento do algoritmo para se ajustar um modelo linear generalizado restrito na linguagem S-Plus e assim obter a estimativa restrita $\tilde{\beta}$ como também outros resultados associados, da mesma forma que o disponível pela função `glm()` do S-Plus.

3.3 Avaliação e seleção de modelos

3.3.1 Avaliação e seleção de MLGs irrestritos

Função desvio

O processo de ajuste de um modelo aos dados pode ser considerado como um meio de substituir um conjunto de valores observados de \mathbf{y} por um conjunto de valores ajustados $\hat{\boldsymbol{\mu}}$ derivados de um modelo envolvendo usualmente um número relativamente pequeno de parâmetros (McCullagh e Nelder, 1989). Cabe então medirmos a discrepância existente entre os y 's originais e aqueles ajustados por meio de $\hat{\boldsymbol{\mu}}$'s. Uma das formas de medir essa discrepância, ou seja, de avaliar a qualidade do ajuste, é comparar a verossimilhança do modelo de interesse ($p < n$ parâmetros) com a verossimilhança do modelo saturado com $p = n$ parâmetros, ou seja, o número máximo possível considerado como uma perfeita descrição dos dados.

Considere $l(\mathbf{y}; \mathbf{y})$ e $l(\hat{\boldsymbol{\beta}}; \mathbf{y})$ como os logaritmos das funções de verossimilhança calculadas nos respectivos máximos do modelo saturado e do modelo com p parâmetros. Caso o modelo de interesse descreva bem os dados, espera-se que $l(\mathbf{y}; \mathbf{y})$ esteja próximo de $l(\hat{\boldsymbol{\beta}}; \mathbf{y})$, ou seja, que a diferença entre eles seja pequena.

Nelder e Wedderburn (1972) propuseram como medida de qualidade do ajuste a função desvio escalonado, dada por

$$D_p^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \phi D_p = 2[l(\mathbf{y}; \mathbf{y}) - l(\hat{\boldsymbol{\beta}}; \mathbf{y})],$$

cuja quantidade $D_p = D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ representa a função desvio definida em (3.6). Um valor pequeno para a função desvio escalonado indica que para um número menor de parâmetros obtém-se um ajuste tão bom quanto o ajuste do modelo saturado. Geralmente compara-se o valor obtido em D_p^* com os percentis de uma distribuição qui-quadrado com $n - p$ graus de liberdade embora D_p^* tenha distribuição assintótica qui-quadrado, sob o modelo, apenas para algumas situações. Por exemplo, D_p^* segue exatamente uma $\chi^2_{(n-p)}$ no caso normal linear. Quando a resposta é binomial tem-se distribuição qui-quadrado apenas para dados agrupados. Para resposta de Poisson o desvio D_p^* segue assintoticamente uma $\chi^2_{(n-p)}$ quando as médias são grandes e para resposta gama isso ocorre se o parâmetro de dispersão ϕ^{-1} for pequeno (ou se ϕ for grande).

Critério de Akaike

Um critério bastante utilizado para selecionar modelos é o critério de Akaike (1974). Neste caso, não há um teste estatístico envolvido e o interesse está em se obter um equilíbrio entre um modelo parcimonioso (com poucos parâmetros) e um ajuste adequado aos dados (que melhora quando aumentamos o número de parâmetros e que é absorvido diretamente pelo logaritmo da função de verossimilhança, $l(\hat{\boldsymbol{\beta}}, \mathbf{y})$, e conseqüentemente por D_p^*). O valor AIC para MLG's é definido por

$$\text{AIC} = D_p^* + 2p,$$

em que D_p^* denota o desvio escalonado do modelo e p é o número de parâmetros do modelo sob investigação. O melhor modelo, segundo este critério, é aquele que minimiza a função AIC.

Para o caso normal, assumindo σ^2 desconhecido, o valor AIC é dado por

$$\text{AIC} = n \log \left\{ \frac{\sum_{i=1}^n (y_i - \hat{\mu}_i)^2}{n} \right\} + 2p.$$

3.3.2 Avaliação e seleção de MLGs restritos

Função desvio

Para o caso restrito, temos que a função desvio escalonada é dada por

$$D_p^*(\mathbf{y}; \tilde{\boldsymbol{\mu}}) = \phi D_p = 2[l(\mathbf{y}; \mathbf{y}) - l(\tilde{\boldsymbol{\beta}}; \mathbf{y})], \quad (3.10)$$

com

$$D_p = D_p(\mathbf{y}; \tilde{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \{y_i(\hat{\theta}_i^0 - \tilde{\theta}_i) + b(\tilde{\theta}_i) - b(\hat{\theta}_i^0)\}$$

representando a função desvio com $\hat{\theta}_i^0 = \theta(y_i)$ e $\tilde{\theta}_i = \theta(\tilde{\mu}_i)$ indicando, respectivamente, as estimativas de máxima verossimilhança de θ_i sob o modelo saturado e sob o modelo restrito com p parâmetros ($p < n$) sujeitos a $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$. Neste caso, compara-se o valor obtido em D_p^* com os percentis de uma distribuição qui-quadrado com $n - (p - q)$ graus de liberdade. Ou seja, a defasagem para $p - q$ graus de liberdade é devido ao fato de estarmos num espaço paramétrico $(p - q)$ -dimensional quando impomos a restrição $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$, desde que \mathbf{C} seja de posto completo q .

Critério de Akaike

Em MLGs restritos, o valor AIC é dado por

$$\text{AIC} = D_p^* + 2(p - q),$$

em que D_p^* denota o desvio escalonado do modelo restrito, p é o número de parâmetros do modelo restrito sob investigação e q é o número de restrições lineares. O melhor modelo, segundo este critério, é aquele que minimiza a função AIC.

Para o caso normal restrito, com σ^2 desconhecido, o valor AIC é dado por

$$\text{AIC} = n \log \left\{ \frac{\sum_{i=1}^n (y_i - \tilde{\mu}_i)^2}{n} \right\} + 2(p - q).$$

3.4 Técnicas de diagnóstico

Selecionado um modelo, o próximo passo consiste em avaliar o seu ajuste, verificando se as suposições assumidas são válidas ou pelo menos se não existem afastamentos sérios das mesmas. Também é interessante detectar pontos aberrantes e/ou influentes do conjunto de dados. Para tanto, utilizamos aproximações que são obtidas para grandes amostras. Apesar de não serem exatos, os resultados encontrados são suficientes para efeitos de diagnóstico, no sentido de destacar os pontos que se diferem demasiadamente dos demais por determinados critérios.

3.4.1 Diagnóstico em MLGs irrestritos

Resíduos

Em MLG's, podemos definir resíduo padronizado da mesma forma que em modelos lineares, porém as principais propriedades não estarão presentes.

Considerando a convergência do processo iterativo de estimação irrestrita dos parâmetros β , chega-se a

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

em que $\hat{\mathbf{z}} = \hat{\boldsymbol{\eta}} + \hat{\mathbf{W}}^{-1/2} \hat{\mathbf{V}}^{-1/2} (\mathbf{y} - \hat{\boldsymbol{\mu}})$.

Tomando-se $\hat{\mathbf{W}}^{1/2} \hat{\mathbf{z}} = \mathbf{z}^*$ e $\hat{\mathbf{W}}^{1/2} \mathbf{X} = \mathbf{X}^*$, temos que

$$\hat{\beta} = (\mathbf{X}^{*T} \mathbf{X}^*)^{-1} \mathbf{X}^{*T} \mathbf{z}^*,$$

ou seja, $\hat{\beta}$ pode ser considerado como uma solução de mínimos quadrados de uma regressão linear de \mathbf{z}^* contra as colunas de \mathbf{X}^* . Nessa regressão fictícia, o valor ajustado é dado por

$$\begin{aligned} \hat{\mathbf{z}}^* &= \mathbf{X}^* \hat{\beta} \\ &= \hat{\mathbf{W}}^{1/2} \mathbf{X} (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{W}} \hat{\mathbf{z}} \\ &= \hat{\mathbf{H}}^* \hat{\mathbf{z}}, \end{aligned}$$

em que $\hat{\mathbf{H}}^* = \hat{\mathbf{W}}^{1/2} \mathbf{X} (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{W}}^{1/2}$. Assim, o resíduo ordinário pode ser expresso por

$$\begin{aligned} \mathbf{e}^* &= \mathbf{z}^* - \hat{\mathbf{z}}^* \\ &= \mathbf{z}^* - \hat{\mathbf{H}}^* \mathbf{z}^* \\ &= (\mathbf{I}_n - \hat{\mathbf{H}}^*) \mathbf{z}^*. \end{aligned}$$

Alternativamente, podemos também escrever $\mathbf{e}^* = \hat{\mathbf{W}}^{1/2}(\hat{\mathbf{z}} - \hat{\boldsymbol{\eta}})$ ou $\mathbf{e}^* = \hat{\mathbf{V}}^{-1/2}(\mathbf{y} - \hat{\boldsymbol{\mu}})$, uma vez que

$$\begin{aligned} \mathbf{e}^* &= (\mathbf{I}_n - \hat{\mathbf{H}}^*) \mathbf{z}^* \\ &= \hat{\mathbf{W}}^{1/2} \hat{\mathbf{z}} - \hat{\mathbf{W}}^{1/2} \mathbf{X} (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{W}} \hat{\mathbf{z}} \\ &= \hat{\mathbf{W}}^{1/2} \hat{\mathbf{z}} - \hat{\mathbf{W}}^{1/2} \mathbf{X} \hat{\boldsymbol{\beta}} \\ &= \hat{\mathbf{W}}^{1/2} \hat{\mathbf{z}} - \hat{\mathbf{W}}^{1/2} \hat{\boldsymbol{\eta}} \\ &= \hat{\mathbf{W}}^{1/2} \{ \hat{\boldsymbol{\eta}} + \hat{\mathbf{W}}^{-1/2} \hat{\mathbf{V}}^{-1/2} (\mathbf{y} - \hat{\boldsymbol{\mu}}) \} - \hat{\mathbf{W}}^{1/2} \hat{\boldsymbol{\eta}} \\ &= \hat{\mathbf{V}}^{-1/2} (\mathbf{y} - \hat{\boldsymbol{\mu}}). \end{aligned}$$

A matriz de variâncias-covariâncias assintótica desse resíduo, com as quantidades avaliadas no parâmetro verdadeiro, é dada por

$$\begin{aligned} \text{Var}(\mathbf{e}^*) &= (\mathbf{I}_n - \mathbf{H}^*) \text{Var}(\mathbf{z}^*) (\mathbf{I}_n - \mathbf{H}^*)^T \\ &= (\mathbf{I}_n - \mathbf{H}^*) \text{Var}(\mathbf{W}^{1/2} \mathbf{z}) (\mathbf{I}_n - \mathbf{H}^*) \\ &= (\mathbf{I}_n - \mathbf{H}^*) \mathbf{W}^{1/2} \text{Var}(\mathbf{z}) \mathbf{W}^{1/2} (\mathbf{I}_n - \mathbf{H}^*) \\ &= (\mathbf{I}_n - \mathbf{H}^*) \mathbf{W}^{1/2} \boldsymbol{\phi}^{-1} \mathbf{W}^{-1} \mathbf{W}^{1/2} (\mathbf{I}_n - \mathbf{H}^*) \\ &= \boldsymbol{\phi}^{-1} (\mathbf{I}_n - \mathbf{H}^*), \end{aligned}$$

uma vez que a matriz $\mathbf{I} - \mathbf{H}^*$ é simétrica e idempotente (ver Apêndice A.2), e que

$$\begin{aligned}\text{Var}(\mathbf{z}) &= \text{Var}\{\boldsymbol{\eta} + \mathbf{W}^{-1/2}\mathbf{V}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})\} \\ &= \text{Var}\{\mathbf{W}^{-1/2}\mathbf{V}^{-1/2}\mathbf{y}\} \\ &= \mathbf{W}^{-1/2}\mathbf{V}^{-1/2}\text{Var}(\mathbf{y})\mathbf{V}^{-1/2}\mathbf{W}^{-1/2} \\ &= \mathbf{W}^{-1/2}\mathbf{V}^{-1/2}\phi^{-1}\mathbf{V}\mathbf{V}^{-1/2}\mathbf{W}^{-1/2} \\ &= \phi^{-1}\mathbf{W}^{-1}.\end{aligned}$$

Portanto, o resíduo padronizado fica, com as quantidades avaliadas em $\hat{\boldsymbol{\mu}}$, dado por

$$\begin{aligned}t_{S_i}^* &= \frac{\hat{V}_i^{-1/2}(y_i - \hat{\mu}_i)}{\sqrt{\phi^{-1}(1 - \hat{h}_{ii}^*)}} \\ &= \frac{\phi^{1/2}(y_i - \hat{\mu}_i)}{\sqrt{\hat{V}_i(1 - \hat{h}_{ii}^*)}}, \quad i = 1, \dots, n,\end{aligned}$$

em que \hat{h}_{ii}^* é o i -ésimo elemento da diagonal principal da matriz $\hat{\mathbf{H}}^*$. O resíduo padronizado $t_{S_i}^*$ não tem as mesmas propriedades que t_{S_i} , como no caso normal, uma vez que $\hat{\boldsymbol{\mu}}$ na prática não é fixo. Em geral $t_{S_i}^*$ tem distribuição empírica assimétrica dificultando a utilização de medidas usuais de diagnóstico.

Contudo, em MLGs, os resíduos mais utilizados se originam das contribuições de cada observação na função desvio dada em (3.10), sendo denominados componentes do desvio. São definidos por

$$d^*(y_i; \hat{\mu}_i) = \phi^{1/2}\text{sinal}(y_i - \hat{\mu}_i)d(y_i; \hat{\mu}_i),$$

em que $d(y_i; \hat{\mu}_i) = \sqrt{2}\{y_i(\hat{\theta}_i^0 - \hat{\theta}_i) + [b(\hat{\theta}_i) - b(\hat{\theta}_i^0)]\}^{1/2}$. A versão padronizada deste resíduo é dada por

$$\begin{aligned}t_{D_i} &= \frac{d^*(y_i; \hat{\mu}_i)}{\sqrt{1 - \hat{h}_{ii}^*}} \\ &= \frac{\phi^{1/2}d(y_i; \hat{\mu}_i)}{\sqrt{1 - \hat{h}_{ii}^*}}.\end{aligned}$$

Através de simulações, Williams (1984) verificou uma boa concordância entre a distribuição empírica de t_{D_i} e a normal padrão para vários MLG's.

O método de envelopes sugerido por Atkinson (1981) que consiste em gerar uma banda de confiança para a distribuição empírica do resíduo t_{S_i} do caso normal pode ser estendido para os resíduos $t_{S_i}^*$ e t_{D_i} .

