

# **Modelos Lineares Generalizados e Extensões**

**Gauss Moutinho Cordeiro**

Departamento de Estatística

CCEN/UFPE

50740-540 Recife, PE

Email: `gauss@de.ufpe.br`

**Clarice G.B. Demétrio**

Departamento de Ciências Exatas, ESALQ, USP

13418-900 Piracicaba, SP

Email: `clarice.demetrio@usp.br`

**Rafael de A. Moral**

Departamento de Ciências Exatas, ESALQ, USP

13418-900 Piracicaba, SP

Email: `rafael.moral@usp.br`

14 de setembro de 2016

## **Prefácio**

Este livro é resultante de vários anos de ensino de cursos e minicursos sobre modelos lineares generalizados e tem como objetivo apresentar noções gerais desses modelos, algumas de suas extensões e aplicações. Enumerar as pessoas a quem devemos agradecimentos é uma tarefa difícil, pois são muitos aqueles que contribuíram de forma direta ou indireta para a elaboração deste material. Agradecemos a Eduardo Bonilha, funcionário do Departamento de Ciências Exatas da ESALQ/USP, o auxílio na digitação, e a todos que nos ajudaram lendo versões anteriores, cuidadosamente, e dando sugestões muito proveitosas. Agradecemos, também, ao CNPq, à CAPES e à FAPESP por financiamentos de projetos que trouxeram contribuições importantes para a elaboração deste livro.

Finalmente, assumimos total responsabilidade pelas imperfeições e solicitamos aos leitores que nos apresentem críticas e sugestões para uma futura edição revisada.

Gauss Moutinho Cordeiro

Clarice Garcia Borges Demétrio

Rafael de Andrade Moral

Piracicaba, setembro de 2016

# Sumário

<b>1</b>	<b>Família exponencial de distribuições</b>	<b>1</b>
1.1	Introdução . . . . .	1
1.2	Família exponencial uniparamétrica . . . . .	2
1.3	Componente aleatório . . . . .	4
1.4	Função geradora de momentos . . . . .	8
1.5	Estatística suficiente . . . . .	12
1.6	Família exponencial multiparamétrica . . . . .	13
1.7	Exercícios . . . . .	16
<b>2</b>	<b>Modelo Linear Generalizado</b>	<b>23</b>
2.1	Introdução . . . . .	23
2.2	Exemplos de motivação . . . . .	26
2.2.1	Ensaios do tipo dose-resposta . . . . .	26
2.2.2	Ensaios de diluição . . . . .	34
2.2.3	Tabelas de contingência . . . . .	37
2.3	Definição . . . . .	41
2.4	Modelos especiais . . . . .	45
2.4.1	Modelo clássico de regressão . . . . .	45
2.4.2	Modelo de Poisson . . . . .	47
2.4.3	Modelo binomial . . . . .	49
2.4.3.1	Dados na forma de proporções . . . . .	50
2.4.3.2	Dados binários agrupados . . . . .	52

2.4.4	Modelo gama . . . . .	53
2.4.5	Modelo normal inverso . . . . .	55
2.4.6	Modelo binomial negativo . . . . .	56
2.4.7	Modelo secante hiperbólico generalizado . . . . .	57
2.4.8	Modelos definidos por transformações . . . . .	58
2.5	Ajuste dos MLGs . . . . .	58
2.5.1	Formulação de modelos . . . . .	59
2.5.2	Ajuste dos modelos . . . . .	63
2.5.3	Inferência . . . . .	65
2.6	Exercícios . . . . .	66
<b>3</b>	<b>Estimação</b>	<b>71</b>
3.1	Estatísticas suficientes . . . . .	71
3.2	O algoritmo de estimação . . . . .	73
3.3	Estimação em modelos especiais . . . . .	79
3.4	Resultados adicionais na estimação . . . . .	81
3.5	Seleção do modelo . . . . .	84
3.6	Considerações sobre a função de verossimilhança . . . . .	87
3.7	Exercícios . . . . .	91
<b>4</b>	<b>Métodos de Inferência</b>	<b>95</b>
4.1	Distribuição dos estimadores dos parâmetros . . . . .	95
4.2	Função desvio e estatística de Pearson generalizada . . . . .	101
4.3	Análise de desvio e seleção de modelos . . . . .	111
4.4	Estimação do parâmetro de dispersão . . . . .	114
4.5	Comparação dos três métodos de estimação do parâmetro de dispersão no modelo gama . . . . .	117
4.6	Testes de hipóteses . . . . .	118
4.6.1	Teste de uma hipótese nula simples . . . . .	119
4.6.2	Teste de uma hipótese nula composta . . . . .	121

4.7	Regiões de confiança . . . . .	123
4.8	Seleção de variáveis explanatórias . . . . .	125
4.9	Exercícios . . . . .	127
<b>5</b>	<b>Resíduos e Diagnósticos</b>	<b>135</b>
5.1	Introdução . . . . .	135
5.2	Técnicas para verificar o ajuste de um modelo . . . . .	136
5.3	Análise de resíduos e diagnósticos para o modelo linear clássico . . . .	137
5.3.1	Tipos de resíduos . . . . .	138
5.3.2	Estatísticas para diagnóstico . . . . .	140
5.3.3	Tipos de gráficos . . . . .	144
5.4	Análise de resíduos e diagnóstico para os MLGs . . . . .	150
5.4.1	Tipos de resíduos . . . . .	151
5.4.2	Tipos de gráficos . . . . .	158
5.5	Método das variáveis explanatórias adicionais na seleção de modelos .	161
5.6	Verificação da função de ligação . . . . .	163
5.7	Verificação da função de variância . . . . .	168
5.8	Verificação das escalas das variáveis explanatórias . . . . .	168
5.9	Verificação de anomalias no componente sistemático, usando-se análise dos resíduos . . . . .	172
5.10	Exercícios . . . . .	175
<b>6</b>	<b>Aplicações a Dados Contínuos</b>	<b>181</b>
6.1	Dados de volume de árvores . . . . .	181
6.2	Dados de gordura no leite . . . . .	188
6.3	Dados de importação Brasileira . . . . .	189
6.4	Dados de tempos de sobrevivência de ratos . . . . .	196
6.5	Dados de assinaturas de TV a cabo . . . . .	198
6.6	Dados de demanda de energia elétrica . . . . .	203

<b>7</b>	<b>Aplicações a Dados Discretos</b>	<b>207</b>
7.1	Dados binários e proporções . . . . .	207
7.1.1	Estimação da dose efetiva e seu intervalo de confiança . . . . .	207
7.1.2	Probabilidade de resposta a uma dose especificada . . . . .	213
7.1.3	Paralelismo entre retas no modelo logístico linear e potência relativa . . . . .	214
7.2	Dados de contagem . . . . .	231
7.2.1	Modelo de Poisson . . . . .	231
7.2.2	Modelos log-lineares para tabelas de contingência . . . . .	238
7.3	Exercícios . . . . .	248

# Capítulo 1

## Família exponencial de distribuições

### 1.1 Introdução

Muitas das distribuições conhecidas podem ser colocadas em uma família paramétrica denominada **família exponencial de distribuições**. Assim, por exemplo, pertencem a essa família as distribuições normal, binomial, binomial negativa, gama, Poisson, normal inversa, multinomial, beta, logarítmica, entre outras. Essa classe de distribuições foi proposta, independentemente, por Koopman, Pitman e Darmois ao estudarem as propriedades de suficiência estatística. Posteriormente, muitos outros aspectos dessa família foram estudados e tornaram-se importantes na teoria moderna de Estatística. O conceito de família exponencial foi introduzido na Estatística por Fisher, mas os modelos da família exponencial surgiram na Mecânica Estatística no final do século XIX e foram desenvolvidos por Maxwell, Boltzmann e Gibbs. A importância dessa família teve maior destaque, na área dos modelos de regressão, a partir do trabalho pioneiro de Nelder e Wedderburn (1972) que definiram os modelos lineares generalizados (MLGs). Na década de 80, esses modelos popularizaram-se, inicialmente, no Reino Unido, e, posteriormente, nos Estados Unidos e por toda Europa ocidental.

## 1.2 Família exponencial uniparamétrica

A **família exponencial uniparamétrica** é caracterizada por uma função (de probabilidade ou densidade) especificada na forma

$$f(x; \theta) = h(x) \exp [\eta(\theta) t(x) - b(\theta)], \quad (1.1)$$

em que as funções  $\eta(\theta)$ ,  $b(\theta)$ ,  $t(x)$  e  $h(x)$  têm valores em subconjuntos dos reais. As funções  $\eta(\theta)$ ,  $b(\theta)$  e  $t(x)$  não são únicas. Por exemplo,  $\eta(\theta)$  pode ser multiplicada por uma constante  $k$  e  $t(x)$  pode ser dividida pela mesma constante. Pelo teorema da fatoração de Neyman-Fisher, a estatística  $t(X)$  é suficiente para  $\theta$ .

Várias distribuições importantes podem ser expressas na forma (1.1), tais como: Poisson, binomial, Rayleigh, normal, gama e normal inversa (as três últimas com a suposição de que um dos parâmetros é conhecido). Cordeiro et al. (1995) apresentaram 24 distribuições na forma (1.1). O suporte da família exponencial (1.1), isto é,  $\{x; f(x; \theta) > 0\}$ , não pode depender de  $\theta$ . Assim, a distribuição uniforme em  $(0, \theta)$  não é um modelo da família exponencial.

É fácil comprovar se uma distribuição pertence, ou não, à família exponencial (1.1), como é demonstrado nos três exemplos que se seguem.

**Exemplo 1.1:** A distribuição de Poisson  $P(\theta)$  de parâmetro  $\theta > 0$ , usada para análise de dados na forma de contagens, tem função de probabilidade

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

e, portanto, é um membro da família exponencial (1.1) com  $\eta(\theta) = \log(\theta)$ ,  $b(\theta) = \theta$ ,  $t(x) = x$  e  $h(x) = 1/x!$ .

**Exemplo 1.2:** A distribuição binomial  $B(m, \theta)$ , com  $0 < \theta < 1$  e  $m$ , o número conhecido de ensaios independentes, é usada para análise de dados na forma de proporções e tem função de probabilidade

$$f(x; \theta) = \binom{m}{x} \theta^x (1 - \theta)^{m-x} = \binom{m}{x} \exp \left[ x \log \left( \frac{\theta}{1 - \theta} \right) + m \log(1 - \theta) \right]$$



com  $\eta(\theta) = \log[\theta/(1 - \theta)]$ ,  $b(\theta) = -m \log(1 - \theta)$ ,  $t(x) = x$  e  $h(x) = \binom{m}{x}$ , sendo, portanto, um membro da família exponencial (1.1).

**Exemplo 1.3:** A distribuição de Rayleigh, usada para análise de dados contínuos positivos, tem função densidade ( $x > 0, \theta > 0$ )

$$f(x; \theta) = \frac{x}{\theta^2} \exp\left(-\frac{x^2}{2\theta^2}\right) = x \exp\left[-\frac{1}{2\theta^2}x^2 - 2\log(\theta)\right]$$

e, portanto, pertence à família exponencial (1.1) com  $\eta(\theta) = -1/(2\theta^2)$ ,  $b(\theta) = 2\log(\theta)$ ,  $t(x) = x^2$  e  $h(x) = x$ .

A família exponencial na forma canônica é definida por (1.1), considerando que as funções  $\eta(\theta)$  e  $t(x)$  são iguais à função identidade, de forma que

$$f(x; \theta) = h(x) \exp[\theta x - b(\theta)]. \quad (1.2)$$

Na parametrização (1.2),  $\theta$  é denominado de **parâmetro canônico**. O logaritmo da função de verossimilhança correspondente a uma única observação no modelo (1.2) é expresso como

$$\ell(\theta) = \theta x - b(\theta) + \log[h(x)]$$

e, portanto, a função escore  $U = U(\theta) = d\ell(\theta)/d\theta$  resulta em  $U = x - b'(\theta)$ .

É fácil verificar das propriedades da função escore,  $E(U) = 0$  e  $\text{Var}(U) = -E[d^2\ell(\theta)/d\theta^2] = -E(U')$  (a última igualdade é a informação de Fisher), que

$$E(X) = b'(\theta) \text{ e } \text{Var}(X) = b''(\theta). \quad (1.3)$$

O simples fato de se calcularem momentos da família exponencial (1.2) em termos de derivadas da função  $b(\theta)$  (denominada função geradora de cumulantes) em relação ao parâmetro canônico  $\theta$  é muito importante na teoria dos modelos lineares generalizados, principalmente, no contexto assintótico.

Suponha que  $X_1, \dots, X_n$  sejam  $n$  variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) seguindo (1.1). A distribuição conjunta de  $X_1, \dots, X_n$

é expressa por

$$f(x_1, \dots, x_n; \theta) = \left[ \prod_{i=1}^n h(x_i) \right] \exp \left[ \eta(\theta) \sum_{i=1}^n t(x_i) - nb(\theta) \right]. \quad (1.4)$$

A equação (1.4) implica que a distribuição conjunta de  $X_1, \dots, X_n$  é, também, um modelo da família exponencial. A estatística suficiente  $\sum_{i=1}^n T(X_i)$  tem dimensão um, qualquer que seja  $n$ .

Geralmente, a estatística suficiente de um modelo da família exponencial tem distribuição, também, pertencente à família exponencial. Por exemplo, se  $X_1, \dots, X_n$  são variáveis aleatórias i.i.d. com distribuição de Poisson  $P(\theta)$ , então a estatística suficiente  $\sum_{i=1}^n T(X_i)$  tem, também, distribuição de Poisson  $P(n\theta)$  e, assim, é um modelo exponencial uniparamétrico.

### 1.3 Componente aleatório

Como será descrito, na Seção 2.3, o componente aleatório de um MLG é definido a partir da família exponencial uniparamétrica na forma canônica (1.2) com um parâmetro extra de perturbação,  $\phi > 0$ , que é uma medida de dispersão da distribuição. Nelder e Wedderburn (1972) ao proporem essa modelagem, conseguiram incorporar distribuições biparamétricas no componente aleatório do modelo. Tem-se

$$f(y; \theta, \phi) = \exp \left\{ \phi^{-1} [y\theta - b(\theta)] + c(y, \phi) \right\}, \quad (1.5)$$

em que  $b(\cdot)$  e  $c(\cdot)$  são funções conhecidas. Quando  $\phi$  é conhecido, a família de distribuições (1.5) é idêntica à família exponencial na forma canônica (1.2). Essa família, também, é conhecida como família exponencial linear. Na Seção 1.4, será demonstrado que o valor esperado e a variância de  $Y$  com distribuição na família (1.5) são

$$E(Y) = \mu = b'(\theta) \text{ e } \text{Var}(Y) = \phi b''(\theta).$$

Observa-se, a partir da expressão da variância, que  $\phi$  é um **parâmetro de dispersão** do modelo e seu inverso  $\phi^{-1}$ , uma **medida de precisão**. A função que

relaciona o parâmetro canônico  $\theta$  com a média  $\mu$  é denotada por  $\theta = q(\mu)$  (inversa da função  $b'(\cdot)$ ). A função da média  $\mu$  na variância é representada por  $b''(\theta) = V(\mu)$ . Denomina-se  $V(\mu)$  de **função de variância**. Observa-se que o parâmetro canônico pode ser obtido de  $\theta = \int V^{-1}(\mu)d\mu$ , pois  $V(\mu) = d\mu/d\theta$ . A Tabela 1.1 apresenta várias distribuições importantes na família (1.5), caracterizando as funções  $b(\theta)$ ,  $c(y, \phi)$ , a média  $\mu(\theta)$  em termos do parâmetro canônico  $\theta$  e a função de variância  $V(\mu)$ . Nessa tabela,  $\Gamma(\cdot)$  é a função gama, isto é,  $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$ ,  $\alpha > 0$ . A família de distribuições (1.5) inclui distribuições que exibem assimetria e de natureza discreta ou contínua, com suportes que são restritos a intervalos do conjunto dos reais, conforme bem exemplificam as distribuições da Tabela 1.1. Essas distribuições serão estudadas no Capítulo 2.

Convém enfatizar que se  $\phi$  não for conhecido, a família (1.5) pode, ou não, pertencer à família exponencial biparamétrica (Seção 1.6). Para (1.5) pertencer à família exponencial biparamétrica quando  $\phi$  é desconhecido, a função  $c(y, \phi)$  deve ser decomposta, segundo Cordeiro e McCullagh (1991), como  $c(y, \phi) = \phi^{-1}d(y) + d_2(\phi) + d_1(y)$ . Esse é o caso das distribuições normal, gama e normal inversa.

Morris (1982) demonstrou que existem apenas seis distribuições na família (1.5) cuja função de variância é uma função, no máximo, quadrática da média. Essas distribuições são normal ( $V = 1$ ), gama ( $V = \mu^2$ ), binomial ( $V = \mu(1 - \mu)$ ), Poisson ( $V = \mu$ ) e binomial negativa ( $V = \mu + \mu^2/k$ ). A sexta, chamada secante hiperbólica generalizada ( $V = 1 + \mu^2$ ), tem função densidade igual a

$$f(y; \theta) = \frac{1}{2} \exp[\theta y + \log(\cos \theta)] \cosh\left(\frac{\pi y}{2}\right), \quad y \in \mathbb{R}, \quad \theta > 0. \quad (1.6)$$

A distribuição secante hiperbólica generalizada (1.6) compete com a distribuição normal na análise de observações contínuas irrestritas. A seguir, apresentam-se duas distribuições que pertencem à família (1.5).

**Exemplo 1.4:** A distribuição normal  $N(\mu, \sigma^2)$ , de média  $\mu \in \mathbb{R}$  e variância  $\sigma^2 > 0$ ,

Tabela 1.1: Algumas distribuições importantes na família (1.5).

Distribuição	$\phi$	$\theta$	$b(\theta)$	$c(y, \phi)$	$\mu(\theta)$	$V(\mu)$
Normal: $N(\mu, \sigma^2)$	$\sigma^2$	$\mu$	$\frac{\theta^2}{2}$	$-\frac{1}{2} \left[ \frac{y^2}{\sigma^2} + \log(2\pi\sigma^2) \right]$	$\theta$	1
Poisson: $P(\mu)$	1	$\log(\mu)$	$e^\theta$	$-\log(y!)$	$e^\theta$	$\mu$
Binomial: $B(m, \pi)$	1	$\log\left(\frac{\mu}{m - \mu}\right)$	$m \log(1 + e^\theta)$	$\log\left(\frac{m}{y}\right)$	$\frac{me^\theta}{1 + e^\theta}$	$\frac{\mu}{m}(m - \mu)$
Binomial negativa: $BN(\mu, k)$	1	$\log\left(\frac{\mu}{\mu + k}\right)$	$-k \log(1 - e^\theta)$	$\log\left[\frac{\Gamma(k + y)}{\Gamma(k)y!}\right]$	$\frac{ke^\theta}{1 - e^\theta}$	$\mu\left(\frac{\mu}{k} + 1\right)$
Gama: $G(\mu, \nu)$	$\nu^{-1}$	$-\frac{1}{\mu}$	$-\log(-\theta)$	$\nu \log(\nu y) - \log(y) - \log \Gamma(\nu)$	$-\frac{1}{\theta}$	$\mu^2$
Normal inversa: $IG(\mu, \sigma^2)$	$\sigma^2$	$-\frac{1}{2\mu^2}$	$-(-2\theta)^{1/2}$	$-\frac{1}{2} \left[ \log(2\pi\sigma^2 y^3) + \frac{1}{\sigma^2 y} \right]$	$(-2\theta)^{-1/2}$	$\mu^3$

tem função densidade de probabilidade (f.d.p.) expressa como

$$f(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(y - \mu)^2}{2\sigma^2} \right].$$

Tem-se, então,

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

obtendo-se os elementos da primeira linha da Tabela 1.1, isto é,

$$\theta = \mu, \quad \phi = \sigma^2, \quad b(\theta) = \frac{\mu^2}{2} = \frac{\theta^2}{2} \quad \text{e} \quad c(y, \phi) = -\frac{1}{2} \left[ \frac{y^2}{\sigma^2} + \log(2\pi\sigma^2) \right],$$

o que demonstra que a distribuição  $N(\mu, \sigma^2)$  pertence à família (1.5).

**Exemplo 1.5:** A distribuição binomial tem função de probabilidade

$$f(y; \pi) = \binom{m}{y} \pi^y (1 - \pi)^{m-y}, \quad \pi \in (0, 1), \quad y = 0, 1, \dots, m.$$

Tem-se, então,

$$\begin{aligned} f(y; \pi) &= \exp \left[ \log \binom{m}{y} + y \log(\pi) + (m - y) \log(1 - \pi) \right] \\ &= \exp \left[ y \log \left( \frac{\pi}{1 - \pi} \right) + m \log(1 - \pi) + \log \binom{m}{y} \right], \end{aligned}$$

obtendo-se os elementos da terceira linha da Tabela 1.1, isto é,

$$\begin{aligned} \phi = 1, \quad \theta = \log \left( \frac{\pi}{1 - \pi} \right) = \log \left( \frac{\mu}{m - \mu} \right), \quad \text{o que implica em} \\ \mu = \frac{me^\theta}{(1 + e^\theta)}, \quad b(\theta) = -m \log(1 - \pi) = m \log(1 + e^\theta) \quad \text{e} \quad c(y, \phi) = \log \binom{m}{y}. \end{aligned}$$

Portanto, a distribuição binomial pertence à família exponencial (1.5).

Outras distribuições importantes podem ser expressas na forma (1.5) como os modelos exponenciais de dispersão, descritos na Seção ??.

## 1.4 Função geradora de momentos

A função geradora de momentos (f.g.m.) da família (1.5) é igual a

$$M(t; \theta, \phi) = E(e^{tY}) = \exp \{ \phi^{-1} [b(\phi t + \theta) - b(\theta)] \}. \quad (1.7)$$

**Prova:** A prova será feita apenas para o caso de variáveis aleatórias contínuas. No caso discreto, basta substituir a integral pelo somatório. Sabe-se que

$$\int_A f(y; \theta, \phi) dy = 1,$$

e, portanto,

$$\int_A \exp \{ \phi^{-1} [\theta y - b(\theta)] + c(y, \phi) \} dy = 1,$$

obtendo-se

$$\int_A \exp [\phi^{-1} \theta y + c(y, \phi)] dy = \exp [\phi^{-1} b(\theta)]. \quad (1.8)$$

Logo,

$$\begin{aligned} M(t; \theta, \phi) &= E(e^{tY}) = \int_A \exp(ty) f(y) dy \\ &= \int_A \exp \{ \phi^{-1} [(\phi t + \theta)y - b(\theta)] + c(y, \phi) \} dy \\ &= \frac{1}{\exp [\phi^{-1} b(\theta)]} \int_A \exp [\phi^{-1} (\phi t + \theta)y + c(y, \phi)] dy \end{aligned}$$

e, usando-se a equação (1.8), tem-se

$$M(t; \theta, \phi) = \exp \{ \phi^{-1} [b(\phi t + \theta) - b(\theta)] \}.$$

A função geradora de cumulantes (f.g.c.) correspondente é, então,

$$\varphi(t; \theta, \phi) = \log[M(t; \theta, \phi)] = \phi^{-1} [b(\phi t + \theta) - b(\theta)]. \quad (1.9)$$

A f.g.c. desempenha um papel muito mais importante do que a f.g.m. na Estatística, pois uma grande parte da teoria assintótica depende de suas propriedades. Diferenciando-se (1.9), sucessivamente, em relação a  $t$ , tem-se

$$\varphi^{(r)}(t; \theta, \phi) = \phi^{r-1} b^{(r)}(\phi t + \theta),$$

em que  $b^{(r)}(\cdot)$  indica a derivada de  $r$ -ésima ordem de  $b(\cdot)$  em relação a  $t$ . Para  $t = 0$ , obtém-se o  $r$ -ésimo cumulante da família (1.5) como

$$\kappa_r = \phi^{r-1} b^{(r)}(\theta). \quad (1.10)$$

Como enfatizado anteriormente, podem ser deduzidos, a partir da equação (1.10), o valor esperado  $\kappa_1$  e a variância  $\kappa_2$  da família (1.5) para  $r = 1$  e  $2$ , respectivamente. Tem-se que  $\kappa_1 = \mu = b'(\theta)$  e  $\kappa_2 = \phi b''(\theta) = \phi d\mu/d\theta$ .

A expressão (1.10) mostra que existe uma relação interessante de recorrência entre os cumulantes da família (1.5), isto é,  $\kappa_{r+1} = \phi d\kappa_r/d\theta$  para  $r = 1, 2, \dots$ . Esse fato é fundamental para a obtenção de propriedades assintóticas dos estimadores de máxima verossimilhança (EMVs) nos MLGs.

Essas expressões podem ser deduzidas, de forma alternativa, usando-se as propriedades da função score. Seja  $\ell = \ell(\theta, \phi) = \log[f(y; \theta, \phi)]$  o logaritmo da função de verossimilhança correspondente a uma única observação em (1.5). Tem-se

$$U = \frac{d\ell}{d\theta} = \phi^{-1}[y - b'(\theta)] \quad \text{e} \quad U' = \frac{d^2\ell}{d\theta^2} = -\phi^{-1}b''(\theta).$$

Logo,

$$E(U) = \phi^{-1} [E(Y) - b'(\theta)] = 0, \quad \text{que implica em } E(Y) = b'(\theta)$$

e, assim,

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

Então,

$$\text{Var}(Y) = \phi b''(\theta).$$

**Exemplo 1.6:** Considerando-se o Exemplo 1.4 (distribuição normal), tem-se que  $\phi = \sigma^2$ ,  $\theta = \mu$  e  $b(\theta) = \theta^2/2$ . Da equação (1.9), obtém-se a f.g.c.

$$\varphi(t) = \frac{1}{\sigma^2} \left[ \frac{(\sigma^2 t + \theta)^2}{2} - \frac{\theta^2}{2} \right] = t\mu + \frac{\sigma^2 t^2}{2}.$$

Note que, diferenciando-se  $\varphi(t)$  e fazendo-se  $t = 0$ , tem-se que  $\kappa_1 = \mu$ ,  $\kappa_2 = \sigma^2$  e  $\kappa_r = 0$ ,  $r \geq 3$ . Assim, todos os cumulantes da distribuição normal de ordem maior do que dois são nulos.

Logo, a f.g.m. é igual a

$$M(t) = \exp \left( t\mu + \frac{\sigma^2 t^2}{2} \right).$$

**Exemplo 1.7:** Considerando-se o Exemplo 1.5 (distribuição binomial), tem-se que  $\phi = 1$ ,  $\theta = \log[\mu/(m - \mu)]$  e  $b(\theta) = -m \log(1 - \pi) = m \log(1 + e^\theta)$ .

Logo, usando-se a f.g.c. (1.9), tem-se

$$\begin{aligned} \varphi(t) &= m [\log(1 + e^{t+\theta}) - \log(1 + e^\theta)] \\ &= \log \left( \frac{1 + e^{t+\theta}}{1 + e^\theta} \right)^m = \log \left( \frac{m - \mu}{m} + \frac{\mu}{m} e^t \right)^m. \end{aligned}$$

Assim, a f.g.m. é

$$M(t) = e^{\varphi(t)} = \left( \frac{m - \mu}{m} + \frac{\mu}{m} e^t \right)^m.$$

A Tabela 1.2 apresenta as funções geradoras de momentos para as distribuições especificadas na Tabela 1.1.

Especificando-se a forma da função  $\mu = q^{-1}(\theta)$ , pode-se demonstrar que a distribuição em (1.5) é univocamente determinada. Assim, uma relação funcional variância-média caracteriza a distribuição na família (1.5). Entretanto, essa relação não caracteriza a distribuição na família exponencial não-linear  $\pi(y; \theta, \phi) = \exp \{ \phi^{-1} [t(y)\theta - b(\theta)] + c(y, \phi) \}$ . Esse fato é comprovado com os três exemplos que se seguem.

**Exemplo 1.8:** Se  $Y$  tem distribuição beta com parâmetros  $\phi^{-1}\mu$  e  $\phi^{-1}(1 - \mu)$  e f.d.p. expressa por

$$f(y; \mu, \phi) = \frac{y^{\phi^{-1}\mu-1} (1 - y)^{\phi^{-1}(1-\mu)-1}}{B[\phi^{-1}\mu, \phi^{-1}(1 - \mu)]},$$



Tabela 1.2: Funções geradoras de momentos para algumas distribuições.

Distribuição	Função geradora de momentos $M(t; \theta, \phi)$
Normal: $N(\mu, \sigma^2)$	$\exp\left(t\mu + \frac{\sigma^2 t^2}{2}\right)$
Poisson: $P(\mu)$	$\exp[\mu(e^t - 1)]$
Binomial: $B(m, \pi)$	$\left(\frac{m - \mu}{m} + \frac{\mu}{m}e^t\right)^m$
Bin. negativa: $BN(\mu, k)$	$\left[1 + \frac{\mu}{k}(1 - e^t)\right]^{-k}$
Gama: $G(\mu, \nu)$	$\left(1 - \frac{t\mu}{\nu}\right)^{-\nu}, \quad t < \frac{\nu}{\mu}$
Normal inversa: $IG(\mu, \sigma^2)$	$\exp\left\{\frac{1}{\sigma^2}\left[\frac{1}{\mu} - \left(\frac{1}{\mu^2} - 2t\sigma^2\right)^{1/2}\right]\right\}, \quad t < \frac{1}{2\sigma^2\mu^2}$

em que  $B(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$  é a função beta completa, tem-se que  $t(y) = \log[y/(1-y)]$ ,  $\theta = \mu$  e  $\text{Var}(Y) = \phi\mu(1-\mu)/(1+\phi)$ , obtendo-se uma função de variância do mesmo tipo que a do modelo binomial.

**Exemplo 1.9:** Se  $Y$  tem distribuição de Euler com média  $\mu$  e f.d.p.

$$f(y; \mu) = \exp\{\mu \log(y) - \mu - \log[\Gamma(\mu)]\},$$

tem-se que  $t(y) = \log(y)$ ,  $\theta = \mu$  e  $\text{Var}(Y) = \mu$ , que é do mesmo tipo que a função de variância do modelo de Poisson.

**Exemplo 1.10:** Se  $Y$  tem distribuição log-normal de parâmetros  $\alpha$  e  $\sigma^2$  e f.d.p.

$$f(y; \alpha, \sigma^2) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left\{-\frac{[\log(y) - \alpha]^2}{2\sigma^2}\right\},$$

então, podem-se obter  $E(Y) = \mu = \exp(\alpha + \sigma^2/2)$ ,  $t(y) = \log(y)$ ,  $\theta = \alpha/\sigma^2$  e  $\text{Var}(Y) = \mu^2[\exp(\sigma^2) - 1]$ , que é do mesmo tipo que a função de variância do modelo

gama.

## 1.5 Estatística suficiente

Uma estatística  $T = T(\mathbf{Y})$  é suficiente para um parâmetro  $\theta$  (que pode ser um vetor) quando resume toda informação sobre esse parâmetro contida na amostra  $\mathbf{Y}$ . Se  $T$  é suficiente para  $\theta$ , então, a distribuição condicional de  $\mathbf{Y}$  dada a estatística  $T(\mathbf{Y})$  é independente de  $\theta$ , isto é,

$$P(\mathbf{Y} = \mathbf{y} | T = t, \theta) = P(\mathbf{Y} = \mathbf{y} | T = t).$$

O critério da fatoração é uma forma conveniente de provar se uma estatística é suficiente. Uma condição necessária e suficiente para  $T$  ser suficiente para um parâmetro  $\theta$  é que a função (densidade ou de probabilidade)  $f_{\mathbf{Y}}(\mathbf{y}; \theta)$  possa ser decomposta como

$$f_{\mathbf{Y}}(\mathbf{y}; \theta) = h(\mathbf{y})g(t, \theta),$$

em que  $t = T(\mathbf{y})$  e  $h(\mathbf{y})$  não dependem de  $\theta$ . Esse resultado é válido para os casos discreto e contínuo.

Seja  $Y_1, \dots, Y_n$  uma amostra aleatória (a.a.) de uma distribuição que pertence à família (1.5). A distribuição conjunta de  $Y_1, \dots, Y_n$  é expressa por

$$\begin{aligned} f(\mathbf{y}; \theta, \phi) &= \prod_{i=1}^n f(y_i; \theta, \phi) = \prod_{i=1}^n \exp \{ \phi^{-1} [y_i \theta - b(\theta)] + c(y_i, \phi) \} \\ &= \exp \left\{ \phi^{-1} \left[ \theta \sum_{i=1}^n y_i - n b(\theta) \right] \right\} \exp \left[ \sum_{i=1}^n c(y_i, \phi) \right]. \end{aligned}$$

Pelo teorema da fatoração de Neyman-Fisher e supondo  $\phi$  conhecido, tem-se que  $T = \sum_{i=1}^n Y_i$  é uma estatística suficiente para  $\theta$ , pois

$$f(\mathbf{y}; \theta, \phi) = g(t, \theta) h(y_1, \dots, y_n),$$

sendo que  $g(t, \theta)$  depende de  $\theta$  e de  $\mathbf{y}$  apenas por meio de  $t$  e  $h(y_1, \dots, y_n)$  independe de  $\theta$ .

Esse fato revela que, se uma distribuição pertence à família exponencial uniparamétrica, então, existe uma estatística suficiente. Na realidade, usando-se o Teorema de Lehmann-Scheffé (Mendenhall et al., 1981) demonstra-se que  $T = \sum_{i=1}^n Y_i$  é uma estatística suficiente minimal.

## 1.6 Família exponencial multiparamétrica

A família exponencial multiparamétrica de dimensão  $k$  é caracterizada por uma função (de probabilidade ou densidade) da forma

$$f(\mathbf{x}; \boldsymbol{\theta}) = h(\mathbf{x}) \exp \left[ \sum_{i=1}^k \eta_i(\boldsymbol{\theta}) t_i(\mathbf{x}) - b(\boldsymbol{\theta}) \right], \quad (1.11)$$

em que  $\boldsymbol{\theta}$  é um vetor de parâmetros, usualmente, de dimensão  $k$ , e as funções  $\eta_i(\boldsymbol{\theta})$ ,  $b(\boldsymbol{\theta})$ ,  $t_i(\mathbf{x})$  e  $h(\mathbf{x})$  têm valores em subconjuntos dos reais. Obviamente, a forma (1.1) é um caso especial de (1.11). Pelo teorema da fatoração, o vetor  $\mathbf{T} = [T_1(\mathbf{X}), \dots, T_k(\mathbf{X})]^T$  é suficiente para o vetor de parâmetros  $\boldsymbol{\theta}$ . Quando  $\eta_i(\boldsymbol{\theta}) = \theta_i$ ,  $i = 1, \dots, k$ , obtém-se de (1.11) a família exponencial na forma canônica com parâmetros canônicos  $\theta_1, \dots, \theta_k$  e estatísticas canônicas  $T_1(\mathbf{X}), \dots, T_k(\mathbf{X})$ . Tem-se,

$$f(\mathbf{x}; \boldsymbol{\theta}) = h(\mathbf{x}) \exp \left[ \sum_{i=1}^k \theta_i t_i(\mathbf{x}) - b(\boldsymbol{\theta}) \right]. \quad (1.12)$$

É fácil verificar (Exercício 12) que as distribuições normal, gama, normal inversa e beta pertencem à família exponencial biparamétrica canônica (1.12) com  $k = 2$ .

Gelfand e Dalal (1990) estudaram a família exponencial biparamétrica  $f(x; \theta, \tau) = h(x) \exp[\theta x + \tau t(x) - b(\theta, \tau)]$ , que é um caso especial de (1.11), com  $k = 2$ . Essa família tem despertado interesse, recentemente, como o componente aleatório dos MLGs superdispersos (Dey et al., 1997). Dois casos especiais importantes dessa família são diretamente obtidos:

- a. a família exponencial canônica uniparamétrica (1.2) surge, naturalmente, quando  $\tau = 0$ ;
- b. o componente aleatório (1.5) dos MLGs é obtido incorporando o parâmetro de dispersão  $\phi$ .

**Exemplo 1.11:** Considere a distribuição multinomial com função de probabilidade

$$f(\mathbf{x}; \boldsymbol{\pi}) = \frac{n!}{x_1! \dots x_k!} \pi_1^{x_1} \dots \pi_k^{x_k},$$

em que  $\sum_{i=1}^k x_i = n$  e  $\sum_{i=1}^k \pi_i = 1$ . Essa distribuição pertence à família exponencial canônica (1.12) com vetor de parâmetros canônicos  $\boldsymbol{\theta} = [\log(\pi_1), \dots, \log(\pi_k)]^T$  e estatística canônica  $\mathbf{T} = (X_1, \dots, X_k)^T$ . Entretanto, devido à restrição  $\sum_{i=1}^k \pi_i = 1$ , a representação mínima da família exponencial é obtida considerando  $\boldsymbol{\theta} = [\log(\pi_1/\pi_k), \dots, \log(\pi_{k-1}/\pi_k)]^T$  e  $\mathbf{T} = (X_1, \dots, X_{k-1})^T$ , ambos vetores de dimensão  $k - 1$ , resultando na família exponencial multiparamétrica de dimensão  $k - 1$

$$f(\mathbf{x}; \boldsymbol{\theta}) = \frac{n!}{x_1! \dots x_k!} \exp \left[ \sum_{i=1}^{k-1} \theta_i x_i - b(\boldsymbol{\theta}) \right], \quad (1.13)$$

com  $\theta_i = \log(\pi_i/\pi_k)$ ,  $i = 1, \dots, k - 1$ , e  $b(\boldsymbol{\theta}) = n \log \left( 1 + \sum_{i=1}^{k-1} e^{\theta_i} \right)$ .

Pode-se demonstrar que os dois primeiros momentos da estatística suficiente  $\mathbf{T} = [T_1(\mathbf{X}), \dots, T_k(\mathbf{X})]^T$  na família exponencial canônica (1.12) são iguais a

$$E(\mathbf{T}) = \frac{\partial b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \quad \text{Cov}(\mathbf{T}) = \frac{\partial^2 b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}. \quad (1.14)$$

As expressões (1.14) generalizam (1.3). Nas equações (1.14), o vetor  $\partial b(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$  de dimensão  $k$  tem um componente típico  $E[T_i(\mathbf{X})] = \partial b(\boldsymbol{\theta})/\partial \theta_i$  e a matriz  $\partial^2 b(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T$  de ordem  $k$  tem como elemento típico  $\text{Cov}(T_i(\mathbf{X}), T_j(\mathbf{X})) = \partial^2 b(\boldsymbol{\theta})/\partial \theta_i \partial \theta_j$ . Assim, os valores esperados e as covariâncias das estatísticas

suficientes do modelo (1.12) são facilmente obtidos por simples diferenciação. A demonstração das equações (1.14) é proposta no Exercício 19.

**Exemplo 1.11 (cont.):** Para o modelo multinomial (1.13), usando as equações (1.14), têm-se

$$E(X_i) = n \frac{\partial}{\partial \theta_i} \log \left( 1 + \sum_{i=1}^{k-1} e^{\theta_i} \right) = \frac{ne^{\theta_i}}{1 + \sum_{i=1}^{k-1} e^{\theta_i}} = \frac{n \frac{\pi_i}{\pi_k}}{1 + \sum_{i=1}^{k-1} \frac{\pi_i}{\pi_k}} = n\pi_i,$$

para  $i = j$

$$\text{Var}(X_i) = n \frac{\partial^2}{\partial \theta_i^2} \log \left( 1 + \sum_{i=1}^{k-1} e^{\theta_i} \right) = n\pi_i(1 - \pi_i).$$

e para  $i \neq j$

$$\text{Cov}(X_i, X_j) = n \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log \left( 1 + \sum_{i=1}^{k-1} e^{\theta_i} \right) = \frac{-ne^{\theta_i} e^{\theta_j}}{\left( 1 + \sum_{i=1}^{k-1} e^{\theta_i} \right)^2} = -n\pi_i \pi_j$$

Finalmente, apresenta-se mais uma distribuição na família exponencial canônica (1.12) com  $k = 2$ , no exemplo que se segue.

**Exemplo 1.12:** Considere a distribuição Gaussiana inversa reparametrizada por  $(\alpha, \beta > 0)$

$$f(x; \alpha, \beta) = \sqrt{\frac{\alpha}{2\pi}} e^{\sqrt{\alpha\beta}x - 3/2} \exp \left[ -\frac{1}{2}(\alpha x^{-1} + \beta x) \right], \quad x > 0.$$

Pode-se escrever essa f.d.p. na forma (1.12) com  $\mathbf{t} = \left( -\frac{1}{2}x^{-1}, -\frac{1}{2}x \right)^T$ ,  $\boldsymbol{\theta} = (\alpha, \beta)^T$  e  $b(\boldsymbol{\theta}) = -\frac{1}{2}\log(\alpha) - \sqrt{\alpha\beta}$ . Usando-se as equações (1.14), obtêm-se, por simples diferenciação,

$$E(X) = \sqrt{\frac{\alpha}{\beta}}, \quad E(X^{-1}) = \alpha^{-1} + \sqrt{\frac{\beta}{\alpha}}$$

e

$$\text{Cov}(X, X^{-1}) = \begin{pmatrix} \alpha^{1/2}\beta^{-3/2} & -(\alpha\beta)^{-1/2} \\ -(\alpha\beta)^{-1/2} & 2\alpha^{-2} + \alpha^{-3/2}\beta^{1/2} \end{pmatrix}.$$

## 1.7 Exercícios

1. Verifique se as distribuições que se seguem pertencem à família (1.5). Obtenha  $\varphi(t)$ ,  $M(t)$ ,  $E(Y)$ ,  $\text{Var}(Y)$  e  $V(\mu)$ .

a) Poisson:  $Y \sim P(\mu)$ ,  $\mu > 0$

$$f(y; \mu) = \frac{e^{-\mu} \mu^y}{y!}, \quad y = 0, 1, 2, \dots;$$

b) Binomial negativa ( $k$  fixo):  $Y \sim \text{BN}(\mu, k)$ ,  $k > 0$ ,  $\mu > 0$

$$f(y; \mu, k) = \frac{\Gamma(k+y)}{\Gamma(k)y!} \frac{\mu^y k^k}{(\mu+k)^{k+y}}, \quad y = 0, 1, 2, \dots;$$

c) Gama:  $Y \sim G(\mu, \nu)$ ,  $\nu > 0$ ,  $\mu > 0$

$$f(y; \mu, \nu) = \frac{\left(\frac{\nu}{\mu}\right)^\nu}{\Gamma(\nu)} y^{\nu-1} \exp\left(-\frac{y\nu}{\mu}\right), \quad y > 0;$$

d) Normal inversa (ou inversa Gaussiana):  $Y \sim \text{IG}(\mu, \sigma^2)$ ,  $\sigma^2 > 0$ ,  $\mu > 0$

$$f(y; \mu, \sigma^2) = \left(\frac{1}{2\pi\sigma^2 y^3}\right)^{1/2} \exp\left[-\frac{(y-\mu)^2}{2\mu^2\sigma^2 y}\right], \quad y > 0.$$

2. Seja  $X$  uma v.a. com distribuição gama  $G(\nu)$  de um parâmetro  $\nu > 0$ , com f.d.p.

$$f(x; \nu) = \frac{x^{\nu-1} e^{-x}}{\Gamma(\nu)}, \quad x > 0.$$

Sendo  $E(X) = \nu$ , mostre que usando-se a transformação  $Y = \frac{X}{\nu}\mu$ , obtém-se a f.d.p. usada no item c) do Exercício 1.

3. Seja  $Y$  uma v.a. com distribuição de Poisson truncada (Ridout e Demétrio, 1992) com parâmetro  $\lambda > 0$ , isto é, com função de probabilidade expressa por

$$f(y; \lambda) = \frac{e^{-\lambda} \lambda^y}{y!(1 - e^{-\lambda})} = \frac{\lambda^y}{y!(e^\lambda - 1)}, \quad y = 1, 2, \dots$$

Mostre que:

a) essa distribuição pertence à família exponencial na forma canônica;

- b)  $E(Y) = \mu = \frac{\lambda}{1 - e^{-\lambda}};$
- c)  $\text{Var}(Y) = \frac{\lambda}{1 - e^{-\lambda}} \left( 1 - \frac{\lambda e^{-\lambda}}{1 - e^{-\lambda}} \right) = \mu(1 + \lambda - \mu);$
- d)  $M(t) = \frac{\exp(\lambda e^t) - 1}{e^\lambda - 1}.$

4. Seja  $Y$  uma v.a. com distribuição binomial truncada (Vieira et al., 2000) com probabilidade de sucesso  $0 < \pi < 1$  e com função de probabilidade expressa por

$$f(y; \pi) = \frac{\binom{m}{y} \pi^y (1 - \pi)^{(m-y)}}{1 - (1 - \pi)^m}, \quad y = 1, \dots, m.$$

Mostre que:

- a) essa distribuição pertence à família exponencial na forma canônica;
- b)  $E(Y) = \mu = \frac{m\pi}{1 - (1 - \pi)^m};$
- c)  $\text{Var}(Y) = \mu[1 + \pi(m - 1) - \mu];$
- d)  $M(t) = \frac{(1 - \pi + \pi e^t)^m - (1 - \pi)^m}{1 - (1 - \pi)^m}.$

5. De acordo com Smyth (1989), uma distribuição contínua pertence à família exponencial se sua f.d.p. pode ser expressa na forma

$$f(y; \theta, \phi) = \exp \left\{ \frac{w}{\phi} [y\theta - b(\theta)] + c(y, \phi) \right\}, \quad (1.15)$$

sendo  $b(\cdot)$  e  $c(\cdot)$  funções conhecidas,  $\phi > 0$ , denominado parâmetro de dispersão, e  $w$ , um peso *a priori*. Se a constante  $\phi$  é desconhecida, então, a expressão (1.15) define uma família exponencial com dois parâmetros apenas se

$$c(y, \phi) = -\frac{w}{\phi} g(y) - \frac{1}{2} s \left( -\frac{w}{\phi} \right) + t(y),$$

sendo  $g(\cdot)$ ,  $s(\cdot)$  e  $t(\cdot)$  funções conhecidas e, nesse caso,  $g'(\cdot)$  deve ser a inversa de  $b'(\cdot)$  tal que  $\theta = g'(\mu)$ . Mostre que isso ocorre para as distribuições normal, normal

inversa e gama.

6. Seja  $Y | P \sim B(m, P)$  e  $P \sim \text{Beta}(\alpha, \beta)$ ,  $\alpha > 0$ ,  $\beta > 0$ ,  $0 < p < 1$ , isto é,

$$f(y | p) = \binom{m}{y} p^y (1-p)^{m-y} \quad \text{e} \quad f(p) = \frac{p^{\alpha-1} (1-p)^{\beta-1}}{B(\alpha, \beta)},$$

sendo  $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$  (Hinde e Demétrio, 1998a). Mostre que:

a) incondicionalmente,  $Y$  tem distribuição beta-binomial com f.d.p. expressa por

$$f(y) = \binom{m}{y} \frac{B(\alpha+y, m+\beta-y)}{B(\alpha, \beta)};$$

b)  $E(Y) = m \frac{\alpha}{\alpha+\beta} = m\pi$  e  $\text{Var}(Y) = m\pi(1-\pi)[1 + \rho(m-1)]$ , sendo  $\rho = \frac{1}{\alpha+\beta+1}$ ;

c) a distribuição beta-binomial não pertence à família (1.5).

7. Seja  $Y_i | Z_i = z_i \sim P(z_i)$ ,  $i = 1, \dots, n$ , isto é,

$$P(Y_i = y_i | Z_i = z_i) = \frac{e^{-z_i} z_i^{y_i}}{y_i!}, \quad y_i = 0, 1, 2, \dots$$

Então, se:

a)  $Z_i \sim G(k, \lambda_i)$ ,  $z_i > 0$ , isto é, com f.d.p. expressa por

$$f(z_i; k, \lambda_i) = \frac{\left(\frac{\lambda_i}{k}\right)^{\lambda_i}}{\Gamma(\lambda_i)} z_i^{\lambda_i-1} \exp\left(-\frac{z_i \lambda_i}{k}\right),$$

mostre que para  $k$  fixo, incondicionalmente,  $Y_i$  tem distribuição binomial negativa, que pertence à família exponencial, com  $E(Y_i) = k\lambda_i^{-1} = \mu_i$  e  $\text{Var}(Y_i) = \mu_i + k^{-1}\mu_i^2$ ;

b)  $Z_i \sim G(k_i, \lambda)$ ,  $z_i > 0$ , isto é, com f.d.p. expressa por

$$f(z_i; k_i, \lambda) = \frac{\left(\frac{\lambda}{k_i}\right)^{\lambda}}{\Gamma(\lambda)} z_i^{\lambda-1} \exp\left(-\frac{z_i \lambda}{k_i}\right),$$



mostre que para  $\lambda$  fixo, incondicionalmente,  $Y_i$  tem distribuição binomial negativa, que não pertence à família exponencial, com  $E(Y_i) = k_i \lambda^{-1} = \mu_i$  e  $\text{Var}(Y_i) = \mu_i + \lambda^{-1} \mu_i = \phi \mu_i$ , sendo  $\phi = 1 + \lambda^{-1}$ .

**8.** Uma forma geral para representar a função de probabilidade da distribuição binomial negativa (Ridout et al., 2001) é expressa por

$$P(Y = y) = \frac{\Gamma\left(y + \frac{\mu^c}{\nu}\right)}{\Gamma\left(\frac{\mu^c}{\nu}\right) y!} \left(1 + \frac{\mu^{c-1}}{\nu}\right)^{-y} (1 + \nu \mu^{1-c})^{-\frac{\mu^c}{\nu}}, \quad y = 0, 1, 2, \dots$$

a) mostre que  $E(Y) = \mu$  e  $\text{Var}(Y) = \mu + \nu \mu^{2-c}$ . Obtenha  $E(Y)$  e  $\text{Var}(Y)$  para os casos mais comuns ( $c = 0$  e  $c = 1$ ) da distribuição binomial negativa;

b) mostre que  $P(Y = y)$  pertence à família (1.5) apenas se  $c = 0$ .

**9.** Uma distribuição para explicar o excesso de zeros em dados de contagens é a distribuição de Poisson inflacionada de zeros, com função de probabilidade igual a

$$P(Y = y) = \begin{cases} \omega + (1 - \omega)e^{-\lambda} & y = 0 \\ (1 - \omega) \frac{e^{-\lambda} \lambda^y}{y!} & y = 1, 2, \dots \end{cases}$$

Mostre que  $E(Y) = (1 - \omega)\lambda = \mu$  e  $\text{Var}(Y) = \mu + \left(\frac{\omega}{1 - \omega}\right) \mu^2$  (Ridout et al., 1998).

**10.** Uma distribuição alternativa para explicar o excesso de zeros em dados na forma de contagens é a distribuição binomial negativa inflacionada de zeros (Ridout et al., 1998), com função de probabilidade expressa por

$$P(Y = y) = \begin{cases} \omega + (1 - \omega) (1 + \alpha \lambda^c)^{-\frac{\lambda^{1-c}}{\alpha}}, & y = 0 \\ (1 - \omega) \frac{\Gamma\left(y + \frac{\lambda^{1-c}}{\alpha}\right)}{y! \Gamma\left(\frac{\lambda^{1-c}}{\alpha}\right)} (1 + \alpha \lambda^c)^{-\frac{\lambda^{1-c}}{\alpha}} \left(1 + \frac{\lambda^{-c}}{\alpha}\right)^{-y}, & y = 1, 2, \dots \end{cases}$$

Mostre que  $E(Y) = (1 - \omega)\lambda$  e  $\text{Var}(Y) = (1 - \omega)\lambda(1 + \omega\lambda + \alpha\lambda^c)$ .

**11.** Obtenha as funções geradoras de momentos e de cumulantes da distribuição secante hiperbólica generalizada definida pela f.d.p. (1.6).

**12.** Mostre que as distribuições normal, gama, normal inversa e beta pertencem à família exponencial canônica biparamétrica (1.12) com  $k = 2$  e identifique  $t_1(\mathbf{x})$ ,  $t_2(\mathbf{x})$ ,  $h(\mathbf{x})$  e  $b(\boldsymbol{\theta})$ .

**13.** No Exercício 12, use as equações (1.14) para calcular  $E(\mathbf{T})$  e  $\text{Cov}(\mathbf{T})$ , sendo  $\mathbf{T} = [T_1(\mathbf{x}), T_2(\mathbf{x})]^T$ .

**14.** A partir das equações (1.14), obtenha  $E[T(\mathbf{X})]$  e  $\text{Var}[T(\mathbf{X})]$  para as 24 distribuições apresentadas por Cordeiro et al. (1995) na família exponencial uniparamétrica (1.1).

**15.** Demonstre as fórmulas de  $E(\mathbf{X})$ ,  $E(\mathbf{X}^{-1})$  e  $\text{Cov}(\mathbf{X}, \mathbf{X}^{-1})$  citadas no Exemplo 1.12.

**16.** Seja  $f(x; \theta) = h(x) \exp[g(x; \theta)]$  uma distribuição uniparamétrica arbitrária. Demonstre que uma condição necessária para ela não pertencer à família exponencial (1.1) é que, dados quatro pontos amostrais  $x_1$ ,  $x_2$ ,  $x_3$  e  $x_4$ , o quociente  $\frac{g(x_1, \theta) - g(x_2, \theta)}{g(x_3, \theta) - g(x_4, \theta)}$  é uma função que depende de  $\theta$ .

**17.** Usando o Exercício 16, mostre que a distribuição de Cauchy  $f(x; \theta) = \frac{1}{\pi [1 + (x - \theta)^2]}$  não é um membro da família exponencial uniparamétrica (1.1).

**18.** Demonstre que para a família exponencial biparamétrica  $f(x; \theta, \tau) = h(x) \exp[\theta x + \tau t(x) - b(\theta, \tau)]$ , tem-se:  $E(X) = b^{(1,0)}$ ,  $\text{Var}(X) = b^{(2,0)}$ ,  $E[T(X)] = b^{(0,1)}$  e  $\text{Cov}[X, T(X)] = b^{(1,1)}$ , sendo que  $b^{(r,s)} = \frac{\partial^{(r+s)} b(\theta, \tau)}{\partial \theta^r \partial \tau^s}$ .

**19.** Considere a família exponencial multiparamétrica na forma canônica (1.12).

Demonstre que os dois primeiros momentos do vetor  $\mathbf{T}$  de estatísticas suficientes são expressos pelas equações (1.14).

**20.** Suponha que  $Y_1$  e  $Y_2$  têm distribuições de Poisson independentes com médias  $\mu$  e  $\rho\mu$ , respectivamente. Mostre que

a)  $Y_+ = Y_1 + Y_2$  tem distribuição de Poisson com média  $\mu(1 + \rho)$ ;

b)  $Y_1|Y_+ = m$  tem distribuição binomial  $B(m, (1 + \rho)^{-1})$ .

**21.** Seja  $X$  uma variável aleatória binomial  $B(m, \theta)$ .

a) Se  $m \rightarrow \infty$  e  $\theta \rightarrow 0$  de modo que  $m\theta = \mu$  permanece constante, demonstre que  $P(X = k) \rightarrow e^{-\mu}\mu^k/k!$ . Esse limite é a base da aproximação da distribuição de Poisson para a distribuição binomial.

b) Demonstre, pela aproximação normal, que

$$P(X = k) \approx \frac{1}{\sqrt{2\pi m\theta(1-\theta)}} \exp \left[ -\frac{(k - m\theta)^2}{2m\theta(1-\theta)} \right].$$

**22.** Obtenha uma expressão geral para o momento central de ordem  $r$  da família de distribuições (1.5), a partir da expressão geral (1.10) dos cumulantes.

**23.** Seja uma distribuição na família exponencial natural com f.d.p. ( $y > 0$ )

$$f(y; \theta) = c(y) \exp[\theta y - b(\theta)]$$

e média  $\mu = \tau(\theta)$ . Demonstre que  $g(y; \theta) = yf(y; \theta)/\tau(\theta)$  é uma nova f.d.p. e calcule suas funções geradoras de momentos e de cumulantes.

**24.** A distribuição logarítmica é definida pela função de probabilidade

$$f(y; \rho) = -\frac{\rho^y}{y \log(1 - \rho)}$$

para  $y = 1, 2, \dots$  e  $0 < \rho < 1$ . Demonstre que essa distribuição pertence à família exponencial e que

$$E(Y) = \frac{\rho}{b(\rho)(1-\rho)} \quad \text{e} \quad \text{Var}(Y) = \frac{\rho[1 - \frac{\rho}{b(\rho)}]}{b(\rho)(1-\rho)^2},$$

sendo  $b(\rho) = -\log(1-\rho)$ .

**25.** Demonstre as fórmulas de recorrência para os momentos ordinários ( $\mu'_r$ ) e centrais ( $\mu_r$ ) da distribuição binomial:

$$\mu'_{r+1} = \mu(1-\mu) \left[ \frac{m\mu'_r}{(1-\mu)} + \frac{d\mu'_r}{d\mu} \right] \quad \text{e} \quad \mu_{r+1} = \mu(1-\mu) \left[ mr\mu_{r-1} + \frac{d\mu_r}{d\mu} \right].$$

**26.** Se  $Y$  tem distribuição exponencial de média unitária, demonstre que a função geradora de momentos de  $Y = \log(X)$  é igual a  $M(t) = \Gamma(1+t)$  e que sua f.d.p. é  $f(y) = \exp(y - e^y)$ .

**27.** Use a expansão de Taylor para verificar que se  $E(X) = \mu$  e  $\text{Var}(X) = \sigma^2$ , então, para qualquer função bem comportada  $G(X)$ , tem-se, para  $\sigma$  suficientemente pequeno,  $\text{Var}[G(X)] = G'(\mu)^2 \sigma^2$ . Mostre que, se  $X \sim B(m, \pi)$ , pode-se estimar  $\text{Var}\{\log[X/(m-X)]\}$  por  $1/x + 1/(m-x)$ , sendo  $x$  o valor observado de  $X$ .

**28.** Demonstre que os cumulantes de uma variável aleatória  $X$  satisfazem  $\kappa_1(a+bX) = a + b\kappa_1(X)$  e  $\kappa_r(a+bX) = b^r \kappa_r(X)$  para  $r \geq 2$ , sendo  $a$  e  $b$  constantes.

## Capítulo 2

# Modelo Linear Generalizado

### 2.1 Introdução

A seleção de modelos é uma parte importante de toda pesquisa em modelagem estatística. Envolve encontrar um modelo que seja o mais simples possível e que descreva bem o processo gerador dos valores observados que surgem em diversas áreas do conhecimento como agricultura, demografia, ecologia, economia, engenharia, geologia, medicina, ciência política, sociologia e zootecnia, entre outras.

Nelder e Wedderburn (1972) mostraram que um conjunto de técnicas estatísticas, comumente estudadas separadamente, podem ser formuladas, de uma maneira unificada, como uma classe de modelos de regressão. A essa teoria unificadora de modelagem estatística, uma extensão dos modelos clássicos de regressão, denominaram de **modelos lineares generalizados** (MLGs). Esses modelos envolvem uma variável resposta univariada, variáveis explanatórias e uma amostra aleatória de  $n$  observações independentes, sendo que

- i) a variável resposta, **componente aleatório** do modelo, tem distribuição pertencente à família de distribuições (1.5) que engloba as distribuições normal, gama e normal inversa para dados contínuos; binomial para proporções; Poisson e binomial negativa para contagens;
- ii) as variáveis explanatórias entram na forma de uma estrutura linear, constituindo o **componente sistemático** do modelo;

- iii) a ligação entre os componentes aleatório e sistemático é feita por meio de uma função adequada como, por exemplo, logarítmica para os modelos log-lineares, denominada **função de ligação**.

O componente sistemático é estabelecido durante o planejamento (fundamental para a obtenção de conclusões confiáveis) do experimento, resultando em modelos de regressão (linear simples, múltipla, etc.), de análise de variância (delineamentos inteiramente casualizados, casualizados em blocos, quadrados latinos com estrutura de tratamentos fatorial, parcelas subdivididas, etc.) e de análise de covariância. O componente aleatório é especificado assim que são definidas as medidas a serem realizadas, que podem ser contínuas ou discretas, exigindo o ajuste de diferentes distribuições. A partir de um mesmo experimento podem ser obtidas medidas de diferentes tipos, como por exemplo, dados de altura de plantas, número de lesões por planta e proporção de plantas doentes.

No modelo clássico de regressão, tem-se

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon},$$

sendo  $\mathbf{Y}$  o vetor, de dimensões  $n \times 1$ , da variável resposta,  $\boldsymbol{\mu} = E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ , o componente sistemático,  $\mathbf{X}$  a matriz do modelo, de dimensões  $n \times p$ ,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ , o vetor dos parâmetros desconhecidos,  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T$ , o componente aleatório com  $\epsilon_i \sim N(0, \sigma^2)$ ,  $i = 1, \dots, n$ . Nesse caso, tem-se que a distribuição normal  $N(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$  de  $\mathbf{Y}$  define o componente aleatório e o vetor de médias  $\boldsymbol{\mu}$  da distribuição normal é igual ao preditor linear que representa o componente sistemático. Essa é a forma mais simples de ligação entre esses dois componentes, sendo denominada de função de ligação identidade.

Em muitos casos, porém, essa estrutura aditiva entre o componente sistemático e o componente aleatório não é verificada. Além disso, não há razão para se restringir à estrutura simples especificada pela função de ligação identidade, nem à distribuição normal para o componente aleatório e à suposição de homogeneidade de variâncias.

Outros modelos foram surgindo e os desenvolvimentos que conduziram a essa visão geral da modelagem estatística remontam a quase dois séculos. Assim, um MLG é definido por uma distribuição de probabilidade, membro da família (1.5) de distribuições, para a variável resposta, um conjunto de variáveis explanatórias descrevendo a estrutura linear do modelo e uma função de ligação entre a média da variável resposta e a estrutura linear. Entre os métodos estatísticos para a análise de dados univariados, que são casos especiais dos MLGs, citam-se:

- (a) modelo clássico de regressão múltipla (Legendre, Gauss, início do século XIX) e modelo de análise de variância para experimentos planejados (Fisher, 1920 a 1935) com o erro aleatório tendo distribuição normal;
- (b) modelo complemento log-log para ensaios de diluição, envolvendo a distribuição binomial (Fisher, 1922);
- (c) modelo probito (Bliss, 1935) para o estudo de proporções, envolvendo a distribuição binomial;
- (d) modelo logístico (Berkson, 1944; Dyke e Patterson, 1952; Rasch, 1960; Cox, 1970) para o estudo de proporções, envolvendo a distribuição binomial;
- (e) modelos log-lineares para análise de dados na forma de contagens em tabelas de contingência, envolvendo as distribuições de Poisson e multinomial (Birch, 1963; Haberman, 1970);
- (f) modelo logístico para tabelas multidimensionais de proporções;
- (g) modelos de testes de vida, envolvendo a distribuição exponencial (Feigl e Zelen, 1965; Zippin e Armitage, 1966; Gasser, 1967);
- (h) polinômios inversos para ensaios de adubação, envolvendo a distribuição normal na escala logarítmica e linearidade na escala inversa (Nelder, 1966);
- (i) modelo de análise de variância com efeitos aleatórios;

- (j) modelo estrutural para dados com distribuição gama;
- (l) modelo de regressão não-simétrica.

Além dessas técnicas usuais, outros modelos podem ser definidos no contexto dos MLGs como, por exemplo, os modelos de Box e Cox (1964) e alguns modelos de séries temporais. Devido ao grande número de métodos estatísticos que engloba, a teoria dos MLGs vem desempenhando um papel importante na Estatística moderna, tanto para especialistas, quanto para não-especialistas. Esses modelos podem ainda representar um meio unificado de ensino da Estatística, em qualquer curso de graduação ou pós-graduação.

Algumas referências para o estudo dos MLGs e extensões são: Cordeiro (1986), McCullagh e Nelder (1989), Firth (1991), Francis et al. (1993), Fahrmeir e Tutz (1994), McCulloch e Searle (2000), Demétrio (2001), Collet (2002), Myers et al. (2002), Paula (2004), Molenberghs e Verbeke (2005), Lee et al. (2006), Hardin e Hilbe (2007), Dobson e Barnett (2008) e Aitkin et al. (2009).

## 2.2 Exemplos de motivação

A seguir, serão apresentados alguns dos modelos que apareceram na literatura, independentemente, e que, conforme será mostrado, podem ser agrupados de acordo com algumas propriedades comuns, o que permite um método unificado para a estimação dos parâmetros.

### 2.2.1 Ensaios do tipo dose-resposta

Ensaios do tipo dose-resposta são aqueles em que uma determinada droga é administrada em  $k$  diferentes doses,  $d_1, \dots, d_k$ , respectivamente, a  $m_1, \dots, m_k$  indivíduos. Suponha que cada indivíduo responde, ou não, à droga, tal que a resposta é quantal (tudo ou nada, isto é, 1 ou 0). Após um período especificado de tempo,  $y_1, \dots, y_k$  indivíduos respondem à droga. Por exemplo, quando um inseticida é apli-



cado a um determinado número de insetos, eles respondem (morrem), ou não (sobrevivem), à dose aplicada. Quando uma droga benéfica é administrada a um grupo de pacientes, eles podem melhorar (sucesso), ou não (fracasso). Dados resultantes desse tipo de ensaio podem ser considerados como provenientes de uma distribuição binomial com probabilidade  $\pi_i$ , que é a probabilidade de ocorrência (sucesso) do evento sob estudo, ou seja, o número de sucessos  $Y_i$  tem distribuição binomial  $B(m_i, \pi_i)$ .

Os objetivos desse tipo de experimento são, em geral, modelar a probabilidade de sucesso  $\pi_i$  como função de variáveis explanatórias e, então, determinar doses efetivas ( $DL_p$ , doses que causam mudança de estado em 100p% dos indivíduos, por exemplo,  $DL_{50}, DL_{90}$ ), comparar potências de diferentes produtos, etc.

### Exemplo 2.1: Toxicidade de Rotenone

Os dados da Tabela 2.1 referem-se a um ensaio de toxicidade de rotenone (Martin, 1942), no delineamento completamente casualizado, em que doses ( $d_i$ ) do inseticida foram aplicadas a  $m_i$  insetos (*Macrosiphoniella sanborni*, pulgão do crisântemo) e, após um certo tempo, foram observados os números ( $y_i$ ) de insetos mortos.

Tabela 2.1: Número de insetos mortos ( $y_i$ ) de  $m_i$  insetos que receberam a dose  $d_i$  de rotenone.

Dose ( $d_i$ )	$m_i$	$y_i$	$p_i$
0,0	49	0	0,00
2,6	50	6	0,12
3,8	48	16	0,33
5,1	46	24	0,52
7,7	49	42	0,86
10,2	50	44	0,88

O interesse do pesquisador estava na determinação das doses letais que ma-

tam 50% ( $DL_{50}$ ) e 90% ( $DL_{90}$ ) dos insetos, para recomendação de aplicação do inseticida no campo. Pode-se observar que o gráfico (Figura 2.1) de dispersão das proporções ( $p_i = y_i/m_i$ ) de insetos mortos versus as doses ( $d_i$ ) tem um aspecto sigmoidal o que orienta a escolha do modelo para  $\pi_i$ .

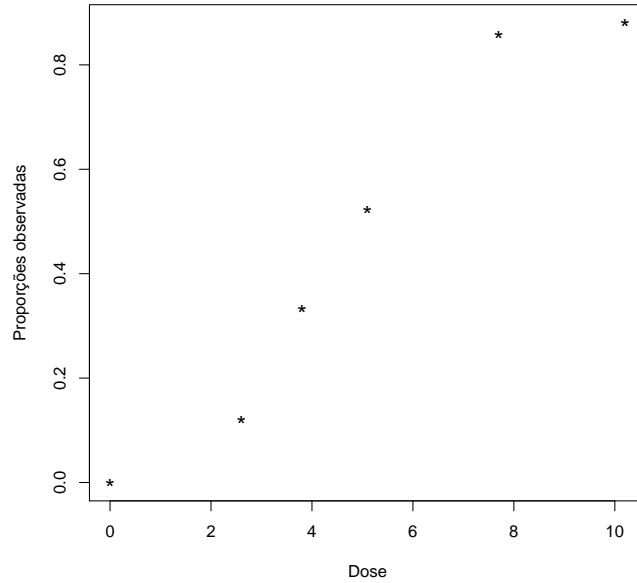


Figura 2.1: Gráfico de dispersão das proporções ( $p_i$ ) versus doses ( $d_i$ ) de rotenone, referentes à Tabela 2.1.

Dois aspectos devem ser considerados nos ensaios de dose-resposta. Um é a intensidade do estímulo que pode ser a dose de uma droga (inseticida, fungicida, herbicida, medicamento) e o outro é o indivíduo (um inseto, um esporo, uma planta, um paciente). O estímulo é aplicado a uma intensidade especificada em unidades de concentração e como resultado uma resposta do indivíduo é obtida. Quando a resposta é binária (0 ou 1), sua ocorrência, ou não, dependerá da intensidade do estímulo aplicado. Para todo indivíduo haverá um certo nível de intensidade abaixo do qual a resposta não ocorre e acima do qual ela ocorre; na terminologia farmacológica e toxicológica, esse valor é denominado tolerância (Ashton, 1972). Essa tolerância varia de um indivíduo para outro da população e, então, há uma

distribuição de tolerâncias à qual pode-se associar uma variável aleatória  $U$  com f.d.p. representada por curvas, simétricas ou assimétricas, dos tipos apresentados na Figura 2.2.

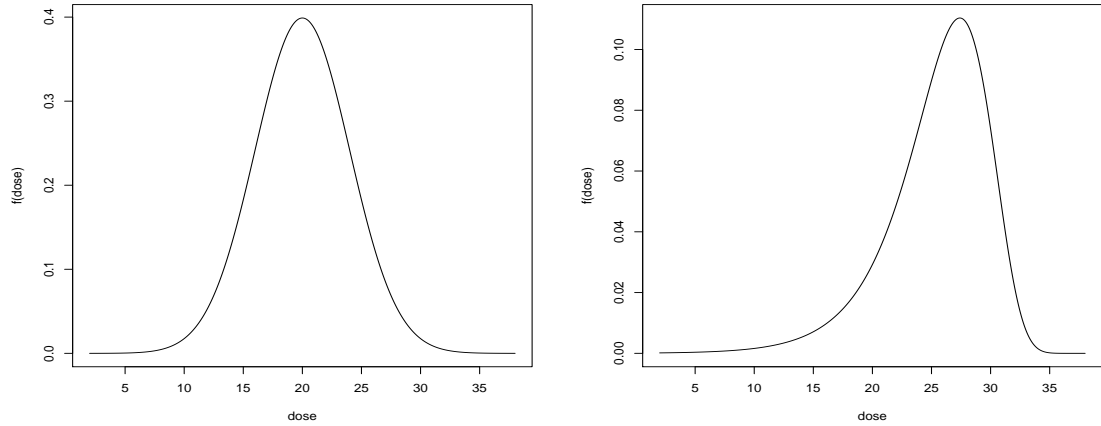


Figura 2.2: Dois tipos de curvas para distribuições de tolerância.

Se a dose  $d$  é dada para a população toda e  $f(u)$  é a função densidade para a distribuição das tolerâncias, todo indivíduo cuja tolerância é menor do que  $d$  responderá à droga, e a probabilidade de que um indivíduo escolhido ao acaso responda à dose, conforme a Figura 2.3, é expressa por

$$\pi = P(U \leq d) = F(d) = \int_{-\infty}^d f(u) du. \quad (2.1)$$

A probabilidade de ocorrer uma resposta (sucesso) é tipicamente nula para valores pequenos de  $d$ , unitária para valores grandes de  $d$  (pois, então, um sucesso é certo) e é uma função estritamente crescente de  $d$ . Essa curva tem as propriedades matemáticas de uma função de distribuição contínua acumulada e exibe a forma sigmoidal típica da Figura 2.3.

Observa-se que nenhum indivíduo responde se a dose é muito pequena e que todos os indivíduos respondem se a dose é muito grande. Essas suposições nem sempre são razoáveis. Pode haver indivíduos que respondem, naturalmente, sem

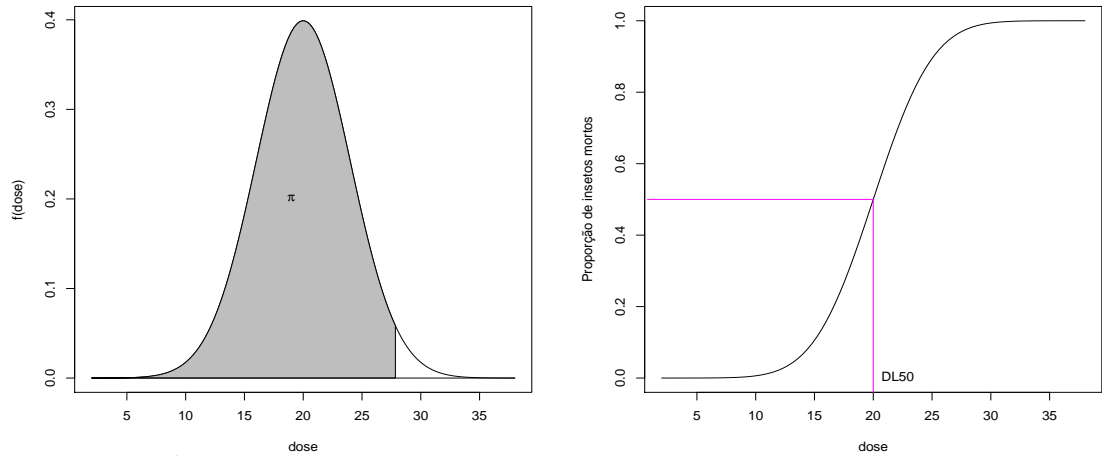


Figura 2.3: Área sob a curva de tolerância e correspondente distribuição acumulada.

a droga (morte natural) e outros que são imunes à droga, o que pode causar um excesso de zeros (Ridout et al., 1998) e uma variabilidade maior do que a esperada (superdispersão) (Hinde e Demétrio, 1998a,b).

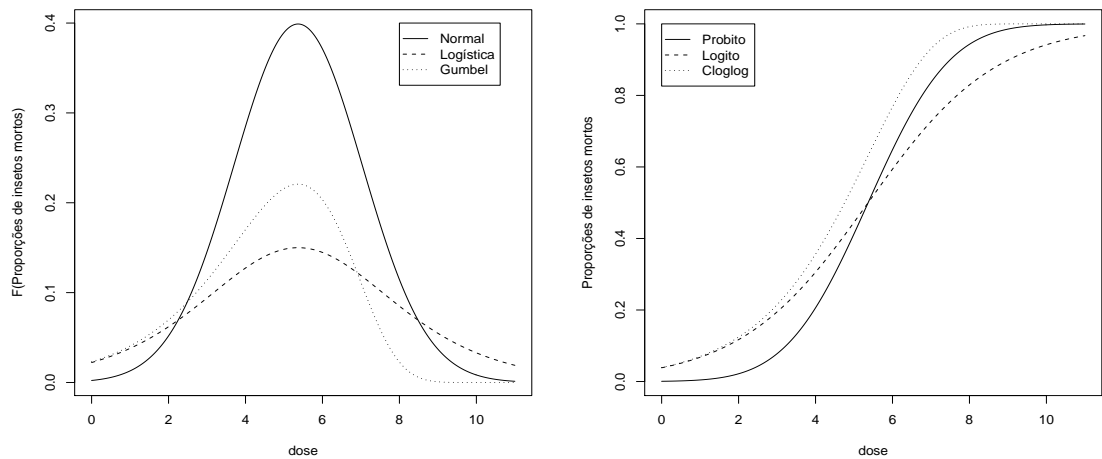


Figura 2.4: Curvas para distribuições de tolerância e correspondentes sigmóides.

O problema, então, consiste em encontrar uma curva sigmoidal que se ajuste bem aos dados e, a partir dela, calcular  $DL_{50}$  e  $DL_{90}$ . Isso pode ser realizado com o uso de modelos não-lineares nos parâmetros. Então, a ideia é se fazer

uma transformação tal que essa curva sigmoideal se transforme em uma reta e, assim, procedimentos comuns de regressão podem ser usados para se estimarem os parâmetros. Na Figura 2.4 são apresentadas as distribuições, mais comumente usadas, e suas correspondentes curvas sigmoideais, cujas expressões e respectivas transformações lineares são apresentadas, a seguir.

i) **Modelo probito** (“**Probability unit**”)

Nesse caso, considera-se que  $U$  tem distribuição normal de média  $\mu \in \mathbb{R}$  e variância  $\sigma^2 > 0$ , isto é,

$$f_U(u; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(u - \mu)^2}{2\sigma^2} \right],$$

e, portanto, com  $Z = \frac{U - \mu}{\sigma} \sim N(0, 1)$ . Então,

$$\pi_i = P(U \leq d_i) = P \left( Z \leq -\frac{\mu}{\sigma} + \frac{1}{\sigma} d_i \right) = P(Z \leq \beta_1 + \beta_2 d_i)$$

para  $\beta_1 = -\mu/\sigma$  e  $\beta_2 = 1/\sigma$ . Logo,

$$\pi_i = \Phi(\beta_1 + \beta_2 d_i)$$

é uma função não-linear em um conjunto linear de parâmetros, em que  $\Phi(\cdot)$  representa a função de distribuição normal padrão. É linearizada por

$$\text{probit}(\pi_i) = \Phi^{-1}(\pi_i) = \beta_1 + \beta_2 d_i.$$

ii) **Modelo logístico** (“**Logistic unit**”)

Nesse caso, considera-se que  $U$  tem distribuição logística com parâmetros  $\mu \in \mathbb{R}$  e  $\tau > 0$ , que é similar à distribuição normal em forma, com caudas um pouco mais longas e tem f.d.p. expressa por

$$f_U(u; \mu, \tau) = \frac{1}{\tau} \frac{\exp \left( \frac{u - \mu}{\tau} \right)}{\left[ 1 + \exp \left( \frac{u - \mu}{\tau} \right) \right]^2},$$

com média  $E(U) = \mu$  e variância  $\sigma^2 = \text{Var}(U) = \pi^2\tau^2/3$ . Fazendo-se  $\beta_1 = -\mu/\tau$  e  $\beta_2 = 1/\tau$ , tem-se

$$f_U(u; \beta_1, \beta_2) = \frac{\beta_2 e^{\beta_1 + \beta_2 u}}{(1 + e^{\beta_1 + \beta_2 u})^2}.$$

Logo,

$$\pi_i = P(U \leq d_i) = F(d_i) = \frac{e^{\beta_1 + \beta_2 d_i}}{1 + e^{\beta_1 + \beta_2 d_i}}$$

é uma função não-linear em um conjunto linear de parâmetros, sendo linearizada por

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_1 + \beta_2 d_i.$$

### iii) Modelo complemento log-log

Nesse caso, considera-se que  $U$  tem distribuição de Gumbel (de valor extremo) com parâmetros  $\alpha$  e  $\tau$ , que é uma distribuição assimétrica, ao contrário das duas anteriores que são simétricas, e tem f.d.p. expressa por

$$f_U(u; \alpha, \tau) = \frac{1}{\tau} \exp\left(\frac{u - \alpha}{\tau}\right) \exp\left[-\exp\left(\frac{u - \alpha}{\tau}\right)\right], \quad \alpha \in \mathbb{R}, \quad \tau > 0,$$

com média  $E(U) = \alpha + \gamma\tau$  e variância  $\sigma^2 = \text{Var}(U) = \pi^2\tau^2/6$ , sendo  $\gamma \approx 0,577216$  o número de Euler definido por  $\gamma = -\psi(1) = \lim_{n \rightarrow \infty} (\sum_{i=1}^n i^{-1} - \log n)$ , em que  $\psi(p) = d \log \Gamma(p) / dp$  é a função digama. Fazendo-se  $\beta_1 = -\alpha/\tau$  e  $\beta_2 = 1/\tau$ , tem-se

$$f_U(u; \beta_1, \beta_2) = \beta_2 \exp(\beta_1 + \beta_2 u - e^{\beta_1 + \beta_2 u}).$$

Logo,

$$\pi_i = P(U \leq d_i) = F(d_i) = 1 - \exp[-\exp(\beta_1 + \beta_2 d_i)]$$

é uma função não-linear em um conjunto linear de parâmetros, sendo linearizada por

$$\log[-\log(1 - \pi_i)] = \beta_1 + \beta_2 d_i.$$

Nesses três exemplos, verifica-se que:

i) a distribuição dos  $Y_i$  (binomial) é um membro da família exponencial, com  $E(Y_i) = \mu_i = m_i \pi_i$ ;

ii) as variáveis explanatórias entram na forma de uma soma linear de seus efeitos sistemáticos, ou seja,

$$\eta_i = \sum_{j=1}^2 x_{ij} \beta_j = \mathbf{x}_i^T \boldsymbol{\beta},$$

sendo  $\mathbf{x}_i^T = (1, d_i)$ ,  $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$  e  $\eta_i$  o preditor linear.

iii) a média  $\mu_i$  é funcionalmente relacionada ao preditor linear, isto é,

$$\eta_i = g\left(\frac{\mu_i}{m_i}\right) = g(\pi_i),$$

que, nos casos analisados, são:

- modelo probito:  $\eta_i = g(\pi_i) = \Phi^{-1}(\pi_i)$ ;
- modelo logístico:  $\eta_i = g(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right)$ ;
- modelo complemento log-log:  $\eta_i = g(\pi_i) = \log[-\log(1 - \pi_i)]$ .

