

3 Modelos Lineares Generalizados

No capítulo 2 foram considerados apenas modelos lineares com distribuição normal e função de ligação identidade. Neste capítulo apresentamos os modelos lineares generalizados (MLG), que foram propostos por Nelder e Wedderburn (1972). Primeiro vemos quais as distribuições de probabilidade usadas nos MLG. Em seguida vemos qual é a estrutura formal dos MLG. Nas seções seguintes mostramos como são feitos a estimação e o teste de significância dos parâmetros, como é verificada a adequação do modelo e também mostraremos o que são a Quase-Verossimilhança, Quase-Verossimilhança Estendida e Quase-Verossimilhança Restrita.

3.1. Distribuições de Probabilidade

Os MLG foram propostos para aplicações onde a variável de resposta y pode ser representada por alguma distribuição da família de exponencial, As funções densidade de probabilidades das distribuições da família de exponencial podem ser expressas na forma:

$$f(y; \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\} \quad (3.1)$$

onde $a(\phi)$, $b(\theta)$ e $c(y, \phi)$ são funções específicas. O parâmetro θ é o parâmetro de localização e ϕ é o parâmetro de dispersão, muitas vezes denominado σ^2 . Pode-se mostrar que a média e a variância da resposta y são dadas por

$$E(y) = \mu = \frac{db(\theta)}{d\theta} \quad \text{e}$$
$$\text{var}(y) = \frac{d^2b(\theta)}{d\theta^2} a(\phi).$$

A parte da variância de Y que não depende de $a(\phi)$ é dada por

$$V(\mu) = \frac{\text{var}(y)}{a(\phi)} = \frac{d^2b(\theta)}{d\theta^2} = \frac{d\mu}{d\theta}$$

que representa a parte da variância de y que depende da sua média μ . A função $V(\mu)$ é denominada função de variância. Assim, a variância de y é um produto de dois fatores, um que depende da média e outro, $a(\phi)$, que não depende.

As principais distribuições pertencentes à família exponencial são: normal, gama, Poisson, binomial, e normal inversa. Apresentamos a seguir a função densidade de probabilidade e as funções $a(\phi)$, $b(\theta)$ e $c(y, \phi)$ específicas de cada uma destas distribuições.

a) Distribuição normal

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

$$= \exp\left[\frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2) - \frac{y^2}{2\sigma^2}\right]$$

Comparando com a Equação (3.1) obtemos:

$$\theta = \mu; \quad b(\theta) = \frac{\mu^2}{2}; \quad a(\phi) = \phi; \quad \phi = \sigma^2; \quad c(y, \phi) = -\frac{1}{2}\ln(2\pi\sigma^2) + \frac{y^2}{2\sigma^2}$$

Média e variância de y :

$$E(y) = \frac{db(\theta)}{d\theta} = \mu; \quad \text{var}(y) = \frac{d^2b(\theta)}{d\theta^2} a(\phi) = \sigma^2$$

b) Distribuição gama

$$f(y; \mu, \alpha) = \frac{1}{\Gamma(\alpha)} \left(\frac{\alpha}{\mu}\right)^\alpha y^{\alpha-1} e^{-\alpha(y/\mu)}$$

$$= \exp\left[\alpha \ln \alpha - \alpha \ln \hat{\mu} + (\alpha - 1) \ln y_i - \frac{\alpha y}{\mu} \ln \Gamma(\alpha)\right]$$

$$= \exp\left[\frac{-y \frac{1}{\mu} - \ln(\mu)}{\frac{1}{\alpha}} + \alpha \ln(\alpha) + (\alpha - 1) \ln(y) - \frac{\alpha y}{\mu} - \ln \Gamma(\alpha)\right]$$

Então

$$\theta = -\frac{1}{\mu}; \quad b(\theta) = -\ln(\mu); \quad a(\phi) = \phi; \quad \phi = \frac{1}{\alpha};$$

$$c(y, \phi) = \alpha \ln(\alpha) + (\alpha - 1) \ln(y) - \frac{\alpha y}{\mu} - \ln \Gamma(\alpha)$$

Média e variância de y :

$$E(y) = \frac{db(\theta)}{d\theta} = \mu; \quad \text{var}(y) = \frac{d^2b(\theta)}{d\theta^2} a(\phi) = \mu^2 \phi = \frac{\mu^2}{\alpha}$$

c) Distribuição de Poisson

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

Então

$$\theta = \ln(\mu); \quad b(\theta) = -\mu; \quad a(\phi) = \phi; \quad \phi = 1; \quad c(y, \phi) = -\ln(y!)$$

Média e variância de y :

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

d) Distribuição binomial

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

Então

$$\theta = \ln \left(\frac{\mu}{1-\mu} \right); \quad b(\theta) = n \ln(1-\mu); \quad a(\phi) = \phi; \quad \phi = 1; \quad c(y, \phi) = \ln \binom{n}{y}$$

Média e variância de y :

$$E(y) = \frac{db(\theta)}{d\theta} = \mu; \quad \text{var}(y) = \frac{d^2b(\theta)}{d\theta^2} a(\phi) = \mu(1-\mu)$$

e) Distribuição normal inversa

$$\begin{aligned} f(y; \mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2 y^3}} \exp \left[\frac{-(y - \hat{\mu})^2}{2\sigma^2 \hat{\mu}^2 y} \right] \\ &= \exp \left[\frac{-y \frac{1}{2\mu^2} + \frac{1}{\mu}}{\sigma^2} - \frac{1}{2y\sigma^2} - \frac{1}{2} \ln(2\pi\sigma^2 y^3) \right] \end{aligned}$$

Então

$$\theta = -\frac{1}{2\mu^2}; \quad b(\theta) = \frac{1}{\mu}; \quad a(\phi) = \phi; \quad \phi = \sigma^2; \quad c(y, \phi) = -\frac{1}{2y\sigma^2} - \frac{1}{2} \ln(2\pi\sigma^2 y^3)$$

Média e variância de y :

$$E(y) = \frac{db(\theta)}{d\theta} = \mu; \quad \text{var}(y) = \frac{d^2b(\theta)}{d\theta^2} a(\phi) = \mu^3 \phi = \mu^3 \sigma^2$$

Observamos que, uma vez escolhida a distribuição de probabilidade, implicitamente são definidas: a função de variância $V(\mu)$, que é a parte da variância da resposta y que depende da média, e o parâmetro de dispersão ϕ , que não depende da média e é constante para os membros da família exponencial. Assim,

$$\text{var}(y) = \phi V(\mu).$$

3.2. Estrutura dos MLG

Considere um experimento com os dados da Tabela 3.1, com n respostas independentes y_i .

Tabela 3.1 - Dados para o Modelo.

x_1	x_2	...	x_k	y
x_{11}	x_{12}	...	x_{1k}	y_1
x_{21}	x_{22}	...	x_{2k}	y_2
.
.
.
x_{n1}	x_{n2}	...	x_{nk}	y_n

Temos então que:

1. Sejam y_1, y_2, \dots, y_n as variáveis de resposta com médias $\mu_1, \mu_2, \dots, \mu_n$.
2. A **distribuição de probabilidade** de y_i é um dos membros da **família exponencial**.
3. A porção sistemática do modelo é composta pelas variáveis de regressão x_1, x_2, \dots, x_k .
4. O modelo é construído com um preditor linear

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k.$$

5. A **função de ligação** $g(\mu_i)$ faz a ligação entre a média μ_i e o preditor linear. A função de ligação define a forma com que os efeitos sistemáticos de x_1, x_2, \dots, x_k são transmitidos para a média.

$$\eta_i = g(\mu_i) = (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik})$$

Note que, no caso particular da regressão linear, a função de ligação é a identidade, ou seja,

$$\eta_i = \mu_i$$

Observe-se que a média da resposta i é

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}) \quad (3.2)$$

A função de ligação é denominada **canônica** quando $\eta_i = \theta_i$

Segundo Myers e Montgomery (2002) a utilização da função de ligação canônica implica em algumas interessantes propriedades, mas isso não quer dizer que ela deva ser utilizada sempre. A sua escolha é conveniente porque não só simplifica as estimativas de máxima verossimilhança dos parâmetros do modelo, mas, também, o cálculo do intervalo de confiança para a média da resposta. Contudo, a conveniência não implica necessariamente em qualidade de ajuste do modelo, o que é mais importante.