Pontos de alavanca

Uma vez que o resíduo ordinário irrestrito de um MLG é dado por $\mathbf{e}^* = (\mathbf{I}_n - \hat{\mathbf{H}}^*)\hat{\mathbf{z}}^*$ e, fazendo uma analogia com o caso normal, observamos que $\hat{\mathbf{H}}^*$ pode desempenhar o papel de matriz de projeção ortogonal local. Assim, sugere-se usar os elementos da diagonal principal dessa matriz, \hat{h}_{ii}^* , como medidas de alavanca embora na realidade não o seja, uma vez que \hat{h}_{ii}^* em geral não coincide com $\frac{\partial \hat{y}_i}{\partial y_i}$ que é a definição de alavanca. Wei et al. (1998) mostram que a quantidade \hat{h}_{ii}^* coincide com a definição de alavanca em MLG's apenas nos casos de resposta contínua e ligação canônica. Portanto, não faz sentido calcular alavanca no caso de resposta discreta, mesmo assim, \hat{h}_{ii}^* tem sido utilizado para detectar pontos de alavanca no modelo linear equivalente ($\hat{\mathbf{z}}^*$ contra as colunas de \mathbf{X}^*). Em geral, adota-se como ponto de corte o valor $\frac{2p}{n}$. Porém há casos em que esse valor não é válido, como na regressão logística em que o ponto de corte de \hat{h}_{ii}^* depende da probabilidade ajustada (vide, por exemplo, Hosmer e Lemeshow, 1989).

O valor de \hat{h}_{ii}^* é expresso como

$$\hat{h}_{ii}^* = \hat{w}_i \mathbf{x}_i^T (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{x}_i,$$

em que $w_i = (d\mu_i/d\eta_i)^2/V_i$. No caso da regressão logística temos que $w_i = n_i \mu_i (1 - \mu_i)$, fazendo com que \hat{h}_{ii}^* dependa das probabilidades ajustadas.

Influência

Uma medida de influência de uma observação num MLG considera o impacto no logaritmo da função de verossimilhança $l(\hat{\beta})$, com a retirada da i -ésima observação. Tal medida baseia-se na região de confiança assintótica de coeficiente $(1 - \alpha)$ para β , dada por

$$\{\beta : 2[l(\hat{\beta}) - l(\beta)] \leq \chi_p^2(1 - \alpha)\}.$$

Essa medida é denominada afastamento da verossimilhança (*likelihood displacement*) e foi apresentada por Cook e Weisberg (1982, 1987), sendo definida por

$$LD_i = 2\{l(\hat{\beta}) - l(\hat{\beta}_{(i)})\}.$$

Não sendo possível obter uma forma analítica para LD_i , é usual utilizar a segunda aproximação por série de Taylor em torno de $\hat{\beta}$. Essa expansão leva ao seguinte:

$$LD_i \cong (\beta - \hat{\beta})^T \{-L''(\hat{\beta})\}(\beta - \hat{\beta}),$$

em que $-L''(\beta)$ denota a matriz observada de Fisher. Substituindo essa matriz pelo correspondente valor esperado e β por $\hat{\beta}_{(i)}$, obtém-se

$$LD_i \cong \phi (\hat{\beta} - \hat{\beta}_{(i)})^T (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X}) (\hat{\beta} - \hat{\beta}_{(i)}).$$

Como em geral não é possível obter uma forma fechada para $\hat{\beta}_{(i)}$, usamos a aproximação de um passo obtida por Pregibon (1981),

$$\hat{\beta}_{(i)}^1 = \hat{\beta} - \frac{\hat{r}_{P_i} \sqrt{\hat{w}_i \phi^{-1}}}{1 - \hat{h}_{ii}^*} (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{x}_i,$$

em que $\hat{r}_{P_i} = \sqrt{\hat{\phi}}(y_i - \hat{\mu}_i) / \sqrt{\hat{V}_i}$ é o resíduo de Pearson estimado. Substituindo $\hat{\beta}_{(i)}$ por $\hat{\beta}_{(i)}^1$, obtemos

$$LD_i \cong \left\{ \frac{\hat{h}_{ii}^*}{1 - \hat{h}_{ii}^*} \right\} t_{S_i}^{*2}.$$

Tal aproximação em geral subestima o verdadeiro valor de LD_i , embora seja suficiente para chamar a atenção para os pontos mais influentes. Outras medidas de diagnóstico baseadas na eliminação de pontos na classe dos MLG's são desenvolvidas por Williams (1987).

3.4.2 Diagnóstico em MLGs restritos

Resíduos

Considerando a convergência do processo iterativo dado por (3.10), e denotando a estimativa por $\tilde{\beta}$, temos

$$\tilde{\beta} = \tilde{\beta}^* - (\mathbf{X}^T \tilde{\mathbf{W}} \mathbf{X})^{-1} \mathbf{C}^T \left\{ \mathbf{C} (\mathbf{X}^T \tilde{\mathbf{W}} \mathbf{X})^{-1} \mathbf{C}^T \right\}^{-1} (\mathbf{C} \tilde{\beta}^* - \mathbf{d}),$$

em que $\tilde{\beta}^*$ é dado por $(\mathbf{X}^T \tilde{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^T \tilde{\mathbf{W}} \tilde{\mathbf{z}}$ com $\tilde{\mathbf{z}} = \tilde{\boldsymbol{\eta}} + \tilde{\mathbf{W}}^{-1/2} \tilde{\mathbf{V}}^{-1/2} (\mathbf{y} - \tilde{\boldsymbol{\mu}})$. Há de se observar que $\tilde{\beta}^*$ assume a mesma forma matemática de $\hat{\beta}$ do caso irrestrito, porém as quantidades agora são avaliadas na estimativa restrita $\tilde{\beta}$. De forma análoga ao caso irrestrito, tomando-se $\mathbf{X}^* = \tilde{\mathbf{W}}^{1/2} \mathbf{X}$ e $\mathbf{z}^* = \tilde{\mathbf{W}}^{1/2} \tilde{\mathbf{z}}$, podemos reescrever a expressão acima como

$$\tilde{\beta} = \tilde{\beta}^* - (\mathbf{X}^{*T} \mathbf{X}^*)^{-1} \mathbf{C}^T \left\{ \mathbf{C} (\mathbf{X}^{*T} \mathbf{X}^*)^{-1} \mathbf{C}^T \right\}^{-1} (\mathbf{C} \tilde{\beta}^* - \mathbf{d}). \quad (3.11)$$

Nota-se pela expressão (3.11) uma similaridade com a estimativa restrita do caso normal linear apresentada em (2.17). Para o nosso interesse em particular, reduzimos aos casos em que $\mathbf{d} = \mathbf{0}$. Assim, a expressão acima toma a forma

$$\tilde{\beta} = \tilde{\beta}^* - (\mathbf{X}^{*T} \mathbf{X}^*)^{-1} \mathbf{C}^T \left\{ \mathbf{C} (\mathbf{X}^{*T} \mathbf{X}^*)^{-1} \mathbf{C}^T \right\}^{-1} \mathbf{C} \tilde{\beta}^*.$$

O vetor de valores restritos ajustados pode ser expresso como

$$\begin{aligned} \tilde{\mathbf{z}}^* &= \mathbf{X}^* \tilde{\beta} \\ &= \mathbf{X}^* (\mathbf{X}^{*T} \mathbf{X}^*)^{-1} \mathbf{X}^{*T} \mathbf{z}^* - \mathbf{X}^* (\mathbf{X}^{*T} \mathbf{X}^*)^{-1} \mathbf{C}^T \left\{ \mathbf{C} (\mathbf{X}^{*T} \mathbf{X}^*)^{-1} \mathbf{C}^T \right\}^{-1} \mathbf{C} (\mathbf{X}^{*T} \mathbf{X}^*)^{-1} \mathbf{X}^{*T} \mathbf{z}^* \\ &\quad + \mathbf{X}^* (\mathbf{X}^{*T} \mathbf{X}^*)^{-1} \mathbf{C}^T \left\{ \mathbf{C} (\mathbf{X}^{*T} \mathbf{X}^*)^{-1} \mathbf{C}^T \right\}^{-1} \mathbf{d}. \end{aligned}$$

Chamando $\mathbf{Z}^* = \mathbf{X}^*(\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{C}^T$, podemos definir quantidades similares ao caso normal linear. Por exemplo, temos que $\mathbf{P}_{\mathbf{Z}^*} = \mathbf{Z}^*(\mathbf{Z}^{*T}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*T} = \tilde{\mathbf{G}}^*$, como também $\mathbf{P}_{\mathbf{X}^*} = \mathbf{X}^*(\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{X}^{*T} = \tilde{\mathbf{H}}^*$. Logo, o vetor de valores restritos ajustados pode ser reescrito como

$$\begin{aligned}\bar{\mathbf{z}}^* &= \tilde{\mathbf{H}}^*\mathbf{z}^* - \mathbf{Z}^*(\mathbf{Z}^{*T}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*T}\mathbf{z}^* + \mathbf{Z}^*(\mathbf{Z}^{*T}\mathbf{Z}^*)^{-1}\mathbf{d} \\ &= \tilde{\mathbf{H}}^*\mathbf{z}^* - \tilde{\mathbf{G}}^*\mathbf{z}^* + \mathbf{Z}^*(\mathbf{Z}^{*T}\mathbf{Z}^*)^{-1}\mathbf{d} \\ &= (\tilde{\mathbf{H}}^* - \tilde{\mathbf{G}}^*)\mathbf{z}^* + \mathbf{Z}^*(\mathbf{Z}^{*T}\mathbf{Z}^*)^{-1}\mathbf{d}.\end{aligned}$$

Em particular, para o caso em que $\mathbf{d} = \mathbf{0}$, a expressão acima toma a forma simplificada

$$\bar{\mathbf{z}}^* = (\tilde{\mathbf{H}}^* - \tilde{\mathbf{G}}^*)\mathbf{z}^*,$$

semelhante a (2.31). Portanto, similarmente ao caso normal linear temos que os resíduos ordinários em MLG's restritos sujeitos a $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ são dados pela diferença entre o valor observado $\mathbf{z}^* = \tilde{\mathbf{W}}^{1/2}\tilde{\mathbf{z}}$ e o respectivo valor ajustado $\bar{\mathbf{z}}^*$ nesta regressão fictícia. Assim, temos

$$\begin{aligned}\mathbf{v}^* &= \mathbf{z}^* - \bar{\mathbf{z}}^* \\ &= (\mathbf{I}_n - \tilde{\mathbf{H}}^* + \tilde{\mathbf{G}}^*)\mathbf{z}^*.\end{aligned}$$

Esse resíduo pode também ser expresso por $\mathbf{v}^* = \tilde{\mathbf{V}}^{-1/2}(\mathbf{y} - \tilde{\boldsymbol{\mu}})$ ou $\mathbf{v}^* = \tilde{\mathbf{W}}^{1/2}(\tilde{\mathbf{z}} - \tilde{\boldsymbol{\eta}})$. Supondo que as quantidades são avaliadas no verdadeiro valor de $\boldsymbol{\beta}$, temos para qualquer valor de \mathbf{d} que

$$\begin{aligned}\text{Var}(\mathbf{v}^*) &= (\mathbf{I}_n - \mathbf{H}^* + \mathbf{G}^*)\text{Var}(\mathbf{z}^*)(\mathbf{I}_n - \mathbf{H}^* + \mathbf{G}^*)^T \\ &= (\mathbf{I}_n - \mathbf{H}^* + \mathbf{G}^*)\text{Var}(\mathbf{W}^{1/2}\mathbf{z})(\mathbf{I}_n - \mathbf{H}^* + \mathbf{G}^*) \\ &= (\mathbf{I}_n - \mathbf{H}^* + \mathbf{G}^*)\mathbf{W}^{1/2}\text{Var}(\mathbf{z})\mathbf{W}^{1/2}(\mathbf{I}_n - \mathbf{H}^* + \mathbf{G}^*) \\ &= (\mathbf{I}_n - \mathbf{H}^* + \mathbf{G}^*)\mathbf{W}^{1/2}\boldsymbol{\phi}^{-1}\mathbf{W}^{-1}\mathbf{W}^{1/2}(\mathbf{I}_n - \mathbf{H}^* + \mathbf{G}^*) \\ &= \boldsymbol{\phi}^{-1}(\mathbf{I}_n - \mathbf{H}^* + \mathbf{G}^*),\end{aligned}$$

uma vez que $\text{Var}(\mathbf{z}) = \phi^{-1}\mathbf{W}^{-1}$ e a matriz $(\mathbf{I}_n - \mathbf{H}^* + \mathbf{G}^*)$ é simétrica e idempotente (ver Apêndice A.3). Em outras palavras, a matriz \mathbf{G}^* projeta num subespaço que está contido no subespaço em que a matriz \mathbf{H}^* projeta.

Logo, o i -ésimo resíduo padronizado toma, com todas as quantidades sendo avaliadas na estimativa restrita $\tilde{\beta}$, a seguinte forma:

$$t_{v_i}^* = \frac{\phi^{1/2}(y_i - \tilde{\mu}_i)}{\sqrt{\tilde{V}_i(1 - \tilde{h}_{ii}^* + \tilde{g}_{ii}^*)}}, \quad i = 1, \dots, n.$$

A versão padronizada do resíduo componente do desvio para o caso restrito fica, portanto, dada por

$$t_{D_i}^* = \frac{\phi^{1/2}d(y_i; \tilde{\mu}_i)}{\sqrt{1 - \tilde{h}_{ii}^* + \tilde{g}_{ii}^*}}.$$

em que $d(y_i; \tilde{\mu}_i) = \sqrt{2}\{y_i(\hat{\theta}_i^0 - \tilde{\theta}_i) + [b(\tilde{\theta}_i) - b(\hat{\theta}_i^0)]\}^{1/2}$.

Pontos de alavanca

Como $\mathbf{v}^* = (\mathbf{I}_n - \tilde{\mathbf{H}}^* + \tilde{\mathbf{G}}^*)\mathbf{z}^*$ podemos, similarmente ao caso normal linear, considerar os elementos da diagonal principal da matriz $(\tilde{\mathbf{H}}^* - \tilde{\mathbf{G}}^*)$ como pontos de alavanca no caso de MLGs restritos, embora não existam garantias de que os mesmos coincidam com $\partial \tilde{y}_i / \partial y_i$ nos casos de resposta contínua, exceto no caso normal (vide Paula, 1999). Como $\text{traço}(\tilde{\mathbf{H}}^* - \tilde{\mathbf{G}}^*) = p - q$, podemos considerar como alavanca os pontos com $(\tilde{h}_{ii}^* - \tilde{g}_{ii}^*)$ superior a $\frac{2(p-q)}{n}$, sendo

$$\tilde{h}_{ii}^* = \tilde{w}_i \mathbf{x}_i^T (\mathbf{X}^T \tilde{\mathbf{W}} \mathbf{X})^{-1} \mathbf{x}_i$$

e

$$\tilde{g}_{ii}^* = \tilde{w}_i \mathbf{x}_i^T (\mathbf{X}^T \tilde{\mathbf{W}} \mathbf{X})^{-1} \mathbf{C}^T \{ \mathbf{C} (\mathbf{X}^T \tilde{\mathbf{W}} \mathbf{X})^{-1} \mathbf{C}^T \}^{-1} \mathbf{C} (\mathbf{X}^T \tilde{\mathbf{W}} \mathbf{X})^{-1} \mathbf{x}_i.$$

Esse critério, contudo, não deve valer no caso de regressão logística como já foi mencionado no caso irrestrito.