Portanto, esses modelos são baseados na família exponencial uniparamétrica (1.2) com médias que são não-lineares em um conjunto de parâmetros lineares, isto é,

- modelo probito:  $\mu_i = m_i \Phi(\beta_1 + \beta_2 d_i)$ ;
- modelo logístico:  $\mu_i = m_i \frac{e^{\beta_1 + \beta_2 d_i}}{1 + e^{\beta_1 + \beta_2 d_i}}$ ;
- modelo complemento log-log:  $\mu_i = m_i \{1 - \exp[-\exp(\beta_1 + \beta_2 d_i)]\}$ .

### 2.2.2 Ensaios de diluição

O uso dos ensaios de diluição é uma prática comum para se estimar a concentração  $\lambda$  de um organismo (número por unidade de volume, de área, de peso etc.) em uma amostra. Quando a contagem direta não é possível, mas a presença ou ausência do organismo em sub-amostras pode ser detectada (Ridout e Fenlon, 1998) pode-se, também, estimar  $\lambda$ . Em geral, registrar a presença, ou ausência, é mais econômico do que fazer a contagem. Por exemplo, pode-se detectar se uma determinada bactéria está presente, ou não, em um líquido por um teste de cor, ou se um fungo está presente, ou não, em uma amostra de solo, plantando-se uma planta susceptível nesse solo e verificando se a planta apresenta sintomas da doença. Esse método está baseado na suposição de que o número de indivíduos presentes segue uma distribuição de Poisson, o que é uma suposição forte e torna-se importante verificar se é verdadeira. Por exemplo, a distribuição espacial de um fungo no solo está longe de ser aleatória e pode ser que o número de indivíduos em diferentes amostras desse solo não tenha a distribuição de Poisson.

Nos ensaios de diluição, a solução original é diluída progressivamente e na  $i$ -ésima diluição são realizadas as contagens (Exemplo 2.2) ou, então, são testadas  $m_i, i = 1, 2, \dots, k$ , sub-amostras das quais  $Y_i$  apresentam resultado positivo para a presença do organismo (Exemplo 2.3). Seja  $\nu_i$  o volume da amostra original que está presente em cada uma das sub-amostras na  $i$ -ésima diluição. Em geral, mas nem sempre, são usadas diluições iguais, de modo que os  $\nu_i$  ficam em progressão geométrica.

Esse método de diluição em série é muito usado em diversas áreas da Biologia. Podem ser tratados, de forma semelhante, os problemas de estimação de:

- a) proporção de sementes doentes em um lote de sementes, em que  $n$  é o tamanho da amostra de sementes,  $\theta$  é a probabilidade de uma semente infectada e

$$\pi = P(\text{pelo menos uma semente doente}) = 1 - (1 - \theta)^n = 1 - e^{n \log(1-\theta)};$$

- b) proporção de um determinado tipo de célula em uma população em estudos de



imunologia;

- c) probabilidade de uma partícula de vírus matar um inseto, nos ensaios de controle biológico;
- d) taxa média de falha de um determinado componente quando os tempos de falha são distribuídos exponencialmente.

**Exemplo 2.2:** Na Tabela 2.2, são apresentados os dados referentes a contagens de partículas de vírus para cinco diluições diferentes, sendo que foram usadas quatro repetições para as quatro primeiras diluições e cinco repetições para a última diluição. O objetivo do experimento é estimar o número de partículas de vírus por unidade de volume.

Tabela 2.2: Números de partículas de vírus para cinco diluições diferentes.

Diluição	Contagens				
0,3162	13	14	17	22	
0,1778	9	14	6	14	
0,1000	4	4	3	5	
0,0562	3	2	1	3	
0,0316	2	1	3	2	2

Fonte: Ridout (1990), notas de aula

**Exemplo 2.3:** Na Tabela 2.3, são apresentados os dados de um ensaio de diluição realizado para determinar o número de esporos de *Bacillus mesentericus* por grama (g) de farinha de batata (Fisher e Yates, 1970). Uma suspensão líquida foi preparada e sujeita a sucessivas diluições para que resultassem soluções com 4, 2, ..., 1/128g de farinha por 100ml de solução. Para cada diluição, consideraram-se cinco amostras de 1ml e foi contado o número de amostras com esporos.

Tabela 2.3: Números de amostras ( $Y$ ) que contêm esporos em cinco amostras para diferentes quantidades (g) de farinha de batata em cada diluição.

g/100 ml	4	2	1	1/2	1/4	1/8	1/16	1/32	1/64	1/128
y	5	5	5	5	4	3	2	2	0	0

O parâmetro de interesse é  $\lambda$ , a concentração de organismos por unidade de volume ( $\nu_i$ ). Se os organismos estão aleatoriamente distribuídos, o número de organismos em uma sub-amostra da  $i$ -ésima diluição segue a distribuição de Poisson com média  $\lambda\nu_i$ , isto é,

$$\mu_i = \lambda\nu_i.$$

Assim, se forem realizadas contagens dos indivíduos após a diluição, tem-se que essa expressão pode ser linearizada, usando-se a função logarítmica, ou seja,

$$\eta_i = \log(\mu_i) = \log(\lambda) + \log(\nu_i) = \beta_1 + \text{offset}, \quad (2.2)$$

sendo que  $\log(\nu_i)$  entra na regressão como variável *offset*, que é um valor conhecido no componente sistemático do modelo.

Quando se observa o número de amostras em que o indivíduo está presente, tem-se  $Y_i \sim B(m_i, \pi_i)$ , desde que as sub-amostras de cada diluição sejam independentes, sendo que a probabilidade  $\pi_i$  de que o organismo esteja presente na sub-amostra  $i$  é expressa como

$$\pi_i = P(\text{pelo menos um organismo presente}) = 1 - \exp(-\lambda\nu_i).$$

Logo,

$$\eta_i = \log[-\log(1 - \pi_i)] = \log(\lambda) + \log(\nu_i) = \beta_1 + \text{offset}. \quad (2.3)$$

Tem-se, em (2.2) e (2.3), que  $\beta_1 = \log(\lambda)$  e  $\log(\nu_i)$  é a variável *offset*. Além disso, em (2.2) tem-se a função de ligação logarítmica para o modelo de Poisson

enquanto que em (2.3) tem-se a função de ligação complemento log-log para o modelo binomial.

Nesses dois exemplos, verifica-se, novamente, que:

- i) a distribuição dos  $Y_i$  (Poisson ou binomial) é um membro da família exponencial uniparamétrica (1.2), com  $E(Y_i) = \mu_i$  (Poisson) ou  $E(Y_i) = \mu_i = m_i \pi_i$  (binomial);
- ii) as variáveis explanatórias entram na forma de uma soma linear de seus efeitos, ou seja,

$$\eta_i = \sum_{j=1}^2 x_{ij} \beta_j = \mathbf{x}_i^T \boldsymbol{\beta},$$

sendo  $\mathbf{x}_i = (1, d_i)^T$ ,  $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$  e  $\eta_i$  o preditor linear.

- iii) a média  $\mu_i$  é funcionalmente relacionada ao preditor linear, isto é,

$$\eta_i = g(\mu_i) \quad \text{ou} \quad \eta_i = g\left(\frac{\mu_i}{m_i}\right) = g(\pi_i)$$

que nos casos analisados foram:

- modelo log-linear:  $\eta_i = g(\mu_i) = \log(\mu_i)$ ;
- modelo complemento log-log:  $\eta_i = g(\pi_i) = \log[-\log(1 - \pi_i)]$ .

Portanto, esses modelos são baseados na família exponencial uniparamétrica (1.2), cujas médias são não-lineares em um conjunto de parâmetros lineares, isto é,

- modelo log-linear:  $\mu_i = e^{\beta_1 + offset}$ ;
- modelo complemento log-log:  $\mu_i = m_i \{1 - \exp[-\exp(\beta_1 + offset)]\}$ ,

sendo  $\beta_2 = 1$  e  $\log(\nu_i) = offset$ .

### 2.2.3 Tabelas de contingência

Dados na forma de contagens são provenientes dos números de eventos (por exemplo, número de brotos por explante), ou então, da frequência de ocorrências

em várias categorias que originam as tabelas de contingência.

**Exemplo 2.4:** Os dados da Tabela 2.4 referem-se a coletas de insetos em armadilhas adesivas de duas cores, em que os indivíduos coletados de uma determinada espécie foram sexados, tendo como objetivo verificar se havia influência da cor da armadilha sobre a atração de machos e fêmeas dessa espécie.

Tabela 2.4: Números de insetos coletados em armadilhas adesivas e sexados.

Armadilha	Machos	Fêmeas	Totais
Alaranjada	246	17	263
Amarela	458	32	490
Totais	704	49	753

Fonte: Silveira Neto et al. (1976)

Tem-se que o número de insetos que chegam às armadilhas, seja do sexo feminino ou do sexo masculino, é um número aleatório, caracterizando uma observação de uma variável com distribuição de Poisson. A hipótese de interesse é a hipótese de independência, isto é, o sexo do inseto não afeta a escolha pela cor da armadilha, condicionado ao total observado.

**Exemplo 2.5:** Em um programa de melhoramento vegetal, um pesquisador coletou um total de 900 plantas de uma progênie da espécie X e classificou-as de acordo com duas respostas: tipo de ciclo (tardio ou precoce) e viriscência (normal ou viriscente), obtendo os resultados apresentados na Tabela 2.5.

Note que aqui, o total de plantas coletadas foi fixado previamente. O interesse, como no exemplo anterior, é estudar se não existe associação entre as duas respostas (hipótese de independência).

Tabela 2.5: Contagens de plantas segregando para dois caracteres: ciclo e viriscência, numa progênie da espécie X.

Ciclo	Normal	Viriscente	Totais
Precoce	453	30	483
Tardio	294	123	417
Totais	747	153	900

**Exemplo 2.6:** Os dados da Tabela 2.6 referem-se a um ensaio de controle da broca do tomate, usando-se quatro tratamentos. Tem-se aqui, também, um caso em que o

Tabela 2.6: Números de tomates sadios e com broca.

Inseticidas	Frutos		Totais
	Sadios	Com broca	
Diazinon	1690	115	1805
Phosdrin	1578	73	1651
Sevin	2061	53	2114
Testemunha	1691	224	1915
Totais	7020	465	7485

Fonte: Silveira Neto et al. (1976)

número total de frutos com broca é uma variável aleatória e, portanto, pode ser estudada pela distribuição de Poisson. A hipótese a ser testada é a da homogeneidade, isto é, a proporção de frutos sadios é a mesma para todos os inseticidas, condicionada aos totais observados de tratamentos.

A distribuição de Poisson é especialmente útil na análise de tabelas de contingência em que as observações consistem de contagens ou frequências nas caselas pelo cruzamento das variáveis resposta e explanatórias.

Considerando-se uma tabela de contingência bidimensional e a hipótese de independência, se  $y_{ij}$  representa o número de observações numa classificação cruzada de dois fatores  $i$  e  $j$  com  $I$  e  $J$  níveis, respectivamente, para  $i = 1, \dots, I$  e  $j = 1, \dots, J$ , então,

$$\mu_{ij} = E(Y_{ij}) = m\pi_{i+}\pi_{+j},$$

em que  $m = \sum_{i=1}^I \sum_{j=1}^J y_{ij}$  e  $\pi_{i+} = \sum_{j=1}^J \pi_{ij}$  e  $\pi_{+j} = \sum_{i=1}^I \pi_{ij}$  são as probabilidades marginais de uma observação pertencer às classes  $i$  e  $j$ , respectivamente. Pode-se, então, supor que  $Y_{ij}$  tem distribuição de Poisson com média  $\mu_{ij}$ .

Verifica-se, então, que uma função logarítmica lineariza esse modelo, isto é,

$$\eta_{ij} = \log(\mu_{ij}) = \log(m) + \log(\pi_{i+}) + \log(\pi_{+j}) = \mu + \alpha_i + \beta_j.$$

Novamente, tem-se:

- i) a distribuição de  $Y_{ij}$  (Poisson) é um membro da família exponencial, com  $E(Y_{ij}) = \mu_{ij}$ ;
- ii) as variáveis explanatórias entram na forma de uma soma linear de seus efeitos, ou seja,

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta},$$

sendo  $\boldsymbol{\eta} = (\eta_{11}, \dots, \eta_{1J}, \dots, \eta_{I1}, \dots, \eta_{IJ})^T$  o preditor linear,  $\mathbf{X}$  uma matriz, de dimensões  $IJ \times (I + J + 1)$ , de variáveis “dummy” e  $\boldsymbol{\beta} = (\mu, \alpha_1, \dots, \alpha_I, \beta_1, \dots, \beta_J)^T$ ;

- iii) a média é funcionalmente relacionada ao preditor linear, isto é,

$$\eta_{ij} = g(\mu_{ij}) = \log(\mu_{ij}).$$

Portanto, tem-se que esses modelos são baseados na família exponencial uniparamétrica (1.2), cujas médias são não-lineares em um conjunto de parâmetros lineares, ou seja,  $\boldsymbol{\mu} = \exp(\boldsymbol{\eta}) = \exp(\mathbf{X}^T\boldsymbol{\beta})$ .

De forma geral, verifica-se que as hipóteses mais comuns para dados dispostos em tabelas de contingência podem ser expressas como modelos multiplicativos para as frequências esperadas das caselas (McCullagh e Nelder, 1989; Agresti, 2002; Paulino e Singer, 2006) e que a média  $\mu$  é obtida como um produto de outras médias marginais. Assim, uma transformação logarítmica do valor esperado lineariza essa parte do modelo.

## 2.3 Definição

Os MLGs podem ser usados quando se tem uma única variável aleatória  $Y$  associada a um conjunto de variáveis explanatórias  $x_1, \dots, x_p$ . Para uma amostra de  $n$  observações  $(y_i, \mathbf{x}_i)$ , em que  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$  é o vetor coluna de variáveis explanatórias, o MLG envolve os três componentes:

- i) **Componente aleatório:** representado por um conjunto de variáveis aleatórias independentes  $Y_1, \dots, Y_n$  obtidas de uma mesma distribuição que faz parte da família de distribuições (1.5) com médias  $\mu_1, \dots, \mu_n$ , ou seja,

$$E(Y_i) = \mu_i, \quad i = 1, \dots, n,$$

sendo  $\phi > 0$  um parâmetro de dispersão e  $\theta_i$  o parâmetro denominado canônico. Então, a função densidade ou de probabilidade de  $Y_i$  é expressa por

$$f(y_i; \theta_i, \phi) = \exp \left\{ \phi^{-1} [y_i \theta_i - b(\theta_i)] + c(y_i, \phi) \right\}, \quad (2.4)$$

sendo  $b(\cdot)$  e  $c(\cdot)$  funções conhecidas. Conforme foi explicado na Seção 1.4,

$$E(Y_i) = \mu_i = b'(\theta_i) \quad \text{e} \quad \text{Var}(Y_i) = \phi b''(\theta_i) = \phi V_i,$$

em que  $V_i = V(\mu_i) = d\mu_i/d\theta_i$  é denominada de função de variância que depende unicamente da média  $\mu_i$ . O parâmetro natural  $\theta_i$  pode ser expresso como

$$\theta_i = \int V_i^{-1} d\mu_i = q(\mu_i), \quad (2.5)$$

sendo  $q(\mu_i)$  uma função conhecida da média  $\mu_i$ . Supondo uma relação funcional para a função de variância  $V(\mu)$ , o parâmetro canônico é obtido da equação (2.5) e a distribuição é univocamente determinada na família exponencial (2.4). A importância da família (2.4) na teoria dos MLGs é que ela permite analisar dados que exibem assimetria, dados de natureza discreta ou contínua e dados que são restritos a um intervalo do conjunto dos reais, como o intervalo  $(0,1)$ .

- ii) **Componente sistemático:** as variáveis explanatórias entram na forma de uma soma linear de seus efeitos

$$\eta_i = \sum_{r=1}^p x_{ir}\beta_r = \mathbf{x}_i^T \boldsymbol{\beta} \quad \text{ou} \quad \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}, \quad (2.6)$$

sendo  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$  a matriz do modelo,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$  o vetor de parâmetros desconhecidos e  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^T$  o preditor linear. Se um parâmetro tem valor conhecido, o termo correspondente na estrutura linear é denominado *offset*, como verificado nos ensaios de diluição (Seção 2.2).

- iii) **Função de ligação:** uma função que relaciona o componente aleatório ao componente sistemático, ou seja, vincula a média ao preditor linear, isto é,

$$\eta_i = g(\mu_i), \quad (2.7)$$

sendo  $g(\cdot)$  uma função monótona e diferenciável.

Assim, verifica-se que para a especificação do modelo, os parâmetros  $\theta_i$  da família de distribuições (2.4) não são de interesse direto (pois há um para cada observação) mas sim um conjunto menor de parâmetros  $\beta_1, \dots, \beta_p$  tais que uma combinação linear dos  $\beta_r$  seja igual à alguma função do valor esperado de  $Y_i$ . Como o parâmetro natural  $\theta_i$  é uma função unívoca da média  $\mu_i$ , pode-se expressar a função de ligação em termos desse parâmetro, isto é,  $\eta_i = g(q^{-1}(\theta_i))$ .

Portanto, uma decisão importante na escolha do MLG é definir os termos do trinômio: (i) distribuição da variável resposta; (ii) matriz do modelo e (iii) função



de ligação. Nesses termos, um MLG é definido por uma distribuição da família (2.4), uma estrutura linear (2.6) e uma função de ligação (2.7). Por exemplo, quando  $\theta = \mu$  e a função de ligação é linear, obtém-se o modelo clássico de regressão como um caso particular. Os modelos log-lineares são deduzidos supondo  $\theta = \log(\mu)$  com função de ligação logarítmica  $\log(\mu) = \eta$ . Torna-se clara, agora, a palavra “generalizado”, significando uma distribuição mais ampla do que a normal para a variável resposta, e uma função não-linear em um conjunto linear de parâmetros conectando a média dessa variável com a parte determinística do modelo.

Observa-se que na definição de um MLG por (2.4), (2.6) e (2.7) não existe, em geral, aditividade entre a média  $\mu$  e o erro aleatório  $\epsilon$  inerente ao experimento, como ocorre no modelo clássico de regressão descrito na Seção 2.1. Define-se no MLG uma distribuição para a variável resposta que representa as observações e não uma distribuição para o erro aleatório  $\epsilon$ .

A escolha da distribuição em (2.4) depende, usualmente, da natureza dos dados (discreta ou contínua) e do seu intervalo de variação (conjunto dos reais, reais positivos ou um intervalo como  $(0,1)$ ). Na escolha da matriz do modelo  $\mathbf{X} = \{x_{ir}\}$ , de dimensões  $n \times p$  e suposta de posto completo,  $x_{ir}$  pode representar a presença ou ausência de um nível de um fator classificado em categorias, ou pode ser o valor de uma covariável. A forma da matriz do modelo representa, matematicamente, o desenho do experimento. A escolha da função de ligação depende do problema em particular e, pelo menos em teoria, cada observação pode ter uma função de ligação diferente.

As funções de ligação usuais são: potência  $\eta = \mu^\lambda$ , em que  $\lambda$  é um número real, logística  $\eta = \log[\mu/(m - \mu)]$ , probito  $\eta = \Phi^{-1}(\mu/m)$ , sendo  $\Phi(\cdot)$  a função de distribuição acumulada (f.d.a.) da distribuição normal padrão e a complemento log-log  $\eta = \log[-\log(1 - \mu/m)]$ , em que  $m$  é o número de ensaios independentes. As três últimas funções de ligação são apropriadas para o modelo binomial, pois transformam o intervalo  $(0,1)$  em  $(-\infty, +\infty)$  (Exercício 1). Casos importantes da função de ligação potência são identidade, recíproca, raiz quadrada e logarítmica,

correspondentes, a  $\lambda = 1, -1, 1/2$  e  $0$ , respectivamente.

Se a função de ligação é escolhida de modo que  $g(\mu_i) = \theta_i = \eta_i$ , o preditor linear modela diretamente o parâmetro canônico  $\theta_i$ , sendo denominada **função de ligação canônica**. Os modelos correspondentes são denominados **canônicos**. Isso resulta, frequentemente, em uma escala adequada para a modelagem com interpretação prática para os parâmetros de regressão, além de vantagens teóricas em termos da existência de um conjunto de estatísticas suficientes para o vetor de parâmetros  $\beta$  e alguma simplificação no algoritmo de estimação. O vetor de estatísticas suficientes para  $\beta$  é  $\mathbf{T} = \mathbf{X}^T \mathbf{Y}$ , com componentes  $T_r = \sum_{i=1}^n x_{ir} Y_i$ ,  $r = 1, \dots, p$ . As funções de ligação canônicas para as principais distribuições estão apresentadas na Tabela 2.7.

Tabela 2.7: Funções de ligação canônicas.

Distribuição	Função de ligação canônica
Normal	Identidade: $\eta = \mu$
Poisson	Logarítmica: $\eta = \log(\mu)$
Binomial	Logística: $\eta = \log\left(\frac{\pi}{1-\pi}\right) = \log\left(\frac{\mu}{m-\mu}\right)$
Gama	Recíproca: $\eta = \frac{1}{\mu}$
Normal inversa	Recíproca do quadrado: $\eta = \frac{1}{\mu^2}$

Deve ser enfatizado que as funções de ligação canônicas conduzem a propriedades estatísticas desejáveis para o modelo, principalmente, no caso de amostras reduzidas. Entretanto, não há nenhuma razão a priori para que os efeitos sistemáticos do modelo sejam aditivos na escala especificada por tais funções. Para o modelo clássico de regressão, a função de ligação canônica é a identidade, pois o preditor linear é igual à média. Essa função de ligação é adequada no sentido em que ambos,  $\eta$  e  $\mu$ , têm valores na reta real. Entretanto, certas restrições surgem quando se trabalha, por exemplo, com a distribuição de Poisson em que  $\mu > 0$  e, portanto, a

função de ligação identidade não deve ser usada, pois  $\hat{\mu}$  poderá ter valores negativos, dependendo dos valores obtidos para  $\hat{\beta}$ . Além disso, dados de contagens dispostos em tabelas de contingência, sob a suposição de independência, conduzem, naturalmente, a efeitos multiplicativos cuja linearização pode ser obtida por meio da função de ligação logarítmica, isto é,  $\eta = \log(\mu)$  e, portanto,  $\mu = e^\eta$  (conforme descrito nos ensaios de diluição da Seção 2.2).

Aranda-Ordaz (1981) propôs a família de funções de ligação para análise de dados na forma de proporções expressa por

$$\eta = \log \left[ \frac{(1 - \pi)^{-\lambda} - 1}{\lambda} \right],$$

sendo  $\lambda$  uma constante desconhecida que tem como casos especiais as funções de ligação logística para  $\lambda = 1$  e complemento log-log quando  $\lambda \rightarrow 0$ .

Uma família importante de funções de ligação, principalmente para dados com média positiva, é a família potência (Exercício 2), especificada por

$$\begin{cases} \frac{\mu^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log \mu & \lambda = 0 \end{cases}$$

ou então,

$$\begin{cases} \mu^\lambda & \lambda \neq 0 \\ \log \mu & \lambda = 0 \end{cases},$$

sendo  $\lambda$  uma constante desconhecida.

## 2.4 Modelos especiais

### 2.4.1 Modelo clássico de regressão

A distribuição normal foi, primeiramente, introduzida por Abraham de Moivre em 1733 como limite da distribuição binomial. A distribuição normal foi, também, deduzida por Laplace em 1774 como uma aproximação para a distribuição hipergeométrica. Em 1778, Laplace tabulou a f.d.a.  $\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^x e^{-t^2/2} dt$  da distribuição normal padronizada. Gauss, em dois artigos publicados em 1809 e

1816, estabeleceu técnicas baseadas na distribuição normal que se tornaram métodos corriqueiros durante o século XIX. No seu artigo de 1816, Gauss deduziu a distribuição normal como a distribuição limite da soma de um número muito grande de erros independentes, podendo assim ser considerado um dos resultados mais antigos do teorema central do limite. Na Seção 1.3 (Exemplo 1.4) é apresentada a f.d.p. da distribuição normal. Existem várias aproximações para calcular a f.d.a.  $\Phi(x)$  da distribuição normal padronizada, vide, por exemplo, Johnson et al. (2004).

A função geradora de momentos da distribuição normal é  $M(t; \mu, \sigma^2) = \exp(\mu t + \sigma^2 t^2/2)$ , sendo, então, seus cumulantes  $\kappa_r = 0$ , para  $r > 2$ . Entre outras características, citam-se: média, moda e mediana iguais a  $\mu$ , coeficientes de assimetria e curtose iguais a 0 e 3, respectivamente,  $r$ -ésimo momento central igual a 0, se  $r$  é ímpar, e  $\frac{\sigma^r r!}{2^{r/2}} \left(\frac{r}{2}\right)!$ , se  $r$  é par.

As origens do modelo clássico de regressão estão nos trabalhos de astronomia de Gauss de 1809 e 1821. O método de mínimos quadrados foi desenvolvido por Legendre em 1805 e por Gauss em 1809 para determinar a órbita do asteroide Ceres. As ideias de obtenção da matriz modelo nos planejamentos dos experimentos surgiram na Estação Experimental de Rothamsted, Inglaterra, com Fisher (1920 a 1935).

O modelo normal  $N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$  para o vetor  $\mathbf{Y}$  da variável resposta, em que  $\mathbf{I}$  é a matriz identidade, é usado na análise de variância com efeitos fixos, como modelo amostral e, mais comumente, como um modelo aproximado para uma distribuição desconhecida. É o caso mais simples do MLG correspondendo a  $\eta = \theta = \mu$ .

Embora a estimação por máxima verossimilhança seja estudada na Seção 3.2, convém salientar que no modelo clássico de regressão, o estimador de máxima verossimilhança (EMV) de  $\boldsymbol{\beta}$ , que coincide com o de mínimos quadrados, é obtido em forma explícita por  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ . A função de verossimilhança só depende dos dados por meio de  $\hat{\boldsymbol{\beta}}$  e da soma de quadrados dos resíduos  $SQR = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$ . Mostra-se que  $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1})$  e  $SQR \sim \sigma^2\chi_{n-p}^2$ . Os testes para os componentes de  $\boldsymbol{\beta}$  são realizados, exatamente, usando-se estatísticas com distribuições  $\chi^2$

e  $F$ .

### 2.4.2 Modelo de Poisson

Em 1837, Poisson publicou a distribuição que tem seu nome, obtendo-a como uma distribuição limite da distribuição binomial. Se a variável aleatória  $Y$  tem distribuição de Poisson,  $P(\mu)$ , com parâmetro  $\mu > 0$ , então sua função de probabilidade é expressa por

$$f(y; \mu) = \frac{e^{-\mu} \mu^y}{y!}, \quad \text{para } y = 0, 1, 2, \dots$$

A f.g.m. é igual a  $M(t; \mu) = \exp\{\mu[\exp(t) - 1]\}$ , sendo todos os cumulantes iguais a  $\mu$  e o  $r$ -ésimo momento central  $\mu_r$  ( $r \geq 2$ ) pode ser calculado pela fórmula de recorrência  $\mu_r = \mu \sum_{i=0}^{r-1} \binom{r-1}{i} \mu_i$ , com  $\mu_0 = 1$ . A moda corresponde ao maior inteiro menor do que  $\mu$  e, para  $\mu$  inteiro, tem valores  $\mu$  e  $\mu - 1$ . Os coeficientes de assimetria e curtose são iguais a  $\mu^{-1/2}$  e  $3 + \mu^{-1}$ , respectivamente. O  $r$ -ésimo momento fatorial é igual a  $E[Y(Y-1)\dots(Y-r+1)] = \mu^r$ .

Quando  $\mu \rightarrow \infty$ , tem-se  $(Y - \mu)\mu^{-1/2} \sim N(0, 1) + O_p(\mu^{-1/2})$ . Em geral, para  $\mu > 9$ , a aproximação da distribuição de Poisson  $P(\mu)$  pela distribuição normal  $N(\mu, \mu)$  é satisfatória. Probabilidades individuais podem ser avaliadas pela expressão aproximada  $P(Y = y) = \Phi(y_2) - \Phi(y_1)$ , sendo  $\Phi(\cdot)$  a f.d.a. da distribuição normal padronizada,  $y_2 = (y - \mu + 0.5)\mu^{-1/2}$  e  $y_1 = (y - \mu - 0.5)\mu^{-1/2}$ . O resultado bastante conhecido  $P(Y \leq y) = P(\chi_{2(1+y)}^2 > 2\mu)$  é, muitas vezes, útil no cálculo da f.d.a. da distribuição de Poisson.

Uma fórmula alternativa aproximada para calcular a f.d.a. da distribuição de Poisson, baseada na f.d.a. da distribuição normal padrão, é  $P(Y \leq y) \approx \Phi[g(y-0.5)]$ , em que

$$g(y) = \begin{cases} 3y^{1/2} - 3y^{1/6}\mu^{1/3} + \mu^{-1/2}/6, & y \neq 0; \\ -(2\mu)^{1/2} + \mu^{-1/2}/6, & y = 0. \end{cases}$$

O modelo de Poisson tem um importante papel na análise de dados em forma de contagens. Suas principais características são:

- a) proporciona, em geral, uma descrição satisfatória de dados experimentais cuja variância é proporcional à média;
- b) pode ser deduzido teoricamente de princípios elementares com um número mínimo de restrições;
- c) se eventos ocorrem independente e aleatoriamente no tempo, com taxa média de ocorrência constante, o modelo determina o número de eventos em um intervalo de tempo especificado.

O modelo de regressão de Poisson desempenha, na análise de dados categorizados, o mesmo papel do modelo normal, na análise de dados contínuos. A diferença fundamental é que a estrutura multiplicativa para as médias do modelo de Poisson é mais apropriada do que a estrutura aditiva das médias do modelo normal. Tem-se constatado, na análise de dados categorizados, que a média  $\mu$  é, geralmente, obtida como um produto de outras médias marginais que se tornam os parâmetros lineares do modelo. A estrutura linear adotada é expressa, na escala do parâmetro canônico da distribuição, por  $\log(\mu) = \eta$ , com os parâmetros  $\beta$  medindo efeitos sobre a escala logarítmica das frequências esperadas. Por exemplo, como já foi mostrado na Seção 2.2, item c, independência de dois fatores numa tabela de contingência  $r \times s$  equivale ao modelo  $\mu_{ij} = \mu_{i+}\mu_{+j}/\mu_{++}$ , com a notação usual para a soma. Isso implica que o logaritmo de  $\mu_{ij}$  é expresso como uma estrutura linear formada pelos efeitos principais dos fatores sem a interação.

O modelo log-linear é definido pela distribuição de Poisson,  $P(\mu)$ , com  $\log(\mu) = \eta = \mathbf{X}\beta$ , sendo um dos casos especiais dos MLGs de maior importância, pelo seu papel na análise de dados categorizados dispostos em tabelas de contingência. Pode-se supor que a tabela de contingência é decorrente de um modelo de Poisson, multinomial ou produto-multinomial, dependendo do planejamento adotado. Para os dois últimos modelos, demonstra-se que isso equivale a um conjunto de distribuições condicionadas de Poisson com a suposição do total das frequências observadas ser fixo (Seção 7.2.2).

Pode-se transformar  $Y$  na forma de contagens e, a seguir, definir modelos alternativos para os dados transformados. Geralmente, usa-se a transformação  $Y^{1/2}$  que estabiliza a variância supondo  $\mu$  grande, ou trata-se  $Y^{2/3}$  como, aproximadamente, normal (Seção ??). Entretanto, nesses casos, ignora-se a natureza discreta dos dados.

### 2.4.3 Modelo binomial

A distribuição binomial foi descrita por James Bernoulli em 1713, embora tenha sido encontrada anteriormente em trabalhos de Pascal. Suponha que  $Y = mP$  tenha distribuição binomial  $B(m, \pi)$ , com função de probabilidade especificada no Exemplo 1.2, sendo que  $P$  representa a proporção de sucessos em  $m$  ensaios independentes com probabilidade de sucesso  $\pi$ . A f.g.m. de  $Y$  é expressa por  $M(t; \pi, m) = \{\pi[\exp(t) - 1] + 1\}^m$  e seus momentos centrais,  $\mu_{2r}$  e  $\mu_{2r+1}$ , são de ordem  $O(m^r)$ , para  $r = 1, 2, \dots$ . O  $r$ -ésimo momento central de  $P$  é, simplesmente,  $m^{-r}\mu_r$ . Todos os cumulantes de  $Y$  são de ordem  $O(m)$  e, portanto,

$$\frac{Y - m\pi}{[m\pi(1 - \pi)]^{1/2}} \sim N(0, 1) + O_p(m^{-1/2}),$$

sendo a taxa de convergência expressa pelo terceiro cumulante padronizado. A moda de  $Y$  pertence ao intervalo  $[(m+1)\pi - 1, (m+1)\pi]$ , e os seus coeficientes de assimetria e curtose são, respectivamente,

$$\frac{(1 - 2\pi)}{[m\pi(1 - \pi)]^{1/2}} \text{ e } 3 - \frac{6}{m} + \frac{1}{m\pi(1 - \pi)}.$$

Quando  $m\pi > 5$  e  $0,1 \leq \pi \leq 0,9$ , ou  $m\pi > 25$ , sendo  $\pi$  qualquer, o modelo binomial  $B(m, \pi)$  pode ser aproximado pelo modelo normal  $N(m\pi, m\pi(1 - \pi))$ . Uma melhor aproximação é obtida de  $P(Y \leq y) = \Phi(y_1) + \phi(y_1)/\{2[m\pi(1 - \pi)]^{1/2}\}$ , em que  $y_1 = (y - m\pi)/[m\pi(1 - \pi)]$  e  $\phi(\cdot)$  é a f.d.p. da distribuição normal padrão, cujo erro é inferior a  $(0,2 + 0,25 | 1 - 2\pi |)/[m\pi(1 - \pi)] + \exp\{-1,5[m\pi(1 - \pi)]^{-1/2}\}$ , se  $m\pi(1 - \pi) \geq 25$ . A aproximação normal com correção de continuidade  $P(Y \leq y) = \Phi(y_2)$ , em que  $y_2 = (y + 0,5 - m\pi)/[m\pi(1 - \pi)]^{1/2}$ , tem erro menor do que  $0,140[m\pi(1 - \pi)]^{-1/2}$  (Cordeiro, 1986).

Se  $y = mp$  é inteiro, um número de aproximações para as probabilidades binomiais são baseadas na equação

$$\begin{aligned} P(Y \geq y) &= \sum_{i=y}^m \binom{m}{i} \pi^i (1-\pi)^{m-i} \\ &= B(y, m-y+1)^{-1} \int_0^\pi t^{y-1} (1-t)^{m-y} dt = I_\pi(y, m-y+1), \end{aligned}$$

em que  $I_\pi(y, m-y+1)$  representa a função razão beta incompleta.

Pode-se, ainda, usar a aproximação da distribuição binomial pela distribuição de Poisson  $P(m\pi)$  quando  $\pi < 0,1$ , o erro da aproximação sendo  $O(m^{-1})$ , ou, então, a fórmula  $P(Y \leq y) = 1 - P\{F[2(y+1), 2(m-y)] < \pi(m-y)/[(1+y)(1-\pi)]\}$ , em que  $F[2(y+1), 2(m-y)]$  representa a distribuição F de Snedecor com  $2(y+1)$  e  $2(m-y)$  graus de liberdade.

Adicionalmente, sejam  $B(y) = \binom{m}{y} \pi^y (1-\pi)^{m-y}$  e  $P(y) = \frac{e^{-\mu} \mu^y}{y!}$ , as probabilidades pontuais das distribuições binomial e de Poisson, respectivamente. Supondo  $\mu = m\pi$  e  $\mu$  fixo, pode-se mostrar, com base na aproximação de Stirling para o fatorial, que quando  $m-y \rightarrow \infty$ ,

$$\frac{B(y)}{P(y)} \approx \left( \frac{m}{m-y} \right)^{1/2}.$$

Esse resultado, também, pode ser comprovado numericamente.

O modelo binomial é usado, principalmente, no estudo de dados na forma de proporções, como nos casos da análise probito (Finney, 1971), logística (ou “logit”) (Ashton, 1972) e complemento log-log (Fisher, 1922) (Seção 2.2), e na análise de dados binários, como na regressão logística linear (Cox, 1970).

#### 2.4.3.1 Dados na forma de proporções

Considera-se o modelo binomial para o estudo de dados na forma de proporções em que são aplicadas doses de uma droga a  $n$  conjuntos de indivíduos, sendo  $m_i$  o número de indivíduos testados no conjunto  $i$ ,  $i = 1, \dots, n$ . Conforme descrito na Seção 2.2, o sucesso de um teste é determinado por uma variável latente  $U$ , denominada *tolerância*, com f.d.a. especificada como  $F(\cdot)$ . Os indivíduos do conjunto



$i$  recebem uma dose fixa  $x_i$  da droga e a probabilidade de sucesso correspondente é expressa como  $\pi_i = P(U \leq x_i) = F(\alpha + \beta x_i)$ , em que  $\alpha$  e  $\beta$  são parâmetros desconhecidos que dependem dos parâmetros da distribuição proposta para  $U$ .

Sejam  $P_1, \dots, P_n$  as proporções de sucessos, supostas independentes, nos conjuntos  $1, \dots, n$ . O modelo para o estudo dessas proporções, no contexto dos MLGs, tem variável resposta  $Y_i = m_i P_i$  com distribuição binomial, função de ligação  $F^{-1}(\cdot)$  e estrutura linear  $\eta_i = \alpha + \beta x_i$ . Convém salientar que é postulada uma relação linear entre alguma função de  $\mu$  e  $x$ , ao invés de uma função de  $P$  e  $x$ . A variância da variável resposta não é constante, como no modelo clássico de regressão, e depende do valor da média.

Vários casos especiais desse modelo binomial são obtidos pela definição da distribuição da tolerância, conforme explicado na Seção 2.2. Se se supõe que a tolerância tem distribuição normal, o modelo correspondente  $\pi_i = \Phi(\alpha + \beta x_i)$  é denominado probito (Finney, 1971). Se se supõe que tem distribuição logística, o modelo  $\pi_i = \exp(\alpha + \beta x_i) / [1 + \exp(\alpha + \beta x_i)]$  é denominado logístico (Berkson, 1944), e quando tem distribuição de valor extremo (Gumbel), a função de ligação  $F^{-1}(\cdot)$  corresponde ao modelo complemento log-log. O modelo logístico, postulando uma regressão linear para  $\log[\pi/(1 - \pi)]$  (“log-odds”), tem sido muito usado na área de Medicina, pois tem uma interpretação simples, enquanto que o probito é o mais usado na área de Entomologia, por influência do artigo de Bliss (1935).

Existe pouca diferença entre as distribuições normal e logística para a tolerância, e, quando essas são re-escaladas adequadamente, por exemplo, para terem as médias e os desvios-padrão iguais, tornam-se bastante similares no intervalo de 0,1 a 0,9. Por essa razão, é, geralmente, difícil diferenciá-las com base no ajuste do modelo. As funções de ligação logística e probito são simétricas em relação ao ponto de inflexão, isto é,  $F^{-1}(\pi) = -F^{-1}(1 - \pi)$ , o que não ocorre com a função de ligação complemento log-log. Essa última função de ligação é mais apropriada para análise de dados sobre incidência de doenças. Para valores de  $\mu$  próximos de zero, as funções de ligação complemento log-log e logística são equivalentes. A família

de funções de ligação de Aranda-Ordaz (1981) com um parâmetro especificada por  $g(\mu; \lambda) = \log\{[(1 - \mu)^{-\lambda} - 1]/\lambda\}$  contém a função de ligação logística ( $\lambda = 1$ ) e a complemento log-log ( $\lambda = 0$ ).

#### 2.4.3.2 Dados binários agrupados

Sejam  $n$  variáveis aleatórias binárias,  $R_1, \dots, R_n$ , tendo somente os valores 0 e 1, classificadas em  $t$  grupos, o grupo  $i$  com  $m_i$  variáveis independentes com probabilidade de sucesso (resposta igual a 1) associada  $\pi_i$ ,  $i = 1, \dots, t$ , sendo  $\sum_{i=1}^t m_i = n$ . Definem-se  $Y_i$  e  $P_i$  como o número e a proporção de sucessos no grupo  $i$ , respectivamente, em que  $Y_i = m_i P_i$  tem distribuição binomial  $B(m_i, \pi_i)$ ,  $i = 1, \dots, t$ . O modelo para experimentos com respostas binárias não-agrupadas corresponde ao caso especial  $m_i = 1$  e  $n = t$ .

O modelo para  $m_i P_i$  com distribuição binomial  $B(m_i, \pi_i)$  e função de ligação  $g(\pi_i) = g(\mu_i/m_i) = \eta_i = \sum_{r=1}^p x_{ir} \beta_r$  pertence à classe dos MLGs devendo a função de ligação ser uma função do intervalo  $(0, 1)$  na reta real. O modelo logístico linear tem como função de ligação  $g(\pi_i) = g(\mu_i/m_i) = \log[\pi_i/(1 - \pi_i)] = \log[\mu_i/(m_i - \mu_i)]$ .

Um modelo alternativo para análise de dados binários agrupados é formulado por variáveis aleatórias independentes  $Z_i = g(Y_i/m_i)$ ,  $i = 1, \dots, t$ . A variável  $Z_i$  tem, aproximadamente, distribuição normal de média  $g(\pi_i)$  e variância  $g'(\pi_i)^2 \pi_i (1 - \pi_i)/m_i$ , desde que  $m_i \rightarrow \infty$  e que  $\pi_i$  não seja próximo de 0 ou 1. Essa variância é, consistentemente, estimada por  $v_i = g'(p_i)^2 p_i (1 - p_i)/m_i$ , substituindo  $\pi_i$  pelo valor amostral  $p_i$  de  $P_i$ .

Considera-se  $\mathbf{z} = (z_1, \dots, z_t)^T$ , em que  $z_i = g(p_i)$ , como realizações de variáveis aleatórias com médias  $E(\mathbf{Z}) = \mathbf{X}\boldsymbol{\beta}$  e estrutura de covariância aproximada  $\mathbf{V} = \text{diag}\{v_1, \dots, v_t\}$ , sendo  $\mathbf{X}$  a matriz do modelo de dimensões  $t \times p$  e  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ . Se não ocorrerem proporções de sucessos iguais a 0 ou 1, o método de mínimos quadrados ponderados, que equivale a minimizar  $(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})$  em relação a  $\boldsymbol{\beta}$ , produzirá o estimador  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{z}$ . Esse estimador é

diferente do EMV de  $\beta$ . Nesse modelo alternativo, testes e regiões de confiança para os parâmetros são obtidos como no contexto do modelo clássico de regressão.

Escolhendo a função de ligação  $g(\cdot)$  como a logística, tem-se  $Z_i = \log[Y_i/(m_i - Y_i)]$ , denominada *transformação logística empírica* de  $Y_i/m_i$ , sendo  $\text{Var}(Z_i)$  estimada por  $m_i/[Y_i(m_i - Y_i)]$ . Uma transformação mais adequada é obtida acrescentando-se 0,5 ao numerador e ao denominador, implicando em

$$Z_i = \log \left( \frac{Y_i + 0,5}{m_i - Y_i + 0,5} \right),$$

pois  $E(Z_i) = \log[\pi_i/(1 - \pi_i)] + O(m_i^{-2})$ , além de ser definida para proporções de sucessos iguais a zero e um. Um estimador não-viesado de  $\text{Var}(Z_i)$  é igual a

$$v_i = \frac{(m_i + 1)(m_i + 2)}{m_i(Y_i + 1)(m_i - Y_i + 1)}.$$

Escolhendo a função de ligação arco seno, tem-se  $Z_i = \arcsen(\sqrt{Y_i/m_i})$ , denominada “transformação angular empírica” que, aproximadamente, estabiliza a variância para  $m_i$  grande. A média e a variância de  $Z_i$  são, aproximadamente, iguais a  $\arcsen(\sqrt{\pi_i})$  e  $1/(4m_i)$ , respectivamente.

#### 2.4.4 Modelo gama

Suponha que  $Y$  tem distribuição gama,  $G(\mu, \phi)$ , com parâmetros positivos  $\mu$  e  $\phi$ , isto é, com f.d.p. expressa por

$$f(y; \mu, \phi) = \frac{\left(\frac{\phi}{\mu}\right)^\phi}{\Gamma(\phi)} y^{\phi-1} \exp\left(-\frac{\phi y}{\mu}\right), \quad y > 0,$$

sendo a média  $\mu$  e o coeficiente de variação igual a  $\sqrt{\phi}$ . Tem-se: f.g.m.  $M(t; \mu, \phi) = (1 - \mu\phi t)^{-\phi^{-1}}$ , se  $t > (\phi\mu)^{-1}$ ,  $r$ -ésimo momento central  $(\mu\phi)^r \prod_{j=0}^{r-1} (j + \phi^{-1})$ ,  $r$ -ésimo cumulante  $(r-1)!\phi^{r-1}\mu^r$  e coeficientes de assimetria e curtose iguais a  $2\sqrt{\phi}$  e  $3 + 6\phi$ , respectivamente. Logo, o modelo gama  $G(\mu, \phi)$  tem o modelo normal como limite quando o parâmetro de dispersão  $\phi \rightarrow 0$ . A moda da distribuição é igual a  $\mu(1 - \phi)$  para  $\phi \leq 1$  e, se  $\phi > 1$ , a função densidade da distribuição gama decresce quando  $y$  cresce.

Se a variável aleatória  $Y$  tem distribuição gama  $G(\mu, \phi)$ , a sua f.d.a. pode ser expressa como

$$P(Y \leq x) = \frac{\gamma(\phi, \phi\mu^{-1}x)}{\Gamma(\phi)},$$

em que  $\gamma(\phi, y) = \int_0^y t^{\phi-1}e^{-t}dt$  é a função gama incompleta. A função  $\Gamma(\phi) = \int_0^\infty t^{\phi-1}e^{-t}dt$  é a função gama. Essas funções estão disponíveis nos principais software estatísticos e encontradas, também, em <http://mathworld.wolfram.com>.

O modelo gama é usado na análise de dados contínuos não-negativos que apresentam uma variância crescente com a média e, principalmente, quando o coeficiente de variação dos dados for, aproximadamente, constante. É, também, aplicado na estimação de componentes de variância de modelos com efeitos aleatórios, e como uma distribuição aproximada de medições físicas, tempos de sobrevivência, etc.

Uma aplicação do modelo gama é na análise de variância com efeitos aleatórios, em que as somas de quadrados, supondo que a variável resposta tem distribuição normal, são proporcionais a variáveis qui-quadrados. Sejam  $k$  somas de quadrados  $SQ_1, \dots, SQ_k$  independentes, tais que  $SQ_i \sim \eta_i \chi_{\nu_i}^2$ , em que  $\nu_i$  é o número de graus de liberdade associado a  $SQ_i$  e  $\eta_i = \sum_{j=1}^p x_{ij}\sigma_j^2$  é uma constante de proporcionalidade expressa como uma combinação linear de  $p$  variâncias desconhecidas  $\sigma_1^2, \dots, \sigma_p^2$ . Como os quadrados médios  $QM_i = SQ_i/\nu_i$  têm distribuição  $(\eta_i/\nu_i)\chi_{\nu_i}^2$ , pode-se considerar  $QM_i, i = 1, \dots, k$ , representando a variável resposta, no contexto dos MLGs, seguindo o modelo  $G(\eta_i, (\nu_i/2)^{-1})$  com função de ligação identidade.

Suponha, agora, que a variável aleatória  $Y$  tem distribuição gama  $G(\mu, \phi)$  com coeficiente de variação  $\sqrt{\phi}$  bastante pequeno. Obtêm-se as aproximações

$$E[\log(Y)] \approx \log(\mu) - \frac{\phi}{2} \text{ e } \text{Var}[\log(Y)] \approx \phi.$$

Assim, ao invés de analisar os dados  $y$ , usando-se o modelo gama  $G(\mu, \phi)$  com função de ligação  $g(\cdot)$ , pode-se ajustar aos logaritmos dos dados um modelo normal alternativo de variância constante  $\phi$  e função de ligação  $g(\exp(\cdot))$ . Além disso, a variância

da variável transformada, isto é,  $\phi = \text{Var}[\log(Y)]$ , pode ser estimada, após o ajuste do modelo normal, por exemplo, pelo quadrado médio dos resíduos.

Finalmente, pode-se demonstrar que o logaritmo da função de verossimilhança do modelo gama  $G(\mu, \phi)$  é, aproximadamente, quadrático, na escala  $\mu^{-1/3}$ , e que a diferença entre o seu máximo e o valor num ponto arbitrário  $\mu$ , é igual a  $9y^{2/3}(y^{-1/3} - \mu^{-1/3})^2/2$  (McCullagh e Nelder, 1989, Seção 7.2). Ainda, tem-se que a variável transformada  $3[(Y/\mu)^{1/3} - 1]$  é, aproximadamente, normal.

### 2.4.5 Modelo normal inverso

A distribuição normal inversa (ou Gaussiana inversa) foi deduzida por Wald e Tweedie em dois artigos publicados, independentemente, em 1947. A f.d.p. da distribuição normal inversa  $IG(\mu, \phi)$  com média  $\mu > 0$  e parâmetro  $\phi > 0$ , representando uma medida de dispersão, é expressa por

$$f(y; \mu, \phi) = (2\pi\phi y^3)^{-1/2} \exp \left[ \frac{-(y - \mu)^2}{2\mu^2\phi y} \right], \quad y > 0.$$

O parâmetro  $\mu$  é, portanto, uma medida de localização e o parâmetro  $\phi$ , uma medida de dispersão igual à razão entre a variância e o cubo da média.

As características da distribuição  $IG(\mu, \phi)$  são: f.g.m.  $M(t; \mu, \phi) = \exp \{(\phi\mu)^{-1}[1 - (1 + 2\mu^2\phi t)^{1/2}]\}$ , cumulantes para  $r \geq 2$  obtidos de  $\kappa_r = 1.3.5 \dots (2r - 1)\mu^{2r-1}\phi^{r-1}$ , coeficientes de assimetria e curtose iguais a  $3\sqrt{\mu\phi}$  e  $3 + 15\mu\phi$ , respectivamente, e moda  $\mu[(1 + 9\mu^2\phi^2/4)^{1/2} - 3\mu\phi/2]$ . A distribuição é unimodal e sua forma depende apenas do valor do produto  $\phi\mu$ . Uma relação importante entre os momentos positivos e negativos é  $E(Y^{-r}) = E(Y^{r+1})/\mu^{2r+1}$ .

A f.d.a. da distribuição normal inversa  $IG(\mu, \phi)$  pode ser determinada a partir da distribuição acumulada da normal  $N(0, 1)$  por  $P(Y \leq y) = \Phi(y_1) + \exp[2/(\phi\mu)]\Phi(y_2)$ , em que  $y_1 = (\phi y)^{-1/2}(-1 + y/\mu)$  e  $y_2 = -(\phi y)^{-1/2}(1 + y/\mu)$ .

A distribuição normal inversa tem distribuição assintótica normal, da mesma forma que a gama, a log-normal e outras distribuições assimétricas. Quando  $\phi \rightarrow 0$ , a distribuição normal inversa  $IG(\mu, \phi)$  é, assintoticamente, normal  $N(\mu, \mu^3\phi)$ .

As aplicações do modelo normal inverso  $IG(\mu, \phi)$  concentram-se no estudo do movimento Browniano de partículas, análise de regressão com dados consideravelmente assimétricos, testes de confiabilidade, análise sequencial e análogo de análise de variância para classificações encaixadas. Outras aplicações incluem análise de tempos, como: duração de greves, tempo de primeira passagem nos passeios aleatórios, tempos de sobrevivência, tempo gasto para injetar uma substância no sistema biológico, etc.

Existem muitas analogias entre os modelos normal e normal inverso. Por exemplo, o dobro do termo do expoente com sinal negativo nas funções densidades normal e normal inversa tem distribuição  $\chi_1^2$ . Um estudo completo do modelo normal inverso  $IG(\mu, \phi)$  é apresentado por Folks e Chhikara (1978).

#### 2.4.6 Modelo binomial negativo

A distribuição binomial negativa com parâmetros  $k > 0$  e  $0 < p < 1$  é definida por

$$P(Y = y) = \binom{k + y - 1}{k - 1} \left( \frac{p}{p + 1} \right)^y \frac{1}{(p + 1)^k},$$

para  $y = 0, 1, 2, \dots$ . O parâmetro  $\mu = kp$  é igual à média e pode ser usado no lugar de  $p$  (Tabela 1.1 e Exercício 1b do Capítulo 1). Quando  $k$  é inteiro, essa distribuição é, também, denominada de distribuição de Pascal. Um caso especial importante é a distribuição geométrica quando  $k = 1$ . Formas especiais da distribuição binomial negativa surgiram, em 1679, com Pascal e Fermat. Gosset (“Student”), em 1907, usou a distribuição binomial negativa para analisar dados na forma de contagens no lugar da distribuição de Poisson.

A f.g.m. é  $M(t) = [1 + p(1 - e^t)]^{-k}$ . A variância é igual a  $\text{Var}(Y) = kp(1 + p)$  e os coeficientes de assimetria e curtose são  $(2p + 1)/\sqrt{kp(p + 1)}$  e  $3 + [1 + 6p(1 + p)]/[kp(1 + p)]$ , respectivamente. Observe-se que a variância pode ser especificada em termos da média como  $\text{Var}(Y) = \mu(1 + \mu/k)$ , o que caracteriza o modelo binomial negativo como um dos modelos adequados para estudar **superdispersão**, isto é,

quando  $\text{Var}(Y) > E(Y)$  (Hinde e Demétrio, 1998a,b).

Pode-se verificar que  $P(Y = y + 1) > P(Y = y)$  quando  $y < \mu(1 - k^{-1}) - 1$  e  $P(Y = y + 1) < P(Y = y)$  quando  $y > \mu(1 - k^{-1}) - 1$ .

A f.d.a. da distribuição binomial negativa  $P(Y \leq y)$  para  $y$  inteiro pode ser determinada a partir da distribuição acumulada da variável aleatória  $X$  com distribuição binomial com parâmetros  $k + y$  e  $(1 + p)^{-1}$  por  $P(Y \leq y) = P(X \geq k)$ . Alternativamente, a distribuição acumulada da binomial negativa pode ser calculada de forma aproximada por

$$P(Y \leq y) \approx e^{-\mu} \sum_{i=0}^y \frac{\mu^i}{i!} - \frac{(y - \mu) k e^{-\mu} \mu^y}{2 (\mu + k) y!}.$$

#### 2.4.7 Modelo secante hiperbólico generalizado

A distribuição secante hiperbólica generalizada (SHG) foi estudada por Morris (1982) no contexto da função de variância da família exponencial, sendo uma função quadrática da média. A f.d.p. é expressa por ( $y \in \mathbb{R}$ )

$$f(y; \mu, \phi) = \exp \left\{ \frac{1}{\phi} [y \arctan \mu - \frac{1}{2} \log(1 + \mu^2)] + c(y, \phi) \right\},$$

sendo

$$c(y, \phi) = \log \left\{ \frac{2^{(1-2\phi)/\phi}}{\pi \phi \Gamma(\phi^{-1})} \right\} - \sum_{j=0}^{\infty} \log \left\{ 1 + \frac{y^2}{(1 + 2j\phi)^2} \right\}.$$

Em relação a outras distribuições na família exponencial, a forma de sua função  $c(y, \phi)$  é bastante complicada. Entretanto, a distribuição SHG pode ser adequada para análise de dados contínuos reais como distribuição alternativa à distribuição normal. A sua função de variância é obtida de  $\theta = \arctan(\mu)$  como  $V = d\mu/d\theta = 1 + \mu^2$ . Morris (1982) demonstrou que na família exponencial (2.4) existem, exatamente, seis distribuições com função de variância quadrática  $V(\mu) = c_0 + c_1\mu + c_2\mu^2$ , a saber: binomial ( $c_0 = 0, c_1 = 1, c_2 = -1$ ), Poisson ( $c_0 = c_2 = 0, c_1 = 1$ ), normal ( $c_0 = 1, c_1 = c_2 = 0$ ), gama ( $c_0 = c_1 = 0, c_2 = 1$ ), binomial negativa ( $c_0 = 0, c_1 = 1, c_2 > 0$ ) e SHG ( $c_0 = c_2 = 1, c_1 = 0$ ).

### 2.4.8 Modelos definidos por transformações

Sejam  $Y_1, \dots, Y_n$  variáveis aleatórias tais que, após alguma transformação  $h(\cdot)$ , as variáveis resultantes  $Z_1, \dots, Z_n$ , em que  $Z_i = h(Y_i)$ , têm distribuições normais de médias  $\mu_1, \dots, \mu_n$  e variância constante  $\sigma^2$ , e que existe uma outra transformação  $g(\cdot)$  produzindo linearidade dos efeitos sistemáticos, isto é,  $g[E(\mathbf{Y})] = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ . Usando-se expansão de Taylor ao redor de  $\mu$  até primeira ordem, tem-se  $g(h^{-1}(\mu)) = \eta$  e, portanto, esses modelos pertencem, de forma aproximada, à classe dos MLGs, com distribuição normal e função de ligação  $g(h^{-1}(\cdot))$ . A variância da variável resposta original pode ser obtida, aproximadamente, de  $\text{Var}(Y) = \sigma^2 / \{h'[g^{-1}(\eta)]^2\}$ .

Usando-se a transformação potência de Box e Cox (1964), pode-se definir uma subclasse de modelos por

$$Z = h(Y) = \phi^{-1}(Y^\phi - 1) \sim N(\mu, \sigma^2)$$

e

$$g[E(Y)] = \phi^{-1}\{[E(Y)]^\theta - 1\},$$

em que  $g[E(\mathbf{Y})] = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ . Aqui,  $\phi = 0$  e  $\theta = 0$  correspondem à transformação logarítmica. Logo, essa subclasse é representada, no contexto dos MLGs, por uma distribuição normal com função de ligação  $\phi^{-1}[(1 + \phi\mu)^{\theta/\phi} - 1] = \eta$ , supondo  $\theta \neq 0$  e  $\phi \neq 0$ . A demonstração segue por expansão em série de Taylor até primeira ordem.

Quando  $\theta = \phi = 1$ , tem-se o modelo clássico de regressão. Um caso importante, denominado *polinômios inversos* (Nelder, 1966), é definido por  $\phi = 0$  e  $\theta = -1$  e, portanto, considera erros normais na escala logarítmica e linearidade na escala inversa, sendo equivalente ao modelo normal  $N(\mu, \sigma^2)$  com função de ligação  $\eta = 1 - \exp(-\mu)$ .

## 2.5 Ajuste dos MLGs

O processo de análise de dados, usando MLGs, pode ser dividido em três etapas: (i) formulação dos modelos; (ii) ajuste dos modelos e (iii) inferência.



Os MLGs formam um ferramental de grande utilidade prática, pois apresentam grande flexibilidade na etapa (i), computação simples em (ii) e critérios razoáveis em (iii). Essas etapas são realizadas sequencialmente. Na análise de dados complexos, após a conclusão da etapa de inferência, pode-se voltar à etapa (i) e escolher outros modelos, a partir de informações mais detalhadas oriundas do estudo feito em (iii).

Uma característica importante dos MLGs é que se supõe independência das variáveis respostas (ou, pelo menos, não-correlação) e, portanto, dados exibindo auto-regressões como as séries temporais, em princípio, podem ser excluídos. Uma segunda característica é que a estrutura da variável resposta é suposta única embora, usualmente, existam várias variáveis explanatórias na estrutura linear desses modelos. Assim, outras técnicas estatísticas devem ser consideradas para analisar dados que ocorrem em planejamentos de experimentos com mais de uma fonte de erro. Ainda, variáveis respostas com distribuições que não pertencem à família (2.4), como a distribuição de Cauchy, e estruturas não-lineares do tipo  $\eta = \sum \beta_j \exp(\alpha_j x_j)$ , a menos que os  $\alpha_j$  sejam conhecidos, devem, também, ser excluídos.

Apresentam-se, agora, as características principais das etapas que formam a metodologia de trabalho com os MLGs.

### 2.5.1 Formulação de modelos

A etapa de formulação dos modelos compreende a escolha de opções para a distribuição de probabilidade da variável resposta, variáveis explanatórias (matriz modelo) e função de ligação para descrever as características principais da variável resposta. Para se escolher razoavelmente a distribuição em (2.4), devem-se examinar cuidadosamente os dados, principalmente quanto aos seguintes pontos básicos: assimetria, natureza contínua ou discreta (por exemplo, contagens) e intervalo de variação.

As distribuições gama e normal inversa são associadas a dados contínuos assimétricos. Se os dados exibem simetria e o intervalo de variação é o conjunto dos

reais, a distribuição normal deve ser escolhida. Entretanto, se os dados têm intervalo de variação em  $(0, \infty)$ , a suposição de normalidade pode ser mais apropriada para alguma transformação dos dados, por exemplo, a logarítmica. Alternativamente, podem-se supor as distribuições normal inversa e gama, cujos intervalos de variação são positivos. Quando os dados apresentam coeficientes de variação constante, o modelo gama deve ser o preferido.

A distribuição de Poisson usada para análise de observações na forma de contagens, também pode ser usada para análise de dados contínuos que apresentam variância, aproximadamente, igual à média. Quando a variância dos dados é maior do que a média (ao invés de igual), pode-se trabalhar com as distribuições gama, normal inversa e binomial negativa. Esse fenômeno é denominado **superdispersão** para distribuições discretas (Hinde e Demétrio, 1998a,b). A escolha entre essas três distribuições pode depender, exclusivamente, da dispersão dos dados. A variância da distribuição binomial negativa ( $V(\mu) = \mu + \mu^2/r$ ) pode ser aproximada, para um intervalo razoável de variação de  $\mu$ , por  $V(\mu) = \lambda\mu$ , em que a função de variância contém um parâmetro multiplicador  $\lambda > 1$ , desconhecido. Portanto, a distribuição de Poisson pode ser empregada para análise de dados que apresentam superdispersão, desde que seja obtida uma estimativa para  $\lambda$ . O fenômeno de subdispersão, em que a variância dos dados é menor do que a média, pode ser tratado usando-se o modelo de Poisson com  $\lambda < 1$ , mas é muito incomum na prática. Nesse caso, o modelo binomial pode ser mais adequado.

A distribuição binomial serve para análise de dados na forma de proporções, podendo ainda ser útil na análise de dados contínuos ou discretos apresentando subdispersão. A superdispersão pode ser modelada usando-se a distribuição binomial com um parâmetro multiplicador na função de variância.

A escolha de uma função de ligação compatível com a distribuição proposta para os dados deve resultar de considerações a priori, exame intensivo dos dados, facilidade de interpretação do modelo e, mais usualmente, uma mistura de tudo isso.

No modelo clássico de regressão, a função de ligação é a identidade no sentido

de que valores esperados e preditores lineares podem ter qualquer valor real. Entretanto, quando os dados representam contagens e a distribuição é de Poisson, a função de ligação identidade, como observado anteriormente, é menos atrativa, pois não restringe os valores esperados ao intervalo  $(0, \infty)$ . Quando efeitos sistemáticos multiplicativos contribuem para as médias dos dados, uma função de ligação logarítmica torna os efeitos aditivos contribuindo para os preditores lineares e, portanto, pode ser a mais apropriada. Analogamente, as funções de ligação adequadas para dados na forma de proporções, devem ser funções de  $(0, 1)$  no conjunto dos reais, como probito, logística, complemento log-log e arco seno. As funções de ligação compatíveis com os modelos gama, normal inverso e binomial negativo devem restringir as médias dos dados ao intervalo  $(0, \infty)$ .

Na Tabela 2.8, apresenta-se a combinação distribuição da variável resposta/função de ligação para os casos especiais dos MLGs  $(a), (b), \dots, (l)$ , descritos na Seção 2.1.

Tabela 2.8: Combinação da distribuição da variável resposta e da função de ligação para os casos especiais de MLGs descritos na Seção 2.1.

Função de ligação	Distribuição				
	Normal	Poisson	Binomial	Gama	Normal Inversa
Identidade	(a)	—	—	(i)	—
Logarítmica	—	(e)	—	—	—
Inversa	(h)	—	—	(g)(j)	—
Inversa do quadrado	—	—	—	—	(l)
Logística	—	—	(d)(f)	—	—
Probit	—	—	(c)	—	—
Complemento log-log	—	—	(b)	—	—
Observação: Para os casos (g), (j) e (l) são escolhidas as funções de ligação mais usuais (canônicas) que correspondem a $\theta = \eta$ .					

Existem funções de ligação que produzem propriedades estatísticas de-

sejáveis para o modelo, particularmente, em pequenas amostras. Essas funções são definidas visando aos seguintes efeitos de forma separada: constância da informação de Fisher e da curvatura do logaritmo da função de verossimilhança, estatísticas suficientes de dimensão mínima, normalização aproximada dos EMVs dos parâmetros lineares e simetria do logaritmo da função de verossimilhança. Nenhuma função de ligação pode produzir todos estes efeitos desejados e, muitas vezes, se existe uma função de ligação superior às demais, ela pode conduzir a dificuldades de interpretação.

A terceira escolha na formulação do modelo é a do conjunto de variáveis explanatórias para representar a estrutura linear do MLG, ou seja, a formação da matriz modelo. Em geral, as variáveis explanatórias escolhidas devem ser não-correlacionadas. Os termos da estrutura linear podem ser *contínuos*, *qualitativos* e *mistos*.

Uma variável explanatória quantitativa (covariável)  $x$ , geralmente, corresponde a um único parâmetro  $\beta$ , contribuindo com o termo  $\beta x$  para o modelo, enquanto uma variável explanatória qualitativa  $A$ , denominada frequentemente de fator, inclui na estrutura linear um conjunto de parâmetros  $\alpha_i$ , em que  $i$  é o índice que representa os níveis do fator. Assim, na estrutura linear  $\eta_i = \alpha_i + \beta x$ , representando grupos distintos de um fator  $A$  mais uma covariável  $x$ , a ordenada varia com o nível do fator, mas a declividade é a mesma. Entretanto, em alguns casos, a declividade deve variar com o nível do fator e, portanto, o termo  $\beta x$  deve ser substituído pelo mais geral  $\beta_i x$ , produzindo  $\eta = \alpha_i + \beta_i x$ . O termo  $\beta_i x$  é denominado misto, pois a declividade associada à variável explanatória é suposta diferente para cada nível do fator.

Frequentemente, as observações são classificadas por dois ou mais fatores simultaneamente e, então, termos representando interações entre os fatores devem ser incluídos no modelo. Uma covariável  $x$  pode ser transformada por uma função não-linear  $h(x)$ , sem prejudicar a linearidade do modelo, desde que  $h(\cdot)$  não contenha parâmetros desconhecidos. Assim, a estrutura linear do modelo pode conter

polinômios em  $x$ . Transformações simples nas variáveis explanatórias podem implicar num grande aperfeiçoamento do componente sistemático do modelo. O caso de funções não-lineares das variáveis explanatórias com parâmetros desconhecidos será descrito na Seção 5.8. Em muitas aplicações, a combinação linear das variáveis explanatórias  $x_1, \dots, x_p$  depende, fortemente, das características do experimento e deve propiciar uma contribuição útil na explicação do comportamento da variável resposta associada às observações  $y$ .

Um MLG é considerado como uma boa representação dos dados se for possível explicar a relação variância/média, satisfatoriamente, e se produzir efeitos aditivos na escala definida pela função de ligação. Um modelo parcimonioso é, também, uma exigência, no sentido de que o número de parâmetros seja tão pequeno quanto possível. Por exemplo, se os dados são classificados por dois ou mais fatores, um modelo parcimonioso deve minimizar o número de interações entre os fatores.

Um ponto fundamental no processo de escolha de um MLG é que não se deve ficar restrito a um único modelo, achando-o mais importante e excluir outros modelos alternativos. É prudente considerar a escolha restrita a um conjunto amplo de modelos estabelecidos por princípios como: facilidade de interpretação, boas previsões anteriores e conhecimento prévio da estrutura dos dados. Algumas características nos dados podem não ser descobertas, mesmo por um modelo muito bom e, portanto, um conjunto razoável de modelos adequados aumenta a possibilidade de se detectarem essas características.

### 2.5.2 Ajuste dos modelos

A etapa de ajuste representa o processo de estimação dos parâmetros lineares dos modelos e de determinadas funções das estimativas desses parâmetros, que representam medidas de adequação dos valores estimados. Vários métodos podem ser usados para estimar os parâmetros dos MLGs. Nesse ponto, convém recordar a citação “nada é tão fácil quanto inventar métodos de estimação” de Sir Ronald Fisher (1925). Como o método de máxima verossimilhança nos MLGs conduz a um

procedimento de estimação bastante simples, esse método é o mais usado.

O algoritmo para a solução das equações de máxima verossimilhança nos MLGs foi desenvolvido por Nelder e Wedderburn (1972) e equivale ao cálculo repetido de uma regressão linear ponderada, como será descrito na Seção 3.2. O algoritmo é similar a um processo iterativo de Newton-Raphson, mas a característica principal é o uso da matriz de valores esperados das derivadas parciais de segunda ordem do logaritmo da função de verossimilhança (informação), em relação aos  $\beta$ , no lugar da matriz correspondente de valores observados. Essa característica foi, primeiramente, desenvolvida por Fisher (1935), para o caso da distribuição binomial com função de ligação probito e o processo é denominado “método score para estimação de parâmetros”.

O algoritmo de Nelder e Wedderburn (1972) tem como casos especiais os algoritmos de Finney (1971), para o cálculo de curvas ajustadas de resposta a um conjunto de doses de um medicamento, e de Haberman (1970), para o cálculo das estimativas nos modelos log-lineares.

Vários programas estatísticos como R, SAS, S-PLUS, STATA e MATLAB apresentam, para cada ajuste, as estimativas dos parâmetros  $\beta$ ,  $\eta$  e  $\mu$  do modelo, resíduos, estruturas de covariância e correlação entre as estimativas e outras funções de interesse.

O algoritmo de estimação nos MLGs é bastante robusto, convergindo rapidamente. Entretanto, pode falhar em convergir de duas maneiras distintas:

- (a) as estimativas dos parâmetros tendem para valores infinitos, embora o máximo do logaritmo da função de verossimilhança esteja convergindo para o valor correto;
- (b) o logaritmo da função de verossimilhança, ao invés de sempre crescer no processo iterativo, começa a decrescer ou oscilar, constituindo uma divergência real.

Quando  $\theta = \eta$ , implicando um modelo com  $p$  estatísticas suficientes mini-

mais, tem-se constatado que exemplos de divergência são muito raros. Ao ocorrer falha do algoritmo, torna-se necessário repetir o procedimento de estimação, a partir dos valores ajustados correntes, usando um modelo diferente, pois a convergência pode ser alcançada (Cordeiro, 1986).

### 2.5.3 Inferência

A etapa de inferência tem como objetivo principal verificar a adequação do modelo como um todo e realizar um estudo detalhado quanto a discrepâncias locais. Essas discrepâncias, quando significativas, podem implicar na escolha de outro modelo, ou em aceitar a existência de observações aberrantes. Em qualquer caso, toda a metodologia de trabalho deverá ser repetida.

Deve-se, nessa etapa, verificar a precisão e a interdependência das estimativas, construir regiões de confiança e testes sobre os parâmetros de interesse, analisar estatisticamente os resíduos e realizar previsões.

A precisão das previsões depende basicamente do modelo selecionado e, portanto, um critério de adequação do ajuste é verificar se a precisão de uma previsão em particular é maximizada. Muitas vezes, é possível otimizar a precisão por simples alteração do componente sistemático do modelo.

Um gráfico dos resíduos padronizados *versus* valores ajustados, sem nenhuma tendência, é um indicativo de que a relação funcional variância/média proposta para os dados é satisfatória. Gráficos dos resíduos *versus* variáveis explanatórias que não estão no modelo são bastante úteis. Se nenhuma variável explanatória adicional for necessária, então não se deverá encontrar qualquer tendência nesses gráficos. Observações com erros grosseiros podem ser detectadas como tendo resíduos grandes e *leverages* pequenos ou resíduos pequenos e *leverages* ( $h$ ) grandes, ou o modelo ajustado deve requerer mais variáveis explanatórias, por exemplo, interações de ordem superior. A inspeção gráfica é um meio poderoso de inferência nos MLGs.

Para verificar o ajuste do MLG, pode-se adotar o critério da razão da verossimilhanças em relação ao modelo saturado e a estatística de Pearson generalizada

(Seção 4.2). Quase toda a parte de inferência nos MLGs é baseada em resultados assintóticos e pouco tem sido estudado sobre a validade desses resultados em amostras muito pequenas.

Um modelo mal ajustado aos dados pode apresentar uma ou mais das seguintes condições: (a) inclusão de um grande número de variáveis explanatórias no modelo, muitas das quais são redundantes e algumas explicando somente um pequeno percentual das observações; (b) formulação de um modelo bastante pobre em variáveis explanatórias, que não revela e nem reflete as características do mecanismo gerador dos dados; (c) as observações mostram-se insuficientes para que falhas do modelo sejam detectadas.

A condição (a) representa uma superparametrização do modelo implicando numa imprecisão das estimativas e (b) é a situação oposta de (a): uma subparametrização que implica em previsões ruins. A terceira condição é um tipo de falha difícil de se detectar, e é devida à combinação inadequada distribuição/função de ligação, que não depende das observações em questão.

## 2.6 Exercícios

1. Para o modelo binomial as funções de ligação mais comuns são: logística, probito e complemento log-log. Compare os valores do preditor linear para essas funções de ligação no intervalo  $(0, 1)$ .

2. Mostre que

$$\lim_{\lambda \rightarrow 0} \frac{\mu^\lambda - 1}{\lambda} = \log(\mu).$$

3. Considere a família de funções de ligação definida por Aranda-Ordaz (1981)

$$\eta = \log \left[ \frac{(1 - \pi)^{-\lambda} - 1}{\lambda} \right], \quad 0 < \pi < 1 \text{ e } \lambda \text{ uma constante.}$$



Mostre que a função de ligação logística é obtida para  $\lambda = 1$  e que quando  $\lambda \rightarrow 0$ , tem-se a função de ligação complemento log-log.

4. Compare os gráficos de  $\eta = \log \left[ \frac{(1 - \mu)^{-\lambda} - 1}{\lambda} \right]$  versus  $\mu$  para  $\lambda = -1, -0.5, 0, 0.5, 1$  e  $2$ .

5. Explique como um modelo de Box-Cox poderia ser formulado no contexto dos MLGs.

6. Demonstre que se  $Y$  tem uma distribuição binomial  $B(m, \pi)$ , então para  $m$  grande  $\text{Var}(\arcsen \sqrt{Y/m})$  é, aproximadamente,  $1/(4m)$ , com o ângulo expresso em radianos. Em que situações uma estrutura linear associada a essa transformação poderá ser adequada?

7. Suponha que  $Y$  tem distribuição binomial  $B(m, \pi)$  e que  $g(Y/m)$  é uma função arbitrária. Calcule o coeficiente de assimetria assintótico de  $g(Y/m)$ . Demonstre que se anula quando  $g(\pi) = \int_0^\pi t^{-1/3}(1-t)^{-1/3}dt$  e, portanto, a variável aleatória definida por  $[g(Y/m) - g(\alpha)]/[\pi^{1/6}(1-\pi)^{1/6}m^{-1/2}]$ , em que  $\alpha = \pi - (1-2\pi)/(6m)$ , tem distribuição próxima da normal reduzida (Cox e Snell, 1968).

8. Sejam  $Y_1$  e  $Y_2$  variáveis aleatórias binomiais de parâmetros  $\pi_1$  e  $\pi_2$  em dois grupos de tamanhos  $m_1$  e  $m_2$ , respectivamente. O número de sucessos  $Y_1$  no primeiro grupo, dado que o número total de sucessos nos dois grupos é  $r$ , tem distribuição hipergeométrica generalizada de parâmetros  $\pi_1, \pi_2, m_1, m_2$  e  $r$ . Demonstre que essa distribuição é um membro da família (2.4) com parâmetro  $\theta = \log\{\pi_1(1-\pi_2)/[\pi_2(1-\pi_1)]\}$ ,  $\phi = 1$  e  $\pi = D_1(\theta)/D_0(\theta)$ , em que  $D_i(\theta) = \sum_x x^i \binom{m_1}{x} \binom{m_2}{r-x} \exp(\theta x)$  para  $i = 0, 1$ . Calcule a expressão do  $r$ -ésimo cumulante dessa distribuição.