Na Tabela 3.2 apresentamos as ligações canônicas da família exponencial.

Tabela 3.2 - Ligações Canônicas para os MLG.

Distribuição	Ligação Canônica
normal	$\eta = \mu$
Poisson	$\eta = \ln \mu$
binomial	$\eta = \ln(\pi / (1 - \pi))$
gama	$\eta = 1 / \mu$
normal inversa	$\eta = 1 / \mu^2$

A seleção da função de ligação em MLG pode ser vista como o equivalente da escolha da transformação da resposta no modelo linear de regressão. Entretanto, é importante ficar claro que a função de ligação transforma μ_i , a média de y_i , e não a resposta.

Como alternativa à função de ligação canônica, pode-se, similarmente à transformação da resposta, definir uma família de funções de ligação de potência

$$\eta = \mu^\lambda \quad \text{para } \lambda \neq 0 \quad \text{e}$$

$$\eta = \ln \mu, \quad \text{para } \lambda = 0$$

3.3. Estimação dos Parâmetros

A estimação dos parâmetros é feita através da maximização da função de log-verossimilhança:

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

Derivando a função L em relação aos parâmetros $\boldsymbol{\beta}$, tem-se

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \frac{dL}{d\theta_i} \frac{d\theta_i}{d\eta_i} \frac{\partial \eta_i}{\partial \boldsymbol{\beta}}.$$

Mas,

$$\frac{dL}{d\theta_i} = \sum_{i=1}^n \frac{1}{a(\phi)} \left[y_i - \frac{db(\theta_i)}{d\theta_i} \right] = \sum_{i=1}^n \frac{1}{a(\phi)} [y_i - \mu_i]$$

e

$$\frac{\partial \eta_i}{\partial \boldsymbol{\beta}} = \mathbf{x}_i.$$

onde \mathbf{x}_i é o vetor das variáveis regressoras para a resposta i .

Então

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \frac{y_i - \mu_i}{a(\phi)} \frac{d\theta_i}{d\eta_i} \mathbf{x}_i$$

Nos MLG, $a(\phi) = \phi = \text{constante}$.

Igualando-se a zero, obtêm-se as equações-escore:

$$\sum_{i=1}^n (y_i - \mu_i) \frac{d\theta_i}{d\eta_i} \mathbf{x}_i = 0 \quad (3.3)$$

Caso haja opção pela ligação canônica $\eta_i = \theta_i$, então, as equações-escore tornam-se mais simples:

$$\sum_{i=1}^n (y_i - \mu_i) \mathbf{x}_i = 0$$

Para estimar os parâmetros temos que resolver as Equações (3.3).

Os aspectos computacionais da solução das equações-escore podem ser encontrados em Myers, Montgomery e Vining (2002), apêndice A6. Estes autores demonstram que a estimação dos parâmetros, pela solução das equações-escore, pode ser obtida com o algoritmo descrito a seguir.

3.3.1. Algoritmo para Estimar os Parâmetros.

O algoritmo para obter a estimativa de máxima verossimilhança é denominado algoritmo dos *mínimos quadrados ponderados iterativo* (MQPI). No algoritmo fazemos uso da equação

$$\hat{\boldsymbol{\beta}}^{(m+1)} = (\mathbf{X}'\mathbf{W}^{(m)}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{(m)}\mathbf{z}^{(m)} \quad (3.4)$$

onde m refere-se à iteração; $\hat{\boldsymbol{\beta}}^{(m)}$ é a estimativa do vetor de parâmetros na iteração m .

$$\hat{\boldsymbol{\beta}}^{(m)} = \begin{bmatrix} \hat{\beta}_1^{(m)} \\ \hat{\beta}_2^{(m)} \\ \vdots \\ \hat{\beta}_k^{(m)} \end{bmatrix}$$

\mathbf{X} é a matriz dos valores das variáveis de regressão

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix}$$

\mathbf{W} é a matriz dos pesos

$$\mathbf{W}^{(m)} = \begin{bmatrix} w_{11}^{(m)} & 0 & \cdots & 0 \\ 0 & w_{22}^{(m)} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & w_{nn}^{(m)} \end{bmatrix}$$

cujos elementos da diagonal, $w_{ii}^{(m)}$, são dados por:

$$w_{ii}^{(m)} = \frac{1}{\text{var}(y_i)} \left(\left(\frac{\partial \mu_i}{\partial \eta_i} \right)^{(m)} \right)^2 \quad i = 1, 2, \dots, n \quad (3.5)$$

e \mathbf{z} é o vetor das variáveis de ajuste na m -ésima iteração

$$\mathbf{z}^{(m)} = \begin{bmatrix} z_1^{(m)} \\ z_2^{(m)} \\ \vdots \\ z_n^{(m)} \end{bmatrix}$$

com seus elementos dados por

$$z_i^{(m)} = \hat{\eta}_i^{(m)} + (y_i - \hat{\mu}_i^{(m)}) \left(\frac{\partial \eta_i}{\partial \mu_i} \right)^{(m)} \quad i = 1, 2, \dots, n. \quad (3.6)$$

Examinando a Equação (3.6), e lembrando que η e μ dependem de β através da Equação (3.2), vemos que a Equação (3.4) é recursiva.

Em cada iteração, o vetor da estimativa dos parâmetros $\hat{\beta}^{(m+1)}$ é calculado como uma função das estimativas anteriores $\hat{\beta}^{(m)}$.

Na experiência do autor desta tese, quando fornecemos uma solução inicial adequada, o algoritmo converge rapidamente. Mais adiante descrevemos como obter uma solução inicial adequada.

A Figura (3.1) mostra as dependências funcionais, e a Figura (3.2) mostra a seqüência de passos do algoritmo, o qual é detalhado mais adiante, na apresentação do Ciclo Iterativo.

