

# **Modelos Lineares Generalizados em Experimentação Agronômica**

**Clarice Garcia Borges Demétrio**



ESALQ/USP – Piracicaba, SP  
Maio/2002

## **PREFÁCIO**

Estas notas são baseadas em vários livros textos citados, duas versões anteriores e têm como objetivo apresentar noções introdutórias de Modelos Lineares Generalizados e algumas aplicações na área de Estatística Aplicada à Experimentação Agronômica.

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 todos e, em especial aos professores: Silvano Cesar da Costa (UEL) e Suely Ruiz Giolo (UFPr) que foram incansáveis e amigos, ajudando-nos através de leitura cuidadosa, sugestões, correção e formatação do texto, elaboração dos programas em SAS e discussões muito proveitosas. Sem o auxílio deles não teríamos terminado de escrever estas notas em tempo hábil.

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

Clarice Garcia Borges Demétrio  
Piracicaba, 1<sup>o</sup> de julho de 2001

## SUMÁRIO

<b>1</b>	<b>Família Exponencial com um parâmetro</b>	<b>1</b>
1.1	Introdução .....	1
1.2	Definição .....	1
1.3	Funções geradoras de momentos e de cumulantes .....	4
1.4	Estatística Suficiente .....	8
1.5	Exercícios .....	9
<b>2</b>	<b>Modelo Linear Generalizado</b>	<b>14</b>
2.1	Introdução .....	14
2.2	Exemplos de motivação .....	16
2.3	Definição .....	25
2.4	Estimação do vetor de parâmetros $\beta$ .....	28
2.5	Um exemplo do algoritmo de estimação .....	33
2.6	Propriedades e distribuição amostral de $\hat{\beta}$ .....	39
2.7	Função <i>deviance</i> e estatística de Pearson $X^2$ generalizada .....	42
2.8	Estimação do parâmetro $\phi$ .....	48
2.9	Análise de <i>deviance</i> e seleção de modelos .....	50
2.10	Testes de hipóteses .....	53
	2.10.1 Teste da razão de verossimilhanças .....	54
	2.10.2 Teste de Wald .....	54
	2.10.3 Teste Escore .....	54
2.11	Intervalos de Confiança .....	57
2.12	Exercícios .....	57
<b>3</b>	<b>Técnicas para Verificação de Ajuste do Modelo</b>	<b>61</b>
3.1	Introdução .....	61
3.2	Técnicas para a verificação do ajuste de um modelo a um conjunto de dados ...	62
3.3	Análise de resíduos e diagnósticos para modelos lineares clássicos .....	62
	3.3.1 Introdução .....	62
	3.3.2 Tipos de resíduos .....	63
	3.3.3 Estatísticas para diagnósticos .....	65
	3.3.4 Tipos de gráficos .....	66
3.4	Análise de resíduos e diagnósticos para modelos lineares generalizados .....	70
	3.4.1 Introdução .....	70
	3.4.2 Tipos de resíduos .....	71
	3.4.3 Tipos de gráficos .....	72
3.5	Verificação da função de ligação .....	73
	3.5.1 Justificativa para o uso de $\hat{\eta}^2$ .....	74
3.6	Verificação da adequação da função de variância .....	75
3.7	Verificação da adequação das escalas das covariáveis .....	75
3.8	Exercícios .....	79

<b>4 Aplicações</b>	82
4.1 Estimação da dose efetiva .....	82
4.2 Paralelismo entre retas no modelo logístico linear .....	83
4.3 Outras aplicações .....	93
<b>Apêndice A</b> – Programas em GLIM para os exemplos do Capítulo 4 .....	100
<b>Apêndice B</b> - Programas em SAS para alguns exemplos do Capítulo 4 .....	105
<b>Referências Bibliográficas</b> .....	109
<b>Lista Adicional de GLM</b> .....	114

# Capítulo 1

## Família Exponencial com um Parâmetro

### 1.1 Introdução

Muitas das distribuições conhecidas podem ser reunidas na chamada família exponencial univariada. Assim, por exemplo, pertencem a essa família as distribuições normal, binomial, binomial negativa, gama, Poisson e normal inversa. Essa classe de famílias de distribuições foi proposta independentemente por Koopman, Pitman e Darmais através do estudo de propriedades de suficiência estatística. Posteriormente, muitos outros aspectos dessa família foram descobertos e tornaram-se importantes na teoria moderna de Estatística. O conceito de família exponencial foi introduzido na Estatística por Fisher (Jørgensen & Labouriau, 1992).

### 1.2 Definição

Seja  $X$  uma variável aleatória (v.a.) cuja f.d.p. (função de probabilidade se  $X$  é discreta ou função densidade de probabilidade se  $X$  é contínua) depende de um único parâmetro  $\tau$ .

Seja a família  $\mathfrak{F} = \{f(x;\tau), \tau \in \Omega \subseteq \mathfrak{R}\}$  de f.d.p.'s. Diz-se que ela é a família exponencial de distribuições com parâmetro  $\tau$  se

$$f(x;\tau) = h(x) t(\tau) e^{q(x)s(\tau)} I_A(x), \quad (1.1)$$

sendo  $h(\cdot)$ ,  $t(\cdot)$ ,  $q(\cdot)$  e  $s(\cdot)$  funções conhecidas e  $I_A(\cdot)$  o indicador do conjunto  $A$  que não pode depender de  $\tau$ . Outra forma com que se apresenta essa família é

$$f(x;\tau) = \exp[q(x)s(\tau) + \ln t(\tau) + \ln h(x)]$$

e, tomando-se  $d(\tau) = \ln t(\tau)$  e  $g(x) = \ln h(x)$ , tem-se:

$$f(x;\tau) = \exp[q(x)s(\tau) + d(\tau) + g(x)] I_A(x). \quad (1.2)$$

Além disso, transformações do tipo 1-1 de variáveis ou de parâmetros não afetam a forma geral de (1.2), isto é, se a distribuição pertence, ou não, à família exponencial (Cox & Hinkley, 1986, pág. 27). Então, desde que  $q(\cdot)$  e  $s(\cdot)$  sejam funções monotônicas, pode-se fazer

$$s(\tau) = \frac{\theta}{a(\phi)} \quad \text{e} \quad q(X) = Y,$$

sendo  $\phi > 0$ , conhecido e fixo. Tem-se, então,

$$f(y; \theta) = \exp\left\{\frac{1}{a(\phi)}[y\theta + d_1(\theta)] + g_1(y; \phi)\right\}$$

em que  $d_1(\cdot)$  e  $g_1(\cdot)$  são funções conhecidas.

Na notação de McCullagh & Nelder (1989), fica

$$f(y; \theta, \phi) = \exp\left\{\frac{1}{a(\phi)}[y\theta - b(\theta)] + c(y; \phi)\right\} I_A(y) \quad (1.3)$$

para  $b(\cdot)$  e  $c(\cdot)$  funções conhecidas e  $\phi > 0$ , suposto conhecido.

Se  $\phi$  for desconhecido  $f(y; \theta, \phi)$  pode pertencer, ou não, à família exponencial com dois parâmetros (ver exercício 1.5.4 e McCullagh & Nelder, 1989, pág. 28).

Fazendo-se ainda  $\phi' = \frac{1}{a(\phi)}$  tem-se a notação usada por Cordeiro (1986), isto é,

$$f(y; \theta, \phi') = \exp\{\phi' [y\theta - b(\theta) + c_1(y; \phi')]\} I_A(y), \quad (1.4)$$

sendo  $c_1(\cdot)$  conhecida.

Tem-se, nestes dois últimos casos, a família exponencial na forma canônica com parâmetro canônico ou natural  $\theta$ . Se há outros parâmetros além de  $\theta$ , eles são olhados como parâmetros de perturbação (*nuisance parameters*). A Tabela 2 mostra os identificadores na família exponencial para as distribuições normal, binomial, Poisson, binomial negativa, gama e normal inversa.

**Exemplo 1:** Seja  $Y$  uma variável aleatória com distribuição normal de média  $\mu$  desconhecida e variância conhecida  $\sigma^2 > 0$ , isto é, com f.d.p.,

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

Desenvolvendo-se essa função de forma a poder ser comparada com a expressão da família exponencial dada por McCullagh & Nelder (1989), tem-se

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

que comparada com (1.3) leva a

$$\begin{aligned}
\theta &= \mu & a(\phi) &= \sigma^2 \\
b(\theta) &= \frac{\mu^2}{2} = \frac{\theta^2}{2} & c(y; \phi) &= -\frac{1}{2} \left[ \frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2) \right]
\end{aligned}$$

o que mostra que a distribuição  $N(\mu, \sigma^2)$  com  $\mu$  desconhecido e  $\sigma^2 > 0$ , conhecido, pertence à família exponencial na forma canônica.

**Exemplo 2:** Seja  $Y$  uma v.a. com distribuição binomial com f.d.p. dada por:

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

Tem-se, então,

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

obtendo-se

$$\begin{aligned}
a(\phi) &= 1 & \theta &= \ln \left(\frac{\pi}{1 - \pi}\right) \rightarrow \pi = \frac{e^\theta}{1 + e^\theta} \\
b(\theta) &= -m \ln(1 - \pi) = m \ln(1 + e^\theta) & c(y; \phi) &= \ln \binom{m}{y}
\end{aligned}$$

e portanto, a distribuição binomial pertence à família exponencial na forma (1.3).

### 1.3 Funções Geradoras de Momentos e de Cumulantes

A função geradora de momentos (f.g.m.) para a família exponencial com um parâmetro, usando-se a notação de McCullagh & Nelder (1989), é dada por:

$$M_Y(t; \theta, \phi) = E[e^{tY}] = \exp\left\{\frac{1}{a(\phi)}\{b[a(\phi)t + \theta] - b(\theta)\}\right\}. \quad (1.5)$$

**Prova:** A prova será feita apenas para o caso de v.a. contínuas. Lembrando-se que:

$$\int_A f(y) dy = 1$$

então,

$$\int_A \exp\left\{\frac{1}{a(\phi)}[\theta y - b(\theta)] + c(y; \phi)\right\} dy = 1$$

ou ainda,

$$\frac{1}{\exp\left(\frac{b(\theta)}{a(\phi)}\right)} \int_A \exp\left[\frac{1}{a(\phi)}\theta y + c(y; \phi)\right] dy = 1,$$

obtendo-se

$$\int_A \exp\left[\frac{1}{a(\phi)}\theta y + c(y; \phi)\right] dy = \exp\left(\frac{b(\theta)}{a(\phi)}\right). \quad (1.6)$$

Logo,

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

e usando-se (1.6) tem-se:

$$M_Y(t; \theta, \phi) = \frac{1}{\exp\left\{\frac{b(\theta)}{a(\phi)}\right\}} \exp\left\{\frac{b[a(\phi)t + \theta]}{a(\phi)}\right\}$$

ou ainda,

$$M_Y(t; \theta, \phi) = \exp \left\{ \frac{1}{a(\phi)} [b(a(\phi)t + \theta) - b(\theta)] \right\}.$$

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

$$\varphi(t; \theta, \phi) = \ln M(t; \theta, \phi) = \frac{1}{a(\phi)} \{b[a(\phi)t + \theta] - b(\theta)\}. \quad (1.7)$$

Derivando-se (1.7), sucessivamente, em relação a  $t$ , tem-se

$$\varphi'(t; \theta, \phi) = \frac{1}{a(\phi)} b'[a(\phi)t + \theta] a(\phi) = b'[a(\phi)t + \theta]$$

$$\varphi''(t; \theta, \phi) = b''[a(\phi)t + \theta] a(\phi)$$

$$\varphi'''(t; \theta, \phi) = b'''[a(\phi)t + \theta] [a(\phi)]^2$$

...

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

e para  $t = 0$ , obtêm-se os cumulantes

$$\kappa_1 = b'(\theta)$$

$$\kappa_2 = a(\phi) b''(\theta)$$

...

$$\kappa_r = [a(\phi)]^{r-1} b^{(r)}(\theta).$$

Verifica-se, portanto, que existe uma relação de recorrência entre os cumulantes da família exponencial. Isto é fundamental na obtenção de propriedades assintóticas dos Modelos Lineares Generalizados.

Os momentos da família exponencial podem ser facilmente obtidos a partir dos cumulantes (Kendall & Stuart, 1969, vol. 1, cap. 3, pág. 70-71).

#### i) Relação entre cumulantes e momentos em relação à origem

$$\kappa_1 = \mu'_1 = \mu$$

$$\kappa_2 = \mu'_2 - (\mu'_1)^2$$

$$\kappa_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2(\mu'_1)^3$$

$$\kappa_4 = \mu'_4 + 4\mu'_3\mu'_1 - 3(\mu'_2)^2 + 12\mu'_2(\mu'_1)^2 - 6(\mu'_1)^4,$$

sendo  $\mu'_r = E(Y^r)$ .

ii) **Relação entre cumulantes e momentos em relação à média**

$$\kappa_2 = \mu_2 = \sigma^2$$

$$\kappa_3 = \mu_3$$

$$\kappa_4 = \mu_4 - 3(\mu_2)^2,$$

sendo  $\mu_r = E\{[Y - E(Y)]^r\}$ .

Portanto, a média e a variância de uma v.a.  $Y$  cuja distribuição pertence à família exponencial, na forma canônica usada por McCullagh & Nelder (1989), são dadas por

$\begin{aligned} \mu &= E(Y) = b'(\theta) \\ \sigma^2 &= \text{Var}(Y) = a(\phi) b''(\theta) \end{aligned}$	<b>IMPORTANT!</b>	(1.8)
---	-------------------	-------

em que  $a(\phi)$ , em geral, pode ser escrita na forma  $a(\phi) = \phi / w$ , sendo  $\phi$  chamado parâmetro de dispersão e  $w$ , peso *a priori*. Além disso,  $b''(\theta) = d\mu/d\theta$  é uma função de  $\mu$  e é representada por  $V(\mu)$ . Logo, a variância de  $Y$  pode ser escrita como

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

Escrevendo-se  $\ell(\theta, \phi; y) = \ln f_Y(y; \theta, \phi)$  como o logaritmo da função de verossimilhança considerado como uma função de  $\theta$  e de  $\phi$ , dado  $y$ , a média e a variância da v.a.  $Y$  podem também ser obtidas facilmente a partir das relações conhecidas (Dobson, 1990, Apêndice 1)

$$E(U) = E\left[\frac{d\ell}{d\theta}\right] = 0$$

e

$$\text{Var}(U) = E(U^2) = E(-U') = E\left[-\frac{d^2\ell}{d\theta^2}\right],$$

sendo  $U$  chamada função score.

Portanto, a partir de

$$\ell = \frac{1}{a(\phi)} [y\theta - b(\theta)] + c(y; \phi)$$

tem-se

$$U = \frac{d\ell}{d\theta} = \frac{1}{a(\phi)} [y - b'(\theta)]$$

e

$$U' = \frac{d^2 \ell}{d\theta^2} = \frac{-1}{a(\phi)} b''(\theta).$$

Logo,

$$E(U) = \frac{1}{a(\phi)} [E(Y) - b'(\theta)] = 0 \rightarrow E(Y) = b'(\theta)$$

e

$$\left. \begin{aligned} \text{Var}(U) = -E(U') &= \frac{1}{a(\phi)} b''(\theta) \\ \text{Var}(U) = E(U^2) &= \frac{1}{[a(\phi)]^2} \text{Var}(Y) \end{aligned} \right\} \rightarrow \text{Var}(Y) = a(\phi) b''(\theta).$$

**Exemplo 3:** Considere o **Exemplo 1**, e obtenha  $\varphi_Y(t)$ ,  $M(t)$ ,  $E(Y)$ ,  $\text{Var}(Y)$  e  $V(\mu)$ . Tem-se que:

$$a(\phi) = \sigma^2, \theta = \mu \text{ e } b(\theta) = \theta^2 / 2.$$

Logo, usando-se (1.5), (1.7) e (1.8), tem-se

$$\begin{aligned} \varphi(t) &= \frac{1}{\sigma^2} \left[ \frac{(\sigma^2 t + \theta)^2}{2} - \frac{\theta^2}{2} \right] = \frac{1}{2\sigma^2} (\sigma^4 t^2 + 2\sigma^2 t\theta + \theta^2 - \theta^2) \\ &= \frac{1}{2} (\sigma^2 t^2 + 2t\theta) = t\mu + \frac{\sigma^2 t^2}{2}, \end{aligned}$$

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

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

$$\text{Var}(Y) = a(\phi) b''(\theta) = \sigma^2 \rightarrow V(\mu) = 1 \text{ (constante).}$$

**Exemplo 4:** Considere o **Exemplo 2**, e obtenha  $\varphi_Y(t)$ ,  $M(t)$ ,  $E(Y)$ ,  $\text{Var}(Y)$  e  $V(\mu)$ . Tem-se que:

$$a(\phi) = 1, \theta = \ln\left(\frac{\pi}{1-\pi}\right) \rightarrow \pi = \frac{e^\theta}{1+e^\theta} \text{ e } b(\theta) = -m \ln(1-\pi) = m \ln(1+e^\theta).$$

Logo, usando-se (1.5), (1.7) e (1.8), tem-se:

$$\begin{aligned}\varphi(t) &= b(t+\theta) - b(\theta) = m[\ln(1+e^{t+\theta}) - \ln(1+e^\theta)] \\ &= \ln\left(\frac{1+e^{t+\theta}}{1+e^\theta}\right)^m = \ln(1-\pi + \pi e^t)^m = \ln[1-\pi + \pi e^t]^m,\end{aligned}$$

$$M(t) = e^{\varphi(t)} = [1-\pi + \pi e^t]^m,$$

$$E(Y) = \mu = b'(\theta) = m \frac{e^\theta}{1+e^\theta} = m\pi,$$

$$\text{Var}(Y) = a(\phi) b''(\theta) = m \frac{(1+e^\theta)e^\theta - e^\theta e^\theta}{(1+e^\theta)^2} = m \frac{e^\theta}{(1+e^\theta)^2} = m\pi(1-\pi) = \frac{1}{m} \mu(m-\mu)$$

e

$$V(\mu) = m\pi(1-\pi) = \frac{1}{m} \mu(m-\mu).$$

#### 1.4 Estatística Suficiente

Seja  $Y_1, Y_2, \dots, Y_n$  uma amostra aleatória (a.a.) de uma distribuição que pertence à família exponencial. A f.d.p. conjunta de  $Y_1, Y_2, \dots, Y_n$  é dada por:

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

Pelo teorema da fatoração de Neyman-Fisher (Silvey, 1975, pág. 27) tem-se que  $T = \sum_{i=1}^n Y_i$  é uma estatística suficiente para  $\theta$ , pois

$$f_Y(y; \theta, \phi) = g(t, \theta)h(y_1, y_2, \dots, y_n),$$

sendo que  $g(t, \theta)$  depende de  $\theta$  e dos  $y$ 's apenas através de  $t$  e  $h(y_1, y_2, \dots, y_n)$  independe de  $\theta$ .

Isto mostra, que sob amostragem aleatória, se uma densidade pertence à família exponencial com um parâmetro, então, existe uma estatística suficiente. Na realidade, usando-se o Teorema de Lehmann-Scheffé (Mendehall *et al.*, 1981, pág. 350) mostra-se que  $T = \sum_{i=1}^n Y_i$  é uma estatística suficiente minimal.

Tabela 1: Funções Geradoras de Momentos para Algumas Distribuições

Distribuição	Funções Geradoras de Momentos
Normal: $N(\mu, \sigma^2)$	$M_Y(t; \theta, \phi) = \exp\left\{t\mu + \frac{\sigma^2 t^2}{2}\right\}$
Poisson: $P(\mu)$	$M_Y(t; \theta, \phi) = \exp\{\mu (e^t - 1)\}$
Binomial: $B(m, \pi)$	$M_Y(t; \theta, \phi) = (1 - \pi + \pi e^t)^m$
Bin. Negativa: $\text{BinNeg}(\mu, k)$	$M_Y(t; \theta, \phi) = \left[1 + \frac{\mu}{k}(1 - e^t)\right]^{-k}, t < -\ln \frac{\mu}{\mu + k}$
Gama: $G(\mu, \nu)$	$M_Y(t; \theta, \phi) = \left(1 - \frac{t\mu}{\nu}\right)^{-\nu}, t < \frac{\nu}{\mu}$
Normal Inversa: $\text{IG}(\mu, \sigma^2)$	$M_Y(t; \theta, \phi) = \exp\left\{\frac{1}{\sigma^2} \left[\frac{1}{\mu} - \left(\frac{1}{\mu^2} - 2t\sigma^2\right)^{\frac{1}{2}}\right]\right\}, t < \frac{1}{2\sigma^2\mu^2}$

## 1.5 Exercícios

**1.5.1** Verifique se as distribuições que se seguem pertencem à família exponencial na forma canônica dada em (1.3). Obtenha  $\varphi_Y(t)$ ,  $M(t)$ ,  $E(Y)$ ,  $\text{Var}(Y)$  e  $V(\mu)$ .

a) Poisson:  $Y \sim P(\mu)$

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

b) Binomial Negativa ( $k$  fixo):  $Y \sim \text{BinNeg}(\mu, k)$ ,  $k > 0$ , conhecido

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

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

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

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

$$f(y; \mu, \sigma^2) = \left( \frac{1}{2\pi\sigma^2 y^3} \right)^{1/2} \exp\left\{ -\frac{(y-\mu)^2}{2\mu^2\sigma^2 y} \right\} I_A(y); A = \mathfrak{R}^+, \mu > 0.$$

**1.5.2** Seja  $X$  uma v.a. com distribuição  $G(v)$ , isto é, com f.d.p.

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

Dado que  $E(X) = v$ , mostre que usando-se a transformação  $Y = \frac{X}{v} \mu$ , obtém-se a f.d.p. usada no item (c) do Exercício 1.5.1.

**1.5.3** Seja  $Y$  uma v.a. com distribuição de Poisson truncada com parâmetro  $\lambda$ , isto é, com f.d.p. dada por:

$$f(y; \lambda) = \frac{e^{-\lambda} \lambda^y}{y!(1 - e^{-\lambda})} = \frac{\lambda^y}{y!(e^\lambda - 1)} I_{\{1, 2, \dots\}}(y), \lambda > 0.$$

Pede-se:

a) mostre que essa distribuição é um membro da família exponencial na forma canônica;

b) mostre que

$$\text{b.1) } E(Y) = \mu = \frac{\lambda}{1 - e^{-\lambda}} = \frac{\lambda e^\lambda}{e^\lambda - 1};$$

$$\text{b.2) } \text{Var}(Y) = \frac{\lambda}{1 - e^{-\lambda}} \left[ 1 - \frac{\lambda e^{-\lambda}}{1 - e^{-\lambda}} \right] = \mu(1 + \lambda - \mu);$$

c) mostre que a f.g.m. de  $Y$  é dada por

$$M_Y(t) = \frac{\exp\{\lambda e^t\} - 1}{e^\lambda - 1}.$$

**1.5.4** De acordo com Smyth (1989), uma distribuição pertence à família exponencial se sua f.d.p. puder ser colocada na forma

$$f(y; \theta, \phi) = \exp\left\{ \frac{w}{\phi} [y\theta - b(\theta)] + c(y; \phi) \right\}, \quad (1.9)$$

sendo  $b(\cdot)$  e  $c(\cdot)$  funções conhecidas, e  $\phi > 0$ , constante, chamado parâmetro de dispersão, conhecido e  $w$ , um peso *a priori*. Se a constante  $\phi$  é desconhecida, então, a expressão (1.9) define uma família exponencial com dois parâmetros apenas se

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

para  $g(\cdot)$ ,  $s(\cdot)$  e  $t(\cdot)$  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.

**1.5.5** Seja  $Y | P \sim \text{Bin}(m, P)$  e  $P \sim \text{Beta}(\alpha, \beta)$ , isto é,

$$f(y | p) = \binom{m}{y} p^y (1-p)^{(m-y)} \quad \text{e} \quad f(p) = \frac{p^{\alpha-1} (1-p)^{\beta-1}}{B(\alpha, \beta)} I_{(0,1)}(p),$$

sendo  $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$ . Pede-se:

a) mostre que, incondicionalmente,  $Y$  tem distribuição beta-binomial com f.d.p. dada por:

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

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

c) mostre que a distribuição beta-binomial não pertence à família exponencial canônica.

**1.5.6** Seja  $Y | Z = z \sim P(z)$ , isto é,

$$P(Y = y | Z = z) = \frac{z^y e^{-z}}{y!}, I_{\{0,1,2,\dots\}}(y)$$

Então, se:

a)  $Z \sim G(k, \lambda)$ , isto é, com f.d.p. dada por

$$f(z; k, \lambda) = \frac{\left(\frac{\lambda}{k}\right)^\lambda}{\Gamma(\lambda)} z^{\lambda-1} \exp\left\{-\frac{z\lambda}{k}\right\} I_{(0, \infty)}(z),$$

mostre que, incondicionalmente,  $Y$  tem distribuição binomial negativa, que pertence à família exponencial, com  $E(Y) = \frac{k}{\lambda} = \mu$  e  $\text{Var}(Y) = \mu + \frac{\mu^2}{k}$ ;

b)  $Z \sim G(r, \lambda)$ , isto é, com f.d.p. dada por

$$f(z; r, \lambda) = \frac{\lambda^r}{\Gamma(r)} z^{r-1} e^{-\lambda z} I_{(0, \infty)}(z)$$

mostre que, incondicionalmente,  $Y$  tem distribuição binomial negativa, que não pertence à família exponencial, com  $E(Y) = \frac{r}{\lambda} = \mu$  e  $\text{Var}(Y) = \mu + \frac{\mu}{\lambda} = \phi\mu$ , sendo  $\phi = 1 + \frac{1}{\lambda}$ .

**1.5.7** Uma forma geral para representar a f.d.p. da distribuição binomial negativa, segundo Saha & Dong (1997) é dada por:

$$\Pr(Y = y) = \frac{\Gamma\left(y + \frac{\mu^c}{\nu}\right)}{\Gamma\left(\frac{\mu^c}{\nu}\right)y!} \left(1 + \frac{\mu^{c-1}}{\nu}\right)^{-y} \left(1 + \nu\mu^{1-c}\right)^{-\frac{\mu^c}{\nu}} \quad y = 0, 1, 2, \dots \quad (1.10)$$

Pede-se:

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 da distribuição binomial negativa, isto é, para  $c = 0$  e  $c = 1$ ;

b) mostre que (1.10) pertence à família exponencial apenas para  $c = 0$ .

**1.5.8** Uma distribuição para explicar o excesso de zeros em dados de contagem é a distribuição de Poisson inflacionada de zeros, com f.d.p. dada por:

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

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

**1.5.9** Uma distribuição alternativa para explicar o excesso de zeros em dados de contagem é a distribuição binomial negativa inflacionada de zeros, com f.d.p. dada por:

$$\Pr(Y = y) = \begin{cases} \omega + (1 - \omega)(1 + \alpha\lambda^c)^{-\frac{\lambda^{1-c}}{\alpha}} & y = 0 \\ (1 - \omega) \frac{\Gamma\left(y + \frac{\lambda^{1-c}}{\alpha}\right)}{y! \Gamma\left(\frac{\lambda^{1-c}}{\alpha}\right)} (1 + \alpha\lambda^c)^{-\frac{\lambda^{1-c}}{\alpha}} \left(1 + \frac{\lambda^{1-c}}{\alpha}\right)^{-y} & y = 1, 2, 3, \dots \end{cases}$$

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

Tabela 2: Identificadores da Família Exponencial para Algumas Distribuições

Distribuição	$a(\phi)$	$\theta$	$b(\theta)$	$c(y;\phi)$	$\mu(\theta)$	$V(\mu)$
Normal $N(\mu, \sigma^2)$	$\sigma^2$	$\mu$	$\frac{\theta^2}{2}$	$-\frac{1}{2} \left[ \frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2) \right]$	$\theta$	1
Poisson $P(\mu)$	1	$\ln \mu$	$e^\theta$	$-\ln y!$	$e^\theta$	$\mu$
Binomial $B(m, \pi)$	1	$\ln \left( \frac{\pi}{1-\pi} \right)$	$m \ln(1 + e^\theta)$	$\ln \binom{m}{y}$	$m \frac{e^\theta}{1 + e^\theta}$	$\frac{1}{m} \mu(m - \mu)$
Bin. Negativa $\text{BinNeg}(\mu, k)$	1	$\ln \left( \frac{\mu}{\mu + k} \right)$	$-k \ln(1 - e^\theta)$	$\ln \left[ \frac{\Gamma(k + y)}{\Gamma(k) y!} \right]$	$k \frac{e^\theta}{1 - e^\theta}$	$\mu \left( \frac{\mu}{k} + 1 \right)$
Gama $G(\mu, \nu)$	$\nu^{-1}$	$-\frac{1}{\mu}$	$-\ln(-\theta)$	$\nu \ln(\nu y) - \ln y - \ln \Gamma(\nu)$	$-\frac{1}{\theta}$	$\mu^2$
Normal Inversa $\text{IG}(\mu, \sigma^2)$	$\sigma^2$	$-\frac{1}{2\mu^2}$	$-(-2\theta)^{1/2}$	$-\frac{1}{2} \left[ \ln(2\pi\sigma^2 y^3) + \frac{1}{\sigma^2 y} \right]$	$(-2\theta)^{-1/2}$	$\mu^3$

## Capítulo 2

### Modelo Linear Generalizado

#### 2.1 Introdução

A seleção de modelos é uma parte importante de toda pesquisa, envolve a procura de um modelo o mais simples possível, razoável, que descreva bem os dados observados. Na maior parte das situações pode-se pensar na variável resposta consistindo de duas partes distintas:

1<sup>a</sup>) **um componente sistemático**, que é estabelecido durante o planejamento (fundamental para a obtenção de conclusões confiáveis) do experimento, resultando em modelos de regressão (linear simples, múltipla, não linear etc), de análise de variância (delineamentos inteiramente casualizados, blocos casualizados, quadrados latinos com estrutura de tratamentos fatorial, parcelas subdivididas etc) e de análise de covariância;

2<sup>a</sup>) **um componente aleatório**, que é estabelecido assim que são definidas as medidas a serem feitas, que podem ser contínuas ou discretas, exigindo o ajuste de distribuições diferentes. Um mesmo experimento pode envolver medidas de diferentes tipos, como por exemplo, dados de altura, número de lesões e proporção de plantas doentes.

No modelo linear clássico tem-se,

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

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, de dimensões  $n \times p$ , do modelo,

$\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$  o vetor dos parâmetros,

$\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$ , o componente aleatório com  $\varepsilon_i \sim N(0, \sigma^2)$ ,  $i = 1, \dots, n$ .

Em muitos casos, porém, essa estrutura aditiva entre o componente sistemático e o componente aleatório não é satisfeita. Além disso, não há razão para se restringir à estrutura simples dada por  $\boldsymbol{\mu} = E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$  para o componente sistemático e nem para se restringir à

distribuição normal para o componente aleatório e à suposição de homogeneidade de variâncias.

Nelder & Wedderburn (1972) propuseram uma teoria unificadora da modelagem estatística a que deram o nome de modelos lineares generalizados (MLG), como uma extensão dos modelos lineares clássicos. Na realidade, eles mostraram que uma série de técnicas comumente estudadas separadamente podem ser reunidas sob o nome de Modelos Lineares Generalizados.

Os desenvolvimentos que levaram a esta visão geral da modelagem estatística, remontam a mais de um século. Um breve histórico (McCullagh & Nelder, 1989; Lindsey, 1997) pode ser traçado:

- regressão linear múltipla, envolvendo distribuição normal (Legendre, Gauss, início do século XIX);
- análise de variância para experimentos planejados, envolvendo distribuição normal (Fisher, 1920 a 1935);
- função de verossimilhança, um procedimento geral para inferência a respeito de qualquer modelo estatístico (Fisher, 1922);
- modelo complemento log-log para ensaios de diluição, envolvendo distribuição binomial (Fisher, 1922);
- família exponencial, uma classe de distribuições com propriedades ótimas (estatísticas suficientes) para a estimação dos parâmetros (Fisher, 1934);
- modelo probit para proporções, envolvendo distribuição binomial (Bliss, 1935);
- modelo logístico para proporções, envolvendo distribuição binomial (Berkson, 1944; Dyke & Patterson, 1952);
- modelo logístico para análise de itens, envolvendo distribuição Bernoulli (Rasch, 1960);
- modelos log-lineares para contagens, envolvendo distribuição Poisson e multinomial (Birch, 1963);
- modelos de regressão para dados de sobrevivência, envolvendo distribuição exponencial (Feigl & Zelen, 1965; Zippin & Armitage, 1966; Gasser, 1967);
- polinômios inversos para ensaios de adubação, envolvendo distribuição gama (Nelder, 1966).

Nelder & Wedderburn (1972) mostraram, então, que a maioria dos problemas estatísticos, que surgem nas áreas de agricultura, demografia, ecologia, economia, geografia, geologia, história, medicina, ciência política, psicologia, sociologia, zootecnia etc, podem ser formulados, de uma maneira unificada, como modelos de regressão. Esses modelos envolvem uma variável resposta univariada, variáveis explicativas e uma amostra aleatória de  $n$  observações, sendo que:

- i) a variável resposta, componente aleatório do modelo, tem uma distribuição pertencente à família exponencial na forma canônica (distribuições normal, gama e normal inversa para dados contínuos; binomial para proporções; Poisson e binomial negativa para contagens);
- ii) as variáveis explicativas, entram na forma de um modelo linear (componente sistemático);
- iii) a ligação entre os componentes aleatório e sistemático é feita através de uma função (por exemplo, logarítmica para os modelos log-lineares).

Outras referências para o estudo de Modelos Lineares Generalizados e extensões são: Cordeiro (1986); Aitkin *et al.* (1989); Dobson (1990); Collett (1991); Firth (1991); Francis *et al.* (1993); Fahrmeir & Tutz (1994); Paula (2000) e McCulloch (2000).

