**Modelos não lineares
de família exponencial revisitados**

Adriana Alvarez Possamai

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
MESTRE EM CIÊNCIAS

Programa: Estatística

Orientador: Prof. Dr. Gilberto Alvarenga Paula

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro do CNPq

São Paulo, outubro de 2009

Modelos não lineares de família exponencial revisitados

Este exemplar corresponde à redação
final da dissertação devidamente corrigida
e defendida por Adriana Alvarez Possamai
e aprovada pela Comissão Julgadora.

Banca Examinadora:

- Prof. Dr. Gilberto Alvarenga Paula (orientador) - IME-USP.
- Profa. Dra. Silvia Nagib Elian - IME-USP.
- Profa. Dra. Hildete Prisco Pinheiro - UNICAMP.

Dedico este trabalho aos meus pais, Terilio e Marialice.

Agradecimentos

Primeiramente agradeço a Deus por ter me dado saúde e coragem.

Aos meus pais por terem conseguido me dar uma boa educação, a minha irmã e o Rafael por sempre estarem presente nas horas difíceis.

Ao meu orientador Prof. Gilberto Alvarenga Paula pelo apoio, compreensão e sugestões.

Ao meu amigo Artur que sempre procurou me ajudar com muita paciência quando tinha dificuldades com o LaTeX e em algumas demonstrações.

Aos meus amigos que conheci durante o mestrado e que me ajudaram sempre que precisei: Fernando, Núbia, Gleiciane, Sandro, Gabriela, Marcos Paulo, Camila, Patricia, Jony,...etc, que compartilharam de alegrias e aflições deste período.

Ao Henrique que ajudou nas dúvidas que tive com o R e ao Enzo, meu chefe, que me liberou várias vezes do trabalho para que eu pudesse terminar a dissertação.

Resumo

O objetivo deste trabalho é fazer uma revisão dos modelos não lineares de família exponencial (Cordeiro & Paula (1989); Wei (1998)) para respostas independentes e apresentar possíveis extensões para o caso de dados correlacionados. Inicialmente são apresentados exemplos ilustrativos, alguns dos quais são reanalisados ao longo do texto. Em seguida são discutidos procedimentos de estimação e testes de hipóteses, tais como apresentação de um processo de estimação que pode ser adaptado ao processo iterativo usado na classe dos modelos lineares generalizados, e alguns resultados assintóticos. Técnicas usuais de diagnóstico, como pontos de alavanca, análise de resíduos e diagnóstico de influência são adaptados para a classe dos modelos não lineares de família exponencial. Extensões para a classe dos modelos não lineares com resposta binomial negativa são também apresentadas. Finalmente, são consideradas duas possíveis extensões dos modelos não lineares de família exponencial para dados correlacionados, através de equações de estimação generalizadas e através de modelagem mista em que efeitos aleatórios em forma linear são adicionados ao componente não linear da parte sistemática do modelo conforme sugerido recentemente por Tang et al. (2006a).

Abstract

The aim of this work is to present a review of the exponential family nonlinear models (Cordeiro & Paula (1989); Wei (1998)) for independent responses and to present possible extensions for the case of correlated data. Firstly, illustrative examples are presented with some of them being reanalyzed along the text. Then, estimation and hypothesis testing procedures, such as the presentation of an iterative process adapted from the one of generalized linear models, and some asymptotic results are discussed. Useful diagnostic techniques, as calculation of leverage measures, residual analysis and influence diagnostics are adapted for the class of exponential family nonlinear models. Extensions to nonlinear negative binomial models are also presented. Finally, two possible extensions for correlated data are considered, by using generalized estimating equations and mixed modeling in which linear random effects are added into the systematic component together with the nonlinear function, as suggested by Tang et al. (2006a).

Sumário

1	Introdução	1
1.1	Exemplos de Motivação	2
1.1.1	Coelhos Europeus	2
1.1.2	Sobrevivência de Pacientes com Leucemia	4
1.1.3	Produção de Gramíneas	7
1.1.4	Mistura de Inseticidas	9
1.1.5	Cálcio Radioativo	11
1.1.6	Produção de Vendas	12
1.1.7	Casos de Câncer de Pulmão	15
1.1.8	Crescimento de Colônias	16
2	Modelos Não Lineares de Família Exponencial	21
2.1	Família Exponencial	21
2.2	Modelos Não Lineares de Família Exponencial	22

2.3	Função Desvio	25
2.4	Estimação dos Parâmetros do Modelo	26
2.4.1	Estimação de β	26
2.4.2	Estimação de ϕ	30
2.5	Distribuição Assintótica de $\hat{\beta}$ e $\hat{\phi}$	32
2.6	Testes de Hipóteses	34
2.7	Região de Confiança Assintótica	38
2.8	Viés de ordem n^{-1}	39
2.9	Binomial Negativa Não Linear	41
3	Métodos de Diagnóstico	47
3.1	Pontos de Alavanca	49
3.2	Resíduos	51
3.3	Influência	54
3.4	Influência Local	57
3.5	Métodos de Diagnóstico na Binomial Negativa	66
4	Aplicações	69
4.1	Coelhos Europeus	69
4.2	Crescimento de Colônias	75

4.3	Cálcio Radioativo	83
4.4	Sobrevivência de Pacientes com Leucemia	87
4.5	Produção de Gramíneas	94
5	Extensões para Dados Correlacionados	97
5.1	Equações de Estimação Generalizadas	97
5.2	Estimação de β	98
5.3	Estruturas de Correlação	100
5.3.1	Não Estruturada	100
5.3.2	Estruturada	100
5.3.3	Modelos Não Lineares de Família Exponencial Mistos	101
6	Conclusões	109
A	Aspectos Computacionais	111
A.1	Coelhos Europeus	111
A.2	Crescimento de Colônias	116
A.3	Cálcio Radioativo	121
A.4	Sobrevivência de Pacientes com Leucemia	124
A.5	Produção de Gramíneas	130

Lista de Tabelas

1.1	Dados dos coelhos europeus ajustados pelo modelo de regressão normal inversa não linear, em que x denota a idade (em dias) e y o peso das lentes (em mg).	4
1.2	Porcentagem de pacientes sobreviventes e não sobreviventes nos grupos com AG positivo (=1) e AG negativo (=0).	5
1.3	Dados de leucemia ajustados pelo modelo de regressão logístico não linear, em que WBC denota o número de células brancas no sangue, AG corresponde à condição morfológica (1:sobrevive, 0: não sobrevive) e n as repetições.	6
1.4	Níveis dos três fatores na produção de gramíneas no litoral de Bermuda.	8
1.5	Produção de gramíneas no litoral de Bermuda segundo os níveis de Nitrogênio, Fósforo e Potássio.	8
1.6	Dados do ensaio com os gafanhotos para o modelo de regressão logístico não linear. . .	10
1.7	Dados da quantidade absorvida de cálcio radioativo para o modelo de regressão normal não linear.	12
1.8	Dados referentes à projeção de vendas e à produção atual de vendas de uma empresa.	15

1.9	Dados sobre estudo de seguimento com médicos britânicos. O nº de mortes por câncer de pulmão é relacionado com o consumo diário de cigarros e a idade.	17
1.10	Descrição do nº de <i>paramecia</i> ao longo do tempo em três colônias de <i>Paramecium aurelium</i> submetidas a um meio nutritivo.	19
2.1	Principais distribuições pertencentes à família exponencial.	23
4.1	Estimativas de máxima verossimilhança com os respectivos erros padrão aproximados obtidos do modelo normal inversa ajustado aos dados sobre Coelhos Europeus.	70
4.2	Variação das estimativas do modelo normal inversa ao excluir as observações citadas na análise de diagnóstico.	74
4.3	Estimativas de máxima verossimilhança com os respectivos erros padrão aproximados obtidos do modelo binomial negativa ajustado aos dados sobre a Colônia A de <i>Paramecium aurelium</i>	75
4.4	Estimativas de máxima verossimilhança com os respectivos erros padrão aproximados obtidos do modelo binomial negativa ajustado aos dados sobre a Colônia B de <i>Paramecium aurelium</i>	76
4.5	Estimativas de máxima verossimilhança com os respectivos erros padrão aproximados obtidos do modelo binomial negativa ajustado aos dados sobre a Colônia C de <i>Paramecium aurelium</i>	76
4.6	Variação das estimativas do modelo binomial negativa para a colônia A ao excluir as observações citadas na análise de diagnóstico.	82

4.7	Variação das estimativas do modelo binomial negativa para a colônia B ao excluir as observações citadas na análise de diagnóstico.	82
4.8	Variação das estimativas do modelo binomial negativa para a colônia C ao excluir as observações citadas na análise de diagnóstico.	82
4.9	Estimativas de máxima verossimilhança com os respectivos erros padrão aproximados obtidos do modelo normal ajustado aos dados sobre Cálcio Radioativo.	83
4.10	Estimativas de máxima verossimilhança com os respectivos erros padrão aproximados obtidos do modelo logístico ajustado aos dados sobre pacientes com Leucemia considerando $\lambda = -0,3$	89
4.11	Estimativas de máxima verossimilhança com os respectivos erros padrão aproximados obtidos do modelo logístico ajustado aos dados sobre pacientes com Leucemia considerando $\lambda = 0$	89
4.12	Estimativas de máxima verossimilhança com os respectivos erros padrão aproximados obtidos do modelo gama ajustado aos dados sobre produção de gramíneas.	94

Lista de Figuras

1.1	Diagrama de dispersão entre o peso das lentes dos olhos (em mg) e a idade (em dias) dos coelhos europeus.	3
1.2	Boxplots da variável WBC para os grupos de pacientes com leucemia que não sobreviveram e que sobreviveram.	7
1.3	Produção de Gramíneas no litoral de Bermuda, em que Y é a produção média dos três anos (1955,1956,1957).	9
1.4	Proporção dos gafanhotos mortos submetidos a log doses de duas drogas, inseticida e sinergista.	11
1.5	Diagrama de dispersão entre a quantidade de cálcio absorvido e o tempo de suspensão.	13
1.6	Diagrama de dispersão entre a projeção de vendas e as vendas reais.	14
1.7	Crescimento de três colônias de <i>Paramecium aurelium</i> submetidas a um meio nutritivo.	18
4.1	Gráfico do modelo normal inversa ajustado aos dados sobre Coelhos Europeus.	71
4.2	Gráficos de diagnóstico referentes ao modelo normal inversa ajustado aos dados sobre Coelhos Europeus.	72

4.3	Gráfico normal de probabilidades (com envelope gerado) para o resíduo componente do desvio referente ao ajuste do modelo normal inversa aos dados sobre Coelhos Europeus.	73
4.4	Gráficos de diagnóstico referentes ao modelo binomial negativa ajustado aos dados sobre a Colônia A de <i>Paramecium aurelium</i>	77
4.5	Gráficos de diagnóstico referentes ao modelo binomial negativa ajustado aos dados sobre a Colônia B de <i>Paramecium aurelium</i>	78
4.6	Gráficos de diagnóstico referentes ao modelo binomial negativa ajustado aos dados sobre a Colônia C de <i>Paramecium aurelium</i>	79
4.7	Gráfico normal de probabilidades (com envelope gerado) para o resíduo componente do desvio referente ao ajuste do modelo binomial negativa aos dados sobre a Colônia A de <i>Paramecium aurelium</i>	80
4.8	Gráfico normal de probabilidades (com envelope gerado) para o resíduo componente do desvio referente ao ajuste do modelo binomial negativa aos dados sobre a Colônia B de <i>Paramecium aurelium</i>	80
4.9	Gráfico normal de probabilidades (com envelope gerado) para o resíduo componente do desvio referente ao ajuste do modelo binomial negativa aos dados sobre a Colônia C de <i>Paramecium aurelium</i>	81
4.10	Gráfico do modelo normal ajustado aos dados sobre Cálcio Radioativo.	84
4.11	Gráficos de diagnóstico referentes ao modelo normal ajustado aos dados sobre Cálcio Radioativo.	85

4.12	Gráfico normal de probabilidades (com envelope gerado) para o resíduo componente do desvio referente ao ajuste do modelo normal aos dados sobre Cálcio Radioativo. . .	86
4.13	Gráfico dos valores dos AIC para cada valor de λ	88
4.14	Gráficos de diagnóstico referentes ao modelo logístico ajustado aos dados sobre pacientes com Leucemia considerando $\lambda = -0,3$	90
4.15	Gráfico normal de probabilidades (com envelope gerado) para o resíduo componente do desvio referente ao ajuste do modelo logístico aos dados sobre pacientes com Leucemia considerando $\lambda = -0,3$	91
4.16	Gráficos de diagnóstico referentes ao modelo logístico ajustado aos dados sobre pacientes com Leucemia considerando $\lambda = 0$	92
4.17	Gráfico normal de probabilidades (com envelope gerado) para o resíduo componente do desvio referente ao ajuste do modelo normal aos dados sobre pacientes com Leucemia considerando $\lambda = 0$	93
4.18	Gráficos de diagnóstico referentes ao modelo logístico ajustado aos dados sobre produção de gramíneas.	95
4.19	Gráfico normal de probabilidades (com envelope gerado) para o resíduo componente do desvio referente ao ajuste do modelo gama aos dados sobre produção de gramíneas. . .	96

Capítulo 1

Introdução

Os modelos não lineares de família exponencial são extensões da classe de modelos lineares generalizados (MLGs) que foram apresentados pela primeira vez por Nelder & Wedderburn (1972). Os autores estenderam o modelo normal linear permitindo alargar as hipóteses admitidas, ou seja, a primeira que a distribuição da variável resposta não necessariamente precisa ser normal, basta pertencer à família exponencial de distribuições, e a segunda que a relação entre o valor médio da variável resposta e a combinação linear dos valores das variáveis explicativas (preditor linear) pode ser estabelecida por qualquer função monótona e diferenciável, chamada função de ligação. Os modelos não lineares de família exponencial admitem preditores não lineares. Vários autores têm tido interesse na inferência desses modelos, dentre eles, tem-se: Cordeiro & Paula (1989) apresentaram uma estatística da razão de verossimilhanças; Cook & Tsai (1990) discutiram aproximações cúbicas para regiões de confiança; Paula (1992) derivou o viés de segunda ordem dos estimadores de máxima verossimilhança; Wei & Shi (1994) estudaram alguns problemas de diagnóstico; Svetliza & Paula (2003) estudaram alguns métodos de diagnóstico em modelos não lineares com resposta binomial negativa, enquanto Jørgensen (1983) e McCullagh (1983) estudaram alguns modelos gerais que incluem os modelos não lineares de família exponencial como casos especiais, Wei (1998) descreve os modelos não lineares de família exponencial com diversos tópicos adicionais tais como técnicas de diagnóstico, teste de hipóteses e medidas de curvatura. Mais recentemente, Kosmidis (2007) estudou a redução do viés nos modelos não lineares de família exponencial para o caso univariado/multivariado. Neste capítulo apresenta-

mos diversos exemplos como motivação do estudo de modelos não lineares de família exponencial. No Capítulo 2 introduzimos os modelos não lineares de família exponencial. Adicionalmente, apresentamos a função desvio, a estimação dos parâmetros, algumas propriedades assintóticas dos estimadores de máxima verossimilhança e testes de hipóteses. Estendemos os tópicos mencionados para os modelos não lineares com resposta binomial negativa. Técnicas usuais de diagnóstico, tais como medidas de pontos de alavanca, análise de resíduos e diagnóstico de influência, particularmente influência local são adaptadas no Capítulo 3 para os modelos não lineares de família exponencial com extensões para os modelos não lineares com resposta binomial negativa. No Capítulo 4 alguns dos exemplos apresentados neste capítulo são reanalisados e no Capítulo 5 discutimos algumas possíveis extensões dos modelos não lineares de família exponencial para dados correlacionados através de equações de estimação generalizadas e modelos mistos. Conclusões e trabalhos futuros são apresentados no último capítulo. Finalmente é apresentado no Apêndice os códigos computacionais escritos na linguagem de programação R que foram utilizados no Capítulo 4 para ajustar os modelos aos bancos de dados reais.

1.1 Exemplos de Motivação

A interpretação dos parâmetros dos modelos não lineares de família exponencial não é facilmente especificada.

1.1.1 Coelhos Europeus

Os dados desse exemplo foram originalmente apresentados em Dudzinski & Mykytowycz (1961) e estudados posteriormente por Ratkowsky (1983) baseados num modelo de regressão normal não linear. Os dados consistem num conjunto de 71 observações em que a variável resposta representa o peso das lentes (em mg) dos olhos de coelhos europeus (*Oryctolagus Cuniculus*) na Austrália e a variável explicativa corresponde à idade (em dias) dos coelhos. Wei (1998) ajustou esses dados utilizando um modelo não linear com resposta normal inversa. Dessa forma, $Y_i \sim \text{NI}(\mu_i, \sigma^2)$ em que

$$\mu_i = \beta_1 - \frac{\beta_2}{x_i + \beta_3},$$

com $i = 1, \dots, 71$, $\mu_i = E(Y_i)$, $\text{Var}(Y_i) = \sigma^2 V(\mu_i)$ com $V(\mu_i) = \mu_i^3$. Os dados desse exemplo encontram-se na Tabela 1.1. Observando a Figura 1.1 percebemos um crescimento não linear do peso dos olhos dos coelhos em função da idade dos coelhos. Além disso, a partir de 400 dias, não há aparentemente ganho no peso das lentes dos coelhos. A variabilidade da resposta também aumenta com a idade dos coelhos.

Figura 1.1: Diagrama de dispersão entre o peso das lentes dos olhos (em mg) e a idade (em dias) dos coelhos europeus.

Tabela 1.1: Dados dos coelhos europeus ajustados pelo modelo de regressão normal inversa não linear, em que x denota a idade (em dias) e y o peso das lentes (em mg).

x	y	x	y	x	y
15	21,66	98	104,30	285	189,66
15	22,75	125	134,90	300	186,09
15	22,30	142	130,68	301	186,70
18	31,25	142	140,58	305	186,80
28	44,79	147	155,30	312	195,10
29	40,55	147	152,20	317	216,41
37	50,25	150	144,50	338	203,23
37	46,88	159	142,15	347	188,38
44	52,03	165	139,81	354	189,70
50	63,47	183	153,22	357	195,31
50	61,13	192	145,72	375	202,63
60	81,00	195	161,10	394	224,82
61	73,09	218	174,18	513	203,30
64	79,09	218	173,03	535	209,70
65	79,51	219	173,54	554	233,90
65	65,31	224	178,86	591	234,70
72	71,90	225	177,68	648	244,30
75	86,10	227	173,73	660	231,00
75	94,60	232	159,98	705	242,40
82	92,50	232	161,29	723	230,77
85	105,00	237	187,07	756	242,57
91	101,70	246	176,13	768	232,12
91	102,90	258	183,40	860	246,70
97	110,00	276	186,26		

Fonte: Dudzinski & Mykytowycz (1961).

1.1.2 Sobrevivência de Pacientes com Leucemia

Os dados de leucemia desse exemplo foram estudados por Cook & Weisberg (1982), Lee (1987) e Lee (1988). Os dados consistem em uma amostra de 33 pacientes que morreram de leucemia aguda. Essa doença é caracterizada pela invasão da medula óssea por glóbulos brancos alterados que se tornam cancerígenos. Existem duas variáveis explicativas: a primeira, a contagem de células brancas no sangue (WBC), é a principal medida da condição inicial do paciente, condições mais severas sendo

Tabela 1.2: Porcentagem de pacientes sobreviventes e não sobreviventes nos grupos com AG positivo (=1) e AG negativo (=0).

	Sobrevive	Não-Sobrevive	%
$AG = 0$	2,01	97,99	100
$AG = 1$	49,91	50,09	100

refletidas por contagens altas; a segunda, classifica cada paciente como AG (=1) para pacientes positivos e AG (=0) para pacientes negativos, em que AG indica a presença ou não de uma certa característica morfológica em WBC. A variável resposta Y é binária (1: sobrevive; 0: não sobrevive) referente à sobrevivência do paciente pelo menos 52 semanas após o diagnóstico. O tamanho amostral considerado foi $n = 33$ [existem 5 pacientes com $WBC = 100000$ os quais foram colocados em dois grupos, um (caso 15) consiste de 3 pacientes com AG positivo (com um sobrevivente), e o outro, (caso 30) consiste em 2 pacientes com AG negativo (com nenhum sobrevivente)]. Cook & Weisberg (1982) ajustaram esses dados utilizando um modelo de regressão logístico linear. Como uma alternativa, Lee (1988) e Wei (1998) consideraram uma transformação na covariável WBC. Então, considerando essa transformação temos o modelo de regressão logístico não linear em que $Y_i \sim B(n_i, p_i)$ com

$$\log \left\{ \frac{p_i}{1 - p_i} \right\} = \beta_0 + \beta_1 AG_i + \beta_2 WBC_i^\lambda, \quad i = 1, \dots, 30,$$

em que $p_i = \Pr(Y_i = 1)$ denota a probabilidade de sobrevivência. Os dados desse exemplo encontram-se na Tabela 1.3.

Observando a Figura 1.2 percebemos uma leve assimetria da variável WBC no primeiro boxplot, possivelmente causada pelos dois pontos aberrantes: caso 29 e caso 30, respectivamente. Já no segundo boxplot apesar de existir uma assimetria, essa parece não ser causada pelo ponto aberrante, caso 15. É notável que a variabilidade no boxplot dos não sobreviventes é muito maior do que dos sobreviventes. Além disso, percebemos através da Tabela 1.2 que quase 98% dos pacientes que não apresentavam a característica morfológica, não sobreviveram.

Tabela 1.3: Dados de leucemia ajustados pelo modelo de regressão logístico não linear, em que WBC denota o número de células brancas no sangue, AG corresponde à condição morfológica (1:sobrevive, 0: não sobrevive) e n as repetições.

Caso	WBC	AG	y	n
1	2300	1	1	1
2	750	1	1	1
3	4300	1	1	1
4	2600	1	1	1
5	6000	1	0	1
6	10500	1	1	1
7	10000	1	1	1
8	17000	1	0	1
9	5400	1	0	1
10	7000	1	1	1
11	9400	1	1	1
12	32000	1	0	1
13	35000	1	0	1
14	52000	1	0	1
15	100000	1	1	3
16	4400	0	1	1
17	3000	0	1	1
18	4000	0	0	1
19	1500	0	0	1
20	9000	0	0	1
21	5300	0	0	1
22	10000	0	0	1
23	19000	0	0	1
24	27000	0	0	1
25	28000	0	0	1
26	31000	0	0	1
27	26000	0	0	1
28	21000	0	0	1
29	79000	0	0	1
30	100000	0	0	2

Fonte: Cook & Weisberg, 1982, p. 193.

Figura 1.2: Boxplots da variável WBC para os grupos de pacientes com leucemia que não sobreviveram e que sobreviveram.

1.1.3 Produção de Gramíneas

Os dados desse exemplo foram originalmente analisados por Welch et al. (1963) e reanalisados por McCullagh & Nelder (1989) e por Wei (1998). O principal interesse é estudar o resultado de experimentos fatoriais ⁴³ com os três principais nutrientes da planta (em lb/acre), x_1 : nitrogênio N, x_2 :

fósforo F e x_3 : potássio P, na produção de gramíneas no litoral de Bermuda, em que Y é a produção média de todos os três anos (1955, 1956, 1957). Os quatro níveis dos três fatores (todos em lb/acre) estão na Tabela 1.4. Observamos no histograma da Figura 1.3 que existe uma bi-modalidade nas classes entre 2 e 3 e 4 e 5 para a produção média de todos os três anos.

Tabela 1.4: Níveis dos três fatores na produção de gramíneas no litoral de Bermuda.

Níveis	1	2	3	4
N	0	100	200	400
F	0	22	44	88
P	0	42	84	168

Tabela 1.5: Produção de gramíneas no litoral de Bermuda segundo os níveis de Nitrogênio, Fósforo e Potássio.

Nitrogênio (N)	Fósforo (F)	Potássio (P)			
		0	1	2	3
0	0	1,98	2,13	2,19	1,97
0	1	2,38	2,24	2,10	2,60
0	2	2,18	2,56	2,22	2,47
0	3	2,22	2,47	2,94	2,48
1	0	3,88	3,91	3,66	4,07
1	1	4,35	4,59	4,47	4,55
1	2	4,14	4,36	4,55	4,35
1	3	4,26	4,72	4,83	4,85
2	0	4,40	4,91	5,10	5,23
2	1	5,01	5,64	5,68	5,60
2	2	4,77	5,69	5,80	6,07
2	3	5,17	5,45	5,85	6,43
3	0	4,43	5,31	5,15	5,87
3	1	4,95	6,27	6,49	6,54
3	2	5,22	6,27	6,35	6,72
3	3	5,66	6,24	7,11	7,32

Fonte: McCullagh & Nelder (1989).

Considerando o modelo gama não linear proposto por McCullagh & Nelder (1989) temos:

$$\mu_i^{-1} = \beta_0 + \beta_1 v_1 + \beta_2 v_2 + \beta_3 v_3,$$

Figura 1.3: Produção de Gramíneas no litoral de Bermuda, em que Y é a produção média dos três anos (1955,1956,1957).

em que $v_i = 1/(x_i + \alpha_i)$, $i = 1, 2, 3$ e $Y_i \sim G(\mu_i, \phi)$. É interessante notar que foi utilizada a ligação recíproca (canônica). Aqui x_i ($i = 1, 2, 3$) são os valores usados de N, F e P, respectivamente, enquanto que α_i são valores desconhecidos no solo. Os dados deste exemplo encontram-se na Tabela 1.5 com os níveis do fator codificados em 0, 1, 2 e 3.

1.1.4 Mistura de Inseticidas

Os dados desse exemplo estão disponíveis em McCullagh & Nelder (1989) que consiste na estimação da mistura do menor gasto de inseticidas e sinergistas. Eles analisaram uma espécie de gafanhoto (*Melanopus sanguinipes*) com o inseticida carbofuran e o sinergista piperonyl butoxide (PB), que aumenta a toxicidade do inseticida. O modelo de regressão logístico não linear, em que a variável

resposta Y_i é o número de gafanhotos mortos, sugerido por Hewlett (1969) assume a seguinte forma:

$$\log \left\{ \frac{p_i}{1 - p_i} \right\} = \alpha + \beta_1 x_{1i} + \frac{\beta_2}{\delta + x_{2i}},$$

em que $Y_i \sim B(m_i, p_i)$, x_{1i} é o log da dose do inseticida, x_{2i} é a log da dose do sinergista PB e p_i é a probabilidade de morte do i -ésimo gafanhoto submetido às dosagens (x_{1i}, x_{2i}) . O efeito do sinergista é dessa maneira modelado assumindo o intercepto anexando um termo hiperbólico tendendo a β_2 para grandes valores de x_2 . A inclinação β_1 é assumida não ser influenciada pela soma de PB. Os dados desse exemplo encontram-se na Tabela 1.6. Observando a Figura 1.4 percebemos que quando foi utilizado sinergista, a proporção de gafanhotos mortos com pequenas quantidades do produto foi maior do que quando utilizou-se inseticida.

Tabela 1.6: Dados do ensaio com os gafanhotos para o modelo de regressão logístico não linear.

Número de mortos, y	Tamanho amostral, m	Dose de inseticida	Dose de sinergista
7	100	4	0
59	200	5	0
115	300	8	0
149	300	10	0
178	300	15	0
229	300	20	0
5	100	2	3,9
43	100	5	3,9
76	100	10	3,9
4	100	2	19,5
57	100	5	19,5
83	100	10	19,5
6	100	2	39,0
57	100	5	39,0
84	100	10	39,0

Fonte: McCullagh & Nelder (1989).

Figura 1.4: Proporção dos gafanhotos mortos submetidos a log doses de duas drogas, inseticida e sinergista.

1.1.5 Cálcio Radioativo

Os dados desse exemplo foram analisados por Rawlings et al. (1998). Howard Grimes, do Departamento de Botânica, da Universidade do Estado da Carolina do Norte, conduziu um experimento para análise bioquímica de armazenamento intracelular e transporte de cálcio através da membrana plasmática. Células ficavam suspensas em uma solução de cálcio radioativo com tempos fixados de

Tabela 1.7: Dados da quantidade absorvida de cálcio radioativo para o modelo de regressão normal não linear.

x	y	x	y
0,45	0,34170	6,10	3,00782
0,45	-0,00438	6,10	2,67061
0,45	0,82531	8,05	3,05959
1,30	1,77967	8,05	3,94321
1,30	0,95384	8,05	3,43726
1,30	0,64080	11,15	4,80735
2,40	1,75136	11,15	3,35583
2,40	1,27497	11,15	2,78309
2,40	1,17332	13,15	5,13825
4,00	3,12273	13,15	4,70274
4,00	2,60958	13,15	4,25702
4,00	2,57429	15,00	3,60407
6,10	3,17881	15,00	4,15029
		15,00	3,42484

Fonte: Rawlings (1998).

0,45 até 15 minutos, x , e então a quantidade de cálcio radioativo Y (em nmoles/mg) absorvida pelas células foi medida numa amostra de 27 observações. Foi considerado um modelo de regressão normal não linear da forma:

$$\mu_i = \beta_0 \{1 - \exp(-\beta_1 x_i)\},$$

em que $Y_i \sim N(\mu_i, \sigma^2)$, $i = 1, \dots, 27$. Os dados desse exemplo encontram-se na Tabela 1.7. Observando a Figura 1.5 percebemos um crescimento logarítmico do tempo de exposição das células no cálcio com a quantidade de cálcio absorvida pelas células.

1.1.6 Produção de Vendas

Os dados desse exemplo foram primeiramente apresentados por Whitmore (1986) e analisados por Wei (1998). Denominaremos de dados de produção de vendas de mercadorias que não estão em estoque. Tem-se ainda que, x_i representa a projeção de vendas totais da i -ésima produção relatada por uma pesquisa organizacional de marketing e Y_i são as correspondentes vendas reais totais de uma

Figura 1.5: Diagrama de dispersão entre a quantidade de cálcio absorvido e o tempo de suspensão.

empresa. Admite-se que a variável resposta, Y_i , tenha uma distribuição normal inversa, isto é,

$$Y_i \sim \text{NI}(\beta x_i^\gamma, k^{-1} x_i^{-\rho}),$$

em que $i = 1, \dots, 20$, $\mu_i = E(Y_i) = \beta \exp\{\gamma \log x_i\}$, $\text{Var}(Y_i) = \sigma_i^2 V(\mu_i)$ com $\sigma_i^{-2} = k x_i^\rho$ e $V(\mu_i) = \mu_i^3$. Neste exemplo, por simplicidade computacional, optou-se por utilizar $\rho = 0$, então $\sigma_i^2 = k^{-1}$ para

todo i . Logo, $Y_i \sim \text{NI}(\beta x_i^\gamma, k^{-1})$ com $\mu_i = \beta x_i^\gamma$ e $\text{Var}(Y_i) = \sigma_i^2 \mu_i^3$ ($\sigma^2 = k^{-1}$). Os dados desse exemplo encontram-se na Tabela 1.8. Observando a Figura 1.6 percebemos que as projeções que foram feitas das vendas, ficaram muito próximas das verdadeiras vendas.

Figura 1.6: Diagrama de dispersão entre a projeção de vendas e as vendas reais.

Tabela 1.8: Dados referentes à projeção de vendas e à produção atual de vendas de uma empresa.

Produção	Vendas	
	Projeção	Atual
1	5959	5673
2	3534	3659
3	2641	2565
4	1965	2182
5	1738	1839
6	1182	1236
7	667	918
8	613	902
9	610	756
10	549	500
11	527	487
12	353	463
13	331	225
14	290	257
15	253	311
16	193	212
17	156	166
18	133	123
19	122	198
20	114	99

Fonte: Whitmore (1986).

1.1.7 Casos de Câncer de Pulmão

Considere o número de mortes de câncer de pulmão, Y_i , como sendo variáveis aleatórias independentes com distribuição de Poisson de médias $\mu_i = E(Y_i) = t_i \lambda_i$, em que t_i representa um tempo particular de exposição, $i = 1, \dots, n$, e λ_i a i -ésima taxa de morte. Uma classe geral de modelos para explicar a taxa de morte de forma aditiva proposta por Breslow & Day (1987) é dada por

$$\lambda_i = \exp(\mathbf{z}_i^\top \boldsymbol{\alpha}) \{1 + \mathbf{x}_i^\top \boldsymbol{\beta}\}, \quad i = 1, \dots, n,$$

em que $\mathbf{z}_i = (z_{i1}, \dots, z_{iq})^\top$ é um vetor com os valores de q variáveis de confundimento, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_q)^\top$, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ representa os valores das p variáveis de exposição e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$. Dessa forma, utilizando os dados de Breslow & Day (1987) de um estudo de seguimento com médicos britânicos, em que Y representa o número de mortes por câncer de pulmão e t a aproximação pessoas anos de observações classificadas pela idade e pelo número de cigarros consumidos por dia, Cordeiro & Paula (1992) ajustaram esses dados utilizando o seguinte modelo aditivo:

$$\log \mu_i = \log t_i + \alpha_0 + \alpha_1 z_i + h(\mathbf{x}_i; \boldsymbol{\beta}),$$

em que $Y_i \sim P(\mu_i)$, $h(\mathbf{x}_i; \boldsymbol{\beta}) = \log(1 + \beta_1 x_i + \beta_2 x_i^2)$, z é a idade média e x o número médio de cigarros consumidos por dia. Adicionalmente, vale ressaltar que Cordeiro & Paula (1992) consideraram $\log t$ como sendo um *offset*¹. Os dados desse exemplo encontram-se na Tabela 1.9.

1.1.8 Crescimento de Colônias

O crescimento de três colônias de *Paramecium aurelium* em um determinado meio nutritivo foi estudado através de curvas de crescimento por Diggle (1990) vide também (Svetliza, 2002). No começo de cada experimento 20 *paramecia* foram colocadas em um tubo com o meio nutritivo a certa temperatura. Cada dia, começando pelo segundo dia, o número de indivíduos foi contado até a estabilização do tamanho das colônias após 10 dias. Seja Y_{ij} o número de indivíduos no i -ésimo dia na j -ésima colônia ($i = 1, \dots, 19$ e $j = 1, 2, 3$). Svetliza (2002) e Svetliza & Paula (2003) assumem que $Y_{ij} \sim \text{BN}(\mu_i, \phi_j)$ tal que

$$\log \mu_{ij} = \exp\{\alpha_j - \exp(\beta_j - \gamma_j x_{ij})\},$$

em que x denota o número de dias. Os dados estão descritos na Tabela 1.10. Observando a Figura 1.7 percebemos que existe um comportamento similar no crescimento de *Paramecium aurelium* em cada colônia.

¹*offset* é uma quantidade que é subtraída do preditor linear em modelos lineares generalizados

Tabela 1.9: Dados sobre estudo de seguimento com médicos britânicos. O nº de mortes por câncer de pulmão é relacionado com o consumo diário de cigarros e a idade.