9. Se  $Y$  tem distribuição de Poisson  $P(\mu)$ , demonstre:

- (a) que o coeficiente de assimetria  $Y^{2/3}$  é de ordem  $\mu^{-1}$  enquanto que aqueles de  $Y$  e  $Y^{1/2}$  são de ordem  $\mu^{-1/2}$ ;

- (b) que o logaritmo da função de verossimilhança para uma única observação é, aproximadamente, quadrático na escala  $\mu^{1/3}$ ;
  - (c) a fórmula do  $r$ -ésimo momento fatorial  $E[Y(Y-1)\dots(Y-r+1)] = \mu^r$ ;
  - (d) a fórmula de recorrência entre os momentos centrais  $\mu_{r+1} = r\mu\mu_{r-1} + \mu d\mu_r/d\mu$ ;
  - (e) que  $2\sqrt{Y}$  tem, aproximadamente, distribuição normal  $N(0, 1)$ .
- 10.** Se  $Y$  tem distribuição gama  $G(\mu, \phi)$ , demonstre que:
- (a) quando  $\phi < 1$ , a função densidade é zero na origem e tem uma única moda no ponto  $\mu(1 - \phi)$ ;
  - (b) o logaritmo da função de verossimilhança para uma única observação é, aproximadamente, quadrático na escala  $\mu^{-1/3}$ ;
  - (c) a variável transformada  $3[(Y/\mu)^{1/3} - 1]$  é, aproximadamente, normal.
- 11.** Se  $Y$  tem distribuição binomial  $B(m, \pi)$ , demonstre que a média e a variância de  $\log[(Y + 0, 5)/(m - Y + 0, 5)]$  são iguais a  $\log[\pi/(1 - \pi)] + O(m^{-2})$  e  $E[(Y + 0, 5)^{-1} + (m - Y + 0, 5)^{-1}] + O(m^{-3})$ , respectivamente.
- 12.** Se  $Y$  tem distribuição de Poisson  $P(\mu)$ , obtenha uma expansão para  $\text{Var}[(Y + c)^{1/2}]$  em potências de  $\mu^{-1}$ , e mostre que o coeficiente de  $\mu^{-1}$  é zero quando  $c = 3/8$ . Ache uma expansão similar para  $\text{Var}[Y^{1/2} + (Y + 1)^{1/2}]$ .
- 13.** Qual é a distribuição da tolerância correspondente à função de ligação  $\arcsen\sqrt{\cdot}$ ?
- 14.** Se  $Y$  tem distribuição binomial  $B(m, \pi)$ , demonstre que os momentos da estatística  $Z = \pm\{2Y \log(Y/\mu) + 2(m - Y) \log[(m - Y)/(m - \mu)]\}^{1/2} + \{(1 - 2\pi)/[m\pi(1 - \pi)]\}^{1/2}/6$  diferem dos correspondentes da distribuição normal reduzida  $N(0, 1)$  com

erro  $O(m^{-1})$ . Essa transformação induz simetria e estabiliza a variância simultaneamente (McCullagh e Nelder, 1989).

**15.** Se  $Y$  tem distribuição binomial  $B(m, \pi)$ , demonstre a expressão aproximada  $P(Y \leq y) = \Phi(y_1)$ , em que  $y_1 = 2m^{1/2}\{\arcsen[(y+3/8)/(m+3/4)]^{1/2} - \arcsen(\pi^{1/2})\}$ .

**16.** Suponha que  $Y \sim B(m, \pi)$ , sendo  $\pi = e^\lambda(1 + e^\lambda)^{-1}$ . Mostre que  $m - Y$  tem distribuição binomial com parâmetro induzido correspondente  $\lambda' = -\lambda$ .

**17.** Demonstre que para a variável aleatória  $Y$  com distribuição de Poisson, tem-se:

$$(a) \ E(Y^{1/2}) \approx \mu^{1/2} \text{ e } \text{Var}(Y^{1/2}) \approx \frac{1}{4};$$

$$(b) \ E(Y^{1/2}) = \mu^{1/2} \left(1 - \frac{1}{8\mu}\right) + O(\mu^{-3/2}) \text{ e } \text{Var}(Y^{1/2}) = \frac{1}{4} \left(1 + \frac{3}{8\mu}\right) + O(\mu^{-3/2});$$

$$(c) \ E(Y^{2/3}) \approx \mu^{2/3} \left(1 - \frac{1}{9\mu}\right) \text{ e } \text{Var}(Y^{2/3}) \approx \frac{4\mu^{1/3}}{9} \left(1 + \frac{1}{6\mu}\right).$$

**18.** Se  $Y$  tem distribuição de Poisson com média  $\mu$ , mostre que:

$$(a) \ P(Y \leq y) = P(\chi_{2(y+1)}^2 > 2\mu);$$

$$(b) \ P(Y \leq y) = \Phi(z) - \phi(z) \left( \frac{z^2 - 1}{6\sqrt{\mu}} + \frac{z^5 - 7z^3 + 3z}{72\mu} \right) + O(\mu^{-3/2}), \text{ em que } z = (y + 0.5 - \mu)\mu^{-1/2} \text{ e } \Phi(\cdot) \text{ e } \phi(\cdot) \text{ são, respectivamente, a f.d.a. e a f.d.p. da distribuição normal reduzida.}$$



# Capítulo 3

## Estimação

### 3.1 Estatísticas suficientes

Seja um MLG definido pelas expressões (2.4), (2.6) e (2.7) e suponha que as observações a serem analisadas sejam representadas pelo vetor  $\mathbf{y} = (y_1, \dots, y_n)^T$ . O logaritmo da função de verossimilhança como função apenas de  $\boldsymbol{\beta}$  (considerando-se o parâmetro de dispersão  $\phi$  conhecido), especificado  $\mathbf{y}$ , é definido por  $\ell(\boldsymbol{\beta}) = \ell(\boldsymbol{\beta}; \mathbf{y})$  e, usando-se a expressão (2.4), tem-se

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^n \ell_i(\theta_i, \phi; y_i) = \phi^{-1} \sum_{i=1}^n [y_i \theta_i - b(\theta_i)] + \sum_{i=1}^n c(y_i, \phi), \quad (3.1)$$

em que  $\theta_i = q(\mu_i)$ ,  $\mu_i = g^{-1}(\eta_i)$  e  $\eta_i = \sum_{r=1}^p x_{ir} \beta_r$ .

A estimação do parâmetro de dispersão  $\phi$  será objeto de estudo na Seção 4.4. Existem  $n$  parâmetros canônicos  $\theta_1, \dots, \theta_n$  e  $n$  médias  $\mu_1, \dots, \mu_n$  que são desconhecidos, mas que são funções de  $p$  parâmetros lineares  $\beta_1, \dots, \beta_p$  do modelo. Deve-se, primeiramente, estimar o vetor  $\boldsymbol{\beta}$  de parâmetros para depois calcular as estimativas do vetor  $\boldsymbol{\mu}$  de médias e do vetor  $\boldsymbol{\theta}$  de parâmetros pelas relações funcionais  $\mu_i = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta})$  e  $\theta_i = q(\mu_i)$ .

Se o intervalo de variação dos dados não depende de parâmetros, pode-se demonstrar para os modelos contínuos (Cox e Hinkley, 1986, Capítulo 9), que todas as derivadas de  $\int \exp[\ell(\boldsymbol{\beta})] d\mathbf{y} = 1$  podem ser computadas dentro do sinal de integração e que o ponto  $\hat{\boldsymbol{\beta}}$  correspondente ao máximo do logaritmo da função

de verossimilhança (3.1) está próximo do vetor  $\boldsymbol{\beta}$  de parâmetros verdadeiros com probabilidade próxima de um. Para os modelos discretos, a integração é substituída pelo somatório. Esse fato ocorre em problemas denominados **regulares**.

Um caso importante dos MLGs surge quando o vetor  $\boldsymbol{\theta}$  de parâmetros canônicos da família (2.4) e o vetor  $\boldsymbol{\eta}$  de preditores lineares em (2.6) são iguais, conduzindo às funções de ligação canônicas. Tem-se  $\theta_i = \eta_i = \sum_{r=1}^p x_{ir}\beta_r$ , para  $i = 1, \dots, n$ . As estatísticas  $S_r = \sum_{i=1}^n x_{ir}Y_i$  para  $r = 1, \dots, p$  são suficientes para os parâmetros  $\beta_1, \dots, \beta_p$  e têm dimensão mínima  $p$ . Sejam  $s_r = \sum_{i=1}^n x_{ir}y_i$  as realizações de  $S_r$ ,  $r = 1, \dots, p$ . Então, a equação (3.1) pode ser escrita na forma

$$\ell(\boldsymbol{\beta}) = \phi^{-1} \left[ \sum_{r=1}^p s_r \beta_r - \sum_{i=1}^n b(\theta_i) \right] + \sum_{i=1}^n c(y_i, \phi)$$

e, portanto,  $\ell(\boldsymbol{\beta})$  tem a seguinte decomposição

$$\ell(\boldsymbol{\beta}) = \ell_1(\mathbf{s}, \boldsymbol{\beta}) + \ell_2(\mathbf{y}),$$

em que  $\ell_1(\mathbf{s}, \boldsymbol{\beta}) = \phi^{-1} \sum_{r=1}^p s_r \beta_r - \phi^{-1} \sum_{i=1}^n b(\sum_{r=1}^p x_{ir}\beta_r)$  e  $\ell_2(\mathbf{y}) = \sum_{i=1}^n c(y_i, \phi)$ .

Pelo teorema da fatoração,  $\mathbf{S} = (S_1, \dots, S_p)^T$  é suficiente de dimensão mínima  $p$  para  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$  e, portanto, ocorre uma redução na dimensão das estatísticas suficientes de  $n$  (o número de observações) para  $p$  (o número de parâmetros a serem estimados). As estatísticas  $S_1, \dots, S_p$  correspondem à maior redução que os dados podem ter, sem qualquer perda de informação relevante para se fazer inferência sobre o vetor  $\boldsymbol{\beta}$  de parâmetros desconhecidos.

Conforme descrito na Seção 2.3, as funções de ligação que produzem estatísticas suficientes de dimensão mínima  $p$  para as diversas distribuições são denominadas *canônicas*. A Tabela 2.7 mostra que essas funções de ligação para os modelos normal, Poisson, binomial, gama e normal inverso são  $\eta = \mu$ ,  $\eta = \log(\mu)$ ,  $\eta = \log[\mu/(m - \mu)]$ ,  $\eta = \mu^{-1}$  e  $\eta = \mu^{-2}$ , respectivamente.

As funções de ligação canônicas produzem propriedades estatísticas de interesse para o modelo, tais como, suficiência, facilidade de cálculo, unicidade dos estimadores de máxima verossimilhança (EMVs) e, em alguns casos, interpretação simples. Em princípio, pode-se trabalhar com as funções de ligação canônicas quando

não existirem indicativos de outra preferível. Entretanto, não existe razão para se considerarem sempre os efeitos sistemáticos como aditivos na escala especificada pela função de ligação canônica. A escolha da função de ligação será descrita, com mais detalhes, na Seção ??.

## 3.2 O algoritmo de estimação

A decisão importante na aplicação do MLG é a escolha do trinômio: distribuição da variável resposta  $\times$  matriz modelo  $\times$  função de ligação. A seleção pode resultar de simples exame dos dados ou de alguma experiência anterior. Inicialmente, considera-se esse trinômio fixo para se obter uma descrição adequada dos dados por meio das estimativas dos parâmetros do modelo. Vários métodos podem ser usados para estimar o vetor  $\boldsymbol{\beta}$  dos parâmetros, inclusive o qui-quadrado mínimo, o Bayesiano e a estimação-M. O último inclui o método de máxima verossimilhança (MV) que tem muitas propriedades ótimas, tais como, consistência e eficiência assintótica.

Neste livro, considera-se apenas o método de MV para estimar os parâmetros lineares  $\beta_1, \dots, \beta_p$  do modelo. O vetor escore é formado pelas derivadas parciais de primeira ordem do logaritmo da função de verossimilhança. Da expressão (3.1) pode-se calcular, pela regra da cadeia, o vetor escore  $\mathbf{U}(\boldsymbol{\beta}) = \partial \ell(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}$  de dimensão  $p$ , com elemento típico  $U_r = \frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_r} = \sum_{i=1}^n \frac{d\ell_i}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_r}$ , pois

$$\begin{aligned}
 \ell(\boldsymbol{\beta}) &= f(\theta_1, \dots, \theta_i, \dots, \theta_n) \\
 &\downarrow \\
 \theta_i &= \int V_i^{-1} d\mu_i = q(\mu_i) \\
 &\downarrow \\
 \mu_i &= g^{-1}(\eta_i) = h(\eta_i) \\
 &\downarrow \\
 \eta_i &= \sum_{r=1}^p x_{ir} \beta_r
 \end{aligned}$$

e, sabendo-se que  $\mu_i = b'(\theta_i)$  e  $d\mu_i/d\theta_i = V_i$ , tem-se

$$U_r = \phi^{-1} \sum_{i=1}^n (y_i - \mu_i) \frac{1}{V_i} \frac{d\mu_i}{d\eta_i} x_{ir} \quad (3.2)$$

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

A estimativa de máxima verossimilhança (EMV)  $\hat{\beta}$  do vetor de parâmetros  $\beta$  é calculada igualando-se  $U_r$  a zero para  $r = 1, \dots, p$ . Em geral, as equações  $U_r = 0$ ,  $r = 1, \dots, p$ , não são lineares e têm que ser resolvidas numericamente por processos iterativos do tipo Newton-Raphson.

O método iterativo de Newton-Raphson para a solução de uma equação  $f(x) = 0$  é baseado na aproximação de Taylor para a função  $f(x)$  na vizinhança do ponto  $x_0$ , ou seja,

$$f(x) = f(x_0) + (x - x_0)f'(x_0) = 0,$$

obtendo-se

$$x = x_0 - \frac{f(x_0)}{f'(x_0)}$$

ou, de uma forma mais geral,

$$x^{(m+1)} = x^{(m)} - \frac{f(x^{(m)})}{f'(x^{(m)})},$$

sendo  $x^{(m+1)}$  o valor de  $x$  no passo  $(m+1)$ ,  $x^{(m)}$  o valor de  $x$  no passo  $m$ ,  $f(x^{(m)})$  a função  $f(x)$  avaliada em  $x^{(m)}$  e  $f'(x^{(m)})$  a derivada da função  $f(x)$  avaliada em  $x^{(m)}$ .

Considerando-se que se deseja obter a solução do sistema de equações  $\mathbf{U} = \mathbf{U}(\beta) = \partial\ell(\beta)/\partial\beta = \mathbf{0}$  e, usando-se a versão multivariada do método de Newton-Raphson, tem-se

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

sendo  $\beta^{(m)}$  e  $\beta^{(m+1)}$  os vetores de parâmetros estimados nos passos  $m$  e  $(m+1)$ , respectivamente,  $\mathbf{U}^{(m)}$  o vetor escore avaliado no passo  $m$ , e  $(\mathbf{J}^{(m)})^{-1}$  a inversa da negativa da matriz de derivadas parciais de segunda ordem de  $\ell(\beta)$ , com elementos  $-\partial^2\ell(\beta)/\partial\beta_r\partial\beta_s$ , avaliada no passo  $m$ .



Quando as derivadas parciais de segunda ordem são avaliadas facilmente, o método de Newton-Raphson é bastante útil. Entretanto, isso nem sempre ocorre e no caso dos MLGs usa-se o método escore de Fisher que, em geral, é mais simples (coincidindo com o método de Newton-Raphson no caso das funções de ligação canônicas). Esse método envolve a substituição da matriz de derivadas parciais de segunda ordem pela matriz de valores esperados das derivadas parciais, isto é, a substituição da matriz de informação observada,  $\mathbf{J}$ , pela matriz de informação esperada de Fisher,  $\mathbf{K}$ . Logo,

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + (\mathbf{K}^{(m)})^{-1} \mathbf{U}^{(m)}, \quad (3.3)$$

sendo que  $\mathbf{K}$  tem elementos típicos expressos por

$$\kappa_{r,s} = -\mathbb{E} \left[ \frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_r \partial \beta_s} \right] = \mathbb{E} \left[ \frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_r} \frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_s} \right],$$

que é a matriz de covariâncias dos  $U_r$ .

Multiplicando-se ambos os membros de (3.3) por  $\mathbf{K}^{(m)}$ , tem-se

$$\mathbf{K}^{(m)} \boldsymbol{\beta}^{(m+1)} = \mathbf{K}^{(m)} \boldsymbol{\beta}^{(m)} + \mathbf{U}^{(m)}. \quad (3.4)$$

O elemento típico  $\kappa_{r,s}$  de  $\mathbf{K}$  é determinado de (3.2), sendo expresso por

$$\kappa_{r,s} = \mathbb{E}(U_r U_s) = \phi^{-2} \sum_{i=1}^n \mathbb{E}(Y_i - \mu_i)^2 \frac{1}{V_i^2} \left( \frac{d\mu_i}{d\eta_i} \right)^2 x_{ir} x_{is}$$

e como  $\text{Var}(Y_i) = \mathbb{E}(Y_i - \mu_i)^2 = \phi V_i$ , obtém-se

$$\kappa_{r,s} = \phi^{-1} \sum_{i=1}^n w_i x_{ir} x_{is},$$

sendo  $w_i = V_i^{-1} (d\mu_i/d\eta_i)^2$  denominada função peso. Logo, a matriz de informação de Fisher para  $\boldsymbol{\beta}$  tem a forma

$$\mathbf{K} = \phi^{-1} \mathbf{X}^T \mathbf{W} \mathbf{X},$$

sendo  $\mathbf{W} = \text{diag}\{w_1, \dots, w_n\}$  uma matriz diagonal de pesos que capta a informação sobre a distribuição e a função de ligação usadas e poderá incluir, também, uma

matriz de pesos a priori. No caso das funções de ligação canônicas tem-se  $w_i = V_i$ , pois  $V_i = V(\mu_i) = d\mu_i/d\eta_i$ . Nota-se que a informação é inversamente proporcional ao parâmetro de dispersão.

O vetor escore  $\mathbf{U} = \mathbf{U}(\boldsymbol{\beta})$  com componentes em (3.2) pode, então, ser expresso na forma

$$\mathbf{U} = \phi^{-1} \mathbf{X}^T \mathbf{W} \mathbf{G} (\mathbf{y} - \boldsymbol{\mu}),$$

em que  $\mathbf{G} = \text{diag}\{d\eta_1/d\mu_1, \dots, d\eta_n/d\mu_n\} = \text{diag}\{g'(\mu_1), \dots, g'(\mu_n)\}$ . Assim, a matriz diagonal  $\mathbf{G}$  é formada pelas derivadas de primeira ordem da função de ligação. Além disso, pode-se notar que  $\mathbf{W} = \mathbf{G}^{-2} \mathbf{V}^{-1}$ , em que  $\mathbf{V} = \text{diag}\{V(\mu_1), \dots, V(\mu_n)\}$ .

Substituindo  $\mathbf{K}$  e  $\mathbf{U}$  em (3.4) e eliminando  $\phi$ , tem-se

$$\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m+1)} = \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m)} + \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{G}^{(m)} (\mathbf{y} - \boldsymbol{\mu}^{(m)}),$$

ou, ainda,

$$\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m+1)} = \mathbf{X}^T \mathbf{W}^{(m)} [\boldsymbol{\eta}^{(m)} + \mathbf{G}^{(m)} (\mathbf{y} - \boldsymbol{\mu}^{(m)})].$$

Define-se a variável dependente ajustada  $\mathbf{z} = \boldsymbol{\eta} + \mathbf{G}(\mathbf{y} - \boldsymbol{\mu})$ . Logo,

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

ou

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{z}^{(m)}. \quad (3.5)$$

A equação matricial (3.5) é válida para qualquer MLG e mostra que a solução das equações de MV equivale a calcular repetidamente uma regressão linear ponderada de uma variável dependente ajustada  $\mathbf{z}$  sobre a matriz  $\mathbf{X}$  usando uma matriz de pesos  $\mathbf{W}$  que se modifica no processo iterativo. As funções de variância e de ligação entram no processo iterativo por meio de  $\mathbf{W}$  e  $\mathbf{z}$ . Tem-se, ainda, que

$$\text{Cov}(\mathbf{z}) = \mathbf{G} \text{Var}(\mathbf{Y}) \mathbf{G} = \phi \mathbf{W}^{-1}, \quad (3.6)$$

isto é, os  $z_i$  não são correlacionados. É importante enfatizar que a equação iterativa (3.5) não depende do parâmetro de dispersão  $\phi$ .

A demonstração da equação (3.5) foi desenvolvida por Nelder e Wedderburn (1972). Eles generalizaram procedimentos iterativos obtidos para casos especiais dos MLGs: probito (Fisher, 1935), log-lineares (Haberman, 1970) e logístico-lineares (Cox, 1972).

A variável dependente ajustada depende da derivada de primeira ordem da função de ligação. Quando a função de ligação é linear ( $\boldsymbol{\eta} = \boldsymbol{\mu}$ ), isto é, a identidade, tem-se  $\mathbf{W} = \mathbf{V}^{-1}$  sendo  $\mathbf{V} = \text{diag}\{V_1, \dots, V_n\}$ ,  $\mathbf{G} = \mathbf{I}$  e  $\mathbf{z} = \mathbf{y}$ , ou seja, a variável dependente ajustada reduz-se ao vetor de observações. Para o modelo normal linear ( $\mathbf{V} = \mathbf{I}, \boldsymbol{\mu} = \boldsymbol{\eta}$ ),  $\mathbf{W}$  é igual à matriz identidade de dimensão  $n$ ,  $\mathbf{z} = \mathbf{y}$  e verifica-se da equação (3.5) que a estimativa  $\hat{\boldsymbol{\beta}}$  reduz-se à fórmula esperada  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ . Esse é o único modelo em que  $\hat{\boldsymbol{\beta}}$  é obtido em forma fechada sem ser necessário um procedimento iterativo.

O método usual para iniciar o processo iterativo é especificar uma estimativa inicial e, sucessivamente, alterá-la até que a convergência seja alcançada e, portanto,  $\boldsymbol{\beta}^{(m+1)}$  aproxime-se de  $\hat{\boldsymbol{\beta}}$  quando  $m$  cresce. Note, contudo, que cada observação pode ser considerada como uma estimativa do seu valor médio, isto é,  $\mu_i^{(1)} = y_i$  e, assim, calcula-se

$$\eta_i^{(1)} = g(\mu_i^{(1)}) = g(y_i) \quad \text{e} \quad w_i^{(1)} = \frac{1}{V(y_i)[g'(y_i)]^2}.$$

Usando-se  $\boldsymbol{\eta}^{(1)}$  como variável resposta,  $\mathbf{X}$ , a matriz do modelo, e  $\mathbf{W}^{(1)}$ , a matriz diagonal de pesos com elementos  $w_i^{(1)}$ , obtém-se o vetor

$$\boldsymbol{\beta}^{(2)} = (\mathbf{X}^T \mathbf{W}^{(1)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(1)} \boldsymbol{\eta}^{(1)}.$$

O algoritmo de estimação, para  $m = 2, \dots, k$ , sendo  $k-1$  o número necessário de iterações para atingir a convergência, pode ser resumido nos seguintes passos:

(1) calcular as estimativas

$$\eta_i^{(m)} = \sum_{r=1}^p x_{ir} \beta_r^{(m)} \quad \text{e} \quad \mu_i^{(m)} = g^{-1}(\eta_i^{(m)});$$

(2) calcular a variável dependente ajustada

$$z_i^{(m)} = \eta_i^{(m)} + (y_i - \mu_i^{(m)})g'(\mu_i^{(m)})$$

e os pesos

$$w_i^{(m)} = \frac{1}{V(\mu_i^{(m)})[g'(\mu_i^{(m)})]^2};$$

(3) calcular

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

voltar ao passo (1) com  $\boldsymbol{\beta}^{(m)} = \boldsymbol{\beta}^{(m+1)}$  e repetir o processo até atingir a convergência, definindo-se, então,  $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(m+1)}$ .

Dentre os muitos existentes, um critério para verificar a convergência do algoritmo iterativo poderia ser

$$\sum_{r=1}^p \left( \frac{\beta_r^{(m+1)} - \beta_r^{(m)}}{\beta_r^{(m)}} \right)^2 < \xi,$$

considerando-se que  $\xi$  é um número positivo suficientemente pequeno. Em geral, esse algoritmo é robusto e converge rapidamente (menos de 10 iterações são suficientes). Entretanto, o critério do desvio é o mais usado e consiste em verificar se  $|\text{desvio}^{(m+1)} - \text{desvio}^{(m)}| < \xi$ , sendo desvio definido na Seção 4.2. Como um exemplo, desenvolve-se, no Apêndice A1, esse algoritmo para o exemplo 2.1, passo a passo, usando-se o R (Team, 2015).

Deve-se ser cauteloso se a função  $g(\cdot)$  não é definida para alguns valores  $y_i$ . Por exemplo, se a função de ligação for especificada por

$$\eta = g(\mu) = \log(\mu)$$

e forem observados valores  $y_i = 0$ , o processo não pode ser iniciado. Um método geral para contornar esse problema é substituir  $y$  por  $y + c$  tal que  $E[g(Y + c)]$  seja o mais próxima possível de  $g(\mu)$ . Para o modelo de Poisson com função de ligação logarítmica, adota-se  $c = 1/2$ . Para o modelo logístico, adota-se  $c = (1 - 2\pi)/2$

e  $\pi = \mu/m$ , sendo  $m$  o índice da distribuição binomial. De uma forma geral, da expansão de Taylor até segunda ordem para  $g(y+c)$  em relação a  $g(\mu)$ , tem-se

$$g(y+c) \approx g(\mu) + (y+c-\mu)g'(\mu) + (y+c-\mu)^2 \frac{g''(\mu)}{2},$$

cujo valor esperado é igual a

$$E[g(Y+c)] \approx g(\mu) + cg'(\mu) + \text{Var}(Y) \frac{g''(\mu)}{2}$$

que implica em

$$c \approx -\frac{1}{2} \text{Var}(Y) \frac{g''(\mu)}{g'(\mu)}.$$

Para pequenas amostras, a equação (3.5) pode divergir. O número de iterações até a convergência depende inteiramente do valor inicial arbitrado para  $\hat{\beta}$ , embora, geralmente, o algoritmo convirja rapidamente. A desvantagem do método tradicional de Newton-Raphson com o uso da matriz observada de derivadas de segunda ordem é que, normalmente, não converge para determinados valores iniciais.

Vários software estatísticos usam o algoritmo iterativo (3.5) para calcular as EMVs  $\hat{\beta}_1, \dots, \hat{\beta}_p$  dos parâmetros lineares do MLG, entre os quais, R, S-PLUS, SAS, GENSTAT e MATLAB.

### 3.3 Estimação em modelos especiais

Para as funções de ligação canônicas ( $w = V = d\mu/d\eta$ ) que produzem os modelos denominados canônicos, as equações de MV têm a seguinte forma, facilmente deduzidas de (3.2),

$$\sum_{i=1}^n x_{ir} y_i = \sum_{i=1}^n x_{ir} \hat{\mu}_i$$

para  $r = 1, \dots, p$ . Em notação matricial, tem-se

$$\mathbf{X}^T \mathbf{y} = \mathbf{X}^T \hat{\boldsymbol{\mu}}. \quad (3.7)$$

Nesse caso, as EMVs dos  $\beta's$  são únicas. Sendo  $\mathbf{S} = (S_1, \dots, S_p)^T$  o vetor de estatísticas suficientes definidas por  $S_r = \sum_{i=1}^n x_{ir} Y_i$ , conforme descrito na Seção 3.1, e  $\mathbf{s} = (s_1, \dots, s_p)^T$  os seus valores amostrais, as equações (3.7) podem ser expressas por

$$E(\mathbf{S}; \hat{\boldsymbol{\mu}}) = \mathbf{s},$$

mostrando que as EMVs das médias  $\mu_1, \dots, \mu_n$  nos modelos canônicos são calculadas igualando-se as estatísticas suficientes minimais aos seus valores esperados.

Se a matriz modelo corresponde a uma estrutura fatorial, consistindo somente de zeros e uns, o modelo pode ser especificado pelas margens que são as estatísticas minimais, cujos valores esperados devem igualar aos totais marginais.

As equações (3.7) são válidas para os seguintes modelos canônicos: modelo clássico de regressão, modelo log-linear, modelo logístico linear, modelo gama com função de ligação recíproca e modelo normal inverso com função de ligação recíproca ao quadrado. Para os modelos canônicos, o ajuste é realizado pelo algoritmo (3.5) com  $\mathbf{W} = \text{diag}\{V_i\}$ ,  $\mathbf{G} = \text{diag}\{V_i^{-1}\}$  e variável dependente ajustada com componente típica expressa por  $z_i = \eta_i + (y_i - \mu_i)/V_i$ .

Nos modelos com respostas binárias, a variável resposta tem distribuição binomial  $B(m_i, \pi_i)$  e o logaritmo da função de verossimilhança em (3.1) pode ser reescrito como

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^n \left[ y_i \log \left( \frac{\mu_i}{m_i - \mu_i} \right) + m_i \log \left( \frac{m_i - \mu_i}{m_i} \right) \right] + \sum_{i=1}^n \log \binom{m_i}{y_i},$$

em que  $\mu_i = m_i \pi_i$ . É importante notar que se  $y_i = 0$ , tem-se como componente típico dessa função  $\ell_i(\boldsymbol{\beta}) = m_i \log[(m_i - \mu_i)/m_i]$  e se  $y_i = m_i$ ,  $\ell_i(\boldsymbol{\beta}) = m_i \log(\mu_i/m_i)$ .

Para o modelo logístico linear, obtém-se  $\eta_i = g(\mu_i) = \log[\mu_i/(m_i - \mu_i)]$ . As iterações em (3.5) são realizadas com matriz de pesos  $\mathbf{W} = \text{diag}\{\mu_i(m_i - \mu_i)/m_i\}$ ,  $\mathbf{G} = \text{diag}\{m_i/[\mu_i(m_i - \mu_i)]\}$  e variável dependente ajustada  $\mathbf{z}$  com componentes iguais a  $z_i = \eta_i + [m_i(y_i - \mu_i)]/[\mu_i(m_i - \mu_i)]$ . O algoritmo (3.5), em geral, converge, exceto quando ocorrem médias ajustadas próximas a zero ou ao índice  $m_i$ .

Nos modelos log-lineares para análise de observações na forma de contagens, a variável resposta tem distribuição de Poisson  $P(\mu_i)$  com função de ligação logarítmica e, portanto,  $\eta_i = \log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$ ,  $i = 1, \dots, n$ . Nesse caso, as iterações em (3.5) são realizadas com matriz de pesos  $\mathbf{W} = \text{diag}\{\mu_i\}$ ,  $\mathbf{G} = \text{diag}\{\mu_i^{-1}\}$  e variável dependente ajustada  $\mathbf{z}$  com componentes iguais a  $z_i = \eta_i + (y_i - \mu_i)/\mu_i$ . Esse caso especial do algoritmo (3.5) foi apresentado, primeiramente, por Haberman (1978).

Para analisar dados contínuos, três modelos são, usualmente, adotados com função de variância potência  $V(\mu) = \mu^\delta$  para  $\delta = 0$  (normal),  $\delta = 2$  (gama) e  $\delta = 3$  (normal inversa). Para a função de variância potência, a matriz  $\mathbf{W}$  no algoritmo (3.5) é igual a  $\mathbf{W} = \text{diag}\{\mu_i^{-\delta}(d\mu_i/d\eta_i)^2\}$  sendo  $\delta$  qualquer real especificado. Outras funções de variância podem ser adotadas no algoritmo (3.5) como aquelas dos modelos de quase-verossimilhança que serão estudados na Seção ???. Por exemplo,  $V(\mu) = \mu^2(1 - \mu)^2$ ,  $V(\mu) = \mu + \delta\mu^2$  (binomial negativo) ou  $V(\mu) = 1 + \mu^2$  (secante hiperbólica generalizada, Seção 1.3).

O algoritmo (3.5) pode ser usado para ajustar inúmeros outros modelos, como aqueles baseados na família exponencial (1.1) que estão descritos em Cordeiro et al. (1995), bastando identificar as funções de variância e de ligação.

### 3.4 Resultados adicionais na estimação

A partir da obtenção da EMV  $\hat{\boldsymbol{\beta}}$  em (3.5), podem-se calcular as EMVs dos preditores lineares  $\hat{\boldsymbol{\eta}} = \mathbf{X}\hat{\boldsymbol{\beta}}$  e das médias  $\hat{\boldsymbol{\mu}} = g^{-1}(\hat{\boldsymbol{\eta}})$ . A EMV do vetor  $\boldsymbol{\theta}$  de parâmetros canônicos é, simplesmente, igual a  $\hat{\boldsymbol{\theta}} = q(\hat{\boldsymbol{\mu}})$ .

A inversa da matriz de informação estimada em  $\hat{\boldsymbol{\beta}}$  representa a estrutura de covariância assintótica de  $\hat{\boldsymbol{\beta}}$ , isto é, a matriz de covariância de  $\hat{\boldsymbol{\beta}}$  quando  $n \rightarrow \infty$ . Logo, a matriz de covariância de  $\hat{\boldsymbol{\beta}}$  é estimada por

$$\widehat{\text{Cov}}(\hat{\boldsymbol{\beta}}) = \phi(\mathbf{X}^T \widehat{\mathbf{W}} \mathbf{X})^{-1}, \quad (3.8)$$

em que  $\widehat{\mathbf{W}}$  é a matriz de pesos  $\mathbf{W}$  avaliada em  $\hat{\boldsymbol{\beta}}$ . Intervalos de confiança assintóticos para os parâmetros  $\beta'$ s podem ser deduzidos da aproximação (3.8). Observa-se que

o parâmetro de dispersão  $\phi$  é um fator multiplicativo na matriz de covariâncias assintóticas de  $\hat{\boldsymbol{\beta}}$ .

A estrutura da covariância assintótica dos EMVs dos preditores lineares em  $\hat{\boldsymbol{\eta}}$  é determinada diretamente de  $\text{Cov}(\hat{\boldsymbol{\eta}}) = \mathbf{X}\text{Cov}(\hat{\boldsymbol{\beta}})\mathbf{X}^T$ . Logo,

$$\widehat{\text{Cov}}(\hat{\boldsymbol{\eta}}) = \phi \mathbf{X}(\mathbf{X}^T \widehat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^T. \quad (3.9)$$

A matriz  $\mathbf{Z} = \{z_{ij}\} = \mathbf{X}(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T$  da expressão (3.9) desempenha um papel importante na teoria assintótica dos MLGs (Cordeiro, 1983; Cordeiro e McCullagh, 1991). Essa matriz surge no cálculo do valor esperado da função desvio (Seção 4.2) até termos de ordem  $O(n^{-1})$  e no valor esperado da estimativa  $\hat{\boldsymbol{\eta}}$  até essa ordem.

A estrutura de covariância assintótica dos EMVs das médias em  $\hat{\boldsymbol{\mu}}$  pode ser calculada expandindo  $\hat{\boldsymbol{\mu}} = g^{-1}(\hat{\boldsymbol{\eta}})$  em série de Taylor. Tem-se

$$\hat{\boldsymbol{\mu}} = g^{-1}(\boldsymbol{\eta}) + \frac{dg^{-1}(\boldsymbol{\eta})}{d\boldsymbol{\eta}}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})$$

e, portanto,

$$\text{Cov}(\hat{\boldsymbol{\mu}}) = \mathbf{G}^{-1} \text{Cov}(\hat{\boldsymbol{\eta}}) \mathbf{G}^{-1}, \quad (3.10)$$

em que a matriz diagonal  $\mathbf{G} = \text{diag}\{d\eta_i/d\mu_i\}$  foi definida na Seção 3.2. Essa matriz é estimada por

$$\widehat{\text{Cov}}(\hat{\boldsymbol{\mu}}) = \phi \widehat{\mathbf{G}}^{-1} \mathbf{X}(\mathbf{X}^T \widehat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^T \widehat{\mathbf{G}}^{-1}.$$

As matrizes  $\text{Cov}(\hat{\boldsymbol{\eta}})$  e  $\text{Cov}(\hat{\boldsymbol{\mu}})$  em (3.9) e (3.10) são de ordem  $O(n^{-1})$ .

Os erros-padrão estimados  $\hat{z}_{ii}^{1/2}$  de  $\hat{\eta}_i$  e os coeficientes de correlação estimados,

$$\widehat{\text{Corr}}(\hat{\eta}_i, \hat{\eta}_j) = \frac{\hat{z}_{ij}}{(\hat{z}_{ii}\hat{z}_{jj})^{1/2}},$$

dos EMVs dos preditores lineares  $\eta_1, \dots, \eta_n$  são resultados aproximados que dependem fortemente do tamanho da amostra. Entretanto, são guias úteis de informação sobre a confiabilidade e a interdependência das estimativas dos preditores lineares



e podem, também, ser usados para obter intervalos de confiança aproximados para esses parâmetros. Para alguns MLGs, é possível achar uma forma fechada para a inversa da matriz de informação e, conseqüentemente, para as estruturas de covariância assintótica das estimativas  $\hat{\beta}$ ,  $\hat{\eta}$  e  $\hat{\mu}$ .

Frequentemente, nos modelos de análise de variância, considera-se que os dados são originados de populações com variâncias iguais. Em termos dos MLGs, isso implica no uso de uma função de ligação  $g(\cdot)$ , tal que  $\mathbf{W}$  não depende da média  $\mu$  e, portanto, que a matriz de informação seja constante. Nesse caso, pelo menos assintoticamente, a matriz de covariâncias das estimativas dos parâmetros lineares é estabilizada.

Essa função de ligação é denominada *estabilizadora* e implica na constância da matriz de pesos do algoritmo de estimação. A função de ligação estabilizadora será vista (como o caso  $\delta = 1/2$ ) na Seção ??, mas pode ser obtida como solução da equação diferencial  $d\mu/d\eta = k d\eta/d\theta$ , sendo  $k$  uma constante arbitrária. Por exemplo, para os modelos gama e Poisson, as soluções dessa equação são o logaritmo e a raiz quadrada, respectivamente. Para as funções de ligação estabilizadoras, é mais fácil obter uma forma fechada para a matriz de informação, que depende inteiramente da matriz modelo, isto é, do delineamento do experimento.

Em muitas situações, os parâmetros de interesse não são aqueles básicos dos MLGs. Seja  $\gamma = (\gamma_1, \dots, \gamma_q)^T$  um vetor de parâmetros, em que  $\gamma_i = h_i(\beta)$ , sendo as funções  $h_i(\cdot)$ ,  $i = 1, \dots, q$ , conhecidas. Supõe-se que essas funções, em geral, não-lineares, são suficientemente bem comportadas. Seja a matriz  $q \times p$  de derivadas  $\mathbf{D} = \{\partial h_i / \partial \beta_j\}$ . As estimativas  $\hat{\gamma}_1, \dots, \hat{\gamma}_q$  podem ser calculadas diretamente de  $\hat{\gamma}_i = h_i(\hat{\beta})$ , para  $i = 1, \dots, q$ . A matriz de covariâncias assintóticas de  $\hat{\gamma}$  é igual a  $\phi \mathbf{D}(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{D}^T$  e deve ser estimada no ponto  $\hat{\beta}$ . Uma aplicação será descrita na Seção ??.

Considere, por exemplo, que após o ajuste de um MLG, tenha-se interesse em estudar as estimativas dos parâmetros  $\gamma$  definidos por um modelo de regressão

assintótico em três parâmetros  $\beta_0$ ,  $\beta_1$  e  $\beta_2$

$$\gamma_r = \beta_0 - \beta_1 \beta_2^{z_r}, \quad r = 1, \dots, q.$$

A matriz  $D$  de dimensões  $q \times 3$  é, portanto, igual a

$$D = \begin{bmatrix} 1 & -\beta_2^{z_1} & -\beta_1 \beta_2^{z_1} \log \beta_2 \\ \dots & \dots & \dots \\ 1 & -\beta_2^{z_q} & -\beta_1 \beta_2^{z_q} \log \beta_2 \end{bmatrix}.$$

### 3.5 Seleção do modelo

É difícil propor uma estratégia geral para o processo de escolha de um MLG a ser ajustado ao vetor de observações. O processo está intimamente relacionado ao problema fundamental da Estatística que, segundo Fisher, é “o que se deve fazer com os dados?”.

Em geral, o algoritmo de ajuste deve ser aplicado não a um MLG isolado mas a vários modelos de um conjunto bem amplo que deve ser, realmente, relevante para a natureza das observações que se deseja analisar. Se o processo é aplicado a um único modelo, não levando em conta possíveis modelos alternativos, existe o risco de não se obter um dos modelos mais adequados aos dados. Esse conjunto de modelos pode ser formulado de várias maneiras:

- (a) definindo uma família de funções de ligação;
- (b) considerando diferentes opções para a escala de medição;
- (c) adicionando (ou retirando) vetores colunas independentes a partir de uma matriz básica original.

Pode-se propor um conjunto de modelos para dados estritamente positivos, usando-se a família potência de funções de ligação  $\eta = g(\mu; \lambda) = (\mu^\lambda - 1)\lambda^{-1}$ , em que  $\lambda$  é um parâmetro que indexa o conjunto. Para dados reais positivos ou negativos, outras famílias podem ser definidas como  $g(\mu; \lambda) = [\exp(\lambda\mu) - 1]\lambda^{-1}$ . A EMV de

$\lambda$ , em geral, define um modelo bastante adequado, porém, muitas vezes, de difícil interpretação.

Nos MLGs, o fator escala não é tão crucial como no modelo clássico de regressão, pois constância da variância e normalidade não são essenciais para a distribuição da variável resposta e, ainda, pode-se achar uma estrutura aditiva aproximada de termos para representar a média da distribuição, usando uma função de ligação apropriada diferente da escala de medição dos dados. Entretanto, não são raros os casos em que os dados devem ser primeiramente transformados para se obter um MLG produzindo um bom ajuste.

Devem-se analisar não somente os dados brutos mas procurar modelos alternativos aplicados aos dados transformados  $z = h(y)$ . O problema crucial é a escolha da função de escala  $h(\cdot)$ . No modelo clássico de regressão, essa escolha visa a combinar, aproximadamente, normalidade e constância da variância do erro aleatório, bem como, aditividade dos efeitos sistemáticos. Entretanto, não existe nenhuma garantia que tal escala  $h(\cdot)$  exista, nem mesmo que produza algumas das propriedades desejadas.

Como uma ilustração, suponha que as observações  $y$  representam contagens, com estrutura de Poisson de média  $\mu$  e que os efeitos sistemáticos dos fatores que classificam os dados sejam multiplicativos. A transformação  $\sqrt{y}$  produz, para valores grandes de  $\mu$ ,  $E(\sqrt{Y}) \doteq \sqrt{\mu}$  e  $\text{Var}(\sqrt{Y}) \doteq 1/4$ , sendo os erros de ordem  $\mu^{-1/2}$ . Portanto, a escala raiz quadrada implica na constância da variância dos dados transformados. Entretanto, se o objetivo é obter uma normalidade aproximada, uma escala preferida deve ser  $h(y) = \sqrt[3]{y^2}$ , pois o coeficiente de assimetria padronizado de  $Y^{2/3}$  é de ordem  $\mu^{-1}$ , ao invés de ordem  $\mu^{-1/2}$  para  $Y$  ou  $Y^{1/2}$  (Seção ??). Ainda, a escala  $h(y) = \log(y)$  é bem melhor para obtenção da aditividade dos efeitos sistemáticos.

Não existe nenhuma escala que produza os três efeitos desejados, embora a escala definida por  $h(y) = (3y^{1/2} - 3y^{1/6}\mu^{1/3} + \mu^{1/2})/6$ , se  $y \neq 0$  e  $h(y) = [-(2\mu)^{1/2} + \mu^{-1/2}]/6$ , se  $y = 0$ , conduza à simetria e constância da variância (McCullagh e Nelder, 1989, Capítulo 6). As probabilidades nas extremidades da distribuição de Poisson

podem ser calculadas por  $P(Y \geq y) \doteq 1 - \Phi[h(y - 1/2)]$ , com erro de ordem  $\mu^{-1}$ , em que  $\Phi(\cdot)$  é a f.d.a. da distribuição normal reduzida.

A terceira parte na seleção do modelo consiste em definir o conjunto de variáveis explanatórias a serem incluídas na estrutura linear. Considere um certo número de possíveis variáveis explanatórias  $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m)}$ , em que cada vetor coluna  $\mathbf{x}^{(r)}$  é de dimensão  $n$ , definindo um conjunto amplo de  $2^m$  modelos. O objetivo é selecionar um modelo de  $p \leq m$  variáveis explanatórias, cujos valores ajustados expliquem adequadamente os dados. Se  $m$  for muito grande, torna-se impraticável o exame de todos esses  $2^m$  modelos, mesmo considerando os avanços da tecnologia computacional.

Um processo simples de seleção é de natureza sequencial, adicionando (ou eliminando) variáveis explanatórias (uma de cada vez) a partir de um modelo original até se obterem modelos adequados. Esse método sequencial tem várias desvantagens, tais como:

(a) modelos potencialmente úteis podem não ser descobertos, se o procedimento é finalizado numa etapa anterior, para o qual nenhuma variável explanatória isolada mostrou-se razoável de ser explorada;

(b) modelos similares (ou mesmo melhores) baseados em subconjuntos de variáveis explanatórias, distantes das variáveis em exame, podem não ser considerados.

Devido aos avanços recentes da estatística computacional, os métodos sequenciais (“stepwise methods”) foram substituídos por procedimentos ótimos de busca de modelos. O procedimento de busca examina, sistematicamente, somente os modelos mais promissores de determinada dimensão  $k$  e, baseado em algum critério, apresenta os resultados de ajuste dos melhores modelos de  $k$  variáveis explanatórias, com  $k$  variando no processo de 1 até o tamanho  $p$  do subconjunto final de modelos considerados bons.

Deve-se sempre tentar eliminar a priori modelos medíocres, observando a estrutura dos dados por meio de análises exploratórias gráficas. Na seleção do modelo,

sempre será feito um balanço entre o grau de complexidade e a qualidade de ajuste do modelo.

### 3.6 Considerações sobre a função de verossimilhança

Expandindo a função suporte  $\ell = \ell(\boldsymbol{\beta})$ , descrita na Seção 3.2, em série multivariada de Taylor ao redor de  $\hat{\boldsymbol{\beta}}$  e notando que  $\mathbf{U}(\hat{\boldsymbol{\beta}}) = 0$ , obtém-se, aproximadamente,

$$\hat{\ell} - \ell \doteq \frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \hat{\mathbf{J}}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}), \quad (3.11)$$

em que  $\hat{\ell} = \ell(\hat{\boldsymbol{\beta}})$  e  $\hat{\mathbf{J}}$  é a informação observada (Seção 3.2) em  $\hat{\boldsymbol{\beta}}$ . Essa equação aproximada revela que a diferença entre o suporte máximo e o suporte num ponto arbitrário, que pode ser considerada como a quantidade de informação dos dados sobre  $\boldsymbol{\beta}$ , é proporcional a  $\hat{\mathbf{J}}$  (isto é, à informação observada no ponto  $\hat{\boldsymbol{\beta}}$ ). O determinante de  $\hat{\mathbf{J}}$  ( $|\hat{\mathbf{J}}|$ ) pode ser interpretado, geometricamente, como a curvatura esférica da superfície suporte no seu máximo. A forma quadrática do lado direito de (3.11) aproxima a superfície suporte por um parabolóide, passando pelo seu ponto de máximo, com a mesma curvatura esférica da superfície nesse ponto. O recíproco de  $|\hat{\mathbf{J}}|$  mede a variabilidade de  $\boldsymbol{\beta}$  ao redor da EMV  $\hat{\boldsymbol{\beta}}$ . E, como esperado, quanto maior a informação sobre  $\boldsymbol{\beta}$  menor será a dispersão de  $\boldsymbol{\beta}$  ao redor de  $\hat{\boldsymbol{\beta}}$ .

A interpretação geométrica desses conceitos é melhor compreendida no caso uniparamétrico, pois (3.11) reduz-se à equação de uma parábola  $\ell \doteq \hat{\ell} - \frac{1}{2}(\beta - \hat{\beta})^2 \hat{J}$ . Uma inspeção gráfica mostrará que essa parábola aproxima a curva suporte, coincidindo no ponto máximo e tendo a mesma curvatura dessa curva em  $\hat{\beta}$ , revelando ainda que quanto maior a curvatura, menor a variação de  $\beta$  em torno de  $\hat{\beta}$ .

A equação (3.11) implica que a função de verossimilhança  $L = L(\boldsymbol{\beta})$  num ponto qualquer  $\boldsymbol{\beta}$  segue, aproximadamente, a expressão

$$L \doteq \hat{L} \exp \left[ -\frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \hat{\mathbf{J}}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right], \quad (3.12)$$

em que  $\hat{L}$  é a função de verossimilhança avaliada em  $\hat{\beta}$ , que representa a forma da curva normal multivariada com média  $\hat{\beta}$  e estrutura de covariância igual a  $\hat{\mathbf{J}}^{-1}$ . Usando-se essa aproximação, pode-se, então, considerar o vetor de parâmetros como se fosse um vetor de variáveis aleatórias tendo distribuição normal multivariada com média igual à EMV  $\hat{\beta}$  e estrutura de covariância  $\hat{\mathbf{J}}^{-1}$ . Quando a função suporte for quadrática, a função de verossimilhança  $L$  terá a forma da distribuição normal multivariada. A forma de  $L$  se aproximará mais da distribuição normal quando  $n$  tender para infinito.

O lado direito de (3.12) é bem interpretado no contexto Bayesiano. Considere qualquer função densidade a priori não-nula para  $\beta$ , por exemplo,  $\pi(\beta)$ . Pelo teorema de Bayes, pode-se escrever a função densidade a posteriori de  $\beta$  como proporcional a  $\pi(\beta)L$ . Quando  $n \rightarrow \infty$ , pois  $\pi(\beta)$  não depende de  $n$ , a função densidade a posteriori de  $\beta$  segue da equação (3.12) com uma constante de proporcionalidade adequada, e, então, pode ser aproximada pela distribuição normal multivariada  $N(\hat{\beta}, \hat{\mathbf{J}}^{-1})$ . A demonstração matemática desse resultado não se insere nos objetivos desse texto. No caso uniparamétrico, a variabilidade de  $\beta$  fica restrita ao intervalo  $|\beta - \hat{\beta}| \leq 3\hat{J}^{-1/2}$  com probabilidade próxima de um.

A fórmula (3.12) mostra a decomposição da função de verossimilhança, pelo menos para  $n$  grande, revelando, pelo teorema da fatoração, a suficiência assintótica da EMV. Conclui-se que, embora as EMVs não sejam necessariamente suficientes para os parâmetros do modelo, essa suficiência será alcançada quando a dimensão do vetor de observações tender para infinito.

Citam-se, aqui, algumas propriedades da matriz de informação. Seja  $\mathbf{K}_y(\beta)$  a informação sobre um vetor paramétrico  $\beta$  contida nos dados  $\mathbf{y}$  obtidos de certo experimento. A informação é aditiva para amostras  $\mathbf{y}$  e  $\mathbf{z}$  independentes, isto é,  $\mathbf{K}_{y+z}(\beta) = \mathbf{K}_y(\beta) + \mathbf{K}_z(\beta)$ . Como  $\hat{\mathbf{U}} = \mathbf{U}(\hat{\beta}) = 0$ , segue-se a relação aproximada (por expansão multivariada de Taylor)

$$\hat{\beta} - \beta \doteq \mathbf{J}^{-1}\mathbf{U} \quad (3.13)$$

entre a EMV  $\hat{\beta}$ , a função escore  $\mathbf{U} = \mathbf{U}(\beta)$  e a informação observada  $\mathbf{J} = \mathbf{J}(\beta)$  avaliadas no ponto  $\beta$  próximo de  $\hat{\beta}$ .

O método de Newton-Raphson, introduzido na Seção 3.2, de cálculo da EMV consiste em usar a equação (3.13) iterativamente. Obtém-se uma nova estimativa  $\beta^{(m+1)}$  a partir de uma estimativa anterior  $\beta^{(m)}$  por meio de

$$\beta^{(m+1)} = \beta^{(m)} + \mathbf{J}^{(m)^{-1}} \mathbf{U}^{(m)}, \quad (3.14)$$

em que quantidades avaliadas na  $m$ -ésima iteração do procedimento iterativo são indicadas com o superescrito  $(m)$ . O processo é, então, repetido a partir de  $\beta^{(m+1)}$  até a distância entre  $\beta^{(m+1)}$  e  $\beta^{(m)}$  se tornar desprezível ou menor do que uma quantidade pequena especificada. Geometricamente, uma iteração do método equivale a ajustar um parabolóide à superfície suporte em  $\beta^{(m)}$ , tendo o mesmo gradiente e curvatura da superfície nesse ponto, e, então, obter o ponto máximo do parabolóide que corresponderá à estimativa atualizada  $\beta^{(m+1)}$ . Quando  $\beta$  é um escalar, a equação (3.14) reduz-se a  $\beta^{(m+1)} = \beta^{(m)} - U^{(m)} / U'^{(m)}$ , sendo  $U' = dU/d\beta$ , que representa o método das tangentes bastante usado para calcular a solução de uma equação não-linear  $\hat{U} = 0$ .

A sequência  $\{\beta^{(m)}; m \geq 1\}$  gerada depende, fundamentalmente, do vetor inicial  $\beta^{(1)}$ , dos valores amostrais e do modelo estatístico e, em determinadas situações, em que  $n$  é pequeno, pode revelar irregularidades específicas nos valores amostrais obtidos do experimento e, portanto, pode não convergir e mesmo divergir da EMV  $\hat{\beta}$ . Mesmo quando há convergência, se a função de verossimilhança tem múltiplas raízes, não há garantia de que o procedimento convirja para a raiz correspondente ao maior valor absoluto da função de verossimilhança. No caso uniparamétrico, se a estimativa inicial  $\beta^{(1)}$  for escolhida próxima de  $\hat{\beta}$  e se  $J^{(m)}$  para  $m \geq 1$  for limitada por um número real positivo, existirá uma chance apreciável que essa sequência convirja para  $\hat{\beta}$ .

A expressão (3.13) tem uma forma alternativa equivalente, assintoticamente, pois pela lei dos grandes números  $\mathbf{J}$  deve convergir para  $\mathbf{K}$  quando  $n \rightarrow \infty$ . Assim,

substituindo a informação observada em (3.13) pela esperada, obtém-se a aproximação de primeira ordem

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \doteq \mathbf{K}^{-1}\mathbf{U}. \quad (3.15)$$

O procedimento iterativo baseado em (3.15) é denominado *método score* de Fisher para parâmetros, isto é,  $\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \mathbf{K}^{(m)-1}\mathbf{U}^{(m)}$ , como foi explicitado na equação (3.3). O aspecto mais trabalhoso dos dois esquemas iterativos é o procedimento de inversão das matrizes  $\mathbf{J}$  e  $\mathbf{K}$ . Ambos os procedimentos são muito sensíveis em relação à estimativa inicial  $\boldsymbol{\beta}^{(1)}$ . Se o vetor  $\boldsymbol{\beta}^{(1)}$  for uma estimativa consistente, ambos os métodos convergirão em apenas um passo para uma estimativa eficiente, assintoticamente.

Existe evidência empírica que o método de Fisher é melhor em termos de convergência do que o método de Newton-Raphson. Ainda, tem a vantagem de incorporar (por meio da matriz de informação) as características específicas do modelo estatístico. Ademais, em muitas situações, é mais fácil determinar a inversa de  $\mathbf{K}$  em forma fechada do que a inversa de  $\mathbf{J}$ , sendo a primeira menos sensível às variações de  $\boldsymbol{\beta}$  do que a segunda. Nesse sentido,  $\mathbf{K}$  pode ser considerada em alguns modelos, aproximadamente, constante em todo o processo iterativo, requerendo que a inversão seja realizada apenas uma vez. Uma vantagem adicional do método score é que  $\mathbf{K}^{-1}$  é usada para calcular aproximações de primeira ordem para as variâncias e covariâncias das estimativas  $\hat{\beta}_1, \dots, \hat{\beta}_p$ .

Os procedimentos iterativos descritos são casos especiais de uma classe de algoritmos iterativos para maximizar o logaritmo da função de verossimilhança  $\ell(\boldsymbol{\beta})$ . Essa classe tem a forma

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} - s^{(m)}\mathbf{Q}^{(m)}\mathbf{U}^{(m)}, \quad (3.16)$$

em que  $s^{(m)}$  é um escalar,  $\mathbf{Q}^{(m)}$  é uma matriz quadrada que determina a direção da mudança de  $\boldsymbol{\beta}^{(m)}$  para  $\boldsymbol{\beta}^{(m+1)}$  e  $\mathbf{U}^{(m)}$  é o vetor gradiente do logaritmo da função de verossimilhança  $\ell(\boldsymbol{\beta})$ , com todas essas quantidades variando no processo iterativo.



Os algoritmos iniciam num ponto  $\beta^{(1)}$  e procedem, por meio da equação (3.16), para calcular aproximações sucessivas para a EMV  $\hat{\beta}$ . Vários algoritmos nessa classe são descritos por Judge et al. (1985). Nos procedimentos iterativos de Newton-Raphson e escore de Fisher,  $s^{(m)}$  é igual a um, e a matriz de direção  $\mathbf{Q}^{(m)}$  é igual à inversa da matriz Hessiana e à inversa do valor esperado dessa matriz, respectivamente. Esses dois procedimentos devem ser iniciados a partir de uma estimativa consistente com o objetivo de se garantir convergência para  $\hat{\beta}$ . A escolha do melhor algoritmo em (3.16) é função da geometria do modelo em consideração e, em geral, não existe um algoritmo superior aos demais em qualquer espectro amplo de problemas de estimação.

### 3.7 Exercícios

1. Defina o algoritmo de estimação especificado em (3.5) para os modelos canônicos relativos às distribuições descritas na Seção 1.3 (Tabela 1.1), calculando  $\mathbf{W}$ ,  $\mathbf{G}$  e  $\mathbf{z}$ .
2. Defina o algoritmo de estimação especificado em (3.5), calculando  $\mathbf{W}$ ,  $\mathbf{G}$  e  $\mathbf{z}$  para os modelos normal, gama, normal inverso e Poisson com função de ligação potência  $\eta = \mu^\lambda$ ,  $\lambda$  conhecido (Cordeiro, 1986). Para o modelo normal, considere, ainda, o caso da função de ligação logarítmica  $\eta = \log(\mu)$ .
3. Defina o algoritmo (3.5), calculando  $\mathbf{W}$ ,  $\mathbf{G}$  e  $\mathbf{z}$ , para o modelo binomial com função de ligação  $\eta = \log\{[(1 - \mu)^{-\lambda} - 1]\lambda^{-1}\}$ ,  $\lambda$  conhecido. Deduza, ainda, as formas do algoritmo para os modelos (c) e (d), definidos na Tabela 2.8 da Seção 2.5.1.
4. Considere a estrutura linear  $\eta_i = \beta x_i$ ,  $i = 1, \dots, n$ , com um único parâmetro  $\beta$  desconhecido e função de ligação  $\eta = (\mu^\lambda - 1)\lambda^{-1}$ ,  $\lambda$  conhecido. Calcule a EMV de  $\beta$  para os modelos normal, Poisson, gama, normal inverso e binomial negativo. Faça o mesmo para o modelo binomial com função de ligação especificada no Exercício 3.

Deduza ainda as estimativas no caso de  $x_1 = x_2 = \dots = x_n$ .

**5.** Para os modelos e funções de ligação citados no Exercício 4, calcule as EMVs de  $\alpha$  e  $\beta$ , considerando a estrutura linear  $\eta_i = \alpha + \beta x_i$ ,  $i = 1, \dots, n$ . Deduza, ainda, a estrutura de covariância aproximada dessas estimativas.

**6.** Caracterize as distribuições log-normal e log-gama no contexto dos MLGs, definindo o algoritmo de ajuste desses modelos com a função de ligação potência  $\eta = \mu^\lambda$ ,  $\lambda$  conhecido.

**7.** Formule o procedimento iterativo de cálculo das estimativas de mínimos quadrados dos parâmetros  $\beta$ 's nos MLGs, que equivale a minimizar  $(\mathbf{y} - \boldsymbol{\mu})^T \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ , em que  $\mathbf{V} = \text{diag}\{V_1, \dots, V_n\}$ , com relação a  $\boldsymbol{\beta}$ . Como aplicação, obtenha essas estimativas nos Exercícios 4 e 5.

**8.** Deduza a forma da matriz de informação para o modelo log-linear associado a uma tabela de contingência com dois fatores sem interação, sendo uma observação por cela. Faça o mesmo para o modelo de Poisson com função de ligação raiz quadrada. Qual a grande vantagem desse último modelo?

**9.** Calcule a forma da matriz de informação para os parâmetros  $\beta$  no modelo de classificação de um fator  $A$  com  $p$  níveis  $g(\mu_i) = \eta_i = \beta + \beta_i^A$ , com  $\beta_+^A = 0$ , considerando a variável resposta como normal, gama, normal inversa e Poisson. Determine as matrizes de covariância assintótica dos estimadores  $\hat{\boldsymbol{\beta}}$ ,  $\hat{\boldsymbol{\eta}}$  e  $\hat{\boldsymbol{\mu}}$ . Deduza as expressões desses estimadores.

**10.** Como o modelo binomial do Exercício 3 poderia ser ajustado se  $\lambda$  fosse desconhecido? E os modelos do Exercício 4, ainda  $\lambda$  desconhecido?

**11.** Sejam variáveis aleatórias  $Y_i$  com distribuições de Poisson  $P(\mu_i)$ ,  $i = 1, \dots, n$ , supostas independentes. Defina-se  $f(\cdot)$  como uma função diferenciável tal que  $[f(\mu +$

$x\mu^{1/2}) - f(\mu)]/\mu^{1/2}f'(\mu) = x + O(\mu^{-1/2})$ , para todo  $x$  com  $\mu \rightarrow \infty$ . Demonstre que a variável aleatória  $[f(Y_i) - f(\mu_i)]/[\mu_i^{1/2}f'(\mu_i)]$  converge em distribuição para a distribuição normal  $N(0, 1)$  quando  $\mu_i \rightarrow \infty$ . Prove que a parte do logaritmo da função de verossimilhança que só depende dos  $\mu_i$  tende, assintoticamente, para  $-2^{-1} \sum_{i=1}^n [f(y_i) - f(\mu_i)]^2 / [y_i f'(y_i)]^2$  quando  $\mu_i \rightarrow \infty, i = 1, \dots, n$ , em que  $y_1, \dots, y_n$  são as realizações dos  $Y$ .

**12.** A probabilidade de sucesso  $\pi = \mu/m$  de uma distribuição binomial  $B(m, \pi)$  depende de uma variável  $x$  de acordo com a relação  $\pi = F(\alpha + \beta x)$ , em que  $F(\cdot)$  é uma função de distribuição acumulada especificada. Considera-se que para os valores  $x_1, \dots, x_n$  de  $x$ ,  $m_1, \dots, m_n$  ensaios independentes foram realizados, sendo obtidas as proporções de sucessos  $p_1, \dots, p_n$ , respectivamente. Compare as estimativas  $\hat{\alpha}$  e  $\hat{\beta}$  para as escolhas de  $F(\cdot)$ : probito, logística, arcsen  $\sqrt{\cdot}$  e complemento log-log.

**13.** Considere a f.d.p.  $f(y) = \exp(-\sum_{i=1}^r \alpha_i y^i)$  com parâmetros  $\alpha_1, \dots, \alpha_r$  desconhecidos. Demonstre que as EMVs e dos momentos desses parâmetros coincidem.

**14.** Considere um modelo log-gama com componente sistemático  $\log(\mu_i) = \alpha + \mathbf{x}_i^T \boldsymbol{\beta}$  e parâmetro de dispersão  $\phi$ . Mostre que

$$E[\log(Y_i)] = \alpha^* + \mathbf{x}_i^T \boldsymbol{\beta}$$

e

$$\text{Var}[\log(Y_i)] = \psi'(\phi^{-1}),$$

em que  $\alpha^* = \alpha + \psi(\phi^{-1}) + \log(\phi)$ . Seja  $\tilde{\boldsymbol{\beta}}$  o estimador de mínimos quadrados de  $\boldsymbol{\beta}$  calculado do ajuste de um modelo de regressão linear aos dados transformados  $\log(y_i), i = 1, \dots, n$ . Mostre que  $\tilde{\boldsymbol{\beta}}$  é um estimador consistente de  $\boldsymbol{\beta}$ .

**15.** Demonstre que a covariância assintótica do EMV  $\hat{\boldsymbol{\beta}}$  de um modelo log-linear é igual a  $\text{Cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}\mathbf{W}\mathbf{X})^{-1}$ , sendo  $\mathbf{W} = \text{diag}\{\mu_1, \dots, \mu_n\}$ .

**16.** Proponha um algoritmo iterativo para calcular a EMV do parâmetro  $\alpha(> 0)$

na função de variância  $V = \mu + \alpha\mu^2$  de um modelo binomial negativo, supondo  $\log(\boldsymbol{\mu}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ .

**17.** Mostre que no modelo binomial negativo definido no exercício 16:

(a) os parâmetros  $\alpha$  e  $\beta$  são ortogonais;

(b)  $\text{Cov}(\hat{\boldsymbol{\beta}}) = \left( \sum_{i=1}^n \frac{\mu_i}{1 + \alpha\mu_i} \mathbf{x}_i \mathbf{x}_i^T \right)^{-1}$  em que  $\mathbf{x}_i$ ,  $i = 1, \dots, n$ , são as linhas da matriz  $\mathbf{X}$  do modelo;

(c)  $\text{Var}(\hat{\alpha}) = \left\{ \sum_{i=1}^n \alpha^{-4} \left[ \log(1 + \alpha\mu_i) - \sum_{j=0}^{y_i-1} \frac{1}{j + \alpha^{-1}} \right]^2 + \sum_{i=1}^n \frac{\mu_i}{\alpha^2(1 + \alpha\mu_i)} \right\}^{-1}$ .

**18.** O estimador de mínimos quadrados não-lineares de um MLG com função de ligação logarítmica minimiza  $\sum_{i=1}^n [y_i - \exp(\mathbf{x}_i^T \boldsymbol{\beta})]^2$ .

(a) Mostre como calcular esse estimador iterativamente.

(b) Calcule a variância assintótica desse estimador.

# Capítulo 4

## Métodos de Inferência

### 4.1 Distribuição dos estimadores dos parâmetros

No modelo clássico de regressão, em que a variável resposta tem distribuição normal e a função de ligação é a identidade, as distribuições dos estimadores dos parâmetros e das estatísticas usadas para verificar a qualidade do ajuste do modelo aos dados podem ser determinadas exatamente. Em geral, porém, a obtenção de distribuições exatas nos MLGs é muito complicada e resultados assintóticos são, rotineiramente, usados. Esses resultados, porém, dependem de algumas condições de regularidade e do número de observações independentes mas, em particular, para os MLGs essas condições são válidas (Fahrmeir e Kaufmann, 1985).

A ideia básica é que se  $\hat{\theta}$  é um estimador consistente para um parâmetro  $\theta$  e  $\text{Var}(\hat{\theta})$  é a variância desse estimador, então, para amostras grandes, tem-se:

i)  $\hat{\theta}$  é assintoticamente imparcial;

ii) a estatística

$$Z_n = \frac{\hat{\theta} - \theta}{\sqrt{\text{Var}(\hat{\theta})}} \rightarrow Z \text{ quando } n \rightarrow \infty, \text{ sendo que } Z \sim N(0, 1)$$

ou, de forma equivalente,

$$Z_n^2 = \frac{(\hat{\theta} - \theta)^2}{\text{Var}(\hat{\theta})} \rightarrow Z^2 \text{ quando } n \rightarrow \infty, \text{ sendo que } Z^2 \sim \chi_1^2.$$

Se  $\hat{\boldsymbol{\theta}}$  é um estimador consistente de um vetor  $\boldsymbol{\theta}$  de  $p$  parâmetros, tem-se, assintoticamente, que

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \mathbf{V}^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim \chi_p^2,$$

sendo  $\mathbf{V}$  a matriz de variâncias e covariâncias de  $\hat{\boldsymbol{\theta}}$ , suposta não-singular. Se  $\mathbf{V}$  é singular, usa-se uma matriz inversa generalizada ou, então, uma reparametrização de forma a se obter uma nova matriz de variâncias e covariâncias não-singular.

Considere-se um MLG definido por uma distribuição em (2.4), uma estrutura linear (2.6) e uma função de ligação (2.7). As propriedades do estimador  $\hat{\boldsymbol{\beta}}$  dos parâmetros lineares do modelo em relação à existência, finitude e unicidade serão apresentadas na Seção ???. Em geral, não é possível obter distribuições exatas para os EMVs e para as estatísticas de testes usadas nos MLGs e, então, trabalha-se com resultados assintóticos. As condições de regularidade que garantem esses resultados são válidas para os MLGs. É fato conhecido que os EMVs têm poucas propriedades que são satisfeitas para todos os tamanhos de amostras, como, por exemplo, suficiência e invariância. As propriedades assintóticas de segunda-ordem de  $\hat{\boldsymbol{\beta}}$ , como o viés de ordem  $O(n^{-1})$  e a sua matriz de covariância de ordem  $O(n^{-2})$ , foram estudadas por Cordeiro e McCullagh (1991) e Cordeiro (2004a,b,c), respectivamente.

Define-se o vetor escore  $\mathbf{U}(\boldsymbol{\beta}) = \partial \ell(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}$  como na Seção 3.2. Como, em problemas regulares (Cox e Hinkley, 1986, Capítulo 9), o vetor escore tem valor esperado zero e estrutura de covariância igual à matriz de informação  $\mathbf{K}$ , tem-se da equação (3.2) que  $E[\mathbf{U}(\boldsymbol{\beta})] = \mathbf{0}$  e

$$\text{Cov}[\mathbf{U}(\boldsymbol{\beta})] = E[\mathbf{U}(\boldsymbol{\beta})\mathbf{U}(\boldsymbol{\beta})^T] = E \left[ \frac{-\partial^2 \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} \right] = \mathbf{K}. \quad (4.1)$$

Conforme demonstrado na Seção 3.2, a matriz de informação para  $\boldsymbol{\beta}$  nos MLGs é expressa por  $\mathbf{K} = \phi^{-1} \mathbf{X}^T \mathbf{W} \mathbf{X}$ .

O teorema central do limite aplicado a  $\mathbf{U}(\boldsymbol{\beta})$  (que equivale a uma soma de variáveis aleatórias independentes) implica que a distribuição assintótica de  $\mathbf{U}(\boldsymbol{\beta})$  é normal  $p$ -variada, isto é,  $N_p(\mathbf{0}, \mathbf{K})$ . Para amostras grandes, a estatística escore de-

finida pela forma quadrática  $S_R = \mathbf{U}(\boldsymbol{\beta})^T \mathbf{K}^{-1} \mathbf{U}(\boldsymbol{\beta})$  tem, aproximadamente, distribuição  $\chi_p^2$  supondo o modelo, com o vetor de parâmetros  $\boldsymbol{\beta}$  especificado, verdadeiro.

De forma resumida têm-se, a seguir, algumas propriedades do estimador  $\hat{\boldsymbol{\beta}}$ :

i) O estimador  $\hat{\boldsymbol{\beta}}$  é assintoticamente não-viesado, isto é, para amostras grandes  $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ . Suponha que o logaritmo da função de verossimilhança tem um único máximo em  $\hat{\boldsymbol{\beta}}$  que está próximo do verdadeiro valor de  $\boldsymbol{\beta}$ . A expansão em série multivariada de Taylor do vetor escore  $\mathbf{U}(\hat{\boldsymbol{\beta}})$  em relação a  $\boldsymbol{\beta}$ , até termos de primeira ordem, substituindo-se a matriz de derivadas parciais de segunda ordem por  $-\mathbf{K}$ , implica em

$$\mathbf{U}(\hat{\boldsymbol{\beta}}) = \mathbf{U}(\boldsymbol{\beta}) - \mathbf{K}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathbf{0},$$

pois  $\hat{\boldsymbol{\beta}}$  é a solução do sistema de equações  $\mathbf{U}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$ . As variáveis aleatórias  $\mathbf{U}(\boldsymbol{\beta})$  e  $\mathbf{K}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$  diferem por quantidades estocásticas de ordem  $O_p(1)$ . Portanto, tem-se até ordem  $n^{-1/2}$  em probabilidade

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \mathbf{K}^{-1} \mathbf{U}(\boldsymbol{\beta}), \quad (4.2)$$

supondo que  $\mathbf{K}$  seja não-singular.

A expressão aproximada (4.2) é de grande importância para a determinação de propriedades do EMV  $\hat{\boldsymbol{\beta}}$ . As variáveis aleatórias  $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$  e  $\mathbf{K}^{-1} \mathbf{U}(\boldsymbol{\beta})$  diferem por variáveis aleatórias de ordem  $n^{-1}$  em probabilidade. Tem-se, então, que

$$E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathbf{K}^{-1} E[\mathbf{U}(\boldsymbol{\beta})] = \mathbf{0} \Rightarrow E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta},$$

pois  $E[\mathbf{U}(\boldsymbol{\beta})] = \mathbf{0}$  e, portanto,  $\hat{\boldsymbol{\beta}}$  é um estimador imparcial para  $\boldsymbol{\beta}$  (pelo menos assintoticamente). Na realidade,  $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + O(n^{-1})$ , sendo que o termo de ordem  $O(n^{-1})$  foi calculado por Cordeiro e McCullagh (1991). Mais recentemente, Cordeiro e Barroso (2007) obtiveram o termo de ordem  $O(n^{-2})$  da expansão de  $E(\hat{\boldsymbol{\beta}})$ .

ii) Denotando-se  $\mathbf{U} = \mathbf{U}(\boldsymbol{\beta})$  e usando (4.1) e (4.2) tem-se que a matriz de variâncias e covariâncias de  $\hat{\boldsymbol{\beta}}$ , para amostras grandes, é expressa por

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T] = \mathbf{K}^{-1} E(\mathbf{U} \mathbf{U}^T) \mathbf{K}^{-1T} = \mathbf{K}^{-1} \mathbf{K} \mathbf{K}^{-1} = \mathbf{K}^{-1},$$

pois  $\mathbf{K}^{-1}$  é simétrica. Na realidade,  $\text{Cov}(\hat{\boldsymbol{\beta}}) = \mathbf{K}^{-1} + O(n^{-2})$ , sendo que o termo matricial de ordem  $O(n^{-2})$  foi calculado por Cordeiro (2004a).

iii) Para amostras grandes, tem-se a aproximação

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{K}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \chi_p^2 \quad (4.3)$$

ou, de forma equivalente,

$$\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \mathbf{K}^{-1}), \quad (4.4)$$

ou seja,  $\hat{\boldsymbol{\beta}}$  tem distribuição assintótica normal multivariada, que é a base para a construção de testes e intervalos de confiança para os parâmetros lineares de um MLG. Para modelos lineares com a variável resposta seguindo a distribuição normal, as equações (4.3) e (4.4) são resultados exatos. Fahrmeir e Kaufmann (1985), em um artigo bastante matemático, desenvolveram condições gerais que garantem a consistência e a normalidade assintótica do EMV  $\hat{\boldsymbol{\beta}}$  nos MLGs.

Para amostras pequenas, como citado em i), o estimador  $\hat{\boldsymbol{\beta}}$  é viesado e torna-se necessário computar o viés de ordem  $n^{-1}$ , que pode ser apreciável. Também, para  $n$  não muito grande, como citado em ii), a estrutura de covariância dos EMVs dos parâmetros lineares difere de  $\mathbf{K}^{-1}$ . Uma demonstração rigorosa dos resultados assintóticos (4.3) e (4.4) exige argumentos do teorema central do limite adaptado ao vetor escore  $\mathbf{U}(\boldsymbol{\beta})$  e da lei fraca dos grandes números aplicada à matriz de informação  $\mathbf{K}$ . Pode-se, então, demonstrar, com mais rigor, a normalidade assintótica de  $\hat{\boldsymbol{\beta}}$ , com média igual ao parâmetro verdadeiro  $\boldsymbol{\beta}$  desconhecido, e com matriz de covariância consistentemente estimada por  $\hat{\mathbf{K}}^{-1} = \phi(\mathbf{X}^T \widehat{\mathbf{W}} \mathbf{X})^{-1}$ , em que  $\widehat{\mathbf{W}}$  é a matriz de pesos  $\mathbf{W}$  avaliada em  $\hat{\boldsymbol{\beta}}$ .

Para as distribuições binomial e de Poisson,  $\phi = 1$ . Se o parâmetro de dispersão  $\phi$  for constante para todas as observações e desconhecido afetará a matriz de covariância assintótica  $\hat{\mathbf{K}}^{-1}$  de  $\hat{\boldsymbol{\beta}}$  mas não o valor de  $\hat{\boldsymbol{\beta}}$ . Na prática, se  $\phi$  for desconhecido, deverá ser substituído por alguma estimativa consistente (Seção 4.4).

A distribuição assintótica normal multivariada  $N_p(\boldsymbol{\beta}, \mathbf{K}^{-1})$  de  $\hat{\boldsymbol{\beta}}$  é a base da construção de testes e intervalos de confiança, em amostras grandes, para os



parâmetros lineares dos MLGs. O erro dessa aproximação para a distribuição de  $\hat{\beta}$  é de ordem  $n^{-1}$  em probabilidade, significando que os cálculos de probabilidade baseados na função de distribuição acumulada da distribuição normal assintótica  $N_p(\beta, K^{-1})$  apresentam erros de ordem de magnitude  $n^{-1}$ .

A distribuição assintótica normal multivariada  $N_p(\beta, K^{-1})$  será uma boa aproximação para a distribuição de  $\hat{\beta}$  se o logaritmo da função de verossimilhança for razoavelmente uma função quadrática. Pelo menos, assintoticamente, todos os logaritmos das funções de verossimilhança têm essa forma. Para amostras pequenas, esse fato pode não ocorrer para  $\beta$ , embora possa existir uma reparametrização  $\gamma = h(\beta)$  que torne o logaritmo da função de verossimilhança uma função, aproximadamente, quadrática. Assim, testes e regiões de confiança mais precisos poderão ser baseados na distribuição assintótica de  $\hat{\gamma} = h(\hat{\beta})$ .

Anscombe (1964), no caso de um único parâmetro  $\beta$ , obtém uma parametrização geral que elimina a assimetria do logaritmo da função de verossimilhança. A solução geral é da forma

$$\gamma = h(\beta) = \int \exp \left[ \frac{1}{3} \int v(\beta) d\beta \right] d\beta, \quad (4.5)$$

em que  $v(\beta) = d^3 \ell(\beta) / d\beta^3 [d^2 \ell(\beta) / d\beta^2]^{-1}$ . Essa transformação tem a propriedade de anular a derivada de terceira ordem do logaritmo da função de verossimilhança, em relação a  $\gamma$ , e, portanto, eliminar a principal contribuição da assimetria.

Para os MLGs, a assimetria do logaritmo da função de verossimilhança pode ser eliminada usando uma função de ligação apropriada, como será explicado na Seção ?? (caso  $\delta = 1/3$ ). Usando-se a expressão (4.5), obtém-se, diretamente,  $\eta = \int \exp \{ \int b'''(\theta) / [3b''(\theta)] d\theta \} d\theta = \int b''(\theta)^{1/3} d\theta$ , a função de ligação que simetriza  $\ell(\beta)$ . Quando a função de ligação é diferente desse caso e se o vetor  $\beta$  tem dimensão maior do que 1, em geral, não é possível anular a assimetria. Em particular, reparametrizações componente a componente  $\gamma_i = h(\beta_i)$ ,  $i = 1, \dots, p$ , não apresentam um bom aperfeiçoamento na forma do logaritmo da função de verossimilhança, a menos que as variáveis explanatórias sejam, mutuamente, ortogonais (Pregibon, 1979).

**Exemplo 4.1:** Seja  $y_1, \dots, y_n$  uma amostra de variáveis aleatórias com distribuição normal  $N(\mu_i, \sigma^2)$ , sendo que  $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ . Considerando a função de ligação identidade  $\eta_i = \mu_i$ , tem-se que  $g'(\mu_i) = 1$ . Além disso,  $V_i = 1$  e, portanto,  $w_i = 1$ . Logo, a matriz de informação é igual a

$$\mathbf{K} = \phi^{-1} \mathbf{X}^T \mathbf{W} \mathbf{X} = \sigma^{-2} \mathbf{X}^T \mathbf{X}$$

e a variável dependente ajustada é  $z_i = y_i$ .