## 2.2 Exemplos de Motivação

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

### a) Ensaios do tipo dose-resposta

Ensaios do tipo dose-resposta, são aqueles em que uma determinada droga é administrada em  $k$  diferentes doses,  $d_1, d_2, \dots, d_k$ , a, respectivamente,  $m_1, m_2, \dots, m_k$  indivíduos, obtendo-se como resposta, após um período especificado,  $y_1, y_2, \dots, y_k$  indivíduos que mudam de estado (ocorrência de um sucesso, por exemplo, morte). Suponha que cada indivíduo responde, ou não, à droga, tal que a resposta é quantal (tudo ou nada, isto é, 1 ou 0). Por exemplo, quando um inseticida é aplicado 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 (falha). 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,  $Y_i \sim Bin(m_i, \pi_i)$ .

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  $p\%$  dos indivíduos, por exemplo,  $DL_{50}$ ,  $DL_{90}$ ), comparar potência de diferentes produtos etc.

**Exemplo 5:** Os dados da Tabela 3 referem-se a um ensaio de toxicidade de rotenone, no delineamento completamente casualizado, em que doses ( $d_i$ ) do inseticida foram aplicadas a  $m_i$  insetos (*Macrosiphoniella sanborni*, pulgão do crisântemo, Martin, 1942) e após um determinado tempo foram observados os números ( $y_i$ ) de insetos mortos.

Tabela 3: Número de insetos mortos ( $y_i$ ) de ( $m_i$ ) insetos que receberam a dose  $d_i$  de rotenone

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

O interesse do pesquisador estava na determinação das doses letais que matam 50% ( $DL_{50}$ ) e 90% ( $DL_{90}$ ) dos insetos, para recomendação de aplicação do inseticida no campo.

Pode-se ver que o gráfico das proporções ( $p_i = y_i / m_i$ ) de insetos mortos versus as doses ( $d_i$ ) tem um aspecto sigmoidal (Figura 1) o que orienta a escolha do modelo para  $\pi_i$ .

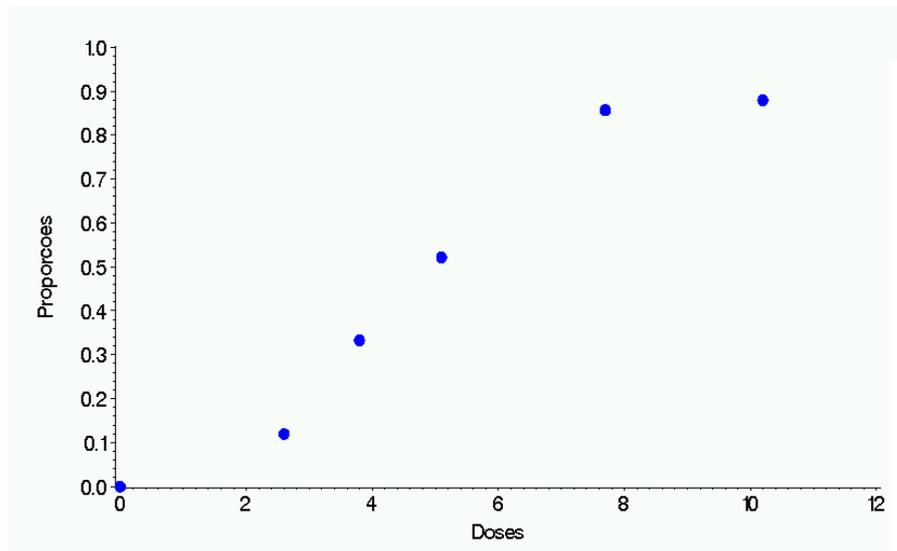


Figura 1: Gráfico das proporções ( $p_i$ ) versus as doses ( $d_i$ ), referentes à Tabela 3.

São dois componentes, portanto, a serem considerados nos ensaios de dose-resposta. Um é a intensidade do estímulo que pode ser a dose de uma droga (inseticida, fungicida, herbicida, medicamento etc) e o outro que é o indivíduo (um inseto, um esporo, uma planta, um paciente etc). O estímulo é aplicado a uma intensidade especificada em unidades de concentração, peso etc 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 este valor é chamado tolerância (Ashton, 1972). Essa tolerância varia de um indivíduo para outro da população e então, há uma distribuição de tolerâncias a qual pode-se associar uma variável aleatória  $U$  com f.d.p. representada por curvas, simétricas ou assimétricas, dos tipos apresentadas na Figura 2.

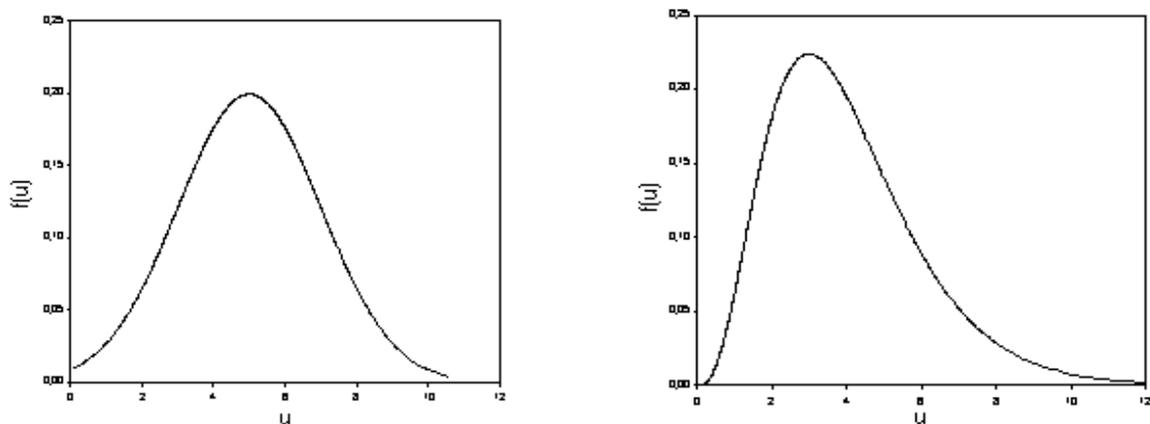


Figura 2: Distribuições de tolerâncias.

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

$$\pi_i = P[U \leq d_i] = F(d_i) = \int_{-\infty}^{d_i} f(u) du. \quad (2.1)$$

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  $d$ . Uma tal curva tem as propriedades matemáticas de uma função de distribuição contínua acumulada e tem as formas sigmoidais típicas (simétrica ou assimétrica) como mostradas na Figura 3.

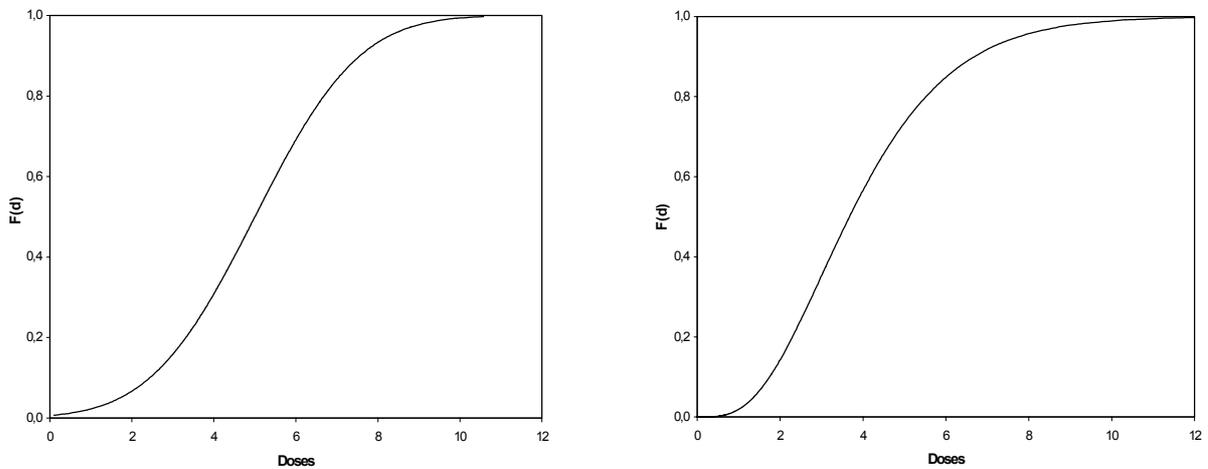


Figura 3: Curvas sigmóides típicas relacionando  $\pi_i$  e dose.

Observe-se que se assume que nenhum indivíduo responderá se a dose é muito pequena e que todos os indivíduos responderão se a dose é grande. Estas suposições nem sempre são razoáveis. Pode haver indivíduos que responderão naturalmente sem a droga (morte natural) e outros que são imunes à droga, o que pode causar um excesso de zeros e uma variabilidade maior do que a esperada (superdispersão).

O problema, então, está em se encontrar uma curva sigmóide que se ajuste bem aos dados e a partir dela obter  $DL_{50}$  e  $DL_{90}$ . O que ocorre, porém, é que são modelos não-lineares nos parâmetros e então, a idéia é se fazer uma transformação tal que essa curva sigmóide se transforme em uma reta e assim, procedimentos comuns de regressão possam ser usados para se estimarem os parâmetros. São apresentadas, a seguir, as curvas sigmóides mais comuns e suas respectivas transformações lineares.

#### i) Modelo *probit* (*Probability unit*)

Nesse caso assume-se que  $U$  tem distribuição normal de média  $\mu$  e variância  $\sigma^2$ , isto é,

$$f_U(u; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(u - \mu)^2}{2\sigma^2}\right], \mu \in \mathcal{R} \text{ e } \sigma^2 > 0,$$

e, portanto,

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

para  $\beta_1 = -\frac{\mu}{\sigma}$  e  $\beta_2 = \frac{1}{\sigma}$ . Portanto, como  $Z \sim N(0,1)$ , tem-se que:

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

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

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

### ii) Modelo logístico (*Logistic unit*)

Nesse caso, assume-se que U tem distribuição logística com parâmetros  $\mu$  e  $\tau$ , que é similar à distribuição normal na forma, com caudas um pouco mais longas e tem f.d.p. dada por:

$$f_U(u; \mu, \tau) = \frac{1}{\tau} \frac{\exp\left(\frac{u - \mu}{\tau}\right)}{\left[1 + \exp\left(\frac{u - \mu}{\tau}\right)\right]^2}, \mu \in \mathfrak{R}, \tau > 0,$$

com média  $E(U) = \mu$  e variância  $\sigma^2 = \text{Var}(U) = \frac{\pi^2 \tau^2}{3}$  (Mood *et al.*, 1974). Fazendo-se,

$\beta_1 = -\frac{\mu}{\tau}$  e  $\beta_2 = \frac{1}{\tau}$ , tem-se:

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

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 e é linearizada por:

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

### iii) Modelo complemento log-log

Nesse caso, assume-se que U tem distribuição 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. dada por:

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

com média  $E(U) = \alpha + \gamma\tau$  e variância  $\sigma^2 = \text{Var}(U) = \frac{\pi^2 \tau^2}{6}$ , sendo  $\gamma \approx 0,577216$  (Mood *et al.*, 1974). Fazendo-se,  $\beta_1 = -\frac{\alpha}{\tau}$  e  $\beta_2 = \frac{1}{\tau}$ , 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 e é linearizada por:

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

Vê-se, então, que esses três exemplos têm em comum

- 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, ou seja,

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

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

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

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

que nos casos vistos foram:

modelo probit:  $\eta_i = g(\pi_i) = \Phi^{-1}(\pi_i)$ ;

modelo logístico:  $\eta_i = g(\pi_i) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right)$ ;

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

Portanto, tem-se que esses modelos são baseados na família exponencial com um parâmetro desconhecido, cujas médias são não-lineares num conjunto de parâmetros lineares, isto é,

$$\text{modelo probit: } \mu_i = m_i \Phi(\eta_i) = m_i \Phi(\beta_1 + \beta_2 d_i);$$

$$\text{modelo logístico: } \mu_i = m_i \frac{e^{\eta_i}}{1 + e^{\eta_i}} = m_i \frac{e^{\beta_1 + \beta_2 d_i}}{1 + e^{\beta_1 + \beta_2 d_i}};$$

$$\text{modelo complemento log-log: } \mu_i = m_i \{1 - \exp\{-\exp(\eta_i)\}\} = m_i \{1 - \exp\{-\exp(\beta_1 + \beta_2 d_i)\}\}.$$

### b) Ensaio de diluição

É prática comum, o uso dos ensaios de diluição para se estimar a concentração de um organismo (número por unidade de volume, de área, de peso etc) em uma amostra quando contagem direta não é possível, mas a presença ou ausência do organismo em sub-amostras pode ser detectada (Ridout, 1998). Não são necessárias, portanto, as contagens dos organismos e, em geral, registrar a presença, ou ausência, fica mais econômico. 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 que o número de indivíduos presentes segue uma distribuição de Poisson, o que é uma suposição forte e é 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 siga a distribuição de Poisson.

Nos ensaios de diluição, a solução original é diluída progressivamente e na  $i$ -ésima diluição são feitas as contagens (**Exemplo 6**) ou, então, são testadas  $m_i$  sub-amostras das quais  $Y_i$  apresentam resultado positivo para a presença do organismo (**Exemplo 7**). Seja  $v_i$  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, tal que os  $v_i$ 's ficam em progressão geométrica.

**Exemplo 6:** A Tabela 4 mostra os dados referentes a contagens de partículas de vírus para 5 diluições diferentes, sendo que foram usadas 4 repetições para as 4 primeiras diluições e 5 repetições para a última diluição. O objetivo do experimento era estimar o número de partículas de vírus por unidade de volume.

Tabela 4: Números de partículas de vírus para 5 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)

**Exemplo 7:** A Tabela 5 mostra os dados de um ensaio de diluição realizado para determinar o número de esporos de *B. mesentericus* por g de farinha de batata (Fisher & 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 foram tomadas 5 amostras de 1ml e foi contado o número de amostras com esporos.

Tabela 5: Números de amostras ( $Y$ ) que contêm esporos em 5 amostras, para diferentes quantias (g) de farinha de batata em cada diluição.

g/100ml	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 ( $v_i$ ). 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  $\lambda v_i$ , isto é,

$$\mu_i = \lambda v_i.$$

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

$$\eta_i = \ln(\mu_i) = \ln(\lambda) + \ln(v_i) = \beta_1 + \text{offset}. \quad (2.2)$$

Quando se observa o número de amostras em que o indivíduo está presente tem-se  $Y_i \sim \text{Bin}(m_i, \pi_i)$ , 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 é dada por:

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

Logo,

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

Tem-se, em (2.2) e (2.3), que  $\beta_1 = \ln(\lambda) \Rightarrow \lambda = \exp(\beta_1)$  e  $\ln(v_i)$  entra como variável *offset*, e, portanto, com coeficiente conhecido que no caso é  $\beta_2 = 1$ . Além disso, para (2.2) tem-se uma função de ligação logarítmica para o modelo Poisson enquanto que para (2.3) tem-se a ligação complemento log-log para o modelo binomial.

Esse método de diluição em série é grandemente utilizado em muitas áreas da Biologia. Podem ser tratados de forma semelhante os problemas de estimação de:

- 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

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

- 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.

Nesse exemplo, vê-se, novamente, que:

- i) a distribuição dos  $Y_i$  (Poisson ou Binomial) é um membro da família exponencial, com  $E(Y_i) = \mu_i$  ou  $E(Y_i) = \mu_i = m_i \pi_i$ ;
- 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,  $\mathbf{x}_i = (1 \quad d_i)^T$ ,  $\boldsymbol{\beta} = (\beta_1 \quad \beta_2)^T$  e  $\eta_i$  o preditor linear;

- iii) a média  $\mu_i$  é funcionalmente **ligada** 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 vistos foram:

modelo log-linear:  $\eta_i = g(\mu_i) = \ln \mu_i$ ;

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

Portanto, tem-se que esses modelos são baseados na família exponencial com um parâmetro desconhecido, cujas médias são não-lineares num conjunto de parâmetros lineares, isto é,

modelo log-linear:  $\mu_i = e^{\eta_i} = e^{\beta_1 + \text{offset}}$ ;

modelo complemento log-log:  $\mu_i = m_i \{1 - \exp\{-\exp(\eta_i)\}\} = m_i \{1 - \exp\{-\exp(\beta_1 + \text{offset})\}\}$ ,

sendo  $\beta_2 = 1$  e  $\ln(v_i) = \text{offset}$ .

### c) Tabelas de contigência

Dados de contagens são oriundos da simples contagem de eventos (por exemplo, número de brotos por explante), ou então, da frequência de ocorrências em várias categorias e que dão origem às tabelas de contigência. Sejam os exemplos:

**Exemplo 8:** Os dados da Tabela 6 referem-se a coletas de insetos em armadilhas adesivas de duas cores, em que os indivíduos coletados de uma determinada espécie foram sexados, com o fim de se verificar a influência da cor da armadilha sobre a atração de machos e fêmeas dessa espécie.

Tabela 6: 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)

Vê-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 da independência, isto é, o sexo do inseto não afeta a escolha pela cor da armadilha.

**Exemplo 9:** Os dados da Tabela 7 referem-se a um ensaio de controle de brocas do fruto do tomateiro através de 4 tratamentos.

Tabela 7: Números de frutos de tomateiro 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)

Tem-se aqui, também, um caso em que o número total de frutos com broca é uma variável aleatória e, portanto, podendo 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.

A distribuição de Poisson desempenha na análise de dados categorizados o mesmo papel que o modelo normal ocupa 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 com erro Normal. Ele 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.

Os valores da variável dependente podem ser considerados como variáveis de Poisson com média  $\mu_i$ , e, portanto,

$$E(Y_i) = \mu_i.$$

Considerando-se uma tabela de contingência bidimensional e a hipótese de independência, se  $y_i$  ocorre nos níveis  $j$  ( $j = 1, \dots, J$ ) e  $k$  ( $k = 1, \dots, K$ ), respectivamente, dos dois fatores de classificação, então,

$$\mu_i = m \pi_j \pi_k'$$

sendo que  $m = \sum_{i=1}^n y_i$  e  $\pi_j$  e  $\pi_k'$  são as probabilidades marginais de uma observação pertencer às classes  $j$  e  $k$ , respectivamente.

De forma semelhante pode ser verificado que, em geral, para dados colocados em tabelas de contingência, as hipóteses mais comuns podem ser expressas como modelos multiplicativos para as frequências esperadas das caselas (McCullagh & Nelder, 1989; Agresti, 1990). Verifica-se, então, que na análise de dados categorizados, de uma forma geral, a média  $\mu_i$  é obtida como um produto de outras médias marginais. Isto sugere que uma transformação logarítmica do valor esperado lineariza essa parte do modelo (daí vem o nome de modelo log-linear). Considerando-se o exemplo dado tem-se:

$$\eta_i = \ln(\mu_i) = \ln m + \ln \pi_j + \ln \pi_k' = \mu + \alpha_j + \beta_k.$$

Novamente, tem-se:

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

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

sendo  $\mathbf{x}_i$  um vetor de variáveis *dummy*,  $\boldsymbol{\beta} = (\mu, \alpha_1, \dots, \alpha_K, \beta_1, \dots, \beta_K)^T$  e  $\eta_i$  o preditor linear;

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

$$\eta_i = g(\mu_i) = \ln(\mu_i).$$

Portanto, tem-se que esses modelos são baseados na família exponencial com um parâmetro desconhecido, cujas médias são não-lineares num conjunto de parâmetros lineares, ou seja,

$$\mu_i = \exp\{\eta_i\} = \exp\{\mathbf{x}_i^T \boldsymbol{\beta}\}.$$

### 2.3 Definição

Os modelos lineares generalizados podem ser usados quando se tem uma única variável aleatória  $Y$  e associado a ela um conjunto de variáveis explicativas  $X_1, X_2, \dots, X_p$ . Para uma amostra de  $n$  observações  $(y_i, \mathbf{x}_i)$  em que  $\mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{pi})^T$  é o vetor coluna de variáveis explicativas, o modelo linear generalizado envolve os três componentes:

- i) **Componente aleatório**: representado por um conjunto de variáveis aleatórias independentes  $Y_1, Y_2, \dots, Y_k$  provenientes de uma mesma distribuição que faz parte da família exponencial na forma canônica com médias  $\mu_1, \mu_2, \dots, \mu_k$ , ou seja,

$$E(Y_i) = \mu_i, \quad i=1, 2, \dots, n,$$

um parâmetro constante de escala, conhecido,  $\phi > 0$  e que depende de um único parâmetro  $\theta_i$ , chamado parâmetro canônico ou natural. A f.d.p. de  $Y_i$  é dada por

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

sendo  $b(\cdot)$  e  $c(\cdot)$  funções conhecidas. Em geral,  $a_i(\phi) = \frac{\phi}{w_i}$ , sendo  $w_i$  pesos *a priori*. Além disso, de (1.8)

$$E(Y_i) = \mu_i = b'(\theta_i)$$

e

$$\text{Var}(Y_i) = a_i(\phi) b''(\theta_i) = a_i(\phi) V(\mu_i) = a_i(\phi) V_i$$

em que  $V_i = \frac{d\mu_i}{d\theta_i}$  é chamada função de variância, e como depende unicamente da média tem-se que o parâmetro natural pode ser expresso como

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

para  $q(\mu_i)$  uma função conhecida de  $\mu_i$ .

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

$$\eta_i = \sum_{j=1}^p x_{ij} \beta_j = \mathbf{x}_i^T \boldsymbol{\beta} \quad \text{ou} \quad \boldsymbol{\eta} = \mathbf{X} \boldsymbol{\beta},$$

sendo  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^T$  a matriz do modelo,  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$  o vetor de parâmetros e  $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)^T$  o preditor linear. Se um parâmetro tem valor conhecido, o termo correspondente na estrutura linear é chamado *offset*, como visto nos ensaios de diluição.

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

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

sendo  $g(\cdot)$  uma função monótona, derivável.

Assim, vê-se que para a especificação do modelo, os parâmetros  $\theta_i$  da família exponencial não são de interesse direto (pois há um para cada observação) mas sim um conjunto menor de parâmetros  $\beta_1, \beta_2, \dots, \beta_p$  tais que uma combinação linear dos  $\beta$ 's seja igual a alguma função do valor esperado de  $Y_i$ .

Portanto, uma decisão importante na escolha do modelo linear generalizado é a escolha do trinômio:

- i) distribuição da variável resposta;
- ii) matriz do modelo e
- iii) função de ligação.

Se a função de ligação é escolhida de tal forma que  $g(\mu_i) = \theta_i$ , o preditor linear modela diretamente o parâmetro canônico e tal função de ligação é chamada ligação canônica. Isto resulta, freqüentemente, 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 os parâmetros  $\beta$ 's e alguma simplificação no algoritmo de estimação. A estatística suficiente é  $\mathbf{T} = \mathbf{X}^T \mathbf{Y}$ , com componentes  $t_j = \sum_{i=1}^n x_{ij} Y_i$ ,  $j = 1, 2, \dots, p$ . As funções de ligação canônicas para as distribuições estudadas estão apresentadas na Tabela 8.

Tabela 8: Funções de ligação canônicas

Distribuição	Ligação canônica
Normal	Identidade: $\eta = \mu$
Poisson	Logarítmica: $\eta = \ln(\mu)$
Binomial	Logística: $\eta = \ln\left(\frac{\pi}{1-\pi}\right) = \ln\left(\frac{\mu}{m-\mu}\right)$
Gama	Recíproca: $\eta = \frac{1}{\mu}$
Normal Inversa	Recíproca <sup>2</sup> : $\eta = \frac{1}{\mu^2}$

Deve ser lembrado, porém, que embora as funções de ligação canônicas levem a propriedades estatísticas desejáveis para o modelo, principalmente no caso de amostras pequenas, não há nenhuma razão *a priori* para que os efeitos sistemáticos do modelo devam ser aditivos na escala dada por tais funções (McCullagh & Nelder, 1989).

Para o modelo linear clássico a função de ligação é chamada identidade, pois o preditor linear é igual à média. Essa função de ligação é adequada no sentido em que ambos,  $\eta$  e  $\mu$ , podem assumir valores na linha 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  $\hat{\eta}$  poderá assumir valores negativos dependendo dos valores obtidos para  $\hat{\beta}$ . Além disso, dados de contagem dispostos em tabelas de contingência, sob a suposição de independência, levam naturalmente a efeitos multiplicativos cuja linearização pode ser obtida através da ligação logarítmica, isto é,  $\eta = \ln \mu$  de onde se tem  $\mu = e^\eta$  (conforme visto em 2.2).

Para a distribuição binomial a restrição é que  $0 < \pi < 1$  e, portanto, uma função de ligação deve satisfazer à condição de transformar o intervalo  $(0,1)$  na linha dos reais. (ver Exercício 2.12.4). Nesse caso as funções de ligação mais comumente encontradas, além da canônica, são a Probit e a complemento log-log:

$$\eta = \Phi^{-1}(\pi) = \Phi^{-1}\left(\frac{\mu}{m}\right)$$

e

$$\eta = \ln[-\ln(1-\pi)] = \ln\left[-\ln\left(1 - \frac{\pi}{m}\right)\right].$$

Aranda-Ordaz (1981) propôs a família de funções de ligação dada por

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

sendo  $\lambda$  uma constante desconhecida e que tem como casos particulares o modelo logístico para  $\lambda = 1$  e o complemento log-log para  $\lambda \rightarrow 0$ .

Uma família importante de funções de ligação, principalmente para dados com média positiva, é a família potência (ver Exercício 2.12.6), especificada por:

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

ou então,

$$\eta = \begin{cases} \mu^\lambda & \lambda \neq 0 \\ \ln \mu & \lambda = 0 \end{cases}$$

sendo  $\lambda$  uma constante desconhecida.

## 2.4 Estimação do vetor de parâmetros $\beta$

O ajuste de um modelo linear generalizado é determinado pelo vetor  $\hat{\beta}$  de estimativas dos parâmetros. O método usado na estimação dos  $\beta$ 's é o método da máxima verossimilhança.

O logaritmo da função de verossimilhança para um conjunto de observações independentes  $y_1, y_2, \dots, y_n$  é dado pela soma das contribuições individuais, isto é,

$$\ell = \ell(\boldsymbol{\theta}; \mathbf{y}) = \sum_{i=1}^n \ell(\theta_i; y_i) = \sum_{i=1}^n \left\{ \frac{1}{a_i(\phi)} [y_i \theta_i - b(\theta_i)] + c(y_i; \phi) \right\},$$

sendo que  $E(Y_i) = \mu_i$ ,  $\eta_i = g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$  e  $\theta_i = \int V_i^{-1} d\mu_i = q(\mu_i)$ .

Uma propriedade da família exponencial de distribuições é que seus elementos satisfazem a condições de regularidade suficientes para assegurar que o máximo global do logaritmo da função de verossimilhança  $\ell(\boldsymbol{\theta}; \mathbf{y})$  é dado unicamente pela solução do sistema de equações  $\mathbf{U}_\theta = \frac{d\ell}{d\boldsymbol{\theta}} = \mathbf{0}$  ou equivalentemente,  $\mathbf{U}_\beta = \frac{d\ell}{d\boldsymbol{\beta}} = \mathbf{0}$  (Cox & Hinkley, 1986, capítulo 9). Tem-se, então, que a função escore é dada por

$$U_j = \sum_{i=1}^n \frac{\partial \ell(\theta_i; y_i, \phi)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial \ell_i}{\partial \beta_j}.$$

Mas,

$$\ell = f(\theta_1, \theta_2, \dots, \theta_i, \dots, \theta_n)$$

↓

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

↓

$$\mu_i = g^{-1}(\eta_i) = h(\eta_i)$$

↓

$$\eta_i = \sum_{j=1}^p x_{ij} \beta_j$$

e, pela regra da cadeia, tem-se

$$\begin{aligned} U_j &= \sum_{i=1}^n \frac{d\ell_i}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} \\ &= \sum_{i=1}^n \frac{1}{a_i(\phi)} [y_i - b'(\theta_i)] \cdot \frac{1}{\frac{d\mu_i}{d\theta_i}} \cdot \frac{d\mu_i}{d\eta_i} \cdot x_{ij}. \end{aligned}$$

Mas, como já foi visto em (1.8)

$$E(Y_i) = \mu_i = b'(\theta_i)$$

e

$$\text{Var}(Y_i) = a_i(\phi) \frac{d\mu_i}{d\theta_i} = a_i(\phi) V(\mu_i) \Rightarrow \frac{d\mu_i}{d\theta_i} = V(\mu_i).$$

Logo,

$$U_j = \frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^n \frac{1}{a_i(\phi)} (y_i - \mu_i) \frac{1}{V(\mu_i)} \frac{d\mu_i}{d\eta_i} x_{ij}. \quad (2.4)$$

Em geral, as equações  $U_j = 0, j = 1, 2, \dots, p$  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  $f(x)$  nas vizinhanças 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  $\mathbf{U}_\beta = \frac{d\ell}{d\beta} = \mathbf{0}$  e usando-se a versão multivariada do método de Newton-Raphson, tem-se:

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

sendo  $\beta^{(m)}$  e  $\beta^{(m+1)}$  os vetores de parâmetros estimados nos passos  $m$  e  $(m+1)$ ,  $\mathbf{U}^{(m)}$  o vetor escore, isto é, o vetor de derivadas parciais de 1ª ordem de  $f(x)$ , com elementos  $\frac{\partial \ell}{\partial \beta_j}$ , avaliado no passo  $m$  e  $(\mathbf{I}_o^{-1})^{(m)}$  a inversa da negativa da matriz de derivadas parciais de 2ª ordem de  $f(x)$ , com elementos  $\frac{-\partial^2 \ell}{\partial \beta_j \partial \beta_k}$ , avaliada no passo  $m$ .

Quando as derivadas de 2ª ordem são obtidas facilmente, o método de Newton-Raphson é bastante útil. Acontece, porém, que isso nem sempre ocorre e no caso dos modelos lineares generalizados 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). Ele envolve a substituição da matriz de derivadas parciais de 2ª ordem pela matriz de valores esperados das derivadas parciais, isto é, a substituição da matriz de informação observada,  $\mathbf{I}_o$ , pela matriz de informação esperada de Fisher,  $\mathfrak{I}$ . Logo,

$$\beta^{(m+1)} = \beta^{(m)} + (\mathfrak{I}^{-1})^{(m)} \mathbf{U}^{(m)} \quad (2.5)$$

sendo que  $\mathfrak{I}$  tem elementos dados por  $\mathfrak{I}_{jk} = E \left[ \frac{-\partial^2 \ell}{\partial \beta_j \partial \beta_k} \right] = E \left[ \frac{\partial \ell}{\partial \beta_j} \frac{\partial \ell}{\partial \beta_k} \right]$ , que é a matriz de covariâncias dos  $U_j$ 's.

Multiplicando-se ambos os lados de (2.5) por  $\mathfrak{I}^{(m)}$  tem-se

$$\mathfrak{I}^{(m)} \beta^{(m+1)} = \mathfrak{I}^{(m)} \beta^{(m)} + \mathbf{U}^{(m)}. \quad (2.6)$$

Assim, usando-se (2.4), obtém-se

$$\begin{aligned}\mathfrak{S}_{jk} &= E(U_j U_k) = \sum_{i=1}^n \frac{1}{[a_i(\phi)]^2} E(Y_i - \mu_i)^2 \frac{1}{[V(\mu_i)]^2} \left[ \frac{d\mu_i}{d\eta_i} \right]^2 x_{ij} x_{ik} \\ &= \sum_{i=1}^n \frac{1}{[a_i(\phi)]^2} a_i(\phi) V(\mu_i) \frac{1}{[V(\mu_i)]^2} \left( \frac{d\mu_i}{d\eta_i} \right)^2 x_{ij} x_{ik} \\ &= \sum_{i=1}^n \frac{1}{a_i(\phi)} \frac{1}{V(\mu_i)} \left( \frac{d\mu_i}{d\eta_i} \right)^2 x_{ij} x_{ik}\end{aligned}$$

e fazendo-se  $a_i(\phi) = \frac{\phi}{w_i}$ , com  $\phi > 0$ , constante,  $w_i$  peso *a priori* e  $W_i = \frac{w_i}{V(\mu_i)} \left( \frac{d\mu_i}{d\eta_i} \right)^2$ , tem-se:

$$\mathfrak{S} = \frac{1}{\phi} \mathbf{X}^T \mathbf{W} \mathbf{X}$$

com elementos  $\mathfrak{S}_{jk} = \sum_{i=1}^n \frac{1}{\phi} x_{ij} W_i x_{ik}$ ,  $\mathbf{X}$ , a matriz do modelo e  $\mathbf{W} = \text{diag}(W_1, W_2, \dots, W_n)$ .

No caso das funções de ligação canônicas  $W_i = w_i V(\mu_i)$ , pois  $\frac{d\theta_i}{d\mu_i} = \frac{d\eta_i}{d\mu_i} = V^{-1}(\mu_i)$ .

Além disso, rearranjando-se os termos de  $U_j$  tem-se

$$U_j = \sum_{i=1}^n \frac{w_i}{\phi} \frac{(y_i - \mu_i)}{V(\mu_i)} \frac{d\mu_i}{d\eta_i} x_{ij} = \sum_{i=1}^n \frac{1}{\phi} x_{ij} W_i \frac{d\eta_i}{d\mu_i} (y_i - \mu_i)$$

e, portanto, o vetor escore  $\mathbf{U}$  fica

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

com  $\Delta = \text{diag} \left\{ \frac{d\eta_1}{d\mu_1}, \frac{d\eta_2}{d\mu_2}, \dots, \frac{d\eta_n}{d\mu_n} \right\} = \text{diag} \{ g'(\mu_1), g'(\mu_2), \dots, g'(\mu_n) \}$ .