Nº de cigarros consumidos por dia	Nº médio consumido		Idade em anos										
			40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79			
0	0	O	0	0	0	1	2	0	0	1	2		
		P-Y	17846,5	15832,5	12226	8905,5	6248	4351	2723,5	1772			
1-4	2,7	O	0	0	0	0	1	0	1	0			
		P-Y	1216	1000,5	853,5	625	509,5	392,5	242	208,5			
5-9	6,6	O	0	0	0	0	0	1	1	2	0		
		P-Y	2041,5	1745	1562,5	1355	1068	843,5	696,5	517,5			
10-14	11,3	O	1	1	2	1	1	2	4	4			
		P-Y	3795,5	3205	2727	2288	1714	1214	862	547			
15-19	16	O	0	1	4	0	2	2	4	5			
		P-Y	4824	3995	3278,5	2466,5	1829,5	1237	683,5	370,5			
20-24	20,4	O	1	1	6	8	13	12	10	7			
		P-Y	7046	6460,5	5583	4357,5	2863,5	1930	1055	512			
25-29	25,4	O	0	2	3	5	4	5	7	4			
		P-Y	2523	2565,5	2620	2108,5	1508,5	974,5	527	209,5			
30-34	30,2	O	1	2	3	6	11	9	2	2			
		P-Y	1715,5	2123	2226,5	1923	1362	763,5	317,5	130			
35-40	38	O	0	0	3	4	7	9	5	2			
		P-Y	892,5	1150	1281	1063	826	515	233	88,5			

Fonte: Breslow & Day (1987).

Figura 1.7: Crescimento de três colônias de *Paramecium aurelium* submetidas a um meio nutritivo.

Tabela 1.10: Descrição do n° de *paramecia* ao longo do tempo em três colônias de *Paramecium aurelium* submetidas a um meio nutritivo.

Dias	Colônia A	Colônia B	Colônia C
0	2	2	2
2	17	15	11
3	29	36	37
4	30	62	67
5	63	84	134
6	185	156	226
7	258	234	306
8	267	348	376
9	392	370	485
10	510	480	530
11	570	520	650
12	650	575	605
13	560	400	580
14	575	545	660
15	650	560	460
16	550	480	650
17	480	510	575
18	520	650	525
19	500	500	550

Para resolver esse tipo de problema apresentado neste capítulo é necessário a utilização de modelos não lineares de família exponencial que será visto no próximo capítulo.

Capítulo 2

Modelos Não Lineares de Família Exponencial

2.1 Família Exponencial

A família exponencial univariada é uma das classes de distribuição mais comumente utilizadas em Estatística, que permite incorporar dados assimétricos, dados discretos ou contínuos, e dados que são restritos a um intervalo do conjunto dos reais. Várias distribuições conhecidas pertencem à família exponencial tais como as distribuições normal, binomial, gama, Poisson, normal inversa e binomial negativa.

A fim de introduzir a família exponencial vamos supor que Y é uma variável aleatória com densidade na forma

$$f(y; \theta, \phi) = \exp\{\phi\{y\theta - b(\theta)\} + c(y, \phi)\}, \quad (2.1)$$

em que $b(\cdot), c(\cdot, \cdot)$ são funções conhecidas e $\phi^{-1} > 0$ é chamado parâmetro de dispersão, que pode ser desconhecido e $-\infty < \theta, y < \infty$. A função geradora de momentos de uma variável aleatória que pertence à família exponencial assume a forma

$$M(t; \theta, \phi) = \exp[\phi\{b(\frac{t}{\phi} + \theta) - b(\theta)\}],$$

que depende da função $b(\cdot)$. Já a função geradora de cumulantes é dada por

$$\log\{M(t; \theta, \phi)\} = \phi\{b(\frac{t}{\phi} + \theta) - b(\theta)\}.$$

Derivando a equação anterior r vezes em relação a t e avaliando em $t = 0$ temos a fórmula geral para o r -ésimo cumulante de y

$$\kappa_r = \phi^{1-r} b^{(r)}(\theta) = \phi^{-1} \partial \kappa_{r-1} / \partial \theta, \quad r \geq 2.$$

Verifica-se, portanto, que existe uma relação de recorrência entre os cumulantes da família exponencial. Isto é muito importante na obtenção de propriedades assintóticas dos MLGs. Os momentos da família exponencial podem ser facilmente obtidos a partir dos cumulantes, vide, por exemplo, (Kendall & Stuart, 1977).

A média e a variância de uma variável aleatória que pertence à família exponencial podem ser obtidas da forma

$$\mu = E(Y) = b'(\theta) \quad \text{e} \quad \sigma^2 = \text{Var}(Y) = \phi^{-1} b''(\theta).$$

Além disso, $b''(\theta) = \partial \mu / \partial \theta$ é uma função unicamente de μ e é representada por $V(\mu)$ (função de variância). Portanto, o parâmetro natural pode ser expresso por uma relação unívoca da média, isto é, $\theta = \int V^{-1} d\mu = q(\mu)$. Dessa forma, a variância de Y pode ser reescrita como

$$\text{Var}(Y) = \phi^{-1} V(\mu).$$

Na Tabela 2.1 temos os resumos das principais distribuições pertencentes à família exponencial.

2.2 Modelos Não Lineares de Família Exponencial

Suponha que $Y_1, \dots, Y_n$ são variáveis aleatórias independentes, em que cada Y_i tem densidade na forma (2.1), com $E(Y_i) = \mu_i$ e $\text{Var}(Y_i) = \phi_i^{-1} V_i$, em que $V_i = V(\mu_i)$. Vamos supor também que

Tabela 2.1: Principais distribuições pertencentes à família exponencial.

Domínio	Distribuição			
	Normal	Poisson	Binomial	Gama
$\Re$	$\Re$	$\{0, 1, 2, \dots\}$	$\{0, 1, \dots, n\}$	$\Re^+$
$b(\theta)$	$\theta^2/2$	e^θ	$\log(1 + e^\theta)$	$-\log(-\theta)$
θ	μ	$\log \mu$	$\log\{\mu/(1 - \mu)\}$	$-\sqrt{-2\theta}$
ϕ	σ^{-2}	1	n	$-1/2\mu^2$
$V(\mu)$	1	μ	$\mu(1 - \mu)$	ϕ
$c(y, \phi)$	$1/2[\log \phi - \log 2\pi]$ $-1/2(y^2/\phi)$	$-\log y!$	$\log\left(\frac{\phi}{\phi y}\right)$	μ^3
$D(y; \hat{\mu})$	$\sum_{i=1}^n (y_i - \hat{\mu}_i)^2$	$2 \sum_{i=1}^n \{y_i \log(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i) \log\{(1 - y_i/n_i)/(1 - \hat{\mu}_i)\}\}$	$2 \sum_{i=1}^k [y_i \log(y_i/n_i \hat{\mu}_i) + (n_i - y_i) \log\{(1 - y_i/n_i)/(1 - \hat{\mu}_i)\}]$	$1/2[\log \phi - \log 2\pi y^3]$ $- \phi/2y$
				$\sum_{i=1}^n (y_i - \hat{\mu}_i)^2 / (y_i \hat{\mu}_i^2)$

CV é o coeficiente de variação.

o parâmetro de dispersão $\phi_i^{-1} > 0$, $i = 1, \dots, n$, é conhecido ou desconhecido, mas é o mesmo para todas as observações. Sejam as variáveis em consideração denotadas por $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ e os valores amostrais por $\mathbf{y} = (y_1, \dots, y_n)^\top$. Os modelos não lineares de família exponencial são definidos por (2.1) e pela componente sistemática

$$g(\mu_i) = \eta_i = f(\mathbf{x}_i; \boldsymbol{\beta}), \quad i = 1, \dots, n, \quad (2.2)$$

em que $g(\cdot)$ é uma função de ligação monótona conhecida e diferenciável no mesmo sentido dos MLGs, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ é um vetor de parâmetros desconhecidos a serem estimados e $f(\cdot; \cdot)$ é uma função contínua, diferenciável e não linear em $\boldsymbol{\beta}$. Aqui $\mathbf{x}_i^\top = (x_{i1}, \dots, x_{iq})$ é um vetor de valores fixados conhecidos de variáveis explicativas associadas com a resposta observada y_i . Vamos assumir identificabilidade no sentido que diferentes $\boldsymbol{\beta}$'s fornecerão diferentes $\boldsymbol{\eta}$'s fazendo com que a matriz de derivadas $\mathbf{J} = \mathbf{J}(\boldsymbol{\beta}) = \partial \boldsymbol{\eta} / \partial \boldsymbol{\beta}^\top$ tenha posto p , para todo $\boldsymbol{\beta}$ em que $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^\top$. Assumimos para (2.1) e (2.2) as condições usuais de regularidade para a função de verossimilhança. É interessante notar que se $f(\mathbf{x}_i; \boldsymbol{\beta}) = \mathbf{x}_i^\top \boldsymbol{\beta}$, então (2.2) representa a classe dos MLGs e $q = p$. Da equação (2.2) temos que o parâmetro $\boldsymbol{\beta}$ pode ser ligado com o parâmetro natural θ_i e a média μ_i como segue:

i) De $\mu_i = b'(\theta_i)$, $g(\mu_i) = g(b'(\theta_i)) = f(\mathbf{x}_i; \boldsymbol{\beta})$, temos que

$$\theta_i = \theta_i(\boldsymbol{\beta}) = (b')^{-1}[g^{-1}\{f(\mathbf{x}_i; \boldsymbol{\beta})\}].$$

Se a função de ligação $g(\cdot)$ é tal que $\theta_i = \eta_i = f(\mathbf{x}_i; \boldsymbol{\beta})$, que significa que $(b')^{-1}\{g^{-1}(\cdot)\}$ é uma função identidade, então tanto a função $g(\cdot)$ e a equação (2.2) são chamadas de ligações canônicas, as quais são mais fáceis de trabalhar, vide, por exemplo, (Fahrmeir & Kaufmann, 1985).

ii) A equação (2.2) pode ser também denotada pela média μ_i ,

$$\mu_i = \mu_i(\boldsymbol{\beta}) = g^{-1}\{f(\mathbf{x}_i; \boldsymbol{\beta})\}.$$

2.3 Função Desvio

O *deviance*, traduzida pela primeira vez como desvio por Cordeiro (1986), é uma importante e bem conhecida estatística discutida em problemas relacionados à família exponencial para avaliar a qualidade do ajuste do modelo. Seja o logaritmo da função de verossimilhança

$$L(\boldsymbol{\mu}; \mathbf{y}) = \sum_{i=1}^n L(\mu_i; y_i) = \sum_{i=1}^n \{\phi[y_i\theta_i - b(\theta_i)] + c(y_i, \phi)\},$$

com $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^\top$, $\mu_i = g^{-1}(\eta_i)$ e $\eta_i = f(\mathbf{x}_i; \boldsymbol{\beta})$.

A função desvio supondo ϕ fixo ou conhecido é definida pela forma

$$D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \phi D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\{L(\mathbf{y}; \mathbf{y}) - L(\hat{\boldsymbol{\mu}}; \mathbf{y})\},$$

que é duas vezes a diferença entre os máximos do logaritmo da função de verossimilhança para os modelos saturado (com n parâmetros) e sob investigação (com p parâmetros) avaliado na estimativa de máxima verossimilhança $\hat{\boldsymbol{\beta}}$. Como essa medida avalia a qualidade do ajuste do modelo postulado, então quanto melhor for o ajuste do modelo aos dados tanto menor será o valor de $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$. Assim, como $L(\hat{\boldsymbol{\mu}}; \mathbf{y}) \leq L(\mathbf{y}; \mathbf{y})$, um modelo bem ajustado aos dados com um valor alto para a função de verossimilhança terá um desvio pequeno.

Podemos reescrever a função $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ não escalonada como

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

em que $\hat{\theta}_i = \theta_i(\hat{\mu}_i)$ e $\hat{\theta}_i^0 = \theta_i(\hat{\mu}_i^0)$ são as estimativas de máxima verossimilhança de θ para os modelos com p parâmetros ($p < n$) e saturado ($p = n$), respectivamente. Usualmente compara-se os valores observados da função desvio com os percentis da distribuição qui-quadrado com $n - p$ graus de liberdade, sendo p o posto da matriz modelo. Entretanto, em geral, $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ não segue uma distribuição χ_{n-p}^2 nem mesmo assintoticamente. Assumindo-se sob a hipótese de que o modelo usado

é verdadeiro, a distribuição binomial, quando k é fixo e $n_i \rightarrow \infty$, $\forall i$ (não vale quando $n_i \mu_i(1 - \mu_i)$ permanece limitado) e a distribuição Poisson, quando $\mu_i \rightarrow \infty$, $\forall i$, tem um $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) \sim \chi_{k-p}^2$ (lembre-se que $\phi = 1$) e um $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) \sim \chi_{n-p}^2$, respectivamente. No caso da distribuição normal, por exemplo, considerando σ^2 conhecido, temos que $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) \sim \sigma^2 \chi_{n-p}^2$.

Nos casos em que $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ depende do parâmetro de dispersão ϕ^{-1} , Jørgensen (1987) mostra que

$$D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) \sim \chi_{n-p}^2, \quad \text{quando } \phi \rightarrow \infty,$$

isto é, quando a dispersão é pequena, fica razoável comparar os valores observados de $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ com os percentis da χ_{n-p}^2 . Lembrando que se $Z \sim \chi_{n-p}^2$, então $E(Z) = n - p$, e portanto, um valor de $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ próximo de $n - p$ pode ser uma indicação de que o modelo ajustado aos dados é adequado. Cordeiro & Paula (1989) apresentam aprimoramentos para a distribuição nula assintótica da estatística da razão de verossimilhanças e consequentemente para a função desvio.

2.4 Estimação dos Parâmetros do Modelo

2.4.1 Estimação de β

Seja $L(\boldsymbol{\theta})$, em que $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi)^\top$, o logaritmo da função de verossimilhança para algum modelo definido por (2.1) expresso na forma

$$L(\boldsymbol{\theta}) = \sum_{i=1}^n \phi \{y_i \theta_i - b(\theta_i)\} + \sum_{i=1}^n c(y_i, \phi). \quad (2.3)$$

Seja a função score para $\boldsymbol{\beta}$ denotada por,

$$\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}},$$

ou simplesmente $\mathbf{U}_{\boldsymbol{\beta}}$. A estimativa de máxima verossimilhança $\hat{\boldsymbol{\beta}}$ pode ser obtida expandindo-se a

função escore em torno de $\beta^{(0)}$ e supondo ϕ fixado em série de Taylor, tal que

$$U_{\beta} \cong U_{\beta}^{(0)} + U_{\beta}^{'(0)}(\beta - \beta^{(0)}), \quad (2.4)$$

em que $U_{\beta}'(\theta) = \partial U_{\beta}(\theta) / \partial \beta^{\top}$, ou seja, $-U_{\beta}'(\theta)$ é a matriz observada de Fisher de β . Resolvendo-se a equação (2.4) chega-se à seguinte solução:

$$\beta^{(1)} = \beta^{(0)} + \{-U_{\beta}^{'(0)}\}^{-1} U_{\beta}^{(0)},$$

e portanto ao processo iterativo

$$\beta^{(m+1)} = \beta^{(m)} + \{-U_{\beta}^{'(m)}\}^{-1} U_{\beta}^{(m)},$$

para $m = 0, 1, \dots$. O método escore de Fisher consiste em substituir $-U_{\beta}'(\theta)$ pelo correspondente valor esperado $K_{\beta\beta}(\theta) = E\{-U_{\beta}'(\theta)\}$, em que $K_{\beta\beta}(\theta)$ é a matriz de informação de Fisher de β . Assim, o processo iterativo para obter $\hat{\beta}$ fica dado por

$$\beta^{(m+1)} = \beta^{(m)} + \{K_{\beta\beta}^{(m)}\}^{-1} U_{\beta}^{(m)}. \quad (2.5)$$

Vamos calcular a seguir as quantidades $U_{\beta}(\theta)$ e $K_{\beta\beta}(\theta)$. Calculamos inicialmente a função

escore para β_j

$$\begin{aligned}
 \mathbf{U}_{\beta_j}(\boldsymbol{\theta}) = \frac{\partial L(\boldsymbol{\theta})}{\partial \beta_j} &= \sum_{i=1}^n \phi \left\{ y_i \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} - \frac{db(\theta_i)}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} \right\} \\
 &= \sum_{i=1}^n \phi \left\{ y_i V_i^{-1} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} - \mu_i V_i^{-1} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} \right\} \\
 &= \sum_{i=1}^n \phi \left\{ y_i \frac{1}{V_i} \frac{d\mu_i}{d\eta_i} \left(\frac{d\mu_i/d\eta_i}{d\mu_i/d\eta_i} \right) \frac{\partial \eta_i}{\partial \beta_j} - \mu_i \frac{1}{V_i} \frac{d\mu_i}{d\eta_i} \left(\frac{d\mu_i/d\eta_i}{d\mu_i/d\eta_i} \right) \frac{\partial \eta_i}{\partial \beta_j} \right\} \\
 &= \sum_{i=1}^n \phi \left\{ y_i \omega_i d_i \frac{\partial \eta_i}{\partial \beta_j} - \mu_i \omega_i d_i \frac{\partial \eta_i}{\partial \beta_j} \right\} \\
 &= \sum_{i=1}^n \phi \left\{ \omega_i d_i \frac{\partial \eta_i}{\partial \beta_j} (y_i - \mu_i) \right\},
 \end{aligned}$$

em que $\omega_i = (d\mu_i/d\eta_i)^2/V_i$ e $d_i = d\eta_i/d\mu_i$. Em notação matricial, temos o seguinte:

$$\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \phi \mathbf{J}^\top \mathbf{W} \mathbf{D}(\mathbf{y} - \boldsymbol{\mu}),$$

em que $\mathbf{J}$ é uma matriz $n \times p$ de posto completo, formada pelos elementos da matriz Jacobiana $\partial \boldsymbol{\eta} / \partial \boldsymbol{\beta}$, $\mathbf{W} = \text{diag}\{\omega_1, \dots, \omega_n\}$ é a matriz de pesos, $\mathbf{D} = \text{diag}\{d\eta_i/d\mu_i\}$, $\mathbf{y} = (y_1, \dots, y_n)^\top$ e $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^\top$. Ainda,

$$\begin{aligned}
 \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_l} &= \phi \sum_{i=1}^n \left\{ (y_i - \mu_i) \frac{d^2 \theta_i}{d\mu_i^2} \left(\frac{d\mu_i}{d\eta_i} \right)^2 \frac{\partial \eta_i}{\partial \beta_l} \frac{\partial \eta_i}{\partial \beta_j} \right\} + \phi \sum_{i=1}^n \left\{ (y_i - \mu_i) \frac{d\theta_i}{d\mu_i} \frac{d^2 \mu_i}{d\eta_i^2} \frac{\partial \eta_i}{\partial \beta_l} \frac{\partial \eta_i}{\partial \beta_j} \right\} \\
 &+ \phi \sum_{i=1}^n \left\{ (y_i - \mu_i) \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial^2 \eta_i}{\partial \beta_j \partial \beta_l} \right\} - \phi \sum_{i=1}^n \left\{ \frac{d\theta_i}{d\mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^2 \frac{\partial \eta_i}{\partial \beta_l} \frac{\partial \eta_i}{\partial \beta_j} \right\},
 \end{aligned}$$

cujo valor esperado fica dado por

$$\mathbb{E} \left\{ \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_l} \right\} = -\phi \sum_{i=1}^n \left\{ \omega_i \frac{\partial \eta_i}{\partial \beta_l} \frac{\partial \eta_i}{\partial \beta_j} \right\}.$$

Portanto, a informação de Fisher para β em forma matricial fica dada por

$$\mathbf{K}_{\beta\beta}(\theta) = \mathbb{E} \left\{ -\frac{\partial^2 L(\theta)}{\partial \beta \partial \beta^\top} \right\} = \phi \mathbf{J}^\top \mathbf{W} \mathbf{J}.$$

Basta agora substituir a função escore e a matriz de informação de Fisher em (2.5) e teremos o seguinte processo iterativo de mínimos quadrados ponderados:

$$\beta^{(m+1)} = \beta^{(m)} + (\mathbf{J}^{(m)\top} \mathbf{W}^{(m)} \mathbf{J}^{(m)})^{-1} \mathbf{J}^{(m)\top} \mathbf{W}^{(m)} \mathbf{D}^{(m)} (\mathbf{y} - \mu^{(m)}).$$

Se pré-multiplicarmos a equação anterior por $(\mathbf{J}^{(m)\top} \mathbf{W}^{(m)} \mathbf{J}^{(m)})$, obtemos

$$(\mathbf{J}^{(m)\top} \mathbf{W}^{(m)} \mathbf{J}^{(m)}) \beta^{(m+1)} = \mathbf{J}^{(m)\top} \mathbf{W}^{(m)} \mathbf{y}^{*(m)}, \quad (2.6)$$

em que $\mathbf{y}^{*(m)} = \mathbf{J}^{(m)} \beta^{(m)} + \mathbf{D}^{(m)} (\mathbf{y} - \mu^{(m)})$. Para conseguirmos implementar (2.6) no S-Plus ou no R basta reescrevermos a variável dependente localmente modificada $\mathbf{y}^*$, na forma $\mathbf{y}^* = \boldsymbol{\eta} - \boldsymbol{\tau} + \mathbf{D}(\mathbf{y} - \mu)$, sendo $\boldsymbol{\eta} = f(\mathbf{J}; \beta)$ e $\boldsymbol{\tau} = f(\mathbf{J}; \beta) - \mathbf{J}\beta$, em que $f(\mathbf{J}; \beta)$ é um vetor $n \times 1$ de componentes $f(\mathbf{x}_i; \beta)$, $i = 1, \dots, n$. Dessa forma, a expressão (2.6) pode ser interpretada como um processo iterativo para ajustar um modelo linear generalizado com matriz modelo $\mathbf{J}$, parte sistemática $g(\mu) = \boldsymbol{\eta}$, função de variância $\mathbf{V}$ e *offset* $\boldsymbol{\tau}$. Relembrando, *offset* é uma quantidade que é subtraída, a cada passo, do preditor $\boldsymbol{\eta}$, vide, por exemplo, (Paula & Cordeiro, 1986). A única diferença com relação aos MLGs é que aqui a matriz $\mathbf{J}$ é modificada a cada passo do processo iterativo.

O seguinte procedimento pode ser executado para encontrar $\hat{\beta}$:

1. Escolher uma estimativa inicial $\beta^{(0)}$;
2. Calcular $\mathbf{J}^{(0)}$, $\boldsymbol{\eta}^{(0)} = f(\mathbf{J}^{(0)}; \beta^{(0)})$ e $\boldsymbol{\tau}^{(0)}$;
3. Ajustar o modelo postulado com parte sistemática $g(\mu) = \boldsymbol{\eta}$, matriz modelo $\mathbf{J}^{(0)}$ e *offset* $\boldsymbol{\tau}^{(0)}$;

4. No cálculo de $\beta^{(1)}$ teremos

$$\mathbf{y}^{*(0)} = \boldsymbol{\eta}^{(0)} - \boldsymbol{\tau}^{(0)} + \mathbf{D}^{(0)}(\mathbf{y} - \boldsymbol{\mu}^{(0)}),$$

que é justamente a variável dependente modificada avaliada em $\beta^{(0)}$;

5. Atualizar $\beta^{(1)}$ e calcular $\mathbf{J}^{(1)}$, $\boldsymbol{\eta}^{(1)}$ e $\boldsymbol{\tau}^{(1)}$. Note que a atualização de $\boldsymbol{\mu}^{(1)}$ é feita automaticamente da relação $\boldsymbol{\mu}^{(1)} = g^{-1}(\boldsymbol{\eta}^{(1)})$;

6. Voltar ao passo (3) até a convergência, obtendo-se $\hat{\beta}$.

Não sabemos sob quais condições este esquema iterativo diverge e quais são as condições para assegurar a existência e unicidade das estimativas dos parâmetros. Porém, vários exemplos aplicados indicam que esse esquema iterativo funciona corretamente. Um critério para verificar a convergência poderia ser, por exemplo, quando a norma das diferenças relativas entre as duas estimativas sucessivas for menor do que um valor ϵ pré-determinado, ou seja, parar o processo iterativo quando $\|(\beta_j^{(m+1)} - \beta_j^{(m)})\|^{1/2} < \epsilon, \forall j$, vide, por exemplo, (Paula & Cordeiro, 1986). Alternativamente, pode-se aplicar outros procedimentos de maximização para a estimação dos parâmetros em modelos não lineares de família exponencial disponíveis em diversos aplicativos, tais como R, SAS, S-Plus, Ox e Mathematica etc.

2.4.2 Estimação de ϕ

Para estimarmos o parâmetro ϕ , iremos admitir que este é o mesmo para todas as observações, isto é, constante. Seja o estimador de máxima verossimilhança $\hat{\phi}$ do parâmetro de escala ϕ . Seja a função escore para ϕ denotada por $U_\phi(\boldsymbol{\theta}) = \partial L(\boldsymbol{\theta}) / \partial \phi$. Igualando $U_\phi(\boldsymbol{\theta})$ a zero obtemos

$$\sum_{i=1}^n \{y_i \hat{\theta}_i - b(\hat{\theta}_i)\} + \sum_{i=1}^n c'(y_i, \hat{\phi}) = 0.$$

Agora, escrevendo em termos do desvio obtemos

$$\sum_{i=1}^n c'(y_i, \hat{\phi}) = \frac{1}{2} D(\mathbf{y}; \hat{\boldsymbol{\mu}}) - \sum_{i=1}^n [y_i \hat{\theta}_i^0 - b(\hat{\theta}_i^0)], \quad (2.7)$$

em que $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ denota o desvio do modelo sob investigação sem o multiplicador ϕ , $\hat{\theta}_i = \theta_i(\hat{\mu}_i)$ e $\hat{\theta}_i^0 = \theta_i(\hat{\mu}_i^0)$ são estimativas de máxima verossimilhança de $\boldsymbol{\theta}$ para os modelos com p parâmetros ($p < n$) e saturado ($p = n$), respectivamente. Apresentaremos a seguir alguns casos particulares da estimação de ϕ .

Gama

No caso gama, tem-se $\theta_i = -1/\mu_i$, $\hat{\theta}_i^0 = -1/y_i$ e $\hat{\theta}_i = -1/\hat{\mu}_i$, pois sabemos que $\hat{\mu}_i^0 = y_i$. Temos também que $b(\hat{\theta}_i^0) = \log(y_i)$, $c(y, \phi) = (\phi - 1) \log y + \phi \log y - \log \Gamma(\phi)$ e $y > 0$. Então, calculando a derivada de $c(y, \phi)$ e substituindo em (2.7) obtemos $\hat{\phi}$ da equação

$$2n[\log \hat{\phi} - \Psi(\hat{\phi})] = D(\mathbf{y}; \hat{\boldsymbol{\mu}}),$$

em que $\Psi(\hat{\phi}) = \Gamma'(\hat{\phi})/\Gamma(\hat{\phi})$ é a função digama. Essa equação deve ser resolvida iterativamente, que equivale ao seguinte procedimento de Newton-Raphson:

$$\phi^{(m+1)} = \phi^{(m)} - \{U'_\phi(\boldsymbol{\theta})\}^{-1} U_\phi^{(m)},$$

em que $U'_\phi(\boldsymbol{\theta}) = \partial U_\phi(\boldsymbol{\theta})/\partial \phi$. Calculando $U_\phi(\boldsymbol{\theta})$ e $U'_\phi(\boldsymbol{\theta})$ para a distribuição gama temos que o processo iterativo acima fica dado por

$$\phi^{(m+1)} = \phi^{(m)} + \left\{ \frac{1}{\Psi'(\phi^{(m)}) - (1/\phi^{(m)})} \right\} \left[-\frac{y}{\mu^{(m)}} + \log \left(\frac{y}{\mu^{(m)}} \right) - \Psi(\phi^{(m)}) + \log \phi^{(m)} + 1 \right],$$

em que $U_\phi = -y/\mu + \log(y/\mu) - \Psi(\phi) + \log \phi + 1$ e $U'_\phi = -\Psi'(\phi) + 1/\phi$.

A estimativa de máxima verossimilhança é obtida na convergência do processo iterativo acima. Essa estimativa pode ser obtida pela *library* MASS (Venables & Ripley, 1999) disponível em S-Plus

e R. Alternativamente, podemos utilizar a estimativa de momentos

$$\tilde{\phi} = \left\{ \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i^2} \right\}^{-1},$$

que é também consistente para ϕ .

Normal

No caso Normal, tem-se $\theta_i = \mu_i$, $\hat{\theta}_i^0 = y_i$ e $\hat{\theta}_i = \hat{\mu}_i$, pois sabemos que $\hat{\mu}_i^0 = y_i$. Temos também que $b(\hat{\theta}_i^0) = y_i^2/2$ e $c(y, \phi) = 1/2[\log \phi - \log 2\pi] - 1/2(y^2/\phi)$. Então, calculando a derivada de $c(y, \phi)$ e substituindo em (2.7) obtemos $\hat{\phi}$ da equação

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

Portanto $\sigma^2 = \hat{\phi}^{-1} = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2/n$. Essa estimativa embora consistente é viesada para n fixo. A estimativa não viesada como bem conhecida é o $s^2 = D(\mathbf{y}; \hat{\boldsymbol{\mu}})/(n - p)$.

Normal Inversa

No caso Normal Inversa, tem-se $\theta_i = -1/2\mu_i^2$, e $\hat{\theta}_i^0 = -1/2y_i^2$ e $\hat{\theta}_i = \hat{\mu}_i$, pois sabemos que $\hat{\mu}_i^0 = y_i$. Temos também que $b(\hat{\theta}_i^0) = -\sqrt{1/y_i^2}$, $c(y, \phi) = 1/2[\log \phi - \log 2\pi y^3] - \phi/2y$ e que $y > 0$. Então, calculando a derivada de $c(y, \phi)$ e substituindo em (2.7) obtemos $\hat{\phi}$ da equação

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

2.5 Distribuição Assintótica de $\hat{\boldsymbol{\beta}}$ e $\hat{\phi}$

Geralmente a obtenção de distribuições exatas é muito complicada e resultados assintóticos são usados. Vários desses resultados assintóticos para a classe dos MLGs podem ser encontrados em McCullagh & Nelder (1983), (Cordeiro (1983), Cordeiro (1987)), Cox & Hinkley (1974), Fahrmeir &

Kaufmann (1985), Dobson (1990) e Sen & Singer (1993). McCullagh & Nelder (1983) e Jørgensen (1987) também apresentaram algumas extensões para classes mais gerais em que vários desses resultados são diretamente aplicáveis aos modelos não lineares de família exponencial.

Tem-se que $\hat{\theta} = (\hat{\beta}^\top, \hat{\phi})^\top$, quando $n \rightarrow \infty$, converge em distribuição, sob certas condições gerais de regularidade, vide, por exemplo, (Sen & Singer, 1993), (Svetliza, 2002), para

$$\begin{pmatrix} \hat{\beta} \\ \hat{\phi} \end{pmatrix} \sim N_{p+1} \left[\begin{pmatrix} \beta \\ \phi \end{pmatrix}, \begin{pmatrix} \phi^{-1}(\mathbf{J}^\top \mathbf{W} \mathbf{J})^{-1} & \mathbf{0} \\ \mathbf{0} & K_{\phi\phi}^{-1}(\theta) \end{pmatrix} \right].$$

A matriz de informação de Fisher total se reduz à matriz bloco diagonal $\mathbf{K}_{\theta\theta} = \text{diag}\{\mathbf{K}_{\beta\beta}(\theta), K_{\phi\phi}(\theta)\}$ em que $\mathbf{K}_{\beta\beta}(\theta) = \phi \mathbf{J}^\top \mathbf{W} \mathbf{J}$ e $K_{\phi\phi}(\theta) = E[-\partial^2 L(\theta)/\partial \phi^2] = E[-\sum_{i=1}^n c''(Y_i, \phi)]$ que é uma função de β e ϕ . Os parâmetros β e ϕ são ortogonais, ou seja, os estimadores de máxima verossimilhança de β e ϕ são assintoticamente independentes.

Portanto, o estimador de máxima verossimilhança $\hat{\beta}$ é assintoticamente normal com média β e matriz de variância-covariância $\phi^{-1}(\mathbf{J}^\top \mathbf{W} \mathbf{J})^{-1}$. Então, intervalos de confiança para um parâmetro específico β_j pode ser obtido da forma

$$\hat{\beta}_j \pm z_\alpha \sqrt{-\hat{k}^{jj}},$$

em que $-\hat{k}^{jj}$ é a variância assintótica estimada de $\hat{\beta}_j$ e $\Phi(-z_\alpha) = \alpha$, sendo $\Phi(\cdot)$ a função de distribuição acumulada da normal padrão. O estimador de máxima verossimilhança $\hat{\phi}$ é assintoticamente normal com média ϕ e variância $K_{\phi\phi}^{-1}(\theta)$. Então, o intervalo de confiança para o parâmetro ϕ pode ser obtido da forma

$$\hat{\phi} \pm z_\alpha \sqrt{\hat{K}_{\phi\phi}^{-1}},$$

em que $\hat{K}_{\phi\phi}^{-1}$ é a variância assintótica estimada de $\hat{\phi}$ e $\Phi(-z_\alpha) = \alpha$, sendo $\Phi(\cdot)$ a função de distribuição acumulada da normal padrão. Apresentamos a seguir alguns casos particulares para $K_{\phi\phi}(\theta)$.

Gama

No caso gama, tem-se $c(y, \phi) = (\phi - 1) \log y + \phi \log \phi - \log \Gamma(\phi)$. Então, $c'(y, \phi) = \log y + \log \phi + 1 - \Psi(\phi)$ e portanto $c''(y, \phi) = -\Psi'(\phi) + 1/\phi$. Logo, $K_{\phi\phi}(\boldsymbol{\theta})$ fica dada por

$$K_{\phi\phi}(\boldsymbol{\theta}) = E\left[-\sum_{i=1}^n c''(Y_i, \phi)\right] = \Psi'(\phi) - \frac{1}{\phi}.$$

Normal

No caso Normal, tem-se $c(y, \phi) = 1/2[\log \phi - \log 2\pi] - 1/2(y^2/\phi)$. Então, $c'(y, \phi) = 1/2\phi - y^2/2$ e portanto $c''(y, \phi) = -1/2\phi^2$. Logo, $K_{\phi\phi}(\boldsymbol{\theta})$ fica dada por

$$K_{\phi\phi}(\boldsymbol{\theta}) = E\left[-\sum_{i=1}^n c''(Y_i, \phi)\right] = \frac{1}{2\phi^2}.$$

Normal Inversa

No caso Normal Inversa, tem-se $c(y, \phi) = 1/2[\log \phi - \log 2\pi y^3] - \phi/2y$. Então, $c'(y, \phi) = 1/2\phi - 1/2y$ e portanto $c''(y, \phi) = -1/2\phi^2$. Logo, $K_{\phi\phi}(\boldsymbol{\theta})$ fica dada por

$$K_{\phi\phi}(\boldsymbol{\theta}) = E\left[-\sum_{i=1}^n c''(Y_i, \phi)\right] = \frac{1}{2\phi^2}.$$

2.6 Testes de Hipóteses **ϕ conhecido**

Muitas vezes estamos interessados em testar hipóteses apenas de um subconjunto de parâmetros, ao invés de testarmos o vetor inteiro. Assumindo ϕ conhecido (ou fixo), suponha uma partição do vetor

de parâmetros β dada por:

$$\beta = (\beta_1^\top, \beta_2^\top)^\top,$$

em que β_1 é o vetor de interesse de dimensão q e β_2 é o vetor de parâmetros de perturbação de dimensão $p - q$. Portanto, podemos estar interessados em testar as hipóteses:

$$H_0 : \beta_1 = \beta_1^0$$

$$H_1 : \beta_1 \neq \beta_1^0,$$

em que β_1^0 é um vetor conhecido. Sejam $\mathbf{J} = [\mathbf{J}_1, \mathbf{J}_2]$ e $\mathbf{U}_\beta = [\mathbf{U}_{\beta_1}^\top, \mathbf{U}_{\beta_2}^\top]^\top$ a matriz Jacobiana $\mathbf{J}$ e o vetor escore de β , particionados seguindo a hipótese anterior, respectivamente. A matriz de informação de Fisher pode ser particionada da forma

$$\mathbf{K}_{\beta\beta} = \begin{pmatrix} \mathbf{K}_{\beta_1\beta_1} & \mathbf{K}_{\beta_1\beta_2} \\ \mathbf{K}_{\beta_2\beta_1} & \mathbf{K}_{\beta_2\beta_2} \end{pmatrix},$$

em que $\mathbf{K}_{\beta_1\beta_1} = \phi \mathbf{J}_1^\top \mathbf{W} \mathbf{J}_1$, $\mathbf{K}_{\beta_1\beta_2} = \mathbf{K}_{\beta_2\beta_1}^\top = \phi \mathbf{J}_1^\top \mathbf{W} \mathbf{J}_2$ e $\mathbf{K}_{\beta_2\beta_2} = \phi \mathbf{J}_2^\top \mathbf{W} \mathbf{J}_2$.