Portanto, o algoritmo de estimação (3.5) reduz-se a

$$\mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^T \mathbf{y}$$

e, desde que  $\mathbf{X}^T \mathbf{X}$  tenha inversa,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, \quad (4.6)$$

que é a solução usual de mínimos quadrados para o modelo clássico de regressão. Tem-se, então,

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

e

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

pois  $E[(\mathbf{Y} - \mathbf{X} \boldsymbol{\beta})(\mathbf{Y} - \mathbf{X} \boldsymbol{\beta})^T] = \sigma^2 \mathbf{I}$ .

Como  $\mathbf{Y} \sim N_n(\mathbf{X} \boldsymbol{\beta}, \sigma^2 \mathbf{I})$  e o vetor  $\hat{\boldsymbol{\beta}}$  dos EMVs é uma transformação linear do vetor  $\mathbf{y}$  em (4.6), conclui-se que o vetor  $\hat{\boldsymbol{\beta}}$  tem exatamente uma distribuição normal multivariada  $N_p(\boldsymbol{\beta}, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1})$ . Logo, tem-se, exatamente, que

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{K} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \chi_p^2,$$

sendo  $\mathbf{K} = \sigma^{-2} \mathbf{X}^T \mathbf{X}$  a matriz de informação.

Os erros-padrão dos EMVs  $\hat{\beta}_1, \dots, \hat{\beta}_p$  são iguais às raízes quadradas dos elementos da diagonal de  $\hat{\mathbf{K}}^{-1}$  e podem apresentar informações valiosas sobre a exatidão desses estimadores. Usa-se aqui a notação  $\mathbf{K}^{-1} = \{\kappa^{r,s}\}$  para a inversa da matriz de informação em que, aproximadamente,  $\text{Cov}(\hat{\beta}_r, \hat{\beta}_s) = \kappa^{r,s}$ . Então, com nível de confiança de  $100(1-\alpha)\%$ , intervalos de confiança para os parâmetros  $\beta'_r s$  são determinados por

$$\hat{\beta}_r \mp z_{1-\alpha/2} \sqrt{\hat{\kappa}^{r,r}},$$

em que  $\hat{\kappa}^{r,r} = \widehat{\text{Var}}(\hat{\beta}_r)$  é o valor de  $\kappa^{r,r}$  em  $\hat{\beta}$  e  $z_{1-\alpha/2}$  é quantil  $1 - \alpha/2$  da distribuição normal padrão. Na prática,  $\phi$  é desconhecido, substituindo-se  $\phi$  por uma estimativa consistente  $\hat{\phi}$  e  $z_{1-\alpha/2}$  por  $t_{n-p, 1-\alpha/2}$ , o quantil  $1 - \alpha/2$  da distribuição  $t$  de Student com  $n - p$  graus de liberdade, para o cálculo dos intervalos de confiança para os parâmetros  $\beta_r$ . Nas Seções 4.6 e 4.7, serão apresentados testes e regiões de confiança construídos com base na função desvio.

A correlação estimada  $\hat{\rho}_{rs}$  entre as estimativas  $\hat{\beta}_r$  e  $\hat{\beta}_s$  segue como

$$\hat{\rho}_{rs} = \widehat{\text{Corr}}(\hat{\beta}_r, \hat{\beta}_s) = \frac{\hat{\kappa}^{r,s}}{\sqrt{\hat{\kappa}^{r,r} \hat{\kappa}^{s,s}}},$$

deduzida a partir da inversa da matriz de informação  $\mathbf{K}$  avaliada em  $\hat{\beta}$ . Essas correlações permitem verificar, pelo menos aproximadamente, a interdependência dos  $\hat{\beta}_r$ .

## 4.2 Função desvio e estatística de Pearson generalizada

O ajuste de um modelo a um conjunto de observações  $\mathbf{y}$  pode ser considerado como uma maneira de substituir  $\mathbf{y}$  por um conjunto de valores estimados  $\hat{\boldsymbol{\mu}}$  para um modelo com um número, relativamente pequeno, de parâmetros. Os  $\hat{\boldsymbol{\mu}}$  não serão exatamente iguais aos  $\mathbf{y}$  e deve-se saber em quanto eles diferem. Isso porque uma discrepância pequena pode ser tolerável enquanto que uma discrepância grande, não.

Assim, admitindo-se uma combinação satisfatória da distribuição da variável resposta e da função de ligação, o objetivo é determinar quantos termos são necessários na estrutura linear para uma descrição razoável dos dados. Um número grande de variáveis explanatórias pode conduzir a um modelo que explique bem os dados mas com um aumento de complexidade na interpretação. Por outro lado, um número pequeno de variáveis explanatórias pode implicar em um modelo de interpretação fácil, porém, que se ajuste pobremente aos dados. O que se deseja na realidade é um modelo intermediário, entre um modelo muito complicado e um modelo pobre em ajuste.

A  $n$  observações podem ser ajustados modelos contendo até  $n$  parâmetros. O modelo mais simples é o **modelo nulo** que tem um único parâmetro, representado por um valor  $\mu$  comum a todos os dados. A matriz do modelo, então, reduz-se a um vetor coluna, formado de 1s. Esse modelo atribui toda a variação entre os  $y$  ao componente aleatório. No modelo nulo, o valor comum para todas as médias dos dados é igual à média amostral, isto é,  $\bar{y} = \sum_{i=1}^n y_i/n$ , mas, em geral, não representa a estrutura dos dados. No outro extremo, está o **modelo saturado** ou **completo** que tem  $n$  parâmetros especificados pelas médias  $\mu_1, \dots, \mu_n$  linearmente independentes, ou seja, correspondendo a uma matriz do modelo igual à matriz identidade de ordem  $n$ . O modelo saturado tem  $n$  parâmetros, um para cada observação, e as EMVs das médias são  $\tilde{\mu}_i = y_i$ , para  $i = 1, \dots, n$ . O til é colocado para diferir das EMVs do MLG com matriz do modelo  $\mathbf{X}$ , de dimensões  $n \times p$ , com  $p < n$ . O modelo saturado atribui toda a variação dos dados ao componente sistemático e, assim, ajusta-se perfeitamente, reproduzindo os próprios dados.

Na prática, o modelo nulo é muito simples e o saturado é não-informativo, pois não sumariza os dados, mas, simplesmente, os repete. Existem dois outros modelos, não tão extremos, quanto os modelos nulo e saturado: o **modelo minimal** que contém o menor número de termos necessários para o ajuste e o **modelo maximal** que inclui o maior número de termos que podem ser considerados. Os termos desses modelos são, geralmente, obtidos por interpretações a priori da estrutura dos dados.

Em geral, trabalha-se com modelos encaixados, e o conjunto de matrizes dos modelos pode, então, ser formado pela inclusão sucessiva de termos ao modelo minimal até se chegar ao modelo maximal. Qualquer modelo com  $p$  parâmetros linearmente independentes, situado entre os modelos minimal e maximal, é denominado de **modelo sob pesquisa** ou **modelo corrente**.

Determinados parâmetros têm que estar no modelo como é o caso, por exemplo, de efeitos de blocos em planejamento de experimentos ou de totais marginais fixados em tabelas de contingência para análise de observações na forma de contagens. Assim, considerando-se um experimento casualizado em blocos, com tratamentos no esquema fatorial com dois fatores, têm-se os preditores lineares:

- nulo:  $\eta_i = \alpha_0$
- minimal:  $\eta_i = \alpha_0 + \beta_\ell$
- maximal:  $\eta_i = \alpha_0 + \beta_\ell + \alpha_j + \gamma_k + (\alpha\gamma)_{jk}$
- saturado:  $\eta_i = \alpha_0 + \beta_\ell + \alpha_j + \gamma_k + (\alpha\gamma)_{jk} + (\beta\alpha)_{\ell j} + (\beta\gamma)_{\ell k} + (\beta\alpha\gamma)_{\ell jk}$ ,

sendo  $\alpha_0$  o efeito associado à uma constante;  $\beta_\ell$  o efeito associado ao bloco  $\ell$ ,  $\ell = 1, \dots, b$ ;  $\alpha_j$  o efeito associado ao  $j$ -ésimo nível do fator A;  $\gamma_k$  o efeito associado ao  $k$ -ésimo nível do fator B;  $(\alpha\gamma)_{jk}$ ,  $(\beta\alpha)_{\ell j}$ ,  $(\beta\gamma)_{\ell k}$ ,  $(\beta\alpha\gamma)_{\ell jk}$  os efeitos associados às interações. O modelo saturado inclui, nesse caso, todas as interações com blocos que não são de interesse prático.

### *Função desvio*

O problema principal de seleção de variáveis explanatórias é determinar a utilidade de um parâmetro extra no modelo corrente (sob pesquisa) ou, então, verificar a falta de ajuste induzida pela sua omissão. Com o objetivo de discriminar entre modelos alternativos, medidas de discrepância devem ser introduzidas para avaliar o ajuste de um modelo. Nelder e Wedderburn (1972) propuseram, como medida de discrepância, a “deviance” (traduzida como **desvio** escalonado por Cordeiro (1986)), cuja expressão é

$$S_p = 2(\hat{\ell}_n - \hat{\ell}_p),$$

sendo  $\hat{\ell}_n$  e  $\hat{\ell}_p$  os máximos do logaritmo da função de verossimilhança para os modelos saturado e corrente (sob pesquisa), respectivamente. Verifica-se que o modelo saturado é usado como base de medida do ajuste de um modelo sob pesquisa (modelo corrente). Do logaritmo da função de verossimilhança (3.1), obtêm-se

$$\hat{\ell}_n = \phi^{-1} \sum_{i=1}^n [y_i \tilde{\theta}_i - b(\tilde{\theta}_i)] + \sum_{i=1}^n c(y_i, \phi)$$

e

$$\hat{\ell}_p = \phi^{-1} \sum_{i=1}^n [y_i \hat{\theta}_i - b(\hat{\theta}_i)] + \sum_{i=1}^n c(y_i, \phi),$$

sendo  $\tilde{\theta}_i = q(y_i)$  e  $\hat{\theta}_i = q(\hat{\mu}_i)$  as EMVs do parâmetro canônico sob os modelos saturado e corrente, respectivamente.

Então, tem-se

$$S_p = \phi^{-1} D_p = 2\phi^{-1} \sum_{i=1}^n [y_i(\tilde{\theta}_i - \hat{\theta}_i) + b(\hat{\theta}_i) - b(\tilde{\theta}_i)], \quad (4.7)$$

em que  $S_p$  e  $D_p$  são denominados de desvio escalonado e desvio, respectivamente. O desvio  $D_p$  é função apenas dos dados  $y$  e das médias ajustadas  $\hat{\mu}$ . O desvio escalonado  $S_p$  depende de  $D_p$  e do parâmetro de dispersão  $\phi$ . Pode-se, ainda, escrever

$$S_p = \phi^{-1} \sum_{i=1}^n d_i^2,$$

sendo que  $d_i^2$  mede a diferença dos logaritmos das funções de verossimilhança observada e ajustada, para a observação  $i$  correspondente, e é denominado **componente do desvio**. A soma deles mede a discrepância total entre os dois modelos na escala logarítmica da verossimilhança. É, portanto, uma medida da distância dos valores ajustados  $\hat{\mu}$  em relação às observações  $y$ , ou de forma equivalente, do modelo corrente em relação ao modelo saturado. Verifica-se que o desvio equivale a uma constante menos duas vezes o máximo do logaritmo da função de verossimilhança para o modelo corrente, isto é,

$$S_p = 2\hat{\ell}_n - 2\hat{\ell}_p = \text{constante} - 2\hat{\ell}_p.$$

O desvio é computado facilmente para qualquer MLG a partir da EMV  $\hat{\boldsymbol{\mu}} = g^{-1}(\mathbf{X}\hat{\boldsymbol{\beta}})$  de  $\boldsymbol{\mu}$ . O desvio é sempre maior do que ou igual a zero, e à medida que variáveis explanatórias entram no componente sistemático, o desvio decresce até se tornar zero para o modelo saturado. Para o teste, define-se o número de graus de liberdade do desvio do modelo por  $\nu = n - p$ , isto é, como o número de observações menos o posto da matriz do modelo sob pesquisa. Em alguns casos especiais, como nos modelos normal e log-linear, o desvio iguala-se às estatísticas comumente usadas nos testes de ajuste.

**Exemplo 4.2:** Seja  $y_1, \dots, y_n$  uma amostra de variáveis aleatórias com distribuição normal  $N(\mu_i, \sigma^2)$ , sendo que  $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ . Tem-se  $\phi = \sigma^2$ ,  $\theta_i = \mu_i$  e  $b(\theta_i) = \frac{\theta_i^2}{2}$ . Logo,

$$\begin{aligned} S_p &= \frac{1}{\sigma^2} \sum_{i=1}^n 2 \left[ y_i(y_i - \hat{\mu}_i) - \frac{y_i^2}{2} + \frac{\hat{\mu}_i^2}{2} \right] = \frac{1}{\sigma^2} \sum_{i=1}^n (2y_i^2 - 2\hat{\mu}_i y_i - y_i^2 + \hat{\mu}_i^2) \\ &= \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 = \frac{SQRes}{\sigma^2}, \end{aligned}$$

que coincide com a estatística clássica  $SQRes = \sum_i (y_i - \hat{\mu}_i)^2$  com  $(n - p)$  graus de liberdade dividida por  $\sigma^2$ .

**Exemplo 4.3:** Sejam  $Y_1, \dots, Y_n$  variáveis aleatórias representando contagens de sucessos em amostras independentes de tamanhos  $m_i$ . Suponha que  $Y_i \sim B(m_i, \pi_i)$ ,  $\phi = 1$ ,  $\theta_i = \log\left(\frac{\mu_i}{m_i - \mu_i}\right)$  e  $b(\theta_i) = m_i \log(1 + e^{\theta_i}) = -m_i \log\left(\frac{m_i - \mu_i}{m_i}\right)$ .

Logo,

$$\begin{aligned} S_p &= 2 \sum_{i=1}^n \left\{ y_i \left[ \log\left(\frac{y_i}{m_i - y_i}\right) - \log\left(\frac{\hat{\mu}_i}{m_i - \hat{\mu}_i}\right) \right] \right\} \\ &\quad + 2 \sum_{i=1}^n \left\{ m_i \log\left(\frac{m_i - y_i}{m_i}\right) - m_i \log\left(\frac{m_i - \hat{\mu}_i}{m_i}\right) \right\} \end{aligned}$$

ou ainda,

$$S_p = 2 \sum_{i=1}^n \left[ y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) + (m_i - y_i) \log\left(\frac{m_i - y_i}{m_i - \hat{\mu}_i}\right) \right].$$

Essa expressão é válida para  $0 < y_i < m_i$ . Se  $y_i = 0$  ou  $y_i = m_i$ , o  $i$ -ésimo termo de  $S_p$  deve ser substituído por  $2m_i \log[m_i/(m_i - \hat{\mu}_i)]$  ou  $2m_i \log(m_i/\hat{\mu}_i)$ , respectivamente (Paula, 2004). Se  $m_i = 1$ , isto é,  $Y_i \sim \text{Bernoulli}(\pi_i)$  e a função de ligação considerada é a logística, a função desvio é apenas uma função das observações e, portanto, não é informativa com relação ao ajuste do modelo aos dados. O mesmo é válido para as funções de ligação probito e complemento log-log.

Para o modelo de Poisson, o desvio tem a forma

$$S_p = 2 \left[ \sum_{i=1}^n y_i \log \left( \frac{y_i}{\hat{\mu}_i} \right) + \sum_{i=1}^n (\hat{\mu}_i - y_i) \right]$$

e, em particular, para os modelos log-lineares a segunda soma é igual a zero, desde que a matriz  $\mathbf{X}$  tenha uma coluna de 1s (Exercício 5 da Seção 4.9). Nesse caso, o desvio é igual à razão de verossimilhanças (denotada por  $G^2$  ou  $Y^2$ ), que é, geralmente, usada nos testes de hipóteses em tabelas de contingência.

Para o modelo gama ( $\theta = -\mu^{-1}$ ) com média  $\mu$  e parâmetro de dispersão  $\phi$  ( $= \text{Var}(Y)/E(Y)^2$ ), a expressão do desvio é

$$S_p = 2\phi^{-1} \sum_{i=1}^n \left[ \log \left( \frac{\hat{\mu}_i}{y_i} \right) + \frac{(y_i - \hat{\mu}_i)}{\hat{\mu}_i} \right],$$

que pode ainda ser simplificada em alguns casos especiais (Exercício 6 da Seção 4.9). Se algum componente do desvio é igual a zero, segundo Paula (2004), pode-se substituir  $D_p$  por

$$D_p = 2c(y) + 2 \sum_{i=1}^n \left[ \log(\hat{\mu}_i) + \frac{y_i}{\hat{\mu}_i} \right],$$

sendo  $c(y)$  uma função arbitrária, porém limitada. Pode ser usada, por exemplo, a expressão  $c(y) = \sum_{i=1}^n \frac{y_i}{1 + y_i}$ . Na Tabela 4.1, apresentam-se os desvios residuais para os principais modelos.

Quanto melhor for o ajuste do MLG aos dados tanto menor será o valor do desvio  $D_p$ . Assim, um modelo bem ajustado aos dados terá uma métrica  $\|\mathbf{y} - \hat{\boldsymbol{\mu}}\|$  pequena, sendo essa métrica definida na escala do logaritmo da função de verossimi-



Tabela 4.1: Desvios residuais para alguns modelos.

Modelo	Desvio
Normal	$D_p = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2$
Binomial	$D_p = 2 \sum_{i=1}^n \left[ y_i \log \left( \frac{y_i}{\hat{\mu}_i} \right) + (m_i - y_i) \log \left( \frac{m_i - y_i}{m_i - \hat{\mu}_i} \right) \right]$
Poisson	$D_p = 2 \sum_{i=1}^n \left[ y_i \log \left( \frac{y_i}{\hat{\mu}_i} \right) + (\hat{\mu}_i - y_i) \right]$
Binomial negativo	$D_p = 2 \sum_{i=1}^n \left[ y_i \log \left( \frac{y_i}{\hat{\mu}_i} \right) + (y_i + k) \log \left( \frac{\hat{\mu}_i + k}{y_i + k} \right) \right]$
Gama	$D_p = 2 \sum_{i=1}^n \left[ \log \left( \frac{\hat{\mu}_i}{y_i} \right) + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \right]$
Normal inverso	$D_p = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{y_i \hat{\mu}_i^2}$

lhança e, portanto, um modelo bem (mal) ajustado aos dados, com uma verossimilhança máxima grande (pequena), tem um pequeno (grande) desvio.

Uma maneira de se obter uma redução do desvio é aumentar o número de parâmetros no preditor linear, o que, porém, significa um aumento do grau de complexidade na interpretação do modelo. Na prática, procuram-se modelos simples com desvios moderados, situados entre os modelos mais complicados e os que se ajustam mal às observações. Para testar a adequação de um MLG, o valor calculado do desvio com  $n - p$  graus de liberdade, sendo  $p$  o posto da matriz do modelo, deve ser comparado com o percentil de alguma distribuição de referência. Para o modelo normal com função de ligação identidade, assumindo-se que o modelo usado é verdadeiro e que  $\sigma^2$  é conhecido, tem-se o resultado exato

$$S_p = \frac{D_p}{\sigma^2} \sim \chi_{n-p}^2.$$

Entretanto, para modelos normais com outras funções de ligação, esse resul-

tado é apenas uma aproximação. Em alguns casos especiais da matriz modelo, com delineamentos experimentais simples, considerando-se as distribuições exponencial (caso especial da gama) e normal inversa, também, podem ser obtidos resultados exatos. No geral, porém, apenas alguns resultados assintóticos estão disponíveis e, em alguns casos, o desvio não tem distribuição  $\chi^2_{n-p}$ , nem mesmo assintoticamente. O desvio modificado por uma correção de Bartlett proposta para os MLGs por Cordeiro (1983, 1987, 1995) tem sido usado para melhorar a sua aproximação pela distribuição  $\chi^2_{n-p}$  de referência. Com efeito, o desvio modificado  $S_p = (n-p)S_p/\widehat{E}(S_p)$ , em que a correção de Bartlett é expressa por  $(n-p)/\widehat{E}(S_p)$  quando  $E(S_p)$  é determinada até termos de ordem  $O(n^{-1})$ , sendo  $\widehat{E}(S_p)$  o valor de  $E(S_p)$  avaliado em  $\hat{\mu}$ , é melhor aproximado pela distribuição  $\chi^2_{n-p}$  de referência do que o próprio desvio  $S_p$ , conforme comprovam os estudos de simulação de Cordeiro (1993).

Considerando-se que o modelo corrente é verdadeiro, para o modelo binomial, quando  $n$  é fixo e  $m_i \rightarrow \infty$ ,  $\forall i$  (não vale quando  $m_i\pi_i(1-\pi_i)$  permanece limitado) e para o modelo de Poisson, quando  $\mu_i \rightarrow \infty$ ,  $\forall i$ , tem-se que (lembre-se que  $\phi = 1$ ):  $S_p = D_p$  é, aproximadamente, distribuído como  $\chi^2_{n-p}$ .

Nos modelos em que  $S_p$  depende do parâmetro de dispersão  $\phi$ , conhecido, Jørgensen (1987) mostra que  $S_p \xrightarrow{D} \chi^2_{n-p}$  quando  $\phi \rightarrow 0$ , isto é, quando o parâmetro de dispersão é pequeno, a aproximação  $\chi^2_{n-p}$  para  $S_p$  é satisfatória. Para o modelo gama, a aproximação da distribuição de  $S_p$  por  $\chi^2_{n-p}$  será tanto melhor quanto mais próximo de um estiver o parâmetro de dispersão. Em geral, porém, não se conhece  $\phi$ , que precisa ser substituído por uma estimativa consistente (Seções 4.4 e 4.5).

Na prática, para  $\phi$  conhecido, contenta-se em testar um MLG comparando-se o valor de  $S_p$  com os percentis da distribuição  $\chi^2_{n-p}$ . Assim, quando

$$S_p = \phi^{-1}D_p \leq \chi^2_{n-p;1-\alpha},$$

ou seja,  $S_p$  é inferior ao valor crítico  $\chi^2_{n-p;1-\alpha}$  da distribuição  $\chi^2_{n-p}$ , pode-se considerar que existem evidências, a um nível aproximado de  $100\alpha\%$  de significância, que o modelo proposto está bem ajustado aos dados, ou ainda, se o valor de  $D_p$  for próximo

do valor esperado  $n - p$  de uma distribuição  $\chi^2_{n-p}$ , pode ser um indicativo de que o modelo ajustado aos dados é adequado.

O desvio  $D_p$  pode ser adotado como um critério de parada do algoritmo de ajuste descrito em (3.5) e, após a convergência, o seu valor com o correspondente número de graus de liberdade podem ser computados.

### *Estatística de Pearson generalizada $X^2$*

Uma outra medida da discrepância do ajuste de um modelo a um conjunto de dados é a estatística de Pearson generalizada  $X_p^2$  cuja expressão é

$$X_p^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}, \quad (4.8)$$

sendo  $V(\hat{\mu}_i)$  a função de variância estimada sob o modelo que está sendo ajustado aos dados. A estatística  $X_p^2$  tem como desvantagem o fato de tratar as observações simetricamente. A fórmula (4.8) da estatística de Pearson generalizada tem uma forma equivalente expressa em termos da variável dependente ajustada do algoritmo (3.5), isto é,

$$X_p^2 = (\mathbf{z} - \hat{\boldsymbol{\eta}})^T \widehat{\mathbf{W}} (\mathbf{z} - \hat{\boldsymbol{\eta}}). \quad (4.9)$$

que para o modelo canônico reduz-se a,

$$X_p^2 = (\mathbf{z} - \hat{\boldsymbol{\eta}})^T \mathbf{V} (\mathbf{z} - \hat{\boldsymbol{\eta}}).$$

Para respostas com distribuição normal,  $X_p^2 = D_p$  e  $\sigma^{-2} X_p^2 = \sigma^{-2} SQRes$  e, então,

$$X_p^2 \sim \sigma^2 \chi^2_{n-p},$$

sendo esse resultado exato somente se a função de ligação for a identidade e  $\sigma^2$  conhecido.

Para dados provenientes das distribuições binomial e de Poisson, em que  $\phi = 1$ ,  $X_p^2$  é a estatística original de Pearson, comumente usada na análise dos

modelos logístico e log-linear em tabelas multidimensionais, e que tem a forma

$$X_p^2 = \sum_{i=1}^n \frac{(o_i - e_i)^2}{e_i},$$

sendo  $o_i$  a frequência observada e  $e_i$  a frequência esperada.

Segundo Jørgensen (2013), o caso do parâmetro  $\phi$  conhecido é relevante apenas para dados discretos ( $\phi = 1$  para as distribuições binomial e de Poisson) e recomenda o uso de  $X_p^2$  ao invés de  $D_p$  como uma medida de qualidade de ajuste, baseado em estudos numéricos e analíticos que mostram que a distribuição  $\chi_{n-p}^2$  limitante é aproximada mais rapidamente (com o aumento do tamanho amostral) para  $X_p^2$  do que para  $D_p$ .

Para as distribuições contínuas não-normais, têm-se apenas resultados assintóticos para  $X_p^2/\phi$ , isto é, a distribuição  $\chi_{n-p}^2$  pode ser usada, somente, como uma aproximação para a distribuição de  $X_p^2/\phi$ , que em muitos casos pode ser inadequada. Quando o parâmetro  $\phi$  é desconhecido, como é o caso das distribuições contínuas, não há teste formal baseado em  $X_p^2$  ou  $D_p$  (Jørgensen, 2013), sendo  $X_p^2$  usado para estimar o parâmetro de dispersão  $\phi$ .

O desvio  $S_p$  tem a grande vantagem como medida de discrepância por ser aditivo para um conjunto de modelos encaixados, enquanto que  $X_p^2$ , em geral, não tem essa propriedade, apesar de ser preferido em relação ao desvio, em muitos casos, pela facilidade de interpretação.

**Exemplo 4.4:** Considere os dados do Exemplo 2.1 da Seção 2.2. A variável resposta tem distribuição binomial, isto é,  $Y_i \sim B(m_i, \pi_i)$ . Adotando-se a função de ligação logística (canônica) e o preditor linear como uma regressão linear simples, isto é,

$$\eta_i = \log \left( \frac{\mu_i}{m_i - \mu_i} \right) = \beta_0 + \beta_1 d_i,$$

tem-se  $S_p = D_p = 10,26$  e  $X_p^2 = 9,70$  com 4 graus de liberdade. Da tabela da distribuição  $\chi^2$ , tem-se  $\chi_{4;0,05}^2 = 9,49$  e  $\chi_{4;0,01}^2 = 13,29$ , indicando que existem evidências, a um nível de significância entre 0,05 e 0,01, que o modelo logístico linear ajusta-se,

razoavelmente, a esse conjunto de dados. Necessita-se, porém, adicionalmente, do teste da hipótese  $H_0 : \beta_1 = 0$ , de uma análise de resíduos e de medidas de diagnóstico.

### 4.3 Análise de desvio e seleção de modelos

A análise de desvio (“Analysis of Deviance” - ANODEV) é uma generalização da análise de variância para os MLGs, visando obter, a partir de uma sequência de modelos encaixados, cada modelo incluindo mais termos do que os anteriores, os efeitos de variáveis explanatórias, fatores e suas interações. Usa-se o desvio como uma medida de discrepância do modelo e forma-se uma tabela de diferenças de desvios.

Seja  $M_{p_1}, M_{p_2}, \dots, M_{p_r}$  uma sequência de modelos encaixados de dimensões respectivas  $p_1 < p_2 < \dots < p_r$ , matrizes dos modelos  $\mathbf{X}_{p_1}, \mathbf{X}_{p_2}, \dots, \mathbf{X}_{p_r}$  e desvios  $D_{p_1} > D_{p_2} > \dots > D_{p_r}$ , tendo os modelos a mesma distribuição e a mesma função de ligação. Essas desigualdades entre os desvios, em geral, não se verificam para a estatística de Pearson  $X_p^2$  generalizada e, por essa razão, a comparação de modelos encaixados é feita, principalmente, baseada na função desvio. Assim, para o caso de um ensaio inteiramente casualizado, com  $r$  repetições e tratamentos no esquema fatorial, com  $a$  níveis para o fator  $A$  e  $b$  níveis para o fator  $B$ , obtêm-se os resultados mostrados na Tabela 4.2.

Dois termos  $A$  e  $B$  são ortogonais se a redução que  $A$  (ou  $B$ ) causa no desvio  $D_p$  é a mesma, esteja  $B$  (ou  $A$ ) incluído, ou não, em  $M_p$ . Em geral, para os MLGs ocorre a não-ortogonalidade dos termos e a interpretação da tabela ANODEV é mais complicada do que a ANOVA usual.

Sejam os modelos encaixados  $M_q$  e  $M_p$  ( $M_q \subset M_p$ ,  $q < p$ ), com  $q$  e  $p$  parâmetros, respectivamente. A estatística  $D_q - D_p$  com  $(p - q)$  graus de liberdade é interpretada como uma medida de variação dos dados, explicada pelos termos que estão em  $M_p$  e não estão em  $M_q$ , incluídos os efeitos dos termos em  $M_q$  e ignorando quaisquer efeitos dos termos que não estão em  $M_p$ . Tem-se, assintoticamente, para

Tabela 4.2: Um exemplo de construção de uma tabela de Análise de Desvio.

Preditor linear	gl	desvio	Dif. de desvios	Dif. de gl	Significado
Nulo	$rab - 1$	$D_1$			
			$D_1 - D_A$	$a - 1$	$A$ ignorando $B$
A	$a(rb - 1)$	$D_A$			
			$D_A - D_{A+B}$	$b - 1$	$B$ incluído $A$
A+B	$a(rb - 1) - (b - 1)$	$D_{A+B}$			
			$D_{A+B} - D_{A*B}$	$(a - 1)(b - 1)$	Interação $AB$ incluídos $A$ e $B$
A+B+A.B	$ab(r - 1)$	$D_{A*B}$			
			$D_{A*B}$	$ab(r - 1)$	Resíduo
Saturado	0	0			

$\phi$  conhecido

$$S_q - S_p = \phi^{-1}(D_q - D_p) \sim \chi_{p-q}^2,$$

em que  $S_q - S_p$  é igual à estatística da razão de verossimilhanças (Seção 4.6).

Nesses termos, quando o modelo com menor número de parâmetros ( $q$ ) é verdadeiro,  $S_q - S_p$  tem distribuição assintótica  $\chi_{p-q}^2$ . Entretanto, cada desvio isolado não é distribuído, assintoticamente, como qui-quadrado. O teorema de Wilks (1937) requer que os espaços de parâmetros, segundo os modelos sob as hipóteses nula e alternativa, sejam de dimensão fixa, enquanto  $n$  cresce e, portanto, não se aplica ao desvio isolado, cujo modelo alternativo é o saturado de dimensão  $n$ .

Se  $\phi$  é desconhecido, deve-se obter uma estimativa  $\hat{\phi}$  consistente, de preferência baseada no modelo maximal (com  $m$  parâmetros), e a inferência pode ser baseada na estatística  $F$  expressa por

$$F = \frac{(D_q - D_p)/(p - q)}{\hat{\phi}} \sim F_{p-q, n-m}.$$

Para modelo normal linear, tem-se que

$$\frac{(SQRes_q - SQRes_p)/(p - q)}{SQRes_m/(n - m)} \sim F_{p-q, n-m},$$

sendo a distribuição F exata.

**Exemplo 4.5: Toxicidade de Rotenone (cont.)**

Considere os dados do Exemplo 2.1 da Seção 2.2. A variável resposta tem distribuição binomial, isto é,  $Y_i \sim B(m_i, \pi_i)$ . Adotando-se a função de ligação logística (canônica) e o preditor linear expresso como uma regressão linear simples, isto é,

$$\eta_i = \log \left( \frac{\mu_i}{m_i - \mu_i} \right) = \beta_0 + \beta_1 d_i,$$

dois modelos encaixados podem ser propostos para a análise desses dados, a saber:

- a) o modelo nulo:  $\eta_i = \beta_0$  e
- b) o modelo de regressão linear:  $\eta_i = \beta_0 + \beta_1 d_i$ .

Tabela 4.3: Desvios e  $X^2$  residuais obtidos para dois modelos encaixados ajustados aos dados da Tabela 2.1.

Preditor linear	g.l.	Desvios	$X^2$
$\eta_i = \beta_0$	5	163,74	135,70
$\eta_i = \beta_0 + \beta_1 d_i$	4	10,26	9,70
$\chi^2_{4;0,05} = 9,49, \quad \chi^2_{4;0,01} = 13,29$			

A Tabela 4.3 apresenta os desvios e os valores da estatística de Pearson generalizada e seus respectivos números de graus de liberdade (g.l.), e a Tabela 4.4, a análise do desvio correspondente.

O exame da Tabela 4.3, confirmando o que foi descrito no Exemplo 4.4, mostra que existem evidências, a um nível de significância entre 0,05 e 0,01, que o modelo logístico linear ajusta-se razoavelmente a esse conjunto de dados, mas rejeita-se o modelo nulo. Pela Tabela 4.4, rejeita-se a hipótese nula  $H_0 : \beta_1 = 0$ , confirmando a adequação do modelo logístico linear. Necessita-se, porém, adicionalmente, de uma análise de resíduos e de diagnóstico.

Tabela 4.4: Análise do Desvio, considerando o modelo logístico linear ajustado aos dados da Tabela 2.1.

Causas de Variação	g.l.	Desvios	Valor $p$
Regressão linear	1	153,48	$< 0,0001$
Resíduo	4	10,26	
Total	5	163,74	
$\chi^2_{1;0,05} = 3,84; \chi^2_{1;0,01} = 6,64$			

Tem-se, ainda, que  $\hat{\beta}_0 = -3,226$  [ $s(\hat{\beta}_0) = 0,3699$ ] e  $\hat{\beta}_1 = 0,6051$  [ $s(\hat{\beta}_1) = 0,0678$ ]. O número esperado de insetos mortos  $\hat{\mu}_i$  para a dose  $d_i$  é expresso por

$$\hat{\mu}_i = m_i \frac{\exp(-3,226 + 0,6051d_i)}{1 + \exp(-3,226 + 0,6051d_i)}.$$

Na Figura 4.1 estão representados a curva do modelo ajustado e os valores observados. Um programa simples em linguagem R (Team, 2015) para obter esses resultados é apresentado no Apêndice B.

## 4.4 Estimação do parâmetro de dispersão

Para as distribuições binomial e Poisson tem-se que o parâmetro de dispersão  $\phi = 1$ . Quando  $\phi$  é desconhecido (distribuições normal, normal inversa e gama), considera-se que seja o mesmo para todas as observações, isto é, constante. Torna-se necessário estimá-lo para obter (conforme descrito na Seção 3.4) os erros-padrão dos  $\hat{\beta}s$ , intervalos de confiança e testes de hipóteses para os  $\beta s$ , etc. Os métodos mais usados para a estimação de  $\phi$  são: método do desvio, método de Pearson e método de máxima verossimilhança.

O método do desvio é baseado na aproximação  $\chi^2_{n-p}$  para o desvio escalonado (4.7). Para um modelo bem ajustado às observações, espera-se, portanto, que o desvio escalonado  $S_p$  tenha valor esperado igual a  $n - p$ . Assim, obtém-se a



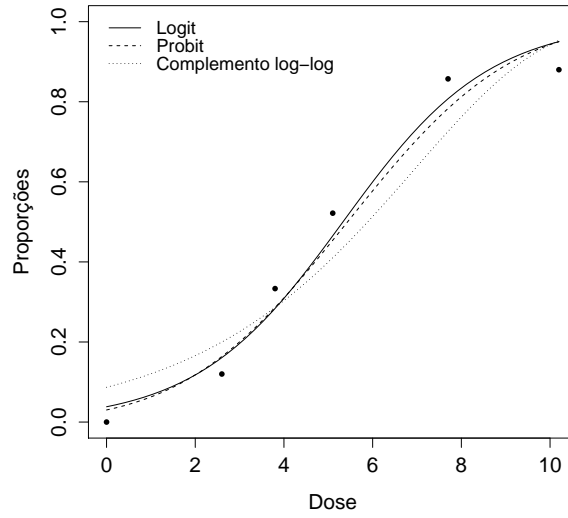


Figura 4.1: Valores observados e curvas ajustadas pelos modelos logístico, probito e complemento log-log lineares aos dados da Tabela 2.1.

estimativa do parâmetro  $\phi$

$$\hat{\phi}_d = \frac{D_p}{n - p}, \quad (4.10)$$

em que o desvio  $D_p$  é calculado de (4.7) como função das observações  $\mathbf{y}$  e dos valores ajustados  $\hat{\boldsymbol{\mu}}$ . O estimador  $\hat{\phi}_d$  é, aproximadamente, não viesado para os modelos normal e normal inverso. Para o modelo normal linear,  $\hat{\phi}_d = \sum (y_i - \hat{\mu}_i)^2 / (n - p)$  é o estimador usual não-viesado de  $\sigma^2$ . Para os modelos gama e normal inverso, as expressões correspondentes dos desvios  $D_p$  estão na Tabela 4.1, possibilitando calcular  $\hat{\phi}_d$  de (4.10).

O método de Pearson é baseado na aproximação da distribuição da estatística de Pearson  $X_p^2$  generalizada (4.8), dividida por  $\phi$ , pela distribuição  $\chi_{n-p}^2$ . Obtém-se, assim, a estimativa de Pearson de  $\phi$

$$\hat{\phi}_P = \frac{1}{n - p} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}. \quad (4.11)$$

Para o modelo normal,  $\hat{\phi}_d = \hat{\phi}_P$ . Para os demais modelos contínuos, esses

estimadores diferem em valor. Os estimadores  $\hat{\phi}_P$  para os modelos gama e normal inverso são deduzidos de (4.11) fazendo-se  $V(\mu) = \mu^2$  e  $V(\mu) = \mu^3$ , respectivamente.

O método de máxima verossimilhança é sempre possível em teoria, mas pode tornar-se complicado computacionalmente quando não existir solução explícita para a EMV. Se  $\phi$  é o mesmo para todas as observações, a EMV de  $\beta$  independe de  $\phi$ . Entretanto, a matriz de variâncias e covariâncias dos  $\hat{\beta}$ s envolve esse parâmetro. Interpretando o logaritmo da função de verossimilhança  $\ell(\beta, \phi)$  como função de  $\beta$  e de  $\phi$ , supondo conhecido  $\mathbf{y}$ , pode-se escrever da equação (3.1)

$$\ell(\beta, \phi) = \phi^{-1} \sum_{i=1}^n [y_i \theta_i - b(\theta_i)] + \sum_{i=1}^n c(y_i, \phi). \quad (4.12)$$

A função escore relativa ao parâmetro  $\phi$  é expressa por

$$U_\phi = \frac{\partial \ell(\beta, \phi)}{\partial \phi} = -\phi^{-2} \sum_{i=1}^n [y_i \theta_i - b(\theta_i)] + \sum_{i=1}^n \frac{dc(y_i, \phi)}{d\phi}.$$

Observa-se que  $U_\phi$  é função de  $\beta$  por meio de  $\theta$  (ou  $\mu$ ) e de  $\phi$ , supondo  $\mathbf{y}$  conhecido. A EMV  $\hat{\phi}$  de  $\phi$  é calculada igualando-se  $\partial \ell(\hat{\beta}, \phi) / \partial \phi$  a zero. Claro que a EMV  $\hat{\phi}$  é função das médias ajustadas  $\hat{\mu}$  e dos dados  $\mathbf{y}$ . Da forma da função  $c(y, \phi)$  especificada na Seção 1.3 (Tabela 1.1), verifica-se facilmente que  $\hat{\phi} = D_p/n$  para os modelos normal e normal inverso. Para o modelo gama, obtém-se a EMV  $\hat{\phi}$  de  $\phi$  como solução da equação não-linear

$$\log(\hat{\phi}^{-1}) - \psi(\hat{\phi}^{-1}) = \frac{D_p}{2n}, \quad (4.13)$$

em que o desvio  $D_p$  é apresentado na Tabela 4.1 e  $\psi(r) = d \log \Gamma(r) / dr$  é a função digama (função psi). Uma aproximação para  $\hat{\phi}$  obtida de (4.13) foi deduzida por Cordeiro e McCullagh (1991) para valores pequenos de  $\phi$

$$\hat{\phi} \approx \frac{2D_p}{n \left[ 1 + \left( 1 + \frac{2D_p}{3n} \right)^{1/2} \right]}.$$

Derivando-se  $U_\phi$  em relação a  $\beta_r$ , tem-se

$$U_{\phi r} = \frac{\partial^2 \ell(\beta, \phi)}{\partial \phi \partial \beta_r} = -\phi^{-2} \sum_{i=1}^n \frac{(y_i - \mu_i)}{V_i} \frac{d\mu_i}{d\eta_i} x_{ir}.$$

Logo,  $E(U_{\phi r}) = 0$ , o que mostra que os parâmetros  $\phi$  e  $\beta$  são ortogonais. Esse fato implica que os EMV de  $\beta$  e  $\phi$  são, assintoticamente, independentes.

Como  $U_{\phi}$  é função de  $\phi$  e  $\mu$ , escreve-se  $U_{\phi} = U_{\phi}(\phi, \mu)$ . Pode-se mostrar que  $2U_{\phi}(\hat{\phi}, \mathbf{y}) = D_p$ , isto é, duas vezes a função escore relativa a  $\phi$  avaliada no ponto  $(\hat{\phi}, \mathbf{y})$  é igual ao desvio do modelo.

## 4.5 Comparação dos três métodos de estimação do parâmetro de dispersão no modelo gama

Nesta seção, comparam-se as três estimativas  $\hat{\phi}_d$ ,  $\hat{\phi}_P$  e  $\hat{\phi}$  de  $\phi$  no modelo gama. Cordeiro e McCullagh (1991) usaram a desigualdade  $\frac{1}{2x} < \log x - \psi(x) < \frac{1}{x}$ , em que  $\psi(\cdot)$  é a função digama, para mostrar que

$$\frac{D_p}{2n} < \hat{\phi} < \frac{D_p}{n}$$

e, portanto,

$$\frac{\hat{\phi}_d(n-p)}{2n} < \hat{\phi} < \frac{\hat{\phi}_d(n-p)}{n}.$$

Logo, para  $n$  grande, a EMV de  $\phi$  deve ficar entre  $\hat{\phi}_d/2$  e  $\hat{\phi}_d$ , ou seja, será menor do que  $\hat{\phi}_d$ .

Para comparar  $\hat{\phi}_d$  e  $\hat{\phi}_P$ , admite-se que a matriz modelo  $\mathbf{X}$  tenha uma coluna de uns relativa ao intercepto. Nesse caso, o desvio  $D_p$  reduz-se a  $D_p = 2 \sum_{i=1}^n \log(\hat{\mu}_i/y_i)$ , pois  $\sum_{i=1}^n (y_i - \hat{\mu}_i)/\hat{\mu}_i = 0$ . Considere a expansão em série de Taylor

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \dots$$

e a função  $f(y_i) = \log(\hat{\mu}_i/y_i)$  com  $x = y_i$  e  $a = \hat{\mu}_i$ . Então,  $f'(y_i) = -y_i^{-1}$ ,  $f''(y_i) = y_i^{-2}$  e  $f'''(y_i) = -2y_i^{-3}$  e

$$f(y_i) = \log\left(\frac{\hat{\mu}_i}{y_i}\right) \approx -\frac{(y_i - \hat{\mu}_i)}{\hat{\mu}_i} + \frac{(y_i - \hat{\mu}_i)^2}{2\hat{\mu}_i^2} - \frac{(y_i - \hat{\mu}_i)^3}{3\hat{\mu}_i^3}.$$

Logo,

$$D_p = 2 \sum_{i=1}^n \log \left( \frac{\hat{\mu}_i}{y_i} \right) \approx -2 \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)}{\hat{\mu}_i} + \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i^2} - \frac{2}{3} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^3}{\hat{\mu}_i^3}. \quad (4.14)$$

O primeiro termo dessa expansão é nulo, pois o MLG tem por hipótese uma coluna de uns. Dividindo a equação (4.14) por  $n - p$  e usando (4.10) e (4.11), tem-se

$$\hat{\phi}_d \approx \hat{\phi}_P - \frac{2}{3(n-p)} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^3}{\hat{\mu}_i^3}.$$

Como a última soma pode ser positiva ou negativa, conclui-se que  $\hat{\phi}_d$  pode ser maior do que, menor do que ou igual a  $\hat{\phi}_P$ . Se o MLG tiver um bom ajuste, as médias ajustadas e as observações serão próximas e, assim,  $\hat{\phi}_d \doteq \hat{\phi}_P$ .

## 4.6 Testes de hipóteses

Os métodos de inferência nos MLGs baseiam-se, fundamentalmente, na teoria de máxima verossimilhança. De acordo com essa teoria, três estatísticas são, usualmente, usadas para testar hipóteses relativas aos parâmetros  $\beta s$ , sendo deduzidas de distribuições assintóticas de funções adequadas dos EMVs dos  $\beta' s$ . Elas são: razão de verossimilhanças, Wald e escore, que são assintoticamente equivalentes.

Sob a hipótese nula e supondo que o parâmetro de dispersão  $\phi$  é conhecido, as três estatísticas convergem para uma variável aleatória com distribuição  $\chi_p^2$ , sendo, porém, a razão de verossimilhanças o critério que define um teste uniformemente mais poderoso. Um estudo comparativo dessas estatísticas pode ser encontrado em Buse (1982) para o caso de hipóteses simples. Dentre outras, referências importantes são Silvey (1975), Cordeiro (1986), ?, McCulloch e Searle (2000) e Paula (2004).

A razão de verossimilhanças para testar componentes do vetor  $\beta$  pode ser obtida como uma diferença de desvios entre modelos encaixados. A estatística de Wald (1943), também denominada de “máxima verossimilhança” por alguns autores, é baseada na distribuição normal assintótica de  $\hat{\beta}$ . A estatística escore (Rao, 1973, Seção 6e) é obtida da função escore introduzida na Seção 3.2.

Dependendo da hipótese a ser testada, em particular, qualquer uma dessas três estatísticas pode ser a mais apropriada. Para hipóteses relativas a um único coeficiente  $\beta_r$ , a estatística de Wald é a mais usada. Para hipóteses relativas a vários coeficientes, a razão de verossimilhanças é, geralmente, preferida. A estatística escore tem sido usada na Bioestatística, com a finalidade de realizar testes como os do tipo de Mantel e Haenszel (1959).

#### 4.6.1 Teste de uma hipótese nula simples

Considere o teste da hipótese nula simples  $H_0 : \beta = \beta_0$  em um MLG supondo  $\phi$  conhecido, em que  $\beta_0$  é um vetor especificado para o vetor  $\beta$  de parâmetros desconhecidos, versus a hipótese alternativa  $H : \beta \neq \beta_0$ . Esse teste não é muito usado, pois, na prática, o interesse é especificar um subconjunto de componentes de  $\beta$ . As três estatísticas para testar  $H_0$  têm as seguintes formas

- razão de verossimilhanças:  $w = 2[\ell(\hat{\beta}) - \ell(\beta_0)]$ ,
- estatística de Wald:  $W = (\hat{\beta} - \beta_0)^T \hat{\mathbf{K}} (\hat{\beta} - \beta_0)$ , e
- estatística escore:  $S_R = \mathbf{U}(\beta_0)^T \mathbf{K}_0^{-1} \mathbf{U}(\beta_0)$ ,

em que  $\ell(\hat{\beta})$  e  $\ell(\beta_0)$  são os valores do logaritmo da função de verossimilhança (3.1) em  $\hat{\beta}$  e  $\beta_0$ , respectivamente,  $\mathbf{U}(\beta_0)$  e  $\mathbf{K}_0$  são o vetor escore e a matriz de informação avaliadas em  $\beta_0$ , e  $\hat{\mathbf{K}}$  a matriz de informação avaliada na EMV  $\hat{\beta}$ . Na estatística de Wald,  $\hat{\mathbf{K}}$  pode ser substituída por  $\mathbf{K}_0$  para definir uma estatística de Wald modificada assintoticamente equivalente. Uma vantagem da estatística escore é que não é necessário calcular a EMV de  $\beta$  segundo hipótese alternativa  $H$ , embora na prática essa estatística seja importante.

As três estatísticas descritas são, assintoticamente, equivalentes e, segundo a hipótese nula  $H_0$ , convergem em distribuição para a variável  $\chi_p^2$ . Se o modelo tem um único parâmetro, usando-se as estatísticas  $\sqrt{S_R}$  e  $\sqrt{W}$ , com um sinal adequado, no lugar de  $S_R$  e  $W$ , obtêm-se testes de mais fácil interpretação.

A estatística escore é definida pela forma quadrática  $S_R = \mathbf{U}^T \mathbf{K}^{-1} \mathbf{U}$ . O vetor escore  $\mathbf{U}$  tem as seguintes propriedades descritas na Seção 4.1:  $E(\mathbf{U}) = \mathbf{0}$  e

$\text{Cov}(\mathbf{U}) = E(\mathbf{U}\mathbf{U}^T) = \mathbf{K}$ . Supondo observações independentes, o vetor escore é definido por uma soma de variáveis aleatórias independentes que, pelo teorema central do limite, tem distribuição assintótica normal  $p$ -dimensional  $N_p(\mathbf{0}, \mathbf{K})$ . Logo, para amostras grandes, a estatística escore  $S_R = \mathbf{U}^T \mathbf{K}^{-1} \mathbf{U}$  converge, assintoticamente, para uma distribuição  $\chi_p^2$  supondo que o modelo com os parâmetros especificados na hipótese nula seja verdadeiro.

**Exemplo 4.6:** Seja  $y_1, \dots, y_n$  uma amostra de variáveis aleatórias com distribuição normal  $N(\mu, \sigma^2)$  com  $\mu$  desconhecido e  $\sigma^2$  conhecido. No contexto dos MLGs, tem-se:

- i) somente um parâmetro de interesse,  $\mu$ ;
- ii) não há variáveis explanatórias e,
- iii) a função de ligação é a identidade  $\eta = \mu$ .

O logaritmo da função de verossimilhança é

$$\ell = \ell(\mu) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 - \frac{n}{2} \log(2\pi\sigma^2),$$

a partir do qual se obtêm:

$$U = \frac{d\ell}{d\mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) = \frac{n}{\sigma^2} (\bar{y} - \mu),$$

$$E(U) = \frac{n}{\sigma^2} [E(\bar{Y}) - \mu] = 0$$

e

$$K = \text{Var}(U) = \frac{n^2}{\sigma^4} \text{Var}(\bar{Y}) = \frac{n}{\sigma^2}.$$

Portanto,

$$S_R = U^T K^{-1} U = \frac{n^2 (\bar{Y} - \mu)^2 \sigma^2}{\sigma^4 n} = \frac{(\bar{Y} - \mu)^2}{\frac{\sigma^2}{n}} \sim \chi_1^2,$$

resultado que pode ser usado para a construção de intervalos de confiança para  $\mu$  para  $\sigma^2$  conhecido.

**Exemplo 4.7:** Suponha que  $Y$  tem distribuição binomial  $B(m, \pi)$ . Então, o logaritmo da função de verossimilhança para uma única observação é

$$\ell(\pi) = \log \binom{m}{y} + y \log(\pi) + (m - y) \log(1 - \pi)$$

e, portanto,

$$U = \frac{d\ell(\pi)}{d\pi} = \frac{y}{\pi} - \frac{(m - y)}{1 - \pi} = \frac{y - m\pi}{\pi(1 - \pi)}.$$

Mas,  $E(Y) = \mu = m\pi$  e  $\text{Var}(Y) = m\pi(1 - \pi) = \frac{\mu}{m}(m - \mu)$ . Logo,

$$E(U) = 0 \quad \text{e} \quad K = \text{Var}(U) = \frac{\text{Var}(Y)}{\pi^2(1 - \pi)^2} = \frac{m}{\pi(1 - \pi)}.$$

Assim,

$$S_R = U^T K^{-1} U = \frac{(Y - m\pi)^2}{\pi^2(1 - \pi)^2} \frac{\pi(1 - \pi)}{m} = \frac{(Y - m\pi)^2}{m\pi(1 - \pi)} = \frac{[Y - E(Y)]^2}{\text{Var}(Y)}$$

que, pelo teorema central do limite, tem distribuição  $\chi_1^2$ , ou, equivalentemente,

$$\frac{Y - E(Y)}{\sqrt{\text{Var}(Y)}} = \frac{\sqrt{m}(Y - \mu)}{\sqrt{\mu(m - \mu)}} \xrightarrow{\mathcal{D}} N(0, 1),$$

resultado que pode ser usado para se fazer inferência sobre  $\mu$ .

#### 4.6.2 Teste de uma hipótese nula composta

Quando se tem um vetor de parâmetros  $\beta$  em um MLG, muitas vezes há interesse em testar apenas um subconjunto de  $\beta$ . Supõe-se que o parâmetro de dispersão  $\phi$  é conhecido. Seja, então, uma partição do vetor de parâmetros  $\beta$  expressa por  $\beta = (\beta_1^T \ \beta_2^T)^T$ , em que  $\beta_1$ , de dimensão  $q$ , é o vetor de interesse e  $\beta_2$ , de dimensão  $(p - q)$ , o vetor de parâmetros de perturbação. De forma semelhante, tem-se a partição da matriz modelo  $\mathbf{X} = (\mathbf{X}_1 \ \mathbf{X}_2)$ , do vetor escore  $\mathbf{U} = \phi^{-1} \mathbf{X}^T \mathbf{W} \mathbf{G}(\mathbf{y} - \boldsymbol{\mu}) = (\mathbf{U}_1^T \ \mathbf{U}_2^T)^T$  com  $\mathbf{U}_1 = \phi^{-1} \mathbf{X}_1^T \mathbf{W} \mathbf{G}(\mathbf{y} - \boldsymbol{\mu})$  e  $\mathbf{U}_2 = \phi^{-1} \mathbf{X}_2^T \mathbf{W} \mathbf{G}(\mathbf{y} - \boldsymbol{\mu})$  e da matriz de informação de Fisher para  $\beta$

$$\mathbf{K} = \phi^{-1} \mathbf{X}^T \mathbf{W} \mathbf{X} = \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{bmatrix},$$

sendo que  $\mathbf{K}_{12} = \mathbf{K}_{21}^T$ .

Usando-se resultados conhecidos de álgebra linear, que envolvem partição de matrizes (Searle, 1982), tem-se, para amostras grandes, a variância assintótica de  $\hat{\beta}_1$ :

$$\text{Cov}(\hat{\beta}_1) = (\mathbf{K}_{11} - \mathbf{K}_{12}\mathbf{K}_{22}^{-1}\mathbf{K}_{21})^{-1} = \phi[\mathbf{X}_1^T\mathbf{W}^{1/2}(\mathbf{I} - \mathbf{P}_2)\mathbf{W}^{1/2}\mathbf{X}_1]^{-1},$$

sendo  $\mathbf{P}_2 = \mathbf{W}^{1/2}\mathbf{X}_2(\mathbf{X}_2^T\mathbf{W}\mathbf{X}_2)^{-1}\mathbf{X}_2^T\mathbf{W}^{1/2}$  a matriz projeção segundo o modelo com matriz  $\mathbf{X}_2$ .

Sejam as hipóteses

$$H_0 : \beta_1 = \beta_{1,0} \text{ versus } H : \beta_1 \neq \beta_{1,0},$$

sendo  $\beta_{1,0}$  um vetor especificado para  $\beta_1$ . Seja  $\hat{\beta} = (\hat{\beta}_1^T \hat{\beta}_2^T)^T$  a EMV de  $\beta$  sem restrição e  $\tilde{\beta} = (\beta_{1,0}^T \tilde{\beta}_2^T)^T$  a EMV restrita de  $\beta$ , em que  $\tilde{\beta}_2$  é a EMV de  $\beta_2$  sob  $H_0$ . A seguir, são definidos os três testes mais usados para testar a hipótese  $H_0$ .

#### (a) Teste da razão de verossimilhanças

Envolve a comparação dos valores do logaritmo da função de verossimilhança maximizada sem restrição ( $\ell(\hat{\beta}_1, \hat{\beta}_2)$ ) e sob  $H_0$  ( $\ell(\beta_{1,0}, \tilde{\beta}_2)$ ), ou, em termos do desvio, a comparação de  $D(\mathbf{y}; \hat{\mu})$  e  $D(\mathbf{y}; \tilde{\mu})$ , em que  $\tilde{\mu} = g^{-1}(\tilde{\eta})$  e  $\tilde{\eta} = \mathbf{X}\tilde{\beta}$ . Esse teste é, geralmente, preferido no caso de hipóteses relativas a vários coeficientes  $\beta_s$ . Se as diferenças são grandes, então, a hipótese  $H_0$  é rejeitada. A estatística da razão de verossimilhanças para esse teste pode ser expressa como uma diferença de desvios

$$w = 2[\ell(\hat{\beta}_1, \hat{\beta}_2) - \ell(\beta_{1,0}, \tilde{\beta}_2)] = \phi^{-1}[D(\mathbf{y}; \tilde{\mu}) - D(\mathbf{y}; \hat{\mu})]. \quad (4.15)$$

Para amostras grandes, rejeita-se  $H_0$ , a um nível de  $100\alpha\%$  de significância, se  $w > \chi_{q,1-\alpha}^2$ .

#### (b) Teste de Wald

É baseado na distribuição normal assintótica de  $\hat{\beta}$ , sendo uma generalização da estatística t de Student (Wald, 1943). É, geralmente, o mais usado no caso de



hipóteses relativas a um único coeficiente  $\beta_r$ . Tem como vantagem, em relação ao teste da razão de verossimilhanças, o fato de não haver necessidade de se calcular a EMV restrita  $\tilde{\beta}_2$ . Como descrito na Seção 4.1, assintoticamente,  $\hat{\beta} \sim N_p(\beta, \mathbf{K}^{-1})$ . Assim, a estatística para esse teste é

$$W = (\hat{\beta}_1 - \beta_{1,0})^T \widehat{\text{Cov}}(\hat{\beta}_1)^{-1} (\hat{\beta}_1 - \beta_{1,0}), \quad (4.16)$$

sendo  $\widehat{\text{Cov}}(\hat{\beta}_1)$  a matriz  $\text{Cov}(\hat{\beta}_1)$  avaliada em  $\hat{\beta} = (\hat{\beta}_1^T \ \hat{\beta}_2^T)^T$ . Para amostras grandes, rejeita-se  $H_0$ , a um nível de  $100\alpha\%$  de significância, se  $W > \chi_{q,1-\alpha}^2$ .

### (c) **Teste escore**

A estatística para esse teste é definida a partir da função escore como

$$S_R = \mathbf{U}_1^T(\tilde{\beta}) \widetilde{\text{Cov}}(\hat{\beta}_1) \mathbf{U}_1(\tilde{\beta}), \quad (4.17)$$

sendo  $\widetilde{\text{Cov}}(\hat{\beta}_1)$  a matriz  $\text{Cov}(\hat{\beta}_1)$  avaliada em  $\tilde{\beta} = (\beta_{1,0}^T \ \tilde{\beta}_2^T)^T$ . Para amostras grandes, rejeita-se  $H_0$ , a um nível de  $100\alpha\%$  de significância, se  $S_R > \chi_{q,1-\alpha}^2$ .

As três estatísticas (4.15), (4.16) e (4.17) diferem por termos de ordem  $O_p(n^{-1})$ . As expansões assintóticas das distribuições dessas três estatísticas são descritas em Cordeiro (1999, Seção 5.7).

Para o cálculo das estatísticas Wald e escore, deve-se obter  $\text{Cov}(\hat{\beta}_1)$  da inversa da matriz de informação subdividida como  $\mathbf{K}$ , ou seja,

$$\text{Cov}(\hat{\beta}) = \mathbf{K}^{-1} = \phi(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} = \begin{pmatrix} \mathbf{K}^{11} & \mathbf{K}^{12} \\ \mathbf{K}^{21} & \mathbf{K}^{22} \end{pmatrix},$$

sendo que  $\mathbf{K}^{12} = \mathbf{K}^{21^T}$ ,  $\text{Cov}(\hat{\beta}_1) = \mathbf{K}^{11}$ ,  $\text{Cov}(\hat{\beta}_2) = \mathbf{K}^{22}$  e  $\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = \mathbf{K}^{12}$ .

## 4.7 Regiões de confiança

Considera-se, nesta seção que o parâmetro  $\phi$  é conhecido ou estimado. Regiões de confiança assintóticas para  $\beta_1$  podem ser construídas usando-se qualquer uma das três estatísticas de teste. A partir da estatística da razão de verossimilhanças, uma região de confiança para  $\beta_1$ , com um coeficiente de confiança de

$100(1 - \alpha)\%$ , inclui todos os valores de  $\beta_1$  tais que:

$$2[\ell(\hat{\beta}_1, \hat{\beta}_2) - \ell(\beta_1, \tilde{\beta}_2)] < \chi_{q,1-\alpha}^2,$$

em que  $\tilde{\beta}_2$  é a EMV de  $\beta_2$  para cada valor de  $\beta_1$  que é testado ser pertencente, ou não, ao intervalo, e  $\chi_{q,1-\alpha}^2$  é o percentil da distribuição  $\chi^2$  com  $q$  graus de liberdade, correspondente a um nível de significância igual a  $100\alpha\%$ .

Usando-se a estatística de Wald, uma região de confiança para  $\beta_1$ , com um coeficiente de confiança de  $100(1 - \alpha)\%$ , inclui todos os valores de  $\beta_1$  tais que:

$$(\hat{\beta}_1 - \beta_1)^T \widehat{\text{Cov}}(\hat{\beta}_1)^{-1} (\hat{\beta}_1 - \beta_1) < \chi_{q,1-\alpha}^2.$$

Alternativamente, regiões de confiança para os parâmetros lineares  $\beta_1, \dots, \beta_p$  de um MLG podem ser construídas, usando-se a função desvio. Deseja-se uma região de confiança aproximada para um conjunto particular de parâmetros  $\beta_1, \dots, \beta_q$  de interesse. Sejam  $S_p$  o desvio do modelo  $M_p$  com todos os  $p$  parâmetros e  $S_{p-q}$  o desvio do modelo  $M_{p-q}$  com  $p - q$  parâmetros linearmente independentes, e os  $q$  parâmetros de interesse tendo valores fixados:  $\beta_r = \beta_r^*$ ,  $r = 1, \dots, q$ . No ajuste do modelo  $M_{p-q}$ , a quantidade  $\sum_{r=1}^q \beta_r^* \mathbf{x}^{(r)}$  representa um *offset* (isto é, uma parte conhecida na estrutura linear do modelo), sendo  $\mathbf{x}^{(r)}$  a  $r$ -ésima coluna da matriz modelo  $\mathbf{X}$  correspondente a  $\beta_r$ .

Uma região aproximada de  $100(1 - \alpha)\%$  de confiança para  $\beta_1, \dots, \beta_q$  é definida pelo conjunto de pontos  $\beta_r^*$ ,  $r = 1, \dots, q$ , não rejeitados pela estatística  $S_{p-q} - S_p$ , isto é, por

$$\{\beta_r^*, r = 1, \dots, q; S_{p-q} - S_p < \chi_{q,1-\alpha}^2\}. \quad (4.18)$$

Embora, na prática, o cálculo dessas regiões de confiança apresente um trabalho considerável, os software R, S-Plus, SAS e MATLAB têm as facilidades necessárias, incluindo o uso de gráficos.

No caso do intervalo de confiança para um único parâmetro  $\beta_r$ , tem-se

$$\{\beta_r^*; S_{p-1} - S_p < \chi_{1,1-\alpha}^2\}, \quad (4.19)$$

em que  $S_{p-1}$  é o desvio do modelo com os parâmetros  $\beta_1, \dots, \beta_{r-1}, \beta_{r+1}, \dots, \beta_p$  e *offset*  $\beta_r^* \mathbf{x}^{(r)}$ . Um outro intervalo aproximado para  $\beta_r$ , simétrico e assintoticamente equivalente a (4.19), pode ser obtido de

$$[\hat{\beta}_r - a_{\alpha/2}(-\hat{\kappa}^{rr})^{1/2}, \hat{\beta}_r + a_{\alpha/2}(-\hat{\kappa}^{rr})^{1/2}], \quad (4.20)$$

em que  $-\hat{\kappa}^{rr}$  é o elemento  $(r, r)$  de  $\hat{\mathbf{K}}^{-1}$  e  $\Phi(-a_{\alpha/2}) = \alpha/2$ , sendo  $\Phi(\cdot)$  a f.d.a. da distribuição normal  $N(0, 1)$ .

A construção de (4.20) é muito mais simples do que (4.19), pois é necessário apenas o ajuste do modelo  $M_p$ . A grande vantagem do uso da equação (4.19), ao invés de (4.20), é de ser independente da parametrização adotada. Por exemplo, com uma parametrização diferente para o parâmetro de interesse  $\gamma_r = h(\beta_r)$ , o intervalo baseado na distribuição normal assintótica de  $\hat{\gamma}_r$  não corresponde exatamente a (4.20). Entretanto, usando (4.19), o intervalo para  $\gamma_r$  pode ser calculado por simples transformação  $\{h(\beta_r^*); S_{p-1} - S_p < \chi_{1,1-\alpha}^2\}$ .

## 4.8 Seleção de variáveis explanatórias

Na prática, é difícil selecionar um conjunto de variáveis explanatórias para formar um modelo parcimonioso, devido aos problemas de ordem combinatória e estatística. O problema de cunho combinatório é selecionar todas as combinações possíveis de variáveis explanatórias que deverão ser testadas para inclusão no preditor linear. O problema estatístico é definir, com a inclusão de um novo termo no preditor linear, o balanço entre o efeito de reduzir a discrepância entre  $\hat{\mu}$  e  $y$  e o fato de se ter um modelo mais complexo.

Outras estatísticas que servem como medidas de comparação da qualidade de ajuste do modelo e o seu grau de complexidade são os critérios de informação de Akaike  $AIC_p = -2\hat{\ell}_p + 2p$  (Akaike, 1974) e de Bayes  $BIC_p = -2\hat{\ell}_p + p \log(n)$  (Schwarz, 1978) que para os MLGs podem ser expressos, respectivamente, como

$$AIC_p = S_p + 2p - 2\hat{\ell}_n. \quad (4.21)$$

e

$$BIC_p = S_p + p \log(n) - 2\hat{\ell}_n. \quad (4.22)$$

Se o modelo envolver um parâmetro de dispersão  $\phi$ , este deve ser estimado, como descrito na Seção 4.4, para calcular um valor numérico em (4.21) e (4.22).

O critério de Akaike foi desenvolvido para estender o método de máxima verossimilhança para a situação de ajustes de vários modelos com diferentes números de parâmetros e para decidir quando parar o ajuste. A estatística (4.21) pode ajudar na seleção de modelos complexos e tem demonstrado produzir soluções razoáveis para muitos problemas de seleção de modelos que não podem ser abordados pela teoria convencional de máxima verossimilhança. Um valor baixo para  $AIC_p$  é considerado como representativo de um melhor ajuste e os modelos são selecionados visando a se obter um mínimo  $AIC_p$ . De forma semelhante interpreta-se  $BIC_p$ .

Uma outra medida de comparação equivalente ao critério de Akaike é

$$C_p^* = S_p + 2p - n = AIC_p + 2\hat{\ell}_n - n. \quad (4.23)$$

Para um MLG isolado é, usualmente, mais simples trabalhar com  $C_p^*$  do que  $AIC_p$ . Para o modelo normal linear com variância constante  $\sigma^2$ ,  $C_p^*$  reduz-se à estatística  $C_p = SQR_p/\tilde{\sigma}^2 + 2p - n$  (Mallows, 1966), em que  $SQR_p = \sum_{\ell=1}^n (y_\ell - \hat{\mu}_\ell)^2$  e  $\tilde{\sigma}^2 = SQR_m/(n - m)$  é, a menos de um coeficiente multiplicador, o resíduo quadrático médio baseado no modelo maximal com  $m$  parâmetros. Nesse caso,  $AIC_p = SQR_p/\tilde{\sigma}^2 + 2p + n \log(2\pi\tilde{\sigma}^2)$ . Nota-se que  $C_m = m$ .

Em geral,  $E(C_p^*) \neq p$ . Para o modelo normal linear com variância conhecida tem-se  $E(C_p^*) = p$ , supondo que o modelo é verdadeiro. Se a variância for desconhecida, o valor esperado de  $C_p^*(= C_p)$  será muito maior do que  $p$ , quando o modelo não se ajustar bem aos dados. Um gráfico de  $C_p^*$  (ou  $AIC_p$ ) *versus*  $p$  fornece uma boa indicação para comparar modelos alternativos. Considerando dois modelos encaixados  $M_q \subset M_p$ ,  $p > q$ , tem-se  $AIC_p - AIC_q = C_p^* - C_q^* = S_p - S_q + 2(p - q)$  e, portanto, supondo  $M_q$  verdadeiro,  $E(AIC_p - AIC_q) = p - q + O(n^{-1})$ .

Na comparação de modelos, sucessivamente, mais ricos, a declividade esperada do segmento de reta unindo  $AIC_p$  com  $AIC_q$  (ou  $C_p^*$  com  $C_q^*$ ) deve ser próxima de um, supondo o modelo mais pobre  $M_q$  verdadeiro. Pares de modelos com declividade observada maior do que um indicam que o modelo maior ( $M_p$ ) não é, significativamente, melhor do que o modelo menor ( $M_q$ ).

Uma outra tentativa para seleção de variáveis explanatórias é minimizar a expressão (Atkinson, 1981)

$$A_p = D_p + \frac{p\alpha}{\phi}, \quad (4.24)$$

em que  $D_p$  é o desvio do modelo  $M_p$  sem o parâmetro de dispersão  $\phi$  e  $\alpha$  é uma constante ou função de  $n$ . Para o cálculo de (4.24),  $\phi$  é estimado como descrito na Seção 4.4. Tem-se  $A_p = [C_p^* + p(\alpha - 2) + n]/p$  e para  $\alpha = 2$ ,  $A_p$  é equivalente a  $C_p^*$  (ou  $AIC_p$ ).

## 4.9 Exercícios

1. Para os modelos normal, gama, normal inverso e Poisson com componentes sistemáticos  $\eta_i = \mu_i^\lambda = \beta_0 + \beta_1 x_i$ , e para o modelo binomial com  $\eta_i = \log\{[(1 - \mu_i)^{-\lambda} - 1]\lambda^{-1}\} = \beta_0 + \beta_1 x_i$ , sendo  $\lambda$  conhecido, calcule: a) as estruturas de covariância assintótica de  $\hat{\beta}$  e  $\hat{\mu}$ ; b) as estatísticas score, de Wald e da razão de verossimilhanças nos testes:  $H_1 : \beta_1 = 0$  versus  $H'_1 : \beta_1 \neq 0$  e  $H_2 : \beta_0 = 0$  versus  $H'_2 : \beta_0 \neq 0$ ; c) intervalos de confiança para os parâmetros  $\beta_0$  e  $\beta_1$ .

2. Sejam  $Y_1, \dots, Y_n$  variáveis binárias independentes e identicamente distribuídas com  $P(Y_i = 1) = 1 - P(Y_i = 0) = \mu$ ,  $0 < \mu < 1$ . A distribuição de  $Y_i$  pertence à família (1.5) com parâmetro natural  $\theta$ . Demonstre que a estatística de Wald para testar  $H_0 : \theta = 0$  versus  $H : \theta \neq 0$  é  $W = [n\hat{\theta}^2 \exp(\hat{\theta})]/[1 + \exp(\hat{\theta})]^2$ , sendo os valores possíveis de  $\hat{\theta}$  iguais a  $\log[t/(n-t)]$ ,  $t = 1, \dots, n-1$ . Quais as formas das estatísticas

escore e da razão de verossimilhanças?

**3.** Deduza as expressões das estatísticas desvio  $D_p$  e  $X_p^2$  de Pearson generalizada para as distribuições descritas no Capítulo 1.

**4.** a) Mostre que para os modelos log-lineares com a matriz do modelo tendo uma coluna de uns, o desvio reduz-se a  $S_p = 2 \sum_{i=1}^n y_i \log(y_i/\hat{\mu}_i)$ ;

b) Mostre que para o modelo gama com índice  $\nu$  e função de ligação potência  $\eta = \mu^\lambda$  ou  $\eta = \log(\mu)$ , nesse último caso a matriz  $\mathbf{X}$  tendo uma coluna de uns, o desvio reduz-se a  $S_p = 2\nu \sum_{i=1}^n \log(\hat{\mu}_i/y_i)$ .

**5.** Mostre que aos dois modelos do exercício 4. se aplica o resultado mais geral  $\sum_{i=1}^n (y_i - \hat{\mu}_i) \hat{\mu}_i V^{-1}(\hat{\mu}_i) = 0$  quando o modelo tem função de ligação  $\eta = \mu^\lambda (\lambda \neq 0)$  ou  $\eta = \log(\mu)$ , nesse último caso,  $\mathbf{X}$  com uma coluna de uns.

**6.** a) Mostre que para o modelo gama simples com índice  $\nu$ , em que todas as médias são iguais, o desvio reduz-se à estatística clássica  $S_1 = 2n\nu \log(\bar{y}/\tilde{y})$ , em que  $\bar{y}$  e  $\tilde{y}$  são as médias aritmética e geométrica dos dados, respectivamente.

b) Mostre que, para um MLG, sendo  $\ell$  o logaritmo da função de verossimilhança total,  $E(\partial^2 \ell / \partial \phi \partial \beta_j) = 0$  e, portanto, os parâmetros  $\phi$  e  $\beta$  são ortogonais.

**7.** Demonstre que a EMV do parâmetro de dispersão  $\phi$  é calculada por

a)  $\hat{\phi} = D_p/n$  (modelos normal e normal inverso);

b)  $\hat{\phi} = \frac{D_p}{n} \left( 1 + \sqrt{1 + \frac{2D_p}{3n}} \right)^{-1}$  (modelo gama, expressão aproximada para  $\phi$  pequeno) (Cordeiro e McCullagh, 1991).

**8.** Considere uma única resposta  $Y \sim B(m, \pi)$ .

a) deduza a expressão para a estatística de Wald  $W = (\hat{\pi} - \pi)^T \hat{K} (\hat{\pi} - \pi)$ , em que  $\hat{\pi}$  é a EMV de  $\pi$  e  $\hat{K}$  é a informação de Fisher estimada em  $\hat{\pi}$ ;

b) deduza a expressão para a estatística escore  $S_R = \tilde{U}^T \tilde{K}^{-1} \tilde{U}$  e verifique que é igual

à estatística de Wald;

c) deduza a expressão para a estatística da razão de verossimilhanças  $w = 2[\ell(\hat{\mu}) - \ell(\mu)]$ ;

d) para amostras grandes, as estatísticas escore, de Wald e da razão de verossimilhanças têm distribuição assintótica  $\chi_1^2$ . Sejam  $m = 10$  e  $y = 3$ . Compare essas estatísticas usando  $\pi = 0,1$ ,  $\pi = 0,3$  e  $\pi = 0,5$ . Quais as conclusões obtidas?

**9.** Seja  $y_1, \dots, y_n$  uma amostra de variáveis aleatórias com distribuição exponencial de média  $\mu$ . Sejam as hipóteses  $H_0 : \mu = \mu_0$  versus  $H : \mu \neq \mu_0$ . Demonstre que:

- a)  $w = 2n \left[ \log \left( \frac{\mu_0}{\bar{y}} \right) + \frac{\bar{y} - \mu_0}{\mu_0} \right]$  (teste da razão de verossimilhanças);
- b)  $W = \frac{n(\bar{y} - \mu_0)^2}{\bar{y}^2}$  (teste de Wald);
- c)  $S_R = \frac{n(\bar{y} - \mu_0)^2}{\mu_0^2}$  (teste escore).

**10.** Sejam  $y_1, \dots, y_n$  uma amostra de variáveis aleatórias independentes com distribuição de Poisson com média  $\mu_i = \mu \rho^{i-1}$  ( $i = 1, \dots, n$ ). Deduza as estatísticas escore, de Wald e da razão de verossimilhanças para os testes das hipóteses que se seguem:

- a)  $H_0 : \mu = \mu_0$  versus  $H : \mu \neq \mu_0$ , quando  $\rho$  é conhecido;
- b)  $H_0 : \rho = \rho_0$  versus  $H : \rho \neq \rho_0$ , quando  $\mu$  é conhecido.

**11.** Considere a estrutura linear  $\eta_i = \beta x_i, i = 1, \dots, n$ , com um único parâmetro  $\beta$  desconhecido e função de ligação  $\eta = (\mu^\lambda - 1)\lambda^{-1}$ ,  $\lambda$  conhecido. Calcule a EMV de  $\beta$ , considerando-se os modelos normal, Poisson, gama, normal inverso e binomial negativo. Faça o mesmo para o modelo binomial com função de ligação  $\eta = \log\{[(1 - \mu)^{-\lambda} - 1]\lambda^{-1}\}$ ,  $\lambda$  conhecido. Calcule, ainda, as estimativas quando  $x_1 = \dots = x_n$ .

**12.** No exercício anterior, considere o teste de  $H_0 : \beta = \beta_0$  versus  $H : \beta \neq \beta_0$ , sendo  $\beta_0$  um valor especificado para o parâmetro desconhecido. Calcule:

- a) a variância assintótica de  $\hat{\beta}$ ;
- b) as estatísticas para os testes da razão de verossimilhanças, Wald e escore;

c) um intervalo de confiança, com um coeficiente de confiança de  $100(1 - \alpha)\%$ , para  $\beta$ ;

d) um intervalo de confiança, com um coeficiente de confiança de  $100(1 - \alpha)\%$ , para uma função  $g(\beta)$  com  $g(\cdot)$  conhecido.

**13.** Seja  $y_1, \dots, y_n$  uma amostra de variáveis aleatórias com distribuição gama  $G(\mu, \phi)$  com média  $\mu$  e parâmetro de dispersão  $\phi$ . Demonstre que:

a) a EMV de  $\phi$  satisfaz  $\log(\hat{\phi}) + \psi(\hat{\phi}^{-1}) = \log(\tilde{y}/\bar{y})$ , sendo  $\bar{y}$  e  $\tilde{y}$  as médias aritmética e geométrica dos dados, respectivamente, e  $\psi(\cdot)$  a função digama;

b) uma solução aproximada é expressa como  $\hat{\phi} = 2(\bar{y} - \tilde{y})/\bar{y}$ .

**14.** Sejam  $Y_i \sim N(\mu_i, \sigma^2)$  e os  $Y_i$ 's independentes,  $i = 1, \dots, n$ , com variância constante desconhecida e  $\mu_i = \exp(\beta x_i)$ . Calcule:

a) a matriz de informação para  $\beta$  e  $\sigma^2$ ;

b) as estatísticas score, de Wald e da razão de verossimilhanças nos seguintes testes:  $H_1 : \beta = \beta^{(0)}$  versus  $H'_1 : \beta \neq \beta^{(0)}$  e  $H_2 : \sigma^2 = \sigma^{(0)2}$  versus  $H'_2 : \sigma^2 \neq \sigma^{(0)2}$ ;

c) intervalos de confiança para  $\beta$  e  $\sigma^2$ .