Logo, substituindo  $\mathfrak{S}$  e  $\mathbf{U}$  em (2.6) tem-se

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

ou, ainda,

$$\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m+1)} = \mathbf{X}^T \mathbf{W}^{(m)} \left[ \mathbf{X} \boldsymbol{\beta}^{(m)} + \Delta^{(m)} (\mathbf{y} - \boldsymbol{\mu})^{(m)} \right]$$

e fazendo-se  $\mathbf{z}^{(m)} = \mathbf{X} \boldsymbol{\beta}^{(m)} + \Delta^{(m)} (\mathbf{y} - \boldsymbol{\mu})^{(m)} = \boldsymbol{\eta}^{(m)} + \Delta^{(m)} (\mathbf{y} - \boldsymbol{\mu})^{(m)}$ , chamada variável dependente ajustada, tem-se

$$\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m+1)} = \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{z}^{(m)} \quad (2.7)$$

ou ainda,

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{z}^{(m)} \quad (2.8)$$

que tem a forma da solução das equações normais, para o modelo linear obtida pelo método dos quadrados mínimos ponderados, exceto que nesse caso a solução  $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(m+1)}$  é obtida por processo numérico iterativo. É importante observar que a expressão (2.8) independe de  $\phi$ .

O método usual para iniciar o processo iterativo é especificar uma estimativa inicial  $\boldsymbol{\beta}^{(0)}$  e sucessivamente alterá-la até que a convergência seja obtida e, portanto,  $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(m+1)}$ . Note, contudo que cada observação pode ser considerada como uma estimativa do seu valor médio, isto é,  $\hat{\mu}_i = y_i$  e, portanto,

$$\hat{\eta}_i = g(\hat{\mu}_i) = g(y_i).$$

Usando-se  $\hat{\eta}$  como a variável dependente e  $\mathbf{X}$ , a matriz do modelo, obtém-se o vetor  $\boldsymbol{\beta}^{(0)}$ . A seguir o algoritmo de estimação pode ser resumido nos seguintes passos:

1) obter as estimativas

$$\eta_i^{(m)} = \sum_{j=1}^p x_{ij} \beta_j^{(m)}$$

e

$$\mu_i^{(m)} = g^{-1}(\eta_i^{(m)});$$

2) obter a variável dependente ajustada

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

e os pesos

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

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 convergência, obtendo-se  $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(m+1)}$ .

Dentre os muitos existentes, um critério para verificar a convergência poderia ser :

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

tomando-se para  $\xi$  um valor suficientemente pequeno. Em geral, esse algoritmo é robusto e converge rapidamente (3 ou 4 iterações são suficientes).

**Observação:** Deve-se tomar cuidado se a função  $g(\cdot)$  não é definida para alguns valores  $y_i$ . Por exemplo, se a função de ligação for dada por

$$\eta = g(\mu) = \ell n \mu$$

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 a mais próxima possível de  $g(\mu)$ . Para o modelo de Poisson com função de ligação logarítmica usa-se  $c = \frac{1}{2}$ .

Para o modelo logístico usa-se  $c = \frac{1 - 2\pi}{2}$ , para  $\pi = \frac{\mu}{m}$  e  $m$ , o índice da distribuição binomial. De uma forma geral, usando-se a expansão de Taylor até 2ª ordem para  $g(y + c)$  em relação a  $g(\mu)$  tem-se:

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

com valor esperado dado por:

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

em que se tem:

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

## 2.5 Um exemplo do algoritmo de estimação

Considere os dados do **Exemplo 5**. A variável resposta tem distribuição binomial, isto é,  $Y_i \sim Bin(m_i, \pi_i)$  e como já foi visto, a f.d.p. de  $Y_i$  pode ser escrita como:

$$f(y_i; \pi_i) = \binom{m_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{(m_i - y_i)} I_A(y_i), \quad \pi_i \in [0, 1], \quad A = \{0, 1, \dots, m_i\}.$$

Logo,

$$\phi = 1, \quad w_i = 1, \quad a_i(\phi) = 1,$$

$$E(Y_i) = \mu_i = b'(\theta_i) = m_i \pi_i,$$

$$\text{Var}(Y_i) = m_i \pi_i (1 - \pi_i) = \frac{1}{m_i} \mu_i (m_i - \mu_i)$$

e

$$V(\mu_i) = m_i \pi_i (1 - \pi_i) = \frac{1}{m_i} \mu_i (m_i - \mu_i).$$

Adotando-se a função de ligação logística (canônica) e o preditor linear dado por uma regressão linear simples, isto é,

$$\eta_i = g\left(\frac{\mu_i}{m_i}\right) = \ln\left(\frac{\mu_i}{m_i - \mu_i}\right) = \beta_1 + \beta_2 d_i$$

tem-se

$$\mu_i = m_i g^{-1}(\eta_i) = m_i \frac{e^{\eta_i}}{1 + e^{\eta_i}},$$

$$\frac{d\eta_i}{d\mu_i} = \frac{(m_i - \mu_i) + \mu_i}{(m_i - \mu_i)^2} \frac{m_i - \mu_i}{\mu_i} = \frac{m_i}{\mu_i (m_i - \mu_i)},$$

$$\mathbf{X} = \begin{bmatrix} 1 & d_1 \\ 1 & d_2 \\ \dots & \dots \\ 1 & d_n \end{bmatrix} \quad \text{e} \quad \boldsymbol{\beta} = (\beta_1, \beta_2)^T.$$

Portanto,

$$z_i = \eta_i + (y_i - \mu_i) \frac{m_i}{\mu_i (m_i - \mu_i)}$$

e

$$W_i = V_i = \frac{1}{m_i} \mu_i (m_i - \mu_i).$$

Ainda,

$$\mathbf{X}^T \mathbf{W} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ d_1 & d_2 & \dots & d_n \end{bmatrix} \begin{bmatrix} W_1 & 0 & \dots & 0 \\ 0 & W_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & W_i \end{bmatrix} = \begin{bmatrix} W_1 & W_2 & \dots & W_n \\ W_1 d_1 & W_2 d_2 & \dots & W_n d_n \end{bmatrix}$$

$$\mathbf{X}^T \mathbf{W} \mathbf{X} = \begin{bmatrix} \sum_{i=1}^n W_i & \sum_{i=1}^n W_i d_i \\ \sum_{i=1}^n W_i d_i & \sum_{i=1}^n W_i d_i^2 \end{bmatrix}, \quad (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} = \frac{1}{\det} \begin{bmatrix} \sum_{i=1}^n W_i d_i^2 & -\sum_{i=1}^n W_i d_i \\ -\sum_{i=1}^n W_i d_i & \sum_{i=1}^n W_i \end{bmatrix},$$

$$\mathbf{X}^T \mathbf{W} \mathbf{z} = \begin{bmatrix} \sum_{i=1}^n W_i z_i \\ \sum_{i=1}^n W_i d_i z_i \end{bmatrix}$$

sendo  $\det = \sum_{i=1}^n W_i \sum_{i=1}^n W_i d_i^2 - \left[ \sum_{i=1}^n W_i d_i \right]^2$ .

Portanto,

$$\boldsymbol{\beta}^{(m+1)} = \begin{bmatrix} \beta_1^{m+1} \\ \beta_{21}^{m+1} \end{bmatrix} = \frac{1}{\det^{(m)}} \begin{bmatrix} \sum_{i=1}^n W_i^{(m)} d_i^2 \sum_{i=1}^n W_i^{(m)} z_i^{(m)} - \sum_{i=1}^n W_i^{(m)} d_i \sum_{i=1}^n W_i^{(m)} d_i z_i^{(m)} \\ \sum_{i=1}^n W_i^{(m)} \sum_{i=1}^n W_i^{(m)} d_i z_i^{(m)} - \sum_{i=1}^n W_i^{(m)} d_i \sum_{i=1}^n W_i^{(m)} z_i^{(m)} \end{bmatrix}.$$

A seguir, apresenta-se um programa em linguagem GLIM (Francis *et al.*, 1993) para o cálculo dos passos intermediários desse algoritmo. Também são apresentados os programas GLIM (*default*) e SAS para se fazer a regressão logística.

```
!      Macro para o processo iterativo
!
!
$M Iter!
$Assign par = beta1, beta2  $!
$Ca eta = beta1 + beta2 * d : mu = m * %Exp(eta)/(1 + %Exp(eta)) !
: z = eta + m * (y-mu)/(mu*(m-mu)) : W = mu*(m-mu)/m : pe = mu/m
: Det = %Cu(W) * %Cu(W*(d**2)) - (%Cu(W*d))**2 !
: beta1 = ( %Cu(W*(d**2)) * %Cu(W*z) - %Cu(W*d) * %Cu(W*d*z) )/Det!
: beta2 = ( %Cu(W) * %Cu(W*d*z) - %CU(W*d) * %Cu(W*z) )/Det!
$Assign npar= beta1, beta2  $!
$Pr : ' ITERACAO NUMERO ' *i I : $Look eta mu z W $!
$PR : 'beta1(' *i I ') = ' beta1 ' beta2(' *i I ') = ' beta2 $!
$CA %W=%Cu((par-npar)/par)**2) : I=I+1 : %N=%IF(%LT(%W,0.0001),0,1) $$E!
!
!
$Slen 6 $Data d y m $Read!
10.2 44 50!
 7.7 42 49!
 5.1 24 46!
 3.8 16 48!
 2.6  6 50!
 0.0  0 49!
$number beta1 beta2 Det I $
!
$Ca y(6)= 0.002 : p= y/m : eta=%log(y/(m-y)) $
!
!
```

```

$Print : '**** RESULTADOS ****' : $Look d m y p eta $!
$Yvar eta $Fit d $Extr %pe $Ca betal=%pe(1) : beta2=%pe(2) $
!
$PR : ' Estimativas iniciais '
: ' betal(0) = ' betal ' Beta2(0) = ' beta2 $VAR 2 par npar $!
!
$CA %N=1 : I = 1 $While %N Iter $!
$CA eta = betal + beta2 * d : mu = m * %Exp(eta)/(1 + %Exp(eta))
: pe = mu/m $!
$Print : '**** RESULTADOS FINAIS ****' : $Look d m y mu p pe $!
!
$Pr: 'USANDO OS RECURSOS DO GLIM' : $!
$Ca y(6)=0 $
$Yvar y $Err B m $!
$Fit d $Display MER $!
$FINISH!

```

[o] \*\*\*\* RESULTADOS \*\*\*\*

```

[o]
[o]          D          M          Y          P          ETA
[o]  1    10.200    50.00    44.000000    0.88000000    1.99243
[o]  2     7.700    49.00    42.000000    0.85714287    1.79176
[o]  3     5.100    46.00    24.000000    0.52173913    0.08701
[o]  4     3.800    48.00    16.000000    0.33333334   -0.69315
[o]  5     2.600    50.00     6.000000    0.12000000   -1.99243
[o]  6     0.000    49.00     0.002000    0.00004082  -10.10639
[o]      deviance = 25.834
[o] residual df = 4
[o]
[o]
[o]  Estimativas iniciais
[o]  betal(0) = -6.680 Beta2(0) = 1.060
[w] -- change to data values affects model
[o]
[o]  ITERACAO NUMERO 1
[o]
[o]          ETA          MU          Z          W
[o]  1     4.131    49.20901   -2.5607    0.77848
[o]  2     1.481    39.92041    1.7620    7.39716
[o]  3    -1.275    10.04794    0.5018    7.85314
[o]  4    -2.653     3.15962    1.6976    2.95164
[o]  5    -3.925     0.96846    1.3735    0.94970
[o]  6    -6.680     0.06144   -7.6489    0.06137
[o]
[o]  betal(1) = 1.131 beta2(1) = -0.0150
[o]
[o]  ITERACAO NUMERO 2
[o]
[o]          ETA          MU          Z          W
[o]  1     0.9776    36.33    1.74971    9.932
[o]  2     1.0152    35.97    1.64579    9.566
[o]  3     1.0543    34.11   -0.09303    8.815
[o]  4     1.0739    35.78   -1.09669    9.111
[o]  5     1.0919    37.44   -2.25019    9.406
[o]  6     1.1310    37.05   -2.96755    9.038
[o]
[o]  betal(2) = -3.065 beta2(2) = 0.5217

```

```

[O] ITERACAO NUMERO 3
[O]
[O]      ETA      MU      Z      W
[O] 1  2.2565  45.260  1.9627  4.290
[O] 2  0.9522  35.356  1.6271  9.845
[O] 3 -0.4042  18.414  0.1017  11.043
[O] 4 -1.0825  12.146 -0.6577  9.073
[O] 5 -1.7085   7.668 -1.9654  6.492
[O] 6 -3.0649   2.184 -4.1106  2.087
[O]
[O] beta1(3) = -3.158  beta2(3) = 0.5876
[O]
[O] ITERACAO NUMERO 4
[O]
[O]      ETA      MU      Z      W
[O] 1  2.8352  47.227  1.60285  2.619
[O] 2  1.3662  39.041  1.73905  7.935
[O] 3 -0.1616  21.146  0.08823  11.425
[O] 4 -0.9254  13.625 -0.68202  9.757
[O] 5 -1.6306   8.188 -1.95007  6.847
[O] 6 -3.1583   1.998 -4.19975  1.916
[O]
[O] beta1(4) = -3.223  beta2(4) = 0.6045
[O]
[O] ITERACAO NUMERO 5
[O]
[O]      ETA      MU      Z      W
[O] 1  2.9431  47.497  1.47252  2.378
[O] 2  1.4318  39.552  1.75280  7.627
[O] 3 -0.1400  21.392  0.08784  11.444
[O] 4 -0.9259  13.620 -0.68197  9.755
[O] 5 -1.6514   8.046 -1.95445  6.751
[O] 6 -3.2232   1.877 -4.26191  1.805
[O]
[O] beta1(5) = -3.225  beta2(5) = 0.6051
[O]
[O]
[O] **** RESULTADOS FINAIS ****
[O]
[O]      D      M      Y      MU      P      PE
[O] 1  10.200  50.00  44.000000  47.505  0.88000000  0.95009
[O] 2   7.700  49.00  42.000000  39.567  0.85714287  0.80748
[O] 3   5.100  46.00  24.000000  21.399  0.52173913  0.46519
[O] 4   3.800  48.00  16.000000  13.619  0.33333334  0.28372
[O] 5   2.600  50.00   6.000000   8.041  0.12000000  0.16081
[O] 6   0.000  49.00   0.002000   1.873  0.00004082  0.03822
[O]
[O] USANDO OS RECURSOS DO GLIM
[O]
[O] scaled deviance = 10.258 at cycle 4
[O]   residual df = 4
[O]
[O] Current model:
[O]
[O]   number of observations in model is 6
[O]
[O]   y-variate  Y
[O]   weight     *
[O]   offset     *

```

```

[o] probability distribution is BINOMIAL
[o] and binomial denominator M
[o] link function is LOGIT
[o] scale parameter is 1.000
[o]
[o] linear model:
[o] terms: 1+D
[o] estimate s.e. parameter
[o] 1 -3.226 0.3699 1
[o] 2 0.6051 0.06781 D
[o] scale parameter 1.000
[o]
[o] unit observed out of fitted residual
[o] 1 44 50 47.505 -2.277
[o] 2 42 49 39.567 0.881
[o] 3 24 46 21.398 0.769
[o] 4 16 48 13.618 0.763
[o] 5 6 50 8.040 -0.785
[o] 6 0 49 1.872 -1.395
[o]
[i] ? $stop

```

### Programa SAS para realizar a regressão logística:

```

options nodate nonumber ps=1000;
Data Doses;
Input dose m y;
datalines;
10.2 50 44
7.7 49 42
5.1 46 24
3.8 48 16
2.6 50 6
0.0 49 0
1.0 ;
proc genmod;
model y/m=dose / dist=b ;
output out=saida p=predito;
run;

Data novo;
set saida;
pobs=y/m;
Yest=m*predito;
run;

proc print data=novo;
run;

```

## 2.6 Propriedades e distribuição amostral de $\hat{\beta}$

No caso particular dos modelos lineares em que as variáveis respostas têm distribuição Normal, as distribuições dos estimadores dos parâmetros e das estatísticas usadas para verificação do ajuste do modelo aos dados podem ser determinadas exatamente. Em geral, porém, a obtenção de distribuições exatas é muito complicada e resultados assintóticos são usados. Esses resultados, porém, são dependentes de várias condições de regularidade e dos tamanhos amostrais. No caso de observações independentes provenientes de distribuições amostrais pertencentes à família exponencial e, em particular, para os modelos lineares generalizados essas condições são satisfeitas (Dobson, 1990; Fahrmeir & Kaufman, 1985; Cox & Hinkley, 1986).

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

- i)  $\hat{\theta}$  é assintoticamente imparcial;
- ii) a estatística

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

ou, de forma equivalente,

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

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

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

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

Um estudo das propriedades (existência, finitude e unicidade) de  $\hat{\beta}$  é apresentado por Cordeiro (1986), às páginas 29 e 30.

De uma forma resumida têm-se, a seguir, algumas propriedades para o estimador  $\hat{\beta}$ :

- i) O estimador  $\hat{\beta}$  é assintoticamente não viesado, isto é, para amostras grandes  $E(\hat{\beta}) = \beta$ .

Suponha que o logaritmo da função de verossimilhança tem um único máximo em  $\hat{\beta}$  que está próximo do verdadeiro valor de  $\beta$ . A aproximação de Taylor até termos de 1ª ordem para o vetor escore  $\mathbf{U}(\hat{\beta})$  em relação a  $\beta$ , substituindo-se a matriz de derivadas parciais por  $-\mathfrak{I}$  (em que  $\mathfrak{I}$  é a matriz de informação de Fisher), é dada por

$$\mathbf{U}(\hat{\beta}) = \mathbf{U}(\beta) - \mathfrak{I}(\hat{\beta} - \beta) = \mathbf{0}$$

pois,  $\hat{\boldsymbol{\beta}}$  é a solução do sistema de equações  $\mathbf{U}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$ . Portanto,

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \mathfrak{Z}^{-1} \mathbf{U}(\boldsymbol{\beta})$$

desde que  $\mathfrak{Z}$  seja não-singular. Tem-se, então, que

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

pois,  $E[\mathbf{U}(\boldsymbol{\beta})] = \mathbf{0}$  e, portanto,  $\hat{\boldsymbol{\beta}}$  é um estimador imparcial para  $\boldsymbol{\beta}$  (pelo menos assintoticamente).

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

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

pois,  $\mathfrak{Z} = E(\mathbf{U}\mathbf{U}^T)$  e  $(\mathfrak{Z}^{-1})^T = \mathfrak{Z}^{-1}$ , pois  $\mathfrak{Z}^{-1}$  é simétrica.

iii) Para amostras grandes, tem-se

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathfrak{Z}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \chi_p^2 \quad (2.9)$$

ou, de forma equivalente,

$$\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \mathfrak{Z}^{-1}) \quad (2.10)$$

que é a base para a construção de testes e intervalos de confiança para os modelos lineares generalizados. Para modelos lineares com variáveis respostas com distribuição normal, (2.9) e (2.10) são exatas.

Para amostras pequenas  $\hat{\boldsymbol{\beta}}$  é tendencioso. Além disso, para  $n$  não muito grande a estrutura de covariâncias das estimativas dos parâmetros lineares difere de  $\mathfrak{Z}^{-1}$ . A matriz  $\mathfrak{Z}$  é consistentemente estimada por

$$\hat{\mathfrak{Z}} = \frac{1}{\phi} \mathbf{X}^T \hat{\mathbf{W}} \mathbf{X},$$

sendo  $\mathbf{X}$  a matriz do modelo,  $\mathbf{W} = \text{diag}\{W_1, W_2, \dots, W_n\}$ ,  $W_i = \frac{w_i}{V(\mu_i)} \left( \frac{d\mu_i}{d\eta_i} \right)^2$  e  $\phi > 0$ ,

constante e conhecido. Para as distribuições binomial e de Poisson  $\phi = 1$ . Se  $\phi$  for constante para todas as observações e desconhecido afetará a estrutura assintótica de  $\hat{\mathfrak{Z}}^{-1}$  (com elementos representados por  $v_{jk}$ ,  $k = j = 1, 2, \dots, p$ ) mas não o valor de  $\hat{\boldsymbol{\beta}}$ . Na prática se  $\phi$  é desconhecido (distribuições normal e normal inversa  $\phi = \sigma^2$  e gama  $\phi = v^{-1}$ ), deve ser substituído por alguma estimativa consistente (ver seção 2.8).

Os erros padrões dos estimadores  $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$  são iguais às raízes quadradas dos elementos da diagonal de  $\hat{\mathfrak{Z}}^{-1}$ , isto é,  $s(\hat{\beta}_j) = v_{jj}$ . Então, intervalos de confiança assintóticos,

com um coeficiente de confiança de 95% de probabilidade, para os parâmetros  $\beta_j$ 's podem ser obtidos por

$$\hat{\beta}_j \pm 1,96 \sqrt{v_{jj}}.$$

A partir de  $\hat{\mathfrak{Z}}^{-1}$ , pode-se calcular a correlação entre  $\hat{\beta}_j$  e  $\hat{\beta}_k$ , ou seja,

$$\hat{\rho}_{jk} = \text{Côrr}(\hat{\beta}_j; \hat{\beta}_k) = \frac{\text{Côv}(\hat{\beta}_j; \hat{\beta}_k)}{\sqrt{\text{Vâr}(\hat{\beta}_j) \text{Vâr}(\hat{\beta}_k)}} = \frac{v_{jk}}{\sqrt{v_{jj} v_{kk}}}$$

que permite verificar, pelo menos aproximadamente, a interdependência dos  $\hat{\beta}_j$ 's.

**Exemplo 10:** Seja  $Y_1, Y_2, \dots, Y_n$  uma amostra aleatória de uma distribuição  $N(\mu_i, \sigma^2)$ , sendo que  $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$  e  $\sigma^2 > 0$ , conhecida. Considerando como função de ligação a identidade, isto é,  $\eta_i = \mu_i$ , tem-se que

$$g'(\mu_i) = \frac{d\eta_i}{d\mu_i} = 1.$$

Além disso,  $V(\mu_i) = 1$ ,  $w_i = 1$  e, portanto,  $W_i = 1$ . Logo, tem-se

$$\mathfrak{Z} = \frac{1}{\phi} \mathbf{X}^T \mathbf{W} \mathbf{X} = \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X}$$

e a variável dependente ajustada

$$z_i = \hat{\eta}_i + g'(\hat{\mu}_i)(y_i - \hat{\mu}_i) = \hat{\mu}_i + y_i - \hat{\mu}_i = y_i.$$

Portanto, o algoritmo de estimação (2.7) fica

$$\frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} = \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{y}$$

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

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

que é a solução usual de mínimos quadrados para Modelos Lineares Clássicos. Tem-se, então, que

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

e

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

pois,  $E[(Y - X\beta)(Y - X\beta)^T] = I\sigma^2$  e  $\mathfrak{S} = \frac{1}{\sigma^2} X^T X$ .

Logo,

$$(\hat{\beta} - \beta)^T \mathfrak{S} (\hat{\beta} - \beta) \sim \chi_p^2. \quad \text{(EXATA!)}$$

## 2.7 Função *Deviance* e estatística de Pearson $X^2$ generalizada

Segundo McCullagh & Nelder (1989), o ajuste de um modelo a um conjunto de dados observados  $y$  pode ser encarado como uma maneira de se substituir  $y$  por um conjunto de valores estimados  $\hat{\mu}$  para um modelo com um número de parâmetros relativamente pequeno. Logicamente os  $\hat{\mu}$ 's não serão exatamente iguais aos  $y$ 's, e a questão, então, que aparece é em quanto eles diferem. Isto 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 (ou covariáveis) pode levar 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 (ou covariáveis) pode levar a um modelo de interpretação fácil, porém, que se ajuste pobremente aos dados. O que se deseja na realidade é um modelo intermediário.

Dadas  $n$  observações, a elas 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 1's. Esse modelo atribui toda a variação entre os  $y$ 's ao componente aleatório. No outro extremo, está o **modelo saturado** ou **completo** que tem  $n$  parâmetros, um para cada observação. Ele atribui toda a variação ao componente sistemático e, portanto, ajusta-se perfeitamente, reproduzindo os próprios dados.

Na prática o modelo nulo é simples demais e o modelo saturado não é informativo, pois não resume os dados, mas simplesmente os repete. Existem, contudo, dois outros modelos limitantes, porém, menos extremos. Certos parâmetros têm que estar no modelo como é o caso, por exemplo, de efeitos de blocos, ou então, totais marginais fixados em tabelas de contingência. O modelo contendo apenas esses parâmetros é chamado **modelo minimal** pois é o modelo que contém o menor número de termos necessários para o ajuste. Por outro lado, o modelo que contém o maior número de termos que podem ser considerados é chamado de **modelo maximal**. Os termos desses modelos extremos são, geralmente, obtidos por interpretações *a priori*, da estrutura dos dados.

Assim, por exemplo, considerando-se um experimento em blocos casualizados, com tratamentos no esquema fatorial com 2 fatores, têm-se os modelos:

$$\text{nulo: } \eta_i = \mu$$

$$\text{minimal: } \eta_i = \mu + \beta_\ell$$

$$\text{maximal: } \eta_i = \mu + \beta_\ell + \alpha_j + \gamma_k + (\alpha\gamma)_{jk}$$

$$\text{saturado: } \eta_i = \mu + \beta_\ell + \alpha_j + \gamma_k + (\alpha\gamma)_{jk} + (\beta\alpha)_{\ell j} + (\beta\gamma)_{\ell k} + (\beta\alpha\gamma)_{\ell jk},$$

sendo,  $\mu$  efeito associado à média geral;

$\beta_\ell$  efeito associado ao bloco  $\ell$ ,  $\ell=1,2,\dots,b$ ;

$\alpha_j$  efeito associado ao  $j$ -ésimo nível do fator A;

$\gamma_k$  efeito associado ao  $k$ -ésimo nível do fator B;

$(\alpha\gamma)_{jk}$ ,  $(\beta\alpha)_{\ell j}$ ,  $(\beta\gamma)_{\ell k}$  e  $(\beta\alpha\gamma)_{\ell jk}$  efeitos associados às interações.

O modelo saturado inclui, nesse caso, todas as interações com blocos que não são de interesse prático.

Em geral, trabalha-se com modelos encaixados e o conjunto de matrizes dos modelos pode, então, ser formado pela adição 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 chamado **modelo corrente** ou **modelo sob pesquisa**. O problema é determinar a utilidade de um parâmetro extra no modelo corrente (sob pesquisa) ou, então, verificar a falta de ajuste induzida pela omissão dele. A fim de discriminar entre modelos, medidas de discrepância devem ser introduzidas para medir o ajuste de um modelo.

Nelder & Wedderburn (1972) propuseram como medida de discrepância a *deviance* (traduzida como desvio por Cordeiro (1986)), com expressão dada por:

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

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. Vê-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, assumindo que  $a_i(\phi) = \frac{\phi}{w_i}$ , tem-se:

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

e

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

sendo  $\tilde{\theta}_i = \tilde{\theta}(y_i)$  e  $\hat{\theta}_i = \hat{\theta}(\hat{\mu}_i)$ , as estimativas do parâmetro canônico sob os modelos saturado e corrente, respectivamente. Tem-se, então,

$$S_p = \frac{1}{\phi} \sum_{i=1}^n 2w_i \{y_i [\tilde{\theta}_i - \hat{\theta}_i] - b(\tilde{\theta}_i) + b(\hat{\theta}_i)\} = \frac{1}{\phi} D_p \quad (2.11)$$

em que  $S_p$  é chamada *scaled deviance* e  $D_p$ , *deviance*. Pode-se, ainda escrever

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

sendo que  $d_i^2$  mede a diferença dos logaritmos das funções de verossimilhanças observada e ajustada, para a observação correspondente e é chamado componente da *deviance*. A soma deles mede a discrepância total entre as duas funções de verossimilhanças. É portanto, uma medida da distância dos valores ajustados  $\hat{\mu}$ 's em relação aos dados observados  $y$ 's, ou de forma equivalente, do modelo corrente em relação ao modelo saturado. Verifica-se que a *deviance* equivale a uma constante menos duas vezes o máximo da função de verossimilhança para o modelo corrente, isto é,

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

**Exemplo 11:** Seja  $Y_1, Y_2, \dots, Y_n$  uma amostra aleatória de uma distribuição  $N(\mu_i, \sigma^2)$ , sendo que  $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$  e  $\sigma^2 > 0$ , conhecida. Considerando como função de ligação a identidade, isto é,  $\eta_i = \mu_i$ , tem-se que

$$\phi = \sigma^2; \quad w_i = 1; \quad \theta_i = \mu_i \quad \text{e} \quad b(\theta_i) = \frac{\theta_i^2}{2} = \frac{\mu_i^2}{2}.$$

Logo,

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

que coincide com a estatística clássica SQRes com  $(n-p)$  graus de liberdade dividida por  $\sigma^2$ .

**Exemplo 12:** Sejam  $Y_i$  variáveis aleatórias representando contagens de sucessos em amostras independentes de tamanhos  $m_i$ . Supondo que  $Y_i \sim \text{Bin}(m_i, \pi_i)$ , então,

$$\phi = 1; \quad w_i = 1; \quad \theta_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \ln\left(\frac{\mu_i}{m_i - \mu_i}\right)$$

e

$$b(\theta_i) = m_i \ln(1 + e^{\theta_i}) = -m_i \ln(1 - \pi_i) = -m_i \ln\left(\frac{m_i - \mu_i}{m_i}\right).$$

Logo,

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

ou ainda,

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

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  $2m_i \ln\left(\frac{m_i}{m_i - \hat{\mu}_i}\right)$  ou  $2m_i \ln\left(\frac{m_i}{\hat{\mu}_i}\right)$ , respectivamente (Paula, 2000).

Se  $m_i = 1$  e a função de ligação considerada é a logística, a função *deviance* é apenas uma função dos dados e, portanto, não é informativa a respeito do ajuste do modelo aos dados (Exercício 2.12.12). O mesmo é válido para as funções de ligação probit e complemento log-log.

As funções *deviance (scaled)* para as distribuições estudadas no capítulo 1 estão na Tabela 9. A *deviance* para a distribuição normal é, simplesmente, a soma de quadrados de resíduos, dividida pela variância  $\sigma^2$ , como visto no **Exemplo 11**, enquanto que para a distribuição de Poisson é a estatística  $G^2$ , usada em modelos log-lineares. Para a distribuição gama, se algum componente é igual a zero, segundo Paula (2000), pode-se substituir  $S_p$  por

$$S_p = 2C(\mathbf{y}) + 2v \sum_{i=1}^n w_i \left[ \ln \hat{\mu}_i + \frac{y_i}{\hat{\mu}_i} \right],$$

sendo  $C(\mathbf{y})$  uma função arbitrária, porém limitada. Pode ser usada, por exemplo, a expressão

$$C(\mathbf{y}) = \sum_{i=1}^n w_i \frac{y_i}{1 + y_i}.$$

Tabela 9: Funções *deviance* para algumas distribuições

Distribuição	<i>Scaled deviance</i>
Normal	$S_p = \frac{1}{\sigma^2} \sum_{i=1}^n w_i (y_i - \hat{\mu}_i)^2$
Binomial	$S_p = 2 \sum_{i=1}^n w_i \left[ y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) + (m_i - y_i) \ln\left(\frac{m_i - y_i}{m_i - \hat{\mu}_i}\right) \right]$
Poisson	$S_p = 2 \sum_{i=1}^n w_i \left[ y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i) \right]$
Binomial negativa	$S_p = 2 \sum_{i=1}^n w_i \left[ y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i + k) \ln\left(\frac{y_i + k}{\hat{\mu}_i + k}\right) \right]$
Gama	$S_p = 2v \sum_{i=1}^n w_i \left[ -\ln\left(\frac{y_i}{\hat{\mu}_i}\right) + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \right]$
Normal Inversa	$S_p = \frac{1}{\sigma^2} \sum_{i=1}^n w_i \frac{(y_i - \hat{\mu}_i)^2}{y_i \hat{\mu}_i^2}$

A *deviance* é sempre maior do que ou igual a zero, e à medida que entram variáveis explanatórias (ou covariáveis) no componente sistemático, decresce até se tornar zero para o modelo saturado. Quanto melhor for o ajuste do modelo aos dados tanto menor será o valor de  $S_p$ . Assim, um modelo bem ajustado aos dados com uma verossimilhança grande tem uma

*deviance* pequena. Uma maneira de se conseguir a diminuição da *deviance* é aumentar o número de parâmetros, o que, porém, significa um aumento do grau de complexidade na interpretação do modelo. Na prática, procuram-se modelos simples com *deviance* moderada, situados entre os modelos mais complicados e os que se ajustam mal aos dados.

Para testar a adequação de um modelo linear generalizado, o valor calculado para a *deviance* (com  $(n-p)$  graus de liberdade, sendo  $p$  o posto da matriz do modelo) deve ser comparado com o percentil de alguma distribuição de probabilidade referência. Para a distribuição normal, assumindo-se que o modelo usado é verdadeiro e que  $\sigma^2$  é conhecido, tem-se que

$$S_p = \frac{D_p}{\sigma^2} \sim \chi_{n-p}^2 \text{ (exata).}$$

Em alguns casos especiais, 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, a *deviance*, não tem distribuição  $\chi_{n-p}^2$ , nem mesmo assintoticamente, e pouco se sabe sobre a adequação de uma aproximação para amostras pequenas, estudos adicionais sendo necessários. Nesse caso, tem sido usada a técnica *bootstrap* como uma possível solução.

Assumindo-se que o modelo usado é verdadeiro, para a distribuição 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 a distribuição de Poisson, quando  $\mu_i \rightarrow \infty, \forall i$ , tem-se que (lembre-se que  $\phi = 1$ ):

$$S_p = D_p \sim \chi_{n-p}^2.$$

Nos casos em que  $S_p$  depende do parâmetro  $\phi$  (conhecido), Jørgensen (1987) mostra que

$$S_p \sim \chi_{n-p}^2, \text{ quando } \phi \rightarrow 0,$$

isto é, quando a dispersão é pequena. Para a distribuição gama a aproximação da distribuição de  $S_p$  de uma  $\chi_{n-p}^2$  será tanto melhor quanto mais próximo de 1 estiver o coeficiente de variação (Paula, 2000). Em geral, porém, não se conhece  $\phi$  e ele precisa ser substituído por uma estimativa consistente.