Sejam $\hat{\beta} = (\hat{\beta}_1^\top, \hat{\beta}_2^\top)^\top$ e $\tilde{\beta} = (\tilde{\beta}_1^{0\top}, \tilde{\beta}_2^\top)^\top$ os estimadores de máxima verossimilhança de β sob H_1 e H_0 , respectivamente. Para testar H_0 contra H_1 podemos aplicar os testes da razão de verossimilhanças, Wald, escore e F os quais serão descritos a seguir.

Teste da Razão de Verossimilhanças

O teste da razão de verossimilhanças envolve a comparação dos valores do logaritmo da função de verossimilhança maximizada sem restrição e sob H_0 . O teste fica simplificado na forma

$$\xi_{RV} = 2\{L(\hat{\beta}_1, \hat{\beta}_2) - L(\beta_1^0, \tilde{\beta}_2)\}.$$

Teste de Wald

O teste de Wald é expresso na forma

$$\xi_W = [\hat{\beta}_1 - \beta_1^0]^\top \hat{\text{Var}}^{-1}(\hat{\beta}_1) [\hat{\beta}_1 - \beta_1^0],$$

em que $\hat{\beta}_1$ sai da partição $\hat{\beta} = (\hat{\beta}_1^\top, \hat{\beta}_2^\top)^\top$. Usando resultados conhecidos de álgebra de matrizes, envolvendo partição de matrizes, vide, por exemplo, (Searle, 1982), tem-se que a variância assintótica de $\hat{\beta}_1$ tem a forma

$$\text{Var}(\hat{\beta}_1) = \phi^{-1} [\mathbf{J}_1^\top \mathbf{W}^{1/2} \mathbf{M}_2 \mathbf{W}^{1/2} \mathbf{J}_1]^{-1},$$

em que $\mathbf{J}_1$ sai da partição $\mathbf{J} = [\mathbf{J}_1, \mathbf{J}_2]$, sendo portanto de dimensão $n \times q$ e $\mathbf{J}_2$ de dimensão $n \times (p-q)$, $\mathbf{M}_2 = \mathbf{I} - \mathbf{H}_2$ e $\mathbf{H}_2 = \mathbf{W}^{1/2} \mathbf{J}_2 (\mathbf{J}_2^\top \mathbf{W} \mathbf{J}_2)^{-1} \mathbf{J}_2^\top \mathbf{W}^{1/2}$ é a matriz de projeção ortogonal de vetores do $\Re^n$ no subespaço gerado pelas colunas da matriz $\mathbf{W}^{1/2} \mathbf{J}_2$.

Teste Escore

O teste escore, também conhecido como teste de Rao, é definido como

$$\xi_{SR} = \mathbf{U}_1(\tilde{\beta})^\top \tilde{\text{Var}}(\hat{\beta}_1) \mathbf{U}_1(\tilde{\beta}),$$

em que $\mathbf{U}_1(\tilde{\beta}) = \partial L(\boldsymbol{\theta}) / \partial \beta_1 |_{\tilde{\beta}} = \phi \tilde{\mathbf{J}}_1^\top \tilde{\mathbf{W}} \tilde{\mathbf{D}}(\mathbf{y} - \tilde{\boldsymbol{\mu}})$ e $\tilde{\text{Var}}(\hat{\beta}_1)$ denota a matriz de variância-covariância assintótica de $\hat{\beta}_1$ avaliada em H_0 .

As estatísticas ξ_{RV} , ξ_W e ξ_{SR} convergem, assintoticamente e sob H_0 , para uma distribuição qui-quadrado central com q graus de liberdade.

Teste F

Podemos definir a estatística F como

$$F = \frac{\{D(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - D(\mathbf{y}; \hat{\boldsymbol{\mu}})\}/p}{D(\mathbf{y}; \hat{\boldsymbol{\mu}})/(n-p)},$$

cuja distribuição nula assintótica é uma $F_{p,(n-p)}$ e $D(\mathbf{y}; \hat{\boldsymbol{\mu}}^0)$ e $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ são as funções desvio correspondentes aos modelos sob H_0 e H_1 , respectivamente, em que $\hat{\boldsymbol{\mu}}^0$ é a estimativa de máxima verossimilhança sob H_0 .

ϕ desconhecido

Suponha ϕ desconhecido e o vetor de parâmetros $\boldsymbol{\theta}$ dado por

$$\boldsymbol{\theta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \phi)^\top,$$

em que $\boldsymbol{\beta}_1$ é o vetor de interesse de dimensão q e $\boldsymbol{\beta}_2$ é o vetor de parâmetros de perturbação de dimensão $p - q$ e ϕ^{-1} é um parâmetro de dispersão. Suponha que iremos testar

$$\begin{aligned} H_0 : \boldsymbol{\beta}_1 &= \boldsymbol{\beta}_1^0 \\ H_1 : \boldsymbol{\beta}_1 &\neq \boldsymbol{\beta}_1^0, \end{aligned}$$

em que $\boldsymbol{\beta}_1^0$ é um vetor conhecido. Considerando a partição do vetor $\boldsymbol{\theta}$ acima, temos a matriz Jacobiana e o vetor escore de $\boldsymbol{\theta}$ dados, respectivamente, por $\mathbf{J} = [\mathbf{J}_1, \mathbf{J}_2]$ e $\mathbf{U}_{\boldsymbol{\theta}} = [\mathbf{U}_{\boldsymbol{\beta}_1}^\top, \mathbf{U}_{\boldsymbol{\beta}_2}^\top, U_\phi]^\top$. Então podemos escrever a matriz de informação de Fisher como

$$\mathbf{K}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \begin{pmatrix} \mathbf{K}_{\boldsymbol{\beta}_1\boldsymbol{\beta}_1} & \mathbf{K}_{\boldsymbol{\beta}_1\boldsymbol{\beta}_2} & \mathbf{0} \\ \mathbf{K}_{\boldsymbol{\beta}_2\boldsymbol{\beta}_1} & \mathbf{K}_{\boldsymbol{\beta}_2\boldsymbol{\beta}_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & K_{\phi\phi} \end{pmatrix},$$

em que $\mathbf{K}_{\boldsymbol{\beta}_1\boldsymbol{\beta}_1} = \phi \mathbf{J}_1^\top \mathbf{W} \mathbf{J}_1$, $\mathbf{K}_{\boldsymbol{\beta}_1\boldsymbol{\beta}_2} = \mathbf{K}_{\boldsymbol{\beta}_2\boldsymbol{\beta}_1}^\top = \phi \mathbf{J}_1^\top \mathbf{W} \mathbf{J}_2$, $\mathbf{K}_{\boldsymbol{\beta}_2\boldsymbol{\beta}_2} = \phi \mathbf{J}_2^\top \mathbf{W} \mathbf{J}_2$ e $K_{\phi\phi} = E[-\sum_{i=1}^n c''(Y_i, \phi)]$.

Sejam $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}_1^\top, \hat{\boldsymbol{\beta}}_2^\top, \hat{\phi})^\top$ e $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\beta}}_1^{0\top}, \tilde{\boldsymbol{\beta}}_2^\top, \tilde{\phi})^\top$ os estimadores de máxima verossimilhança de $\boldsymbol{\theta}$ sob H_1 e H_0 , respectivamente. Assim, a estatística da razão de verossimilhanças fica expressa na forma

$$\xi_{RV} = 2\{L(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\phi}) - L(\boldsymbol{\beta}_1^0, \tilde{\boldsymbol{\beta}}_2, \tilde{\phi})\}.$$

A estatística de Wald, por sua vez, fica dada por

$$\xi_W = \widehat{\phi}[\widehat{\beta}_1 - \beta_1^0]^\top [\widehat{\mathbf{J}}_1^\top \widehat{\mathbf{W}}^{1/2} \widehat{\mathbf{M}}_2 \widehat{\mathbf{W}}^{1/2} \widehat{\mathbf{J}}_1] [\widehat{\beta}_1 - \beta_1^0],$$

em que $\widehat{\mathbf{J}}_1, \widehat{\mathbf{W}}$ e $\widehat{\mathbf{M}}_2$ são os mesmos descritos na seção anterior, avaliados em H_1 .

Já a estatística de escore fica dada por

$$\xi_{SR} = \mathbf{U}_1(\widetilde{\beta})^\top \widetilde{\text{Var}}(\widehat{\beta}_1) \mathbf{U}_1(\widetilde{\beta}),$$

em que $\mathbf{U}_1(\widetilde{\beta})$ e $\widetilde{\text{Var}}(\widehat{\beta}_1)$ são os mesmos descritos na seção anterior, avaliados em H_0 .

2.7 Região de Confiança Assintótica

Supondo ϕ conhecido, uma região assintótica de confiança para β baseada no teste de Wald e com coeficiente de confiança $(1 - \alpha)$ é dada por

$$[\beta; (\widehat{\beta} - \beta)^\top (\widehat{\mathbf{J}}^\top \widehat{\mathbf{W}} \widehat{\mathbf{J}}) (\widehat{\beta} - \beta) \leq \phi^{-1} \chi_p^2(1 - \alpha)],$$

em que $\chi_p^2(1 - \alpha)$ denota o percentil $(1 - \alpha)$ de uma distribuição qui-quadrado com p graus de liberdade. Como essa região pode depender da parametrização utilizada quando η é não linear, vide, por exemplo, (Ratkowsky, 1983), pode ser mais conveniente, nesses casos, construir a região utilizando uma das estatísticas invariantes. Em particular, se a estatística da razão de verossimilhanças for escolhida, a região assintótica fica dada por

$$[\beta; 2\{L(\widehat{\beta}) - L(\beta)\} \leq \chi_p^2(1 - \alpha)].$$

Se estamos interessados num subconjunto β_1 , q -dimensional, a região assintótica de confiança utilizando as estatísticas de Wald e da razão de verossimilhanças ficam, respectivamente, dadas por

$$[\beta; (\widehat{\beta}_1 - \beta)^\top \widehat{\text{Var}}^{-1}(\widehat{\beta}_1) (\widehat{\beta}_1 - \beta) \leq \phi^{-1} \chi_q^2(1 - \alpha)] \quad \text{e} \quad [\beta; 2\{L(\widehat{\beta}) - L(\beta, \widehat{\beta}_2(\beta))\} \leq \chi_q^2(1 - \alpha)],$$

em que β é aqui q -dimensional e $\hat{\beta}_2(\beta)$ é a estimativa de máxima verossimilhança de β_2 dado β , vide, por exemplo, (Seber & Wild, 1989). Regiões invariantes de confiança para β podem também ser construídas de forma similar usando as estatísticas F e de escore.

2.8 Viés de ordem n^{-1}

Como é bem conhecido da regressão normal não linear os estimadores de mínimos quadrados são viesados para uma amostra finita e o viés depende do tamanho da amostra e da parametrização usada, vide, por exemplo, (Ratkowsky, 1983). Assim, a escolha de uma parametrização conveniente para $\eta(\beta)$ no sentido de produzir vieses pequenos para $\hat{\beta}$ tem sido um tópico relevante de pesquisa em modelos normais não lineares. Box (1971) apresenta uma expressão fechada para o viés de ordem n^{-1} em modelos normais não lineares. Paula (1992) usando expansões de Cox & Snell (1968) mostra que o viés de ordem n^{-1} para o estimador de máxima verossimilhança $\hat{\beta}$ em modelos não lineares de família exponencial pode ser expresso como sendo a solução de mínimos quadrados de uma regressão linear ponderada. Ou seja, se denotarmos por $b(\hat{\beta})$ o viés de ordem n^{-1} para $\hat{\beta}$, Paula (1992) mostra que

$$b(\hat{\beta}) = (J^\top W J)^{-1} J^\top W (\xi_1 + \xi_2), \quad (2.8)$$

em que $\xi_1 = -(2\phi)^{-1} Z_d W^{-1} F \mathbf{1}$, $\xi_2 = -(2\phi)^{-1} D \mathbf{1}$, e $\mathbf{1}$ é um vetor $n \times 1$ de uns, Z_d é uma matriz $n \times n$ com elementos $z_{ll} = j_l^\top (J^\top W J)^{-1} j_l$, j_l sendo a l -ésima linha de J , $D = \text{diag}\{d_1, \dots, d_n\}$, $d_l = \text{tr}\{J_l (J^\top W J)^{-1}\}$, J_l é uma matriz $p \times p$ com elementos $\partial^2 f(x_l; \beta) / \partial \beta_r \partial \beta_s$ ($r, s = 1, \dots, p$) e $F = \text{diag}\{f_1, \dots, f_n\}$ tal que $f_l = V_l^{-1} (d\mu_l / d\eta_l) (d^2 \mu_l / d\eta_l^2)$. Portanto, $b(\hat{\beta})$ pode ser obtido como sendo a solução de mínimos quadrados ponderados da regressão linear de $\xi_1 + \xi_2$ sobre as colunas de J com matriz de pesos W .

Em particular, para MLGs tem-se que $\xi_2 = 0$ e $J = X$ de modo que o viés de ordem n^{-1} para $\hat{\beta}$ assume a forma (Cordeiro & McCullagh, 1991)

$$b(\hat{\beta}) = (X^\top W X)^{-1} X^\top W \xi_1,$$

em que $z_{ll} = x_l^\top (X^\top W X)^{-1} x_l$, x_l sendo a l -ésima linha da matriz modelo X . Já para os modelos

normais não lineares temos que $\boldsymbol{\xi}_1 = 0$ e $\mathbf{W} = \mathbf{I}$, ficando o viés de ordem n^{-1} para $\hat{\boldsymbol{\beta}}$ expresso na forma (Box, 1971)

$$\mathbf{b}(\hat{\boldsymbol{\beta}}) = (\mathbf{J}^\top \mathbf{J})^{-1} \mathbf{J}^\top \boldsymbol{\xi}_2,$$

em que $d_l = \text{tr}\{\mathbf{J}_l(\mathbf{J}^\top \mathbf{J})^{-1}\}$.

A estimativa de máxima verossimilhança corrigida pelo viés estimado de ordem n^{-1} fica assim dada por

$$\hat{\boldsymbol{\beta}}^* = \hat{\boldsymbol{\beta}} - \hat{\mathbf{b}}(\hat{\boldsymbol{\beta}}),$$

em que $\hat{\mathbf{b}}(\hat{\boldsymbol{\beta}})$ é o viés de ordem n^{-1} avaliado em $\hat{\boldsymbol{\beta}}$.

Similarmente é possível obter uma expressão em forma fechada para o viés de ordem n^{-1} para $\hat{\phi}$. Usando a relação $c(y, \phi) = d(\phi) + \phi a(y) + u(y)$ válida para algumas distribuições da família exponencial tais como normal, normal inversa e gama, Cordeiro & McCullagh (1991) mostram que

$$b(\hat{\phi}) = \{nd''(\phi)\}^{-1} \left\{ \frac{d'''(\phi)}{2d''(\phi)} - (2\phi)^{-1}p \right\}, \quad (2.9)$$

em que $d''(\phi)$ e $d'''(\phi)$ denotam, respectivamente, as derivadas de segunda e terceira ordens de $d(\phi)$ com relação a ϕ . Por exemplo, para um modelo gama temos, $d(\phi) = \phi \log \phi - \log \Gamma(\phi)$, $d''(\phi) = 1/\phi - \Psi'(\phi)$ e $d'''(\phi) = -1 - \Psi''(\phi)$, sendo que $\Psi(\cdot) = \Gamma'(\cdot)/\Gamma(\cdot)$ denota a função digama e $\Psi', \Psi'', \dots$ são as derivadas sucessivas dessa função. Assim, substituindo os valores na equação acima temos

$$b(\hat{\phi}) = \left\{ \frac{n}{\phi} - n\Psi'(\phi) \right\}^{-1} \left\{ \frac{-(1 + \Psi''(\phi))}{2\left(\frac{1}{\phi} - \Psi'(\phi)\right)} - (2\phi)^{-1}p \right\}.$$

A estimativa de máxima verossimilhança corrigida fica portanto dada por

$$\hat{\phi}^* = \hat{\phi} - \hat{b}(\hat{\phi}),$$

em que $\hat{b}(\hat{\phi})$ é o viés de ordem n^{-1} avaliado em $\hat{\phi}$.

2.9 Binomial Negativa Não Linear

A distribuição binomial negativa é comumente utilizada para ajustar dados de contagem em que a variância é maior do que a média, ou seja, quando ocorre o fenômeno de superdispersão ou sobredispersão. Para construir a distribuição binomial negativa podemos supor que a distribuição condicional de $Y|Z = z \sim P(z)$ enquanto que $Z \sim G(\mu, \nu)$. Daí segue que

$$E(Y) = E\{E(Y|Z)\} = \mu \quad \text{e}$$

$$\begin{aligned} \text{Var}(Y) &= E\{\text{Var}(Y|Z)\} + \text{Var}\{E(Y|Z)\} \\ &= \mu + \frac{\mu^2}{\nu}, \end{aligned}$$

portanto temos $\text{Var}(Y) > E(Y)$, uma vez que $\nu > 0$.

Para obter a distribuição marginal de $\mathbf{Y}$ temos que resolver a integral abaixo

$$f_Y(y; \mu, \nu) = \int_0^\infty g(y|z)h(z; \mu, \nu)dz, \quad (2.10)$$

em que $g(y|z)$ é a função de probabilidades de uma Poisson de média z e $h(z; \mu, \nu)$ é a função densidade de uma $G(\mu, \nu)$. A solução de (2.10) foi demonstrada em Svetliza (2002) e é dada por

$$f_Y(y; \mu, \nu) = \frac{\Gamma(y + \nu)}{\Gamma(y + 1)\Gamma(\nu)} \left(\frac{\mu}{\mu + \nu} \right)^y \left(\frac{\nu}{\mu + \nu} \right)^\nu, \quad (2.11)$$

para $y = 0, 1, 2, \dots$. Tem-se portanto em (2.11) a função de probabilidades de uma distribuição binomial negativa de média μ e parâmetro de dispersão $\nu > 0$ desconhecido. Denotaremos $Y \sim \text{BN}(\mu, \nu)$. Quando ν é conhecido é possível mostrar que Y pertence à família exponencial de distribuições, com $\theta = \log \{\mu/\mu + \nu\}$, $b(\theta) = -\nu \log \{\nu/\mu + \nu\}$, $c(y, \nu) = \Gamma(y + \nu)/\Gamma(y + 1)\Gamma(\nu)$ e $y > 0$.

Vamos supor agora que $Y_1, \dots, Y_n$ são variáveis aleatórias independentes tais que $Y_i \sim \text{BN}(\mu_i, \nu)$, ν é desconhecido. Os modelos não lineares com resposta binomial negativa foram in-

troduzidos por Svetliza (2002) (ver também, Svetliza & Paula (2003)) e são definidos supondo parte sistemática dada por

$$g(\mu_i) = \eta_i = f(x_i; \boldsymbol{\beta}),$$

em que $g(\cdot)$ e $f(\cdot; \cdot)$ são definidos como na classe dos modelos não lineares de família exponencial. Quando ν é conhecido os modelos não lineares com resposta binomial negativa fazem parte da classe dos modelos não lineares de família exponencial.

Seja $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \nu)^\top$ então o logaritmo da função de verossimilhança para o modelo binomial negativa fica dado por

$$L(\boldsymbol{\theta}) = \sum_{i=1}^n \left[\log \left\{ \frac{\Gamma(y_i + \nu)}{y_i! \Gamma(\nu)} \right\} + y_i \log \left\{ \frac{\mu_i}{\mu_i + \nu} \right\} + \nu \log \left\{ \frac{\nu}{\mu_i + \nu} \right\} \right],$$

em que $\mu_i = g^{-1}\{f(x_i; \boldsymbol{\beta})\}$ e $\Gamma(\cdot)$ é a função gama.

Função Desvio e Escore

A função desvio supondo ν fixo ou conhecido fica dada por

$$D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \left[\nu \log \left\{ \frac{\hat{\mu}_i + \nu}{y_i + \nu} \right\} + y_i \log \left\{ \frac{y_i(\hat{\mu}_i + \nu)}{\hat{\mu}_i(y_i + \nu)} \right\} \right],$$

em que $\mu_i = g^{-1}\{f(x_i; \hat{\boldsymbol{\beta}})\}$. Na prática ν é substituído pela estimativa de máxima verossimilhança $\hat{\nu}$. Para ν grande e μ_i grande $\forall_i$ tem-se que $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) \sim \chi_{n-p}^2$. Estamos assumindo na expressão acima que $y_i > 0, \forall_i$. Quando $y_i = 0$ para algum i , o i -ésimo componente do desvio fica dado por

$$2\nu \log \left(\frac{y_i + \nu}{\hat{\mu}_i + \nu} \right).$$

As funções escore para $\boldsymbol{\beta}$ e ν ficam, respectivamente, dadas por

$$\mathbf{U}_\beta(\boldsymbol{\theta}) = \mathbf{J}^\top \mathbf{W} \mathbf{D}(\mathbf{y} - \boldsymbol{\mu}) \quad \text{e} \quad U_\nu(\boldsymbol{\theta}) = \sum_{i=1}^n \left[\Psi(\nu + y_i) - \Psi(\nu) - \frac{(y_i + \nu)}{(\nu + \mu_i)} + \log \left\{ \frac{\nu}{\nu + \mu_i} \right\} + 1 \right],$$

em que $\mathbf{J} = \partial \boldsymbol{\eta} / \partial \boldsymbol{\beta}$, $\mathbf{W} = \text{diag}\{\omega_1, \dots, \omega_n\}$ com $\omega_i = (d\mu_i/d\eta_i)^2/(\mu_i + \mu_i^2\nu^{-1})$, $\mathbf{D} = \text{diag}\{d\eta_i/d\mu_i\}$, $\mathbf{y} = (y_1, \dots, y_n)^\top$ e $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^\top$. Aqui também temos a ortogonalidade entre $\boldsymbol{\beta}$ e ν , de modo que $\mathbf{K}_{\beta\nu}(\boldsymbol{\theta}) = \mathbf{K}_{\nu\beta}^\top(\boldsymbol{\theta}) = \mathbf{0}$. Além disso, mostra-se que

$$\mathbf{K}_{\beta\beta}(\boldsymbol{\theta}) = \mathbf{E} \left\{ -\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \right\} = \mathbf{J}^\top \mathbf{W} \mathbf{J}$$

e

$$K_{\nu\nu}(\boldsymbol{\theta}) = \mathbf{E} \left\{ -\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \nu^2} \right\} = \sum_{i=1}^n \left\{ \sum_{j=0}^{\infty} (\nu + j)^{-2} \mathbf{P}_r(Y_i \geq j) - \frac{\nu^{-1} \mu_i}{\mu_i + \nu} \right\},$$

em que $Y_i \sim \text{BN}(\mu_i, \nu)$.

Estimação de $\boldsymbol{\beta}$ e ν

O processo iterativo conjunto para obter as estimativas de máxima verossimilhança $\hat{\boldsymbol{\beta}}$ e $\hat{\nu}$ é dado por

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{J}^{(m)\top} \mathbf{W}^{(m)} \mathbf{J}^{(m)})^{-1} \mathbf{J}^{(m)\top} \mathbf{W}^{(m)} \mathbf{y}^{*(m)} \quad (2.12)$$

e

$$\nu^{(m+1)} = \nu^{(m)} - \{U'_\nu(\boldsymbol{\theta})\}^{-1} U'_\nu(\boldsymbol{\theta}), \quad (2.13)$$

em que $\mathbf{y}^{*(m)} = \mathbf{J}^{(m)} \boldsymbol{\beta}^{(m)} + \mathbf{D}^{(m)}(\mathbf{y} - \boldsymbol{\mu}^{(m)})$ e $U'_\nu(\boldsymbol{\theta}) = \partial U_\nu(\boldsymbol{\theta}) / \partial \nu$. É preciso inicializar o processo acima com valores iniciais $\boldsymbol{\beta}^{(0)}$ e $\nu^{(0)}$.

Similarmente ao procedimento iterativo descrito na Seção 2.4 para obter a estimativa de máxima verossimilhança $\hat{\boldsymbol{\beta}}$ nos modelos não lineares de família exponencial, é possível desenvolver um procedimento iterativo para obter $\hat{\nu}$ e $\hat{\boldsymbol{\beta}}$ em modelos não lineares com resposta binomial negativa sendo que para conseguirmos implementar (2.12) e (2.13) no S-Plus ou no R basta utilizarmos a

library MASS (Venables & Ripley, 1999) e a facilidade *offset*, reescrevendo a variável dependente localmente modificada $\mathbf{y}^*$, na forma $\mathbf{y}^* = \boldsymbol{\eta} - \boldsymbol{\tau} + \mathbf{D}(\mathbf{y} - \boldsymbol{\mu})$, sendo $\boldsymbol{\eta} = f(\mathbf{J}; \boldsymbol{\beta})$ e $\boldsymbol{\tau} = f(\mathbf{J}; \boldsymbol{\beta}) - \mathbf{J}\boldsymbol{\beta}$, em que $f(\mathbf{J}; \boldsymbol{\beta})$ é um vetor $n \times 1$ de componentes $f(\mathbf{x}_i; \boldsymbol{\beta})$, $i = 1, \dots, n$. Dessa forma, a expressão (2.12) pode ser interpretada como um processo iterativo para ajustar um modelo linear generalizado com matriz modelo $\mathbf{J}$, parte sistemática $g(\boldsymbol{\mu}) = \boldsymbol{\eta}$, função de variância $\mathbf{V} = \boldsymbol{\mu}^2 \nu^{-1} + \boldsymbol{\mu}$ e $\boldsymbol{\tau}$ como sendo o *offset*, o qual é uma quantidade que é subtraída, a cada passo, do preditor $\boldsymbol{\eta}$, vide, por exemplo, (Cordeiro & Paula, 1989). A única diferença com relação aos MLGs é que aqui a matriz $\mathbf{J}$ e ν são também modificadas a cada passo do processo iterativo com ν sendo modificado do processo iterativo (2.13). Procedimentos alternativos para encontrar as estimativas de máxima verossimilhança $\hat{\boldsymbol{\beta}}$ e $\hat{\nu}$ estão disponíveis em vários aplicativos como R, S-Plus, SAS, Ox e Mathematica.

Distribuição Assintótica de $\hat{\boldsymbol{\beta}}$ e $\hat{\nu}$

Similarmente como descrito na Seção 2.5 tem-se que $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\nu})^\top$, quando $n \rightarrow \infty$, converge em distribuição, sob certas condições gerais de regularidade, vide, por exemplo, (Sen & Singer (1993), Svetliza (2002)), para

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\nu} \end{pmatrix} \sim N_{p+1} \left[\begin{pmatrix} \boldsymbol{\beta} \\ \phi \end{pmatrix}, \begin{pmatrix} (\mathbf{J}^\top \mathbf{W} \mathbf{J})^{-1} & \mathbf{0} \\ \mathbf{0} & K_{\nu\nu}^{-1}(\boldsymbol{\theta}) \end{pmatrix} \right].$$

A matriz de informação de Fisher total se reduz à matriz bloco diagonal $\mathbf{K}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \text{diag}\{\mathbf{K}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}), K_{\nu\nu}(\boldsymbol{\theta})\}$ em que $\mathbf{K}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}) = \mathbf{J}^\top \mathbf{W} \mathbf{J}$ e $K_{\nu\nu}(\boldsymbol{\theta}) = \sum_{i=1}^n \left\{ \sum_{j=0}^{\infty} (\nu + j)^{-2} P_r(Y_i \geq j) - \nu^{-1} \mu_i / (\mu_i + \nu) \right\}$ que é uma função de $\boldsymbol{\beta}$ e ν . Os parâmetros $\boldsymbol{\beta}$ e ν são ortogonais, ou seja, os estimadores de máxima verossimilhança de $\boldsymbol{\beta}$ e ν são assintoticamente independentes.

Região de Confiança Assintótica

Da mesma maneira como descrito na Seção 2.7 podemos encontrar uma região assintótica para $\boldsymbol{\beta}$ quando ν for conhecido baseada no teste de Wald e com coeficiente de confiança $(1 - \alpha)$, a qual é

dada por

$$[\beta; (\hat{\beta} - \beta)^\top (\hat{\mathbf{J}}^\top \widehat{\mathbf{W}} \hat{\mathbf{J}}) (\hat{\beta} - \beta) \leq \nu^{-1} \chi_p^2(1 - \alpha)],$$

em que $\chi_p^2(1 - \alpha)$ denota o percentil $(1 - \alpha)$ de uma distribuição qui-quadrado com p graus de liberdade. Se utilizarmos a estatística da razão de verossimilhanças por não depender da parametrização utilizada quando $\boldsymbol{\eta}$ é não linear, vide, por exemplo, (Ratkowsky, 1983), então a região assintótica fica dada por

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

Viés de ordem n^{-1}

Se considerarmos ν conhecido, podemos denotar o viés de ordem n^{-1} para $\hat{\beta}$ nos modelos não lineares de família exponencial como o mesmo viés mostrado por Paula (1992) e citado na Seção 2.8 pela equação (2.8).

Teste de hipótese

Assim como descrito na Seção 2.6 e assumindo $\phi = \nu$ desconhecido, temos as estatísticas da razão de verossimilhanças, de Wald e de escore dadas respectivamente por:

$$\xi_{RV} = 2\{L(\hat{\beta}_1, \hat{\beta}_2, \hat{\nu}) - L(\beta_1^0, \tilde{\beta}_2, \tilde{\nu})\},$$

$$\xi_W = [\hat{\beta}_1 - \beta_1^0]^\top \hat{\text{Var}}^{-1}(\hat{\beta}_1) [\hat{\beta}_1 - \beta_1^0] \quad \text{e}$$

$$\xi_{SR} = \mathbf{U}_1(\tilde{\beta})^\top \tilde{\text{Var}}(\hat{\beta}_1) \mathbf{U}_1(\tilde{\beta}),$$

em que $\hat{\text{Var}}^{-1}(\hat{\beta}_1)$ e $\tilde{\text{Var}}(\hat{\beta}_1)$ são os mesmos descritos na Seção 2.6 e $\hat{\text{Var}}(\hat{\beta}_1)$ e $\tilde{\text{Var}}(\hat{\beta}_1)$ denotam as matrizes de variância-covariância assintóticas de $\hat{\beta}_1$ avaliadas em H_1 e H_0 , respectivamente.

Os modelos não lineares de família exponencial são deduzidos da mesma forma como em regressão linear, ou seja, através das suposições iniciais que devem ser verificadas após o ajustamento do modelo. Para tanto são utilizadas técnicas de diagnóstico que serão vistas no próximo capítulo.

Capítulo 3

Métodos de Diagnóstico

As técnicas de diagnóstico iniciaram com a análise de resíduos para detectar a presença de pontos aberrantes e verificar se a distribuição proposta para a variável resposta é adequada. Os trabalhos mais conhecidos e usados até hoje são os de Anscombe (1953) e Cox & Snell (1968). Um outro interesse, é avaliar possíveis afastamentos das suposições iniciais, especialmente na parte aleatória e na parte sistemática, bem como avaliar alguma interferência desproporcional nos resultados do ajuste do modelo. Com esse objetivo foram desenvolvidas outras técnicas de diagnóstico, além da análise de resíduo.

Em modelos lineares generalizados, por exemplo, McCullagh (1987), Pregibon (1981), Williams (1987) e Paula (1995) apresentaram versões padronizadas para o resíduo componente do desvio. Davison & Gigli (1989) estenderam a proposta de Cox & Snell (1968) e definiram uma forma mais geral de padronização para o resíduo componente do desvio para distribuições contínuas, mesmo quando a função de distribuição acumulada não é expressa em forma fechada, enquanto Farhrmeir & Tutz (1994) estenderam o trabalho de McCullagh (1987) para modelos mais gerais, não pertencentes à família exponencial de distribuições. Therneau et al. (1990) consideram o resíduo componente do desvio em processos de contagem e Souza & Paula (2002) usaram o método proposto por Davison & Gigli (1989) a fim de obterem uma forma padronizada para o resíduo componente do desvio em modelos de regressão von Mises, os quais têm sido aplicados na análise de dados circulares. Recentemente, Leiva et al. (2007), Ortega et al. (2008) e Barros et al. (2008) propuseram resíduos

tipo martingale em modelos paramétricos para a análise de dados de sobrevivência.

Um outro tópico importante na análise de diagnóstico é detectar pontos de alavanca, ou seja, aquelas observações que exercem um peso desproporcional no próprio valor ajustado. Hoaglin & Welsch (1978) apresentaram a matriz de projeção $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$, em que $\mathbf{X}$ denota a matriz modelo, que motivou a definição de pontos de alavanca em modelos lineares os quais são determinados pelos elementos da diagonal principal dessa matriz. Extensões da definição de pontos de alavanca para modelos normais não lineares são apresentadas em Saint Laurent & Cook (1992). Wei et al. (1998) generalizaram a definição de pontos de alavanca para modelos mais gerais cuja variável resposta seja contínua. Paula (1999) discutiu pontos de alavanca em modelos de regressão com parâmetros restritos e com extensões para os MLGs.

Uma outra técnica baseada em perturbações no modelo, particularmente na eliminação de pontos, talvez seja a mais conhecida para avaliar o impacto nas estimativas dos parâmetros com a retirada de uma observação do modelo. A distância de Cook (1977) foi desenvolvida para modelos normais lineares sendo rapidamente estendida e assimilada para diversas classes de modelos. Por exemplo, Moolgavkar et al. (1984) apresentaram resultados para uma classe geral de modelos não lineares, com aplicações aos estudos emparelhados de caso e controle. Storer & Crowley (1985) estudaram as mudanças nas estimativas dos modelos condicionais de análise de sobrevivência. Ross (1987) discutiu a geometria da exclusão de casos em regressão não linear. Cook et al. (1988) compararam o afastamento da verossimilhança com medidas tradicionais de exclusão de pontos tais como a distância de Cook. Davison & Tsai (1992) e Cordeiro & Paula (1992) estenderam a distância de Cook para modelos que não pertencem à família exponencial de distribuições, enquanto que Galea et al. (2000) investigaram a distância de Cook em modelos elípticos multivariados.