**15.** Sejam  $Y_i \sim P(\mu_i)$  com  $\mu_i = \mu \rho^{i-1}$ ,  $i = 1, \dots, n$ . Calcule as estatísticas score, de Wald e da razão de verossimilhanças nos seguintes testes:

a) de  $H_0 : \mu = \mu^{(0)}$  versus  $H : \mu \neq \mu^{(0)}$  para os casos de  $\rho$  conhecido e desconhecido;

b) de  $H_0 : \rho = \rho^{(0)}$  versus  $H : \rho \neq \rho^{(0)}$  para os casos de  $\mu$  conhecido e desconhecido.

**16.** Sejam  $Y_1, \dots, Y_n$  variáveis aleatórias independentes com distribuição gama  $G(\mu_i, \phi)$ , sendo  $\phi$  o parâmetro de dispersão, com  $\mu_i = \phi^{-1} \exp(-\alpha - \beta x_i)$ , em que  $\phi$ ,  $\alpha$  e  $\beta$  são parâmetros desconhecidos, e os  $x_i$ 's são valores especificados,  $i = 1, \dots, n$ . Calcule estatísticas adequadas para os seguintes testes:

a)  $H_1 : \beta = 0$  versus  $H'_1 : \beta \neq 0$ ;

b)  $H_2 : \phi = \phi^{(0)}$  versus  $H'_2 : \phi \neq \phi^{(0)}$ ;



c)  $H_3 : \alpha = 0$  versus  $H'_3 : \alpha \neq 0$ .

**17.** Seja  $y_1, \dots, y_n$  uma amostra de variáveis aleatórias com distribuição normal  $N(\mu_i, \sigma^2)$ , com  $\sigma^2$  conhecido, e  $\mu_i = \alpha + \beta \exp(-\gamma x_i)$ ,  $i = 1, \dots, n$ , em que  $\alpha$ ,  $\beta$  e  $\gamma$  são parâmetros desconhecidos.

a) Calcule intervalos de confiança para  $\alpha$ ,  $\beta$  e  $\gamma$ ;

b) Teste  $H_0 : \gamma = 0$  versus  $H : \gamma \neq 0$  por meio das estatísticas escore, de Wald e da razão de verossimilhanças;

c) Como proceder em a) e b) se  $\sigma^2$  for desconhecido?

**18.** Sejam  $Y_1, \dots, Y_n$  variáveis aleatórias independentes e identicamente distribuídas como normal  $N(\mu, \sigma^2)$ . Defina-se  $Z_i = |Y_i|$ ,  $i = 1, \dots, n$ . Demonstrar que a razão de verossimilhanças no teste de  $H_0 : \mu = 0$  versus  $H : \mu \neq 0$ ,  $\sigma^2$  desconhecido, é, assintoticamente, equivalente ao teste baseado em valores grandes da estatística  $T = \sum_{i=1}^n Z_i^4 / \sum_{i=1}^n Z_i^2$ , que é uma estimativa do coeficiente de curtose de  $Z$ .

**19.** Considere o teste do exercício anterior. Demonstre as expressões das estatísticas escore  $S_R = n\bar{y}^2/\hat{\sigma}^2$  e de Wald  $W = n\bar{y}^2/\hat{\sigma}^2$ , em que  $\bar{y}$  é a média dos  $y_i$ s,  $\hat{\sigma}^2 = \sum_{i=1}^n (y_i - \bar{y})^2/n$  e  $\tilde{\sigma}^2 = \hat{\sigma}^2 + \bar{y}^2$  são as estimativas de  $\sigma^2$ , segundo  $H$  e  $H_0$ , respectivamente, e que, segundo  $H_0$ ,  $E(W) = (n-1)/(n-3)$  e  $E(S_R) = 1 + O(n^{-2})$ .

**20.** Sejam  $k$  amostras independentes de tamanhos  $n_i$  ( $i = 1, \dots, k$ ;  $n_i \geq 2$ ) obtidas de populações normais diferentes de médias  $\mu_i$  e variâncias  $\sigma_i^2$ ,  $i = 1, \dots, k$ . Formule o critério da razão de verossimilhanças para o teste de homogeneidade de variâncias,  $H_0 : \sigma_1^2 = \dots = \sigma_k^2$  versus  $H : \sigma_i^2$  não é constante. Como realizar esse teste na prática?

**21.** Seja um MLG com estrutura linear  $\eta_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2$  e função de ligação  $g(\cdot)$  conhecida. Determine as estatísticas nos testes da razão de verossimilhanças, de Wald e escore, para as hipóteses:

- a)  $H_0 : \beta_2 = \beta_3 = 0$  versus  $H : H_0$  é falsa;  
 b)  $H_0 : \beta_2 = 0$  versus  $H : \beta_2 \neq 0$ ;  
 c)  $H_0 : \beta_3 = 0$  versus  $H : \beta_3 \neq 0$ .

**22.** Considere uma tabela de contingência  $r \times s$ , em que  $Y_{ij}$  tem distribuição de Poisson  $P(\mu_{ij})$ ,  $i = 1, \dots, r$ ,  $j = 1, \dots, s$ . Para o teste da hipótese de independência linha-coluna versus uma alternativa geral, calcule a forma das estatísticas escore, de Wald e da razão de verossimilhanças.

**23.** Considere o problema de testar  $H_0 : \mu = 0$  versus  $H : \mu \neq 0$  numa distribuição normal  $N(\mu, \sigma^2)$  com  $\sigma^2$  desconhecido. Compare as estatísticas dos testes escore, de Wald e da razão de verossimilhanças entre si e com a distribuição  $\chi^2$  assintótica.

**24.** Seja uma distribuição multinomial com probabilidades  $\pi_1, \dots, \pi_m$  dependendo de um parâmetro  $\theta$  desconhecido. Considere uma amostra de tamanho  $n$ . Calcule a forma das estatísticas dos testes da razão de verossimilhanças, escore e de Wald para testar as hipóteses  $H_0 : \theta = \theta^{(0)}$  versus  $H : \theta \neq \theta^{(0)}$ , sendo  $\theta^{(0)}$  um valor especificado.

**25.** A estatística escore pode ser usada para escolher um entre dois modelos separados. Sejam  $Y_1, \dots, Y_n$  variáveis aleatórias independentes com  $Y_i$  tendo distribuição normal  $N(\mu_i, \sigma^2)$ , com  $\mu_i = \beta x_i$  ou  $\mu_i = \gamma z_i$ ,  $i = 1, \dots, n$ , sendo todos os parâmetros desconhecidos e os  $x_i$ s e os  $z_i$ s conhecidos. Proponha um teste baseado na estatística escore para escolher uma dentre essas estruturas.

**26.** Sejam  $Y_1, \dots, Y_n$  variáveis aleatórias independentes sendo que  $Y_i$ ,  $i = 1, \dots, n$ , tem distribuição binomial negativa inflacionada de zeros (BNIZ) com  $P(Y_i = y_i)$  especificada no Exercício 10 do Capítulo 1, em que  $\log(\lambda) = X\beta$ ,  $\log[(\omega/(1-\omega))] = Z\gamma$ ,  $X$  e  $Z$  são matrizes de variáveis explanatórias e  $\beta$  e  $\gamma$  vetores de parâmetros.

- a) Mostre que a estatística escore para testar a hipótese  $H_0$ : PIZ versus  $H$  : BNIZ,

isto é,  $H_0 : \alpha = 0$  versus  $H_1 : \alpha > 0$  é expressa como  $T = S\sqrt{\hat{\kappa}^{\alpha\alpha}}$ , em que  $S = \frac{1}{2} \sum_i \hat{\lambda}_i^{c-1} \left\{ \left[ (y_i - \hat{\lambda}_i)^2 - y_i \right] - I_{(y_i=0)} \hat{\lambda}_i^2 \hat{\omega}_i / \hat{p}_{0,i} \right\}$ ,  $\hat{\kappa}^{\alpha\alpha}$  é o elemento superior esquerdo da inversa da matriz de informação de Fisher

$$\mathbf{K} = \begin{bmatrix} \kappa_{\alpha\alpha} & K_{\alpha\beta} & K_{\alpha\gamma} \\ K_{\alpha\beta} & K_{\beta\beta} & K_{\beta\gamma} \\ K_{\alpha\gamma} & K_{\beta\gamma} & K_{\gamma\gamma} \end{bmatrix}$$

avaliada na EMV sob  $H_0$ . Note que  $\kappa_{\alpha\alpha}$  é um escalar, e que os outros elementos são, em geral, matrizes com dimensões determinadas pelas dimensões dos vetores de parâmetros  $\boldsymbol{\beta}$  e  $\boldsymbol{\gamma}$ . No limite, quando  $\alpha \rightarrow \infty$ , os elementos típicos da matriz de informação são deduzidos por Ridout et al. (2001).

b) Mostre que, no caso particular em que não há variáveis explanatórias para  $\lambda$  e  $\omega$ , o teste escore simplifica-se para

$$T = \frac{\sum_i \left[ (y_i - \hat{\lambda})^2 - y_i \right] - n\hat{\lambda}^2\hat{\omega}}{\hat{\lambda} \sqrt{n(1 - \hat{\omega}) \left( 2 - \frac{\hat{\lambda}^2}{e^{\hat{\lambda}} - 1 - \hat{\lambda}} \right)}}.$$



# Capítulo 5

## Resíduos e Diagnósticos

### 5.1 Introdução

A escolha de um MLG engloba três etapas principais: i) definição da distribuição (que determina a função de variância); ii) definição da função de ligação; iii) definição da matriz do modelo.

Na prática, porém, pode ocorrer que após uma escolha cuidadosa de um modelo e subsequente ajuste a um conjunto de observações, o resultado obtido seja insatisfatório. Isso decorre em função de algum desvio sistemático entre as observações e os valores ajustados ou, então, porque uma ou mais observações são discrepantes em relação às demais.

Desvios sistemáticos podem surgir pela escolha inadequada da função de variância (falsa distribuição populacional para a variável resposta), da função de ligação e da matriz do modelo, ou ainda pela definição errada da escala da variável dependente ou das variáveis explanatórias (um parâmetro importante que está sendo omitido no modelo) ou mesmo porque algumas observações se mostram dependentes ou exibem alguma forma de correlação serial. Discrepâncias isoladas podem ocorrer porque os pontos estão nos extremos da amplitude de validade da variável explanatória (uma ou mais observações não pertencendo à distribuição proposta para a variável resposta), ou porque eles estão realmente errados como resultado de uma leitura incorreta ou uma transcrição mal feita, ou ainda porque algum fator não controlado influenciou a sua obtenção.

Na prática, em geral, há uma combinação dos diferentes tipos de falhas. Assim, por exemplo, a detecção de uma escolha incorreta da função de ligação pode ocorrer porque ela está realmente errada ou porque uma ou mais variáveis explanatórias estão na escala errada ou devido à presença de alguns pontos discrepantes. Esse fato faz com que a verificação da adequação de um modelo para um determinado conjunto de observações seja um processo realmente difícil. Maiores detalhes sobre análise de resíduos e diagnósticos podem ser encontrados em Atkinson (1985), Cordeiro (1986), Atkinson et al. (1989), McCullagh e Nelder (1989), Francis et al. (1993) e Paula (2004).

## 5.2 Técnicas para verificar o ajuste de um modelo

As técnicas usadas com esse objetivo podem ser formais ou informais. As informais baseiam-se em exames visuais de gráficos para detectar padrões, ou então, pontos discrepantes. As formais envolvem especificar o modelo sob pesquisa em uma classe mais ampla pela inclusão de um parâmetro (ou vetor de parâmetros) extra  $\gamma$ . As mais usadas são baseadas nos testes da razão de verossimilhanças e escore. Parâmetros extras podem aparecer devido à:

- inclusão de uma variável explanatória adicional;
- inclusão de uma variável explanatória  $x$  em uma família  $h(x, \gamma)$  indexada por um parâmetro  $\gamma$ , sendo um exemplo a família de Box-Cox;
- inclusão de uma função de ligação  $g(\mu)$  em uma família mais ampla  $g(\mu, \gamma)$ , sendo um exemplo a família de Aranda-Ordaz (1981), especificada no Exercício 3 do Capítulo 2;
- inclusão de uma variável construída, por exemplo  $\hat{\eta}^2$ , a partir do ajuste original, para o teste de adequação da função de ligação;
- inclusão de uma variável *dummy* tendo o valor 1 (um) para a unidade discrepante e 0 (zero) para as demais. Isso é equivalente a eliminar essa observação

do conjunto de dados, fazer a análise com a observação discrepante e sem ela e verificar, então, se a mudança no valor do desvio é significativa, ou não. Ambos, porém, dependem da localização do(s) ponto(s) discrepante(s).

### 5.3 Análise de resíduos e diagnósticos para o modelo linear clássico

No modelo clássico de regressão  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ , os elementos  $\epsilon_i$  do vetor  $\boldsymbol{\epsilon}$  são as diferenças entre os valores observados  $y_i$  e aqueles esperados  $\mu_i$  pelo modelo. Esses elementos são denominados erros aleatórios (ou ruídos brancos) e considera-se que os  $\epsilon_i$  são independentes e, além disso, que  $\epsilon_i$  tem distribuição normal com média zero e variância  $\sigma^2$ , isto é,  $N(0, \sigma^2)$ . Esses termos representam a variação natural dos dados, mas, também, podem ser interpretados como o efeito cumulativo de fatores que não foram considerados no modelo. Se as pressuposições do modelo são violadas, a análise resultante pode conduzir a resultados duvidosos. Esse tipo de violação do modelo origina as falhas denominadas sistemáticas (não linearidade, não-normalidade, heterocedasticidade, não-independência, etc). Outro fato bastante comum é a presença de pontos atípicos (falhas isoladas), que podem influenciar, ou não, no ajuste do modelo. Eles podem surgir de várias maneiras. Segundo Atkinson (1985), algumas possibilidades são:

- devido a erros grosseiros na variável resposta ou nas variáveis explanatórias, por medidas erradas ou registro da observação, ou ainda, erros de transcrição;
- observação proveniente de uma condição distinta das demais;
- modelo mal especificado (falta de uma ou mais variáveis explanatórias, modelo inadequado, etc);
- escala usada de forma errada, isto é, talvez os dados sejam melhor descritos após uma transformação, do tipo logarítmica ou raiz quadrada, por exemplo;

- a parte sistemática do modelo e a escala estão corretas, mas a distribuição da resposta tem uma cauda mais longa do que a distribuição normal.

A partir de um conjunto de observações e ajustando-se um determinado modelo com  $p$  parâmetros linearmente independentes, para verificar as pressuposições devem ser considerados como elementos básicos:

- os valores estimados (ou ajustados)  $\hat{\mu}_i$ ;
- os resíduos ordinários  $r_i = y_i - \hat{\mu}_i$ ;
- a variância residual estimada (ou quadrado médio residual),  $\hat{\sigma}^2 = s^2 = \text{QMRes} = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 / (n - p)$ ;
- os elementos da diagonal (*leverage*) da matriz de projeção  $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ , isto é,

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

sendo  $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$ .

Uma ideia importante, também, é a da deleção (*deletion*), isto é, a comparação do ajuste do modelo escolhido, considerando-se todos os pontos, com o ajuste do mesmo modelo sem os pontos atípicos. As estatísticas obtidas pela omissão de um certo ponto  $i$  são denotadas com um índice entre parênteses. Assim, por exemplo,  $s_{(i)}^2$  representa a variância residual estimada para o modelo ajustado, excluído o ponto  $i$ .

### 5.3.1 Tipos de resíduos

Vale destacar que os resíduos têm papel fundamental na verificação do ajuste de um modelo. Vários tipos de resíduos foram propostos na literatura (Cook e Weisberg, 1982; Atkinson, 1985).

#### a) Resíduos ordinários



Os resíduos do processo de ajuste por mínimos quadrados são definidos por

$$r_i = y_i - \hat{\mu}_i.$$

Enquanto os erros  $\epsilon_i$  são independentes e têm a mesma variância, o mesmo não ocorre com os resíduos obtidos a partir do ajuste do modelo, usando-se mínimos quadrados. Tem-se,

$$\text{Var}(\mathbf{R}) = \text{Var}[(\mathbf{I} - \mathbf{H})\mathbf{Y}] = \sigma^2(\mathbf{I} - \mathbf{H}).$$

Em particular, a variância do  $i$ -ésimo resíduo é igual a  $\text{Var}(R_i) = \sigma^2(1 - h_{ii})$  e a covariância dos resíduos relativos às observações  $i$  e  $j$  é  $\text{Cov}(R_i, R_j) = -\sigma^2 h_{ij}$ .

Assim, o uso dos resíduos ordinários pode não ser adequado devido à heterogeneidade das variâncias e falta de independência. Então, foram propostas diferentes padronizações para minimizar esse problema.

**b) Resíduos estudentizados internamente** (*Studentized residuals*)

Considerando-se  $s^2 = \text{QMRes}$  como a estimativa de  $\sigma^2$ , tem-se que um estimador não viesado para  $\text{Var}(r_i)$  é expresso por

$$\widehat{\text{Var}}(R_i) = (1 - h_{ii})s^2 = (1 - h_{ii}) \text{QMRes}.$$

Como  $E(R_i) = E(Y_i - \hat{\mu}_i) = 0$ , o resíduo estudentizado internamente é igual a

$$\text{rsi}_i = \frac{r_i}{s\sqrt{(1 - h_{ii})}} = \frac{y_i - \hat{\mu}_i}{\sqrt{(1 - h_{ii}) \text{QMRes}}}.$$

Esses resíduos são mais sensíveis do que os anteriores por considerarem variâncias distintas. Entretanto, um valor discrepante pode alterar profundamente a variância residual dependendo do modo como se afasta do grupo maior das observações. Além disso, o numerador e o denominador dessa expressão são variáveis dependentes, isto é,  $\text{Cov}(R_i, \text{QMRes}) \neq 0$ .

**c) Resíduos estudentizados externamente** (*jackknifed residuals, deletion residuals, externally Studentized residuals, RStudent*)

Para garantir a independência entre o numerador e o denominador, na padronização dos resíduos, define-se o resíduo estudentizado externamente, como

$$\text{rse}_{(i)} = \frac{r_i}{s_{(i)} \sqrt{(1 - h_{ii})}},$$

sendo  $s_{(i)}^2$  o quadrado médio residual livre da influência da observação  $i$ , ou seja, a estimativa de  $\sigma^2$ , omitindo-se a observação  $i$ . Atkinson (1985) demonstra que

$$\text{rse}_{(i)} = \text{rsi}_i \sqrt{\frac{n - p - 1}{n - p - \text{rsi}_i^2}},$$

sendo  $p$  o número de parâmetros independentes do modelo e  $\text{rsi}_i$  definido no item b).

A vantagem de usar o resíduo  $\text{rse}_{(i)}$  é que, sob normalidade, tem distribuição  $t$  de Student com  $(n - p - 1)$  graus de liberdade. Embora não seja recomendada a prática de testes de significância na análise de resíduos, sugere-se que a  $i$ -ésima observação requer atenção especial se  $|\text{rse}_{(i)}|$  for maior do que o  $100[1 - \alpha/(2n)]$ -ésimo percentil da distribuição  $t$  com  $(n - p - 1)$  graus de liberdade, sendo que o nível de significância  $\alpha$  é dividido por  $n$  por ser esse o número de observações sob análise.

### 5.3.2 Estatísticas para diagnóstico

Discrepâncias isoladas (pontos atípicos) podem ser caracterizadas por terem  $h_{ii}$  e/ou resíduos grandes, serem inconsistentes e/ou influentes (McCullagh e Nelder, 1989, p. 404). Uma observação inconsistente é aquela que se afasta da tendência geral das demais. Quando uma observação está distante das outras em termos das variáveis explanatórias, ela pode ser, ou não, influente. Uma observação influente é aquela cuja omissão do conjunto de dados resulta em mudanças substanciais nas estatísticas de diagnóstico do modelo. Essa observação pode ser um *outlier* (observação aberrante), ou não. Uma observação pode ser influente de diversas maneiras, isto é,

- no ajuste geral do modelo;
- no conjunto das estimativas dos parâmetros;
- na estimativa de um determinado parâmetro;

- na escolha de uma transformação da variável resposta ou de uma variável explanatória.

As estatísticas mais usadas para verificar pontos atípicos são:

- medida de *leverage*:  $h_{ii}$ ;
- medida de inconsistência:  $\text{rse}_{(i)}$ ;
- medida de influência sobre o parâmetro  $\beta_j$ :  $\text{DFBetaS}_{(i)}$  para  $\beta_j$ ;
- medidas de influência geral:  $\text{DFFitS}_{(i)}$ ,  $D_{(i)}$  ou  $C_{(i)}$ .

De uma forma geral, pode-se classificar uma observação como:

- ponto inconsistente: ponto com  $|\text{rse}_{(i)}|$  grande, isto é, tal que  $|\text{rse}_{(i)}| \geq t_{1-\alpha/(2n); n-p-1}$ , com nível de significância igual a  $\alpha$ ;
- ponto de alavanca: ponto com  $h_{ii}$  grande, isto é, tal que  $h_{ii} \geq 2p/n$  (Belsley et al., 1980, p. 17). Pode ser classificado como bom, quando consistente, ou ruim, quando inconsistente;
- *outlier*: ponto inconsistente com *leverage* pequeno, ou seja, com  $\text{rse}_{(i)}$  grande e  $h_{ii}$  pequeno;
- ponto influente: ponto com  $\text{DFFitS}_{(i)}$ ,  $C_{(i)}$ ,  $D_{(i)}$  ou  $\text{DFBetaS}_{(i)}$  grande, como explicado a seguir.

A seguir, são descritas com mais detalhes as estatísticas citadas.

#### a) Elementos da diagonal da matriz de projeção $\mathbf{H}$ ( $h_{ii}$ , *leverage*)

A distância de uma observação em relação às demais é medida por  $h_{ii}$  (medida de *leverage*). No caso particular da regressão linear simples, usando-se a variável centrada  $x_i = X_i - \bar{X}$ , tem-se:

$$\mathbf{H} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdots & \cdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{\sum_{i=1}^n x_i^2} \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix}$$

e, portanto,

$$h_{ii} = \frac{1}{n} + \frac{x_i^2}{\sum_{i=1}^n x_i^2} = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum_{i=1}^n x_i^2}, \text{ elementos da diagonal de } \mathbf{H} \text{ e}$$

$$h_{ij} = \frac{1}{n} + \frac{x_i x_j}{\sum_{i=1}^n x_i^2} = \frac{1}{n} + \frac{(X_i - \bar{X})(X_j - \bar{X})}{\sum_{i=1}^n x_i^2}, \text{ elementos fora da diagonal de } \mathbf{H},$$

o que mostra que à medida que  $X_i$  se afasta de  $\bar{X}$ , o valor de  $h_{ii}$  aumenta e que seu valor mínimo é  $1/n$ . Esse valor mínimo ocorre para todos os modelos que incluem uma constante. No caso em que o modelo de regressão passa pela origem, o valor mínimo de  $h_{ii}$  é 0 para uma observação  $X_i = 0$ . O valor máximo de  $h_{ii}$  é 1, ocorrendo quando o modelo ajustado é irrelevante para a predição em  $X_i$  e o resíduo é igual a 0. Sendo  $\mathbf{H}$  uma matriz de projeção, tem-se  $\mathbf{H} = \mathbf{H}^2$  e, portanto,

$$h_{ii} = \sum_{j=1}^n h_{ij}^2 = h_{ii}^2 + \sum_{j \neq i} h_{ij}^2$$

concluindo-se que  $0 \leq h_{ii} \leq 1$  e  $\sum_{j=1}^n h_{ij} = 1$ . Além disso,

$$\text{tr}(\mathbf{H}) = \text{tr}[\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] = \text{tr}[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X}] = \text{tr}(\mathbf{I}_p) = \sum_{i=1}^n h_{ii} = p,$$

e, então, o valor médio de  $h_{ii}$  é  $p/n$ .

No processo de ajuste, como  $\hat{\boldsymbol{\mu}} = \mathbf{H}\mathbf{y}$ , tem-se

$$\hat{\mu}_i = \sum_{j=1}^n h_{ij} y_j = h_{i1} y_1 + \dots + h_{ii} y_i + \dots + h_{in} y_n \text{ com } 1 \leq i \leq n.$$

Verifica-se, portanto, que o valor ajustado  $\hat{\mu}_i$  é uma média ponderada dos valores observados e que o peso de ponderação é o valor de  $h_{ij}$ . Assim, o elemento da diagonal de  $\mathbf{H}$  é o peso com que a observação  $y_i$  participa do processo de obtenção do valor ajustado  $\hat{\mu}_i$ . Valores de  $h_{ii} \geq 2p/n$  indicam observações que merecem uma análise mais apurada (Belsley et al., 1980, p. 17).

## b) DFBeta e DFBetaS

Essas estatísticas são importantes quando o coeficiente de regressão tem um significado prático. A estatística  $\text{DFBeta}_{(i)}$  mede a alteração no vetor estimado  $\hat{\boldsymbol{\beta}}$  ao

se retirar a  $i$ -ésima observação da análise, isto é,

$$\text{DFBeta}_{(i)} = \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)} = \frac{r_i}{(1 - h_{ii})} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i,$$

ou ainda, considerando que  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = \mathbf{C} \mathbf{y}$ , em que  $\mathbf{C} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$  é uma matriz  $p \times n$ , tem-se

$$\text{DFBeta}_{(i)} = \frac{r_i}{(1 - h_{ii})} \mathbf{c}_i^T, \quad i = 1, \dots, n,$$

sendo  $\mathbf{c}_i^T$  a  $i$ -ésima linha de  $\mathbf{C}$ . Então,

$$\text{DFBeta}_{j(i)} = \frac{r_i}{(1 - h_{ii})} c_{ji}, \quad i = 1, \dots, n, \quad j = 0, \dots, p - 1.$$

Cook e Weisberg (1982) propuseram curvas empíricas para o estudo dessa medida. Como  $\text{Cov}(\hat{\boldsymbol{\beta}}) = \mathbf{C} \text{Var}(\mathbf{Y}) \mathbf{C}^T$ , a versão estudentizada de  $\text{DFBeta}_{j(i)}$  reduz-se a

$$\text{DFBetaS}_{j(i)} = \frac{c_{ji}}{(\sum c_{ji}^2) s_{(i)}} \frac{r_i}{(1 - h_{ii})}.$$

### c) **DFFit** e **DFFitS**

A estatística DFFit e sua versão estudentizada DFFitS medem a alteração decorrente no valor ajustado pela eliminação da observação  $i$ . São expressas como

$$\text{DFFit}_{(i)} = \mathbf{x}_i^T (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}) = \hat{\mu}_i - \hat{\mu}_{(i)}$$

e

$$\text{DFFitS}_{(i)} = \frac{\text{DFFit}_{(i)}}{\sqrt{h_{ii} s_{(i)}^2}} = \frac{\mathbf{x}_i^T (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})}{\sqrt{h_{ii} s_{(i)}^2}} = \frac{1}{\sqrt{h_{ii} s_{(i)}^2}} \frac{r_i}{(1 - h_{ii})} \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i$$

ou, ainda,

$$\text{DFFitS}_{(i)} = \left( \frac{h_{ii}}{1 - h_{ii}} \right)^{\frac{1}{2}} \frac{r_i}{s_{(i)} (1 - h_{ii})^{\frac{1}{2}}} = \left( \frac{h_{ii}}{1 - h_{ii}} \right)^{\frac{1}{2}} \text{rse}_{(i)},$$

sendo o quociente  $h_{ii}/(1 - h_{ii})$ , denominado potencial de influência, uma medida da distância do ponto  $\mathbf{x}_i$  em relação às demais observações. Nota-se que DFFitS pode

ser grande quando  $h_{ii}$  é grande ou quando o resíduo estudentizado externamente é grande. Valores absolutos excedendo  $2\sqrt{p/n}$  podem identificar observações influentes (Belsley et al., 1980, p. 28).

#### d) Distância de Cook

Uma medida de afastamento do vetor de estimativas resultante da eliminação da observação  $i$  é a distância de Cook. Tem uma expressão muito semelhante ao DFFitS mas que usa como estimativa da variância residual aquela obtida com todas as  $n$  observações, ou ainda, considera o resíduo estudentizado internamente. É expressa por

$$D_{(i)} = \frac{(\hat{\beta} - \hat{\beta}_{(i)})^T (\mathbf{X}^T \mathbf{X}) (\hat{\beta} - \hat{\beta}_{(i)})}{ps^2} = \frac{h_{ii}}{(1 - h_{ii})^2} \frac{r_i^2}{ps^2} = \left[ \frac{r_i}{(1 - h_{ii})^{\frac{1}{2}} s} \right]^2 \frac{h_{ii}}{p(1 - h_{ii})}$$

ou, ainda,

$$D_{(i)} = \frac{h_{ii} \text{rsi}_i^2}{p(1 - h_{ii})}.$$

#### e) Distância de Cook modificada

Atkinson (1981, p.25) sugere uma modificação para a distância de Cook

$$C_{(i)} = \left[ \frac{(n - p)}{p} \frac{h_{ii}}{(1 - h_{ii})} \right]^{\frac{1}{2}} |\text{rse}_{(i)}| = \left( \frac{n - p}{p} \right)^{\frac{1}{2}} \text{DFFitS}_{(i)}.$$

### 5.3.3 Tipos de gráficos

Dentre os gráficos mais usados para auxiliar na análise de resíduos e diagnósticos para os modelos lineares clássicos, destacam-se a) **Valores observados ( $y$ ) versus variáveis explanatórias ( $x_j$ )**

Esse tipo de gráfico indica a relação que pode existir entre a variável dependente e as diversas variáveis explanatórias e, também, pode mostrar a presença de heterocedasticidade. Pode, porém, conduzir a uma ideia falsa no caso de muitas variáveis explanatórias (a não ser que haja ortogonalidade entre todas).

**b) Variável explanatória  $x_j$  versus variável explanatória  $x_{j'}$** 

Esse tipo de gráfico pode indicar a estrutura que pode existir entre duas variáveis explanatórias e, também, pode mostrar a presença de heterocedasticidade. Pode, porém, conduzir a uma ideia falsa no caso de muitas variáveis explanatórias (a não ser que haja ortogonalidade entre todas).

**c) Resíduos versus variáveis explanatórias não incluídas ( $x_{fora}$ )**

Pode revelar se existe uma relação entre os resíduos do modelo ajustado e uma variável ainda não incluída no modelo e, também, pode mostrar a presença de heterocedasticidade. Pode, porém, conduzir a uma ideia falsa no caso de muitas variáveis explanatórias (a não ser que haja ortogonalidade entre todas). Uma alternativa melhor é o gráfico da variável adicionada (*added variable plot*).

**d) Resíduos versus variáveis explanatórias incluídas ( $x_{dentro}$ )**

Pode mostrar se ainda existe uma relação sistemática entre os resíduos e a variável  $x_j$  que está incluída no modelo, isto é, por exemplo se  $x_{dentro}^2$  deve ser incluída. O padrão para esse tipo de gráfico é uma distribuição aleatória de média zero e amplitude constante. Desvios sistemáticos podem indicar:

- escolha errada da variável explanatória,
- falta de termo quadrático (ou de ordem superior),
- escala errada da variável explanatória.

Pode, também, conduzir a uma ideia falsa no caso de muitas variáveis explanatórias (a não ser que haja ortogonalidade entre todas). Uma alternativa melhor é o gráfico dos resíduos parciais (*partial residual plot*).

**e) Resíduos versus valores ajustados**

O padrão para esse tipo de gráfico é uma distribuição aleatória de média zero e amplitude constante. Pode mostrar heterogeneidade de variâncias e pontos discrepantes.

f) **Gráficos de índices**

Consistem em se plotarem medidas de diagnósticos  $d_i$  versus  $i$ . Servem para localizar observações com resíduos,  $h_{ii}$  (*leverage*), distâncias de Cook modificadas, etc, grandes.

g) **Gráfico da variável adicionada ou da regressão parcial** (*added variable plot*)

Embora os gráficos dos resíduos *versus* variáveis não incluídas no modelo possam indicar a necessidade de variáveis extras no modelo, a interpretação exata deles não é clara. A dificuldade reside em que, a menos que a variável explanatória, considerada para inclusão, seja ortogonal a todas as variáveis que estão incluídas no modelo, o coeficiente angular do gráfico dos resíduos não é o mesmo que o coeficiente angular no modelo ajustado, incluindo a variável em questão.

Esse tipo de gráfico pode ser usado para detectar a relação de  $\mathbf{y}$  com uma variável explanatória  $\mathbf{u}$ , ainda não incluída no modelo, livre do efeito de outras variáveis, e como isso é influenciado por observações individuais. Note que  $\mathbf{u}$  pode ser, também, uma variável construída para verificar a necessidade de uma transformação para a variável resposta e/ou para as variáveis explanatórias. No caso do modelo linear geral, tem-se

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta} + \gamma\mathbf{u},$$

sendo  $\mathbf{u}$  uma variável a ser adicionada e  $\gamma$ , o parâmetro escalar adicional. O interesse está em se saber se  $\gamma = 0$ , isto é, se não há necessidade de se incluir a variável  $\mathbf{u}$  no modelo. A partir do sistema de equações normais,  $\mathbf{X}^{*T}\mathbf{X}^*\hat{\boldsymbol{\beta}} = \mathbf{X}^{*T}\mathbf{y}$ , em que  $\mathbf{X}^* = [\mathbf{X} \ \mathbf{u}]$  e  $\hat{\boldsymbol{\beta}} = [\hat{\boldsymbol{\beta}} \ \hat{\gamma}]^T$ , tem-se

$$\begin{bmatrix} \mathbf{X}^T \\ \mathbf{u}^T \end{bmatrix} \begin{bmatrix} \mathbf{X} & \mathbf{u} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{u}^T\mathbf{y} \end{bmatrix} \Rightarrow \begin{cases} \mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}^T\mathbf{u}\hat{\gamma} = \mathbf{X}^T\mathbf{y} \\ \mathbf{u}^T\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{u}^T\mathbf{u}\hat{\gamma} = \mathbf{u}^T\mathbf{y} \end{cases}$$

e, portanto,

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



e

$$\hat{\gamma} = \frac{\mathbf{u}^T(\mathbf{I} - \mathbf{H})\mathbf{y}}{\mathbf{u}^T(\mathbf{I} - \mathbf{H})\mathbf{u}} = \frac{\mathbf{u}^T(\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H})\mathbf{y}}{\mathbf{u}^T(\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H})\mathbf{u}} = \frac{\mathbf{u}^{*T}\mathbf{r}}{\mathbf{u}^{*T}\mathbf{u}^*},$$

que é o coeficiente angular de uma reta que passa pela origem, sendo  $\mathbf{r} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{I} - \mathbf{H})\mathbf{y}$  o vetor dos resíduos de  $\mathbf{y}$  ajustado para  $\mathbf{X}$ ,  $\mathbf{u}^* = (\mathbf{I} - \mathbf{H})\mathbf{u}$  o vetor dos resíduos de  $\mathbf{u}$  ajustado para  $\mathbf{X}$  e  $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ .

O gráfico da variável adicionada de  $\mathbf{r}$  *versus*  $\mathbf{u}^*$ , portanto, tem coeficiente angular  $\hat{\gamma}$  (diferente do gráfico de  $\mathbf{r}$  *versus*  $\mathbf{u}$ ) e é calculado a partir dos resíduos ordinários da regressão de  $\mathbf{y}$  como função de todas as variáveis explanatórias, exceto  $\mathbf{u} = x_j$ , *versus* os resíduos ordinários da regressão de  $\mathbf{u} = x_j$  como função das mesmas variáveis explanatórias usadas para analisar  $\mathbf{y}$ . Assim, por exemplo, para um modelo com três variáveis explanatórias, o gráfico da variável adicionada para  $x_3$  é obtido a partir de duas regressões lineares

$$\hat{\boldsymbol{\mu}} = \hat{\beta}_0 + \hat{\beta}_1\mathbf{x}_1 + \hat{\beta}_2\mathbf{x}_2 \Rightarrow \mathbf{r} = \mathbf{y} - \hat{\boldsymbol{\mu}}$$

e

$$\hat{\mathbf{x}}_3 = \hat{\beta}'_0 + \hat{\beta}'_1\mathbf{x}_1 + \hat{\beta}'_2\mathbf{x}_2 \Rightarrow \mathbf{u}^* = \mathbf{x}_3 - \hat{\mathbf{x}}_3.$$

O padrão nulo do gráfico de  $\mathbf{r}$  *versus*  $\mathbf{u}^*$  indicará a não necessidade de inclusão da variável  $\mathbf{u}$ .

#### h) Gráfico de resíduos parciais ou gráfico de resíduos mais componente (*partial residual plot*)

Se se deseja detectar uma estrutura omitida, tal como uma forma diferente de dependência em  $\mathbf{u}$  (que poderá ser, por exemplo, um termo quadrático no modelo), um gráfico usando  $\mathbf{u}$  pode ser de maior interesse. Esse gráfico consiste em se plotarem os resíduos do modelo  $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta} + \gamma\mathbf{u}$  mais  $\hat{\gamma}\mathbf{u}$  (isto é, os resíduos aumentados  $\tilde{\mathbf{r}} = \mathbf{r} + \hat{\gamma}\mathbf{u}$ ) *versus*  $\mathbf{u}$ . O padrão nulo desse gráfico indicará a não necessidade de inclusão da variável  $\mathbf{u}$ .

i) **Gráficos normal e semi-normal de probabilidades** (*normal plots e half normal plots*)

O gráfico normal de probabilidades destaca-se por dois aspectos (Weisberg, 2005):

- identificação da distribuição originária dos dados e
- identificação de valores que se destacam no conjunto.

Seja uma amostra aleatória de tamanho  $n$ . As estatísticas de ordem correspondentes aos resíduos padronizados obtidos a partir do ajuste de um determinado modelo são  $d_{(1)}, \dots, d_{(i)}, \dots, d_{(n)}$ . O fundamento geral para a construção do gráfico normal de probabilidades é que se os valores de uma dada amostra provêm de uma distribuição normal, então os valores das estatísticas de ordem e os  $z_i$  correspondentes, obtidos da distribuição normal padrão, são linearmente relacionados. Portanto, o gráfico de  $d_{(i)}$  versus  $z_i$  deve ser, aproximadamente, uma reta. Formatos aproximados comuns que indicam ausência de normalidade são:

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

Esses gráficos, na realidade, são muito dependentes do número de observações, atingindo a estabilidade quando o número de observações é grande (em torno de 300). Para a construção desse gráfico, seguem-se as etapas:

a) ajuste um determinado modelo a um conjunto de dados e obtenha  $d_{(i)}$ , os valores ordenados de uma certa estatística de diagnóstico (resíduos, distância de Cook,  $h_{ii}$ , etc);

b) a partir da estatística de ordem na posição  $(i)$ , calcule a respectiva probabilidade acumulada  $p_i$  e o respectivo quantil, ou seja, o inverso da função de distribuição normal  $\Phi(\cdot)$  no ponto  $p_i$ . Essa probabilidade  $p_i$  é, em geral, aproximada

por

$$p_i = \frac{i - c}{n - 2c + 1}$$

sendo  $0 < c < 1$ . Diversos valores têm sido propostos para a constante  $c$ . Vários autores recomendam a utilização de  $c = 3/8$ , ficando, então,

$$z_i = \Phi^{-1} \left( \frac{i - 0,375}{n + 0,25} \right), \text{ para } i = 1, \dots, n.$$

c) coloque, em um gráfico,  $d_{(i)}$  versus  $z_i$ .

Esse gráfico tem, também, o nome de *Q-Q plot*, por relacionar os valores de um quantil amostral ( $d_{(i)}$ ) versus os valores do quantil correspondente da distribuição normal ( $z_i$ ). Se forem plotados os valores  $|d_{(i)}|$  versus  $z_i$ , em que  $z_i = \Phi^{-1}[(i + n - 0,125)/(2n + 0,5)]$ , tem-se o gráfico semi-normal de probabilidades

McCullagh e Nelder (1989) sugerem o uso do gráfico normal de probabilidades para os resíduos e o gráfico semi-normal de probabilidades para medidas positivas como é o caso de  $h_{ii}$  e da distância de Cook modificada. No caso do gráfico normal de probabilidades para os resíduos, espera-se que, na ausência de pontos discrepantes, o aspecto seja linear mas não há razão para se esperar que o mesmo ocorra quando são usados  $h_{ii}$  ou a distância de Cook modificada. Os valores extremos aparecerão nos extremos do gráfico, possivelmente com valores que desviam da tendência indicada pelos demais.

Para auxiliar na interpretação do gráfico semi-normal de probabilidades, Atkinson (1985) propôs a adição de um envelope simulado, que é obtido, seguindo-se as etapas:

a) ajuste um determinado modelo a um conjunto de dados e obtenha  $d_{(i)}$ , os valores absolutos ordenados de uma certa estatística de diagnóstico (resíduos, distância de Cook,  $h_{ii}$ , etc);

b) simule 19 amostras da variável resposta, usando as estimativas obtidas após um determinado modelo ser ajustado aos dados e os mesmos valores para as variáveis explanatórias;

- c) ajuste o mesmo modelo a cada uma das 19 amostras e calcule os valores absolutos ordenados da estatística de diagnóstico de interesse,  $d_{j(i)}^*$ ,  $j = 1, \dots, 19$ ,  $i = 1, \dots, n$ ;
- d) para cada  $i$ , calcule a média, o mínimo e o máximo dos  $d_{j(i)}^*$ ;
- e) coloque as quantidades determinadas no item anterior em um gráfico de  $d_{(i)}$  versus  $z_i$ .

Esse envelope é tal que, sob o modelo correto, as estatísticas (resíduos, *leverage*, distância de Cook, etc) obtidas a partir das observações ficam inseridas no envelope. Moral et al. (2015) apresentam um pacote com funções que possibilitam a construção desses gráficos para uma grande variedade de modelos, usando o software R. Aplicações desse pacote a diversos conjuntos de dados entomológicos podem ser encontradas em Demétrio et al. (2014).

#### j) **Valores observados ( $y$ ) ou resíduos versus tempo**

Mesmo que o tempo não seja uma variável incluída no modelo, gráficos de respostas ( $y$ ) ou de resíduos versus tempo devem ser apresentados sempre que possível. Esse tipo de gráfico pode conduzir à detecção de padrões não revelados, devido ao tempo ou, então, a alguma variável muito correlacionada com o tempo.

## 5.4 Análise de resíduos e diagnóstico para os MLGs

As técnicas usadas para análise de resíduos e diagnóstico para os MLGs são semelhantes às aquelas usadas para o modelo clássico de regressão, com algumas adaptações. Assim, por exemplo, na verificação da pressuposição de linearidade para o modelo clássico de regressão, usam-se os vetores  $\mathbf{y}$  e  $\hat{\boldsymbol{\mu}}$ , enquanto que para o MLG devem ser usados  $\hat{\mathbf{z}}$ , a variável dependente ajustada estimada, e  $\hat{\boldsymbol{\eta}}$ , o preditor linear estimado. A variância residual  $s^2$  é substituída por uma estimativa consistente do

parâmetro de dispersão  $\phi$  e a matriz de projeção  $\mathbf{H}$  é definida por

$$\mathbf{H} = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2}, \quad (5.1)$$

o que é equivalente a substituir  $\mathbf{X}$  por  $\mathbf{W}^{1/2} \mathbf{X}$ . Nota-se que  $\mathbf{H}$ , agora, depende das variáveis explanatórias, da função de ligação e da função de variância, tornando mais difícil a interpretação da medida de *leverage*. Demonstra-se que

$$\mathbf{V}^{-1/2}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \cong \mathbf{H} \mathbf{V}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu}), \quad (5.2)$$

sendo  $\mathbf{V} = \text{diag}\{V(\mu_i)\}$ . A equação (5.2) indica que  $\mathbf{H}$  mede a influência em unidades estudentizadas de  $\mathbf{y}$  sobre  $\hat{\boldsymbol{\mu}}$  (McCullagh e Nelder, 1989, p. 397).

A matriz  $\mathbf{H}$  em (5.1) desempenha um papel importante na análise dos resíduos nos MLGs e tem as propriedades  $\text{tr}(\mathbf{H}) = p$  e  $0 \leq h_{ii} \leq 1$ , descritas na Seção 5.3.2 no contexto do modelo clássico de regressão. Outra matriz importante de projeção é definida como  $\mathbf{I} - \mathbf{H} = \mathbf{I} - \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2} = \mathbf{I} - \mathbf{W}^{1/2} \mathbf{Z} \mathbf{W}^{1/2}$ , em que  $\mathbf{Z} = \{z_{ij}\} = \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T$  é a matriz de covariância assintótica de  $\hat{\boldsymbol{\eta}}$  (ver Seção 3.4), sendo  $\mathbf{W} = \text{diag}\{g'(\mu_i)^{-2} V(\mu_i)^{-1}\} = \mathbf{V}^{-1} \mathbf{G}^{-2}$ . Essa aproximação está correta até termos de ordem  $O(n^{-1})$  (Cordeiro, 2004b).

### 5.4.1 Tipos de resíduos

Em geral, a escolha do tipo de resíduo a ser usado depende, basicamente, do tipo de anomalia que se deseja detectar no modelo. O resíduo  $r_i$  deve expressar uma discrepância (distância) entre a observação  $y_i$  e o seu valor ajustado  $\hat{\mu}_i$ , isto é,

$$r_i = h_i(y_i, \hat{\mu}_i), \quad (5.3)$$

em que  $h_i$  é uma função adequada de fácil interpretação, usualmente escolhida para estabilizar a variância e/ou induzir simetria na distribuição amostral de  $R_i$  (Cox e Snell, 1968). Em geral, usa-se a mesma função  $h_i(\cdot) = h(\cdot)$  para todas as observações.

Na expressão geral dos resíduos  $r_i = h_i(y_i, \hat{\mu}_i)$ , em (5.3), a função  $h_i$  deve ser escolhida visando a satisfazer as propriedades de segunda ordem:  $E(R_i) = 0$ ,

$\text{Var}(R_i) = \text{constante}$  e  $\text{Cov}(R_i, R_j) = 0$ ,  $i \neq j$ , pois, em muitos casos, essas condições são suficientes para especificar a forma da distribuição de  $R_i$ . Cox e Snell (1968) apresentam fórmulas gerais para  $E(R_i)$ ,  $\text{Cov}(R_i, R_j)$  e  $\text{Var}(R_i)$  até termos de ordem  $n^{-1}$ , válidas para qualquer função  $h_i$  especificada. Essas fórmulas possibilitam calcular resíduos modificados cujas distribuições são melhor aproximadas pelas distribuições de probabilidade de referência.

O número de observações para estimar os parâmetros do modelo e proporcionar informações sobre a distribuição de probabilidade dos resíduos deve ser grande. Frequentemente, o número de parâmetros,  $p$ , é pequeno comparado com o número de observações,  $n$ , e as combinações de parâmetros são estimadas com erro padrão de ordem  $n^{-1/2}$ . Nesse caso, o resíduo  $R_i$  difere do **erro verdadeiro**  $\epsilon_i = h_i(y_i, \mu_i)$  por uma quantidade de ordem  $n^{-1/2}$  em probabilidade, e muitas propriedades estatísticas dos  $R_i$  são equivalentes às propriedades respectivas dos  $\epsilon_i$  (Cordeiro, 2004b).

Em geral, a distribuição exata de  $R_i$  não é conhecida e trabalha-se com resultados assintóticos, tais como, valor esperado  $E(R_i)$  e variância  $\text{Var}(R_i)$  até ordem  $n^{-1}$ . Cordeiro (2004b) baseia-se nos resultados de Cox e Snell (1968) para calcular expressões matriciais aproximadas, até ordem  $n^{-1}$ , para os valores esperados, variâncias e covariâncias dos resíduos de Pearson, válidas para qualquer MLG. Essas expressões dependem das funções de ligação e de variância e de suas duas primeiras derivadas. Theil (1965) sugere usar uma combinação linear dos resíduos  $\mathbf{R}^* = \mathbf{C}\mathbf{R}$ , no lugar de  $\mathbf{R} = (R_1, \dots, R_n)^T$ , para testes e gráficos, em que  $\mathbf{C}$  é uma matriz  $(n-p) \times n$ , escolhida de modo que  $\mathbf{R}^*$  tenha, aproximadamente, distribuição normal multivariada  $N(\mathbf{0}, \sigma^2 \mathbf{I})$ .

Em geral, a definição da função  $h_i$  depende, basicamente, do tipo de anomalia que se deseja detectar no modelo. As escolhas mais comuns de  $h_i$  são  $r_i = (y_i - \hat{\mu}_i)/[\text{Var}(Y_i)]^{1/2}$  e  $r_i = (y_i - \hat{\mu}_i)/[\text{Var}(Y_i - \hat{\mu}_i)]^{1/2}$ , a primeira forma sendo a mais usual, em que as expressões da variância nos denominadores são estimadas segundo o modelo sob pesquisa. Escolhendo  $h_i$ , adequadamente, essas anomalias podem ser encontradas, usando-se gráficos como descritos em 5.3.3. Apresentam-se,

a seguir, os tipos de resíduos mais comuns para os MLGs.

### a) Resíduos ordinários

Embora esses resíduos não sejam de interesse direto para os MLGs, entram na definição de outros resíduos. São definidos por

$$r_i = y_i - \hat{\mu}_i.$$

### b) Resíduos de Pearson

O resíduo de Pearson, definido por

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\hat{V}_i^{1/2}}, \quad (5.4)$$

que é um componente da estatística de Pearson generalizada  $X_p^2 = \sum_{i=1}^n r_i^{P^2}$  especificada em (4.8). Podem-se incorporar pesos a priori na fórmula (5.4). Para o modelo linear clássico, o resíduo de Pearson coincide com o resíduo ordinário.

Além disso, a partir da expressão (4.9), o vetor the resíduos de Pearson pode ser expresso, matriciamente, por

$$\mathbf{r}_p = \widehat{\mathbf{V}}^{-1/2}(\mathbf{y} - \hat{\boldsymbol{\mu}}) = \widehat{\mathbf{W}}^{1/2}(\hat{\mathbf{z}} - \hat{\boldsymbol{\eta}}) \quad (5.5)$$

em que  $\mathbf{V} = \text{diag}\{V_1, \dots, V_n\}$ ,  $\mathbf{W} = \text{diag}\{g'(\mu_i)^{-2}V_i^{-1}\}$  e  $\hat{\mathbf{z}} = \hat{\boldsymbol{\eta}} + \widehat{\mathbf{G}}(\mathbf{y} - \hat{\boldsymbol{\mu}})$  com  $\mathbf{G} = \text{diag}\{g'(\mu_i)\}$ .

A desvantagem do resíduo de Pearson é que sua distribuição é, geralmente, bastante assimétrica para modelos não-normais. Para os modelos log-lineares tem-se que  $r_i^P = (y_i - \hat{\mu}_i)\hat{\mu}_i^{-1/2}$  e Haberman (1974) sugere a correção  $r_i^{P*} = r_i^P / (1 - \hat{\mu}_i \hat{z}_{ii})^{1/2}$ , em que  $\mathbf{Z} = \{z_{ij}\} = \mathbf{X}(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T$  com  $\mathbf{W} = \text{diag}\{\mu_i\}$ , para tornar a distribuição do resíduo  $R_i^{P*}$ , aproximadamente, normal  $N(0, 1)$ . A variância média dos  $R_i^{P*}$  é  $1 - (1 + p)/n$ .

Cordeiro (2004b) apresenta expressões para a média e a variância de  $R_i^P$  válidas até ordem  $n^{-1}$ . Partindo do resultado do algoritmo de estimação da Seção 3.2, a expressão (3.5) avaliada na EMV  $\hat{\beta}$  implica em

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

sendo  $\hat{\mathbf{z}} = \hat{\boldsymbol{\eta}} + \widehat{\mathbf{G}}(\mathbf{y} - \hat{\boldsymbol{\mu}})$ . Logo, usando-se a definição da matriz  $\mathbf{Z}$ , tem-se

$$\hat{\mathbf{z}} - \hat{\boldsymbol{\eta}} = (\mathbf{I} - \widehat{\mathbf{Z}} \widehat{\mathbf{W}}) \hat{\mathbf{z}}.$$

Supondo que  $\mathbf{Z}$  e  $\mathbf{W}$  são tais que, aproximadamente, pelo menos o produto  $\widehat{\mathbf{Z}} \widehat{\mathbf{W}}$  é constante, pode-se escrever

$$\text{Cov}(\hat{\mathbf{z}} - \hat{\boldsymbol{\eta}}) \approx (\mathbf{I} - \mathbf{Z} \mathbf{W}) \text{Cov}(\hat{\mathbf{z}}) (\mathbf{I} - \mathbf{Z} \mathbf{W})^T$$

e como  $\text{Cov}(\hat{\mathbf{z}}) = \phi \mathbf{W}^{-1}$ , conforme equação (3.6) da Seção 3.2, tem-se

$$\text{Cov}(\hat{\mathbf{z}} - \hat{\boldsymbol{\eta}}) \approx \phi \mathbf{W}^{-1/2} (\mathbf{I} - \mathbf{H}) \mathbf{W}^{-1/2}.$$

Logo,

$$\text{Cov}(\mathbf{r}_p) = \text{Cov}[\widehat{\mathbf{W}}^{1/2}(\hat{\mathbf{z}} - \hat{\boldsymbol{\eta}})] \approx \phi \mathbf{H}_2, \quad (5.6)$$

em que  $\mathbf{H}_2 = \mathbf{I} - \mathbf{W}^{1/2} \mathbf{Z} \mathbf{W}^{1/2}$ .

Convém enfatizar que Cordeiro (2004b) mostra que a expressão de  $\text{Cov}[\widehat{\mathbf{W}}^{1/2}(\hat{\mathbf{z}} - \hat{\boldsymbol{\eta}})]$  está correta até ordem  $n^{-1}$  para os elementos fora da diagonal, mas não para os elementos da diagonal.

### c) Resíduos de Anscombe

Anscombe (1953) apresenta uma definição geral de resíduos, usando uma transformação  $N(y_i)$  da observação  $y_i$ , escolhida visando tornar a sua distribuição o mais próxima possível da distribuição normal. Barndorff-Nielsen (1978) demonstra que, para os MLGs,  $N(\cdot)$  é calculada por  $N(\mu) = \int V^{-1/3} d\mu$ . Como  $N'(\mu)(V/\phi)^{1/2}$  é



a aproximação de primeira ordem do desvio padrão de  $N(y)$ , o resíduo de Anscombe, visando à normalização e à estabilização da variância, é expresso por

$$A_i = \frac{N(y_i) - N(\hat{\mu}_i)}{N'(\hat{\mu}_i)\hat{V}_i^{1/2}}, \quad (5.7)$$

que para os modelos de Poisson, binomial, gama e normal inverso assumem as expressões apresentados na Tabela 5.1. Cox e Snell (1968) expressam esse resíduo para a distribuição binomial, usando a função beta incompleta. Da definição do resíduo de Anscombe, conclui-se que a transformação aplicada aos dados para normalizar os resíduos é a mesma que, aplicada às médias das observações, normaliza a distribuição de  $\hat{\beta}$  (vide equação (??), caso  $\delta = 2/3$ ).

Tabela 5.1: Expressões para os resíduos de Pearson ( $r^P$ ), de Anscombe ( $A$ ) e o definido como a raiz quadrada do componente do desvio ( $r^D$ ) para algumas distribuições, em que  $N(\mu) = \int [\mu(m - \mu)]^{-1/3} d\mu$  e  $\delta$  representa o sinal de  $(y - \hat{\mu})$

Distribuições	$r^P$	$A$	$r^D$
Poisson	$\frac{y - \hat{\mu}}{\mu}$	$3 \frac{y^{2/3} - \hat{\mu}^{2/3}}{2\hat{\mu}^{1/6}}$	$\delta \left[ 2 \left( y \log \frac{y}{\hat{\mu}} + \hat{\mu} - y \right) \right]^{1/2}$
Binomial	$\frac{y - \hat{\mu}}{m\hat{\pi}(1 - \hat{\pi})}$	$m^{1/2} \frac{N(y) - N(\hat{\mu})}{[\hat{\mu}(m - \hat{\mu})]^{1/6}}$	$\delta \left\{ 2m \left[ y \log \frac{y}{\hat{\mu}} + (1 - y) \log \frac{1 - y}{1 - \hat{\mu}} \right] \right\}^{1/2}$
Gama	$\frac{y - \hat{\mu}}{\mu^2}$	$3 \frac{y^{1/3} - \hat{\mu}^{1/3}}{\hat{\mu}^{1/3}}$	$\delta \left\{ 2 \left[ \log \frac{\hat{\mu}}{y} + \frac{y - \hat{\mu}}{\hat{\mu}} \right] \right\}^{1/2}$
Normal inverso	$\frac{y - \hat{\mu}}{\mu^3}$	$\frac{\log y - \log \hat{\mu}}{\hat{\mu}^{1/2}}$	$\frac{y - \hat{\mu}}{y^{1/2}\hat{\mu}}$

#### d) Resíduos de Pearson estudentizados internamente

A partir da expressão (5.6) tem-se que  $\text{Var}(r_i^P) \approx \phi(1 - \hat{h}_{ii})$ . Logo, os resíduos de Pearson estudentizados internamente são definidos por

$$r_i^{P'} = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\phi}V(\hat{\mu}_i)(1 - \hat{h}_{ii})}}, \quad (5.8)$$

sendo  $h_{ii}$  o  $i$ -ésimo elemento da diagonal da matriz definida em (5.1). Os resíduos  $r_i^{P'}$  definidos pela expressão (5.8) apresentam propriedades razoáveis de segunda

ordem mas podem ter distribuições bem diferentes da distribuição normal; têm, aproximadamente, variância igual a um quando o parâmetro de dispersão  $\phi \rightarrow 0$ .

#### e) Componentes do desvio

Outro tipo de resíduos usados para os MLGs são as raízes quadradas dos componentes do desvio com o sinal igual ao sinal de  $y_i - \hat{\mu}_i$ , isto é,

$$r_i^D = \text{sinal}(y_i - \hat{\mu}_i) \sqrt{2[v(y_i) - v(\hat{\mu}_i) + q(\hat{\mu}_i)(\hat{\mu}_i - y_i)]^{1/2}}, \quad (5.9)$$

em que a função  $v(x) = xq(x) - b(q(x))$  é expressa em termos das funções  $b(\cdot)$  e  $q(\cdot)$  definidas na Seção 1.3. Para os modelos de Poisson, binomial, gama e normal inverso, os resíduos definidos como as raízes quadradas dos componentes do desvio estão apresentados na Tabela 5.1.

O resíduo  $r_i^D$  representa uma distância da observação  $y_i$  ao seu valor ajustado  $\hat{\mu}_i$ , medida na escala do logaritmo da função de verossimilhança. Tem-se  $D_p = \sum_{i=1}^n r_i^{D^2}$ . Um valor grande para  $r_i^D$  indica que a  $i$ -ésima observação está mal ajustada pelo modelo. Pregibon (1979) demonstra que, se existe uma transformação  $h_i$  que normaliza a distribuição do resíduo  $R_i = h_i(y_i, \hat{\mu}_i)$ , então as raízes quadradas dos componentes do desvio são resíduos que exibem as mesmas propriedades induzidas por essa transformação. Assim, os resíduos  $r_i^D$  podem ser considerados, aproximadamente, como variáveis aleatórias normais reduzidas e, conseqüentemente,  $r_i^{D^2}$  como tendo, aproximadamente, uma distribuição  $\chi_1^2$ .

As vantagens dos resíduos (5.9) são: a) não requerem o conhecimento da função normalizadora; b) computação simples após o ajuste do MLG; c) são definidos para todas as observações, até mesmo para observações censuradas, desde que contribuam para o logaritmo da função de verossimilhança.

#### f) Componentes do desvio estudentizados internamente

Pregibon (1979) sugere trabalhar com as raízes quadradas dos componentes do desvio estudentizadas internamente, isto é, resíduos definidos a partir da equação

(5.9) e escritos como

$$r_i^{D'} = \frac{r_i^D}{\sqrt{\hat{\phi}(1 - \hat{h}_{ii})}}. \quad (5.10)$$

Entretanto, a adequação da distribuição normal para aproximar a distribuição de  $r_i^{D'}$  ainda continua sendo um problema a ser pesquisado.

g) **Resíduos estudentizados externamente** (*deletion residuals, jackknifed residuals*)

A obtenção de resíduos estudentizados externamente pode ser computacionalmente intensiva e muito demorada. Em função disso, foram propostas aproximações usando-se apenas a 1ª iteração do ajuste de um determinado modelo de interesse (Pregibon, 1981; Williams, 1987; McCullagh e Nelder, 1989). Isso significa substituir a expressão (5.8) por

$${}_1r_{(i)}^{P'} = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\phi}_{(i)}V(\hat{\mu}_i)(1 - \hat{h}_{ii})}}, \quad (5.11)$$

o que mede a mudança na estatística de Pearson generalizada  $X_p^2$  dividida por  $\phi$ , pela omissão do ponto  $i$ . A aproximação análoga para a mudança no desvio foi obtida por Williams (1987)

$${}_1r_{(i)}^J = \text{sign}(y_i - \hat{\mu}_i) \sqrt{(1 - h_{ii})({}_1r_i^D)^2 + h_{ii}({}_1r_i^P)^2}$$

em que  ${}_1r_{(i)}^{P'}$  é dado pela expressão (5.11) e

$${}_1r_{(i)}^{D'} = \frac{r_i^D}{\sqrt{\hat{\phi}_{(i)}(1 - \hat{h}_{ii})}}.$$

Os resíduos de Pearson, de Anscombe e componentes do desvio, expressos em (5.4), (5.7) e (5.9), respectivamente, são os mais importantes nas aplicações

dos MLGs. No modelo normal, nenhuma distinção é feita entre esses três tipos de resíduos. Para modelos bem ajustados, as diferenças entre  $r_i^D$  e  $r_i^P$  devem ser pequenas. Entretanto, para os modelos mal-ajustados e/ou para observações aberrantes, podem ocorrer diferenças consideráveis entre esses resíduos. Embora os resíduos, definidos por (5.7) e (5.9), apresentem formas bem diferentes para modelos não-normais, os seus valores, especificados  $y$  e  $\hat{\mu}$ , são similares. Admite-se que  $\hat{\mu} = cy$ , em que  $c$  é um real qualquer. Seja A/D o quociente entre o resíduo de Anscombe (A) e aquele definido como a raiz quadrada do componente do desvio (D). Para os modelos de Poisson, gama e normal inverso, esse quociente é igual a  $3\delta(1 - c^{2/3})/(2\sqrt{2})c^{1/6}(c - 1 - \log c)^{1/2}$ ,  $3\delta(1 - c^{1/3})c^{1/6}/\sqrt{2}(c \log c + 1 - c)^{1/2}$  e  $c^{1/2} \log c/(c - 1)$ , respectivamente, em que  $\delta = +1(-1)$  quando  $c < 1(> 1)$ . Na Tabela 5.2, são apresentados valores do quociente A/D para esses três modelos. Dessa tabela, conclui-se que esses dois resíduos são, aproximadamente, equivalentes. Essa equivalência poderia ainda ser determinada por expansões em série de Taylor. McCullagh e Nelder (1989) comparam os resíduos de Pearson, de Anscombe e aqueles definidos como a raiz quadrada dos componentes do desvio, para o modelo de Poisson.

Definindo-se uma distribuição teórica conveniente para os resíduos, podem-se aplicar as diversas técnicas analíticas e gráficas para detectar desvios do modelo sob pesquisa. Em geral, o resíduo definido como a raiz quadrada do componente do desvio, studentizado ou não, é melhor para verificação de ajustes de modelos pois tem propriedades melhores (Pierce e Schafer, 1986).

### 5.4.2 Tipos de gráficos

Os gráficos mais usados para auxiliar na análise de resíduos e diagnósticos para os MLGs são basicamente os mesmos apresentados na Seção 5.3.3 com pequenas modificações e com interpretações semelhantes.

#### a) Resíduos *versus* alguma função dos valores ajustados

É recomendado o gráfico de algum tipo de resíduo studentizado ( $r_i^{P'}$  ou

Tabela 5.2: Relação A/D entre o resíduo de Anscombe e o definido como a raiz quadrada do componente do desvio, para três modelos.

c	Poisson	gama	normal inverso
0,1	1,0314	0,9462	0,8090
0,2	1,0145	0,9741	0,8997
0,4	1,0043	0,9918	0,9658
0,6	1,0014	0,9977	0,9892
0,8	1,0010	0,9994	0,9979
2,0	1,0019	0,9958	0,9802
3,0	1,0048	0,9896	0,9514
5,0	1,0093	0,9790	0,8997
10,0	1,0169	0,9598	0,8090

$r_i^{D'}$ ) *versus*  $\hat{\eta}_i$ , ou então, *versus* os valores ajustados transformados de tal forma a se ter variância constante para a distribuição em uso (vide Seção ??). Assim, usar, no eixo das abscissas,  $\hat{\mu}_i$  para a distribuição normal,  $2\sqrt{\hat{\mu}_i}$  para a Poisson,  $2\arcsen\sqrt{\hat{\mu}_i/m_i}$  para a binomial,  $2\log(\hat{\mu}_i)$  para a gama e  $-2\hat{\mu}_i^{-1/2}$  para a normal inversa. O padrão nulo desse gráfico é uma distribuição dos resíduos em torno de zero com amplitude constante. Desvios sistemáticos podem apresentar algum tipo de curvatura ou, então, mudança sistemática da amplitude com o valor ajustado.

#### b) Resíduos *versus* variáveis explanatórias não incluídas

Esse gráfico pode mostrar se existe uma relação entre os resíduos do modelo ajustado e uma variável ainda não incluída no modelo. Uma alternativa melhor para esse tipo de gráfico é o gráfico da variável adicionada (*added variable plot*). O padrão nulo desse gráfico é uma distribuição dos resíduos em torno de zero com amplitude constante.

c) **Resíduos versus variáveis explanatórias já incluídas**

Esse gráfico pode mostrar se ainda existe uma relação sistemática entre os resíduos e uma variável que está incluída no modelo. Uma alternativa melhor é o gráfico de resíduos parciais (*partial residual plot*). O padrão nulo para esse tipo de gráfico é uma distribuição aleatória de média zero e amplitude constante.

d) **Gráfico da variável adicionada ou da regressão parcial** (*added variable plot*)

Para a obtenção do gráfico da variável adicionada, inicialmente, ajusta-se o modelo linear generalizado com preditor linear  $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ . Em seguida, faz-se o gráfico de  $\widehat{\mathbf{W}}^{-1/2}\mathbf{s}$  versus  $(\mathbf{I} - \widehat{\mathbf{H}})\widehat{\mathbf{W}}^{1/2}\mathbf{u}$ , sendo  $\mathbf{s}$  o vetor com elementos estimados por

$$s_i = \frac{(y_i - \hat{\mu}_i) \frac{d\mu_i}{d\eta_i}}{V(\hat{\mu}_i)}$$

e  $\mathbf{u}$  o vetor com os valores da variável a ser adicionada (Wang, 1985). Aqui  $\widehat{\mathbf{W}}^{-1/2}\mathbf{s}$  representa o vetor de elementos  $(y_i - \hat{\mu}_i)V(\hat{\mu}_i)^{-1/2}$  (resíduo de Pearson generalizado da regressão ponderada de  $\mathbf{y}$  em relação a  $\mathbf{X}$  com matriz de pesos estimada  $\widehat{\mathbf{W}}$ ) e  $(\mathbf{I} - \widehat{\mathbf{H}})\widehat{\mathbf{W}}^{1/2}\mathbf{u}$  representa os resíduos da regressão ponderada de  $\mathbf{u}$  em relação a  $\mathbf{X}$  com matriz de pesos estimada  $\widehat{\mathbf{W}}$ . O padrão nulo para esse tipo de gráfico é uma distribuição aleatória de média zero e amplitude constante.

e) **Gráfico de resíduos parciais ou gráfico de resíduos mais componente** (*partial residual plot*)

Para a obtenção do gráfico de resíduos parciais, inicialmente, ajusta-se o MLG com preditor linear  $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \gamma\mathbf{u}$ , obtendo-se  $\widehat{\mathbf{W}}^{-1}\mathbf{s}$  e  $\hat{\gamma}$ . Em seguida, faz-se o gráfico de  $\widehat{\mathbf{W}}^{-1}\mathbf{s} + \hat{\gamma}\mathbf{u}$  versus  $\mathbf{u}$  (Wang, 1987). O padrão nulo desse gráfico é linear com coeficiente angular  $\hat{\gamma}$  se a escala da variável  $\mathbf{u}$  está adequada. A forma desse gráfico pode sugerir uma escala alternativa para  $\mathbf{u}$ .

**f) Gráficos de índices**

Consistem em se plotarem medidas de diagnósticos  $d_i$  versus  $i$ . Servem para localizar observações com resíduo, *leverage* ( $h_{ii}$ ), distância de Cook modificada, etc, grandes.

**g) Gráficos normal e semi-normal de probabilidades (*normal plots e half normal plots*)**

Esses gráficos são construídos da mesma maneira que para o modelo clássico de regressão, usando-se, porém, a distribuição pertinente. Moral et al. (2015) apresentam um pacote com funções que possibilitam a construção desses gráficos para uma grande variedade de modelos, usando o software R.

**h) Valores observados ou resíduos *versus* tempo**

Mesmo que o tempo não seja uma variável incluída no modelo, gráficos de valores observados ( $y$ ) ou de resíduos *versus* tempo devem ser construídos sempre que possível. Esse tipo de gráfico pode conduzir à detecção de padrões concebidos a priori, devido ao tempo ou, então, a alguma variável muito correlacionada com o tempo.

## 5.5 Método das variáveis explanatórias adicionais na seleção de modelos

O método das variáveis explanatórias adicionais (Pregibon, 1979, Capítulo 3), consiste em aumentar a estrutura linear do modelo, usando-se variáveis explanatórias bastante adequadas para representar anomalias específicas no MLG usual. A forma mais comum do método tem origem no trabalho de Box e Tidwell (1962), que consideraram uma regressão com parâmetros não-lineares nas variáveis explanatórias. Se existir, no preditor, uma função  $h(x; \gamma)$ , em que  $\gamma$  é não-linear em  $x$ , expande-se essa função em série de Taylor ao redor de um valor próximo conhecido  $\gamma^{(o)}$  tornando

$\gamma$  um parâmetro linear na variável explanatória adicional  $\partial h(x; \gamma) / \partial \gamma \big|_{\gamma=\gamma^{(o)}}$ .

No método, a estrutura linear do modelo aumentado é do tipo

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}, \quad (5.12)$$

em que  $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_q)$ , sendo  $\mathbf{z}_r$  um vetor coluna de dimensão  $n$  conhecido e  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)^T$ . Em casos especiais, as colunas  $\mathbf{z}_r$  podem ser funções do ajuste do modelo usual, isto é,  $\mathbf{Z} = \mathbf{Z}(\mathbf{X}\hat{\boldsymbol{\beta}})$ , ou funções específicas das variáveis explanatórias originais  $\mathbf{z}_r = \mathbf{z}_r(\mathbf{x}^{(r)})$ .

A importância das variáveis explanatórias adicionais é expressa pela diferença dos desvios dos modelos  $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$  e (5.12). Se a adição das variáveis explanatórias  $\mathbf{Z}\boldsymbol{\gamma}$  altera substancialmente o ajuste, as anomalias em questão afetam, seriamente, o modelo original. Em geral, quando isso ocorre, as formas das variáveis explanatórias adicionais produzem uma ação corretiva.

Um bom exemplo do uso de uma variável explanatória adicional está no teste de Tukey (1949) de um grau de liberdade para verificar a não-aditividade de um modelo. Em termos de MLGs, considera-se  $(\mathbf{X}\hat{\boldsymbol{\beta}}) \odot (\mathbf{X}\hat{\boldsymbol{\beta}})$ , em que  $\odot$  é o produto termo a termo, como uma variável adicional e, se no ajuste do modelo aumentado, o coeficiente dessa variável explanatória for significativamente diferente de zero, aceita-se a não-aditividade no modelo original. Uma transformação do tipo potência da variável resposta pode ser uma medida corretiva para eliminar a não-aditividade.

Para verificar se a escala de uma variável explanatória isolada  $\mathbf{x}^{(r)}$  está correta, o teste de Tukey considera  $\hat{\beta}_r^2(\mathbf{x}^{(r)} \odot \mathbf{x}^{(r)})$ , em que  $\hat{\beta}_r$  é o coeficiente estimado de  $\mathbf{x}^{(r)}$ , como uma variável adicional. Quando o coeficiente associado a essa variável explanatória, no ajuste do modelo aumentado, for estatisticamente zero, aceita-se a linearidade de  $\boldsymbol{\eta}$  em  $\mathbf{x}^{(r)}$ .

Pregibon (1979) recomenda um método gráfico alternativo, baseado na estatística  $\mathbf{v}_r = \hat{\beta}_r \mathbf{x}^{(r)} + \hat{\mathbf{z}} - \hat{\boldsymbol{\eta}} = \hat{\beta}_r \mathbf{x}^{(r)} + \hat{\mathbf{H}}(\mathbf{y} - \hat{\boldsymbol{\mu}})$ , que representa uma medida da linearidade da variável explanatória  $\mathbf{x}^{(r)}$ . A estatística  $\mathbf{v}_r$  é, simplesmente, um resíduo parcial generalizado para a variável explanatória  $\mathbf{x}^{(r)}$ , expresso na escala da



variável dependente modificada  $\mathbf{z}$ . A escala de  $\mathbf{x}^{(r)}$  é considerada correta se o gráfico de  $\mathbf{v}_r$  *versus*  $\mathbf{x}^{(r)}$  é, aproximadamente, linear. Caso contrário, a forma do gráfico deve sugerir a ação corretiva.

A inferência sobre  $\boldsymbol{\gamma}$  pode ser realizada a partir da redução do desvio do modelo com a inclusão de  $\mathbf{Z}\boldsymbol{\gamma}$ , ou por meio da distribuição normal assintótica de  $\hat{\boldsymbol{\gamma}}$ , de média igual ao parâmetro verdadeiro  $\boldsymbol{\gamma}$  e matriz de covariância expressa por

$$(\mathbf{Z}^T \mathbf{W} \mathbf{Z})^{-1} + \mathbf{L}(\mathbf{X}^T \mathbf{W} \mathbf{X} - \mathbf{X}^T \mathbf{W} \mathbf{Z} \mathbf{L})^{-1} \mathbf{L}^T,$$

em que  $\mathbf{L} = (\mathbf{Z}^T \mathbf{W} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{W} \mathbf{X}$ . O método das variáveis explanatórias adicionais é bastante usado para estimar a função de ligação e para identificar observações que não são importantes para o modelo.

## 5.6 Verificação da função de ligação

Na Seção ??, serão apresentadas várias funções de ligação com objetivos diferentes. Muitas vezes, para um conjunto particular de observações, pode ser difícil decidir qual a melhor função de ligação que pode não pertencer a uma família especificada.

Um método informal para verificar a adequação da função de ligação usada é o gráfico da variável dependente ajustada estimada  $\hat{\mathbf{z}} = \hat{\boldsymbol{\eta}} + \hat{\mathbf{G}}(\mathbf{y} - \hat{\boldsymbol{\mu}})$  *versus* o preditor linear estimado  $\hat{\boldsymbol{\eta}}$ . O padrão nulo é uma reta, indicando a adequação da função de ligação usada. O gráfico da variável adicionada, também, pode ser usado, considerando-se  $\mathbf{u} = \hat{\boldsymbol{\eta}} \odot \hat{\boldsymbol{\eta}}$ , em que  $\odot$  significa produto termo a termo. O padrão nulo indicará que a função de ligação usada é adequada.

Para funções de ligação na família potência, uma curvatura para cima no gráfico indica que deve ser usada uma função de ligação com expoente maior, enquanto que uma curvatura para baixo indica um expoente menor. Esse tipo de gráfico não é adequado para dados binários.

Existem dois métodos formais para verificar a adequação da função de ligação utilizada:

i) o mais simples consiste em se adicionar  $\mathbf{u} = \hat{\boldsymbol{\eta}} \odot \hat{\boldsymbol{\eta}}$  como uma variável explanatória extra e examinar a mudança ocorrida no desvio, o que equivale ao teste da razão de verossimilhanças. Se ocorrer uma diminuição drástica há evidência de que a função de ligação é insatisfatória. Pode-se usar, também, o teste escore;

ii) outro método formal de estimação da função de ligação, desenvolvido por Pregibon (1980), usa variáveis explanatórias adicionais, obtidas de uma linearização da função de ligação. Seja a função de ligação  $g(\boldsymbol{\mu}; \boldsymbol{\lambda}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$  dependente de um conjunto de parâmetros  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_r)^T$ , supostos desconhecidos. Uma família de funções de ligação com um único parâmetro é a família potência  $g(\mu; \lambda) = (\mu^\lambda - 1)/\lambda$  ou  $\mu^\lambda$ . Incerteza sobre a função de ligação é mais comum com dados contínuos que têm distribuição gama e com proporções cujo número de sucessos segue a distribuição binomial. Assim, por exemplo, para observações com distribuição gama, pode-se usar a família de funções de ligação  $\eta = \mu^\lambda$ . Para dados com distribuição binomial, pode-se usar a família de funções de ligação  $\eta = \log[(1 - \pi)^{-\lambda} - 1]/\lambda$  de Aranda-Ordaz (1981) que tem como casos especiais a função de ligação logística para  $\lambda = 1$  e a complemento log-log quando  $\lambda \rightarrow 0$ . Em geral, usa-se o método do logaritmo da função de verossimilhança perfilada para se estimar  $\lambda$ . Para o modelo clássico de regressão, esse teste equivale ao teste proposto por Tukey (1949) para não-aditividade.

Um teste aproximado da hipótese nula composta  $H_0 : \boldsymbol{\lambda} = \boldsymbol{\lambda}^{(0)}$ , em que  $\boldsymbol{\lambda}^{(0)}$  é um valor especificado para  $\boldsymbol{\lambda}$ , versus  $H : \boldsymbol{\lambda} \neq \boldsymbol{\lambda}^{(0)}$ , pode ser deduzido expandindo  $g(\boldsymbol{\mu}; \boldsymbol{\lambda}) = \boldsymbol{\eta}$  em série de Taylor ao redor de  $\boldsymbol{\lambda}^{(0)}$  até primeira ordem. Tem-se

$$g(\boldsymbol{\mu}; \boldsymbol{\lambda}) = g(\boldsymbol{\mu}; \boldsymbol{\lambda}^{(0)}) + \mathbf{D}(\boldsymbol{\mu}; \boldsymbol{\lambda}^{(0)})(\boldsymbol{\lambda} - \boldsymbol{\lambda}^{(0)}), \quad (5.13)$$

em que  $\mathbf{D}(\boldsymbol{\mu}; \boldsymbol{\lambda}) = \frac{\partial g(\boldsymbol{\mu}; \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}$  é uma matriz de dimensões  $n \times r$  que depende de  $\boldsymbol{\beta}$  e  $\boldsymbol{\lambda}$ . Seja  $\hat{\boldsymbol{\beta}}_0$  a EMV de  $\boldsymbol{\beta}$  calculada do ajuste do modelo  $g(\boldsymbol{\mu}; \boldsymbol{\lambda}^{(0)}) = \mathbf{X}\boldsymbol{\beta}$  e  $\hat{\boldsymbol{\mu}}_0 = g^{-1}(\mathbf{X}\hat{\boldsymbol{\beta}}_0; \boldsymbol{\lambda}^{(0)})$ . Estima-se  $\mathbf{D}(\boldsymbol{\mu}; \boldsymbol{\lambda}^{(0)})$  por  $\hat{\mathbf{D}}^{(0)} = \mathbf{D}(\hat{\boldsymbol{\mu}}_0; \boldsymbol{\lambda}^{(0)})$ .