Na prática, contenta-se em testar um modelo linear generalizado, sem muito rigor, comparando-se o valor  $S_p$  com os percentis da distribuição  $\chi_{n-p}^2$  (McCullagh & Nelder, 1989). Assim, nos casos em que é possível a aproximação de uma  $\chi_{n-p}^2$ , tem-se que se

$$S_p \leq \chi_{n-p;\alpha}^2$$

pode-se considerar que existem evidências, a um nível aproximado de  $100\alpha\%$  de probabilidade, que o modelo proposto está bem ajustado aos dados. Ou ainda, lembrando que

se  $Z \sim \chi_{n-p}^2$ , então,  $E(Z) = n - p$ , e, portanto, um valor de  $S_p$  próximo de  $(n - p)$  pode ser uma indicação de que o modelo ajustado aos dados é adequado.

Outra medida da discrepância de ajuste de um modelo a um conjunto de dados é a estatística de Pearson  $X^2$  generalizada cuja expressão é dada por:

$$X^2 = \sum_{i=1}^n w_i \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)},$$

sendo  $V(\hat{\mu}_i)$  a função de variância estimada sob o modelo que está sendo ajustado aos dados.

Para respostas com distribuição normal,  $X^2 = \text{SQRes}$  e

$$\frac{X^2}{\sigma^2} \sim \chi_{n-p}^2 \text{ (exata).}$$

Para dados provenientes das distribuições binomial e de Poisson, em que  $\phi = 1$ ,  $X^2$  é a estatística original de Pearson, comumente usada na análise dos modelos logístico e log-linear para tabelas multidimensionais e que pode ser escrita na forma

$$X^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i},$$

sendo  $O_i$  a frequência observada e  $E_i$  a frequência esperada.

Para as distribuições não-normais, têm-se apenas resultados assintóticos, isto é, a distribuição  $\chi_{n-p}^2$  pode ser usada, somente, como uma aproximação, que em muitos casos pode ser pobre. Além disso,  $X^2$  tem como desvantagem o fato de tratar os  $y_i$ 's simetricamente. Em muitos casos, é preferida em relação à *deviance*, por facilidade de interpretação.

**Exemplo 13:** Considere os dados do **Exemplo 5**. A variável resposta tem distribuição binomial, isto é,  $Y_i \sim \text{Bin}(m_i, \pi_i)$ . Adotando-se a função de ligação logística (canônica) e o preditor linear dado por uma regressão linear simples, isto é,

$$\eta_i = \ln\left(\frac{\mu_i}{m_i - \mu_i}\right) = \beta_1 + \beta_2 d_i$$

tem-se  $S_p = D_p = 10,26$  e  $X^2 = 9,70$  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 5% e 1% de probabilidade que o modelo logístico linear ajusta-se razoavelmente a esse conjunto de dados. Necessita-se, porém, adicionalmente, do teste da hipótese  $H_0 : \beta_2 = 0$ , de uma análise de resíduos e de diagnósticos.

## 2.8 Estimação do parâmetro $\phi$

Para as distribuições binomial e Poisson tem-se  $\phi = 1$ . Quando desconhecido (distribuições normal e normal inversa  $\phi = \sigma^2$  e gama  $\phi = \nu^{-1}$ ), admite-se que seja o mesmo para todas as observações, isto é, constante. Necessária se faz sua estimação para a obtenção (conforme visto na seção 2.6) dos erros padrões 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 da máxima verossimilhança, método dos momentos e perfil de verossimilhança.

O método da máxima verossimilhança é sempre possível em teoria, mas pode-se tornar intratável computacionalmente quando não existe solução explícita. Se  $\phi$  é o mesmo para todas as observações, a estimativa de máxima verossimilhança de  $\beta$  independe de  $\phi$  mas já a matriz de variâncias e covariâncias dos  $\beta$ 's envolve esse parâmetro. Interpretando o logaritmo da função de verossimilhança  $\ell(\boldsymbol{\beta}, \phi; \mathbf{y})$  como função de  $\beta$  e de  $\phi$ , dado  $\mathbf{y}$ , a estimativa de máxima verossimilhança para  $\phi$  é obtida, fazendo-se

$$\frac{\partial \ell(\boldsymbol{\beta}, \phi)}{\partial \phi} = 0.$$

Para as distribuições normal e normal inversa tem-se

$$\hat{\phi} = \frac{1}{n} D_p.$$

**Exemplo 14:** Seja  $Y_1, Y_2, \dots, Y_n$  uma amostra aleatória de uma distribuição  $N(\mu_i, \sigma^2)$ . Então, o logaritmo da função de verossimilhança fica

$$\ell = -\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\phi} - \frac{n}{2} \ln(2\pi\phi)$$

cuja derivada em relação a  $\phi$  é

$$\frac{\partial \ell}{\partial \phi} = \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\phi^2} - \frac{n}{2\phi}$$

e fazendo-se  $\frac{\partial \ell}{\partial \phi} = 0$  tem-se

$$\hat{\phi} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 = \frac{1}{n} D_p.$$

Conforme se verifica a estimativa de máxima verossimilhança para  $\phi$ , no caso do modelo normal, é exata. No caso da distribuição gama, Cordeiro (1986), usando a expansão de Taylor para a função digama ( $\psi(\phi) = \Gamma'(\phi)/\Gamma(\phi)$ ), para  $\phi$  grande, propõe a aproximação (Exercício 2.12.15)

$$\frac{1}{\hat{\phi}} = \frac{n}{2D_p} \left[ 1 + \sqrt{1 + \frac{2D_p}{3n}} \right]$$

e para dados com grande dispersão ela é inconsistente.

Verifica-se que (Paula, 2000):

$$\text{Var}(\hat{\phi}^{-1}) = -\sum_{i=1}^n c''(y_i, \phi).$$

Além disso, pode-se mostrar que

$$E\left(\frac{\partial^2 \ell(\boldsymbol{\beta}, \phi)}{\partial \beta_j \partial \phi}\right) = 0$$

e, portanto, os parâmetros  $\boldsymbol{\beta}$  e  $\phi$  são ortogonais (Exercício 2.12.14).

O método dos momentos fornece uma outra estimativa, também não consistente, para  $\phi$ . Esse método baseia-se no fato (que nem sempre é verdadeiro) de que  $S_p \sim \chi_{n-p}^2$ . Logo,

$$E(S_p) = \frac{1}{\phi} E(D_p) \cong n - p$$

e, portanto,

$$\tilde{\phi} = \frac{D_p}{n - p},$$

sendo  $D_p$  a *deviance* sob o modelo corrente (sob pesquisa).

Uma estimativa considerada melhor que essa é

$$\tilde{\phi} = \frac{D_m}{n - m},$$

sendo  $D_m$  a *deviance* sob o modelo maximal. Espera-se que para um modelo maximal bem ajustado aos dados  $S_m$  tenha um valor mais próximo da esperança da qui-quadrado de

referência, isto é,  $E(S_m) = \frac{1}{\phi} E(D_m) \cong n - m$ . Para o modelo normal

$$\tilde{\phi} = \frac{D_m}{n - m} = \frac{SQRes}{n - m}$$

é a estimativa usual de  $\sigma^2$  e é não viesada, mas para os outros modelos isso não acontece, em geral.

Uma outra maneira de se estimar  $\phi$  é baseada na estatística de Pearson  $X^2$  generalizada (Jørgensen, 1987) e é dada por

$$\phi^* = \frac{X^2}{n - m}$$

que nem sempre é imparcial, porém, é consistente.

Barndorff-Nielsen (1988) propôs o método do perfil de verossimilhança modificado (*modified profile likelihood*) para estimar  $\phi$ .

## 2.9 Análise de *deviance* e seleção de modelos

A análise de *deviance* (ANODEV) é uma generalização da análise da variância para os modelos lineares generalizados, visando obter, a partir de uma seqüência de modelos, cada um incluindo mais termos do que os anteriores, os efeitos de fatores, covariáveis e suas interações. Dada uma seqüência de modelos encaixados, utiliza-se a *deviance* como uma medida de discrepância do modelo e forma-se uma tabela de diferença de *deviances*.

Seja  $M_{p_1}, M_{p_2}, \dots, M_{p_r}$  uma seqüência 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 *deviances*  $D_{p_1} > D_{p_2} > \dots > D_{p_r}$ , tendo os modelos a mesma distribuição e a mesma função de ligação. Estas desigualdades entre as *deviances*, em geral, não se verificam para a estatística de Pearson  $X^2$  generalizada e, por esta razão, a comparação de modelos encaixados é feita, principalmente, via função *deviance*. 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 10.

Tabela 10: Um exemplo de construção de uma tabela de Análise de *Deviance*

Modelo	G.L.	Deviance	Dif. de <i>deviances</i>	Dif. de G.L.	Significado
Nulo	$rab-1$	$D_1$			
A	$a(rb-1)$	$D_A$	$D_1$	$a-1$	$A$ ignorando $B$
A+B	$a(rb-1)-(b-1)$	$D_{A+B}$	$D_A - D_{A+B}$	$b-1$	$B$ incluído $A$
A+B+A.B	$ab(r-1)$	$D_{A*B}$	$D_{A+B} - D_{A*B}$	$(a-1)(b-1)$	Interação $AB$ incluídos $A$ e $B$
Saturado	0	0	$D_{A*B}$		Resíduo

Dois termos  $A$  e  $B$  são ortogonais se a redução que  $A$  (ou  $B$ ) causa na *deviance*  $D_p$  é a mesma, esteja  $B$  (ou  $A$ ) incluído, ou não, em  $M_p$ . Em geral, para os modelos lineares generalizados ocorre a não-ortogonalidade dos termos e a interpretação da tabela ANODEV é mais complicada do que a ANOVA usual.

Sejam os modelos  $M_p$  e  $M_q$  ( $p < q$ ) com  $p$  e  $q$  parâmetros, respectivamente. A estatística  $D_p - D_q$  com  $(q - p)$  graus de liberdade é interpretada como uma medida de variação dos dados, explicada pelos termos que estão em  $M_q$  e não estão em  $M_p$ , incluídos os efeitos dos termos em  $M_p$  e ignorando quaisquer efeitos dos termos que não estão em  $M_q$ . Tem-se, assintoticamente, para  $\phi$  conhecido, que

$$S_p - S_q = \frac{1}{\phi} (D_p - D_q) \sim \chi_{q-p}^2,$$

que é simplesmente o teste da razão de verossimilhanças (ver seção 2.10). Se  $\phi$  é desconhecido, deve-se obter uma estimativa  $\hat{\phi}$  consistente, de preferência baseada no modelo maximal (com  $m$  parâmetros), e inferência pode ser baseada na estatística  $F$ , dada por

$$F = \frac{(D_p - D_q)/(q - p)}{\hat{\phi}} \sim F_{q-p, n-m}.$$

Para a distribuição normal, tem-se

$$\frac{(\text{SQRes}_p - \text{SQRes}_q)/(q - p)}{\text{SQRes}_m/(n - m)} \sim F(q - p, n - m) \text{ (exata).}$$

**Exemplo 15:** Considere os dados do **Exemplo 5**. A variável resposta tem distribuição binomial, isto é,  $Y_i \sim \text{Bin}(m_i, \pi_i)$ . Adotando-se a função de ligação logística (canônica) e o preditor linear dado por uma regressão linear simples, isto é,

$$\eta_i = \ln\left(\frac{\mu_i}{m_i - \mu_i}\right) = \beta_1 + \beta_2 d_i,$$

dois modelos encaixados podem ser propostos para a análise desses dados, a saber:

- a) modelo nulo:  $\eta_i = \beta_1$  e
- b) modelo de regressão linear:  $\eta_i = \beta_1 + \beta_2 d_i$ .

A Tabela 11 apresenta as *deviances* e seus respectivos números de graus de liberdade (g.l.), correspondentes a eles e a Tabela 12, a Análise de *deviance*.

Tabela 11: *Deviances* e  $X^2$  residuais

Modelo	g.l.	Deviances	$X^2$
$\eta_i = \beta_1$	5	163,74	135,70
$\eta_i = \beta_1 + \beta_2 d_i$	4	10,26	9,70

$\chi_{4;0,05}^2 = 9,49$ ;  $\chi_{4;0,01}^2 = 13,29$

O exame da Tabela 11, confirmando o que já foi visto no **Exemplo 13**, mostra que existem evidências, a um nível de significância entre 5% e 1% de probabilidade, que o modelo logístico linear ajusta-se razoavelmente a esse conjunto de dados, mas rejeita-se o modelo nulo. Pelo exame da Tabela 12 rejeita-se a hipótese  $H_0: \beta_2 = 0$ , confirmando a adequação do modelo logístico linear. Necessita-se, porém, adicionalmente, de uma análise de resíduos e de diagnósticos.

Tabela 12: Análise de *Deviance*

Causa de variação	g.l.	<i>Deviances</i>	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  $\hat{\beta}_1 = -3,226$  ( $s(\hat{\beta}_1) = 0,3699$ ) e  $\hat{\beta}_2 = 0,6051$  ( $s(\hat{\beta}_2) = 0,0678$ ). O número esperado de insetos mortos  $\hat{\mu}_i$  para a dose  $d_i$  é dado por:

$$\hat{\mu}_i = m_i \frac{\exp(-3,226 + 0,6051d_i)}{1 + \exp(-3,226 + 0,6051d_i)}.$$

Na Figura 4 estão representados o modelo ajustado e os valores observados.

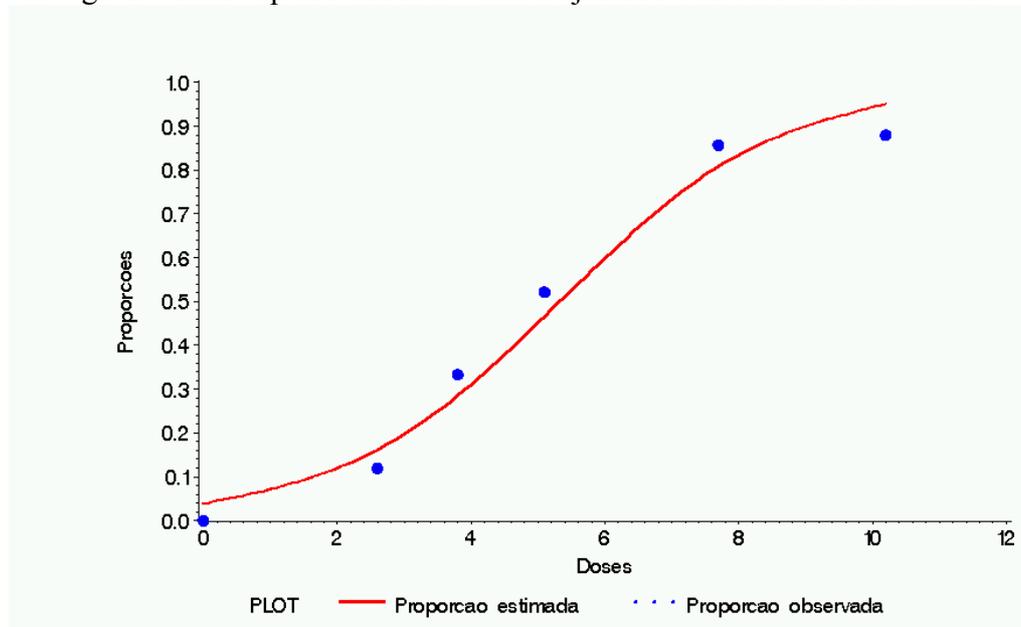


Figura 4: Valores observados e modelo ajustado

Um programa simples em linguagem GLIM (Francis *et al.*, 1993) para a obtenção desses resultados é dado a seguir.

```

$Slen 6 $Data d y m $Read!
10.2 44 50!
 7.7 42 49!
 5.1 24 46!
 3.8 16 48!
 2.6  6 50!
 0.0  0 49!
$Yvar y $Err B m $!
$Fit $Print : 'X2= ' %X2 : $
$Fit + d $Print : 'X2= ' %X2 : $
$Display MER $!
$FINISH!

```

## 2.10 Testes de hipóteses

Os métodos de inferência nos modelos lineares generalizados baseiam-se, fundamentalmente, na teoria de máxima verossimilhança. De acordo com esta teoria, existem três estatísticas para testar hipóteses relativas aos parâmetros  $\beta$ 's, que são deduzidas de distribuições assintóticas de funções adequadas das estimativas dos  $\beta$ 's. São elas:

- i) razão de verossimilhanças;
- ii) Wald e
- iii) escore,

assintoticamente equivalentes e, sob  $H_0$  e para  $\phi$  conhecido, convergem para uma variável 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, pág.108-122), Cordeiro (1986), Dobson (1990), Paula (2000) e McCulloch (2001).

Quando se tem um vetor de parâmetros, muitas vezes há interesse no teste de hipóteses de apenas um subconjunto deles. Seja, então, uma partição do vetor de parâmetros  $\beta$  dada 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 *nuisance*. De forma semelhante tem-se a partição da matriz do modelo  $\mathbf{X} = [\mathbf{X}_1 \ \mathbf{X}_2]$ , do vetor escore  $\mathbf{U} = \phi^{-1} \mathbf{X}^T \mathbf{W} \Delta(\mathbf{y} - \boldsymbol{\mu}) = [\mathbf{U}_1^T \ \mathbf{U}_2^T]^T$  com  $\mathbf{U}_1 = \phi^{-1} \mathbf{X}_1^T \mathbf{W} \Delta(\mathbf{y} - \boldsymbol{\mu})$  e da matriz de informação de Fisher para  $\hat{\beta}$

$$\mathfrak{I} = \frac{1}{\phi} \mathbf{X}^T \mathbf{W} \mathbf{X} = \begin{bmatrix} \mathfrak{I}_{11} & \mathfrak{I}_{12} \\ \mathfrak{I}_{21} & \mathfrak{I}_{22} \end{bmatrix},$$

sendo que  $\mathfrak{I}_{12} = \mathfrak{I}_{21}^T$ .

Usando-se resultados conhecidos de álgebra de matrizes, envolvendo partição de matrizes (Searle, 1982), tem-se, para amostras grandes, a variância assintótica de  $\hat{\beta}_1$ :

$$\text{Var}(\hat{\beta}_1) = (\mathfrak{I}_{11} - \mathfrak{I}_{12} \mathfrak{I}_{22}^{-1} \mathfrak{I}_{21})^{-1} = \phi [\mathbf{X}_1^T \mathbf{W}^{1/2} (\mathbf{I} - \mathbf{H}_2) \mathbf{W}^{1/2} \mathbf{X}_1]^{-1},$$

sendo  $\mathbf{H}_2 = \mathbf{W}^{1/2} \mathbf{X}_2 (\mathbf{X}_2^T \mathbf{W} \mathbf{X}_2)^{-1} \mathbf{X}_2^T \mathbf{W}^{1/2}$ .

Sejam as hipóteses

$$H_0 : \beta_1 = \beta_{1,0} \text{ versus } H_a : \beta_1 \neq \beta_{1,0},$$

sendo  $\beta_{1,0}$  um valor especificado para  $\beta_1$ . Seja  $\hat{\beta} = [\hat{\beta}_1^T \ \hat{\beta}_2^T]^T$  o estimador de máxima verossimilhança para  $\beta$  sem restrição e  $\hat{\beta}_0 = [\beta_{1,0}^T \ \hat{\beta}_{2,0}^T]^T$  em que  $\hat{\beta}_{2,0}$  é o estimador de

máxima verossimilhança para  $\beta_2$  sob  $H_0$ . A seguir são definidos os três testes mais usados para testar a hipótese  $H_0$ .

### 2.10.1 Teste da razão de verossimilhanças

Envolve a comparação dos valores do logaritmo da função de verossimilhança maximizada sem restrição ( $\ell(\hat{\beta}_1, \hat{\beta}_2; \mathbf{y})$ ) e sob  $H_0$  ( $\ell(\beta_{1,0}, \hat{\beta}_{2,0}; \mathbf{y})$ ), ou, em termos de *deviance*, a comparação de  $D(\mathbf{y}; \hat{\mu})$  e  $D(\mathbf{y}; \hat{\mu}_0)$  em que  $\hat{\mu}_0 = \mathbf{g}^{-1}(\hat{\eta}_0)$  e  $\hat{\eta}_0 = \mathbf{X}\hat{\beta}_0$ . Esse teste é, geralmente, preferido no caso de hipóteses relativas a vários coeficientes  $\beta$ 's. Se as diferenças são grandes, então,  $H_0$  é rejeitada. A estatística para esse teste é dada por:

$$\Lambda = -2 \ln \lambda = 2[\ell(\hat{\beta}_1, \hat{\beta}_2; \mathbf{y}) - \ell(\beta_{1,0}, \hat{\beta}_{2,0}; \mathbf{y})] = \frac{1}{\phi} [D(\mathbf{y}; \hat{\mu}_0) - D(\mathbf{y}; \hat{\mu})].$$

Para amostras grandes, rejeita-se  $H_0$ , a um nível de  $100\alpha\%$  de probabilidade, se  $\Lambda > \chi_{q,1-\alpha}^2$ .

### 2.10.2 Teste de Wald

É baseado na distribuição normal assintótica de  $\hat{\beta}$  e é uma generalização da estatística *t* de Student (Wald, 1943). É, geralmente, o mais usado no caso de hipóteses relativas a um único coeficiente  $\beta_j$ . Tem como vantagem, em relação ao teste da razão de verossimilhanças, o fato de não haver necessidade de se calcular  $\hat{\beta}_{2,0}$ . Já foi visto que, assintoticamente:

$$\hat{\beta} \sim N_p(\beta, \Sigma^{-1}).$$

Assim, a estatística para esse teste é

$$W = (\hat{\beta}_1 - \beta_{1,0})^T [\hat{\text{Var}}(\hat{\beta}_1)]^{-1} (\hat{\beta}_1 - \beta_{1,0}),$$

sendo  $\hat{\text{Var}}(\hat{\beta}_1)$  a  $\text{Var}(\hat{\beta}_1)$  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 probabilidade, se  $W > \chi_{q,1-\alpha}^2$ .

### 2.10.3 Teste escore

Obtido a partir da função escore, tem sido muito usado na Bioestatística. A estatística de teste nesse caso é dada por:

$$E = \mathbf{U}_1^T(\hat{\beta}_0) \hat{\text{Var}}_0(\hat{\beta}_1) \mathbf{U}_1(\hat{\beta}_0),$$

sendo  $\hat{\text{Var}}_0(\hat{\beta}_1)$  a  $\text{Var}(\hat{\beta}_1)$  avaliada em  $\hat{\beta}_0 = [\beta_{1,0}^T \hat{\beta}_{2,0}^T]^T$ . Para amostras grandes, rejeita-se  $H_0$ , a um nível de  $100\alpha\%$  de probabilidade, se  $E > \chi_{q,1-\alpha}^2$ .

**Caso particular:** No caso em que há interesse no teste de hipótese do vetor  $\boldsymbol{\beta}$  como um todo, isto é, no teste das hipóteses

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \text{ versus } H_a : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$$

o vetor  $\boldsymbol{\beta}_2$  desaparece e  $\boldsymbol{\beta}_1 = \boldsymbol{\beta}$  ( $q$  passa a ser igual a  $p$ ), e têm-se as expressões:

$$\begin{aligned} \text{i) teste da razão de verossimilhanças: } \Lambda &= -2 \ln \lambda = 2[\ell(\hat{\boldsymbol{\beta}}; \mathbf{y}) - \ell(\boldsymbol{\beta}_0; \mathbf{y})] \\ &= \frac{1}{\phi} [D(\mathbf{y}; \boldsymbol{\mu}_0) - D(\mathbf{y}; \hat{\boldsymbol{\mu}})]; \end{aligned}$$

$$\text{ii) teste de Wald: } W = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \hat{\mathfrak{I}} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0),$$

sendo  $\hat{\mathfrak{I}}$  a matriz de informação de Fisher avaliada em  $\hat{\boldsymbol{\beta}}$ ;

$$\text{iii) teste escore: } E = \mathbf{U}^T (\boldsymbol{\beta}_0) \mathfrak{I}_0^{-1} \mathbf{U}(\boldsymbol{\beta}_0),$$

sendo  $\mathfrak{I}_0$  a matriz de informação de Fisher avaliada em  $\boldsymbol{\beta}_0$ .

**Exemplo 16:** Conforme já foi visto, a função escore em relação a  $\beta_j$  é dada por:

$$U_j = \frac{\partial \ell}{\partial \beta_j}, \quad j = 1, 2, \dots, p,$$

sendo  $\ell$  o logaritmo da função de verossimilhança e  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$  o vetor de parâmetros do componente linear do modelo.

Além disso,  $E(\mathbf{U}) = \mathbf{0}$  e  $Cov(\mathbf{U}) = E(\mathbf{U}\mathbf{U}^T) = \mathfrak{I}$ , sendo  $\mathbf{U} = (U_1, U_2, \dots, U_p)^T$  e  $\mathfrak{I}$  a matriz de informação de Fisher. Então, pelo Teorema Central do Limite a distribuição assintótica de  $\mathbf{U} = \mathbf{U}(\boldsymbol{\beta})$  é normal  $p$ -dimensional, isto é,  $N_p(\mathbf{0}, \mathfrak{I})$  e, portanto, para amostras grandes a estatística escore

$$E = \mathbf{U}^T \mathfrak{I}^{-1} \mathbf{U}$$

converge assintoticamente para uma distribuição  $\chi_p^2$ , desde que  $\mathfrak{I}$  seja não-singular e supondo o modelo com os parâmetros especificados, verdadeiro. (Cordeiro, 1986; Dobson, 1990).

**Exemplo 17:** Seja  $Y_1, Y_2, \dots, Y_n$  uma amostra aleatória de uma distribuição  $N(\mu, \sigma^2)$  com  $\mu$  desconhecido e  $\sigma^2$  conhecido. Visto como modelo linear generalizado tem-se:

- i) somente um parâmetro de interesse,  $\mu$ ;
- ii) não há variáveis explanatórias e,
- iii) a função de ligação é a identidade, isto é,  $\eta = \mu$ .

O logaritmo da função de verossimilhança é

$$\ell = \ell(\mu; y_1, \dots, y_n) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 - \frac{n}{2} \ln(2\pi\sigma^2),$$

a partir da 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

$$\mathfrak{I} = \text{Var}(U) = \frac{n^2}{(\sigma^2)^2} \text{Var}(\bar{Y}) = \frac{n^2}{(\sigma^2)^2} \frac{\sigma^2}{n} = \frac{n}{\sigma^2}.$$

Portanto,

$$E = U^T \mathfrak{I}^{-1} U = \frac{n^2 (\bar{Y} - \mu)^2}{(\sigma^2)^2} \frac{\sigma^2}{n} = \frac{(\bar{Y} - \mu)^2}{\frac{\sigma^2}{n}} \sim \chi_1^2.$$

Esse resultado pode ser usado para a obtenção de intervalos de confiança para  $\mu$ .

**Exemplo 18:** Seja  $Y \sim \text{Bin}(m, \pi)$ . Então, conforme já foi visto,

$$\ell(\pi; y) = \ln \binom{m}{y} + y \ln \pi + (m-y) \ln(1-\pi)$$

e, portanto,

$$U = \frac{d\ell}{d\pi} = \frac{y}{\pi} - \frac{(m-y)}{1-\pi} = \frac{y - m\pi}{\pi(1-\pi)}.$$

Mas,  $E(Y) = \mu = m\pi$  e  $\text{Var}(Y) = m\pi(1-\pi) = \frac{1}{m} \mu(m - \mu)$ . Logo,

$$E(U) = 0$$

e

$$\mathfrak{I} = \text{Var}(U) = \frac{1}{\pi^2(1-\pi)^2} \text{Var}(Y) = \frac{m}{\pi(1-\pi)}.$$

Logo,

$$E = U^T \mathfrak{I}^{-1} U = \frac{(Y - m\pi)^2}{\pi^2(1-\pi)^2} \frac{\pi(1-\pi)}{m} = \frac{(Y - m\pi)^2}{m\pi(1-\pi)} = \frac{[Y - E(Y)]^2}{\text{Var}(Y)}$$

que, pelo teorema central do limite, tem distribuição  $\chi_1^2$ , ou, equivalentemente

$$\frac{\sqrt{m}(Y - \mu)}{\sqrt{\mu(m - \mu)}} = \frac{Y - E(Y)}{\sqrt{Var(Y)}} \xrightarrow{D} N(0,1).$$

Esse resultado pode ser usado para se fazer inferência a respeito de  $\mu$ .

## 2.11 Intervalos de Confiança

Intervalos de confiança assintóticos para  $\beta_1$  podem ser construídos, usando-se qualquer uma das três estatísticas de teste. A partir da estatística de teste 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:

$$2[\ell(\hat{\beta}_1, \hat{\beta}_2; \mathbf{y}) - \ell(\beta_1, \hat{\beta}_{2,1}; \mathbf{y})] < \chi_{q,1-\alpha}^2,$$

sendo  $\hat{\beta}_{2,1}$  a estimativa de máxima verossimilhança de  $\beta_2$  para cada valor de  $\beta_1$  que é testado ser pertencente, ou não, ao intervalo.

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:

$$(\hat{\beta}_1 - \beta_1)^T [\hat{Var}(\hat{\beta}_1)]^{-1} (\hat{\beta}_1 - \beta_1) < \chi_{q,1-\alpha}^2.$$

## 2.12 Exercícios

**2.12.1** Obter o algoritmo de estimação dado em (2.7) para os  $\beta$ 's para o caso particular das funções de ligação canônicas para as distribuições estudadas no Capítulo 1, calculando  $\mathbf{W}$ ,  $\Delta$  e  $\mathbf{z}$ .

**2.12.2** Considere os dados do **Exemplo 6** e estime a concentração de vírus  $\lambda$  na solução original. Use o algoritmo de estimação dado em (2.7) e os recursos do GLIM ou SAS.

**2.12.3** Definir o algoritmo de estimação dado em (2.7), calculando  $\mathbf{W}$ ,  $\Delta$  e  $\mathbf{z}$  para os modelos com erros Normal, gama, Normal inverso e Poisson com ligação potência  $\eta = \mu^\lambda$ ,  $\lambda$  conhecido. (Cordeiro, 1986).

**2.12.4** Para dados com distribuição binomial as funções de ligação mais comuns são: logística, Probit e complemento log-log. Comparar os valores do preditor linear para essas funções de ligação no intervalo (0,1).

**2.12.5** Mostre que

$$\lim_{\lambda \rightarrow 0} \frac{\mu^\lambda - 1}{\lambda} = \ln \mu.$$

**2.12.6** Considere a família de ligações dada por Aranda-Ordaz (1981)

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

Mostre que o modelo logístico é dado para  $\lambda = 1$  e que para  $\lambda \rightarrow 0$  tem-se a função de ligação complemento log-log.

**2.12.7** Obtenha a expressão da estatística  $X^2$  de Pearson para as distribuições estudadas no Capítulo 1.

**2.12.8** Obtenha a expressão da *deviance* para as distribuições estudadas no Capítulo 1.

**2.12.9** Mostre que para os modelos log-lineares com matriz modelo tendo uma coluna de 1's, a *deviance* reduz-se a:

$$S_p = 2 \sum_{i=1}^n y_i \ln \left( \frac{y_i}{\hat{\mu}_i} \right) \quad (\text{Cordeiro, 1986, pág. 79}).$$

**2.12.10** Mostre que para o modelo gama com ligação potência  $\eta = \mu^\lambda$  ou  $\eta = \ln \mu$ , neste último caso a matriz  $\mathbf{X}$  tendo uma coluna de 1's, a *deviance* reduz-se a:

$$S_p = 2 \sum_{i=1}^n v_i \ln \left( \frac{\hat{\mu}_i}{y_i} \right) \quad (\text{Cordeiro, 1986, pág. 80}).$$

**2.12.11** Os últimos dois exercícios são casos particulares do resultado mais geral

$$2 \sum_{i=1}^n \frac{1}{a_i(\phi)} (y_i - \hat{\mu}_i) \hat{\mu}_i V^{-1}(\mu_i) = 0$$

quando o modelo tem ligação  $\eta = \mu^\lambda$  ( $\lambda \neq 0$ ) ou  $\eta = \ln \mu$ , neste último caso,  $\mathbf{X}$  com uma coluna de 1's. Mostre esse resultado (Nelder & Wedderburn, 1972; Cordeiro, 1986, pág. 80).

**2.12.12** Comente o resultado obtido para a função *deviance* para dados binários com ligação canônica. Use como referência bibliográfica Williams (1983).

**2.12.13** Mostre que para o modelo gama simples, em que todas as médias são iguais, a deviance reduz-se à estatística clássica

$$S_1 = 2n v \ln\left(\frac{\bar{y}}{\tilde{y}}\right),$$

sendo  $\bar{y}$  e  $\tilde{y}$ , respectivamente, as médias aritmética e geométrica dos dados  $y_1, y_2, \dots, y_n$  (Cordeiro, 1986, pág. 80).

**2.12.14** Mostre que

$$E\left(\frac{\partial^2 \ell}{\partial \phi \partial \beta_j}\right) = 0$$

e que, portanto, os parâmetros  $\phi$  e  $\beta$  são ortogonais.