Quando excluímos individualmente um ponto, podemos deixar de detectar pontos conjuntamente discrepantes. Esse efeito é conhecido como *masking effect*. Porém, a exclusão múltipla de pontos não é muito comum principalmente pelo custo computacional envolvido. Entretanto, Fung (1993) e Peña & Yohai (1999) apresentaram vários procedimentos robustos na exclusão de pontos com custos computacionais relativamente baixos.

A influência local proposta por Cook (1986) tornou-se uma ferramenta muito popular para avaliar a influência conjunta das observações sob pequenas perturbações (mudanças) no modelo ou nos dados, ao invés da avaliação da retirada individual ou conjunta de pontos. Nos últimos 20 anos inúmeros artigos foram publicados nesse assunto. Por exemplo, sob erros normais, Lawrence (1988) investigou a aplicação de influência local em modelos lineares com parâmetros na transformação da resposta. Beckman et al. (1987) apresentaram estudos de influência em modelos de análise de variância com efeito misto, Tsai & Wu (1992) investigaram influência local em modelos autoregressivos de primeira ordem e modelos heteroscedásticos e Paula (1993) aplicou influência local em modelos lineares com restrições nos parâmetros. Saindo da classe de erros normais, Pettitt & Bin Daud (1989) investigaram influência local em modelos de regressão de Cox com riscos proporcionais. Escobar & Meeker (1992) adaptaram o método de influência local numa classe paramétrica de modelos para análise de sobrevivência; Kim (1995), O'Hara Hines et al. (1992) e Pan et al. (1997) aplicaram métodos de influência local em regressão multivariada; Galea et al. (1997), Liu (2000), Galea et al. (2003) e Osorio et al. (2007) apresentaram estudos de influência local em modelos de contornos elípticos; Kwan & Fung (1998) aplicaram influência local em análise fatorial; Gu & Fung (1998) em análise de correlação canônica; Paula (1996) em modelos próprios de dispersão e Ortega et al. (2003) em modelos log-gama generalizados com dados censurados. Rancel & Sierra (2001) fizeram uma revisão de influência local e Tang et al. (2001) investigaram a influência local em modelos de dispersão reprodutivos. Recentemente, Barros et al. (2008) desenvolveram métodos de influência local em modelos Birnbaum-Saunders com erros t de Student e dados censurados.

3.1 Pontos de Alavanca

Uma possível definição de ponto de alavanca em modelos não lineares de família exponencial é baseada na analogia entre o estimador de máxima verossimilhança $\hat{\beta}$ do modelo e a solução de mínimos quadrados de uma regressão normal ponderada. Para ver isso, note que na convergência do processo iterativo dado em (2.5), tem-se o seguinte:

$$\hat{\beta} = (\hat{\mathbf{J}}^T \mathbf{W} \hat{\mathbf{J}})^{-1} \hat{\mathbf{J}}^T \mathbf{W} \mathbf{y}^*,$$

em que $\mathbf{W} = \text{diag}\{\omega_1, \dots, \omega_n\}$ é a matriz de pesos com $\omega_i = (d\mu_i/d\eta_i)^2/V_i$, $\mathbf{y}^* = \mathbf{J}\hat{\boldsymbol{\beta}} + \hat{\mathbf{D}}(\mathbf{y} - \hat{\boldsymbol{\mu}})$ pode ser interpretado como a solução de mínimos quadrados da regressão linear de $\widehat{\mathbf{W}}^{1/2}\mathbf{y}^*$ contra as colunas de $\widehat{\mathbf{W}}^{1/2}\hat{\mathbf{J}}$. Dessa maneira, a matriz de projeção ortogonal no subespaço vetorial tangente à solução de mínimos quadrados da regressão linear de $\mathbf{y}^*$ contra $\hat{\mathbf{J}}$ com pesos $\widehat{\mathbf{W}}$ pode ser expressa na forma

$$\hat{\mathbf{H}} = \widehat{\mathbf{W}}^{1/2}\hat{\mathbf{J}}(\hat{\mathbf{J}}^\top \widehat{\mathbf{W}}\hat{\mathbf{J}})^{-1}\hat{\mathbf{J}}^\top \widehat{\mathbf{W}}^{1/2}. \quad (3.1)$$

Sugere-se a utilização dos elementos da diagonal principal de $\hat{\mathbf{H}}$ para detectar a presença de pontos de alavanca. O i -ésimo elemento da diagonal principal de $\hat{\mathbf{H}}$ fica dado por

$$\hat{h}_{ii} = \hat{w}_i \hat{\mathbf{j}}_i^\top (\hat{\mathbf{J}}^\top \widehat{\mathbf{W}}\hat{\mathbf{J}})^{-1} \hat{\mathbf{j}}_i,$$

em que $\hat{\mathbf{j}}_i^\top$ denota a i -ésima linha da matriz $\mathbf{J}$.

Em particular, para ligação canônica, temos que

$$\hat{\mathbf{H}} = \hat{\mathbf{V}}^{1/2}\hat{\mathbf{J}}(\hat{\mathbf{J}}^\top \hat{\mathbf{V}}\hat{\mathbf{J}})^{-1}\hat{\mathbf{J}}^\top \hat{\mathbf{V}}^{1/2},$$

e portanto $\hat{h}_{ii} = \hat{v}_i \hat{\mathbf{j}}_i^\top (\hat{\mathbf{J}}^\top \hat{\mathbf{V}}\hat{\mathbf{J}})^{-1} \hat{\mathbf{j}}_i$, em que $\hat{\mathbf{V}} = \text{diag}\{\hat{\mathbf{V}}_1, \dots, \hat{\mathbf{V}}_n\}$.

Uma outra maneira de avaliar a influência do valor observado y_i sobre o próprio valor ajustado $\hat{y}_i$ pode ser bem representada pela derivada $\partial \hat{y}_i / \partial y_i$, que coincide com $\hat{h}_{ii} = \mathbf{x}_i^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_i$ no modelo normal linear e foi discutida por diversos autores, vide, por exemplo, (Hoaglin & Welsch (1978); Cook & Weisberg (1982); Emerson et al. (1984); Saint Laurent & Cook (1992) e Wei et al. (1998)) que propuseram uma forma bastante geral para $\partial \hat{\mathbf{y}} / \partial \mathbf{y}^\top$ quando a resposta é contínua, podendo a mesma ser aplicada em diversas situações de estimação até mesmo em modelos não lineares de família exponencial. A matriz $\partial \hat{\mathbf{y}} / \partial \mathbf{y}^\top$ de dimensão $n \times n$, a qual denominaram de matriz de alavancas generalizadas, pode ser obtida da forma geral

$$\widehat{\mathbf{GL}} = \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{y}^\top} = \mathbf{D}_\beta (-\ddot{\mathbf{L}}_{\beta\beta})^{-1} \ddot{\mathbf{L}}_{\beta\mathbf{y}} \Big|_{\hat{\boldsymbol{\beta}}},$$

em que $\mathbf{D}_\beta = \partial \boldsymbol{\mu} / \partial \boldsymbol{\beta}^\top$, $\ddot{\mathbf{L}}_{\beta\beta} = \partial^2 L(\boldsymbol{\beta}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top$ e $\ddot{\mathbf{L}}_{\beta\mathbf{y}} = \partial^2 L(\boldsymbol{\beta}) / \partial \boldsymbol{\beta} \partial \mathbf{y}^\top$. Mostra-se facilmente que

$$\mathbf{D}_\beta = \mathbf{N}\mathbf{J} \quad \text{e} \quad \ddot{\mathbf{L}}_{\beta\mathbf{y}} = \phi \mathbf{J}^\top \mathbf{V}^{-1} \mathbf{N},$$

em que $\mathbf{N} = \text{diag}\{d\mu_1/d\eta_1, \dots, d\mu_n/d\eta_n\}$. Se substituirmos $-\ddot{\mathbf{L}}_{\beta\beta}$ pelo seu valor esperado $\phi(\mathbf{J}^\top \mathbf{W} \mathbf{J})$, obtemos aproximadamente

$$\widehat{\mathbf{GL}} = \widehat{\mathbf{N}} \widehat{\mathbf{J}} (\widehat{\mathbf{J}}^\top \widehat{\mathbf{W}} \widehat{\mathbf{J}})^{-1} \widehat{\mathbf{J}}^\top \widehat{\mathbf{V}}^{-1} \widehat{\mathbf{N}}.$$

Assim, o elemento $\widehat{\mathbf{GL}}_{ii}$ pode ser expresso na forma

$$\widehat{\mathbf{GL}}_{ii} = \widehat{w}_i \widehat{\mathbf{j}}_i^\top (\widehat{\mathbf{J}}^\top \widehat{\mathbf{W}} \widehat{\mathbf{J}})^{-1} \widehat{\mathbf{j}}_i,$$

em que $w_i = (d\mu_i/d\eta_i)^2/V_i$. Em particular, para ligação canônica em que $-\ddot{\mathbf{L}}_{\beta\beta} = \phi(\mathbf{J}^\top \mathbf{V} \mathbf{J})$ obtemos exatamente

$$\widehat{\mathbf{GL}} = \widehat{\mathbf{V}} \widehat{\mathbf{J}} (\widehat{\mathbf{J}}^\top \widehat{\mathbf{V}} \widehat{\mathbf{J}})^{-1} \widehat{\mathbf{J}}^\top.$$

Portanto, $\widehat{\mathbf{GL}}_{ii} = \widehat{V}_i \widehat{\mathbf{j}}_i^\top (\widehat{\mathbf{J}}^\top \widehat{\mathbf{V}} \widehat{\mathbf{J}})^{-1} \widehat{\mathbf{j}}_i$ que coincide exatamente com $\hat{h}_{ii}$ para ligação canônica.

Assim, para grandes amostras espera-se que $\widehat{\mathbf{GL}}_{ii}$ e $\hat{h}_{ii}$ estejam muito próximos. Essas duas quantidades coincidem para ligação canônica. Como $\widehat{\mathbf{W}}$ e $\widehat{\mathbf{J}}$ dependem de $\widehat{\boldsymbol{\mu}}$, sugere-se os gráficos de $\hat{h}_{ii}$ ou $\widehat{\mathbf{GL}}_{ii}$ contra os valores ajustados para detectar pontos de alavanca. Deve-se olhar com mais atenção aquelas observações com valores relativamente altos para $\hat{h}_{ii}$ ou $\widehat{\mathbf{GL}}_{ii}$. Detalhes sobre o cálculo da matriz de alavancas generalizadas em modelos não lineares de família exponencial podem ser encontrados em Andrade (2004).

3.2 Resíduos

O resíduo para a i -ésima observação pode ser definido como uma função $r_i = r(y_i, \widehat{\mu}_i)$ que procura medir a discrepância entre o valor observado y_i e o valor ajustado $\widehat{y}_i$. O sinal de r_i indica a direção dessa discrepância.

A definição mais conhecida de resíduos no diagnóstico de modelos lineares e não lineares é a do resíduo ordinário, definido por $r_i = y_i - \hat{\mu}_i$. Porém, esses resíduos geralmente não possuem distribuição aproximadamente normal, exceto na regressão normal linear, o que dificulta a utilização das técnicas de diagnóstico existentes, as quais utilizam a distribuição normal padrão.

A definição de resíduo ordinário padronizado, também conhecido como resíduo de Pearson, é a seguinte:

$$r_{pi} = \frac{\phi^{1/2}(y_i - \hat{\mu}_i)}{\hat{V}_i^{1/2}}.$$

Esse resíduo é em geral viesado para modelos não lineares, dificultando o uso de procedimentos usuais de diagnóstico para detectar pontos aberrantes bem como afastamentos das suposições feitas para o modelo. Em particular, para modelos não lineares Cook & Tsai (1985) estudam o viés do resíduo ordinário e propõem um resíduo alternativo, denominado resíduo projetado, cujas propriedades estão mais próximas das propriedades do resíduo ordinário no caso normal linear. Por outro lado, Huet et al. (1996) propõem a utilização de métodos de *bootstrap* para estudar a distribuição empírica do resíduo r_{pi} em modelos normais não lineares. Acreditamos que essa metodologia possa ser estendida para o estudo de r_{pi} em modelos não lineares de família exponencial.

Outro resíduo bem conhecido é o resíduo ordinário studentizado que é definido na forma

$$t_{Si} = \frac{\phi^{1/2}(y_i - \hat{\mu}_i)}{\sqrt{\hat{V}_i(1 - \hat{h}_{ii})}},$$

com $\hat{h}_{ii}$ definido na seção anterior. A distribuição desse resíduo em MLGs é, em geral, assimétrica, mesmo para grandes amostras como mostrado por (Williams, 1984). Para modelos não lineares de família exponencial além de assimetria em geral observada para esse resíduo, dificultando o uso de técnicas tradicionais de análise de resíduos, tem-se também a possibilidade de vieses acentuados dependendo da parametrização adotada para o modelo. O uso de técnicas tipo *bootstrap* para estudar a distribuição empírica do resíduo é também recomendado neste caso.

O resíduo de Anscombe (1953) é definido da forma

$$t_{Ai} = \frac{\phi^{1/2}\{\varphi(y_i) - \varphi(\hat{\mu}_i)\}}{\hat{V}^{1/2}(\hat{\mu}_i)\varphi'(\hat{\mu}_i)},$$

em que $\varphi(\cdot)$ é uma transformação que busca normalidade e variância constante na distribuição de Y_i . Esse resíduo foi proposto a fim de contornar o problema da não normalidade. Porém, essa transformação é difícil de ser encontrada, havendo poucos modelos, vide, por exemplo, (Cox & Snell, 1968), para os quais $\varphi(\cdot)$ é conhecida.

Pregibon (1979) propôs para MLGs um resíduo que foi definido a partir dos componentes da função desvio e é definido na forma

$$t_{Di} = \frac{d^*(y_i, \hat{\mu}_i)}{\sqrt{1 - \hat{h}_{ii}}} = \frac{\phi^{1/2}d(y_i, \hat{\mu}_i)}{\sqrt{1 - \hat{h}_{ii}}},$$

em que $d(y_i, \hat{\mu}_i) = \text{sign}(y_i - \hat{\mu}_i)\sqrt{2}\{y_i(\hat{\theta}_i^0 - \hat{\theta}_i) + (b(\hat{\theta}_i) - b(\hat{\theta}_i^0))\}^{1/2}$, $\text{sign}(y_i - \hat{\mu}_i)$ é o sinal de $y_i - \hat{\mu}_i$. A distribuição empírica desse resíduo se aproxima em geral da normalidade ou pelo menos da simetria, como verificado por Williams (1984), para vários MLGs.

McCullagh (1987) mostra para MLGs que a distribuição de probabilidades de

$$\frac{d^*(Y_i; \mu_i) + \rho_{3i}/6}{\sqrt{1 + (14\rho_{3i}^2 - 9\rho_{4i})/36}},$$

é aproximadamente $N(0,1)$, em que ρ_{3i} e ρ_{4i} são coeficientes de assimetria e curtose de $\partial \mathbf{L}(\eta_i)/\partial \eta_i$, respectivamente. Cordeiro (1983) usa resultados de Cox & Snell (1968) para mostrar que

$$E\{d^*(Y_i; \mu_i)\} = 0 \quad \text{e} \quad \text{Var}\{d^*(Y_i; \mu_i)\} = 1 - h_{ii},$$

em que os termos negligenciados são de $O(n^{-1})$. Esses resultados reforçam a necessidade do uso da padronização $\sqrt{1 - \hat{h}_{ii}}$ para o resíduo $d^*(y_i; \hat{\mu}_i)$. Svetliza (2002) estende os resultados obtidos por (Cordeiro, 1983) para modelos não lineares com resposta binomial negativa. A extensão desses

resultados para modelos não lineares de família exponencial deve ser direta.

Um outro tipo de resíduo proposto por Williams (1987) para MLGs é interpretado como uma média ponderada entre o i -ésimo componente da função desvio e o correspondente resíduo studentizado, sendo da forma

$$t_{Gi} = \text{sign}(y_i, \hat{\mu}_i) \{ (1 - \hat{h}_{ii}) t_{Di}^2 + \hat{h}_{ii} t_{Si}^2 \}^{1/2}.$$

Williams (1987) verificou através de simulações de Monte Carlo para alguns MLGs, que t_{Gi} tem esperança diferente de zero, variância excedendo um, assimetria desprezível e alguma curtose; mesmo assim, o gráfico normal de probabilidades mostrou-se eficiente no diagnóstico de irregularidades com o ajuste, e foi mais revelador, quando utilizou-se das sugestões de Atkinson (1981), Atkinson (1985) que acrescentou bandas de confiança obtidos através de simulações com o modelo ajustado. Esse resíduo pode também ser aplicado em modelos não lineares de família exponencial, embora estudos de simulação sobre a distribuição empírica do mesmo não sejam conhecidos.

3.3 Influência

Um tópico de grande importância na análise de diagnóstico é a detecção de observações influentes, isto é, pontos que exercem um peso desproporcional nas estimativas dos parâmetros ou até mesmo na significância dos parâmetros. As técnicas mais conhecidas para detectar esse tipo de influência são baseadas na eliminação de um único ponto, as quais procuram avaliar o impacto da retirada de uma observação particular nas estimativas da regressão. Durante a década de 70 surgiram várias propostas relacionadas com a influência das observações nas estimativas dos coeficientes do modelo normal linear.

Suponha ϕ conhecido ou fixo e que o logaritmo da função de verossimilhança seja expresso na forma

$$L_\delta(\beta) = \sum_{j=1}^n \delta_j L_j(\beta), \quad (3.2)$$

em que $L_j(\beta)$ denota o logaritmo da função de verossimilhança correspondente à j -ésima observação e δ_j é um tipo de perturbação, definido tal que $0 \leq \delta_j \leq 1$. Quando $\delta_j = 1, \forall_j$, significa que não há

perturbação no modelo e, quando $\delta_j = 0$, significa que a j -ésima observação foi excluída. Então, o processo iterativo para obter $\hat{\beta}_\delta$ fica dado por

$$(\mathbf{J}^{(m)\top} \mathbf{W}^{(m)} \Delta \mathbf{J}^{(m)}) \beta_\delta^{(m+1)} = \mathbf{J}^{(m)\top} \mathbf{W}^{(m)} \Delta \mathbf{y}^{*(m)}, \quad (3.3)$$

em que $m = 0, 1, \dots$, $\mathbf{y}^{*(m)} = \mathbf{J}^{(m)} \beta_\delta^{(m)} + \mathbf{D}^{(m)}(\mathbf{y} - \boldsymbol{\mu}^{(m)})$, Δ é uma matriz $(n \times n)$ diagonal de 1's com δ na i -ésima posição e $\hat{\beta}_\delta$ é a estimativa de máxima verossimilhança de β supondo (3.2).

Uma medida natural da influência da perturbação feita no logartimo da função de verossimilhança em β é a diferença entre $\hat{\beta} - \hat{\beta}_\delta$. Como o cálculo dessa diferença, com δ variando para todas as observações, requer $n + 1$ repetições do processo iterativo, o que para n grande por ser trabalhoso computacionalmente, e como a taxa de convergência do processo de Newton-Raphson é pelo menos quadrática quando o mesmo é iniciado num valor consistente, a solução encontrada para a estimação da diferença acima foi obter a estimativa de primeiro passo $\hat{\beta}_\delta^1$ em (3.2) fazendo $\beta_\delta^{(0)} = \hat{\beta}$. Dessa forma, podemos expressar $\hat{\beta}_\delta^1$ como

$$\hat{\beta}_\delta^1 = \hat{\beta} - \frac{\hat{r}_{pi} \sqrt{\hat{w}_i \phi^{-1}} (1 - \delta)}{1 - (1 - \delta) \hat{h}_{ii}} (\hat{\mathbf{J}}^\top \widehat{\mathbf{W}} \hat{\mathbf{J}})^{-1} \hat{\mathbf{j}}_i, \quad (3.4)$$

sendo que as quantidades do lado direito de (3.4) são computadas em $\hat{\beta}$. A demonstração de (3.4) pode ser encontrada em Paula (1988). Em particular, para $\delta = 0$, o que significa que o i -ésimo ponto foi excluído, (3.4) fica expresso na forma simplificada

$$\hat{\beta}_{(i)}^1 = \hat{\beta} - \frac{\hat{r}_{pi} \sqrt{\hat{w}_i \phi^{-1}}}{1 - \hat{h}_{ii}} (\hat{\mathbf{J}}^\top \widehat{\mathbf{W}} \hat{\mathbf{J}})^{-1} \hat{\mathbf{j}}_i. \quad (3.5)$$

Quando tivermos pequenas amostras, a aproximação (3.5) poderá ficar muito afastada de $\hat{\beta}_{(i)}$ se $L(\beta)$ não for localmente quadrática, vide, por exemplo, (Cook & Weisberg, 1982). Entretanto, diversos estudos mostraram que essa aproximação é suficiente para destacar os pontos mais influentes, vide, por exemplo, (Pregibon (1981) e Storer & Crowley (1985)).

Uma outra medida de influência, conhecida como distância de Cook, foi proposta por Cook

(1977) para modelos normais lineares e rapidamente assimilada e estendida para diversas classes de modelos, sendo definida na classe de modelos não lineares de família exponencial por

$$D_i = (\hat{\beta} - \hat{\beta}_{(i)})^\top (\phi \hat{\mathbf{J}}^\top \widehat{\mathbf{W}} \hat{\mathbf{J}}) (\hat{\beta} - \hat{\beta}_{(i)}). \quad (3.6)$$

Quando substituimos $\hat{\beta}_{(i)}$ pela estimativa de primeiro passo $\hat{\beta}_{(i)}^1$ em (3.6), temos que

$$D_i^1 = \left\{ \frac{\hat{h}_{ii}}{1 - \hat{h}_{ii}} \right\} t_{Si}^2.$$

O gráfico de D_i^1 tem sido útil para avaliar a influência das observações nas estimativas dos parâmetros.

Finalmente, uma medida de influência que utiliza o logaritmo da função de verossimilhança em $\hat{\beta}$ e $\hat{\beta}_{(i)}$, respectivamente, a qual é denominada afastamento da verossimilhança (*likelihood displacement*), vide, por exemplo, (Cook & Weisberg, 1982), é definida por

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

É usual utilizar a aproximação de um passo por série de Taylor em torno de $\hat{\beta}$ pois não é possível encontrar uma forma analítica para LD_i . Então, LD_i fica expressa na forma

$$LD_i^1 \cong (\beta - \hat{\beta})^\top \{-\ddot{\mathbf{L}}(\hat{\beta})\}(\beta - \hat{\beta}).$$

Se substituirmos $-\ddot{\mathbf{L}}(\hat{\beta})$ pelo correspondente valor esperado, e β por $\hat{\beta}_{(i)}^1$, LD_i fica reexpressa na forma

$$LD_i^1 \cong (\hat{\beta} - \hat{\beta}_{(i)}^1)^\top (\phi \hat{\mathbf{J}}^\top \widehat{\mathbf{W}} \hat{\mathbf{J}}) (\hat{\beta} - \hat{\beta}_{(i)}^1),$$

que coincide com D_i .

O afastamento da verossimilhança pode ser baseado na região assintótica de confiança com

nível de confiança $(1 - \alpha)$ para β , vide, por exemplo, (Cox & Hinkley, 1974)

$$[\beta; 2\{L(\hat{\beta}) - L(\beta)\} \leq \chi_p^2(1 - \alpha)],$$

em que $\chi_{p;(1-\alpha)}^2$ é o percentil $(1 - \alpha)$ da distribuição qui-quadrado com p graus de liberdade. É interessante notar que como LD_i^1 coincide com D_i , então, nos casos que $L(\beta)$ for aproximadamente quadrática, D_i^1 poderá também ter seus valores comparados aos níveis nominais da χ_p^2 .

3.4 Influência Local

Uma das idéias mais inovadoras na área de diagnóstico em modelos de regressão foi apresentada por Cook (1986), que consiste em avaliar o comportamento de alguma medida de influência segundo pequenas perturbações (mudanças) no modelo ou nos dados, ao invés da retirada individual ou conjunta de pontos. Segundo Billor & Loynes (1993), perturbação é qualquer arranjo nas suposições do modelo ou nos dados para constatar alguma mudança substancial que ocorre nos resultados da análise. Essa metodologia, denominada influência local, teve grande receptividade entre os usuários e pesquisadores de regressão, havendo inúmeras publicações no assunto. Por exemplo, Beckman et al. (1987) investigaram a aplicação do método para modelos mistos não lineares, Thomas & Cook (1990) discutiram influência local em modelos lineares generalizados, Lee & Wang (1996) usaram o método para fazer análises de sensibilidade em modelos de equações estruturais, Paula (1996) utilizou influência local em modelos próprios de dispersão, Wei (1998) desenvolveu a aplicação da aproximação para modelos não lineares de família exponencial, Tang et al. (2001) estudaram influência local em modelos de dispersão reprodutivo não linear, entre outros autores. Além disso, outros autores estudaram também a relação entre a curvatura normal de Cook (1986) e outras estatísticas; vide, por exemplo, (Saint Laurent & Cook (1992) e Schall & Dunne (1992)).

A fim de introduzir a metodologia, suponha que $L(\theta)$ denota o logaritmo da função de verossimilhança do modelo, sendo $\theta = (\beta^\top, \phi)^\top$. Seja δ um vetor $n \times 1$ de perturbações, restrito a algum subconjunto aberto $\Omega \subset \mathbb{R}^n$. As perturbações são aplicadas no logaritmo da função de verossimilhança. Assumiremos em particular, o esquema de perturbação de casos em que o logaritmo da

função de verossimilhança toma a forma

$$L(\boldsymbol{\theta} \mid \boldsymbol{\delta}) = \sum_{i=1}^n \delta_i \log f(y_i; \boldsymbol{\theta}),$$

em que $0 \leq \delta_i \leq 1$ é um tipo de perturbação e $\boldsymbol{\delta}_0 = (1, \dots, 1)^\top$ é o vetor de não perturbação. Assumiremos existentes as derivadas até segunda ordem de $L(\boldsymbol{\theta} \mid \boldsymbol{\delta})$. Note que $L(\boldsymbol{\theta} \mid \boldsymbol{\delta}_0) = L(\boldsymbol{\theta})$. Para calcular a influência da perturbação $\boldsymbol{\delta}$ na estimativa de máxima verossimilhança $\hat{\boldsymbol{\theta}}$, o afastamento da verossimilhança é geralmente preferível (Pregibon, 1981); (Cook & Weisberg, 1982). Segundo Cook (1986) e Escobar & Meeker (1992), consideramos

$$LD(\boldsymbol{\delta}) = 2\{L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}_\delta)\},$$

em que $\hat{\boldsymbol{\theta}}_\delta$ denota a estimativa de máxima verossimilhança de $\boldsymbol{\theta}$ sob o modelo $L(\boldsymbol{\theta} \mid \boldsymbol{\delta})$. Com a definição acima, tem-se que $LD(\boldsymbol{\delta}) \geq 0$.

A idéia de influência local foi introduzida por Cook (1986) e consiste em investigar o comportamento da função de afastamento da verossimilhança $LD(\boldsymbol{\delta})$ em uma vizinhança de $\boldsymbol{\delta}_0$, que é o ponto onde as duas funções de (log)-verossimilhança coincidem, isto é, $L(\boldsymbol{\theta}) = L(\boldsymbol{\theta} \mid \boldsymbol{\delta})$. O procedimento consiste em selecionar uma direção unitária $\boldsymbol{\ell}$, $\|\boldsymbol{\ell}\| = 1$, e considerar o gráfico de $LD(\boldsymbol{\delta}_0 + a\boldsymbol{\ell})$ contra a , em que $a \in \mathbb{R}$. Esse gráfico é denominado linha projetada. Como $L(\boldsymbol{\delta}_0) = 0$, o gráfico de $LD(\boldsymbol{\delta}_0 + a\boldsymbol{\ell})$ tem um mínimo local em $a = 0$. A curvatura normal $C_{\boldsymbol{\ell}}(\boldsymbol{\theta})$ é uma caracterização de $LD(\boldsymbol{\delta}_0 + a\boldsymbol{\ell})$ em torno de $a = 0$. Essa curvatura é interpretada como o inverso do raio do melhor círculo ajustado em $a = 0$. Uma sugestão é considerar a direção $\boldsymbol{\ell}_{\max}$ que corresponde à maior curvatura $C_{\boldsymbol{\ell}_{\max}}(\boldsymbol{\theta})$. O gráfico de $|\boldsymbol{\ell}_{\max}|$ contra a ordem das observações pode revelar aqueles casos que sob pequenas perturbações exercem notável influência em $LD(\boldsymbol{\delta})$. Tais casos podem ser responsáveis por mudanças substanciais nas estimativas dos parâmetros. Então, seria prudente olhar com mais cuidado esses casos a fim de entender melhor a influência dos mesmos.

Cook (1986) mostra que a curvatura normal na direção ℓ toma a forma

$$C_{\ell}(\theta) = 2 \mid \ell^{\top} \Delta^{\top} \ddot{L}_{\hat{\theta}\hat{\theta}}^{-1} \Delta \ell \mid,$$

em que $-\ddot{L}_{\hat{\theta}\hat{\theta}}$ é a matriz de informação de Fisher de θ e Δ é uma matriz $(p+1) \times n$ com elementos dados por $\Delta_{ji} = \partial^2 L(\theta \mid \delta) / \partial \theta_i \partial \delta_j$, avaliados em $\theta = \hat{\theta}$ e $\delta = \delta_0$, $i = 1, \dots, n$ e $j = 1, \dots, p, p+1$. É interessante notar que o máximo de $\ell^{\top} B \ell$, em que $B = \Delta^{\top} (-\ddot{L}_{\hat{\theta}\hat{\theta}}^{-1}) \Delta$, corresponde ao maior autovalor (em valor absoluto) da matriz B . Então, $C_{\ell_{\max}}$ é o maior autovalor da matriz B e $\ell_{\max}$ é o correspondente autovetor. Dessa maneira, o gráfico de $\mid \ell_{\max} \mid$ contra a ordem das observações para a matriz $\Delta^{\top} (-\ddot{L}_{\hat{\theta}\hat{\theta}}^{-1}) \Delta$, pode revelar aqueles pontos com maior influência na vizinhança de $LD(\delta_0)$. Tais pontos podem causar mudanças substanciais nas estimativas dos parâmetros sob pequenas perturbações no modelo. Deve-se olhar também os componentes do segundo autovetor quando $C_{\ell_{\max}}$ não for muito maior do que o segundo autovalor, pois é provável que nesse caso, o segundo autovetor destaque algum tipo de influência particular das observações nas estimativas.

Lesaffre & Verbeke (1998) sugerem um gráfico alternativo que considera também a curvatura na direção da i -ésima observação que corresponde em computar a curvatura avaliada no vetor ℓ_i de dimensão $n \times 1$ formado por zeros com um na i -ésima posição. Nesse caso, deve-se padronizar $C_i = C_i / \sum_{j=1}^n C_j$. Uma sugestão é prestar mais atenção naqueles pontos tais que $C_i > 2\bar{C}$, em que $\bar{C} = \sum_{j=1}^n C_j / n$.

No entanto, se o interesse está somente em obter a influência no vetor $\hat{\beta}$, a curvatura normal na direção ℓ é dada por $C_{\ell}(\beta) = 2 \mid \ell^{\top} \Delta^{\top} (\ddot{L}_{\hat{\theta}\hat{\theta}}^{-1} - B_1) \Delta \ell \mid$, sendo

$$B_1 = \begin{pmatrix} 0 & 0 \\ 0 & \ddot{L}_{\hat{\phi}\hat{\phi}}^{-1} \end{pmatrix},$$

com $-\ddot{L}_{\hat{\phi}\hat{\phi}}^{-1}$ denotando a informação de Fisher observada de ϕ avaliado em $\hat{\theta}$. O gráfico do maior autovetor de $\Delta^{\top} (\ddot{L}_{\hat{\theta}\hat{\theta}}^{-1} - B_1) \Delta$ contra a ordem das observações pode mostrar aquelas observações mais influentes em $\hat{\beta}$. Da mesma forma, a curvatura normal para o parâmetro de dispersão ϕ na

direção ℓ é dada por $C_\ell(\phi) = 2 \mid \ell^\top \Delta^\top (\ddot{L}_{\hat{\theta}\hat{\theta}}^{-1} - B_2) \Delta \ell \mid$, com

$$B_2 = \begin{pmatrix} \ddot{L}_{\hat{\beta}\hat{\beta}}^{-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix},$$

em que $-\ddot{L}_{\hat{\beta}\hat{\beta}}^{-1}$ é a informação de Fisher observada de β avaliada em $\hat{\theta}$. A influência local das observações em $\hat{\phi}$ pode ser obtida considerando o gráfico de $\ell_{\max}$ contra a ordem das observações para a matriz $\Delta^\top (\ddot{L}_{\hat{\theta}\hat{\theta}}^{-1} - B_2) \Delta$.

Vamos obter a seguir a matriz Δ para alguns esquemas de perturbação usuais. A matriz Δ será denotada por $\Delta = (\Delta_\beta^\top, \Delta_\phi^\top)^\top$ em que $\Delta_\beta = \partial^2 L(\theta \mid \delta) / \partial \beta \partial \delta$ e $\Delta_\phi = \partial^2 L(\theta \mid \delta) / \partial \phi \partial \delta$ ambas avaliadas em δ_0 e $\hat{\theta}$.

Ponderação de Casos

Para esta situação o logaritmo da função de verossimilhança perturbada será expresso na forma

$$L(\theta \mid \delta) = \sum_{i=1}^n \delta_i L_i(\theta),$$

em que $0 \leq \delta_i \leq 1$ e $\delta_0 = (1, \dots, 1)^\top$ é o vetor de não perturbação. Note que $L(\theta \mid \delta_0) = L(\theta)$. Considere $L(\theta)$ como descrito na equação (2.4). Então, substituindo na equação acima temos que

$$L(\theta \mid \delta) = \sum_{i=1}^n \delta_i \phi \{y_i \theta_i - b(\theta_i)\} + \sum_{i=1}^n \delta_i c(y_i, \phi).$$

Para obtermos a matriz Δ calculamos inicialmente a primeira e segunda derivadas de $L(\theta \mid \delta)$ em relação a β_i e δ_i respectivamente. Daí segue que

$$\begin{aligned}
\frac{\partial L(\boldsymbol{\theta} \mid \boldsymbol{\delta})}{\partial \beta_j} &= \sum_{i=1}^n \delta_i \phi \left\{ y_i \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} - \frac{db(\theta_i)}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} \right\} \\
&= \sum_{i=1}^n \delta_i \phi \left\{ y_i V_i^{-1} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} - \mu_i V_i^{-1} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} \right\} \\
&= \sum_{i=1}^n \delta_i \phi \left\{ y_i \frac{1}{V_i} \frac{d\mu_i}{d\eta_i} \left(\frac{d\mu_i/d\eta_i}{d\mu_i/d\eta_i} \right) \frac{\partial \eta_i}{\partial \beta_j} - \mu_i \frac{1}{V_i} \frac{d\mu_i}{d\eta_i} \left(\frac{d\mu_i/d\eta_i}{d\mu_i/d\eta_i} \right) \frac{\partial \eta_i}{\partial \beta_j} \right\} \\
&= \sum_{i=1}^n \delta_i \phi \left\{ y_i \omega_i d_i \frac{\partial \eta_i}{\partial \beta_j} - \mu_i \omega_i d_i \frac{\partial \eta_i}{\partial \beta_j} \right\} \\
&= \sum_{i=1}^n \delta_i \phi \left\{ \omega_i d_i \frac{\partial \eta_i}{\partial \beta_j} (y_i - \mu_i) \right\},
\end{aligned}$$

em que $\omega_i = (d\mu_i/d\eta_i)^2/V_i$ e $d_i = d\eta_i/d\mu_i$. Agora derivando o resultado acima em relação δ_i temos

$$\left. \frac{\partial^2 L(\boldsymbol{\theta} \mid \boldsymbol{\delta})}{\partial \beta_j \partial \delta_i} \right|_{\boldsymbol{\delta}_0} = \phi \omega_i d_i j_{ij} (y_i - \mu_i).$$

Dessa forma, $\boldsymbol{\Delta}_{\boldsymbol{\beta}} = \{\Delta_{ji}\}$, pode ser escrito como

$$\Delta_{ji} = \phi \omega_i d_i j_{ij} (y_i - \mu_i),$$

que deve ser avaliado em $\hat{\boldsymbol{\theta}}$, em que $i = 1, \dots, n$ e $j = 1, \dots, p$. Assim, temos que $\Delta_{ji} = \hat{\phi} \hat{\omega}_i \hat{d}_i \hat{j}_{ij} (y_i - \hat{\mu}_i)$. Portanto $\boldsymbol{\Delta}_{\boldsymbol{\beta}}$ é uma matriz $p \times n$ de elementos Δ_{ji} .