Influência

O afastamento da verossimilhança no caso restrito, fazendo uma analogia com o caso irrestrito, fica dado por

$$LD_i^c = 2\{L(\tilde{\beta}) - L(\tilde{\beta}_{(i)})\},$$

cuja aproximação quadrática é dada por

$$LD_i^c \cong \phi (\tilde{\beta} - \tilde{\beta}_{(i)})^T (\mathbf{X}^T \tilde{\mathbf{W}} \mathbf{X}) (\tilde{\beta} - \tilde{\beta}_{(i)}).$$

Após algumas manipulações algébricas podemos mostrar que a aproximação de um passo no caso restrito assume a forma

$$\tilde{\beta}_{(i)}^1 = \tilde{\beta} - \frac{v_i^*}{1 - \tilde{h}_{ii}^* + \tilde{g}_{ii}^*} (\mathbf{X}^T \tilde{\mathbf{W}} \mathbf{X})^{-1} (\mathbf{x}_i - \tilde{\mathbf{a}}_i),$$

em que $\tilde{\mathbf{a}}_i = \mathbf{C}^T (\tilde{\mathbf{Z}}^T \tilde{\mathbf{Z}})^{-1} \tilde{\mathbf{z}}_i$.

Assim,

$$LD_i^c \cong \left\{ \frac{\tilde{h}_{ii}^* - \tilde{g}_{ii}^*}{1 - \tilde{h}_{ii}^* + \tilde{g}_{ii}^*} \right\} t_{v_i}^{*2}, \quad i = 1, \dots, n.$$

3.4.3 Envelopes

Para a construção de envelopes em MLGs pode-se tanto utilizar os resíduos studentizados (t_{S_i} ou $t_{S_i}^*$) quanto os resíduos componentes do desvio (t_{D_i} ou $t_{D_i}^*$). Em Paula(2004) temos os programas desenvolvidos com o software S-Plus para MLG's irrestritos. No caso de MLGs restritos na forma $\mathbf{C}\beta = \mathbf{0}$ a adaptação é simples, sendo que no Apêndice E temos o programa para casos com distribuição normal e gama.

3.4.4 Resumo das medidas de diagnóstico para MLG's

Na Tabela 3.1 temos um resumo das medidas de diagnóstico citadas para os casos irrestrito e restrito em modelos lineares generalizados. Nota-se uma analogia com os modelos lineares apresentados na Tabela 2.1.

Tabela 3.1 Medidas de diagnóstico para MLG
irrestritos e restritos na forma $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$

| Medida | caso irrestrito | caso restrito |
|-------------------|--|--|
| valor ajustado | $\hat{\mathbf{z}}^* = \hat{\mathbf{H}}^* \mathbf{z}^*$ | $\tilde{\mathbf{z}}^* = (\tilde{\mathbf{H}}^* - \tilde{\mathbf{G}}^*) \mathbf{z}^*$ |
| res. ordinário | $\mathbf{e}^* = (\mathbf{I}_n - \hat{\mathbf{H}}^*) \mathbf{z}^*$ | $\mathbf{v}^* = (\mathbf{I}_n - \tilde{\mathbf{H}}^* + \tilde{\mathbf{G}}^*) \mathbf{z}^*$ |
| res. padronizado | $t_{S_i}^* = \frac{\phi^{1/2}(y_i - \hat{\mu}_i)}{\sqrt{\hat{V}_i(1 - \hat{h}_{ii}^*)}}$ | $t_{v_i}^* = \frac{\phi^{1/2}(y_i - \tilde{\mu}_i)}{\sqrt{\tilde{V}_i(1 - \tilde{h}_{ii}^* + \tilde{g}_{ii}^*)}}$ |
| res. comp. desvio | $t_{D_i} = \frac{\phi^{1/2}d(y_i; \hat{\mu}_i)}{\sqrt{1 - \hat{h}_{ii}^*}}$ | $t_{D_i}^* = \frac{\phi^{1/2}d(y_i; \tilde{\mu}_i)}{\sqrt{1 - \tilde{h}_{ii}^* + \tilde{g}_{ii}^*}}$ |
| alavanca, (corte) | $h_{ii}^*, \left(\frac{2p}{n}\right)$ | $h_{ii}^* - g_{ii}^*, \left(\frac{2(p-q)}{n}\right)$ |
| desloc. veross. | $LD_i \cong \left\{ \frac{\hat{h}_{ii}^*}{1 - \hat{h}_{ii}^*} \right\} t_{S_i}^{*2}$ | $LD_i^c \cong \left\{ \frac{\tilde{h}_{ii}^* - \tilde{g}_{ii}^*}{1 - \tilde{h}_{ii}^* + \tilde{g}_{ii}^*} \right\} t_{v_i}^{*2}$ |
| critério Akaike | $AIC = D_p^* + 2p$ | $AIC^c = D_p^c + 2(p - q)$ |

3.4.5 Extensões

Podemos também utilizar a teoria desenvolvida nesse capítulo para compor a parte sistemática de um modelo linear generalizado em que o preditor apresenta uma estrutura polinomial na forma de covariáveis, ou seja,

$$\boldsymbol{\eta} = \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{\varepsilon},$$

em que \mathbf{X}_1 é uma matriz $n \times p_1$ com valores de variáveis explicativas (grupos, por exemplo) e \mathbf{X}_2 é uma matriz $n \times (p - p_1)$ com a representação polinomial. Tomando-se $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ e $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$, pode-se utilizar todos os resultados desenvolvidos anteriormente para essa estrutura, inclusive sujeitá-la a restrições do tipo $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$.

Capítulo 4

Aplicações

Neste capítulo retomaremos o exemplo citado na introdução e apresentaremos um outro para ilustrar as técnicas propostas de modelagem e diagnóstico para modelos segmentados restritos. Outros exemplos didáticos são encontrados em Tarpey e Holcomb (2000).

4.1 Crescimento de crianças

Este exemplo consiste de um estudo de crescimento de crianças americanas em idade pré-escolar. Os dados referem-se a informações de razão entre o peso (libras) e a altura (cm) de 72 meninos na faixa etária de 0,5 a 71,5 meses (Eubank, 1988) e estão no Apêndice B.1. O gráfico de dispersão com esses dados está na Figura 4.1, onde podemos notar uma mudança na tendência da função ao longo da faixa-etária. Do nascimento até o 16^o mês há um crescimento acelerado, a partir do qual segue um ritmo bem menos acentuado.

A princípio, modelos de regressão polinomial com suposição de $\varepsilon_i \sim N(0, \sigma^2)$ poderiam ser sugeridos. A expressão para esses modelos é dada por

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_i^j + \varepsilon_i, \quad i = 1, \dots, n, \quad (4.1)$$

em que k é o grau do polinômio. Ajustes de graus $k = 1, \dots, 5$ foram realizados, uma vez que encontrávamos tendências nos gráficos que avaliavam a homoscedasticidade (as Figuras 4.2a-c apresentam os gráficos para $k = 3, 4$ e 5). Com um polinômio de grau $k = 5$ observamos que houve uma estabilização na variância dos resíduos, porém o modelo contava com 6 parâmetros.

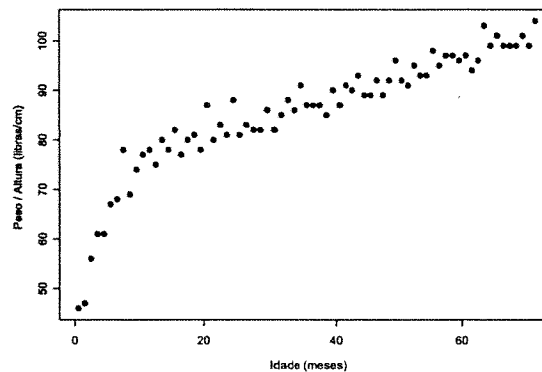


Figura 4.1: Diagrama de dispersão do crescimento de crianças segundo a idade.

Gráficos de envelopes para os resíduos studentizados t_{v_i} 's também foram gerados e podem ser observados nas Figuras 4.3a-c. Paralelamente, apresentamos na Tabela 4.1 algumas medidas obtidas com esses modelos.

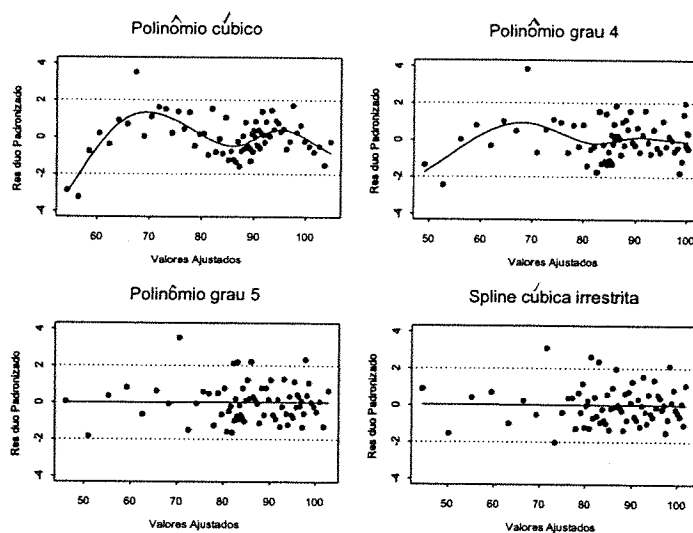


Figura 4.2: Gráficos de resíduos contra os valores ajustados para os modelos polinomiais de graus 3, 4, 5 e *spline* cúbica irrestrita.

Tabela 4.1 Medidas de seleção e de qualidade de ajuste para os modelos polinomiais de grau 3, 4, 5 e para a *spline* cúbica irrestrita.

| Modelo | PRESS | R^2 ajustado | s(g.l) | AIC |
|---------------------------------------|--------|----------------|----------|--------|
| Polinômio de grau 3 | 914,12 | 0,92 | 3,36(68) | 777,73 |
| Polinômio de grau 4 | 589,32 | 0,95 | 2,70(67) | 499,19 |
| Polinômio de grau 5 | 465,79 | 0,96 | 2,43(66) | 402,26 |
| Spline cúbica irrestrita (modelo 4.4) | 431,33 | 0,96 | 2,34(67) | 377,59 |

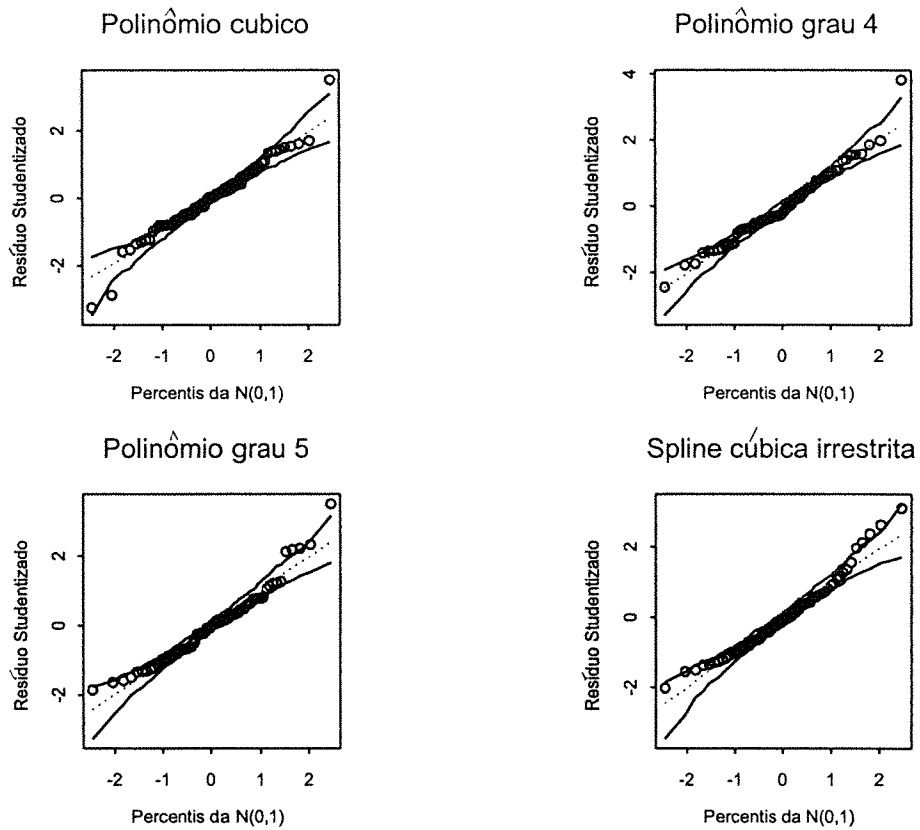


Figura 4.3: Gráficos normais de probabilidade para os modelos polinomiais de graus 3, 4, 5 e *spline* cúbica irrestrita.

Uma alternativa de ajuste para esse exemplo é encontrada em Tarpey (2000) que sugere a utilização de modelos normais *spline* cúbicos com um ponto de mudança em $t = 16$. Algumas propostas de modelagem por *splines* contínuas com variados graus de suavização foram testados. O modelo contínuo com 1ª derivada não-contínua é dado por

$$y_i = \beta_{00} + \beta_{01}x_i + \beta_{02}x_i^2 + \beta_{03}x_i^3 + \beta_{11}(x_i - 16)_+^1 + \beta_{12}(x_i - 16)_+^2 + \beta_{13}(x_i - 16)_+^3 + \varepsilon_i, \quad (4.2)$$

em que y_i é a razão entre o peso e a altura da i -ésima criança, x_i é a respectiva idade, $i = 1, \dots, 72$, e $\varepsilon_i \sim N(0, \sigma^2)$.