**2.12.15** Mostre que a estimativa de máxima verossimilhança do parâmetro de dispersão  $\phi$  é dada por:

$$\text{a) } \hat{\phi} = \frac{D_p}{n} \quad (\text{Normal e Normal Inversa})$$

$$\text{b) } \frac{1}{\hat{\phi}} = \frac{n}{2D_p} \left[ 1 + \sqrt{1 + \frac{2D_p}{3n}} \right] \quad (\text{Gama, apenas aproximada}).$$

**2.12.16** Considere uma única resposta  $Y \sim \text{Bin}(m, \pi)$ . Pede-se:

- obtenha a expressão para a estatística de Wald  $(\hat{\pi} - \pi)^T \mathfrak{I}(\hat{\pi} - \pi)$ , em que  $\hat{\pi}$  é a estimativa de máxima verossimilhança de  $\pi$  e  $\mathfrak{I}$  é a matriz de informação de Fisher;
- obtenha a expressão para a estatística score  $U^T \mathfrak{I}^{-1} U$  e verifique que é igual à estatística de Wald;
- obtenha a expressão para a estatística da razão de verossimilhanças  $\Lambda = 2[\ell(\hat{\mu}; y) - \ell(\mu; y)]$ ;
- para amostras grandes as estatísticas score, de Wald e da razão de verossimilhanças têm distribuição assintótica  $\chi_1^2$ . Sejam  $m = 10$  e  $y = 3$  e usando i)  $\pi = 0,1$ ; ii)  $\pi = 0,3$  e iii)  $\pi = 0,5$ , compare-as. Quais as conclusões obtidas? (Dobson, 1990).

**2.12.17** Seja  $Y_1, Y_2, \dots, Y_n$  uma amostra aleatória de uma distribuição exponencial de média  $\mu$ . Sejam as hipóteses

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_a : \mu \neq \mu_0.$$

Mostre que:

$$\text{a) LR} = -2n \left[ \ln \left( \frac{\bar{y}}{\mu_0} \right) + \frac{\mu_0 - \bar{y}}{\mu_0} \right] \quad (\text{teste da razão de verossimilhanças});$$

$$\text{b) } W = \frac{n(\bar{y} - \mu_0)^2}{\bar{y}^2} \quad (\text{teste de Wald});$$

$$\text{c) } E = \frac{n(\bar{y} - \mu_0)^2}{\mu_0^2} \quad (\text{teste escore}).$$

**2.12.18** Sejam  $Y_1, Y_2, \dots, Y_n$  variáveis independentes com distribuição de Poisson com média  $\mu_i = \mu \rho^{i-1}$  ( $i = 1, \dots, n$ ). Obter as estatísticas escore, de Wald e da razão de verossimilhanças para o teste das hipóteses que se seguem:

$$\text{a) } H_0 : \mu = \mu_0 \quad \text{vs} \quad H_a : \mu \neq \mu_0, \text{ quando } \rho \text{ é conhecido};$$

$$\text{b) } H_0 : \rho = \rho_0 \quad \text{vs} \quad H_a : \rho \neq \rho_0, \text{ quando } \mu \text{ é conhecido. (Cordeiro, 1986, pág. 57).}$$

**2.12.19** Considere a estrutura linear  $\eta_i = \beta x_i$ ,  $i = 1, \dots, n$  com um único parâmetro  $\beta$  desconhecido e ligação  $\eta = (\mu^\lambda - 1)\lambda^{-1}$ ,  $\lambda$  conhecido. Obter o estimador de máxima verossimilhança para  $\beta$ , considerando-se os modelos normal, Poisson, gama, normal inverso e binomial negativo. Fazer o mesmo para o modelo binomial com ligação  $\eta = \ln\{[(1-\mu)^{-\lambda} - 1]\lambda^{-1}\}$ ,  $\lambda$  conhecido. Obter, ainda as estimativas no caso de  $x_1 = x_2 = \dots = x_n$ . (Cordeiro, 1986, pág. 36).

**2.12.20** No exercício anterior, considere o teste de:

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \text{ versus } H_a : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0,$$

sendo  $\boldsymbol{\beta}_0$  um valor especificado para o parâmetro desconhecido. Obtenha:

- a) a variância assintótica para  $\hat{\boldsymbol{\beta}}$ ;
- 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  $\boldsymbol{\beta}$ ;
- d) um intervalo de confiança, com um coeficiente de confiança de  $100(1-\alpha)\%$ , para uma função  $g(\boldsymbol{\beta})$  com  $g(\cdot)$  conhecido. (Cordeiro, 1986, pág. 56).

**2.12.21** Considere o **Exemplo 5** e obtenha os intervalos de confiança, com um coeficiente de confiança de 95%, para  $\boldsymbol{\beta}$ , usando os testes de Wald e o da razão de verossimilhanças.

## Capítulo 3

# Técnicas para Verificação de Ajuste do Modelo

### 3.1 Introdução

A escolha de um modelo linear generalizado envolve três passos:

- 1<sup>o</sup>) definição da distribuição (que determina a função de variância);
- 2<sup>o</sup>) definição da função de ligação;
- 3<sup>o</sup>) definição da matriz do modelo.

Na prática, porém, pode acontecer que após uma escolha cuidadosa de um modelo e subsequente ajuste a um conjunto de dados o resultado obtido seja insatisfatório. Isso pode ocorrer em função de algum desvio sistemático entre valores observados e valores ajustados ou, então, porque um ou mais valores são discrepantes em relação aos demais.

Desvios sistemáticos podem ser provocados pela escolha inadequada da função de variância, da função de ligação e da matriz do modelo, ou ainda pela definição errada da escala da variável dependente ou das covariáveis.

Discrepâncias isoladas podem ocorrer ou porque os pontos estão nos extremos da amplitude de validade da covariável, ou porque eles estão realmente errados como resultado de uma leitura errada ou uma transcrição mal feita, ou ainda por que algum fator não controlado influenciou na sua obtenção.

Na prática, em geral, ocorre que há uma interação dos diferentes tipos de falhas. Assim, por exemplo, a detecção de uma escolha errada da função de ligação pode ocorrer porque ela está realmente errada ou porque uma ou mais covariáveis estão na escala errada ou devido à presença de alguns pontos discrepantes. Isso faz com que a verificação da adequação de um modelo para um determinado conjunto de dados seja um processo realmente difícil.

Maiores detalhes podem ser vistos em Atkinson (1985), Atkinson *et al.* (1989), Cordeiro (1986), McCullagh & Nelder (1989), Francis, Green & Payne (1993), Paula (2000).

### 3.2 Técnicas para a verificação do ajuste de um modelo a um conjunto de dados

As técnicas usadas para esse fim podem ser formais ou informais.

As informais baseiam-se em exames visuais de gráficos para a detecção de padrões, ou então, de pontos discrepantes.

As formais envolvem aninhar o modelo sob pesquisa em uma classe maior 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 a:

- inclusão de uma covariável adicional;
- aninhamento de uma covariável  $x$  em uma família  $h(x; \gamma)$  indexada por um parâmetro  $\gamma$ , sendo um exemplo a família de Box-Cox;
- aninhamento de uma função de ligação  $g(\mu)$  em uma família maior  $g(\mu; \gamma)$ , sendo um exemplo a família de Aranda-Ordaz (Exercício 2.12.5)
- 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* tomando 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, a fazer a análise com a observação discrepante e sem ela e verificar se a mudança no valor da *deviance* é significativa, ou não. Ambos, porém, dependem da localização do(s) ponto(s) discrepante(s).

### 3.3 Análise de resíduos e diagnósticos para modelos lineares clássicos

#### 3.3.1 Introdução

No modelo linear  $\mathbf{Y} = \mathbf{X}\beta + \varepsilon = \mu + \varepsilon$  os elementos  $\varepsilon_i$  do vetor  $\varepsilon$  são as diferenças entre os valores observados dos  $Y_i$ 's e aqueles esperados pelo modelo ( $\mu_i$ 's). Esses elementos são chamados de erros e admite-se que os  $\varepsilon_i$ 's são independentes e, além disso,  $\varepsilon_i \sim N(0, \sigma^2)$ . Esses termos representam a variação natural dos dados, mas podem, também, 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 levar a resultados duvidosos. Este tipo de violação do modelo dá origem às chamadas **falhas 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. 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, modelo inadequado etc).
- escala usada errada, talvez os dados sejam melhor descritos após uma transformação, do tipo logarítmica ou raiz quadrada;
- 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.

Dado um conjunto de observações e ajustado um determinado modelo a ele, para a verificação das pressuposições devem ser considerados como material básico:

- os valores estimados (ou ajustados)  $\hat{\mu}_i$ ;
- os resíduos  $r_i = y_i - \hat{\mu}_i$ ;
- a variância residual estimada,  $\hat{\sigma}^2 = s^2 = \text{QMRes}$ ;
- os elementos da diagonal (*leverage*) da matriz de projeção  $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ , isto é,

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

sendo  $\mathbf{x}_i^T = [x_{i1} \ x_{i2} \ \dots \ x_{ik}]$ .

Uma idéia 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  $i$ .

### 3.3.2 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 & Weisberg, 1982; Atkinson, 1985; Miazaki & Stangenhuis, 1994).

- a) **Resíduos ordinários:** Os resíduos do processo de ajustamento por mínimos quadrados são dados por:

$$r_i = \hat{\varepsilon}_i = y_i - \hat{\mu}_i.$$

Enquanto os erros  $\varepsilon_i$ 's são independentes e com a mesma variância, o mesmo não ocorre com os resíduos do ajuste do modelo através de mínimos quadrados, isto é,

$$\text{Var}(\hat{\varepsilon}) = \text{Var}[(\mathbf{I} - \mathbf{H})\mathbf{Y}] = (\mathbf{I} - \mathbf{H}) \sigma^2.$$

Em particular, a variância do  $i$ -ésimo resíduo é dada por:

$$\text{Var}(r_i) = \text{Var}(\hat{\varepsilon}_i) = (1 - h_{ii}) \sigma^2$$

e a covariância dos resíduos dos  $i$ -ésimo e  $i'$ -ésimo resíduos é dada por:

$$\text{Cov}(r_i, r_{i'}) = \text{Cov}(\hat{\varepsilon}_i, \hat{\varepsilon}_{i'}) = -h_{ii'} \sigma^2.$$

Assim, usar  $r_i = \hat{\varepsilon}_i$  pode não ser adequado devido à heterogeneidade de variâncias. Foram, então, propostas diferentes padronizações para sanar esse problema.

- b) **Resíduos estudentizados internamente (*Studentized residual*)**: Considerando-se  $s^2 = \text{QMRes}$  como a estimativa de  $\sigma^2$ , tem-se que um estimador não tendencioso para  $\text{Var}(\hat{\varepsilon}_i)$  é dado por:

$$\widehat{\text{Var}}(\hat{\varepsilon}_i) = (1 - h_i)s^2 = (1 - h_i)\text{QMRes}$$

e, como  $E(r_i) = E(\hat{\varepsilon}_i) = E(Y_i - \hat{\mu}_i) = 0$ , então o resíduo estudentizado internamente é dado por:

$$rsi_i = \frac{\hat{\varepsilon}_i}{\sqrt{\widehat{\text{Var}}(\hat{\varepsilon}_i)}} = \frac{r_i}{\sqrt{(1 - h_i)s^2}} = \frac{y_i - \hat{u}_i}{\sqrt{(1 - h_i)\text{QMRes}}}$$

Esses resíduos são mais sensíveis 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, numerador e denominador são variáveis dependentes, isto é,  $\text{Cov}(\hat{\varepsilon}_i, \text{QMRes}) \neq 0$ .

- c) **Resíduos padronizados externamente (*jackknifed residuals, deletion residuals, externally studentized residual, RStudent*)**: Para garantir a independência do numerador e denominador na padronização dos resíduos, define-se o resíduo estudentizado externamente, como:

$$rse_i = \frac{r_i}{s_{(i)}\sqrt{(1 - h_i)}},$$

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  $i$ . Prova-se que:

$$rse_i = rsi_i \sqrt{\frac{n - p - 1}{n - p - rsi_i^2}},$$

sendo  $p$  o número de parâmetros independentes.

A vantagem de usar  $rse_i$  é que, sob normalidade, ele 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 seja merecedora de atenção especial se  $|rse_i|$  for maior do que o  $100\left(1 - \frac{\alpha}{2n}\right)$ -ésimo percentil da distribuição  $t$  com  $(n - p - 1)$  graus de liberdade, sendo que  $\alpha$ , o nível de confiança, é dividido por  $n$  por ser este o número de pontos sob análise.

### 3.3.3 Estatísticas para diagnósticos

Discrepâncias isoladas (pontos atípicos) podem ser caracterizadas por ter  $h$  e/ou resíduo grandes, ser inconsistente e/ou ser influente (ver pág 404, McCullagh & Nelder, 1989). Em geral, pode-se classificar uma observação como:

- i) **ponto de alavanca (bom ou ruim):**  $h$  alto;
- ii) **inconsistente:** o ponto não segue a tendência dos dados;
- iii) **outlier:**  $h$  baixo e resíduo grande;
- iv) **influyente:** afeta, de forma significativa, o ajuste do modelo.

Assim, uma observação influente é aquela cuja omissão do conjunto de dados resulta em mudanças substanciais em certos aspectos do modelo. Ela pode ser um *outlier*, ou não. Uma observação pode ser influente de diversas maneiras, isto é,

- no ajuste geral do modelo;
- no conjunto de estimativas dos parâmetros;
- na estimativa de um determinado parâmetro;
- na escolha de uma transformação de uma variável explanatória.

As estatísticas mais utilizadas para a verificação de pontos atípicos são:

a) **Elementos da diagonal da matriz de projeção  $\mathbf{H}$  ( $h_i$ , *leverage*):** Quando uma observação está distante das outras em termos das variáveis explanatórias ela pode ser, ou não, influente. A distância de uma observação em relação às demais é medida pelo  $h$  (medida de *leverage*).

No processo de ajuste, como  $\hat{\mu} = \mathbf{H}\mathbf{Y}$ , tem-se

$$\hat{\mu}_i = \sum_{j=1}^n h_{ij} Y_j = h_{i1} Y_1 + h_{i2} Y_2 + \dots + h_{ii} Y_i + \dots + h_{in} Y_n, \quad 1 \leq i \leq n$$

sendo que  $h_i = h_{ii} = \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i$  dependente da matriz do delineamento. Mas, como  $\mathbf{H}$  é uma matriz de projeção, verifica-se que

$$h_i = h_{ii} = \sum_{i'=1}^n h_{ii'}^2 = h_{ii}^2 + \sum_{i' \neq i} h_{ii'}^2$$

e, além disso,  $0 \leq h_{ii} \leq 1$ ,  $\sum_{i'=1}^n h_{ii'} = 1$  e  $r(\mathbf{H}) = \text{tr}(\mathbf{H}) = \sum_{i=1}^n h_{ii} = p$ .

Vê-se, portanto, que o valor ajustado  $\hat{\mu}_i$  é a média ponderada dos valores observados e que o peso de ponderação é o valor de  $h_{ii}$ . 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 com  $h_{ii} > \frac{2p}{n}$ , segundo Belsley *et al.* (1980, pág. 17) indicam observações que merecem uma análise mais apurada.

b) **DFBeta:** É importante quando o coeficiente de regressão tem um significado prático. Mede a alteração em  $\hat{\beta}$  ao se retirar o  $i$ -ésimo ponto da análise. É dado por:

$$\text{DFBeta}_{(i)} = \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)} = \frac{1}{(1-h_i)} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}'_i \hat{\boldsymbol{\epsilon}}_i.$$

Não tem interpretação simples. Cook & Weisberg (1982) propuseram curvas empíricas para o estudo dessa medida.

c) **DFFitS** - Mede a alteração provocada no valor ajustado pela retirada da observação  $i$ .

$$\text{DFFitS}_{(i)} = \frac{\text{DFFit}_{(i)}}{\sqrt{h_i s_{(i)}^2}} = \frac{\mathbf{x}_i (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})}{\sqrt{h_i s_{(i)}^2}} = \frac{\hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\mu}}_{(i)}}{\sqrt{h_i s_{(i)}^2}} = \frac{1}{\sqrt{h_i s_{(i)}^2} (1-h_i)} \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i r_i$$

ou, ainda,

$$\text{DFFitS}_{(i)} = \left( \frac{h_i}{1-h_i} \right)^{\frac{1}{2}} \frac{r_i}{s_{(i)} (1-h_i)^{\frac{1}{2}}} = \left( \frac{h_i}{1-h_i} \right)^{\frac{1}{2}} rse_i$$

sendo o quociente  $\frac{h_i}{1-h_i}$  chamado potencial de influência e é uma medida da distância do ponto  $x$  em relação às demais observações. Belsley *et al.* (1980, pág. 28) sugerem que valores absolutos excedendo  $2\sqrt{\frac{p}{n}}$  podem identificar observações influentes.

d) **Distância de Cook**: É também uma medida de afastamento do vetor de estimativas provocado pela retirada da observação  $i$ . É uma expressão muito semelhante ao DFFitS mas que usa como estimativa da variância residual aquela obtida com todas as  $n$  observações, ou ainda, usa o resíduo estudentizado internamente. É dada por:

$$D_{(i)} = \frac{(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})^T (\mathbf{X}^T \mathbf{X}) (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})}{ps^2} = \frac{h_i}{(1-h_i)^2} \frac{r_i^2}{ps^2} = \left( \frac{r_i}{(1-h_i)^{\frac{1}{2}} s} \right)^2 \frac{h_i}{(1-h_i)} \frac{1}{p}$$

ou, ainda,

$$D_{(i)} = \frac{h_i}{(1-h_i)} \frac{1}{p} rsi_i^2.$$

e) **Distância de Cook modificada**: Atkinson (1981) sugeriu uma modificação para a distância de Cook

$$C_i = \left( \frac{n-p}{p} \frac{h_i}{1-h_i} \right)^{\frac{1}{2}} |rse_i|.$$

### 3.3.4 Tipos de gráficos

a) **Valores observados ( $y$ ) vs variáveis explanatórias ( $x_j$ )**: Esse tipo de gráfico indica a estrutura que pode existir entre a variável dependente e as diversas covariáveis. Pode indicar, também, a presença de heterocedasticidade. Pode, porém, levar a uma idéia falsa no

caso de muitas covariáveis (a não ser que haja ortogonalidade entre todas). O padrão nulo deste gráfico é uma distribuição dos resíduos em torno de zero com amplitude constante.

b) **Variável explanatória  $x_j$  vs variável explanatória  $x_j$** : Esse tipo de gráfico indica a estrutura que pode existir entre duas variáveis explanatórias. Pode indicar, também, a presença de heterocedasticidade. Pode, porém, levar a uma idéia falsa no caso de muitas covariáveis (a não ser que haja ortogonalidade entre todas). O padrão nulo deste gráfico é uma distribuição dos resíduos em torno de zero com amplitude constante.

c) **Resíduos vs variáveis explanatórias não incluídas ( $x_{\text{fora}}$ )**: Pode mostrar se existe uma relação entre os resíduos do modelo ajustado e uma variável ainda não incluída no modelo. Pode mostrar, também, a evidência de heterocedasticidade. Pode levar, porém, ao mesmo tipo de problema apontado nos itens (a) e (b). Uma alternativa melhor para esse tipo de gráfico é o gráfico da variável adicionada (*added variable plot*). O padrão nulo deste gráfico é uma distribuição dos resíduos em torno de zero com amplitude constante.

d) **Resíduos vs variáveis explanatórias incluídas ( $x_{\text{dentro}}$ )**: Pode mostrar se ainda existe uma relação sistemática entre os resíduos e a variável  $x_j$  já incluída, isto é, por exemplo se  $x_{\text{dentro}}^2$  deve ser incluída. Esse tipo de gráfico apresenta o mesmo tipo de problema que o citado nos itens (a), (b) e (c). Alternativa melhor para isso é o gráfico de resíduos parciais (*partial residual plot*). O padrão para esse tipo de gráfico é uma distribuição aleatória de média 0 e amplitude constante. Desvios sistemáticos podem indicar:

- escolha errada da variável explanatória;
- termo quadrático (ou ordem superior) faltando;
- escala errada da variável explanatória.

e) **Resíduos vs valores ajustados**: O padrão para esse tipo de gráfico é uma distribuição aleatória de média 0 e amplitude constante.

f) **Gráficos de índices**: servem para localizar observações com resíduo, *leverage* ( $h$ ), distância de Cook modificada etc, grandes.

g) **Gráfico da variável adicionada ou da regressão parcial (*added variable plot*)**: Embora os gráficos de resíduos vs variáveis não incluídas no modelo possam indicar a necessidade de variáveis extras no modelo, a interpretação exata deles não é clara. A dificuldade está em que, a menos que a variável explanatória, considerada para inclusão, seja ortogonal a todas as variáveis que já estão no modelo, o coeficiente angular do gráfico de 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 com uma variável explanatória e como isto é influenciado por observações individuais. No caso do modelo linear geral, tem-se

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma} \mathbf{u}$$

sendo  $\mathbf{u}$  a variável a ser adicionada (pode ser uma variável construída). O interesse está em se saber se  $\boldsymbol{\gamma} = 0$ , isto é, se não há necessidade de se incluir a variável  $\mathbf{u}$  no modelo. A partir do sistema de equações normais

$$\begin{bmatrix} \mathbf{X}^T \\ \mathbf{u}^T \end{bmatrix} \begin{bmatrix} \mathbf{X} & \mathbf{u} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{Y} \\ \mathbf{u}^T \mathbf{Y} \end{bmatrix} \Rightarrow \begin{cases} \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{X}^T \mathbf{u} \hat{\gamma} = \mathbf{X}^T \mathbf{Y} \\ \mathbf{u}^T \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{u}^T \mathbf{u} \hat{\gamma} = \mathbf{u}^T \mathbf{Y} \end{cases}$$

tem-se

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

e

$$\hat{\gamma} = \frac{\mathbf{u}^T (\mathbf{I} - \mathbf{H}) \mathbf{Y}}{\mathbf{u}^T (\mathbf{I} - \mathbf{H}) \mathbf{u}} = \frac{\mathbf{u}^{*T} \mathbf{r}}{\mathbf{u}^{*T} \mathbf{u}^*}$$

que é o coeficiente angular de uma reta que passa pela origem e em que  $\mathbf{r} = \mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}} = (\mathbf{I} - \mathbf{H}) \mathbf{Y}$  são os resíduos de  $\mathbf{Y}$  ajustado para  $\mathbf{X}$  e  $\mathbf{u}^* = (\mathbf{I} - \mathbf{H}) \mathbf{u}$  são os resíduos de  $\mathbf{u}$  ajustado para  $\mathbf{X}$ .

O gráfico da variável adicionada (*added variable plot*) de  $\mathbf{r}$  versus  $\mathbf{u}^*$  tem coeficiente angular  $\hat{\gamma}$  (diferente de  $\mathbf{r}$  vs  $\mathbf{u}$ ). Ele pode mostrar, também, como a evidência para a inclusão de  $\mathbf{u}$  depende de observações individuais. Esse gráfico, portanto, é obtido a partir dos resíduos ordinários da regressão de  $\mathbf{Y}$  como função de todas as covariáveis exceto  $\mathbf{u} = \mathbf{x}_j$  versus os resíduos ordinários da regressão de  $\mathbf{u} = \mathbf{x}_j$  como função das mesmas covariáveis usadas para modelar  $\mathbf{Y}$ . Assim, por exemplo, para um modelo com 3 covariáveis, o gráfico da variável adicionada para  $\mathbf{x}_3$  é obtido a partir de

$$\hat{\boldsymbol{\mu}} = \hat{\beta}_0 + \hat{\beta}_1 \mathbf{x}_1 + \hat{\beta}_2 \mathbf{x}_2 \Rightarrow \mathbf{r} = \mathbf{Y} - \hat{\boldsymbol{\mu}}$$

e

$$\hat{\mathbf{x}}_3 = \hat{\beta}_0 + \hat{\beta}_1 \mathbf{x}_1 + \hat{\beta}_2 \mathbf{x}_2 \Rightarrow \mathbf{u}^* = \mathbf{x}_3 - \hat{\mathbf{x}}_3.$$

**h) Gráfico de resíduos parciais ou gráfico de resíduos mais componente (*partial residual plot*):** Se o interesse está em se detectar uma estrutura omitida, tal como uma forma diferente de dependência em  $\mathbf{u}$ , um gráfico usando  $\mathbf{u}$  pode servir melhor. Esse gráfico também, tem coeficiente angular  $\hat{\gamma}$ . Consiste em se plotarem os resíduos do modelo  $E(\mathbf{Y}) = \mathbf{X} \boldsymbol{\beta} + \gamma \mathbf{u}$  mais  $\hat{\gamma} \mathbf{u}$  versus  $\mathbf{u}$ , isto é, no gráfico de  $\tilde{\mathbf{r}} = \mathbf{r} + \hat{\gamma} \mathbf{u}$  versus  $\mathbf{u}$ . Por isso ele, também, é chamado de gráfico do resíduo mais componente.

**i) Gráfico normal e semi-normal de probabilidades (*normal plots e half normal plots*):** Segundo Weisberg (1985) o gráfico normal de probabilidades destaca-se por dois aspectos:

- 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 obtidos a partir do ajuste de um determinado modelo a essa amostra são  $d_{(1)}, d_{(2)}, \dots, d_{(n)}$ .

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  $z_i$ 's correspondentes, obtidos da distribuição normal padrão são

linearmente relacionados. Portanto, o gráfico dos valores  $d_{(i)}$  versus  $z_i$  deve ser, aproximadamente, uma reta. Formatos aproximados comuns que indicam ausência de normalidade são:

- **S (Esse)**: indica distribuições com caudas muito curtas, isto é, distribuições cujos valores estão muito próximos da média;
- **S invertido (Esse 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 os passos:

- i) 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,  $h$  etc);
- ii) dada a 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

$$p_i = \frac{i - c}{n - 2c + 1}$$

para  $0 < c < 1$ . Diversos valores têm sido propostos para a constante  $c$ . Vários autores recomendam a utilização de  $c = \frac{3}{8}$ , ficando, então,

$$z_i = \Phi^{-1}\left(\frac{i - 0,375}{n + 0,25}\right), \quad i = 1, 2, \dots, n.$$

- iii) 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.

A construção do gráfico semi-normal de probabilidades (*half normal plot*) é o resultado do conjunto de pontos obtidos por  $|d_{(i)}|$  versus  $z_i$ , sendo  $z_i = \Phi^{-1}\left(\frac{i + n - 0,125}{2n + 0,5}\right)$ .

McCullagh & Nelder (1989) sugerem o uso do gráfico normal de probabilidades (*normal plot*) para resíduos e o gráfico semi-normal de probabilidades (*half normal plot*) para medidas positivas como é o caso de  $h$  (medida de *leverage*) e da distância de Cook modificada.

No caso do gráfico normal de probabilidades para resíduos, espera-se que na ausência de pontos discrepantes, o aspecto seja linear, mas não há razão para se esperar que o mesmo aconteça quando são usados  $h$  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 (*half normal plot*), Atkinson (1985) propôs a adição de um envelope simulado. Este envelope é tal que sob o modelo correto as quantias (resíduos, *leverage*, distância de Cook etc) obtidas a partir dos dados observados caem dentro do envelope. Esse gráfico é obtido seguindo-se os passos:

i) 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, *leverage* etc);

ii) simule 19 amostras da variável resposta, usando as estimativas obtidas após um determinado modelo ter sido ajustado aos dados e os mesmos valores para as variáveis explanatórias;

iii) ajuste o mesmo modelo a cada uma das 19 amostras e calcule os valores absolutos ordenados da estatística de diagnóstico de interesse,  $d_{j(i)}^*, j = 1, \dots, 19, i = 1, \dots, n$ ;

iv) para cada  $i$ , calcule a média, o mínimo e o máximo de  $d_{j(i)}^*$ ;

v) coloque em um gráfico as quantidades obtidas no item anterior e  $d_{(i)}^*$  contra  $z_i$ .

Demétrio & Hinde (1997) apresentam um conjunto de macros que permitem fazer esses gráficos para uma grande variedade de modelos, usando o GLIM, enquanto que Paula (2000) faz o mesmo em SPlus.

j) **Valores observados 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 feitos sempre que possível. Esse tipo de gráfico pode levar à detecção de padrões não suspeitados, devido ao tempo ou, então, a alguma variável altamente correlacionada com tempo.

### 3.4 Análise de resíduos e diagnósticos para modelos lineares generalizados

#### 3.4.1 Introdução

As técnicas usadas para análise de resíduos e diagnósticos para modelos lineares generalizados são semelhantes às usadas para modelos lineares clássicos, com algumas adaptações. Assim, por exemplo, na verificação da pressuposição de linearidade para o modelo linear clássico usam-se os vetores  $\mathbf{y}$  e  $\hat{\boldsymbol{\mu}}$  enquanto que para o modelo linear generalizado devem ser usados  $\mathbf{z}$ , a variável dependente ajustada, e  $\hat{\boldsymbol{\eta}}$ , o preditor linear. A variância residual  $s^2$  é substituída por uma estimativa consistente do parâmetro  $\phi$  e a matriz  $\mathbf{H}$  torna-se:

$$\mathbf{H} = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2}$$

o que é equivalente a substituir  $\mathbf{X}$  por  $\mathbf{W}^{1/2} \mathbf{X}$ . Note-se que, agora  $\mathbf{H}$  depende das variáveis explanatórias, da função de ligação e da função de variância, tornando mais difícil a interpretação da medida de *leverage*. Pode ser mostrado que:

$$\mathbf{V}^{-1/2} (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \cong \mathbf{H} \mathbf{V}^{-1/2} (\mathbf{Y} - \boldsymbol{\mu}),$$

sendo  $\mathbf{V} = \text{diag}\{V(\mu_i)\}$ . Isso mostra que  $\mathbf{H}$  mede a influência em unidades estudentizadas de  $\mathbf{Y}$  sobre  $\hat{\boldsymbol{\mu}}$ .

### 3.4.2 Tipos de resíduos

Os tipos de resíduos mais usados para os modelos lineares generalizados são:

a) **resíduos ordinários**

$$r_i = y_i - \hat{\mu}_i;$$

b) **resíduos de Pearson generalizados**

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{\frac{\hat{\phi}}{w_i} V(\hat{\mu}_i)}},$$

sendo  $\hat{\phi}$  uma estimativa consistente do parâmetro  $\phi$  e  $w_i$  um peso *a priori* (na maior parte dos casos igual a 1);

c) **resíduos de Pearson generalizados estudentizados internamente**

$$r_i^{P'} = \frac{y_i - \hat{\mu}_i}{\sqrt{\frac{\hat{\phi}}{w_i} V(\hat{\mu}_i)(1 - h_i)}},$$

sendo  $h_i$  elemento da diagonal da matriz  $\mathbf{H}$ ;

d) **componentes da deviance**

$$r_i^D = \text{sign}(y_i - \hat{\mu}_i) \sqrt{\frac{2w_i}{\phi} [y_i(\tilde{\theta}_i - \hat{\theta}_i) + b(\tilde{\theta}_i) - b(\hat{\theta}_i)]};$$

e) **componentes da deviance estudentizado internamente** (Cox & Snell, 1968)

$$r_i^{D'} = \frac{r_i^D}{\sqrt{\hat{\phi}(1 - h_i)}};$$

f) **componentes da deviance estudentizado externamente** (*Jackknifed residuals, deletion residuals*, Williams, 1987; Pregibon, 1981).

$$r_i^{D^*} = \text{sign}(y_i - \hat{\mu}_i) \sqrt{(1 - h_i) ({}_1r_i^{D'})^2 + h_i ({}_1r_i^{P'})^2},$$

sendo que o índice 1 significa 1ª iteração.

### 3.4.3 Tipos de gráficos

São basicamente os mesmos gráficos já estudados em (3.3.4) com algumas modificações e com intepretações semelhantes.

a) **Resíduos versus alguma função dos valores ajustados:** É recomendado o gráfico de algum tipo de resíduo estudentizado versus  $\hat{\eta}_i$ , ou então, versus os valores ajustados transformados de tal forma a se ter variância constante (McCullagh & Nelder, 1989) para a distribuição em uso. Assim, usar  $\hat{\mu}_i$  para a distribuição normal,  $2\sqrt{\hat{\mu}_i}$  para a Poisson,  $2\arcseno(\hat{\mu}_i / m_i)$  para a binomial,  $2\ln(\hat{\mu}_i)$  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 ser algum tipo de curvatura ou, então, mudança sistemática da amplitude com o valor ajustado. Não tem significado para dados binários (Bernoulli).

b) **Resíduos versus variáveis explanatórias não incluídas:** 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 deste gráfico é uma distribuição dos resíduos em torno de zero com amplitude constante.

c) **Resíduos versus variáveis explanatórias já incluídas:** Pode mostrar se ainda existe uma relação sistemática entre os resíduos e uma variável já incluída. Alternativa melhor para isso é o gráfico de resíduos parciais (*partial residual plot*). O padrão nulo para esse tipo de gráfico é uma distribuição aleatória de média 0 e amplitude constante.

d) **Gráfico da variável adicionada ou da regressão parcial (*added variable plot*):** Inicialmente, ajusta-se o modelo com preditor linear  $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ . Em seguida, faz-se o gráfico de

$\mathbf{W}^{-1/2}\mathbf{s}$  versus  $(\mathbf{I} - \mathbf{H})\mathbf{W}^{1/2}\mathbf{u}$ , sendo  $\mathbf{s}$  o vetor com elementos  $s_i = \frac{y_i - \hat{\mu}_i}{a_i(\phi)V(\hat{\mu}_i)} \frac{d\mu_i}{d\eta_i}$  (Wang,

1985). Aqui  $\mathbf{W}^{-1/2}\mathbf{s}$  representa o vetor de elementos  $\frac{y_i - \hat{\mu}_i}{\sqrt{a_i(\phi)V(\hat{\mu}_i)}}$  (resíduo de Pearson

generalizado da regressão ponderada de  $\mathbf{Y}$  em relação a  $\mathbf{X}$  com matriz de pesos  $\mathbf{W}$ ) e  $(\mathbf{I} - \mathbf{H})\mathbf{W}^{1/2}\mathbf{u}$  representa os resíduos da regressão ponderada de  $\mathbf{u}$  em relação a  $\mathbf{X}$  com matriz de pesos  $\mathbf{W}$ .

e) **Gráfico de resíduos parciais ou gráfico de resíduos mais componente (*partial residual plot*):** Inicialmente ajusta-se o modelo com preditor linear  $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \gamma \mathbf{u}$ , obtendo-se  $\mathbf{W}^{-1}\mathbf{s}$  e  $\hat{\gamma}$ . Em seguida, faz-se o gráfico de  $\mathbf{W}^{-1}\mathbf{s} + \gamma \mathbf{u}$  versus  $\mathbf{u}$  (Wang, 1987).

f) **Gráficos de índices:** servem para localizar observações com resíduo, *leverage* ( $h$ ), distância de Cook modificada etc, grandes.

g) **Gráfico normal e semi-normal de probabilidades (*normal plots e half normal plots*)**: construídos da mesma forma que para os modelos lineares clássicos, usando-se, porém, a distribuição pertinente.

h) **Valores observados 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 feitos sempre que possível. Esse tipo de gráfico pode levar à detecção de padrões não suspeitados, devido ao tempo ou, então, a alguma variável altamente correlacionada com tempo.