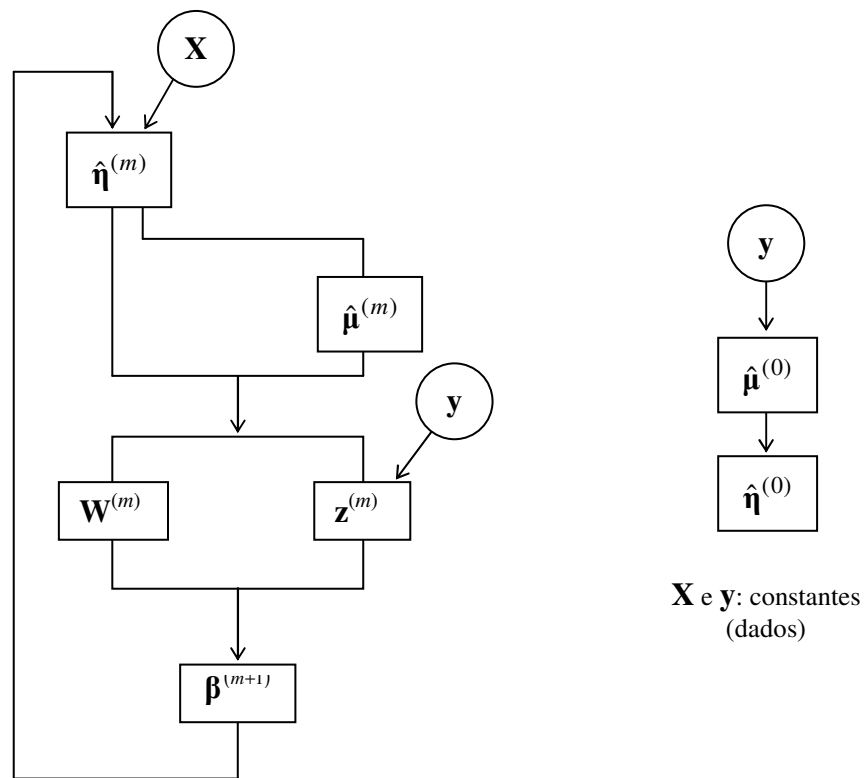


Figura 3.1 – Dependências Funcionais

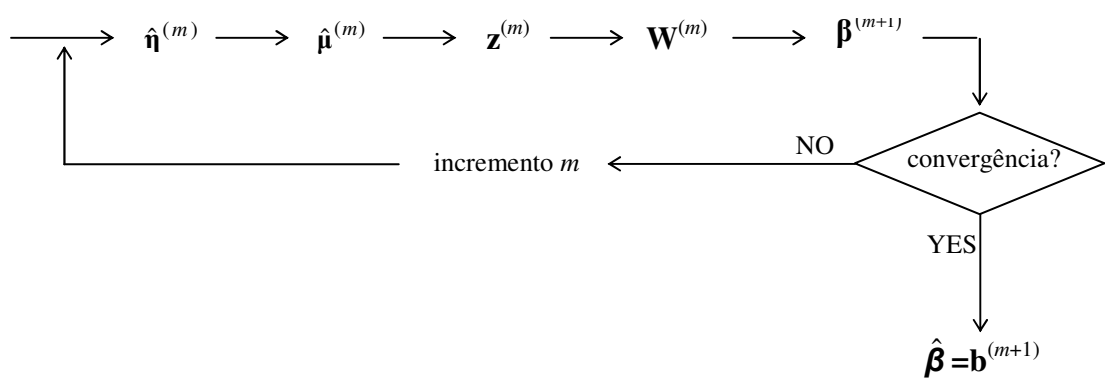


Figure 3.2 – Algoritmo MQPI

3.3.1.1. Ciclo Iterativo

Começamos com a descrição do ciclo iterativo; e então, mostramos como é feita a estimativa inicial de $\hat{\beta}^{(1)}$ para ser usada na primeira iteração. Seja a m -ésima iteração para estimar $\hat{\beta}^{(m)}$, o vetor de parâmetros, proveniente das estimativas de η , μ , z and W . Esta iteração termina com a atualização do estimador $\hat{\beta}^{(m+1)}$, que vai ser usado na próxima iteração.

Passo 1: Calcular o vetor $\hat{\eta}^{(m)}$:

$$\hat{\eta}^{(m)} = \begin{bmatrix} \hat{\eta}_1^{(m)} \\ \hat{\eta}_2^{(m)} \\ \vdots \\ \hat{\eta}_k^{(m)} \end{bmatrix} = \mathbf{X}\hat{\beta}^{(m)}$$

Passo 2: Calcular o vetor $\hat{\mu}^{(m)}$

$$\hat{\mu}^{(m)} = \begin{bmatrix} \hat{\mu}_1^{(m)} \\ \hat{\mu}_2^{(m)} \\ \vdots \\ \hat{\mu}_k^{(m)} \end{bmatrix}$$

onde cada $\hat{\mu}_i$ é obtida com a Equação (3.2)

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}).$$

Passo 3: Calcular o vetor z .

$$z_i^{(m)} = \hat{\eta}_i^{(m)} + (y_i - \hat{\mu}_i^{(m)}) \left(\frac{\partial \eta_i}{\partial \mu_i} \right)^{(m)} \quad i = 1, 2, \dots, n$$

Passo 4: Calcular os elementos da diagonal da matriz W , dados pela Equação (3.5):

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

Passo 5: Atualizar a estimativa do vetor $\hat{\beta}$:

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

Passo 6: Testar a convergência do algoritmo, verificando se alguma medida apropriada (previamente definida) da distância entre $\hat{\beta}^{(m+1)}$ e $\hat{\beta}^{(m)}$ é menor do que uma tolerância especificada. Se já houve convergência, FIM do algoritmo. Se não, incrementar m e retornar ao Passo 1, para a próxima iteração.

Não há uma maneira única e geral para obter a estimativa inicial $\hat{\beta}^{(1)}$. A mais comum — a qual sempre adotamos — é estimar $\hat{\beta}^{(1)}$ a partir de uma estimativa inicial $\hat{\mu}^{(0)}$ para o vetor da média μ . Uma estimativa inicial muito usada é $\hat{\mu}^{(0)} = \mathbf{y}$.

Note que no Ciclo Iterativo a obtenção de $\hat{\mu}^{(m)}$ constitui o Passo 2; portanto, a estimativa inicial $\hat{\beta}^{(1)}$ pode ser obtida a partir da estimativa inicial $\hat{\mu}^{(0)}$, executando os Passos 3, 4 e 5. Também é necessário calcular os valores de $\hat{\eta}_i^{(0)} = g(\hat{\mu}_i^{(0)})$, porque o Passo 3 requer uma estimativa de η .

Para um exemplo, ver Vieira e Epprecht (2004), que inclusive mostram como aplicar o algoritmo usando uma planilha *Excel*. São mostradas todas as iterações do algoritmo até a convergência.

3.4. Teste de Significância dos Coeficientes

Seja $\hat{\beta}$ o vetor dos parâmetros estimados. Para a família exponencial, pode-se demonstrar que a matriz de covariância de $\hat{\beta}$ é

$$\text{cov}(\hat{\beta}) = [\mathbf{X}'\mathbf{W}\mathbf{X}]^{-1} \sigma^2 \quad (3.7)$$

onde \mathbf{W} é a matriz diagonal cujos elementos são dados pela Equação (3.5).

A título de comparação, observe-se que, para o caso do modelo linear com MQ, a matriz de variância-covariância de $\hat{\beta}$ é $\text{cov}(\hat{\beta}) = [\mathbf{X}'\mathbf{X}]^{-1} \sigma^2$

Para esta seleção de parâmetros, em modelos lineares com distribuição normal, é comum usar a estatística

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\text{var}(\hat{\beta}_j)}}$$

que segue a distribuição *t*-Student. Alguns *softwares* fornecem esta estatística para testar parâmetros em modelos não normais, e fornecem o P-valor referente à distribuição *t* para a decisão de incluir ou não o parâmetro no modelo. Acontece que para modelos não normais a distribuição *t* é uma aproximação para a distribuição da estatística, aproximação esta que pode não ser boa mesmo com amostras grandes, com resultados enganosos (Lindsey 1997).

Entretanto, a estatística t_0 pode ser útil para indicar coeficientes significativos ou não significativos. Um valor elevado de t_0 , digamos maior do que três, é uma indicação de significância, em geral, para qualquer distribuição de probabilidade. Por outro lado, um valor pequeno de t_0 , digamos menor do que um, é uma indicação de não significância, em geral, para qualquer distribuição de probabilidade.

3.4.1. Deviance

Para testar a significância dos coeficientes, Atkinson e Riani (2000), Lindsey (1997) e McCullagh e Nelder (1989) recomendam usar a função desvio (*deviance*).

A *deviance* está para o método dos MLG como a soma dos quadrados dos resíduos está para o método dos MQ. A *deviance* de um modelo qualquer é definida como sendo o desvio deste modelo em relação ao modelo saturado, conforme a definição:

$$D = -2 \ln \left[\frac{L_{\text{Mod}}}{L_{\text{Sat}}} \right]$$

onde L_{Mod} é a função de máxima verossimilhança do modelo em questão e L_{Sat} é a função de máxima verossimilhança do modelo saturado, que é o modelo para o qual os valores ajustados $\hat{\mu}_i$ são iguais às respostas observadas y_i . Podemos então escrever

$$D(y_i, \hat{\mu}_i) = -2 \ln \left[\frac{L(y_i, \hat{\mu}_i)}{L(y_i, y_i)} \right] = -2 [\ln L(y_i, \hat{\mu}_i) - \ln L(y_i, y_i)] \quad (3.8)$$

As *deviances* para os membros da família exponencial são:

Distribuição normal:
$$\sum_{i=1}^n (y_i - \hat{\mu}_i)^2$$

Distribuição de Poisson:
$$2 \sum_{i=1}^n [y_i \ln(y_i / \hat{\mu}_i) - (y_i - \hat{\mu}_i)]$$

Distribuição binomial:

$$2 \sum_{i=1}^n \{y_i \ln(y_i / \hat{\mu}_i) + (m - y_i) \ln[(m - y_i) / (m - \hat{\mu}_i)]\}$$

Distribuição gama:
$$2 \sum_{i=1}^n [-\ln(y_i / \hat{\mu}_i) + (y_i - \hat{\mu}_i) / \hat{\mu}_i]$$

Distribuição normal inversa
$$\sum_{i=1}^n (y_i - \hat{\mu}_i)^2 / (\hat{\mu}_i y_i)$$

Observe-se que no caso da distribuição normal a *deviance* é a soma dos quadrados dos resíduos.