Calculando agora a primeira e a segunda derivadas de $L(\boldsymbol{\theta} \mid \boldsymbol{\delta})$ em relação a ϕ e δ_i respectivamente, temos

$$\begin{aligned}
\frac{\partial L(\boldsymbol{\theta} \mid \boldsymbol{\delta})}{\partial \phi} &= \sum_{i=1}^n \delta_i \{y_i \theta_i - b(\theta_i)\} + \sum_{i=1}^n \delta_i c'(y_i, \phi), \\
\left. \frac{\partial^2 L(\boldsymbol{\theta} \mid \boldsymbol{\delta})}{\partial \phi \partial \delta_i} \right|_{\boldsymbol{\delta}_0} &= y_i \theta_i - b(\theta_i) + c'(y_i, \phi).
\end{aligned}$$

Dessa forma, $\Delta_{\phi i} = y_i \theta_i - b(\theta_i) + c'(y_i, \phi)$ que deve ser avaliado em $\hat{\theta}$. Aqui temos que $\Delta_{\phi i} = y_i \hat{\theta}_i - b(\hat{\theta}_i) + c'(y_i, \hat{\phi})$. Portanto $\Delta_{\phi} = \{\Delta_{\phi i}\}$ é um vetor $1 \times n$ de elementos $\Delta_{\phi i}$. Assim, $\Delta = (\Delta_{\beta}^{\top}, \Delta_{\phi}^{\top})^{\top}$ é uma matriz $(p+1) \times n$.

Perturbação na Resposta

Vamos supor y_i uma resposta contínua de um modelo não linear de família exponencial e a seguinte perturbação:

$$y_{i\delta} = y_i + \delta_i,$$

em que δ_i é um valor arbitrário real. Aqui o vetor de não perturbação é dado por $\delta_0 = (0, \dots, 0)^{\top}$. Então o logaritmo da função de verossimilhança é dado por

$$L(\theta \mid \delta) = \sum_{i=1}^n \phi\{(y_i + \delta_i)\theta_i - b(\theta_i)\} + \sum_{i=1}^n c((y_i + \delta_i), \phi).$$

Calculando inicialmente a primeira e segunda derivadas de $L(\theta \mid \delta)$ em relação a β_j e δ_i , respectivamente, temos que

$$\begin{aligned} \frac{\partial L(\theta \mid \delta)}{\partial \beta_j} &= \sum_{i=1}^n \phi \left\{ (y_i + \delta_i) \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} - \frac{db(\theta_i)}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} \right\} \\ &= \sum_{i=1}^n \phi \left\{ (y_i + \delta_i) V_i^{-1} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} - \mu_i V_i^{-1} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} \right\} \\ &= \sum_{i=1}^n \phi \left\{ (y_i + \delta_i) \frac{1}{V_i} \frac{d\mu_i}{d\eta_i} \left(\frac{d\mu_i/d\eta_i}{d\mu_i/d\eta_i} \right) \frac{\partial \eta_i}{\partial \beta_j} - \mu_i \frac{1}{V_i} \frac{d\mu_i}{d\eta_i} \left(\frac{d\mu_i/d\eta_i}{d\mu_i/d\eta_i} \right) \frac{\partial \eta_i}{\partial \beta_j} \right\} \\ &= \sum_{i=1}^n \phi \left\{ (y_i + \delta_i) \omega_i d_i \frac{\partial \eta_i}{\partial \beta_j} - \mu_i \omega_i d_i \frac{\partial \eta_i}{\partial \beta_j} \right\} \\ &= \sum_{i=1}^n \phi \left\{ \omega_i d_i \frac{\partial \eta_i}{\partial \beta_j} (y_i + \delta_i - \mu_i) \right\}, \end{aligned}$$

em que $\omega_i = (d\mu_i/d\eta_i)^2/V_i$ e $d_i = d\eta_i/d\mu_i$. Agora derivando o resultado acima em relação δ_i temos

$$\left. \frac{\partial^2 L(\boldsymbol{\theta} \mid \boldsymbol{\delta})}{\partial \beta_j \partial \delta_i} \right|_{\boldsymbol{\delta}_0} = \phi \omega_i d_i j_{ij}$$

Dessa forma, $\boldsymbol{\Delta}_\beta = \{\Delta_{ji}\}$, pode ser escrito como

$$\Delta_{ji} = \phi \omega_i d_i j_{ij},$$

que deve ser avaliado em $\hat{\boldsymbol{\theta}}$, em que $i = 1, \dots, n$ e $j = 1, \dots, p$. Assim, temos que $\Delta_{ji} = \hat{\phi} \hat{\omega}_i \hat{d}_i \hat{j}_{ij}$. Portanto $\boldsymbol{\Delta}_\beta$ é uma matriz $p \times n$ de elementos Δ_{ji} .

Calculando agora a primeira e a segunda derivadas de $L(\boldsymbol{\theta} \mid \boldsymbol{\delta})$ em relação a ϕ e δ_i respectivamente, temos

$$\begin{aligned} \frac{\partial L(\boldsymbol{\theta} \mid \boldsymbol{\delta})}{\partial \phi} &= \sum_{i=1}^n \{(y_i + \delta_i)\theta_i - b(\theta_i)\} + \sum_{i=1}^n c'(y_i + \delta_i, \phi) \\ \frac{\partial^2 L(\boldsymbol{\theta} \mid \boldsymbol{\delta})}{\partial \phi \partial \delta_i} &= \theta_i + c''(y_i + \delta_i, \phi). \end{aligned}$$

Dessa forma, $\Delta_{\phi i} = \hat{\theta}_i + c''(y_i, \hat{\phi})$ já calculado em $\boldsymbol{\delta}_0$ e $\hat{\boldsymbol{\theta}}$. Portanto $\boldsymbol{\Delta}_\phi = \{\Delta_{\phi i}\}$ é um vetor $1 \times n$ de elementos $\Delta_{\phi i}$ e $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\beta^\top, \boldsymbol{\Delta}_\phi^\top)^\top$ é uma matriz $(p+1) \times n$.

Perturbação na Variável Explicativa

Vamos supor $\mathbf{x}_i^\top$ é um vetor de valores fixados conhecidos de variáveis explicativas de um modelo não linear de família exponencial e estamos perturbando a m -ésima variável explicativa, com a seguinte perturbação:

$$x_{im\delta} = x_{im} + \delta_i,$$

em que δ_i é um valor arbitrário real. Aqui o vetor de não perturbação é dado por $\boldsymbol{\delta}_0 = (0, \dots, 0)^\top$. Então, o logaritmo da função de verossimilhança é dado por

$$L(\boldsymbol{\theta} \mid \boldsymbol{\delta}) = \sum_{i=1}^n \phi\{y_i \theta_{i\delta} - b(\theta_{i\delta})\} + \sum_{i=1}^n c(y_i, \phi).$$

Calculando inicialmente a primeira e segunda derivadas de $L(\boldsymbol{\theta} \mid \boldsymbol{\delta})$ em relação a β_j e δ_i , respectivamente, temos que

$$\begin{aligned} \frac{\partial L(\boldsymbol{\theta} \mid \boldsymbol{\delta})}{\partial \beta_j} &= \sum_{i=1}^n \phi \left\{ y_i \frac{d\theta_{i\delta}}{d\mu_{i\delta}} \frac{d\mu_{i\delta}}{d\eta_{i\delta}} \frac{\partial \eta_{i\delta}}{\partial \beta_j} - \frac{db(\theta_{i\delta})}{d\theta_{i\delta}} \frac{d\theta_{i\delta}}{d\mu_{i\delta}} \frac{d\mu_{i\delta}}{d\eta_{i\delta}} \frac{\partial \eta_{i\delta}}{\partial \beta_j} \right\} \\ &= \sum_{i=1}^n \phi \left\{ y_i V_{i\delta}^{-1} \frac{d\mu_{i\delta}}{d\eta_{i\delta}} \frac{\partial \eta_{i\delta}}{\partial \beta_j} - \mu_{i\delta} V_{i\delta}^{-1} \frac{d\mu_{i\delta}}{d\eta_{i\delta}} \frac{\partial \eta_{i\delta}}{\partial \beta_j} \right\} \\ &= \sum_{i=1}^n \phi \left\{ y_i \frac{1}{V_{i\delta}} \frac{d\mu_{i\delta}}{d\eta_{i\delta}} \left(\frac{d\mu_{i\delta}/d\eta_{i\delta}}{d\mu_{i\delta}/d\eta_{i\delta}} \right) \frac{\partial \eta_{i\delta}}{\partial \beta_j} - \mu_{i\delta} \frac{1}{V_{i\delta}} \frac{d\mu_{i\delta}}{d\eta_{i\delta}} \left(\frac{d\mu_{i\delta}/d\eta_{i\delta}}{d\mu_{i\delta}/d\eta_{i\delta}} \right) \frac{\partial \eta_{i\delta}}{\partial \beta_j} \right\} \\ &= \sum_{i=1}^n \phi \left\{ y_i \omega_{i\delta} d_{i\delta} \frac{\partial \eta_{i\delta}}{\partial \beta_j} - \mu_{i\delta} \omega_{i\delta} d_{i\delta} \frac{\partial \eta_{i\delta}}{\partial \beta_j} \right\} \\ &= \sum_{i=1}^n \phi \left\{ \omega_{i\delta} d_{i\delta} \frac{\partial \eta_{i\delta}}{\partial \beta_j} (y_i - \mu_{i\delta}) \right\}, \end{aligned}$$

em que $\omega_{i\delta} = (d\mu_{i\delta}/d\eta_{i\delta})^2/V_{i\delta}$ e $d_{i\delta} = d\eta_{i\delta}/d\mu_{i\delta}$. Agora derivando o resultado acima em relação δ_i temos

$$\begin{aligned} \frac{\partial^2 L(\boldsymbol{\theta} \mid \boldsymbol{\delta})}{\partial \beta_j \partial \delta_i} &= \phi \left\{ \left(\frac{d\omega_{i\delta}}{d\delta_i} \right) d_{i\delta} \frac{\partial \eta_{i\delta}}{\partial \beta_j} (y_i - \mu_{i\delta}) + \omega_{i\delta} \frac{d}{d\delta_i} \left(d_{i\delta} \frac{\partial \eta_{i\delta}}{\partial \beta_j} (y_i - \mu_{i\delta}) \right) \right\} \\ &= \phi \left\{ \left(\frac{d\omega_{i\delta}}{d\delta_i} \right) d_{i\delta} \frac{\partial \eta_{i\delta}}{\partial \beta_j} (y_i - \mu_{i\delta}) + \omega_{i\delta} \frac{dd_{i\delta}}{d\delta_i} \left(\frac{\partial \eta_{i\delta}}{\partial \beta_j} \right) (y_i - \mu_{i\delta}) + d_{i\delta} \frac{d}{d\delta_i} \left(\frac{\partial \eta_{i\delta}}{\partial \beta_j} (y_i - \mu_{i\delta}) \right) \right\} \\ &= \phi \left\{ \left(\frac{d\omega_{i\delta}}{d\delta_i} \right) d_{i\delta} \frac{\partial \eta_{i\delta}}{\partial \beta_j} (y_i - \mu_{i\delta}) \right\} + \phi \left\{ \omega_{i\delta} \left(\frac{d}{d\delta_i} d_{i\delta} \right) \left(\frac{\partial \eta_{i\delta}}{\partial \beta_j} \right) (y_i - \mu_{i\delta}) \right\} \\ &+ \phi \left\{ \omega_{i\delta} d_{i\delta} \left(\frac{d}{d\delta_i} \left(\frac{\partial \eta_{i\delta}}{\partial \beta_j} \right) (y_i - \mu_{i\delta}) \right) \right\} - \phi \left\{ \omega_{i\delta} d_{i\delta} \left(\frac{\partial \eta_{i\delta}}{\partial \beta_j} \right) \left(\frac{d}{d\delta_i} \mu_{i\delta} \right) \right\}. \end{aligned}$$

Dessa forma, $\mathbf{\Delta}_{\beta} = \{\Delta_{ji}\}$, pode ser escrito como

$$\begin{aligned}\Delta_{ji} &= \phi \left\{ \left(\frac{d\omega_{i\delta}}{d\delta_i} \right) d_{i\delta} \frac{\partial \eta_{i\delta}}{\partial \beta_j} (y_i - \mu_{i\delta}) \right\} + \phi \left\{ \omega_{i\delta} \left(\frac{d}{d\delta_i} d_{i\delta} \right) \left(\frac{\partial \eta_{i\delta}}{\partial \beta_j} \right) (y_i - \mu_{i\delta}) \right\} \\ &+ \phi \left\{ \omega_{i\delta} d_{i\delta} \left(\frac{d}{d\delta_i} \left(\frac{\partial \eta_{i\delta}}{\partial \beta_j} \right) (y_i - \mu_{i\delta}) \right) \right\} - \phi \left\{ \omega_{i\delta} d_{i\delta} \left(\frac{\partial \eta_{i\delta}}{\partial \beta_j} \right) \left(\frac{d}{d\delta_i} \mu_{i\delta} \right) \right\},\end{aligned}$$

que deve ser avaliado em δ_0 e $\hat{\theta}$, em que $i = 1, \dots, n$ e $j = 1, \dots, p$. Portanto $\mathbf{\Delta}_{\beta}$ é uma matriz $p \times n$ de elementos Δ_{ji} .

Calculando agora a primeira e a segunda derivadas de $L(\theta \mid \delta)$ em relação a ϕ e δ_i respectivamente, temos

$$\frac{\partial L(\theta \mid \delta)}{\partial \phi} = \sum_{i=1}^n \{y_i \theta_{i\delta} - b(\theta_{i\delta})\} + \sum_{i=1}^n c'(y_i, \phi),$$

$$\begin{aligned}\frac{\partial^2 L(\theta \mid \delta)}{\partial \phi \partial \delta_i} &= y_i \frac{d\theta_{i\delta}}{d\mu_{i\delta}} \frac{d\mu_{i\delta}}{d\eta_{i\delta}} \frac{\partial \eta_{i\delta}}{\partial \beta_j} - \frac{db(\theta_{i\delta})}{d\theta_{i\delta}} \frac{d\theta_{i\delta}}{d\mu_{i\delta}} \frac{d\mu_{i\delta}}{d\eta_{i\delta}} \frac{\partial \eta_{i\delta}}{\partial \beta_j} \\ &= y_i V_{i\delta}^{-1} \frac{d\mu_{i\delta}}{d\eta_{i\delta}} \frac{\partial \eta_{i\delta}}{\partial \beta_j} - \mu_{i\delta} V_{i\delta}^{-1} \frac{d\mu_{i\delta}}{d\eta_{i\delta}} \frac{\partial \eta_{i\delta}}{\partial \beta_j} \\ &= y_i \frac{1}{V_{i\delta}} \frac{d\mu_{i\delta}}{d\eta_{i\delta}} \left(\frac{d\mu_{i\delta}/d\eta_{i\delta}}{d\mu_{i\delta}/d\eta_{i\delta}} \right) \frac{\partial \eta_{i\delta}}{\partial \beta_j} - \mu_{i\delta} \frac{1}{V_{i\delta}} \frac{d\mu_{i\delta}}{d\eta_{i\delta}} \left(\frac{d\mu_{i\delta}/d\eta_{i\delta}}{d\mu_{i\delta}/d\eta_{i\delta}} \right) \frac{\partial \eta_{i\delta}}{\partial \beta_j} \\ &= y_i \omega_{i\delta} d_{i\delta} \frac{\partial \eta_{i\delta}}{\partial \beta_j} - \mu_{i\delta} \omega_{i\delta} d_{i\delta} \frac{\partial \eta_{i\delta}}{\partial \beta_j} \\ &= \omega_{i\delta} d_{i\delta} \frac{\partial \eta_{i\delta}}{\partial \beta_j} (y_i - \mu_{i\delta}),\end{aligned}$$

em que $\omega_{i\delta} = (d\mu_{i\delta}/d\eta_{i\delta})^2/V_{i\delta}$ e $d_{i\delta} = d\eta_{i\delta}/d\mu_{i\delta}$.

Dessa forma, $\Delta_{\phi i} = \omega_{i\delta} d_{i\delta} (\partial \eta_{i\delta} / \partial \beta_j) (y_i - \mu_{i\delta})$ que deve ser avaliado em δ_0 e $\hat{\theta}$. Em particular temos $\Delta_{\phi i} = \hat{\omega}_i \hat{d}_i \hat{j}_{ij} (y_i - \hat{\mu}_i)$. Portanto $\mathbf{\Delta}_{\phi} = \{\Delta_{\phi i}\}$ é um vetor $1 \times n$ de elementos $\Delta_{\phi i}$ e $\mathbf{\Delta} = (\mathbf{\Delta}_{\beta}^{\top}, \mathbf{\Delta}_{\phi}^{\top})^{\top}$ é uma matriz $(p+1) \times n$.

3.5 Métodos de Diagnóstico na Binomial Negativa

Pontos de Alavanca

Nos modelos não lineares de família exponencial com resposta binomial negativa a matriz de projeção ortogonal é a mesma dada na equação (3.1) com

$$\begin{aligned} h_{ii} &= w\mathbf{j}_i^\top (\mathbf{J}^\top \mathbf{W} \mathbf{J})^{-1} \mathbf{j}_i \\ &= \frac{(d\mu_i/d\eta_i)^2}{(\mu_i^2 \nu^{-1} + \mu_i)} \mathbf{j}_i^\top (\mathbf{J}^\top \mathbf{W} \mathbf{J})^{-1} \mathbf{j}_i, \end{aligned}$$

em que $\mathbf{j}_i^\top$ denota a i -ésima linha de $\mathbf{J}$.

Resíduo

O resíduo componente do desvio, assumindo ν conhecido ou fixo fica expresso na forma

$$d(y_i, \hat{\mu}_i, \nu) = \text{sign}(y_i - \hat{\mu}_i) \sqrt{2} \left[\nu \log \left\{ \frac{\hat{\mu}_i + \nu}{y_i + \nu} \right\} + y_i \log \left\{ \frac{y_i(\hat{\mu}_i + \nu)}{\hat{\mu}_i(y_i + \nu)} \right\} \right]^{1/2},$$

para $i = 1, \dots, n$.

Svetliza (2002) mostra que aproximadamente $E\{d^*(Y_i; \mu_i, \nu)\} = 0$ e $\text{Var}\{d^*(Y_i; \mu_i, \nu)\} = 1 - h_{ii}$ para os modelos não lineares com resposta binomial negativa. Então, podemos escrever o resíduo componente do desvio padronizado como

$$d^*(y_i, \hat{\mu}_i, \nu) = \frac{\pm d(y_i, \hat{\mu}_i, \nu)}{\sqrt{1 - \hat{h}_{ii}}}.$$

Influência Local

Assumindo em particular o esquema de perturbação de casos, a função logaritmo de verossimilhança assume a seguinte forma

$$L(\boldsymbol{\theta} \mid \boldsymbol{\delta}) = \sum_{i=1}^n \delta_i \left[\log \left\{ \frac{\Gamma(y_i + \nu)}{y_i! \Gamma(\nu)} \right\} + y_i \log \left\{ \frac{\mu_i}{\mu_i + \nu} \right\} + \nu \log \left\{ \frac{\nu}{\mu_i + \nu} \right\} \right],$$

em que $0 \leq \delta_i \leq 1$ e $\boldsymbol{\delta}_0 = (1, \dots, 1)^\top$ é o vetor de não perturbação.

A matriz $\boldsymbol{\Delta}_\beta = \{\Delta_{ji}\}$ fica dada por

$$\Delta_{ji} = \left. \frac{\partial^2 L(\boldsymbol{\theta} \mid \boldsymbol{\delta})}{\partial \beta_j \partial \delta_i} \right|_{\boldsymbol{\delta}_0} = \omega_i d_i j_{ij} (y_i - \mu_i),$$

que deve ser avaliado em $\widehat{\boldsymbol{\theta}}$, em que $\omega_i = (d\mu_i/d\eta_i)^2/(\mu_i + \mu_i^2\nu^{-1})$, $d_i = (d\eta_i/d\mu_i)$, com $i = 1, \dots, n$ e $j = 1, \dots, p$. Assim, temos que $\Delta_{ji} = \widehat{w}_i \widehat{d}_i \widehat{j}_{ij} (y_i - \widehat{\mu}_i)$. Já $\Delta_{\nu i}$ fica dado por

$$\Delta_{\nu i} = \Psi(\widehat{\nu} + y_i) - \Psi(\widehat{\nu}) - \frac{(\widehat{\nu} + y_i)}{(\widehat{\nu} + \widehat{\mu}_i)} + \log \widehat{\nu} + 1 - \log(\widehat{\nu} + \widehat{\mu}_i),$$

com $i = 1, \dots, n$. Portanto, temos que $\boldsymbol{\Delta}_\nu = \{\Delta_{\nu i}\}$ é um vetor de dimensão $1 \times n$ e $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\beta^\top, \boldsymbol{\Delta}_\nu^\top)^\top$ é uma matriz de dimensão $(p+1) \times n$.

Como a resposta Y_i não é contínua sob a distribuição binomial negativa não faz sentido aplicação de perturbação na variável resposta, contudo podemos considerar esquemas de perturbação em variáveis explicativas contínuas.

Capítulo 4

Aplicações

Neste Capítulo apresentamos cinco dos exemplos mostrados no Capítulo 1. Para encontrarmos as estimativas dos parâmetros utilizamos o algoritmo proposto no Capítulo 2 em que declaramos $\tau = f(\mathbf{J}; \beta) - \mathbf{J}\beta$ como sendo um *offset* e apresentamos o critério de informação de Akaike - AIC (Akaike, 1974). Além disso, calculamos o viés das estimativas utilizando a equação (2.8). Finalmente construímos os gráficos de diagnóstico e de envelope para cada exemplo. Toda essa parte computacional foi resolvida utilizando o software R e os comandos encontram-se no Apêndice A.

4.1 Coelhos Europeus

Os dados consistem num conjunto de 71 observações. A variável resposta, Y , representa o peso das lentes (em mg) dos olhos de coelhos europeus (*Oryctolagus Cuniculus*) na Austrália. Uma variável explicativa foi considerada, x , e corresponde a idade (em dias) dos coelhos. Vamos assumir o mesmo modelo proposto por Wei (1998), em que Y tem distribuição normal inversa, isto é, $Y_i \sim \text{NI}(\mu_i, \sigma^2)$, em que a parte sistemática é dada por

$$\mu_i = \beta_1 - \frac{\beta_2}{x_i + \beta_3}, \quad i = 1, \dots, 71.$$

As estimativas dos parâmetros, viés, erros padrão (assintóticos), estatística z e p -valores são apresentadas na Tabela 4.1.

Tabela 4.1: Estimativas de máxima verossimilhança com os respectivos erros padrão aproximados obtidos do modelo normal inversa ajustado aos dados sobre Coelhos Europeus.

Parâmetro	Estimativa	Viés	Erro Padrão	Estatística z	p -valor
β_1	5,63	0,0003	0,025	224,96	$< 0,0001$
β_2	128,53	0,1645	6,094	21,09	$< 0,0001$
β_3	36,78	0,0504	2,209	16,65	$< 0,0001$
AIC	-161,74				

Note que todos os parâmetros são individualmente significativos (p -valor $< 0,0001$) ao nível de significância de 5%. Portanto, têm-se indícios de que a relação funcional proposta tende a ser uma boa aproximação para explicar a média da variável resposta. Na Figura 4.1 apresentamos os dados observados juntamente com o modelo ajustado.

Com o objetivo de detectar a existência de observações aberrantes ou influentes com alguma interferência desproporcional nos resultados do ajuste, bem como verificar possíveis afastamentos das suposições feitas para o modelo e para as partes aleatória e sistemática apresentamos, na Figura 4.2 uma análise de diagnóstico e na Figura 4.3 apresentamos o gráfico de envelope. Note que os pontos atípicos estão todos situados no primeiro quartil dos dados, indicando uma maior sensibilidade dessas observações com relação ao ajuste do modelo.

Destacou-se na Figura 4.2(b) as observações 1, 4, 5, 16 e 17 como pontos influentes. Nota-se que as mesmas observações são pontos aberrantes (Figura 4.2(c)). Já observando a Figura 4.2(d) temos que a ligação parece ser adequada para este modelo, devido a aleatoriedade dos pontos dentro da faixa $(-2; 2)$. As descrições dessas observações são as seguintes: #1 corresponde ao coelho com peso das lentes dos olhos de 21,66 mg e com 15 dias de idade; #4 corresponde ao coelho com peso das lentes dos olhos de 31,25 mg e com 18 dias de idade; #5 corresponde ao coelho com peso das lentes dos olhos de 44,79 mg e com 28 dias de idade; #16 corresponde ao coelho com peso das lentes dos olhos de 65,31 mg e com 65 dias de idade; #17 corresponde ao coelho com peso das lentes dos olhos de 71,90 mg e com 72 dias de idade.

Observe na Figura 4.3 que alguns pontos estão fora da banda de confiança dando indícios de

Figura 4.1: Gráfico do modelo normal inversa ajustado aos dados sobre Coelhos Europeus.

que a distribuição postulada para o peso das lentes (em mg) dos olhos dos coelhos europeus talvez não esteja adequada.

Com o objetivo de analisar o impacto das observações detectadas na Figura 4.2 sobre as estimativas dos parâmetros, realizamos uma análise confirmatória. Para isso, consideramos a variação

Figura 4.2: Gráficos de diagnóstico referentes ao modelo normal inversa ajustado aos dados sobre Coelhos Europeus.

percentual das estimativas definidas por

$$VP = \frac{\hat{\beta}_j - \hat{\beta}_{j(m)}}{\hat{\beta}_j} \times 100,$$

em que $\hat{\beta}$ é a estimativa de β com todas as observações e $\hat{\beta}_{(m)}$ denota a estimativa de β retirando a

Figura 4.3: Gráfico normal de probabilidades (com envelope gerado) para o resíduo componente do desvio referente ao ajuste do modelo normal inversa aos dados sobre Coelhos Europeus.

m -ésima observação.

Primeiramente, reajustamos o modelo eliminando individualmente cada observação (1, 4, 5, 16 e 17). Em seguida, reajustamos o modelo eliminando conjuntamente as observações detectadas na análise de diagnóstico. As estimativas dos parâmetros dos modelos podem ser vistas na Tabela 4.2. Observe que todos os parâmetros continuam sendo individualmente significativos ao nível de significância de 1% mesmo excluindo todas as observações. A maior variação percentual (VP) ocorreu em β_3 quando excluímos todos os pontos conjuntamente. No entanto, não foi superior a 6%.

Tabela 4.2: Variação das estimativas do modelo normal inversa ao excluir as observações citadas na análise de diagnóstico.

	Estimativas			p-valor			VP (%)			AIC
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3	
Obs										
#1	5,64 (0,024)	131,79 (6,307)	38,49 (2,363)	< 0,0001	< 0,0001	< 0,0001	-0,144	-2,535	-4,648	-162,34
#4	5,62 (0,021)	126,50 (5,225)	35,36 (1,888)	< 0,0001	< 0,0001	< 0,0001	0,069	1,579	3,868	-174,81
#5	5,64 (0,023)	130,90 (5,795)	37,35 (2,073)	< 0,0001	< 0,0001	< 0,0001	-0,141	-1,843	-1,545	-168,38
#16	5,62 (0,023)	125,37 (5,641)	35,63 (2,053)	< 0,0001	< 0,0001	< 0,0001	0,139	2,461	3,134	-168,38
#17	5,62 (0,023)	125,91 (5,739)	35,81 (2,087)	< 0,0001	< 0,0001	< 0,0001	0,108	2,037	2,640	-166,21
Todas	5,62 (0,015)	124,69 (3,822)	34,60 (1,422)	< 0,0001	< 0,0001	< 0,0001	0,099	2,985	5,943	-208,83

Como não houveram mudanças inferenciais em nenhum dos casos optou-se pelo modelo inicial. Uma possível melhoria no ajuste desse modelo seria utilizar um modelo que considerasse heterocedasticidade proposto por Wei (1998).

4.2 Crescimento de Colônias

Os dados consistem no crescimento de três colônias de *Paramecium aurelium* em um determinado meio nutritivo. No começo de cada experimento, 20 *paramecia* foram colocadas em um tubo com o meio nutritivo a certa temperatura. Cada dia, começando pelo segundo dia, o número de indivíduos foi contado até a estabilização do tamanho das colônias após 10 dias. Iremos considerar cada colônia em separado. Seja Y_{ij} o número de indivíduos no i -ésimo dia ($i = 1, \dots, 19$), em que j representa as colônias A, B e C, respectivamente, e x denota o número de dias. Vamos assumir o mesmo modelo proposto por Svetliza & Paula (2003), em que Y tem distribuição binomial negativa, isto é, $Y_{ij} \sim \text{BN}(\mu_i, \phi_j)$, em que a parte sistemática é dada por

$$\log \mu_{ij} = \exp\{\alpha_j - \exp(\beta_j - \gamma_j x_{ij})\}, \quad i = 1, \dots, 19,$$

com $j = 1, 2, 3$. As estimativas dos parâmetros, viés, erros padrão (assintóticos), estatística z e p -valores para as colônias A, B e C, são apresentadas nas Tabela 4.3, Tabela 4.4 e Tabela 4.5, respectivamente.

Tabela 4.3: Estimativas de máxima verossimilhança com os respectivos erros padrão aproximados obtidos do modelo binomial negativa ajustado aos dados sobre a Colônia A de *Paramecium aurelium*.

Parâmetro	Estimativa	Viés	Erro Padrão	Estatística z	p -valor
α	1,85	0,0000	0,010	178,74	$< 0,0001$
β	0,71	0,0000	0,135	5,29	$< 0,0001$
γ	0,36	0,0000	0,032	11,16	$< 0,0001$
AIC	200,50				

Observe que todos os parâmetros são individualmente significativos (p -valor $< 0,0001$) ao nível de significância de 1%. Portanto, têm-se indícios de que a relação funcional proposta tende a ser

Tabela 4.4: Estimativas de máxima verossimilhança com os respectivos erros padrão aproximados obtidos do modelo binomial negativa ajustado aos dados sobre a Colônia B de *Paramecium aurelium*.

Parâmetro	Estimativa	Viés	Erro Padrão	Estatística z	p -valor
α	1,84	0,0000	0,008	230,71	$< 0,0001$
β	0,62	0,0000	0,118	5,23	$< 0,0001$
γ	0,36	0,0000	0,028	12,92	$< 0,0001$
AIC	189,14				

Tabela 4.5: Estimativas de máxima verossimilhança com os respectivos erros padrão aproximados obtidos do modelo binomial negativa ajustado aos dados sobre a Colônia C de *Paramecium aurelium*.

Parâmetro	Estimativa	Viés	Erro Padrão	Estatística z	p -valor
α	1,85	0,0000	0,005	310,88	$< 0,0001$
β	0,86	0,0000	0,130	6,59	$< 0,0001$
γ	0,44	0,0000	0,031	14,30	$< 0,0001$
AIC	187,40				

uma boa aproximação para explicar a média da variável resposta. Além disso, percebemos que as estimativas das três colônias deram bem próximas, o que já era esperado e o valor das estimativas do viés deu bem próximo de zero.

Com o objetivo de detectar a existência de observações aberrantes e influentes com alguma interferência desproporcional nos resultados do ajuste, bem como verificar possíveis afastamentos das suposições feitas para o modelo e para as partes aleatória e sistemática, apresentamos nas Figuras 4.4, 4.5 e 4.6 uma análise de diagnóstico para as colônias A, B e C, respectivamente. Os gráficos de envelope dos modelos para as colônias A, B e C, respectivamente, estão nas Figuras 4.7, 4.8 e 4.9.

Destacou-se na Figura 4.4(b) as observações 2 e 5 como possíveis pontos influentes. Nota-se que as mesmas observações são pontos aberrantes (Figura 4.4(c)). As descrições dessas observações são as seguintes: #2 que corresponde a 17 *paramecia* que cresceram em dois dias no meio nutritivo; #5 que corresponde a 63 *paramecia* que cresceram em cinco dias no meio nutritivo. Na Figura 4.5 destacou-se as observações 3 e 5 que correspondem às seguintes observações: #3 que corresponde a 36 *paramecia* que cresceram em três dias no meio nutritivo e #5 que corresponde a 84 *paramecia*

Figura 4.4: Gráficos de diagnóstico referentes ao modelo binomial negativa ajustado aos dados sobre a Colônia A de *Paramecium aurelium*.

Figura 4.5: Gráficos de diagnóstico referentes ao modelo binomial negativa ajustado aos dados sobre a Colônia B de *Paramecium aurelium*.

Figura 4.6: Gráficos de diagnóstico referentes ao modelo binomial negativa ajustado aos dados sobre a Colônia C de *Paramecium aurelium*.

Figura 4.7: Gráfico normal de probabilidades (com envelope gerado) para o resíduo componente do desvio referente ao ajuste do modelo binomial negativa aos dados sobre a Colônia A de *Paramecium aurelium*.

Figura 4.8: Gráfico normal de probabilidades (com envelope gerado) para o resíduo componente do desvio referente ao ajuste do modelo binomial negativa aos dados sobre a Colônia B de *Paramecium aurelium*.

Figura 4.9: Gráfico normal de probabilidades (com envelope gerado) para o resíduo componente do desvio referente ao ajuste do modelo binomial negativa aos dados sobre a Colônia C de *Paramecium aurelium*.

que cresceram em cinco dias no meio nutritivo. Por outro lado, na Figura 4.6 as observações 3 e 15 foram destacadas e correspondem às seguintes observações: #3 que corresponde a 37 *paramecia* que cresceram em três dias no meio nutritivo e #15 que corresponde a 460 *paramecia* que cresceram em quinze dias no meio nutritivo.