Se a expansão (5.13) for adequada, pode-se considerar a estrutura linear

$$g(\boldsymbol{\mu}; \boldsymbol{\lambda}^{(0)}) = \begin{pmatrix} \mathbf{X} & -\hat{\mathbf{D}}^{(0)} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\lambda} \end{pmatrix} + \hat{\mathbf{D}}^{(0)} \boldsymbol{\lambda}^{(0)} \quad (5.14)$$

como uma aproximação de  $g(\boldsymbol{\mu}; \boldsymbol{\lambda}) = \mathbf{X}\boldsymbol{\beta}$  com  $\boldsymbol{\lambda}$  desconhecido.

Na estrutura (5.14), o vetor de parâmetros  $\boldsymbol{\lambda}$  aparece como linear nas variáveis adicionais  $-\hat{\mathbf{D}}^{(0)}$  e o preditor linear envolve  $\hat{\mathbf{D}}^{(0)}\boldsymbol{\lambda}^{(0)}$  como *offset*. Essas variáveis adicionais representam uma medida da distância da função de ligação definida por  $\boldsymbol{\lambda}^{(0)}$  à função de ligação verdadeira. A inferência sobre  $\boldsymbol{\lambda}$  pode ser realizada de maneira análoga a  $\boldsymbol{\beta}$ , como descrito na Seção 4.6.2.

Logo, testar  $H_0 : \boldsymbol{\lambda} = \boldsymbol{\lambda}^{(0)}$  versus  $H : \boldsymbol{\lambda} \neq \boldsymbol{\lambda}^{(0)}$  corresponde, aproximadamente, a comparar os modelos  $\mathbf{X}$  e  $(\mathbf{X} - \hat{\mathbf{D}}^{(0)})$ , ambos tendo a mesma função de ligação  $g(\boldsymbol{\mu}; \boldsymbol{\lambda}^{(0)}) = \boldsymbol{\eta}$ . Se a diferença de desvios entre esses modelos é maior do que  $\chi_r^2(\alpha)$ , rejeita-se a hipótese nula  $H_0$ .

A aproximação do teste depende fortemente da linearização (5.13). Quando o  $\boldsymbol{\lambda}$  verdadeiro estiver distante de  $\boldsymbol{\lambda}^{(0)}$ , não existirá garantia de convergência no ajuste de (5.14) e, mesmo convergindo, a estimativa de  $\boldsymbol{\lambda}$  obtida pode diferir substancialmente do valor correto de sua EMV. Para calcular uma melhor aproximação dessa estimativa, o processo (5.14) deverá ser repetido com as variáveis explanatórias adicionais sendo reestimadas a cada etapa, a partir das estimativas correspondentes de  $\boldsymbol{\beta}$  e  $\boldsymbol{\lambda}$ .

Um processo alternativo para obter uma boa estimativa de  $\boldsymbol{\lambda}$  é considerar  $\boldsymbol{\lambda}$  fixado e pertencendo a um conjunto amplo de valores arbitrários e, então, computar o desvio  $S_p(\boldsymbol{\lambda})$  como função de  $\boldsymbol{\lambda}$ . Traça-se o gráfico da superfície  $S_p(\boldsymbol{\lambda})$  versus  $\boldsymbol{\lambda}$ , escolhendo a estimativa  $\tilde{\boldsymbol{\lambda}}$  correspondente ao valor mínimo de  $S_p(\boldsymbol{\lambda})$  nesse conjunto. Se  $\boldsymbol{\lambda}$  é unidimensional, o processo é bastante simples, caso contrário, pode ser impraticável. Uma região de  $100(1 - \alpha)\%$  de confiança para  $\boldsymbol{\lambda}$  é determinada no gráfico por  $\{\boldsymbol{\lambda}; S_p(\boldsymbol{\lambda}) - S_p(\tilde{\boldsymbol{\lambda}}) \leq \chi_r^2(\alpha)\}$ , sendo independente da parametrização adotada. Um teste de  $H_0 : \boldsymbol{\lambda} = \boldsymbol{\lambda}^{(0)}$  pode ser baseado nessa região. Pode-se calcular, numericamente, a EMV de  $\boldsymbol{\lambda}$ , embora com uma maior complexidade computacional.

A verificação da adequação da função de ligação é, inevitavelmente, afetada pela falha em estabelecer escalas corretas para as variáveis explanatórias no preditor linear. Em particular, se o teste formal construído pela adição de  $\hat{\boldsymbol{\eta}} \odot \hat{\boldsymbol{\eta}}$  ao preditor

linear produz uma redução significativa no desvio do modelo, isso pode indicar uma função de ligação errada ou escalas erradas para as variáveis explanatórias ou ambas. Observações atípicas, também, podem afetar a escolha da função de ligação.

**Exemplo 5.1:** Seja a função de ligação  $g_0(\boldsymbol{\mu}) = g(\boldsymbol{\mu}, \lambda_0) = X\boldsymbol{\beta}$ , incluída em uma família paramétrica  $g(\boldsymbol{\mu}, \lambda)$ , indexada pelo parâmetro escalar  $\lambda$ , por exemplo,

$$g(\boldsymbol{\mu}, \lambda) = \begin{cases} \frac{\boldsymbol{\mu}^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log(\boldsymbol{\mu}) & \lambda = 0 \end{cases} \quad (5.15)$$

que inclui as funções de ligação identidade, logarítmica, etc, ou então, a família de Aranda-Ordaz,

$$\boldsymbol{\mu} = \begin{cases} 1 - (1 + \lambda e^\eta)^{-\frac{1}{\lambda}} & \lambda e^\eta > -1 \\ 1 & c.c. \end{cases}$$

que inclui as funções de ligação logística, complemento log-log, etc.

A expansão de Taylor, no caso unidimensional, para  $g(\boldsymbol{\mu}, \lambda)$  ao redor de um valor conhecido  $\lambda_0$ , produz

$$g(\boldsymbol{\mu}, \lambda) \simeq g(\boldsymbol{\mu}, \lambda_0) + (\lambda - \lambda_0)\mathbf{u}(\lambda_0) = X\boldsymbol{\beta} + \gamma\mathbf{u}(\lambda_0),$$

$$\text{em que } \mathbf{u}(\lambda_0) = \left. \frac{\partial g(\boldsymbol{\mu}, \lambda)}{\partial \lambda} \right|_{\lambda=\lambda_0}.$$

De uma forma geral, usa-se  $\mathbf{u} = \hat{\boldsymbol{\eta}} \odot \hat{\boldsymbol{\eta}}$ , cuja justificativa é mostrada a seguir, como variável adicionada ao preditor linear do modelo para o teste de adequação da função de ligação de um MLG.

Suponha que a função de ligação considerada é  $\boldsymbol{\eta} = g(\boldsymbol{\mu})$  e que a função de ligação verdadeira seja  $g^*(\boldsymbol{\mu})$ . Então,

$$g(\boldsymbol{\mu}) = g[g^{*-1}(\boldsymbol{\eta})] = h(\boldsymbol{\eta}).$$

A hipótese nula é  $H_0 : h(\boldsymbol{\eta}) = \boldsymbol{\eta}$  e a alternativa é  $H : h(\boldsymbol{\eta}) = \text{não linear}$ .

Fazendo-se a expansão de  $g(\boldsymbol{\mu})$  em série de Taylor, tem-se

$$g(\boldsymbol{\mu}) \simeq h(0) + h'(0)\boldsymbol{\eta} + \frac{h''(0)}{2}\boldsymbol{\eta} \odot \boldsymbol{\eta}$$

e, então, a variável adicionada é  $\hat{\eta} \odot \hat{\eta}$ , desde que o modelo tenha termos para o qual a média geral seja marginal.

**Exemplo 5.2:** Considere os dados do Exemplo 2.1. A variável resposta tem distribuição binomial, isto é,  $Y_i \sim B(m_i, \pi_i)$ . Adotando-se a função de ligação logística (canônica) e os preditores lineares expressos por

$$\eta_i = \log \left( \frac{\mu_i}{m_i - \mu_i} \right) = \beta_1 + \beta_2 d_i,$$

e

$$\eta_i = \log \left( \frac{\mu_i}{m_i - \mu_i} \right) = \beta_1 + \beta_2 d_i + \gamma u_i,$$

sendo  $u_i = \hat{\eta}_i^2$ , usa-se a diferença de desvios para testar a adequação da função de ligação, obtendo-se os resultados da Tabela 5.3. Verifica-se que se rejeita a hipótese nula  $H_0 : \gamma = 0$ , ao nível de 5% de significância, indicando que a função de ligação logística não é adequada. A estimativa para  $\gamma$  é  $\hat{\gamma} = -0,2087$  com erro padrão 0,0757.

Tabela 5.3: Análise de desvio e teste da função de ligação para os dados do Exemplo 2.1.

Causa de variação	g.l.	Desvio	Valor de p
Regressão linear	1	153,480	< 0,0001
Função de ligação	1	9,185	0,0024
Novo Resíduo	3	1,073	
(Resíduo)	4	10,260	0,0527
Total	5	163,740	

Fazendo-se uma análise de resíduos, verifica-se que a primeira observação é discrepante. Eliminando-a e refazendo-se o teste para a função de ligação, a hipótese nula  $H_0 : \gamma = 0$  não é rejeitada, indicando a adequação da função de ligação logística.

Tem-se, então,  $\hat{\gamma} = 0,0757$  com erro padrão 0,086 e,

$$\eta_i = \log \left( \frac{\hat{\mu}_i}{m_i - \hat{\mu}_i} \right) = -3,5823 + 0,7506d_i.$$

## 5.7 Verificação da função de variância

Um método informal para verificar a adequação da função de variância (que é definida ao se escolher uma determinada distribuição) é o gráfico dos resíduos absolutos *versus* os valores ajustados transformados em uma escala com variância constante (vide Seção 5.4.2, item a). O padrão nulo para esse tipo de gráfico é uma distribuição aleatória de média zero e amplitude constante. A escolha errada da função de variância mostrará uma tendência na média. Em geral, a não adequação da função de variância será considerada como superdispersão (Hinde e Demétrio, 1998a,b).

Um método formal para verificar a adequação da função de variância consiste em indexar essa função por um parâmetro  $\lambda$  e fazer um teste de hipótese  $H_0 : \lambda = \lambda_0$ , usando-se os testes da razão de verossimilhanças e escore. Assim, por exemplo, pode-se usar  $V(\mu) = \mu^\lambda$  e observar como o ajuste varia em função de  $\lambda$ . Em geral, usa-se o logaritmo da função de verossimilhança perfilada para se estimar  $\lambda$ .

Para se compararem ajustes de modelos com diferentes funções de variância, o desvio não pode ser usado, e há necessidade de se usar a teoria de quase verossimilhança estendida, que será discutida na Seção ??.

A verificação da adequação da função de variância é, inevitavelmente, afetada pela falha em estabelecer escalas corretas para as variáveis explanatórias no preditor linear, escolha errada da função de ligação e observações atípicas.

## 5.8 Verificação das escalas das variáveis explanatórias

O gráfico de resíduos parciais é uma ferramenta importante para saber se um termo  $\beta x$  no preditor linear pode ser melhor expresso como  $\beta h(x; \lambda)$  para alguma função monótona  $h(\cdot; \lambda)$ . Nos MLGs, o vetor de resíduos parciais (ou resíduos + componente) é especificado por

$$\tilde{\mathbf{r}} = \hat{\mathbf{z}} - \hat{\boldsymbol{\eta}} + \hat{\boldsymbol{\gamma}}\mathbf{x},$$

sendo  $\hat{\mathbf{z}}$  a variável dependente ajustada estimada,  $\hat{\boldsymbol{\eta}}$  o preditor linear estimado e  $\hat{\boldsymbol{\gamma}}$  a estimativa do parâmetro referente à variável explanatória  $\mathbf{x}$ .

O gráfico de  $\tilde{\mathbf{r}}$  *versus*  $\mathbf{x}$  conduz a um método informal. Se a escala de  $\mathbf{x}$  é satisfatória, o gráfico deve ser, aproximadamente, linear. Se não, sua forma pode sugerir um modelo alternativo. Poderão, entretanto, ocorrer distorções se as escalas das outras variáveis explanatórias estiverem erradas e, então, pode ser necessário analisar gráficos de resíduos parciais para diversos  $x$ s.

Um método formal consiste em colocar  $x$  em uma família  $h(\cdot; \lambda)$  indexada por  $\lambda$ ; calcular, então, o logaritmo da função de verossimilhança maximizada para um conjunto de valores de  $\lambda$  e determinar  $\hat{\lambda}$  como aquele valor que conduz a um logaritmo da função de verossimilhança maximal (método da verossimilhança perfilada). O ajuste para  $\hat{\lambda}$  será, então, comparado com o ajuste para a escolha inicial  $\lambda_0$  que, em geral, é um. Esse procedimento pode ser usado para vários  $x$ s simultaneamente e é, particularmente, útil quando se têm as mesmas dimensões físicas, tal que seja necessária uma transformação comum. A família mais comum de transformações é a família de Box e Cox (1964) expressa por  $h(x; \lambda) = (x^\lambda - 1)/\lambda$ , se  $\lambda \neq 0$ , e  $h(x; \lambda) = \log(x)$ , se  $\lambda = 0$ .

Um método informal para o estudo de uma única variável explanatória implica na forma  $\mathbf{u}(\lambda_0) = dh(\boldsymbol{\mu}, \lambda)/d\lambda|_{\lambda=\lambda_0}$  que é, então, usada como variável adicional para o teste de adequação da escala usada para a variável explanatória de interesse. Pode-se, então, fazer o gráfico dos resíduos parciais, como descrito na Seção 5.4.2, item e). Essa mesma variável  $u$  construída pode ser usada como uma variável adicional no modelo para o teste da hipótese  $H_0 : \lambda = \lambda_0$  (o que equivale ao

teste de  $H_0 : \gamma = 0$ ) que, se não rejeitada, indicará a adequação da escala escolhida para a variável explanatória de interesse.

### Exemplo 5.3: Transformação para a variável dependente

Seja a família de transformações de Box-Cox normalizada

$$\mathbf{z}(\lambda) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \begin{cases} \frac{\mathbf{y}^\lambda - 1}{\lambda \dot{y}^{\lambda-1}} + \boldsymbol{\epsilon} & \lambda \neq 0 \\ \dot{y} \log(\mathbf{y}) + \boldsymbol{\epsilon} & \lambda = 0, \end{cases}$$

sendo  $\dot{y}$  a média geométrica das observações. A expansão de  $\mathbf{z}(\lambda)$  em série de Taylor em relação a  $\lambda_0$ , suposto conhecido, é

$$\mathbf{z}(\lambda) \approx \mathbf{z}(\lambda_0) + (\lambda - \lambda_0)\mathbf{u}(\lambda_0),$$

sendo  $\mathbf{u}(\lambda_0) = \left. \frac{d\mathbf{z}(\lambda)}{d\lambda} \right|_{\lambda=\lambda_0}$ . Então, o modelo linear aproximado é

$$\mathbf{z}(\lambda_0) = \mathbf{z}(\lambda) - (\lambda - \lambda_0)\mathbf{u}(\lambda_0) = \mathbf{X}\boldsymbol{\beta} + \gamma\mathbf{u} + \boldsymbol{\epsilon}.$$

Mas  $\mathbf{z}(\lambda) = \frac{\mathbf{y}^\lambda - 1}{\lambda \dot{y}^{\lambda-1}} + \boldsymbol{\epsilon}$  e, portanto,

$$\mathbf{u}(\lambda) = \frac{d\mathbf{z}(\lambda)}{d\lambda} = \frac{\mathbf{y}^\lambda \log(\mathbf{y}) - (\mathbf{y}^\lambda - 1)[\lambda^{-1} + \log(\dot{y})]}{\lambda \dot{y}^{\lambda-1}}.$$

O interesse, em geral, está em testar alguns valores de  $\lambda$ , tais como  $\lambda_0 = 1$  (sem transformação) e  $\lambda_0 = 0$  (transformação logarítmica). Desde que são necessários apenas os resíduos de  $\mathbf{u}(\lambda)$ , então, constantes podem ser ignoradas se  $\boldsymbol{\beta}$  contém uma constante. Então,

$$\mathbf{u}(1) = \mathbf{y} \left[ \log\left(\frac{\mathbf{y}}{\dot{y}}\right) - 1 \right], \text{ variável construída para testar se } \lambda_0 = 1$$

e

$$\mathbf{u}(0) = \dot{y} \log(\mathbf{y}) \left[ \frac{\log(\mathbf{y})}{2} - \log(\dot{y}) \right], \text{ variável construída para testar se } \lambda_0 = 0.$$

Como  $-\gamma = \lambda - \lambda_0$ , tem-se que uma estimativa para  $\lambda$  pode ser obtida como  $\hat{\lambda} = \lambda_0 - \hat{\gamma}$ . Usa-se, em geral, um valor para  $\lambda$  próximo de  $\hat{\lambda}$  que tenha uma



interpretação prática.

**Exemplo 5.4: Transformação para as variáveis explanatórias**

Se em lugar de transformar  $y$  houver necessidade de transformar  $x_k$ , tem-se que o componente sistemático mais amplo é especificado por

$$E(\mathbf{Y}) = \beta_0 + \sum_{j \neq k} \beta_j \mathbf{x}_j + \beta_k \mathbf{x}_k^\lambda.$$

A expansão de  $\mathbf{z}(\lambda)$  em série de Taylor em relação a  $\lambda_0$ , suposto conhecido, é

$$\mathbf{z}(\lambda) \approx \mathbf{z}(\lambda_0) + (\lambda - \lambda_0) \left. \frac{d\mathbf{z}(\lambda)}{d\lambda} \right|_{\lambda=\lambda_0}.$$

Então,

$$\mathbf{z}(\lambda) \approx \beta_0 + \sum_{j \neq k} \beta_j \mathbf{x}_j + \beta_k \mathbf{x}_k^{\lambda_0} + \beta_k (\lambda - \lambda_0) \mathbf{x}_k^{\lambda_0} \log(\mathbf{x}_k) = \beta_0 + \sum_{j \neq k} \beta_j \mathbf{x}_j + \beta_k \mathbf{x}_k^{\lambda_0} + \gamma \mathbf{u}(\lambda_0),$$

pois  $d\mathbf{z}(\lambda)/d\lambda = \beta_k \mathbf{x}_k^\lambda \log(\mathbf{x}_k)$ . Portanto, testar  $\lambda = \lambda_0$  é equivalente a testar  $\gamma = 0$  para a regressão com a variável construída  $\mathbf{u}_x(\lambda_0) = \mathbf{x}_k^{\lambda_0} \log(\mathbf{x}_k)$  com  $\mathbf{x}_k^{\lambda_0}$  incluída no modelo. Para  $\lambda_0 = 1$ , tem-se

$$E(\mathbf{Y}) = \beta_0 + \sum_{j \neq k} \beta_j \mathbf{x}_j + \beta_k \mathbf{x}_k + \beta_k (\lambda - 1) \mathbf{x}_k \log(\mathbf{x}_k) = \mathbf{X}\boldsymbol{\beta} + \gamma \mathbf{u}_x,$$

sendo  $\mathbf{u}_x = \mathbf{x}_k \log(\mathbf{x}_k)$  e  $\gamma = \beta_k (\lambda - 1)$ . Portanto, faz-se a regressão de  $\mathbf{Y}$  em relação a todos os  $\mathbf{x}_j$ ,  $j = 1, \dots, p-1$ , e a  $\mathbf{u}_x = \mathbf{x}_k \log(\mathbf{x}_k)$ . Rejeita-se  $H_0 : \lambda = 1$  se  $t = \hat{\gamma}/\text{se}(\hat{\gamma}) > t_{n-p-1, \gamma/2}$  e, nesse caso, mostra-se que a escala das observações não está adequada. A contribuição de observações individuais pode ser examinada, usando-se o gráfico da variável adicionada.

**Exemplo 5.5: Transformação simultânea para as variáveis resposta e explanatórias**

Para a transformação simultânea da variável resposta e das  $p-1$  variáveis explanatórias (exceto a constante  $1^\lambda = 1$ ), o modelo para um vetor de  $p$  parâmetros

$\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)^T$  é

$$\mathbf{z}(\lambda_p) = \beta_0 + \sum_{j=1}^{p-1} \beta_j \mathbf{x}_j^{\lambda_j} + \boldsymbol{\epsilon}. \quad (5.16)$$

Na equação (5.16) cada variável, incluindo a variável resposta, pode ter um parâmetro de transformação diferente. De forma semelhante aos Exemplos 5.3 e 5.4, a expansão de Taylor desse modelo ao redor de um  $\lambda_0$  comum, suposto conhecido, é

$$\mathbf{z}(\lambda_0) = \beta_0 - (\lambda_p - \lambda_0) \mathbf{u}(\lambda_0) + \sum_{j=1}^{p-1} \beta_j \mathbf{x}_j^{\lambda_0} + \sum_{j=1}^{p-1} (\lambda_j - \lambda_0) \beta_j \mathbf{x}_j^{\lambda_0} \log(\mathbf{x}_j) + \boldsymbol{\epsilon}.$$

Para o caso de  $\lambda_0 = 1$ , tem-se

$$\mathbf{z}(\lambda_0) = \beta_0 + \sum_{j=1}^{p-1} \beta_j \mathbf{x}_j + \gamma \left[ \sum_{j=1}^{p-1} \beta_j \mathbf{x}_j \log(\mathbf{x}_j) - \mathbf{u}(1) \right] + \boldsymbol{\epsilon},$$

sendo  $\gamma = \lambda - 1$  e definindo a variável construída por

$$\mathbf{u}_{xy}(1) = \beta_0 + \sum_{j=1}^{p-1} \hat{\beta}_j \mathbf{x}_j \log(\mathbf{x}_j) - \mathbf{y} \left[ \log \left( \frac{\mathbf{y}}{\dot{\mathbf{y}}} \right) - 1 \right]$$

usando-se as estimativas  $\hat{\beta}_j$  de mínimos quadrados do modelo sem transformação no lugar dos  $\beta_j$ . Rejeita-se  $H_0 : \boldsymbol{\lambda} = \mathbf{1}$  se  $t = \hat{\gamma}/\text{se}(\hat{\gamma}) > t_{n-p-1, \gamma/2}$  e, nesse caso, mostra-se que a escala das observações e das variáveis explanatórias não está adequada. A contribuição de observações individuais pode ser examinada, usando-se o gráfico da variável adicionada.

## 5.9 Verificação de anomalias no componente sistemático, usando-se análise dos resíduos

Considera-se um MLG com distribuição na família (1.5) e componente sistemático  $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$ . As possíveis anomalias no componente aleatório do modelo podem ser descobertas pelos gráficos i'), ii') e iii') descritos na Seção 5.4.1, desde que os resíduos sejam definidos apropriadamente. Nesta seção, apresenta-se uma técnica

geral para verificar anomalias no componente sistemático do modelo definido pelas equações (2.5) e (2.6).

Considera-se que o componente sistemático correto contém uma variável explanatória  $\mathbf{z}$  adicional (Seção 5.5) e um parâmetro escalar  $\gamma$ , isto é,

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + h(\mathbf{z}; \gamma), \quad (5.17)$$

em que  $h(\mathbf{z}; \gamma)$  pode representar:

- a) um termo adicional em uma ou mais variáveis explanatórias originais, por exemplo:  $h(\mathbf{z}; \gamma) = \gamma x_j^2$  ou  $h(\mathbf{z}; \gamma) = \gamma x_j x_k$ ;
- b) uma contribuição linear ou não-linear de alguma variável explanatória omitida, por exemplo:  $h(\mathbf{z}; \gamma) = \gamma \mathbf{z}$  ou  $h(\mathbf{z}; \gamma) = \mathbf{z}^\gamma$ .

O objetivo é definir resíduos modificados  $\tilde{\mathbf{R}}$  para o modelo ajustado  $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$  tais que  $E(\tilde{\mathbf{R}}) = h(\mathbf{z}; \gamma)$ . Se isso acontecer, um gráfico de  $\tilde{\mathbf{R}}$  *versus*  $\mathbf{z}$ , desprezando a variação aleatória, exibirá a função  $h(\mathbf{z}; \gamma)$ .

Para fixar ideias, considere o modelo normal linear e os resíduos ordinários usuais:  $\mathbf{R} = \mathbf{y} - \hat{\boldsymbol{\mu}} = [\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] \mathbf{y} = (\mathbf{I} - \mathbf{H}) \mathbf{y}$ . Supondo que o componente sistemático correto é (5.17), tem-se  $\mathbf{R} = (\mathbf{I} - \mathbf{H})[\mathbf{X}\boldsymbol{\beta} + h(\mathbf{z}; \gamma) + \boldsymbol{\varepsilon}]$ , em que  $\boldsymbol{\varepsilon}$  é um ruído branco. Como  $\mathbf{X}$  é ortogonal a  $\mathbf{I} - \mathbf{H}$ , tem-se  $\mathbf{R} = (\mathbf{I} - \mathbf{H})h(\mathbf{z}; \gamma) + \boldsymbol{\varepsilon}$  e, portanto,  $E(\mathbf{R}) = (\mathbf{I} - \mathbf{H})h(\mathbf{z}; \gamma)$ . Assim, um gráfico de  $\mathbf{R}$  *versus*  $\mathbf{z}$  não apresentará nenhuma semelhança com  $h(\mathbf{z}; \gamma)$ . Entretanto, se  $h(\mathbf{z}; \gamma)$  for, aproximadamente, linear, um gráfico de  $\mathbf{R}$  *versus*  $(\mathbf{I} - \mathbf{H})\mathbf{z}$  poderá ser usado. A declividade da reta de mínimos quadrados ajustada aos pontos desse gráfico proporcionará uma estimativa de  $\gamma$  no modelo (5.17). Se a declividade for próxima de zero, o modelo  $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$  poderá ser aceito ao invés de (5.17).

Para o modelo normal linear, supondo  $h(\mathbf{z}; \gamma)$ , aproximadamente, linear, Larsen e McCleary (1972) definem *resíduos parciais* por

$$\tilde{\mathbf{R}} = \mathbf{y} - \hat{\boldsymbol{\mu}} + \hat{\gamma} \mathbf{H} \mathbf{z} = (\mathbf{I} - \mathbf{H}) \mathbf{y} + \hat{\gamma} \mathbf{H} \mathbf{z}, \quad (5.18)$$

em que  $\hat{\gamma}$  é a estimativa de mínimos quadrados de  $\gamma$  baseada na regressão de  $\mathbf{y} - \hat{\boldsymbol{\mu}}$  sobre a matriz  $(\mathbf{I} - \mathbf{H})\mathbf{z}$ , isto é,  $\hat{\gamma} = [\mathbf{z}^T(\mathbf{I} - \mathbf{H})\mathbf{z}]^{-1}\mathbf{z}^T(\mathbf{I} - \mathbf{H})(\mathbf{y} - \hat{\boldsymbol{\mu}})$ , com  $\mathbf{z} = (z_1, \dots, z_n)^T$ . Pode-se demonstrar que os resíduos parciais (5.18) podem ser expressos como combinações lineares dos resíduos  $\mathbf{y} - \hat{\boldsymbol{\mu}}$  e, também, como combinações lineares das observações  $\mathbf{y}$ .

Ainda, no modelo normal linear, a noção de resíduos parciais pode ser estendida para determinar se variáveis explanatórias, com contribuições não-lineares, estão omissas no componente sistemático do modelo. Suponha, agora, que  $\boldsymbol{\gamma}$  seja um vetor de parâmetros. Isso é possível, desde que a função  $h(\mathbf{z}; \boldsymbol{\gamma})$  possa ser aproximada por um polinômio de grau baixo, isto é,  $h(\mathbf{z}; \boldsymbol{\gamma}) \approx \mathbf{T}\boldsymbol{\gamma}$ , em que  $\mathbf{T} = \mathbf{T}(\mathbf{z}) = (\mathbf{z}, \mathbf{z}^{(2)}, \mathbf{z}^{(3)} \dots)$  com  $\mathbf{z}^{(i)} = (z_1^i, \dots, z_n^i)^T$ .

Com essa aproximação, definem-se os *resíduos aumentados* de Andrews e Pregibon (1978), por uma expressão análoga a (5.18),

$$\tilde{\mathbf{R}} = \mathbf{y} - \hat{\boldsymbol{\mu}} + \mathbf{H}\mathbf{T}\hat{\boldsymbol{\gamma}} = (\mathbf{I} - \mathbf{H})\mathbf{y} + \mathbf{H}\mathbf{T}\hat{\boldsymbol{\gamma}}, \quad (5.19)$$

em que  $\hat{\boldsymbol{\gamma}}$  é a estimativa de mínimos quadrados de  $\boldsymbol{\gamma}$  na regressão linear de  $\mathbf{y} - \hat{\boldsymbol{\mu}}$  sobre  $(\mathbf{I} - \mathbf{H})\mathbf{T}$ , isto é,  $\hat{\boldsymbol{\gamma}} = [\mathbf{T}^T(\mathbf{I} - \mathbf{H})\mathbf{T}]^{-1}\mathbf{T}^T(\mathbf{I} - \mathbf{H})(\mathbf{y} - \hat{\boldsymbol{\mu}})$ .

Tem-se  $E(\tilde{\mathbf{R}}) = \mathbf{T}\boldsymbol{\gamma} \approx h(\mathbf{z}; \boldsymbol{\gamma})$  e, portanto, exceto por variações aleatórias, um gráfico de  $\tilde{\mathbf{R}}$  versus  $\mathbf{z}$  poderá exibir a forma da função  $h(\mathbf{z}; \boldsymbol{\gamma})$ .

Para os MLGs os resíduos aumentados podem ser definidos a partir de resíduos medidos na escala linear

$$\mathbf{R} = \hat{\mathbf{z}} - \hat{\boldsymbol{\eta}} = (\mathbf{I} - \widehat{\mathbf{Z}\mathbf{W}})\hat{\mathbf{z}}. \quad (5.20)$$

Essa expressão foi introduzida na Seção ???. Aqui, estima-se  $\boldsymbol{\gamma}$  ajustando o modelo aumentado  $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{T}\boldsymbol{\gamma}$  aos dados. Isso determinará opções de aperfeiçoamento da estrutura linear do modelo. O ajuste de polinômios de graus elevados é, numericamente, bastante instável, sendo melhor considerar no máximo  $\mathbf{T} = (\mathbf{z}, \mathbf{z}^{(2)}, \mathbf{z}^{(3)})$ .

Tem-se  $\mathbf{R} = (\mathbf{I} - \widehat{\mathbf{Z}\mathbf{W}})(\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{T}\hat{\boldsymbol{\gamma}} + \boldsymbol{\varepsilon}) = (\mathbf{I} - \widehat{\mathbf{Z}\mathbf{W}})(\mathbf{T}\hat{\boldsymbol{\gamma}} + \boldsymbol{\varepsilon})$  e, portanto, os resíduos aumentados nos MLG são expressos por

$$\tilde{\mathbf{R}} = \mathbf{R} + \widehat{\mathbf{Z}\mathbf{W}}\mathbf{T}\hat{\boldsymbol{\gamma}} \quad (5.21)$$

e têm valores esperados próximos de  $h(\mathbf{z}; \boldsymbol{\gamma})$ . Na fórmula (5.21) as estimativas de  $\mathbf{Z}$  e  $\mathbf{W}$  são segundo o modelo reduzido  $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$ .

A expressão (5.19) é um caso especial de (5.21) quando  $\mathbf{W}$  é igual à matriz identidade. Um gráfico de  $\tilde{\mathbf{R}}$  *versus*  $\mathbf{z}$  poderá indicar se essa variável explanatória deve estar incluída no modelo e, se isso acontecer, poderá ainda sugerir a forma de inclusão. Não se devem comparar os resíduos aumentados em (5.21) com os resíduos ordinários  $\mathbf{R}$ , pois os primeiros são baseados no ajuste do modelo aumentado.

A análise gráfica dos resíduos aumentados pode ser bastante útil nos estágios preliminares de seleção de variáveis explanatórias, quando se têm muitas dessas variáveis para serem consideradas. A formação do componente sistemático pode ser feita, passo a passo, com a introdução de uma única variável explanatória, a cada passo, pelo método descrito.

Para determinar a contribuição de uma variável explanatória  $\mathbf{x}_i = (x_{i1}, \dots, x_{in})^T$  da própria matrix  $\mathbf{X}$  no ajuste do modelo reduzido  $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$  aos dados, pode-se trabalhar com os resíduos parciais generalizados

$$v_i = \hat{z}_i - \hat{\eta}_i + \hat{\beta}_j x_{ij}. \quad (5.22)$$

Os resíduos (5.22), descritos na Seção 5.8, são muito mais simples de serem computados do que os resíduos aumentados definidos em (5.21).

## 5.10 Exercícios

1. Compare os resíduos de Anscombe, Pearson e como raiz quadrada do componente do desvio, para o modelo de Poisson. Como sugestão suponha  $\hat{\mu} = cy$  e varie  $c$ , por exemplo, 0(0.2)2(0.5)10. Faça o mesmo para os modelos binomial, gama e normal inverso.

2. Defina os resíduos de Anscombe, Pearson e como raiz quadrada do componente

do desvio para o modelo binomial negativo, comparando-os em algum modelo.

**3.** Seja um MLG com estrutura linear  $\eta_i = \alpha + \beta x_i + x_i^\gamma$  e função de ligação  $g(\cdot)$  conhecida.

(a) Formule, por meio da função desvio, critérios para os seguintes testes:  $H_1 : \gamma = \gamma^{(0)}$  versus  $H'_1 : \gamma \neq \gamma^{(0)}$ ;  $H_2 : \beta = \beta^{(0)}, \gamma = \gamma^{(0)}$  versus  $H'_2 : \beta \neq \beta^{(0)}, \gamma = \gamma^{(0)}$  e versus  $H''_2 : \beta \neq \beta^{(0)}, \gamma \neq \gamma^{(0)}$ ;  $H_3 : \beta = \beta^{(0)}$  versus  $H_3 : \beta \neq \beta^{(0)}$ ;

(b) como obter um intervalo de confiança para  $\gamma$  usando a função desvio?

(c) se a função de ligação dependesse de um parâmetro  $\lambda$  desconhecido, como determinar critérios para os testes citados?

**4.** Os dados da Tabela ?? (Ryan et al., 1976, p. 329) do Apêndice A.1 referem-se a medidas de diâmetro a 4,5 pés acima do solo ( $D$ , polegadas) e altura ( $H$ , pés) de 21 cerejeiras (*black cherry*) em pé e de volume ( $V$ , pés cúbicos) de árvores derrubadas. O objetivo desse tipo de experimento é verificar de que forma essas variáveis estão relacionadas para poder prever o volume de madeira em uma área de floresta (*Allegheny National Forest*), usando medidas nas árvores em pé. Pede-se:

a) faça os gráficos de variáveis adicionadas para  $H$  e  $D$ ;

b) faça os gráficos de resíduos parciais para  $H$  e  $D$ ;

c) faça as transformações  $LV = \log(V)$ ,  $LH = \log(H)$  e  $LD = \log(D)$  e repita os gráficos dos itens (a) e (b);

d) verifique se existem pontos discrepantes em ambas as escalas;

e) usando

$$\mathbf{u}(1) = \sum_{j=2}^p \hat{\beta}_j \mathbf{x}_j \log(\mathbf{x}_j) - \mathbf{y} \left[ \log \left( \frac{\mathbf{y}}{\hat{\mathbf{y}}} \right) - 1 \right],$$

obtido como no Exemplo 5.5 da Seção 5.8, como variável adicionada, verifique que há necessidade da transformação simultânea de  $V$ ,  $H$  e  $D$ .

**5.** Considerando os dados da Tabela 5.4 que se referem à mortalidade de escaravelhos após 5 h de exposição a diferentes doses de bissulfeto de carbono ( $CS^2$ ),

Tabela 5.4: Número de insetos mortos ( $y_i$ ) de  $m_i$  insetos após 5 h de exposição a diferentes doses de  $CS^2$ .

$\log(\text{Dose}) (d_i)$	$m_i$	$y_i$
1,6907	59	6
1,7242	60	13
1,7552	62	18
1,7842	56	28
1,8113	63	52
1,8369	59	53
1,8610	62	61
1,8839	60	60

- a) ajuste o modelo logístico linear e faça o teste para a função de ligação;
- b) ajuste o modelo complemento log-log e faça o teste para a função de ligação;
- c) faça o gráfico da variável adicionada para os itens a) e b);
- d) verifique se há necessidade de transformação para a variável dose usando o gráfico de resíduos parciais.

**6.** Os dados da Tabela 5.5 (Phelps, 1982) são provenientes de um experimento casualizado em três blocos em que foram usadas como tratamentos oito doses de um inseticida fosforado e foram contadas quantas ( $y$ ) cenouras estavam danificadas de totais de  $m$  cenouras.

- a) ajuste o modelo logístico linear e faça o teste para a função de ligação;
- b) ajuste o modelo complemento log-log e faça o teste para a função de ligação;
- c) faça o gráfico da variável adicionada para os itens (a) e (b);
- d) usando a família de funções de ligação de Aranda-Ordaz, obtenha a variável construída e estime  $\lambda$ ;
- e) ajuste o modelo logístico com preditor linear quadrático e faça o teste para a

Tabela 5.5: Número de cenouras danificadas ( $y_i$ ) de  $m_i$  cenouras (Phelps, 1982).

log(Dose)	Bloco I		Bloco II		Bloco III	
$d_i$	$m_i$	$y_i$	$m_i$	$y_i$	$m_i$	$y_i$
1,52	10	35	17	38	10	34
1,64	16	42	10	40	10	38
1,76	8	50	8	33	5	36
1,88	6	42	8	39	3	35
2,00	9	35	5	47	2	49
2,12	9	42	17	42	1	40
2,24	1	32	6	35	3	22
2,36	2	28	4	35	2	31

função de ligação.

7. Considere a família (5.15) de funções de ligação. Mostre que a variável construída para o teste da hipótese  $H_0 : \lambda = 0$  é expressa por (Atkinson, 1985, p. 238)

$$\mathbf{u}(\lambda_0) = \frac{dh(\boldsymbol{\mu}, \lambda)}{d\lambda} \Big|_{\lambda=0} = -\frac{\log(\hat{\boldsymbol{\mu}}) \odot \log(\hat{\boldsymbol{\mu}})}{2} = -\frac{\hat{\boldsymbol{\eta}} \odot \hat{\boldsymbol{\eta}}}{2},$$

em que  $\odot$  representa o produto termo a termo.

8. Seja  $Y_i \sim B(m_i, \mu_i)$  com a notação usual  $\boldsymbol{\mu} = g^{-1}(\mathbf{X}\boldsymbol{\beta})$ ,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ , etc. Demonstre que os resíduos podem ser definidos por

$$\frac{[G(Y_i/m_i) - G'(\hat{\mu}_i)]}{G'(\hat{\mu}_i)} \left[ \frac{\hat{\mu}_i(1 - \hat{\mu}_i)}{m_i} \right]^{1/2}.$$

Quais as vantagens das escolhas  $G(\mu) = \mu$ ,  $G(\mu) = \log[\mu/(1 - \mu)]$  e  $G(\mu) = \int_0^\mu x^{-1/3}(1 - x)^{-1/3}dx$ ?

9. No modelo normal linear com estrutura para a média especificada por  $\boldsymbol{\mu} = E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta} + g(\mathbf{z}; \boldsymbol{\gamma})$ , sendo a função  $g(\mathbf{z}; \boldsymbol{\gamma})$  aproximadamente linear, demonstre



que os resíduos parciais  $\hat{\mathbf{R}} = (\mathbf{I} - \mathbf{H})\mathbf{y} + \mathbf{H}\mathbf{z}\hat{\gamma}$ , em que  $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$  é a matriz de projeção, podem ser expressos como combinações lineares dos resíduos ordinários  $\mathbf{y} - \hat{\boldsymbol{\mu}}$  e, também, como combinações lineares dos dados  $\mathbf{y}$ .

**10.** Demonstre as fórmulas das estatísticas para diagnóstico do MLG apresentadas na Seção 5.3.2.

**11.** Os resíduos  $r_i^{P'}$  definidos em (??) são, também, denominados resíduos de Student (W.S. Gosset). Calcule expressões para  $a_0^{(1)}$ ,  $b_i$  e  $c_i$  em função desses resíduos.

**12.** Seja um modelo normal, gama ou normal inverso com componente usual  $g(\boldsymbol{\mu}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$  e suponha que o parâmetro  $\phi$  seja constante para todas as observações, embora desconhecido. Determine, usando a função desvio, critérios para os seguintes testes:

(a)  $\phi = \phi^{(0)}$  versus  $\phi \neq \phi^{(0)}$ ; (b)  $\boldsymbol{\beta} = \boldsymbol{\beta}^{(0)}$  versus  $\boldsymbol{\beta} \neq \boldsymbol{\beta}^{(0)}$  (Cordeiro, 1986).



# Capítulo 6

## Aplicações a Dados Contínuos

Neste capítulo, apresentam-se análises dos seguintes conjuntos de dados contínuos: volume de árvores, gordura no leite, importação brasileira, tempos de sobrevivência de ratos, assinaturas de TV a cabo, demanda de energia elétrica e tempo de funcionamento de um transformador.

### 6.1 Dados de volume de árvores

Os dados da Tabela ??, do Apêndice A, referem-se a medidas de diâmetro a 4,5 pés acima do solo ( $D$ , polegadas) e de altura ( $H$ , pés) de 31 cerejeiras (“black cherry”) em pé e de volume ( $V$ , pés cúbicos) de árvores derrubadas (Ryan et al., 1976) em uma área da floresta *Allegheny National Forest*. O objetivo desse tipo de experimento é prever o volume de madeira a ser extraída como função de medidas diâmetro e altura nas árvores em pé.

A Figura 6.1 mostra os gráficos de dispersão das variáveis duas a duas para os dados observados sem transformação e com transformação logarítmica. Pode-se verificar que existe alguma relação funcional mais forte entre volume e diâmetro à altura do peito do que entre volume e altura. Além disso, as observações da variável altura têm variabilidade maior do que as observações da variável diâmetro à altura do peito.

Como um primeiro modelo ( $M_1$ ) para a análise desses dados, supõe-se que a variável resposta é  $Y = \mu + \varepsilon_1$ , em que  $Y = V$  e  $\varepsilon_1 \sim N(0, \sigma_1^2)$  e, portanto,

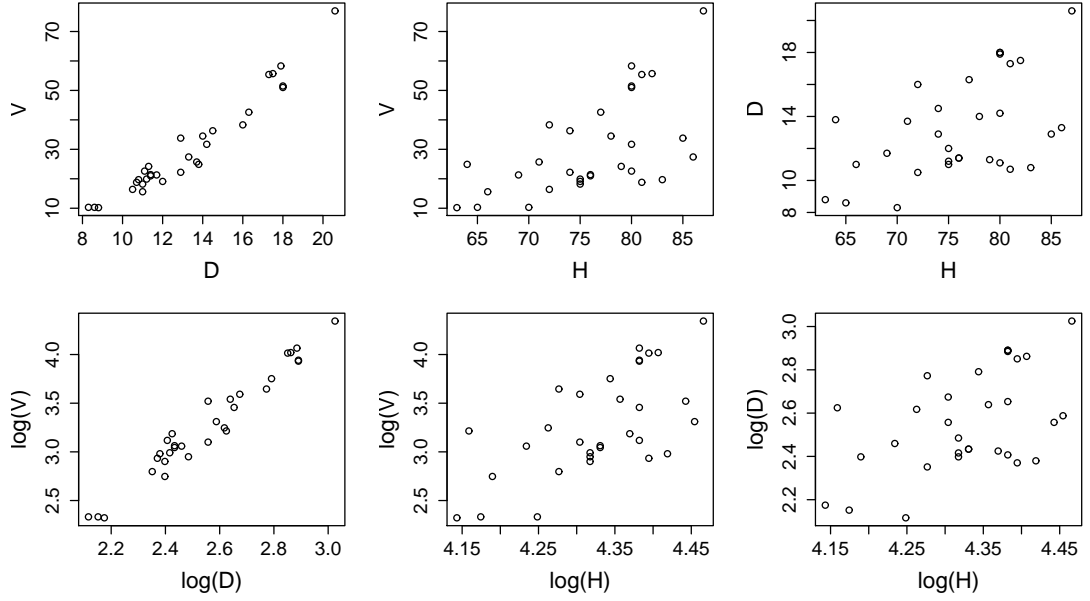


Figura 6.1: Gráfico de dispersão para os dados de cerejeiras - valores observados e transformados na escala logarítmica.

$Y \sim N(\mu, \sigma_1^2)$ , a função de ligação é a identidade,  $\eta = \mu$ , e o preditor linear é expresso por

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2, \quad (6.1)$$

em que  $x_1 = D$  e  $x_2 = H$ .

Um segundo modelo ( $M_2$ ) baseia-se no fato de que o volume é proporcional ao produto do diâmetro à altura do peito pela altura, isto é,  $V \approx \gamma_0 D^{\beta_1} H^{\beta_2}$  e, portanto,  $\log(V) \approx \beta_0 + \beta_1 \log(D) + \beta_2 \log(H)$ . Então, pode-se supor que para a variável resposta transformada  $Y = \mu + \varepsilon_2$ , em que  $Y = \log(V)$  e  $\varepsilon_2 \sim N(0, \sigma_2^2)$  e, portanto,  $Y \sim N(\mu, \sigma_2^2)$ , a função de ligação é a identidade,  $\eta = \mu$ , e o preditor linear é expresso por (6.1) com  $x_1 = \log(D)$  e  $x_2 = \log(H)$ .

Como um terceiro modelo ( $M_3$ ), supõe-se que a variável resposta é  $Y = \mu + \varepsilon_3$ , em que  $Y = V$ ,  $\mu = \gamma_0 D^{\beta_1} H^{\beta_2}$  e  $\varepsilon_3 \sim N(0, \sigma_3^2)$  e, portanto,  $Y \sim N(\mu, \sigma_3^2)$ , a função de ligação é a logarítmica,  $\eta = \log(\mu)$ , e o preditor linear é expresso por (6.1) com  $x_1 = \log(D)$  e  $x_2 = \log(H)$ .

Após o ajuste desses três modelos, foram feitos: a) os gráficos de valores observados versus valores ajustados; b) os gráficos dos resíduos estudentizados versus valores ajustados para os modelos M1 e M2 e dos componentes do desvio versus valores ajustados para o modelo M3 e c) gráficos normais de probabilidade com envelopes simulados para os resíduos estudentizados para os modelos M1 e M2 e dos componentes do desvio, usando-se o pacote `hnp` do software R (Demétrio et al., 2014; Moral et al., 2015). Verificou-se que existem evidências de um bom ajuste de todos eles, conforme Figura 6.2.

As Tabelas 6.1 e 6.2 mostram os resultados obtidos, considerando-se diversos submodelos para o preditor linear, para a análise dos dados sem transformação ( $M_1$ ) e com transformação logarítmica ( $M_2$ ). Verifica-se que existem evidências, ao nível de 1% de significância, que os efeitos tanto do diâmetro à altura do peito como da altura são significativos, sendo que o efeito do diâmetro à altura do peito é maior do que o da altura, tanto para o caso de dados não transformados como para transformados. Entretanto, é muito mais forte no caso de dados transformados. É importante, lembrar, também, que o teste para o modelo com ambas as variáveis (regressão parcial) simultaneamente tem um nível de significância conjunto, enquanto que na análise sequencial, não se sabe o nível conjunto de significância dos testes. Há evidências, portanto, de que ambas as variáveis explanatórias altura e diâmetro são necessárias para explicar o volume e que o melhor ajuste é obtido com os dados transformados. Testes  $t$  (equivalentes aos testes  $F$ ) e intervalos de confiança para os parâmetros e intervalos de previsão para  $Y$  podem, então, ser calculados.

Há necessidade, porém, de um estudo mais detalhado, fazendo-se uma análise dos resíduos e de diagnóstico, para a escolha do modelo final. Conforme pode-se verificar na Figura 6.3, há indicação de que o modelo  $M_1$  não se ajusta bem às observações. No gráfico dos valores ajustados *versus* valores observados, destacam-se como pontos extremos as observações 1, 2, 3 e 31, enquanto que no gráfico dos valores absolutos de DFFitS *versus* índices, destaca-se a observação 31. No gráfico normal de probabilidades, com envelope de simulação, destacam-se as observações

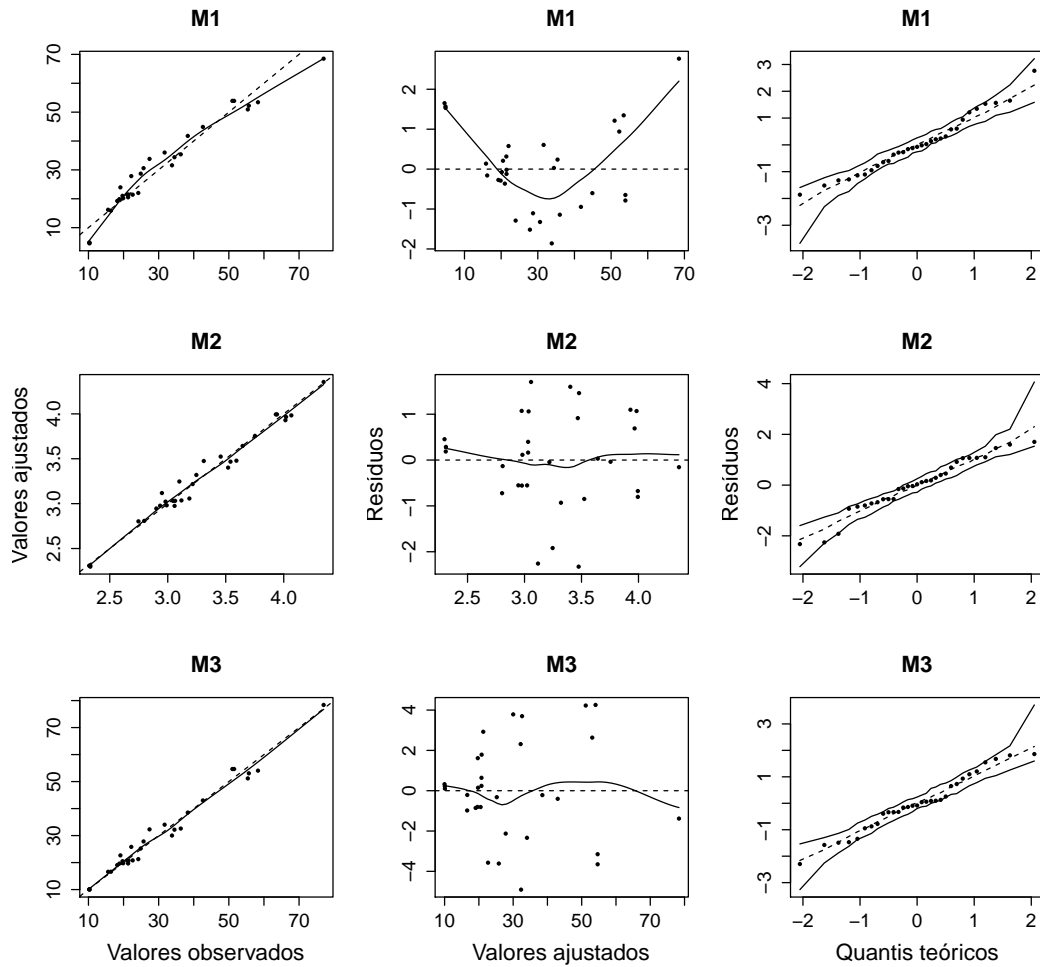


Figura 6.2: Gráficos de valores observados versus valores ajustados; gráficos dos resíduos estudentizados versus valores ajustados para os modelos M1 e M2 e dos componentes do desvio versus valores ajustados para o modelo M3 e gráficos normais de probabilidade com envelopes simulados para os resíduos estudentizados para os modelos M1 e M2 e dos componentes do desvio, para os dados de cerejeiras.

18 e 31. O gráfico para a escolha de uma transformação na família Box-Cox mostra um intervalo de confiança para o parâmetro  $\lambda$  que não inclui o valor  $\lambda = 1$ , indicando uma escala inadequada para a variável resposta. É interessante notar que as árvores 1, 2 e 3 são aquelas de menores volumes, enquanto que a árvore 31 é a de maior volume.

Conforme pode-se verificar na Figura 6.4, há indicação de que o modelo

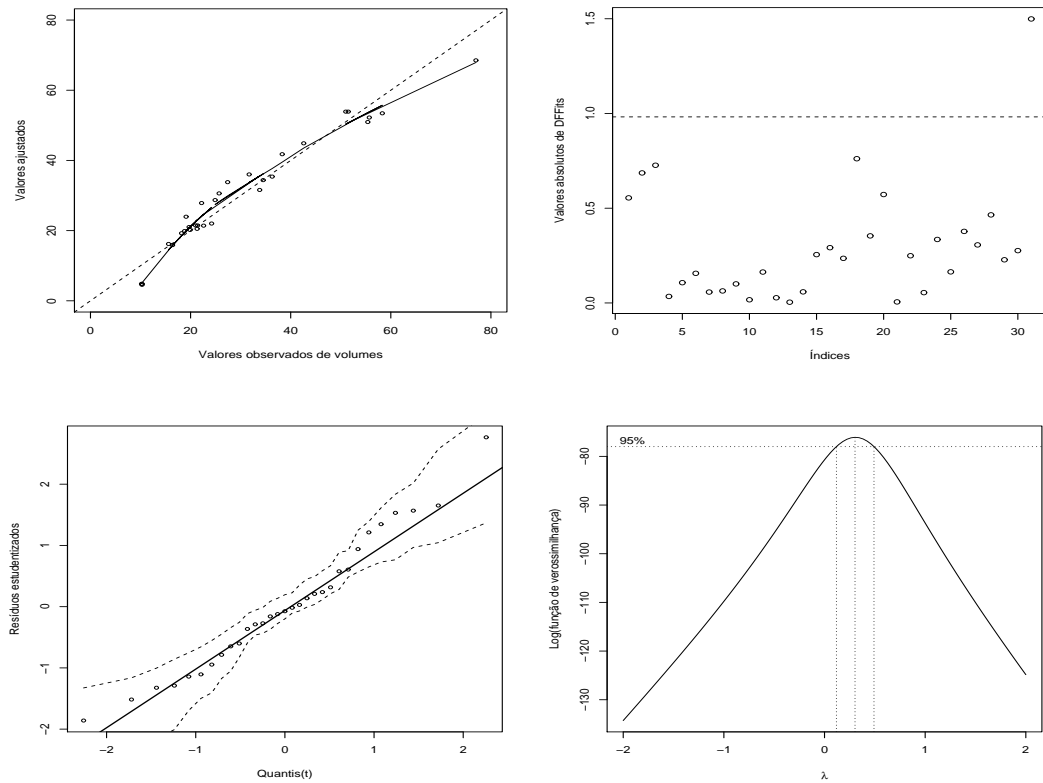


Figura 6.3: Gráficos de valores ajustados (modelo  $M_1$ ) *versus* valores observados, valores absolutos de DFFits *versus* índices, gráfico normal de probabilidades com envelope de simulação e gráfico para escolha da transformação na família Box-Cox (dados de árvores não transformados).

$M_2$  ajusta-se bem às observações. No gráfico de valores ajustados *versus* valores observados, continuam destacando-se como pontos extremos as observações 1, 2, 3 e 31, enquanto que no gráfico de valores absolutos de DFFits *versus* índices, destaca-se a observação 18. No gráfico normal de probabilidades com envelope de simulação, destacam-se as observações 11 e 18. O gráfico para a escolha de uma transformação na família Box-Cox mostra um intervalo de confiança para o parâmetro  $\lambda$  que inclui o valor  $\lambda = 1$ , indicando uma escala adequada para a variável resposta. O programa para as análises foi desenvolvido em R e encontra-se no Apêndice B.1.

Tabela 6.1: Análise de variância, teste  $F$  e estimativas - Dados de árvores sem transformação ( $M_1$ ).

$\eta = \beta_0 + \beta_1 D$				
Causas de variação	G.L.	S.Q.	Q.M.	F
DAP	1	7.581,8	7.581,8	419,4 **
Resíduo	29	524,3	18,1	
Total	30	8.106,1		
$\hat{V} = -36,94 + 5,066D$ $R^2 = 0,935$ $\bar{R}^2 = 0,933$ $s(\hat{\beta}_0) = 3,36$ e $s(\hat{\beta}_1) = 0,247$				
$\eta = \beta_0 + \beta_2 H$				
Causas de variação	G.L.	S.Q.	Q.M.	F
Altura	1	2.901,2	2.901,2	16,2 **
Resíduo	29	5.204,9	179,5	
Total	30	8.106,1		
$\hat{V} = -87,12 + 1,543H$ $R^2 = 0,358$ $\bar{R}^2 = 0,336$ $s(\hat{\beta}_0) = 29,27$ e $s(\hat{\beta}_2) = 0,384$				
$\eta = \beta_0 + \beta_1 D + \beta_2 H$ - Parcial				
Causas de variação	G.L.	S.Q.	Q.M.	F
DAP e Altura	2	7.684,4	3.842,2	255,0 **
Resíduo	28	421,9	15,1	
Total	30	8.106,1		
$\hat{V} = -57,99 + 4,708D + 0,339H$ $R^2 = 0,948$ $\bar{R}^2 = 0,944$ $s(\hat{\beta}_0) = 8,64$ , $s(\hat{\beta}_1) = 0,264$ e $s(\hat{\beta}_2) = 0,130$				
$\eta = \beta_0 + \beta_1 D + \beta_2 H$ - Sequencial				
Causas de variação	G.L.	S.Q.	Q.M.	F
DAP	1	7.581,8	7.581,8	503,1 **
Altura DAP	1	102,4	102,4	6,8*
Resíduo	28	421,9	15,1	
Total	30	8.106,1		
$\eta = \beta_0 + \beta_1 D + \beta_2 H$ - Sequencial				
Causas de variação	G.L.	S.Q.	Q.M.	F
Altura	1	2.901,2	2.901,2	192,5 **
DAP Altura	1	4.783,0	4.783,0	317,4 **
Resíduo	28	421,9	15,1	
Total	30	8.106,1		
$F_{1,29;0,05} = 4,18$ , $F_{2,28;0,05} = 3,34$ e $F_{1,28;0,05} = 4,20$ $F_{1,29;0,01} = 7,60$ , $F_{2,28;0,01} = 5,45$ e $F_{1,28;0,01} = 7,64$				



Tabela 6.2: Análise de variância, teste  $F$  e estimativas - Dados de árvores transformados ( $M_2$ ).

$\eta = \beta_0 + \beta_1 \log(D)$				
Causas de variação	G.L.	S.Q.	Q.M.	F
DAP	1	7,9254	7,9254	599,7 **
Resíduo	29	0,3832	0,0132	
Total	30	8,3087		
$\widehat{\log(V)} = -2,353 + 2,2 \log(D) \quad R^2 = 0,954 \quad \bar{R}^2 = 0,952$ $s(\hat{\beta}_0) = 0,231$ e $s(\hat{\beta}_1) = 0,089$				
$\eta = \beta_0 + \beta_2 \log(H)$				
Causas de variação	G.L.	S.Q.	Q.M.	F
Altura	1	3,496	3,496	21,06 **
Resíduo	29	4,8130	0,166	
Total	30	8,3087		
$\widehat{\log(V)} = -13,96 + 3,982 \log(H) \quad R^2 = 0,421 \quad \bar{R}^2 = 0,401$ $s(\hat{\beta}_0) = 3,76$ e $s(\hat{\beta}_2) = 0,868$				
$\eta = \beta_0 + \beta_1 \log(D) + \beta_2 \log(H)$ - Parcial				
Causas de variação	G.L.	S.Q.	Q.M.	F
DAP e Altura	2	8,1228	4,0614	615,36 **
Resíduo	28	0,1855	0,0066	
Total	30	8,3087		
$\widehat{\log(V)} = -6,632 + 1,983 \log(D) + 1,117 \log(H) \quad R^2 = 0,978$ $\bar{R}^2 = 0,976 \quad s(\hat{\beta}_0) = 0,799, s(\hat{\beta}_1) = 0,0075$ e $s(\hat{\beta}_2) = 0,204$				
$\eta = \beta_0 + \beta_1 \log(D) + \beta_2 \log(H)$ - Sequencial				
Causas de variação	G.L.	S.Q.	Q.M.	F
DAP	1	7,9254	7,9254	1196,5 **
Altura DAP	1	0,1978	0,1978	
Resíduo	28	0,1855	0,0066	
Total	30	8,3087		
$\eta = \beta_0 + \beta_1 \log(D) + \beta_2 \log(H)$ - Sequencial				
Causas de variação	G.L.	S.Q.	Q.M.	F
Altura	1	3,4957	3,4957	527,8 **
DAP Altura	1	4,6275	4,6275	
Resíduo	28	0,1855	0,0066	
Total	30	8,3087		
$F_{1,29;0,05} = 4,18, F_{2,28;0,05} = 3,34$ e $F_{1,28;0,05} = 4,20$ $F_{1,29;0,01} = 7,60, F_{2,28;0,01} = 5,45$ e $F_{1,28;0,01} = 7,64$				

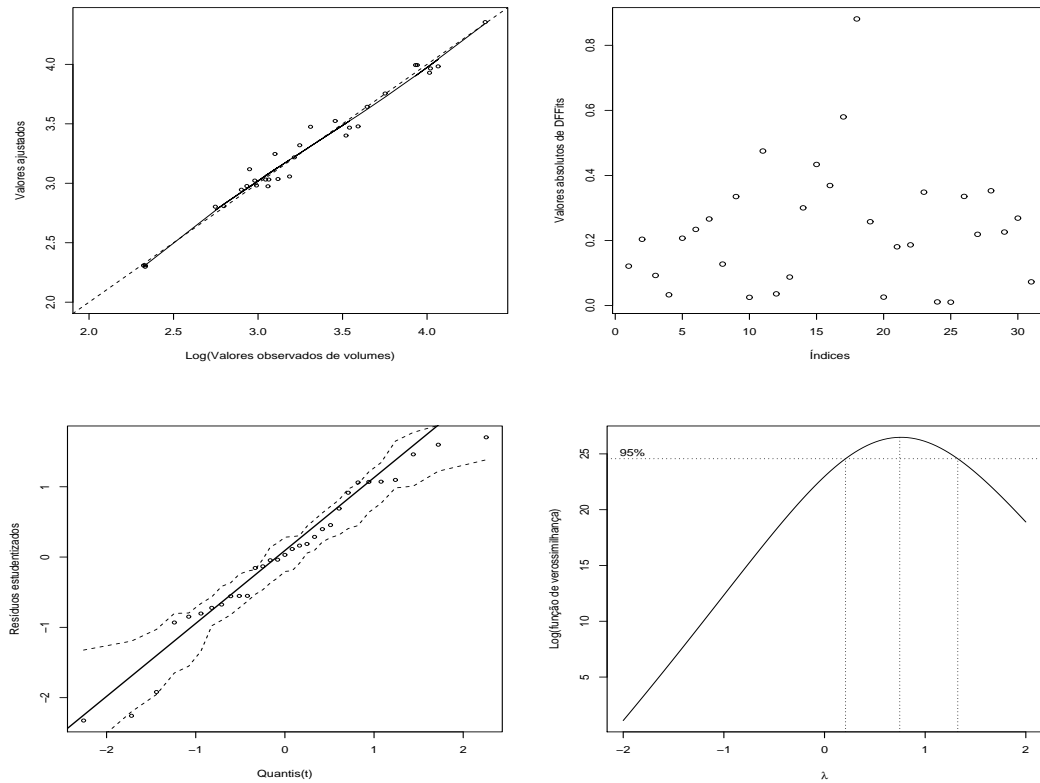


Figura 6.4: Gráficos de valores ajustados (modelo  $M_2$ ) *versus*  $\log(\text{valores observados})$ , valores absolutos de DFFits *versus* índices, gráfico normal de probabilidades com envelope de simulação e gráfico para escolha da transformação na família Box-Cox (dados de árvores transformados).

## 6.2 Dados de gordura no leite

A Tabela 6.3 refere-se a produções médias diárias de gordura (kg/dia) no leite de uma única vaca durante 35 semanas (McCulloch, 2001). É comum supor que a produção média de gordura  $Y_i$  tem distribuição com média

$$\mu_i = \alpha t_i^\beta e^{\gamma t_i},$$

em que  $t$  representa a semana e  $\alpha$ ,  $\beta$  e  $\gamma$  são parâmetros desconhecidos.

Portanto, usando-se a função de ligação logarítmica, tem-se

$$\log(\mu_i) = \log(\alpha) + \beta \log(t_i) + \gamma t_i.$$

Tabela 6.3: Produções médias diárias de gordura (kg/dia) do leite de uma vaca.

0.31	0.39	0.50	0.58	0.59	0.64	0.68
0.66	0.67	0.70	0.72	0.68	0.65	0.64
0.57	0.48	0.46	0.45	0.31	0.33	0.36
0.30	0.26	0.34	0.29	0.31	0.29	0.20
0.15	0.18	0.11	0.07	0.06	0.01	0.01

Pode-se supor, ainda, que  $Y_i \sim N(\mu_i, \tau^2)$ , isto é,

$$Y_i = \mu_i + \delta_i = \alpha t_i^\beta e^{\gamma t_i} + \delta_i,$$

em que  $\delta_i \sim N(0, \tau^2)$ . Isso equivale ao MLG em que a variável resposta  $Y$  tem distribuição normal com função de ligação logarítmica,  $\eta_i = \log(\mu_i)$ , e preditor linear que é igual a  $\log(\alpha) + \beta \log(t_i) + \gamma t_i$ .

Entretanto, na prática é comum supor que  $\log(Y_i) \sim N(\log(\mu_i), \sigma^2)$ , isto é,

$$\log(Y_i) = \log(\mu_i) + \epsilon_i = \log(\alpha) + \beta \log(t_i) + \gamma t_i + \epsilon_i,$$

em que  $\epsilon_i \sim N(0, \sigma^2)$ . Isso equivale ao MLG em que a variável resposta  $\log(Y)$  tem distribuição normal com função de ligação identidade,  $\eta_i = \mu_i$ , e mesmo preditor linear  $\log(\alpha) + \beta \log(t_i) + \gamma t_i$ .

A Figura 6.5 mostra que a distribuição normal com função de ligação logarítmica produz um melhor ajuste do que adotar uma transformação logarítmica dos dados e supor uma distribuição normal com função de ligação identidade. Isso é confirmado nos gráficos de valores ajustados *versus* valores observados apresentado na Figura 6.6. O programa para as análises foi desenvolvido em R e encontra-se no Apêndice B.2.

## 6.3 Dados de importação Brasileira

O impacto das variáveis explanatórias que influenciam a balança comercial tem sido amplamente discutido após a abertura econômica diante do processo de

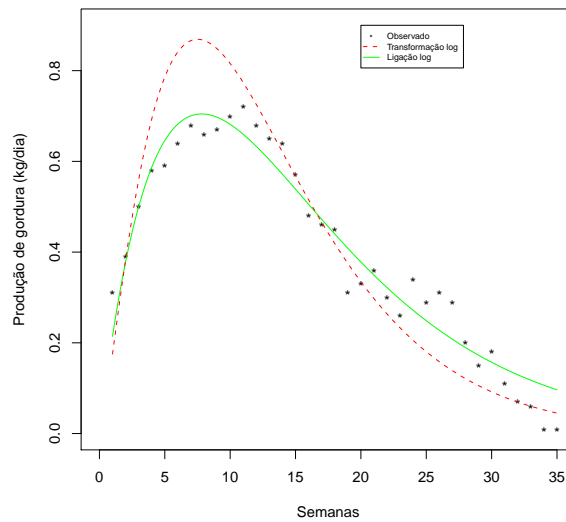


Figura 6.5: Valores observados e curvas ajustadas (dados da Tabela 6.3).

inserção da economia brasileira na globalização dos anos 90. Do ponto de vista da política econômica, é importante identificar esses impactos, bem como, o efeito dinâmico de políticas monetárias e cambiais frente aos setores que se relacionam com o comércio internacional. Dentro desse contexto, há um interesse particular em examinar, detalhadamente, a dinâmica da desvalorização e/ou valorização cambial sobre as importações, face à evidência empírica no sentido de que esse efeito possa ser negativo.

Os dados da Tabela ?? do Apêndice A referem-se a importações brasileiras (IM) em milhões de dólares, taxa de câmbio (TCI) e o Produto Interno Bruto representando a renda nacional (RN). Os dados são trimestrais das contas externas do Brasil no período de 1980 a 1998 (Banco Central). A taxa de câmbio representa a relação entre reais e dólar, isto é, quantos reais são gastos para comprar um dólar americano e, por fim, a renda nacional em número índice (dez90 = 100).

Na Figura 6.7, apresentam-se os gráficos de dispersão das variáveis duas a duas para os dados observados brutos e transformados por logaritmo e pelo seu inverso. Verifica-se que existe uma relação não-linear entre IM e as variáveis expla-

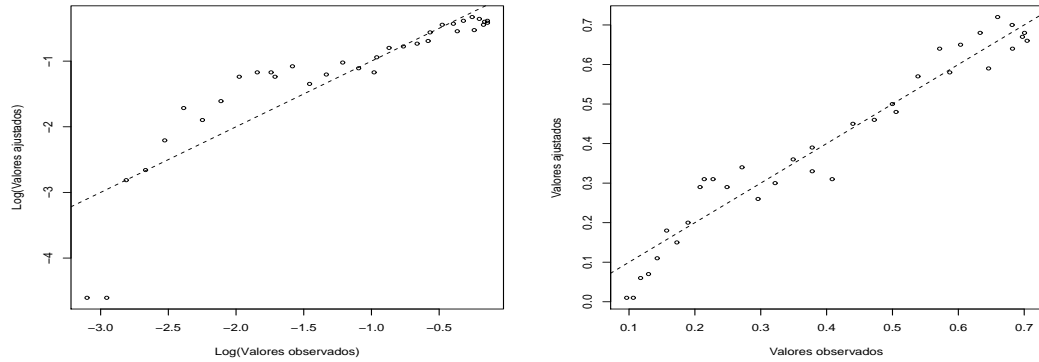


Figura 6.6: Gráficos de valores ajustados *versus* valores observados obtidos para o modelo normal para  $\log(Y)$  com função de ligação identidade e para o modelo normal para  $Y$  com função de ligação logarítmica (dados da Tabela 6.3).

natórias TCI e RN. Essa relação não-linear diminui, porém, não desaparece quando se usa o logaritmo da variável resposta. Entretanto, a transformação  $1/IM$  parece linearizar a relação com as variáveis explanatórias. Nota-se, contudo, a presença de heterogeneidade de variâncias em todos os casos, principalmente, relacionada à variável RN. Portanto, espera-se, de antemão, uma falta de ajuste do modelo com a distribuição normal. A relação entre TCI e RN não parece muito forte e mostra uma variabilidade acentuada.

Como um primeiro modelo ( $M_1$ ), supõe-se que a variável resposta é  $IM = \mu + \varepsilon_1$ , em que  $\varepsilon_1 \sim N(0, \sigma_1^2)$  e, portanto,  $IM \sim N(\mu, \sigma_1^2)$ , a função de ligação é a identidade,  $\eta = \mu$ , e o preditor linear é expresso por

$$\eta_i = \beta_0 + \beta_1 TCI_i + \beta_2 RN_i, \quad i = 1, \dots, 74. \quad (6.2)$$

Como esperado, a Figura 6.8 mostra a falta de ajuste desse modelo, por meio dos gráficos dos valores observados *versus* valores ajustados, resíduos estudentizados *versus* valores ajustados e resíduos estudentizados *versus* TCI e RN. Verifica-se que as suposições de homocedasticidade e independência dos erros aleatórios são violadas, além de indicar a necessidade de um componente não-linear. Como um segundo modelo ( $M_2$ ), adiciona-se  $TCI^2$  e  $RN^2$  ao preditor linear da equação (6.2), obtendo-

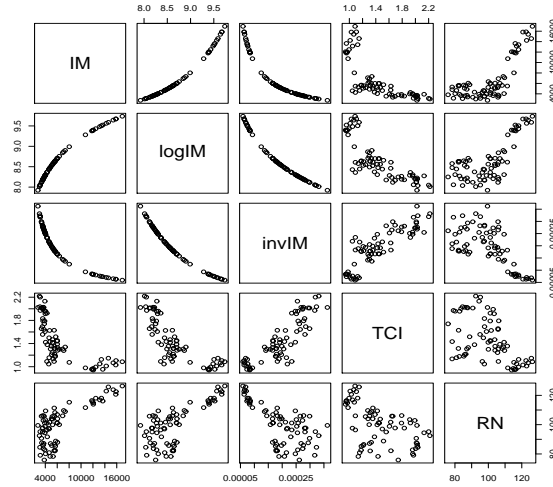


Figura 6.7: Gráficos de dispersão - valores observados e transformados e variáveis explanatórias (dados da Tabela ??).

se

$$\eta_i = \beta_0 + \beta_1 TCI_i + \beta_2 RN_i + \beta_3 TCI^2 + \beta_4 RN^2, \quad i = 1, \dots, 74. \quad (6.3)$$

O problema de não-linearidade desaparece, mas permanece a heterogeneidade das variâncias, como indicado por gráficos de resíduos (não mostrados).

Como um terceiro modelo ( $M_3$ ), supõe-se que  $\log(IM) = \mu + \varepsilon_2$ , em que  $\varepsilon_2 \sim N(0, \sigma_2^2)$  e, portanto,  $\log(IM) \sim N(\mu, \sigma_2^2)$ , que a função de ligação é a identidade,  $\eta = \mu$ , e com preditor linear (6.2). Como um quarto modelo ( $M_4$ ), supõe-se que  $IM = \mu + \varepsilon_3$ , em que  $\varepsilon_3 \sim N(0, \sigma_3^2)$  e, portanto,  $IM \sim N(\mu, \sigma_3^2)$ , a função de ligação é a logarítmica,  $\eta = \log(\mu)$ , e com preditor linear (6.2). Gráficos de resíduos (não mostrados) indicam que os modelos  $M_3$  e  $M_4$  não se ajustam bem aos dados. O gráfico normal de probabilidades (Figura 6.9) confirma a falta de um bom ajuste dos modelos  $M_1$ ,  $M_3$  e  $M_4$ .

É importante observar que o desvio para modelos com distribuições contínuas depende da escala dos dados como mostra a Tabela 6.4, não servindo, portanto, como estatística para verificar o ajuste do modelo. No caso da distribuição normal, com função de ligação identidade, os desvios residuais são denominados de

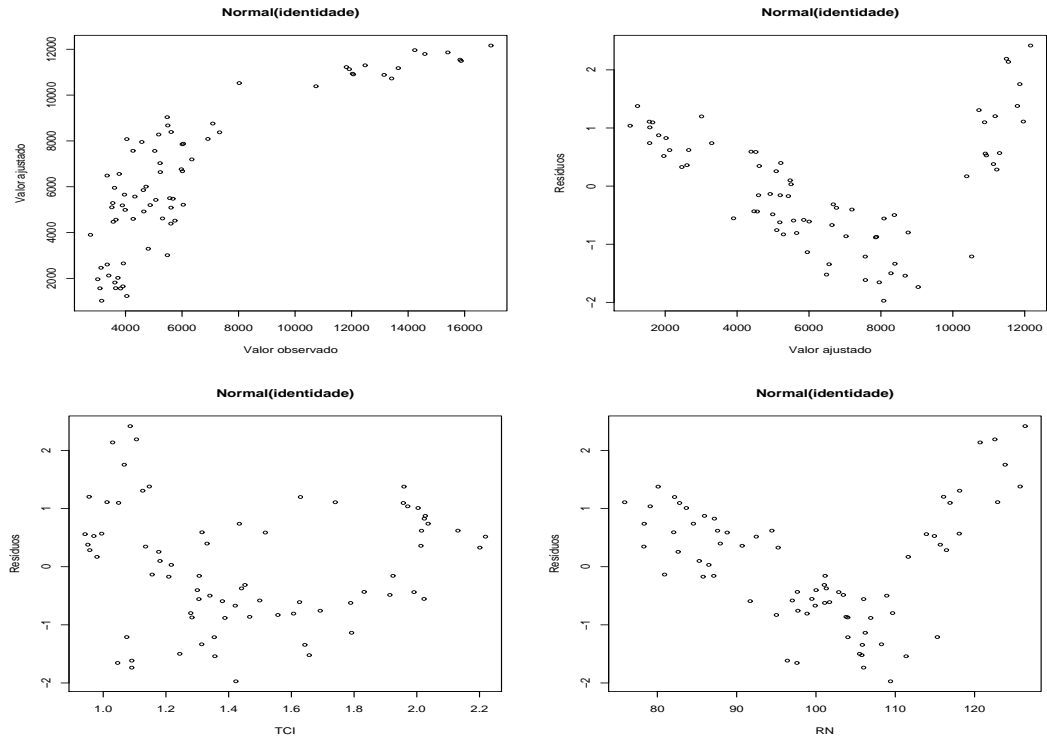


Figura 6.8: Gráficos para verificação do ajuste do modelo  $M_1$  (dados da Tabela ??).

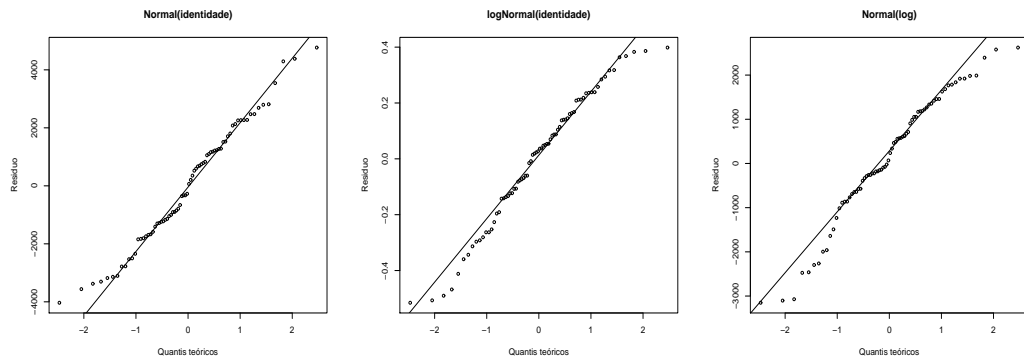


Figura 6.9: Gráficos normais de probabilidade, modelos  $M_1$ ,  $M_3$  e  $M_4$  (dados da Tabela ??).

somas de quadrados residuais. Verifica-se, também, que a estatística AIC depende da escala da variável resposta (IM ou  $\log(\text{IM})$ ). Em geral, os valores do logaritmo da função de verossimilhança e, portanto, da estatística AIC, são calculados sem os

termos constantes e para escalas diferentes não há uma forma de normalizá-los de forma a serem comparáveis. Valores de AIC são apenas comparáveis na mesma escala da variável resposta. O uso de diferentes funções de ligação não altera o número de parâmetros mas muda o aspecto estrutural do modelo, modificando a função de verossimilhança. Entretanto, é possível normalizar a função de verossimilhança de modo que comparações para diferentes funções de ligação tenham sentido.

Tabela 6.4: Resumo do ajuste do modelo normal para diferentes funções de ligação.