Lindsey (1997), pág. 212, demonstra que $D(y, \hat{\mu}_i)$ tem, assintoticamente, distribuição χ^2 com $n-p$ graus de liberdade, sendo n o número de observações e p , o número de parâmetros do modelo.

O procedimento recomendado é a análise de deviance, proposta, mas não detalhada, por McCullagh e Nelder (1989). Este procedimento é o equivalente, para os MLG, à análise de variância nos modelos baseados nos MQ. Como foi visto em MQ, usamos a soma dos quadrados dos resíduos para testar a significância dos estimadores. Em MLG vamos usar o teste de razão de log-verossimilhança, ou diferença de *deviance*, de dois modelos.

Lindsey (1997), pg. 214, recomenda o procedimento de análise de *deviance* com base na diferença de *deviances*, como descrito a seguir.

Suponha o Modelo 1 como sendo o saturado, o Modelo 2 com $p+1$ parâmetros $\beta_0, \dots, \beta_{p-1}, \beta_p$ e o Modelo 3, aninhado no Modelo 2, com p parâmetros $\beta_0, \dots, \beta_{p-1}$. Vimos que as *deviances* dos modelos 2 e 3 são:

$$D_2(y_i, \hat{\mu}_i^{(2)}) = -2[\ln L_2(y_i, \hat{\mu}_i^{(2)}) - \ln L_1(y_i, y_i)]$$

$$D_3(y_i, \hat{\mu}_i^{(3)}) = -2\{\ln L_3(y_i, \hat{\mu}_i^{(3)}) - \ln L_1(y_i, y_i)\}$$

Então, a diferença entre os desvios deviances dos modelos 3 e 2 é:

$$\begin{aligned} D_3(y_i, \hat{\mu}_i^{(3)}) - D_2(y_i, \hat{\mu}_i^{(2)}) &= -2[\ln L_3(y_i, \hat{\mu}_i^{(3)})] + 2[\ln L_2(y_i, \hat{\mu}_i^{(2)})] \\ &= -2\ln \left[\frac{L_3(y_i, \hat{\mu}_i^{(3)})}{L_2(y_i, \hat{\mu}_i^{(2)})} \right] \end{aligned}$$

Lindsey demonstra que esta diferença segue, aproximadamente, a distribuição χ_1^2 quando o modelo 3 é correto.

Ademais, $D_3(y_i, \hat{\mu}_i)$ segue, aproximadamente, a distribuição χ_{n-p}^2 quando o Modelo 3 é correto.

Portanto, se o Modelo 3 for correto o quociente abaixo segue a distribuição $F_{1,n-p}$.

$$F_0 = \frac{(D_3(y_i, \hat{\mu}_i) - D_2(y_i, \hat{\mu}_i))}{\frac{1}{D_3(y_i, \hat{\mu}_i)}} \sim F_{1,n-p} \quad (3.9)$$

Então, para uma seqüência de k modelos aninhados, podemos calcular as deviances: $D_j(y_i, \hat{\mu}_i)$ ($j = 1, 2, \dots, k$) e proceder aos testes de significância construindo uma tabela de ANODE (*ANalysis Of Deviance*) similarmente à tabela de ANOVA. Um exemplo, mais adiante, será ilustrativo.

Exemplo 3.1

Na Tabela 3.3 apresentamos os dados de um experimento fatorial 2^4 que foi gerado a partir do modelo

$$\mu = \frac{1}{0,040 + 0,008x_1 + 0,010x_2 + 0,012x_3 + 0,005x_1x_3 + 0,002x_2x_3} \quad (3.10)$$

Portanto, a função de ligação é a inversa.

A distribuição de probabilidade escolhida foi gama com parâmetro de dispersão $\phi = 0,01$. Para gerar aleatoriamente os dados, usamos a inversa da distribuição gama acumulada (INVGAMA) da planilha *Excel*, que usa a parametrização com α e β , onde $\alpha = 1/\phi = 100$ e $\beta_i = \mu_i/\alpha = \mu_i/100$.

Para isto necessitamos de 16 números aleatórios uniformes entre zero e um, os quais foram gerados no software *S-Plus* a partir da semente 435187. Portanto, temos que

$$y_i = \text{INVGAMA}\{\text{Uniforme } [0, 1]; 100; \beta_i\}$$

Tabela 3.3 - Dados do Experimento.

x_1	x_2	x_3	x_4	μ_i	Uniforme [0, 1]	β_i	y_i
-1	-1	-1	-1	58,83	0,8835	0,5882	65,92
1	-1	-1	-1	43,48	0,4619	0,4348	42,92
-1	1	-1	-1	30,31	0,4748	0,3030	30,02
1	1	-1	-1	25,65	0,3281	0,2564	24,44
-1	-1	1	-1	37,04	0,3201	0,3704	35,22
1	-1	1	-1	18,87	0,0260	0,1887	15,38
-1	1	1	-1	19,61	0,5019	0,1961	19,56
1	1	1	-1	12,99	0,6284	0,1299	13,38
-1	-1	-1	1	58,83	0,2063	0,5882	53,95
1	-1	-1	1	43,48	0,2466	0,4348	40,43
-1	1	-1	1	30,31	0,8101	0,3030	32,94
1	1	-1	1	25,65	0,2938	0,2564	24,20
-1	-1	1	1	37,04	0,8468	0,3704	40,83
1	-1	1	1	18,87	0,1717	0,1887	17,08
-1	1	1	1	19,61	0,1524	0,1961	17,61
1	1	1	1	12,99	0,1274	0,1299	11,53

Na Tabela 3.4, fornecida pelo software *ARC*, apresentamos as estimativas dos coeficientes, o erro-padrão da estimativa e o quociente entre a estimativa e o erro-padrão. Os quocientes entre as estimativas dos coeficientes de x_4 , x_1x_2 , x_1x_4 , x_2x_4 e x_3x_4 e seus respectivos erros-padrão têm valores inferiores a um, indicando que esses coeficientes não são significativos. Entretanto, esta indicação não é um teste formal. Vamos proceder à ANODE, que é um teste formal.

Tabela 3.4 - Estimativa dos Coeficientes e Erro Padrão.

Coefficiente	Estimativa	Erro Padrão	Estimativa/ErroPadrão
Constante	0,042024	0,001278	32,88
X1	0,010068	0,001251	8,05
X2	0,010259	0,001195	8,59
X3	0,013654	0,001275	10,71
X4	-0,000055	0,001147	-0,05
X1X2	-0,000036	0,001025	-0,04
X1X3	0,006192	0,001163	5,32
X1X4	-0,000041	0,000895	-0,05
X2X3	0,002367	0,001138	2,08
X2X4	-0,000113	0,000968	-0,12
X3X4	-0,000832	0,000956	-0,87

As variáveis x_1 , x_2 e x_3 são claramente significativas. Começando com o modelo com estas três variáveis vamos adicionando os termos que mais contribuem para o decréscimo no desvio *deviance* (coluna 1 da tabela 3.5). Esta seqüência e os respectivos deviances (coluna 3) são fornecidos pelo software *ARC*. Calculamos a diferença no deviance, ocasionada pelo acréscimo de cada termo (coluna 4), calculamos a estatística de teste F_0 (coluna 5) e o P-valor desta estatística (coluna 6).

Tabela 3.5 - Análise de Deviance (ANODE).