A função $(x - 16)_+^j$, para $j = 1, \dots, 3$, é definida por

$$(x - 16)_+^j = \begin{cases} 0, & \text{se } x \leq 16 \\ (x - 16)^j, & \text{se } x > 16. \end{cases}$$

Outras *splines* polinomiais foram construídas eliminando-se alguns termos de (4.2) e aumentando-se o grau de suavização do modelo, como segue abaixo:

$$y_i = \beta_{00} + \beta_{01}x_i + \beta_{02}x_i^2 + \beta_{03}x_i^3 + \beta_{12}(x_i - 16)_+^2 + \beta_{13}(x_i - 16)_+^3 + \varepsilon_i \quad (4.3)$$

e

$$y_i = \beta_{00} + \beta_{01}x_i + \beta_{02}x_i^2 + \beta_{03}x_i^3 + \beta_{13}(x_i - 16)_+^3 + \varepsilon_i, \quad i = 1, \dots, n. \quad (4.4)$$

O modelo (4.3) garante a continuidade da 1ª derivada. Em (4.4) temos a modelagem mais suave dos três modelos, apresentando 1ª e 2ª derivadas contínuas. Os modelos (4.2) e (4.3) apresentaram algumas estimativas não significativas (ver Tabelas 4.2 e 4.3). A *spline* polinomial cúbica dada por (4.4) tem todas as estimativas significativas (Tabela 4.4) e medidas que indicam que o ajuste por esse modelo é superior ao polinomial não segmentado (ver Tabela 4.1 e Figuras 4.3d e 4.2d). O gráfico com envelope também parece acomodar melhor a observação 72. Nesse modelo, a resposta é cúbica para qualquer idade, sendo algebricamente dada por

$$y_i = \begin{cases} \beta_{00} + \beta_{01}x_i + \beta_{02}x_i^2 + \beta_{03}x_i^3, & \text{para } x_i \leq 16, \\ \beta_{00} - 4096\beta_{13} + (\beta_{01} + 768\beta_{13})x_i + (\beta_{02} - 48\beta_{13})x_i^2 \\ \quad + (\beta_{03} + \beta_{13})x_i^3, & \text{para } x_i > 16. \end{cases} \quad (4.5)$$

Tabela 4.2 Estimativas dos parâmetros referentes ao modelo (4.2) *spline* cúbica irrestrita com 1ª derivada descontínua.

| Parâmetro | Estimativa(E) | D.Padrão | E/D.Padrão | P-valor |
|--------------|-----------------------|----------------------|------------|---------|
| β_{00} | 41,17 | 2,36 | 17,43 | 0,00 |
| β_{01} | 6,52 | 1,21 | 5,37 | 0,00 |
| β_{02} | -0,39 | 0,16 | -2,38 | 0,02 |
| β_{03} | $8,6 \times 10^{-3}$ | $6,3 \times 10^{-3}$ | 1,36 | 0,17 |
| β_{11} | -0,13 | 0,95 | -0,14 | 0,89 |
| β_{12} | -0,02 | 0,14 | -0,12 | 0,90 |
| β_{13} | $-8,6 \times 10^{-3}$ | $6,4 \times 10^{-3}$ | -1,35 | 0,18 |

Tabela 4.3 Estimativas dos parâmetros referentes ao modelo (4.3) *spline* cúbica irrestrita com 1ª derivada contínua.

| Parâmetro | Estimativa(E) | D.Padrão | E/D.Padrão | P-valor |
|--------------|-----------------------|----------------------|------------|---------|
| β_{00} | 41,30 | 2,14 | 19,25 | 0,00 |
| β_{01} | 6,40 | 0,92 | 6,98 | 0,00 |
| β_{02} | -0,37 | 0,10 | -3,67 | 0,00 |
| β_{03} | $7,8 \times 10^{-3}$ | $3,2 \times 10^{-3}$ | 2,42 | 0,02 |
| β_{12} | 6×10^{-4} | $5,8 \times 10^{-2}$ | 0,01 | 0,99 |
| β_{13} | $-7,8 \times 10^{-3}$ | $3,2 \times 10^{-3}$ | -2,45 | 0,02 |

Tabela 4.4 Estimativas dos parâmetros referentes ao modelo (4.4) *spline* cúbica irrestrita com 1ª e 2ª derivadas contínuas.

| Parâmetro | Estimativa(E) | D.Padrão | E/D.Padrão | P-valor |
|--------------|-----------------------|----------------------|------------|---------|
| β_{00} | 41,29 | 1,77 | 23,36 | 0,00 |
| β_{01} | 6,41 | 0,51 | 12,70 | 0,00 |
| β_{02} | -0,38 | 0,04 | -9,77 | 0,00 |
| β_{03} | $7,9 \times 10^{-3}$ | $8,7 \times 10^{-3}$ | 9,02 | 0,00 |
| β_{13} | $-7,8 \times 10^{-3}$ | $9,1 \times 10^{-3}$ | -8,62 | 0,00 |

Para ilustrar, mostramos na Figura 4.4 as curvas ajustadas no diagrama de dispersão referentes aos modelos polinomiais de graus 3, 4 e 5, assim como da *spline* cúbica. Aparentemente, todas as curvas poderiam ajustar-se aos dados, porém como vimos pelas medidas apresentadas, o modelo segmentado parece mais interessante.

Tarpey (2000) propõe ainda para esse mesmo conjunto de dados um modelo alternativo restrito. A proposta seria ajustar um modelo segmentado dado primeiramente por uma equação cúbica e então, a partir de $x = 16$, complementar com uma equação linear, uma vez que o comportamento da variável resposta parece mudar a partir do 16º mês. Para tanto, um modelo *spline* cúbico-linear contínuo poderá ser construído a partir de (4.4) uma vez que será um caso particular desse quando ocorrerem

$$\beta_{02} - 48\beta_{13} = 0 \quad \text{e} \quad \beta_{03} + \beta_{13} = 0.$$

Para isso, determinamos os coeficientes do modelo (4.4) sujeito às restrições acima, que serão exatamente aquelas que anularão os coeficientes de x_i^2 e x_i^3 para crianças com idade superior a 16 meses ($x_i > 16$), como observamos ao analisar a função(4.5).

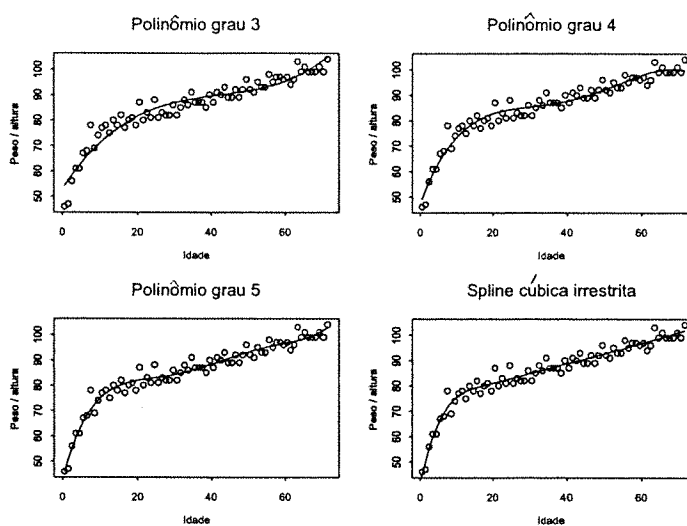


Figura 4.4: Diagramas de dispersão dos dados de crescimento com curvas ajustadas.

Para a utilização dos métodos apresentados, as restrições deverão estar na forma matricial $C\beta = d$, em que

$$C = \begin{bmatrix} 0 & 0 & 1 & 0 & -48 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix},$$

$$\beta = [\beta_{00}, \beta_{01}, \beta_{02}, \beta_{03}, \beta_{13}]^T \quad e$$

$$d = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Por sua vez, a matriz modelo será dada por

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & 0 \\ 1 & x_2 & x_2^2 & x_2^3 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{16} & x_{16}^2 & x_{16}^3 & 0 \\ 1 & x_{17} & x_{17}^2 & x_{17}^3 & (x_{17} - 16)^3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{72} & x_{72}^2 & x_{72}^3 & (x_{72} - 16)^3 \end{bmatrix}.$$

O modelo foi ajustado utilizando-se estimação restrita por mínimos quadrados, sendo que as estimativas encontradas estão na Tabela 4.5.

Tabela 4.5 Estimativas dos parâmetros referentes ao modelo normal cúbico-linear restrito.

| Parâmetro | Estimativa(E) | D.Padrão | E/D.Padrão | P-valor |
|--------------|-----------------------|--------------------|------------|---------|
| β_{00} | 41,48 | 1.74 | 23,76 | 0,00 |
| β_{01} | 6,32 | 0.49 | 12,68 | 0,00 |
| β_{02} | -0,37 | 0.04 | -9,72 | 0,00 |
| β_{03} | $7,7 \times 10^{-3}$ | 9×10^{-4} | 8,95 | 0,00 |
| β_{13} | $-7,7 \times 10^{-3}$ | 9×10^{-4} | -8,53 | 0,00 |

Tabela 4.6 Medidas comparativas e desvio-padrão estimado (graus de liberdade) dos modelos normal irrestrito e restrito.

| Modelo | PRESS | R^2 ajustado | AIC | s(g.l.) |
|------------------|--------|----------------|--------|----------|
| irrestrito (4.4) | 431,32 | 0,96 | 377,59 | 2,34(67) |
| restrito | 405,12 | 0,96 | 375,35 | 2,31(69) |

Nota-se que as estimativas dos modelos restrito e irrestrito estão bem próximas. Pela Tabela 4.6 observamos que todos os critérios de seleção e comparação de modelos indicam que o modelo restrito está melhor ajustado. Realizou-se também um teste de Wald para se comparar o modelo restrito (H_0) com o irrestrito (H_1) que é expresso através das seguintes hipóteses:

$$H_0 : C\beta = 0 \quad \text{contra} \quad H_1 : C\beta \neq 0.$$

A estatística de Wald apresentou o valor de 0,3209 com dois graus de liberdade ($P = 0,8517$), o que nos leva a não rejeitar H_0 .

Pelo gráfico de envelope do modelo restrito, apresentado na Figura 4.5, nota-se que o modelo não está inadequado embora apareça uma leve tendência sistemática que pode indicar falta de algum termo quadrático na parte sistemática, ou que a distribuição do erro não é simétrica.

A Figura 4.6 apresenta alguns gráficos de diagnóstico para o modelo normal restrito. Nesses gráficos destacam-se: a observação #2 é a que mais influencia as estimativas dos parâmetros, além de ser alavanca; observação #8 aparece como aberrante e influente e a observação #1, que além de ser alavanca é também influente. Pela Figura 4.6d, não parece haver indícios de heteroscedasticidade.

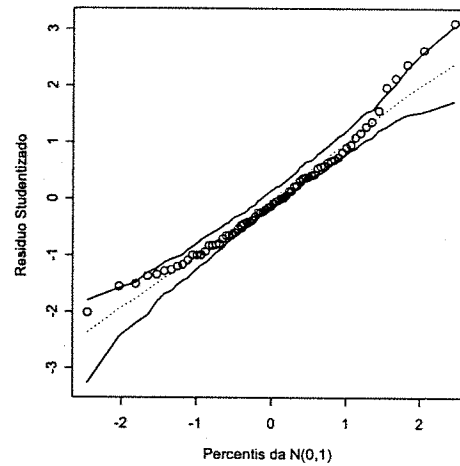


Figura 4.5: Gráfico normal de probabilidade do modelo cúbico-linear restrito.

O impacto sofrido pelas estimativas dos coeficientes ao eliminarmos as observações atípicas do conjunto de dados, pode ser visto na Tabela 4.7, onde apresentamos uma análise confirmatória. Espera-se num ajuste “ideal”, que cada observação contribua com $1/72 = 0,0139 = 1,39\%$ para a estimativa de cada coeficiente. Como podemos ver, é com a retirada da observação #2 que as estimativas apresentam o maior impacto. As observações #1 e #2 referem-se a duas crianças de 0,5 e 1,5 meses com valor da variável resposta inferior ao esperado pelo crescimento das crianças nos demais meses. Já as observação #8 e #21 referem-se a crianças com razão peso altura superior ao valor médio esperado para os 7,5 e 20,5 meses, respectivamente. Contudo, a eliminação individual dessas observações não muda os resultados inferenciais.

Tabela 4.7 Variações nas estimativas do modelo normal cúbico-linear restrito com a eliminação individual das observações suspeitas.

| Parâmetro | sem #1 | sem #2 | sem #8 | sem #21 |
|--------------|---------|---------|--------|---------|
| β_{00} | -2,43 % | 3,47 % | 0,56% | 0,46% |
| β_{01} | 3,25 % | -4,50 % | -0,24% | -1,35% |
| β_{02} | -3,50 % | 4,83 % | 0,37% | 1,53% |
| β_{03} | 3,50 % | -4,83 % | -0,37% | -1,53% |
| β_{13} | -3,50 % | 4,83 % | 0,37% | 1,53% |

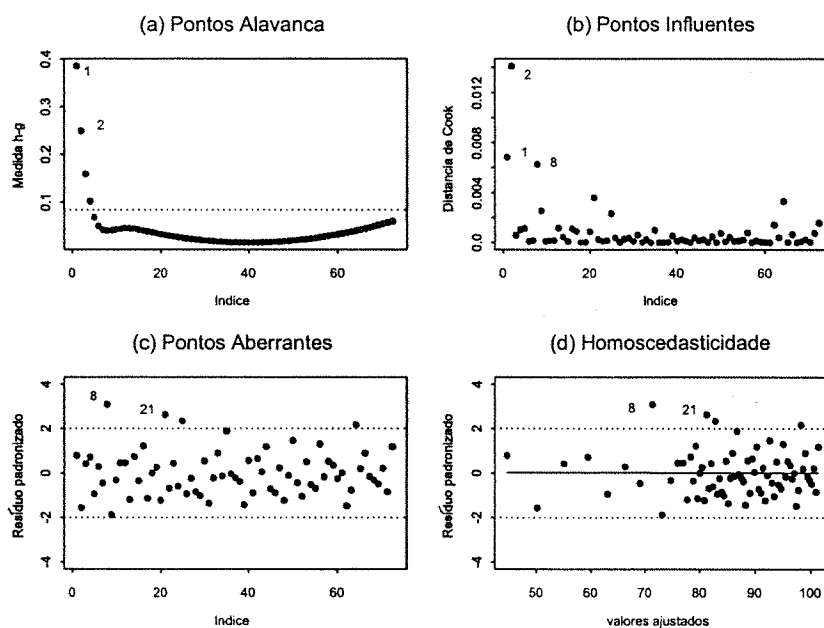


Figura 4.6: Gráficos de diagnóstico do modelo normal cúbico-linear restrito.

Ajuste via MLG

Com o intuito apenas de comparação, apresentamos a seguir para o mesmo conjunto de dados um ajuste através de modelos lineares generalizados. Como a resposta é a razão de duas medidas (peso e altura) assumida com distribuição normal, podemos pensar que a resposta tenha uma outra distribuição, gama, por exemplo. Supondo essa distribuição para a variável resposta, teríamos um modelo linear generalizado. Assim, nosso modelo seria composto por 2 componentes:

1) parte aleatória com Y_i tendo distribuição gama de média μ_i e coeficiente de variação $\phi^{-1/2}$, denotamos $Y_i \sim G(\mu_i, \phi)$;

2) componente sistemática dada por

$$g(\mu_i) = \eta_i,$$

em que

$$\eta_i = \beta_{00} + \beta_{01}x_i + \beta_{02}x_i^2 + \beta_{03}x_i^3 + \beta_{13}(x_i - 16)_+^3, \quad (4.6)$$

é o preditor linear que poderá estar sujeito a restrições na forma $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$.

Na Tabela 4.8 apresentamos o desvio escalonado $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ para três funções de ligação, recíproca, logarítmica e identidade. Segundo este critério não observamos diferenças importantes entre os três modelos. O valor alto para $\hat{\phi}$ confirma a simetria para a distribuição da resposta.

Os gráficos com envelopes para os modelos gama irrestritos são apresentados nas Figuras 4.7a-c e sugerem uma ligeira superioridade para o modelo gama com ligação identidade, confirmando assim o bom ajuste já observado para o modelo normal.