### 3.5 Verificação da função de ligação

Um método informal para isso é o gráfico da variável dependente ajustada  $\mathbf{z}$  contra o preditor linear estimado  $\hat{\boldsymbol{\eta}}$ . O padrão nulo é uma reta. O gráfico da variável adicionada também pode ser usado, considerando-se  $\mathbf{u} = \hat{\boldsymbol{\eta}}^2$ , sendo que 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 serve para dados binários.

Existem dois métodos formais para a verificação da adequacidade da função de ligação utilizada:

1<sup>o</sup>) o mais simples consiste em se adicionar  $\hat{\boldsymbol{\eta}}^2$  como uma covariável extra e examinar a mudança ocorrida na *deviance* 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;

2<sup>o</sup>) outro método formal consiste em indexar a família de ligações por um parâmetro  $\lambda$  e fazer um teste da hipótese  $H_0 : \lambda = \lambda_0$ . Para isso podem ser usados os testes da razão de verossimilhanças e o escore. Incerteza sobre a função de ligação é mais comum com dados contínuos que têm distribuição gama e com proporções tendo erros binomiais. Assim, por exemplo, para dados 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 de Aranda-Ordaz (1981)  $\eta = \ln\{[(1 - \pi)^{-\lambda} - 1]/\lambda\}$  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 perfil de verossimilhança para se estimar  $\lambda$ . Para o modelo linear clássico esse teste equivale ao teste proposto por Tukey (1949) para não-aditividade.

A verificação da adequacidade 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  $\hat{\boldsymbol{\eta}}^2$  ao preditor linear indica desvio do modelo, isto pode indicar função de ligação errada ou escalas erradas para as variáveis explanatórias ou ambas. Pontos atípicos, também, podem afetar a escolha da função de ligação.

### 3.5.1 Justificativa para o uso de $\hat{\eta}^2$

Suponha que a função de ligação usada foi  $\eta = g(\mu)$  e que a verdadeira seja  $\eta = g^*(\mu)$ . Então,

$$g(\mu) = g\{g^{*-1}(\eta)\} = h(\eta).$$

A hipótese nula é  $H_0 : h(\eta) = \eta$  e a alternativa é  $H_a : h(\eta) = \text{não-linear}$ . Usando-se a expansão de Taylor para  $g(\mu)$  tem-se:

$$g(\mu) \cong h(0) + \eta h'(0) + \eta^2 \frac{h''(0)}{2} + \dots$$

o que mostra que a variável adicionada é  $\hat{\eta}^2$ , desde que o modelo tenha termos para o qual a média geral é marginal.

**Exemplo 19:** Considere os dados do **Exemplo 5**. A variável resposta tem distribuição binomial, isto é,  $Y_i \sim \text{Bin}(m_i, \pi_i)$ . Adotando-se a função de ligação logística (canônica) e os preditores lineares dados por:

$$\eta_i = \ln\left(\frac{\mu_i}{m_i - \mu_i}\right) = \beta_1 + \beta_2 d_i,$$

e

$$\eta_i = \ln\left(\frac{\mu_i}{m_i - \mu_i}\right) = \beta_1 + \beta_2 d_i + \gamma u_i,$$

sendo  $u_i = \hat{\eta}_i^2$ , usa-se a diferença de *deviances* para testar a adequação da função de ligação, obtendo-se os resultados da Tabela 13. Verifica-se que se rejeita a hipótese  $H_0 : \gamma = 0$ , ao nível de 5% de probabilidade, indicando que a ligação logística não é adequada. A estimativa para  $\gamma$  é  $\hat{\gamma} = -0,2087$  (0,0757).

Tabela 13: Análise de *Deviance*

Causa de variação	g.l.	<i>Deviances</i>	Valor 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	
Total	5	163,740	

$$\chi_{1;0,05}^2 = 3,84 ; \chi_{1;0,01}^2 = 6,64$$

Fazendo-se uma análise de resíduos verifica-se que a 1ª observação é discrepante. Eliminando-a e refazendo-se o teste para a função de ligação a hipótese  $H_0 : \gamma = 0$  não é rejeitada, indicando a adequação da função de ligação logística. Tem-se, então,  $\hat{\gamma} = -0,1140$  (0,086) e,

$$\hat{\eta}_i = \ln\left(\frac{\hat{\mu}_i}{m_i - \hat{\mu}_i}\right) = -3,5823 + 0,7506 d_i.$$

### 3.6 Verificação da adequação da função de variância

Um método informal para a adequação da função de variância (que é definida ao se escolher uma determinada distribuição) é o gráfico dos resíduos absolutos contra os valores ajustados transformados em uma escala com variância constante (como usado em 3.4.3, item a). O padrão nulo para esse tipo de gráfico é uma distribuição aleatória de média 0 (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 tratada como superdispersão (Hinde & Demétrio, 1998).

Um método formal para a verificação da adequação da função de variância usada consiste em indexar a função de variância por um parâmetro  $\lambda$  e fazer um teste da hipótese  $H_0 : \lambda = \lambda_0$ . Para isso podem ser usados os testes da razão de verossimilhanças e o escore. Assim, por exemplo, pode-se usar  $V(\mu) = \mu^\lambda$  e observar como o ajuste varia em função da variação de  $\lambda$ . Em geral, usa-se o método do perfil de verossimilhança para se estimar  $\lambda$ .

Para a comparação de ajustes com diferentes funções de variância, a *deviance* não pode mais ser usada, há necessidade de se usar a teoria de quase verossimilhança estendida (Nelder & Pregibon, 1987)

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 pontos atípicos.

### 3.7 Verificação da adequação das escalas das covariáveis

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  $h(\cdot; \lambda)$ . Em modelos lineares generalizados o resíduo parcial é definido por:

$$u = z - \hat{\eta} + \hat{\gamma}x,$$

sendo  $z$  a variável dependente ajustada,  $\hat{\eta}$  o preditor linear ajustado e  $\hat{\gamma}$  a estimativa do parâmetro para a variável explanatória  $x$ .

O gráfico de  $u$  contra  $x$  leva a um método informal. Se a escala de  $x$  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, tal que pode ser necessário olhar gráficos de resíduos parciais para diversos  $x$ 's.

Um método formal consiste em colocar  $x$  em uma família  $z(\cdot; \lambda)$  indexada por  $\lambda$ ; calcular, então, a *deviance* para um conjunto de valores de  $\lambda$  e determinar  $\hat{\lambda}$  como aquele valor que leva a uma *deviance* mínima (método do perfil de verossimilhança). O ajuste para  $\hat{\lambda}$  será, então, comparado com o ajuste para a escolha inicial  $\lambda_0$  que em geral é 1. Este procedimento pode ser usado para vários  $x$ 's simultaneamente e é particularmente útil quando têm as mesmas dimensões físicas, tal que é necessária uma transformação comum. A família mais comum de transformação é a família de Box & Cox (1964) dada por

$$h(x; \lambda) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{para } \lambda \neq 0 \\ \ln x & \text{para } \lambda = 0 \end{cases}.$$

Um método informal para o estudo de uma única covariável toma a forma de  $\mathbf{u}(\lambda_0) = \left. \frac{d\mathbf{z}(\lambda)}{d\lambda} \right|_{\lambda=\lambda_0}$  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 de resíduos parciais, como já discutido em 3.4.3, item (e).

Essa mesma variável  $\mathbf{u}$  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 adequacidade da escala escolhida para a variável explanatória de interesse.

**Exemplo 20: Transformação para a variável dependente** – Seja a família de transformações de Box-Cox normalizada

$$\mathbf{z}(\lambda) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \begin{cases} \frac{\mathbf{y}^\lambda - 1}{\lambda \dot{y}^{\lambda-1}} & \text{para } \lambda \neq 0 \\ \dot{y} \ln \mathbf{y} & \text{para } \lambda = 0 \end{cases}$$

sendo  $\dot{y}$  a média geométrica das observações. A expansão de  $\mathbf{z}(\lambda)$  em uma série de Taylor em relação a  $\lambda_0$ , conhecido, é dada por:

$$\mathbf{z}(\lambda) \cong \mathbf{z}(\lambda_0) + (\lambda - \lambda_0) \mathbf{u}(\lambda_0),$$

sendo  $\mathbf{u}(\lambda_0) = \left. \frac{d\mathbf{z}(\lambda)}{d\lambda} \right|_{\lambda=\lambda_0}$ . Então,

$$\mathbf{z}(\lambda_0) = \mathbf{z}(\lambda) - (\lambda - \lambda_0) \mathbf{u}(\lambda_0) + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \gamma \mathbf{u} + \boldsymbol{\varepsilon}.$$

Mas  $\mathbf{z}(\lambda) = \frac{\mathbf{y}^\lambda - 1}{\lambda \dot{y}^{\lambda-1}}$  e, portanto,

$$\mathbf{u}(\lambda) = \frac{d\mathbf{z}(\lambda)}{d\lambda} = \frac{y^\lambda \ln y - (y^\lambda - 1)(\lambda^{-1} + \ln \dot{y})}{\lambda \dot{y}^{\lambda-1}}.$$

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 resíduos de  $\mathbf{u}(\lambda)$ , então, constantes podem ser ignoradas se  $\boldsymbol{\beta}$  contém uma constante. Então,

$$\mathbf{u}(1) = y \left[ \ln \left( \frac{\mathbf{y}}{\dot{y}} \right) - 1 \right], \text{ variável construída para testar se } \lambda_0 = 1$$

e

$$\mathbf{u}(0) = \dot{y} \ln \mathbf{y} \left[ \frac{\ln \mathbf{y}}{2} - \ln \dot{y} \right], \text{ variável construída para testar se } \lambda_0 = 0.$$

Como  $-\gamma = \lambda - \lambda_0$ , tem-se que uma estimativa para  $\lambda$  pode ser obtida por  $\hat{\lambda} = \lambda_0 - \hat{\gamma}$ . Usa-se, em geral, um valor para  $\lambda$  próximo de  $\hat{\lambda}$  que tenha uma interpretação prática.

**Exemplo 21: Transformação para as variáveis explanatórias** – Se em lugar de transformar  $y$  houver necessidade de transformar  $x_j$ , tem-se que a família maior é:

$$E(\mathbf{Y}) = \sum_{j \neq j'} \beta_j \mathbf{x}_j + \beta_{j'} \mathbf{x}_{j'}^\lambda = \mathbf{z}(\lambda).$$

A expansão de  $\mathbf{z}(\lambda)$  em uma série de Taylor em relação a  $\lambda_0$ , conhecido, é dada por:

$$\mathbf{z}(\lambda) \cong \mathbf{z}(\lambda_0) + (\lambda - \lambda_0) \mathbf{u}(\lambda_0)$$

sendo  $\mathbf{u}(\lambda_0) = \left. \frac{d\mathbf{z}(\lambda)}{d\lambda} \right|_{\lambda=\lambda_0}$ . Então,

$$\mathbf{z}(\lambda) \cong \sum_{j \neq j'} \beta_j x_j + \beta_{j'} x_{j'}^{\lambda_0} + \beta_{j'} (\lambda - \lambda_0) x_{j'}^{\lambda_0} \ln x_{j'} = \sum_{j \neq j'} \beta_j x_j + \beta_{j'} x_{j'}^{\lambda_0} + \gamma \mathbf{u}(\lambda_0)$$

pois  $\frac{d\mathbf{z}(\lambda)}{d\lambda} \cong \beta_{j'} x_{j'}^{\lambda_0} \ln x_{j'}$ . Portanto, testar  $\lambda = \lambda_0$  é equivalente a testar  $\gamma = 0$  para a regressão com a variável construída  $\mathbf{u}(\lambda_0) = x_{j'}^{\lambda_0} \ln x_{j'}$ , com  $x_{j'}^{\lambda_0}$  já no modelo. Para  $\lambda_0 = 1$  tem-se

$$E(\mathbf{Y}) = \sum_{j \neq j'} \beta_j x_j + \beta_{j'} x_{j'} + \beta_{j'} (\lambda - 1) \ln x_{j'} = \mathbf{X}\boldsymbol{\beta} + \gamma \mathbf{u},$$

sendo  $\mathbf{u}(\lambda) = x_{j'} \ln x_{j'}$ .

**Exemplo 22: Transformação simultânea para as variáveis resposta e explanatórias** – Para a transformação simultânea das variáveis resposta e explanatórias (exceto a constante  $1^\lambda = 1$ ) à mesma potência, a variável construída  $\mathbf{u}(\lambda_0)$  para  $\lambda_0 = 1$  é:

$$\mathbf{u}(1) = \sum_{j=2}^p \hat{\beta}_j x_j \ln x_j - y \left[ \ln \left( \frac{y}{\dot{y}} \right) - 1 \right].$$

**Observação:** Comandos GLIM para obtenção de alguns gráficos:

i) Gráficos da variável adicionada ou da regressão parcial

```
! Entrada de dados....
$Yvar Y $
$Fit X1+X2 $
$Extract %rs $Ca R= %rs $
$Yvar X3 $
$Fit . $
$Extract %rs $Ca U = %rs $
$Graph R U $
```

ii) Gráfico de resíduos parciais ou de resíduos mais componente

```
$Yvar Y $
$Fit X<2> $
$Extract %pe %rs $Ca R* = %rs + %pe(2) $
$Graph R* X $
```

iii) Gráfico normal de probabilidades (*normal plot*)

```
$Input %plc qplot $
$use qplot $
```

colocado logo após o ajuste do modelo desejado.

iv) Gráfico semi normal de probabilidades (*half normal plot*)

```
$Yvar Y $
$Fit <modelo> $
$Input 'hnp.mac' $
$Use hnp y $
```

### 3.8 Exercícios

**3.8.1** Os dados da Tabela 14 referem-se à mortalidade de escaravelhos após 5 h de exposição a diferentes doses de  $CS^2$ .

Tabela 14: Mortalidade de escaravelhos

Doses	$m_i$	$Y_i$
1,69	59	6
1,72	60	13
1,76	62	18
1,78	56	28
1,81	63	52
1,84	59	53
1,86	62	61
1,88	60	60

Fonte: Bliss (1935)

Pede-se:

- ajuste o modelo logístico linear e faça o teste para a função de ligação;
- ajuste o modelo complemento log-log e faça o teste para a função de ligação;
- faça o gráfico da variável adicionada para os itens (a) e (b);
- verifique se há necessidade de transformação para a variável dose, usando o gráfico de resíduos parciais.

**3.8.2** Os dados da Tabela 15 referem-se a medidas de diâmetro à altura do peito ( $X_1$ ) e altura ( $X_2$ ) de árvores em pé e de volumes ( $Y$ ) das árvores derrubadas. O objetivo desse tipo de experimento é verificar de que forma essas variáveis estão relacionadas para, através de medidas nas árvores em pé, poder se prever o volume de madeira em uma área de floresta.

Pede-se:

- fazer os gráficos de variáveis adicionadas para  $X_1$  e  $X_2$ ;
- fazer os gráficos de resíduos parciais para  $X_1$  e  $X_2$ ;
- fazer as transformações  $LY = \log(Y)$ ,  $LX_1 = \log(X_1)$  e  $LX_2 = \log(X_2)$  e repetir os gráficos dos itens (a) e (b);
- verificar se existem pontos discrepantes em ambas as escalas;
- usando  $u(1) = \sum_{j=2}^p \hat{\beta}_j x_j \ln x_j - y \left[ \ln \left( \frac{y}{\hat{y}} \right) - 1 \right]$  obtido no **Exemplo 22**, como variável adicionada, verifique que há necessidade da transformação simultânea de  $Y$ ,  $X_1$  e  $X_2$ ;

Tabela 15: Medidas de árvores em uma área florestal.

Amostra	X <sub>1</sub>	X <sub>2</sub>	Y
1	8,3	70	10,3
2	8,6	65	10,3
3	8,8	63	10,2
4	10,5	72	16,4
5	10,7	81	18,8
6	10,8	83	19,7
7	11,0	66	15,6
8	11,0	75	18,2
9	11,1	80	22,6
10	11,2	75	19,9
11	11,3	79	24,2
12	11,4	76	21,0
13	11,4	76	21,4
14	11,7	69	21,3
15	12,0	75	19,1
16	12,9	74	22,2
17	12,9	85	33,8
18	13,3	86	27,4
19	13,7	71	25,7
20	13,8	64	24,9
21	14,0	78	34,5
22	14,2	80	31,7
23	14,5	74	36,3
24	16,0	72	38,3
25	16,3	77	42,6
26	17,3	81	55,4
27	17,5	82	55,7
28	17,9	80	58,3
29	18,0	80	51,5
30	18,0	80	51,0
31	20,6	87	77,0

Fonte: Ryan *et al.* (1976).

**3.8.3** Os dados da Tabela 16 são provenientes de um experimento em delineamento blocos ao acaso em que foram usadas como tratamentos 8 doses de um inseticida fosforado e foram contadas quantas (Y) cenouras estavam danificadas de totais de M cenouras.

Tabela 16: Número de cenouras danificadas (Y) de M cenouras

Log (dose) X	Bloco I		Bloco II		Bloco III	
	Y	M	Y	M	Y	M
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

Fonte: Phelps (1982).

Pede-se:

- 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 função de ligação.

**3.8.4** Considere

$$\eta = g(\mu; \lambda) = \begin{cases} \frac{\mu^\lambda - 1}{\lambda} & \text{para } \lambda \neq 0 \\ \ln \mu & \text{para } \lambda = 0 \end{cases}.$$

Mostre que a variável construída para o teste da hipótese  $H_0: \lambda = 0$  é dada por

$$\mathbf{u}(0) = \left. \frac{dg(\mu; \lambda)}{d\lambda} \right|_{\lambda=0} = -\frac{\ln^2 \hat{\mu}}{2} = -\frac{\hat{\eta}^2}{2} \quad (\text{Atkinson, 1985, pág. 238}).$$

## Capítulo 4

### Aplicações

#### 4.1 Estimação da Dose Efetiva

Como já foi visto no 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, probit e complemento log-log. Tais modelos, ajustados a conjuntos de dados, no caso em que o preditor linear é uma regressão linear simples, podem ser usados para sumarizá-los através do par de estimativas  $(\hat{\beta}_1, \hat{\beta}_2)$  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 muitos casos, porém, o interesse está na determinação de estimativas de doses efetivas,  $\theta_p$  (DE<sub>100p</sub>), que são doses, as quais sob o modelo ajustado causam uma mudança de estado em 100p% dos indivíduos. Um exemplo muito comum é a determinação da DL<sub>50</sub> (também chamada dose mediana) que é a dose que causa 50% de mortalidade dos indivíduos.

Assim, para os modelos citados tem-se:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \hat{\beta}_1 + \hat{\beta}_2 \hat{\theta}_p \Rightarrow \hat{\theta}_p = \frac{1}{\hat{\beta}_2} \left[ \ln\left(\frac{p}{1-p}\right) - \hat{\beta}_1 \right], \text{ logístico;}$$

$$\text{Probit}(p) = \Phi^{-1}(p) = \hat{\beta}_1 + \hat{\beta}_2 \hat{\theta}_p \Rightarrow \hat{\theta}_p = \frac{1}{\hat{\beta}_2} [\Phi^{-1}(p) - \hat{\beta}_1], \text{ probit;}$$

$$\ln[-\ln(1-p)] = \hat{\beta}_1 + \hat{\beta}_2 \hat{\theta}_p \Rightarrow \hat{\theta}_p = \frac{1}{\hat{\beta}_2} \{ \ln[-\ln(1-p)] - \hat{\beta}_1 \}, \text{ complemento log-log.}$$

**Observação:** Para qualquer modelo simétrico verifica-se que  $\hat{\theta}_{50} = -\frac{\hat{\beta}_1}{\hat{\beta}_2}$ .

De uma forma geral, tem-se:

$$\hat{\theta}_p = \frac{1}{\hat{\beta}_2} [F^{-1}(p) - \hat{\beta}_1],$$

sendo  $p = F(\hat{\beta}_1 + \hat{\beta}_2 \hat{\theta}_p)$ . Se  $\hat{\theta}_p = \ln(\hat{d}_p)$ , então,  $\hat{d}_p = \exp\{\hat{\theta}_p\}$ .

Para o modelo assimétrico proposto por Aranda-Ordaz (1981), tem-se

$$F(\hat{\beta}_1 + \hat{\beta}_2 \hat{\theta}_p) = 1 - \left(1 + \lambda e^{\hat{\beta}_1 + \hat{\beta}_2 \hat{\theta}_p}\right)^{-1/\lambda} = p.$$

Então,

$$1 - p = \left(1 + \lambda e^{\hat{\beta}_1 + \hat{\beta}_2 \hat{\theta}_p}\right)^{-1/\lambda} \Rightarrow (1 - p)^{-\lambda} = 1 + \lambda e^{\hat{\beta}_1 + \hat{\beta}_2 \hat{\theta}_p}$$

ou, ainda,

$$\frac{1}{\lambda(1-p)^\lambda} - \frac{1}{\lambda} = e^{\hat{\beta}_1 + \hat{\beta}_2 \hat{\theta}_p} \Rightarrow \hat{\beta}_1 + \hat{\beta}_2 \hat{\theta}_p = \ln\left(\frac{1}{\lambda(1-p)^\lambda} - \frac{1}{\lambda}\right) = F^{-1}(p).$$

Portanto,

$$\hat{\theta}_p = \frac{1}{\hat{\beta}_2} \left[ \ln\left(\frac{1}{\lambda(1-p)^\lambda} - \frac{1}{\lambda}\right) - \hat{\beta}_1 \right]$$

e, para  $p = 0,50$ ,

$$\hat{\theta}_{50} = \frac{1}{\hat{\beta}_2} \left[ \ln\left(\frac{1}{\lambda(I/2)^\lambda} - \frac{1}{\lambda}\right) - \hat{\beta}_1 \right] \Rightarrow \hat{\theta}_{50} = \frac{1}{\hat{\beta}_2} \left[ \ln\left(\frac{2^\lambda - 1}{\lambda}\right) - \hat{\beta}_1 \right].$$

Dentre os métodos usados para a construção de intervalos de confiança para doses efetivas, os mais comuns são: o método Delta, o de Fieller e o da razão de verossimilhanças (perfil de verossimilhanças) (Collett, 1991; Morgan, 1992).

Usando-se os dados do **Exemplo 5**, e calculando-se a dose letal que mata 50% dos insetos e os intervalos de confiança com um coeficiente de confiança de 90% de probabilidade, pelos 3 métodos obtiveram-se os resultados:

i) dose letal:  $\hat{\theta}_{50} = 5,3$ ;

ii) intervalos de confiança:  $\begin{cases} \text{Fieller: } 4,8 < \theta_{50} < 5,9 \\ \text{Delta: } 4,8 < \theta_{50} < 5,9 \\ \text{Perfil de verossimilhança: } 5,0 < \theta_{50} < 5,7. \end{cases}$

## 4.2 Paralelismo entre retas no modelo logístico linear

Na área de toxicologia é muito comum o interesse na comparação da eficiência de produtos (fungicidas, inseticidas, herbicidas, medicamentos etc). Considerando-se o modelo logístico linear com uma variável quantitativa  $x$  (dose ou  $\ln(\text{dose})$ ) e  $J$  produtos a serem testados, os preditores lineares a serem considerados são:

i)  $\eta = \ln\left(\frac{\pi}{1-\pi}\right) = \alpha_j + \beta_j x$  - retas concorrentes;

ii)  $\eta = \ln\left(\frac{\pi}{1-\pi}\right) = \alpha_j + \beta x$  - retas paralelas;

iii)  $\eta = \ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta_j x$  - retas com intercepto comum;

iv)  $\eta = \ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta x$  - retas coincidentes,

sendo que  $j = 1, 2, \dots, J$ . O ajuste desses modelos aos dados é testado através das diferenças das *deviances* 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 por:

$$\text{constante} = \ln\left(\frac{p}{1-p}\right) = \hat{\alpha}_j + \hat{\beta} \hat{\theta}_j^{(p)}, j = 1, 2, \dots, J.$$

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  $x = \ln(\text{dose}) = \ln(d)$ , então,

$$\frac{\hat{\alpha}_j - \hat{\alpha}_{j'}}{\hat{\beta}} = \ln\left\{\frac{\hat{d}_{j'}^{(p)}}{\hat{d}_j^{(p)}}\right\} = \ln \hat{\rho}_{jj'} \Rightarrow \hat{\rho}_{jj'} = \exp\left(\frac{\hat{\alpha}_j - \hat{\alpha}_{j'}}{\hat{\beta}}\right) = \frac{\text{DE50}_{j'}}{\text{DE50}_j},$$

sendo  $\hat{\rho}_{jj'}$  a estimativa da eficiência relativa  $\rho_{jj'}$  do produto  $j$  em relação ao  $j'$  e  $\ln \hat{\rho}_{jj'} = \ln \hat{d}_{j'}^{(p)} - \ln \hat{d}_j^{(p)}$  medindo a 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 obtidos pelos métodos Delta, de Fieller e da razão de verossimilhanças (perfil de verossimilhanças) (Collett, 1991; Morgan, 1992),

**Exemplo 4.1: Resistência a *cypermethrin*** - Amostras de 20 insetos, *Heliothis virescens* (praga do algodão), resistentes a *cypermethrin*, foram expostas a doses crescentes do inseticida, dois dias depois da emergência da pupa. Após 72h foram contados os números de insetos mortos e os resultados obtidos estão na Tabela 17.

Tabela 17: Mortalidade do *Heliothis virescens* após exposição a *cypermethrin*

Doses	Número de insetos mortos ( $Y_i$ )	
	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

Fonte: Collett (1991)

Considerações:

- **variável resposta:**  $Y_i$  = número de insetos mortos em amostras de tamanho  $m_i = 20$ ;
- **distribuição:** Binomial;
- **parte sistemática:** inteiramente casualizado no esquema fatorial, modelos de regressão;
- **objetivo:** determinação de doses letais para machos e fêmeas e verificação se são diferentes.

A Tabela 18 apresenta as *deviances* e  $X^2$  residuais e seus respectivos números de graus de liberdade (g.l.) e a Tabela 19, a Análise de *deviance*.

Tabela 18: *Deviances* e  $X^2$  residuais

Modelo	g.l.	<i>Deviances</i>	Valor p	$X^2$	Valor p
$\eta = \alpha$	11	124,88	< 0,0001	101,4	< 0,0001
$\eta = \text{sexo}_j$	10	118,80	< 0,0001	97,4	< 0,0001
$\eta = \text{dose}_k$	6	15,15	0,0191	12,9	0,0446
$\eta = \text{sexo}_j + \text{dose}_k$	5	5,01	0,4146	3,7	0,5933

Vê-se que existem evidências de que o modelo com preditor linear com dois fatores, sexo (com dois níveis,  $j = 1, 2$ ) e dose (com 6 níveis,  $k = 1, \dots, 6$ , em princípio sem levar em consideração o fato de serem quantitativos), ajusta-se bem aos dados, enquanto que os modelos mais simples, não.

Tabela 19: Análise de *Deviance*

Causa de variação	g.l.	<i>Deviances</i>	Valor p
Sexo	1	6,08	0,0137
Doses Sexo	5	113,79	< 0,0001
Doses	5	109,72	< 0,0001
Sexo Doses	1	10,14	0,0015
Resíduo	5	5,01	0,4146
Total	11	124,88	

Pela Tabela 19 verifica-se que há evidências para efeito significativo de sexo e de dose. Note-se, ainda, que as *deviances* 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. Pode-se, ainda, tentar uma simplificação desse modelo, considerando que dose é um fator quantitativo. Se for usado como preditor linear um polinômio com  $x = \text{dose}$ , verifica-se que há necessidade 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 as *deviances* e  $X^2$  residuais estão apresentados na Tabela 20.

Tabela 20: *Deviances* e  $X^2$  residuais

Modelo	g.l.	<i>Deviances</i>	Valor p	$X^2$	Valor p
$\eta = \alpha$	11	124,88	<0,0001	101,4	<0,0001
$\eta = \alpha + \beta x$	10	16,98	0,0748	14,8	0,1395
$\eta = \alpha + \beta_j x$	9	5,04	0,8308	3,5	0,9411
$\eta = \alpha_j + \beta x$	9	6,76	0,6621	5,3	0,8074
$\eta = \alpha_j + \beta_j x$	8	4,99	0,7586	3,5	0,8991

Pela Tabela 20, vê-se que existem evidências que os modelos com retas concorrentes, paralelas e com intercepto comum ajustam-se bem aos dados. Tem-se, ainda, que as diferenças de *deviances* 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 ( $5,04 - 4,99 = 0,05$ ), ambas com 1 grau de liberdade, não são estatisticamente significativas. Utilizando de parcimônia e facilidade de interpretação opta-se pelo modelo de retas paralelas. A Tabela 21 traz a Análise de *deviance* para o modelo escolhido.

Tabela 21: Análise de *Deviance*

Causa de variação	g.l.	<i>Deviances</i>	Valor p
Sexo	1	6,08	0,0137
Regressão linear	1	112,04	< 0,0001
Resíduo	9	6,76	0,6600
Total	11	124,88	

A partir do modelo escolhido obtêm-se, então, respectivamente, para machos e fêmeas:

$$\hat{\eta} = \ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = -2,372 + 1,064 \log_2(\text{dose}) \Rightarrow DL_{50} = 4,68 \text{ e } DL_{90} = 19,62$$

e

$$\hat{\eta} = \ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = -3,473 + 1,064 \log_2(\text{dose}) \Rightarrow DL_{50} = 9,60 \text{ e } DL_{90} = 40,18.$$

Verifica-se que as fêmeas são mais resistentes, pois para matar 100p% das fêmeas há necessidade de uma dose 2 vezes maior do que para matar 100p% dos machos. Vê-se, que a dose letal para  $p = 0,9$  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 estiver na estimação dessa dose há necessidade de se aumentar a amplitude de doses para fêmeas em um novo experimento. Necessária se faz ainda uma análise de resíduos e diagnósticos.

A Figura 5 mostra o gráfico das curvas ajustadas e os valores observados.

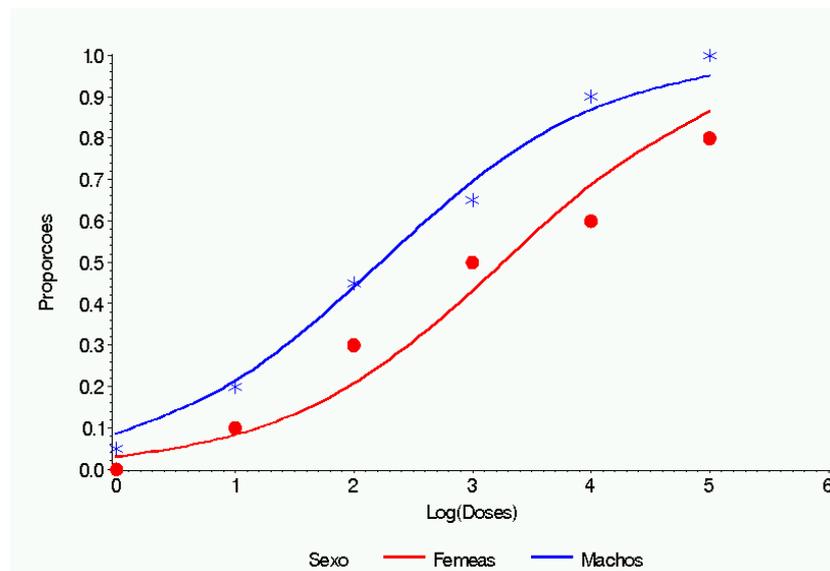


Figura 5: valores observados e modelos ajustados

**Exemplo 4.2: Mortalidade do besouro da farinha** - Grupos de insetos (*Tribolium confusum*, praga da farinha) foram expostos a doses (mg/l) crescentes de bisulfeto de carbono ( $CS_2$ ). Depois de 4 horas foram contados os números de insetos mortos e os resultados obtidos estão na Tabela 22.

Tabela 22: Mortalidade de besouros após 72h de exposição a  $CS_2$ 

Dose de $CS_2$ $d_i$	Repetição 1		Repetição 2	
	$y_i$	$m_i$	$y_i$	$m_i$
49,06	2	29	4	30
52,99	7	30	6	30
56,91	9	28	9	34
60,84	14	27	14	29
64,76	23	30	29	33
68,69	29	31	24	28
72,61	29	30	32	32
76,54	29	29	31	31

Fonte: Collett (1991)

Considerações:

- **variável resposta:**  $Y_i$  = número de insetos mortos em amostras de tamanho  $m_i$  ;
- **distribuição:** binomial;
- **parte sistemática:** inteiramente casualizado, modelos de regressão;
- **objetivo:** determinação de doses letais.

A Tabela 23 apresenta as *deviances* e  $X^2$  residuais e seus respectivos números de graus de liberdade (g.l.) e a Tabela 24, a Análise de *deviance*, considerando-se o **modelo logístico**.