(1)	(2)	(3)	(4)	(5)	(6)
Termo Adicionado	Graus de Liberdade	Deviance	Diferença no Deviance	Estatística F_0	P-valor
X1-X2-X3	12	0.52145			
X1X3	11	0.12636	0.39509	34.39	0.00008
X2X3	10	0.07274	0.05362	7.37	0.02011
X3X4	9	0.05991	0.01283	1.93	0.19518
X2X4	8	0.05974	0.00017	0.02	0.88412
X1X2	7	0.05973	0.00002	0.00	0.96705
X4	6	0.05971	0.00001	0.00	0.97092
X1X4	5	0.05969	0.00002	0.00	0.96512

As interações x_1x_3 e x_2x_3 têm P-valor inferior a 0,05, sendo consideradas significativas. Reajustando o modelo com as variáveis x_1 , x_2 , x_3 , x_1x_3 e x_2x_3 , obtemos o modelo:

$$\hat{\mu} = \frac{1}{0,04200 + 0,0101x_1 + 0,0103x_2 + 0,0137x_3 + 0,0062x_1x_3 + 0,0024x_2x_3}$$

3.4.2. Estimação do Parâmetro de Dispersão

Em regressão linear estima-se σ^2 com

$$\hat{\sigma}^2 = \frac{SS_E}{n - p}.$$

Atkinson e Riani (2000), pág. 197, afirmam que o estimador análogo para os MLG é

$$\hat{\sigma}^2 = \hat{\phi} = \frac{D(y_i, \hat{\mu}_i)}{n - p}.$$

Na Seção 4.4.1. fornecemos os estimadores de máxima verossimilhança de ϕ , para cada uma das distribuições da família exponencial. A estimativa do parâmetro de dispersão para o Exemplo 3.1 é

$$\hat{\phi} = \frac{D(y_i, \hat{\mu}_i)}{n - p} = \frac{0,07274}{16 - 10} = 0,0121,$$

que está de acordo com o valor $\phi = 0,01$, com que os dados foram gerados.

McCullagh e Nelder (1989), pág. 296, afirmam que, para a distribuição gama, este estimador é extremamente sensível à erros de arredondamento, quando as observações são próximas de zero. Eles recomendam o estimador pelo método dos momentos

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

onde X^2 é a estatística de Pearson.

3.5. Adequação do Modelo

Como em regressão linear, nos MLG, resíduos também são utilizados para verificar a adequação do modelo.

Para os MLG, McCullagh e Nelder (1989), pág 39 definem vários tipos de resíduos. Entretanto, Pierce e Schafer (1989), citados por Lee e Nelder (1998), mostraram que, para o caso das distribuições da família exponencial, o resíduo *deviance* é o que mais se aproxima da distribuição normal e recomendam o resíduo *deviance* studentizado para verificar a adequação do modelo. Esses tipos de resíduos serão descritos a seguir.

3.5.1. Resíduo *Deviance*

Para cada resposta y_i pode-se definir a *deviance* $d_i = D_i(y_i, \hat{\mu}_i)$. Como nos MLG a *deviance* é usada como medida de discrepância, então cada unidade i contribui com uma quantidade d_i , de tal modo que

$$\sum_1^n d_i = D(y_i, \hat{\mu}_i).$$

Define-se então o resíduo *deviance* correspondente a cada resposta:

$$r_{Di} = \text{sign}(y_i - \hat{\mu}_i) \sqrt{d_i} \quad (3.11)$$

3.5.2. Resíduo *Deviance* Studentizado

Nos MLG a matriz chapéu é dada por (McCullagh e Nelder 1989, pág 37):

$$\mathbf{H} = \mathbf{W}^{\frac{1}{2}} \mathbf{X} (\mathbf{X}' \mathbf{W} \mathbf{X})^{-1} \mathbf{X}' \mathbf{W}^{\frac{1}{2}}$$

onde a matriz \mathbf{W} é matriz diagonal, com os elementos da diagonal principal dados por

$$w_i = \frac{1}{\text{var}(\mu_i)} \left(\frac{d\mu_i}{d\eta_i} \right)^2.$$

Os resíduos studentizados são então definidos como

$$r_i = \frac{r_{Di}}{\sqrt{\hat{\phi}^2 (1 - h_{ii})}} \quad (3.13)$$

onde h_{ii} é o i -ésimo elemento da diagonal da matriz \mathbf{H} e $\hat{\phi}^2 = D(y_i, \mu_i)/(n - p)$ é a estimativa do parâmetro de dispersão.

A adequação do modelo e a existência de observações atípicas podem ser observadas com o gráfico de probabilidade normal dos resíduos *deviance* studentizados.

Para o Exemplo 3.1, apresentamos na Figura 3.3 esse gráfico com envelope, fornecido pelo software *ARC*.

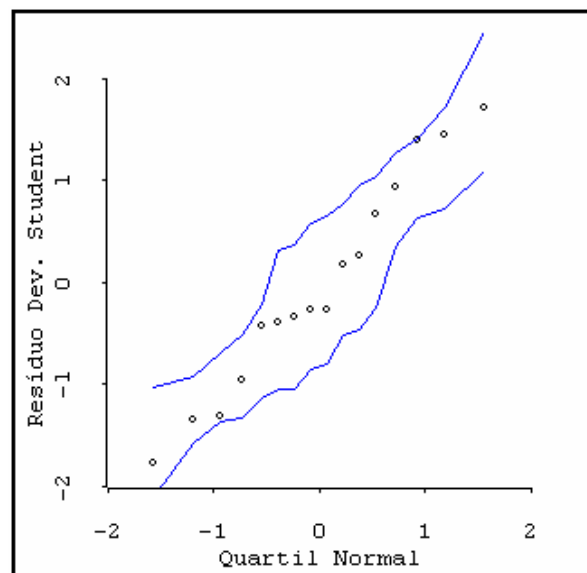


Figura 3.3 – Gráfico de Probabilidade Normal com Envelope

Não observamos pontos muito fora do envelope e não observamos pontos muito fora do alinhamento. Por conseguinte, não há indicação de observações atípicas nem de que o modelo seja inadequado.

3.5.3. Verificação da Adequação da Função de Ligação

A função de ligação é verificada através do gráfico dos resíduos studentizados *versus* valores ajustados. Na Figura 3.4 temos esse gráfico para o Exemplo 3.1, fornecido pelo software *ARC*.

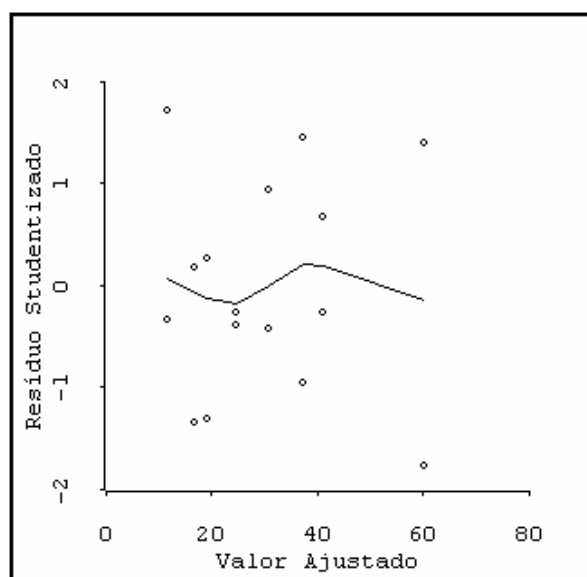


Figura 3.4 – Gráfico do Resíduo *Studentizado versus* Valor Ajustado

No gráfico em questão, os resíduos apresentam-se de forma desestruturada; isto é, eles não contêm nenhum padrão óbvio, apresentando-se aleatoriamente distribuídos. A linha resultante do amortecimento (*lowess*) é aproximadamente horizontal e próxima da reta horizontal de ordenada zero, indicando que a função de ligação é correta.

3.5.4. Verificação da Adequação da Função de Variância

A função de variância é verificada através do gráfico do valor absoluto dos resíduos studentizados *versus* valores ajustados. Na Figura 3.5 temos esse gráfico para o Exemplo 3.1, fornecido pelo software *ARC*.