Modelo	Variável resposta	F. Ligação	Desvio Residual	$\hat{\sigma}$	AIC
$M_1$	IM	$\mu$	315765900	2108,9	1347,7
$M_3$	$\log(\text{IM})$	$\mu$	4,0	0,2382	2,6
$M_4$	IM	$\log(\mu)$	146543227	1436,7	1290,9

Tabela 6.5: Resumo do ajuste do modelo gama para funções de ligação inversa (com dois preditores lineares diferentes), logarítmica e identidade.

Modelo	F. Ligação	$\hat{\phi}$	Desvio	AIC
$M_5$	$1/\mu$	0,0307	2,294	1240,9
$M_8$	$1/\mu$	0,0236	1,705	1220,9
$M_6$	$\log(\mu)$	0,0524	3,908	1280,6
$M_7$	$\mu$	0,0892	6,191	1315,0

Modelos alternativos podem ser usados, supondo-se que  $IM_i \sim G(\mu_i, \phi)$  com as funções de ligação canônica (inversa,  $M_5$ ), logarítmica ( $M_6$ ) e identidade ( $M_7$ ) e preditor linear (6.2). A Tabela 6.5 mostra um resumo, considerando o ajuste para esses três casos e para o caso da distribuição gama e função de ligação canônica (inversa), acrescentando-se  $RN^2$  no preditor linear ( $M_8$ ). Observa-se que os modelos com menores AIC são aqueles com distribuição gama e função de ligação canônica (inversa). A Figura 6.10 apresenta os gráficos dos valores ajustados *versus* valores observados e os gráficos normais de probabilidade dos modelos  $M_5$ ,  $M_6$  e  $M_7$ .

A Figura 6.11 apresenta os gráficos dos valores ajustados *versus* valores observados, dos resíduos *versus* valores ajustados e o normal de probabilidade para



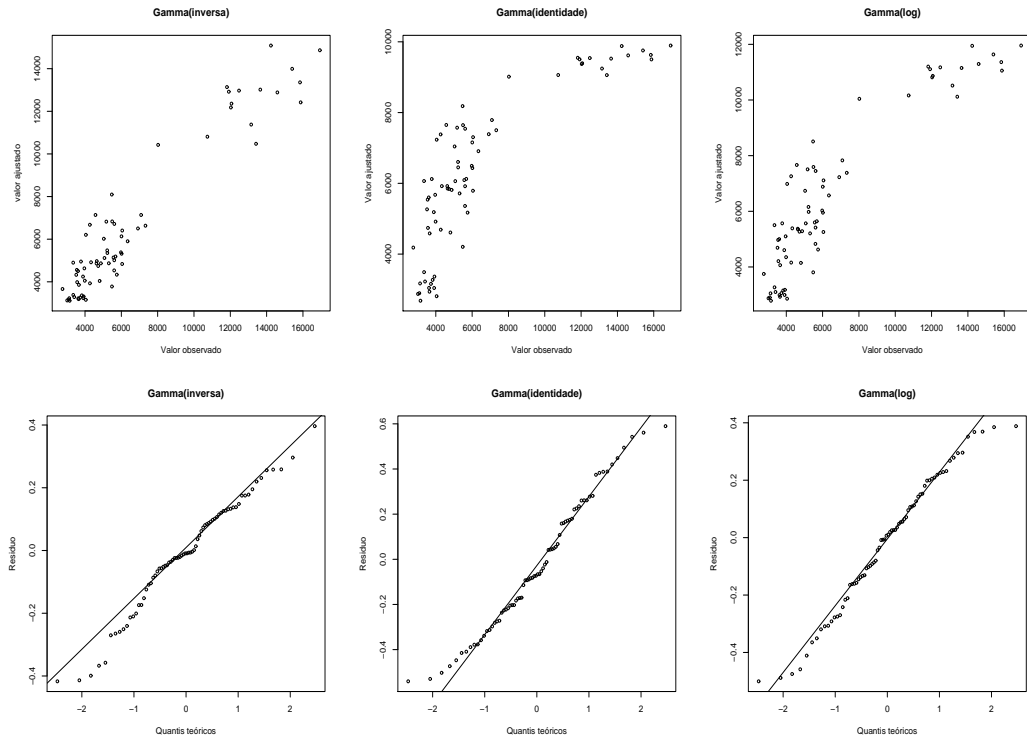


Figura 6.10: Gráficos dos valores ajustados *versus* valores observados, modelos  $M_5$ ,  $M_6$  e  $M_7$  (dados da Tabela ??).

o modelo  $M_8$ . Entretanto, seria interessante completar com estudos de simulação e testes “bootstrap” para a escolha do melhor modelo. Outros estudos referentes a pontos discrepantes e/ou influentes são necessários. Um resumo das estatísticas para o modelo escolhido encontra-se na Tabela 6.5.

Note que para o teste das variáveis que entrarão no modelo são usados polinômios ortogonais. Entretanto, para obtenção dos coeficientes, há necessidade de se usar o preditor linear (6.2) sendo, portanto, estimado por

$$\hat{\eta}_i = \hat{\mu}_i^{-1} = -0,0004888 + 0,0001393 \text{ TCI} + 0,00001180 \text{ RN} - 0,00000006903 \text{ RN}^2.$$

O programa para as análises foi desenvolvido em R e encontra-se no Apêndice B.3.

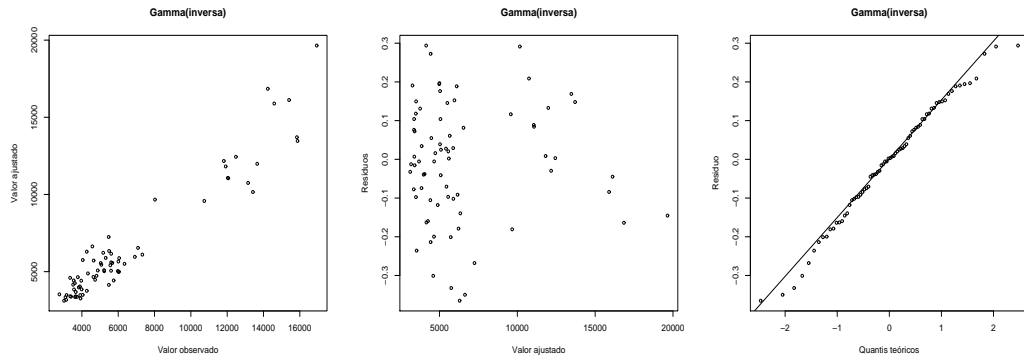


Figura 6.11: Gráficos dos valores ajustados *versus* valores observados, dos valores ajustados *versus* resíduos e o normal de probabilidade para o modelo  $M_8$  (dados da Tabela ??).

Tabela 6.6: Resumo do ajuste do modelo gama com função de ligação inversa  $1/\mu$  ( $M_8$ ).

Parâmetro	Estimativa	e.p.	t	$\Pr(> t )$
(Intercepto)	-1,188e-05	1,574e-05	-0,755	0,453
TCI	1,393e-04	1,162e-05	11,990	< 2e-16 ***
poly(RN,2)1	-2,383e-04	2,987e-05	-7,977	2,01e-11 ***
poly(RN,2)2	-1,076e-04	2,100e-05	-5,124	2,53e-06 ***

## 6.4 Dados de tempos de sobrevivência de ratos

Os dados da Tabela 6.7 do Apêndice A referem-se a tempos de sobrevivência de ratos após envenenamento com quatro tipos de venenos e três diferentes tratamentos (Box e Cox, 1964). Como pode ser constatado na Figura 6.12, os dados sem transformação apresentam heterogeneidade de variâncias que é amenizada quando se usam os inversos dos valores observados ou os valores observados elevados à potência  $-3/4$ .

Seja, inicialmente, o modelo

$$Y_{ij} = \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ij},$$

em que  $Y_{ij}$  representa o tempo de sobrevivência do rato que recebeu o veneno  $i$  e o tratamento  $j$ ,  $\alpha_i$  representa o efeito do veneno  $i$ ,  $\beta_j$  representa o efeito do tratamento  $j$ ,

Tabela 6.7: Tempos de sobrevivência de ratos após envenenamento.

Tempo	Tipo	Trat.	Tempo	Tipo	Trat.	Tempo	Tipo	Trat.	Tempo	Tipo	Trat.
0,31	1	1	0,45	1	1	0,46	1	1	0,43	1	1
0,82	1	2	1,10	1	2	0,88	1	2	0,72	1	2
0,43	1	3	0,45	1	3	0,63	1	3	0,76	1	3
0,45	1	4	0,71	1	4	0,66	1	4	0,62	1	4
0,36	2	1	0,29	2	1	0,4	2	1	0,23	2	1
0,92	2	2	0,61	2	2	0,49	2	2	1,24	2	2
0,44	2	3	0,35	2	3	0,31	2	3	0,40	2	3
0,56	2	4	1,02	2	4	0,71	2	4	0,38	2	4
0,22	3	1	0,21	3	1	0,18	3	1	0,23	3	1
0,30	3	2	0,37	3	2	0,38	3	2	0,29	3	2
0,23	3	3	0,25	3	3	0,24	3	3	0,22	3	3
0,30	3	4	0,36	3	4	0,31	3	4	0,33	3	4

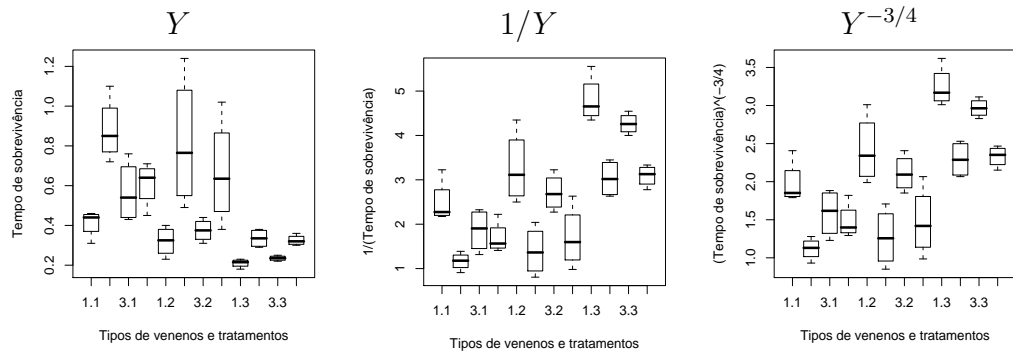


Figura 6.12: Box-plots para as observações da Tabela 6.7, com e sem transformação.

$\alpha\beta_{ij}$  representa o efeito da interação do veneno  $i$  com o tratamento  $j$ ,  $\varepsilon_{ij} \sim N(0, \sigma^2)$ . O gráfico para verificar a necessidade de uma transformação na família Box-Cox indica que  $\hat{\lambda} = -0,75$ , conforme mostra a Figura 6.13. Entretanto, o valor  $\hat{\lambda} = -1$  está no intervalo de confiança e  $1/Y$  tem uma melhor interpretação nesse caso, isto é, representa a taxa de mortalidade.

Ajustando-se, também, os modelos

$$\frac{1}{Y_{ij}} = \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ij}$$

e

$$Y_{ij}^{-3/4} = \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ij} + \delta_{ij},$$

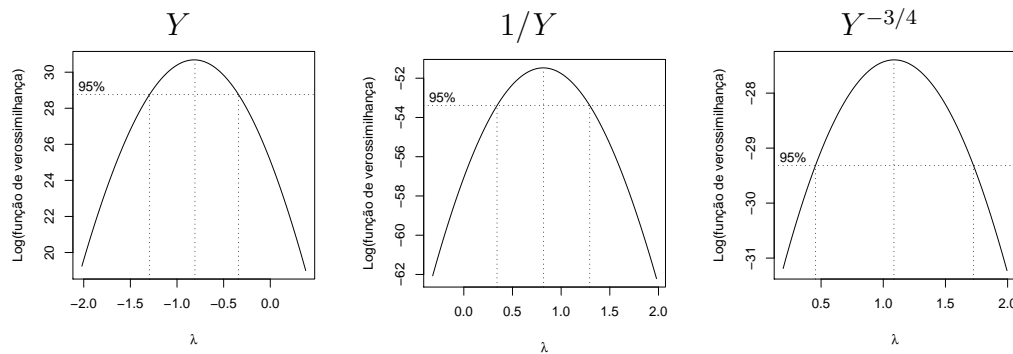


Figura 6.13: Gráficos para escolha de transformação na família Box-Cox (dados da Tabela 6.7).

em que  $\epsilon_{ij} \sim N(0, \tau^2)$  e  $\delta_{ij} \sim N(0, \zeta^2)$ , obtêm-se os outros dois gráficos da Figura 6.13, mostrando que o valor  $\hat{\lambda} = 1$  está incluído no intervalo de confiança e que, portanto, ambas as transformações tornam a escala da variável resposta adequada. A Figura 6.14 mostra os gráficos dos valores ajustados *versus* valores observados sem e com transformação, dos valores ajustados *versus* resíduos e gráficos normais de probabilidades. Esses gráficos revelam, claramente, a falta de ajuste para o caso do modelo normal para a variável sem transformação e que ambas as transformações resolvem o problema de heterogeneidade de variâncias e da falta de normalidade da variável resposta. Outros modelos, supondo distribuição normal com função de ligação inversa, distribuições gama e normal inversa, foram usados e apresentaram resultados piores.

Os resultados da Tabela 6.8 mostram que, em ambos os casos, existem evidências do efeito significativo do tipo de veneno e do tratamento mas não da interação entre eles. Entretanto, a evidência é muito mais forte para o caso em que foram feitas as transformações  $1/Y$  e  $Y^{-3/4}$ . O programa para as análises foi desenvolvido em R e encontra-se no Apêndice B.4.

## 6.5 Dados de assinaturas de TV a cabo

Os dados da Tabela ?? do Apêndice A referem-se ao número de assinantes (em milhares) de TV a Cabo ( $y$ ) em 40 áreas metropolitanas (Ramanathan, 1993), número de domicílios (em milhares) na área ( $x_1$ ), renda per capita (em US\$) por domicílio com TV

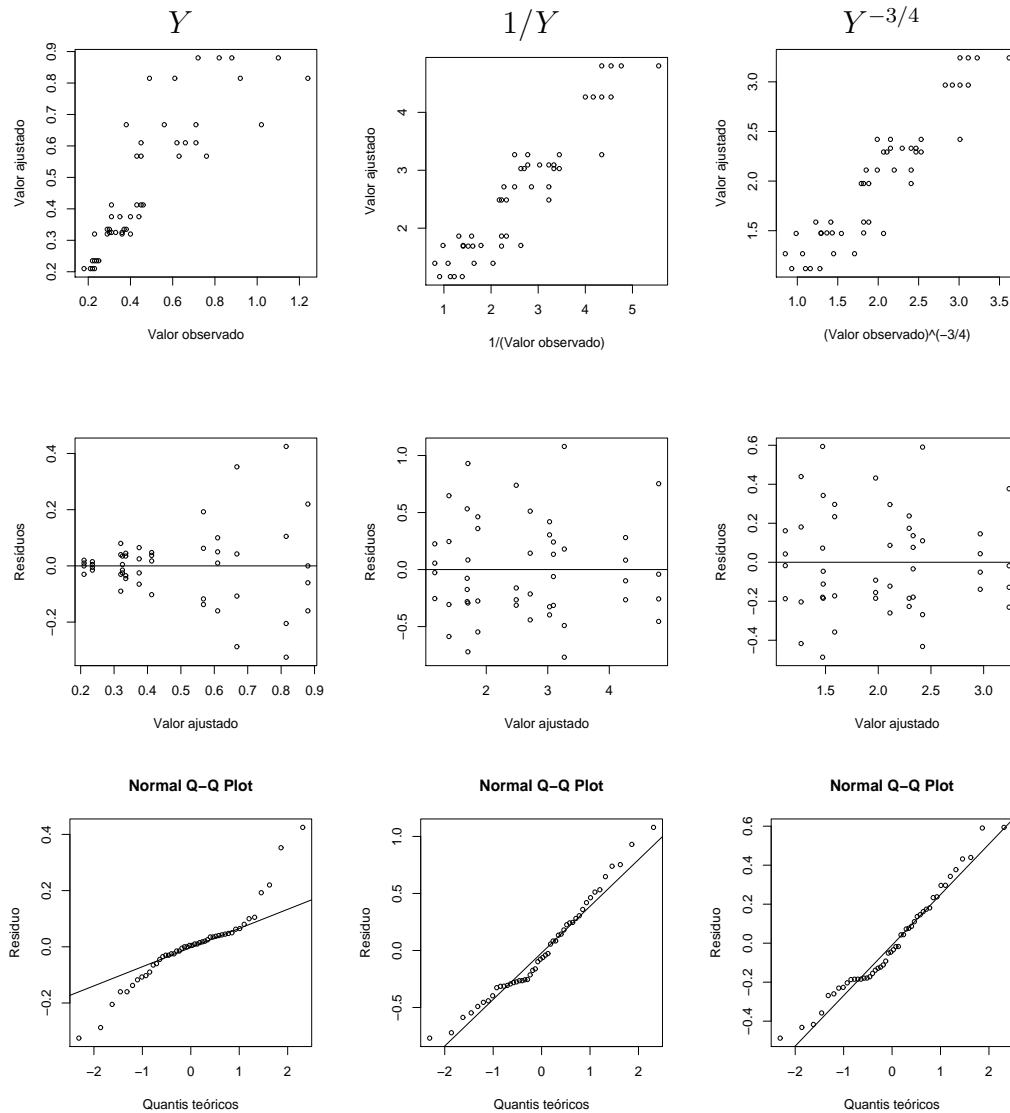


Figura 6.14: Gráficos dos valores ajustados *versus* valores observados sem e com transformação, dos resíduos *versus* valores ajustados e gráficos normais de probabilidades (dados da Tabela 6.7).

a cabo ( $x_2$ ), taxa de instalação ( $x_3$ ), custo médio mensal de manutenção ( $x_4$ ), número de canais a cabo disponíveis na área ( $x_5$ ) e número de canais não pagos com sinal de boa qualidade disponíveis na área ( $x_6$ ). O interesse está em analisar o número de assinantes (variável resposta) como função das demais variáveis explanatórias.

A Figura 6.15 mostra os gráficos de dispersão das variáveis duas a duas sem

Tabela 6.8: Análise de variância para os tempos de sobrevivência de ratos após envenenamento, sem e com transformação inversa, descritos na Tabela 6.7.

Fonte	GL	Tempo			1/Tempo			Tempo <sup>-3/4</sup>		
		SQ	QM	F	SQ	QM	F	SQ	QM	F
Tipo	2	1,0330	0,5165	23,27**	34,877	17,439	72,46**	11,9261	5,9630	68,45**
Tratamento	3	0,9212	0,3071	16,71**	20,414	6,805	28,35**	7,1579	2,3860	27,39**
Interação	6	0,2501	0,0417	1,88	1,571	0,262	1,09	0,4859	0,0810	0,93
Resíduo	36	0,8007	0,0222		8,643	0,240		3,1361	0,0871	

transformação e com transformação logarítmica. Nota-se que existe uma relação linear forte entre  $\log(y)$  e  $\log(x_1)$  e mais fraca de  $y$  e  $x_1$  e de  $\log(y)$  e  $\log(x_5)$ , sem muita evidência de relação entre  $y$  ou  $\log(y)$  e as outras variáveis explanatórias. Há evidências, também, de relação entre as variáveis  $x_2$ ,  $x_4$ ,  $x_5$  e  $x_6$ , o que pode mascarar a relação entre a variável resposta  $Y$  e as variáveis explanatórias.

A variável resposta  $Y$ , embora discreta, tem valores muito grandes o que justifica o uso da distribuição normal para  $Y$  ou  $\log(Y)$ . Supondo-se que as demais variáveis sejam constantes, espera-se que o número de assinaturas de TV a cabo seja proporcional ao número de domicílios na área ( $x_1$ ). Além disso, espera-se que outras variáveis afetem a média de uma forma multiplicativa. Isso sugere um modelo linear para  $\log(Y)$  com pelo menos  $\log(x_1)$  como um dos preditores com um coeficiente próximo de um, de acordo com a Figura 6.15.

Ajustando-se, aos dados da Tabela ??, o modelo  $M_1$

$$Y_i = \beta_0 + \beta_1 \log(x_{1i}) + \beta_2 \log(x_{2i}) + \beta_3 \log(x_{3i}) + \beta_4 \log(x_{4i}) + \beta_5 \log(x_{5i}) + \beta_6 \log(x_{6i}) + \varepsilon_i,$$

em que  $\varepsilon_i \sim N(0, \sigma^2)$ , o gráfico na Figura 6.16 para a família de transformações Box-Cox evidencia a necessidade da transformação  $\log(Y)$ , pois  $\lambda = 0$  pertence ao intervalo de confiança.

Considerando-se o modelo  $M_2$

$$\log(Y_i) = \beta_0 + \beta_1 \log(x_{1i}) + \beta_2 \log(x_{2i}) + \beta_3 \log(x_{3i}) + \beta_4 \log(x_{4i}) + \beta_5 \log(x_{5i}) + \beta_6 \log(x_{6i}) + \epsilon_i,$$

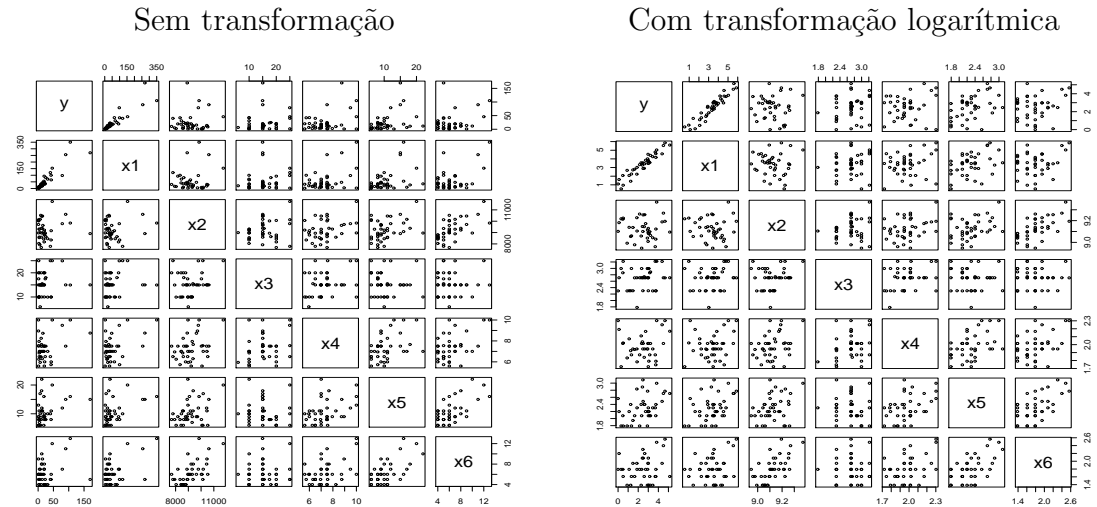


Figura 6.15: Gráfico de dispersão - valores observados e transformados na escala logarítmica (dados da Tabela ??).

em que  $\epsilon_i \sim N(0, \tau^2)$ , verifica-se um melhor ajuste aos dados da Tabela ?. As estimativas dos parâmetros para o modelo  $M_2$  com seus erros-padrão encontram-se na Tabela 6.9, revelando a não significância de  $\log(x_{3i})$ .

Tabela 6.9: Resumo do ajuste do modelo  $M_2$  com e sem  $\log(x_3)$  (Dados da Tabela ??).

Parâmetro	Estimativa	e.p.	t	Pr(> t )	Estimativa	e.p.	t	Pr(> t )
Intercepto	-16,44	7,46	-2,20	0,03 *	-18,27	7,45	-2,45	0,02 *
$\log(x_1)$	0,96	0,05	17,67	<0,001 ***	0,98	0,05	18,38	<0,001 ***
$\log(x_2)$	1,86	0,86	2,16	0,04 *	2,12	0,85	2,50	0,02 *
$\log(x_3)$	0,28	0,20	1,40	0,17	—	—	—	—
$\log(x_4)$	-1,10	0,45	-2,44	0,02 *	-0,90	0,43	-2,08	0,04 *
$\log(x_5)$	0,61	0,23	2,61	0,01 *	0,51	0,23	2,26	0,03 *
$\log(x_6)$	-0,70	0,28	-2,49	0,02 *	-0,73	0,28	-2,59	0,01 *

Eliminando-se  $\log(x_3)$  do modelo  $M_2$ , as estimativas dos parâmetros com seus erros-padrão estão na Tabela 6.9. Os gráficos na Figura 6.17 dos valores ajustados *versus*  $\log(y)$ , dos resíduos *versus* valores ajustados e dos quantis observados *versus* quantis teóricos revelam o bom ajuste desse modelo. No gráfico dos resíduos *versus* valores ajustados, há três observações discrepantes, 11, 14 e 26.

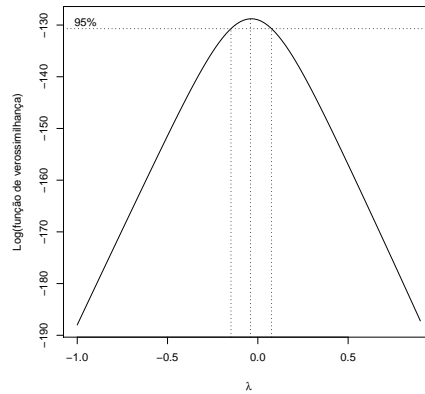


Figura 6.16: Gráfico dos valores ajustados *versus* valores observados e gráfico para a família de transformações Box-Cox, modelo ( $M_1$ ) (dados da Tabela ??).

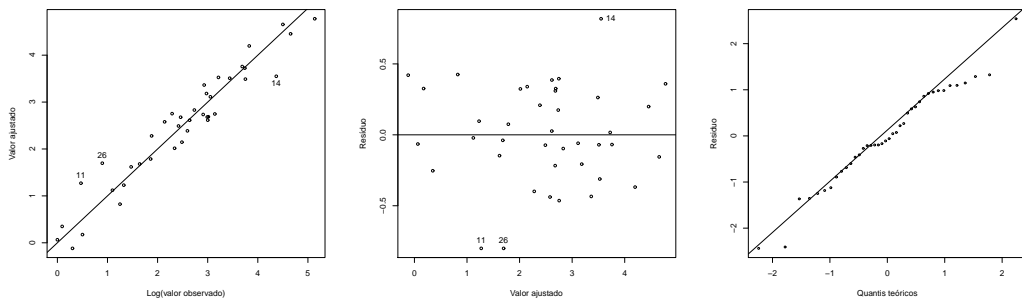


Figura 6.17: Gráfico dos valores ajustados *versus* log(valores observados), dos resíduos *versus* valores ajustados e dos quantis observados *versus* quantis teóricos, modelo  $M_2$ , sem  $\log(x_3)$  (dados da Tabela ??).

Logo, o valor estimado de  $y_i$  pode ser calculado por

$$\hat{\mu}_i = \exp[-18,27 + 0,98 \log(x_{1i}) + 2,12 \log(x_{2i}) - 0,90 \log(x_{4i}) + 0,51 \log(x_{5i}) - 0,73 \log(x_{6i})].$$

Nota-se que o coeficiente de  $\log(x_1)$  está muito próximo de um, como previsto. Tem-se, portanto, evidência de que aumentando-se o número de domicílios (em milhares) na área ( $x_1$ ), a renda per capita (em US\$) por domicílio com TV a cabo ( $x_2$ ) e o número de canais a cabo disponíveis na área ( $x_5$ ) há um aumento no número de assinantes e, também, que o aumento no custo médio mensal de manutenção ( $x_4$ ) e o aumento do número de canais não pagos com sinal de boa qualidade disponíveis na área ( $x_6$ ) causam



um decréscimo no número de assinantes. Assim, para esse modelo, tem-se que para cada aumento de uma unidade no número de domicílios em unidade logarítmica, há um aumento de 0,98 unidades no número de assinantes, mantidas constantes as demais covariáveis. Entretanto, para cada aumento de uma unidade no número de canais não pagos com sinal de boa qualidade disponíveis na área ( $x_6$ ), corresponderá uma redução de 0,73 unidades no número de assinantes. O programa para as análises foi desenvolvido em R e encontra-se no Apêndice B.5.

## 6.6 Dados de demanda de energia elétrica

Os dados da Tabela ?? do Apêndice A referem-se à variável resposta demanda de eletricidade agregada per capita para o setor residencial (ELAR), como função das variáveis explanatórias preço médio da eletricidade para o setor residencial (PER), preço do gás natural para o setor residencial (PGR) e renda per capita (RECA). As variáveis binárias D1, D2, D3 e D4 representam os trimestres, enquanto que a variável Tri representa o fator trimestre. Os dados foram coletados no primeiro trimestre de 1961 até o quarto trimestre de 1983, com o total de 92 observações.

A Figura 6.18 mostra a dispersão da variável resposta eletricidade agregada per capita para o setor residencial (ELAR) *versus* as demais variáveis preço médio da eletricidade para o setor residencial (PER), preço do gás natural para o setor residencial (PGR) e renda per capita (RECA). Nota-se que existe uma relação forte entre ELAR e PER e entre ELAR e RECA que depende do trimestre. Há evidência de uma demanda maior de eletricidade nos trimestres 1 (verão) e 3 (inverno) e menor nos trimestres 2 (outono) e 4 (primavera). Além disso, há um crescimento de demanda de eletricidade ao longo dos anos com aumento da variabilidade. Há evidências, também, de relação entre as variáveis PER e RECA, o que pode mascarar a relação entre a variável resposta e as variáveis explanatórias.

O ajuste do modelo será iniciado usando erro normal e as ligações identidade e logarítmica, respectivamente. Abaixo temos o resultados do valor do desvio e graus de liberdade do modelo. Observa-se através da Tabela 1 que o modelo que melhor ajusta os dados é o com erro normal e ligação logarítmica (desvio = 0.17169).

*Tabela 1 - Desvio e graus de liberdade dos*

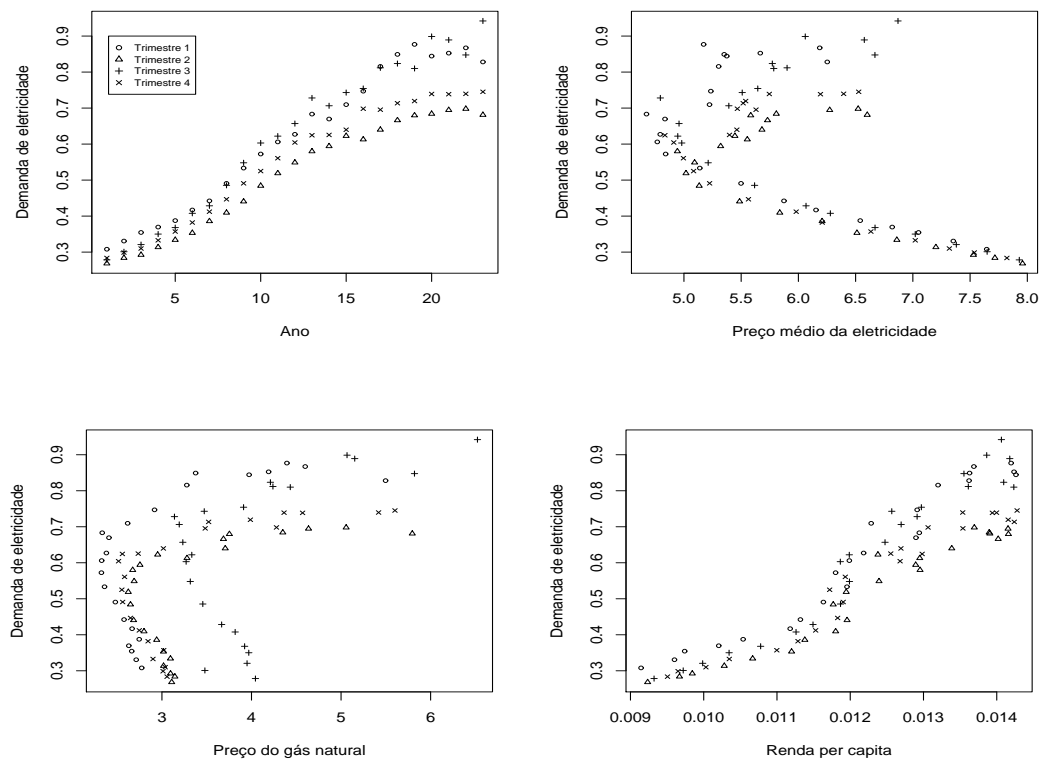


Figura 6.18: Gráficos de dispersão dos valores observados da demanda de eletricidade *versus* valores observados do preço médio da eletricidade, do preço do gás natural e da renda per capita (dados da Tabela ??).