Tabela 23: *Deviances* e  $X^2$  residuais (Logístico)

Modelo	g.l.	<i>Deviances</i>	Valor de $p$	$X^2$	Valor de $p$
$\eta = \alpha$	15	289,14	< 0,0001	241,00	< 0,0001
$\eta = \alpha + \beta_1 x$	14	12,51	0,5654	10,84	0,6985
$\eta = \alpha + \beta_1 x + \gamma$ LP2	13	7,93	0,8475	7,36	0,8826
$\eta = \alpha + \beta_1 x + \beta_2 x^2$	13	7,93	0,8475	7,36	0,8826
$\eta = \alpha + \beta_1 x + \beta_2 x^2 + \gamma$ LP2	12	7,88	0,7944	7,39	0,8308
$\eta = \alpha_k$	8	4,94	0,7639	4,52	0,8074

Vê-se que, em uma primeira análise, existem evidências de que os modelos com preditores lineares dados por uma regressão linear simples, por uma regressão linear quadrática e aquele usando dose como fator qualitativo ajustam-se bem aos dados. Entretanto, ao se considerar a diferença de *deviances* para o modelo com preditor dado por uma regressão linear simples e por uma regressão linear quadrática ( $12,51 - 7,93 = 4,58$ ) verifica-se que existem evidências de um ajuste mais significativo da regressão linear quadrática. Essa diferença pode ser explicada, pois como foi visto, a aproximação assintótica da distribuição da *deviance* residual de uma distribuição  $\chi^2$  pode ser pobre. Já a distribuição da diferença de *deviances* residuais tem propriedades melhores. Verifica-se, ainda que o teste para a função de ligação após a regressão linear (coincide, neste caso, com o teste para a adição do termo quadrático) indica a não adequação da função de ligação logística, enquanto que após a regressão linear quadrática indica a adequação. Necessária se faz ainda uma análise de resíduos e diagnósticos.

Tabela 24: Análise de *Deviance* (Logístico)

Causa de variação	g.l.	<i>Deviances</i>	Valor de $p$
Regressão linear	1	276,60	< 0,0001
(Função de ligação)	(1)	(4,58)	0,0323
Regressão quadrática	1	4,58	0,0323
(Função de ligação)	(1)	(0,04)	0,8415
(Desvios)	(5)	(2,99)	0,7015
(Doses)	(7)	(284,20)	< 0,0001
Resíduo	13	7,93	0,8481
Total	15	289,14	

Tem-se, então, a partir do modelo logístico com preditor linear dado por uma regressão linear quadrática

$$\hat{\eta} = \ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = 7,968 - 0,5166 \text{ dose} + 0,006372 \text{ dose}^2 \Rightarrow \text{LD}_{50} = 60,35 \text{ e } \text{LD}_{90} = 67,80.$$

Note que para a obtenção das doses letais há a necessidade de se resolver uma equação de segundo grau. Como solução de

$$\ln\left(\frac{0,5}{1-0,5}\right) = \ln(1) = 0 = 7,968 - 0,5166\theta + 0,006372\theta^2$$

obtêm-se os valores 60,35 e 20,72, e como solução de

$$\ln\left(\frac{0,9}{1-0,9}\right) = \ln(9) = 7,968 - 0,5166\theta + 0,006372\theta^2,$$

os valores 67,80 e 13,36.

A Tabela 25 apresenta as *deviances* e  $X^2$  residuais e seus respectivos números de graus de liberdade (g.l.) e a Tabela 26, a Análise de *deviance*, considerando o **modelo complemento log-log**.

Tabela 25: *Deviances* e  $X^2$  residuais (complemento log-log)

Modelo	g.l.	<i>Deviances</i>	Valor de $p$	$X^2$	Valor de $p$
$\eta = \alpha$	15	289,14	< 0,0001	241,00	< 0,0001
$\eta = \alpha + \beta_1 x$	14	8,67	0,8516	8,62	0,8546
$\eta = \alpha + \beta_1 x + \gamma$ LP2	13	8,30	0,8235	7,87	0,8519
$\eta = \alpha_k$	8	4,94	0,7639	4,52	0,8074

Vê-se que, em uma primeira análise, existem evidências de que os modelos com preditores lineares dados por uma regressão linear simples e aquele usando dose como fator qualitativo ajustam-se bem aos dados. Além disso, ao se considerar a diferença de *deviances* para o modelo com preditor dado por uma regressão linear simples e dose como fator qualitativo ( $8,67 - 4,94 = 3,73$ ) existem evidências de que ela não é significativa. Também, o teste para a função de ligação após a regressão linear indica a adequação da função de ligação complemento log-log. Necessária se faz ainda uma análise de resíduos e diagnósticos.

Tabela 26: Análise de *Deviance* (Complemento log-log)

Causa de variação	g.l.	<i>Deviances</i>	Valor de $p$
Regressão linear	1	280,50	< 0,0001
(Função de ligação)	(1)	(0,37)	0,5430
(Desvios)	(6)	(3,73)	0,7131
(Doses)	(7)	(284,20)	< 0,0001
Resíduo	14	8,67	0,8516
Total	15	289,14	

Tem-se, então, a partir do modelo complemento log-log com preditor linear dado por uma regressão linear simples

$$\hat{\eta} = \ln[-\ln(1 - \hat{\pi})] = -9,755 + 0,1554 \text{dose} \Rightarrow LD_{50} = 60,40 \text{ e } LD_{90} = 68,13.$$

A Figura 6 mostra o gráfico das curvas ajustadas usando o modelo logístico com preditor linear dado por uma regressão linear quadrática e o modelo complemento log-log com preditor linear dado por uma regressão linear simples e os valores observados.

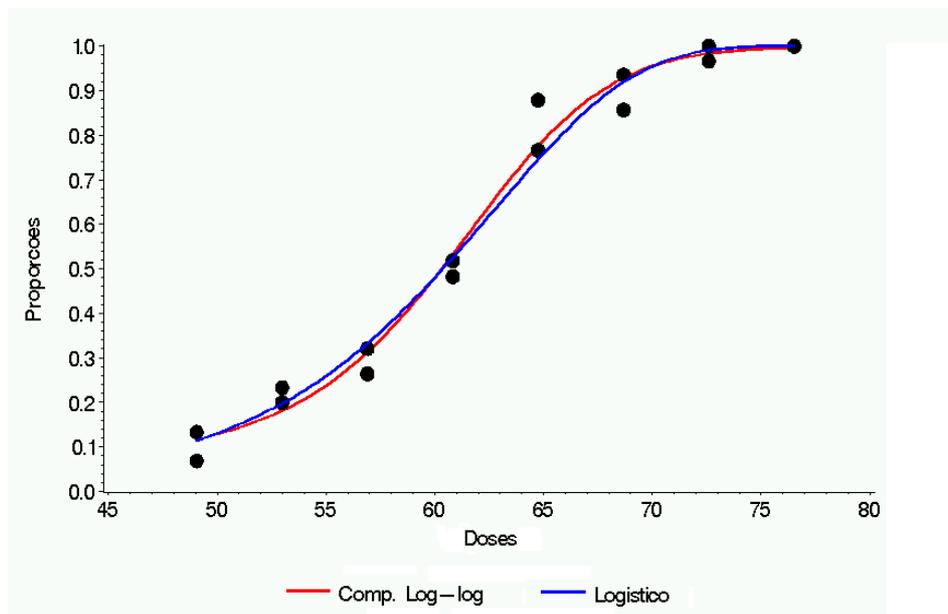


Figura 6. Curvas ajustadas e valores observados

**Exemplo 4.3:** Os dados da Tabela 27 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 foi 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 27: Número de frutos produzidos no ano anterior e o número de gemas

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
Crispin	1	93	43	0,46
Crispin	2	147	59	0,40
Crispin	3	149	57	0,38
Crispin	4	151	43	0,28
Cox	0	34	12	0,35
Cox	1	92	15	0,16
Cox	2	133	18	0,14
Cox	3	146	14	0,10
Cox	4	111	9	0,08
Golden Delicious	0	21	6	0,29
Golden Delicious	1	89	20	0,22
Golden Delicious	2	118	20	0,17
Golden Delicious	3	124	21	0,10
Golden Delicious	4	81	4	0,00

Fonte: Ridout (1991)

Considerações:

- **variável resposta:**  $Y_i$  = número de gemas florais em totais de  $n_i$  gemas;
- **distribuição:** binomial;
- **ligação:** função logística  $\eta = \ln\left(\frac{\pi}{1-\pi}\right)$ ;
- **parte sistemática:** delineamento inteiramente casualizado e modelos de análise de variância com  $\eta = \alpha_j + \beta_j x$ , isto é, supondo uma equação linear com coeficientes linear e angular diferentes para as três variedades, sendo  $x$ , o número de frutos no ano anterior;
- **objetivo:** verificar se o número de frutos de um ano tem influência na produção do ano seguinte e se isso depende da variedade.

A Tabela 28 apresenta as *deviances* e  $X^2$  residuais e seus respectivos números de graus de liberdade (g.l.) e a Tabela 29, a Análise de *deviance*.

Tabela 28: *Deviances* e  $X^2$  residuais

Modelo	g.l.	<i>Deviances</i>	Valor p	$X^2$	Valor p
$\eta = \alpha$	14	182,16	< 0,0001	181,10	< 0,0001
$\eta = \alpha + \beta x$	13	138,99	< 0,0001	139,00	< 0,0001
$\eta = \alpha_j$	12	53,04	< 0,0001	54,51	< 0,0001
$\eta = \alpha + \beta_j x$	11	31,08	0,0011	30,97	0,0011
$\eta = \alpha_j + \beta x$	11	8,80	0,6403	8,67	0,6523
$\eta = \alpha_j + \beta_j x$	9	7,87	0,5473	7,76	0,5585

Vê-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. Além disso, ao se considerar a diferença de *deviances* para o modelo com retas paralelas e com retas concorrentes ( $8,80 - 7,87 = 0,93$ ) existem evidências de que ela não é significativa. A Tabela 29 traz a Análise de *deviance* para o modelo de retas paralelas. Vê-se que existem evidências para o efeito de variedades e para o efeito de regressão linear.

Tabela 29: Análise de *Deviance*

Causa de variação	g.l.	<i>Deviances</i>	Valor p
Variedades	2	129,12	<0,0001
Regressão linear	1	44,24	<0,0001
Resíduo	11	8,80	0,6403
Total	14	182,16	

Tem-se, também, que

$$\begin{aligned}\hat{\alpha}_2 - \hat{\alpha}_1 &= -1,510 \text{ e } s(\hat{\alpha}_2 - \hat{\alpha}_1) = 0,1566 \\ \hat{\alpha}_3 - \hat{\alpha}_1 &= -1,258 \text{ e } s(\hat{\alpha}_3 - \hat{\alpha}_1) = 0,1563 \\ \hat{\alpha}_3 - \hat{\alpha}_2 &= 0,252\end{aligned}$$

e, usando-se as estimativas da matriz de variâncias e covariâncias,

$$\begin{aligned}V(\hat{\alpha}_3 - \hat{\alpha}_2) &= \text{Var}(\hat{\alpha}_2 - \hat{\alpha}_1) + \text{Var}(\hat{\alpha}_3 - \hat{\alpha}_1) - 2\text{Cov}\left(\hat{\alpha}_2 - \hat{\alpha}_1, \hat{\alpha}_3 - \hat{\alpha}_1\right) \\ &= 0,02452 + 0,02443 - 2 \cdot 0,00725 = 0,03445.\end{aligned}$$

e, portanto,

$$s(\hat{\alpha}_3 - \hat{\alpha}_2) = 0,1856.$$

Logo,

$$\begin{aligned}t(\hat{\alpha}_2 - \hat{\alpha}_1) &= \frac{-1,510}{0,1566} = -9,64 \\ t(\hat{\alpha}_3 - \hat{\alpha}_1) &= \frac{-1,258}{0,1563} = -8,05 \\ t(\hat{\alpha}_3 - \hat{\alpha}_2) &= \frac{0,252}{0,1856} = 1,36,\end{aligned}$$

que comparados com o valor tabelado  $t_{11;0,05} = 2,20$  da distribuição t de Student mostram que as variedades Cox e Golden Delicious não diferem entre si (como os contrastes feitos não são ortogonais, não se tem o nível conjunto de significância para as comparações feitas). Devido a esse resultado um novo modelo foi usado considerando-se os dados das variedades 2 e 3 como se fosse uma única variedade. A diferença entre as *deviances* residuais,  $(10,64 - 8,80 = 1,84)$  indica que existem evidências de que as variedades 2 e 3 comportam-se de forma semelhante. Os resultados obtidos para a análise de *deviance* estão na Tabela 30.

Tabela 30: Análise de *Deviance*

Causas de Variação	g.l.	<i>Deviances</i>	Valor p
Variedades	1	127,17	<0,0001
Regressão linear	1	44,35	<0,0001
Resíduo	12	10,64	0,556
Total	14	182,16	

Logo, a proporção esperada de gemas florais pode ser calculada por

$$\hat{\pi}_i = \begin{cases} \frac{e^{0,3605-0,3302x}}{1 + e^{0,3605-0,3302x}} & \text{para } i = 1, \dots, 5 \\ \frac{e^{-1,0285-0,3302x}}{1 + e^{-1,0285-0,3302x}} & \text{para } i = 6, \dots, 15. \end{cases}$$

A Figura 7 mostra os gráficos das curvas ajustadas usando-se o modelo logístico, considerando-se três e dois interceptos diferentes, e os valores observados.

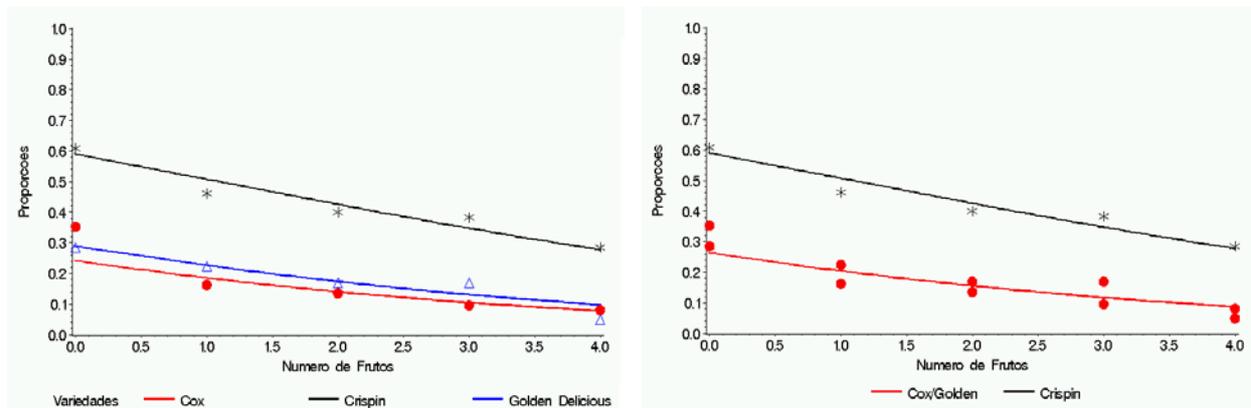


Figura 7. Modelos logístico para três e dois interceptos diferentes e os valores observados

### 4.3 Outras Aplicações

**Exemplo 4.4:** Os dados apresentados na Tabela 31 referem-se a um experimento no delineamento em blocos casualizados com cultura de tecidos de maçãs. Os tratamentos no esquema fatorial 2 x 3 x 3 foram:

- A: 2 tipos de citocinina (BAP, TDZ)
- B: 3 níveis de citocinina (5,0; 1,0 e 0,1)
- C: 3 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 3 fatores) e o explante. A variável resposta, Y, é binária, isto é,

$$Y = \begin{cases} 1 & \text{se o explante regenerou após 4 semanas} \\ 0 & \text{em caso contrário.} \end{cases}$$

Considerações:

- **variável resposta:**  $Y_i$  = número de insetos mortos em amostras de tamanho  $m_i = 1$ ;
- **distribuição:** Bernoulli (caso particular da Binomial);
- **ligação:** função logística;
- **parte sistemática:** delineamento em blocos casualizados e modelos de análise de variância com preditor linear:  $\eta = \mu + \delta_{\perp} + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk}$ ;
- **objetivo:** verificar a influência dos fatores sobre a regeneração e se existe interação entre eles.

Tabela 31: Dados de um experimento com cultura de tecidos de maçãs

Citocinina		Auxina	Blocos									
Tipo	Nível		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
TDZ	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
	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

Fonte: Ridout (1991)

A Tabela 32 apresenta as *deviances* e  $X^2$  residuais e seus respectivos números de graus de liberdade (g.l.) e a Tabela 33, a Análise de *deviance*, considerando-se o **modelo logístico**.

Como foi observado em 2.7 a *deviance* residual não é informativa para a verificação da adequação dos modelos para dados binários, pois é apenas uma função dos dados (Exercício 2.12.12). A diferença entre *deviances*, porém, pode ser utilizada. Verifica-se, então, que apenas o Tipo de Citocinina e o Tipo de Auxina têm influência significativa na regeneração de tecidos de maçã e além do mais existe uma interação entre esses dois fatores, conforme mostra a Tabela 33, para os modelos encaixados na seqüência.

Tabela 32: *Deviances* e  $X^2$  residuais (logístico)

Modelo	g.l.	<i>Deviances</i>	$X^2$
1	179	202,44	180,0
B1	170	193,79	180,0
B1+A	169	165,65	179,5
B1+B	168	189,72	183,4
B1+C	168	187,07	178,9
B1+A+B	167	160,83	181,8
B1+A+C	167	157,57	207,1
B1+B+C	166	182,83	181,5
B1+A*B	165	157,57	189,2
B1+A*C	165	147,82	210,3
B1+B*C	163	142,27	214,0
B1+A*B+C	163	149,03	221,0
B1+A*C+B	163	142,27	214,0
B1+B*C+A	161	148,08	203,1
B1+A*B+A*C	161	138,66	197,9
B1+A*B+B*C	159	141,71	547,8
B1+B*C+A*C	159	137,05	232,7
B1+A*B*C	153	127,01	152,4

Tabela 33: Análise de *Deviance*

Causas de Variação	g.l.	<i>Deviance</i>	Valor p
Blocos	9	8,6	0,4749
Tipo de Citocinina (A)	1	28,1	< 0,0001
Nível de Citocinina (B A)	2	4,8	0,0907
Auxina (C A,B)	2	8,4	0,0149
AB (A,B,C)	2	3,4	0,1827
AC (A*B,C)	2	10,4	0,0055
BC (A*B,A*C)	4	6,0	0,1991
ABC	4	5,6	0,2311
Resíduo	153	127,0	0,9383
Total	179	202,4	

Foi, então, considerado o modelo reduzido com preditor linear:

$$\eta = \mu + \delta_{\ell} + \alpha_j + \gamma_k + \alpha\gamma_{jk}$$

e os resultados estão na Tabela 34. Necessária se faz, ainda, uma análise de resíduos e diagnósticos.

Tabela 34: Análise de *Deviance*

Causas de Variação	g.l.	<i>Deviances</i>	Valor p
Blocos	9	8,6	0,4749
Tipo de Citocinina (A)	1	28,1	< 0,0001
Auxina (C A)	2	8,1	0,0174
AC	2	9,8	0,0074
Resíduo	165	147,8	0,8276
Total	179	202,4	

A seguir tem-se o quadro de médias considerando-se os fatores tipo de citocinina e tipo de auxina.

Tabela 35: Quadro de médias

Auxina	Tipo de Citocinina	
	BAP	TDZ
NAA	<b>0,33</b>	0,93
IBA	0,67	0,97
2-4D	0,77	0,83

Verifica-se que o dado da casela marcada é o responsável pela interação significativa entre os dois fatores. Na realidade esse experimento foi repetido mais 4 vezes e a interação não foi significativa.

**Exemplo 4.5:** Os dados apresentados na Tabela 36 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 x 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, alguns explantes morreram. No caso em que morreram todos os explantes, o recipiente foi eliminado do experimento, pois algumas dessas mortes podem ter sido devido à contaminação com bactérias o que não está relacionado com o tratamento. No caso em que houve 1 ou 2 mortes no recipiente os dados foram considerados e, usou-se \* no lugar do dado perdido.

Considerações:

- **variável resposta:**  $Y_i$  = número de brotos;
- **distribuição:** Poisson;
- **ligação:** função logarítmica,  $\eta = \ln(\mu)$ ;
- **parte sistemática:** delineamento inteiramente casualizado, modelos de análise de variância com preditor linear:  $\eta = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_k$
- **objetivo:** verificar se existe interação entre meio de cultura e quantidade de hormônio e se influenciam o número de brotos.

Tabela 36: Números de brotos 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 2 1	2 2 3	1 4 4	2 4 *	1 * *	
B	X1	4 1 4 0 4 2	5 4 5 6 2 2	5 4 3 3 3 *	3 4 4 1 5 *	2 3 2	1 0 1	
B	X2	2 2 1 0 8 2	2 4 4 2 4 *	2 3 0 3 1 *	0 0 4 10 * *	<b>12 0 4</b>	1 0 4	
C	X1	0 2 0 2 2 2	1 1 3 2 0 2	5 3 3	3 2 1	2 2 2	0 2 2	
C	X2	2 2 3	<b>11 6 5</b>	5 3 4	6 4 *	4 4 *	3 3 *	

Fonte: Ridout (1991)

A Tabela 37 apresenta a Análise de *deviance*, considerando-se o modelo proposto.

Tabela 37: Análise de *Deviance*

Causas de Variação	g.l.	<i>Deviances</i>	Valor p
Meio de cultura (M)	2	0,42	0,8106
Níveis de hormônio (H)	1	5,20	0,0226
Interação M x H	2	14,93	0,0006
Entre recipientes	38	61,99	0,0083
Entre pl. d. recipientes	73	94,70	0,0448
Total	116	177,31	

Verifica-se que a diferença entre as *deviances* obtidas para Entre recipientes e Entre plantas dentro de recipientes não é significativa. Adotou-se, então, como preditor linear

$$\eta = \mu + \alpha_i + \beta_j + \alpha\beta_{ij},$$

obtendo-se os resultados da Tabela 38.

Tabela 38: Análise de *Deviances*

Causas de Variação	g.l.	<i>Deviances</i>	Valor p
Meios de cultura (M)	2	0,42	0,8106
Níveis de hormônio (H)	1	5,20	0,0226
Interação M x H	2	14,93	0,0006
Resíduo	111	156,76	0,0028
Total	116		

A *deviance* 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 diagnósticos como complementação. Vê-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 39 verifica-se que a interação está sendo significativa devido ao meio de cultura C. O exame da Tabela 36, indica duas parcelas em destaque cuja influência na análise mereceria ser melhor estudada.

Tabela 39: 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 4.6:** Considere os dados da Tabela 4, que se referem a contagens de partículas de vírus para 5 diluições diferentes, sendo que foram usadas 4 repetições para as 4 primeiras diluições e 5 repetições para a última diluição. O objetivo do experimento era estimar o número de partículas de vírus por unidade de volume.

A análise desses dados foi feita inicialmente considerando-se, um modelo do tipo discutido em (2.2) para ensaios de diluição em que se tem

- **distribuição:** Poisson

- **ligação:** logarítmica

- **preditor linear:**  $\eta = \ln \lambda_i + \ln x_i = \beta_i + \text{variável "offset"}$ ,

porém, com a diferença de se estar supondo que os coeficientes escalares variam dependendo da concentração. Os resultados obtidos foram

Tabela 40: Análise de *Deviance*

Causas de Variação	g.l.	<i>Deviance</i>	Valor p
Diluições	4	3,168	0,5301
Resíduo	16	10,381	0,8460
Total	20	13,549	

Verifica-se que não há necessidade de se utilizarem coeficientes  $\beta$ 's diferentes para as diferentes diluições, isto é, pode-se considerar como preditor linear

$$\eta = \log \lambda + \log x_i = \beta + \text{variável "offset"}.$$

Tem-se, então, que

$$\hat{\beta} = \log \hat{\lambda} = 3,955 \Rightarrow \hat{\lambda} = e^{3,955} = 52,2,$$

isto é, 52 partículas de vírus por unidade de volume. Um intervalo de confiança para  $\lambda$  com um coeficiente de confiança de 95% de probabilidade é dado por

$$3,955 \pm 1,96 \cdot 0,083 \Rightarrow \text{I.C.}(\lambda)_{0,95} : (44,36; 61,42).$$

A Figura 8 a seguir apresenta os valores observados e o modelo ajustado.

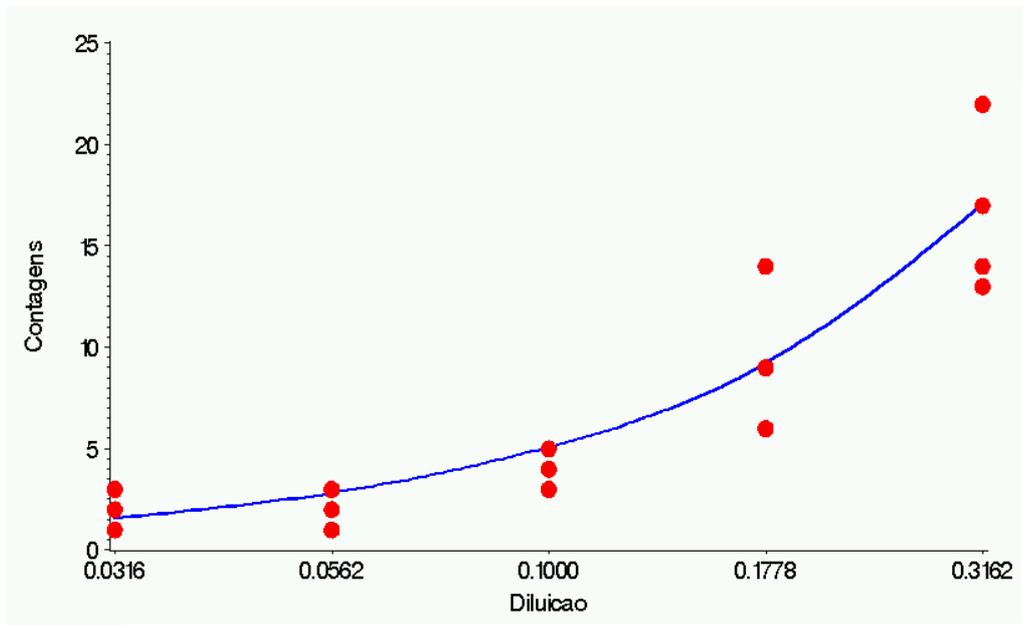


Figura 8. Valores observados e modelo ajustado.

# APÊNDICE A

## A.1 Programa Glim para os dados do exemplo 4.1 - Capítulo 4.

```
! Exemplo 3.7 - Collett pag. 75
$slen 12 $data r $read
  1  4  9 13 18 20
  0  2  6 10 12 16
$gfactor sexo 2 dose 6 $
$ca n=20 : ld=dose-1 :d=2**ld $
$yvar r $err b n $
$fit $pr 'X2= ' %X2 $
$fit +sexo $pr 'X2= ' %X2 $
$fit +dose $pr 'X2= ' %X2 $
$fit : +dose $pr 'X2= ' %X2 $
$fit +sexo $
$fit sexo : +d : +d<2> : +d<3> : +d<4> : +dose $
!
$pr 'Usando log(dose) - base 2'$
$pr 'Modelo com interceptos comuns' $
$fit : +sexo.ld $pr 'X2= ' %X2$
$pr 'Modelo de retas coincidentes' $
$fit : +ld $pr 'X2= ' %X2 $
$pr 'Modelo de retas concorrentes' $
$fit sexo : +sexo.ld $pr 'X2= ' %X2$
$pr 'Modelo de retas paralelas'$
$fit sexo-1 : +ld $pr 'X2= ' %X2 $d e $
$extr %pe $
$number d50M d50F d90M d90F $
$ca d50M=2**(-%pe(1)/%pe(3)) : d50F=2**(-%pe(2)/%pe(3))$
$pr 'LD50 (macho)= ' d50M : 'LD50 (femea)= ' d50F $
$ca d90M=2**((%log(9)-%pe(1))/%pe(3))
  : d90F=2**((%log(9)-%pe(2))/%pe(3)) $
$pr 'LD90 (macho)= ' d90M : 'LD90 (femea)= ' d90F $
!
$number z $ca z=%log(32)/%log(2) $
$ca p=r/n : $
$assign pld=0,0.05...z$
  : pld=pld,pld
  : psex=pld $
$ca psex=%gl(2,%len(pld)/2) $
$fac psex 2 $
$predict sexo=psex ld=pld $
$gstyle 2 colour 1 linetype 0 symbol 5$
$graph (c='post' ) p,%pfv ld,pld 1,2,10,11 sexo,psex $
$ca lgtp=%log(p/(1-p))/%log(2)
  : lgtfv=%log(%pfv/(1-%pfv))/%log(2) $
$graph lgtp,lgtfv ld,pld 1,2,10,11 sexo,psex $
$finish
```

**A.2 Programa Glim para os dados do exemplo 4.2 - Capítulo 4.**

```

! *** CS2.glm ***
! Collett (1991), pag. 109
$slen 16 $
$data y n $read
  2 29 4 30
  7 30 6 30
  9 28 9 34
14 27 14 29
23 30 29 33
29 31 24 28
29 30 32 32
29 29 31 31
$assign X=49.06, 52.99, 56.91, 60.84, 64.76, 68.69, 72.61, 76.54$
$ca i=%gl(8,2) : d=X(i) : p=y/n $
$ca d2=d**2 $
$gfactor dose 8 $
$yvar y $error B n $
$pr 'Ligacao logistica' $
$link G !ligacao logistica
$fit $pr 'X2= ' %X2 $
$fit +d $pr 'X2= ' %X2 $
$ca LP2=%lp**2 $
$fit +LP2 $pr 'X2= ' %X2 $d e $
$fit d : +d2 $pr 'X2= ' %X2 $
$d e $
$ca LP2=%lp**2 $
$fit +LP2 $pr 'X2= ' %X2 $d e $
!
$fit : +d<1> : + d<2> $pr 'X2= ' %X2 $
$assign pd=49, 49.5...77 $
$predict d=pd $ca fv1=%pfv $
$fit +dose $pr 'X2= ' %X2 $
$fit : +dose$
!
$pr 'Ligacao complemento log-log'$
$link c $ !ligacao complemento log-log
$fit $pr 'X2= ' %X2 $
$fit + d $pr 'X2= ' %X2 $
$ca LP2=%lp**2 $
$fit + LP2 $pr 'X2= ' %X2 $
$fit d : + dose $pr 'X2= ' %X2 $
$fit d $d e $extr %pe $
$predict d=pd $ca fv2=%pfv $
$graph(c='post' h='dose' v='proporcao') p, fv1, fv2 d, pd, pd 1, 10, 11 $
$number ld50 ld90$
$ca ld50=(%log(%log(2))-%pe(1))/%pe(2)$
$pr 'LD50= ' ld50 $
$ca ld90=(%log(%log(10))-%pe(1))/%pe(2)$
$pr 'LD90= ' ld90 $
$fit : +dose$
$return

```

### A.3 Programa Glim para os dados do exemplo 4.3 - Capítulo 4.

```

! Exemplo 5.2 - Demetrio pag. 105
$slen 15 $data x Y N $read
0 42 69
1 43 93
2 59 147
3 57 149
4 43 151
0 12 34
1 15 92
2 18 133
3 14 146
4 9 111
0 6 21
1 20 89
2 20 118
3 21 124
4 4 81
$gfactor var 3 $
$yvar Y $err b N $
$pr 'Modelo de retas concorrentes'$
$fit $pr 'X2= ' %X2 $
$fit +var $pr 'X2= ' %X2 $
$fit +var.X $pr 'X2= '%X2$
$pr 'Modelo de retas coincidentes'$
$fit $fit +X $pr 'X2= ' %X2$
$pr 'Modelo com interceptos comuns '$
$fit $fit +var.X $pr 'X2= ' %X2$
$pr 'Modelo de retas paralelas'$
$fit var-1 : +X $pr 'X2= ' %X2$d e$
$d e $
$ca p=y/n $
$gstyle 2 colour 1 linetype 0 symbol 5$
$gstyle 3 colour 1 linetype 0 symbol 2$
$assign PX=0,0.1...4.0$
$assign Px=Px,Px,Px$
$assign Pv=Px$
$ca pv=%gl(3,%len(Px)/3) $
$fac pv 3$
$predict var=pv x=px$
$graph (h='numero de frutos' v='logit(proporcao)')
  p,%pfv x,px 1,2,3,10,11,12 var, pv $

$factor var 2 $assign var=1,1,1,1,1,2,2,2,2,2,2,2,2,2,2 $
$fit : +var : +x $
$fit var-1+x $d e$
!
$del pv : px$
$assign PX=0,0.1...4.0$
$assign Px=Px,Px$
$assign Pv=Px$
$ca pv=%gl(2,%len(Px)/2) $
$fac pv 2$
$predict var=pv x=px$
$graph (c='post' h='numero de frutos' v='proporcao')
  p,%pfv x,px 1,2,10,11 var, pv $
$return
$finish