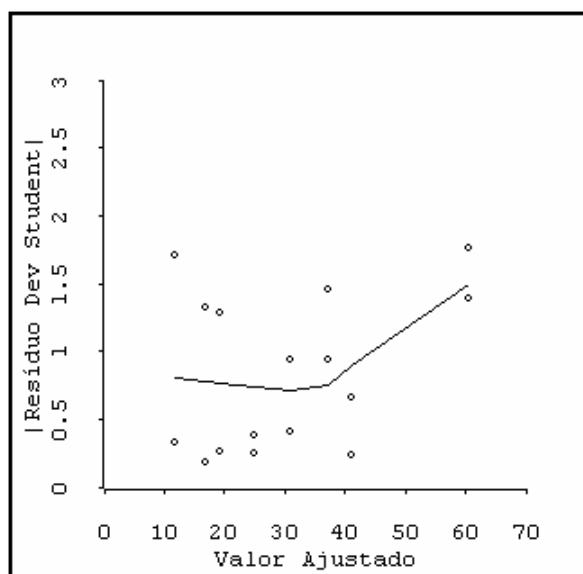


Figura 3.5 – Valor Absoluto do Resíduo *Studentizado Versus* Valor Ajustado

Como foi visto na Seção 3.2, a função de variância geralmente é definida como uma função de potência da média:

$$\text{var}(\mu) = \mu^\lambda$$

quando a linha *lowess* cresce sistematicamente, da esquerda para a direita, com o aumento da média, indica que deve-se usar um maior valor para λ do que o valor correspondente à distribuição que foi usada no modelo, e quando decresce sistematicamente, indica a adequação de um menor valor para λ . Para o gráfico em questão, a linha resultante do amortecimento (*lowess*) apresenta crescimento sistemático da esquerda para a direita a partir do 14º resíduo. Entretanto, observamos que ela é ocasionada essencialmente pelos dois maiores resíduos (em valor absoluto), numa região em que os pontos são mais esparsos e, portanto, a confiança na forma da linha *lowess* é menor. Portanto, não devemos considerar como indicação de função de variância incorreta.

3.5.5. Distância de Cook

A distância de Cook para os MLG é fornecida por Atkinson e Riani (2000):

$$D_i = \frac{r_{pi}^2 h_{ii}}{p \hat{\phi} (1 - h_{ii})^2} \quad (3.14)$$

onde p é o número de parâmetros e r_{pi}^2 é o resíduo de Pearson studentizado.

$$r'_{pi} = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\phi} \text{var}(\mu_i)(1 - h_{ii})}}$$

Na Figura 3.6, fornecido pelo software *ARC*, temos o gráfico da distância de Cook. Não há indicação de observação influente.

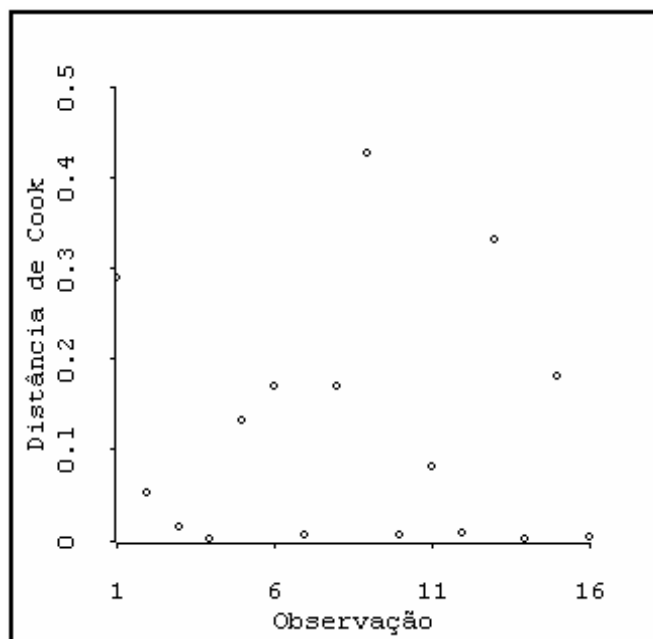


Figura 3.6 – Gráfico da Distância de Cook

3.5.6. Forward Search

Na Figura 3.7, apresentamos os resultados da FS da *deviance* (à esquerda) e da estimativa do parâmetro de dispersão σ (à direita). As duas estatísticas crescem sem saltos, o que indica não haver observações atípicas ou influentes mascaradas.

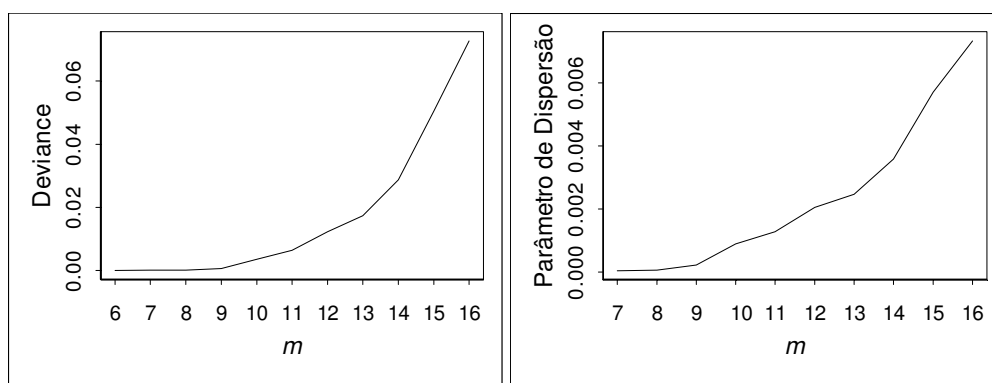


Figura 3.7 – Gráfico da FS para a *Deviance* e $\hat{\phi}$.

Na Figura 3.8, apresentamos o gráfico dos resultados da FS para os resíduos *deviance*. Durante toda a FS não há resíduos que se destaquem. Não há indicação de observações atípicas mascaradas.

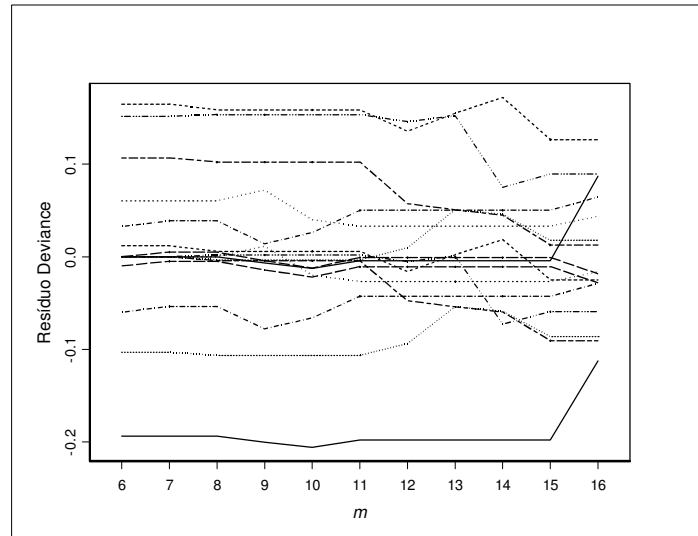


Figura 3.8 – Gráfico da FS para o Resíduo Deviance

3.6. Quase-Verossimilhança

Até aqui, para a definição dos modelos, consideramos unicamente as distribuições da família exponencial. Muitas vezes observa-se que não é adequado escolher um dos membros de tal família. Para superar este problema, Wedderburn (1974), citado por McCullagh e Nelder (1989), Cap. 9, define a função de quase-verossimilhança, que tem certas propriedades da função de log-verossimilhança. A idéia é usar os mínimos quadrados ponderados pela variância de y_i :

$$SQP = \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\text{var}(y_i)} = \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\phi V(\mu_i)}$$

Para minimizar SQP, derivamos a expressão acima em relação aos coeficientes:

$$\frac{dSQP}{d\beta} = \sum_{i=1}^n \frac{y_i - \mu_i}{\phi V(\mu_i)} \frac{d\mu_i}{d\beta}$$

Igualando a derivada a zero, obtemos as equações-escore

$$\sum_{i=1}^n \frac{(y_i - \mu_i)}{\phi V(\mu_i)} \frac{d\mu_i}{d\beta} = \mathbf{0} \quad (3.15)$$

Para estimar os coeficientes, temos que resolver as Equações (3.15). Mais adiante abordaremos a resolução destas equações.