Tabela 4.8 Desvio escalonado dos modelos *spline* gama com preditor(4.6) para algumas funções de ligação.

| Ligação | $g(\mu)$ | $\hat{\phi}$ | $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ |
|-------------|-----------------|--------------|---|
| recíproca | $\frac{1}{\mu}$ | 1168,95 | 72,0103 |
| logarítmica | $\log \mu$ | 1264,49 | 72,0095 |
| identidade | μ | 1280,39 | 72,0094 |

A partir dos modelos gama *spline* cúbicos podemos impor restrições lineares na forma $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$ ao preditor linear, e ajustar um MLG restrito. As quantidades \mathbf{C} , $\boldsymbol{\beta}$ e \mathbf{d} são as mesmas do ajuste normal linear restrito. Através do software S-Plus, após implantar a macro `mlg.rest` (Cysneiros, 1997) e definir as matrizes $\mathbf{C}_{res} = \mathbf{C}$, $\mathbf{sol} = \mathbf{d}$, podemos realizar os ajustes para os modelos gama com ligação recíproca, logarítmica e identidade da seguinte maneira:

```
eubank.reg <- glm(y ~ 1 + x + x^2 + x^3 + (x - t)^3, method = "mlg.rest",
family = gama(lig), Cres = Cres, sol = sol),
```

em que *lig* deverá ser substituída por alguma ligação (*inverse*, *log* ou *identity*) e *t* pelo valor do ponto de mudança. A Tabela 4.9 apresenta os desvios escalonados restritos $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ e valores para ϕ , tal que ϕ^{-1} é o parâmetro de dispersão para esses modelos. Observando essas medidas, não notamos diferenças importantes entre os valores dos três desvios. Os gráficos de envelope para esses modelos restritos foram gerados (vide programa no Apêndice E) e são apresentados nas Figuras 4.8a-c. Novamente nota-se uma ligeira superioridade do modelo gama com ligação identidade. Ademais, por simplicidade nas interpretações dos parâmetros iremos continuar as análises com esse modelo.

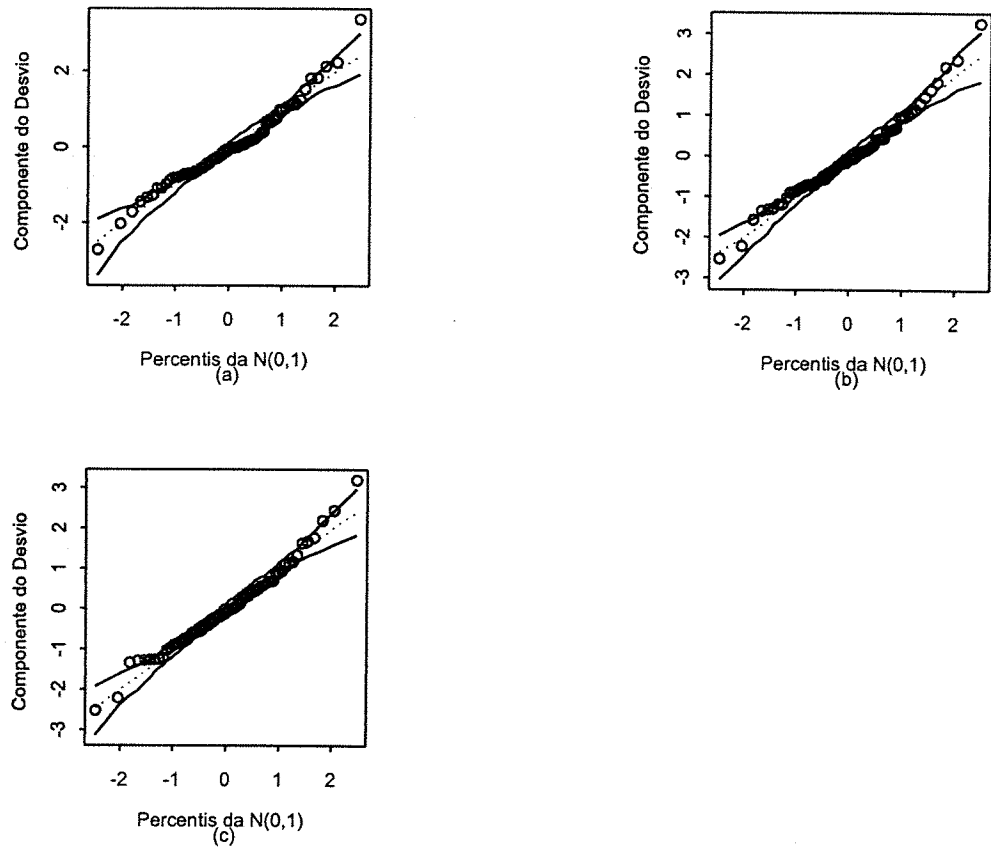


Figura 4.7: Gráficos normais de probabilidade para os modelos gama com preditor (4.6) e ligação (a) inversa, (b) logarítmica e (c) identidade.

Tabela 4.9 Desvio escalonado dos modelos *splines* gama com preditor cúbico-linear para algumas funções de ligação.

| Ligação | $g(\mu)$ | $\bar{\phi}$ | $D^*(\mathbf{y}; \bar{\boldsymbol{\mu}})$ |
|-------------|-----------------|--------------|---|
| recíproca | $\frac{1}{\mu}$ | 1151,82 | 72,0104 |
| logarítmica | $\log \mu$ | 1257,62 | 72,0095 |
| identidade | μ | 1276,35 | 72,0094 |

Os gráficos de diagnóstico para o modelo gama com preditor restrito e ligação identidade podem ser vistos na Figura 4.9. Fazendo uma comparação com a Figura 4.6, referente ao modelo normal restrito, notamos que sob o modelo gama apenas as observações #1 e #2 aparecem como possivelmente influentes. A eliminação individual dessas observações teve variações semelhantes às aquelas apresentadas na Tabela 4.7. Na seqüência temos as estimativas para os modelos gama identidade irrestrito (Tabela 4.10) e restrito (Tabela 4.11). Nota-se que ao utilizarmos a ligação identidade encontramos praticamente as mesmas estimativas obtidas para o caso normal (Tabela 4.4), uma vez que para valores grandes de ϕ há uma aproximação para a distribuição normal.

Para ilustrar, apresentamos na Figura 4.10 os diagramas de dispersão com curvas ajustadas pelos modelos normal *spline* cúbica, normal *spline* cúbica-linear e gama *spline* cúbica-linear com ligação identidade. Em termos gráficos, as diferenças entre esses modelos parecem muito sutis.

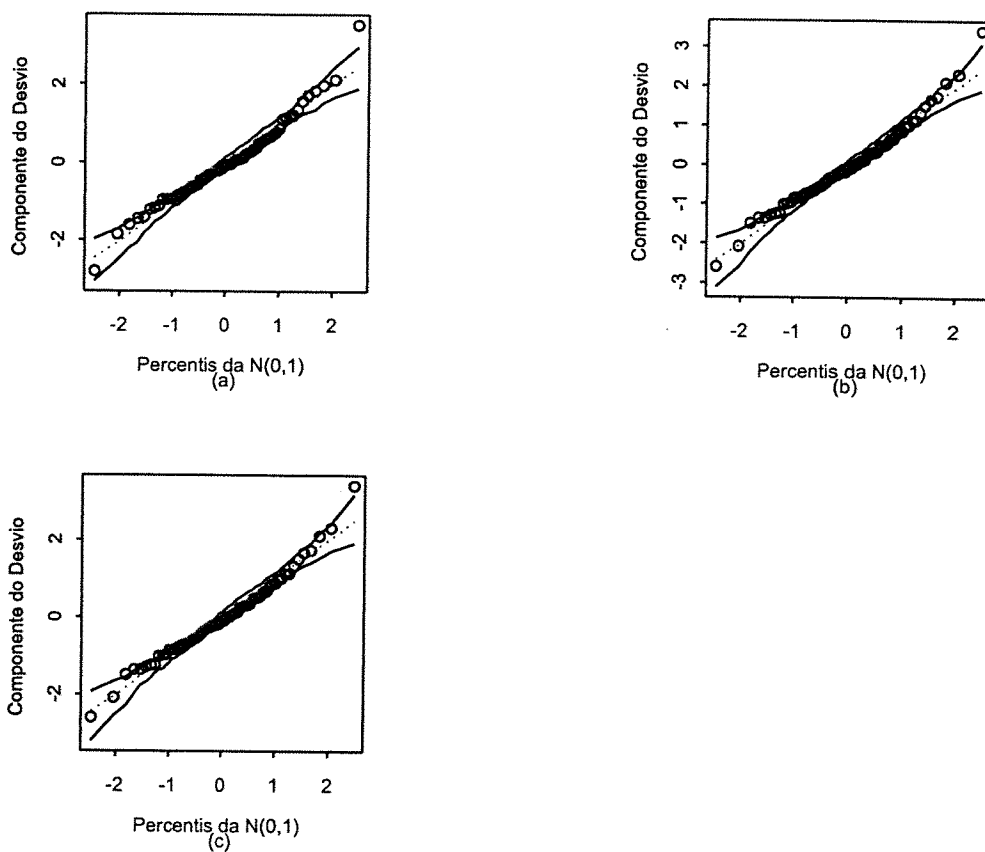


Figura 4.8: Gráficos normais de probabilidade para os modelos gama com preditor cúbico-linear e ligação (a) inversa, (b) logarítmica e (c) identidade.

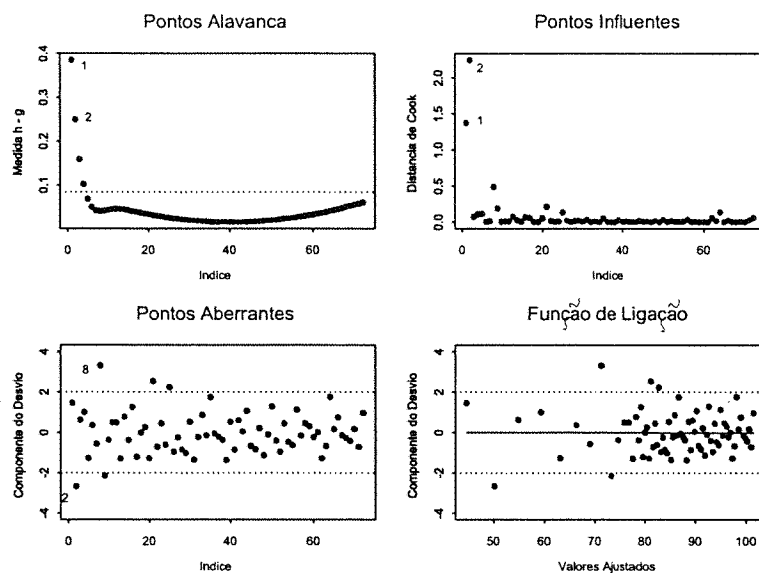


Figura 4.9: Gráficos de diagnóstico do modelo gama restrito com ligação identidade.

Tabela 4.10 Estimativas dos parâmetros referentes ao modelo gama irrestrito com preditor na forma (4.6) e ligação identidade.

| Parâmetro | Estimativa(E) | D.Padrão | E/D.Padrão | P-valor |
|-------------------|------------------------|--------------------|------------|---------|
| β_{00} | 41,29 | 1,77 | 23,36 | 0,00 |
| β_{01} | 6,41 | 0,50 | 12,70 | 0,00 |
| β_{02} | -0,37 | 0,04 | -9,77 | 0,00 |
| β_{03} | $7,85 \times 10^{-3}$ | 8×10^{-4} | 8,98 | 0,00 |
| β_{13} | $-7,84 \times 10^{-3}$ | 9×10^{-4} | -8,56 | 0,00 |
| ϕ | 1280,39 | 213,37 | 6,00 | 0,00 |
| Desvio escalonado | 72,01 | | | |

Tabela 4.11 Estimativas dos parâmetros referentes ao modelo gama com preditor cúbico-linear e ligação identidade.

| Parâmetro | Estimativa(E) | D.Padrão | E/D.Padrão | P-valor |
|-------------------|-----------------------|----------------------|------------|---------|
| β_{00} | 41,48 | 1,74 | 23,76 | 0,00 |
| β_{01} | 6,32 | 0,49 | 12,68 | 0,00 |
| β_{02} | -0,37 | 0,04 | -9,72 | 0,00 |
| β_{03} | $7,7 \times 10^{-3}$ | $8,6 \times 10^{-4}$ | 8,95 | 0,00 |
| β_{13} | $-7,7 \times 10^{-3}$ | 9×10^{-4} | -8,53 | 0,00 |
| ϕ | 1.276,35 | 212,70 | 6,00 | 0,00 |
| Desvio escalonado | 72,01 | | | |

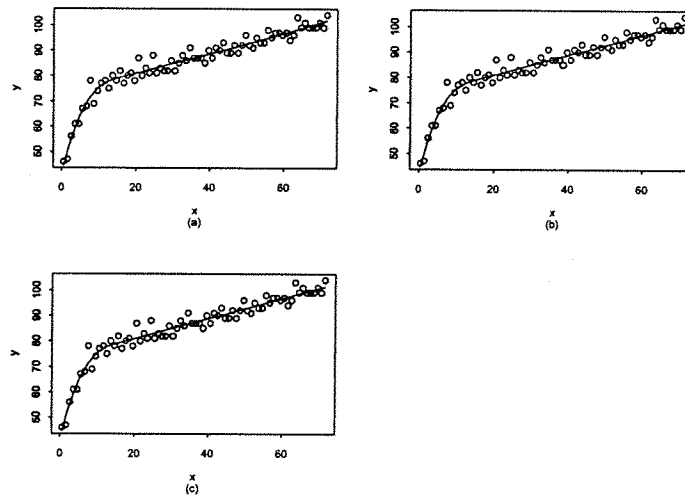


Figura 4.10: Gráficos de dispersão com curva ajustada para os modelos normal *spline* cúbica(a), cúbica-linear(b) e gama *spline* cúbica-linear com ligação identidade(c).

4.2 Diabetes *mellitus*

Os dados deste exemplo referem-se a um estudo sobre os fatores que afetam o padrão de diabetes *mellitus* tipo I em 41 crianças dependentes de insulina, particularmente no que se refere à secreção residual de insulina (Sockett et al., 1987). Embora o conjunto de dados apresente vários fatores, concentraremos nosso interesse em investigar a dependência do nível de serum C-peptídeo (pmol/ml) em relação à idade (anos). Hastie e Tibishirani (1990) utilizam esse exemplo para o estudo de suavizadores em modelos aditivos generalizados. Neste trabalho o objetivo será fornecer uma abordagem alternativa, uma aplicação paramétrica através de modelos lineares generalizados com restrições nos parâmetros.

Na Figura 4.11 apresentamos o gráfico do logaritmo da concentração de C-peptídeo (pmol/ml) no diagnóstico (y_i) contra idade (x_i) das crianças.