```

**A.4 Programa Glim para os dados do exemplo 4.4 - Capítulo 4.**

```

$slen 180 $data y $read
1 1 0 0 1 0 1 0 1 1
0 1 1 1 1 1 0 1 1 1
1 1 1 1 1 1 1 0 0 1
0 0 0 0 0 0 0 0 0 0
1 1 1 0 0 1 1 0 1 1
1 0 1 1 0 1 1 1 1 1
0 0 1 1 1 0 1 0 0 0
0 0 0 1 1 1 1 0 1 0
0 0 1 1 1 1 1 0 1 1
1 1 1 1 1 0 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 0 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 0
1 1 1 1 1 1 1 0 1 1
1 1 1 1 1 1 0 1 1 1
0 0 1 0 1 1 1 1 1 1
$gfactor A 2 B 3 C 3 Bl 10 $
$ca n=1$
$yvar y $err Bin n$
$fit $pr 'X2= ' %X2 $
$fit : +Bl $pr 'X2= ' %X2 $
$fit Bl : +A $pr 'X2= ' %X2 $
$fit Bl : +B $pr 'X2= ' %X2 $
$fit Bl : +C $pr 'X2= ' %X2 $
$fit Bl +A : +B $pr 'X2= ' %X2 $
$fit Bl +B : +A$pr 'X2= ' %X2 $
$fit Bl +A : +C $pr 'X2= ' %X2 $
$fit Bl +C : +A$pr 'X2= ' %X2 $
$fit Bl +B : +C $pr 'X2= ' %X2 $
$fit Bl +C : +B$pr 'X2= ' %X2 $
$fit Bl +A+B : +A.B $pr 'X2= ' %X2 $
$fit Bl +A+C : +A.C $pr 'X2= ' %X2 $
$fit Bl +B+C : +B.C $pr 'X2= ' %X2 $
$fit Bl +A*B : +C $pr 'X2= ' %X2 $
$fit Bl +A*C : +B $pr 'X2= ' %X2 $
$fit Bl +B*C : +A $pr 'X2= ' %X2 $
$fit Bl +A*B+C : + A.C $pr 'X2= ' %X2 $
$fit Bl +A*B+C : + B.C $pr 'X2= ' %X2 $
$fit Bl +A*C+B : + A.B $pr 'X2= ' %X2 $
$fit Bl +A*C+B : + B.C $pr 'X2= ' %X2 $
$fit Bl +B*C+A : + A.C $pr 'X2= ' %X2 $
$fit Bl +B*C+A : + A.B $pr 'X2= ' %X2 $
$fit Bl : +A*B*C $pr 'X2= ' %X2 $fit -A.B.C$

$pr 'Modelo final'$
$fit Bl : +A : +C : +A.C $

$return

```

### A.5 Programa Glim para os dados do exemplo 4.5 - Capítulo 4.

```

$slen 132 $data y $read
4 5 2      1 2 5      2 2 -1
3 5 3      2 2 1      2 2 3      1 4 4      2 4 -1      1 -1 -1
2 -1 -1
4 1 4      5 4 5      5 4 3      3 4 4      2 3 2      1 0 1
0 4 2      6 2 2      3 3 -1      1 5 -1
2 2 1      2 4 4      2 3 0      0 0 4      12 0 4      1 0 4
0 8 2      2 4 -1      3 1 -1      10 -1 -1
0 2 0      1 1 3      5 3 3      3 2 1      2 2 2      0 2 2
2 2 2      2 0 2
2 2 3      11 6 5      5 3 4      6 4 -1      4 4 -1      3 3 -1
$ca w=%ge(y,0) $ca y=y+%eq(y,-1) $
$factor meio 3 horm 2 $
$var 9 A : 21 B : 30 C D : 24 e : 18 f $
$ca a=1: b=1 : c=2 : d=2 : e=3 : f=3 $
$assign meio=a,b,c,d,e,f $
$ca b=2 : c=1 : d=2 : e=1 : f=2 $
$assign horm=a,b,c,d,e,f $
$fac jar 44 $ca jar =%gl(44,3) $
$yvar y $wei w $err P$
$fit : +meio : +horm : +meio.horm $
$d e $
$fit + jar $
$return

```

### A.6 Programa Glim para os dados do exemplo 4.6 - Capítulo 4.

```

$slen 21 $data y $read
13 14 17 22
9 14 6 14
4 4 3 5
3 2 1 3
2 1 3 2 2
$data x $read
0.3162 0.3162 0.3162 0.3162
0.1778 0.1778 0.1778 0.1778
0.1000 0.1000 0.1000 0.1000
0.0562 0.0562 0.0562 0.0562
0.0316 0.0316 0.0316 0.0316 0.0316
$fac fx 5 $assign fx=1,1,1,1,2,2,2,2,3,3,3,3,4,4,4,4,5,5,5,5,5 $
$ca logx=%log(x) $ ! Calcula o offset
$yvar y $err P $off logx $
$fit $d e $
$fit +fx $
$return

```

## APÊNDICE B

### B.1 Programas SAS para os dados do exemplo 4.1 - capítulo 4

#### B.1.1 - Para esquema fatorial

```

data collett;
input doses $ sexo $ y n;
datalines;
  1.0 m 1 20
  2.0 m 4 20
  4.0 m 9 20
  8.0 m 13 20
 16.0 m 18 20
 32.0 m 20 20
  1.0 f 0 20
  2.0 f 2 20
  4.0 f 6 20
  8.0 f 10 20
 16.0 f 12 20
 32.0 f 16 20
;

proc genmod;
class sexo doses;
  model y/n= / dist=b type1 type3;
run;

proc genmod;
class sexo doses;
  model y/n=sexo / dist=b type1 type3;
run;

proc genmod;
class sexo doses;
  model y/n=doses / dist=b type1 type3;
run;

proc genmod;
class sexo doses;
  model y/n=sexo doses / dist=b type1 type3;
run;

proc genmod;
class sexo doses;
  model y/n=sexo|doses / dist=b type1 type3;
run;

```

#### B.1.2 - Para modelo de regressão

```

data collett;
input doses sexo $ y n;
ld=log(doses);
datalines;
  1.0 m 1 20
  2.0 m 4 20
  4.0 m 9 20
  8.0 m 13 20
 16.0 m 18 20
 32.0 m 20 20
  1.0 f 0 20
  2.0 f 2 20
  4.0 f 6 20
  8.0 f 10 20
 16.0 f 12 20
 32.0 f 16 20
;

proc genmod;
class sexo;
  model y/n= / dist=b type1 type3;
run;

proc genmod;
class sexo;
  model y/n=sexo sexo*ld /dist=b type1 type3;
run;

proc genmod;
class sexo;
  model y/n=sexo ld / noint dist=b type1 type3;
run;

proc genmod;
class sexo;
  model y/n=sexo*ld / dist=b type1 type3;
run;

proc genmod;
class sexo;
  model y/n=ld/ dist=b type1 type3;
run;

```

## B.2 Programas SAS para os dados do exemplo 4.3 - capítulo 4

### B.3.1 - Para Cox, Golden e Crispin diferentes

```

data macieira;
input var $ X N Y;
p=y/n;
datalines;
Cox 0 34 12
Cox 1 92 15
Cox 2 133 18
Cox 3 146 14
Cox 4 111 9
Golden 0 21 6
Golden 1 89 20
Golden 2 118 20
Golden 3 124 21
Golden 4 81 4
Crispin 0 69 42
Crispin 1 93 43
Crispin 2 147 59
Crispin 3 149 57
Crispin 4 151 43
;
title 'Modelo Completo com Intercepto';
proc genmod order=data;
class var;
model y/n=var|x /dist=bin typel type3 covb;
run;
title;

title 'Modelo Completo sem Intercepto';
proc genmod order=data;
class var;
model y/n=var|x /noint dist=bin typel type3
covb;
run;
title;

title 'Modelo Reduzido com Intercepto';
proc genmod order=data;
class var;
model y/n=var x / dist=bin typel type3 covb;
contrast 'Golden vs Cox' var -1 1 0;
contrast 'Cox vs Crispin' var 1 0 -1;
contrast 'Golden vs Crispin' var 0 1 -1;
estimate 'Golden vs Cox' var -1 1 0;
estimate 'Cox vs Crispin' var 1 0 -1;
estimate 'Golden vs Crispin' var 0 1 -1;
run;
title;

title 'Modelo Reduzido sem Intercepto';
proc genmod order=data;
class var;
model y/n=var x/noint dist=bin typel type3 covb;
contrast 'Golden vs Cox' var -1 1 0;
contrast 'Cox vs Crispin' var 1 0 -1;
contrast 'Golden vs Crispin' var 0 1 -1;
estimate 'Golden vs Cox' var -1 1 0;
estimate 'Cox vs Crispin' var 1 0 -1;
estimate 'Golden vs Crispin' var 0 1 -1;
run;
title;

```

### B.3.2 - Para Cox e Golden reunidas

```

data macieira2;
input var $ X N Y;
p=y/n;
datalines;
Crispin 0 69 42
Crispin 1 93 43
Crispin 2 147 59
Crispin 3 149 57
Crispin 4 151 43
CoxGold 0 34 12
CoxGold 1 92 15
CoxGold 2 133 18
CoxGold 3 146 14
CoxGold 4 111 9
CoxGold 0 21 6
CoxGold 1 89 20
CoxGold 2 118 20
CoxGold 3 124 21
CoxGold 4 81 4
;
title 'Modelo Reduzido 1 com Intercepto';
proc genmod order=data;
class var;
model y/n=var x /dist=bin typel type3 covb;
estimate 'Crispin vs CoxGold' var 1 -1;
run;
title;

title 'Modelo Reduzido 1 sem Intercepto';
proc genmod order=data;
class var;
model y/n=var x /noint dist=bin typel type3 covb;
estimate 'Crispin vs CoxGold' var 1 -1;
run;
title;

```

### B.3 Programa SAS para os dados do exemplo 4.5 - capítulo 4

```

DATA EXE53;
INPUT Meio $ Hormonio $ Erecip $ Drecip $
      Brotos;
DATALINES;
A      X1      1      1      4
A      X1      1      2      5
A      X1      1      3      2
A      X1      2      4      1
A      X1      2      5      2
A      X1      2      6      5
A      X1      3      7      2
A      X1      3      8      2
A      X1      3      9      .
A      X2      4      10     3
A      X2      4      11     5
A      X2      4      12     3
A      X2      5      13     2
A      X2      5      14     2
A      X2      5      15     1
A      X2      6      16     2
A      X2      6      17     2
A      X2      6      18     3
A      X2      7      19     1
A      X2      7      20     4
A      X2      7      21     4
A      X2      8      22     2
A      X2      8      23     4
A      X2      8      24     .
A      X2      9      25     1
A      X2      9      26     .
A      X2      9      27     .
A      X2     10      28     2
A      X2     10      29     .
A      X2     10      30     .
B      X1     11      31     4
B      X1     11      32     1
B      X1     11      33     4
B      X1     12      34     5
B      X1     12      35     4
B      X1     12      36     5
B      X1     13      37     5
B      X1     13      38     4
B      X1     13      39     3
B      X1     14      40     3
B      X1     14      41     4
B      X1     14      42     4
B      X1     15      43     2
B      X1     15      44     3
B      X1     15      45     2
B      X1     16      46     1
B      X1     16      47     0
B      X1     16      48     1
B      X1     17      49     0
B      X1     17      50     4
B      X1     17      51     2
B      X1     18      52     6
B      X1     18      53     2
B      X1     18      54     2
B      X1     19      55     3
B      X1     19      56     3
B      X1     19      57     .
B      X1     20      58     1
B      X1     20      59     5
B      X1     20      60     .
B      X2     21      61     2
B      X2     21      62     2
B      X2     21      63     1
B      X2     22      64     2
B      X2     22      65     4
B      X2     22      66     4
B      X2     23      67     2
B      X2     23      68     3
B      X2     23      69     0
B      X2     24      70     0
B      X2     24      71     0
B      X2     24      72     4
B      X2     25      73     12
B      X2     25      74     0
B      X2     25      75     4
;
B      X2     26      76      1
B      X2     26      77      0
B      X2     26      78      4
B      X2     27      79      0
B      X2     27      80      8
B      X2     27      81      2
B      X2     28      82      2
B      X2     28      83      4
B      X2     28      84      .
B      X2     29      85      3
B      X2     29      86      1
B      X2     29      87      .
B      X2     30      88      10
B      X2     30      89      .
B      X2     30      90      .
C      X1     31      91      0
C      X1     31      92      2
C      X1     31      93      0
C      X1     32      94      1
C      X1     32      95      1
C      X1     32      96      3
C      X1     33      97      5
C      X1     33      98      3
C      X1     33      99      3
C      X1     34     100      3
C      X1     34     101      2
C      X1     34     102      1
C      X1     35     103      2
C      X1     35     104      2
C      X1     35     105      2
C      X1     36     106      0
C      X1     36     107      2
C      X1     36     108      2
C      X1     37     109      2
C      X1     37     110      2
C      X1     37     111      2
C      X1     38     112      2
C      X1     38     113      0
C      X1     38     114      2
C      X2     39     115      2
C      X2     39     116      2
C      X2     39     117      3
C      X2     40     118      11
C      X2     40     119      6
C      X2     40     120      5
C      X2     41     121      5
C      X2     41     122      3
C      X2     41     123      4
C      X2     42     124      6
C      X2     42     125      4
C      X2     42     126      .
C      X2     43     127      4
C      X2     43     128      4
C      X2     43     129      .
C      X2     44     130      3
C      X2     44     131      3
C      X2     44     132      .
;
RUN;
title 'Modelo com Efeito Entre e Dentro de
      Recipientes';
PROC GENMOD;
  CLASS Meio Hormonio Erecip Drecip;
  MODEL BROTOS=MEIO|HORMONIO Erecip Drecip
        /DIST=P TYPE1 TYPE3;
RUN;
title;
PROC GENMOD;
  CLASS Meio Hormonio Erecip Drecip;
  MODEL BROTOS=meio|hormonio /DIST=P TYPE1
        TYPE3;
Lsmeans meio|hormonio;
RUN;

```

## B.4 Programa SAS para os dados do exemplo 4.6 - capítulo 4

```

data ex54;
input diluicao valor rep;
of=log(valor);
datalines;
0.0316 2
0.0316 1
0.0316 3
0.0316 2
0.0316 2
0.0562 3
0.0562 2
0.0562 1
0.0562 3
0.1000 4
0.1000 4
0.1000 3
0.1000 5
0.1778 9
0.1778 14
0.1778 6
0.1778 14
0.3162 13
0.3162 14
0.3162 17
0.3162 22
;

proc genmod order=data;
class diluicao;
model valor=diluicao of /dist=poisson type1
type3 offset=of;
output out=saida p=predito;
run;
proc print data=saida;
run;

Goptions vsize=6 hsize=6 ;

Symbol1 c=black i=none v=dot l=1 h=1 w=1;
Symbol2 c=black i=spline v=none l=1 h=2 w=1;

Axis1 w=1 label=(a=90 'Contagens') order=(0
to 25 by 5);
Axis2 w=1;

PROC GPLOT data=saida;
plot (valor predito)*Diluicao / overlay
nolegend vaxis=axis1 haxis=axis2 noframe;
label Y='Contagens' X='Diluição';
run;

```

## Referências Bibliográficas

- AGRESTI, A. **Categorical Data Analysis**. New York: John Wiley & Sons, 1990. 558p.
- AITKIN, M.; ANDERSON, D.; FRANCIS, B.; HINDE, J. **Statistical Modelling in GLIM**. New York: Oxford University Press, 1989. 374p.
- ARANDA-ORDAZ, F. J. On two families of transformations to additivity for binary response data. **Biometrika**, **68**, p.357-63, 1981.
- ASHTON, W. D. **The Logit Transformation with Special Reference to its Uses in Bioassay**, London: Griffin, 1972. 88p.
- ATKINSON, A. C. Two graphical displays for outlying and influential observations in regression. **Biometrika**, **68**, p.13-20, 1981.
- ATKINSON, A. C. **Plots, Transformations and Regression**. New York: Oxford University Press, 1985. 282p.
- ATKINSON, A. C.; DAVISON, A. C.; NELDER, J. A.; O'BRIEN, C. M. **Model Checking**. London: Imperial College, 1989.
- BARNDORFF-NIELSEN, O. E. **Parametric Statistical Models and Likelihood**. Lecture Notes in Statistics, **50**, New York: Springer-Verlag, 1988.
- BELSLEY, D. A.; KUH, E.; WELSCH, R. E. **Regression Diagnostics**. New York: John Wiley & Sons, 1980.
- BERKSON, J. Application of the logistic function to bioassay. **Journal of the American Statistical Association**, **39**, p.357-365, 1944.
- BIRCH, M. W. Maximum likelihood in three-way contingency tables. **Journal of the Royal Statistical Society B**, **25**, p.220-233, 1963.
- BLISS, C. I. The calculation of the dosage-mortality curve. **Annals Applied Biology**, **22**, p.134-167, 1935.
- BOX, G. E. P.; COX, D. R. An analysis of transformations. **Journal of the Royal Statistical Society B**, **26**, p.211-52, 1964.
- BUSE, A. The likelihood ratio, Wald and Lagrange multiplier tests: An expository note. **The American Statistician**, **36**, 3, p.153-57, 1982.
- COLLETT, D. **Modelling Binary Data**. London: Chapman & Hall, 1991. 369p.
- COOK, R. D.; WEISBERG, S. **Residuals and Influence in Regression**. New York: Chapman and Hall, 1982.

- COX, D. R.; SNELL, E. J. A general definition of residuals (with discussion). **Journal of the Royal Statistical Society B**, **30**, p.248-275, 1968.
- COX, D. R.; HINKLEY, D.V. **Theoretical Statistics**. Cambridge: Chapman and Hall, 3<sup>rd</sup> ed., 1986. 174p.
- CORDEIRO, G.M. **Modelos Lineares Generalizados**. Campinas, VII SINAPE, 1986. 286p.
- DYKE, G. V.; PATTERSON, H. D. Analysis of factorial arrangements when the data are proportions. **Biometrics**, **8**, p.1-12, 1952.
- DEMÉTRIO, C. G. B.; HINDE, J. P. Half normal plots and overdispersion. **GLIM newsletter**, **27**, p.19-26, 1997.
- DOBSON, A. J. **An Introduction to Statistical Modelling**. London: Chapman and Hall, 2<sup>nd</sup> ed., 1990. 174p.
- FAHRMEIR, L; KAUFMAN, H. Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. **Annals of Statistics**, **13**, p.342-68, 1985.
- FAHRMEIR, L.; TUTZ; G. **Multivariate Statistical Modelling based on Generalized Linear Models**. New York,: Springer-Verlag, 1994.
- FEIGL, P.; ZELEN, M. Estimation of exponential survival probabilities with concomitant information. **Biometrics**, **21**, p.826-838, 1965.
- FIRTH, D. Generalized Linear models. In D. Hinkley, N., N. Reid, and E. Snell (Eds.), **Statistical Theory and Modelling**, p.55-82, Chapman & Hall, 1991.
- FISHER, R. A. On the mathematical foundations of theoretical statistics. **Philosophical Transactions of the Royal Society**, **222**, p.309-368, 1922.
- FISHER, R. A. Two new properties of mathematical likelihood. **Proceedings of the Royal Society A** , **144**, p.285-307, 1934.
- FISHER, R. A.; YATES, F. **Statistical Tables for Biological, Agricultural and Medical Research**. Edinburgh: Oliver & Boyd, 1970.
- FRANCIS, B. J.; GREEN, M.; PAYNE, C. **The GLIM System. Release 4 Manual**. New York: Oxford University Press, 1993.
- GASSER, M. Exponential survival with covariance. **Journal of the American Statistical Association**, **62**, p.561-568, 1967.
- HINDE, J. P.; DEMÉTRIO, C. G. B. Overdispersion: Models and Estimation. **13<sup>o</sup> SINAPE - Simpósio Nacional de Probabilidade e Estatística**, Caxambu, MG, 1998.

- JØRGENSEN, B. Exponential dispersion models (with discussion). **Journal of the Royal Statistical Society B**, **49**, p.127-62, 1987.
- JØRGENSEN, B.; LABOURIAU, R. S. **Famílias Exponenciais e Inferência Teórica**. Rio de Janeiro: Intituto de Matemática Pura e Aplicada, 1992. 264p.
- KENDALL, M. G.; STUART, A. **The Advanced Theory of Statistics - Distribution Theory**. New York: Hafner, 3<sup>rd</sup> ed., 1969, 3v.
- LINDSEY, J. K. **Applying Generalized Linear Models**. New York: Springer-Verlag, 1997. 256p.
- MARTIN, J.T. 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**, p 69-81, 1942.
- McCULLAGH, P.; NELDER, J. A. **Generalized Linear Models**. London: Chapman and Hall, 2<sup>nd</sup> ed., 1989. 511p.
- McCULLOCH, C. E.; SEARLE, S.R. **Generalized, Linear, and Mixed Models**. New York: John Wiley & Sons, 2000. 325p.
- MENDENHALL, W.; SCHEAFFER, R. L.; WACKERLY, D. D. **Mathematical Statistics with Applications**. Boston, Duxbury, 2<sup>nd</sup> ed., 1981. 686p.
- MIAZAKI, E.; STANGENHAUS, G. Métodos para Detecção de Dados Atípicos. **11<sup>o</sup> SINAPE – Simpósio Nacional de Probabilidade e Estatística**. Belo Horizonte – MG, 1994.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the Theory of Statistics**. McGraw-Hill, 3<sup>rd</sup> ed., 1974. 564p.
- MORGAN, B. J. T. **Analysis of Quantal Response Data**. London: Chapman & Hall, 1992.
- NELDER, J. A. Inverse polynomials, a useful group of multifactor response functions. **Biometrics**, **22**, p.128-141, 1966.
- NELDER, J. A.; WEDDERBURN, R. W. M. Generalized Linear Models. **Journal of the Royal Statistical Society A**, **135**, 3, p.370-84, 1972.
- NELDER, J. A.; PREGIBON, D. An extended quasi-likelihood function. **Biometrika**, **74**, 221-232, 1987.
- PAULA, G. A. **Modelos de Regressão com Apoio Computacional**. São Paulo: IME/USP. 2000.

- PHELPS, K. 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, **14**, p.155-163. New York: Springer-Verlag, 1982.
- PREGIBON, D. Logistic regression diagnostics. **Ann. Statist.**, **9**, p.705-724, 1981.
- RASCH, G. **Probabilistic Models for Some Intelligence and Attainment Tests**. Copenhagen: Danmarks Paedagogiske Institut., 1960.
- RIDOUT, M. S. Non-convergence of Fisher's method of scoring - A simple example, **GLIM Newsletter**, **20**, p.8-11, 1990.
- RIDOUT, M. S. Using Generalized Linear Models to Analyze Data from Agricultural, and Horticultural Experiments. **Minicurso** lecionado no Departamento de Matemática e Estatística da ESALQ/USP, 1991 (não publicado).
- RIDOUT, M. S.; FENLON, J. **Statistics in Microbiology**. East Malling: Notes for workshop. 1998.
- RYAN, T. A.; JOINER, B. L.; RYAN, B. F. **Minitab Student Handbook**. North Scituate, Mass: Duxbury Press, 1976.
- SAHA, A.; DONG, D. Estimating nested count data models. **Oxford Bulletin of Economics & Statistics**, **59**, p.423-430, 1997.
- SEARLE, S. R. **Matrix Algebra Useful for Statistics**. New York: John Wiley & Sons, 1982
- SAS Institute. **SAS/STAT User's Guide 8.0**. Cary, N. C.: SAS Institute Inc., 1999.
- SILVEIRA NETO, S.; NAKANO, O.; BARBIN, D.; VILLA NOVA, N. A. **Manual de Ecologia dos Insetos**. São Paulo: Ed. Agronômica 'Ceres', 1976. 419p.
- SILVEY, S. D. **Statistical Inference**. London: Chapman and Hall, 2<sup>nd</sup> ed., 1975. 191p.
- SMYTH, G. K. Generalized linear models with varying dispersion. **Journal of the Royal Statistical Society B**, London, **51**, 1, p.47-60, 1989.
- TUKEY, J. W. One degree of freedom for non-additivity. **Biometrics**, **5**, p.232-242, 1949.
- WALD, A. Tests os statistical hypotheses concerning several parameters when the number of observations is large. **Trans. Amer. Math. Soc.**, **54**, p.426-482, 1943.
- WANG, P. C. Adding a variable in generalized linear models. **Technometrics**, **27**, p.273-276, 1985.

- WANG, P. C. Residuals plots for detecting non-linearity in generalized linear models. **Technometrics**, **29**, p.435-438, 1987.
- WEISBERG, S. **Applied Linear Regression**. New York: John Wiley & Sons, 2<sup>nd</sup> ed. 1985. 324p.
- WILLIAMS, D.A. The Use of the deviance to test the goodness of fit of a logistic-linear model to binary data". **The GLIM Newsletter**, **6**, p.60-62, 1983.
- WILLIAMS, D.A. Generalized Linear model diagnostics using the deviance and "single-case deletions". **Applied Statistics**, **36**, p.181-191, 1987.
- ZIPPIN C.; ARMITAGE, P. Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter. **Biometrics**, **22**, p.665-672, 1966

## LISTA ADICIONAL DE MODELOS LINEARES GENERALIZADOS

**Prof. Clarice G. B. Demétrio**

**1.5.1** Suponha que  $X$  tenha distribuição de Poisson com parâmetro  $\Lambda$ . Suponha ainda que  $\Lambda$  seja uma variável aleatória com distribuição exponencial de parâmetro  $\mu$ . Pede-se:

- a) a distribuição marginal de  $X$ ;
- b) a densidade condicional de  $\Lambda$  dado  $X = k$ .

**1.5.2** O número de acidentes em que se envolve um motorista durante um ano é uma variável aleatória  $Y$  tendo distribuição de Poisson com parâmetro  $\Lambda$ . Suponha ainda que  $\Lambda$  seja uma variável aleatória com distribuição Gama  $(\alpha, \beta)$ . Pede-se:

- a) a distribuição marginal de  $Y$ ;
- b) a densidade condicional de  $\Lambda$  dado  $Y = k$ ;
- c) a média e a variância de  $Y$  usando argumentos de esperança e variância condicionais;
- d) o coeficiente de correlação entre  $Y$  e  $\Lambda$ .

**1.5.3** Uma variável aleatória  $X$  com distribuição Beta tem densidade

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} I_{(0,1)}(x).$$

Seja  $Y$  uma variável aleatória cuja distribuição condicionada a  $X = p$  é Binomial de parâmetros  $N$  e  $p$ , pede-se:

- a) a distribuição marginal de  $Y$ ;
- b) a densidade condicional de  $X$  dado  $Y = k$ ;
- c) a média e a variância de  $Y$  usando argumentos de esperança e variância condicionais;
- d) o coeficiente de correlação entre  $Y$  e  $X$ .

**1.5.4** Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória de  $X \sim B(p)$ . Suponha que  $p$  é uma variável aleatória com distribuição Beta  $(\alpha, \beta)$ .

- a) Determine a distribuição condicional de  $p$  dado  $X_1, X_2, \dots, X_n$ .
- b) Calcule a esperança condicional de  $p$  dado  $X_1, X_2, \dots, X_n$ .

**1.5.4** As variáveis  $X$  e  $Y$  tem distribuição conjunta dada por  $f(x, y) = 2 I_A(x, y)$ , onde  $A = \{0 < y < x < 1\}$ . Achar:

a) as distribuições condicionais de  $X$  dado  $Y = 1/2$  e de  $Y$  dado  $X = 1/3$ ;

b)  $\text{Cov}(X, Y)$ ;

c) as marginais de  $X$  e de  $Y$ . Elas são independentes?

**1.5.6** Seja  $X_1, X_2, \dots$  uma seqüência de variáveis aleatórias independentes e identicamente distribuídas segundo uma distribuição exponencial de parâmetro  $\lambda > 0$ . Defina  $Y = 0$  se  $N = 0$  e  $Y = X_1 + X_2 + \dots + X_N$  se  $N = 1, 2, 3, \dots$ , onde  $N$  é uma variável tendo distribuição Geométrica de parâmetro  $p$ , isto é,  $P(N = n) = pq^{n-1} I_A(n)$ ,  $A = \{1, 2, 3, \dots\}$ . Ache a distribuição de  $Y$ .

**1.5.7** Seja  $X \sim B(N, p)$  onde  $N \sim \text{Bin}(m, a)$ . Mostre que  $X \sim \text{Bin}(m, pa)$ .

**1.5.8** Suponha que  $X | N = n \sim \text{Bin}(n, p)$  e que  $N \sim \text{Poisson}(\lambda)$ . Mostre que  $X \sim \text{Poisson}(\lambda p)$ .

**1.5.9** Suponha que  $X | R = r \sim \text{Bin}(r, p)$  e que  $R \sim \text{Geométrica}(\beta)$ , sendo  $A = \{1, 2, 3, \dots\}$ . Mostre que a distribuição marginal de  $X$  é geométrica com parâmetro  $c = (1 - \beta).p / (1 - \beta p)$  e  $A = \{0, 1, 2, 3, \dots\}$ .

Obs.:  $A = \{x | f(x) = P(X = x) > 0\}$ .

**1.5.10** O número de acidentes por semana em uma usina de açúcar é uma variável aleatória com distribuição de Poisson com média 2. O número de empregados acidentados em diferentes acidentes são independentemente distribuídos com média 3 e variância 4. Determine a média e a variância do número de empregados acidentados por semana.

**1.5.11** Solomon (1983) detalha o seguinte modelo biológico. Suponha que cada um de um número aleatório,  $N$ , de insetos ponha  $X_i$  ovos, onde as  $X_i$ 's são independentes e identicamente distribuídas. O número total de ovos postos é :

$$H = \sum_{i=1}^N X_i .$$

É muito comum supor que  $N \sim \text{Poisson}(\lambda)$ . Além disso vamos supor ainda que cada  $X_i$  tenha uma distribuição logarítmica de parâmetro  $p$ , isto é,:

$$P(X = x) = - \frac{1}{\ln(p)} \frac{(1-p)^x}{x} I_A(x), \text{ sendo } A = \{1, 2, 3, \dots\}.$$

Mostre que a distribuição de  $H \sim \text{Bin}(r, p)$ , em que  $r = -\lambda / \ln(p)$ .

**1.5.12** Se  $Y | X = x \sim N(x, x^2)$  e  $X \sim U(0, 1)$ . Achar:

- a)  $E(Y)$  e  $\text{Var}(Y)$ ;
- b) a distribuição conjunta de  $(X, Y)$ ;
- c) a marginal de  $Y$ .

**1.5.13** Seja  $f(x,y) = \frac{2}{n(n+1)} I_A(x,y)$ , sendo  $A = \{1, 2, \dots, n\} \times \{1, 2, \dots, n\}$ . Achar:

- a)  $E(X | Y = y)$  e  $E(Y | X = x)$ ;
- b)  $\rho(X, Y)$ .

**1.5.14** Seja  $f(x, y) = (x + y) I_A(x, y)$ , sendo  $A = (0, 1) \times (0, 1)$ . Calcular:

- a)  $E(X)$ ,  $\text{Var}(X)$ ,  $E(Y)$ ,  $\text{Var}(Y)$ ;
- b)  $E(X | Y = y)$  e  $\text{Var}(X | Y = y)$ .

**1.5.15** Sejam  $X$  e  $Y$  variáveis aleatórias relacionadas com ovelhas de um rebanho da raça Corriedale assim definidas.

- $X$  : Cordeiros nascidos por ovelha parida;
- $Y$  : Cordeiros desmamados por ovelha parida.

A distribuição conjunta de  $(X, Y)$  é dada por:

y	x	
	1	2
0	0,15	0,05
1	0,50	0,20
2	0	0,10

- a) determine  $E(X)$ ,  $E(Y)$ ,  $\text{Var}(X)$ ,  $\text{Var}(Y)$ ,  $\text{cov}(X, Y)$ ;
- b) ache a distribuição condicional de  $Y | X = x$  e a de  $X | Y = y$ ;
- c) ache a distribuição de:  $V = E(X | Y)$ ,  $U = E(X | Y)$ ,  $W = \text{Var}(X | Y)$  e  $T = \text{Var}(Y | X)$ ;
- d) verifique que:  $E[E(Y | X)] = E(Y)$  e  $\text{Var}[E(Y | X)] + E[\text{Var}(Y | X)] = \text{Var}(Y)$ .

**1.5.16** Remessas diárias contendo um número variável de peças (N) chegam a uma loja de acordo com a seguinte lei de probabilidade

n	10	11	12	13	14	15
P(N = n)	0,05	0,10	0,10	0,20	0,35	0,20

A probabilidade de que qualquer peça em particular seja defeituosa é a mesma para todas as peças igual a 0,10. Se X representar o número de peças defeituosas recebidas pela loja durante o dia, qual será a média e a variância de X?

**1.5.17** Suponha que o fornecimento energético (quilowatts) a uma companhia hidrelétrica, durante um período especificado, seja uma variável aleatória X, a qual admitiremos ter uma distribuição uniforme sobre [10, 30]. A demanda da potência (quilowatts) é também uma variável aleatória Y, que admitiremos ser uniformemente distribuída sobre [10,20]. Deste modo, em média, mais potência é fornecida do que é demandada, porque  $E(X) = 20$  e  $E(Y) = 15$ . Para cada quilowatt fornecido, a companhia tem um lucro de US\$ 0,03. Se a demanda exceder a oferta, a companhia obterá potência adicional de outra fonte, tendo um lucro de US\$ 0,01 por quilowatt desta potência fornecida. Qual será o lucro esperado, durante o período especificado?

**1.5.18** Suponha que o peso do filho de um homem com x quilos seja uma variável aleatória normalmente distribuída com média (x +1) e variância 4. Você vai prever o peso do filho de um homem com 72 quilos. Qual a melhor a previsão?

**1.5.19** Um rato está preso em um labirinto. Nesse local há 3 portas. A primeira porta leva a um túnel que após 3 minutos o levará para a liberdade. A segunda porta o levará a um outro túnel que o trará de volta ao labirinto após 5 minutos. A terceira porta o levará para um outro túnel que o trará de volta após 7 minutos. Se supusermos que em todas as vezes o ratinho escolherá uma das portas com igual probabilidade determine a média e a variância do tempo gasto pelo rato até alcançar a liberdade?

**1.5.20** A função densidade conjunta de (X,Y) é dada por

$$f(x,y) = x \exp[-x(y + 1)] I_A(x,y), \quad A = (0, \infty)^2$$

- Qual a densidade condicional de Y | X = x? E de X | Y = y?
- Qual a densidade de  $V = E(X/Y)$ ?