A função de quase-verossimilhança para uma observação é

$$Q_i(\mu_i, y_i) = \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi V(t)} dt \quad (3.16)$$

e, para as n observações a função de quase-verossimilhança é

$$Q(\boldsymbol{\mu}, \mathbf{y}) = \sum_{i=1}^n Q_i(\mu_i, y_i) = \sum_{i=1}^n \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi V(t)} dt.$$

As derivadas dessa função em relação aos coeficientes reproduzirão a Equação (3.15). Portanto a função de quase-verossimilhança corresponde à função de log-verossimilhança. A diferença está em que quando se usa a função de quase-verossimilhança para estimar os coeficientes, apenas se define a relação da variância da resposta com a média da resposta, não sendo necessário definir uma distribuição de probabilidade. Para chegar à equação (3.15) não é preciso usar as propriedades dos membros da família exponencial. Em particular, se fizermos $\partial \mu_i / \partial \theta_i = V(\mu_i)$, obteremos a Equação 3.3.

Ademais, na Equação (3.15), a função-escore

$$U_i = \frac{y_i - \mu_i}{\phi V(\mu_i)},$$

tem as seguintes propriedades:

$$\begin{aligned} E(U_i) &= E\left[\frac{y_i - \mu_i}{\phi V(\mu_i)}\right] = 0, \\ \text{var}(U_i) &= \text{var}\left[\frac{y_i - \mu_i}{\phi V(\mu_i)}\right] = \frac{\text{var}(y_i)}{[\phi V(\mu_i)]^2} = \frac{\phi V(\mu_i)}{[\phi V(\mu_i)]^2} = \frac{1}{\phi V(\mu_i)} \\ -E\left(\frac{\partial U_i}{\partial \mu_i}\right) &= \frac{1}{\phi V(\mu_i)}, \end{aligned}$$

que são propriedades das funções-escore das derivadas da função de log-verossimilhança.

Como foi visto, a relação da função de variância com a média pode ser representada por uma função de potência

$$V(\mu_i) = \mu_i^t$$

Usando a Equação (3.16) temos que:

Para $t = 0$ a função de quase-verossimilhança é

$$Q_i(y_i, \mu_i) = -\frac{(y_i - \mu_i)^2}{2}$$

Para $t = 1$ a função quase-verossimilhança é

$$Q_i(y_i, \mu_i) = y_i \ln(\mu_i) - \mu_i$$

Para $t = 2$ a função quase-verossimilhança é

$$Q_i(y_i, \mu_i) = -\frac{y_i}{\mu_i} - \ln(\mu_i)$$

Para $t = 3$ a função quase-verossimilhança é

$$Q_i(y_i, \mu_i) = -\frac{y_i}{2\mu_i^2} + \frac{1}{\mu_i}$$

Para $t \neq 0, 1$ e 2 a função quase-verossimilhança é

$$Q_i(y_i, \mu_i) = \mu_i^{-t} \left(\frac{\mu_i y_i}{1-t} - \frac{\mu_i^2}{2-t} \right)$$

Para a função de variância $V(\mu) = \mu(1-\mu)$ a função quase-verossimilhança é

$$Q_i(y_i, \mu_i) = y \ln \left(\frac{\mu_i}{1-\mu_i} \right) + \ln(1-\mu_i)$$

Para $t = 0, 1, 2$ e 3 , as funções de quase-verossimilhança correspondem, respectivamente, às funções de log-verossimilhança das distribuições normal, de Poisson, gama e normal inversa. Para $V(\mu) = \mu(1-\mu)$, que é a função de variância da binomial, a função de quase-verossimilhança corresponde a função de log-verossimilhança da binomial.

3.6.1. Estimação dos Coeficientes

Como acabamos de ver, nos casos acima, a maximização da função de quase-verossimilhança produz as mesmas estimativas dos coeficientes, portanto, pode-se usar o mesmo algoritmo de estimativa dos coeficientes, visto na Seção 3.3.1.1.

Para outros valores da função de variância, que não são iguais aos da família exponencial, McCullagh e Nelder (1989), pág. 327, descrevem um algoritmo iterativo, similar ao algoritmo visto na Seção 3.3.1.1, e fornecem a matriz de variância-covariância dos estimadores dos coeficientes:

$$\text{cov}(\hat{\beta}) \cong [\mathbf{D}'\mathbf{W}\mathbf{D}]^{-1} \sigma^2 \quad (3.17)$$

onde \mathbf{D} é uma matriz $n \times p$, cujos elementos são $\partial \mu_i / \partial \beta_j$.

3.6.2. Estimação do Parâmetro de Dispersão

Segundo McCullagh e Nelder (1989), pág. 328, a estimativa do parâmetro de dispersão é

$$\tilde{\phi} = \frac{1}{n-p} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} = \frac{X^2}{n-p} \quad (3.18)$$

onde X^2 é a estatística de Pearson.

3.6.3. Significância dos Coeficientes

Para testar a significância dos coeficientes tem-se a estatística *quase-deviance*.

A *quase-deviance* está para a modelagem pela função de *quase-verossimilhança* como a *deviance* para os MLG. Por analogia, a *quase-deviance* de um modelo qualquer é definida como sendo o desvio deste modelo em relação ao modelo saturado, conforme a definição:

$$D_i(y_i, \hat{\mu}_i) = -2\phi[Q_i(y_i, \hat{\mu}_i) - Q_i(y_i, y_i)] = -2\phi[Q_i(y_i, \hat{\mu}_i)] = 2 \int_{\mu_i}^{y_i} \frac{y_i - \hat{\mu}_i}{V(\hat{\mu}_i)} \quad (3.19)$$

onde $Q_i(y_i, \hat{\mu}_i)$ é a função de máxima *quase-verossimilhança* do modelo em questão e $Q_i(y_i, y_i)$ é a função de máxima *quase-verossimilhança* do modelo saturado, que é o modelo para o qual os valores ajustados $\hat{\mu}_i$ são iguais às respostas observadas y_i .

O procedimento recomendado é a análise de *quase-deviance*. Este procedimento é o equivalente à análise de *deviance* para os modelos baseados nos MLG, como foi visto na Seção 3.4.1.

3.6.4. Adequação do Modelo

Como nos MLG, resíduos são utilizados para verificar a adequação do modelo. Para o método da *quase-verossimilhança*, Lee e Nelder (1998), recomendam utilizar o resíduo *deviance* studentizado, apresentado na Seção 3.5.2:

$$r_i = \frac{r_{Di}}{\sqrt{\hat{\phi}^2(1-h_{ii})}}$$

Os gráficos para verificar a adequação dos modelos são os apresentados na Seção 3.5.

Exemplo 3.1 (cont.)

Neste exemplo os dados foram gerados a partir do modelo

$$\mu = \frac{1}{0,040 + 0,008x_1 + 0,010x_2 + 0,012x_3 + 0,005x_1x_3 + 0,002x_2x_3}$$

A distribuição de probabilidade escolhida foi a distribuição gama com parâmetro de dispersão $\phi = 0,01$, cuja função de variância é $V(\mu) = \mu^2$. O modelo ajustado foi

$$\hat{\mu} = \frac{1}{0,0420 + 0,0101x_1 + 0,0103x_2 + 0,0137x_3 + 0,0062x_1x_3 + 0,0024x_2x_3}$$

Como ilustração, vamos usar o mesmo modelo e, para estimar os parâmetros, a maximização da função de quase-verossimilhança, com a função de variância $V(\mu) = \mu$.

O software *S-Plus* forneceu o seguinte resultado para o modelo da média:

$$\hat{\mu} = \frac{1}{0,0422 + 0,0102x_1 + 0,0104x_2 + 0,0139x_3 + 0,0064x_1x_3 + 0,0027x_2x_3}$$

Observamos que os dois modelos são similares. A modelagem com as funções de variância μ^2 e μ , resultaram, praticamente, na mesma estimativa dos coeficientes.