A seguir, apresentamos uma análise confirmatória. Primeiramente, reajustamos o modelo eliminando individualmente cada observação descrita anteriormente para as colônias A, B e C, respectivamente. Em seguida, reajustamos o modelo eliminando conjuntamente todas as observações detectadas na análise de diagnóstico. As estimativas dos parâmetros dos modelos para as colônias A, B e C, respectivamente, podem ser vistas nas Tabelas 4.6, 4.7 e 4.8. Observe que todos os parâmetros continuam sendo individualmente significativos ao nível de significância de 1% mesmo excluindo todas as observações. As maiores variações percentuais aconteceram para β excluindo a observação #2 para a colônia A e a observação #3 para as colônias B e C.

Como não houveram mudanças inferenciais em nenhum dos casos optou-se pelo modelo inicial, ou seja, com todas as observações.

Tabela 4.6: Variação das estimativas do modelo binomial negativa para a colônia A ao excluir as observações citadas na análise de diagnóstico.

	Estimativas			p-valor			VP (%)			AIC
	α	β	γ	α	β	γ	α	β	γ	
Obs										
#2	1,85 (0,009)	0,92 (0,175)	0,40 (0,038)	< 0,0001	< 0,0001	< 0,0001	0,218	-28,984	-10,359	191,46
#5	1,85 (0,008)	0,72 (0,129)	0,37 (0,031)	< 0,0001	< 0,0001	< 0,0001	0,197	-0,850	-4,322	187,12
Todas	1,85 (0,008)	0,90 (0,168)	0,41 (0,036)	< 0,0001	< 0,0001	< 0,0001	0,353	-26,063	-13,096	178,80

Tabela 4.7: Variação das estimativas do modelo binomial negativa para a colônia B ao excluir as observações citadas na análise de diagnóstico.

	Estimativas			p-valor			VP (%)			AIC
	α	β	γ	α	β	γ	α	β	γ	
Obs										
#3	1,84 (0,007)	0,70 (0,142)	0,37 (0,030)	< 0,0001	< 0,0001	< 0,0001	0,042	-12,719	-3,227	182,06
#5	1,84 (0,007)	0,61 (0,117)	0,37 (0,028)	< 0,0001	< 0,0001	< 0,0001	0,110	0,777	-2,405	178,72
Todas	1,84 (0,007)	0,67 (0,141)	0,37 (0,030)	< 0,0001	< 0,0001	< 0,0001	0,128	-8,193	-4,444	172,30

Tabela 4.8: Variação das estimativas do modelo binomial negativa para a colônia C ao excluir as observações citadas na análise de diagnóstico.

	Estimativas			p-valor			VP (%)			AIC
	α	β	γ	α	β	γ	α	β	γ	
Obs										
#3	1,85 (0,005)	0,96 (0,162)	0,46 (0,035)	< 0,0001	< 0,0001	< 0,0001	0,037	-12,091	-3,797	180,21
#15	1,86 (0,005)	0,85 (0,118)	0,43 (0,027)	< 0,0001	< 0,0001	< 0,0001	-0,233	1,028	1,478	170,77
Todas	1,86 (0,005)	0,95 (0,147)	0,45 (0,030)	< 0,0001	< 0,0001	< 0,0001	-0,191	-10,942	-2,296	163,43

4.3 Cálcio Radioativo

Os dados consistem num experimento para análise bioquímica de armazenamento intracelular e transporte de cálcio através da membrana plasmática. Uma variável explicativa foi considerada, x , e consiste nas células que ficavam suspensas em uma solução de cálcio radioativo com tempos fixados de 0,45 até 15 minutos. A variável resposta Y representa a quantidade de cálcio radioativo (em nmoles/mg) absorvida pelas células. Vamos assumir o mesmo modelo proposto por Rawlings et al. (1998), em que Y tem distribuição Normal, isto é, $Y_i \sim N(\mu_i, \sigma^2)$, sendo μ_i dada pela relação não linear

$$\mu_i = \beta_0 \{1 - \exp(-\beta_1 x_i)\}, \quad i = 1, \dots, 27.$$

As estimativas dos parâmetros, viés, erros padrão (assintóticos), estatística z e p -valores são apresentadas na Tabela 4.9. Note que todos os parâmetros são individualmente significativos (p -valor $< 0,0001$) ao nível de significância de 1% e que as estimativas apresentaram um leve viés. Portanto, têm-se indícios de que a relação funcional proposta tende a ser uma boa aproximação para explicar a média da variável resposta. Na Figura 4.10 apresentamos os dados observados juntamente com o modelo ajustado e notamos que a medida que vai aumentando o tempo de suspensão na solução de cálcio radioativo, a variabilidade aumenta, dando um leve indício de heterocedasticidade.

Tabela 4.9: Estimativas de máxima verossimilhança com os respectivos erros padrão aproximados obtidos do modelo normal ajustado aos dados sobre Cálcio Radioativo.

Parâmetro	Estimativa	Viés	Erro Padrão	Estatística z	p -valor
β_0	4,19	0,0473	0,322	13,02	$< 0,0001$
β_1	0,22	0,0022	0,048	4,64	$< 0,0001$
AIC	56,26				

Com o objetivo de detectar a existência de observações aberrantes e influentes com alguma interferência desproporcional nos resultados do ajuste, bem como verificar possíveis afastamentos das suposições feitas para o modelo e para as partes aleatória e sistemática, apresentamos na Figura 4.11 uma análise de diagnóstico e na Figura 4.12 apresentamos o gráfico de envelope.

Figura 4.10: Gráfico do modelo normal ajustado aos dados sobre Cálcio Radioativo.

Figura 4.11: Gráficos de diagnóstico referentes ao modelo normal ajustado aos dados sobre Cálcio Radioativo.

Figura 4.12: Gráfico normal de probabilidades (com envelope gerado) para o resíduo componente do desvio referente ao ajuste do modelo normal aos dados sobre Cálcio Radioativo.

Note através da Figura 4.12 uma boa acomodação dos pontos dentro do envelope, ou seja, não há indícios de afastamentos da suposição de distribuição normal para a quantidade de cálcio radioativo (em nmoles/mg) absorvida pelas células apesar da indicação de heterocedasticidade observada na Figura 4.10. Portanto optamos pelo modelo inicial.

4.4 Sobrevivência de Pacientes com Leucemia

Os dados consistem de uma amostra de 33 pacientes que morreram de leucemia aguda. Duas variáveis explicativas foram consideradas: WBC – contagem de células brancas no sangue; e AG – presença ou não de uma certa característica morfológica em WBC. A variável resposta, Y , é binária (1: sobrevive; 0: não sobrevive) e refere-se à sobrevivência do paciente pelo menos 52 semanas após o diagnóstico. Vamos assumir o mesmo modelo proposto por Lee (1988) e Wei (1998), em que Y tem distribuição binomial, isto é, $Y_i \sim B(n_i, p_i)$, em que a parte sistemática é dada por

$$\log \left\{ \frac{p_i}{1 - p_i} \right\} = \beta_0 + \beta_1 \text{AG}_i + \beta_2 \text{WBC}_i^\lambda, \quad i = 1, \dots, 30,$$

em que $p_i = \Pr(Y_i = 1)$ denota a probabilidade de sobrevivência. De acordo com o modelo anterior, se λ for muito grande então β_2 deverá ser pequeno para compensar o efeito exponencial que λ exerce sobre a variável WBC e vice versa.

O algoritmo proposto no Capítulo 2, usado para estimar conjuntamente os parâmetros β_0 , β_1 , β_2 e λ do modelo proposto, não convergiu. Uma forma simples de contornar esse problema é considerar uma região (grade de valores) para λ e estimar o modelo

$$\eta_i = \beta_0 + \beta_1 \text{AG}_i + \beta_2 \text{WBC}_i^\lambda,$$

para cada λ na grade de valores e, por exemplo, tomar λ tal que o modelo apresente o menor valor para o AIC. A Figura 4.13 apresenta uma grade de valores para λ versus o AIC dos modelos ajustados. Nota que $\lambda = -0,3$ refere-se ao modelo com o menor valor para o AIC.

Observe pela Figura 4.13 que o valor de λ está muito próximo de zero. Se fizermos $\lambda = 0$ no

Figura 4.13: Gráfico dos valores dos AIC para cada valor de λ .

modelo inicial considerado, temos o modelo

$$\eta_i = \beta_0 + \beta_1 \text{AG}_i + \beta_2 \log(\text{WBC}_i),$$

que representa um modelo linear generalizado. Assim, para decidirmos qual modelo se adequa melhor

Tabela 4.10: Estimativas de máxima verossimilhança com os respectivos erros padrão aproximados obtidos do modelo logístico ajustado aos dados sobre pacientes com Leucemia considerando $\lambda = -0,3$.

Parâmetro	Estimativa	Viés	Erro Padrão	Estatística z	p -valor
β_0	-5,61	-0,0000	2,164	-2,59	0,009
β_1	2,59	-0,0000	1,113	2,32	0,019
β_2	53,06	-0,0000	25,542	2,07	0,037
AIC	31,91				

Tabela 4.11: Estimativas de máxima verossimilhança com os respectivos erros padrão aproximados obtidos do modelo logístico ajustado aos dados sobre pacientes com Leucemia considerando $\lambda = 0$.

Parâmetro	Estimativa	Viés	Erro Padrão	Estatística z	p -valor
β_0	7,06	-0,0000	4,263	1,65	0,097
β_1	2,50	-0,0000	1,066	2,35	0,018
β_2	-0,98	-0,0000	0,483	-2,04	0,040
AIC	32,29				

aos dados, iremos ajustar os modelos

$$\eta_i = \begin{cases} \beta_0 + \beta_1 AG_i + \beta_2 \log(WBC_i), & \text{se } \lambda = 0 \\ \beta_0 + \beta_1 AG_i + \beta_2 WBC_i^\lambda, & \text{se } \lambda \neq 0, \end{cases}$$

e verificar, através de uma análise de diagnóstico, se algum deles é superior ao outro. As estimativas dos parâmetros, viés, erros padrão (assintóticos), estatística z e p -valores para os parâmetros com $\lambda = -0,3$ e $\lambda = 0$ são apresentadas nas Tabelas 4.10 e 4.11, respectivamente. Nota que todos os parâmetros do modelo com $\lambda = -0,3$ são significativos (p -valor $< 0,05$) ao nível de significância de 5%, já para o modelo com $\lambda = 0$, todos os parâmetros são significativos ao nível de significância de 10%. Observamos também que o valor do AIC dos dois modelos ficaram bem próximos e que a estimativa do viés dos dois modelos ficou bem próxima de zero.

Os gráficos de diagnóstico para o modelo estimado com $\lambda = -0,3$ são apresentados nas Figuras 4.14 e 4.15. A Figura 4.14 indica que existem poucos pontos influentes o que é uma indicação que o modelo com $\lambda = -0,3$ parece representar bem o problema. Os gráficos de pontos de alavanca e pontos

influentes indicam que o ponto 19 tende a apresentar maior influência nas estimativas dos parâmetros do modelo que os outros pontos, no entanto, esse efeito não é grande. A Figura 4.15 apresenta o envelope construído com base nos resíduos do modelo e fornece evidências de que o modelo se adequa bem aos dados, já que não é observado nenhum ponto discrepante nesta figura.

Figura 4.14: Gráficos de diagnóstico referentes ao modelo logístico ajustado aos dados sobre pacientes com Leucemia considerando $\lambda = -0,3$.

Os gráficos de diagnóstico para o modelo estimado com $\lambda = 0$ são apresentados nas Figuras

Figura 4.15: Gráfico normal de probabilidades (com envelope gerado) para o resíduo componente do desvio referente ao ajuste do modelo logístico aos dados sobre pacientes com Leucemia considerando $\lambda = -0,3$.

4.16 e 4.17. Nota-se que esse modelo tende a apresentar mais observações influentes, porém, não é observado grande diferença desse modelo em relação ao outro modelo. Assim, na prática, os dois modelos representam bem os dados e optariamos pelo modelo linear, ou seja, com $\lambda = 0$ pois é um modelo mais simples de ser estimado.

Figura 4.16: Gráficos de diagnóstico referentes ao modelo logístico ajustado aos dados sobre pacientes com Leucemia considerando $\lambda = 0$.

Figura 4.17: Gráfico normal de probabilidades (com envelope gerado) para o resíduo componente do desvio referente ao ajuste do modelo normal aos dados sobre pacientes com Leucemia considerando $\lambda = 0$.

4.5 Produção de Gramíneas

O principal interesse desse exemplo é estudar o resultado de experimentos fatoriais com os três principais nutrientes da planta (em lb/acre) na produção de gramíneas no litoral de Bermuda. Três variáveis explicativas foram consideradas: x_1 – nitrogênio N, x_2 – fósforo F e x_3 – potássio P. A variável resposta, Y , é a produção média de todos os três anos (1955, 1956, 1957). Vamos assumir o mesmo modelo proposto por Wei (1998), em que Y tem distribuição Gama, isto é, $Y_i \sim G(\mu_i, \phi)$, em que a parte sistemática é dada por

$$\mu_i^{-1} = \beta_0 + \beta_1 v_1 + \beta_2 v_2 + \beta_3 v_3,$$

com $v_i = 1/(x_i + \alpha_i)$ e $i = 1, 2, 3$. Aqui x_i ($i = 1, 2, 3$) são os valores usados de N, F e P, respectivamente, enquanto que α_i são valores desconhecidos no solo. As estimativas dos parâmetros, viés, erros padrão (assintóticos), estatística z e p -valores são apresentadas na Tabela 4.12.

Tabela 4.12: Estimativas de máxima verossimilhança com os respectivos erros padrão aproximados obtidos do modelo gama ajustado aos dados sobre produção de gramíneas.

Parâmetro	Estimativa	Viés	Erro Padrão	Estatística z	p -valor
β_0	0,09	-0,0019	0,009	10,03	< 0,0001
β_1	13,15	0,0458	1,359	9,67	< 0,0001
β_2	0,69	0,1264	0,455	1,51	0,1343
β_3	1,35	0,3116	0,975	1,39	0,1698
α_1	44,61	0,1141	4,214	10,58	< 0,0001
α_2	15,41	1,6653	8,444	1,82	0,0732
α_3	32,79	4,4138	19,408	1,69	0,0966
AIC	12,17				

Observando a Tabela 4.12 verificamos que o parâmetro β_0 é altamente significativo (p -valor <0,001) e o parâmetro β_1 , associado com o composto Nitrogênio, também é altamente significativo. Já os parâmetros β_2 , associado com o composto Fósforo, e β_3 , associado com o composto Potássio, parecem não ter relação linear com o preditor linear (p -valor >0,10), no entanto, ainda mantêm alguma relação não linear com o preditor ao nível de significância de 10% (α_2 e α_3). Os gráficos de

diagnóstico para o modelo estimado são apresentados nas Figuras 4.18 e 4.19. Através da Figura 4.19 observamos um ponto fora da banda de confiança. No entanto, está dentro dos 5% esperado, ou seja, o modelo está bem ajustado.

Figura 4.18: Gráficos de diagnóstico referentes ao modelo logístico ajustado aos dados sobre produção de gramíneas.

Figura 4.19: Gráfico normal de probabilidades (com envelope gerado) para o resíduo componente do desvio referente ao ajuste do modelo gama aos dados sobre produção de gramíneas.

Nem sempre podemos considerar independência entre as observações. Dessa forma, extensões para o caso de dados correlacionados são apresentados no próximo capítulo.

Capítulo 5

Extensões para Dados Correlacionados

Os modelos lineares generalizados são definidos com a suposição de respostas independentes, todavia extensões para o caso de dados correlacionados têm sido propostas por vários autores nos últimos 20 anos. É o caso, por exemplo, das equações de estimação generalizadas (EEGs) desenvolvidas por Liang & Zeger (1986) que estenderam os modelos de quase verossimilhança para dados correlacionados. Os modelos lineares generalizados mistos (MLGMs) (Breslow & Day (1993); McCulloch & Searle (2001)) constituem outra forma de tratar dados correlacionados na classe dos MLGs, em que o preditor linear é também formado por efeitos aleatórios cuja distribuição é assumida gaussiana. Lee & Nelder (1996); Lee & Nelder (2001) propuseram uma flexibilização para a distribuição dos efeitos aleatórios na classe dos MLGMs, contudo sugerem que a estimação seja feita sob a estrutura hierárquica em que os efeitos aleatórios são estimados conjuntamente com os coeficientes da regressão.

Neste capítulo faremos uma breve discussão sobre as possíveis extensões das EEGs para o caso em que o preditor é não linear e apresentaremos uma proposta de inclusão de efeitos aleatórios nos modelos não lineares de família exponencial mistos.

5.1 Equações de Estimação Generalizadas

Seguindo notação similar àquela apresentada em Paula (2004) denotaremos por $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^\top$ o vetor resposta multivariado para a i -ésima unidade experimental, $i = 1, \dots, n$, em que Y_{it} é a

observação do indivíduo i no instante t , $t = 1, \dots, n_i$, e assumiremos que a distribuição marginal de Y_{it} é da forma dada em (2.1) sendo que com $E(Y_{it}) = \mu_{it}$, $\text{Var}(Y_{it}) = \phi_i^{-1} V_{it}$, e $V_{it} = d\mu_{it}/d\theta_{it}$ é a função de variância e $\phi_i^{-1} > 0$ é o parâmetro de dispersão, em geral desconhecido. Podemos definir um modelo não linear de família exponencial para cada instante t supondo que a componente sistemática é dada por

$$g(\mu_{it}) = f(\mathbf{x}_{it}; \boldsymbol{\beta}),$$

em que $g(\cdot)$ e $f(\cdot; \cdot)$ são definidas como na Seção 2.2, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ contém os parâmetros a serem estimados e $\mathbf{x}_{it} = (x_{it1}, \dots, x_{itq})^\top$ contém os valores de variáveis explicativas que podem ser dependentes do tempo no caso de dados longitudinais.

Como em geral não é possível conhecer a distribuição de $\mathbf{Y}_i$ podemos fazer suposições sobre a estrutura de correlação intraclasse, ou seja, dentro de cada unidade experimental. Assim, similarmente às EEGs vamos supor que

$$\text{Var}(\mathbf{Y}_i) = \phi^{-1} \mathbf{V}_i^{1/2} \mathbf{R}_i \mathbf{V}_i^{1/2},$$

em que $\mathbf{V}_i^{1/2} = \text{diag}\{\sqrt{V_{i1}}, \dots, \sqrt{V_{in_i}}\}$ e $\mathbf{R}_i$ é uma matriz $n_i \times n_i$ conhecida como matriz trabalho que pode ser estruturada ou não estruturada. No caso de $\mathbf{R}_i$ ser uma matriz estruturada é usual supor que $\mathbf{R}_i = \mathbf{R}_i(\boldsymbol{\rho})$, em que $\boldsymbol{\rho} = (\rho_1, \dots, \rho_r)^\top$ é um vetor de parâmetros de perturbação que não dependem de $\boldsymbol{\beta}$.

5.2 Estimação de $\boldsymbol{\beta}$

Para os modelos não lineares de família exponencial as EEGs serão dadas por

$$S_\beta(\hat{\boldsymbol{\beta}}_G) = \mathbf{0},$$

em que $S_\beta(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{J}_i^\top \boldsymbol{\Omega}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i)$, com $\mathbf{J}_i = \partial \boldsymbol{\eta}_i / \partial \boldsymbol{\beta}^\top$, sendo $\boldsymbol{\eta}_i = (f(x_{i1}; \boldsymbol{\beta}), \dots, f(x_{in_i}; \boldsymbol{\beta}))^\top$, $\boldsymbol{\Omega}_i = \phi^{-1} \mathbf{V}_i^{1/2} \mathbf{R}_i(\boldsymbol{\rho}) \mathbf{V}_i^{1/2}$, $\mathbf{y}_i = (Y_{i1}, \dots, Y_{in_i})^\top$ e $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{in_i})^\top$. Note que $\mathbf{J}_i$ é uma matriz de dimensão $n_i \times p$ assumida de posto completo.

A estimação de β é feita separadamente de ρ e ϕ que são estimados de forma consistente, por exemplo, através do estimador pelo método dos momentos. Assim, um processo iterativo condicional a ρ e ϕ para estimar β pode ser expresso na forma abaixo

$$\beta^{(m+1)} = \beta^{(m)} + \left\{ \sum_{i=1}^n \mathbf{J}_i^{(m)\top} \boldsymbol{\Omega}_i^{-(m)} \mathbf{J}_i^{(m)\top} \right\}^{-1} \left[\sum_{i=1}^n \mathbf{J}_i^{(m)\top} \boldsymbol{\Omega}_i^{-(m)} (\mathbf{y}_i - \boldsymbol{\mu}_i^{(m)}) \right], \quad (5.1)$$

para $m = 0, 1, 2, \dots$. Valores iniciais para β são necessários no processo de estimação.

Supondo que $\hat{\rho}$ e $\hat{\phi}$ são estimadores consistentes (por exemplo através do método de momentos) é razoável supor que para n grande

$$\sqrt{n}(\hat{\beta}_G - \beta) \xrightarrow{d} N_p(\mathbf{0}, \Sigma),$$

em que $\Sigma = \lim_{n \rightarrow \infty} [n \mathbf{H}_1^{-1}(\beta) \mathbf{H}_2(\beta) \mathbf{H}_1^{-1}(\beta)]$.

Portanto, para grandes amostras $\hat{\beta}_G$ tem distribuição aproximadamente normal de média β e matriz de variância-covariância $\text{Var}(\hat{\beta}_G)$ dada por

$$\text{Var}(\hat{\beta}_G) = \mathbf{H}_1^{-1}(\hat{\beta}_G) \mathbf{H}_2(\hat{\beta}_G) \mathbf{H}_1^{-1}(\hat{\beta}_G), \quad (5.2)$$

com $\mathbf{H}_1(\beta) = \sum_{i=1}^n \{ \mathbf{J}_i^\top \boldsymbol{\Omega}_i^{-1} \mathbf{J}_i \}$ e $\mathbf{H}_2(\beta) = \sum_{i=1}^n \{ \mathbf{J}_i^\top \boldsymbol{\Omega}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) (\mathbf{y}_i - \boldsymbol{\mu}_i)^\top \boldsymbol{\Omega}_i^{-1} \mathbf{J}_i \}$.

O estimador dado por (5.2) é conhecido como estimador *sandwich* sendo robusto contra especificações incorretas da matriz trabalho $\mathbf{R}_i(\rho)$.

5.3 Estruturas de Correlação

5.3.1 Não Estruturada

Quando a matriz trabalho $\mathbf{R}_i(\boldsymbol{\rho})$ é não estruturada o vetor $\boldsymbol{\rho}$ terá dimensão $n_i(n_i-1)/2$ e o (l, j) -ésimo elemento da matriz $\mathbf{R}_i(\boldsymbol{\rho})$ será estimado por $\hat{R}_{ilj}$ que assumirá a seguinte expressão:

$$\hat{R}_{ilj} = \frac{1}{n-p} \sum_{i=1}^n \frac{(y_{il} - \hat{\mu}_{il})(y_{ij} - \hat{\mu}_{ij})}{\sqrt{\hat{V}_{il}}\sqrt{\hat{V}_{ij}}}.$$

5.3.2 Estruturada

Simétrica ou Permutável

A estrutura simétrica assume que o (l, j) -ésimo elemento da matriz $\mathbf{R}_i(\rho)$ é definido por

$$R_{ilj}(\rho) = \begin{cases} 1 & \text{se } l = j \\ \rho & \text{se } l \neq j. \end{cases}$$

Assim, a matriz $\mathbf{R}_i(\rho)$ que tem dimensão $n_i \times n_i$ fica dada por

$$\begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \vdots & \dots & \dots & \ddots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{bmatrix}.$$

Uma estimativa consistente para ρ é dada por

$$\hat{\rho} = \frac{1}{n} \sum_{i=1}^n \frac{1}{n_i(n_i-1)} \sum_{t \neq t'} \hat{r}_{it} \hat{r}_{it'},$$

em que $\hat{r}_{it} = (y_{it} - \hat{\mu}_{it}) / \sqrt{V(\hat{\mu}_{it})}$ é o resíduo de Pearson não escalonado.

Autoregressiva AR(1)

A estrutura autoregressiva de ordem 1 assume que o (l, j) -ésimo elemento da matriz $\mathbf{R}_i(\rho)$ assume a forma $R_{ilj}(\rho) = \rho^{|l-j|}$. Uma estimativa consistente para o parâmetro ρ é dada por

$$\hat{\rho} = \frac{1}{n} \sum_{i=1}^n \frac{1}{n_i(n_i - 1)} \sum_{t=1}^{n_i-1} \hat{r}_{it} \hat{r}_{i(t+1)}.$$

Para finalizar, uma estimativa consistente para ϕ é dada por

$$\hat{\phi} = \frac{1}{n} \sum_{i=1}^n \frac{1}{n_i} \sum_{t=1}^{n_i} \hat{r}_{it}^2.$$

Assim, teremos o seguinte processo de estimação para os parâmetros β , ρ e ϕ :

1. Atribuir valores iniciais $\beta^{(0)}$;
2. Calcular as médias $\mu_{it}^{(0)} = g^{-1}\{f(\mathbf{x}_{it}; \beta^{(0)})\}$ e consequentemente as quantidades $\phi^{(0)}$ e $\rho^{(0)}$;
3. Ir ao processo iterativo (5.1) e obter $\beta^{(1)}$;
4. Repetir os passos (1)-(3) até a convergência.

5.3.3 Modelos Não Lineares de Família Exponencial Mistos

Novamente iremos denotar por $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^\top$ o vetor resposta para a i -ésima unidade experimental e similarmente a McCulloch & Searle (2001) vamos propor a seguinte estrutura hierárquica:

1. $Y_{it} | \mathbf{b}_i \stackrel{ind}{\sim} \text{FE}(\mu_{it}, \phi)$;
2. $g(\mu_{it}) = f(\mathbf{x}_{it}; \beta) + \mathbf{z}_{it}^\top \mathbf{b}_i$;

$$3. \mathbf{b}_i \stackrel{ind}{\sim} N_s(\mathbf{0}, \mathbf{D}),$$

ou seja, estamos assumindo que a distribuição condicional de Y_{it} dado o vetor de efeitos aleatórios $\mathbf{b}_i = (b_{i1}, \dots, b_{is})^\top$ segue uma distribuição na família exponencial de distribuições de média μ_{it} e parâmetro de dispersão ϕ^{-1} . Além disso, temos agora que a média μ_{it} depende, através da função de ligação $g(\cdot)$, da parte fixa não linear $f(\mathbf{x}_{it}; \boldsymbol{\beta})$ e da parte aleatória dada por $\mathbf{z}_{it}\mathbf{b}_i$, em que $\mathbf{z}_{it}^\top = (z_{it1}, \dots, z_{its})^\top$ contém valores de variáveis explicativas e $\mathbf{b}_i$ é o vetor de efeitos aleatórios cuja distribuição é assumida normal de média zero e matriz de variância-covariância $\mathbf{D}$. A estrutura hierárquica (1)-(3) foi proposta recentemente por Tang et al. (2006b) para a classe dos modelos não lineares de família exponencial.

Cálculo de momentos

Segundo McCulloch & Searle (2001) temos que a esperança e a variância de Y_{it} e a covariância entre Y_{it} e $Y_{it'}$ com $t \neq t'$ são obtidas da seguinte forma:

$$\begin{aligned} E(Y_{it}) &= E[E(Y_{it}|\mathbf{b}_i)] \\ &= E(\mu_{it}) \\ &= E[g^{-1}\{f(\mathbf{x}_{it}; \boldsymbol{\beta}) + \mathbf{z}_{it}^\top \mathbf{b}_i\}], \end{aligned}$$

$$\begin{aligned} \text{Var}(Y_{it}) &= \text{Var}[E(Y_{it}|\mathbf{b}_i)] + E[\text{Var}(Y_{it}|\mathbf{b}_i)] \\ &= \text{Var}(\mu_{it}) + E[\phi^{-1}V(\mu_{it})] \\ &= \text{Var}[g^{-1}\{f(\mathbf{x}_{it}; \boldsymbol{\beta}) + \mathbf{z}_{it}^\top \mathbf{b}_i\}] + E[\phi^{-1}V\{g^{-1}(f(\mathbf{x}_{it}; \boldsymbol{\beta}) + \mathbf{z}_{it}^\top \mathbf{b}_i)\}] \end{aligned}$$

e

$$\begin{aligned}
\text{Cov}(Y_{it}, Y_{it'}) &= \text{Cov}\{E(Y_{it}|\mathbf{b}_i), E(Y_{it'}|\mathbf{b}_i)\} + E\{\text{Cov}(Y_{it}|\mathbf{b}_i, Y_{it'}|\mathbf{b}_i)\} \\
&= \text{Cov}(\mu_{it}, \mu_{it'}) + 0 \\
&= \text{Cov}[g^{-1}\{f(\mathbf{x}_{it}; \boldsymbol{\beta}) + \mathbf{z}_{it}^\top \mathbf{b}_i\}, g^{-1}\{f(\mathbf{x}_{it'}; \boldsymbol{\beta}) + \mathbf{z}_{it'}^\top \mathbf{b}_i\}]
\end{aligned}$$

para $t \neq t'$. Por exemplo, para ligação identidade $g(\mu) = \mu$ temos algumas simplificações

$$\begin{aligned}
E(Y_{it}) &= E[f(\mathbf{x}_{it}; \boldsymbol{\beta}) + \mathbf{z}_{it}^\top \mathbf{b}_i] \\
&= f(\mathbf{x}_{it}; \boldsymbol{\beta}),
\end{aligned}$$

$$\begin{aligned}
\text{Var}(Y_{it}) &= \text{Var}[f(\mathbf{x}_{it}; \boldsymbol{\beta}) + \mathbf{z}_{it}^\top \mathbf{b}_i] + \phi^{-1} E[V(f(\mathbf{x}_{it}; \boldsymbol{\beta}) + \mathbf{z}_{it}^\top \mathbf{b}_i)] \\
&= \mathbf{z}_{it}^\top \mathbf{D} \mathbf{z}_{it} + \phi^{-1} E[V(f(\mathbf{x}_{it}; \boldsymbol{\beta}) + \mathbf{z}_{it}^\top \mathbf{b}_i)]
\end{aligned}$$

e

$$\begin{aligned}
\text{Cov}(Y_{it}, Y_{it'}) &= \text{Cov}(f(\mathbf{x}_{it}; \boldsymbol{\beta}) + \mathbf{z}_{it}^\top \mathbf{b}_i, f(\mathbf{x}_{it'}; \boldsymbol{\beta}) + \mathbf{z}_{it'}^\top \mathbf{b}_i) \\
&= \text{Cov}(\mathbf{z}_{it}^\top \mathbf{b}_i, \mathbf{z}_{it'}^\top \mathbf{b}_i) \\
&= \mathbf{z}_{it}^\top \mathbf{D} \mathbf{z}_{it'}
\end{aligned}$$

para $t \neq t'$.

Estimação

Vamos supor que a matriz $\mathbf{D}$ seja estruturada, por exemplo, $\mathbf{D} = \text{diag}\{d_1, \dots, d_s\}$. Assim, os parâmetros a serem estimados são dados por $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi, \mathbf{d}^\top)^\top$ em que $\mathbf{d} = (d_1, \dots, d_s)^\top$. Ou seja, teremos $p + 1 + s$ parâmetros para serem estimados.

Denotaremos por $f_i(y_{ij}|\mathbf{b}_i)$ a função densidade de probabilidades (f.d.p.) de $Y_{ij}|\mathbf{b}_i$, por $f(\mathbf{b}_i)$

a f.d.p. de $\mathbf{b}_i$ e por $f(y_{ij}, \mathbf{b}_i)$ a f.d.p. conjunta de $(Y_{ij}, \mathbf{b}_i)$. Assim, a f.d.p. marginal de $\mathbf{Y}_i$ fica dada por

$$f(\mathbf{y}_i) = \int \left\{ \prod_{j=1}^{n_i} f_i(y_{ij} | \mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i \right\}, \quad (5.3)$$

em que $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^\top$. Em geral (5.3) não tem forma fechada sendo necessário a utilização de métodos de integração numérica, tais como Laplace e quadratura de Gauss-Hermite, para obter uma expressão aproximada para a f.d.p. $f(\mathbf{y}_i)$. Em McCulloch & Searle (2001) há uma descrição detalhada desses métodos com aplicações em modelos lineares generalizados mistos. Tang et al. (2006b) aplicam o método de Laplace em modelos não lineares de família exponencial mistos, porém não se conhece aplicações de outras metodologias de integração numérica nesta classe.

A inferência para os parâmetros $\boldsymbol{\beta}$, ϕ e $\mathbf{d}$ deverá ser feita através da f.d.p. marginal de $\mathbf{Y} = (\mathbf{Y}_1^\top, \dots, \mathbf{Y}_n^\top)^\top$ que será dada por

$$f(\mathbf{y}) = \prod_{i=1}^n f_i(\mathbf{y}_i),$$

em que $\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top)^\top$. Portanto, o logaritmo da função de verossimilhança para $\boldsymbol{\theta}$ será dado por

$$\begin{aligned} L(\boldsymbol{\theta}) &= \log f(\mathbf{y}) \\ &= \sum_{i=1}^n \log f_i(\mathbf{y}_i). \end{aligned}$$

Obviamente que $L(\boldsymbol{\theta})$ deverá em geral ter uma forma aproximada dependendo do método utilizado para obter $f_i(\mathbf{y}_i)$. As estimativas de máxima verossimilhança para $\boldsymbol{\beta}$, ϕ e $\mathbf{d}$ serão obtidas resolvendo-se as equações $\mathbf{U}_\beta(\boldsymbol{\theta}) = \mathbf{0}$, $\mathbf{U}_\phi(\boldsymbol{\theta}) = 0$ e $\mathbf{U}_\mathbf{d}(\boldsymbol{\theta}) = \mathbf{0}$, em que

$$\mathbf{U}_\beta(\boldsymbol{\theta}) = \frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}},$$

$$U_\phi(\boldsymbol{\theta}) = \frac{\partial L(\boldsymbol{\theta})}{\partial \phi}$$

e

$$U_d(\boldsymbol{\theta}) = \frac{\partial L(\boldsymbol{\theta})}{\partial \mathbf{d}},$$

respectivamente.

Métodos usuais tais como Newton-Raphson poderão ser utilizados para resolver essas equações. A fim de obter estimativas para os erros padrão das estimativas de máxima verossimilhança $\hat{\boldsymbol{\beta}}$, $\hat{\phi}$ e $\hat{\mathbf{d}}$ pode-se utilizar a matriz observada de Fisher $-\ddot{\mathbf{L}}_{\theta\theta}$, em que

$$\ddot{\mathbf{L}}_{\theta\theta} = \begin{bmatrix} \ddot{\mathbf{L}}_{\beta\beta} & \ddot{\mathbf{L}}_{\beta\phi} & \ddot{\mathbf{L}}_{\beta d} \\ \ddot{\mathbf{L}}_{\phi\beta} & \ddot{\mathbf{L}}_{\phi\phi} & \ddot{\mathbf{L}}_{\phi d} \\ \ddot{\mathbf{L}}_{d\beta} & \ddot{\mathbf{L}}_{d\phi} & \ddot{\mathbf{L}}_{dd} \end{bmatrix},$$

com $\ddot{\mathbf{L}}_{\beta\beta} = \partial^2 L(\boldsymbol{\theta}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top$, $\ddot{\mathbf{L}}_{\beta\phi} = \partial^2 L(\boldsymbol{\theta}) / \partial \boldsymbol{\beta} \partial \phi$, $\ddot{\mathbf{L}}_{\beta d} = \partial^2 L(\boldsymbol{\theta}) / \partial \boldsymbol{\beta} \partial \mathbf{d}^\top$, $\ddot{\mathbf{L}}_{\phi\phi} = \partial^2 L(\boldsymbol{\theta}) / \partial \phi^2$, $\ddot{\mathbf{L}}_{\phi d} = \partial^2 L(\boldsymbol{\theta}) / \partial \phi \partial \mathbf{d}^\top$ e $\ddot{\mathbf{L}}_{dd} = \partial^2 L(\boldsymbol{\theta}) / \partial \mathbf{d} \partial \mathbf{d}^\top$.