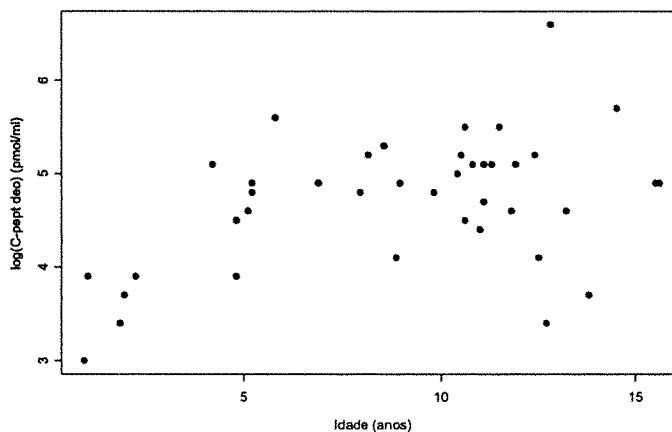


Figura 4.11: Diagrama de dispersão para os dados de logaritmo(C-peptideo) contra a idade de crianças com diabetes *mellitus* tipo I.

Nota-se uma tendência crescente na variabilidade da concentração de C-peptídeo com o passar da idade. Essa característica faz-nos propor para a variável resposta uma distribuição gama de média μ_i , parâmetro de dispersão ϕ^{-1} e ligação logarítmica. Assim, teremos um modelo heteroscedástico de média μ_i e desvio padrão $\mu_i/\sqrt{\phi}$.

Propomos inicialmente um modelo gama polinomial quadrático, com preditor linear dado por

$$\eta_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2, \quad \text{para } i = 1, \dots, 41. \quad (4.7)$$

As estimativas para esse modelo estão na Tabela 4.12.

Tabela 4.12 Estimativas dos parâmetros referentes ao modelo gama polinomial quadrático.

| Parâmetro | Estimativa(E) | D.Padrão | E/D.Padrão | P-valor |
|-------------------|---------------------|----------------------|------------|---------|
| β_0 | 1,20 | 0,07 | 16,15 | 0,00 |
| β_1 | 0,08 | 0,02 | 3,94 | 0,00 |
| β_2 | -4×10^{-3} | $1,2 \times 10^{-3}$ | -3,11 | 0,00 |
| ϕ | 66,80 | 14,72 | 4,54 | 0,00 |
| Desvio escalonado | 41,1023 | | | |

Nesse modelo as estimativas são todas significativas e o desvio escalonado encontrado foi de 41,10 com 38 graus de liberdade ($P = 0,3363$). Para esse ajuste não há indícios de que a ligação seja incorreta, como vemos na Figura 4.12. O gráfico com envelope encontra-se na Figura 4.13.

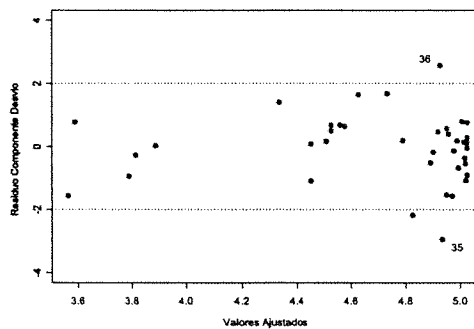


Figura 4.12: Gráfico de adequação da função de ligação para o modelo gama polinomial quadrático.

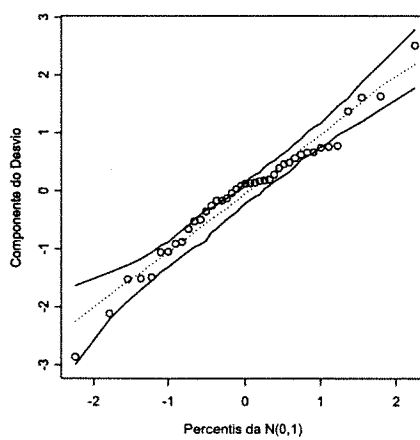


Figura 4.13: Gráfico normal de probabilidade para o modelo gama polinomial quadrático.

Na sequência, propomos um polinômio de grau 3 com o seguinte preditor:

$$\eta_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3, \quad \text{para } i = 1, \dots, 41.$$

Porém, o ajuste para esse modelo apresenta estimativa não significativa para β_3 . Uma outra alternativa seria dividir os dados e ajustar funções diferentes, uma para cada subconjunto. Propõe-se então, um modelo com distribuição gama, preditor linear segmentado (*spline*) quadrático e um ponto de mudança em $t = 6$, dado por

$$\eta_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 (x_i - 6)^2, \quad \text{para } i = 1, \dots, 41. \quad (4.8)$$

Nesse modelo, o preditor tem grau 2 para qualquer idade, sendo algebricamente dado por

$$\eta_i = \begin{cases} \beta_0 + \beta_1 x_i + \beta_2 x_i^2, & \text{para } x_i \leq 6, \\ \beta_0 + 36\beta_3 + (\beta_1 - 12\beta_3)x_i + (\beta_2 + \beta_3)x_i^2, & \text{para } x_i > 6. \end{cases} \quad (4.9)$$

As estimativas dos parâmetros estão na Tabela 4.13. Como podemos ver, esse modelo apresenta a estimativa para β_3 não significativa ao nível de 5%, o que faria com que retornássemos ao modelo gama polinomial quadrático. Porém, vamos investir um pouco mais numa possível relação segmentada entre as variáveis. Acrescentando restrições aos parâmetros, podemos obter um modelo gama *spline* quadrático-linear restrito. Parte-se do modelo segmentado quadrático dado por (4.7) e impõe-se a restrição linear $\beta_2 - \beta_3 = 0$, que é da forma $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$, para obter linearidade no segundo segmento. Nesse caso temos que $\mathbf{C} = [0, 0, 1, 1]$, $\boldsymbol{\beta}^T = [\beta_0, \beta_1, \beta_2, \beta_3]$ e $\mathbf{d} = 0$.

No S-Plus utilizamos a macro `glm.rest` (Cysneiros, 1997) através do comando

```
diabetes.reg < -glm(y ~ 1 + x + x^2 + (x - t)^2, method = "mlg.rest",
family = gama(link = log), Cres = Cres, sol = sol),
```

em que $\text{Cres} = \mathbf{C}$, $\text{sol} = \mathbf{d}$. Ajustando-se o modelo, obtemos todas as estimativas significativas a um nível de 5% como vemos na Tabela 4.14. O gráfico com envelope está na Figura 4.15 e dá indícios de que houve melhora no ajuste, se compararmos com a Figura 4.13. Porém, o desvio escalonado se manteve praticamente o mesmo, chegando ao valor 41,09 com 38 graus de liberdade ($P = 0,3365$).

Tabela 4.13 Estimativas dos parâmetros referentes ao modelo gama *spline* quadrático com $t = 6$.

| Parâmetro | Estimativa(E) | D.Padrão | E/D.Padrão | P-valor |
|-------------------|-----------------------|----------------------|------------|---------|
| β_0 | 1,05 | 0,11 | 9,52 | 0,00 |
| β_1 | 0,17 | 0,05 | 3,10 | 0,00 |
| β_2 | $-1,4 \times 10^{-2}$ | $5,6 \times 10^{-3}$ | -2,39 | 0,02 |
| β_3 | $1,3 \times 10^{-2}$ | $7,3 \times 10^{-3}$ | 1,77 | 0,08 |
| ϕ | 72,39 | 15,95 | 4,54 | 0,00 |
| Desvio escalonado | 41,0944 | | | |

Tabela 4.14 Estimativas dos parâmetros referentes ao modelo gama *spline* quadrático-linear com $t = 6$.

| Parâmetro | Estimativa (E) | D.Padrão | E/D.padrão | P-valor |
|-------------------|-----------------------|----------------------|------------|---------|
| β_0 | 1,04 | 0,11 | 9,53 | 0,00 |
| β_1 | 0,18 | 0,05 | 3,32 | 0,00 |
| β_2 | $-1,5 \times 10^{-2}$ | $5,6 \times 10^{-3}$ | -2,64 | 0,01 |
| β_3 | $1,5 \times 10^{-2}$ | $7,2 \times 10^{-3}$ | 2,05 | 0,04 |
| ϕ | 72,23 | 15,92 | 4,54 | 0,00 |
| Desvio escalonado | 41,0946 | | | |

Na Figura 4.14 podemos ver as curvas ajustadas para os modelos gama quadrático e

spline quadrático-linear. O modelo *spline* restrito aparece como alternativa interessante à tendência apresentada pelos dados, e baseados nele, prosseguiremos nossa análise. Um outro modelo, como a *spline* linear também seria possível (ver Figura 4.14-c).

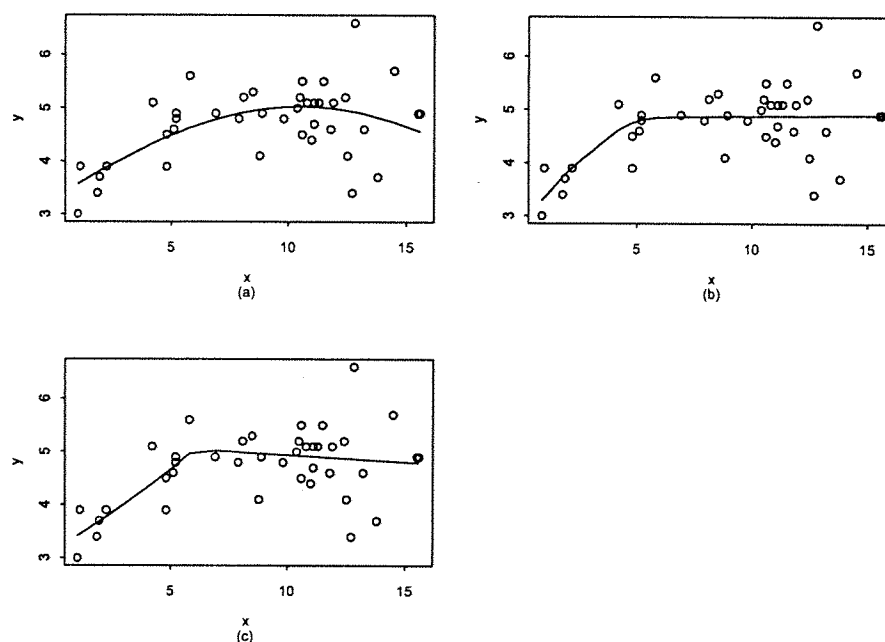


Figura 4.14: Gráfico de dispersão com curvas ajustadas para os modelos gama quadrático(a) e *spline* quadrático-linear(b).

A escolha do ponto de mudança (*knot* em $t = 6$) privilegiou a minimização do desvio escalonado e informações contidas nos próprios dados, que foram obtidas ao utilizarmos o suavizador *spline* no gráfico de dispersão (ver Figura 4.16), obtido diretamente no S-Plus através do comando `lines(smooth.spline(x,y))`.

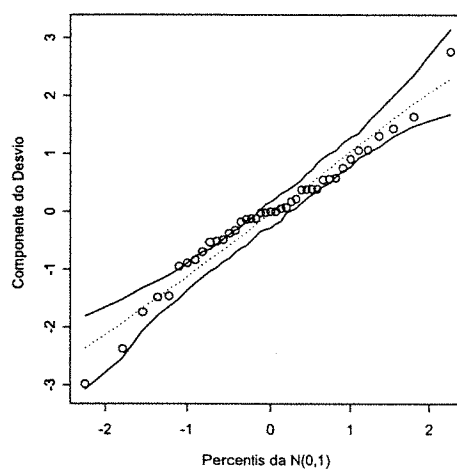


Figura 4.15: Gráfico normal de probabilidades para o modelo gama *spline* quadrático-linear.

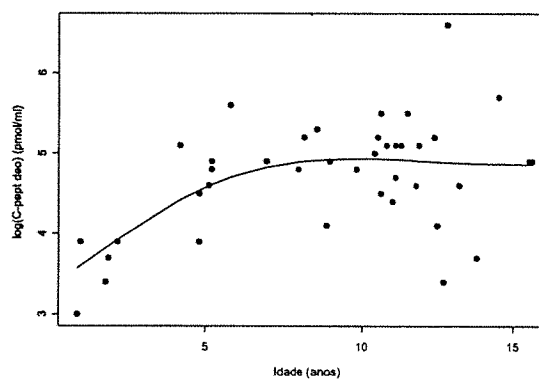


Figura 4.16: Diagrama de dispersão para os dados de log(C-peptídeo) contra a idade com suavizador *spline* ajustado.

Gráficos de diagnóstico para o modelo gama *spline* quadrático-linear estão na Figura 4.17 e destacam quatro observações: #1, #2, #35 e #36. As duas primeiras como candidatas a pontos de alavanca, as duas últimas como aberrantes e as de número #2 e #36 como influentes. Realizando ajustes após eliminarmos uma observação por vez, obtemos variações nas estimativas dos parâmetros que podem ser conferidas na Tabela 4.15. Uma vez que nosso conjunto de dados compõem-se de uma amostra de $n = 41$ crianças, espera-se que cada observação tenha uma contribuição individual de 2,44% ($1/41 = 0,0244$). Logo, vemos que a 2ª observação é a que mais influencia as estimativas dos parâmetros. Contudo, a retirada individual de cada ponto não leva a mudanças inferenciais.

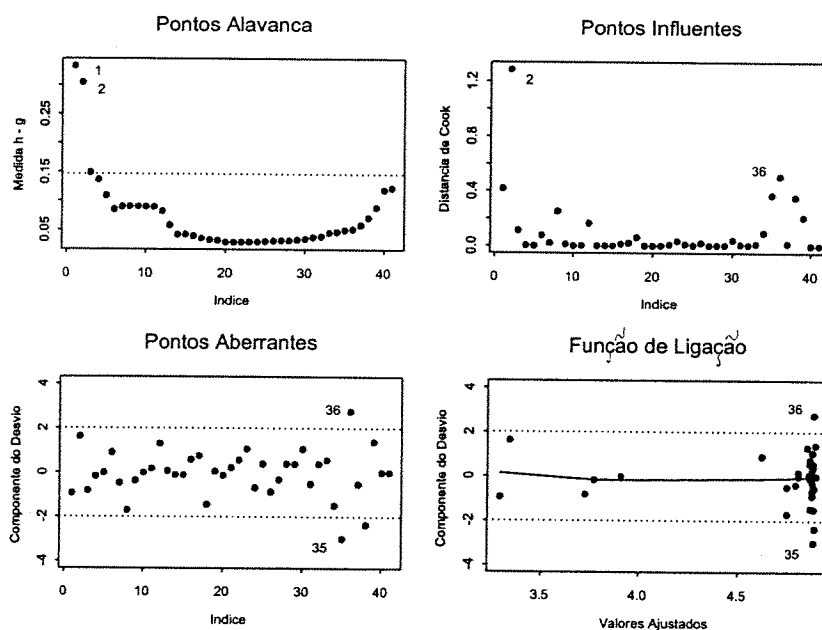


Figura 4.17: Gráficos de diagnóstico para o modelo gama *spline* quadrático-linear restrito, $t = 6$.