A estatística de Pearson é $X^2 = 2,3870$ e a estimativa do parâmetro de dispersão é

$$\hat{\phi} = \frac{X^2}{n - p} = \frac{2,3870}{16 - 6} = 0,2387$$

que é mais que o dobro do valor “real” 0,01. Isto se deve à função de variância adotada que foi μ e não a real μ^2 .

Na Figura 3.9 apresentamos o gráfico dos resíduos *deviance* studentizados *versus* valores ajustados e o gráfico do valor absoluto dos resíduos de *deviance* studentizados *versus* valores ajustados, ambos fornecidos pelo software *ARC*.

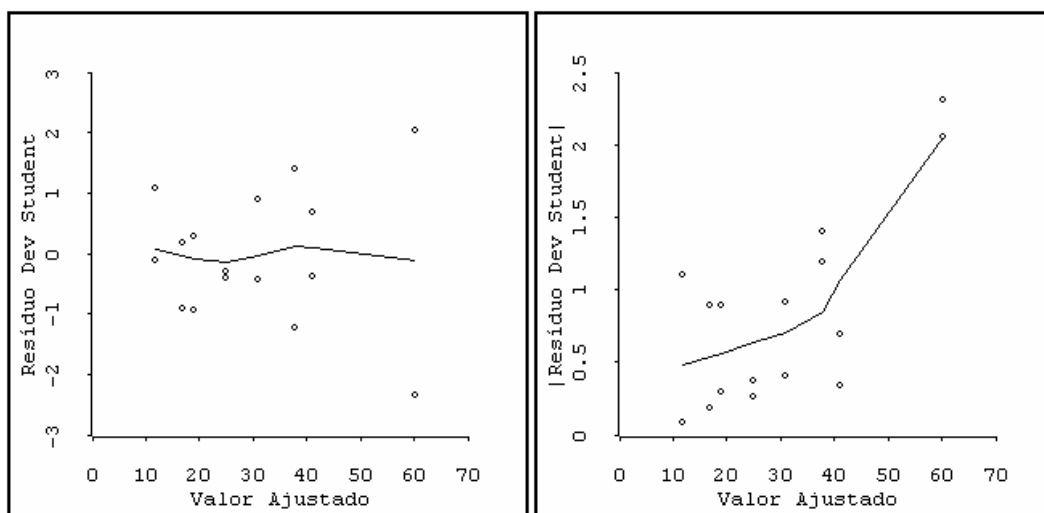


Figura 3.9 - Gráfico dos Resíduos *Deviance Studentizados versus* Valores Ajustados (à esquerda) e Gráfico Valor Absoluto do Resíduo de *Deviance Studentizados Versus* Valor Ajustado (à direita).

No gráfico à esquerda os resíduos apresentam-se de forma desestruturada; isto é, eles não contêm nenhum padrão evidente, apresentando-se aleatoriamente distribuídos. A linha resultante do amortecimento (*lowess*) é aproximadamente horizontal e próxima da reta horizontal de ordenada zero, indicando que a função de ligação é adequada.

No gráfico à direita a linha resultante do amortecimento (*lowess*) apresenta crescimento da esquerda para a direita. Isto indica que o expoente da média na função de variância deve ser aumentado. Isto se deve à função de variância do modelo, $V(\mu) = \mu$, que não é a “real”, $V(\mu) = \mu^2$.

3.7. Quase-Verossimilhança Estendida

A função de quase-verossimilhança não permite representar os casos em que o parâmetro de dispersão ϕ_i varia conforme cada tratamento. Para esses casos, e McCullagh e Nelder (1989) propõem a função de quase-verossimilhança estendida (*extended quasi-likelihood*):

$$-2Q^+ = \sum_{i=1}^n \left\{ \frac{d_i}{\phi_i} + \ln[2\pi\phi_i V(y_i)] \right\} \quad (3.21)$$

onde d_i é o quase-deviance, dado por

$$d_i = 2 \int_{\mu_i}^{y_i} \frac{y_i - t}{V(t)} dt .$$

Na verdade, a utilização da QVE pode ser vista como um artifício para os casos em que a função de variância não “explica” completamente a variabilidade da resposta, quando então o parâmetro de dispersão não é constante para cada tratamento (como nos MLG), mas depende dos fatores x_{ij} . Um modelo para a dispersão é então construído para estabelecer esta relação de dependência. Esta aplicação de QVE será vista no Capítulo 5 na modelagem conjunta da média e da variância.

3.7.1. Estimação dos Coeficientes

A maximização da função de quase-verossimilhança estendida (QVE) com relação a β será obtida com os estimadores de quase-verossimilhança (QV), com pesos $1/\phi_i$, cujas equações-escore são

$$\frac{\partial Q^+}{\partial \beta} = \sum_{i=1}^n \frac{(y_i - \mu_i)}{\phi_i V(\mu_i)} \frac{\partial \mu_i}{\partial \beta} = \mathbf{0}$$

ou ainda:

$$\sum_{i=1}^n \frac{y_i - \mu_i}{\phi_i V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} \mathbf{x}_i = \mathbf{0} \quad (3.22)$$

As equações-escore de QVE são as mesmas equações de QV com pesos conhecidos $1/\phi_i$. Portanto, sabendo-se o valor dos pesos, para estimar os coeficientes, procede-se da mesma forma que em QV.

Na Seção 3.6.1. vimos que, para as distribuições da família exponencial, a maximização da função de QV e os MLG produzem as mesmas estimativas dos coeficientes, portanto, pode-se usar o mesmo algoritmo de estimação dos coeficientes, visto na Seção 3.3.1.1.

Como as equações-escore de QVE são as mesmas equações de quase-verossimilhança (QV) com pesos conhecidos $1/\phi_i$, podemos usar o mesmo algoritmo, visto na Seção 3.3.1.1, para resolver as Equações (3.22).

A única diferença está na Equação (3.5) que passa a ser

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

3.7.2. Estimação do Parâmetro de Dispersão

Na quase-verossimilhança estendida o parâmetro de dispersão ϕ_i é diferente para cada resposta, e é estimado com um modelo construído para a dispersão, o que será vista na Seção 5.4 na modelagem conjunta da média e da variância.

3.7.3. Significância dos Coeficientes

McCullagh e Nelder (1989), pág. 351, afirmam que a função de quase-verossimilhança estendida tem as mesmas propriedades que a função de quase-verossimilhança. Portanto, para testar a significância dos coeficientes adota-se o mesmo procedimento descrito na Seção 3.6.2, com função de quase-verossimilhança estendida no lugar da função de quase-verossimilhança.

3.7.4. Adequação do Modelo

Como nos MLG, resíduos são utilizados para verificar a adequação do modelo. Para o método da quase-verossimilhança Lee e Nelder (1998) recomendam os resíduos *deviance* studentizados, apresentado na Seção 3.5.2. Os gráficos para verificar a adequação são os apresentados nas seções 3.5.2, 3.5.3, 3.5.4, 3.5.5 e 3.5.6.

3.8. Máxima verossimilhança restrita.

Em experimentos altamente fracionados, nos quais o número de dados é pequeno em relação ao número de parâmetros do modelo para a média, os estimadores dos parâmetros do modelo de dispersão podem ter vieses acentuados, devido à escassez de graus de liberdade para modelagem da dispersão (Lee e Nelder, 1998). Neste caso, procedimentos de ajuste devem ser adotados para o modelo da média. Lee e Nelder citam vários tipos de ajuste e propõem uso da técnica de máxima verossimilhança restrita (MVR), com a maximização da seguinte função de verossimilhança

$$-2Q_B = \sum_{i=1}^n \left\{ \frac{d_i^*}{\phi_i} + \ln[2\pi\phi_i V(y_i)] \right\} \quad (3.23)$$

onde $d_i^* = d_i/(1-h_i)$ e h_i , $i = 1, 2, \dots, n$ são os elementos da diagonal da matriz chapéu (*hat matrix*) \mathbf{H} , que é a matriz de projeção dos valores ajustados sobre os valores observados para a média, ou seja: $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$.

No Capítulo 5 veremos uma aplicação de máxima verossimilhança restrita.