Assim, para grandes amostras, é razoável supor sob condições usuais de regularidade para $L(\boldsymbol{\theta})$ que $\hat{\boldsymbol{\theta}}$ tenha distribuição aproximadamente normal de média $\boldsymbol{\theta}$ e matriz de variância-covariância dada por

$$\text{Var}(\hat{\boldsymbol{\theta}}) = [-\ddot{\mathbf{L}}_{\theta\theta}]^{-1}.$$

Modelo com Intercepto Aleatório

Um caso particular dos modelos dados por (1)-(3) são os modelos não lineares de família exponencial com intercepto aleatório, descritos abaixo:

1. $Y_{it}|b_i \stackrel{ind}{\sim} \text{FE}(\mu_{it}, \phi);$
2. $g(\mu_{it}) = f(\mathbf{x}_{it}; \boldsymbol{\beta}) + b_i;$
3. $b_i \stackrel{ind}{\sim} N(0, \sigma_e^2).$

Neste caso o objetivo principal do efeito aleatório b_i é introduzir correlação dentro de cada unidade experimental. Ou seja, fazer com que os elementos Y_{it} e $Y_{it'}$ sejam correlacionados para $t \neq t'$. Quando $\sigma_e^2 = 0$ o modelo (1)-(3) acima reduz-se ao modelo não linear de família exponencial.

Neste caso particular teremos os seguintes momentos:

$$E(Y_{it}) = E[g^{-1}\{f(\mathbf{x}_{it}; \boldsymbol{\beta}) + b_i\}],$$

$$\text{Var}(Y_{it}) = \text{Var}[g^{-1}\{f(\mathbf{x}_{it}; \boldsymbol{\beta}) + b_i\}] + \phi^{-1}E[V\{g^{-1}(f(\mathbf{x}_{it}; \boldsymbol{\beta}) + b_i)\}]$$

e

$$\text{Cov}(Y_{it}, Y_{it'}) = \text{Cov}[g^{-1}\{f(\mathbf{x}_{it}; \boldsymbol{\beta}) + b_i\}, g^{-1}\{f(\mathbf{x}_{it'}; \boldsymbol{\beta}) + b_i\}].$$

Em particular para ligação identidade temos que

$$E(Y_{it}) = f(\mathbf{x}_{it}; \boldsymbol{\beta}),$$

$$\text{Var}(Y_{it}) = \sigma_e^2 + \phi^{-1}E[V(f(\mathbf{x}_{it}; \boldsymbol{\beta}) + b_i)]$$

e

$$\text{Cov}(Y_{it}, Y_{it'}) = \sigma_e^2.$$

Para $V(\mu) = \mu$ teremos que

$$\text{Var}(Y_{it}) = \sigma_e^2 + \phi^{-1}f(\mathbf{x}_{it}; \boldsymbol{\beta})$$

e portanto

$$\rho_{tt'} = \text{Corr}(Y_{it}, Y_{it'}) = \frac{\sigma_e^2}{\sqrt{\sigma_e^2 + \phi^{-1}f(\mathbf{x}_{it}; \boldsymbol{\beta})}\sqrt{\sigma_e^2 + \phi^{-1}f(\mathbf{x}_{it'}; \boldsymbol{\beta})}}.$$

O logaritmo da verossimilhança neste caso de intercepto aleatório fica dado por

$$L(\boldsymbol{\theta}) = \sum_{i=1}^n \log \int_{-\infty}^{\infty} \prod_{j=1}^{n_i} f_i(y_{ij}|b_i) f(b_i) db_i,$$

ou seja, teremos apenas uma única integral para ser resolvida numericamente.

Métodos de diagnóstico de deleção de pontos e influência local utilizando a aproximação de Laplace são apresentados em Tang et al. (2006b).

Para ilustrar um exemplo de modelo não linear de família exponencial com intercepto aleatório vamos considerar o exemplo apresentado na Seção 4.2 sobre o crescimento de colônias de *Paramecium aurelium*. Como para cada tipo de colônia o crescimento ocorre no mesmo meio nutritivo é bastante razoável pensarmos na suposição de correlação entre o número de indivíduos de dois tempos diferentes, isto é, entre Y_{ij} e $Y_{i'j}$ para $i \neq i'$.

Assim, podemos propor o seguinte modelo não linear de família exponencial de intercepto aleatório:

1. $Y_{ij}|b_j \stackrel{ind}{\sim} P(\mu_{ij});$
2. $\log \mu_{ij} = \eta_{ij} + b_j;$
3. $b_j \stackrel{ind}{\sim} N(0, \sigma_e^2),$

em que $i = 1, \dots, 19$, $j = 1, 2, 3$ e $\eta_{ij} = \exp\{\alpha_j - \exp(\beta_j - \gamma_j x_{ij})\}$.

Portanto, temos que

$$\begin{aligned} E(Y_{ij}) &= E[\exp(\eta_{ij} + b_j)] \\ &= \exp(\eta_{ij})E(e^{b_j}), \end{aligned}$$

$$\begin{aligned} \text{Var}(Y_{ij}) &= \text{Var}[\exp(\eta_{ij} + b_j)] + E[\exp(\eta_{ij} + b_j)]^2 \\ &= \exp(2\eta_{ij})\text{Var}(e^{b_j}) + \exp(2\eta_{ij})E(e^{b_j})^2 \end{aligned}$$

e

$$\begin{aligned}\text{Cov}(Y_{ij}, Y_{i'j}) &= \text{Cov}[\exp(\eta_{ij} + b_j), \exp(\eta_{i'j} + b_j)] \\ &= \exp(\eta_{ij} + \eta_{i'j}) \text{Var}(e^{b_j}) \quad \text{para } i \neq i' .\end{aligned}$$

Como é bem conhecido se $b \sim N(0, \sigma^2)$ então e^b segue uma log-normal de média $e^{\sigma^2/2}$ e variância $e^{\sigma^2}(e^{\sigma^2} - 1)$. Assim, para o modelo definido por (1)-(3) temos o seguinte:

$$E(Y_{ij}) = e^{\eta_{ij}} e^{\sigma_e^2/2} = e^{\frac{\sigma_e^2}{2} + \eta_{ij}},$$

$$\begin{aligned}\text{Var}(Y_{ij}) &= e^{2\eta_{ij}} e^{\sigma_e^2} (e^{\sigma_e^2} - 1) + e^{\eta_{ij}} e^{\sigma_e^2/2} \\ &= e^{2\eta_{ij}} \{e^{2\sigma_e^2} - e^{\sigma_e^2} + e^{-\eta_{ij} + \sigma_e^2/2}\},\end{aligned}$$

portanto temos contemplada a sobredispersão uma vez que

$$\text{Var}(Y_{ij}) > E(Y_{ij}).$$

A covariância entre Y_{ij} e $Y_{i'j}$ fica dada por $\text{Cov}(Y_{ij}, Y_{i'j}) = e^{\eta_{ij} + \eta_{i'j}} e^{\sigma_e^2} (e^{\sigma_e^2} - 1)$ e daí segue que a correlação entre Y_{ij} e $Y_{i'j}$ para $i \neq i'$ fica dada por

$$\begin{aligned}\rho_{ii'} &= \frac{e^{\eta_{ij} + \eta_{i'j}} e^{\sigma_e^2} (e^{\sigma_e^2} - 1)}{\sqrt{e^{2\eta_{ij}} \{e^{2\sigma_e^2} - e^{\sigma_e^2} + e^{-\eta_{ij} + \sigma_e^2/2}\}} \sqrt{e^{2\eta_{i'j}} \{e^{2\sigma_e^2} - e^{\sigma_e^2} + e^{-\eta_{i'j} + \sigma_e^2/2}\}}} \\ &= \frac{e^{\sigma_e^2} (e^{\sigma_e^2} - 1)}{\sqrt{(e^{2\sigma_e^2} - e^{\sigma_e^2} + e^{-\eta_{ij} + \sigma_e^2/2})} \sqrt{(e^{2\sigma_e^2} - e^{\sigma_e^2} + e^{-\eta_{i'j} + \sigma_e^2/2})}}.\end{aligned}$$

Note que quando $\sigma_e^2 = 0$ temos que $\rho_{ii'} = 0$ e portanto voltamos ao modelo não linear de família exponencial com respostas independentes.

Capítulo 6

Conclusões

Os modelos não lineares de família exponencial são extensões da classe de modelos lineares generalizados. A diferença fundamental entre esses dois modelos é que os modelos não lineares de família exponencial admitem preditores não lineares, ou seja, a componente sistemática é dada por $f(\mathbf{x}_i; \boldsymbol{\beta})$ que é uma função contínua, diferenciável e não linear em $\boldsymbol{\beta}$. Recentemente os modelos não lineares de família exponencial foram revisitados por Kosmidis (2007) cujo artigo sairá na *Biometrika*, o que mostra a importância desses modelos até hoje. Para conseguirmos implementar esse tipo de modelo devemos declarar a parte não linear como sendo um *offset*. Como visto no Capítulo 4, esse método de estimação proposto por Paula & Cordeiro (1986) mostrou-se bem eficiente na estimação dos parâmetros dos cinco exemplos apresentados. Já o viés pode ser expresso como sendo a solução de mínimos quadrados de uma regressão linear ponderada, assim como mostrado por Paula (1992) e foi diretamente calculado para os exemplos práticos. As técnicas de diagnóstico são as mesmas válidas para os modelos lineares generalizados sendo que a matriz do modelo aqui é dada por $\mathbf{J} = \partial \boldsymbol{\eta} / \partial \boldsymbol{\beta}^\top$.

Na análise de dados correlacionados não podemos utilizar os modelos de regressão clássicos que têm como uma das pressuposições básicas a independência entre as observações. Uma forma de trabalhar a correlação entre as observações é aplicar a metodologia de Equações de Estimação Generalizadas em que a estrutura de correlação é modelada através da inclusão de uma matriz trabalho. Uma outra maneira é a inclusão de efeitos aleatórios no preditor linear que chamamos de modelos lineares generalizados mistos. Nesses modelos as variáveis explicativas são complementadas

com um vetor $\mathbf{b}$ de efeitos aleatórios, em que os componentes de $\mathbf{y}$ condicionados a esses efeitos são considerados independentes. A principal dificuldade nesses modelos é encontrar a f.d.p. marginal de $\mathbf{Y}_i$ pois precisamos de métodos de integração numérica para resolver essa integral. Um possível trabalho futuro poderia estar relacionado na obtenção de uma expressão aproximada para a f.d.p. marginal de $\mathbf{Y}_i$ nos modelos não lineares de família exponencial mistos utilizando a quadratura de Gauss-Hermite ao invés da aproximação de Laplace como feito por Tang et al. (2006b).

Apêndice A

Aspectos Computacionais

Para o ajuste dos modelos estudados no Capítulo 4 foram utilizadas rotinas existentes no programa R, para a obtenção das estimativas de máxima verossimilhança, viés, gráfico de diagnóstico e gráfico normal de probabilidades para o resíduo componente do desvio.

A.1 Coelhos Europeus

ENTRADAS:

- beta1: Valor Inicial para estimação do primeiro parâmetro do modelo inverso gaussiano
- beta2: Valor Inicial para estimação do segundo parâmetro do modelo inverso gaussiano
- beta3: Valor Inicial para estimação do terceiro parâmetro do modelo inverso gaussiano
- n: número de iterações do algoritmo de estimação

SAÍDAS:

- betas: vetor de parâmetros estimados (sem o parâmetro de dispersão)
- phi: parâmetro de dispersão estimado pelo modelo.

```
library(MASS)
```

```
caminho_dados<- "C:\\Documents and Settings\\Wxp\\Meus documentos\\dri\\dissertacao\\coelhos"
```

```
setwd(caminho_dados)
```

```
dados <- read.csv("dados_coelho.csv",sep=";",header=T)
```

```
## Variável explicativa do modelo de coelhos Europeus : Idade dos coelhos (em dias)
```

```
x <- dados[,2]
```

```
## Variável resposta do modelo de coelhos Europeus : Peso Lentes Olhos (em mg)
```

```
y <- dados[,1]
```

```

## Valores iniciais para os parâmetros (entrada do modelo)
beta1 <- 15
beta2 <- 300
beta3 <- 30
betas <- c(beta1, beta2, beta3)
## Número de iterações do algoritmo de estimação
n <- 1000
for (i in 1:n){
  ## Matriz de Jacobiano do modelo (Derivada de Eta com relação à Beta)
  J1 <- rep(1,dim(dados)[1])
  J2 <- -1/(x+beta3)
  J3 <- beta2/((x+beta3)^2)
  J <- cbind(J1, J2, J3)
  Eta <- beta1 -(beta2/(x+beta3))
  W <- diag(1/(Eta^3))
  tal <- Eta - J%*%betas
  ## Aproximação inicial considerando todas as observações do modelo
  fit.model<- glm(log(y) ~ J1 + J2 + J3 -1 + offset(tal),family=inverse.gaussian(link ="identity"))
  #,subset=-c(1))
  betas <- coef(fit.model)
  beta1 <- betas[1]
  beta2 <- betas[2]
  beta3 <- betas[3]}
betas
## RESUMO DO AJUSTE DO GLM
summary(fit.model)
## Parâmetro de Dispersão
phi <- 1/summary(fit.model)$dispersion

## Cálculo do Viés
aux <- solve(t(J)%*%W%*%J)
tam_n <- length(y)
Zd <- array(0,c(tam_n,tam_n))
for (l in 1:tam_n){
  Zd[l,l] <- t(J[l,])%*%aux%*%J[l,]}

```

```

F <- array(0,c(tam_n,tam_n))
qsi_1 <- -(2*phi)^(-1)*Zd%%solve(W)%%F%%rep(1,tam_n)
p<-3
Jl <- array(0,c(p,p))
D <- diag(1,tam_n)
for (l in 1:tam_n){
  Jl[1,1] <- 0
  Jl[2,2] <- 0
  Jl[3,3] <- -2*beta2/((x[l] + beta3)^3)
  Jl[1,2] <- 0
  Jl[1,3] <- 0
  Jl[2,1] <- 0
  Jl[3,1] <- 0
  Jl[2,3] <- 1/((x[l] + beta3)^2)
  Jl[3,2] <- 1/((x[l] + beta3)^2)
  aux1 <- Jl%%aux
  D[l,l] <- sum(diag(aux1))}
qsi_2 <- -(2*phi)^(-1)*D%%rep(1,tam_n)
vicio <- aux%%t(J)%%W%%(qsi_1 + qsi_2)
betas_corrigidos <- betas - vicio

## IMPLANTAÇÃO DAS TÉCNICAS DE DIAGNÓSTICO (utilizado para todos os outros exemplos)
Matriz_Projeto <- model.matrix(fit.model)
n_linhas <- nrow(Matriz_Projeto)
n_parametros <- ncol(Matriz_Projeto)
pesos_w <- fit.model$weights
Matriz_Pesos_W <- diag(pesos_w)
# W^(.5) X (X' W X)^(-1) X' W X' W^(.5)
H <- solve(t(Matriz_Projeto)%%Matriz_Pesos_W%%Matriz_Projeto)
H <- sqrt(Matriz_Pesos_W)%%Matriz_Projeto%%H%%t(Matriz_Projeto)%%sqrt(Matriz_Pesos_W)
h <- diag(H)

## Resíduos versão Desvio(utilizado para todos os outros exemplos)
rd <- resid(fit.model,type="deviance")
## Resíduos padronizados (studentizados)(utilizado para todos os outros exemplos)

```

```

td <- rd*sqrt(phi/(1-h))
## Resíduos versão Pearson(utilizado para todos os outros exemplos)
rp <- resid(fit.model,type="pearson")
## Resíduos padronizados de Pearson (studentizados)(utilizado para todos os outros exemplos)
ts <- rp*sqrt(phi/(1-h))
## Likelihood Displacement (Distância de Cook)(utilizado para todos os outros exemplos)
LD <- h*(ts^ 2)/(1 - h)

## TÉCNICAS GRÁFICAS RECOMENDADAS(utilizado para todos os outros exemplos)
par(mfrow=c(2,2))
## Gráfico dos Valores ajustados contra os valores da diagonal da matriz chapéu
plot(fitted(fit.model),h,xlab="Valores Ajustados", ylab="Medida h", main="Pontos de Alavanca",
pch=16,ylim=c(0,1))
title(sub="(a)")
identify(fitted(fit.model),h)
## Gráfico da Distância de Coook Contra o tempo
plot(LD,xlab="Indice", ylab="Distancia de Cook", main="Pontos Influentes", pch=16)
title(sub="(b)")
identify(LD)
## Gráfico dos Valores ajustados contra o Resíduo do Desvio
plot(fitted(fit.model),td,xlab="Valores Ajustados", ylab="Componente do Desvio",
main="Pontos Aberrantes",
ylim=range(td),pch=16)
title(sub="(c)")
abline(2,0,lty=2)
abline(-2,0,lty=2)
identify(fitted(fit.model),td)
## Gráfico dos Valores ajustados contra o Resíduo de Pearson
plot(predict(fit.model),td,xlab="Preditor Linear", ylab="Componente do Desvio",
main="Função de Ligação",
ylim=range(td), pch=16)
title(sub="(d)")
abline(2,0,lty=2)
abline(-2,0,lty=2)
identify(predict(fit.model),td)

```

```

## Gráfico do modelo ajustado com os dados
z<-exp(fitted(fit.model))
plot(x,y,xlab="Idade dos coelhos (em dias)",ylab="Peso das lentes dos olhos dos coelhos (em mg)",
pch=16)
lines(x,z,xlab="Idade dos coelhos (em dias)",ylab="Peso das lentes dos olhos dos coelhos (em mg)",
type="l")

## Gráfico do modelo ajustado com os dados com a transformação nos dados
z<-(fitted(fit.model))
plot(x,log(y),xlab="x", ylab="y", pch=16)
lines(x,z,xlab="x", ylab="y",type="l")

## Gráfico de Envelope
library(statmod)

## Número de simulações do resíduo para construir o Intervalo de Confiança para o modelo
n_simulacoes <- 1000

## Matriz que armazenará os resíduos simulados com a distribuição postulada
residuo_simulado <- matrix(0,n_linhas,n_simulacoes)

## Utiliza a média ajustada do modelo para verificar a qualidade do ajuste
media_ajustada <- predict(fit.model,type="response")

for(i in 1:n_simulacoes){
##Simula a resposta com distribuição Normal inversa, média como sendo a média ajustada pelo modelo
## e parâmetro de dispersão estimado (Passo 1 do algoritmo)
resposta_simulada <- rinvgauss(n_linhas, media_ajustada ,phi)

## Ajusta o modelo postulado para a variável simulada (Passo 2 do algoritmo)
ajuste <- glm(resposta_simulada ~ Matriz_Projeto, family=inverse.gaussian(link=identity))
pesos_w <- ajuste$weights
Matriz_Pesos_W <- diag(pesos_w)

#  $W^{(.5)} X (X' W X)^{-1} X' W X' W^{(.5)}$ 
H <- solve(t(Matriz_Projeto)%*%Matriz_Pesos_W)%*%Matriz_Projeto)
H <- sqrt(Matriz_Pesos_W)%*%Matriz_Projeto)%*%H)%*%t(Matriz_Projeto)%*%sqrt(Matriz_Pesos_W)
h <- diag(H)

phi_sim <- 1/summary(ajuste)$dispersion

## Calcular o vetor de resíduos studentizados para cada simulação
## (Passo 3 e 4 do algoritmo)
residuo_simulado[,i] <- sort(resid(ajuste,type="deviance")*sqrt(phi_sim/(1-h)))}

```

```

## Passo 5 do algoritmo
## Pegar os valores "mínimos" e "máximos" dos resíduos studentizados simulados
## (são utilizados o mínimo e o P1 e P2 e máximo P99 e P98) para suavizar as bandas de confiança
e_100 <- apply(residuo_simulado,1,quantile,1)
e_99 <- apply(residuo_simulado,1,quantile,.99)
e_98 <- apply(residuo_simulado,1,quantile,.98)
e_00 <- apply(residuo_simulado,1,quantile,.00)
e_01 <- apply(residuo_simulado,1,quantile,.01)
e_02 <- apply(residuo_simulado,1,quantile,.02)
## Constrói a faixa como uma média
## Passo 6 do algoritmo
faixa_max<- (e_100+e_99+e_98)/3
faixa_min<- (e_00+e_01+e_02)/3
media_residuos <- apply(residuo_simulado,1,mean)
faixa <- range(td,e_00,e_100)
## Constrói o gráfico de envelope e coloca os limites estimados através de simulação
par(mfrow=c(1,1))
par(pty="s")
qqnorm(td,xlab="Percentis da N(0,1)",ylab="Componente do Desvio", ylim=faixa, pch=16)
#identify(qqnorm(td,xlab="Percentis da N(0,1)",
#ylab="Componente do Desvio", ylim=faixa, pch=16))
par(new=T)
qqnorm(faixa_min,axes=F,xlab="",ylab="",type="l",ylim=faixa,lty=1)
par(new=T)
qqnorm(faixa_max,axes=F,xlab="",ylab="", type="l",ylim=faixa,lty=1)
par(new=T)
qqnorm(media_residuos,axes=F,xlab="", ylab="", type="l",ylim=faixa,lty=2)

```

A.2 Crescimento de Colônias

ENTRADAS:

- alfa: Valor Inicial para estimação do primeiro parâmetro do modelo binomial negativa
- beta: Valor Inicial para estimação do segundo parâmetro do modelo binomial negativa
- gama: Valor Inicial para estimação do terceiro parâmetro do modelo binomial negativa
- n: número de iterações do algoritmo de estimação


```
## SAÍDAS:
- phi: parâmetro de dispersão estimado pelo modelo.
library(MASS)
rm(list=ls(all=T))
dados <- read.table("C:\\Documents and Settings\\Wxp\\Meus documentos\\dri\\dissertacao\\colonias\\
dados_colonia.txt",header=T)
attach(dados)
names(dados)
x <- dias#[-c(3,5)] #tirando as observações 3 e 5
#b <- b#[-c(3,5)] #tirando as observações 3 e 5
#a <- a#[-c(2,5)] #tirando as observações 2 e 5
c <- c#[-c(3,15)] #tirando as observações 3 e 15
n <- length(x)
teta<-2
eta<-matrix(20,n,1)
start<-matrix(1,n,1)
param<-matrix(1,3,1)
## Valores iniciais para os parâmetros (entrada do modelo)
alfa<-10
beta<-2
gama<-0.2
ff<-function(x,alfa,beta,gama){exp(alfa-exp(beta-gama*x))}
expr3<-exp((beta-(gama*x)))
expr5<-exp((alfa-expr3))
grad<-array(0,c(length(x),3), list(NULL,c("alfa","beta","gama")))
grad[, "alfa"]<-expr5
grad[, "beta"]<- -(expr5*expr3)
grad[, "gama"]<-expr5*(expr3*x)
param[1,1]<-alfa
param[2,1]<-beta
param[3,1]<-gama
ofs<-ff(x,alfa,beta,gama)-grad%*%param
## ESTIMAÇÃO DE PARÂMETROS PARA O MODELO COM TODAS AS OBSERVAÇÕES
#####AQUI DEVEMOS SUBSTITUIR POR COLONIA A, B E C NO AJUSTE DO MODELO
for (kk in 1:20){
```

```

t0<-teta
Theta<-teta
ofs<-ff(x,alfa,beta,gama)-grad%*%param
init.theta<-teta
fit.model<-glm.nb(c~ -1+grad+offset(ofs),control=glm.control(maxit = 40),init.theta=teta,link=log)
cat("numero=",kk,"")
Beta<-summary(fit.model)$coef
#print(Beta)
alfa<-summary(fit.model)$coef[1,1]
beta<-summary(fit.model)$coef[2,1]
gama<-summary(fit.model)$coef[3,1]
param[1,1]<-alfa
param[2,1]<-beta
param[3,1]<-gama
ff<- function(x,alfa,beta,gama){
exp(alfa-exp(beta-gama*x))}
expr3<-exp((beta-(gama*x)))
expr5<-exp((alfa-expr3))
grad<-array(0,c(length(x),3), list(NULL,c("alfa","beta","gama")))
grad[, "alfa"]<-expr5
grad[, "beta"]<- -(expr5*expr3)
grad[, "gama"]<-expr5*(expr3*x)
deviance<-fit.model$deviance
ofs<-ff(x,alfa,beta,gama)-grad%*%param
eta<-(fit.model)$fitted
teta<-summary(fit.model)$theta
to<-teta}

## RESUMO DO AJUSTE DO GLM
summary(fit.model)

## Parâmetro de Dispersão
phi <- fit.model$theta

## Cálculo do Viés
mu <- exp(grad%*%param)
tam_n <- length(c)#mudar para cada colonia

```

```

W <- diag(as.vector(mu/(mu/phi+1)))
J <- grad
aux <- solve(t(J)%*%W%*%J)
Zd <- array(0,c(tam_n,tam_n))
for (l in 1:tam_n){
  Zd[l,l] <- t(J[l,])%*%aux%*%J[l,]}
F <- array(0,c(tam_n,tam_n))
for (i in 1:tam_n){
  F[i,i] <- W[i,i]}
qsi_1 <- -(2*phi)^(-1)*Zd%*%solve(W)%*%F%*%rep(1,tam_n)
p<-3
Jl <- array(0,c(p,p))
D <- diag(1,tam_n)
for (l in 1:tam_n){
  Jl[1,1] <- exp(alfa - exp(beta-gama*x[l]))
  Jl[2,2] <- exp(alfa - exp(beta-gama*x[l]))*exp(2*(beta-gama*x[l])) -
exp(alfa-exp(beta-gama*x[l]))*exp(beta-gama*x[l])
  Jl[3,3] <- exp(alfa - exp(beta-gama*x[l]))*x[l]^2*exp(2*(beta-gama*x[l])) -
exp(alfa - exp(beta-gama*x[l]))*x[l]^2*exp(beta-gama*x[l])
  Jl[1,2] <- Jl[2,1] <- -exp(alfa - exp(beta-gama*x[l]))*exp(beta-gama*x[l])
  Jl[1,3] <- Jl[3,1] <- exp(alfa-exp(beta-gama*x[l]))*exp(beta-gama*x[l])*x[l]
  Jl[2,3] <- Jl[3,2] <- -exp(alfa-exp(beta-gama*x[l]))*exp(2*(beta-gama*x[l]))*x[l] +
x[l]*exp(beta-gama*x[l])*exp(alfa-exp(beta-gama*x[l]))
  aux1 <- Jl%*%aux
  D[l,l] <- sum(diag(aux1))}
qsi_2 <- -(2*phi)^(-1)*D%*%rep(1,tam_n)
vicio <- aux%*%t(J)%*%W%*%(qsi_1 + qsi_2)
betas_corrigidos <- c(alfa, beta, gama) - vicio

## Gráfico de Envelope
library(statmod)

## Número de simulações do resíduo para construir o Intervalo de Confiança para o modelo
n_simulacoes <- 1000
Matriz_Projeto <- model.matrix(fit.modelo)
n_linhas <- nrow(Matriz_Projeto)

```

```

e <- matrix(0,n_linhas,100)
for(i in 1:n_simulacoes){
  resp <- rnegbin(n_linhas, fitted(fit.model),phi)
  fit <- glm.nb(resp ~ Matriz_Projeto)
  pesos_w <- fit$weights
  Matriz_Pesos_W <- diag(pesos_w)
  #  $W^{(.5)} X (X' W X)^{-1} X' W X' W^{(.5)}$ 
  H <- solve(t(Matriz_Projeto)%*%Matriz_Pesos_W%*%Matriz_Projeto)
  H <- sqrt(Matriz_Pesos_W)%*%Matriz_Projeto%*%H%*%t(Matriz_Projeto)%*%sqrt(Matriz_Pesos_W)
  h <- diag(H)
  e[,i] <- sort(resid(fit,type="deviance")/sqrt(1-h))}
#
e1 <- numeric(n_linhas)
e2 <- numeric(n_linhas)
#
for(i in 1:n_linhas){
  eo <- sort(e[,i])
  e1[i] <- (eo[2]+eo[3])/2
  e2[i] <- (eo[97]+eo[98])/2}
#
med <- apply(e,1,mean)
faixa <- range(rd,e1,e2)
par(pty="s")
qqnorm(rd,xlab="Percentis da N(0,1)",
ylab="Componente do Desvio", ylim=faixa, pch=16)
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)

```

A.3 Cálcio Radioativo

```
## ENTRADAS:
- beta0: Valor Inicial para estimação do segundo parâmetro do modelo gaussiano
- beta1: Valor Inicial para estimação do terceiro parâmetro do modelo gaussiano
- n: número de iterações do algoritmo de estimação
## SAÍDAS:
- betas: vetor de parâmetros estimados (sem o parâmetro de dispersão)
- phi: parâmetro de dispersão estimado pelo modelo.
library(MASS)
caminho_dados<- "C:\\Documents and Settings\\Wxp\\Meus documentos\\dri\\dissertacao\\calcio"
setwd(caminho_dados)
dados <- read.csv("dados_calcio_radioativo.csv",sep=";",header=T)
## Variável explicativa do modelo de cálcio radioativo : tempo em suspensão em uma
solução de cálcio radioativo (em minutos)
x <- dados[,1]
## Variável resposta do modelo de cálcio radioativo : qdade de cálcio radioativo
absorvida pelas células (em nmoles/mg)
y <- dados[,2]
## RESUMO DO AJUSTE DO GLM
beta0<- 4 #matriz_betas[dev==min(dev),1]
beta1<- .3#matriz_betas[dev==min(dev),2]
m <- 300
betas<- c(beta0,beta1)
matriz_betas_reduzida <- array(1,c(m,length(betas)))
## Estimação do GLM começa aqui
aic <- numeric(0)
beta0 <- betas[1]
beta1 <- betas[2]
for (j in 1:m){
## Matriz de Jacobiano do modelo (Derivada de Eta com relação à Beta)
J1 <- rep(1,length(x)) - exp(-beta1*x)
J2 <- +beta0*x*exp(-beta1*x)
J <- cbind(J1, J2)
Eta <- beta0 - beta0*exp(-beta1*x)
```

```

tal <- Eta - J%*%betas
fit.model <- glm(y ~ J1 + J2 - 1 + offset(tal),family=gaussian(link = "identity"))
summary(fit.model)
aic<- c(aic, summary(fit.model)$aic)
betas <- coef(fit.model)
matriz_betas_reduzida[j,] <- betas
beta0<- coef(fit.model)[1]
beta1<- coef(fit.model)[2]}
summary(fit.model)
## Parâmetro de Dispersão
phi <- 1/summary(fit.model)$dispersion

## Cálculo do Viés
tam_n <- length(y)
Eta <- beta0 - beta0*exp(-beta1*x)
aux <- solve(t(J)%*%J)
p<-2
Jl <- array(0,c(p,p))
D <- diag(1,tam_n)
for (l in 1:tam_n){
  Jl[2,1] <- J[1,2] <- x[l]*exp(-beta1*x[l])
  Jl[2,2] <- -beta0*x[l]^2*exp(-beta1*x[l])
  aux1 <- Jl%*%aux
  D[l,1] <- sum(diag(aux1))}
qsi_2 <- -(2*phi)^(-1)*D%*%rep(1,tam_n)
vicio <- aux%*%t(J)%*%(qsi_2)
betas_corrigidos <- betas - vicio

## Gráfico de Envelope
library(statmod)
## Número de simulações do resíduo para construir o Intervalo de Confiança para o modelo
n_simulacoes <- 1000
## Matriz que armazenará os resíduos simulados com a distribuição postulada
residuo_simulado <- matrix(0,n_linhas,n_simulacoes)
## Utiliza a média ajustada do modelo para verificar a qualidade do ajuste

```

```

media_ajustada <- predict(fit.model,type="response")
for(i in 1:n_simulacoes){
##Simula a resposta com distribuição Normal , média como sendo a média ajustada pelo modelo
## e parâmetro de dispersão estimado (Passo 1 do algoritmo)
resposta_simulada <- rnorm(n_linhas, media_ajustada ,phi)
## Ajusta o modelo postulado para a variável simulada (Passo 2 do algoritmo)
ajuste <- glm(resposta_simulada ~ Matriz_Projeto, family=gaussian(link="identity"))
pesos_w <- ajuste$weights
Matriz_Pesos_W <- diag(pesos_w)
#  $W^{(.5)} X (X' W X)^{-1} X' W X' W^{(.5)}$ 
H <- solve(t(Matriz_Projeto)%*%Matriz_Pesos_W%*%Matriz_Projeto)
H <- sqrt(Matriz_Pesos_W)%*%Matriz_Projeto%*%H%*%t(Matriz_Projeto)%*%sqrt(Matriz_Pesos_W)
h <- diag(H)
phi_sim <- 1/summary(ajuste)$dispersion
## Calcular o vetor de resíduos studentizados para cada simulação (Passo 3 e 4 do algoritmo)
residuo_simulado[,i] <- sort(resid(ajuste,type="deviance")*sqrt(phi_sim/(1-h))))}
## Passo 5 do algoritmo
## Pegar os valores "mínimos" e "máximos" dos resíduos studentizados simulados
## (são utilizados o mínimo e o P1 e P2 e máximo P99 e P98) para suavizar as bandas de confiança
e_100 <- apply(residuo_simulado,1,quantile,1)
e_99 <- apply(residuo_simulado,1,quantile,.99)
e_98 <- apply(residuo_simulado,1,quantile,.98)
e_00 <- apply(residuo_simulado,1,quantile,.00)
e_01 <- apply(residuo_simulado,1,quantile,.01)
e_02 <- apply(residuo_simulado,1,quantile,.02)
## Constrói a faixa como uma média
## Passo 6 do algoritmo
faixa_max<- (e_100+e_99+e_98)/3
faixa_min<- (e_00+e_01+e_02)/3
media_residuos <- apply(residuo_simulado,1,mean)
faixa <- range(td,e_00,e_100)
## Constrói o gráfico de envelope e coloca os limites estimados através de simulação
par(mfrow=c(1,1))
par(pty="s")
qqnorm(td,xlab="Percentis da N(0,1)",ylab="Componente do Desvio", ylim=faixa, pch=16)

```

```
#identify(qqnorm(td,xlab="Percentis da N(0,1)",
#ylab="Componente do Desvio", ylim=faixa, pch=16))
par(new=T)
qqnorm(faixa_min,axes=F,xlab="",ylab="",type="l",ylim=faixa,lty=1)
par(new=T)
qqnorm(faixa_max,axes=F,xlab="",ylab="", type="l",ylim=faixa,lty=1)
par(new=T)
qqnorm(media_residuos,axes=F,xlab="", ylab="", type="l",ylim=faixa,lty=2)
```