O envelope com a eliminação da 2ª observação do conjunto de dados está na Figura 4.18. Nota-se que com a exclusão dessa observação, há interferências no modelo adotado, uma vez que os pontos não se acomodam da mesma forma que o conjunto de dados completo (Figura 4.15). Os demais gráficos de diagnóstico (Figura 4.19) apontam os mesmos pontos anteriores (a defasagem na numeração das observações é devido à exclusão do ponto #2). As estimativas para o modelo ajustado sem a observação #2 estão na Tabela 4.16.

Embora não tenha havido uma evidente superioridade do modelo gama *spline* quadrático-linear restrito em relação ao gama polinomial quadrático, o modelo restrito reflete melhor o comportamento dos dados observados na Figura 4.16.

Tabela 4.15 Variação nas estimativas dos parâmetros do modelo gama *spline* quadrático-linear com *knot* em $t = 6$, ao excluirmos uma observação de cada vez.

| Parâmetro | sem #1 | sem #2 | sem #35 | sem #36 |
|-----------|--------|---------|---------|---------|
| β_0 | -5,80% | 10,98% | -0,59% | 0,75% |
| β_1 | 12,51% | -23,27% | 2,88% | -3,64% |
| β_2 | 12,98% | -24,07% | 4,36% | -5,43% |
| β_3 | 12,98% | -24,07% | 4,36% | -5,43% |

Tabela 4.16 Estimativas dos parâmetros referentes ao modelo gama *spline* quadrático-linear com $t = 6$, após excluirmos a 2ª observação.

| Parâmetro | Estimativa (E) | D.Padrão | E/D.padrão | P-valor |
|-------------------|----------------|----------------------|------------|---------|
| β_0 | 0,93 | 0,13 | 7,30 | 0,00 |
| β_1 | 0,22 | 0,06 | 3,73 | 0,00 |
| β_2 | -0,02 | 6×10^{-3} | -3,06 | 0,00 |
| β_3 | 0,02 | $7,6 \times 10^{-3}$ | 2,44 | 0,02 |
| ϕ | 75,57 | 41,09 | 4,48 | 0,00 |
| Desvio escalonado | 41,09438 | | | |

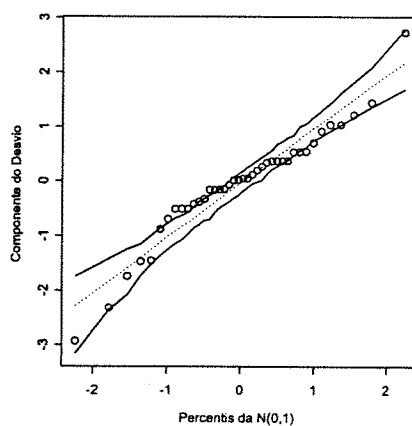


Figura 4.18: Gráfico normal de probabilidades para o modelo gama *spline* quadrático-linear com a 2ª observação excluída.

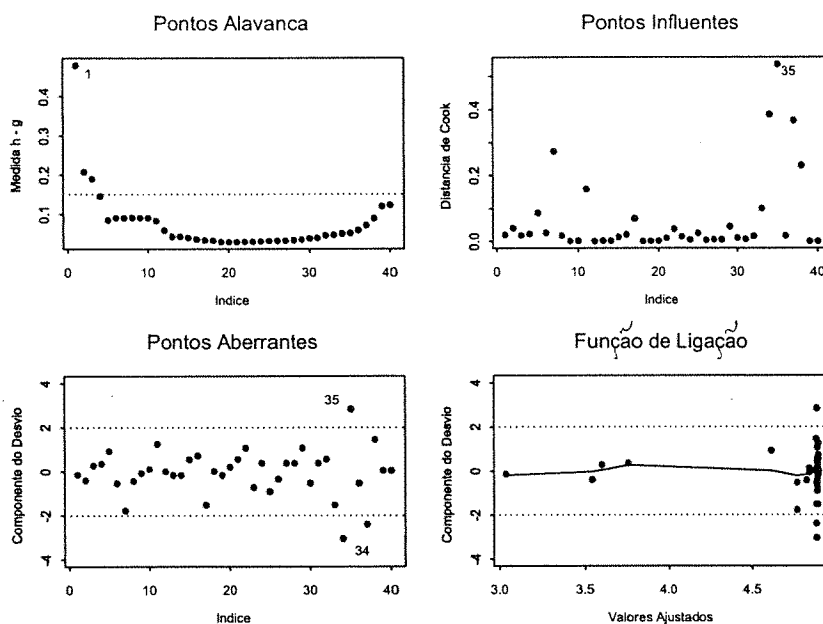


Figura 4.19: Gráficos de diagnóstico para o modelo gama *spline* quadrático-linear restrito, $t = 6$, com a 2^a observação excluída.

Capítulo 5

Considerações Finais

5.1 Conclusões

Esse trabalho iniciou-se com a leitura do artigo de Tarpey (2000), que desenvolve a derivação da medida PRESS para modelos normais restritos. Nesse artigo, o autor ilustra com um exemplo de *spline* restrita, o que nos instigou a trabalhar com esses tipos de modelagem. Para os modelos segmentados normal linear e linear generalizado (MLG) irrestrito, as medidas de diagnóstico já estavam desenvolvidas (vide por exemplo, Paula, 2004), porém necessitávamos de resultados para os casos restritos. Então, fizemos extensões de algumas medidas tradicionais e adaptamos os programas de análise de diagnóstico e geração de envelopes para o caso em que restrições na forma de igualdades lineares eram impostas. Em especial, desenvolvemos a expressão para a estimativa do vetor de parâmetros de um modelo normal restrito quando o i -ésimo ponto é excluído, dada por (2.35).

A construção dos modelos polinomiais segmentados necessita de inúmeros ajustes, incluindo-se ou retirando-se termos, variando-se os graus dos segmentos e comparando-os com polinômios mais simples. A determinação do(s) ponto(s) de mudança (*knot*) quando

não há informação teórica prévia que justifique a sua localização foi também obtida por “tentativa e erro”. As propostas para o “melhor” modelo buscaram sempre um parecer obtido através das informações das medidas de qualidade de ajuste, significância das estimativas, gráficos de homoscedasticidade e de envelopes e bom senso.

Durante o desenvolvimento dos exemplos notamos que nem sempre as medidas de qualidade de ajuste como PRESS e desvio levavam necessariamente aos melhores ajustes. O uso de envelopes conjuntamente, traz uma visão global do modelo e muitas vezes foi o fator preponderante.

5.2 Estudos futuros

Algumas propostas de pesquisas futuras foram surgindo ao longo do trabalho. Algumas delas incluem:

- 1) Estender o tipo de segmentação apresentada para outros modelos (modelos elípticos, por exemplo);
- 2) Obter a expressão para medidas de influência local em modelos restritos na forma de igualdades e desigualdades lineares. Também outras medidas de diagnóstico para esses modelos poderão ser desenvolvidas;
- 3) Estimar os *knots*;
- 4) Adaptar os resultados para casos em que n é pequeno;
- 5) Aplicar a teoria apresentada em situações que envolvam outras distribuições, assim como, em modelos com dados categorizados.

Apêndice A

Resultados utilizados

A.1 Teorema de Sherman-Morrison-Woodbury

Considere uma matriz quadrada não-singular \mathbf{A} , de dimensão $p \times p$, e um vetor coluna p -dimensional. Pelo Teorema de Sherman-Morrison-Woodbury (vide, por exemplo, Rao, 1973), temos que

$$(\mathbf{A} - \mathbf{z}\mathbf{z}^T)^{-1} = \mathbf{A}^{-1} + \frac{\mathbf{A}^{-1}\mathbf{z}\mathbf{z}^T\mathbf{A}^{-1}}{1 - \mathbf{z}^T\mathbf{A}^{-1}\mathbf{z}}$$

Para nosso interesse, em particular, \mathbf{A} é a matriz $\mathbf{X}^T\mathbf{X}$ e o vetor \mathbf{z}^T é a i -ésima linha da matriz \mathbf{X} . Assim, a matriz $(\mathbf{A} - \mathbf{z}\mathbf{z}^T)$ torna-se a matriz $\mathbf{X}^T\mathbf{X}$ com o i -ésimo ponto dos dados excluídos.

A.2 Resultado 1

$I - H^*$ é simétrica e idempotente

$$\begin{aligned} (I - H^*)(I - H^*) &= I - H^* - H^* + H^*H^* \\ &= I - H^* - H^* + H^* \\ &= I - H^*. \end{aligned}$$

uma vez que H^* também é simétrica e idempotente. (Seção 3.4.1)

A.3 Resultado 2

A matriz $(I - H^* + G^*)$ é simétrica e idempotente, como demonstrado abaixo:

$$\begin{aligned} (I - H^* + G^*)(I - H^* + G^*) &= I - H^* + G^* - H^* + H^*H^* - H^*G^* + G^* - G^*H^* \\ &\quad + G^*G^* \\ &= I - H^* + G^* - H^* + H^* - G^* + G^* - G^* + G^* \\ &= I - H^* + G^*, \end{aligned}$$

uma vez que $H^*H^* = H^*$ e $G^*H^* = H^*G^* = G^*G^* = G^*$. (Seção 3.4.2)

Apêndice B

Dados dos Exemplos do Capítulo 4

B.1 Crescimento de crianças

Tabela B.1 Medidas da razão entre o peso e a altura de crianças em libras/cm (y) e idades em meses (x)

| Obs | x | y | Obs | x | y | Obs | x | y |
|-----|------|------|-----|------|------|-----|------|------|
| 1 | 0,5 | 0,46 | 13 | 12,5 | 0,75 | 25 | 24,5 | 0,88 |
| 2 | 1,5 | 0,47 | 14 | 13,5 | 0,80 | 26 | 25,5 | 0,81 |
| 3 | 2,5 | 0,56 | 15 | 14,5 | 0,78 | 27 | 26,5 | 0,83 |
| 4 | 3,5 | 0,61 | 16 | 15,5 | 0,82 | 28 | 27,5 | 0,82 |
| 5 | 4,5 | 0,61 | 17 | 16,5 | 0,77 | 29 | 28,5 | 0,82 |
| 6 | 5,5 | 0,67 | 18 | 17,5 | 0,80 | 30 | 29,5 | 0,86 |
| 7 | 6,5 | 0,68 | 19 | 18,5 | 0,81 | 31 | 30,5 | 0,82 |
| 8 | 7,5 | 0,78 | 20 | 19,5 | 0,78 | 32 | 31,5 | 0,85 |
| 9 | 8,5 | 0,69 | 21 | 20,5 | 0,87 | 33 | 32,5 | 0,88 |
| 10 | 9,5 | 0,74 | 22 | 21,5 | 0,80 | 34 | 33,5 | 0,86 |
| 11 | 10,5 | 0,77 | 23 | 22,5 | 0,83 | 35 | 34,5 | 0,91 |
| 12 | 11,5 | 0,78 | 24 | 23,5 | 0,81 | 36 | 35,5 | 0,87 |

continua

Tabela B.1 Medidas ...

(continuação)

| Obs | x | y | Obs | x | y | Obs | x | y |
|-----|------|------|-----|------|------|-----|------|------|
| 37 | 36,5 | 0,87 | 49 | 48,5 | 0,92 | 61 | 60,5 | 0,97 |
| 38 | 37,5 | 0,87 | 50 | 49,5 | 0,96 | 62 | 61,5 | 0,94 |
| 39 | 38,5 | 0,85 | 51 | 50,5 | 0,92 | 63 | 62,5 | 0,96 |
| 40 | 39,5 | 0,90 | 52 | 51,5 | 0,91 | 64 | 63,5 | 1,03 |
| 41 | 40,5 | 0,87 | 53 | 52,5 | 0,95 | 65 | 64,5 | 0,99 |
| 42 | 41,5 | 0,91 | 54 | 53,5 | 0,93 | 66 | 65,5 | 1,01 |
| 43 | 42,5 | 0,90 | 55 | 54,5 | 0,93 | 67 | 66,5 | 0,99 |
| 44 | 43,5 | 0,93 | 56 | 55,5 | 0,98 | 68 | 67,5 | 0,99 |
| 45 | 44,5 | 0,89 | 57 | 56,5 | 0,95 | 69 | 68,5 | 0,99 |
| 46 | 45,5 | 0,89 | 58 | 57,5 | 0,97 | 70 | 69,5 | 1,01 |
| 47 | 46,5 | 0,92 | 59 | 58,5 | 0,97 | 71 | 70,5 | 0,99 |
| 48 | 47,5 | 0,89 | 60 | 59,5 | 0,96 | 72 | 71,5 | 1,04 |

B.2 Diabetes *mellitus*

Tabela B.2

Medidas de $\log(\text{C-peptideo})(y)$ de pacientes com diabetes *mellitus*
e idades (x)

| Obs | x | y | Obs | x | y | Obs | x | y |
|-----|-----|-----|-----|------|-----|-----|------|-----|
| 1 | 0,9 | 3,0 | 15 | 7,9 | 4,8 | 29 | 11,3 | 5,1 |
| 2 | 1,0 | 3,9 | 16 | 8,1 | 5,2 | 30 | 11,5 | 5,5 |
| 3 | 1,8 | 3,4 | 17 | 8,5 | 5,3 | 31 | 11,8 | 4,6 |
| 4 | 1,9 | 3,7 | 18 | 8,8 | 4,1 | 32 | 11,9 | 5,1 |
| 5 | 2,2 | 3,9 | 19 | 8,9 | 4,9 | 33 | 12,4 | 5,2 |
| 6 | 4,2 | 5,1 | 20 | 9,8 | 4,8 | 34 | 12,5 | 4,1 |
| 7 | 4,8 | 4,5 | 21 | 10,4 | 5,0 | 35 | 12,7 | 3,4 |
| 8 | 4,8 | 3,9 | 22 | 10,5 | 5,2 | 36 | 12,8 | 6,6 |
| 9 | 5,1 | 4,6 | 23 | 10,6 | 5,5 | 37 | 13,2 | 4,6 |
| 10 | 5,2 | 4,8 | 24 | 10,6 | 4,5 | 38 | 13,8 | 3,7 |
| 11 | 5,2 | 4,9 | 25 | 10,8 | 5,1 | 39 | 14,5 | 5,7 |
| 12 | 5,8 | 5,6 | 26 | 11,0 | 4,4 | 40 | 15,5 | 4,9 |
| 13 | 6,9 | 4,9 | 27 | 11,1 | 4,7 | 41 | 15,6 | 4,9 |
| 14 | 7,9 | 4,8 | 28 | 11,1 | 5,1 | | | |