*modelos ajustados erro normal.*

Ligação	Desvio	DF
Identidade	0.21417	85
Logarítmica	0.17169	85

Abaixo temos as estimativas do modelo com erro normal e ligação logarítmica. Observe que todas as variáveis foram significativas ao nível de 5% de significância.

```
>fit2<-glm(elar~per+pgr+reca+DD1+DD2+DD3,
family = gaussian(link = "log"))
> summary(fit2)
```

```
Call: glm(formula = elar ~ per + pgr + reca + DD1 + DD2 + DD3,
family = gaussian(link = "log"))
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-8.904e-02	-3.255e-02	8.283e-05	2.854e-02	1.037e-01

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-2.22778	0.23949	-9.302	1.32e-14 ***
per	-0.11247	0.02396	-4.694	1.02e-05 ***
pgr	0.02012	3.628	0.000486	163.04261 14.15700
reca	11.517	< 2e-16 ***	DD1	0.12624 0.02217 5.693
DD1	1.74e-07 ***	DD2	-0.04949	0.02409 -2.054 0.043050 *
DD2	DD3	0.11021	0.02369	4.652 1.20e-05 ***
DD3				

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.002019894)

Null deviance: 3.20412 on 91 degrees of freedom  
Residual deviance: 0.17169 on 85 degrees of freedom AIC: -301.03

Number of Fisher Scoring iterations: 4

Para verificar se o modelo ajustado é razoável, fazemos o gráfico de dispersão dos valores observados versus os valores ajustados. Observa-se através da Figura 1 que é razoável o ajuste aplicado. Além disso, ainda na Figura 1, temos que a distribuição normal para os erros ordenados de Anscombe é aceita; os resíduos de Pearson apresentam uma distribuição aleatória quando feita a sua dispersão versus os valores ajustados, indicando assim, que os resíduos são não correlacionados, ou seja, a hipótese de independência e

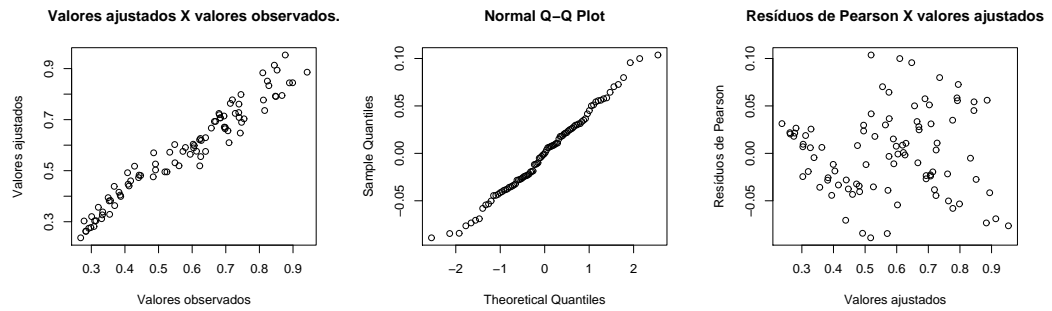


Figura 6.19: Gráfico dos valores ajustados *versus*  $\log(\text{valores observados})$ , dos resíduos *versus* valores ajustados e dos quantis observados *versus* quantis teóricos, modelo  $\log(y_i) = \beta_0 + \beta_1 \log(x_1) + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \epsilon_i$ , Tabela ??.

variância constante para os resíduos são aceitas.

Assim, concluímos que o modelo para ajustar a demanda de energia elétrica é dada por:

$$\log(\text{elar}) = -2.228 - 0.1125\text{per} + 0.073\text{pgr} + 163\text{reca} + 0.1262\text{DD1} - 0.04949\text{DD2} + 0.1102\text{DD3}$$

# Capítulo 7

## Aplicações a Dados Discretos

Neste capítulo, serão apresentadas diversas aplicações com dados na forma de proporções e na forma de contagens. Os programas em R estão no Apêndice B.

### 7.1 Dados binários e proporções

#### 7.1.1 Estimação da dose efetiva e seu intervalo de confiança

Como foi descrito na Seção 2.2 do Capítulo 2, ensaios do tipo dose-resposta são muito usados na área de toxicologia. Em geral, os dados resultantes são proporções e os modelos mais usados são logístico, probito e complemento log-log, que no caso em que o preditor linear é uma regressão linear simples, podem ser expressos por

$$F^{-1}(\pi_i) = \beta_0 + \beta_1 x_i, \quad (7.1)$$

em que  $\pi_i$  é a probabilidade de sucesso do evento sob estudo,  $F(\cdot)$  uma f.d.a. de interesse e  $x_i$  é a variável explanatória. Esses modelos, ajustados a conjuntos de dados, podem ser usados para sumarizá-los, nesse caso, pelo par de estimativas  $(\hat{\beta}_0, \hat{\beta}_1)$  dos parâmetros e formam a base para comparação de diferentes conjuntos de dados (Morgan, 1992). Assim, por exemplo, podem ser usados para a comparação de potência de diferentes produtos (inseticidas, fungicidas, herbicidas, etc).

Em geral, porém, o interesse está na determinação de estimativas de doses efetivas,  $\theta_p$  ( $DE_{100p}$ ), que são doses que, sob o modelo ajustado, causam uma mudança de estado em 100p% dos indivíduos. Um exemplo muito comum é a determinação da  $DL_{50}$  (também chamada dose mediana) que é a dose que causa 50% de uma mudança de estado (por

exemplo, mortalidade) dos indivíduos. De (7.1) para um valor  $p$  especificado, tem-se

$$F^{-1}(p) = \beta_0 + \beta_1 \theta_p, \quad (7.2)$$

sendo que  $\theta_p$  representa a dose efetiva. Portanto, de uma forma geral, a estimativa da dose efetiva  $\theta_p$  é calculada usando-se a expressão

$$\hat{\theta}_p = \frac{F^{-1}(p) - \hat{\beta}_0}{\hat{\beta}_1} = g(\hat{\beta}_0, \hat{\beta}_1), \quad (7.3)$$

que para os modelos comumente usados transforma-se em

$$\text{Logístico : } \text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \hat{\beta}_0 + \hat{\beta}_1 \hat{\theta}_p \Rightarrow \hat{\theta}_p = \frac{1}{\hat{\beta}_1} \left[ \log\left(\frac{p}{1-p}\right) - \hat{\beta}_0 \right]$$

$$\text{Probit : } \text{probit}(p) = \Phi^{-1}(p) = \hat{\beta}_0 + \hat{\beta}_1 \hat{\theta}_p \Rightarrow \hat{\theta}_p = \frac{1}{\hat{\beta}_1} [\Phi^{-1}(p) - \hat{\beta}_0]$$

$$\text{Clog-log : } \log[-\log(1-p)] = \hat{\beta}_0 + \hat{\beta}_1 \hat{\theta}_p \Rightarrow \hat{\theta}_p = \frac{1}{\hat{\beta}_1} \{\log[-\log(1-p)] - \hat{\beta}_0\}$$

$$\text{Aranda-Ordaz : } \log\left[\frac{1 - (1-p)^\lambda}{\lambda(1-p)^\lambda}\right] = \hat{\beta}_0 + \hat{\beta}_1 \hat{\theta}_p \Rightarrow \hat{\theta}_p = \frac{1}{\hat{\beta}_1} \left\{ \log\left[\frac{1 - (1-p)^\lambda}{\lambda(1-p)^\lambda}\right] - \hat{\beta}_0 \right\}.$$

O modelo de Aranda-Ordaz tem como casos particulares o modelo logístico ( $\lambda = 1$ ) e o modelo complemento log-log ( $\lambda \rightarrow 0$ ). Uma estimativa de  $\lambda$  pode ser obtida pelo método de perfil de verossimilhança.

Se  $p = 0,50$ , verifica-se que, para qualquer modelo simétrico, portanto, incluindo os modelos logístico e probito, a dose efetiva é obtida por

$$\hat{\theta}_{50} = -\frac{\hat{\beta}_0}{\hat{\beta}_1}$$

enquanto que para o modelo complemento log-log é expressa por

$$\hat{\theta}_{50} = \frac{\log(\log 2) - \hat{\beta}_0}{\hat{\beta}_1}$$

e, para o modelo de Aranda-Ordaz, como

$$\hat{\theta}_{50} = \frac{1}{\hat{\beta}_1} \left[ \log\left(\frac{2^\lambda - 1}{\lambda}\right) - \hat{\beta}_0 \right].$$

É importante notar que se o modelo está como função do logaritmo, em uma base  $b$  qualquer, da dose, então,  $\hat{\theta}_p = \log_b(\hat{d}_p)$  e, portanto, a dose efetiva é obtida fazendo-se  $\hat{d}_p = b^{\hat{\theta}_p}$ .

A estimativa por ponto de uma dose efetiva nem sempre é suficiente, havendo necessidade de se obterem intervalos de confiança para  $\theta_p$ . Entretanto,  $\hat{\theta}_p$ , calculada pela expressão (7.3), é um quociente de variáveis aleatórias que, assintoticamente, têm distribuição normal, isto é,  $\hat{\beta}_0 \sim N(\beta_0, \text{Var}(\hat{\beta}_0))$ ,  $\hat{\beta}_1 \sim N(\beta_1, \text{Var}(\hat{\beta}_1))$  e  $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \neq 0$ , isto é,  $\hat{\beta} \sim N(\beta, \mathbf{V})$  (Seção 4.1), em que  $\mathbf{V} = \text{Cov}(\hat{\beta})$  é a matriz de variâncias e covariâncias dos estimadores dos parâmetros (inversa da matriz de informação de Fisher). Os métodos aproximados mais comumente usados para a construção de intervalos de confiança para doses efetivas são: o método Delta, o de Fieller e o da razão de verossimilhanças (Morgan, 1992; Collet, 2002).

### Método Delta

A variância assintótica do EMV de uma função escalar  $g(\hat{\beta})$  de um vetor  $\hat{\beta}$ , de dimensão  $p$ , de estimativas de parâmetros desconhecidos, quando a matriz de covariâncias de  $\hat{\beta}$  é conhecida, pode ser calculada usando-se o método delta. Esse método é baseado na expansão de Taylor até primeira ordem e supõe que, segundo condições gerais de regularidade, a distribuição assintótica do EMV  $\hat{\beta}$  é  $N_p(\beta, \mathbf{V})$ , sendo  $\mathbf{V}$  obtida pela inversa da matriz de informação.

Tem-se, supondo que as derivadas parciais  $\partial g(\beta)/\partial \beta_r$  são contínuas e não todas nulas em  $\hat{\beta}$ ,

$$g(\hat{\beta}) \xrightarrow{D} N(g(\beta), \sigma^2),$$

em que  $\sigma^2 = \text{Var}[g(\hat{\beta})] = \gamma^T \mathbf{V}_\beta \gamma$  e  $\gamma = (\partial g/\partial \beta_1, \dots, \partial g/\partial \beta_p)^T$ . Na prática,  $g(\beta)$ ,  $\mathbf{V}$  e  $\gamma$  são estimados em  $\hat{\beta}$ , para que sejam realizados testes de hipóteses e contruídos intervalos de confiança para  $g(\beta)$ , baseando-se na aproximação normal  $N(g(\beta), \sigma^2)$ .

Ilustra-se, agora, o método delta para estimar a variância da dose efetiva de um tratamento correspondente a uma taxa especificada 100p% de mortalidade em um experimento de dose-resposta. Suponha que a matriz de informação  $\mathbf{K}$  de  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)^T$  é estimada. Seja  $\hat{\mathbf{K}}^{-1}$  a sua inversa, especificada por

$$\hat{\mathbf{K}}^{-1} = \hat{\mathbf{V}}_\beta = \begin{bmatrix} \hat{\kappa}^{\beta_0, \beta_0} & \hat{\kappa}^{\beta_0, \beta_1} \\ \hat{\kappa}^{\beta_0, \beta_1} & \hat{\kappa}^{\beta_1, \beta_1} \end{bmatrix} = \begin{bmatrix} \hat{v}_{11} & \hat{v}_{12} \\ \hat{v}_{12} & \hat{v}_{22} \end{bmatrix}.$$

De acordo com o método delta, fazendo-se uma expansão de Taylor de primeira

ordem para a expressão (7.3) de  $g(\hat{\beta}_0, \hat{\beta}_1)$  em torno de  $(\beta_0, \beta_1)$ , tem-se

$$\hat{\theta}_p = g(\hat{\beta}_0, \hat{\beta}_1) \approx g(\beta_0, \beta_1) + (\hat{\beta}_0 - \beta_0) \frac{\partial g(\hat{\beta}_0, \hat{\beta}_1)}{\partial \hat{\beta}_0} \Big|_{(\beta_0, \beta_1)} + (\hat{\beta}_1 - \beta_1) \frac{\partial g(\hat{\beta}_0, \hat{\beta}_1)}{\partial \hat{\beta}_1} \Big|_{(\beta_0, \beta_1)},$$

$$\text{em que } \gamma^T = \left( \frac{\partial g(\hat{\beta}_0, \hat{\beta}_1)}{\partial \hat{\beta}_0} \Big|_{(\beta_0, \beta_1)}, \frac{\partial g(\hat{\beta}_0, \hat{\beta}_1)}{\partial \hat{\beta}_1} \Big|_{(\beta_0, \beta_1)} \right) = \left( -\frac{1}{\beta_1}, -\frac{F^{-1}(p) - \beta_0}{\beta_1^2} \right).$$

Logo, a estimativa de  $\text{Var}(\hat{\theta}_p) = \gamma^T \mathbf{V} \gamma$  é expressa por

$$\begin{aligned} \widehat{\text{Var}}(\hat{\theta}_p) &= \frac{1}{\hat{\beta}_1^2} \{ \widehat{\text{Var}}(\hat{\beta}_0) + \hat{\theta}_p^2 \widehat{\text{Var}}(\hat{\beta}_1) + 2\hat{\theta}_p \widehat{\text{Cov}}(\hat{\beta}_0, \hat{\beta}_1) \} \\ &= \frac{1}{\hat{\beta}_1^2} (\hat{v}_{11} + 2\hat{\theta}_p \hat{v}_{12} + \hat{\theta}_p^2 \hat{v}_{22}). \end{aligned} \quad (7.4)$$

Pelo método delta, supõe-se que a distribuição assintótica de  $\hat{\theta}_p = g(\hat{\beta})$  obtida usando-se a expressão (7.3) é normal  $N(\theta_p, \sigma^2)$  e, portanto, um intervalo de confiança  $100(1 - \gamma)\%$  aproximado para a dose efetiva  $\theta_p$  é expresso por

$$IC(\theta_p) : \hat{\theta}_p \mp z_{\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{\theta}_p)}, \quad (7.5)$$

em que  $z_{\alpha/2}$  é o  $(1 - \gamma/2)$  percentil da distribuição normal reduzida.

Uma desvantagem desse método é que o intervalo de confiança é sempre simétrico, o que pode ser desfavorável à estimação de doses efetivas extremas correspondentes a valores de  $p$  próximos de zero ou um. Pode, também, levar a limites inferiores negativos. Além disso, está baseado na distribuição normal assintótica de  $g(\hat{\beta})$ .

### Método baseado no teorema de Fieller

O teorema de Fieller é um resultado geral que permite a obtenção de intervalos de confiança para razões de duas variáveis aleatórias normalmente distribuídas, que é o caso da dose efetiva  $\theta_p$ . Esse teorema especifica a distribuição da soma de variáveis aleatórias que têm distribuição normal bivariada. O EMV  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)^T$  tem, assintoticamente, distribuição normal bivariada de média  $\beta$  e matriz de covariâncias  $\mathbf{V} = \mathbf{K}^{-1}$ .

A partir da equação (7.3) pode-se construir a função  $\hat{\psi} = \hat{\beta}_0 + \hat{\beta}_1 \theta_p - F^{-1}(p)$ . Então, usando-se a expressão (7.2), tem-se

$$E(\hat{\psi}) = \beta_0 + \beta_1 \theta_p - F^{-1}(p) = 0$$



e

$$\text{Var}(\hat{\psi}) = \text{Var}(\hat{\beta}_0) + 2\theta_p \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) + \theta_p^2 \text{Var}(\hat{\beta}_1) = v_{11} + 2\theta_p v_{12} + \theta_p^2 v_{22}.$$

Portanto,  $\hat{\beta}_0 + \hat{\beta}_1 \theta_p - F^{-1}(p) \sim N(0, \text{Var}(\hat{\psi}))$  e

$$\frac{\hat{\beta}_0 + \hat{\beta}_1 \theta_p - F^{-1}(p)}{\sqrt{\text{Var}(\hat{\beta}_0) + 2\theta_p \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) + \theta_p^2 \text{Var}(\hat{\beta}_1)}} = \frac{\hat{\beta}_0 + \hat{\beta}_1 \theta_p - F^{-1}(p)}{\sqrt{v_{11} + 2\theta_p v_{12} + \theta_p^2 v_{22}}} \sim N(0, 1).$$

Logo, um intervalo de confiança para  $\theta_p$ , com um coeficiente de confiança  $100(1 - \gamma)\%$ , pode ser expresso pelo conjunto de valores de  $\theta_p$  que satisfazem a inequação

$$\frac{[\hat{\beta}_0 + \hat{\beta}_1 \theta_p - F^{-1}(p)]^2}{v_{11} + 2\theta_p v_{12} + \theta_p^2 v_{22}} \leq z_{\alpha/2}^2,$$

isto é,

$$[\hat{\beta}_0 + \hat{\beta}_1 \theta_p - F^{-1}(p)]^2 - (v_{11} + 2\theta_p v_{12} + \theta_p^2 v_{22}) z_{\alpha/2}^2 = 0 \quad (7.6)$$

sendo que, substituindo-se  $v_{11}$ ,  $v_{12}$  e  $v_{22}$  por suas estimativas, os limites do intervalo de confiança igualam às raízes da correspondente equação de segundo grau. No caso de raízes complexas, o intervalo não existirá. Em geral, os resultados são semelhantes aos calculados pelo método delta. Se a variável explanatória for  $x = \log_b(\text{dose})$ , a dose efetiva é obtida fazendo-se  $\hat{d}_p = b^{\hat{\theta}_p}$  e, de forma semelhante, os limites do intervalo de confiança para  $b^{\theta_p}$ .

### Método baseado na razão de verossimilhanças

Uma terceira alternativa para se obter um intervalo de confiança para uma dose efetiva  $\theta_p$  é baseada na estatística da razão de verossimilhanças (Morgan, 1992). O componente sistemático usual (7.1), especificado  $p$ , pode ser reparametrizado em termos de  $\beta_1$  e  $\theta_p$ , isto é,

$$F^{-1}(\pi_i) = \beta_0 + \beta_1 x_i = \beta_0 + \beta_1 \theta_p - \beta_1 \theta_p + \beta_1 x_i = F^{-1}(p) + \beta_1 (x_i - \theta_p), \quad (7.7)$$

pois a dose efetiva satisfaz a condição dada pela equação (7.2). O componente sistemático (7.7) possibilita especificar  $F^{-1}(p)$  como *offset*.

Supondo-se  $n$  amostras de tamanhos  $m_i$ , o logaritmo da função de verossimilhança expresso em termos de  $\beta_1$  e  $\theta_p$  é

$$\ell(\beta_1, \theta_p) = \sum_{i=1}^n [y_i \log \pi_i + (m_i - y_i) \log(1 - \pi_i)], \quad (7.8)$$

em que  $y_i = 0, 1, \dots, m_i$  e  $\pi_i = F[F^{-1}(p) + \beta_1(x_i - \theta_p)]$ . Pode-se, então, maximizar (7.8) para obter as EMVs  $\hat{\beta}_1$  e  $\hat{\theta}_p$  e seu valor máximo  $\hat{\ell} = \ell(\hat{\beta}_1, \hat{\theta}_p)$ .

Um intervalo de confiança para  $\theta_p$  pode ser baseado na razão de verossimilhanças

$$w(\theta_p) = 2[\ell(\hat{\beta}_1, \hat{\theta}_p) - \ell(\tilde{\beta}_1, \theta_p)],$$

em que  $\tilde{\beta}_1$  é a EMV restrita de  $\beta_1$  fixado  $\theta_p = \theta$ , isto é, sob a hipótese de nulidade  $H_0 : \theta_p = \theta$ . Como, assintoticamente,  $w(\theta_p)$  tem distribuição  $\chi_1^2$ , tem-se que um intervalo com  $100(1 - \gamma)\%$  de confiança para  $\theta_p$  é formado por todos os valores de  $\theta$  para os quais  $H_0$  não é rejeitada em favor da hipótese alternativa  $H_0 : \theta_p \neq \theta$ , isto é, que verificam a inequação

$$\ell(\tilde{\beta}_1, \theta) \geq \ell(\hat{\beta}_1, \hat{\theta}_p) - \frac{1}{2}\chi_{1,\gamma}^2 \Leftrightarrow \text{desvio}(\tilde{\beta}_1, \theta) \leq \text{desvio}(\hat{\beta}_1, \hat{\theta}_p) + \chi_{1,\gamma}^2,$$

em que  $\chi_{1,\gamma}^2$  é o percentil  $(1 - \gamma)$  da distribuição qui-quadrado com 1 grau de liberdade.

### Exemplo 7.1: Toxicidade de Rotenone (cont.)

Usando-se os dados do **Exemplo 4.5**, tem-se que as estimativas dos parâmetros são  $\hat{\beta}_0 = -3,2257$  e  $\hat{\beta}_1 = 0,6051$  com  $\widehat{\text{Var}}(\hat{\beta}_0) = 0,1368$ ,  $\widehat{\text{Var}}(\hat{\beta}_1) = 0,0046$  e  $\widehat{\text{Cov}}(\hat{\beta}_0, \hat{\beta}_1) = -0,0227$ . Portanto, a dose que mata 50% dos insetos é dada por

$$\hat{\theta}_{50} = \frac{3,2257}{0,6051} = 5,3$$

com variância estimada, obtida a partir da equação (7.4),

$$\widehat{\text{Var}}(\hat{\theta}_{50}) = \frac{0,1368 - 2 \times 5,3 \times 0,0227 + 5,3^2 \times 0,0046}{0,6051^2} = 0,0708$$

enquanto que os intervalos de 95% de confiança, obtidos pelos três métodos, são

$$\text{i) Método Delta: } 4,8 < \theta_{50} < 5,9,$$

obtido a partir da equação (7.5), isto é, de

$$5,3 \pm 1,96 \times \sqrt{0,0708}$$

$$\text{ii) Método de Fieller: } 4,8 < \theta_{50} < 5,9$$

obtido a partir da solução da equação (7.6), isto é, de

$$(-3,2257 + 0,6051 \times \theta)^2 - (0,1368 - 2 \times \theta \times 0,0227 + 5,3^2 \times 0,0046) \times 1,96^2 = 0$$

iii) Método da razão de verossimilhanças:  $5, 0 < \theta_{50} < 5, 7$ .

Na Figura 7.1, apresenta-se o perfil de verossimilhanças para esses dados com a estimativa de  $DL_{50}$  e o respectivo intervalo de confiança com 95% de confiança.

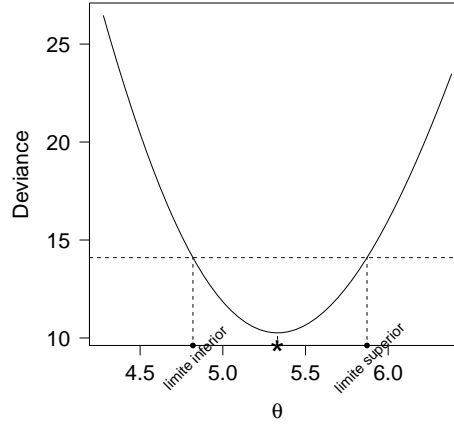


Figura 7.1: Perfil de verossimilhança para os dados de mortalidade de Rotenone, com a estimativa e o intervalo de confiança com 95% de confiança para  $DL_{50}$  e

### 7.1.2 Probabilidade de resposta a uma dose especificada

Considere que a dose  $d$  recebida por um indivíduo  $i$  não é fixada, como descrito na Seção 2.2, mas corresponde à soma de uma dose nominal  $x$  mais um erro aleatório  $\epsilon \sim N(0, \sigma^2)$ , isto é,  $d = x + \epsilon$ . Se a distribuição da tolerância  $T$  (independe do erro aleatório), tem a forma usual

$$P(T \leq d) = F(d) = \Phi(\beta'_0 + \beta'_1 d),$$

pode-se concluir que  $T \sim N(-\beta'_0/\beta'_1, \beta_1'^2)$ . Logo,  $T - \epsilon$  tem distribuição normal de média  $-\beta'_0/\beta'_1$  e variância aumentada

$$\text{Var}(T - \epsilon) = \beta_1'^2 + \sigma^2.$$

Então,

$$P(T \leq d) = P(T - \epsilon \leq x) = P\left(\frac{T - \epsilon + \beta'_0/\beta'_1}{(\beta_1'^2 + \sigma^2)^{1/2}} \leq \frac{x + \beta'_0/\beta'_1}{(\beta_1'^2 + \sigma^2)^{1/2}}\right)$$

e, portanto,

$$F(d) = \Phi\left(\frac{\beta'_0 + \beta'_1 x}{(\beta_1'^2 + \sigma^2)^{1/2}}\right). \quad (7.9)$$

A equação (7.9) revela uma forma computacional simples de calcular a probabilidade de resposta a uma dose  $d$ . Na prática, procede-se à análise usual de ajuste do modelo, considerando  $d$  fixo, para o cálculo das EMVs de  $\beta_0$  e  $\beta_1$ . Para usar a equação (7.9), obtêm-se as correções  $\beta'_0 = \beta_0 c$  e  $\beta'_1 = \beta_1 c$ , em que  $c = (1 - \beta_1^2 \sigma^2)^{-1/2}$ , supondo  $\beta_1 < \sigma^{-1}$  e  $\sigma^2$  conhecido.

### 7.1.3 Paralelismo entre retas no modelo logístico linear e potência relativa

Na área de toxicologia é muito comum o interesse na comparação da eficiência de produtos (fungicidas, inseticidas, herbicidas, medicamentos, etc) ou tratamentos. Considerando-se o modelo logístico linear com uma variável quantitativa  $x$  (dose ou  $\log(\text{dose})$ ) e  $k$  produtos a serem testados, os preditores lineares considerados são:

Retas concorrentes :	$\text{logit}(p_{ij}) = \alpha_j + \beta_j \log(\text{dose}_i)$
Retas paralelas :	$\text{logit}(p_{ij}) = \alpha_j + \beta \log(\text{dose}_i)$
Retas com intercepto comum :	$\text{logit}(p_{ij}) = \alpha + \beta_j \log(\text{dose}_i)$
Retas coincidentes :	$\text{logit}(p_{ij}) = \alpha + \beta \log(\text{dose}_i)$

para  $j = 1, \dots, k$ . O ajuste desses modelos aos dados é verificado usando-se as diferenças dos desvios residuais. No caso em que existem evidências de que o modelo de retas paralelas ajusta-se bem aos dados, tem-se, então, que a dose efetiva ( $\hat{\theta}_j^{(p)}$ ) para 100p% dos indivíduos é obtida a partir de:

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \hat{\alpha}_j + \hat{\beta} \hat{\theta}_j^{(p)}, \quad j = 1, \dots, k.$$

Portanto, para  $j \neq j'$ , tem-se

$$\frac{\hat{\alpha}_j - \hat{\alpha}_{j'}}{\hat{\beta}} = \hat{\theta}_{j'}^{(p)} - \hat{\theta}_j^{(p)}.$$

Se  $x = \log(d)$ , então,

$$\frac{\hat{\alpha}_j - \hat{\alpha}_{j'}}{\hat{\beta}} = \log\left(\frac{\hat{d}_j^{(p)}}{\hat{d}_{j'}^{(p)}}\right) = \log(\hat{\rho}_{jj'}) \Rightarrow \hat{\rho}_{jj'} = \exp\left(\frac{\hat{\alpha}_j - \hat{\alpha}_{j'}}{\hat{\beta}}\right) = \frac{DE_{j'}^{(50)}}{DE_j^{(50)}},$$

sendo  $\hat{\rho}_{jj'}$  a estimativa da eficiência (potência) relativa  $\rho_{jj'}$  do produto  $j$  em relação ao  $j'$  e  $\log[\hat{d}_j^{(p)}] - \log[\hat{d}_{j'}^{(p)}]$ , uma medida da diferença horizontal entre as duas retas paralelas. Portanto,  $\rho_{jj'}$  é a razão de duas doses igualmente efetivas. Intervalos de confiança para  $\rho_{jj'}$  podem ser calculados pelos métodos Delta, de Fieller e da razão de verossimilhanças (perfil de verossimilhanças) (Morgan, 1992; Collet, 2002).

### Exemplo 7.2: Resistência a cipermetrina

Como parte de estudos de resistência de insetos a piretróides, um experimento completamente casualizado foi conduzido. Amostras de 20 mariposas de *Heliothis virescens* (praga de algodão, soja, tabaco), de cada sexo, foram expostas a doses crescentes de cipermetrina, dois dias depois da emergência da pupa (Collet, 2002). Após 72h, foram contados os números de insetos mortos cujos resultados estão na Tabela 7.1.

Tabela 7.1: Números de insetos mortos em amostras de 20 insetos, machos e fêmeas, expostos a doses ( $d_i$ ) crescentes de cipermetrina

Doses ( $d_i$ )	Número de insetos mortos	
	Machos	Fêmeas
1,0	1	0
2,0	4	2
4,0	9	6
8,0	13	10
16,0	18	12
32,0	20	16

Considerando-se que a variável aleatória número de insetos mortos  $Y_i$  tem distribuição binomial, isto é,  $Y_i \sim B(20, \pi_i)$ , ajusta-se, inicialmente, o modelo logístico, assumindo para o preditor linear um fatorial  $2 \times 6$  e seus submodelos, sendo sexo um fator com dois níveis, e dose um fator com 6 níveis (em princípio, sem considerar o fato de serem quantitativos).

Na Tabela 7.2, são apresentados os desvios residuais, estatísticas  $X^2$  para os diversos modelos e seus respectivos números de graus de liberdade (g.l.). Verifica-se que existem evidências de que o modelo com preditor linear com dois fatores aditivos, ajusta-se bem aos dados, enquanto que os modelos mais simples, não. Não há, portanto, evidência

Tabela 7.2: Desvios e estatísticas  $X^2$  residuais e respectivos números de graus de liberdade e níveis descritivos ( $p$ ), para os dados de mortalidade de *Heliothis virescens* (Tabela 7.1).

Modelo	g.l.	Desvios	$p$	$X^2$	$p$
Constante	11	124,88	$< 0,0001$	101,42	$< 0,0001$
Sexo	10	118,80	$< 0,0001$	97,39	$< 0,0001$
Dose	6	15,15	0,0191	12,93	0,0441
Sexo + Dose	5	5,01	0,4143	3,70	0,5932
Dose*Sexo	0	0		0	

de efeito de interação entre os dois fatores. Por outro lado, a análise de desvios (Tabela 7.3) evidencia o efeito significativo de sexo e de dose. Nota-se, ainda, que os desvios para sexo ignorando dose e, para sexo ajustado para dose, são diferentes devido à não ortogonalidade por se estar considerando a distribuição binomial. O mesmo ocorre para dose ignorando sexo e para dose ajustada para sexo.

Tabela 7.3: Análise de Desvios para os dados de mortalidade de *Heliothis virescens* (Tabela 7.1), considerando-se sexo e doses como fatores.

Fontes	g.l.	Desvios	$p$	Fontes	g.l.	Desvios	$p$
Sexo	1	6,08	0,0137	Dose	5	109,72	$< 0,0001$
Dose Sexo	5	113,79	$< 0,0001$	Sexo Dose	1	10,14	0,0014
Resíduo	5	5,01		Resíduo	5	5,01	
Total	11	124,9			11	124,9	

É importante observar que a fonte de variação “Resíduo” na realidade representa uma composição de erros aleatórios e interação, pois o experimento foi conduzido com apenas uma repetição de cada combinação dos níveis dos fatores Sexo e Dose de cipermetrina.

Pode-se, ainda, tentar uma simplificação desse modelo, considerando que dose é um fator quantitativo e que há interesse em se estudar a relação entre a proporção de insetos mortos e a dose de cipermetrina, para ambos os sexos. Se for usado como preditor linear regressão polinomial para  $x = \text{dose}$ , verifica-se que há necessidade de um polinômio de grau 3. Como, porém, as doses estão em progressão geométrica é conveniente usar

como variável regressora  $x = \log_2(\text{dose})$ , considerando-se os modelos de retas concorrentes, paralelas, com intercepto comum e coincidentes. Os resultados para o desvio e para a estatística  $X^2$  residuais estão apresentados na Tabela 7.4.

Tabela 7.4: Desvios e estatísticas  $X^2$  residuais e respectivos números de graus de liberdade e níveis descritivos ( $p$ ), para os dados de mortalidade de *Heliothis virescens* (Tabela 7.1).

Modelo	g.l.	Desvios	$p$	$X^2$	$p$
Constante	11	124,88	$< 0,0001$	101,42	$< 0,0001$
Const. + log(Dose)	10	16,98	0,0747	14,76	0,1409
Const. + Sexo.log(Dose)	9	5,04	0,8304	3,50	0,9410
Sexo + log(dose)	9	6,76	0,6624	5,31	0,8069
Sexo + Sexo.log(dose)	8	4,99	0,7582	3,50	0,8988

Pela Tabela 7.4, nota-se que existem evidências que os modelos com retas concorrentes, paralelas e com intercepto comum ajustam-se aos dados. Adicionalmente, tem-se que as diferenças de desvios, ambas com um grau de liberdade, entre os modelos com retas paralelas e retas concorrentes ( $6,76 - 4,99 = 1,77$ ) e entre os modelos com intercepto comum e retas concorrentes ( $5,04 - 4,99 = 0,05$ ) não são estatisticamente significativas. Utilizando de parcimônia e da facilidade de interpretação prática, pode-se optar pelo modelo de retas paralelas, com análise de desvios apresentada na Tabela 7.5.

Tabela 7.5: Análise de Desvios para os dados de mortalidade de *Heliothis virescens* (Tabela 7.1), considerando-se o modelo de retas paralelas.

Fontes de Variação	g.l.	Desvios	$p$
Sexo	1	6,08	0,0137
Regressão Linear	1	112,04	$< 0,0001$
Resíduo	9	6,76	
Total	11	124,88	

É importante observar que a fonte de variação “Resíduo” na realidade representa uma composição de erros aleatórios e falta de ajuste (não-linearidade), pois o experimento

foi conduzido com apenas uma repetição de cada combinação dos níveis dos fatores Sexo e Dose de cipermetrina. Pelo gráfico meio-normal de probabilidade com envelopes simulados (Figura 7.2(a)) confirma-se que esse modelo é adequado.

A partir do modelo linear logístico escolhido, obtêm-se, então, respectivamente, as equações para machos e fêmeas:

$$\begin{aligned}\text{Machos : } \log \left( \frac{\hat{p}_i}{1 - \hat{p}_i} \right) &= -2,3724 + 1,0642 \log_2(\text{dose}_i) \\ \text{Fêmeas : } \log \left( \frac{\hat{p}_i}{1 - \hat{p}_i} \right) &= -3,4732 + 1,0642 \log_2(\text{dose}_i)\end{aligned}$$

cujas curvas com as proporções observadas estão representadas na Figura 7.2(b).

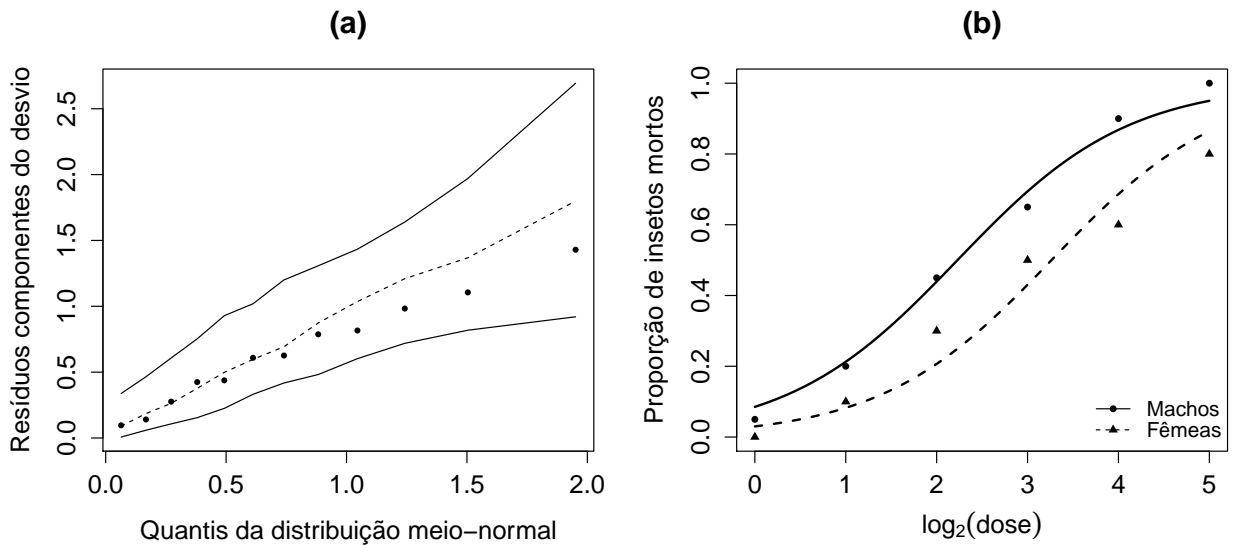


Figura 7.2: Modelo logístico linear para  $\log_2(\text{dose})$  de cipermetrina – (a) gráfico meio-normal de probabilidade com envelopes simulados; (b) curvas ajustadas com as proporções observadas.

A partir dos coeficientes estimados, podem-se estimar as doses que matam, por exemplo, 50% dos insetos

$$\begin{aligned}\text{Machos : } \log_2(\widehat{DL}_{50}) &= \frac{2,3724}{1,0642} = 2,2336 \Rightarrow \widehat{DL}_{50} = 4,69 \\ \text{Fêmeas : } \log_2(\widehat{DL}_{50}) &= \frac{3,4732}{1,0642} = 3,2636 \Rightarrow \widehat{DL}_{50} = 9,60\end{aligned}$$

com erros padrões iguais a  $\sqrt{0,0527} = 0,2298$  and  $\sqrt{0,0511} = 0,2260$ , respectivamente.



Os intervalos de confiança para as doses que matam 50% dos insetos, são dados por:

	Machos	Fêmeas
Método Delta :	$[2^{1,7864}, 2^{2,6721}] = [3, 45; 6, 37]$	$[2^{1,8133}, 2^{2,7139}] = [7, 03; 13, 12]$
Método de Fieller :	$[2^{1,7685}, 2^{2,6813}] = [3, 41; 6, 41]$	$[2^{2,8122}, 2^{2,7406}] = [7, 02; 13, 37]$
Método de PV :	$[2^{1,7771}, 2^{2,6743}] = [3, 43; 6, 38]$	$[2^{2,8176}, 2^{3,7315}] = [7, 05; 13, 28]$

Verifica-se que as fêmeas são mais tolerantes, pois para matar 100p% das fêmeas há necessidade de uma dose duas vezes maior do que para matar 100p% dos machos. Pode-se verificar que a dose letal correspondente a  $p = 0,9$  para as fêmeas está fora do intervalo estudado, o que é perigoso, pois acima da dose 32 não se sabe se o comportamento será o mesmo. Se o interesse for estimar essa dose, há necessidade de se aumentar a amplitude de doses para fêmeas em um novo experimento.

### Exemplo 7.3: Potência relativa - Mortalidade do besouro da farinha

Grupos de insetos (*Tribolium castaneum*, praga de grãos) foram expostos a doses (mg/l) crescentes de DDT,  $\gamma$ -BHC e mistura dos dois. Depois de 6 dias, foram contados os números de insetos mortos cujos resultados estão na Tabela 7.6 (Collet, 2002).

Tabela 7.6: Proporções de insetos mortos quando expostos a doses crescentes de DDT,  $\gamma$ -BHC e mistura dos dois.

Inseticida	Doses					
	2,00	2,64	3,48	4,59	6,06	8,00
DDT	3/50	5/49	19/47	19/50	24/49	35/50
$\gamma$ -BHC	2/50	14/49	20/50	27/50	41/50	40/50
DDT + $\gamma$ -BHC	28/50	37/50	46/50	48/50	48/50	50/50

Considerando-se que a variável aleatória número de insetos mortos  $Y_i$  tem distribuição binomial, isto é,  $Y_i \sim B(m_i, \pi_i)$ , ajusta-se, inicialmente, o modelo logístico assumindo para o preditor linear um fatorial  $3 \times 6$  e seus submodelos, sendo inseticida um fator com três níveis, e dose um fator com 6 níveis (em princípio, sem considerar o fato de serem quantitativos). Na Tabela 7.7, são apresentados os desvios residuais, estatísticas  $X^2$  para os diversos modelos e seus respectivos números de graus de liberdade (g.l.).

Tabela 7.7: Desvios e estatísticas  $X^2$  residuais e respectivos números de graus de liberdade e níveis descritivos ( $p$ ), para os dados de mortalidade de *Tribolium castaneum* (Tabela 7.6).

Modelo	g.l.	Desvios	$p$	$X^2$	$p$
Constante	17	413,65	$< 0,0001$	347,14	$< 0,0001$
Inseticida	15	234,71	$< 0,0001$	215,05	$< 0,0001$
Dose	12	242,64	$< 0,0001$	218,95	$< 0,0001$
Inseticida + Dose	10	12,86	0,2317	11,80	0,2989

Verifica-se que existem evidências de que o modelo com preditor linear com dois fatores aditivos, inseticida (com três níveis) e dose (com 6 níveis, em princípio sem levar em consideração o fato de serem quantitativos), ajusta-se bem aos dados, enquanto que os modelos mais simples, não. Não há, portanto, evidência de efeito de interação entre os dois fatores. Por outro lado, a análise de desvios (Tabela 7.8) evidencia o efeito significativo de inseticida e de dose. Nota-se, ainda, que os desvios para inseticida ignorando dose e, para inseticida ajustado para dose, são diferentes devido à não ortogonalidade por se estar considerando a distribuição binomial. O mesmo ocorre para dose ignorando inseticida e para dose ajustada para inseticida.

Tabela 7.8: Análise de Desvios para os dados de mortalidade de *Tribolium castaneum* (Tabela 7.6), considerando-se inseticida e dose como fatores.

Fontes	g.l.	Desvios	$p$	Fontes	g.l.	Desvios	$p$
Inseticida	2	178,93	$< 0,0001$	Dose	5	171,00	$< 0,0001$
Dose Inseticida	5	221,85	$< 0,0001$	Inseticida Dose	2	229,78	$< 0,0001$
Resíduo	10	12,86		Resíduo	10	12,86	
Total	17	413,65		Total	17	413,65	

É importante observar que a fonte de variação “Resíduo” na realidade representa uma composição de erros aleatórios e interação, pois o experimento foi conduzido com apenas uma repetição de cada combinação dos níveis dos fatores inseticida e dose.

Pode-se, ainda, tentar uma simplificação desse modelo, considerando que dose é um fator quantitativo e que há interesse em se estudar a relação entre a proporção de

insetos mortos e a dose, para os três produtos inseticidas. Pode-se usar como preditor linear uma regressão polinomial para  $x = \log(\text{dose})$ , considerando-se os modelos de retas concorrentes, paralelas, com intercepto comum e coincidentes. Os resultados para o desvio e para a estatística  $X^2$  residuais estão apresentados na Tabela 7.9.

Tabela 7.9: Desvios e estatísticas  $X^2$  residuais e respectivos números de graus de liberdade e níveis descritivos ( $p$ ), para os dados de mortalidade de *Tribolium castaneum* (Tabela 7.6).

Preditor	g.l.	Desvios	$p$	$X^2$	$p$
$\alpha$	17	413,65	$< 0,0001$	347,14	$< 0,0001$
$\alpha + \beta x$	16	246,83	$< 0,0001$	219,84	$< 0,0001$
$\alpha + \beta_j x$	14	24,71	0,0375	28,04	0,0141
$\alpha_j + \beta x$	14	21,28	0,0946	20,32	0,1203
$\alpha_j + \beta_j x$	12	17,89	0,1191	17,61	0,1280

Pela Tabela 7.9, nota-se que existem evidências que os modelos com retas concorrentes, paralelas e com intercepto comum ajustam-se aos dados. Adicionalmente, tem-se que as diferenças de desvios, ambas com dois graus de liberdade, entre os modelos com intercepto comum e retas concorrentes ( $24,71 - 17,89 = 6,82$ ,  $p = 0,03299$ ) e entre os modelos com retas paralelas e retas concorrentes ( $21,28 - 17,89 = 3,39$ ,  $p = 0,1834$ ) mostram evidências que significativas no primeiro caso e não significativa no segundo. Utilizando de parcimônia, opta-se pelo modelo de retas paralelas, com análise de desvios apresentada na Tabela 7.10.

Tabela 7.10: Análise de Desvios para os dados de mortalidade de *Tribolium castaneum* (Tabela 7.6), considerando-se inseticida e dose como fatores.

Fontes de Variação	g.l.	Desvios	$p$
Inseticida	2	178,93	$< 0,0001$
Regressão Linear	1	213,43	$< 0,0001$
Resíduo	14	21,28	
Total	17	413,65	

É importante observar que a fonte de variação “Resíduo” na realidade representa

uma composição de erros aleatórios e falta de ajuste (não-linearidade), pois o experimento foi conduzido com apenas uma repetição de cada combinação dos níveis dos fatores inseticida e dose. Pelo gráfico meio-normal de probabilidade com envelopes simulados (Figura 7.3(a)) confirma-se que o modelo de retas paralelas é adequado.

A partir do modelo linear logístico escolhido, obtêm-se, então, respectivamente, as equações para os três produtos inseticidas:

$$\begin{aligned} \text{DDT :} \quad & \log \left( \frac{\hat{p}_i}{1 - \hat{p}_i} \right) = -4,5553 + 2,6958 \log(\text{dose}_i) \\ \gamma - \text{BHC :} \quad & \log \left( \frac{\hat{p}_i}{1 - \hat{p}_i} \right) = -3,8425 + 2,6958 \log(\text{dose}_i) \\ \text{DDT} + \gamma - \text{BHC :} \quad & \log \left( \frac{\hat{p}_i}{1 - \hat{p}_i} \right) = -1,4248 + 2,6958 \log(\text{dose}_i) \end{aligned}$$

cujas curvas com as proporções observadas estão representadas na Figura 7.3(b).

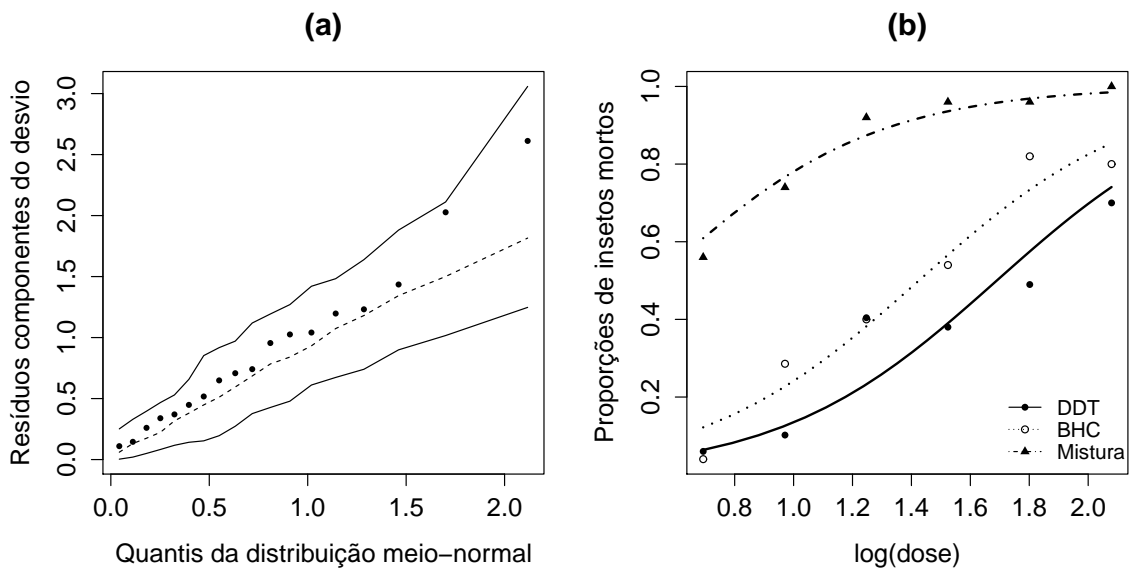


Figura 7.3: Modelo logístico linear para  $\log(\text{dose})$  dos inseticidas – (a) gráfico meio-normal de probabilidade com envelopes simulados; (b) curvas ajustadas com as proporções observadas.

A partir dos coeficientes estimados, podem-se estimar as doses que matam, por

exemplo, 50% dos insetos

$$\begin{aligned} \text{DDT :} \quad \log(\widehat{DL}_{50}) &= \frac{4,5553}{2,6958} = 1,69 \Rightarrow \widehat{DL}_{50} = 5,42 \\ \gamma - \text{BHC :} \quad \log(\widehat{LD}_{50}) &= \frac{3,8425}{2,6958} = 1,43 \Rightarrow \widehat{DL}_{50} = 4,16 \\ \text{DDT} + \gamma - \text{BHC :} \quad \log(\widehat{DL}_{50}) &= \frac{1,4248}{2,6958} = 0,53 \Rightarrow \widehat{DL}_{50} = 1,70 \end{aligned}$$

com erros padrões iguais a  $\sqrt{0,0029} = 0,0537$ ,  $\sqrt{0,0025} = 0,05048$  and  $\sqrt{0,0060} = 0,0773$ , respectivamente.

Os intervalos de confiança, com um nível de confiança de 95%, para as doses que matam 50% dos insetos, são dados por:

	DDT	BHC	Mistura
Delta :	[4,89; 6,02]	[3,77; 4,59]	[1,46; 1,97]
Fieller :	[4,89; 6,06]	[3,76; 4,60]	[1,44; 1,95]
PV :	[4,89; 6,05]	[3,77; 4,60]	[1,43; 1,95]

e as potências relativas

$$\begin{aligned} \text{da mistura em relação ao DDT :} \quad & \frac{4,16}{1,70} = 2,45 \\ \text{da mistura em relação ao } \gamma - \text{BHC :} \quad & \frac{5,42}{1,70} = 3,19 \end{aligned}$$

mostrando evidência de **sinergismo**, isto é, a mistura dos inseticidas potencializa o efeito de ação.

Entretanto, pela Figura 7.3, sugere-se que seja testada a coincidência das retas para os inseticidas DDT e  $\gamma$ -BHC. Para esse teste, supõe-se as observações referentes a esses inseticidas são provenientes de inseticidas de efeitos semelhantes, isto é, ajusta-se um modelo de retas paralelas para dois (e não três) tratamentos, o primeiro com 10 observações e o segundo com cinco observações, obtendo-se um desvio residual de 34,52 com 15 graus de liberdade. Esse valor comparado com o valor obtido na Tabela 7.9, é igual a  $34,52 - 21,28 = 13,24$  com um grau de liberdade, mostrando evidência de rejeição da hipótese de nulidade que os inseticidas DDT e  $\gamma$ -BHC têm efeitos semelhantes. Portanto, permanecem os resultados obtidos.

#### Exemplo 7.4: Proporções de gemas florais de macieiras

Os dados da Tabela 7.11 referem-se a um experimento em que gemas de galhos de três variedades de macieiras foram classificadas em florais ou vegetativas. Para cada

variedade, os galhos foram agrupados de acordo com o número de frutos (de 0 a 4) produzidos no ano anterior. O objetivo do experimento era estudar a relação entre a proporção de gemas florais e o número de frutos produzidos no ano anterior e verificar se essa relação era diferente para as variedades estudadas.

Tabela 7.11: Número de frutos produzidos no ano anterior e o número de gemas em galhos de macieiras de três variedades.

Variedades	Número de frutos no ano anterior (X)	Número total de gemas (N)	Número de gemas florais (Y)	Proporção de gemas florais (P)
Crispin	0	69	42	0,61
	1	93	43	0,46
	2	147	59	0,40
	3	149	57	0,38
	4	151	43	0,28
Cox	0	34	12	0,35
	1	92	15	0,16
	2	133	18	0,14
	3	146	14	0,10
	4	111	9	0,08
Golden Delicious	0	21	6	0,29
	1	89	20	0,22
	2	118	20	0,17
	3	124	21	0,10
	4	81	4	0,00

A variável resposta,  $Y_i$ , é o número de gemas florais em totais de  $m_i$  gemas, e, portanto, a distribuição a ser considerada é a binomial. Como função de ligação pode ser usada a logística,  $g(\mu_i/m_i) = \log[\mu_i/(m_i - \mu_i)]$ , e como parte sistemática tem-se um delineamento inteiramente casualizado com fatores, variedades (qualitativo) e número de frutos no ano anterior (quantitativo). Os preditores lineares a serem considerados são retas que podem ser concorrentes, paralelas, com intercepto comum ou coincidentes, para as três variedades.

Na Tabela 7.12, são apresentados os desvios e as estatísticas  $X^2$  residuais e seus respectivos números de graus de liberdade (g.l.). Verifica-se que existem evidências de que os modelos com retas paralelas e com retas concorrentes ajustam-se bem aos dados, enquanto que os outros modelos, não. Adicionalmente, tem-se que as diferenças de desvios, ambas com dois graus de liberdade, entre os modelos com intercepto comum e retas

Tabela 7.12: Desvios e estatísticas  $X^2$  residuais e respectivos números de graus de liberdade e níveis descritivos ( $p$ ), para proporções de gemas florais em galhos de macieiras como função do número de frutos produzidos no ano anterior (Tabela 7.11).

Preditor linear	g.l.	Desvios	$p$	$X^2$	$p$
$\alpha$	14	182,16	$< 0,0001$	181,06	$< 0,0001$
$\alpha + \beta x$	13	138,99	$< 0,0001$	139,04	$< 0,0001$
$\alpha_j$	12	53,04	$< 0,0001$	54,51	$< 0,0001$
$\alpha + \beta_j x$	11	31,08	0,0011	30,97	0,0011
$\alpha_j + \beta x$	11	8,80	0,6400	8,67	0,6524
$\alpha_j + \beta_j x$	9	7,87	0,5469	7,76	0,5585

concorrentes ( $31,08 - 7,87 = 23,21$ ,  $p > 0,001$ ) e entre os modelos com retas paralelas e retas concorrentes ( $8,80 - 7,87 = 0,93$ ,  $p = 0,6281$ ) mostram evidências significativa no primeiro caso e não significativa no segundo. Utilizando de parcimônia, opta-se pelo modelo de retas paralelas, com análise de desvios apresentada na Tabela 7.13. Observa-se que existem evidências para o efeito de variedades e para o efeito de regressão linear.

Tabela 7.13: Análise de Desvios para proporções de gemas florais em galhos de macieiras como função do número de frutos produzidos no ano anterior (Tabela 7.11), considerando o preditor linear de retas paralelas.

Fontes de Variação	g.l.	Desvios	$p$
Variedades	2	129,12	$< 0,0001$
Regressão linear	1	44,24	$< 0,0001$
Resíduo	11	8,80	
Total	14	182,16	

Pelo gráfico meio-normal de probabilidade com envelope simulado (Figura 7.4(a)) confirma-se que o preditor linear de retas paralelas é adequado, obtendo-se, então, respectivamente, as equações para o cálculo das proporções estimadas para as três variedades:

$$\begin{aligned}
\text{Crispin :} \quad \hat{\pi}_i &= \frac{e^{0,3605-0,3302x_i}}{1 - e^{0,3605-0,3302x_i}} \\
\text{Cox :} \quad \hat{\pi}_i &= \frac{e^{-1,1491-0,3302x_i}}{1 - e^{-1,1491-0,3302x_i}} \\
\text{Golden Delicious :} \quad \hat{\pi}_i &= \frac{e^{-0,8979-0,3302x_i}}{1 - e^{-0,8979-0,3302x_i}}
\end{aligned}$$

cujas curvas com as proporções observadas estão representadas na Figura 7.4(b).

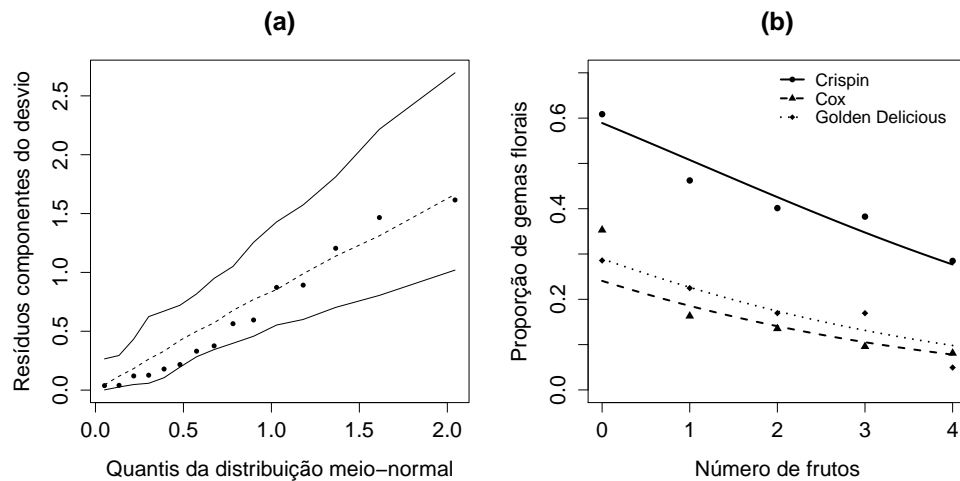


Figura 7.4: Modelo logístico para proporções de gemas florais em galhos de macieiras como função do número de frutos produzidos no ano anterior, considerando o preditor linear de retas paralelas – (a) gráfico meio-normal de probabilidade com envelope simulado; (b) curvas ajustadas com as proporções observadas.

Verifica-se que as curvas para as variedades Cox e Golden Delicious estão muito próximas e, portanto, é interessante testar se elas diferem estatisticamente. Um novo modelo foi usado, considerando-se os dados das variedades Cox e Golden Delicious como se fosse uma única variedade. A diferença entre os desvios residuais,  $(10,64 - 8,80 = 1,84, p = 0,1760)$  indica que existem evidências de que as variedades Cox e Golden Delicious comportam-se de forma semelhante. Os resultados obtidos para a análise de desvio estão na Tabela 7.14.

Pelo gráfico meio-normal de probabilidade com envelope simulado (Figura 7.5(a)) confirma-se que o preditor linear de retas paralelas é adequado, obtendo-se, então, respec-



Tabela 7.14: Análise de Desvios para proporções de gemas florais em galhos de macieiras como função do número de frutos produzidos no ano anterior (Tabela 7.11), considerando o preditor linear de retas paralelas e agrupamento das variedades Cox e Golden Delicious.

Fontes de Variação g.l. Desvios		$p$
Variedades	1	127,17 < 0,0001
Regressão linear	1	44,35 < 0,0001
Resíduo	12	10,64
Total	14	182,16

tivamente, as equações para o cálculo das proporções estimadas para as “duas” variedades:

$$\begin{aligned} \text{Crispin :} \quad \hat{\pi}_i &= \frac{e^{0,3605-0,3302x_i}}{1 - e^{0,3605-0,3302x_i}} \\ \text{Golden Delicious e Cox :} \quad \hat{\pi}_i &= \frac{e^{-1,0286-0,3302x_i}}{1 - e^{-1,0286-0,3302x_i}} \end{aligned}$$

cujas curvas com as proporções observadas estão representadas na Figura 7.5(b).

O resultado evidencia, portanto, que a proporção de gemas florais decresce em função do número de frutos produzidos no ano anterior e que a variedade Crispin produz mais gemas florais do que as variedades Cox e Golden Delicious. Esse é um fenômeno conhecido como alelopatia em Agricultura, em que muitas culturas são bienais, isto é, produzem bem a cada dois anos.

### Exemplo 7.5: Cultura de tecidos de macieiras

Os dados apresentados na Tabela 7.15 referem-se a um experimento de cultura de tecidos de macieiras. O delineamento experimental utilizado foi o casualizado em blocos com os tratamentos, no esquema fatorial  $2 \times 3 \times 3$ , sendo os fatores: A - dois tipos de citocinina (BAP, TDZ); B - três doses de citocinina (5,0; 1,0 e 0,1) e C - três tipos de auxina (NAA, IBA, 2-4D). Cada parcela do experimento era constituída de um recipiente em que era colocado o meio de cultura (de acordo com a combinação dos níveis dos três fatores) e o explante. O objetivo desse experimento foi verificar a influência dos fatores e se existia interação entre eles sobre a regeneração dos explantes.

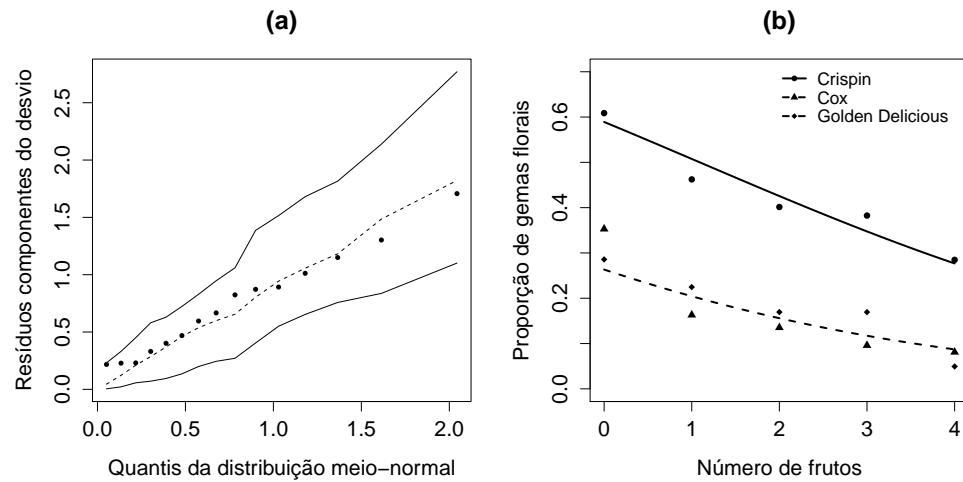


Figura 7.5: Modelo logístico para proporções de gemas florais em galhos de macieiras como função do número de frutos produzidos no ano anterior, considerando o preditor linear de retas paralelas e agrupamento das variedades Cox e Golden Delicious– (a) gráfico meio-normal de probabilidade com envelope simulado; (b) curvas ajustadas com as proporções observadas.

A variável resposta,  $Y$ , é binária, isto é,

$$Y = \begin{cases} 1 & \text{o explante regenerou após 4 semanas} \\ 0 & \text{em caso contrário} \end{cases}$$

e, portanto, a distribuição a ser considerada é a Bernoulli (caso particular da binomial). Como função de ligação pode ser usada a logística  $g(\mu) = \log[\mu/(1 - \mu)]$  e como parte sistemática tem-se um delineamento casualizado em blocos com três fatores, isto é, com preditor linear:

$$\eta = \delta_l + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk}. \quad (7.10)$$

em que  $\delta_l$ ,  $l = 1, 2, \dots, 10$ , representa o efeito do bloco  $l$ ;  $\alpha_i$ ,  $i = 1, 2, 3$ , o efeito do tipo  $i$  de citocinina;  $\beta_j$ ,  $j = 1, 2$ , o efeito da dose  $j$  de citocinina;  $\gamma_k$ ,  $k = 1, 2, 3$ , o efeito do tipo  $k$  de auxina e os demais representam efeitos de interações duplas e tripla.

Na Tabela 7.16, são apresentados os desvios e  $X^2$  residuais e seus respectivos números de graus de liberdade (g.l.). É importante lembrar que no caso de dados binários

Tabela 7.15: Dados de regeneração de tecidos de macieiras.

Citocinina			Blocos									
Tipo	Dose	Auxina	1	2	3	4	5	6	7	8	9	10
BAP	5,0	NAA	1	1	0	0	1	0	1	0	1	1
		IBA	0	1	1	1	1	1	0	1	1	1
		2-4D	1	1	1	1	1	1	1	0	0	1
	1,0	NAA	0	0	0	0	0	0	0	0	0	0
		IBA	1	1	1	0	0	1	1	0	1	1
		2-4D	1	0	1	1	0	1	1	1	1	1
	0,1	NAA	0	0	1	1	1	0	1	0	0	0
		IBA	0	0	0	1	1	1	1	0	1	0
		2-4D	0	0	1	1	1	1	1	0	1	1
	5,0	NAA	1	1	1	1	1	0	1	1	1	1
		IBA	1	1	1	1	1	1	1	1	1	1
		2-4D	1	0	1	1	1	1	1	1	1	1
TDZ	1,0	NAA	1	1	1	1	1	1	1	1	1	1
		IBA	1	1	1	1	1	1	1	1	1	1
		2-4D	1	1	1	1	1	1	1	1	1	0
	0,1	NAA	1	1	1	1	1	1	1	0	1	1
		IBA	1	1	1	1	1	1	0	1	1	1
		2-4D	0	0	1	0	1	1	1	1	1	1

(Seção 4.2) o desvio residual não é informativo para verificar a adequação dos modelos, pois é apenas uma função das observações. A diferença entre desvios, porém, pode ser utilizada e comparada com o percentil de uma distribuição  $\chi^2_\nu$  com  $\nu$  graus de liberdade (Seção 4.2).

Na Tabela 7.17, é apresentada uma sequência para análise de desvios, dentre as muitas possíveis. Verifica-se, então, que existem evidências que apenas os tipos de citocinina e de auxina têm influência significativa na regeneração de tecidos de macieiras e, além do mais, existe uma interação entre esses dois fatores, conforme mostra a Tabela 7.17, para os modelos encaixados. Considera-se, então, o modelo reduzido com preditor linear:

$$\eta = \delta_l + \alpha_i + \gamma_k + \alpha\gamma_{ik} \quad (7.11)$$

com os resultados apresentados na Tabela 7.18. Na Figura 7.6 apresenta-se o gráfico meio-normal de probabilidade com envelope simulado, mostando bom ajuste do modelo.

Considerando-se, então, os fatores tipo de citocinina e tipo de auxina, as pro-

Tabela 7.16: Desvios e estatísticas  $X^2$  residuais e respectivos números de graus de liberdade (g.l.) e níveis descritivos ( $p$ ), para os dados de regeneração de tecidos de macieiras (Tabela 7.15).

Modelo	g.l.	Desvios	$X^2$
1	179	202,44	180,0
Bl	170	193,79	180,0
Bl+A	169	165,65	179,5
Bl+B	168	189,72	183,4
Bl+C	168	187,07	178,9
Bl+A+B	167	160,83	181,8
Bl+A+C	167	157,57	207,1
Bl+B+C	166	182,83	181,5
Bl+A*B	165	157,57	189,4
Bl+A*C	165	147,82	210,5
Bl+B*C	162	179,27	186,0
Bl+A*B+C	163	149,03	221,3
Bl+A*C+B	163	142,27	214,2
Bl+B*C+A	161	148,08	203,1
Bl+A*B+A*C	161	138,66	197,9
Bl+A*B+B*C	159	141,71	547,9
Bl+B*C+A*C	159	137,05	232,7
Bl+A*B+A*C+B*C	157	132,65	213,5
Bl+A*B*C	153	127,00	152,4

porções de regeneração de tecidos estão apresentadas na Tabela 7.19. Verifica-se que a observação da casela em negrito é responsável pela interação significativa entre os dois fatores. Na realidade, esse experimento foi repetido outras quatro vezes e não houve evidência de interação significativa.

**Exemplo 7.6:** Toxicidade a dissulfeto de carbono gasoso

Os dados que se seguem referem-se a números de insetos mortos em amostras de tamanhos  $m_i$  depois de cinco horas de exposição a diferentes doses de dissulfeto de carbono gasoso (Bliss, 1935; Pregibon, 1980).

**FALTA TERMINAR ESTE EXEMPLO**

Tabela 7.17: Análise de Desvios para os dados de regeneração de tecidos de macieiras (Tabela 7.15), usando-se o preditor linear (7.10).

Fontes de Variação	g.l.	Desvios	$p$
Blocos	9	8,6	0,470
Tipo de Citocinina (A)	1	28,1	< 0,0001
Dose de Citocinina (B A)	2	4,8	0,090
Auxina (C A,B)	2	8,4	0,015
AB (A,B,C)	2	3,4	0,180
AC (A*B,C)	2	10,4	0,006
BC (A*B,A*C)	4	6,0	0,198
ABC	4	5,6	0,227
Resíduo	153	127,0	0,938
Total	179	202,4	

Tabela 7.18: Análise de Desvios para os dados de regeneração de tecidos de macieiras, usando-se o preditor linear (7.11).

Fontes de Variação	g.l.	Desvios	$p$
Blocos	9	8,6	0,470
Tipo de Citocinina (A)	1	28,1	< 0,0001
Auxina (C A)	2	8,1	0,018
AC (A,C)	2	9,7	0,008
Resíduo	165	147,8	0,827
Total	179	202,4	

## 7.2 Dados de contagem

### 7.2.1 Modelo de Poisson

A distribuição de Poisson tem um papel importante na análise de observações na forma de contagens. Se eventos ocorrem independente e aleatoriamente no tempo, com taxa média de ocorrência constante, o modelo determina o número de eventos num tempo especificado. Assim, se  $Y_i$ ,  $i = 1, \dots, n$ , são contagens com médias  $\mu_i$ , o modelo padrão Poisson assume que  $Y_i \sim \text{Pois}(\mu_i)$  com função de variância  $\text{Var}(Y_i) = \text{E}(Y_i) = \mu_i$ . A função de ligação canônica é a logarítmica

$$g(\mu_i) = \log(\mu_i) = \eta_i.$$

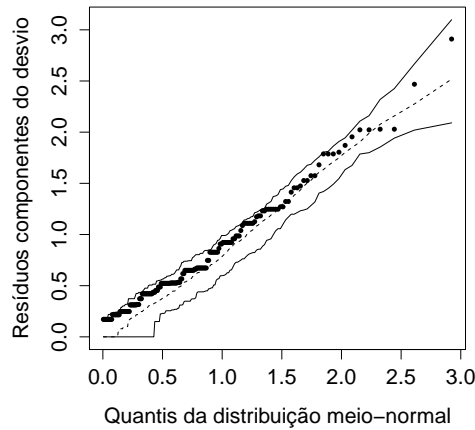


Figura 7.6: Gráfico meio-normal de probabilidade com envelope simulado para os dados de regeneração de tecidos de macieiras, usando-se o modelo logístico.

Tabela 7.19: Tabela de médias, considerando-se os fatores tipo de citocinina e tipo de auxina, para os dados de regeneração de tecidos de macieiras (Tabela 7.15).

Tipo de Citocinina	Tipo de Auxina		
	NAA	IBA	2-4D
BAP	<b>0,33</b>	0,67	0,77
TDZ	0,93	0,97	0,83

No caso de períodos (ou volumes, áreas, etc) de observação diferentes, tem-se

$$Y_i \sim \text{Pois}(t_i \lambda_i)$$

com  $\text{Var}(Y_i) = E(Y_i) = \mu_i = t_i \lambda_i$ . Tomando um modelo log-linear para as taxas,

$$\log(\lambda_i) = \mathbf{x}_i^T \boldsymbol{\beta}$$

resulta no seguinte modelo log-linear para as médias Poisson

$$\log(\mu_i) = \log(t_i \lambda_i) = \log(t_i) + \mathbf{x}_i^T \boldsymbol{\beta},$$

em que  $\log(t_i)$  é incluído como um termo fixo, ou *offset*, no modelo.

Tabela 7.20: Números de insetos mortos em amostras de tamanhos  $m_i$  depois de cinco horas de exposição a diferentes doses de dissulfeto de carbono gasoso.

Log <sub>10</sub> dosagem	Número de insetos expostos	Número de insetos mortos
1.6907	59	6
1.7242	60	13
1.7552	62	18
1.7842	56	28
1.8113	63	52
1.8369	59	53
1.8610	62	61
1.8839	60	60

Por outro, neste caso, pode-se, também, fazer análise de regressão Poisson ponderada com pesos  $t_i$  para as razões

$$R_i = \frac{Y_i}{t_i},$$

pois,

$$E[R_i] = \frac{t_i \lambda_i}{t_i} = \lambda_i = e^{\mathbf{x}_i^T \beta} = \mu_i^R$$

e

$$\text{Var}[R_i] = \frac{1}{t_i^2} \text{Var}(Y_i) = \frac{t_i \lambda_i}{t_i^2} = \frac{\lambda_i}{t_i} = \frac{\mu_i^R}{t_i}.$$

A seguir, serão apresentados alguns exemplos de aplicação.

### Exemplo 7.6: Armazenamento de microorganismos

Na Tabela 7.21, são apresentados valores de concentrações de bactérias (contagens por área fixa) feitas no congelamento inicial ( $-70^\circ\text{C}$ ) e após 1, 2, 6 e 12 meses (Francis et al., 1993).

Tabela 7.21: Concentrações de bactérias por área fixa

Tempo	0	1	2	6	12
Contagem	31	26	19	15	20

Pode-se supor, inicialmente, que  $Y_i$ , o número de bactérias por área fixa, segue a distribuição de Poisson com média  $\mu_i$ , isto é,  $Y_i \sim P(\mu_i)$ . Além disso, em geral, espera-se que a contagem média decresça com o tempo, isto é,

$$\mu_i \propto \frac{1}{(\text{tempo}_i)^\gamma}$$

e, portanto, pode-se escrever

$$\log(\mu_i) = \beta_0 + \beta_1 \log(\text{tempo}_i + 0,1), \quad (7.12)$$

sendo a constante 0,1 adicionada para evitar problemas com o tempo zero.

Tabela 7.22: Desvios residuais e estatísticas  $X^2$  e respectivos números de graus de liberdade (g.l.) e níveis descritivos ( $p$ ), para os dados de concentrações de bactérias por área fixa (Tabela 7.21).

Preditor linear	g.l.	Desvios	$p$	$X^2$	$p$
$\beta_0$	4	7,0672	0,1324	7,1532	0,1280
$\beta_0 + \beta_1 \log(\text{tempo}_i + 0,1)$	3	1,8338	0,6076	1,8203	0,6105

Na Tabela 7.22, são apresentados os desvios residuais e as estatísticas  $X^2$  e seus respectivos números de graus de liberdade e níveis descritivos ( $p$ ). Observa-se que existem evidências de bom ajuste de ambos os modelos. Entretanto, o número de graus de liberdade associado aos desvios residuais é muito pequeno, o que torna duvidosa a aproximação assintótica da distribuição das estatísticas desvios residuais e das estatísticas  $X^2$  de uma distribuição  $\chi^2$ .

Tabela 7.23: Análise de desvios para os dados de concentrações de bactérias por área fixa (Tabela 7.21), usando-se o preditor linear (7.12).

Fontes de Variação	g.l.	Desvios	$p$
Regressão linear	1	5,2334	0,0222
Resíduo	3	1,8338	
Total	4	7,0672	



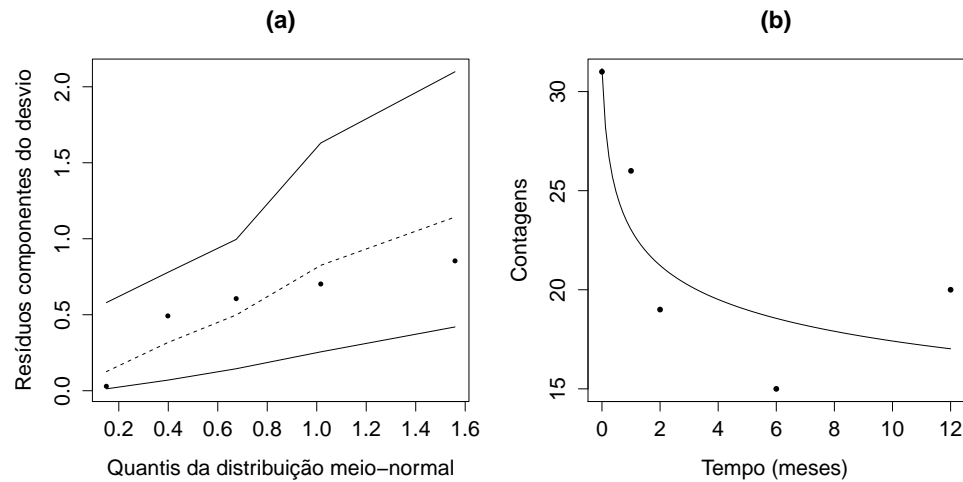


Figura 7.7: Modelo log-linear para os dados de concentrações de bactérias por área fixa (Tabela 7.21), considerando o preditor linear (7.12) – (a) gráfico meio-normal de probabilidade com envelope simulado; (b) curvas ajustadas com as observações.

Adicionalmente, tem-se que de acordo com a análise de desvios apresentada na Tabela 7.23, há evidências do efeito significativo da regressão linear. Pelo gráfico meio-normal de probabilidade com envelope simulado (Figura 7.7(a)) confirma-se que o preditor linear (7.12) é adequado, obtendo-se, então, a equação da curva ajustada

$$\hat{\mu}_i = e^{3,1489 - 0,1261 \log(\text{tempo}_i + 0,1)},$$

que pode ser observada na Figura 7.7(b) juntamente com os valores observados.

### Exemplo 7.7: Número de brotos em um estudo de micropropagação de macieiras

Os dados apresentados na Tabela 7.24 referem-se ao número de brotos produzidos por explante em um experimento de micropropagação. O delineamento experimental utilizado foi o inteiramente casualizado com os tratamentos no esquema fatorial  $3 \times 2$ , isto é, 3 meios de cultura aos quais era adicionada uma quantia de hormônio (2 níveis, X1: quantia pequena e X2: quantia grande). As parcelas eram constituídas de recipientes com 3 explantes e os dados estão apresentados em grupos de 3 para indicar os recipientes diferentes. Inicialmente, havia 10 recipientes (portanto, 30 explantes) para cada tratamento, porém,

Tabela 7.24: Números de brotos de macieiras por explante.

Meio de Cultura	Hormônio	Número de Brotos														
A	X1	4	5	2	1	2	5	2	2	*						
A	X2	3	5	3	2	2	1	2	2	3	1	4	4	2	4	*
		2	*	*												
B	X1	4	1	4	5	4	5	5	4	3	3	4	4	2	3	2
		0	4	2	6	2	2	3	3	*	1	5	*			
B	X2	2	2	1	2	4	4	2	3	0	0	0	4	<b>12</b>	<b>0</b>	<b>4</b>
		0	8	2	2	4	*	3	1	*	10	*	*			
C	X1	0	2	0	1	1	3	5	3	3	3	2	1	2	2	2
		2	2	2	2	0	2									
C	X2	2	2	3	<b>11</b>	<b>6</b>	<b>5</b>	5	3	4	6	4	*	4	4	*

alguns explantes morreram. No caso em que morreram todos os explantes, o recipiente foi eliminado do experimento, pois algumas dessas mortes podem decorrer da contaminação com bactérias o que não está relacionado com o tratamento. No caso em que houve uma ou duas mortes no recipiente, os dados foram considerados e usou-se \* no lugar do dado faltante. O objetivo é verificar se existe interação entre meio de cultura e quantidade de hormônio e se influenciam o número de brotos.

A variável resposta,  $Y_i$ , é o número de de brotos, e, portanto, a distribuição a ser considerada é a Poisson. Como função de ligação pode ser usada a logarítmica  $g(\mu) = \log(\mu)$  e como parte sistemática tem-se um delineamento inteiramente casualizado com  $k$  (1,2,3) plantas por parcela e como fatores, meio de cultura e quantidade de hormônio. O preditor linear inicial a ser considerado é

$$\eta = \alpha_i + \beta_j + \alpha\beta_{ij} + \xi_k,$$

para o qual é apresentada a análise de desvios, na Tabela 7.25. O desvio residual (94, 76,  $p = 0,0444$ ) com 73 graus de liberdade, mostra evidência marginal de falta de ajuste.

#### FALTA REVER o Final desse exemplo

Verifica-se que a diferença entre os desvios obtidos para “entre recipientes” e “entre plantas dentro de recipientes” (156, 76 – 94, 76 = 62, 0), com (111 – 73 = 38) graus de liberdade, é significativa. Adotou-se, então, como preditor linear  $\eta = \alpha_i + \beta_j + \alpha\beta_{ij}$ , obtendo-se os resultados da Tabela 7.26.

Tabela 7.25: Análise de Desvios.

Fontes de Variação	g.l.	Desvios	<i>p</i>
Meio de cultura (M)	2	0,42	0,810
Níveis de hormônio (H)	1	5,20	0,023
Interação M x H	2	14,93	0,001
Entre recipientes	38	61,99	0,008
Entre pl. d. recipientes	73	94,70	0,044
Total	116	177,31	

Tabela 7.26: Análise de Desvios.

Fontes de Variação	g.l.	Desvios	<i>p</i>
Meio de cultura (M)	2	0,42	0,810
Níveis de hormônio (H)	1	5,20	0,023
Interação M x H	2	14,93	0,001
Resíduo	111	156,76	0,003
Total	116	177,31	

O desvio residual mostra que existem evidências de que o modelo usado está se ajustando relativamente bem aos dados. Há necessidade, porém, de se utilizarem outras técnicas de diagnóstico como complementação. Observa-se, ainda, que a interação entre meios de cultura e níveis de hormônio é significativa. Ao se observar o quadro de médias apresentado na Tabela 7.27, verifica-se que a interação está sendo significativa devido ao meio de cultura C. O exame da Tabela 7.24 indica duas parcelas em destaque cuja influência na análise mereceria ser melhor estudada.

Tabela 7.27: Quadro de médias.

Meios de Cultura	Níveis de Hormônio		Médias
	Baixo	Alto	
A	2,9	2,6	2,7
B	3,0	2,9	2,9
C	<b>1,8</b>	4,3	2,8
Médias	2,5	3,2	

**Exemplo 7.8: Números de espécies de plantas**

Os dados da Tabela 7.28 (Crawley, 2007) referem-se a números de espécies de plantas em parcelas com diferentes quantidades de biomassa (variável contínua) e solos com diferentes pH (fator com três níveis: alto, médio e baixo). O interesse desse estudo é verificar se existe relação entre o número de espécies de plantas e a quantidade de biomassa e se essa relação é diferente para os solos com diferentes níveis de pH.

Pode-se supor que variável resposta, número de espécies de plantas, tem distribuição Poisson com média  $\mu_i$ , isto é,  $Y_i \sim P(\mu_i)$ , com função de ligação logarítmica. Para o preditor linear, adotam-se os modelos de retas paralelas e de retas concorrentes, isto é,

$$\text{Modelo 1 : } \eta_i = \beta_{1j} + \beta_2 \text{ biomassa}_i, \quad j = 1, 2, 3$$

$$\text{Modelo 2 : } \eta_i = \beta_{1j} + \beta_{2j} \text{ biomassa}_i, \quad j = 1, 2, 3$$

O desvio residual para o modelo de retas paralelas é 99,24 com 86 graus de liberdade, enquanto que para o modelo de retas concorrentes é 83,20 ( $p = 0,50$ ) com 84 graus de liberdade, sendo 16,04 a diferença de desvios com 2 graus de liberdade, mostrando evidência significativa ( $p = 0,00033$ ) em favor do modelo de retas concorrentes. Além disso, existe evidência de efeito significativo do fator pH (desvio 187,23 com 2 graus de liberdade,  $p < 0,001$ ) e de regressão linear de biomassa dentro de pH (desvio 181,92 com 3 graus de liberdade,  $p < 0,001$ ). Por outro lado, o modelo de retas concorrentes está bem ajustado aos dados conforme mostra o gráfico meio-normal de probabilidades (Figura 7.8a), obtendo-se, então, as equações das curvas ajustadas

$$\text{Solo de pH alto : } \hat{\mu}_i = e^{(3,7681 - 0,1071 \text{ biomassa}_i)}$$

$$\text{Solo de pH médio : } \hat{\mu}_i = e^{(3,4367 - 0,1390 \text{ biomassa}_i)}$$

$$\text{Solo de pH baixo : } \hat{\mu}_i = e^{(2,9526 - 0,2622 \text{ biomassa}_i)}$$

que estão representadas na Figura 7.8b, juntamente com os valores observados.

**7.2.2 Modelos log-lineares para tabelas de contingência**

O modelo de Poisson desempenha na análise de dados categorizados o mesmo papel que o modelo normal na análise de dados contínuos. A diferença fundamental está em que a estrutura multiplicativa para as médias do modelo de Poisson é mais apropriada do que a estrutura aditiva do modelo normal. Ele é especialmente útil na análise de tabelas de

Tabela 7.28: Números de espécies ( $Y$ ) de plantas em parcelas com diferentes quantidades de biomassa e solos com diferentes pH (Crawley, 2007).

pH	Biomassa	$Y$	pH	Biomassa	$Y$	pH	Biomassa	$Y$
high	0,469	30	mid	0,176	29	low	0,101	18
high	1,731	39	mid	1,377	30	low	0,139	19
high	2,090	44	mid	2,551	21	low	0,864	15
high	3,926	35	mid	3,000	18	low	1,293	19
high	4,367	25	mid	4,906	13	low	2,469	12
high	5,482	29	mid	5,343	13	low	2,367	11
high	6,685	23	mid	7,700	9	low	2,629	15
high	7,512	18	mid	0,554	24	low	3,252	9
high	8,132	19	mid	1,990	26	low	4,417	3
high	9,572	12	mid	2,913	26	low	4,781	2
high	0,087	39	mid	3,216	20	low	0,050	18
high	1,237	35	mid	4,980	21	low	0,483	19
high	2,532	30	mid	5,659	15	low	0,653	13
high	3,408	30	mid	8,100	8	low	1,555	9
high	4,605	33	mid	0,740	31	low	1,672	8
high	5,368	20	mid	1,527	28	low	2,870	14
high	6,561	26	mid	2,232	18	low	2,511	13
high	7,242	36	mid	3,885	16	low	3,498	4
high	8,504	18	mid	4,627	19	low	3,679	8
high	9,391	7	mid	5,121	20	low	4,832	2
high	0,765	39	mid	8,300	6	low	0,290	17
high	1,176	39	mid	0,511	25	low	0,078	14
high	2,325	34	mid	1,478	23	low	1,429	15
high	3,223	31	mid	2,935	25	low	1,121	17
high	4,136	24	mid	3,505	22	low	1,508	9
high	5,137	25	mid	4,618	15	low	2,326	8
high	6,422	20	mid	5,697	11	low	2,996	12
high	7,066	21	mid	6,093	17	low	3,538	14
high	8,746	12	mid	0,730	24	low	4,365	7
high	9,982	11	mid	1,158	27	low	4,871	3

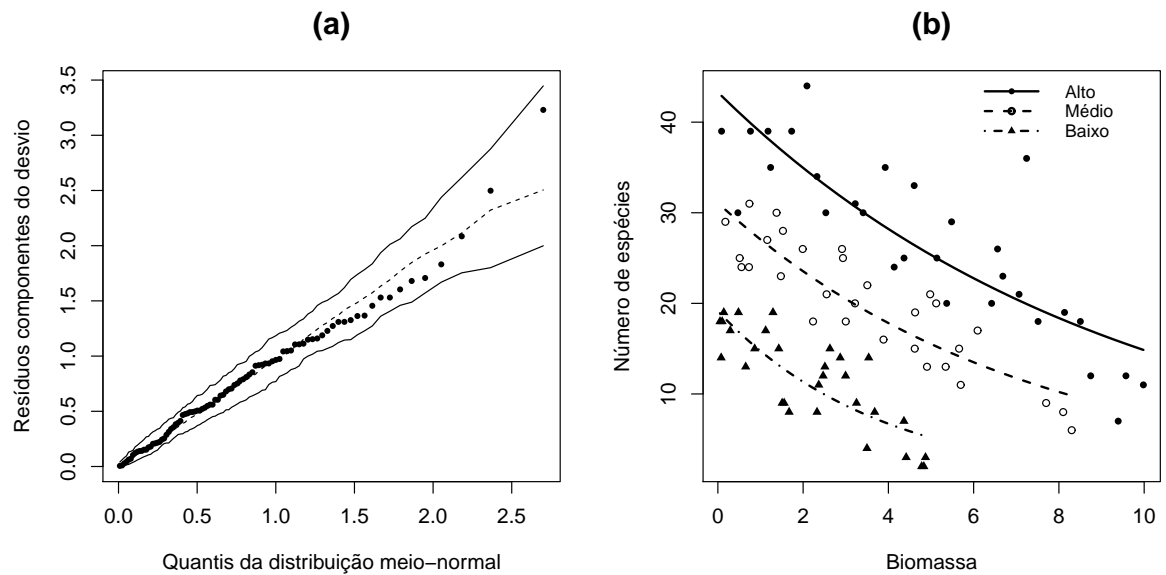


Figura 7.8: Modelo log-linear para número de espécies de plantas como função da quantidade de biomassa e diferentes pH, considerando o preditor linear de retas concorrentes – (a) gráfico meio-normal de probabilidade com envelope simulado; (b) curvas ajustadas com as proporções observadas.

contingência em que as observações consistem de contagens ou frequências nas caselas pelo cruzamento das variáveis resposta(s) e explanatórias e as hipóteses mais comuns podem ser expressas como modelos multiplicativos para as frequências esperadas das caselas (McCullagh e Nelder, 1989; Agresti, 2002), conforme ilustração de um caso simples na Seção 2.2.3.

Entretanto, é importante entender como o planejamento do estudo afeta a escolha dos modelos probabilísticos a serem ajustados aos dados (Dobson e Barnett, 2008). Diferente dos modelos vistos até agora, modelos lineares generalizados para dados categorizados podem ser definidos quando há mais de uma variável resposta, isto é, pode-se considerar o uso de um modelo univariado para análise de dados multivariados, como ilustrado a seguir.

Na Tabela 2.4 do Exemplo 2.4 que apresenta números de insetos coletados em armadilhas adesivas de duas cores e sexados, têm-se duas respostas: tipo de armadilha escolhida e sexo do inseto. Note que o total de insetos coletados não é fixado previamente.

O interesse, nesse caso, é estudar se não existe associação entre tipo de armadilha escolhida e sexo do inseto (hipótese de independência), condicionado ao total  $N = n$ .

Na Tabela 2.5 do Exemplo 2.5 que apresenta contagens de plantas segregando para dois caracteres, ciclo (tardio ou precoce) e viriscência (normal ou viriscente), numa progênie da espécie X, fixado um total de 900 plantas, têm-se duas respostas: ciclo e viriscência. Como no exemplo anterior, o interesse é estudar se não existe associação entre as duas respostas (hipótese de independência)

Na Tabela 2.6 do Exemplo 2.6 que apresenta números de tomates sadios e com broca, de acordo com quatro inseticidas aplicados nos tomateiros, tem-se uma variável resposta (fruto sadio ou não) e uma variável explanatória (tipo de inseticida). Note que o total de tomates examinados não é fixado previamente. Nesse caso, o interesse é verificar se a proporção de tomates sadios é a mesma para todos os inseticidas (hipótese de homogeneidade), condicionando ao conhecimento dos totais de tomates examinados em cada inseticida.

De uma forma mais geral, seja, inicialmente, uma tabela de contingência com duas entradas, como mostrado na Tabela 7.34, podendo o total ( $y_{..}$ ) ser fixo ou aleatório, o mesmo podendo ocorrer com totais de linhas ( $y_{i.}$ ) ou colunas ( $y_{.j}$ ).

Tabela 7.29: Tabela de contingência com duas entradas

A	B				
	1	2	...	b	
1	$y_{11}$	$y_{12}$	...	$y_{1b}$	$y_{1.}$
2	$y_{21}$	$y_{22}$	...	$y_{2b}$	$y_{2.}$
...	...	...	...	...	...
$a$	$y_{a1}$	$y_{a2}$	...	$y_{ab}$	$y_{a.}$
	$y_{.1}$	$y_{.2}$	...	$y_{.b}$	$y_{..}$

Três situações, então, podem ocorrer:

a) *Caso mais simples em que total geral é aleatório*

Assume-se que as variáveis aleatórias  $Y_{ij}$  são independentes com distribuição de

Poisson de média  $\mu_{ij}$ , isto é,  $Y_{ij} \sim P(\mu_{ij})$ . A distribuição conjunta dos  $Y_{ij}$  é obtida por

$$f(\mathbf{y}; \boldsymbol{\mu}) = \prod_{i=1}^a \prod_{j=1}^b \frac{\mu_{ij}^{y_{ij}} e^{-\mu_{ij}}}{y_{ij}!}$$

em que  $E(Y_{ij}) = \mu_{ij}$ . A função de ligação canônica para o modelo Poisson é a logarítmica.

b) *Caso em que o total geral é (condicionalmente) fixo*

Nesse caso, tem-se que a soma de variáveis aleatórias independentes com distribuição de Poisson  $N = \sum_{i=1}^a \sum_{j=1}^b Y_{ij} = Y_{..} \sim P(\mu)$ , em que  $\mu = \sum_{i=1}^a \sum_{j=1}^b \mu_{ij}$ . Portanto, a distribuição conjunta de  $(Y_{11}, Y_{12}, \dots, Y_{a,b} | N = y_{..})$  é obtida por

$$f(\mathbf{y} | N = y_{..}) = \prod_{i=1}^a \prod_{j=1}^b \frac{\mu_{ij}^{y_{ij}} e^{-\mu_{ij}}}{y_{ij}!} \frac{y_{..}!}{\mu^{y_{..}} e^{-\mu}} = y_{..}! \prod_{i=1}^a \prod_{j=1}^b \left( \frac{\mu_{ij}}{\mu} \right)^{y_{ij}} \frac{1}{y_{ij}!} = y_{..}! \prod_{i=1}^a \prod_{j=1}^b \frac{\pi_{ij}^{y_{ij}}}{y_{ij}!},$$

isto é,

$$(Y_{11}, Y_{12}, \dots, Y_{a,b} | N = y_{..}) \sim \text{Multinomial}(\pi_{11}, \pi_{12}, \dots, \pi_{a,b}),$$

em que  $0 \leq \pi_{ij} = \frac{\mu_{ij}}{\mu} \leq 1$ ,  $\sum_{i=1}^a \sum_{j=1}^b \pi_{ij} = 1$  e

$$E(Y_{ij}) = \mu_{ij} = y_{..} \pi_{ij}. \quad (7.13)$$

Sob a hipótese de independência,  $H_0 : \pi_{ij} = \pi_{i.} \pi_{.j}$ , a equação (7.13) fica,

$$\mu_{ij} = y_{..} \pi_{ij} = y_{..} \pi_{i.} \pi_{.j} = \frac{1}{y_{..}} \mu_{i.} \mu_{.j}$$

e, portanto,

$$\log(\mu_{ij}) = \log(y_{..}) + \log(\pi_{i.}) + \log(\pi_{.j}) = \beta + \alpha_i + \gamma_j.$$

Por analogia, com análise de variância, o modelo maximal (nesse caso, saturado) pode ser escrito como

$$\log(\mu_{ij}) = \beta + \alpha_i + \gamma_j + \alpha\gamma_{ij},$$

correspondendo ao modelo  $\mu_{ij} = y_{..} \pi_{ij}$ . Assim, a hipótese de independência  $H_0 : \pi_{ij} = \pi_{i.} \pi_{.j}, \forall i, \forall j$  é equivalente à hipótese de não existência de interação, isto é,  $H_0 : \alpha\gamma_{ij} = 0, \forall i, \forall j$ . O modelo minimal, nesse caso, é

$$\log(\mu_{ij}) = \beta$$



visto que apenas o total geral é fixado.

c) *Caso em que os totais de linhas (ou colunas) são (condicionalmente) fixos*

A distribuição para cada linha (ou coluna) é multinomial e as linhas (ou colunas) são supostamente independentes. Assumindo, então, totais de linhas fixos, a distribuição conjunta dos  $(Y_{11}, Y_{12}, \dots, Y_{a,b} | Y_{1.} = y_{1.}, \dots, Y_{a.} = y_{a.})$  é obtida por

$$\begin{aligned} f(\mathbf{y} | Y_{1.} = y_{1.}, \dots, Y_{a.} = y_{a.}) &= \prod_{i=1}^a \frac{y_{i.}!}{\mu_{i.}^{y_{i.}} e^{-\mu_{i.}}} \prod_{j=1}^b \frac{\mu_{ij}^{y_{ij}} e^{-\mu_{ij}}}{y_{ij}!} \\ &= \prod_{i=1}^a y_{i.}! \prod_{j=1}^b \left( \frac{\mu_{ij}}{\mu_{i.}} \right)^{y_{ij}} \frac{1}{y_{ij}!} = \prod_{i=1}^a y_{i.}! \prod_{j=1}^b \frac{\pi_{ij}^{y_{ij}}}{y_{ij}!}, \end{aligned}$$

isto é,

$$(Y_{11}, \dots, Y_{a,b} | Y_{1.} = y_{1.}, Y_{2.} = y_{2.}, \dots, Y_{a.} = y_{a.}) \sim \text{Produto multinomial}(\pi_{11}, \pi_{12}, \dots, \pi_{a,b}),$$

em que  $0 \leq \pi_{ij} = \frac{\mu_{ij}}{\mu_{i.}} \leq 1$ ,  $\sum_{j=1}^b \pi_{ij} = 1$  e

$$E(Y_{ij}) = \mu_{ij} = y_{i.} \pi_{ij}. \quad (7.14)$$

Sob a hipótese de homogeneidade,  $H_0 : \pi_{ij} = \pi_{.j}, \forall j$ , a equação (7.14) fica

$$\mu_{ij} = y_{i.} \pi_{ij} = y_{i.} \pi_{.j} = \frac{1}{y_{..}} y_{i.} \mu_{.j}$$

e, portanto,

$$\log(\mu_{ij}) = \log(y_{i.}) + \log(\pi_{.j}) = \beta + \alpha_i + \gamma_j.$$

Novamente, por analogia com análise de variância, o modelo maximal (saturado) pode ser escrito como

$$\log(\mu_{ij}) = \beta + \alpha_i + \gamma_j + \alpha\gamma_{ij},$$

correspondendo ao modelo  $\mu_{ij} = y_{i.} \pi_{ij}$ . Assim, a hipótese de homogeneidade  $H_0 : \pi_{ij} = \pi_{.j}, \forall j$  é equivalente à hipótese de não existência de interação, isto é,  $H_0 : \alpha\gamma_{ij} = 0, \forall j$ . O modelo minimal, nesse caso, é

$$\log(\mu_{ij}) = \beta + \alpha_i$$

visto que os totais marginais de linhas são fixados.

É importante notar que, nos dois últimos casos, os modelos aditivos obtidos sob as hipóteses de nulidade, permitem que os totais observados das tabelas sejam reproduzidos. De forma geral, pode ser verificado que as hipóteses mais comuns para dados dispostos em tabelas de contingência podem ser expressas como modelos multiplicativos para as frequências esperadas das caselas (McCullagh e Nelder, 1989; Agresti, 2002; Paulino e Singer, 2006) e que a média  $\mu$  é obtida como um produto de outras médias marginais. Assim, uma transformação logarítmica do valor esperado lineariza essa parte do modelo (daí vem a terminologia de modelo log-linear). A seguir, serão apresentados alguns casos particulares.

### *Modelos log-lineares para tabelas $2 \times 2$*

Considere o caso particular de uma tabela de contingência  $2 \times 2$ , como mostrado na Tabela 7.30, em que os  $y_{ij}$  são contagens associadas aos fatores  $A$  e  $B$ . Tem-se que o total geral da tabela pode ser aleatório (Exemplo 2.4) ou fixo (Exemplo 2.5). Entretanto, em ambos os casos, o interesse está em verificar se existe associação entre os fatores  $A$  e  $B$ , condicionado ao total  $N = y_{..}$ .

Tabela 7.30: Tabela de contingência  $2 \times 2$

A	B		
	1	2	
1	$y_{11}$	$y_{12}$	$y_{1.}$
2	$y_{21}$	$y_{22}$	$y_{2.}$
	$y_{.1}$	$y_{.2}$	$y_{..}$

Uma forma de se medir a associação entre dois fatores  $A$  e  $B$  é por meio da razão de chances. A chance de um evento é expressa pela razão entre a probabilidade de sucesso e a probabilidade de insucesso, isto é,  $\frac{\pi}{1 - \pi}$  enquanto que a razão de chances entre dois eventos é expressa por  $\psi = \frac{\pi_2(1 - \pi_1)}{\pi_1(1 - \pi_2)} = \frac{\mu_{11} \mu_{22}}{\mu_{12} \mu_{21}}$ , que para a Tabela 7.30 é estimada por

$$\hat{\psi} = \frac{y_{11} \times y_{22}}{y_{12} \times y_{21}}.$$

O interesse está em se saber se as chances são iguais para ambos os eventos, isto é, testar

a hipótese  $H_0 : \psi = 1$ . Isso corresponde, ao teste de independência para tabelas de contingência, como será mostrado a seguir.

Por outro lado, como mostrado, pode-se usar um modelo log-linear com preditores lineares para o modelo de independência e o modelo saturado.

(i) **Modelo de independência:** representado simbolicamente por A+B

Como descrito, o preditor linear para o modelo sob independência dos fatores  $A$  e  $B$  pode ser expresso por

$$\log(\mu_{ij}) = \mu + \alpha_i + \beta_j = \lambda + \lambda_i^A + \lambda_j^B, \quad i, j = 1, 2 \quad (7.15)$$

com  $\lambda_1^A = \lambda_1^B = 0$ , isto é, com preditor linear resumido na Tabela 7.31. Portanto, sob

Tabela 7.31: Preditor linear para tabela de contingência  $2 \times 2$ , sob independência

A	B	
	1	2
1	$\lambda$	$\lambda + \lambda_2^B$
2	$\lambda + \lambda_2^A$	$\lambda + \lambda_2^A + \lambda_2^B$

independência, o logaritmo da razão das chances teórica é igual 0, isto é,

$$\begin{aligned} \log(\psi) &= \log(\mu_{11}) + \log(\mu_{22}) - \log(\mu_{12}) - \log(\mu_{21}) \\ &= (\lambda + \lambda_2^A + \lambda_2^B) + \lambda - (\lambda + \lambda_2^B) - (\lambda + \lambda_2^A) = 0, \end{aligned}$$

e a razão das chances é  $\psi = 1$ .

A partir do preditor linear  $\log(\mu_{ij})$  dado pela equação (7.15), definem-se a matriz  $\mathbf{X}$  e o vetor  $\boldsymbol{\beta}$  por

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \lambda \\ \lambda_2^A \\ \lambda_2^B \end{bmatrix}$$

Usando-se a equação (3.7), as estimativas de  $\lambda$ ,  $\lambda_2^A$  e  $\lambda_2^B$  são obtidas a partir de

$$y_{..} = \hat{\mu}_{..} = e^{\hat{\lambda}} + e^{\hat{\lambda} + \hat{\lambda}_2^A} + e^{\hat{\lambda} + \hat{\lambda}_2^B} + e^{\hat{\lambda} + \hat{\lambda}_2^A + \hat{\lambda}_2^B} = e^{\hat{\lambda}}(1 + e^{\hat{\lambda}_2^A})(1 + e^{\hat{\lambda}_2^B}) \quad (7.16)$$

$$y_{2.} = \hat{\mu}_{2.} = e^{\hat{\lambda} + \hat{\lambda}_2^A} + e^{\hat{\lambda} + \hat{\lambda}_2^A + \hat{\lambda}_2^B} = e^{\hat{\lambda} + \hat{\lambda}_2^A}(1 + e^{\hat{\lambda}_2^B}) \quad (7.17)$$

e

$$y_{.2} = \hat{\mu}_{.2} = e^{\hat{\lambda} + \hat{\lambda}_2^B} + e^{\hat{\lambda} + \hat{\lambda}_2^A + \hat{\lambda}_2^B} = e^{\hat{\lambda} + \hat{\lambda}_2^B} (1 + e^{\hat{\lambda}_2^A}). \quad (7.18)$$

Dividindo-se (7.16) por (7.18), (7.16) por (7.17) e (7.16) pelo produto de (7.17) e (7.18), após algumas operações algébricas, obtêm-se

$$\hat{\lambda} = \log \left( \frac{y_{1.} y_{.1}}{y_{..}} \right), \quad \hat{\lambda}_2^A = \log \left( \frac{y_{2.}}{y_{1.}} \right) \quad \text{e} \quad \hat{\lambda}_2^B = \log \left( \frac{y_{.2}}{y_{.1}} \right).$$

Além disso, a partir de (7.16), tem-se

$$\log y_{..} = \hat{\lambda} + \log \left( 1 + \frac{y_{2.}}{y_{1.}} \right) + \log \left( 1 + \frac{y_{.2}}{y_{.1}} \right) = \hat{\lambda} + \log \left( \frac{y_{..}}{y_{1.}} \right) + \log \left( \frac{y_{..}}{y_{.1}} \right)$$

implicando em

$$\hat{\lambda} = \log \left( y_{..} \frac{y_{1.} y_{.1}}{y_{..} y_{..}} \right) = \log \left( \frac{y_{1.} y_{.1}}{y_{..}} \right),$$

isto é,

$$e^{\hat{\lambda}} = \hat{\mu}_{11} = y_{..} \frac{y_{1.} y_{.1}}{y_{..} y_{..}} = y_{..} \hat{\pi}_{1.} \hat{\pi}_{.1},$$

confirmando a hipótese de independência. De forma semelhante, obtêm-se

$$\hat{\mu}_{12} = e^{\hat{\lambda} + \hat{\lambda}_2^B} = e^{\hat{\lambda}} e^{\hat{\lambda}_2^B} = y_{..} \frac{y_{1.} y_{.1} y_{.2}}{y_{..} y_{..} y_{.1}} = y_{..} \frac{y_{1.} y_{.2}}{y_{..} y_{..}} = y_{..} \hat{\pi}_{1.} \hat{\pi}_{.2}.$$

(ii) **Modelo saturado ou de interação:** representado simbolicamente por  $A*B \equiv A+B+A.B$

O preditor linear, nesse caso, é expresso como

$$\log(\mu_{ij}) = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}, \quad i, j = 1, 2,$$

com  $\lambda_1^A = \lambda_1^B = \lambda_{1j}^{AB} = \lambda_{i1}^{AB} = 0$ , isto é, o preditor linear  $\log(\mu_{ij})$  conforme o quadro que se segue

A	B	
	1	2
1	$\lambda$	$\lambda + \lambda_2^B$
2	$\lambda + \lambda_2^A$	$\lambda + \lambda_2^A + \lambda_2^B + \lambda_{22}^{AB}$

A partir desse preditor linear, definem-se a matriz  $\mathbf{X}$  e o vetor  $\boldsymbol{\beta}$  por

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \lambda \\ \lambda_2^A \\ \lambda_2^B \\ \lambda_{22}^{AB} \end{bmatrix}.$$

Pode-se mostrar que

$$\hat{\lambda} = \log(m_{11}), \quad \hat{\lambda}_2^A = \log\left(\frac{y_{21}}{y_{11}}\right), \quad \hat{\lambda}_2^B = \log\left(\frac{y_{12}}{y_{11}}\right) \quad \text{e} \quad \hat{\lambda}_{22}^{AB} = \log\left(\frac{y_{22}y_{11}}{y_{12}y_{21}}\right) = \log(\hat{\psi}).$$

Tem-se, portanto, que o logaritmo da razão de chances corresponde ao parâmetro de interação e testar a hipótese  $H_0 : \psi = 1 \Rightarrow \log(\psi) = 0$  é o mesmo que testar o efeito da interação no modelo log-linear.

### Exemplo 7.9: Coletas de insetos em armadilhas adesivas

Considere os dados descritos no Exemplo 2.4 em que os insetos, de uma determinada espécie, coletados em armadilhas adesivas de duas cores são sexados, tendo como objetivo verificar se há influência da cor da armadilha sobre a atração de machos e fêmeas. Tem-se,

$$\text{Razão de chances observada} = \hat{\psi} = \frac{246 \times 32}{458 \times 17} = 1,01.$$

Na Tabela 7.32, apresentam-se os desvios e as estatísticas  $X^2$  residuais e seus respectivos números de graus de liberdade (g.l.), considerando-se o modelo log-linear.

Tabela 7.32: Desvios e estatísticas  $X^2$

Model	d.f.	Desvios	$X^2$
Cor da armadilha + sexo	1	0,001254	0,001252
Cor da armadilha * sexo	0	0	

Observa-se que existem evidências que o modelo de independência ajusta-se bem aos dados. Como esperado, o modelo de interação (saturado) tem desvio e estatística  $X^2$  iguais a zero. As estimativas dos parâmetros do modelo saturado são:

Tabela 7.33: Estimativas

parâmetro	estimativa	e.p.	estimativa	e.p.
1	5,505	0,0638	5,505	0,0624
2	0,622	0,0790	0,622	0,0764
3	-2,672	0,2508	2,665	0,1478
4	0,011	0,3104		

É importante notar que o modelo saturado reproduz os dados e que o logaritmo da razão de chances ajustada é  $\log(\hat{\psi}) = 0,011$ , resultando em  $\hat{\psi} = \exp(0,011) = 1,01$ .

As estimativas para o modelo de efeitos principais (independência) são:

	estimativa	e.p.	parâmetro
1	5,505	0,0624	1
2	0,622	0,0764	armcor(2)
3	-2,665	0,1478	sexo(2)

Nota-se, agora, que o logaritmo da razão de chances é zero. Pela Tabela 7.32, tem-se que a diferença de desvios é 0,00125 ( $p = 0,9117$ ), não significativa, isto é, existem evidências para não se rejeitar a hipótese que a razão de chances é igual a 1, isto é, não há associação entre sexo do inseto e preferência por cor de armadilha adesiva.

### Modelos log-lineares para tabelas $a \times b$

De uma forma mais geral, seja, inicialmente, uma tabela de contingência com duas entradas, como mostrado na Tabela 7.34, podendo o total ( $y_{..}$ ) ser fixo ou aleatório, o mesmo podendo ocorrer com totais de linhas ( $y_{i.}$ ) ou colunas ( $y_{.j}$ ).

Tabela 7.34: Tabela de contingência com duas entradas

A	B				
	1	2	...	K	
1	$y_{11}$	$y_{12}$	...	$y_{1K}$	$y_{1.}$
2	$y_{21}$	$y_{22}$	...	$y_{2K}$	$y_{2.}$
...	...	...	...	...	...
$J$	$y_{J1}$	$y_{J2}$	...	$y_{JK}$	$y_{J.}$
	$y_{.1}$	$y_{.2}$	...	$y_{.K}$	$y_{..}$

## 7.3 Exercícios

1. Use o método delta para estimar as doses efetivas  $\theta_p$  de uma droga correspondente ao valor 100p% da taxa de mortalidade para os modelos probito e complemento-log-log.
2. Calcule a matriz de covariância assintótica das estimativas de máxima verossimilhança de  $\beta_0$  e  $\beta_1$  no modelo (7.1), considerando para  $g(\cdot)$  as funções de ligação: logito, probito e complemento log-log. Apresente as equações de estimação dos modelos descritos.
3. Apresente as fórmulas da razão de verossimilhanças para construir intervalos de confiança para a dose efetiva  $\theta_p$  nos modelos logito, probito e complemento log-log.
4. Aranda-Ordaz (1981) propôs o seguinte modelo para representar a probabilidade de sucesso

$$\pi(x) = 1 - (1 + \lambda e^{\beta_0 + \beta_1 x})^{-1/\lambda}$$

para  $\lambda e^{\beta_0 + \beta_1 x} > 1$ . (a) Estimar os parâmetros  $\beta_0$ ,  $\beta_1$  e  $\lambda$  por máxima verossimilhança. (b) Usar o método delta para construir os intervalos de confiança para a dose efetiva  $\theta_p$ .

**completar com mais exemplos**





# Referências Bibliográficas

- Agresti, A. (2002). *Categorical Data Analysis*. John Wiley & Sons, New York, second edition.
- Aitkin, M.; Francis, B.; Hinde, J.; Darnell, R. (2009). *Statistical modelling in R*. Oxford University Press, Oxford.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Auto Cntl AC-19*, **6**, 716–723.
- Andrews, D. F.; Pregibon, D. (1978). Finding the outliers that matter. *Journal of the Royal Statistical Society B*, **40**, 87–93.
- Anscombe, F. J. (1953). Contribution to the discussion of h. hotelling’s paper. *J. R. Statist. Soc. B*, **15**, 229–230.
- Anscombe, F. J. (1964). Normal likelihood functions. *Ann. Inst. Statist. Math.*, **16**, 1–19.
- Aranda-Ordaz, F. (1981). On the families of transformations to additivity for binary response data. *Biometrika*, **68**, 357–363.
- Ashton, W. D. (1972). *The Logit Transformation with Special Reference to its Uses in Bioassay*. Griffin, London.
- Atkinson, A. C. (1981). Robustness, transformations and two graphical displays for outlying and influential observations in regression. *Biometrika*, **68**, 13–20.
- Atkinson, A. C. (1985). *Transformations and Regression*. Oxford University Press, Oxford.

- Atkinson, A. C.; Davison, A. C.; Nelder, J. A.; O'Brien, C. M. (1989). *Model Checking*. Imperial College, London.
- Barndorff-Nielsen, O. E. (1978). *Information and exponential families in statistical theory*. John Wiley & Sons, New York.
- Belsley, D. A.; Kuh, E.; Welsch, R. E. (1980). *Regression diagnostics: identifying influential data and sources of collinearity*. John Wiley, New York.
- Berkson, J. (1944). Application of the logistic function to bioassay. *J. R. Statist. Soc. B*, **39**, 357–365.
- Birch, M. W. (1963). Maximum likelihood in three-way contingency tables. *J. R. Statist. Soc. B*, **25**, 220–233.
- Bliss, C. I. (1935). The calculator of the dosage-mortality curve. *Ann. Appl. Biol.*, **22**, 134–167.
- Box, G. E. P.; Cox, D. R. (1964). An analysis of transformation. *J. R. Statist. Soc. B*, **26**, 211–252.
- Box, G. E. P.; Tidwell, P. W. (1962). Transformations of the independent variables. *Technometrics*, **4**, 531–550.
- Buse, A. (1982). The likelihood ratio, wald and lagrange multiplier tests: An expository note. *The American Statistician*, **36**, 153–157.
- Collet, D. (2002). *Modelling binary data*. Chapman and Hall, London, second edition.
- Cook, R. D.; Weisberg, S. (1982). *Residuals and influence in regression*. Chapman and Hall, London.
- Cordeiro, G. M. (1983). Improved likelihood ratio statistics for generalized linear models. *J. Roy. Statist. Soc. B*, **45**, 401–413.
- Cordeiro, G. M. (1986). *Modelos lineares generalizados*. VII SINAPE, UNICAMP.

- Cordeiro, G. M. (1987). On the corrections to the likelihood ratio statistics. *Biometrika*, **74**, 265–274.
- Cordeiro, G. M. (1993). Bartlett corrections and bias correction for two heteroscedastic regression models. *Communications in Statistics, Theory and Methods*, **22**, 169–188.
- Cordeiro, G. M. (1995). Performance of a bartlett-type modification for the deviance. *Journal of Statistical Computation and Simulation*, **51**, 385–403.
- Cordeiro, G. M. (1999). *Introdução à teoria assintótica*. 22º Colóquio Brasileiro de Matemática, IMPA.
- Cordeiro, G. M. (2004a). Corrected likelihood ratio tests in symmetric nonlinear regression models. *Journal of Statistical Computation and Simulation*, **74**, 609–620.
- Cordeiro, G. M. (2004b). On pearson’s residuals in generalized linear models. *Statistics and Probability Letters*, **66**, 213–219.
- Cordeiro, G. M. (2004c). Second-order covariance matrix of maximum likelihood estimates in generalized linear models. *Statistics and Probability Letters*, **66**, 153–160.
- Cordeiro, G. M.; Barroso, L. P. (2007). A third-order bias corrected estimate in generalized linear models. *Test*, **16**, 76–89.
- Cordeiro, G. M.; Cribari-Neto, F.; Aubin, E. Q.; Ferrari, S. L. P. (1995). Bartlett corrections for one-parameter exponential family models. *Journal of Statistical Computation and Simulation*, **53**, 211–231.
- Cordeiro, G. M.; McCullagh, P. (1991). Bias correction in generalized linear models. *J. Roy. Statist. Soc. B*, **53**, 629–643.
- Cox, D. R. (1970). *Analysis of binary data*. Chapman and Hall, London.
- Cox, D. R. (1972). Regression models and life tables (with discussion). *J. R. Statist. Soc. B*, **74**, 187–220.
- Cox, D. R.; Hinkley, D. V. (1986). *Theoretical Statistics*. University Press, Cambridge.

- Cox, D. R.; Snell, E. J. (1968). A general definition of residual (with discussion). *J. R. Statist. Soc. B*, **30**, 248–275.
- Crawley, M. J. (2007). *The R Book*. John Wiley, Sussex.
- Demétrio, C. G. B. (2001). *Modelos Lineares Generalizados em Experimentação Agronômica*. ESALQ/USP, Piracicaba.
- Demétrio, C. G. B.; Hinde, J.; Moral, R. A. (2014). Models for overdispersed data in entomology. In Ferreira, C. P.; Godoy, W. A. C., editors, *Ecological models applied to entomology*, pages 219–259. Springer.
- Dey, D. K.; Gelfand, A. E.; Peng, F. (1997). Overdispersion generalized linear models. *Journal of Statistical Planning and Inference*, **68**, 93–107.
- Dobson, A. J.; Barnett, A. G. (2008). *An Introduction to Generalized Linear Models*. Chapman & Hall/CRC, London, third edition.
- Dyke, G. V.; Patterson, H. D. (1952). Analysis of factorial arrangements when the data are proportions. *Biometrics*, **8**, 1–12.
- Fahrmeir, L.; Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics*, **13**, 342–368.
- Fahrmeir, L.; Tutz, G. (1994). *Multivariate Statistical Modelling based on Generalized Linear Models*. Springer-Verlag, New York.
- Feigl, P.; Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics*, **21**, 826–838.
- Finney, D. J. (1971). *Probit Analysis*. Cambridge University Press, London, third edition.
- Firth, D. (1991). Generalized linear models. In Hinkley, D.; Reid, N.; Snell, E., editors, *Statistical Theory and Modelling*, pages 55–82. Chapman & Hall.
- Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society*, **222**, 309–368.

- Fisher, R. A. (1925). *Statistical methods for research workers*. Oliver and Boyd, Edinburgh.
- Fisher, R. A. (1935). The case of zero survivors (appendix to bliss, c.i. (1935)). *Ann. Appl. Biol.*, **22**, 164–165.
- Fisher, R. A.; Yates, F. (1970). *Statistical Tables for Biological, Agricultural and Medical Research*. Oliver and Boyd, Edinburgh.
- Folks, J. L.; Chhikara, R. S. (1978). The inverse gaussian distribution and its statistical application, a review. *J. R. Statist. Soc. B*, **40**, 263–289.
- Francis, B.; Green, M.; Payne, C. (1993). *The GLIM system generalized linear interactive modelling*. Oxford University Press, Oxford.
- Gasser, M. (1967). Exponential survival with covariance. *Journal of the American Statistical Association*, **62**, 561–568.
- Gelfand, A. E.; Dalal, S. R. (1990). A note on overdispersed exponential families. *Biometrika*, **77**, 55–64.
- Haberman, S. (1970). *The general log-linear model. PhD dissertation*. Univ. of Chicago Press, Chicago, Illinois.
- Haberman, S. (1974). *The analysis of frequency data*. Univ. of Chicago Press, Chicago, Illinois.
- Haberman, S. (1978). *Analysis of quantitative data*, volume 1. Academic Press, New York.
- Hardin, J. W.; Hilbe, J. M. (2007). *Generalized Linear Models and Extensions*. Stata, Texas, 2 edition.
- Hinde, J.; Demétrio, C. G. B. (1998a). *Overdispersion: Models and Estimation*. XIII SINAPE, São Paulo.
- Hinde, J.; Demétrio, C. G. B. (1998b). Overdispersion: Models and estimation. *Computational Statistics and Data Analysis*, **27**, 151–170.

- Johnson, N. L.; Kotz, S.; Balakrishnan, N. (2004). *Continuous univariate distributions*. John Wiley & Sons, New York, second edition.
- Jørgensen, B. (1987). Exponential dispersion models (with discussion). *J. R. Statist. Soc. B*, **49**, 127–162.
- Jørgensen, B. (2013). Generalized linear models. *Encyclopedia of Environmetrics*.
- Judge, G. G.; Griffiths, W. E.; Hill, R. C.; Lütkepohl, H.; Lee, T.-C. (1985). *The theory and practice of Econometrics*. John Wiley & Sons, New York.
- Larsen, W. A.; McCleary, S. J. (1972). The use of partial residual plots in regression analysis. *Technometrics*, **14**, 781–790.
- Lee, Y.; Nelder, J. A.; Pawitan, Y. (2006). *Generalized Linear Models with Random Effects. Unified Analysis via H-likelihood*. Chapman & Hall/CRC, London.
- Mallows, C. L. (1966). *Choosing a subset regression*. Presented at Annual A.S.A. Meetings, Los Angeles.
- Mantel, N.; Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J. Nat. Cancer Inst.*, **22**, 719–748.
- Martin, J. T. (1942). The problem of the evaluation of rotenone-containing plants. vi: The toxicity of 1-elliptone and of poisons applied jointly, with further observations on the rotenone equivalent method of assessing the toxicity of derris root. *Annals of Applied Biology*, **29**, 69–81.
- McCullagh, P.; Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall, London, second edition.
- McCulloch, C. E.; Searle, S. R. (2000). *Generalized, Linear, and Mixed Models*. John Wiley & Sons, New York.
- Mendenhall, P.; Scheaffer, R. L.; Wackerly, D. D. (1981). *Mathematical Statistics with Applications*. Duxbury, Boston.

- Molenberghs, G.; Verbeke, G. (2005). *Models for discrete longitudinal data*. Springer-Verlag, New York.
- Moral, R. A.; Hinde, J.; Demétrio, C. G. B. (2015). *hnp: Half-Normal Plots with Simulation Envelopes*. R package version 1.1.
- Morgan, C. N. (1992). *Analysis of Quantal Response Data*. Chapman and Hall, London.
- Morris, C. N. (1982). Natural exponential families with quadratic variance functions: statistical theory. *Annals of Statistics*, **11**, 515–529.
- Myers, R. H.; Montgomery, D. C.; Vining, G. G. (2002). *Generalized Linear Models: With Applications in Engineering and the Sciences*. John Wiley, New York.
- Nelder, J. A. (1966). Inverse polynomials, a useful group of multifactor response functions. *Biometrics*, **22**, 128–141.
- Nelder, J. A.; Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, A*, **135**, 370–384.
- Paula, G. A. (2004). *Modelos de Regressão com Apoio Computacional*. IME/USP, São Paulo.
- Paulino, C. D.; Singer, J. M. (2006). *Análise de dados categorizados*. Editora Edgard Blücher, São Paulo.
- Phelps, K. (1982). Use of the complementary log-log function to describe dose response relationship in insecticide evaluation field trials. In *GLIM 82: Proceedings of the International Conference on Generalized Linear Models. Lecture notes in Statistics*, volume 14, pages 155–163. Springer-Verlag, New York.
- Pierce, D. A.; Schafer, D. W. (1986). Residual in generalized linear models. *Journal of the American Statistical Association*, **81**, 977–986.
- Pregibon, D. (1979). *Data analytic methods for generalized linear models. PhD Thesis*. University of Toronto, Toronto.

- Pregibon, D. (1980). Goodness of link tests for generalized linear models. *Appl. Statist.*, **29**, 15–24.
- Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics*, **9**, 705–724.
- Rao, C. R. (1973). *Linear statistical inference and its applications*. John Wiley, New York.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Danmarks Paedagogiske Institut, Copenhagen.
- Ridout, M. S. (1990). *Using Generalized Linear Models to Analyze Data from Agricultural, and Horticultural Experiments*. Departamento de Matemática e Estatística da ESALQ/USP, Piracicaba (não publicado).
- Ridout, M. S.; Demétrio, C. G. B. (1992). Generalized linear models for positive count data. *Revista de Matemática e Estatística*, **10**, 139–148.
- Ridout, M. S.; Demétrio, C. G. B.; Hinde, J. (1998). Models for count data with many zeros. *Proceedings of XIXth International Biometrics Conference, Cape Town, Invited Papers*, pages . 179–192.
- Ridout, M. S.; Fenlon, J. (1998). *Statistics in Microbiology*. Horticultural Station, East Malling (Notes for workshop).
- Ridout, M. S.; Hinde, J.; Demétrio, C. G. B. (2001). A score test for testing a zero-inflated poisson regression model against zero-inflated negative binomial alternatives. *Biometrics*, **57**, 219–223.
- Ryan, B. F.; Joiner, B. L.; Ryan Jr., T. A. (1976). *Minitab Student Handbook*. Duxbury Press, New York.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.
- Searle, S. (1982). *Linear models*. John Wiley, New York.
- Silveira Neto, S.; Nakano, O.; Barbin, D.; Villa Nova, N. (1976). *Manual de Ecologia dos Insetos*. Ed. Agronômica 'Ceres', São Paulo.



- Silvey, S. (1975). *Statistical Inference*. Chapman and Hall', London, second edition.
- Smyth, G. (1989). Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society B*, **51**, 47–60.
- Team, R. C. (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Theil, H. (1965). The analysis of disturbances in regression analysis. *Journal of the American Statistical Association*, **60**, 1067–1079.
- Tukey, J. (1949). One degree of freedom for non-additivity. *Biometrics*, **5**, 232–242.
- Vieira, A.; Hinde, J.; Demétrio, C. (2000). Zero-inflated proportion data models applied to a biological control assay. *Journal of Applied Statistics*, **27**, 373–389.
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.*, **54**, 426–482.
- Wang, P. (1985). Adding a variable in generalized linear models. *Technometrics*, **27**, 273–276.
- Wang, P. (1987). Residual plots for detecting nonlinearity in generalized linear models. *Technometrics*, **29**, 435–438.
- Weisberg, S. (2005). *Applied linear regression*. John Wiley, New York, third edition.
- Wilks, S. (1937). The large sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.*, **9**, 60–62.
- Williams, D. (1987). Generalized linear model diagnostics using the deviance and single-case deletions. *Applied Statistics*, **36**, 181–191.
- Zippin, C.; Armitage, P. (1966). Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter. *Biometrics*, **22**, 665–672.