A.4 Sobrevivência de Pacientes com Leucemia

ENTRADAS:

- beta1: Valor Inicial para estimação do primeiro parâmetro do modelo logístico
- beta2: Valor Inicial para estimação do segundo parâmetro do modelo logístico
- beta3: Valor Inicial para estimação do terceiro parâmetro do modelo logístico
- lambda: Valor Inicial para estimação do quarto parâmetro do modelo logístico
- n: número de iterações do algoritmo de estimação

SAÍDAS:

- betas: vetor de parâmetros estimados (sem o parâmetro de dispersão)
- phi: parâmetro de dispersão estimado pelo modelo.

```
library(MASS)
```

```
caminho_dados<- "C:\\Documents and Settings\\Wxp\\Meus documentos\\dri\\dissertacao\\leucemia"
```

```
setwd(caminho_dados)
```

```
dados <- read.csv("dados_leucemia.csv",sep=";",header=T)
```

Variável explicativa do modelo de pacientes com leucemia : Presença de certa característica Morfológica: AG (negativo ou positivo).

```
x1 <- dados$AG
```

Variável explicativa do modelo de pacientes com leucemia : Contagem de células brancas no sangue: WBC

```
x2 <- dados$WBC
```

Variável resposta do modelo de pacientes com leucemia : Vivo ou morto

```
y <- dados$Morte.Leucemia
```

Valores iniciais para os parâmetros (entrada do modelo)

```
beta1 <- 1
```

```
beta2 <- 1
```



```

beta3 <- 1
lambda <- 1
betas <- c(beta1, beta2, beta3, lambda )
grid_lambda <- seq(-5,5,0.05)
## Número de iterações do algoritmo de estimação
n <- 1000
## ESTIMAÇÃO DE PARÂMETROS
## Ajuste do modelo via fórmula paramétrica
(o software não consegue resolver o problema de estimação)
for (j in 1:60){
  ## Matriz de Jacobiano do modelo (Derivada de Eta com relação à Beta)
  J1 <- rep(1,dim(dados)[1])
  J2 <- x1
  J3 <- x2^lambda
  J4 <- beta3*x2^lambda*log(x2)
  J <- cbind(J1, J2, J3, J4)
  Eta <- beta1 + beta2*x1 + beta3*x2^lambda
  tal <- Eta - J%*betas
  u <- exp(Eta)/(1+ exp(Eta))
  W <- diag(1/(u^3*(1-u)^3))
  D <- diag(u*(1-u))
  y_1 <- Eta - tal + D%*(y - u)
  betas <- solve(t(J)%*W%*J)%*t(J)%*W%*y_1
  beta1 <- betas[1]
  beta2 <- betas[2]
  beta3 <- betas[3]
  lambda <- betas[4]}
## Ajuste do modelo baseado numa grade de valores para lambda
library(fUtilities)
matriz_betas <- array(0,c(length(grid_lambda),4))
matriz_erro_pad_betas <- array(0,c(length(grid_lambda),4))
aic <- numeric(0)
dev <- numeric(0)
beta1 <- 1
beta2 <- 1

```

```

beta3 <- 1
lambda <- 1
betas <- c(beta1, beta2, beta3)
for (i in grid_lambda){
  for (j in 1:60){
    ## Matriz de Jacobiano do modelo (Derivada de Eta com relação à Beta)
    J1 <- rep(1,dim(dados)[1])
    J2 <- x1
    if (i!=0){
      J3 <- x2^i
    } else {
      J3 <- log(x2)}
    J <- cbind(J1, J2, J3)
    Eta <- beta1 + beta2*x1 + beta3*J3
    tal <- Eta - J%*%betas
    fit.model<- glm(y ~ J1 + J2 + J3 - 1 + offset(tal),family=binomial(link = "logit"))
    betas <- coef(fit.model)
    beta1 <- betas[1]
    beta2 <- betas[2]
    beta3 <- betas[3]}
    matriz_betas[seq(1,length(grid_lambda))[grid_lambda==i],]<- c(betas,i)
    matriz_erro_pad_betas[seq(1,length(grid_lambda))[grid_lambda==i],]<-
    c(summary(fit.model)$coefficients[,2],0)
    aic <- c(aic, summary(fit.model)$aic)
    dev <- c(dev, summary(fit.model)$deviance)}
    inicio<- 1
    fim <- sum(x1)
    x2_ord <- x2[inicio:fim][rank(x2[inicio:fim])]
    y_ord <- y[inicio:fim][rank(x2[inicio:fim])]
    inicio<- (sum(x1)+1)
    fim <- length(x1)
    x2_ord <- c(x2_ord,x2[inicio:fim][rank(x2[inicio:fim])])
    y_ord <- c(y_ord, y[inicio:fim][rank(x2[inicio:fim])])
    eta_medio_estimado<- array(0,c(length(x1),dim(matriz_betas)[1]))
    for (i in 1:dim(matriz_betas)[1]){

```

```

if (matriz_betas[i,4]==0) {
a<- matriz_betas[i,1] + matriz_betas[i,2]*x1 + matriz_betas[i,3]*log(x2_ord)
} else {
a<- matriz_betas[i,1] + matriz_betas[i,2]*x1 + matriz_betas[i,3]*x2_ord^matriz_betas[i,4]}
eta_medio_estimado[,i] <- exp(a)/(1+exp(a))}
#grafico dos valores de lambda versus AIC
plot(grid_lambda,aic,main="",xlab="Valores de Lambda", ylab="AIC",t="l", las=1)
#abline(v=seq(min(grid_lambda),max(grid_lambda),length.out =15) ,col=gray(0.8))
#abline(h=seq(min(aic),max(aic),length.out =15) ,col=gray(0.8))
lines(grid_lambda,aic,main="",xlab="Valores de Lambda", ylab="AIC",t="l")
abline(v=-0.3, lty=2)
mtext(expression(paste(lambda," = - 0.3")),line=-4, adj=0, at=-1.5)
#mtext(expression(paste(hat(lambda)," = - 0.3")),line=-4, adj=0, at=-1.5)
## RESUMO DO AJUSTE DO GLM
beta1 <- 1
beta2 <- 1
beta3 <- 1
betas <- c(beta1, beta2, beta3)
### Fixa o valor de lambda final para que seja feita análise de diagnóstico do modelo
i<- 0#grid_lambda[aic==min(aic)] #com i = 0, estima modelo com log(WBC)
for (j in 1:60){
## Matriz de Jacobiano do modelo (Derivada de Eta com relação à Beta)
J1 <- rep(1,dim(dados)[1])
J2 <- x1
if (i!=0){
J3 <- x2^i
} else {
J3 <- log(x2)}
J <- cbind(J1, J2, J3)
Eta <- beta1 + beta2*x1 + beta3*J3
tal <- Eta - J%*%betas
fit.model<- glm(y ~ J1 + J2 + J3 - 1 + offset(tal),family=binomial(link = "logit"))
betas <- coef(fit.model)
beta1 <- betas[1]
beta2 <- betas[2]

```

```

beta3 <- betas[3]}

## Parâmetro de Dispersão
phi <- 1/summary(fit.model)$dispersion
summary(fit.model)

## Cálculo do Viés
lambda <- i
tam_n <- length(y)
Eta <- beta1 + beta2*x1 + beta3*x2^lambda
u <- exp(Eta)/(1+ exp(Eta))
W <- diag(1/(u^3*(1-u)^3))
aux <- solve(t(J)%*%W)%*%J
Zd <- array(0,c(tam_n,tam_n))
for (l in 1:tam_n){
  Zd[l,l] <- t(J[l,])%*%aux%*%J[l,]}
F <- array(0,c(tam_n,tam_n))
for (l in 1:tam_n){
  Vl <- u[l]*(1-u[l])
  dmu_deta <- exp(Eta[l])/(1+exp(Eta[l])) - exp(2*Eta[l])/((1+exp(Eta[l]))^2)
  d2mu_deta2 <- dmu_deta - 2*exp(2*Eta[l])/((1+exp(Eta[l]))^2) + 2*exp(3*Eta[l])/((1+exp(Eta[l]))^3)
  F[l,l] <- 1/Vl*dmu_deta*d2mu_deta2}
qsi_1 <- -(2*phi)^(-1)*Zd%*%solve(W)%*%F%*%rep(1,tam_n)
p<-3
Jl <- array(0,c(p,p))
D <- diag(1,tam_n)
for (l in 1:tam_n){
  aux1 <- Jl%*%aux
  D[l,l] <- sum(diag(aux1))}
qsi_2 <- -(2*phi)^(-1)*D%*%rep(1,tam_n)
vicio <- aux%*%t(J)%*%W)%*(qsi_1 + qsi_2)
betas_corrigidos <- betas - vicio

## Gráfico de Envelope
library(statmod)

## Número de simulações do resíduo para construir o Intervalo de Confiança para o modelo

```

```

n_simulacoes <- 1000

## Matriz que armazenará os resíduos simulados com a distribuição postulada
residuo_simulado <- matrix(0,n_linhas,n_simulacoes)

## Utiliza a média ajustada do modelo para verificar a qualidade do ajuste
media_ajustada <- predict(fit.modelo,type="response")

for(i in 1:n_simulacoes){

##Simula a resposta com distribuição Normal inversa, média como sendo a média ajustada pelo modelo
## e parâmetro de dispersão estimado (Passo 1 do algoritmo)
resposta_simulada <- rbinom(n_linhas, 1, prob = media_ajustada)

## Ajusta o modelo postulado para a variável simulada (Passo 2 do algoritmo)
ajuste <- glm(resposta_simulada ~ Matriz_Projeto, family=binomial(link = "logit"))

pesos_w <- ajuste$weights
Matriz_Pesos_W <- diag(pesos_w)

#  $W^{(.5)} X (X' W X)^{-1} X' W X' W^{(.5)}$ 
H <- solve(t(Matriz_Projeto)%*%Matriz_Pesos_W)%*%Matriz_Projeto)
H <- sqrt(Matriz_Pesos_W)%*%Matriz_Projeto)%*%H)%*%t(Matriz_Projeto)%*%sqrt(Matriz_Pesos_W)
h <- diag(H)

phi_sim <- 1/summary(ajuste)$dispersion

## Calcular o vetor de resíduos studentizados para cada simulação
## (Passo 3 e 4 do algoritmo)
residuo_simulado[,i] <- sort(resid(ajuste,type="deviance")*sqrt(phi_sim/(1-h)))}

## Passo 5 do algoritmo

## Pegar os valores "mínimos" e "máximos" dos resíduos studentizados simulados
## (são utilizados o mínimo e o P1 e P2 e máximo P99 e P98) para suavizar as bandas de confiança
e_100 <- apply(residuo_simulado,1,quantile,1)
e_99 <- apply(residuo_simulado,1,quantile,.99)
e_98 <- apply(residuo_simulado,1,quantile,.98)
e_00 <- apply(residuo_simulado,1,quantile,.00)
e_01 <- apply(residuo_simulado,1,quantile,.01)
e_02 <- apply(residuo_simulado,1,quantile,.02)

## Constrói a faixa como uma média

## Passo 6 do algoritmo
faixa_max<- (e_100+e_99+e_98)/3
faixa_min<- (e_00+e_01+e_02)/3
media_residuos <- apply(residuo_simulado,1,mean)

```

```

faixa <- range(td,e_00,e_100)
## Constrói o gráfico de envelope e coloca os limites estimados através de simulação
par(mfrow=c(1,1))
par(pty="s")
qqnorm(td,xlab="Percentis da N(0,1)",ylab="Componente do Desvio", ylim=faixa, pch=16)
#identify(qqnorm(td,xlab="Percentis da N(0,1)",
#ylab="Componente do Desvio", ylim=faixa, pch=16))
par(new=T)
qqnorm(faixa_min,axes=F,xlab="",ylab="",type="l",ylim=faixa,lty=1)
par(new=T)
qqnorm(faixa_max,axes=F,xlab="",ylab="", type="l",ylim=faixa,lty=1)
par(new=T)
qqnorm(media_residuos,axes=F,xlab="", ylab="", type="l",ylim=faixa,lty=2)

```

A.5 Produção de Gramíneas

```

## ENTRADAS:
- beta0: Valor Inicial para estimação do primeiro parâmetro do modelo gama
- beta1: Valor Inicial para estimação do segundo parâmetro do modelo gama
- beta2: Valor Inicial para estimação do terceiro parâmetro do modelo gama
- beta3: Valor Inicial para estimação do quarto parâmetro do modelo gama
- alpha1: Valor Inicial para estimação do quinto parâmetro do modelo gama
- aplha2: Valor Inicial para estimação do sexto parâmetro do modelo gama
- alpha3: Valor Inicial para estimação do setimo parâmetro do modelo gama
- n: número de iterações do algoritmo de estimação
## SAÍDAS:
- betas: vetor de parâmetros estimados (sem o parâmetro de dispersão)
- phi: parâmetro de dispersão estimado pelo modelo.
library(MASS)
caminho_dados<- "C:\\Documents and Settings\\Wxp\\Meus documentos\\dri\\dissertacao\\gramineas"
setwd(caminho_dados)
dados <- read.csv("dados_gramineas.csv",sep=";",header=T)
dados<- dados
## Variável explicativa do modelo de PRODUÇÃO DE GRAMÍNEAS
x1 <- dados[,1]

```

```

x2 <- dados[,2]
x3 <- dados[,3]
## Variável resposta do modelo de PRODUÇÃO DE GRAMÍNEAS
y <- dados[,4]
## Número de iterações do algoritmo de estimação
n <- 1000
## RESUMO DO AJUSTE DO GLM
beta0<- 1 #matriz_betas[dev==min(dev),1]
beta1<- 1#matriz_betas[dev==min(dev),2]
beta2<- 1#matriz_betas[dev==min(dev),3]
beta3<- 1#matriz_betas[dev==min(dev),3]
alpha1<- 10#matriz_betas[dev==min(dev),3]
alpha2<- 10#matriz_betas[dev==min(dev),3]
alpha3<- 10#matriz_betas[dev==min(dev),3]
betas <- c(beta0, beta1, beta2, beta3, alpha1, alpha2,alpha3)
matriz_betas1 <- array(1,c(300,7))
m<- 300
for (j in 1:m){
## Matriz de Jacobiano do modelo (Derivada de Eta com relação à Beta)
J1 <- rep(1,dim(dados)[1])
J2 <- 1/(x1+alpha1)
J3 <- 1/(x2+alpha2)
J4 <- 1/(x3+alpha3)
J5 <- -beta1/((x1+alpha1)^2)
J6 <- -beta2/((x2+alpha2)^2)
J7 <- -beta3/((x3+alpha3)^2)
J <- cbind(J1, J2, J3, J4, J5, J6, J7)
Eta <- beta0 + beta1/(x1+alpha1) + beta2/(x2+alpha2) + beta3/(x3+alpha3)
tal <- Eta - J%*%betas
W <- diag(1/Eta^2)
fit.model <- glm(y ~ J1 + J2 + J3 + J4 + J5 + J6 + J7 - 1 + offset(tal),
family=Gamma(link = "inverse"))#,subset=-c(36))
summary(fit.model)
betas <- coef(fit.model)
matriz_betas1[j,] <- betas

```

```

beta0<- coef(fit.model)[1]
beta1<- coef(fit.model)[2]
beta2<- coef(fit.model)[3]
beta3<- coef(fit.model)[4]
alpha1<- coef(fit.model)[5]
alpha2<- coef(fit.model)[6]
alpha3<- coef(fit.model)[7]}
summary(fit.model)
## Parâmetro de Dispersão
phi <- 1/summary(fit.model)$dispersion

## Cálculo do Viés
tam_n <- length(y)
aux <- solve(t(J)%*%W%*%J)
Zd <- array(0,c(tam_n,tam_n))
for (l in 1:tam_n){
Zd[l,1] <- t(J[l,])%*%aux%*%J[l,]}
F <- array(0,c(tam_n,tam_n))
for (l in 1:tam_n){
F[l,1] <- -2/(Eta[l]^3)}
qsi_1 <- -(2*phi)^(-1)*Zd%*%solve(W)%*%F%*%rep(1,tam_n)
p<-7
Jl <- array(0,c(p,p))
D <- diag(1,tam_n)
for (l in 1:tam_n){
Jl[2,5] <- Jl[5,2] <- -1/((x1[l]+alpha1)^2)
Jl[3,6] <- Jl[6,3] <- -1/((x2[l]+alpha2)^2)
Jl[4,7] <- Jl[7,4] <- -1/((x3[l]+alpha3)^2)
Jl[5,5] <- 2*beta1/((x1[l]+alpha1)^3)
Jl[6,6] <- 2*beta2/((x2[l]+alpha2)^3)
Jl[7,7] <- 2*beta3/((x3[l]+alpha3)^3)
aux1 <- Jl%*%aux
D[l,1] <- sum(diag(aux1))}
qsi_2 <- -(2*phi)^(-1)*D%*%rep(1,tam_n)
vicio <- aux%*%t(J)%*%W%*%(qsi_1 + qsi_2)

```



```

betas_corrigidos <- betas - vicio

## Gráfico de Envelope
library(statmod)
## Número de simulações do resíduo para construir o Intervalo de Confiança para o modelo
n_simulacoes <- 1000
## Matriz que armazenará os resíduos simulados com a distribuição postulada
residuo_simulado <- matrix(0,n_linhas,n_simulacoes)
## Utiliza a média ajustada do modelo para verificar a qualidade do ajuste
media_ajustada <- predict(fit.model,type="response")
for(i in 1:n_simulacoes){
##Simula a resposta com distribuição Normal inversa, média como sendo a média ajustada pelo modelo
## e parâmetro de dispersão estimado (Passo 1 do algoritmo)
resposta_simulada <- rgamma(n_linhas, media_ajustada ,phi)
## Ajusta o modelo postulado para a variável simulada (Passo 2 do algoritmo)
ajuste <- glm(resposta_simulada ~ Matriz_Projeto, family=Gamma(link=inverse))
pesos_w <- ajuste$weights
Matriz_Pesos_W <- diag(pesos_w)
#  $W^{(.5)} X (X' W X)^{-1} X' W X' W^{(.5)}$ 
H <- solve(t(Matriz_Projeto)%*%Matriz_Pesos_W%*%Matriz_Projeto)
H <- sqrt(Matriz_Pesos_W)%*%Matriz_Projeto%*%H%*%t(Matriz_Projeto)%*%sqrt(Matriz_Pesos_W)
h <- diag(H)
phi_sim <- 1/summary(ajuste)$dispersion
## Calcular o vetor de resíduos studentizados para cada simulação (Passo 3 e 4 do algoritmo)
residuo_simulado[,i] <- sort(resid(ajuste,type="deviance")*sqrt(phi_sim/(1-h))))}
## Passo 5 do algoritmo
## Pegar os valores "mínimos" e "máximos" dos resíduos studentizados simulados
## (são utilizados o mínimo e o P1 e P2 e máximo P99 e P98) para suavizar as bandas de confiança
e_100 <- apply(residuo_simulado,1,quantile,1)
e_99 <- apply(residuo_simulado,1,quantile,.99)
e_98 <- apply(residuo_simulado,1,quantile,.98)
e_00 <- apply(residuo_simulado,1,quantile,.00)
e_01 <- apply(residuo_simulado,1,quantile,.01)
e_02 <- apply(residuo_simulado,1,quantile,.02)
## Constrói a faixa como uma média

```

```
## Passo 6 do algoritmo
faixa_max<- (e_100+e_99+e_98)/3
faixa_min<- (e_00+e_01+e_02)/3
media_residuos <- apply(residuo_simulado,1,mean)
faixa <- range(td,e_00,e_100)
## Constrói o gráfico de envelope e coloca os limites estimados através de simulação
par(mfrow=c(1,1))
par(pty="s")
identify(qqnorm(td,xlab="Percentis da N(0,1)",ylab="Componente do Desvio", ylim=faixa, pch=16))
#identify(qqnorm(td,xlab="Percentis da N(0,1)",
#ylab="Componente do Desvio", ylim=faixa, pch=16))
par(new=T)
qqnorm(faixa_min,axes=F,xlab="",ylab="",type="l",ylim=faixa,lty=1)
par(new=T)
qqnorm(faixa_max,axes=F,xlab="",ylab="", type="l",ylim=faixa,lty=1)
par(new=T)
qqnorm(media_residuos,axes=F,xlab="", ylab="", type="l",ylim=faixa,lty=2)
```

Referências Bibliográficas

- Akaike, H. (1974). A new look at statistical model identification, *IEEE Transactions on Automatic Control* **AU-19**: 716–722.
- Andrade, F. C. (2004). *Pontos de Alavanca em Regressão*, Dissertação de Mestrado - IME-USP.
- Anscombe, F. J. (1953). Contribution to the discussion of H. Hotelling's paper, *Journal of the Royal Statistical Society B* **15**: 229–230.
- 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.
- Barros, M., Paula, G. A. & Leiva, V. (2008). A new class of survival regression models with heavy-tailed errors: robustness and diagnostics, *Lifetime Data Analysis* **14**: 316–332.
- Beckman, R. J., Nachtsheim, C. J. & Cook, R. D. (1987). Diagnostics for mixed-model analysis of variance, *Technometrics* **29**: 413–426.
- Billor, N. & Loynes, R. M. (1993). Local influence: a new approach, *Communications in Statistics, Theory and Methods* **22**: 1595–1611.
- Box, M. J. (1971). Bias in nonlinear estimation (with discussion), *Journal of the Royal Statistical Society B* **32**: 171–201.

- Breslow, N. E. & Day, N. E. (1987). *Statistical Methods in Cancer Research*, Vol. 2, IARC Scientific Publication.
- Breslow, N. E. & Day, N. E. (1993). Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association* **88**: 9–25.
- Cook, R. D. (1977). Detection of influential observation in linear regression, *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., Peña, D. & Weisberg, S. (1988). The likelihood displacement: A unifying principle for influence measures, *Communications in Statistics, Theory and Methods* **17**: 623–640.
- Cook, R. D. & Tsai, C. L. (1985). Residuals in nonlinear regression, *Biometrika* **72**: 23–29.
- Cook, R. D. & Tsai, C. L. (1990). Diagnostics for assessing the accuracy of normal approximations in exponential family nonlinear models, *Journal of American Statistical Association* **85**: 770–777.
- Cook, R. D. & Weisberg, S. (1982). *Residuals and Influence in Regression*, Chapman and Hall.
- Cordeiro, G. M. (1983). Improved likelihood ratio test for generalized linear models, *Journal of the Royal Statistical Society B* **45**: 404–413.
- Cordeiro, G. M. (1986). *Modelos Lineares Generalizados*, VII Simpósio Nacional de Probabilidade e Estatística.
- Cordeiro, G. M. (1987). On the corrections to the likelihood ratio statistics, *Biometrika* **74**: 265–274.
- Cordeiro, G. M. & McCullagh, P. (1991). Bias correction in generalized linear models, *Journal of the Royal Statistical Society B* **53**: 629–643.
- Cordeiro, G. M. & Paula, G. A. (1989). Improved likelihood ratio statistics for exponential family nonlinear models, *Biometrika* **76**: 93–100.

- Cordeiro, G. M. & Paula, G. A. (1992). Estimation, large-sample parametric tests and diagnostics for non-exponential family nonlinear models, *Communications in Statistics - Simulation and Computation* **21**: 149–172.
- Cox, D. R. & Hinkley, D. V. (1974). *Theoretical Statistics*, Chapman and Hall.
- Cox, D. R. & Snell, E. J. (1968). A general definition of residuals (with discussion), *Journal of the Royal Statistical Society B* **30**: 248–275.
- Davison, A. C. & Gigli, A. (1989). Deviance residuals and normal scores plots, *Biometrika* **76**: 211–221.
- Davison, A. C. & Tsai, C.-L. (1992). Regression model diagnostics, *International Statistical Review* **60**: 337–353.
- Diggle, P. J. (1990). *Times Series. A Biostatistical Introduction*, Oxford University Press.
- Dobson, A. J. (1990). *An Introduction to Generalized Linear Models*, Chapman and Hall.
- Dudzinski, M. L. & Mykytowycz, R. (1961). The eye lens as an indicator of age in the wild rabbit in Australia, *CSIRO WILDL.* **6**: 156–159.
- Emerson, J. D., Hoaglin, D. C. & Kempthorne, P. J. (1984). Leverage in least squares additive-plus-multiplicate fits for two-way tables, *Journal of the American Statistical Association* **79**: 329–335.
- Escobar, L. A. & Meeker, W. Q. (1992). Assessing influence in regression analysis with censored data, *Biometrics* **48**: 507–528.
- Fahrmeir, L. & Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models, *Annals of Statistics* **13**: 342–368.
- Fahrmeir, L. & Tutz, G. (1994). *Multivariate Statistical Modelling based on Generalized Linear Models*, Springer.
- Fung, W. K. (1993). Unmasking outliers and leverage points: A confirmation, *Journal of American Statistical Association* **88**: 515–519.

- Galea, M., Paula, G. A. & Bolfarine, H. (1997). Local influence in elliptical linear regression models, *The Statistician* **46**: 71–79.
- Galea, M., Riquelme, M. & Paula, G. A. (2000). Diagnostic methods in elliptical linear regression models, *Brazilian Journal of Probability and Statistics* **14**: 167–184.
- Galea, M., Uribe-Opazo, M. & Paula, G. A. (2003). On influence diagnostic in univariate elliptical linear regression models, *Statistical Papers* **44**: 23–45.
- Gu, H. & Fung, W. K. (1998). Assessing local influence in canonical correlation analysis, *Annals of the Institute of Statistical Mathematics* **50**: 755–772.
- Hewlett, P. S. (1969). Measurement of the potencies of drug mixtures, *Biometrics* **25**: 477–487.
- Hoaglin, D. C. & Welsch, R. E. (1978). The hat matrix in regression and anova, *The American Statistician* **32**: 17–22.
- Huet, S., Bouvier, A., Gruet, M. A. & Jolivet, E. (1996). *Statistical Tools for Nonlinear Regression*, Springer.
- Jørgensen, B. (1983). Maximum likelihood estimation and large-sample inference for generalized linear and nonlinear regression models, *Biometrika* **70**: 19–28.
- Jørgensen, B. (1987). Exponential dispersion models(with discussion), *Journal of the Royal Statistical Society B* **49**: 127–162.
- Kendall, M. G. & Stuart, A. (1977). *The Advanced Theory of Statistics*, 4th edn, C. Griffin.
- Kim, M. G. (1995). Local influence in multivariate regression, *Communications in Statistics, Theory and Methods* **20**: 1271–1278.
- Kosmidis, I. (2007). *Bias Reduction in Exponential Family Nonlinear Models*, Thesis - The University of Warwick.
- Kwan, C. W. & Fung, W. K. (1998). Assessing local influence for specific restricted likelihood: Applications to factor analysis, *Psychometrika* **63**: 35–46.

- Lawrence, A. J. (1988). Regression transformation diagnostics using local influence, *Journal of American Statistical Association* **84**: 125–141.
- Lee, A. H. (1987). Diagnostics displays for assessing leverage and influence in generalized linear models, *Australian Journal Statistics* **29**: 233–243.
- Lee, A. H. (1988). Assessing partial influence in generalized linear models, *Biometrics* **44**: 71–77.
- Lee, S. Y. & Wang, S. J. (1996). Sensitivity analysis of structural equation models, *Psychometrika* **61**: 93–108.
- Lee, Y. & Nelder, J. A. (1996). Hierarchical generalized linear models, *Journal of the Royal Statistical Society B* **58**: 619–678.
- Lee, Y. & Nelder, J. A. (2001). Hierarchical generalized linear models: a synthesis of generalized linear models, random-effect models and structured dispersion, *Biometrika* **88**: 987–1006.
- Leiva, V., Barros, M., Paula, G. A. & Galea, M. (2007). Influence diagnostic in log-Birnbaum-Saunders regression models with censored data, *Computational Statistics and Data Analysis* **51**: 5694–5707.
- Lesaffre, E. & Verbeke, G. (1998). Local influence in linear mixed models, *Biometrics* **54**: 570–582.
- Liang, K. Y. & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika* **73**: 13–22.
- Liu, S. Z. (2000). On local influence for elliptical linear models, *Statistical Papers* **41**: 211–224.
- McCullagh, P. (1983). Quasi-likelihood functions, *Annals of Statistics* **11**: 59–67.
- McCullagh, P. (1987). *Tensor Methods in Statistical*, Chapman and Hall.
- McCullagh, P. & Nelder, J. A. (1983). *Generalized Linear Models*, Chapman and Hall.
- McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models*, 2nd edn, Chapman and Hall.

- McCulloch, C. E. & Searle, S. R. (2001). *Linear and Generalized Linear Mixed Models*, Wiley.
- Moolgavkar, S. H., Lustbader, E. D. & Venzon, D. J. (1984). A geometric approach to non-linear regression diagnostics with application to matched case-control studies, *Annals of Statistics* **12**: 816–826.
- Nelder, J. A. & Wedderburn, R. W. M. (1972). Generalized linear models, *Journal of the Royal Statistical Society A* **135**: 370–384.
- O'Hara Hines, R. J., Lawless, J. F. & Carter, E. M. (1992). Diagnostics for a cumulative multinomial generalized linear model with application to grouped toxicological mortality data, *Journal of American Statistical Association* **87**: 1059–1069.
- Ortega, E. M. M., Bolfarine, H. & Paula, G. A. (2003). Influence diagnostic in generalized log-gamma regression models, *Computational Statistics and Data Analysis* **42**: 165–186.
- Ortega, E. M. M., Paula, G. A. & Bolfarine, H. (2008). Deviance residuals in generalised log-gamma regression models with censored observations, *Journal of Statistical Computation and Simulation* **78**: 747–764.
- Osorio, F., Paula, G. A. & Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure, *Computational Statistics and Data Analysis Amsterdam* **51**: 4354–4368.
- Pan, J. X., Fang, K. T. & von Rosen (1997). Local influence assessment in the growth curve model with unstructured covariance, *Journal of Statistical Planning and Inference* **62**: 263–278.
- Paula, G. A. (1988). *Correções de Bartlett e Métodos de Diagnóstico para Modelos Não Lineares de Família Exponencial*, Tese de Doutorado - IME-USP.
- Paula, G. A. (1992). Bias correction for exponential family nonlinear models, *Journal of Statistical Computation and Simulation* **40**: 43–54.
- 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* **40**: 43–54.
- Paula, G. A. (1996). Influence diagnostic in proper dispersion models, *Australian Journal of Statistics* **38**: 307–316.
- 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*, IME-USP.
- Paula, G. A. & Cordeiro, G. M. (1986). Alguns modelos não lineares via o glim, *Atas do VII Simpósio Nacional de Probabilidades e Estatística* pp. 204–217.
- Peña, D. & Yohai, V. (1999). A fast procedure for outlier diagnostics in large regression problems, *Journal of American Statistical Association* **94**: 434–445.
- Pettitt, A. N. & Bin Daud, I. (1989). Case-weight measures of influence for proportional hazards regression, *Applied Statistics* **38**: 51–67.
- Pregibon, D. (1979). *Data Analytic Methods for Generalized Linear Models*, Thesis PhD - University of Toronto.
- Pregibon, D. (1981). Logistic regression diagnostics, *Annals of Statistics* **9**: 705–724.
- Rancel, M. M. S. & Sierra, M. A. G. (2001). Regression diagnostic using local influence: a review, *Communications in Statistics, Theory and Methods* **30**: 799–813.
- Ratkowsky, D. A. (1983). *Nonlinear Regression Modelling*, Marcel Dekker.
- Rawlings, J. O., Pantula, S. G. & Dickey, D. A. (1998). *Applied Regression Analysis : A Research Tool*, 2nd edn, Springer.
- Ross, W. H. (1987). The geometry of case deletion and the assessment of influence in nonlinear regression, *Canadian Journal of Statistics* **15**: 91–103.

- Saint Laurent, R. T. & Cook, R. D. (1992). Leverage and superleverage in nonlinear regression, *Journal of the American Statistical Association* **87**: 985–990.
- Schall, R. & Dunne, T. T. (1992). A note on the relationship between parameter collinearity and local influence, *Biometrika* **79**: 399–404.
- Searle, S. R. (1982). *Matrix Algebra Useful for Statistics*, Wiley.
- Seber, G. A. F. & Wild, C. J. (1989). *Nonlinear Regression*, Wiley.
- Sen, P. K. & Singer, J. M. (1993). *Large Sample Methods in Statistics: An Introduction with Applications*, Chapman and Hall.
- Souza, F. A. M. & Paula, G. A. (2002). Deviance residuals for an angular response, *Australian and New Zealand Journal of Statistics* **44**: 345–356.
- Storer, B. E. & Crowley, J. (1985). A diagnostic for Cox regression and general conditional likelihoods, *Journal of the American Statistical Association* **80**: 139–147.
- Svetliza, C. F. (2002). *Modelos Não Lineares com Resposta Binomial Negativa*, Tese de Doutorado - IME-USP.
- Svetliza, C. F. & Paula, G. A. (2003). Diagnostics in nonlinear negative binomial models, *Communications in Statistics, Theory and Methods* **32**: 1227–1250.
- Tang, N. S., Wei, B. C. & Wang, X. R. (2001). Local influence in nonlinear reproductive dispersion models, *Communications in Statistics, Theory and Methods* **30**: 435–449.
- Tang, N. S., Wei, B. C. & Zhang, W. Z. (2006a). Influence diagnostics in nonlinear reproductive dispersion mixed models, *Statistics* **40**: 227–246.
- Tang, N. S., Wei, B. C. & Zhang, W. Z. (2006b). Influence diagnostics in nonlinear reproductive dispersion mixed models, *Statistics* **40**: 227–246.
- Therneau, T. M., Grambsch, P. M. & Fleming, T. R. (1990). Martingale based residuals for survival models, *Biometrika* **77**: 147–160.

- Thomas, W. & Cook, R. D. (1990). Assessing influence on predictions from generalized linear models, *Technometrics* **32**: 59–65.
- Tsai, C. H. & Wu, X. (1992). Assessing local influence in linear regression models with first-order autoregressive or heteroscedastic error structure, *Statistics and Probability Letters* **14**: 247–252.
- Venables, W. N. & Ripley, B. D. (1999). *Modern Applied Statistics with S-Plus*, 3rd edn, Springer.
- Wei, B. C. (1998). *Exponential Family Nonlinear Models: Lecture Notes in Statistics*, Vol. 130, Springer.
- Wei, B. C., Hu, Y. Q. & Fung, W. K. (1998). Generalized leverage and its applications, *Scandinavian Journal of Statistics* **25**: 25–37.
- Wei, B. C. & Shi, J. Q. (1994). On statistical models in regression diagnostics, *Annals of Mathematical Statistics* **46**: 267–278.
- Welch, L. F., Adams, W. E. & Corman, J. L. (1963). Yield response surfaces, isoquants and economic fertilizer optima for coastal bermuda grass, *Agron. J.* **55**: 63–67.
- Whitmore, G. A. (1986). Inverse gaussian ratio estimation, *Applied Statistics* **35**: 8–15.
- Williams, D. A. (1984). Residuals in generalized linear models, *In: Proceedings of the 12th. International Biometrics Conference* pp. 59–68.
- Williams, D. A. (1987). Generalized linear model diagnostic using the deviance and single case deletion, *Applied Statistics* **36**: 181–191.