Nota: os dados originais foram reordenados em função da variável x

Apêndice C

Programa para diagnóstico em modelos normais restritos

```
# Definir previamente: y(resposta), C(restrições), d(constantes), fit.model(ajuste lm)
# Ativar a library Matrix
X <- model.matrix(fit.model)
X <- as.Matrix(X)
Q <- solve(t(X)%*%X)
betahat <- Q%*%t(X)%*%y
n <- nrow(X)
p <- ncol(X)
q <- nrow(C)
H <- X%*%solve(t(X)%*%X)%*%t(X)
h <- diag(H)
betatil <- betahat + Q%*%t(C)%*%solve(C%*%Q%*%t(C))%*%(d - C%*%betahat)
G <- X%*%Q%*%t(C)%*%solve(C%*%Q%*%t(C))%*%C%*%Q%*%t(X)
g <- diag(G)
```

```
ytil <- (H-G)%*%y
eitol <- (y-ytil)/(1-h+g)
pressc <- sum(eitol^2)
Z <- X%*%Q%*%t(C)
Z <- as.Matrix(Z)
ZZi <- solve(t(Z)%*%Z)
v <- y-ytil
si <- lm.influence(fit.model)$sigma
tvis <- (v/si)*(1-h+g)^(-0.5)
lms <- summary(fit.model)
s <- lms$sigma
SQR <- sum(e*e)
SQRc <- SQR+t(C%*%betahat-d)%*%ZZi%*%(C%*%betahat-d)
sc2 <- SQRc/(n-p+q)
sc <- sqrt(sc2)
dic <- (v^2*(h-g))/((p-q)*sc^2*(1-h+g)^2)
win.graph()
par(mfrow=c(2,2))
plot(h-g,xlab="Indice", ylab="Medida h-g",
main="Pontos Alavanca", pch=16)
cut <- 2*(p-q)/n
abline(cut,0,lty=2)
identify(h-g,n=6)
plot(dic,xlab="Indice", ylab="Distancia de Cook ",
main="Pontos Influentes", pch=16)
identify(dic,n=7)
plot(tvis,xlab="Indice", ylab="Residuo padronizado",
```

```
main="Pontos Aberrantes", ylim=c(-4,4), pch=16)
  abline(2,0,lty=2)
  abline(-2,0,lty=2)
  identify(tvis,n=6)
  plot(ytil,tvis,xlab="valores ajustados",ylab="Residuo padronizado",
main="Homoscedasticidade",ylim=c(-4,4),pch=16)
  abline(2,0,lty=2)
  abline(-2,0,lty=2)
  identify(ytil,tvis,n=6)
  lines(smooth.spline(ytil,tvis))
```

Apêndice D

Programa para diagnóstico em modelos gama restritos

```
# definir previamente as quantidades: C, d e q como objetos do S-Plus
# implantar a sub-rotina: glmr.txt (Cysneiros, 1997)
# definir um objeto fmrest como sendo um ajuste glm Gamma restrito
# ativar librarys mass e Matrix
X <- model.matrix(fmrest)
X <- as.Matrix(X)
n <- nrow(X)
p <- ncol(X)
q <- nrow(C)
In <- diag(n)
mutil <- fmrest$fitted.values
Vtil <- mutil*mutil
etatil <- fmrest$linear.predictors
wtil <- fmrest$weights
```

```
wtilr <- sqrt(wtil)
wtilri <- 1/(sqrt(wtil))
Wtil <- diag(wtil)
Wtil <- diag(wtil)
Wmtil <- -diag(wtilr)
Wmmtil <- -diag(wtilri)
Wtil <- as.Matrix(Wtil)
Xstil <- -Wmtil%*%X
Xstil <- as.Matrix(Xstil)
Qstil <- -solve(t(Xstil)%*%Xstil)
Hstil <- -Xstil%*%Qstil%*%t(Xstil)
hstil <- diag(Hstil)
Zs <- -Xstil%*%Qstil%*%t(C)
Zs <- as.Matrix(Zs)
ZZsi <- -solve(t(Zs)%*%Zs)
Gstil <- -Zs%*%ZZsi%*%t(Zs)
gstil <- diag(Gstil)
Vvtil <- -mutil*mutil
Vvtil <- -1/sqrt(Vvtil)
Vtil <- -diag(Vvtil)
Vrtil <- -sqrt(Vvtil)
Vmtil <- -diag(Vrtil)
Vmmtil <- -diag(Vvtil)
rpStil <- -Vmmtil%*%(y-mutil)
ztil <- -etatil+Wmmtil%*%rpStil
ztils <- -Wmtil%*%ztil
vs <- -(In-Hstil+Gstil)%*%ztils
```



```
fital <- gamma.shape(fmrest)$alpha
tvis <- vs*sqrt(fital/(1-hstil+gstil))
tds <- resid(fmrest,type="deviance")*sqrt(fital/(1-hstil+gstil))
dis <- ((hstil-gstil)/(1-hstil+gstil))*tvis*tvis
win.graph()
par(mfrow=c(2,2))
plot(hstil-gstil,xlab="Indice", ylab="Medida h - g", main="Pontos Alavanca",
pch=16)
cut <- 2*(p-q)/n
abline(cut,0,lty=2)
identify(hstil-gstil,n=6)
plot(dis,xlab="Indice", ylab="Distancia de Cook", main="Pontos Influentes",pch=16)
identify(dis,n=6)
plot(tds,xlab="Indice", ylab="Componente do Desvio",
main="Pontos Aberrantes", ylim=c(-4,4),pch=16)
abline(2,0,lty=2)
abline(-2,0,lty=2)
identify(tds,n=6)
plot(fitted(fmrest),tds,xlab="Valores Ajustados", ylab="Componente do Desvio",
main="Funcao de Ligacao", ylim=c(-4,4), pch=16)
abline(2,0,lty=2)
abline(-2,0,lty=2)
identify(fitted(fmrest),tds,n=5)
lines(smooth.spline(fitted(fmrest),tds))
```

Apêndice E

Programas de Envelopes

Apresentamos neste Apêndice alguns programas de envelopes usados para gerar os gráficos normais de probabilidades para as distribuições normal e gama para os casos onde há restrições nos parâmetros. Pequenas adaptações podem ser feitas para obter-se esses envelopes para outras distribuições da família exponencial. Nesses programas são geradas 100 amostras e o coeficiente da banda de confiança gerada que é de 90%. Para estimar-se os parâmetros de dispersão e os coeficientes da regressão na distribuição gama, é necessário que as librarys mass e matrix sejam disponibilizadas no programa antes de obtermos os envelopes. Também a macro `glm.rest`(Cysneiros, 1997) foi anteriormente implementada no S-Plus.

DISTRIBUIÇÃO NORMAL

```
X <- model.matrix(fit.model)
X <- as.Matrix(X)
n <- nrow(X)
p <- ncol(X)
Q <- solve(t(X)%*%X)
betahat <- Q%*%t(X)%*%y
```

```

betatil <- betahat + Q%*%t(C)%*%solve(C%*%Q%*%t(C))%*%(d-C%*%betahat)
H <- X%*%Q%*%t(X)
h <- diag(H)
G <- X%*%Q%*%t(C)%*%solve(C%*%Q%*%t(C))%*%C%*%Q%*%t(X)
g <- diag(G)
s <- summary(fit.model)$sigma
si <- lm.influence(fit.model)$sigma
v <- y-X%*%betatil
tvis <- v/(si*sqrt(1-h+g))
#
ident <- diag(n)
epsilon <- matrix(0,n,100)
e <- matrix(0,n,100)
e1 <- numeric(n)
e2 <- numeric(n)
#
for ( i in 1:100) {
epsilon[,i] <- rnorm(n,0,1)
e[,i] <- (ident - H + G)%*%epsilon[,i]
u <- diag(ident - H + G)
e[,i] <- e[,i]/sqrt(u)
e[,i] <- sort(e[,i]) }
#
for ( i in 1:n) {
eo <- sort(e[i,])
e1[i] <- eo[5]
e2[i] <- eo[95] }

```

```

#
med <- apply(e,1,mean)
faixa <- range(tvis,e1,e2)
par(pty='s ')
qqnorm(tsi, xlab='Percentis da N(0,1)',
+ ylab = 'Residuo Studentizado ', ylim=faixa)
par(new=T)
qqnorm(e1,axes=F,xlab='', ylab= '', type='l ', ylim=faixa, lty=1)
par(new=T)
qqnorm(e2,axes=F,xlab='', ylab='', type='l ', ylim=faixa, lty=1)
par(new=T)
qqnorm(med,axes=F,xlab='', ylab='', type='l ', ylim=faixa, lty=2)

```

DISTRIBUÇÃO GAMA

```

X <- model.matrix(fit.model)
X <- as.Matrix(X)
n <- nrow(X)
p <- ncol(X)
w <- fit.model$weights
W <- diag(w)
H <- solve(t(X)%*%W%*%X)
H <- sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
h <- diag(H)
ro <- resid(fit.model,type="response")
fi <- gamma.shape(fit.model)$alpha
td <- resid(fit.model,type="deviance")*sqrt(fi/(1-h))
#

```

```
e <- matrix(0,n,100)
for (i in 1:100) {
  resp <- rgamma(n,fi)
  resp <- (fitted(fit.model)/fi)*resp
  fit <- glm( resp ~ -1+ X, family=Gamma(link=ligacao),method='glm.rest',
Cres=Cres,sol=sol)
  w <- fit$weights
  W <- diag(w)
  H <- solve(t(X)%%W%%X)
  H <- sqrt(W)%%X%%H%%t(X)%%sqrt(W)
  h <- diag(H)
  ro <- resid(fit, type= "response ")
  phi <- gamma.shape(fit)$alpha
  e[,i] <- sort(resid(fit, type= "deviance")*sqrt(phi/(1-h))) }
#
e1 <- numeric(n)
e2 <- numeric(n)
#
for (i in 1:n) {
  eo <- sort(e[i,])
  e1[i] <- eo[5]
  e2[i] <- eo[95]
#
med <- apply(e,1,mean)
faixa <- range(td,e1,e2)
#
par(pty= "s ")
```

```
qqnorm(td, xlab="Percentis da N(0,1)", ylab="Componente do Desvio  
+ Padronizado", ylim=faixa)  
par(new=T)  
qqnorm(e1, axes=F, xlab="", ylab="", type="l", ylim=faixa, lty=1)  
par(new=T)  
qqnorm(e2, axes=F, xlab="", ylab="", type="l", ylim=faixa, lty=1)  
par(new=T)  
qqnorm(med, axes=F, xlab="", ylab="", type="l", ylim=faixa, lty=2)
```

Bibliografia

- Akaike, H. (1974). A new look at statistical model identification. *IEEE Transactions on Automatic Control* **AU-19** 716-722.
- Allen, D. M. (1974). The relationship between variable selection and data augmentation and a method for prediction. *Technometrics* **16**, 125-127.
- Atkinson, A. C. (1981). Two graphical display for outlying and influential observations in regression. *Biometrika* **68**, 13-20.
- Atkinson, A. C. (1985). *Plots, Transformations and Regressions*. Oxford Statistical Science Series, Oxford.
- Belsley, D. A., Kuh, E. e Welsch, R. E. (1980). *Regression Diagnostics*. John Wiley, New York.
- Buse, A., e Lim, L. (1977). Cubic splines as a special case of restricted least squares. *Journal of the American Statistical Association*, **72**, 64-68.
- Cook, R. D. (1977). Detection of influential observations in linear regressions. *Technometrics* **19**, 15-18.
- Cook, R. D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society B* **48**, 133-169.

- Cook, R. D. e Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall, London.
- Cordeiro, G. M. (1986). *Modelos Lineares Generalizados*. Livro texto de minicurso, VII Simpósio Nacional de Probabilidade e Estatística, UNICAMP, Campinas, SP.
- Cordeiro, G. M. e McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society B* **53**, 629-643.
- Cysneiros, F. J. A. (1997). *Estimação e testes em modelos lineares generalizados com restrições nos parâmetros na forma de desigualdades lineares*. Dissertação de Mestrado, IME-USP.
- David, H. A. (1981). *Order Statistics*. John Wiley, New York.
- Draper, N. R. e Smith, H. (1981). *Applied regression analysis*. John Wiley, New York.
- Ellenberg, J. H. (1973). The joint distribution of the standardized least squares residuals from a general linear regression. *J. Amer. Statist. Assoc.* **68**, 941-943.
- Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression*. Marcel Dekker, New York.
- Fahrmeir, L. e Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Annals of Statistics* **13**, 342-368.
- Gallant, A. R., e Fuller, W. A. (1973). Fitting segmented polynomial regression models whose join points have to be estimated. *Journal of the American Statistical Association* **68**, 144-147.
- Hastie, T. e Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall, London.

- Hosmer, D. W. e Lemeshow, S. (1989). *Applied Logistic Regression*. John Wiley, New York.
- Margolin, B. H. (1977). The distribution of internally studentized statistics via Laplace transform inversion. *Biometrika* **64**, 573-82.
- McCullagh, P. e Nelder, J. A. (1989). *Generalized Linear Models, 2nd. Edition*. Chapman and Hall, London.
- Montgomery, D. C. e Peck, E. A. (1992). *Introduction to Linear Regression Analysis*. John Wiley, New York.
- Nelder, J. A. e Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society A* **135**, 370-384.
- Nyquist, H. (1991). Restricted estimation of restricted generalized linear models. *Applied Statistics* **40**, 133-141.
- Paula, G. A. (1993). Assessing local influence in restricted regression models. *Computational Statistics and Data Analysis* **16**, 63-79.
- Paula, G. A. (1995). Influence and residuals in restricted generalized linear models. *Journal of Statistical Computation and Simulation* **51**, 315-352.
- Paula, G. A. (1997). *Estimação e Testes em Modelos de Regressão com Parâmetros Restritos*. Livro texto de minicurso da 5ª Escola de Modelos de Regressão, realizada de 26 a 28-02-97 em Campos do Jordão, SP.
- Paula, G. A. (1999). Leverage in inequality constrained regression models. *The Statistician* **48**, 529-538.
- Paula, G. A. (2004). Modelos de regressão com apoio computacional. Versão preliminar, IME-USP, 253 p.

- Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics* **9**, 705-724.
- Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. John Wiley, New York.
- Ryan, B. F. e Joiner, B. L. (1994). *Minitab Handbook, Third Edition*. Duxbury Press, Belmont.
- Smith, P. L. (1979). Splines as a useful and convenient statistical tool. *The American Statistician* **33**, 57-62.
- Socketk, E. B., Daneman, D., Clarson, C. e Ehrich, R. M. (1987). Factor affecting and patterns of residual insulin secretion during the first year of type I (insulin dependent) diabetes mellitus in children. *Diabet.* **30**, 453-459.
- Tarpey, T. (2000). A note on the prediction sum of squares statistic for restricted least squares. *The American Statistician* **54**, 116-118.
- Tarpey, T. e Holcomb, J. (2000). Spline Bottles. *The American Statistician* **54**, 129-135.
- Wei, B.C., Hu, Y.Q. e Fung, W.K. (1998). Generalized leverage and its applications. *Scandinavian Journal of Statistics* **25**, 25-37.
- Williams, D. A. (1984). Residuals in generalized linear models. *In: Proceedings of the 12th. International Biometrics Conference, Tokyo*, pp. 59-68.
- Williams, D. A. (1987). Generalized linear model diagnostic using the deviance and single case deletion. *Applied Statistics* **36**, 181-191.
- Wold, S. (1971). Analysis of kinetic data by means of spline functions. *Chemica Scripta* **1**, 97-102.
- Wold, S. (1974). Spline functions in data analysis. *Technometrics* **16**, 1-11.